

*Nodos de origen, proceso y resultado
de IBM SPSS Modeler 17*

IBM

Nota

Antes de utilizar esta información y el producto que soporta, lea la información que se incluye en el apartado “Avisos” en la página 383.

Información de producto

Esta edición se aplica a la versión 17, release 0, modificación 0 de IBM(r) SPSS(r) Modeler y a todos los releases y modificaciones posteriores mientras no se indique lo contrario en nuevas ediciones.

Contenido

Prefacio vii

Capítulo 1. Acerca de IBM SPSS Modeler 1

Productos IBM SPSS Modeler	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
IBM SPSS Modeler ServerAdaptadores paraIBM SPSS Collaboration and Deployment Services	2
Ediciones de IBM SPSS Modeler	2
Documentación de IBM SPSS Modeler	3
Documentación de SPSS Modeler Professional	3
Documentación de SPSS Modeler Premium	4
Ejemplos de aplicaciones	5
Carpeta Demos	5

Capítulo 2. Nodos de origen 7

Resumen	7
Configuración de almacenamiento de campos y formato	9
Almacenamiento de listas y niveles de medición asociados	11
Nodo origen Analytic Server	12
Selección de un origen de datos	13
Modificación de credenciales	13
Nodos soportados	13
Nodo de origen Base de datos	17
Configuración de las opciones de nodos Base de datos	19
Adición de una conexión de base de datos	20
Especificación de valores predefinidos para una conexión de base de datos	20
Selección de una tabla de la base de datos	23
Realización de consultas a la base de datos	24
Nodo Archivo variable	25
Opciones de configuración para el nodo Archivo variable.	26
Importación de los datos geoespaciales al nodo Archivo variable.	28
Nodo Archivo fijo	29
Configuración de opciones para el nodo Archivo fijo	29
Data Collection Nodo	31
Opciones de archivo de importación de Data Collection	31
Propiedades de metadatos de importación de IBM SPSS Data Collection	34
Cadena de conexión a la base de datos	35
Propiedades avanzadas	35
Importación de conjuntos de respuestas múltiples Notas de importación de columna de IBM SPSS Data Collection	35
Nodo de origen de IBM Cognos BI	36

Iconos de objeto de cognos	36
Importación de datos de Cognos	37
Importación de informes de Cognos	38
Conexiones de Cognos	38
Selección de ubicación de Cognos	39
Especificación de parámetros de datos o informes Nodo Origen de IBM Cognos TM1	39
Importación de datos de IBM Cognos TM1	40
Nodo de origen SAS	41
Configuración de opciones para el nodo de origen SAS	41
Nodo de origen de Excel	42
Nodo Origen XML	43
Selección entre múltiples elementos raíz	44
Eliminación de espacios no deseados de los datos de origen XML	44
Nodo Datos Usuario	44
Configuración de opciones para el nodo Datos Usuario.	45
Nodo Generar simulación	49
Configuración de las opciones del nodo Generar simulación.	50
Clonar campo	56
Ajustar detalles	56
Especificar parámetros.	57
Distribuciones	59
nodo Vista de datos	62
Opciones de configuración del nodo Vista de datos	63
Nodo de origen geoespacial	64
Opciones de configuración para el nodo Origen geoespacial	64
Nodo Enterprise View	65
Opciones de configuración del nodo Enterprise View	66
Conexiones de Enterprise View	66
Selección de la DPD	67
Selección de la tabla	67
Pestañas de nodos de origen comunes	67
Definición de niveles de medición en el nodo Origen	68
Filtrado de campos desde el nodo de origen	69

Capítulo 3. Nodos de operaciones con registros 71

Conceptos básicos de las operaciones con registros	71
Nodo Seleccionar	72
Nodo Muestrear	73
Opciones del nodo Muestrear	74
Configuración de Clúster y estratificación	75
Tamaños de muestra por estratos	77
Nodo Equilibrar	77
Opciones de configuración del nodo Equilibrar	78
Nodo Agregar	78
Opciones de configuración del nodo Agregar	79

Valores de optimización de la agregación	81	Derivación de múltiples campos	144
Nodo Adición de RFM	82	Opciones de configuración de derivación como fórmula	145
Opciones de configuración del nodo Adición de RFM.	82	Opciones de configuración de derivación como marca	147
Nodo Ordenar	83	Establecimiento de opciones de derivación nominal	147
Configuración de optimización de ordenación	84	Opciones de configuración de derivación como estado	148
Nodo Fundir	84	Opciones de configuración de derivación como recuento	148
Tipos de uniones	84	Opciones de configuración de derivación como condicional	149
Especificación de un método de fusión y claves	86	Nueva codificación de valores con el nodo Derivar	149
Selección de datos para uniones parciales	87	Nodo Rellenar	149
Especificación de condiciones para una fusión	87	Conversión del almacenamiento mediante el nodo Rellenar	150
Especificación de condiciones clasificadas para una fusión	88	Nodo Reclasificar	151
Filtrados de campos del nodo Fundir.	90	Opciones de configuración del nodo Reclasificar	151
Configuración del orden de entrada y etiquetado	90	Reclasificación de múltiples campos	152
Configuración de optimización de fusión	91	Almacenamiento y nivel de medición de los campos reclasificados.	152
Nodo Añadir	92	Nodo Anonimizar	153
Opciones de configuración de Añadir.	92	Opciones de configuración del nodo Anonimizar	154
nodo Distinguir	93	Anonimización de valores de campo	154
Configuración de optimización de Distinguir	95	Nodo Intervalos	155
Valores compuestos diferenciados	95	Opciones de configuración del nodo Intervalos	156
Nodo de Generación de análisis TS	98	Intervalos de ancho fijo	156
Opciones de campo de Generación de análisis TS	98	Cuantiles (igual frecuencia o igual suma)	157
Opciones del modelo de Generación de análisis TS	99	Clasificar casos	159
Criterios del modelizador experto de series temporales	100	Media/desviación estándar.	159
Criterios de suavizado exponencial de series temporales	100	Intervalos óptimos.	160
Criterios ARIMA de series temporales	101	Vista preliminar de intervalos generados	161
Funciones de transferencia	102	Nodo Análisis de RFM	161
Tratamiento de valores atípicos	103	Configuración del nodo Análisis de RFM	162
Opciones de despliegue de Generación de análisis TS	104	Intervalos del nodo Análisis de RFM	163
Nodo Cajas-espacio-tiempo.	105	Nodo Conjunto.	163
Definición de la densidad de una caja-espacio-tiempo	107	Configuración de nodo Conjunto	164
Capítulo 4. Nodos de operaciones con campos	109	Nodo Partición	165
Conceptos básicos de las operaciones con campos	109	Opciones del nodo Partición	165
Preparación automática de datos	111	Nodo Marcas	167
Pestaña Campos	113	Opciones de configuración del nodo Marcas	167
Pestaña Configuración	113	Nodo Reestructurar	167
Pestaña análisis.	118	Opciones de configuración del nodo Reestructurar	168
Generación de un nodo Derivar	124	Nodo Transponer	169
Nodo Tipo	125	Opciones de configuración del nodo Transponer	169
Niveles de medición	127	Nodo de intervalos de tiempo.	170
Conversión de datos continuos	130	Especificación de intervalos de tiempo	171
¿Qué es la instanciación?	130	Opciones de generación en intervalos de tiempo	172
Valores de campos.	131	Período de estimación	173
Definición de valores perdidos	135	Previsiones	173
Comprobación de valores de tipo.	136	Intervalos admitidos	175
Definición del rol de campo	136	Nodo Historial	182
Copia de atributos de tipo	137	Opciones de configuración del nodo Historial	182
Pestaña Configuración de Formato de campo	138	Nodo Reorg. campos	183
Cambio de nombre o filtrado de campos	139	Opciones de configuración de Reorg. campos	183
Opciones de configuración de filtrado	140	Nodo Intervalos de tiempo de AS	184
Nodo Derivar	142	Intervalo de tiempo de AS: Opciones de campos	184
Opciones básicas de configuración del nodo Derivar	143		

Intervalo de tiempo de AS: Opciones de generación	185
Nodo Reproyección	185
Establecer opciones para el nodo Reproyectar	186
Capítulo 5. Nodos Gráfico	187
Características comunes de nodos Gráfico	187
Aspectos estéticos, superposiciones, paneles y animación	188
Uso de la pestaña Resultados	189
Uso de la pestaña Anotaciones	190
Gráficos 3D	190
Nodo Tablero	192
Tablero Pestaña Básico	192
Tablero Pestaña Detallado	196
Tipos de visualización de Graphboard preincorporado disponibles.	198
Creación de visualizaciones de mapas	205
Ejemplos de Tableros	206
Pestaña Aspecto de Tablero.	216
Configuración de la ubicación de las plantillas, hojas de estilo y mapas	218
Administración de plantillas, hojas de estilo y archivos de mapas.	219
Conversión y distribución de mapas en formato shapefile	219
Conceptos claves de los mapas	220
Uso de la utilidad de conversión de mapas	221
Distribución de archivos de mapas	226
Nodo Gráfico	227
Pestaña de nodo Gráfico.	229
Pestaña Opciones de Gráfico	230
Pestaña Aspecto de Gráfico.	232
Utilización de un gráfico	232
Nodo G. múltiple	233
Pestaña Gráfico de G. múltiple	233
Pestaña Aspecto de G. múltiple	235
Utilización de un gráfico múltiple	235
Nodo Gráfico de tiempo.	236
Pestaña Gráfico de Tiempo	237
Pestaña Aspecto de Gráfico de Tiempo	238
Utilización de un gráfico de tiempo	238
Nodo Distribución.	239
Pestaña Gráfico de Distribución	239
Pestaña Aspecto de Distribución	240
Utilización de un nodo Distribución.	240
Nodo Histograma	243
Pestaña Gráfico de Histograma	243
Pestaña Opciones de Histograma	243
Pestaña Aspecto de Histograma	244
Uso de histogramas	244
Nodo Colección	245
Pestaña Gráfico de Colección	245
Pestaña Opciones de Colección	246
Pestaña Aspecto de Colección	246
Utilización de un gráfico de colección	247
Nodo Malla	248
Pestaña Gráfico de Malla	249
Pestaña Opciones de Malla	250
Pestaña Aspecto de Malla	252
Utilización de un gráfico de malla	252

Nodo Evaluación	256
Pestaña Gráfico de Evaluación.	260
Pestaña Opciones de Evaluación	262
Pestaña Aspecto de Evaluación	263
Lectura de los resultados de una evaluación del modelo	263
Uso de un diagrama de evaluación	264
Nodo Visualización de mapas	265
Pestaña de gráfico de Visualización de mapas	265
Pestaña Aspecto de visualización de correlación	269
Exploración de gráficos	269
Uso de tramos	270
Usar regiones	273
Uso de elementos con marca	275
Generación de nodos a partir de gráficos	276
Edición de visualizaciones	279
Reglas generales de edición de visualizaciones	280
Edición y formato de texto	281
Cambio de colores, tramas, estilos de guión y transparencia	281
Rotación y cambio de la forma y la relación de aspecto de los elementos de punto	282
Cambio de tamaño de los elementos gráficos	282
Especificación de márgenes y relleno	283
Formato de numeración	283
Cambio de la configuración del eje y la escala	284
Edición de categorías.	285
Cambio de la orientación de paneles	287
Transformación del sistema de coordenadas	287
Cambio de estadísticos y elementos gráficos	288
Cambio de la posición de la leyenda	290
Copia de una visualización y de datos de visualización	291
Atajos de teclado	291
Adición de títulos y notas al pie	291
Uso de hojas de estilo de gráficos	292
Impresión, almacenamiento, copia y exportación de gráficos	293
Capítulo 6. Nodos de resultado.	295
Conceptos básicos sobre nodos de resultados.	295
Administración de resultados	296
Visualización de resultados.	296
Publicar en Web	297
Visualización de resultados en un explorador HTML.	298
Exportación de resultados	298
Selección de casillas y columnas.	299
Nodo Tabla	299
Pestaña Configuración del nodo Tabla	299
Pestaña de formato del nodo Tabla	300
Pestaña Resultado del nodo de resultados.	300
Explorador de tablas	301
Nodo Matriz	302
Pestaña Configuración del nodo Matriz.	302
Pestaña Aspecto del nodo Matriz.	303
Explorador de resultados de nodo Matriz.	304
Nodo Análisis	304
Pestaña Análisis del nodo Análisis	305
Explorador de resultados de análisis.	306
Nodo Auditoría de datos	308

Pestaña Configuración del nodo Auditoría de datos	308
Pestaña Calidad de Auditoría de datos	309
Explorador de resultados de auditoría de datos	310
Nodo Transformación	315
Pestaña Opciones del nodo Transformar	315
Pestaña Resultado del nodo Transformar	316
Visor de resultados del nodo Transformar	316
Nodo Estadísticos	318
Pestaña Configuración del nodo Estadísticos	318
Explorador de resultados de estadísticos	319
Nodo Medias	320
Comparación de medias para grupos independientes	320
Comparación de medias entre campos relacionados	321
Opciones del nodo Medias	321
Explorador de resultados del nodo Medias	321
Nodo Informe	323
Pestaña Plantilla del nodo Informe	323
Explorador de resultados del nodo Informe	324
Nodo Val. globales	324
Pestaña Configuración del nodo Val. globales	325
Nodo Ajustar simulación	325
Ajuste de distribuciones	326
Pestaña Valores del nodo Ajustar simulación	327
Nodo Evaluación de simulación	328
Pestaña Configuración del nodo Evaluación de simulación	328
Salida del nodo Evaluación de simulación	331
Aplicaciones de ayuda de IBM SPSS Statistics	336

Capítulo 7. Nodos de exportación . . . 339

Conceptos básicos sobre nodos de exportación	339
Nodo Exportar base de datos	340
Pestaña Exportar del nodo Base de datos	340
Opciones de fusión Exportar base de datos	341
Opciones de Esquema de exportación de base de datos	342
Opciones de índice de exportación de base de datos	345
Opciones avanzadas de exportación de base de datos	347
Programación del cargador masivo	348
nodo de exportación Archivo sin formato	355
Pestaña Exportar del nodo Archivo sin formato	355
Nodos de exportación IBM SPSS Data Collection	356
Nodos de exportación Analytic Server	356
Nodo Exportar de IBM Cognos BI	357
Conexión de Cognos	357
Conexión ODBC	358
Nodo Exportar de IBM Cognos TM1	359
Conexión con un cubo de IBM Cognos TM1 para exportar datos	360
Correlación de datos de IBM Cognos TM1 para exportación	360

Nodo Exportar SAS	361
Pestaña Exportar del nodo Exportar SAS	361
Nodo de exportación Excel	362
Pestaña Exportar del nodo Excel	362
Nodo de exportación XML	363
Escritura de datos XML	363
Opciones de correlación de registros en XML	364
Opciones de correlación de campos a XML	364
Presentación preliminar de correlación a XML	364

Capítulo 8. Nodos de IBM SPSS Statistics 365

Conceptos básicos de nodos de IBM SPSS Statistics	365
Nodo de archivos estadísticos	366
Nodo Transformación de Statistics	367
Pestaña Sintaxis del nodo Transformación de Statistics	367
Sintaxis permitida	368
Nodo Modelo de Statistics	370
Pestaña Modelo del nodo Modelos de Statistics	370
Nodo Modelo de Statistics - Resumen del nugget de modelo	370
Nodo Resultados de Statistics	371
Pestaña Sintaxis del nodo Resultados de Statistics	371
Pestaña Resultados del nodo Resultados de Statistics	372
Nodo Exportación de Statistics	373
Pestaña Exportar del nodo Exportar Statistics	373
Cambio de nombre o filtrado de campos para IBM SPSS Statistics	374

Capítulo 9. Supernodos 375

Conceptos básicos de supernodos	375
Tipos de supernodos	375
Supernodos de origen	375
Supernodos de proceso	375
Supernodos terminales	376
Creación de supernodos	376
Anidación de supernodos	377
Bloqueo de Supernodos	377
Bloqueo y desbloqueo de un supernodo	377
Edición de un supernodo bloqueado	378
Edición de Supernodos	378
Modificación de los tipos de Supernodos	378
Anotación y cambio de nombre de Supernodos	378
Parámetros de Supernodos	379
Supernodos y almacenamiento en caché	381
Supernodos y scripts	381
Almacenamiento y carga de supernodos	382

Avisos 383

Marcas comerciales	384
------------------------------	-----

Índice. 389

Prefacio

IBM® SPSS Modeler es el conjunto de programas de minería de datos de IBM Corp. orientado a las empresas. SPSS Modeler ayuda a las organizaciones a mejorar la relación con sus clientes y los ciudadanos a través de la comprensión profunda de los datos. Las organizaciones utilizan la comprensión que les ofrece SPSS Modeler para retener a los clientes más rentables, identificar las oportunidades de venta cruzada, atraer a nuevos clientes, detectar el fraude, reducir el riesgo y mejorar la prestación de servicios del gobierno.

La interfaz visual de SPSS Modeler invita a los usuarios a aplicar su experiencia empresarial específica, lo que deriva en modelos predictivos más eficaces y la reducción del tiempo necesario para encontrar soluciones. SPSS Modeler ofrece muchas técnicas de modelado tales como predicciones, clasificaciones, segmentación y algoritmos de detección de asociaciones. Una vez que se crean los modelos, IBM SPSS Modeler Solution Publisher permite su distribución en toda la empresa a los encargados de tomar las decisiones o a una base de datos.

Acerca de IBM Business Analytics

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que confían los encargados de la toma de decisiones para mejorar el rendimiento comercial. Un conjunto integral de inteligencia empresarial, análisis predictivo, rendimiento financiero y gestión de estrategias y aplicaciones de análisis que ofrece una perspectiva clara, inmediata e interactiva del rendimiento actual y la capacidad de predecir resultados futuros. En combinación con extensas soluciones sectoriales, prácticas probadas y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir el máximo de productividad, automatizar las decisiones de forma fiable y alcanzar mejores resultados.

Como parte de esta familia, el software de análisis predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para lograr mejores resultados comerciales. Los clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología IBM SPSS como ventaja ante la competencia para atraer, retener y hacer crecer a los clientes, reduciendo al mismo tiempo el fraude y el riesgo. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas, capaces de dirigir y automatizar decisiones para alcanzar los objetivos comerciales y lograr una ventaja considerable sobre la competencia. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

Asistencia técnica

Hay asistencia técnica disponible para los clientes de mantenimiento. Los clientes podrán ponerse en contacto con el servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en los entornos de hardware admitidos. Para ponerse en contacto con el servicio de asistencia técnica, consulte el sitio web de IBM Corp. en <http://www.ibm.com/support>. Tenga a mano su acuerdo de asistencia y esté preparado para identificarse a sí mismo y a su organización al solicitar ayuda.

Capítulo 1. Acerca de IBM SPSS Modeler

IBM SPSS Modeler es un conjunto de herramientas de minería de datos que permite desarrollar rápidamente modelos predictivos mediante técnicas empresariales y desplegarlos en operaciones empresariales para mejorar la toma de decisiones. Con un diseño que sigue el modelo CRISP-DM, estándar del sector, IBM SPSS Modeler admite el proceso completo de minería de datos, desde los propios datos hasta obtener los mejores resultados empresariales.

IBM SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico. Los métodos disponibles en la paleta de modelado permiten derivar nueva información procedente de los datos y desarrollar modelos predictivos. Cada método tiene ciertos puntos fuertes y es más adecuado para determinados tipos de problemas.

SPSS Modeler puede adquirirse como producto independiente o utilizarse como cliente junto con SPSS Modeler Server. También hay disponible cierto número de opciones adicionales que se resumen en las siguientes secciones. Si desea obtener más información, consulte <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Productos IBM SPSS Modeler

La familia de productos IBM SPSS Modeler y su software asociado se componen de lo siguiente:

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server adaptadores para IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler es una versión con todas las funcionalidades del producto que puede instalar y ejecutar en su ordenador personal. Puede ejecutar SPSS Modeler en modo local como un producto independiente o utilizarla en modo distribuido junto con IBM SPSS Modeler Server para mejorar el rendimiento a la hora de trabajar con grandes conjuntos de datos.

Con SPSS Modeler, puede crear modelos predictivos precisos de forma rápida e intuitiva sin necesidad de programación. Mediante su exclusiva interfaz visual, podrá visualizar fácilmente el proceso de minería de datos. Con ayuda del análisis avanzado incrustado en el producto podrá detectar patrones y tendencias en sus datos que anteriormente estaban ocultos. Podrá modelar los resultados y comprender los factores que influyen en ellos, lo que le permitirá aprovechar oportunidades comerciales y mitigar los riesgos.

SPSS Modeler está disponible en dos ediciones: SPSS Modeler Professional y SPSS Modeler Premium. Consulte el tema “Ediciones de IBM SPSS Modeler” en la página 2 para obtener más información.

IBM SPSS Modeler Server

SPSS Modeler utiliza una arquitectura de cliente/servidor para distribuir peticiones de cliente para operaciones que requieren un uso intensivo de los recursos a un software de servidor de gran potencia, lo que proporciona un rendimiento más rápido con conjuntos de datos de mayor volumen.

SPSS Modeler Server es un producto con licencia independiente que se ejecuta de manera continua en modo de análisis distribuido en un host de servidor junto con una o más instalaciones de IBM SPSS

Modeler. De esta forma, SPSS Modeler Server ofrece un rendimiento mucho mayor en los conjuntos de datos de gran tamaño, ya que las operaciones que requieren mucha memoria se pueden realizar en el servidor sin tener que descargar datos en el equipo del cliente. IBM SPSS Modeler Server también proporciona soporte para las funciones de optimización de SQL y de modelado en la base de datos, lo que ofrece ventajas adicionales de rendimiento y automatización.

IBM SPSS Modeler Administration Console

Modeler Administration Console es una aplicación gráfica para administrar muchas de las opciones de configuración de SPSS Modeler Server, las cuales también pueden configurarse a través de un archivo de opciones. La aplicación proporciona una interfaz de usuario de la consola para supervisar y configurar las instalaciones de SPSS Modeler Server y está disponible de forma completamente gratuita para los clientes actuales de SPSS Modeler Server. La aplicación solamente se puede instalar en los ordenadores con Windows; sin embargo, puede administrar un servidor que esté instalado en cualquier plataforma compatible.

IBM SPSS Modeler Batch

Aunque la minería de datos suele ser un proceso interactivo, también es posible ejecutar SPSS Modeler desde una línea de comandos, sin necesidad de la interfaz gráfica del usuario. Por ejemplo, puede que tenga tareas repetitivas o cuya ejecución sea de larga duración que quiera realizar sin intervención del usuario. SPSS Modeler Batch es una versión especial del producto que proporciona soporte para todas las prestaciones de análisis de SPSS Modeler sin acceso a la interfaz de usuario habitual. SPSS Modeler Server debe utilizar SPSS Modeler Batch.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher es una herramienta que le permite crear una versión empaquetada de una ruta de SPSS Modeler que se puede ejecutar en un motor de tiempo de ejecución externo o incrustado en una aplicación externa. De este modo, podrá publicar y desplegar rutas completas de SPSS Modeler para utilizarlas en entornos que no tengan SPSS Modeler instalado. SPSS Modeler Solution Publisher se distribuye como parte del servicio IBM SPSS Collaboration and Deployment Services - Puntuación, para el que se necesita una licencia independiente. Con esta licencia, recibirá Entorno de ejecución de SPSS Modeler Solution Publisher, que le permite ejecutar las rutas publicadas.

Para obtener más información sobre SPSS Modeler Solution Publisher, consulte la documentación de IBM SPSS Collaboration and Deployment Services. El Knowledge Center de IBM SPSS Collaboration and Deployment Services contiene secciones denominadas "IBM SPSS Modeler Solution Publisher" e "IBM SPSS Analytics Toolkit".

IBM SPSS Modeler Server Adaptadores para IBM SPSS Collaboration and Deployment Services

Tiene a su disposición un determinado número de adaptadores para IBM SPSS Collaboration and Deployment Services que permiten que SPSS Modeler y SPSS Modeler Server interactúen con un repositorio de IBM SPSS Collaboration and Deployment Services. De este modo, varios usuarios podrán compartir una ruta de SPSS Modeler desplegada en el repositorio, o bien se podrá acceder a ella desde la aplicación cliente de baja intensidad IBM SPSS Modeler Advantage. Debe instalar el adaptador en el sistema donde se aloje el repositorio.

Ediciones de IBM SPSS Modeler

SPSS Modeler está disponible en las siguientes ediciones.

SPSS Modeler Professional

SPSS Modeler Professional proporciona todas las herramientas que necesita para trabajar con la mayoría de los tipos de datos estructurados, como los comportamientos e interacciones registrados en los sistemas

de CRM, datos demográficos, comportamientos de compra y datos de ventas.

SPSS Modeler Premium

SPSS Modeler Premium es un producto con licencia independiente que amplía SPSS Modeler Professional para poder trabajar con datos especializados, como los utilizados para el análisis de entidades o las redes sociales, así como con datos de texto no estructurados. SPSS Modeler Premium está formado por los siguientes componentes:

IBM SPSS Modeler Entity Analytics incorpora una dimensión adicional al análisis predictivo de IBM SPSS Modeler predictive analytics. Mientras que el análisis predictivo trata de predecir comportamientos futuros a partir de datos del pasado, el análisis de entidades se centra en mejorar la coherencia de los datos actuales mediante la resolución de conflictos de identidades dentro de los propios registros. La identidad de un individuo, una organización, un objeto o cualquier otra entidad puede estar expuesta a ambigüedades. La resolución de identidades puede ser vital en diversos campos, entre los que se incluyen la gestión de la relación con el cliente, la detección de fraudes, la lucha contra el blanqueo de dinero y la seguridad nacional e internacional.

IBM SPSS Modeler Social Network Analysis transforma la información sobre relaciones en campos que caracterizan el comportamiento social de individuos y grupos. Mediante el uso de datos que describen las relaciones subyacentes de las redes sociales, IBM SPSS Modeler Social Network Analysis identifica a los líderes sociales que influyen en el comportamiento de otros en la red. Además, puede determinar qué personas se ven más afectadas por otros participantes de la red. Al combinar estos resultados con otras medidas, puede crear perfiles completos de individuos en los que basar sus modelos predictivos. Los modelos que incluyan esta información social tendrán un mejor rendimiento que los modelos que no la incluyan.

IBM SPSS Modeler Text Analytics utiliza tecnologías de lingüística avanzada y Procesamiento del lenguaje natural (PLN) para procesar con rapidez una gran variedad de datos de texto sin estructurar, extraer y organizar los conceptos clave y agruparlos en categorías. Las categorías y conceptos extraídos se pueden combinar con los datos estructurados existentes, como pueden ser datos demográficos, y se pueden aplicar para modelar utilizando el conjunto completo de herramientas de minería de datos de IBM SPSS Modeler para tomar decisiones mejores y más certeras.

Documentación de IBM SPSS Modeler

Tiene a su disposición documentación en formato de ayuda en línea desde el menú Ayuda de SPSS Modeler. Se incluye documentación para SPSS Modeler, SPSS Modeler Server y SPSS Modeler Solution Publisher, así como el Manual de aplicaciones y otros materiales de apoyo.

La documentación completa de cada producto (incluidas las instrucciones de instalación) en formato PDF está disponible en la carpeta *\Documentation* en cada DVD del producto. Los documentos de instalación también se pueden descargar de la web en <http://www.ibm.com/support/docview.wss?uid=swg27043831>.

la documentación en ambos formatos también está disponible en el Knowledge Center de SPSS Modeler, en http://www-01.ibm.com/support/knowledgecenter/SS3RA7_17.0.0.0.

Documentación de SPSS Modeler Professional

El conjunto de documentación de SPSS Modeler Professional (excluidas las instrucciones de instalación) es el siguiente.

- **IBM SPSS Modeler Manual del usuario.** Introducción general sobre cómo usar SPSS Modeler, incluyendo cómo crear rutas de datos, tratar valores perdidos, crear expresiones CLEM, trabajar con proyectos e informes y empaquetar rutas para su despliegue en IBM SPSS Collaboration and Deployment Services, Predictive Applications o IBM SPSS Modeler Advantage.

- **Nodos de origen, proceso y resultado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para leer, procesar y dar salida a datos en diferentes formatos. En la práctica, esto implica todos los nodos que no sean nodos de modelado.
- **Nodos de Modelado de IBM SPSS Modeler.** Descripciones de todos los nodos utilizados para crear modelos de minería de datos. IBM SPSS Modeler ofrece una gran variedad de métodos de modelado procedentes del aprendizaje automático, la inteligencia artificial y el estadístico.
- **Manual de algoritmos de IBM SPSS Modeler.** Descripciones de los fundamentos matemáticos de los métodos de modelado que se utilizan en IBM SPSS Modeler. Esta guía está disponible únicamente en formato PDF.
- **Manual de aplicaciones de IBM SPSS Modeler.** Los ejemplos de esta guía ofrecen introducciones breves y concisas a métodos y técnicas de modelado específicos. También tiene a su disposición una versión en línea de este manual en el menú Ayuda. Consulte el tema “Ejemplos de aplicaciones” en la página 5 para obtener más información.
- **IBM SPSS Modeler Automatización y scripts de Python.** Información sobre la automatización del sistema a través de scripts en Python, incluidas las propiedades que se pueden utilizar para manipular nodos y rutas.
- **IBM SPSS Modeler Manual de despliegue.** Información sobre la ejecución de rutas y escenarios de IBM SPSS Modeler como pasos en trabajos de procesamiento en IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **Guía del desarrollador de IBM SPSS Modeler CLEF.** CLEF ofrece la capacidad de integrar programas de otros fabricantes, como rutinas de procesamiento de datos o algoritmos de modelado como nodos en IBM SPSS Modeler.
- **Manual de minería interna de bases de datos de IBM SPSS Modeler.** Este manual incluye información sobre cómo utilizar la potencia de su base de datos, tanto para mejorar su rendimiento como para ampliar su oferta de capacidades analíticas a través de algoritmos de terceros.
- **Guía de administración de IBM SPSS Modeler Server y su rendimiento.** Información sobre la configuración y administración de IBM SPSS Modeler Server.
- **Manual del usuario de IBM SPSS Modeler Administration Console.** Información sobre cómo instalar y utilizar la interfaz de usuario de la consola para supervisar y configurar IBM SPSS Modeler Server. La consola se implementa como complemento de la aplicación Gestor de despliegue.
- **Manual CRISP-DM de IBM SPSS Modeler.** Manual que explica paso a paso cómo utilizar la metodología de CRISP-DM en la minería de datos con SPSS Modeler.
- **IBM SPSS Modeler Batch Manual del usuario.** Guía completa de cómo utilizar IBM SPSS Modeler en modo por lotes, incluida información detallada sobre la ejecución del modo por lotes y argumentos de línea de comandos. Esta guía está disponible únicamente en formato PDF.

Documentación de SPSS Modeler Premium

El conjunto de documentación de SPSS Modeler Premium (excluidas las instrucciones de instalación) es el siguiente.

- **IBM SPSS Modeler Entity Analytics Guía del usuario.** Información sobre cómo utilizar el análisis de entidades con SPSS Modeler, que cubre la instalación y configuración de repositorios, nodos de análisis de entidades y tareas administrativas.
- **IBM SPSS Modeler Social Network Analysis Guía del usuario.** Una guía para realizar análisis de redes sociales con SPSS Modeler, incluido el análisis de grupos y el análisis de difusión.
- **SPSS Modeler Text Analytics Manual del usuario.** Información sobre cómo utilizar el análisis de texto con SPSS Modeler, que cubre los nodos de minería de texto, programa interactivo, plantillas y otros recursos.

Ejemplos de aplicaciones

Mientras que las herramientas de minería de datos de SPSS Modeler pueden ayudar a resolver una amplia variedad de problemas organizativos y empresariales, los ejemplos de la aplicación ofrecen introducciones breves y adaptadas de técnicas y métodos de modelado específicos. Los conjuntos de datos utilizados aquí son mucho más pequeños que los enormes almacenes de datos gestionados por algunos analizadores de datos, pero los conceptos y métodos implicados deberían ser escalables a las aplicaciones reales.

Para acceder a los ejemplos pulsando **Ejemplos de aplicación** en el menú Ayuda de SPSS Modeler. Los archivos de datos y rutas de ejemplo se instalan en la carpeta *Demos* en el directorio de instalación del producto. Consulte el tema “Carpeta Demos” para obtener más información.

Ejemplos de modelado de bases de datos. Consulte los ejemplos que figuran en el *Manual de minería interna de bases de datos de IBM SPSS Modeler*.

Ejemplos de scripts. Consulte los ejemplos en la *Guía de scripts y automatización de IBM SPSS Modeler*.

Carpeta Demos

Los archivos de datos y rutas de ejemplo utilizados con los ejemplos de la aplicación se instalan en la carpeta *Demos* en el directorio de instalación del producto. También puede acceder a esta carpeta desde el grupo de programas IBM SPSS Modeler del menú Inicio de Windows o pulsando *Demos* en la lista de directorios recientes del cuadro de diálogo Abrir archivo.

Capítulo 2. Nodos de origen

Resumen

Los nodos de origen permiten importar los datos almacenados en distintos formatos, entre los que se incluyen archivos planos, IBM SPSS Statistics (.sav), SAS, Microsoft Excel y bases de datos relacionales compatibles con ODBC. También puede generar datos sintéticos mediante el nodo Datos Usuario.

La paleta Orígenes contiene los siguientes nodos:



El origen de Analytic Server permite ejecutar una ruta en un sistema de archivos distribuido de Hadoop (HDFS en sus siglas inglesas). La información de un origen de datos de Analytic Server puede proceder de diversos lugares, como archivos de texto y bases de datos. Consulte el tema “Nodo origen Analytic Server” en la página 12 para obtener más información.



El nodo Base de datos se puede utilizar para importar datos desde otros paquetes mediante ODBC (del inglés, Open Database Connectivity), incluidos Microsoft SQL Server, DB2, Oracle, etc. Consulte el tema “Nodo de origen Base de datos” en la página 17 para obtener más información.



El nodo Archivo variable lee datos desde los archivos de texto de campo libre, es decir, campos cuyos registros contienen un número constante de campos, pero un número variado de caracteres. Este nodo resulta también útil para los archivos con texto de cabecera de longitud fija y determinados tipos de anotaciones. Consulte el tema “Nodo Archivo variable” en la página 25 para obtener más información.



El nodo Archivo fijo importa datos desde archivos de texto de campo fijo; esto es, archivos cuyos campos no están delimitados pero empiezan en la misma posición y tienen una longitud fija. Los datos heredados o generados por la máquina se suelen almacenar en formato de campo fijo. Consulte el tema “Nodo Archivo fijo” en la página 29 para obtener más información.



El nodo Archivo de estadísticas lee los datos desde un formato de archivo .sav o .zsav que utiliza IBM SPSS Statistics y archivos caché guardados en IBM SPSS Modeler, que también puede utilizar el mismo formato.



El nodo IBM SPSS Data Collection importa datos de encuesta a partir de varios formatos utilizados por el software de investigación de mercado conforme al modelo de datos de IBM SPSS Data Collection. Se debe instalar la biblioteca de desarrollador de IBM SPSS Data Collection para utilizar este nodo. Consulte el tema “Data Collection Nodo” en la página 31 para obtener más información.



El nodo de origen de IBM Cognos BI importa datos desde las bases de datos de Cognos BI.



El nodo fuente de IBM Cognos TM1 importa datos de las bases de datos de Cognos TM1.



El nodo Archivo SAS importa datos SAS a IBM SPSS Modeler. Consulte el tema “Nodo de origen SAS” en la página 41 para obtener más información.



El nodo Excel importa datos de cualquier versión de Microsoft Excel. No es necesario un origen de datos ODBC. Consulte el tema “Nodo de origen de Excel” en la página 42 para obtener más información.



El nodo de origen XML importa datos en formato XML en la ruta. Puede importar un único archivo o todos los archivos en un directorio. Puede especificar opcionalmente un archivo de esquema para leer la estructura XML.



El nodo Datos de usuario proporciona una manera fácil de crear datos sintéticos, ya sea partiendo de cero o modificando los datos existentes. Esto resulta útil, por ejemplo, cuando desee crear un conjunto de datos de comprobación para el modelado. Consulte el tema “Nodo Datos Usuario” en la página 44 para obtener más información.



El nodo Generar simulación permite, de forma sencilla, generar datos simulados —ya sea a partir de cero utilizando distribuciones estadísticas especificadas por el usuario o de forma automática utilizando las distribuciones obtenidas de la ejecución de un nodo Ajustar simulación sobre datos históricos existentes. Esto es útil cuando se desea evaluar el resultado de un modelo predictivo en presencia de incertidumbre en las entradas de modelo.



El nodo Ver datos puede utilizarse para acceder a los orígenes de datos definidos en las vistas de datos analíticos de IBM SPSS Collaboration and Deployment Services. Una vista de datos analíticos define una interfaz estándar para acceder a los datos y asocia varios orígenes de datos físicos con dicha interfaz. Consulte el tema “nodo Vista de datos” en la página 62 para obtener más información.



Utilice el nodo de origen Geoespacial para introducir datos de mapas o espaciales en la sesión de minería de datos. Consulte el tema “Nodo de origen geoespacial” en la página 64 para obtener más información.



Nota: El nodo Enterprise View fue sustituido en IBM SPSS Modeler 16.0 por el nodo Vista de datos. Todavía se admite el nodo Enterprise View para rutas guardadas en ediciones anteriores. No obstante, al actualizar o crear nuevas rutas, recomendamos utilizar el nodo Vista de datos. El nodo Enterprise View crea una conexión a IBM SPSS Collaboration and Deployment Services Repository, lo que permite leer los datos de Enterprise View en una ruta y empaquetar un modelo en un escenario al que otros usuarios pueden acceder desde el repositorio. Consulte el tema “Nodo Enterprise View” en la página 65 para obtener más información.

Para iniciar una ruta, añada un nodo de origen al lienzo de rutas. A continuación, pulse dos veces en el nodo para abrir su cuadro de diálogo. Las distintas pestañas del cuadro de diálogo permiten leer datos, ver los campos y valores así como establecer distintas opciones, incluidos filtros, tipos de datos, rol del campo y comprobación de los valores perdidos.

Configuración de almacenamiento de campos y formato

Las opciones de la pestaña Datos para los nodos Archivo fijo, Archivo variable, Origen XML y Datos usuario permiten especificar el tipo de almacenamiento para los campos, como se importan o crean en IBM SPSS Modeler. Para los nodos Archivo fijo, Archivo variable y Datos Usuario, también puede especificar el formato de campo u otros metadatos.

Para leer datos de otros orígenes, el almacenamiento viene determinado de manera automática aunque se puede modificar mediante una función de conversión, como `to_integer`, en un nodo Rellenar o un nodo Derivar.

Campo Utilice la columna **Campo** para ver y seleccionar campos en el conjunto de datos actual.

Ignorar Seleccione la casilla de verificación en la columna **Ignorar** para activar las opciones de las columnas **Almacenamiento** y **Formato de entrada**.

Almacenamiento de datos

Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es distinto del nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*. Aunque el almacenamiento se debe determinar en el origen, el nivel de medición se puede cambiar utilizando un nodo Tipo en cualquier punto de la ruta. Consulte el tema “Niveles de medición” en la página 127 para obtener más información.

Los tipos de almacenamiento disponibles son los siguientes:

- **Cadena:** Se utiliza para campos que contienen datos no numéricos, también llamados datos alfanuméricos. Una cadena puede incluir cualquier secuencia de caracteres, como *fred*, *Clase 2* o *1234*. Tenga en cuenta que los números de las cadenas no se pueden utilizar en los cálculos.
- **Entero:** Un campo cuyos valores son números enteros.
- **Real:** Los valores son números que pueden incluir decimales (no está limitado a números enteros). El formato de presentación se especifica en el cuadro de diálogo Propiedades de ruta y puede quedar anulado para campos individuales en un nodo Tipo (pestaña Formato).
- **Fecha** Valores de fecha especificados en un formato estándar como año, mes y día (por ejemplo, 2007-09-26). El formato determinado se especifica en el cuadro de diálogo Propiedades de ruta.
- **Tiempo:** El tiempo medido como una duración. Por ejemplo, una llamada de servicio que dure 1 hora, 26 minutos y 38 segundos se puede representar como 01:26:38, según el formato de hora actual que se especifica en el cuadro de diálogo Propiedades de ruta.
- **Indicación de fecha y hora:** Valores que incluyen un componente de fecha y hora, por ejemplo 2007-09-26 09:04:00, en función de los formatos actuales de fecha y hora del cuadro de diálogo Propiedades de ruta. Tenga en cuenta que es posible que los valores de marca de tiempo se tengan que escribir entre comillas dobles para garantizar que se interpretan como un valor único en lugar de valores de fecha y hora diferentes. (Es aplicable, por ejemplo cuando introduce valores en un nodo Datos Usuario.)
- **Lista:** Introducido en SPSS Modeler versión 17, junto con los nuevos niveles de medición de Geoespacial y Colección, el campo de almacenamiento de lista contiene varios valores para un único registro. Existen versiones de lista de todos los otros tipos de almacenamiento.

Tabla 1. Iconos de tipo de almacenamiento de lista

Icono	Tipo de almacenamiento
[A]	Lista de cadenas
[I]	Lista de enteros
[R]	Lista de reales
[T]	Lista de tiempo
[D]	Lista de fechas
[DT]	Lista de indicaciones de fecha y hora
[]	Lista con una profundidad mayor que cero

Asimismo, para utilizarlo con el nivel de medición Colección, existen versiones de los siguientes niveles de medición.

Tabla 2. Iconos de nivel de medición de lista

Icono	Nivel de medición
[C]	lista de continuos
[C]	Lista de categóricos
[M]	Lista de marcas
[N]	Lista de nominales
[O]	Lista de ordinales

Las listas se pueden importar a SPSS Modeler en uno de los tres nodos de origen (Analytic Server, Geoespacial o Archivo variable), o se pueden crear en sus rutas utilizando los nodos operativos de los campos Derivar o Rellenar.

Para obtener más información acerca de las listas y su interacción con los niveles de medición Colección y Geoespacial, consulte el tema “Almacenamiento de listas y niveles de medición asociados” en la página 11

Conversiones del almacenamiento. Puede convertir el almacenamiento de un campo mediante una serie de funciones de conversión, como `to_string` y `to_integer`, en un nodo Rellenar. Consulte el tema “Conversión del almacenamiento mediante el nodo Rellenar” en la página 150 para obtener más información. Tenga en cuenta que las funciones de conversión (y otras muchas funciones que requieren un tipo específico de entrada, como un valor de fecha u hora) dependen de los formatos actuales especificados en el cuadro de diálogo de propiedades de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003*, *Feb 2003*, (etc.) para almacenar datos, seleccione **MES AAAA** como formato de fecha predeterminado de la ruta. Las funciones de conversión también están disponibles en el nodo Derivar para la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos. Consulte el tema “Nueva codificación de valores con el nodo Derivar” en la página 149 para obtener más información.

Lectura de datos mezclados. Tenga en cuenta que al leer los campos con almacenamiento numérico (entero, real, hora, marca de tiempo o fecha), cualquier valor no numérico se establece como nulo o perdido por el sistema. Esto ocurre porque, a diferencia de otras aplicaciones, IBM SPSS Modeler no permite la combinación de tipos de almacenamiento dentro de un campo. Para evitarlo, cualquier campo con datos mezclados debe leerse como cadenas, ya sea cambiando el tipo de almacenamiento en el nodo de origen o en la aplicación externa, según sea necesario.

Formato de entrada de campo (únicamente los nodos Archivo fijo, Archivo variable y Datos usuario)

Para todos los tipos de almacenamiento, excepto Cadena y Entero, puede especificar las opciones de formato para el campo seleccionado mediante la lista desplegable. Por ejemplo, al fusionar datos desde distintas configuraciones regionales, puede que necesite especificar un punto (.) como el separador decimal para un campo, mientras que otro necesitará una coma como separador.

Las opciones de entrada especificadas en el nodo de origen omiten las opciones de formato especificadas en el cuadro de diálogo de propiedades de la ruta; sin embargo, no permanecen más adelante en la ruta. Pretenden realizar el análisis de la entrada correctamente, según el conocimiento que se tenga de los datos. Los formatos especificados se utilizan como un manual para realizar el análisis de los datos tal y como se leen en IBM SPSS Modeler y no para determinar cómo se debe asignar el formato una vez leídos en IBM SPSS Modeler. Para especificar el formato en una base por campo en otro lugar de la ruta, utilice la pestaña Formato de un nodo Tipo. Consulte el tema “Pestaña Configuración de Formato de campo” en la página 138 para obtener más información.

Las opciones varían en función del tipo de almacenamiento. Por ejemplo, para el tipo de almacenamiento Real, puede seleccionar **Punto (.)** o **Coma (,)** como separador decimal. Para los campos de marca de tiempo, se abrirá un cuadro de diálogo independiente cuando seleccione **Especificar** de la lista desplegable. Consulte el tema “Opciones de configuración de Formato de campo” en la página 138 para obtener más información.

Para todos los tipos de almacenamiento, puede seleccionar también **Valor predeterminado de la ruta** de forma que pueda utilizar los valores predeterminados de la ruta para la importación. La configuración de la ruta se especifica en el cuadro de diálogo de propiedades de la ruta.

Opciones adicionales

Hay otras opciones que se pueden especificar mediante la pestaña Datos:

- Si desea ver la configuración del almacenamiento para los datos que ya no están conectados a través del nodo actual (por ejemplo, datos de entrenamiento), seleccione **Ver configuración de campos no utilizados**. Puede borrar los campos heredados pulsando en **Borrar**.
- Mientras trabaja en este cuadro de diálogo, pulse en **Actualizar** en cualquier momento para volver a cargar los campos desde el origen de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de origen o cuando se trabaja entre pestañas en el cuadro de diálogo.

Almacenamiento de listas y niveles de medición asociados

Introducido en SPSS Modeler versión 17, para trabajar con los nuevos niveles de medición de Geoespacial y Colección, el campo de almacenamiento de lista contiene varios valores para un único registro. Las listas se encierran entre corchetes ([]). Los ejemplos de listas son: [1,2,4,16] y ["abc", "def"].

Las listas se pueden importar a SPSS Modeler en uno de los tres nodos de origen (Analytic Server, Geoespacial o Archivo variable), que se crean en sus rutas utilizando los nodos operativos de los campos Derivar o Rellenar, o que se generan mediante el nodo Fusión cuando se utiliza el método de fusión Condición clasificada.

Se considera que las listas tienen una profundidad, por ejemplo, una lista simple con elementos incluidos entre corchetes, con el formato [1,3], se registra en IBM SPSS Modeler con una profundidad de cero. Además de las listas simples con una profundidad de cero, puede utilizar listas anidadas en las que cada valor contenido en una lista es una lista en sí mismo.

La profundidad de una lista anidada depende del nivel de medición asociado. En el caso de Sin tipo no existe un límite de profundidad, en el caso de Colección el límite es cero, y para Geoespacial la profundidad debe estar entre cero y dos inclusive, en función del número de elementos anidados.

En las listas de profundidad cero puede establecer el nivel de medición en Geoespacial o en Colección. Estos dos niveles son niveles de medición padre y en el recuadro de diálogo Valores puede establecer la información acerca del subnivel de medición. El subnivel de medición de una colección determina el nivel de medición de los elementos contenidos en dicha lista. Todos los niveles de medición, salvo Sin tipo y Geoespacial, están disponibles como subniveles para las colecciones. El nivel de medición Geoespacial tiene seis subniveles de Punto, Cadena de líneas, Polígono, Multipunto, Cadena multilínea y Multipolígono. Para obtener más información, consulte "Subniveles de medición geoespacial" en la página 129.

Nota: El nivel de medición Colección solo se puede utilizar con las listas de profundidad 0, el nivel de medición Geoespacial solo se puede utilizar con las listas de una profundidad máxima de 2 y el nivel de medición Sin tipo se puede utilizar con cualquier profundidad de lista.

El siguiente ejemplo muestra la diferencia entre una lista de profundidad cero y una lista anidada utilizando la estructura de los subniveles de medición geoespacial Punto y Cadena de líneas:

- El subnivel de medición geoespacial Punto tiene una profundidad de campo de cero:
[1,3] dos coordenadas
[1,3,-1] tres coordenadas
- El subnivel de medición geoespacial de Cadena de líneas tiene una profundidad de campo de uno:
[[1,3], [5,0]] dos coordenadas
[[1,3,-1], [5,0,8]] tres coordenadas

El campo Punto (con una profundidad de cero) es una lista normal en la que cada valor consta de dos o tres coordenadas. El campo Cadena de líneas (con una profundidad de uno) es una lista de puntos en la que cada punto consta de una serie adicional de valores de lista.

Para obtener más información acerca de la creación de listas, consulte "Derivación de un campo de lista o geoespacial" en la página 146.

Nodo origen Analytic Server

El origen de Analytic Server permite ejecutar una ruta en un sistema de archivos distribuido de Hadoop (HDFS en sus siglas inglesas). La información de un origen de datos de Analytic Server puede proceder de diversos lugares, entre los que se incluyen:

- archivos de texto en HDFS
- bases de datos
- HCatalog

Normalmente, una secuencia con un origen de Analytic Server se ejecuta en HDFS. Sin embargo, si una secuencia contiene un nodo cuya ejecución no esté soportada en HDFS, se "inyectará de vuelta" tanta ruta como sea posible a Analytic Server, y entonces SPSS Modeler Server intentará procesar el resto de la ruta. Los bloques de datos muy grandes tendrán que dividirse en submuestras colocando, por ejemplo, un nodo Muestra en la ruta.

Origen de datos. Suponiendo que el administrador de SPSS Modeler Server haya establecido una conexión, seleccione el origen de datos que contenga los datos que desee utilizar. Un origen de datos contiene los archivos y los metadatos asociados a ese origen. Pulse **Seleccionar** para visualizar una lista de los orígenes de datos disponibles. Consulte el tema “Selección de un origen de datos” para obtener más información.

Si necesita crear un origen de datos o editar uno existente, pulse **Lanzar editor de orígenes de datos...**

Selección de un origen de datos

La tabla Orígenes de datos muestra una lista de los orígenes de datos disponibles. Seleccione el origen que desee y pulse **Aceptar**.

Pulse **Mostrar propietario** para visualizar el propietario del origen de datos.

Filtrar por permite filtrar el listado de orígenes de datos por **Palabra clave**, aplicando los criterios de filtro al nombre del origen de datos, a su descripción o al **Propietario**. Puede especificar una combinación de caracteres de cadena, numéricos o de comodín como criterio de búsqueda, tal como se describe a continuación. En la cadena de búsqueda se distingue entre mayúsculas y minúsculas. Pulse **Renovar** para actualizar la tabla Orígenes de datos.

– Se puede utilizar un carácter de subrayado para representar cualquier carácter individual en la cadena de búsqueda.

% Se puede utilizar un signo de porcentaje para representar cualquier secuencia de cero o más caracteres en la cadena de búsqueda.

Modificación de credenciales

Si las credenciales de acceso a Analytic Server son distintas de las credenciales de acceso a SPSS Modeler Server, será necesario especificar las credenciales de Analytic Server cuando se ejecute una ruta en Analytic Server. Si desconoce sus credenciales, póngase en contacto con el administrador del servidor-

Nodos soportados

La ejecución de muchos nodos de SPSS Modeler está soportada en HDFS, aunque es posible exista alguna diferencia en la ejecución de determinados nodos, mientras que otros ni siquiera están soportados en la actualidad. Este tema detalla en nivel de soporte actual.

Tema general

- Analytic Server no acepta algunos caracteres que normalmente se aceptan en el interior de un nombre de campo entrecomillado de Modeler.
- Para que una ruta de Modeler ejecute en Analytic Server, debe empezar con uno o más nodos Origen de Analytic Server y terminar con un único nodo de modelado o de exportación de Analytic Server.
- Se recomienda definir el almacenamiento de destinos continuos como real en lugar de entero. Los modelos de puntuación siempre escriben valores reales en los archivos de datos de salida de los destinos continuos, mientras que el modelo de datos de salida de las puntuaciones se ajusta al almacenamiento del destino. Por tanto, si un destino continuo tiene un almacenamiento entero, se producirá una discordancia entre los valores escritos y el modelo de datos de las puntuaciones, y dicha discordancia provocará errores cuando se intenten leer los datos puntuados.

Origen

- Una ruta que comience con cualquier cosa que no sea un nodo de origen de Analytic Server ejecutará en local.

Registrar operaciones

Se soportan todas las operaciones de registro con la excepción de los nodos Resolución de TS y Cuadros de espacio-tiempo. Las siguientes son notas adicionales acerca de las funciones de nodos soportadas.

Seleccionar

- Soporta el mismo conjunto de funciones que soporta el Nodo de derivación.

Muestrear

- No está soportado el muestreo a nivel de bloque.
- No se soportan los métodos de muestreo complejo.

Agregar

- Las claves contiguas no están soportadas. Si está reutilizando una ruta existente que se ha configurado para ordenar los datos y, a continuación, utiliza este valor en el nodo Agregar, cambie la ruta para eliminar el nodo Ordenar.
- Las estadísticas de orden (mediana, primer cuartil, tercer cuartil) se calculan de forma aproximada y están soportadas mediante el separador Optimización.

Ordenar

- La pestaña Optimización no está soportada.

En un entorno distribuido, hay un número limitado de operaciones que conservan el orden de registros establecido por el nodo Ordenar.

- Un nodo Ordenar seguido de uno Exportar produce un origen de datos ordenado.
- Un nodo Ordenar seguido de uno Muestrear con muestreo de **Primer** registro devuelve los N primeros registros.
- Uno nodo Ordenar seguido de un nodo de modelado que tenga el objetivo **Optimizar para conjuntos de datos muy grandes** (Red neuronal, Lineal, Árbol C y R, Quest o CHAID) es un patrón útil para mezclar aleatoriamente los registros mediante una ordenación basada en una clave de número aleatorio a fin de evitar que puedan introducirse sesgos en el algoritmo de construcción de modelos si los registros originales están ordenados.

En general, debe colocarse un nodo Ordenar lo más cerca posible de las operaciones que necesitan los registros ordenados.

Fusionar

- La fusión por Orden no está soportada.
- La pestaña Optimización no está soportada.
- En la actualidad no se soporta colocar un nodo Ordenar o un nugget de modelo entre un nodo Origen de Analytic Server y un nodo Fusionar. Normalmente es posible especificar un nodo Seleccionar para sustituir la funcionalidad del nodo Muestrear.
- Analytic Server no realiza uniones sobre claves de cadena vacías, es decir, si una de las claves por las que se está efectuando la fusión contiene cadenas vacías, los registros que contengan la cadena vacía se descartarán de la salida fusionada.
- Las operaciones de fusión son relativamente lentas. Si se dispone de espacio suficiente en HDFS, puede ser mucho más rápido fusionar una única vez los orígenes de datos y utilizar el origen fusionado en las rutas posteriores en lugar de fusionar los orígenes de datos en cada ruta.

Transformación R

La sintaxis R en el nodo debe constar de operaciones de tipo "registrar cada vez".

Las operaciones de campo

Se soportan todas las operaciones de campo con la excepción de los nodos Transponer, Intervalos de tiempo e Historial. Las siguientes son notas adicionales acerca de las funciones de nodos soportadas.

Preparación automática de datos

- No se soporta el entrenamiento del nodo. Se soporta la aplicación a los datos nuevos de transformaciones de un nodo Preparación de datos automática entrenado.

Tipo

- La columna Comprobación no está soportada.
- La pestaña Formato no está soportada.

Derivar

- Se soportan todas las funciones de Derivar, a excepción de las funciones de secuencia.
- Los campos divididos no pueden derivarse en la misma ruta que los utiliza como divisiones. En tal caso será necesario crear dos rutas: una que derive el campo dividido y una que utilice el campo como divisiones.
- No se puede utilizar un campo de marcas como tal en una comparación; es decir, `if (campoMarcas) then ... endif` provocará un error; la solución consiste en utilizar `if (campoMarca=valorVerdadero) then ... endif`
- Al utilizar el operador `**`, se recomienda especificar el exponente como un número real como, por ejemplo, `x**2.0`, en lugar de `x**2`, para comparar los resultados en Modeler

Rellenar

- Soporta el mismo conjunto de funciones soportado por el Nodo de derivación.

En intervalo

La siguiente funcionalidad no está soportada.

- Intervalos óptimos.
- Rangos.
- Cuantiles -> Creación de cuantiles: suma de valores.
- Cuantiles -> Empates: mantener en actual y asignar aleatoriamente.
- Cuantiles -> N personalizados: valores por encima de 100 y cualquier valor N donde el 100% de N no sea igual a cero.

Análisis de RFM

- No se soporta la opción Mantener en actual para tratar los empates. La actualidad de RFM, la frecuencia y las puntuaciones monetarias no siempre coincidirán con las que Modeler calcula a partir de los mismos datos. Los rangos de puntuación serán los mismos, pero las asignaciones de puntuación (números de intervalo) pueden diferir en uno.

Gráficos

Se soportan todos los nodos Gráfico.

Modelado

Se da soporte a los siguientes nodos de modelado: Lineal, Red neuronal, C y RT, Chaid, Quest, TCM, Bietápico-AS, STP y Reglas de asociación. Las siguientes son notas adicionales acerca de las funciones de estos nodos.

Lineal Al crear modelos basados en datos masivos, lo normal es cambiar el objetivo a Conjuntos de datos muy grandes o bien especificar divisiones.

- No se soporta el entrenamiento continuado de modelos de PSM existentes.
- El objetivo Generación de modelo estándar solo se recomienda si los campos divididos se definen de modo que el número de registros de cada división no sea demasiado

grande, donde la definición de "demasiado grande" depende de la potencia de los nodos individuales del clúster de Hadoop. Por contra, también debe procurarse que las divisiones no se definan tan pequeñas que contengan demasiados pocos registros como para construir un modelo.

- No se soporta el objetivo de Potenciación.
- No se soporta el objetivo de Agregación autodocimante.
- No se recomienda el objetivo de Conjuntos de datos muy grandes cuando hay pocos registros; con frecuencia no se generará un modelo, o el modelo generado estará degradado.
- No se soporta la preparación de datos automática. Esto puede dar lugar a problemas cuando se intente construir un modelo a partir de datos con muchos valores ausentes; normalmente dichos datos se imputarían como parte de la preparación de datos automática. Una solución consistiría en utilizar un modelo de árbol o una red neuronal con el valor Avanzado seleccionado para imputar los valores ausentes.
- La estadística de precisión no se calcula para los modelos divididos.

Red neuronal

Al crear modelos basados en datos masivos, lo normal es cambiar el objetivo a Conjuntos de datos muy grandes o bien especificar divisiones.

- No se soporta el entrenamiento continuado de modelos de PSM o estándar existentes.
- El objetivo Generación de modelo estándar solo se recomienda si los campos divididos se definen de modo que el número de registros de cada división no sea demasiado grande, donde la definición de "demasiado grande" depende de la potencia de los nodos individuales del clúster de Hadoop. Por contra, también debe procurarse que las divisiones no se definan tan pequeñas que contengan demasiados pocos registros como para construir un modelo.
- No se soporta el objetivo de Potenciación.
- No se soporta el objetivo de Agregación autodocimante.
- No se recomienda el objetivo de Conjuntos de datos muy grandes cuando hay pocos registros; con frecuencia no se generará un modelo, o el modelo generado estará degradado.
- Cuando falten muchos valores en los datos, utilice el valor Avanzado para imputar los valores que faltan.
- La estadística de precisión no se calcula para los modelos divididos.

Árbol C&R, CHAID y Quest

Al crear modelos basados en datos masivos, lo normal es cambiar el objetivo a Conjuntos de datos muy grandes o bien especificar divisiones.

- No se soporta el entrenamiento continuado de modelos de PSM existentes.
- El objetivo Generación de modelo estándar solo se recomienda si los campos divididos se definen de modo que el número de registros de cada división no sea demasiado grande, donde la definición de "demasiado grande" depende de la potencia de los nodos individuales del clúster de Hadoop. Por contra, también debe procurarse que las divisiones no se definan tan pequeñas que contengan demasiados pocos registros como para construir un modelo.
- No se soporta el objetivo de Potenciación.
- No se soporta el objetivo de Agregación autodocimante.
- No se recomienda el objetivo de Conjuntos de datos muy grandes cuando hay pocos registros; con frecuencia no se generará un modelo, o el modelo generado estará degradado.
- No se soportan las sesiones interactivas.
- La estadística de precisión no se calcula para los modelos divididos.

Puntuación de modelos

Todos los modelos soportados para el modelado también están soportados para la puntuación. Asimismo, se soportan los nuggets de modelos incorporados localmente para la puntuación: C y RT, Quest, CHAID, Lineal y Red neuronal (independientemente de si el modelo es estándar, de potenciación, de agregación autocorimante o para conjuntos de datos muy grandes), Regresión, C5.0, Logístico, Genlin, GLMM, Cox, SVM, Red de Bayes, Bietápico, KNN, Lista de decisiones, Discriminante, Autoaprendizaje, Detección de anomalías, Apriori, Carma, K-Medias, Kohonen, R y Minería de texto.

- No se puntuarán las propensiones brutas ni las ajustadas. A modo de solución alternativa, puede lograrse el mismo efecto calculando manualmente la propensión bruta utilizando un nodo Derivar con la siguiente expresión: `if 'valor-pronosticado' == 'valor-de-interés' then 'prob-de-ese-valor' else 1-'prob-de-ese-valor' endif`
- Al puntuar un modelo, Analytic Server no comprueba si todos los campos utilizados en el modelo están presentes en el conjunto de datos, así que asegúrese de que sea así antes de ejecutar en Analytic Server

R La sintaxis R en el nugget debe constar de operaciones de tipo "registrar cada vez".

Resultado

Se soportan los nodos Matriz, Análisis, Auditar datos, Transformar, Estadísticas y Medias.

Se da soporte al nodo Tabla escribiendo un origen de datos de Analytic Server temporal que contenga los resultados de las operaciones en sentido ascendente. A continuación, el nodo Tabla pagina por el contenido de dicho origen de datos.

Exportar

Una secuencia puede comenzar con un nodo origen de Analytic Server y terminar con un nodo de exportación distinto del nodo de exportación de Analytic Server, pero los datos se moverán de HDFS a SPSS Modeler Server y por último a la ubicación de exportación.

Nodo de origen Base de datos

El nodo de origen de base de datos se puede utilizar para importar datos desde otros paquetes mediante ODBC (Open Database Connectivity), incluidos Microsoft SQL Server, DB2, Oracle, etc.

Para leer o escribir en una base de datos, debe tener un origen de datos ODBC instalado y configurado para la base de datos pertinente, con los permisos de lectura o escritura, según sea necesario. IBM SPSS Data Access Pack incluye un conjunto de controladores ODBC que puede utilizarse con este fin. Estos controladores están disponibles en IBM SPSS Data Access Pack DVD o en el sitio de descarga. Si tiene alguna pregunta acerca de la creación o configuración de permisos de los orígenes de datos ODBC, póngase en contacto con el administrador de la base de datos.

Controladores ODBC compatibles

Si desea obtener la información más reciente acerca de las bases de datos y los controladores ODBC compatibles y cuyo uso se ha comprobado con IBM SPSS Modeler 17, vea las matrices de compatibilidad de productos disponibles en el sitio de asistencia corporativo (<http://www.ibm.com/support>).

Dónde instalar los controladores

Nota: Los controladores ODBC se deben instalar y configurar en todos los ordenadores donde vaya a realizarse el procesamiento.

- Si desea ejecutar IBM SPSS Modeler en modo local (independiente), deberá instalar los controladores en el ordenador local.
- Si ejecuta IBM SPSS Modeler en modo distribuido en un servidor IBM SPSS Modeler Server remoto, los controladores ODBC se deberán instalar en el ordenador en el que se encuentre IBM SPSS Modeler

Server. En el caso de IBM SPSS Modeler Server en sistemas UNIX consulte también "Configuración de controladores ODBC en sistemas UNIX" más adelante en esta sección.

- Si necesita poder acceder a los mismos orígenes de datos desde IBM SPSS Modeler y IBM SPSS Modeler Server, deberá instalar los controladores ODBC en ambos ordenadores.
- Si ejecuta IBM SPSS Modeler con los Servicios de Terminal Server, los controladores ODBC se deberán instalar en el servidor de Servicios de Terminal Server en el que ha instalado IBM SPSS Modeler.

Configuración de controladores ODBC en sistemas UNIX

Nota: Si va a trabajar con SAP HANA o IBM DB2, consulte la siguiente sección para obtener detalles de configuración.

De forma predeterminada, Driver Manager de DataDirect no está configurado para IBM SPSS Modeler Server en sistemas UNIX. Para configurar UNIX para que cargue Driver Manager de DataDirect, introduzca los siguientes comandos:

```
cd <directorio_instalación_modeler_server>/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

Al realizar esta acción, se elimina el enlace predeterminado y se crea un enlace a Driver Manager de DataDirect.

Configuración de controladores ODBC para SAP HANA e IBM DB2 en sistemas UNIX

La conexión con SAP HANA o IBM DB2, utilizando el controlador ODBC de IBM DB2, desde IBM SPSS Modeler Server en un entorno UNIX o Linux requiere la configuración de SPSS Modeler Server para que utilice un nuevo derivador ODBC, `libspssodbc_datadirect_utf16.so`. Para configurar SPSS Modeler Server para que utilice este derivador, realice los pasos siguientes:

1. Vaya a la carpeta `<directorio_instalación>/bin` de SPSS Modeler Server.
2. Elimine en archivo de enlace `libspssodbc.so` existente utilizando el comando siguiente:

```
rm -f libspssodbc.so
```
3. Enlace el nuevo derivador con `libspssodbc.so` utilizando el comando siguiente:

```
ln -s libspssodbc_datadirect_utf16.so libspssodbc.so
```
4. Configure el script de inicio de SPSS Modeler Server, `modelersrv.sh`, para extraer el archivo de entorno `odbc.sh` de IBM SPSS Data Access Pack añadiendo la línea siguiente en `modelersrv.sh`:

```
./<vía_instalación_SDAP>/odbc.sh
```

Donde `<vía_instalación_SDAP>` es la ruta completa a la instalación de IBM SPSS Data Access Pack.

5. Reinicie SPSS Modeler Server.
6. Añada la definición de parámetro siguiente para el DSN en el archivo `odbc.ini` para evitar desbordamientos de almacenamiento intermedio durante la conexión:

```
DriverUnicodeType=1
```

Nota: El derivador `libspssodbc_datadirect_utf16.so` también es compatible con los demás controladores ODBC de SPSS Modeler Server admitidos.

Acceso a datos desde una base de datos

Para acceder a datos desde una base de datos, siga estos pasos.

- Instale un controlador ODBC y configure un origen de datos en la base de datos que desee utilizar.

- En el cuadro de diálogo del nodo Base de datos, conéctese a una base de datos mediante el modo Tabla o Consulta SQL.
- Seleccione una tabla de la base de datos.
- Mediante el cuadro de diálogo del nodo Base de datos, puede modificar los tipos de usos y filtrar los campos de datos.

En los temas de la documentación relacionados se proporciona información más detallada acerca de los pasos anteriores.

Nota: Si invoca los procedimientos almacenados (SP) de base de datos desde SPSS Modeler, es posible que vea un único campo de resultado denominado RowsAffected, en lugar del resultado previsto de los SP. Esto sucede cuando ODBC no devuelve información suficiente para poder determinar el modelo de datos de salida del SP. SPSS Modeler tiene únicamente un soporte limitado para los SP que devuelven resultados y se recomienda que, en lugar de utilizar los SP, extraiga SELECT de SP y utilice cualquiera de las acciones siguientes.

- Cree una vista basada en SELECT y seleccione la vista en el nodo Origen de base de datos
- Utilice SELECT directamente en el nodo Origen de base de datos.

Configuración de las opciones de nodos Base de datos

Puede utilizar las opciones de la pestaña Datos del cuadro de diálogo del nodo de origen de base de datos para poder acceder a una base de datos y leer los datos desde la tabla seleccionada.

Modo. Seleccione **Tabla** para conectarse a una tabla mediante los controles de los cuadros de diálogo.

Seleccione **Consulta SQL** para realizar consultas a la base de datos seleccionada a continuación mediante SQL. Consulte el tema “Realización de consultas a la base de datos” en la página 24 para obtener más información.

Origen de datos. Para los modos Tabla y Consulta SQL, puede introducir un nombre en el campo Origen de datos o seleccionar **Añadir nueva conexión de base de datos** de la lista desplegable.

Las siguientes opciones se utilizan para conectarse a una base de datos y seleccionar una tabla mediante el cuadro de diálogo:

Nombre de tabla. Si conoce el nombre de la tabla a la que desea acceder, introdúzcalo en el campo Nombre de tabla. En caso contrario, pulse en el botón **Seleccionar** para abrir un cuadro de diálogo que muestre las tablas disponibles.

Poner entre comillas nombres de columna y tabla. Indique si desea que los nombres de columna y tabla aparezcan entre comillas cuando las consultas se envíen a la base de datos (en el caso, por ejemplo, de que contengan espacios o signos de puntuación).

- La opción **Cuando sea necesario** pondrá entre comillas los nombres de campos y tablas *sólo* si incluyen caracteres no estándar. Los caracteres no estándar incluyen caracteres no ASCII, caracteres de espacio y los no alfanuméricos distintos del punto (.).
- Seleccione **Siempre** si desea que *todos* los nombres de campos y tablas aparezcan entre comillas.
- Seleccione **Nunca** si *no* desea que los nombres de campos y tablas aparezcan entre comillas.

Eliminar espacios iniciales y finales. Seleccione las opciones para descartar los espacios iniciales y finales en las cadenas.

Nota: Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios finales.

Lectura de cadenas vacías desde Oracle. Al leer o grabar en una base de datos Oracle, debe tener en cuenta que, a diferencia de IBM SPSS Modeler y la mayoría del resto de base de datos, Oracle trata y almacena valores de cadenas vacíos tanto como valores equivalentes hasta nulos. Esto significa que los mismos datos extraídos desde una base de datos de Oracle pueden comportarse de manera diferente a cuando se extraen desde otra base de datos o de un archivo, y pueden devolver resultados diferentes.

Adición de una conexión de base de datos

Para abrir una base de datos, en primer lugar seleccione el origen de datos al que desea conectarse. En la pestaña Datos, seleccione **Añadir nueva conexión de base de datos** de la lista desplegable Origen de datos.

Esto abrirá el cuadro de diálogo Conexiones a la base de datos.

Nota: Otro modo de abrir este diálogo, desde el menú principal, es seleccionar: **Herramientas > Bases de datos...**

Orígenes de datos. Muestra los orígenes de datos disponibles. Desplácese hacia abajo si no encuentra la base de datos deseada. Una vez haya seleccionado un origen de datos y haya introducido las contraseñas, pulse en **Conectar**. Pulse en **Actualizar** para actualizar la lista.

Nombre de usuario. Si el origen de datos está protegido por una contraseña, introduzca su nombre de usuario.

Contraseña. Si el origen de datos está protegido por una contraseña, introduzca su contraseña.

Credencial. Si se ha configurado una credencial en IBM SPSS Collaboration and Deployment Services, puede seleccionar esta opción para examinar el repositorio en su búsqueda. El nombre de usuario y la contraseña de la credencial deben coincidir con el nombre de usuario y la contraseña necesarios para acceder a la base de datos.

Conexiones. Muestra bases de datos conectadas actualmente.

- **Valor predeterminado.** Puede seleccionar una conexión como el valor predeterminado. Al hacerlo, el origen de base de datos o los nodos de exportación tendrán esta conexión predefinida como su origen de datos, aunque puede editarse si lo desea.
- **Guardar.** Puede seleccionar una o más conexiones que quiera volver a mostrar en sesiones posteriores.
- **Origen de datos.** Las cadenas de conexión para las bases de datos conectadas actualmente.
- **Predefinido.** Indica (con el carácter *) si se han especificado valores predefinidos para la conexión de base de datos. Para especificar valores predefinidos, pulse en esta columna en la fila correspondiente a la conexión de base de datos y seleccione Especificar en la lista. Consulte el tema “Especificación de valores predefinidos para una conexión de base de datos” para obtener más información.

Para eliminar las conexiones, seleccione una de la lista y pulse en **Eliminar**.

Una vez haya completado sus selecciones, pulse en **Aceptar**.

Especificación de valores predefinidos para una conexión de base de datos

Para algunas bases de datos, puede especificar cierto número de ajustes predeterminado para una conexión de base de datos. Todos los ajustes se aplican a la exportación de base de datos.

Los tipos de bases de datos que admiten esta característica son los siguientes:

- IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior. Consulte el tema “Configuración para IBM DB2 InfoSphere Warehouse” en la página 21 para obtener más información.

- SQL Server 2008 o posterior, Enterprise Edition y Developer Edition. Consulte el tema “Configuración para SQL Server” para obtener más información.
- Oracle 10g y 11gR1 o posterior, Enterprise Edition o Personal Edition. Consulte el tema “Configuración para Oracle” para obtener más información.
- IBM Netezza, IBM DB2 para z/OS y Teradata se conectan a una base de datos o esquema de un modo similar. Consulte el tema “Valores para IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW y Teradata” en la página 22 para obtener más información.

Si está conectado a una base de datos o esquema que no admite esta característica, verá el mensaje **No se puede configurar ningún valor predeterminado para esta conexión de base de datos.**

Configuración para IBM DB2 InfoSphere Warehouse

Estos ajustes se muestran para IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior.

Espacio de tabla. El espacio de tabla que debe utilizarse para la exportación. Los administradores de la base de datos pueden crear o configurar espacios de tabla con particiones. Recomendamos que seleccione uno de estos espacios de tabla (en lugar del valor predeterminado) para utilizarlo para la exportación de base de datos.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión (por ejemplo, el equivalente a `CREATE TABLE MYTABLE(...) COMPRESS YES;` en SQL).

No registrar actualizaciones. Si está seleccionado, evita el registro al crear tablas e insertar datos (el equivalente de `CREATE TABLE MYTABLE(...) NOT LOGGED INITIALLY;` en SQL).

Configuración para SQL Server

Estos ajustes se muestran para SQL Server 2008 o posterior, Enterprise Edition y Developer Edition.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Fila.** Activa la compresión a nivel de fila (por ejemplo, el equivalente de `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW);` en SQL).
- **Página.** Activa la compresión a nivel de página (por ejemplo, `CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE);` en SQL).

Configuración para Oracle

Configuración de Oracle 10g

Estos ajustes se muestran para Oracle 10g Enterprise Edition o Personal Edition.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión. Para esta versión de la base de datos, únicamente está disponible la compresión básica (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL).

Configuración de Oracle 11gR1

Estos ajustes se muestran para Oracle 11g R1 Enterprise Edition o Personal Edition.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, `CREATE TABLE MYTABLE(...) COMPRESS;` en SQL). En este caso, tiene el mismo efecto que la opción **Operaciones de carga directa.**

- **Operaciones de carga directa.** Activa la compresión únicamente para operaciones de inserción masivas (de ruta directa) (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR DIRECT_LOAD OPERATIONS; en SQL).
- **Todas las operaciones.** Activa la compresión para todas las operaciones (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR ALL OPERATIONS; en SQL).

Configuración de Oracle 11gR2: opción Básica

Estos ajustes se muestran para Oracle 11g R2 Enterprise Edition o Personal Edition que utilice la opción Básica.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción **Básica**.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).

Configuración de Oracle 11gR2: opción Avanzada

Estos ajustes se muestran para Oracle 11g R2 Enterprise Edition o Personal Edition que utilice la opción Avanzada.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción **Básica**.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).
- **OLTP.** Activa la compresión OLTP (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP; en SQL).
- **Consulta baja/alta.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para la consulta (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY LOW; o CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY HIGH; en SQL). La compresión para consulta es de utilidad en entornos de almacenamiento de datos; HIGH proporciona una relación de compresión mayor que LOW.
- **Archivo bajo/alto.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para el archivo (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE LOW; o CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE HIGH; en SQL). La compresión para archivo es de utilidad para comprimir datos que se almacenarán durante largos períodos de tiempo; HIGH proporciona una relación de compresión mayor que LOW.

Valores para IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW y Teradata

Cuando especifica valores predefinidos para IBM Netezza, IBM DB2 for z/OS, IBM DB2 LUW o Teradata, se le solicitará que seleccione los elementos siguientes:

Utilizar base de datos del adaptador de puntuación del servidor o Utilizar esquema del adaptador de puntuación del servidor. Si se selecciona, habilita la opción **Base de datos del adaptador de puntuación del servidor** o la opción **Esquema del adaptador de puntuación del servidor**.

Base de datos del adaptador de puntuación del servidor o Esquema del adaptador de puntuación del servidor: En la lista desplegable, seleccione la conexión que necesite.

Asimismo, para Teradata, también puede establecer los detalles de Query Banding para proporcionar metadatos adicionales como ayuda con elementos, tales como la gestión de carga de trabajo, la clasificación, identificación y resolución de consultas, y el seguimiento del uso de la base de datos.

Comprobar ortografía de Query Banding. Seleccione esta opción si Query Banding se establecerá una sola vez durante el periodo que trabaja con la conexión de la base de datos de Teradata (**Por sesión**), o si se establecerá cada vez que ejecute una ruta (**Por transacción**).

Nota: Si establece Query Banding en una ruta, se perderán las bandas si copia dicha ruta en otra máquina. Para impedirlo, puede utilizar scripts para ejecutar su ruta y utilizar la palabra clave *querybanding* en el script para aplicar los valores que requiera.

Valores predeterminados necesarios para acceder a SIBM.SYSROUTINES en IBM DB2 for z/OS y en IBM DB2 LUW

Los usuarios requieren acceso a las tablas del sistema relacionadas para poder recuperar la información registrada de las rutinas externas del adaptador de puntuación de UDF. SPSS Modeler Server requiere acceso a las tablas siguientes:

Netezza: UDF

- `_V_FUNCTION`

Netezza: En Minería de base de datos

- `_V_DATABASE`

Teradata: UDF/QueryBinding

- `DBC.Functions`
- `DBC.USERS`

DB2 LUW: Puntuación de UDF, Valores predeterminados de base de datos

- `SYSCAT.TABLESPACES`
- `SYSCAT.SCHEMATA`
- `SYSIBM.SYSDUMMY1`
- `SYSCAT.TABLES`

DB2 for i Series: Puntuación de UDF, Valores predeterminados de base de datos

- `QSYS2.SYSROUTINES`

DB2/z: Puntuación de UDF

- `SYSIBM.SYSDUMMY1`
- `SYSIBM.SYSDUMMYU`
- `SYSIBM.SYSROUTINES`
- `SYSIBM.SYSTABLES`
- `SYSIBM.SYSVIEWS`

Selección de una tabla de la base de datos

Una vez se haya conectado a un origen de datos, puede elegir importar campos desde una tabla o vista determinada. Desde la pestaña Datos del cuadro de diálogo Base de datos, puede especificar el nombre de una tabla en el campo Nombre de tabla o pulsar **Seleccionar** para abrir el cuadro de diálogo Seleccionar tabla/vista, que muestra las tablas y vistas disponibles.

Mostrar propietario de tabla. Indique si un origen de datos necesita que se especifique el propietario de una tabla antes de acceder a ella. Anule la selección de esta opción para los orígenes de datos que no tienen este requisito.

Note: las bases de datos de Oracle y SAS suelen requerir que muestre al propietario de tabla.

Tablas/vistas. Seleccione la tabla o vista que se vaya a importar.

Mostrar. Muestra las columnas del origen de datos a la que está conectado actualmente. Pulse en una de las opciones siguientes para personalizar la vista de las tablas disponibles:

- Pulse en **Tablas de usuario** para ver las tablas normales de la base de datos creadas por los usuarios de la base de datos.
- Pulse en **Tablas de sistema** para ver las tablas de la base de datos que posee el sistema (por ejemplo, las tablas que ofrecen información acerca de la base de datos, como detalles de índices). Esta opción puede utilizarse para ver las pestañas utilizadas en las bases de datos de Excel. (Tenga en cuenta que también existe un nodo de origen de Excel independiente. Consulte el tema “Nodo de origen de Excel” en la página 42 para obtener más información).
- Pulse en **Vistas** para ver las tablas virtuales basadas en una consulta que implica una o más tablas normales.
- Pulse en **Sinónimos** para ver los sinónimos creados en la base de datos para las tablas existentes.

Filtros de nombre/propietario. Estos campos le permiten filtrar la lista de tablas mostradas por nombre o propietario. Por ejemplo, escriba *SYS* para mostrar solamente las tablas con este propietario. Para las búsquedas con comodín, se puede utilizar un guión bajo (*_*) para representar un carácter simple y un signo de porcentaje (*%*) puede representar una secuencia de cero o más caracteres.

Establecer como valor predeterminado. Guarda la configuración actual como el valor predeterminado para el usuario actual. Esta configuración se restaurará en el futuro cuando un usuario abra un nuevo cuadro de diálogo de selector de tablas *para el mismo nombre de origen de datos e inicio de sesión sólo del usuario*.

Realización de consultas a la base de datos

Una vez se haya conectado a un origen de datos, podrá elegir importar campos mediante consultas SQL. En el cuadro de diálogo principal, seleccione **Consulta SQL** como modo de conexión. Esto añadirá una ventana de editor de consultas al cuadro de diálogo. Mediante el editor de consultas, puede crear o cargar una o más consultas SQL cuyo conjunto de resultados se leerá en la ruta de datos.

Si especifica múltiples consultas SQL, sepárelas con punto y coma (;) y asegúrese de que no hay varias declaraciones SELECT.

Para cancelar y cerrar la ventana de editor de consultas, seleccione **Tabla** como modo de conexión.

Puede incluir parámetros de ruta de SPSS Modeler (un tipo de variable definida por el usuario) en la consulta SQL. Consulte el tema “Uso de parámetros de ruta en una consulta SQL” en la página 25 para obtener más información.

Cargar consulta. Pulse para abrir el explorador de archivos, que puede utilizar para cargar una consulta guardada anteriormente.

Guardar consulta. Pulse para abrir el cuadro de diálogo Guardar consulta, que puede utilizar para guardar la consulta actual.

Importar valor predeterminado. Pulse para importar una instrucción SELECT DE SQL de ejemplo generada de forma automática mediante la tabla y columnas seleccionadas en el cuadro de diálogo.

Borrar. Borre el contenido del área de trabajo. Utilice esta opción cuando desee comenzar de nuevo.

Dividir texto. La opción principal **Nunca** significa que la consulta se enviará a la base de datos como un todo. De forma alternativa, puede elegir **Cuando sea necesario**, lo que significa que SPSS Modeler intentará analizar la consulta e identificar si hay sentencias SQL que deban enviarse a la base de datos una tras otra.

Uso de parámetros de ruta en una consulta SQL

Al escribir una consulta SQL para importar campos, puede incluir parámetros de ruta de SPSS Modeler que se hayan definido anteriormente. Se admiten todos los tipos de parámetros de ruta.

La siguiente tabla muestra cómo se interpretarán algunos ejemplos de parámetros de ruta en la consulta SQL.

Tabla 3. Ejemplos de parámetros de ruta.

Nombre de parámetro de ruta (ejemplo)	Almacenamiento	Valor del parámetro de ruta	Interpretado como
PString	Cadena	ss	'ss'
PInt	Entero	5	5
PReal	Real	5.5	5.5
PTime	Hora	23:05:01	t{'23:05:01'}
PDate	Fecha	2011-03-02	d{'2011-03-02'}
PTimeStamp	Marca de tiempo	2011-03-02 23:05:01	ts{'2011-03-02 23:05:01'}
PColumn	Desconocido	IntValue	IntValue

En la consulta SQL, se especifica un parámetro de ruta del mismo modo que en una expresión CLEM, concretamente por '\$P-<nombre_parámetro>', donde <nombre_parámetro> es el nombre que se ha definido para el parámetro de ruta.

Cuando se hace referencia a un campo, el tipo de almacenamiento debe definirse como Desconocido, y el valor del parámetro debe ir entre comillas si es necesario. Por lo tanto, utilizando los ejemplos mostrados en la tabla, si introdujo la consulta SQL:

```
select "IntValue" from Table1 where "IntValue" < '$P-PInt';
```

se evaluará como:

```
select "IntValue" from Table1 where "IntValue" < 5;
```

Si necesita hacer referencia al campo IntValue mediante el parámetro PColumn, deberá especificar la consulta de la manera siguiente para obtener el mismo resultado:

```
select "IntValue" from Table1 where "'$P-PColumn'" < '$P-PInt';
```

Nodo Archivo variable

Puede utilizar los nodos Archivo variable para leer los datos desde los archivos de texto de campo libre (archivos cuyos registros contienen un número constante de campos pero un número variado de caracteres), también conocido como archivos de texto delimitado. Este tipo de nodo resulta también útil para los archivos con texto de cabecera de longitud fija y determinados tipos de anotaciones. Los registros se leerán uno a uno y pasarán por la ruta hasta que se lea todo el archivo.

Nota relativa a la lectura de datos geoespaciales

Si el nodo contiene datos geoespaciales, y el nodo se creó como una exportación a partir de un archivo sin formato, debe realizar algunos pasos adicionales para configurar los metadatos geoespaciales. Si desea obtener más información, consulte “Importación de los datos geoespaciales al nodo Archivo variable” en la página 28.

Notas para la lectura en datos de texto delimitados

- Los registros se deben delimitar por un carácter de nueva línea en el extremo de cada línea. El carácter de nueva línea no debe utilizarse con ningún otro fin (como dentro de cualquier valor o nombre de campo). Los espacios iniciales y finales se deberían eliminar para ahorrar espacio, aunque no es esencial. De forma opcional, estos espacios se pueden eliminar a través del nodo.
- Los campos se deben delimitar por una coma u otro carácter que idealmente se utilice solamente como delimitador, es decir, que no aparezca en valores o nombres de campos. Si esto no es posible, todos los campos de texto se pueden incluir entre signos de comillas dobles, si ninguno de los nombres de campo ni los valores de texto contiene un signo de comillas dobles. Si los valores o nombres de campo contienen signos de comillas dobles, los campos de texto deben delimitarse con signos de comillas simples como alternativa pero, al igual que en el caso anterior, si no se utilizan signos de comillas simples dentro de los valores. Si no se pueden utilizar signos de comillas simples ni dobles, se deberán cambiar los valores de texto para eliminar o sustituir el carácter delimitador o los signos de comillas simples o dobles.
- Cada fila, incluida la fila de la cabecera, debe contener el mismo número de campos.
- La primera línea debería contener los nombres de campo. Si no es así, deseleccione **Leer nombres de campo desde archivo** para proporcionar a cada campo un nombre genérico como Field1, Field2, etc.
- La segunda línea debería contener el primer registro de los datos. No debe haber línea en blanco ni comentarios.
- Los valores numéricos no deben incluir el separador de millares ni el símbolo de agrupación, sin la coma en 3,000.00, por ejemplo. El indicador decimal (punto o punto y aparte en EE.UU. o el Reino Unido) sólo debe utilizarse cuando sea necesario.
- Los valores de fecha y hora deben estar en uno de los formatos reconocidos en el cuadro de diálogo Opciones de ruta, como DD/MM/AAAA o HH:MM:SS. Todos los campos de fecha y hora del archivo deberían seguir el mismo formato y cualquier campo que contenga una fecha debe utilizar el mismo formato para todos los valores dentro de ese campo.

Opciones de configuración para el nodo Archivo variable

Las opciones se establecen en la pestaña Archivo del cuadro de diálogo del nodo Archivo variable.

Archivo: especifique el nombre del archivo. Puede especificar un nombre de archivo o pulsar el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará cuando seleccione un archivo y su contenido se mostrará con los delimitadores del panel inferior.

El texto de muestra representado desde el origen de datos se puede copiar y pegar en los controles siguientes: caracteres de comentario EOL y delimitadores especificados por el usuario. Utilice Ctrl-C y Ctrl-V para copiar y pegar.

Leer nombres de campo desde archivo: seleccionada de forma predeterminada, esta opción trata la primera fila del archivo de datos como etiquetas para la columna. Si la primera fila no es una cabecera, anule la selección para proporcionar a cada campo de forma automática un nombre genérico, como *Campo1*, *Campo2*, para el número de campos del conjunto de datos.

Especificar número de campos. Especifique el número de campos de cada registro. El número de campos puede detectarse de forma automática siempre que los registros hayan finalizado en nueva línea. También puede establecer el número de forma manual.

Omitir caracteres de cabecera. Especifique los caracteres que desea ignorar al principio del primer registro.

Caracteres de comentario de EOL. Especifique los caracteres, como # o !, para indicar anotaciones en los datos. Si uno de estos caracteres aparece en el archivo de datos, se ignorará todo lo que aparezca hasta aquí, sin incluir el siguiente carácter de nueva línea.

Eliminar espacios iniciales y finales. Seleccione las opciones para descartar los espacios iniciales y finales en las cadenas de importación.

Nota: Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios finales.

Caracteres no válidos. Seleccione **Descartar** para eliminar los caracteres no válidos de origen de datos. Seleccione **Reemplazar con** para reemplazar los caracteres no válidos con el símbolo especificado (solamente un carácter). Los caracteres no válidos son los caracteres nulos o los que no existen en el método de codificación especificado.

Codificación. Especifica el método de codificación de textos utilizado. Puede elegir entre valor predeterminado del sistema, el valor predeterminado de la ruta o UTF-8.

- El valor predeterminado del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta predeterminada se especifica en el cuadro de diálogo Propiedades de ruta.

Símbolo decimal: seleccione el tipo de separador decimal que se utiliza en el origen de datos. El **Valor predeterminado de la ruta** es el carácter seleccionado de la pestaña Opciones del cuadro de diálogo de propiedades de la ruta. En el resto de los casos, seleccione **Punto (.)** o **Coma (,)** para leer todos los datos en este cuadro de diálogo utilizando el carácter elegido como el separador decimal.

El delimitador de línea es el carácter de nueva línea: para utilizar el carácter de nueva línea como el delimitador de línea en lugar de un delimitador de campo, seleccione esta opción. Por ejemplo, esto puede ser de utilidad si hay un número impar de delimitadores en una fila que hace que haya ajuste de línea. Tenga en cuenta que seleccionar esta opción significa que no podrá seleccionar **Nueva línea** en la lista Delimitadores.

Nota: Si selecciona esta opción, se eliminarán los valores de espacios en blanco al final de las filas de datos.

Delimitadores. Mediante las casillas de verificación mostradas para este control, puede especificar los caracteres, como la coma (,), que desea que definan los límites de los campos del archivo. Asimismo, puede especificar más de un delimitador, como " | " para los registros que utilizan varios delimitadores. El delimitador predeterminado es la coma.

Note: si la coma se ha definido también como el separador decimal, los valores predeterminados no funcionarán aquí. En aquellos casos en los que la coma sea el delimitador del campo y el separador decimal, seleccione **Otros** en la lista Delimitadores. A continuación, especifique una coma en el campo de entrada.

Seleccione **Permitir múltiples delimitadores vacíos** para tratar varios caracteres delimitadores vacíos adyacentes como un solo delimitador. Por ejemplo, si un valor de datos aparece seguido de cuatro espacios y otro valor de datos, este grupo será tratado como dos campos en lugar de cinco.

Líneas que explorar para columna y tipo: especifique cuántas líneas y columnas se van a explorar para los tipos de datos especificados.

Reconocer fechas y horas automáticamente: para que IBM SPSS Modeler intente reconocer automáticamente entradas de datos como fechas u horas, active esta casilla de verificación. Por ejemplo, significa que una entrada del tipo 07-11-1965 se identificará como fecha y una del tipo 02:35:58 se identificará como hora; sin embargo, las entradas ambiguas como 07111965 ó 023558 se mostrarán como enteros ya que no hay delimitadores entre los números.

Nota: Para evitar posibles problemas de datos cuando utilice archivos de datos de versiones anteriores de IBM SPSS Modeler, esta casilla se desactiva de forma predeterminada para la información guardada en versiones anteriores a 13.

Tratar corchetes como listas: si marca este recuadro de selección, los datos incluidos entre corchetes de apertura y cierre se tratarán como un único valor, incluso si el contenido incluye caracteres delimitadores como comas y comillas dobles. Por ejemplo, esto podría incluir dos o tres datos geoespaciales, en los que las coordenadas incluidas entre corchetes se procesan como un único elemento de lista. Si desea obtener más información, consulte “Importación de los datos geoespaciales al nodo Archivo variable”

Comillas. Si utiliza las listas desplegables, podrá especificar las comillas simples y dobles tratadas en la importación. Puede seleccionar **Descartar** todas las comillas, **Incluir como texto** incluyéndolas en el valor del campo o **Emparejar y descartar** para hacer coincidir las parejas de comillas y eliminarlas. Si alguna comilla no coincide, recibirá un mensaje de error. Tanto **Descartar** como **Emparejar y descartar** almacenan el valor del campo (sin comillas) como una cadena.

Mientras trabaja en este cuadro de diálogo, pulse en **Actualizar** en cualquier momento para volver a cargar los campos desde el origen de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de origen o cuando se trabaja entre pestañas en el cuadro de diálogo.

Importación de los datos geoespaciales al nodo Archivo variable

Si el nodo contiene datos geoespaciales, se ha creado como una exportación desde un archivo sin formato y se utiliza en la misma ruta en la que se ha creado, el nodo retiene los metadatos geoespaciales y no se necesitan pasos de configuración adicionales.

Sin embargo, si el nodo se ha exportado y utilizado en una ruta diferente, los datos de la lista geoespacial se convierten automáticamente a un formato de cadena. Para restaurar el tipo de almacenamiento de lista y los metadatos geoespaciales asociados deberá realizar algunos pasos adicionales.

Para obtener más información acerca de las listas, consulte “Almacenamiento de listas y niveles de medición asociados” en la página 11.

Para obtener información acerca de los detalles que puede establecer como metadatos geoespaciales, consulte “Subniveles de medición geoespacial” en la página 129.

Para configurar los metadatos geoespaciales, siga estos pasos.

1. En el separador Archivo del nodo Archivo variable, seleccione el recuadro de selección **Tratar corchetes como listas**. La selección de este recuadro de selección significa que los datos incluidos entre los corchetes de apertura y cierre se tratan como un único valor, incluso si el contenido incluye caracteres delimitadores, tales como las comas y las comillas dobles. Si no se selecciona este recuadro de selección significa que sus datos se leerán como el tipo de almacenamiento de cadena y cualquier coma en el campo se procesará como un delimitador, por lo que su estructura de datos se interpretará de forma incorrecta.
2. Si los datos incluye comillas simples o dobles, seleccione la opción **Emparejar y descartar** en los campos **Comillas simples** y **Comillas dobles**, según corresponda.
3. En el separador Datos del nodo Archivo variable, para los campos de datos geoespaciales, seleccione el recuadro de selección **Sustituir** y cambie el tipo de **Almacenamiento** de una cadena a una lista.

4. De forma predeterminada, el tipo de **Almacenamiento** de lista se establece como una *Lista de reales* y el tipo de almacenamiento de valores subyacentes del campo de lista se establece en *Real*. Para cambiar el tipo de almacenamiento de valores subyacentes o la profundidad, pulse **Especificar...** para visualizar el recuadro de subdiálogo Almacenamiento.
5. En el recuadro de subdiálogo Almacenamiento puede modificar los valores siguientes:
 - **Almacenamiento:** Especifique el tipo de almacenamiento general del campo de datos. De forma predeterminada, el tipo de almacenamiento se establece en Lista. Sin embargo, la lista desplegable contiene todos los otros tipos de almacenamiento (Cadena, Entero, Real, Fecha, Hora e Indicación de la hora). Si selecciona cualquier tipo de almacenamiento distinto a Lista, las opciones **Almacenamiento de valores** y **Profundidad** no están disponibles.
 - **Almacenamiento de valores:** Especifique los tipos de almacenamiento de los elementos de la lista, a diferencia de especificar el campo en su totalidad. Cuando importa campos geoespaciales, los únicos tipos de almacenamiento relevantes son Real y entero, siendo Real el valor predeterminado.
 - **Profundidad:** Especifique la profundidad del campo de lista. La profundidad necesaria depende del tipo del campo geoespacial y sigue este criterio:
 - Punto – 0
 - Cadena de líneas – 1
 - Polígono – 1
 - Multipunto – 1
 - Cadena multilínea – 2
 - Multipolígono – 2

Nota: Debe conocer el tipo del campo geoespacial que está volviendo a convertir en una lista y la profundidad necesaria para dicho tipo de campo. Si se establece incorrectamente esta información, no es posible utilizar el campo.
6. En el separador Tipos del nodo Archivo variable, para el campo de datos geoespaciales, asegúrese de que la casilla **Medición** contiene el nivel de medición correcto. Para cambiar el nivel de medición, en la casilla **Medición**, pulse **Especificar...** para visualizar el recuadro de diálogo Valores.
7. En el recuadro de diálogo Valores, se muestran **Medición**, **Almacenamiento** y **Profundidad** para la lista. Seleccione la opción **Especificar valores y etiquetas** y, desde la lista desplegable **Tipo**, seleccione el tipo correcto para **Medición**. En función del **Tipo**, es posible que se le soliciten más detalles, tales como si los datos representan 2 o 3 dimensiones y qué sistema de coordenadas se utiliza.

Nodo Archivo fijo

Puede utilizar los nodos Archivo fijo para importar los datos desde los archivos de texto de campo fijo (archivos cuyos campos no están delimitados pero empiezan en la misma posición y tienen una longitud fija). Los datos heredados o generados por la máquina se suelen almacenar en formato de campo fijo. Mediante la pestaña Archivo del nodo Archivo fijo, podrá especificar fácilmente la posición y longitud de las columnas en los datos.

Configuración de opciones para el nodo Archivo fijo

La pestaña Archivo del nodo Archivo fijo le permite introducir datos en IBM SPSS Modeler y especificar la posición de las columnas y la longitud de los registros. Con el panel de presentación preliminar de los datos en el centro del cuadro de diálogo, puede pulsar para añadir flechas que especifiquen los puntos de separación entre los campos.

Archivo. Especifique el nombre del archivo. Puede especificar un nombre de archivo o pulsar el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará cuando haya seleccionado un archivo y su contenido se mostrará con los delimitadores en el panel inferior.

El panel de presentación preliminar de los datos se puede utilizar para especificar la posición y longitud de la columna. La regla situada en la parte superior de la ventana de presentación preliminar ayuda a medir la longitud de las variables y especifica el punto de separación entre ellas. Puede especificar las líneas del punto de separación pulsando en el área de la regla por encima de los campos. Los puntos de separación se pueden mover arrastrándolos y se pueden descartar arrastrándolos fuera de la región de presentación preliminar de los datos.

- Cada línea del punto de separación añade de forma automática un nuevo campo a la tabla de campos inferior.
- Las posiciones de inicio que indican las flechas se añaden de forma automática a la columna Inicio de la tabla inferior.

Organizado por líneas. Indique si desea omitir el carácter de nueva línea al final de cada registro.

Omitir líneas de cabecera. Especifique las líneas que desea ignorar al principio del primer registro. Esto resulta útil para ignorar las cabeceras de columna.

Longitud de registro. Especifique el número de caracteres de cada registro.

Campo. Todos los campos definidos para este archivo de datos se muestran aquí. Hay dos formas de definir los campos:

- Especifique los campos de forma interactiva utilizando el panel anterior de presentación preliminar de los datos.
- Especifique los campos de forma manual añadiendo las filas de campos vacíos a la tabla inferior. Pulse en el botón a la derecha del panel de campos para añadir nuevos campos. A continuación, en el campo vacío, escriba un nombre de campo, una posición de inicio y una longitud. Estas opciones añadirán flechas de forma automática al panel de presentación preliminar de los datos, lo cual se puede ajustar fácilmente.

Para eliminar un campo definido previamente, seleccione el campo en la lista y pulse en el botón rojo de eliminación.

Inicio. Especifique la posición del primer carácter del campo. Por ejemplo, si el segundo campo de un registro comienza en el decimosexto carácter, deberá introducir 16 como punto de partida.

Longitud. Especifique cuántos caracteres hay en el valor más largo de cada campo. Esto determina el punto de corte para el siguiente campo.

Eliminar espacios iniciales y finales. Seleccione descartar los espacios iniciales y finales en las cadenas de importación.

Nota: Las comparaciones entre cadenas que utilizan y no utilizan retrotracción SQL pueden generar diferentes resultados donde existen espacios finales.

Caracteres no válidos. Seleccione **Descartar** para eliminar los caracteres no válidos de la entrada de datos. Seleccione **Reemplazar con** para reemplazar los caracteres no válidos con el símbolo especificado (solamente un carácter). Los caracteres no válidos son los caracteres nulos (0) o los que no existen en la codificación actual.

Codificación. Especifica el método de codificación de textos utilizado. Puede elegir entre valor predeterminado del sistema, el valor predeterminado de la ruta o UTF-8.

- El valor predeterminado del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta predeterminada se especifica en el cuadro de diálogo Propiedades de ruta.

Símbolo decimal. Seleccione el tipo de separador decimal utilizado en el origen de datos. El **Valor predeterminado de la ruta** es el carácter seleccionado de la pestaña Opciones del cuadro de diálogo de propiedades de la ruta. En el resto de los casos, seleccione **Punto (.)** o **Coma (,)** para leer todos los datos en este cuadro de diálogo utilizando el carácter elegido como el separador decimal.

Reconocer automáticamente fechas y horas. Para que IBM SPSS Modeler intente reconocer automáticamente entradas de datos como fechas u horas, active esta casilla de verificación. Por ejemplo, significa que una entrada del tipo 07-11-1965 se identificará como fecha y una del tipo 02:35:58 se identificará como hora; sin embargo, las entradas ambiguas como 07111965 ó 023558 se mostrarán como enteros ya que no hay delimitadores entre los números.

Note: para evitar posibles problemas de datos cuando utilice archivos de datos de versiones anteriores de IBM SPSS Modeler, esta casilla se desactiva de forma predeterminada para la información guardada en versiones anteriores a 13.

Líneas que explorar en busca del tipo. Especifique cuántas líneas se van a explorar para los tipos de datos especificados.

Mientras trabaja en este cuadro de diálogo, pulse en **Actualizar** en cualquier momento para volver a cargar los campos desde el origen de datos. Esto resulta útil cuando se modifican las conexiones de datos en el nodo de origen o cuando se trabaja entre pestañas en el cuadro de diálogo.

Data Collection Nodo

Los nodos de origen de Data Collection importan datos de encuesta basándose en Survey Reporter Developer Kit de IBM SPSS Data Collection utilizado por el software de investigación de mercados de IBM Corp.. Este formato distingue los **datos de casos** (las respuestas reales a las preguntas recogidas durante una encuesta), de los **metadatos** que describen cómo se recopilan y organizan los datos de casos. Los metadatos constan de información como textos de preguntas, nombres de variables y descripciones, definiciones de variable de respuestas múltiples, traducciones de las cadenas de texto y la definición de la estructura de los datos de casos.

Note: este nodo requiere IBM SPSS Data Collection Survey Reporter Developer Kit, que se distribuye con los productos de software de IBM SPSS Data Collection en IBM Corp.. Para obtener más información, consulte la página Web de IBM SPSS Data Collection en . Además de la instalación de Developer Kit, no se requieren configuraciones adicionales.

Comentarios

- Los datos de encuesta se leen solamente desde el formato VDATA tabular y plano o desde los orígenes en el formato HDATA jerárquico si incluyen un origen de metadatos (requiere IBM SPSS Data Collection 4.5 o superiores).
- Los tipos se instancian con la información de los metadatos.
- Cuando se importan los datos de encuesta a IBM SPSS Modeler, las preguntas se representan como campos, con un registro para cada encuestado.

Opciones de archivo de importación de Data Collection

La pestaña Archivo del nodo Data Collection permite especificar opciones para los metadatos y los datos de casos que desea importar.

Configuración de metadatos

Note: Para ver la lista completa de tipos de archivos de proveedores, es necesario que instale Survey Reporter Developer Kit de IBM SPSS Data Collection, disponible con el software IBM SPSS Data Collection. Si desea obtener más información, consulte la página Web de IBM SPSS Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>.

Proveedor de metadatos. Los datos de encuesta se pueden importar desde distintos formatos compatibles con el software IBM SPSS Data Collection Survey Reporter Developer Kit. Entre los tipos de proveedor disponibles se encuentran:

- **DataCollectionMDD.** Lee metadatos de un archivo de definición de cuestionario (*.mdd*). Es el formato estándar de IBM SPSS Data Collection Data Model.
- **Base de datos de ADO.** Lee datos de casos y metadatos de archivos ADO. Especifique el nombre y la ubicación del archivo *.adoinfo* que contiene los metadatos. El nombre interno de este DSC es *mrADODsc*.
- **Base de datos de In2data.** Lee metadatos y datos de casos de In2data. El nombre interno de este DSC es *mrI2dDsc*.
- **Archivo de registro de recopilación de datos.** Lee metadatos de un archivo de registro estándar de IBM SPSS Data Collection. Normalmente, los archivos de registro tienen una extensión de nombre de archivo *.tmp*. Sin embargo, algunos archivos de registro pueden tener otra extensión de nombre de archivo. Si es necesario, puede cambiar el nombre del archivo para que tenga una extensión de nombre de archivo *.tmp*. El nombre interno de este DSC es *mrLogDsc*.
- **Archivo de definiciones de Quancept.** Convierte los metadatos en un script de Quancept. Especifique el nombre del archivo *.qdi* de Quancept. El nombre interno de este DSC es *mrQdiDrsDsc*.
- **Base de datos de Quanvert.** Lee metadatos y datos de casos de Quanvert. Especifique el nombre y la ubicación del archivo *.qinfo* o el archivo *.pkd*. El nombre interno de este DSC es *mrQvDsc*.
- **Base de datos de participación de recopilación de datos.** Lee tablas de muestra y de historial del proyecto y crea variables categóricas derivadas correspondientes a las columnas de esas tablas. El nombre interno de este DSC es *mrSampleReportingMDSC*.
- **Archivo Statistics.** Lee metadatos y datos de casos de un archivo IBM SPSS Statistics *.sav* de SPSS. Escribe datos de casos en un archivo IBM SPSS Statistics *.sav* para su análisis en IBM SPSS Statistics. Escribe metadatos desde un archivo *.sav* de IBM SPSS Statistics a un archivo *.mdd*. El nombre interno de este DSC es *mrSavDsc*.
- **Archivo Surveycraft.** Lee metadatos y datos de caso de SurveyCraft. Especifique el nombre del archivo *.vq* de SurveyCraft. El nombre interno de este DSC es *mrSCDsc*.
- **Archivo de scripts de recopilación de datos.** Lee metadatos en un archivo *mrScriptMetadata*. Normalmente, estos archivos tienen una extensión de nombre de archivo *.mdd* o *.dms*. El nombre interno de este DSC es *mrScriptMDSC*.
- **Archivo XML triple-S** Lee metadatos de un archivo triple-S en formato XML. El nombre interno de este DSC es *mrTripleSDsc*.

Propiedades de metadatos. Opcionalmente, seleccione **Propiedades** para especificar la versión de encuesta que se va a importar así como el idioma, contexto y tipo de etiqueta que se van a utilizar. Consulte el tema “Propiedades de metadatos de importación de IBM SPSS Data Collection” en la página 34 para obtener más información.

Configuración de datos de casos

Note: Para ver la lista completa de tipos de archivos de proveedores, es necesario que instale Survey Reporter Developer Kit de IBM SPSS Data Collection, disponible con el software IBM SPSS Data Collection. Si desea obtener más información, consulte la página Web de IBM SPSS Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>.

Obtener configuración de datos de casos. Al leer metadatos solamente desde archivos *.mdd*, pulse en **Obtener configuración de datos de casos** para determinar los orígenes de datos de casos que están asociadas a los metadatos seleccionados, junto con la configuración específica necesaria para acceder a un origen determinado. Esta opción solamente se encuentra disponible para archivos *.mdd*.

Proveedor de datos de casos. Los siguientes tipos de proveedor son compatibles:

- **Base de datos de ADO.** Lee datos de casos mediante la interfaz de Microsoft ADO. Seleccione UDL de OLE-DB para el tipo de datos de casos y especifique una cadena de conexión en el campo UDL de casos de datos. Consulte el tema “Cadena de conexión a la base de datos” en la página 35 para obtener más información. El nombre interno de este componente es *mrADODsc*.
- **Archivo de texto delimitado (Excel).** Lee datos de casos de un archivo delimitado por comas (.CSV), de manera que se pueda abrir con Excel. El nombre interno es *mrCsvDsc*.
- **Archivo de datos de recopilación de datos.** Lee datos de casos de un archivo de formato de datos de IBM SPSS Data Collection nativo (IBM SPSS Data Collection 4.5 en adelante). El nombre interno es *mrDataFileDsc*.
- **Base de datos de In2data.** Lee metadatos y datos de casos de un archivo de base de datos de In2data (.i2d). El nombre interno es *mrI2dDsc*.
- **Archivo de registro de recopilación de datos.** Lee datos de casos de un archivo de registro de IBM SPSS Data Collection estándar. Normalmente, los archivos de registro tienen una extensión de nombre de archivo .tmp. Sin embargo, algunos archivos de registro pueden tener otra extensión de nombre de archivo. Si es necesario, puede cambiar el nombre del archivo para que tenga una extensión de nombre de archivo .tmp. El nombre interno es *mrLogDsc*.
- **Archivo de datos de Quantum.** Lee datos de casos de cualquier archivo ASCII con formato Quantum (.dat). El nombre interno es *mrPunchDsc*.
- **Archivo de datos de Quancept.** Lee datos de casos de un archivo .drs, .drz o .dru de Quancept. El nombre interno es *mrQdiDrsDsc*.
- **Base de datos de Quanvert.** Lee datos de casos de un archivo *qvinfo* o .pkd de Quanvert. El nombre interno es *mrQvDsc*.
- **Base de datos de recopilación de datos (MS SQL Server).** Lee datos de casos en una base de datos relacional de Microsoft SQL Server. Consulte el tema “Cadena de conexión a la base de datos” en la página 35 para obtener más información. El nombre interno es *mrRdbDsc2*.
- **Archivo Statistics.** Lee datos de casos desde un archivo IBM SPSS Statistics .sav. El nombre interno es *mrSavDsc*.
- **Archivo Surveycraft.** Lee datos de casos de un archivo .qdt de SurveyCraft. Tanto los archivos .vq como los archivos .qdt deben estar en el mismo directorio, con acceso de lectura y escritura para ambos archivos. Los archivos no se crean así de forma predeterminada al utilizar SurveyCraft, por lo que se deben mover para importar los datos de SurveyCraft. El nombre interno es *mrScDsc*.
- **Archivo de datos triple-S.** Lee datos de casos de un archivo de datos Triple-S, en formato de longitud fija o delimitado por comas. El nombre interno es *mr TripleDsc*.
- **XML de recopilación de datos.** Lee datos de casos de un archivo de datos XML de IBM SPSS Data Collection. Normalmente, este formato se puede utilizar para transferir datos de casos de una ubicación a otra. El nombre interno es *mrXmlDsc*.

Tipo de datos de casos. Especifica si los datos de casos se leen desde un archivo, carpeta, UDL de OLE-DB o DSN de ODBC, y actualiza las opciones del cuadro de diálogo que correspondan. Las opciones válidas dependen del tipo de proveedor. Puede especificar las opciones de conexión ODBC u OLE-DB para los proveedores de bases de datos. Consulte el tema “Cadena de conexión a la base de datos” en la página 35 para obtener más información.

Proyecto de datos de casos. Al leer datos de casos de una base de datos de IBM SPSS Data Collection, puede escribir el nombre del proyecto. Para el resto de tipos de datos de casos, esta configuración se deberá dejar en blanco.

Importación de variable

Importar variables del sistema. Especifica si las variables del sistema se importan, incluidas las variables que indican el estado de intervención (en proceso, completado, fecha de finalización, etc.). Puede seleccionar **Ninguna**, **Todos** o **Común**.

Importar variables "Códigos". Controla la importación de variables que representan los códigos utilizados para "otras" respuestas abiertas para las variables categóricas.

Importar variables "SourceFile". Controla la importación de variables que contienen nombres de archivo de imágenes de respuestas exploradas.

Importar variables de respuestas múltiples como. Las variables de respuestas múltiples se pueden importar como varios campos de marcas (un conjunto de dicotomía múltiple), que es el método predeterminado de las nuevas rutas. Las rutas creadas en las versiones de IBM SPSS Modeler anteriores a la 12.0 importaban respuestas múltiples en un solo campo, con valores separados por comas. El método anterior aún se admite para permitir que las rutas existentes se ejecuten igual que anteriormente, pero se recomienda la actualización de las rutas más antiguas para utilizar el nuevo método. Consulte el tema "Importación de conjuntos de respuestas múltiples" en la página 35 para obtener más información.

Propiedades de metadatos de importación de IBM SPSS Data Collection

Al importar datos de encuesta de IBM SPSS Data Collection en el cuadro de diálogo Propiedades de metadatos, puede especificar la versión de encuesta que se va a importar así como el idioma, contexto y tipo de etiqueta que se van a utilizar. Tenga en cuenta que sólo se puede importar un idioma, contexto y tipo de etiqueta al mismo tiempo.

Versión. Cada versión de encuesta puede considerarse como una instantánea de los metadatos que se utilizan para recopilar un conjunto de datos de casos determinado. Como se realizan cambios en los cuestionarios, se pueden crear varias versiones. Puede importar la última versión, todas las versiones o una versión específica.

- **Todas las versiones.** Seleccione esta opción si desea utilizar una combinación (superconjunto) de todas las versiones disponibles. (En ocasiones, esto se denomina superversión). Cuando existe un conflicto entre las versiones, las versiones más recientes suelen tener prioridad sobre las versiones anteriores. Por ejemplo, si una etiqueta de categoría es diferente en alguna de las versiones, se utilizará el texto de la última versión.
- **Última versión.** Seleccione esta opción si desea utilizar la versión más reciente.
- **Especificar versión.** Seleccione esta opción si desea utilizar una versión de encuesta determinada.

Seleccionar todas las versiones es útil cuando desea exportar datos de casos para más de una versión y se han hecho cambios en las definiciones de variable y categoría, lo que supone que los datos de casos recopilados con una versión no son válidos en otra versión. Seleccionar todas las versiones para las que desea exportar los datos de casos significa que podrá exportar los datos de casos recopilados con las distintas versiones al mismo tiempo sin que se produzcan errores de validación debido a las diferencias entre versiones. Sin embargo, según los cambios de versión, puede producirse algún error de validación.

Idioma. Se pueden almacenar preguntas y textos asociados en varios idiomas en los metadatos. Puede utilizar el idioma predeterminado para la encuesta o especificar un idioma determinado. Si un elemento no está disponible en el idioma especificado, se utilizará el valor predeterminado.

Contexto. Seleccione el contexto de usuario que desee utilizar. El contexto de usuario controla los textos que se muestran. Por ejemplo, seleccione **Pregunta** para mostrar textos de preguntas o **Análisis** para mostrar textos más cortos adecuados para visualizarlos al analizar los datos.

Tipo de etiqueta. Enumera los tipos de etiqueta que se han definido. El valor predeterminado es **etiqueta**, que se utiliza para textos de preguntas en el contexto de usuario Pregunta y las descripciones de variables en el contexto de usuario Análisis. Se pueden definir otros tipos de etiquetas para las instrucciones, descripciones, etc.

Cadena de conexión a la base de datos

Al utilizar el nodo IBM SPSS Data Collection para importar datos de casos de una base de datos a través de OLE-DB o ODBC, seleccione **Edición** en la pestaña Archivo para acceder al cuadro de diálogo de cadena de conexión, que permite personalizar la cadena de conexión que se pasa al proveedor para ajustar la conexión.

Propiedades avanzadas

Al utilizar el nodo IBM SPSS Data Collection para importar datos de casos de una base de datos que requiere un inicio de sesión explícito, seleccione **Avanzado** para proporcionar un ID de usuario y una contraseña para acceder al origen de datos.

Importación de conjuntos de respuestas múltiples

Las variables de respuestas múltiples se pueden importar de IBM SPSS Data Collection como conjuntos de dicotomía múltiple con un campo de marca independiente para cada posible valor de la variable. Por ejemplo, si se pide a los encuestados que seleccionen qué museos han visitado de una lista, el conjunto incluiría un campo de marca independiente para cada museo enumerado.

Tras importar los datos, puede añadir o editar conjuntos de respuestas múltiples desde cualquier nodo que incluya una pestaña Filtro. Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.

Importación de respuestas múltiples en un único campo (para rutas creadas en versiones anteriores)

En las versiones de IBM SPSS Modeler anteriores en lugar de importar respuestas múltiples como se ha descrito antes, se importaban en un solo campo, con valores separados por comas. Este método aún se admite para compatibilizar las rutas existentes, pero se recomienda que cualquier ruta se actualice para utilizar el nuevo método.

Notas de importación de columna de IBM SPSS Data Collection

Las columnas de los datos de IBM SPSS Data Collection se leen en IBM SPSS Modeler tal y como se resume en la siguiente tabla.

Tabla 4. Resumen de importación de columnas de IBM SPSS Data Collection

Tipo de columna de IBM SPSS Data Collection	Almacenamiento de IBM SPSS Modeler	Nivel de medición
Marca booleana (sí/no)	Cadena	Marca (valores 0 y 1)
Categorico	Cadena	Nominal
Marca de fecha o tiempo	Marca de tiempo	Continuo
Dobles (valor de punto flotante en un rango especificado)	Real	Continuo
Largo (valor de punto flotante en un rango especificado)	Entero	Continuo
Texto (descripción de texto libre)	Cadena	Sin tipo
Nivel (indica cuadrículas o bucles de una pregunta)	No sucede en VDATA y no se importa a IBM SPSS Modeler	
Objeto (datos binarios como un facsímil que muestra texto escrito a mano o una grabación de voz)	No se importa a IBM SPSS Modeler	
Ninguno (tipo desconocido)	No se importa a IBM SPSS Modeler	

Tabla 4. Resumen de importación de columnas de IBM SPSS Data Collection (continuación)

Tipo de columna de IBM SPSS Data Collection	Almacenamiento de IBM SPSS Modeler	Nivel de medición
Columna de encuestados en serie (asocia un ID exclusivo a cada encuestado)	Entero	Sin tipo

Para evitar posibles incoherencias entre las etiquetas de valor que se leen desde los metadatos y los valores reales, todos los valores de los metadatos se convierten a minúsculas. Por ejemplo, el valor de la etiqueta *E1720_years* se convertiría a *e1720_years*.

Nodo de origen de IBM Cognos BI

El nodo de origen de IBM Cognos BI le permite introducir los datos de la base de datos de Cognos BI o informes de lista individuales en la sesión de minería de datos. De este modo, puede combinar las características de inteligencia empresarial de Cognos con las capacidades de análisis predictivo de IBM SPSS Modeler. Puede importar datos relacionales, relacionales modelados dimensionalmente (DMR) y OLAP.

Desde una conexión de servidor de Cognos, primero seleccione una ubicación a la que importar datos o informes. Una ubicación contiene un modelo de Cognos y todas las carpetas, consultas, informes, vistas, atajos, URL y definiciones de trabajo asociados a dicho modelo. Un modelo de Cognos define reglas empresariales, descripciones de datos, relaciones de datos, dimensiones y jerarquías empresariales y otras tareas administrativas.

Si está importando datos, seleccione los objetos que quiere importar desde el paquete seleccionado. Entre los objetos que puede importar se incluyen asuntos de consulta (que representan tablas de bases de datos) o elementos de consulta individuales (que representan las columnas de tabla). Consulte el tema "Iconos de objeto de cognos" para obtener más información.

Si el paquete tiene filtros definidos, puede importar uno o más de ellos. Si un filtro que ha importado está asociado con datos importados, ese filtro se aplica antes de importar los datos. *Note:* los datos que se van a importar deben tener el formato UTF-8.

Si importa un informe, seleccione un paquete o una carpeta de un paquete que contenga uno o más informes. A continuación, seleccione el informe individual que desee importar. *Nota:* solo se pueden importar los informes de lista individuales; las listas múltiples no son compatibles.

Si los parámetros se han definido, para un objeto de datos o un informe, puede especificar valores para estos parámetros antes de importar el objeto o el informe.

Iconos de objeto de cognos

Los diferentes tipos de objeto que puede importar de una base de datos de Cognos BI se representan por diferentes iconos, tal y como muestra la siguiente tabla.

Tabla 5. Iconos de objeto de cognos.













Icono	Objeto
	Paquete
	Espacio de nombre
	Asunto de consulta

Tabla 5. Iconos de objeto de cognos (continuación).

Icono	Objeto
	Elemento de consulta
	Dimensión de medida
	Medida
	Dimensión
	Jerarquía de nivel
	Nivel
	Filtro
	Informe
	Cálculo autónomo

Importación de datos de Cognos

Para importar datos de una base de datos de IBM Cognos BI, en la pestaña Datos del cuadro de diálogo IBM Cognos BI, asegúrese de que **Modo** está establecido en **Datos**.

Conexión. Pulse en el botón **Editar** para mostrar un cuadro de diálogo donde puede definir los detalles de una nueva conexión de Cognos desde la que importar datos o informes. Si ya ha iniciado sesión en un servidor de Cognos mediante IBM SPSS Modeler, también puede editar los detalles de la conexión actual. Consulte el tema “Conexiones de Cognos” en la página 38 para obtener más información.

Ubicación. Cuando haya establecido la conexión de servidor de Cognos, pulse en el botón **Editar** junto a este campo para mostrar una lista de paquetes disponibles desde los que importar contenido. Consulte el tema “Selección de ubicación de Cognos” en la página 39 para obtener más información.

Contenido. Muestra el nombre del paquete seleccionado, junto con los espacios de nombre asociados al paquete. Pulse dos veces en un espacio de nombre para mostrar los objetos que puede importar. Los diferentes tipos de objeto se muestran con diferentes iconos. Consulte el tema “Iconos de objeto de cognos” en la página 36 para obtener más información.

Para seleccionar un objeto para importar, seleccione el objeto y pulse la flecha superior de las dos que se encuentran a la derecha para mover el objeto al panel **Campos que van a importarse**. Al seleccionar un asunto de consulta se importan todos sus elementos de consulta. Si pulsa dos veces en un asunto de consulta, lo expandirá, de modo que podrá seleccionar uno o más de sus elementos de consulta individuales. Puede realizar varias selecciones con Ctrl-pulsar (seleccionar elementos individuales), Mayús-pulsar (seleccionar un bloque de elementos) y Ctrl-A (seleccionar todos los elementos).

Para seleccionar un filtro que se aplicará (si el paquete tiene filtros definidos), navegue hasta el filtro en el panel de contenido, seleccione el filtro y pulse la flecha inferior de las dos que se encuentran a la derecha para mover el filtro al panel **Filtros que se aplicarán**. Puede realizar múltiples selecciones dejando pulsada la tecla Ctrl (para seleccionar filtros individuales) y la tecla Mayús (para seleccionar un bloque de filtros).

Campos que van a importarse. Enumera los objetos de base de datos que ha seleccionado para importar a IBM SPSS Modeler para su procesamiento. Si ya no necesita un objeto específico, selecciónelo y pulse en la flecha izquierda para devolverlo al panel **Contenido**. Puede realizar varias selecciones del mismo modo que para **Contenido**.

Filtros que se aplicarán. Enumera los filtros que haya seleccionado para aplicar a los datos antes de importarse. Si ya no necesita un filtro concreto específico, selecciónelo y pulse en la flecha izquierda para devolverlo al panel **Contenido**. Puede realizar varias selecciones del mismo modo que para **Contenido**.

Parámetros. Si este botón está activado, el objeto seleccionado tiene parámetros definidos. Puede utilizar los parámetros para realizar ajustes (por ejemplo, realizar un cálculo parametrizado) antes de importar los datos. Si los parámetros se definen pero no se proporciona el parámetro predeterminado, el botón muestra un triángulo de advertencia. Pulse el botón para mostrar los parámetros y, opcionalmente, editarlos. Si el botón está desactivado, el informe no tiene parámetros definidos.

Agregar datos antes de importar. Seleccione esta casilla si desea importar datos agregados en lugar de datos sin procesar.

Importación de informes de Cognos

Para importar un informe predefinido de una base de datos de IBM Cognos BI, en la pestaña Datos del cuadro de diálogo IBM Cognos BI, asegúrese de que **Modo** está establecido en **Informe**. *Nota:* solo se pueden importar los informes de lista individuales; las listas múltiples no son compatibles.

Conexión. Pulse en el botón **Editar** para mostrar un cuadro de diálogo donde puede definir los detalles de una nueva conexión de Cognos desde la que importar datos o informes. Si ya ha iniciado sesión en un servidor de Cognos mediante IBM SPSS Modeler, también puede editar los detalles de la conexión actual. Consulte el tema “Conexiones de Cognos” para obtener más información.

Ubicación. Cuando haya establecido la conexión de servidor de Cognos, pulse en el botón **Editar** junto a este campo para mostrar una lista de paquetes disponibles desde los que importar contenido. Consulte el tema “Selección de ubicación de Cognos” en la página 39 para obtener más información.

Contenido. Muestra el nombre del paquete o de la carpeta seleccionada que contiene los informes. Desplácese hasta un informe específico, selecciónelo y pulse la flecha derecha para traer el informe al campo **Informe que se importará**.

Informe que se importará. Indica el informe que ha seleccionado para importar a IBM SPSS Modeler. Si ya no necesita el informe, selecciónelo y pulse en la flecha izquierda para devolverlo al panel **Contenido** o traer un informe diferente a este campo.

Parámetros. Si este botón está activado, el informe seleccionado tiene los parámetros definidos. Puede utilizar parámetros para realizar ajustes antes de importar el informe (por ejemplo, especificando una fecha de inicio y de fin para los datos del informe). Si los parámetros se definen pero no se proporciona el parámetro predeterminado, el botón muestra un triángulo de advertencia. Pulse el botón para mostrar los parámetros y, opcionalmente, editarlos. Si el botón está desactivado, el informe no tiene parámetros definidos.

Conexiones de Cognos

El cuadro de diálogo Conexiones de Cognos le permite seleccionar el servidor de Cognos BI desde el que importar o exportar objetos de base de datos.

URL del servidor Cognos. Escriba la URL del servidor de Cognos BI desde la que realizará la importación o exportación. Este es el valor de la propiedad de entorno "URI de envío externo" de la configuración IBM Cognos del servidor de Cognos BI. Póngase en contacto con el administrador del sistema de Cognos si no está seguro de qué URL utilizar.

Modo. Seleccione **Establecer credenciales** si desea registrarse con un espacio de nombre, nombre de usuario y contraseña de Cognos específicos (por ejemplo, como administrador). Seleccione **Utilizar conexión anónima** para registrarse sin credenciales de usuario, en cuyo caso no cumplimentará los otros campos.

ID de espacio de nombres. Especifique el proveedor de autenticación de seguridad de Cognos utilizado para registrarse en el servidor. El proveedor de autenticación se utiliza para definir y mantener usuarios, grupos y roles, así como para controlar el proceso de autenticación. Tenga en cuenta que es el ID de espacio de nombres, no el nombre del espacio de nombres (el ID no siempre es igual que el Nombre).

Nombre de usuario. Introduzca el nombre de usuario de Cognos con el que se inicia sesión en el servidor.

Contraseña. Introduzca la contraseña asociada al nombre de usuario especificado.

Guardar como valor predeterminado. Pulse en este botón para almacenar esta configuración como predeterminada, para no tener que volver a introducirla cada vez que abra el nodo.

Selección de ubicación de Cognos

El cuadro de diálogo Especificar ubicación permite seleccionar un paquete de Cognos desde el que importar datos o un paquete o carpeta desde el que importar informes.

Carpetas públicas. Si importa datos, enumera los paquetes y carpetas disponibles desde el servidor seleccionado. Seleccione el paquete que quiera utilizar y pulse en **Aceptar**. Solamente puede seleccionar un paquete por cada nodo de origen de Cognos BI.

Si importa informes, enumera los paquetes y carpetas que contienen los informes disponibles desde el servidor seleccionado. Seleccione un paquete o carpeta de informe y pulse en **Aceptar**. Solamente puede seleccionar un paquete o carpeta de informe por nodo de origen de Cognos BI, aunque las carpetas de informe pueden contener otras carpetas de informes e informes individuales.

Especificación de parámetros de datos o informes

Si los parámetros se han definido en Cognos BI, para un objeto de datos o un informe, puede especificar valores para estos parámetros antes de importar el objeto o el informe. Un ejemplo de parámetros para un informe podría ser las fechas de inicio y fin de los contenidos del informe.

Nombre. El nombre del parámetro, tal y como se especifica en la base de datos de Cognos BI.

Tipo. Una descripción del parámetro.

Valor. El valor que se asignará al parámetro. Para introducir o editar un valor, pulse dos veces en su casilla de la tabla. Los valores no se validan aquí, por lo que cualquier valor no válido se detecta en el periodo de ejecución.

Eliminar automáticamente de la tabla de parámetros no válidos. Esta opción se selecciona de forma predeterminada y eliminará cualquier parámetro no válido encontrado dentro del objeto de datos o informe.

Nodo Origen de IBM Cognos TM1

El nodo Origen de IBM Cognos TM1 le permite introducir datos de Cognos TM1 en una sesión de minería de datos. De este modo, se pueden combinar las características de planificación empresarial de Cognos con las capacidades de análisis predictivo de IBM SPSS Modeler. Puede importarse una versión aplanada del cubo OLAP multidimensional.

Nota: El usuario de TM1 necesita los permisos siguientes: Privilegio de escritura de cubos, Privilegio de lectura de procesos SPSS, Privilegio de lectura de dimensiones, Privilegio de escritura de elementos de dimensión y Privilegio de escritura de archivos .b1b.

Hay que modificar los datos de TM1 antes de importar los datos.

Nota: los datos que se van a importar deben tener el formato UTF-8.

Desde una conexión de gestión de rendimiento de IBM Cognos, seleccione en primer lugar un servidor de TM1 desde el que importará los datos; un servidor contiene uno o más cubos de TM1. Luego seleccione el cubo que necesite y dentro del cubo seleccione las columnas y filas que desee importar.

Nota: Para poder utilizar los nodos de origen o exportación de TM1 en SPSS Modeler, debe copiar los siguientes tres procesos desde IBM SPSS Modeler al servidor de TM1: ExportToSPSS.pro, ImportFromSPSS.pro y SPSSCreateNewMeasures.pro. Para añadir estos procesos al servidor TM1, debe copiarlos en el directorio de datos del servidor de TM1 y reiniciar el servidor de TM1. Estos archivos están disponibles desde el directorio siguiente: <directorio de instalación de Modeler>/ext/bin/pasw.tml/scripts.

Importación de datos de IBM Cognos TM1

Para importar datos desde una base de datos de IBM Cognos TM1, en la pestaña Datos del cuadro de diálogo IBM Cognos TM1 seleccione el sistema de gestión de rendimiento de TM1 adecuado y los detalles del servidor, cubo y datos asociados.

Nota: Antes de importar datos debe realizarse un procesamiento previo en TM1 para asegurar que los datos estén en un formato reconocible a IBM SPSS Modeler. Esto implica filtrar los datos utilizando el Editor de subconjuntos que la vista tenga el tamaño y la forma adecuados para la importación.

Tenga en cuenta que los valores de cero (0) que se importan desde TM1 se tratarán como valores "nulos" (TM1 no distingue entre los valores de cero y los valores en blanco). Asimismo, tenga en cuenta que los datos no numéricos (o metadatos) de las *dimensiones regulares* se pueden importar a IBM SPSS Modeler. Pero no está soportada la importación de *mediciones* no numéricas.

Sistema PM: Escriba el URL del sistema de gestión de rendimiento donde está instalado el servidor de TM1 al que desea conectarse. El sistema de gestión del rendimiento se define como un URL único para todos los servidores TM1. Desde este URL, se pueden descubrir y acceder a todos los servidores de IBM Cognos TM1 que están instalados y en ejecución en el entorno.

Servidor TM1: Cuando haya establecido la conexión con el sistema de gestión de rendimiento, seleccione el servidor que contiene los datos que desea importar y pulse **Iniciar sesión**. Si no se ha conectado anteriormente a este servidor, se le solicitará que especifique el **Nombre de usuario** y la **Contraseña**; de forma alternativa, puede seleccionar otro servidor.

Seleccione una vista de cubo de TM1 a importar: Muestra el nombre de los cubos del servidor TM1 desde el que se pueden importar los datos. Pulse dos veces un cubo para mostrar los datos de vista que se pueden importar.

Nota: Solo se pueden importar a IBM SPSS Modeler los cubos con una dimensión.

Para seleccionar datos a importar, seleccione la vista y pulse en la flecha derecha para moverla al panel **Ver para importar**. Si la vista deseada no está visible, haga doble clic en un cubo para expandir su lista de vistas.

Dimensión de columna: Lista el nombre de la dimensión de columna de los datos seleccionados para importar. Recorra la lista de niveles y seleccione el que necesite.

Dimensiones de fila. Lista el nombre de la dimensión de fila de los datos seleccionados para importar. Recorra la lista de niveles y seleccione el que necesite.

Dimensiones de contexto: Solo de visualización. Muestra las dimensiones de contexto que se relacionan con las filas y columnas seleccionadas.

Nodo de origen SAS

Note: Esta característica está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo de origen SAS le permite introducir los datos SAS en la sesión de minería de datos. Puede importar cuatro tipos de archivos:

- SAS para Windows/OS2 (.sd2)
- SAS para UNIX (.ssd)
- Archivo de transporte SAS (.tpt)
- SAS versión 7/8/9 (.sas7bdat)

Una vez se han importado los datos, se mantienen todas las variables y no se cambia ningún tipo de variable. Se han seleccionado todos los casos.

Configuración de opciones para el nodo de origen SAS

Importar. Seleccione el tipo de archivo SAS que se va a transportar. Puede seleccionar **SAS para Windows/OS2 (.sd2)**, **SAS para UNIX (.SSD)**, **Archivo de transporte SAS (.tpt)** o **SAS versión 7/8/9 (.sas7bdat)**.

Importar archivo. Especifique el nombre del archivo. Puede especificar un nombre de archivo o pulsar el botón de puntos suspensivos (...) para ir a la ubicación del archivo.

Miembro. Seleccione un miembro para importar desde el archivo de transporte SAS seleccionado anteriormente. Puede introducir un nombre de archivo o pulsar en **Seleccionar** para buscar en todos los miembros del archivo.

Leer formatos de usuario de un archivo de datos SAS. Seleccione leer formatos de usuario. Los archivos SAS almacenan datos y formatos de datos (como etiquetas de variables) en archivos diferentes. La mayoría de las veces, deseará importar también los formatos. No obstante, si tiene un conjunto de datos grande, puede que desee anular la selección de esta opción para ahorrar memoria.

Archivo de formato. Si se necesita un archivo de formato, este cuadro de diálogo estará activado. Puede especificar un nombre de archivo o pulsar el botón de puntos suspensivos (...) para ir a la ubicación del archivo.

Nombres de variable. Seleccione un método para gestionar etiquetas y nombres de variables en la importación desde un archivo SAS. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en IBM SPSS Modeler y se pueden volver a exportar para su uso en SAS.

- **Leer nombres y etiquetas.** Seleccione para leer las etiquetas y nombres de variables en IBM SPSS Modeler. De forma predeterminada, esta opción está seleccionada y los nombres de variables se muestran en el nodo Tipo. Las etiquetas se pueden mostrar en el generador de expresiones, gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta.
- **Leer etiquetas como nombres.** Seleccione para leer las etiquetas descriptivas de variables desde el archivo SAS en lugar de los nombres abreviados de campos y utilice estas etiquetas como nombres de variables en IBM SPSS Modeler.

Nodo de origen de Excel

El nodo de origen de Excel permite importar datos de cualquier versión de Microsoft Excel.

Tipo de archivo. Seleccione el tipo de archivo de Excel que está importando.

Importar archivo. Especifica el nombre y la ubicación del archivo de hoja de cálculo que se va a importar.

Usar rango con nombre. Permite especificar un rango con nombre de casillas tal y como se define en la hoja de trabajo de Excel. Pulse el botón de puntos suspensivos (...) para realizar una selección en la lista de hojas de trabajo disponibles. Si se utiliza un rango con nombre, no se podrá aplicar otra configuración de rango de datos y hoja de trabajo y, en consecuencia, estará desactivada.

Seleccionar hoja de trabajo. Especifique la hoja de trabajo que se va a importar, por índice o por nombre.

- **Por índice.** Especifique el valor de índice para la hoja de trabajo que desea importar, comenzando por 0 para la primera hoja de trabajo, 1 para la segunda hoja de trabajo y así sucesivamente.
- **Por nombre.** Especifique el nombre de la hoja de trabajo que desea importar. Pulse en el botón de puntos suspensivos (...) para seleccionar de la lista de hojas de trabajo disponibles.

Intervalo en hoja de trabajo. Puede importar datos que comiencen por la primera fila que no esté en blanco o con un rango de casillas explícito.

- **El intervalo comienza en la primera fila no en blanco.** Busca la primera casilla que no esté en blanco y la utiliza como la esquina superior izquierda del rango de datos.
- **Rango explícito de casillas.** Le permite especificar un intervalo explícito por fila y columna. Por ejemplo, para especificar el intervalo de Excel A1:D5, puede introducir A1 en el primer campo y D5 en el segundo (o, de forma alternativa, R1C1 y R5C4). Se devolverán todas las filas del rango especificado, incluidas las filas en blanco.

En filas en blanco. Si se encuentra más de una fila en blanco, puede seleccionar si desea **Parar de leer** o seleccionar **Devolver filas en blanco** para continuar leyendo todos los datos hasta el final de la hoja de trabajo, incluidas las filas en blanco.

Primera fila tiene nombres de columnas. Indica que la primera fila del rango concreto debería usarse como nombres de campo (columna). Si no se selecciona, los nombres de campo se generan automáticamente.

Almacenamiento de campo y Nivel de medición

Al leer valores de Excel, los campos con almacenamiento numérico se leen con un nivel de medición de *Continuo* de forma predeterminada y los campos de cadena se leen como *Nominales*. Puede cambiar el nivel de medición de forma manual (continuo en lugar de nominal) en la pestaña Tipo, pero el almacenamiento viene determinado de forma automática (aunque puede cambiarse mediante la función de conversión, como `to_integer`, en un nodo Rellenar o un nodo Derivar si es necesario). Consulte el tema “Configuración de almacenamiento de campos y formato” en la página 9 para obtener más información.

De forma predeterminada, los campos con una mezcla de valores numéricos y de cadena se leen como números, lo que significa que cualquier valor de cadena se establecerá en valores nulos (valores perdidos del sistema) en IBM SPSS Modeler. Esto ocurre porque, a diferencia de Excel, IBM SPSS Modeler no permite tipos de almacenamiento mezclado dentro de un campo. Para evitarlo, puede establecer de forma manual el formato de casillas a **Texto** en la hoja de cálculo de Excel, lo que hará que todos los valores (incluidos los números) se lean como cadenas.

Nodo Origen XML

Note: Esta característica está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo Origen XML le permite importar datos de un archivo en formato XML a una ruta de IBM SPSS Modeler. XML es un lenguaje estándar de intercambio de datos y, para la mayoría de organizaciones es el formato que eligen para ello. Por ejemplo, es posible que una agencia tributaria quiera analizar datos de devoluciones de impuestos que se han enviado en línea que tienen datos en formato XML (consulte <http://www.w3.org/standards/xml/>).

La importación de datos XML a una ruta de IBM SPSS Modeler le permite realizar una amplia gama de funciones de análisis predictivos en el origen. Los datos XML se analizan en formato tabular en el que las columnas se corresponden con los diferentes niveles de anidación de los elementos y atributos XML. Los elementos XML se muestran en formato de XPath (consulte <http://www.w3.org/TR/xpath20/>).

Lectura de un archivo único. De forma predeterminada, IBM SPSS Modeler lee un único archivo, que especifique en el campo **Origen de datos XML**.

Lectura de todos los archivos XML en un directorio. Seleccione esta opción si desea leer todos los archivos XML de un directorio concreto. Especifique la ubicación en el campo **Directorio** que aparece. Seleccione la casilla de verificación **Incluir subdirectorios** para leer más archivos XML de todos los subdirectorios del directorio especificado.

Origen de datos XML. Introduzca la ruta y el nombre completo del archivo de origen XML que desee importar o utilice el botón Examinar para buscar el archivo.

Esquema XML. (Opcional) Especifique la ruta completa y el nombre del archivo de un archivo XSD o DTD desde el que se leerá la estructura XML o utilice el botón Examinar para buscar el archivo. Si deja este campo en blanco, se leerá la estructura desde el archivo de origen XML. Un archivo XSD o DTD puede tener más de un elemento raíz. En este caso, cuando seleccione un campo diferente, aparecerá un cuadro de diálogo en el que podrá seleccionar el elemento raíz que desee utilizar. Consulte el tema “Selección entre múltiples elementos raíz” en la página 44 para obtener más información.

Estructura XML. Un árbol de jerarquía que muestra la estructura del archivo de origen XML (o el esquema, si lo ha especificado en el campo **Esquema XML**). Para definir un límite del registro, seleccione un elemento y pulse en el botón de flecha derecha para copiar el elemento en el campo **Registros**.

Mostrar atributos. Muestra u oculta los atributos de los elementos XML en el campo **estructura XML**.

Registros (expresión XPath). Muestra la sintaxis XPath de un elemento copiado desde el campo de estructura XML. Este elemento se resalta en la estructura XML y define los límites del registro. Cada vez que este elemento se encuentra en el archivo de origen se crea un nuevo registro. Si el campo está vacío, el primer elemento hijo de la raíz se utiliza como el límite del registro.

Leer todos los datos. De forma predeterminada, todos los datos del archivo de origen se leen en la ruta.

Especificar los datos que se leerán. Seleccione esta opción si desea importar elementos individuales, atributos o ambos. Si selecciona esta opción activará la pestaña Campos donde puede especificar los datos que desea importar.

Campos. Esta tabla contienen los elementos y atributos seleccionados para importar, si ha seleccionado la opción **Especificar los datos que se leerán**. Puede introducir la sintaxis de XPath o un elemento o atributo directamente en la columna XPath o seleccionar un elemento o atributo en la estructura XML y pulsar en el botón de flecha derecha para copiar el elemento en la tabla. Para copiar todos los elementos hijo y atributos de un elemento, seleccione el elemento en la estructura XML y pulse en el botón de doble flecha a la derecha.

- **XPath.** La sintaxis XPath de los elementos que se importarán.
- **Ubicación.** La ubicación de la estructura XML de los elementos que se importarán. **Ruta fijada** muestra la ruta del elemento relativo al elemento resaltado en la estructura XML (o el primer elemento hijo en la raíz, si no se ha resaltado ningún elemento). **Cualquier ubicación** denota un elemento del nombre asignado a cualquier ubicación de la estructura XML. **Personalizado** aparece si introduce una ubicación directamente en la columna XPath.

Selección entre múltiples elementos raíz

Mientras un archivo XML formado adecuadamente solamente puede tener un único elemento raíz, un archivo XSD o DTD puede contener múltiples raíces. Si una de las raíces coincide con la del archivo de origen XML, se utilizará ese elemento raíz, mientras que si no es así deberá seleccionar uno para utilizarlo.

Seleccione la raíz que desea mostrar. Seleccione el elemento raíz que desea utilizar. El predeterminado es el primer elemento raíz en la estructura XSD o DTD.

Eliminación de espacios no deseados de los datos de origen XML

Es posible utilizar la combinación de caracteres [CR] [LF] para implementar saltos de línea en los datos del origen XML. En algunos casos, estos saltos de línea pueden aparecer en medio de una cadena de texto, por ejemplo:

