

IBM SPSS Categories 20

Jacqueline J. Meulman

Willem J. Heiser



Nota: Antes de utilizar esta información y el producto que admite, lea la información general en Avisos el p. 307.

Esta edición se aplica a IBM® SPSS® Statistics 20 y a todas las versiones y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.

Materiales bajo licencia: Propiedad de IBM

© Copyright IBM Corporation 1989, 2011.

Derechos restringidos para los usuarios del gobierno de Estados Unidos: Uso, duplicación o revelación restringidos por GSA ADP Schedule Contract con IBM Corp.

Prefacio

IBM® SPSS® Statistics es un sistema global para el análisis de datos. El módulo adicional opcional Categories proporciona las técnicas de análisis adicionales que se describen en este manual. El módulo adicional Categories se debe utilizar con el sistema básico de SPSS Statistics y está completamente integrado en dicho sistema.

Acerca de IBM Business Analytics

IBM Business Analytics proporciona información completa, coherente y precisa en la que confían para mejorar el rendimiento de su negocio quienes toman las decisiones. Un conjunto de documentos que incluye [inteligencia comercial](#), [análisis predictivo](#), [rendimiento financiero y gestión de estrategias](#) y [aplicaciones analíticas](#) proporciona ideas claras e inmediatas del rendimiento actual y la habilidad para predecir resultados futuros. Combinado con numerosas soluciones para empresas, prácticas de eficacia demostrada y servicios profesionales, las organizaciones de cualquier tamaño pueden conseguir la más alta productividad, automatizar decisiones con seguridad y obtener mejores resultados.

Como parte de estos documentos, IBM SPSS Predictive Analytics ayuda a las organizaciones a predecir situaciones futuras y a actuar de forma proactiva con esa información para mejorar sus resultados. Clientes comerciales, gubernamentales y académicos de todo el mundo confían en la tecnología IBM SPSS como mejora competitiva para atraer, conservar y aumentar la clientela reduciendo el fraude y los riesgos. Al incorporar IBM SPSS a sus operaciones diarias, las organizaciones se convierten en empresas predictivas – capaces de dirigir y automatizar decisiones para conseguir los objetivos de la empresa y lograr una mejora competitiva y ostensible. Para obtener más información o contactar con un representante, visite <http://www.ibm.com/spss>.

Asistencia técnica

El servicio de asistencia técnica está a disposición de todos los clientes de mantenimiento. Los clientes podrán ponerse en contacto con este servicio de asistencia técnica si desean recibir ayuda sobre la utilización de los productos de IBM Corp. o sobre la instalación en alguno de los entornos de hardware admitidos. Para contactar con el servicio de asistencia técnica, visite el sitio Web de IBM Corp. en <http://www.ibm.com/support>. Tenga a mano su identificación, la de su organización y su contrato de asistencia cuando solicite ayuda.

Asistencia técnica para estudiantes:

Si usted es un estudiante que utiliza una versión académica o para estudiantes de cualquier producto de software IBM SPSS, consulte nuestras páginas especiales en línea de [Soluciones educativas](#) (<http://www.ibm.com/spss/rd/students/>) para estudiantes. Si usted es estudiante y utiliza una copia proporcionada por la universidad del software IBM SPSS, póngase en contacto con el coordinador del producto IBM SPSS en su universidad.

Servicio de atención al cliente

Si tiene preguntas referentes a su envío o cuenta, póngase en contacto con su oficina local. Recuerde tener preparado su número de serie para identificarse.

Cursos de preparación

IBM Corp. ofrece cursos de preparación, tanto públicos como in situ. Todos los cursos incluyen talleres prácticos. Los cursos tendrán lugar periódicamente en las principales ciudades. Si desea más información sobre estos seminarios, visite <http://www.ibm.com/software/analytics/spss/training>.

Publicaciones adicionales

Los documentos *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* y *SPSS Statistics: Advanced Statistical Procedures Companion*, escritos por Marija Norušis y publicados por Prentice Hall, están disponibles y se recomiendan como material adicional. Estas publicaciones cubren los procedimientos estadísticos del módulo SPSS Statistics Base, el módulo Advanced Statistics y el módulo Regression. Tanto si da sus primeros pasos en el análisis de datos como si ya está preparado para las aplicaciones más avanzadas, estos libros le ayudarán a aprovechar al máximo las funciones ofrecidas por IBM® SPSS® Statistics. Si desea información adicional sobre el contenido de la publicación o muestras de capítulos, consulte el sitio web de la autora: <http://www.norusis.com>

Agradecimientos

Los procedimientos de escalamiento óptimo y su implementación en IBM® SPSS® Statistics han sido desarrollados por el grupo Data Theory Scaling System Group (DTSS), formado por miembros de los departamentos de educación y psicología de la facultad de ciencias sociales y del comportamiento de la Universidad de Leiden.

Willem Heiser, Jacqueline Meulman, Gerda van den Berg y Patrick Groenen colaboraron en los procedimientos originales de 1990. Jacqueline Meulman y Peter Neufeglise participaron en el desarrollo de procedimientos de regresión categórica, análisis de correspondencias, análisis de componentes principales categórico y escalamiento multidimensional. Además, Anita van der Kooij contribuyó especialmente a CATREG, CORRESPONDENCE y CATPCA. Willem Heiser, Jacques Commandeur, Frank Busing, Gerda van den Berg y Patrick Groenen participaron en el desarrollo del procedimiento PROXSCAL. Frank Busing, Willem Heiser, Patrick Groenen y Peter Neufeglise participaron en el desarrollo del procedimiento PREFSCAL.

Contenido

Parte I: Manual del usuario

1 **Introducción a los procedimientos de escalamiento óptimo de datos categóricos** **1**

¿Qué es el escalamiento óptimo?	1
¿Por qué utilizar el escalamiento óptimo?	1
Nivel de escalamiento óptimo y nivel de medida	2
Selección del nivel de escalamiento óptimo	3
Gráficos de transformación	3
Códigos de la categoría	4
¿Qué procedimiento es el mejor para la aplicación?	6
Regresión categórica	7
Análisis de componentes principales categórico	8
Análisis de correlación canónica no lineal	9
Análisis de correspondencias	10
Análisis de correspondencias múltiple	11
Multidimensional Scaling	12
Desplegamiento multidimensional	13
Relación de aspecto en gráficos de escalamiento óptimo	13
Lecturas recomendadas	13

2 **Regresión categórica (CATREG)** **15**

Para definir la escala en regresión categórica	16
Discretización de regresión categórica	18
Valores perdidos de regresión categórica	19
Opciones de regresión categórica	20
Regularización de regresión categórica	21
Resultados de la regresión categórica	22
Regresión categórica: Guardar	24
Gráficos de transformación en regresión categórica	25
Funciones adicionales del comando CATREG	26

3 *Análisis de componentes principales categórico (CATPCA)* 27

Definir escala y ponderación en CATPCA	29
Análisis de componentes principales categórico: Discretización	31
Análisis de componentes principales categórico: Valores perdidos	32
Análisis de componentes principales categórico: Opciones	33
Análisis de componentes principales categórico: Resultados	35
Análisis de componentes principales categórico: Guardar	37
Análisis de componentes principales categórico: Gráficos de objetos	38
Análisis de componentes principales categórico: Gráficos de categorías	39
Análisis de componentes principales categórico: gráfico de saturaciones	40
Funciones adicionales del comando CATPCA	41

4 *Análisis de correlación canónica no lineal (OVERALS)* 42

Definir rango y escala	45
Definir rango	45
Análisis de correlación canónica no lineal: Opciones	46
Funciones adicionales del comando OVERALS	47

5 *Análisis de correspondencias* 49

Definir rango de filas en Análisis de correspondencias	50
Definir rango de columnas en Análisis de correspondencias	51
Análisis de correspondencias: Modelo	52
Análisis de correspondencias: Estadísticos	54
Análisis de correspondencias: Gráficos	55
Funciones adicionales del comando CORRESPONDENCE	57

6 *Análisis de correspondencias múltiple* 58

Definir ponderación de la variable en el análisis de correspondencias múltiple	60
Discretización en el análisis de correspondencias múltiple	60
Valores perdidos en el análisis de correspondencias múltiple	61
Opciones de análisis de correspondencias múltiple	63

Resultados del análisis de correspondencias múltiple	65
Análisis de correspondencias múltiple: Guardar	66
Análisis de correspondencias múltiple: Gráficos de objetos	67
Análisis de correspondencias múltiple: Gráficos de variables.	67
Funciones adicionales del comando MULTIPLE CORRESPONDENCE.	68
7 Escalamiento multidimensional (PROXSCAL)	70
Proximidades en matrices a través de columnas.	72
Proximidades en columnas.	73
Proximidades en una columna	74
Crear proximidades de los datos.	75
Crear la medida a partir de los datos.	76
Definir un modelo de escalamiento multidimensional	77
Escalamiento multidimensional: Restricciones	78
Escalamiento multidimensional: Opciones	79
Escalamiento multidimensional: Gráficos, Versión 1	81
Escalamiento multidimensional: Gráficos, Versión 2	82
Escalamiento multidimensional: Resultados	82
Funciones adicionales del comando PROXSCAL	84
8 Desplegamiento multidimensional (PREFSCAL)	85
Definir un modelo de desplegado multidimensional	86
Restricciones del desplegado multidimensional	88
Opciones de desplegado multidimensional.	89
Gráficos de desplegado multidimensional	91
Resultados del desplegado multidimensional	92
Funciones adicionales del comando PREFSCAL	94

Parte II: Ejemplos

9 Regresión categórica 96

Ejemplo: Datos de limpiadores de alfombras	96
Análisis de regresión lineal ordinario	97
Análisis de regresión categórica	103
Ejemplo: Datos sobre el ozono	115
Discretización de las variables	116
Selección del tipo de transformación	116
Optimalidad de las cuantificaciones	129
Efectos de las transformaciones	131
Lecturas recomendadas	139

10 Análisis de componentes principales categórico 141

Ejemplo: Examen de las interrelaciones de los sistemas sociales	141
Ejecución del análisis	142
Número de dimensiones	146
Cuantificaciones	147
Puntuaciones de los objetos	149
Saturaciones en componentes	150
Dimensiones adicionales	152
Ejemplo: sintomatología de los trastornos de la alimentación	154
Ejecución del análisis	155
Gráficos de transformación	167
Resumen del modelo	169
Saturaciones en componentes	170
Puntuaciones de los objetos	171
Examen de la estructura del transcurso de la enfermedad	173
Lecturas recomendadas	189

11 Análisis de correlación canónica no lineal 191

Ejemplo: Un análisis de los resultados de una encuesta	191
Examen de los datos	192
Explicación de la similitud entre los conjuntos	198
Saturaciones en componentes	202

Gráficos de transformación	203
Coordenadas de categoría simples frente a categorías múltiples	205
Centroides y centroides proyectados	206
Un análisis alternativo	209
Sugerencias generales	214
Lecturas recomendadas	215

12 *Análisis de correspondencias* 216

Correspondence analysis	216
Normalización	217
Ejemplo: Percepciones de marcas de café	218
Ejecución del análisis	218
Dimensionalidad	222
Contribuciones	223
Gráficos	224
Normalización simétrica	225
Lecturas recomendadas	227

13 *Análisis de correspondencias múltiple* 228

Ejemplo: Características de la ferretería	228
Ejecución del análisis	228
Resumen del modelo	231
Puntuaciones de los objetos	232
Medidas de discriminación	233
Cuantificaciones de categorías	234
Una observación más detallada de las puntuaciones de los objetos	236
Omisión de valores atípicos	239
Lecturas recomendadas	242

14 *Multidimensional Scaling* 244

Ejemplo: Un examen de los términos de parentesco	244
Elección del número de dimensiones	245
Una solución tridimensional	251

Una solución tridimensional con transformaciones que no son por defecto	257
Discusión	260
Lecturas recomendadas	260

15 despliegamiento multidimensional

262

Ejemplo: Preferencias del elemento de desayuno	262
Creación de una solución degenerada	262
Medidas	265
Espacio común	266
Ejecución de un análisis no degenerado	267
Medidas	268
Espacio común	269
Ejemplo: Despliegamiento triple de las preferencias de los elementos del desayuno	269
Ejecución del análisis	270
Medidas	274
Espacio común	275
Espacios individuales	276
Uso de una configuración inicial diferente	279
Medidas	281
Espacio común	282
Espacios individuales	283
Ejemplo: Examen de lo apropiado de conducta y situación	285
Ejecución del análisis	285
Medidas	290
Espacio común	291
Transformaciones de proximidades	292
Cambio de la transformación de proximidades (ordinal)	292
Medidas	294
Espacio común	295
Transformaciones de proximidades	296
Lecturas recomendadas	296

Apéndices

A Archivos muestrales ***297***

B Avisos ***307***

Bibliografía ***310***

Índice ***315***

Parte I:
Manual del usuario

Introducción a los procedimientos de escalamiento óptimo de datos categóricos

Los procedimientos de categorías utilizan el escalamiento óptimo para analizar datos que son difíciles o imposibles de analizar mediante los procedimientos estadísticos estándar. Este capítulo describe lo que hace cada procedimiento, las situaciones en las que cada procedimiento es más apropiado, las relaciones entre los procedimientos y las relaciones de estos procedimientos con sus procedimientos estadísticos análogos estándar.

Nota: estos procedimientos y su implementación en IBM® SPSS® Statistics han sido desarrollados por el grupo Data Theory Scaling System Group (DTSS), formado por miembros de los departamentos de educación y psicología de la facultad de ciencias sociales y del comportamiento de la Universidad de Leiden.

¿Qué es el escalamiento óptimo?

La idea que subyace tras el escalamiento óptimo es asignar cuantificaciones numéricas a las categorías de cada variable, lo que permite utilizar los procedimientos estándar para obtener una solución con las variables cuantificadas.

Los valores de escala óptimos se asignan a las categorías de cada variable de acuerdo con el criterio de optimización del procedimiento que se esté utilizando. A diferencia de las etiquetas originales de las variables nominales u ordinales del análisis, estos valores de escala tienen propiedades métricas.

En la mayoría de los procedimientos de categorías, la cuantificación óptima de cada variable escalada se obtiene mediante un método iterativo denominado **mínimos cuadrados alternantes** en el que, después de que se utilicen las cuantificaciones actuales para encontrar una solución, las cuantificaciones se actualizan utilizando dicha solución. A continuación, se utilizan las cuantificaciones actualizadas para buscar una nueva solución, que a su vez se utiliza para actualizar las cuantificaciones y así sucesivamente, hasta que se alcanza algún criterio que indica al proceso que finalice.

¿Por qué utilizar el escalamiento óptimo?

Los datos categóricos se utilizan con frecuencia en los estudios de mercado, los estudios de encuestas y la investigación en las ciencias sociales y del comportamiento. De hecho, muchos investigadores trabajan casi exclusivamente con datos categóricos.

Aunque existen adaptaciones de la mayoría de los modelos estándar que permiten analizar específicamente datos categóricos, con frecuencia no funcionan bien con conjuntos de datos con las siguientes características:

- Observaciones insuficientes
- Demasiadas variables
- Demasiados valores por cada variable

Mediante la cuantificación de categorías, las técnicas de escalamiento óptimo evitan los problemas de estas situaciones. Además, son muy útiles incluso cuando es apropiado utilizar técnicas especializadas.

En vez de interpretar las estimaciones de los parámetros, la interpretación de los resultados del escalamiento óptimo muchas veces se basa en representaciones gráficas. Las técnicas de escalamiento óptimo ofrecen excelentes análisis exploratorios, que complementan también a otros modelos de IBM® SPSS® Statistics. Mediante el acotamiento del objetivo de la investigación, la visualización de los datos mediante el escalamiento óptimo puede formar la base de un análisis que se centre en la interpretación de los parámetros del modelo.

Nivel de escalamiento óptimo y nivel de medida

Este concepto puede resultar muy confuso cuando se utilizan por primera vez los procedimientos de categorías. Al especificar el nivel, no se especifica el nivel al que se *miden* las variables, sino el nivel al que se *escalan*. La idea es que las variables que se van a cuantificar pueden tener relaciones no lineales independientemente de cómo se midan.

En categorías, hay tres niveles básicos de medidas:

- El nivel **nominal** implica que los valores de una variable representan categorías desordenadas. Algunos ejemplos de variables que pueden ser nominales serían la región, el área del código postal, la confesión religiosa y las categorías con varias opciones.
- El nivel **ordinal** implica que los valores de una variable representan categorías ordenadas. Entre los ejemplos se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- El nivel **numérico** implica que los valores de una variable representan categorías ordenadas con una métrica significativa, de modo que las comparaciones de distancia entre categorías son adecuadas. Entre los ejemplos se incluyen la edad en años y los ingresos en dólares.

Por ejemplo, suponga que las variables *región*, *trabajo* y *edad* se codifican como se muestra en la siguiente tabla.

Tabla 1-1
Esquema de codificación para región, trabajo y edad

Código de región	Valor de región	Código de trabajo	Valor de trabajo	Edad
1	Norte	1	trabajador en prácticas	20
2	Sur	2	vendedor	22
3	Este	3	administrador	25
4	Oeste			27

Los valores mostrados representan las categorías de cada variable. *Región* sería una variable nominal. Hay cuatro categorías de *región*, sin ningún orden intrínseco. Los valores del 1 al 4 sencillamente representan las cuatro categorías; el esquema de codificación es completamente arbitrario. *Trabajo*, por otra parte, se podría considerar como variable ordinal. Las categorías originales forman una progresión desde trabajador en prácticas hasta administrador. Los códigos mayores representan un trabajo superior en la escala corporativa. Sin embargo, sólo se conoce la información acerca del orden, no se puede decir nada acerca de la distancia existente entre categorías adyacentes. Por el contrario, *edad* se podría considerar como una variable numérica. En el caso de *edad*, las distancias entre los valores son intrínsecamente significativas. La distancia entre 20 y 22 es la misma que la que hay entre 25 y 27, mientras que la distancia entre 22 y 25 es superior a las dos anteriores.

Selección del nivel de escalamiento óptimo

Es importante comprender que no hay propiedades intrínsecas de una variable que predefinan automáticamente el nivel de escalamiento óptimo que se debería definir para ella. Puede explorar los datos de cualquier manera lógica que facilite la interpretación. Mediante el análisis de una variable de nivel numérico a nivel ordinal, por ejemplo, el uso de una transformación no lineal puede permitir una solución con menos dimensiones.

Los dos ejemplos siguientes ilustran cómo el nivel de medida “obvio” no siempre es el mejor nivel de escalamiento óptimo. Supongamos que una variable ordena objetos en dos grupos de edad. Aunque la edad se puede escalar como una variable numérica, puede ocurrir que en las personas menores de 25 años, la seguridad tenga una relación positiva con la edad, mientras que en las personas mayores de 60 años, la seguridad tenga una relación negativa con la edad. En este caso, puede ser mejor tratar la edad como una variable nominal.

Tomemos otro ejemplo, una variable que ordena a las personas por preferencia política parece ser básicamente nominal. Sin embargo, si se ordenan los partidos desde la izquierda política hasta la derecha política, puede que le interese que la cuantificación de los partidos respete este orden mediante un nivel ordinal de análisis.

Aunque no haya propiedades predefinidas de una variable que la coloquen exclusivamente en un nivel o en otro, existen algunas normas generales que pueden ayudar al usuario inexperto. Con la cuantificación nominal simple, habitualmente no se conoce el orden de las categorías pero se desea que el análisis imponga una. Si se conoce el orden de las categorías, debería intentarse la cuantificación ordinal. Si las categorías no se pueden ordenar, puede intentar la cuantificación nominal múltiple.

Gráficos de transformación

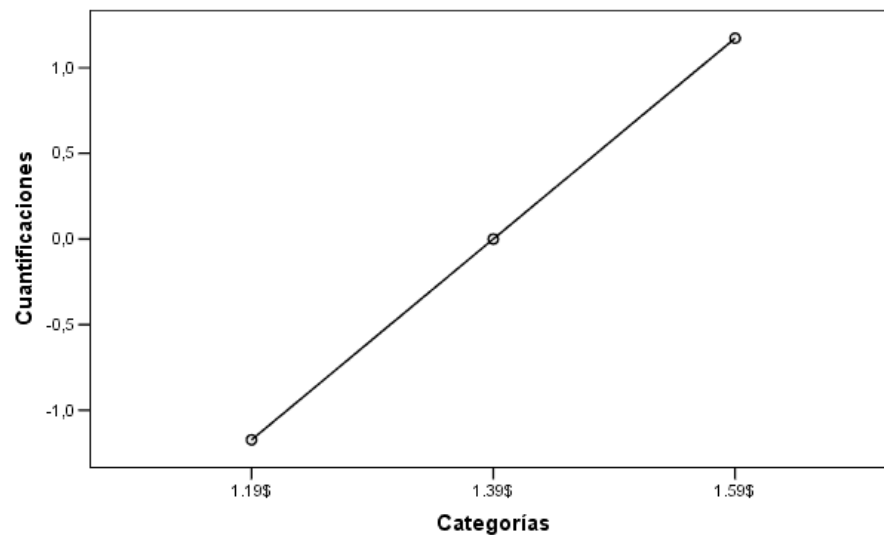
Los diferentes niveles en los que se puede escalar cada variable imponen diferentes restricciones sobre las cuantificaciones. Los gráficos de transformación ilustran la relación entre las cuantificaciones y las categorías originales que resultan del nivel de escalamiento óptimo seleccionado. Por ejemplo, se genera un gráfico de transformación lineal cuando una variable se trata como numérica. Las variables tratadas como ordinales generan gráficos de transformación no decreciente. Los gráficos de transformación de variables tratadas nominalmente que tienen forma de U (o la inversa) muestran una relación cuadrática. Las variables nominales también pueden

generar gráficos de transformación sin tendencias aparentes mediante el cambio completo del orden de las categorías. La siguiente figura muestra un gráfico de transformación de ejemplo.

Los gráficos de transformación son especialmente adecuados para determinar si funciona bien el nivel de escalamiento óptimo seleccionado. Si varias categorías reciben cuantificaciones similares, la agrupación de estas categorías en una categoría puede estar justificada. Otra posibilidad, si una variable tratada como nominal recibe cuantificaciones que muestran una tendencia creciente, una transformación ordinal puede generar un ajuste similar. Si esta tendencia es lineal, el tratamiento numérico puede resultar adecuado. Sin embargo, si la agrupación de categorías o el cambio de los niveles de escalamiento están justificados, el análisis no cambiará de manera significativa.

Figura 1-1

Transformación del gráfico del precio (numérico)



Códigos de la categoría

Se debe tener cierto cuidado al codificar las variables categóricas ya que algunos esquemas de codificación pueden generar resultados no deseados o análisis incompletos. En la siguiente tabla se muestran algunos posibles esquemas de codificación para el *trabajo*.

Tabla 1-2

Esquemas de codificación alternativos para el trabajo

Categoría	A	B	C	D
trabajador en prácticas	1	1	5	1
vendedor	2	2	6	5
administrador	3	7	7	3

Algunos procedimientos de categorías requieren que se defina el rango de cada variable. Todo valor fuera de este rango se tratará como un valor perdido. El valor de categoría mínima es siempre 1. El valor de categoría máxima lo indica el usuario. Este valor no es el *número* de categorías de una variable, sino el valor *mayor* de la categoría. Por ejemplo, en la tabla, el esquema A tiene un valor de categoría máxima de 3 y el esquema B tiene un valor de categoría máxima de 7, aunque ambos esquemas codifican las mismas tres categorías.

El rango de la variable determina las categorías que se omitirán del análisis. Todas las categorías con códigos fuera del rango definido se omitirán del análisis. Se trata de un método sencillo de omitir categorías, pero puede generar análisis no deseados. Una categoría máxima incorrectamente definida puede omitir categorías *válidas* del análisis. Por ejemplo, para el esquema B, si se define que el valor de categoría máxima sea 3 indica que *trabajo* tiene categorías codificadas de 1 a 3; la categoría *administrador* se tratará como valor perdido. Como no hay ninguna categoría que se haya codificado realmente como 3, la tercera categoría del análisis no contendrá ningún caso. Si desea omitir todas las categorías de administrador, este análisis sería adecuado. Sin embargo, si desea incluir a los administradores, la categoría máxima deberá definirse como 7 y los valores perdidos deberán codificarse con valor por encima de 7 o por debajo de 1.

Para las variables tratadas como nominales u ordinales, el rango de las categorías no afecta a los resultados. Para las variables nominales, es importante sólo la etiqueta y no el valor asociado con dicha etiqueta. Para las variables ordinales, se conserva el orden de las categorías en las cuantificaciones; los propios valores de las categorías no son importantes. Todos los esquemas de codificación que produzcan el mismo orden de categorías tendrán resultados idénticos. Por ejemplo, los tres primeros esquemas de la tabla son funcionalmente equivalentes si *trabajo* se analiza a nivel ordinal. El orden de las categorías es idéntico en estos esquemas. El esquema D, por otra parte, invierte las categorías segunda y tercera y generará resultados diferentes de los de los otros esquemas.

Aunque muchos esquemas de codificación de una variable son funcionalmente equivalentes, se prefieren aquellos esquemas con pequeñas diferencias entre los códigos, ya que los códigos afectan a la cantidad de resultados que genera un procedimiento. Todas las categorías codificadas con valores entre 1 y el máximo definido por el usuario son válidas. Si alguna de estas categorías está vacía, las cuantificaciones correspondientes serán valores perdidos por el sistema o 0, dependiendo del procedimiento. Aunque ninguna de estas asignaciones afecta a los análisis, se generarán los resultados correspondientes a estas categorías. Por tanto, para el esquema B, *trabajo* tiene cuatro categorías que reciben valores perdidos por el sistema. Para el esquema C, también hay cuatro categorías que reciben indicadores de valores perdidos por el sistema. Por el contrario, para el esquema A no hay ninguna cuantificación de valores perdidos por el sistema. El uso de números enteros consecutivos como códigos de las variables tratadas como nominales u ordinales genera una cantidad mucho menor de resultados sin que éstos se vean afectados.

Los esquemas de codificación de las variables tratadas como numéricas son mucho más restringidos que cuando se trata de una variable ordinal. Para estas variables, las diferencias entre las categorías consecutivas son importantes. La siguiente tabla muestra tres esquemas de codificación para *edad*.

Tabla 1-3
Esquemas de codificación alternativos para la edad

Categoría	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

Toda recodificación de las variables numéricas debe conservar las diferencias entre las categorías. El uso de los valores originales es un método para asegurarse de la conservación de las diferencias. Sin embargo, esto puede generar que muchas categorías tengan indicadores de valores perdidos

por el sistema. Por ejemplo, el esquema A utiliza los valores originales observados. Para todos los procedimientos de categorías salvo para el análisis de correspondencias, el valor de categoría máxima es 27 y el valor de categoría mínimo se establece en 1. Las primeras 19 categorías están vacías y reciben indicadores de valores perdidos por el sistema. Los resultados pueden volverse rápidamente bastante engorrosos si la categoría máxima es mucho mayor que 1 y hay muchas categorías vacías entre 1 y el máximo.

Para reducir la cantidad de resultados, se puede aplicar una recodificación. Sin embargo, en el caso numérico, la función Recodificación automática no se debe utilizar. La codificación en enteros consecutivos generará diferencias de 1 entre todas las categorías consecutivas y, por tanto, todas las cuantificaciones tendrán idéntica separación. Las características métricas que se consideraban importantes al tratar una variable como numérica desaparecen al recodificarla mediante enteros consecutivos. Por ejemplo, el esquema C de la tabla corresponde a la recodificación automática de *edad*. La diferencia entre las categorías 22 y 25 ha cambiado de tres a uno y las cuantificaciones reflejarán esta última diferencia.

Un esquema de recodificación alternativa que conserva las diferencias entre las categorías consiste en restar el valor de la categoría más pequeña a cada categoría y sumar 1 a cada diferencia. El esquema B surge de esta transformación. El valor de categoría más pequeño, 20, se ha restado a cada categoría y se sumado 1 a cada resultado. Los códigos transformados tienen un mínimo de 1 y todas las diferencias son idénticas a los datos originales. El valor de categoría máxima es ahora 8 y se eliminan todas las cuantificaciones de cero antes de la primera cuantificación que no es cero. Aún así, las cuantificaciones que no son cero correspondientes a cada categoría que resultan del esquema B son idénticas a las cuantificaciones del esquema A.

¿Qué procedimiento es el mejor para la aplicación?

Las técnicas que se incluyen en cuatro de estos procedimientos (Análisis de correspondencias, Análisis de correspondencias múltiple, Análisis de componentes principales categórico y Análisis de correlación canónica no lineal) corresponden al área general del análisis de datos multivariantes conocido como **reducción de dimensiones**. Es decir, las relaciones entre las variables se representan en un número pequeño de dimensiones (digamos, dos o tres) con tanta frecuencia como sea posible. Esto permite describir las estructuras o patrones de las relaciones que serían extremadamente difícil de descubrir con su riqueza y complejidad original. En las aplicaciones de investigación de mercados, estas técnicas pueden ser una forma de **asignación perceptual**. Una importante ventaja de estos procedimientos es que se adaptan a los datos con diferentes niveles de escalamiento óptimo.

La regresión categórica describe la relación entre una variable de respuesta categórica y una combinación de predictores categóricos. La influencia de cada predictor sobre la variable de respuesta la describe la correspondiente ponderación de regresión. Como en los demás procedimientos, los datos se pueden analizar con diferentes niveles de escalamiento óptimo.

El escalamiento multidimensional y el desplegamiento multidimensional describe las relaciones entre objetos en un espacio conceptual de pocas dimensiones utilizando las proximidades entre los objetos.

A continuación, se indican unas breves normas generales para cada uno de los procedimientos:

- Utilice la regresión categórica para pronosticar los valores de una variable dependiente categórica a partir de una combinación de variables independientes categóricas.

- Utilice el análisis de componentes principales categórico para tener en cuenta los patrones de variación de un único conjunto de variables con varios tipos de niveles de escalamiento óptimos.
- Utilice el análisis de correlación canónica no lineal para evaluar el grado de correlación entre dos o más conjuntos de variables de varios tipos de niveles de escalamiento óptimo.
- Utilice el análisis de correspondencias para analizar tablas de contingencia de doble clasificación o datos que se puedan expresar como una tabla de doble clasificación, como la preferencia por una u otra marca o datos de opción sociométricos.
- Utilice el análisis de correspondencias múltiple para analizar una matriz de datos multivariantes categóricos cuando no se está dispuesto a suponer que todas las variables se analizan a nivel nominal.
- Utilice el escalamiento multidimensional para analizar datos de proximidad para buscar una representación de un solo conjunto de los objetos en un espacio de pocas dimensiones.
- Utilice el despliegamiento multidimensional para analizar datos de proximidad para buscar una representación de dos conjuntos de los objetos en un espacio de pocas dimensiones.

Regresión categórica

El uso de la regresión categórica es especialmente adecuado cuando el objetivo del análisis es pronosticar una variable dependiente (de respuesta) a partir de un conjunto de variables (predictoras) independientes. Al igual que ocurre con todos los procedimientos de escalamiento óptimo, los valores de escala se asignan a cada categoría de cada variable de manera que estos valores sean óptimos respecto a la regresión. La solución de una regresión categórica maximiza la correlación de los cuadrados entre la respuesta transformada y la combinación ponderada de los predictores transformados.

Relación con otros procedimientos de categorías. La regresión categórica con escalamiento óptimo se puede comparar con el análisis de correlación canónica mediante escalamiento óptimo con dos conjuntos, uno de los cuales sólo contiene la variable dependiente. En esta última técnica, la similitud de los conjuntos se determina mediante la comparación de cada conjunto con una variable desconocida que se encuentra en algún sitio entre todos los conjuntos. En la regresión categórica, la similitud de la respuesta transformada y la combinación lineal de los predictores transformados se evalúa directamente.

Relación con las técnicas estándar. En la regresión lineal estándar, las variables categóricas se pueden recodificar como variables indicadoras o se pueden tratar de la misma manera que las variables de nivel de intervalo. En el primer método, el modelo contiene una pendiente y un punto de corte diferentes para cada combinación de los niveles de las variables categóricas. Esto implica que es necesario interpretar un gran número de parámetros. En el segundo enfoque, sólo se estima un parámetro para cada variable. Sin embargo, la naturaleza arbitraria de las codificaciones de categorías hace imposibles las generalizaciones.

Si algunas de las variables no son continuas, habrá disponibles análisis alternativos. Si la respuesta es continua y los predictores son categóricos, se suele utilizar el análisis de la varianza. Si la respuesta es categórica y los predictores son continuos, puede que sean adecuados el análisis discriminante o la regresión logística. Si la respuesta y los predictores son ambos categóricos, se suelen utilizar modelos loglineales.

La regresión con el escalamiento óptimo ofrece tres niveles de escalamiento para cada variable. Las combinaciones de estos niveles pueden tener en cuenta una gran variedad de relaciones no lineales para las que cualquier método “estándar” no se adapte bien. Por tanto, el escalamiento óptimo ofrece una mayor flexibilidad que los enfoques estándar con una mínima complejidad adicional.

Además, las transformaciones no lineales de los predictores suelen reducir las dependencias entre los predictores. Si compara los autovalores de la matriz de correlaciones de los predictores con los autovalores de la matriz de correlaciones de los predictores escalados óptimamente, este último conjunto normalmente será menos variable que el anterior. Dicho de otra manera, en la regresión categórica, el escalamiento óptimo hace que los autovalores mayores de la matriz de correlaciones de los predictores sean más pequeños y que los autovalores más pequeños sean más grandes.

Análisis de componentes principales categórico

El uso del análisis de componentes principales categórico es especialmente adecuado cuando se desea tener en cuenta los patrones de variación de un único conjunto de variables con varios tipos de niveles de escalamiento óptimos. Esta técnica intenta reducir la dimensionalidad de un conjunto de variables al mismo tiempo que tiene en cuenta toda la variación que sea posible. Se asignan valores de escala a cada categoría de cada variable de manera que estos valores sean óptimos respecto a la solución de componentes principales. Los objetos del análisis reciben puntuaciones de componentes en función de los datos cuantificados. Los gráficos de las puntuaciones de componentes revelan los patrones de los objetos del análisis y pueden revelar objetos poco habituales en los datos. La solución de un análisis de componentes principales categórico maximiza las correlaciones de las puntuaciones de objetos con cada una de las variables cuantificadas para el número de componentes (dimensiones) especificado.

Una aplicación importante de los componentes principales categóricos es el estudio de los datos de preferencias, en el que los encuestados clasifican o evalúan un número de elementos de acuerdo con sus preferencias. En la configuración de datos de IBM® SPSS® Statistics habitual, las filas son individuos, las columnas son medidas de los elementos y las puntuaciones a través de las filas son puntuaciones de preferencias (en una escala de 0 a 10, por ejemplo), lo que hace que los datos sean condicionales respecto a las filas. Para los datos de preferencias, es posible que quiera tratar a los individuos como variables. Mediante el procedimiento Transponer, puede transponer los datos. Los evaluadores se convierten en las variables y todas las variables se declaran como ordinales. No hay ningún problema en utilizar más variables que objetos en CATPCA.

Relación con otros procedimientos de categorías. Si todas las variables se declaran como nominales múltiples, el análisis de componentes principales categórico produce un análisis equivalente a un análisis de correspondencias múltiple ejecutado sobre las mismas variables. Por tanto, el análisis de componentes principales categórico se puede considerar como un tipo de análisis de correspondencias múltiple en el que algunas de las variables se declaran como ordinales o numéricas.

Relación con las técnicas estándar. Si todas las variables se escalan a nivel numérico, el análisis de componentes principales categórico es equivalente al análisis de componentes principales estándar.

De manera más general, el análisis de componentes principales categóricos es una alternativa al cálculo de las correlaciones entre las escalas no numéricas y su análisis mediante un enfoque de análisis de factores o de componentes principales estándar. El uso simplista del coeficiente de correlación de Pearson habitual como medida de la asociación de los datos ordinales puede llevar a un sesgo de importancia a la hora de estimar las correlaciones.

Análisis de correlación canónica no lineal

El análisis de correlación canónica no lineal es un procedimiento muy general con numerosas aplicaciones diferentes. El objetivo del análisis de correlación canónica no lineal es analizar las relaciones existentes entre dos o más conjuntos de variables en vez de entre las propias variables, como ocurre en el análisis de componentes principales. Por ejemplo, imaginemos que tiene dos conjuntos de variables, donde un conjunto de variables pueden contener elementos de información demográfica general en un conjunto de encuestados, mientras que un segundo puede contener las respuestas a un conjunto de elementos de actitud. Los niveles de escalamiento del análisis pueden ser cualquier combinación de nominal, ordinal y numérico. El análisis de correlación canónica mediante escalamiento óptimo determina la similitud existente entre los conjuntos mediante la comparación de manera simultánea de las variables canónicas de cada conjunto con un conjunto de puntuaciones de compromiso asignadas a los objetos.

Relación con otros procedimientos de categorías. Si hay dos o más conjuntos de variables con únicamente una variable por conjunto, el análisis de correlación canónica mediante escalamiento óptimo es equivalente al análisis de componentes principales mediante escalamiento óptimo. Si todas las variables en un análisis de una variable por conjunto son nominales múltiples, el análisis de correlación canónica mediante escalamiento óptimo es equivalente al análisis de correspondencias múltiple. Si hay dos conjuntos de variables, uno de los cuales contiene únicamente una variable, el análisis de correlación canónica categórica mediante escalamiento óptimo es equivalente a una regresión categórica mediante escalamiento óptimo.

Relación con las técnicas estándar. El análisis de correlación canónica típico es una técnica estadística que busca una combinación lineal de un conjunto de variables y una combinación lineal de un segundo conjunto de variables de correlación máxima. Dado este conjunto de combinaciones lineales, el análisis de correlación canónica puede buscar otros conjuntos independientes de combinaciones lineales, a los que se denominan variables canónicas, hasta un número máximo que es igual que el número de variables del conjunto más pequeño.

Si hay dos conjuntos de variables en el análisis y todas las variables se definen como numéricas, el análisis de correlación canónica mediante escalamiento óptimo es equivalente a un análisis de correlación canónica típico. Aunque IBM® SPSS® Statistics no tiene un procedimiento de análisis de correlación canónica, muchos de los estadísticos pertinentes se pueden obtener mediante un análisis multivariante de la varianza.

El análisis de correlación canónica mediante escalamiento óptimo tiene otras aplicaciones. Si tiene dos conjuntos de variables y uno de los conjuntos contiene una variable nominal declarada como nominal simple, los resultados del análisis de correlación canónica mediante escalamiento óptimo se pueden interpretar de manera similar al análisis de regresión. Si considera que la variable es nominal múltiple, el análisis mediante escalamiento óptimo es una alternativa al análisis discriminante. La agrupación de las variables en más de dos conjuntos proporciona varias maneras de analizar los datos.

Análisis de correspondencias

El objetivo del análisis de correspondencias es hacer gráficos de dispersión biespacial para las tablas de correspondencias. En una tabla de correspondencias, las variables de filas y de columnas se supone que representan categorías desordenadas; por tanto, se utiliza siempre el nivel de escalamiento óptimo nominal. Sólo se examina la información nominal de ambas variables. Es decir, la única consideración es el hecho de que algunos objetos se encuentran en la misma categoría, mientras que otros no. No se hace ninguna suposición acerca de la distancia o el orden entre las categorías de la misma variable.

Un uso específico del análisis de correspondencias es el análisis de las tablas de contingencia de doble clasificación. Si una tabla tiene r filas activas y c columnas activas, el número de dimensiones de la solución de análisis de correspondencias es el mínimo de r menos 1 o c menos 1, el que sea inferior. Dicho de otra manera, puede representar perfectamente las categorías de fila o las categorías de columna de una tabla de contingencias en un espacio de dimensiones. En lo que se refiere a la práctica, sin embargo, resulta preferible representar las categorías de fila y de columna de una tabla de doble clasificación en un espacio de pocas dimensiones, por ejemplo de dos dimensiones, ya que los gráficos bidimensionales son de más fácil comprensión que las representaciones espaciales multidimensionales.

Cuando se utiliza un número menor que el número máximo de dimensiones posibles, los estadísticos generados por el análisis describen la bondad con la que se representan las categorías de fila y de columna en la representación de pocas dimensiones. Siempre que la calidad de la representación de la solución de dos dimensiones sea buena, puede examinar los gráficos de los puntos de fila y de los puntos de columna para conocer las categorías de la variable de fila que son similares, las categorías de la variable de columna que son similares y las categorías de fila y de columna que son similares entre sí.

Relación con otros procedimientos de categorías. El análisis de correspondencias simple se limita a las tablas de doble clasificación. Si hay más de dos variables de interés, puede combinar las variables para crear variables de interacción. Por ejemplo, para las variables *región*, *trabajo* y *edad*, puede combinar *región* y *trabajo* para crear una nueva variable *retrabajo* con las 12 categorías que aparecen en la siguiente tabla. Esta nueva variable forma una tabla bidimensional con *edad* (12 filas, 4 columnas), que se pueden analizar en el análisis de correspondencias.

Tabla 1-4
Combinaciones de *región* y *trabajo*

Código de categoría	Definición de categoría	Código de categoría	Definición de categoría
1	Norte, trabajador en prácticas	7	Este, trabajador en prácticas
2	Norte, vendedor	8	Este, vendedor
3	Norte, administrador	9	Este, administrador
4	Sur, trabajador en prácticas	10	Oeste, trabajador en prácticas
5	Sur, vendedor	11	Oeste, vendedor
6	Sur, administrador	12	Oeste, administrador

Un defecto de este enfoque es que se puede combinar cualquier par de variables. Podemos combinar *trabajo* y *edad*, lo que generaría otra variable de 12 categorías. También podemos combinar *región* y *edad*, lo que generaría una nueva variable de 16 categorías. Cada una de estas

variables de interacción forma una tabla de doble clasificación con la variable restante. El análisis de correspondencias de estas tres tablas no generará resultados idénticos, aunque cada uno de los enfoques es válido. Además, si hay cuatro o más variables, se pueden generar las tablas de doble clasificación que comparen una variable de interacción con otra variable de interacción. El número de posibles tablas que se pueden analizar puede ser bastante grande, incluso cuando hay pocas variables. Puede seleccionar una de estas tablas para analizarla o puede analizarlas todas. Otra posibilidad consiste en utilizar el procedimiento Análisis de correspondencias múltiple para examinar todas las variables simultáneamente sin que sea necesario crear variables de interacción.

Relación con las técnicas estándar. El procedimiento Tablas de contingencia también se puede utilizar para analizar tablas de contingencia, con independencia como un enfoque común en los análisis. Sin embargo, incluso en las tablas pequeñas, la detección de las desviaciones de la independencia puede ser difícil. La utilidad del análisis de correspondencias reside en la presentación de tales patrones para tablas de doble clasificación de cualquier tamaño. Si hay una asociación entre las variables de fila y de columna (es decir, el valor de chi-cuadrado es significativo) el análisis de correspondencias puede ayudar a revelar la naturaleza de la relación.

Análisis de correspondencias múltiple

El análisis de correspondencias múltiple intenta generar una solución en la que los objetos de la misma categoría se representan cerca los unos de los otros y los objetos de categorías diferentes se representan alejados los unos de los otros. Cada objeto se encuentra lo más cerca posible de los puntos de categoría para las categorías que se aplican a dicho objeto. De esta manera, las categorías dividen los objetos en subgrupos homogéneos. Las variables se consideran homogéneas cuando clasifican objetos de las mismas categorías en los mismos subgrupos.

Para una solución de una sola dimensión, el análisis de correspondencias múltiple asigna valores de escala óptimos (cuantificaciones de las categorías) a cada categoría de cada variable de forma que en general, como media, las categorías tengan una máxima dispersión. Para una solución de dos dimensiones, el análisis de correspondencias múltiple busca un segundo conjunto de cuantificaciones de las categorías de cada variable no relacionada con el primer conjunto, volviendo a intentar maximizar la dispersión y así sucesivamente. Como las categorías de una variable reciben tantas puntuaciones como dimensiones haya, se supone que las variables del análisis son nominales múltiples en el nivel de escalamiento óptimo.

El análisis de correspondencias múltiple también asigna puntuaciones a los objetos del análisis de manera que las cuantificaciones de las categorías son los promedios, o los centroides, de las puntuaciones de los objetos de dicha categoría.

Relación con otros procedimientos de categorías. El análisis de correspondencias múltiple también se conoce como análisis de homogeneidad o escalamiento dual. Proporciona resultados comparables, pero no idénticos, al análisis de correspondencias cuando hay sólo dos variables. El análisis de correspondencias genera resultados únicos que resumen el ajuste y la calidad de la representación de la solución, incluida la información sobre la estabilidad. Por tanto, el análisis de correspondencias habitualmente resulta preferible al análisis de correspondencias múltiple en el caso de dos variables. Otra diferencia entre los dos procedimientos es que la entrada al análisis de correspondencias múltiple es una matriz de datos, donde las filas son objetos y las columnas son variables, mientras que la entrada al análisis de correspondencias puede ser la misma matriz de datos, una matriz de proximidades general o una tabla de contingencias conjunta, que es una

matriz agregada en la que tanto las filas como las columnas representan categorías de variables. El análisis de correspondencias múltiple también se puede considerar como un análisis de componentes principales de datos escalados al nivel nominal múltiple.

Relación con las técnicas estándar. El análisis de correspondencias múltiple se puede considerar como el análisis de una tabla de contingencia de clasificación múltiple. Las tablas de contingencia de clasificación múltiple también se pueden analizar con el procedimiento Tablas de contingencia, pero Tablas de contingencia proporciona estadísticos de resumen independientes para cada categoría de cada variable de control. Con el análisis de correspondencias múltiple, con frecuencia es posible resumir la relación entre todas las variables con un único gráfico bidimensional. Un uso avanzado del análisis de correspondencias múltiple es sustituir los valores originales de la categoría por los valores de la escala óptimos de la primera dimensión y realizar un análisis multivariante secundario. Ya que el análisis de correspondencias múltiple sustituye las etiquetas de categoría por valores de escala numéricos, se pueden aplicar muchos procedimientos diferentes que requieren datos numéricos tras el análisis de correspondencias múltiple. Por ejemplo, el procedimiento Análisis factorial produce un primer componente principal que es equivalente a la primera dimensión del análisis de correspondencias múltiple. Las puntuaciones de los componentes de la primera dimensión son iguales que las puntuaciones de los objetos y las saturaciones en componentes al cuadrado son iguales que las medidas de discriminación. La segunda dimensión del análisis de correspondencias múltiple, sin embargo, no es igual que la segunda dimensión del análisis factorial.

Multidimensional Scaling

El uso del escalamiento multidimensional es especialmente apropiado cuando el objetivo del análisis es buscar la estructura de un conjunto de medidas de distancia entre un solo conjunto de objetos o casos. Esto se logra asignando las observaciones a posiciones específicas en un espacio conceptual de pocas dimensiones, de modo que las distancias entre los puntos en el espacio concuerden al máximo con las similitudes (o disimilitudes) dadas. El resultado es una representación de mínimos cuadrados de los objetos en dicho espacio de pocas dimensiones que, en muchos casos, le ayudará a entender mejor los datos.

Relación con otros procedimientos de categorías. Cuando tiene datos multivariantes a partir de los que se crean distancias y, a continuación, los analiza con el escalamiento multidimensional, los resultados son similares al análisis de los datos mediante el análisis de componentes principales categóricos con la normalización principal de objetos. Este tipo de PCA también se conoce como análisis de coordenadas principales.

Relación con las técnicas estándar. El procedimiento Escalamiento multidimensional (PROXSCAL) de categorías ofrece varias mejoras sobre el procedimiento de escalamiento disponible en la opción Estadísticas básicas (ALSCAL). PROXSCAL ofrece un algoritmo acelerado para ciertos modelos y permite colocar restricciones en el espacio común. Además, PROXSCAL intenta minimizar el stress bruto normalizado, en vez del S-stress (también denominado **tensión**). El stress bruto normalizado se prefiere en general ya que es una medida basada en las distancias, mientras que el S-stress se basa en los cuadrados de las distancias.

Desplegamiento multidimensional

El uso del despliegamiento multidimensional es especialmente apropiado cuando el objetivo del análisis es buscar la estructura de un conjunto de medidas de distancia entre dos conjuntos de objetos (denominados objetos de fila y de columna). Esto se logra asignando las observaciones a posiciones específicas en un espacio conceptual de pocas dimensiones, de modo que las distancias entre los puntos en el espacio concuerden al máximo con las similaridades (o disimilaridades) dadas. El resultado es una representación de mínimos cuadrados de los objetos de fila y de columna en dicho espacio de pocas dimensiones que, en muchos casos, le ayudará a entender mejor los datos.

Relación con otros procedimientos de categorías. Si los datos constan de distancias entre un único conjunto de objetos (una matriz simétrica cuadrada), utilice Escalamiento multidimensional.

Relación con las técnicas estándar. El procedimiento Desplegamiento multidimensional (PREFSCAL) de categorías ofrece varias mejoras sobre la funcionalidad de despliegamiento disponible en la opción Estadísticas básicas (a través de ALSCAL). PREFSCAL permite colocar restricciones en el espacio común; es más, PREFSCAL intenta minimizar una medida de tensión penalizada que le ayuda a evitar las soluciones degeneradas (a las que los algoritmos antiguos son propensos).

Relación de aspecto en gráficos de escalamiento óptimo

La relación de aspecto en los gráficos de escalamiento óptimo es isótropa. En un gráfico bidimensional, la distancia que representa una unidad en la dimensión 1 es igual que la distancia que representa una unidad en la dimensión 2. Si cambia el rango de una dimensión en un diagrama bidimensional, el sistema cambiará el tamaño de la otra dimensión para mantener iguales las distancias físicas. La relación de aspecto isótropa no se puede omitir para los procedimientos de escalamiento óptimo.

Lecturas recomendadas

Si desea obtener información general sobre las técnicas de escalamiento óptimo consulte los siguientes textos:

Barlow, R. E., D. J. Bartholomew, D. J. Bremner, y H. D. Brunk. 1972. *Statistical inference under order restrictions*. Nueva York: John Wiley and Sons.

Benzécri, J. P. 1969. Statistical analysis as a tool to make patterns emerge from data. En: *Methodologies of Pattern Recognition*, S. Watanabe, ed. Nueva York: Academic Press.

Bishop, Y. M., S. E. Feinberg, y P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: MIT Press.

De Leeuw, J. 1984. The Gifi system of nonlinear multivariate analysis. En: *Data Analysis and Informatics III*, E. Diday, et al., ed..

De Leeuw, J. 1990. Multivariate analysis with optimal scaling. En: *Progress in Multivariate Analysis*, S. Das Gupta, y J. Sethuraman, eds. Calcuta: Indian Statistical Institute.

- De Leeuw, J., y J. Van Rijckeversel. 1980. HOMALS and PRINCALS—Some generalizations of principal components analysis. En: *Data Analysis and Informatics*, E. Diday, et al., ed. Ámsterdam: North-Holland.
- De Leeuw, J., F. W. Young, y Y. Takane. 1976. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .
- Gifi, A. 1990. *Nonlinear multivariate analysis*. Chichester: John Wiley and Sons.
- Heiser, W. J., y J. J. Meulman. 1995. Nonlinear methods for the analysis of homogeneity and heterogeneity. En: *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, ed. Oxford: Oxford University Press.
- Israëls, A. 1987. *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.
- Krzanowski, W. J., y F. H. C. Marriott. 1994. *Multivariate analysis: Part I, distributions, ordination and inference*. Londres: Edward Arnold.
- Lebart, L., A. Morineau, y K. M. Warwick. 1984. *Multivariate descriptive statistical analysis*. Nueva York: John Wiley and Sons.
- Máx, J. 1960. Quantizing for minimum distortion. *Proceedings IEEE (Information Theory)*, 6, .
- Meulman, J. J. 1986. *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
- Meulman, J. J. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika*, 57, .
- Nishisato, S. 1980. *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Nishisato, S. 1994. *Elements of dual scaling: An introduction to practical data analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Rao, C. R. 1973. *Linear statistical inference and its applications*, 2nd ed. Nueva York: John Wiley and Sons.
- Rao, C. R. 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. En: *Multivariate Analysis, Vol. 5*, P. R. Krishnaiah, ed. Ámsterdam: North-Holland.
- Roskam, E. E. 1968. *Metric analysis of ordinal data in psychology*. Voorschoten: VAM.
- Shepard, R. N. 1966. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 3, .
- Wolter, K. M. 1985. *Introduction to variance estimation*. Berlín: Springer-Verlag.
- Young, F. W. 1981. Quantitative analysis of qualitative data. *Psychometrika*, 46, .

Regresión categórica (CATREG)

La **regresión categórica** cuantifica los datos categóricos mediante la asignación de valores numéricos a las categorías, obteniéndose una ecuación de regresión lineal óptima para las variables transformadas. La regresión categórica se conoce también por el acrónimo CATREG, del inglés *categorical regression* (regresión categórica).

El análisis de regresión lineal ordinario implica minimizar las diferencias de la suma de los cuadrados entre una variable de respuesta (la dependiente) y una combinación ponderada de las variables predictoras (las independientes). Las variables son normalmente cuantitativas, con los datos categóricos (nominales) recodificados como variables binarias o de contraste. Como resultado, las variables categóricas sirven para separar grupos de casos y la técnica estima conjuntos separados de parámetros para cada grupo. Los coeficientes estimados reflejan cómo los cambios en los predictores afectan a la respuesta. El pronóstico de la respuesta es posible para cualquier combinación de los valores predictores.

Un método alternativo incluye la regresión de la respuesta respecto a los propios valores predictores categóricos. Como consecuencia, se estima un coeficiente para cada variable. Sin embargo, para las variables categóricas, los valores categóricos son arbitrarios. La codificación de las categorías de diferentes maneras proporciona diferentes coeficientes, dificultando las comparaciones entre los análisis de las mismas variables.

CATREG amplía el método típico mediante un escalamiento de las variables nominales, ordinales y numéricas simultáneamente. El procedimiento cuantifica las variables categóricas de manera que las cuantificaciones reflejen las características de las categorías originales. El procedimiento trata a las variables categóricas cuantificadas como si fueran variables numéricas. La utilización de transformaciones no lineales permite a las variables ser analizadas en varios niveles para encontrar el modelo que más se ajusta.

Ejemplo. La regresión categórica se puede utilizar para describir cómo la satisfacción laboral depende de la categoría laboral, de la región geográfica y del número de desplazamientos. Observará que los niveles más altos de satisfacción corresponden a ejecutivos y a un número bajo de desplazamientos. La ecuación de regresión resultante se puede utilizar para pronosticar la satisfacción laboral para cualquier combinación de las tres variables independientes.

Estadísticos y gráficos. Frecuencias, coeficientes de regresión, tabla ANOVA, historial de iteraciones, cuantificaciones de categorías, correlaciones entre predictores no transformados, correlaciones entre predictores transformados, gráficos de residuos y gráficos de transformación.

Datos. CATREG trabaja con variables indicadoras de categorías. Los indicadores de las categorías deben ser enteros positivos. Puede utilizar el cuadro de diálogo Discretización para convertir variables con valores fraccionarios y variables de cadena en enteros positivos.

Supuestos. Sólo se permite una variable de respuesta, pero el número máximo de predictores es 200. Los datos deben contener al menos tres casos válidos y el número de casos válidos debe ser superior al número de variables predictoras más uno.

Procedimientos relacionados. CATREG es equivalente al análisis de correlación canónica categórico mediante escalamiento óptimo (OVERALS) con dos conjuntos, uno de los cuales contiene sólo una variable. Si se escalan todas las variables a nivel numérico, el análisis se corresponderá con el análisis de regresión múltiple típico.

Para obtener una regresión categórica

- ▶ Elija en los menús:
Analizar > Regresión > Escalamiento óptimo (CATREG)...

Figura 2-1
Cuadro de diálogo Regresión categórica

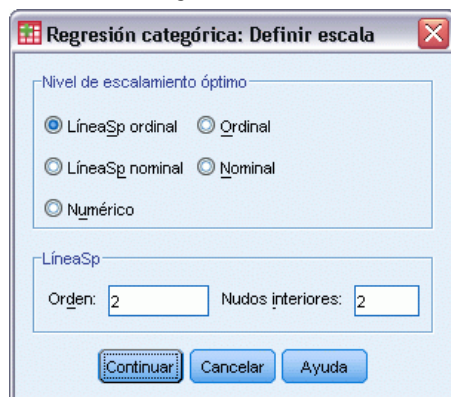


- ▶ Seleccione la variable dependiente y la variable o variables independientes.
- ▶ Pulse en Aceptar.
Si lo desea, cambie el nivel de escalamiento de cada variable.

Para definir la escala en regresión categórica

Se puede definir el nivel de escalamiento óptimo de las variables dependientes e independientes. Por defecto, se escalan como líneasSp (ordinales) monotónicas de segundo orden con dos nudos interiores. Asimismo, se puede definir la ponderación para las variables del análisis.

Figura 2-2
Cuadro de diálogo Definir escala



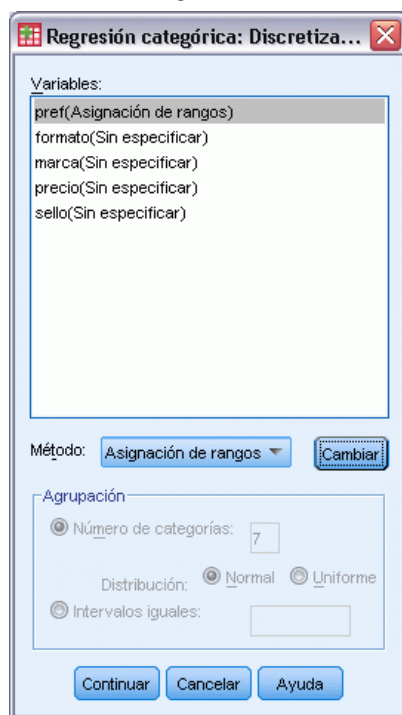
Nivel de escalamiento óptimo. También se puede seleccionar el nivel de escalamiento para cuantificar cada variable.

- **LíneaSp ordinal.** El orden de las categorías de la variable observada se conserva en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante es un polinomio monótonico por tramos suave del orden seleccionado. Las partes se especifican por el número de nudos interiores definido por el usuario y su posición es determinada por el procedimiento en función del número de nudos interiores.
- **LíneaSp nominal.** La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en categorías. No se conserva el orden de las categorías de la variable observada. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante es un polinomio, posiblemente monótonico, por tramos suave del orden seleccionado. Las partes se especifican por el número de nudos interiores definido por el usuario y su posición es determinada por el procedimiento en función del número de nudos interiores.
- **Ordinal.** El orden de las categorías de la variable observada se conserva en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante se ajusta mejor que la transformación de líneaSp ordinal pero la suavidad es menor.
- **Nominal.** La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en categorías. No se conserva el orden de las categorías de la variable observada. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante se ajusta mejor que la transformación de líneaSp nominal pero la suavidad es menor.
- **Numérico.** Las categorías se tratan como que están ordenadas y espaciadas uniformemente (a nivel de intervalo). El orden de las categorías y la equidistancia entre los números de las categorías de la variable observada se conservan en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. Cuando todas las variables están a nivel numérico, el análisis es análogo al análisis de componentes principales típico.

Discretización de regresión categórica

El cuadro de diálogo Discretización permite seleccionar un método para recodificar las variables. Las variables con valores fraccionarios se agrupan en siete categorías (o en el número de valores diferentes de la variable si dicho número es inferior a siete) con una distribución aproximadamente normal, si no se especifica lo contrario. Las variables de cadena se convierten siempre en enteros positivos mediante la asignación de indicadores de categoría en función del orden alfanumérico ascendente. La discretización de las variables de cadena se aplica a estos enteros resultantes. Por defecto, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Figura 2-3
Cuadro de diálogo Discretización



Método. Seleccione entre agrupación, asignación de rangos y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.
- **Asignación de rangos.** La variable se discretiza mediante la asignación de rangos a los casos.
- **Multiplicación.** Los valores actuales de la variable se tipifican, multiplican por 10, redondean y se les suma una constante de manera que el menor valor discretizado sea 1.

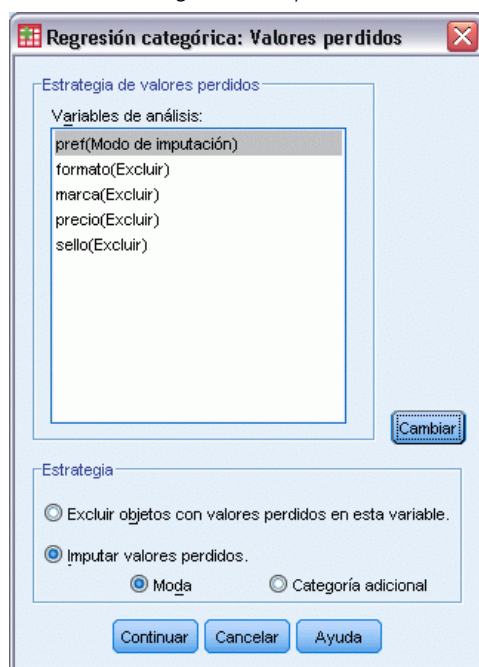
Agrupación. Se encuentran disponibles las siguientes opciones al discretizar variables por agrupación:

- **Número de categorías.** Especifique un número de categorías y si los valores de la variable deben seguir una distribución aproximadamente normal o uniforme en dichas categorías.
- **Intervalos iguales.** Las variables se recodifican en las categorías definidas por dichos intervalos de igual tamaño. Se debe especificar la longitud de los intervalos.

Valores perdidos de regresión categórica

El cuadro de diálogo Valores perdidos permite seleccionar la estrategia para el tratamiento de los valores perdidos en las variables de análisis y las suplementarias.

Figura 2-4
Cuadro de diálogo Valores perdidos



Estrategia. Seleccione excluir los objetos con valores perdidos (eliminación por lista) o imputar los valores perdidos (tratamiento activo).

- **Excluir objetos con valores perdidos en esta variable.** Los objetos con valores perdidos en la variable seleccionada se excluyen del análisis. Esta estrategia no está disponible para las variables suplementarias.
- **Imputar valores perdidos.** Los objetos con valores perdidos en la variable seleccionada tendrán dichos valores imputados. Se puede seleccionar el método de imputación. Seleccione **Moda** para reemplazar los valores perdidos por la categoría más frecuente. Cuando existen varias modas, se utiliza la que tiene el indicador de categoría más pequeño. Seleccione **Categoría adicional** para reemplazar los valores perdidos con la misma cuantificación de una categoría adicional. Esto implica que los objetos con un valor perdido en esta variable se consideran que pertenecen a la misma categoría (la adicional).

Opciones de regresión categórica

El cuadro de diálogo Opciones permite seleccionar el estilo para la configuración inicial, especificar los criterios de iteración y convergencia, seleccionar los objetos suplementarios y definir el etiquetado de los gráficos.

Figura 2-5
Cuadro de diálogo Opciones

Objetos suplementarios. Permite especificar los objetos que se tratarán como suplementarios. Sólo hay que escribir el número de un objeto suplementario (o especificar un intervalo de casos) y pulsar en Añadir. No es posible ponderar los objetos suplementarios (se ignoran las ponderaciones especificadas).

Configuración inicial. Si no se trata ninguna variable como nominal, seleccione la configuración Numérica. Si al menos una variable se trata como nominal, seleccione la configuración Aleatoria.

Además, si al menos una variable tiene un nivel de escalamiento ordinal o líneaSp ordinal, el algoritmo normal de ajuste de modelo puede dar lugar a una solución inferior a la óptima. Si selecciona Múltiples inicios sistemáticos con todos los patrones de signos de prueba posibles siempre encontrará la solución óptima, pero el tiempo de procesamiento necesario aumentará rápidamente, al aumentar el número de variables ordinales y de líneaSp ordinal del conjunto de datos. Puede reducir el número de patrones de prueba especificando un porcentaje de pérdida de umbral de varianza, donde mientras mayor sea el umbral, más patrones de signos se excluirán.

Con esta opción, no se garantiza obtener la solución óptima, pero se reduce la posibilidad de obtener una solución que no sea óptima. Además, si no se encuentra la solución óptima, disminuye la posibilidad de que la solución encontrada sea muy diferente de la solución óptima. Cuando se solicitan múltiples inicios sistemáticos, los signos de los coeficientes de regresión para cada inicio se escriben en un archivo de datos de IBM® SPSS® Statistics externo en la sesión actual. [Si desea obtener más información, consulte el tema Regresión categórica: Guardar el p. 24.](#)

Los resultados de una ejecución anterior con múltiples inicios sistemáticos le permiten Utilizar signos fijos para los coeficientes de regresión. Los signos (indicados por 1 y -1) deben estar en una fila del conjunto de datos o archivo especificado. El número de inicio de valor entero representa el número de caso de la fila de este archivo que contiene los signos que se utilizarán.

Criterios. Se puede especificar el número máximo de iteraciones que la regresión puede realizar durante los cálculos. También puede seleccionar un valor para el criterio de convergencia. La regresión detiene la iteración si la diferencia del ajuste total entre la dos últimas iteraciones es menor que el valor de convergencia o si se ha alcanzado el número máximo de iteraciones.

Etiquetar gráficos con. Permite especificar si se utilizarán en los gráficos las etiquetas de variable y las etiquetas de valor o los nombres de variable y los valores. También se puede especificar una longitud máxima para las etiquetas.

Regularización de regresión categórica

Figura 2-6
Cuadro de diálogo Regularización

Regresión categórica: Regularización

Método

Ninguno

Regresión de cresta

Mínimo: 0.0 Máximo: 1.0 Incremento: 0.02

Lazo

Mínimo: 0.0 Máximo: 1.0 Incremento: 0.02

Red elástica

Regresión de cresta: Mínimo: 0.0 Máximo: 1.0 Incremento: 0.1

Lazo: Mínimo: 0.0 Máximo: 1.0 Incremento: 0.02

Gráficos de red elástica

Generar todos los gráficos de red elástica posibles

Generar gráficos de red elástica para algunas penalizaciones de cresta

Rango de valores

Primero: 0.2

Último: 0.6

Valor único

Valor:

Valores de penalización de cresta

Lista de valores de penalización

Añadir

Cambiar

Eliminar

Mostrar gráficos de regularización

Continuar Cancelar Ayuda

Método. Los métodos de regularización pueden mejorar el error predictivo del modelo reduciendo la variabilidad de las estimaciones de coeficiente de regresión reduciendo las estimaciones hacia 0. Lazo y Red elástica reducen algunas estimaciones de coeficientes a exactamente 0, ofreciendo así una forma de selección de variables. Cuando se solicita un método de regularización, el modelo y los coeficientes regularizados para cada valor de coeficiente de penalización se escriben en un archivo o conjunto de datos de IBM® SPSS® Statistics externo en la sesión actual. [Si desea obtener más información, consulte el tema Regresión categórica: Guardar el p. 24.](#)

- **Regresión contraída** La regresión contraída reduce los coeficientes introduciendo un término de penalización igual a la suma de coeficientes cuadrados por un **coeficiente de penalización**. Este coeficiente puede variar de 0 (sin penalización) a 1; el procedimiento buscará el “mejor” valor de penalización si especifica un rango e incremento.
- **Lazo.** El término de penalización de Lazo se basa en la suma de coeficientes absolutos y la especificación de un coeficiente de penalización es similar al de la regresión contraída; sin embargo, Lazo supone un procesamiento más extenso.
- **Red elástica.** La Red elástica simplemente combina las penalizaciones de Lazo y regresión contraída, y buscará en la cuadrícula de valores especificada para encontrar los “mejores” coeficientes de penalización de Lazo y de regresión contraída. En un par concreto de penalizaciones de Lazo y regresión contraída, la red elástica no requiere mucho más procesamiento que Lazo.

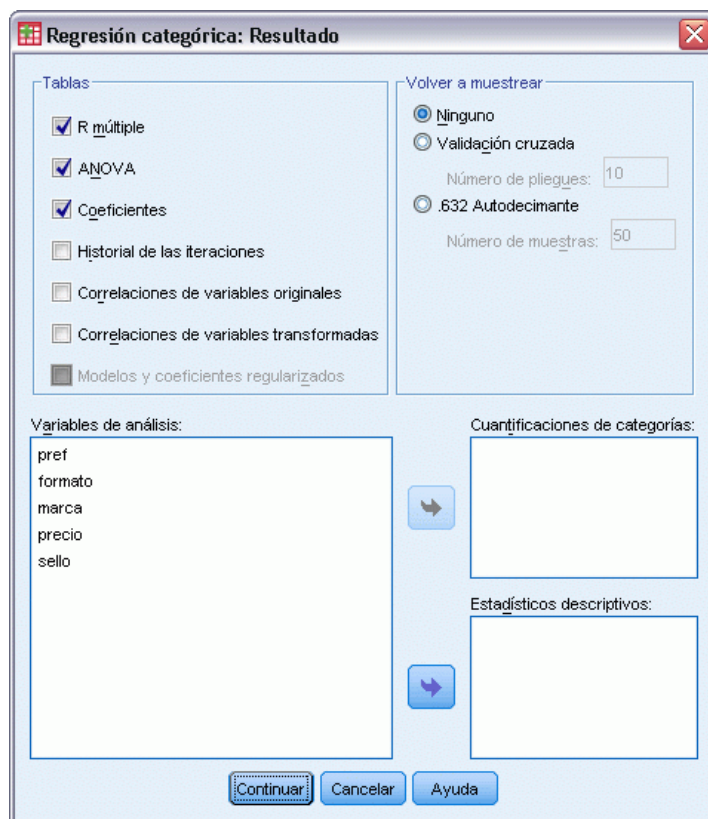
Gráficos de regularización de visualización. Son gráficos de los coeficientes de regresión en comparación con la penalización de regularización. Cuando busque un rango de valores para el “mejor” coeficiente de penalización, ofrecerá una vista de la forma en que los coeficientes cambian en dicho rango.

Gráficos de red elástica. En el método de Red elástica se producen gráficos de regularización distintos mediante los valores de la penalización de regresión contraída. Todos los gráficos posibles utiliza todos los valores del rango determinados por los valores mínimo y máximo especificados de penalización de regresión contraída. Para algunas penalizaciones de r. contraída le permite especificar un subconjunto de valores en el rango determinado por el mínimo y el máximo. Sólo hay que escribir el número de un valor de penalización (o especificar un rango de valores) y pulsar en Añadir.

Resultados de la regresión categórica

El cuadro de diálogo Resultados permite seleccionar los estadísticos que aparecerán en los resultados.

Figura 2-7
Cuadro de diálogo Resultados



Tablas. Genera tablas correspondientes a:

- **R múltiple.** Incluye R^2 , R^2 corregida y R^2 corregida que tiene en cuenta el escalamiento óptimo.
- **ANOVA.** Esta opción incluye las sumas de cuadrados de la regresión y de los residuos, las medias cuadráticas y F . Se muestran dos tablas ANOVA: una con tantos grados de libertad para la regresión como predictores haya y otra con los grados de libertad de la regresión teniendo en cuenta el escalamiento óptimo.
- **Coeficientes.** Esta opción produce tres tablas: una tabla de coeficientes que incluye betas, error típico de las betas, valores t y significación; una tabla de escalamiento óptimo de coeficientes con el error típico de las betas que tiene en cuenta los grados de libertad del escalamiento óptimo; y una tabla con las correlaciones de orden cero, parciales y semiparciales, la medida de importancia relativa de Pratt de los predictores transformados y la tolerancia antes y después de la transformación.
- **Historial de iteraciones.** Para cada iteración, incluyendo los valores iniciales del algoritmo, aparecen la R múltiple y el error de la regresión. El incremento de la R múltiple aparece a partir de la primera iteración.
- **Correlaciones de variables originales.** Aparece una matriz que muestra las correlaciones entre las variables no transformadas.

- **Correlaciones de variables transformadas.** Aparece una matriz que muestra las correlaciones entre las variables transformadas.
- **Coefficientes y modelos regularizados.** Muestra valores de penalización, R cuadrado y los coeficientes de regresión de cada modelo regularizado. Si se especifica un método de repetición de muestreo o si se especifican objetos suplementarios (casos de prueba), también mostrará el error de predicción o MSE de prueba.

Repetición de muestreo. Los métodos de repetición de muestreo le dan una estimación del error de predicción del modelo.

- **Validación cruzada.** La validación cruzada divide la muestra en un número de submuestras o pliegues. A continuación, se generan los modelos de regresión categórica, que no incluyen los datos de cada submuestra. El primer modelo se basa en todos los casos excepto los correspondientes al primer pliegue de la muestra; el segundo modelo se basa en todos los casos excepto los del segundo pliegue de la muestra y así sucesivamente. Para cada modelo se calcula el error de predicción aplicando el modelo a la submuestra que se excluyó al generarse este.
- **Autodocimante .632.** Con el autodocimante, las observaciones se trazan aleatoriamente a partir de los datos con reposición, repitiendo este proceso una serie de veces para obtener muestras de autodocimante. Se ajusta un modelo para cada muestra de autodocimante, y el error de predicción de cada modelo se estima según este modelo ajustado y se aplica después a los casos que no se encuentran en la muestra de autodocimante.

Cuantificaciones de categorías. Aparecen tablas que muestran los valores transformados de las variables seleccionadas.

Estadísticos descriptivos. Aparecen tablas que muestran las frecuencias, los valores perdidos y los modos de las variables seleccionadas.

Regresión categórica: Guardar

El cuadro de diálogo Guardar permite guardar valores pronosticados, residuos y valores transformados en un conjunto de datos activo y/o guardar datos discretizados, valores transformados, modelos y coeficientes regularizados y signos de coeficientes de regresión en un archivo de datos o conjunto de datos de IBM® SPSS® Statistics externo en la sesión actual.

- Los conjuntos de datos están disponibles durante la sesión actual, pero no así en las sesiones posteriores, a menos que los haya guardado explícitamente como archivos de datos. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.
- Los nombres de archivo o de conjunto de datos deben ser distintos para cada tipo de dato guardado.

Figura 2-8
Cuadro de diálogo Guardar

Regresión categórica: Guardar

Guardar valores pronosticados en el conjunto de datos activo

Guardar residuos en el conjunto de datos activo

Datos discretizados

Crear datos discretizados

Crear un nuevo conjunto de datos

Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Examinar...

Modelos y coeficientes regularizados

Crear un nuevo conjunto de datos

Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Examinar...

Variables transformadas

Guardar variables transformadas en el conjunto de datos activo

Guardar las variables transformadas a un nuevo conjunto de datos o archivo

Crear un nuevo conjunto de datos

Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Examinar...

Continuar Cancelar Ayuda

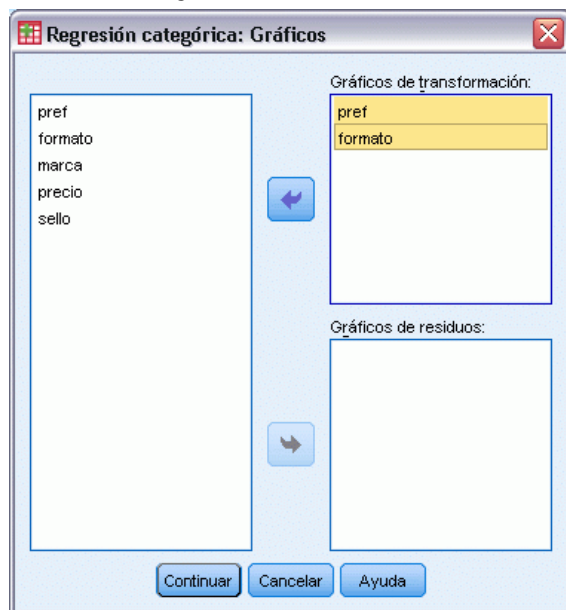
Los coeficientes y modelos regularizados se guardan siempre que se selecciona un método de regularización en el cuadro de diálogo [Regularización](#). Por defecto, el procedimiento crea un nuevo conjunto de datos con un nombre único, aunque por supuesto podrá especificar el nombre que desee seleccionando o escribiendo en un archivo externo.

Los signos de coeficientes de regresión se guardan siempre que se utilizan múltiples inicios sistemáticos como configuración inicial en el cuadro de diálogo [Opciones](#). Por defecto, el procedimiento crea un nuevo conjunto de datos con un nombre único, aunque por supuesto podrá especificar el nombre que desee seleccionando o escribiendo en un archivo externo.

Gráficos de transformación en regresión categórica

El cuadro de diálogo Gráficos permite especificar las variables que generarán gráficos de transformación y de residuos.

Figura 2-9
Cuadro de diálogo Gráficos



Gráficos de transformación. Para cada una de estas variables, se representan las cuantificaciones de las categorías respecto a los valores de categoría originales. Las categorías vacías aparecen en el eje horizontal pero no afectan a los cálculos. Estas categorías se identifican por las rupturas en la línea que conecta las cuantificaciones.

Gráficos de residuos. Para cada una de estas variables, se representan los residuos (calculados para la variable dependiente pronosticada de todos los predictores salvo del predictor en cuestión) respecto a los indicadores de las categorías y las cuantificaciones de las categorías óptimas multiplicadas por beta respecto a los indicadores de las categorías.

Funciones adicionales del comando CATREG

Se puede personalizar la regresión categórica si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando CATREG resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar nombres de raíz para las variables transformadas al guardarlas en el conjunto de datos activo (con el subcomando *SAVE*).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Análisis de componentes principales categórico (CATPCA)

Este procedimiento cuantifica simultáneamente las variables categóricas a la vez que reduce la dimensionalidad de los datos. El análisis de componentes principales categórico se conoce también por el acrónimo CATPCA, del inglés *CAT*egorical Principal Components Analysis.

El objetivo de los análisis de componentes principales es la reducción de un conjunto original de variables en un conjunto más pequeño de componentes no correlacionados que representen la mayor parte de la información encontrada en las variables originales. La técnica es más útil cuando un extenso número de variables impide una interpretación eficaz de las relaciones entre los objetos (sujetos y unidades). Al reducir la dimensionalidad, se interpreta un pequeño número de componentes en lugar de un extenso número de variables.

El análisis típico de componentes principales asume relaciones lineales entre las variables numéricas. Por otra parte, el método de escalamiento óptimo permite escalar las variables a diferentes niveles. Las variables categóricas se cuantifican de forma óptima en la dimensionalidad especificada. Como resultado, se pueden modelar relaciones no lineales entre las variables.

Ejemplo. El análisis de componentes principales categórico se puede utilizar para representar gráficamente la relación entre la categoría laboral, la división laboral, la provincia, el número de desplazamientos (alto, medio y bajo) y la satisfacción laboral. Observará que con dos dimensiones se puede explicar una gran cantidad de varianza. La primera dimensión podría separar la categoría laboral de la provincia, mientras que la segunda dimensión podría separar la división laboral del número de desplazamientos. También podrá observar que la alta satisfacción laboral está relacionada con un número medio de desplazamientos.

Estadísticos y gráficos. Frecuencias, valores perdidos, nivel de escalamiento óptimo, moda, varianza explicada por: las coordenadas del centroide, las coordenadas de vector, total por variable y total por dimensión; saturaciones en componentes para las variables cuantificadas por los vectores, cuantificaciones y coordenadas de categoría, historial de iteraciones, correlaciones entre las variables transformadas y los autovalores de la matriz de correlaciones, correlaciones entre las variables originales y los autovalores de la matriz de correlaciones, puntuaciones de objetos, gráficos de categorías, gráficos de categorías conjuntas, gráficos de transformación, gráficos de residuos, gráficos de centroides proyectados, gráficos de objetos, diagramas de dispersión biespaciales, diagramas de dispersión triespaciales y gráficos de las saturaciones en componentes.

Datos. Los valores de las variables de cadena se convierten en enteros positivos por orden alfabético ascendente. Los valores perdidos definidos por el usuario, los valores perdidos del sistema y los valores menores que 1 se consideran valores perdidos; se puede añadir una constante o recodificar las variables con valores inferiores a 1 para evitar que se pierdan los mismos.

Supuestos. Los datos deben contener al menos tres casos válidos. El análisis se basa en datos enteros positivos. La opción de discretización categorizará de forma automática una variable con valores fraccionarios, agrupando sus valores en categorías con una distribución casi “normal” y

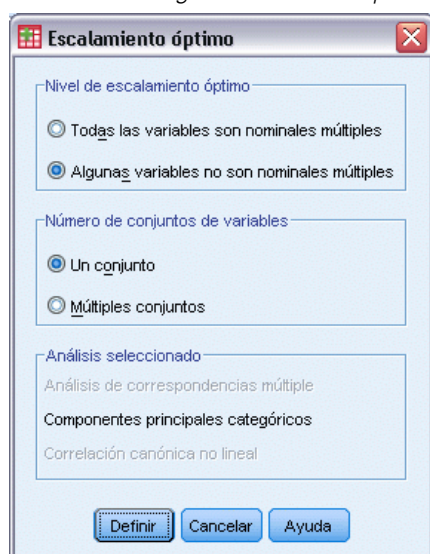
convertirá de forma automática los valores de las variables de cadena en enteros positivos. Se pueden especificar otros esquemas de discretización.

Procedimientos relacionados. Si se escalan todas las variables a nivel numérico, el análisis se corresponderá con el análisis de componentes principales típico. Hay funciones de representación alternativas que están disponibles si se utilizan las variables transformadas en un análisis de componentes principales lineal típico. Si todas las variables tienen un nivel de escalamiento nominal múltiple, el análisis de componentes principales categórico es idéntico al análisis de correspondencias múltiple (HOMALS). Si hay conjuntos de variables que son de interés, se debe utilizar el análisis de correlación canónica categórico (no lineal).

Para obtener un Análisis de componentes principales categórico

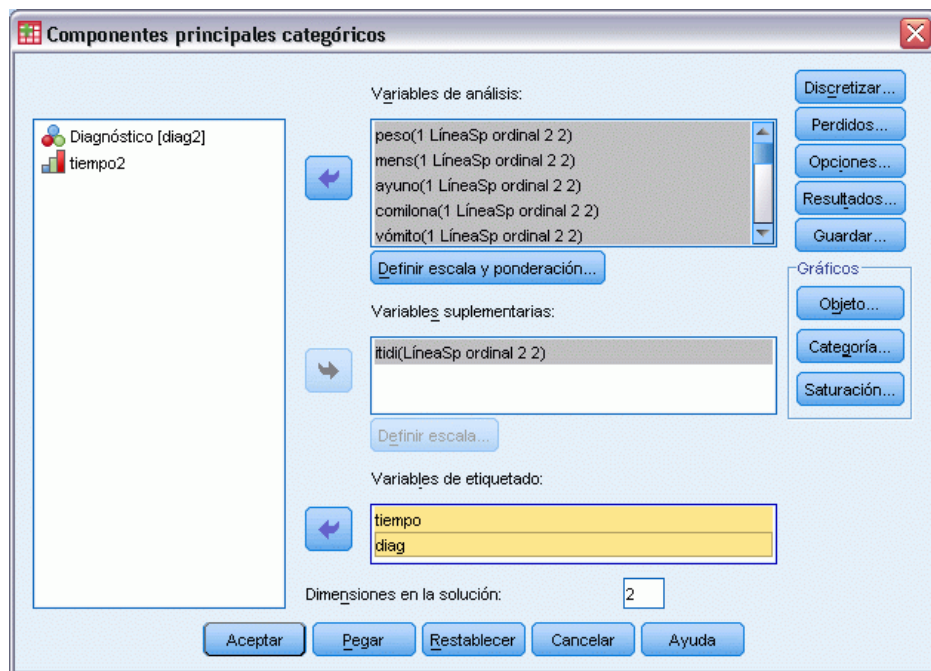
- ▶ Elija en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 3-1
Cuadro de diálogo Escalamiento óptimo



- ▶ Seleccione *Alguna variable no es nominal múltiple*.
- ▶ Seleccione *Un conjunto*.
- ▶ Pulse en *Definir*.

Figura 3-2
Cuadro de diálogo Componentes principales categórico



- Seleccione al menos dos variables de análisis y especifique el número de dimensiones en la solución.

- Pulse en Aceptar.

Si lo desea, puede especificar variables suplementarias, que serán ajustadas sobre la solución encontrada, o variables de etiqueta para los gráficos.

Definir escala y ponderación en CATPCA

Se puede definir el nivel de escalamiento óptimo de las variables del análisis y de las suplementarias. Por defecto, se escalan como líneasSp (ordinales) monotónicas de segundo orden con dos nudos interiores. Asimismo, se puede definir la ponderación para las variables del análisis.

Figura 3-3
Definir escala y ponderación

Ponderación de la variable. Se puede definir una ponderación para cada variable. El valor especificado debe ser un entero positivo. El valor por defecto es 1.

Nivel de escalamiento óptimo. Asimismo, se puede seleccionar el nivel de escalamiento que se utilizará para cuantificar cada variable.

- **LíneaSp ordinal.** El orden de las categorías de la variable observada se conserva en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante es un polinomio monotónico por tramos suave del orden seleccionado. Las partes se especifican por el número de nudos interiores definido por el usuario y su posición es determinada por el procedimiento en función del número de nudos interiores.
- **LíneaSp nominal.** La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en categorías. No se conserva el orden de las categorías de la variable observada. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante es un polinomio, posiblemente monotónico, por tramos suave del orden seleccionado. Las partes se especifican por el número de nudos interiores definido por el usuario y su posición es determinada por el procedimiento en función del número de nudos interiores.
- **Nominal múltiple.** La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en categorías. No se conserva el orden de las categorías de la variable observada. Los puntos de categoría estarán en el centroide de los objetos para las categorías particulares. El término *Múltiple* indica que se obtienen diferentes conjuntos de cuantificaciones para cada dimensión.
- **Ordinal.** El orden de las categorías de la variable observada se conserva en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. La transformación resultante se ajusta mejor que la transformación de líneaSp ordinal pero la suavidad es menor.
- **Nominal.** La única información de la variable observada que se conserva en la variable escalada óptimamente es la agrupación de los objetos en categorías. No se conserva el orden de las categorías de la variable observada. Los puntos de categoría estarán sobre una

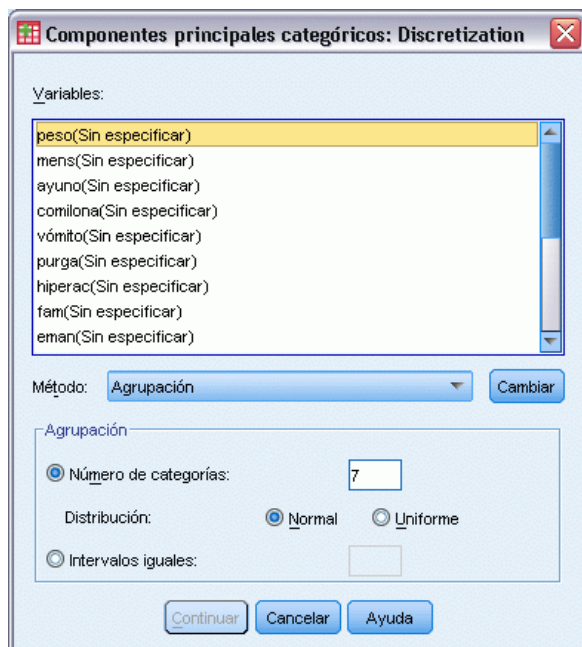
recta (vector) que pasa por el origen. La transformación resultante se ajusta mejor que la transformación de línea nominal pero la suavidad es menor.

- **Numérico.** Las categorías se tratan como que están ordenadas y espaciadas uniformemente (a nivel de intervalo). El orden de las categorías y la equidistancia entre los números de las categorías de la variable observada se conservan en la variable escalada óptimamente. Los puntos de categoría estarán sobre una recta (vector) que pasa por el origen. Cuando todas las variables están a nivel numérico, el análisis es análogo al análisis de componentes principales típico.

Análisis de componentes principales categórico: Discretización

El cuadro de diálogo Discretización permite seleccionar un método para recodificar las variables. Las variables con valores fraccionarios se agrupan en siete categorías (o en el número de valores diferentes de la variable si dicho número es inferior a siete) con una distribución aproximadamente normal, si no se especifica lo contrario. Las variables de cadena se convierten siempre en enteros positivos mediante la asignación de indicadores de categoría en función del orden alfanumérico ascendente. La discretización de las variables de cadena se aplica a estos enteros resultantes. Por defecto, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Figura 3-4
Cuadro de diálogo Discretización



Método. Seleccione entre agrupación, asignación de rangos y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.

- **Asignación de rangos.** La variable se discretiza mediante la asignación de rangos a los casos.
- **Multipliación.** Los valores actuales de la variable se tipifican, multiplican por 10, redondean y se les suma una constante de manera que el menor valor discretizado sea 1.

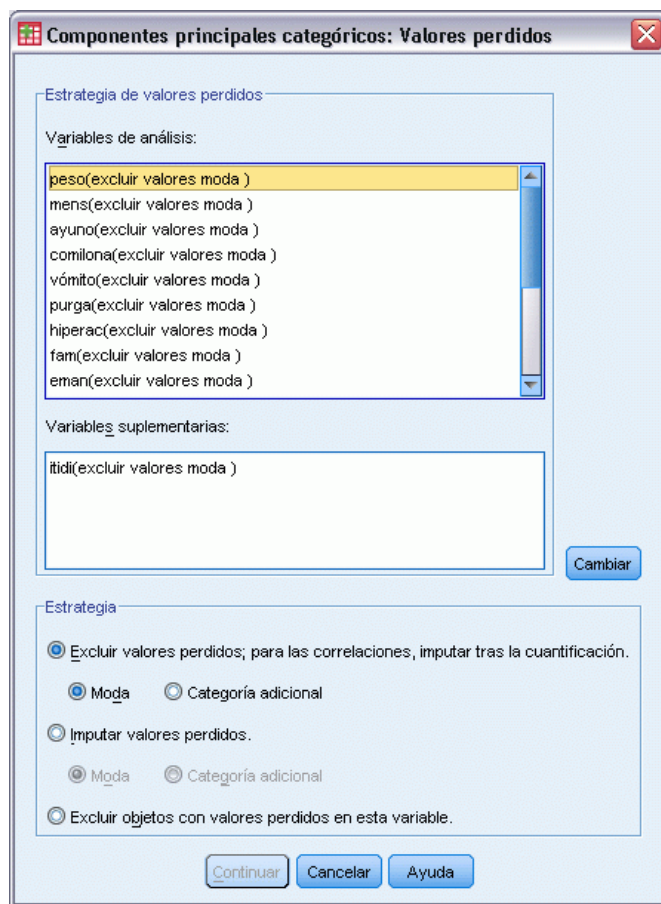
Agrupación. Se encuentran disponibles las siguientes opciones cuando se discretizan variables por agrupación:

- **Número de categorías.** Especifique un número de categorías y si los valores de la variable deben seguir una distribución aproximadamente normal o uniforme en dichas categorías.
- **Intervalos iguales.** Las variables se recodifican en las categorías definidas por dichos intervalos de igual tamaño. Se debe especificar la longitud de los intervalos.

Análisis de componentes principales categórico: Valores perdidos

El cuadro de diálogo Valores perdidos permite seleccionar la estrategia para el tratamiento de los valores perdidos en las variables de análisis y las suplementarias.

Figura 3-5
Cuadro de diálogo Valores perdidos



Estrategia. Seleccione excluir los valores perdidos (tratamiento pasivo), imputar los valores perdidos (tratamiento activo) o excluir objetos con valores perdidos (eliminación por lista).

- **Excluir valores perdidos; para las correlaciones, imputar tras la cuantificación.** Los objetos con valores perdidos en la variable seleccionada no contribuyen en el análisis de esta variable. Si a todas las variables se les aplica tratamiento pasivo, los objetos con valores perdidos en todas las variables se tratarán como suplementarios. Si se especifican correlaciones en el cuadro de diálogo Resultados, tras el análisis, los valores perdidos se imputarán con la categoría más frecuente, o moda, de la variable para las correlaciones de las variables originales. Para las correlaciones de la variable escalada óptimamente, se puede seleccionar el método de imputación. Seleccione Moda para reemplazar los valores perdidos por la moda de la variable escalada óptimamente. Seleccione Categoría adicional para reemplazar los valores perdidos por la cuantificación de una categoría adicional. Esto implica que los objetos con un valor perdido en esta variable se consideran que pertenecen a la misma categoría (la adicional).
- **Imputar valores perdidos.** Los objetos con valores perdidos en la variable seleccionada tendrán dichos valores imputados. Se puede seleccionar el método de imputación. Seleccione Moda para reemplazar los valores perdidos por la categoría más frecuente. Cuando existen varias modas, se utiliza la que tiene el indicador de categoría más pequeño. Seleccione Categoría adicional para reemplazar los valores perdidos con la misma cuantificación de una categoría adicional. Esto implica que los objetos con un valor perdido en esta variable se consideran que pertenecen a la misma categoría (la adicional).
- **Excluir objetos con valores perdidos en esta variable.** Los objetos con valores perdidos en la variable seleccionada se excluyen del análisis. Esta estrategia no está disponible para las variables suplementarias.

Análisis de componentes principales categórico: Opciones

El cuadro de diálogo de opciones permite seleccionar la configuración inicial, especificar los criterios de iteración y convergencia, seleccionar un método de normalización, elegir el método para etiquetar los gráficos y especificar objetos suplementarios.

Figura 3-6
Cuadro de diálogo Opciones

Objetos suplementarios. Especifique el número de caso del objeto, o bien los números de caso primero y último de un rango de objetos que desee convertir en suplementario y, a continuación, pulse en Añadir. Continúe hasta que haya especificado todos los objetos suplementarios. Si se especifica un objeto como suplementario, se ignorarán las ponderaciones de caso para dicho objeto.

Método de normalización. Se puede especificar una de las cinco opciones para normalizar las puntuaciones de objeto y las variables. Sólo se puede utilizar un método de normalización en un análisis dado.

- **Principal por variable.** Esta opción optimiza la asociación entre las variables. Las coordenadas de las variables en el espacio de los objetos son las saturaciones en componentes (las correlaciones con componentes principales, como son las dimensiones y las puntuaciones de los objetos). Esta opción es útil cuando el interés principal está en la correlación entre las variables.
- **Principal por objeto.** Esta opción optimiza las distancias entre los objetos. Esta opción es útil cuando el interés principal está en las diferencias y similitudes entre los objetos.
- **Simétrico.** Se utiliza esta opción de normalización si el interés principal está en la relación entre objetos y variables.

- **Independiente.** Se utiliza esta opción de normalización si se desea examinar por separado las distancias entre los objetos y las correlaciones entre las variables.
- **Personalizado.** Se puede especificar cualquier valor real en el intervalo cerrado $[-1, 1]$. Un valor 1 es igual al método Principal por objeto, un valor 0 es igual al método Simétrico y un valor -1 es igual al método Principal por variable. Si se especifica un valor mayor que -1 y menor que 1, se puede distribuir el autovalor entre los objetos y las variables. Este método es útil para generar diagramas de dispersión biespaciales y trispaciales a medida.

Criterios. Se puede especificar el número máximo de iteraciones que el procedimiento puede realizar durante los cálculos. También puede seleccionar un valor para el criterio de convergencia. El algoritmo detiene la iteración si la diferencia del ajuste total entre la dos últimas iteraciones es menor que el valor de convergencia o si se ha alcanzado el número máximo de iteraciones.

Etiquetar gráficos con. Permite especificar si se utilizarán en los gráficos las etiquetas de variable y las etiquetas de valor o los nombres de variable y los valores. También se puede especificar una longitud máxima para las etiquetas.

Dimensiones del gráfico. Permite controlar las dimensiones que se muestran en los resultados.

- **Muestra todas las dimensiones de la solución.** Todas las dimensiones de la solución se muestran en un diagrama de dispersión matricial.
- **Restringe el número de dimensiones** Las dimensiones mostradas se restringen a los pares representados. Si restringe las dimensiones, deberá seleccionar las dimensiones menor y mayor que se van a representar. La dimensión menor puede variar desde 1 hasta el número de dimensiones de la solución menos 1 y se representa respecto a las dimensiones mayores. El valor de la dimensión mayor puede oscilar variar desde 2 hasta el número de dimensiones de la solución e indica la dimensión mayor que se utilizará al representar los pares de dimensiones. Esta especificación se aplica a todos los gráficos multidimensionales solicitados.

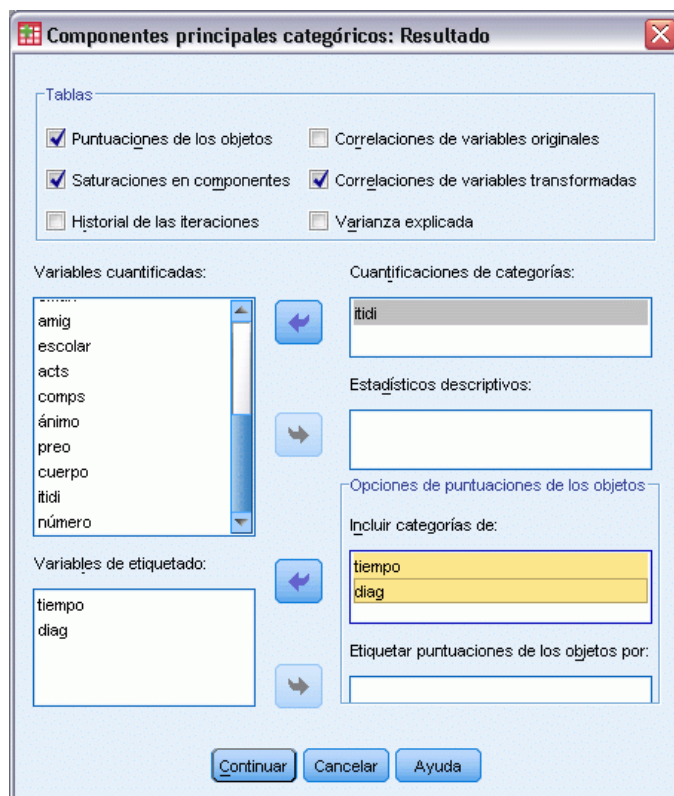
Configuración. Se pueden leer datos de un archivo que contenga las coordenadas de una configuración. La primera variable del archivo deberá contener las coordenadas para la primera dimensión, la segunda variable las coordenadas para la segunda dimensión, y así sucesivamente.

- **Inicial.** La configuración del archivo especificado se utilizará como el punto inicial del análisis.
- **Fija.** La configuración del archivo especificado se utilizará para ajustar las variables. Las variables que se ajustan se deben seleccionar como variables de análisis, pero al ser la configuración fija, se tratan como variables suplementarias (de manera que no es necesario seleccionarlas como variables suplementarias).

Análisis de componentes principales categórico: Resultados

El cuadro de diálogo Resultados permite generar tablas para las puntuaciones de los objetos, las saturaciones en componentes, el historial de iteraciones, las correlaciones de las variables originales y de las transformadas, la varianza explicada por variable y por dimensión, las cuantificaciones de las categorías para las variables seleccionadas y estadísticos descriptivos para las variables seleccionadas.

Figura 3-7
Cuadro de diálogo Resultados



Puntuaciones de los objetos. Muestra las puntuaciones de los objetos y tiene las siguientes opciones:

- **Incluir categorías de.** Muestra los indicadores de las categorías de las variables de análisis seleccionadas.
- **Etiquetar puntuaciones de los objetos por.** De la lista de variables especificadas como variables de etiquetado, se puede seleccionar una para etiquetar los objetos.

Saturaciones en componentes. Muestra las saturaciones en componentes para todas las variables que no recibieron niveles de escalamiento nominal múltiple.

Historial de iteraciones. En cada iteración, se muestra la varianza explicada, la pérdida y el incremento en la varianza explicada.

Correlaciones de variables originales. Muestra la matriz de correlaciones de las variables originales y los autovalores de dicha matriz.

Correlaciones de variables transformadas. Muestra la matriz de correlaciones de las variables transformadas (mediante escalamiento óptimo) y los autovalores de dicha matriz.

Varianza explicada. Muestra la cantidad de varianza explicada por las coordenadas de los centroides, las coordenadas de vectores y total (coordenadas de centroides y de vectores combinadas) por variable y por dimensión.

Cuantificaciones de categorías. Muestra las cuantificaciones de las categorías y las coordenadas para cada dimensión de las variables seleccionadas.

Estadísticos descriptivos. Muestra las frecuencias, el número de valores perdidos y la moda de las variables seleccionadas.

Análisis de componentes principales categórico: Guardar

El cuadro de diálogo Guardar permite guardar datos discretizados, puntuaciones de objeto, valores transformados y aproximaciones en un archivo de datos o conjunto de datos de IBM® SPSS® Statistics. También puede guardar estos valores transformados, puntuaciones de objetos y aproximaciones en el conjunto de datos activo.

- Los conjuntos de datos están disponibles durante la sesión actual, pero no así en las sesiones posteriores, a menos que los haya guardado explícitamente como archivos de datos. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.
- Los nombres de archivo o de conjunto de datos deben ser distintos para cada tipo de dato guardado.
- Si guarda puntuaciones de objeto o valores transformados en el conjunto de datos activo, podrá especificar el número de dimensiones nominales múltiples.

Figura 3-8
Cuadro de diálogo Guardar

Componentes principales categóricos: Guardar

Datos discretizados

- Crear datos discretizados
- Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
- Escribir un nuevo archivo de datos

Variables transformadas

- Guardar en el conjunto de datos activo
- Crear variables
- Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
- Escribir un nuevo archivo de datos

Puntuaciones de los objetos

- Guardar en el conjunto de datos activo
- Crear puntuaciones de los objetos
- Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
- Escribir un nuevo archivo de datos

Aproximaciones

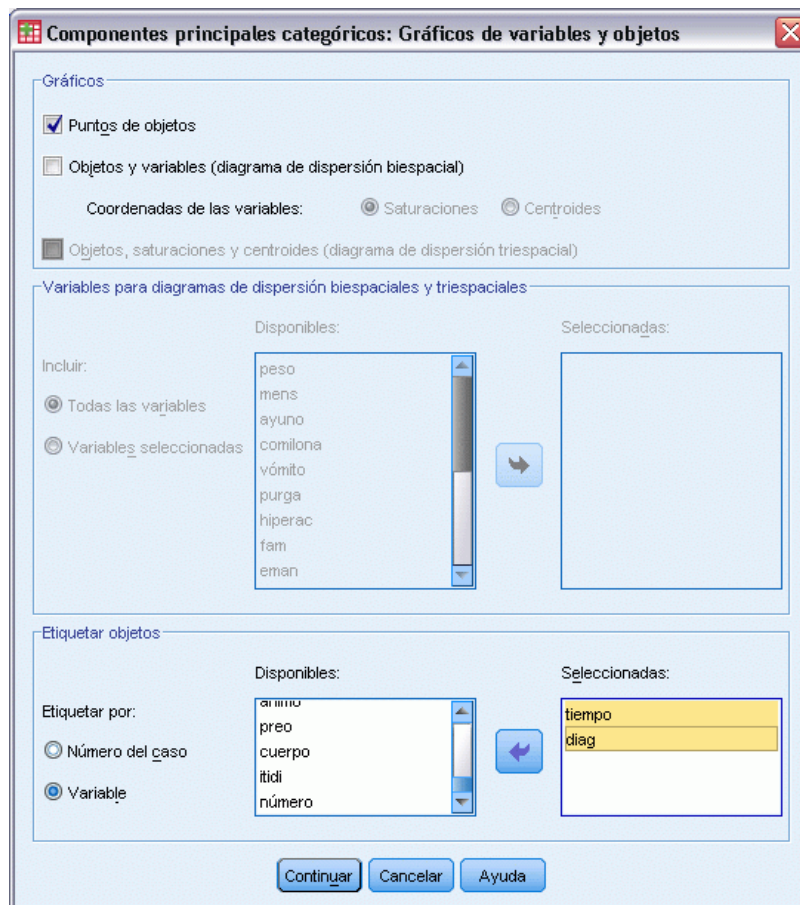
- Guardar en conjunto de datos activo
- Crear aproximaciones
- Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
- Escribir un nuevo archivo de datos

Dimensiones nominales múltiples: Todo Primero:

Análisis de componentes principales categórico: Gráficos de objetos

El cuadro de diálogo Gráficos de variables y objetos permite especificar los tipos de gráficos deseados y las variables para las que se generarán los gráficos.

Figura 3-9
Gráficos de variables y objetos



Puntos de objetos. Se muestra un gráfico de los puntos de objetos.

Objetos y variables (gráfico biespacial). Los puntos de objetos se representan con la selección realizada de las coordenadas de las variables (saturaciones en componentes o centroides de las variables).

Objetos, saturaciones y centroides (gráfico triespacial). Los puntos de objetos se representan con los centroides de las variables con un nivel de escalamiento nominal múltiple y las saturaciones en componentes de las otras variables.

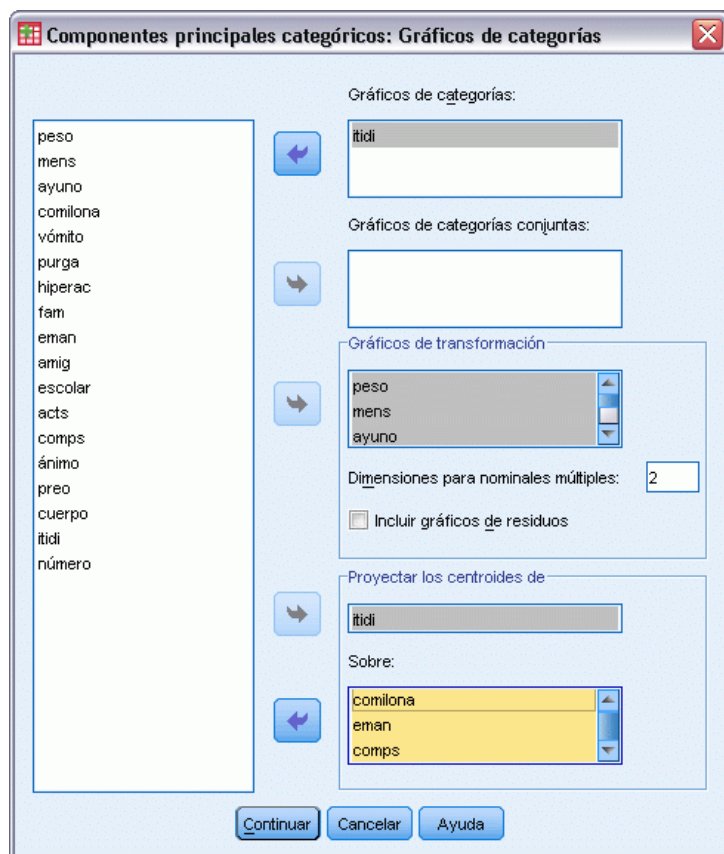
Variables para gráficos biespaciales y triespaciales. Puede utilizar todas las variables para los gráficos de dispersión biespacial y triespacial o seleccionar un subconjunto.