```
<description>Una descripción detallada de la creación de aplicaciones[CR] [LF]
con XML.</description>
```

Estos saltos de línea pueden no ser visibles cuando se abre el archivo en algunas aplicaciones, por ejemplo un navegador web. Sin embargo, cuando se leen los datos en la ruta a través del nodo de origen XML, los saltos de línea se convierten en una serie de caracteres de espacio.

Puede corregir esto usando un nodo Rellenar para quitar los espacios no deseados:

Aquí hay un ejemplo de cómo puede conseguirlo:

1. Conecte un nodo Rellenar al nodo Origen XML.
2. Abra el nodo Rellenar y utilice el selector de campos para seleccionar el campo con espacios no deseados.
3. Defina **Reemplazar** como **En función de una condición** y defina **Condición** como **true**.
4. En el campo **Reemplazar con**, introduzca `replace(" ", "", @FIELD)` y pulse Aceptar.
5. Conecte un nodo Tabla al nodo Rellenar y ejecute la ruta.

En los resultados del nodo Tabla, el texto aparecerá sin espacios adicionales.

Nodo Datos Usuario

El nodo Datos Usuario le facilita la creación de los datos sintéticos, bien partiendo desde cero o modificando los datos existentes. Esto resulta útil, por ejemplo, cuando desee crear un conjunto de datos de comprobación para el modelado.

Creación de datos partiendo desde cero

El nodo Datos Usuario está disponible en la paleta Orígenes y se puede añadir directamente al lienzo de rutas.

1. Pulse en la pestaña **Orígenes** de la paleta de nodos.
2. Arrastre y suelte o pulse dos veces para añadir el nodo Datos Usuario al lienzo de rutas.
3. Pulse dos veces para abrir el cuadro de diálogo y especifique los campos y valores.

Nota: los nodos Datos de usuario que se seleccionan desde la paleta Orígenes estarán completamente en blanco, sin campos ni información de datos. Esto permite crear datos sintéticos partiendo totalmente desde cero.

Generación de datos a partir de un origen de datos existente

También puede generar un nodo Datos Usuario a partir de un nodo no terminal de la ruta:

1. Indique en qué punto de la ruta desea reemplazar el nodo.
2. Pulse con el botón derecho en el nodo que introducirá sus datos en el nodo Datos Usuario y seleccione **Generar nodo de datos de usuario** en el menú.
3. El nodo Datos usuario aparece con todos los procesos que se encuentran debajo adjuntos, sustituyendo en este punto el nodo existente en la ruta de datos. Una vez generado, el nodo hereda toda la estructura de datos y la información sobre el tipo de campo (si está disponible) de los metadatos.

Note: si los datos no se han ejecutado a través de todos los nodos de la ruta, los nodos no estarán completamente instanciados, lo que significa que puede que los valores de datos y almacenamiento no estén disponibles cuando se realice la sustitución por un nodo Datos Usuario.

Configuración de opciones para el nodo Datos Usuario

El cuadro de diálogo para el nodo Datos Usuario contiene varias herramientas que puede utilizar para introducir valores y definir la estructura de datos para los datos sintéticos. Para un nodo generado, la tabla de la pestaña Datos contiene nombres de campos del origen de datos original. Para un nodo añadido desde la paleta Orígenes, la tabla está vacía. Si utiliza las opciones de la tabla, podrá realizar las siguientes tareas:

- Añadir nuevos campos mediante el botón Añadir un nuevo campo, situado en la parte derecha de la tabla.
- Cambiar el nombre de los campos existentes.
- Especificar el almacenamiento de datos para cada campo.
- Especificar valores.
- Cambiar el orden de los campos en la visualización.

Introducción de datos

Puede especificar valores o insertarlos desde el conjunto de datos original para cada campo, mediante el botón Selector de valores, situado en la parte derecha de la tabla. Consulte las reglas descritas a continuación, si desea obtener más información acerca de la especificación de valores. También puede dejar el campo vacío: los campos vacíos se rellenan con el espacio nulo del sistema (\$null\$).

Para especificar valores de cadena, simplemente introdúzcalos en la columna Valores separados por espacios:

Fred Ethel Martin

Cadenas que incluyen espacios que se pueden escribir entre comillas dobles:

"Bill Smith" "Fred Martin" "Jack Jones"

Para los campos numéricos, puede introducir varios valores de la misma forma (con espacios entre ellos):

10 12 14 16 18 20

O puede especificar la misma serie de números estableciendo los límites (10, 20) y los pasos entre ellos (2). Con este método, se escribirá lo siguiente:

10,20,2

Estos dos métodos se pueden combinar insertando uno dentro de otro, como se indica a continuación:

1 5 7 10,20,2 21 23

Esta entrada generará los siguientes valores:

1 5 7 10 12 14 16 18 20 21 23

Los valores de fecha y hora se pueden introducir mediante el formato predeterminado actual seleccionado en el cuadro de diálogo Propiedades de ruta, por ejemplo:

11:04:00 11:05:00 11:06:00

2007-03-14 2007-03-15 2007-03-16

Para los valores de marca de tiempo, que tienen un componente de fecha y hora, se deben utilizar comillas dobles:

"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"

Para obtener información adicional consulte los comentarios sobre el almacenamiento de datos a continuación.

Generar datos. Permite especificar cómo se generan los registros al ejecutar la ruta.

- **Todas las combinaciones.** Genera registros que contienen todas las combinaciones posibles de los valores de campo, por lo tanto, cada valor de campo aparecerá en varios registros. En ocasiones puede generar más datos de los deseados, por lo que es recomendable seguir este nodo con un nodo de muestra.
- **En el orden.** Genera registros en el orden en que se han especificado los valores de campo. Cada valor de campo aparece solamente en un registro. El número de registros será igual al número más grande de valores de un único campo. Cuando los campos tienen un número menor, se insertan valores no definidos (\$null\$).

Mostrar ejemplo

Por ejemplo, las entradas siguientes generarán los registros mostrados en los dos ejemplos de la tabla siguiente.

- **Edad.** 30,60,10
- **PS.** BAJA
- **Colesterol.** NORMAL ALTO
- **Medicamento.** (se deja vacío)

Tabla 6. Campo Generar datos definido como Todas las combinaciones.

Edad	PS	Colesterol	Medicamento
30	BAJA	NORMAL	\$null\$
30	BAJA	SUPERIOR	\$null\$
40	BAJA	NORMAL	\$null\$
40	BAJA	SUPERIOR	\$null\$
50	BAJA	NORMAL	\$null\$
50	BAJA	SUPERIOR	\$null\$
60	BAJA	NORMAL	\$null\$
60	BAJA	SUPERIOR	\$null\$

Tabla 7. Campo Generar datos definido como En el orden.

Edad	PS	Colesterol	Medicamento
30	BAJA	NORMAL	\$null\$
40	\$null\$	SUPERIOR	\$null\$
50	\$null\$	\$null\$	\$null\$
60	\$null\$	\$null\$	\$null\$

Almacenamiento de datos

Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es distinto del nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*. Aunque el almacenamiento se debe determinar en el origen, el nivel de medición se puede cambiar utilizando un nodo Tipo en cualquier punto de la ruta. Consulte el tema “Niveles de medición” en la página 127 para obtener más información.




Los tipos de almacenamiento disponibles son los siguientes:

- **Cadena:** Se utiliza para campos que contienen datos no numéricos, también llamados datos alfanuméricos. Una cadena puede incluir cualquier secuencia de caracteres, como *fred*, *Clase 2* o *1234*. Tenga en cuenta que los números de las cadenas no se pueden utilizar en los cálculos.
- **Entero:** Un campo cuyos valores son números enteros.
- **Real:** Los valores son números que pueden incluir decimales (no está limitado a números enteros). El formato de presentación se especifica en el cuadro de diálogo Propiedades de ruta y puede quedar anulado para campos individuales en un nodo Tipo (pestaña Formato).
- **Fecha:** Valores de fecha especificados en un formato estándar como año, mes y día (por ejemplo, 2007-09-26). El formato determinado se especifica en el cuadro de diálogo Propiedades de ruta.
- **Tiempo:** El tiempo medido como una duración. Por ejemplo, una llamada de servicio que dure 1 hora, 26 minutos y 38 segundos se puede representar como 01:26:38, según el formato de hora actual que se especifica en el cuadro de diálogo Propiedades de ruta.
- **Indicación de fecha y hora:** Valores que incluyen un componente de fecha y hora, por ejemplo 2007-09-26 09:04:00, en función de los formatos actuales de fecha y hora del cuadro de diálogo Propiedades de ruta. Tenga en cuenta que es posible que los valores de marca de tiempo se tengan que escribir entre comillas dobles para garantizar que se interpretan como un valor único en lugar de valores de fecha y hora diferentes. (Es aplicable, por ejemplo cuando introduce valores en un nodo Datos Usuario.)
- **Lista:** Introducido en SPSS Modeler versión 17, junto con los nuevos niveles de medición de Geoespacial y Colección, el campo de almacenamiento de lista contiene varios valores para un único registro. Existen versiones de lista de todos los otros tipos de almacenamiento.

Tabla 8. Iconos de tipo de almacenamiento de lista






Icono	Tipo de almacenamiento
[A]	Lista de cadenas
[0]	Lista de enteros
[.0]	Lista de reales
[0:]	Lista de tiempo

Tabla 8. Iconos de tipo de almacenamiento de lista (continuación)

Icono	Tipo de almacenamiento
	Lista de fechas
	Lista de indicaciones de fecha y hora
	Lista con una profundidad mayor que cero

Asimismo, para utilizarlo con el nivel de medición Colección, existen versiones de los siguientes niveles de medición.

Tabla 9. Iconos de nivel de medición de lista

Icono	Nivel de medición
	lista de continuos
	Lista de categóricos
	Lista de marcas
	Lista de nominales
	Lista de ordinales

Las listas se pueden importar a SPSS Modeler en uno de los tres nodos de origen (Analytic Server, Geoespacial o Archivo variable), o se pueden crear en sus rutas utilizando los nodos operativos de los campos Derivar o Rellenar.

Para obtener más información acerca de las listas y su interacción con los niveles de medición Colección y Geoespacial, consulte el tema “Almacenamiento de listas y niveles de medición asociados” en la página 11

Conversiones del almacenamiento. Puede convertir el almacenamiento de un campo mediante una serie de funciones de conversión, como `to_string` y `to_integer`, en un nodo Rellenar. Consulte el tema “Conversión del almacenamiento mediante el nodo Rellenar” en la página 150 para obtener más información. Tenga en cuenta que las funciones de conversión (y otras muchas funciones que requieren un tipo específico de entrada, como un valor de fecha u hora) dependen de los formatos actuales especificados en el cuadro de diálogo de propiedades de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003*, *Feb 2003*, (etc.) para almacenar datos, seleccione **MES AAAA** como formato de fecha predeterminado de la ruta. Las funciones de conversión también están disponibles en el nodo Derivar para la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos. Consulte el tema “Nueva codificación de valores con el nodo Derivar” en la página 149 para obtener más información.

Lectura de datos mezclados. Tenga en cuenta que al leer los campos con almacenamiento numérico (entero, real, hora, marca de tiempo o fecha), cualquier valor no numérico se establece como nulo o perdido por el sistema. Esto ocurre porque, a diferencia de otras aplicaciones, IBM SPSS Modeler no permite la combinación de tipos de almacenamiento dentro de un campo. Para evitarlo, cualquier campo con datos mezclados debe leerse como cadenas, ya sea cambiando el tipo de almacenamiento en el nodo de origen o en la aplicación externa, según sea necesario.

Note: los nodos Datos Usuario generados pueden contener ya información de almacenamiento, recogida del nodo de origen si está instanciado. Un nodo que no está instanciado no contiene información acerca del tipo de uso ni del almacenamiento.

Reglas para la especificación de valores

Para los campos simbólicos, debe dejar espacios entre varios valores, como:

ALTA NORMAL BAJA

Para los campos numéricos, puede introducir varios valores de la misma forma (con espacios entre ellos):

10 12 14 16 18 20

O puede especificar la misma serie de números estableciendo los límites (10, 20) y los pasos entre ellos (2). Con este método, se escribirá lo siguiente:

10,20,2

Estos dos métodos se pueden combinar insertando uno dentro de otro, como se indica a continuación:

1 5 7 10,20,2 21 23

Esta entrada generará los siguientes valores:

1 5 7 10 12 14 16 18 20 21 23

Nodo Generar simulación

El nodo Generar simulación permite, de forma sencilla, generar datos simulados, ya sea sin datos históricos utilizando distribuciones estadísticas especificadas por el usuario o de forma automática utilizando las distribuciones obtenidas de la ejecución de un nodo Ajustar simulación sobre datos históricos existentes. La generación de datos simulados resulta de utilidad cuando se desea evaluar el resultado de un modelo predictivo en presencia de incertidumbre en las entradas de modelo.

Creación de datos sin datos históricos

El nodo Generar simulación está disponible en la paleta Orígenes y se puede añadir directamente al lienzo de rutas.

1. Pulse en la pestaña **Orígenes** de la paleta de nodos.
2. Arrastre y suelte o pulse dos veces para añadir el nodo Generar simulación al lienzo de rutas.
3. Pulse dos veces para abrir el cuadro de diálogo y especifique campos, tipos de almacenamiento, distribuciones estadísticas y parámetros de distribución.

Nota: los nodos Generación de simulación que se seleccionan desde la paleta Orígenes estarán completamente en blanco, sin campos ni información de distribución. Esto permite crear datos simulados completos sin utilizar datos históricos.

Generación de datos simulados utilizando datos históricos existentes

Un nodo Generar simulación también puede crearse ejecutando un nodo terminal Ajustar simulación:

1. Pulse con el botón derecho en el nodo Ajustar simulación y seleccione **Ejecutar** en el menú.
2. Aparecerá el nodo Generar simulación en el lienzo de ruta con un enlace de actualización al nodo Ajustar simulación.

- Una vez generado, el nodo Generar simulación hereda todos los campos, tipos de almacenamiento e información de distribución estadística del nodo Ajustar simulación.

Definición de un enlace de actualización a un nodo de ajuste de simulación

Puede crearse un enlace entre un nodo Generar simulación y un nodo Ajustar simulación. Esto es útil si se desea actualizar uno o más campos con la información de la distribución que mejor se ajuste, determinada mediante el ajuste a datos históricos.

- Pulse con el botón derecho en el nodo Generar simulación.
- En el menú, seleccione **Definir enlace de actualización**. El cursor cambiará al cursor de enlaces.
- Pulse en otro nodo. Si este nodo es un nodo Ajustar simulación, se creará un enlace. Si este nodo no es un nodo Ajustar simulación, no se creará ningún enlace y el cursor volverá a ser un cursor normal.

Si los campos del nodo Ajustar simulación son distintos de los del nodo Generar simulación, se mostrará un mensaje en el que se informa de que hay una discrepancia.

Cuando el nodo Ajustar simulación se utiliza para actualizar el nodo Generar simulación enlazado, el resultado dependerá de si los mismos campos están presentes en ambos nodos y de si los campos están desbloqueados en el nodo Generación de simulación. Los resultados de actualizar un nodo Ajustar simulación se muestran en la tabla siguiente.

Tabla 10. Resultados de la actualización de un nodo Ajustar simulación

Campo en nodo Generar simulación	Campo en nodo Ajustar simulación	
	Presente	Perdido
Presente y desbloqueado.	El campo se sobrescribe.	El campo se suprime.
Ausente.	El campo se añade.	No cambia.
Presente y bloqueado.	La distribución del campo no se sobrescribe. Se actualizan la información del cuadro de diálogo Detalles del ajuste y las correlaciones.	El campo no se sobrescribe. Las correlaciones se establecen a cero.
Se selecciona la casilla de verificación No borrar Min/Max al reajustar.	Se sobrescribe el campo, salvo los valores en la columna Min,Max.	
Se selecciona la casilla de verificación No recalculan correlaciones al reajustar.	Si el campo está desbloqueado, se sobrescribirá. Las correlaciones no se sobrescriben.	

Eliminación de un enlace de actualización a un nodo de ajuste de simulación

Puede eliminarse un enlace entre un nodo Generar simulación y un nodo Ajustar simulación siguiendo los pasos siguientes:

- Pulse con el botón derecho en el nodo Generar simulación.
- En el menú, seleccione **Eliminar enlace de actualización**. Se eliminará el enlace.

Configuración de las opciones del nodo Generar simulación

Las opciones de la pestaña Datos del cuadro de diálogo del nodo Generar simulación pueden utilizarse para hacer lo siguiente:

- Ver, especificar y editar la información de distribución estadística de los campos.
- Ver, especificar y editar las correlaciones entre campos.
- Especificar el número de iteraciones y casos a simular.

Seleccionar un elemento. Permite conmutar entre las tres vistas del nodo Generar simulación: Campos simulados, Correlaciones y Opciones avanzadas.

Vista Campos simulados

Si el nodo Generar simulación se ha generado o actualizado desde un nodo Ajustar simulación utilizando datos históricos, en la vista Campos simulados podrá verse y editarse la información de distribución estadística de cada campo. En la pestaña **Tipos** del nodo Generar simulación se copia la siguiente información de cada campo procedente del nodo Ajustar simulación:

- Nivel de medición
- Valores
- Perdido
- Comprobación
- Función

Si no se dispone de datos históricos, podrán definirse campos y especificar sus distribuciones seleccionando un tipo de almacenamiento y seleccionando un tipo de distribución y especificando los parámetros necesarios. La generación de datos de esta forma significa que la información relativa al nivel de medición de cada campo no estará disponible mientras no se instancien los datos, por ejemplo, en la pestaña **Tipos** o en un nodo Tipo.

La vista Campos simulados contiene varias herramientas que pueden utilizarse para realizar las siguientes tareas:

- Añadir y eliminar campos.
- Cambiar el orden de los campos en la visualización.
- Especificar un tipo de almacenamiento para cada campo.
- Especificar una distribución estadística para cada campo.
- Especificar valores de parámetro para la distribución estadística de cada campo.

Campos simulados. Esta tabla contendrá una fila vacía si el nodo Generar simulación se ha añadido al lienzo de rutas desde la paleta Orígenes. Cuando se edita esta fila, se añade una fila vacía al final de la tabla. Si el nodo Generar simulación se ha creado desde un nodo Ajustar simulación, esta tabla contendrá una fila por cada campo de los datos históricos. Pueden añadirse filas adicionales a la tabla pulsando en el icono **Añadir nuevo campo**.

La tabla Campos simulados consta de las columnas siguientes:

- **Campo.** Contiene los nombres de los campos. Los nombres de campo pueden editarse escribiendo en las celdas.
- **Almacenamiento.** Las celdas de esta columna contienen una lista desplegable de tipos de almacenamiento. Los tipos de almacenamiento disponibles son **Cadena, Entero, Real, Hora, Fecha, e Indicación de fecha y hora**. La elección del tipo de almacenamiento determinará qué distribuciones están disponibles en la columna Distribución. Si el nodo Generar simulación se ha creado a partir de un nodo Ajustar simulación, el tipo de almacenamiento se copiará del nodo Ajustar simulación.

Nota: En el caso de los campos con tipos de almacenamiento de fecha y hora, deberán especificarse los parámetros de distribución como enteros. Por ejemplo, para especificar una fecha de media de 1 de enero de 1970, se utiliza el entero 0. El entero con signo representa el número de segundos transcurridos desde (o antes) de la medianoche del 1 de enero de 1970.

- **Estado.** Los iconos en la columna de estado indican el estado del ajuste de cada campo.



No se ha especificado ninguna distribución para el campo o faltan uno o más parámetros de distribución. Para ejecutar la simulación, deberá especificarse una distribución para este campo y especificarse valores válidos para los parámetros.



El campo se establece a la distribución que mejor se ajuste.

Nota: Este icono solo podrá mostrarse si el nodo Generar simulación se crea a partir de un nodo Ajustar simulación.



La distribución que mejor se ajusta se ha sustituido con una distribución alternativa procedente del subcuadro de diálogo Detalles del ajuste. Consulte el tema “Ajustar detalles” en la página 56 para obtener más información.



La distribución se ha especificado o editado manualmente, y es posible que incluya un parámetro especificado a más de un nivel.

- **Bloqueado.** El bloqueo de un campo simulado mediante la selección de la casilla de verificación que tiene el icono de bloqueo excluye el campo de la actualización automática que lleva a cabo el nodo Ajustar simulación. Esto resulta muy útil cuando se especifica una distribución manualmente y se desea asegurar que no se vea afectada por el ajuste de distribución automático que tiene lugar cuando se ejecuta un nodo Ajustar simulación enlazado.
- **Distribución.** Las celdas de esta columna contienen una lista desplegable de distribuciones estadísticas. La elección del tipo de almacenamiento determinará qué distribuciones estarán disponibles en esta columna para un campo dado. Consulte el tema “Distribuciones” en la página 59 para obtener más información.

Nota: La distribución fija no puede especificarse para todos los campos. Si se desea que cada campo de los datos generados sea fijo, podrá utilizarse un nodo Entrada de usuario seguido de un nodo Equilibrar.

- **Parámetros.** Los parámetros de distribución asociados con cada distribución ajustada se muestran en esta columna. Varios valores de un parámetro se separan mediante comas. La especificación de varios valores para un parámetro genera múltiples iteraciones para la simulación. Consulte el tema “Iteraciones” en la página 59 para obtener más información. Si faltan parámetros, quedará reflejado en el icono que aparece en la columna Estado. Para especificar valores para los parámetros, se pulsa en esta columna en la fila correspondiente al campo en cuestión y se selecciona **Especificar** en la lista. Esto abrirá el subcuadro de diálogo Especificar parámetros. Consulte el tema “Especificar parámetros” en la página 57 para obtener más información. Esta columna estará inhabilitada si se ha seleccionado Empírico en la columna Distribución.
- **Mín, Max.** En esta columna, en algunas distribuciones, puede especificarse un valor mínimo, un valor máximo o ambos para los datos simulados. Los datos simulados que sean menores que el valor mínimo y mayores que el valor máximo serán rechazados aun siendo válidos para la distribución especificada. Para especificar valores mínimo y máximo, se pulsa en esta columna en la fila correspondiente al campo en cuestión y se selecciona **Especificar** en la lista. Esto abrirá el subcuadro de diálogo Especificar parámetros. Consulte el tema “Especificar parámetros” en la página 57 para obtener más información. Esta columna estará inhabilitada si se ha seleccionado Empírico en la columna Distribución.

Utilizar mejor ajuste. Solo está habilitado si el nodo Generar simulación se ha creado automáticamente a partir de un nodo Ajustar simulación utilizando datos históricos y está seleccionada una única fila en la

tabla Campos simulados. Sustituye la información del campo en la fila seleccionada con la información de la distribución que mejor ajusta del campo. Si se ha editado la información de la fila seleccionada, cuando se pulse este botón se restablecerá la información de forma que vuelva a la distribución de mejor ajuste determinada a partir del nodo Ajustar simulación.

Detalles de ajuste. Solo está habilitado si el nodo Generar simulación se ha creado de forma automática a partir de un nodo Ajustar simulación. Abre el subcuadro de diálogo Detalles de ajuste. Consulte el tema “Ajustar detalles” en la página 56 para obtener más información.

Pueden llevarse a cabo diversas tareas de utilidad utilizando los iconos situados a la derecha de la vista Campos simulados. Estos iconos se describen en la tabla siguiente.

Tabla 11. Iconos en la vista Campos simulados.





Icono	Ayuda flotante	Descripción
	Editar parámetros de distribución	Solo está habilitado cuando se selecciona una única fila en la tabla Campos simulados. Abre el subcuadro de diálogo Especificar parámetros de la fila seleccionada. Consulte el tema “Especificar parámetros” en la página 57 para obtener más información.
	Añadir nuevo campo	Solo está habilitado cuando se selecciona una única fila en la tabla Campos simulados. Añade una fila vacía nueva al final de la tabla Campos simulados.
	Crear varias copias	Solo está habilitado cuando se selecciona una única fila en la tabla Campos simulados. Abre el subcuadro de diálogo Clonar campo. Consulte el tema “Clonar campo” en la página 56 para obtener más información.
	Suprimir campo seleccionado	Suprime la fila seleccionada en la tabla Campos simulados.
	Mover al principio	Sólo está habilitado si la fila seleccionada no está ya en la fila superior de la tabla Campos simulados. Mueve la fila seleccionada al principio de la tabla Campos simulados. Esta acción afecta al orden de los campos de los datos simulados.
	Subir	Sólo está habilitado si la fila seleccionada no es la fila superior de la tabla Campos simulados. Sube la fila seleccionada una posición en la tabla Campos simulados. Esta acción afecta al orden de los campos de los datos simulados.

Tabla 11. Iconos en la vista Campos simulados (continuación).

Icono	Ayuda flotante	Descripción
	Bajar	Sólo está habilitado si la fila seleccionada no es la fila inferior de la tabla Campos simulados. Baja la fila seleccionada una posición en la tabla Campos simulados. Esta acción afecta al orden de los campos de los datos simulados.
	Mover al final	Sólo está habilitado si la fila seleccionada no está ya en la fila inferior de la tabla Campos simulados. Mueve la fila seleccionada al final de la tabla Campos simulados. Esta acción afecta al orden de los campos de los datos simulados.

No borrar Min ni Max al reajustar. Cuando está seleccionado, los valores mínimo y máximo no se sobrescriben cuando se actualizan las distribuciones ejecutando un nodo conectado Ajuste de simulación.

Vista Correlaciones

Con frecuencia, se sabe que los campos de entrada de los modelos de predicción están correlacionados -por ejemplo, la altura y el peso. Las correlaciones entre campos que van a simularse deben tenerse en cuenta para asegurar que los valores simulados conserven dichas correlaciones.

Si el nodo Generar simulación se ha generado o actualizado desde un nodo Ajustar simulación utilizando datos históricos, en la vista Correlaciones podrán verse y editarse las correlaciones calculadas entre pares de campos. Si no se dispone de datos históricos, podrán especificarse las correlaciones manualmente a partir del conocimiento que se tenga sobre cómo se correlacionan los campos.

Nota: Antes de generarse los datos, se comprueba automáticamente si la matriz de correlaciones es semidefinida positiva y puede, por tanto, invertirse. Una matriz puede invertirse si sus columnas son linealmente independientes. Si la matriz de correlaciones no puede invertirse, se ajustará automáticamente para que pueda invertirse.

Puede optarse por mostrar las correlaciones en formato de matriz o de lista.

Matriz de correlaciones. Muestra las correlaciones existentes entre pares de campos de una matriz. Se listan los nombres de campo en orden alfabético bajando por la izquierda y a lo largo de la parte superior de la matriz. Solo pueden editarse las celdas situadas por debajo de la diagonal; debe especificarse un valor entre -1,000 y 1,000, ambos inclusive. La celda situada por encima de la diagonal se actualiza cuando el foco se aparta de su celda especular situada por debajo de la diagonal; ambas celdas mostrarán entonces el mismo valor. Las celdas diagonales siempre están inhabilitadas y siempre tienen una correlación de 1,000. El valor predeterminado de las celdas restantes es 0,000. Un valor de 0,000 indica que no existe ninguna correlación entre el par de campos asociado. En la matriz solo se incluyen campos continuos y ordinales. No se muestran en la tabla los campos nominales, categóricos y de marca, así como los campos asignados a la distribución Fija.

Lista de correlaciones. Muestra las correlaciones existentes entre pares de campos de una tabla. Cada fila de la tabla muestra la correlación existente entre un par de campos. Las filas no pueden añadirse ni suprimirse. Las columnas con los encabezados Campo 1 y Campo 2 contienen los nombres de campo, que no pueden editarse. La columna Correlación contiene las correlaciones, que pueden editarse; debe especificarse un valor comprendido entre -1,000 y 1,000, ambos inclusive. El valor predeterminado de

todas las celdas es 0,000. En la lista solo se incluyen campos continuos y ordinales. No se muestran en la lista los campos nominales, categóricos y de marca, así como los campos asignados a la distribución Fija.

Restablecer correlaciones. Abre el cuadro de diálogo Restablecer correlaciones. Si se dispone de datos históricos, podrá optarse por una de las tres opciones siguientes:

- **Ajustadas.** Sustituye las correlaciones actuales con las calculadas utilizando los datos históricos.
- **Ceros.** Sustituye las correlaciones actuales con ceros.
- **Cancelar.** Cierra el recuadro de diálogo. Las correlaciones no cambian.

Si no se dispone de datos históricos, pero se han efectuado cambios a las correlaciones, podrá optarse por sustituir las correlaciones actuales con ceros o cancelar.

Mostrar como. Seleccione **Tabla** para visualizar las correlaciones como una matriz. Seleccione **Lista** para visualizar las correlaciones como una lista.

No recalcular correlaciones al ajustar. Seleccione esta opción si desea especificar manualmente las correlaciones y evitar que se sobrescriban cuando las distribuciones se ajusten automáticamente mediante un nodo Ajustar simulación y datos históricos.

Utilizar tabla de contingencias multivía para las entradas con una distribución categórica. De forma predeterminada, todos los campos con una distribución categórica se incluyen en una tabla de contingencias (o tabla de contingencias multivía, en función del número de campos con una distribución categórica). La tabla de contingencias se construye, al igual que las correlaciones, cuando se ejecuta un nodo Ajustar simulación. La tabla de contingencias no puede verse. Cuando esta opción está seleccionada, los campos con una distribución categórica se simulan utilizando los porcentajes reales de la tabla de contingencias. Es decir, las asociaciones que pueda haber entre campos nominales se regeneran en los nuevos datos simulados. Cuando se deselecciona esta opción, los campos con distribuciones categóricas se simulan utilizando los porcentajes esperados de la tabla de contingencias. Si se modifica un campo, dicho campo se elimina de la tabla de contingencias.

Vista Opciones avanzadas

Número de casos a simular. Muestra las opciones de especificación del número de casos a simular y cómo se nombrarán las iteraciones.

- **Número máximo de casos.** Especifica el número máximo de casos de datos simulados, y los valores de destino asociados, que se generarán. El valor predeterminado es 100 000, el valor mínimo es 1000 y el valor máximo es 2 147 483 647.
- **Iteraciones.** Este número se calcula automáticamente y no se puede editar. Una iteración se crea automáticamente cada vez que un parámetro de distribución tiene varios valores especificados.
- **Filas totales.** Sólo está habilitado cuando el número de iteraciones es mayor que 1. El número se calcula automáticamente utilizando la ecuación mostrada y no se puede editar.
- **Crear campo de iteración.** Sólo está habilitado cuando el número de iteraciones es mayor que 1. Cuando se selecciona, el campo **Nombre** está habilitado. Consulte el tema "Iteraciones" en la página 59 para obtener más información.
- **Nombre.** Solo está habilitado cuando la casilla de verificación **Crear campo de iteración** está seleccionada y el número de iteraciones es mayor que 1. El nombre del campo de iteración se edita escribiendo en este campo de texto. Consulte el tema "Iteraciones" en la página 59 para obtener más información.

Semilla aleatoria. Al establecer una semilla aleatoria podrá replicar las simulaciones.

- **Replicar resultados.** Cuando está seleccionado, el botón **Generar** y el campo **Semilla aleatoria** están habilitados.

- **Semilla aleatoria.** Sólo está habilitado cuando la casilla de verificación **Replicar resultados** está seleccionada. En este campo se puede especificar el entero que se va a utilizar como semilla aleatoria. El valor predeterminado es 629111597.
- **Generar.** Sólo está habilitado cuando la casilla de verificación **Replicar resultados** está seleccionada. Crea un entero pseudoaleatorio entre 1 y 999999999, ambos inclusive, en el campo **Semilla aleatoria**.

Clonar campo

En el cuadro de diálogo Clonar Campo puede especificarse cuántas copias del campo seleccionado crear y cómo se denomina cada copia. Es útil tener varias copias de campos cuando se investiga efectos compuestos como, por ejemplo, tipos de interés o tasas de crecimiento a lo largo de una serie de períodos de tiempo sucesivos.

La barra de título del cuadro de diálogo contiene el nombre del campo seleccionado.

Número de copias a efectuar. Contiene el número de copias del campo a crear. Pulse en las flechas para seleccionar el número de copias a crear. El número mínimo de copias es 1 y el máximo es 512. El número de copias se establece inicialmente a 10.

Caracter(es) de sufijo de la copia Contiene los caracteres que se añaden al final del nombre de campo de cada copia. Estos caracteres separan el nombre de campo del número de copia. Los caracteres de sufijo pueden editarse escribiendo en este campo. Este campo puede dejarse vacío; en tal caso no habrá caracteres entre el nombre de campo y el número de copia. El carácter predeterminado es el guión bajo.

Número de copia inicial. Contiene el número de sufijo de la primera copia. Pulse en las flechas para seleccionar el número de copia inicial. El número de copia inicial mínimo es 1 y el máximo 1000. El número de copia inicial predeterminado es 1.

Salto de número de copia. Contiene el incremento de los números de sufijo. Pulse en las flechas para seleccionar el incremento. El incremento mínimo es 1 y el máximo 255. El incremento se establece inicialmente a 1.

Campos. Contiene una vista previa de los nombres de campo de las copias, que se actualiza al editarse alguno de los campos del cuadro de diálogo Clonar Campo. Este texto se genera automáticamente y no se puede editar.

Aceptar. Genera todas las copias tal y como se ha especificado en el cuadro de diálogo. Las copias se añaden a la tabla Campos simulados en el cuadro de diálogo del nodo Generar simulación, inmediatamente debajo de la fila que contiene el campo copiado.

Cancelar. Cierra el recuadro de diálogo. Se descartan los cambios que puedan haberse efectuado.

Ajustar detalles

El cuadro de diálogo Detalles de ajuste solo está disponible si el nodo Generar simulación se ha creado o actualizado mediante la ejecución de un nodo Ajustar simulación. Muestra los resultados del ajuste automático de distribución del campo seleccionado. Las distribuciones se ordenan por la bondad del ajuste, apareciendo en primer lugar la distribución con mejor ajuste. En este cuadro de diálogo pueden realizarse las siguientes tareas:

- Examinar las distribuciones ajustadas a los datos históricos.
- Seleccionar una de las distribuciones ajustadas.

Campo. Contiene el nombre del campo seleccionado. Este texto no se puede editar.

Tratar como (medición). Muestra el tipo de medición del campo seleccionado. Se toma de la tabla Campos simulados del cuadro de diálogo del nodo Generar simulación. El tipo de medición puede cambiarse pulsando la flecha y seleccionando un tipo de medición de la lista desplegable. Hay tres opciones: **Continuo**, **Nominal** y **Ordinal**.

Distribuciones. La tabla Distribuciones muestra todas las distribuciones adecuadas para el tipo de medición. Las distribuciones que se han ajustado a los datos históricos están ordenadas por la bondad del ajuste, de mejor a peor ajuste. La bondad del ajuste viene determinada por la estadística de ajuste seleccionada en el nodo Ajustar simulación. Las distribuciones que no se han ajustado a los datos históricos aparecen listadas en la tabla en orden alfabético debajo de las distribuciones ajustadas.

La tabla de distribuciones muestra las siguientes columnas:

- **Utilizar.** El botón de selección seleccionado indica qué distribución está actualmente seleccionada para el campo. Puede sustituirse la distribución de mejor ajuste seleccionando el botón de selección de la distribución deseada en la columna Utilizar. La selección de un botón de selección en la columna Utilizar también muestra una representación gráfica de la distribución superpuesta sobre un histograma (o gráfica de barras) de los datos históricos del campo seleccionado. Solo puede seleccionarse una distribución a la vez.
- **Distribución.** Contiene el nombre de la distribución. Esta columna no se puede editar.
- **Estadísticas de ajuste.** Contiene las estadísticas del ajuste calculado de la distribución. Esta columna no se puede editar. El contenido de la celda depende del tipo de medición del campo:
 - **Continuo.** Contiene los resultados de las pruebas de Anderson-Darling y Kolmogorov-Smirnoff. También se muestran los valores p asociados a las pruebas. La estadística de ajuste seleccionada como criterio de Bondad del ajuste en el nodo Ajustar simulación se muestra en primer lugar y se utiliza para ordenar las distribuciones. Las estadísticas de Anderson-Darling se muestran como $A=vala P=valp$. Las estadísticas de Kolmogorov-Smirnoff se muestran como $K=valk P=valp$. Cuando no se pueda calcular una estadística, se mostrará un punto en lugar de un número.
 - **Nominal y Ordinal.** Contiene los resultados de la prueba chi-cuadrado. El valor p asociado a la prueba también se muestra. Las estadísticas se muestran como $Chi-Sq=val P=valp$. Si la distribución no se ha ajustado, se mostrará No ajustada. Si la distribución no puede ajustarse matemáticamente, se mostrará No puede ajustarse.

Nota: la celda siempre está vacía en la distribución Empírica.

- **Parámetros.** Contiene los parámetros de distribución asociados con cada distribución ajustada. Los parámetros se muestran como *nombre_parámetro = valor_parámetro*, con los parámetros separados por un único espacio. En el caso de la distribución categórica, los nombres de parámetro son las categorías y los valores de parámetro son las probabilidades asociadas. Si la distribución no se ha ajustado a los datos históricos, la celda estará vacía. Esta columna no se puede editar.

Miniatura de histograma. Muestra una representación gráfica de la distribución seleccionada superpuesta al histograma de los datos históricos del campo seleccionado.

Miniatura de distribución. Muestra una explicación y una ilustración de la distribución seleccionada.

Aceptar. Cierra el cuadro de diálogo y actualiza los valores de las columnas Medición, Distribución, Parámetros y Min, Max de la tabla Campos simulados del campo seleccionado con la información de la distribución seleccionada. El icono de la columna Estado también se actualiza para indicar si la distribución seleccionada es la distribución que mejor se ajusta a los datos.

Cancelar. Cierra el recuadro de diálogo. Se descartan los cambios que puedan haberse efectuado.

Especificar parámetros

En el cuadro de diálogo Especificar parámetros se pueden especificar manualmente los valores de los parámetros para la distribución del campo seleccionado. También se puede elegir una distribución diferente para el campo seleccionado.

El cuadro de diálogo Especificar parámetros puede abrirse de tres formas:

- Efectúe una doble pulsación en un nombre de campo de la tabla Campos simulados en el cuadro de diálogo del nodo Generar simulación.
- Pulse en la columna Parámetros o Min, Max de la tabla Campos simulados y seleccione **Especificar** en la lista.
- En la tabla Campos simulados, seleccione una fila y pulse en el icono **Editar parámetros de distribución**.

Campo. Contiene el nombre del campo seleccionado. Este texto no se puede editar.

Distribución. Contiene la distribución del campo seleccionado. Se toma de la tabla Campos simulados. La distribución puede cambiarse pulsando la flecha y seleccionando una distribución de la lista desplegable. Las distribuciones disponibles dependen del tipo de almacenamiento del campo seleccionado.

Caras. Esta opción solo está disponible cuando la distribución de dado está seleccionada en el campo **Distribución**. Pulse en las flechas para seleccionar el número de caras, o categorías, en las que hay que dividir el campo. El número mínimo de caras es de dos, y el máximo es de 20. El número de caras se establece inicialmente a 6.

Parámetros de distribución. La tabla Parámetros de distribución contiene una fila por cada parámetro de la distribución elegida.

Nota: La distribución utiliza un parámetro rate con un parámetro de forma $\alpha = k$ y un parámetro de escala inversa $\beta = 1/\theta$.

La tabla contiene dos columnas:

- **Parámetro.** Contiene los nombres de los parámetros. Esta columna no se puede editar.
- **Valor(es).** Contiene los valores de los parámetros. Si el nodo Generar simulación se ha creado o actualizado desde un nodo Ajustar simulación, las celdas de esta columna contendrán los valores de los parámetros determinados mediante el ajuste de la distribución a los datos históricos. Si el nodo Generar simulación se ha añadido al lienzo de rutas desde la paleta de nodos Origen, las celdas de esta columna estarán vacías. Los valores se pueden editar escribiendo en las celdas. Consulte el tema “Distribuciones” en la página 59 para obtener más información sobre los parámetros necesarios para cada distribución y los valores de parámetro aceptables.

Varios valores de un parámetro deben estar separados por coma. La especificación de varios valores para un parámetro definirá múltiples iteraciones de la simulación. Solo se pueden especificar varios valores para un único parámetro.

Nota: En el caso de los campos con tipos de almacenamiento de fecha y hora, deberán especificarse los parámetros de distribución como enteros. Por ejemplo, para especificar una fecha media de 1 de enero de 1970, se utiliza el entero 0.

Nota: Cuando está seleccionada la distribución de dado, la tabla de parámetros de distribución es ligeramente diferente. La tabla contiene una fila por cada lado (o categoría). La tabla contiene una columna Valor y una columna Probabilidad. La columna Valor contiene una etiqueta por cada categoría. Los valores predeterminados de las etiquetas son los enteros 1-N, donde N es el número de caras. Las etiquetas se pueden editar escribiendo en las celdas. Puede escribirse cualquier valor en dichas celdas. Si desea utilizar un valor que no sea un número, el tipo de almacenamiento del campo de datos deberá cambiarse a cadena si el tipo de almacenamiento no fuera ya cadena. La columna probabilidad contiene la probabilidad de cada categoría. Las probabilidades no pueden editarse y se calculan como $1/N$.

Vista previa. Muestra una gráfica de ejemplo de la distribución, basada en los parámetros especificados. Si se han especificado dos o más valores para un parámetro, se mostrarán gráficas de ejemplo para cada

valor del parámetro. Si hay datos históricos disponibles para el campo seleccionado, la gráfica de la distribución se superpondrá al histograma de los datos históricos.

Valores opcionales. Estas opciones se utilizan para especificar un valor mínimo, un valor máximo o ambos para los datos simulados. Los datos simulados que sean menores que el valor mínimo y mayores que el valor máximo serán rechazados aun siendo válidos para la distribución especificada.

- **Especificar mínimo.** Seleccione esta opción para habilitar el campo **Rechazar valores por debajo de**. La casilla de verificación se inhabilita si se selecciona la distribución Empírica.
- **Rechazar valores por debajo de.** Solo se habilita si se selecciona **Especificar mínimo**. Especifique el valor mínimo de los datos simulados. Todo valor simulado menor que este valor será rechazado.
- **Especificar máximo.** Seleccione esta opción para habilitar el campo **Rechazar valores por encima de**. La casilla de verificación se inhabilita si se selecciona la distribución Empírica.
- **Rechazar valores por encima de.** Solo se habilita si se selecciona **Especificar máximo**. Especifique el valor máximo de los datos simulados. Todo valor simulado mayor que este valor será rechazado.

Aceptar. Cierra el cuadro de diálogo y actualiza los valores de las columnas Distribución, Parámetros y Min, Max de la tabla Campos simulados del campo seleccionado. El icono de la columna Estado también se actualiza para indicar la distribución seleccionada.

Cancelar. Cierra el recuadro de diálogo. Se descartan los cambios que puedan haberse efectuado.

Iteraciones

Si se ha especificado más de un valor para un campo fijo o un parámetro de distribución, se generará un conjunto independiente de casos simulados –a todos los efectos, una simulación independiente- para cada valor especificado. Esto permite investigar el efecto de variar el campo o parámetro. Cada conjunto de casos simulados se denomina *iteración*. En los datos simulados, las iteraciones se apilan.

Si el recuadro de selección **Crear campo de iteración** en la vista Opciones avanzadas del diálogo del nodo Generar simulación está seleccionado, se añadirá un campo de iteración a los datos simulados como un campo nominal con almacenamiento numérico. El nombre de este campo puede editarse escribiendo en el campo **Nombre** de la vista Opciones avanzadas. Este campo contiene una etiqueta que indica a qué iteración pertenece cada caso simulado. El formato de las etiquetas depende del tipo de iteración:

- **Iteración de un campo fijo.** La etiqueta es el nombre del campo seguido de un signo igual, seguido del valor del campo de esa iteración, es decir
nombre_campo = valor_campo
- **Iteración de un parámetro de distribución.** La etiqueta es el nombre del campo seguido de dos puntos, seguido del nombre del parámetro iterado, seguido de un signo igual, seguido del valor del parámetro de esa iteración, es decir
nombre_campo:nombre_parámetro = valor_parámetro
- **Iteración de un parámetro de distribución de una distribución de rango o categórica.** La etiqueta es el nombre del campo, seguido de dos puntos, seguido de "Iteration", seguido del número de iteración, es decir
nombre_campo: Iteration número_iteración

Distribuciones

Puede especificarse de forma manual la distribución de probabilidad de cualquier campo abriendo el cuadro de diálogo Especificar parámetros de dicho campo, seleccionando la distribución deseada de la lista **Distribución** y especificando los parámetros de distribución en la tabla **Parámetros de distribución**. A continuación se incluyen algunas notas sobre distribuciones concretas:

- **Categóricas.** La distribución categórica describe un campo de entrada que tiene un número fijo de valores numéricos conocidos como categorías. Cada categoría tiene una probabilidad asociada, de forma que la suma de las probabilidades de todas las categorías es igual a uno.

Nota: Si se especifican probabilidades para las categorías que no suman 1, se recibirá un aviso.

- **Binomial negativa - Errores.** Describe la distribución del número de fallos en una secuencia de intentos antes de detectar un número especificado de éxitos. El parámetro *Umbral* es el número especificado de éxitos y el parámetro *Probabilidad* es la probabilidad de éxito en un intento determinado.
- **Binomial negativa - Intentos.** Describe la distribución del número de intentos necesarios para detectar un número especificado de éxitos. El parámetro *Umbral* es el número especificado de éxitos y el parámetro *Probabilidad* es la probabilidad de éxito en un intento determinado.
- **Rango.** Esta distribución consiste en un conjunto de intervalos con una probabilidad asignada a cada intervalo de forma que la suma de probabilidades en todos los intervalos es igual a 1. Los valores en un intervalo concreto se generan a partir de una distribución uniforme definida en ese intervalo. Los intervalos se especifican introduciendo un valor mínimo, un valor máximo y una probabilidad asociada.

Por ejemplo, se cree que el coste de una materia prima tiene una probabilidad del 40% de caer en el rango de 10 € - 15 € la unidad, y una probabilidad del 60% de caer en el rango de 15 € - 20 € la unidad. Se modelaría el coste con una distribución del rango de ambos intervalos [10 - 15] y [15 - 20], definiendo la probabilidad asociada con el primer intervalo a 0,4 y la probabilidad asociada con el segundo intervalo a 0,6. Los intervalos no tienen que ser contiguos y se pueden solapar. Por ejemplo, podría especificar los intervalos 10 - 15 y 20 - 25 o 10 - 15 y 13 - 16.

- **Weibull.** El parámetro *Ubicación* es un parámetro de ubicación opcional que especifica dónde se ubica el origen de la distribución.

La siguiente tabla muestra las distribuciones que están disponibles para ajustar la distribución personalizada y los posibles valores de los parámetros. Algunas de estas distribuciones están disponibles para un ajuste personalizado a determinados tipos de almacenamiento, incluso si el nodo Ajustar simulación no los ajusta automáticamente a dichos tipos de almacenamiento.

Tabla 12. Distribuciones disponibles en el ajuste personalizado

Distribución	Tipo de almacenamiento soportado para el ajuste personalizado	Parámetros	Límites del parámetro	Notas
Bernoulli	Entero, real, fecha y hora	Probabilidad	$0 \leq \text{Probabilidad} \leq 1$	
Beta	Entero, real, fecha y hora	Forma 1 Forma 2 Mínimo Máximo	≥ 0 ≥ 0 $< \text{Máximo}$ $> \text{Mínimo}$	El mínimo y el máximo son opcionales.
Binomial	Entero, real, fecha y hora	Número de intentos (n) Probabilidad Mínimo Máximo	> 0 , entero $0 \leq \text{Probabilidad} \leq 1$ $< \text{Máximo}$ $> \text{Mínimo}$	El número de intentos debe ser un entero. El mínimo y el máximo son opcionales.
Categoría	Entero, real, fecha y hora, cadena	Nombre de categoría (o etiqueta)	$0 \leq \text{Valor} \leq 1$	El valor es la probabilidad de la categoría. Los valores deben sumar 1, de lo contrario se generará un aviso.
Dice	Entero, cadena	Caras	$2 \leq \text{Caras} \leq 20$	La probabilidad de cada categoría (cara) se calcula como $1/N$, donde N es el número de caras. Las probabilidades no pueden editarse.

Tabla 12. Distribuciones disponibles en el ajuste personalizado (continuación)

Distribución	Tipo de almacenamiento soportado para el ajuste personalizado	Parámetros	Límites del parámetro	Notas
Empírica	Entero, real, fecha y hora			La distribución empírica no puede editarse ni seleccionarse como tipo. La distribución empírica solo estará disponible donde hayan datos históricos.
Exponencial	Entero, real, fecha y hora	Escala Mínimo Máximo	> 0 $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales.
Fija	Entero, real, fecha y hora, cadena	Valor		La distribución fija no puede especificarse para todos los campos. Si se desea que cada campo de los datos generados sea fijo, podrá utilizarse un nodo Entrada de usuario seguido de un nodo Equilibrar.
Gamma	Entero, real, fecha y hora	Formas Escala Mínimo Máximo	≥ 0 ≥ 0 $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales. La distribución utiliza un parámetro rate con un parámetro de forma $\alpha = k$ y un parámetro de escala inversa $\beta = 1/\theta$.
Logarítmica normal	Entero, real, fecha y hora	Forma 1 Forma 2 Mínimo Máximo	≥ 0 ≥ 0 $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales.
Binomial negativa: fallos	Entero, real, fecha y hora	Umbral Probabilidad Mínimo Máximo	≥ 0 $0 \leq Probabilidad \leq 1$ $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales.
Binomial negativa: intentos	Entero, real, fecha y hora	Umbral Probabilidad Mínimo Máximo	≥ 0 $0 \leq Probabilidad \leq 1$ $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales.
Normales	Entero, real, fecha y hora	Media Desviación típica Mínimo Máximo	≥ 0 > 0 $< Máximo$ $> Mínimo$	El mínimo y el máximo son opcionales.

Tabla 12. Distribuciones disponibles en el ajuste personalizado (continuación)

Distribución	Tipo de almacenamiento soportado para el ajuste personalizado	Parámetros	Límites del parámetro	Notas
Poisson	Entero, real, fecha y hora	Media Mínimo Máximo	≥ 0 $< \text{Máximo}$ $> \text{Mínimo}$	El mínimo y el máximo son opcionales.
Rango	Entero, real, fecha y hora	Comienzo(X) Fin(X) Probabilidad(X)	$0 \leq \text{Valor} \leq 1$	X es el índice de cada intervalo. Los valores de probabilidad deben sumar 1.
Triangular	Entero, real, fecha y hora	Modo Mínimo Máximo	$\text{Mínimo} \leq \text{Valor} \leq \text{Máximo}$ $< \text{Máximo}$ $> \text{Mínimo}$	
Uniforme	Entero, real, fecha y hora	Mínimo Máximo	$< \text{Máximo}$ $> \text{Mínimo}$	
Weibull	Entero, real, fecha y hora	Tasa Escala Ubicación Mínimo Máximo	> 0 > 0 ≥ 0 $< \text{Máximo}$ $> \text{Mínimo}$	Ubicación, máximo y mínimo son opcionales.

nodo Vista de datos

Utilice el nodo Ver datos para incluir en la ruta datos definidos en una vista de datos analíticos de IBM SPSS Collaboration and Deployment Services. una vista de datos analíticos define una estructura de acceso a datos que describe las entidades utilizadas en modelos predictivos y reglas de negocio. La vista asocia la estructura de datos a orígenes de datos físicos para el análisis.

El análisis predictivo requiere datos organizados en tablas donde cada fila se corresponde con una entidad para la que se realizan predicciones. Cada columna de una tabla representa un atributo cuantificable de la entidad. Algunos atributos pueden derivarse mediante una agregación de los valores de otro atributo. Por ejemplo, las filas de una tabla podrían representar clientes con columnas correspondientes al nombre de cliente, sexo, código postal, y el número de veces que el cliente ha tenido una compra superior a 500 € en el año pasado. La última columna se deriva del historial de pedidos de cliente, que normalmente se almacena en una o más tablas relacionadas.

El proceso de análisis predictivo implica la utilización de distintos conjuntos de datos a lo largo del ciclo de vida de un modelo. Durante el desarrollo inicial de un modelo predictivo, se utilizan datos históricos que a menudo tienen resultados conocidos para el evento que se predice. Para evaluar la eficacia y precisión del modelo, se valida un modelo candidato contra datos distintos. Tras validar el modelo, se despliega en producción para generar puntuaciones para varias entidades en un proceso por lotes o para entidades únicas de un proceso en tiempo real. Si se combina el modelo con reglas de negocio en un proceso de gestión de decisiones, se utilizan datos simulados para validar los resultados de la combinación. Sin embargo, aunque los datos utilizados difieren entre las fases del proceso de desarrollo del modelo, cada conjunto de datos deberá proporcionar el mismo conjunto de atributos para el modelo. El conjunto de atributos permanece constante; los registros de datos analizados cambian.

Una vista de datos analíticos consta de los componentes siguientes que abordan necesidades específicas de la analítica predictiva:

- Un esquema de vista de datos que define la interfaz de acceso a los datos como un conjunto de atributos organizados en tablas relacionadas. Los atributos del modelo pueden derivarse de otros atributos.
- Uno o varios planes de acceso a datos que proporcionen los atributos del modelo de datos con valores físicos. Los datos disponibles al modelo de datos se controlan especificando qué plan de acceso a datos está activo para una determinada aplicación.

Importante: Para utilizar el nodo Ver datos, antes deberá instalarse y configurarse un IBM SPSS Collaboration and Deployment Services Repository en el sitio. La vista de datos analíticos referenciada en el nodo suele crearse y almacenarse en el repositorio mediante IBM SPSS Collaboration and Deployment Services Deployment Manager.

Opciones de configuración del nodo Vista de datos

Las opciones de la pestaña **Datos** del cuadro de diálogo del nodo Vista de datos se utilizan para especificar la configuración de los datos de una vista de datos analíticos seleccionada en IBM SPSS Collaboration and Deployment Services Repository.

Vista de datos analíticos. Pulse en el botón de puntos suspensivos (...) para seleccionar la vista de datos analíticos. Si en ese momento no se está conectado a un servidor de repositorios, especifique el URL del servidor en el cuadro de diálogo Servidor de repositorios, pulse **Aceptar** y especifique sus credenciales de conexión en el cuadro de diálogo Credenciales de repositorio. Para obtener más información sobre el inicio de sesión en el repositorio y la recuperación de objetos, consulte la Guía del usuario de IBM SPSS Modeler.

Nombre de tabla. Seleccione una tabla del modelo de datos en la vista de datos analíticos. Cada tabla del modelo de datos representa un concepto o entidad que participa en el proceso de análisis predictivo. Los campos de las tablas se corresponden con atributos de las entidades representadas por las tablas. Por ejemplo, si están analizando los pedidos de los clientes, el modelo de datos podría incluir una tabla para los clientes y una tabla para los pedidos. La tabla de clientes puede tener atributos para el identificador, la edad, el sexo, el estado civil y el país de residencia del cliente. La tabla de pedidos puede tener atributos para el identificador de pedido, el número de elementos del pedido, el coste total y el identificador del cliente que ha formalizado el pedido. El atributo del identificador de cliente puede utilizarse para asociar los clientes de la tabla de clientes con sus pedidos en la tabla de pedidos.

Plan de acceso a datos. Seleccione un plan de acceso a datos en la vista de datos analíticos. Un plan de acceso a datos asocia las tablas de modelo de datos de una vista de datos analíticos a orígenes de datos físicos. Una vista de datos analíticos suele contener varios planes de acceso a datos. Cuando se cambia el plan de acceso a datos utilizado, se cambian los datos utilizados por la ruta. Por ejemplo, si la vista de datos analíticos contiene un plan de acceso a datos para entrenar un modelo y un plan de acceso a datos para probar un modelo, puede cambiarse del entrenamiento de datos a la prueba de datos cambiando el plan de acceso a datos utilizado.

Atributos opcionales. Si la aplicación que utiliza la vista de datos analíticos no necesita un determinado atributo, podrá marcarse dicho atributo como opcional. A diferencia de los atributos obligatorios, los atributos opcionales pueden incluir valores nulos. Puede que tenga que ajustar la aplicación para que incluya el tratamiento de valores nulos de los atributos opcionales. Por ejemplo, cuando se invoca una regla de negocio creada en IBM Operational Decision Manager, IBM Analytical Decision Management consulta el servicio de reglas para determinar qué entradas se necesitan. Si el registro a puntuar contiene un valor nulo en cualquiera de los campos necesarios del servicio de reglas, la regla no se invocará y los campos de salida de la regla se rellenarán con valores predeterminados. Si un campo opcional contiene un valor nulo, se invocará la regla. La regla puede comprobar la existencia de valores nulos para controlar el procesamiento.

Para especificar un atributo como opcional, pulse **Atributos opcionales** y seleccione los atributos que sean opcionales.

Incluir datos XML en el campo. Esta opción se selecciona para crear un campo que contenga los datos XML del modelo de objeto ejecutable por cada fila de datos. Esta información es obligatoria si los datos se van a utilizar con IBM Operational Decision Manager. Especifique el nombre de este nuevo campo.

Nodo de origen geoespacial

El nodo origen geoespacial se utiliza para aportar los datos de mapa o geoespaciales a su sesión de minería de datos. Puede importar datos de una de las dos formas siguientes:

- En un archivo de formas (.shp)
- Conectándose a un servidor ESRI que contenga un sistema de archivos jerárquicos que incluya archivos de mapas.

Nota: Es posible que necesite permisos de acceso para poder conectarse al servidor del servicio de mapas.

Los modelos STP (Spatio-Temporal Prediction) pueden incluir elementos de mapas o espaciales en sus pronósticos. Para obtener más información acerca de estos modelos, consulte el tema "Spatio-Temporal Prediction modeling node" en la sección Time Series Models de la publicación Modeler Modeling Nodes (ModelerModelingNodes.pdf).

Opciones de configuración para el nodo Origen geoespacial

Tipo de origen de datos: Puede importar datos desde un **Archivo de formas** (.shp) o conectar con un **Servicio de mapas**.

Si está utilizando un **Archivo de formas**, escriba el nombre de archivo y la vía de acceso al mismo o explore para seleccionar el archivo.

Nota: Los datos de formas requieren un archivo .shp y un archivo .dbf. Los dos archivos deben tener el mismo nombre y deben estar en la misma carpeta. El archivo .dbf se importa automáticamente cuando selecciona el archivo .shp. Además, es posible que exista un archivo .prj que especifique el sistema de coordenadas para los datos de formas.

Si está utilizando un **Servicio de mapas**, escriba el URL del servicio y pulse **Conectar**. Después de conectarse al servicio, se muestran las capas de dicho servicio en la parte inferior del recuadro de diálogo en una estructura de árbol en el panel **Mapas disponibles**. Expanda el árbol y seleccione la capa que requiera.

Nota: Es posible que se le solicite un nombre de usuario y una contraseña cuando seleccione el servicio de mapas.

Definición automática de los datos geoespaciales

De forma predeterminada, SPSS Modeler define automáticamente, siempre que sea posible, cualquier campo de datos geoespaciales del nodo de origen con los metadatos correctos. Es posible que los metadatos incluyan el nivel de medición del campo geoespacial (tal como Punto o Polígono) y el sistema de coordenadas que utilizan los campos, incluidos detalles como el punto de origen (por ejemplo, la latitud 0, la longitud 0) y las unidades de medición. Para obtener más información acerca de los niveles de medición, consulte "Subniveles de medición geoespacial" en la página 129.

Los archivos .shp y .dbf que componen el archivo de formas contienen un campo de identificador común que se utiliza como una clave. Por ejemplo, el archivo .shp puede contener países, con el campo de nombre de país como identificador, y el archivo .dbf puede contener información acerca de dichos países y el nombre del país también se utiliza como identificador.

Nota: Si el sistema de coordenadas no es el mismo que el sistema de coordenadas de SPSS Modeler predeterminado, es posible que pueda volver a proyectar los datos para utilizar el sistema de coordenadas necesario. Si desea obtener más información, consulte "Nodo Reproyección" en la página 185.

Nodo Enterprise View

Nota: El nodo Enterprise View se ha sustituido en IBM SPSS Modeler 16.0 por el nodo Vista de datos. Todavía se admite el nodo Enterprise View para rutas guardadas en ediciones anteriores. No obstante, al actualizar o crear nuevas rutas, se recomienda utilizar el nodo Vista de datos.

El nodo Enterprise View permite crear y mantener una conexión entre una sesión de IBM SPSS Modeler y una Enterprise View en un IBM SPSS Collaboration and Deployment Services Repository compartido. Al hacerlo, es posible leer datos de Enterprise View en una ruta de IBM SPSS Modeler y empaquetar un modelo de IBM SPSS Modeler en un escenario al que pueden acceder otros usuarios del repositorio compartido.

Un **escenario** es un archivo que contiene una ruta de IBM SPSS Modeler con modelos y nodos específicos, así como propiedades adicionales que permiten su despliegue en IBM SPSS Collaboration and Deployment Services Repository con fines de puntuación o actualización automática del modelo. El uso de los nodos Enterprise View con escenario garantiza que, en una situación con varios usuarios, todos los usuarios trabajen con los mismos datos. Una **conexión** es un enlace de una sesión de IBM SPSS Modeler con Enterprise View en IBM SPSS Collaboration and Deployment Services Repository.

La **Enterprise View** es un conjunto completo de datos que pertenecen a una organización, independientemente de dónde se encuentren físicamente los datos. Cada conexión está formada por una selección específica de una única **vista de aplicación** (subconjunto de Enterprise View adaptado a una aplicación concreta), una **definición de proveedor de datos** (DPD, enlaza las tablas y columnas lógicas de la vista de aplicación con un origen de datos físico) y un **entorno** (identifica las columnas concretas que deben asociarse a los segmentos empresariales definidos). Las definiciones de Enterprise View, vistas de aplicación y DPD se almacenan y versionan en el repositorio, aunque los datos reales se encuentran en una o más bases de datos u otros orígenes externos.

Una vez que se ha establecido una conexión, se especifica la **tabla** de la vista de aplicación con la que se desea trabajar en IBM SPSS Modeler. En una vista de aplicación, una tabla es una vista lógica que consta de algunas o todas las columnas de una o más tablas físicas de una o más bases de datos físicas. Por tanto, el nodo Enterprise View permite ver los registros procedentes de varias tablas de bases de datos como una única tabla en IBM SPSS Modeler.

Requisitos

- Para utilizar el nodo Enterprise View, debe instalar y configurar IBM SPSS Collaboration and Deployment Services Repository en el sitio, con una Enterprise View, unas vistas de aplicación y DPD ya definidos.
- Además, es necesario instalar Controlador de IBM SPSS Collaboration and Deployment Services Enterprise View Driver en cada ordenador que se utilice para modificar o ejecutar la ruta. Para Windows, basta con instalar el controlador en el ordenador donde se ha instalado IBM SPSS Modeler o IBM SPSS Modeler Server y no será necesario realizar ninguna configuración adicional del controlador. En UNIX, es necesario añadir una referencia al script *pev.sh* al script de inicio. Póngase en contacto con el administrador local si desea obtener más información acerca de la instalación del controlador Controlador de IBM SPSS Collaboration and Deployment Services Enterprise View Driver.
- Una DPD se define respecto a un determinado origen de datos ODBC. Para usar una DPD de IBM SPSS Modeler, debe tener un origen de datos ODBC definido en el host del servidor de IBM SPSS Modeler que tenga el mismo nombre (y que conecte con el mismo almacén de datos) que a la que se hace referencia en la DPD.

Opciones de configuración del nodo Enterprise View

Puede utilizar las opciones de la pestaña Datos del cuadro de diálogo Enterprise View para:

- Seleccionar una conexión de repositorio existente.
- Editar una conexión de repositorio existente.
- Crear una nueva conexión de repositorio.
- Seleccionar una tabla de vista de aplicación.

Consulte la *Guía del administrador de IBM SPSS Collaboration and Deployment Services* para obtener más información sobre cómo trabajar con repositorios.

Conexión. La lista desplegable proporciona opciones para seleccionar una conexión de repositorio existente, editar una conexión existente o añadir una conexión. Si ya ha iniciado sesión en un repositorio utilizando IBM SPSS Modeler, al elegir la opción **Añadir/editar una conexión** aparece el cuadro de diálogo Conexiones de Enterprise View, en el que puede definir o editar los datos necesarios para la conexión actual. Si no ha iniciado sesión, al seleccionar esta opción aparecerá el cuadro de diálogo de inicio de sesión del repositorio.

Para obtener información sobre el inicio de sesión en el repositorio, consulte el *Manual del usuario de IBM SPSS Modeler*.

Una vez que se ha establecido una conexión a un repositorio, dicha conexión permanecerá hasta que salga de IBM SPSS Modeler. Otros nodos de la misma ruta pueden compartir la misma conexión, pero será necesario crear una nueva conexión para cada nueva ruta.

Si se inicia sesión correctamente, aparecerá el cuadro de diálogo Conexiones de Enterprise View.

Nombre de tabla. Este campo está vacío inicialmente y no puede rellenarse hasta que se cree una conexión. Si conoce el nombre de la tabla de vista de aplicación a la que desea acceder, introdúzcalo en el campo Nombre de tabla. En caso contrario, pulse en el botón **Seleccionar** para abrir un cuadro de diálogo que muestre las tablas de vista de aplicación disponibles.

Conexiones de Enterprise View

Este cuadro de diálogo permite definir o editar los detalles necesarios para la conexión de repositorio. Puede especificar las siguientes opciones:

- Vista de aplicación y versión
- Entorno
- Definición de proveedor de datos (DPD)
- Descripción de la conexión

Conexiones. Muestra las conexiones de repositorio existentes.

- **Añadir una nueva conexión.** Muestra el cuadro de diálogo Recuperar objeto, desde el que puede buscar y seleccionar una vista de aplicación del repositorio.
- **Copiar la conexión seleccionada.** Realiza una copia de la conexión seleccionada, lo que evita tener que volver a buscar la misma vista de aplicación.
- **Eliminar la conexión seleccionada.** Elimina la conexión seleccionada de la lista.

Detalles de la conexión. Muestra la vista de aplicación, la etiqueta de versión, el entorno, la DPD y texto descriptivo de la conexión actualmente seleccionada en el panel de conexiones.

- **Vista de aplicación.** La lista desplegable muestra la vista de aplicación seleccionada, si la hay. Si se han realizado conexiones con otras vistas de aplicación en la sesión actual, también aparecerán en la lista desplegable. Pulse en el botón Examinar que aparece al lado para buscar otras vistas de aplicación en el repositorio.

- **Etiqueta de la versión.** Los campos desplegables muestran todas las etiquetas de versión definidas para la vista de aplicación especificada. Las etiquetas de versión ayudan a identificar las versiones de los objetos del repositorio. Por ejemplo, puede haber dos versiones de una vista de aplicación determinada. Con las etiquetas, puede especificar la etiqueta **PRUEBA** para la versión utilizada en el entorno de desarrollo y la etiqueta **PRODUCCIÓN** para la versión utilizada en el entorno de producción. Seleccione la etiqueta adecuada.

Nota: Las etiquetas no deben incluir paréntesis como "(" o ")", corchetes como "[" o "]" ni caracteres de doble corchete (doble byte en japonés); de lo contrario, el nombre de tabla no se muestra en el separador Datos del recuadro de diálogo Enterprise View.

- **Entorno.** El campo desplegable muestra todos los entornos válidos. La configuración del entorno determina las DPD que estarán disponibles, especificando por tanto las columnas concretas que deberán asociarse a los segmentos empresariales definidos. Por ejemplo, cuando se selecciona **Analytic**, solamente se devuelven las columnas de la vista de aplicación definidas como **Analytic**. El entorno predeterminado es **Analytic**; también puede seleccionar **Operational**.
- **Proveedor de datos.** La lista desplegable muestra los nombres de hasta diez definiciones de proveedor de datos para la vista de aplicación seleccionada. Solamente se muestran las DPD que hacen referencia a la vista de aplicación seleccionada. Pulse en el botón Examinar que aparece al lado para ver el nombre y la ruta de todas las DPD relacionadas con la vista de aplicación actual.
- **Descripción.** Texto descriptivo sobre la conexión de repositorio. Este texto se utilizará para el nombre de la conexión; al pulsar en **Aceptar** el texto aparecerá en la lista desplegable Conexión y en la barra de título del cuadro de diálogo de Enterprise View, así como en la etiqueta del nodo Enterprise View, en el lienzo.

Selección de la DPD

El cuadro de diálogo Seleccionar proveedor de datos muestra el nombre y la ruta de todas las DPD que hacen referencia a la vista de aplicación actual.

Las vistas de aplicación pueden tener varias DPD para servir de ayuda en las diferentes fases de un proyecto. Por ejemplo, los datos históricos utilizados para generar un modelo pueden proceder de una base de datos, mientras que los datos operativos proceden de otra.

Una DPD se define respecto a un determinado origen de datos ODBC. Para usar una DPD de IBM SPSS Modeler, debe tener un origen de datos ODBC definido en el host del servidor de IBM SPSS Modeler que tenga el mismo nombre (y que conecte con el mismo almacén de datos) que a la que se hace referencia en la DPD.

Para elegir la DPD con la que desea trabajar, elija su nombre en la lista y pulse en **Aceptar**.

Selección de la tabla

El cuadro de diálogo Seleccionar tabla muestra todas las tablas a las que se hace referencia en la vista de aplicación actual. Este cuadro de diálogo aparece vacío si no se ha establecido ninguna conexión a IBM SPSS Collaboration and Deployment Services Repository.

Para elegir la tabla con la que desea trabajar, elija su nombre en la lista y pulse en **Aceptar**.

Pestañas de nodos de origen comunes

Las siguientes opciones se pueden especificar para todos los nodos de origen pulsando en la pestaña correspondiente:

- **Pestaña Datos.** Se utiliza para cambiar el tipo de almacenamiento predeterminado.
- **Pestaña Filtro.** Se utiliza para eliminar los campos de datos o cambiarles el nombre. Esta pestaña ofrece la misma funcionalidad que el nodo Filtrar. Consulte el tema “Opciones de configuración de filtrado” en la página 140 para obtener más información.

- **Pestaña Tipos.** Se utiliza para definir niveles de medición. Esta pestaña ofrece la misma funcionalidad que el nodo Tipo.
- **Pestaña Anotaciones.** Se utiliza para todos los nodos, esta pestaña proporciona opciones para cambiar el nombre de los nodos, ofrecer herramientas personalizadas y almacenar una anotación más extensa.

Definición de niveles de medición en el nodo Origen

Las propiedades de campo se pueden especificar en un nodo de origen o en un nodo Tipo independiente. La funcionalidad es similar en ambos nodos. Se trata de las siguientes propiedades:

- **Campo:** pulse dos veces en cualquier nombre de campo para especificar etiquetas de valor y de campo para los datos de IBM SPSS Modeler. Por ejemplo, los metadatos de campo importados de IBM SPSS Statistics se pueden ver o modificar aquí. Del mismo modo, puede crear nuevas etiquetas para campos y sus valores. Las etiquetas que se especifican aquí se muestran en IBM SPSS Modeler dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.
- **Medición:** este es el nivel de medición que se usa para describir características de los datos en un campo determinado. Si se conocen todos los detalles de un campo, éste se denomina **completamente instanciado**. Si desea obtener más información, consulte “Niveles de medición” en la página 127.

Nota: El nivel de medición de un campo difiere del tipo de almacenamiento de un campo, lo cual indica si los datos están almacenados como cadenas, números enteros, números reales, fechas, horas o marcas de tiempo.

- **Valores:** esta columna le permite especificar las opciones para leer los valores de datos desde el conjunto de datos, o utilizar la opción **Especificar** para especificar los niveles de medición y los valores en un cuadro de diálogo separado. También puede optar por pasar los campos sin leer sus valores. Si desea obtener más información, consulte “Valores de campos” en la página 131.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Perdidos:** se utiliza para especificar cómo se tratan los valores perdidos de un campo. Si desea obtener más información, consulte “Definición de valores perdidos” en la página 135.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Comprobar:** en esta columna puede configurar opciones que garanticen que los valores de campo se ajusten a los valores o rangos especificados. Si desea obtener más información, consulte “Comprobación de valores de tipo” en la página 136.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Rol:** se usa para indicar a los nodos de modelado si los campos serán **Entrada** (campos predictores) o **Objetivo** (campos predichos) para un proceso de aprendizaje automático. **Ambas** y **Ninguna** también son roles disponibles, junto con **Partición**, que indica un campo utilizado para crear particiones de registros en muestras independientes para entrenamiento, comprobación y validación. El valor **Split** especifica que se generarán modelos diferentes para cada valor posible del campo. Si desea obtener más información, consulte “Definición del rol de campo” en la página 136.

Consulte el tema “Nodo Tipo” en la página 125 para obtener más información.

Cuándo crear una instancia en el nodo de origen

Puede conocer dos formas acerca del almacenamiento de datos y los valores de los campos. Esta **instanciación** se puede producir en el nodo de origen, cuando introduzca primero los datos en IBM SPSS Modeler, o insertando el nodo Tipo en la ruta de datos.

La instanciación en el nodo de origen resulta útil cuando:

- El conjunto de datos es pequeño.

- Tiene previsto derivar nuevos campos mediante el generador de expresiones (la instanciación hace que los valores de campos estén disponibles desde el generador de expresiones).

Normalmente, si el conjunto de datos no es muy grande y no tiene intención de añadir más adelante ningún campo a la ruta, la instanciación en el nodo de origen es el método más conveniente.

Filtrado de campos desde el nodo de origen

La pestaña Filtro en el cuadro de diálogo del nodo de origen le permitirá excluir los campos de las operaciones que se encuentren debajo, según el examen inicial de los datos. Esto resulta útil, por ejemplo, cuando hay campos duplicados en los datos o si ya conoce los datos como para excluir los campos irrelevantes. Si lo prefiere, puede añadir un nodo Filtrar independiente en la ruta. La funcionalidad es similar en ambos casos. Consulte el tema “Opciones de configuración de filtrado” en la página 140 para obtener más información.

Capítulo 3. Nodos de operaciones con registros

Conceptos básicos de las operaciones con registros

Los nodos de operaciones con registros se usan para realizar cambios en los datos a nivel de registro. Estas operaciones son importantes durante las fases **Comprensión de los datos** y **Preparación de los datos** del análisis porque permiten adaptar los datos a las necesidades particulares de su negocio.

Por ejemplo, según los resultados de la auditoría de datos realizada con el nodo Auditoría de datos (paleta Resultado), puede que desee fusionar los registros de las compras realizadas por los clientes durante los últimos tres meses. Con el nodo Fundir, puede fusionar registros basándose en los valores de un campo clave, como el *ID de cliente*. Así mismo, puede descubrir que es imposible administrar una base de datos con información sobre visitas al sitio Web con más de un millón de registros. Con el nodo Muestrear, puede seleccionar un subconjunto de datos para utilizarlo en el modelado.

La paleta Operaciones con registros contiene los siguientes nodos:



El nodo Seleccionar selecciona o descarta un subconjunto de registros de la ruta de datos en función de una condición específica. Por ejemplo, podría seleccionar los registros que pertenezcan a una región de ventas determinada.



El nodo Muestrear selecciona un subconjunto de registros. Se admite una variedad de tipos de muestras, entre las que se incluyen las muestras estratificadas, agrupadas en clústeres y no aleatorias (estructuradas). El muestreo puede ser de gran utilidad para mejorar el rendimiento y para seleccionar grupos de registros o transacciones relacionadas para un análisis.



El nodo Equilibrar corrige los desequilibrios de un conjunto de datos para que cumpla una condición determinada. La directiva de equilibrado ajusta la proporción de registros si una condición es verdadera por el factor determinado.



El nodo Agregar reemplaza una secuencia de registros de entrada con registros de salida agregados y resumidos.



El nodo Adición de RFM (actualidad, frecuencia, monetario) permite tomar datos de transacciones históricas de clientes, deshacerse de los datos no utilizados y combinar todos los datos de transacciones restantes en una única fila que indica cuándo hizo negociaciones con los clientes por última vez, cuántas transacciones hicieron y el valor monetario total de dichas transacciones.



Los nodos Ordenar organizan registros en orden ascendente o descendente atendiendo a los valores de uno o varios campos.



El nodo Fundir toma varios registros de entrada y crea un registro de salida único que contiene todos o algunos de los campos de entrada. Es útil para fusionar datos desde diferentes orígenes, como datos de clientes internos y datos demográficos adquiridos.



El nodo Añadir concatena conjuntos de registros. Es útil para combinar conjuntos de datos con estructuras parecidas, pero con datos diferentes.



El nodo Distinguir se puede usar para eliminar registros duplicados pasando el primero de los registros distintos a la ruta de datos o descartando el primer registro y pasando cualquier duplicado a la ruta de datos en su lugar.



El nodo Generación de análisis TS construye y puntúa modelos de series temporales en un paso, sin la necesidad de un nodo Intervalos de tiempo.

Algunos de los nodos de la paleta Operaciones con registros requieren el uso de una expresión CLEM. Si está familiarizado con CLEM, puede introducir una expresión en el campo. Sin embargo, todos los campos de expresión ofrecen un botón que abre el generador de expresiones CLEM, que ayuda a crear estas expresiones de forma automática.



Figura 1. Botón Generador de expresiones

Nodo Seleccionar

Puede usar el nodo Seleccionar para elegir o descartar un subconjunto de registros de la ruta de datos basándose en una condición especial, como PS (presión sanguínea) = "HIGH" (alta).

Moda. Especifica si los registros que cumplen la condición se incluirán o se excluirán de la ruta de datos.

- **Incluir.** Seleccione esta opción para incluir los registros que reúnan la condición de selección.
- **Descartar.** Seleccione esta opción para excluir los registros que reúnan la condición de selección.

Condición. Muestra la condición de selección que se usará para comprobar cada registro, lo que se especifica usando una expresión CLEM. Puede escribir una expresión en la ventana o bien usar el generador de expresiones pulsando en el botón de calculadora (Generador de expresiones) ubicado en la parte derecha de la ventana.

Si selecciona descartar los registros basados en una condición, como los siguientes:

```
(var1='valor1' and var2='valor2')
```

el nodo Seleccionar también descarta de forma predeterminada los registros con valores nulos para todos los campos de selección. Para evitarlo, añada la siguiente condición a la original:

```
and not(@NULL(var1) and @NULL(var2))
```

Los nodos Seleccionar también se usan para seleccionar una proporción de registros. No obstante, por lo general, se usaría un nodo diferente, el nodo Muestrear, para esta operación. Sin embargo, si la condición

que desea especificar es más compleja que los parámetros que se proporcionan, puede crearla usando el nodo Seleccionar. Por ejemplo, podría usar una condición como ésta:

```
BP = "HIGH" and random(10) <= 4
```

Con ella se seleccionará aproximadamente el 40% de los registros que muestran alta presión sanguínea y pasarlos a la parte de abajo para sus análisis posteriores.

Nodo Muestrear

Puede utilizar los nodos Muestrear para seleccionar un subconjunto de análisis o especificar una proporción de registros que descartar. Se admite una variedad de tipos de muestras, entre las que se incluyen las muestras estratificadas, agrupadas en clústeres y no aleatorias (estructuradas). El muestreo se puede utilizar por varias razones:

- Para mejorar el rendimiento por modelos de estimación en un subconjunto de datos. Los modelos estimados de una muestra son normalmente tan precisos como los derivados del conjunto de datos completo y pueden serlo más si el rendimiento mejorado permite experimentar con diferentes métodos que de lo contrario no intentaría.
- Para seleccionar grupos de registros relacionados o transacciones de análisis, como la selección de todos los elementos en un carro de la compra en línea (o cesta de la compra) o todas las propiedades en una vecindad específica.
- Para identificar las unidades o los casos de inspección aleatoria en busca de la garantía de calidad, la prevención de fraudes o la seguridad.

Nota: Si simplemente desea realizar una partición de sus datos en muestras de comprobación y entrenamiento con fines de validación, puede utilizar un nodo Partición. Consulte el tema “Nodo Partición” en la página 165 para obtener más información.

Tipos de muestras

Muestras agrupadas en clústeres. Grupos o clústeres de muestra en lugar de unidades individuales. Por ejemplo, supongamos que tiene un archivo de datos con un registro por alumno. Si ha agrupado por colegio y el tamaño de la muestra es 50 %, se seleccionará el 50 % de colegios y se tomarán todos los alumnos de cada colegio seleccionado. Se rechazarán los alumnos de los colegios no seleccionados. Como media, esperaría una selección del 50 % de alumnos aproximadamente, pero dado que los colegios tienen distintos tamaños, el porcentaje no será exacto. De forma similar, podría agrupar elementos del carro de la compra por ID de transacciones para asegurarse de que se conservan todos los elementos de las transacciones seleccionadas. Para un ejemplo que agrupe las propiedades por ciudad, consulte la ruta de ejemplo *complexsample_property.str*.

Muestras estratificadas. Selecciona muestras de forma independiente dentro de subgrupos no superpuestos de población o estratos. Por ejemplo, puede asegurar que se toman hombres y mujeres en proporciones iguales o que se representan todas las regiones o grupos socioeconómicos dentro de una población urbana. También puede especificar un tamaño de muestra diferente para cada estrato (por ejemplo, si piensa que un grupo se ha representado en inferioridad de condiciones en los datos originales). Para un ejemplo que estratifique las propiedades por provincia, consulte la ruta de ejemplo *complexsample_property.str*.

Muestreo sistemático o 1-de cada-n. Cuando es difícil de obtener una selección aleatoria, las unidades se pueden muestrear de forma sistemática (a un intervalo fijo) o secuencial.

Ponderaciones muestrales. Las ponderaciones de muestreo se calculan automáticamente al extraer una muestra compleja y corresponde aproximadamente a la "frecuencia" que cada unidad muestreada representa en los datos originales. Por consiguiente, la suma de las ponderaciones sobre la muestra debe calcular el tamaño de los datos originales.

Marco de muestreo

Un marco de muestreo define el origen potencial de casos que se deben incluir en una muestra o estudio. En algunos casos, puede ser factible identificar cada miembro individual de una población e incluir cualquiera de ellos en una muestra, por ejemplo, cuando realizan muestreos de elementos extraídos de una línea de producción. La mayoría de las veces no podrá acceder a cada caso posible. Por ejemplo, no puede estar seguro de quién votará en una elección hasta que se haya producido la elección. En este caso, podría utilizar el registro electoral como marco de muestreo, incluso si algunas personas registradas no votan y algunas personas podrían votar a pesar de no haberse incluido en la lista en el momento que comprobó el registro. Cualquiera fuera del marco de muestreo no tiene posibilidades de poder incluirse en la muestra. La cuestión de si su marco de muestreo es lo bastante cercano a la población que intenta evaluar debe estudiarse para cada caso real.

Opciones del nodo Muestrear

Puede seleccionar el método **Simple** o **Complejo** según se adapte mejor a sus necesidades.

Opciones de muestreo simple

El método Simple permite seleccionar un porcentaje aleatorio de registros, seleccionar registros contiguos o seleccionar cada n° registro.

Moda. Seleccione si desea pasar (incluir) o descartar (excluir) registros de las siguientes maneras:

- **Incluir muestra.** Incluye los registros seleccionados en la ruta de datos y descarta todos los demás. Por ejemplo, si configura el modo como **Incluir muestra** y establece el valor de **1-de cada-n** en 5, se incluirá cada quinto registro, produciendo un conjunto de datos que representa un quinto del tamaño original aproximadamente. Es el modo predeterminado cuando se realiza un muestreo de datos y el único modo posible con el método complejo.
- **Descartar muestra.** Excluye los registros seleccionados e incluye todos los demás. Por ejemplo, si configura el modo en **Descartar muestra** y establece el valor de **1-de cada-n** en 5, se descartará cada quinto informe. Este nodo sólo está disponible con el método simple.

Muestra. Seleccione el método de muestreo de entre las opciones siguientes:

- **Primeros.** Seleccione esta opción para usar muestro de datos contiguos. Por ejemplo, si el tamaño máximo de la muestra se define como 10000, se seleccionarán los primeros 10.000 registros.
- **1-de cada-n.** Seleccione esta opción para muestrear datos pasando o descartando cada *enésimo* registro. Por ejemplo, si n se define como 5, se seleccionarán cada cinco registros.
- **Aleatorio %.** Seleccione esta opción para muestrear un porcentaje aleatorio de los datos. Por ejemplo, si el porcentaje se establece en 20, se pasará a la ruta de datos o se descartará el 20% de los datos, dependiendo del modo seleccionado. Use el campo para especificar un porcentaje de muestreo. También puede especificar un valor de semilla usando el control **Establecer semilla aleatoria**.

Utilizar muestreo de nivel de bloque (sólo en base de datos). Esta opción sólo está habilitada si selecciona el muestreo de porcentaje aleatorio al realizar minería interna de bases de datos en una base de datos de Oracle o IBM DB2. En estas circunstancias, el muestreo de nivel de bloque puede ser más eficaz.

Tamaño máximo de muestra. Especifica el número máximo de registros que se incluyen en la muestra. Esta opción es redundante, por lo que se desactiva cuando se seleccionan **Primeros** e **Incluir**. Asimismo tenga en cuenta que cuando se utiliza junto con la opción **Aleatorio %**, esta configuración puede evitar que se seleccionen determinados registros. Por ejemplo, si tiene 10 millones de registros en la base de datos y selecciona el 50 % de ellos con un tamaño de muestra máximo de tres millones de registros, sólo se seleccionará el 50 % de los primeros seis millones de registros y los cuatro millones restantes no tendrán probabilidades de ser seleccionados. Para evitar esta limitación, seleccione el método de muestreo **Complejo** y solicite una muestra aleatoria de tres millones de registros sin especificar un clúster o variable de estratificación.

Opciones de muestreo complejo

Las opciones de muestreo complejo permiten un control más exhaustivo de la muestra, incluidas las muestras agrupadas en clústeres, estratificadas y ponderadas junto con otras opciones.

Clúster y estratificación. Permite especificar campos de ponderación de entrada, estratificación y clúster en caso necesario. Consulte el tema “Configuración de Clúster y estratificación” para obtener más información.

Tipo de muestra.

- **Aleatorios.** Selecciona clústeres o registros aleatoriamente dentro de cada estrato.
- **Systematic.** Selecciona registros en un intervalo fijo. Esta opción funciona como el método *1 de cada n*, salvo que la posición del primer registro cambia dependiendo de una semilla aleatoria. El valor de *n* se determina automáticamente en base al tamaño de la muestra o la proporción.

Unidades de muestra. Puede seleccionar proporciones o recuentos como unidades de muestra básicas.

Tamaño muestral. Puede especificar el tamaño de la muestra de varias formas:

- **Fija.** Permite especificar el tamaño general de la muestra como un recuento o proporción.
- **Personalizado.** Permite especificar el tamaño de la muestra para cada subgrupo o estrato. Esta opción sólo está disponible si se ha especificado un campo de estratificación en el subcuadro de diálogo Clúster y estratificación.
- **Variable.** Permite al usuario seleccionar un campo que define el tamaño de la muestra para cada subgrupo o estrato. Este campo debe tener el mismo valor para cada registro dentro de un estrato particular; por ejemplo, si la muestra se estratifica por provincia, todos los registros con *county = Surrey* deben tener el mismo valor. El campo debe ser numérico y sus valores deben coincidir con las unidades de muestra seleccionadas. Para las proporciones, los valores deben ser mayores que 0 y menor que 1; para los recuentos, el valor mínimo es 1.

Muestra mínima por estrato. Especifica un número mínimo de registros (o un número mínimo de clústeres si se especifica un campo de clúster).

Muestra máxima por estrato. Especifica un número máximo de registros o clústeres. Si selecciona esta opción sin especificar un clúster o campo de estratificación, se seleccionará una muestra aleatoria o sistemática del tamaño especificado.

Establecer semilla aleatoria. Cuando se muestran o particionan registros a partir de un porcentaje aleatorio, esta opción permite duplicar los mismos resultados en otra sesión. Al especificar el valor inicial utilizado por el generador de números aleatorios, puede garantizar que se asignan los mismos registros cada vez que se ejecuta el nodo. Introduzca el valor de semilla deseado, o pulse en el botón **Generar** para generar automáticamente un valor aleatorio. Si no se selecciona esta opción, se generará una muestra diferente cada vez que se ejecute el nodo.

Note: cuando se utiliza la opción **Establecer semilla aleatoria** con registros leídos de una base de datos, puede ser necesario un nodo Ordenar, antes del muestreo con el fin de garantizar el mismo resultado cada vez que se ejecute el nodo. Esto se debe a que la semilla aleatoria depende del orden de registros, sin estar garantizado que sea el mismo en una base de datos relacional. Consulte el tema “Nodo Ordenar” en la página 83 para obtener más información.

Configuración de Clúster y estratificación

El cuadro de diálogo Clúster y estratificación permite seleccionar campos de ponderación, estratificación y clúster cuando extrae una muestra compleja.

Clústeres. Especifica un campo categórico empleado para agrupar registros. Los registros se muestrean en base a su pertenencia al clúster, incluyendo unos clústeres y otros no. Pero si se incluye cualquier registro de un clúster determinado, se incluirán todos. Por ejemplo, al analizar las asociaciones de producto en los carros de la compra, podría agrupar elementos por ID de transacciones para asegurarse de que se conservan todos los elementos de las transacciones seleccionadas. En lugar de muestrear registros, lo cual destruiría la información sobre los elementos que se venden conjuntamente, puede muestrear transacciones para asegurarse de que se conservan todos los registros de las transacciones seleccionadas.

Estratificar por. Especifica un campo categórico utilizado para estratificar los registros de manera que las muestras se seleccionan de forma independiente dentro de subgrupos no superpuestos de población o estratos. Si selecciona una muestra del 50 % estratificada por género, por ejemplo, se tomarán dos muestras del 50 %, una para hombres y otra para mujeres. Por ejemplo, los estratos pueden ser grupos socioeconómicos, categorías laborales, grupos de edades o grupos étnicos, permitiendo garantizar los tamaños de muestra adecuados para los subgrupos de interés. Si hay el triple de mujeres que de hombres en el conjunto de datos original, esta relación se mantendrá realizando muestreos de forma separada para cada grupo. También se pueden especificar varios campos de estratificación (por ejemplo, mediante el muestreo de líneas de producto dentro de regiones o viceversa).

Nota: si estratifica por un campo que tiene valores perdidos (los valores perdidos del sistema o nulos, cadenas vacías, espacios blancos y valores perdidos definidos por el usuario o en blanco), no podrá especificar tamaños de muestra personalizados por estratos. Si desea utilizar los tamaños de muestras personalizados al estratificar por campo con valores en blanco o perdidos, tendrá que especificarlos antes.

Utilizar ponderación de entrada. Especifica un campo empleado para ponderar registros antes del muestreo. Por ejemplo, si el campo de ponderación tiene valores comprendidos entre 1 y 5, los registros ponderados con 5 tienen una probabilidad de cinco de ser seleccionados. Los valores de este campo se sobrescribirán con las ponderaciones de salida final generadas por el nodo (consulte el siguiente párrafo).

Nueva ponderación de resultados. Especifica el nombre del campo donde se incluyen las ponderaciones finales si no se especifica ningún campo de ponderación de entrada. (Si se especifica un campo de ponderación de entrada, sus valores se sustituyen por las ponderaciones finales como se indica arriba y no se crea ningún campo de ponderación de resultados independiente.) Los valores de ponderación de resultados indican el número de registros representados por cada registro muestreado en los datos originales. La suma de los valores ponderados ofrece una estimación del tamaño de la muestra. Por ejemplo, si se toma una muestra aleatoria del 10 %, la ponderación de resultados será 10 para todos los registros, indicando que cada registro muestreado representa aproximadamente 10 registros en los datos originales. En una muestra estratificada o ponderada, los valores de ponderación de resultados pueden variar en base a la proporción de la muestra para cada estrato.

Comentarios

- El muestreo clúster es útil si no puede obtener una lista completa de la población que desea muestrear, pero puede obtener listas completas para determinados grupos o clústeres. También se utiliza cuando una muestra aleatoria produce una lista de sujetos de comprobación con los que sería poco práctico ponerse en contacto. Por ejemplo, sería más fácil visitar a todos los granjeros de una provincia que a una selección de granjeros de distintas provincias dentro del mismo país.
- Puede especificar tanto el clúster como los campos de estratificación para muestrear los clústeres de forma independiente para cada estrato. Por ejemplo, puede muestrear valores de la propiedad estratificados por provincia y el clúster por ciudad dentro de cada provincia. Así, garantizará que una muestra independiente de ciudades se extrae de cada provincia. Algunas ciudades se incluirán y otras no, pero cada ciudad que se incluya, se considerarán todas las propiedades de la ciudad.
- Para seleccionar una muestra aleatoria de unidades dentro de cada clúster, puede encadenar dos nodos Muestrear. Por ejemplo, primero podría muestrear municipios estratificados por provincia como se ha descrito anteriormente. A continuación, se añadiría un segundo nodo Muestrear y seleccionaría *ciudad* como campo de estratificación, lo que le permitiría muestrear una proporción de registros de cada municipio.

- En los casos en los que se requiere una combinación de campos para identificar clústeres de forma exclusiva, se puede generar un nuevo campo mediante un nodo Derivar. Por ejemplo, si varias tiendas utilizan el mismo sistema de numeración para las transacciones, podría derivar un nuevo campo que concatene la tienda y los ID de transacción.

Tamaños de muestra por estratos

Cuando extraiga una muestra estratificada, la opción predeterminada es muestrear la misma proporción de registros o clústeres de cada estrato. Si un grupo supera a otro por un factor de 3, por ejemplo, normalmente querrá mantener la misma relación en la muestra. Sin embargo, si no es el caso, puede especificar el tamaño de la muestra de forma independiente para cada estrato.

El cuadro de diálogo Tamaños de muestra por estratos enumera cada valor del campo de estratificación, permitiendo sobrescribir el valor predeterminado para ese estrato. Si se seleccionan varios campos de estratificación, se enumeran todas las combinaciones de valores posibles, permitiendo especificar el tamaño de cada grupo étnico dentro de cada ciudad, por ejemplo, o cada ciudad dentro de cada provincia. Los tamaños se especifican como proporciones o recuentos, según determina la configuración actual del nodo Muestrear.

Especificación de tamaños de muestra por estratos

1. En el nodo Muestrear, seleccione **Complejo** y seleccione uno o más campos de estratificación. Consulte el tema “Configuración de Clúster y estratificación” en la página 75 para obtener más información.
2. Seleccione **Personalizado** y **Especificar tamaños**.
3. En el cuadro de diálogo Tamaños de muestra por estratos, pulse en el botón **Leer valores** en la parte inferior izquierda para rellenar la presentación. En caso necesario, es posible que tenga que instanciar los valores en un nodo Tipo o de origen anterior de la ruta. Consulte el tema “¿Qué es la instanciación?” en la página 130 para obtener más información.
4. Pulse en cualquier fila para sobrescribir el tamaño predeterminado de ese estrato.

Notas sobre el tamaño de la muestra

Los tamaños de muestra personalizados pueden ser útiles si los distintos estratos tienen varianzas diferentes, por ejemplo, para hacer que los tamaños de la muestra sean proporcionales a la desviación estándar. (Si los casos dentro del estrato son más variados, tendrá que muestrear más estratos para obtener una muestra representativa.) O si un estrato es pequeño, es posible que desee utilizar una proporción de muestra mayor para asegurar que se ha incluido un número mínimo de observaciones.

Note: Si estratifica por un campo que tiene valores perdidos (los valores perdidos del sistema o nulos, cadenas vacías, espacios blancos y valores perdidos definidos por el usuario o en blanco), no podrá especificar tamaños de muestra personalizados por estratos. Si desea utilizar los tamaños de muestras personalizados al estratificar por campo con valores en blanco o perdidos, tendrá que especificarlos antes.

Nodo Equilibrar

Los nodos Equilibrar se pueden usar para corregir los desequilibrios de los conjuntos de datos de modo que cumplan determinados criterios de comprobación. Por ejemplo, imagine que un conjunto de datos contiene sólo dos valores, *bajo* o *alto*, y que el 90% de los casos es *bajo* y sólo el 10% de ellos es *alto*. Muchas técnicas de modelado presentan problemas con estos datos sesgados, puesto que tenderán a aprender sólo el resultado *bajo* y omitirán el valor *alto*, puesto que es más inusual. Si los datos están bien balanceados con aproximadamente el mismo número de resultados de *bajo* y *alto*, los modelos tendrán más posibilidades de encontrar patrones que hagan la distinción entre los dos grupos. En este caso, un nodo Equilibrar resulta útil para la creación de una directiva de equilibrado que reduzca los casos con un resultado *bajo*.

El equilibrado se lleva a cabo mediante el duplicado y posterior descarte de registros basándose en las condiciones especificadas. Los registros para los que no se establece ninguna condición siempre se pasan. Como este proceso funciona duplicando y descartando registros, la secuencia original de los datos se pierde en las operaciones efectuadas más abajo. Asegúrese de derivar cualquier valor relacionado con la secuencia antes de añadir un nodo Equilibrar a la ruta de datos.

Note: Los nodos Equilibrar se pueden generar de forma automática desde histogramas y gráficos de distribución. Por ejemplo, puede equilibrar sus datos para mostrar proporciones iguales en todas las categorías de un campo categórico, como se muestra en un gráfico de distribución.

Ejemplo. Cuando genere una ruta de RFM para identificar clientes recientes que han respondido de forma positiva a campañas de marketing anteriores, el departamento de marketing de una empresa de ventas utiliza un nodo Equilibrar para equilibrar las diferencias entre las respuestas verdaderas y falsas de los datos.

Opciones de configuración del nodo Equilibrar

Directivas de equilibrado de registros. Enumera las directivas de equilibrado actuales. Cada directiva incluye un factor y una condición que indica al software que "aumente la proporción de registros por un factor determinado si la condición es verdadera." Un factor inferior a 1,0 significa que la proporción de registros indicada se reducirá. Por ejemplo, si desea reducir el número de registros en los que el medicamento Y es el adecuado para el tratamiento, podría crear una directiva de equilibrado con un factor 0,7 y una condición Drug = "drugY". Esta directiva significa que el número de registros en los que el medicamento Y es el apropiado para el tratamiento se reducirá al 70% en todas las operaciones siguientes.

Note: los factores de equilibrado para la reducción se deben especificar con cuatro posiciones decimales. Los factores con valores inferiores a 0,0001 generarán un error, ya que los resultados no se calculan correctamente.

- **Cree condiciones** pulsando en el botón situado a la derecha del campo de texto. Así se insertará una fila vacía para introducir nuevas condiciones. Si desea crear una expresión CLEM para la condición, pulse en el botón Generador de expresiones.
- **Elimine directivas** usando el botón de eliminación rojo.
- **Ordene directivas** usando los botones de flecha arriba y abajo.

Sólo datos de entrenamiento de equilibrado Si un campo de partición está presente en la ruta, esta opción equilibra solamente los datos de la partición de entrenamiento. En especial, puede ser útil si genera puntuaciones ajustadas de propensión, que requiere una comprobación desequilibrada o partición de validación. Si no hay campos de partición presentes en la secuencia (o si se especifican varios campos de partición), esta opción se ignora y se equilibran todos los datos.

Nodo Agregar

La agregación es una tarea de preparación de datos que se usa con frecuencia para reducir el tamaño de un conjunto de datos. Antes de comenzar a añadir elementos, debe dedicar un tiempo a limpiar los datos, concentrándose especialmente en los valores perdidos. Una vez que haya realizado la agregación, es posible que se pierda información potencialmente útil sobre valores perdidos.

El nodo Agregar se puede usar para reemplazar una secuencia de registros de entrada con registros de salida agregados de resumen. Por ejemplo, podría tener un conjunto de registros de entradas de ventas como el que se muestra en la tabla siguiente:

Tabla 13. Ejemplo de registro de entradas de ventas

Edad	Sexo	Región	Rama	Ventas
23	M	S	8	4

Tabla 13. Ejemplo de registro de entradas de ventas (continuación)

Edad	Sexo	Región	Rama	Ventas
45	M	S	16	4
37	M	S	8	5
30	M	S	5	7
44	M	N	4	9
25	M	N	2	11
29	F	S	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

Se pueden agregar estos registros con *Sexo* y *Región* como campos clave. A continuación puede agregar *Edad* con el modo **Media** y *Ventas* con el modo **Suma**. Seleccione **Incluir recuento de registros en campo** en el cuadro de diálogo del nodo Agregar para que el resultado agregado sea el que se muestra en la tabla siguiente:

Tabla 14. Ejemplo de registro agregado

Edad (media)	Sexo	Región	Ventas (suma)	Recuento de registros
35,5	F	N	25	4
29	F	S	6	1
34,5	M	N	20	2
33.75	M	S	20	4

Del ejemplo se entiende, por ejemplo, que la edad media de las cuatro mujeres del personal de ventas en la región norte es de 35,5, y que la suma de sus ventas totales es de 25 unidades.

Note: los campos como *Rama* se descartan automáticamente cuando no se especifica ningún modo de agregación.

Opciones de configuración del nodo Agregar

En el nodo Agregar puede especificar lo siguiente.

- Una o más claves de registro que se utilizarán como categorías para la agregación
- Uno o más campos de agregación para calcular los valores de agregación
- Uno o más modos de agregación (tipos de agregación) para obtener el resultado de cada campo de agregación

También puede especificar los modos de agregación predeterminados que se utilizarán para utilizar los campos recién añadidos y utilizar expresiones (similares a las fórmulas) para categorizar la agregación.

Tengan en cuenta que para mejorar el rendimiento, las operaciones de agregación pueden mejorarse si se activa el procesamiento en paralelo.

Campos clave. Enumera los campos que se pueden usar como categorías para la agregación. Tanto los campos (numéricos) como los campos categóricos se pueden usar como claves. Si selecciona más de un campo clave, los valores se combinarán para producir un valor clave de la agregación de registros. Se generará un registro agregado para cada campo clave exclusivo. Por ejemplo, si *Sexo* y *Región* son los

campos clave, cada combinación exclusiva de *V* y *M* con las regiones *N* y *S* (cuatro combinaciones exclusivas) tendrá un registro agregado. Para añadir un campo clave, utilice el botón de selección de campos de la parte derecha de la ventana.

El resto del cuadro de diálogo se divide en dos áreas principales: **Agregados básicos** y **Expresiones de agregados**.

Agregados básicos

Agregar campos. Enumera los campos cuyos valores se añadirán, así como los modos de agregación seleccionados. Para añadir campos a esta lista, utilice el botón de selección de campos de la parte derecha. Los siguientes modos de agregación están disponibles:

Nota: Algunos de los modos no son aplicables a campos no numéricos (por ejemplo, **Sum** para un campo de fecha/hora). Los modos que no se pueden utilizar con un campo de agregación seleccionada están desactivados.

- **Suma.** Seleccione esta opción para devolver los resultados sumados de cada combinación de campos clave. La suma es el total de todos los valores, a lo largo de todos los casos que no tengan valores perdidos.
- **Media.** Seleccione esta opción para devolver el valor de la media de cada combinación de campos clave. La media es una medida de la tendencia central, y es el promedio aritmético (la suma dividida por el número de casos).
- **Mín.** Seleccione esta opción para devolver los valores mínimos de cada combinación de campos clave.
- **Máx.** Seleccione esta opción para devolver los valores máximos de cada combinación de campos clave.
- **D.T.** Seleccione esta opción para devolver la desviación estándar de cada combinación de campos clave. La desviación estándar es una medida de la dispersión en torno a la media, igual a la raíz cuadrada de la medición de la varianza.
- **Mediana.** Seleccione esta opción para devolver el valor de la mediana de cada combinación de campos clave. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos). También se denomina percentil 50° o 2° cuartil.
- **Recuento.** Seleccione esta opción para devolver el recuento de valores no nulos de de cada combinación de campos clave.
- **Varianza.** Seleccione esta opción para devolver el valor de la varianza de cada combinación de campos clave. La varianza es una medida de dispersión en torno a la media, igual a la suma de las desviaciones al cuadrado respecto a la media, dividida por el número de casos menos 1.
- **1° cuartil.** Seleccione esta opción para devolver los valores del 1° cuartil (percentil 25°) de cada combinación de campos clave.
- **3° cuartil.** Seleccione esta opción para devolver los valores del 3° cuartil (percentil 75°) de cada combinación de campos clave.

Nota: Cuando se ejecuta una secuencia que contiene un nodo **Agregar**, los valores que se devuelven para el primer y tercer Cuartiles cuando se vuelve a enviar el SQL a una base de datos Oracle pueden ser diferentes de los que se devuelven en la modalidad nativa.

Modo predeterminado. Especifique el modo de agregación predeterminado que se debe usar con los campos recientemente añadidos. Si utiliza con frecuencia la misma agregación, seleccione uno o varios modos aquí y use el botón **Aplicar** a todos de la parte derecha para que los modos seleccionados se apliquen a todos los campos de la lista.

Nueva extensión de nombre de campo. Seleccione esta opción para añadir un sufijo o prefijo, como "1" o "nuevo", a los campos agregados duplicados. Por ejemplo, el resultado de una agregación de valores mínimos en el campo *Edad* producirá un nombre de campo denominado *Edad_Mín_1* se seleccionó la

opción de sufijo y especifica "1" como extensión. *Note:* las extensiones de agregación como *_Mín* o *Máx_* se añaden automáticamente al nuevo campo, indicando el tipo de agregación realizada. Seleccione **Sufijo** o **Prefijo** para indicar el estilo de extensión que prefiera.

Incluir recuento de registros en campo. Seleccione esta opción para incluir un campo adicional en cada registro de salida con el nombre *Recuento_registros*, de forma predeterminada. Este campo indica la cantidad de registros de entrada añadidos para conformar cada registro añadido. Asigne un nombre personalizado a este campo escribiéndolo directamente en el cuadro de edición.

Nota: cuando se calculan agregados pero no se incluyen en el recuento de registros, se ejecutan valores nulos del sistema. Por otro lado, se incluyen valores vacíos en la agregación y el recuento de registros. Para excluir los valores vacíos, puede usar el nodo Rellenar, que permite sustituir los espacios vacíos por valores nulos. También puede eliminar los espacios vacíos usando un nodo Seleccionar.

Expresiones de agregados

Las expresiones son similares a las fórmulas que se crean a partir de valores, nombres de campos, operadores y funciones. A diferencia de las funciones que funcionan en un solo registro cada vez, las expresiones de agregado operan en un grupo, conjunto o colección de registros.

Nota: Solo puede crear expresiones de agregados si la ruta incluye una conexión de base de datos (mediante un nodo de origen Base de datos).

Las nuevas expresiones se crean como campos derivados. Para crear una expresión utilice las funciones de *Agregados de base de datos* que están disponibles en el Constructor de expresiones.

Para obtener más información acerca del Constructor de expresiones, consulte la publicación IBM SPSS Modeler Guía del usuario (ModelerUsersGuide.pdf).

Tenga en cuenta que existe una conexión entre los **Campos de claves** y cualquier expresión de agregado que cree, ya que las expresiones de agregados se agrupan por el campo de clave.

Las expresiones de agregados válidas son las que se evalúan en los resultados de la agregación. Los siguientes son dos ejemplos de expresiones de agregados válidas y las reglas por las que se rigen, como se indica a continuación:

- Puede utilizar las funciones escalar para combinar varias funciones de agregación conjuntamente para generar un solo resultado de agregación. Por ejemplo:
`max(C01) - min(C01)`
- Una función de agregación puede operar en el resultado de varias funciones escalares. Por ejemplo:
`sum (C01*C01)`

Valores de optimización de la agregación

En el separador Optimización, especifique lo siguiente.

Las claves son contiguas. Seleccione esta opción si sabe que todos los registros con los mismos valores clave se agrupan en la entrada (por ejemplo, si la entrada se clasifica en los campos clave). Con ello puede mejorar el rendimiento.

Permitir aproximación para medianas y cuartiles. Las estadísticas de orden (mediana, primer cuartil y tercer cuartil) no se soportan de forma simultánea cuando se procesan datos en Analytic Server. Si utiliza Analytic Server, puede seleccionar este recuadro de selección para utilizar, en su lugar, un valor aproximado para estas estadísticas, el cual se calcula agrupando en intervalos los datos y, a continuación, calculando una estimación para las estadísticas basada en la distribución entre los intervalos. De forma predeterminada, esta opción no está seleccionada.

Número de intervalos. Solo está disponible si selecciona el recuadro de selección **Permitir aproximación para medianas y cuartiles**. Seleccione el número de intervalos que se ha de utilizar cuando se calculan las estadísticas. El número de intervalos afecta el **Porcentaje de error máximo**. De forma predeterminada, el número de intervalos es 1000, el cual corresponde a un error máximo del porcentaje de 0,1 del rango.

Nodo Adición de RFM

El nodo Adición de RFM (actualidad, frecuencia, monetario) permite tomar datos de transacciones históricas de clientes, deshacerse de los datos no utilizados y combinar todos los datos de transacciones restantes en una única fila, utilizando el ID de cliente exclusivo como clave, que enumera cuándo hizo negociaciones con los clientes por última vez (actualidad), cuántas transacciones hicieron (frecuencia) y el valor total de dichas transacciones (monetario).

Antes de comenzar a añadir elementos, debe dedicar un tiempo a limpiar los datos, concentrándose especialmente en los valores perdidos.

Una vez que haya identificado y transformado los datos mediante el nodo Adición de RFM, puede utilizar un nodo Análisis de RFM para realizar más análisis. Consulte el tema “Nodo Análisis de RFM” en la página 161 para obtener más información.

Tenga en cuenta que una vez ejecutado el archivo de datos mediante el nodo Adición de RFM, no tendrá ningún valor objetivo, por consiguiente, antes de poder utilizarlo como entradas para análisis predictivos adicionales con nodos de modelado como C5.0 o CHAID, tendrá que fusionarlos con otros datos de cliente (por ejemplo, haciendo coincidir los ID de cliente). Consulte el tema “Nodo Fundir” en la página 84 para obtener más información.

Los nodos Adición de RFM y Análisis de RFM de IBM SPSS Modeler están configurados para utilizar intervalos independientes; es decir, clasifican y establecen en intervalos los datos de cada medida de actividades recientes, frecuencia y valor monetario, sin considerar sus valores o las otras dos medidas.

Opciones de configuración del nodo Adición de RFM

La pestaña Configuración del nodo Adición de RFM contiene los campos siguientes.

Calcular actualidad relativa a. Especifica la fecha a partir de la que se calculará la actualidad de las transacciones. Puede ser una **Fecha fija** que introduzca o **Fecha de hoy**, definida por su sistema. **Fecha de hoy** se introduce de forma predeterminada y se actualiza automáticamente cuando se ejecuta el nodo.

Los ID son contiguos. Si los datos se han clasificado previamente de forma que todos los registros con el mismo ID aparecen en la ruta de datos, seleccione esta opción para que el procesamiento sea más rápido. Si los datos no se han clasificado previamente (o no lo sabe a ciencia cierta), no active esta opción y el nodo clasificará los datos automáticamente.

ID. Seleccione el campo que desea utilizar para identificar el cliente y sus transacciones. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha.

Fecha. Seleccione el campo de fecha que se utilizará para calcular la actualidad. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha.

Recuerde que para ello se necesita un campo de fecha o marca de tiempo con el formato apropiado para utilizarlo como entrada. Por ejemplo, si tiene un campo de cadena con valores como *Ene 2007*, *Feb 2007*, etc., puede convertirlo en un campo de fecha mediante el nodo Rellenar y la función `to_date()`. Consulte el tema “Conversión del almacenamiento mediante el nodo Rellenar” en la página 150 para obtener más información.

Valor. Selecciona el campo que se utilizará para calcular el valor monetario total de las transacciones del cliente. Para mostrar los campos que puede seleccionar, utilice el botón Selector de campos de la parte derecha. *Note:* Este valor debe ser numérico.

Nueva extensión de nombre de campo. Seleccione esta opción para añadir un sufijo o prefijo, como "12_month", a los campos actualidad, frecuencia y monetario generados recientemente. Seleccione **Sufijo** o **Prefijo** para indicar el estilo de extensión que prefiera. Por ejemplo, puede ser útil cuando examine varios períodos de tiempo.

Descartar registros con valor inferior a. En caso necesario, puede especificar un valor mínimo por debajo del cual no se utilice ningún detalle de transacción al calcular los totales de RFM. Las unidades del valor se relacionan con el campo **Valor** seleccionado.

Incluir sólo transacciones recientes. Si analiza una base de datos grande, puede especificar que sólo se utilicen los últimos registros. Puede seleccionar utilizar los datos registrados incluso después de una determinada fecha o dentro de un período reciente:

- **Fecha de transacción después de.** Especifica la fecha de transacción tras la que se incluirán los registros en su análisis.
- **Transacción dentro del último.** Especifica el número y tipo de períodos (días, semanas, meses o años) desde la fecha **Calcular actualidad relativa a** tras la cual se incluirán los registros en su análisis.

Guardar fecha de la segunda transacción más reciente. Si desea conocer la fecha de la segunda transacción más reciente para cada cliente, seleccione esta casilla. Asimismo, puede seleccionar la casilla **Guardar fecha de la tercera transacción más reciente.** Por ejemplo, puede ayudarle a identificar clientes que hayan podido realizar muchas transacciones hace bastante tiempo, pero sólo una transacción reciente.

Nodo Ordenar

Los nodos Ordenar se pueden usar para organizar registros en orden ascendente o descendente atendiendo a los valores de uno o varios campos. Por ejemplo, los nodos Ordenar se usan con frecuencia para ver y seleccionar registros con los valores de datos más comunes. Generalmente, primero se añaden los datos usando el nodo Agregar y, a continuación se usa el nodo Ordenar para organizar los datos agregados en el orden descendente del recuento de registros. Si se muestran estos resultados en una tabla, se facilita la exploración de los datos y la toma de decisiones, como la selección de registros de los 10 mejores clientes.

La pestaña Configuración del nodo Ordenar contiene los campos siguientes.

Ordenar por. Todos los campos seleccionados como claves de ordenación se muestran en una tabla. Un campo clave funciona mejor en la ordenación si es numérico.

- Para **añadir campos** a esta lista, utilice el botón de selección de campos de la parte derecha.
- **Seleccione un orden** pulsando en las flechas **Ascendente** o **Descendente** de la columna *Orden* de la tabla.
- **Elimine campos** usando el botón de eliminación rojo.
- **Ordene directivas** usando los botones de flecha arriba y abajo.

Orden de clasificación predeterminado. Seleccione **Ascendente** o **Descendente** para determinar el orden de clasificación predeterminado cuando se añadan nuevos campos a la tabla.

Nota: El nodo Ordenar no se aplica si hay un nodo Distinct bajo la ruta de modelo. Para obtener información sobre el nodo Distinct, consulte "nodo Distinguir" en la página 93.

Configuración de optimización de ordenación

Si trabaja con datos que ya se han ordenado por determinados campos clave, puede especificar los campos que ya se han ordenado y permitir que el sistema ordene el resto de los datos de una manera más eficaz. Por ejemplo, desea ordenar por *Edad* (orden descendente) y *Droga* (orden ascendente) pero sabe que los datos ya están ordenados por *Edad* (orden descendente).

Los datos están preordenados. Determina si los datos ya se han ordenado en función de uno o más campos.

Especificar orden de clasificación existente. Determina los campos que ya están ordenados. Puede añadir campos a la lista mediante el cuadro de diálogo Seleccionar campos. En la columna *Orden*, especifique si cada campo se ordena en orden ascendente o descendente. Si especifica varios campos, asegúrese de que los enumera en el orden de clasificación correcto. Utilice las flechas de la derecha de la lista para organizar los campos en el orden correcto. Si comete un error al especificar el orden de clasificación existente, aparecerá un error cuando ejecute la ruta que muestra el número de registro donde la clasificación no coincide con la que especificó.

Note: La velocidad de clasificación puede mejorarse al activar el procesamiento paralelo.

Nodo Fundir

La función de un nodo Fundir es tomar varios registros de entrada para crear un registro de salida que contenga todos o algunos de los campos de entrada. Se trata de una operación útil cuando se desean fusionar datos de diferentes orígenes, como datos de clientes internos y datos demográficos adquiridos. Se pueden fusionar datos de los siguientes modos.

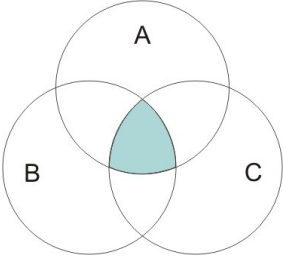
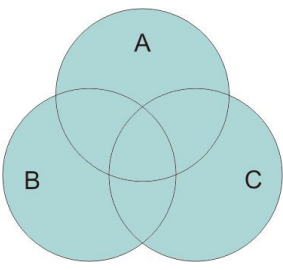
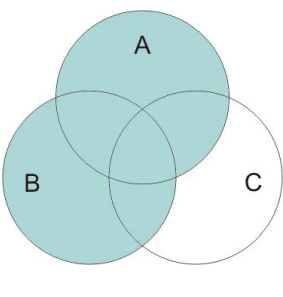
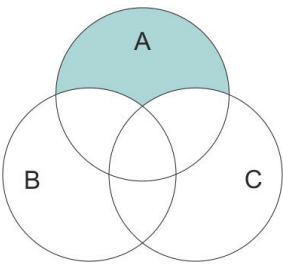
- Fusionar por **Orden** concatena registros correspondientes procedentes de todos los orígenes en el orden de entrada hasta vaciar el origen de datos más pequeño. Si se usa esta opción, es importante haber ordenado previamente los datos con un nodo Ordenar.
- Fusionar usando un campo **Clave**, como el *ID de cliente*, para especificar cómo relacionar los registros procedentes de un origen de datos con los procedentes de otros. Hay varias posibilidades de unión, incluidas la unión interior, la exterior, la exterior parcial y la anti-unión. Consulte el tema “Tipos de uniones” para obtener más información.
- Fusionar por **Condición** hace que se pueda especificar una condición que ha de cumplirse para que pueda realizarse la fusión. Se puede especificar la condición directamente en el nodo, o crear la condición utilizando el generador de expresiones.
- Fusionar por **Condición de clasificación** es una unión externa de lado izquierdo en la cual especifica una condición que se debe satisfacer para que se lleve a cabo la fusión y una expresión de clasificación que establece el orden de menor a mayor. Utilizada con mayor frecuencia para fusionar datos geoespaciales, se puede especificar la condición directamente en el nodo, o crear la condición utilizando el generador de expresiones.

Tipos de uniones

Cuando se utiliza un campo clave para la fusión de datos, resulta útil dedicar cierto tiempo a decidir qué registros se deben excluir y cuáles se deben incluir. Existe una gran variedad de uniones que se describen detalladamente a continuación.

Los dos tipos básicos de uniones se conocen como interior y exterior. Estos métodos suelen utilizarse para fusionar tablas de conjuntos de datos relacionados según los valores comunes de un campo clave, como *ID de cliente*. Las uniones interiores permiten la realización de fusiones limpias y conjuntos de datos de salida que incluyen únicamente registros completos. Las uniones exteriores también incluyen registros completos procedentes de los datos fusionados, pero también permiten la inclusión de datos exclusivos procedentes de una o varias tablas de entrada.

A continuación se describen con más detalle los tipos de unión permitidos.

	<p>Una unión interior incluye únicamente registros en los que un valor de campo clave es común a todas las tablas de entrada. Es decir, los registros no coincidentes no se incluyen en el conjunto de datos de salida.</p>
	<p>En una unión exterior total se incluyen todos los registros, tanto coincidentes como no coincidentes, procedentes de las tablas de entrada. Las uniones exteriores izquierda y derecha se denominan uniones exteriores parciales y se describen a continuación.</p>
	<p>En una unión exterior parcial se incluyen todos los registros coincidentes por un campo clave y los registros no coincidentes procedentes de las tablas especificadas. (Es decir, todos los registros de algunas etiquetas y sólo los registros coincidentes de otras.) Se pueden seleccionar tablas (como A y B en la ilustración) que incluir en la unión exterior usando el botón Seleccionar de la pestaña Fundir. Las uniones parciales también se denominan uniones exteriores izquierdas o derechas cuando sólo se fusionan dos tablas. Puesto que IBM SPSS Modeler permite la fusión de más de dos tablas, nos referimos a este tipo de unión como exterior parcial.</p>
	<p>Una anti-unión incluye únicamente registros no coincidentes de la primera tabla de entrada (la A en la ilustración). Este tipo de unión se opone a la interior y no incluye registros completos en el conjunto de datos resultante.</p>

Por ejemplo, si tiene información acerca de las granjas de un conjunto de datos y de reclamaciones del seguro relacionadas con las granjas en otro, puede establecer una correspondencia entre los registros del primer origen de datos y los del segundo mediante la opción Correlacionar a.

Para determinar si un cliente de la muestra de granjas ha rellenado una reclamación del seguro, utilice la opción de unión interior para devolver una lista que muestre dónde coinciden los ID de las dos muestras.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	clairtype	clairvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

Figura 2. Resultado de la muestra para una fusión de unión interior

Al utilizar la opción de unión exterior completa, se obtienen tanto los registros coincidentes como los no coincidentes de las tablas de entrada. El valor perdido del sistema (\$null\$) se utilizará para los valores incompletos.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalu
1	id601	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	decomm...	74703.1C
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	\$null\$	\$nul
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	arable_de	122135.0

Figura 3. Resultado de la muestra de una fusión de unión exterior completa

En una unión exterior parcial se incluyen todos los registros coincidentes por un campo clave y los registros no coincidentes procedentes de las tablas especificadas. La tabla muestra todos los registros coincidentes del campo ID así como los registros coincidentes del primer conjunto de datos.

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	\$null\$	\$null\$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	\$null\$	\$null\$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	\$null\$	\$null\$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	\$null\$	\$null\$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	\$null\$	\$null\$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	\$null\$	\$null\$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

Figura 4. Resultado de la muestra de una fusión de unión exterior parcial

Si utiliza la opción anti-unión, la tabla devolverá sólo los registros no coincidentes para la primera tabla de entrada.

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

Figura 5. Resultado de la muestra para una fusión anti-unión

Especificación de un método de fusión y claves

La pestaña Fundir del nodo Fundir contiene los campos siguientes.

Método de fusión: seleccione el método a utilizar para fusionar registros. La selección de **Claves** o **Condición** activa la mitad inferior del cuadro de diálogo.

- **Orden:** fusiona los registros de tal manera que el *enésimo* registro de cada entrada se fusione para producir el *enésimo* registro de salida. Cuando algún registro se queda sin registro de entrada correspondiente, no se producen más registros de salida. Esto quiere decir que el número de registros creados equivale al de registros en el conjunto de datos de menor tamaño.
- **Claves:** usa un campo clave, como el *ID de transacción*, para fusionar registros que compartan el mismo valor en este campo. Equivale a una "equi-unión" de base de datos. Si un valor clave tiene lugar más de una vez, se devuelven todas las combinaciones posibles. Por ejemplo, si registros con el mismo valor de campo clave *A* contienen valores diferentes *B*, *C* y *D* en otros campos, los campos fusionados producen un registro independiente para cada combinación de *A* con el valor *B*, *A* con el valor *C* y *A* con el valor *D*.

Note: los valores nulos no se consideran idénticos en el método de fusión por claves, por lo que no se unen.

- **Condición:** utilice esta opción para especificar una condición para la fusión. Si desea obtener más información, consulte "Especificación de condiciones para una fusión" en la página 87.

- **Condición de clasificación:** utilice esta opción para especificar si debe fusionarse cada emparejamiento de filas de los conjuntos de datos primario y todos los secundarios; utilice la expresión de clasificación para ordenar las coincidencias múltiples de menor a mayor. Si desea obtener más información, consulte “Especificación de condiciones clasificadas para una fusión” en la página 88.

Claves posibles: sólo enumera los campos con nombres de campo que coincidan exactamente en todos los orígenes de datos de entrada. Seleccione un campo de esta lista y use el botón de flecha para añadirlo como campo clave utilizado para la fusión de registros. Es posible usar más de un campo clave. Puede cambiar el nombre de los campos de entrada no coincidentes mediante un nodo Filtrar o la pestaña Filtro de un nodo de origen.

Claves para fusión: enumera todos los campos usados para fusionar registros desde todos los orígenes de datos de entrada basándose en valores de los campos clave. Para quitar una clave de la lista, selecciónela y use el botón de flecha para devolverla a la lista Claves posibles. Cuando se seleccionan varios campos clave, la opción de la parte inferior se desactiva.

Combinar campos de clave duplicados: cuando se seleccionan arriba varios campos clave, esta opción garantiza que sólo haya un campo de salida con dicho nombre. Esta opción está habilitada de forma predeterminada, salvo en el caso en que las rutas se hayan importado de versiones anteriores de IBM SPSS Modeler. Cuando la opción está desactivada, es preciso cambiar el nombre de los campos clave duplicados o excluirlos mediante la pestaña Filtro, en el cuadro de diálogo del nodo Fusionar.

Incluir sólo registros coincidentes (unión interna): seleccione esta opción para fusionar sólo registros completos.

Incluir registros coincidentes y no coincidentes (unión exterior completa): seleccione esta opción para realizar una "unión exterior completa". Esto quiere decir que si los valores del campo clave no están presentes en todas las tablas de entrada, los registros incompletos se conservarán. El valor no definido (\$null\$) se añade al campo clave y se incluye en el registro de salida.

Incluir registros coincidentes y no coincidentes seleccionados (unión exterior parcial): seleccione esta opción para llevar a cabo una "unión exterior parcial" de las tablas que seleccione en un cuadro de diálogo secundario. Pulse en **Seleccionar** para especificar tablas de las que conservar registros incompletos en la fusión.

Incluir registros del primer conjunto de datos que no coinciden con ningún otro (anti-unión): seleccione esta opción para llevar a cabo un tipo de "anti-unión" donde sólo se pasen a la parte de abajo los registros no coincidentes del primer conjunto de datos. Puede especificar el orden de los conjuntos de datos de entrada mediante las flechas de la pestaña Entradas. Este tipo de unión no incluye registros completos en el conjunto de datos de salida. Si desea obtener más información, consulte “Tipos de uniones” en la página 84.

Selección de datos para uniones parciales

Para llevar a cabo una unión exterior parcial, debe seleccionar las tablas para las que se deben retener registros incompletos. Por ejemplo, puede que desee conservar todos los registros de una tabla de clientes y sólo los registros de la tabla de préstamos hipotecarios.

Columna Unión exterior. En la columna *Unión exterior*, seleccione los conjuntos de datos que desee incluir íntegramente. Si desea realizar una unión parcial, se conservarán los registros superpuestos, así como los registros incompletos de los conjuntos de datos aquí seleccionados. Consulte el tema “Tipos de uniones” en la página 84 para obtener más información.

Especificación de condiciones para una fusión

Al establecer el método de fusión en **Condición**, se pueden especificar una o más condiciones que se deben cumplir para que se pueda realizar la fusión.

Se pueden introducir las condiciones directamente en el campo Condición, o crearlas con la ayuda del generador de expresiones, pulsando en el símbolo de calculadora a la derecha del campo.

Añadir etiquetas a nombres de campo duplicados para evitar conflictos de fusión: si dos o más conjuntos de datos que se van a fusionar contienen los mismos nombres de campo, marque este recuadro de selección para añadir una etiqueta de prefijo distinta al inicio de las cabeceras de columna de campo. Por ejemplo, si hay dos campos denominados *Nombre*, el resultado de la fusión contendría *1_Nombre* y *2_Nombre*. Si la etiqueta se ha renombrado en el origen de datos, se utilizará el nombre nuevo en lugar de la etiqueta de prefijo numerada. Si no marca este recuadro de selección y existen nombres duplicados en los datos, se mostrará un aviso a la derecha del recuadro de selección.

Especificación de condiciones clasificadas para una fusión

Una fusión de condiciones clasificadas se puede considerar como una fusión de unión exterior izquierda por condición. El lado izquierdo de la fusión es el conjunto de datos primario en el que cada registro es un suceso. Por ejemplo, en un modelo que se utiliza para buscar patrones de datos de crímenes, cada registro del conjunto de datos primario debe ser un crimen y su información asociada (ubicación, tipo, etc.). En este ejemplo, el extremo derecho puede contener los conjuntos de datos geoespaciales relevantes.

La fusión utiliza una condición de fusión y una expresión de clasificación. La condición de fusión puede utilizar una función geoespacial, tal como *within* o *close_to*. Durante la fusión, todos los campos de los conjuntos de datos de la derecha se añaden al conjunto de datos de la izquierda pero si existen múltiples coincidencias se genera un campo de lista. Por ejemplo:

- Lado izquierdo: Datos de crímenes
- Lado derecho: Conjunto de datos de condados y conjunto de datos de carreteras
- Condiciones de fusión: Los condados de *within* y las carreteras de *close_to* de los datos de crímenes, junto con una definición de lo que cuenta en *close_to*.

En este ejemplo, si se produce un crimen dentro de la distancia *close_to* necesaria de las tres carreteras (y el número de coincidencias que se ha de devolver se establece en al menos tres), entonces se devuelven las tres carreteras como un elemento de lista.

Si se establece el método de fusión en **Condición clasificada**, se pueden especificar una o varias condiciones que se deben cumplir para que se pueda realizar la fusión.

Conjunto de datos primario Seleccione el conjunto de datos primario para la fusión. Los campos de los otros conjuntos de datos se añaden al conjunto de datos que seleccione. Esto se puede considerar como el lado izquierdo de una fusión de unión externa.

Cuando selecciona un conjunto de datos primario, todos los otros conjuntos de datos de entrada que se conectan al nodo Fundir se listan automáticamente en la tabla **Fusiones**.

Añadir etiquetas a los nombres de campo duplicados para evitar conflictos de fusión Si dos o más conjuntos de datos que se han de fusionar contienen los mismos nombres de campos, seleccione este recuadro de selección para añadir un código de prefijo al inicio de las cabeceras de columnas de los campos. Por ejemplo, si existen dos campos *Name*, el resultado de la fusión contendrá *1_Name* y *2_Name*. Si se cambia el nombre del código en el origen de datos, se utiliza el nombre nuevo en lugar del código con el prefijo numerado. Si no selecciona este recuadro de selección y existen nombres duplicados en los datos, se muestra un aviso a la derecha del recuadro de selección.

Fusiones

Conjunto de datos

Muestra el nombre de los conjuntos de datos secundarios conectados como entradas al nodo Fundir. De forma predeterminada, si existe más de un conjunto de datos secundario, éstos se listarán en el orden en que se conecten al nodo Fundir.

Condición de fusión

Escriba las condiciones exclusivas para fusionar cada uno de los conjuntos de datos en la tabla con los datos primarios. Puede escribir las condiciones directamente en la casilla o puede crearlas con la ayuda del generador de expresiones, pulsando en el símbolo de calculadora a la derecha de la casilla. Por ejemplo, puede utilizar predicados geoespaciales para crear una condición de fusión que sustituya los datos de crímenes de un conjunto de datos dentro de los datos del condado de otro conjunto de datos. La condición de fusión predeterminada depende del nivel de medición geoespacial, como se muestra en la lista siguiente.

- Punto, Cadena de líneas, Multipunto, Cadena multilínea: condición predeterminada de *close_to*.
- Polígono, Multipolígono: condición predeterminada de *within*.

Para obtener más información acerca de estos niveles, consulte “Subniveles de medición geoespacial” en la página 129.

Si un conjunto de datos contiene varios campos geoespaciales de tipos diferentes, la condición predeterminada que se utilizará depende del primer nivel de medición que se encuentre en los datos, en el orden descendente siguiente.

- Punto
- Cadena de líneas
- Polígono

Nota: No hay valores predeterminados disponibles cuando existe un campo de datos geoespaciales en la base de datos secundaria.

Expresión de clasificación

Especifica una expresión para clasificar la fusión de los conjuntos de datos. Esta expresión se utiliza para clasificar varias coincidencias en un orden basado en el criterio de clasificación. Puede escribir las condiciones directamente en la casilla o puede crearlas con la ayuda del generador de expresiones, pulsando en el símbolo de calculadora a la derecha de la casilla.

En el generador de expresiones se proporcionan expresiones de clasificación de distancia y área predeterminadas y ambas se clasifican de menor a mayor, lo que significa que, por ejemplo, la coincidencia principal para la distancia es el valor más bajo. Un ejemplo de clasificación por distancia es cuando el conjunto de datos primario contiene crímenes y su ubicación asociada y cada conjunto de datos contiene objetos con ubicaciones. En este caso, la distancia entre los crímenes y los objetos se pueden utilizar como criterio de clasificación. La expresión de clasificación predeterminada depende del nivel de medición geoespacial, como se muestra en la lista siguiente.

- Punto, Cadena de líneas, Multipunto, Cadena multilínea: La condición predeterminada es *distancia*.
- Polígono, Multipolígono: La expresión predeterminada es *área*.

Nota: No hay valores predeterminados disponibles cuando existe un campo de datos geoespaciales en la base de datos secundaria.

Número de coincidencias

Especifique el número de coincidencias que se han de devolver, en función de las expresiones de condición y de clasificación. El número predeterminado de coincidencias depende del nivel de medición del conjunto de datos secundario, como se muestra en la lista siguiente. Sin embargo, puede pulsar dos veces la casilla para escribir su propio valor, hasta un máximo de 100.

- Punto, Cadena de líneas, Multipunto, Cadena multilínea: El valor predeterminado de 3.
- Polígono, Multipolígono: El valor predeterminado de 1.
- El conjunto de datos que no contiene campos geoespaciales: El valor predeterminado de 1.

Por ejemplo, si se configura una fusión basada en un **Condición de fusión** de *close_to* y una **Expresión de clasificación** de *distance*, las tres coincidencias principales (las más cercanas) desde los conjuntos de datos secundarios a cada uno de los registros del conjunto de datos primario, se devuelven como valores en el campo de lista resultante.

Filtrados de campos del nodo Fundir

Los nodos Fundir incorporan una forma cómoda de filtrar o cambiar el nombre a los campos resultantes de una fusión con varios orígenes de datos. Pulse en la pestaña **Filtro** del cuadro de diálogo para seleccionar las opciones de filtrado.

Las opciones que aquí se presentan son casi idénticas a las del nodo Filtrar. Existen, por tanto, opciones adicionales que aquí no se detallan y que están disponibles en el nodo Filtrar. Consulte el tema "Cambio de nombre o filtrado de campos" en la página 139 para obtener más información.

Campo. Muestra los campos de entrada de los orígenes de datos conectados actualmente.

Etiqueta. Enumera los nombres (o números) de etiqueta asociados con el enlace del origen de datos. Pulse en la pestaña **Entradas** para modificar los enlaces activos a este nodo Fundir.

Nodo de origen. Muestra el nodo de origen cuyos datos se están fusionando.

Nodo conectado. Muestra el nombre del nodo que está conectado al nodo Fundir. Con frecuencia, la minería de datos complejos requiere la realización de varias fusiones o de operaciones adicionales que pueden incluir el mismo nodo de origen. El nombre del nodo conectado proporciona una forma de diferenciarlas.

Filtro. Muestra las conexiones actuales entre el campo de entrada y salida. Las conexiones activas muestran una flecha no interrumpida. Las conexiones con una X roja indican los campos filtrados.

Campo. Enumera los campos de salida tras la fusión o adición. Los campos duplicados se muestran en rojo. Pulse en el campo Filtro para desactivar los campos duplicados.

Ver campos actuales. Seleccione esta opción para ver información sobre los campos seleccionados para ser campos clave.

Ver configuración de campos no utilizados. Seleccione esta opción para ver información sobre los campos que no están en uso actualmente.

Configuración del orden de entrada y etiquetado

El uso de la pestaña Entradas de los cuadros de diálogo de los nodos Fundir y Añadir permite especificar el orden de los orígenes de datos de entrada y efectuar modificaciones a los nombres de etiqueta de cada origen.

Etiquetas y orden de los conjuntos de datos de entrada. Seleccione esta opción para fusionar o añadir sólo registros completos.

- **Etiqueta.** Enumera los nombres de etiqueta actuales de cada origen de datos de entrada. Los nombres de etiqueta, o **etiquetas**, constituyen un modo exclusivo de identificar los enlaces de los datos en la operación de fusionar o añadir. Por ejemplo, imagine una masa de agua procedente de varias tuberías que se encuentran en un punto para continuar por una sola tubería. Los datos fluyen del mismo modo en IBM SPSS Modeler; el punto de fusión es a veces una compleja interacción entre varios orígenes de datos. Las etiquetas ofrecen una forma de administrar las entradas ("tuberías") en un nodo Fundir o Añadir de tal modo que, si éste se guarda o desconecta, los enlaces permanecen y son fácilmente identificables.

Cuando conecta datos adicionales en un nodo Fundir o Añadir, se crean automáticamente etiquetas predeterminadas a partir de números que representan el orden en que los nodos se conectaron. Este orden no está relacionado con el orden de los campos en los conjuntos de datos de entrada o salida. Se puede cambiar el orden predeterminado de las etiquetas especificando un nuevo nombre en la columna *Etiqueta*.

- **Nodo de origen.** Muestra el nodo de origen cuyos datos se están combinando.
- **Nodo conectado.** Muestra el nombre del nodo que está conectado al nodo Fundir o Añadir. Con frecuencia, la minería de datos complejos requiere la realización de varias operaciones de fusión que pueden incluir el mismo nodo de origen. El nombre del nodo conectado proporciona una forma de diferenciarlas.
- **Campos.** Detalla el número de campos de cada origen de datos.

Ver etiquetas actuales. Seleccione esta opción para ver las etiquetas que están usando de forma activa los nodos Fundir o Añadir. En otras palabras, las etiquetas actuales identifican enlaces al nodo que contienen los datos que fluyen. Volviendo a la metáfora de las tuberías, las etiquetas actuales son las tuberías con flujo de agua activo.

Ver configuración de etiquetas no utilizadas. Seleccione esta opción para ver etiquetas o enlaces previamente usados para la conexión con los nodos Fundir o Añadir pero que actualmente no están conectados a ningún origen de datos. Serían las tuberías vacías aún intactas de un sistema de fontanería. Puede optar por conectar estas "tuberías" a un nuevo origen o quitarlas. Para quitar etiquetas no utilizadas del nodo, pulse en **Borrar**. Al hacerlo, se borrarán todas las etiquetas no utilizadas de una vez.

Configuración de optimización de fusión

El sistema ofrece dos opciones que pueden ayudar a fusionar datos de una manera más eficaz en ciertas situaciones. Estas opciones permiten optimizar la fusión cuando el conjunto de datos de entrada es significativamente mayor que los otros conjuntos de datos o cuando los datos ya se han ordenado por todos o algunos de los campos clave que utiliza para la fusión.

Nota: Las optimizaciones de este separador sólo se aplican a la ejecución del nodo nativo de IBM SPSS Modeler; es decir, cuando el nodo Fundir no retrotrae en SQL. Los valores de optimización no afectan a la generación SQL.

Un conjunto de datos de entrada es relativamente grande. Seleccione esta opción para indicar que uno de los conjuntos de datos de entrada es mucho mayor que los otros. El sistema almacenará en la caché los conjuntos de datos de menor tamaño de la memoria y, a continuación, realizará la fusión al procesar el conjunto de datos grande sin almacenarla en caché ni ordenarla. Este tipo de unión se suele utilizar con los datos que se diseñan mediante un esquema de estrella o de diseño similar, donde existe una tabla central de gran tamaño de datos compartidos (por ejemplo, en los datos transaccionales). Si selecciona esta opción, pulse en **Seleccionar** para especificar el conjunto de datos grande. Tenga en cuenta que sólo puede seleccionar *un* conjunto de datos grande. La siguiente tabla resume las uniones que se pueden optimizar con este método.

Tabla 15. Resumen de optimizaciones de unión.

Tipo de unión	¿Se puede optimizar para un conjunto de datos de entrada grande?
Interior	Sí
Parcial	Sí, si no hay ningún registro incompleto en el conjunto de datos grande.
Completa	No
Anti-unión	Sí, si el conjunto de datos grande es la primera entrada.

Todas las entradas ya están ordenadas por campos clave. Seleccione esta opción para indicar que los datos de entrada ya están ordenados por uno o más de los campos clave que está utilizando para la fusión. Asegúrese de que *todos* los conjuntos de datos de entrada están ordenados.

Especificar orden de clasificación existente. Determina los campos que ya están ordenados. Puede añadir campos a la lista mediante el cuadro de diálogo Seleccionar campos. Sólo puede realizar la selección de los campos clave que se utilizan para la fusión (especificados en la pestaña Fundir). En la columna *Orden*, especifique si cada campo se ordena en orden ascendente o descendente. Si especifica varios campos, asegúrese de que los enumera en el orden de clasificación correcto. Utilice las flechas de la derecha de la lista para organizar los campos en el orden correcto. Si comete un error al especificar el orden de clasificación existente, aparecerá un error cuando ejecute la ruta que muestra el número de registro donde la clasificación no coincide con la que especificó.

Dependiendo de la sensibilidad de caracteres del método de ordenación utilizado por la base de datos, puede que la optimización no funcione correctamente cuando la base de datos ordene una o más entradas. Por ejemplo, si tiene dos entradas de las que una distingue entre mayúsculas y minúsculas y la otra no, los resultados de la ordenación podrían ser diferentes. La optimización de fusión hace que los registros se procesen mediante su ordenación. Como resultado, si las entradas se ordena utilizando métodos de ordenación diferentes, el nodo Fundir indica un error y muestra el número de registro cuando la ordenación no es coherente. Cuando todas las entradas son de un origen o se ordenan mediante ordenaciones inclusivas mutuamente, los registros se pueden combinar correctamente.

Note: La velocidad de fusión puede mejorarse al activar el procesamiento paralelo.

Nodo Añadir

Los nodos Añadir se pueden usar para concatenar conjuntos de registros. A diferencia con los nodos Fundir, que une registros de diferentes orígenes, los nodos Añadir leen y pasan a la parte de abajo todos los registros de un único origen hasta que no quede ninguno. A continuación, los registros procedentes del siguiente origen se leen usando la misma estructura de datos (número de registros, número de campos, etc.) que la entrada primera (o primaria). Cuando el origen primario contiene más campos que otro registro de entrada, se usa la cadena de valor nulo del sistema (\$null\$) para los valores incompletos.

Los nodos Añadir son útiles para combinar conjuntos de datos con estructuras similares pero datos diferentes. Por ejemplo, podría tener datos de transacción almacenados en diferentes archivos para diferentes períodos (un archivo de datos de venta para el mes de marzo y otro para el mes de abril, por ejemplo). Suponiendo que tengan la misma estructura (los mismos campos en el mismo orden), el nodo Añadir los une en un archivo de gran tamaño que se puede analizar.

Note: Con el fin de añadir los archivos, los niveles de medición de campo deben coincidir. Por ejemplo, un campo *Nominal* no puede añadirse a un campo cuyo nivel de medición sea *Continuo*.

Opciones de configuración de Añadir

Coincidir campos por. Seleccione un método que utilizar para asignar los campos que añadir.

- **Posición.** Seleccione esta opción para añadir los conjuntos de datos basándose en la posición que tienen los campos en el origen de datos principal. Cuando use este método, los datos deben estar ordenados para garantizar una correcta adición.
- **Nombre.** Seleccione esta opción para añadir los conjuntos de datos basándose en el nombre que tienen los campos en los conjuntos de datos de entrada. Seleccione también **Coincidir mayúsculas y minúsculas** para activar la sensibilidad de caracteres en la coincidencia de nombres de campos.

Campo de salida. Enumera los nodos de origen conectados al nodo Añadir. El primero de la lista es el origen de entrada primario. Los campos de la lista se pueden ordenar pulsando en la cabecera de columna. Esta organización no ordena realmente los campos del conjunto de datos.

Incluir campos de. Seleccione **Sólo conjunto de datos principal** para producir campos de salida basándose en los campos del conjunto de datos principal. El conjunto de datos principal es la primera entrada, especificada en la pestaña Entradas. Seleccione **Todos los conjuntos de datos** para producir

campos de salida para todos los campos en todos los conjuntos de datos, independientemente de si existe un campo coincidente en todos los conjuntos de datos de entrada.

Etiquetar registros incluyendo el conjunto de datos de origen en el campo. Seleccione esta opción para añadir un campo de información adicional al archivo de resultados cuyos valores indican el conjunto de datos de origen de cada registro. Especifique un nombre en el cuadro de texto. El nombre predeterminado es *Entrada*.

nodo Distinguir

Los registros duplicados en una base de datos deben eliminarse antes de comenzar con la minería de datos. Por ejemplo, en una base de datos de marketing, las personas pueden aparecer muchas veces con diferentes direcciones o diferente información de empresas. El nodo Distinguir puede utilizarse para buscar o eliminar registros en los datos o para crear un registro compuesto único a partir de un grupo de registros duplicados.

Para utilizar el nodo Distinguir, primero debe definir un conjunto de campos clave que determinan cuándo se considera que dos registros están duplicados.

Si no selecciona todos los campos como campos clave, dos registros duplicados quizá no sean realmente idénticos porque pueden diferir en los valores de los campos restantes. En este caso, también puede definir un orden de clasificación que se aplique dentro de cada grupo de registros duplicados. Este orden de clasificación le proporciona un control detallado sobre qué registro se trata como el primero dentro de un grupo. De lo contrario, todos los duplicados se consideran intercambiables y puede seleccionarse cualquier registro. La orden de entrada de los registros no se tiene en cuenta, por lo tanto, no es útil utilizar un nodo Ordenar en sentido ascendente (consulte "Ordenación de registros dentro del nodo Distinguir" a continuación).

Moda. Especifica si se crea un registro compuesto o si se incluye o excluye (descarta) el primer registro.

- **Crear un registro compuesto para cada grupo.** Proporciona una forma de agregar campos no numéricos. Al seleccionarse esta opción se habilita la pestaña Compuesto, en la que se especifica cómo crear los registros compuestos. Para obtener más información, consulte "Valores compuestos diferenciados" en la página 95.
- **Incluir sólo el primer registro en cada grupo.** Selecciona el primer registro de cada grupo de registros duplicados y descarta el resto. El *primer* registro se determina por el orden de clasificación definido a continuación, y no por el orden de entrada de los registros.
- **Descartar sólo el primer registro en cada grupo.** Descarta el primer registro de cada grupo de registros duplicados y selecciona el resto. El *primer* registro se determina por el orden de clasificación definido a continuación, y no por el orden de entrada de los registros. Esta opción resulta útil para *buscar* duplicados en los datos con el fin de examinarlos posteriormente en la ruta.

Campos clave para el agrupamiento. Enumera el campo o campos utilizados para determinar si los registros son idénticos. Tiene la posibilidad de:

- Para añadir campos a esta lista, utilice el botón selector de campos de la parte derecha.
- Elimine campos de la lista utilizando el botón X rojo (eliminar).

En grupos, ordenar registros por. Enumera los campos utilizados para especificar el modo en que se ordenan los registros dentro de cada grupo de duplicados, y si se ordenan de forma ascendente o descendente. Tiene la posibilidad de:

- Para añadir campos a esta lista, utilice el botón selector de campos de la parte derecha.
- Elimine campos de la lista utilizando el botón X rojo (eliminar).
- Mueva campos mediante los botones de arriba o abajo si está ordenando por más de un campo.

Debe especificar un orden de clasificación si ha optado por incluir o excluir el primer registro de cada grupo, y le importa qué registro se trata como el primero.

También puede que desee especificar un orden de clasificación si ha optado por crear un registro compuesto, para determinadas opciones en la pestaña Compuesto. Para obtener más información, consulte “Valores compuestos diferenciados” en la página 95.

Orden de clasificación predeterminado. Especifique si, de forma predeterminada, los registros se ordenan de forma **Ascendente** o **Descendente** de los valores de clave de ordenación.

Ordenación de registros dentro del nodo Distinguir

Si el orden de los registros dentro de un grupo de duplicados es importante para usted, entonces debe especificar el orden utilizando la opción **En los grupos, clasificar registros por** en el nodo Distinguir. No se base en un nodo Ordenar en sentido ascendente. Recuerde que el orden de entrada de los registros no se tiene en cuenta -- sólo el orden especificado dentro del nodo.

Si no especifica ningún campo de clasificación (o especifica campos de clasificación insuficientes), los registros dentro de cada grupo de duplicados estarán desordenados (u ordenados de forma incompleta) y los resultados pueden ser imprevisibles.

Por ejemplo, supongamos que tiene un gran conjunto de registros que pertenecen a varias máquinas. El registro contiene datos como los siguientes:

Tabla 16. Datos de registro de máquina

Marca de tiempo	Máquina	Temperatura
17:00:22	Máquina A	31
13:11:30	Máquina B	26
16:49:59	Máquina A	30
18:06:30	Máquina X	32
16:17:33	Máquina A	29
19:59:04	Máquina C	35
19:20:55	Máquina Y	34
15:36:14	Máquina X	28
12:30:41	Máquina Y	25
14:45:49	Máquina C	27
19:42:00	Máquina B	34
20:51:09	Máquina Y	36
19:07:23	Máquina X	33

Para reducir el número de registros al último registro para cada máquina, utilice Máquina como el campo de clave y utilice Marca de tiempo como el campo de ordenación (en orden descendente). El orden de entrada no afecta al resultado porque la selección de clasificación específica cuáles de las diversas filas de una Máquina concreta se van a devolver y el resultado de datos final sería el siguiente:

Tabla 17. Datos de registro de máquina ordenados

Marca de tiempo	Máquina	Temperatura
17:00:22	Máquina A	31
19:42:00	Máquina B	34
19:59:04	Máquina C	35

Tabla 17. Datos de registro de máquina ordenados (continuación)

Marca de tiempo	Máquina	Temperatura
19:07:23	Máquina X	33
20:51:09	Máquina Y	36

Configuración de optimización de Distinguir

Si los datos en los que trabaja sólo tienen un pequeño número de registros o ya se han ordenado, puede optimizar la forma en la que se gestionan para que IBM SPSS Modeler procese los datos de forma más eficiente.

Note: Si selecciona **El conjunto de datos de entrada tiene un pequeño número de claves distintas** o la generación SQL para el nodo, se puede devolver cualquier fila dentro del valor de clave distinta; para controlar la fila que se devuelve, tiene que especificar el orden de clasificación mediante **En grupos, ordenar registros por** de la pestaña Configuración. Las opciones de optimización no afectan a los resultados ofrecidos por el nodo Distinguir siempre que haya especificado un orden de clasificación en la pestaña Configuración.

El conjunto de datos de entrada tiene un pequeño número de claves distintas. Seleccione esta opción si sólo tiene un pequeño número de registros, un pequeño número de valores exclusivos del campo(s) clave o ambos. Con ello puede mejorar el rendimiento.

El conjunto de datos de entrada ya está ordenado por los campos de agrupación y los campos de clasificación en la pestaña Configuración. Seleccione esta opción únicamente si los datos ya están ordenados según los campos enumerados en **En grupos, ordenar registros por** en la pestaña Configuración, y si el orden de clasificación ascendente o descendente de los datos es el mismo. Con ello puede mejorar el rendimiento.

Desactivar generación de SQL. Seleccione esta opción para desactivar la generación de SQL para el nodo.

Valores compuestos diferenciados

Si los datos sobre los que se trabaja tienen varios registros (por ejemplo, para la misma persona), podrá optimizarse la forma en que se manejan los datos creando un único registro compuesto, o agregado, para procesarlo. Si se tiene instalado IBM SPSS Modeler Entity Analytics, también podrá utilizarse esto para combinar, o aplanar, registros duplicados que sean salida de SPSS Entity Analytics.

Nota: Esta pestaña solo está disponible cuando se selecciona **Crear un registro compuesto para cada grupo** en la pestaña Valores.

Por ejemplo, suponga que SPSS Entity Analytics marca tres registros como pertenecientes a la misma entidad, tal y como se muestra en la tabla siguiente.

Tabla 18. Ejemplo de varios registros de la misma entidad.

\$EA-ID	Nombre	Edad	Banco	Nivel académico	Deuda total
0003	Bob Jones	27	K	Escuela	27000
0003	Robert Jones	35	N	relación	42000
0003	Robbie Jones	27	D	Doctorado	7000

Lo que se desea es agregar estos tres registros en un único registro que luego pueda utilizarse en un punto posterior de la ruta. Podría utilizarse el nodo Agregar para sumar la deuda total y calcular una

edad media; sin embargo, no se pueden promediar detalles tales como nombres, bancos, etc. Si se especifican qué detalles deben utilizarse para crear un registro compuesto, podrá derivarse un único registro.

A partir de la tabla podría crearse un registro compuesto seleccionando los detalles siguientes.

- Para el **Nombre** utilice el registro Primero.
- Para la **Edad** quédese con el Más alto.
- Para el **Banco** concatene todos los valores sin delimitador.
- Para la **Titulación de más nivel** quédese con el Primero encontrado en la lista (Doctorado Licenciatura Escuela)
- Para **Deuda** quédese con el Total.

Al combinar (o agregar) estos detalles se obtiene un único registro compuesto que contiene los detalles siguientes.

- Nombre: Bob Jones.
- Edad: 35.
- Banco: KND.
- Titulación de más nivel: doctorado.
- Deuda: 76000.

Esto ofrece el mejor perfil de Bob Jones, una persona con formación de doctorado, con al menos 35 años de edad, con al menos tres cuentas bancarias conocidas y una elevada deuda total.

Definición de opciones en la pestaña Compuesto

Campo. Esta columna muestra todos los campos, excepto los campos clave del modelo de datos, en su orden de clasificación natural; si el nodo no está conectado, no se mostrará ningún campo. Para ordenar las filas alfabéticamente por nombre de campo, pulse en la cabecera de la columna. Puede seleccionarse más de una fila mediante Mayús y pulsación o Ctrl y pulsación. Además, si se pulsa con el botón derecho en un campo, se visualizará un menú desde el que se puede optar por seleccionar todas las filas, ordenar las filas por orden ascendente o descendente del nombre o valor del campo, seleccionar los campos por medición o tipo de almacenamiento, o seleccionar un valor para añadir automáticamente la misma entrada **Rellenar con valores basados en** a cada fila seleccionada.

Rellenar con valores basados en. Seleccione el tipo del valor que se utilizará en el registro compuesto para el **Campo**. Las opciones disponibles dependen del tipo de campo.

- En el caso de los campos de rango numérico puede elegirse entre:
 - Primer registro del grupo.
 - Último registro del grupo.
 - Total.
 - Media.
 - Mínimo.
 - Máximo.
 - Personalizado.
- En el caso de campos de fecha u hora puede elegirse entre:
 - Primer registro del grupo.
 - Último registro del grupo.
 - Más antiguo
 - Más reciente
 - Personalizado.

- En el caso de campos de tipo cadena o sin tipo puede elegirse entre:
 - Primer registro del grupo.
 - Último registro del grupo.
 - Primer alfanumérico.
 - Último alfanumérico.
 - Personalizado.

En cada caso podrá utilizarse la opción **Personalizado** para tener un mayor control sobre qué valor se utiliza para rellenar el registro compuesto. Para obtener más información, consulte “Compuesto distinguido - Pestaña Personalizado”.

Incluir recuento de registros en campo. Seleccione esta opción para incluir un campo adicional en cada registro de salida, llamado Record_Count de forma predeterminada. Este campo indica la cantidad de registros de entrada añadidos para conformar cada registro añadido. Para crear un nombre personalizado para este campo, escríbalo en el campo de edición.

Compuesto distinguido - Pestaña Personalizado

El cuadro de diálogo Relleno personalizado ofrece un mayor control sobre el valor utilizado para completar el nuevo registro compuesto. Tenga en cuenta que los datos deberán instanciarse antes de utilizar esta opción si solo se está personalizando una única fila Campo en la pestaña Compuesto.

Nota: Este cuadro de diálogo solo está disponible cuando se selecciona el valor Personalizado en la columna **Rellenar con valores basados en** en la pestaña Compuesto.

Según el tipo de campo, podrá elegirse una de las opciones siguientes.

- **Seleccionar por frecuencia.** Elija un valor basado en la frecuencia con la que aparece en el registro de datos.

Nota: No está disponible en campos con un tipo Continuo, Sin tipo o Fecha/hora.

- **Utilizar.** Seleccione Más frecuente o Menos frecuente.
- **Empates.** Si hay dos o más registros con la misma frecuencia de aparición, especifique cómo seleccionar el registro necesario. Puede elegirse una de cuatro opciones: Utilizar primero, Utilizar último, Utilizar más bajo o Utilizar más alto.
- **Incluye valor (T/F).** Seleccione esta opción para convertir un campo a un distintivo que identifique si alguno de los registros de un grupo tiene especificado un valor. Luego podrá seleccionar el **Valor** de la lista de valores del campo seleccionado.

Nota: No está disponible si se selecciona más de una fila Campo en la pestaña Compuesto.

- **Primera coincidencia de la lista.** Seleccione esta opción para priorizar qué valor dar al registro compuesto. Luego podrá seleccionar uno de los **Elementos** de la lista de valores del campo seleccionado.

Nota: No está disponible si se selecciona más de una fila Campo en la pestaña Compuesto.

- **Concatenar valores.** Seleccione esta opción para retener todos los valores de un grupo concatenándolos en una cadena. Debe especificarse el delimitador que se va a utilizar entre cada valor.

Nota: Esta es la única opción disponible si se seleccionan una o más filas Campo con tipo Continuo, Sin tipo o Fecha/Hora.

- **Utilizar delimitador.** Puede optarse por utilizar un **Espacio** o una **Coma** como valor delimitador en la cadena concatenada. De forma alternativa, en el campo **Otros**, podrá especificar su propio carácter delimitador.

Nota: Solo está disponible si se selecciona la opción **Concatenar valores**.

Nodo de Generación de análisis TS

El nodo Generación de análisis de series temporales (TS) puede utilizarse para construir y puntuar modelos de series temporales en un único paso, y puede utilizarse con o sin un nodo Intervalos de tiempo en un punto anterior en la ruta. Por cada campo de destino se construye un modelo de series temporales independiente; sin embargo, no se añaden nuggets de modelo a la paleta de modelos generados y la información del modelo no puede examinarse.

No es necesario utilizar un nodo Intervalos de tiempo con el nodo Generación de análisis TS siempre y cuando se cumplan las condiciones siguientes:

- Los datos están espaciados uniformemente.
- Los datos tienen etiquetas de serie temporal.
- No es necesario añadir registros adicionales.
- Los datos tienen una *periodicidad* estándar como, por ejemplo, cinco días a la semana u ocho horas al día.

Los métodos para modelar datos de series temporales precisan de un intervalo uniforme entre cada medición. Los valores perdidos deben indicarse mediante filas vacías. Si sus datos aún no cumplen este requisito, un nodo Intervalos de tiempo situado en un punto anterior de la ruta podrá transformar los valores según sea necesario. Consulte el tema “Nodo de intervalos de tiempo” en la página 170 para obtener más información.

Deben tenerse en cuenta estas otras cuestiones en relación con el nodo Serie temporal:

- Los campos deben ser numéricos.
- Los campos de fecha no se pueden utilizar como entrada.
- Las particiones se ignoran.

El nodo Generar análisis TS estima modelos de suavizado exponencial, modelos autorregresivos integrados de media móvil (ARIMA) univariados y modelos ARIMA (o de función de transferencia) multivariados para series temporales y genera previsiones a partir de los datos de series temporales. También está disponible un modelizador experto, que intenta identificar y estimar de forma automática el modelo de suavizado ARIMA o exponencial que mejor se ajuste a uno o más campos de destino. Para obtener más información sobre los modelos de series en general, consulte el tema *Modelos de series temporales* en el documento *Nodos de modelado de IBM SPSS Modeler 17*.

La utilización de un nodo Generación de análisis TS está soportada en un entorno de despliegue de generación de análisis a través de IBM SPSS Modeler Solution Publisher, utilizando el servicio de puntuación de IBM SPSS Collaboration and Deployment Services o IBM InfoSphere Warehouse.

Opciones de campo de Generación de análisis TS

En la pestaña Campos se especifican los campos que se van a utilizar para generar el modelo. Para generar un modelo, primero se deben especificar los campos que se desea usar como objetivos y como entradas. El nodo Generación de análisis TS utiliza habitualmente información procedente de un nodo Tipo anterior en la ruta. Si utiliza un nodo Tipo para seleccionar campos de entrada y objetivo, no es necesario cambiar nada en esta pestaña.

Utilizar configuración del nodo Tipo. Esta opción permite indicar al nodo que use la información de campo de un nodo Tipo situado en un punto anterior de la ruta. Este es el método predeterminado.

Utilizar configuración personalizada. Esta opción permite indicar al nodo que use la información de campo especificada aquí en lugar de la proporcionada en nodos Tipo situados en cualquier punto anterior de la ruta. Después de seleccionar esta opción, especifique los campos siguientes. Tenga en cuenta que los campos almacenados como fechas no se aceptan como campos objetivo o de entrada.

- **Destino.** Seleccione uno o varios campos objetivo. Se trata de una acción similar a establecer un rol de un campo a *Objetivo* en un nodo Tipo. Los campos objetivos de un modelo de series temporales debe tener un nivel de medición de *Continuo*. Se crea un modelo distinto para cada campo objetivo. Un campo objetivo tiene en cuenta todos los campos *Entrada* especificados (excepto dicho campo) como posibles entradas. Por lo tanto, el mismo campo puede incluirse en ambas listas; ese campo se utilizará como posible entrada para todos los modelos, excepto en el que actúa como objetivo.
- **Entradas.** Seleccione los campos de entrada. Se trata de una acción similar a establecer un rol de un campo a *Entrada* en un nodo Tipo. Los campos de entrada de un modelo de serie temporal deben ser numéricos.

Opciones del modelo de Generación de análisis TS

Método. Puede elegir Modelizador experto, Suavizado exponencial o ARIMA. Seleccione **Criterios** para especificar opciones para el método seleccionado.

- **Modelizador experto.** Seleccione esta opción para utilizar el modelizador experto, que busca automáticamente el modelo que mejor se ajusta a cada serie.
- **Suavizado exponencial.** Utilice esta opción para especificar un modelo de suavizado exponencial personalizado.
- **ARIMA.** Utilice esta opción para especificar un modelo ARIMA personalizado.

Derivar campos para los límites de confianza superior e inferior. Esta opción se selecciona para generar campos de intervalos de confianza para las predicciones del modelo y las autocorrelaciones residuales. Cuando está seleccionada, el control **Amplitud de límite de confianza (%)** está habilitado.

Amplitud de límite de confianza (%). Especifica los intervalos de confianza que se calculan. Puede especificarse cualquier valor positivo inferior a 100. De forma predeterminada, se utiliza un intervalo de confianza del 95%.

Opciones de intervalos de tiempo

En esta sección del cuadro de diálogo se puede elegir entre utilizar especificaciones para las estimaciones y previsiones de un nodo Intervalos de tiempo situado en un punto anterior de la ruta o especificar estos valores en el nodo Generación de análisis TS.

Utilizar valores del nodo Intervalos de tiempo. Esta opción se selecciona para utilizar información relativa a las especificaciones de estimaciones y predicciones efectuadas en un nodo Intervalos de tiempo situado en un punto anterior de la ruta.

En la primera línea se indica si hay algún registro que se excluya del modelo o se utilice como caso reservado.

En la segunda línea aparece información sobre todos los períodos de previsión especificados en el nodo Intervalos de tiempo.

Si en la primera línea aparece **No se ha definido ningún intervalo de tiempo**, esto significa que no hay ningún nodo Intervalos de tiempo conectado. Esta situación provocará un error cuando se intente ejecutar la ruta. Si se selecciona **Utilizar valores del nodo Intervalos de tiempo** deberá incluirse un nodo Intervalos de tiempo en un punto de la ruta anterior al nodo Generación de análisis TS.

Especificar valores. Esta opción se selecciona para especificar los valores de cálculos y previsiones si no hay ningún nodo Intervalos de tiempo en un punto de la ruta anterior al nodo Generación de análisis TS.

Estimar hasta. Seleccione **Última observación válida** para estimar el modelo utilizando todos los registros hasta la última observación válida. Los registros utilizados en la previsión empiezan a partir de la primera observación no válida. Las observaciones *Válidas* son aquellas en las que los campos de destino

tienen valores no nulos. Seleccione **Incluir anteriores al último registro** para reservar varios de los registros más recientes para la predicción. Puede que los registros más recientes tengan valores para los campos de destino, aunque no es necesario. Si **Incluir anteriores al último registro** está seleccionada, se habilita el control **Desplazamiento** donde puede especificarse el número de registros a utilizar en la predicción. El número predeterminado de registros es cuatro, y el máximo es de 99.

Nota: Los registros que se van a utilizar en la previsión deberán tener valores para todos los campos de entrada.

Intervalo de tiempo. Especifica el intervalo de tiempo utilizado en estimaciones y predicciones. Las siguientes opciones están disponibles: años, trimestres, meses, semanas, días, horas, minutos y segundos.

Criterios del modelizador experto de series temporales

Tipo de modelo. Las siguientes opciones están disponibles:

- **Todos los modelos.** El modelizador experto tiene en cuenta tanto los modelos ARIMA como los modelos de suavizado exponencial.
- **Sólo modelos de suavizado exponencial.** El modelizador experto sólo tiene en cuenta los modelos de suavizado exponencial.
- **Sólo modelos ARIMA.** El modelizador experto sólo tiene en cuenta los modelos ARIMA.

El modelizador experto considera modelos estacionales. Esta opción sólo está activada si se ha definido una periodicidad para el conjunto de datos activo. Si esta opción está seleccionada, el modelizador experto tiene en cuenta los modelos tanto estacionales como no estacionales. Si esta opción no está activada, el modelizador experto sólo tiene en cuenta los modelos no estacionales.

Eventos e intervenciones. Permite designar determinados campos de entrada como campos de eventos o intervención. De esta forma, se identifica el campo que contiene datos de series temporales afectados por eventos (situaciones recurrentes predecibles, como promociones de ventas) o intervenciones (incidentes puntuales, como apagones o huelgas). El modelizador experto no tiene en cuenta las funciones de transferencia arbitrarias para entradas identificadas como campos de evento o intervención.

Los campos de entrada deben tener un nivel de medición de *Marca*, *Nominal* u *Ordinal* y deben ser numéricos (por ejemplo, 1/0, no True/False, para un campo marca), antes de que aparezcan en esta lista.

Valores atípicos

Detectar valores atípicos de forma automática. De forma predeterminada, no se realiza la detección automática de valores atípicos. Seleccione esta opción para realizar una detección automática de valores atípicos y, a continuación, seleccione los tipos de valores atípicos que desee. Consulte el tema "Tratamiento de valores atípicos" en la página 103 para obtener más información.

Criterios de suavizado exponencial de series temporales

Tipo de modelo. Los modelos de suavizado exponencial se clasifican como estacionales o no estacionales¹. Los modelos estacionales sólo están disponibles si la periodicidad es estacional. Las periodicidades estacionales son: períodos cíclicos, años, trimestres, meses, días por semana, horas al día, minutos al día y segundos por día.

- **Simple.** Este modelo es adecuado para las series sin tendencia ni estacionalidad. Su único parámetro de suavizado relevante es el nivel. El suavizado exponencial simple es muy similar a un ARIMA con cero órdenes de autorregresión, un orden de diferenciación, un orden de media móvil y ninguna constante.

1. Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1-28.

- **Tendencia lineal de Holt.** Este modelo es adecuado para las series con una tendencia lineal y sin estacionalidad. Sus parámetros de suavizado relevantes son el nivel y la tendencia y, en este modelo, no están restringidos por sus valores respectivos. El modelo de Holt es más general que el de Brown, pero puede tardar más en calcular estimaciones para series grandes. El suavizado exponencial de Holt es muy similar a un ARIMA con cero órdenes de autorregresión, dos órdenes de diferenciación y dos órdenes de media móvil.
- **Tendencia lineal de Brown.** Este modelo es adecuado para las series con una tendencia lineal y sin estacionalidad. Sus parámetros de suavizado relevantes son el nivel y la tendencia, pero, en este modelo, se supone que son iguales. Por ello, el modelo de Brown es un caso especial del modelo de Holt. El suavizado exponencial de Brown es muy similar a un ARIMA con cero órdenes de autorregresión, dos órdenes de diferenciación y dos órdenes de media móvil, siendo el coeficiente del segundo orden de la media móvil igual a la mitad del coeficiente del primer orden al cuadrado.
- **Tendencia amortiguada.** Este modelo es adecuado para las series con una tendencia lineal que va desapareciendo y sin estacionalidad. Sus parámetros de suavizado relevantes son el nivel, la tendencia y la tendencia de amortiguación. El suavizado exponencial amortiguado es muy similar a un ARIMA con cero órdenes de autorregresión, un orden de diferenciación y dos órdenes de media móvil.
- **Estacional simple.** Este modelo es adecuado para las series sin una tendencia y un efecto estacional constante a lo largo del tiempo. Sus parámetros de suavizado relevantes son el nivel y la estacionalidad. El suavizado exponencial estacional es muy similar a un ARIMA con cero órdenes de autorregresión, un orden de diferenciación, un orden de diferenciación estacional y los órdenes 1, p y $p+1$ de media móvil, donde p es el número de períodos contenidos en un intervalo estacional. En el caso de los datos mensuales, $p = 12$.
- **Aditivo de Winters.** Este modelo es adecuado para las series con una tendencia lineal y un efecto estacional constante a lo largo del tiempo. Sus parámetros de suavizado relevantes son el nivel, la tendencia y la estacionalidad. El suavizado exponencial aditivo de Winters es muy similar a un ARIMA con cero órdenes de autorregresión, un orden de diferenciación, un orden de diferenciación estacional y los órdenes $p+1$ de media móvil, donde p es el número de períodos contenidos en un intervalo estacional. En el caso de los datos mensuales, $p = 12$.
- **Multiplicativo de Winters.** Este modelo es adecuado para series en las que haya una tendencia lineal y con un efecto estacional que cambie en función de la magnitud de las series. Sus parámetros de suavizado relevantes son el nivel, la tendencia y la estacionalidad. El modelo de suavizado exponencial multiplicativo de Winters no es similar a ningún modelo ARIMA.

Transformación destino. Puede especificar una transformación para que se lleve a cabo en cada variable dependiente antes de su modelado. Consulte el tema *Transformaciones de series* en el documento *IBM SPSS Modeler 17 Modelado de nodos* para obtener más información.

- **Ninguno.** No se lleva a cabo ninguna transformación.
- **Raíz cuadrada.** Se realiza una transformación de raíz cuadrada.
- **Logaritmo natural.** Se realiza una transformación logarítmica natural.

Criterios ARIMA de series temporales

El nodo Serie temporal permite crear modelos ARIMA estacionales o no estacionales personalizados -también conocidos como modelos Box-Jenkins- con o sin un conjunto fijo de variables de entrada (predictoras) ². Puede definir funciones de transferencia para algunas o todas las variables de entrada y especificar la detección automática de valores atípicos o especificar un conjunto explícito de valores atípicos.

Todas las variables de entrada especificadas se incluyen en el modelo de manera explícita, a diferencia de lo que ocurre al utilizar el modelizador experto, donde las variables de entrada sólo se incluyen si tienen una relación estadísticamente significativa con la variable objetivo.

2. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

Modelo

La pestaña Modelo le permite especificar la estructura de un modelo ARIMA personalizado.

Órdenes ARIMA. Escriba valores para los distintos componentes ARIMA del modelo en las casillas correspondientes de la cuadrícula Estructura. Todos los valores deben ser enteros no negativos. Para los componentes autorregresivos y de media móvil, el valor representa el orden máximo. Todos los órdenes inferiores positivos se incluyen en el modelo. Por ejemplo, si especifica 2, el modelo incluye los órdenes 2 y 1. Las casillas de la columna Estacional sólo se activan si se ha definido una periodicidad para el conjunto de datos activo.

- **Autorregresivo (p).** Es el número de órdenes autorregresivos del modelo. Los órdenes autorregresivos especifican los valores previos de la serie utilizados para predecir los valores actuales. Por ejemplo, un orden autorregresivo igual a 2 especifica que se van a utilizar los valores de la serie correspondientes a dos períodos de tiempo del pasado para predecir el valor actual.
- **Diferencia (d).** Especifica el orden de diferenciación aplicado a la serie antes de estimar los modelos. La diferenciación es necesaria si hay tendencias (las series con tendencias suelen ser no estacionarias y el modelado de ARIMA asume la estacionariedad) y se utiliza para eliminar su efecto. El orden de diferenciación se corresponde con el grado de la tendencia de la serie (la diferenciación de primer orden representa las tendencias lineales, la diferenciación de segundo orden representa las tendencias cuadráticas, etc.).
- **Media móvil (q).** Es el número de órdenes de media móvil presentes en el modelo. Los órdenes de media móvil especifican el modo en que se utilizan las desviaciones de la media de la serie para los valores previos con el fin de predecir los valores actuales. Por ejemplo, los órdenes de media móvil de 1 y 2 especifican que las desviaciones respecto al valor medio de la serie de cada uno de los dos últimos períodos de tiempo se tienen en cuenta al predecir los valores actuales de la serie.

Órdenes estacionales. Los componentes estacionales autorregresivos, de media móvil y de diferenciación tienen los mismos roles que los componentes no estacionales correspondientes. No obstante, en el caso de los órdenes estacionales, los valores de la serie actual se ven afectados por los valores de la serie anterior separados por uno o más períodos estacionales. Por ejemplo, para los datos mensuales (período estacional de 12), un orden estacional de 1 significa que el valor de la serie actual se ve afectado por el valor de la serie 12 períodos antes del actual. Un orden estacional de 1 para los datos mensuales equivale a la especificación de un orden no estacional de 12.

Transformación destino. Puede especificar una transformación para que se lleve a cabo en cada variable objetivo antes de su modelado. Consulte el tema *Transformación de series* en el documento *IBM SPSS Modeler 17 Modelado de nodos* para obtener más información.

- **Ninguno.** No se lleva a cabo ninguna transformación.
- **Raíz cuadrada.** Se realiza una transformación de raíz cuadrada.
- **Logaritmo natural.** Se realiza una transformación logarítmica natural.

Incluir constante en modelo. La inclusión de una constante es estándar a menos que esté seguro de que el valor de la media global de la serie es 0. Se recomienda la exclusión de la constante si se aplica la diferenciación.

Funciones de transferencia

La pestaña Funciones de transferencia le permite definir funciones de transferencia para algunos o todos los campos de entrada. Las funciones de transferencia le permiten especificar el modo en que se utilizan los valores anteriores de estos campos para predecir valores futuros de la serie objetivo.

La pestaña solo se muestra si los campos de entrada (con el rol establecido a *Entrada*) están especificados en el nodo Tipo o en la pestaña Campos del nodo Generación de análisis TS (seleccione **Utilizar valores—entradas personalizados**).

En la lista superior aparecen todos los campos de entrada. El resto de la información que aparece en este cuadro de diálogo es específica del campo de entrada seleccionado en la lista.

Órdenes de la función de transferencia. Escriba valores para los distintos componentes de la función de transferencia en las casillas correspondientes de la cuadrícula Estructura. Todos los valores deben ser enteros no negativos. Para los componentes de numerador y denominador, el valor representa el orden máximo. Todos los órdenes inferiores positivos se incluyen en el modelo. Además, el orden 0 siempre se incluye para los componentes de numerador. Por ejemplo, si especifica 2 para el numerador, el modelo incluye los órdenes 2, 1 y 0. Si especifica 3 para el denominador, el modelo incluye los órdenes 3, 2 y 1. Las casillas de la columna Estacional sólo se activan si se ha definido una periodicidad para el conjunto de datos activo.

Numerador. El orden de numerador de la función de transferencia especifica los valores previos de la serie independiente (predictora) seleccionada que se utilizan para predecir los valores actuales de la serie dependiente. Por ejemplo, un orden de numerador de 1 especifica que se utiliza el valor de una serie independiente de un período anterior (además del valor actual de la serie independiente) para predecir el valor actual de cada serie dependiente.

Denominador. El orden de denominador de la función de transferencia especifica cómo se utilizan las desviaciones respecto a la media de la serie para los valores previos de la serie independiente (predictora) seleccionada para predecir los valores actuales de la serie dependiente. Por ejemplo, un orden de denominador de 1 especifica que las desviaciones respecto al valor medio de una serie independiente para un período de tiempo anterior se tienen en cuenta al predecir el valor actual de cada serie dependiente.

Diferencia. Especifica el orden de diferenciación aplicado a la serie independiente (predictora) seleccionada antes de estimar los modelos. La diferenciación es necesaria si hay tendencias y se utiliza para eliminar su efecto.

Órdenes estacionales. Los componentes estacionales de numerador, denominador y diferenciación tienen los mismos roles que los componentes no estacionales correspondientes. No obstante, en el caso de los órdenes estacionales, los valores de la serie actual se ven afectados por los valores de la serie anterior separados por uno o más períodos estacionales. Por ejemplo, para los datos mensuales (período estacional de 12), un orden estacional de 1 significa que el valor de la serie actual se ve afectado por el valor de la serie 12 períodos antes del actual. Un orden estacional de 1 para los datos mensuales equivale a la especificación de un orden no estacional de 12.

Retardo. Establecer un retardo provoca que la influencia del campo de entrada se retrase según el número de intervalos especificados. Por ejemplo, si el retardo se establece en 5, el valor de la variable de entrada en el tiempo t no afecta a las previsiones hasta que han transcurrido cinco períodos ($t + 5$).

Transformación. La especificación de una función de transferencia para un conjunto de variables independientes también incluye una transformación opcional que se puede aplicar a dichas variables.

- **Ninguno.** No se lleva a cabo ninguna transformación.
- **Raíz cuadrada.** Se realiza una transformación de raíz cuadrada.
- **Logaritmo natural.** Se realiza una transformación logarítmica natural.

Tratamiento de valores atípicos

La pestaña Valores atípicos proporciona una serie de opciones para el manejo de valores atípicos en los datos de ³.

3. Pena, D., G. C. Tiao y R. S. Tsay, eds. 2001. *A course in time series analysis*. New York: John Wiley and Sons.

No detectar valores atípicos ni modelarlos. De forma predeterminada, los valores atípicos no se detectan ni modelan. Seleccione esta opción para desactivar la detección o el modelado de valores atípicos.

Detectar valores atípicos de forma automática. Seleccione esta opción para realizar una detección automática de valores atípicos y seleccione uno o más de los tipos de valores atípicos que se muestran.

Tipo de los valores atípicos a detectar. Seleccione los tipos de valores atípicos que desea detectar. Los tipos soportados son:

- Aditivo (predeterminado).
- Cambio de nivel (predeterminado).
- Innovador.
- Transitorio.
- Aditivo estacional.
- Tendencia local.
- Parche aditivo.

Consulte el tema *Valores atípicos* en el documento *IBM SPSS Modeler 17 Modelado de nodos* para obtener más información.

Opciones de despliegue de Generación de análisis TS

Si se utiliza un nodo Generación de análisis TS en un entorno de despliegue a través de IBM SPSS Modeler Solution Publisher, de forma predeterminada, cada vez que se utiliza un modelo de serie temporal para puntuar registros, se reconstruye el modelo y se realizan nuevas predicciones. Este comportamiento predeterminado puede cambiarse utilizando las opciones de la pestaña Despliegue para especificar cuándo se regeneran los modelos de series temporales.

Nota: Los valores de la pestaña Despliegue se ignoran en el uso normal del nodo Generación de análisis TS en IBM SPSS Modeler.

Forzar regeneración completa cuando sea necesario (puede cambiar el tipo de modelo). Seleccione esta opción si desea que los modelos de series temporales se regeneren solo si hay un número especificado de registros nuevos. La regeneración del modelo al tercer o cuarto registro nuevo en lugar de hacerlo con cada registro nuevo puede mejorar el rendimiento. La selección de esta opción habilita el control **Número mínimo de registros nuevos para que sea necesaria la regeneración**.

Número mínimo de registros para que sea necesaria la regeneración. Puede elegirse cuándo se regeneran los modelos de series temporales en función del número de registros nuevos. Seleccione una de las opciones siguientes:

- Si está seleccionada la opción **Utilizar porcentaje de registros (%)**, los modelos de series temporales se regenerarán cuando un porcentaje especificado del número total de registros sean nuevos registros. El porcentaje predeterminado es del 10%.
- Si está seleccionada la opción **Utilizar número de registros absoluto**, el modelo de series temporales se regenerará cuando un número especificado de registros sean nuevos registros. El número mínimo de registros nuevos que se pueden especificar para que se regenere el modelo es de uno, y el máximo es de 100. El número predeterminado de registros es cuatro.

Campo para identificar nuevos registros. Seleccione qué campo se observa para detectar cuándo hay registros nuevos. El campo que se elija deberá ser el que ordena los registros en la serie temporal. Solo pueden elegirse campos continuos.

Nodo Cajas-espacio-tiempo

Las cajas-espacio-tiempo (STB) son una extensión de las ubicaciones espaciales demarcadas mediante Geohash. En concreto, una STB es una cadena alfanumérica que representa una región de forma regular de espacio y tiempo.

Por ejemplo, la STB `dr5ru7|2013-01-01 00:00:00|2013-01-01 00:15:00` consta de las tres partes siguientes:

- El geohash `dr5ru7`
- La indicación de fecha y hora inicial `2013-01-01 00:00:00`
- La indicación de fecha y hora final `2013-01-01 00:15:00`

Un ejemplo sería utilizar la información espacio-temporal para reforzar la certeza de que dos entidades son iguales al encontrarse virtualmente en el mismo lugar al mismo tiempo. De forma alternativa, podría mejorarse la precisión de la identificación de relaciones mostrando que dos entidades están relacionadas debido a su proximidad en el espacio y en el tiempo.

Pueden utilizarse los modos **Registros individuales** o **Lugares habituales** según convenga. Ambos modos requieren los mismos detalles básicos, como se indica a continuación:

Campo de latitud. Seleccione el campo que identifica la latitud (en sistema de coordenadas WGS84).

Campo de longitud. Seleccione el campo que identifica la longitud (en sistema de coordenadas WGS84).

Campo de indicación de la hora. Seleccione el campo que identifica la fecha u hora.

Opciones de un registro individual

Este modo se utiliza para añadir un campo adicional a un registro para identificar su ubicación en un momento dado.

Derivar. Seleccione una o más densidades de espacio y tiempo a partir de las cuales derivar el nuevo campo. Para obtener más información, consulte “Definición de la densidad de una caja-espacio-tiempo” en la página 107.

Extensión de nombre de campo. Permite escribir la extensión que desea añadir a los nombres de los campos nuevos. Puede optar por añadir esta extensión como un **Sufijo** o como un **Prefijo**.

Opciones de los lugares comunes

Puede pensarse en un lugar común como en una ubicación y/o hora en la que una entidad se encuentra de forma continua o repetida. Por ejemplo, esto podría utilizarse para identificar un vehículo que realiza rutas de transporte regulares e identificar posibles desviaciones de la norma.

El detector de lugares comunes supervisa el movimiento de entidades y condiciones de marcas en los que se observa que una entidad está “pasando tiempo” en el área. El detector de lugares comunes asigna automáticamente cada lugar común marcado a uno o más STB, y utiliza la entidad incorporada en la memoria y el seguimiento de eventos para detectar lugares comunes con una eficiencia óptima.

Densidad de la STB. Seleccione la densidad del espacio y del tiempo a partir de la cual derivar el nuevo campo. Por ejemplo, el valor `STB_GH4_10MINS` correspondería a un cuadro de geohash de cuatro caracteres de un tamaño aproximado de 20 km por 20 km y una ventana de tiempo de 10 minutos. Para obtener más información, consulte “Definición de la densidad de una caja-espacio-tiempo” en la página 107.

Campo de ID de entidad. Seleccione la entidad a utilizar como identificador del lugar común. Este campo de ID identifica el evento.

Número mínimo de eventos. Un evento es una fila en los datos. Seleccione el número mínimo de apariciones de un evento para que se considere que la entidad frecuenta un lugar común. Un lugar común también debe calificar basándose en el siguiente campo **El tiempo de permanencia es por lo menos**.

El tiempo de permanencia es por lo menos. Especifique el tiempo mínimo que una entidad debe permanecer en la misma ubicación. Esto puede ayudar, por ejemplo, a excluir un coche que espera en un semáforo de ser considerado como que frecuenta un lugar común. Un lugar común también debe calificar basándose en el campo anterior **Número mínimo de eventos**.

A continuación se muestran más detalles sobre lo que califica como un lugar común:

Supongamos que e_1, \dots, e_n indica todos los eventos ordenados temporalmente que se han recibido de un ID de entidad durante un periodo de tiempo (t_1, t_n) . Estos eventos califican como un lugar común si:

- $n \geq$ número mínimo de eventos
- $t_n - t_1 \geq$ tiempo mínimo de permanencia
- Todos los eventos e_1, \dots, e_n se producen en el mismo STB

Permitir que los lugares comunes rebasen los límites de una STB. Si esta opción está seleccionada, la definición de un lugar común es menos estricta y podría incluir, por ejemplo, una entidad que frecuente más de una caja-espacio-tiempo. Por ejemplo, si las STB se definen como horas enteras, la selección de esta opción reconocería como válida una entidad que frecuentase un lugar común durante una hora, incluso si dicha hora constara de los treinta minutos anteriores a la media noche y de los treinta minutos posteriores a la media noche. Si esta opción no está seleccionada, el 100% de tiempo en que se frecuenta un lugar común deberá entrar dentro de una única STB.

Mínima proporción de eventos que determina la calidad de lugar común de una caja temporal (%).

Sólo está disponible si **Permitir que los lugares comunes rebasen los límites de una STB** está seleccionada. Esto se usa para controlar el grado en que un lugar común notificado en una STB puede solaparse con otra. Seleccione la proporción mínima de eventos que debe tener lugar dentro de una única STB para identificar un lugar común. Si se establece en un 25%, y la proporción de eventos es del 26%, podrá considerarse un lugar común.

Por ejemplo, suponga que configura el detector de lugares comunes para que requiera al menos dos eventos (número mínimo de eventos = 2) y un tiempo mantenido contiguo de al menos 2 minutos en un cuadro de espacio de 4 bytes geohash y un cuadro de tiempo de 10 minutos (STB_NAME = STB_GH4_10MINS). Cuando se detecta un lugar común, la entidad se mantiene en el mismo cuadro de espacio de 4 bytes geohash mientras que los tres eventos que califican se producen dentro de un intervalo de tiempo de 10 minutos entre las 16:57 y las 17:07, a las 16:58, 17:01 y 17:03. El valor de porcentaje de cuadro de tiempo que califica especifica que la STB dé crédito para el lugar común, de la siguiente manera:

- **100%.** Se notifica un lugar común en el cuadro de tiempo 17:00 - 17:10 y no en el cuadro de tiempo 16:50 - 17:00 (los eventos 17:01 y 17:03 cumplen todas las condiciones necesarias para un lugar común y el 100% de estos eventos se han producido en el cuadro de tiempo 17:00 - 17:10).
- **50%.** Se notifican los lugares comunes en los dos cuadros de tiempo (los eventos 17:01 y 17:03 cumplen todas las condiciones necesarias para un lugar común y al menos el 50% de estos eventos se han producido en el cuadro de tiempo 17:50 - 17:00 y al menos el 50% de estos eventos se han producido en el cuadro de tiempo 17:00 - 17:10).
- **0%.** Se notifican los lugares comunes en los dos cuadros de tiempo.

Cuando se especifica 0%, los informes de lugar común incluyen las STB que representan cada cuadro de tiempo que se incluye en la duración de clasificación. La duración de clasificación debe ser menor o igual que la duración correspondiente del cuadro de tiempo en la STB. En otras palabras, nunca debe haber una configuración en la que una STB de 10 minutos está configurada en tándem con una duración de clasificación de 20 minutos.

Se informa de un lugar común tan pronto como se cumplen las condiciones de clasificación, y no se informa más de una vez por cada STB. Supongamos que tres eventos califican para un lugar común y se han producido un total de 10 eventos dentro de una duración de clasificación, todos dentro de la misma STB. En ese caso, se notifica el lugar común cuando se produce el tercer evento de clasificación. Ninguno de los siete eventos adicionales generan un informe de lugar común.

Nota:

- Los datos de evento en memoria del detector de lugares comunes no se comparten entre procesos. Por lo tanto, una entidad determinada tiene afinidad con un nodo de detector de lugares comunes particular. Es decir, los datos de movimiento de entrada para una entidad deben pasarse siempre coherentemente al nodo de detector de lugares comunes que realiza el seguimiento de esa entidad, que normalmente es el mismo nodo en toda la ejecución.
- Los datos de evento en memoria del detector de lugares comunes son volátiles. Cuando se sale y se reinicia el detector de lugares comunes, se pierden los lugares comunes de trabajo en curso. Esto significa que detener y reiniciar el proceso puede hacer que el sistema no notifique lugares comunes reales. Una solución potencial es reproducir algunos de los datos de movimiento históricos (por ejemplo, ir 48 horas hacia atrás y reproducir los registros de movimiento que son aplicables a cualquier nodo que se ha reiniciado).
- Al detector de lugares comunes se le deben proporcionar datos en orden secuencial de tiempo.

Definición de la densidad de una caja-espacio-tiempo

Elija el tamaño (densidad) de las Cajas-Espacio-Tiempo (STB) especificando el área física y el tiempo transcurrido a incluir en cada una de ellas.

Geodensidad. Seleccione el tamaño del área a incluir en cada STB.

Intervalo de tiempo. Seleccione el número de horas a incluir en cada STB.

Nombre de campo. Con el prefijo STB, esto se completa automáticamente en función de las selecciones de los dos campos anteriores.

Capítulo 4. Nodos de operaciones con campos

Conceptos básicos de las operaciones con campos

Una vez que haya realizado una exploración de datos inicial, es posible que tenga que seleccionar, limpiar o construir datos para preparar el análisis. La paleta Operaciones con campos contiene muchos nodos útiles para esta transformación y preparación.

Por ejemplo, con un nodo Derivar se puede crear un atributo que no se representa en los datos en la actualidad. También se puede utilizar un nodo Intervalos para recodificar automáticamente valores de campos en análisis objetivos. Con frecuencia utilizará los nodos Tipo, ya que permiten asignar valores, un nivel de medición y un rol de modelado a cada campo del conjunto de datos. Estas operaciones son útiles para gestionar valores perdidos y modelado posterior en la ruta.

La paleta Operaciones con campos contiene los siguientes nodos:



El nodo de preparación automática de datos (ADP) puede analizar sus datos e identificar los valores fijos, cribar los campos problemáticos o que no serán útiles y derivar nuevos atributos cuando sea necesario y mejorar el rendimiento mediante técnicas de cribado y muestreo inteligente. Puede utilizar el nodo de forma totalmente automática, permitiendo que el nodo seleccione y aplique valores fijos, o bien puede tener una vista previa de los cambios antes de que se apliquen y aceptarlos o rechazarlos.



El nodo Tipo especifica propiedades y metadatos de campo. Por ejemplo, puede especificar un nivel de medición (continuo, nominal, ordinal o marca) para cada campo, establecer las opciones para gestionar valores perdidos y nulos del sistema, establecer el rol de un campo con fines de modelado, especificar las etiquetas de valor y campo y especificar los valores de un campo.



El nodo Filtrar filtra (descarta) campos, vuelve a nombrar campos y correlaciona campos de nodo de origen a otro.



El nodo Derivar modifica los valores de datos o crea campos nuevos desde uno o más campos existentes. Crea campos del tipo fórmula, marca, nominal, estado, recuento y condicional.



El nodo Conjunto combina dos o más nugget de modelo para obtener predicciones más precisas que pueden conseguirse de cualquier modelo.



El nodo Rellenar sustituye valores de campos y cambia el almacenamiento. Puede sustituir los valores en función de una condición CLEM, como @BLANK(@FIELD). También puede sustituir todos los espacios vacíos o valores nulos por un valor específico. Un nodo Rellenar suelen utilizarse junto con un nodo Tipo para sustituir valores perdidos.



El nodo Anonimizar transforma la manera en que se representan los nombres y los valores de los campos a partir de ese punto de la ruta, lo que permite disfrazar los datos originales. Puede resultar útil si desea permitir que otros usuarios generen modelos utilizando datos confidenciales, como los nombres de los clientes u otros detalles.



El nodo Reclasificar transforma un conjunto de valores categóricos en otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis.



El nodo Intervalos crea automáticamente nuevos campos nominales (conjunto) en función de los valores de uno o más campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos como desviaciones desde la media. Una vez creados los intervalos para el campo nuevo, puede generar un nodo Derivar en función de los puntos de corte.



El nodo Análisis de RFM (actualidad, frecuencia, monetario) permite determinar cuantitativamente qué clientes son los mejores examinando cuándo ha sido la compra más reciente de un cliente (actualidad), cuántas veces suele comprar (frecuencia) y cuánto gasta el cliente en todas las transacciones (valor monetario).



El nodo Partición genera un campo de partición, que divide los datos en subconjuntos diferentes para las fases de entrenamiento, comprobación y validación en la generación del modelo.



El nodo Marcas deriva varios campos de marcas en función de los valores categóricos definidos para uno o más campos nominales.



El nodo Reestructurar convierte un campo nominal o marca en un grupo de campos que se puede rellenar con los valores todavía de otro campo. Por ejemplo, para un campo determinado llamado *tipo de pago*, con valores de *crédito*, *efectivo*, y *débito*, se crearían tres campos nuevos (*crédito*, *efectivo*, *débito*), que contendría cada uno el valor del pago real realizado.



El nodo Transponer intercambia los datos en filas y columnas de manera que los registros se conviertan en campos y los campos en registros.



El nodo Intervalos de tiempo especifica los intervalos y genera etiquetas (si es necesario) para modelar los datos de series temporales. Si los valores no están espaciados de manera uniforme, el nodo puede rellenar o agregar valores según sea necesario para crear un intervalo uniforme entre registros.



El nodo Intervalos de tiempo original no es compatible con el Servidor de análisis (AS). El nodo Intervalos de tiempo AS (nuevo en SPSS Modeler versión 17.0) contiene un subconjunto de las funciones del nodo Intervalos de tiempo existente que se puede utilizar con Servidor de análisis.

Utilice el nodo Intervalos de tiempo AS para especificar intervalos y derivar un campo de tiempo nuevo para la estimación o previsión. Se admite una gama completa de intervalos de tiempo que abarca desde segundos a años.



El nodo Historial se utiliza para crear campos nuevos que contienen datos de los campos de registros anteriores. Los nodos Historial se suelen utilizar para los datos secuenciales, como los datos de series temporales. Antes de utilizar un nodo Historial, puede desear ordenar los datos utilizando un nodo Ordenar.



El nodo Reorg. campos define el orden natural utilizado para mostrar los campos en la parte posterior de la ruta. Este orden afecta a la visualización de los campos en diversas ubicaciones, como las tablas, las listas y el selector de campos. Esta operación resulta útil al trabajar con conjuntos de datos amplios que hacen más visibles los campos de interés.



En SPSS Modeler, elementos como las funciones espaciales Creador de expresiones, el nodo de Predicción espacio-temporal (STP) y el nodo Visualización de mapas utilizan el sistema de coordenadas proyectado. Utilice el nodo Reproyectar para cambiar el sistema de coordenadas de los datos que importa y que utilizan un sistema de coordenadas geográficas.

Algunos de estos nodos pueden generarse directamente desde el informe de auditoría creado por el nodo Auditoría de datos. Consulte el tema “Generación de otros nodos para la preparación de datos” en la página 315 para obtener más información.

Preparación automática de datos

La preparación de los datos para su análisis es uno de los pasos más importantes en cualquier proyecto y, tradicionalmente, uno de los que más tiempo requieren. La Preparación automática de datos (ADP) controla las tareas automáticamente, analizando los datos e identificando problemas, cribando campos problemáticos o sin posibilidades de ser útiles, derivando nuevos atributos cuando sea necesario y mejorando el rendimiento mediante técnicas de cribado inteligente. Puede utilizar el algoritmo de una forma totalmente **automática**, permitiendo seleccionar y aplicar soluciones; o de forma **interactiva**, previniendo los cambios antes de que se realicen y aceptarlos o rechazarlos según sea necesario.

ADP permite hacer que sus datos estén listos para la generación de modelos de forma rápida y fácil, sin necesidad de tener conocimientos previos de los conceptos previos implicados. Los modelos tienden a crearse y puntuarse con mayor rapidez; además, el uso de ADP mejora la solidez de los procesos de modelado automatizados, como actualización de modelos y análisis Champion/Challenger.

Nota: cuando el ADP prepara un campo para su análisis, crea un nuevo campo con los ajustes o transformaciones, en vez de reemplazar los valores y propiedades existentes del campo anterior. El campo anterior no se usa en más análisis, su rol se define como Ninguno.

Ejemplo. Una correduría de seguros con recursos limitados para investigar las reclamaciones de seguros de los asegurados desea crear un modelo para etiquetar las reclamaciones sospechosas y potencialmente fraudulentas. Antes de construir el modelo, leerán los datos para el modelado mediante la preparación automática de datos. Como desean revisar las transformaciones propuestas antes de que se apliquen las transformaciones, utilizarán la preparación automática de datos en modo interactivo.

Un grupo del sector del automóvil desea realizar un seguimiento de las ventas de diversos vehículos a motor. Para poder identificar los modelos como mejor y peor rendimiento, desean establecer una relación entre las ventas de vehículos y las características de los vehículos. Utilizarán la preparación automática de datos para preparar los datos para el análisis y crearán modelos utilizando la preparación "anterior" y "posterior" de datos para ver cómo difieren los resultados.

¿Cuál es su objetivo? Preparación automática de datos recomienda ejecutar pasos para la preparación de datos que afectan a la velocidad con la que el resto de algoritmos pueden generar modelos y mejorar el potencial predictivo de esos modelos. Pueden incluir la transformación, construcción y selección de características. El destino también puede transformarse. Puede especificar las prioridades de generación de modelos en las que se deben centrar el proceso de preparación de datos.

- **Equilibrar velocidad y precisión.** Esta opción prepara los datos para dar igual prioridad a la velocidad con la que se procesan los datos por algoritmos de creación de modelos y la precisión de las predicciones.
- **Optimizar velocidad.** Esta opción prepara los datos para dar prioridad a la velocidad con la que se procesan los datos por los algoritmos de generación de modelos. Cuando esté trabajando con conjuntos de datos de gran tamaño o busque una respuesta rápida, seleccione esta opción.
- **Optimizar precisión.** Esta opción prepara los datos para dar prioridad a la precisión de las predicciones producidas por los algoritmos de generación de modelos.
- **Análisis personalizado.** Seleccione esta opción si desea cambiar manualmente el algoritmo de la pestaña Configuración. Tenga en cuenta que esta configuración se selecciona automáticamente si realiza cambios posteriores a muchas opciones de la pestaña Configuración que sean incompatibles con los de otros objetivos.

Formación del nodo

El nodo ADP se implementa como nodo de proceso y funciona de forma parecida al nodo Tipo; el **entrenamiento** del nodo ADP se corresponde con la instanciación en el nodo Tipo. Una vez realizado el análisis, las transformaciones especificadas se aplican a los datos sin realizar más análisis siempre que el modelo de datos anteriores no cambie. Al igual que los nodos Tipo y Filtrar, si el nodo ADP está desconectado recuerda el modelo de datos y las transformaciones, de modo que si vuelve a conectarse no hay que volver a entrenarlo; esto le permite entrenarlo en un subconjunto de datos normales y a continuación copiarlo o desplegarlo para utilizarlo en datos activos con la frecuencia necesaria.

Uso de la barra de herramientas

La barra de herramientas le permite ejecutar y actualizar la visualización del análisis de datos y generar nodos que puede utilizar junto con los datos originales.

- **Generar** Desde este menú puede generar un nodo Filtrar o Derivar. Tenga en cuenta que este menú sólo está disponible cuando se muestra un análisis en la pestaña Análisis.

El nodo Filtrar elimina los campos de entrada transformados. Si configura el nodo ADP para que deje los campos de entrada originales en el conjunto de datos, se restaurará el conjunto original de entradas que le permitirán interpretar el campo de puntuación en cuanto a las entradas. Por ejemplo, puede ser de gran utilidad si desea producir un gráfico del campo de puntuación con respecto a diversas entradas.

El nodo Derivar puede restaurar el conjunto de datos original y las unidades destino. Sólo puede generar un nodo Derivar si el nodo ADP contiene un análisis que vuelve a trazar un destino de rango (es decir, el cambio de escala Box-Cox está seleccionado en el panel Preparar entradas y objetivo). No puede generar un nodo Derivar si el destino no es un rango o si el cambio de escala Box-Cox no está seleccionado. Consulte el tema "Generación de un nodo Derivar" en la página 124 para obtener más información.

- **Ver** Contiene opciones que controlan lo que se muestra en la pestaña Análisis. Se incluyen los controles de edición de gráficos y las selecciones de visualización para el panel principal y las vistas vinculadas.

- **Presentación preliminar** Muestra las transformaciones que se aplicarán a los datos de entrada.
- **Analizar datos** Inicia un análisis mediante la configuración actual y muestra los resultados en la pestaña Análisis.
- **Borrar análisis** Elimina el análisis existente (sólo está disponible cuando existe un análisis actual).

Estado del nodo

El estado del nodo ADP en el lienzo de IBM SPSS Modeler se indica mediante una flecha o marca en el icono que muestra si el análisis se ha realizado o no.

Pestaña Campos

Para generar un modelo, primero se deben especificar los campos que se desea usar como objetivos y como entradas. Salvo algunas excepciones, todos los nodos de modelado usarán la información de los campos procedente de un nodo Tipo anterior en la ruta. Si utiliza un nodo Tipo para seleccionar campos de entrada y objetivo, no es necesario cambiar nada en esta pestaña.

Utilizar configuración del nodo Tipo. Esta opción permite indicar al nodo que use la información de campo de un nodo Tipo situado en un punto anterior de la ruta. Este es el método predeterminado.

Utilizar configuración personalizada. Esta opción permite indicar al nodo que use la información de campo especificada aquí en lugar de la proporcionada en nodos Tipo situados en cualquier punto anterior de la ruta. Después de seleccionar esta opción, especifique los campos siguientes si es necesario.

Objetivo. En los modelos que requieran uno o varios campos objetivo, selecciónelos. Se trata de una acción similar a establecer el rol del campo en *Objetivo* en un nodo Tipo.

Entradas. Seleccione el campo(s) de entrada. Se trata de una acción similar a establecer el rol del campo en *Entrada* en un nodo Tipo.

Pestaña Configuración

La pestaña Configuración contiene diferentes grupos de ajustes que puede modificar para ajustar con precisión la forma en que el algoritmo procesa sus datos. Si realiza algún cambio en la configuración predeterminada que sea incompatible con el resto de objetivos, la pestaña Objetivo se actualiza automáticamente para seleccionar la opción **Personalizar análisis**.

Configuración de campo

Utilizar campo de frecuencia Esta opción le permite seleccionar un campo como ponderación de frecuencia. Utilícela si los registros de sus datos de formación representan más de una unidad cada uno, por ejemplo, si está utilizando datos agregados. Los valores del campo deben ser el número de unidades representadas por cada registro.

Utilizar campo de ponderación Esta opción le permite seleccionar un campo como ponderación de casos. Las ponderaciones de casos se usan para contabilizar las diferencias existentes en la varianza entre los niveles del campo de salida.

Cómo gestionar los campos excluidos del modelado Especifica lo que ocurre a los campos excluidos; puede optar por filtrarlos de los datos o simplemente definir su *Rol* como **Ninguno**.

Note: Esta acción también se aplicará al destino si se transforma. Por ejemplo, si la nueva versión derivada del objetivo se utiliza como el campo **Destino**, el objetivo original se filtra o se establece como **Ninguno**.

Si los campos entrantes no coinciden con el análisis existente Especifique lo que ocurre si uno o más campos de entrada obligatorios faltan del conjunto de datos entrante cuando ejecuta un nodo ADP entrenado.

- **Detener la ejecución y mantener el análisis existente** Esto detiene el proceso de ejecución, conserva la información del análisis actual y muestra un error.
- **Borrar el análisis existente y analizar los nuevos datos** Esto borra el análisis existente, analiza los datos entrantes y aplica las transformaciones recomendadas a dichos datos.

Preparar fechas y horas

Muchos algoritmos no pueden tratar directamente los detalles de fecha y hora; estas configuraciones permiten derivar nuevos datos de duración que pueden utilizarse como entradas de modelo de fechas y horas de sus datos existentes. Los campos que contienen las fechas y las horas se deben predefinir con los tipos de almacenamiento de fecha u hora. Los campos de fecha y hora originales no se recomiendan como entradas de modelo posteriores a la preparación automática de datos.

Preparar fechas y horas para el modelado. Si cancela la selección de esta opción se desactivan todos los demás controles de Preparar fechas y horas mientras se mantienen las selecciones.

Calcular tiempo transcurrido hasta fecha de referencia. Esto produce el número de años/meses/días desde una fecha de referencia para cada variable que contenga fechas.

- **Fecha de referencia.** Especifique la fecha desde la que se calculará la duración en lo relativo a la información de fecha de los datos de entrada. Si selecciona **Fecha de hoy**, la fecha actual del sistema se utilizará siempre que se ejecute el nodo ADP. Para utilizar una fecha específica, seleccione **Fecha fija** e introduzca la fecha obligatoria. La fecha actual se especifica automáticamente en el campo **Fecha fija** cuando se crea el nodo por primera vez.
- **Unidades de duración de fecha.** Especifique si el nodo debería decidir automáticamente sobre la unidad de duraciones de fecha o establezca **Unidades fijas** como Años, Meses o Días.

Calcular tiempo transcurrido hasta hora de referencia. Esto produce el número de horas/minutos/segundos desde una hora de referencia para cada variable que contenga horas.

- **Hora de referencia.** Especifique la hora desde la que se calculará la duración en lo relativo a la información de hora de los datos de entrada. Si selecciona **Hora actual**, la hora actual del sistema se utilizará siempre que se ejecute el nodo ADP. Para utilizar una hora específica, seleccione **Hora fija** e introduzca los detalles obligatorios. La hora actual se introduce automáticamente en el campo **Hora fija** cuando se crea el nodo por primera vez.
- **Unidades de duración de tiempo.** Especifique si el nodo debería decidir automáticamente sobre la unidad de duraciones de hora o establezca **Unidades fijas** como Horas, Minutos o Segundos.

Extraer elementos temporales cíclicos. Utilice esta configuración para dividir un único campo de fecha o de hora en uno o más campos. Por ejemplo, si selecciona las tres casillas de verificación de fecha, el campo de fecha de entrada "1954-05-23" se dividirá en tres campos: 1954, 5 y 23, cada uno con el sufijo definido en el panel **Nombres de campos**, y el campo de fecha original se ignorará.

- **Extraer de fechas.** Para cualquier entrada de fecha, especifique si desea extraer años, meses, días o cualquier combinación.
- **Extraer de horas.** Para cualquier entrada de hora, especifique si desea extraer horas, minutos, segundos o cualquier combinación.

Excluir campos

Los datos de mala calidad pueden afectar a la precisión de sus predicciones; por lo tanto, puede especificar el nivel de calidad aceptable de las características de entrada. Todos los campos que no sean constantes o les falte el 100% de los valores se excluirán automáticamente.

Excluir campos de entrada de baja calidad. Si cancela la selección de esta opción se desactivan todos los demás controles de Excluir campos mientras se mantienen las selecciones.

Excluir campos con demasiados valores perdidos. Los campos con un porcentaje de valores perdidos mayor que el porcentaje especificado se eliminan de análisis posteriores. Especifique un valor superior o igual a 0 (que equivale a cancelar la selección de esta opción) y menor o igual a 100, aunque los campos que tienen valores que faltan se excluyen automáticamente. El valor por defecto es 50.

Excluir campos nominales con demasiadas categorías exclusivas. Los campos nominales con un número de categorías superior al especificado se eliminarán de análisis posteriores. Especifique un número entero positivo. El valor predeterminado es 100. Esto resulta útil para eliminar automáticamente campos que contengan información registros exclusivos para el modelado, como ID, dirección o nombre.

Excluir campos categóricos con demasiados valores en una única categoría. Los campos nominales y ordinales con una categoría con un porcentaje de registros superior al especificado se eliminarán de análisis posteriores. Especifique un valor superior o igual a 0 (que equivale a cancelar la selección de esta opción) y menor o igual a 100, aunque los campos constantes se excluyen automáticamente. El valor predeterminado es 95.

Preparación de entradas y destinos

Ya que los datos nunca están en el estado perfecto para su procesamiento, puede que desee ajustar algunas configuraciones antes de ejecutar un análisis. Por ejemplo, esto puede incluir la eliminación de valores atípicos, la especificación de cómo tratar los valores perdidos o el ajuste del tipo.

Note: Si cambia los valores de este panel, la pestaña **Objetivos** se actualiza automáticamente para seleccionar la opción **Análisis personalizado**.

Preparar los campos de entrada y destino para el modelado. Activa y desactiva todos los campos del panel.

Ajustar el tipo y mejorar la calidad de los datos. Para las entradas y el destino puede especificar varias transformaciones de datos independientemente; esto es así porque tal vez no desee cambiar los valores del destino. Por ejemplo, una predicción de los ingresos en dólares tiene más sentido que una predicción medida en $\log(\text{dollars})$. Además, si el destino tiene valores perdidos, el hecho de rellenar los valores perdidos no supone una ganancia predictiva, mientras que rellenando los valores perdidos en las entradas se puede permitir que algunos algoritmos procesen información que de otro modo se perdería.

Las configuraciones adicionales para estas transformaciones, como el valor de corte atípico, son comunes para el destino y las entradas.

Puede seleccionar la siguiente configuración para las entradas, para el destino o para ambos:

- **Ajustar el tipo de campos numéricos.** Seleccione esta opción para determinar si los campos numéricos con un nivel de medición *Ordinal* puede convertirse a *Continuo* o viceversa. Puede especificar los valores de umbral mínimo y máximo para controlar la conversión.
- **Reordenar campos nominales.** Seleccione esta opción para ordenar los campos nominales (conjunto) de menor a mayor según su categoría.
- **Sustituir valores extremos en campos continuos.** Especifique si desea sustituir los valores atípicos; utilice esta opción junto con la opción **Método de sustitución de valores atípicos** que aparece a continuación.
- **Campos continuos: sustituir valores perdidos por la media.** Seleccione esta opción para sustituir los valores de las características continuas (rango).
- **Campos nominales: sustituir valores perdidos por el modo.** Seleccione esta opción para sustituir los valores perdidos de características nominales (conjunto).
- **Campos ordinales: sustituir valores perdidos por la mediana.** Seleccione esta opción para sustituir los valores perdidos de características ordinales (conjunto ordenado).

Número máximo de valores de campos ordinales. Especifique el umbral para redefinir los campos ordinales (conjunto ordenado) como continuo (rango). El valor predeterminado es 10, por lo que si un campo ordinal tiene más de 10 categorías, se redefine como continuo (rango).

Número mínimo de valores de campos continuos. Especifique el umbral para redefinir los campos de escalas o continuo (rango) como ordinal (conjunto ordenado). El valor predeterminado es 5, por lo que si un campo continuo tiene menos de 5 valores, se redefine como ordinal (conjunto ordenado).

Valor de corte atípico. Especifique el criterio de corte atípico, medido en desviaciones estándar; el valor predeterminado es 3.

Método de sustitución de valores atípicos. Seleccione si los valores atípicos deben sustituirse mediante el recorte (forzar) con el valor de corte o si deben eliminarse y establecerse como valores perdidos. Todos los valores atípicos establecidos como valores perdidos siguen la configuración de gestión de valores perdidos seleccionada anteriormente.

Poner todos los campos de entrada continuos en una escala común. Para normalizar los campos de entrada continuos, seleccione esta casilla de verificación y seleccione el método de normalización. El valor predeterminado es **Transformación de puntuaciones z**, donde puede especificar la **Media final**, que tiene un valor predeterminado de 0, y la **Desviación estándar final**, que tiene un valor predeterminado de 1. También puede decidir utilizar **Transformación mínima/máxima** y especificar los valores mínimo y máximo, que tienen valores predeterminados de 0 y 100, respectivamente.

Este campo es especialmente útil cuando selecciona **Realizar creación de características** en el panel Crear y seleccionar características.

Cambiar la escala de un objetivo continuo con una transformación de Box-Cox. Para normalizar un campo objetivo continuo (escala o rango), seleccione esta casilla de verificación. La transformación de Box-Cox tiene valores predeterminados de 0 para la **Media final** y de 1 para la **Desviación estándar final**.

Note: Si decide normalizar el destino, la dimensión del destino se transformará. En este caso, puede que necesite generar un nodo Derivar para aplicar una transformación inversa para volver a cambiar las unidades transformadas a un formato reconocible para procesarlo con posterioridad. Consulte el tema “Generación de un nodo Derivar” en la página 124 para obtener más información.

Creación y selección de características

Para mejorar el poder predictivo de sus datos, puede transformar los campos de entrada o crear otros nuevos basados en los campos existentes.

Note: Si cambia los valores de este panel, la pestaña **Objetivos** se actualiza automáticamente para seleccionar la opción **Análisis personalizado**.

Transformar, crear y seleccionar campos de entrada para mejorar el poder predictivo. Activa y desactiva todos los campos del panel.

Combinar categorías dispersas para aprovechar al máximo la asociación con el destino. Seleccione esta opción para realizar un modelo más parsimonioso reduciendo el número de variables que deben procesarse junto con el destino. Si es necesario, cambie el valor de probabilidad del valor predeterminado de 0,05.

Tenga en cuenta que si todas las categorías se combinan en una, las versiones original y derivada del campo se excluyen porque no tienen ningún valor como predictor.

Si no hay ningún destino, combine las categorías dispersas según los recuentos. Si está tratando con datos que no tienen destino, puede fusionar las categorías dispersas de características ordinales (conjunto

ordenado), nominales (conjunto) o ambas. Especifique el porcentaje mínimo de casos o registros de los datos que identifica las categorías que deben fusionarse; el valor predeterminado es 10.

Las categorías se fusionan mediante las reglas siguientes:

- La fusión no se realiza en campos binarios.
- Si sólo hay dos categorías durante la fusión, ésta se detiene.
- Si no hay ninguna categoría original ni ninguna categoría creada durante la fusión con menos del porcentaje mínimo especificado de casos, la fusión se detiene.

Desechar campos discontinuos al preservar el poder predictivo. Si el conjunto de datos incluye un destino categórico, puede crear un intervalo para entradas continuas con asociaciones fuertes para mejorar el rendimiento del procesamiento. Si es necesario, cambie el valor de probabilidad para los subconjuntos homogéneos del valor predeterminado de 0,05.

Si la operación de creación de intervalos da como resultado un único intervalo para un campo específico, las versiones original y con intervalos del campo se excluyen porque no tienen ningún valor como predictor.

Note: Los intervalos en ADP son diferentes de intervalos óptimos utilizados en otras partes de IBM SPSS Modeler. Intervalos óptimos utiliza entropía de información para convertir una variable continua en una variable categórica; necesita ordenar los datos y almacenarlo todo en memoria. ADP utiliza subconjuntos homogéneos para agrupar una variable continua, significa que intervalos ADP no necesita ordenar los datos ni almacenar los datos en memoria. El uso del método de subconjunto homogéneo para agrupar una variable continua significa que el número de categorías después de la agrupación es siempre menor o igual que el número de categorías del destino.

Realizar selección de características. Seleccione esta opción para eliminar características con un coeficiente de correlación bajo. Si es necesario, cambie el valor de probabilidad del valor predeterminado de 0,05.

Esta opción sólo se aplica a características de entrada continuas en las que el destino sea continuo y a características de entrada categóricas.

Realizar construcción de características. Seleccione esta opción para derivar nuevas características de una combinación de varias características existentes (que luego se descartan del modelado).

Esta opción sólo es aplicable a características de entrada continuas en las que el destino es continuo o en las que no hay destino.

Nombres de campos

Para identificar fácilmente las características nuevas y transformadas, ADP crea y aplica nombres, prefijos o sufijos básicos nuevos. Puede modificar estos nombres para que sean más relevantes para sus propias necesidades y datos. Si desea especificar otras etiquetas, deberá realizarlo en un nodo Tipo posterior.

Campos transformados y contruidos. Especifique las extensiones de nombre que se aplicarán a campos de entrada y de destino transformado.

Tenga en cuenta que en el nodo ADP, si establece que los campos de cadena no contengan nada puede provocar un error dependiendo de cómo decida gestionar los campos sin utilizar. Si **Cómo gestionar los campos excluidos del modelado** se establece como **Filtrar campos no utilizados** en el panel Configuración de campos de la pestaña Configuración, las extensiones de nombres de entradas y el destino pueden establecerse como nada. Los campos originales se filtran y los campos transformados se guardan sobre ellos; en este caso, los nuevos campos transformados tendrán el mismo nombre que el original.

Sin embargo si selecciona **Definir la dirección de campos no utilizados a 'Ninguna'**, las extensiones de nombre vacías o nulas para el destino y las entradas provocarán un error porque estará intentando crear nombres de campo duplicados.

Además, especifique el nombre de prefijo que se aplicará a todas las características que se creen mediante la configuración de Crear y seleccionar. El nuevo nombre se crea adjuntando un sufijo numérico a este nombre de raíz de prefijo. El formato del número depende de cuántas nuevas características se deriven, por ejemplo:

- 1-9 características creadas se denominarán: característica1 a característica9.
- 10-99 características creadas se denominarán: característica01 a característica99.
- 100-999 características creadas se denominarán: característica001 a característica999, etc.

De esta forma se garantiza que las características creadas se ordenen de forma adecuada independientemente de cuántas sean.

Duraciones calculadas de fechas y horas. Especifique las extensiones de nombre que se aplicarán a duraciones calculadas a partir de fechas y horas.

Elementos cíclicos extraídos de fechas y horas. Especifique las extensiones de nombre que se aplicarán a elementos cíclicos extraídos de fechas y horas.

Pestaña análisis

1. Cuando haya terminado con la configuración del nodo ADP, incluyendo las modificaciones realizadas en las pestañas **Objetivos**, **Campos** y **Configuración**, pulse **Analizar datos**; el algoritmo aplica la configuración a las entradas de datos y muestra los resultados en la pestaña **Análisis**.

La pestaña **Análisis** contiene resultados tabulares y gráficos que resumen el procesamiento de sus datos y muestra recomendaciones acerca de cómo los datos se pueden modificar o mejorar para establecer la puntuación. Puede revisar y aceptar o rechazar esas recomendaciones.

La pestaña **Análisis** se compone de dos paneles, la vista principal en la parte izquierda y la vista relacionada o auxiliar de la derecha. Hay tres vistas principales:

- Resumen de procesamiento de campos (la configuración predeterminada). Consulte el tema “Resumen de procesamiento de campo” en la página 119 para obtener más información.
- Campos. Consulte el tema “Campos” en la página 119 para obtener más información.
- Resumen de acciones. Consulte el tema “Resumen de acciones” en la página 120 para obtener más información.

Hay cuatro vistas relacionadas/auxiliares:

- Poder predictivo (la configuración predeterminada). Consulte el tema “Poder predictivo” en la página 121 para obtener más información.
- Tabla de campos. Consulte el tema “Tabla de campos” en la página 121 para obtener más información.
- Detalles de campo. Consulte el tema “Detalles de campo” en la página 121 para obtener más información.
- Detalles de acción. Consulte el tema “Detalles de acción” en la página 122 para obtener más información.

Enlaces entre vistas

En la vista principal, el texto subrayado de las tablas controla la visualización en la vista vinculada. Si pulsa el texto podrá obtener detalles de un campo concreto, conjunto de campos o paso de procesamiento. El enlace que ha seleccionado aparece en color más oscuro; de esta forma podrá identificar la conexión entre el contenido de los dos paneles de vista.

Restablecimiento de las vistas

Para volver a mostrar las recomendaciones de análisis originales y abandonar los cambios que haya realizado en las vistas de análisis, pulse **Restablecer** en la parte inferior del panel de vista principal.

Resumen de procesamiento de campo

La tabla Resumen de procesamiento de campos proporciona una instantánea del impacto total previsto de procesamiento, incluyendo los cambios en el estado y el número de características creadas.

Tenga en cuenta que no se crea un modelo realmente, por lo que no existe una medida ni un gráfico del cambio con el poder predictivo total antes y después de la preparación de los datos. Por contra, puede visualizar los gráficos de poder predictivo de los predictores individuales recomendados.

La tabla muestra la siguiente información:

- El número de campos de destino.
- El número de predictores (de entrada) originales.
- Los predictores recomendados para su uso en el análisis y modelado. Incluye el número total de campos recomendados; el número de campos originales sin transformar recomendados; el número de campos transformados recomendados (excluyendo las versiones intermedias de campos, campos derivados de predictores de fecha y hora y predictores creados); el número de campos recomendados derivados de los campos de fecha/hora; y el número de predictores creados recomendados.
- El número de predictores de entrada no recomendados para su uso en cualquier formulario, ya sea en su formato original, como campo derivado o como entrada en un predictor creado.

Si cualquiera de la información de los **Campos** está subrayada, pulse para visualizar más detalles en una vista vinculada. Los detalles de **Destino**, **Características de entrada** y **Características de entrada no utilizadas** se muestran en la vista vinculada Tabla de campos. Consulte el tema “Tabla de campos” en la página 121 para obtener más información. Las **Características recomendadas para su uso en el análisis** se muestran en la vista vinculada Poder predictivo. Consulte el tema “Poder predictivo” en la página 121 para obtener más información.

Campos

La vista principal Campos muestra los campos procesados y si el modo ADP recomienda su uso en modelos posteriores. Puede omitir la recomendación de cualquier campo; por ejemplo, para excluir las características creadas o incluir características que el nodo ADP recomienda excluir. Si un campo se ha transformado, puede decidir si acepta la transformación sugerida o utiliza la versión original.

La vista Campos tiene dos tablas, una para el destino y otra para los predictores procesados o creados.

Tabla Destino

La tabla **Destino** sólo se muestra si se ha definido un destino en los datos.

La tabla contiene dos columnas:

- **Nombre.** Es el nombre de la etiqueta o del campo de destino; el nombre del original se utiliza siempre, incluso si el campo se ha transformado.
- **Nivel de medición.** Muestra el icono que representa el nivel de medición; pase el ratón por encima del icono para mostrar una etiqueta (continuo, ordinal, nominal, etc.) que describe los datos.
Si el destino se ha transformado, la columna **Nivel de medición** refleja la versión final transformada.
Note: no puede desactivar las transformaciones del destino.

Tabla Predictores

La tabla **Predictores** se muestra siempre. Cada fila de la tabla representa un campo. De forma predeterminada, las filas se clasifican en orden descendente de potencia predictiva.

En características ordinarias, el nombre original siempre se utiliza como el nombre de la fila. Las versiones original y derivada de los campos de fecha/hora aparecen en la tabla (en filas separadas); la tabla también incluye los predictores creados.

Tenga en cuenta que las versiones transformadas de los campos que aparecen en la tabla siempre representan las versiones finales.

De forma predeterminada, sólo se muestran los campos recomendados en la tabla Predictores. Para mostrar el resto de campos, seleccione el cuadro **Incluir campos no recomendados en la tabla** encima de la tabla; estos campos se mostrarán en la parte inferior de la tabla.

La tabla muestra las siguientes columnas:

- **Versión de uso.** Muestra una lista desplegable que controla si un campo se utilizará posteriormente y si se utilizarán las transformaciones sugeridas. De forma predeterminada, la lista desplegable refleja las recomendaciones.

Para los predictores comunes que se han transformado, la lista desplegable contiene tres opciones: **Transformadas, Original y No utilizar.**

Para los predictores comunes sin transformar, las opciones son: **Original y No utilizar.**

Para campos derivados de fecha/hora y predictores contruidos, las opciones son: **Transformadas y No utilizar.**

Para los campos de fecha originales, la lista desplegable está desactivada y definida a **No utilizar.**

Nota: Para los predictores con versiones originales y transformadas, si cambia entre las versiones **Original** y **Transformadas**, se actualiza automáticamente la configuración de **Nivel de medición** y **Poder predictivo** de esas características.

- **Nombre.** Cada nombre de campo es un enlace. Pulse un nombre para ver más información acerca del campo en la vista vinculada. Consulte el tema “Detalles de campo” en la página 121 para obtener más información.
- **Nivel de medición.** Muestra el icono que representa el tipo de datos; pase el ratón por encima del icono para mostrar una etiqueta (continuo, ordinal, nominal, etc.) que describe los datos.
- **Poder predictivo.** El poder predictivo sólo se muestra en los campos que ADP recomienda. Esta columna no se muestra si no hay un destino definido. La potencia predictiva varía de 0 a 1, siendo los mayores valores los que indican "mejores" predictores. En general, la potencia predictiva resulta útil para comparar predictores con un análisis ADP, aunque los valores de potencia predictiva no deben compararse en distintos análisis.

Resumen de acciones

En cada acción realizada por la preparación automática de datos, los predictores de entrada se transforman y/o se filtran; los campos que sobreviven una acción se utilizarán en la acción siguiente. Los campos que sobreviven hasta el último paso se recomiendan para su uso en modelado, mientras que los predictores creados y transformados se filtran.

El Resumen de acciones es una sencilla tabla que enumera las acciones de procesamiento realizadas por ADP. Si alguna **Acción** está subrayada, pulse para ver más detalles en una vista vinculada sobre las acciones que se realizan. Consulte el tema “Detalles de acción” en la página 122 para obtener más información.

Note: Sólo se muestran las versiones transformadas originales y finales de cada campo, no las versiones intermedias utilizadas durante el análisis.

Poder predictivo

Se muestra de forma predeterminada cuando el análisis se ejecuta por primera vez o cuando selecciona **Predictores recomendados para su uso en análisis** en la vista principal Resumen del procesamiento de campos, el gráfico muestra el poder predictivo de los predictores recomendados. Los campos se ordenan según el poder predictivo, con el campo de mayor valor en la parte superior.

En versiones transformadas de predictores comunes, el nombre del campo refleja su elección del sufijo en el panel Nombres de campos de la pestaña Configuración, por ejemplo: *_transformed*.

Los iconos de Nivel de medición se muestran después de los nombres de campo individuales.

El poder predictivo de cada predictor recomendado se calcula a partir de una regresión lineal o un modelo simple Bayes, dependiendo de si el destino es continuo o categórico.

Tabla de campos

La vista Tabla de campos aparece cuando pulsa en **Destino**, **Predictores** o **Predictores no utilizados** en la vista principal Resumen del procesamiento de campos y muestra una tabla simple con las características relevantes.

La tabla contiene dos columnas:

- **Nombre.** Nombre del predictor.

En los destinos se utiliza el nombre original o la etiqueta del campo, incluso si el destino se ha transformado.

En versiones transformadas de predictores comunes, el nombre refleja su elección del sufijo en el panel Nombres de campos de la pestaña Configuración, por ejemplo: *_transformed*.

En los campos derivados de las fechas y horas se utiliza el nombre de la versión final transformada; por ejemplo: *bdate_years*.

En los predictores creados, se utiliza el nombre del predictor creado, por ejemplo: *Predictor1*.

- **Nivel de medición.** Muestra el icono que representa el tipo de datos.

En Destino, el **Nivel de medición** siempre refleja la versión transformada (si el destino se ha transformado); por ejemplo, si se ha cambiado de ordinal (conjunto ordenado) a continuo (rango, escala) o viceversa.

Detalles de campo

Se muestra cuando pulsa cualquier **Nombre** en la vista principal Campos, la vista Detalles de campo contiene distribución, valores perdidos y gráficos de poder predictivo (si procede) del campo seleccionado. Además, también se muestran el historial de procesamiento del campo y el nombre del campo transformado (si procede).

En cada conjunto de gráficos, las dos versiones se muestran juntas para comparar el campo con y sin las transformaciones aplicadas; si no existe una versión transformada del campo, se muestra un gráfico de la versión original únicamente. En los campos de fecha u hora derivados y en los predictores creados, los gráficos sólo se muestran para el nuevo predictor.

Note: Si se excluye un campo porque tiene demasiadas categorías, solo se muestra el historial de procesamiento.

Gráfico Distribución

La distribución de campos continuos se muestra como una curva normal superpuesta y una línea de referencia vertical para el valor medio; los campos categóricos se muestran como un gráfico de barras.

Los histogramas se etiquetan y muestran la desviación y asimetría estándar; sin embargo, la asimetría no se muestra si el número de valores es 2 o menos si la varianza del campo original es inferior a 10-20.

Pase el ratón por encima del gráfico para mostrar la media de los histogramas o el número y el porcentaje del número total de registros para las categorías en gráficos de barras.

Gráfico de valor perdido

Los gráficos circulares comparan el porcentaje de valores perdidos con y sin transformaciones aplicadas; las etiquetas de gráficos muestran el porcentaje.

Si el nodo ADP ha ejecutado la gestión de valores perdidos, el gráfico circular posterior a la transformación también incluye el valor de sustitución como una etiqueta, es decir, el valor que se utiliza en lugar de los valores perdidos.

Pase el ratón por encima del gráfico para mostrar el valor perdido y el porcentaje del número total de registros.

Gráfico de poder predictivo

En los campos recomendados, los gráficos de barras muestran el poder predictivo antes y después de la transformación. Si el destino se ha transformado, el poder predictivo se calcula con respecto al destino transformado.

Note: Los gráficos de poder predictivo no se muestran si se define el destino o si el pulsa el destino en el panel de vista principal.

Pase el ratón por encima del gráfico para mostrar el valor del poder predictivo.

Tabla Historial de procesamiento

La tabla muestra cómo se ha derivado la versión transformada de un campo. Las acciones que realiza ADP aparecen en el orden en que se ejecutan; sin embargo, en algunos pasos se han realizado varias acciones en un campo concreto.

Note: Esta tabla no se muestra para los campos que no se han transformado.

La información de la tabla se divide en dos o tres columnas:

- **Acción.** Nombre de la acción. Por ejemplo, Predictores continuos. Consulte el tema “Detalles de acción” para obtener más información.
- **Detalles.** La lista de procesos ejecutados. Por ejemplo, Transformar a unidades estándar.
- **Función.** Sólo se muestra para predictores creados y se muestra la combinación lineal de campos de entrada, por ejemplo, $0,06 * \text{edad} + 1,21 * \text{altura}$.

Detalles de acción

La vista vinculada Detalles de acción aparece cuando selecciona cualquier **Acción** subrayada en la vista principal Resumen de acciones y muestra los datos comunes y específicos de cada paso de procesamiento realizado; los detalles específicos de la acción se muestran primero.

En cada acción, se utiliza la descripción como título en la parte superior de la vista vinculada. Los detalles específicos de la acción se muestran bajo el título y pueden incluir detalles sobre el número de predictores derivados, reestructuración de campos, transformaciones de destinos, categorías fusionadas o reordenadas y predictores creados o excluidos.

A medida que se procesa cada acción, puede cambiar el número de predictores utilizados en el procesamiento, por ejemplo a medida que se excluyen o fusionan los predictores.

Note: Si se ha desactivado una acción o si no se ha especificado un destino, aparece un mensaje de error en lugar de los detalles de la acción cuando pulsa en la acción en la vista principal Resumen de acciones.

Existen nueve acciones disponibles; sin embargo, no todas están necesariamente activas para cada análisis.

Tabla Campos de texto

La tabla muestra el número de:

- Valores en blanco finales recortados.
- Predictores excluidos del análisis.

Tabla Predictores de fecha y hora

La tabla muestra el número de:

- Duraciones derivadas de los predictores de fecha y hora.
- Elementos de fecha y hora.
- Predictores derivados de fecha y hora, en total.

La fecha u hora de referencia se muestra como nota al pie si se han calculado algunas de las duraciones de fecha.

Tabla Cribado de predictores

La tabla muestra el número de los siguientes predictores excluidos del procesamiento:

- Constantes.
- Predictores con demasiados valores perdidos.
- Predictores con demasiados casos en una única categoría.
- Campos nominales (conjuntos) con demasiadas categorías.
- Predictores cribados, en total.

Tabla Comprobación de nivel de medición

La tabla muestra los números de reestructuración de campos, que se dividen en:

- Reestructuración de campos ordinales (conjuntos ordenados) como campos continuos.
- Reestructuración de campos continuos como campos ordinales.
- Reestructuración de números totales.

Si los campos de entrada (destinos o predictores) no eran continuos u ordinales, se muestra como nota al pie.

Tabla Valores atípicos

La tabla muestra cómo se han tratado los valores atípicos.

- El número de campos continuos donde se han encontrado y suprimido valores atípicos o el número de campos continuos donde se han encontrado valores atípicos y se han definido como perdidos, dependiendo de su configuración en el panel Preparar entradas y destino en la pestaña Configuración.
- Se ha excluido el número de campos continuos porque eran constantes, después del tratamiento de los valores atípicos.

Una nota al pie muestra el valor de corte atípico; mientras que se muestra otra nota al pie si no hay campos de entrada continuos (destino o predictores).

Tabla Valores perdidos

La tabla muestra el número de campos con valores perdidos sustituidos y desglosados en:

- Objetivo. Esta fila no se muestra si no se han especificado destinos.
- Predictores. Pueden desglosarse por el número de nominales (conjunto), ordinales (conjunto ordenado) y continuas.
- El número total de valores perdidos sustituidos.

Tabla Destino

La tabla muestra si se ha transformado el destino, que se muestra como:

- Transformación de Box-Cox a normalidad. Se desglosa a su vez en columnas que muestran los criterios especificados (media y la desviación estándar) y Lambda.
- Categorías de destino reordenadas para mejorar la estabilidad.

Tabla Predictores categóricos

La tabla muestra el número de predictores categóricos:

- Cuyas categorías se han reordenado de menor a mayor para mejorar la estabilidad.
- Características cuyas categorías se han fusionado para aumentar al máximo su asociación con el destino.
- Características cuyas categorías se han fusionado para tratar categorías dispersas.
- Características cuyas categorías se han excluido por su asociación baja con el destino.
- Características cuyas categorías se han excluido porque eran constantes después de la fusión.

Se muestra una nota al pie si no se han introducido predictores categóricos.

Tabla Predictores continuos

Hay dos tablas. La primera muestra uno de los siguientes números de transformaciones:

- Valores de predictores transformados a unidades estándar. Además, muestra el número de predictores transformados, la media especificada y la desviación estándar.
- Valores de predictores correlacionados a un rango común. Además, muestra el número de predictores transformados utilizando una transformación mínima-máxima, así como los valores mínimo y máximo especificados.
- Valores de predictores en intervalos y el número de predictores en intervalos.

La segunda tabla muestra los detalles de creación de espacio predictor, mostrados como el número de predictores:

- Construido.
- Características cuyas categorías se han excluido por su asociación baja con el destino.
- Características cuyas categorías se han excluido porque eran constantes después de la agrupación.
- Excluido por ser constante tras la construcción.

Se muestra una nota al pie si no se han introducido predictores continuos.

Generación de un nodo Derivar

Cuando genera un nodo Derivar, aplica la transformación de destino inversa al campo de puntuación. De forma predeterminada, el nodo introduce el nombre del campo de puntuación que debería producir un modelador automático (como Clasificador automático o Autonumérico) o el nodo Conjunto. Si se ha

transformado un destino (rango) de escala, el campo de puntuación se muestra en unidades transformadas; por ejemplo, $\log(\$)$ en lugar de $\$$. Para interpretar y utilizar los resultados, debe convertir el valor predicho a la escala original.

Note: Sólo puede generar un nodo Derivar si el nodo ADP contiene un análisis que vuelve a trazar un destino de rango (es decir, el cambio de escala Box-Cox está seleccionado en el panel Preparar entradas y objetivo). No puede generar un nodo Derivar si el destino no es un rango o si el cambio de escala Box-Cox no está seleccionado.

El nodo Derivar se crea en modo Múltiple y utiliza @FIELD en la expresión para que pueda agregar el destino transformado si fuera necesario. Por ejemplo, utilizando los siguientes detalles:

- Nombre de campo objetivo: respuesta
- Nombre de campo objetivo transformado: respuesta_transformada
- Nombre de campo de puntuación: \$XR-respuesta_transformada

El nodo Derivar crearía un nuevo campo: \$XR-respuesta_transformada_inversa.

Nota: Si no utiliza un nodo de modelado automático o Conjunto, necesitará modificar el nodo Derivar para transformar el campo de puntuación correcto de su modelo.

Destinos continuos normalizados

De forma predeterminada, si selecciona la casilla de verificación **Cambiar la escala de objetivo continuo con transformación de Box-Cox** en el panel Preparar entradas y objetivo, transforma el objetivo y crea un nuevo campo, será el destino de su generación de modelos. Por ejemplo, si su destino original era *respuesta*, el nuevo destino será *respuesta_transformada*; los modelos posteriores al nodo ADP seleccionarán este nuevo destino automáticamente.

Sin embargo, puede causar algunos problemas, dependiendo del destino original. Por ejemplo, si el destino era *Edad*, los valores del nuevo destino no serán *Años*, sino una versión transformada de *Años*. Significa que no puede ver los resultados e interpretarlos ya que no están en unidades reconocibles. En este caso puede aplicar una transformación inversa que convertirá sus unidades transformadas en lo que eran originalmente. Para ello:

1. Después de hacer pulsar **Analizar datos** para ejecutar el análisis ADP, seleccione *Nodo Derivar* en el menú *Generar*.
2. Coloque el nodo Derivar después del nugget en el lienzo del modelo.

El nodo Derivar restaurará el campo de puntuación a las dimensiones original para que la predicción esté en los valores originales de *Años*.

De forma predeterminada, el nodo Derivar transforma el campo de puntuación generado por un modelo de conjunto o de modelado automático. Si está creando un modelo individual, necesita modificar el nodo Derivar para derivarlo de su campo de puntuación real. Si desea evaluar su modelo, debe añadir el destino transformado a **Derivar de** en el nodo Derivar. Aplica la misma transformación inversa al destino y cualquier nodo Evaluación o Análisis posterior los utilizará correctamente siempre que los modifique para utilizar los nombres de campos en lugar de los metadatos.

Si también desea restaurar el nombre original, puede utilizar un nodo Filtrar para eliminar el campo de destino original si aún sigue allí y cambia el nombre del destino y los campos de puntuación.

Nodo Tipo

Las propiedades de campo se pueden especificar en un nodo de origen o en un nodo Tipo independiente. La funcionalidad es similar en ambos nodos. Se trata de las siguientes propiedades:

- **Campo:** pulse dos veces en cualquier nombre de campo para especificar etiquetas de valor y de campo para los datos de IBM SPSS Modeler. Por ejemplo, los metadatos de campo importados de IBM SPSS Statistics se pueden ver o modificar aquí. Del mismo modo, puede crear nuevas etiquetas para campos y sus valores. Las etiquetas que se especifican aquí se muestran en IBM SPSS Modeler dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.
- **Medición:** este es el nivel de medición que se usa para describir características de los datos en un campo determinado. Si se conocen todos los detalles de un campo, éste se denomina **completamente instanciado**. Si desea obtener más información, consulte “Niveles de medición” en la página 127.

Nota: El nivel de medición de un campo difiere del tipo de almacenamiento de un campo, lo cual indica si los datos están almacenados como cadenas, números enteros, números reales, fechas, horas o marcas de tiempo.

- **Valores:** esta columna le permite especificar las opciones para leer los valores de datos desde el conjunto de datos, o utilizar la opción **Especificar** para especificar los niveles de medición y los valores en un cuadro de diálogo separado. También puede optar por pasar los campos sin leer sus valores. Si desea obtener más información, consulte “Valores de campos” en la página 131.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Perdidos:** se utiliza para especificar cómo se tratan los valores perdidos de un campo. Si desea obtener más información, consulte “Definición de valores perdidos” en la página 135.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Comprobar:** en esta columna puede configurar opciones que garanticen que los valores de campo se ajusten a los valores o rangos especificados. Si desea obtener más información, consulte “Comprobación de valores de tipo” en la página 136.

Nota: No es posible modificar la celda de esta columna si la entrada de **Campo** correspondiente contiene una lista.

- **Rol:** se usa para indicar a los nodos de modelado si los campos serán **Entrada** (campos predictores) o **Objetivo** (campos predichos) para un proceso de aprendizaje automático. **Ambas** y **Ninguna** también son roles disponibles, junto con **Partición**, que indica un campo utilizado para crear particiones de registros en muestras independientes para entrenamiento, comprobación y validación. El valor **Split** especifica que se generarán modelos diferentes para cada valor posible del campo. Si desea obtener más información, consulte “Definición del rol de campo” en la página 136.

Se pueden especificar otras opciones con la ventana del nodo Tipo:

- Con el botón de menú de herramientas puede elegir **Omitir campos exclusivos** una vez instanciado un nodo Tipo (bien a través de las especificaciones, la lectura de los valores o la ejecución de la ruta). Al omitir los campos exclusivos, se omiten automáticamente los campos con un solo valor.
- Con el botón de menú de herramientas puede elegir **Omitir conjuntos grandes** una vez instanciado un nodo Tipo. Al omitir conjuntos grandes, se omiten automáticamente los conjuntos con un número elevado de miembros.
- Con el botón de menú de herramientas puede elegir **Convertir enteros continuos a ordinal** una vez instanciado un nodo Tipo. Consulte el tema “Conversión de datos continuos” en la página 130 para obtener más información.
- Con el botón de menú de herramientas puede generar un nodo Filtrar para descartar los campos seleccionados.
- Con los botones de gafas puede establecer los valores predeterminados de todos los campos en Leer o Pasar. La pestaña Tipos del nodo de origen pasa los campos de forma predeterminada, mientras que el propio nodo Tipo lee los valores predeterminados.

- Con el botón **Borrar valores** puede borrar los cambios de los valores del campo realizados en este nodo (valores no heredados) y volver a leer valores de las operaciones anteriores a la ruta. Esta opción es útil para restablecer los cambios que pueda haber realizado a los campos específicos anteriores en la ruta.
- Con el botón **Borrar todos los valores** puede restablecer los valores de **todos** los campos que se leen en el nodo. Esta opción establece de forma eficaz la columna **Valores** en **Leer** para todos los campos. También resulta útil para restablecer los valores de todos los campos y volver a leer los valores y los tipos de las operaciones anteriores en la ruta.
- En el menú contextual puede elegir **Copiar** los atributos de un campo a otro. Consulte el tema “Copia de atributos de tipo” en la página 137 para obtener más información.
- Con la opción **Ver configuración de campos no utilizados** puede ver la configuración de los tipos que ya no están presentes en los datos o que estuvieron conectados alguna vez a este nodo Tipo. Esta opción es útil al reutilizar un nodo Tipo para conjuntos de datos que han cambiado.

Niveles de medición

El nivel de medición (antes conocido como "tipo de datos" o "tipo de uso") describe el uso de los campos de datos en IBM SPSS Modeler. El nivel de medición puede especificarse en la pestaña Tipos de un nodo de origen o Tipo. Por ejemplo, puede que desee definir el nivel de medición para un campo de enteros con valores de 1 y 0 para una *Marca*. Esto suele indicar que 1=*Verdadero* y 0=*Falso*.

Almacenamiento frente a medición. Tenga en cuenta que el nivel de medición de un campo difiere de su tipo de almacenamiento, que indica si los datos están almacenados como una cadena, un número entero, un número real, una fecha, una hora o una marca de tiempo. Los tipos de datos se pueden modificar en cualquier punto de una ruta mediante un nodo Tipo, mientras que el almacenamiento ha de definirse en el origen, al leer los datos en IBM SPSS Modeler (si bien posteriormente puede cambiarse a través de una función de conversión). Consulte el tema “Configuración de almacenamiento de campos y formato” en la página 9 para obtener más información.

Algunos nodos de modelado indican los tipos de nivel de medición permitidos para sus campos de entrada y objetivo por iconos en su pestaña Campos.

Iconos de nivel de medición

Tabla 19. Iconos de nivel de medición










Icono	Nivel de medición
	Valor predeterminado
	Continuo
	Categorico
	Distintivo
	Nominal
	Ordinal
	Sin tipo
	Colección

Tabla 19. Iconos de nivel de medición (continuación)

Icono	Nivel de medición
	Geoespacial

Los siguientes niveles de medición están disponibles:

- **Valor predeterminado:** Los datos cuyo tipo de almacenamiento y valores son desconocidos (por ejemplo, porque todavía no se han leído) se muestran como <Valor predeterminado>.
- **Continuo:** Se usa para describir valores numéricos, como el rango de 0–100 o de 0,75–1,25. Un valor continuo pueden ser un número entero, un número real o la fecha/hora.
- **Catagórico:** Se utiliza en el caso de los valores de cadenas, cuando se desconoce un número exacto de valores distintos. Se trata de un tipo de datos **sin instanciar**, lo que significa que toda la información posible acerca del almacenamiento y utilización de los datos aún no se conoce. Una vez leídos los datos, el nivel de medición será *Marca*, *Nominal* o *Sin tipo*, dependiendo del número máximo de campos nominales especificado en el cuadro de diálogo Propiedades de la ruta.
- **Marca** Se utiliza para datos con dos valores distintos que indican la presencia o ausencia de un rasgo, como true y false, Sí y No o 0 y 1. Los valores utilizados pueden variar, pero uno siempre debe designarse como el valor verdadero y el otro como el valor falso. Los datos pueden representarse como texto, número entero, número real, fecha, hora o marca de tiempo.
- **Nominal:** Se usa para describir datos con varios valores distintos, cada uno tratado como un miembro de un conjunto (por ejemplo pequeño/medio/grande). Los datos nominales pueden tener cualquier almacenamiento: numérico, cadena o fecha/hora. Tenga en cuenta que el establecimiento del nivel de medida a *Nominal* no cambia de forma automática los valores a almacenamiento de cadena.
- **Ordinal:** Se usa para describir datos con varios valores diferentes que tienen un orden inherente. Por ejemplo, las categorías salariales o las clasificaciones de satisfacción pueden escribirse como datos ordinales. El orden se define a través del orden de clasificación natural de los elementos de datos. Por ejemplo, 1, 3, 5 es el orden de clasificación predeterminado de un conjunto de números enteros, mientras que ALTO, BAJO, NORMAL (orden alfabético ascendente) es el orden de un conjunto de cadenas. El nivel de medición ordinal le permite definir un conjunto de datos catagóricos como datos ordinales para la visualización, generación de modelos y la exportación a otras aplicaciones (como IBM SPSS Statistics), que reconoce los datos ordinales como un tipo distinto. Puede utilizar un campo ordinal en cualquier punto donde pueda utilizarse un campo nominal. Además, los campos de cualquier tipo de almacenamiento (real, entero, cadena, fecha, hora, etc.) pueden definirse como ordinal.
- **Sin tipo:** Se utiliza para datos que no se ajustan a ninguno de los tipos anteriores, para campos con un único valor, o para datos nominales donde el conjunto tiene más miembros que el máximo definido. Resulta también útil con los casos en los que, de lo contrario, el nivel de medición sería un conjunto con demasiados miembros (como un número de cuenta). Cuando seleccione **Sin tipo** para un campo, el rol se establece automáticamente en **Ninguno**, con el **ID de registro** como la única alternativa. El tamaño máximo predeterminado de los conjuntos es de 250 valores exclusivos. Este número puede ajustarse o desactivarse en la pestaña Opciones del cuadro de diálogo Propiedades de la ruta, que puede abrirse desde el menú Herramientas.
- **Recopilación:** Se utiliza para identificar los datos que no son geoespaciales y que se registran en una lista. Una recopilación es, en la práctica, un campo de lista con una profundidad de cero, en la que los elementos de dicha lista tienen uno de los otros niveles de medición.
Para obtener más información acerca de las listas, consulte el tema Almacenamiento de lista y niveles de medición asociados en la sección Nodos de origen de la publicación Nodos de origen, proceso y resultado de SPSS Modeler.
- **Geoespacial:** Se utiliza con el tipo de almacenamiento de lista para identificar los datos geoespaciales. Las listas pueden ser el campo Lista de enteros o Lista de reales con una profundidad de lista entre cero y dos, inclusive.
Para obtener más información, consulte el tema Subniveles de mediciones geoespaciales, en la sección Nodo Tipo, de la publicación Nodos de origen, proceso y resultado de SPSS Modeler.

Puede especificar manualmente los niveles de medición o bien dejar que el software lea los datos y determine el nivel de medición según los valores que lee.

Asimismo, si tiene varios campos de datos continuos que deberían tratarse como datos categóricos, puede seleccionar una opción para convertirlos. Consulte el tema “Conversión de datos continuos” en la página 130 para obtener más información.

Para utilizar la definición automática de tipos

1. En el nodo Tipo o en la pestaña Tipos de un nodo de origen, establezca la columna *Valores* en <Leer> para los campos deseados. Así se pondrán los metadatos a disposición de todos los nodos que se encuentran a continuación en la ruta. Puede establecer rápidamente todos los campos en <Leer> o <Pasasr> usando los botones de gafas del cuadro de diálogo.
2. Pulse en **Leer valores** para leer inmediatamente los valores del origen de datos.

Para establecer manualmente el nivel de medición de un campo

1. Seleccione un campo de la tabla.
2. En la lista desplegable de la columna *Medición* seleccione un nivel de medición para el campo.
3. Si lo prefiere, puede pulsar Ctrl-A o Ctrl mientras pulsa en un campo para seleccionar varios campos antes de utilizar la lista desplegable con objeto de seleccionar un nivel de medición.

Subniveles de medición geoespacial

El nivel de medición geoespacial, que se utiliza con el tipo de almacenamiento de lista, tiene seis subniveles que se utilizan para identificar diferentes tipos de datos geoespaciales.

- **Punto:** Identifica una ubicación específica. Por ejemplo, el centro de una ciudad.
- **Polígono:** Una serie de puntos que identifica un límite individual de una región y su ubicación. Por ejemplo, un condado.
- **Cadena de líneas:** La Cadena de líneas, a la que también se hace referencia como Polilínea o simplemente Línea, es una serie de puntos que identifican la ruta de una línea. Por ejemplo, una Cadena de líneas puede ser un elemento fijo, tal como una carretera, un río o una vía férrea, o puede ser la ruta de algo que se mueve, tal como la ruta de vuelo de un avión o la ruta de un barco.
- **Multipunto:** Se utiliza cuando cada fila de sus datos contiene varios puntos por región. Por ejemplo, si cada fila representa una calle de una ciudad, se pueden utilizar los múltiples puntos de cada calle para identificar cada farola de la calle.
- **Multipolígono:** Se utiliza cuando cada fila de sus datos contiene varios polígonos. Por ejemplo, si cada fila representa el contorno de un condado, los Estados Unidos se pueden registrar como varios polígonos para identificar las diferentes áreas, tales como el continente, Alaska y Hawaii.
- **Cadena multilínea:** Se utiliza cuando cada fila de sus datos contiene varias líneas. Dado que no se pueden ramificar las líneas, puede utilizar la Cadena multilínea para identificar un grupo de líneas. Por ejemplo, datos como los canales navegables o la red ferroviaria de cada país.

Estos subniveles de medición se utilizan con el tipo de almacenamiento de lista. Si desea obtener más información, consulte “Almacenamiento de listas y niveles de medición asociados” en la página 11.

Restricciones

Debe tener en cuenta algunas restricciones cuando utilice los datos geoespaciales.







- El sistema de coordenadas puede afectar el formato de los datos. Por ejemplo, un sistema de coordenadas proyectado utiliza los valores de coordenadas x, y, z (cuando es necesario), mientras que un sistema de coordenadas geográfico utiliza los valores de longitud, latitud y (cuando es necesario) un valor para la altitud o la profundidad.

Para obtener más información acerca de los sistemas de coordenadas, consulte el tema Configuración de opciones geoespaciales para corrientes en la sección Trabajar con corrientes de la publicación SPSS Modeler Guía del usuario.

- Una Cadena de líneas no se puede cruzar sobre sí misma.
- Un Polígono no se cierra a sí mismo. En cada Polígono debe asegurarse de que define el primer punto y el último punto como el mismo.
- La dirección de los datos de un MultiPolígono es importante. La dirección de las agujas del reloj indica una forma sólida y la dirección contraria a las agujas del reloj indica un agujero. Por ejemplo, si registra un área de una comarca que contiene lagos, el límite del área de tierra se puede registrar en la dirección de las agujas del reloj y la forma del lago en la dirección contraria a las agujas del reloj.
- Un Polígono no se puede interseccionar consigo mismo. Un ejemplo de esta intersección será si intenta trazar el límite del polígono como una línea continua con el formato de la Figura 8.
- Los Multipolígonos no se pueden solapar entre sí.

Iconos de subnivel de medición geoespacial

Tabla 20. Iconos de subnivel de medición geoespacial

Icono	Nivel de medición
	Punto
	Polígono
	Cadena de líneas
	Multipunto
	Multipolígono
	Cadena multilínea

Conversión de datos continuos

Si se trata un tipo de datos categóricos como continuos, la calidad de un modelo puede verse muy afectada, especialmente si se trata del campo de destino; por ejemplo, podría producirse un modelo de regresión en lugar de un modelo binario. Para evitarlo, puede convertir rangos enteros en tipos categóricos como *Ordinal* o *Marca*.

1. Desde el botón Operaciones y menú Generar (con el símbolo de herramienta), seleccione **Convertir enteros continuos a ordinal**. Aparecerá el cuadro de diálogo Valores de conversión.
2. Especifique el tamaño del rango que se convertirá automáticamente; esto es aplicable a cualquier rango inferior e igual al tamaño que introduzca.
3. Pulse en **Aceptar**. Los rangos afectados se convierten a *Marca* u *Ordinal* y se muestran en la pestaña Tipos del nodo Tipo.

Resultados de la conversión

- Cuando un campo *Continuo* con almacenamiento entero cambia a *Ordinal*, los valores inferior y superior se amplían para incluir todos los valores enteros desde el valor inferior hasta el valor superior. Por ejemplo, si el rango es 1, 5, el conjunto de valores es 1, 2, 3, 4, 5.
- Cuando el campo *Continuo* cambia a *Marca*, los valores inferior y superior se convierten en los valores false y true del campo Marca.

¿Qué es la instanciación?

La **instanciación** es el proceso de leer o especificar información, como los valores y el tipo de almacenamiento de un campo de datos. Para optimizar los recursos del sistema, la instanciación es un

proceso dirigido por el usuario, que le indica al software que lea los valores especificando opciones en la pestaña Tipos de un nodo de origen o ejecutando los datos a través de un nodo Tipo.

- Los datos con tipos desconocidos también se denominan **sin instanciar**. Los datos cuyos valores y tipo de almacenamiento son desconocidos se muestran en la columna *Medición* de la pestaña Tipos como **<Valor predeterminado>**.
- Cuando se dispone de alguna información acerca del almacenamiento de un campo, como cadena o numérico, los datos se denominan **parcialmente instanciados**. **Categorico** o **Continuo** son niveles de medición parcialmente instanciados. Por ejemplo, **Categorico** especifica que el campo es simbólico, pero no se conoce si es nominal, ordinal o marca.
- Cuando se conocen todos los detalles de un tipo, incluidos los valores, se muestra un nivel de medición (nominal, ordinal, marca o continuo) **completamente instanciado** en esta columna. *Note*: el tipo *Continuo* se utiliza tanto para los campos de datos parcialmente instanciados como para los campos de datos completamente instanciados. Los datos continuos pueden ser números enteros o reales.

Durante la ejecución de una ruta de datos con un nodo Tipo, los tipos sin instanciar se vuelven parcialmente instanciados, en función de los valores de datos iniciales. Una vez que todos los datos pasan por el nodo, se vuelven completamente instanciados a menos que los valores estuvieran establecidos en **<Pasar>**. Si la ejecución se interrumpe, los datos permanecerán como parcialmente instanciados. Una vez instanciada la pestaña Tipos, los valores de un campo se vuelven estáticos en ese punto de la ruta. Esto significa que cualquier cambio anterior a la ruta no afectará a los valores de un campo determinado, incluso si se vuelve a ejecutar la ruta. Para cambiar o actualizar los valores en función de los nuevos datos o manipulaciones añadidas, debe editarlos en la propia pestaña Tipos o establecer el valor de un campo en **<Leer>** o **<Leer +>**.

Cuándo se puede crear una instancia

Normalmente, si el conjunto de datos no es muy grande y no tiene intención de añadir más adelante ningún campo a la ruta, la instanciación en el nodo de origen es el método más conveniente. Sin embargo, la instanciación en un nodo Tipo independiente resulta útil cuando:

- El conjunto de datos es grande y la ruta filtra un subconjunto anterior al nodo Tipo.
- Los datos se han filtrado en la ruta.
- Los datos se han fusionado o añadido en la ruta.
- Los campos de datos nuevos se derivan durante el procesamiento.

Valores de campos

Utilizando la columna *Valores* de la pestaña Tipos, se pueden leer automáticamente los valores desde los datos o especificar niveles de medición y valores en un cuadro de diálogo independiente.

Las opciones disponibles en la lista desplegable Valores proporcionan instrucciones para la definición automática de tipos, como se muestra en la siguiente tabla.

Tabla 21. Instrucciones para la definición automática de tipos

Opción	Función
<Leer>	Los datos se leerán cuando el nodo se ejecute.
<Leer+>	Los datos se leerán y se añadirán a los datos actuales (si existe alguno).
<Pasar>	No se lee ningún dato.
<Actual>	Mantiene los valores de datos actuales.
Especificar...	Se ejecuta un cuadro de diálogo independiente en el que especificar los valores y las opciones de nivel de medición.

Al ejecutar un nodo Tipo o pulsar en **Leer valores**, los valores se leerán y se escribirán automáticamente desde el origen de datos en función de la selección. Estos valores también se pueden especificar manualmente con la opción Especificar o al pulsar dos veces en una casilla de la columna *Campo*.

Una vez realizados los cambios a los campos en el nodo Tipo, puede restablecer la información de los valores con los siguientes botones de la barra de herramientas del cuadro de diálogo.

- Con el botón **Borrar valores** puede borrar los cambios de los valores del campo realizados en este nodo (valores no heredados) y volver a leer valores de las operaciones anteriores a la ruta. Esta opción es útil para restablecer los cambios que pueda haber realizado a los campos específicos anteriores en la ruta.
- Con el botón **Borrar todos los valores** puede restablecer los valores de **todos** los campos que se leen en el nodo. Esta opción establece de forma eficaz la columna *Valores* en **Leer** para todos los campos. También resulta útil para restablecer los valores de todos los campos y volver a leer los valores y los niveles de medición de las operaciones anteriores en la ruta.

Utilización del cuadro de diálogo de valores

Si pulsa en la columna **Valores** o **Perdidos** de la pestaña Tipos, aparecerá una lista desplegable de valores predefinidos. Si selecciona la opción **Especificar...** en la lista, se abre un recuadro de diálogo independiente donde puede establecer opciones para leer, especificar, etiquetar y gestionar valores para el campo seleccionado.

Muchos de los controles son comunes a todos los tipos de datos. Aquí se describen estos controles comunes.

Medición: Muestra el nivel de medición seleccionado actualmente. Puede cambiar la definición para reflejar la forma en la que piensa utilizar los datos. Por ejemplo, si un campo llamado `día_de_la_semana` contiene números que representan días individuales, es posible que desee cambiar este tipo a datos nominales para crear un nodo Distribución que examine cada categoría individualmente.

Almacenamiento: Muestra el tipo de almacenamiento, si se conoce. Los tipos de almacenamiento no se ven afectados por el nivel de medición que seleccione. Para alterar el tipo de almacenamiento, puede utilizar la pestaña Datos de los nodos de origen Archivo fijo y Archivo variable, o una función de conversión en un nodo Rellenar.

Campo de modelo: Para campos generados como resultado de puntuar un nugget de modelo, también se pueden visualizar los detalles del campo de modelo. Incluyen el nombre del campo objetivo así como el rol del campo de modelado (ya sea un valor predicho, probabilidad, propensión, etc.).

Valores: Permite seleccionar un método para determinar los valores para el campo seleccionado. Las selecciones que realice aquí omiten cualquier selección efectuada anteriormente en la columna **Valores** del cuadro de diálogo del nodo Tipo. Las opciones para leer los valores son las siguientes:

- **Leer de datos:** Seleccione esta opción para leer los valores cuando el nodo se ejecuta. Esta opción es la misma que **<Leer>**.
- **Pasar:** Omite la lectura de los datos del campo actual. Esta opción es la misma que **<Passar>**.
- **Especificar valores y etiquetas:** Estas opciones se utilizan para especificar los valores y las etiquetas del campo seleccionado. Utilizada junto con la comprobación de valor, esta opción permite especificar valores en función de sus conocimientos del campo actual. Esta opción activa controles exclusivos para cada tipo de campo. Las opciones de los valores y las etiquetas se tratan individualmente en los temas posteriores.

Nota: No se pueden especificar valores o etiquetas para un campo cuyo nivel de medición sea Sin tipo o **<Valor predeterminado>**.

- **Extender valores a partir de los datos:** Seleccione esta opción para añadir los datos actuales con los valores que especifique aquí. Por ejemplo, si `campo_1` tiene un rango desde (0,10) e introduce un rango

de valores desde (8,16), el rango se extiende al añadir el 16, sin eliminar el mínimo original. El nuevo rango sería (0,16). Al seleccionar esta opción, se establece automáticamente la opción de definición automática de tipos en <Leer+>.

Longitud máxima de lista: Solo está disponible para los datos con un nivel de medición de Geoespacial o Colección. Establezca la longitud máxima de la lista especificando el número de elementos que puede contener la lista.

Comprobar valores: Permite seleccionar un método para forzar los valores de modo que se ajusten a los valores continuos, de distintivo, de código o nominales especificados. Esta opción corresponde a la columna **Comprobar** del cuadro de diálogo del nodo Tipo y la configuración aquí realizada omite la del cuadro de diálogo. Junto con la opción **Especificar valores y etiquetas**, la comprobación de valores permite ajustar los valores de los datos según los valores previstos. Por ejemplo, si se especifican valores como 1, 0 y, a continuación, se utiliza la opción **Descartar**, puede descartar todos los registros con valores distintos de 1 o 0.

Definir vacíos: Seleccione esta opción para activar los siguientes controles que permiten declarar los valores que faltan o los espacios en blanco en sus datos.

- **Valores perdidos:** Utilice esta tabla para definir valores específicos (por ejemplo, 99 o 0) como espacios en blanco. El valor debe ser adecuado al tipo de almacenamiento del campo.
- **Rango:** Se utiliza para especificar un rango de valores perdidos, por ejemplo, las edades 1–17 o mayores de 65. Si un valor límite se deja en blanco, el rango no tendrá límite. Por ejemplo, si se especifica un límite inferior a 100 sin ningún límite superior, todos los valores superiores o iguales a 100 se definirán como perdidos. Los valores límite son inclusivos. Por ejemplo, un rango con un límite inferior a 5 y un límite superior de, 10 incluye 5 y 10 en la definición de rango. Se puede definir un rango de valores perdidos para cualquier tipo de almacenamiento, incluidas la fecha/hora y la cadena (en cuyo caso se utilizará el orden de clasificación alfabético para determinar si un valor está dentro del rango).
- **Espacio en blanco/nulo:** También puede especificar valores nulos del sistema, que se muestran en los datos como \$null\$, y espacios en blanco, que son valores de cadenas sin caracteres visibles.

Nota: El nodo Tipo también trata cadenas vacías como espacios en blanco con fines analíticos, si bien internamente éstas se almacenan de forma distinta y posiblemente deban gestionarse de otra manera en determinados casos.

Nota: Para codificar espacios en blanco como no definidos o como \$null\$, se debe utilizar el nodo Rellenar.

Descripción: Utilice este recuadro de texto para especificar una etiqueta de campo. Estas etiquetas aparecen en diferentes ubicaciones, tales como gráficos, tablas, salida y navegadores de modelos, en función de las selecciones que realice en el recuadro de diálogo Propiedades de ruta.

Especificación de valores y etiquetas para datos continuos

El nivel de medición *Continuo* se utiliza para campos numéricos. Existen tres tipos de almacenamiento para datos continuos:

- Real
- Entero
- Fecha/hora

Se utiliza el mismo cuadro de diálogo para editar todos los campos continuos; solamente se muestra el tipo de almacenamiento como referencia.

Especificación de valores

Los siguientes controles son exclusivos de los campos continuos y se utilizan para especificar un rango de valores:

Inferior. Especifica un límite inferior para el rango de valores.

Superior. Especifica un límite superior para el rango de valores.

Especificación de etiquetas

Puede especificar etiquetas para cualquier valor de un campo de rango. Pulse en el botón **Etiquetas** para abrir un cuadro de diálogo independiente con el fin de especificar las etiquetas de valores.

Subcuadro de diálogo Valores y etiquetas: Al pulsar en la opción **Etiquetas** del cuadro de diálogo Valores de un campo de rango, se abre un cuadro de diálogo nuevo en el que pueden especificar etiquetas para cualquier valor del rango.

Puede utilizar las columnas *Valores* y *Etiquetas* de esta tabla para definir pares de etiquetas y valores. A continuación se muestran los pares definidos actualmente. Puede añadir pares de etiquetas nuevos pulsando en una casilla vacía e introduciendo un valor y su correspondiente etiqueta. *Note:* al añadir pares de valor/valor-etiqueta a esta tabla no se añade ningún valor nuevo al campo. En su lugar, simplemente se crean metadatos para el valor del campo.

Las etiquetas que se especifican en el nodo Tipo se muestran en muchos lugares (como Información sobre herramientas, etiquetas de resultados, etc.) dependiendo de las selecciones efectuadas en el cuadro de diálogo de propiedades de la ruta.

Especificación de valores y etiquetas para datos nominales y ordinales

Los niveles de medición nominales (conjunto) y ordinales (conjunto ordenado) indican que los valores de datos se utilizan de forma discreta como miembros de un conjunto. Los tipos de almacenamiento de un conjunto pueden ser una cadena, un número entero, un número real o la fecha/hora.

Los siguientes controles son exclusivos de los campos nominales y ordinales y se utilizan para especificar valores y etiquetas:

Valores. La columna *Valores* de la tabla permite especificar valores en función del conocimiento del campo actual. Mediante esta tabla, puede introducir valores esperados para el campo y comprobar la conformidad del conjunto de datos para estos valores utilizando la lista desplegable Comprobar valores. Puede modificar valores existentes y reordenar o eliminar valores con los botones de flecha o de eliminación.

Etiquetas La columna *Etiquetas* permite especificar etiquetas para cada valor del conjunto. Estas etiquetas aparecen en distintas ubicaciones (gráficos, tablas, resultados y exploradores de modelos), dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.

Especificación de valores para una marca

Los campos de marca se utilizan para mostrar datos que tienen dos valores distintos. Los tipos de almacenamiento de las marcas pueden ser una cadena, un número entero, un número real o la fecha/hora.

Verdadero. Permite especificar un valor de marca para el campo cuando se cumple la condición.

Falso. Permite especificar un valor de marca para el campo cuando no se cumple la condición.

Etiquetas Permite especificar etiquetas para cada valor del campo de marcas. Estas etiquetas aparecen en distintas ubicaciones (gráficos, tablas, resultados y exploradores de modelos), dependiendo de las selecciones realizadas en el cuadro de diálogo de propiedades de la ruta.

Especificación de valores para Data Collection

Los campos de colección se utilizan para mostrar datos no geoespaciales que se encuentran en una lista.

El único elemento que puede establecer para el nivel de **Medición** de colección es la **Medición de lista**. De forma predeterminada, esta medida está establecida en Sin tipo, pero puede seleccionar otro valor para establecer el nivel de medición de los elementos dentro de la lista. Puede seleccionar una de las opciones siguientes:

- Sin tipo
- Continuo
- Nominal
- Ordinal
- Distintivo

Especificación de valores para los datos geoespaciales

Los campos geoespaciales se utilizan para mostrar datos geoespaciales en una lista.

En el nivel de **Medición** Geoespacial, puede seleccionar las opciones siguientes para establecer el nivel de medición de los elementos contenidos en la lista:

Tipo: Seleccione el subnivel de medición del campo geoespacial. Los subniveles disponibles se determinan según la profundidad del campo de lista: Punto (profundidad cero), Cadena de líneas (profundidad uno) y Polígono (profundidad uno).

Para obtener más información acerca de los subniveles, consulte “Subniveles de medición geoespacial” en la página 129.

Para obtener más información acerca de las profundidades de las listas, consulte “Almacenamiento de listas y niveles de medición asociados” en la página 11.

Sistema de coordenadas: Esta opción solo está disponible si ha modificado el nivel de medición desde un nivel no geoespacial al nivel geoespacial. Para aplicar un sistema de coordenadas a sus datos geoespaciales, seleccione este recuadro de selección. De forma predeterminada, se muestra el sistema de coordenadas en el panel **Herramientas > Propiedades de ruta > Opciones > Geoespacial**. Para utilizar un sistema de coordenadas diferente, pulse el botón **Cambiar** para visualizar el recuadro de diálogo Seleccionar sistema de coordenadas y seleccione el sistema que necesite.

Para obtener más información acerca de los sistemas de coordenadas, consulte el tema Configuración de opciones geoespaciales para corrientes en la sección Trabajar con corrientes de la publicación SPSS Modeler Guía del usuario.

Definición de valores perdidos

La columna **Missing** de la pestaña Tipos indica si se ha definido la gestión de valores perdidos para un campo. Los ajustes posibles son:

On (*). Indica que se ha definido la gestión de los valores perdidos para este campo. Puede hacerse mediante un nodo Rellenar posterior o mediante una especificación explícita utilizando la opción Specify.

Off. Este campo no tiene definida la gestión de valores perdidos.

Especificar. Seleccione esta opción para abrir un cuadro de diálogo donde puede declarar que los valores explícitos se consideren como valores perdidos para este campo.

Comprobación de valores de tipo

Al activar la opción *Comprobar* en cada campo, se examinan todos los valores de los campos para determinar si se ajustan a la configuración de tipos actual o a los valores especificados en el cuadro de diálogo *Especificar valores*. Esta opción es útil para limpiar los conjuntos de datos y reducir el tamaño de un conjunto de datos en una sola operación.

La configuración de la columna *Comprobar* del cuadro de diálogo del nodo *Tipo* determina lo que ocurre si se descubre un valor que se encuentre fuera de los límites del tipo. Para cambiar la configuración de *Comprobar* en un campo, utilice la lista desplegable de ese campo en la columna *Comprobar*. Para establecer la configuración de *Comprobar* para todos los campos, pulse en la columna *Campo* y pulse *Ctrl-A*. A continuación, utilice la lista desplegable de cualquier campo de la columna *Comprobar*.

Éstas son las configuraciones de *Comprobar* disponibles:

Ninguno. Los valores pasan sin comprobación. Ésta es el ajuste predeterminado.

Anular. Permite cambiar los valores que se encuentran fuera de los límites del valor nulo del sistema (\$null\$).

Forzar. Los campos cuyos niveles de medición están completamente instanciados se comprueban para determinar los valores que están fuera de los rangos especificados. Los valores no especificados se convertirán en un valor válido para este nivel de medición según las siguientes reglas:

- En el caso de las marcas, cualquier valor distinto de verdadero y falso se convierte en el valor falso.
- En el caso de los conjuntos (nominales u ordinales), cualquier valor desconocido se convierte en el primer miembro de los valores del conjunto.
- Los números que superen el límite superior de un rango se sustituyen por el límite superior.
- Los números que estén por debajo del límite inferior de un rango se sustituyen por el límite inferior.
- Los valores nulos de un rango toman el valor del punto medio de ese rango.

Descartar. Cuando se encuentran valores no válidos, el registro entero se descarta.

Avisar. El número de elementos no válidos se cuenta y se indica en el cuadro de diálogo de propiedades de la ruta tras leer todos los datos.

Cancelar. El primer valor no válido que se encuentra detiene la ejecución de la ruta. Se informa del error en el cuadro de diálogo de propiedades de la ruta.

Definición del rol de campo

El rol de un campo especifica cómo se utiliza en la generación de modelos; por ejemplo, si un campo es de entrada o si es un campo objetivo (lo que se desea predecir).

Note: Los roles *Partición*, *Frecuencia* e *ID* de registro pueden aplicarse por separado a un único campo.

Éstos son los roles disponibles:

Entrada. El campo se utiliza como una entrada en el aprendizaje de las máquinas (un campo predictor).

Objetivo. El campo se utiliza como un campo de salida o un campo objetivo para el aprendizaje de las máquinas (uno de los campos que el modelo intentará predecir).

Ambas. El campo se utilizará como campo de entrada y de salida por el nodo *Apriori*. El resto de los nodos de modelado omiten el campo.

Ninguno. El aprendizaje de las máquinas omitirá el campo. Los campos cuyo nivel de medición se haya definido a **Sin tipo** se establecen automáticamente en **Ninguna** en la columna *Rol*.

Partición. Indica un campo utilizado para dividir los datos en muestras independientes para entrenamiento, comprobación y validación (opcional). El campo debe ser un tipo de conjunto instanciado con dos o más valores positivos (tal y como se define en el cuadro de diálogo Valores de campos). El primer valor representa la muestra de entrenamiento, el segundo representa la muestra de comprobación y el tercero (si está presente) representa la muestra de validación. Cualquier valor adicional se omite y los campos de marcas no se pueden usar. Tenga en cuenta que para utilizar la partición en un análisis, ésta debe estar activada en la pestaña Opciones de modelo en el nodo de análisis o de generación de modelos adecuado. Los registros con valores nulos para el campo de partición se excluyen del análisis cuando la partición está activada. Si se han definido varios campos de partición en la ruta, se deberá especificar un campo de partición simple en la pestaña Campos en todos los nodos de modelado correspondientes. Si no existe ningún campo apropiado en los datos, puede crear uno usando un nodo Partición o un nodo Derivar. Consulte el tema “Nodo Partición” en la página 165 para obtener más información.

Dividir. (Campos nominales, ordinales o marca únicamente) Especifica que se creará un modelo para cada valor posible del campo.

Frecuencia. (Campos numéricos únicamente) La configuración de este rol permite que se pueda utilizar el valor del campo como un factor de ponderación de frecuencia para el registro. Esta característica es admitida únicamente por los modelos Árbol C&, CHAID, QUEST y lineales; el resto de los nodos ignoran este rol. La ponderación de frecuencias está activada por la opción **Utilizar ponderaciones de frecuencia** en la pestaña Campos de los nodos de modelado que admiten la característica.

ID de registro. El campo se utilizará como el identificador de registros exclusivo. La mayoría de los nodos ignoran esta característica; sin embargo es compatible con los modelos lineales y es necesaria para los nodos de minería interna de bases de datos de IBM Netezza.

Copia de atributos de tipo

Puede copiar con facilidad los atributos de un tipo (como los valores, las opciones de comprobación y los valores perdidos) desde un campo a otro.

1. Pulse con el botón derecho en el campo cuyos atributos desea copiar.
2. En el menú contextual, seleccione **Copiar**.
3. Pulse con el botón derecho en los campos cuyos atributos desea cambiar.
4. En el menú contextual, seleccione **Pegado especial**. *Nota:* puede seleccionar varios campos manteniendo pulsada la tecla Ctrl mientras pulsa en un campo o con la opción **Seleccionar campos** del menú contextual.

Se abrirá un cuadro de diálogo nuevo desde el que puede seleccionar los atributos específicos que desea pegar. Si pega los datos en varios campos, las opciones que seleccione aquí se aplicarán en todos los campos objetivo.

Pegar los siguientes atributos. Elija un elemento de la lista que aparece a continuación para pegar atributos de un campo a otro.

- **Tipo.** Seleccione para pegar el nivel de medición.
- **Valores.** Seleccione esta opción para pegar los valores de campo.
- **Perdidos.** Seleccione esta opción para pegar los valores perdidos.
- **Comprobar.** Seleccione esta opción para pegar las opciones de comprobación de valor.
- **Rol.** Seleccione esta opción para pegar el rol de un campo.

Pestaña Configuración de Formato de campo

La pestaña Formato de los nodos Tabla y Tipo muestra una lista de campos actuales o no utilizados así como opciones de formato para cada campo. A continuación se muestra una descripción de cada columna de la tabla de formato de campo:

Campo. Muestra el nombre del campo seleccionado.

Formato. Al pulsar dos veces en una casilla de esta columna, puede especificar el formato de los campos de forma individual con el cuadro de diálogo que se abre. Consulte el tema “Opciones de configuración de Formato de campo” para obtener más información. El formato especificado aquí altera temporalmente el especificado en las propiedades de ruta globales.

Nota: los nodos Exportar Estadísticas y Resultados de Estadísticas exportan archivos *.sav* que incluyen formatos por campos en sus metadatos. Si se especifica un formato por campos que no sea compatible con el formato de archivo IBM SPSS Statistics *.sav*, el nodo utilizará el formato predeterminado de IBM SPSS Statistics.

Justificar. Utilice esta columna para especificar la forma en que los valores deben justificarse en la columna Tabla. La configuración predeterminada es **Automático**, que justifica a la izquierda los valores simbólicos y a la derecha los numéricos. Puede omitir los valores predeterminados seleccionando **Izquierda**, **Derecha** o **Centro**.

Ancho de columna. De forma predeterminada, los anchos de columna se calculan automáticamente en función de los valores del campo. Para omitir el cálculo de ancho automático, pulse en una casilla de la tabla y utilice la lista desplegable para seleccionar un ancho nuevo. Para especificar un ancho personalizado que no aparezca en la lista, abra el subcuadro de diálogo de formatos de campo pulsando dos veces en una casilla de la tabla de la columna Campo o Formato. También puede pulsar con el botón derecho en una casilla y seleccionar **Definir formato**.

Ver campos actuales. De forma predeterminada, el cuadro de diálogo muestra la lista de los campos que están activos actualmente. Para ver la lista de campos no utilizados, seleccione **Ver configuración de campos no utilizados**.

Menú contextual. El menú contextual de esta pestaña proporciona distintas opciones de actualización de selección y configuración. Pulse en una columna con el botón derecho del ratón para visualizar este menú.

- **Seleccionar todo.** Selecciona todos los campos.
- **No seleccionar.** Borra la selección.
- **Seleccionar campos.** Selecciona los campos en función del tipo o las características de almacenamiento. Las opciones son **Seleccionar categórico**, **Seleccionar continuo** (numérico), **Seleccionar sin tipo**, **Seleccionar cadenas**, **Seleccionar números** o **Seleccionar fecha y hora**. Consulte el tema “Niveles de medición” en la página 127 para obtener más información.
- **Definir formato.** Abre un subcuadro de diálogo para especificar las opciones de fecha, hora y decimales por campos.
- **Definir justificación.** Establece la justificación de los campos seleccionados. Las opciones son **Automático**, **Centro**, **Izquierda** o **Derecha**.
- **Definir ancho de columna.** Establece el ancho de campo de los campos seleccionados. Especifique **Automático** para leer el ancho de los datos. También puede establecer el ancho del campo en 5, 10, 20, 30, 50, 100 o 200.

Opciones de configuración de Formato de campo

El formato de campo se especifica en un subcuadro de diálogo disponible en la pestaña Formato de los nodos Tipo y Tabla. Si ha seleccionado más de un campo antes de abrir este cuadro de diálogo, la

configuración del primer campo de la selección se utiliza para todos los campos. Al pulsar en **Aceptar** después de realizar las especificaciones, se aplicará esta configuración a todos los campos seleccionados en la pestaña Formato.

Las siguientes opciones están disponibles por campos. Muchas de estas configuraciones también se pueden especificar en el cuadro de diálogo de propiedades de la ruta. Cualquier configuración realizada en el nivel de campo omite la configuración predeterminada especificada para la ruta.

Formato de fecha. Seleccione un formato de fecha que usar en los campos de almacenamiento de fechas o cuando las funciones de fecha de CLEM interpreten las cadenas como fechas.

Formato de hora. Seleccione un formato de hora que utilizar en los campos de almacenamiento de horas o cuando las funciones de hora de CLEM interpreten las cadenas como horas.

Formato de presentación de los números. Se puede elegir entre los formatos de presentación estándar (####,###), científico (#,###E+##) o de moneda (\$###,##).

Símbolo decimal. Seleccione una coma (,) o un punto (.) como separador decimal.

Símbolo de agrupación. Para los formatos de presentación de los números, seleccione el símbolo utilizado para agrupar valores (por ejemplo, el punto en 3.000,00). Entre las opciones se incluyen la ausencia de puntuación, el punto, la coma, el espacio y definido por el entorno local (en cuyo caso se utiliza el valor predeterminado del entorno local actual).

Cifras decimales (estándar, científico, moneda, exportación). Para los formatos de presentación de los números, esta opción especifica el número de cifras decimales que se utilizan cuando se visualizan números reales. Esta opción se especifica de forma independiente en cada formato de presentación.

Justificar. Especifica la forma en que los valores deben justificarse en la columna Tabla. La configuración predeterminada es **Automático**, que justifica a la izquierda los valores simbólicos y a la derecha los numéricos. Puede omitir los valores predeterminados seleccionando Izquierda, Derecha o Centro.

Ancho de columna. De forma predeterminada, los anchos de columna se calculan automáticamente en función de los valores del campo. Puede especificar un ancho personalizado en intervalos de cinco usando las flechas situadas a la derecha del cuadro de lista.

Cambio de nombre o filtrado de campos

Puede cambiar el nombre o excluir campos en cualquier punto de una ruta. Por ejemplo, como investigador médico, es posible que no esté interesado en el nivel de potasio (datos de nivel de campo) de los pacientes (datos de nivel de registro); por ello, puede filtrar el campo K (potasio). Se puede hacer utilizando un nodo Filtrar diferente o utilizando la pestaña Filtro en un nodo de origen o resultado. Esta funcionalidad es la misma con independencia del nodo del que se accede.

- En los nodos de origen, como Archivo variable, Archivo fijo, Archivo Estadísticas y XML, puede cambiar el nombre de los campos o filtrarlos a medida que los datos se leen en IBM SPSS Modeler.
- Al utilizar un nodo Filtrar, puede cambiar el nombre de los campos o filtrarlos en cualquier punto de la ruta.
- Desde los nodos Exportar Estadísticas, Transformación de Estadísticas, Modelo Estadísticas y Resultados de Estadísticas puede filtrar los campos o cambiarles el nombre para que cumplan los estándares de nomenclatura de IBM SPSS Statistics. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.
- Puede utilizar el nodo Filtrar en cualquiera de los nodos anteriores para definir o editar conjuntos de respuestas múltiples. Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.

- Finalmente, puede utilizar un nodo Filtrar para correlacionar campos de un nodo de origen a otro.

Opciones de configuración de filtrado

La tabla utilizada en la pestaña Filtro muestra el nombre de cada campo tal y como entra en el nodo, así como el nombre de cada uno de ellos tal y como sale. Puede utilizar las opciones de esta tabla para filtrar o cambiar el nombre de los campos que están duplicados o no son necesarios para posteriores operaciones.

- **Campo.** Muestra los campos de entrada de los orígenes de datos conectados actualmente.
- **Filtro.** Muestra el estado del filtro de todos los campos de entrada. Los campos filtrados incluyen una X roja en la columna, indicativo de que este campo no se pasará a la parte posterior de la ruta. Pulse en la columna *Filtro* del campo seleccionado para activar o desactivar el filtrado. También puede seleccionar opciones para varios campos simultáneamente manteniendo pulsada la tecla Mayús mientras realiza la selección.
- **Campo.** Muestra los campos tal y como salen del nodo Filtrar. Los nombres duplicados se muestran en rojo. Puede editar los nombres de campos pulsando en esta columna e introduciendo un nombre nuevo. También se pueden eliminar campos pulsando en la columna *Filtro* para desactivar los campos duplicados.

Todas las columnas de la tabla se pueden organizar pulsando en la cabecera de la columna.

Ver campos actuales. Seleccione esta opción para ver los campos de los conjuntos de datos conectados de forma activa con el nodo Filtrar. Esta opción está seleccionada de forma predeterminada y es el método más común para utilizar los nodos Filtrar.

Ver configuración de campos no utilizados. Seleccione esta opción para ver los campos de los conjuntos de datos que han estado conectados con anterioridad al nodo Filtrar pero ya no lo están. Esta opción resulta útil al copiar nodos Filtrar de una ruta a otra o cuando se almacenan o se vuelven a cargar nodos Filtrar.

Menú del botón Filtro

Pulse en el botón Filtro en la esquina superior izquierda del cuadro de diálogo para acceder a un menú que proporciona diferentes accesos directos y otras opciones.

Puede seleccionar:

- Eliminar todos los campos.
- Incluir todos los campos.
- Conmutar todos los campos.
- Eliminar duplicados. *Note:* al seleccionar esta opción, se eliminan todas las instancias del nombre duplicado, incluida la primera.
- Cambie el nombre de los campos y conjuntos de respuestas múltiples para que cumplan con otras aplicaciones. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.
- Truncar nombres de campos.
- Anonimizar campo y conjunto de respuestas múltiples.
- Utilizar nombres de campos de entrada.
- Editar conjuntos de respuestas múltiples Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.
- Establecer el estado del filtro predeterminado.

También puede utilizar los botones de flechas situados en la parte superior del cuadro de diálogo para determinar si desea incluir o descartar campos de forma predeterminada. Esta opción es útil para los

conjuntos de datos grandes en los que sólo se incluyen unos pocos campos a partir de ese punto de la ruta. Por ejemplo, puede seleccionar únicamente los campos que quiera conservar e indicar que se descarte el resto (en lugar de seleccionar uno a uno todos los campos que se vayan a descartar).

Truncado de nombres de campos

En el menú del botón Filtro (esquina superior izquierda de la pestaña Filtro), puede seleccionar truncar nombres de campo.

Longitud máxima. Permite especificar un número de caracteres para limitar la longitud de los nombres de campos.

Número de dígitos. Si los nombres de campos, cuando están truncados, dejan de ser exclusivos, se truncarán más y se distinguirán mediante dígitos añadidos. Puede especificar el número de dígitos que utilizar. Utilice las flechas para ajustar el número.

Por ejemplo, la siguiente tabla ilustra cómo se truncan los nombres de campos de un conjunto de datos médicos mediante la configuración predeterminada (longitud máxima = 8 y número de dígitos = 2).

Tabla 22. Truncamiento de nombres de campo

Nombres de campos	Nombres de campos truncados
Entrada de paciente 1	Pacien01
Entrada de paciente 2	Pacien02
Ritmo cardíaco	RitmoCar
PS	PS

Anonimización de nombres de campos

Puede anonimizar los nombres de los campos de cualquier nodo que incluya una pestaña Filtro pulsando en el menú del botón Filtro en la esquina superior izquierda y seleccionando **Anonimizar nombres de campos**. Los nombres de los campos anonimizados están compuestos por un prefijo de cadena y un valor numérico exclusivo.

Es posible anonimizar los nombres de: Seleccione **Sólo campos seleccionados** para anonimizar sólo los nombres de los campos seleccionados en la pestaña Filtro. La opción seleccionada predeterminada es **Todos los campos**, que anonimiza todos los nombres de campos.

Prefijo de nombres de campo: El prefijo predeterminado para los nombres de campo anonimizados es **anon_**; seleccione **Personalizado** y escriba su propio prefijo si desea uno diferente.

Anonimizar conjuntos de respuestas múltiples. Anonimiza los nombres de conjuntos de respuestas múltiples de la misma manera como campos. Consulte el tema “Edición de conjuntos de respuestas múltiples” para obtener más información.

Para restaurar los nombres de campo originales, seleccione **Utilizar nombres de campos de entrada** en el menú del botón de filtro.

Edición de conjuntos de respuestas múltiples

Puede añadir o editar conjuntos de respuestas múltiples de cualquier nodo que incluya una pestaña Filtro pulsando en el menú del botón Filtro en la esquina superior izquierda y seleccionando **Editar conjuntos de respuestas múltiples**.

Los conjuntos de respuestas múltiples se utilizan para registrar datos que pueden tener más de un valor para cada caso, por ejemplo, si pregunta a los encuestados los museos que han visitado o las revistas que

han leído. Los conjuntos de respuestas múltiples se pueden importar a IBM SPSS Modeler mediante un nodo de origen Data Collection o un nodo de origen de archivo Statistics y se pueden definir en IBM SPSS Modeler mediante un nodo Filtrar.

Pulse en **Nuevo** para crear un conjunto de respuestas múltiples nuevo o pulse en **Editar** para modificar un conjunto existente.

Nombre y etiqueta. Especifica el nombre y la descripción del conjunto.

Tipo. Las preguntas de respuestas múltiples se pueden gestionar de dos formas:

- **Conjunto de dicotomía múltiple.** Se crea un campo de marca independiente para cada respuesta posible, además si hay 10 revistas, habrá 10 campos de marca, cada una con valores como 0 ó 1 para *true* o *false*. El valor de recuento permite especificar el valor que se recontará como true. Este método es muy útil si desea dejar que los encuestados puedan seleccionar todas las opciones aplicables.
- **Conjunto de categoría múltiple.** Se crea un campo nominal para cada respuesta, hasta el número máximo de respuestas de un encuestado concreto. Cada campo nominal tiene valores que representan las posibles respuestas, como 1 para *Muy Interesante*, 2 para *National Geographic* y 3 para *PC Actual*. Este método es de gran utilidad si desea limitar el número de respuestas; por ejemplo, si le pide a los encuestados que seleccionen entre las tres revistas que leen con mayor frecuencia.

Campos del conjunto. Utilice los iconos a la derecha para añadir o eliminar campos.

Comentarios

- Todos los campos incluidos en un conjunto de respuestas múltiples deben estar en la misma ubicación.
- Los conjuntos son distintos de los campos que incluyen. Por ejemplo, si elimina un conjunto, no se eliminarán los campos que incluyen, únicamente los enlaces entre los campos. El conjunto seguirá siendo visible desde el punto anterior de su eliminación, pero no desde un punto posterior.
- Si cambia el nombre de los campos mediante un nodo Filtrar (directamente en la pestaña o seleccionando las opciones Cambiar nombre para IBM SPSS Statistics, **Truncar** o **Anonimizar** del menú Filtro), también se actualizarán todas las referencias a estos campos utilizados en conjuntos de respuestas múltiples. Sin embargo, los campos de cualquier conjunto de respuestas múltiples que el nodo Filtrar active, no se eliminarán del conjunto de respuestas múltiples. Aunque estos campos ya no sean visibles en la ruta, el conjunto de respuestas múltiples seguirá haciendo referencia, lo que puede ser de gran ayuda cuando realice exportaciones, por ejemplo.

Nodo Derivar

Una de las características más eficaces de IBM SPSS Modeler es la capacidad de modificar valores de datos y derivar campos nuevos a partir de datos existentes. Durante proyectos minería de datos de larga duración, es común realizar varias derivaciones, como extraer un ID de cliente a partir de una cadena de datos del registro Web o crear un valor de por vida de clientes basado en los datos demográficos y de transacción. Todas estas transformaciones pueden realizarse, utilizando diversos nodos de operaciones con campos.

Existen diversos nodos que ofrecen la capacidad de derivar campos nuevos:



El nodo Derivar modifica los valores de datos o crea campos nuevos desde uno o más campos existentes. Crea campos del tipo fórmula, marca, nominal, estado, recuento y condicional.



El nodo Reclasificar transforma un conjunto de valores categóricos en otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis.



El nodo Intervalos crea automáticamente nuevos campos nominales (conjunto) en función de los valores de uno o más campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos como desviaciones desde la media. Una vez creados los intervalos para el campo nuevo, puede generar un nodo Derivar en función de los puntos de corte.



El nodo Marcas deriva varios campos de marcas en función de los valores categóricos definidos para uno o más campos nominales.



El nodo Reestructurar convierte un campo nominal o marca en un grupo de campos que se puede rellenar con los valores todavía de otro campo. Por ejemplo, para un campo determinado llamado *tipo de pago*, con valores de *crédito*, *efectivo*, y *débito*, se crearían tres campos nuevos (*crédito*, *efectivo*, *débito*), que contendría cada uno el valor del pago real realizado.



El nodo Historial se utiliza para crear campos nuevos que contienen datos de los campos de registros anteriores. Los nodos Historial se suelen utilizar para los datos secuenciales, como los datos de series temporales. Antes de utilizar un nodo Historial, puede desear ordenar los datos utilizando un nodo Ordenar.

Utilización del nodo Derivar

Al utilizar el nodo Derivar, puede crear seis tipos de campos nuevos desde uno o más campos existentes:

- **Fórmula.** El campo nuevo es el resultado de una expresión CLEM arbitraria.
- **Marca.** El campo nuevo es una marca que representa una condición especificada.
- **Nominal.** El campo nuevo es nominal, lo cual supone que sus miembros conforman un grupo de valores especificados.
- **Estado.** El campo nuevo es uno de dos estados. El cambio entre estos estados los desencadena una condición especificada.
- **Recuento.** El campo nuevo está basado en el número de veces que una condición es verdadera.
- **Condicional.** El campo nuevo es el valor de una de las dos expresiones, dependiendo del valor de una condición.

Cada uno de estos nodos contiene un conjunto de opciones especiales en el cuadro de diálogo del nodo Derivar. Estas opciones se describen en los siguientes temas.

Opciones básicas de configuración del nodo Derivar

En la parte superior del cuadro de diálogo de los nodos Derivar se incluyen varias opciones para seleccionar el tipo de nodo Derivar que se necesita.

Modo. Seleccione **Único** o **Múltiple**, dependiendo de si desea derivar múltiples campos. Al seleccionar **Múltiple**, el cuadro de diálogo muestra opciones para derivar múltiples campos.

Derivar campo. En el caso de los nodos Derivar únicos, especifique el nombre del campo que desea derivar y añádale a cada registro. El nombre predeterminado es DerivarN, donde N es el número de nodos Derivar que se han creado durante la sesión actual.

Derivar como. Seleccione un tipo de nodo Derivar, como Fórmula o Nominal, de la lista desplegable. Para cada tipo, se crea un campo nuevo en función de las condiciones que se especifiquen en el cuadro de diálogo de cada tipo.

Al seleccionar una opción de la lista desplegable, se añade un conjunto nuevo de controles al cuadro de diálogo principal de acuerdo con las propiedades de cada tipo de nodo Derivar.

Tipo de campo. Seleccione un nivel de medición (como continuo, categórico o marca) para el nodo recién derivado. Esta opción es común para todas las formas de nodos Derivar.

Note: para derivar campos nuevos, se suele requerir el uso de funciones especiales o expresiones matemáticas. Para facilitar la creación de estas expresiones, existe un generador de expresiones en el cuadro de diálogo para todos los tipos de nodos Derivar, que ofrece comprobación de reglas y una lista completa de expresiones del lenguaje para la manipulación de expresiones CLEM.

Derivación de múltiples campos

La selección del modo **Múltiple** en un nodo Derivar permite derivar múltiples campos en función de la misma condición dentro de un mismo nodo. Esta característica permite ahorrar tiempo cuando se desea realizar transformaciones idénticas en varios campos del conjunto de datos. Por ejemplo, si desea generar un modelo de regresión que prediga el salario actual en función del salario inicial y la experiencia previa, puede resultar útil aplicar una transformación de registro a las tres variables desviadas. En lugar de añadir un nodo Derivar nuevo a cada transformación, puede aplicar la misma función a todos los campos a la vez. Sólo hay que seleccionar todos los campos desde los que se va a derivar un campo nuevo y, a continuación, escribir la expresión de derivación con la función @FIELD dentro de los paréntesis del campo.

Note: la función @FIELD es una herramienta importante para derivar múltiples campos al mismo tiempo. Permite hacer referencia a los contenidos del campo actual o los campos sin especificar el nombre de campo exacto. Por ejemplo, una expresión CLEM utilizada para aplicar una transformación de registro a múltiples campos es registro(@FIELD).

Las siguientes opciones se añaden al cuadro de diálogo al seleccionar el modo **Múltiple**:

Derivar de. Permite utilizar el selector de campos para elegir los campos a partir de los cuales derivar campos nuevos. Se generará un campo de salida para cada campo seleccionado. *Note:* los campos seleccionados no tienen que coincidir en el tipo de almacenamiento; sin embargo, la operación Derivar fallará si la condición no es válida para *todos* los campos.

Extensión de nombre de campo. Permite escribir la extensión que desea añadir a los nombres de los campos nuevos. Por ejemplo, en el caso de un campo nuevo que contenga el registro *Salario Actual*, podría añadir la extensión *reg_* al nombre de campo, que resultaría en *reg_Salario Actual*. Utilice los botones de radio para seleccionar si desea añadir la extensión como un prefijo (al principio) o como un sufijo (al final) del nombre de los campos. El nombre predeterminado es DerivarN, donde N es el número de nodos Derivar que se han creado durante la sesión actual.

Como en el caso del nodo Derivar en modo único, debe crear una expresión para utilizarla al derivar un campo nuevo. Existen distintas opciones para crear una condición en función del tipo de operación de derivación seleccionada. Estas opciones se describen en los siguientes temas. Para crear una expresión, sólo hay que escribir en los campos de la fórmula o utilizar el generador de expresiones, al que se accede mediante el botón de calculadora. Recuerde utilizar la función @FIELD al hacer referencia a las manipulaciones en múltiples campos.

Selección de varios campos

Todos los nodos que realizan operaciones en múltiples campos de entrada, como Derivar (modo múltiple), Agregar, Ordenar, G. múltiple y Gráfico de tiempo, permiten seleccionar los campos de forma sencilla a través del siguiente cuadro de diálogo Seleccionar campos.

Ordenar por. Puede ordenar campos disponibles para su visualización eligiendo una de las siguientes opciones:

- **Natural.** El orden de los campos es aquél en que pasaron desde la parte anterior de la ruta de datos al nodo actual.
- **Nombre.** Ordena los campos siguiendo un orden alfabético para su visualización.
- **Tipo.** Ordena los campos en función de su nivel de medición. Esta opción es útil cuando se seleccionan campos con un nivel de medición en particular.

Seleccione los campos de la lista de uno en uno o mantenga pulsada la tecla Mayús o Ctrl mientras selecciona otros campos para seleccionar varios campos. También puede utilizar los botones que se muestran bajo la lista para seleccionar grupos de campos en función de su nivel de medición o seleccionar y anular la selección de todos los campos de la tabla.

Opciones de configuración de derivación como fórmula

Los nodos de derivación como fórmula crean un campo nuevo para cada registro de un conjunto de datos a partir de los resultados de una expresión CLEM. Esta expresión no puede ser condicional. Para derivar valores que se basan en una expresión condicional, utilice la marca o el tipo condicional del nodo Derivar.

Fórmula Permite especificar una fórmula mediante el lenguaje CLEM con el fin de derivar un valor para el campo nuevo.

Nota: Puesto que SPSS Modeler no puede saber qué nivel de submedición debe utilizarse para un campo de lista derivada, para los niveles de medición de Colección y Geospacial, puede pulsar en **Especificar...** para abrir el cuadro de diálogo Valores y establecer el nivel de submedición necesario. Si desea obtener más información, consulte “Configuración de los valores de lista derivados”.

Configuración de los valores de lista derivados

El recuadro de diálogo Valores se muestra cuando selecciona **Especificar...** en la lista desplegable **Tipo de campo** de la fórmula del nodo Derivar. En este recuadro de diálogo, se establecen los valores de los niveles de submedición que se utilizarán para los niveles de medición de **Tipo de campo** de la fórmula, ya sean geoespaciales o de colección.

Medición: Seleccione **Colección** o **Geoespacial**. Si selecciona cualquier otro nivel de medición, el recuadro de diálogo muestra un mensaje que indica que no hay valores editables.

Colección

El único elemento que puede establecer para el nivel de **Medición** de colección es **Listar medición**. De forma predeterminada, esta medición se establece en Sin tipo, pero puede seleccionar cualquier otro valor para establecer el nivel de medición de los elementos contenidos en la lista. Puede seleccionar una de las opciones siguientes:

- Sin tipo
- Categórico
- Continuo
- Nominal
- Ordinal
- Marca

Geoespacial

En el nivel de **Medición**, puede seleccionar las opciones siguientes para establecer el nivel de medición de los elementos contenidos en la lista:

Tipo: Seleccione el subnivel de medición del campo geoespacial. Los subniveles disponibles se determinan según la profundidad del campo de lista. Los valores predeterminados son los siguientes:

- Punto (profundidad cero)
- Cadena de líneas (profundidad uno)
- Polígono (profundidad uno)
- Multipunto (profundidad uno)
- Cadena multilínea (profundidad dos)
- Multipolígono (profundidad dos)

Para obtener más información acerca de los subniveles, consulte el tema Subniveles de mediciones geoespaciales en la sección Nodo Tipo, en la publicación Nodos de origen, proceso y resultado de SPSS Modeler.

Para obtener más información acerca de las profundidades de listas, consulte el tema Almacenamiento de lista y niveles de medición asociados en la publicación Nodos de origen, proceso y resultado de SPSS Modeler.

Sistema de coordenadas: Esta opción solo está disponible si ha modificado el nivel de medición desde un nivel no geoespacial al nivel geoespacial. Para aplicar un sistema de coordenadas a sus datos geoespaciales, seleccione este recuadro de selección. De forma predeterminada, se muestra el sistema de coordenadas en el panel **Herramientas > Propiedades de ruta > Opciones > Geoespacial**. Para utilizar un sistema de coordenadas diferente, pulse el botón **Cambiar** para visualizar el recuadro de diálogo Seleccionar sistema de coordenadas y seleccione el sistema que coincida con sus datos.

Para obtener más información acerca de los sistemas de coordenadas, consulte el tema Configuración de opciones geoespaciales para corrientes en la sección Trabajar con corrientes de la publicación SPSS Modeler Guía del usuario.

Derivación de un campo de lista o geoespacial

En algunas ocasiones los datos se deben registrar como un elemento de lista en SPSS Modeler con los atributos erróneos. Por ejemplo, como campos geoespaciales separados, tal como una coordenada x más una coordenada y, o como una longitud y latitud, en filas individuales de un archivo .csv. En esta situación, debe combinar campos individuales en un único campo de lista. Un modo de hacerlo es mediante el nodo Derivar.

Nota: Debe saber cuál es el campo x (o longitud) y cuál es el campo y (o latitud) cuando combina los datos geoespaciales. Debe combinar los datos de modo que el campo de lista resultante tenga el orden de los elementos: [x, y] o [Longitud, Latitud], que son los formatos estándar para las coordenadas geoespaciales.

Los pasos siguientes muestran un sencillo ejemplo de cómo derivar un campo de lista.

1. En su ruta, adjunte un nodo Derivar al nodo Origen.
2. En la pestaña Configuración del nodo Derivar, seleccione **Fórmula** en la lista **Derivar como**.
3. En **Tipo de campo**, seleccione **Colección** para una lista no geoespacial o **Geoespacial**. De forma predeterminada, SPSS Modeler utiliza un método de "mejor acierto" para establecer el detalle de lista correcto. Puede seleccionar **Especificar...** para abrir el recuadro de diálogo Valores. En este recuadro de diálogo, puede especificar información adicional acerca de los datos de su lista. Por ejemplo, en una lista geoespacial puede cambiar el nivel de medición.

4. En el panel **Fórmula**, especifique la fórmula para combinar sus datos en el formato de lista correcto. De forma alternativa, pulse el botón de la calculadora para abrir el generador de expresiones. Un ejemplo sencillo de una fórmula para derivar una lista es [x, y], donde x e y son campos separados en el origen de datos. El nuevo campo derivado que se crea es una lista en la que el valor de cada registro son los valores x e y concatenados para dicho registro.

Nota: Los campos que se combinan en una lista de este modo deben tener el mismo tipo de almacenamiento.

Para obtener más información acerca de las listas y de las profundidades de las listas, consulte “Almacenamiento de listas y niveles de medición asociados” en la página 11.

Opciones de configuración de derivación como marca

Los nodos de derivación como marca se utilizan para indicar una condición específica, como una alta presión sanguínea o una inactividad de cuenta del cliente. Se crea un campo de marcas para cada registro y, cuando se cumple la condición verdadera, el valor de marca verdadero se añade al campo.

Valor para Verdadero. Permite especificar un valor para incluirlo en el campo de marcas para los registros que cumplen la condición que se especifica a continuación. El valor predeterminado es T (del inglés 'True').

Valor para Falso. Permite especificar un valor para incluirlo en el campo de marcas de los registros que *no* cumplen la condición que se especifica a continuación. El valor predeterminado es F.

Verdadero si. Permite especificar una condición CLEM para evaluar ciertos valores de cada registro y proporcionar un valor verdadero o falso al registro (definido anteriormente). Tenga en cuenta que el valor verdadero se asignará a los registros en caso de que existan valores numéricos no falsos.

Note: para devolver una cadena vacía, es necesario escribir comillas de apertura y cierre sin nada entre ellas, como "". Las cadenas vacías se suelen utilizar a menudo, por ejemplo, como el valor falso para permitir que los valores verdaderos destaquen con más claridad en la tabla. De igual manera, las comillas deben utilizarse si desea un valor de cadena que, de otro modo, se trataría como número.

Ejemplo

En releases de IBM SPSS Modeler anteriores a 12.0, se importaban respuestas múltiples en un solo campo, con valores separados por comas. Por ejemplo:

```
museo_de_diseño,instituto_de_moda_y_textil  
museo_de_diseño  
museo_arqueológico  
$null$  
museo_nacional_de_arte,museo_nacional_de_ciencias,otro
```

Para preparar los datos para el análisis, puede utilizar la función un nodo `hassubstring` para generar un campo de marcas independiente para cada respuesta con una expresión como:

```
hassubstring(museums,"museo_de_diseño")
```

Establecimiento de opciones de derivación nominal

Los nodos de derivación nominal se utilizan para ejecutar un conjunto de condiciones CLEM que determine la condición que satisface cada registro. Cuando se cumple una condición para cada registro, se añadirá un valor (que indica el conjunto de condiciones que se cumple) al nuevo campo derivado.

Valor predeterminado. Permite especificar un valor que utilizar en el campo nuevo si no se cumple ninguna de las condiciones.

Definir campo como. Permite especificar un valor que introducir en el campo nuevo cuando se cumple una determinada condición. Cada valor de la lista tiene una condición asociada que se puede especificar en la columna adyacente.

Si esta condición es verdadera. Permite especificar una condición para cada miembro en el campo de conjuntos que se va a enumerar. Utilice el generador de expresiones para seleccionar de entre los campos y las funciones disponibles. Puede utilizar los botones de flecha y eliminación para reorganizar o eliminar las condiciones.

Las condiciones funcionan comprobando los valores de un campo determinado del conjunto de datos. Al comprobar cada condición, los valores especificados con anterioridad se asignan al nuevo campo para indicar la condición que se cumple, (caso de haber alguna). Si no se cumple ninguna de las condiciones, se utiliza el valor predeterminado.

Opciones de configuración de derivación como estado

En cierto modo, los nodos de derivación como estado son similares a los nodos de derivación como marca. Los nodos de derivación como marca establecen valores en función del cumplimiento de una condición *única* del registro actual, pero los nodos de derivación de estado pueden cambiar los valores de un campo dependiendo del modo en que cumplen *dos condiciones independientes*. Esto quiere decir que el valor cambiará (activado o desactivado) según se cumplan las condiciones.

Estado inicial. Permite seleccionar si se va a dar a cada registro del nuevo campo el valor **Activado** o **Desactivado** en un principio. Tenga en cuenta que este valor puede cambiar cada condición que se cumpla.

Valor "Activado". Permite especificar el valor del campo nuevo cuando se cumple la condición Activado.

Cambiar a "Activado" cuando. Permite especificar una condición CLEM que cambia el estado a Activado cuando la condición es verdadera. Pulse en el botón de calculadora para abrir el generador de expresiones.

Valor "Desactivado". Especifica el valor del campo nuevo cuando se cumple la condición Desactivado.

Cambiar a "Desactivado" cuando. Permite especificar una condición CLEM que cambia el estado a Desactivado cuando la condición sea falsa. Pulse en el botón de calculadora para abrir el generador de expresiones.

Note: para especificar una cadena vacía, es necesario escribir comillas de apertura y cierre sin nada entre ellas, como "". De igual manera, las comillas deben utilizarse si desea un valor de cadena que, de otro modo, se trataría como número.

Opciones de configuración de derivación como recuento

Los nodos de derivación como recuento se utilizan para aplicar una serie de condiciones a los valores de un campo numérico en el conjunto de datos. Al cumplirse cada condición, el valor del campo de recuento derivado aumenta en un incremento determinado. Este tipo de nodo Derivar resulta útil para los datos de series temporales.

Valor inicial. Establece un valor que utilizar en la ejecución del nuevo campo. El valor inicial debe ser una constante numérica. Utilice los botones de flecha para aumentar o reducir el valor.

Incrementar cuando. Permite especificar la condición CLEM que, de cumplirse, cambia el valor derivado en función del número especificado en Incrementar en. Pulse en el botón de calculadora para abrir el generador de expresiones.

Incrementar en. Permite especificar el valor utilizado para aumentar el recuento. Puede utilizar una constante numérica o el resultado de una expresión CLEM.

Restablecer si. Permite especificar una condición que, de cumplirse, restablece el valor derivado al valor inicial. Pulse en el botón de calculadora para abrir el generador de expresiones.

Opciones de configuración de derivación como condicional

Los nodos de derivación como condicional utilizan una serie de instrucciones Si-Entonces-En caso contrario para derivar el valor del campo nuevo.

Si. Permite especificar una condición CLEM que se evaluará en cada registro tras la ejecución. Si la condición es verdadera (o no falsa, en el caso de los números), el nuevo campo tomará el valor especificado a continuación por la expresión Entonces. Pulse en el botón de calculadora para abrir el generador de expresiones.

Entonces. Permite especificar un valor o expresión CLEM para el nuevo campo cuando la instrucción Si anterior es verdadera (o no falsa). Pulse en el botón de calculadora para abrir el generador de expresiones.

En caso contrario. Permite especificar un valor o expresión CLEM para el nuevo campo cuando la instrucción Si anterior es falsa. Pulse en el botón de calculadora para abrir el generador de expresiones.

Nueva codificación de valores con el nodo Derivar

Los nodos Derivar sirven también para recodificar valores, por ejemplo, convirtiendo un campo de cadena con valores categóricos en un campo nominal (conjunto) numérico.

1. En Derivar como, seleccione el tipo de campo (Nominal, Marca, etc.) según sea adecuado.
2. Especifique las condiciones para recodificar los valores. Por ejemplo, puede establecer el valor en 1 si `Drug='drugA'`, 2 si `Drug='drugB'`, etc.

Nodo Rellenar

Los nodos Rellenar se utilizan para sustituir valores de campos y cambiar el almacenamiento. Puede sustituir los valores en función de una condición CLEM, como `@BLANK(FIELD)`. También puede sustituir todos los espacios vacíos o valores nulos por un valor específico. Los nodos Rellenar suelen utilizarse junto con el nodo Tipo para sustituir valores perdidos. Por ejemplo, puede rellenar espacios vacíos con el valor medio de un campo especificando una expresión del tipo `@GLOBAL_MEAN`. Esta expresión rellenará todos los espacios vacíos con el valor medio tal y como lo calculó un nodo de valores globales.

Rellenar campos. El selector de campos (botón situado a la derecha del campo de texto) permite seleccionar campos del conjunto de datos cuyos valores se examinarán y sustituirán. El comportamiento predeterminado consiste en sustituir los valores en función de la condición y por las expresiones especificadas a continuación. También puede seleccionar un método alternativo de sustitución con las opciones de sustitución que se muestran a continuación.

Note: al seleccionar varios campos para sustituirlos por un valor definido por el usuario, es importante que los tipos de campo sean similares (todos numéricos o simbólicos).

Reemplazar. Seleccione esta opción para reemplazar los valores de los campos seleccionados mediante uno de estos métodos:

- **En función de una condición.** Esta opción activa el campo Condición y el generador de expresiones para crear una expresión utilizada como una condición de sustitución por el valor especificado.
- **Siempre.** Sustituye todos los valores del campo seleccionado. Por ejemplo, podría utilizar esta opción para convertir el almacenamiento de los ingresos en una cadena mediante la siguiente expresión CLEM: `(to_string(income))`.

- **Valor vacío.** Sustituye todos los valores vacíos especificados por el usuario en el campo seleccionado. La condición estándar @BLANK(@FIELD) se utiliza para seleccionar valores vacíos. *Note:* puede definir valores vacíos mediante la pestaña Tipos del nodo de origen o con un nodo Tipo.
- **Valores nulos.** Sustituye todos los valores nulos del sistema del campo seleccionado. La condición estándar @NULL(@FIELD) se utiliza para seleccionar valores nulos.
- **Valores vacíos y nulos.** Sustituye los valores vacíos y los valores nulos del sistema del campo seleccionado. Esta opción es útil cuando no se está seguro de si los valores nulos se han definido como valores perdidos.

Condición. Esta opción está disponible cuando se ha seleccionado la opción **En función de una condición**. Utilice este cuadro de texto para especificar una expresión CLEM para evaluar los campos seleccionados. Pulse en el botón de calculadora para abrir el generador de expresiones.

Reemplazar con. Permite especificar una expresión CLEM para dar un nuevo valor a los campos seleccionados. También puede reemplazar el valor por un valor nulo escribiendo undef en el cuadro de texto. Pulse en el botón de calculadora para abrir el generador de expresiones.

Note: cuando los campos seleccionados son cadenas, debe reemplazarlos con un valor de cadena. Al utilizar el valor predeterminado 0 u otro valor numérico como valor de sustitución para los campos de cadena, se producirá un error.

Conversión del almacenamiento mediante el nodo Rellenar

Al utilizar la condición Reemplazar de un nodo Rellenar, puede convertir fácilmente el almacenamiento del campo de uno o varios campos. Por ejemplo, al utilizar la función de conversión to_integer, se puede convertir la cadena *ingreso* en valor entero con la siguiente expresión CLEM: to_integer(income).

El generador de expresiones permite ver funciones de conversión disponibles y crear automáticamente una expresión CLEM. En la lista desplegable de funciones, seleccione **Conversión** para ver una lista de funciones de conversión de almacenamiento. Las funciones de conversión disponibles son:

- to_integer(ELEMENTO)
- to_real(ELEMENTO)
- to_number(ELEMENTO)
- to_string(ELEMENTO)
- to_time(ELEMENTO)
- to_timestamp(ELEMENTO)
- to_date(ELEMENTO)
- to_datetime(ELEMENTO)

Conversión de valores de fecha y hora. Tenga en cuenta que las funciones de conversión (y muchas otras funciones que requieren un tipo específico de entrada como un valor de fecha y hora) dependen de los formatos actuales especificados en el cuadro de diálogo Opciones de ruta. Por ejemplo, si desea convertir un campo de cadena con los valores *Ene 2003, Feb 2003*, etc. en un almacenamiento de fecha, seleccione **MES AAAA** como formato de fecha predeterminado de la ruta.

Las funciones de conversión también están disponibles en el nodo Derivar para la conversión temporal durante un cálculo de derivación. También puede utilizar el nodo Derivar para realizar otras manipulaciones, como volver a codificar los campos con valores categóricos. Consulte el tema “Nueva codificación de valores con el nodo Derivar” en la página 149 para obtener más información.

Nodo Reclasificar

El nodo Reclasificar permite la transformación desde un conjunto de valores categóricos a otro. La reclasificación es útil para contraer categorías o reagrupar datos para su análisis. Por ejemplo, se pueden reclasificar los valores *Producto* en tres grupos, como por ejemplo *Utensilios de cocina, Baño y ropa de cama y Electrodomésticos*. A menudo, esta operación se realiza directamente desde un nodo Distribución agrupando valores y generando un nodo Reclasificar. Consulte el tema “Utilización de un nodo Distribución” en la página 240 para obtener más información.

La reclasificación puede realizarse mediante uno o varios campos simbólicos. También puede sustituir los nuevos valores por el campo existente o generar un campo nuevo.

Cuándo utilizar un nodo Reclasificar

Antes de utilizar un nodo Reclasificar, tenga en cuenta si existe otro nodo de operaciones con campos que sea más adecuado para la tarea en cuestión:

- Para transformar rangos numéricos en conjuntos con un método automático, como por ejemplo rangos o percentiles, se debe utilizar un nodo Intervalos. Consulte el tema “Nodo Intervalos” en la página 155 para obtener más información.
- Para clasificar rangos numéricos en conjuntos manualmente, se debe utilizar un nodo Derivar. Por ejemplo, si desea contraer valores de salario en categorías de rango de salario específicas, se debe utilizar un nodo Derivar para definir cada categoría manualmente.
- Para crear uno o más campos de marcas en función de los valores de un campo categórico, como por ejemplo *Mortgage_type*, se debe utilizar un nodo Marcas.
- Para convertir un campo categórico en almacenamiento numérico, puede usar un nodo Derivar. Así, podría convertir los valores *No* y *Sí* a 0 y 1 respectivamente. Consulte el tema “Nueva codificación de valores con el nodo Derivar” en la página 149 para obtener más información.

Opciones de configuración del nodo Reclasificar

Hay tres pasos para utilizar el nodo Reclasificar:

1. Primero, seleccione si desea reclasificar varios campos o un campo único.
2. A continuación, elija si desea recodificar en el campo existente o crear un campo nuevo.
3. A continuación, utilice las opciones dinámicas del cuadro de diálogo del nodo Reclasificar para correlacionar conjuntos como desee.

Modo. Seleccione **Único** para reclasificar las categorías de un campo. Seleccione **Múltiple** para activar las opciones que permiten la transformación de varios campos al mismo tiempo.

Reclasificar como. Seleccione **Campo nuevo** para conservar el campo nominal original y derivar un campo adicional que contenga los valores clasificados. Seleccione **Campo existente** para sobrescribir los valores en el campo original con las nuevas clasificaciones. Se trata fundamentalmente de una operación de “relleno”.

Una vez especificado el modo y las opciones de sustitución, debe seleccionar el campo de transformación y especificar los nuevos valores de clasificación mediante las opciones dinámicas de la mitad inferior del cuadro de diálogo. Estas opciones varían en función del modo seleccionado con anterioridad.

Reclasificar campos. Utilice el botón de selección de campos de la parte derecha para seleccionar un campo discreto (modo Único) o más campos categóricos (modo Múltiple).

Nombre del campo nuevo. Permite especificar un nombre para el nuevo campo nominal que contiene valores recodificados. Esta opción está disponible solamente en modo Único si se ha seleccionado **Campo nuevo** con anterioridad. Cuando **Campo existente** está seleccionado, el nombre del campo original se conserva. Al trabajar en modo Múltiple, esta opción se sustituye por controles para especificar una extensión añadida a cada nuevo campo. Consulte el tema “Reclasificación de múltiples campos” en la página 152 para obtener más información.

Valores de reclasificación. Esta tabla permite establecer una correlación clara entre los antiguos valores del conjunto y los que se especifican aquí.

- **Valor original.** Esta columna enumera los valores existentes para los campos seleccionados.
 - **Valor nuevo.** Utilice esta columna para escribir nuevos valores de categoría o seleccionar uno de la lista desplegable. Al generar automáticamente un nodo Reclasificar con los valores del cuadro Distribución, estos valores se incluyen en la lista desplegable. Esto permite correlacionar rápidamente los valores existentes a un conjunto de valores conocido. Por ejemplo, en ocasiones las organizaciones de salud agrupan los diagnósticos de forma distinta en función de la configuración de red o local. Después de una fusión o una adquisición, todas las partes deben reclasificar los datos nuevos e incluso los existentes de forma coherente. En lugar de escribir manualmente cada valor objetivo de una larga lista, puede leer la lista maestra de valores en IBM SPSS Modeler, ejecutar un cuadro Distribución para el campo *Diagnóstico* y generar un nodo Reclasificar (valores) para este campo directamente desde el cuadro. Este proceso hará que todos los valores objetivo de Diagnóstico estén disponibles en la lista desplegable Valores nuevos.
4. Pulse en **Obtener** para leer los valores originales de uno o varios campos seleccionados con anterioridad.
 5. Pulse en **Copiar** para pegar valores originales en la columna *Valor nuevo* en el caso de los campos que aún no se han correlacionado. Los valores originales sin correlación se añaden a la lista desplegable.
 6. Pulse en **Borrar nuevo** para borrar todas las especificaciones de la columna *Valor nuevo*. *Note:* esta opción no elimina los valores de la lista desplegable.
 7. Pulse en **Automático** para generar automáticamente números enteros consecutivos para cada uno de los valores originales. Sólo se pueden generar valores enteros (no valores reales como 1,5 o 2,5).

Por ejemplo, puede generar automáticamente números de ID de producto consecutivos para los nombres de productos o números de cursos para ofertas de clasificación universitaria. Esta función se corresponde con la transformación de la recodificación automática para los conjuntos de IBM SPSS Statistics.

Para valores no especificados usar. Esta opción se utiliza para rellenar valores no especificados del campo nuevo. También puede conservar el valor original seleccionando **Valor original** o especificar un valor predeterminado.

Reclasificación de múltiples campos

Para correlacionar valores de categorías para más de un campo a la vez, active el modo **Múltiple**. Esto permite establecer nuevas configuraciones en el cuadro de diálogo Reclasificar, que se describen a continuación.

Reclasificar campos. Utilice el botón de selección de campos de la parte derecha para seleccionar los campos que desea transformar. El botón de selección de campos permite seleccionar todos los campos a la vez o campos de tipos similares, como nominal o marca.

Extensión de nombre de campo. Al recodificar múltiples campos simultáneamente, es más eficaz especificar una extensión común añadida a los todos los campos nuevos en lugar de los nombres de campos individuales. Especifique una extensión del tipo *_recode* y seleccione si desea añadir esta extensión antes o después de los nombres de campos originales.

Almacenamiento y nivel de medición de los campos reclasificados

El nodo Reclasificar siempre crea un campo nominal desde la operación de recodificación. En algunos casos, esto puede cambiar el nivel de medición de campo al utilizar el modo de reclasificación **Campo existente**.

El almacenamiento de los campos nuevos (la forma de *almacenar* los datos en lugar de cómo se *utilizan*) se calcula dependiendo de las siguientes opciones de la pestaña Configuración:

- Si se establece que los valores sin especificar utilicen un valor predeterminado, el tipo de almacenamiento vendrá determinado al examinar los nuevos valores y el valor predeterminado y determinar el almacenamiento adecuado. Por ejemplo, si todos los valores se pueden analizar como números enteros, el campo tendrá un tipo de almacenamiento entero.
- Si los valores no especificados se configuren para utilizar valores originales, el tipo de almacenamiento se basará en el almacenamiento del campo original. Si todos los valores pueden analizarse como el almacenamiento del campo original, dicho almacenamiento se conservará; de lo contrario, el almacenamiento vendrá determinado por el tipo de almacenamiento más adecuado que cubra tanto valores antiguos como nuevos. Por ejemplo, al reclasificar un conjunto de números enteros { 1, 2, 3, 4, 5 } con la reclasificación 4 => 0, 5 => 0 se genera un conjunto de números enteros nuevos { 1, 2, 3, 0 }, mientras que con la reclasificación 4 => "Sobre 3", 5 => "Sobre 3" se genera el conjunto de serie {"1", "2", "3", "Sobre 3"}.

Nota: si el tipo original estaba sin instanciar, el nuevo tipo será también sin instanciar.

Nodo Anonimizar

El nodo Anonimizar permite ocultar nombres de campo, valores de campo o ambos, cuando se trabaja con datos que se van a incluir en un modelo que se encuentra en un punto posterior de la ruta al nodo. De esta forma, el modelo generado se puede distribuir libremente (por ejemplo, al servicio de asistencia técnica) sin peligro de que usuarios no autorizados puedan acceder a datos confidenciales como registros de empleados o registros médicos de pacientes.

Según dónde coloque el nodo Anonimizar en la ruta, puede que necesite realizar cambios en otros nodos. Por ejemplo, si inserta un nodo Anonimizar antes de un nodo Seleccionar, será necesario cambiar los criterios de selección del nodo Seleccionar para que actúen sobre los valores que ahora han sido anonimizados.

El método que se debe utilizar para la anonimización depende de varios factores. En los nombres de campos y en todos los valores de campo, excepto los de niveles de medición continuo, los datos se sustituyen por una ruta con el siguiente formato:

prefijo_Sn

donde *prefijo_* es, bien una cadena especificada por el usuario, o bien la cadena predeterminada *anon_*, y *n* es un valor entero que empieza en 0 y aumenta en cada valor exclusivo (por ejemplo, *anon_S0*, *anon_S1*, etc.)

Los valores de campos de tipo Continuo deben transformarse, ya que los rangos numéricos trabajan con valores enteros o reales en lugar de cadenas. Como tal, pueden anonimarse transformando solamente el rango en otro diferente, ocultando de esta forma los datos originales. La transformación de un valor *x* en el rango se realiza de la siguiente forma:

$$A*(x + B)$$

donde:

A es un factor de escala que debe ser mayor que 0.

B es un desplazamiento de traducción que se suma a los valores.

Ejemplo

En el caso del campo *EDAD*, donde el factor de escala *A* se establece en 7 y el desplazamiento de traducción *B* en 3, los valores para *EDAD* se transforman en:

$$7*(EDAD + 3)$$

Opciones de configuración del nodo Anonimizar

Aquí, puede seleccionar los campos cuyos valores se van a ocultar a partir de ese punto de la ruta.

Recuerde que los campos se deben instanciar antes del nodo Anonimizar para que se puedan realizar las operaciones de anonimización. Para instanciar los datos, pulse en el botón **Leer valores** de un nodo Tipo o en la pestaña Tipos de un nodo de origen.

Campo. Muestra los campos del conjunto de datos actual. Si ya se han anonimizado algunos nombres de campo, se muestran los nombres anonimizados.

Medición. El nivel de medición del campo.

Anonimizar valores. Seleccione uno o más campos, pulse en esta columna y elija **Sí** para anonimizar el valor del campo mediante el prefijo predeterminado **anon_**; elija **Especificar** para mostrar un cuadro de diálogo donde podrá escribir su propio prefijo o, en el caso de valores de campo de tipo *Continuo*, especificar si la transformación de los valores de campo va a utilizar valores aleatorios o especificados por el usuario. Tenga en cuenta que no se pueden especificar en la misma operación los tipos de campo *Continuo* y no *Continuo*, sino que deberá hacerlo por separado para cada tipo de campo.

Ver campos actuales. Seleccione esta opción para ver los campos de los conjuntos de datos conectados de forma activa con el nodo Anonimizar. Esta opción está seleccionada de forma predeterminada.

Ver configuración de campos no utilizados. Seleccione esta opción para ver los campos de los conjuntos de datos que han estado conectados con anterioridad al nodo pero ya no lo están. Esta opción resulta útil al copiar nodos de una ruta a otra o cuando se guardan y se vuelven a cargar nodos.

Especificación del método de anonimización de los valores de campo

El cuadro de diálogo Reemplazar valores permite elegir si se va a utilizar el prefijo predeterminado para los valores de campo anonimizados o un prefijo personalizado. Al pulsar en **Aceptar** en este cuadro de diálogo, la opción Anonimizar valores de la pestaña Configuración cambia a **Sí** para el campo o campos seleccionados.

Prefijo de valores de campo. El prefijo predeterminado para los valores de campo anonimizados es **anon_**; seleccione **Personalizado** y escriba su propio prefijo si desea uno diferente.

El cuadro de diálogo Transformar valores se muestra solamente para los campos de tipo Continuo y permite especificar si en la transformación de los valores de campo se van a utilizar valores aleatorios o especificados por el usuario.

Aleatorios. Seleccione esta opción para utilizar valores aleatorios para la transformación. **Establecer semilla aleatoria** está seleccionado de forma predeterminada; especifique un valor en el campo **Semilla** o utilice el valor predeterminado.

Fija. Seleccione esta opción para especificar sus propios valores aleatorios para la transformación.

- **Escalar por.** El número por el que se multiplicarán los valores de campo en la transformación. El valor mínimo es 1 y el máximo suele ser 10, pero puede reducirse para evitar desbordamientos.
- **Trasladar por.** El número que se sumará a los valores de campo en la transformación. El valor mínimo es 0 y el máximo suele ser 1000, pero puede reducirse para evitar desbordamientos.

Anonimización de valores de campo

Los campos seleccionados para la anonimización en la pestaña Configuración tienen sus valores anonimizados:

- cuando se ejecuta la ruta que contiene el nodo Anonimizar
- cuando se previsualizan los valores

Para ver la presentación preliminar de los valores, pulse en el botón **Anonimizar valores** en la pestaña Valores anonimizados. A continuación, seleccione un nombre de campo desde la lista desplegable.

Si el nivel de medición es Continuo, la pantalla muestra:

- valores mínimo y máximo del rango original
- ecuación utilizada para transformar los valores

Si el nivel de medición es diferente de Continuo, la pantalla muestra los valores original y anonimizado de este campo.

Cuando la visualización aparece con un fondo amarillo, indica que la configuración para el campo seleccionado ha cambiado desde la última anonimización de los valores o que se han realizado cambios a los datos en un punto de la ruta anterior al nodo Anonimizar y, por lo tanto, los valores de anonimización puede que ya no sean correctos. Se muestra el conjunto actual de valores; vuelva a pulsar en el botón **Anonimizar valores** para generar un nuevo conjunto de valores de acuerdo con la configuración actual.

Anonimizar valores. Crea valores anonimizados para el campo seleccionado y los muestra en la tabla. Si está utilizando semillas aleatorias para un campo de tipo Continuo, pulse varias veces en este botón para crear un conjunto de valores diferente cada vez.

Borrar valores. Borra los valores originales y anonimizados de la tabla.

Nodo Intervalos

El nodo Intervalos permite crear automáticamente nuevos campos nominales en función de los valores de uno o varios campos continuos (rango numérico) existentes. Por ejemplo, puede transformar un campo de ingresos continuo en un campo categórico nuevo que contenga grupos de ingresos con el mismo ancho, o como desviaciones desde la media. Si lo prefiere, puede seleccionar un campo categórico "supervisor" para conservar la fuerza de la asociación original entre los dos campos.

Los intervalos pueden resultar útiles por varias razones, entre ellas:

- **Requisitos de algoritmos.** Algunos algoritmos como Bayesiano ingenuo y Regresión logística requieren entradas categóricas.
- **Rendimiento.** El rendimiento de algoritmos como los de regresión logística multinomial es mayor si se reduce el número de valores distintos de los campos de entrada. Por ejemplo, utilice el valor de la media o la mediana para cada intervalo en lugar de los valores originales.
- **Privacidad de los datos.** La información personal confidencial, como los salarios, se puede registrar en rangos en lugar de cifras salariales reales para proteger la privacidad.

Están disponibles diversos métodos de creación de intervalos. Una vez creados los intervalos para el campo nuevo, puede generar un nodo Derivar en función de los puntos de corte.

Cuándo utilizar un nodo Intervalos

Antes de utilizar un nodo Intervalos, considere si existe otra técnica más adecuada para la tarea en cuestión:

- Para especificar manualmente los puntos de corte para las categorías, como rangos de salario predefinidos específicos, utilice un nodo Derivar. Consulte el tema "Nodo Derivar" en la página 142 para obtener más información.
- Para crear nuevas categorías para conjuntos existentes, utilice un nodo Reclasificar. Consulte el tema "Nodo Reclasificar" en la página 151 para obtener más información.

Gestión de valores perdidos

El nodo Intervalos gestiona valores perdidos de las siguientes formas:

- **Elementos vacíos especificados por el usuario.** Los valores perdidos especificados como valores vacíos se incluyen durante la transformación. Por ejemplo, si designó -99 para indicar un valor vacío usando el nodo Tipo, este valor se incluirá en el proceso de intervalos. Para omitir los valores vacíos durante el intervalo, se debe usar el nodo Rellenar, que permite sustituir valores vacíos por valores nulos del sistema.
- **Valores perdidos del sistema (\$null\$).** Los valores nulos se omiten durante la transformación de intervalos y permanecen así tras la transformación.

La pestaña Configuración proporciona opciones para las técnicas disponibles. La pestaña Ver muestra puntos de corte establecidos para los datos ejecutados previamente en el nodo.

Opciones de configuración del nodo Intervalos

Con el nodo Intervalos puede generar intervalos (categorías) automáticamente mediante las siguientes técnicas:

- Intervalos de ancho fijo
- Cuantiles (igual frecuencia o igual suma)
- Media y desviación estándar
- Rangos
- Optimizado respecto a un campo categórico "supervisor"

La mitad inferior del cuadro de diálogo cambia de forma dinámica en función del método de intervalos seleccionado.

Campos de intervalos. Los campos continuos (rango numérico) pendientes de transformación se muestran aquí. El nodo Intervalos permite segmentar varios campos simultáneamente. Añada campos utilizando los botones de la derecha.

Método de intervalos. Permite seleccionar el método utilizado para determinar los puntos de corte de los intervalos de campo nuevos (categorías). Los temas siguientes describen las opciones disponibles en cada caso.

Umbrales de intervalo. Especifica cómo se calculan los umbrales de intervalo.

- **Recalcular siempre.** Los puntos de corte y las asignaciones de umbrales siempre se recalculan cuando se ejecuta el nodo.
- **Leer de la pestaña Agrupar valores en intervalos si está disponible.** Los puntos de corte y las asignaciones de umbrales se calculan únicamente cuando sea necesario (por ejemplo, cuando se han añadido nuevos datos).

Los siguientes temas detallan las opciones de los métodos de intervalos disponibles.

Intervalos de ancho fijo

Al seleccionar **Ancho fijo** como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

Extensión del nombre. Permite especificar una extensión que utilizar al generar campos. *_INT* es la extensión predeterminada. También puede determinar si la extensión se debe añadir al principio (**prefijo**) o al final (**sufijo**) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos_INT*.

Ancho de intervalo. Especifique un valor (entero o real) para calcular la "anchura" del intervalo. Por ejemplo, puede utilizar el valor predeterminado 10 para segmentar el campo *Edad*. Dado que *Edad* tiene un rango de 18 a 65, los intervalos generados serían los que se muestran en la tabla siguiente:

Tabla 23. Intervalos de Edad con el rango 18-65

Intervalo 1	Intervalo 2	Intervalo 3	Intervalo 4	Intervalo 5	Intervalo 6
>=13 a <23	>=23 a <33	>=33 a <43	>=43 a <53	>=53 a <63	>=63 a <73

El inicio de los intervalos se calcula mediante el valor menor explorado menos la mitad del ancho de intervalo (tal y como se especifica). Por ejemplo, en los intervalos que se muestra más arriba, 13 se utiliza para iniciar los intervalos según el siguiente cálculo: $18 [\text{menor valor de datos}] - 5 [0,5 \times (\text{Anchura de intervalo } 10)] = 13$.

Número de intervalos. Utilice esta opción para especificar un número entero para determinar el número de intervalos de ancho fijo (categorías) para los nuevos campos.

Una vez ejecutado el nodo Intervalos en una ruta, puede ver los umbrales de intervalo generados pulsando en la pestaña **Presentación preliminar** del cuadro de diálogo del nodo Intervalos. Consulte el tema “Vista preliminar de intervalos generados” en la página 161 para obtener más información.

Cuantiles (igual frecuencia o igual suma)

Con el método de intervalos de cuantil se crean campos nominales que pueden servir para dividir registros explorados en grupos de percentiles (o cuantiles, deciles, etc.) para que, de este modo, cada grupo contenga el mismo número de registros, o bien la suma de los valores de cada uno de ellos sea la misma. Los registros se clasifican en orden ascendente en función del valor del campo de intervalo especificado, y así, a los registros con los valores más bajos de la variable de intervalo seleccionada se les asigna un rango de 1, al siguiente conjunto de registros un rango de 2, y así sucesivamente. Los valores de umbral de cada intervalo se generan automáticamente según los datos y el método de mosaico empleados.

Extensión del nombre de cuantil. Permite especificar una extensión utilizada para los campos generados con p-tilas estándar. La extensión predeterminada es `_TILE` más *N*, donde *N* es el número de cuantil. También puede determinar si la extensión se debe añadir al principio (**prefijo**) o al final (**sufijo**) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado `ingresos_INT4`.

Extensión de cuantil personalizada. Permite especificar una extensión utilizada para un rango de cuantiles personalizados. El valor predeterminado es `_TILN`. Tenga en cuenta que, en este caso, la *N* no se sustituye por el número personalizado.

Los p-tilas disponibles son:

- **Cuartil.** Genera cuatro intervalos, cada uno con el 25% de los casos.
- **Quintil.** Genera cinco intervalos, cada uno con el 20% de los casos.
- **Decil.** Genera 10 intervalos, cada uno con el 10% de los casos.
- **Veintil.** Genera 20 intervalos, cada uno con el 5% de los casos.
- **Percentil.** Genera 100 intervalos, cada uno con el 1% de los casos.
- **N personalizado.** Seleccione esta opción para especificar el número de intervalos. Por ejemplo, un valor de 3 produciría 3 categorías agrupadas (2 puntos de corte), cada una de las cuales contendría el 33,3% los casos.

Tenga en cuenta que si hay menos valores discretos en los datos que cantidad de cuantiles especificados, no se utilizarán todos los cuantiles. En estos casos la nueva distribución suele reflejar la distribución original de los datos.

Método de mosaico. Especifica el método utilizado para asignar registros a los intervalos.

- **Recuento de registros.** Trata de asignar el mismo número de registros a cada intervalo.

- **Suma de los valores.** Trata de asignar registros a intervalos de forma que la suma de los valores de cada intervalo sea la misma. Al intentar dirigir los esfuerzos de ventas, por ejemplo, este método puede servir para asignar posibles clientes a grupos de deciles en función del valor de cada registro, de forma que los posibles clientes con los valores más altos se incluyan en el intervalo superior. Así, una empresa farmacéutica puede clasificar a los médicos en grupos de deciles en función del número de prescripciones que realizan. Si bien cada decil contendría aproximadamente el mismo número de scripts, el número de individuos que contribuyen en dichos scripts no sería el mismo, concentrándose los individuos que realizan más scripts en el decil 10. Observe que en este método se da por hecho que todos los valores son superiores a cero, de modo que puede arrojar resultados inesperados en caso de que no sea así.

Empates. Una condición de empate se produce cuando los valores en ambas partes de un punto de corte son idénticos. Por ejemplo, si está asignando deciles y más del 10% de los registros tienen el mismo valor para el campo de intervalo, no todos podrán encajar en el mismo intervalo a menos que se fuerce el umbral de una forma u otra. Así, los empates pueden pasar al siguiente intervalo o permanecer en el actual, pero deberán resolverse de manera que todos los registros con valores idénticos se encuentren en el mismo intervalo, aun cuando esto suponga que algunos intervalos contengan más registros de lo previsto. En consecuencia, los umbrales de los intervalos siguientes también pueden ajustarse, lo que hará que los valores se asignen de manera diferente para el mismo conjunto de números según el método empleado para solucionar el empate.

- **Añadir al siguiente.** Seleccione esta opción para mover los valores de empate hacia el siguiente intervalo superior.
- **Mantener en el actual.** Mantiene los valores de empate en el intervalo actual (más bajo). Este método puede dar como resultado un menor número total de intervalos creados.
- **Asignar aleatoriamente.** Seleccione esta opción para relacionar los valores aleatoriamente a un intervalo. Se intenta mantener así el mismo número de registros en cada intervalo.

Ejemplo: Mosaico por recuento de registros

La tabla siguiente ilustra cómo los valores de campo simplificados se clasifican como cuartiles cuando se usa el mosaico por recuento de registros. Tenga en cuenta que los resultados variarán en función de la opción de empate seleccionada.

Tabla 24. Ejemplo de mosaico por recuento de registros.

Valores	Añadir al siguiente	Mantener en el actual
10	1	1
13	2	1
15	3	2
15	3	2
20	4	3

El número de elementos por intervalo se calcula así:

número total de valor / número de cuantiles

En el ejemplo simplificado anterior, el número de elementos deseado por intervalo es 1,25 (5 valores / 4 cuartiles). El valor 13 (siendo el número de valor 2) divide el umbral del recuento deseado de 1,25 y, por lo tanto, se trata de forma distinta en función de la opción de empate seleccionada. En el modo **Añadir al siguiente**, se añade al intervalo 2. En el modo **Mantener en el actual**, se deja en el intervalo 1, forzando el rango de valores del intervalo 4 fuera de los valores de datos existentes. Como resultado, solamente se crearán tres intervalos y los umbrales para cada uno de ellos se ajustarán en consecuencia, como se muestra en la tabla siguiente.

Tabla 25. Resultado del ejemplo de intervalos.

Intervalo	Inferior	Superior
1	>=10	<15
2	>=15	<20
3	>=20	<=20

Note: la velocidad de creación de intervalos por cuantiles puede ser mayor si se activa el procesamiento paralelo.

Clasificar casos

Al seleccionar **Rangos** como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

La clasificación crea nuevos campos que contienen rangos, rangos fraccionales y valores de percentiles para campos numéricos en función de las opciones especificadas más adelante.

Orden de clasificación. Permite seleccionar **Ascendente** (el valor más bajo está marcado con 1) o **Descendente** (el valor más alto está marcado con 1).

Rango. Seleccione esta opción para clasificar casos en orden ascendente o descendente tal y como se especifica anteriormente. El rango de valores del campo nuevo será 1– N , donde N es el número de valores específicos del campo original. Los valores empatados toman la media de sus rangos.

Rango fraccional. Seleccione esta opción para clasificar casos donde el valor del campo nuevo es igual al rango dividido por la suma de las ponderaciones de los casos que no están perdidos. Los rangos fraccionales están dentro del rango 0–1.

Rango fraccional como porcentaje. Cada rango se divide por el número de registros que tienes valores válidos y se multiplica por 100. Los rangos fraccionales de porcentaje están dentro del rango 1–100.

Extensión. En todas las opciones de rangos puede crear extensiones personalizadas y especificar si la extensión se debe añadir al principio (**prefijo**) o al final (**sufijo**) del nombre del campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos_P_RANGO*.

Media/desviación estándar

Al seleccionar **Media/desviación estándar** como método de intervalos, se muestra un conjunto de opciones en el cuadro de diálogo.

Este método genera uno o varios campos nuevos con categorías de unión en función de los valores de la media y la desviación estándar de la distribución de los campos especificados. Seleccione el número de desviaciones que desee utilizar a continuación.

Extensión del nombre. Permite especificar una extensión que utilizar al generar campos. *_DTINT* es la extensión predeterminada. También puede determinar si la extensión se debe añadir al principio (**prefijo**) o al final (**sufijo**) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado *ingresos_DTINT*.

- **+/- 1 desviación estándar.** Seleccione esta opción para generar tres intervalos.
- **+/- 2 desviaciones estándar.** Seleccione esta opción para generar cinco intervalos.
- **+/- 3 desviaciones estándar.** Seleccione esta opción para generar siete intervalos.

Por ejemplo, al seleccionar +/-1 desviación estándar, se generan tres intervalos, como se muestra en la tabla siguiente:

Tabla 26. Ejemplo de intervalo con desviación estándar.

Intervalo 1	Intervalo 2	Intervalo 3
$x < (\text{Media} - \text{Desv. estándar})$	$(\text{Media} - \text{Desv. estándar}) \leq x \leq (\text{Media} + \text{Desv. estándar})$	$x > (\text{Media} + \text{Desv. estándar})$

En una distribución normal, el 68% de los casos entran dentro de una desviación estándar de la media, el 95% en dos desviaciones estándar y el 99% en tres desviaciones estándar. No obstante, tenga en cuenta que al crear categorías de uniones en función de las desviaciones estándar, algunos intervalos pueden definirse fuera del rango de datos real e incluso fuera del rango de los valores de datos posibles (por ejemplo, un rango de salario negativo).

Intervalos óptimos

Si el campo que desea dividir en intervalos está fuertemente asociado a otro campo categórico, puede seleccionar el campo categórico como "supervisor" para crear los intervalos de forma que se conserve la fuerza de la asociación original entre los dos campos.

Por ejemplo, suponga que ha utilizado el análisis de clústeres para agrupar estados basados en los porcentajes de impago de préstamos hipotecarios, con los porcentajes más altos en el primer clúster. En este caso, podría seleccionar *Porcentaje en mora* y *Porcentaje de ejecuciones hipotecarias* como campos de intervalos, y el campo de pertenencia al clúster generado por el modelo, como campo supervisor.

Extensión de nombre: especifique una extensión para utilizarla en los campos generados y especifique si debe añadirse al inicio (**Prefijo**) o al final (**Sufijo**) del nombre de campo. Por ejemplo, puede generar un campo nuevo denominado `enmora_ÓPTIMO` y otro denominado `enejecución_ÓPTIMO`.

Campo supervisor: campo categórico empleado para construir los intervalos.

Procesar campos previamente para mejorar el rendimiento con conjuntos de datos grandes: indica si se debe realizar un procesamiento previo para simplificar los intervalos óptimos. Este método agrupa los valores de escala en un número mayor de intervalos mediante un sencillo método de intervalos sin supervisión, representa los valores dentro de cada intervalo mediante la media y ajusta la ponderación de casos en consecuencia antes de continuar con los intervalos supervisados. En términos prácticos, este método logra un equilibrio entre el grado de precisión y la velocidad, por lo que resulta recomendable para conjuntos de datos de gran tamaño. También puede especificar el número máximo de intervalos en que cualquier variable debe terminar después del preprocesamiento cuando se utiliza esta opción.

Fusionar los intervalos que tienen un recuento de casos relativamente pequeño con un vecino de mayor tamaño. Si está activada, indica que se ha fusionado un intervalo si el cociente de su tamaño (número de casos) entre el del intervalo del vecino es inferior al del umbral especificado; tenga en cuenta que los umbrales mayores pueden resultar en una fusión de más elementos.

Configuración de punto de corte

El cuadro de diálogo Configuración de punto de corte permite especificar opciones avanzadas para el algoritmo de intervalos óptimos. Estas opciones indican al algoritmo cómo calcular los intervalos utilizando el campo objetivo.

Límites de intervalos. Puede especificar si deben incluirse ($\text{inferior} \leq x$) o excluirse ($\text{inferior} < x$) los límites inferior y superior.

Primer y último intervalo. Tanto para el primero como para el último intervalo, puede especificar si los intervalos deben tener límites (puntos de datos superior e inferior) o no tenerlos (tendiendo hacia el infinito positivo o negativo).

Vista preliminar de intervalos generados

La pestaña Agrupar valores en intervalos del nodo Intervalos permite ver los umbrales de los intervalos generados. Mediante el menú Generar, también puede generar un nodo Derivar que puede utilizar para aplicar estos umbrales de un conjunto de datos a otro.

Campos en intervalos. Utilice la lista desplegable para seleccionar un campo que visualizar. Los nombres de campos mostrados utilizan el nombre de campo original para mayor claridad.

Cuantil. Utilice la lista desplegable para seleccionar un cuantil que visualizar, como 10 o 100. Esta opción está disponible sólo cuando los intervalos se generan al utilizar el método de mosaico (igual frecuencia o igual suma).

Umbrales de intervalo. Aquí se muestran los valores de umbral para cada intervalo generado, junto con el número de registros correspondiente a cada intervalo. Solamente en el método de intervalos óptimos, se muestra el número de registros de cada intervalo como porcentaje del total. Recuerde que los umbrales no tienen aplicación cuando se usa el método de intervalos de rangos.

Leer valores. Lee los valores agrupados en intervalos del conjunto de datos. Recuerde que los umbrales también se sobrescribirán cuando los datos nuevos se ejecuten en la ruta.

Generación de un nodo Derivar

Puede utilizar el menú Generar para crear un nodo Derivar basado en los umbrales actuales. Esto resulta útil para aplicar los umbrales de intervalo establecidos en un conjunto de datos a otro. Además, una vez se conocen estos puntos de división, la operación Derivar es más eficiente (más rápida) que una operación Intervalos al trabajar con conjuntos de datos grandes.

Nodo Análisis de RFM

El nodo Análisis de RFM (actualidad, frecuencia, monetario) permite determinar cuantitativamente qué clientes son los mejores examinando cuándo ha sido la compra más reciente de un cliente (actualidad), cuántas veces suele comprar (frecuencia) y cuánto gasta el cliente en todas las transacciones (valor monetario).

La razón del análisis de RFM es que los clientes que han adquirido el producto o servicio una vez tienen más posibilidades de volver a comprarlo. Los datos del cliente categorizado se separan en un número de intervalos, con los criterios ajustados a sus necesidades. En cada uno de los intervalos se asigna una puntuación a los clientes y las puntuaciones se combinan para proporcionar una puntuación RFM global. Esta puntuación es una representación de la pertenencia del cliente a los intervalos creados por cada uno de los parámetros RFM. Los datos de los intervalos pueden ser suficientes para satisfacer sus necesidades, por ejemplo, identificando a los clientes más frecuentes o de mayor valor; además, se pueden pasar a una ruta para su posterior modelado y análisis.

Tenga en cuenta, sin embargo, que aunque la posibilidad de analizar y clasificar puntuaciones RFM es una herramienta de gran utilidad, debe tener en cuenta diferentes factores a la hora de utilizarla. Puede tener la tentación de dirigirse a los clientes con las puntuaciones más altas; sin embargo, si se excede en esta actitud, puede provocar que el efecto contrario que opten por no volver a adquirir el producto. También cabe recordar que los clientes con menos puntuación no se deben despreciar, sino intentar que suban de categoría. Por el contrario, las puntuaciones más altas no son necesariamente un indicativo de mayores posibilidades de venta, dependiendo del mercado. Por ejemplo, un cliente en el intervalo 5 de actualidad (que ha adquirido un producto muy recientemente), no es el mejor objetivo para una empresa que venda productos de un alto precio o muy duraderos como coches o televisores.

Note: Dependiendo de cómo almacene sus datos, es posible que necesite ejecutar el nodo Análisis de RFM, junto con el nodo Adición de RFM para transformar los datos en un formato utilizable. Por

ejemplo, los datos de entrada deben estar en el formato del cliente, con una fila por cliente; si los datos del clientes están en formato de transacciones, utilice un nodo Adición de RFM anterior para derivar los campos de actualidad, frecuencia y monetario. Consulte el tema “Nodo Adición de RFM” en la página 82 para obtener más información.

Los nodos Adición de RFM y Análisis de RFM de IBM SPSS Modeler están configurados para utilizar intervalos independientes; es decir, clasifican y establecen en intervalos los datos de cada medida de actividades recientes, frecuencia y valor monetario, sin considerar sus valores o las otras dos medidas.

Configuración del nodo Análisis de RFM

Actualidad. Con el selector de campos (botón situado a la derecha del campo de texto), seleccione el campo de actualidad. Puede ser una fecha, marca de tiempo o un número simple. Tenga en cuenta que si una fecha o marca de tiempo representa la fecha de la transacción más reciente, el mayor valor se considera el más reciente; si se especifica un número, representa el tiempo transcurrido desde la transacción más reciente y el menor valor se considera el más reciente.

Note: Si el nodo Análisis de RFM está precedido en la ruta por un nodo Adición de RFM, los campos de actualidad, frecuencia y monetario generados por el nodo Adición de RFM de deben seleccionar como entradas en el nodo Análisis de RFM.

Frecuencia. Si utiliza el selector de campos, seleccione el campo de frecuencia que se utilizará.

Monetario. Si utiliza el selector de campos, seleccione el campo de monetario que se utilizará.

Número de intervalos. En cada uno de los tres tipos de entrada, seleccione el número de intervalos que se crearán. Toma como valor predeterminado 5.

Note: El número mínimo de intervalos es 2 y el máximo 9.

Ponderación. De forma predeterminada, se da mayor importancia a los datos recientes, seguidos de frecuencia y monetario, cuando calcula puntuaciones. Si fuera necesario, puede modificar la ponderación que afecta a uno o varios de los tipos de datos para modificar al que se le concede mayor importancia.

La puntuación RFM se calcula del siguiente modo: (puntuación de actualidad x ponderación de actualidad) + (puntuación de frecuencia x ponderación de frecuencia) + (puntuación de monetario x ponderación de monetario).

Empates. Especifique cómo se distribuirán por intervalos las puntuaciones idénticas (empataadas). Las opciones son:

- **Añadir al siguiente.** Seleccione esta opción para mover los valores de empate hacia el siguiente intervalo superior.
- **Mantener en el actual.** Mantiene los valores de empate en el intervalo actual (más bajo). Este método puede dar como resultado un menor número total de intervalos creados. (Éste es el valor predeterminado.)

Umbrales de intervalo. Especifique si las puntuaciones RFM y las asignaciones de umbrales se vuelven a computar siempre cuando se ejecuta el nodo, o que sólo se computan cuando sea necesario (por ejemplo, si se han añadido nuevos datos). Si selecciona **Leer de la pestaña Agrupar valores en intervalos si está disponible**, puede editar los puntos de corte superior e inferior de los diferentes intervalos en la pestaña Agrupar valores en intervalos.

Cuando se ejecuta, el nodo Análisis de RFM distribuye por intervalos los campos actualidad, frecuencia y monetarios brutos y añade los siguientes nuevos campos al conjunto de datos:

- Puntuación de actualidad. Un rango (valor de intervalo) para actualidad

- Puntuación de frecuencia. Un rango (valor de intervalo) para frecuencia
- Puntuación de monetario. Un rango (valor de intervalo) para monetario
- Puntuación de RFM. La suma ponderada de las puntuaciones de actualidad, frecuencia y monetario.

Añadir valores atípicos a intervalos finales. Si selecciona esta casilla de verificación, los registros por debajo del intervalo más inferior se añaden al intervalo inferior y los registros por encima, se añaden al intervalo superior; de otro modo, tendrán un valor nulo. Esta casilla solamente está disponible si selecciona **Leer de la pestaña Agrupar valores en intervalos si está disponible**.

Intervalos del nodo Análisis de RFM

La pestaña Agrupar valores permite visualizar y, en algunos casos, modificar los umbrales de los intervalos generados.

Note: Solamente puede modificar los valores de esta pestaña si selecciona **Leer de la pestaña Agrupar valores en intervalos si está disponible** en la pestaña Configuración.

Campos en intervalos. Utilice la lista desplegable para seleccionar un campo para dividir en valores. Los valores disponibles son los que seleccione de la pestaña Configuración.

Tabla de valores en intervalos. Aquí se muestran los valores de umbrales para cada intervalo generado. Si selecciona **Leer de la pestaña Agrupar valores en intervalos si está disponible** en la pestaña Configuración, puede modificar los puntos de corte superior e inferior de cada intervalo pulsando dos veces en la casilla correspondiente.

Leer valores. Lee valores de intervalos del conjunto de datos y completa los valores de intervalo de la tabla. Tenga en cuenta que si selecciona **Recalcular siempre** en la pestaña Configuración, los umbrales de los intervalos se sobrescribirán cuando los nuevos datos se ejecuten en la ruta.

Nodo Conjunto

El nodo Conjunto combina dos o más nugget de modelo para obtener predicciones más precisas que pueden conseguirse de los modelos individuales. Al combinar predicciones de varios modelos, pueden evitarse las limitaciones en modelos individuales que dan como resultado una precisión global superior. Los modelos combinados de esta forma suelen ejecutarse tan bien como el mejor de los modelos individuales y, en ocasiones, mejor.

Esta combinación de nodos ocurre de forma automática en los nodos de modelado automático Clasificador automático, Clúster automático y Autonumérico.

Después de utilizar un nodo Conjunto, puede utilizar un nodo Análisis o Evaluación para comparar la precisión de los resultados combinados con cada uno de los modelos de entrada. Para ello, asegúrese de que la opción **Filtrar campos generados por modelos de conjunto** no está seleccionada en la pestaña Configuración en el nodo Conjunto.

Campos de resultado

Cada nodo Conjunto genera un campo que contiene las puntuaciones combinadas. El nombre se basa en el campo objetivo especificado y con el prefijo $XF_$, $XS_$ o $XR_$, dependiendo del nivel de medición (marca, nominal (conjunto) o continuo (rango), respectivamente). Por ejemplo, si el objetivo es un campo de marca denominado *respuesta*, el campo de resultado sería $XF_{respuesta}$.

Confianza o campos de propensión. Para los campos nominales y marca, se crean campos de propensión o confianza adicionales en función del método de conjunto, tal y como se detalla en la tabla siguiente.

Tabla 27. Creación de campos del método de conjunto.

Método de conjunto	Nombre de campo
Voting Votación ponderada de confianza Votación ponderada de propensión en bruto Votación ponderada de propensión ajustada Mayor confianza gana	\$XFC_<campo>
Propensión en bruto media	\$XFRP_<campo>
Propensión en bruto ajustada media	\$XFAP_<campo>

Configuración de nodo Conjunto

Campo objetivo de conjunto. Seleccione un único campo que se utilice como objetivo por dos o más modelos anteriores. Los modelos anteriores pueden utilizar objetivos marca, nominales o continuos, pero al menos dos de estos modelos deben compartir el mismo objetivo para combinar las puntuaciones.

Filtrar campos generados por modelos de conjunto. Elimina los resultados de todos los campos adicionales que generan los modelos individuales que contiene el nodo Conjunto. Seleccione esta casilla de verificación si solamente está interesado en la puntuación combinada de todos los modelos de entrada. Asegúrese de que esta opción no está seleccionada si, por ejemplo, desea utilizar un nodo Análisis o Evaluación para comparar la precisión de la puntuación combinada con la de cada uno de los modelos de entrada individuales.

Las configuraciones disponibles dependen del nivel de medición del campo seleccionado como objetivo.

Objetivos continuos

En un objetivo continuo, se calculará el promedio de las puntuaciones. Es el único método disponible para combinar puntuaciones.

Al promediar promedios o estimaciones, el nodo Conjunto utiliza un cálculo de error estándar para calcular la diferencia entre los valores medidos o estimados y los valores true, y para mostrar si las estimaciones coinciden. Los cálculos de error estándar se generan para los nuevos modelos de forma predeterminada; sin embargo, puede cancelar la selección de la casilla de verificación de los modelos existentes, por ejemplo, si se van a regenerar.

Objetivos categóricos

Los objetivos categóricos admiten diferentes métodos, incluyendo **votación**, que funciona cuadrando el número de veces que cada posible valor predicho se elige y seleccionando el valor con mayor valor total. Por ejemplo, si tres de los cinco modelos predicen *sí* y las otras dos predicen *no, sí* gana por 3 a 2. Además, los votos se pueden **ponderar** en función del valor de confianza o propensión de cada predicción. Las ponderaciones se suman y el valor con el mayor total se vuelve a seleccionar. La confianza de la predicción final es la suma de las ponderaciones del valor seleccionado dividido por el número de modelos que se incluyen en el conjunto.

Todos los campos categóricos. Los campos marca y nominales admiten los métodos siguientes:

- Voting
- Votación ponderada de confianza
- Mayor confianza gana

Campos de marcas únicamente. En los campos de marca únicamente, también existen disponibles diferentes métodos según su propensión:

- Votación ponderada de propensión en bruto
- Votación ponderada de propensión ajustada
- Propensión en bruto media
- Propensión ajustada media

Empates de votaciones. En los métodos de votación puede especificar cómo se resuelven los empates.

- **Selección aleatoria.** Se seleccionará aleatoriamente uno de los valores empatados.
- **La mayor confianza.** El valor empatado predicho con el mayor índice de confianza gana. Tenga en cuenta que no es necesariamente la misma opción que la del índice de mayor confianza de todos los valores predichos.
- **Propensión en bruto o ajustada (campos con marcas únicamente).** El valor empatado predicho con la mayor propensión absoluta, donde la propensión absoluta se calcula como:

$$\frac{\text{abs}(0,5 - \text{propensión})}{2}$$

O, en el caso de la propensión ajustada:

$$\text{abs}(0,5 - \text{propensión ajustada}) * 2$$

Nodo Partición

Los nodos Partición se utilizan para generar un campo de partición que divide los datos en subconjuntos o muestras independientes para las fases de entrenamiento, comprobación y validación en la generación del modelo. Si usa una muestra para generar el modelo y otra muestra distinta para probarlo, podrá obtener una buena indicación de la bondad del modelo a la hora de generalizarlo a conjuntos de datos de mayor tamaño similares a los datos actuales.

El nodo Partición genera un campo nominal con el rol definido a **Partición**. Si lo prefiere, en el caso de que un campo adecuado ya exista en los datos, puede designarse como una partición utilizando un nodo Tipo. En este caso no se requiere ningún nodo Partición independiente. Se puede utilizar como partición cualquier campo nominal instanciado con dos o tres valores, pero no los campos con marcas. Consulte el tema “Definición del rol de campo” en la página 136 para obtener más información.

En una ruta se pueden definir múltiples campos de partición pero, de hacerlo, será necesario seleccionar un campo de partición simple en la pestaña Campos de cada nodo de modelado que utilice la partición. (Si solamente hay una partición, se usará automáticamente siempre que se active la partición.)

Activación de la partición. Para utilizar la partición en un análisis, ésta debe estar activada en la pestaña Opciones de modelo en el nodo de análisis o de generación de modelos adecuado. Si se anula esta opción, se posibilita la desactivación de la partición sin eliminar el campo.

Para crear un campo de partición en función de algún otro criterio como un rango de fechas o una ubicación, también puede utilizar un nodo Derivar. Consulte el tema “Nodo Derivar” en la página 142 para obtener más información.

Ejemplo. Cuando genere una ruta de RFM para identificar clientes recientes que han respondido de forma positiva a campañas de marketing anteriores, el departamento de marketing de una empresa de ventas utiliza un nodo Partición para dividir los datos en particiones de prueba y entrenamiento.

Opciones del nodo Partición

Campo de partición. Especifica el nombre del campo creado por el nodo.

Particiones. Puede dividir los datos en dos muestras (Entrenamiento y comprobación) o en tres (Entrenamiento, comprobación y validación).

- **Entrenamiento y comprobación.** Divide los datos en dos particiones de muestra, lo que permite entrenar el modelo con una muestra y comprobarlo con otra.
- **Entrenamiento, comprobación y validación.** Divide los datos en tres particiones de muestra, lo que permite entrenar el modelo con una muestra, comprobarlo y ajustarlo con otra y validar los resultados con otra. Así se reduce el tamaño de cada partición en consecuencia y puede ser más adecuado al trabajar con un conjunto de datos muy grande.

Tamaño de partición. Especifica el tamaño relativo de cada partición. Si la suma de los tamaños de partición es inferior al 100%, los registros no incluidos en una partición se descartarán. Por ejemplo, si un usuario tiene 10 millones de registros y tamaños de partición del 5% para el entrenamiento y el 10% para la prueba, después de ejecutar el nodo, debería haber unos 500.000 registros de entrenamiento y un millón de registros de prueba, con el resto descartado.

Valores. Especifica los valores utilizados para representar cada muestra de partición de los datos.

- **Utiliza valores definidos por el sistema (1", "2" y "3").** Utiliza un número entero para representar cada partición; por ejemplo, todos los registros que se encuentran en la muestra de entrenamiento tienen un valor de 1 para el campo de partición. Así se garantiza que los datos sean intercambiables entre configuraciones regionales y que si el campo de partición se reinstancia en otro lugar (por ejemplo, al leer los datos de una base de datos), se mantendrá el orden de clasificación (así, 1 seguirá representando la partición de entrenamiento). Sin embargo, los valores requieren alguna interpretación.
- **Añadir etiquetas a los valores definidos por el sistema.** Combina el número entero con una etiqueta; por ejemplo, los registros de partición de entrenamiento tienen un valor de 1_Entrenamiento. Así es posible identificar la identidad de los valores y mantener el orden de clasificación. Sin embargo, los valores son específicos de una configuración regional específica.
- **Utilizar etiquetas como valores.** Utiliza la etiqueta sin números enteros; por ejemplo, **Entrenamiento**. Esto permite especificar los valores al editar las etiquetas. Sin embargo, los datos de configuración se vuelven específicos de la configuración regional y, al volver a crear una instancia de una columna de partición, los valores se colocan en su orden natural, que puede no corresponder con su orden "semántico".

Semilla. Solo está disponible si se selecciona **Asignación de partición repetible**. Cuando se muestran o particionan registros a partir de un porcentaje aleatorio, esta opción permite duplicar los mismos resultados en otra sesión. Al especificar el valor inicial utilizado por el generador de números aleatorios, puede garantizar que se asignan los mismos registros cada vez que se ejecuta el nodo. Introduzca el valor de semilla deseado, o pulse en el botón **Generar** para generar automáticamente un valor aleatorio. Si no se selecciona esta opción, se generará una muestra diferente cada vez que se ejecute el nodo.

Nota: Cuando se utiliza la opción **Semilla** con registros leídos en una base de datos, es posible que antes del muestreo sea necesario un nodo Ordenar para garantizar el mismo resultado cada vez que se ejecuta el nodo. Esto se debe a que la semilla aleatoria depende del orden de registros, sin estar garantizado que sea el mismo en una base de datos relacional. Consulte el tema "Nodo Ordenar" en la página 83 para obtener más información.

Utilizar campo exclusivo para asignar particiones. Solo está disponible si se selecciona **Asignación de partición repetible**. (Solamente para bases de datos de nivel 1.) Marque este cuadro para utilizar retroacción SQL para asignar registros a particiones. En la lista desplegable, seleccione un campo con valores exclusivos (como un campo de ID) para garantizar que se asignen los registros de forma aleatoria pero repetible.

Los niveles de la base de datos se explican en la descripción del nodo de origen de base de datos. Consulte el tema "Nodo de origen Base de datos" en la página 17 para obtener más información.

Generación de nodos Seleccionar

Mediante el menú Generar del nodo Partición se puede generar automáticamente un nodo Seleccionar para cada partición. Por ejemplo, se pueden seleccionar todos los registros de la partición de prueba para obtener una evaluación o análisis más profundos al utilizar solamente esta partición.

Nodo Marcas

El nodo Marcas se utiliza para derivar campos de marcas en función de los valores categóricos definidos para uno o más campos nominales. Por ejemplo, su conjunto de datos puede contener un campo nominal, *PS* (presión sanguínea), con los valores *Alta*, *Normal* y *Baja*. Para manipular los datos de una forma más sencilla, puede crear un campo de marcas para la presión sanguínea alta, que indicará si el paciente tiene la presión sanguínea alta o no.

Opciones de configuración del nodo Marcas

Campos de conjuntos. Enumera todos los campos de datos con un nivel de medición de *Nominal* (conjunto). Seleccione uno de la lista para mostrar los valores en el conjunto. Puede elegir entre estos valores para crear un campo de marcas. Tenga en cuenta que, para poder ver los campos nominales disponibles (y sus valores), los datos se deben instanciar completamente utilizando un nodo Tipo o de origen en un punto anterior de la ruta. Consulte el tema "Nodo Tipo" en la página 125 para obtener más información.

Extensión de nombre de campo. Seleccione esta opción para activar los controles y especificar una extensión que añadir como sufijo o prefijo al nuevo campo de marcas. De forma predeterminada, los nombres de los campos nuevos se crean automáticamente combinando el nombre del campo original con el valor del campo en una etiqueta, como *Nombredecampo_valordecampo*.

Valores del conjunto disponibles. Aquí se muestran los valores del conjunto seleccionado anteriormente. Seleccione uno o varios valores para los que desea generar marcas. Por ejemplo, si los valores de un campo denominado *presión_sanguínea* son *Alta*, *Media* y *Baja*, puede seleccionar *Alta* y añadirla a la lista de la derecha. Así se creará un campo con una marca para los registros que tendrá un valor indicativo de alta presión sanguínea.

Crear campos de marcas. Aquí se enumeran los campos de marcas creados recientemente. Puede especificar opciones para nombrar el nuevo campo mediante los controles de extensión de nombre de campo.

Valor para Verdadero. Permite especificar el valor para verdadero utilizado por el nodo al configurar una marca. Este valor es *V* de forma predeterminada.

Valor para Falso. Permite especificar el valor para falso utilizado por el nodo al configurar una marca. Este valor es *F* de forma predeterminada.

Claves de agregación. Seleccione esta opción para agrupar registros en función de los campos clave especificados a continuación. Cuando se selecciona **Claves de agregación**, todos los campos de marcas de un grupo se "activan" si *cualquier* registro se ha configurado como verdadero. Utilice el selector de campos para especificar los campos clave que se utilizarán para agregar registros.

Nodo Reestructurar

El nodo Reestructurar puede usarse para generar varios campos a partir de los valores de un campo nominal o de marcas. Estos campos recién generados pueden contener valores de otro campo o marcas numéricas (0 y 1). La funcionalidad de este nodo es similar a la que posee el nodo Marcas. Sin embargo, ofrece mayor flexibilidad. Con él, podrá crear campos de cualquier tipo (incluidas marcas numéricas) a partir de los valores de otro campo. Tras ello, puede llevar a cabo una agregación o cualquier otra

manipulación con otros nodos a partir de ese punto de la ruta. (El nodo Marcas permite agregar campos en un paso, lo cual es conveniente si está creando campos de marcas.)

Por ejemplo, el siguiente conjunto de datos contiene un campo nominal, *Cuenta*, con los valores *Ahorros* y *Cheque*. Se registra el saldo de apertura y el actual de cada cuenta, y son varios los clientes que tienen varias cuentas de cada tipo. Imaginemos que desea saber si cada cliente tiene un tipo de cuenta concreto y, de ser así, la cantidad de dinero que contiene cada tipo de cuenta. Usará el nodo Reestructurar para generar un campo para cada uno de los valores de *Cuenta* y seleccionará *Saldo_actual* como el valor. Cada campo nuevo se rellenará con el saldo actual del registro correspondiente.

Tabla 28. Datos de muestra antes de la reestructuración.

IDclien	Cuenta	Saldo_ab	cheque_cuent
12701	Cheque	1000	1005,32
12702	Ahorros	100	144,51
12703	Ahorros	300	321,20
12703	Ahorros	150	204,51
12703	Cheque	1200	586,32

Tabla 29. Datos de muestra tras la reestructuración.

IDclien	Cuenta	Saldo_ab	cheque_cuent	Saldo_actual_ cheque_cuent	Saldo_actual_ ahorros_cuent
12701	Cheque	1000	1005,32	1005,32	\$null\$
12702	Ahorros	100	144,51	\$null\$	144,51
12703	Ahorros	300	321,20	\$null\$	321,20
12703	Ahorros	150	204,51	\$null\$	204,51
12703	Cheque	1200	586,32	586,32	\$null\$

Uso del nodo Reestructurar con el nodo Agregar

En muchas ocasiones, puede que quiera emparejar el nodo Reestructurar con un nodo Agregar. En el ejemplo anterior, un cliente (con el ID 12703) tiene tres cuentas. Puede usar un nodo Agregar para calcular el saldo total de cada tipo de cuenta. El campo clave es *IDclien* y los campos agregados son los nuevos campos reestructurados, *Saldo_actual_cheque_cuent* y *Saldo_actual_ahorros_cuent*. La siguiente tabla recoge los resultados.

Tabla 30. Datos de muestra tras la reestructuración y agregación.

IDclien	Recuento_registros	Suma_saldo_actual_ ahorros_cuent	Suma_saldo_actual_ ahorros_cuent
12701	1	1005,32	\$null\$
12702	1	\$null\$	144,51
12703	3	586,32	525,71

Opciones de configuración del nodo Reestructurar

Campos disponibles. Enumera todos los campos de datos con un nivel de medición de *Nominal* (conjunto) o *Marca*. Seleccione uno de la lista para mostrar los valores en el conjunto o marca y, a continuación, elija de entre estos valores para crear los campos reestructurados. Tenga en cuenta que, para poder ver los campos disponibles (y sus valores), los datos se deben instanciar completamente utilizando un nodo Tipo o un nodo de origen anterior de la ruta. Consulte el tema “Nodo Tipo” en la página 125 para obtener más información.

Valores disponibles. Aquí se muestran los valores del conjunto seleccionado anteriormente. Seleccione uno o varios valores para los que desea generar campos reestructurados. Por ejemplo, si los valores en un campo denominado *Presión sanguínea* son *Alta*, *Media* y *Baja*, puede seleccionar *Alta* y añadir esta opción a la lista de la derecha. De esta forma, se creará un campo con un valor especificado (consulte a continuación) para los registros con el valor *Alta*.

Crear campos reestructurados. Aquí se enumeran los campos reestructurados recién creados. De forma predeterminada, los nombres de los campos nuevos se crean automáticamente combinando el nombre del campo original con el valor del campo en una etiqueta, como *Nombredcampo_valordecampo*.

Incluir nombres de campo. Anule la selección para eliminar el nombre original del campo como prefijo de los nombres de los nuevos campos.

Usar valores de otros campos. Especifique uno o varios campos cuyos valores se vayan a usar para rellenar los campos reestructurados. Utilice el selector de campos para seleccionar uno o varios campos. Por cada campo seleccionado se creará un nuevo campo. El nombre del campo de valor se añadirá al nombre del campo reestructurado; por ejemplo, *PS_alta_edad* o *PS_baja_edad*. Todos los campos nuevos heredan el tipo del campo de valor original.

Crear marcas de valores numéricos. Seleccione para rellenar los nuevos campos con marcas de valores numéricos (0 para falso y 1 para verdadero), en lugar de usar un valor de otro campo.

Nodo Transponer

De forma predeterminada, las columnas son campos y las filas, registros u observaciones. En caso de que sea necesario, puede usarse un nodo Transponer para intercambiar los datos de las filas y columnas de forma que los campos se convierten en registros y los registros, en campos. Por ejemplo, si tiene datos de series temporales donde cada serie es una fila y no una columna, puede transponer los datos antes de proceder al análisis.

Opciones de configuración del nodo Transponer

Nombres de campos nuevos

Se pueden generar nuevos nombres de campo automáticamente a partir de un prefijo concreto, o bien se pueden leer desde un campo existente en los datos.

Usar prefijo. Con esta opción se generan nuevos nombres de campos a partir del prefijo especificado (*Campo1*, *Campo2*, etc.). Puede personalizar el prefijo como considere conveniente. Con esta opción debe indicar el número de campos que se va a crear, independientemente del número de filas en los datos originales. Por ejemplo, si **Número de campos nuevos** se establece en 100, se descartarán todos los datos que no estén contenidos en las 100 primeras filas. Si hay menos de 100 filas en los datos originales, algunos de los campos serán nulos. (Puede aumentar el número de campos según sea necesario, pero el propósito de esta configuración reside en evitar que millones de registros se transpongan a millones de campos, dado que con esto se obtendría un resultado imposible de administrar.)

Por ejemplo, suponga que tiene datos con series en filas y un campo diferente (columna) para cada mes. Puede transponerlos para que cada serie esté en un campo diferente, con una fila para cada mes.

Leer de campo. Lee nombres de campo de un campo existente. Con esta opción, el número de campos nuevos se establece a partir de los datos y hasta el máximo establecido. Cada valor del campo seleccionado se convierte en un nuevo campo en los datos de salida. El campo seleccionado admite cualquier tipo de almacenamiento (número entero, cadena, fecha, etc.), pero, para evitar nombres de campos duplicados, cada valor del campo seleccionado debe ser exclusivo (en otras palabras, el número de valores ha de coincidir con el número de filas). En caso de que existan nombres de campos duplicados, se mostrará una advertencia.

- **Leer valores.** Si el campo seleccionado no se ha instanciado, seleccione esta opción para rellenar la lista de nuevos nombres de campos. En caso de que el campo sí esté instanciado, este paso no será necesario.
- **Número máximo de valores para leer.** Al leer nombres de campos desde los datos, se especifica un límite superior con el fin de evitar que se genere un número desmesurado de campos. (Tal y como se ha indicado anteriormente, la transposición de un millón de registros en un millón de campos produciría un resultado imposible de manejar.)

Por ejemplo, si la primera columna de los datos especifica el nombre de cada serie, puede utilizar estos valores como nombres de campo en los datos transpuestos.

Transponer. De forma predeterminada, únicamente los campos continuos (rango numérico) se transponen (ya sean de número entero o de almacenamiento real). De forma opcional, puede elegir un subconjunto de campos numéricos o transponer campos de cadena en su lugar. No obstante, todos los campos transpuestos deben ser del mismo tipo de almacenamiento (numérico o de cadena, pero nunca ambos), dado que la mezcla de campos de entrada generaría valores mezclados en cada columna de resultados y, en consecuencia, se incumpliría la norma que establece que todos los valores de un campo deben tener el mismo almacenamiento. El resto de tipos de almacenamiento (fecha, hora, marca de tiempo) no puede transponerse.

- **Todos los numéricos.** Transpone todos los campos numéricos (número entero o almacenamiento real). El número de filas del resultado coincide con el número de campos numéricos en los datos originales.
- **Todas las cadenas.** Transpone todos los campos de cadena.
- **Personalizado.** Permite seleccionar un subconjunto de campos numéricos. El número de filas del resultado coincide con el número de campos seleccionado. *Note:* esta opción solamente se encuentra disponible para los campos numéricos.

Nombre ID de fila. Especifica el nombre del campo de ID de fila creado por el nodo. Los valores de este campo se establecen a partir de los nombres de los campos en los datos originales.

Sugerencia: al transponer datos de series temporales de filas a columnas, si los datos originales incluyen una fila (como fecha, mes o año) que etiqueta el período para cada medición, asegúrese de que estas etiquetas se leen en IBM SPSS Modeler como nombres de campo (tal y como se describe en los ejemplos anteriores, que muestran el mes o la fecha como nombres de campo en los datos originales, respectivamente) en lugar de incluir la etiqueta en la primera fila de datos. De esta forma evitará mezclar etiquetas y valores en cada columna (lo que forzaría que los números se leyesen como cadenas, ya que no se pueden mezclar tipos de almacenamiento dentro de una columna).

Nodo de intervalos de tiempo

El nodo Intervalos de tiempo permite especificar intervalos y generar etiquetas para los datos que se utilizarán en un nodo de modelado de series temporales, en un nodo Generación de análisis TS o en un nodo Gráfico de tiempo con fines de estimación o previsión. Se admite una gama completa de intervalos de tiempo que abarca desde segundos a años. Por ejemplo, si tiene una serie con mediciones diarias que empieza el 3 de enero de 2005, podrá etiquetar registros a partir de esa fecha, siendo la segunda fila el 4 de enero, y así sucesivamente. También puede indicar la periodicidad, por ejemplo, cinco días a la semana u ocho horas al día.

Además, puede especificar el rango de registros que se va a utilizar para la estimación. Puede elegir entre excluir los registros más recientes de la serie o especificar casos reservados. De esta forma, puede probar el modelo reservando los registros más recientes de los datos de series temporales para comparar sus valores conocidos con los valores estimados para dichos períodos.

También puede especificar cuántos períodos de tiempo desea prever en el futuro y los valores futuros que se van a utilizar en la previsión en nodos de modelado de serie temporal o en nodos Generación de análisis TS a partir de ese punto de la ruta.

El nodo Intervalos de tiempo genera un campo *EtiquetaTiempo* en el formato adecuado para el intervalo y el período, así como un campo *ÍndiceTiempo*, que asigna un número entero exclusivo a cada registro. Puede que además se generen varios campos adicionales según el intervalo o período seleccionado (como el minuto o el segundo al que corresponde una medición).

Puede rellenar o agregar valores según sea necesario con el fin de asegurarse de que las mediciones están espaciadas uniformemente. Los métodos para modelar datos de series temporales precisan de un intervalo uniforme entre cada medición. Los valores perdidos deben indicarse mediante filas vacías. Si sus datos no cumplen con este requisito, el nodo puede transformarlos para que lo hagan.

Comentarios

- Es posible que los intervalos periódicos no coincidan con el tiempo real. Así, por ejemplo, una serie basada en una semana de trabajo estándar compuesta por cinco días consideraría el intervalo entre viernes y lunes como un solo día.
- El nodo de intervalos de tiempo da por hecho que cada serie corresponde a un campo o una columna, con una fila por medición. Si procede, puede transponer los datos para cumplir este requisito. Consulte el tema “Nodo Transponer” en la página 169 para obtener más información.
- En cuanto a aquellas series que no están espaciadas uniformemente, puede especificar un campo que identifique la fecha o la hora de cada medición. Recuerde que para ello se necesita un campo de fecha, hora o marca de tiempo con el formato apropiado para utilizarlo como entrada. Si es preciso, puede convertir a este formato un campo existente (por ejemplo, un campo de cadena o de etiqueta) mediante un nodo Rellenar. Consulte el tema “Conversión del almacenamiento mediante el nodo Rellenar” en la página 150 para obtener más información.
- A la hora de ver los detalles de la etiqueta y los campos de índice generados, a veces resulta práctico activar la visualización de etiquetas de valores. Por ejemplo, al visualizar una tabla con valores generados para los datos mensuales, puede pulsar en el icono de etiquetas de valores de la barra de herramientas para ver *enero, febrero, marzo*, etc. en lugar de *1, 2, 3*, etc.



Figura 6. Icono de etiquetas de valores

Especificación de intervalos de tiempo

Con la pestaña Intervalos podrá indicar el intervalo y la periodicidad para crear o etiquetar las series. La configuración particular depende del intervalo seleccionado: Así, si selecciona **Horas por día**, podrá establecer el número de días a la semana, el día en el que cada semana empieza, el número de horas al día y la hora a la que cada día empieza. Consulte el tema “Intervalos admitidos” en la página 175 para obtener más información.

Etiquetado o creación de la serie

Puede etiquetar los registros de manera consecutiva o crear la serie en función de una fecha, marca de tiempo o campo de tiempo concretos.

- **Iniciar etiquetado desde el primer registro.** Especifique la fecha y/o hora de inicio para etiquetar registros consecutivos. Si, por ejemplo, va a etiquetar las horas por día, especificaría la fecha y la hora a la que empezaría la serie, con un solo registro por cada hora sucesiva. Aparte de añadir etiquetas, este método no altera los datos originales. En su lugar, supone que los registros están espaciados uniformemente, con un intervalo uniforme entre cada medición. Cualquier medición perdida deberá indicarse mediante filas vacías en los datos.
- **Generar a partir de los datos.** En cuanto a aquellas series que no están espaciadas uniformemente, puede especificar un campo que identifique la fecha o la hora de cada medición. Recuerde que para ello se necesita un campo de fecha, hora o marca de tiempo con el formato apropiado para utilizarlo como entrada. Por ejemplo, si tiene un campo de cadena con valores como *Ene 2000, Feb 2000*, etc.,

puede convertirlo en un campo de fecha mediante el nodo Rellenar. Consulte el tema “Conversión del almacenamiento mediante el nodo Rellenar” en la página 150 para obtener más información. La opción **Generar a partir de los datos** también transforma los datos para que coincidan con el intervalo especificado rellenando o agregando registros cuando proceda; así, acumulando semanas en meses o sustituyendo los registros perdidos por espacios en blanco o valores extrapolados. Puede especificar las funciones utilizadas para rellenar o agregar registros en la pestaña Generar. Consulte el tema “Opciones de generación en intervalos de tiempo” para obtener más información.

Nueva extensión de nombre de campo. Permite especificar un prefijo o un sufijo que se aplica a todos los campos que el nodo genera. Por ejemplo, si se usa el prefijo predeterminado *\$TI*, los campos que el nodo cree se denominarán *\$TI_ÍndiceTiempo*, *\$TI_EtiquetaTiempo*, etc.

Formato de fecha. Especifica el formato para el campo *EtiquetaTiempo* que el nodo ha creado según proceda con respecto al intervalo actual. La disponibilidad de esta opción depende de la selección actual.

Formato de hora. Especifica el formato para el campo *EtiquetaTiempo* que el nodo ha creado según proceda con respecto al intervalo actual. La disponibilidad de esta opción depende de la selección actual.

Opciones de generación en intervalos de tiempo

Con la pestaña Generar del nodo Intervalos de tiempo podrá especificar opciones para agregar o rellenar campos de manera que coincidan con el intervalo concreto. Esta configuración se aplica únicamente si se selecciona la opción **Generar a partir de los datos** de la pestaña Intervalos. Por ejemplo, si tiene una mezcla de datos semanales y mensuales, podría agregar o "acumular" los valores semanales con el fin de obtener un intervalo mensual uniforme. Otra opción consiste en establecer el intervalo en semanal y rellenar la serie insertando valores en blanco para las semanas que faltan, o bien extrapolando los valores que faltan mediante una función de relleno específica.

Cuando se rellenan o agregan datos, los campos de marca de tiempo o de fecha que existan se sustituyen eficazmente con los campos *EtiquetaTiempo* e *ÍndiceTiempo* generados y se eliminan del resultado. Los campos sin tipos se eliminan igualmente. Se conservan aquellos campos que miden el tiempo como una duración (como por ejemplo un campo que mide la duración de una llamada de servicio y no la hora a la que la llamada se inició), siempre y cuando se almacenen internamente como campos de tiempo y no de marca de tiempo. Consulte el tema “Configuración de almacenamiento de campos y formato” en la página 9 para obtener más información. Otros campos se agregan en función de las opciones especificadas en la pestaña Generar.

- **Usar funciones y campos predeterminados.** Especifica que todos los campos han de agregarse o rellenarse según sea necesario, a excepción de los campos sin tipo, los de marca de tiempo y los de fecha, tal y como se ha indicado anteriormente. La función predeterminada que se aplica depende del nivel de medición; por ejemplo, los campos continuos se agregarán utilizando la media, mientras que los campos nominales se agregarán utilizando la moda. El valor predeterminado para uno o varios niveles de medición puede cambiarse en la parte inferior del cuadro de diálogo.
- **Especificar funciones y campos.** Permite especificar los campos que van a rellenarse o agregarse, así como la función empleada para ambas acciones. Los campos que no se hayan seleccionado se eliminarán del resultado. Use los iconos de la derecha para añadir o eliminar campos de la tabla, o bien pulse en la casilla de la columna pertinente para cambiar la función de agregación o relleno utilizada para sustituir la función predeterminada. Los campos sin tipo se excluyen de la lista y no se pueden añadir a la tabla.

Valores predeterminados. Especifica las funciones de agregación y relleno que se usan de forma predeterminada para los distintos tipos de campos. Estos valores predeterminados se aplican cuando se selecciona **Usar funciones y campos predeterminados** y, asimismo, constituye la opción predeterminada inicial para los nuevos campos que se añaden a la tabla. (Si se cambian los valores predeterminados, la configuración existente en la tabla no se verá alterada, pero sí afectará a los campos que se añadan con posterioridad.)

Función de agregación. Las funciones de agregación disponibles son:

- **Continuo.** Las funciones disponibles para los campos continuos incluyen **Media, Suma, Moda, Mín y Máx.**
- **Nominal.** Las opciones incluyen **Moda, Primeros y Último.** Primeros indica los primeros valores no nulos (ordenados por fecha) en el grupo de agregación, mientras que Último señala el último valor no nulo.
- **Marca.** Las opciones incluyen **Verdadero si alguno es verdadero, Moda, Primeros y Último.**

Función de relleno. Las funciones de relleno disponibles son:

- **Continuo.** Las opciones incluyen **Vacío y Media de los puntos recientes,** que indica los tres valores no nulos más recientes antes del período de tiempo que se va a crear. Si no existen tres valores, el nuevo estará vacío. Los valores recientes sólo incluyen valores reales. Un valor relleno creado previamente no se tendrá en cuenta en la búsqueda de un valor no nulo.
- **Nominal. Vacío y Valor más reciente.** Con más reciente se hace referencia al valor no nulo más reciente antes del período de tiempo que se va a crear. De nuevo, solamente se tendrán en cuenta los valores reales en la búsqueda de un valor reciente.
- **Marca.** Las opciones incluyen **Vacío, Verdadera y Falsa.**

Número máximo de registros del conjunto de datos resultante. Especifica un límite máximo de número de registros creados, ya que, de lo contrario, podría ser muy elevado, en concreto cuando el intervalo de tiempo se establece en segundos (ya sea deliberadamente o no). Por ejemplo, una serie con sólo dos valores (1 ene, 2000 y 1 ene, 2001) generaría 31.536.000 registros si se rellenan en segundos (60 segundos x 60 minutos x 24 horas x 365 días). El sistema detendrá el proceso y mostrará una advertencia en caso de que se supere el máximo especificado.

Campo Recuento

Cuando se agregan o rellenan valores, se crea un campo *Recuento* nuevo que indica el número de registros implicados en la definición del nuevo registro. Si, por ejemplo, se han agregado valores semanales en un solo mes, el recuento sería 4. En el caso de un registro relleno, el recuento es 0. El nombre del campo es *Recuento* más el prefijo o sufijo especificado en la pestaña Intervalos.

Período de estimación

La pestaña Estimación del nodo Intervalos de tiempo permite especificar el intervalo de registros utilizado en la estimación del modelo, además de los casos reservados. Esta configuración se puede sobrescribir en los nodos de modelado posteriores según sea necesario, pero es conveniente especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Comenzar estimación. Puede comenzar el período de estimación al principio de los datos o excluir los valores antiguos que tal vez no sean muy útiles para las previsiones. Dependiendo de los datos, puede que al reducir el período de estimación se acelere el rendimiento (y disminuya el tiempo dedicado a la preparación de los datos) sin que suponga una pérdida significativa en la precisión de las previsiones.

Terminar estimación. Puede estimar el modelo utilizando todos los registros hasta el final de los datos, o "reservar" los registros más recientes para evaluar el modelo. En el caso siguiente, en realidad está "haciendo previsiones" valores que ya conoce, lo que permite comparar los valores observados y predictivos para evaluar la eficacia del modelo.

Previsiones

La pestaña Previsión del nodo Intervalos de tiempo permite especificar el número de registros que se desea predecir y los valores futuros que se van a utilizar en la previsión mediante nodos de modelado de serie temporal o nodos Generación de análisis TS a partir de ese punto de la ruta. Esta configuración se

puede sobrescribir en los nodos posteriores según sea necesario, pero puede resultar más cómodo especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Extender registros en el futuro. Especifica el número de registros para los que se harán previsiones más allá del período de estimación. Tenga en cuenta que estos registros pueden ser o no "previsiones" según el número de reservas que se haya especificado en la pestaña Estimación.

Campo de indicador de futuro. Etiqueta del campo generado que indica si un registro contiene datos de previsiones. El valor predeterminado de la etiqueta es *\$TI_Futuro*.

Valores futuros que se utilizarán en la previsión. Si utiliza campos predictores (Dirección = Entrada), para cada registro que desee prever (con el rol definido a *Entrada*) deberá especificar los valores estimados para el período de predicción para cada predictor. Puede introducir los valores manualmente o seleccionarlos de una lista.

- **Campo.** Pulse en el botón selector de campos y seleccione los campos que se utilizarán como predictores. Tenga en cuenta que los campos aquí seleccionados podrán utilizarse o no en el modelado; para utilizar un campo como predictor debe seleccionarse en un nodo de modelado posterior. Este cuadro de diálogo permite especificar los valores futuros de forma cómoda para que puedan compartirlos varios nodos de modelado posteriores sin tener que especificarlos por separado en cada nodo. Recuerde también que la lista de campos disponibles puede restringirse mediante las selecciones realizadas en la pestaña Generar. Por ejemplo, si se selecciona **Especificar funciones y campos** en la pestaña Generar, se eliminarán de la ruta todos los campos que no se hayan agregado o rellenado y no podrán utilizarse en el modelado.

Note: si se especifican valores futuros para un campo que ya no está disponible en la ruta (debido a que se ha eliminado o a las selecciones actualizadas realizadas en la pestaña Generar) el campo se mostrará en rojo en la pestaña Previsión.

- **Valores.** Para cada campo, puede seleccionar en una lista de funciones o pulsar en **Especificar** para introducir los valores manualmente o desde una lista de valores predefinidos. Si los campos predictores se refieren a los elementos que están bajo su control, o que se conocen de otra forma, debe introducir los valores manualmente. Por ejemplo, si está haciendo previsiones de los ingresos del próximo mes para un hotel a partir del número de reservas, puede especificar el número de reservas que tiene para ese período. Por el contrario, si un campo predictor está relacionado con algo que escapa a su control, como el precio de las acciones, puede utilizar una función como el valor más reciente o la media de los puntos recientes.

Las funciones disponibles dependen del nivel de medición del campo.

Tabla 31. Funciones disponibles para los niveles de medición.

Nivel de medición	Funciones
Campo continuo o nominal	Vacío Media de los puntos recientes Valor más reciente Especifica
Campo de marcas	Vacío Valor más reciente True False Especifica

Media de los puntos recientes: calcula el valor futuro a partir de la media de los tres últimos puntos de datos.

Valor más reciente: establece el valor futuro en el punto de datos más reciente.

Verdadero/falso: establece el valor futuro de un campo de marcas en Verdadero o Falso según lo especificado.

Especificar: abre un cuadro de diálogo que permite especificar los valores futuros manualmente o desde una lista predefinida.

Valores futuros

Aquí puede especificar valores futuros para utilizarlos en las previsiones realizadas en nodos de modelado de series temporales o en nodos de generación de análisis TS en puntos posteriores de la ruta. Esta configuración se puede sobrescribir en los nodos posteriores según sea necesario, pero puede resultar más cómodo especificarla aquí en lugar de hacerlo para cada nodo individualmente.

Puede introducir los valores manualmente o pulsar en el botón selector situado en el lado derecho del cuadro de diálogo para seleccionarlos de una lista de valores definida para el campo actual.

El número de valores futuros que se puede especificar corresponde al número de registros para los que está extendiendo la serie temporal en el futuro.

Intervalos admitidos

El nodo Intervalos de tiempo admite una amplia variedad de intervalos, desde segundos a años, así como períodos cíclicos (por ejemplo, estacionales) y no cíclicos. Puede especificar el intervalo en el campo Intervalos de tiempo, en la pestaña Intervalos.

Períodos

Seleccione **Períodos** para etiquetar una serie no cíclica existente que no coincida con ninguno de los otros intervalos especificados. La serie debe hallarse ya en el orden correcto, con un intervalo uniforme entre cada medición. La opción **Generar a partir de los datos** no estará disponible si se selecciona este intervalo.

Resultado de muestra

Los registros se etiquetan de forma incremental en función del valor de inicio especificado (*Período 1*, *Período 2*, etc.). Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 32. Ejemplos de creación de campos nuevos

\$TI_ÍndiceTiempo (Entero)	\$TI_EtiquetaTiempo (Cadena)	\$TI_Período (Entero)
1	Período 1	1
2	Periodo 2	2
3	Periodo 3	3
4	Periodo 4	4
5	Periodo 5	5

Períodos cíclicos

Seleccione **Períodos cíclicos** para etiquetar una serie existente con un ciclo repetitivo que no se ajuste a uno de los intervalos estándar. Por ejemplo, podría usar esta opción si tiene solamente 10 meses en el año fiscal. La serie debe hallarse ya en el orden correcto, con un intervalo uniforme entre cada medición. (La opción **Generar a partir de los datos** no estará disponible si se selecciona este intervalo.)

Resultado de muestra

Los registros se etiquetan de forma incremental en función del período y ciclo de inicio especificados (*Ciclo 1, Período 1, Ciclo 1, Período 2*, etc.). Por ejemplo, si el número de períodos por ciclo está definido como 3, se crearán nuevos campos como se muestra en la tabla siguiente.

Tabla 33. Ejemplos de creación de campos nuevos.

\$TI_ÍndiceTiempo (Entero)	\$TI_EtiquetaTiempo (Cadena)	\$TI_Ciclo (Entero)	\$TI_Período (Entero)
1	Ciclo 1, Período 1	1	1
2	Ciclo 1, Período 2	1	2
3	Ciclo 1, Período 3	1	3
4	Ciclo 2, Período 1	2	1
5	Ciclo 2, Período 2	2	2

Años

En relación con los años, se puede especificar el año de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para especificar un campo de marca de tiempo o de fecha que identifique el año de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 34. Ejemplos de creación de campos nuevos

\$TI_ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)
1	2000	2000
2	2001	2001
3	2002	2002
4	2003	2003
5	2004	2004

Trimestres

En el caso de una serie trimestral, puede especificar el mes en el que comienza el año fiscal. Asimismo, puede indicar el año o el trimestre de inicio (así, T1 2000) para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de fecha que identifique el trimestre y el año de cada registro.

Resultado de muestra

En relación con un año fiscal que comienza en enero, se crearían y rellenarían campos nuevos como se muestra en la tabla siguiente:

Tabla 35. Ejemplos de creación de campos nuevos

\$TI_ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)	\$TI-Trimestre (Entero con etiquetas)
1	T1 2000	2000	1 (T1)
2	T2 2000	2000	2 (T2)
3	T3 2000	2000	3 (T3)
4	T4 2000	2000	4 (T4)
5	T1 2001	2001	1 (T1)

En caso de que el año comience en un mes que no sea enero, los nuevos campos serán los que se muestran en la tabla siguiente (suponiendo que un año fiscal empieza en julio). Para ver las etiquetas que

identifican los meses de cada trimestre, active la visualización de las etiquetas de los valores pulsando en el icono de la barra de herramientas.



Figura 7. icono de etiquetas de valores

Tabla 36. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Año (Entero)	\$TI-Trimestre (Entero con etiquetas)
1	T1 2000/2001	1	1 (T1 Jul-Sep)
2	T2 2000/2001	1	2 (T2 Oct-Dic)
3	T3 2000/2001	1	3 (T3 Ene-Mar)
4	T4 2000/2001	1	4 (T4 Abr-Jun)
5	T1 2001/2002	2	1 (T1 Jul-Sep)

Meses

Puede seleccionar el año y mes de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de fecha que señale el mes de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 37. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Año (Entero)	\$TI-Meses (Entero con etiquetas)
1	Enero de 2000	2000	1 (Enero)
2	Feb 2000	2000	2 (Febrero)
3	Mar 2000	2000	3 (Marzo)
4	Abr 2000	2000	4 (Abril)
5	May 2000	2000	5 (Mayo)

Semanas (no periódicas)

Para una serie semanal, puede seleccionar el día de la semana en el que desea que empiece el ciclo.

Tenga en cuenta que esta opción solamente puede ser no periódica, ya que diferentes meses, trimestres o años, no tienen necesariamente el mismo número de semanas. Si embargo, es muy sencillo agregar o rellenar datos con marcas de tiempo en un nivel semanal para modelos no periódicos.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 38. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Semana (Entero)
1	1999-12-27	1
2	2000-01-03	2

Tabla 38. Ejemplos de creación de campos nuevos (continuación)

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Semana (Entero)
3	2000-01-10	3
4	2000-01-17	4
5	2000-01-24	5

El campo \$TI-EtiquetaTiempo para una semana muestra el primer día de la semana. En la tabla anterior, el usuario inicia el etiquetado el 1 de enero de 2000. Sin embargo, la semana empieza en Lunes, y el 1 de enero de 2000 es Sábado. Así, la semana que incluye el 1 de enero empieza el 27 de diciembre de 1999, y es la etiqueta del primer punto.

El formato de fecha determina las cadenas generadas para el campo \$TI-EtiquetaTiempo.

Días por semana

En el caso de las mediciones por días que se encuentran dentro de un ciclo semanal, puede especificar el número de días a la semana, así como el día en que cada semana empieza. Puede especificar la fecha de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de fecha que señale la fecha de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 39. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)	\$TI-Semana (Entero)	\$TI-Día (Entero con etiquetas)
1	Ene 5 2005	1	3 (Miércoles)
2	Ene 6 2005	1	4 (Jueves)
3	Ene 7 2005	1	5 (Viernes)
4	Ene 10 2005	2	1 (Lunes)
5	Ene 11 2005	2	2 (Martes)

Nota: la semana siempre comienza en 1 para el primer período de tiempo y no sigue un ciclo en función del calendario. Por lo tanto, a la semana 52 le sigue la semana 53, 54, etc. La semana no refleja la semana del año, sino simplemente el número de incrementos semanales que la serie contiene.

Días (no periódicos)

Escoja días no periódicos en caso de que tenga mediciones diarias que no se ajusten a un ciclo semanal regular. Puede especificar la fecha de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de fecha que señale la fecha de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 40. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)
1	Ene 5 2005
2	Ene 6 2005

Tabla 40. Ejemplos de creación de campos nuevos (continuación)

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Fecha)
3	Ene 7 2005
4	Ene 8 2005
5	Ene 9 2005

Horas por día

En el caso de las mediciones por horas que se ajusten a un ciclo diario, puede especificar el número de días a la semana, el de horas al día (como un día laboral de ocho horas, por ejemplo), el día en que la semana comienza y la hora a la que cada día empieza. Las horas pueden especificarse incluso por minutos de acuerdo a un reloj de 24 horas (por ejemplo, 14:05 = 2:05 p.m.).

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 41. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Marca de tiempo)	\$TI-Día (Entero con etiquetas)	\$TI-Hora (Entero con etiquetas)
1	Ene 5 2005 8:00	3 (Miércoles)	8 (8:00)
2	Ene 5 2005 9:00	3 (Miércoles)	9 (9:00)
3	Ene 5 2005 10:00	3 (Miércoles)	10 (10:00)
4	Ene 5 2005 11:00	3 (Miércoles)	11 (11:00)
5	Ene 5 2005 12:00	3 (Miércoles)	12 (12:00)

Horas (no periódicas)

Escoja esta opción en caso de que tenga mediciones por horas que no se ajusten a un ciclo diario regular. Puede especificar la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

Las horas se basan en un reloj de 24 horas (13:00 = 1:00 p.m.) y no se basan en ciclos de 24 horas (es decir, a la hora 24 le sigue la hora 25).

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 42. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Hora (Entero con etiquetas)
1	8:00	8 (8:00)
2	9:00	9 (9:00)
3	10:00	10 (10:00)
4	11:00	11 (11:00)
5	12:00	12 (12:00)

Minutos por día

En el caso de las mediciones por minuto que se encuentren dentro de un ciclo diario, puede especificar el número de días por semana, el día en el que empieza la semana, el número de horas al día y la hora a la que el día comienza. Las horas se especifican de acuerdo a un reloj de 24 horas y pueden especificarse incluso por minutos y segundos utilizando dos puntos (por ejemplo, 2:05:17 p.m. = 14:05:17). También puede especificar el número de minutos de incremento (cada minuto, cada dos minutos, etc., donde el incremento debe ser un valor múltiplo de 60).

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 43. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Marca de tiempo)	\$TI-Minuto
1	2005-01-05 08:00:00	0
2	2005-01-05 08:01:00	1
3	2005-01-05 08:02:00	2
4	2005-01-05 08:03:00	3
5	2005-01-05 08:04:00	4

Minutos (no periódicos)

Escoja esta opción en caso de que tenga mediciones por minuto que no se ajusten a un ciclo diario regular. Puede especificar el número de minutos de incremento (cada minuto, cada dos minutos, etc., donde el valor especificado debe ser múltiplo de 60).

Puede especificar la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 44. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Minuto
1	8:00	0
2	8:01	1
3	8:02	2
4	8:03	3
5	8:04	4

- La cadena *EtiquetaTiempo* se crea intercalando dos puntos entre la hora y los minutos. Los ciclos no son de 24 horas; es decir, a la hora 24 le sigue la hora 25.
- Los minutos se incrementan según el valor especificado en el cuadro de diálogo. Por ejemplo, si el incremento es 2, la *EtiquetaTiempo* será 8:00, 8:02, etc., mientras que los minutos serán 0, 2, y así sucesivamente.

Segundos por día

En el caso de los intervalos en segundos que se encuentren dentro de un ciclo diario, puede especificar el número de días por semana, el día en el que empieza la semana, el número de horas al día y la hora a la

que el día comienza. Las horas se determinan de acuerdo a un reloj de 24 horas y pueden especificarse incluso por minutos y segundos utilizando dos puntos (por ejemplo, 2:05:17 p.m. = 14:05:17). Asimismo, puede especificar el número de segundos de incremento (cada segundo, cada dos segundos, etc., donde el valor especificado debe ser múltiplo de 60).

Puede especificar la fecha y la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo que señale la fecha y hora de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 45. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Marca de tiempo)	\$TI-Minuto	\$TI-Segundo
1	2005-01-05 08:00:00	0	0
2	2005-01-05 08:00:01	0	1
3	2005-01-05 08:00:02	0	2
4	2005-01-05 08:00:03	0	3
5	2005-01-05 08:00:04	0	4

Segundos (no periódicos)

Escoja esta opción en caso de que tenga mediciones por segundo que no se ajusten a un ciclo diario regular. Puede especificar el número de segundos de incremento (cada segundo, cada dos segundos, etc., donde el valor especificado debe ser múltiplo de 60).

Especifique la hora de inicio para etiquetar registros consecutivos o seleccionar **Generar a partir de los datos** para elegir un campo de marca de tiempo o de hora que señale la hora de cada registro.

Resultado de muestra

Se crean campos nuevos, como se muestra en la tabla siguiente.

Tabla 46. Ejemplos de creación de campos nuevos

\$TI-ÍndiceTiempo (Entero)	\$TI-EtiquetaTiempo (Cadena)	\$TI-Minuto	\$TI-Segundo
1	8:00:00	0	0
2	8:00:01	0	1
3	8:00:02	0	2
4	8:00:03	0	3
5	8:00:04	0	4

- La cadena *EtiquetaTiempo* se crea intercalando dos puntos entre la hora y los minutos y entre los minutos y los segundos. Los ciclos no son de 24 horas; es decir, a la hora 24 le sigue la hora 25.
- Los segundos aumentan según el número que se haya especificado como incremento. Si el incremento es 2, la *EtiquetaTiempo* será 08:00:00, 08:00:02, etc., mientras que los segundos serán 0, 2, y así sucesivamente.

Nodo Historial

Los nodos Historial se suelen utilizar para los datos secuenciales, como los datos de series temporales. Se utilizan para crear campos nuevos que contienen datos de los campos de registros anteriores. Al utilizar un nodo Historial, es posible que desee tener los datos ordenados previamente por un campo determinado. Puede utilizar un nodo Ordenar para hacerlo.

Opciones de configuración del nodo Historial

Campos seleccionados. El selector de campos (botón situado a la derecha del campo de texto) permite seleccionar campos de los que obtener un historial. Cada campo seleccionado se utiliza para crear campos nuevos para todos los registros del conjunto de datos.

Desplazamiento. Permite especificar el último registro anterior al registro actual desde el que desea extraer valores del campo histórico. Por ejemplo, si el valor de Desplazamiento es 3, a medida que cada registro pasa por este nodo, los valores del campo del tercer registro anterior se incluirán en el registro actual. Utilice la opción Amplitud para especificar el límite hasta el que se pueden extraer registros. Utilice las flechas para ajustar el valor de desplazamiento.

Amplitud. Permite especificar la cantidad de registros anteriores de los que desea extraer valores. Por ejemplo, si se establece Desplazamiento en 3 y Amplitud en 5, cada registro que pase por el nodo tendrá cinco campos añadidos para cada campo especificado en la lista Campos seleccionados. Esto significa que cuando el nodo procesa el registro 10, los campos se añaden desde el registro 7 al 3. Utilice las flechas para ajustar el valor de amplitud.

Cuando no esté disponible el historial. Permite seleccionar una de las siguientes opciones para tratar los registros que no tengan valores de historial. Suele hacer referencia a los primeros registros de la parte superior del conjunto de datos, de los que no hay registros previos que utilizar como historial.

- **Descartar registros.** Seleccione esta opción para descartar registros donde no existe ningún valor de historial disponible para el campo seleccionado.
- **Dejar historial sin definir.** Seleccione esta opción para conservar registros donde no existe ningún valor de historial disponible. El campo historial se rellenará con un valor no definido que se muestra como \$null\$.
- **Rellenar valores con.** Permite especificar un valor o cadena que utilizar en el caso de los registros en los que no existen valores de historial disponibles. El valor de sustitución predeterminado es *undef*, el valor nulo del sistema. Los valores nulos se muestran con la cadena \$null\$.

Al seleccionar un valor de sustitución, tenga en cuenta las siguientes reglas para realizar una ejecución correcta.

- Los campos seleccionados deben ser del mismo tipo de almacenamiento.
- Si todos los campos seleccionados tienen un almacenamiento numérico, el valor de sustitución debe analizarse como un número entero.
- Si todos los campos seleccionados tienen un almacenamiento real, el valor de sustitución debe analizarse como un número real.
- Si todos los campos seleccionados tienen un almacenamiento simbólico, el valor de sustitución debe analizarse como una cadena.
- Si todos los campos seleccionados tienen un almacenamiento de fecha/hora, el valor de sustitución debe analizarse como un campo de fecha/hora.

Si ninguna de las condiciones anteriores se cumple, recibirá un error al ejecutar el nodo Historial.

Nodo Reorg. campos

El nodo Reorg. campos permite definir el orden natural utilizado para mostrar los campos en la parte posterior de la ruta. Este orden afecta a la visualización de los campos en diversas ubicaciones, como las tablas, las listas y el selector de campos. Esta operación resulta útil, por ejemplo, al trabajar con conjuntos de datos amplios para aumentar la visibilidad de los campos de interés.

Opciones de configuración de Reorg. campos

Existen dos opciones de reorganización de los campos: orden personalizado y ordenación automática.

Orden personalizado

Seleccione **Orden personalizado** para activar una tabla de nombres de campos y tipos donde pueda ver todos los campos y utilizar los botones de flechas para crear un orden personalizado.

Para reordenar campos:

1. Seleccione un campo de la tabla. Mantenga pulsada la tecla Ctrl mientras selecciona en otro campo para seleccionar series de campos.
2. Utilice los botones de flechas sencillas para desplazar los campos una fila arriba o abajo.
3. Utilice los botones de flecha para mover los campos a la parte superior o inferior de la lista.
4. Especifique el orden de los campos no incluidos desplazando la fila separadora hacia arriba o hacia abajo (con el texto [otros campos]).

Más información en sobre [otros campos]

Otros campos. El objetivo de la fila separadora [otros campos] es dividir la tabla en dos mitades.

- Los campos que aparecen por encima de la fila separadora se ordenarán (como aparecen en la tabla) en la parte superior de todos los órdenes naturales utilizados para mostrar los campos hacia abajo de este nodo.
- Los campos que aparecen por debajo de la fila separadora se ordenarán (como aparecen en la tabla) en la parte inferior de todos los órdenes naturales utilizados para mostrar los campos hacia abajo de este nodo.

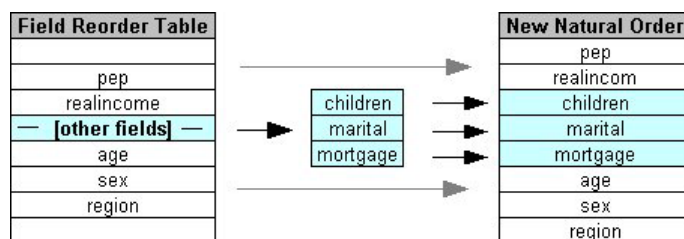


Figura 8. Diagrama que ilustra cómo se incorporan "otros campos" al nuevo orden de campos

- Los demás campos que no aparecen en la tabla de reorganización de campos se incluyen entre los campos "superior" e "inferior" tal y como queda indicado por la ubicación de la fila separadora.

Las opciones de ordenación personalizada adicionales son:

- Ordenar los campos en orden ascendente o descendente pulsando en las flechas ubicadas sobre cada cabecera de columna (**Tipo**, **Nombre** y **Almacenamiento**). Al ordenar por columna, los campos que no están especificados aquí (indicados por la fila [otros campos]) se ordenan en último lugar en orden natural.

- Pulsar en **Borrar no utilizados** para eliminar todos los campos no utilizados del nodo Reorg. campos. Los campos no utilizados se muestran en la tabla con letras de color rojo. Esto indica que el campo se ha eliminado en las operaciones anteriores.
- Especificar orden para cualquier campo nuevo (se muestra con un icono luminoso para indicar un campo nuevo o sin especificar). Al pulsar en **Aceptar** o **Aplicar**, el icono desaparece.

Note: si los campos se añaden hacia arriba (parte anterior de la ruta) tras aplicar un orden personalizado, los campos nuevos se añaden al final de la lista personalizada.

Ordenación automática

Seleccione **Ordenación automática** para especificar un parámetro de ordenación. Las opciones del cuadro de diálogo cambian de forma dinámica para proporcionar opciones de ordenación automática.

Ordenar por. Permite seleccionar una de las tres formas de ordenar los campos leídos en el nodo Reorg. campos. Los botones de flecha indican si el orden será ascendente o descendente. Seleccione uno para realizar un cambio.

- Nombre
- Tipo
- Almacenamiento

Los campos del nodo Reorg. campos añadidos anteriormente tras haber aplicado la ordenación automática se colocarán automáticamente en su posición correcta en función del tipo de ordenación seleccionado.

Nodo Intervalos de tiempo de AS

El nodo Intervalos de tiempo original en SPSS Modeler no es compatible con Analytic Server (AS). El nodo Intervalos de tiempo de AS (una novedad en SPSS Modeler release 17.0) contiene un subconjunto de las funciones del nodo Intervalos de tiempo existente que se puede utilizar con Analytic Server.

Utilice el nodo Intervalos de tiempo de AS para especificar intervalos y derivar un nuevo campo de tiempo para la estimación o el pronóstico. Se da soporte a un rango completo de intervalos de tiempo, desde segundos a años.

Utilice el nodo para derivar un nuevo campo de tiempo. El nuevo campo tiene el mismo tipo de almacenamiento que el campo de tiempo de entrada que seleccione. El nodo genera los siguientes elementos:

- El campo especificado en el separador Campos como el **Campo de tiempo**, junto con el prefijo o sufijo seleccionado. De forma predeterminada, el sufijo es \$TI_.
- Los campos especificados en el separador Campos como los **Campos de dimensión**.
- Los campos especificados en el separador Campos como los **Campos para agregar**.

También se pueden generar varios campos adicionales según el intervalo o periodo seleccionado, tal como el minuto o el segundo al que corresponde una medición.

Intervalo de tiempo de AS: Opciones de campos

Utilice el separador Campos del nodo Intervalos de tiempo de AS para seleccionar los datos de los que se derivará el intervalo de tiempo nuevo.

Campos: Muestra todos los campos de entrada para el nodo con sus iconos de tipo de medición. Todos los campos de tiempo tienen un tipo de medición "continuo". Seleccione el campo que se ha de utilizar como entrada.

Campo de tiempo: Muestra el campo de entrada del que se derivará el intervalo de tiempo nuevo. Solo se permite un único campo continuo. El campo los utiliza el nodo Intervalos de tiempo como la clave de agregación para convertir el intervalo. El nuevo campo tiene el mismo tipo de almacenamiento que el campo de tiempo de entrada seleccionado. Si selecciona un campo de entero, se considera que se trata de un índice de tiempo.

Campos de dimensiones: Opcionalmente, puede añadir campos aquí para crear series de tiempo individuales basadas en los valores de los campos. Por ejemplo, en el caso de los datos geoespaciales puede utilizar un campo de punto como una dimensión. En este ejemplo, la salida de datos del nodo Intervalos de tiempo de AS se clasifica en series de tiempo para cada valor de punto en el campo de punto.

Campos para agregar: Seleccione los campos que se han de agregar como parte de la modificación del campo de tiempo. Solo los campos que seleccione aquí estarán disponibles en el separador Generar para la tabla **Valores personalizados para los campos especificados**. Los campos que no se incluyan se filtrarán y no contendrán los datos de salida del nodo. Esto significa que cualquier campo que permanezca en la lista **Campos** se filtrará y no incluirá datos.

Intervalo de tiempo de AS: Opciones de generación

Utilice el separador Generar para especificar las opciones para cambiar el intervalo de tiempo y el modo en que se agregan campos en los datos, en función de su tipo de medición.

Cuando agrega los datos, cualquier campo de fecha, hora o indicación de la hora existente se sustituyen por los campos generados y se descartan en el resultado. Los otros campos se agregan en función de las opciones que especifique en este separador.

Intervalo de tiempo: Seleccione el intervalo y la periodicidad para generar las series. Si desea obtener más información, consulte “Intervalos admitidos” en la página 175.

Valores predeterminados: Seleccione la agregación predeterminada que se aplicará a los datos de tipos diferentes. El valor predeterminado que se aplicará depende del nivel de medición. Por ejemplo, los campos continuos se agregarán mediante la suma, mientras que los campos nominales se agregarán utilizando el modo. Puede establecer el valor predeterminado para 3 niveles de medición diferentes:

- **Continuo:** Las funciones disponibles para los campos continuos son **Suma, Media, Mín, Máx, Mediana, Primer cuartil** y **Tercer cuartil**.
- **Nominal:** Las opciones son **Modo, Mín** y **Máx**.
- **Marca:** Las opciones son **Verdadero si alguno es verdadero** o **Falso si alguno es falso**.

Valores especificados para los campos especificados: Puede especificar excepciones para los valores de agregación predeterminados para campos individuales. Use los iconos de la derecha para añadir o eliminar campos de la tabla, o pulse la casilla de la columna pertinente para cambiar la función de agregación de dicho campo. Los campos sin tipo se excluyen de la lista y no se pueden añadir a la tabla.

Extensión de nombre para campos nuevos: Especifique el **Prefijo** o el **Sufijo** que se aplicará a todos los campos generados por el nodo.

Nodo Reproyección

Cuando se trabaja con datos geoespaciales o datos de mapa, dos de las formas más habituales que se utilizan para identificar coordenadas son los sistemas de coordenadas proyectadas y los sistemas de coordenadas geográficas. Dentro de IBM SPSS Modeler, elementos tales como las funciones espaciales del creador de expresiones, el nodo STP (Spatio-Temporal Prediction) y el nodo de visualización de mapa utilizan el sistema de coordenadas proyectadas. Por tanto, cualquier dato importado que se registre con un sistema de coordenadas geográficas se debe re proyectar. Cuando es posible, los campos geoespaciales (campos con un nivel de medición geoespacial) se re proyectan automáticamente cuando se utilizan (no

cuando se importan). Cuando un campo no se pueda reproyectar automáticamente, utilice el nodo **Reproyectar** para cambiar el sistema de coordenadas del campo. Cuando la reproyección se realiza de esta manera, puede corregir situaciones en las que se produce un error debido al uso de un sistema de coordenadas incorrecto.

La lista siguiente muestra situaciones de ejemplo en las que puede ser necesario reproyectar para cambiar el sistema de coordenadas:

- *Añadir* Si intenta añadir dos conjuntos de datos con sistemas de coordenadas diferentes para un campo geoespacial, SPSS Modeler muestra el mensaje de error siguiente: Los sistemas de coordenadas de <Campo1> y <Campo2> no son compatibles. Reproyecte uno o ambos campos para utilizar el mismo sistema de coordenadas.
<Campo1> y <Campo2> son los nombres de los campos geoespaciales que provocaron el error.
- *Expresión if/else* Si utiliza una expresión que contiene una sentencia if/else con campos geoespaciales o tipos de retorno en ambas partes de la expresión, pero con sistemas de coordenadas diferentes, SPSS Modeler muestra el mensaje de error siguiente: La expresión condicional contiene sistemas de coordenadas incompatibles: <arg1> y <arg2>.
<arg1> y <arg2> son los argumentos de then o else que devuelven un tipo geoespacial con sistemas de coordenadas diferentes.
- *Crear una lista de campos geoespaciales* Para crear una lista de numerosos campos geoespaciales, todos los argumentos de campo geoespacial que se proporcionan a la expresión de lista deben estar en el mismo sistema de coordenadas. De lo contrario, se muestra este mensaje de error: Los sistemas de coordenadas de <Campo1> y <Campo2> no son compatibles. Reproyecte uno o ambos campos para utilizar el mismo sistema de coordenadas.

Para obtener más información sobre sistemas de coordenadas, consulte el tema "Establecer opciones geoespaciales para secuencias" en la sección "Utilización de secuencias" de la Guía del usuario de SPSS Modeler.

Establecer opciones para el nodo **Reproyectar**

Campos

Campos geoespaciales

De forma predeterminada, esta lista está vacía. Puede mover campos geoespaciales a esta lista desde la lista **Campos para reproyectar** para asegurarse de que esos campos no se reproyecten.

Campos para reproyectar

De forma predeterminada, esta lista contiene todos los campos geoespaciales que son datos de entrada para este nodo. Todos los campos de esta lista se reproyectan al sistema de coordenadas que define en el área **Sistema de coordenadas**.

Sistema de coordenadas

Valor predeterminado de secuencia

Seleccione esta opción para utilizar el sistema de coordenadas predeterminado.

Especificar

Si selecciona esta opción, puede utilizar el botón **Cambiar** para abrir el cuadro de diálogo **Seleccionar sistema de coordenadas** y elegir el sistema de coordenadas que se debe utilizar para la reproyección.

Para obtener más información sobre sistemas de coordenadas, consulte el tema "Establecer opciones geoespaciales para secuencias" en la sección "Utilización de secuencias" de la Guía del usuario de SPSS Modeler.

Capítulo 5. Nodos Gráfico

Características comunes de nodos Gráfico

Muchas fases del proceso de minería de datos utilizan gráficos y diagramas para explorar los datos introducidos en IBM SPSS Modeler. Por ejemplo, puede conectar un nodo Gráfico o Distribución a un origen de datos para comprender mejor los tipos de datos y las distribuciones. A continuación puede llevar a cabo manipulaciones de registros y campos para preparar los datos para las operaciones de modelado posteriores de la ruta. Otro uso común de los gráficos consiste en comprobar la distribución y las relaciones entre campos recién derivados.

La paleta Gráficos contiene los siguientes nodos:



El nodo Tablero ofrece muchos tipos diferentes de gráficos en un único nodo. Con este nodo puede seleccionar los campos de datos que desee explorar y seleccionar un gráfico de los disponibles para los datos seleccionados. El nodo filtra automáticamente cualquier tipo de gráfico que no sea compatible con las selecciones de campo.



El nodo Gráfico muestra la relación entre los campos numéricos. Puede crear un gráfico mediante puntos (un diagrama de dispersión) o líneas.



El nodo Distribución muestra las instancias de valores simbólicos (categóricos), como el tipo de hipoteca o el género. Normalmente, podría usar el nodo Distribución para mostrar los desequilibrios de los datos, que pueden rectificarse mediante el nodo Equilibrar antes de crear un modelo.



El nodo Histograma muestra las instancias de valores de los campos numéricos. Se suele utilizar para explorar los datos antes de las manipulaciones y la generación de modelos. Al igual que con el nodo Distribución, con frecuencia el nodo Histograma detecta desequilibrios en los datos.



El nodo Colección muestra la distribución de valores de un campo numérico relativo a los valores de otro. (Crea gráficos parecidos a los histogramas.) Es útil para ilustrar una variable o un campo cuyos valores cambian con el tiempo. Con los gráficos 3D también puede incluir un eje simbólico que muestra las distribuciones por categoría.



El nodo G. múltiple crea un gráfico que muestra varios campos Y sobre un campo X único. Los campos Y están representados como líneas coloreadas; cada uno equivale a un nodo Gráfico con el estilo establecido en **Línea** y el Modo para X establecido en **Ordenar**. Los gráficos múltiples son útiles cuando quiere explorar la fluctuación de varias variables a través del tiempo.



El nodo Malla muestra la fuerza de las relaciones entre los valores de dos o más campos simbólicos (categóricos). El gráfico utiliza líneas de varios anchos para indicar la fuerza de la conexión. Podría utilizar un nodo Malla, por ejemplo, para explorar las relaciones existentes entre la compra de un conjunto de elementos en un sitio de comercio electrónico.



El nodo Gráfico de tiempo muestra uno o más conjuntos de datos de series temporales. Normalmente, primero se utilizaría un nodo Intervalos de tiempo para crear un campo *EtiquetaTiempo*, que se utilizaría para etiquetar el eje *x*.



El nodo Evaluación ayuda a evaluar y comparar modelos predictivos. El diagrama de evaluación muestra la calidad con que los modelos predicen resultados particulares. Ordena registros en función del valor predicho y la confianza de la predicción. Divide el registro en grupos de igual tamaño (**cuantiles**) y, a continuación, representa el valor del criterio de negocio de cada cuantil de mayor a menor. El gráfico muestra múltiples modelos como líneas independientes.



El nodo Visualización de mapas puede aceptar varias conexiones de entrada y visualizar datos geospaciales en un mapa como una serie de capas. Cada capa es un único campo geoespacial; por ejemplo, la capa base podría ser el mapa de un país, y por encima puede tener una capa de carreteras, una capa de ríos y una capa de ciudades.

Una vez añadido un nodo de gráficos a una ruta, puede pulsar dos veces en el nodo para abrir un cuadro de diálogo para especificar opciones. La mayoría de los gráficos contienen un número de opciones exclusivas presentadas en una o varias pestañas. También existen varias opciones de pestaña comunes a todos los gráficos. Los siguientes secciones contienen más información acerca de estas opciones comunes.

Una vez configuradas las opciones de un nodo de gráficos, puede ejecutarlo desde el cuadro de diálogo o como parte de una ruta. En la ventana del gráfico que se genera, puede crear nodos Derivar (Conjunto y Marca) y Seleccionar basados en una selección o región de datos, estableciendo los datos en subconjuntos de forma eficaz. Por ejemplo, puede utilizar esta potente característica para identificar y excluir valores atípicos.

Aspectos estéticos, superposiciones, paneles y animación

Superposiciones y aspectos estéticos

Los aspectos estéticos (y superposiciones) añaden dimensionalidad a una visualización. El efecto de un aspecto estético (agrupación, agrupación en clústeres o apilación) depende del tipo de visualización, el tipo de campo (variable) y el tipo y estadístico del elemento gráfico. Por ejemplo, puede utilizarse un campo categórico de color para agrupar puntos en un diagrama de dispersión o para crear las pilas de un gráfico de barras apiladas. También puede utilizarse un rango numérico continuo de color para indicar los valores del rango para cada punto de un diagrama de dispersión.

Debería experimentar con los aspectos estéticos y superposiciones para descubrir qué cubre sus necesidades. Las siguientes descripciones pueden ayudarle a elegir lo correcto.

Note: No todos los aspectos estéticos o superposiciones están disponibles para todos los tipos de visualización.

- **Color.** Cuando un campo categórico define el color, divide la visualización basándose en las categorías individuales, correspondiendo cada color a una categoría distinta. Cuando el color es un rango numérico continuo, el color varía basándose en el valor del campo de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un registro o caso y se utiliza un campo de rango para el color, éste varía dependiendo de la *media* del campo de rango.
- **Forma.** La forma se define a través de un campo categórico que divide la visualización en elementos de formas diferentes, uno por cada categoría.
- **Transparencia.** Cuando un campo categórico define la transparencia, divide la visualización basándose en las categorías individuales, correspondiendo cada nivel de transparencia a cada categoría. Cuando la transparencia es un rango numérico continuo, la transparencia varía basándose en el valor del campo

de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un registro o caso y se utiliza un campo de rango para la transparencia, el color varía dependiendo de la *media* del campo de rango. En el valor más alto, los elementos gráficos son totalmente transparentes. En el valor más bajo, son totalmente opacos.

- **Etiqueta de datos.** Las etiquetas de datos se definen mediante cualquier tipo de campo cuyos valores se utilicen para crear etiquetas que se adjunten a los elementos gráficos.
- **Tamaño.** Cuando un campo categórico define el tamaño, divide la visualización basándose en las categorías individuales, correspondiendo cada tamaño a una categoría distinta. Cuando el tamaño es un rango numérico continuo, el tamaño varía basándose en el valor del campo de rango. Si el elemento gráfico (por ejemplo, una barra o un cuadro) representa a más de un registro o caso y se utiliza un campo de rango para el tamaño, éste varía dependiendo de la *media* del campo de rango.

Paneles y animación

Paneles. La opción Paneles, también conocida como facetas, crea una tabla de gráficos. Se genera un gráfico por cada categoría en los campos de paneles, pero todos los paneles aparecerán simultáneamente. Paneles es de utilidad para comprobar si la visualización está sujeta a las condiciones de los campos de paneles. Por ejemplo, puede incluir un histograma en un panel según género para determinar si las distribuciones de frecuencia son iguales entre hombres y mujeres. Es decir, puede comprobar si el salario está sujeto a diferencias de sexo. Seleccione un campo categórico para los paneles.

Animación. La opción Animación se parece a la opción Paneles en cuanto a que se crean varios gráficos a partir de los valores del campo de animación, pero estos gráficos no se muestran juntos. Más bien, utiliza los controles en el modo de exploración para animar el resultado y desplazarse por una secuencia de gráficos individuales. Además, a diferencia de la opción Paneles, Animación no necesita un campo categórico. Puede especificar un campo continuo cuyos valores estén divididos en rangos automáticamente. Puede variar el tamaño del rango con los controles de animación en el modo de exploración. No todas las visualizaciones ofrecen animación.

Uso de la pestaña Resultados

En el caso de todos los tipos de gráficos, puede especificar las siguientes opciones para el nombre de archivo y la visualización de los gráficos generados.

Note: Los gráficos de nodo Distribución tienen una configuración adicional.

Nombre del resultado. Especifica el nombre del gráfico generado cuando se ejecuta el nodo. **Automático** selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar **Personalizado** para especificar un nombre diferente.

Salida a pantalla. Seleccione esta opción para generar y mostrar el gráfico en una nueva ventana.

Salida a archivo. Seleccione para guardar el resultado como un archivo.

- **Gráfico de resultados.** Seleccione para producir resultados en un formato de gráfico. Disponibles sólo en los nodos Distribución.
- **Tabla de resultados.** Seleccione para producir resultados en un formato de tabla. Disponibles sólo en los nodos Distribución.
- **Nombre de archivo.** Permite especificar un nombre de archivo para el gráfico o tabla generados. Utilice el botón de puntos suspensivos (...) para especificar una ubicación y un archivo específicos.
- **Tipo de archivo.** Especifique el tipo de archivo de la lista desplegable. Para todos los nodos de gráfico, excepto el nodo Distribución con una opción **Tabla de resultados**, los tipos de archivo de gráficos disponibles son los siguientes:
 - Mapa de bits (.*bmp*)
 - PNG (.*png*)

- Objeto Resultado (.cou)
- JPEG (.jpg)
- HTML (.html)
- Documento ViZml (.xml) para utilizarlo en otras aplicaciones de IBM SPSS Statistics.

Para la opción **Tabla de resultados** en el nodo Distribución, los tipos de archivo disponibles son los siguientes:

- Datos delimitados por tabulaciones (.tab)
- Datos delimitados por comas (.csv)
- HTML (.html)
- Objeto Resultado (.cou)

Paginar resultados. Al guardar la salida como HTML, esta opción se activa para permitir controlar el tamaño de cada página HTML. (Se aplica solamente al nodo Distribución.)

Líneas por página. Cuando se selecciona **Paginar resultados**, esta opción se activa para permitir determinar la longitud de cada página HTML. La configuración predeterminada es de 400 líneas. (Se aplica solamente al nodo Distribución.)

Uso de la pestaña Anotaciones

Se utiliza para todos los nodos, esta pestaña proporciona opciones para cambiar el nombre de los nodos, ofrecer herramientas personalizadas y almacenar una anotación más extensa.

Gráficos 3D

Los gráficos y los gráficos de colección de IBM SPSS Modeler tienen la capacidad de mostrar información en un tercer eje. Así se dispone de una mayor flexibilidad para visualizar los datos para seleccionar subconjuntos o derivar campos nuevos para el modelado.

Una vez haya creado un gráfico 3D, puede pulsar en él y arrastrar el ratón para girarlo y poder verlo desde cualquier ángulo.

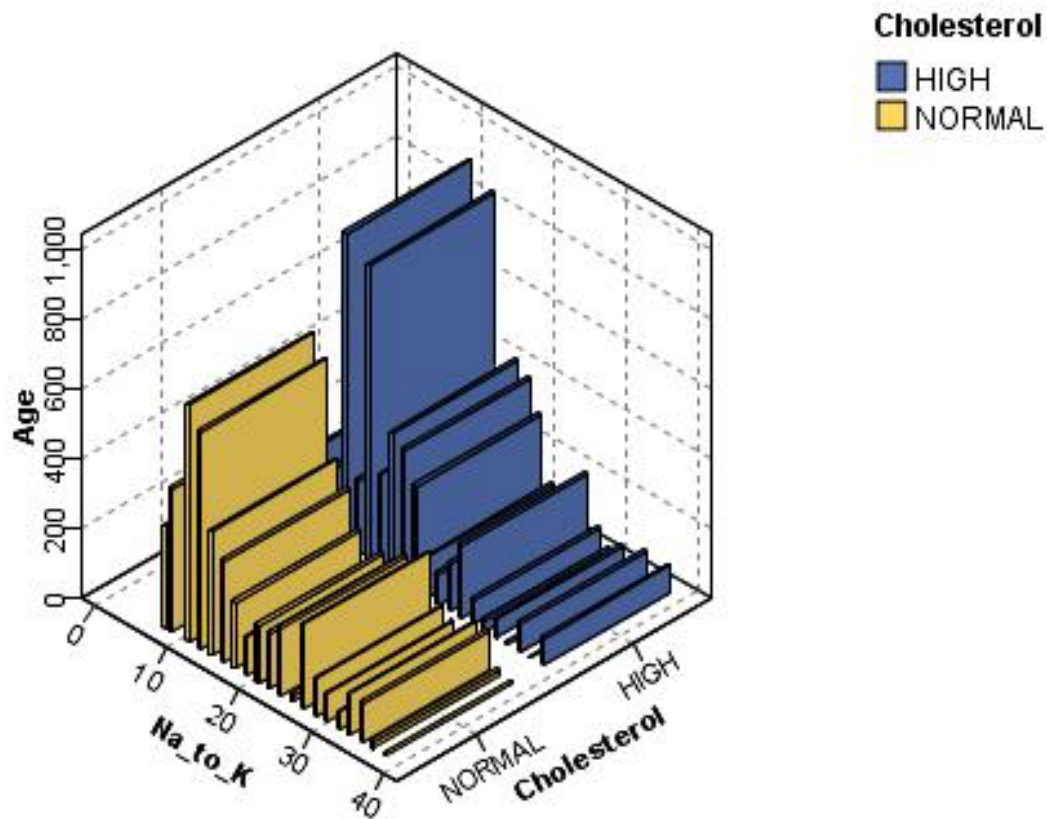


Figura 9. Gráfico de colección con ejes x, y, z

Existen dos maneras de crear gráficos 3-D en IBM SPSS Modeler: representar información en un tercer eje (gráficos 3D verdaderos) y mostrar gráficos con efectos 3D. Ambos métodos están disponibles para gráficos y colecciones.

Para representar información en un tercer eje

1. En el cuadro de diálogo del nodo de gráficos, pulse en la pestaña **Gráfico**.
2. Pulse en el botón 3D para activar las opciones del eje z.
3. Utilice el botón de elección de campos para seleccionar un campo para el eje z. En algunos casos solamente se permiten los campos simbólicos. El selector de campos mostrará los campos adecuados.

Para añadir efectos 3D a un gráfico

1. Una vez creado un gráfico, pulse en la pestaña **Gráfico** en la ventana de resultados.
2. Pulse en el botón 3D para cambiar la vista a un gráfico tridimensional.

Nodo Tablero

El nodo Tablero permite seleccionar distintos resultados de gráficos (gráficos de barras, gráficos circulares, histogramas, diagramas de dispersión, mapas de calor, etc.) en un único nodo. Empieza en la primera pestaña seleccionando los campos de datos que desea explorar y, a continuación, el nodo le proporciona una selección de tipos de gráficos que se adaptan a sus datos. El nodo filtra automáticamente cualquier tipo de gráfico que no sea compatible con las selecciones de campo. Puede definir opciones de gráfico más avanzadas o detalladas en la pestaña Detallado.

Note: Debe conectar el nodo Tablero a una ruta con los datos para poder editar el nodo o seleccionar tipos de gráficos.

Existen dos botones que permiten controlar las plantillas de visualización (y hojas de estilo y mapas) disponibles:

Administrar. Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Consulte el tema “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219 para obtener más información.

Ubicación. Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Consulte el tema “Configuración de la ubicación de las plantillas, hojas de estilo y mapas” en la página 218 para obtener más información.

Tablero Pestaña Básico

Si no está seguro de qué tipo de visualización representará mejor sus datos, utilice la pestaña Básico. Cuando seleccione sus datos, aparecerá un subconjunto de tipos de visualizaciones adecuados para los datos. Para obtener ejemplos, consulte “Ejemplos de Tableros ” en la página 206.

1. Seleccione uno o más campos (variables) de la lista. Mantenga pulsada la tecla Ctrl para seleccionar varios campos.
Tenga en cuenta que el nivel de medición del campo determina el tipo de visualizaciones disponibles. Puede cambiar el nivel de medición pulsando con el botón derecho en el campo de la lista y seleccionando una opción. Para obtener más información sobre los tipos de nivel de medición disponibles, consulte “Tipos de campos (variables)” en la página 194.
2. Seleccione un tipo de visualización. Para obtener descripciones de los tipos disponibles, consulte “Tipos de visualización de Graphboard preincorporado disponibles” en la página 198.
3. En el caso de ciertas visualizaciones, puede seleccionar un estadístico de resumen. Hay diferentes subconjuntos de estadísticos disponibles dependiendo de si el estadístico está basado en recuentos o si se ha calculado a partir de un campo continuo. Los estadísticos disponibles también dependen de la misma plantilla. Una lista completa de los estadísticos que pueden estar disponibles sigue al paso siguiente.
4. Si desea definir más opciones, como los campos de aspectos estéticos y paneles opcionales, pulse en **Detallado**. Consulte el tema “Tablero Pestaña Detallado” en la página 196 para obtener más información.

Estadísticos de resumen calculados a partir de un campo continuo

- *Media.* Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- *Mediana.* Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

- *Moda*. El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.
- *Mínimo*. Se trata del valor menor de una variable numérica.
- *Máximo*. Se trata del valor mayor de una variable numérica.
- *Rango*. Diferencia entre los valores mínimos y máximos.
- *Medio intervalo*. La mitad del rango, es decir, el valor cuya diferencia con el mínimo es igual a la diferencia con el máximo.
- *Suma*. Suma o total de todos los valores, a lo largo de todos los casos con valores no perdidos.
- *Suma acumulada*. La suma acumulada de los valores. Cada elemento del gráfico muestra la suma de un subgrupo más la suma total de los grupos anteriores.
- *Porcentaje de la suma*. El porcentaje en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos.
- *Suma del porcentaje acumulado*. El porcentaje acumulativo en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.
- *Varianza*. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media dividida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- *Desviación estándar*. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- *Error estándar*. Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación estándar de la distribución muestral de un estadístico. Por ejemplo, el error estándar de la media es la desviación estándar de las medias muestrales.
- *Curtosis*. Es una medida del grado en que las observaciones se agrupan en torno a un punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.
- *Asimetría*. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Los siguientes estadísticos de región pueden resultar en más de un elemento gráfico por subgrupo. Si utiliza los elementos de intervalo, área, o gráficos de borde, un estadístico de región da como resultado un elemento gráfico que muestra el rango. El resto de elementos gráficos dan como resultado dos elementos diferentes, uno que muestra el inicio del rango y el otro que muestra el final del rango.

- **Región: Rango**. El rango entre los valores mínimo y máximo.
- **Región: 95% de intervalo de confianza para la media**. Un rango de valores con un 95% de opciones de incluir la media poblacional.
- **Región: 95% de intervalo de confianza individual**. Un rango de valores con un 95% de opciones de incluir el valor predictivo especificando el caso individual.
- **Región: 1 Desviación estándar por encima/debajo de la media**. Un rango de valores entre 1 *desviación estándar* por encima y por debajo de la *media*.

- **Región: 1 Error estándar por encima/debajo de la media.** Un rango de valores entre 1 *error estándar* por encima y por debajo de la *media*.

Estadísticos de resumen basados en recuentos

- **Recuento.** Número de filas/casos.
- **Recuento acumulado.** El número acumulado de filas/casos. Cada elemento del gráfico muestra el recuento de un subgrupo más el recuento total de los grupos anteriores.
- **Porcentaje de recuento.** El porcentaje de filas/casos de cada subgrupo en comparación con el número total de filas/casos.
- **Porcentaje acumulado de recuento.** El porcentaje acumulado de filas/casos de cada subgrupo en comparación con el número total de filas/casos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.

Tipos de campos (variables)

Los iconos que aparecen junto a los campos en las listas de campos para indicar el tipo de campo y los tipos de datos. Los iconos también identifican múltiples conjuntos de respuestas.

Tabla 47. Iconos de nivel de medición.

Nivel de medición	Numérico	Cadena	Fecha	Hora
Continuo		n/d		
Conjunto ordenado				
Definir				

Tabla 48. Iconos de conjunto de respuestas múltiples.

Tipo de conjunto de respuestas múltiples	Icono
Conjuntos de respuestas múltiples, categorías múltiples	
Conjunto de respuestas múltiples, dicotomías múltiples	

Nivel de medición

Un nivel de medición de campo es importante cuando crea una visualización. A continuación se incluye una descripción de los niveles de medición. Puede cambiar temporalmente el nivel de medición pulsando con el botón derecho en un campo en una lista de campos y seleccionando una opción. En la mayoría de los casos, debe considerar únicamente las dos clasificaciones más amplias de los campos, categóricas y continuas:

Catagóricas. Datos con un número limitado de valores o categorías distintos (por ejemplo, sexo o religión). Los campos catagóricos pueden ser campos de cadena (alfanuméricos) o campos numéricos que utilizan códigos numéricos para representar categorías (por ejemplo, 0 = *hombre* y 1 = *mujer*). También se denominan datos cualitativos. Los conjuntos, conjuntos ordenados y marcas son campos catagóricos.

- *Establecer*. Un campo/variable cuyos valores representan categorías que no obedecen a una clasificación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa. También se conoce como variable nominal.
- *Conjunto ordenado*. Un campo/variable cuyos valores representan categorías con alguna clasificación intrínseca. Por ejemplo, los niveles de satisfacción con un servicio, que abarquen desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de conjuntos ordenados se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias. También se conoce como variable ordinal.
- *Marca*. Un campo/variable con dos valores distintos, como Sí y No o 1 y 2. También se conoce como una variable dicotómica o binaria.

Continua. Datos medidos en un intervalo o escala de cociente, donde los datos indican el orden de valores y la distancia entre valores. Por ejemplo, un salario de 72.195 \$ es superior que un salario de 52.398 \$ y la distancia entre ambos valores es de 19.797 \$. También se denominan datos cuantitativos, de escala o de rango numérico.

Los campos categóricos definen categorías en la visualización, normalmente para dibujar diferentes elementos gráficos o para agrupar elementos gráficos. Los campos continuos se suelen resumir en categorías de campos categóricos. Por ejemplo, una visualización predeterminada de ingresos de categorías de género mostrará los ingresos medios de hombres y los ingresos medios de mujeres. Los valores brutos de los campos continuos también se pueden representar, como en un diagrama de dispersión. Por ejemplo, un diagrama de dispersión puede mostrar el salario actual y el salario inicial de cada caso. Un campo categórico se puede usar para agrupar los casos por género.

Tipos de datos

El nivel de medición no es la única propiedad de un campo que determina su tipo. Un campo también se guarda como un tipo de datos específico. Los tipos de datos posibles son cadenas (datos no numéricos como letras), valores numéricos (números reales) y fechas. Al contrario que el nivel de medición, un tipo de datos de campo no se puede modificar temporalmente. Debe cambiar la forma en la que se guardan los datos en el conjunto de datos originales.

Conjuntos de resp. múltiples

Algunos archivos de datos admiten un tipo especial de "campo" llamado **conjunto de respuestas múltiples**. Los conjuntos de respuestas múltiples no son "campos" realmente en el sentido normal. Los conjuntos de respuestas múltiples utilizan múltiples campos para registrar respuestas a preguntas donde el encuestado puede ofrecer más de una respuesta. Los conjuntos de respuestas múltiples se tratan como campos categóricos y la mayoría de las acciones que puede realizar con los campos categóricos, también las puede realizar con los conjuntos de respuestas múltiples.

Los conjuntos de respuestas múltiples pueden ser conjuntos de dicotomía múltiple o conjuntos de categoría múltiple.

conjuntos de dicotomías múltiples. Un conjunto de dicotomía múltiple consta de varios campos de dicotomía: campos con sólo dos valores posibles de naturaleza sí/no, presente/ausente, marcado/no marcado. Aunque los campos no puedan ser estrictamente dicotómicos, todos los campos del conjunto están codificados de la misma forma.

Por ejemplo, una encuesta proporciona cinco respuestas posibles para la pregunta, "¿En cuál de los siguientes orígenes confía para las noticias?" El encuestado puede indicar varias opciones seleccionando una casilla de verificación junto a cada opción. Las cinco respuestas serán cinco campos en los archivos de datos, con el código 0 para *No* (no marcado) y 1 para *Sí* (marcado).

Conjuntos de categoría múltiple. Un conjunto de categoría múltiple está formado por múltiples campos, todos ellos codificados de la misma forma, con frecuencia con varias categorías de respuesta posibles. Por ejemplo, una pregunta de la encuesta es, "Nombre hasta tres nacionalidades que describan su herencia étnica". Puede haber hasta trescientas respuestas posibles, pero para la codificación, la lista se limita a las 40 nacionalidades más comunes, con el resto de categorías definidas como "otras". En un archivo de datos, las tres opciones son tres campos, cada uno con 41 categorías (40 nacionalidades codificadas y una categoría "otras").

Tablero Pestaña Detallado

Utilice la pestaña Detallado cuando sepa el tipo de visualización que desea crear o cuando quiera añadir aspectos estéticos, paneles y/o animaciones opcionales a una visualización. Por ejemplo, consulte "Ejemplos de Tableros" en la página 206.

1. Si había seleccionado un tipo de visualización en la pestaña Básico, se mostrará aquí. Si no es así, seleccione uno de la lista desplegable. Para obtener información sobre los tipos de visualización, consulte "Tipos de visualización de Graphboard preincorporado disponibles" en la página 198.
2. Justo a la derecha de la imagen en miniatura de la visualización encontrará controles para especificar los campos (variables) necesarios para el tipo de visualización. Debe especificar todos estos campos.
3. En el caso de ciertas visualizaciones, puede seleccionar un estadístico de resumen. En algunos casos (como los gráficos de barras), puede utilizar una de estas opciones de resumen para el aspecto estético de la transparencia. Para obtener descripciones de los estadísticos de resumen, consulte "Tablero Pestaña Básico" en la página 192.
4. Puede seleccionar uno o más de los aspectos estéticos opcionales. Estos aspectos pueden aumentar las dimensiones permitiéndole incluir otros campos en la visualización. Por ejemplo, puede utilizar un campo para modificar el tamaño de los puntos de un diagrama de dispersión. Para obtener más información acerca de los aspectos estéticos opcionales, consulte "Aspectos estéticos, superposiciones, paneles y animación" en la página 188. Recuerde que el aspecto estético de la transparencia no se admite en los scripts.
5. Si está creando una visualización de mapa, el grupo **Archivos de mapa** muestra el archivo o archivos de mapas que se usarán. Si hay un archivo de mapa predeterminado se muestra este archivo. Para cambiar el archivo de mapa, pulse **Seleccionar archivo de mapa** para que se muestre el cuadro de diálogo Seleccionar mapas. También puede especificar el archivo de mapa predeterminado en este cuadro de diálogo. Consulte el tema "Seleccionar archivos de mapas para visualizaciones de mapas" para obtener más información.
6. Puede seleccionar una o más opciones de paneles o animación. Para obtener más información sobre las opciones de paneles y animación, consulte "Aspectos estéticos, superposiciones, paneles y animación" en la página 188.

Seleccionar archivos de mapas para visualizaciones de mapas

Si selecciona una plantilla de visualización de mapa, necesita un archivo de mapa que defina la información geográfica para dibujar el mapa. Si hay un archivo de mapa predeterminado se usará para la visualización de mapas. Para elegir otro archivo de mapa, pulse **Seleccionar archivo de mapa** en la pestaña Detallado para que se muestre el cuadro de diálogo Seleccionar mapas.

El cuadro de diálogo Seleccionar mapas le permite elegir un archivo de mapa primario y uno de referencia. Los archivos de mapas definen la información geográfica para dibujar el mapa. Su aplicación se instala con un conjunto de archivos de mapas estándar. Si tiene otros archivos ESRI shapefile y quiere usarlos, primero tiene que convertirlos a archivos SMZ. Consulte el tema "Conversión y distribución de mapas en formato shapefile" en la página 219 para obtener más información. Tras convertir el mapa, pulse **Administrar...** en el cuadro de diálogo Selector de plantillas para importar el mapa en el Sistema de administración de modo que esté disponible en el cuadro de diálogo Seleccionar mapas.

A continuación se muestran algunos puntos que hay que considerar al especificar archivos de mapas:

- Todas las plantillas de mapas necesitan al menos un archivo de mapas.
- Normalmente el archivo de mapa enlaza un atributo de la leyenda con la clave de datos.

- Si la plantilla no necesita una leyenda que enlace con una clave de datos, necesita un archivo de mapa de referencia y campos que especifiquen las coordenadas (tales como longitud y latitud) para dibujar los elementos en el mapa de referencia.
- Las plantillas de mapas superpuestas necesitan dos mapas: un archivo de mapa primario y uno de referencia. El mapa de referencia se dibuja primero para que quede detrás del mapa primario.

Para obtener más información sobre términos de mapas, tales como atributos y características, consulte “Conceptos claves de los mapas” en la página 220.

Archivo de mapa. Puede seleccionar cualquier archivo de mapa que se encuentre en el Sistema de administración. Aquí se incluyen los archivos de mapa instalados previamente y los importados por usted. Para obtener más información acerca de la gestión de archivos de mapas, consulte “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219.

Leyenda. Especifique el atributo que quiere usar como clave para enlazar el archivo de mapa a la clave de datos.

Guarde como predeterminados este archivo y configuración de mapa. Seleccione esta casilla de verificación si quiere usar el archivo de mapa seleccionado como predeterminado. Si especifica un archivo de mapa como predeterminado no tendrá que especificar un archivo de mapa cada vez que cree una visualización de mapa.

Clave de datos. Este control muestra el mismo valor tal y como aparece en Selector de plantilla de la pestaña Detallado. Se incluye aquí por si necesita cambiar la clave si ha seleccionado un archivo de mapa específico.

Mostrar todas las características del mapa en la visualización. Cuando esta opción está activada, todas las características del mapa están representadas en la visualización, incluso si no coincide ningún valor de clave de datos. Si quiere ver tan sólo las características de las que hay datos, borre esta opción. Las características identificadas en las leyendas de la lista **Claves de mapa no coincidentes** no se mostrarán en la visualización.

Comparar valores de datos y mapa. La leyenda y la clave de datos se enlazan respectivamente para crear la visualización del mapa. La leyenda y la clave de datos deben extraerse del mismo dominio (por ejemplo, países y regiones). Pulse **Comparar** para comprobar que coinciden los valores de la leyenda y la clave de datos. El icono que se muestra le informa del estado de la comparación. Estos iconos se describen a continuación. Si se ha hecho una comparación y hay valores de la clave de datos que no coinciden con los de la leyenda, los valores de la clave de datos aparecen en la lista **Claves de datos no coincidentes**. En la lista **Claves de mapa no coincidentes**, también puede ver los valores de leyenda que no tienen valores de la clave de datos coincidentes. Si **Mostrar todas las características del mapa en la visualización** no está activada, las características identificadas por estos valores de leyenda no se mostrarán.

Tabla 49. Iconos de comparación.





Icono	Descripción
	No se ha realizado ninguna comparación. Éste es el estado predeterminado antes de pulsar en Comparar . Debe tener cuidado ya que no sabe si los valores de la clave de datos y la leyenda coinciden.
	Se ha realizado una comparación y los valores de la clave de datos y la leyenda coinciden por completo. Para cada valor de la clave de datos hay una característica coincidente identificada por la leyenda.

Tabla 49. Iconos de comparación (continuación).

Icono	Descripción
	Se ha realizado una comparación y algunos valores de la clave de datos y de la leyenda no coinciden. Para algunos valores de la clave de datos, no hay ninguna característica coincidente identificada por la leyenda. Debe tener cuidado. Si continúa, la visualización del mapa no incluirá todos los valores.
	Se ha realizado una comparación y no hay valores de la clave de datos y de la leyenda que coincidan. Debe elegir otra clave de datos u otra leyenda, ya que si continúa no se representará ningún mapa.

Tipos de visualización de Graphboard preincorporado disponibles

Puede crear diferentes tipos de vistas. Los siguientes tipos preincorporados están disponibles en las pestañas Básico y Detallado. Algunas de las descripciones para las plantillas (especialmente las plantillas de mapas) identifican los campos (variables) especificados en la pestaña Detallado utilizando **texto especial**.

Tabla 50. Tipos de gráficos disponibles.

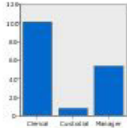
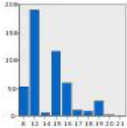
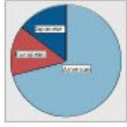
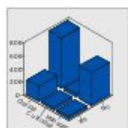
Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Barras</p> <p>Calcula una estadística de resumen de un campo numérico continuo y muestra los resultados en forma de barras de cada categoría de un campo categórico.</p> <p><i>Requiere:</i> un campo categórico y un campo continuo.</p>		<p>Barras de recuentos</p> <p>Muestra la proporción de filas/casos en cada categoría de un campo categórico como barras. También puede utilizar el nodo Gráfico de distribución para crear este gráfico. Dicho nodo ofrece opciones adicionales. Consulte el tema “Nodo Distribución” en la página 239 para obtener más información.</p> <p><i>Requiere:</i> un único campo categórico.</p>
	<p>Sectores</p> <p>Calcula la suma de un campo numérico continuo y muestra la proporción de esa suma distribuida en cada categoría de un campo categórico como porciones de un gráfico circular.</p> <p><i>Requiere:</i> un campo categórico y un campo continuo.</p>		<p>Sectores de recuentos</p> <p>Muestra la proporción de filas o casos en cada categoría de un campo categórico como porciones de un gráfico circular.</p> <p><i>Requiere:</i> un único campo categórico.</p>

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico



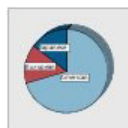
Descripción

Barras en 3D

Calcula un estadístico de resumen de un campo numérico continuo y muestra los resultados del cruce de categorías entre dos campos categóricos.

Requiere: un par de campos categóricos y un campo continuo.

Icono de gráfico

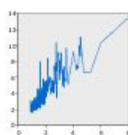


Descripción

Sectores en 3D

Es igual que el gráfico Sectores excepto por el efecto en 3D adicional.

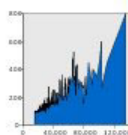
Requiere: un campo categórico y un campo continuo.



Línea

Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja una línea que conecta dichos valores. También puede utilizar el nodo Gráfico para crear un gráfico de línea. Dicho nodo ofrece opciones adicionales. Consulte el tema "Nodo Gráfico" en la página 227 para obtener más información. Las

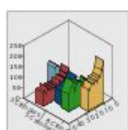
Requiere: un par de campos de cualquier tipo.



Área

Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja un área que conecta dichos valores. La diferencia entre un gráfico de línea y uno de áreas es mínima, ya que el de áreas presenta una línea con el espacio que aparece debajo coloreado. Sin embargo, si utiliza un aspecto estético de color, el resultado será una división simple de la línea y un apilado del área.

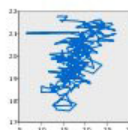
Requiere: un par de campos de cualquier tipo.



Áreas en 3D

Muestra los valores de un campo representado en contraposición a los valores de otro y divididos según un campo categórico. Se obtiene un elemento de área para cada categoría.