Etiquetar objetos. Se puede elegir que los objetos se etiqueten con las categorías de las variables seleccionadas (se pueden seleccionar entre los valores del indicador de categoría o las etiquetas de valor, en el cuadro de diálogo Opciones) o con sus números de caso. Se genera un gráfico por cada variable si se selecciona Variable.

Análisis de componentes principales categórico: Gráficos de categorías

El cuadro de diálogo Gráficos de categorías permite especificar los tipos de gráficos deseados y las variables para las que se generarán los gráficos.

Figura 3-10
Cuadro de diálogo Gráficos de categorías



Gráficos de categorías. Para cada variable seleccionada, se representa un gráfico de las coordenadas de vector y del centroide. Para las variables con nivel de escalamiento nominal múltiple, las categorías están sobre los centroides de los objetos para las categorías particulares. Para todos los demás niveles de escalamiento, las categorías están sobre un vector que pasa por el origen.

Gráficos de categorías conjuntas. Este es un único gráfico con el centroide y las coordenadas de vector de cada variable seleccionada.

Gráficos de transformación. Muestra un gráfico de las cuantificaciones de las categorías óptimas en oposición a los indicadores de las categorías. Se puede especificar el número de dimensiones deseado para las variables con nivel de escalamiento nominal múltiple; se generará un gráfico para cada dimensión. También se puede seleccionar si se muestran los gráficos de los residuos para cada variable seleccionada.

Proyectar los centroides de. Se puede seleccionar una variable y proyectar sus centroides sobre las variables seleccionadas. Las variables con niveles de escalamiento nominal múltiple no se pueden seleccionar para la proyección. Al solicitar este gráfico, aparece una tabla con las coordenadas de los centroides proyectados.

Análisis de componentes principales categórico: gráfico de saturaciones

El cuadro de diálogo Gráficos de saturaciones permite especificar las variables que se incluirán en el gráfico, así como si se van a incluir o no centroides en el gráfico.

Figura 3-11
Cuadro de diálogo Gráficos de saturaciones



Mostrar saturaciones en componentes. Si está seleccionado, se muestra un gráfico de las saturaciones en componentes.

Variables de saturaciones. Puede utilizar todas las variables para el gráfico de saturaciones en componentes o bien seleccionar un subconjunto.

Incluir centroides. Las variables con un nivel de escalamiento nominal múltiple no tienen saturaciones en componentes, pero se pueden incluir los centroides de dichas variables en el gráfico. Puede utilizar todas las variables nominales múltiples o bien seleccionar un subconjunto.

Funciones adicionales del comando CATPCA

Se puede personalizar el análisis de componentes principales categórico si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando de CATPCA resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar nombres de raíz para las variables transformadas, puntuaciones de objetos y aproximaciones al guardarlas en el conjunto de datos activo (con el subcomando `SAVE`).
- Especificar una longitud máxima de las etiquetas para cada gráfico por separado (con el subcomando `PLOT`).
- Especificar una lista de variables distinta para los gráficos de residuos (con el subcomando `PLOT`).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Análisis de correlación canónica no lineal (OVERALS)

El análisis de correlación canónica no lineal coincide con el análisis de correlación canónica categórico mediante escalamiento óptimo. El propósito de este procedimiento es determinar la similitud entre los conjuntos de variables categóricas. El análisis de correlación canónica no lineal se conoce también por el acrónimo OVERALS.

El análisis de correlación canónica típico es una extensión de la regresión múltiple, en la que el segundo conjunto no contiene una única variable de respuesta, sino varias. El objetivo es explicar el máximo posible de la varianza sobre las relaciones existentes entre dos conjuntos de variables numéricas en un espacio de pocas dimensiones. Inicialmente, las variables de cada conjunto se combinan linealmente de forma que las combinaciones lineales tengan una correlación máxima entre sí. Una vez dadas estas combinaciones, se establece que las combinaciones lineales subsiguientes no estén correlacionadas con las combinaciones anteriores y que también tengan la mayor correlación posible.

El enfoque de escalamiento óptimo expande el análisis estándar de tres formas decisivas. Primera: OVERALS permite más de dos conjuntos de variables. Segunda: las variables se pueden escalar como nominales, ordinales o numéricas. Como resultado, se pueden analizar relaciones no lineales entre las variables. Finalmente, en lugar de maximizar las correlaciones entre los conjuntos de variables, los conjuntos se comparan con un conjunto de compromiso desconocido definido por las puntuaciones de los objetos.

Ejemplo. El análisis de correlación canónica categórica mediante escalamiento óptimo se puede utilizar para representar gráficamente la relación entre un conjunto de variables que contienen la categoría laboral y los años de formación y otro conjunto de variables con la región de residencia y el sexo. Podemos encontrar que los años de formación y la región de residencia discriminan mejor que las variables restantes. También podemos encontrar que los años de formación es la variable que mejor discrimina en la primera dimensión.

Estadísticos y gráficos. Frecuencias, centroides, historial de iteraciones, puntuaciones de objeto, cuantificaciones de categoría, ponderaciones, saturaciones en las componentes, ajuste simple y múltiple, gráficos de las puntuaciones de objeto, gráficos de las coordenadas de categoría, gráficos de las saturaciones en las componentes, gráficos de los centroides de categoría y gráficos de transformación.

Datos. Utilice enteros para codificar las variables categóricas (nivel de escalamiento nominal u ordinal). Para minimizar los resultados, utilice enteros consecutivos, comenzando por el 1, para codificar cada variable. Las variables escaladas a nivel numérico no deben ser recodificadas en enteros consecutivos. Para minimizar los resultados, en cada variable escalada a nivel numérico, sustraiga el menor valor observado a todos los valores y súmele 1. Los valores fraccionarios se truncarán tras el decimal.

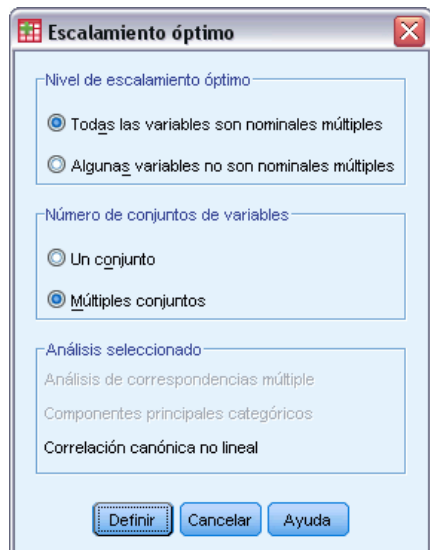
Supuestos. Las variables se pueden clasificar en dos o más conjuntos. Las variables del análisis se escalan como nominales múltiples, nominales simples, ordinales o numéricas. El número máximo de dimensiones que se utiliza en el procedimiento depende del nivel de escalamiento óptimo de las variables. Si todas las variables se especifican como ordinales, nominales simples o numéricas, el número máximo de dimensiones es el menor de los dos valores siguientes: el número de observaciones menos 1 o el número total de variables. Sin embargo, si sólo se definen dos conjuntos de variables, el número máximo de dimensiones es el número de variables en el conjunto más pequeño. Si algunas variables son nominales múltiples, el número máximo de dimensiones es el número total de categorías nominales múltiples más el número de variables nominales no múltiples menos el número de variables nominales múltiples. Por ejemplo, si el análisis incluye cinco variables, una de las cuales es nominal múltiple con cuatro categorías, el número máximo de dimensiones será $(4 + 4 - 1)$ o 7. Si se especifica un número mayor que el máximo, se utilizará el valor máximo.

Procedimientos relacionados. Si cada conjunto contiene una variable, el Análisis de correlación canónica no lineal será equivalente al Análisis de componentes principales mediante escalamiento óptimo. Si todas estas variables son nominales múltiples, el análisis coincidirá con el análisis de correspondencias múltiple. Si hay dos conjuntos de variables implicados y uno de ellos contiene sólo una variable, el análisis será idéntico a la Regresión categórica mediante escalamiento óptimo.

Para obtener un análisis de correlación canónica no lineal

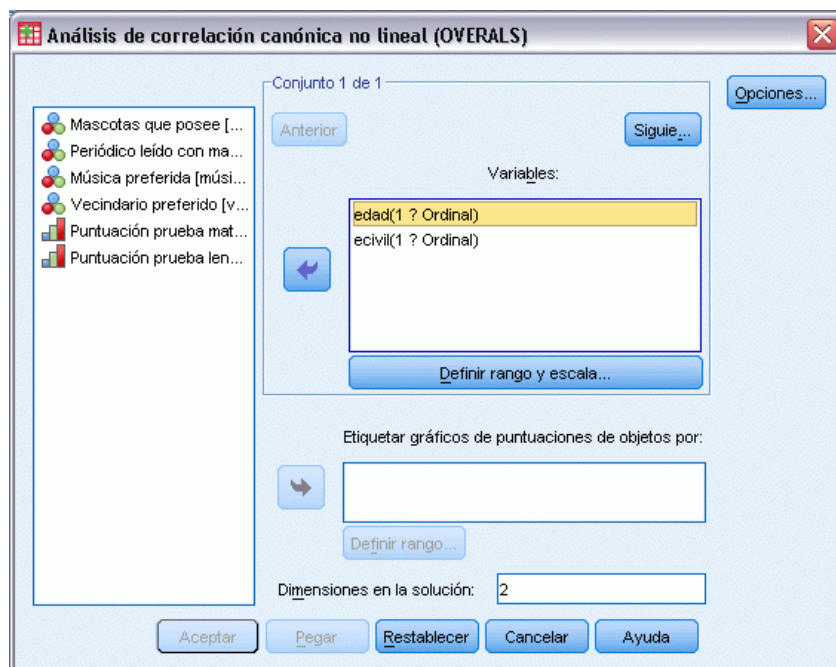
- ▶ Elija en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 4-1
Cuadro de diálogo Escalamiento óptimo



- ▶ Seleccione Todas las variables son nominales múltiples o alguna variable no es nominal múltiple.
- ▶ Seleccione Múltiples conjuntos.
- ▶ Pulse en Definir.

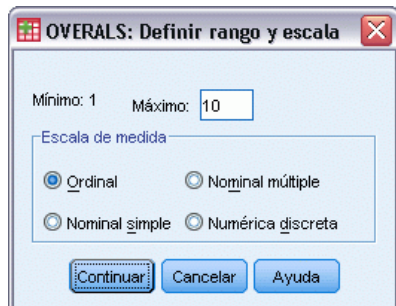
Figura 4-2
Cuadro de diálogo *Análisis de correlación canónica no lineal (OVERALS)*



- ▶ Defina al menos dos conjuntos de variables. Seleccione la variable o variables que desee incluir en el primer conjunto. Para desplazarse al siguiente conjunto, pulse en *Siguiente* y seleccione las variables que desee incluir en el segundo conjunto. Puede añadir los conjuntos adicionales que desee. Pulse en *Anterior* para volver al conjunto de variables definido anteriormente.
- ▶ Defina el rango de valores y la escala de medida (nivel de escalamiento óptimo) para cada variable seleccionada.
- ▶ Pulse en *Aceptar*.
- ▶ Si lo desea:
 - Seleccionar una o más variables para proporcionar etiquetas de punto en los gráficos de las puntuaciones de objeto. Cada variable genera un gráfico diferente, con los puntos etiquetados mediante los valores de dicha variable. Debe definir un rango para cada una de las variables de etiquetado de los gráficos. Cuando se usa el cuadro de diálogo, no se puede utilizar una misma variable en el análisis y como variable de etiquetado. Si se desea etiquetar el gráfico de las puntuaciones de objeto con una variable ya utilizada en el análisis, utilice la función *Calcular* (disponible en el menú *Transformar*) para crear una copia de dicha variable. Utilice la nueva variable para etiquetar el gráfico. Alternativamente, se puede utilizar la sintaxis de comandos.
 - Especifique el número de dimensiones que desea en la solución. En general, seleccione el menor número de dimensiones que necesite para explicar la mayor parte de la variación. Si el análisis incluye más de dos dimensiones, se generarán gráficos tridimensionales de las tres primeras dimensiones. Si se edita el gráfico, se pueden representar otras dimensiones.

Definir rango y escala

Figura 4-3
Cuadro de diálogo Definir rango y escala



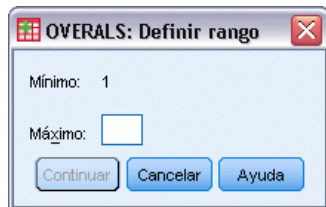
Debe definir un rango para cada variable. El valor máximo especificado debe ser un entero. En el análisis, se truncarán los valores de los datos fraccionarios. Se ignorará en el análisis cualquier valor de categoría que esté fuera del rango especificado. Para minimizar el resultado, utilice la función Recodificación automática (disponible en el menú Transformar) para crear categorías consecutivas, comenzando en el 1, para las variables tratadas como nominales u ordinales. No se recomienda la recodificación en enteros consecutivos de las variables que se escalan a nivel numérico. Para minimizar el resultado para variables tratadas como numéricas, para cada variable sustraiga el valor mínimo de cada valor y añádale 1.

Asimismo, se puede seleccionar el nivel de escalamiento que se utilizará para cuantificar cada variable.

- **Ordinal.** El orden de las categorías de la variable observada se conserva en la variable cuantificada.
- **Nominal simple.** En la variable cuantificada, los objetos que se encuentran en la misma categoría reciben la misma puntuación.
- **Nominal múltiple.** Las cuantificaciones pueden ser distintas para cada dimensión.
- **Numérica discreta.** Las categorías se tratan como si estuvieran ordenadas y espaciadas uniformemente. Las diferencias entre los números de las categorías y el orden de las categorías de la variable observada se conservan en la variable cuantificada.

Definir rango

Figura 4-4
Cuadro de diálogo Definir rango



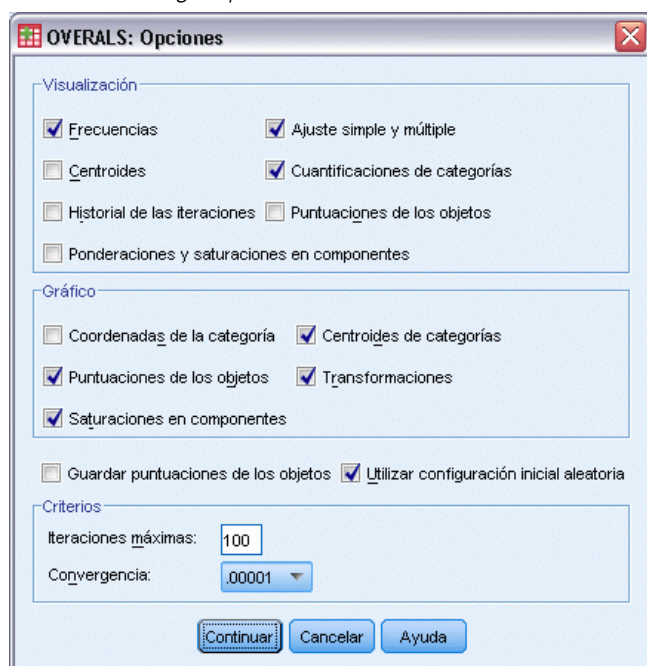
Debe definir un rango para cada variable. El valor máximo especificado debe ser un entero. En el análisis, se truncarán los valores de los datos fraccionarios. Se ignorará en el análisis cualquier valor de categoría que esté fuera del rango especificado. Para minimizar los resultados, utilice la función Recodificación automática (disponible en el menú Transformar) para crear categorías consecutivas, comenzando en el 1.

También debe definir un rango para cada una de las variables utilizadas para etiquetar los gráficos de las puntuaciones de objeto. Sin embargo, las etiquetas para las categorías con valores de los datos fuera del rango definido para la variable sí aparecen en los gráficos.

Análisis de correlación canónica no lineal: Opciones

El cuadro de diálogo Opciones permite seleccionar gráficos y estadísticos opcionales, guardar las puntuaciones de objeto como variables nuevas en conjunto de datos activo, especificar los criterios de iteración y convergencia y especificar una configuración inicial para el análisis.

Figura 4-5
Cuadro de diálogo Opciones



Mostrar. Los estadísticos disponibles incluyen las frecuencias marginales (los recuentos), los centroides, el historial de iteraciones, las ponderaciones y las saturaciones en las componentes, las cuantificaciones de las categorías, las puntuaciones de objeto y los estadísticos de ajuste simple y múltiple.

- **Centroides.** Las cuantificaciones de las categorías y los promedios proyectados y reales de las puntuaciones de los objetos para los objetos (los casos) incluidos en cada conjunto, para aquellos que pertenecen a la misma categoría de la variable.

- **Ponderaciones y saturaciones en componentes.** Los coeficientes de regresión en cada dimensión para toda variable cuantificada de un conjunto, donde las puntuaciones de los objetos se regresan sobre las variables cuantificadas; y la proyección de cada variable cuantificada en el espacio de los objetos. Proporciona una indicación de la contribución que cada variable aporta a la dimensión dentro de cada conjunto.
- **Ajustes simple y múltiple.** Medidas de bondad de ajuste simple y múltiple de las coordenadas de las categorías/cuantificaciones de las categorías, respecto a los objetos.
- **Cuantificaciones de categorías.** Valores de escalamiento óptimo asignados a las categorías de una variable.
- **Puntuaciones de los objetos.** Puntuación óptima que se asigna a un objeto (caso) en una dimensión particular.

Gráfico. Puede generar gráficos de las coordenadas de las categorías, las puntuaciones de objeto, las saturaciones en las componentes, los centroides de las categorías y las transformaciones.

Guardar puntuaciones de los objetos. Puede guardar las puntuaciones de los objetos como nuevas variables en el conjunto de datos activo. Las puntuaciones de objeto se guardan para el número de dimensiones especificadas en el cuadro de diálogo principal.

Utilizar configuración inicial aleatoria. Una configuración inicial aleatoria debe utilizarse en el caso de que algunas o todas las variables sean nominales simples. Si esta opción no se selecciona, se utiliza una configuración inicial anidada.

Criterios. Puede especificar el número máximo de iteraciones que el análisis de correlación canónica no lineal puede realizar durante los cálculos. También puede seleccionar un valor para el criterio de convergencia. El análisis detiene la iteración si la diferencia del ajuste total entre las dos últimas iteraciones es menor que el valor de convergencia o si se ha alcanzado el número máximo de iteraciones.

Funciones adicionales del comando OVERALS

Se puede personalizar el análisis de correlación canónica no lineal si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando OVERALS resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar los pares de dimensiones que se van a representar, en lugar de representar todas las dimensiones extraídas (mediante la palabra clave `NDIM` en el subcomando `PLOT`).
- Especificar el número de caracteres de la etiqueta del valor utilizado para etiquetar los puntos en los gráficos (mediante el subcomando `PLOT`).
- Designar más de cinco variables como variables de etiqueta para los gráficos de las puntuaciones de objeto (mediante el subcomando `PLOT`).
- Seleccionar las variables utilizadas en el análisis también como variables de etiqueta para los gráficos de las puntuaciones de objeto (mediante el subcomando `PLOT`).
- Seleccionar variables para proporcionar etiquetas de punto para el gráfico de las puntuaciones de cuantificación (mediante el subcomando `PLOT`).
- Especificar el número de casos que se van a incluir en el análisis si no desea utilizar todos los casos existentes en el conjunto de datos activo (mediante el subcomando `NOOBSERVATIONS`).

- Especificar nombres de raíz para las variables creadas al guardar las puntuaciones de objeto (mediante el subcomando `SAVE`).
- Especificar el número de dimensiones que se van a guardar, en lugar de guardar todas las dimensiones extraídas (mediante el subcomando `SAVE`).
- Escribir las cuantificaciones de las categorías en un archivo matricial (utilizando el subcomando `MATRIX`).
- Generar gráficos de baja resolución más fáciles de leer que los gráficos de alta resolución habituales (mediante el comando `SET`).
- Generar gráficos de transformación y de los centroides sólo para las variables especificadas (mediante el subcomando `PLOT`).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Análisis de correspondencias

Uno de los objetivos del análisis de correspondencias es describir las relaciones existentes entre dos variables nominales, recogidas en una tabla de correspondencias, sobre un espacio de pocas dimensiones, mientras que al mismo tiempo se describen las relaciones entre las categorías de cada variable. Para cada variable, las distancias sobre un gráfico entre los puntos de categorías reflejan las relaciones entre las categorías, con las categorías similares representadas próximas unas a otras. La proyección de los puntos de una variable sobre el vector desde el origen hasta un punto de categoría de la otra variable describe la relación entre ambas variables.

El análisis de las tablas de contingencia a menudo incluye examinar los perfiles de fila y de columna, así como contrastar la independencia a través del estadístico de chi-cuadrado. Sin embargo, el número de perfiles puede ser bastante grande y la prueba de chi-cuadrado no revelará la estructura de la dependencia. El procedimiento Tablas de contingencia ofrece varias medidas y pruebas de asociación pero no puede representar gráficamente ninguna relación entre las variables.

El análisis factorial es una técnica típica para describir las relaciones existentes entre variables en un espacio de pocas dimensiones. Sin embargo, el análisis factorial requiere datos de intervalo y el número de observaciones debe ser cinco veces el número de variables. Por su parte, el análisis de correspondencias asume que las variables son nominales y permite describir las relaciones entre las categorías de cada variable, así como la relación entre las variables. Además, el análisis de correspondencias se puede utilizar para analizar cualquier tabla de medidas de correspondencia que sean positivas.

Ejemplo. El análisis de correspondencias se puede utilizar para representar gráficamente la relación entre la categoría laboral y el hábito de fumar. Observará que, en relación al tabaco, el comportamiento de los subdirectores difiere del de las secretarías, pero el comportamiento de éstas no difiere del de los directores. Asimismo observará, que el tabaquismo pronunciado se encuentra asociado a los subdirectores, mientras que el tabaquismo leve se encuentra asociado a las secretarías.

Estadísticos y gráficos. Medidas de correspondencia, perfiles de fila y de columna, valores propios, puntuaciones de fila y de columna, inercia, masa, estadísticos de confianza para las puntuaciones de fila y de columna, estadísticos de confianza para los valores propios, gráficos de transformación, gráficos de los puntos de fila, gráficos de los puntos de columna y diagramas de dispersión biespaciales.

Datos. Las variables categóricas que se van a analizar se encuentran escaladas a nivel nominal. Para los datos agregados o para una medida de correspondencia distinta de las frecuencias, utilice una variable de ponderación con valores de similitud positivos. De manera alternativa, para datos tabulares, utilice la sintaxis para leer la tabla.

Supuestos. El máximo número de dimensiones utilizado en el procedimiento depende del número de categorías activas de fila y de columna y del número de restricciones de igualdad. Si no se utilizan criterios de igualdad y todas las categorías son activas, la dimensionalidad máxima es igual al número de categorías de la variable con menos categorías menos uno. Por ejemplo, si una variable dispone de cinco categorías y la otra de cuatro, el número máximo de dimensiones es

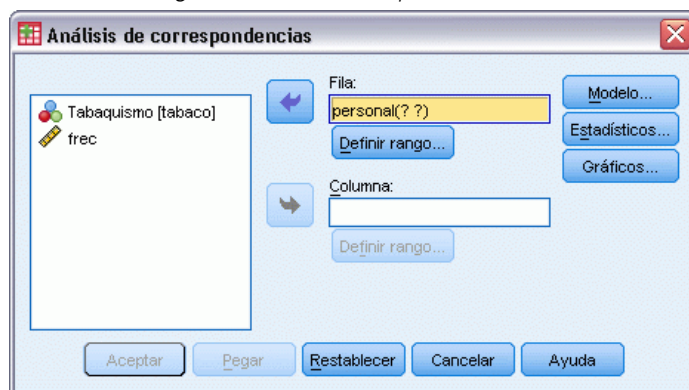
tres. Las categorías suplementarias no son activas. Por ejemplo, si una variable dispone de cinco categorías, dos de las cuales son suplementarias, y la otra variable dispone de cuatro categorías, el número máximo de dimensiones es dos. Considere todos los conjuntos de categorías con restricción de igualdad como una única categoría. Por ejemplo, si una variable dispone de cinco categorías, tres de las cuales tienen restricción de igualdad, dicha variable se debe tratar como si tuviera tres categorías en el momento de calcular la dimensionalidad máxima. Dos de las categorías no tienen restricción y la tercera corresponde a las tres categorías restringidas. Si se especifica un número de dimensiones superior al máximo, se utilizará el valor máximo.

Procedimientos relacionados. Si hay implicadas más de dos variables, utilice el análisis de correspondencias múltiple. Si se deben escalar las variables de forma ordinal, utilice el análisis de componentes principales categórico.

Para obtener un análisis de correspondencias

- ▶ Elija en los menús:
Analizar > Reducción de dimensiones > Análisis de correspondencias...

Figura 5-1
Cuadro de diálogo Análisis de correspondencias



- ▶ Seleccione una variable de filas.
- ▶ Seleccione una variable de columnas.
- ▶ Defina los rangos para las variables.
- ▶ Pulse en Aceptar.

Definir rango de filas en Análisis de correspondencias

Debe definir un rango para la variable de filas. Los valores mínimo y máximo especificados deben ser números enteros. En el análisis, se truncarán los valores de los datos fraccionarios. Se ignorará en el análisis cualquier valor de categoría que esté fuera del rango especificado.

Figura 5-2
Cuadro de diálogo Definir rango de filas

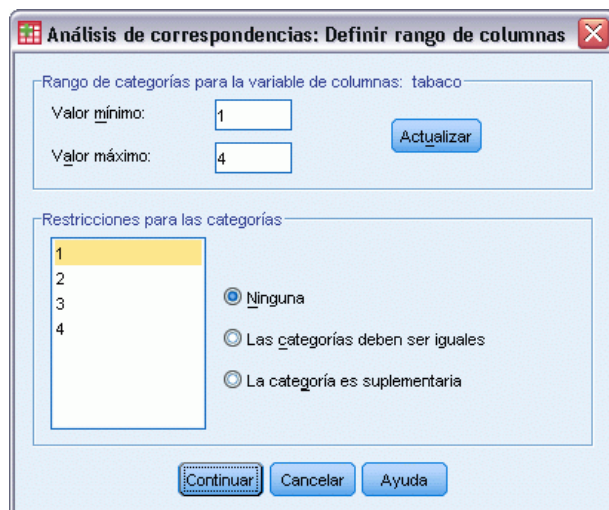
Inicialmente, todas las variables estarán sin restringir y activas. Puede restringir las categorías de fila para igualarlas a otras categorías de fila o puede definir cualquier categoría de fila como suplementaria.

- **Las categorías deben ser iguales.** Las puntuaciones de las categorías deben ser iguales. Utilice las restricciones de igualdad si el orden obtenido para las categorías no es el deseado o si no se corresponde con lo intuitivo. El máximo número de categorías de fila que se puede restringir para que sean consideradas iguales es el número total de categorías de fila activas menos 1. Utilice la sintaxis para imponer restricciones de igualdad a diferentes conjuntos de categorías. Por ejemplo, utilice la sintaxis para imponer la restricción de que sean consideradas iguales las categorías 1 y 2 y, por otra parte, que sean consideradas iguales las categorías 3 y 4.
- **La categoría es suplementaria.** Las categorías suplementarias no influyen en el análisis pero se representan en el espacio definido por las categorías activas. Las categorías suplementarias no juegan ningún papel en la definición de las dimensiones. El número máximo de categorías de fila suplementarias es el número total de categorías de fila menos 2.

Definir rango de columnas en Análisis de correspondencias

Debe definir un rango para la variable de columnas. Los valores mínimo y máximo especificados deben ser números enteros. En el análisis, se truncarán los valores de los datos fraccionarios. Se ignorará en el análisis cualquier valor de categoría que esté fuera del rango especificado.

Figura 5-3
Cuadro de diálogo Definir rango de columnas



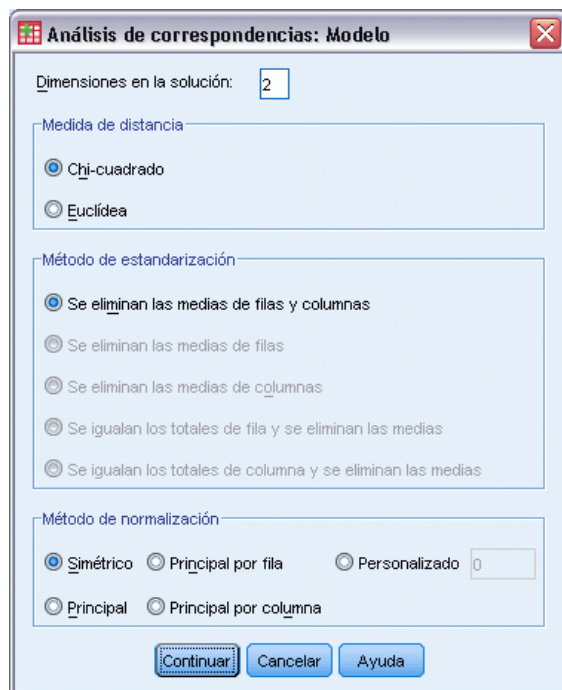
Inicialmente, todas las variables estarán sin restringir y activas. Puede restringir las categorías de columna para igualarlas a otras categorías de columna o puede definir cualquier categoría de columna como suplementaria.

- **Las categorías deben ser iguales.** Las puntuaciones de las categorías deben ser iguales. Utilice las restricciones de igualdad si el orden obtenido para las categorías no es el deseado o si no se corresponde con lo intuitivo. El número máximo de categorías de columna al que se puede imponer la restricción de igualdad es el número total de categorías de columna activas menos 1. Utilice la sintaxis para imponer restricciones de igualdad a diferentes conjuntos de categorías. Por ejemplo, utilice la sintaxis para imponer la restricción de que sean consideradas iguales las categorías 1 y 2 y, por otra parte, que sean consideradas iguales las categorías 3 y 4.
- **La categoría es suplementaria.** Las categorías suplementarias no influyen en el análisis pero se representan en el espacio definido por las categorías activas. Las categorías suplementarias no juegan ningún papel en la definición de las dimensiones. El número máximo de categorías de columna suplementarias es el número total de categorías de columna menos 2.

Análisis de correspondencias: Modelo

El cuadro de diálogo Modelo permite especificar el número de dimensiones, la medida de distancia, el método de estandarización y el método de normalización.

Figura 5-4
Cuadro de diálogo Modelo



Dimensiones en la solución. Especifique el número de dimensiones. En general, seleccione el menor número de dimensiones que necesite para explicar la mayor parte de la variación. El máximo número de dimensiones depende del número de categorías activas utilizadas en el análisis y de las restricciones de igualdad. El máximo número de dimensiones es el menor entre:

- El número de categorías de fila activas menos el número de categorías de fila con restricción de igualdad, más el número de conjuntos de categorías de fila que se han restringido.
- El número de categorías de columna activas menos el número de categorías de columna con restricción de igualdad, más el número de conjuntos de categorías de columna que se han restringido.

Medida de distancia. Puede seleccionar la medida de distancia entre las filas y columnas de la tabla de correspondencias. Seleccione una de las siguientes opciones:

- **Chi-cuadrado.** Utiliza una distancia ponderada entre los perfiles, donde la ponderación es la masa de las filas o de las columnas. Esta distancia es necesaria para el análisis de correspondencias típico.
- **Euclídea.** Utiliza la raíz cuadrada de la suma de los cuadrados de las diferencias entre los pares de filas y entre los pares de columnas.

Método de estandarización. Seleccione una de las siguientes opciones:

- **Se eliminan las medias de filas y columnas.** Se centran las filas y las columnas. Este método es necesario para el análisis de correspondencias típico.
- **Se eliminan las medias de filas.** Sólo se centran las filas.
- **Se eliminan las medias de columnas.** Sólo se centran las columnas.

- **Se igualan los totales de fila y se eliminan las medias.** Antes de centrar las filas, se igualan los márgenes de fila.
- **Se igualan los totales de columna y se eliminan las medias.** Antes de centrar las columnas, se igualan los márgenes de columna.

Método de normalización. Seleccione una de las siguientes opciones:

- **Simétrico.** Para cada dimensión, las puntuaciones de fila son la media ponderada de las puntuaciones de columna divididas por el valor propio coincidente y las puntuaciones de columna son la media ponderada de las puntuaciones de fila divididas por el valor propio coincidente. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de las dos variables.
- **Principal.** Las distancias entre los puntos de fila y los puntos de columna son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Utilice este método si desea examinar las diferencias entre las categorías de una o de ambas variables en lugar de las diferencias entre las dos variables.
- **Principal por fila.** Las distancias entre los puntos de fila son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Las puntuaciones de fila son la media ponderada de las puntuaciones de columna. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de la variable de filas.
- **Principal por columna.** Las distancias entre los puntos de columna son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Las puntuaciones de columna son la media ponderada de las puntuaciones de fila. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de la variable de columnas.
- **Personalizado.** Debe especificar un valor entre -1 y 1 . El valor -1 corresponde a Principal por columna. El valor 1 corresponde a Principal por fila. El valor 0 corresponde a simétrico. Todos los demás valores dispersan la inercia entre las puntuaciones de columna y de fila en diferentes grados. Este método es útil para generar diagramas de dispersión biespaciales a medida.

Análisis de correspondencias: Estadísticos

El cuadro de diálogo Estadísticos permite especificar los resultados numéricos producidos.

Figura 5-5
Cuadro de diálogo Estadísticos

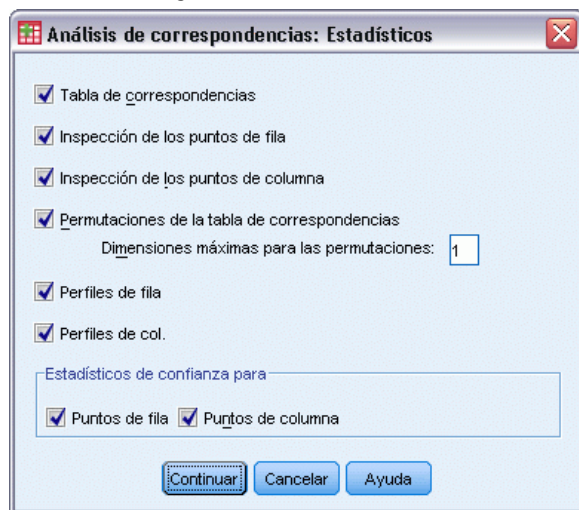


Tabla de correspondencias. Es la tabla de contingencia de las variables de entrada con los totales marginales de fila y columna.

Inspección de los puntos de fila. Para cada categoría de fila, las puntuaciones, la masa, la inercia, la contribución a la inercia de la dimensión y la contribución de la dimensión a la inercia del punto.

Inspección de los puntos de columna. Para cada categoría de columna, las puntuaciones, la masa, la inercia, la contribución a la inercia de la dimensión y la contribución de la dimensión a la inercia del punto.

Perfiles de fila. Para cada categoría de fila, la distribución a través de las categorías de la variable de columna.

Perfiles de col. Para cada categoría de columna, la distribución a través de las categorías de la variable de fila.

Permutaciones de la tabla de correspondencias. La tabla de correspondencias reorganizada de tal manera que las filas y las columnas estén en orden ascendente de acuerdo con las puntuaciones en la primera dimensión, Si lo desea, puede especificar el número de la dimensión máxima para el que se generarán las tablas permutadas. Se generará una tabla permutada para cada dimensión desde 1 hasta el número especificado.

Estadísticos de confianza para puntos de fila. Incluye la desviación típica y las correlaciones para todos los puntos de fila no suplementarios.

Estadísticos de confianza para puntos de columna. Incluye la desviación típica y las correlaciones para todos los puntos de columna no suplementarios.

Análisis de correspondencias: Gráficos

El cuadro de diálogo Gráficos permite especificar qué gráficos se van a generar.

Figura 5-6
Cuadro de diálogo Gráficos



Diagramas de dispersión. Produce una matriz de todos los gráficos por parejas de las dimensiones. Los diagramas de dispersión disponibles incluyen:

- **Diagrama de dispersión biespacial.** Produce una matriz de diagramas conjuntos de los puntos de fila y de columna. Si está seleccionada la normalización principal, el diagrama de dispersión biespacial no estará disponible.
- **Puntos de fila.** Produce una matriz de diagramas de los puntos de fila.
- **Puntos de columna.** Produce una matriz de diagramas de los puntos de columna.

Si lo desea, puede especificar el número de caracteres de etiqueta de valor que se va a utilizar al etiquetar los puntos. Este valor debe ser un entero no negativo menor o igual que 20.

Gráficos de línea. Produce un gráfico para cada dimensión de la variable seleccionada. Los gráficos de líneas disponibles incluyen:

- **Categorías de fila transformadas.** Produce un gráfico de los valores originales para las categorías de fila frente a las puntuaciones de fila correspondientes.
- **Categorías de columna transformadas.** Produce un gráfico de los valores originales para las categorías de columna frente a las puntuaciones de columna correspondientes.

Si lo desea, puede especificar el número de caracteres de etiqueta de valor que se va a utilizar al etiquetar los ejes de categorías. Este valor debe ser un entero no negativo menor o igual que 20.

Dimensiones del gráfico. Permite controlar las dimensiones que se muestran en los resultados.

- **Muestra todas las dimensiones de la solución.** Todas las dimensiones de la solución se muestran en un diagrama de dispersión matricial.
- **Restringe el número de dimensiones** Las dimensiones mostradas se restringen a los pares representados. Si restringe las dimensiones, deberá seleccionar las dimensiones menor y mayor que se van a representar. La dimensión menor puede variar desde 1 hasta el número de dimensiones de la solución menos 1 y se representa respecto a las dimensiones mayores. El valor de la dimensión mayor puede oscilar variar desde 2 hasta el número de dimensiones de la solución e indica la dimensión mayor que se utilizará al representar los pares de dimensiones. Esta especificación se aplica a todos los gráficos multidimensionales solicitados.

Funciones adicionales del comando CORRESPONDENCE

Se puede personalizar el análisis de correspondencias si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando CORRESPONDENCE resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar datos tabulares como entrada en lugar de utilizar datos por caso (mediante el subcomando TABLE = ALL).
- Especificar el número de caracteres de etiqueta de valor que se utilizan al etiquetar los puntos para cada tipo de diagrama de dispersión matricial o diagrama de dispersión biespacial matricial (mediante el subcomando PLOT).
- Especificar el número de caracteres de etiqueta de valor que se utilizan al etiquetar los puntos para cada tipo de gráfico de líneas (mediante el subcomando PLOT).
- Escribir una matriz de puntuaciones de fila y de columna en un archivo de datos matriciales (mediante el subcomando OUTFILE).
- Escribir una matriz de estadísticos de confianza (varianzas y covarianzas) para los valores propios y las puntuaciones en un archivo de datos matriciales (mediante el subcomando OUTFILE).
- Especificar varios conjuntos de categorías para igualar (mediante el subcomando EQUAL).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Análisis de correspondencias múltiple

El análisis de correspondencias múltiple cuantifica los datos nominales (categóricos) mediante la asignación de valores numéricos a los casos (objetos) y a las categorías, de manera que los objetos de la misma categoría estén cerca los unos de los otros y los objetos de categorías diferentes estén alejados los unos de los otros. Cada objeto se encuentra lo más cerca posible de los puntos de categoría para las categorías que se aplican a dicho objeto. De esta manera, las categorías dividen los objetos en subgrupos homogéneos. Las variables se consideran homogéneas cuando clasifican objetos de las mismas categorías en los mismos subgrupos.

Ejemplo. El análisis de correspondencias múltiple se puede utilizar para representar gráficamente la relación entre la categoría laboral, la clasificación étnica y el género. Puede que encontremos que la clasificación étnica y el género son capaces de discriminar entre las personas, pero no así la categoría laboral. También puede que encontremos que las categorías Latino y Afro-americano son similares entre sí.

Estadísticos y gráficos. Puntuaciones de objetos, medidas de discriminación, historial de iteraciones, correlaciones de las variables originales y transformadas, cuantificaciones de categorías, estadísticos descriptivos, gráficos de puntos de objetos, gráficos de dispersión biespacial, gráficos de categorías, gráficos de categorías conjuntas, gráficos de transformación y gráficos de medidas de discriminación.

Datos. Los valores de las variables de cadena se convierten en enteros positivos por orden alfabético ascendente. Los valores perdidos definidos por el usuario, los valores perdidos del sistema y los valores menores que 1 se consideran valores perdidos; se puede añadir una constante o recodificar las variables con valores inferiores a 1 para evitar que se pierdan los mismos.

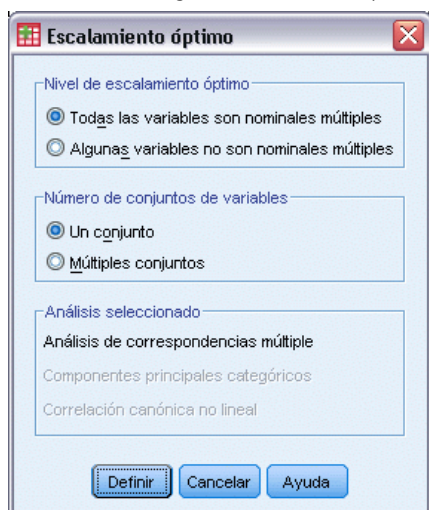
Supuestos. Todas las variables tienen un nivel de escalamiento nominal múltiple. Los datos deben contener al menos tres casos válidos. El análisis se basa en datos enteros positivos. La opción de discretización categorizará de forma automática una variable con valores fraccionarios, agrupando sus valores en categorías con una distribución casi normal y convertirá de forma automática los valores de las variables de cadena en enteros positivos. Se pueden especificar otros esquemas de discretización.

Procedimientos relacionados. Para dos variables, el análisis de correspondencias múltiple es análogo al análisis de correspondencias. Si piensa que las variables poseen propiedades ordinales o numéricas, se debe utilizar el análisis de componentes principales categórico. Si hay conjuntos de variables que son de interés, se debe utilizar el análisis de correlación canónica no lineal.

Para obtener un análisis de correspondencias múltiple

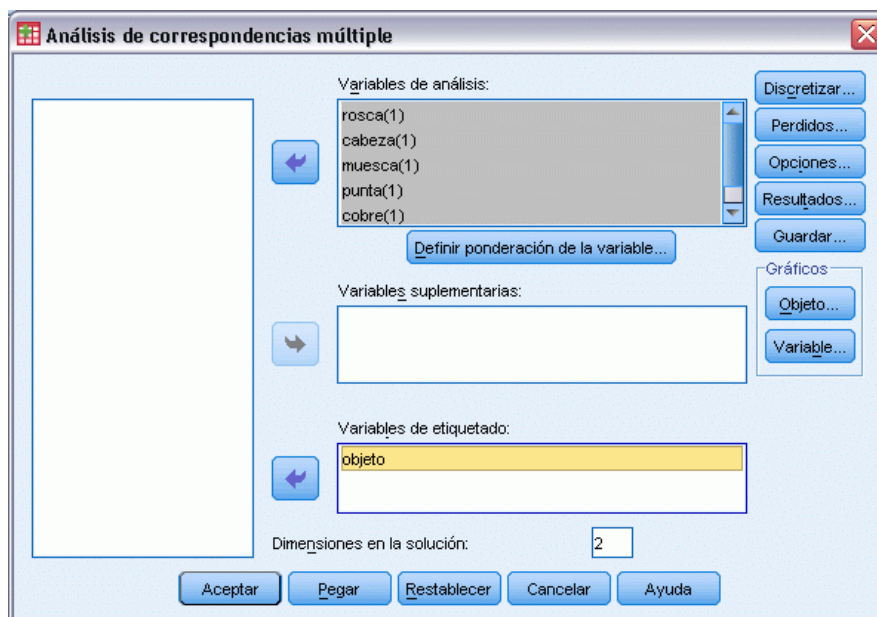
- ▶ Elija en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 6-1
Cuadro de diálogo Escalamiento óptimo



- ▶ Seleccione Todas las variables son nominales múltiples.
- ▶ Seleccione Un conjunto.
- ▶ Pulse en Definir.

Figura 6-2
Cuadro de diálogo Análisis de correspondencias múltiple



- ▶ Seleccione al menos dos variables de análisis y especifique el número de dimensiones en la solución.
- ▶ Pulse en Aceptar.

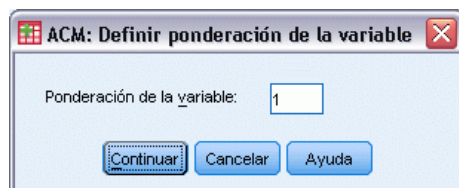
Si lo desea, puede especificar variables suplementarias, que serán ajustadas sobre la solución encontrada, o variables de etiqueta para los gráficos.

Definir ponderación de la variable en el análisis de correspondencias múltiple

Se puede definir la ponderación para las variables del análisis.

Figura 6-3

Cuadro de diálogo Definir ponderación de la variable

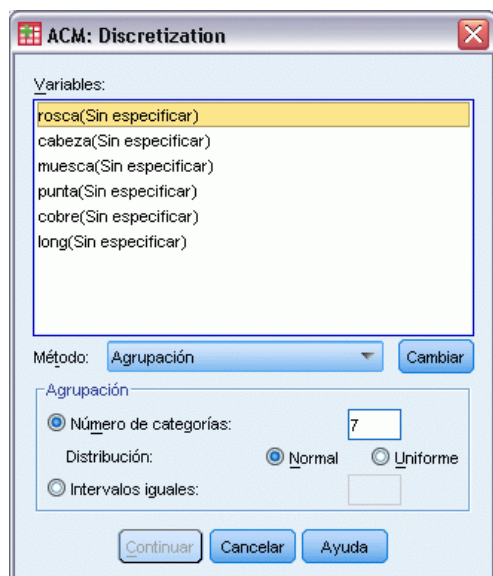


Ponderación de la variable. Se puede definir una ponderación para cada variable. El valor especificado debe ser un entero positivo. El valor por defecto es 1.

Discretización en el análisis de correspondencias múltiple

El cuadro de diálogo Discretización permite seleccionar un método para recodificar las variables. Las variables con valores fraccionarios se agrupan en siete categorías (o en el número de valores diferentes de la variable si dicho número es inferior a siete) con una distribución aproximadamente normal, si no se especifica lo contrario. Las variables de cadena se convierten siempre en enteros positivos mediante la asignación de indicadores de categoría en función del orden alfanumérico ascendente. La discretización de las variables de cadena se aplica a estos enteros resultantes. Por defecto, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Figura 6-4
Cuadro de diálogo Discretización



Método. Seleccione entre agrupación, asignación de rangos y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.
- **Asignación de rangos.** La variable se discretiza mediante la asignación de rangos a los casos.
- **Multiplicación.** Los valores actuales de la variable se tipifican, multiplican por 10, redondean y se les suma una constante de manera que el menor valor discretizado sea 1.

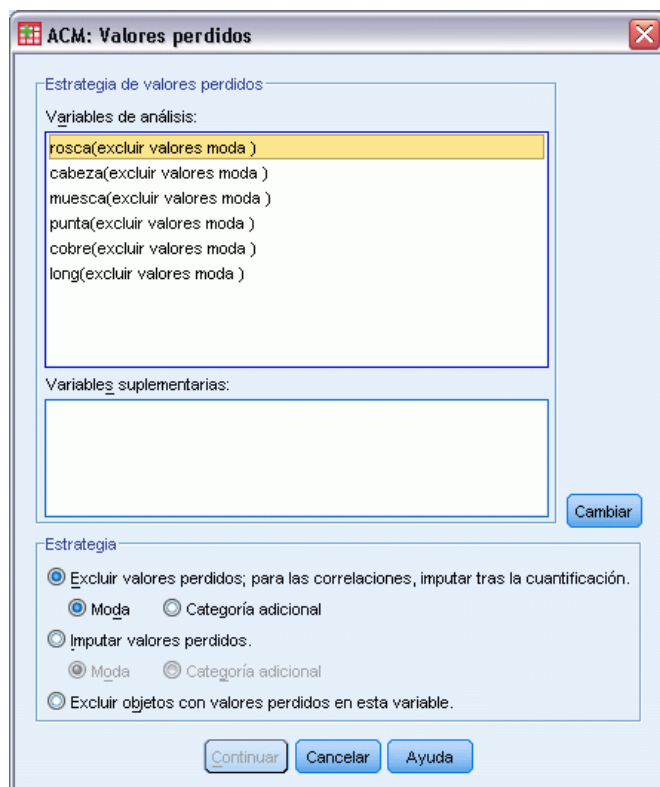
Agrupación. Se encuentran disponibles las siguientes opciones al discretizar variables por agrupación:

- **Número de categorías.** Especifique un número de categorías y si los valores de la variable deben seguir una distribución aproximadamente normal o uniforme en dichas categorías.
- **Intervalos iguales.** Las variables se recodifican en las categorías definidas por dichos intervalos de igual tamaño. Se debe especificar la longitud de los intervalos.

Valores perdidos en el análisis de correspondencias múltiple

El cuadro de diálogo Valores perdidos permite seleccionar la estrategia para el tratamiento de los valores perdidos en las variables de análisis y las suplementarias.

Figura 6-5
Cuadro de diálogo Valores perdidos



Estrategia de valores perdidos. Seleccione excluir los valores perdidos (tratamiento pasivo), imputar los valores perdidos (tratamiento activo) o excluir objetos con valores perdidos (eliminación por lista).

- **Excluir valores perdidos; para las correlaciones, imputar tras la cuantificación.** Los objetos con valores perdidos en la variable seleccionada no contribuyen en el análisis de esta variable. Si a todas las variables se les aplica tratamiento pasivo, los objetos con valores perdidos en todas las variables se tratarán como suplementarios. Si se especifican correlaciones en el cuadro de diálogo Resultados, tras el análisis, los valores perdidos se imputarán con la categoría más frecuente, o moda, de la variable para las correlaciones de las variables originales. Para las correlaciones de la variable escalada óptimamente, se puede seleccionar el método de imputación. Seleccione Moda para reemplazar los valores perdidos por la moda de la variable escalada óptimamente. Seleccione Categoría adicional para reemplazar los valores perdidos por la cuantificación de una categoría adicional. Esto implica que los objetos con un valor perdido en esta variable se consideran que pertenecen a la misma categoría (la adicional).
- **Imputar valores perdidos.** Los objetos con valores perdidos en la variable seleccionada tendrán dichos valores imputados. Se puede seleccionar el método de imputación. Seleccione Moda para reemplazar los valores perdidos por la categoría más frecuente. Cuando existen varias modas, se utiliza la que tiene el indicador de categoría más pequeño. Seleccione Categoría adicional para reemplazar los valores perdidos con la misma cuantificación de una categoría

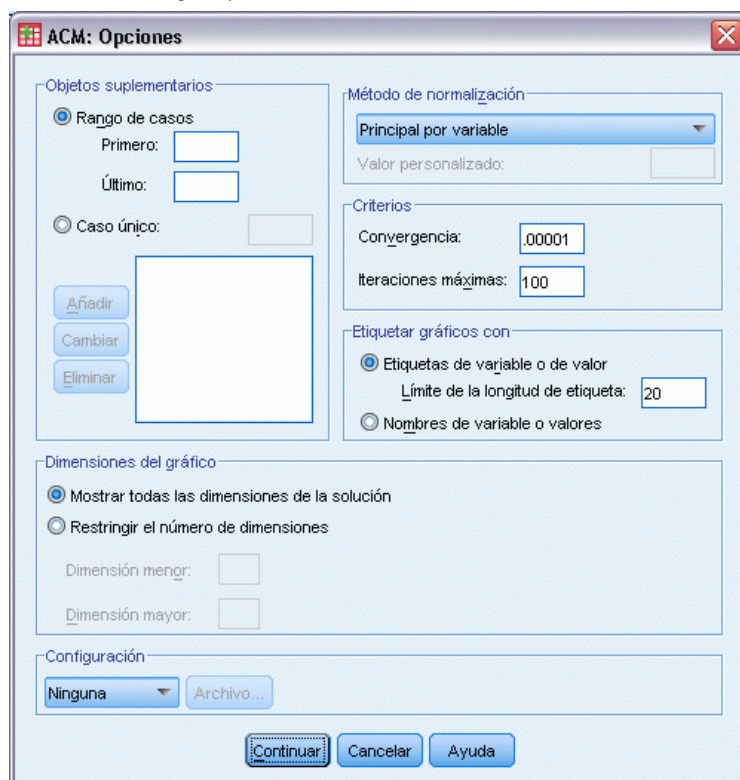
adicional. Esto implica que los objetos con un valor perdido en esta variable se consideran que pertenecen a la misma categoría (la adicional).

- **Excluir objetos con valores perdidos en esta variable.** Los objetos con valores perdidos en la variable seleccionada se excluyen del análisis. Esta estrategia no está disponible para las variables suplementarias.

Opciones de análisis de correspondencias múltiple

El cuadro de diálogo de opciones permite seleccionar la configuración inicial, especificar los criterios de iteración y convergencia, seleccionar un método de normalización, elegir el método para etiquetar los gráficos y especificar objetos suplementarios.

Figura 6-6
Cuadro de diálogo Opciones



Objetos suplementarios. Especifique el número de caso del objeto (o bien los números de caso primero y último de un rango de objetos) que desee convertir en suplementario y, a continuación, pulse en Añadir. Continúe hasta que haya especificado todos los objetos suplementarios. Si se especifica un objeto como suplementario, se ignorarán las ponderaciones de caso para dicho objeto.

Método de normalización. Se puede especificar una de las cinco opciones para normalizar las puntuaciones de objeto y las variables. Sólo se puede utilizar un método de normalización en un análisis dado.

- **Principal por variable.** Esta opción optimiza la asociación entre las variables. Las coordenadas de las variables en el espacio de los objetos son las saturaciones en componentes (las correlaciones con componentes principales, como son las dimensiones y las puntuaciones de los objetos). Esta opción es útil cuando el interés principal está en la correlación entre las variables.
- **Principal por objeto.** Esta opción optimiza las distancias entre los objetos. Esta opción es útil cuando el interés principal está en las diferencias y similitudes entre los objetos.
- **Simétrico.** Se utiliza esta opción de normalización si el interés principal está en la relación entre objetos y variables.
- **Independiente.** Se utiliza esta opción de normalización si se desea examinar por separado las distancias entre los objetos y las correlaciones entre las variables.
- **Personalizado.** Se puede especificar cualquier valor real en el intervalo cerrado $[-1, 1]$. Un valor 1 es igual al método Principal por objeto, un valor 0 es igual al método Simétrico y un valor -1 es igual al método Principal por variable. Si se especifica un valor mayor que -1 y menor que 1, se puede distribuir el autovalor entre los objetos y las variables. Este método es útil para generar diagramas de dispersión biespaciales y trispaciales a medida.

Criterios. Se puede especificar el número máximo de iteraciones que el procedimiento puede realizar durante los cálculos. También puede seleccionar un valor para el criterio de convergencia. El algoritmo detiene la iteración si la diferencia del ajuste total entre la dos últimas iteraciones es menor que el valor de convergencia o si se ha alcanzado el número máximo de iteraciones.

Etiquetar gráficos con. Permite especificar si se utilizarán en los gráficos las etiquetas de variable y las etiquetas de valor o los nombres de variable y los valores. También se puede especificar una longitud máxima para las etiquetas.

Dimensiones del gráfico. Permite controlar las dimensiones que se muestran en los resultados.

- **Muestra todas las dimensiones de la solución.** Todas las dimensiones de la solución se muestran en un diagrama de dispersión matricial.
- **Restringe el número de dimensiones** Las dimensiones mostradas se restringen a los pares representados. Si restringe las dimensiones, deberá seleccionar las dimensiones menor y mayor que se van a representar. La dimensión menor puede variar desde 1 hasta el número de dimensiones de la solución menos 1 y se representa respecto a las dimensiones mayores. El valor de la dimensión mayor puede oscilar variar desde 2 hasta el número de dimensiones de la solución e indica la dimensión mayor que se utilizará al representar los pares de dimensiones. Esta especificación se aplica a todos los gráficos multidimensionales solicitados.

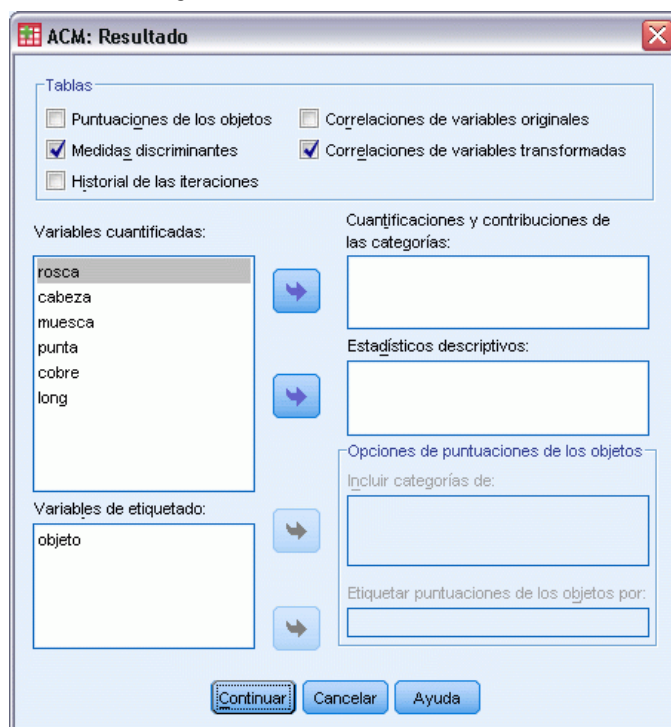
Configuración. Se pueden leer datos de un archivo que contenga las coordenadas de una configuración. La primera variable del archivo deberá contener las coordenadas para la primera dimensión, la segunda variable las coordenadas para la segunda dimensión, y así sucesivamente.

- **Inicial.** La configuración del archivo especificado se utilizará como el punto inicial del análisis.
- **Fija.** La configuración del archivo especificado se utilizará para ajustar las variables. Las variables que se ajustan se deben seleccionar como variables de análisis, pero, al ser la configuración fija, se tratan como variables suplementarias (de manera que no es necesario seleccionarlas como variables suplementarias).

Resultados del análisis de correspondencias múltiple

El cuadro de diálogo permite generar tablas para las puntuaciones de los objetos, las medidas de discriminación, el historial de iteraciones, las correlaciones de las variables originales y transformadas, las cuantificaciones de categorías para las variables seleccionadas y los estadísticos descriptivos de las variables seleccionadas.

Figura 6-7
Cuadro de diálogo Resultados



Puntuaciones de los objetos. Muestra las puntuaciones de los objetos (incluidas la masa, la inercia y las contribuciones) y tiene las siguientes opciones:

- **Incluir categorías de.** Muestra los indicadores de las categorías de las variables de análisis seleccionadas.
- **Etiquetar puntuaciones de los objetos por.** De la lista de variables especificadas como variables de etiquetado, se puede seleccionar una para etiquetar los objetos.

Medidas de discriminación. Muestra las medidas de discriminación por variable y por dimensión.

Historial de iteraciones. En cada iteración, se muestra la varianza explicada, la pérdida y el incremento en la varianza explicada.

Correlaciones de variables originales. Muestra la matriz de correlaciones de las variables originales y los autovalores de dicha matriz.

Correlaciones de variables transformadas. Muestra la matriz de correlaciones de las variables transformadas (mediante escalamiento óptimo) y los autovalores de dicha matriz.

Cuantificaciones de categorías y contribuciones. Muestra las cuantificaciones de las categorías (coordenadas), incluidas la masa, la inercia y las contribuciones) para cada dimensión de las variables seleccionadas.

Nota: las coordenadas y contribuciones (incluyendo masa e inercia) se muestran en capas diferentes de los resultados de la tabla pivote, mostrando las coordenadas por defecto. Para mostrar la contribuciones, active (pulse dos veces) la tabla y seleccione Contribuciones de la lista desplegable Capas.

Estadísticos descriptivos. Muestra las frecuencias, el número de valores perdidos y la moda de las variables seleccionadas.

Análisis de correspondencias múltiple: Guardar

El cuadro de diálogo Guardar permite guardar datos discretizados, puntuaciones de objeto, valores transformados en un archivo de datos o conjunto de datos de IBM® SPSS® Statistics. También puede guardar estos valores transformados y puntuaciones de objetos en el conjunto de datos activo.

- Los conjuntos de datos están disponibles durante la sesión actual, pero no así en las sesiones posteriores, a menos que los haya guardado explícitamente como archivos de datos. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables.
- Los nombres de archivo o de conjunto de datos deben ser distintos para cada tipo de dato guardado.
- Si guarda puntuaciones de objeto o valores transformados en el conjunto de datos activo, podrá especificar el número de dimensiones nominales múltiples.

Figura 6-8
Cuadro de diálogo Guardar

ACM: Guardar

Datos discretizados

Crear datos discretizados

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Variables transformadas

Guardar variables transformadas en el conjunto de datos activo

Crear variables transformadas

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Puntuaciones de los objetos

Guardar puntuaciones de los objetos en el conjunto de datos activo

Crear puntuaciones de los objetos

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos

Dimensiones nominales múltiples: Todo Primero:

Análisis de correspondencias múltiple: Gráficos de objetos

El cuadro de diálogo Gráficos de objetos permite especificar los tipos de gráficos deseados y las variables que se van a representar

Figura 6-9
Cuadro de diálogo Gráficos de objetos



Puntos de objetos. Se muestra un gráfico de los puntos de objetos.

Objetos y centroides (diagrama de dispersión biespacial). Los puntos de objetos se representan con los centroides de las variables.

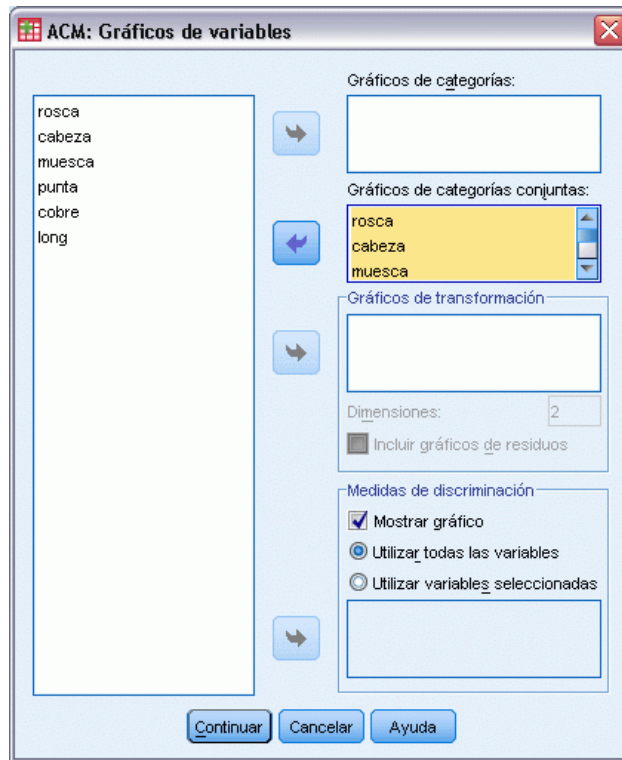
Variables de diagrama de dispersión biespacial. Puede utilizar todas las variables para los gráficos de dispersión biespacial o seleccionar un subconjunto.

Etiquetar objetos. Se puede elegir que los objetos se etiqueten con las categorías de las variables seleccionadas (se pueden seleccionar entre los valores del indicador de categoría o las etiquetas de valor, en el cuadro de diálogo Opciones) o con sus números de caso. Se genera un gráfico por cada variable si se selecciona Variable.

Análisis de correspondencias múltiple: Gráficos de variables

El cuadro de diálogo Gráficos de variables permite especificar los tipos de gráficos deseados y las variables que se van a representar.

Figura 6-10
Cuadro de diálogo Gráficos de variables



Gráficos de categorías. Para cada variable seleccionada, se representa un gráfico de las coordenadas del centroide. Las categorías se encuentran en los centroides de los objetos de las categorías particulares.

Gráficos de categorías conjuntas. Este es un único gráfico de las coordenadas del centroide de cada variable seleccionada.

Gráficos de transformación. Muestra un gráfico de las cuantificaciones de las categorías óptimas en oposición a los indicadores de las categorías. Puede especificar el número de dimensiones deseado; se generará un gráfico para cada dimensión. También se puede seleccionar si se muestran los gráficos de los residuos para cada variable seleccionada.

Medidas de discriminación. Genera un único gráfico de las medidas de discriminación de las variables seleccionadas.

Funciones adicionales del comando MULTIPLE CORRESPONDENCE

Se puede personalizar el análisis de correspondencias múltiple si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando `MULTIPLE CORRESPONDENCE`. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar nombres de raíz para las variables transformadas, puntuaciones de objetos y aproximaciones al guardarlas en el conjunto de datos activo (con el subcomando `SAVE`).

- Especificar una longitud máxima de las etiquetas para cada gráfico por separado (con el subcomando `PLOT`).
- Especificar una lista de variables distinta para los gráficos de residuos (con el subcomando `PLOT`).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Escalamiento multidimensional (PROXSCAL)

El escalamiento multidimensional trata de encontrar la estructura existente en un conjunto de medidas de proximidades entre objetos. Este proceso se logra asignando las observaciones a posiciones específicas en un espacio conceptual de pocas dimensiones, de modo que las distancias entre los puntos en el espacio concuerden al máximo con las similaridades (o disimilaridades) dadas. El resultado es una representación de mínimos cuadrados de los objetos en dicho espacio de pocas dimensiones que, en muchos casos, le ayudará a entender mejor los datos.

Ejemplo. El escalamiento multidimensional puede ser muy útil en la determinación de relaciones perceptuales. Por ejemplo, al considerar la imagen de un producto, se puede llevar a cabo un estudio para obtener un conjunto de datos que describa la similaridad percibida (o proximidad) de este producto con el de la competencia. Mediante estas proximidades y las variables independientes (como el precio), puede intentar determinar las variables que son importantes en la visión que el público tiene del producto, así como ajustar la imagen de acuerdo con ello.

Estadísticos y gráficos. Historial de iteraciones, medidas de stress, descomposición del stress, coordenadas del espacio común, distancias entre objetos dentro de la configuración final, ponderaciones del espacio individual, espacios individuales, proximidades transformadas, variables independientes transformadas, gráficos del stress, diagramas de dispersión del espacio común, diagramas de dispersión de la ponderación del espacio individual, diagramas de dispersión de los espacios individuales, gráficos de transformación, gráficos residuales de Shepard y gráficos de transformación de las variables independientes.

Datos. Los datos se pueden suministrar en forma de matrices de proximidades o como variables que se convierten en matrices de proximidades. Las matrices pueden tener formato en columna o a través de las columnas. Las proximidades se pueden tratar con los niveles de escalamiento de razón, de intervalo, ordinal o de líneaSp.

Supuestos. Se deben especificar al menos tres variables. El número de dimensiones no puede superar al número de objetos menos uno. Se omite la reducción de dimensionalidad si se combina con varios inicios aleatorios. Si sólo se especifica una fuente, todos los modelos son equivalentes al modelo identidad, por consiguiente, el análisis toma por defecto el modelo identidad.

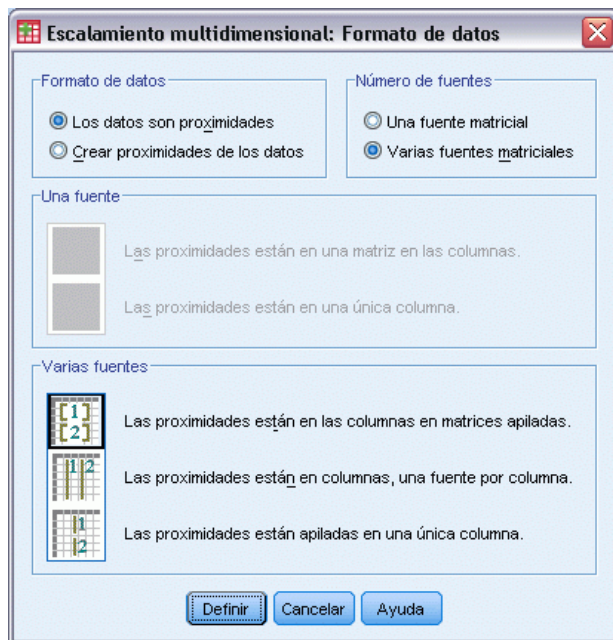
Procedimientos relacionados. Si se escalan todas las variables a nivel numérico, el análisis se corresponderá con el análisis de escalamiento multidimensional típico.

Para obtener un escalamiento multidimensional

- En los menús, seleccione:
Analizar > Escala > Escalamiento multidimensional (PROXSCAL)...

Accederá al cuadro de diálogo Formato de datos.

Figura 7-1
Cuadro de diálogo Formato de datos



- Especifique el formato de los datos:

Formato de datos. Especifique si los datos son medidas de proximidad o si desea crear las proximidades a partir de los datos.

Número de fuentes. Si los datos son proximidades, debe especificar si dispone de una fuente única o de varias fuentes de medidas de proximidad.

Una fuente. Si hay una sola fuente de proximidades, especifique si el conjunto de datos se encuentra en un formato con las proximidades en una matriz a través de las columnas o en una única columna con dos variables diferentes para identificar la fila y la columna de cada proximidad.

- **Las proximidades están en una matriz a través de columnas.** La matriz de proximidades está distribuida en un número de columnas igual al número de objetos. Lo que conduce al cuadro de diálogo Proximidades en matrices a través de columnas.
- **Las proximidades están en una sola columna.** Las matrices de proximidades se contraen en una única columna o variable. Se necesitan dos variables adicionales, que identifican la fila y la columna de cada casilla. Lo que conduce al cuadro de diálogo Proximidades en una columna.

Varias fuentes. Si hay varias fuentes de proximidades, especifique si el conjunto de datos se encuentra en un formato con las proximidades a través de las columnas en matrices apiladas, en varias columnas con una fuente por cada columna o en una única columna.

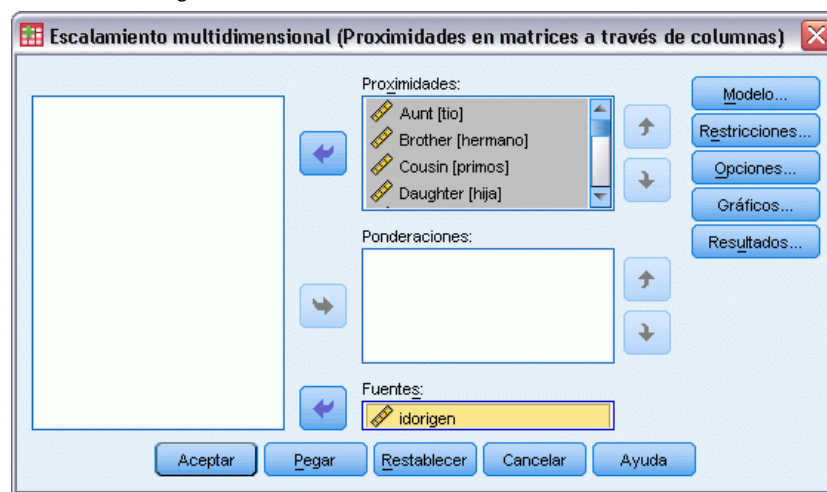
- **Las proximidades se encuentran en matrices apiladas en las columnas.** Las matrices de proximidades están distribuidas a lo largo de un número de columnas igual al número de objetos y apiladas una sobre la otra a lo largo de un número de filas igual al número de objetos por el número de fuentes. Lo que conduce al cuadro de diálogo Proximidades en matrices a través de columnas.
 - **Las proximidades están en columnas, una fuente por columna.** Las matrices de proximidades se contraen en varias columnas o variables. Se necesitan dos variables adicionales, que identifican la fila y la columna de cada casilla. Lo que conduce al cuadro de diálogo Proximidades en columnas.
 - **Las proximidades se apilan en una sola columna.** Las matrices de proximidades se contraen en una única columna o variable. Se necesitan tres variables adicionales, que identifican la fila, la columna y la fuente de cada casilla. Lo que conduce al cuadro de diálogo Proximidades en una columna.
- Pulse en Definir.

Proximidades en matrices a través de columnas

Si selecciona el modelo para los datos con las proximidades en matrices, para una o varias fuentes, en el cuadro de diálogo Formato de datos, el cuadro de diálogo principal aparecerá de la siguiente manera:

Figura 7-2

Cuadro de diálogo Proximidades en matrices a través de columnas



- Seleccione tres o más variables de proximidades. (Asegúrese de que el orden de las variables en la lista coincide con el orden de las columnas para las proximidades.)
- Si lo desea, seleccione un número de variables de ponderaciones igual al número de variables de proximidades. (Asegúrese de que el orden de las ponderaciones coincide con el orden de las proximidades a las que ponderan.)

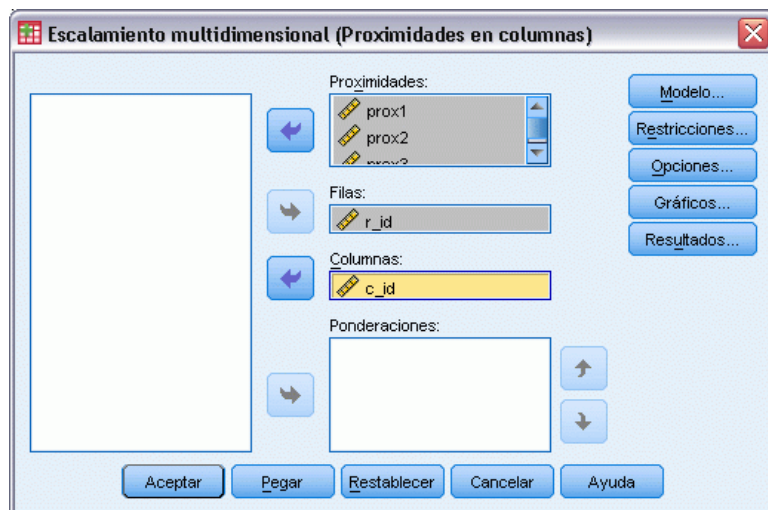
- ▶ Como alternativa, si existen varias fuentes, seleccione una variable de fuentes. (El número de casos en cada variable de proximidades debe ser igual al número de variables de proximidades multiplicado por el número de fuentes.)

Además, puede definir un modelo para el escalamiento multidimensional, establecer restricciones en el espacio común, establecer criterios de convergencia, especificar la configuración inicial que se va a utilizar y seleccionar gráficos y resultados.

Proximidades en columnas

Si selecciona el modelo de varias columnas para varias fuentes, en el cuadro de diálogo Formato de datos, el cuadro de diálogo principal aparecerá de la siguiente manera:

Figura 7-3
Cuadro de diálogo Proximidades en columnas



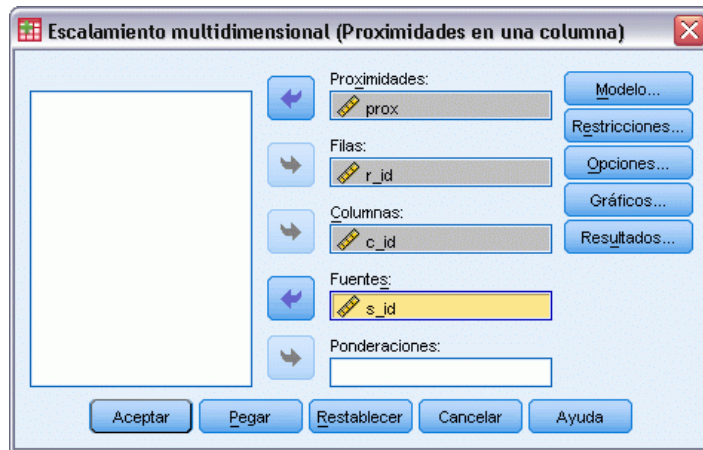
- ▶ Seleccione dos o más variables de proximidades. (Se asume que cada variable es una matriz de proximidades de una fuente diferente.)
- ▶ Seleccione una variable de fila para definir las ubicaciones de fila para las proximidades en cada variable de proximidades.
- ▶ Seleccione una variable de columna para definir las ubicaciones de columna para las proximidades en cada una de las variables de proximidades. (Las casillas de la matriz de proximidades que no reciban una designación fila/columna se tratarán como pérdidas.)
- ▶ Si lo desea, seleccione un número de variables de ponderaciones igual al número de variables de proximidades.

Además, puede definir un modelo para el escalamiento multidimensional, establecer restricciones en el espacio común, establecer criterios de convergencia, especificar la configuración inicial que se va a utilizar y seleccionar gráficos y resultados.

Proximidades en una columna

Si selecciona el modelo de una columna para una o varias fuentes en el cuadro de diálogo Formato de datos, el cuadro de diálogo principal aparecerá de la siguiente manera:

Figura 7-4
Cuadro de diálogo Proximidades en una columna



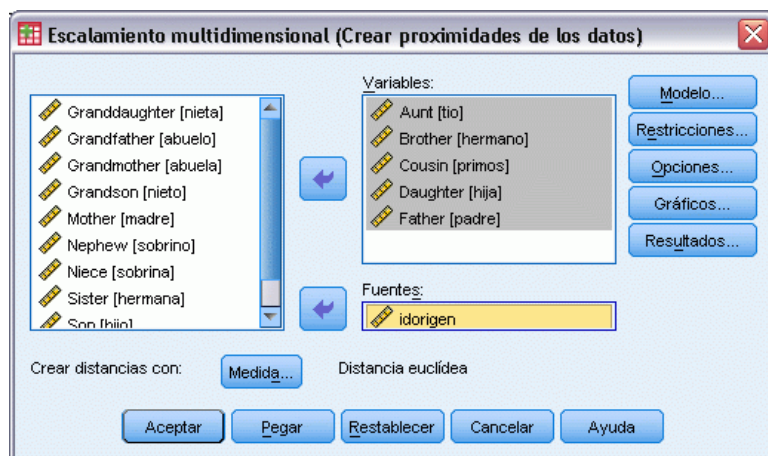
- ▶ Seleccione una variable de proximidades. (Se asume que es una o más matrices de proximidades.)
- ▶ Seleccione una variable de fila para definir las ubicaciones de fila para las proximidades en la variable de proximidades.
- ▶ Seleccione una variable de columna para definir las ubicaciones de columna para las proximidades en la variable de proximidades.
- ▶ Si existen varias fuentes, seleccione una variable de fuentes. (Para cada fuente, las casillas de la matriz de proximidades que no reciban una designación fila/columna se tratarán como perdidas.)
- ▶ Si lo desea, seleccione una variable de ponderaciones.

Además, puede definir un modelo para el escalamiento multidimensional, establecer restricciones en el espacio común, establecer criterios de convergencia, especificar la configuración inicial que se va a utilizar y seleccionar gráficos y resultados.

Crear proximidades de los datos

Si decide crear proximidades a partir de los datos, en el cuadro de diálogo Formato de datos, el cuadro de diálogo principal aparecerá de la siguiente manera:

Figura 7-5
Cuadro de diálogo Crear proximidades de los datos

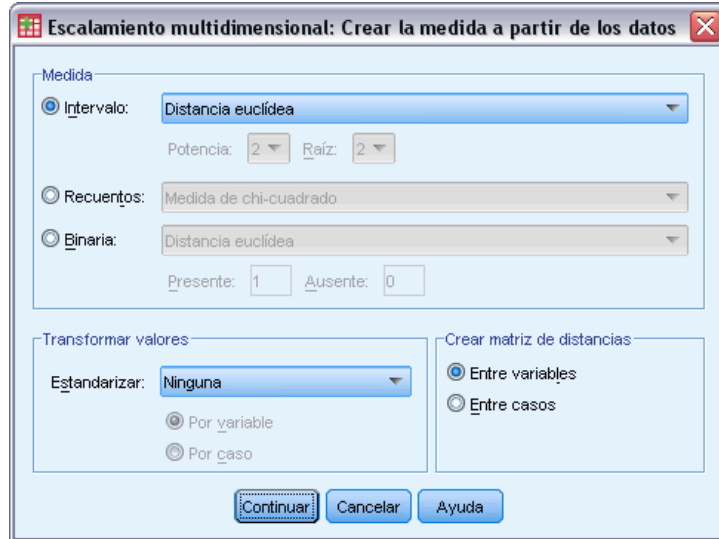


- ▶ Si crea distancias entre variables (consulte el cuadro de diálogo Crear la medida a partir de los datos), seleccione al menos tres variables. Estas variables se utilizarán para crear la matriz de proximidades (o matrices, si hay varias fuentes). Si crea las distancias entre los casos, sólo será necesaria una variable.
- ▶ Si existen varias fuentes, seleccione una variable de fuentes.
- ▶ Si lo desea, seleccione una medida para crear proximidades.

Además, puede definir un modelo para el escalamiento multidimensional, establecer restricciones en el espacio común, establecer criterios de convergencia, especificar la configuración inicial que se va a utilizar y seleccionar gráficos y resultados.

Crear la medida a partir de los datos

Figura 7-6
Cuadro de diálogo Crear la medida a partir de los datos



El escalamiento multidimensional utiliza datos de disimilaridad para crear una solución de escalamiento. Si los datos son datos multivariantes (los valores de las variables que se han medido), debe crear los datos de disimilaridad para poder calcular una solución de escalamiento multidimensional. Puede especificar los detalles para la creación de las medidas de disimilaridad a partir de los datos.

Medida. Le permite especificar la medida de disimilaridad para el análisis. Seleccione una opción del grupo Medida que se corresponda con el tipo de datos y, a continuación, seleccione una de las medidas de la lista desplegable correspondiente a ese tipo de medida. Las opciones disponibles son:

- **Intervalo.** Distancia euclídea, Distancia euclídea al cuadrado, Chebychev, Bloque, Minkowski o Personalizada.
- **Contar apariciones.** Medida de chi-cuadrado o Medida de phi-cuadrado.
- **Binaria.** Distancia euclídea, Distancia euclídea al cuadrado, Diferencia de tamaño, Diferencia de configuración, Varianza o Lance y Williams.

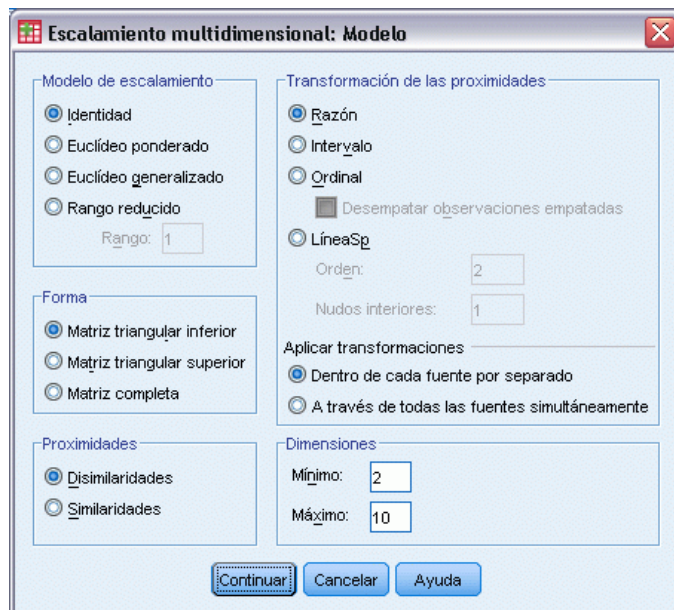
Crear matriz de distancias. Le permite elegir la unidad de análisis. Las opciones son Entre variables o Entre casos.

Transformar valores. En determinados casos, como cuando las variables se miden en escalas muy distintas, puede que desee tipificar los valores antes de calcular las proximidades (no es aplicable a datos binarios). Seleccione un método de estandarización de la lista desplegable Estandarizar (si no se requiere ninguna estandarización, seleccione Ninguna).

Definir un modelo de escalamiento multidimensional

El cuadro de diálogo modelo permite especificar un modelo de escalamiento, su número de dimensiones máximo y mínimo, la estructura de la matriz de proximidades, la transformación que se va a utilizar en las proximidades y si las proximidades se transforman dentro de cada fuente por separado o de forma incondicional en la fuente.

Figura 7-7
Cuadro de diálogo Modelo



Modelo de escalamiento. Seleccione entre las siguientes alternativas:

- **Identidad.** Todas las fuentes tienen la misma configuración.
- **Euclídeo ponderado.** Éste es un modelo de diferencias individuales. Cada fuente dispone de un espacio individual en el que cada dimensión del espacio común se pondera de forma diferenciada.
- **Euclídeo generalizado.** Éste es un modelo de diferencias individuales. Cada fuente dispone de un espacio individual que es igual a la rotación del espacio común, seguida de una ponderación diferencial de las dimensiones.
- **Rango reducido.** Este modelo es un modelo Euclídeo generalizado para el que puede especificar el rango del espacio individual. Debe especificar un rango que sea mayor o igual que 1 y menor que el máximo número de dimensiones.

Forma. Especifique si las proximidades deben tomarse de la parte triangular inferior o de la triangular superior de la matriz de proximidades. Puede especificar que se utilice la matriz completa, en cuyo caso se analizará la suma ponderada de la parte triangular superior y de la parte triangular inferior. En cualquier caso, se debe especificar la matriz completa, incluyendo la diagonal, aunque sólo se utilicen las partes especificadas.

Proximidades. Especifique si la matriz de proximidades contiene medidas de similitud o disimilitud.

Transformaciones de proximidades. Seleccione entre las siguientes alternativas:

- **Razón.** Las proximidades transformadas son proporcionales a las proximidades originales. Sólo se permite para proximidades con un valor positivo.
- **Intervalo.** Las proximidades transformadas son proporcionales a las proximidades originales, más un término de intersección. La intersección asegura que todas las proximidades transformadas sean positivas.
- **Ordinal.** Las proximidades transformadas tienen el mismo orden que las proximidades originales. Puede especificar si para las proximidades empatadas se debe mantener el empate o si se deben desempatar.
- **LíneaSp.** Las proximidades transformadas son una transformación polinómica por partes no decreciente suave de las proximidades originales. Puede especificar el grado del polinomio y el número de nudos interiores.

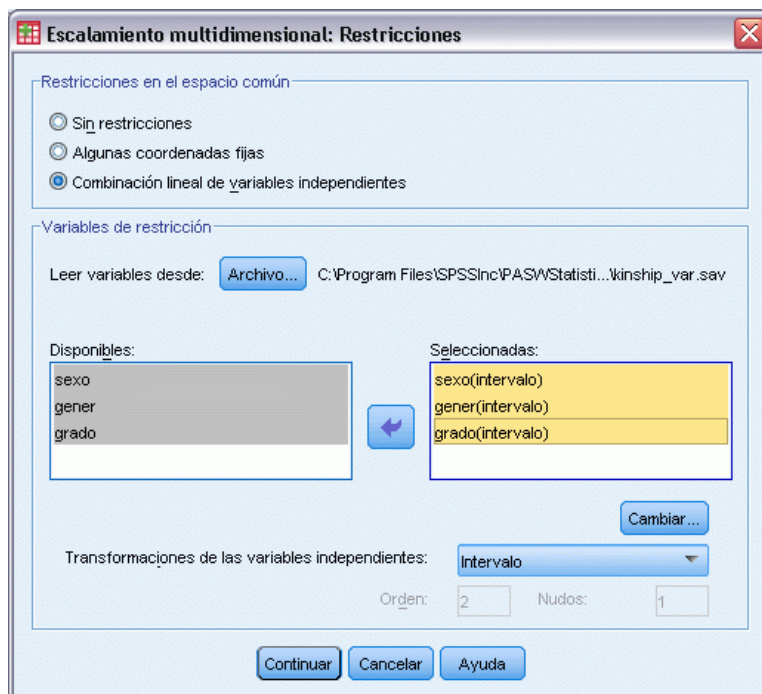
Aplicar transformaciones. Especifica si sólo las proximidades dentro de cada fuente se comparan entre sí o si las comparaciones son incondicionales respecto a la fuente.

Dimensiones. Por defecto, se calcula una solución en dos dimensiones (Mínimo=2, Máximo=2). Puede seleccionar un entero mínimo y otro máximo entre 1 y el número de objetos menos 1 (siempre y cuando el mínimo sea menor o igual que el máximo). El procedimiento calcula una solución en las dimensiones máximas y, a continuación, reduce la dimensionalidad por pasos hasta llegar al mínimo.

Escalamiento multidimensional: Restricciones

El cuadro de diálogo Restricciones permite establecer restricciones en el espacio común.

Figura 7-8
Cuadro de diálogo Restricciones



Restricciones en el espacio común. Especifique el tipo de restricciones que desee.

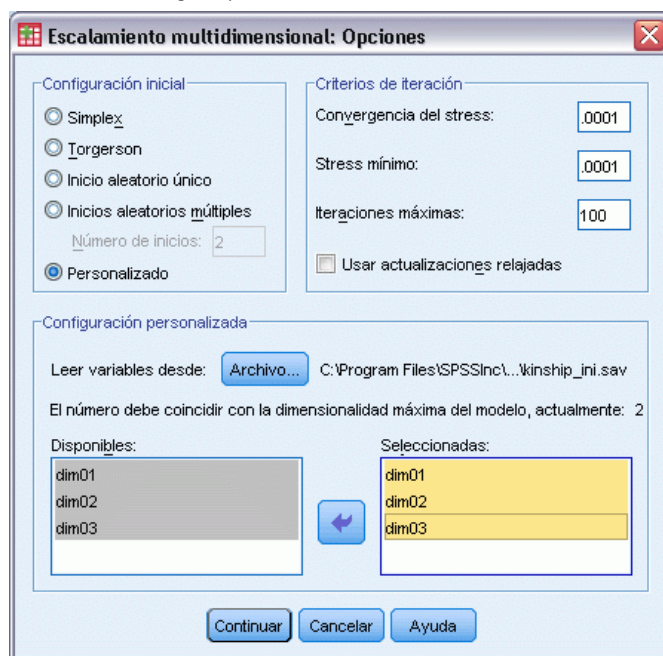
- **Sin restricciones.** No se impone ninguna restricción en el espacio común.
- **Algunas coordenadas fijas.** La primera variable seleccionada incluye las coordenadas de los objetos en la primera dimensión, la segunda variable se corresponde con las coordenadas de la segunda dimensión, y así sucesivamente. Un valor perdido indica que una coordenada en una dimensión está libre. El número de variables seleccionadas debe ser igual al número máximo de dimensiones solicitadas.
- **Combinación lineal de variables independientes.** El espacio común se restringe hasta ser una combinación lineal de las variables seleccionadas.

Variables de restricción. Seleccione las variables que definen las restricciones en el espacio común. Si se especifica una combinación lineal, se debe especificar una transformación de intervalo, nominal, ordinal o de líneaSp para las variables de restricción. En cualquier caso, el número de casos para cada variable debe ser igual al número de objetos.

Escalamiento multidimensional: Opciones

El cuadro de diálogo Opciones permite seleccionar el estilo para la configuración inicial, especificar los criterios de iteración y convergencia y seleccionar entre actualizaciones estándar o relajadas.

Figura 7-9
Cuadro de diálogo Opciones



Configuración inicial. Seleccione una de las siguientes opciones:

- **Simplex.** Los objetos se colocan a la misma distancia unos de otros en la máxima dimensión. Se lleva a cabo una iteración para mejorar esta configuración de gran dimensionalidad, seguida de una operación de reducción de dimensiones para obtener una configuración inicial que tenga el número máximo de dimensiones especificadas en el cuadro de diálogo Modelo.
- **Torgerson.** Se utiliza una solución de escalamiento clásico como la configuración inicial.
- **Inicio aleatorio único.** Se selecciona una configuración aleatoriamente.
- **Inicios aleatorios múltiples.** Se seleccionan varias configuraciones aleatoriamente y se utiliza la que tenga menor stress bruto normalizado como la configuración inicial.
- **Personalizado.** Seleccione variables que contengan las coordenadas de su propia configuración inicial. El número de variables seleccionadas debe ser igual al número máximo de dimensiones especificadas, correspondiendo la primera variable a las coordenadas en la dimensión 1, la segunda variable a las coordenadas en la dimensión 2 y así sucesivamente. El número de casos en cada variable debe ser igual al número de objetos.

Criterios de iteración. Especifique los valores de los criterios de iteración.

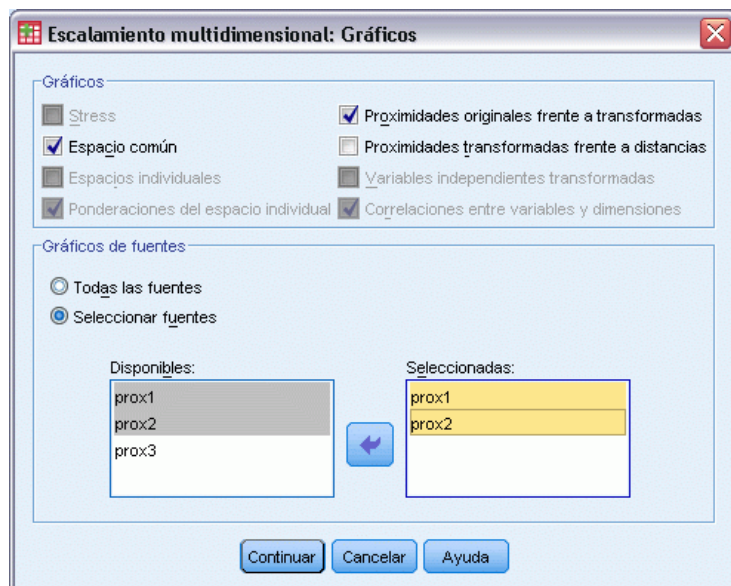
- **Convergencia del stress.** El algoritmo detendrá la iteración cuando la diferencia entre valores consecutivos del stress bruto normalizado sea menor que el número aquí especificado, que debe estar entre 0,0 y 1,0.
- **Stress mínimo.** El algoritmo detendrá la iteración cuando el stress bruto normalizado sea menor que el número aquí especificado, que debe estar entre 0,0 y 1,0.

- **Iteraciones máximas.** El algoritmo realizará el número de iteraciones aquí especificado, a no ser que antes se satisfaga alguno de los criterios anteriores.
- **Usar actualizaciones relajadas.** Las actualizaciones relajadas acelerarán el algoritmo; no se pueden utilizar con modelos que no sean el de identidad, ni tampoco con restricciones.

Escalamiento multidimensional: Gráficos, Versión 1

El cuadro de diálogo Gráficos permite especificar los gráficos que se van a generar. Si tiene el formato de los datos: Proximidades en las columnas, aparecerá el siguiente cuadro de diálogo. Para los gráficos Ponderaciones del espacio individual, Proximidades originales frente a transformadas y Proximidades transformadas frente a distancias, puede especificar las fuentes para las que se generarán los gráficos. La lista de fuentes disponibles es la lista de variables de proximidades en el cuadro de diálogo principal.

Figura 7-10
Cuadro de diálogo Gráficos, versión 1



Stress. Se genera un gráfico del stress bruto normalizado frente a las dimensiones. Este gráfico se genera sólo si el número máximo de dimensiones es mayor que el número mínimo de dimensiones.

Espacio común. Aparecerá un diagrama de dispersión matricial de las coordenadas del espacio común.

Espacios individuales. Para cada fuente, las coordenadas de los espacios individuales aparecerán diagramas de dispersión matriciales. Esto es posible sólo si se especifica uno de los modelos de diferencias individuales en el cuadro de diálogo Modelo.

Ponderaciones del espacio individual. Se genera un diagrama de dispersión de las ponderaciones del espacio individual. Esto es posible sólo si se especifica uno de los modelos de diferencias individuales en el cuadro de diálogo Modelo. Para el modelo Euclídeo ponderado, las ponderaciones se imprimen en gráficos, con una dimensión en cada eje. Para el modelo Euclídeo generalizado, se genera un gráfico por cada dimensión, indicando la rotación y la ponderación de

dicha dimensión, El modelo de rango reducido genera el mismo gráfico que el modelo Euclídeo generalizado, pero reduce el número de dimensiones para los espacios individuales.

Proximidades originales frente a transformadas. Se generan gráficos de las proximidades originales frente a las transformadas.

Proximidades transformadas frente a distancias. Se representan las proximidades transformadas frente a las distancias.

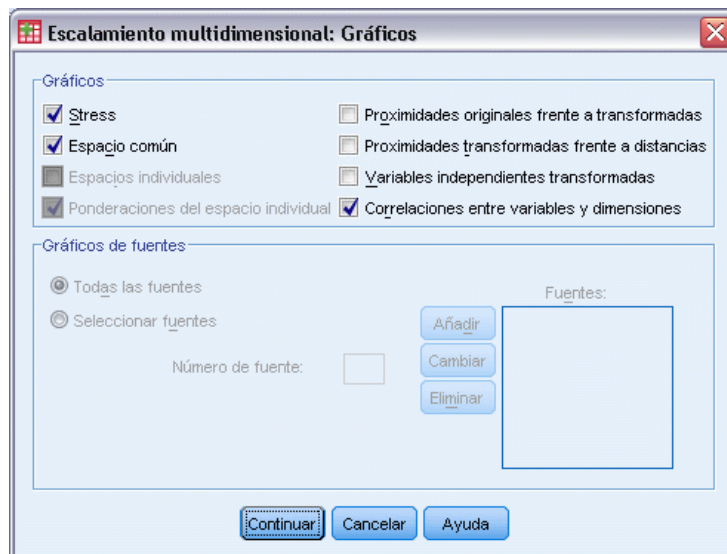
Variables independientes transformadas. Se generan gráficos de transformación para las variables independientes.

Correlaciones entre variables y dimensiones. Aparece un gráfico de las correlaciones entre las variables independientes y las dimensiones del espacio común.

Escalamiento multidimensional: Gráficos, Versión 2

El cuadro de diálogo Gráficos permite especificar los gráficos que se van a generar. Si el formato de los datos no es Proximidades en las columnas, aparecerá el siguiente cuadro de diálogo Gráficos. Para los gráficos Ponderaciones del espacio individual, Proximidades originales frente a transformadas y Proximidades transformadas frente a distancias, puede especificar las fuentes para las que se generarán los gráficos. Los números de fuentes introducidos deben ser valores de la variable de fuentes especificada en el cuadro de diálogo principal y el rango debe estar entre 1 y el número de fuentes.

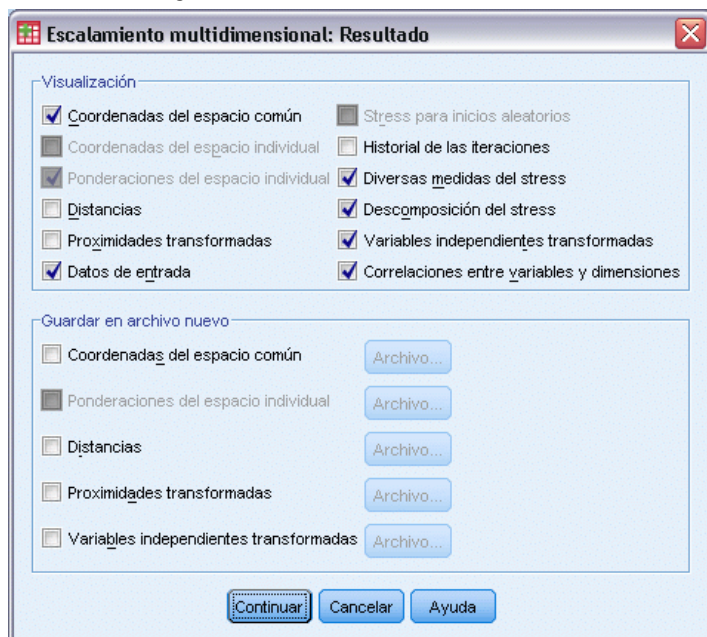
Figura 7-11
Cuadro de diálogo Gráficos, versión 2



Escalamiento multidimensional: Resultados

El cuadro de diálogo Resultados permite controlar la cantidad de resultados mostrados y guardar parte en archivos independientes.

Figura 7-12
Cuadro de diálogo Resultados



Mostrar. Seleccione uno o varios de los siguientes elementos para su visualización:

- **Coordenadas del espacio común.** Muestra las coordenadas del espacio común.
- **Coordenadas del espacio individual.** Las coordenadas del espacio individual sólo aparecen si no se trata del modelo identidad.
- **Ponderaciones del espacio individual.** Muestra las ponderaciones del espacio individual sólo si se especifican modelos de diferencias individuales. En función del modelo, las ponderaciones del espacio se descomponen en ponderaciones de rotación y de dimensión, que también se muestran.
- **Distancias.** Muestra las distancias entre los objetos en la configuración.
- **Proximidades transformadas.** Muestra las proximidades transformadas entre los objetos en la configuración.
- **Datos de entrada.** Incluye las proximidades originales y, si aparecen, las ponderaciones de los datos, la configuración inicial y las coordenadas fijas de las variables independientes.
- **Stress para inicios aleatorios.** Muestra la semilla de aleatorización y el valor del stress bruto normalizado para cada inicio aleatorio.
- **Historial de iteraciones.** Muestra el historial de iteraciones del algoritmo principal.
- **Diversas medidas del stress.** Muestra distintos valores del stress. La tabla contiene valores para el stress bruto normalizado, Stress-I, Stress-II, S-Stress, la Dispersión explicada (DAF),
- **Descomposición del stress.** Muestra la descomposición por orígenes y por objetos del stress bruto normalizado final, incluyendo el promedio por objeto y por origen.

- **Variables independientes transformadas.** Si se ha seleccionado una restricción de combinación lineal, se muestran las variables independientes transformadas y las ponderaciones de regresión correspondientes.
- **Correlaciones entre variables y dimensiones.** Si se ha seleccionado una restricción de combinación lineal, se muestran las correlaciones entre las variables independientes y las dimensiones del espacio común.

Guardar en archivo nuevo. Puede guardar las coordenadas del espacio común, las ponderaciones del espacio individual, las distancias, las proximidades transformadas y las variables independientes transformadas en distintos archivos de datos de IBM® SPSS® Statistics.

Funciones adicionales del comando PROXSCAL

Se puede personalizar el escalamiento multidimensional del análisis de proximidades si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando PROXSCAL resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especificar una lista de variables distinta para los gráficos de transformación y los de residuos (con el subcomando PLOT).
- Especificar listas de variables distintas para los gráficos de ponderaciones del espacio individual, los de transformaciones y los de residuos (con el subcomando PLOT).
- Especificar un subconjunto de los gráficos de transformación de las variables independientes que se van a mostrar (con el subcomando PLOT).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Desplegamiento multidimensional (PREFSCAL)

El procedimiento Desplegamiento multidimensional intenta buscar una escala cuantitativa común que permita examinar visualmente las relaciones entre dos conjuntos de objetos.

Ejemplos. Ha pedido a 21 individuos que clasifiquen 15 elementos de desayuno por orden de preferencia, del 1 al 15. Con el desplegado multidimensional, puede determinar que los individuos discriminan entre los elementos de desayuno de dos formas principales: entre pan tierno y crujiente, y entre elementos con mucha grasa o con poca grasa.

Por otro lado, también ha pedido a un grupo de conductores que evalúen 26 modelos de coches respecto a 10 atributos, en una escala de 6 puntos, desde 1=“completamente falso” hasta 6=“totalmente cierto”. Los valores promediados respecto a los individuos se toman como similitudes. Utilice el desplegado multidimensional para encontrar conglomerados de modelos similares y los atributos con los que más se asocian.

Estadísticos y gráficos. El procedimiento Desplegamiento multidimensional puede producir historial de iteraciones, medidas de stress, descomposición del stress, coordenadas del espacio común, distancias entre objetos dentro de la configuración final, ponderaciones del espacio individual, espacios individuales, proximidades transformadas, gráficos del stress, diagramas de dispersión del espacio común, diagramas de dispersión de la ponderación del espacio individual, diagramas de dispersión de los espacios individuales, gráficos de transformación y gráficos residuales de Shepard.

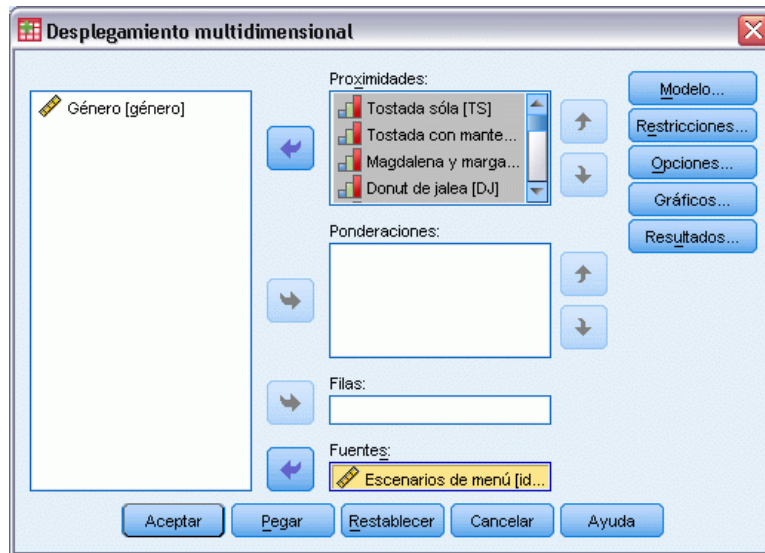
Datos. Los datos se suministran en forma de matrices de proximidades rectangulares. Cada columna se considera como un objeto de columna independiente. Cada fila de una matriz de proximidades se considera como un objeto de fila independiente. Cuando hay varias fuentes de proximidades, las matrices se apilan.