Requiere: un campo categórico y un par de campos de cualquier tipo.



Ruta

Muestra los valores de un campo en contraposición a los valores de otro, con una línea que conecta dichos valores en el orden en el que aparecen en el conjunto de datos original. El orden es la principal diferencia entre una ruta y una línea.

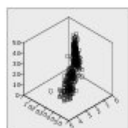
Requiere: un par de campos de cualquier tipo.

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Barras acumuladas</p> <p>Calcula un estadístico de resumen de un campo para cada valor de otro campo y dibuja una banda que conecta dichos valores. Una barra acumulada es básicamente una línea con efectos en 3D. Realmente no es un gráfico en 3D.</p> <p><i>Requiere:</i> un par de campos de cualquier tipo.</p>		<p>Superficie</p> <p>Muestra los valores de tres campos representados en contraposición a los valores de otro, con una superficie que conecta dichos valores.</p> <p><i>Requiere:</i> tres campos de cualquier tipo.</p>
	<p>Gráfico de dispersión</p> <p>Muestra los valores de un campo representado en contraposición a los valores de otro. Este gráfico puede destacar la relación entre los campos (si existe). También puede utilizar el nodo Gráfico para crear un diagrama de dispersión. Dicho nodo ofrece opciones adicionales. Consulte el tema "Nodo Gráfico" en la página 227 para obtener más información. Las</p> <p><i>Requiere:</i> un par de campos de cualquier tipo.</p>		<p>Gráfico de burbujas</p> <p>Igual que el diagrama de dispersión básico, muestra los valores de un campo representado en contraposición a los valores de otro. La diferencia es que los valores de un tercer campo se utilizan para modificar el tamaño de los puntos individuales.</p> <p><i>Requiere:</i> tres campos de cualquier tipo.</p>
	<p>Diagrama de dispersión en intervalos</p> <p>Igual que el diagrama de dispersión básico, muestra los valores de un campo representado en contraposición a los valores de otro. La diferencia es que los valores similares se agrupan en intervalos y que se utiliza el color o tamaño para indicar el número de casos de cada intervalo.</p> <p><i>Requiere:</i> un par de campos continuos.</p>		<p>Diagrama de dispersión en intervalos hexagonales</p> <p>Consulte la descripción de Diagrama de dispersión en intervalos. La diferencia es la forma de los intervalos subyacentes, que son hexágonos en lugar de círculos. El aspecto del diagrama de dispersión en intervalos hexagonales es parecido al del diagrama de dispersión en intervalos. Sin embargo, el número de valores de cada intervalo variará dependiendo del gráfico debido a la forma de los intervalos subyacentes.</p> <p><i>Requiere:</i> un par de campos continuos.</p>

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico



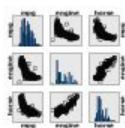
Descripción

Diagrama de dispersión en 3D

Muestra los valores de tres campos representados en contraposición a los valores de cada uno de ellos. Este gráfico puede destacar la relación entre los campos (si existe). También puede utilizar el nodo Gráfico para crear un diagrama de dispersión en 3D. Dicho nodo ofrece opciones adicionales. Consulte el tema “Nodo Gráfico” en la página 227 para obtener más información. Las

Requiere: tres campos de cualquier tipo.

Icono de gráfico

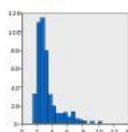


Descripción

Matriz de diagramas de dispersión (SPLOM)

Muestra los valores de un campo representado en contraposición a los valores de otro para cada campo. Una SPLOM es igual que una tabla de diagramas de dispersión. La SPLOM también incluye un histograma de cada campo.

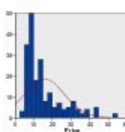
Requiere: dos o más campos continuos.



Histograma

Muestra la distribución de la frecuencia de un campo. Un histograma puede ayudarle a determinar el tipo de distribución y comprobar si ésta es asimétrica. También puede utilizar el nodo de gráfico Histograma para crear este gráfico. Dicho nodo ofrece opciones adicionales. Consulte el tema “Pestaña Gráfico de Histograma” en la página 243 para obtener más información.

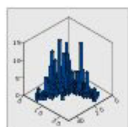
Requiere: un único campo de cualquier tipo.



Histograma con distribución normal

Muestra la distribución de la frecuencia de un campo continuo con una curva superpuesta de la distribución normal.

Requiere: un único campo continuo.



Histograma en 3D

Muestra la distribución de la frecuencia de un par de campos continuos.

Requiere: un par de campos continuos.



Densidad en 3D

Muestra la distribución de la frecuencia de un par de campos continuos. Es similar a un histograma en 3D, con la única diferencia de que se utiliza una superficie en lugar de barras para mostrar la distribución.

Requiere: un par de campos continuos.

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Gráfico de puntos</p> <p>Muestra los casos o filas individuales y los apila en los puntos de datos distintos del eje x. Este gráfico es parecido a un histograma en cuanto a que muestra la distribución de los datos; sin embargo, éste presenta cada caso o fila en lugar de un recuento total de un intervalo específico (rango de valores).</p> <p><i>Requiere:</i> un único campo de cualquier tipo.</p>		<p>Gráfico de puntos en 2D</p> <p>Muestra los casos o filas individuales y los apila en los distintos puntos de datos del eje y para cada categoría de un campo categórico.</p> <p><i>Requiere:</i> un campo categórico y un campo continuo.</p>
	<p>Diagrama de caja</p> <p>Calcula los cinco estadísticos (mínimo, primer cuartil, mediana, tercer cuartil y máximo) de un campo continuo para cada categoría de un campo categórico. Los resultados se muestran como elementos de un gráfico de caja o esquema. Los gráficos de caja pueden ayudarle a ver cómo varía la distribución de los datos continuos entre las categorías.</p> <p><i>Requiere:</i> un campo categórico y un campo continuo.</p>		<p>Mapa de calor</p> <p>Calcula la media de un campo continuo del cruce de categorías entre dos campos categóricos.</p> <p><i>Requiere:</i> un par de campos categóricos y un campo continuo.</p>
	<p>Paralelo</p> <p>Crea ejes paralelos de cada campo y dibuja una línea a través del valor de campo para cada fila o caso de los datos.</p> <p><i>Requiere:</i> dos o más campos continuos.</p>		<p>Coropleta de recuentos</p> <p>Calcula el recuento para cada categoría de un campo categórico (Clave de datos) y dibuja un mapa que utiliza la saturación de color para representar los recuentos de las funciones de mapa que corresponden a las categorías.</p> <p><i>Requiere:</i> un campo categórico. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos.</p>



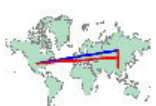
Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Coropleta de medias/medianas/sumas</p> <p>Calcula la media, la mediana o la suma de un campo continuo (Color) para cada categoría de un campo categórico (Clave de datos) y dibuja un mapa que utiliza la saturación de color para representar las estadísticas calculadas en las funciones de mapa que corresponden a las categorías.</p> <p><i>Requiere:</i> un campo categórico y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>		<p>Coropleta de valores</p> <p>Dibuja un mapa que utiliza colores para representar valores de un campo categórico (Color) para las funciones de mapa que corresponden a los valores definidos por otro campo categórico (Clave de datos). Si existen varios valores categóricos del campo Color para cada característica, se utiliza el valor modal.</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>
	<p>Coordenadas de una coropleta de recuentos</p> <p>Es similar a la coropleta de recuentos, excepto en que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar puntos en el mapa de coropleta.</p> <p><i>Requiere:</i> un campo categórico y un par de campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>		<p>Coordenadas de una coropleta de medias/medianas/sumas</p> <p>Es similar a la Coropleta de medias/medianas/sumas excepto en que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar puntos en el mapa de coropleta.</p> <p><i>Requiere:</i> un campo categórico y tres campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>
	<p>Coordenadas en un coroplético de valores</p> <p>Es similar a la coropleta de valores, excepto en que hay dos campos continuos adicionales (Longitud y Latitud) que identifican las coordenadas para dibujar puntos en el mapa de coropleta.</p> <p><i>Requiere:</i> un par de campos categóricos y un par de campos continuos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>		<p>Barras de recuentos sobre un mapa</p> <p>Calcula la proporción de filas o casos en cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos) y dibuja un mapa y los gráficos de barras en el centro de cada mapa.</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Barras sobre un mapa</p> <p>Calcula una estadística de resumen para un campo continuo (Valores) y muestra los resultados de cada categoría de un campo categórico (Categorías) para cada función de mapa (Clave de datos) como diagramas de barras situado en el centro de cada función de mapa.</p> <p><i>Requiere:</i> un par de campos categóricos y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>		<p>Gráfico circular de recuentos de un mapa</p> <p>Muestra la proporción de filas o casos en cada categoría de un campo categórico (Categorías) para cada característica del mapa (Clave de datos) y dibuja un mapa y las proporciones en forma de porciones de un gráfico circular en el centro de cada función de mapa.</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>
	<p>Gráfico circular sobre un mapa</p> <p>Calcula la suma de un campo continuo (Valores) en cada categoría de un campo categórico (Categorías) para cada función de mapa (Clave de datos) y dibuja un mapa y las sumas como porciones de un diagrama circular en el centro de cada función de mapa.</p> <p><i>Requiere:</i> un par de campos categóricos y un campo continuo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>		<p>Gráfico de líneas sobre un mapa</p> <p>Calcula un estadístico de resumen de un campo continuo (Y) para cada valor de otro campo (X) para cada característica del mapa (Clave de datos) y dibuja un mapa y los gráficos de líneas que conectan los valores en el centro de cada mapa.</p> <p><i>Requiere:</i> un campo categórico y un par de campos de cualquier tipo. Un archivo de mapas cuya clave coincide con las categorías de la clave de datos).</p>
	<p>Coordenadas sobre un mapa de referencia</p> <p>Dibuja un mapa y puntos utilizando campos continuos (Longitud y Latitud) que identifican las coordenadas de los puntos.</p> <p><i>Requiere:</i> un par de campos de rango. Un archivo de mapas.</p>		<p>Flechas en un mapa de referencia</p> <p>Dibuja un mapa y flechas utilizando campos continuos que identifican los puntos iniciales (Longitud inicial y Latitud inicial) y los puntos finales (Longitud final y Latitud final) para cada flecha. Cada registro o caso en los resultados de los datos en una flecha del mapa.</p> <p><i>Requiere:</i> cuatro campos continuos. Un archivo de mapas.</p>

Tabla 50. Tipos de gráficos disponibles (continuación).

Icono de gráfico	Descripción	Icono de gráfico	Descripción
	<p>Mapa de superposición de puntos</p> <p>Dibuja un mapa de referencia y superpone sobre él otro mapa de puntos con puntos coloreadas por un campo categórico (Color).</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapa de puntos cuya clave coincide con las categorías Clave de datos. Un archivo de mapas de referencia.</p>		<p>Mapa de superposición de polígonos</p> <p>Dibuja un mapa de referencia y superpone sobre él otro mapa de polígonos con polígonos coloreados por un campo categórico (Color).</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapas de polígono cuya clave coincide con las categorías Clave de datos. Un archivo de mapas de referencia.</p>
	<p>Mapa de superposición de líneas</p> <p>Dibuja un mapa de referencia y superpone sobre él otro mapa de líneas con líneas coloreadas por un campo categórico (Color).</p> <p><i>Requiere:</i> un par de campos categóricos. Un archivo de mapas de líneas cuya clave coincide con las categorías de la clave de datos). Un archivo de mapas de referencia.</p>		

Creación de visualizaciones de mapas

Para muchas visualizaciones, debe realizar sólo dos elecciones: campos (variables) de interés y una plantilla para visualizar esos campos. No se requieren elecciones o acciones adicionales. Las visualizaciones de mapas requieren como mínimo un paso adicional: seleccionar un archivo de mapas que defina la información geográfica para la visualización de mapas.

Los pasos básicos para la creación de un mapa simple son los siguientes:

1. Seleccione los campos de interés en la pestaña Básico. Para obtener información acerca del tipo y número de campos necesarios para varias visualizaciones de mapas, consulte "Tipos de visualización de Graphboard preincorporado disponibles" en la página 198.
2. Seleccione una plantilla del mapa.
3. Pulse en la pestaña Detallado.
4. Compruebe que tanto la **Clave de datos** como otras listas desplegables obligatorias se ha establecido en los campos correctos.
5. En el grupo Archivos de mapas, pulse **Seleccionar un archivo de mapas**.
6. Seleccione el cuadro de diálogo Seleccionar mapas para seleccionar la leyenda así como el archivo de mapas. Los valores de la leyenda deben coincidir con los valores del campo que especifican **las claves de datos**. Puede utilizar el botón **Comparar** para comparar esos valores. Si selecciona una superposición de plantillas de mapas, deberá seleccionar un mapa de referencia. El mapa de referencia

no tiene clave de los datos. Se utiliza como fondo para el mapa principal. Si desea obtener más información acerca del cuadro de diálogo Seleccionar mapas, consulte "Seleccionar archivos de mapas para visualizaciones de mapas" en la página 196.

7. Pulse **Aceptar** para cerrar el cuadro de diálogo Seleccionar mapas.
8. En el Selector de plantillas de tablero, pulse **Ejecutar** para crear la visualización del mapa.

Ejemplos de Tableros

Esta sección incluye varios ejemplos diferentes para mostrar las opciones disponibles. Los ejemplos también ofrecen información para interpretar las visualizaciones resultantes.

Estos ejemplos usan la ruta denominada *graphboard.str*, que hace referencia a los archivos de datos, denominados *employee_data.sav*, *customer_subset.sav* y *worldsales.sav*. Estos archivos están disponibles en la carpeta *Demos* de cualquier IBM SPSS Modeler instalación Client. Puede acceder desde el grupo de programas IBM SPSS Modeler en el menú Inicio de Windows. El archivo *graphboard.str* se encuentra en la carpeta *streams*.

Se recomienda que lea los ejemplos en el orden presentado. Los ejemplos siguientes se basan en los anteriores.

Ejemplo: Gráfico de barras con un estadístico de resumen

Crearemos un gráfico de barras que resume un campo o variable numérico continuo para cada categoría de una variable de conjunto o categórica. De forma específica, crearemos un gráfico de barras que muestre el salario medio de hombres y mujeres.

Éste y varios de los ejemplos siguientes utilizan *Employee data*, que es un conjunto de datos hipotético con información sobre los trabajadores de una empresa.

1. Añada un nodo de origen Archivo de Statistics que haga referencia a *employee_data.sav*.
2. Añada un nodo Tablero y ábralo para editarlo.
3. En la pestaña Básico, seleccione *Género* y *Salario actual*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
4. Seleccione **Barras**.
5. En la lista desplegable Resumen, seleccione **Media**.
6. Pulse en **Ejecutar**.
7. En la pantalla resultante, pulse el botón de la barra de herramientas "Mostrar campo y etiquetas de valores" (el segundo del grupo de dos en el centro de la barra de herramientas).

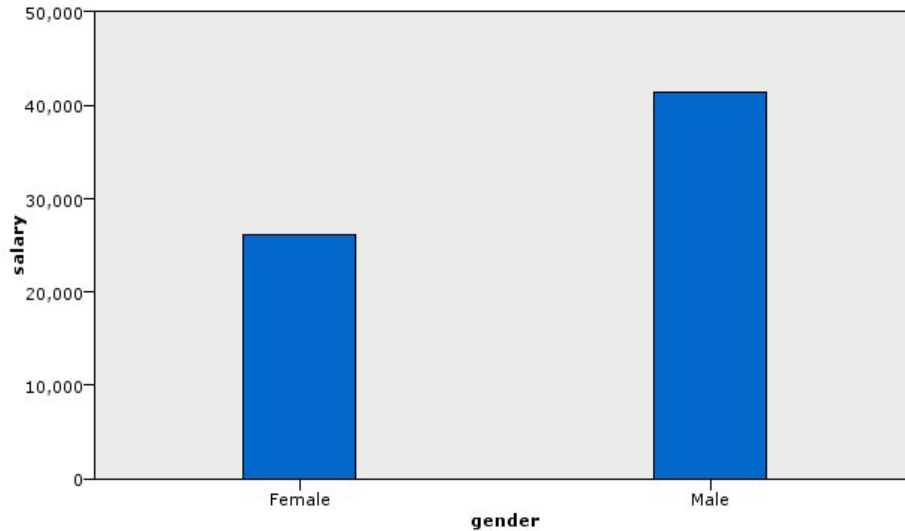


Figura 10. Gráfico de barras con un estadístico de resumen

Podemos observar lo siguiente:

- Basándonos en la altura de las barras, está claro que el salario medio de los hombres es mayor que el salario medio de las mujeres.

Ejemplo: Gráfico de barras apiladas con un estadístico de resumen

Ahora crearemos un gráfico de barras apiladas para ver si la diferencia en el salario medio entre hombres y mujeres está relacionada con el tipo de trabajo. Tal vez las mujeres, en promedio, se dedican a ciertos trabajos más que los hombres.

Nota: Este ejemplo usa *Employee data*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Categoría laboral* y *Salario actual*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **Barras**.
4. En la lista Resumen, seleccione **Media**.
5. Pulse en la pestaña Detallado. Recuerde que sus selecciones en la pestaña anterior se reflejarán aquí.
6. En el grupo Aspectos estéticos opcionales, seleccione *sexo* en la lista desplegable Color.
7. Pulse en **Ejecutar**.

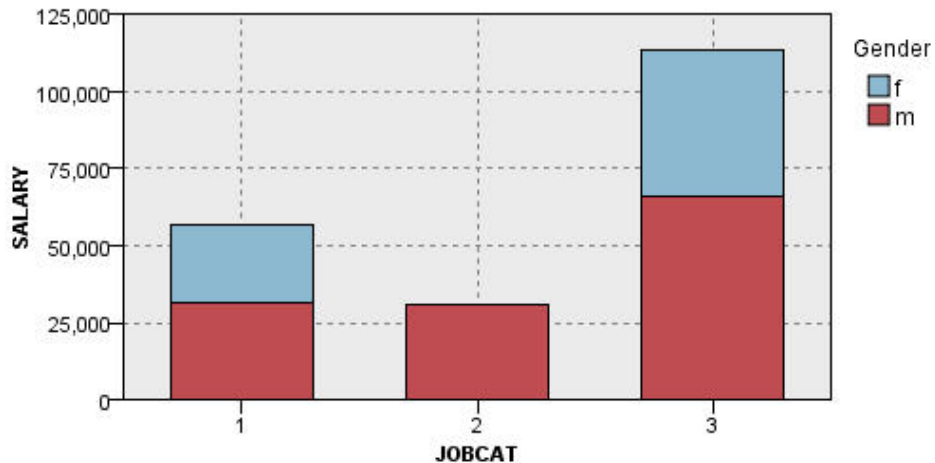


Figura 11. Gráfico de barras apiladas

Podemos observar lo siguiente:

- La diferencia en los salarios medios de cada tipo de trabajo no parece ser tan grande como lo era en el gráfico de barras que comparaba los salarios medios de todos los hombres y mujeres. Quizás haya un número variable de hombres y mujeres en cada grupo. Puede comprobarlo creando un gráfico de barras de recuentos.
- Independientemente del tipo de trabajo, el salario medio de los hombres siempre es mayor que el de las mujeres.

Ejemplo: Histograma panelado

Crearemos un histograma panelado por sexos para poder comparar las distribuciones de frecuencia del salario de hombres y mujeres. La distribución de frecuencia muestra cuántos casos o filas se encuentran dentro de rangos de salario específicos. El histograma panelado puede ayudarnos a analizar aun más la diferencia de salarios entre los dos sexos.

Note: Este ejemplo usa *Employee data*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Salario actual*.
3. Seleccione **Histograma**.
4. Pulse en la pestaña Detallado.
5. En el grupo Paneles y animación, seleccione *sexo* en la lista desplegable A través del panel.
6. Pulse en **Ejecutar**.

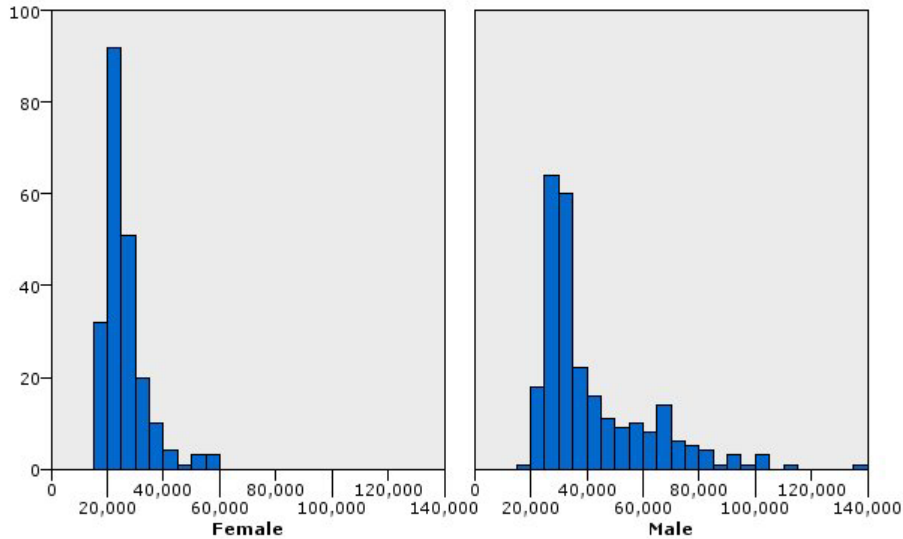


Figura 12. Histograma panelado

Podemos observar lo siguiente:

- Ninguna de las distribuciones de frecuencia es normal. Es decir, los histogramas no parecen curvas de campana, como debería ser si los datos estuvieran distribuidos con normalidad.
- Las barras más altas están en el lado izquierdo de cada gráfico. Por lo tanto, tanto en hombres como en mujeres, hay un mayor número de personas con salarios más bajos.
- Las distribuciones de frecuencia de salarios entre hombres y mujeres no son iguales. Observe la forma de los histogramas. Hay más hombres con salarios más altos que mujeres con salarios más altos.

Ejemplo: Gráfico de puntos panelado

Como un histograma, un gráfico de puntos muestra la distribución de un rango numérico continuo. A diferencia de un histograma, que muestra los recuentos de rangos de datos en intervalos, un gráfico de puntos muestra todos los casos o filas de los datos. Por lo tanto, un gráfico de puntos ofrece una mayor cantidad de detalles en comparación con el histograma. De hecho, el uso de un gráfico de puntos puede ser el punto de inicio preferido cuando se analizan distribuciones de frecuencia.

Note: Este ejemplo usa *Employee data*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Salario actual*.
3. Seleccione **Gráfico de puntos**.
4. Pulse en la pestaña Detallado.
5. En el grupo Paneles y animación, seleccione *sexo* en la lista desplegable A través del panel.
6. Pulse en **Ejecutar**.
7. Maximice la ventana de resultados para ver el gráfico con mayor claridad.

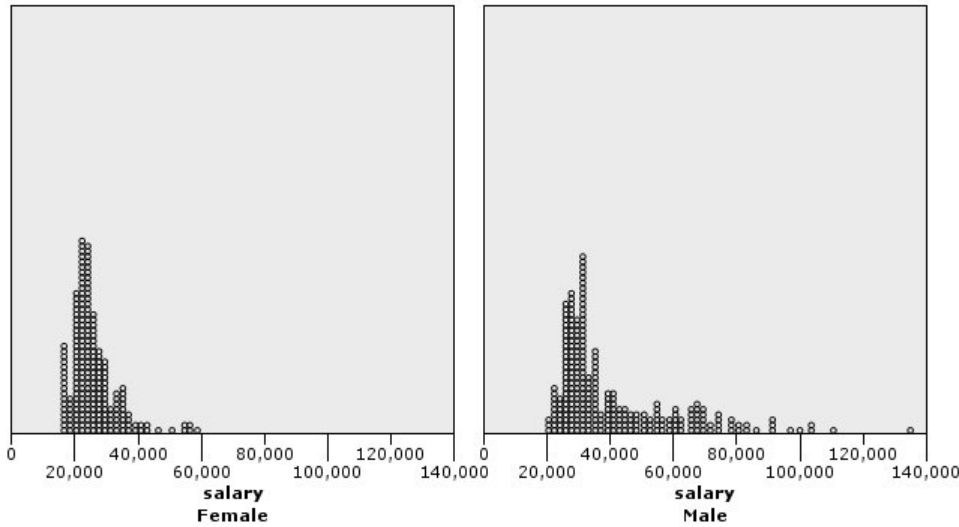


Figura 13. Gráfico de puntos panelado

En comparación con el histograma (consulte “Ejemplo: Histograma panelado” en la página 208), podemos observar lo siguiente:

- El pico de los 20.000 que apareció en el histograma de mujeres es menos drástico en el gráfico de puntos. Hay muchos casos o filas concentrados alrededor de dicho valor, pero la mayoría de los valores están más cerca de los 25.000. Este nivel de detalle no es aparente en el histograma.
- Aunque el histograma de los hombres sugiere que el salario medio de los hombres desciende gradualmente tras los 40.000, el gráfico de puntos muestra que la distribución es bastante uniforme tras este valor, hasta los 80.000. En cualquier valor de salario dentro de ese rango, hay tres o más hombres que obtienen dicho salario en particular.

Ejemplo: Gráfico de caja

Un diagrama de cajas es otra visualización que resultará útil para saber cómo se distribuyen los datos. Un diagrama de cajas contiene varias medidas estadísticas que estudiaremos tras crear la visualización.

Note: Este ejemplo usa *Employee data*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Género* y *Salario actual* . (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **Gráfico de caja**.
4. Pulse en **Ejecutar**.

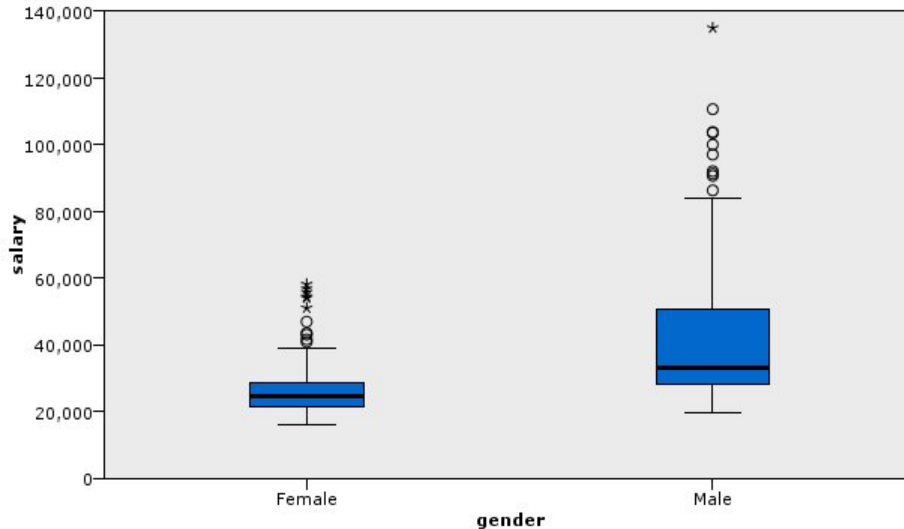


Figura 14. Diagrama de caja

Veamos las distintas partes del gráfico de caja:

- La línea oscura que se encuentra en la mitad de las cajas es la mediana del *salario*. La mitad de los casos o filas tiene un valor superior a la mediana y la otra mitad tiene un valor inferior. Igual que la media, la mediana es una medida de tendencia central. A diferencia de la media, se ve menos influida por los casos o filas con valores extremos. En este ejemplo, la mediana es inferior a la media (compárelo con “Ejemplo: Gráfico de barras con un estadístico de resumen” en la página 206). La diferencia entre la media y la mediana indica que hay menos casos o filas con valores extremos que elevan la media. Es decir, hay un par de trabajadores con grandes salarios.
- La parte inferior de la caja indica el percentil 25. El veinticinco por ciento de los casos o filas tienen valores por debajo del percentil 25. La parte superior de la caja representa el percentil 75. El veinticinco por ciento de los casos o filas tienen valores por encima del percentil 75. Esto significa que el 50 % de los casos o filas se encuentran dentro de la caja. La caja es mucho más pequeña para mujeres que para hombres. Es una pista que indica que el *salario* varía menos entre las mujeres que entre los hombres. Las partes superior e inferior de la caja suelen denominarse **bisagras**.
- Las barras en forma de T que salen de las cajas se denominan **cercas internas** o **patillas** o **bigotes**. Tienen una extensión de 1,5 veces la altura de la caja o, si no hay ningún caso o fila con valor en dicho rango, hasta los valores mínimo y máximo. Si los datos se distribuyen con normalidad, se espera que aproximadamente el 95 % de los datos se encuentre entre las cercas internas. En este ejemplo, las cercas internas tienen una extensión menor para las mujeres que para los hombres, otro indicativo de que el *salario* varía menos entre las mujeres que entre los hombres.
- Los puntos son **valores atípicos**. Se definen como valores que no están dentro de las cercas internas. Los valores atípicos son valores extremos. Los asteriscos o estrellas son **valores atípicos extremos**. Representan casos o filas con valores superiores a la altura de las cajas multiplicada por tres. Existen varios valores atípicos tanto para mujeres como para hombres. Recuerde que la media es superior que la mediana. La media de mayor tamaño es el resultado de estos valores atípicos.

Ejemplo: Gráfico circular

Ahora utilizaremos un conjunto de datos diferente para explorar otros tipos de visualización. El conjunto de datos es *customer_subset*, que es un archivo de datos hipotéticos con información sobre clientes.

Primero crearemos un gráfico circular para comprobar la proporción de clientes en diferentes regiones geográficas.

1. Añada un nodo de origen Archivo de Statistics que haga referencia a *customer_subset.sav*.
2. Añada un nodo Tablero y ábralo para editarlo.

3. En la pestaña Básico, seleccione *Indicador geográfico*.
4. Seleccione **Sectores de recuentos**.
5. Pulse en **Ejecutar**.

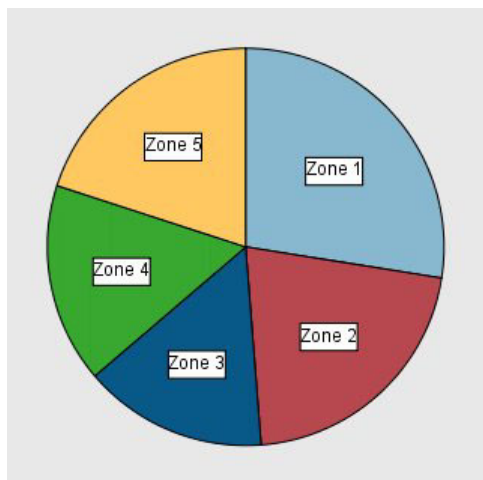


Figura 15. Gráfico circular

Podemos observar lo siguiente:

- La zona 1 tiene más clientes que el resto de zonas.
- Los clientes se distribuyen equitativamente entre el resto de zonas.

Ejemplo: Mapa de calor

Ahora crearemos un mapa de calor categórico para comprobar la media de ingresos de los clientes en diferentes regiones geográficas y grupos de edad.

Note: Este ejemplo utiliza *customer_subset*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Indicador geográfico*, *Categoría de edad* e *Ingresos del hogar en miles*, en ese orden. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **Mapa de calor**.
4. Pulse en **Ejecutar**.
5. En la ventana de resultados, pulse el botón de la barra de herramientas "Mostrar campo y etiquetas de valores" (el botón de la derecha de los dos en el centro de la barra de herramientas).

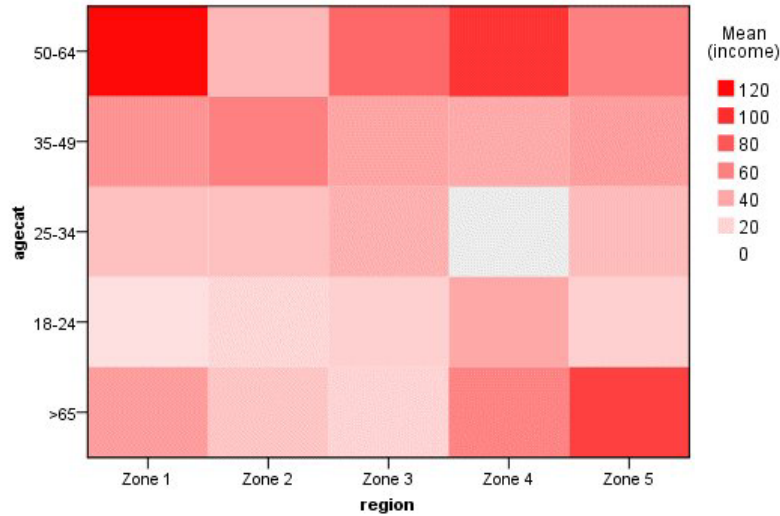


Figura 16. Mapa de calor categórico

Podemos observar lo siguiente:

- Un mapa de calor es como una tabla que utiliza colores en lugar de números para representar los valores de las casillas. El color rojo intenso y brillante indica el valor más alto, mientras que el gris indica un valor bajo. El valor de cada casilla es la media del campo o variable continuos de cada par de categorías.
- Excepto en las zonas 2 y 5, el grupo de clientes cuya edad está entre 50 y 64 tiene una media de ingresos por hogar superior a la de otros grupos.
- No hay clientes con una edad entre 25 y 34 en la zona 4.

Ejemplo: Matriz de diagramas de dispersión (SPLOM)

Crearemos una matriz de diagramas de dispersión de distintas variables para determinar si existen relaciones entre las variables del conjunto de datos.

Note: Este ejemplo utiliza *customer_subset*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Edad en años*, *Ingresos del hogar en miles* y *Deuda de tarjeta de crédito en miles*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **SPLOM**.
4. Pulse en **Ejecutar**.
5. Maximice la ventana de resultados para ver la matriz con mayor claridad.

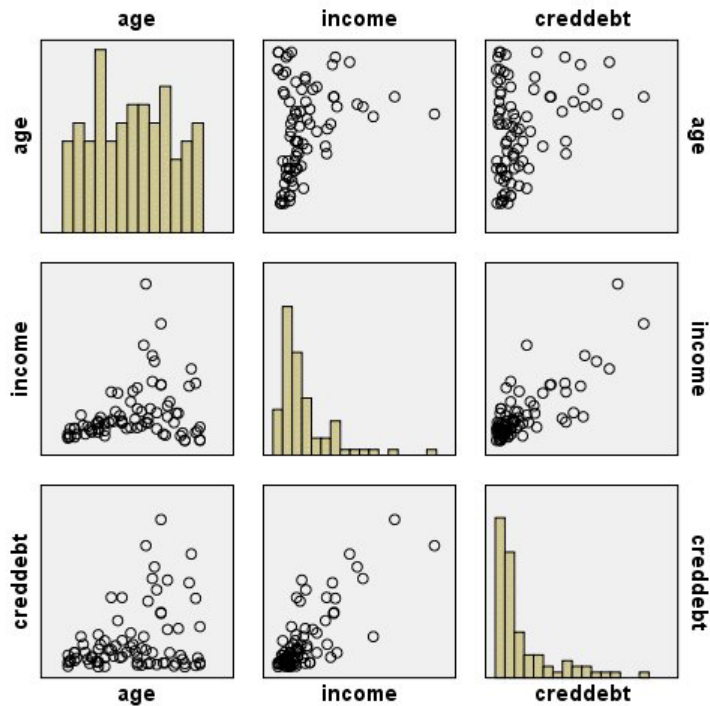


Figura 17. Matriz de diagramas de dispersión (SPLOM)

Podemos observar lo siguiente:

- Los histogramas que aparecen en la diagonal muestran la distribución de cada variable de la SPLOM. El histograma de *edad* aparece en la casilla superior izquierda, el de *ingresos* en la casilla central y el de *deudacred* en la casilla inferior derecha. Ninguna de las variables aparece distribuida con normalidad. Es decir, ninguno de los histogramas parece una curva de campana. Además, tenga en cuenta que los histogramas de *ingresos* y *deudacred* son positivamente asimétricos.
- No parece que exista ninguna relación entre *edad* y el resto de variables.
- Hay una relación lineal entre *ingresos* y *deudacred*. Es decir, el valor de *deudacred* aumenta a la par que el valor de *ingresos*. Puede que desee crear diagramas de dispersión individuales de estas variables y las otras variables relacionadas para ampliar las relaciones.

Ejemplo: Coropleta (mapa de color) de sumas

Ahora crearemos una visualización de mapas. Luego, en el siguiente ejemplo, crearemos una variación de esta visualización. El conjunto de datos *worldsales*, que es un archivo de datos hipotéticos que contiene ingresos por ventas clasificados por continentes y productos.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Continente* e *Ingreso*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **Coropleta de sumas**.
4. Pulse en la pestaña Detallado.
5. En el grupo Aspectos estéticos opcionales, seleccione *Continente* en la lista desplegable Etiquetas de datos.
6. En el grupo Archivos de mapas, pulse **Seleccionar un archivo de mapas**.
7. En el cuadro de diálogo Seleccionar mapas, compruebe que **Mapa** está configurada en *Continentes* y **Leyenda** en *CONTINENTE*.

8. En los grupos Comparar mapa y Valores de datos, haga clic en **Comparar** para asegurarse de que las leyendas coinciden con las claves de datos. En este ejemplo, todos los valores de las claves de datos han coincidido con las leyendas y las características. También podemos ver que ningún dato coincide con Oceanía.
9. En el cuadro de diálogo Seleccionar mapas, haga clic en **Aceptar**.
10. Pulse en **Ejecutar**.



Figura 18. Coropleta de sumas

Desde la visualización de mapas, podemos ver fácilmente que los ingresos son superiores en Norteamérica e inferiores en Sudamérica y África. Cada continente está etiquetado porque utilizamos *Continente* para la estética de las etiquetas de datos.

Ejemplo: Gráficas de barras en un mapa

Este ejemplo muestra cómo se dividen los ingresos por producto en cada continente.

Note: Este ejemplo utiliza *worldsales*.

1. Añada un nodo Tablero y ábralo para editarlo.
2. En la pestaña Básico, seleccione *Continente*, *Producto*, e *Ingreso*. (Mantenga pulsada la tecla Ctrl mientras pulsa para seleccionar varios campos o variables.)
3. Seleccione **Barras de un mapa**.

4. Pulse en la pestaña Detallado.

Si utiliza más de un campo de un tipo específico, es importante comprobar que cada campo está asignado al intervalo correcto.

5. En la lista desplegable Categorías, seleccione *Producto*.
6. En la lista desplegable Valores, seleccione *Ingreso*.
7. En la lista desplegable Clave de datos, seleccione *Continente*.

8. En la lista desplegable Resumen, seleccione *Suma*.
9. En el grupo Archivos de mapas, pulse **Seleccionar un archivo de mapas**.
10. En el cuadro de diálogo Seleccionar mapas, compruebe que **Mapa** está configurada en *Continentes* y **Leyenda** en *CONTINENTE*.
11. En los grupos Comparar mapa y Valores de datos, haga clic en **Comparar** para asegurarse de que las leyendas coinciden con las claves de datos. En este ejemplo, todos los valores de las claves de datos han coincidido con las leyendas y las características. También podemos ver que ningún dato coincide con Oceanía.
12. En el cuadro de diálogo Seleccionar mapas, haga clic en **Aceptar**.
13. Pulse en **Ejecutar**.
14. Maximice la ventana de resultados para ver la pantalla con mayor claridad.

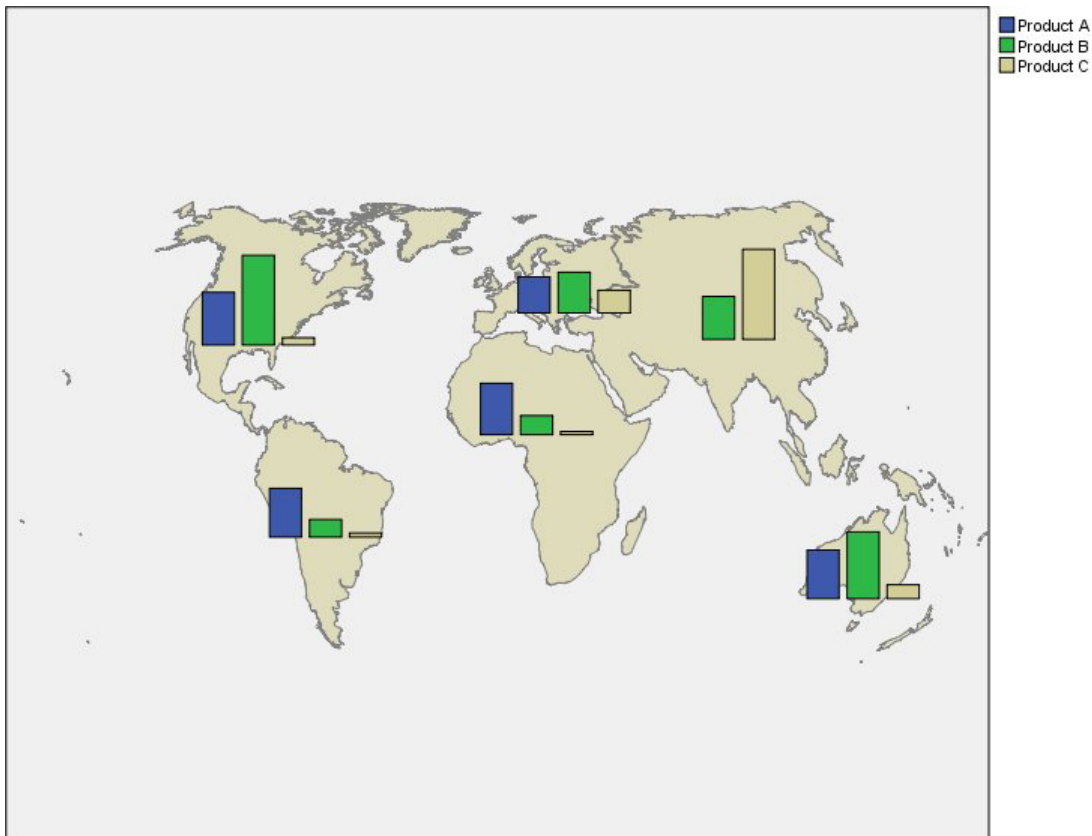


Figura 19. Gráficos de barras de un Mapa

Podemos observar lo siguiente:

- La distribución del total de ingresos en productos es muy similar en Sudamérica y África.
- *Producto C* genera los menores ingresos en todo el mundo excepto en Asia.
- No hay ingresos o ingresos mínimos del *Producto A* en Asia.

Pestaña Aspecto de Tablero

Puede especificar las opciones de aspecto antes de crear gráficos.

Opciones generales de aspecto

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Muestreo. Especifique un método para conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el número predeterminado de registros. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar la opción **Muestra**. Si lo prefiere, puede trazar todos los puntos de datos seleccionando **Utilizar todos los datos**, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

Opciones de aspecto de la hoja de estilo

Existen dos botones que permiten controlar las plantillas de visualización (y hojas de estilo y mapas) disponibles:

Administrar. Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Consulte el tema “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219 para obtener más información.

Ubicación. Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Consulte el tema “Configuración de la ubicación de las plantillas, hojas de estilo y mapas” en la página 218 para obtener más información.

El siguiente ejemplo muestra dónde se colocan las opciones de aspecto en el gráfico. (*Nota:* no todos los gráficos utilizan todas estas opciones.)

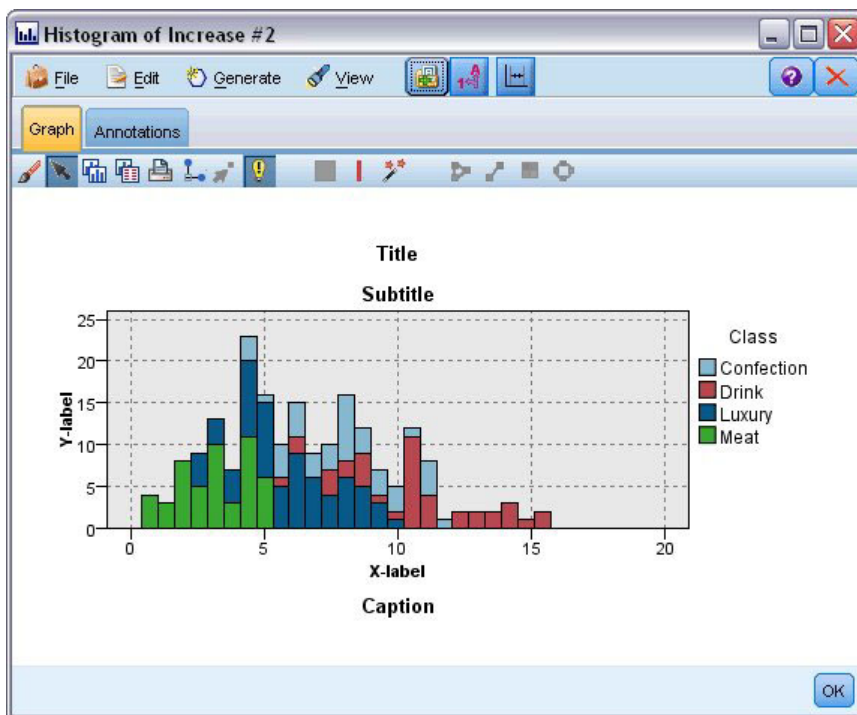


Figura 20. Posición de las diversas opciones de aspecto de gráficos

Configuración de la ubicación de las plantillas, hojas de estilo y mapas

Las plantillas y hojas de estilo de visualización y los archivos de mapas se guardan en una carpeta local específica o en el IBM SPSS Collaboration and Deployment Services Repository. Al seleccionar plantillas, hojas de estilo y mapas, sólo se muestran los incorporados en esta ubicación. Al mantener todas las plantillas, hojas de estilo y archivos de mapas en una ubicación, las aplicaciones de IBM SPSS pueden acceder a ellas fácilmente. Si desea obtener más información sobre cómo agregar más plantillas, hojas de estilo y archivos de mapas a esta ubicación, consulte “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219.

Cómo se configura la ubicación de las plantillas, hojas de estilo y archivos de mapas

1. En el cuadro de diálogo de plantillas o de hojas de estilo, pulse **Ubicación...** para abrir el cuadro de diálogo Plantillas, hojas de estilo y mapas.
2. Seleccione una opción para la ubicación predeterminada de las plantillas, hojas de estilo y mapas:
Máquina local. Las plantillas, hojas de estilo y archivos de mapas se guardan en una carpeta concreta o en su máquina local. En Windows XP, esta carpeta es *C:\Documents and Settings\\Application Data\SPSSInc\Graphboard*. La carpeta no se puede cambiar.
IBM SPSS Collaboration and Deployment Services Repository. Las plantillas, hojas de estilo y archivos de mapas se guardan en una carpeta especificada por el usuario del IBM SPSS Collaboration and Deployment Services Repository. Para identificar la carpeta específica, pulse **Carpeta**. Si desea obtener más información, consulte “Uso del IBM SPSS Collaboration and Deployment Services Repository como ubicación de las plantillas, hojas de estilo y mapas”.
3. Pulse en **Aceptar**.

Uso del IBM SPSS Collaboration and Deployment Services Repository como ubicación de las plantillas, hojas de estilo y mapas

Las plantillas y hojas de estilo de visualización se pueden almacenar en el IBM SPSS Collaboration and Deployment Services Repository. Esta ubicación es una carpeta específica en el IBM SPSS Collaboration and Deployment Services Repository. Si se define como la ubicación predeterminada, todas las plantillas, hojas de estilo y archivos de mapas de esta ubicación estarán disponibles para su selección.

Cómo se configura una carpeta en el IBM SPSS Collaboration and Deployment Services Repository como la ubicación de las plantillas, hojas de estilo y archivos de mapas

1. En un cuadro de diálogo con un botón Ubicación, pulse **Ubicación...**
2. Seleccione IBM SPSS Collaboration and Deployment Services Repository.
3. Pulse en **Carpeta**.
Note: Si aún no está conectado al IBM SPSS Collaboration and Deployment Services Repository, se le solicitará la información de conexión.
4. En el cuadro de diálogo Seleccionar carpeta, seleccione la carpeta en la que se guardarán las plantillas, hojas de estilo y archivos de mapas.
5. Opcionalmente, puede seleccionar una etiqueta en **Recuperar etiqueta**. Sólo se mostrarán las plantillas, hojas de estilo y archivos de mapas con esa etiqueta.
6. Si busca una carpeta que contiene una plantilla u hoja de estilo concreta, es posible que desee buscar la plantilla, hoja de estilo o archivo de mapa en la pestaña Buscar. El cuadro de diálogo Seleccionar carpeta selecciona automáticamente la carpeta en la que se encuentra la plantilla, hoja de estilo o archivo de mapa.
7. Pulse **Seleccionar carpeta**.

Administración de plantillas, hojas de estilo y archivos de mapas

Puede administrar las plantillas, hojas de estilo y archivos de mapas en la ubicación local de su equipo utilizando el cuadro de diálogo Administrar plantillas, hojas de estilo y mapas. Este cuadro de diálogo le ayuda a importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y archivos de mapas de visualización en la ubicación local de su equipo.

Pulse **Administrar...** en uno de los cuadros de diálogo de selección de las plantillas, hojas de estilo o mapas.

Cuadro de diálogo Administrar plantillas, hojas de estilo y mapas...

La pestaña Plantilla muestra todas las plantillas locales. La pestaña Hoja de estilo muestra todas las hojas de estilo locales, así como visualizaciones de ejemplo con datos de muestra. Puede seleccionar una de las hojas de estilo que aplicarán sus estilos a las visualizaciones de ejemplo. Consulte el tema “Aplicación de hojas de estilo” en la página 292 para obtener más información. La pestaña Mapa muestra una lista de todos los archivos de mapas locales. Esta pestaña también muestra las leyendas incluyendo valores de muestra, comentarios si se hizo alguno al crear el mapa y una vista previa del mapa.

Los botones siguientes funcionan en cualquier pestaña que esté activada actualmente.

Importar. Importa una plantilla, hoja de estilo o archivo de mapas de visualización desde el sistema de archivos. Si importa una plantilla, hoja de estilo o archivo de mapas, estará disponible para la aplicación de IBM SPSS. Si otro usuario ha enviado una plantilla, hoja de estilo o archivo de mapas, se importa antes de utilizarlo en su aplicación.

Exportar. Exporta una plantilla, hoja de estilo o archivo de mapas de visualización al sistema de archivos. Exporte una plantilla, hoja de estilo o archivo de mapas si desea enviarlo a otro usuario.

Cambiar nombre. Cambia el nombre de la plantilla, hoja de estilo o archivo de mapas de visualización seleccionado. No puede cambiar un nombre por uno que ya esté en uso.

Exportar leyenda. Exporta las leyendas como un archivo de valores separados por comas (CSV). Este botón solo está activado en la pestaña Mapa.

Eliminar. Elimina la plantilla, hoja de estilo o archivo de mapas de visualización seleccionado. Puede seleccionar varias plantillas, hojas de estilo o archivos de mapas dejando pulsado Ctrl . No puede deshacer esta acción, por lo que debe proceder con cuidado.

Conversión y distribución de mapas en formato shapefile

El selector de plantillas de tablero le permite crear visualizaciones de mapas a partir de la combinación de una plantilla de visualización y de un archivo SMZ. Los archivos SMZ se parecen a los archivos con formato shapefile ESRI (formato de archivo SHP) en los que se contiene la información geográfica para dibujar un mapa (por ejemplo, las fronteras de los países), pero están optimizados para la visualización de mapas. El selector de plantillas de tablero tiene preinstalados varios archivos SMZ. Si ya cuenta con un formato ESRI shapefile y desea utilizarlo para visualizar mapas, primero debe convertir el formato shapefile a un archivo con formato SMZ mediante la utilidad de conversión de mapas. La utilidad de conversión de mapas admite formatos ESRI shapefile con puntos, polilíneas o polígonos (tipos de formas 1, 3 y 5) que contengan una única capa.

Además de convertir formatos ESRI shapefile, la utilidad de conversión de mapas le permite modificar el nivel de detalle de un mapa, cambiar las etiquetas de características, combinar y mover características, entre muchos otros cambios opcionales. También puede utilizar la utilidad de conversión de mapas para modificar un archivo SMZ existente (incluyendo los ya instalados).

Edición de archivos SMZ preinstalados

1. Exporte el archivo SMZ del sistema de gestión. Consulte el tema “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219 para obtener más información.
2. Utilice la utilidad de conversión de mapas para abrir y editar el archivo SMZ exportado. Se recomienda que guarde el archivo con un nombre diferente. Consulte el tema “Uso de la utilidad de conversión de mapas” en la página 221 para obtener más información.
3. Importe el archivo SMZ modificado en el sistema de gestión. Consulte el tema “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219 para obtener más información.

Recursos adicionales para archivos de mapas

Los datos geoespaciales en el formato de archivo SHP, que se deben utilizar para cubrir sus necesidades de correlación desde varios orígenes tanto privados como públicos. Consulte con los sitios Web de administraciones públicas si busca datos gratuitos. En este producto, muchas de las plantillas están basadas en datos disponibles para el público y que se obtienen en GeoCommons () y el Bureau of the Census de EE.UU. (<http://www.census.gov>).

AVISO IMPORTANTE: la información relativa a los productos que no son de IBM se ha obtenido de los proveedores de esos productos, de sus publicaciones anunciadas o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión de su rendimiento, su compatibilidad o cualquier otra reclamación relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos. Cualquier referencia a sitios Web que no sean de IBM en esta información solamente es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no son parte de los materiales de este programa de IBM, a menos que esté indicado en un archivo de avisos junto a este programa de IBM y el uso de los sitios de materiales será bajo su propia responsabilidad.

Conceptos claves de los mapas

Comprender algunos conceptos clave relacionados con el formato shapefile le ayudará a utilizar de manera eficaz la utilidad de conversión de mapas.

Un **shapefile** proporciona la información geográfica para dibujar un mapa. Existen tres tipos de formatos shapefile compatibles con la utilidad de conversión de mapas:

- **Punto.** El formato shapefile identifica las localizaciones de los puntos como, por ejemplo, las ciudades.
- **Polilínea.** El formato shapefile identifica las rutas y sus localizaciones como, por ejemplo, los ríos.
- **Polígono.** El formato shapefile identifica las regiones limítrofes y sus localizaciones, como los países.

Por lo general utilizará un formato shapefile de polígonos. Los mapas de coropletas se crean a partir de los formatos shapefile de polígonos. Los mapas de coropletas utilizan colores para representar un valor dentro de polígonos individuales (regiones). Los formatos shapefile de puntos y polilíneas se superponen normalmente en un formato shapefile de polígonos. Un ejemplo de ello es un formato shapeline de puntos de las ciudades de EE.UU. superpuesto sobre un formato shapeline con polígonos de los estados de EE.UU.

Un shapefile se compone de **características**. Las características son las entidades geográficas individuales. Por ejemplo, las características pueden ser países, estados, ciudades, etc. El shapefile también contiene datos acerca de las características. Estos datos se almacenan en **atributos**. Los atributos son similares a los campos o variables en un archivo de datos. Hay al menos un atributo que es la **leyenda del mapa** de la característica. La leyenda del mapa puede ser una etiqueta como, por ejemplo, el nombre de un país o de un estado. La leyenda del mapa es lo que vinculará a una variable o campo en un archivo de datos para crear la visualización de un mapa.

Tenga en cuenta que solamente podrá retener el atributo o los atributos clave en el archivo SMZ. La utilidad de conversión de mapas no puede guardar atributos adicionales. Esto significa que necesitará crear varios archivos SMZ si quiere agregarlos a distintos niveles. Por ejemplo, si quiere agregar estados y regiones de EE.UU., necesitará distintos archivos SMZ: uno que contenga una clave que identifique los estados y otro que contenga una clave que identifique las regiones.

Uso de la utilidad de conversión de mapas

Uso de la utilidad de conversión de mapas

Seleccione en los menús:

Herramientas > Utilidad de conversión de mapas

La utilidad de conversión de mapas consta de cuatro escenas (pasos) principales. Uno de los pasos también consta de pasos secundarios para un control más detallado de la edición del archivo de mapas.

Paso 1: Elegir los archivos de destino y de origen

Primero tiene que seleccionar un archivo de mapas de origen y uno de destino del archivo del mapa convertido. Necesitará los archivos *.shp* y *.dbf* del formato shapefile.

Seleccionar un archivo *.shp* (ESRI) o *.smz* para la conversión. Vaya a un archivo de mapas existente de su equipo. Este es el archivo que convertirá y guardará como un archivo SMZ. El archivo *.dbf* del formato shapefile *debe* almacenarse en la misma ubicación con un nombre base de archivo que coincida con el archivo *.shp*. El archivo *.dbf* es necesario porque contiene la información del atributo del archivo *.shp*.

Definir un destino y un nombre de archivo para el archivo del mapa convertido. Introduzca una ruta y un nombre de archivo para el archivo SMZ que se creará a partir del origen del mapa original.

- **Importar al selector de plantillas.** Además de guardar un archivo en el sistema de archivos, puede añadir opcionalmente el mapa a la lista Gestionar del selector de plantillas. Si selecciona esta opción, el mapa estará disponible automáticamente en el selector de plantillas para productos de IBM SPSS instalados en su equipo. Si no importa al selector de plantillas ahora, deberá realizar la importación manualmente posteriormente. Si desea obtener más información acerca de la importación de mapas al sistema de gestión del selector de plantillas, consulte “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219.

Paso 2: Elegir la leyenda del mapa

Ahora elegirá las leyendas del mapa que va a incluir en el archivo SMZ. A continuación, puede cambiar algunas opciones que afectarán a la presentación del mapa. Los pasos posteriores en la utilidad de conversión de mapas incluyen una presentación preliminar del mapa. Las opciones de presentación que elija se utilizarán para generar la presentación preliminar del mapa.

Elegir la leyenda principal del mapa. Seleccione el atributo que es la clave principal para la identificación y el etiquetado de las características del mapa. Por ejemplo, la leyenda principal de un mapa del mundo puede ser el atributo que identifica los nombres de los países. La clave principal también vinculará sus datos a las características del mapa, así que asegúrese de que los valores (etiquetas) del atributo que elija coinciden con los valores de sus datos. Al elegir un atributo, se muestran las etiquetas a modo de ejemplo. Si necesita cambiar estas etiquetas, podrá hacerlo en el último paso.

Elegir las leyendas adicionales que vaya a incluir. Además de la leyenda principal del mapa, elija otros atributos clave que desee incluir en el archivo SMZ generado. Por ejemplo, algunos atributos pueden contener etiquetas traducidas. Si quiere que los datos se codifiquen en otros idiomas, puede que desee preservar estos atributos. Tenga en cuenta que solamente puede elegir aquellas claves adicionales que representen las mismas características que la clave principal. Por ejemplo, si la clave principal fuese los nombres completos de los estados de EE.UU., puede seleccionar solamente aquellas claves alternativas que representen los estados de EE.UU., como las abreviaturas de los estados.

Suavizar automáticamente el mapa. Los shapefiles con polígonos contienen normalmente demasiados puntos de datos y demasiados detalles para las visualizaciones estadísticas de los mapas. El exceso de detalles puede distraer e impactar de forma negativa en el rendimiento. Puede reducir el nivel de detalles y generalizar el mapa con el suavizado. El mapa será más nítido y se representará más rápidamente. Cuando se suaviza el mapa automáticamente, el ángulo máximo es de 15 grados y el porcentaje que hay que mantener es 99. Para obtener más información sobre estos valores, consulte “Suavizar el mapa”. Tenga en cuenta que tiene la oportunidad de aplicar el suavizado adicional posteriormente en otro paso.

Eliminar los límites entre los polígonos en la misma característica. Algunas características pueden incluir características secundarias con límites internos en las características principales de interés. Por ejemplo, un mapa del mundo con los continentes puede contener límites internos para los países dentro de cada continente. Si selecciona esta opción, los límites internos no aparecerán en el mapa. En el ejemplo del mapa del mundo con los continentes, si selecciona esta opción se eliminarían los límites de los países y se mantendrían las fronteras de los continentes.

Paso 3: Editar el mapa

Ahora que ha especificado las opciones básicas del mapa, puede editar opciones más específicas. Estas modificaciones son opcionales. Este paso de la utilidad de conversión de mapas le guía a través de las tareas asociadas y muestra una presentación preliminar del mapa para que pueda verificar sus cambios. En función del tipo de formato shapefile (punto, polilínea o polígono) y del sistema de coordenadas, puede que algunas tareas no estén disponibles.

Cada tarea tiene los siguientes controles comunes en la parte izquierda de la utilidad de conversión de mapas.

Mostrar las etiquetas del mapa. De forma predeterminada, las etiquetas de características no se muestran en la presentación preliminar. Puede elegir mostrar las etiquetas. Aunque las etiquetas son útiles para identificar características, pueden interferir con la selección directa del mapa de presentación preliminar. Active esta opción cuando lo necesite; por ejemplo, al editar las etiquetas de características.

Aplicar color a la presentación preliminar del mapa. De forma predeterminada, el mapa de presentación preliminar muestra las áreas con un único color. Todas las características son del mismo color. Puede disponer de una gama de colores asignados a las características individuales del mapa. Esta opción puede ser útil para distinguir distintas características en el mapa. Será de especial ayuda si combina características y desea ver cómo se representan las nuevas características en la presentación preliminar.

Cada tarea tiene el siguiente control común en la parte derecha de la utilidad de conversión de mapas.

Deshacer. Pulse **Deshacer** para volver al último estado. Puede deshacer un máximo de 100 cambios.

Suavizar el mapa: Los shapefiles con polígonos contienen normalmente demasiados puntos de datos y demasiados detalles para las visualizaciones estadísticas de los mapas. El exceso de detalles puede distraer e impactar de forma negativa en el rendimiento. Puede reducir el nivel de detalles y generalizar el mapa con el suavizado. El mapa será más nítido y se representará más rápidamente. Esta opción no está disponible para los mapas de puntos y de polilíneas.

Ángulo máximo El ángulo máximo, que debe ser un valor comprendido entre 1 y 20, especifica la tolerancia al suavizado de conjuntos de puntos casi lineales. Un valor más alto permite más tolerancia al suavizado lineal y mostrará posteriormente más puntos, dando como resultado un mapa más generalizado. Para aplicar el suavizado lineal, la utilidad de conversión de mapas comprueba el ángulo interior formado por cada conjunto de tres puntos en el mapa. Si 180 menos el ángulo da un resultado inferior al valor especificado, la utilidad de conversión de mapas mostrará el punto medio. Es decir, la utilidad de conversión de mapas comprueba si la línea formada por los tres puntos es una línea casi recta. Si es así, la utilidad de conversión de mapas considerará la línea como una línea recta entre los extremos y mostrará el punto medio.

Porcentaje que hay que mantener. El porcentaje que hay que mantener, que debe ser un valor comprendido entre 90 y 100, determina la cantidad de terreno que hay que mantener cuando se suaviza el mapa. Esta opción afecta solamente a aquellas características que tienen varios polígonos, como podría ser el caso de una característica que incluye islas. Si el área total de una característica menos un polígono es superior al porcentaje especificado del área original, la utilidad de conversión de mapas mostrará el polígono del mapa. La utilidad de conversión de mapas nunca eliminará todos los polígonos de una característica. Es decir, siempre habrá al menos un polígono de la característica, con independencia de la cantidad de suavizado aplicada.

Cuando seleccione un ángulo máximo y un porcentaje, pulse en **Aplicar**. Se actualizará la presentación preliminar con los cambios de suavizado. Si necesita suavizar de nuevo el mapa, repita el proceso hasta obtener el nivel de homogeneidad deseado. Tenga en cuenta que existe un límite para aplicar el suavizado. Si suaviza repetidamente, llegará a un punto en el que no podrá aplicar más suavizado al mapa.

Editar las etiquetas de características: Puede editar las etiquetas de características según sea necesario (quizás para igualar sus datos) y también puede volver a posicionar las etiquetas en el mapa. Incluso si piensa que no necesita modificar las etiquetas, debería revisarlas antes de crear visualizaciones del mapa. Debido a que las etiquetas no se muestran de forma predeterminada en la presentación preliminar, puede que también desee seleccionar **Mostrar las etiquetas en el mapa** para mostrarlas.

Claves. Seleccione la clave que contenga las etiquetas de características que desee revisar y/o editar.

Características. Esta lista muestra las etiquetas de característica presentes en la clave seleccionada. Para editar la etiqueta, pulse dos veces sobre ella en la lista. Si se muestran las etiquetas del mapa, también puede pulsar dos veces en las etiquetas de características directamente en la presentación preliminar del mapa. Si desea comparar las etiquetas con un archivo de datos real, pulse en **Comparar**.

X/Y. Estos cuadros de texto incluyen el punto central actual de la etiqueta seleccionada de la característica del mapa. Las unidades se muestran en las coordenadas del mapa. Éstas pueden ser coordenadas locales, coordenadas cartesianas (por ejemplo, el sistema estatal de coordenadas planas) o coordenadas geográficas (donde **X** es la longitud e **Y** la latitud). Introducir las coordenadas para la nueva posición de la etiqueta. Si se muestran las etiquetas, también puede pulsar y arrastrar una etiqueta del mapa para moverla. Se actualizarán los cuadros de texto con la nueva posición.

Comparar. Si tiene un archivo de datos que contenga los valores de los datos que se pretenden igualar a las etiquetas de características de una clave determinada, pulse en **Comparar** para mostrar el cuadro de diálogo Comparar con un origen de datos externo. En este cuadro de diálogo podrá abrir el archivo de datos y comparar sus valores directamente con aquellos valores de las etiquetas de características de la leyenda del mapa.

Cuadro de diálogo Comparar con un origen de datos externo: El cuadro de diálogo Comparar con un origen de datos externo, le permite abrir un archivo de valores separados por tabuladores (con una extensión *.txt*) o un archivo de valores separados por comas (con una extensión *.csv*) o un archivo de datos con formato de IBM SPSS Statistics (con una extensión *.sav*). Cuando se abre el archivo, puede seleccionar un campo en el archivo de datos para compararlo con las etiquetas de características de una leyenda específica del mapa. A continuación, puede corregir todas las discrepancias en el archivo del mapa.

Campos en el archivo de datos. Seleccione el campo cuyos valores desee comparar con las etiquetas de características. Si la primera fila del archivo *.txt* o *.csv* contiene etiquetas descriptivas de cada campo, marque **Utilizar primera fila como etiquetas de columna**. De lo contrario, se identificará cada campo según su posición en el archivo de datos (por ejemplo, "Columna 1", "Columna 2", etc.).

Clave para comparar. Seleccione la leyenda del mapa cuyas etiquetas de características desee comparar con los valores del campo del archivo de datos.

Comparar. Pulse cuando esté preparado para comparar los valores.

Resultados comparativos. De forma predeterminada, la tabla de resultados comparativos solamente muestra los valores del campos no coincidentes en el archivo de datos. La aplicación intenta buscar una etiqueta de características relacionada, normalmente comprobando los espacios insertados o inexistentes. Pulse en la lista desplegable en la columna *Etiqueta del mapa* para hacer coincidir la etiqueta de características en el archivo del mapa con el valor del campo mostrado. Si no hay ninguna etiqueta de características correspondiente en su archivo de mapas, seleccione *Dejar sin marcar*. Si desea ver todos los valores del campo, incluso aquellos que ya coinciden con una etiqueta de características, borre la selección de **Mostrar solamente los casos no coincidentes**. Puede que desee realizar esta acción para anular una o más coincidencias.

Puede utilizar cada característica una sola vez para igualarla a un valor de campo. Si desea igualar varias características con un único valor de campo, puede combinar las características y, a continuación, igualar la nueva característica combinada con el valor del campo. Para obtener más información acerca de las características de fusión, consulte “Combinar características”.

Combinar características: La combinación de características es útil para crear regiones más grandes en un mapa. Por ejemplo, si estuviese convirtiendo un mapa de estados, podría combinar los estados (las características en este ejemplo) con regiones más grandes del Norte, Sur, Este y Oeste.

Claves. Seleccione la leyenda del mapa que contenga las etiquetas de características que le ayudarán a identificar las características que desea combinar.

Características. Pulse en la primera característica que desee combinar. Utilice la tecla Ctrl para pulsar en las otras características que desee combinar. Tenga en cuenta que también se seleccionarán las características en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl directamente en las características de la presentación preliminar del mapa además de seleccionarlas de la lista.

Después de seleccionar las características que desee combinar, pulse en **Combinar** para mostrar el cuadro de diálogo Poner nombre a la característica combinada, donde podrá aplicar una etiqueta a la nueva característica. Puede que desee marcar **Colorear la presentación preliminar del mapa** después de combinar las características para asegurarse de que los resultados obtenidos son los que esperaba.

Después de combinar las características, puede que también quiera mover la etiqueta de la nueva característica. Puede realizar esa acción en la tarea *Editar las etiquetas de características*. Consulte el tema “Editar las etiquetas de características” en la página 223 para obtener más información.

Cuadro de diálogo Poner nombre a la característica combinada: El cuadro de diálogo Poner nombre a la característica combinada, le permite asignar etiquetas a la nueva característica combinada.

La tabla Etiquetas muestra información de cada clave en el archivo de mapas y le permite asignar una etiqueta para cada clave.

Nueva etiqueta. Introduzca una nueva etiqueta para la característica combinada para asignarla a la leyenda específica del mapa.

Clave. La leyenda del mapa a la que va a asignar la nueva etiqueta.

Etiquetas anteriores. Las etiquetas de las características que se combinarán en la nueva característica.

Eliminar los límites entre los polígonos de la misma función. Compruebe esta opción para eliminar los límites de las características que se combinaron de forma conjunta. Por ejemplo, si ha combinado estados con regiones geográficas, esta opción eliminaría los límites en torno a los estados individuales.

Mover características: Puede mover las características en el mapa. Esto puede ser útil cuando desea unir características como, por ejemplo, una península e islas alejadas.

Claves. Seleccione la leyenda del mapa que contenga las etiquetas de características que le ayudarán a identificar las características que desea mover.

Características. Pulse en la característica que desee mover. Tenga en cuenta que se seleccionará la característica en la presentación preliminar del mapa. También puede pulsar directamente en la característica en la presentación preliminar del mapa.

X/Y. Estos cuadros de texto incluyen el punto central actual de la característica del mapa. Las unidades se muestran en las coordenadas del mapa. Éstas pueden ser coordenadas locales, coordenadas cartesianas (por ejemplo, el sistema estatal de coordenadas planas) o coordenadas geográficas (donde X es la longitud e Y la latitud). Introducir las coordenadas para la nueva posición de la característica. También puede pulsar y arrastrar una característica del mapa para moverla. Se actualizarán los cuadros de texto con la nueva posición.

Eliminar características: Puede eliminar las características no deseadas del mapa. Esto puede resultarle útil cuando desee eliminar algunas perturbaciones eliminando las características que no sean de interés en la visualización del mapa.

Claves. Seleccione la leyenda del mapa que contenga las etiquetas de características que le ayudarán a identificar las características que desea eliminar.

Características. Pulse en la característica que desee eliminar. Si desea eliminar varias características a la vez, pulse con la tecla Ctrl en las características adicionales. Tenga en cuenta que también se seleccionarán las características en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl en las características directamente en la presentación preliminar del mapa además de seleccionarlas de la lista.

Eliminar elementos individuales: Además de eliminar características completas, puede eliminar algunos de los elementos individuales que las componen como, por ejemplo, lagos y pequeñas islas. Esta opción no está disponible para los mapas de puntos.

Elementos. Pulse en los elementos que desee eliminar. Si desea eliminar varios elementos a la vez, pulse con la tecla Ctrl en los elementos adicionales. Tenga en cuenta que también se seleccionarán los elementos en la presentación preliminar del mapa. Puede pulsar utilizando la tecla Ctrl en los elementos directamente en la presentación preliminar del mapa además de seleccionarlos de la lista. Debido a que la lista de nombres de elementos no es descriptiva (cada elemento se asigna a un número dentro de la característica), debería comprobar la selección en la presentación preliminar del mapa para asegurarse de que ha seleccionado los elementos deseados.

Definir la proyección:

La proyección del mapa especifica la forma en la que se representa La Tierra tridimensional en dos dimensiones. Todas las proyecciones causan distorsiones. Sin embargo, algunas proyecciones son más apropiadas dependiendo de si está visualizando un mapa global o un mapa más local. Además, algunas proyecciones preservan la forma de las características originales. Las proyecciones que preservan la forma son proyecciones conformes. Esta opción está disponible solamente para los mapas con coordenadas geográficas (longitud y latitud).

A diferencia de otras opciones de la utilidad de conversión de mapas, puede cambiarse la proyección después de haber creado una visualización del mapa.

Proyección. Seleccione una proyección del mapa. Si va a crear un mapa global o hemisférico, utilice la proyección *local*, la proyección de *Mercator* o la proyección de *Winkel Tripel*. Para las áreas más pequeñas,

utilice la proyección *local*, la proyección *cónica conforme de Lambert* o la proyección *transversa de Mercator*. Todas las proyecciones utilizan el elipsoide WGS83 de los datos.

- La proyección **local** se utiliza siempre que se haya creado el mapa con un sistema de coordenadas local como, por ejemplo, el State Plane Coordinate System (sistema estatal de coordenadas planas). Estos sistemas de coordenadas se definen por las coordenadas cartesianas en lugar de las coordenadas geográficas (longitud y latitud). En la proyección local, las líneas horizontales y verticales se distribuyen por igual en un sistema de coordenadas cartesianas. La proyección local no es una proyección conforme.
- La proyección de **Mercator** es una proyección conforme para los mapas globales. Las líneas horizontales y verticales son rectas y siempre perpendiculares unas a otras. Tenga en cuenta que la proyección de Mercator se extiende hasta el infinito a medida que se va acercando a los polos Norte y Sur, de modo que no podrá utilizarla si su mapa incluye el Polo Norte o el Polo Sur. La distorsión es mayor cuando el mapa se aproxima a estos límites.
- La proyección de **Winkel Tripel** es una proyección no conforme para los mapas globales. Aunque no es una proyección conforme, ofrece un buen balance entre forma y tamaño. A excepción del Ecuador y del Meridiano de Greenwich, todas las líneas son curvas. Esta proyección es una buena opción si su mapa global incluye el Polo Norte o el Polo Sur.
- Como su nombre indica, la proyección **cónica conforme de Lambert** es una proyección conforme y se utiliza para mapas de masas continentales o masas de tierra más pequeñas que son más grandes por el Este y el Oeste que por el Norte y el Sur.
- La proyección **transversa de Mercator** es otra proyección conforme para los mapas de masas continentales y de masas de tierra más pequeñas. Utilice esta proyección para las masas de tierra que son más grandes por el Norte y el Sur en comparación con el Este y el Oeste.

Paso 4: Finalizar

En este punto puede añadir un comentario para describir el archivo de mapas y así como crear un archivo de mapas de muestra a partir de las leyendas del mapa.

Leyendas del mapa. Si hay varias claves en el archivo de mapas, seleccione una leyenda del mapa cuyas etiquetas de características desee mostrar en la presentación preliminar. Si crea un archivo de datos a partir del mapa, se utilizarán estas etiquetas para los valores de datos.

Comentar. Introduzca un comentario que describa el mapa o que ofrezca información adicional que pueda ser relevante para sus usuarios, como el origen del formato shapefile original. El comentario aparecerá en el sistema de gestión del selector de plantillas de tablero.

Crear un conjunto de datos a partir de las etiquetas de características. Marque esta opción si desea crear un archivo de datos a partir de las etiquetas de características que se muestran. Al pulsar **Examinar...**, podrá especificar una ubicación así como un nombre de archivo. Si agrega una extensión de *.txt*, se guardará el archivo como un archivo de valores separados por tabulaciones. Si agrega una extensión de *.csv*, se guardará el archivo como un archivo de valores separados por comas. Si añade una extensión *.sav*, el archivo se guardará en formato IBM SPSS Statistics. SAV es el formato predeterminado si no se especifica ninguna extensión.

Distribución de archivos de mapas

En el primer paso de la utilidad de conversión de mapas, se selecciona una ubicación en la que guardar el archivo SMZ convertido. Puede que también haya seleccionado añadir el mapa al sistema de gestión del selector de plantillas de tablero. Si elige guardar en el sistema de gestión, el mapa estará disponible en cualquier producto de IBM SPSS que ejecute en el mismo equipo.

Para distribuir el mapa a otros usuarios, necesitará enviarles el archivo SMZ. A continuación, esos usuarios pueden utilizar el sistema de gestión para importar el mapa. Simplemente puede enviar el archivo cuya ubicación ha especificado en el paso 1. Si desea enviar un archivo que se encuentra en el sistema de gestión, primero debe exportarlo:

1. En el selector de plantillas, pulse en **Gestionar...**
2. Pulse en la pestaña Mapa.
3. Seleccione el mapa que desee distribuir.
4. Pulse en **Exportar...** y elija una ubicación donde desee guardar el archivo.

Ahora puede enviar el archivo de mapas físico a otros usuarios. Los usuarios necesitarán invertir este proceso e importar el mapa al sistema de gestión.

Nodo Gráfico

Los nodos Gráfico muestran la relación entre los campos numéricos. Puede crear un gráfico con puntos (también denominado diagrama de dispersión), o puede utilizar líneas. Puede crear tres tipos de gráficos de líneas especificando un valor de Modo para X en el cuadro de diálogo.

Modo para X = Ordenar

Al establecer el Modo para X en **Ordenar**, los datos se clasifican por valores en el campo representado por el eje x. Así se ejecuta una sola línea de izquierda a derecha en el gráfico. Al utilizar un campo nominal como una superposición, se producen varias líneas de diferentes tonos que se ejecutan de izquierda a derecha en el gráfico.

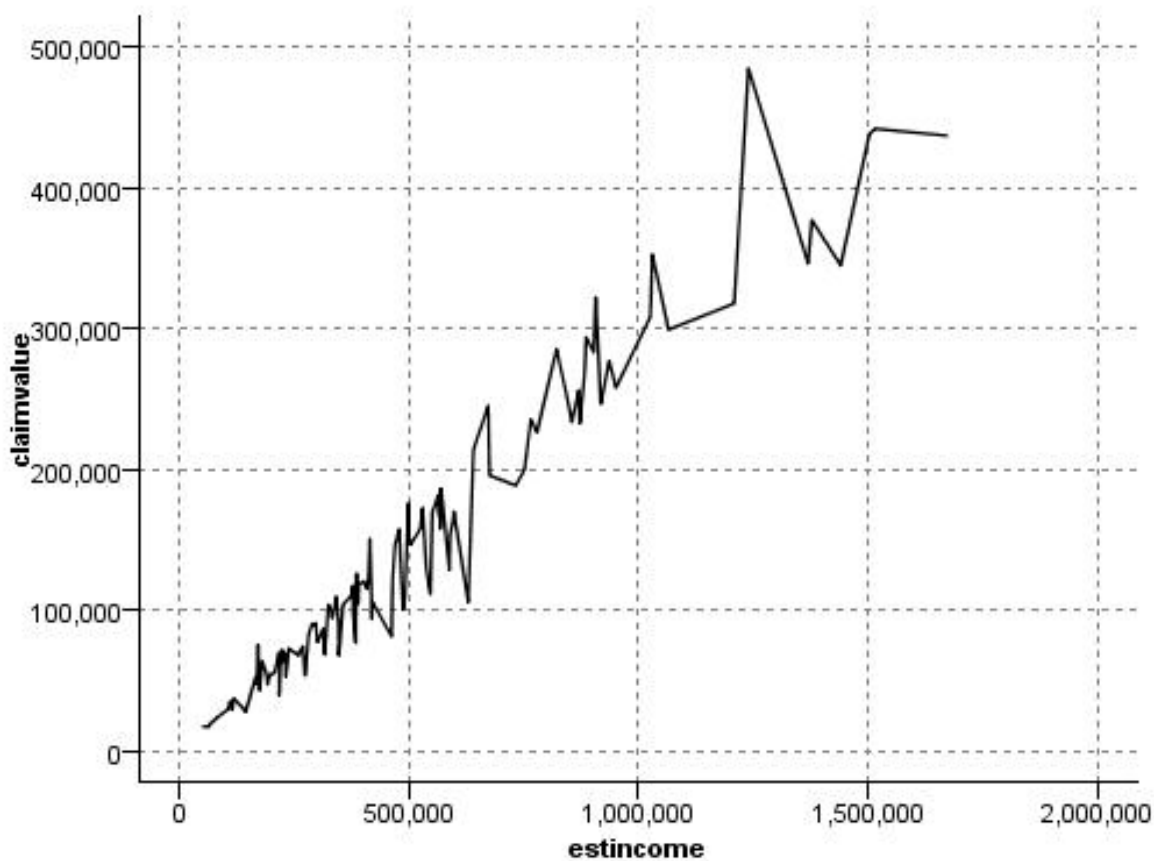


Figura 21. Gráfico de líneas con el Modo para X establecido en Ordenar

Modo para X = Superponer

Al establecer el Modo para X en **Superponer**, se crean varios gráficos de línea en el mismo gráfico. Los datos no se ordenan en el caso de un gráfico de superposición; siempre que los valores del eje x aumenten, los datos se representarán en una sola línea. Si los valores disminuyen, comienza una línea nueva. Por ejemplo, si x se mueve de 0 a 100, los valores de y se representarán en una sola línea. Cuando x cae por debajo de 100, se representa una línea nueva además de la primera. El gráfico terminado puede tener numerosos gráficos que resultan útiles para comparar varias series de valores de y . El tipo de gráfico es útil para los datos con un componente temporal periódico, como una demanda de electricidad en períodos sucesivos de 24 horas.

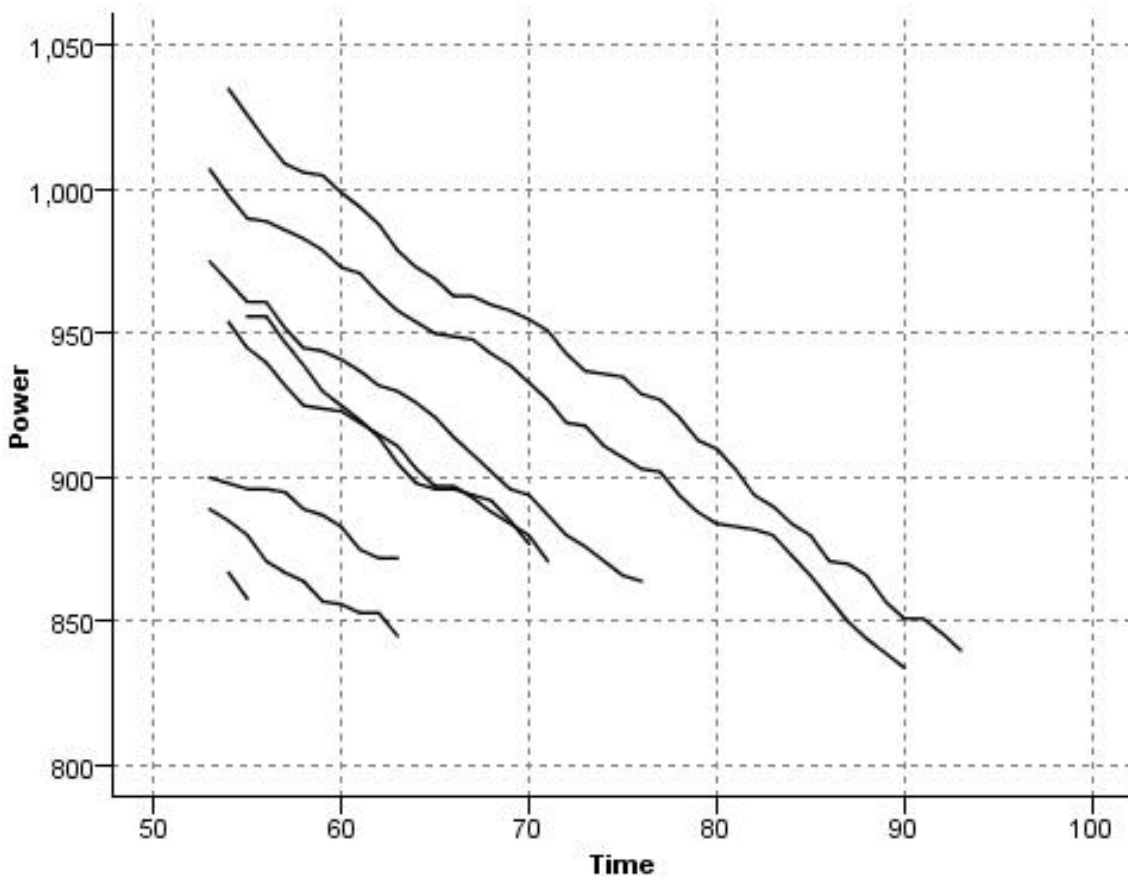


Figura 22. Gráfico de líneas con el Modo para X establecido en Superponer

Modo para X = Como se lee

Al establecer el Modo para X en **Como se lee**, los valores de x e y se representan como se leen desde el origen de datos. Esta opción es útil para los datos con un componente de serie temporal donde el interés recae sobre tendencias o patrones que dependen del orden de los datos. Puede que sea necesario ordenar los datos antes de crear este tipo de gráfico. También puede ser útil comparar dos gráficos similares con el Modo para X establecido en **Ordenar** y **Como se lee** para determinar hasta qué punto el patrón depende del orden.

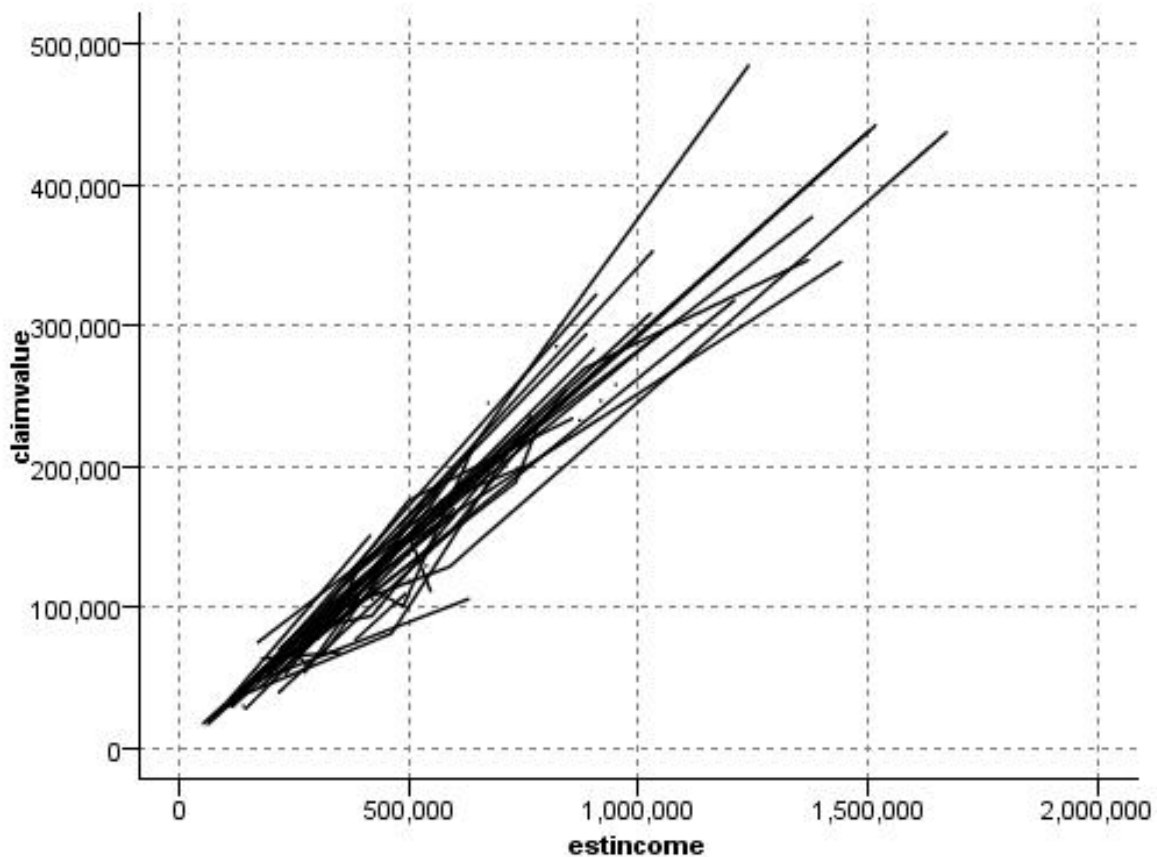


Figura 23. Gráfico de línea mostrado anteriormente como Ordenar, ejecutado de nuevo con el Modo para X establecido en Como se lee

También puede utilizar el nodo Tablero para crear diagramas de dispersión y gráficos de línea. No obstante, tendrá más opciones para seleccionar en este nodo. Consulte el tema “Tipos de visualización de Graphboard preincorporado disponibles” en la página 198 para obtener más información.

Pestaña de nodo Gráfico

Los gráficos muestran valores de un campo Y frente a los valores de un campo X. A menudo, estos campos corresponden a una variable dependiente y a una variable independiente respectivamente.

Campo X. Desde la lista, seleccione el campo que desea visualizar en el eje horizontal x .

Campo Y. Desde la lista, seleccione el campo que desea visualizar en el eje vertical y .

Campo Z. Al pulsar en el botón 3D del diagrama, puede seleccionar un campo de la lista que se muestra en el eje z .

Superponer. Existen varias formas de ilustrar las categorías de los valores de datos. Por ejemplo, puede utilizar *maincrop* (cosecha principal) como una superposición de colores para indicar los valores *estincome* y *claimvalue* para la cosecha principal cultivada por solicitantes de reclamaciones. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Tipo de superposición. Determina si se muestra una función superpuesta o suavizamiento. Las opciones suavizamiento y superpuesta se calculan siempre en función de y .

- **Ninguno.** No se muestra ningún campo de superposición.
- **Suavizamiento.** Muestra una línea de ajuste suavizada calculada con regresión de mínimos cuadrados ponderados localmente (LOESS). Este método calcula de forma eficaz una serie de regresiones, cada una enfocada a una pequeña zona del gráfico. Esto produce una serie de líneas de regresión "locales" que, a continuación, se unen para crear una curva suavizada.

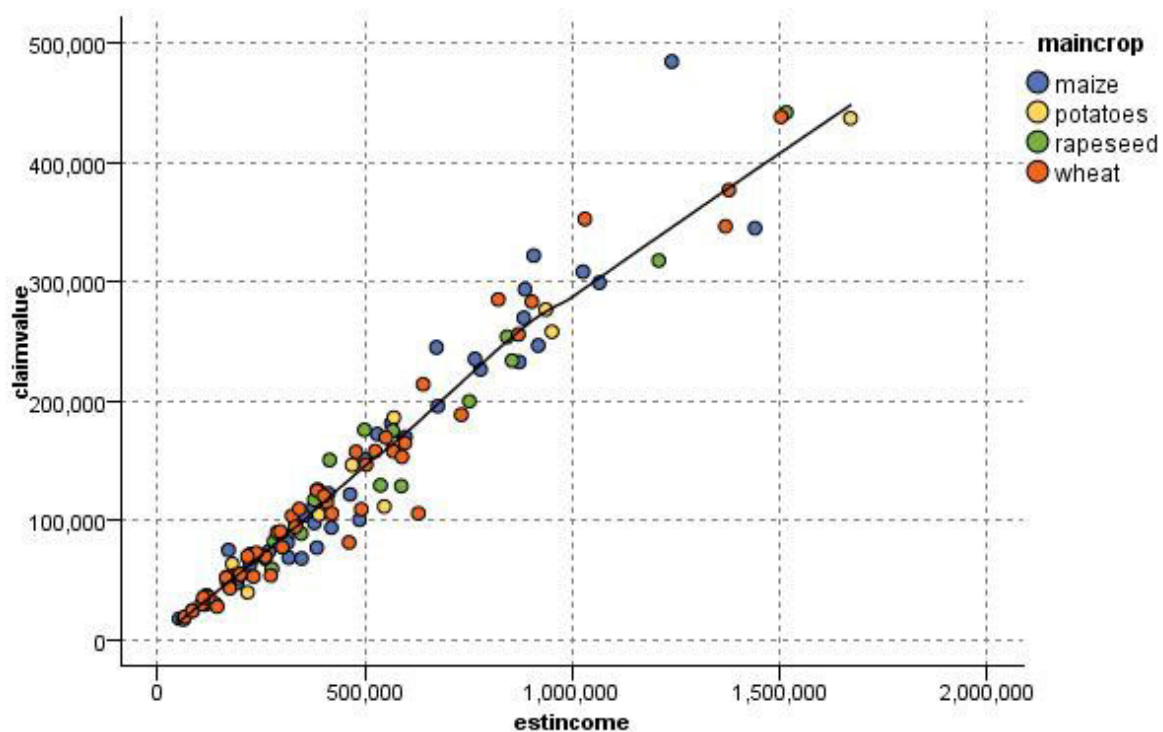


Figura 24. Gráfico con un campo de superposición de suavizamiento LOESS

- **Función.** Seleccione esta opción para especificar una función conocida que comparar con los valores reales. Por ejemplo, para comparar valores reales con valores predichos, puede representar la función $y = x$ como una superposición. Especifique una función para $y =$ en el cuadro de texto. La función predeterminada es $y = x$, pero puede especificar cualquiera, como una función cuadrática o una expresión arbitraria, en cuanto a x .

Note: las funciones superpuestas no están disponibles para un panel o gráfico de animación.

Una vez establecidas las opciones de un gráfico, puede ejecutar el gráfico directamente desde el cuadro de diálogo pulsando en **Ejecutar**. Sin embargo, es posible que desee utilizar la pestaña Opciones para establecer especificaciones adicionales, como intervalos, Modo para X y estilo.

Pestaña Opciones de Gráfico

Estilo. Permite seleccionar **Punto** o **Línea** para el estilo de gráfico. Al seleccionar **Línea**, se activa el control **Modo para X**. Si selecciona **Punto** utilizará un símbolo de adición (+) como forma de punto predeterminada. Una vez creado el gráfico, podrá cambiar la forma del punto y alterar su tamaño.

Modo para X. En el caso de los gráficos de líneas, debe elegir un Modo para X que defina el estilo del gráfico de líneas. Seleccione **Ordenar**, **Superponer** o **Como se lee**. En el caso de **Superponer** o **Como se**

lee debe especificar un tamaño del conjunto de datos máximo utilizado para muestrear los primeros n registros. De lo contrario, se utilizará el valor predeterminado de 2,000.

Rango de X automático. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores **Mín.** y **Máx.** que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan de forma predeterminada para activar la generación rápida de gráficos.

Rango de Y automático. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores **Mín.** y **Máx.** que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan de forma predeterminada para activar la generación rápida de gráficos.

Rango de Z automático. Sólo cuando se especifica un gráfico en 3D en la pestaña Gráfico. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores **Mín.** y **Máx.** que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan de forma predeterminada para activar la generación rápida de gráficos.

Perturbar. Conocido también como **agitación**, la perturbación es útil para los gráficos de puntos de un conjunto de datos en el que se repiten muchos valores. Para ver una disposición clara de los valores, puede utilizar la perturbación para distribuir los puntos aleatoriamente alrededor del valor real.

Nota para los usuarios de versiones anteriores de IBM SPSS Modeler: el valor de perturbación utilizado en un gráfico emplea una métrica distinta en este release de IBM SPSS Modeler. En versiones anteriores el valor tiene un número real, pero ahora es una proporción del tamaño del marco. Esto significa que es bastante probable que los valores de perturbación de las rutas antiguas sean demasiado altos. Para esta versión, cualquier valor de perturbación distinto de cero se convertirá en el valor 0,2.

Número máximo de registros que se van a dibujar. Permite especificar un método para trazar conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el valor predeterminado de 2.000 registros. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar las opciones **Intervalo** o **Muestra**. Si lo prefiere, puede trazar todos los puntos de datos seleccionando **Utilizar todos los datos**, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

Note: cuando la opción de Modo para X es **Superponer** o **Como se lee**, estas opciones están desactivadas y solamente se utilizan los primeros n registros.

- **Intervalo.** Seleccione esta opción para activar los intervalos cuando el conjunto de datos contiene más registros de los especificados. Los intervalos dividen el gráfico en cuadrículas finas antes del trazado en sí y realizan el recuento del número de puntos que aparecerán de cada una de las casillas de la cuadrícula. En el gráfico final se traza un punto por casilla en el centroide del intervalo (media de todas las ubicaciones de los puntos del intervalo). El tamaño de los símbolos trazados indica el número de puntos de esa región (a menos que haya utilizado el tamaño como una superposición). Al utilizar el centroide y el tamaño para representar el número de puntos, el gráfico en intervalos se convierte en una forma superior de representar conjuntos de datos grandes ya que previene un trazado excesivo en regiones densas (masas de color sin diferenciar) y reduce los artefactos de símbolos (patrones artificiales de densidad). Los artefactos de símbolos se producen cuando ciertos símbolos (en particular el símbolo de adición [+]) choca de tal manera que produce áreas densas que no existen en los datos sin procesar.
- **Muestra.** Seleccione esta opción para muestrear los datos de forma aleatoria en los registros introducidos en el campo de texto. El valor predeterminado es 2,000.

Pestaña Aspecto de Gráfico

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Z. Disponible sólo para gráficos 3D, permite aceptar la etiqueta de eje z generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta personalizada.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Utilización de un gráfico

Los gráficos y multigráficos son esencialmente gráficos de X contra Y . Por ejemplo, si explora un fraude potencial en las solicitudes de subvenciones para agricultura (tal y como se ilustra en el archivo *fraud.str* de la carpeta *Demos*, en el directorio de instalación de IBM SPSS Modeler), puede que desee trazar el ingreso reclamado en la solicitud para compararlo con el ingreso estimado por una red neuronal. Al utilizar una superposición, como el tipo de cosecha, se ilustrará si existe una relación entre las reclamaciones (valor o número) y el tipo de cosecha.

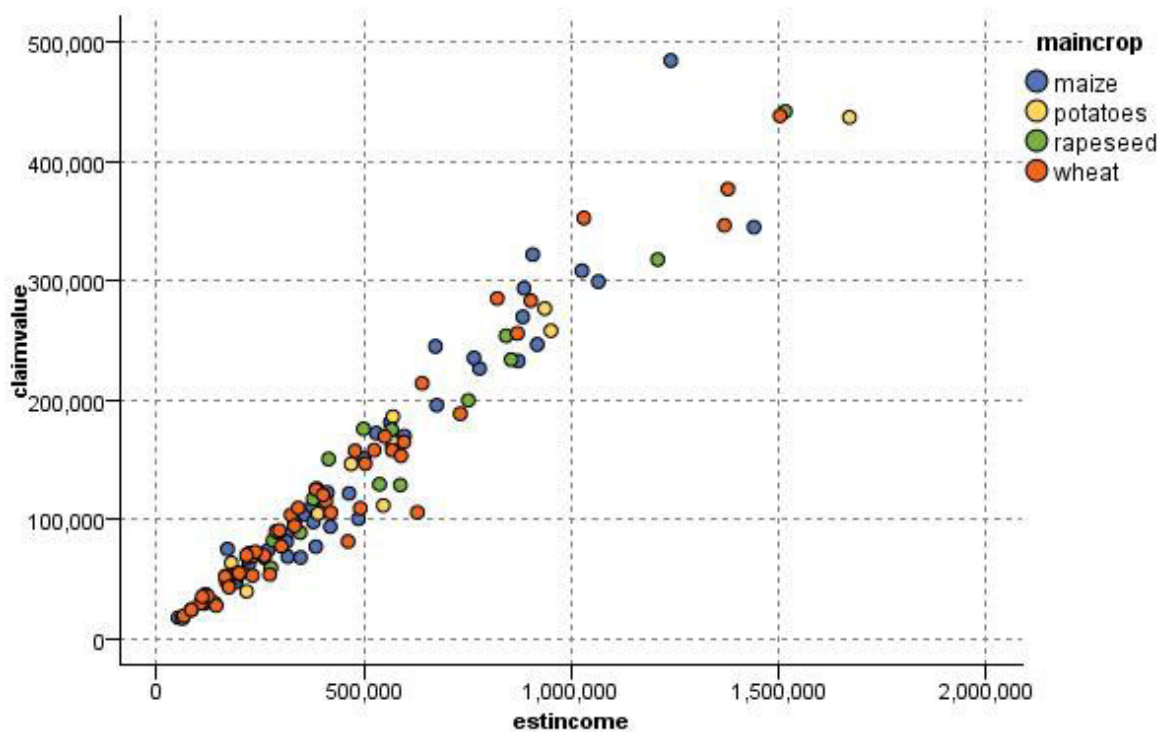


Figura 25. Gráfico de la relación entre el ingreso estimado y el valor de reclamación con el tipo de cosecha principal como una superposición

Como los gráficos, los gráficos múltiples y los diagramas de evaluación son representaciones bidimensionales de Y frente a X , es fácil interactuar con ellos definiendo regiones, marcando elementos o incluso dibujando tramos. También puede generar nodos para los datos representados por dichos tramos, elemento o regiones. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información.

Nodo G. múltiple

Un gráfico múltiple es un tipo especial de gráfico que muestra varios campos Y sobre un único campo X . Los campos Y están trazados como líneas coloreadas y cada uno equivale a un nodo Gráfico con el estilo establecido en **Línea** y Modo para X establecido en **Ordenar**. Los gráficos múltiples son útiles cuando se tienen datos de una secuencia temporal y se desea explorar la fluctuación de diversas variables durante un período de tiempo.

Pestaña Gráfico de G. múltiple

Campo X. Desde la lista, seleccione el campo que desea visualizar en el eje horizontal x .

Campos Y. Seleccione uno o varios campos de la lista para mostrarlo en el rango de los valores del campo X . Utilice el botón del selector de campos para seleccionar múltiples campos. Pulse en el botón de eliminación para eliminar campos de la lista.

Superponer. Existen varias formas de ilustrar las categorías de los valores de datos. Por ejemplo, puede utilizar una superposición de animaciones para mostrar múltiples gráficos para cada valor de los datos. Es útil para los conjuntos que contienen más de 10 categorías. Cuando se utiliza para conjuntos de más

de 15 categorías, podrá apreciar una disminución del rendimiento. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Normalizar. Seleccione esta opción para escalar todos los valores de Y para el rango 0–1 al mostrarlos en el gráfico. La normalización ayuda a explorar la relación entre líneas que, de lo contrario, estaría oculta debido las diferencias existentes en el rango de valores de cada serie, y se recomienda cuando se representan varias líneas en el mismo gráfico, o cuando se contrastan representaciones en paneles a cada lado. (La normalización no es necesaria cuando todos los valores de datos se encuentran en un rango parecido.)

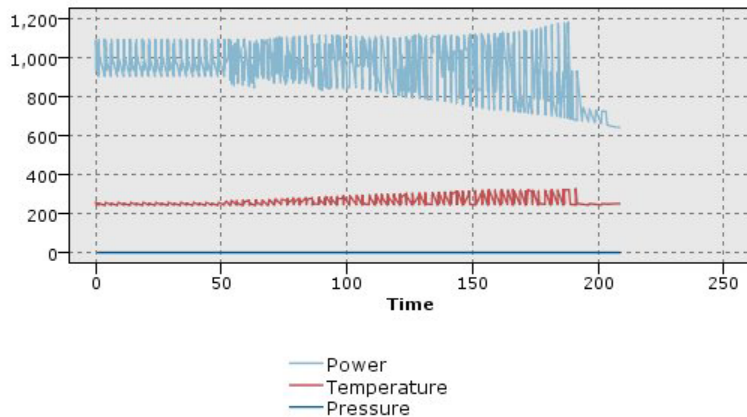


Figura 26. Gráfico múltiple estándar que muestra la fluctuación de la planta de energía durante un período de tiempo (tenga en cuenta que sin la normalización, resulta imposible ver el gráfico Pressure)

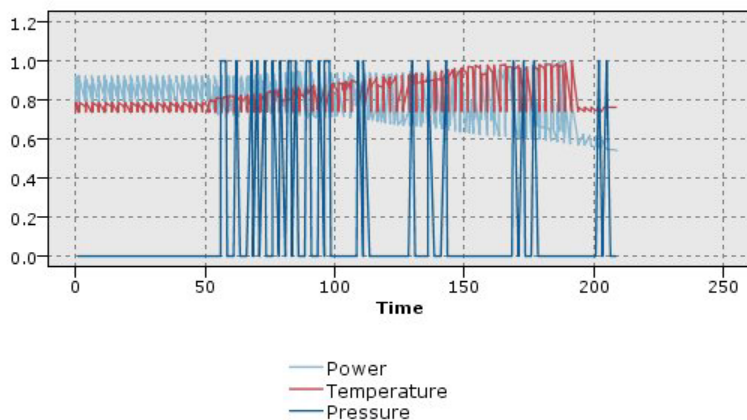


Figura 27. Gráficos múltiples normalizados que muestran un gráfico para Pressure

Función superpuesta. Seleccione esta opción para especificar una función conocida que comparar con los valores reales. Por ejemplo, para comparar valores reales con valores predichos, puede representar la función $y = x$ como una superposición. Especifique una función para $y =$ en el cuadro de texto. La función predeterminada es $y = x$, pero puede especificar cualquiera, como una función cuadrática o una expresión arbitraria, en cuanto a x .

Note: las funciones superpuestas no están disponibles para un panel o gráfico de animación.

Cuando el número de registros es mayor que. Permite especificar un método para trazar conjuntos de datos grandes. Puede especificar un tamaño de conjunto de datos máximo o utilizar el valor predeterminado de 2.000 puntos. El rendimiento mejora en el caso de conjuntos de datos grandes al seleccionar las opciones **Intervalo** o **Muestra**. Si lo prefiere, puede trazar todos los puntos de datos seleccionando **Utilizar todos los datos**, pero debe tener en cuenta que esto puede hacer que el rendimiento del software descienda radicalmente.

Note: cuando la opción de Modo para X es **Superponer** o **Como se lee**, estas opciones están desactivadas y solamente se utilizan los primeros n registros.

- **Intervalo.** Seleccione esta opción para activar los intervalos cuando el conjunto de datos contiene más registros de los especificados. Los intervalos dividen el gráfico en cuadrículas finas antes de realizar el trazado y realiza el recuento del número de conexiones que aparecerán en cada una de las casillas de la cuadrícula. En el gráfico final se utiliza una conexión por casilla en el centroide del intervalo (media de todos los puntos de conexión del intervalo).
- **Muestra.** Seleccione esta opción para muestrear los datos de forma aleatoria en el número de registros especificado.

Pestaña Aspecto de G. múltiple

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Utilización de un gráfico múltiple

Los gráficos y multigráficos son esencialmente gráficos de X contra Y . Por ejemplo, si explora un fraude potencial en las solicitudes de subvenciones para agricultura (tal y como se ilustra en el archivo *fraud.str* de la carpeta *Demos*, en el directorio de instalación de IBM SPSS Modeler), puede que desee trazar el ingreso reclamado en la solicitud para compararlo con el ingreso estimado por una red neuronal. Al utilizar una superposición, como el tipo de cosecha, se ilustrará si existe una relación entre las reclamaciones (valor o número) y el tipo de cosecha.

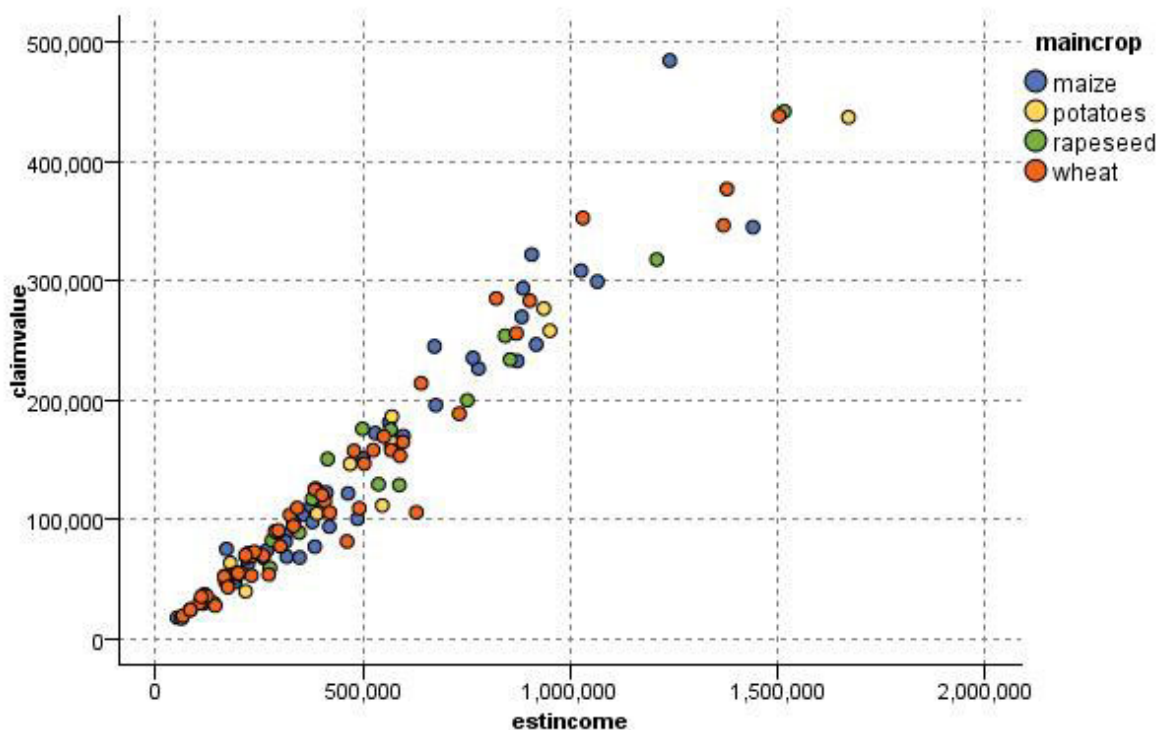


Figura 28. Gráfico de la relación entre el ingreso estimado y el valor de reclamación con el tipo de cosecha principal como una superposición

Como los gráficos, los gráficos múltiples y los diagramas de evaluación son representaciones bidimensionales de Y frente a X , es fácil interactuar con ellos definiendo regiones, marcando elementos o incluso dibujando tramos. También puede generar nodos para los datos representados por dichos tramos, elemento o regiones. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información.

Nodo Gráfico de tiempo

Los nodos Gráfico de tiempo permiten ver una o más series representadas durante un período de tiempo. La serie que represente debe contener valores numéricos que deben producirse durante un intervalo de tiempo en el que los períodos son uniformes. Se suele utilizar un nodo Intervalos de tiempo antes de un nodo Gráfico de tiempo para crear un campo *EtiquetaTiempo*, que se utiliza de forma predeterminada para etiquetar el eje x en los gráficos. Consulte el tema “Nodo de intervalos de tiempo” en la página 170 para obtener más información.

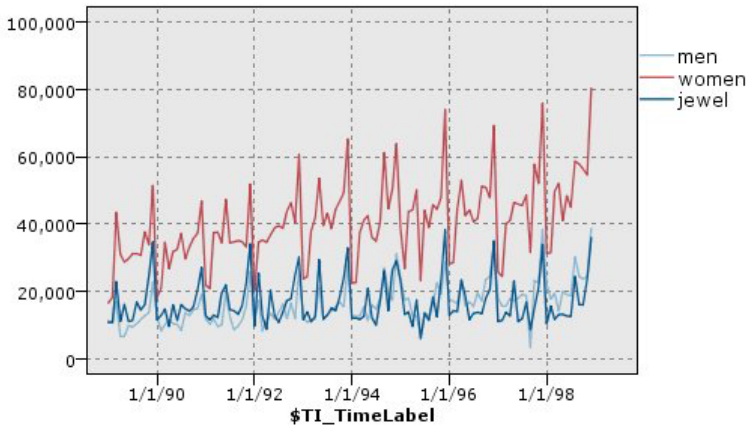


Figura 29. Representación de las ventas de ropa y joyería para hombre y mujer durante un período de tiempo

Creación de eventos e intervenciones

Puede crear campos de eventos e intervenciones a partir de los resultados del gráfico de tiempo mediante la generación de un nodo Derivar (marca o nominal) desde los menús contextuales. Por ejemplo, podría crear un campo de evento en el caso de una huelga de trenes, donde el estado sea Verdadero si se produce el evento y Falso en el caso contrario. En un campo de intervención, por ejemplo para una subida de precios, podría utilizar una derivación como recuento para identificar la fecha de la subida, con 0 para el precio anterior y 1 para el nuevo. Consulte el tema “Nodo Derivar” en la página 142 para obtener más información.

Pestaña Gráfico de Tiempo

Gráfico. Permite seleccionar el modo de representación de los datos de series.

- **Serie seleccionada.** Representa los valores para las series temporales seleccionadas. Si selecciona esta opción para representar los intervalos de confianza, anule la selección de la casilla de verificación **Normalizar**.
- **Modelos de serie temporal seleccionada.** Utilizada junto con un modelo Serie temporal, esta opción representa todos los campos relacionados (valores reales y predichos, e intervalos de confianza) para una o más series temporales seleccionadas. Esta opción deshabilita algunas de las demás opciones del cuadro de diálogo. Ésta es la opción más adecuada para la representación de intervalos de confianza.

Serie. Seleccione uno o más campos con los datos de series temporales que desea representar. Los datos deben ser numéricos.

Etiqueta de eje X. Elija la etiqueta predeterminada o un campo único para utilizarlo como etiqueta del eje x en los gráficos. Si selecciona Valor predeterminado, el sistema utilizará el campo EtiquetaTiempo creado a partir de números enteros anteriores de la ruta o secuenciales del nodo Intervalos de tiempo si no hay ningún nodo Intervalos de tiempo. Consulte el tema “Nodo de intervalos de tiempo” en la página 170 para obtener más información.

Mostrar series en paneles independientes. Determina si cada serie se muestra en un panel independiente. Como alternativa, si no selecciona ningún panel, todas las series temporales se representarán en el mismo gráfico y los suavizamientos no estarán disponibles. Al representar todas las series temporales en el mismo gráfico, cada serie se representará en un color diferente.

Normalizar. Seleccione esta opción para escalar todos los valores de Y para el rango 0–1 al mostrarlos en el gráfico. La normalización ayuda a explorar la relación entre líneas que, de lo contrario, estaría oculta debido las diferencias existentes en el rango de valores de cada serie, y se recomienda cuando se

representan varias líneas en el mismo gráfico, o cuando se contrastan representaciones en paneles a cada lado. (La normalización no es necesaria cuando todos los valores de datos se encuentran en un rango parecido.)

Mostrar. Seleccione uno o más elementos para mostrarlos en el gráfico. Puede elegir entre líneas, puntos y suavizamientos (LOESS). Los suavizamientos están disponibles solamente si muestra las series en paneles independientes. De forma predeterminada, se selecciona el elemento de línea. Asegúrese de que selecciona al menos un elemento de gráfico antes de ejecutar el nodo de gráficos; de lo contrario el sistema devolverá un error indicando que no ha seleccionado ningún elemento para representar.

Limitar registros. Seleccione esta opción si desea limitar el número de registros representados. Especifique el número de registros, leídos desde el inicio del archivo de datos, que se representarán en la opción **Número máximo de registros que se van a dibujar**. De forma predeterminada, este número se establece en 2.000. Si desea representar los últimos n registros del archivo de datos, puede utilizar un nodo Ordenar antes de este nodo para organizar los registros en orden descendente por tiempo.

Pestaña Aspecto de Gráfico de Tiempo

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Diseño. Sólo en el caso de los gráficos de tiempo, puede especificar si los valores de tiempo se representan en un eje horizontal o vertical.

Utilización de un gráfico de tiempo

Una vez creado un gráfico de tiempo, existen varias opciones para ajustar la visualización del gráfico y generar nodos para su posterior análisis. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información.

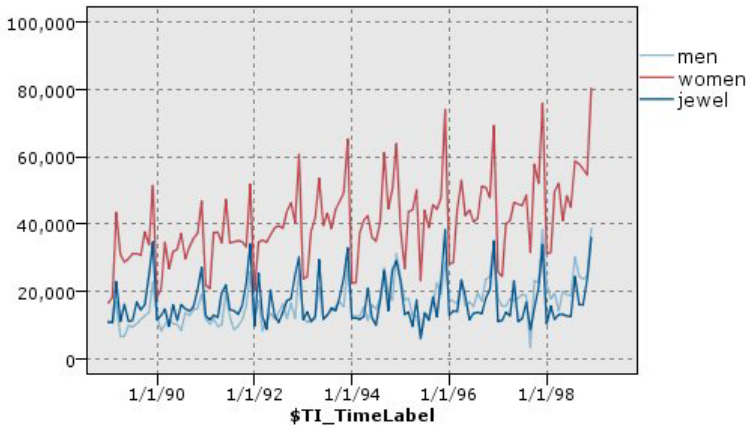


Figura 30. Representación de las ventas de ropa y joyería para hombre y mujer durante un período de tiempo

Una vez creado un gráfico de tiempo, definidos los tramos y examinados los resultados, puede utilizar las opciones del menú Generar y el menú contextual para crear nodos Seleccionar o Derivar. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

Nodo Distribución

Los gráficos o tablas de distribución muestran las instancias de valores simbólicos (no numéricos), como un género o un tipo de hipoteca, en un conjunto de datos. El uso habitual del nodo Distribución consiste en mostrar los desequilibrios de los datos que pueden rectificarse mediante el nodo Equilibrar antes de crear un modelo. Puede generar automáticamente un nodo Equilibrar mediante el menú Generar en la ventana de gráfico o tabla de distribución.

También puede utilizar el nodo Tablero para crear gráficos de barras de recuentos. No obstante, tendrá más opciones para seleccionar en este nodo. Consulte el tema “Tipos de visualización de Graphboard preincorporado disponibles” en la página 198 para obtener más información.

Note: para mostrar las instancias de valores numéricos, se debe utilizar un nodo Histograma.

Pestaña Gráfico de Distribución

Gráfico. Permite seleccionar el tipo de distribución. Seleccione **Campos seleccionados** para mostrar la distribución del campo seleccionado. Seleccione **Todas las marcas (valores verdaderos)** para mostrar la distribución de los valores verdaderos de los campos de marcas del conjunto de datos.

Campo. Permite seleccionar un campo nominal o marca para el que se va a mostrar la distribución de los valores. Solamente aparecen en la lista los campos que no se han establecido específicamente como numéricos.

Superponer. Permite seleccionar un campo nominal o marca para utilizarlo como superposición de colores, ilustrando la distribución de sus valores dentro de cada valor del campo seleccionado. Por ejemplo, puede utilizar la respuesta de la campaña de marketing (*pep*) como una superposición para el número de hijos (*hijos*) con objeto de ilustrar la respuesta por tamaño de familia. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Normalizar por color. Seleccione esta opción para escalar barras de tal forma que todas las barras ocupen el ancho total del gráfico. Los valores de superposición corresponden a una proporción de cada barra, lo que simplifica las comparaciones entre las categorías.

Ordenar. Permite seleccionar el método utilizado para mostrar valores en el gráfico de distribución. Seleccione **Alfabético** para utilizar el orden alfabético o **Por recuento** para enumerar los valores en orden de instancia decreciente.

Escala proporcional. Seleccione esta opción para escalar la distribución de valores de tal forma que el valor con el mayor recuento ocupe la anchura total del gráfico. El resto de barras se escalan en función de este valor. Al anular la selección de esta opción, las barras se escalan en función de los recuentos totales de cada valor.

Pestaña Aspecto de Distribución

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Utilización de un nodo Distribución

Los nodos Distribución se utilizan para mostrar la distribución de valores simbólicos en un conjunto de datos. Se suelen utilizar antes que los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio. Por ejemplo, si las instancias de los encuestados sin hijos se suceden con mucha más frecuencia que otros tipos de encuestados, tal vez desee reducir estas instancias para que se pueda generar una regla más útil en posteriores operaciones de minería de datos. Los nodos Distribución ayudan a examinar y tomar decisiones acerca de tales desequilibrios.

El nodo Distribución es extraño porque genera tanto un gráfico como una tabla para analizar los datos.

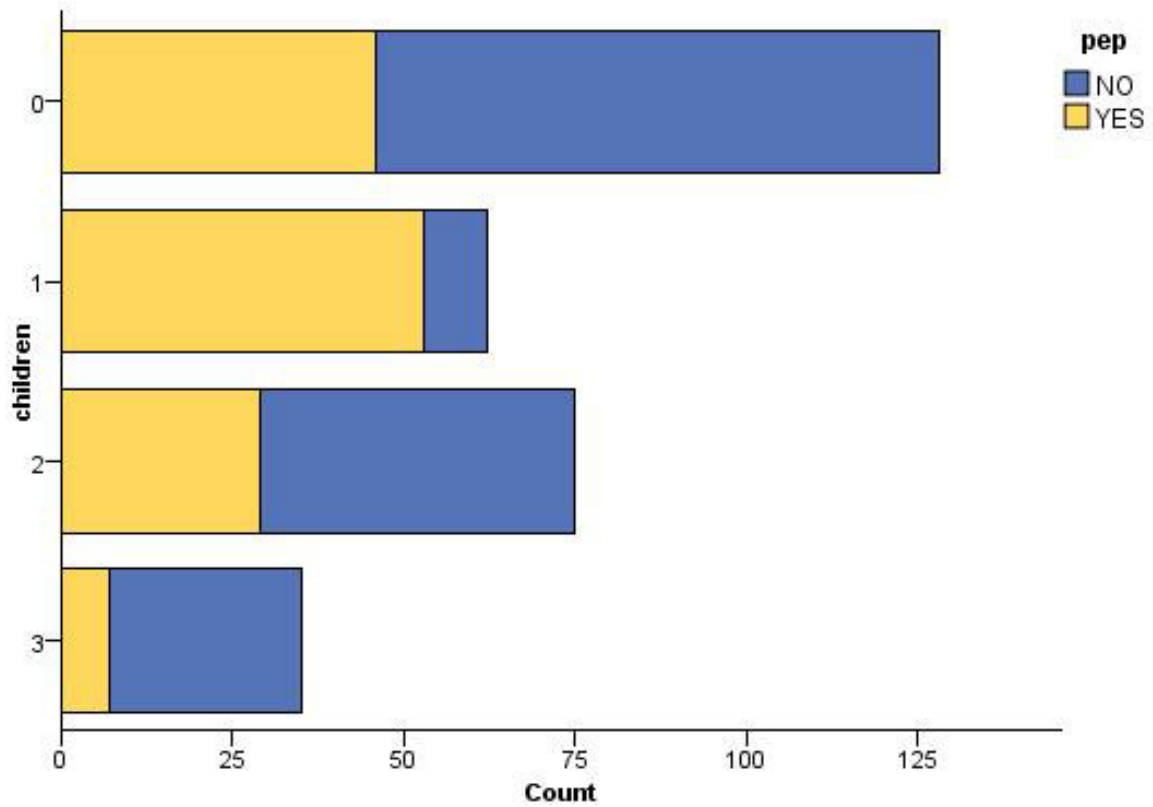


Figura 31. Gráficos de distribución con la proporción de número de personas con hijos con respuesta a una campaña de marketing

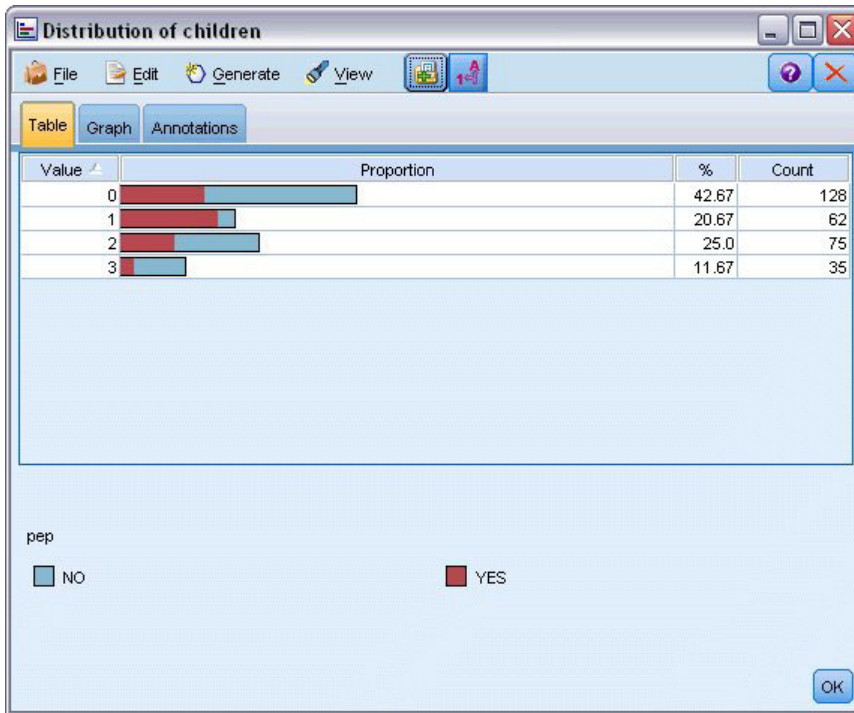


Figura 32. Tabla de distribución con la proporción de personas con o sin hijos con respuesta a una campaña de marketing

Una vez creado un gráfico y una tabla de distribución y examinados los resultados, puede utilizar las opciones de los menús para agrupar valores, copiar valores y generar un número de nodos para la preparación de datos. Además, puede copiar o exportar la información del gráfico y la tabla para utilizarla en otras aplicaciones como MS Word o MS PowerPoint. Consulte el tema “Impresión, almacenamiento, copia y exportación de gráficos” en la página 293 para obtener más información.

Para seleccionar y copiar valores desde una tabla de distribución

1. Mantenga pulsado el botón del ratón mientras lo arrastra sobre las filas para seleccionar un conjunto de valores. También puede utilizar la opción **Seleccionar todo** del menú Edición.
2. En el menú Edición, seleccione **Copiar tabla** o **Copiar tabla (inc. nombres de campos)**.
3. Péguelo en el portapapeles o en la aplicación deseada.

Note: las barras no se copian directamente. En su lugar, se copian los valores de la tabla. Esto implica que los valores superpuestos no se mostrarán en la tabla copiada.

Para agrupar valores desde una tabla de distribución

1. Seleccione valores que desea agrupar manteniendo pulsada la tecla Ctrl mientras pulsa con el ratón en los valores.
2. En el menú Edición, elija **Agrupar**.

Note: Al agrupar y desagrupar valores, el gráfico de la pestaña Gráfico se vuelve a generar de forma automática para mostrar los cambios.

También puede:

- Desagrupar valores seleccionando el nombre del grupo en la lista de distribución y eligiendo **Desagrupar** en el menú Edición.
- Editar grupos seleccionando el nombre del grupo de la lista de distribución y eligiendo **Editar grupo** en el menú Edición. Se abrirá un cuadro de diálogo donde se pueden cambiar los valores hacia y desde el grupo.

Opciones del menú Generar

Puede utilizar opciones del menú Generar para seleccionar un subconjunto de datos, derivar un campo de marcas, reagrupar valores, reclasificar valores o equilibrar los datos de un gráfico o una tabla. Estas operaciones generan un nodo de preparación de datos y lo colocan en el lienzo de rutas. Para utilizar el nodo de generación, conéctelo a una ruta existente. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

Nodo Histograma

Los nodos Histograma muestran las instancias de valores de los campos numéricos. Se suelen utilizar para explorar los datos antes de las manipulaciones y la generación de modelos. Al igual que con el nodo Distribución, con frecuencia los nodos de histogramas se utilizan para detectar desequilibrios en los datos. Al mismo tiempo puede utilizar el nodo Tablero para producir un histograma, tiene más opciones para seleccionar a partir de este nodo. Consulte el tema “Tipos de visualización de Graphboard preincorporado disponibles” en la página 198 para obtener más información.

Note: Para mostrar las instancias de valores para campos simbólicos, se debe utilizar un nodo Distribución.

Pestaña Gráfico de Histograma

Campo. Permite seleccionar un campo numérico para el que se va a mostrar la distribución de los valores. Solamente aparecen en la lista los campos que no se han definido específicamente como simbólicos (categóricos).

Superponer. Permite seleccionar un campo simbólico con objeto de mostrar categorías de valores para el campo especificado. Al seleccionar un campo de superposición, el histograma se convierte en un gráfico apilado donde los colores representan distintas categorías del campo de superposición. El nodo Histograma tiene tres tipos de superposiciones: color, panel y animación. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Pestaña Opciones de Histograma

Rango de X automático. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores **Mín.** y **Máx.** que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan de forma predeterminada para activar la generación rápida de gráficos.

Intervalos. Seleccione **Por número** o **Por ancho**.

- Seleccione **Por número** para mostrar un número fijo de barras de colección cuya anchura depende del rango especificado anteriormente y el número de buckets especificado. Indique el número de intervalos que se utilizarán en el gráfico en la opción **Número de intervalos**. Utilice las flechas para ajustar el número.
- Seleccione **Por ancho** para crear un gráfico con barras de un ancho fijo. El número de intervalos depende del ancho especificado y el rango de valores. Indique la anchura de las barras en la opción **Ancho de intervalo**.

Normalizar por color. Seleccione esta opción para ajustar todas las barras a la misma altura, mostrando los valores superpuestos como un porcentaje del número total de casos de cada barra.

Mostrar curva normal. Seleccione esta opción para añadir una curva normal al gráfico, mostrando la media y la varianza de los datos.

Separar tramos para cada color Seleccione esta opción para mostrar cada valor superpuesto como un tramo independiente en el gráfico.

Pestaña Aspecto de Histograma

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Uso de histogramas

Los histogramas muestran la distribución de valores en un campo numérico cuyos valores se extienden por el eje x . Los histogramas funcionan de forma similar a los gráficos de colecciones. Las colecciones muestran la distribución de los valores de un campo numérico *relativo a los valores de otro*, en lugar de las instancias de los valores de un solo campo.

Una vez haya creado un gráfico, puede examinar los resultados y definir los tramos para dividir los valores a lo largo del eje x o definir regiones. También puede marcar elementos dentro del gráfico. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información.

Puede utilizar opciones del menú Generar para crear nodos Equilibrar, Seleccionar o Derivar mediante los datos de gráfico o más específicamente dentro de tramos, regiones o elementos marcados. Este tipo de gráfico se suele utilizar antes de los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio al generar un nodo Equilibrar desde el gráfico que desea utilizar en la ruta. También puede generar un nodo de derivación como marca para añadir un campo que muestre el tramo donde se encuentra cada registro o un nodo Seleccionar para seleccionar todos los registros de un rango de valores o un conjunto determinado. Dichas operaciones ayudan a centrarse en un subconjunto de datos determinado para su mejor exploración. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

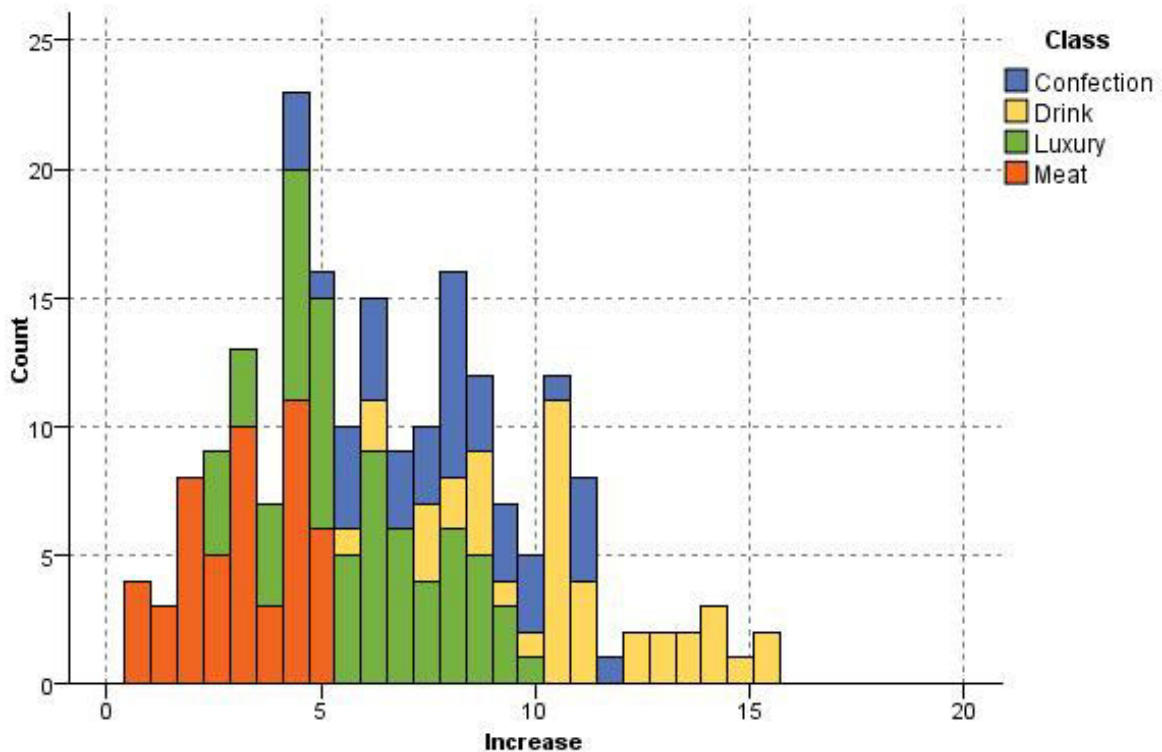


Figura 33. Histograma con la distribución de compras incrementadas por categoría debido a la promoción

Nodo Colección

Las colecciones son similares a los histogramas salvo por el hecho de que las colecciones muestran la distribución de los valores de un campo numérico relativo a los valores de otro, en lugar de las instancias de los valores de un solo campo. Las colecciones son útiles para ilustrar una variable o un campo cuyos valores cambian con el tiempo. Con los gráficos 3D también puede incluir un eje simbólico que muestra las distribuciones por categoría. Las colecciones bidimensionales se muestran como gráficos de barras apilados, con superposiciones donde se utilicen. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Pestaña Gráfico de Colección

Recolectar. Permite seleccionar un campo cuyos valores se recopilarán y mostrarán en el rango de valores para el campo especificado en **Sobre**. Solamente se enumeran los campos definidos como simbólicos.

Sobre. Permite seleccionar un campo cuyos valores se utilizan para mostrar el campo especificado en **Recolectar**.

Por. Activada al crear un gráfico 3D, esta opción permite seleccionar un campo nominal o marca utilizado para mostrar el campo de colección por categorías.

Operación. Permite seleccionar lo que representa cada barra del gráfico de colección. Las opciones son **Suma, Media, Máx, Mín y Desviación estándar**.

Superponer. Permite seleccionar un campo simbólico con objeto de mostrar categorías de valores para el campo especificado. Al seleccionar un campo de superposición, la colección se convierte y se crean múltiples barras de distintos colores para cada categoría. Este nodo tiene tres tipos de superposiciones: color, panel y animación. Consulte el tema “Aspectos estéticos, superposiciones, paneles y animación” en la página 188 para obtener más información.

Pestaña Opciones de Colección

Rango de X automático. Seleccione esta opción para utilizar todo el rango de valores de los datos en el eje. Anule la selección para utilizar un subconjunto explícito de valores en función de los valores **Mín.** y **Máx.** que especificó. Introduzca los valores o utilice las flechas. Los rangos automáticos se seleccionan de forma predeterminada para activar la generación rápida de gráficos.

Intervalos. Seleccione **Por número** o **Por ancho**.

- Seleccione **Por número** para mostrar un número fijo de barras de colección cuya anchura depende del rango especificado anteriormente y el número de buckets especificado. Indique el número de intervalos que se utilizarán en el gráfico en la opción **Número de intervalos**. Utilice las flechas para ajustar el número.
- Seleccione **Por ancho** para crear un gráfico con barras de un ancho fijo. El número de intervalos depende del ancho especificado y el rango de valores. Indique la anchura de las barras en la opción **Ancho de intervalo**.

Pestaña Aspecto de Colección

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Etiqueta Sobre. Permite aceptar la etiqueta generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta Recolectar. Permite aceptar la etiqueta generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta Por. Permite aceptar la etiqueta generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

El siguiente ejemplo muestra dónde se colocan las opciones de aspecto en una versión 3D del gráfico.

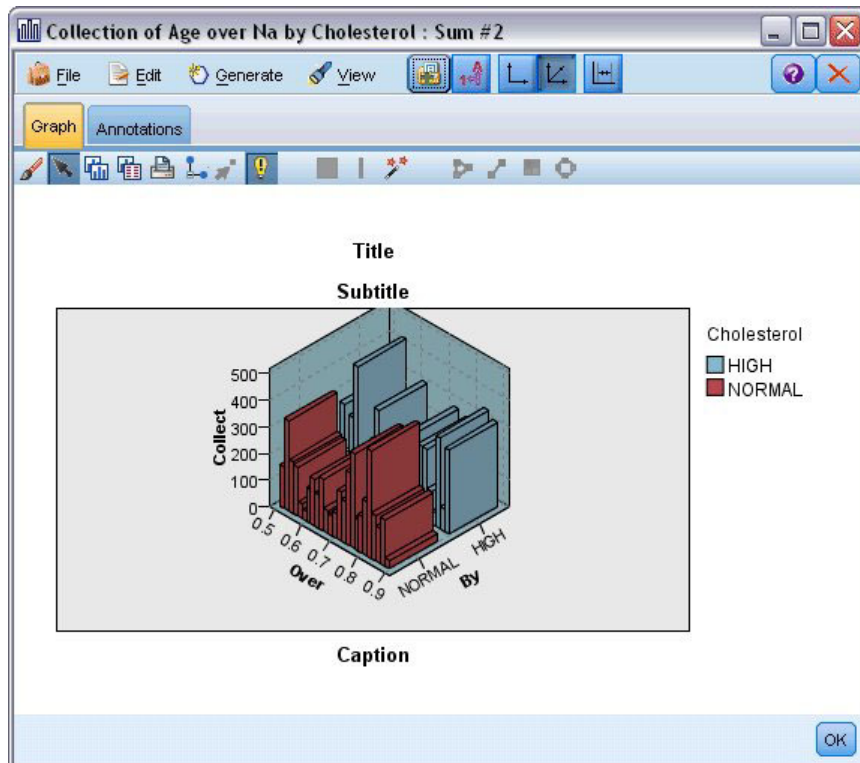


Figura 34. Posición de las opciones de aspecto de gráfico en gráfico de colección 3D

Utilización de un gráfico de colección

Las colecciones muestran la distribución de los valores de un campo numérico *relativo a los valores de otro*, en lugar de las instancias de los valores de un solo campo. Los histogramas funcionan de forma similar a los gráficos de colecciones. Los histogramas muestran la distribución de valores en un campo numérico cuyos valores se extienden por el eje x .

Una vez haya creado un gráfico, puede examinar los resultados y definir los tramos para dividir los valores a lo largo del eje x o definir regiones. También puede marcar elementos dentro del gráfico. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información.

Puede utilizar opciones del menú Generar para crear nodos Equilibrar, Seleccionar o Derivar mediante los datos de gráfico o más específicamente dentro de tramos, regiones o elementos marcados. Este tipo de gráfico se suele utilizar antes de los nodos de manipulación para explorar los datos y corregir cualquier desequilibrio al generar un nodo Equilibrar desde el gráfico que desea utilizar en la ruta. También puede generar un nodo de derivación como marca para añadir un campo que muestre el tramo donde se encuentra cada registro o un nodo Seleccionar para seleccionar todos los registros de un rango de valores o un conjunto determinado. Dichas operaciones ayudan a centrarse en un subconjunto de datos determinado para su mejor exploración. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

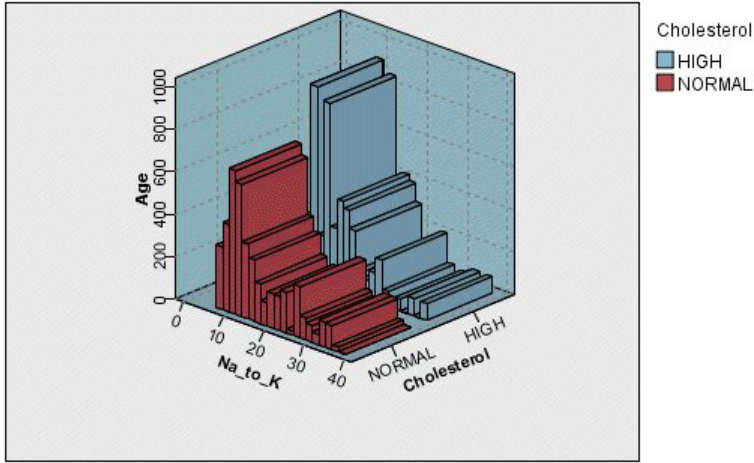


Figura 35. Gráficos de colección 3D mostrando la suma de Na_K sobre Age para niveles de colesterol altos y normales

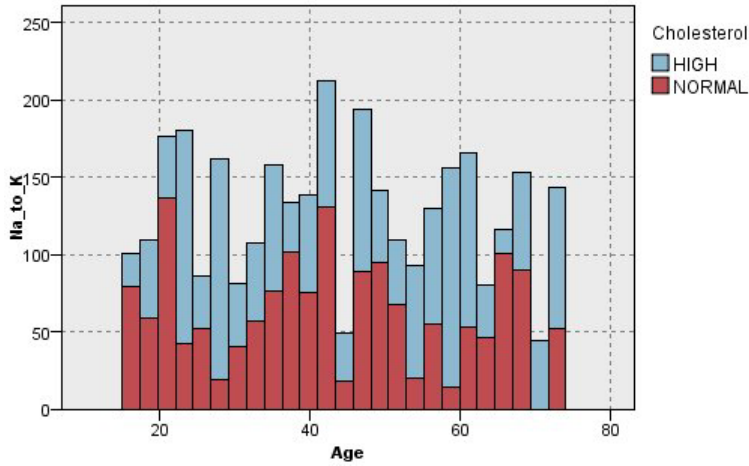


Figura 36. Gráfico de colección sin eje z pero con el colesterol como superposición de colores

Nodo Malla

Los nodos Malla muestran la fuerza de las relaciones entre los valores de dos o más campos simbólicos. El gráfico muestra las conexiones usando varios tipos de líneas para indicar la fuerza de conexión. Puede utilizar un nodo Malla, por ejemplo, para explorar las relaciones existentes entre la compra de varios artículos en un sitio de comercio electrónico o un punto de venta al por menor tradicional.

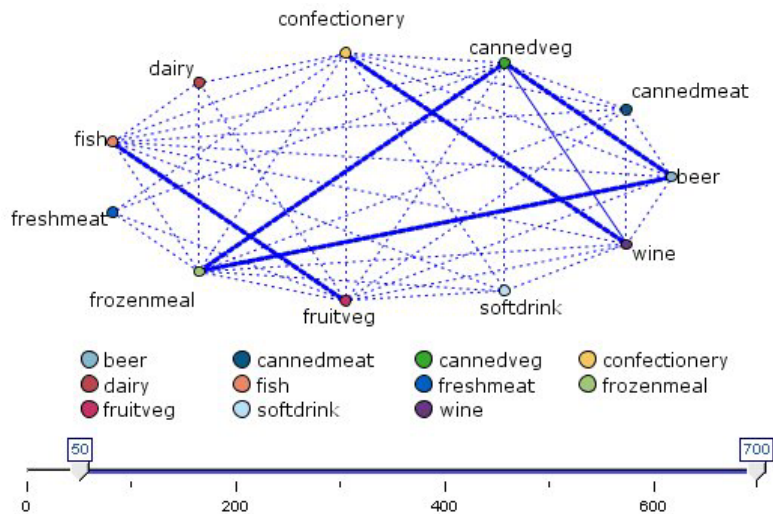


Figura 37. Gráfico de malla que muestra las relaciones entre la compra de comestibles

Mallas direccionales

Los nodos de mallas direccionales son similares a los nodos Mallas en cuanto a que muestran la fuerza de las relaciones entre campos simbólicos. Sin embargo, los gráficos de mallas direccionales muestran solamente las conexiones de uno o más campos de origen (Desde) con un único campo de destino (Hacia). Las conexiones son unidireccionales en el sentido de que son conexiones de una sola dirección.

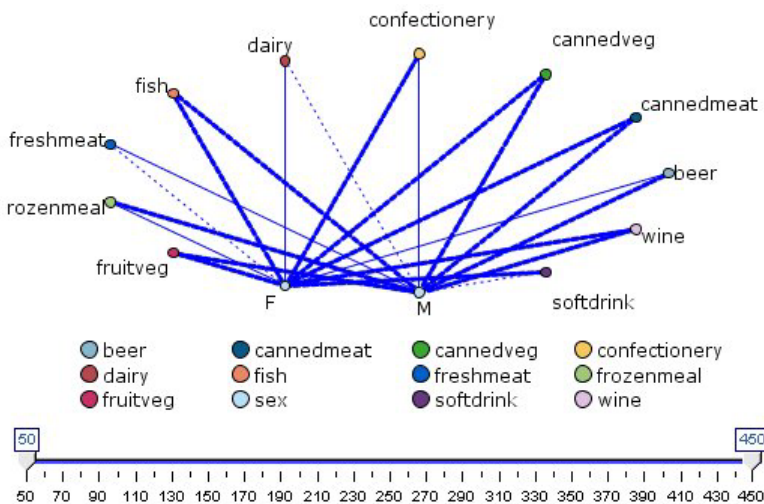


Figura 38. Gráfico de malla direccional que muestra la relación existente entre la compra de comestibles y el género

A semejanza de los nodos Malla, el gráfico muestra las conexiones usando varios tipos de líneas para indicar la fuerza de conexión. Puede utilizar un nodo de malla direccional, por ejemplo, para explorar las relaciones entre el género y una propensión a ciertos artículos de compra.

Pestaña Gráfico de Malla

Malla. Seleccione esta opción para crear un gráfico de malla que ilustre la fuerza de las relaciones entre todos los campos especificados.

Malla direccional. Seleccione esta opción para crear un gráfico de malla direccional que ilustre la fuerza de las relaciones entre múltiples campos y los valores de un campo, como el género o la religión. Cuando

esta opción está seleccionada, se activa un Campo Hacia y el nombre del control de campos que aparece a continuación cambia a Campos Desde para ofrecer mayor claridad.

Campo Hacia (solamente malla direccional). Permite seleccionar un campo nominal o marca utilizados para crear una malla direccional. Solamente se enumeran los campos establecidos explícitamente como numéricos.

Campos/Campos Desde. Permite seleccionar campos para crear un gráfico de malla. Solamente se enumeran los campos establecidos explícitamente como numéricos. Utilice el selector de campos para elegir varios campos o campos por tipo.

Note: en las mallas direccionales, este control se utiliza para seleccionar campos Desde.

Mostrar solamente marcas verdaderas. Seleccione esta opción para mostrar solamente marcas verdaderas para un campo de marcas. Esta opción simplifica la visualización de la malla y suele utilizarse para datos donde las instancias de valores positivos son de especial importancia.

Los valores de líneas son. Seleccione un tipo de umbral de la lista desplegable.

- **Absolutos** establece umbrales en función del número de registros que tengan cada par de valores.
- **Porcentajes globales** muestra el número de casos absolutos representados por el enlace como una proporción de todas las instancias de cada par de valores representados en el gráfico de malla.
- **Porcentajes del campo/valor menor y Porcentajes del campo/valor mayor** indican el campo o valor que se debe utilizar para evaluar los porcentajes. Por ejemplo, suponga que 100 registros tienen el valor *drugY* para el campo *Drug* y sólo 10 tienen el valor *LOW* para el campo *BP*. Si siete registros tienen ambos valores *drugY* y *LOW*, este porcentaje es 70% o 7%, en función del campo al que se está haciendo referencia, (*PS*) más pequeño o (*Drug*) más grande.

Note: en el caso de gráficos de mallas direccionales, las opciones tercera y cuarta anteriores no están disponibles. En su lugar, puede seleccionar **Porcentaje del campo/valor "Hasta"** y **Porcentaje del campo/valor "Desde"**.

Los enlaces fuertes aparecen más gruesos. Esta opción, seleccionada de forma predeterminada, es la forma estándar de ver el tipo de enlace entre los campos.

Los enlaces débiles aparecen más gruesos. Seleccione esta opción para invertir el significado de los enlaces indicados por líneas gruesas. Esta opción se suele utilizar para la detección de fraudes o el examen de valores atípicos.

Pestaña Opciones de Malla

La pestaña Opciones de los nodos Malla contiene un número de opciones adicionales para personalizar el gráfico de resultados.

Número de enlaces. Las siguientes opciones se utilizan para controlar el número de enlaces que se muestra en el gráfico de resultados. Algunas de estas opciones, como **Enlaces débiles por encima de** y **Enlaces fuertes por encima de** están también disponibles en la ventana del gráfico de resultado. También puede utilizar un control deslizante en el gráfico final para ajustar el número de enlaces mostrado.

- **Número máximo de enlaces que se van a mostrar.** Especifique el número máximo de enlaces que mostrar en el gráfico de resultados. Utilice las flechas para ajustar el valor.
- **Mostrar sólo enlaces por encima de.** Especifique un valor mínimo para el que mostrar una conexión en la malla. Utilice las flechas para ajustar el valor.
- **Mostrar todos los enlaces.** Especifique esta opción para mostrar todos los enlaces independientemente de los valores mínimo y máximo. Al seleccionar esta opción, el tiempo de procesamiento puede aumentar si existe un número de campos elevado.

Descartar si hay muy pocos registros. Seleccione esta opción para omitir las conexiones admitidas por muy pocos registros. Establezca el umbral para esta opción especificando un número en **Núm. mínimo de registros/líneas**.

Descartar si hay demasiados registros. Seleccione esta opción para omitir conexiones fuertemente admitidas. Introduzca un número en **Núm. máx de registros/líneas**.

Enlaces débiles por debajo de. Permite especificar un número para indicar el umbral para las conexiones débiles (líneas punteadas) y las conexiones habituales (líneas normales). Todas las conexiones que estén por debajo de este valor se consideran débiles.

Enlaces fuertes por encima de. Permite especificar un umbral para las conexiones fuertes (líneas fuertes) y las conexiones habituales (líneas normales). Todas las conexiones que estén por encima de este valor se consideran fuertes.

Tamaño de enlace. Permite especificar opciones para controlar el tamaño de los enlaces:

- **El tamaño de enlace varía de manera continua.** Seleccione esta opción para mostrar un rango de tamaños de enlace que refleje la variación de las fuerzas de conexión en función de los valores de datos reales.
- **El tamaño de enlace muestra las categorías fuerte/normal/débil.** Seleccione esta opción para mostrar tres fuerzas de conexión (fuerte, normal y débil). Los puntos de corte para estas categorías pueden especificarse con anterioridad y en el gráfico final.

Visualización de malla. Permite seleccionar un tipo de malla que mostrar:

- **Diseño circular.** Seleccione esta opción para utilizar la visualización de malla estándar.
- **Diseño de red.** Seleccione esta opción para utilizar un algoritmo para agrupar los enlaces más fuertes. Esta opción está pensada para resaltar enlaces fuertes usando diferenciaciones espaciales y líneas ponderadas.
- **Diseño direccional.** Seleccione esta opción para crear un diseño de malla direccional que utilice la selección **Campo Hacia** de la pestaña Gráfico como centro de la dirección.
- **Diseño de cuadrícula.** Seleccione esta opción para crear un diseño de malla en un patrón de cuadrícula con espacios regulares.

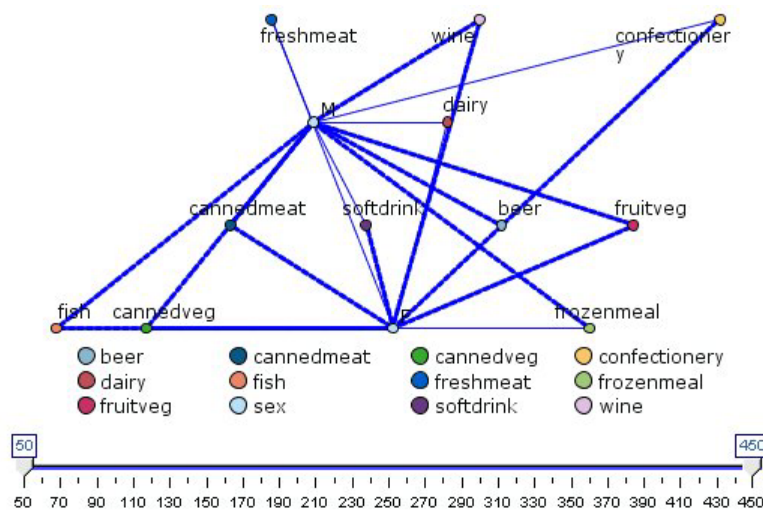


Figura 39. Gráfico Malla que muestra conexiones fuertes desde congelados y lata_veg a otros comestibles

Pestaña Aspecto de Malla

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Mostrar leyenda. Puede especificar si se visualiza la leyenda. En el caso de los gráficos con un número elevado de campos, el aspecto del gráfico puede mejorar al ocultar la leyenda.

Utilizar etiquetas como nodos. Puede incluir el texto de la etiqueta en cada nodo en lugar de mostrar las etiquetas adyacentes. En el caso de los gráficos con un número pequeño de campos, el gráfico puede resultar más legible.

Relationship between gender and grocery purchases

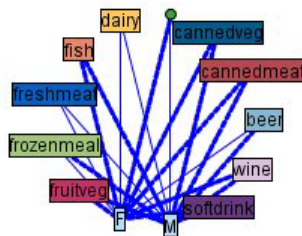


Figura 40. Gráfico Malla que muestra etiquetas como nodos

Utilización de un gráfico de malla

Los nodos Malla se utilizan para ilustrar la fuerza de las relaciones existentes entre los valores de dos o más campos simbólicos. Las conexiones se muestran en un gráfico con distintos tipos de líneas para indicar conexiones de creciente fuerza. Por ejemplo, puede utilizar un nodo Malla para explorar la relación existente entre los niveles de colesterol, la presión sanguínea y los medicamentos que resultaron eficaces en el tratamiento de las enfermedades de los pacientes.

- Las conexiones fuertes se indican con una línea gruesa. Así se indica que los dos valores están fuertemente relacionados y se deben explorar más.
- Las conexiones medias se muestran con una línea de grosor normal.
- Las conexiones débiles se muestran con una línea punteada.
- Si no se muestra ninguna línea entre dos valores, éstos nunca se dan en el mismo registro o esta combinación se produce en un número de registros inferior al umbral especificado en el cuadro de diálogo del nodo Malla.

Una vez creado un nodo Malla, existen varias opciones para ajustar la visualización del gráfico y generar nodos para su posterior análisis.

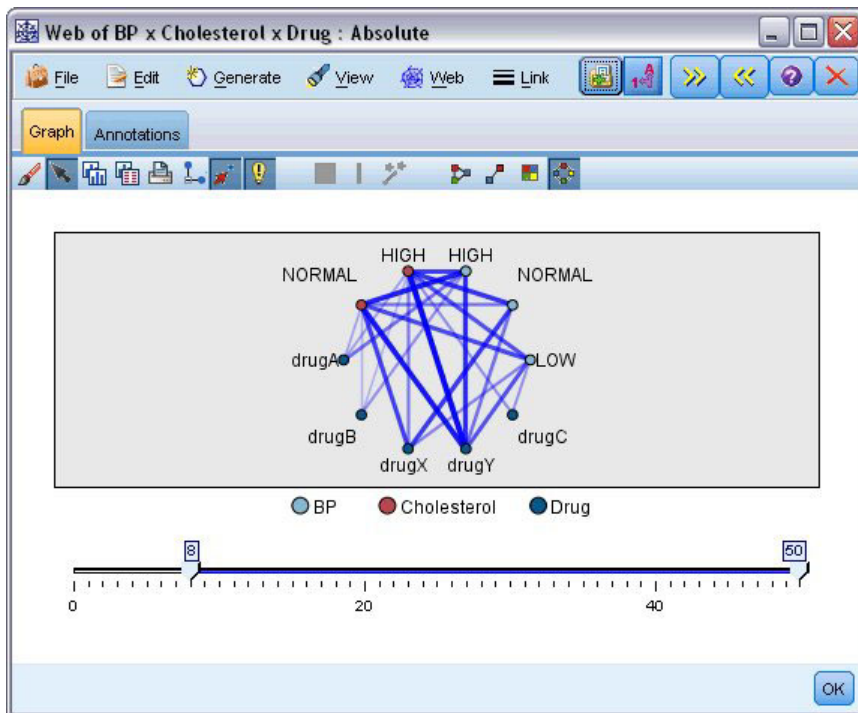


Figura 41. Gráfico de malla que indica un número de relaciones fuertes, como presión sanguínea normal con drugX y colesterol alto con drugY.

En el caso de los nodos Malla y los nodos de mallas direccionales, se puede:

- Cambiar el diseño de la visualización de la malla.
- Ocultar puntos para simplificar la visualización.
- Cambiar los umbrales que controlan los estilos de línea.
- Resaltar las líneas situadas entre los valores para indicar una relación "seleccionada".
- Generar un nodo Seleccionar para uno o varios registros "seleccionados" o un nodo de derivación como marca asociado con una o varias relaciones en la malla.

Para ajustar puntos

- **Mueva** puntos pulsando con el ratón en ellos y arrastrándolos a una nueva ubicación. La malla se volverá a generar para reflejar la nueva ubicación.
- **Oculte** puntos de la malla pulsando con el botón derecho del ratón en ellos y seleccionando **Ocultar u Ocultar y volver a planear** en el menú contextual. **Ocultar** solamente oculta el punto seleccionado y cualquier línea asociada a él. **Ocultar y volver a planear** vuelve a dibujar la malla, ajustándola a los cambios realizados. Los movimientos manuales se deshacen.
- **Muestre** todos los puntos ocultos seleccionando **Mostrar todo** o **Mostrar todo y volver a planear** en el menú de malla de la ventana del gráfico. Al seleccionar **Mostrar todo y volver a planear**, se vuelve a dibujar la malla y se ajusta para incluir todos los puntos ocultos previamente y sus conexiones.

Para seleccionar o "resaltar" líneas

Las líneas seleccionadas están resaltadas en rojo.

1. Para seleccionar una única línea, pulse con el botón izquierdo sobre la línea.
2. Para seleccionar varias líneas, realice una de las siguientes acciones:
 - Utilizando el cursor, dibuje un círculo alrededor de los puntos de las líneas que desee seleccionar.
 - Mantenga pulsada la tecla Ctrl y pulse con el botón izquierdo las líneas que desee seleccionar.

Puede anular la selección de todas las líneas seleccionadas pulsando sobre el fondo del gráfico o seleccionando **Borrar selección** en el menú Malla de la ventana del gráfico.

Para ver la malla mediante un diseño diferente

En el menú Malla, seleccione **Diseño circular**, **Diseño de red**, **Diseño direccional** o **Diseño de cuadrícula** para cambiar el diseño del gráfico.

Para activar o desactivar el control deslizante de enlaces

En el menú Ver, seleccione **Control deslizante de enlaces**.

Para seleccionar o marcar registros de una relación simple

1. Pulse con el botón derecho del ratón en la línea que representa la relación de interés.
2. En el menú contextual, elija **Generar nodo Seleccionar para el enlace** o **Generar nodo Derivar para el enlace**.

Automáticamente se añade un nodo Seleccionar o un nodo Derivar al lienzo de rutas con las opciones y condiciones especificadas adecuadas:

- El nodo Seleccionar selecciona todos los registros de una relación dada.
- El nodo Derivar genera una marca que indica si la relación seleccionada es verdadera para los registros de todo el conjunto de datos. El campo de marcas toma su nombre de la unión de dos valores de la relación unidos por guión bajo (por ejemplo *LOW_drugC* o *drugC_LOW*).

Para seleccionar o marcar registros para un grupo de relaciones

1. Seleccione las líneas de la visualización de malla que representan relaciones de interés.
 2. En el menú Generar de la ventana del gráfico, elija **Nodo Seleccionar ("Y")**, **Nodo Seleccionar ("O")**, **Nodo Derivar ("Y")** o **Nodo Derivar ("O")**.
- El nodo "O" proporciona la disyunción de condiciones. Esto significa que el nodo se aplicará a los registros para los que cualquiera de las relaciones seleccionadas sea válida.
 - El nodo "Y" proporciona la conjunción de condiciones. Esto significa que el nodo se aplicará solamente a los registros para los que todas las relaciones seleccionadas sean válidas. Se producirá un error si cualquiera de las relaciones seleccionadas es mutuamente excluyente.

Una vez completada la selección, se añade un nodo Seleccionar o un nodo Derivar automáticamente al lienzo de rutas con las opciones y condiciones especificadas adecuadas.

Ajuste de umbrales de malla

Una vez creado el gráfico de malla, puede ajustar los umbrales que controlan los estilos de línea mediante la barra de herramientas para cambiar la línea mínima visible. También puede ver opciones de umbrales adicionales al pulsar en el botón de flecha doble amarilla de la barra de herramientas para expandir la ventana de gráfico de malla. A continuación, pulse en la pestaña **Controles** para ver opciones adicionales.

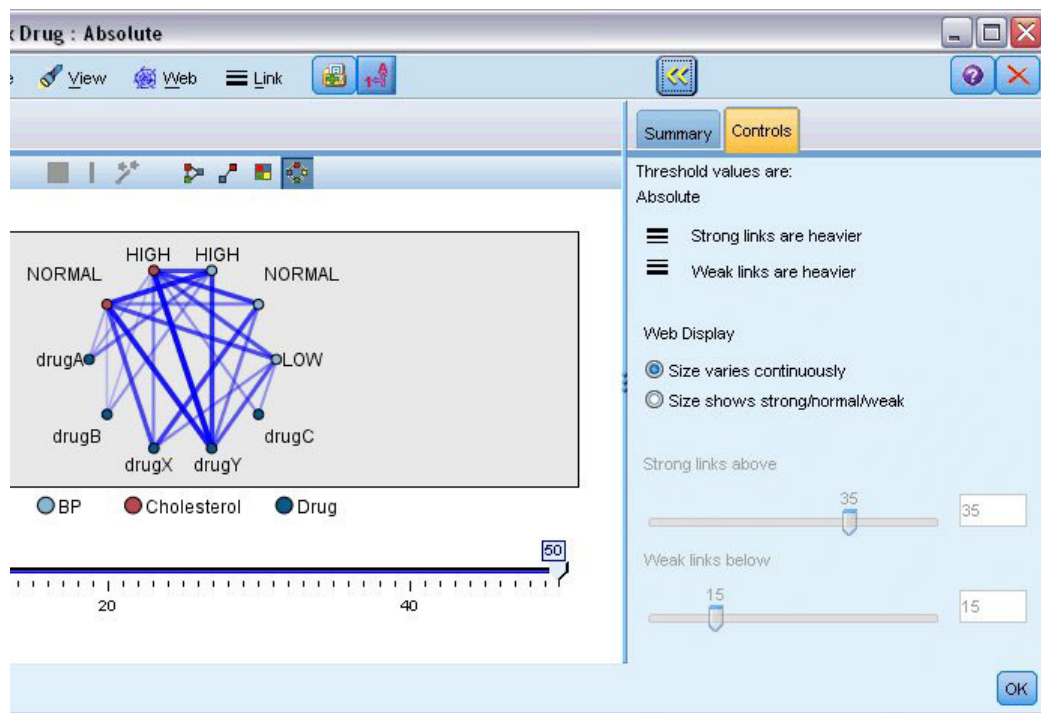


Figura 42. Ventana expandida que muestra las opciones de umbral y presentación

Los valores de umbral son. Muestra el tipo de umbral seleccionado durante la creación en el cuadro de diálogo del nodo Malla.

Los enlaces fuertes aparecen más gruesos. Esta opción, seleccionada de forma predeterminada, es la forma estándar de ver el tipo de enlace entre los campos.

Los enlaces débiles aparecen más gruesos. Seleccione esta opción para invertir el significado de los enlaces indicados por líneas gruesas. Esta opción se suele utilizar para la detección de fraudes o el examen de valores atípicos.

Visualización de malla. Permite especificar las opciones para controlar el tamaño de los gráficos de resultado:

- **El tamaño varía de manera continua.** Seleccione esta opción para mostrar un rango de tamaños de enlace que refleje la variación de las fuerzas de conexión en función de los valores de datos reales.
- **El tamaño se muestra fuerte/normal/débil.** Seleccione esta opción para mostrar tres fuerzas de conexión (fuerte, normal y débil). Los puntos de corte para estas categorías pueden especificarse con anterioridad y en el gráfico final.

Enlaces fuertes por encima de. Permite especificar un umbral para las conexiones fuertes (líneas fuertes) y las conexiones habituales (líneas normales). Todas las conexiones que estén por encima de este valor se consideran fuertes. Utilice el control deslizante para ajustar el valor o introducir un número en el campo.

Enlaces débiles por debajo de. Permite especificar un número para indicar el umbral para las conexiones débiles (líneas punteadas) y las conexiones habituales (líneas normales). Todas las conexiones que estén por debajo de este valor se consideran débiles. Utilice el control deslizante para ajustar el valor o introducir un número en el campo.

Una vez ajustados los umbrales de la malla, puede volver a planear o a dibujar la visualización de la malla con los valores de umbral nuevos a través del menú de malla ubicado en la barra de herramientas de gráfico de malla. Una vez encontrada la configuración que muestra los patrones más significativos,

puede actualizar la configuración original en el nodo Malla (también denominado nodo de malla padre) seleccionando **Actualizar nodo padre** en el menú Malla de la ventana del gráfico.

Creación de un resumen de malla

Puede crear un documento de resumen de malla que enumere enlaces fuertes, medios y débiles pulsando en el botón de flecha doble amarilla de la barra de herramientas para expandir la ventana del gráfico de malla. A continuación, pulse en la pestaña **Resumen** para ver las tablas de cada tipo de enlace. Las tablas pueden expandirse y contraerse mediante los botones de conmutación de cada una.

Para imprimir el resumen, seleccione lo siguiente desde el menú en la ventana del gráfico Web:

Archivo > Imprimir resumen

Nodo Evaluación

El nodo Evaluación ofrece una forma sencilla de evaluar y comparar modelos predictivos para elegir el mejor modelo para su aplicación. Los diagramas de evaluación muestran el comportamiento de los modelos prediciendo determinados resultados. Funcionan ordenando los registros en función del valor predicho y confianza de la predicción, dividiendo los registros en grupos de igual tamaño (**cuantiles**) y, a continuación, dibujando el valor del criterio de negocio de cada cuantil, del más alto al más bajo. El gráfico muestra múltiples modelos como líneas independientes.

Los resultados se gestionan definiendo un valor o rango de valores específicos como un **acierto**. Los aciertos suelen indicar algún tipo de éxito (como una venta a un cliente) o un evento de interés (como un diagnóstico médico específico). Puede definir criterios de aciertos en la pestaña Opciones del cuadro de diálogo o utilizar los criterios de aciertos predeterminados como se indica a continuación:

- Los campos de salida de **marcas** son directos; los aciertos corresponden a valores *verdaderos*.
- Para los campos de resultado **nominales**, el primer valor del conjunto define un acierto.
- Para los campos de resultado **Continuos**, los aciertos equivalen a valores mayores que el punto medio del rango del campo.

Existen seis tipos de diagramas de evaluación, cada uno de ellos con el énfasis puesto en un criterio de evaluación diferente.

Ganancias

Las ganancias se definen como la proporción de aciertos totales que se produce en cada cuantil. Las ganancias se calculan como el resultado de: (número de aciertos en cuantil / número total de aciertos) x 100%.

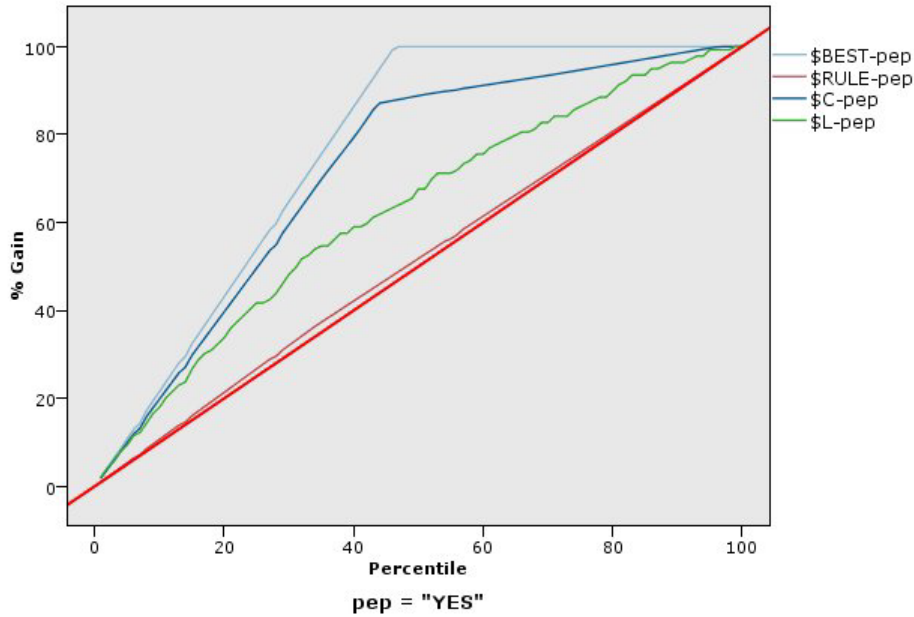


Figura 43. Gráfico de ganancias (acumulado) con línea base que incluye mejor línea y regla de negocio

Elevación

La elevación compara el porcentaje de registros de cada cuantil que supone aciertos con el porcentaje global de aciertos de los datos de entrenamiento. Se calcula como el resultado de: (aciertos del cuantil / registros del cuantil) / (aciertos totales / registros totales).

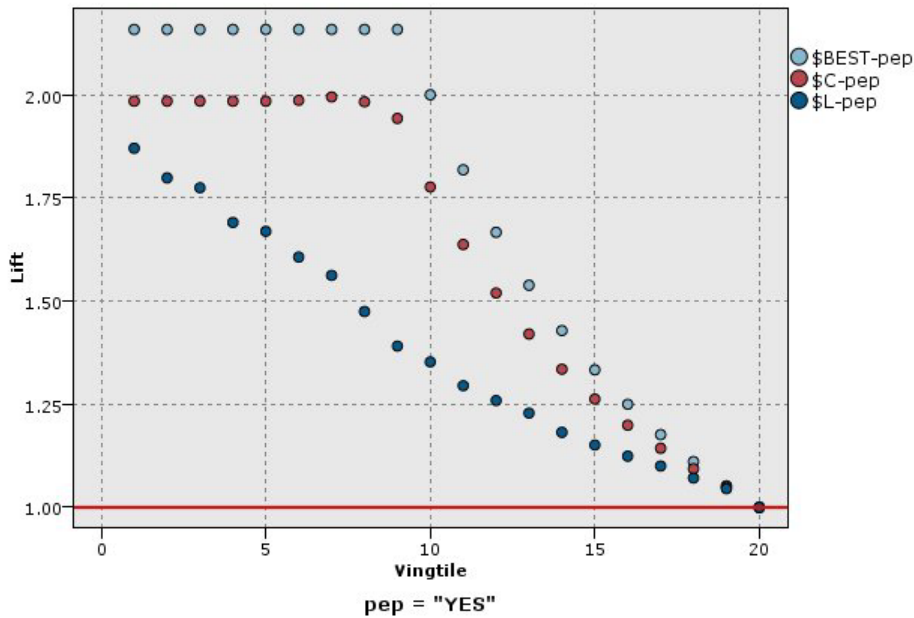


Figura 44. Gráfico de elevación (acumulado) mediante puntos y mejor línea.

Respuesta

La respuesta es sencillamente el porcentaje de registros del cuantil que son aciertos. La respuesta se calcula como el resultado de: (aciertos del cuantil / registros del cuantil) X 100%.

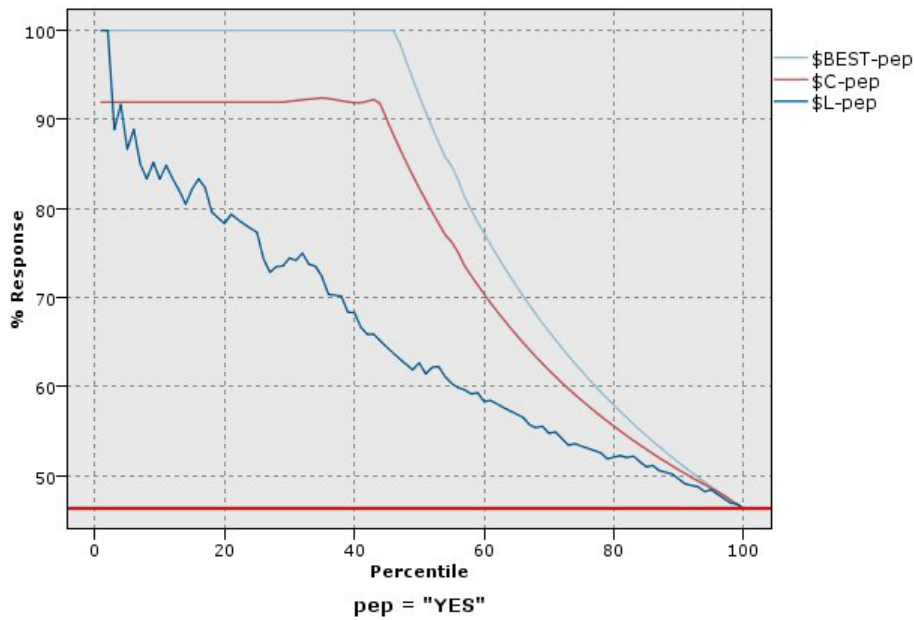


Figura 45. Gráfico de respuestas (acumulado) con mejor línea

Beneficio

El beneficio es igual a los **ingresos** de cada registro menos el **coste** del registro. Los beneficios de un cuantil son la suma de los beneficios de todos los registros del cuantil. Se asume que los ingresos se aplican sólo a los aciertos, pero los costes se aplican a todos los registros. Los beneficios y los costes se pueden fijar o estar definidos por campos en los datos. Los beneficios se calculan como el resultado de: (suma de los ingresos de los registros del cuantil – suma de los costes de los registros del cuantil).

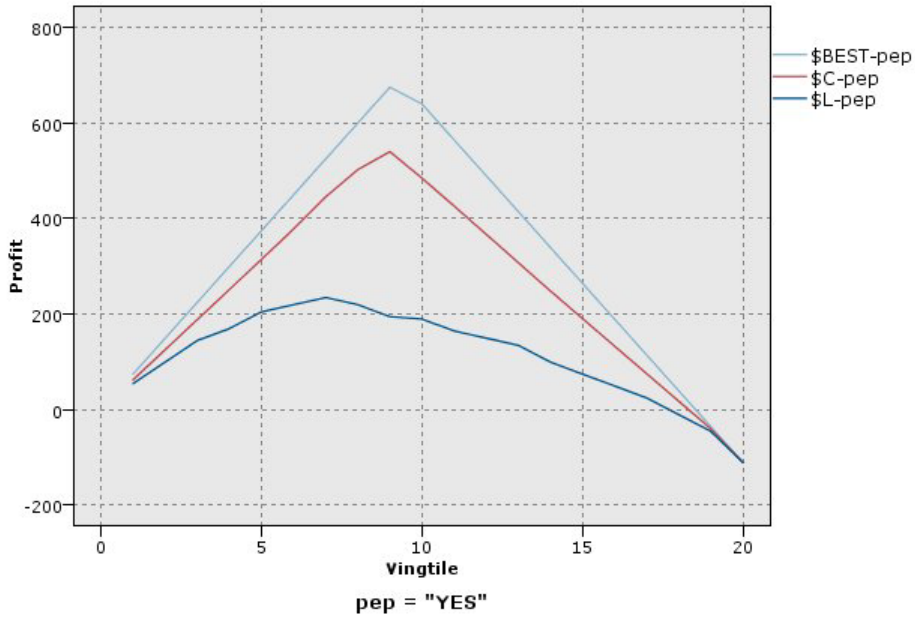


Figura 46. Gráfico de beneficios (acumulado) con mejor línea

Rentabilidad de la inversión

La rentabilidad de la inversión (ROI, del inglés 'Return On Investment') es similar al beneficio en cuanto a que implica la definición de ingresos y costes. La rentabilidad de la inversión compara los beneficios con los costes del cuantil. La rentabilidad de la inversión se calcula como el resultado de: (beneficios del cuantil/costes del cuantil) X 100%.

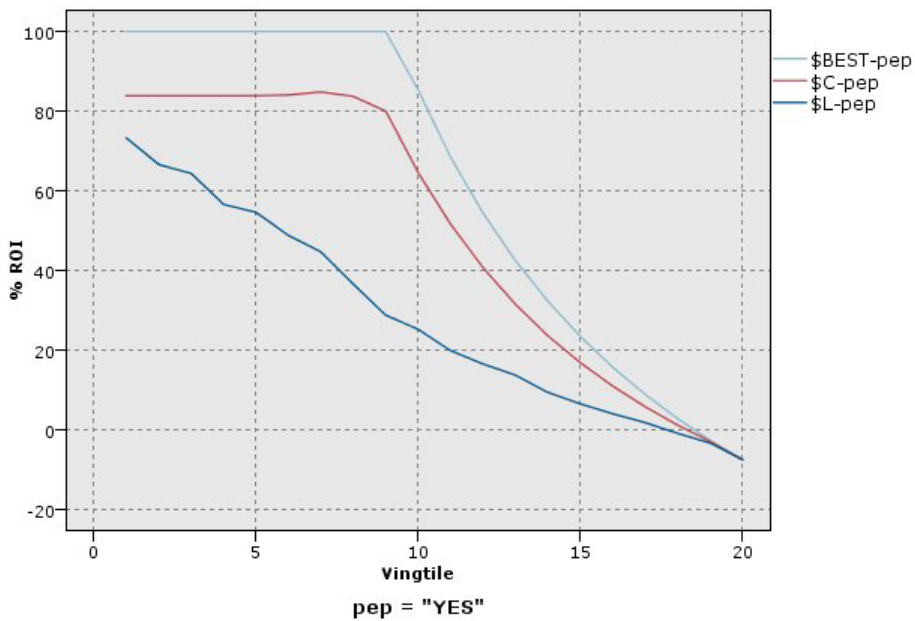


Figura 47. Gráfico de rentabilidad de la inversión (acumulado) con mejor línea

Gráficas ROC

La ROC (característica operador receptor) solo puede utilizarse con clasificadores binarios. La ROC puede utilizarse para visualizar, organizar y seleccionar clasificadores en función de su rendimiento. Una gráfica ROC representa la tasa de auténticos positivos (o sensibilidad) frente a la tasa de falsos positivos del clasificador. Una gráfica ROC representa el equilibrio relativo entre beneficios (auténticos positivos) y costes (falsos positivos). Un positivo auténtico es una instancia que es un acierto y está clasificada como tal. Por tanto, la tasa de positivos auténticos se calcula como el número de positivos auténticos / número de instancias que realmente son aciertos. Un falso positivo es una instancia que es un fallo y está clasificada como tal. Por tanto, la tasa de falsos positivos se calcula como el número de falsos positivos / número de instancias que realmente son fallos.

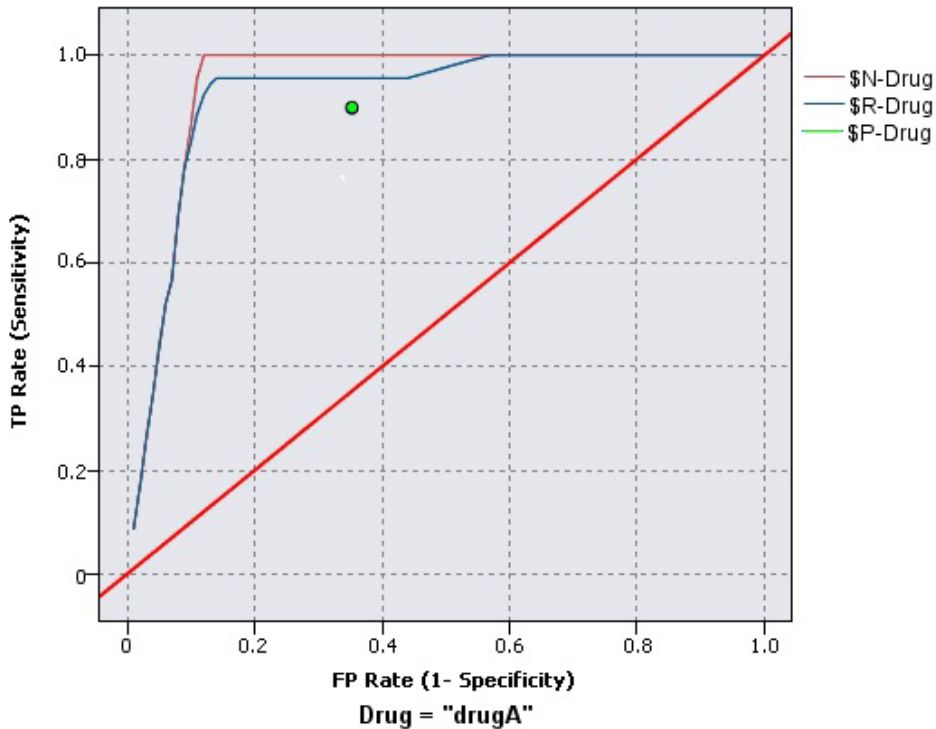


Figura 48. Gráfica ROC con la mejor línea

Los diagramas de evaluación también pueden ser acumulados, de forma que cada punto equivalga al valor del cuantil correspondiente más todos los cuantiles mayores. Los gráficos acumulados suelen mostrar mejor el rendimiento global de modelos, mientras que los gráficos no acumulados suelen ser mejores para indicar determinadas áreas de problemas para los modelos.

Pestaña Gráfico de Evaluación

Tipo de gráfico. Seleccione uno de los tipos siguientes: **Ganancias**, **Respuesta**, **Elevación**, **Beneficio**, **ROI** (retorno de la inversión) o **ROC** (característica operador receptor).

Gráfico acumulado. Seleccione esta opción para crear un gráfico acumulado. Los valores de los gráficos acumulados se dibujan para cada cuantil más todos los cuantiles mayores. (**Representación acumulada** no está disponible en las gráficas ROC).

Incluir línea base. Seleccione esta opción para incluir una línea base en el gráfico, indicando una distribución aleatoria perfecta de los aciertos donde la confianza se vuelve irrelevante. (**Incluir línea base** no está disponible en la gráfica ROI ni en la de beneficios).

Incluir la mejor línea. Seleccione esta opción para incluir una mejor línea en el gráfico, que indique una confianza perfecta (donde los aciertos = 100% de los casos). (**Incluir la línea mejor**) no está disponible en las gráficas ROC).

Utilizar criterios de beneficio en todos los tipos de gráfica. Esta opción se selecciona para utilizar criterios de beneficio (coste, ingresos y peso) al calcular las medidas de evaluación en lugar del recuento normal de aciertos. Para los modelos con determinados destinos numéricos, tales como un modelo que pronostica los ingresos obtenidos desde un cliente como respuesta a una oferta, el valor del campo de destino ofrece una mejor medida del rendimiento del modelo que el recuento de aciertos. La selección de esta opción habilita los campos **Costes**, **Ingresos** y **Peso** en las gráficas Ganancias, Respuesta y Elevación. Para utilizar el criterio de beneficios de estos tres tipos de diagramas, se le recomienda que establezca **Ingresos** como campo de destino, **Coste** en 0.0, de modo que el beneficio sea igual a los ingresos, y que especifique una condición de acierto definido por usuario cuyo valor sea "true", de modo que todos los registros se recuenten como aciertos. (**Utilizar criterios de beneficio en todos los tipos de gráfica** no está disponible en las gráficas ROC).

Buscar campos predichos/predictores utilizando. Seleccione **Metadatos de campos de salida del modelo** para buscar los campos predichos del gráfico mediante metadatos, o seleccione **Formato del nombre del campo** para buscarlos por nombre.

Campos Puntuación de Gráfico. Seleccione esta casilla de verificación para activar el selector de campos de puntuación. A continuación, seleccione uno o más rangos o campos de puntuación continuos, es decir, campos que no sean modelos estrictamente predictivos que puedan ser útiles para clasificar registros en términos de propensión para ser un acierto. El nodo Evaluación puede comparar cualquier combinación de uno o más campos de puntuación o modelos más predictivos. Un ejemplo típico puede ser comparar varios campos RFM con su mejor modelo predictivo.

Objetivo. Seleccione el campo objetivo mediante el selector de campo. Seleccione cualquier campo de nominal o marca instanciado o defina el campo con dos o más valores.

Note: Este campo objetivo solamente se aplica a los campos de puntuación (modelos predictivos definidos por sus propios objetivos) y se ignora si se define un criterio de acierto personalizado en la pestaña Opciones.

Dividir por partición. Si utiliza un campo de partición para dividir los registros en muestreos de entrenamiento, comprobación y validación, seleccione esta opción para mostrar un diagrama de evaluación independiente para cada partición. Consulte el tema "Nodo Partición" en la página 165 para obtener más información.

Note: al dividir en particiones, los registros con valores nulos en el campo de partición se excluyen de la evaluación. Esto nunca debería ser un problema si se utiliza un nodo Partición ya que este tipo de nodos no genera valores nulos.

Gráfico. Permite seleccionar el tamaño de los cuantiles para dibujarlos en el gráfico desde la lista desplegable. Las opciones son **Cuartiles**, **Quintiles**, **Deciles**, **Veintiles**, **Percentiles** y **1000-tiles**. (**Representar** no está disponible en las gráficas ROC).

Estilo. Seleccione **Línea** o **Punto**.

En todos los tipos de gráfica salvo las gráficas ROC, los controles adicionales permiten especificar costes, ingresos y ponderaciones.

- **Costes.** Permite especificar el coste asociado con cada registro. Puede seleccionar costes **fijos** o **variables**. En el caso de los costes fijos, especifique el valor del coste. En el caso de los costes variables, pulse en el selector de campos para elegir un campo de costes. (**Costes** no está disponible en las gráficas ROC).

- **Ingresos.** Permite especificar los ingresos asociados con cada registro que representa un acierto. Puede seleccionar costes **fijos** o **variables**. En el caso de los ingresos fijos, especifique el valor del ingreso. En el caso de los ingresos variables, pulse en el selector de campos para elegir un campo de ingresos. (**Ingresos** no está disponible en las gráficas ROC).
- **Ponderación.** Si los registros de los datos representan más de una unidad, puede utilizar ponderaciones de frecuencias para ajustar los resultados. Especifique la ponderación asociada con cada registro mediante valores **fijos** o **variables**. En el caso de las ponderaciones fijas, especifique el valor de ponderación (el número de unidades por registro). En el caso de ponderaciones variables, pulse en el selector de campos para elegir un campo de ponderaciones. (**Peso** no está disponible en las gráficas ROC).

Pestaña Opciones de Evaluación

La pestaña Opciones de los diagramas de evaluación proporciona flexibilidad al definir aciertos, criterios de puntuación y reglas de negocio mostradas en el gráfico. También puede establecer opciones para exportar los resultados de la evaluación del modelo.

Acierto definido por el usuario. Seleccione esta opción para especificar una condición personalizada con objeto de indicar un acierto. Esta opción resulta útil para definir el resultado de interés en lugar de deducirlo del tipo de campo objetivo y el orden de valores.

- **Condición.** Cuando se selecciona **Acierto definido por el usuario**, se debe especificar una expresión CLEM para una condición de acierto. Por ejemplo, @TARGET = "SÍ" es una condición válida que indica que un valor *Sí* para el campo objetivo se tomará como un acierto en la evaluación. La condición especificada se utilizará para todos los campos objetivo. Para crear una condición, escriba en el campo o utilice el generador de expresiones para generar una expresión de condición. Si los datos están instanciados, puede insertar valores directamente desde el generador de expresiones.

Puntuación definida por el usuario. Seleccione esta opción para especificar una condición utilizada para puntuar casos antes de asignarlos a los cuantiles. La puntuación predeterminada se calcula a partir del valor predicho y la confianza. Utilice el campo Expresión para crear una expresión de puntuación personalizada.

- **Expresión.** Permite especificar una expresión CLEM utilizada para la puntuación. Por ejemplo, si un resultado numérico en el rango 0–1 se ordena de forma que los valores más bajos sean mejores que los más altos, puede definir un acierto como @TARGET < 0,5 y la puntuación asociada como 1 - @PREDICTED. La expresión de puntuación debe producir un valor numérico. Para crear una condición, escriba en el campo o utilice el generador de expresiones para generar una expresión de condición.

Incluir regla de negocio. Seleccione esta opción para especificar una condición de regla que refleje los criterios de interés. Por ejemplo, tal vez le interese que se muestre una regla para todos los casos en que hipoteca = "Y" e ingresos >= 33000. Las reglas de negocio se dibujan en el gráfico y se etiquetan en la clave como *Regla*. (**Incluir regla de negocio** no se soporta en las gráficas ROC).

- **Condición.** Permite especificar una expresión CLEM utilizada para definir una regla de negocio en el gráfico de resultados. Simplemente escriba en el campo o utilice el generador de expresiones para generar una expresión de condición. Si los datos están instanciados, puede insertar valores directamente desde el generador de expresiones.

Exportar resultados en archivo. Seleccione esta opción para exportar los resultados de la evaluación del modelo a un archivo de texto delimitado. Puede leer este archivo para realizar análisis especializados en los valores calculados. Establezca las siguientes opciones de exportación:

- **Nombre de archivo.** Permite introducir el nombre del archivo de resultados. Utilice el botón de puntos suspensivos (...) para seleccionar la carpeta deseada.
- **Delimitador.** Permite introducir un carácter, como una coma o un espacio, que utilizar como delimitador de campos.

Incluir nombres de campo. Seleccione esta opción para incluir nombres de campos como la primera línea del archivo de resultados.

Nueva línea después de cada registro. Seleccione esta opción para comenzar cada registro en una nueva línea.

Pestaña Aspecto de Evaluación

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Texto. Permite aceptar la etiqueta generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de X. Permite aceptar la etiqueta de eje x (horizontal) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Etiqueta de Y. Permite aceptar la etiqueta del eje y (vertical) generada automáticamente o seleccionar **Personalizado** para especificar una etiqueta.

Mostrar línea de cuadrícula. Seleccionada de forma predeterminada, esta opción muestra una línea de cuadrícula detrás del gráfico o el diagrama que permite determinar con mayor facilidad los puntos de corte de la región y el tramo. Las líneas de cuadrícula siempre se muestran en blanco a menos que el fondo del gráfico sea blanco; en cuyo caso se muestran en gris.

Lectura de los resultados de una evaluación del modelo

La interpretación de un diagrama de evaluación depende hasta cierto punto del tipo de gráfico, pero hay algunas características comunes a todos los diagramas de evaluación. En el caso de los gráficos acumulados, las líneas más altas indican modelos mejores, especialmente en la parte izquierda del gráfico. En muchos casos, al comparar múltiples modelos las líneas se cruzan para que uno de ellos sea más alto en una parte del gráfico y otro sea más alto en otra parte distinta. En este caso, debe tener en cuenta la porción de la muestra que desee (que define un punto en el eje x) al decidir el modelo que va a elegir.

La mayoría de los gráficos no acumulados será muy similar. Para obtener buenos modelos, los gráficos no acumulados deben ser altos en la parte izquierda del gráfico y bajos en la parte derecha. (Si un gráfico no acumulado muestra un patrón de sierra, puede suavizarlo reduciendo el número de cuantiles que se van a dibujar y volviendo a ejecutarlo.) Los descensos de la parte izquierda del gráfico o los picos de la parte derecha pueden indicar áreas donde el modelo tiene una predicción pobre. Una línea plana que cruza el gráfico indica un modelo que no proporciona ninguna información.

Ganancias. Los gráficos de ganancias acumulados siempre comienzan en el 0% y terminan en el 100% de izquierda a derecha. Para obtener un buen modelo, el gráfico de ganancias debe aumentar considerablemente hacia el 100% y, a continuación, descender. Los modelos que no proporcionan información mostrarán una diagonal desde la esquina inferior izquierda a la superior derecha (que se muestra en el gráfico si se selecciona **Incluir línea base**).

Elevación. Los gráficos de elevación acumulados tienden a comenzar por encima de 1,0 y descender gradualmente hasta alcanzar 1,0 de izquierda a derecha. El borde derecho del gráfico representa el conjunto de datos completo, de modo que el cociente de aciertos en cantidades acumuladas con respecto a los aciertos de datos sea 1,0. Para obtener un buen modelo válido, la elevación debe comenzar muy por encima de 1,0 en la izquierda, permanecer en una meseta alta a medida que se mueve hacia la derecha y, a continuación, descender rápidamente hacia el 1,0 en la parte derecha del gráfico. En el caso de un

modelo que no proporciona información, la línea se mantendrá alrededor del 1,0 en todo el gráfico. (Si la opción **Incluir línea base** está seleccionada, aparece una línea horizontal en el 1,0 en el gráfico como referencia.)

Respuestas. Los gráficos de respuestas acumulados tienden a ser muy similares a los gráficos de elevación salvo por la escala. Los gráficos de respuestas suelen comenzar cerca del 100% y descender gradualmente hasta alcanzar la tasa de respuesta global (total de aciertos / total de registros) en el borde derecho del gráfico. Para obtener un buen modelo, la línea debe comenzar cerca de 100% en la izquierda, permanecer en una meseta alta a medida que se mueve a la derecha y, a continuación, descender rápidamente hacia la tasa de respuesta global en la parte derecha del gráfico. En el caso de un modelo que no proporciona información, la línea se mantendrá alrededor de la tasa de respuesta global en todo el gráfico. (Si la opción **Incluir línea base** está seleccionada, aparecerá una línea horizontal en la tasa de respuesta global en el gráfico como referencia.)

Beneficios. Los gráficos de beneficios acumulados muestran la suma de beneficios a medida que aumenta el tamaño de la muestra seleccionada, de izquierda a derecha. Los gráficos de beneficios suelen comenzar cerca del 0, aumentar rápidamente hacia la derecha hasta que alcanzan un pico o meseta en el centro y, a continuación, descender hacia el borde derecho del gráfico. Para obtener un modelo válido, los beneficios deben mostrar un pico bien definido en algún punto del centro del gráfico. En el caso de un modelo que no proporciona ninguna información, la línea será bastante recta y puede ascender, descender o permanecer en el mismo nivel en función de la estructura de costes/ingresos que se aplique.

Rentabilidad de la inversión. Los gráficos de rentabilidad de la inversión tienden a ser similares a los gráficos de respuestas y a los gráficos de elevación excepto por la escala. Los gráficos de rentabilidad de la inversión suelen comenzar por encima del 0% y descender gradualmente hasta alcanzar la rentabilidad de la inversión global de todo el conjunto de datos (que puede ser un valor negativo). Para obtener un buen modelo, la línea debe comenzar muy por encima de 0%, permanecer en una meseta alta a medida que se mueve a la derecha y, a continuación, descender rápidamente hacia la rentabilidad de la inversión global en la parte derecha del gráfico. En el caso de un modelo que no proporciona información, la línea debe mantenerse alrededor del valor de rentabilidad de la inversión global.

Gráficas ROC. Las curvas ROC suelen tener la forma de una gráfica de ganancias acumuladas. La curva empieza en la coordenada (0,0) y termina en la coordenada (1,1) según avanza de izquierda a derecha. Una gráfica que sube bruscamente hacia la coordenada (0,1) y luego se estabiliza es indicadora de un buen clasificador. Un modelo que clasifica instancias aleatoriamente como aciertos o fallos seguirá la diagonal que va desde el extremo izquierdo inferior al derecho superior (mostrado en la gráfica si **Incluir línea base** está seleccionada). Si no se proporciona ningún campo de confianza para un modelo, dicho modelo se trazará como un punto único. El clasificador con el umbral óptimo de clasificación se encuentra más cercano a la coordenada (0,1) o esquina superior izquierda de la gráfica. Esta ubicación representa un gran número de instancias que están correctamente clasificadas como aciertos, y un número bajo de instancias que están incorrectamente clasificadas como aciertos. Los puntos por encima de la línea diagonal representan buenos resultados de clasificación. Los puntos por debajo de la línea diagonal representan malos resultados de clasificación, peores que si las instancias se hubieran clasificado aleatoriamente.

Uso de un diagrama de evaluación

El uso del ratón para explorar un diagrama de evaluación es similar al de un histograma o un gráfico de colección. El eje x representa puntuaciones de modelos en las cantidades especificadas, como los veintiles o los deciles.

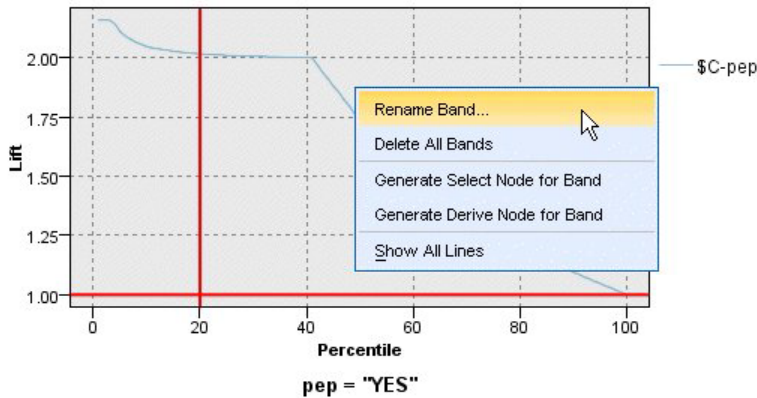


Figura 49. Cómo trabajar con un diagrama de evaluación

Puede dividir el eje x en tramos como lo haría con un histograma usando el icono de división, para mostrar opciones de división automática del eje en tramos iguales. Consulte el tema “Exploración de gráficos” en la página 269 para obtener más información. Puede editar los límites de los tramos manualmente seleccionando **Tramos del gráfico** en el menú Edición.

Una vez creado un diagrama de evaluación, definidos los tramos y examinados los resultados, puede utilizar las opciones del menú Generar y el menú contextual para crear nodos automáticamente en función de las selecciones del gráfico. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

Al generar nodos desde un diagrama de evaluación, se le solicitará que seleccione un modelo simple de entre todos los modelos disponibles en el gráfico.

Seleccione un modelo y pulse en **Aceptar** para generar el nuevo nodo en el lienzo de rutas.

Nodo Visualización de mapas

El nodo Visualización de mapas puede aceptar varias conexiones de entrada y visualizar datos geoespaciales en un mapa como una serie de capas. Cada capa es un único campo geoespacial; por ejemplo, la capa base podría ser el mapa de un país, y por encima puede tener una capa de carreteras, una capa de ríos y una capa de ciudades.

Aunque la mayoría de los conjuntos de datos geoespaciales normalmente contienen un solo campo geoespacial, cuando hay varios campos geoespaciales en una sola entrada, puede elegir qué campos visualizar. Dos campos de la misma conexión de entrada no se pueden visualizar al mismo tiempo; sin embargo, puede copiar y pegar la conexión entrante y mostrar un campo diferente de cada uno.

Pestaña de gráfico de Visualización de mapas

Capas

Esta tabla muestra información sobre las entradas al nodo de mapas. El orden de las capas dicta el orden en el que se muestran las capas en la presentación preliminar del mapa y la salida visual cuando se ejecuta el nodo. La fila superior de la tabla es la capa ‘superior’ y la fila inferior es la capa ‘base’; en otras palabras, cada capa se muestra en el mapa en frente de la capa directamente debajo de ella en la tabla.

Nota: Cuando una capa en la tabla contiene un campo geoespacial tridimensional, sólo se trazan los ejes x e y . El eje z se ignora.

Nombre

Los nombres se crean automáticamente para cada capa y se componen utilizando el formato siguiente: código[nodo de origen:nodo conectado]. De forma predeterminada, la etiqueta se muestra como un número, el 1 para la primera entrada que está conectada, el 2 para la segunda entrada, etcétera. Si es necesario, pulse el botón **Editar capa** para cambiar la etiqueta en el cuadro de diálogo Cambiar opciones de capa de mapa. Por ejemplo, puede cambiar la etiqueta a "carreteras" o "ciudades" para reflejar la entrada de datos.

Tipo Muestra el icono de tipo de medición del campo geoespacial seleccionado como la capa. Si los datos de entrada contienen varios campos con un tipo de medición geoespacial, la selección predeterminada utiliza el orden de clasificación siguiente:

1. Punto
2. Cadena lineal
3. Polígono
4. Multipunto
5. Cadena multilínea
6. Multipolígono

Nota: Si hay dos campos con el mismo tipo de medición, de forma predeterminada se selecciona el primer campo (alfabéticamente por nombre).

Símbolo

Nota: Esta columna sólo se completa para los campos Punto y Multipunto.

Muestra el símbolo que se utiliza para los campos Punto o Multipunto. Si es necesario, pulse el botón **Editar capa** para cambiar el símbolo en el cuadro de diálogo Cambiar opciones de capa de mapa.

Color Muestra el color que se ha seleccionado para representar la capa en el mapa. Si es necesario, pulse el botón **Editar capa** para cambiar el color en el cuadro de diálogo Cambiar opciones de capa de mapa. El color se aplica a elementos diferentes según el tipo de medición.

- Para Puntos o Multipuntos, el color se aplica al símbolo de la capa.
- Para Cadenas lineales y Polígonos, el color se aplica a la forma completa. Los Polígonos siempre tienen un contorno de color negro; el color que se muestra en la columna es el color que se utiliza para rellenar la forma.

Vista previa

Este panel muestra una vista previa de la selección actual de entradas en la tabla **Capas**. La vista previa tiene en cuenta el orden de los valores de capa, símbolo, color y cualquier otro valor de visualización asociado a las capas y, cuando es posible, actualiza la visualización cuando cambian los valores. Si cambia detalles en algún otro lugar de la ruta, como los campos geoespaciales a utilizar como capas, o si modifica detalles como las funciones de agregación asociadas, quizá deba pulsar el botón **Actualizar datos** para actualizar la vista previa.

Utilice la **Vista previa** para establecer los valores de visualización antes de ejecutar la ruta. Para protegerse contra los retardos de tiempo que podría causar la utilización de un conjunto de datos de gran tamaño, la vista previa toma una muestra de cada capa y crea una visualización de los primeros 100 registros.

Cambiar las capas de mapa

Puede utilizar el diálogo Cambiar opciones de capa de mapa para modificar varios detalles de cualquier capa que se muestra en la pestaña **Gráfico** del nodo Visualización de mapas.

Detalles de entrada

Código

De forma predeterminada, la etiqueta es un número; puede sustituir este número por una etiqueta más significativa para facilitarle la identificación de la capa en el mapa. Por ejemplo, la etiqueta puede ser el nombre de la entrada de datos, por ejemplo, "Ciudades".

Campo Capa

Si tiene más de un campo geoespacial en los datos de entrada, utilice esta opción para seleccionar el campo que desea visualizar como una capa en el mapa.

De forma predeterminada, las capas entre las que puede elegir están en el orden de clasificación siguiente.

- Punto
- Cadena lineal
- Polígono
- Multipunto
- Cadena multilínea
- Multipolígono

Valores de visualización

Utilizar agrupación hexagonal

Nota: Esta opción sólo afecta a los campos punto y multipunto.

La agrupación hexagonal combina los puntos adyacentes (en función de sus coordenadas x e y) en un único punto para mostrar en el mapa. Este único punto se muestra como un hexágono, pero se representa efectivamente como un polígono.

Debido a que el hexágono se representa como un polígono, los campos punto con agrupación hexagonal activada se tratan como polígonos. Esto significa que si elige **Ordenar por tipo** en el cuadro de diálogo de nodo de mapa, las capas de punto que tienes aplicada la agrupación hexagonal se representan por encima de las capas de polígono pero por debajo de las capas de Cadena lineal y Punto.

Si utiliza la agrupación hexagonal para un campo multipunto, el campo primero se convierte en un campo punto agrupando los valores multipunto para calcular el punto central. Los puntos centrales se utilizan para calcular las agrupaciones hexagonales.

Agregación

Nota: Esta columna solo está disponible cuando se selecciona el recuadro **Utilizar agrupación hexagonal** y también se selecciona una **Superposición**.

Si selecciona un campo **Superposición** para una capa de puntos que utiliza la agrupación hexagonal, todos los valores de dicho campo deben agregarse para todos los puntos en el hexágono. Especifique una función de agregación para los campos de superposición que desee aplicar al mapa. Las funciones de agregación disponibles dependen del tipo de medición.

- Funciones de agregación para un tipo de medición Continua, con almacenamiento Real o Entero :
 - Sum
 - Media
 - Mín
 - Máx
 - Mediana
 - Primer cuartil

- Tercer cuartil
- Funciones de agregación para un tipo de medición Continua, con almacenamiento Hora, Fecha o Marca de hora:
 - Media
 - Mín
 - Máx
- Funciones de agregación para tipos de medición Nominal o Categórico:
 - Modo
 - Mín
 - Máx
- Funciones de agregación para un tipo de medición Marca:
 - True si alguno es True
 - False, si alguno es False

Color

Utilice esta opción para elegir un color estándar para aplicarlo a todas las características del campo geoespacial, o un campo de superposición, que pone color a las características en función de los valores de otro campo en los datos.

Si selecciona **Estándar**, puede elegir un color de la paleta de colores que se muestra en el panel **Orden de colores de categorías gráficas** en la pestaña de visualización del cuadro de diálogo Opciones de usuario.

Si selecciona **Superposición**, puede elegir cualquier campo del origen de datos que contiene el campo geoespacial que se ha seleccionado como el **campo Capa**.

- Para los campos de superposición nominales o categóricos, la paleta de colores de la que puede elegir es la misma que se muestra para las opciones de color **Estándar**.
- Para los campos de superposición continuo y ordinal se muestra una segunda lista desplegable, en la que puede seleccionar un color. Cuando selecciona un color, la superposición se aplica variando la saturación de ese color según los valores en el campo continuo u ordinal. El valor más alto utiliza el color seleccionado de la lista desplegable y los valores más bajos se muestran mediante saturaciones correspondientemente más bajas.

Símbolo

Nota: Sólo habilitado para tipos de medición punto y multipunto.

Utilice esta opción para elegir si desea tener un símbolo **Estándar**, que se aplica a todos los registros de su campo geoespacial, o un símbolo de **Superposición**, que cambia el icono de símbolo para los puntos según los valores de otro campo en los datos.

Si selecciona **Estándar**, puede elegir uno de los símbolos predeterminados de una lista desplegable para representar los datos de punto en el mapa.

Si selecciona **Superposición**, puede elegir cualquier campo nominal, ordinal o categórico del origen de datos que contiene el campo geoespacial que se ha seleccionado como el **campo Capa**. Para cada valor en el campo de superposición, se muestra un símbolo diferente muestra en el mapa.

Por ejemplo, los datos pueden contener un campo de punto que representa la ubicación de las tiendas y la superposición podría ser un campo de tipo de tienda. En este ejemplo, todas las tiendas de alimentos podrían identificarse en el mapa mediante el símbolo de una cruz y todas las tiendas de electrónica mediante el símbolo de un cuadrado.

Tamaño

Nota: Sólo habilitado para tipos de medición punto, multipunto, cadena lineal y cadena multilínea.

Utilice esta opción para elegir si desea tener un tamaño **Estándar**, que se aplica a todos los registros de su campo geoespacial, o un tamaño de **Superposición**, que cambia el tamaño del icono de símbolo o el grosor de línea según los valores de otro campo en los datos.

Si selecciona **Estándar**, puede elegir un valor de ancho de píxel. Las opciones disponibles son 1, 2, 3, 4, 5, 10, 20 o 30.

Si selecciona **Superposición**, puede elegir cualquier campo del origen de datos que contiene el campo geoespacial que se ha seleccionado como el **campo Capa**. El grosor de la línea o el punto varía en función del valor del campo seleccionado.

Transparencia

Utilice esta opción para elegir si desea tener una transparencia **Estándar**, que se aplica a todos los registros de su campo geoespacial, o una transparencia de **Superposición**, que cambia la transparencia del símbolo, de la línea o del polígono según los valores de otro campo en los datos.

Si selecciona **Estándar**, puede elegir a partir de una selección de niveles de transparencia que empieza en 0% (opaco) y aumenta en incrementos de 10% hasta el 100% (transparente).

Si selecciona **Superposición**, puede elegir cualquier campo del origen de datos que contiene el campo geoespacial que se ha seleccionado como el **campo Capa**. En el mapa se muestra un nivel de transparencia diferente para cada valor en el campo de superposición. La transparencia se aplica al color elegido en la lista desplegable de color para el punto, la línea o el polígono.

Etiqueta de datos

Nota: Esta opción no está disponible si selecciona la casilla de verificación **Utilizar agrupación hexagonal**.

Utilice esta opción para seleccionar un campo para utilizarlo como etiquetas de datos en el mapa. Por ejemplo, si se aplica a una capa de polígono, la etiqueta de datos puede ser el campo de nombre, que contiene el nombre de cada polígono. Si selecciona el campo de nombre, esos nombres se muestran en el mapa.

Pestaña Aspecto de visualización de correlación

Puede especificar las opciones de aspecto antes de crear gráficos.

Título. Especifique el texto que debe utilizarse para el título del gráfico.

Subtítulo. Especifique el texto que debe utilizarse para el subtítulo del gráfico.

Pie. Especifique el texto que debe utilizarse para el pie del gráfico.

Exploración de gráficos

Mientras que el modo de edición le permite editar el diseño y apariencia del gráfico, Explorar modo permite explorar analíticamente los datos y valores que se representan en el gráfico. El objetivo principal de la exploración es analizar los datos e identificar los valores utilizando tramos, regiones y marcaciones para generar nodos Seleccionar, Derivar o Equilibrar. Para seleccionar este modo, seleccione **Ver > Explorar modo** de los menús (o pulse en el icono de la barra de herramientas).

Aunque algunos gráficos pueden utilizar todas las herramientas de exploración, otros sólo aceptan uno. Explorar modo incluye:

- Definición y edición de tramos, que se utilizan para dividir los valores en una escala del eje x . Consulte el tema "Uso de tramos" en la página 270 para obtener más información.

- Definición y edición de regiones, que se utilizan para identificar un grupo de valores en el área rectangular. Consulte el tema “Usar regiones” en la página 273 para obtener más información.
- Marcación y desmarcación de elementos para controlar la selección de los valores que se pueden utilizar para generar un nodo Seleccionar o Derivar. Consulte el tema “Uso de elementos con marca” en la página 275 para obtener más información.
- Generación de nodos utilizando los valores identificados por tramos, regiones, elementos con marca y enlaces Web que se utilizarán en la ruta. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

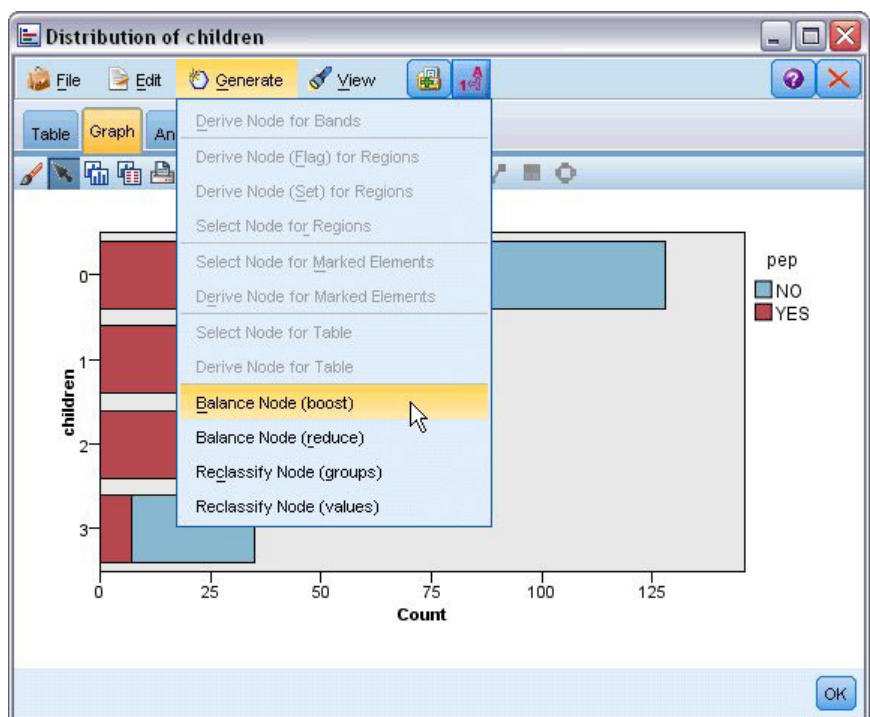


Figura 50. Gráfico con el menú Generar

Uso de tramos

En un gráfico con un campo de escala en el eje x , puede dibujar líneas de tramos verticales para dividir el rango de los valores en el eje x . Si un gráfico tiene múltiples paneles, una línea de tramo dibujada en un panel también se representa en el resto de paneles.

No todos los gráficos admiten tramos. Algunos de estos gráficos que pueden contener tramos incluyen: histogramas, gráficos de barras y distribuciones, gráficos (línea, dispersión, hora, etc.), colecciones y diagramas de evaluación. En gráficos con paneles, los tramos aparecen en todos los paneles. Y en algunos casos, en una SPLOM, verá una línea de tramo horizontal ya que el eje en el que el tramo de campo/variable se ha dibujado se ha desplazado.

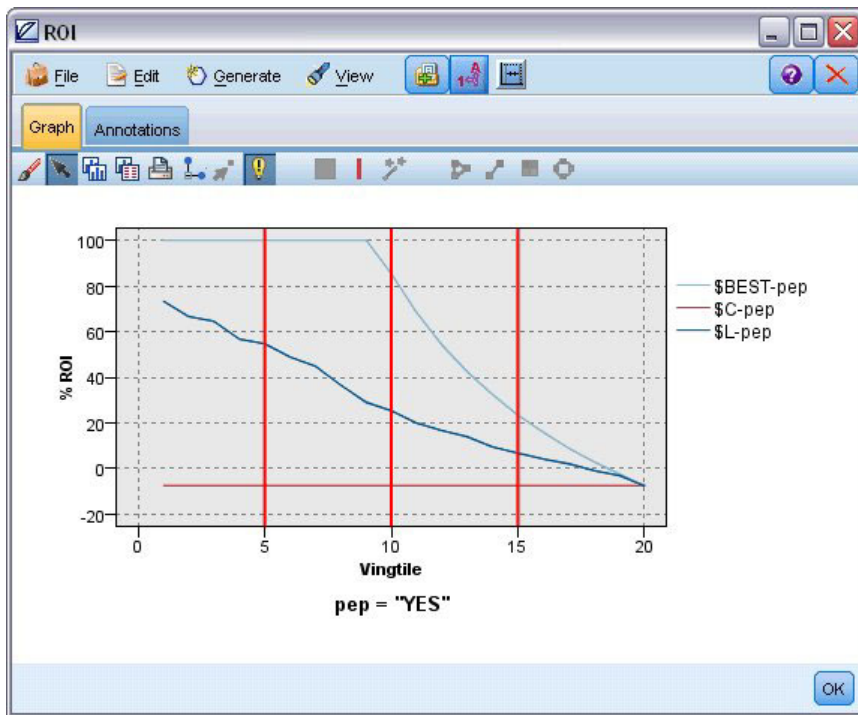


Figura 51. Gráfico con tres tramos

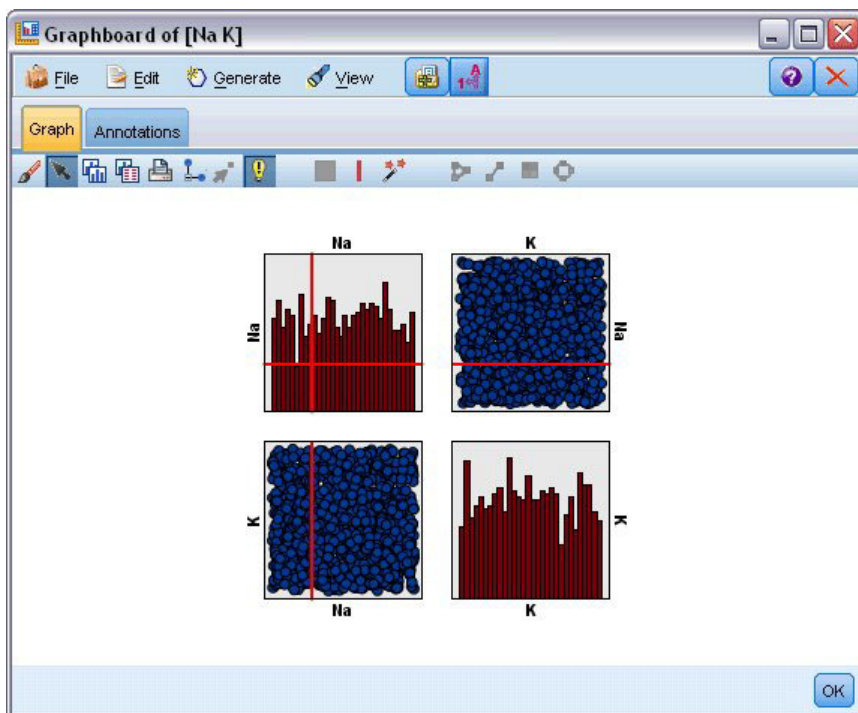


Figura 52. SPLOM con tramos

Definición de tramos

En un gráfico sin tramos, si añade una línea de tramo dividirá el gráfico en dos tramos. La línea de tramo representa el punto de inicio, que también se denomina límite inferior, del segundo tramo cuando

lea el gráfico de izquierda a derecha. Del mismo modo, en un gráfico con dos tramos, si añade una línea de tramo dividirá uno de los tramos en dos, con lo que tendrá tres tramos. De forma predeterminada, los tramos se denominan *tramoN*, donde *N* es el número de tramos de izquierda a derecha del eje *x*.

Una vez haya definido un tramo, puede arrastrar y soltar el tramo para volverlo a colocar en el eje *x*. Podrá ver más accesos directos si pulsa con el botón derecho en el tramo en tareas como cambiar el nombre, eliminar o generar nodos para ese tramo concreto.

Para definir tramos:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Dibujar tramo.



Figura 53. Botón de la barra de herramientas Dibujar tramos

3. En un gráfico que acepte tramos, pulse en el punto de valor del eje *x* en el que desea definir una línea de tramos.

Note: También puede pulsar en el icono de la barra de herramientas **Dividir el gráfico en tramos** e introducir el número de tramos iguales que desea y pulsar en **Dividir**.



Figura 54. Icono de división utilizado para expandir la barra de herramientas con opciones para dividir en tramos