Supuestos. Se deben especificar al menos dos variables. El número de dimensiones de la solución no puede superar el número de objetos menos uno. Si sólo se especifica una fuente, todos los modelos son equivalentes al modelo identidad, por consiguiente, el análisis toma por defecto el modelo identidad.

Para obtener un desplegado multidimensional

- ▶ En los menús, seleccione:
Analizar > Escala > Desplegamiento multidimensional (PREFSCAL)...

Figura 8-1
Cuadro de diálogo principal Desplegamiento multidimensional



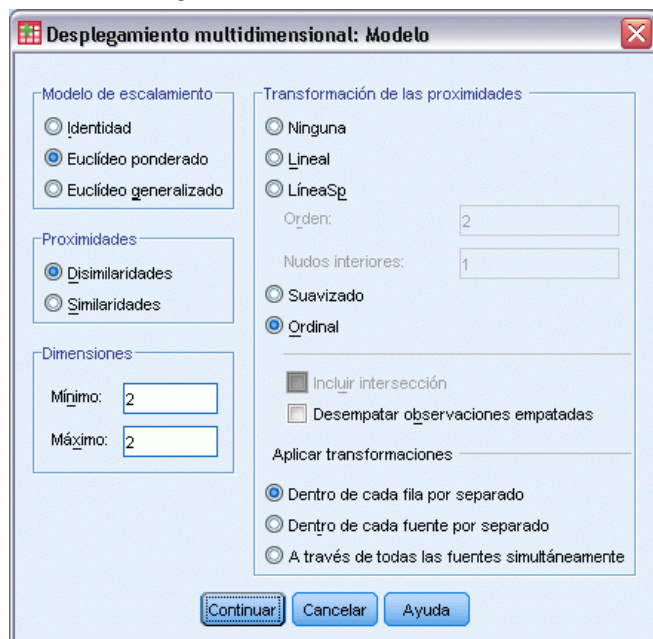
- ▶ Seleccione dos o más variables que identifiquen las columnas en la matriz de proximidades rectangular. Cada variable representa un objeto de columna distinto.
- ▶ Si lo desea, seleccione un número de variables de ponderación que sea igual al número de variables de objeto de columna. El orden de las variables de ponderación debe coincidir con el orden de los objetos de columna a los que ponderan.
- ▶ Si lo desea, seleccione una variable de filas. Los valores (o etiquetas de valor) de esta variable se utilizan para etiquetar los objetos de fila en los resultados.
- ▶ Si existen varias fuentes, puede seleccionar una variable de fuentes, si lo desea. El número de casos del archivo de datos debe ser igual al número de objetos de fila multiplicado por el número de fuentes.

Además, puede definir un modelo para el despliegamiento multidimensional, establecer restricciones en el espacio común, establecer criterios de convergencia, especificar la configuración inicial que se va a utilizar y seleccionar gráficos y resultados.

Definir un modelo de despliegamiento multidimensional

El cuadro de diálogo Modelo permite especificar un modelo de escalamiento, su número de dimensiones máximo y mínimo, la estructura de la matriz de proximidades, la transformación que se va a utilizar en las proximidades y si las proximidades se transforman de forma condicional sobre la fila, de forma condicional sobre la fuente o de forma incondicional en la fila.

Figura 8-2
Cuadro de diálogo Modelo



Modelo de escalamiento. Seleccione entre las siguientes alternativas:

- **Identidad.** Todas las fuentes tienen la misma configuración.
- **Euclídeo ponderado.** Éste es un modelo de diferencias individuales. Cada fuente dispone de un espacio individual en el que cada dimensión del espacio común se pondera de forma diferenciada.
- **Euclídeo generalizado.** Éste es un modelo de diferencias individuales. Cada fuente dispone de un espacio individual que es igual a la rotación del espacio común, seguida de una ponderación diferencial de las dimensiones.

Proximidades. Especifique si la matriz de proximidades contiene medidas de similitud o disimilitud.

Dimensiones. Por defecto, se calcula una solución en dos dimensiones (Mínimo=2, Máximo=2). Puede seleccionar un entero mínimo y otro máximo entre 1 y el número de objetos menos 1, siempre y cuando el mínimo sea menor o igual que el máximo. El procedimiento calcula una solución en las dimensiones máximas y, a continuación, reduce la dimensionalidad por pasos hasta llegar al mínimo.

Transformaciones de proximidades. Seleccione entre las siguientes alternativas:

- **Ninguno.** Las proximidades no se transforman. Si lo desea, puede seleccionar Incluir intersección, en cuyo caso las proximidades se pueden desplazar según un término constante.
- **Lineal.** Las proximidades transformadas son proporcionales a las proximidades originales, es decir, la función de transformación estima una pendiente y la intersección se fija en 0. También se denomina transformación de la razón. Si lo desea, puede seleccionar Incluir intersección, en cuyo caso las proximidades también se pueden desplazar según un término constante. También se le denomina una transformación de intervalo.

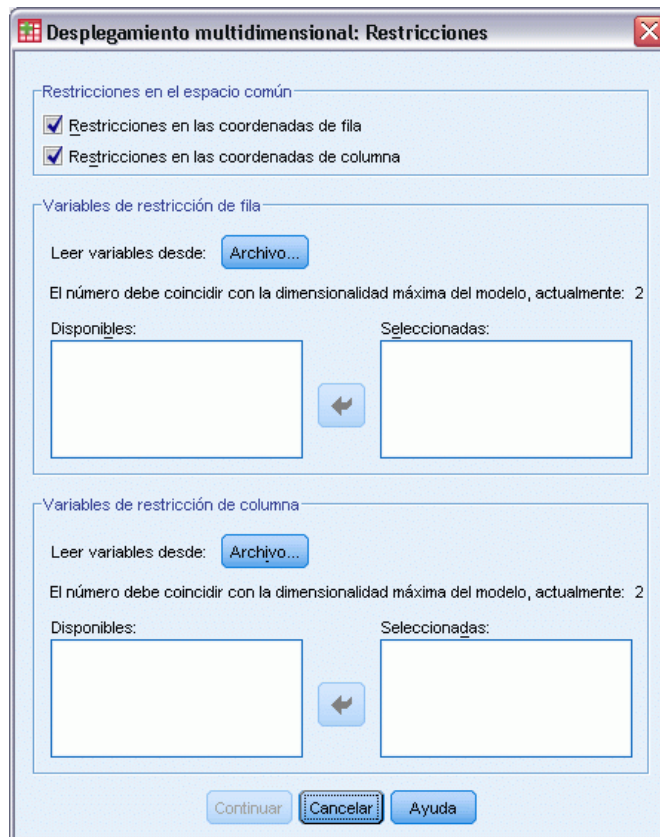
- **LíneaSp.** Las proximidades transformadas son una transformación polinómica por partes no decreciente suave de las proximidades originales. Puede especificar el orden del polinomio y el número de nudos interiores. Si lo desea, puede seleccionar Incluir intersección, en cuyo caso las proximidades también se pueden desplazar según un término constante.
- **Suavizado.** Las proximidades transformadas tienen el mismo orden que las originales, incluida una restricción que tiene en cuenta las diferencias entre los valores posteriores. El resultado es una transformación “ordinal suavizada”. Puede especificar si para las proximidades empatadas se debe mantener el empate o si se deben desempatar.
- **Ordinal.** Las proximidades transformadas tienen el mismo orden que las proximidades originales. Puede especificar si para las proximidades empatadas se debe mantener el empate o si se deben desempatar.

Aplicar transformaciones. Especifique si se compararán entre sí sólo las proximidades dentro de cada fila o sólo las proximidades dentro de cada fuente, o las comparaciones será incondicionales en la fila o en la fuente, es decir, si las transformaciones se realizarán por fila, por fuente o sobre todas las proximidades a la vez.

Restricciones del despliegamiento multidimensional

El cuadro de diálogo Restricciones permite establecer restricciones en el espacio común.

Figura 8-3
Cuadro de diálogo Restricciones



Restricciones en el espacio común. Puede fijar las coordenadas de los objetos de fila y/o columna en el espacio común.

Variables de restricción de fila/columna. Seleccione el archivo que contiene las restricciones y seleccione las variables que definen las restricciones en el espacio común. La primera variable seleccionada incluye las coordenadas de los objetos en la primera dimensión, la segunda variable corresponde a las coordenadas de la segunda dimensión, etc. Un valor perdido indica que una coordenada en una dimensión está libre. El número de variables seleccionadas debe ser igual al número máximo de dimensiones solicitadas. El número de casos de cada variable debe ser igual al número de objetos.

Opciones de despliegamiento multidimensional

El cuadro de diálogo Opciones permite seleccionar el estilo para la configuración inicial, especificar los criterios de iteración y convergencia, y establecer el término de penalización para el stress.

Figura 8-4
Cuadro de diálogo Opciones

Configuración inicial. Seleccione una de las siguientes opciones:

- **Clásica.** La matriz de proximidades rectangular se utiliza para complementar los intrabloques (valores entre filas y entre columnas) de la matriz MDS simétrica completa. Después de formarse la matriz completa, se utiliza una solución de escalamiento clásico como

configuración inicial. Los intrabloques se pueden rellenar mediante la imputación, utilizando la desigualdad de triángulo o las distancias de Spearman.

- **Ross-Cliff.** El inicio Ross-Cliff utiliza los resultados de una descomposición de valor propio en la matriz de proximidades cuadrada y doblemente centrada como los valores iniciales de los objetos de fila y columna.
- **Correspondencia.** El inicio de correspondencia utiliza los resultados de un análisis de correspondencias en los datos invertidos (similaridades en lugar de disimilaridades) con la normalización simétrica de las puntuaciones de fila y columna.
- **Centroides.** El procedimiento comienza colocando los objetos de fila en la configuración con una descomposición de autovalor. A continuación, los objetos de columna se colocan en el centroide de las opciones específicas. Para el número de opciones, especifique un entero positivo entre 1 y el número de variables de proximidad.
- **Inicios aleatorios múltiples.** Las soluciones se calculan para varias configuraciones iniciales seleccionadas aleatoriamente y se muestra como la mejor solución la que tenga la tensión penalizada más baja.
- **Personalizado.** Puede seleccionar variables que contengan las coordenadas de su propia configuración inicial. El número de variables seleccionadas debe ser igual al número máximo de dimensiones especificadas, correspondiendo la primera variable a las coordenadas en la dimensión 1, la segunda variable a las coordenadas en la dimensión 2 y así sucesivamente. El número de casos en cada variable debe ser igual al número combinado de objetos de fila y columna. Las coordenadas de fila y columna deben estar apiladas, con las coordenadas de columna situadas después de las coordenadas de fila.

Criterios de iteración. Especifique los valores de los criterios de iteración.

- **Convergencia del stress.** El algoritmo detendrá la iteración cuando la diferencia relativa entre valores consecutivos de tensión penalizada sea menor que el número aquí especificado, que debe ser no negativo.
- **Stress mínimo.** El algoritmo detendrá la iteración cuando la tensión penalizada sea menor que el número aquí especificado, que debe ser no negativo.
- **Iteraciones máximas.** El algoritmo realizará el número de iteraciones aquí especificado, a no ser que antes se satisfaga alguno de los criterios anteriores.

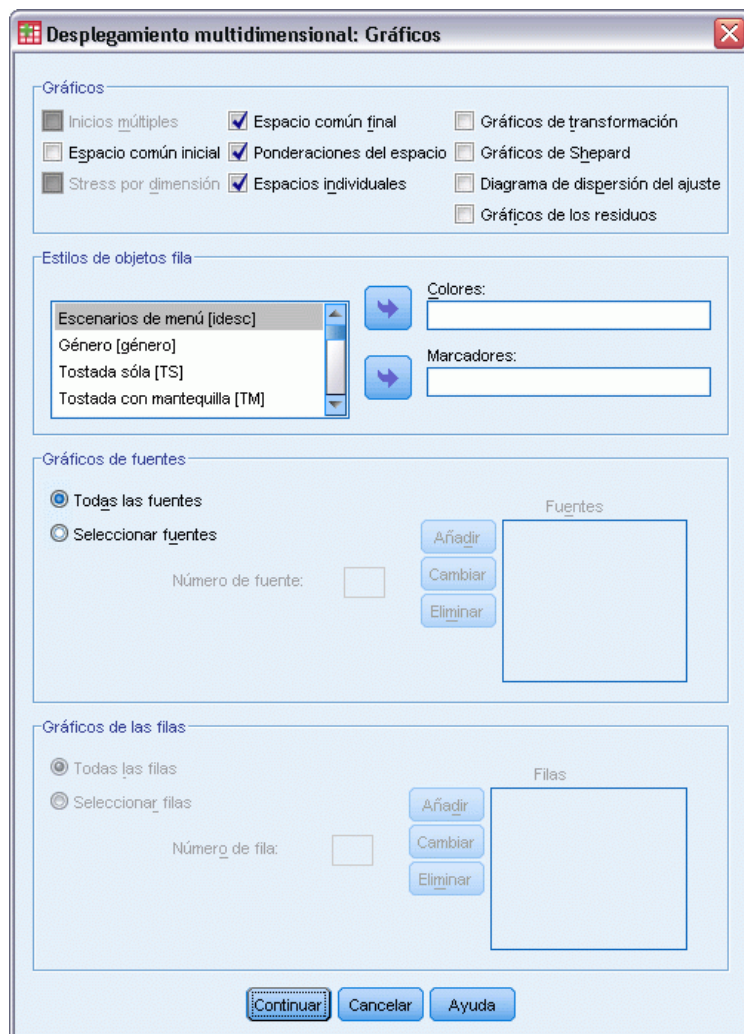
Término de penalización. El algoritmo intenta minimizar la tensión penalizada, una medida de bondad de ajuste igual al producto del Stress-I de Kruskal y un término de penalización basado en el coeficiente de variación de las proximidades transformadas. Estos controles permiten especificar la magnitud y el rango del término de penalización.

- **Magnitud.** Cuanto menor sea el valor del parámetro de magnitud, más fuerte será la penalización. Especifique un valor entre 0,0 y 1,0.
- **Rango.** Este parámetro establece el momento en el que la penalización se vuelve activa. Si se establece en 0,0, la penalización está inactiva. Si se aumenta el valor, el algoritmo buscará una solución con mayor variación entre las proximidades transformadas. Especifique un valor no negativo.

Gráficos de despliegamiento multidimensional

El cuadro de diálogo Gráficos permite especificar los gráficos que se van a generar.

Figura 8-5
Cuadro de diálogo Gráficos



Diagramas. Se encuentran disponibles los siguientes gráficos:

- **Inicios múltiples.** Muestra un histograma apilado de tensión penalizada que indica tanto la tensión como la penalización.
- **Espacio común inicial.** Muestra un diagrama de dispersión matricial de las coordenadas del espacio común inicial.
- **Stress por dimensión.** Genera un gráfico lineal de tensión penalizada frente a dimensionalidad. Este gráfico se genera sólo si el número máximo de dimensiones es mayor que el número mínimo de dimensiones.

- **Espacio común final.** Aparecerá un diagrama de dispersión matricial de las coordenadas del espacio común.
- **Ponderaciones del espacio.** Se genera un diagrama de dispersión de las ponderaciones del espacio individual. Esto es posible sólo si se especifica uno de los modelos de diferencias individuales en el cuadro de diálogo Modelo. Para el modelo Euclídeo ponderado, las ponderaciones de todas las fuentes se muestran en un gráfico, con una dimensión en cada eje. Para el modelo Euclídeo generalizado, se genera un gráfico por cada dimensión, indicando la rotación y la ponderación de dicha dimensión para cada fuente.
- **Espacios individuales.** Se muestra un diagrama de dispersión matricial de las coordenadas del espacio individual de cada fuente. Esto es posible sólo si se especifica uno de los modelos de diferencias individuales en el cuadro de diálogo Modelo.
- **Gráficos de transformación.** Se genera un diagrama de dispersión de las proximidades originales frente a las transformadas. Según cómo se apliquen las transformaciones, se asignará un color distinto a cada fila o fuente. Una transformación incondicional genera un solo color.
- **Gráficos de Shepard.** Las proximidades originales frente a las distancias y las proximidades transformadas. Las distancias se indican con puntos y las proximidades transformadas se indican con una línea. Según cómo se apliquen las transformaciones, se generará una línea distinta para cada fila o fuente. Una transformación incondicional produce una línea.
- **Diagrama de dispersión del ajuste.** Se muestra un diagrama de dispersión de las proximidades transformadas frente a las distancias. Si se especifican varias fuentes, se asignará un color distinto a cada fuente.
- **Gráficos de los residuos.** Se muestra un diagrama de dispersión de las proximidades transformadas frente a los residuos (proximidades transformadas menos las distancias). Si se especifican varias fuentes, se asignará un color distinto a cada fuente.

Estilos de objetos fila. Permiten tener un mayor control de la visualización de los objetos de fila en los gráficos. Los valores de la variable de colores opcional se utilizan para mostrar sucesivamente todos los colores. Los valores de la variable de marcadores opcionales se utilizan para mostrar sucesivamente todos los posibles marcadores.

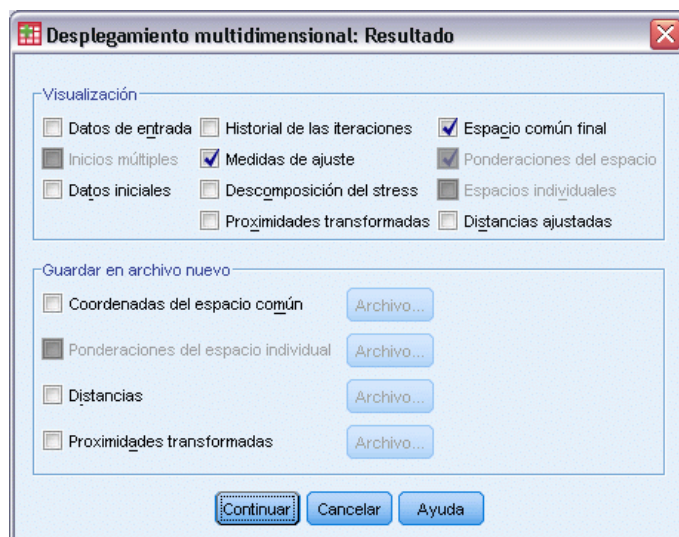
Gráficos de fuentes. Para los gráficos de espacios individuales, de diagrama de dispersión del ajuste y de residuos —y en caso de que las transformaciones se apliquen por fuente, para los gráficos de transformación y de Shepard— se pueden especificar las fuentes para las que se deben generar los gráficos. Los números de fuentes introducidos deben ser valores de la variable de fuentes especificada en el cuadro de diálogo principal y el rango debe estar entre 1 y el número de fuentes.

Gráficos de las filas. Si se aplican transformaciones por fila, para los gráficos de transformación y gráficos de Shepard, puede especificar la fila para la que se deben generar los gráficos. Los números de fila especificados deben estar comprendidos entre 1 y el número de filas.

Resultados del despliegamiento multidimensional

El cuadro de diálogo Resultados permite controlar la cantidad de resultados mostrados y guardar parte en archivos independientes.

Figura 8-6
Cuadro de diálogo Resultados



Mostrar. Seleccione una o varias de las siguientes opciones para su visualización:

- **Datos de entrada.** Incluye las proximidades originales y, si aparecen, las ponderaciones de los datos, la configuración inicial y las coordenadas fijas.
- **Inicios múltiples.** Muestra la semilla de aleatorización y el valor de la tensión penalizada de cada inicio aleatorio.
- **Datos iniciales.** Muestra las coordenadas del espacio común inicial.
- **Historial de iteraciones.** Muestra el historial de iteraciones del algoritmo principal.
- **Medidas de ajuste.** Muestra diferentes medidas. La tabla contiene varias medidas de bondad de ajuste, maldad de ajuste, correlación, variación y medidas de no degeneración.
- **Descomposición del stress.** Muestra la descomposición de objetos, filas y fuentes de tensión penalizada, incluidas las desviaciones típicas y las medias de fila, columna y fuente.
- **Proximidades transformadas.** Muestra las proximidades transformadas.
- **Espacio común final.** Muestra las coordenadas del espacio común.
- **Ponderaciones del espacio.** Muestra las ponderaciones del espacio individual. Esta opción está disponible únicamente si se especifica uno de los modelos de diferencias individuales. En función del modelo, las ponderaciones del espacio se descomponen en ponderaciones de rotación y de dimensión, que también se muestran.
- **Espacios individuales.** Se muestran las coordenadas de los espacios individuales. Esta opción está disponible únicamente si se especifica uno de los modelos de diferencias individuales.
- **Distancias ajustadas.** Muestra las distancias entre los objetos en la configuración.

Guardar en archivo nuevo. Puede guardar las coordenadas del espacio común, las ponderaciones del espacio individual, las distancias y las proximidades transformadas en distintos archivos de datos de IBM® SPSS® Statistics.

Funciones adicionales del comando PREFSCAL

Se puede personalizar el despliegamiento multidimensional del análisis de proximidades si se pegan las selecciones en una ventana de sintaxis y se edita la sintaxis del comando `PREFSCAL` resultante. Con el lenguaje de sintaxis de comandos también podrá:

- Especifique varias listas de fuentes para gráficos de espacios individuales, diagramas de dispersión del ajuste y gráficos de residuos —y en el caso de transformaciones condicionales de matriz, para gráficos de transformación y gráficos de Shepard— cuando varias fuentes están disponibles (con el subcomando `PLOT`).
- Especifique varias listas de filas para los gráficos de transformación y los gráficos de Shepard en caso de transformaciones condicionales de filas (subcomando `PLOT`).
- Especifique un número de filas en lugar de una variable de ID de fila (con el subcomando `INPUT`).
- Especifique un número de fuentes en lugar de una variable de ID de fuente (con el subcomando `INPUT`).

Si desea información detallada sobre la sintaxis, consulte la referencia de sintaxis de comandos (*Command Syntax Reference*).

Parte II: Ejemplos

Regresión categórica

El objetivo de la regresión categórica con el escalamiento óptimo es describir la relación entre una variable de respuesta y un conjunto de predictores. Mediante la cuantificación de esta relación, los valores de la respuesta se pueden predecir para cualquier combinación de predictores.

En este capítulo, utilizaremos dos ejemplos para ilustrar los análisis que implica la regresión con escalamiento óptimo. El primer ejemplo utiliza un pequeño conjunto de datos para ilustrar los conceptos básicos. El segundo ejemplo utiliza un conjunto mucho más grande de variables y observaciones en un ejemplo práctico.

Ejemplo: Datos de limpiadores de alfombras

En un ejemplo muy conocido (Green y Wind, 1973), una compañía interesada en sacar al mercado un nuevo limpiador de alfombras desea examinar la influencia de cinco factores sobre la preferencia del consumidor: diseño del producto, marca comercial, precio, sello de *buen producto para el hogar* y garantía de devolución del importe. Hay tres niveles de factores para el diseño del producto, cada uno con una diferente colocación del cepillo del aplicador; tres nombres comerciales (*K2R*, *Glory* y *Bissell*); tres niveles de precios; y dos niveles (no o sí) para los dos últimos factores. La siguiente tabla muestra las variables utilizadas en el estudio de limpiadores de alfombras, con sus etiquetas de variable y valores.

Tabla 9-1
Variables explicativas en el estudio de limpiadores de alfombras

Nombre de variable	Etiqueta de variable	Etiqueta de valor
<i>formato</i>	Diseño del producto	A*, B*, C*
<i>marca</i>	Nombre comercial	K2R, Glory, Bissell
<i>precio</i>	Precio	1,19 \$, 1,39 \$, 1,59 \$
<i>sello</i>	Sello de <i>buen producto para el hogar</i>	No, sí
<i>importe</i>	Garantía de devolución del importe	No, sí

Diez consumidores clasificaron 22 perfiles definidos por estos factores. La variable *Preferencia* contiene el rango de las clasificaciones medias de cada perfil. Las clasificaciones inferiores corresponden a preferencias elevadas. Esta variable refleja una medida global de la preferencia de cada perfil. Mediante la regresión categórica, exploraremos cómo estos cinco factores están relacionados con la preferencia. Este conjunto de datos se puede encontrar en *carpet.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20.](#)

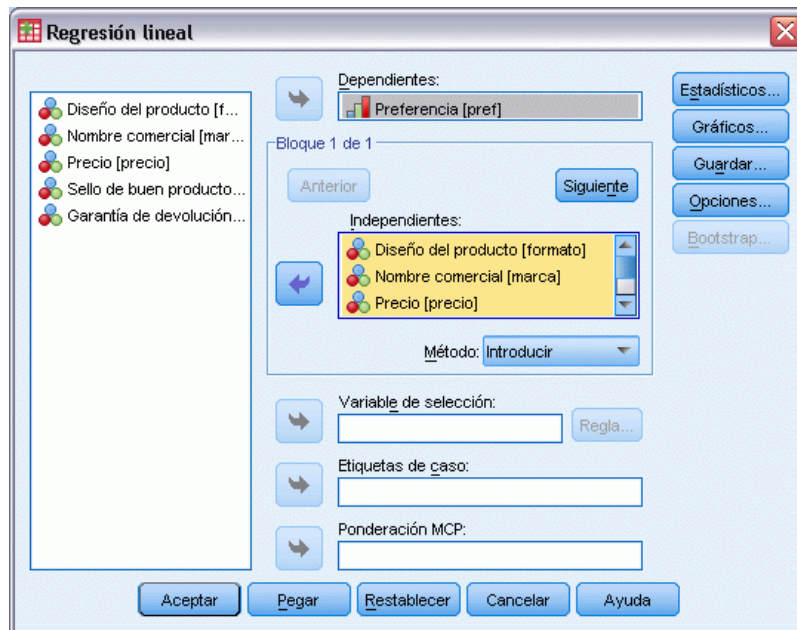
Análisis de regresión lineal ordinario

- ▶ Para generar resultados de regresión lineal típica, elija en los menús:
Analizar > Regresión > Lineales...

Nota: Esta función requiere la opción Statistics Base.

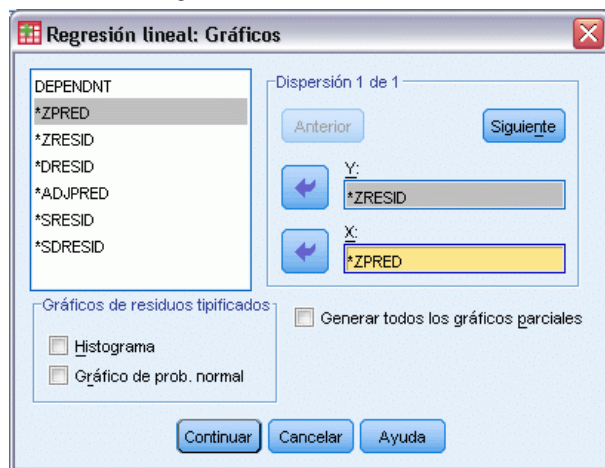
Figura 9-1

Cuadro de diálogo Regresión lineal



- ▶ Seleccione *Preferencia* como la variable dependiente.
- ▶ Seleccione desde *Diseño del producto* hasta *Garantía de devolución del importe* como las variables independientes.
- ▶ Pulse en Gráficos.

Figura 9-2
Cuadro de diálogo Gráficos



- ▶ Seleccione **ZRESID* como la variable del eje *Y*.
- ▶ Seleccione **ZPRED* como la variable del eje *X*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Guardar en el cuadro de diálogo Regresión lineal.

Figura 9-3
Cuadro de diálogo Guardar

- ▶ Seleccione Tipificados en el grupo Residuos.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Regresión lineal.

Resumen del modelo

Figura 9-4
Resumen del modelo para la regresión lineal típica

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,841 ^a	,707	,615	3,99810

a. Variables predictoras: (Constante), Garantía de devolución del importe, Precio, Sello de buen producto para el hogar, Nombre comercial, Diseño del producto

b. Variable dependiente: Preferencia

El método estándar para describir las relaciones en este problema es la regresión lineal. La medida más habitual de la bondad de ajuste del modelo de regresión a los datos es R^2 . Este estadístico representa la parte de la varianza de la respuesta que queda explicada por la combinación

ponderada de predictores. Cuanto más cerca esté R^2 de 1, mejor será el ajuste del modelo. La regresión de *Preferencia* sobre los cinco predictores genera una R^2 de 0,707, lo que indica que aproximadamente el 71% de la varianza de las clasificaciones de preferencia queda explicada por las variables predictoras en la regresión lineal.

Coeficientes

En la tabla se muestran los coeficientes tipificados. El signo del coeficiente indica si la respuesta pronosticada aumenta o disminuye cuando aumenta el predictor, manteniéndose constantes todos los demás predictores. En el caso de datos categóricos, la codificación de categorías determina el significado de un aumento en un predictor. Por ejemplo, un aumento de *Garantía de devolución del importe*, *Diseño del producto* o *Sello de buen producto para el hogar* producirán una disminución en la clasificación de preferencia pronosticada. *Garantía de devolución del importe* se codifica como 1 para *sin garantía de devolución del importe* y 2 para *garantía de devolución del importe*. Un aumento de *Garantía de devolución del importe* corresponde a la incorporación de una garantía de devolución del importe. Por tanto, al añadir una garantía de devolución del importe se reduce la clasificación de preferencia pronosticada, lo que corresponde a un aumento en la preferencia pronosticada.

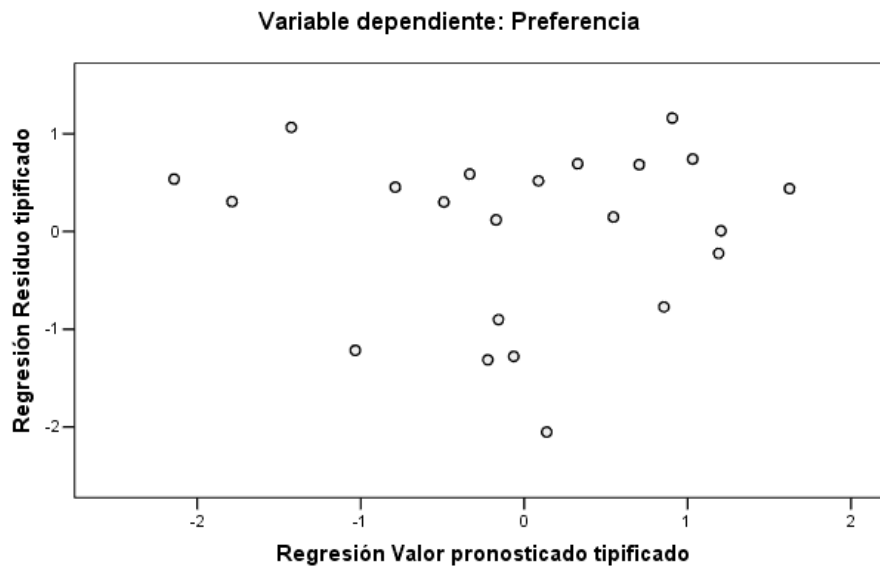
Figura 9-5
Coeficientes de regresión

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	22,529	5,177		4,352	,000
	Diseño del producto	-4,159	1,036	-,560	-4,015	,001
	Nombre comercial	,429	1,054	,056	,407	,689
	Precio	2,703	1,009	,366	2,681	,016
	Sello de buen producto para el hogar	-4,314	1,780	-,330	-2,423	,028
	Garantía de devolución del importe	-2,779	1,921	-,197	-1,447	,167

El valor del coeficiente refleja la cantidad de cambio en la clasificación de preferencia pronosticada. Mediante coeficientes tipificados, las interpretaciones se basan en las desviaciones típicas de las variables. Cada coeficiente indica el número de desviaciones típicas que la respuesta pronosticada cambia para un cambio de una desviación típica en un predictor, si todos los demás predictores permanecen constantes. Por ejemplo, un cambio de una desviación típica en *Nombre comercial* produce un aumento en la preferencia pronosticada de 0,056 desviaciones típicas. La desviación típica de *Preferencia* es 6,44, por lo que *Preferencia* aumenta en $0,056 \times 6,44 = 0,361$. Los cambios en *Diseño del producto* producen los mayores cambios en la preferencia pronosticada.

Diagrama de dispersión de los residuos

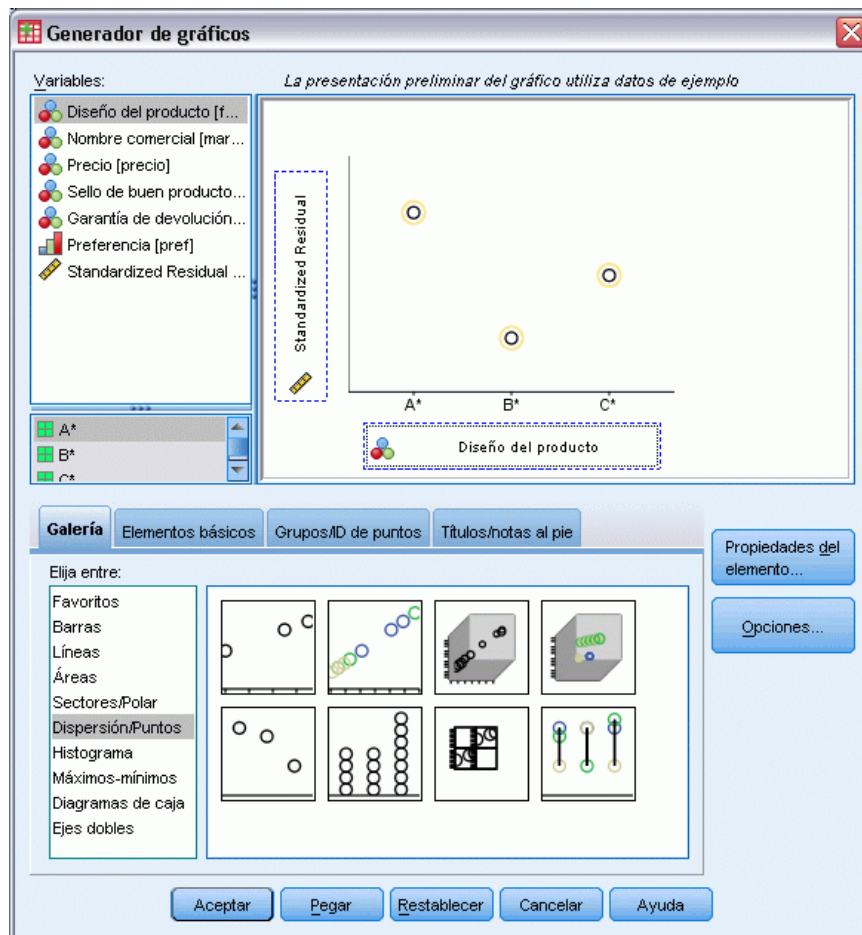
Figura 9-6
Residuos frente a valores pronosticados



Los residuos tipificados se representan frente a los valores pronosticados tipificados. No debe aparecer ningún patrón si el modelo se ajusta correctamente. Aquí vemos una forma de U en la que tanto los valores pronosticados bajos como altos tienen residuos positivos. Los valores pronosticados tipificados cercanos a 0 tienden a tener residuos negativos.

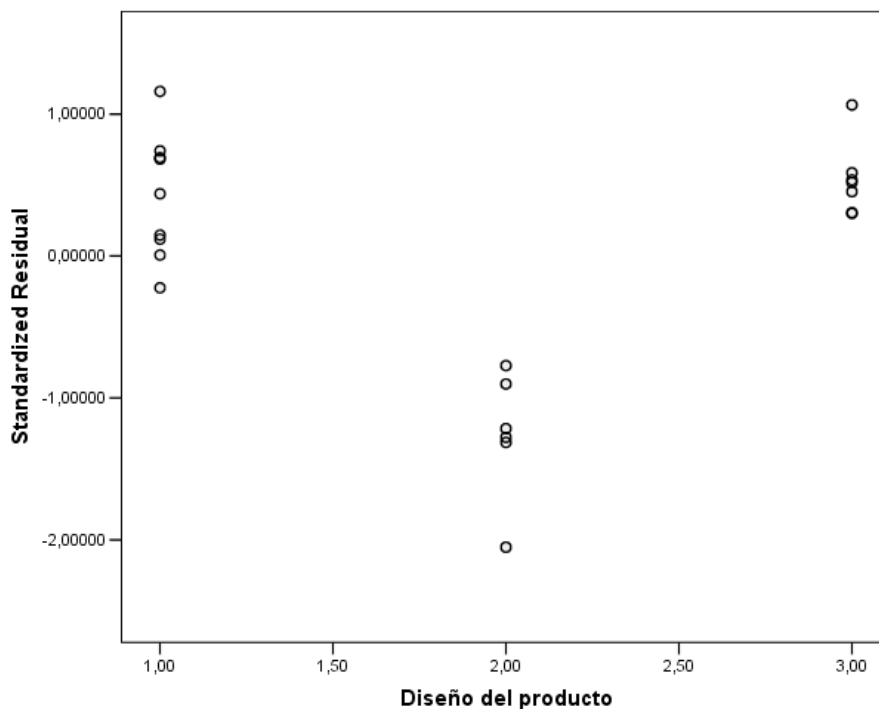
- Para generar un diagrama de dispersión de los residuos por el predictor *Diseño del producto*, elija en los menús:
Gráficos > Generador de gráficos...

Figura 9-7
Generador de gráficos



- ▶ Seleccione la galería Dispersión/Puntos y elija Dispersión simple.
- ▶ Seleccione *Residuo tipificado* como variable *y* y *Diseño del producto* como variable *x*.
- ▶ Pulse en Aceptar.

Figura 9-8
Residuos frente a diseño del producto



La forma de U es más pronunciada en el gráfico de los residuos tipificados frente al formato. Todos los residuos del diseño B* son negativos, mientras que todos menos uno de los residuos son positivos para los otros dos diseños. Como el modelo de regresión lineal ajusta un parámetro para cada variable, no es posible capturar la relación mediante el método típico.

Análisis de regresión categórica

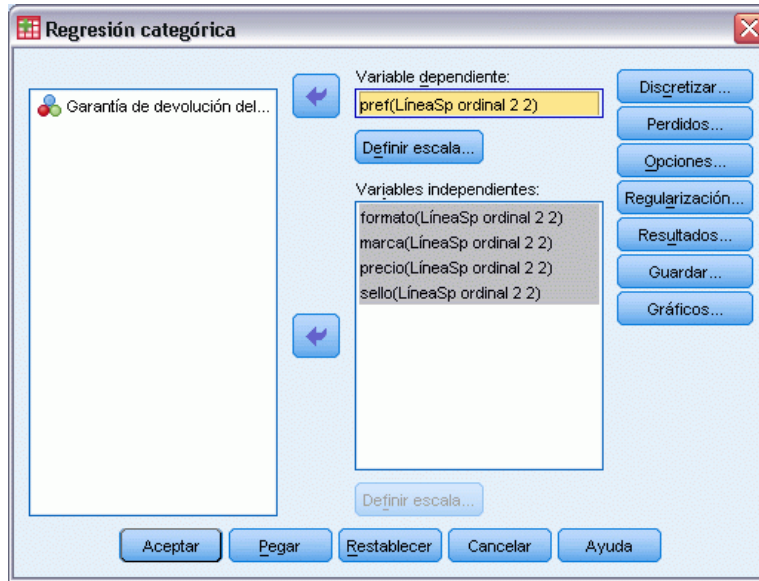
La naturaleza categórica de las variables y la relación no lineal entre *Preferencia* y *Diseño del producto* sugiere que la regresión sobre las puntuaciones óptimas puede funcionar mejor que la regresión típica. La forma de U de los gráficos de residuos indica que se debe utilizar un tratamiento nominal de *Diseño del producto*. Todos los demás predictores se tratarán a nivel de escalamiento numérico.

La variable de respuesta justifica una consideración especial. Desea pronosticar los valores de *Preferencia*. Por tanto, resulta deseable recuperar tantas propiedades de sus categorías como sea posible en las cuantificaciones. El uso del nivel de escalamiento ordinal o nominal ignora las diferencias entre las categorías de respuesta. Sin embargo, la transformación lineal de las categorías de respuesta conserva las diferencias de categoría. Por consiguiente, se prefiere en general el escalamiento numérico de la respuesta y se utilizará en este caso.

Ejecución del análisis

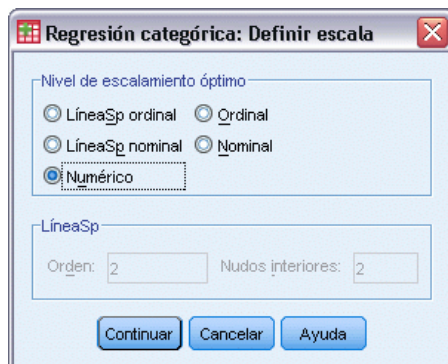
- ▶ Para ejecutar un análisis de Regresión categórica, elija en los menús: Analizar > Regresión > Escalamiento óptimo (CATREG)...

Figura 9-9
Cuadro de diálogo Regresión categórica



- ▶ Seleccione *Preferencia* como la variable dependiente.
- ▶ Seleccione desde *Diseño del producto* hasta *Garantía de devolución del importe* como las variables independientes.
- ▶ Seleccione *Preferencia* y pulse en Definir escala.

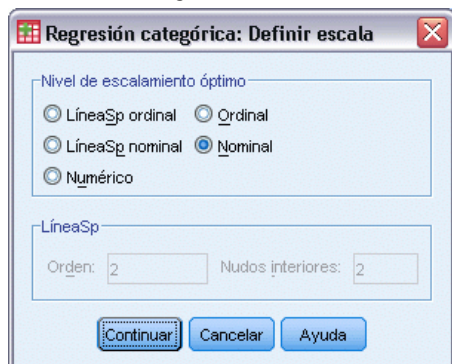
Figura 9-10
Cuadro de diálogo Definir escala



- ▶ Seleccione Numérico como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.

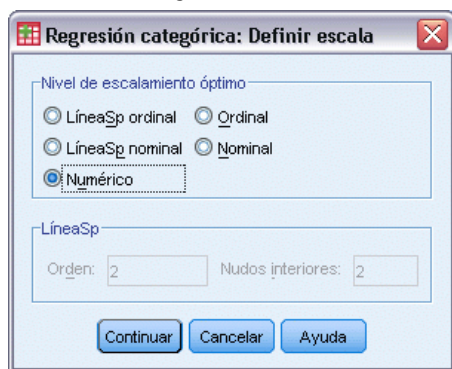
- ▶ Seleccione *Diseño del producto* y pulse en Definir escala en el cuadro de diálogo Regresión categórica.

Figura 9-11
Cuadro de diálogo Definir escala



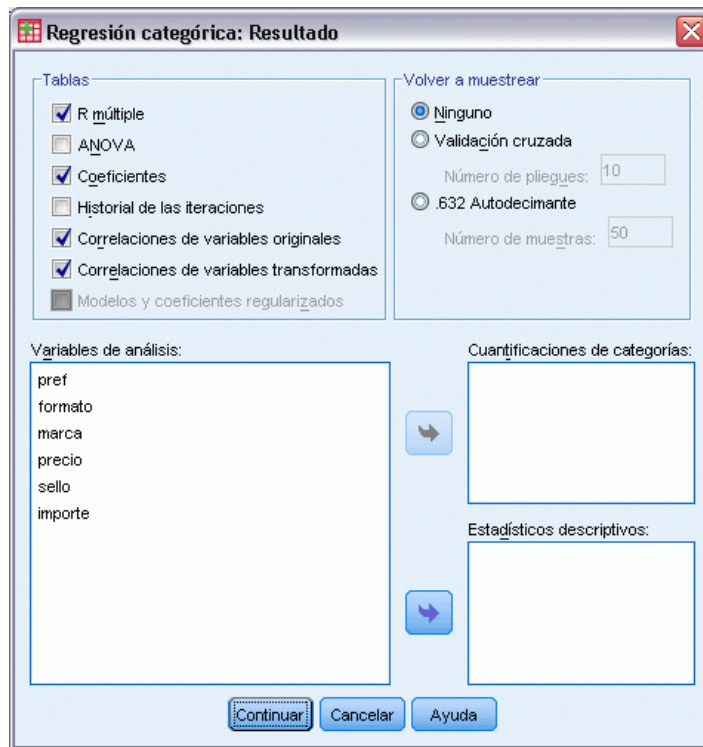
- ▶ Seleccione Nominal como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Seleccione desde *Nombre comercial* hasta *Garantía de devolución del importe* y pulse en Definir escala en el cuadro de diálogo Regresión categórica.

Figura 9-12
Cuadro de diálogo Definir escala



- ▶ Seleccione Numérico como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Pulse en Resultados en el cuadro de diálogo Regresión categórica.

Figura 9-13
Cuadro de diálogo Resultados



- ▶ Seleccione Correlaciones de variables originales y Correlaciones de variables transformadas.
- ▶ Anule la selección de ANOVA.
- ▶ Pulse en Continuar.
- ▶ Pulse en Guardar en el cuadro de diálogo Regresión categórica.

Figura 9-14
Cuadro de diálogo Guardar

Regresión categórica: Guardar

Guardar valores pronosticados en el conjunto de datos activo
 Guardar residuos en el conjunto de datos activo

Datos discretizados

Crear datos discretizados
 Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
 Escribir un nuevo archivo de datos

Modelos y coeficientes regularizados

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
 Escribir un nuevo archivo de datos

Variables transformadas

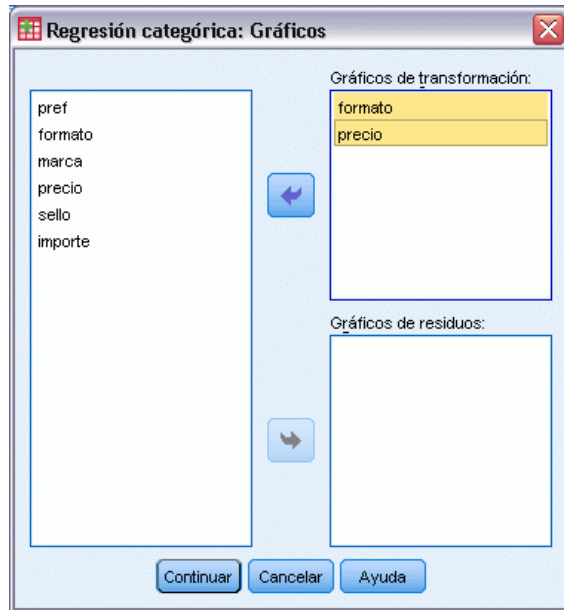
Guardar variables transformadas en el conjunto de datos activo
 Guardar las variables transformadas a un nuevo conjunto de datos o archivo
 Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
 Escribir un nuevo archivo de datos

Signos de coeficientes de regresión

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:
 Escribir un nuevo archivo de datos

- ▶ Seleccione Guardar residuos en el conjunto de datos activo.
- ▶ Seleccione Guardar variables transformadas en el conjunto de datos activo en el grupo Variables transformadas.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Regresión categórica.

Figura 9-15
Cuadro de diálogo Gráficos



- ▶ Elija crear gráficos de transformación para *Diseño del producto* y *Precio*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Regresión categórica.

Intercorrelaciones

Las intercorrelaciones entre los predictores son útiles para identificar la multicolinealidad en la regresión. Las variables que están muy correlacionadas producirán estimaciones inestables de la regresión. Sin embargo, debido a su alta correlación, la omisión de una de ellas del modelo sólo afecta de manera mínima al pronóstico. La varianza de la respuesta que se puede explicar mediante la variable omitida sigue estando explicada por la variable correlacionada que permanece. Sin embargo, las correlaciones de orden cero son sensibles a los valores atípicos y además tampoco puede identificar la multicolinealidad debido a una alta correlación entre un predictor y una combinación de otros predictores.

Figura 9-16
Correlaciones de los predictores originales

	Diseño del producto	Nombre comercial	Precio	Sello de buen producto para el hogar	Garantía de devolución del importe
Diseño del producto	1,000	-,189	-,126	,081	,066
Nombre comercial	-,189	1,000	,065	-,042	-,034
Precio	-,126	,065	1,000	,000	,000
Sello de buen producto para el hogar	,081	-,042	,000	1,000	-,039
Garantía de devolución del importe	,066	-,034	,000	-,039	1,000
Dimensión	1	2	3	4	5
Autovalores	1,291	1,038	,980	,905	,785

Figura 9-17
Correlaciones de los predictores transformados

	Diseño del producto	Nombre comercial	Precio	Sello de buen producto para el hogar	Garantía de devolución del importe
Diseño del producto	1,000	-,156	-,089	,032	,102
Nombre comercial	-,156	1,000	,065	-,042	-,034
Precio	-,089	,065	1,000	,000	,000
Sello de buen producto para el hogar	,032	-,042	,000	1,000	-,039
Garantía de devolución del importe	,102	-,034	,000	-,039	1,000
Dimensión	1	2	3	4	5
Autovalores	1,248	1,043	,983	,905	,821

Se muestran las intercorrelaciones de los predictores tanto para los predictores transformados como sin transformar. Todos los valores están cerca de 0, lo que indica que la multicolinealidad entre variables individuales no supone una preocupación.

Observe que sólo las correlaciones que cambian implican a *Diseño del producto*. Como el resto de los predictores se tratan numéricamente, las diferencias entre las categorías y el orden de las categorías se conservan para estas variables. Por tanto, las correlaciones no pueden cambiar.

Coefficientes y ajuste del modelo

El procedimiento Regresión categórica produce una R^2 de 0,948, lo que indica que casi el 95% de la varianza de las clasificaciones de preferencia transformada queda explicada por la regresión sobre los predictores transformados óptimamente. La transformación de los predictores mejora el ajuste respecto al método típico.

Figura 9-18
Resumen del modelo de regresión categórica

R múltiple	R cuadrado	R cuadrado corregida
,974	,948	,927

Variable dependiente: Preferencia
Predictores: Diseño del producto Nombre comercial Precio Sello de buen producto para el hogar Garantía de devolución del importe

La siguiente tabla muestra los coeficientes de regresión tipificados. La regresión categórica tipifica las variables, por lo que sólo se informará de los coeficientes tipificados. Estos valores aparecen divididos por sus correspondientes errores típicos, lo que produce una prueba F para cada variable. Sin embargo, la prueba para cada variable depende del resto de los predictores que haya en el modelo. Es decir, la prueba determina si la omisión de una variable predictora del modelo manteniendo todos los demás predictores empeorará significativamente las capacidades de pronóstico del modelo. Estos valores no se deben utilizar para omitir varias variables de una vez para un modelo posterior. Además, los mínimos cuadrados alternantes optimizan las cuantificaciones, lo que implica que estas pruebas deben interpretarse de manera conservadora.

Figura 9-19
Coeficientes tipificados de los predictores transformados

	Coeficientes tipificados		gl	F	Sig.
	Beta	Error típ.			
Diseño del producto	-,748	,060	2	155,289	,000
Nombre comercial	,045	,060	1	,578	,459
Precio	,371	,059	1	39,312	,000
Sello de buen producto para el hogar	-,350	,059	1	35,299	,000
Garantía de devolución del importe	-,159	,059	1	7,175	,017

Variable dependiente: Preferencia

El coeficiente mayor aparece para *Diseño del producto*. Un aumento de una desviación estándar en *Diseño del producto* produce una disminución de desviación típica de 0,748 en la clasificación de preferencia pronosticada. Sin embargo, *Diseño del producto* se trata de manera nominal, por lo que un aumento en las cuantificaciones no tiene por qué corresponder con un aumento en los códigos de las categorías originales.

Los coeficientes tipificados suelen interpretarse como un reflejo de la importancia de cada predictor. Sin embargo, los coeficientes de regresión no pueden describir completamente el impacto de un predictor ni las relaciones entre los predictores. Hay que utilizar otros estadísticos junto con los coeficientes tipificados para explorar completamente los efectos de los predictores.

Correlaciones e importancia

Para interpretar las contribuciones de los predictores a la regresión, no basta con examinar únicamente los coeficientes de regresión. Además, es necesario examinar las correlaciones, las correlaciones parciales y las correlaciones semiparciales. La siguiente tabla contiene estas medidas de correlación para cada variable.

La correlación de orden cero es la correlación entre el predictor transformado y la respuesta transformada. Para este dato, la correlación más grande aparece para *Diseño del producto*. Sin embargo, si puede explicar parte de la variación en el predictor o en la respuesta, obtendrá una mejor representación de lo bien que funciona el predictor.

Figura 9-20
Correlaciones de orden cero, semiparciales y parciales (variables transformadas)

	Correlaciones			Importancia	Tolerancia	
	Orden cero	Parcial	Semiparcial		Después de la transformación	Antes de la transformación
Diseño del producto	-,816	-,955	-,733	,644	,959	,942
Nombre comercial	,206	,193	,045	,010	,971	,961
Precio	,440	,851	,369	,172	,989	,982
Sello de buen producto para el hogar	-,370	-,838	-,349	,137	,996	,991
Garantía de devolución del importe	-,223	-,569	-,158	,037	,987	,993

Variable dependiente: Preferencia

Las demás variables del modelo pueden afectar negativamente a la capacidad de un determinado predictor de pronosticar la respuesta. El coeficiente de correlación parcial elimina los efectos lineales de los demás predictores tanto del predictor como de la respuesta. Esta medida iguala la correlación entre los residuos de la regresión del predictor respecto a los demás predictores y los residuos de la regresión de la respuesta respecto a los otros predictores. La correlación parcial de los cuadrados corresponde a la proporción de la varianza explicada relativa a la varianza residual de la respuesta que queda tras eliminar los efectos de las demás variables. Por ejemplo, *Diseño del producto* tiene una correlación parcial de $-0,955$. Si se eliminan los efectos de las demás variables, *Diseño del producto* explica $(-0,955)^2 = 0,91 = 91\%$ de la variación en las clasificaciones de preferencia. Tanto *Precio* como *Sello de buen producto para el hogar* también explican una gran parte de la varianza si se eliminan los efectos de las demás variables.

Como alternativa a la eliminación de los efectos de las variables de la respuesta y de un predictor, puede eliminar los efectos de únicamente el predictor. La correlación entre la respuesta y los residuos de la regresión de un predictor respecto a los otros predictores es la correlación semiparcial. Si elevamos este valor al cuadrado obtenemos una medida de la proporción de la varianza explicada relativa a la varianza total de la respuesta. Si elimina los efectos de *Nombre comercial*, *Sello de buen producto para el hogar*, *Garantía de devolución del importe* y *Precio* de *Diseño del producto*, la parte restante de *Diseño del producto* explica $(-0,733)^2 = 0,54 = 54\%$ de la variación de las clasificaciones de preferencia.

Importancia

Además de los coeficientes de regresión y de las correlaciones, la medida de Pratt de la importancia relativa (Pratt, 1987) también ayuda a interpretar las contribuciones de los predictores a la regresión. Las importancias individuales elevadas relativas a las otras importancias corresponden a predictores que son fundamentales para la regresión. Además, la presencia de variables supresoras queda indicada por una baja importancia para una variable que tiene un coeficiente de tamaño similar a los predictores importantes.

En contraste con los coeficientes de regresión, esta medida define la importancia de los predictores de manera aditiva, es decir, la importancia de un conjunto de predictores es la suma de las importancias individuales de cada uno de los predictores. La medida de Pratt es igual al producto del coeficiente de regresión y la correlación de orden cero de un predictor. Estos productos se suman a R^2 , por lo que si se dividen por R^2 , producen una suma de 1. El conjunto de productores *Diseño del producto* y *Nombre comercial*, por ejemplo, tiene una importancia de 0,654. La importancia más grande corresponde a *Diseño del producto*, con *Diseño del producto*,

Precio y Sello de buen producto para el hogar y explica el 95% de la importancia para esta combinación de predictores.

Multicolinealidad

Las correlaciones grandes entre los predictores reducirán drásticamente la estabilidad de un modelo de regresión. Los predictores correlacionados producirán estimaciones inestables de los parámetros. La tolerancia refleja hasta qué punto están relacionadas linealmente las variables independientes entre sí. La tolerancia de una variable es la proporción de su varianza no explicada por las otras variables independientes de la ecuación. Si los demás predictores pueden explicar gran parte de la varianza de un predictor, dicho predictor no es necesario en el modelo. Un valor de la tolerancia cercano a 1 indica que la variable no se puede pronosticar muy bien a partir de los demás predictores. Por el contrario, una variable con una tolerancia muy baja contribuye con poca información a un modelo y puede causar problemas de cálculo. Además, valores grandes negativos de la medida de importancia de Pratt indican multicolinealidad.

Todas las medidas de tolerancia son muy altas. Ninguno de los predictores queda pronosticado muy bien por los demás predictores, por lo que no existe multicolinealidad.

Gráficos de transformación

La representación de los valores de categoría originales respecto a sus correspondientes cuantificaciones puede mostrar tendencias que podrían no detectarse en una lista con las cuantificaciones. A estos gráficos se les denomina con frecuencia gráficos de transformación. Hay que prestar atención a las categorías que reciben cuantificaciones similares. Estas categorías afectan a la respuesta pronosticada de la misma manera. Sin embargo, el tipo de transformación determina el aspecto básico del gráfico.

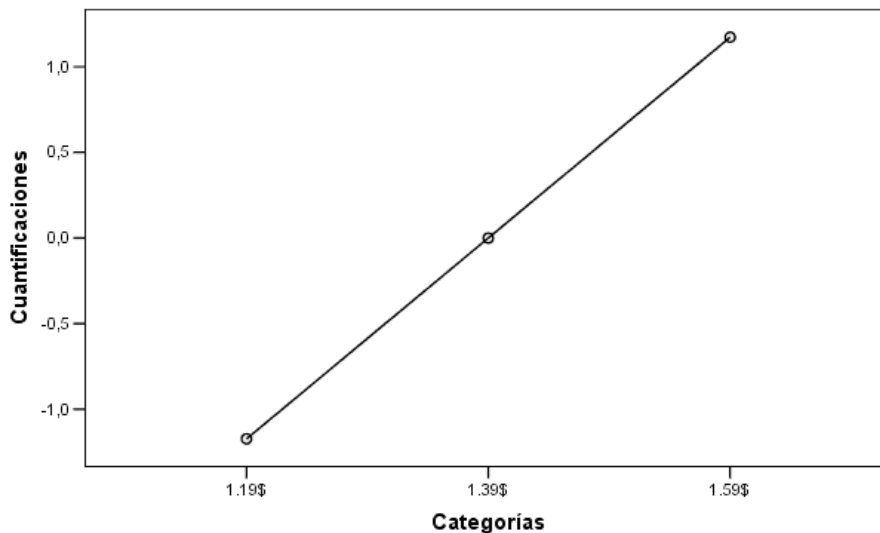
Las variables tratadas como resultado numérico en una relación lineal entre las cuantificaciones y las categorías originales, correspondientes a una línea recta en el gráfico de transformación. El orden y la diferencia entre las categorías originales se conservan en las cuantificaciones.

El orden de las cuantificaciones para las variables tratadas como ordinales corresponde al orden de las categorías originales. Sin embargo, no se conservan las diferencias entre las categorías. Como resultado, el gráfico de transformación no es decreciente pero no tiene por qué ser una línea recta. Si categorías consecutivas corresponden a cuantificaciones similares, la distinción de categorías tal vez no sea necesaria y se puedan combinar las categorías. Tales categorías producen una meseta en el gráfico de transformación. Sin embargo, este patrón también puede obtenerse imponiendo una estructura ordinal en una variable que deba tratarse como nominal. Si el tratamiento nominal posterior de la variable muestra el mismo patrón, la combinación de categorías está justificada. Además, si las cuantificaciones de una variable tratada como ordinal quedan situadas a lo largo de una línea recta, es posible que resulte más adecuada una transformación numérica.

Para variables tratadas como nominales, el orden de las categorías a lo largo del eje horizontal corresponde al orden de los códigos utilizado para representar las categorías. Las interpretaciones del orden de las categorías o de la distancia entre las categorías carecen de fundamento. El gráfico puede tomar cualquier forma lineal o no. Si aparece una tendencia creciente, debería intentarse un tratamiento ordinal. Si el gráfico de transformación nominal muestra una tendencia nominal, puede que sea más adecuada una transformación numérica.

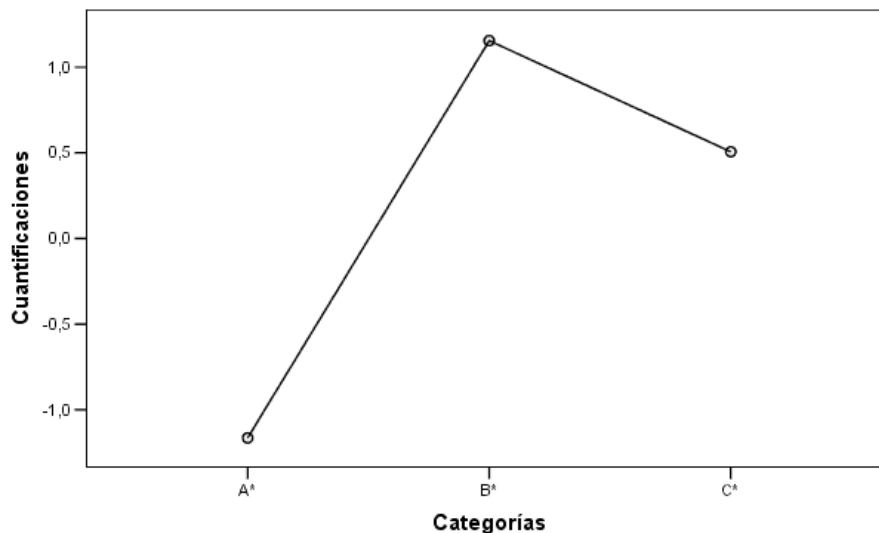
La siguiente figura muestra el gráfico de transformación para *Precio*, que se trató como numérica. Observe que el orden de las categorías a lo largo de la línea recta corresponde al orden de las categorías originales. Además, la diferencia entre las cuantificaciones para 1,19 \$ y 1,39 \$ (-1,173 y 0) es la misma que la diferencia entre las cuantificaciones para 1,39 \$ y 1,59 \$ (0 y 1.173). El hecho de que las categorías 1 y 3 se encuentran a la misma distancia de la categoría 2 se conserva en las cuantificaciones.

Figura 9-21
Transformación del gráfico del precio (numérico)



La transformación nominal de *Diseño del producto* produce el siguiente gráfico de transformación. Observe la nítida forma no lineal en la que la segunda categoría tiene la mayor cuantificación. En términos de la regresión, la segunda categoría disminuye la clasificación de preferencia pronosticada, mientras que las categorías primera y tercera tienen el efecto opuesto.

Figura 9-22
Gráfico de transformación para diseño del producto (nominal)

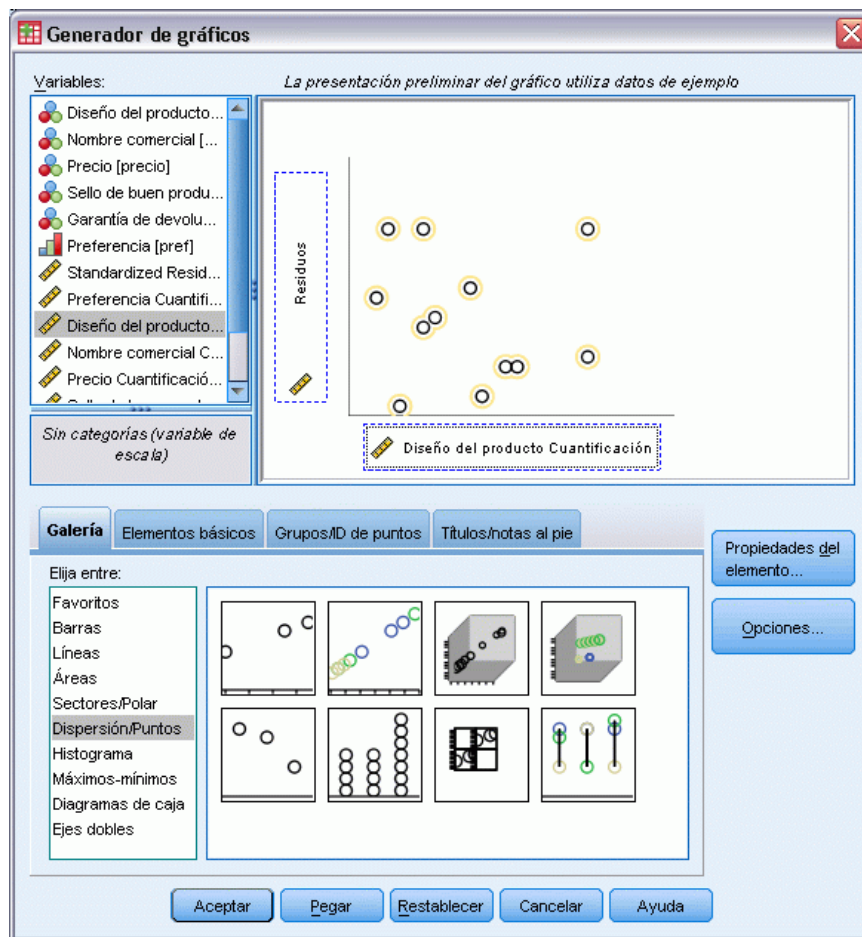


Análisis de los residuos

El uso de los datos transformados y de los residuos que se han guardado en el archivo de datos de trabajo permite crear un diagrama de dispersión de los valores pronosticados por los valores transformados de *Diseño del producto*.

Para obtener este diagrama de dispersión, vuelva a Generador de gráficos y pulse en Restablecer para borrar las selecciones anteriores y restaurar las opciones por defecto.

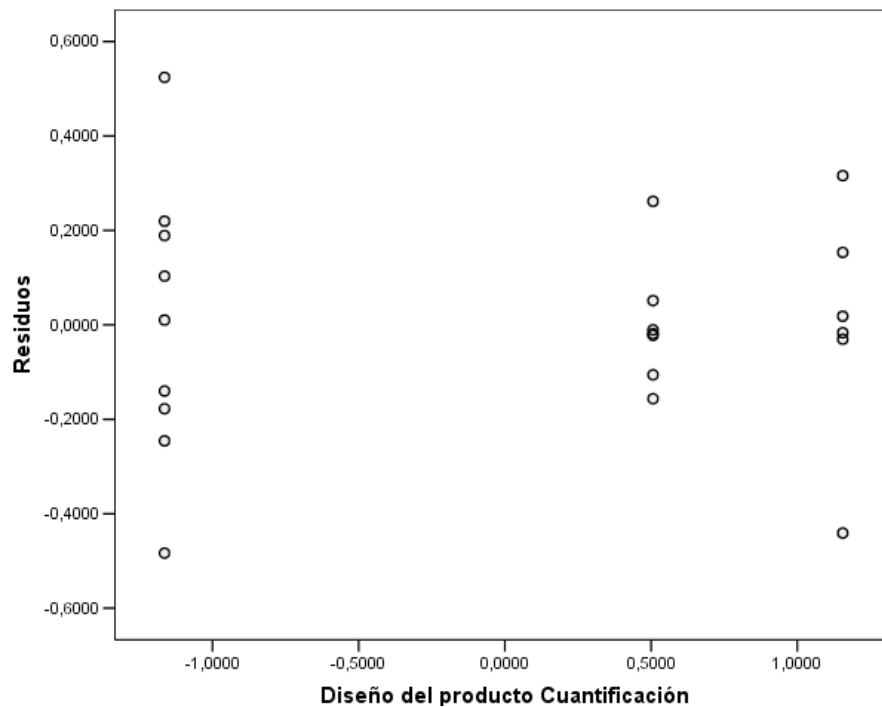
Figura 9-23
Generador de gráficos



- ▶ Seleccione la galería Dispersión/Puntos y elija Dispersión simple.
- ▶ Seleccione *Residual* como la variable del eje *Y*.
- ▶ Seleccione *Diseño del producto Cuantificación* como variable del eje *X*.
- ▶ Pulse en Aceptar.

El diagrama de dispersión muestra los residuos tipificados representados frente a las puntuaciones óptimas para *Diseño del producto*. Todos los residuos se encuentran comprendidos entre dos desviaciones típicas de 0. Una dispersión aleatoria de puntos sustituye a la forma de U que aparece en el diagrama de dispersión de la regresión lineal típica. Las posibilidades de pronóstico se mejoran mediante la cuantificación óptima de las categorías.

Figura 9-24
Residuos de la regresión categórica



Ejemplo: Datos sobre el ozono

En este ejemplo, utilizaremos un conjunto de datos de mayor tamaño para ilustrar la selección y los efectos de las transformaciones de escalamiento óptimo. Los datos incluyen 330 observaciones en seis variables meteorológicas anteriormente analizadas por Breiman y Friedman (Breiman y Friedman, 1985), y Hastie y Tibshirani (Hastie y Tibshirani, 1990), entre otros. La siguiente tabla describe las variables originales. La regresión categórica intenta pronosticar la concentración de ozono a partir de las variables restantes. Los investigadores anteriores han encontrado que no hay linealidad entre estas variables, lo que dificulta los métodos mediante regresión típica.

Tabla 9-2
Variables originales

Variable	Descripción
<i>ozono</i>	nivel diario de ozono; categorizado en una de 38 categorías
<i>abi</i>	altura de base de inversión
<i>gdp</i>	gradiente de presión (mm Hg)

Variable	Descripción
<i>vis</i>	visibilidad (millas)
<i>temp</i>	temperatura (grados F)
<i>dda</i>	día del año

Este conjunto de datos se puede encontrar en *ozone.sav*. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en *IBM SPSS Categories 20*.

Discretización de las variables

Si una variable tiene más categorías de las que se pueden interpretar en la práctica, deberá modificar las categorías mediante el cuadro de diálogo Discretización para reducir un rango de categorías a un número más manejable.

La variable *Día del año* tiene un valor mínimo de 3 y un valor máximo de 365. El uso de esta variable en una regresión categórica corresponde a utilizar una variable con 365 categorías. De manera similar, *Visibilidad (millas)* oscila de 0 a 350. Para simplificar la interpretación de los análisis, se pueden discretizar estas variables en intervalos iguales de longitud 10.

La variable *Altura de base de inversión* oscila de 111 a 5000. Una variable con este número de categoría produce relaciones muy complejas. Sin embargo, si se discretiza esta variable en intervalos iguales de longitud 100 se obtendrán aproximadamente 50 categorías. El uso de una variable de 50 categorías en vez de una variable de 5000 categorías simplifica las interpretaciones de manera significativa.

Gradiente de presión (mm Hg) oscila de -69 a 107. El procedimiento omite todas las categorías codificadas con números negativos del análisis, pero si se discretiza esta variable en intervalos iguales de longitud 10 se obtendrán aproximadamente 19 categorías.

Temperatura (grados F) oscila de 25 a 93 en la escala Fahrenheit. Para analizar los datos como si estuvieran en la escala Celsius, discretice esta variable en intervalos iguales de longitud 1,8.

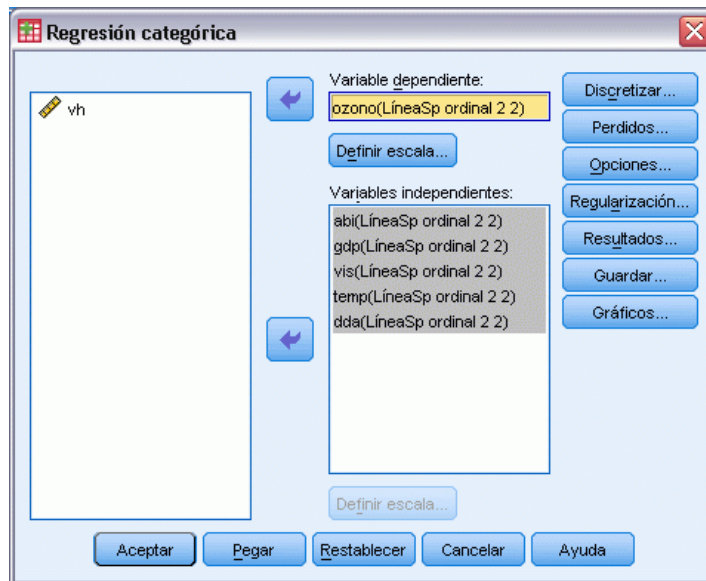
Puede ser deseable utilizar diferentes discretizaciones para las variables. Las decisiones que se han tomado aquí son puramente subjetivas. Si desea utilizar menos categorías, elija intervalos de mayor tamaño. Por ejemplo, *Día del año* se podría haber dividido en meses del año o en estaciones.

Selección del tipo de transformación

Cada variable se puede analizar a uno de varios niveles diferentes. Sin embargo, como el objetivo es el pronóstico de la respuesta, debe escalar la respuesta “tal cual” utilizando el nivel de escalamiento óptimo numérico. Por consiguiente, el orden y las diferencias entre las categorías se conservarán en la variable transformada.

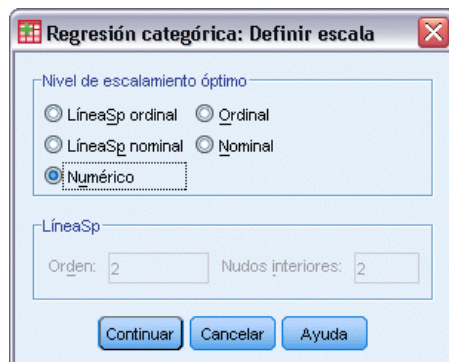
- Para ejecutar un análisis de Regresión categórica, elija en los menús:
Analizar > Regresión > Escalamiento óptimo (CATREG)...

Figura 9-25
Cuadro de diálogo Regresión categórica



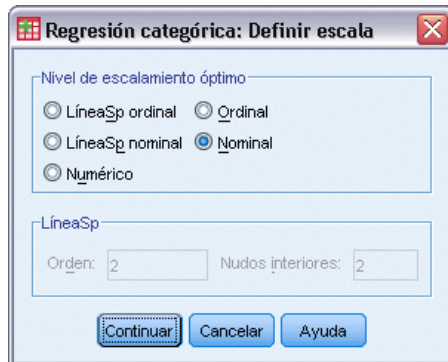
- ▶ Seleccione *Nivel diario de ozono* como la variable dependiente.
- ▶ Seleccione desde *Altura de base de inversión* hasta *Día del año* como las variables independientes.
- ▶ Seleccione *Nivel diario de ozono* y pulse en Definir escala.

Figura 9-26
Cuadro de diálogo Definir escala



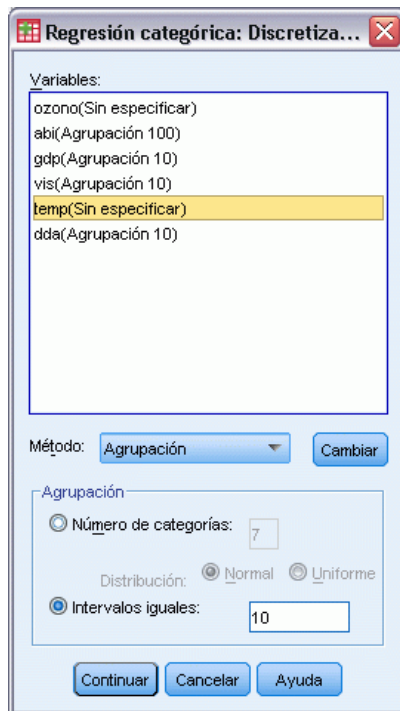
- ▶ Seleccione Numérico como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Seleccione desde *Altura de base de inversión* hasta *Día del año* y pulse en Definir escala en el cuadro de diálogo Regresión categórica.

Figura 9-27
Cuadro de diálogo Definir escala



- ▶ Seleccione Nominal como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Pulse en Discretizar en el cuadro de diálogo Regresión categórica.

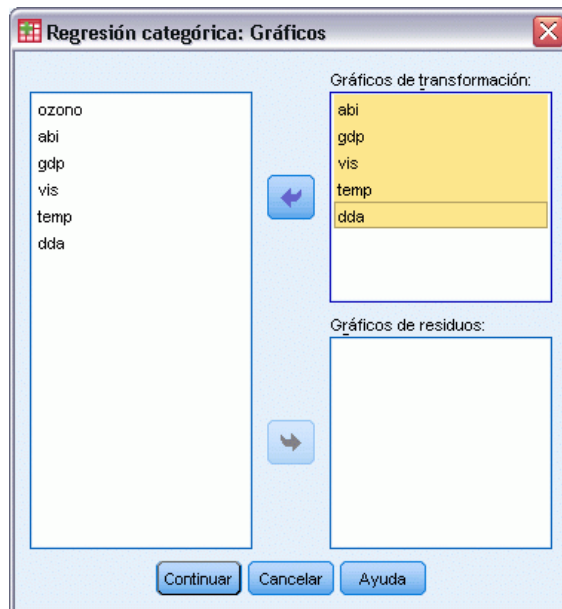
Figura 9-28
Cuadro de diálogo Discretización



- ▶ Seleccione *abi*.
- ▶ Seleccione Intervalos iguales y escriba 100 como longitud del intervalo.
- ▶ Pulse en Cambiar.
- ▶ Seleccione *dgp*, *vis* y *dda*.

- ▶ Escriba 10 como longitud del intervalo.
- ▶ Pulse en Cambiar.
- ▶ Seleccione *temp*.
- ▶ Escriba 1,8 como longitud del intervalo.
- ▶ Pulse en Cambiar.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Regresión categórica.

Figura 9-29
Cuadro de diálogo Gráficos



- ▶ Seleccione gráficos de transformación para las variables *Altura de base de inversión* hasta *Día del año*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Regresión categórica.

Figura 9-30
Resumen del modelo

	R múltiple	R cuadrado	R cuadrado corregida	Error de predicción aparente
Datos estandarizados	.938	.880	.785	.120

Variable dependiente: ozono
Predictores: abi gdp vis temp dda

El tratamiento de todos los predictores como nominales produce una R^2 de 0,880. Esta gran cantidad de varianza explicada no resulta sorprendente ya que el tratamiento nominal no impone ninguna restricción sobre las cuantificaciones. Sin embargo, la interpretación de los resultados puede ser bastante difícil.

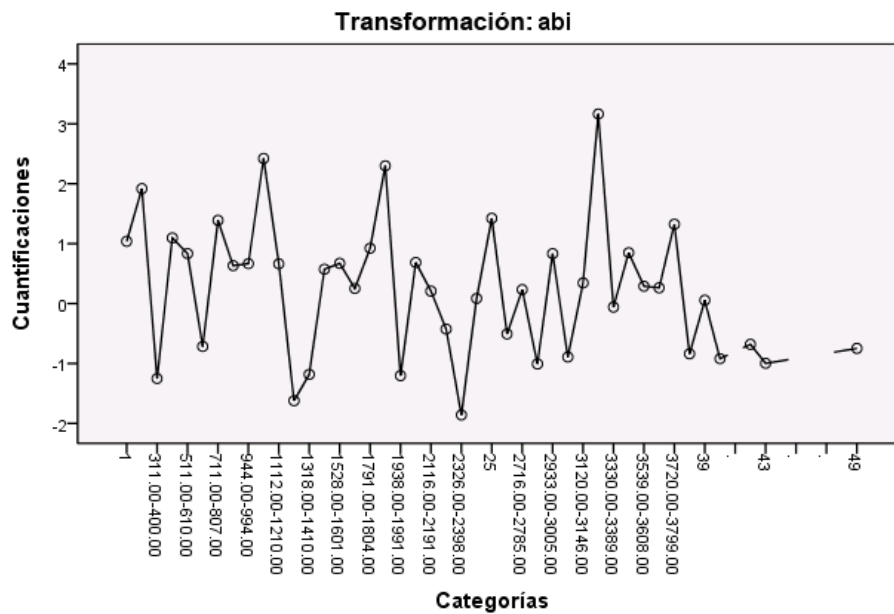
Figura 9-31
Coeficientes de regresión (todos los predictores son nominales)

	Coeficientes tipificados		gl	F	Sig.
	Beta	Bootstrap (1000) Estimación de error típico			
abi	.297	.053	42	31.047	.000
gdp	.326	.055	16	34.793	.000
vis	.229	.050	17	21.465	.000
temp	.577	.091	35	40.562	.000
dda	.420	.070	36	36.171	.000

Variable dependiente: ozono

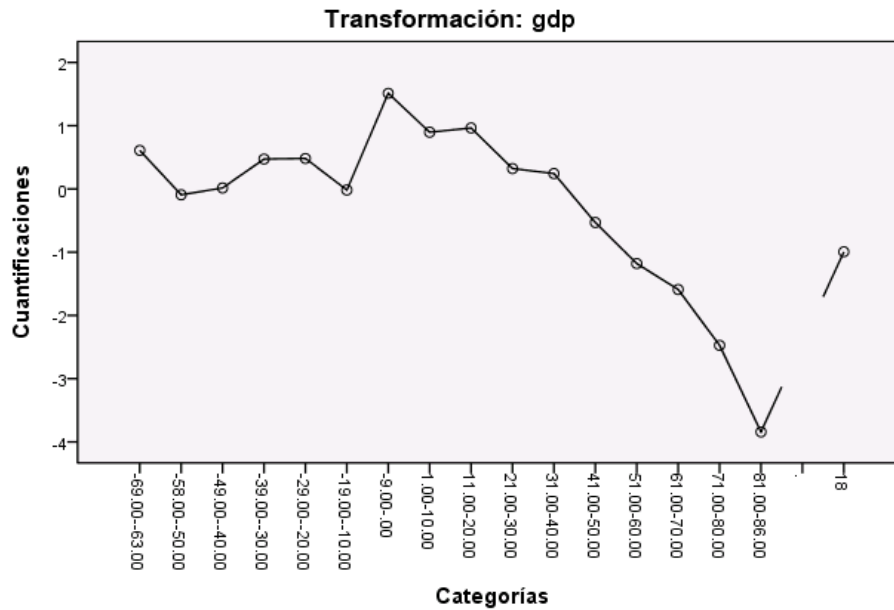
Esta tabla muestra los coeficientes de regresión tipificados de los predictores. Un error que se comete con frecuencia al interpretar estos valores radica en centrarse en los coeficientes mientras que se ignoran las cuantificaciones. No puede simplemente afirmar que, por ejemplo, un valor positivo de *Altura de base de inversión* implica que a medida que aumenta el predictor, aumenta el *Ozono* pronosticado. Todas las interpretaciones deben ser relativas a las variables transformadas, de modo que a medida que aumentan las cuantificaciones de *Altura de base de inversión*, aumenta el *Ozono* pronosticado. Para examinar los efectos de las variables originales, debe relacionar las categorías con las cuantificaciones.

Figura 9-32
Gráfico de transformación para altura de base de inversión (nominal)



El gráfico de transformación de *Altura de base de inversión* no parece mostrar ningún patrón. Como mostraba el aspecto irregular del gráfico, el desplazamiento de las categorías inferiores a las categorías superiores produce fluctuaciones en las cuantificaciones en ambas direcciones. Por tanto, la descripción de los efectos de esta variable requiere centrarse en las categorías individuales. La imposición de restricciones ordinales o lineales sobre las cuantificaciones de esta variable puede reducir considerablemente el ajuste.

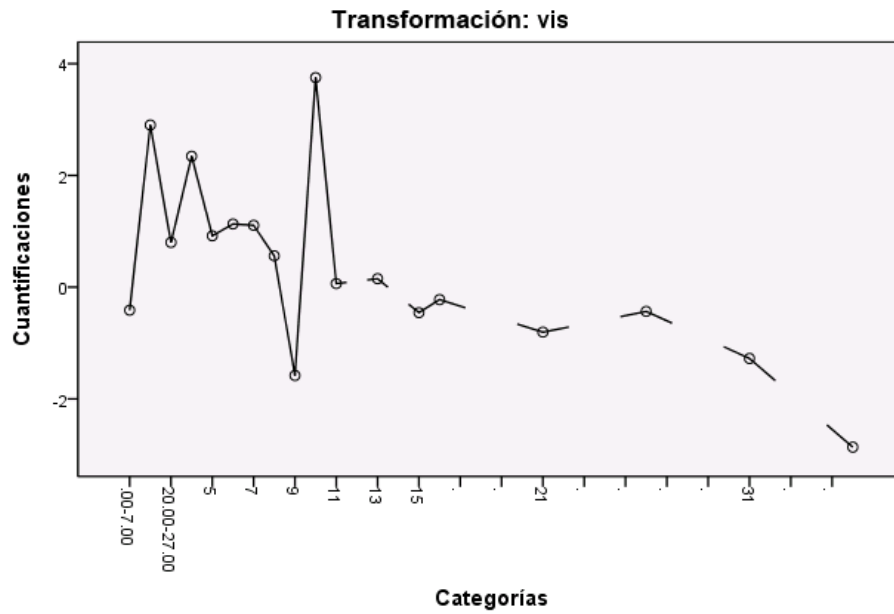
Figura 9-33
Gráfico de transformación de gradiente de presión (nominal)



Esta figura muestra el gráfico de transformación de *Gradiente de presión*. Las categorías discretizadas iniciales (desde 1 hasta 6) reciben pequeñas cuantificaciones y, por tanto, contribuyen mínimamente a la respuesta pronosticada. Las siguientes tres categorías reciben valores positivos algo más altos, lo que produce un aumento moderado del ozono pronosticado.

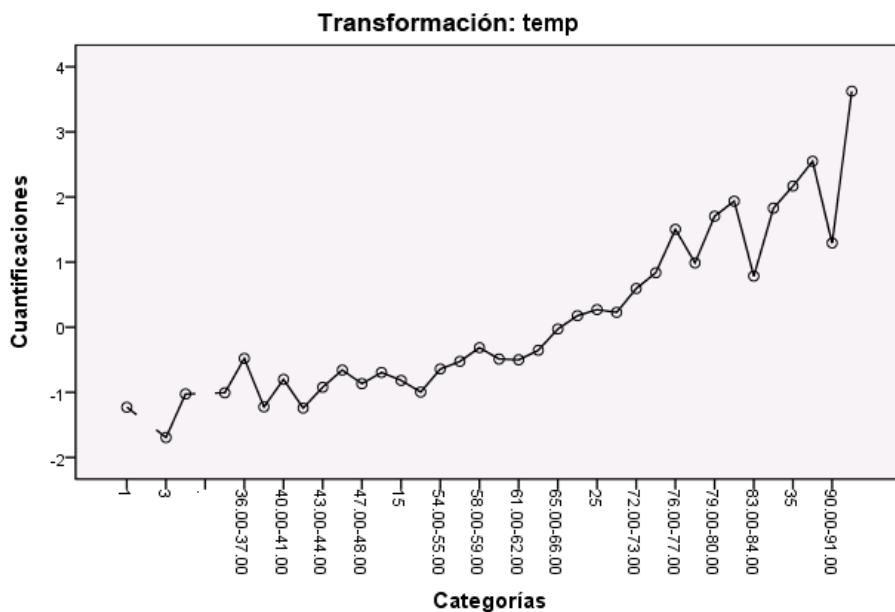
Las cuantificaciones disminuyen hasta la categoría 16, donde *Gradiente de presión* tiene el mayor efecto decreciente sobre el ozono pronosticado. Aunque la línea aumenta después de esta categoría, el uso de un nivel de escalamiento ordinal para *Gradiente de presión* puede que no reduzca de manera significativa el ajuste, al tiempo que simplifica las interpretaciones de los efectos. Sin embargo, la medida de la importancia de 0,04 y el coeficiente de regresión de *Gradiente de presión* indican que esta variable no es muy útil en la regresión.

Figura 9-34
Gráfico de transformación de visibilidad (nominal)



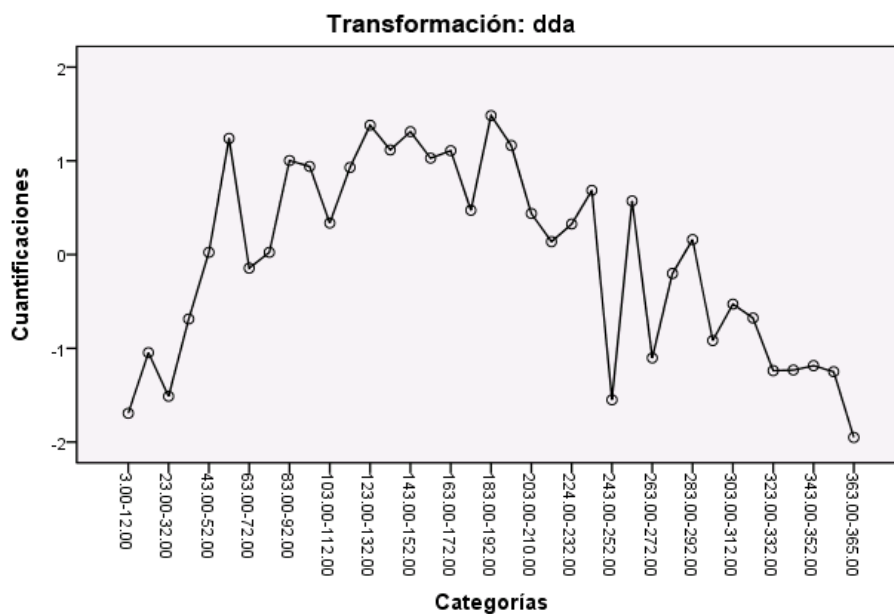
El gráfico de transformación de *Visibilidad*, al igual que el de *Altura de base de inversión*, no parece mostrar ningún patrón. La imposición de restricciones ordinales o lineales sobre las cuantificaciones de esta variable puede reducir considerablemente el ajuste.

Figura 9-35
Gráfico de transformación de temperatura (nominal)



El gráfico de transformación de *Temperatura* muestra otro patrón. A medida que aumentan las categorías, las cuantificaciones tienden a aumentar. Como resultado, a medida que aumenta *Temperatura*, el ozono pronosticado tiende a aumentar. Este patrón sugiere escalar *Temperatura* a nivel ordinal.

Figura 9-36
Gráfico de transformación de día del año (nominal)

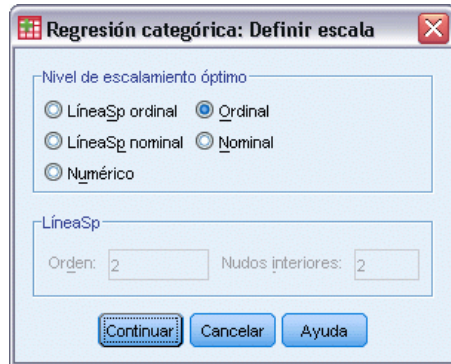


Esta figura muestra el gráfico de transformación de *Día del año*. Las cuantificaciones tienden a aumentar hasta el punto medio del gráfico, en el que tienden a reducirse, produciendo una forma de U. Considerando el signo del coeficiente de regresión para *Día del año*, las categorías iniciales reciben cuantificaciones que tienen un efecto decreciente sobre el ozono pronosticado. En las categorías intermedias, el efecto de las cuantificaciones sobre el ozono pronosticado aumenta, alcanzando su punto máximo cerca del punto medio del gráfico.

A partir de ese punto, las cuantificaciones tienden a disminuir el ozono pronosticado. Aunque la línea es bastante irregular, sigue siendo posible identificar la forma general. Por tanto, los gráficos de transformación sugieren escalar *Temperatura* a nivel ordinal mientras que se siguen escalando todos los demás predictores de manera nominal.

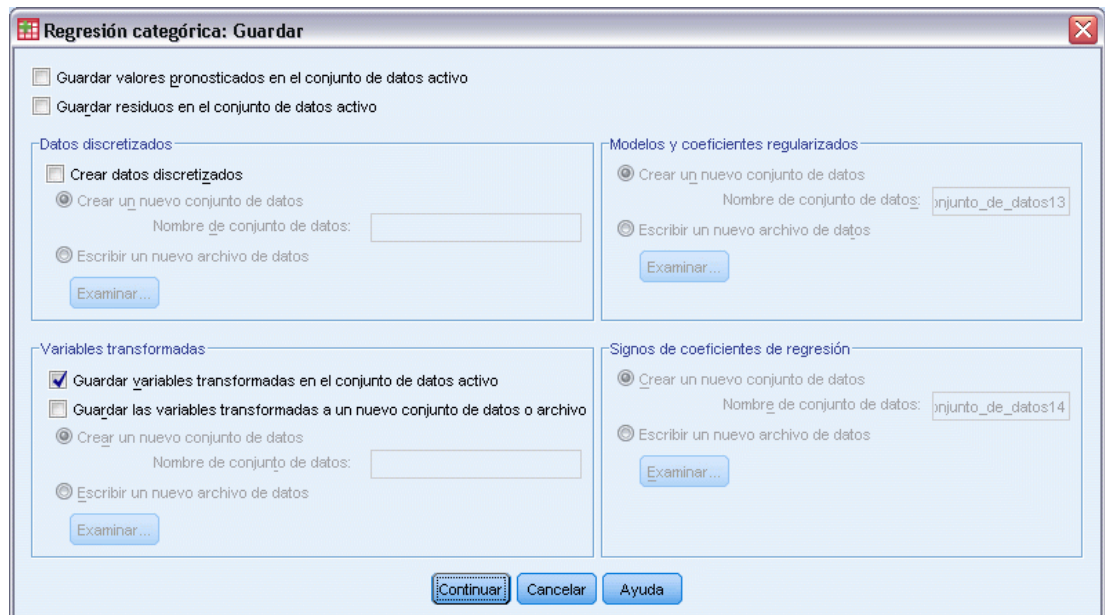
Para volver a calcular la regresión y escalar *Temperatura* a nivel ordinal, vuelva al cuadro de diálogo Regresión categórica.

Figura 9-37
Cuadro de diálogo Definir escala



- ▶ Seleccione *Temperatura* y pulse en Definir escala.
- ▶ Seleccione Ordinal como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Pulse en Guardar en el cuadro de diálogo Regresión categórica.

Figura 9-38
Cuadro de diálogo Guardar



- ▶ Seleccione Guardar variables transformadas en el conjunto de datos activo en el grupo Variables transformadas.
- ▶ Pulse en Continuar.

- Pulse en Aceptar en el cuadro de diálogo Regresión categórica.

Figura 9-39

Resumen del modelo de regresión con *Temperatura* (ordinal)

Resumen del modelo				
	R múltiple	R cuadrado	R cuadrado corregida	Error de predicción aparente
Datos estandarizados	.934	.872	.787	.128

Variable dependiente: ozono
Predictores: abi gdp vis temp dda

Este modelo produce una R^2 de 0,872, por lo que la varianza explicada disminuye de manera insignificante cuando se impone a las cuantificaciones de *Temperatura* la restricción de estar ordenadas.

Figura 9-40

Coefficientes de regresión con *Temperatura* (ordinal)

	Coefficientes tipificados		gl	F	Sig.
	Beta	Bootstrap (1000) Estimación de error típico			
abi	.298	.042	42	51.376	.000
gdp	.301	.047	16	41.882	.000
vis	.224	.044	17	25.659	.000
temp	.609	.084	21	52.113	.000
dda	.373	.053	36	50.330	.000

Variable dependiente: ozono

Esta tabla muestra los coeficientes del modelo en el que *Temperatura* se escala como ordinal. Al comparar los coeficientes con los del modelo en el que *Temperatura* se escalaba como nominal, no se ve ningún cambio de importancia.

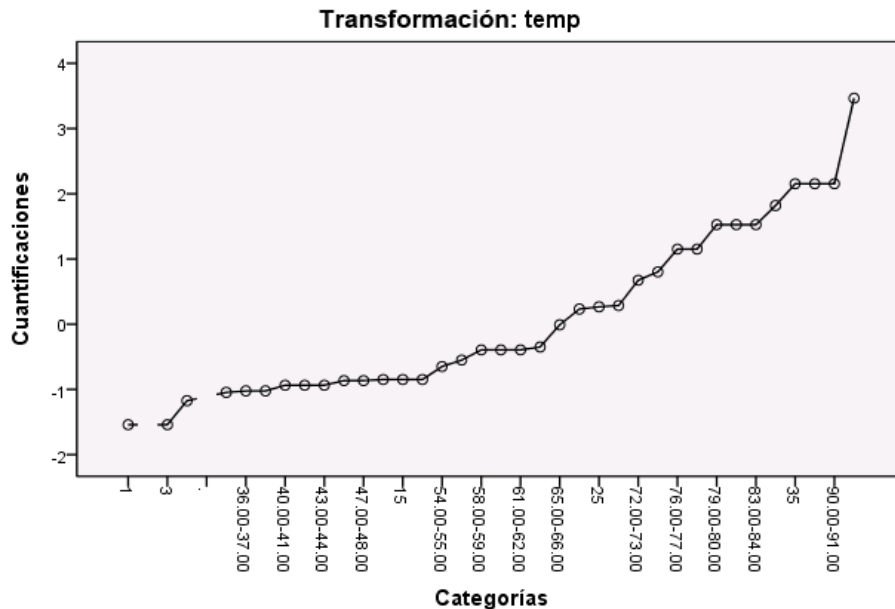
Figura 9-41
Correlaciones, importancia y tolerancia

	Correlaciones y tolerancia				Tolerancia	
	Correlaciones			Importancia	Después de la transformación	Antes de la transformación
	Orden cero	Parcial	Semiparcial		n	n
abi	.438	.627	.288	.150	.930	.596
gdp	.128	.606	.272	.044	.815	.858
vis	.365	.518	.216	.094	.933	.752
temp	.804	.843	.559	.562	.842	.580
dda	.352	.677	.329	.151	.777	.802

Variable dependiente: ozono

Además, las medidas de la importancia sugieren que *Temperatura* sigue siendo mucho más importante en la regresión que las demás variables. Ahora, no obstante, como resultado del nivel de escalamiento ordinal de *Temperatura* y el coeficiente de regresión positivo, puede afirmar que a medida que aumenta *Temperatura*, aumenta el ozono pronosticado.

Figura 9-42
Gráfico de transformación de temperatura (ordinal)



El gráfico de transformación ilustra la restricción ordinal sobre las cuantificaciones de *Temperatura*. La línea irregular de la transformación nominal se ve aquí sustituida por una suave línea ascendente. Además, ya no aparece ninguna larga meseta, lo que indica que no es necesario agrupar categorías.

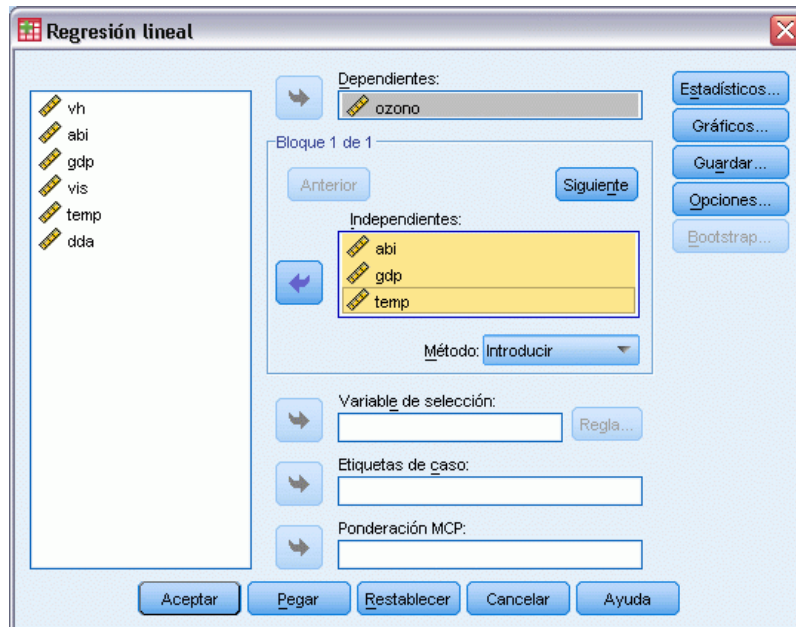
Optimalidad de las cuantificaciones

Las variables transformadas de una regresión categórica se pueden utilizar en la regresión lineal típica, lo que produciría resultados idénticos. Sin embargo, las cuantificaciones son óptimas sólo para el modelo que las generó. El uso de un subconjunto de predictores en la regresión lineal no corresponde a una regresión con escalamiento óptimo sobre el mismo subconjunto.

Por ejemplo, la regresión categórica que se ha calculado tiene una R^2 de 0,875. Ha guardado las variables transformadas, por lo que para ajustar una regresión lineal utilizando únicamente *Temperatura*, *Gradiente de presión* y *Altura de base de inversión* como predictores, elija en los menús:

Analizar > Regresión > Lineales...

Figura 9-43
Cuadro de diálogo Regresión lineal



- ▶ Seleccione *Nivel diario de ozono Cuantificación* como la variable dependiente.
- ▶ Seleccione *Altura de base de inversión Cuantificación*, *Gradiente de presión (mm Hg) Cuantificación* y *Temperatura (grados F) Cuantificación* como variables independientes.
- ▶ Pulse en Aceptar.

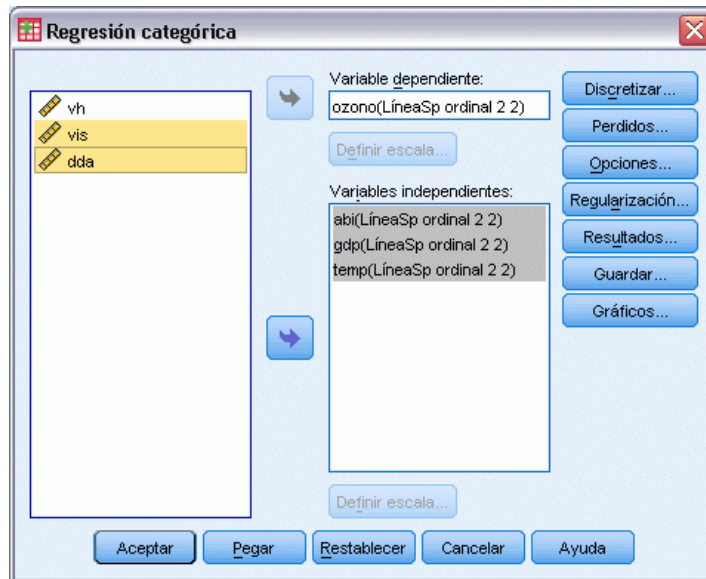
Figura 9-44
Resumen del modelo de regresión con un subconjunto de predictores escalados de manera óptima

Resumen del modelo				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	.856 ^a	.732	.729	4.16711

a. Variables predictoras: (Constante), Temp Cuantificación, Gdp Cuantificación, Abi Cuantificación

Utilizando las cuantificaciones para la respuesta, *Temperatura*, *Gradiente de presión* y *Altura de base de inversión* en una regresión lineal típica se obtiene un ajuste del 0,732. Para comparar esto con el ajuste de una regresión categórica que utilice únicamente estos tres predictores, vuelva al cuadro de diálogo Regresión categórica.

Figura 9-45
Cuadro de diálogo Regresión categórica



- ▶ Anule la selección de *Visibilidad (millas)* y *Día del año* como variables independientes.
- ▶ Pulse en Aceptar.

Figura 9-46
Resumen del modelo de regresión categórica con tres predictores

	R múltiple	R cuadrado	R cuadrado corregida	Error de predicción aparente
Datos estandarizados	.892	.796	.735	.204

Variable dependiente: ozono
Predictores: abi gdp temp

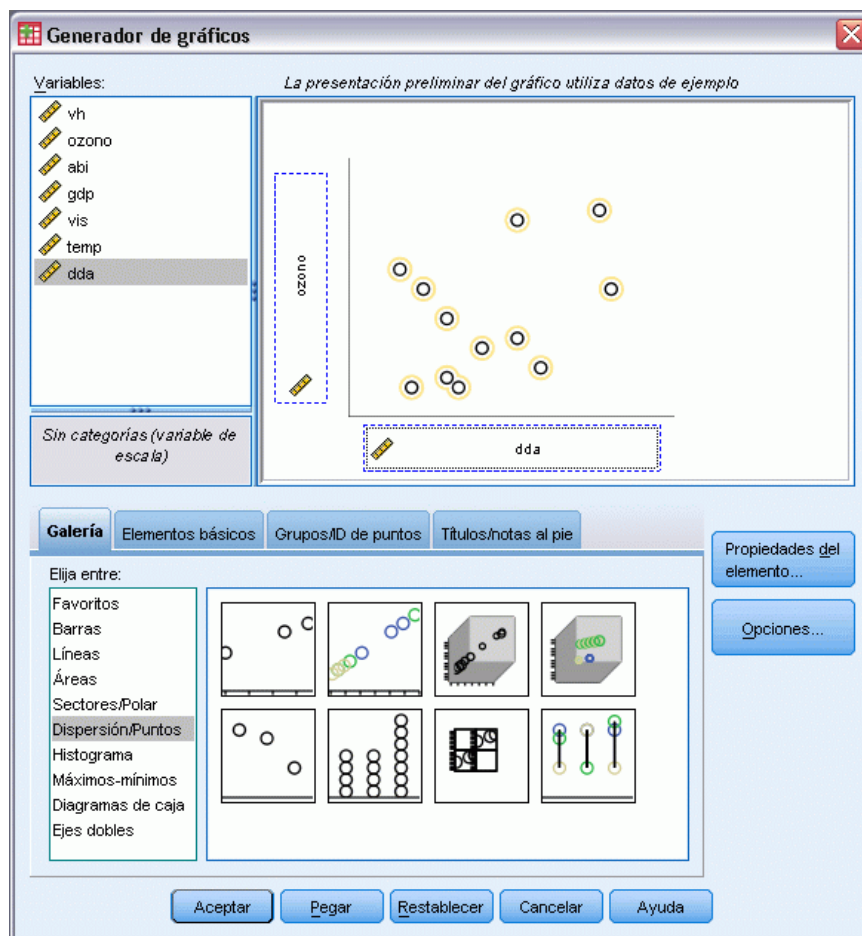
El análisis de regresión categórica tiene un ajuste de 0,796, que es mejor que el ajuste de 0,732. Se demuestra así la propiedad de los escalamientos de que las cuantificaciones obtenidas en la regresión original son únicamente óptimas cuando se incluyen las cinco variables en el modelo.