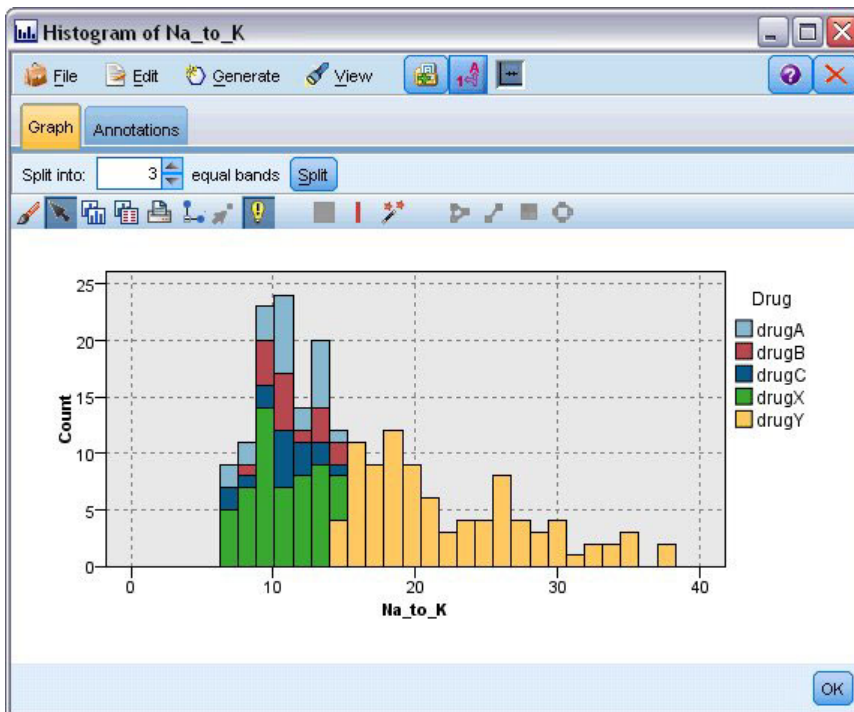


Figura 55. Creación de la barra de herramientas de tramos iguales con tramos activados

Edición, cambio de nombre y eliminación de tramos

Puede modificar las propiedades de tramos existentes en el cuadro de diálogo Modificar tramos del gráfico o mediante los menús contextuales del propio gráfico.

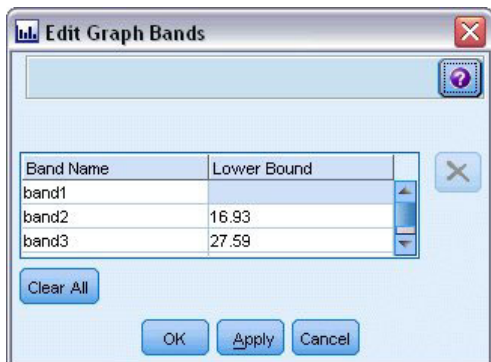


Figura 56. Cuadro de diálogo Modificar tramos del gráfico

Para modificar tramos:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Dibujar tramo.
3. En los menús, seleccione **Edición > Tramos del gráfico**. Se abrirá el cuadro de diálogo Modificar tramos del gráfico.
4. Si su gráfico contiene múltiples campos (como gráficos SPLOM), puede seleccionar el campo que desee en la lista desplegable.
5. Añada un tramo nuevo introduciendo un nombre y límite inferior. Pulse la tecla Intro para iniciar una fila nueva.
6. Edite el límite de un tramo ajustando el valor de **Límite inferior**.
7. Cambie el nombre de un tramo introduciendo un nombre de tramo nuevo.
8. Elimine un tramo seleccionando la línea en la tabla y pulsando en el botón de eliminación.
9. Pulse en **Aceptar** para aplicar los cambios y cerrar el cuadro de diálogo.

Note: También puede eliminar y cambiar el nombre de los tramos directamente en el gráfico pulsando con el botón derecho en la línea del tramo y seleccionando la opción de los menús contextuales.

Usar regiones

En cualquier gráfico con dos ejes de escala (o rango), puede dibujar regiones para agrupar valores en un área rectangular, denominada región. Una **región** es un área del gráfico que se describe por sus valores X e Y mínimo y máximo. Si un gráfico tiene múltiples paneles, una región dibujada en un panel también se representa en el resto de paneles.

No todos los gráficos admiten regiones. Algunos de los gráficos que aceptan regiones son: gráficos (línea, dispersión, burbujas, tiempo, etc.), SPLOM y colecciones. Estas regiones se dibujan en espacios de X,Y y no se pueden definir en gráficos 1-D, 3-D o animados. En gráficos con paneles, las regiones aparecen en todos los paneles. Con una matriz de diagrama de dispersión (SPLOM), aparecerá una región correspondiente en los gráficos superiores correspondientes pero no en los gráficos diagonales, ya que sólo muestran un campo de escala.

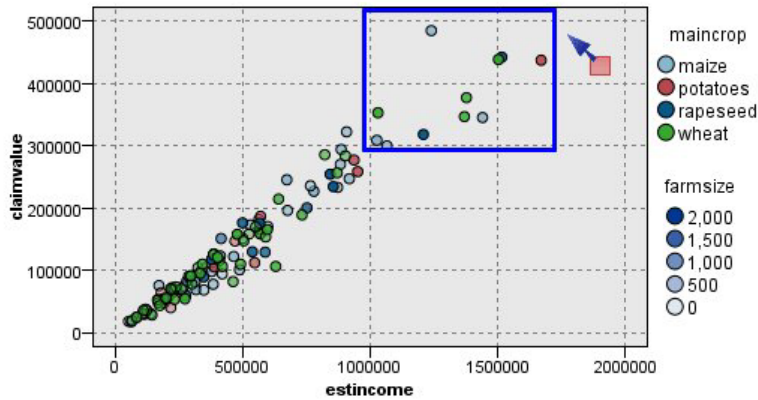


Figura 57. Definición de una región de valores de reclamación altos

Definición de regiones

Cuando crea una región, crea un grupo de valores. De forma predeterminada, todas las regiones nuevas se denominan *Región<N>*, donde *N* corresponde al número de regiones ya creadas.

Una vez haya definido una región, pulse con el botón derecho en la línea de la región para acceder a algunos accesos directos básicos. Sin embargo, podrá ver más accesos directos si pulsa con el botón derecho en la región (no en la línea) en tareas como cambiar el nombre, eliminar o generar nodos Seleccionar y Derivar para esa región concreta.

Puede seleccionar subconjuntos de registros en función de su inclusión en una región determinada o en una de varias regiones. También puede incorporar información de región a un registro generando un nodo Derivar para marcar registros en función de su inclusión en una región. Consulte el tema “Generación de nodos a partir de gráficos” en la página 276 para obtener más información.

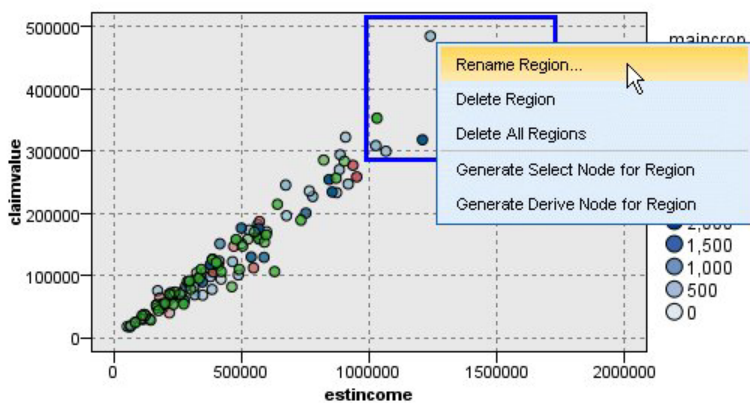


Figura 58. Exploración de la región de valores de reclamación altos

Para definir regiones:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Dibujar región.



Figura 59. Botón de la barra de herramientas Dibujar región

3. En un gráfico que acepte regiones, pulse el ratón y arrástrelo para dibujar la región rectangular.

Edición, cambio de nombre y eliminación de regiones

Puede modificar las propiedades de regiones existentes en el cuadro de diálogo Modificar regiones del gráfico o mediante los menús contextuales del propio gráfico.

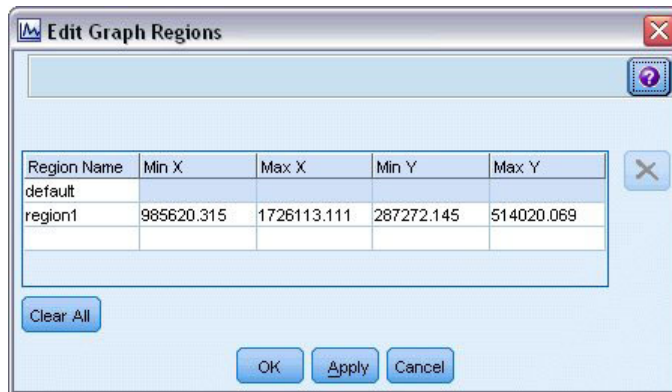


Figura 60. Especificación de propiedades de las regiones definidas

Para modificar regiones:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Dibujar región.
3. En los menús, seleccione **Edición > Regiones del gráfico**. Se abrirá el cuadro de diálogo Modificar regiones del gráfico.
4. Si tiene múltiples campos en el gráfico (por ejemplo, gráficos SPLOM), debe definir el campo de la región en las columnas *Campo A* y *Campo B*.
5. Añada una nueva región en una nueva línea introduciendo un nombre, seleccionando los campos de nombre (si procede) y definiendo los límites máximo y mínimo de cada campo. Pulse la tecla Intro para iniciar una fila nueva.
6. Modifique los límites de la región existente ajustando los valores **Min** y **Max** de *A* y *B*.
7. Cambie el nombre de una región cambiando el nombre de la región en la tabla.
8. Elimine una región seleccionando la línea en la tabla y pulsando en el botón de eliminación.
9. Pulse en **Aceptar** para aplicar los cambios y cerrar el cuadro de diálogo.

Note: También puede eliminar y cambiar el nombre de las regiones directamente en el gráfico pulsando con el botón derecho en la línea de la región y seleccionando la opción de los menús contextuales.

Uso de elementos con marca

Puede marcar elementos como barras, sectores y puntos en cualquier gráfico. Las líneas, áreas y superficies no se pueden marcar en gráficos, salvo en gráficos de tiempo, gráficos múltiples y diagramas de evaluación, ya que las líneas hacen referencia a los campos en esos casos. Siempre que marque un elemento, está resaltando todos los datos que representa ese elemento. En cualquier gráfico en que se represente el mismo caso en más de un lugar (como SPLOM), marcación es sinónimo de cepillado. Puede marcar elementos en gráficos y en tramos y regiones. Siempre que marque un elemento y vuelva a modo

de edición, la marcación seguirá siendo visible.

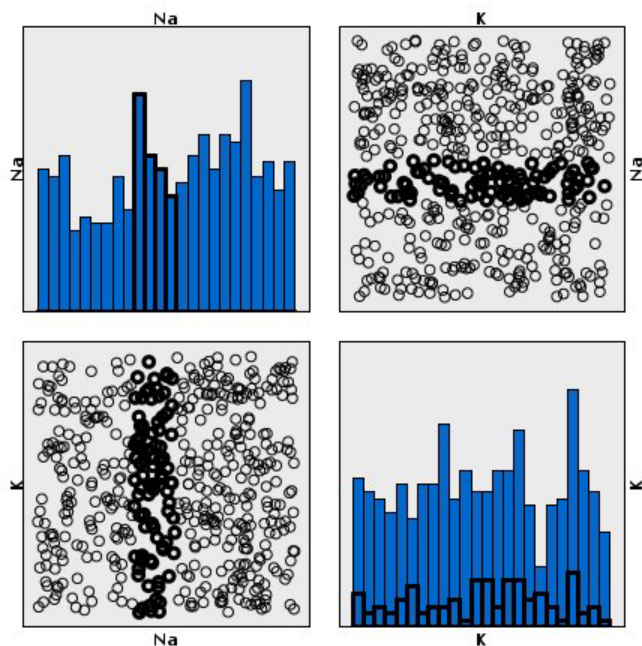


Figura 61. Marcación de elementos en una matriz SPLOM

Puede marcar y eliminar la marca pulsando en los elementos del gráfico. Cuando pulsa en un elemento para marcarlo por primera vez, el elemento aparecerá con un borde grueso para indicar que se ha marcado. Si vuelve a pulsar en el elemento, el borde desaparecerá y el elemento ya no estará marcado. Para marcar múltiples elementos, puede mantener pulsada la tecla Ctrl mientras pulsa los elementos o puede arrastrar el ratón por los elementos que desea marcar con la "varita mágica". Recuerde que si pulsa en otra área o elemento sin mantener pulsada la tecla Ctrl, se eliminará la marca de todos los elementos marcados anteriormente.

Puede generar los nodos Seleccionar y Derivar de los elementos marcados en su gráfico. Consulte el tema "Generación de nodos a partir de gráficos" para obtener más información.

Para marcar elementos:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Marcar elementos.
3. Pulse en el elemento que necesita o pulse y arrastre el ratón para dibujar una línea alrededor de la región con varios elementos.

Generación de nodos a partir de gráficos

Una de las características más útiles de los gráficos de IBM SPSS Modeler es la capacidad de generar nodos a partir de un gráfico o una selección en el gráfico. Por ejemplo, en un gráfico de tiempo, puede generar los nodos Derivar y Seleccionar basados en una selección o región de datos, estableciendo los datos en subconjuntos de forma eficaz. Por ejemplo, puede utilizar esta potente característica para identificar y excluir valores atípicos.

Cuando dibuje un tramo, también puede generar un nodo Derivar. En gráficos con dos ejes de escala, puede generar los nodos Seleccionar o Derivar de las regiones dibujadas en el gráfico. En gráficos con elementos marcados, puede generar nodos Derivar, Seleccionar y, en algunos casos, nodos Filtrar a partir

de estos elementos. Cualquier gráfico con una distribución de cuentas puede generar el nodo Equilibrar.

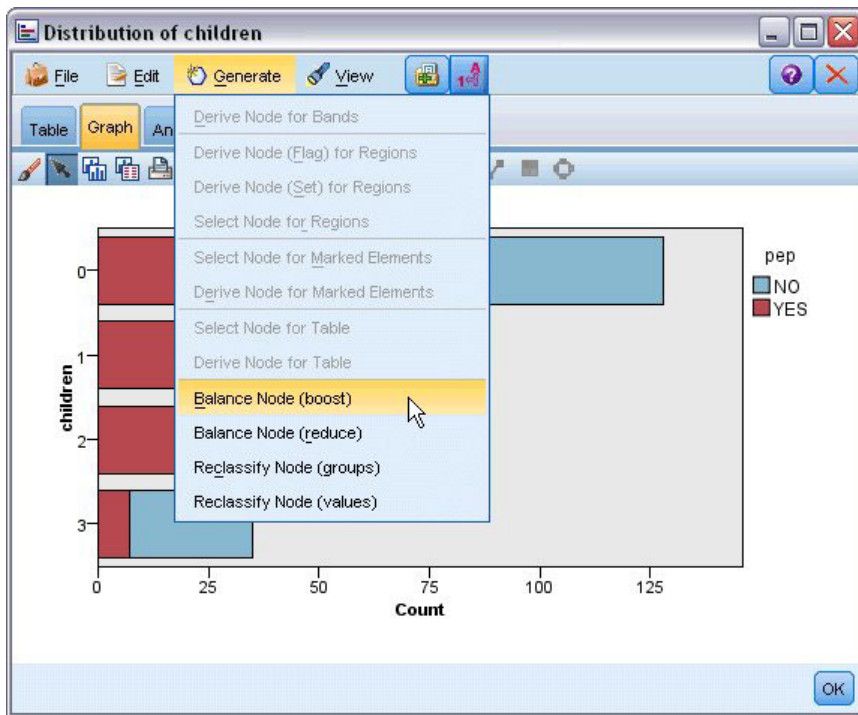


Figura 62. Gráfico con el menú Generar

Siempre que genere un nodo, se coloca directamente en el lienzo de rutas, por lo que puede conectarlo a una ruta existente. Los nodos siguientes pueden generarse a partir de gráficos: Seleccionar, Derivar, Equilibrar, Filtro y Reclasificar.

Nodos Seleccionar

Los nodos Seleccionar se pueden generar para comprobar la inclusión de registros en una región y la exclusión de todos los registros ajenos a esa región o al revés para el procesamiento posterior.

- **Tramos.** Puede generar un nodo Seleccionar que incluya o excluya los registros de ese tramo. **Nodo seleccionar para tramos únicamente** sólo está disponible en los menús contextuales, ya que necesita seleccionar el tramo que va a utilizar en el nodo Seleccionar.
- **Regiones.** Puede generar un nodo Seleccionar que incluya o excluya los registros de una región.
- **Elementos marcados.** Puede generar nodos Seleccionar para capturar los registros correspondientes a los elementos marcados o enlaces de gráfico de mallas.

Nodos Derivar

Los nodos derivar se pueden generar a partir de regiones, tramos y elementos marcados. Todos los gráficos pueden producir nodos Derivar. En el caso de los diagramas de evaluación, se abre un cuadro de diálogo para seleccionar el modelo. En el caso de gráficos de malla, puede seleccionar los nodos **Derivar ("Y")** y **Derivar ("O")**.

- **Tramos.** Puede generar un nodo Derivar que produzca una categoría para cada intervalo marcado en el eje, utilizando los nombres de tramos que aparecen en el cuadro de diálogo Editar tramos como nombres de categoría.
- **Regiones.** Puede generar un nodo Derivar (**Derivar como marca**) que crea un campo de marca denominado *en_región* con las marcas definidas como *T* para los registros dentro de cualquier región y *F* para registros fuera de todas las regiones. También puede generar un nodo Derivar [**Derivar como**

conjunto] que produce un conjunto con un valor para cada región con un nuevo campo denominado *región* para cada registro, que toma como su valor el nombre de la región en la que se encuentran los registros. Los registros que no están en ninguna región reciben el nombre de la región predeterminada. Los nombres de los valores son los nombres de la región que aparecen en el cuadro de diálogo Editar regiones.

- **Elementos marcados.** Puede generar un nodo Derivar que calcula una marca que es *True* para todos los elementos marcados y *False* para el resto de registros.

Nodos Equilibrar

Los nodos Equilibrar se pueden generar para corregir desequilibrios en los datos, como reducir la frecuencia de valores comunes [utilice la opción de menú **Nodo Equilibrar (reducir)**] o para aumentar las instancias de los valores menos frecuentes [utilice la opción de menú **Nodo Equilibrar (aumentar)**]. La generación del nodo Equilibrar está activada para cualquier gráfico que muestra una distribución de cuentas, como Histograma, Puntos, Colección, Barras de recuentos, Sectores de recuentos y G. múltiple.

Nodos Filtrar

Los nodos Filtrar se pueden generar para cambiar el nombre o filtrar los campos en función de las líneas o nodos marcados en el gráfico. En el caso de los diagramas de evaluación, la mejor línea de ajuste no genera un nodo Filtrar.

Nodos Reclasificar

Los nodos Reclasificar se pueden generar para volver a codificar los valores. Esta opción se utiliza para los gráficos de distribución. Puede generar un nodo Reclasificar para **grupos** para volver a codificar valores concretos de un campo representado en función de su inclusión en un grupo (seleccione los grupos mediante Ctrl+pulsar en la pestaña **Tablas**). También puede generar un nodo Reclasificar para **valores** para volver a codificar datos en un conjunto existente de numerosos valores, como reclasificar datos en un conjunto estándar de valores para fusionar datos financieros de varias empresas para su análisis.

Note: Si los valores se definen previamente, puede leerlos en IBM SPSS Modeler como un archivo sin formato y utilizar una distribución para mostrar todos los valores. A continuación, puede generar un nodo Reclasificar (valores) para este campo directamente desde el gráfico. Si lo hace, colocará todos los valores objetivo en la columna *Valores nuevos* del nodo Reclasificar (lista desplegable).

Cuando establece opciones para el nodo Reclasificar, la tabla habilita una correlación clara entre los valores establecidos antiguos y los valores nuevos que especifique:

- **Valor original.** Esta columna enumera los valores existentes para los campos seleccionados.
- **Valor nuevo.** Utilice esta columna para escribir nuevos valores de categoría o seleccionar uno de la lista desplegable. Al generar automáticamente un nodo Reclasificar con los valores del cuadro Distribución, estos valores se incluyen en la lista desplegable. Esto permite correlacionar rápidamente los valores existentes a un conjunto de valores conocido. Por ejemplo, en ocasiones las organizaciones de salud agrupan los diagnósticos de forma distinta en función de la configuración de red o local. Después de una fusión o una adquisición, todas las partes deben reclasificar los datos nuevos e incluso los existentes de forma coherente. En lugar de escribir manualmente cada valor objetivo de una larga lista, puede leer la lista maestra de valores en IBM SPSS Modeler, ejecutar un cuadro Distribución para el campo *Diagnóstico* y generar un nodo Reclasificar (valores) para este campo directamente desde el cuadro. Este proceso hará que todos los valores objetivo de Diagnóstico estén disponibles en la lista desplegable Valores nuevos.

Para obtener más información sobre el Nodo Reclasificar, consulte “Opciones de configuración del nodo Reclasificar” en la página 151.

Generación de nodos a partir de gráficos

Puede utilizar el menú Generar en la ventana del resultado de gráfico para generar nodos. El nodo generado se colocará en el lienzo de rutas. Para utilizar el nodo, conéctelo a una ruta existente.

Para generar un nodo a partir de un gráfico:

1. Compruebe que está en Explorar modo. Seleccione **Ver > Explorar modo**.
2. En la barra de herramientas Explorar modo, pulse en el botón Región.
3. Defina los tramos, regiones o cualquier elemento marcado que necesite para generar el nodo.
4. En el menú Generar, seleccione el tipo de nodo que desea producir. Sólo están activados los que son posibles.

Note: También puede generar nodos directamente del gráfico pulsando con el botón derecho en el gráfico y seleccionando la opción de generación que desee de los menús contextuales.

Edición de visualizaciones

Mientras el modo de exploración le permite explorar analíticamente los datos y valores representados por la visualización, el modo de edición le permite cambiar el diseño y aspecto de la visualización. Por ejemplo, puede cambiar las fuentes y colores para que coincidan con el manual de estilo de su organización. Para seleccionar este modo, seleccione **Ver > Modo edición** de los menús (o pulse en el icono de la barra de herramientas).

En el modo de edición hay varias barras de herramientas que afectan a distintos aspectos del diseño de la visualización. Si no utiliza algunas de ellas, puede ocultarlas para ampliar el espacio del cuadro de diálogo en el que aparece el gráfico. Para seleccionar o anular la selección de las barras de herramientas, pulse el nombre de la barra de herramientas correspondiente en el menú Ver.

Nota: para añadir más detalle a sus visualizaciones, puede aplicar un título, una nota al pie y etiquetas de ejes. Consulte el tema “Adición de títulos y notas al pie” en la página 291 para obtener más información.

Tiene varias opciones para editar una visualización en el **Modo de edición**. Tiene la posibilidad de:

- Editar texto y darle formato.
- Cambiar el color de relleno, la transparencia y el patrón de marcos y elementos gráficos.
- Cambiar el color y el estilo de guión de los bordes y las líneas.
- Girar y cambiar la forma y la relación de aspecto de los elementos de punto.
- Cambiar el tamaño de los elementos gráficos (como barras y puntos).
- Ajustar el espacio que rodea a los elementos mediante márgenes y relleno.
- Especificar un formato para los números.
- Cambiar la configuración del eje y la escala.
- Ordenar, excluir y contraer categorías en un eje categórico.
- Definir la orientación de los paneles.
- Aplicar transformaciones a un sistema de coordenadas.
- Cambiar estadísticos, tipos de elementos gráficos y modificadores de colisión.
- Cambiar la posición de la leyenda.
- Aplicar hojas de estilo de visualización.

Los siguientes temas describen cómo se pueden realizar estas distintas tareas. También se recomienda leer las reglas generales de edición de gráficos.

Para cambiar al modo de edición

Seleccione en los menús:

Ver > Modo de edición

Reglas generales de edición de visualizaciones

Modo edición

Todas las ediciones se realizan en modo de edición. Para activarlo, elija en los menús:

Ver > Modo de edición

Selección

Las opciones disponibles para edición dependen de la selección. Se activan diferentes opciones de la barra de herramientas y la paleta de propiedades según lo que se seleccione. Sólo se aplican a la selección actual los elementos activados. Por ejemplo, si se selecciona un eje, las pestañas Escalas, Marcas principales y Marcas secundarias estarán disponibles en la paleta de propiedades.

A continuación mostramos algunos consejos para seleccionar elementos de la visualización:

- Pulse en un elemento para seleccionarlo.
- Seleccione un elemento gráfico (como los puntos de un diagrama de dispersión o las barras de un gráfico de barras) pulsando una vez. Tras la selección inicial, vuelva a pulsar para reducir la selección a grupos de elementos gráficos o a un único elemento gráfico.
- Pulse en Esc para anular la selección de todo.

Paletas

Cuando se selecciona un elemento de la visualización, las diferentes paletas se actualizan para reflejar la selección. Las paletas contienen controles para realizar las modificaciones en la selección. Las paletas pueden ser barras de herramientas o un panel con múltiples controles y pestañas. Las paletas pueden estar ocultas, por lo que hay que asegurarse de que se muestra la paleta adecuada para realizar las ediciones. Compruebe en el menú Ver las paletas que se muestran actualmente.

Puede cambiar la posición de las paletas pulsando y arrastrando el espacio vacío de una paleta de barra de herramientas o en la parte izquierda de otras paletas. La información visual le permite saber dónde anclar la paleta. En el caso de paletas que no son de la barra de herramientas, también puede hacer en el botón cerrar para ocultar la paleta y en el botón desanclar para mostrar la paleta en una ventana diferente. Pulse el botón ayuda para mostrar la ayuda de la paleta específica.

Configuración automática

Algunos parámetros de configuración incluyen la opción **-automático-**. Esto indica que se aplican valores automáticos. Los ajustes automáticos que se utilizarán dependen de la visualización específica y los valores de los datos. Puede introducir un valor que sustituya a la configuración automática. Si desea restaurar la configuración automática, elimine el valor actual y pulse en Entrar. El parámetro volverá a mostrar la opción **-automático-**.

Eliminación y ocultación de elementos

Puede eliminar u ocultar varios elementos en la visualización. Por ejemplo, puede ocultar la leyenda o la etiqueta del eje. Seleccione un elemento y pulse la tecla Supr para eliminarlo. Si no se permite eliminar ese elemento, no se realizará ninguna acción. Si elimina un elemento por accidente, pulse Ctrl+Z para deshacer la eliminación.

Estado

Algunas barras de herramientas muestran el estado de la selección actual, mientras que otras no. La paleta de propiedades siempre muestra el estado. Si una barra de herramientas *no* muestra el estado, se indicará en el tema que describe dicha barra de herramientas.

Edición y formato de texto

Puede editar texto directamente y cambiar el formato de un bloque de texto entero. Tenga en cuenta que no podrá editar texto directamente enlazado a valores de datos. Por ejemplo, no podrá editar una etiqueta de marca, ya que su contenido se deriva de los datos subyacentes. Sin embargo, puede formatear cualquier texto de la visualización.

Cómo editar texto directamente

1. Pulse dos veces en el bloque de texto. De esta forma se selecciona todo el texto. Todas las barras de texto se desactivarán en este momento, ya que no podrá cambiar ninguna otra parte de la visualización mientras edita el texto.
2. Escriba sobre el texto que desea sustituir. También puede volver a pulsar en el texto para que aparezca un cursor. Coloque el cursor en la posición deseada y escriba el texto adicional.

Cómo dar formato al texto

1. Seleccione el marco que contiene el texto. No pulse dos veces en el texto.
2. Use la barra de herramientas de fuentes para dar formato al texto. Si la barra de herramientas no está activada, asegúrese de seleccionar sólo el *marco* que contiene el texto. Si el texto está seleccionado, se desactivará la barra de herramientas.

Puede cambiar distintas características de la fuente:

- Color
- Familia (por ejemplo, Arial o Verdana)
- Tamaño (la unidad es pt, a menos que se indique otra distinta, como pc)
- Ponderación
- Alineación relativa al marco del texto

El formato se aplica a todo el texto de un marco. No puede cambiar el formato de letras o palabras individuales de ningún bloque de texto concreto.

Cambio de colores, tramas, estilos de guión y transparencia

Muchos elementos diferentes de una visualización tienen un relleno y un borde. El ejemplo más obvio es la barra de un gráfico de barras. El color de las barras es el color de relleno. También pueden llevar un borde negro y continuo alrededor.

Hay otros elementos menos obvios en la visualización que tienen colores de relleno. Si el color de relleno es transparente, es posible que no se vea que existe dicho relleno. Tomemos como ejemplo el texto de una etiqueta de eje. Da la sensación de que el texto es "flotante", pero en realidad aparece en un marco con color de relleno transparente. Puede ver el marco si selecciona la etiqueta del eje.

Cualquier marco de la visualización puede tener un estilo de relleno y borde, incluyendo el marco alrededor de toda la visualización. Además, cualquier relleno tiene un nivel de opacidad/transparencia asociado que puede ajustarse.

Para cambiar de colores, tramas, estilos de guión y transparencia

1. Seleccione el elemento cuyo formato desea cambiar. Por ejemplo, seleccione las barras de un gráfico de barras o un marco que contenga texto. Si la visualización está dividida por un campo o variable categórica, puede seleccionar también el grupo que corresponda a una categoría individual. con el fin

de cambiar el aspecto estético predeterminado asignado a dicho grupo. Por ejemplo, puede cambiar el color de uno de los grupos de apilado de un gráfico de barras apiladas.

2. Use la barra de herramientas de colores para cambiar el color de relleno o del borde, o bien la trama de relleno.

Note: esta barra de herramientas no muestra el estado de la selección actual.

Para cambiar un color o relleno, puede pulsar el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción. En el caso de los colores, observe que hay uno representado como blanco y atravesado por una línea diagonal roja. Se trata del color transparente. Puede usarlo, por ejemplo, para ocultar los bordes de las barras de un histograma.

- El primer botón controla el color de relleno. Si el color está asociado con un campo continuo u ordinal, este botón cambia el color de relleno del color asociado con el mayor valor de los datos. Puede utilizar la pestaña Color de la paleta de propiedades para cambiar el color asociado con el menor valor y los datos que faltan. El color de los elementos irán cambiado gradualmente desde el color bajo al alto a medida que aumenten los datos subyacentes.
 - El segundo botón controla el color del borde.
 - El tercer botón controla la trama de relleno. La trama de relleno utiliza el color del borde. Por tanto, la trama de relleno sólo se ve si el color del borde es visible.
 - El cuarto control es una barra deslizante y un cuadro de texto que controlan la opacidad del color y el patrón de relleno. Si el porcentaje es menor, habrá menor opacidad y más transparencia. 100% significa que es totalmente opaco (sin transparencia).
3. Para cambiar el estilo de guión de un borde o una línea, use la barra de herramientas de líneas.

Note: esta barra de herramientas no muestra el estado de la selección actual.

Al igual que con la otra barra de herramientas, puede pulsar en el botón para seleccionar la opción mostrada o en la flecha desplegable para seleccionar otra opción.

Rotación y cambio de la forma y la relación de aspecto de los elementos de punto

Puede girar elementos de punto, asignar una forma predefinida diferente o cambiar la relación de aspecto (la relación entre anchura y altura).

Cómo modificar elementos de punto

1. Seleccione los elementos de punto. No puede girar ni cambiar la forma y relación de aspecto de elementos de punto individuales.
2. Use la barra de herramientas de símbolos para modificar los puntos.
 - El primer botón permite cambiar la forma de los puntos. Pulse en la flecha desplegable y seleccione una forma predefinida.
 - Con el segundo botón puede girar los puntos hasta una posición específica como si lo hiciera con un compás. Pulse en la flecha desplegable y, a continuación, arrastre la punta hasta la posición deseada.
 - El tercero permite cambiar la relación de aspecto. Pulse en la flecha desplegable y, a continuación, arrastre el rectángulo que aparece. La forma del rectángulo representa la relación de aspecto.

Cambio de tamaño de los elementos gráficos

Puede cambiar el tamaño de los elementos gráficos de la visualización. Estos incluyen barras, líneas y puntos entre otros. Si el tamaño del elemento gráfico depende de una variable o un campo, el tamaño especificado es el *mínimo*.

Cómo cambiar el tamaño de los elementos gráficos

1. Seleccione los elementos gráficos cuyo tamaño desea cambiar.

- Use el control deslizante o introduzca un tamaño específico para la opción disponible en la barra de herramientas de símbolos. La unidad es píxeles, a menos que se indique otra distinta (consulte a continuación la lista completa de abreviaturas de unidad). También puede especificar un porcentaje (como 30 %), lo que significa que un elemento gráfico usa el porcentaje especificado del espacio disponible. El espacio disponible dependerá del tipo de elemento gráfico y la visualización específica.

Tabla 51. Abreviaturas de unidad válidas

Abreviatura	Unidad
cm	centímetro
en	pulgada
mm	milímetro
pc	pica
pt	punto
px	pixel

Especificación de márgenes y relleno

Si hay demasiado espacio o un espacio insuficiente alrededor o dentro de un marco de la visualización, puede cambiar este margen y los ajustes de relleno. El **margen** es la cantidad de espacio existente entre el marco y otros elementos que lo rodean. El **relleno** es la cantidad de espacio existente entre el borde del marco y el *contenido* de dicho marco.

Cómo especificar los márgenes y el relleno

- Seleccione el marco cuyos márgenes y relleno desea especificar. Puede tratarse de un marco de texto, el marco que rodea a la leyenda o incluso el marco de datos que muestra los elementos gráficos (como barras y puntos).
- Use la pestaña Márgenes de la paleta de propiedades para especificar la configuración. Todos los tamaños se expresan en píxeles, a menos que se indique una unidad distinta (como cm o pulg).

Formato de numeración

Puede especificar el formato de los números de las etiquetas señalizadoras en el eje continuo o las etiquetas de valor de datos que muestran un número. Por ejemplo, puede especificar que los números que aparecen en las etiquetas de marcas se muestren en miles.

Para especificar formatos de numeración

- Seleccione las etiquetas señalizadoras del eje continuo o las etiquetas de valor de datos si contienen números.
- Pulse en la pestaña **Formato** de la paleta Propiedades.
- Seleccione las opciones de formato de numeración que desee:
 - Prefijo.** Carácter que se mostrará al comienzo del número. Introduzca, por ejemplo, un signo de dólar (\$) si los números se refieren a salarios en dólares estadounidenses.
 - Sufijo.** Carácter que se mostrará al final del número. Introduzca, por ejemplo, un signo de porcentaje (%) si los números equivalen a porcentajes.
 - Dígitos de entero mínimo.** Número mínimo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real no contiene el número mínimo de dígitos, la parte entera del valor se rellenará con ceros.
 - Dígitos de entero máx.** Número máximo de dígitos que mostrar en la parte entera de una representación decimal. Si el valor real supera el número máximo de dígitos, la parte entera del valor se rellenará con asteriscos.

Dígitos de decimal mínimo. Número mínimo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real no contiene el número mínimo de dígitos, la parte decimal del valor se rellenará con ceros.

Dígitos de decimal máx. Número máximo de dígitos que mostrar en la parte decimal de una representación decimal o científica. Si el valor real supera el número máximo de dígitos, la parte decimal se redondeará al correspondiente número de dígitos.

Científica. Determina si se muestran o no los números en notación científica. La notación científica resulta útil para números muy grandes o muy pequeños. **-auto-** permite a la aplicación determinar cuándo es adecuado utilizar la notación científica.

Escala. Un factor de escala, que es un número por el que se divide el valor original. Utilice un factor de escala si los números son grandes, pero no desea que la etiqueta se extienda demasiado para adaptarse al número. Si cambia el formato de numeración de las etiquetas señalizadoras, asegúrese de editar el título del eje para indicar cómo se debe interpretar el número. Supongamos, por ejemplo, que el eje de escala muestra salarios y las etiquetas son 30.000, 50.000 y 70.000. Puede introducir un factor de escala de 1000 para mostrar 30, 50 y 70. A continuación, debe editar el título del eje para incluir el texto en miles.

Paréntesis -ve. Indica si deben incluirse paréntesis alrededor de los valores negativos.

Agrupación. Indica si se debe mostrar un carácter entre grupos de dígitos. La configuración regional actual de su equipo determinará qué carácter se utiliza para la agrupación de dígitos.

Cambio de la configuración del eje y la escala

Existen varias opciones para modificar ejes y escalas.

Cómo cambiar la configuración del eje y la escala

1. Seleccione cualquier parte del eje (por ejemplo, la etiqueta del eje o las etiquetas de marca).
2. Use las pestañas Escalas, Marcas principales y Marcas secundarias de la paleta de propiedades para cambiar la configuración del eje y la escala.

Pestaña Escalas

Nota: la pestaña Escalas no aparece para gráficos en los que los datos estén agregados previamente (por ejemplo, histogramas).

Tipo. Especifica si la escala es lineal o transformada. Las transformaciones de escala permiten comprender los datos o realizar los supuestos necesarios para la inferencia estadística. En los diagramas de dispersión, puede utilizar una escala transformada si la relación entre las variables o los campos independientes y dependientes es no lineal. Las transformaciones de escala también se pueden utilizar para lograr que un histograma asimétrico sea más simétrico y se parezca a una distribución normal. Tenga en cuenta que sólo se transforma la escala en que se muestran los datos; los datos reales no se transforman.

- **Lineal.** Especifica una escala lineal sin transformar.
- **Log.** Especifica una escala transformada de logaritmo en base 10. Para adaptarse a los valores cero y negativos, esta transformación usa una versión modificada de la función logarítmica. Esta función de "logaritmo seguro" se define como $\text{signo}(x) * \log(1 + \text{abs}(x))$. Por tanto, $\text{logaritmoSeguro}(-99)$ equivale a:
$$\text{signo}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$
- **potencia** Especifica una escala transformada de potencia con un exponente de 0,5. Para adaptarse a los valores negativos, esta transformación usa una versión modificada de la función exponencial. Esta función de "potencia segura" se define como $\text{signo}(x) * \text{pot}(\text{abs}(x), 0,5)$. Por tanto, $\text{potenciaSegura}(-100)$ equivale a:
$$\text{signo}(-100) * \text{pot}(\text{abs}(-100), 0,5) = -1 * \text{pot}(100, 0,5) = -1 * 10 = -10$$

Mín/Máx/Bajo redondeado/Alto redondeado. Especifica el rango de la escala. Si selecciona **Bajo redondeado** y **Alto redondeado**, la aplicación podrá seleccionar una escala apropiada de acuerdo con los datos. Los valores mínimo y máximo están redondeados porque suelen ser valores enteros mayores o menores que los valores de datos máximos y mínimos. Por ejemplo, si los datos oscilan entre 4 y 92, los valores alto y bajo redondeados para la escala pueden estar comprendidos entre 0 y 100 en lugar de los valores mínimo y máximo reales de los datos. Preste atención para no definir un rango demasiado pequeño que oculte elementos importantes. Tenga en cuenta también que no puede definir un valor mínimo y máximo explícito si está seleccionada la opción **Incluir cero**.

Margen bajo/Margen alto. Crea márgenes en los extremos inferior y/o superior del eje. El margen aparece en posición perpendicular con respecto al eje seleccionado. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg). Por ejemplo, si define **Margen alto** en 5 para el eje vertical, se extenderá un margen horizontal de 5 px a lo largo de la parte superior del marco de datos.

Inversa. Especifica si la escala está invertida.

Incluir cero. Indica que la escala debe incluir el valor 0. Esta opción suele utilizarse en gráficos de barras para asegurarse de que las barras empiezan en el 0, y no en un valor cercano a la altura de la barra más pequeña. Si se selecciona esta opción, **Mín** y **Máx** se desactivan ya que no se pueden definir los valores mínimo y máximo del rango de la escala.

Pestañas Marcas principales/Marcas secundarias

Las **marcas** o **marcas de comprobación** son líneas que aparecen sobre un eje. Indican valores en intervalos o categorías específicos. Las **marcas principales** son marcas de comprobación con etiquetas. Son también más largas que las otras marcas de comprobación. Las **marcas secundarias** son marcas de comprobación que aparecen entre las marcas de comprobación principales. Algunas opciones son específicas del tipo de marca, pero la mayoría de las opciones están disponibles para marcas principales y secundarias.

Mostrar marcas. Especifica si aparecen marcas mayores o menores en el gráfico.

Mostrar líneas de cuadrícula. Especifica si las líneas de cuadrícula aparecen en las marcas principales o secundarias. Las **Líneas de cuadrícula** son líneas que atraviesan todo el gráfico de eje a eje.

Posición. Especifica la posición de las marcas de comprobación en relación con el eje.

Longitud. Especifica la longitud de las marcas de comprobación. La unidad es píxeles, a menos que se indique otra distinta (como cm o pulg).

Base. *Sólo se aplica a las marcas principales.* Especifica el valor en el que aparece la primera marca principal.

Delta. *Sólo se aplica a las marcas principales.* Especifica la diferencia entre marcas principales. Esto es, las marcas principales aparecerán cada n valores, donde n es el valor delta.

Divisiones. *Sólo se aplica a las marcas secundarias.* Especifica el número de divisiones de marcas secundarias existentes entre las marcas principales. El número de marcas secundarias es igual al número de divisiones menos uno. Por ejemplo, supongamos que hay marcas principales en 0 y 100. Si especifica 2 como número de divisiones de marcas secundarias, habrá *una* marca secundaria en 50, dividiendo el intervalo de 0–100, lo que creará *dos* divisiones.

Edición de categorías

Puede editar las categorías en un eje categórico de varias formas:

- Cambie el orden de clasificación para mostrar las categorías.
- Excluya categorías específicas.

- Añada una categoría que no aparezca en el conjunto de datos.
- Contraiga/combine categorías pequeñas en una categoría.

Cómo cambiar el orden de clasificación de categorías

1. Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

Note: Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de IBM SPSS Modeler, seleccione **Categorías**.

2. En la paleta Categorías, seleccione una opción de ordenación de la lista desplegable.

Personalizado. Ordena las categorías basándose en el orden en el que aparecen en la paleta. Utilice los botones de flecha para mover categorías a la parte superior de la lista, arriba, abajo y a la parte inferior de la lista.

Datos. Ordena las categorías basándose en el orden en el que aparecen en el conjunto de datos.

Nombre. Ordena las categorías alfabéticamente, utilizando los nombres como aparecen en la paleta. Esto puede ser tanto el valor como la etiqueta, dependiendo de si está seleccionado el botón de la barra de herramientas para mostrar valores y etiquetas.

Valor. Ordena las categorías según el valor de datos subyacente, utilizando los valores que aparecen entre paréntesis en la paleta. Sólo los orígenes de datos con metadatos (como los archivos de datos de IBM SPSS Statistics) admiten esta opción.

Estadístico. Ordena las categorías basándose en el estadístico calculado para cada categoría. Algunos ejemplos de estadísticos son recuentos, porcentajes y medias. Esta opción sólo está disponible si se utiliza un estadístico en el gráfico.

Para añadir una categoría

De forma predeterminada, únicamente las categorías que aparecen en el conjunto de datos están disponibles. Puede añadir una categoría a la vista si fuera necesario.

1. Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

Note: Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de IBM SPSS Modeler, seleccione **Categorías**.

2. En la paleta de categorías, pulse en el botón Añadir categoría:



Figura 63. Botón Añadir categoría

3. En el cuadro de diálogo Añadir una nueva categoría, introduzca un nombre para la categoría.
4. Pulse en **Aceptar**.

Cómo excluir categorías específicas

1. Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

Note: Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de IBM SPSS Modeler, seleccione **Categorías**.

2. En la paleta Categorías, seleccione un nombre de categoría en la lista Incluir y, a continuación, pulse en el botón X. Para mover la categoría hacia atrás, seleccione su nombre en la lista Excluido y, a continuación, pulse en la flecha situada a la derecha de la lista.

Cómo contraer/ combinar categorías pequeñas

Puede combinar categorías que sean tan pequeñas que no necesite mostrarlas por separado. Por ejemplo, si tiene un gráfico circular con muchas categorías, considere la opción de contraer categorías con un

porcentaje inferior a 10. Contraer sólo está disponible para las estadísticas que son aditivas. Por ejemplo, no puede añadir medias porque las medias no son aditivas. Por lo tanto, no se pueden combinar/contraer categorías utilizando una media.

1. Seleccione un eje categórico. La paleta Categorías muestra las categorías en el eje.

Note: Si la paleta no está visible, asegúrese de haberla activado. En el menú Ver de IBM SPSS Modeler, seleccione **Categorías**.

2. En la paleta Categorías, seleccione **Contraer** y especifique un porcentaje. Cualquier categoría cuyo porcentaje del total sea inferior al número especificado se combinará en una categoría. El porcentaje se basa en el estadístico mostrado en el gráfico. La contracción sólo está disponible para estadísticos basados en recuentos y de suma.

Cambio de la orientación de paneles

Si utiliza paneles en su visualización, puede cambiar su orientación.

Cómo cambiar la orientación de los paneles

1. Seleccione cualquier parte de la visualización.
2. Pulse en la pestaña **Paneles** de la paleta Propiedades.
3. Seleccione una opción en **Diseño**:

Tabla. Organiza los paneles como una tabla, de manera que hay una fila o columna asignada a cada valor individual.

Transpuesto. Organiza los paneles como una tabla, aunque también intercambia las filas y columnas originales. Esta opción no equivale a transponer el gráfico en sí. Observe que los ejes x e y no cambian al seleccionar esta opción.

Lista. Organiza los paneles como una lista, de manera que cada casilla representa una combinación de valores. Las columnas y filas ya no estarán asignadas a valores individuales. Esta opción permite que los paneles se ajusten si es necesario.

Transformación del sistema de coordenadas

Muchas visualizaciones se muestran en un sistema de coordenadas plano y rectangular. Puede transformar el sistema de coordenadas según sea necesario. Por ejemplo puede aplicar una transformación polar al sistema de coordenadas, añadir efectos de sombra en sentido oblicuo y transponer los ejes. También puede deshacer cualquiera de estas transformaciones si ya se aplicaron a la visualización actual. Por ejemplo, se dibuja un gráfico circular en un sistema de coordenadas polares. Puede deshacer la transformación polar y mostrar el gráfico circular como una única barra apilada en un sistema de coordenadas rectangular.

Para transformar el sistema de coordenadas

1. Seleccione el sistema de coordenadas que desee transformar. El sistema de coordenadas se selecciona eligiendo el marco que rodee el gráfico individual.
2. Pulse en la pestaña **Coordenadas** de la paleta Propiedades.
3. Seleccione las transformaciones que desee aplicar al sistema de coordenadas. También puede cancelar la selección de una transformación para deshacerla.

Transpuesto. El cambio de orientación de los ejes se denomina **transposición**. Es como intercambiar los ejes vertical y horizontal en una visualización 2-D.

Polar. Una transformación polar dibuja los elementos gráficos con un ángulo y una distancia específicos desde el centro del gráfico. Un gráfico circular es una visualización de una dimensión con una transformación polar que dibuja las barras individuales con ángulos específicos. Un gráfico radial

es una visualización 2-D con una transformación polar que dibuja elementos gráficos con ángulos y distancias específicos desde el centro del gráfico. Una visualización 3-D también incluye una dimensión de profundidad adicional.

Oblicuo. Una transformación oblicua añade un efecto 3-D a los elementos gráficos. Esta transformación añade profundidad a los elementos gráficos, pero su profundidad es meramente decorativa. No se ve influenciado por valores de datos concretos.

Misma proporción. Si se aplica la misma proporción se especifica que la misma distancia en cada escala represente la misma diferencia de valores de datos. Por ejemplo, 2 cm en ambas escalas representan una diferencia de 1000.

% del recuadro antes de transformación . Si tras la transformación se recortan los ejes, puede que desee añadir recuadros al gráfico antes de aplicar la transformación. Los recuadros reducen las dimensiones en un porcentaje antes de aplicar transformaciones al sistema de coordenadas. Tiene control sobre las dimensiones x inferior, x superior, y inferior e y superior, en ese orden.

% del recuadro tras la transformación . Si desea cambiar la relación de aspecto del gráfico, puede añadirle recuadros tras aplicar transformación. Los recuadros reducen las dimensiones en un porcentaje después de aplicar transformaciones al sistema de coordenadas. Estos recuadros también pueden aplicarse aunque no se haya realizado ninguna transformación en el gráfico. Tiene control sobre las dimensiones x inferior, x superior, y inferior e y superior, en ese orden.

Cambio de estadísticos y elementos gráficos

Puede convertir un elemento gráfico a otro tipo, cambiar el estadístico utilizado para dibujar el elemento gráfico o especificar el modificador de colisión que determina lo que ocurre cuando los elementos gráficos se solapan.

Para convertir un elemento gráfico

1. Seleccione el elemento gráfico que desee convertir.
2. Pulse en la pestaña **Elemento** de la paleta Propiedades.
3. Seleccione un nuevo tipo de elemento gráfico en la lista Tipo.

Tabla 52. Tipos de elementos gráficos

Tipo de elemento gráfico	Descripción
Point	Un marcador que identifica un punto de datos específico. Un elemento de punto se utiliza en diagramas de dispersión y otras visualizaciones relacionadas.
Intervalo	Una forma rectangular dibujada con un valor de datos específico y que rellena el espacio entre un origen y otro valor de datos. Se utiliza un elemento de intervalo en gráficos de barras e histogramas.
Line	Una línea que conecta los valores de datos.
Path	Una línea que conecta los valores de datos en el orden en que aparecen en el conjunto de datos.
Area	Una línea que conecta los elementos de datos con el área entre la línea y un origen relleno.
Polígono	Una forma con varias caras que encierra una región de datos. Se puede utilizar un elemento poligonal en un diagrama de dispersión en intervalos o en un mapa.
Esquema	Un elemento que consta de un cuadro con bigotes y marcadores indicando valores atípicos. Se utilizan elementos de esquema para gráficos de caja.

Para cambiar el estadístico

1. Seleccione el elemento gráfico cuyo estadístico desee cambiar.
2. Pulse en la pestaña **Elemento** de la paleta Propiedades.

Estadísticos de resumen calculados a partir de un campo continuo

- *Media*. Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.
- *Mediana*. Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).
- *Moda*. El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.
- *Mínimo*. Se trata del valor menor de una variable numérica.
- *Máximo*. Se trata del valor mayor de una variable numérica.
- *Rango*. Diferencia entre los valores mínimos y máximos.
- *Medio intervalo*. La mitad del rango, es decir, el valor cuya diferencia con el mínimo es igual a la diferencia con el máximo.
- *Suma*. Suma o total de todos los valores, a lo largo de todos los casos con valores no perdidos.
- *Suma acumulada*. La suma acumulada de los valores. Cada elemento del gráfico muestra la suma de un subgrupo más la suma total de los grupos anteriores.
- *Porcentaje de la suma*. El porcentaje en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos.
- *Suma del porcentaje acumulado*. El porcentaje acumulativo en cada subgrupo basado en un campo sumado en comparación con la suma en todos los grupos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.
- *Varianza*. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media dividida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.
- *Desviación estándar*. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.
- *Error estándar*. Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación estándar de la distribución muestral de un estadístico. Por ejemplo, el error estándar de la media es la desviación estándar de las medias muestrales.
- *Curtosis*. Es una medida del grado en que las observaciones se agrupan en torno a un punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.
- *Asimetría*. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Los siguientes estadísticos de región pueden resultar en más de un elemento gráfico por subgrupo. Si utiliza los elementos de intervalo, área, o gráficos de borde, un estadístico de región da como resultado un elemento gráfico que muestra el rango. El resto de elementos gráficos dan como resultado dos elementos diferentes, uno que muestra el inicio del rango y el otro que muestra el final del rango.

- **Región: Rango.** El rango entre los valores mínimo y máximo.
- **Región: 95% de intervalo de confianza para la media.** Un rango de valores con un 95% de opciones de incluir la media poblacional.
- **Región: 95% de intervalo de confianza individual.** Un rango de valores con un 95% de opciones de incluir el valor predictivo especificando el caso individual.
- **Región: 1 Desviación estándar por encima/debajo de la media.** Un rango de valores entre 1 *desviación estándar* por encima y por debajo de la *media*.
- **Región: 1 Error estándar por encima/debajo de la media.** Un rango de valores entre 1 *error estándar* por encima y por debajo de la *media*.

Estadísticos de resumen basados en recuentos

- **Recuento.** Número de filas/casos.
- **Recuento acumulado.** El número acumulado de filas/casos. Cada elemento del gráfico muestra el recuento de un subgrupo más el recuento total de los grupos anteriores.
- **Porcentaje de recuento.** El porcentaje de filas/casos de cada subgrupo en comparación con el número total de filas/casos.
- **Porcentaje acumulado de recuento.** El porcentaje acumulado de filas/casos de cada subgrupo en comparación con el número total de filas/casos. Cada elemento del gráfico muestra el porcentaje de un subgrupo más el porcentaje total de los grupos anteriores.

Para especificar el modificador de colisión

El modificador de colisión determina lo que ocurre cuando los elementos gráficos se superponen.

1. Seleccione el elemento gráfico cuyo modificador de colisión desee especificar.
2. Pulse en la pestaña **Elemento** de la paleta Propiedades.
3. En la lista desplegable Modificador, seleccione un modificador de colisión. **-auto-** permite a la aplicación determinar qué modificador de colisión es adecuado para el tipo de elemento gráfico y estadístico.

Superpuesto. Dibuje elementos gráficos encima de otros si tienen el mismo valor.

Pila. Elementos gráficos de pila que normalmente se superpondrían al tener los mismos valores de datos.

Esquivar. Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan simétricamente. Es decir, los elementos gráficos se trasladan a lados opuestos de una posición central. Esquivar es similar a la agrupación en clústeres.

Pila. Mueve los elementos gráficos junto a otros elementos gráficos que aparecen con el mismo valor, en vez de superponerlos. Los elementos gráficos se ordenan asimétricamente. Es decir, los elementos gráficos se apilan unos sobre otros, con el elemento gráfico más bajo colocado en un valor específico de la escala.

Perturbar (normal) . Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución normal.

Perturbar (uniforme) . Recoloca los elementos gráficos en el mismo valor de datos aleatoriamente utilizando la distribución uniforme.

Cambio de la posición de la leyenda

Si un gráfico incluye una leyenda, ésta suele mostrarse a la derecha de un gráfico. Si lo desea, puede cambiar esta posición.

Cómo cambiar la posición de la leyenda

1. Seleccione la leyenda.
2. Pulse en la pestaña **Leyenda** de la paleta Propiedades.

3. Seleccione una posición.

Copia de una visualización y de datos de visualización

La paleta General incluye botones para copiar la visualización y sus datos.



Figura 64. Botón Copiar visualización

Copia de visualización. Esta acción copia la visualización en el portapapeles como una imagen. Hay múltiples formatos de imagen disponibles. Cuando pegue la imagen en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos de imagen disponibles para pegar.



Figura 65. Botón Copiar datos de visualización

Copia de los datos de visualización. Esta acción copia los datos subyacentes utilizados para dibujar la visualización. Los datos se copian en el portapapeles como texto normal o como texto con formato HTML. Cuando pegue los datos en otra aplicación, puede seleccionar una opción de pegado especial para seleccionar uno de los formatos disponibles para pegar.

Atajos de teclado

Tabla 53. Atajos de teclado

Tecla de acceso directo	Función
Ctrl + Espacio	Cambiar entre modo de exploración y el modo de edición
Suprimir	Eliminar un elemento de visualización
Ctrl + Z	Deshacer
Ctrl + Y	Rehacer
F2	Mostrar esquema para seleccionar los elementos del gráfico

Adición de títulos y notas al pie

En todos los tipos de gráfico, es posible añadir títulos, notas al pie o etiquetas de eje exclusivos que ayuden a identificar lo que aparece en el gráfico.

Adición de títulos a gráficos

1. Desde los menús, seleccione **Edición > Añadir título de gráfico**. Se mostrará un cuadro de texto que contiene <TITLE> sobre el gráfico.
2. Compruebe que está en el modo de edición. Desde los menús, seleccione **Ver > Modo edición**.
3. Pulse dos veces en el texto <TITLE>.
4. Escriba el título que desee y pulse Volver.

Adición de notas al pie del gráfico

1. Desde los menús, seleccione **Edición > Añadir nota al pie del gráfico**. Se mostrará un cuadro de texto que contiene <FOOTNOTE> debajo del gráfico.

2. Compruebe que está en el modo de edición. Desde los menús, seleccione **Ver > Modo edición**.
3. Pulse dos veces en el texto <FOOTNOTE>.
4. Escriba el título que desee y pulse Volver.

Uso de hojas de estilo de gráficos

La información de visualización de gráfico básica, como colores, fuentes, símbolos y grosor de líneas, se controla en una hoja de estilo. Se suministra una hoja de estilo predeterminada con IBM SPSS Modeler, sin embargo, puede modificarla si lo necesita. Por ejemplo, es posible que desee utilizar en los gráficos el esquema de color corporativo que suele utilizar en las presentaciones. Consulte el tema “Edición de visualizaciones” en la página 279 para obtener más información.

En los nodos de gráfico, puede utilizar el modo de edición para realizar cambios de estilo al aspecto de un gráfico. Puede utilizar el menú **Edición > Estilos** para guardar los cambios como una hoja de estilo para aplicarlos a todos los gráficos que genere posteriormente desde el nodo de gráfico actual, o utilizarlo como nueva hoja de estilo predeterminada para todos los gráficos que genere con IBM SPSS Modeler.

Existen cinco opciones de hoja de estilo disponibles en la opción **Estilos** del menú Edición.

- **Cambiar a hoja de estilo.** Muestra una lista de diferentes estilos de hojas almacenados que puede seleccionar para cambiar la apariencia de los gráficos. Consulte el tema “Aplicación de hojas de estilo” para obtener más información.
- **Almacenar estilos en nodo.** Almacena las modificaciones realizadas a los estilos del gráfico seleccionado para poder aplicarlas a los futuros gráficos que se generen desde el mismo nodo de gráfico de la ruta actual.
- **Almacenar estilos como valor predeterminado.** Almacena las modificaciones realizadas a los estilos del gráfico seleccionado para poder aplicarlas a los futuros gráficos que se generen desde cualquier nodo de gráfico de cualquier ruta. Al seleccionar esta opción, puede utilizar **Aplicar estilos predeterminados** para que todos los gráficos existentes utilicen los mismos estilos.
- **Aplicar estilos predeterminados.** Esta opción cambia los estilos del gráfico seleccionado por aquellos que se han guardado como estilos predeterminados.
- **Aplicar estilos originales.** Esta opción vuelve a cambiar los estilos del gráfico seleccionado por aquellos que se suministran de forma predeterminada.

Aplicación de hojas de estilo

Puede aplicar una hoja de estilo de visualización que especifica propiedades de estilo de la visualización. Por ejemplo, la hoja de estilo puede definir fuentes, estilos de guión y colores, entre otras opciones. Hasta cierto punto, las hojas de estilo proporcionan un acceso directo a las modificaciones que tenga que realizar manualmente. Tenga en cuenta, sin embargo, que una hoja de estilo se limita a cambios de *estilo*. Otros cambios, como la posición de la leyenda o el rango de la escala no se guardan en la hoja de estilo.

Cómo se aplica una hoja de estilo

1. Seleccione en los menús:
Editar > Estilos > Cambiar a hoja de estilo
2. Utilice el cuadro de diálogo Cambiar a hoja de estilo para seleccionar una hoja de estilo.
3. Pulse **Aplicar** para aplicar la hoja de estilo a la visualización sin cerrar el cuadro de diálogo. Pulse **Aceptar** para la hoja de estilo y cerrar el cuadro de diálogo.

Cuadro de diálogo Cambiar a/Seleccionar hoja de estilo

La tabla de la parte superior del cuadro de diálogo muestra todas las hojas de estilo de visualización que están actualmente disponibles. Algunas hojas de estilo están preinstaladas, mientras que otras se pueden haber creado en IBM SPSS Visualization Designer (como producto independiente).

La parte inferior del cuadro de diálogo muestra visualizaciones de ejemplo con datos de muestra. Seleccione una de las hojas de estilo que aplicarán sus estilos a las visualizaciones de ejemplo. Estos ejemplos le ayudan a determinar cómo afectarán las hojas de estilo a su visualización real.

El cuadro de diálogo ofrece las siguientes opciones.

Estilos existentes. De forma predeterminada, una hoja de estilo puede sobrescribir todos los estilos de la visualización. Es posible modificar este comportamiento.

- **Sobrescribir todos los estilos.** Cuando se aplica la hoja de estilo, se sobrescriben todos los estilos de la visualización, incluyendo aquellos estilos que se modifican durante la sesión de modificación actual.
- **Conservar estilos modificados.** Cuando se aplica la hoja de estilo, se sobrescriben sólo aquellos estilos que *no* se han modificado en la visualización durante la sesión de modificación actual. Los estilos que se han modificado durante la sesión de modificación actual se conservan.

Administrar. Administra plantillas, hojas de estilo y mapas de visualización en su ordenador. Puede importar, exportar, cambiar el nombre y eliminar plantillas, hojas de estilo y mapas de visualización en su máquina local. Consulte el tema “Administración de plantillas, hojas de estilo y archivos de mapas” en la página 219 para obtener más información.

Ubicación. Cambia la ubicación donde están almacenadas las plantillas, hojas de estilo y mapas de visualización. La ubicación actual se enumera a la derecha del botón. Consulte el tema “Configuración de la ubicación de las plantillas, hojas de estilo y mapas” en la página 218 para obtener más información.

Impresión, almacenamiento, copia y exportación de gráficos

Cada gráfico tiene un número de opciones que permiten guardar o imprimir el gráfico o exportarlo a otro formato. La mayoría de estas opciones están disponibles en el menú Archivo. Adicionalmente, en el menú Edición, puede seleccionar copiar el gráfico o los datos que contienen para utilizarlos en otras aplicaciones.

Impresión

Para imprimir el gráfico, utilice el elemento de menú o el botón **Imprimir**. Antes de imprimir, puede utilizar **Configurar página** y **Presentación preliminar** para configurar las opciones de impresión y previsualizar el resultado.

Almacenamiento de gráficos

Para guardar el gráfico en un archivo de resultados de IBM SPSS Modeler (*.cou), seleccione **Archivo > Guardar** o **Archivo > Guardar como** en los menús.

o

Para guardar el gráfico en el repositorio, elija **Archivo > Almacenar resultado** en los menús.

Copia de gráficos

Para copiar el gráfico para utilizarlo en otra aplicación, como MS Word o MS PowerPoint, elija **Edición > Copiar gráfico** en los menús.

Copia de datos

Para copiar los datos para utilizarlos en otra aplicación, como MS Excel o MS Word, elija **Edición > Copiar datos** en los menús. De forma predeterminada, los datos aparecen con formato HTML. Utilice **Pegado especial** en la otra aplicación para ver otras opciones de formato al pegar.

Exportación de gráficos

La opción **Exportar gráfico** permite exportar el gráfico en uno de los formatos siguientes: mapa de bits (*.bmp*), JPEG (*.jpg*), PNG (*.png*), HTML (*.html*) o documento ViZml (*.xml*) para utilizarlo en otras aplicaciones de IBM SPSS Statistics.

Para exportar gráficos, seleccione **Archivo > Exportar gráfico** en los menús y seleccione el formato.

Exportación de tablas

La opción **Exportar tabla** permite exportar la tabla en uno de los formatos siguientes: delimitado por tabulaciones (*.tab*), delimitado por comas (*.csv*) o HTML (*.html*).

Para exportar tablas, seleccione **Archivo > Exportar tabla** en los menús y seleccione el formato.

Capítulo 6. Nodos de resultado

Conceptos básicos sobre nodos de resultados

Los nodos de resultados ofrecen los medios para obtener información acerca de los datos y los modelos. También proporcionan un mecanismo para exportar datos en varios formatos y poder interactuar con otras herramientas de software.

Los siguientes nodos de resultado están disponibles:



El nodo Tabla muestra los datos en formato de tabla, que también se puede escribir en un archivo. Esto es útil en cualquier momento en que necesite inspeccionar sus valores de datos o exportarlos en un formato fácilmente legible.



El nodo Matriz crea una tabla que muestra las relaciones entre campos. Se suele utilizar normalmente para mostrar las relaciones entre dos campos simbólicos, pero también puede mostrar relaciones entre campos de marcas o entre campos numéricos.



El nodo Análisis evalúa la capacidad de los modelos predictivos para generar predicciones precisas. Los nodos Análisis realizan varias comparaciones entre los valores predichos y los valores reales para uno o más nugget de modelo. También pueden comparar modelos predictivos entre sí.



El nodo Auditoría de datos permite echar un primer vistazo exhaustivo a los datos, incluyendo estadísticos de resumen, histogramas y distribución para cada campo, así como información sobre valores atípicos, valores perdidos y extremos. Los resultados se muestran en una matriz fácil de leer que se puede ordenar y utilizar para generar nodos de preparación de datos y gráficos de tamaño completo.



El nodo Transformación permite seleccionar y previsualizar los resultados de las transformaciones antes de aplicarlas a los campos seleccionados.



El nodo Estadísticos ofrece información básica de resumen acerca de los campos numéricos. Calcula estadísticos de resumen para campos individuales y correlaciones entre campos.



El nodo Medias compara las medias de grupos independientes o de pares de campos relacionados para probar si existen diferencias significativas. Por ejemplo, puede comparar los ingresos medios antes y después de poner en marcha una promoción o comparar los ingresos de los clientes que no recibieron esa promoción con los que sí lo hicieron.



El nodo Informe crea informes con formato que contienen texto fijo, así como datos y otras expresiones derivadas de los datos. Puede especificar el formato del informe utilizando plantillas de texto para definir el texto fijo y las construcciones de resultados de datos. Puede proporcionar formato de texto personalizado utilizando etiquetas HTML de la plantilla y configurando opciones en la pestaña Resultado. Puede incluir valores de datos y otros resultados condicionales mediante el uso de expresiones CLEM en la plantilla.



El nodo Val. globales explora los datos y calcula los valores de resumen que se pueden utilizar en expresiones CLEM. Por ejemplo, puede utilizar este nodo para calcular estadísticos para un campo denominado *edad* y, a continuación, utilizar la media global de *edad* en expresiones CLEM insertando la función @GLOBAL_MEAN(*edad*).



El nodo Ajustar simulación examina la distribución estadística de los datos en cada campo y genera (o actualiza) un nodo Generar simulación, asignando a cada campo la mejor distribución de ajuste. El nodo Generar simulación podrá entonces utilizarse para generar datos simulados.



El nodo Evaluación de simulación evalúa un campo de destino especificado predicho, y presenta la información de correlación y distribución sobre el campo de destino.

Administración de resultados

El gestor de resultados muestra los diagramas, gráficos y tablas generados durante una sesión de IBM SPSS Modeler. Siempre puede volver a abrir un resultado pulsando dos veces en él en el gestor, sin tener que volver a ejecutar la ruta o el nodo correspondiente.

Para ver el gestor de resultados

Abra el menú Ver y elija **Gestores**. Pulse en la pestaña **Resultados**.

Desde el gestor de resultados, puede:

- Visualizar objetos de resultados existentes, como histogramas, diagramas de evaluación y tablas.
- Cambiar el nombre de los objetos de resultados.
- Guardar los objetos de resultados en disco o en IBM SPSS Collaboration and Deployment Services Repository (si está disponible).
- Añadir archivos de resultados al proyecto actual.
- Eliminar objetos de resultados no almacenados desde la sesión actual.
- Abrir los objetos de resultados guardados o recuperarlos de IBM SPSS Collaboration and Deployment Services Repository (si está disponible).

Para acceder a estas opciones, pulse con el botón derecho del ratón en la pestaña Resultados.

Visualización de resultados

El resultado en pantalla se muestra en una ventana del explorador de resultados. La ventana del explorador de resultados tiene su propio conjunto de menús que permiten imprimir y guardar el resultado, así como exportarlo a otro formato. Tenga en cuenta que las opciones específicas pueden variar dependiendo del tipo de resultado.

Impresión, almacenamiento y exportación de datos. Hay más información disponible a continuación:

- Para imprimir el resultado, utilice la opción del menú o el botón **Imprimir**. Antes de imprimir, puede utilizar **Configurar página** y **Presentación preliminar** para configurar las opciones de impresión y previsualizar el resultado.
- Para guardar el resultado en un archivo de salida de IBM SPSS Modeler (.cou), seleccione **Guardar** o **Guardar como** en el menú Archivo.
- Para guardar los resultados en otro formato, como texto o HTML, elija **Exportar** en el menú Archivo. Consulte el tema “Exportación de resultados” en la página 298 para obtener más información.
Tenga en cuenta que sólo puede seleccionar estos formatos si la salida contiene datos que pueden ser razonablemente exportados de esa manera. Por ejemplo, el contenido de un árbol de decisión se pueden exportar como texto, pero el contenido de un modelo K-media no tiene sentido como texto.
- Para guardar el resultado en un repositorio compartido para que otros usuarios puedan visualizarlo utilizando el Portal de despliegue de IBM SPSS Collaboration and Deployment Services, seleccione **Publicar en Web** en el menú Archivo. Tenga en cuenta que esta opción requiere una licencia distinta de IBM SPSS Collaboration and Deployment Services.

Selección de casillas y columnas. El menú Edición contiene varias opciones para seleccionar, anular la selección y copiar casillas y columnas, según considere adecuado para el tipo de resultado en cuestión. Consulte el tema “Selección de casillas y columnas.” en la página 299 para obtener más información.

Generación de nuevos nodos. El menú Generar permite generar nuevos nodos basados en el contenido del explorador de resultados. Las opciones varían dependiendo del tipo de resultado y los elementos del resultado actualmente seleccionados. Para obtener detalles acerca de las opciones de generación de nodos para un tipo particular de resultado, consulte la documentación de ese resultado.

Publicar en Web

La característica Publicar en Web le permite publicar ciertos tipos de resultados de rutas en un IBM SPSS Collaboration and Deployment Services Repository central compartido que conforma la base de IBM SPSS Collaboration and Deployment Services. Si utiliza esta opción, otros usuarios que deseen ver sus resultados podrán hacerlo utilizando un acceso a Internet y una cuenta de IBM SPSS Collaboration and Deployment Services, no necesitan tener instalado IBM SPSS Modeler.

La siguiente tabla enumera los nodos de IBM SPSS Modeler que son compatibles con la característica Publicar en Web. Los resultados de estos nodos se almacenan en IBM SPSS Collaboration and Deployment Services Repository en formato de objeto de resultado (.cou), y pueden verse directamente en Portal de despliegue de IBM SPSS Collaboration and Deployment Services.

El resto de los tipos de resultados pueden verse únicamente si la aplicación apropiada (por ej. IBM SPSS Modeler para objetos de ruta) está instalada en la máquina del usuario.

Tabla 54. Nodos que son compatibles con Publicar en Web.

Tipo de nodo	Nodo
Gráficos	todos
Resultado	Tabla
	Matriz
	Auditoría de datos
	Transformar
	Medias
	Análisis
	Estadísticos
	Informe (HTML)
IBM SPSS Statistics	Resultado Statistics

Publicación de resultados en la Web

Para publicar resultados en la Web:

1. En una ruta de IBM SPSS Modeler, ejecute uno de los nodos enumerados en la tabla. Al hacerlo se creará un objeto de resultado (por ejemplo, una tabla, una matriz o un objeto de informe) en una nueva ventana.
2. Desde la ventana de objeto de resultado, seleccione:
Archivo > Publicar en Web
Nota: si desea exportar archivos HTML simples para su uso en un navegador Web estándar, seleccione **Exportar** en el menú Archivo y seleccione **HTML**.
3. Conecte con IBM SPSS Collaboration and Deployment Services Repository.
Cuando se haya conectado correctamente, se mostrará el cuadro de diálogo Repositorio: Almacenar, que ofrece diversas opciones de almacenamiento.
4. Cuando haya seleccionado las opciones de almacenamiento que desea, pulse en **Almacenar**.

Visualización de resultados publicados en la Web

Debe disponer de una cuenta de IBM SPSS Collaboration and Deployment Services para utilizar esta característica. Si tiene la aplicación apropiada instalada para el tipo de objeto que desea ver (por ejemplo, IBM SPSS Modeler o IBM SPSS Statistics), el resultado se muestra dentro de la aplicación en vez de en el navegador.

Para ver resultados publicados en la Web:

1. Introduzca la siguiente dirección en su navegador: *http://<host_repos>:<puerto_repos>/peb*
donde *host_repos* y *puerto_repos* representan el nombre de host y el número de puerto del host de IBM SPSS Collaboration and Deployment Services.
2. Introduzca los detalles de inicio de sesión de su cuenta de IBM SPSS Collaboration and Deployment Services.
3. Pulse en **Repositorio de contenidos**.
4. Busque el objeto que desee visualizar.
5. pulse en el nombre de objeto. En algunos tipos de objetos, como gráficos, puede haber una demora mientras se representa el objeto en el navegador.

Visualización de resultados en un explorador HTML

En la pestaña Avanzado de los nuggets de modelo Regresión lineal, Regresión logística y PCA/Factorial, puede ver la información en un explorador diferente, como Internet Explorer. La información aparece en formato HTML, lo que le permite guardarla y reutilizarla en otros sitios, como una intranet de empresa, o un sitio de Internet.

Para mostrar la información en un explorador, pulse en el botón situado debajo del icono de modelo, en la esquina superior izquierda de la pestaña Avanzado del nugget del modelo.

Exportación de resultados

En la ventana del explorador de resultados, puede elegir exportar el resultado a otro formato, como texto o HTML. Los formatos de exportación varían dependiendo del tipo de resultado, pero en general son parecidos a las opciones de tipo de archivo disponibles si selecciona **Guardar en un archivo** en el nodo utilizado para generar el resultado.

Nota: Sólo puede seleccionar estos formatos si la salida contiene datos que pueden ser razonablemente exportados de esa manera. Por ejemplo, el contenido de un árbol de decisión se pueden exportar como texto, pero el contenido de un modelo K-media no tiene sentido como texto.

Para exportar resultados

1. En el explorador de resultados, abra el menú Archivo y elija **Exportar**. A continuación, elija el tipo de archivo que desea crear:
 - **Delimitado por tabulaciones (*.tab)**. Esta opción genera un archivo de texto con formato que contiene los valores de datos. Este estilo suele ser útil para generar una representación de texto sin formato de la información que se puede importar a otras aplicaciones. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
 - **Delimitado por comas (*.dat)**. Esta opción genera un archivo de texto delimitado por comas que contiene los valores de datos. Este estilo suele ser útil como una manera rápida de generar un archivo de datos que se puede importar a hojas de cálculo u otras aplicaciones de análisis de datos. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
 - **Delimitado por tabuladores transpuesto (*.tab)**. Esta opción es idéntica a la opción Delimitado por tabuladores, pero los datos se transponen de manera que las filas representan los campos y las columnas representan los registros.
 - **Delimitado por comas transpuesto (*.dat)**. Esta opción es idéntica a la opción Delimitado por comas, pero los datos se transponen de manera que las filas representan los campos y las columnas representan los registros.
 - **HTML (*.html)** Esta opción escribe resultado con formato HTML en un archivo o archivos.

Selección de casillas y columnas.

Una serie de nodos, incluidos los nodos Tabla, Matriz y Medias, generan un resultado tabular. Estas tablas de resultados se pueden ver y manipular de formas similares, incluyendo seleccionar casillas, copiar toda o parte de la tabla al Portapapeles, generar nuevos nodos a partir de la selección actual y guardar e imprimir la tabla.

Selección de casillas. Para seleccionar una casilla, pulse en ella. Para seleccionar un intervalo rectangular de casillas, pulse en una esquina del intervalo que desee, arrastre el ratón a la otra esquina del intervalo y suelte el botón del ratón. Para seleccionar toda una columna, pulse en su cabecera. Para seleccionar varias columnas, pulse la tecla Mayús o Ctrl mientras pulsa en las cabeceras de las columnas.

Cuando realiza una nueva selección, la vieja selección se borra. Si mantiene pulsada la tecla Ctrl mientras realiza la selección, podrá añadir la nueva selección a cualquier selección existente en lugar de borrar la antigua. Puede utilizar este método para seleccionar varias regiones no contiguas de la tabla. El menú Edición también contiene las opciones **Seleccionar todo** y **Borrar selección**.

Reordenación de columnas. Los exploradores de resultados de los nodos Medias y Tabla permiten mover las columnas en la tabla pulsando en la cabecera de columna y arrastrándola hasta la ubicación deseada. Sólo puede mover una columna a la vez.

Nodo Tabla

El nodo Tabla crea una tabla que enumera los valores de sus datos. Se incluyen todos los campos y todos los valores de la ruta, lo que facilita la inspección de sus valores de datos o su exportación en un formato fácilmente legible. También puede resaltar registros que cumplan ciertas condiciones.

Pestaña Configuración del nodo Tabla

Resaltar registros en los que. Puede resaltar registros en la tabla introduciendo una expresión CLEM verdadera para los registros que se van a resaltar. Esta opción se activa solo cuando se selecciona **Salida a pantalla**.

Pestaña de formato del nodo Tabla

La pestaña Formato contiene opciones utilizadas para especificar el formato por campos. Esta pestaña se comparte con el nodo Tipo. Consulte el tema “Pestaña Configuración de Formato de campo” en la página 138 para obtener más información.

Pestaña Resultado del nodo de resultados

Para los nodos que generan resultados en estilo de tabla, la pestaña Resultados permite especificar el formato y ubicación de los resultados.

Nombre del resultado. Especifica el nombre del resultado generado cuando se ejecuta el nodo.

Automático selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar **Personalizado** para especificar un nombre diferente.

Salida a pantalla (valor predeterminado). Crea un objeto de resultados para ver en línea. El objeto de resultados aparecerá en la pestaña Resultados de la ventana del gestor al ejecutar el nodo de resultados.

Salida a archivo. Almacena el resultado en un archivo cuando se ejecuta el nodo. Si selecciona esta opción, introduzca un nombre de archivo (o navegue a un directorio y especifique un nombre de archivo mediante el botón Buscar el archivo) y seleccione un tipo de archivo. Tenga en cuenta que algunos tipos de archivos pueden no estar disponibles para ciertos tipos de resultados.

Los datos son de resultado en el formato de codificación predeterminado del sistema, que se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.

- **Datos (delimitados por tabulaciones) (*.tab).** Esta opción genera un archivo de texto con formato que contiene los valores de datos. Este estilo suele ser útil para generar una representación de texto sin formato de la información que se puede importar a otras aplicaciones. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
- **Datos (delimitados por comas) (*.dat).** Esta opción genera un archivo de texto delimitado por comas que contiene los valores de datos. Este estilo suele ser útil como una manera rápida de generar un archivo de datos que se puede importar a hojas de cálculo u otras aplicaciones de análisis de datos. Esta opción está disponible para los nodos Tabla, Matriz y Medias.
- **HTML (*.html)** Esta opción escribe resultado con formato HTML en un archivo o archivos. Para resultados tabulares (de los nodos Tabla, Matriz o Medias), un conjunto de archivos HTML contiene un panel de contenido que muestra nombres de campos y los datos en una tabla HTML. La tabla se puede dividir en varios archivos HTML si el número de filas de la tabla supera la especificación de **Líneas por página**. En ese caso, el panel de contenido incluye enlaces a todas las páginas de la tabla y ofrece medios para desplazarse por ella. Para resultados no tabulares, se crea un solo archivo HTML que contiene los resultados del nodo.

Note: Si los resultados HTML contienen sólo formato para la primera página, seleccione **Paginar resultados** y ajuste la especificación **Líneas por página** para incluir todos los resultados en una sola página. O, si la plantilla de resultados para los nodos como el nodo Informe contiene etiquetas HTML personalizadas, asegúrese de que ha especificado **Personalizado** como tipo de formato.

- **Archivo de texto (*.txt)** Esta opción genera un archivo de texto que contiene los resultados. Este estilo suele ser útil para generar resultados que se pueden importar a otras aplicaciones, como procesadores de texto o software de presentación. Esta opción no se encuentra disponible para algunos nodos.
- **Objeto de resultados (*.cou).** Los objetos de resultados guardados en este formato se pueden abrir y ver en IBM SPSS Modeler, añadir a los proyectos y publicar y seguir mediante IBM SPSS Collaboration and Deployment Services Repository.

Vista de resultado. Para el nodo Medias, puede especificar si se ve de forma predeterminada el resultado avanzado o simple. Tenga en cuenta que también puede cambiar de una vista a otra al explorar el resultado generado. Consulte el tema “Explorador de resultados del nodo Medias” en la página 321 para obtener más información.

Formato. Para el nodo Informe, puede seleccionar si al resultado se aplica el formato automáticamente o utilizando el HTML incluido en la plantilla. Seleccione **Personalizado** para permitir formato HTML en la plantilla.

Título. Para el nodo Informe, puede especificar un texto de título opcional que aparecerá en la parte superior del resultado del informe.

Resaltar texto insertado. Para el nodo Informe, seleccione esta opción para resaltar el texto generado por expresiones CLEM en la plantilla Informe. Consulte el tema “Pestaña Plantilla del nodo Informe” en la página 323 para obtener más información. Esta opción no es recomendable cuando se utiliza el formato **Personalizado**.

Líneas por página. Para el nodo Informe, especifique un número de líneas para incluir en cada página durante el formato **Automático** del informe de resultados.

Transponer datos. Esta opción transpone los datos antes de exportarlos, de manera que las filas representan campos y las columnas registros.

Note: para tablas grandes, las opciones anteriores pueden ser algo ineficaces, especialmente cuando se trabaja con un servidor remoto. En tales casos, utilizar un nodo de resultado de archivo proporciona un rendimiento mucho mayor. Consulte el tema “nodo de exportación Archivo sin formato” en la página 355 para obtener más información.

Explorador de tablas

El explorador de tablas muestra datos tabulares y permite realizar operaciones estándar incluyendo seleccionar y copiar casillas, reordenar columnas y guardar e imprimir la tabla. Consulte el tema “Selección de casillas y columnas.” en la página 299 para obtener más información. Son las mismas operaciones que puede realizar cuando visualiza la vista previa de los datos en un nodo.

Exportación de datos de la tabla. Puede exportar datos desde el navegador de tablas seleccionando:

Archivo > Exportar

Consulte el tema “Exportación de resultados” en la página 298 para obtener más información.

Los datos se exportan en el formato de codificación predeterminado del sistema, que se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.

Búsqueda en la tabla. El botón de búsqueda (con el icono de los prismáticos) de la barra de herramientas principal activa la barra de herramientas de búsqueda, permitiéndole buscar valores específicos en la tabla. Puede buscar hacia adelante o hacia atrás en la tabla, puede especificar una búsqueda que distinga entre mayúsculas y minúsculas (el botón **Aa**) y puede interrumpir el proceso de búsqueda con el botón de interrupción de búsqueda.

Generación de nuevos nodos. El menú Generar contiene operaciones de generación de nodos.

- **Nodo Seleccionar ("Registros").** Genera un nodo Seleccionar que selecciona los registros para los que se selecciona cualquier casilla de la tabla.
- **Seleccionar ("Y").** Genera un nodo Seleccionar que selecciona los registros que contienen *todos* los valores seleccionados en la tabla.
- **Seleccionar "O").** Genera un nodo Seleccionar que selecciona los registros que contienen *cualquier* valor seleccionado en la tabla.
- **Derivar ("Registros")** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros para los que se selecciona cualquier casilla de la tabla y *F* para los registros restantes.

- **Derivar ("Y").** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros que contienen *todos* los valores seleccionados en la tabla y *F* para los registros restantes.
- **Derivar ("O").** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene *T* para los registros que contienen *cualquiera* de los valores seleccionados en la tabla y *F* para los registros restantes.

Nodo Matriz

El nodo Matriz permite crear una tabla que muestra las relaciones entre campos. Se suele utilizar normalmente para mostrar las relaciones entre dos campos categóricos (marca, nominal u ordinal), pero también se puede utilizar para mostrar relaciones entre campos continuos (rango numérico).

Pestaña Configuración del nodo Matriz

La pestaña Configuración permite especificar opciones para la estructura de la matriz.

Campos. Seleccione un tipo de selección de campo de entre las siguientes opciones:

- **Seleccionado.** Esta opción permite seleccionar un campo categórico para las filas y otro para las columnas de la matriz. Las filas y columnas de la matriz están definidas por la lista de valores para el campo categórico seleccionado. Las casillas de la matriz contienen los estadísticos de resumen seleccionados más abajo.
- **Todas las marcas (valores verdaderos).** Esta opción solicita una matriz con una fila y una columna para cada campo de marca en los datos. Las casillas de la matriz contienen los recuentos de positivos dobles para cada combinación de marcas. En otras palabras, para una fila correspondiente a *pan comprado* y una columna correspondiente a *queso comprado*, la casilla en la intersección de esa fila y esa columna contiene el número de registros para los que *pan comprado* y *queso comprado* son verdaderos.
- **Todos los numéricos.** Esta opción solicita una matriz con una fila y una columna para cada campo numérico. Las casillas de la matriz representan la suma de los productos cruzados para el par de campos correspondiente. En otras palabras, para cada casilla de la matriz, los valores para el campo de fila y el campo de columna se multiplican por cada registro y, a continuación, se suman en los registros.

Incluir valores perdidos. Incluye los valores perdidos por el sistema (\$null\$) o por el usuario (vacíos) en los resultados de fila y columna. Por ejemplo, si el valor *N/A* se ha definido como perdido por el usuario para el campo de columna seleccionado, se incluirá en la tabla una columna diferente etiquetada con *N/A* (suponiendo que este valor realmente aparezca en los datos) al igual que cualquier otra categoría. Si se anula la selección de esta opción, la columna *N/A* se excluirá independientemente de la frecuencia con que aparezca.

Nota: la opción de incluir valores perdidos sólo se aplica cuando los campos seleccionados son de tabulación cruzada. Los valores vacíos se correlacionan a \$null\$ y se excluyen de la agregación para el campo de la función cuando el modo está **Seleccionado** y el contenido está definido como **Función**, y para todos los campos numéricos cuando el modo está definido como **Todos los numéricos**.

Contenidos de casillas. Si ha seleccionado los campos **Seleccionado** anteriormente, puede especificar el estadístico que se va a utilizar en las casillas de la matriz. Seleccione un estadístico basado en los recuentos o seleccione un campo de superposición para resumir los valores de un campo numérico según los valores de los campos de fila y columna.

- **Tabulaciones cruzadas.** Los valores de las casillas son recuentos y/o porcentajes de los recuentos con la combinación de valores correspondiente. Puede especificar qué resúmenes de tabulación cruzada desea utilizando las opciones de la pestaña Aspecto. El valor chi-cuadrado global es muestra junto con la significación. Consulte el tema "Explorador de resultados de nodo Matriz." en la página 304 para obtener más información.

- **Función.** Si selecciona una función de resumen, los valores de las casillas son una función de los valores del campo de superposición seleccionados para los casos con valores de fila y columna adecuados. Por ejemplo, si el campo de fila es *Región*, el campo de columna es *Producto* y el campo de superposición es *Ingresos*, entonces la casilla de la fila *Nordeste* y de la columna *Artilugio* contendrán la suma (o media, mínimo o máximo) de los ingresos provenientes de la venta de artilugios en la región nordeste. La función de resumen predeterminada es **Media**. Puede seleccionar otra función para resumir el campo de la función. Las opciones incluyen **Media**, **Suma**, **D.T.** (desviación estándar), **Máx** (máximo) y **Mín** (mínimo).

Pestaña Aspecto del nodo Matriz

La pestaña Aspecto permite controlar las opciones de ordenación y resaltado de la matriz, así como los estadísticos presentados para las matrices de tabulación cruzada.

Filas y columnas. Controla la ordenación de las cabeceras de fila y columna en la matriz. El valor predeterminado es **No ordenado**. Seleccione **Ascendente** o **Descendente** para ordenar las cabeceras de fila y columna en la dirección especificada.

Superponer. Permite resaltar los valores extremos de la matriz. Los valores se resaltan según los recuentos de las casillas (para matrices de tabulación cruzada) o los valores calculados (para matrices de función).

- **Resaltar mejores.** Puede solicitar que se resalten (en rojo) los valores más altos de la matriz. Especifique el número de valores que desee resaltar.
- **Resaltar peores.** Puede solicitar que se resalten (en verde) los valores más bajos de la matriz. Especifique el número de valores que desee resaltar.

Note: para las dos opciones de resaltado, los empates pueden hacer que se resalten más valores de los solicitados. Por ejemplo, si tiene una matriz con seis ceros en las casillas y solicita **Resaltar peores 5**, se resaltarán los seis ceros.

Contenido de las casillas de la tabulación cruzada. Para las tabulaciones cruzadas, puede especificar los estadísticos de resumen contenidos en la matriz para las matrices de tabulación cruzada. Estas opciones no están disponibles si se selecciona la opción **Todos los numéricos** o la opción **Función** en la pestaña Configuración.

- **Recuentos.** Las casillas incluyen el número de registros con el valor de la fila que tiene el valor de la columna correspondiente. Éste es el único contenido predeterminado de la casilla.
- **Valores esperados.** El valor esperado para el número de registros de la casilla, suponiendo que no haya relación entre las filas y las columnas. Los valores esperados se basan en la siguiente fórmula:

$$p(\text{valor de fila}) * p(\text{valor de columna}) * \text{número total de registros}$$
- **Residuos.** Diferencia entre los valores observados y esperados.
- **Porcentaje de fila.** Porcentaje de todos los registros con el valor de la fila que tiene el valor de la columna correspondiente. La suma de porcentajes es 100 en las filas.
- **Porcentaje de columna.** Porcentaje de todos los registros con el valor de la columna que tiene el valor de la fila correspondiente. La suma de porcentajes es 100 en las columnas.
- **Porcentaje de total.** Porcentaje de todos los registros con la combinación de valor de columna y valor de fila. La suma de porcentajes es 100 en toda la matriz.
- **Incluir totales de fila y columna.** Añade una fila y una columna a la matriz para los totales de columna y fila.
- **Aplicar configuración.** (Sólo en el explorador de resultados). Permite cambiar el aspecto de los resultados del nodo Matriz sin necesidad de cerrar y volver a abrir el explorador de resultados. Realice los cambios necesarios en esta pestaña del explorador de resultados, pulse en este botón y, a continuación, seleccione la pestaña Matriz para ver el efecto de los cambios.

Explorador de resultados de nodo Matriz.

El explorador de matrices muestra datos de tabulación cruzada y permite realizar operaciones en la matriz, incluida la selección de casillas, copia de la matriz en el Portapapeles en parte o en su totalidad, generación de nuevos nodos basados en la selección de la matriz y almacenamiento e impresión de la matriz. El explorador de matrices también se puede utilizar para mostrar resultados de determinados modelos, como los modelos de bayesiano ingenuo de Oracle.

Los menús Edición y Archivo proporcionan las opciones habituales para impresión, almacenamiento y exportación de resultados, así como selección y copia de datos. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Chi-cuadrado. Para una tabulación cruzada de dos campos categóricos, el chi-cuadrado de Pearson global aparece también bajo la tabla. Esta prueba indica la probabilidad de que los dos campos no estén relacionados, a partir de la diferencia entre los recuentos observados y los recuentos que esperaría si no existiera relación. Por ejemplo, si no existiera relación entre la satisfacción del cliente y la ubicación de la tienda, se esperarían unas tasas de satisfacción similares en todas las tiendas. Pero si los clientes de ciertas tiendas presentasen sistemáticamente tasas superiores a las de otras, se podría sospechar que no es coincidencia. A mayor diferencia, menor probabilidad de que el resultado sea un error de muestreo aleatorio solo.

- La prueba de chi-cuadrado indica la probabilidad de que dos campos no estén relacionados, en cuyo caso cualquier diferencia entre las frecuencias observadas y esperadas es el resultado del azar solo. Si esta probabilidad es muy pequeña (normalmente menor del 5%), la relación entre los dos campos se considera significativa.
- Si sólo hay una columna o una fila (una prueba de chi-cuadrado unidireccional), el número de grados de libertad es el número de casillas menos uno. En una chi-cuadrado bidireccional, el número de grados de libertad es el número de filas menos el número de columnas menos uno.
- Tenga cuidado al interpretar el estadístico chi cuadrado si alguna de las frecuencias de las casillas esperadas es menor de cinco.
- La prueba de chi-cuadrado sólo está disponible para una tabulación cruzada de dos campos. (Cuando en la pestaña Configuración se selecciona **Todas las marcas** o **Todos los numéricos**, esta prueba no se muestra.)

Menú Generar. El menú Generar contiene operaciones de generación de nodos. Estas operaciones sólo están disponibles para matrices de tabulación cruzada y debe haber al menos una casilla seleccionada en la matriz.

- **Nodo Seleccionar** Genera un nodo Seleccionar que selecciona los registros que coinciden con cualquier casilla de la matriz.
- **Nodo Derivar (Marca).** Genera un nodo Derivar para crear un nuevo campo de marca. El campo de marca contiene T para los registros que coinciden con cualquier casilla seleccionada en la matriz y F para los registros restantes.
- **Nodo Derivar (Conjunto).** Genera un nodo Derivar para crear un nuevo campo nominal. El campo nominal contiene una categoría para cada conjunto contiguo de casillas seleccionadas de la matriz.

Nodo Análisis

El nodo Análisis permite evaluar la capacidad de un modelo para generar predicciones precisas. Los nodos Análisis realizan varias comparaciones entre los valores predichos y los valores (su campo de objetivo) reales para uno o más nugget de modelo. Los nodos Análisis también se pueden utilizar para comparar modelos predictivos con otros modelos predictivos.

Cuando ejecuta un nodo Análisis, se añade automáticamente un resumen de los resultados de análisis a la sección Análisis de la pestaña Resumen para cada nugget de modelo de la ruta ejecutada. Los resultados de análisis detallados aparecen en la pestaña Resultados de la ventana del gestor o se pueden escribir directamente en un archivo.

Nota: dado que los nodos Análisis comparan valores predichos con valores reales, sólo son útiles con modelos supervisados (aquellos que requieren un campo de objetivo). Para los modelos sin supervisar como los algoritmos de clústeres, no existen resultados reales disponibles como base de comparación.

Pestaña Análisis del nodo Análisis

La pestaña Análisis permite especificar los detalles del análisis.

Matrices de coincidencias (para objetivos simbólicos o categóricos). Muestra el patrón de coincidencias entre cada campo generado (predicho) y su campo objetivo para objetivos categóricos (marca, nominal u ordinal). Se muestra una tabla con filas definidas por valores reales y columnas definidas por valores predichos, con el número de registros que tienen ese patrón en cada casilla. Esto es útil para identificar errores sistemáticos en las predicciones. Si existe más de un campo generado relacionado con el mismo campo de salida pero generado por modelos distintos, los casos en los que estos campos concuerdan y no concuerdan se cuentan y se muestran los totales. En los casos en los que concuerdan, se muestra otro conjunto de estadísticos correcto/incorrecto.

Evaluación del rendimiento. Muestra estadísticos de evaluación del rendimiento para modelos con resultados categóricos. Este estadístico, mostrado para cada categoría de los campos de salida, es una medida del contenido de información medio (en bits) del modelo para predecir registros pertenecientes a dicha categoría. Se tiene en cuenta la dificultad del problema de clasificación, de forma que las predicciones precisas para categorías inusuales obtendrán un índice de evaluación del rendimiento mayor que las predicciones precisas para categorías comunes. Si el modelo no hace más que adivinar una categoría, el índice de evaluación del rendimiento para esa categoría será 0.

Métricas de evaluación (AUC & Gini, clasificadores binarios solo). En el caso de los clasificadores binarios, esta opción informa de las métricas de evaluación de coeficiente Gini y AUC (Area Under Curve, área bajo la curva). Ambas métricas de evaluación se calculan de forma conjunta para cada modelo binario. Los valores de las métricas se notifican en una tabla en el navegador de salida de análisis.

La métrica de evaluación AUC se calcula como el área bajo una curva ROC (Receiver Operator Characteristic, característica de operador receptor) y es una representación escalar del rendimiento esperado de un clasificador. El AUC se sitúa siempre entre 0 y 1, y cuanto más alto es el valor, mejor es el clasificador. Una curva ROC diagonal entre las coordenadas (0,0) y (1,1) representa un clasificador aleatorio y tiene un AUC de 0,5. Así pues, un clasificador realista no tendrá un AUC de menos de 0,5.

La métrica de evaluación de coeficiente Gini se utiliza a veces como métrica de evaluación alternativa a la AUC, y ambas medidas están estrechamente relacionadas. El coeficiente Gini se calcula como dos veces el área comprendida entre la curva ROC y la diagonal, o como $Gini = 2AUC - 1$. El coeficiente Gini está siempre entre 0 y 1, y cuanto mayor es el número, mejor es el clasificador. El coeficiente Gini será negativo en el improbable caso de que la curva ROC esté por debajo de la diagonal.

Cifras de confianza (si están disponibles). Para los modelos que generan un campo de confianza, esta opción muestra estadísticos de los valores de confianza y su relación con las predicciones. Existen dos configuraciones para esta opción:

- **Umbral para.** Muestra el nivel de confianza por encima del cual la precisión será el porcentaje especificado.
- **Mejora en la precisión.** Muestra el nivel de confianza por encima del cual la precisión se mejorará en el valor especificado. Por ejemplo, si la precisión total es del 90% y esta opción se establece en 2,0, el valor mostrado será la confianza necesaria para una precisión del 95%.

Buscar campos predichos/predictores utilizando. Determina cómo se establece la correspondencia entre los campos predichos y el campo objetivo original.

- **Metadatos de campos de salida del modelo.** Establece la correspondencia entre los campos predichos y el objetivo basándose en información del campo Modelo, lo que permite coincidencias incluso si se

ha cambiado el nombre de un archivo predicho. También se puede acceder a la información del campo Modelo para cualquier campo predicho desde el cuadro de diálogo Valores mediante un nodo Tipo. Consulte el tema “Utilización del cuadro de diálogo de valores” en la página 132 para obtener más información.

- **Formato del nombre del campo.** Establece la correspondencia entre campos basándose en la convención de nomenclatura. Por ejemplo, los valores predichos generados por un nugget de modelo C5.0 para un objetivo denominado *respuesta* deben encontrarse en un campo denominado *\$C-respuesta*.

Separar por partición. Si se utiliza un campo de partición para dividir los registros en muestreos de entrenamiento, comprobación y validación, seleccione esta opción para mostrar los resultados de forma separada en cada partición. Consulte el tema “Nodo Partición” en la página 165 para obtener más información.

Nota: al separar en particiones, los registros con valores nulos en el campo de partición se excluyen del análisis. Esto nunca debería ser un problema si se utiliza un nodo Partición ya que este tipo de nodos no genera valores nulos.

Análisis definido por el usuario. Puede especificar su propio cálculo de análisis para utilizarlo al evaluar sus modelos. Utilice expresiones CLEM para especificar lo que debe calcularse para cada registro y cómo combinar las puntuaciones de nivel de registro en una puntuación global. Utilice las funciones @TARGET y @PREDICTED para hacer referencia al valor objetivo (resultado real) y al valor predicho, respectivamente.

- **Si.** Especifique una expresión condicional si necesita utilizar cálculos distintos dependiendo de alguna condición.
- **Entonces.** Especifique el cálculo si la condición Si es verdadera.
- **En caso contrario.** Especifique el cálculo si la condición Si es falsa.
- **Utilizar.** Seleccione un estadístico para calcular una puntuación global a partir de puntuaciones individuales.

Desglosar análisis por campos. Muestra los campos categóricos disponibles para desglosar el análisis. Además del análisis global, se mostrará un análisis independiente para cada categoría de campo de desglose.

Explorador de resultados de análisis

El explorador de resultados de análisis muestra los resultados de ejecutar un nodo Análisis. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Cuando se examina el resultado del análisis por primera vez, los resultados se expanden. Para ocultar los resultados cuando haya terminado de consultarlos, utilice el control de expansión situado a la izquierda del elemento para contraer los resultados específicos que desee ocultar o pulse en el botón **Contraer todo** para contraer todos los resultados. Para volver a ver los resultados tras contraerlos, utilice el control de expansión situado a la izquierda del elemento para mostrar los resultados o pulse en el botón **Expandir todo** para mostrar todos los resultados.

Resultados para el campo de salida. El resultado del análisis contiene una sección para cada campo de salida con un campo de predicción correspondiente creado por un modelo generado.

Comparando. Dentro de la sección del campo de salida hay una subsección para cada campo de predicción asociada con dicho campo de salida. Para campos de salida categóricos, el nivel superior de esta sección contiene una tabla que muestra el número y el porcentaje de predicciones correctas e incorrectas y el número total de registros en la ruta. Para los campos de salida numéricos, esta sección muestra la siguiente información:

- **Error mínimo.** Muestra el error mínimo (diferencia entre los valores observados y los predichos).

- **Error máximo.** Muestra el error máximo.
- **Error promedio.** Muestra la media (promedio) de errores en todos los registros. Esto indica si hay un **sesgo** sistemático (una tendencia mayor a sobrestimar que a subestimar, o viceversa) en el modelo.
- **Error absoluto promedio.** Muestra la media de valores absolutos de los errores en todos los registros. Indica la magnitud de error media, independientemente de la dirección.
- **Desviación estándar.** Muestra la desviación estándar de los errores.
- **Correlación lineal.** Muestra la correlación lineal entre los valores reales y los predichos. Esta estadística varía entre $-1,0$ y $1,0$. Los valores cercanos a $+1,0$ indican una fuerte asociación positiva, de forma que los valores predichos altos están asociados con valores reales altos y los valores predichos bajos están asociados con valores reales bajos. Los valores cercanos a $-1,0$ indican una fuerte asociación negativa, de forma que los valores predichos altos están asociados con valores reales bajos y viceversa. Los valores cercanos a $0,0$ indican una asociación débil, de forma que los valores predichos son más o menos independientes de los valores reales. *Note:* Una entrada en blanco aquí indica que la correlación lineal no se puede computar en este caso porque los valores reales o los valores predichos son constantes.
- **Ocurrencias.** Muestra el número de registros utilizados en el análisis.

Matriz de coincidencias. Para los campos de salida categóricos, si ha solicitado una matriz de coincidencias en las opciones de análisis, aparecerá una subsección aquí que contenga la matriz. Las filas representan valores observados reales y las columnas valores predichos. La casilla de la tabla indica el número de registros para cada combinación de valores reales y predichos.

Evaluación del rendimiento. Para los campos de salida categóricos, si ha solicitado estadísticos de evaluación del rendimiento en las opciones de análisis, los resultados de evaluación del rendimiento aparecerán aquí. Cada categoría de resultado aparece con su estadístico de evaluación del rendimiento.

Informe de valores de confianza. Para los campos de salida categóricos, si ha solicitado valores de confianza en las opciones de análisis, los valores aparecen aquí. Los siguientes estadísticos se muestran para los valores de confianza modelo:

- **Rango.** Muestra el rango (los valores máximos y mínimos) de valores de confianza para registros en los datos de la ruta.
- **Media para correctos.** Muestra la confianza media para los registros que se han clasificado correctamente.
- **Media para incorrectos.** Muestra la confianza media para los registros que se han clasificado de forma incorrecta.
- **Siempre correctos por encima de.** Muestra el umbral de confianza por encima del cual las predicciones son siempre correctas y el porcentaje de casos que cumplen este criterio.
- **Siempre incorrectos por debajo de.** Muestra el umbral de confianza por debajo del cual las predicciones son siempre incorrectas y muestra el porcentaje de casos que cumplen este criterio.
- **X% Precisión por encima de.** Muestra el nivel de confianza en el que la precisión es X%. X es aproximadamente el valor especificado para **Umbral para** en las opciones de Análisis. Para algunos modelos y conjuntos de datos, no se puede elegir un valor de confianza que ofrezca el umbral exacto especificado en las opciones (normalmente debido a los clústeres de casos similares con el mismo valor de confianza cerca del umbral). El umbral que se muestra es el valor más cercano al criterio de precisión especificado que se puede obtener con un solo umbral de valor de confianza.
- **X Veces correctas por encima de.** Muestra el valor de confianza en el cual la precisión es X veces mejor de lo que es para el conjunto de datos global. X es aproximadamente el valor especificado para **Mejora en la precisión** en las opciones de Análisis.

Acuerdo entre. Si dos o más modelos generados que predicen el mismo campo de salida se incluyen en la ruta, también verá estadísticos en el **acuerdo** entre predicciones generadas por los modelos. Esto incluye el número y el porcentaje de registros en los que concuerdan las predicciones (para campos de resultado categóricos) o estadísticos de resumen de errores (para campos de salida continuos). Para

campos categóricos, incluye un análisis de las predicciones comparadas con los valores reales para el subconjunto de registros en los que concuerdan los modelos (generan el mismo valor predicho).

Métricas de evaluación. En el caso de los clasificadores binarios, si se han solicitado métricas de evaluación en las opciones de análisis, se mostrarán las métricas de evaluación de coeficiente de Gini y AUC en una tabla de esta sección. La tabla tiene una fila por cada modelo clasificador binario. La tabla de métricas de evaluación se muestra por cada campo de salida y no por cada modelo.

Nodo Auditoría de datos

El nodo Auditoría de datos permite echar un primer vistazo exhaustivo a los datos introducidos en IBM SPSS Modeler, presentados en una matriz fácil de leer que se puede ordenar y utilizar para generar nodos de preparación de datos y gráficos de tamaño completo.

- La pestaña Auditar muestra un informe que proporciona estadísticos de resumen, histogramas y gráficos de distribución que pueden ser útiles para obtener una primera idea de los datos. El informe también muestra el icono de almacenamiento delante del nombre de campo.
- La pestaña Calidad del informe de auditoría muestra información sobre valores extremos, atípicos y perdidos, y proporciona herramientas para tratar dichos valores.

Uso del nodo Auditoría de datos

El nodo Auditoría de datos se puede conectar directamente a un nodo de origen o a nodos posteriores a un nodo Tipo instanciado. También puede generar varios nodos de preparación de datos en base a los resultados. Por ejemplo, puede generar un nodo Filtrar que excluya los campos con demasiados valores perdidos como para que sean útiles para el modelado, y generar un Supernodo que impute los valores perdidos de uno de los campos que permanecen o de todos ellos. Aquí es donde se demuestra la potencia real de la auditoría, que permite no sólo evaluar el estado actual de los datos, sino también realizar acciones en función de la evaluación.

Cribado o muestreo de datos. Dado que una auditoría inicial es especialmente efectiva cuando se trabaja con “datos de gran tamaño”, puede utilizarse un nodo Muestrear para reducir el tiempo de procesamiento durante la exploración inicial, mediante la selección de un solo subconjunto de registros. El nodo Auditoría de datos también se puede utilizar junto con nodos como Selección de características y Detección de anomalías en las fases de exploración del análisis.

Pestaña Configuración del nodo Auditoría de datos

La pestaña Configuración permite especificar los parámetros básicos para la auditoría.

Valor predeterminado. Puede conectar fácilmente el nodo a la ruta y pulsar en **Ejecutar** para generar un informe de auditoría para todos los campos con la configuración predeterminada, de la siguiente forma:

- Si no se ha configurado el nodo Tipo, todos los campos se incluyen en el informe.
- Si hay configuración del nodo Tipo (independientemente de si está o no instanciada) todos los campos *Entrada*, *Objetivo* y *Ambos* se incluyen en la representación. Si hay un solo campo *Objetivo*, se utiliza como campo de superposición. Si hay más de un campo *Objetivo* especificado, no se especifica ninguna superposición predeterminada.

Utilizar campos personalizados. Elija esta opción para seleccionar manualmente los campos. Utilice el botón de selección de campos de la derecha para seleccionar los campos de forma individual o por tipo.

Campo de superposición. El campo de superposición se emplea para dibujar los gráficos en miniatura que aparecen en el informe de auditoría. En los campos continuos (rango numérico), también se calculan los estadísticos bivariados (covarianza y correlación). Si se presenta un solo campo *Objetivo* basado en la

configuración del nodo Tipo, se utiliza como campo de superposición predeterminada, tal como se describe arriba. Si lo prefiere, puede seleccionar **Utilizar campos personalizados** para especificar una superposición.

Mostrar. Permite especificar si los gráficos van a estar disponibles en los resultados y seleccionar los estadísticos que se mostrarán de forma predeterminada.

- **Gráficos.** Muestra un gráfico para cada campo seleccionado; un gráfico de distribución (de barras), un histograma o un diagrama de dispersión, según los datos. Los gráficos aparecen en miniatura en el informe inicial, pero también se pueden generar gráficos a tamaño completo y nodos de gráficos. Consulte el tema “Explorador de resultados de auditoría de datos” en la página 310 para obtener más información.
- **Estadísticos básicos/avanzados.** Especifica el nivel de estadísticos que se muestran de forma predeterminada en los resultados. Esta configuración determina la representación inicial, ya que todos los estadísticos están disponibles en los resultados, independientemente de esta configuración. Consulte el tema “Mostrar estadísticos” en la página 311 para obtener más información.

Mediana y moda. Calcula la mediana y la moda para todos los campos del informe. Recuerde que cuando se trabaja con grandes conjuntos de datos, estos estadísticos pueden aumentar el tiempo de procesamiento, puesto que su cálculo requiere más tiempo. Sólo en el caso de la mediana, ocasionalmente el valor mostrado puede estar basado en una muestra de 2.000 registros (en lugar del conjunto de datos completo). Este muestreo se realiza por campos cuando, de otro modo, podrían excederse los límites de memoria. Cuando se realice el muestreo, los resultados se etiquetarán como tales en la salida (*Mediana de muestras* en lugar de sólo *Mediana*). El resto de estadísticos se calcula siempre utilizando el conjunto de datos completo.

Campos vacíos o sin tipo. Cuando se utilizan con datos instanciados, los campos sin tipo no se incluyen en el informe de auditoría. Para incluir los campos sin tipo (incluidos los campos vacíos), seleccione **Borrar todos los valores** en los nodos Tipo anteriores de la ruta. De esta forma se garantiza que los datos no estén instanciados, lo que hace que todos los campos se incluyan en el informe. Por ejemplo, puede resultar útil para obtener una lista completa de todos los campos o para generar un nodo Filtrar que excluya los campos vacíos. Consulte el tema “Filtrado de campos con datos perdidos” en la página 314 para obtener más información.

Pestaña Calidad de Auditoría de datos

La pestaña Calidad del nodo Auditoría de datos proporciona opciones para tratar los valores perdidos, atípicos y extremos.

Valores perdidos

- **Recuento de registros con valores válidos.** Seleccione esta opción para mostrar el número de registros con valores válidos para cada campo evaluado. Recuerde que los valores nulos (no definidos), los valores vacíos, los espacios en blanco y las cadenas vacías siempre se tratan como valores no válidos.
- **Desglosar recuentos de registros con valores no válidos.** Seleccione esta opción para mostrar el número de registros con cada tipo de valor no válido para cada campo.

Valores atípicos y extremos

Métodos de detección de valores atípicos y extremos. Se admiten dos métodos:

Desviación estándar de la media. Detecta los valores atípicos y extremos a partir del número de desviaciones estándar de la media. Por ejemplo, si tiene un campo con una media de 100 y una desviación estándar de 10, podría especificar 3,0 para indicar que cualquier valor inferior a 70 o superior a 130 debe tratarse como atípico.

Rango intercuartil. Detecta los valores atípicos y extremos a partir del rango intercuartil, que es el intervalo en el que caen los dos cuartiles centrales (entre los percentiles 25 y 75). Por ejemplo, basándonos en la configuración predeterminada de 1,5, el umbral inferior para los valores atípicos sería $Q1 - 1,5 * IQR$ y el umbral superior $Q3 + 1,5 * IQR$. Recuerde que el uso de esta opción puede ralentizar el rendimiento en grandes conjuntos de datos.

Explorador de resultados de auditoría de datos

El explorador de auditoría de datos es una potente herramienta que permite obtener una visión general de los datos. La pestaña Auditar muestra gráficos en miniatura, iconos de almacenamiento y estadísticos para todos los campos, mientras que la pestaña Calidad muestra información sobre los valores atípicos, extremos y perdidos. Basándose en los estadísticos de resumen y los gráficos iniciales, puede decidir recodificar un campo numérico, derivar un nuevo campo o volver a clasificar los valores de un campo nominal. O puede que desee explorar más utilizando una visualización más sofisticada. Puede hacerlo desde el explorador de informes de auditoría, utilizando el menú Generar para crear varios nodos que pueden emplearse para transformar o visualizar los datos.

- Ordenar columnas pulsando en la cabecera de la columna, o reordenarlas mediante la función de arrastrar y soltar. También se admiten la mayoría de las operaciones estándar de salida. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.
- Ver valores y rangos para los campos pulsando dos veces en un campo de las columnas Medición y Exclusivos.
- Utilice la barra de herramientas o el menú Edición para mostrar u ocultar etiquetas de valor, o seleccione los estadísticos que desea mostrar. Consulte el tema “Mostrar estadísticos” en la página 311 para obtener más información.
- Verifique los iconos de almacenamiento a la izquierda de los nombres de campo. Almacenamiento describe la forma en la que los datos se almacenan en un campo. Por ejemplo, un campo con valores 1 y 0 almacena datos enteros. Esto es distinto del nivel de medición, que describe el uso de los datos y no afecta al almacenamiento. Consulte el tema “Configuración de almacenamiento de campos y formato” en la página 9 para obtener más información.

Visualización y generación de gráficos

Si no hay ninguna superposición seleccionada, la pestaña Auditar muestra gráficos de barras (para campos nominales o marca) o histogramas (campos continuos).

Para la superposición de un campo nominal o marca, los gráficos se colorean con los valores de la superposición.

Para una superposición de campo nominal, se generan diagramas de dispersión bidimensionales en lugar de histogramas y barras unidimensionales. En ese caso, el eje x correlaciona al campo de superposición, permitiendo ver la misma escala en todos los ejes x conforme lee la tabla.

- Para los campos nominales o marca, mantenga el cursor del ratón sobre una barra para mostrar el valor subyacente o la etiqueta en un texto de información.
- Para los campos nominales o marca, utilice la barra de herramientas para cambiar la orientación de los gráficos en miniatura entre horizontal y vertical.
- Para generar un gráfico a tamaño completo a partir de una miniatura, pulse dos veces en la miniatura o seleccione la miniatura y elija **Resultado gráfico** en el menú Generar. *Note:* si hay un gráfico en miniatura basado en datos muestreados, el gráfico generado contendrá todos los casos si la ruta de datos original todavía está abierta.

Sólo puede generar un gráfico si el nodo Auditoría de datos que creó el resultado está conectado a la ruta.

- Para generar el correspondiente nodo de gráficos, seleccione uno o más campos en la pestaña Auditar y seleccione **Nodo Gráfico** en el menú Generar. El nodo resultante se añade al lienzo de rutas y se puede utilizar para volver a crear el gráfico cada vez que se ejecute la ruta.

- si un conjunto de superposiciones tiene más de 100 valores, aparece una advertencia y la superposición no se incluye.

Mostrar estadísticos

El cuadro de diálogo Mostrar estadísticos permite seleccionar los estadísticos que se muestran en la pestaña Auditar. La configuración inicial se especifica en el nodo Auditoría de datos. Consulte el tema "Pestaña Configuración del nodo Auditoría de datos" en la página 308 para obtener más información.

Mínimo. Se trata del valor menor de una variable numérica.

Máximo. Se trata del valor mayor de una variable numérica.

Suma. Suma o total de todos los valores, a lo largo de todos los casos con valores no perdidos.

Rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

Media. Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra, extraídas éstas de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).

desviación estándar. Una medida de la dispersión en torno a la media, igual a la raíz cuadrada de la varianza. La desviación estándar se mide en las mismas unidades que la variable original.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media dividida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

Error estándar de asimetría. La razón de la asimetría sobre su error estándar puede utilizarse como una prueba de normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda

Curtosis. Es una medida del grado en que las observaciones se agrupan en torno a un punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.

Error estándar de curtosis. La razón de la curtosis sobre su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución

normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

Exclusivo. Evalúa todos los efectos simultáneamente, corrigiendo cada efecto por todos los demás efectos de cualquier tipo.

Válidos. Los casos válidos que no tienen un valor perdido del sistema ni un valor definido como perdido del usuario. Tenga en cuenta que los valores nulos (no definidos), los valores vacíos, los espacios en blanco y las cadenas vacías siempre se tratan como valores no válidos.

Mediana. Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

Moda. El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.

Recuerde que la mediana y la moda se eliminan de forma predeterminada para mejorar el rendimiento, pero se pueden seleccionar en la pestaña Configuración del nodo Auditoría de datos. Consulte el tema "Pestaña Configuración del nodo Auditoría de datos" en la página 308 para obtener más información.

Estadísticos para superposiciones

Si se está utilizando un campo de superposición continuo (rango numérico), también están disponibles los siguientes estadísticos:

Covarianza. Medida no tipificada del grado de asociación entre dos variables igual a la desviación del producto vectorial dividido por $N-1$.

Pestaña Calidad del explorador de auditoría de datos

La pestaña Calidad del explorador de auditoría de datos muestra los resultados del análisis de calidad de los datos y permite especificar los tratamientos para los valores atípicos, extremos y perdidos.

Introducción de valores perdidos: El informe de auditoría muestra el porcentaje de registros completos para cada campo, junto con el número de valores válidos, nulos y vacíos. Puede seleccionar imputar valores perdidos para campos específicos, como considere oportuno y, a continuación, generar un Supernodo para aplicar estas transformaciones.

1. En la columna **Imputar perdidos**, especifique el tipo de valores que desea imputar, si los hay. Puede elegir entre imputar valores vacíos, nulos o ambos, o especificar una condición o expresión personalizada que seleccione los valores que se van a imputar.

Hay varios tipos de valores perdidos reconocidos por IBM SPSS Modeler:

- **Valores nulos o perdidos por el sistema.** Son valores que no son de cadena y que se han dejado en blanco en la base de datos o el archivo de origen y que no se han definido específicamente como "perdidos" en un nodo Tipo o de origen. Los valores perdidos del sistema se muestran como **\$null\$**. Tenga en cuenta que las cadenas vacías no se consideran nulas en IBM SPSS Modeler, aunque se pueden tratar como tales en algunas bases de datos.
- **Cadenas vacías y espacios en blanco.** Los valores de cadenas vacías y los espacios en blanco (cadenas con caracteres no visibles) se tratan como distintos de los valores nulos. Las cadenas vacías se tratan como equivalentes al espacio en blanco en la mayoría de los casos. Por ejemplo, si selecciona la opción de tratar los espacios en blanco como vacíos en un nodo Tipo o de origen, esta configuración se aplica a las cadenas vacías también.
- **Valores vacíos o perdidos definidos por el usuario.** Se trata de valores como desconocidos, 99 o -1, que se definen explícitamente como perdidos en el nodo Tipo o de origen. Si lo prefiere,

también puede elegir tratar los valores nulos o espacios en blanco como vacíos, lo que permite que se les marque para un tratamiento especial y que se excluyan de la mayoría de los cálculos. Por ejemplo, puede utilizar la función @BLANK para tratar estos valores junto con otros tipos de valores perdidos, como vacíos.

2. En la columna **Método**, especifique el método que desea utilizar.

Los siguientes métodos están disponibles para introducir valores perdidos:

Fija. Sustituye un valor fijo (ya sea la media de campo, el punto medio del rango o un constante que especifique).

Aleatorios. Sustituye un valor aleatorio basado en una distribución uniforme o normal.

Expresión. Permite especificar una expresión personalizada. Por ejemplo, podría sustituir los valores con una variable global creada por el nodo Val. globales.

Algoritmo. Sustituye un valor predicho por un modelo basado en el algoritmo C&RT. En cada campo imputado con este método, habrá un modelo C&RT independiente, junto con un nodo Rellenar que sustituye valores vacíos y nulos con el valor que predice el modelo. A continuación, se utiliza un nodo Filtrar para eliminar los campos de predicción generados por el modelo.

3. Para generar un Supernodo de valores perdidos, elija en los menús:

Generar > Supernodo de valores perdidos.

Se visualizará el cuadro de diálogo Supernodo de valores perdidos.

4. Seleccione **Todos los campos** o **Sólo campos seleccionados**, y especifique un tamaño de muestra si lo desea. (La muestra especificada es un porcentaje, de forma predeterminada se muestrea el 10 % de todos los registros).
5. Pulse en **Aceptar** para añadir el Supernodo generado al lienzo de rutas.
6. Conecte el Supernodo a la ruta para aplicar las transformaciones.

Dentro del Supernodo, se utiliza una combinación de nugget de modelo, Rellenar y Filtrar, según sea apropiado. Para comprender cómo funciona, puede editar el Supernodo y pulsar en **Acercar**, tras lo que podrá añadir, modificar o eliminar nodos específicos dentro del Supernodo para ajustar el comportamiento.

Tratamiento de los valores atípicos y extremos: El informe de auditoría muestra el número de valores atípicos y extremos para cada campo según las opciones de detección especificadas en el nodo Auditoría de datos. Consulte el tema “Pestaña Calidad de Auditoría de datos” en la página 309 para obtener más información. Puede seleccionar forzar, descartar o anular estos valores para campos específicos, según considere oportuno y, a continuación, generar un Supernodo para aplicar las transformaciones.

1. En la columna **Acción**, especifique cómo desea tratar los valores atípicos y extremos para campos específicos.

Hay disponibles las siguientes acciones para el tratamiento de valores atípicos y extremos:

- **Forzar.** Sustituye los valores atípicos y extremos por el valor más cercano que no se consideraría extremo. Por ejemplo, si un valor atípico se define como todo lo que queda por encima o por debajo de tres desviaciones estándar, todos los valores atípicos se sustituirían por el valor mayor o menor de este intervalo.
- **Descartar.** Descarta los registros con valores atípicos o extremos para el campo especificado.
- **Anular.** Sustituye los valores atípicos y extremos con valores nulos o perdidos por el sistema.
- **Forzar valores atípicos/descartar extremos.** Descarta sólo los valores extremos.
- **Forzar valores atípicos/anular extremos.** Anula sólo los valores extremos.

2. Para generar el Supernodo, seleccione en los menús:

Generar > Supernodo de valores atípicos y extremos

Se visualizará el cuadro de diálogo Supernodo de valores atípicos.

3. Seleccione **Todos los campos** o **Sólo campos seleccionados** y, a continuación, pulse en **Aceptar** para añadir el Supernodo generado al lienzo de rutas.

4. Conecte el Supernodo a la ruta para aplicar las transformaciones.

Si lo desea, puede editar el Supernodo y utilizar la opción de acercar para examinarlo o realizar cambios. Dentro del Supernodo, los valores se descartan, fuerzan o anulan mediante varios nodos Seleccionar o Rellenar, según proceda.

Filtrado de campos con datos perdidos: Desde el explorador de auditoría de datos, puede crear un nuevo nodo Filtrar según los resultados del análisis de calidad utilizando el cuadro de diálogo generar filtro desde calidad.

Moda. Seleccione la operación que desea para los campos especificados, ya sea **Incluir** o **Excluir**.

- **Campos seleccionados.** El nodo Filtrar incluirá o excluirá los campos seleccionados en la pestaña Calidad. Por ejemplo, puede ordenar la tabla por la columna % **Completo**. Para ello, mantenga pulsada la tecla Mayús mientras pulsa en los campos menos completos y, a continuación, genere un nodo Filtrar que excluya dichos campos.
- **Campos con un porcentaje de calidad superior a %.** El nodo Filtrar incluirá/excluirá campos en los que el porcentaje de registros completos sea mayor que el umbral especificado. El umbral predeterminado es 50%.

Filtrado de campos vacíos o sin tipo

Recuerde que una vez instanciados los valores de datos, los campos vacíos o sin tipo se excluyen de los resultados de auditoría y de la mayoría de los otros resultados de IBM SPSS Modeler. Estos campos se ignoran con fines de modelado, pero pueden recargar o hinchar los datos. Si es así, puede utilizar el explorador de auditoría de datos para generar un nodo Filtrar que elimine estos campos de la ruta.

1. Para asegurarse de que todos los campos se incluyen en la auditoría, incluidos los campos vacíos o sin tipo, pulse en **Borrar todos los valores** en el nodo Tipo o de origen anterior de la ruta, o establezca Valores en <Pasar> para todos los campos.
2. En el explorador de auditoría de datos, ordene por la columna % **Completo**, seleccione los campos que tienen cero valores válidos (u otro umbral) y utilice el menú Generar para generar un nodo Filtrar que se pueda añadir a la ruta.

Selección de registros con datos perdidos: Desde el explorador de auditoría de datos, puede crear un nuevo nodo Seleccionar según los resultados del análisis de calidad.

1. En el navegador Auditoría de datos, seleccione la pestaña Calidad.
2. Elija en el menú:

Generar > Nodo Seleccionar valores perdidos

Aparecerá el cuadro de diálogo Generar nodo Seleccionar.

Seleccionar si el registro es. Especifique si el registro debe mantenerse cuando es **Válido** o **No válido**.

Buscar valores no válidos en. Especifique dónde deben buscarse los valores no válidos.

- **Todos los campos.** El nodo Seleccionar buscará valores no válidos en todos los campos.
- **Campos seleccionados en la tabla.** El nodo Seleccionar comprobará sólo los campos seleccionados actualmente en la tabla de resultados de calidad.
- **Campos con un porcentaje de calidad superior a %.** El nodo Seleccionar comprobará los campos en los que el porcentaje de registros completos es mayor que el umbral especificado. El umbral predeterminado es 50%.

Un registro se considerará no válido si se encuentra un valor no válido en. Especifique la condición para identificar un registro no válido.

- **Cualquiera de los campos anteriores.** El nodo Seleccionar considerará un registro como no válido si *cualquiera* de los campos especificados anteriormente contiene un valor no válido para ese registro.

- **Todos los campos anteriores.** El nodo Seleccionar considerará un registro como no válido sólo si *todos* los campos especificados anteriormente contienen un valor no válido para ese registro.

Generación de otros nodos para la preparación de datos

Se pueden generar varios nodos utilizados en la preparación de datos directamente desde el explorador de auditoría de datos, incluidos los nodos Intervalos, Reclasificar y Derivar. Por ejemplo:

- Puede derivar un nuevo campo basado en los valores de *claimvalue* y *farmincome* seleccionando ambos en el informe de auditoría y seleccionando **Derivar** en el menú Generar. El nuevo nodo se añade al lienzo de rutas.
- De forma similar, puede determinar, según los resultados de auditoría, si recodificar *farmincome* en intervalos de percentiles ofrece un análisis más centrado. Para generar un nodo Intervalos, seleccione la fila de campo de la representación y seleccione **Intervalos** del menú Generar.

Una vez generado el nodo y añadido al lienzo de rutas, debe adjuntarlo a la ruta y abrirlo para especificar opciones para los campos seleccionados.

Nodo Transformación

La normalización de los campos de entrada es un importante paso previo a la utilización de las técnicas de puntuación tradicionales, como la regresión, la regresión logística y el análisis discriminante. Estas técnicas implican ciertos supuestos sobre las distribuciones normales de datos que pueden no ser ciertas para muchos archivos de datos sin procesar. Un método para trabajar con datos reales es aplicar transformaciones para convertir un elemento de los datos brutos hacia una distribución más normal. Además, los campos normalizados se pueden comparar fácilmente entre sí: por ejemplo, los ingresos y la edad son escalas completamente diferentes en un archivo de datos sin procesar pero, al normalizarse, se puede interpretar fácilmente el impacto relativo sobre cada uno.

El nodo Transformación proporciona un visor de resultados que permite realizar una rápida evaluación visual de la mejor transformación que se puede utilizar. Con sólo un vistazo, puede ver si las variables están distribuidas normalmente y, si es necesario, seleccionar la transformación necesaria y aplicarla. Puede elegir varios campos y realizar una transformación en cada uno.

Tras seleccionar las transformaciones preferidas para los campos, puede generar nodos Derivar o Rellenar para realizar las transformaciones y conectar estos nodos a la ruta. El nodo Derivar crea nuevos campos, mientras que el nodo Rellenar transforma los campos ya existentes. Consulte el tema “Generación de gráficos” en la página 317 para obtener más información.

Pestaña Campos del nodo Transformar

En la pestaña Campos, puede especificar los campos de los datos que desea utilizar para ver las posibles transformaciones y aplicarlas. Sólo es posible transformar los campos numéricos. Pulse en el botón selector de campos y seleccione uno o varios campos numéricos de la lista.

Pestaña Opciones del nodo Transformar

La pestaña Opciones le permite especificar el tipo de transformaciones que desea incluir. Puede elegir entre incluir todas las transformaciones disponibles o seleccionarlas individualmente.

En este último caso, también puede introducir un número con el fin de desplazar los datos para las transformaciones inversas y logarítmicas. Esto resulta útil en situaciones en las que una gran proporción de ceros en los datos sesgaría los resultados de la media y la desviación estándar.

Por ejemplo, supongamos que tiene un campo llamado *SALDO* que tiene algunos valores cero y desea utilizar la transformación inversa sobre él. Para evitar sesgos no deseados, tendría que seleccionar **Inversa (1/x)** y escribir un 1 en el campo **Usar desplazamiento de datos**. (Recuerde que este desplazamiento no está relacionado con el de la función de secuencia @0FFSET de IBM SPSS Modeler.)

Todas las fórmulas. Indica que todas las transformaciones disponibles deben calcularse y mostrarse en el resultado.

Fórmulas seleccionadas. Permite seleccionar las diferentes transformaciones que se van a calcular y mostrar en el resultado.

- **Inversa (1/x).** Indica que la transformación inversa debe mostrarse en el resultado.
- **Log (log n).** Indica que la transformación del \log_n debe mostrarse en el resultado.
- **Log (log 10).** Indica que la transformación del \log_{10} debe mostrarse en el resultado.
- **Exponencial.** Indica que la transformación exponencial (e^x) debe mostrarse en el resultado.
- **Raíz cuadrada.** Indica que la transformación de raíz cuadrada debe mostrarse en el resultado.

Pestaña Resultado del nodo Transformar

La pestaña Resultado permite especificar el formato y la ubicación del resultado. Puede elegir entre mostrar los resultados en pantalla o enviarlos a uno de los tipos de archivo estándar. Consulte el tema “Pestaña Resultado del nodo de resultados” en la página 300 para obtener más información.

Visor de resultados del nodo Transformar

El visor de resultados permite ver los resultados obtenidos tras ejecutar el nodo Transformar. El visor es una potente herramienta que muestra múltiples transformaciones por campo en vistas en miniatura de la transformación, lo que le permite comparar rápidamente los campos. Puede utilizar las opciones del menú Archivo para guardar, exportar o imprimir el resultado. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Para cada transformación (distinta a la seleccionada), se muestra una leyenda debajo con el formato:

Media (desviación estándar)

Generación de nodos para las transformaciones

El visor de resultados ofrece un punto de partida útil para la preparación de datos. Por ejemplo, puede desear normalizar el campo *EDAD* con el fin de utilizar una técnica de puntuación (como la regresión logística o el análisis discriminante) que suponga una distribución normal. Según los gráficos iniciales y a los estadísticos de resumen, es posible que decida transformar el campo *EDAD* de acuerdo con una distribución determinada (por ejemplo, log). Tras seleccionar la distribución deseada, puede generar un nodo Derivar con una transformación estandarizada para utilizarla para puntuar.

Desde el visor de resultados puede generar los siguientes nodos de operaciones con campos:

- Derivar
- Rellenar

El nodo Derivar crea nuevos campos con las transformaciones deseadas, mientras que el nodo Rellenar transforma los campos existentes. Los nodos se colocan en el lienzo en forma de un Supernodo.

Si selecciona la misma transformación para campos diferentes, un nodo Derivar o Rellenar contiene las fórmulas de dicho tipo de transformación para todos los campos a los que se aplica la transformación. Por ejemplo, supongamos que ha seleccionado los campos y transformaciones mostrados en la tabla siguiente para generar un nodo Derivar.

Tabla 55. Ejemplo de generación de nodo Derivar.

Campo	Transformación
<i>EDAD</i>	Distribución actual
<i>INGRESOS</i>	Log
<i>SALDO_PENDIENTE</i>	Inversa

Tabla 55. Ejemplo de generación de nodo Derivar (continuación).

SALDO	Inversa
-------	---------

El Supernodo contiene los siguientes nodos:

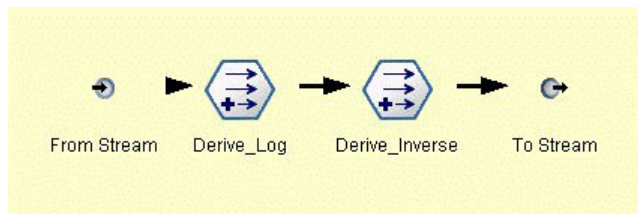


Figura 66. Supernodo en el lienzo

En este ejemplo, el nodo Derive_Log tiene la fórmula de log para el campo *INGRESOS* y el nodo Derive_Inverse tiene las fórmulas inversas para los campos *SALDO_PENDIENTE* y *SALDO*.

Generación de un nodo

1. Para cada campo del visor de resultados, seleccione la transformación deseada.
2. En el menú Generar, seleccione **Nodo Derivar** o **Nodo Rellenar** como desee.

Al hacerlo, se muestra el cuadro de diálogo Generar nodo Derivar o Generar nodo Rellenar, según corresponda.

Seleccione **Transformación no estandarizada** o **Transformación estandarizada (puntuación z)**, según sus preferencias. La segunda opción aplica una puntuación *z* a la transformación; las puntuaciones *z* representan valores como la distancia desde la media de la variable en desviaciones estándar. Por ejemplo, si aplica la transformación logarítmica al campo *EDAD* y selecciona una transformación estandarizada, la ecuación final para el nodo generado será:

$$(\log(\text{EDAD}) - \text{Media}) / \text{DT}$$

Una vez generado el nodo y representado en el lienzo de rutas:

1. Conéctelo con la ruta.
2. Para un Supernodo, si lo desea pulse dos veces en el nodo para ver su contenido.
3. Si lo prefiere, pulse dos veces en un nodo Derivar o Rellenar para modificar las opciones del campo seleccionado.

Generación de gráficos: Puede generar resultados como histogramas a tamaño completo a partir de los histogramas en miniatura del visor de resultados.

Generación de un gráfico

1. Pulse dos veces en un gráfico en miniatura del visor de resultados.

o

Seleccione un gráfico en miniatura del visor de resultados.

2. En el menú Generar, elija **Resultado gráfico**.

Al hacerlo, se muestra el histograma con una curva de distribución normal superpuesta, lo que le permite comparar en qué medida se ajusta cada transformación disponible a una distribución normal.

Note: Sólo puede generar un gráfico si el nodo Transformar que creó el resultado está conectado a la ruta.

Otras operaciones: Desde el visor de resultados, también puede:

- Ordenar la cuadrícula de salida por la columna Campo.

- Exportar los resultados a un archivo HTML. Consulte el tema “Exportación de resultados” en la página 298 para obtener más información.

Nodo Estadísticos

El nodo Estadísticos ofrece información de resumen básica acerca de los campos numéricos. Puede obtener estadísticos de resumen para campos individuales y correlaciones entre campos.

Pestaña Configuración del nodo Estadísticos

Examinar. Seleccione el campo o los campos para los que desee estadísticos de resumen individuales. Puede seleccionar varios campos.

Estadísticos. Seleccione los estadísticos para el informe. Las opciones disponibles incluyen **Recuento**, **Media**, **Suma**, **Mín**, **Máx**, **Rango**, **Varianza**, **Desv. estándar**, **Error estándar de la media**, **Mediana** y **Moda**.

Correlacionar. Seleccione el campo o los campos que desee correlacionar. Puede seleccionar varios campos. Cuando los campos de correlación están seleccionados, la correlación entre cada campo Examinar y los campos de correlación se mostrará en el resultado.

Configuración de correlación. Puede especificar las opciones de visualización de la fuerza de correlación de los resultados.

Configuración de correlación

IBM SPSS Modeler puede caracterizar correlaciones con etiquetas descriptivas para ayudarlo a resaltar relaciones importantes. La **correlación** mide la fuerza de la relación entre dos campos continuos (rango numérico). Toma valores entre -1.0 y 1.0 . Los valores cercanos a $+1.0$ indican una fuerte asociación positiva, de forma que los valores altos de un campo están asociados con los valores altos del otro y los valores bajos están asociados con los valores bajos. Los valores cercanos a -1.0 indican una asociación negativa fuerte, de forma que los valores altos para un campo están asociados con los valores bajos para el otro y viceversa. Los valores cercanos a $0,0$ indican una asociación débil, de forma que los valores para los dos campos son más o menos independientes.

Mediante el cuadro de diálogo Configuración de correlación, puede controlar la visualización de las etiquetas de correlación, cambiar los umbrales que definen las categorías y cambiar las etiquetas utilizadas para cada rango. Ya que la forma de caracterizar los valores de correlación depende en gran parte del dominio del problema, puede que desee personalizar los rangos y las etiquetas para que se ajusten a su situación en particular.

Mostrar etiquetas de fuerza de correlación en salida. Esta opción está seleccionada de forma predeterminada. Anule la selección para anular las etiquetas descriptivas del resultado.

Fuerza de correlación. Hay dos opciones para definir y etiquetar la fuerza de las correlaciones:

- **Definir fuerza de correlación por importancia (1-p).** Etiqueta las correlaciones según la importancia, que se define como 1 menos la significación o 1 menos la probabilidad de que la diferencia de las medias se pueda explicar por el azar solo. Cuanto más cerca está dicho valor a 1, mayor es la probabilidad de que los dos campos *no* sean independientes; es decir, que exista alguna relación entre ellos. El etiquetado de correlaciones por importancia se recomienda generalmente en vez del valor absoluto ya que tiene en cuenta la variabilidad de los datos; por ejemplo, un coeficiente de $0,6$ puede ser altamente significativo en un conjunto de datos y nada significativo en otro. De forma predeterminada, los valores de importancia $0,0$ y $0,9$ se etiquetan como *Débil*, los valores entre $0,9$ y $0,95$ se etiquetan como *Media* y entre $0,95$ y $1,0$ se etiquetan como *Fuerte*.
- **Definir fuerza de correlación por valor absoluto.** Etiqueta las correlaciones según el valor absoluto del coeficiente de correlación de Pearson, que varía entre -1 y 1 , como se ha descrito anteriormente. Cuanto más cerca de 1 está el valor de esta medida, más fuerte es la correlación. De forma

predeterminada, las correlaciones entre 0,0 y 0,3333 (en valores absolutos) se etiquetan como *Débil*, entre 0,3333 y 0,6666 se etiquetan como *Media* y entre 0,6666 y 1,0 se etiquetan como *Fuerte*. Tenga en cuenta, sin embargo, que la significación de cualquier valor es difícil de generalizar de un conjunto de datos a otro. Por ello, se recomienda en la mayoría de los casos definir correlaciones según la probabilidad en lugar de por el valor absoluto.

Explorador de resultados de estadísticos

El explorador de resultados del nodo Estadísticos muestra los resultados de análisis estadísticos y permite realizar operaciones, incluida la selección de campos, la generación de nuevos nodos basados en la selección y el almacenamiento y la impresión de resultados. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo y las opciones de edición normales están disponibles en el menú Edición. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Cuando se examina el resultado de estadísticos por primera vez, los resultados se expanden. Para ocultar los resultados cuando haya terminado de consultarlos, utilice el control de expansión situado a la izquierda del elemento para contraer los resultados específicos que desee ocultar o pulse en el botón **Contraer todo** para contraer todos los resultados. Para volver a ver los resultados tras contraerlos, utilice el control de expansión situado a la izquierda del elemento para mostrar los resultados o pulse en el botón **Expandir todo** para mostrar todos los resultados.

El resultado contiene una sección para cada campo *Examinar* que contiene una tabla de los estadísticos solicitados.

- **Recuento.** Número de registros con valores válidos para el campo.
- **Media.** Valor medio (promedio) para el campo en todos los registros.
- **Suma.** Suma de valores para el campo en todos los registros.
- **Mín.** Valor mínimo para el campo.
- **Máx.** Valor máximo para el campo.
- **Rango.** Diferencia entre los valores mínimos y máximos.
- **Varianza.** Medida de la variabilidad en los valores de un campo. Se calcula tomando la diferencia entre cada valor y la media global, calculando el cuadrado, sumando todos los valores y dividiendo por el número de registros.
- **Desviación estándar.** Otra medida de variabilidad en los valores de un campo, calculada como la raíz cuadrada de la varianza.
- **Error estándar de la media.** Medida de la incertidumbre en la estimación de la media del campo si se asume que la media se aplica a los datos nuevos.
- **Mediana.** El valor del "centro" para el campo, es decir, el valor que separa la mitad superior de los datos de la mitad inferior de los datos (según los valores del campo).
- **Moda.** El valor único más común de los datos.

Correlaciones. Si ha especificado campos correlacionados, el resultado también contiene una sección que muestra la correlación de Pearson entre el campo Examinar y cada campo correlacionado y etiquetas descriptivas opcionales para los valores de correlación. Consulte el tema “Configuración de correlación” en la página 318 para obtener más información.

Menú Generar. El menú Generar contiene operaciones de generación de nodos.

- **Filtro.** Genera un nodo Filtrar que filtra campos que no están correlacionados o que están débilmente correlacionados con otros campos.

Generación de un nodo Filtrar a partir de Estadísticos

El nodo Filtrar generado a partir del explorador de resultados de estadísticos filtrará los campos según sus correlaciones con otros campos. Funciona ordenando las correlaciones según el valor absoluto,

tomando las correlaciones mayores (según el criterio establecido en el cuadro de diálogo Generar filtro a partir de Estadísticos) y creando un filtro que pasa todos los campos que aparecen en cualquiera de esas correlaciones de gran tamaño.

Modo. Decida cómo seleccionar las correlaciones. **Incluir** hace que los campos que aparecen en las correlaciones especificadas se mantengan. **Excluir** hace que se filtren los campos.

Incluir/excluir campos que aparecen en. Definir el criterio para seleccionar correlaciones.

- **Número especificado de correlaciones.** Selecciona el número especificado de correlaciones e incluye/excluye los campos que aparecen en cualquiera de esas correlaciones.
- **Porcentaje especificado de correlaciones (%).** Selecciona el porcentaje especificado ($n\%$) de correlaciones e incluye/excluye los campos que aparecen en cualquiera de esas correlaciones.
- **Correlaciones mayores que.** Selecciona correlaciones con un valor absoluto mayor que el umbral especificado.

Nodo Medias

El nodo Medias compara las medias de grupos independientes o de pares de campos relacionados para probar si existen diferencias significativas. Por ejemplo, puede comparar los ingresos medios antes y después de llevar a cabo una promoción o comparar los ingresos de los clientes que no recibieron esa promoción con los que sí lo hicieron.

Puede comparar medias de dos maneras distintas, dependiendo de los datos:

- **Entre grupos dentro de un campo.** Para comparar grupos independientes, seleccione un campo de prueba y un campo de agrupación. Por ejemplo, se puede excluir una muestra de clientes "reservados" al enviar una promoción y comparar los ingresos medios de ese grupo con los del resto. En este caso, podría especificar un campo de prueba único que indique el ingreso de cada cliente, con un campo nominal o marca que indique si recibieron la oferta. Las muestras son independientes en el sentido de que cada registro se asigna a un grupo o a otro, y no hay manera de vincular un miembro específico de un grupo con un miembro específico de otro. También puede especificar un campo nominal con más de dos valores para comparar las medias de varios grupos. Cuando se ejecuta, el nodo calcula una prueba ANOVA unidireccional de los campos seleccionados. En los casos en que sólo hay dos grupos de campos, los resultados de ANOVA unidireccional son básicamente los mismos que los de una prueba t de muestras independientes. Consulte el tema "Comparación de medias para grupos independientes" para obtener más información.
- **Entre pares de campos.** Al comparar las medias de dos campos relacionados, los grupos deben estar relacionados de alguna manera para que los resultados tengan sentido. Por ejemplo, podría comparar los ingresos medios de un mismo grupo de clientes antes y después de llevar a cabo una promoción o comparar las tasas de uso de un servicio entre pares marido-mujer para ver si son diferentes. Cada registro contiene dos medidas diferentes pero relacionadas que se pueden comparar de manera lógica. Cuando se ejecuta, el nodo calcula una prueba t de muestras relacionadas en cada par de campos seleccionado. Consulte el tema "Comparación de medias entre campos relacionados" en la página 321 para obtener más información.

Comparación de medias para grupos independientes

Seleccione **Entre grupos de un campo** en el nodo Medias para comparar la media de dos o más grupos independientes.

Campo de agrupación. Seleccione una marca numérica o un campo nominal con dos o más valores distintos que divida los registros en los grupos que se quieren comparar, por ejemplo, los que recibieron la oferta frente a los que no la recibieron. Independientemente del número de campos de prueba, sólo se puede seleccionar un campo de agrupamiento.

Probar campos. Seleccione uno o más campos numéricos que contengan las medidas que desea analizar. Se realizará una prueba diferente para cada campo seleccionado. Por ejemplo, podría analizar el impacto de una promoción determinada sobre el uso, ingresos y abandono de clientes.

Comparación de medias entre campos relacionados

Seleccionar **Entre pares de campos** en el nodo Medias para comparar medias entre campos distintos. Los campos deben estar relacionados de alguna manera para que los resultados tengan sentido, como los ingresos antes y después de una promoción. También se pueden seleccionar varios pares de campos.

Campo uno. Seleccione un campo numérico que contenga la primera medida que se desea comparar. En un estudio de antes y después, este sería el campo "antes".

Campo dos. Seleccione el segundo campo que desea comparar.

Añadir. Añade el par seleccionado a la lista Comprobar pares de campos.

Repita la selección de campos según sea necesario para añadir varios pares a la lista.

Configuración de correlación. Permite especificar las opciones para etiquetar la fuerza de las correlaciones. Consulte el tema "Configuración de correlación" en la página 318 para obtener más información.

Opciones del nodo Medias

La pestaña Opciones permite establecer el umbral de los valores p que se utiliza para etiquetar los resultados como importantes, secundarios o sin importancia. También puede editar la etiqueta de cada clasificación. La importancia se mide en una escala de porcentaje y se puede definir en líneas generales como 1 menos la probabilidad de obtener un resultado (como la diferencia en las medias de dos campos) tan extremo o más que el resultado observado debido únicamente al azar. Por ejemplo, un valor p superior a 0,95 indica menos de un 5% de probabilidad de que el resultado se pueda explicar únicamente por el azar.

Etiquetas de importancia. Puede editar las etiquetas utilizadas para etiquetar cada par de campos o grupo del resultado. Las etiquetas predeterminadas son *importante*, *secundario* y *sin importancia*.

Valores de corte. Especifica el umbral para cada rango. Normalmente, los valores p superiores a 0,95 se considerarían como importantes, mientras que los valores inferiores a 0,9 se considerarían como sin importancia, pero estos umbrales se pueden ajustar según sea necesario.

Nota: las medidas de importancia están disponibles en una serie de nodos. Los cálculos específicos dependen del nodo y del tipo de objetivo y campos de entrada utilizados, aunque los valores se pueden seguir comparando ya que todos se miden en una escala de porcentajes.

Explorador de resultados del nodo Medias

El explorador de resultados de medias muestra datos de tabulación cruzada y permite realizar operaciones estándar, incluyendo seleccionar y copiar la tabla fila por fila, ordenar por cualquier columna y guardar e imprimir la tabla. Consulte el tema "Visualización de resultados" en la página 296 para obtener más información.

La información específica de la tabla depende del tipo de comparación (grupos de un campo o diferentes campos).

Ordenar por. Permite ordenar los resultados por una columna específica. Pulse en la flecha de arriba o abajo para cambiar la dirección del orden. También puede pulsar en cualquier cabecera de columna para ordenar por esa columna. (Para cambiar la dirección del orden dentro de esa columna, vuelva a pulsar.)

Ver. Puede elegir **Simple** o **Avanzado** para controlar el nivel de detalle de la visualización. La vista avanzada incluye toda la información de la vista simple pero con detalles adicionales.

Resultados de las medias al comparar los grupos de un campo

Al comparar los grupos de un campo, el nombre del campo de agrupamiento aparece encima de la tabla de resultados y se presentan las medias y los estadísticos relacionados de forma independiente para cada grupo. La tabla incluye una fila independiente para cada campo de prueba.

Se muestran las siguientes columnas:

- **Campo.** Muestra los nombres de los campos de prueba seleccionados.
- **Medias por grupo.** Muestra la media para cada categoría del campo de agrupamiento. Por ejemplo, puede comparar los grupos que recibieron una oferta especial (*Nueva promoción*) con aquellos que no lo hicieron (*Estándar*). En la vista avanzada también aparecen la desviación estándar, el error estándar y el recuento.
- **Importancia.** Muestra la etiqueta y el valor de importancia. Consulte el tema “Opciones del nodo Medias” en la página 321 para obtener más información.

Salida avanzada

En la vista avanzada, se muestran las siguientes columnas adicionales.

- **Prueba F.** Esta prueba se basa en el cociente de la varianza entre los grupos y la varianza de cada grupo. Si las medias son las mismas para todos los grupos, se esperaría una razón *F* cercana a 1 ya que son estimaciones de la misma varianza poblacional. Cuanto mayor es dicha razón, mayor es la variación entre grupos y mayor la probabilidad de que exista una diferencia significativa.
- **gl.** Muestra los grados de libertad.

Resultados de las medias al comparar pares de campos

Al comparar campos diferentes, la tabla de resultados incluye una fila para cada par de campos seleccionado.

- **Campo uno/dos.** Muestra el nombre del primer y segundo campo de cada par. En la vista avanzada también aparecen la desviación estándar, el error estándar y el recuento.
- **Media uno/dos.** Muestra la media de cada campo, respectivamente.
- **Correlación.** Mide la fuerza de la relación entre dos campos continuos (rango numérico). Los valores cercanos a +1,0 indican una fuerte asociación positiva y los valores cercanos a -1,0 indican una fuerte asociación negativa. Consulte el tema “Configuración de correlación” en la página 318 para obtener más información.
- **Diferencia de medias.** Muestra la diferencia entre las medias de los dos campos.
- **Importancia.** Muestra la etiqueta y el valor de importancia. Consulte el tema “Opciones del nodo Medias” en la página 321 para obtener más información.

Salida avanzada

Un resultado avanzado añade las siguientes columnas:

Intervalo de confianza de 95%. Límites inferior y superior del intervalo en el que la media verdadera se encuentra dentro del 95% de todas las muestras posibles de este tamaño a partir de esta población.

Prueba T. El estadístico *t* se obtiene dividiendo la diferencia de medias por su error estándar. Cuanto mayor es el valor absoluto de este estadístico, mayor es la probabilidad de que las medias no sean iguales.

gl. Muestra los grados de libertad del estadístico.

Nodo Informe

El nodo Informe permite crear informes con formato que contienen texto fijo, así como datos y otras expresiones derivadas de los datos. Puede especificar el formato del informe utilizando plantillas de texto para definir el texto fijo y las construcciones de resultados de datos. Puede proporcionar formato de texto personalizado utilizando etiquetas HTML de la plantilla y configurando opciones en la pestaña Resultado. Los valores de los datos y otros resultados condicionales se incluyen en el informe utilizando expresiones CLEM de la plantilla.

Alternativas al nodo Informe

El nodo Informe se suele utilizar para enumerar registros o resultados de casos de una ruta, como todos los registros que cumplan una condición determinada. A este respecto, se puede considerar como una alternativa menos estructurada que el nodo Tabla.

- Si desea obtener un informe que enumere la información de campo o cualquier otro dato que esté definido en la ruta en lugar de los propios datos (como definiciones de campo especificadas en un nodo Tipo) puede utilizar un script en su lugar.
- Para generar un informe que incluya varios objetos de resultados (como una colección de modelos, tablas y gráficos generados por una o más rutas) y que se pueden mostrar en varios formatos (incluido texto, HTML y Microsoft Word/Office) se puede utilizar un proyecto de IBM SPSS Modeler.
- Para generar una lista de nombres de campos sin utilizar scripts, puede utilizar un nodo Tabla precedido por un nodo Muestrear que descarte todos los registros. Así, se producirá una tabla sin filas, que se puede transponer al exportarla para generar una lista de nombres de campos en una única columna. (Seleccione **Transponer datos** en la pestaña Resultado del nodo Tabla para realizar esta acción.)

Pestaña Plantilla del nodo Informe

Creación de una plantilla. Para definir el contenido del informe, cree una plantilla en la pestaña Plantilla del nodo Informe. La plantilla consta de líneas de texto, cada una de las cuales especifica algo sobre el contenido del informe y algunas líneas de etiqueta especiales utilizadas para indicar el ámbito de las líneas de contenido. Con cada línea de contenido, las expresiones CLEM entre corchetes ([]) se evalúan antes de que la línea se envíe al informe. Existen tres ámbitos posibles para una línea en la plantilla:

Fija. Las líneas que no están marcadas se consideran fijas. Las líneas fijas se copian en el informe sólo una vez, después de la evaluación de cualquier expresión que contengan. Por ejemplo, la línea Este informe fue generado: [@TODAY]

copiará una sola línea que contiene el texto y la fecha actual en el informe.

Globales (iterar ALL). Las líneas que se encuentran entre las etiquetas especiales #ALL y # se copian en el informe una vez para cada registro de datos de entrada. Las expresiones de CLEM (entre paréntesis) se evalúan según el registro actual para cada línea de resultado. Por ejemplo, las líneas