Efectos de las transformaciones

La transformación de las variables convierte una relación que no es lineal entre la respuesta original y el conjunto original de predictores, en lineal para las variables transformadas. Sin embargo, cuando hay varios predictores, las relaciones por parejas aparecen confundidas por las demás variables del modelo.

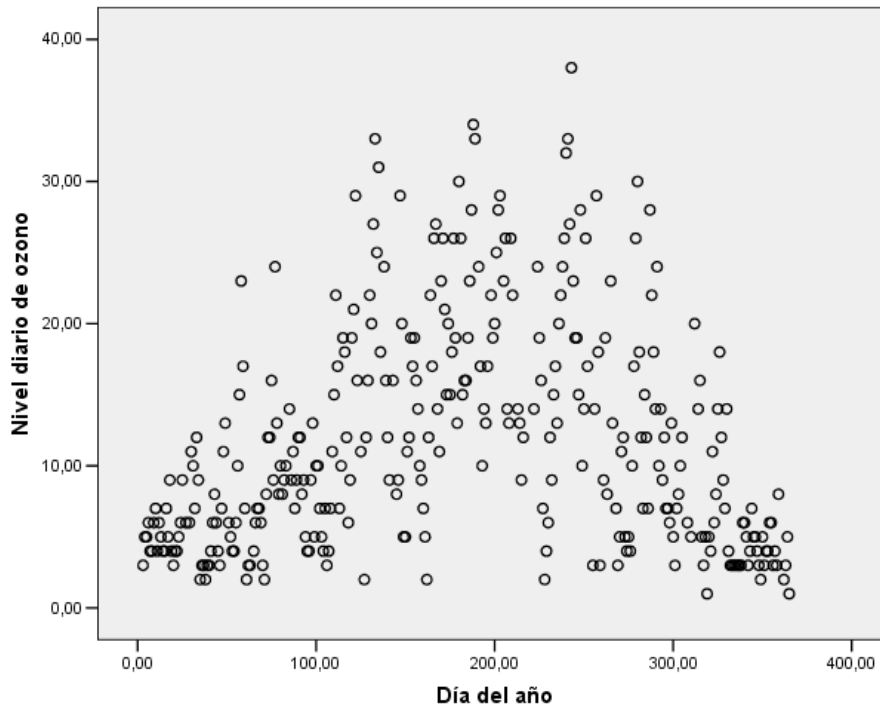
Para centrar el análisis en la relación entre *Nivel diario de ozono* y *Día del año*, comenzaremos observando el diagrama de dispersión. Seleccione en los menús:
Gráficos > Generador de gráficos...

Figura 9-47
Cuadro de diálogo Generador de gráficos



- ▶ Seleccione la galería Dispersión/Puntos y elija Dispersión simple.
- ▶ Seleccione *Nivel diario de ozono* como variable y y *Día del año* como variable x .
- ▶ Pulse en Aceptar.

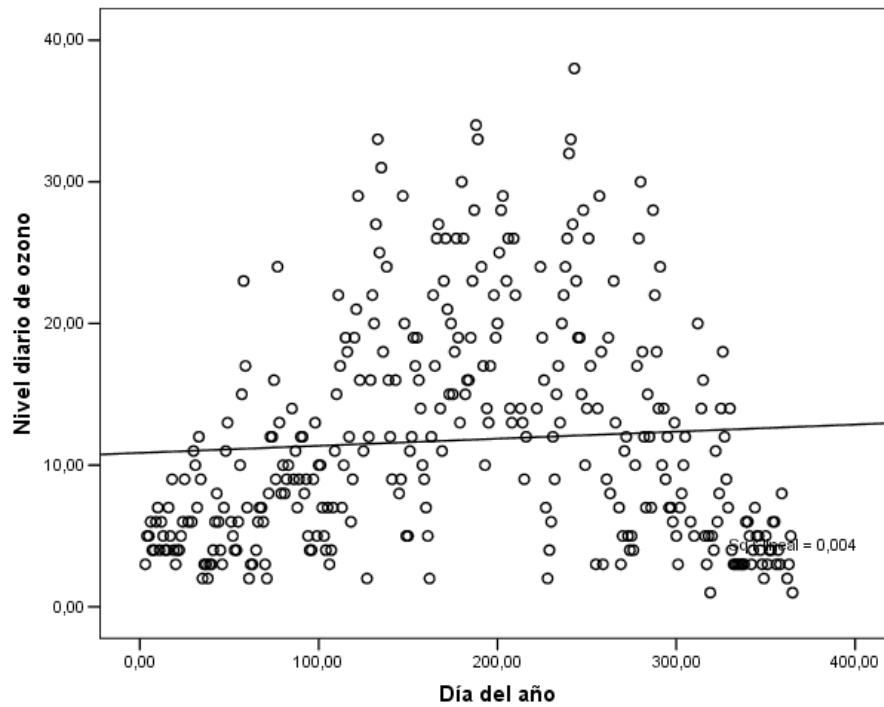
Figura 9-48
Diagrama de dispersión del nivel de ozono diario por día del año



Esta figura ilustra la relación entre *Nivel diario de ozono* y *Día del año*. A medida que *Día del año* aumenta hasta aproximadamente 200, también aumenta *Nivel diario de ozono*. No obstante, para los valores *Día del año* valores mayores que 200, el valor de *Nivel diario de ozono* disminuye. Este patrón de U invertida sugiere una relación cuadrática entre las dos variables. Una regresión lineal no puede capturar esta relación.

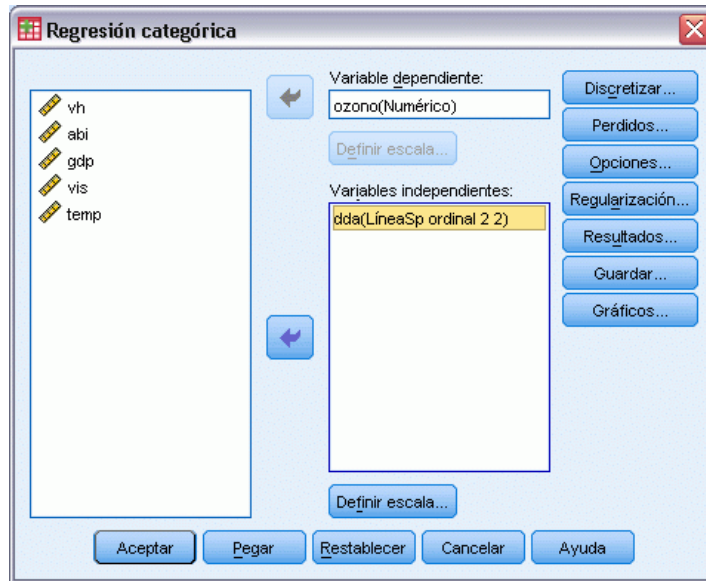
- ▶ Para ver una línea de mejor ajuste superpuesta sobre los puntos del diagrama de dispersión, active el gráfico pulsando dos veces en él.
- ▶ Seleccione un punto en el Editor de gráficos.
- ▶ Pulse en la herramienta Añadir línea de ajuste total y, a continuación, cierre el Editor de gráficos.

Figura 9-49
Diagrama de dispersión con una línea de ajuste



Una regresión lineal de *Nivel diario de ozono* sobre *Día del año* produce una R^2 de 0,004. Este ajuste sugiere que *Día del año* no tiene ningún valor predictivo para *Nivel diario de ozono*. Esto no resulta sorprendente, teniendo en cuenta el patrón de la figura. Utilizando el escalamiento óptimo, sin embargo, puede hacer lineal la relación cuadrática y utilizar el *Día del año* para pronosticar la respuesta.

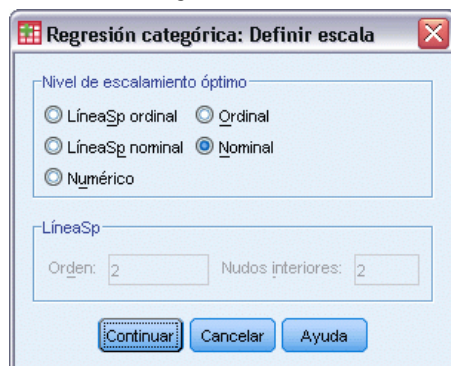
Figura 9-50
Cuadro de diálogo Regresión categórica



Para obtener una regresión categórica de *Nivel diario de ozono* sobre *Día del año*, vuelva al cuadro de diálogo Regresión categórica.

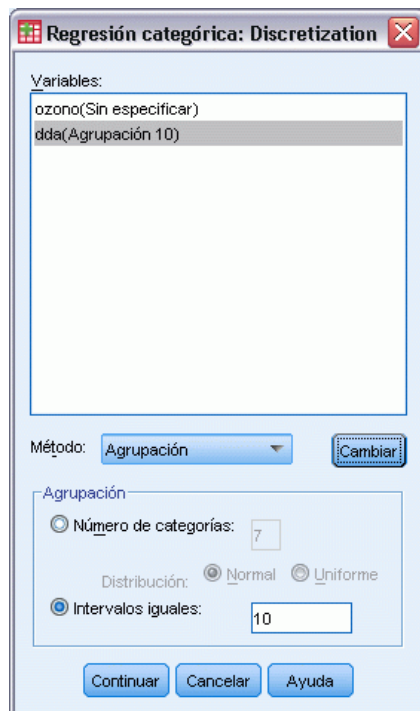
- ▶ Anule la selección desde *Altura de base de inversión* hasta *Temperatura (grados F)* como variables independientes.
- ▶ Seleccione *Día del año* como una variable independiente.
- ▶ Pulse en Definir escala.

Figura 9-51
Cuadro de diálogo Definir escala



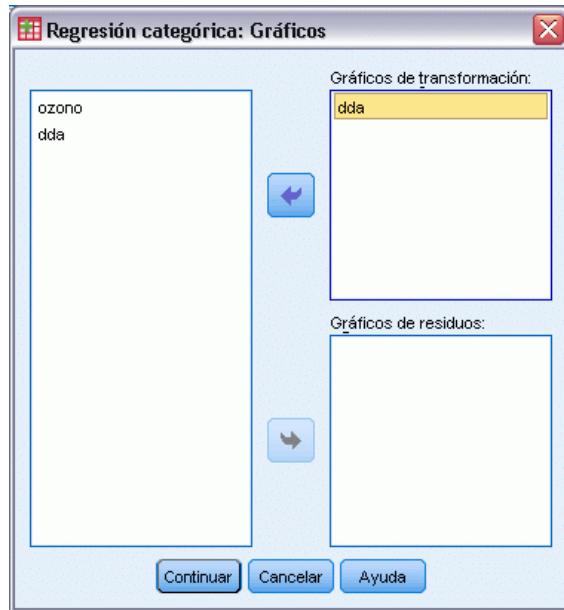
- ▶ Seleccione Nominal como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Pulse en Discretizar en el cuadro de diálogo Regresión categórica.

Figura 9-52
Cuadro de diálogo Discretización



- ▶ Seleccione *dda*.
- ▶ Seleccione Intervalos iguales.
- ▶ Escriba 10 como longitud del intervalo.
- ▶ Pulse en Cambiar.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Regresión categórica.

Figura 9-53
Cuadro de diálogo Gráficos



- ▶ Seleccione *dda* para los gráficos de transformación.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Regresión categórica.

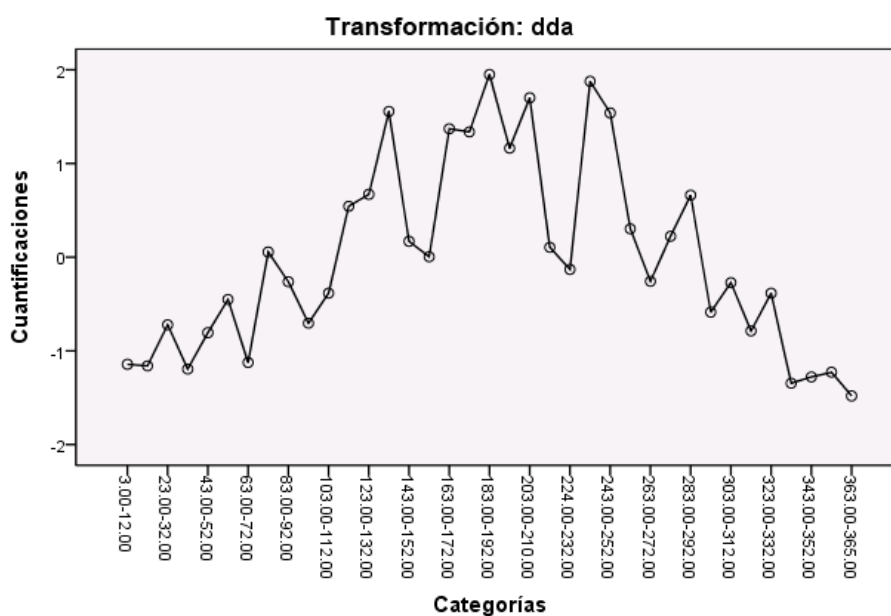
Figura 9-54
Resumen del modelo de regresión categórica del Nivel diario de ozono con respecto a Día del año

	R múltiple	R cuadrado	R cuadrado corregida	Error de predicción aparente
Datos estandarizados	.741	.549	.494	.451

Variable dependiente: ozono
Predictor: dda

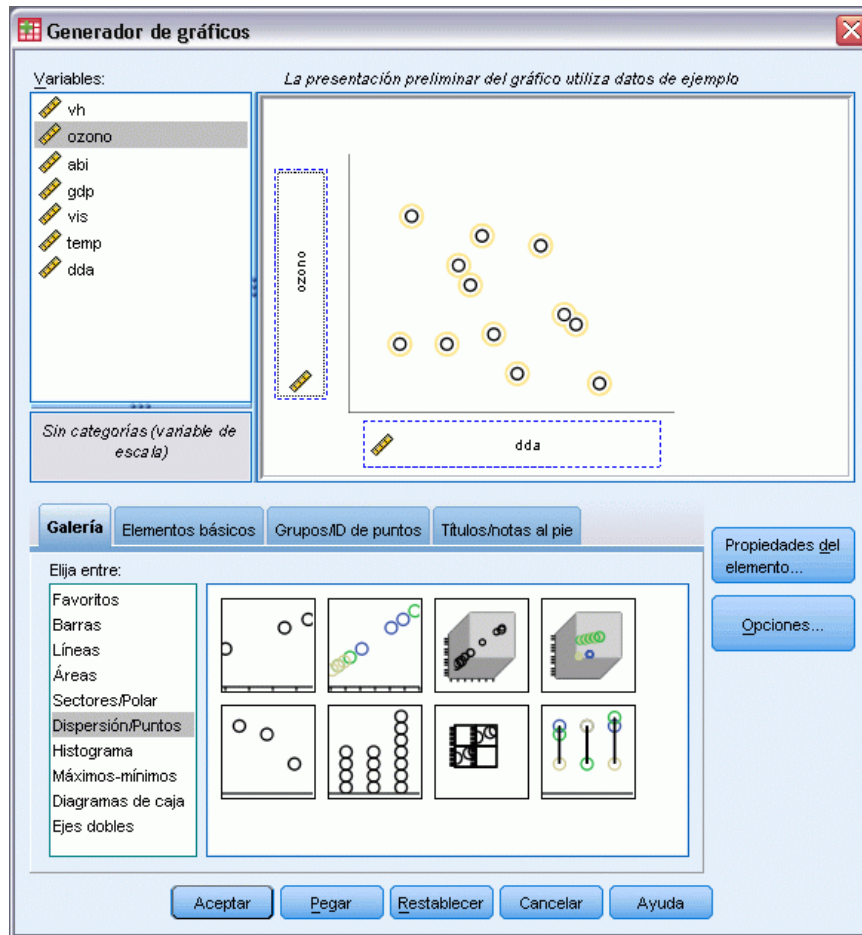
La regresión de escalamiento óptimo trata *Nivel diario de ozono* como numérico y *Día del año* como nominal. Se produce como resultado una R^2 de 0,549. Aunque sólo el 55% de la variación en *Nivel diario de ozono* queda explicada por la regresión categórica, se trata de una mejora substancial respecto a la regresión original. La transformación *Día del año* permite pronosticar *Nivel diario de ozono*.

Figura 9-55
Gráfico de transformación de día del año (nominal)



Esta figura muestra el gráfico de transformación de *Día del año*. Ambos extremos de *Día del año* reciben cuantificaciones negativas, mientras que los valores centrales tienen cuantificaciones positivas. Mediante la aplicación de esta transformación, los valores inferiores y superiores de *Día del año* tienen efectos similares sobre los valores pronosticados de *Nivel diario de ozono*.

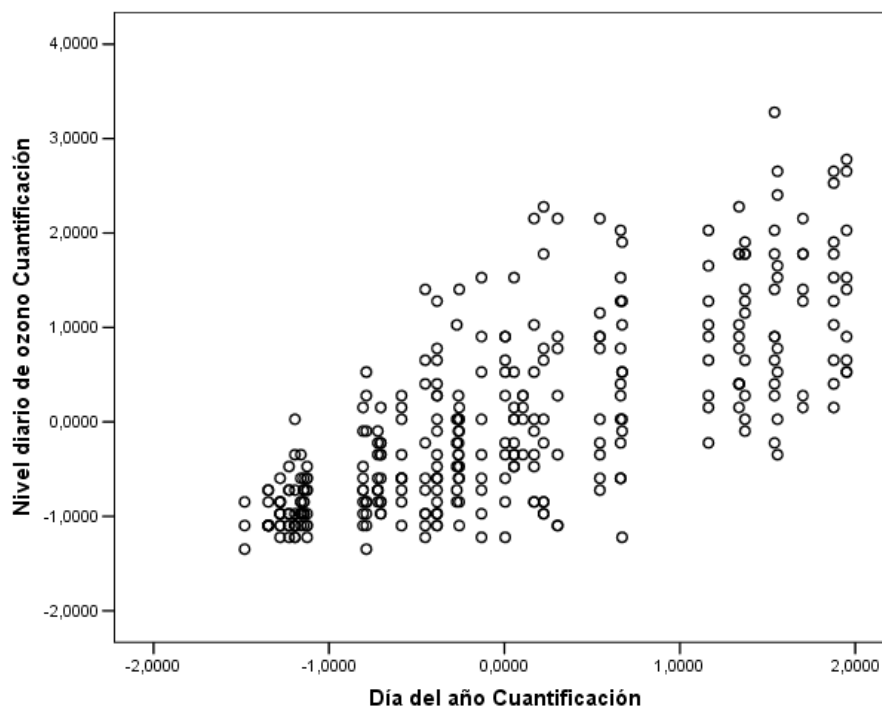
Figura 9-56
Generador de gráficos



Para ver un diagrama de dispersión de las variables transformadas, vuelva a Generador de cuadros y pulse en Restablecer para borrar las selecciones anteriores.

- ▶ Seleccione la galería Dispersión/Puntos y elija Dispersión simple.
- ▶ Seleccione *Nivel diario de ozono Cuantificación [TRA1_3]* como variable y *Día del año Cuantificación [TRA2_3]* como variable x .
- ▶ Pulse en Aceptar.

Figura 9-57
Diagrama de dispersión de las variables transformadas



Esta figura describe la relación entre las variables transformadas. Una tendencia creciente sustituye a la U invertida. La línea de regresión tiene una pendiente positiva, lo que indica que a medida que el valor transformado de *Día del año* aumenta, también lo hace *Nivel diario de ozono*. El uso del escalamiento óptimo convierte en lineal la relación y permite realizar interpretaciones que de otra manera pasarían desapercibidas.

Lecturas recomendadas

Consulte los siguientes textos si desea obtener más información sobre la regresión categórica:

Buja, A. 1990. Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics*, 18, .

Hastie, T., R. Tibshirani, y A. Buja. 1994. Flexible discriminant analysis. *Journal of the American Statistical Association*, 89, .

Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 2, .

Kruskal, J. B. 1965. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society Series B*, 27, .

Meulman, J. J. 2003. Prediction and classification in nonlinear data analysis: Something old, something new, something borrowed, something blue. *Psychometrika*, 4, .

Ramsay, J. O. 1989. Monotone regression splines in action. *Statistical Science*, 4, .

Van der Kooij, A. J., y J. J. Meulman. 1997. MURALS: Multiple regression and optimal scaling using alternating least squares. En: *Softstat '97*, F. Faulbaum, y W. Bandilla, eds. Stuttgart: Gustav Fisher.

Winsberg, S., y J. O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika*, 67, .

Winsberg, S., y J. O. Ramsay. 1983. Monotone spline transformations for dimension reduction. *Psychometrika*, 48, .

Young, F. W., J. De Leeuw, y Y. Takane. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .

Análisis de componentes principales categórico

El análisis de componentes principales categórico se puede considerar como un método de reducción de dimensiones. Se analiza un conjunto de variables para descubrir las dimensiones más importantes de la variación. El conjunto de datos original se puede sustituir por un nuevo conjunto de datos más pequeño con una pérdida mínima de información. El método descubre las relaciones entre las variables, entre los casos y entre las variables y los casos.

El criterio utilizado por el análisis de componentes principales categórico para cuantificar los datos observados es que las puntuaciones de los objetos (las puntuaciones de los componentes) deben tener grandes correlaciones con cada una de las variables cuantificadas. Una solución es buena en tanto que satisface este criterio.

Se presentarán dos ejemplos del análisis de componentes principales categórico. El primero utiliza un conjunto de datos bastante pequeño para ilustrar los conceptos básicos y las interpretaciones asociadas con el procedimiento. El segundo ejemplo examina una aplicación práctica.

Ejemplo: Examen de las interrelaciones de los sistemas sociales

Este ejemplo examina la adaptación de Guttman (Guttman, 1968) de una tabla de Bell (Bell, 1961). Los datos también están comentados por Lingoes (Lingoes, 1968).

Bell presentó una tabla para ilustrar posibles grupos sociales. Guttman utilizó parte de esta tabla, en la que se cruzaron cinco variables que describían elementos como la interacción social, los sentimientos de pertenencia a un grupo, proximidad física de los miembros y grado de formalización de la relación con siete grupos sociales teóricos, incluidos multitudes (por ejemplo, las personas que acuden a un partido de fútbol), espectadores (por ejemplo, las personas que acuden a un teatro o de una conferencia), públicos (por ejemplo, los lectores de periódicos o los espectadores de televisión), muchedumbres (como una multitud pero con una interacción mucho más intensa), grupos primarios (íntimos), grupos secundarios (voluntarios) y la comunidad moderna (confederación débil que resulta de la proximidad cercana física y de la necesidad de servicios especializados).

La siguiente tabla muestra las variables del conjunto de datos que resulta de la clasificación en siete grupos sociales utilizada en los datos de Guttman-Bell, con sus etiquetas de variable y las etiquetas de valor (categorías) asociadas con los niveles de cada variable. Este conjunto de datos se puede encontrar en *guttman.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20](#). Además de seleccionar las variables que se incluirán en el cálculo del análisis de componentes principales categórico, puede seleccionar las variables que se utilizarán para etiquetar a los objetos en los gráficos. En este ejemplo, las primeras cinco variables de los datos se incluyen en el análisis, mientras que el conglomerado se utiliza de manera exclusiva como variable de etiquetado. Cuando especifique un análisis de componentes

principales categórico, debe especificar el nivel de escalamiento óptimo para cada variable de análisis. En este ejemplo, se especifica un nivel ordinal para todas las variables del análisis.

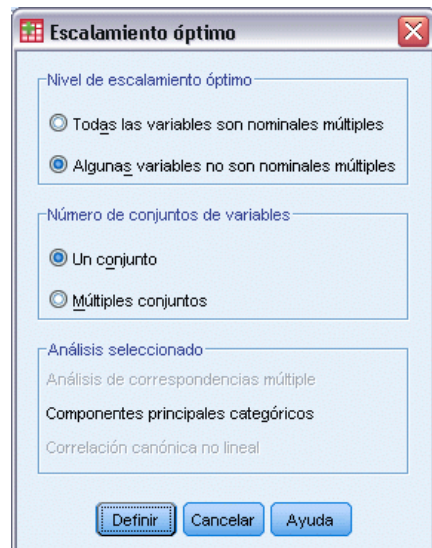
Tabla 10-1
Variables del conjunto de datos Guttman-Bell

Nombre de variable	Etiqueta de variable	Etiqueta de valor
<i>intesidd</i>	Intensidad	Ligera, baja, moderada, alta
<i>frecuenc</i>	Frecuencia	Ligera, no recurrente, no frecuente, frecuente
<i>pertenci</i>	Pertenencia	Ninguno, ligero, variable, alto
<i>proxmdad</i>	Proximidad	Distante, cercana
<i>formal</i>	Formalismo	Sin relación, formal, informal
<i>conglomerado</i>		Multitudes, espectadores, públicos, muchedumbres, grupos primarios, grupos secundarios, comunidad moderna

Ejecución del análisis

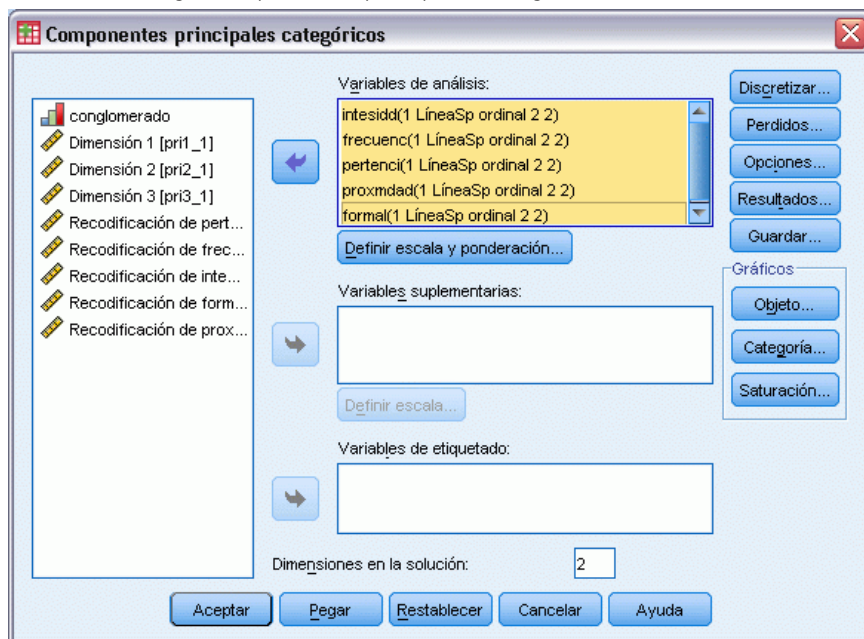
- Para generar los resultados de componentes principales categóricos de este conjunto de datos, seleccione en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 10-1
Cuadro de diálogo Escalamiento óptimo



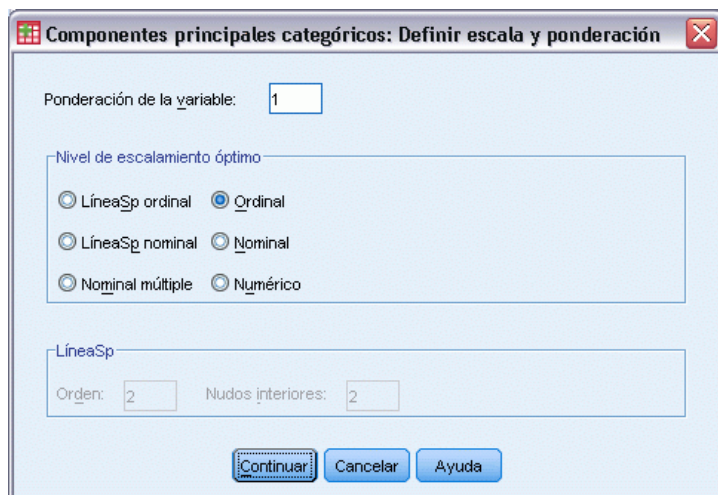
- Seleccione Alguna variable no es nominal múltiple en el grupo Nivel para escalamiento óptimo.
- Pulse en Definir.

Figura 10-2
Cuadro de diálogo Componentes principales categórico



- ▶ Seleccione de *Intensidad* a *Formalismo* como variables de análisis.
- ▶ Pulse en Definir escala y ponderación.

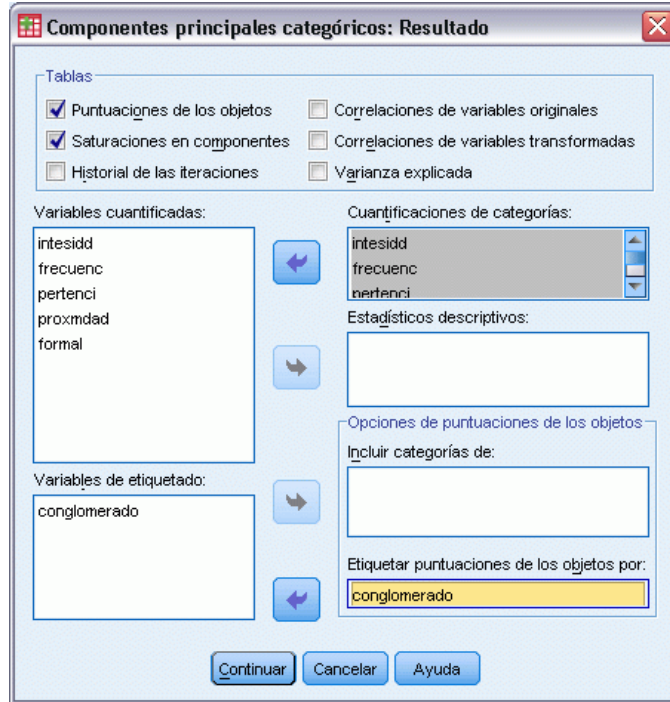
Figura 10-3
Definir escala y ponderación



- ▶ Seleccione Ordinal en el grupo Nivel para escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Seleccione *conglomerado* como variable de etiquetado en el cuadro de diálogo Componentes principales categóricos.

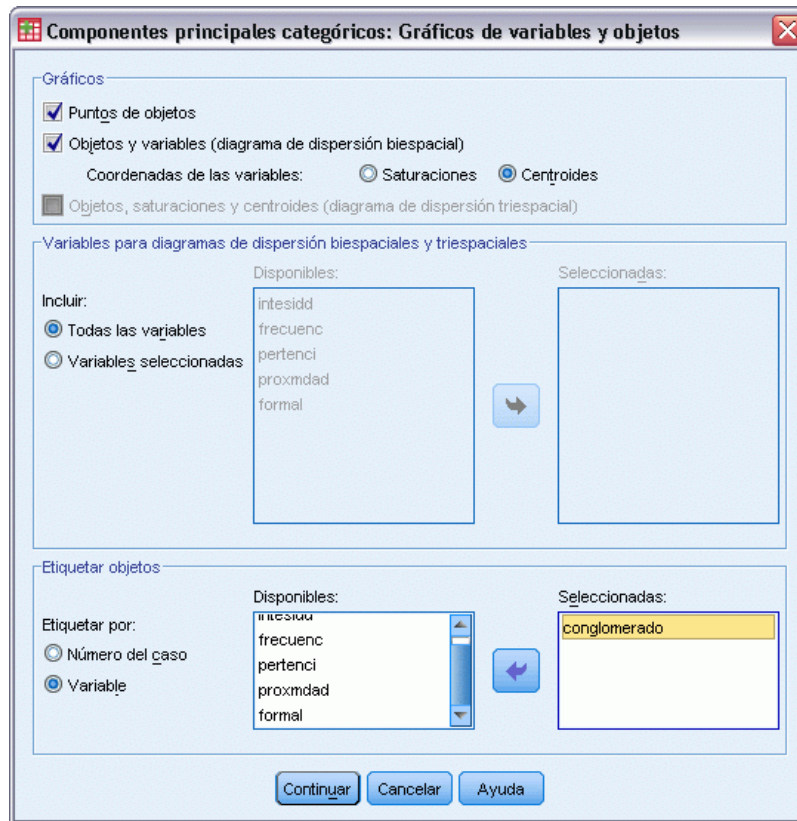
- ▶ Pulse en Resultados.

Figura 10-4
Cuadro de diálogo Resultados



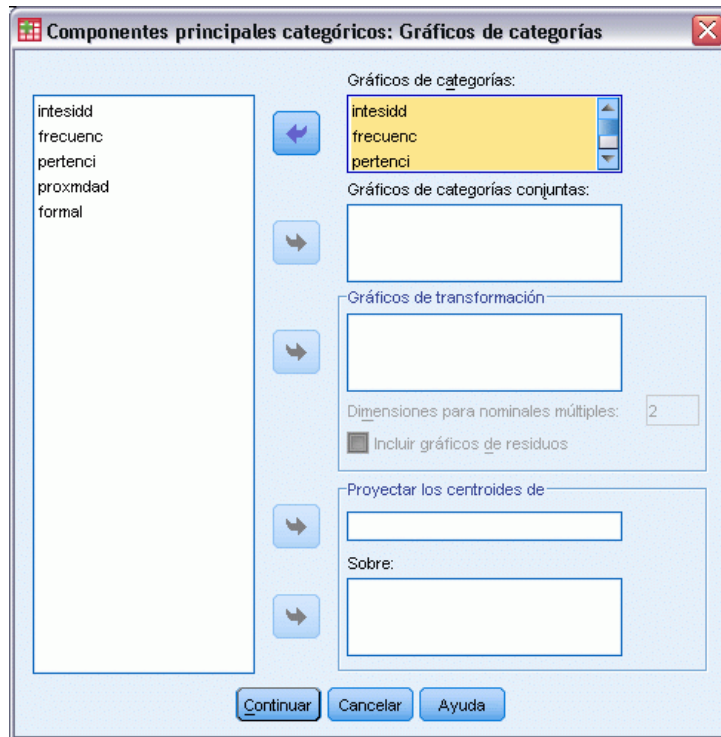
- ▶ Seleccione Puntuaciones de los objetos y desactive Correlaciones de variables transformadas en el grupo Tablas.
- ▶ Seleccione generar cuantificaciones de las categorías desde *intesidd* (*Intensidad de la interacción*) hasta *formal* (*Grado de formalización de la relación*).
- ▶ Seleccione etiquetar las puntuaciones de los objetos por *conglomerado*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Objeto en el grupo Gráficos del cuadro de diálogo Componentes principales categóricos.

Figura 10-5
Gráficos de variables y objetos



- ▶ Seleccione **Objetos y variables** (gráfico biespacial) en el grupo **Gráficos**.
- ▶ Seleccione **etiquetar los objetos por Variable** en el grupo **Etiquetar objetos** y, a continuación, seleccione *conglomerado* como la variable por la que se etiquetarán los objetos.
- ▶ Pulse en **Continuar**.
- ▶ Pulse en **Categoría** en el grupo **Gráficos** del cuadro de diálogo **Componentes principales categóricos**.

Figura 10-6
Cuadro de diálogo Gráficos de categorías



- ▶ Seleccione generar gráficos de categorías conjuntas para desde *intensidd* (*Intensidad de la interacción*) hasta *formal* (*Grado de formalización de la relación*).
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Componentes principales categóricos.

Número de dimensiones

Estas figuras muestran parte de los resultados iniciales del análisis de componentes principales categórico. Tras el historial de iteraciones del algoritmo, aparecerá el resumen del modelo, incluidos los autovalores de cada dimensión. Estos autovalores son equivalentes a los del análisis de componentes principales clásico. Son las medidas de qué parte de la varianza se explica por cada una de las dimensiones.

Figura 10-7
Historial de las iteraciones

Número de iteraciones	Varianza explicada		Pérdida		
	Total	Incremento	Total	Coordenadas de centroide	Restricción del centroide a las coordenadas del vector
0 ^a	4,515315	,000000	5,484685	4,075583	1,409101
31 ^b	4,726009	,000008	5,273991	4,273795	1,000196

a. La iteración 0 muestra los estadísticos de la solución con todas las variables, excepto las variables con un nivel de escalamiento óptimo nominal múltiple, consideradas como numéricas.

Figura 10-8
Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	% de la varianza
1	,881	3,389	67,774
2	,315	1,337	26,746
Total	,986 ^a	4,726	94,520

a. El Alfa de Cronbach Total está basado en los autovalores totales.

Los autovalores se pueden utilizar como indicación del número de dimensiones que son necesarias. En este ejemplo, se utilizó el número por defecto de dimensiones, 2. ¿Es éste el número correcto? Como norma general, cuando todas las variables son nominales, ordinales o numéricas únicas, el autovalor de una dimensión debe ser mayor que 1. Ya que la solución de dos dimensiones resuelve el 94,52% de la varianza, una tercera dimensión probablemente no añadiría mucha más información.

Para variables nominales múltiples, no hay ninguna norma general sencilla apropiada para determinar el número adecuado de dimensiones. Si el número de variables se sustituye por el número total de categorías menos el número de variables, la regla anterior seguirá siendo válida. Pero esta regla independiente, probablemente permita utilizar más dimensiones de las necesarias. Al elegir el número de dimensiones, la directriz más útil es que el número sea suficientemente pequeño para que sean posibles interpretaciones que tengan sentido. La tabla del resumen del modelo también muestra el alfa de Cronbach (una medida de la fiabilidad), la cual maximiza el procedimiento.

Cuantificaciones

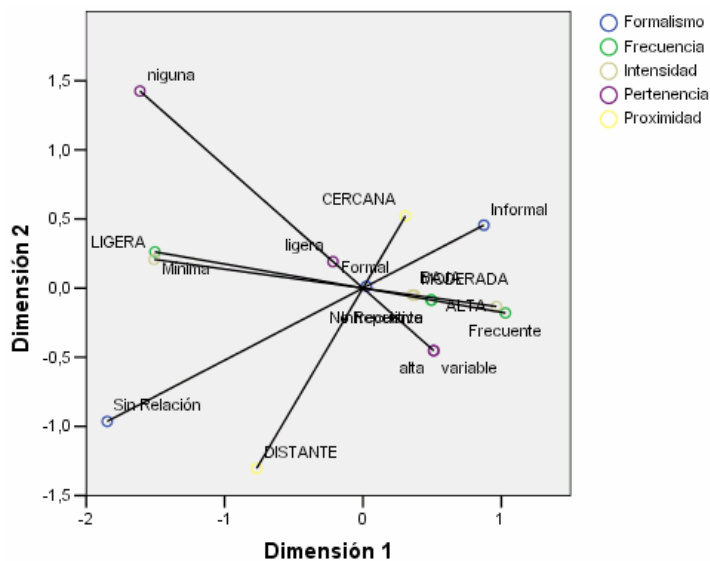
Para cada variable se muestran las cuantificaciones, las coordenadas del vector y las coordenadas del centroide para cada una de las dimensiones. Las cuantificaciones son los valores asignados a cada categoría. Las coordenadas del centroide son la media de las puntuaciones de los objetos de la misma categoría. Las coordenadas del vector son las coordenadas de las categorías cuando es necesario que estén en una línea, representando a la variable en el espacio de los objetos. Esto es necesario para las variables con el nivel de escalamiento ordinal y numérico.

Figura 10-9
Cuantificaciones de la intensidad de la interacción

Categoría	Frecuencia	Cuantificación	Coordenadas de centroide		Coordenadas del vector	
			Dimensión		Dimensión	
			1	2	1	2
LIGERA	2	-1,530	-1,496	,308	-1,510	,208
BAJA	2	,362	,392	,202	,358	-,049
MODERADA	1	,379	,188	-1,408	,374	-,051
ALTA	2	,978	1,010	,194	,965	-,133

Mirando las cuantificaciones en el gráfico conjunto de los puntos de categorías, podrá ver que algunas de las categorías de algunas variables no aparecen tan claramente diferenciadas por el análisis de componentes principales categórico como sería de esperar si el nivel hubiese sido realmente ordinal. Las variables *Intensidad* y *Frecuencia*, por ejemplo, tienen cuantificaciones iguales o muy similares para las de las dos categorías intermedias. Este tipo de resultado puede sugerir intentar otros análisis de componentes principales categóricos, tal vez con algunas categorías unidas o tal vez con un diferente nivel de análisis, como nominal (múltiple).

Figura 10-10
Puntos de categorías del gráfico conjunto



El gráfico conjunto de los puntos de categorías se parece al gráfico de las saturaciones en componentes, pero también muestra dónde se encuentran los puntos finales que corresponden a las cuantificaciones menores (por ejemplo, *ligera* para *Intensidad* y *ninguna* para *Sentimiento*). Las dos variables que miden la interacción, *Intensidad* y *Frecuencia*, aparecen muy juntas y explican la parte de la varianza de la dimensión 1. *Formalismo* también aparece cerca de *Proximidad*.

Si nos centramos en los puntos de categoría, será posible ver las relaciones todavía con mayor claridad. No sólo están cerca *Intensidad* y *Frecuencia*, sino que las direcciones de sus escalas son similares; es decir, la intensidad ligera está cerca de la frecuencia ligera y la interacción frecuente está cerca de la intensidad alta de interacción. También se puede ver que la proximidad

cercana física parece ir unida a un tipo informal de relación, mientras que la distancia física está relacionada con que no haya ninguna relación.

Puntuaciones de los objetos

También puede solicitar un listado y un gráfico de las puntuaciones de los objetos. El gráfico de las puntuaciones de los objetos puede ser útil para detectar valores atípicos, detectar grupos habituales de objetos y descubrir algunos patrones especiales.

La tabla de puntuaciones de los objetos muestra el listado de las puntuaciones de los objetos etiquetados por grupo social para los datos de Guttman-Bell. Si se examinan los valores de los puntos de los objetos, puede identificar objetos concretos en el gráfico.

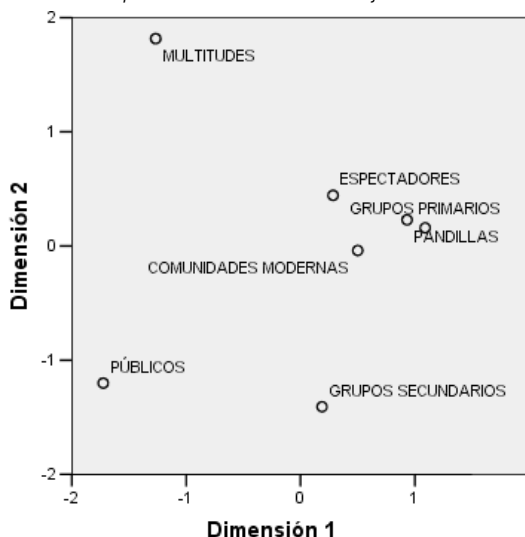
Figura 10-11
Puntuaciones de los objetos

conglomerado	Dimensión	
	1	2
MULTITUDES	-1,266	1,816
ESPECTADORES	,284	,444
PÚBLICOS	-1,726	-1,201
PANDILLAS	,931	,229
GRUPOS PRIMARIOS	1,089	,159
GRUPOS SECUNDARIOS	,188	-1,408
COMUNIDADES MODERNAS	,500	-,039

Normalización principal por variable.

La primera dimensión parece separar *MULTITUDES* y *PÚBLICOS*, que tienen puntuaciones negativas relativamente grandes, de *MUCHEDUMBRES* y *GRUPOS PRIMARIOS*, que tienen puntuaciones positivas relativamente grandes. La segunda dimensión presenta tres grupos: *PÚBLICOS* y *GRUPOS SECUNDARIOS* con valores negativos grandes, *MUCHEDUMBRES* con valores positivos grandes y el resto de los otros grupos sociales entre ellos. Resulta más fácil ver esto mediante la observación del gráfico de las puntuaciones de los objetos.

Figura 10-12
Gráfico de puntuaciones de los objetos

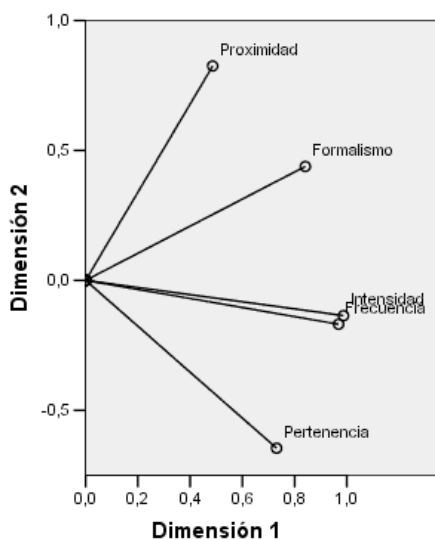


En el gráfico, puede ver *PÚBLICOS* y *GRUPOS SECUNDARIOS* en la parte inferior, *MULTITUDES* en la parte superior y el resto de los grupos sociales en el medio. La observación de los patrones en los objetos individuales depende de la información adicional disponible para las unidades de análisis. En este caso, conoce la clasificación de los objetos. En otros casos, puede utilizar variables suplementarias para etiquetar los objetos. También puede ver que el análisis de componentes principales categórico no separa *MUCHEDUMBRES* de *GRUPOS PRIMARIOS*. Aunque la mayoría de las personas no consideran a sus familias como muchedumbres, en las variables utilizadas, estos dos grupos recibieron la misma puntuación nada menos que en cuatro de las cinco variables. Obviamente, puede que desee estudiar los posibles defectos de las variables y las categorías que se han utilizado. Por ejemplo, la alta intensidad de la interacción y de las relaciones informales probablemente tengan significados diferentes para estos dos grupos. Otra posibilidad consiste en considerar una solución con más dimensiones.

Saturaciones en componentes

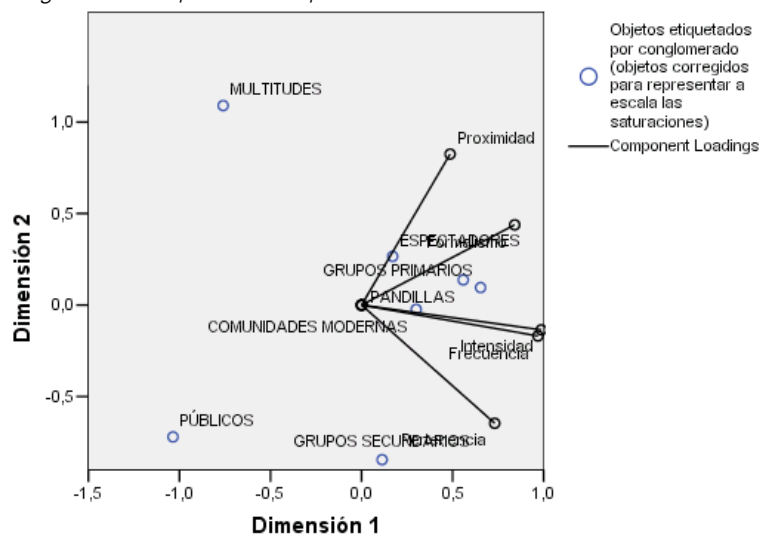
Esta figura muestra el gráfico de las saturaciones en componentes. Los vectores (líneas) son relativamente largos, lo que vuelve a indicar que las primeras dos dimensiones explican la mayoría de la varianza de todas las variables cuantificadas. En la primera dimensión, todas las variables tienen saturaciones en componentes altas (positivas). La segunda dimensión está correlacionada principalmente con las variables cuantificadas *Sentimiento* y *Proximidad*, en direcciones opuestas. Por tanto, los objetos con una puntuación negativa grande en la dimensión 2 tendrán una alta puntuación en el sentimiento de pertenencia y una baja puntuación en la proximidad física. La segunda dimensión, por tanto, muestra un contraste entre estas dos variables al mismo tiempo que tiene escasa relación con las variables cuantificadas *Intensidad* y *Frecuencia*.

Figura 10-13
Saturaciones en componentes



Para examinar la relación entre los objetos y las variables, observe el diagrama de dispersión biespacial de los objetos y las saturaciones en componentes. El vector de una variable apunta en la dirección de la categoría mayor de la variable. Por ejemplo, para *Proximidad* y *Pertenencia* las categorías mayores son *cercano* y *alto*, respectivamente. Por tanto, *MUCHEDUMBRES* se caracterizan por la proximidad física cercana y por la ausencia de sentimiento de pertenencia, mientras que *GRUPOS SECUNDARIOS*, por proximidad física distante y alto sentimiento de pertenencia.

Figura 10-14
Diagrama de dispersión biespacial



Dimensiones adicionales

El aumento del número de dimensiones aumentará la cantidad de variación explicada y puede servir para detectar diferencias ocultas en las soluciones con un número menor de dimensiones. Como se observó previamente, no se puede separar con dos dimensiones *MUCHEDUMBRES* y *GRUPOS PRIMARIOS*. Sin embargo, el aumento de la dimensionalidad tal vez permita diferenciar estos dos grupos.

Ejecución del análisis

- ▶ Para obtener una solución tridimensional, vuelva al cuadro de diálogo Componentes principales categóricos.
- ▶ Escriba 3 como número de dimensiones de la solución.
- ▶ Pulse en Aceptar en el cuadro de diálogo Componentes principales categóricos.

Resumen del modelo

Figura 10-15
Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	% de la varianza
1	,885	3,424	68,480
2	-,232	,844	16,871
3	-,459	,732	14,649
Total	1,000 ^a	5,000	99,999

a. El Alfa de Cronbach Total está basado en los autovalores totales.

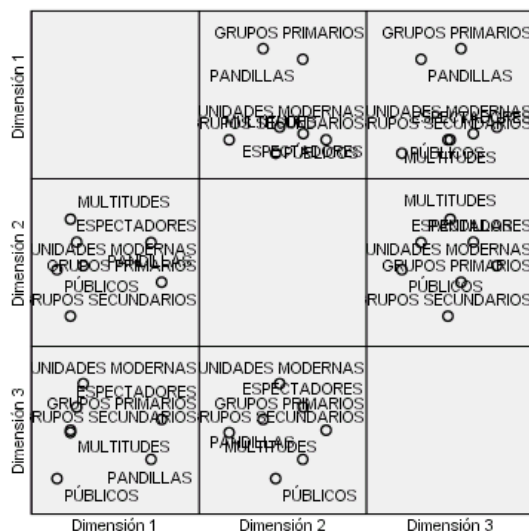
Una solución tridimensional tiene los autovalores 3,424, 0,844 y 0,732, que explican prácticamente toda la varianza.

Puntuaciones de los objetos

Las puntuaciones de la solución tridimensional se representan en un diagrama de dispersión matricial. En un diagrama de dispersión matricial, todas las dimensiones se representan respecto a las demás dimensiones en una serie de diagramas de dispersión bidimensionales. Observe que los dos primeros autovalores en tres dimensiones no son iguales que los autovalores de la solución de dos dimensiones; dicho de otra manera, las soluciones no están anidadas. Como los autovalores de las dimensiones 2 y 3 ahora son menores que 1 (lo que hace que la alfa de Cronbach sea negativa), resulta preferible la solución de dos dimensiones. La solución tridimensional se incluye a modo ilustrativo.

Figura 10-16

Diagrama de dispersión matricial de las puntuaciones de los objetos de tres dimensiones



La fila superior de gráficos muestra que la primera dimensión separa *GRUPOS PRIMARIOS* y *MUCHEDUMBRES* respecto a los demás grupos. Observe que el orden de los objetos a lo largo del eje vertical no cambia en ninguno de los gráficos de la fila superior; cada uno de estos gráficos utiliza la dimensión 1 como eje Y.

La fila intermedia de gráficos permite la interpretación de la dimensión 2. La segunda dimensión ha cambiado ligeramente respecto a la solución de dos dimensiones. Anteriormente, la segunda dimensión presentaba tres grupos diferenciados, pero ahora los objetos están más dispersos a lo largo del eje.

La tercera dimensión ayuda a separar *MUCHEDUMBRES* de *GRUPOS PRIMARIOS*, lo cual no ocurría en la solución de dos dimensiones.

Ahora observaremos con más detalle los gráficos de la dimensión 2 respecto a la dimensión 3 y de la dimensión 1 respecto a la dimensión 2. En el plano definido por las dimensiones 2 y 3, los objetos forman aproximadamente un rectángulo, con *MUCHEDUMBRES*, *COMUNIDAD MODERNA*, *GRUPOS SECUNDARIOS* y *PÚBLICOS* en los vértices. En este plano, *MUCHEDUMBRES* y *GRUPOS PRIMARIOS* parecen ser combinaciones convexas de *PÚBLICO-MUCHEDUMBRES* y *GRUPOS SECUNDARIOS-COMUNIDAD MODERNA*, respectivamente. Sin embargo, como se ha mencionado previamente, se separan de los demás grupos a lo largo de la dimensión 1. *ESPECTADORES* no se separa de los demás grupos a lo largo de la dimensión 1 y parece ser una combinación de *MUCHEDUMBRES* y *COMUNIDAD MODERNA*.

Saturaciones en componentes

Figura 10-17
Saturaciones en componentes en tres dimensiones

	Dimensión		
	1	2	3
Intensidad	,980	-,005	-,201
Frecuencia	,521	-,643	,561
Pertenencia	,980	-,002	-,197
Proximidad	,519	,656	,549
Formalismo	,981	,004	-,193

Normalización principal por variable.