```
#ALL
Para el registro [@INDEX], el valor de AGE es [AGE]
#
```

incluiría una línea para cada registro indicando el número del registro y la edad.

Para generar una lista de todos los registros:

```
#ALL
[Age] [Sex] [Cholesterol] [BP]
#
```

Condicional (iterar WHERE). Las líneas que se encuentran entre las etiquetas especiales #WHERE <condición> y # se copian en el informe una vez para cada registro en el que se cumple la condición especificada. La condición es una expresión CLEM. (En la condición WHERE, los paréntesis son opcionales.) Por ejemplo, las líneas

```
#WHERE [SEX = 'M']
#WHERE [SEX = 'M'] El hombre del registro nº [@INDEX] tiene [AGE] años.
#
```

escribirá una línea en el archivo para cada registro con un valor *M* para sexo. El informe completo contendrá las líneas fijas, globales y condicionales definidas al aplicar la plantilla a los datos de entrada.

Puede especificar opciones para mostrar o guardar los resultados mediante la pestaña Resultado, común a varios tipos de nodos de resultados. Consulte el tema “Pestaña Resultado del nodo de resultados” en la página 300 para obtener más información.

Salida de datos en formato XML o HTML

Puede incluir etiquetas XML o HTML directamente en la plantilla para escribir informes en cualquiera de estos formatos. Por ejemplo, la siguiente plantilla genera una tabla HTML.

Este informe está escrito en HTML.

Sólo se incluyen los registros donde la edad es superior a 60.

```
<HTML>
  <TABLE border="2">
    <TR>
      <TD>Edad</TD>
      <TD>PS</TD>
      <TD>Colesterol</TD>
      <TD>Medicamento</TD>
    </TR>

    #WHERE Age > 60
    <TR>
      <TD>[Edad]</TD>
      <TD>[PS]</TD>
      <TD>[Colesterol]</TD>
      <TD>[Medicamento]</TD>
    </TR>
  #
</TABLE>
</HTML>
```

Explorador de resultados del nodo Informe

El explorador de informes muestra el contenido del informe generado. Las opciones normales de almacenamiento, exportación e impresión están disponibles en el menú Archivo y las opciones de edición normales están disponibles en el menú Edición. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Nodo Val. globales

El nodo Val. globales explora los datos y calcula los valores de resumen que se pueden utilizar en expresiones CLEM. Por ejemplo, puede utilizar un nodo Val. globales para calcular estadísticos para un campo denominado *edad* y, a continuación, utilizar la media global de *edad* en expresiones CLEM insertando la función @GLOBAL_MEAN(*edad*).

Pestaña Configuración del nodo Val. globales

Valores globales que se van a crear. Seleccione el campo o los campos para los que desea que haya valores globales disponibles. Puede seleccionar varios campos. Para cada campo, especifique los estadísticos para calcular asegurándose de que los estadísticos que desea están seleccionados en las columnas junto al nombre del campo.

- **MEDIA.** Valor medio (promedio) para el campo en todos los registros.
- **SUMA.** Suma de valores para el campo en todos los registros.
- **MÍN.** Valor mínimo para el campo.
- **MÁX.** Valor máximo para el campo.
- **D.T.** La desviación estándar, medida de variabilidad en los valores de un campo, se calcula como la raíz cuadrada de la varianza.

Operaciones predeterminadas. Las opciones seleccionadas aquí se utilizarán cuando se añadan nuevos campos a la lista de valores globales anterior. Para cambiar el conjunto de estadísticos predeterminado, seleccione o anule la selección de los estadísticos de la forma adecuada. También puede utilizar el botón **Aplicar** para aplicar las operaciones predeterminadas a todos los campos de la lista.

Nota: Algunas operaciones no son aplicables a campos no numéricos (por ejemplo, Sum para un campo de fecha/hora). Las operaciones que no se pueden utilizar con un campo seleccionado se inhabilitan

Borrar todos los valores globales antes de la ejecución. Seleccione esta opción para quitar todos los valores globales antes de calcular los valores nuevos. Si esta opción no está seleccionada, los valores recién calculados sustituirán a los valores antiguos, pero los valores globales que no se han vuelto a calcular permanecerán, también, disponibles.

Mostrar presentación preliminar de los valores globales creados después de la ejecución. Si selecciona esta opción, la pestaña Globales del cuadro de diálogo de propiedades de la ruta aparecerá tras la ejecución para mostrar los valores globales calculados.

Nodo Ajustar simulación

El nodo Ajustar simulación ajusta un conjunto de distribuciones estadísticas candidatas a cada campo de los datos. El ajuste de cada distribución a un campo se evalúa mediante un criterio de bondad de ajuste. Cuando se ejecuta un nodo Ajustar simulación, se construye un nodo Generar simulación (o se actualiza un nodo existente). A cada campo se le asigna la distribución que mejor se le ajusta. El nodo Generar simulación podrá entonces utilizarse para generar datos simulados para cada campo.

Aunque el nodo Ajustar simulación es un nodo terminal, no añade un modelo a la paleta de modelos generados ni añade una salida o gráfica a la pestaña de resultados, ni exporta datos.

Nota: Si los datos históricos están dispersos (es decir, hay muchos datos ausentes), puede que al componente de ajuste le resulte difícil encontrar suficientes valores válidos para ajustar las distribuciones a los datos. En aquellos casos en los que los datos estén dispersos, antes de procederse con el ajuste deberán eliminarse los campos dispersos si no son necesarios o imputar los valores que faltan. Mediante las opciones de la pestaña **Calidad** del nodo Auditoría de datos, puede verse el número de registros completos, identificarse los campos dispersos y seleccionarse un método de imputación. Si el número de registros es insuficiente para ajustar la distribución, podrá utilizarse un nodo Equilibrar para aumentar el número de registros.

Utilización de un nodo Ajustar simulación para crear automáticamente un nodo Generar simulación

La primera vez que se ejecuta el nodo Ajustar simulación, se crea un nodo Generar simulación con un enlace de actualización al nodo Ajustar simulación. Si se vuelve a ejecutar el nodo Ajustar simulación, solo se creará un nodo Generar simulación si se ha eliminado el enlace de actualización. Un nodo Ajustar

simulación también puede utilizarse para actualizar un nodo Generar simulación conectado. El resultado dependerá de si los mismos campos están presentes en ambos nodos y de si los campos están desbloqueados en el nodo Generación de simulación. Consulte el tema “Nodo Generar simulación” en la página 49 para obtener más información.

Un nodo Ajustar simulación solo puede tener un enlace de actualización a un nodo Generar simulación. Para definir un enlace de actualización a un nodo Generar simulación, siga estos pasos:

1. Pulse con el botón derecho en el nodo Ajustar simulación.
2. En el menú, seleccione **Definir enlace de actualización**.
3. Pulse en el nodo Generar simulación al que desee definir un enlace de actualización.

Para eliminar un enlace de actualización entre un nodo Ajustar simulación y un nodo Generar simulación, pulse con el botón derecho en el enlace de actualización y seleccione **Eliminar enlace**.

Ajuste de distribuciones

Una distribución estadística es la frecuencia teórica de la ocurrencia de los valores que puede tomar una variable. En el nodo Ajustar simulación, se compara un conjunto de distribuciones estadísticas teóricas con cada uno de los campos de datos. Las distribuciones disponibles al ajuste se describen en el tema “Distribuciones” en la página 59. Los parámetros de la distribución teórica se ajustan para dar el mejor ajuste a los datos conforme una medición de la bondad del ajuste, ya sea el criterio de Anderson-Darling o el criterio de Kolmogorov-Smirnov. El resultado del ajuste de distribución llevado cabo por el nodo Ajustar simulación muestra qué distribuciones se han ajustado, las mejores estimaciones de parámetros para cada distribución y en qué medida la distribución se ajusta a los datos. Durante el ajuste de distribución, también se calculan las correlaciones entre los campos con tipos de almacenamiento numérico y las contingencias entre campos con una distribución categórica. Los resultados del ajuste de distribución se utilizan para crear un nodo Generar simulación.

Antes de que ajustarse las distribuciones a los datos, se examinan los primeros 1000 registros en busca de datos ausentes. Si faltan demasiados valores, el ajuste de distribuciones no será posible. En tal caso, deberá decidir cuál de las opciones siguientes procede:

- Utilizar un nodo anterior en la ruta para eliminar los registros a los que les falten valores.
- Utilizar un nodo anterior en la ruta para asignar los valores que falten.

El ajuste de distribuciones no excluye los valores de usuario ausente. Si los datos tienen valores de usuario ausente y se desea excluirlos del ajuste de distribuciones, dichos valores deberían establecerse a valores perdidos del sistema.

El rol de un campo no se tiene en cuenta cuando se ajustan las distribuciones. Por ejemplo, los campos con el rol **Destino** se tratan del igual que los campos con los roles de **Entrada**, **Ninguno**, **Ambos**, **Partición División**, **Frecuencia** e **ID**.

Los campos se tratan de forma diferente durante el ajuste de distribuciones según su tipo de almacenamiento y nivel de medición. El tratamiento de los campos durante el ajuste de distribuciones se describen en la tabla siguiente.

Tabla 56. Ajuste de distribuciones en función del tipo de almacenamiento y del nivel de medición de los campos

Tipo de almacenamiento	Nivel de medición					
	Continuo	Categorico	Distintivo	Nominal	Ordinal	Sin tipo
Cadena	Imposible		Se ajustan las distribuciones categórica, dice y fija.			

Tabla 56. Ajuste de distribuciones en función del tipo de almacenamiento y del nivel de medición de los campos (continuación)

Tipo de almacenamiento	Nivel de medición			
Entero	Se ajustan todas las distribuciones. Se calculan las correlaciones y las contingencias.	Se ajusta la distribución categórica. No se calculan las correlaciones.	Se ajustan las distribuciones binomial, binomial negativa y Poisson, y se calculan las correlaciones.	El campo se ignora y no se pasa al nodo Generar simulación.
Real				
Hora				
Fecha				
Marca de tiempo				
Desconocido	El tipo de almacenamiento correspondiente se determina a partir de los datos.			

Los campos con el nivel de medición ordinal se tratan como campos continuos y se incluyen en la tabla de correlaciones del nodo Generar simulación. Si se desea ajustar una distribución distinta de la binomial, binomial negativa o Poisson a un campo ordinal, deberá cambiarse a continuo el nivel de medición del campo. Si se ha definido previamente una etiqueta para cada valor de un campo ordinal y luego se cambia el nivel de medición a continuo, se perderán las etiquetas.

Durante el ajuste de distribución, los campos que tienen valores únicos no se tratan de forma distinta de los campos con varios valores. Los campos con tipo de almacenamiento hora, fecha o fecha y hora se tratan como numéricos.

Ajuste de distribuciones a campos de división

Si los datos contienen un campo de división y se desea que el ajuste de distribución se lleve a cabo de forma independiente para cada división, deberán transformarse los datos utilizando un nodo Reestructurar anterior en la ruta. Mediante un nodo Reestructurar, genere un campo nuevo para cada valor del campo de división. Estos datos reestructurados podrán utilizarse luego en un ajuste de distribución en el nodo Ajustar simulación.

Pestaña Valores del nodo Ajustar simulación

Nombre del nodo origen. Puede generarse de forma automática el nombre del nodo generado (o actualizado) Generar simulación seleccionando **Auto**. El nombre generado automáticamente es el nombre del nodo Ajustar simulación si se ha especificado un nombre personalizado (o Sim Gen si no se ha especificado ningún nombre personalizado en el nodo Ajustar simulación). Seleccione **Personalizado** para especificar un nombre personalizado en el campo de texto adyacente. A menos que se edite el campo de texto, el nombre personalizado predeterminado será Sim Gen.

Opciones de Ajuste Con estas opciones se puede especificar cómo se ajustan las distribuciones a los campos y cómo se evalúa el ajuste de las distribuciones.

- **Número de casos de la muestra.** Especifica el número de casos a utilizar cuando se ajustan las distribuciones a los campos del conjunto de datos. Seleccione **Todos los casos** para ajustar las distribuciones a todos los registros de los datos. Si el conjunto de datos es muy grande, puede que le interese limitar el número de casos utilizados en el ajuste de la distribución. Seleccione **Limitar a los N primeros casos** para utilizar solo los N primeros casos. Pulse en las flechas para especificar el número de casos a utilizar. De forma alternativa, puede utilizarse un nodo anterior en la ruta para tomar una muestra aleatoria de registros para el ajuste de la distribución.
- **Bondad de los criterios de ajuste (solo campos continuos).** En el caso de los campos continuos, seleccione la prueba de bondad de ajuste de Anderson-Darling o la prueba de Kolmogorov-Smirnoff para clasificar las distribuciones cuando se ajusten distribuciones a los campos. La prueba Anderson-Darling está seleccionada de forma predeterminada y se recomienda especialmente cuando se desea asegurar el mejor ajuste posible en las regiones de los extremos. Ambas estadísticas se calculan para cada distribución candidata, pero solo se utilizará la estadística seleccionada para ordenar las distribuciones y determinar la distribución que mejor se ajusta.
- **Intervalos (solo distribuciones empíricas)** En el caso de los campos continuos, la distribución Empírica es la función de distribución acumulada de los datos históricos. Es la probabilidad de cada valor o rango de valores y se deriva directamente de los datos. Puede especificarse el número de intervalos utilizados para calcular la distribución Empírica de los campos continuos pulsando en las flechas. El valor predeterminado es 100 y el máximo es 1.000.
- **Campo de ponderación (opcional).** Si el conjunto de datos contiene un campo de ponderación, pulse en el icono selector de campo y seleccione el campo de ponderación en la lista. El campo de ponderación se excluye luego del proceso de ajuste de la distribución. La lista muestra todos los campos del conjunto de datos que tienen el nivel de medición continuo. Solo se puede seleccionar un campo de ponderación.

Nodo Evaluación de simulación

El nodo Evaluación de simulación es un nodo terminal que evalúa un campo especificado, proporciona una distribución del campo y produce gráficas de distribuciones y correlaciones. Este nodo se utiliza principalmente para evaluar campos continuos. Por lo tanto, complementa la gráfica de evaluación, generada por un nodo Evaluación y útil para evaluar los campos discretos. Otra diferencia es que el nodo Evaluación de simulación evalúa una única predicción en varias iteraciones, mientras que el nodo Evaluación evalúa varias predicciones, cada una con una única iteración. Las iteraciones se generan cuando se especifica más de un valor para un parámetro de distribución en el nodo Generar simulación. Consulte el tema “Iteraciones” en la página 59 para obtener más información.

El nodo Evaluación de simulación está diseñado para ser utilizado con datos obtenidos de los nodos Ajustar simulación y Generar simulación. No obstante, el nodo se puede utilizar con cualquier otro nodo. Puede colocarse cualquier número de pasos de procesamiento entre el nodo Generar simulación y el nodo Evaluación de simulación.

Importante: El nodo Evaluación de simulación requiere un mínimo de 1000 registros con valores válidos para el campo de destino.

Pestaña Configuración del nodo Evaluación de simulación

En la pestaña Configuración del nodo Evaluación de simulación puede especificarse el rol de cada campo del conjunto de datos y personalizarse la salida generada por la simulación.

Seleccionar un elemento. Permite alternar las tres vistas del nodo Evaluación de simulación: Campos, Funciones de densidad y Resultados.

Vista Campos

Campo de destino. Se trata de un campo necesario. Pulse en la flecha para seleccionar el campo de destino del conjunto de datos de la lista desplegable. El campo seleccionado puede tener un nivel de medición continuo, ordinal o nominal, pero no puede tener un nivel de medición de fecha o sin especificar.

Campo de iteración (opcional). Si los datos tienen un campo de iteración que indica a qué iteración pertenece cada registro de los datos, deberá seleccionarse aquí. Esto significa que cada iteración se evaluará de forma independiente. Solo podrán seleccionarse los campos con niveles de medición continuo, ordinal o nominal.

Los datos de entrada ya están ordenados por iteración. Sólo está habilitado si se ha especificado un campo de iteración en el campo **Campo de iteración (opcional)**. Seleccione esta opción solo si está seguro de que los datos de entrada ya están ordenados por el campo de iteración especificado en **Campo de iteración (opcional)**.

Número máximo de iteraciones que se van a representar. Sólo está habilitado si se ha especificado un campo de iteración en el campo **Campo de iteración (opcional)**. Pulse en las flechas para especificar el número de iteraciones a representar. La especificación de este número evita que se intenten representar demasiadas iteraciones en una única gráfica, lo que dificultaría la interpretación de la misma. El nivel más bajo a que puede establecerse el número máximo de iteraciones es 2; el más alto, 50. El número máximo de iteraciones a representar está establecido inicialmente a 10.

Campos de entrada del tornado de correlación. La gráfica de tornado de correlación es una gráfica de barras que muestra los coeficientes de correlación entre el objetivo especificado y cada una de las entradas especificadas. Pulse en el icono del selector de campos para seleccionar, de una lista de entradas simuladas disponibles, los campos de entrada a incluir en la gráfica de tornado. Solo podrán seleccionarse los campos de entrada con niveles de medición continuo y ordinal. Los campos de entrada nominal, sin tipo y de fecha no aparecen en la lista y no pueden seleccionarse.

Vista de funciones de densidad

Mediante las opciones de esta vista, pueden personalizarse la salida de las funciones de densidad de probabilidad y las funciones de distribución acumulada de los objetivos continuos, así como las gráficas de barras de los valores predichos de los objetivos categóricos.

Funciones de densidad. Las funciones de densidad son las medias principales de la prueba del conjunto de resultados de su simulación.

- **Función de densidad de probabilidad (PDF).** Esta opción se selecciona para producir una función de densidad de probabilidad para el campo de destino. La función de densidad de probabilidad muestra la distribución de los valores de destino. La función de densidad de probabilidad puede utilizarse para determinar la probabilidad de que el objetivo se encuentre dentro de una determinada región. En el caso de los objetivos categóricos (objetivos con un nivel de medición nominal u ordinal) se genera una gráfica de barras que muestra el porcentaje de casos que entran en cada categoría del objetivo.
- **Función de distribución acumulada (CDF).** Esta opción se selecciona para producir una función de distribución acumulada para el campo de destino. La función de distribución acumulada muestra la probabilidad de que el valor del destino sea menor o igual que un valor especificado. Solo está disponible para destinos continuos.

Líneas de referencia (continuos). Estas opciones están habilitadas si **Función de densidad de probabilidad (PDF)**, **Función de distribución acumulada (CDF)** o ambas están seleccionadas. Con estas opciones, pueden añadirse varias líneas verticales de referencia fijas a funciones de densidad de probabilidad y funciones de distribución acumulada.

- **Media.** Esta opción se selecciona para añadir una línea de referencia en el valor medio del campo de destino.

- **Mediana.** Esta opción se selecciona para añadir una línea de referencia en el valor de la mediana del campo de destino.
- **Desviaciones estándar.** Esta opción se selecciona para añadir líneas de referencia en más-menos un número especificado de desviaciones estándar de la media del campo objetivo. Al seleccionarse esta opción se habilita el campo contiguo **Número**. Pulse en las flechas para especificar el número de desviaciones estándar. El número mínimo de desviaciones estándar es 1 y el máximo es 10. El número de desviaciones estándar se establece inicialmente a 3.
- **Percentiles.** Esta opción se selecciona para añadir líneas de referencia en dos valores de percentil de la distribución del campo de destino. La selección de esta opción habilita los campos de texto adjuntos **Inferior** y **Superior**. Por ejemplo, si se especifica un valor de 90 en el campo de texto **Superior** se añadiría una línea de referencia en el percentil 90º del objetivo, que es el valor por debajo del cual entra el 90% de las observaciones. Del mismo modo, un valor de 10 en el campo de texto **Inferior** representa el percentil décimo del objetivo, que es el valor por debajo del cual entra el 10% de las observaciones.
- **Líneas de referencia personalizadas.** Esta opción se selecciona para añadir líneas de referencia en valores especificados a lo largo del eje horizontal. La selección de esta opción habilita la tabla adjunta **Valores**. Cada vez que se especifica un número válido en la tabla **Valores**, se añade una nueva fila vacía al final de la tabla. Un número *válido* es un número que entra en el rango de valores del campo de destino.

Nota: Cuando varias funciones de densidad o funciones de distribución (de varias iteraciones) se visualizan en una única gráfica, las líneas de referencia (menos las líneas personalizadas) se aplican de forma independiente a cada función.

Objetivo categórico (solo PDF). Estas opciones solo se habilitan si está seleccionada la opción **Función de densidad de probabilidad (PDF)**.

- **Valores de categoría a incluir en el informe.** En el caso de los modelos con campos de destino categóricos, el resultado del modelo es un conjunto de las probabilidades predichas, una por cada categoría, en las que entra el valor de destino de cada categoría. La categoría con la probabilidad más elevada se considera como la categoría predicha y se utiliza en la generación de la gráfica de barras de la función de densidad de probabilidad. Seleccione **Categoría predicha** para generar la gráfica de barras. Seleccione **Probabilidades predichas** para generar histogramas de la distribución de las probabilidades predichas para cada una de las categorías del campo de destino. También puede seleccionarse **Ambos** para generar ambos tipos de gráfica.
- **Agrupamiento para el análisis de sensibilidad.** Las simulaciones que incluyen iteraciones de análisis de sensibilidad generan un campo de destino independiente (o campo de destino predicho a partir de un modelo) por cada iteración definida mediante el análisis. Existe una iteración por cada valor del parámetro de distribución que se varía. Cuando hay iteraciones presentes, la gráfica de barras de la categoría predicha de un campo de destino categórico se muestra como una gráfica de barras agrupada que incluye los resultados de todas las iteraciones. Seleccione **Agrupar categorías** o **Agrupar iteraciones**.

Vista de resultados

Valores de percentil de las distribuciones de destino. Con estas opciones, puede optarse por crear una tabla de valores de percentil de las distribuciones de destino y especificarse los percentiles a visualizar.

Crear una tabla de valores de percentil. En el caso de los campos de destino continuos, esta opción se selecciona para obtener una tabla de percentiles especificados de las distribuciones de destino. Seleccione una de las opciones siguientes para especificar los percentiles.

- **Cuartiles.** Los cuartiles son los percentiles 25º, 50º y 75º de la distribución del campo de destino. Las observaciones se dividen en cuatro grupos de igual tamaño.

- **Intervalos.** Si desea un número de grupos iguales distinto de cuatro, seleccione **Intervalos**. Al seleccionarse esta opción se habilita el campo contiguo **Número**. Pulse en las flechas para especificar el número de intervalos. El número mínimo de intervalos es 2 y el máximo es 100. El número de intervalos se establece inicialmente a 10.
- **Percentiles personalizados.** Seleccione **Percentiles personalizados** para especificar percentiles individuales como, por ejemplo, el percentil 99^o. La selección de esta opción habilita la tabla adjunta **Valores**. Cada vez que se especifica un número válido, entre 1 y 100, en la tabla **Valores**, se añade una nueva fila vacía al final de la tabla.

Salida del nodo Evaluación de simulación

Cuando se ejecuta un nodo Evaluación de simulación, la salida se añade al Gestor de resultados. El explorador de resultados de Evaluación de simulación muestra los resultados de ejecutar un nodo Evaluación de simulación. Las opciones habituales de guardado, exportación e impresión están disponibles en el menú **Archivo** y las opciones de edición habituales están disponibles en el menú **Editar**. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información. El menú **Ver** sólo está habilitado si una de las gráficas está seleccionada. No está habilitado para la tabla de distribución ni para las salidas de información. En el menú **Ver** puede seleccionarse **Modalidad de edición** para cambiar el diseño y la apariencia de la gráfica, o **Modalidad de exploración** para explorar los datos y valores representados en la gráfica. El modo estático fija las líneas de referencia de la gráfica (y los controles deslizantes) en sus posiciones actuales para que no puedan moverse. El modo estático es el único modo en el que se puede copiar, exportar o imprimir la gráfica con sus líneas de referencia. Para seleccionar este modo, pulse **Modo estático** en el menú **Ver**.

La ventana del explorador de resultados de Evaluación de simulación consta de dos paneles. A la izquierda de la ventana hay un panel de navegación que muestra representaciones en miniatura de las gráficas que se generan al ejecutar el nodo Evaluación de simulación. Cuando una miniatura está seleccionada, el resultado de la gráfica se muestra en el panel de la derecha de la ventana.

Panel de navegación

El panel de navegación del navegador de salida contiene las miniaturas de las gráficas generadas a partir de una simulación. Las miniaturas que aparecen en el panel de navegación dependen del nivel de medición del campo de destino, y en las opciones que están seleccionadas en el cuadro de diálogo del nodo Evaluación de simulación. En la tabla siguiente se proporcionan las descripciones de las miniaturas.

Tabla 57. Miniaturas del panel de navegación

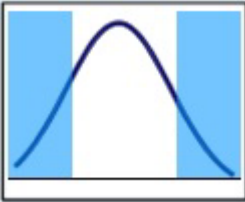
Miniatura	Descripción	Comentarios
	Función de densidad de probabilidad	Esta miniatura solo se muestra si el nivel de medición del campo de destino es continuo y Función de densidad de probabilidad (PDF) está seleccionada en la vista Funciones de densidad del cuadro de diálogo del nodo Evaluación de simulación. Si el nivel de medición del campo de destino es categórico, esta miniatura no se mostrará.

Tabla 57. Miniaturas del panel de navegación (continuación)

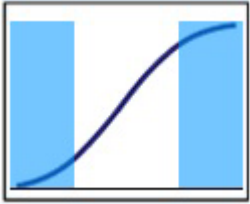
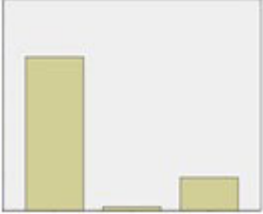
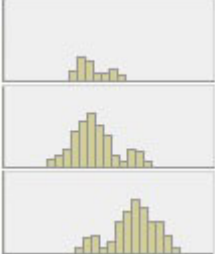
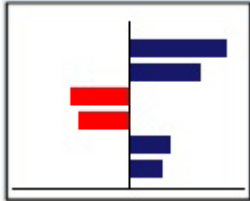
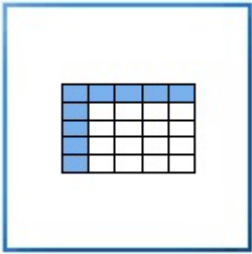

Miniatura	Descripción	Comentarios
	Función de distribución acumulada	<p>Esta miniatura solo se muestra si el nivel de medición del campo de destino es continuo y Función de distribución acumulada (CDF) está seleccionada en la vista Funciones de densidad del cuadro de diálogo del nodo Evaluación de simulación.</p> <p>Si el nivel de medición del campo de destino es categórico, esta miniatura no se mostrará.</p>
	Valores de categoría predicha	<p>Esta miniatura solo se muestra si el nivel de medición del campo de destino es categórico, está seleccionada la Función de densidad de probabilidad (PDF) en el cuadro de diálogo del nodo Evaluación de simulación y Categoría predicha, o bien Ambos, está seleccionada en el área Valores de categoría a notificar.</p> <p>Si el nivel de medición del campo de destino es continuo, esta miniatura no se mostrará.</p>
	Probabilidades de categoría predichas	<p>Esta miniatura solo se muestra si el nivel de medición del campo de destino es categórico, está seleccionada la Función de densidad de probabilidad (PDF) en el cuadro de diálogo del nodo Evaluación de simulación y Probabilidades predichas, o bien Ambos, está seleccionada en el área Valores de categoría a notificar.</p> <p>Si el nivel de medición del campo de destino es continuo, esta miniatura no se mostrará.</p>
	Gráficas de tornado	<p>Esta miniatura solo se muestra si uno o más campos de entrada están seleccionados en el campo Campos de entrada del tornado de correlación en la vista Campos del cuadro de diálogo del nodo Evaluación de simulación.</p>

Tabla 57. Miniaturas del panel de navegación (continuación)

Miniatura	Descripción	Comentarios
	Tabla de distribución	Esta miniatura solo se muestra si el nivel de medición del campo de destino es continuo y Crear una tabla de valores de percentil está seleccionada en la vista Salidas del cuadro de diálogo del nodo Evaluación de simulación. El menú Vista está inhabilitado para esta gráfica. Si el nivel de medición del campo de destino es categórico, esta miniatura no se mostrará.
	Información	Esta miniatura siempre se muestra. El menú Vista está inhabilitado para esta salida.

Resultado de gráfica

Los tipos de gráfica de resultados disponibles dependen del nivel de medición del campo de destino, de si se utiliza un campo de iteración y de las opciones seleccionadas en el cuadro de diálogo del nodo Evaluación de simulación. Una serie de gráficas generadas a partir de una simulación tiene características interactivas que pueden utilizarse para personalizar la visualización. Las características interactivas están disponibles pulsando **Opciones de gráfica**. Todos los gráficos de simulación son visualizaciones de tablero.

Gráficas de funciones de densidad de probabilidad de destinos continuos. Esta gráfica muestra la probabilidad y la frecuencia, con la escala de probabilidad en el eje vertical izquierdo y la escala de frecuencia en el eje vertical derecho. La gráfica tiene dos líneas de referencia verticales deslizantes que dividen la gráfica en regiones independientes. La tabla situada debajo de la gráfica muestra el porcentaje de la distribución situado en cada una de las regiones. Si se visualizan varias funciones de densidad en el mismo diagrama (debido a las iteraciones), la tabla tendrá una fila independiente para las probabilidades asociadas a cada función de densidad y una columna adicional que contiene el nombre de la iteración y un color asociado a cada función de densidad. Las iteraciones se listan en la tabla por orden alfabético de la etiqueta de la iteración. Si no hay disponible ninguna etiqueta de iteración, se utilizará el valor de la iteración en su lugar. La tabla no puede editarse.

Cada una de las líneas de referencia tiene un control deslizante (triángulo invertido) que puede utilizarse para mover la línea con facilidad. Cada control deslizante tiene una etiqueta que indica su posición actual. De forma predeterminada, los controles deslizantes se sitúan en los percentiles 5° y 95° de la distribución. Si hay varias iteraciones, los controles deslizantes se sitúan en los percentiles 5° y 95° de la primera distribución que aparece listada en la tabla. No se pueden mover las líneas de forma que se crucen entre sí.

Hay una serie de características adicionales disponibles al pulsar **Opciones de gráfica**. En concreto, se pueden establecer de forma explícita las posiciones de los controles deslizantes, añadir líneas de

referencia fijas y cambiar la vista de gráfica de curva continua a histograma. Consulte el tema “Opciones de gráfica” en la página 335 para obtener más información. Pulse con el botón derecho en la gráfica para copiar o exportar la gráfica.

Gráficas de funciones de distribución acumulada de destinos continuos. Esta gráfica tiene las dos mismas líneas de referencia verticales movibles y la tabla asociada descritas en el apartado de la gráfica de función de densidad de probabilidad. Los controles deslizantes y la tabla se comportan de la misma forma que la función de densidad de probabilidad cuando hay varias iteraciones. Se utilizan los mismos colores que en las funciones de distribución para identificar qué función de densidad pertenece a cada iteración.

Esta gráfica también proporciona acceso al cuadro de diálogo Opciones de gráfica, que permite establecer de forma explícita las posiciones de los controles deslizantes y líneas de referencia fijas, y especificar si la función de distribución acumulada se visualiza como una función que incrementa (comportamiento predeterminado) o una función que decreciente. Consulte el tema “Opciones de gráfica” en la página 335 para obtener más información. Pulse con el botón derecho en la gráfica para copiar, exportar o editar la gráfica. Al seleccionar **Editar** se abre la gráfica en una ventana del editor de tablero flotante.

Gráfica de valores de categoría predichos para objetivos categóricos. En el caso de los campos objetivo categóricos, una gráfica de barras muestra los valores predichos. Los valores predichos se muestran como el porcentaje del campo de destino que se prevé que va a entrar en cada categoría. En el caso de los campos de destino categóricos con iteraciones de análisis de sensibilidad, los resultados de la categoría de destino predicha se muestran como una gráfica de barras agrupadas que incluye los resultados de todas las iteraciones. La gráfica se agrupa por categoría o por iteración, dependiendo de qué opción se ha seleccionado en el área **Agrupación del análisis de sensibilidad** de la vista Funciones de densidad del cuadro de diálogo del nodo Evaluación de simulación. Pulse con el botón derecho en la gráfica para copiar, exportar o editar la gráfica. Al seleccionar **Editar** se abre la gráfica en una ventana del editor de tablero flotante.

Gráfica de probabilidades de categoría predichas de objetivos categóricos. En el caso de los campos de destino categóricos, un histograma muestra la distribución de las probabilidades predichas para cada una de las categorías del objetivo. En el caso de campos de destino categóricos con iteraciones de análisis de sensibilidad, los histogramas se visualizan por categoría o por iteración, dependiendo de qué opción se ha seleccionado en el área **Agrupación del análisis de sensibilidad** de la vista Funciones de densidad del cuadro de diálogo del nodo Evaluación de simulación. Si los histogramas están agrupados por categoría, una lista desplegable que contiene las etiquetas de iteración permitirá elegir qué iteración se visualiza. También puede seleccionarse la iteración a visualizar pulsando con el botón derecho del ratón en la gráfica y seleccionando la iteración en el submenú **Iteración**. Si los histogramas están agrupados por iteración, una lista desplegable que contiene los nombres de categoría permitirá elegir qué categoría se visualiza. También puede seleccionarse la categoría a visualizar pulsando con el botón derecho del ratón en la gráfica y seleccionando la categoría en el submenú **Categoría**.

Esta gráfica solo está disponible para un subconjunto de modelos y, en el nugget de modelo, la opción para generar todas las probabilidades de grupo debe estar seleccionada. Por ejemplo, en el nugget de modelo logístico, debe seleccionarse **Añadir todas las probabilidades**. Los siguientes nuggets de modelo soportan esta opción:

- Logístico, SVM, Bayes, Red neural y KNN.
- Modelos de minería en base de datos de DB2/ISW para regresión logística, árboles de decisión y naïve Bayes.

De forma predeterminada, la opción de generar todas las probabilidades de grupo no está seleccionada en estos nuggets de modelo.

Gráficas de tornado. Una gráfica de tornado es una gráfica de barras que muestra la sensibilidad del campo de destino a cada una de las entradas especificadas. La sensibilidad se mide por la correlación de destino con cada entrada. El título de la gráfica contiene el nombre del campo de destino. Cada barra del

diagrama representa la correlación entre el campo de destino y un campo de entrada. Las entradas simuladas incluidas en la gráfica son entradas que están seleccionadas en el campo **Campos de entrada del tornado de correlación** en la vista Campos del cuadro de diálogo del nodo Evaluación de simulación. Cada barra se etiqueta con el valor de correlación. Las barras se ordenan por el valor absoluto de las correlaciones, de mayor a menor valor. Si hay iteraciones, se genera una gráfica independiente por cada iteración. Cada gráfica tiene un subtítulo que contiene el nombre de la iteración.

Tabla de distribución. Esta tabla contiene el valor del campo de destino bajo el cual cae el porcentaje de observaciones especificado. La tabla contiene una fila por cada valor de percentil especificado en la vista Salidas del cuadro de diálogo del nodo Simulación de evaluación. Los valores de percentil pueden ser cuartiles, un número diferente de percentiles espaciados regularmente o percentiles especificados individualmente. La tabla de distribución contiene una columna por cada iteración.

Información. Esta sección proporciona un resumen general de los campos y registros que se utilizan en la evaluación. También muestra los campos de entrada y los recuentos de registro, desglosados por iteración.

Opciones de gráfica

En el cuadro de diálogo Opciones de gráfica se puede personalizar la visualización de gráficas activadas de funciones de densidad de probabilidad y funciones de distribución acumulada generadas a partir de una simulación.

Vista. La lista desplegable **Vista** sólo se aplica a la gráfica de función de densidad de probabilidad. Puede utilizarse para alternar la vista de gráfica entre una curva continua y un histograma. Esta característica está inhabilitada cuando se visualizan varias funciones de densidad (de múltiples iteraciones) en el mismo diagrama. Cuando hay varias funciones de densidad, estas sólo pueden visualizarse como curvas continuas.

Ordenar. La lista desplegable **Orden** sólo se aplica a la gráfica de función de distribución acumulada. Especifica si la función de distribución acumulada se muestra como una función ascendente (la opción predeterminada) o una función descendente. Cuando se visualiza como una función descendente, el valor de la función en un determinado punto del eje horizontal es la probabilidad de que el campo de destino se encuentre a la derecha de dicho punto.

Posiciones de control deslizante. El campo de texto **Superior** contiene la posición actual de la línea de referencia que se desliza a la derecha. El campo de texto **Inferior** contiene la posición actual de la línea de referencia que se desliza a la izquierda. Las posiciones de los controles deslizantes pueden establecerse explícitamente especificando valores en los campos de texto **Superior** e **Inferior**. El valor en el campo de texto **Inferior** debe ser estrictamente menor que el valor del campo de texto **Superior**. La línea de referencia izquierda puede eliminarse seleccionando **-Infinito**, con lo que se establece la posición de forma efectiva a infinito negativo. Esta acción inhabilita el campo de texto **Inferior**. La línea de referencia derecha puede eliminarse seleccionando **Infinito**, con lo que se establece su posición de forma efectiva a infinito. Esta acción inhabilita el campo de texto **Superior**. No se pueden eliminar ambas líneas de referencia; la selección de **-Infinito** inhabilita el recuadro de selección **Infinito** y viceversa.

Líneas de referencia. Pueden añadirse varias líneas verticales de referencia fijas a funciones de densidad de probabilidad y funciones de distribución acumulada.

- **Media.** Puede añadirse una línea de referencia en la media del campo de destino.
- **Mediana.** Puede añadirse una línea de referencia en la mediana del campo de destino.
- **Desviaciones estándar.** Pueden añadirse líneas de referencia en más-menos un número especificado de desviaciones estándar de la media del campo objetivo. Puede especificarse el número de desviaciones estándar que se va a utilizar en el campo de texto adyacente. El número mínimo de desviaciones estándar es 1, y el máximo es 10. El número de desviaciones estándar se establece inicialmente a 3.
- **Percentiles.** Pueden añadirse líneas de referencia en uno o dos valores de percentil de la distribución del campo de destino especificando valores en los campos de texto **Parte inferior** y **Parte superior**. Por

ejemplo, un valor de 95 en el campo de texto **Parte superior** representa el percentil 95^o, que es el valor por debajo del cual se sitúa el 95% de las observaciones. De igual manera, un valor de 5 en el campo de texto **Parte inferior** representa el quinto percentil, que es el valor por debajo del cual se sitúa el 5% de las observaciones. En el caso del campo de texto de la **Parte inferior**, el valor de percentil mínimo es de 0 y el máximo 49. En el caso del campo de texto de la **Parte superior**, el valor de percentil mínimo es de 50 y el máximo 100.

- **Posiciones personalizadas.** Pueden añadirse líneas de referencia a los valores especificados en el eje horizontal. Pueden eliminarse líneas de referencia personalizadas suprimiendo la entrada de la cuadrícula.

Cuando pulsa **Aceptar**, los controles deslizantes, las etiquetas sobre los controles deslizantes, las líneas de referencia y la tabla situada debajo del diagrama se actualizan para reflejar las opciones seleccionadas en el cuadro de diálogo Opciones de gráfica. Pulse **Cancelar** para cerrar el cuadro de diálogo sin realizar ningún cambio. Las líneas de referencia pueden eliminarse deseleccionando la selección asociada en el diálogo Opciones de gráfica y pulsando **Aceptar**.

Nota: Cuando se muestran varias funciones de densidad o funciones de distribución en una única gráfica (debido a los resultados de las iteraciones de análisis de sensibilidad), las líneas de referencia (que no sean líneas personalizadas) se aplican por separado a cada función. Solo se muestran las líneas de la primera iteración. Las etiquetas de línea de referencia incluyen la etiqueta de iteración. La etiqueta de la iteración se deriva de un punto anterior en la ruta, normalmente de un nodo Generar simulación. Si no hay disponible ninguna etiqueta de iteración, se utilizará el valor de la iteración en su lugar. Las opciones **Media**, **Mediana**, **Desviaciones estándar** y **Percentiles** están inhabilitadas en el caso de funciones de distribución acumulada con varias iteraciones.

Aplicaciones de ayuda de IBM SPSS Statistics

Si tiene instalada una versión compatible de IBM SPSS Statistics en su ordenador con la licencia necesaria, puede configurar IBM SPSS Modeler para procesar datos con funcionalidad IBM SPSS Statistics mediante los nodos Transformación de Estadísticas, Modelo de Estadísticas Resultados de Estadísticas o Exportar de Estadísticas.

Para obtener información sobre la compatibilidad de productos con la versión actual de IBM SPSS Modeler, consulte el sitio de soporte de la empresa en <http://www.ibm.com/support>.

Para configurar IBM SPSS Modeler para que funcione con IBM SPSS Statistics y otras aplicaciones, seleccione:

Herramientas > Opciones > Aplicaciones de ayuda

IBM SPSS Statistics Interactive. Introduzca la ruta completa y el nombre del comando (por ejemplo, *C:\Archivos de programa\IBM\SPSS\Statistics\<nn>\stats.exe*) que desea utilizar al iniciar IBM SPSS Statistics directamente en un archivo de datos producido por el nodo de exportación de Statistics. Consulte el tema “Nodo Exportación de Statistics” en la página 373 para obtener más información.

Conexión. Si el servidor IBM SPSS Statistics se encuentra en el mismo servidor que el servidor de IBM SPSS Modeler Server, puede activar una conexión entre las dos aplicaciones, lo que aumenta la eficacia al dejar los datos en el servidor durante el análisis. Seleccione **Servidor** para activar la opción **Puerto** más abajo. La configuración predeterminada es **Local**.

Puerto. Especifique el puerto del servidor para el servidor IBM SPSS Statistics.

Utilidad de localización de IBM SPSS Statistics. Para permitir que IBM SPSS Modeler utilice los nodos Transformación de Statistics, Modelo de Statistics y Resultados de Statistics, debe tener instalada una copia de IBM SPSS Statistics con licencia en el ordenador donde se ejecute la ruta.

- Si ejecuta IBM SPSS Modeler en modo local (autónomo), la copia con licencia de IBM SPSS Statistics debe estar en el equipo local. Pulse este botón para especificar la ubicación de la instalación local de IBM SPSS Statistics que desea utilizar para la licencia.
- Además, si se ejecuta en modo distribuido en un servidor de IBM SPSS Modeler Server remoto, también necesita ejecutar una utilidad en el host de IBM SPSS Modeler Server para crear el archivo *statistics.ini*, que indica a IBM SPSS Statistics la ruta de instalación de IBM SPSS Modeler Server. Para ello, en el indicador de comandos, cambie al directorio IBM SPSS Modeler Server *bin* y, en Windows, ejecute:

```
statisticsutility -location=<IBM SPSS Statistics_ruta de instalación>/
```

De forma alternativa, en UNIX, ejecute:

```
./statisticsutility -location=<IBM SPSS Statistics_ruta de instalación>/bin
```

Si no tiene una copia con licencia de IBM SPSS Statistics en su máquina local, podrá seguir ejecutando nodo Estadísticos de archivo en un servidor de IBM SPSS Statistics, pero si intenta ejecutar otros nodos de IBM SPSS Statistics, se mostrará un mensaje de error.

Comentarios

Si tiene problemas para ejecutar los nodos Procedimiento de IBM SPSS Statistics, tenga en cuenta las siguientes sugerencias:

- Si los nombres de campos utilizados en IBM SPSS Modeler tienen más de ocho caracteres (para versiones anteriores a IBM SPSS Statistics 12.0), más de 64 caracteres (para IBM SPSS Statistics 12.0 y versiones posteriores) o contienen caracteres no válidos, es necesario cambiarles el nombre o truncarlos antes de leerlos en IBM SPSS Statistics. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.
- Si IBM SPSS Statistics se instaló después que IBM SPSS Modeler, puede que necesite especificar la ubicación de IBM SPSS Statistics, tal y como se explicó anteriormente.

Capítulo 7. Nodos de exportación

Conceptos básicos sobre nodos de exportación

Los nodos de exportación proporcionan un mecanismo para exportar datos en varios formatos y poder interactuar con otras herramientas de software.

Los siguientes nodos de exportación están disponibles:



El nodo Exportar base de datos escribe datos en orígenes de datos relacionales compatibles con ODBC. Para escribir en un origen de datos ODBC, el origen de datos debe existir y debe tener permiso para escribir en él.



El nodo Archivo sin formato produce datos en un archivo de texto delimitado. Esto es útil para exportar datos que se pueden leer con otro software de hoja de cálculo o de análisis.



El nodo Exportar estadísticas genera datos de salida en formato *.sav* o *.zsav* de IBM SPSS Statistics. Los archivos *.sav* o *.zsav* se pueden leer con IBM SPSS Statistics Base y otros productos. Este es también el formato utilizado para los archivos caché de IBM SPSS Modeler.



El nodo de exportación IBM SPSS Data Collection abre los datos en el formato utilizado por el software de investigación de mercados IBM SPSS Data Collection. Se debe instalar la biblioteca de datos de IBM SPSS Data Collection para utilizar este nodo.



El nodo Exportar de IBM Cognos BI exporta datos en un formato que pueden leer las bases de datos de Cognos BI.



El nodo Exportar IBM Cognos TM1 exporta datos en un formato que pueden leer las bases de datos de Cognos TM1.



El nodo Exportar SAS produce datos en formato SAS, para leerlos en SAS o en un paquete de software compatible con SAS. Están disponibles tres formatos de archivo SAS: SAS para Windows/OS2, SAS para UNIX o SAS versión 7/8.



El nodo de exportación Excel ofrece los datos resultantes en formato de Microsoft Excel (*.xls*). Si lo desea, puede elegir iniciar automáticamente Excel y abrir el archivo exportado cuando se ejecute el nodo.



El nodo de exportación XML exporta datos a un archivo en formato XML. También puede crear un nodo de origen XML para leer los datos exportados a la ruta.

Nodo Exportar base de datos

Puede usar nodos Base de datos para escribir datos en orígenes de datos relacionales compatibles con ODBC, que se explican en la descripción del nodo de origen de base de datos. Consulte el tema “Nodo de origen Base de datos” en la página 17 para obtener más información.

Siga estos pasos generales para escribir datos en una base de datos:

1. Instale un controlador ODBC y configure un origen de datos en la base de datos que desee utilizar.
2. En la pestaña Exportar del nodo Base de datos, especifique el origen de datos y la tabla donde desea escribir. Puede crear una tabla nueva o insertar datos en una existente.
3. Especifique opciones adicionales según sea necesario.

Estos pasos se describen de forma más detallada en los siguientes temas.

Pestaña Exportar del nodo Base de datos

Origen de datos. Muestra el origen de datos seleccionado. Introduzca el nombre o selecciónelo de la lista desplegable. Si no ve en la lista la base de datos que desea, seleccione **Añadir nueva conexión de base de datos** y busque la base de datos desde el cuadro de diálogo Conexiones a la base de datos. Consulte el tema “Adición de una conexión de base de datos” en la página 20 para obtener más información.

Nombre de tabla. Introduzca el nombre de la tabla a la que desee enviar los datos. Si selecciona la opción **Insertar en tabla**, puede seleccionar una tabla existente de la base de datos pulsando en el botón **Seleccionar**.

Crear tabla. Seleccione esta opción para crear una nueva tabla de base de datos o para sobrescribir una tabla de base de datos existente.

Insertar en tabla. Seleccione esta opción para insertar los datos como nuevas filas en una tabla de base de datos existente.

Tabla Fusión. (Donde esté disponible) Seleccione esta opción para actualizar columnas de base de datos seleccionadas con valores de los campos de datos de origen correspondiente. Si selecciona esta opción activará el botón **Fusión**, que muestra un cuadro de diálogo donde podrá correlacionar campos de orígenes de datos a columnas de base de datos.

Eliminar tabla existente. Seleccione esta opción para eliminar cualquier tabla existente con el mismo nombre al crear una tabla nueva.

Eliminar filas existentes. Seleccione esta opción para eliminar filas existentes de la tabla antes de exportar cuando se inserte en una tabla.

Note: si está seleccionada alguna de las dos opciones anteriores, recibirá el mensaje **Advertencia de sobrescritura** cuando ejecute el nodo. Para suprimir las advertencias, anule la selección de **Avisar cuando un nodo sobrescriba una tabla de la base de datos** en la pestaña Notificaciones del cuadro de diálogo Opciones de usuario.

Tamaño de cadena predeterminado. Los campos marcados como sin tipo en un nodo Tipo anterior de la ruta se escriben en la base de datos como campos de cadena. Especifique el tamaño de las cadenas que se van a utilizar para campos sin tipo.

Pulse en **Esquema** para abrir un cuadro de diálogo donde podrá establecer diversas opciones de exportación (para bases de datos que admitan esta característica), establecer los tipos de datos de SQL para los campos y especificar la clave primaria para la indización de la base de datos. Consulte el tema “Opciones de Esquema de exportación de base de datos” en la página 342 para obtener más información.

Pulse en **Índices** para especificar las opciones de indización de la tabla exportada para mejorar el rendimiento de la base de datos. Consulte el tema “Opciones de índice de exportación de base de datos” en la página 345 para obtener más información.

Pulse en **Avanzado** para especificar opciones de carga masiva y de confirmación de base de datos. Consulte el tema “Opciones avanzadas de exportación de base de datos” en la página 347 para obtener más información.

Poner entre comillas nombres de columna y tabla. Seleccione opciones utilizadas al enviar la instrucción CREATE TABLE a la base de datos. Las tablas o columnas con espacios o caracteres que no sean estándar deben ponerse entre comillas.

- **Cuando sea necesario.** Selecciónelo para permitir que IBM SPSS Modeler determine automáticamente cuándo son necesarias las comillas de forma individual.
- **Siempre.** Para poner siempre entre comillas los nombres de tabla y columna.
- **Nunca.** Para desactivar la utilización de comillas.

Generar un nodo de importación para este dato. Para generar un nodo de origen de Base de datos para los datos tal y como se exportan en la tabla y el origen de datos especificado. Al ejecutarse, este nodo se añade al lienzo de rutas.

Opciones de fusión Exportar base de datos

Este cuadro de diálogo le permite correlacionar campos desde los datos de origen a columnas en la tabla de base de datos de destino. Si un campo de datos de origen se correlaciona a una columna de base de datos, el valor de columna se sustituye por el valor de los datos de origen cuando se ejecuta la ruta. Los campos de origen sin correlacionar se quedan sin cambiar en la base de datos.

Correlacionar campos. Aquí puede especificar la correlación entre campos de datos de origen y columnas de base de datos. Los campos de datos de origen con el mismo nombre que las columnas de la base de datos se correlacionan automáticamente.

- **Correlacionar.** Correlaciona un campo de datos de origen en la lista de campos en la parte izquierda del botón a una columna de base de datos seleccionada en la lista de la derecha. Puede correlacionar uno o más campos a la vez, pero el número de entradas seleccionadas en ambas listas deben ser las mismas.
- **Cancelar correlación.** Elimina la correlación para una o más columnas de base de datos seleccionadas. Este botón se activa cuando se selecciona una columna de base de datos o campo en la tabla que aparece a la derecha del cuadro de diálogo.
- **Añadir.** Añade uno o más campos de datos de origen seleccionados en la lista de campos a la izquierda del botón a la lista de la derecha, lista para la correlación. Este botón se activa cuando selecciona un campo de la lista de la izquierda y no existe ningún campo con ese nombre en la lista de la derecha. Al pulsar en este botón se correlaciona el campo seleccionado a una nueva columna de base de datos con el mismo nombre. La palabra <NEW> aparece tras el nombre de columna de base de datos para indicar que se trata de un nuevo campo.

Fusionar filas. Usa un campo clave, como el *ID de transacción*, para fusionar registros que compartan el mismo valor en este campo. Equivale a una "equi-uniión" de base de datos. Los valores clave deben ser aquellos de las claves primarias; es decir, deben ser exclusivos y no pueden contener valores nulos.

- **Claves posibles.** Enumera todos los campos encontrados en todos los orígenes de datos de entrada. Seleccione uno o más campos de esta lista y utilice el botón de flecha para añadirlos como campos clave para fusionar los registros. Cualquier campo de correlación con una columna de base de datos

correlacionada está disponible como clave, salvo que los campos se añadan como nuevas columnas de bases de datos (se muestran con <NEW> tras el nombre) no están disponibles.

- **Claves para fusión.** Enumera todos los campos usados para fusionar registros desde todos los orígenes de datos de entrada basándose en valores de los campos clave. Para quitar una clave de la lista, selecciónela y use el botón de flecha para devolverla a la lista Claves posibles. Cuando se seleccionan varios campos clave, la opción de la parte inferior se desactiva.
- **Incluir únicamente registros que hay en la base de datos.** Realiza una unión parcial; si el registro está en la base de datos y la ruta, se actualizarán los campos correlacionados.
- **Añade registros a la base de datos.** Realiza una unión exterior; se fusionarán todos los registros de la ruta (si existe el mismo registro en la base de datos) o se añadirán (si el registro no existe en la base de datos).

Para correlacionar un campo de datos de origen a una columna de base de datos nueva

1. Pulse en el nombre de campo de origen en la lista de la izquierda, en **Correlacionar campos**.
2. Pulse en el botón **Añadir** para comprobar la correlación.

Para correlacionar un campo de datos de origen a una columna de base de datos existente

1. Pulse en el nombre de campo de origen en la lista de la izquierda, en **Correlacionar campos**.
2. Pulse en el nombre de la columna en **Columna de base de datos** en la parte derecha.
3. Pulse en el botón **Correlacionar** para completar la correlación.

Para eliminar una correlación

1. En la lista de la derecha, debajo de Campo, pulse en el nombre del campo cuya correlación desea eliminar.
2. Pulse en el botón **Cancelar correlación**.

Para cancelar la selección de un campo en cualquiera de las listas

Mantenga pulsada la tecla CTRL y pulse el nombre del campo.

Opciones de Esquema de exportación de base de datos

En el cuadro de diálogo Esquema de exportación de base de datos, puede establecer opciones para la exportación de base de datos (para bases de datos que admitan estas opciones), establecer tipos de datos SQL para sus campos, especificar qué campos son las claves primarias y personalizar la instrucción CREATE TABLE generada al exportar.

El cuadro de diálogo tiene varias partes:

- La sección de la parte superior (si se muestra) contiene opciones para la exportación a una base de datos que admita estas opciones. Esta sección no se muestra si no está conectado a dicha base de datos.
- El campo de texto de la parte central muestra la plantilla utilizada para generar el comando CREATE TABLE, que de forma predeterminada tiene el siguiente formato:
CREATE TABLE <table-name> <(table columns)>
- La tabla de la parte inferior le permite especificar el tipo de datos SQL para cada campo e indicar qué campos son claves primarias como se explica a continuación. El cuadro de diálogo genera automáticamente los valores de los parámetros <table-name> y <(table columns)> basados en las especificaciones de la tabla.

Configuración de opciones de exportación de base de datos

Si esta sección se muestra, puede especificar determinados ajustes para exportar a la base de datos. Los tipos de bases de datos que admiten esta característica son los siguientes:

- IBM InfoSphere Warehouse ejecutándose en DB2 9.1 o posterior. Consulte el tema “Opciones para IBM DB2 InfoSphere Warehouse” para obtener más información.
- SQL Server 2008 o posterior, Enterprise Edition y Developer Edition. Consulte el tema “Opciones para SQL Server” en la página 344 para obtener más información.
- Oracle 10g y 11gR1 o posterior, Enterprise Edition o Personal Edition. Consulte el tema “Opciones para Oracle” en la página 344 para obtener más información.

Personalización de las instrucciones CREATE TABLE

Utilizando la parte de campo de texto de este cuadro de diálogo, puede añadir opciones adicionales específicas de la base de datos a la instrucción CREATE TABLE.

1. Seleccione la casilla de verificación **Personalizar comando CREAM TABLA** para activar la ventana de texto.
2. Añada cualquier opción específica de la base de datos a la instrucción. Asegúrese de mantener el texto de los parámetros <table-name> y (<table-columns>) ya que IBM SPSS Modeler los sustituye por las definiciones de columna y nombre real de tabla.

Configuración de tipos de datos SQL

De forma predeterminada, IBM SPSS Modeler permitirá que el servidor de la base de datos asigne tipos de datos SQL automáticamente. Para sobrescribir el tipo automático para un campo, busque la fila correspondiente al campo y seleccione el tipo que desee de la lista desplegable en la columna *Tipo* de la tabla de esquema. Puede utilizar Mayús y pulsar para seleccionar más de una fila.

Para tipos con una longitud, precisión o argumento de escala (BINARY, VARBINARY, CHAR, VARCHAR, NUMERIC y NUMBER), debe especificar una longitud en lugar de permitir que el servidor de la base de datos asigne una longitud automática. Por ejemplo, especificar un valor adecuado, como VARCHAR(25), para la longitud garantiza que el tipo de almacenamiento en IBM SPSS Modeler se sobrescribirá si esa es su intención. Para sobrescribir la asignación automática, seleccione **Especificar** de la lista desplegable Tipo y sustituya la definición de tipo con la instrucción de definición de tipo de SQL que desee.

La forma más fácil de hacerlo es seleccionando primero el tipo más cercano a la definición de tipo que desee y, a continuación, seleccionando **Especificar** para modificar esa definición. Por ejemplo, para establecer un tipo de datos de SQL en VARCHAR(25), establezca primero el tipo en **VARCHAR(longitud)** desde la lista desplegable Tipo y, a continuación, seleccione **Especificar** y sustituya la longitud del texto por el valor 25.

Claves primarias

Si una o más columnas de la tabla exportada debe tener un valor exclusivo o una combinación de valores exclusiva para cada fila, puede indicarlo seleccionando la casilla de verificación **Clave primaria** para cada campo al que se aplique. La mayoría de las bases de datos no permiten que se modifique la tabla de ninguna manera que invalide una restricción de clave primaria y crearán automáticamente un índice sobre la clave primaria para ayudar a aplicar dicha restricción. (Si lo desea, puede crear índices para otros campos en el cuadro de diálogo Índices. Consulte el tema “Opciones de índice de exportación de base de datos” en la página 345 para obtener más información.)

Opciones para IBM DB2 InfoSphere Warehouse

Espacio de tabla. El espacio de tabla que debe utilizarse para la exportación. Los administradores de la base de datos pueden crear o configurar espacios de tabla con particiones. Recomendamos que seleccione uno de estos espacios de tabla (en lugar del valor predeterminado) para utilizarlo para la exportación de base de datos.

Dividir los datos por campo. Especifica el campo de entrada que debe utilizarse en la partición.

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión (por ejemplo, el equivalente a CREATE TABLE MYTABLE(...) COMPRESS YES; en SQL).

Opciones para SQL Server

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Fila.** Activa la compresión a nivel de fila (por ejemplo, el equivalente de CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = ROW); en SQL).
- **Página.** Activa la compresión a nivel de página (por ejemplo, CREATE TABLE MYTABLE(...) WITH (DATA_COMPRESSION = PAGE); en SQL).

Opciones para Oracle

Configuración de Oracle 10g

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión. Para esta versión de la base de datos, únicamente está disponible la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL).

Configuración de Oracle 11gR1

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción **Operaciones de carga directa**.
- **Operaciones de carga directa.** Activa la compresión únicamente para operaciones de inserción masivas (de ruta directa) (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR DIRECT_LOAD OPERATIONS; en SQL).
- **Todas las operaciones.** Activa la compresión para todas las operaciones (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR ALL OPERATIONS; en SQL).

Configuración de Oracle 11gR2: opción Básica

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción **Básica**.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).

Configuración de Oracle 11gR2: opción Avanzada

Utilizar compresión. Si está seleccionado, crea tablas para la exportación con compresión.

Compresión para. Elija el nivel de compresión.

- **Valor predeterminado.** Activa la compresión predeterminada (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS; en SQL). En este caso, tiene el mismo efecto que la opción **Básica**.
- **Básica.** Activa la compresión básica (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS BASIC; en SQL).
- **OLTP.** Activa la compresión OLTP (por ejemplo, CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP; en SQL).

- **Consulta baja/alta.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para la consulta (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY LOW;` o `CREATE TABLE MYTABLE(...)COMPRESS FOR QUERY HIGH;` en SQL). La compresión para consulta es de utilidad en entornos de almacenamiento de datos; HIGH proporciona una relación de compresión mayor que LOW.
- **Archivo bajo/alto.** (Servidores Exadata únicamente) Activa la compresión de columna híbrida para el archivo (por ejemplo, `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE LOW;` o `CREATE TABLE MYTABLE(...)COMPRESS FOR ARCHIVE HIGH;` en SQL). La compresión para archivo es de utilidad para comprimir datos que se almacenarán durante largos períodos de tiempo; HIGH proporciona una relación de compresión mayor que LOW.

Opciones de índice de exportación de base de datos

El cuadro de diálogo Índices le permite crear índices en tablas de bases de datos exportadas desde IBM SPSS Modeler. Puede especificar los conjuntos de campo que desea incluir y personalizar el comando `CREATE INDEX`, según sea necesario.

El cuadro de diálogo tiene dos partes:

- El campo de texto de la parte superior muestra una plantilla que se puede utilizar para generar uno o más comandos `CREATE INDEX`, que, de forma predeterminada, tienen el formato:
`CREATE INDEX <nombre-índice> ON <nombre-tabla>`
- La tabla de la parte inferior del cuadro de diálogo le permite añadir especificaciones para cada índice que desee crear. En cada índice, especifique el nombre del índice y los campos o columnas que desee incluir. El cuadro de diálogo genera automáticamente los valores de los parámetros `<index-name>` y `<table-name>` correctamente.

Por ejemplo, el SQL generado para un índice único de los campos *empid* y *deptid* tendría un aspecto similar a este:

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)
```

Puede añadir varias filas para crear varios índices. Se generará un comando `CREATE INDEX` diferente para cada fila.

Personalización del comando `CREATE INDEX`

Si lo desea, puede personalizar el comando `CREATE INDEX` para todos los índices o sólo para un índice específico, lo que le proporciona la flexibilidad para adaptarse a opciones o requisitos específicos de la base de datos y aplicar personalizaciones a todos los índices o solamente a unos determinados, según sea necesario.

- Seleccione **Personalizar comando `CREATE INDEX`** en la parte superior del cuadro de diálogo para modificar la plantilla que se utiliza para todos los índices añadidos posteriormente. Tenga en cuenta que los cambios no se aplicarán automáticamente a los índices que ya se hayan añadido a la tabla.
- Seleccione una o más filas de la tabla y, a continuación, pulse en **Actualizar índices seleccionados** en la parte superior del cuadro de diálogo para aplicar las personalizaciones actuales a todas las filas seleccionadas.
- Seleccione la casilla de verificación **Personalizar** de cada fila para modificar la plantilla del comando solamente para tal índice.

Tenga en cuenta que los valores de los parámetros `<index-name>` y `<table-name>` se generan automáticamente a través del cuadro de diálogo según las especificaciones de la tabla y no se pueden editar directamente.

Palabra clave BITMAP. Si está utilizando una base de datos Oracle, puede personalizar la plantilla para crear un índice de mapa de bits en lugar de un índice estándar, de la siguiente manera:

```
CREATE BITMAP INDEX <index-name> ON <table-name>
```

Los índices de mapas de bits pueden ser útiles para la indexación de columnas con una pequeña cantidad de valores distintos. El SQL resultante tendría una apariencia similar a esto:

```
CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)
```

palabra clave UNIQUE. La mayoría de las bases de datos admiten la palabra clave UNIQUE en el comando CREATE INDEX, Esto aplica una restricción de exclusividad parecida a la restricción de clave primaria de la tabla subyacente.

```
CREATE UNIQUE INDEX <index-name> ON <table-name>
```

Tenga en cuenta que para los campos ya designados como claves primarias, no es necesaria esta especificación. La mayoría de las bases de datos crearán automáticamente un índice para cualquier campo especificado como campo clave primario dentro del comando CREATE TABLE, por lo que no es necesario crear explícitamente índices en estos campos. Consulte el tema “Opciones de Esquema de exportación de base de datos” en la página 342 para obtener más información.

palabra clave FILLFACTOR. Algunos parámetros físicos para el índice se pueden ajustar con precisión. Por ejemplo, SQL Server permite al usuario equilibrar el tamaño del índice (después de la creación inicial) respecto a los costes de mantenimiento a medida que se realizan cambios posteriores a la tabla.

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20
```

Otros comentarios

- Si ya existe un índice con el nombre especificado, la creación del índice fallará. Cualquier fallo se tratará en principio como una advertencia, lo que permite que se creen los índices siguientes y que después se vuelvan a informar como errores en el registro de mensajes una vez que se hayan intentado todos los índices.
- Para un rendimiento óptimo, los índices se deberían crear una vez que se hayan cargado los datos en la tabla. Los índices deben contener al menos una columna.
- Antes de ejecutar el nodo, puede obtener una presentación preliminar del SQL generado en el registro de mensajes.
- Las opciones que permiten especificar las claves primarias y los índices no están disponibles para las tablas temporales escritas en la base de datos (es decir, cuando el almacenamiento en caché del nodo está activo). Sin embargo, el sistema puede crear índices en la tabla temporal según sea necesario dependiendo de la manera en que se usen los datos en los nodos posteriores en la ruta. Por ejemplo, si los datos en caché se unen posteriormente mediante una columna *DEPT*, tendría sentido indexar la tabla en caché en dicha columna.

Optimización de consultas e índices

En algunos sistemas de gestión de bases de datos, una vez que se ha creado, cargado e indexado una base de datos, es necesario un paso más para que el optimizador pueda utilizar los índices para acelerar la ejecución de consultas en la tabla nueva. Por ejemplo, en Oracle, el optimizador de consultas basado en costes necesita que se analice una tabla antes de que se puedan utilizar los índices en la optimización de consultas. El archivo de propiedades interno ODBC de Oracle (no visible para el usuario) contiene una opción para que esto suceda, de la siguiente manera:

```
# Define el SQL que se ejecutará después de que se hayan creado una tabla
#y todos los números de índices asociados
y se haya rellenado
table_analysis_sql, 'ANALYZE TABLE <table-name> COMPUTE STATISTICS'
```

Este paso se ejecuta cada vez que se crea una tabla en Oracle (independientemente de que se definan índices o claves primarias). Si fuera necesario, el archivo de propiedades ODBC para bases de datos adicionales se puede personalizar de una forma parecida. Póngase en contacto con la asistencia técnica.

Opciones avanzadas de exportación de base de datos

Al pulsar en el botón Avanzado en los cuadros de diálogo del nodo de exportación de la base de datos, se visualiza un nuevo cuadro de diálogo en el que puede especificar los detalles técnicos para exportar resultados a una base de datos.

Utilizar confirmación por lotes. Para desactivar las confirmaciones fila por fila en la base de datos.

Tamaño de lote. Especifica el número de registros para enviar a la base de datos antes de confirmar en la memoria. Al reducir este número se proporciona una integridad de datos mayor al precio de velocidades de transferencia más lentas. Puede que desee ajustar este número para que su base de datos tenga un rendimiento óptimo.

Opciones de InfoSphere Warehouse. Solamente se muestra si está conectado a una base de datos de InfoSphere Warehouse (IBM DB2 9.7 o posterior). **No registrar actualizaciones** le permite evitar eventos de registro al crear tablas e insertar datos.

Utilizar carga masiva. Especifica un método para la carga masiva de datos a la base de datos directamente desde IBM SPSS Modeler. Es necesario algo de experimentación para seleccionar las opciones de carga masiva que son apropiadas para un escenario determinado.

- **A través de ODBC.** Seleccione utilizar la API de ODBC para ejecutar inserciones de varias filas con mayor eficacia que la exportación normal a la base de datos. Seleccione entre enlace a lo largo de las filas o de las columnas en las opciones siguientes.
- **A través de un cargador externo.** Seleccione utilizar un programa cargador masivo específico para su base de datos. Si selecciona esta opción se activan varias de las opciones siguientes.

Opciones avanzadas de ODBC. Estas opciones sólo están disponibles cuando **A través de ODBC** está seleccionado. Tenga en cuenta que puede que no todos los controladores ODBC admitan esta funcionalidad.

- **A lo largo de las filas.** Seleccione un enlace a lo largo de las filas para utilizar la llamada `SQLBulkOperations` para cargar datos a la base de datos. El enlace a lo largo de las filas suele mejorar la velocidad en comparación con el uso de inserciones parametrizadas que insertan datos según una base de registro por registro.
- **A lo largo de las columnas.** Seleccione utilizar un enlace a lo largo de las columnas para cargar datos a la base de datos. Un enlace a lo largo de las columnas mejora el rendimiento enlazando cada columna de la base de datos (en una instrucción `INSERT` parametrizada) con una matriz de valores N . Al ejecutar la instrucción `INSERT` una vez se insertan N filas en la base de datos. Este método puede aumentar el rendimiento de forma espectacular.

Opciones del cargador externo. Cuando se especifica **A través de un cargador externo**, se muestran varias opciones para exportar el conjunto de datos a un archivo y especificar y ejecutar un programa cargador personalizado para cargar los datos desde ese archivo a la base de datos. IBM SPSS Modeler puede interactuar con los cargadores externos para muchos sistemas de base de datos populares. Se han incluido varios scripts con el software que están disponibles junto con documentación técnica en el subdirectorio *scripts*. Tenga en cuenta que para utilizar esta funcionalidad, Python 2.7 debe estar instalado en el mismo ordenador que IBM SPSS Modeler o IBM SPSS Modeler Server y el parámetro `python_exe_path` debe estar establecido en el archivo *options.cfg*. Consulte el tema “Programación del cargador masivo” en la página 348 para obtener más información.

- **Utilizar delimitador.** Especifica qué carácter delimitador debe utilizarse en el archivo exportado. Seleccione **Tabulador** para delimitar con el tabulador y **Espacio** para delimitar con espacios. Seleccione **Otros** para especificar otro carácter, como una coma (,).
- **Especificar archivo de datos.** Para introducir la ruta que se va a utilizar para el archivo de datos escrito durante la carga masiva. De forma predeterminada, se crea un archivo temporal en el directorio temporal del servidor.

- **Especificar programa cargador.** Para especificar el programa de carga masiva. De forma predeterminada, el software busca el subdirectorio *scripts* de la instalación de IBM SPSS Modeler en un script Python para ejecutarlo para una base de datos determinada. Se han incluido varios scripts con el software que están disponibles junto con documentación técnica en el subdirectorio *scripts*.
- **Generar registro.** Para generar un archivo de registro en el directorio especificado. El archivo de registro contiene información de error y es útil si la operación de carga masiva falla.
- **Comprobar el tamaño de la tabla.** Para realizar una comprobación de tabla que garantice que el aumento del tamaño de la tabla se corresponda con el número de filas exportadas desde IBM SPSS Modeler.
- **Opciones adicionales del cargador.** Especifica argumentos adicionales para el programa cargador. Utilice comillas para los argumentos que contengan espacios.

Las comillas se incluyen en los argumentos opcionales con una barra inclinada invertida. Por ejemplo, la opción especificada como `-comment "Esto es un \"comentario\""` incluye tanto el distintivo `-comment` como el comentario en sí representado como `Esto es un "comentario"`.

Se puede incluir una sola barra inclinada invertida utilizando otra barra inclinada invertida. Por ejemplo, la opción especificada como `-specialdir "C:\\Test Scripts\\"` incluye el distintivo `-specialdir` y el directorio representado como `C:\Test Scripts\`.

Programación del cargador masivo

El nodo Exportar base de datos incluye opciones de carga masiva en el cuadro de diálogo Opciones avanzadas. Los programas de cargadores masivos se pueden utilizar para cargar datos desde un archivo de texto a una base de datos.

La opción **Utilizar carga masiva (a través de un cargador externo)** configura IBM SPSS Modeler para que realice tres cosas:

- Crear cualquier tabla de base de datos necesaria.
- Exportar los datos a un archivo de texto.
- Invocar un programa cargador masivo para cargar los datos desde dicho archivo a la tabla de la base de datos.

Normalmente, el programa cargador masivo no es la propia utilidad de carga de la base de datos (por ejemplo, la utilidad `sqlldr` de Oracle), sino un pequeño script o programa que forma los argumentos correctos, crea cualquier archivo auxiliar específico de la base de datos (como un archivo de control) y, a continuación, invoca la utilidad de carga de la base de datos. La información de las siguientes secciones le ayudará a editar un cargador masivo existente.

Si lo prefiere, puede escribir su propio programa de carga masiva. Consulte el tema “Desarrollo de programas de carga masiva” en la página 352 para obtener más información. Tenga en cuenta que esto no está cubierto bajo un acuerdo de soporte técnico estándar y debe ponerse en contacto con un representante de servicios de IBM para obtener ayuda.

Scripts de carga masiva

IBM SPSS Modeler se entrega con varios programas cargadores masivos para bases de datos diferentes que se implementan utilizando scripts Python. Al ejecutar una ruta que contenga un nodo Exportar base de datos con la opción **A través de un cargador externo** seleccionada, IBM SPSS Modeler crea una tabla de base de datos (en caso necesario) a través de ODBC, exporta los datos a un archivo temporal en el host que ejecuta IBM SPSS Modeler Server y, a continuación, invoca el script de carga masiva. A su vez, el script ejecuta utilidades proporcionadas por el proveedor de DBMS para cargar datos de los archivos temporales a la base de datos.

Note: La instalación de IBM SPSS Modeler no incluye un intérprete de tiempo de ejecución de Python, por lo que se requiere una instalación independiente de Python. Consulte el tema “Opciones avanzadas de exportación de base de datos” en la página 347 para obtener más información.

Se proporcionan scripts (en la carpeta `\scripts` del directorio de instalación de IBM SPSS Modeler) para las bases de datos indicadas en la tabla siguiente.

Tabla 58. Scripts de cargador masivo proporcionados.

Base de datos	Nombre de script	Más información
IBM DB2	<code>db2_loader.py</code>	Consulte el tema “Carga masiva de datos a bases de datos IBM DB2” para obtener más información.
IBM Netezza	<code>netezza_loader.py</code>	Consulte el tema “Carga masiva de datos a bases de datos IBM Netezza” en la página 350 para obtener más información.
Oracle	<code>oracle_loader.py</code>	Consulte el tema “Carga masiva de datos a bases de datos Oracle” en la página 350 para obtener más información.
SQL Server	<code>mssql_loader.py</code>	Consulte el tema “Carga masiva de datos a bases de datos SQL Server” en la página 351 para obtener más información.
Teradata	<code>teradata_loader.py</code>	Consulte el tema “Carga masiva de datos a bases de datos Teradata” en la página 352 para obtener más información.

Carga masiva de datos a bases de datos IBM DB2

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM SPSS Modeler a una base de datos IBM DB2 utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Asegúrese de que la utilidad de procesador de línea de comandos (CLP) de DB2 está instalada

El script `db2_loader.py` invoca el comando DB2 LOAD. Asegúrese de que el procesador de línea de comandos (`db2` en UNIX, `db2cmd` en Windows) está instalado en el servidor en el que se ejecuta `db2_loader.py` (normalmente, el host que ejecuta IBM SPSS Modeler Server).

Compruebe si el nombre de alias de la base de datos local coincide con el nombre de la base de datos real

El alias de la base de datos local DB2 es el nombre usado por el software cliente de DB2 para hacer referencia a una base de datos en una instancia de DB2 local o remota. Si el alias de la base de datos local no es igual que el nombre de la base de datos remota, proporcione la opción adicional del cargador:

```
-alias <local_database_alias>
```

Por ejemplo, la base de datos remota se llama STARS en el host GALAXY, pero el alias de la base de datos local DB2 en el host que ejecuta IBM SPSS Modeler Server es STARS_GALAXY. Utilice la opción adicional del cargador

```
-alias STARS_GALAXY
```

Codificación de datos con caracteres que no son ASCII

Si está realizando una carga masiva de datos que no están en formato ASCII, debe asegurarse de que la variable de página de códigos en la sección de configuración de `db2_loader.py` está configurada correctamente en su sistema.

Cadenas en blanco

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

Carga masiva de datos a bases de datos IBM Netezza

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM SPSS Modeler a una base de datos IBM Netezza utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Asegúrese de que la utilidad *nzload* de Netezza está instalada

El script *netezza_loader.py* invoca la utilidad *nzload* de Netezza. Asegúrese de que *nzload* está instalado y se ha configurado correctamente en el servidor en el que se va a ejecutar *netezza_loader.py*.

Exportación de datos que no están en ASCII

Si exporta contenido que no está en formato ASCII, puede que necesite añadir `-encoding UTF8` al campo de **opciones adicionales del cargador** en el cuadro de diálogo Exportar BD: Opciones avanzadas. De este modo, se asegura que se cargan correctamente los datos que no están en ASCII.

Datos de formato de fecha, hora y marca de tiempo

En las propiedades de ruta, establezca el formato de fecha **DD-MM-AAAA** y el formato de hora **HH:MM:SS**.

Cadenas en blanco

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

Orden de columnas diferente en la tabla de destino y la ruta al insertar datos en una tabla existente

Si el orden de las columnas de la ruta es distinto al de la tabla de destino, los valores de datos se insertarán en columnas equivocadas. Utilice el nodo Reorg. campos para garantizar que el orden de las columnas de la ruta coincide con el orden de la tabla de destino. Consulte el tema “Nodo Reorg. campos” en la página 183 para obtener más información.

Seguimiento del progreso de *nzload*

Cuando ejecute IBM SPSS Modeler en modo local, añada `-sts` al campo de **opciones adicionales del cargador**, en el cuadro de diálogo Exportar BD: Opciones avanzadas, para ver mensajes de estado cada 10.000 filas en la ventana de comando abierta por la utilidad *nzload*.

Carga masiva de datos a bases de datos Oracle

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM SPSS Modeler a una base de datos Oracle utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Asegúrese de que la utilidad *sqlldr* de Oracle está instalada

El script *oracle_loader.py* invoca la utilidad *sqlldr* de Oracle. Tenga en cuenta que *sqlldr* no se incluye automáticamente en el cliente de Oracle. Asegúrese de que *sqlldr* está instalado en el servidor en el que se va a ejecutar *oracle_loader.py*.

Especifique el SID o el nombre de servicio de la base de datos

Si está exportando datos a un servidor de Oracle no local o su servidor de Oracle local tiene múltiples bases de datos, debe especificar lo siguiente en el campo de **opciones adicionales del cargador**, en el cuadro de diálogo Exportar BD: Opciones avanzadas, para pasar el SID o nombre de servicio:

`-database <SID>`

Edición de la sección de configuración en `oracle_loader.py`

En sistemas UNIX (opcional en Windows), edite la sección de configuración al inicio del script `oracle_loader.py`. Aquí, los valores para las variables de entorno `ORACLE_SID`, `NLS_LANG`, `TNS_ADMIN` y `ORACLE_HOME` se pueden especificar, si procede, así como la ruta completa de la utilidad `sqlldr`.

Datos de formato de fecha, hora y marca de tiempo

En las propiedades de ruta, normalmente debe establecer el formato de fecha **AAAA-MM-DD** y el formato de hora **HH:MM:SS**.

Si necesita utilizar formatos de fecha y hora distintos a los anteriores, consulte la documentación de Oracle y edite el archivo de script `oracle_loader.py`.

Codificación de datos con caracteres que no son ASCII

Si está realizando una carga masiva de datos que no están en formato ASCII, debe asegurarse de que la variable de entorno `NLS_LANG` está configurada correctamente en su sistema. Esto es leído por la utilidad de cargador `sqlldr` de Oracle. Por ejemplo, el valor correcto de `NLS_LANG` para Shift-JIS en Windows es `Japanese_Japan.JA16SJIS`. Si necesita más información acerca de `NLS_LANG`, consulte la documentación de Oracle.

Cadenas en blanco

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

Carga masiva de datos a bases de datos SQL Server

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM SPSS Modeler a una base de datos SQL Server utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Asegúrese de que la utilidad `bcp.exe` de SQL Server está instalada

El script `mssql_loader.py` invoca la utilidad `bcp.exe` de SQL Server. Asegúrese de que `bcp.exe` está instalado en el servidor donde se va a ejecutar `mssql_loader.py`.

Los espacios no sirven como delimitadores

Evite elegir espacios como delimitadores en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Se recomienda seleccionar la opción **Comprobar el tamaño de la tabla**

Le recomendamos que active la opción **Comprobar tamaño de la tabla** en el cuadro de diálogo Exportar BD: Opciones avanzadas. No siempre se detectan los errores en el proceso de carga masiva, y si se activa esta opción, se realiza una comprobación adicional para verificar que se ha cargado el número correcto de filas.

Cadenas en blanco

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

Especificar la instancia con nombre del servidor de SQL totalmente calificado

Puede haber casos en los que SPSS Modeler no pueda acceder al servidor SQL debido a un nombre de host no calificado y muestre el error siguiente:

Se ha detectado un error al ejecutar el cargador masivo externo.
Puede que el archivo de registro proporcione más detalles.

Para corregir este error, añada la cadena siguiente, incluyendo las comillas dobles, en el campo

Opciones adicionales del cargador:

```
"-S mhreboot.spss.com\SQLEXPRESS"
```

Carga masiva de datos a bases de datos Teradata

Los siguientes puntos pueden ayudarle a configurar la carga masiva de IBM SPSS Modeler a una base de datos Teradata utilizando la opción de cargador externo en el cuadro de diálogo Exportar BD: Opciones avanzadas.

Asegúrese de que la utilidad *fastload* de Teradata está instalada

El script *teradata_loader.py* invoca la utilidad *fastload* de Teradata. Asegúrese de que *fastload* se ha instalado y configurado correctamente en el servidor en el que se ejecutará *teradata_loader.py*.

Solo es posible realizar cargas masivas de datos a tablas vacías

Solo puede utilizar tablas vacías como destino de una carga masiva. Si una tabla de destino contiene datos previos a la carga masiva, la operación fallará.

Datos de formato de fecha, hora y marca de tiempo

En las propiedades de ruta, establezca el formato de fecha **AAAA-MM-DD** y el formato de hora **HH:MM:SS**.

Cadenas en blanco

Las cadenas en blanco se exportan a la base de datos como valores nulos (NULL).

ID de procesos de Teradata (tdpid)

De forma predeterminada, *fastload* exporta los datos al sistema Teradata con *tdpid=dbc*. Normalmente, habrá una entrada en el archivo HOSTS que asocia *dbccop1* con la dirección IP del servidor de Teradata. Para utilizar un servidor diferente, especifique lo siguiente en el campo de **opciones adicionales del cargador** del cuadro de diálogo Exportar BD: Opciones avanzadas para pasar el *tdpid* de este servidor:
`-tdpid <id>`

Espacios en nombres de columnas y tablas

Si el nombre de una columna o tabla contiene espacios, fallará la operación de carga masiva. Si es posible, cambie el nombre de la columna o tabla eliminando los espacios.

Desarrollo de programas de carga masiva

Aquí se explica cómo desarrollar un programa cargador masivo que se pueda ejecutar desde IBM SPSS Modeler para cargar datos de un archivo de texto a una base de datos. Tenga en cuenta que esto no está cubierto bajo un acuerdo de soporte técnico estándar y debe ponerse en contacto con un representante de servicios de IBM para obtener ayuda.

Utilización de Python para crear programas cargadores masivos

De forma predeterminada, IBM SPSS Modeler busca un programa cargador masivo predeterminado basado en el tipo de base de datos. Consulte la Tabla 58 en la página 349.

Puede utilizar el script *test_loader.py* para ayudar al desarrollo de programas cargadores por lotes. Consulte el tema “Prueba de programas de carga masiva” en la página 354 para obtener más información.

Objetos pasados al programa cargador masivo

IBM SPSS Modeler escribe dos archivos que se han pasado al programa cargador masivo.

- **Archivo de datos.** Contiene los datos que se tienen que cargar, en formato de texto.
- **Archivo de esquema.** Es un archivo XML que describe los nombres y tipos de las columnas, y proporciona información acerca del formato del archivo de datos (por ejemplo, qué carácter se emplea como delimitador entre campos).

Además, IBM SPSS Modeler pasa otra información, como el nombre de la tabla, el nombre de usuario y la contraseña como argumentos al invocar el programa cargador masivo.

Note: Para señalar la finalización satisfactoria a IBM SPSS Modeler, el programa cargador masivo debería eliminar el archivo de esquema.

Argumentos pasados al programa cargador masivo

Los argumentos pasados al programa se muestran en la siguiente tabla.

Tabla 59. Argumentos pasados al cargador masivo.

Argumento	Descripción
schemafile	Ruta del archivo de esquema.
data file	Ruta del archivo de datos.
servername	Nombre de servidor de DBMS (puede estar en blanco).
databasename	Nombre de la base de datos dentro del servidor de DBMS (puede estar en blanco).
username	Nombre de usuario para acceder a la base de datos.
password	Contraseña para acceder a la base de datos.
tablename	Nombre de la tabla para cargar.
ownername	Nombre del propietario de la tabla (también se denomina nombre de esquema).
logfile name	Nombre del archivo de registro (si está en blanco, no se crea ningún archivo de registro).
rowcount	Número de filas del conjunto de datos.

Cualquier opción especificada en el campo **Opciones adicionales del cargador** del cuadro de diálogo Exportar BD: Opciones avanzadas se pasa al programa cargador masivo tras estos argumentos estándar.

Formato del archivo de datos

Los datos se escriben en el archivo de datos con formato de texto; cada campo se separa con un carácter delimitador especificado en el cuadro de diálogo Exportar BD: Opciones avanzadas. A continuación se muestra un ejemplo del aspecto de un archivo delimitado por tabulaciones.

```
48 F HIGH NORMAL 0.692623 0.055369 drugA
15 M NORMAL HIGH 0.678247 0.040851 drugY
37 M HIGH NORMAL 0.538192 0.069780 drugA
35 F HIGH HIGH 0.635680 0.068481 drugA
```

El archivo se escribe con la codificación local utilizada por IBM SPSS Modeler Server (o IBM SPSS Modeler si no está conectado a IBM SPSS Modeler Server). Algunos formatos están controlados a través de la configuración de la ruta de IBM SPSS Modeler.

Formato del archivo de esquema

El archivo de esquema es un archivo XML que describe el archivo de datos. A continuación se muestra un ejemplo que acompaña al archivo de datos previo.

```

<?xml version="1.0" encoding="UTF-8"?>
<DBSCHEMA version="1.0">
  <table delimiter="\t" commit_every="10000" date_format="YYYY-MM-DD" time_format="HH:MM:SS"
append_existing="false" delete_datafile="false">
  <column name="Age" encoded_name="416765" type="integer"/>
  <column name="Sex" encoded_name="536578" type="char" size="1"/>
  <column name="BP" encoded_name="4250" type="char" size="6"/>
  <column name="Cholesterol" encoded_name="43686F6C65737465726F6C" type="char" size="6"/>
  <column name="Na" encoded_name="4E61" type="real"/>
  <column name="K" encoded_name="4B" type="real"/>
  <column name="Drug" encoded_name="44727567" type="char" size="5"/>
</table>
</DBSCHEMA>

```

Las dos tablas siguientes enumeran los atributos de los elementos <table> y <column> del archivo de esquema.

Tabla 60. Atributos del elemento <table>.

Atributo	Descripción
delimiter	Carácter delimitador de campo (TAB se representa como \t).
commit_every	Intervalo de tamaño de lote (como en el cuadro de diálogo Exportar BD: Opciones avanzadas).
date_format	Formato usado para representar fechas.
time_format	Formato usado para representar horas.
append_existing	true si la tabla que debe cargarse ya contiene datos; false de lo contrario.
delete_datafile	true si el programa de carga masiva debe eliminar el archivo de datos al completar la carga.

Tabla 61. Atributos del elemento <column>.

Atributo	Descripción
name	Nombre de columna.
encoded_name	El nombre de columna convertido a la misma codificación que el archivo de datos y la salida como una serie de números hexadecimales de dos dígitos.
type	El tipo de datos de la columna: integer, real, char, time, date o datetime.
size	Para el tipo de datos char, la anchura máxima de la columna en caracteres.

Prueba de programas de carga masiva

Puede probar la carga masiva utilizando el script de prueba *test_loader.py* incluido en la carpeta *\scripts* del directorio de instalación de IBM SPSS Modeler. Esto resulta de utilidad al intentar desarrollar, depurar o solucionar problemas de los programas o scripts de carga masiva para su uso con IBM SPSS Modeler.

Para utilizar el script de prueba, siga estos pasos.

1. Ejecute el script *test_loader.py* para copiar los archivos de esquema y de datos a los archivos *schema.xml* y *data.txt*, y cree un archivo por lotes de Windows (*test.bat*).
2. Edite el archivo *test.bat* para seleccionar el programa o script de carga masiva que desea probar.
3. Ejecute *test.bat* desde un shell de comandos para probar el programa o script de carga masiva elegido.

Note: Al ejecutar *test.bat* no se carga ningún dato en la base de datos.

nodo de exportación Archivo sin formato

El nodo de exportación de archivo sin formato permite escribir datos en un archivo de texto delimitado. Esto es útil para exportar datos que se pueden leer con otro software de hoja de cálculo o de análisis.

Si los datos contienen información geoespacial, puede exportarlos como un archivo sin formato y, si genera un nodo de origen Archivo de variables para su uso dentro de la misma ruta, se conservan todos los metadatos geoespaciales, de almacenamiento y de medición en el nuevo nodo de origen. No obstante, si exporta los datos y luego los importa en una ruta distinta, debe realizar pasos adicionales para establecer los metadatos geoespaciales en el nuevo nodo de origen. Para obtener más información, consulte el tema “Nodo Archivo variable” en la página 25.

Nota: No puede escribir archivos con el formato de memoria caché antiguo, ya que IBM SPSS Modeler ya no utiliza dicho formato para almacenar archivos en memoria caché. Los archivos de memoria caché de IBM SPSS Modeler se guardan ahora en formato *.sav* de IBM SPSS Statistics, que puede escribir utilizando un nodo de exportación de Estadísticas. Para obtener más información, consulte el tema “Nodo Exportación de Statistics” en la página 373.

Pestaña Exportar del nodo Archivo sin formato

Exportar archivo. Especifica el nombre del archivo. Escriba el nombre del archivo o pulse en el botón de selección de archivos para acceder a la ubicación del archivo.

Modo escritura. Si selecciona **Sobrescribir**, se sobrescribirá cualquier dato existente en el archivo especificado. Si selecciona **Añadir**, el resultado se añadirá al final del archivo existente, conservando cualquier dato que contenga.

- **Incluir nombres de campo.** Si selecciona esta opción, los nombres de los campos se escribirán en la primera línea del archivo de resultados. Esta opción sólo está disponible para el modo escritura **Sobrescribir**.

Nueva línea después de cada registro. Si selecciona esta opción, cada registro se escribirá en una línea nueva del archivo de resultados.

Separador de campos. Especifica el carácter que desee insertar entre los valores de campo del archivo de texto generado. Las opciones son **Coma**, **Tabulador**, **Espacio** y **Otro**. Si selecciona **Otro**, introduzca los caracteres delimitadores que desee en el cuadro de texto.

Comillas para los símbolos. Especifica el tipo de comillas que vaya a utilizar para los valores de campos simbólicos. Las opciones son **Ninguna** (los valores no irán entre comillas), **Simples (')**, **Dobles (")** y **Otro**. Si selecciona **Otro**, introduzca los caracteres de entrecorillado que desee en el cuadro de texto.

Codificación. Especifica el método de codificación de textos utilizado. Puede elegir entre valor predeterminado del sistema, el valor predeterminado de la ruta o UTF-8.

- El valor predeterminado del sistema se especifica en el Panel de control de Windows o, si lo ejecuta en modo distribuido, en el equipo servidor.
- La ruta predeterminada se especifica en el cuadro de diálogo Propiedades de ruta.

Símbolo decimal. Especifica cómo deben representarse los decimales en los datos.

- **Valor predeterminado de la ruta.** Se utilizará el separador decimal definido por la configuración del valor predeterminado de la ruta actual. Este será normalmente el separador decimal definido por la configuración regional del ordenador.
- **Punto (.)**. Se utilizará el punto como separador decimal.
- **Coma (,)**. Se utilizará la coma como separador decimal.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen Archivo var. que leerá el archivo de datos exportado. Consulte el tema “Nodo Archivo variable” en la página 25 para obtener más información.

Nodos de exportación IBM SPSS Data Collection

El nodo de exportación IBM SPSS Data Collection guarda datos en el formato utilizado por el software de investigación de mercados IBM SPSS Data Collection, en función del modelo de datos de IBM SPSS Data Collection. Este formato distingue los datos de casos (las respuestas reales a las preguntas recogidas durante una encuesta), de los metadatos que describen cómo se recopilan y organizan los datos de casos. Los metadatos constan de información como textos de preguntas, nombres de variables y descripciones, conjuntos de respuestas múltiples, traducciones de los diversos textos y las definiciones de la estructura de los datos de casos. Consulte el tema “Data Collection Nodo” en la página 31 para obtener más información.

Note: este nodo requiere tener instalado IBM SPSS Data Collection Data Model versión 4.0 o posterior, que se distribuye con los productos de software de IBM SPSS Data Collection. Si desea obtener más información, consulte la página Web de IBM SPSS Data Collection en <http://www.ibm.com/software/analytics/spss/products/data-collection/>. Además de la instalación de Modelo de datos, no se requieren configuraciones adicionales.

Archivo de metadatos. Especifica el nombre del archivo de definición de cuestionario (*.mdd*) donde se guardarán los metadatos exportados. Se crea un cuestionario predeterminado basándose en la información sobre el tipo de campo. Por ejemplo, un campo nominal (conjunto) podría representarse como una única pregunta con la descripción de campo utilizada como el texto de la pregunta y una casilla de verificación independiente para cada valor definido.

Fundir metadatos. Especifica si los metadatos sobrescribirán las versiones existentes o se combinarán con los metadatos existentes. Si se selecciona la opción de combinación, se creará una nueva versión cada vez que se ejecute la ruta. Esto permite realizar un seguimiento de las versiones de un cuestionario a medida que se va modificando. Cada versión puede considerarse como una instantánea de los metadatos que se utilizan para recopilar un conjunto de datos de casos determinado.

Activar variables de sistema. Especifica si las variables de sistema se incluyen en el archivo *.mdd* exportado. Entre ellas se incluyen variables como *Respondent.Serial*, *Respondent.Origin* y *DataCollection.StartTime*.

Configuración de datos de casos. Especifica el archivo de datos IBM SPSS Statistics (*.sav*) donde se exportan los datos de casos. Tenga en cuenta que todas las restricciones sobre los nombres de variables y valores son aplicables, así que, por ejemplo, puede que necesite cambiar a la pestaña Filtro y utilizar la opción “Cambiar nombre para IBM SPSS Statistics” en el menú de opciones Filtro para corregir los caracteres que no son válidos en los nombres de campo.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen IBM SPSS Data Collection que leerá el archivo de datos exportado.

Conjuntos de respuestas múltiples. Cualquier conjunto de respuestas múltiples definido en la ruta se conservará automáticamente cuando se exporte el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.

Nodos de exportación Analytic Server

El nodo de exportación de Analytic Server permite escribir datos desde el análisis a un origen de datos de Analytic Server existente. Este puede ser, por ejemplo, archivos de texto en un sistema de archivos distribuido de Hadoop (HDFS) o una base de datos.

Una ruta con un nodo Exportar de Analytic Server suele comenzar también con nodos Origen de Analytic Server, y se envía a Analytic Server y se ejecuta en HDFS. De forma alternativa, una ruta con orígenes de datos "locales" puede terminar con un nodo Exportar de Analytic Server para cargar bloques de datos relativamente pequeños (de no más de 100.000 registros) para usarlos en Analytic Server.

Origen de datos. Seleccione el origen de datos que contenga los datos que desee. Un origen de datos contiene los archivos y los metadatos asociados a ese origen. Pulse **Seleccionar** para visualizar una lista de los orígenes de datos disponibles. Consulte el tema "Selección de un origen de datos" en la página 13 para obtener más información.

Si necesita crear un origen de datos o editar uno existente, pulse **Lanzar editor de orígenes de datos...**

Modo. Seleccione **Añadir** para añadir al origen de datos existente, o **Sobrescribir** para sustituir el contenido del origen de datos.

Generar un nodo Importación para estos datos. Selecciónelo para generar un nodo origen para los datos como se han exportado al origen de datos especificado. Este nodo se añadirá al lienzo de la ruta.

Nodo Exportar de IBM Cognos BI

El nodo Exportar de IBM Cognos BI le permite exportar datos desde una ruta de IBM SPSS Modeler a Cognos BI, en formato UTF-8. De este modo, Cognos BI puede utilizar los datos transformados o puntuados desde IBM SPSS Modeler. Por ejemplo, podría utilizar Cognos BI Report Studio para crear un informe basado en los datos exportados, incluidas las predicciones y los valores de confianza. El informe luego puede guardarse en el servidor de Cognos BI y distribuirse a usuarios de Cognos BI.

Note: Solamente puede exportar datos relacionales, no datos OLAP.

Para exportar datos a Cognos BI, debe especificar lo siguiente:

- Conexión de Cognos: la conexión al servidor de Cognos BI
- Conexión ODBC: la conexión al servidor de datos de Cognos que utiliza el servidor de Cognos BI

En la conexión Cognos especifica un origen de datos de Cognos que se utilizará. Este origen de datos debe utilizar el mismo inicio de sesión que el origen de datos ODBC.

Exporta los datos de ruta reales al servidor de datos y los metadatos de paquete al servidor de Cognos BI.

Al igual que con cualquier otro nodo de exportación, también puede utilizar la pestaña Publicar del cuadro de diálogo del nodo para publicar la ruta para su despliegue mediante IBM SPSS Modeler Solution Publisher.

Conexión de Cognos

Aquí es donde se especifica la conexión al servidor de Cognos BI que quiera utilizar para la exportación. El procedimiento implica la exportación de los metadatos a un nuevo paquete del servidor de Cognos BI, mientras que los datos de ruta se exportan al servidor de datos de Cognos.

Conexión. Pulse en el botón **Editar** para mostrar un cuadro de diálogo donde puede definir la URL y otros detalles del servidor de Cognos BI al que quiere exportar los datos. Si ya ha iniciado sesión en un servidor de Cognos BI mediante IBM SPSS Modeler, también puede editar los detalles de la conexión actual. Consulte el tema "Conexiones de Cognos" en la página 38 para obtener más información.

Origen de datos. El nombre del origen de datos de Cognos (normalmente una base de datos) a la que está exportando los datos. La lista desplegable muestra todos los orígenes de datos de Cognos a los que puede acceder desde la conexión actual. Pulse en el botón **Actualizar** para actualizar la lista.

Carpeta. La ruta y el nombre de la carpeta del servidor de Cognos BI donde se creará el paquete de exportación.

Nombre del paquete. El nombre del paquete de la carpeta especificada que debe contener los metadatos exportados. Éste debe ser un nuevo paquete con un único asunto de consulta; no puede exportar a un paquete existente.

Modo. Especifica cómo quiere realizar la exportación:

- **Publicar paquete ahora.** (valor predeterminado) Realiza la operación de exportación tan pronto como pulsa en **Ejecutar**.
- **Exportar script de acción.** Crea un script XML que puede ejecutar más tarde (por ejemplo, utilizando Framework Manager) para realizar la exportación. Escriba la ruta y el nombre de archivo del script en el campo **Archivo** o utilice el botón **Editar** para especificar el nombre y la ubicación del archivo de script.

Generar un nodo de importación para este dato. Seleccione esta opción para generar un nodo de origen para los datos tal y como se exportan en la tabla y el origen de datos especificados. Cuando pulsa en **Ejecutar**, este nodo se añade al lienzo de rutas.

Conexión ODBC

Aquí especifica la conexión al servidor de datos de Cognos (es decir, la base de datos) al que deben exportarse los datos de ruta.

Note: Debe asegurarse de que el origen de datos que especifique aquí apunta al mismo que se ha especificado en el panel **Conexiones de Cognos**. También debe asegurarse de que el origen de datos de conexión Cognos utiliza el mismo inicio de sesión que el origen de datos ODBC.

Origen de datos. Muestra el origen de datos seleccionado. Introduzca el nombre o selecciónelo de la lista desplegable. Si no ve en la lista la base de datos que desea, seleccione **Añadir nueva conexión de base de datos** y busque la base de datos desde el cuadro de diálogo Conexiones a la base de datos. Consulte el tema “Adición de una conexión de base de datos” en la página 20 para obtener más información.

Nombre de tabla. Introduzca el nombre de la tabla a la que desee enviar los datos. Si selecciona la opción **Insertar en tabla**, puede seleccionar una tabla existente de la base de datos pulsando en el botón **Seleccionar**.

Crear tabla. Seleccione esta opción para crear una nueva tabla de base de datos o para sobrescribir una tabla de base de datos existente.

Insertar en tabla. Seleccione esta opción para insertar los datos como nuevas filas en una tabla de base de datos existente.

Tabla Fusión. (Donde esté disponible) Seleccione esta opción para actualizar columnas de base de datos seleccionadas con valores de los campos de datos de origen correspondiente. Si selecciona esta opción activará el botón **Fusión**, que muestra un cuadro de diálogo donde podrá correlacionar campos de orígenes de datos a columnas de base de datos.

Eliminar tabla existente. Seleccione esta opción para eliminar cualquier tabla existente con el mismo nombre al crear una tabla nueva.

Eliminar filas existentes. Seleccione esta opción para eliminar filas existentes de la tabla antes de exportar cuando se inserte en una tabla.

Note: si está seleccionada alguna de las dos opciones anteriores, recibirá el mensaje **Advertencia de sobrescritura** cuando ejecute el nodo. Para suprimir las advertencias, anule la selección de **Avisar cuando un nodo sobrescriba una tabla de la base de datos** en la pestaña Notificaciones del cuadro de diálogo Opciones de usuario.

Tamaño de cadena predeterminado. Los campos marcados como sin tipo en un nodo Tipo anterior de la ruta se escriben en la base de datos como campos de cadena. Especifique el tamaño de las cadenas que se van a utilizar para campos sin tipo.

Pulse en **Esquema** para abrir un cuadro de diálogo donde podrá establecer diversas opciones de exportación (para bases de datos que admitan esta característica), establecer los tipos de datos de SQL para los campos y especificar la clave primaria para la indización de la base de datos. Consulte el tema “Opciones de Esquema de exportación de base de datos” en la página 342 para obtener más información.

Pulse en **Índices** para especificar las opciones de indización de la tabla exportada para mejorar el rendimiento de la base de datos. Consulte el tema “Opciones de índice de exportación de base de datos” en la página 345 para obtener más información.

Pulse en **Avanzado** para especificar opciones de carga masiva y de confirmación de base de datos. Consulte el tema “Opciones avanzadas de exportación de base de datos” en la página 347 para obtener más información.

Poner entre comillas nombres de columna y tabla. Seleccione opciones utilizadas al enviar la instrucción CREATE TABLE a la base de datos. Las tablas o columnas con espacios o caracteres que no sean estándar deben ponerse entre comillas.

- **Cuando sea necesario.** Selecciónelo para permitir que IBM SPSS Modeler determine automáticamente cuándo son necesarias las comillas de forma individual.
- **Siempre.** Para poner siempre entre comillas los nombres de tabla y columna.
- **Nunca.** Para desactivar la utilización de comillas.

Generar un nodo de importación para este dato. Seleccione esta opción para generar un nodo de origen para los datos tal y como se exportan en la tabla y el origen de datos especificados. Cuando pulsa en **Ejecutar**, este nodo se añade al lienzo de rutas.

Nodo Exportar de IBM Cognos TM1

El nodo Exportar de IBM Cognos BI permite exportar datos desde una ruta de IBM SPSS Modeler a Cognos TM1. De este modo, Cognos BI puede utilizar los datos transformados o puntuados desde IBM SPSS Modeler.

Nota: Solo pueden exportarse medidas, no datos de dimensión de contexto; de forma alternativa, pueden añadirse nuevos elementos al cubo.

Para exportar datos a Cognos BI, debe especificar lo siguiente:

- La conexión con el servidor de Cognos TM1.
- El cubo al que se van a exportar los datos.
- La correlación entre los nombres de datos de SPSS y las dimensiones y medidas de TM1 equivalentes.

Nota: El usuario de TM1 necesita los permisos siguientes: Privilegio de escritura de cubos, Privilegio de lectura de procesos SPSS, Privilegio de lectura de dimensiones, Privilegio de escritura de elementos de dimensión y Privilegio de escritura de archivos .b1b.

Al igual que con cualquier otro nodo de exportación, también puede utilizar la pestaña Publicar del cuadro de diálogo del nodo para publicar la ruta para su despliegue mediante IBM SPSS Modeler Solution Publisher.

Nota: Para poder utilizar los nodos de origen o exportación de TM1 en SPSS Modeler, debe copiar los siguientes tres procesos desde Modeler al servidor de TM1: `ExportToSPSS.pro`, `ImportFromSPSS.pro` y `SPSSCreateNewMeasures.pro`. Para añadir estos procesos al servidor TM1, debe copiarlos en el directorio de datos del servidor de TM1 y reiniciar el servidor de TM1. Estos archivos están disponibles desde el directorio siguiente: `<directorio de instalación de Modeler>/ext/bin/pasw.tm1/scripts`.

Conexión con un cubo de IBM Cognos TM1 para exportar datos

El primer paso de la exportación de datos a una base de datos de IBM Cognos TM1 es seleccionar el sistema de gestión de rendimiento TM1 relevante, así como el servidor y el cubo asociados en la pestaña **Conexión** del cuadro de diálogo IBM Cognos TM1.

Nota: Sólo se descartarán los valores "null" reales al exportar los datos a TM1. Los valores cero (0) se exportarán como valores válidos. Tenga en cuenta también que únicamente los campos que tienen un tipo de almacenamiento de *cadena* pueden correlacionarse con dimensiones en la pestaña **Correlación**. Antes de exportar a TM1, debe utilizar el cliente de IBM SPSS Modeler para convertir los tipos de datos que no sean cadena a cadena.

Sistema PM. Escriba el URL del sistema de gestión del rendimiento en el que está instalado el servidor TM1 al que desea conectarse. El sistema de gestión del rendimiento se define como un URL único para todos los servidores TM1. Desde este URL, es posible descubrir y acceder a todos los servidores IBM Cognos TM1 que se han instalado y se están ejecutando en el entorno.

Servidor TM1. Una vez que haya establecido la conexión con el sistema de gestión del rendimiento, seleccione el servidor que contiene los datos que desee importar y pulse **Iniciar sesión**. Si no se ha conectado anteriormente a este servidor, se le solicitará que especifique el **Nombre de usuario** y la **Contraseña**; de forma alternativa, puede seleccionar otro servidor.

Seleccionar un cubo TM1 para exportar. Muestra el nombre de los cubos del servidor TM1 donde puede exportar datos.

Para seleccionar los datos que se han de exportar, seleccione el cubo y pulse la flecha derecha para mover la vista al campo **Exportar a cubo**. Cuando haya seleccionado el cubo, utilice la pestaña **Correlación** para correlacionar las dimensiones y medidas de TM1 con los campos relevantes de SPSS o con un valor fijo (operación *Porción*).

Correlación de datos de IBM Cognos TM1 para exportación

Después de seleccionar el sistema de gestión de rendimiento de IBM Cognos TM1 y el servidor y cubo TM1 asociados, utilice la pestaña **Correlación** del cuadro de diálogo **Exportación de IBM Cognos TM1** para correlacionar dimensiones y medidas TM1 en campos SPSS o dimensiones TM1 de porción en un valor fijo.

Nota: Sólo los campos con el tipo de almacenamiento *string* pueden correlacionarse con dimensiones. Antes de exportar a TM1, debe utilizar el cliente de IBM SPSS Modeler para convertir en series los tipos de datos distintos de serie.

Campos. Lista los nombres de campos de datos del archivo de datos SPSS que están disponibles para la exportación.

Dimensiones TM1. Muestra el cubo TM1 seleccionado en la pestaña **Conexión** junto con sus dimensiones regulares y de medida. Seleccione el nombre de la dimensión TM1 según sea para correlación con el campo de datos de SPSS o para porción en un valor fijo. Por cubos con muchos elementos, puede ser útil

utilizar los botones en la parte inferior de la lista para filtrar las dimensiones que se muestran. Por ejemplo, puede mostrar todas las dimensiones, solo las dimensiones regulares o solo las dimensiones de medida.

Las siguientes opciones están disponibles en la pestaña Correlación.

Crear porción en. Los campos que se han dividido están etiquetadas con una **S**. *Crear porción en* es un término utilizado por IBM Cognos TM1 para fijar una dimensión en un elemento específico. Vea el ejemplo siguiente de datos importados desde TM1. Observe que las tres primeras columnas tienen valores que se repiten para cada fila (por ejemplo, *región* tiene el valor fijo *mundo* en cada fila). En lugar de exportar estos datos a TM1 para cada fila de datos, puede utilizar la característica **Crear porción en** para fijar el valor de la dimensión (*región* en este ejemplo) en el elemento denominado *mundo*.

Tabla 62. Ejemplo de creación de porción

región	actvsbud	modelo	mes	unidades	margen bruto	ventas	costes variables
mundo	varianza	A	Enero	72	-117,66	-157,11	-39,45
mundo	varianza	A	Febrero	4	-72,09	360,29	432,38
mundo	varianza	A	Marzo	-4	-539,21	45,52	584,73

Cuando selecciona una dimensión y pulsa **Crear porción en**, aparece un diálogo que muestra los miembros de hoja de la dimensión seleccionada. Solo puede crear porciones en miembros de hoja.

Correlacionar. Correlaciona el campo de datos de SPSS seleccionado en la dimensión o medida TM1 seleccionada (una dimensión regular o una medida específica de la dimensión de medida). Los campos correlacionados se etiquetan con una **M**.

Cancelar correlación. Anula la correlación del campo de datos de SPSS seleccionado de la dimensión o medida TM1 seleccionada. Tenga en cuenta que sólo puede anular una correlación individual a la vez. El campo de datos de SPSS correlacionado se moverá de nuevo a la columna de la izquierda.

Crear nueva. Crea una nueva medida en la dimensión de medida TM1. Aparecerá un diálogo que le permite especificar un nombre para la nueva medida TM1. Esta opción sólo está disponible para dimensiones de medida, no para dimensiones regulares.

Para obtener más información acerca de TM1, consulte la documentación de IBM Cognos Insight en <http://pic.dhe.ibm.com/infocenter/ctm1/v10r2m0/index.jsp>.

Nodo Exportar SAS

Note: Esta característica está disponible en SPSS Modeler Professional y SPSS Modeler Premium.

El nodo Exportar SAS permite escribir datos en formato SAS para leerlos en SAS o en un paquete de software compatible con SAS. Puede exportar en tres formatos de archivo SAS: SAS para Windows/OS2, SAS para UNIX o SAS versión 7/8.

Pestaña Exportar del nodo Exportar SAS

Exportar archivo. Especifique el nombre del archivo. Escriba el nombre del archivo o pulse en el botón de selección de archivos para acceder a la ubicación del archivo.

Exportar. Especifique el formato del archivo de exportación. Las opciones son **SAS para Windows/OS2**, **SAS para UNIX** o **SAS versión 7/8**.

Exportar nombres de campo como variables. Seleccione opciones para exportar nombres de campos y etiquetas desde IBM SPSS Modeler con objeto de utilizarlos con SAS.

- **nombres.** Para exportar los nombres y las etiquetas de campo de IBM SPSS Modeler. Los nombres se exportan como nombres de variables de SAS, mientras que las etiquetas se exportan como etiquetas de variables de SAS.
- **etiquetas.** Seleccione esta opción para utilizar los nombres de campos de IBM SPSS Modeler como etiquetas de variables en SAS. IBM SPSS Modeler permite caracteres en nombres de campos que no son válidos en nombres de variables de SAS. Para evitar la posibilidad de crear nombres SAS no válidos, seleccione **Nombres** en su lugar.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen SAS que leerá el archivo de datos exportado. Consulte el tema “Nodo de origen SAS” en la página 41 para obtener más información.

Nodo de exportación Excel

El nodo de exportación Excel ofrece los datos resultantes en formato de Microsoft Excel (.xls). Si lo desea, puede elegir iniciar automáticamente Excel y abrir el archivo exportado cuando se ejecute el nodo.

Pestaña Exportar del nodo Excel

Nombre del archivo. Escriba el nombre del archivo o pulse en el botón de selección de archivos para acceder a la ubicación del archivo. El nombre de archivo predeterminado es *excelexp.xls*.

Tipo de archivo. Seleccione el tipo de archivo de Excel que está exportando.

Crear archivo nuevo. Crea un nuevo archivo de Excel.

Insertar en el archivo existente. El contenido se sustituye comenzando en la casilla que designa el campo **Iniciar en celda**. El resto de casillas de la hoja de cálculo se quedan con su contenido original.

Incluir nombres de campo. Determina si los nombres de campos deben incluirse en la primera fila de la hoja de trabajo.

Iniciar en celda. La ubicación de la casilla se utiliza para el primer registro de exportación (o el nombre del primer campo si **Incluir nombres de campo** está seleccionado). Los datos se cumplimentan a la derecha y abajo desde la casilla inicial.

Seleccionar hoja de trabajo. Especifica la hoja de trabajo en la que desea exportar los datos. Puede identificar la hoja de datos por índice o por nombre:

- **Por índice.** Si crea un nuevo archivo, especifique un número entre 0 y 9 para identificar la hoja de trabajo a la que desea exportar, comenzando por 0 para la primera hoja de trabajo, 1 para la segunda, etcétera. Puede utilizar los valores de 10 o más si ya existe una única hoja de trabajo en esta posición.
- **Por nombre.** Si crea un nuevo archivo, especifique el nombre que se utiliza para la hoja de trabajo. Si lo inserta en un archivo existente, los datos se insertan en esta hoja de trabajo si existe, de lo contrario se crea una nueva hoja de trabajo con este nombre.

Iniciar Excel. Especifica si Excel se inicia automáticamente en el archivo exportado cuando se ejecuta el nodo. Tenga en cuenta que al trabajar en modo distribuido respecto a IBM SPSS Modeler Server, el resultado se guarda en el sistema de archivos del servidor y Excel se inicia en el cliente con una copia del archivo exportado.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen Excel que leerá el archivo de datos exportado. Consulte el tema “Nodo de origen de Excel” en la página 42 para obtener más información.

Nodo de exportación XML

El nodo Exportar XML permite obtener datos de resultados en formato XML, utilizando codificación UTF-8. También puede crear un nodo de origen XML para leer los datos exportados a la ruta.

Archivo de exportación XML. La ruta completa y el nombre del archivo XML al que desea exportar los datos.

Utilizar esquema XML. Seleccione esta casilla de verificación si desea utilizar un esquema o DTD para controlar la estructura de los datos exportados. Al hacerlo se activa el botón **Mapa** que se describe a continuación.

Si no utiliza un esquema o DTD, se utiliza la siguiente estructura para los datos exportados:

```
<records>
  <registro>
    <nombrecampo1>valor</nombrecampo1>
    <nombrecampo2>valor</nombrecampo2>
    :
    <nombrecampoN>valor</nombrecampoN>
  </record>
<record>
  :
  :
</record>
  :
  :
</records>
```

Los espacios en un nombre de campo se sustituyen por guiones bajos; por ejemplo, "Mi campo" es <Mi_campo>.

Correlacionar. Si ha optado por utilizar un esquema XML, este botón abre un cuadro de diálogo donde puede especificar la parte de la estructura XML que se debe utilizar para iniciar cada nuevo registro. Consulte el tema "Opciones de correlación de registros en XML" en la página 364 para obtener más información.

Campos correlacionados. Indica el número de campos que se han correlacionado.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen de XML que leerá el archivo de datos exportado en la ruta. Consulte el tema "Nodo Origen XML" en la página 43 para obtener más información.

Escritura de datos XML

Si se especifica un elemento XML, el valor del campo se coloca en la etiqueta del elemento:

```
<elemento>valor</elemento>
```

Cuando se correlaciona un atributo, el valor del campo se coloca como el valor del atributo:

```
<elemento atributo="valor">
```

Si un campo se correlaciona a un elemento por encima del elemento <records>, el campo sólo se asigna una vez y será una constante para todos los registros. El valor de este elemento proviene del primer registro.

Si se va a escribir un valor nulo, se realiza especificando contenido vacío. En elementos es:

```
<elemento></elemento>
```

En atributos, es:

```
<elemento atributo="">
```

Opciones de correlación de registros en XML

La pestaña Registros permite especificar la parte de la estructura XML que se utilizará para iniciar cada nuevo registro. Para correlacionar correctamente a un esquema, necesita especificar el delimitador del registro.

Estructura XML. Un árbol jerárquico que muestra la estructura del esquema XML especificado en la pantalla anterior.

Registros (expresión XPath). Para definir el delimitador del registro, seleccione un elemento de la estructura XML y pulse en el botón de flecha derecha. Cada vez que este elemento se encuentra en los datos de origen, se crea un nuevo registro en el archivo de resultados.

Note: Si selecciona el elemento de raíz en la estructura XML, solamente se puede escribir un registro y el resto se omiten.

Opciones de correlación de campos a XML

La pestaña Campos se utiliza para correlacionar campos en el conjunto de datos con elementos o atributos en la estructura XML cuando se utiliza un archivo de esquema.

Los nombres de campos que coinciden con un elemento o atributo se correlacionan automáticamente siempre y cuando el nombre del elemento o atributo sea exclusivo. Además, si hay un elemento y un atributo con el nombre field1, no se produce la correlación automática. Si sólo hay un elemento en la estructura denominado field1, se correlaciona automáticamente en la ruta un campo con ese nombre.

Campos. La lista de campos del modelo. Seleccione uno o más campos como la parte de origen de la correlación. Puede utilizar estos botones en la parte inferior de la lista para seleccionar todos los campos o todos los campos con un nivel de medición concreto.

Estructura XML. Seleccione un elemento en la estructura XML como destino de la correlación. Para crear la correlación, pulse Correlacionar. Se muestra la correlación. El número de campos que se han correlacionado de esta forma se muestran bajo la lista.

Para eliminar una correlación, seleccione el elemento en la lista de estructuras XML y pulse en **Cancelar correlación**.

Mostrar atributos. Muestra u oculta los atributos, si los tiene, de los elementos en la estructura.

Presentación preliminar de correlación a XML

En la pestaña Presentación preliminar, pulse en **Actualizar** para ver una vista previa del XML que se escribirá.

Si la correlación es incorrecta, vuelva a la pestaña Registros o Campos para corregir los errores y vuelva a pulsar en **Actualizar** para ver el resultado.

Capítulo 8. Nodos de IBM SPSS Statistics

Conceptos básicos de nodos de IBM SPSS Statistics

Para complementar IBM SPSS Modeler y su capacidad de minería de datos, IBM SPSS Statistics le ofrece la capacidad de realizar análisis estadístico y gestión de datos más profundos.

Si tiene una copia compatible con licencia de IBM SPSS Statistics instalada, puede conectarla desde IBM SPSS Modeler y realizar una manipulación y análisis complejos y en pasos múltiples de datos que de otro modo no sería posible con IBM SPSS Modeler. Para usuarios avanzados también existe la opción de modificar más los análisis utilizando sintaxis de comandos. Consulte las notas de publicación para obtener información sobre compatibilidad de la versión.

Si están disponibles, los nodos de IBM SPSS Statistics se muestran en una parte exclusiva de la paleta de nodos.

Note: Recomendamos que instancie sus datos en un nodo Tipo antes de utilizar los nodos Transformar, Modelo o Resultado de IBM SPSS Statistics. Esto también es un requisito cuando se utiliza el comando de sintaxis AUTORECODE.

La paleta de IBM SPSS Statistics contiene los siguientes nodos:



El nodo Archivo de estadísticas lee los datos desde un formato de archivo *.sav* o *.zsav* que utiliza IBM SPSS Statistics y archivos caché guardados en IBM SPSS Modeler, que también puede utilizar el mismo formato.



El nodo Transformación Statistics ejecuta una selección de comandos de sintaxis de IBM SPSS Statistics en los orígenes de datos de IBM SPSS Modeler. Este nodo requiere una copia de IBM SPSS Statistics con licencia.



El nodo Modelo Statistics permite analizar y trabajar con sus datos ejecutando los procedimientos de IBM SPSS Statistics que producen PMML. Este nodo requiere una copia de IBM SPSS Statistics con licencia.



El nodo Resultados de Statistics le permite llamar a un procedimiento de IBM SPSS Statistics para analizar los datos de IBM SPSS Modeler. Se puede acceder a una gran variedad de procedimientos analíticos de IBM SPSS Statistics. Este nodo requiere una copia de IBM SPSS Statistics con licencia.



El nodo Exportar estadísticas genera datos de salida en formato *.sav* o *.zsav* de IBM SPSS Statistics. Los archivos *.sav* o *.zsav* se pueden leer con IBM SPSS Statistics Base y otros productos. Este es también el formato utilizado para los archivos caché de IBM SPSS Modeler.

Note: Si su copia de SPSS Statistics solamente tiene licencia para un único usuario y ejecuta una ruta con dos o más ramas, las cuales contienen un nodo de SPSS Statistics cada una, puede que reciba un error de

licencia. Esto se produce cuando la sesión de SPSS Statistics para una rama no se ha terminado antes de que la sesión de otra rama intenta comenzar. Si es posible, rediseñe la ruta para que varias ramas con nodos de SPSS Statistics no se ejecuten en paralelo.

Nodo de archivos estadísticos

Puede utilizar el nodo Archivo de estadísticas para leer los datos directamente desde un archivo de IBM SPSS Statistics guardado (.sav o .zsav). Este formato se utiliza ahora para reemplazar el archivo de caché de las versiones anteriores de IBM SPSS Modeler. Si desea importar un archivo de caché guardado, debe utilizar el nodo Archivo de IBM SPSS Statistics.

Importar archivo. Especifique el nombre del archivo. Puede especificar un nombre de archivo o pulsar el botón de puntos suspensivos (...) para seleccionar un archivo. La ruta de archivo se mostrará una vez haya seleccionado un archivo.

Archivo cifrado mediante contraseña. Seleccione esta casilla si sabe que el archivo está protegido mediante contraseña; se le solicitará que especifique la **Contraseña**. Si el archivo está protegido mediante contraseña y no especifica dicha contraseña, se mostrará un mensaje de aviso si se intenta cambiar a otra pestaña, renovar los datos, previsualizar el contenido del nodo o ejecutar una ruta que contenga el nodo.

Nota: Los archivos protegidos por contraseña solo pueden abrirse con IBM SPSS Modeler versión 16 o superior.

Nombres de variable. Seleccione un método de gestión de etiquetas y nombres de variables tras la importación desde un archivo .sav o .zsav de IBM SPSS Statistics. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en IBM SPSS Modeler y se pueden volver a exportar para su uso en IBM SPSS Statistics.

- **Leer nombres y etiquetas.** Seleccione para leer las etiquetas y nombres de variables en IBM SPSS Modeler. De forma predeterminada, esta opción está seleccionada y los nombres de variables se muestran en el nodo Tipo. Las etiquetas se pueden mostrar en gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta. De forma predeterminada, la representación de etiquetas en los resultados está desactivada.
- **Leer etiquetas como nombres.** Seleccione para leer las etiquetas descriptivas de variables desde el archivo .sav o .zsav de IBM SPSS Statistics en lugar de los nombres abreviados de campos y utilice estas etiquetas como nombres de variables en IBM SPSS Modeler.

Valores. Seleccione un método de gestión de valores y etiquetas tras la importación desde un archivo .sav o .zsav de IBM SPSS Statistics. Los metadatos seleccionados para incluirlos aquí permanecen durante el trabajo en IBM SPSS Modeler y se pueden volver a exportar para su uso en IBM SPSS Statistics.

- **Leer datos y etiquetas.** Seleccione para leer valores reales y etiquetas de valor en IBM SPSS Modeler. De forma predeterminada, esta opción está seleccionada y los propios valores se muestran en el nodo Tipo. Las etiquetas de valor se pueden mostrar en el generador de expresiones, gráficos, exploradores de modelos y otros tipos de resultados, dependiendo de las opciones especificadas en el cuadro de diálogo de propiedades de la ruta.
- **Leer etiquetas como datos.** Indique si desea utilizar las etiquetas de valor del archivo .sav o .zsav en lugar de los códigos numéricos o simbólicos utilizados para representar los valores. Por ejemplo, si selecciona esta opción para datos con un campo de sexo cuyos valores 1 y 2 representan realmente *masculino* y *femenino* respectivamente, convertirá el campo en una cadena e importará *masculino* y *femenino* como los valores reales.

Es importante tener en cuenta los valores perdidos en los datos de IBM SPSS Statistics antes de seleccionar esta opción. Por ejemplo, si un campo numérico utiliza etiquetas sólo para valores perdidos (0 = *Sin respuesta*, -99 = *Desconocido*), al seleccionar la opción anterior se importarán sólo las etiquetas de valor *Sin respuesta* y *Desconocido* y el campo se convertirá en una cadena. En estos casos, debe importar los propios valores y establecer los valores perdidos en un nodo Tipo.

Utilice la información de formato de campo para determinar el almacenamiento. Si esta casilla está deseleccionada, los valores de campo que tienen el formato en el archivo *.sav* como enteros (por ej. los campos especificados como *Fn.0* en la Vista de variables en IBM SPSS Statistics) se importan utilizando el almacenamiento de enteros. El resto de los valores de campos excepto las cadenas se importan como números reales.

Si esta casilla está seleccionada (valor predeterminado), todos los valores de campos excepto las cadenas se importan como números reales, ya tengan formato en el archivo *.sav* como enteros o no.

Conjuntos de respuestas múltiples. Cualquier conjunto de respuestas múltiples definido en el archivo IBM SPSS Statistics se conservará cuando se importe el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.

Nodo Transformación de Statistics

El nodo Transformación de Statistics permite llevar a cabo transformaciones de datos utilizando la sintaxis de comandos de IBM SPSS Statistics, lo que permite llevar a cabo las transformaciones que no admite IBM SPSS Modeler y automatizar transformaciones complejas que requieren varios pasos, incluida la creación de varios campos desde un único nodo. Es similar al nodo Resultados de Statistics, la diferencia es que los datos se devuelven a IBM SPSS Modeler para su posterior análisis, mientras que en el nodo Resultados los datos se devuelven como objetos de resultados solicitados (por ejemplo, gráficos o tablas).

Debe tener una versión compatible de IBM SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Consulte el tema “Aplicaciones de ayuda de IBM SPSS Statistics” en la página 336 para obtener más información. Consulte las Notas del release para obtener información de compatibilidad.

Si es necesario, puede utilizar la pestaña Filtro para filtrar los campos o cambiar su nombre para que cumplan los estándares de nomenclatura de IBM SPSS Statistics. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.

Referencia de sintaxis. Para obtener detalles sobre procedimientos específicos de IBM SPSS Statistics, consulte la *Referencia de sintaxis de comandos de IBM SPSS Statistics*, que se incluye con su copia del software IBM SPSS Statistics. Para ver la guía desde la pestaña Sintaxis, seleccione la opción **Editor de sintaxis** y pulse el botón Iniciar ayuda de sintaxis de IBM SPSS Statistics.

Note: este nodo no admite toda la sintaxis de IBM SPSS Statistics. Consulte el tema “Sintaxis permitida” en la página 368 para obtener más información.

Pestaña Sintaxis del nodo Transformación de Statistics

Opción de cuadro de diálogo de IBM SPSS Statistics

Si no está familiarizado con la sintaxis de IBM SPSS Statistics para un procedimiento, el modo más sencillo de crear sintaxis en IBM SPSS Modeler es seleccionar la opción **Cuadro de diálogo IBM SPSS Statistics**, seleccionar el cuadro de diálogo del procedimiento, cumplimentar el cuadro de diálogo y pulsar en Aceptar. Al realizar estas acciones, colocará la sintaxis en la pestaña Sintaxis del nodo de IBM SPSS Statistics que está utilizando en IBM SPSS Modeler. A continuación, podrá ejecutar la ruta para obtener los resultados del procedimiento.

Opción del editor Sintaxis de IBM SPSS Statistics

Comprobar. Tras introducir los comandos de sintaxis en la parte superior del cuadro de diálogo, utilice este botón para validar las entradas. Todos los errores de sintaxis aparecen identificados en la parte inferior del cuadro de diálogo.

Para garantizar que el proceso de comprobación no tarda demasiado, al validar la sintaxis, se comprueba una muestra representativa de los datos para comprobar que las entradas son válidas, en lugar de comprobar el conjunto de datos completo.

Sintaxis permitida

Si tiene una gran cantidad de sintaxis heredada de IBM SPSS Statistics o está familiarizado con las características de preparación de datos de IBM SPSS Statistics, puede utilizar el nodo Transformación de Statistics para ejecutar muchas de las transformaciones existentes. A modo de orientación, el nodo permite transformar los datos de maneras predecibles (por ejemplo, ejecutando bucles de comandos o cambiando, añadiendo, ordenando, filtrando o seleccionando datos).

Estos son algunos ejemplos de comandos que se pueden llevar a cabo:

- Calcular números aleatorios según una distribución binomial:
`COMPUTE nuevavar = RV.BINOM(10000,0.1)`
- Recodificar una variable en una nueva variable:
`RECODE Edad (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO EdadRecodificada`
- Sustituir valores perdidos:
`RMV Edad_1=SMEAN(Edad)`

A continuación se indica la sintaxis de IBM SPSS Statistics admitida por el nodo Transformación de Statistics:

Nombre de comando

ADD VALUE LABELS
APPLY DICTIONARY
AUTORECODE
BREAK
CD
CLEAR MODEL PROGRAMS
CLEAR TIME PROGRAM
CLEAR TRANSFORMATIONS
COMPUTE
COUNT
CREATE
DATE
DEFINE-!ENDDDEFINE
DELETE VARIABLES
DO IF
DO REPEAT
ELSE
ELSE IF
END CASE
END FILE

Nombre de comando

END IF
END INPUT PROGRAM
END LOOP
END REPEAT
EXECUTE
FILE HANDLE
FILE LABEL
FILE TYPE-END FILE TYPE
FILTER
FORMATS
IF
INCLUDE
INPUT PROGRAM-END INPUT PROGRAM
INSERT
LEAVE
LOOP-END LOOP
MATRIX-END MATRIX
MISSING VALUES
N OF CASES
NUMERIC
PERMISSIONS
PRESERVE
RANK
RECODE
RENAME VARIABLES
RESTORE
RMV
SAMPLE
SELECT IF
SET
SORT CASES
STRING
SUBTITLE
TEMPORARY
TITLE
UPDATE
V2C
VALIDATEDATA
VALUE LABELS
VARIABLE ATTRIBUTE
VARSTOCASES

Nombre de comando

VECTOR

Nodo Modelo de Statistics

El nodo Modelo Statistics permite analizar y trabajar con sus datos ejecutando los procedimientos de IBM SPSS Statistics que producen PMML. Los nuggets de modelo que cree pueden utilizarse luego de la manera habitual en las rutas de IBM SPSS Modeler para su puntuación y otras operaciones.

Debe tener una versión compatible de IBM SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Consulte el tema “Aplicaciones de ayuda de IBM SPSS Statistics” en la página 336 para obtener más información. Consulte las Notas del release para obtener información de compatibilidad.

Los procedimientos analíticos de IBM SPSS Statistics disponibles dependen del tipo de licencia que disponga.

Pestaña Modelo del nodo Modelos de Statistics

Nombre del modelo. Puede generar el nombre del modelo de forma automática basándose en el campo objetivo o de ID (o en el nombre del tipo de modelo si se especifica ningún campo objetivo), o bien especificar un nombre personalizado.

Seleccione un cuadro de diálogo. Pulse para mostrar una lista de procedimientos de IBM SPSS Statistics disponibles que puede seleccionar y ejecutar. La lista sólo muestra aquellos procedimientos que producen PMML y para los que tiene licencia, y no incluye los procedimientos escritos por el usuario.

1. Pulse en el procedimiento necesario; aparecerá el cuadro de diálogo de IBM SPSS Statistics correspondiente.
2. En el cuadro de diálogo de IBM SPSS Statistics, introduzca los detalles para el procedimiento.
3. Pulse en **Aceptar** para volver al nodo Modelos de Statistics; la sintaxis de IBM SPSS Statistics se muestra en la pestaña Modelo.
4. Para volver al cuadro de diálogo de IBM SPSS Statistics en cualquier momento, como por ejemplo para modificar su consulta, pulse en el cuadro de diálogo de IBM SPSS Statistics a la derecha del botón de selección de procedimiento.

Nodo Modelo de Statistics - Resumen del nugget de modelo

Cuando ejecuta el nodo Modelo de Statistics, se ejecuta el procedimiento asociado de IBM SPSS Statistics y se crea un nugget de modelo que puede utilizar en las rutas de IBM SPSS Modeler para su puntuación.

La pestaña Resumen de un nugget de modelo muestra información sobre los campos, la configuración de creación y el proceso de estimación del modelo. Los resultados se muestran en una vista de árbol que se puede expandir o contraer pulsando los elementos específicos.

El botón **Ver modelo** muestra los resultados en un formato modificado del Visor de resultados de IBM SPSS Statistics. Para obtener más información sobre este visor, consulte la documentación de IBM SPSS Statistics.

Las opciones normales de exportación e impresión están disponibles en el menú Archivo. Consulte el tema “Visualización de resultados” en la página 296 para obtener más información.

Nodo Resultados de Statistics

El nodo Resultados de Statistics le permite llamar a un procedimiento de IBM SPSS Statistics para analizar los datos de IBM SPSS Modeler. Puede ver los resultados en una ventana del explorador o guardarlos en el formato de archivo de resultados de IBM SPSS Statistics. Se puede acceder a una gran variedad de procedimientos analíticos de IBM SPSS Statistics desde IBM SPSS Modeler.

Debe tener una versión compatible de IBM SPSS Statistics instalada en su equipo y con licencia para poder usar este nodo. Consulte el tema “Aplicaciones de ayuda de IBM SPSS Statistics” en la página 336 para obtener más información. Consulte las Notas del release para obtener información de compatibilidad.

Si es necesario, puede utilizar la pestaña Filtro para filtrar los campos o cambiar su nombre para que cumplan los estándares de nomenclatura de IBM SPSS Statistics. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.

Referencia de sintaxis. Para obtener detalles sobre procedimientos específicos de IBM SPSS Statistics, consulte la *Referencia de sintaxis de comandos de IBM SPSS Statistics*, que se incluye con su copia del software IBM SPSS Statistics. Para ver la guía desde la pestaña Sintaxis, seleccione la opción **Editor de sintaxis** y pulse el botón Iniciar ayuda de sintaxis de IBM SPSS Statistics.

Pestaña Sintaxis del nodo Resultados de Statistics

Use esta pestaña para crear sintaxis para el procedimiento de IBM SPSS Statistics que desea utilizar para analizar los datos. La sintaxis se compone de dos partes: una **instrucción** y las **opciones** asociadas. La instrucción especifica el análisis o la operación que se va a realizar y los campos que se van a utilizar. Las opciones especifican todo lo demás, incluyendo los estadísticos que se van a mostrar, los campos derivados que se van a guardar, etc.

Opción de cuadro de diálogo de IBM SPSS Statistics

Si no está familiarizado con la sintaxis de IBM SPSS Statistics para un procedimiento, el modo más sencillo de crear sintaxis en IBM SPSS Modeler es seleccionar la opción **Cuadro de diálogo IBM SPSS Statistics**, seleccionar el cuadro de diálogo del procedimiento, cumplimentar el cuadro de diálogo y pulsar en Aceptar. Al realizar estas acciones, colocará la sintaxis en la pestaña Sintaxis del nodo de IBM SPSS Statistics que está utilizando en IBM SPSS Modeler. A continuación, podrá ejecutar la ruta para obtener los resultados del procedimiento.

También puede generar un nodo de origen Archivo de Statistics para importar los datos resultantes. Esto es de utilidad, por ejemplo, si un procedimiento escribe campos como puntuaciones para el conjunto de datos activo además de mostrar los resultados.

Para crear la sintaxis:

1. Pulse en el botón **Seleccionar un cuadro de diálogo...**
2. Seleccione una de las opciones:
 - **Analizar.** Enumera el contenido del menú Analizar de IBM SPSS Statistics; seleccione el procedimiento que desee utilizar.
 - **Otro.** Si aparece, enumera los cuadros de diálogo creados por el generador de cuadros de diálogo personalizados en IBM SPSS Statistics, así como otros cuadros de diálogo de IBM SPSS Statistics que no aparecen en el menú Analizar y para los que tiene licencia. Si no se especifica ningún cuadro de diálogo aplicable, esta opción no aparece.

Note: Los cuadros de diálogo Preparación de datos automática no se muestran.

Si tiene un cuadro de diálogo personalizado de IBM SPSS Statistics que crea nuevos campos, estos campos no se podrán utilizar en IBM SPSS Modeler porque el nodo Resultados de Statistics es un nodo terminal.

También puede marcar la casilla **Generar un nodo de importación para los datos resultantes** para crear un nodo de origen Archivo de Statistics que puede utilizarse para importar los datos resultantes en otra ruta. El nodo se coloca en el lienzo de pantalla, con los datos incluidos en el archivo *.sav* especificados en el campo **Archivo** (la ubicación predeterminada es el directorio de instalación de IBM SPSS Modeler).

Opción del editor Sintaxis

Para guardar la sintaxis que se ha creado para un procedimiento de uso frecuente:

1. Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
2. En el menú seleccione **Guardar** o **Guardar como**.
3. Guarde el archivo como un archivo *sps*.

Para usar archivos de sintaxis creados previamente, sustituya el contenido actual, si los hay, del editor Sintaxis:

1. Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
2. Seleccione **Abrir** en el menú.
3. Al seleccionar un archivo *.sps* se pegará el contenido en la pestaña Sintaxis del nodo Resultados.

Para insertar una sintaxis guardada previamente sin sustituir el contenido actual:

1. Pulse en el botón Opciones de archivo (el primero de la barra de herramientas).
2. Seleccione **Insertar** en el menú.
3. Al seleccionar un archivo *.sps* se pegará el contenido en el nodo Resultados en el punto especificado por el cursor.

También puede marcar la casilla **Generar un nodo de importación para los datos resultantes** para crear un nodo de origen Archivo de Statistics que puede utilizarse para importar los datos resultantes en otra ruta. El nodo se coloca en el lienzo de pantalla, con los datos incluidos en el archivo *.sav* especificados en el campo **Archivo** (la ubicación predeterminada es el directorio de instalación de IBM SPSS Modeler).

Al pulsar en **Ejecutar**, se muestran los resultados en el Visor de resultados de IBM SPSS Statistics. Para obtener más información sobre el visor, consulte la documentación de IBM SPSS Statistics.

Pestaña Resultados del nodo Resultados de Statistics

La pestaña Resultado permite especificar el formato y la ubicación del resultado. Puede elegir entre mostrar los resultados en pantalla o enviarlos a uno de los tipos de archivo disponibles.

Nombre del resultado. Especifica el nombre del resultado generado cuando se ejecuta el nodo. **Automático** selecciona un nombre según el nodo que genera el resultado. De forma opcional, puede seleccionar **Personalizado** para especificar un nombre diferente.

Salida a pantalla (valor predeterminado). Crea un objeto de resultados para ver en línea. El objeto de resultados aparecerá en la pestaña Resultados de la ventana del gestor al ejecutar el nodo de resultados.

Salida a archivo. Almacena el resultado en un archivo cuando se ejecuta el nodo. Si selecciona esta opción, introduzca un nombre de archivo en el campo **Nombre de archivo** (o navegue a un directorio y especifique un nombre de archivo mediante el botón Buscar el archivo) y seleccione un tipo de archivo.

Tipo de archivo. Seleccione el tipo de archivo al que desee enviar los resultados.

- **Documento HTML (*.html)** Escribe los resultados en un formato HTML.

- **IBM SPSS Statistics Archivo de visor (*.spv).** Escribe los resultados en un formato que puede leer el Visor de resultados de IBM SPSS Statistics.
- **IBM SPSS Statistics Archivo de informes web (*.spw).** Escribe los resultados en formato de informes web de IBM SPSS Statistics, que pueden publicarse en un repositorio de un IBM SPSS Collaboration and Deployment Services y verse posteriormente en un navegador web. Consulte el tema “Publicar en Web” en la página 297 para obtener más información.

Nota: si selecciona **Salida a la pantalla**, la directiva VIEWER=NO de IBM SPSS Statistics OMS no tiene ningún efecto; además, el las API de script (módulo *Básico* y *Python SpssClient*) no están disponibles en IBM SPSS Modeler.

Nodo Exportación de Statistics

El nodo Exportar de Statistics permite exportar datos en formato *.sav* de IBM SPSS Statistics. Los archivos IBM SPSS Statistics *.sav* se pueden leer con IBM SPSS Statistics Base y otros módulos. Este es también el formato utilizado para los archivos caché de IBM SPSS Modeler.

El hecho de correlacionar nombres de campos de IBM SPSS Modeler con nombres de variables de IBM SPSS Statistics puede provocar errores, ya que los nombres de variables de IBM SPSS Statistics están limitados a 64 caracteres y no pueden incluir determinados caracteres, como espacios, el signo del dólar (\$) y guiones (–). Existen dos modos de ajustar estas restricciones:

- Puede cambiar el nombre de los campos ajustándose a los requisitos del nombre de las variables de IBM SPSS Statistics pulsando en la pestaña Filtro. Consulte el tema “Cambio de nombre o filtrado de campos para IBM SPSS Statistics” en la página 374 para obtener más información.
- Elija exportar las etiquetas y los nombres de campos desde IBM SPSS Modeler.

Note: IBM SPSS Modeler escribe archivos *.sav* en formato Unicode UTF-8. IBM SPSS Statistics sólo admite archivos en formato Unicode UTF-8 de la versión 16.0 en adelante. Para evitar la posibilidad de corrupción de datos de archivos *.sav* guardados con codificación Unicode, no se deberían utilizar en versiones de IBM SPSS Statistics previas a 16.0. Si desea más información, consulte la ayuda de IBM SPSS Statistics.

Conjuntos de respuestas múltiples. Cualquier conjunto de respuestas múltiples definido en la ruta se conservará automáticamente cuando se exporte el archivo. Puede definir o editar conjuntos de respuestas múltiples en un nodo con una pestaña Filtro. Consulte el tema “Edición de conjuntos de respuestas múltiples” en la página 141 para obtener más información.

Pestaña Exportar del nodo Exportar Statistics

Exportar archivo. Especifica el nombre del archivo. Escriba el nombre del archivo o pulse en el botón de selección de archivos para acceder a la ubicación del archivo.

Tipo de archivo. Selecciona si el archivo debe guardarse en formato normal *.sav* o en formato comprimido *.zsav*.

Cifrar archivo con contraseña. Para proteger el archivo con una contraseña, seleccione esta casilla; se le solicitará que especifique y confirme la **Contraseña** en un cuadro de diálogo aparte.

Nota: Los archivos protegidos por contraseña solo pueden abrirse con IBM SPSS Modeler versión 16 o superior, o con IBM SPSS Statistics versión 21 o superior.

Exportar nombres de campo como variables. Especifica un método de gestión de etiquetas y nombres de variables tras la exportación desde IBM SPSS Modeler a un archivo *.sav* o *.zsav* de IBM SPSS Statistics.

- **nombres.** Para exportar los nombres y las etiquetas de campo de IBM SPSS Modeler. Los nombres se exportan como nombres de variables de IBM SPSS Statistics, mientras que las etiquetas se exportan como etiquetas de variables de IBM SPSS Statistics.
- **etiquetas.** Seleccione esta opción para utilizar los nombres de campos de IBM SPSS Modeler como etiquetas de variables en IBM SPSS Statistics. IBM SPSS Modeler permite caracteres en nombres de campos que no son válidos en nombres de variables de IBM SPSS Statistics. Para evitar crear posibles nombres de IBM SPSS Statistics no válidos, seleccione nombres en **Exportar nombres de campo como variables** en su lugar o utilice la pestaña Filtro para ajustar los nombres de campos.

Abrir aplicación. Si IBM SPSS Statistics está instalado en su ordenador, puede seleccionar esta opción para activarlos directamente en el archivo de datos almacenado. Las opciones de ejecución de cada una de las aplicaciones se deben especificar en el cuadro de diálogo Aplicaciones de ayuda. Consulte el tema “Aplicaciones de ayuda de IBM SPSS Statistics” en la página 336 para obtener más información. Si solo desea crear un archivo *.sav* o *.zsav* de IBM SPSS Statistics sin abrir un programa externo, deselectione esta opción.

Generar un nodo de importación para este dato. Seleccione esta opción para generar automáticamente un nodo de origen Archivo de Statistics que leerá el archivo de datos exportado. Consulte el tema “Nodo de archivos estadísticos” en la página 366 para obtener más información.

Cambio de nombre o filtrado de campos para IBM SPSS Statistics

Antes de exportar o desplegar los datos de IBM SPSS Modeler a aplicaciones externas como IBM SPSS Statistics, puede que sea necesario cambiar o ajustar los nombres de los campos. Los cuadros de diálogo Transformación de Statistics, Resultados de Statistics y Exportar Statistics contienen una pestaña Filtro para facilitar este proceso.

En otras secciones se ofrece una descripción básica de la funcionalidad de la pestaña Filtro. Consulte el tema “Opciones de configuración de filtrado” en la página 140 para obtener más información. Este tema ofrece sugerencias para leer datos en IBM SPSS Statistics.

Para ajustar los nombres de campo para conformar a las convenciones de nomenclatura de IBM SPSS Statistics:

1. En la pestaña Filtro, pulse en el botón de la barra de herramientas del menú Opciones de filtro (el primero de la barra de herramientas).
2. Seleccione Cambiar nombre para IBM SPSS Statistics.
3. En el cuadro de diálogo Cambiar nombre para IBM SPSS Statistics puede seleccionar sustituir los caracteres no válidos de los nombres de archivo con un carácter de **almohadilla (#)** o un **guión bajo** (_).

Cambiar nombre de conjuntos de respuestas múltiples. Seleccione esta opción si desea ajustar los nombres de conjuntos de respuestas múltiples, que pueden importarse a IBM SPSS Modeler mediante un nodo de origen Archivo de Statistics. Se usan para registrar datos que pueden tener más de un valor para cada caso, como en respuestas de encuestas.

Capítulo 9. Supernodos

Conceptos básicos de supernodos

El que cada nodo tenga una función claramente definida es uno de los motivos por los que resulta tan fácil aprender la interfaz de programación visual de IBM SPSS Modeler. Sin embargo, para procesamientos complejos puede ser necesaria una larga secuencia. Con el tiempo, esto puede recargar el lienzo de rutas y puede resultar difícil seguir los diagramas de rutas. Hay dos formas de evitar que una ruta larga y compleja se recargue:

- Puede dividir una secuencia de procesamiento en varias rutas introduciéndose una en la otra. La primera ruta, por ejemplo, crea un archivo de datos que la segunda utiliza como entrada. La segunda crea un archivo que la tercera utiliza como entrada y así sucesivamente. Puede administrar varias rutas guardándolas en un **proyecto**. Un proyecto proporciona organización para varias rutas y sus salidas. Sin embargo, un archivo de proyecto sólo contiene una referencia a los objetos que incluye y aún tendrá que administrar varios archivos de rutas.
- Puede crear un **Supernodo** como alternativa más simplificada al trabajar con procesos de ruta complejos.

Los supernodos agrupan varios nodos en uno solo encapsulando las secciones de una ruta de datos. Esto proporciona numerosas ventajas al minero de datos:

- Las rutas estarán más ordenadas y serán más manejables.
- Los nodos se podrán unir en un supernodo específico de negocio.
- Los supernodos se podrán exportar a bibliotecas para reutilizarlos en varios proyectos de minería de datos.

Tipos de supernodos

Los supernodos se representan en la ruta de datos mediante un icono de estrella. El icono aparece sombreado para representar el tipo de supernodo y la dirección en la que debe fluir la ruta o desde la que debe fluir.

Hay tres tipos de supernodos:

- Supernodos de origen
- Supernodos de proceso
- Supernodos terminales

Supernodos de origen

Los supernodos de origen contienen un origen de datos como un nodo de origen normal y se pueden utilizar en cualquier lugar al igual que un nodo de origen normal. La parte izquierda del supernodo de origen está sombreada para indicar que está "cerrado" en esa parte y que los datos deben fluir hacia abajo *desde* el supernodo.

Los supernodos de origen tienen sólo un punto de conexión a la derecha, mostrando que los datos dejan el supernodo y fluyen hacia la ruta.

Supernodos de proceso

Los supernodos de proceso sólo contienen nodos de proceso y no están sombreados para mostrar que los datos pueden fluir *dentro* y *fuera* de este tipo de supernodo.

Los supernodos de proceso tienen puntos de conexión a la derecha y a la izquierda, mostrando que los datos entran en el supernodo y lo dejan para volver a la ruta. Aunque los supernodos pueden contener fragmentos de rutas adicionales e incluso rutas extra, los dos puntos de conexión deben fluir a través de una sola ruta de acceso conectando los puntos *De ruta* y *A ruta*.

Note: los supernodos de proceso a veces se denominan *Supernodos de manipulación*.

Supernodos terminales

Los supernodos terminales contienen uno o varios nodos terminales (gráfico, tabla, etc.) y se pueden utilizar de la misma forma que un nodo terminal. Un supernodo terminal aparece sombreado en la parte derecha para indicar que está "cerrado" en esa parte y que los datos pueden fluir sólo *en* un supernodo terminal.

Los supernodos terminales sólo tienen un punto de conexión a la derecha, mostrando que los datos entran en el supernodo desde la ruta y finalizan dentro del supernodo.

Los supernodos terminales también pueden contener scripts que se utilizan para especificar el orden de ejecución para todos los nodos terminales dentro del supernodo. Consulte el tema "Supernodos y scripts" en la página 381 para obtener más información.

Creación de supernodos

La creación de un supernodo "reduce" la ruta de datos encapsulando varios nodos en uno. Una vez se ha creado o cargado una ruta en el lienzo, hay varias formas de crear un supernodo.

Selección múltiple

La forma más sencilla de crear un supernodo es seleccionando todos los nodos que desee encapsular:

1. Utilice el ratón para seleccionar varios nodos en el lienzo de rutas. También puede utilizar Mayús-pulsar para seleccionar una ruta o la sección de una ruta. *Note:* los nodos que seleccione deben ser de una ruta continua o bífida. No puede seleccionar nodos que no sean adyacentes o no estén conectados de alguna forma.
2. A continuación, utilizando uno de los tres métodos siguientes, encapsule los nodos seleccionados:
 - Pulse en el icono Supernodo (sombreado con forma de estrella) de la barra de herramientas.
 - Pulse con el botón derecho en el supernodo y en el menú contextual seleccione lo siguiente:
Crear supernodo > Desde selección
 - En el menú de Supernodo, seleccione:
Crear supernodo > Desde selección

Estas tres opciones encapsulan los nodos en un supernodo sombreado para reflejar el tipo (de origen, de proceso o terminal) según el contenido.

Selección única

También puede crear un supernodo seleccionando un solo nodo y utilizando las opciones de menú para determinar el inicio y finalización del supernodo o encapsulando todo lo que hay debajo del nodo seleccionado.

1. Pulse en el nodo que determine el inicio de la encapsulación.
2. En el menú de Supernodo, seleccione:
Crear Supernodo > Desde aquí

Los supernodos también se pueden crear de forma más interactiva seleccionando el inicio y el final de la sección de la ruta para encapsular los nodos:

1. Pulse en el primer o último nodo que desee incluir en el supernodo.
2. En el menú de Supernodo, seleccione:
Crear supernodo > Seleccionar...
3. Si lo prefiere, puede utilizar las opciones del menú contextual pulsando con el botón derecho en el nodo deseado.
4. El cursor se convierte en un icono de Supernodo, que indica que debe seleccionar otro punto de la ruta. Desplácese hacia arriba o hacia abajo al "otro extremo" del fragmento del supernodo y pulse en un nodo. Esta acción reemplazará todos los nodos entre ellos con el icono de estrella del supernodo.

Note: los nodos que seleccione deben ser de una ruta continua o bífida. No puede seleccionar nodos que no sean adyacentes o no estén conectados de alguna forma.

Anidación de supernodos

Los supernodos se pueden anidar en otros supernodos. Las mismas reglas para cada tipo de supernodo (de origen, de proceso y terminal) se aplican a los supernodos anidados. Por ejemplo, un supernodo de proceso con anidación debe tener un flujo continuo de datos a través de todos los supernodos anidados para que siga siendo un supernodo de proceso. Si uno de los supernodos anidados es terminal, los datos no seguirán fluyendo a través de la jerarquía.

Los supernodos terminales y de origen pueden contener otros tipos de supernodos anidados pero se aplican las mismas reglas básicas para la creación de supernodos.

Bloqueo de Supernodos

Una vez haya creado un supernodo, puede bloquearlo con una contraseña para evitar que se modifique. Por ejemplo, puede hacer esto si crea rutas o partes de rutas, como plantillas de valor fijo para el uso de otros usuarios en su organización que tengan menos experiencia con consultas de configuración de IBM SPSS Modeler.

Cuando se bloquea un supernodo, los usuarios podrán introducir valores en la pestaña Parámetros que se hayan definido y se puede ejecutar un supernodo sin introducir la contraseña.

Note: Las operaciones de bloqueo y desbloqueo no se pueden ejecutar utilizando scripts.

Bloqueo y desbloqueo de un supernodo

Aviso: las contraseñas perdidas no pueden recuperarse.

Puede bloquear o desbloquear un supernodo desde cualquiera de las tres pestañas.

1. Pulse en **Bloquear nodo**.
2. Introduzca y confirme la contraseña.
3. Pulse en **Aceptar**.

Un supernodo protegido por contraseña se identifica en el lienzo de rutas por un pequeño icono de candado en la parte superior izquierda del icono del supernodo.

Desbloquear un supernodo

1. Para eliminar permanentemente la protección por contraseña, pulse en **Desbloquear nodo**. Se le solicitará la contraseña.
2. Introduzca la contraseña y pulse en **Aceptar**. El supernodo ya no estará protegido por contraseña y el icono de candado ya no aparecerá junto al icono de la ruta.

Edición de un supernodo bloqueado

Si intenta definir los parámetros o ampliar el zoom para ver un supernodo bloqueado, se le solicitará la contraseña.

Introduzca la contraseña y pulse en **Aceptar**.

Ahora podrá editar las definiciones de los parámetros y ampliar y reducir el zoom cuando lo necesite hasta que cierre la ruta en la que se encuentra el supernodo.

Tenga en cuenta que no se elimina la protección por contraseña, sólo permite acceder a trabajar con el supernodo. Consulte el tema “Bloqueo y desbloqueo de un supernodo” en la página 377 para obtener más información.

Edición de Supernodos

Una vez haya creado un supernodo, puede examinarlo más detenidamente ampliando el zoom; si el supernodo está bloqueado, se le pedirá que introduzca la contraseña. Consulte el tema “Edición de un supernodo bloqueado” para obtener más información.

Para ver el contenido de un supernodo, puede utilizar el icono acercar de la barra de herramientas de IBM SPSS Modeler o el método siguiente:

1. Pulse con el botón derecho en un supernodo.
2. En el menú contextual, seleccione **Acercar**.

El contenido del supernodo seleccionado se mostrará en un entorno de IBM SPSS Modeler algo diferente, con conectores que muestran el flujo de datos a través de la ruta o del fragmento de ruta. En este nivel del lienzo de rutas, hay varias tareas que puede realizar:

- Modificar el tipo de supernodo: origen, proceso o terminal.
- Crear parámetros o editar los valores de un parámetro. Los parámetros se utilizan en scripts y expresiones CLEM.
- Especificar las opciones de almacenamiento en caché para el supernodo y los subnodos.
- Crear o modificar un script de supernodo (sólo para supernodos terminales).

Modificación de los tipos de Supernodos

En algunas circunstancias, conviene modificar el tipo de un supernodo. Esta opción está disponible sólo cuando se acerque a un supernodo y sólo se aplica al supernodo en este nivel. Los tres tipos de supernodos se describen en la siguiente tabla.

Tabla 63. Tipos de supernodo.

Tipo de supernodo	Descripción
Supernodo de origen	Una conexión saliente
Supernodo de proceso	Dos conexiones: una entrante y otra saliente
Supernodo terminal	Una conexión entrante

Para cambiar el tipo de un supernodo

1. Asegúrese de que se ha acercado al supernodo.
2. En el menú Supernodo seleccione **Tipo de Supernodo** y elija el tipo.

Anotación y cambio de nombre de Supernodos

Puede cambiar el nombre de un supernodo como aparece en la ruta y realizar anotaciones utilizadas en un proyecto o informe. Para acceder a estas propiedades:

- Pulse con el botón derecho en un supernodo (alejamiento) y seleccione **Cambiar nombre y anotar**.
- Si lo prefiere, en el menú Supernodo seleccione **Cambiar nombre y anotar**. Esta opción está disponible tanto en el modo de acercamiento como en el de alejamiento.

En ambos casos, se abrirá un cuadro de diálogo con la pestaña Anotaciones seleccionada. Utilice estas opciones para personalizar el nombre mostrado en el lienzo de rutas y proporcionar documentación relacionada con las operaciones del supernodo.

Uso de comentarios con supernodos

Si crea un supernodo de un nodo o nugget de comentario, debe incluir el comentario en la selección para crear el nodo si desea que aparezca en el supernodo. Si omite el comentario de la selección, el comentario permanecerá en la ruta cuando cree el supernodo.

Si expande el nodo que incluye los comentarios, los comentarios se quedarán en su lugar antes de que se cree el supernodo.

Si expande un supernodo que incluye objetos comentados, pero los comentarios no se incluyen en el supernodo, los objetos se reinstancian a su ubicación original, sin sus comentarios

Parámetros de Supernodos

En IBM SPSS Modeler, puede establecer variables definidas por el usuario, como *Minvalue*, cuyos valores se pueden especificar cuando se utilicen en scripts o en expresiones CLEM. Estas variables se denominan **parámetros**. Puede establecer parámetros para las rutas, las sesiones y los supernodos. Los parámetros establecidos para un supernodo estarán disponibles al generar expresiones CLEM en dicho supernodo o en los nodos anidados. Los parámetros establecidos para los supernodos anidados no están disponibles en el supernodo padre.

Hay dos pasos para crear y configurar los parámetros para los supernodos:

1. Defina los parámetros para el supernodo.
2. A continuación, especifique el valor para cada parámetro del supernodo.

Estos parámetros se podrán utilizar en expresiones CLEM para los nodos encapsulados.

Definición de los parámetros de Supernodos

Los parámetros de un supernodo se pueden definir tanto en un modo de alejamiento como de acercamiento. Los parámetros definidos se aplican a todos los nodos encapsulados. Para definir los parámetros de un supernodo, primero deberá acceder a la pestaña Parámetros del cuadro de diálogo Supernodo. Utilice uno de los siguientes métodos para abrir el cuadro de diálogo:

- Pulse dos veces en un supernodo de la ruta.
- En el menú Supernodo, seleccione **Definir parámetros**.
- Si lo prefiere, cuando se acerque a un supernodo, seleccione **Definir parámetros** en el menú contextual.

Una vez abierto el cuadro de diálogo, la pestaña Parámetros estará visible con los parámetros definidos previamente.

Para definir un nuevo parámetro

Pulse en el botón **Definir parámetros** para abrir el cuadro de diálogo.

Nombre. Los nombres de los parámetros se enumeran aquí. Se puede crear un parámetro nuevo escribiendo un nombre en este campo. Por ejemplo, si desea crear un parámetro para la temperatura mínima, puede escribir *minvalue*. No incluya el prefijo \$P- que denota un parámetro en las expresiones CLEM. Este nombre se utiliza para su representación en el generador de expresiones de CLEM.

Nombre largo. Indica el nombre descriptivo de cada parámetro creado.

Almacenamiento. Seleccione un tipo de almacenamiento de la lista. Indica cómo se almacenan los valores de datos en el parámetro. Por ejemplo, cuando trabaje con valores con ceros iniciales que desee conservar (como 008), debe seleccionar **Cadena** como el tipo de almacenamiento. De lo contrario, los ceros se eliminarán del valor. Los tipos de almacenamiento disponibles son cadena, entero, real, hora, fecha y marca de tiempo. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar utilizando la notación estándar ISO tal y como se detalla en el siguiente párrafo.

Valor. Indica el valor actual para cada parámetro. Ajuste el parámetro como desee. Tenga en cuenta que para los parámetros de fecha, los valores se deben especificar en la notación estándar ISO (que es AAAA-MM-DD). No se aceptan fechas especificadas en otros formatos.

Tipo (opcional). Si desea desplegar la ruta en una aplicación externa, seleccione un nivel de medición de la lista. En caso contrario, se recomienda dejar la columna *Tipo* tal como está. Si desea especificar restricciones de valores para el parámetro, como límites superiores e inferiores para un rango numérico, seleccione **Especificar** en la lista.

Tenga en cuenta que las opciones de tipo, almacenamiento y nombre largo se pueden establecer para los parámetros sólo a través de la interfaz de usuario. Estas opciones no se pueden establecer utilizando scripts.

Pulse en las flechas de la derecha para mover el parámetro seleccionado hacia arriba o hacia abajo en la lista de parámetros disponibles. Utilice el botón de eliminación (marcado con una X) para eliminar el parámetro seleccionado.

Configuración de valores para los parámetros de Supernodo

Una vez haya definido los parámetros para un supernodo, podrá especificar los valores mediante los parámetros en un script o expresión CLEM.

Para especificar los parámetros de un supernodo

1. Pulse dos veces en el icono Supernodo para abrir el cuadro de diálogo Supernodo.
2. Si lo prefiere, en el menú Supernodo, seleccione **Definir parámetros**.
3. Pulse en la pestaña **Parámetros**. *Note:* los campos de este cuadro de diálogo se definen pulsando en el botón **Definir parámetros** de esta pestaña.
4. Introduzca un valor en el cuadro de texto para cada parámetro creado. Por ejemplo, puede definir el valor *minvalue* para un umbral determinado de interés. Este parámetro se podrá utilizar en numerosas operaciones, como la selección de registros por encima o debajo de este umbral para realizar una exploración más a fondo.

Uso de parámetros de Supernodo para acceder a las propiedades de los nodos

Los parámetros de supernodo también se pueden utilizar para definir las propiedades de los nodos (asimismo denominadas **parámetros de intervalo**) para los nodos encapsulados. Por ejemplo, imagine que desea especificar que un supernodo entrena un nodo Red neuronal encapsulado para un período de tiempo determinado mediante una muestra aleatoria de los datos disponibles. Con los parámetros, podrá especificar los valores para el período de tiempo y la muestra de porcentaje.

Supongamos que el Supernodo de ejemplo contiene un nodo Muestrear denominado *Sample* y un nodo Red neuronal denominado *Train*. Puede utilizar los cuadros de diálogo del nodo para especificar la configuración **Muestrear** del nodo Muestrear como **Aleatorio %** y la configuración **Detener cuando** del nodo Red neuronal como **Tiempo**. Una vez se hayan especificado estas opciones, podrá acceder a las propiedades de los nodos y determinar los valores específicos del supernodo. En el cuadro de diálogo Supernodo, pulse en **Definir parámetros** y cree los parámetros mostrados en la tabla siguiente.

Tabla 64. Parámetros a crear

Parámetro	Valor	Nombre completo
Train.time	5	Tiempo para entrenar (minutos)
Sample.random	10	Muestra aleatoria de porcentaje

Note: los nombres de los parámetros, como *Sample.random*, utilizan una sintaxis correcta para referirse a las propiedades de los nodos, donde *Sample* representa el nombre del nodo y *random* es la propiedad.

Una vez haya definido estos parámetros, podrá modificar fácilmente los valores para las propiedades de los dos nodos Muestrear y Red neuronal sin necesidad de volver a abrir los cuadros de diálogo. En su lugar, sólo tiene que seleccionar **Definir parámetros** en el menú Supernodo para acceder a la pestaña Parámetros del cuadro de diálogo Supernodo, donde puede especificar valores nuevos para **Aleatorio %** y **Tiempo**. Esto resulta muy útil a la hora de explorar los datos durante las numerosas iteraciones de la generación del modelo.

Supernodos y almacenamiento en caché

Desde un supernodo, todos los nodos excepto los terminales se pueden almacenar en caché. El almacenamiento en caché se controla pulsando con el botón derecho en un nodo y seleccionando una de las diversas opciones del menú contextual Caché. Esta opción de menú está disponible desde el exterior de un supernodo y para los nodos encapsulados en un supernodo.

Hay varias directrices para las cachés del supernodo:

- Si hay algún nodo encapsulado en un supernodo que tenga el almacenamiento en caché activado, el supernodo también lo tendrá.
- La desactivación de la caché en un supernodo desactiva la caché de *todos* los nodos encapsulados.
- La activación del almacenamiento en caché en un supernodo, activa realmente la caché en el último subnodo en el que se puede realizar dicho almacenamiento. Es decir, si el último subnodo es un nodo Seleccionar, la caché se activará para dicho nodo. Si el último subnodo es un nodo terminal (que no permite el almacenamiento en caché), el siguiente nodo por encima que admita el almacenamiento en caché, estará activado.
- Una vez haya definido las cachés para los subnodos de un supernodo, las actividades por encima del nodo de cachés, como añadir o editar nodos, vaciarán las cachés.

Supernodos y scripts

Puede usar el lenguaje de scripts de IBM SPSS Modeler para escribir programas simples que manipulen y ejecuten el contenido de un supernodo terminal. Por ejemplo, puede que le resulte útil especificar el orden de ejecución de una ruta compleja. A modo de ejemplo, si un supernodo contiene un nodo Val. globales que debe ejecutarse antes que un nodo Gráfico, podrá crear un script que ejecute primero el nodo Val. globales. Los valores calculados por este nodo, como la media o la desviación estándar, se pueden usar posteriormente cuando se ejecute el nodo Gráfico.

La pestaña Script del cuadro de diálogo Supernodo sólo estará disponible para los supernodos terminales.

Para abrir el cuadro de diálogo de scripts para un supernodo terminal

- Pulse con el botón derecho en el lienzo del supernodo y seleccione **Script de Supernodo**.
- También puede seleccionar, tanto en el modo de acercamiento como en el de alejamiento, la opción **Script de Supernodo** del menú Supernodo.

Note: los scripts de supernodo sólo se ejecutan con la ruta y el supernodo cuando haya seleccionado **Ejecutar este script** en el cuadro de diálogo.

Las opciones específicas de los scripts y su uso en IBM SPSS Modeler se describen en la *Guía de scripts y automatización*, disponible en IBM SPSS Modeler DVD.

Almacenamiento y carga de supernodos

Una de las ventajas de los supernodos es que se pueden guardar y reutilizar en otras rutas. Al guardar y cargar supernodos, tenga en cuenta que usan una extensión *.slb*.

Para guardar un supernodo

1. Acérquese al supernodo.
2. En el menú Supernodo, seleccione **Guardar Supernodo**.
3. Especifique un nombre de archivo y directorio en el cuadro de diálogo.
4. Seleccione si desea añadir el supernodo guardado al proyecto actual.
5. Pulse en **Guardar**.

Para cargar un supernodo

1. En el menú Insertar de la ventana de IBM SPSS Modeler, seleccione **Supernodo**.
2. Seleccione un archivo de Supernodo (*.slb*) en el directorio actual o examine uno diferente.
3. Pulse en **Cargar**.

Note: los supernodos importados tiene los valores predeterminados para todos los parámetros. Para cambiar los parámetros, pulse dos veces en un supernodo del lienzo de rutas.

Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Es posible que IBM no ofrezca los productos, servicios o funciones que se tratan en este documento en otros países. Póngase en contacto con el representante local de IBM si desea obtener información sobre los productos y servicios disponibles en su zona. Las referencias a un producto, programa o servicio IBM no pretenden establecer ni implicar que sólo se pueda utilizar dicho producto, programa o servicio IBM. Se puede utilizar en su lugar cualquier producto, programa o servicio funcionalmente equivalente que no vulnere ningún derecho de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

Puede que IBM tenga patentes o solicitudes de patente pendientes que cubran la materia descrita en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing
Ley de propiedad intelectual y legal
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japón

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplica a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar en cualquier momento mejoras o cambios en los productos o programas descritos en esta publicación sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información solamente es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales incluidos en dichos sitios web no forman parte de los materiales de este producto de IBM y el uso de dichos sitios web se realiza por cuenta y riesgo del usuario.

IBM puede utilizar o distribuir cualquier información que proporcione en la forma que crea adecuada sin incurrir por ello en ninguna obligación con el cliente.

Los propietarios de licencia de este programa que deseen tener información sobre el mismo con el fin de: (i) intercambiar información entre programas creados de forma independiente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deberán ponerse en contacto con:

Tel.
900 100 400
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
Estados Unidos de América

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

El programa con licencia descrito en este documento y todo el material bajo licencia disponible para el mismo se proporciona por parte de IBM bajo los términos de los acuerdos IBM Customer Agreement, IBM International Program License Agreement o cualquier acuerdo equivalente entre las dos partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas mediciones se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas medidas sean las mismas en los sistemas comerciales. Además, es posible que algunas mediciones hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado dichos productos y no puede confirmar la precisión en cuanto al rendimiento, la compatibilidad u otras características relacionadas con productos que no son de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM u otras empresas. Encontrará una lista actual de marcas registradas de IBM en la Web en "Información de copyright y marca registrada" en www.ibm.com/legal/copytrade.shtml.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y logotipos basados en Java son marcas comerciales o marcas registradas de Oracle y/o sus afiliados.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

Glosario

A

Asimetría. Medida de la asimetría de una distribución. La distribución normal es simétrica y tiene un valor de sesgo de 0. Una distribución con un sesgo positivo significativo tiene una cola derecha larga. Una distribución que tenga una asimetría negativa significativa tiene una cola izquierda larga. Como regla aproximada, un valor de la asimetría mayor que el doble de su error estándar se asume que indica una desviación de la simetría.

C

Covarianza. Medida no tipificada del grado de asociación entre dos variables igual a la desviación del producto vectorial dividido por $N-1$.

Curtosis. Es una medida del grado en que las observaciones se agrupan en torno a un punto central. Para una distribución normal, el valor del estadístico de curtosis es 0. Una curtosis positiva indica que, con respecto a una distribución normal, las observaciones se concentran más en el centro de la distribución y presentan colas más estrechas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución leptocúrtica son más gruesas con respecto a una distribución normal. Una curtosis negativa indica que, con respecto a una distribución normal, las observaciones se concentran menos y presentan colas más gruesas hasta los valores extremos de la distribución, en cuyo punto las colas de la distribución platicúrtica son más estrechas con respecto a una distribución normal.

D

desviación estándar. Una medida de la dispersión en torno a la media, igual a la raíz cuadrada de la varianza. La desviación estándar se mide en las mismas unidades que la variable original.

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos se encuentra dentro de una desviación estándar de la media y el 95% queda entre dos desviaciones estándar. Por ejemplo, si la edad media es de 45 años, con una desviación estándar de 10, el 95% de los casos estaría entre los 25 y 65 en una distribución normal.

E

Error estándar. Medida de cuánto puede variar el valor de un estadístico de contraste de muestra en muestra. Es la desviación estándar de la distribución muestral de un estadístico. Por ejemplo, el error estándar de la media es la desviación estándar de las medias muestrales.

Error estándar de asimetría. La razón de la asimetría sobre su error estándar puede utilizarse como una prueba de normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la asimetría indica una cola larga a la derecha; un valor extremo y negativo indica una cola larga por la izquierda.

Error estándar de curtosis. La razón de la curtosis sobre su error estándar puede utilizarse como prueba de normalidad (es decir, se puede rechazar la normalidad si la razón es menor que -2 o mayor que +2). Un valor grande y positivo para la curtosis indica que las colas son más largas que las de una distribución normal; por el contrario, un valor extremo y negativo indica que las colas son más cortas (llegando a tener forma de caja como en la distribución uniforme).

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra, extraídas éstas de la misma distribución. Puede utilizarse para comparar de forma aproximada la media observada respecto a un valor hipotetizado (es decir, se puede concluir que los dos valores son distintos si la diferencia entre ellos, dividida por el error estándar, es menor que -2 o mayor que +2).

Exclusivo. Evalúa todos los efectos simultáneamente, corrigiendo cada efecto por todos los demás efectos de cualquier tipo.

M

Máximo. Se trata del valor mayor de una variable numérica.

Media. Una medida de tendencia central. El promedio aritmético, la suma dividida por el número de casos.

Mediana. Es el valor por encima y por debajo del cual se encuentran la mitad de los casos, el percentil 50. Si hay un número par de casos, la mediana es la media de los dos valores centrales, cuando los casos se ordenan en orden ascendente o descendente. La mediana es una medida de tendencia central que no es sensible a los valores atípicos (a diferencia de la media, que puede resultar afectada por unos pocos valores extremadamente altos o bajos).

Mínimo. Se trata del valor menor de una variable numérica.

Moda. El valor que ocurre con mayor frecuencia. Si varios valores comparten la mayor frecuencia de aparición, cada uno de ellos es un modo.

R

Rango. Diferencia entre los valores mayor y menor de una variable numérica; el máximo menos el mínimo.

S

Suma. Suma o total de todos los valores, a lo largo de todos los casos con valores no perdidos.

V

Válidos. Los casos válidos que no tienen un valor perdido del sistema ni un valor definido como perdido del usuario.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media dividida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las de la variable en cuestión.

Índice

A

- aciertos
 - opciones del diagrama de evaluación 262
- actualidad
 - configuración de fecha relativa 82
- acumulación de datos de series temporales 172
- adición
 - registros 78
- adición de paneles 188
- agregación de datos de series temporales 172, 184, 185
- agregación de registros 167
- ajustes -automáticos- 280
- almacenamiento 132
 - conversión 149, 150
 - objetos de resultados 296, 300
 - resultados 296
- almacenamiento de campos
 - conversión 149
- animación en gráficos 188
- anonimización nombres de campos 141
- ANOVA
 - nodo Medias 320
- ANOVA unidireccional
 - nodo Medias 320
- anti-uni3n 84
- a3o fiscal
 - nodo Intervalos de tiempo 176
- apertura
 - objetos de resultados 296
- aplicaciones de ayuda 336
- aproximaci3n cuartil 81
- aproximaci3n mediana 81
- archivo de datos employee_data.sav 367
- archivos .dbf 64
- archivos .sav 366
- archivos .sd2 (SAS) 41
- archivos .shp 64
- archivos .slb 382
- archivos .ssd (SAS) 41
- archivos .tpt (SAS) 41
- archivos .zsv 366
- archivos DAT
 - almacenamiento 300
 - exportaci3n 298, 362
- archivos de datos de IBM SPSS Statistics
 - importaci3n de datos de encuesta 31
- archivos de Excel
 - exportaci3n 362
- archivos de formato 41
- archivos de mapas
 - cambio del nombre 219
 - eliminaci3n 219
 - exportaci3n 219
 - importaci3n 219
 - seleccionar en el selector de plantillas de tablero 196
 - ubicaci3n 218

- archivos de resultados
 - almacenamiento 300
- archivos de texto 25
 - exportaci3n 362
- archivos de transporte
 - Nodo de origen SAS 41
- archivos delimitados por comas
 - almacenamiento 300
 - exportaci3n 298, 362
- archivos ESRI 219
- archivos planos 25
- archivos SMZ
 - cambio del nombre 219
 - conceptos b3sicos 219
 - creaci3n 219
 - edici3n de archivos SMZ preinstalados 219
 - eliminaci3n 219
 - exportaci3n 219
 - importaci3n 219
 - preinstalados 219
- archivos XLS
 - exportaci3n 362
- asignaci3n de tipos de datos 109
- atributos
 - en mapas 220
- atributos de campo 137
- atributos de tipo 137
- auditoría
 - auditoría de datos iniciales 308
 - Nodo Auditoría de datos 308
- autorregresi3n
 - modelos ARIMA 101

B

- base de datos
 - carga masiva 347, 348
- bases de datos de ADO
 - importaci3n 31
- bases de datos de gran tama3o 71
 - realizaci3n de una auditoría de datos 308
- bases de datos de In2data
 - importaci3n 31
- bases de datos de Quantvert
 - importaci3n 31
- bloqueo de supernodos 377
- borrar valores 68
- b3squeda
 - explorador de tablas 301

C

- caché
 - Supernodos 381
- Cadena de línas 129
- Cadena multilínea 129
- calcular duraciones
 - preparaci3n autom3tica de datos 114

- c3lculo de duraciones
 - preparaci3n autom3tica de datos 114
- calidad de datos
 - explorador de auditoría de datos 312
- cambio del nombre
 - archivos de mapas 219
 - campos para exportar 374
 - hojas de estilo de visualizaci3n 219
 - plantillas de visualizaci3n 219
- cambio del nombre de los objetos de resultados. 296
- Campo EtiquetaTiempo
 - nodo Intervalos de tiempo 171
- Campo ÍndiceTiempo
 - nodo Intervalos de tiempo 171
- campo Recuento
 - nodo Intervalos de tiempo 172
 - series temporales de agregaci3n o relleno 172
- campos
 - anonimizaci3n datos 153
 - derivaci3n de m3ltiples campos 144
 - etiquetas de valor y de campo 132
 - reordenaci3n 183
 - selecci3n m3ltiple 145
 - transposici3n 169
- campos clave 79, 167
- campos claves primarios
 - nodo Exportar base de datos 342
- campos de correlaciones 341
- campos de partici3n 136, 165
- capas en mapas geoespaciales 265
- caracteres de comentario
 - en archivos variables 26
- caracteres de EOL 26
- características
 - en mapas 220
- carga masiva 347, 348
- casos reservados
 - modelado de series temporales 173
- chi-cuadrado
 - nodo Matriz 304
- Chi-cuadrado de Pearson
 - nodo Matriz 304
- cifras decimales
 - formatos de presentaci3n 138
- clasificar casos 159
- clave de recuento para la agregaci3n 79
- claves contiguas 81
- cl3ster 279, 288
- codificaci3n de variables dummy 167
- Cognos, consulte IBM Cognos BI 38
- comando CREATE INDEX 345
- combinaci3n de datos 92
 - de varios archivos 84
- comentarios
 - uso con supernodos 378
- comillas
 - importaci3n de archivos de texto 26
 - para exportar base de datos 340
- comprobaci3n de tipos 136

- condiciones
 - clasificadas 88
 - especificación de una serie 148
 - especificaciones para una fusión 87
- condiciones clasificadas
 - especificaciones para una fusión 88
- conexiones
 - al repositorio de IBM SPSS Collaboration and Deployment Services 65
- conexiones a la base de datos
 - definición 20
 - valores predefinidos 20
- conjunto de datos principal 92
- conjuntos
 - conversión a marcas 167
 - transformación 151, 152
- conjuntos de categorías múltiples 141
- conjuntos de dicotomía múltiple 141
- conjuntos de respuestas múltiples
 - conjuntos de categorías múltiples 141
 - conjuntos de dicotomía múltiple 141
 - definición 141
 - eliminación 141
 - en visualizaciones 194
 - nodo de origen de IBM SPSS Statistics 366
 - nodo de origen Recopilación de datos de IBM SPSS 31, 35
- consultas
 - Nodo de origen Base de datos 17, 19
- Consultas SQL
 - Nodo de origen Base de datos 17, 19, 24, 25
- conversión de conjuntos a marcas 167
- convertir niveles de medición 130
- coordenadas polares 287
- copia de atributos de tipo 137
- copia de visualizaciones 291
- coropleta
 - ejemplo 214
- correlación
 - datos para exportar a IBM Cognos TM1 360
- correlaciones 318
 - etiquetas descriptivas 318
 - nodo Medias 322
 - probabilidad 318
 - resultado de estadísticos 319
 - significación 318
 - valor absoluto 318
- Correlaciones de Pearson
 - nodo Medias 322
 - resultado de estadísticos 319
- costes
 - diagramas de evaluación 260
- creación
 - campos nuevos 142, 143
- CRISP-DM
 - comprensión de los datos 7
- cuantiles
 - nodo Intervalos 157

D

- datos
 - agregación 78
 - almacenamiento 149, 150
 - anonimización 153
 - auditoría 308
 - comprensión 71
 - exploración 308
 - preparación 71
 - tipo de almacenamiento 132
- datos anuales
 - nodo Intervalos de tiempo 176
- datos categóricos 130
- datos continuos 130, 133
- Datos CSV
 - importación 31
- datos de casos
 - nodo de origen de IBM SPSS Data Collection 31
 - nodo de origen Recopilación de datos de IBM SPSS 31
- datos de encuesta
 - importación 31, 35
 - nodo de origen de IBM SPSS Data Collection 31
- datos de encuesta de IBM SPSS Data Collection
 - importación 31
- datos de investigación de mercados
 - importación 31, 35
 - nodo de origen Recopilación de datos de IBM SPSS 31, 35
- datos de muestreo 77
- datos de Quancept
 - importación 31
- datos de Quantum
 - importación 31
- datos de resumen 78
- datos de serie temporal
 - agregación 170, 172, 184
 - casos reservados 173
 - definición 170, 171, 172, 173
 - etiquetado 170, 171, 172, 173
 - generación a partir de datos 172
 - intervalos 171
 - período de estimación 173
 - relleno 170, 172
- datos de Surveycraft
 - importación 31
- datos de texto de campo fijo 29
- datos de texto de campo libre 25
- datos de texto delimitado 25
- datos decrecientes 72, 73
- datos desequilibrados 77
- datos geoespaciales 25
 - derivación 146
 - en archivos variables 28
 - exportar 355
 - fusión 88
 - fusión de condiciones clasificadas 88
 - importar 26
 - listas en archivos de variables 28
 - restricciones 129
- datos geoespaciales, reproyección de 185
- datos geoespaciales en mapas 265
- datos insesgados 77

- datos mensuales
 - nodo Intervalos de tiempo 177
- datos nominales 134
- datos ordinales 134
- datos retardados 182
- datos semanales
 - nodo Intervalos de tiempo 177
- datos sesgados 77
- datos simulados
 - nodo Generar simulación 49
- datos sintéticos
 - nodo de datos de usuario 44
- datos trimestrales
 - nodo Intervalos de tiempo 176
- Datos triple-S
 - importación 31
- definición automática de tipos 127, 131
- definición de densidad en Cajas-Espacio-Tiempo 107
- Definiciones de proveedor de datos 65
- delimitadores 26, 347
- densidad
 - 3-D 198
 - densidad en 3D 198
- derivación múltiple 144
- desbloqueo de supernodos 377
- descarte
 - campos 139
- desviación estándar
 - nodo Intervalos 159
 - nodo Val. globales 325
 - resultado de estadísticos 319
- desviación estándar para la agregación 79
- diagrama de dispersión 198
 - 3-D 198
 - en intervalos 198
 - en intervalos hexagonales 198
- diagrama de dispersión en 3D 198
- diagrama de dispersión en intervalos 198
 - intervalos hexagonales 198
- diagrama de dispersión en intervalos hexagonales 198
- diagramas de dispersión 227, 233
- dibujos de asociaciones 248
- dirección de campos 136
- diseño de red para gráficos Malla 250
- diseño direccional para gráficos Malla 250
- distribución 243
- documentación 3
- documentos MDD
 - importación 31
- DPD 65
- duplicados
 - campos 84, 140
 - registros 93

E

- edición de gráficos
 - tamaño de los elementos gráficos 282
- edición de visualizaciones 279
 - adición de efectos 3-D 287
 - categorías 285
 - colores y tramas 281

- edición de visualizaciones (*continuación*)
 - combinación de categorías 285
 - configuración automática 280
 - contracción de categorías 285
 - ejes 284
 - escalas 284
 - estilo de guión 281
 - exclusión de categorías 285
 - forma de puntos 282
 - formatos de numeración 283
 - márgenes 283
 - ordenación de categorías 285
 - paneles 287
 - posición de la leyenda 290
 - reglas 280
 - relación de aspecto de puntos 282
 - relleno 283
 - rotación de puntos 282
 - selección 280
 - texto 281
 - transformación de los sistemas de coordenadas 287
 - transparencia 281
 - transponer 287
 - editor de consultas
 - Nodo de origen Base de datos 24, 25
 - ejecución
 - especificación del orden de 381
 - ejemplos
 - conceptos básicos 5
 - Manual de aplicaciones 3
 - ejemplos de aplicaciones 3
 - elementos de hora cíclicos
 - preparación automática de datos 114
 - elementos de resultado 333
 - elementos gráficos
 - cambio 288
 - conversión 288
 - modificadores de colisión 288
 - eliminación
 - archivos de mapas 219
 - hojas de estilo de visualización 219
 - objetos de resultados 296
 - plantillas de visualización 219
 - empates
 - nodo Intervalos 157
 - encapsulación de nodos 376
 - enlace a lo largo de las columnas 347
 - enlace a lo largo de las filas 347
 - enlaces
 - nodo Malla 250
 - entradas múltiples 84
 - error estándar de la media
 - resultado de estadísticos 319
 - escenario 65
 - esquema
 - nodo Exportar base de datos 342
 - esquivar 279, 288
 - establecer semilla aleatoria
 - registros de muestreo 165
 - estadístico de la evaluación del rendimiento 305
 - estadístico F
 - nodo Medias 322
 - estadísticos
 - edición de visualizaciones 288
 - Nodo Auditoría de datos 308
 - estadísticos (*continuación*)
 - nodo Matriz 302
 - estadísticos de resumen
 - Nodo Auditoría de datos 308
 - etiquetas 84, 90, 134
 - especificación 132, 133, 134
 - exportación 361, 373
 - importación 41, 366
 - etiquetas de campo
 - etiquetado de registros en el resultado 136
 - etiquetas de valor
 - Nodo Archivo Statistics 366
 - etiquetas de variable
 - Nodo Archivo Statistics 366
 - Nodo Exportación de Statistics 373
 - evaluación de modelos 304
 - evaluación del modelo 256
 - eventos
 - creación 236
 - Excel
 - inicio desde IBM SPSS Modeler 362
 - exclusión de campos no utilizados
 - preparación automática de datos 113
 - exploración de datos
 - Nodo Auditoría de datos 308
 - exploración de gráficos 269
 - marcación de elementos 275
 - regiones 273
 - tramos de gráficos 270
 - varita mágica 275
 - explorador de análisis
 - interpretación 306
 - explorador de auditoría de datos
 - generación de gráficos 315
 - generación de nodos 315
 - Menú Archivo 310
 - Menú Edición 310
 - explorador de calidad
 - generación de nodos Filtrar 314
 - generación de nodos Seleccionar 314
 - explorador de estadísticos
 - generación de nodos Filtrar 319
 - interpretación 319
 - Menú Generar 319
 - explorador de informes 324
 - explorador de matrices
 - Menú Generar 304
 - explorador de tablas
 - búsqueda 301
 - Menú Generar 301
 - reordenación de columnas 299, 301
 - selección de casillas 299, 301
 - exportación
 - archivos de mapas 219
 - datos procedentes de IBM Cognos TM1 360
 - hojas de estilo de visualización 219
 - plantillas de visualización 219
 - resultados 298
 - Supernodos 382
 - exportación de Analytic Server 356
 - exportación de cifras decimales 138
 - exportación de datos
 - a Excel 362
 - a IBM SPSS Statistics 373
 - a una base de datos 340
 - exportación de datos (*continuación*)
 - archivos DAT 362
 - formato de archivo sin formato 355
 - Formato SAS 361
 - Formato XML 363
 - geoespacial 355
 - nodo Exportar de IBM Cognos BI 38, 357, 358
 - Nodo Exportar de IBM Cognos TM1 359
 - texto 362
 - expresiones CLEM 71
 - extensión
 - campo derivado 144
- ## F
- factores de equilibrado 78
 - factores de escala 78
 - fecha/hora 127
 - fechas
 - formatos de configuración 138
 - filas en blanco
 - archivos de Excel 42
 - filtrado de campos 90, 139
 - para IBM SPSS Statistics 374
 - formato de almacenamiento de lista 11
 - formato de presentación científica 138
 - formato de presentación de las monedas 138
 - Formato HDATA
 - nodo de origen Recopilación de datos de IBM SPSS 31
 - Formato VDATA
 - nodo de origen de IBM SPSS Data Collection 31
 - formatos
 - datos 9
 - formatos de almacenamiento 9
 - formatos de hora 138
 - formatos de presentación
 - científica 138
 - cifras decimales 138
 - moneda 138
 - números 138
 - símbolo de agrupación 138
 - formatos de presentación de los números 138
 - formatos de resultados 300
 - formatos shapefile 219
 - fórmula de derivación de campo 145
 - forzado de valores 136
 - frecuencias
 - nodo Intervalos 157
 - función de los más recientes
 - series temporales de relleno 172
 - función de media de los más recientes
 - series temporales de relleno 172
 - función hassubstring 147
 - función Máx
 - agregación de series temporales 172, 184, 185
 - función Media
 - agregación de series temporales 172, 184, 185

función Mín
 agregación de series temporales 172, 184, 185

función Moda
 agregación de series temporales 172, 184, 185

función Primeros
 agregación de series temporales 172

función Suma
 agregación de series temporales 172, 184, 185

función Último
 agregación de series temporales 172

función Vacío
 series temporales de relleno 172

función Verdadero si alguno es verdadero
 agregación de series temporales 172, 184, 185

funciones de transferencia 102
 delay 102
 órdenes de denominador 102
 órdenes de diferencia 102
 órdenes de numerador 102
 órdenes estacionales 102

fusión de orden 84

G

generación de marcas 167, 168

generación de nodos a partir de gráficos 276
 nodos Derivar 276
 Nodos Equilibrar 276
 Nodos Filtrar 276
 Nodos Reclasificar 276
 nodos Seleccionar 276

Generador de expresiones 71

Geoespacial
 configuración de opciones de importación 64

gestión de valores perdidos 109

gestor de resultados 296

gestores
 pestaña Resultados 296

grados de libertad
 nodo Matriz 304
 nodo Medias 322

gráfico circular 198
 3-D 198
 de un mapa 198
 ejemplo 211
 mediante recuentos 198

gráfico circular en 3D 198

gráfico de áreas 198
 3-D 198

gráfico de áreas en 3D
 descripción 198

gráfico de barras 198
 3-D 198
 de recuentos 198
 de un mapa 198
 ejemplo 206, 207

gráfico de barras acumuladas 198

gráfico de barras apiladas
 ejemplo 207

gráfico de barras en 3D 198

gráfico de burbujas 198

gráfico de caja 198
 ejemplo 210

gráfico de coordenadas paralelas 198

gráfico de línea 198
 de un mapa 198

gráfico de máximos-mínimos 198

gráfico de máximos-mínimos-cierre 198

gráfico de puntos 198
 2D 198
 ejemplo 209

gráfico de puntos en 2D 198

gráfico de ruta 198

gráfico de superficie 198

gráficos
 3-D 190
 almacenamiento 293
 almacenamiento de diseños modificados 292
 almacenamiento de modificaciones de diseño 292
 almacenamiento de resultados 300
 colecciones 245
 copia 293
 de tablero 192
 diagramas de evaluación 256
 distribuciones 239
 eliminación de regiones 273
 esquema de color
 predeterminado 292
 etiquetas de eje 291
 exploración 269
 exportación 293
 generación de nodos 276
 generación desde una auditoría de datos 315
 gráfico múltiple 233
 gráficos 227
 histogramas 243
 hoja de estilo 292
 impresión 293
 mallas 248
 nota al pie 291
 pestaña anotaciones 190
 pestaña resultados 189
 regiones 273
 rotación de una imagen 3D 190
 tamaño de los elementos gráficos 282
 título 291
 tramos 270
 variables de series temporales 236
 visualización de mapas 265

gráficos 3D 190

gráficos de asociación 248

gráficos de beneficios 256, 263

gráficos de elevación 256, 263

gráficos de ganancias 256, 263

gráficos de líneas 227, 233

gráficos de puntos 227, 233

gráficos de respuestas 256, 263

H

histograma 198
 3-D 198
 ejemplo 208

histograma en 3D 198

hojas de cálculo
 importación desde Excel 42

hojas de estilo
 cambio del nombre 219
 eliminación 219
 exportación 219
 importación 219

hojas de estilo de visualización
 aplicación 292
 cambio del nombre 219
 eliminación 219
 exportación 219
 importación 219
 ubicación 218

HTML
 almacenamiento de resultados 300

I

IBM SPSS Modeler 1
 documentación 3

IBM SPSS Modeler Server 1

IBM SPSS Statistics
 inicio desde IBM SPSS Modeler 336, 371, 373
 nombres de campos válidos 374
 ubicación de la licencia 336

iconos, IBM Cognos BI 36

igual frecuencia
 nodo Intervalos 157

importación
 archivos de mapas 219
 datos de IBM Cognos BI 37
 datos procedentes de IBM Cognos TM1 40
 hojas de estilo de visualización 219
 informes de IBM Cognos BI 38
 plantillas de visualización 219
 Supernodos 382

importancia
 comparación de medias 321
 nodo Medias 322

impresión de resultados 296

incrementos por minuto
 nodo Intervalos de tiempo 180

incrementos por segundo
 nodo Intervalos de tiempo 180, 181

indexación de tablas de bases de datos 345

índices BITMAP
 tablas de bases de datos 345

informe de calidad
 explorador de auditoría de datos 312

informes
 almacenamiento de resultados 300

ingresos
 diagramas de evaluación 260

instanciación 127, 130, 131
 nodo de origen 68

instrucciones if-then-else caso contrario 149

intervalos
 datos de serie temporal 170, 184

intervalos de confianza
 nodo Medias 322

intervalos de cuartil 157

intervalos de decil 157

intervalos de percentil 157
intervalos de quintil 157
intervalos de veintil 157
intervalos óptimos 160
intervalos supervisados 160
intervenciones
 creación 236

J

jitter 279, 288

L

lengua
 nodo de origen Recopilación de datos
 de IBM SPSS 34
leyenda
 posición 290
línea base
 opciones del diagrama de
 evaluación 260
lista 11, 127
 derivación 146
 longitud máxima 132
 niveles de medición geoespacial 129
 profundidad 11
listas
 tipo de datos de colección 135
 tipo de datos geoespaciales 135
listas en archivos de variables 28

M

mapa
 colores 198
 con flechas 198
 con gráficos circulares 198
 con gráficos de barras 198
 con gráficos de líneas 198
 con puntos 198
 superponer 198
mapa de calor 198
 ejemplo 212
mapa de color 198
 ejemplo 214
mapa de coordenadas 198
mapa de coropletas 198
mapa de flujos 198
mapa de superposición 198
mapas
 combinación de características 224
 conversión de los formatos ESRI
 shapefile 219
 distribución 226
 eliminación de características 225
 eliminación de elementos
 individuales 225
 etiquetas de características 223
 mover características 225
 proyección 225
 reducir 221, 222
 suavizar 221, 222
mapas en formato shapefile
 con la utilización del selector de
 plantillas de tablero 219

mapas en formato shapefile (*continuación*)
 conceptos 220
 edición de mapas SMZ
 preinstalados 219
 tipos 220
marca de tiempo 127
marcación de elementos 273, 275
marco de muestreo 73
matriz de coincidencias
 Nodo Análisis 305
matriz de diagramas de dispersión
 ejemplo 213, 215
matriz de diagramas de dispersión
 (SPLOM) 198
máximo
 nodo Val. globales 325
 resultado de estadísticos 319
media
 nodo Intervalos 159
 nodo Val. globales 325
 resultado de estadísticos 319
media/desviación estándar
 uso para segmentar campos 159
media móvil
 modelos ARIMA 101
mediana
 resultado de estadísticos 319
medias
 comparación 320, 321
mediciones diarias
 nodo Intervalos de tiempo 178
mediciones por horas
 nodo Intervalos de tiempo 179
mejor línea
 opciones del diagrama de
 evaluación 260
metadatos 132
 nodo de origen Recopilación de datos
 de IBM SPSS 31
método clave 84
miembro (importar SAS)
 configuración 41
mínimo
 nodo Val. globales 325
 resultado de estadísticos 319
mode
 resultado de estadísticos 319
modelizador experto
 criterios en modelos de series
 temporales 100
modelo del proceso CRISP-DM
 preparación de datos 109
modelos
 anonimizar datos para 153
 ARIMA 101
modelos ARIMA
 constant 101
 criterios en modelos de series
 temporales 101
 funciones de transferencia 102
 órdenes autorregresivos 101
 órdenes de diferenciación 101
 órdenes de media móvil 101
 órdenes estacionales 101
 valores atípicos 103
modelos de IBM SPSS Statistics 370
 acerca de 370

modelos de IBM SPSS Statistics
 (*continuación*)
 detalles avanzados de nugget 370
 nugget de modelo 370
 opciones de modelo 370
modelos de series temporales
 criterios ARIMA 101
 criterios de suavizado
 exponencial 100
 criterios del modelizador experto 100
 funciones de transferencia 102
 periodicidad 102
 transformación de serie 102
 valores atípicos 103
modificación de valores de datos 142
modificadores de colisión 288
muestras agrupadas en clústeres 73, 74,
 75
muestras de comprobación
 partición de datos 165
muestras de entrenamiento
 equilibrado 78
 partición de datos 165
muestras de validación
 partición de datos 165
muestras estratificadas 73, 74, 75, 77
muestras no aleatorias 73, 74
muestras ponderadas 75
muestras sistemáticas 73, 74
muestreo 1-de cada-n 74
muestreo de datos contiguos 74
Multipolígono 129
Multipunto 129

N

navegación 331
nivel de medición
 cambio de visualizaciones 192
 colección 11, 135
 definido 127
 en visualizaciones 194
 geoespacial 11, 129, 135, 145
 recopilación 145
 restricciones en los datos
 geoespaciales 129
nivel de medición de colección 135, 145
nivel de medición geoespacial 135, 145
niveles de medición geoespacial 11, 127,
 129
Nodo Adición de RFM
 conceptos básicos 82
 opciones de configuración 82
nodo Agregar
 aproximación para cuartil 81
 aproximación para mediana 81
 conceptos básicos 78
 configuración de optimización 81
 opciones de configuración 79
 procesamiento paralelo 79
 rendimiento 79
Nodo Análisis 304
 pestaña análisis 305
 pestaña Resultado 300
Nodo Análisis de RFM
 conceptos básicos 161
 configuración 162

- Nodo Análisis de RFM (*continuación*)
 - valores de intervalos 163
- nodo Anonimizar
 - conceptos básicos 153
 - creación de valores anonimizados 154
 - opciones de configuración 154
- nodo Añadir
 - campos de etiquetado 90
 - coincidencia de campos 92
 - conceptos básicos 92
 - opciones de configuración 92
- nodo Archivo fijo
 - conceptos básicos 29
 - opciones de configuración 29
 - reconocimiento automático de fecha 29
- Nodo Archivo Statistics 366
- nodo Archivo var. 25
 - opciones de configuración 26
 - reconocimiento automático de fecha 26
- nodo Archivo variable
 - importar datos geoespaciales 28
 - metadatos geoespaciales 28
- Nodo Auditoría de datos 308
 - pestaña Configuración 308
 - pestaña Resultado 300
- nodo Cajas-Espacio-Tiempo
 - definición de densidad 107
- Nodo Cajas-espacio-tiempo
 - conceptos básicos 105
- nodo Colección 245
 - pestaña Aspecto 246
 - pestaña Opciones 245, 246
 - utilización del gráfico 247
- Nodo Conjunto
 - campos de salida 163
 - combinación de puntuaciones 163
- Nodo de ajuste de distribución 325
 - ajuste de distribución 326
 - configuración de resultados 327
 - pestaña Configuración 327
- nodo de archivo caché 366
- nodo de datos de usuario
 - conceptos básicos 44
 - opciones de configuración 45
- nodo de exportación Archivo sin formato 355
 - pestaña Exportar 355
- nodo de exportación de IBM Cognos TM1
 - correlación de datos de exportación 360
 - exportación de datos 360
- nodo de exportación de ODBC. Consulte el nodo Exportar base de datos 340
- nodo de exportación de Recopilación de datos de IBM SPSS 356
- Nodo de exportación Excel 362
- nodo de exportación XML 363
- Nodo de origen Base de datos 17
 - Consultas SQL 19
 - editor de consultas 24, 25
 - selección de tablas y vistas 23
- Nodo de origen de Excel 42
- Nodo de origen de IBM Cognos BI 36, 38, 39
 - Nodo de origen de IBM Cognos BI (*continuación*)
 - iconos 36
 - importación de datos 37
 - importación de informes 38
 - nodo de origen de IBM SPSS Data Collection 31
 - Nodo de origen de Microsoft Excel 42
 - Nodo de origen geoespacial
 - archivos .dbf 64
 - archivos .shp 64
 - Servicio de mapas 64
 - nodo de origen Recopilación de datos de IBM SPSS 31, 35
 - archivos de metadatos 31
 - archivos de registro 31
 - configuración de conexión de base de datos 35
 - conjuntos de respuestas múltiples 35
 - lengua 34
 - tipos de etiquetas 34
 - Nodo de origen SAS
 - archivos .sd2 (SAS) 41
 - archivos .ssd (SAS) 41
 - archivos .tpt (SAS) 41
 - archivos de transporte 41
 - Nodo de origen XML 43
 - Nodo de preparación de datos automática 111
 - nodo de transformación 315
 - nodo de visualización de mapas 265
 - opciones de cambio de capa 266
 - pestaña Gráfico 265
 - nodo Derivar
 - conceptos básicos 142
 - condicional 149
 - conversión de almacenamiento de campos 149
 - derivación de un campo de lista 146
 - derivación de un campo geoespacial 146
 - derivación múltiple 144
 - estado 148
 - formula 145
 - generación a partir de enlaces de gráficos Malla 252
 - generación a partir de gráficos 276
 - generación a partir de preparación automática de datos 124
 - generación desde intervalos 155
 - generación desde un nodo Intervalos 161
 - marca 147
 - nominal 147
 - nueva codificación de valores 149
 - opciones de configuración 143
 - recuento 148
 - valor de fórmula 145
 - valores de colección 145
 - valores geoespaciales 145
- nodo Distinguir
 - conceptos básicos 93
 - configuración de optimización 95
 - ordenación de registros 93
 - valores compuestos 95
 - valores de compuesto 97
- nodo Distribución 239
 - nodo Distribución (*continuación*)
 - pestaña Aspecto 240
 - pestaña Gráfico 239
 - utilización de la tabla 240
 - utilización del gráfico 240
 - nodo Enterprise View 65
 - nodo Equilibrar
 - conceptos básicos 77
 - generación a partir de gráficos 276
 - opciones de configuración 78
 - nodo Estadísticos 318
 - correlaciones 318
 - estadísticos 318
 - etiquetas de correlación 318
 - pestaña Configuración 318
 - pestaña Resultado 300
 - nodo Evaluación 256
 - condición de acierto 262
 - expresión de puntuación 262
 - pestaña Aspecto 263
 - pestaña Gráfico 260
 - pestaña Opciones 262
 - regla de negocio 262
 - resultados de la lectura 263
 - utilización del gráfico 264
 - nodo Evaluación de simulación
 - pestaña Configuración 328
 - Nodo Evaluación de simulación 328, 331, 333, 335
 - configuración de resultados 328
 - nodo evaluation
 - pestaña Aspecto 269
 - Nodo Exportación de Statistics 373
 - pestaña Exportar 373
 - nodo Exportar base de datos 340
 - correlación de campos de datos de origen a columnas de base de datos 341
 - esquema 342
 - indexación de tablas 345
 - nombre de tabla 340
 - opciones de fusión 341
 - origen de datos 340
 - pestaña Exportar 340
 - nodo Exportar de IBM Cognos BI 38, 357, 358
 - Nodo Exportar de IBM Cognos TM1 359
 - Nodo Exportar SAS 361
 - nodo Filtrar
 - conceptos básicos 139
 - conjuntos de respuestas múltiples 141
 - opciones de configuración 140
 - nodo Fundir 84
 - campos de etiquetado 90
 - conceptos básicos 84
 - configuración de optimización 91
 - filtrado de campos 90
 - opciones de configuración 86, 87, 88
 - nodo G. múltiple 233
 - pestaña Aspecto 235
 - pestaña Gráfico 233
 - utilización del gráfico 235
 - nodo Generación de análisis TS
 - conceptos básicos 98
 - opciones de campos 98
 - opciones de despliegue 104

- nodo Generación de análisis TS (continuación)
 - opciones de modelo 99
- nodo Generar simulación
 - conceptos básicos 49
 - opciones de configuración 50
- nodo Gráfico 227
 - pestaña Aspecto 232
 - pestaña Gráfico 229
 - pestaña Opciones 230
 - utilización del gráfico 232
- nodo Gráfico de tiempo 236
 - pestaña Aspecto 238
 - pestaña Gráfico 237
 - utilización del gráfico 238
- nodo Histograma 243
 - pestaña Aspecto 244
 - pestaña Gráfico 243
 - utilización del gráfico 244
- nodo Historial 182
 - conceptos básicos 182
- nodo Importar a Excel
 - generación desde resultados 362
- nodo Informe 323
 - pestaña Plantilla 323
 - pestaña Resultado 300
- nodo Intervalos
 - conceptos básicos 155
 - igual frecuencia 157
 - igual suma 157
 - intervalos de ancho fijo 156
 - intervalos de desviación estándar/media 159
 - opciones de configuración 156
 - óptimos 160
 - rangos 159
 - vista preliminar de intervalos 161
- nodo Intervalos de tiempo 171, 172, 173
 - Analytic Server 184
 - conceptos básicos 170
- Nodo Intervalos de tiempo de Analytic Server
 - visión general 184
- nodo Intervalos de tiempo de AS 184, 185
 - visión general 184
- nodo Malla 248
 - ajuste de puntos 252
 - ajuste de umbrales 254
 - cambiar diseño 252
 - control deslizante 252
 - control deslizante de enlaces 252
 - definición de enlaces 250
 - pestaña Aspecto 252
 - pestaña Gráfico 249
 - pestaña Opciones 250
 - resumen malla 256
 - utilización del gráfico 252
- Nodo Marcas 167
- nodo Matriz 302
 - explorador de resultados 304
 - ordenación de filas y columnas 303
 - pestaña Aspecto 303
 - pestaña Configuración 302
 - pestaña Resultado 300
 - porcentajes de fila 303
 - porcentajes de la columna 303
- nodo Matriz (continuación)
 - resultado 303
 - tabulación cruzada 303
- nodo Medias 320
 - campos relacionados 321
 - explorador de resultados 321, 322
 - grupos independientes 320
 - importancia 321
 - pestaña Resultado 300
- nodo Muestrear
 - marco de muestreo 73
 - muestras agrupadas en clústeres 73, 74, 75
 - muestras aleatorias 73, 74
 - muestras estratificadas 73, 74, 75, 77
 - muestras no aleatorias 73, 74
 - muestras ponderadas 75
 - muestras sistemáticas 73, 74
 - Tamaños de muestra por estratos 77
- nodo Ordenar
 - conceptos básicos 83
 - configuración de optimización 84
- nodo Origen de IBM Cognos TM1 39
 - importación de datos 40
- Nodo Origen geoespacial
 - archivos .dbf 64
 - archivos .shp 64
 - Servicio de mapas 64
- nodo Partición 165
- nodo Reclassificar 151, 152
 - conceptos básicos 151, 155
 - generación desde una distribución 240
- nodo Reestructurar 167, 168
 - con nodo Agregar 167
- nodo Rellenar
 - conceptos básicos 149
- nodo Reorg. campos 183
 - opciones de configuración 183
 - orden personalizado 183
 - ordenación automática 183
- nodo Resultados de IBM SPSS Statistics
 - Pestaña Resultados 372
- Nodo Resultados de Statistics 371
 - pestaña Sintaxis 371
- nodo Seleccionar
 - conceptos básicos 72
 - generación a partir de enlaces de gráficos Malla 252
 - generación a partir de gráficos 276
- nodo Tabla 299
 - configuración de resultados 299
 - pestaña Configuración 299
 - pestaña Resultado 300
- Nodo Tablero 192
 - pestaña Aspecto 216
- nodo Tipo
 - borrado de valores 68
 - conceptos básicos 125
 - configuración del rol de modelado 136
 - copia de tipos 137
 - datos continuos 133
 - datos nominales 134
 - datos ordinales 134
 - opciones de configuración 127, 129, 130
- nodo Tipo (continuación)
 - tipo de campo de marcas 134
 - tipo de datos de colección 135
 - tipo de datos geoespaciales 135
 - tratamiento de los elementos vacíos 132
- Nodo Transformación de Statistics 367
 - opciones de configuración 367
 - pestaña Sintaxis 367
 - sintaxis permitida 368
- nodo Transponer 169
 - campos de cadena 169
 - campos numéricos 169
 - nombres de campos 169
- nodo Val. globales 324
 - pestaña Configuración 325
- nodo Vista de datos 62
 - opciones de configuración 63
- nodos de exportación 339
 - exportación de Analytic Server 356
- nodos de gráficos 187
 - animación 188
 - Colección 245
 - Distribución 239
 - Evaluación 256
 - G. múltiple 233
 - Gráfico 227
 - Gráfico de tiempo 236
 - Histograma 243
 - Malla 248
 - paneles 188
 - superposiciones 188
 - Tablero 192
 - visualización de mapas 265
- nodos de IBM SPSS Statistics 365
- nodos de operaciones con campos 109
 - generación desde una auditoría de datos 315
- nodos de operaciones con registros 71
 - nodo Intervalos de tiempo 170
 - nodo Intervalos de tiempo de AS 184
- nodos de origen
 - conceptos básicos 7
 - nodo Archivo fijo 29
 - Nodo Archivo Statistics 366
 - nodo Archivo var. 25
 - nodo de datos de usuario 44, 45
 - Nodo de origen Base de datos 17
 - Nodo de origen de Excel 42
 - Nodo de origen de IBM Cognos BI 36, 38, 39
 - Nodo de origen geoespacial 64
 - Nodo de origen SAS 41
 - Nodo de origen XML 43
 - nodo Enterprise View 65
 - nodo Generar simulación 49, 50
 - nodo Origen de IBM Cognos TM1 39
 - origen de Analytic Server 12
 - tipos de instanciación 68
- nodos de resultados 295, 299, 302, 304, 308, 318, 323, 324, 325, 326, 327, 328, 331, 333, 335, 371
 - pestaña Resultado 300
 - publicar en web 297
- nombres de campos 141
 - anonimización 141

- nombres de campos (*continuación*)
 - exportación de datos 340, 355, 361, 373
- nombres de variables
 - exportación de datos 340, 355, 361, 373
- normalizar objetivo continuo 115, 124
- normalizar valores
 - nodos de gráficos 233, 237
- nulos 132
 - en tablas Matriz 302

O

- ocultación datos para utilizarlos en un modelo 153
- ODBC
 - carga masiva a través de 347, 348
 - conexión para el nodo Exportar de IBM Cognos BI 358
 - Nodo de origen Base de datos 17
- opciones
 - IBM SPSS Statistics 336
- opciones de capa en mapas 266
- opciones de fusión, exportación de base de datos 341
- Opciones de gráfica 335
- opciones de modelo
 - nodo Modelo de Statistics 370
- Oracle 17
- orden ascendente 83
- orden de columna
 - explorador de tablas 299, 301
- orden de datos de entrada 90
- orden de ejecución
 - especificación 381
- orden descendente 83
- orden natural
 - alteración 183
- ordenación
 - campos 183
 - campos ordenados previamente 84, 95
 - nodo Distinguir 93
 - registros 83
- ordenación de datos 83, 183
- órdenes estacionales
 - modelos ARIMA 101
- origen de Analytic Server 12
- orígenes de datos
 - conexiones a la base de datos 20

P

- Palabra clave FILLFACTOR
 - indexación de tablas de bases de datos 345
- Palabra clave UNIQUE
 - indexación de tablas de bases de datos 345
- paletas
 - desplazamiento 280
 - ocultación 280
 - presentación 280
- parámetros
 - configuración de los supernodos 379

- parámetros (*continuación*)
 - en IBM Cognos BI 39
 - propiedades de los nodos 380
 - Supernodos 379, 380
- parámetros de ruta 24, 25
- parámetros de Supernodos 379, 380
- partición de datos 165
 - diagramas de evaluación 260
 - Nodo Análisis 305
- periodicidad
 - datos de serie temporal 170, 184
 - Modelizador de series temporales 102
- período de estimación 173
- períodos
 - nodo Intervalos de tiempo 175
- períodos cíclicos
 - nodo Intervalos de tiempo 175
- perturbación 230
- pestaña Sintaxis
 - Nodo Resultados de Statistics 371
- pila 279, 288
- planes de acceso de datos 63
- plantillas
 - cambio del nombre 219
 - eliminación 219
 - exportación 219
 - importación 219
 - nodo Informe 323
- plantillas de visualización
 - cambio del nombre 219
 - eliminación 219
 - exportación 219
 - importación 219
 - ubicación 218
- Polígono 129
- ponderaciones
 - diagramas de evaluación 260
- preparación automática de datos
 - análisis de campos 119
 - campos 113
 - configuración de campos 113
 - creación 116
 - detalles de acción 122
 - detalles de campo 121
 - enlaces entre vistas 118
 - excluir campos 114
 - exclusión de campos no utilizados 113
 - generación del nodo derivar 124
 - nombrar campos 117
 - normalizar objetivo continuo 115, 124
 - objetivos 111
 - poder predictivo 121
 - preparación de destinos 115
 - preparación de entradas 115
 - preparar fechas y horas 114
 - restablecer vistas 118
 - resumen de acciones 120
 - resumen de procesamiento de campos 119
 - selección de características 116
 - tabla de campos 121
 - vista de modelo 118
- preparación de datos geoespaciales
 - Reproyectar, nodo 186

- Primer cuartil
 - agregación de series temporales 184, 185
- procesamiento paralelo
 - fusión 91
 - nodo Agregar 79
 - ordenación 84
- profundidad de lista 11
- propiedades
 - nodo 380
- propiedades de los nodos 380
- Prueba t
 - muestras independientes 320
 - muestras relacionadas 321
 - nodo Medias 320, 321, 322
- publicar en web 297
- Punto 129
- puntos de corte
 - nodo Intervalos 155
- puntuación
 - opciones del diagrama de evaluación 262
- puntuaciones ajustadas de propensión
 - equilibrado de datos 78
- puntuaciones de propensión
 - equilibrado de datos 78
- Python
 - scripts de carga masiva 347, 348

Q

- Query Banding
 - Teradata 22

R

- rango
 - resultado de estadísticos 319
- rangos 127
 - valores perdidos 132
- rangos de casillas
 - archivos de Excel 42
- rangos enteros 133
- rangos fraccionales 159
- rangos reales 133
- recodificación 151, 155
- recodificación automática 151
- reconocimiento automático de fecha 26, 29
- reconocimiento de fecha 26, 29
- recuentos
 - nodo Intervalos 157
 - resultado de estadísticos 319
- reestructuración de los datos 167
- regiones en gráficos 273
- registro
 - etiquetas 136
 - longitud 29
 - recuentos 79
- registros
 - fusión 84
 - transposición 169
- registros compuestos 95
 - valores personalizados 97
- registros concatenados 92
- registros exclusivos 93

- registros incompletos 86
- regla de negocio
 - opciones del diagrama de evaluación 262
- regresión de mínimos cuadrados ponderados localmente
 - nodo Gráfico 229
- relleno de datos de series temporales 172
- rendimiento
 - datos de muestreo 73
 - fusión 91
 - nodo Agregar 79
 - nodos Derivar 161
 - nodos Intervalos 161
 - ordenación 84
- Repositorio de IBM SPSS Collaboration and Deployment Services
 - conexión con 65
 - uso como ubicación de plantillas de visualización, hojas de estilo y mapas 218
- Reproyección, nodo 185
- reproyección de datos de mapa 185
- Reproyectar, nodo 186
- residuos
 - nodo Matriz 303
- restricciones en los datos geoespaciales 129
- resultado de gráfica 333
- resultado de la matriz
 - almacenamiento como texto 300
- resultado en formato HTML
 - nodo Informe 323
 - ver en explorador 298
- resultado en formato XML
 - nodo Informe 323
- resultado tabular
 - reordenación de columnas 299
 - selección de casillas 299
- resultados 331, 333
 - almacenamiento 296
 - exportación 298
 - generación de nodos nuevos desde 296
 - HTML 298
 - impresión 296
- ROI
 - gráficos 256, 263
- roles
 - especificación de campos 136
- roles de modelado
 - especificación de campos 136
- rotación de gráficos 3D 190

S

- SAS
 - configuración de opciones de importación 41
- scripts
 - Supernodos 381
- selección de filas (casos) 72
- servicio de mapas
 - Nodo de origen geoespacial 64
 - Nodo Origen geoespacial 64
- servidor ESRI 64

- significación
 - fuerza de correlación 318
- símbolo de agrupación
 - formatos de presentación de los números 138
- símbolo decimal 26
 - formatos de presentación de los números 138
 - nodo de exportación Archivo sin formato 355
- Sintaxis de XPath 43
- sistema de coordenadas geográficas 185
- sistema de coordenadas proyectadas 185
- sistemas de coordenadas
 - transformación 287
- smoother
 - nodo Gráfico 229
- sobrescritura de tablas de base de datos 340
- SPLOM 198
 - ejemplo 213, 215
- suavizado exponencial
 - criterios en modelos de series temporales 100
- suavizamiento LOESS
 - nodo Gráfico 229
- suavizamiento LOESS Consulte suavizamiento LOESS
 - nodo Gráfico 229
- suma
 - nodo Val. globales 325
 - resultado de estadísticos 319
- Supernodos 375
 - acercamiento 378
 - almacenamiento 382
 - anidación 377
 - bloqueo 377
 - carga 382
 - configuración de los parámetros 379
 - creación 376
 - creación de cachés para 381
 - desbloqueo 377
 - edición 378
 - protección por contraseña 377, 378
 - scripts 381
 - supernodos de origen 375
 - supernodos de proceso 375
 - supernodos terminales 376
 - tipos de 375
 - uso de comentarios con 378
- superposición de gráfico de colores 188
- superposición de gráfico de formas 188
- superposición de gráfico de paneles 188
- superposición de gráfico de tamaño 188
- superposición de gráficos 188
- sustitución de valores de campos 149

T

- tablas
 - almacenamiento como texto 300
 - almacenamiento de resultados 300
 - uniones 84
- tablero
 - tipos de gráfico 198
- tabulación cruzada
 - nodo Matriz 302, 303

- tamaño de confirmación 347
- TeamTest
 - modelos ARIMA 101
- Teradata
 - Query Banding 22
- tercer cuartil
 - agregación de series temporales 184
- Tercer cuartil
 - agregación de series temporales 185
- texto
 - datos 25, 29
 - delimitado 25
- tipo 9
- tipo de almacenamiento de lista 28
- tipo de colección 135
- tipo de conjunto 127
- tipo de marca 127, 134
- tipo de uso 9, 127
- tipo geoespacial 135
- tipos de almacenamiento
 - lista 28
- tipos de campos
 - en visualizaciones 194
- tipos de datos 29, 109, 127
 - instanciación 130
- tipos de etiquetas
 - nodo de origen Recopilación de datos de IBM SPSS 34
- tipos de gráfico
 - tablero 198
- tipos de variables
 - en visualizaciones 194
- tramos en gráficos 270
- transformación de diferenciación
 - modelos ARIMA 101
- transformación de diferenciación estacional
 - modelos ARIMA 101
- transformación de raíz cuadrada
 - Modelizador de series temporales 102
- transformación logarítmica
 - Modelizador de series temporales 102
- transformación logarítmica natural
 - Modelizador de series temporales 102
- transformaciones
 - reclasificar 151, 155
 - recodificación 151, 155
- transparencia en gráficos 188
- transposición de datos 169
- tratamiento de los elementos vacíos 132
 - nodo Intervalos 156
 - valores de relleno 149
- truncado de nombres de campos 140, 141

U

- umbrales
 - visualización de umbrales de intervalos 161
- unión de conjuntos de datos 92
- unión exterior 84
- unión interior 84
- uniones 84, 86

- uniones (*continuación*)
 - exteriores parciales 87
- uniones parciales 84, 87
- Utilidad de conversión de mapas 219, 221

V

- vacíos
 - en tablas Matriz 302
- valor clave para la agregación 79
- valor de cuartil para la agregación 79, 81
- valor de la media de los registros 78
- valor de la media para la agregación 79
- valor de la mediana para la agregación 79, 81
- valor de semilla
 - muestreo y registros 165
- valor de semilla aleatoria
 - registros de muestreo 165
- valor de varianza para la agregación 79
- valor máximo para la agregación 79
- valor mínimo para la agregación 79
- valor p
 - importancia 321
- valores
 - especificación 132
 - etiquetas de valor y de campo 132
 - lectura 131
- valores atípicos
 - en modelos de series temporales 103
 - modelos ARIMA 103
- valores atípicos aditivos
 - Modelizador de series temporales 103
- valores atípicos aditivos estacionales
 - Modelizador de series temporales 103
- valores atípicos de cambio de nivel
 - Modelizador de series temporales 103
- valores atípicos de tendencia local
 - Modelizador de series temporales 103
- valores atípicos innovadores
 - Modelizador de series temporales 103
- valores atípicos transitorios
 - Modelizador de series temporales 103
- valores ausentes
 - en nodos Agregar 78
- valores de agrupación 240
- valores de colección para la fórmula de derivación 145
- valores de selección 270, 273, 275
- valores definidos por el usuario como perdidos
 - en tablas Matriz 302
- valores esperados
 - nodo Matriz 303
- valores falsos 134
- valores geoespaciales para la fórmula de derivación 145
- valores globales 324
- valores no definidos 86

- valores nulos
 - en tablas Matriz 302
- valores para la fórmula de derivación 145
- valores perdidos 109, 132, 135
 - en tablas Matriz 302
- valores perdidos del sistema
 - en tablas Matriz 302
- valores predefinidos, conexión de base de datos 20
- valores sumados 79
- valores vacíos
 - en tablas Matriz 302
- valores verdaderos 134
- Variables de códigos
 - nodo de origen Recopilación de datos de IBM SPSS 31
- variables de series temporales 182
- Variables del sistema
 - nodo de origen Recopilación de datos de IBM SPSS 31
- Variables SourceFile
 - nodo de origen Recopilación de datos de IBM SPSS 31
- varianza
 - resultado de estadísticos 319
- varios campos
 - selección 145
- varita mágica en gráficos 275
- ver
 - resultado en formato HTML en un explorador 298
- vista de modelo
 - en preparación automática de datos 118
- vistas de datos analíticos 63
- visualización
 - gráficos y diagramas 187
- visualización de mapas
 - ejemplo 214
- visualizaciones
 - categorías 285
 - colores y tramas 281
 - copia 291
 - edición 279
 - ejes 284
 - escalas 284
 - estilos de guión 281
 - forma de puntos 282
 - formatos de numeración 283
 - márgenes 283
 - modo de edición 279
 - paneles 285, 287
 - posición de la leyenda 290
 - relación de aspecto de puntos 282
 - relleno 283
 - rotación de puntos 282
 - texto 281
 - transformación de los sistemas de coordenadas 287
 - transparencia 281
 - transponer 285, 287
- visualizaciones de mapas
 - creación 205

Z

- zoom 378



Impreso en España