Conocer la diferenciación existente entre los objetos no nos informa de qué variables corresponden a qué dimensiones. Para ello, se utilizan las saturaciones en componentes. La primera dimensión corresponde principalmente a *Pertenencia*, *Intensidad* y *Formalismo*; la segunda dimensión diferencia *Frecuencia* y *Proximidad*; y la tercera dimensión separa éstas de las otras.

Ejemplo: sintomatología de los trastornos de la alimentación

Los trastornos de la alimentación son enfermedades que producen un debilitamiento y que están asociadas con problemas con la conducta de alimentación, una grave distorsión de la imagen del propio cuerpo y una obsesión con el peso que afecta simultáneamente tanto al cuerpo como a la mente. Millones de personas se ven afectados todos los años, estando especialmente en peligro los adolescentes. Hay disponibles tratamientos, la mayoría de los cuales son útiles cuando se detecta este problema de manera temprana.

Un profesional sanitario puede intentar diagnosticar un trastorno de la alimentación mediante una evaluación psicológica y médica. Sin embargo, puede ser difícil asignarle a un paciente uno de los diferentes tipos de trastornos de la alimentación porque no hay ninguna sintomatología normalizada del comportamiento anoréxico o bulímico. ¿Existen síntomas que diferencien claramente a los pacientes en estos cuatro grupos? ¿Qué síntomas tienen en común?

Para intentar responder a estas preguntas, los investigadores (Van der Ham, Meulman, Van Strien, y Van Engeland, 1997) realizaron un estudio de 55 adolescentes con trastornos de la alimentación conocidos, como se muestran en la siguiente tabla.

Tabla 10-2
Diagnósticos de los pacientes

Diagnóstico	Número de pacientes
Anorexia nerviosa	25
Anorexia con bulimia nerviosa	9
Bulimia nerviosa tras anorexia	14
Trastorno atípico de la alimentación	7
Total	55

Cada paciente fue examinado cuatro veces durante cuatro años, lo que representa un total de 220 observaciones. En cada observación, se puntuó a los pacientes para cada uno de los 16 síntomas que aparecen en la siguiente tabla. Faltan las puntuaciones de los síntomas para el paciente 71 en el tiempo 2, el paciente 76 en el tiempo 2 y el paciente 47 en el tiempo 3, lo que nos deja 217

observaciones válidas. Los datos se pueden encontrar en *anorectic.sav*. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en *IBM SPSS Categories 20*.

Tabla 10-3

Subescalas de Morgan-Russell modificadas que miden el bienestar

Nombre de variable	Etiqueta de variable	Extremo inferior (puntuación1)	Extremo superior (puntuación 3 ó 4)
<i>weight</i>	Peso corporal	Fuera del intervalo normal	Normal
<i>mens</i>	Menstruación	Amenorrea	Períodos regulares
<i>fast</i>	Restricción de ingesta de alimentos (ayuno)	Menor de 1200 calorías	Comidas normales/regulares
<i>binge</i>	Comilonas excesivas	Más de una vez a la semana	Sin comilonas excesivas
<i>vomit</i>	Vómitos	Más de una vez a la semana	Sin vómitos
<i>purge</i>	Purgas	Más de una vez a la semana	Sin purgas
<i>hyper</i>	Hiperactividad	No puede descansar	Sin hiperactividad
<i>fami</i>	Relaciones familiares	Pobres	Buenos
<i>eman</i>	Emancipación de la familia	Muy dependiente	Adecuada
<i>frie</i>	Amigos	Sin buenos amigos	Dos o más buenos amigos
<i>school</i>	Historial escolar/laboral	Ha interrumpido escuela/trabajo	Registro de moderado a bueno
<i>satt</i>	Actitud sexual	Inadecuado	Adecuada
<i>sbeh</i>	Comportamiento sexual	Inadecuado	Puede disfrutar del sexo
<i>mood</i>	Estado mental (ánimo)	Muy deprimido	Normal
<i>preo</i>	Preocupación por la comida y el peso	Completa	Sin preocupación
<i>body</i>	Percepción corporal	Trastornada	Normal

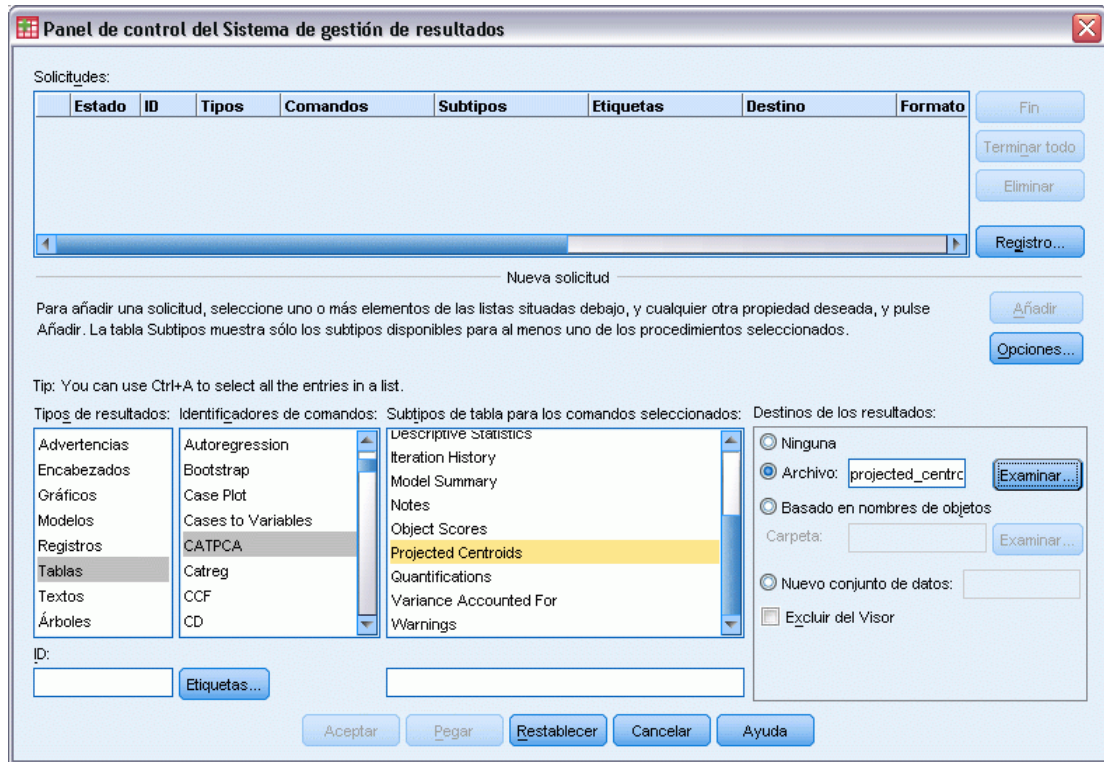
El análisis de componentes principales es ideal para esta situación ya que el objetivo del estudio es determinar las relaciones entre los síntomas y las diferentes clases de trastornos de la alimentación. Además, es probable que el análisis de componentes principales categórico resulte más útil que el análisis de componentes principales clásico ya que los síntomas se puntúan en una escala ordinal.

Ejecución del análisis

Para examinar correctamente la estructura del transcurso de la enfermedad para cada diagnóstico, probablemente desee que los resultados de la tabla de centroides proyectados estén disponibles para los diagramas de dispersión. Puede llevar esto a cabo mediante el Sistema de gestión de resultados.

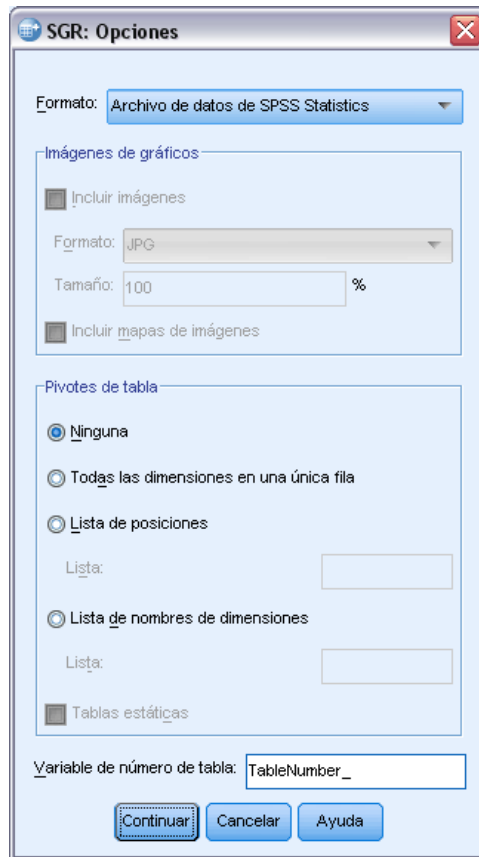
- Para iniciar una solicitud de SGR, elija en los menús:
Utilidades > Panel de control de SGR...

Figura 10-18
Panel de control del Sistema de gestión de resultados



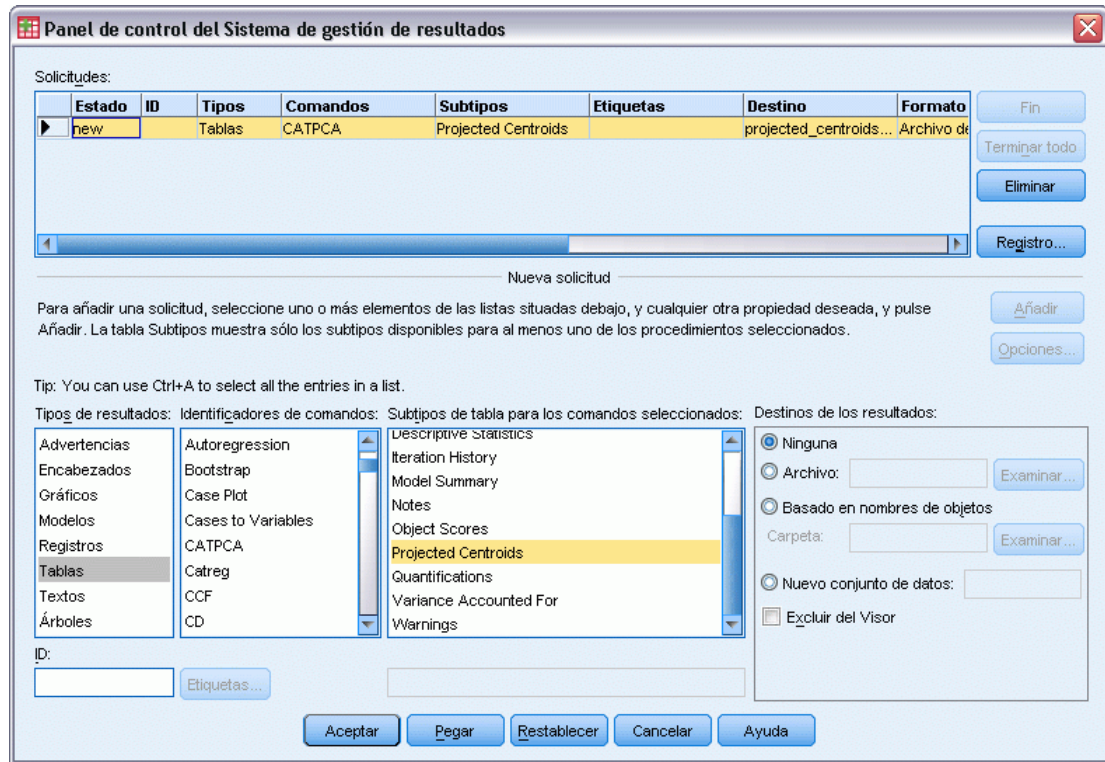
- ▶ Seleccione Tablas como tipo de resultado.
- ▶ Seleccione CATPCA como comando.
- ▶ Seleccione Centroides proyectados como tipo de tabla.
- ▶ Seleccione Archivo en el grupo Destinos de los resultados y escriba centroides_proyectados.sav como nombre de archivo.
- ▶ Pulse en Opciones.

Figura 10-19
Cuadro de diálogo Opciones



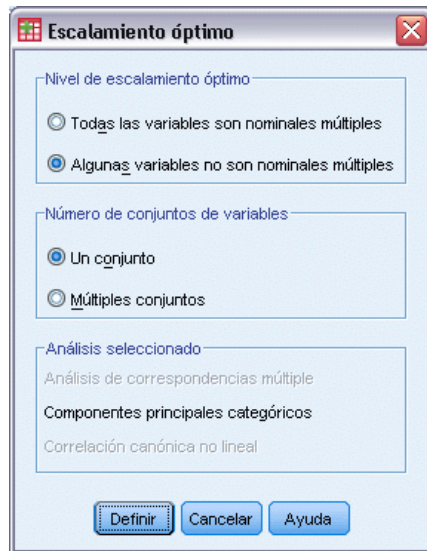
- ▶ Seleccione IBM® SPSS® Statistics Archivo de datos como formato de los resultados.
- ▶ Escriba NúmeroTabla_1 como variable de número de tabla.
- ▶ Pulse en Continuar.

Figura 10-20
Panel de control del Sistema de gestión de resultados



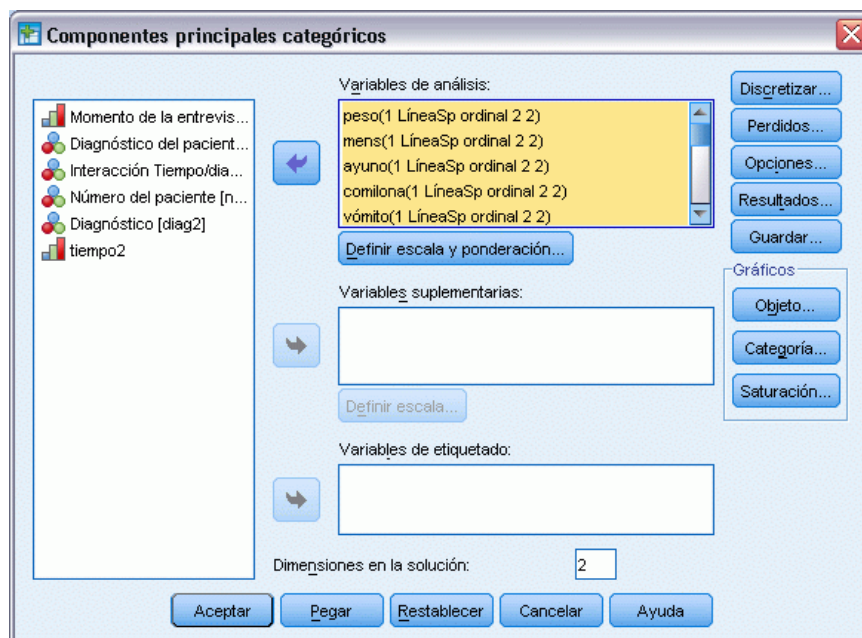
- ▶ Pulse en Añadir.
- ▶ Pulse en Aceptar y, a continuación, en Aceptar para confirmar la sesión de SGR.
El Sistema de gestión de resultados ya estará configurado para escribir los resultados de la tabla de centroides proyectados en el archivo *centroides_proyectados.sav*.
- ▶ Para generar los resultados de componentes principales categóricos de este conjunto de datos, seleccione en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 10-21
Cuadro de diálogo Escalamiento óptimo



- ▶ Seleccione *Algunas variables no es nominal múltiple* en el grupo Nivel para escalamiento óptimo.
- ▶ Pulse en Definir.

Figura 10-22
Cuadro de diálogo Componentes principales categórico



- ▶ Seleccione de *Peso corporal* a *Percepción corporal* como variables de análisis.
- ▶ Pulse en Definir escala y ponderación.

Figura 10-23
Definir escala y ponderación

Componentes principales categóricos: Definir escala y ponderación

Ponderación de la variable: 1

Nivel de escalamiento óptimo

LíneaSp ordinal Ordinal

LíneaSp nominal Nominal

Nominal múltiple Numérico

LíneaSp

Orden: 2 Nudos interiores: 2

Continuar Cancelar Ayuda

- ▶ Seleccione Ordinal como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.
- ▶ Seleccione *Interacción tiempo/diagnóstico* como variable suplementaria y pulse en Definir escala en el cuadro de diálogo Componentes principales categóricos.

Figura 10-24
Cuadro de diálogo Definir escala

Componentes principales categóricos: Definir escala

Nivel de escalamiento óptimo

LíneaSp ordinal Ordinal

LíneaSp nominal Nominal

Nominal múltiple Numérico

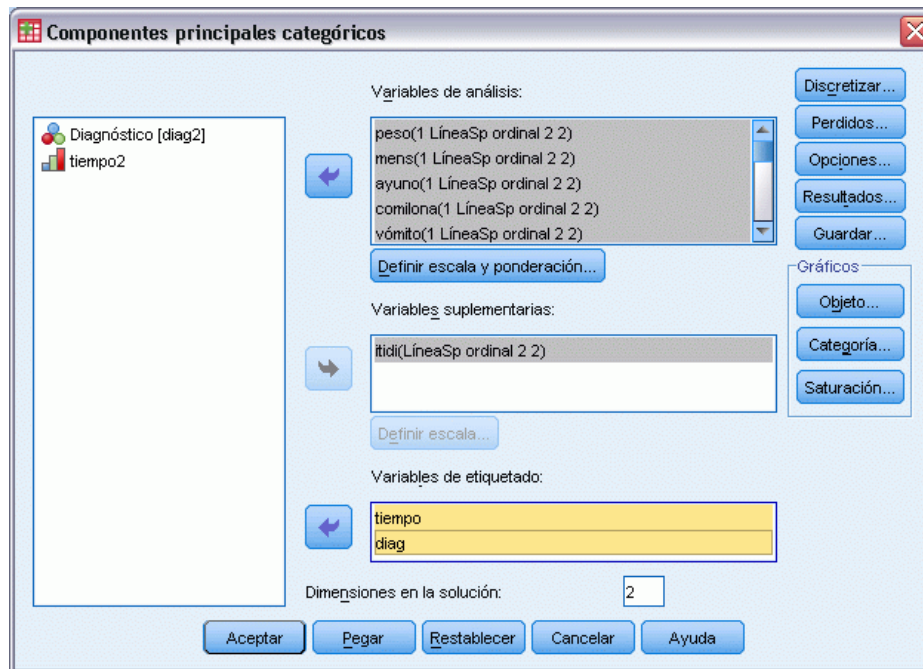
LíneaSp

Orden: 2 Nudos interiores: 2

Continuar Cancelar Ayuda

- ▶ Seleccione Nominal múltiple como nivel de escalamiento óptimo.
- ▶ Pulse en Continuar.

Figura 10-25
Cuadro de diálogo Componentes principales categórico



- ▶ Seleccione desde *Momento de la entrevista* hasta *Número del paciente* como variables de etiquetado.
- ▶ Pulse en Opciones.

Figura 10-26
Cuadro de diálogo Opciones

Componentes principales categóricos: Opciones

Objetos suplementarios

Rango de casos

Primero:

Último:

Caso único:

Añadir

Cambiar

Eliminar

Método de normalización

Principal por variable

Valor personalizado:

Criterios

Convergencia:

Iteraciones máximas:

Etiquetar gráficos con

Etiquetas de variable o de valor

Límite de la longitud de etiqueta:

Nombres de variable o valores

Dimensiones del gráfico

Mostrar todas las dimensiones de la solución

Restringir el número de dimensiones

Dimensión menor:

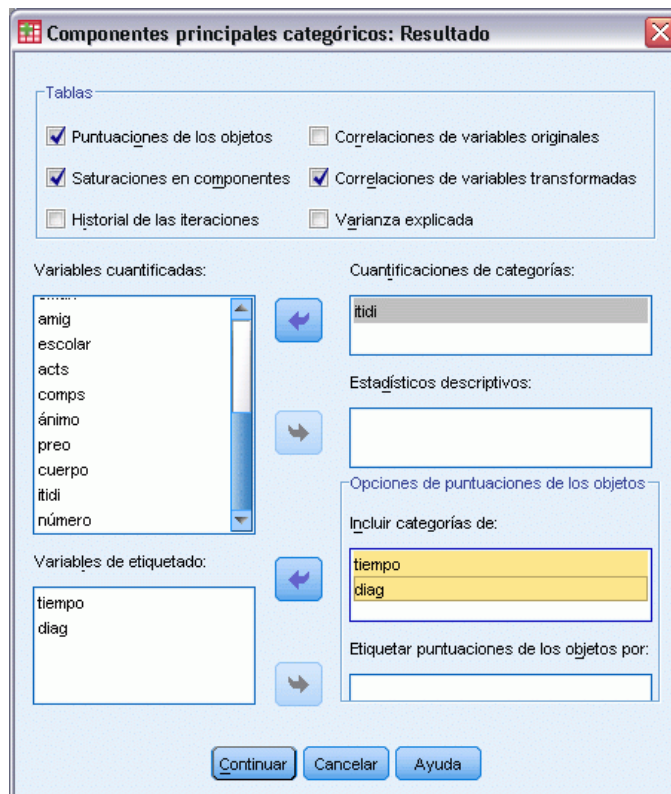
Dimensión mayor:

Configuración

Ninguna

- ▶ Seleccione etiquetar los gráficos por Nombres de variable o valores.
- ▶ Pulse en Continuar.
- ▶ Pulse en Resultados en el cuadro de diálogo Componentes principales categóricos.

Figura 10-27
Cuadro de diálogo Resultados



- ▶ Seleccione Puntuaciones de los objetos en el grupo Tablas.
- ▶ Solicite las cuantificaciones de las categorías por *tidi*.
- ▶ Seleccione incluir las categorías de *time*, *diag* y *number*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Guardar en el cuadro de diálogo Componentes principales categóricos.

Figura 10-28
Cuadro de diálogo Guardar

Componentes principales categóricos: Guardar

Datos discretizados

Crear datos discretizados

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos
Archivo...

Variables transformadas

Guardar en el conjunto de datos activo

Crear variables

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos
Archivo...

Puntuaciones de los objetos

Guardar en el conjunto de datos activo

Crear puntuaciones de los objetos

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

Escribir un nuevo archivo de datos
Archivo...

Aproximaciones

Guardar en conjunto de datos activo

Crear aproximaciones

Crear un nuevo conjunto de datos
Nombre de conjunto de datos:

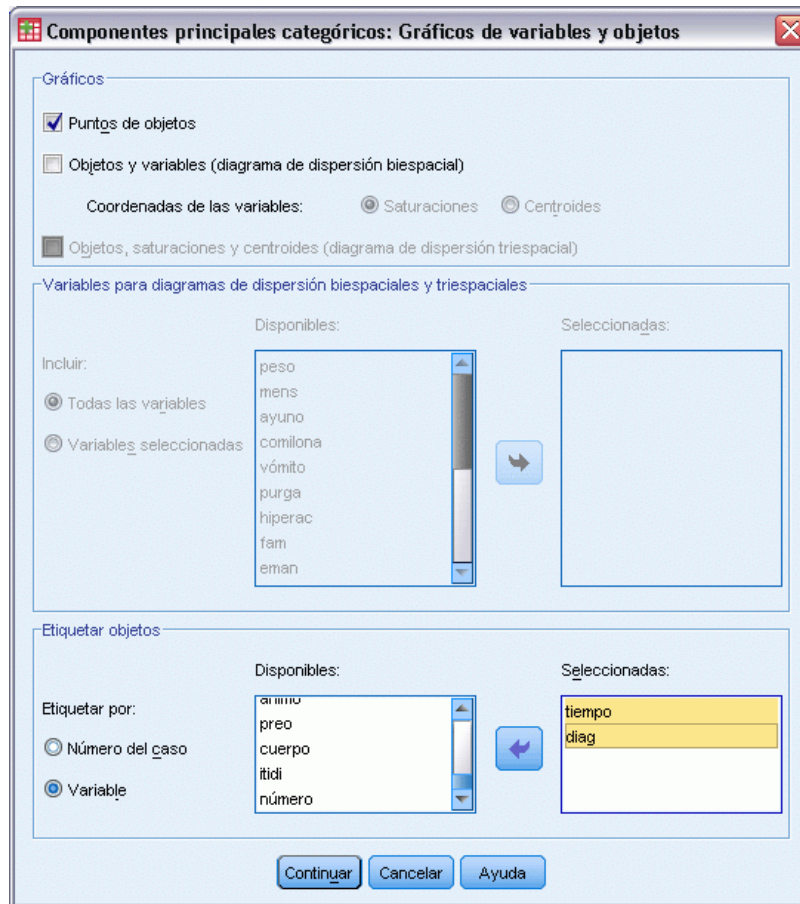
Escribir un nuevo archivo de datos
Archivo...

Dimensiones nominales múltiples: Todo Primero:

Continuar Cancelar Ayuda

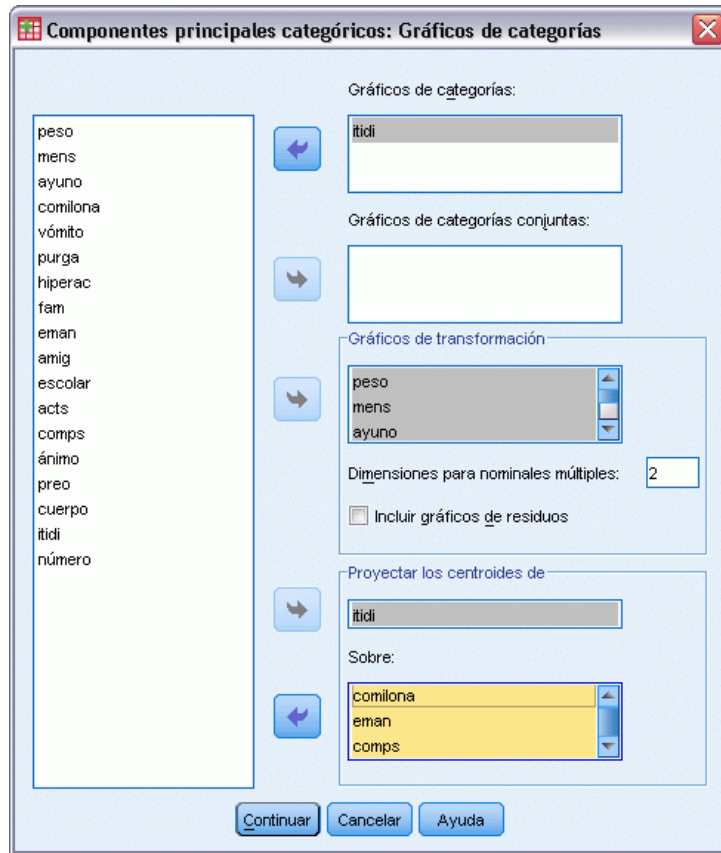
- ▶ En el grupo Variables transformadas, seleccione Guardar en el conjunto de datos activo.
- ▶ Pulse en Continuar.
- ▶ Pulse en Objeto en el cuadro de diálogo Componentes principales categóricos.

Figura 10-29
Gráficos de variables y objetos



- ▶ Seleccione etiquetar los objetos por Variable.
- ▶ Seleccione *time* y *diag* como las variables por las que se etiquetarán los objetos.
- ▶ Pulse en Continuar.
- ▶ Pulse en Categoría en el cuadro de diálogo Componentes principales categóricos.

Figura 10-30
Cuadro de diálogo Gráficos de categorías

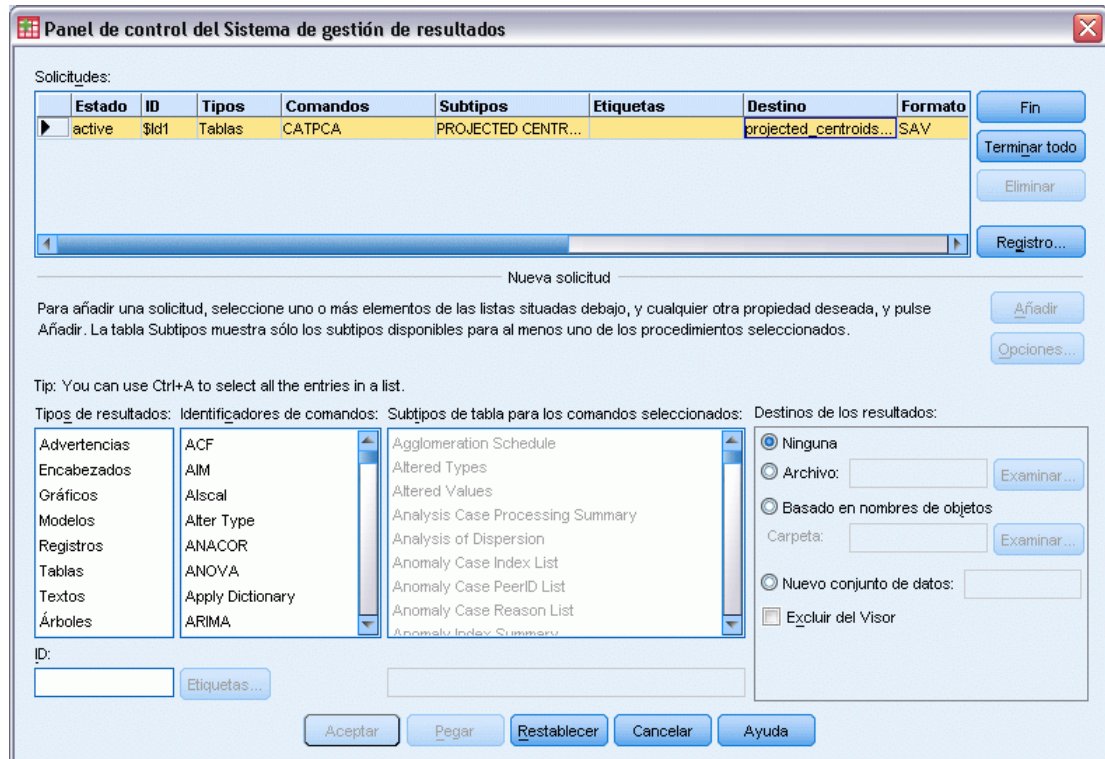


- ▶ Solicite gráficos de categorías para *tidi*.
- ▶ Solicite gráficos de transformación para de *weight* a *body*.
- ▶ Elija proyectar los centroides de *tidi* en *binge*, *satt* y *preo*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Componentes principales categóricos.

El procedimiento genera como resultados las puntuaciones de los sujetos (con media 0 y varianza la unidad) y cuantificaciones de las categorías que maximizan la correlación cuadrática media de las puntuaciones de los sujetos y de las variables transformadas. En el análisis actual, las cuantificaciones de las categorías se limitaban a reflejar la información ordinal.

Por último, para escribir la información de la tabla de centroides proyectados en *centroides_proyectados.sav*, es necesario finalizar la solicitud de SGR. Vuelva a acceder al panel de control del SGR.

Figura 10-31
Panel de control del Sistema de gestión de resultados

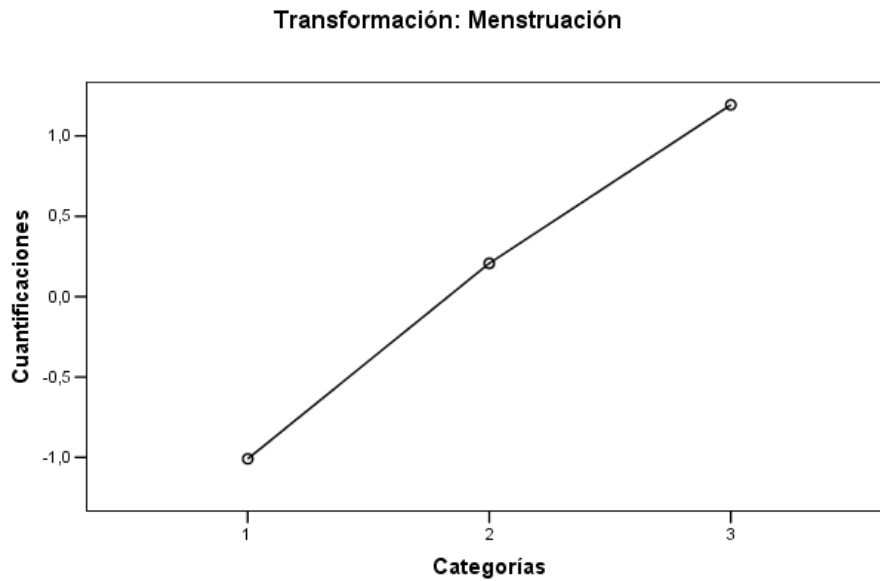


- ▶ Pulse en Terminar.
- ▶ Pulse en Aceptar y, a continuación, en Aceptar para confirmar.

Gráficos de transformación

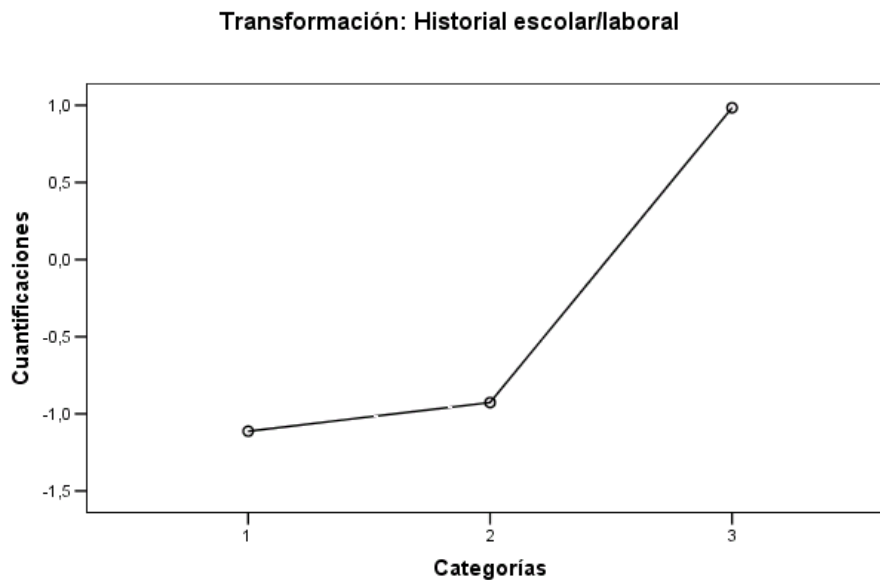
Los gráficos de transformación muestran el número de categoría original en los ejes horizontales; los ejes verticales indican las cuantificaciones óptimas.

Figura 10-32
Gráfico de transformación para menstruación



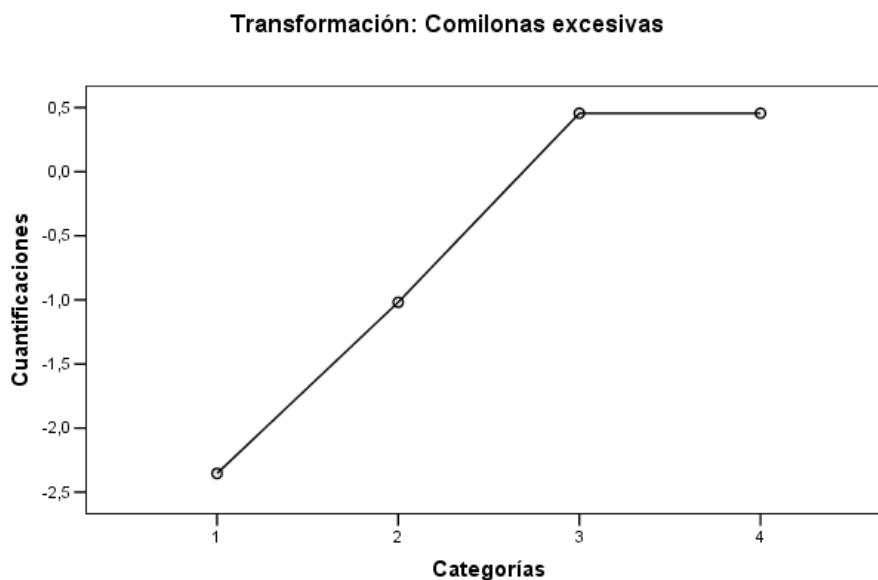
Algunas variables, como *Menstruación*, consiguen transformaciones prácticamente lineales, por lo que en este análisis se pueden interpretar como numéricas.

Figura 10-33
Gráfico de transformación para Historial escolar/laboral



Las cuantificaciones de otras variables como *Historial escolar/laboral* no han obtenido transformaciones lineales y deben interpretarse a nivel de escalamiento ordinal. La diferencia entre las categorías segunda y tercera es mucho más importante que entre las categorías primera y segunda.

Figura 10-34
Gráfico de transformación para comilonas excesivas



Un caso interesante surge en las cuantificaciones de *Comilonas excesivas*. La transformación obtenida es lineal para las categorías de la 1 a la 3, pero los valores cuantificados de las categorías 3 y 4 son iguales. Este resultado muestra que las puntuaciones de 3 y 4 no hacen diferencia entre los pacientes y sugieren que se podría utilizar el nivel de escalamiento numérico en una solución de dos componentes si se recodifican los valores de 4 como 3.

Resumen del modelo

Figura 10-35
Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	% de la varianza
1	,874	5,550	34,690
2	,522	1,957	12,234
Total	,925 ^a	7,508	46,924

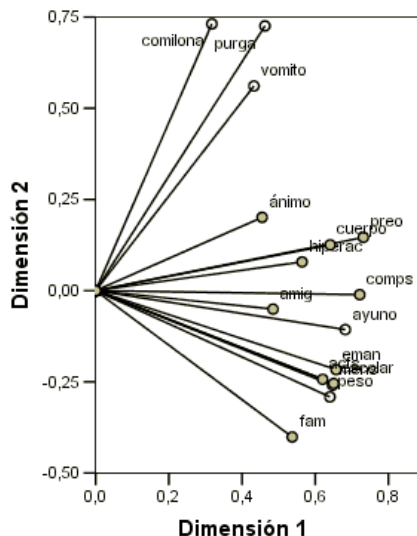
a. El Alfa de Cronbach Total está basado en los autovalores totales.

Para ver la bondad de ajuste del modelo a los datos, observemos el resumen del modelo. Aproximadamente el 47% de la varianza total queda explicada por el modelo de dos componentes, el 35% por la primera dimensión y el 12% por la segunda. Por tanto, prácticamente la mitad de la variabilidad de los objetos individuales queda explicada por el modelo de dos componentes.

Saturaciones en componentes

Para comenzar a interpretar las dos dimensiones de la solución, observemos las saturaciones en las componentes. Todas las variables tienen una saturación en componentes positiva en la primera dimensión, lo que indica que hay un factor común que tiene una correlación positiva con todas las variables.

Figura 10-36
Gráfico de saturaciones en componentes



La segunda dimensión separa las variables. Las variables *Comilonas excesivas*, *Vómitos* y *Purgas* forman un grupo que tiene grandes saturaciones positivas en la segunda dimensión. Estos síntomas suelen considerarse representativos del comportamiento bulímico.

Las variables *Emancipación de la familia*, *Historial escolar/laboral*, *Actitud sexual*, *Peso corporal* y *Menstruación* constituyen otro grupo y se puede incluir *Restricción de ingesta de alimentos (ayuno)* y *Relaciones familiares* en este grupo, ya que sus vectores se encuentran cerca del conglomerado principal y estas variables se consideran síntomas de anorexia (ayuno, peso, menstruación) o son de naturaleza psicosocial (emancipación, registro de escuela/trabajo, actitud sexual, relaciones familiares). Los vectores de este grupo son ortogonales (perpendiculares) a los vectores de binge, vomit y purge, lo que indica que este conjunto de variables no tiene correlación con el conjunto de las variables bulímicas.

Las variables *Amigos*, *Estado mental (ánimo)* e *Hiperactividad* no parecen ajustarse muy bien en la solución. Puede comprobarlo en el gráfico observando las longitudes de cada vector. La longitud del vector de una determinada variable corresponde a su ajuste y estas variables son las que tienen los vectores más cortos. En base a una solución de dos componentes, probablemente se eliminarían estas variables de una propuesta de sintomatología de los trastornos de la alimentación. Sin embargo, es posible que se ajusten mejor a una solución con un mayor número de dimensiones.

Las variables *Comportamiento sexual*, *Preocupación por la comida y el peso* y *Percepción corporal* forman otro grupo teórico de síntomas, que pertenecen a la experiencia que el paciente tiene de su propio cuerpo. Aunque están correlacionadas con los dos grupos ortogonales de variables, estas variables tienen vectores bastante largos y están asociados fuertemente con la

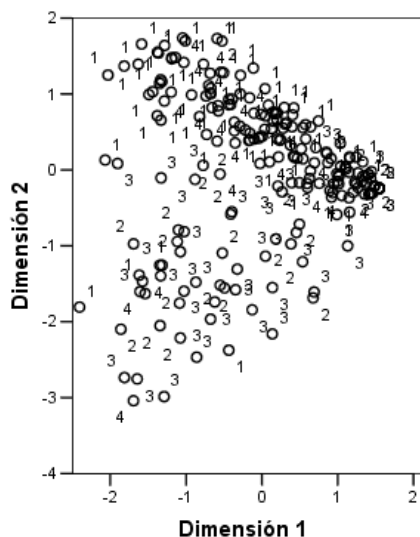
primera dimensión y, por tanto, pueden proporcionar cierta información útil acerca del factor “común”.

Puntuaciones de los objetos

La siguiente figura muestra un gráfico de las puntuaciones de los objetos, en el que los sujetos aparecen etiquetados con la categoría de su diagnóstico.

Figura 10-37

Gráficos con las puntuaciones de los objetos etiquetados por su diagnóstico

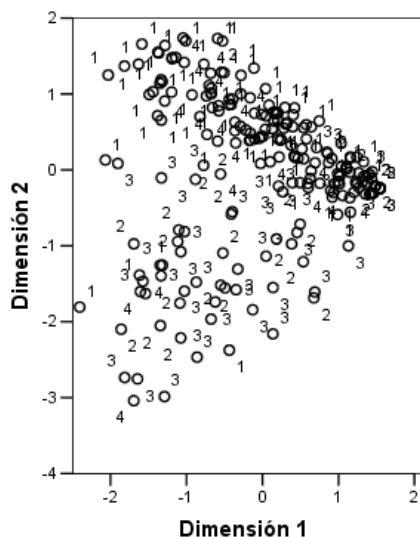


Este gráfico no ayuda a interpretar la primera dimensión ya que los pacientes no aparecen diferenciados por el diagnóstico a lo largo del mismo. Sin embargo, hay cierta información acerca de la segunda dimensión. Los sujetos de anorexia (1) y los pacientes con un trastorno atípico de la alimentación (4) forman un grupo, situado por encima de los sujetos con algún tipo de bulimia (2 y 3). Por tanto, la segunda dimensión diferencia los pacientes bulímicos de los demás, como ya ha visto en la sección anterior (las variables del grupo bulímico tienen saturaciones en componentes grandes y positivas en la segunda dimensión). Esto tiene sentido, ya que las saturaciones en componentes de los síntomas que tradicionalmente se asocian a la bulimia tienen valores grandes en la segunda dimensión.

Esta figura muestra un gráfico de las puntuaciones de los objetos, en el que los sujetos aparecen etiquetados con el tiempo en el que se realizó el diagnóstico.

Figura 10-38

Puntuaciones de los objetos etiquetados por el momento de la entrevista



El etiquetado de las puntuaciones de los objetos por el tiempo muestra que la primera dimensión tiene una relación con el tiempo ya que parece haber una progresión de tiempos de diagnóstico desde el 1 que se encuentra más a la izquierda y los otros a la derecha. Observe que puede conectar los puntos temporales de este gráfico guardando las puntuaciones de los objetos y creando un diagrama de dispersión utilizando las puntuaciones de la dimensión 1 en el eje X , las puntuaciones de la dimensión 2 en el eje Y y estableciendo los marcados mediante los números de los pacientes.

Si se compara el gráfico de las puntuaciones de los objetos etiquetados por tiempo con el gráfico etiquetado por diagnóstico podemos obtener cierta información acerca de los objetos menos habituales. Por ejemplo, en el gráfico etiquetado por tiempo, hay un paciente cuyo diagnóstico en el tiempo 4 se encuentra a la izquierda de los demás puntos del gráfico. Esto no es habitual ya que la tendencia general de los puntos es que en los tiempos posteriores se encuentren a la derecha. Curiosamente, este punto que parece estar descolocado en el tiempo también tiene un diagnóstico poco habitual, ya que se trata de un paciente anoréxico cuya puntuación coloca al paciente en el conglomerado bulímico. Si observamos la tabla con las puntuaciones de los objetos, veremos que se trata del paciente 43, diagnosticado con anorexia nerviosa, cuyas puntuaciones de los objetos se muestran en la siguiente tabla.

Tabla 10-4

Puntuaciones de los objetos para el paciente 43

Hora	Dimensión 1	Dimensión 2
1	-2.031	1.250
2	-2.067	0.131
3	-1.575	-1.467
4	-2.405	-1.807

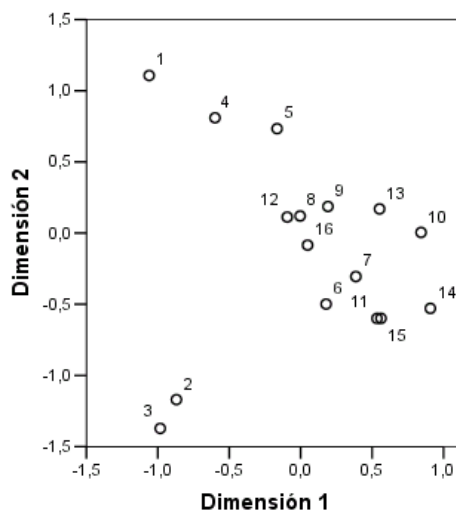
Las puntuaciones del paciente en el tiempo 1 son prototípicas de la anorexia, con una elevada puntuación negativa en la dimensión 1 correspondiente a la pobre imagen del cuerpo y una puntuación positiva en la dimensión 2 correspondiente a síntomas de anorexia o pobre comportamiento psicosocial. Sin embargo, a diferencia de la mayoría de los pacientes, hay poco o nulo progreso en la dimensión 1. En la dimensión 2, parece haber cierto progreso hacia “normal” (alrededor de 0, entre comportamiento anoréxico y bulímico), pero a continuación el paciente pasa a presentar síntomas de bulimia.

Examen de la estructura del transcurso de la enfermedad

Para buscar más información acerca de cómo se encuentran relacionadas las dos dimensiones para las cuatro características de diagnóstico y los cuatro puntos temporales, se ha creado una variable suplementaria *Interacción tiempo/diagnóstico* mediante una clasificación cruzada de las cuatro categorías de *Diagnóstico del paciente* y las cuatro categorías de *Momento de la entrevista*. Por tanto, *Interacción tiempo/diagnóstico* tiene 16 categorías, donde la primera categoría indica los pacientes de anorexia nerviosa en su primera visita. La quinta categoría indica los pacientes de la anorexia nerviosa en el punto temporal 2, etc. y la categoría decimosexta indica los pacientes de un trastorno atípico de la alimentación en el punto temporal 4. El uso de la variable suplementaria *Interacción tiempo/diagnóstico* permite el estudio de los transcurros de la enfermedad para los diferentes grupos con el tiempo. A la variable se le dio un nivel de escalamiento nominal múltiple y los puntos de categoría se muestran en la siguiente figura.

Figura 10-39

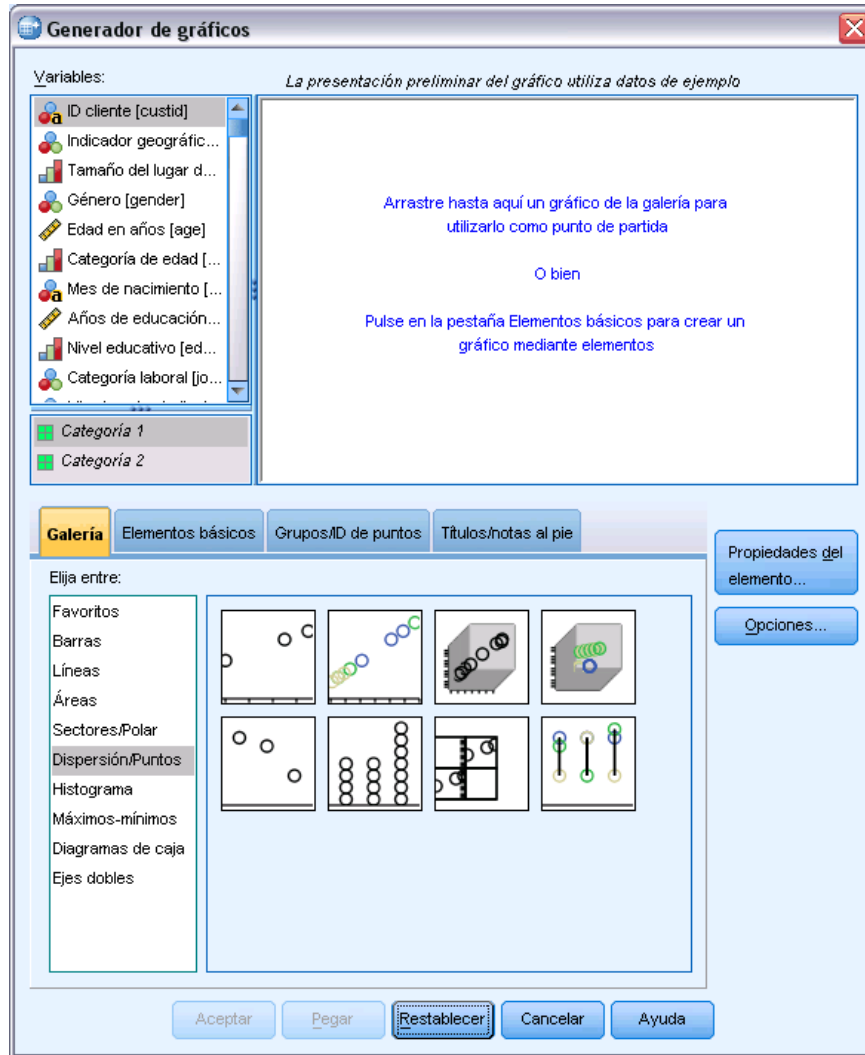
Puntos de categoría para la interacción tiempo/diagnóstico



Parte de la estructura se ve directamente en este gráfico: las categorías de diagnóstico en el punto temporal 1 diferencian claramente la anorexia nerviosa y el trastorno atípico de la alimentación de la anorexia nerviosa con la bulimia nerviosa y la bulimia nerviosa tras anorexia nerviosa en la segunda dimensión. Aparte de esto, resulta un poco más difícil ver los patrones.

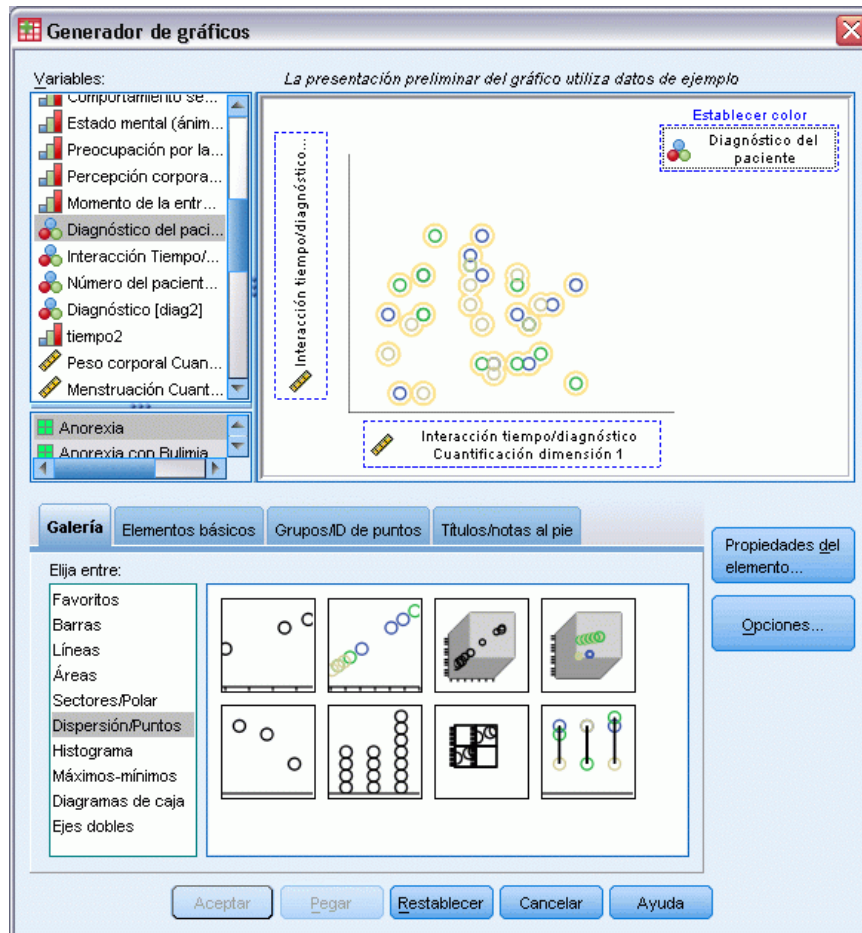
No obstante, puede conseguir que los patrones sean visibles con mayor facilidad si crea un diagrama de dispersión basado en las cuantificaciones. Para ello, elija en los menús: Gráficos > Generador de gráficos...

Figura 10-40
Galería Dispersión/Puntos



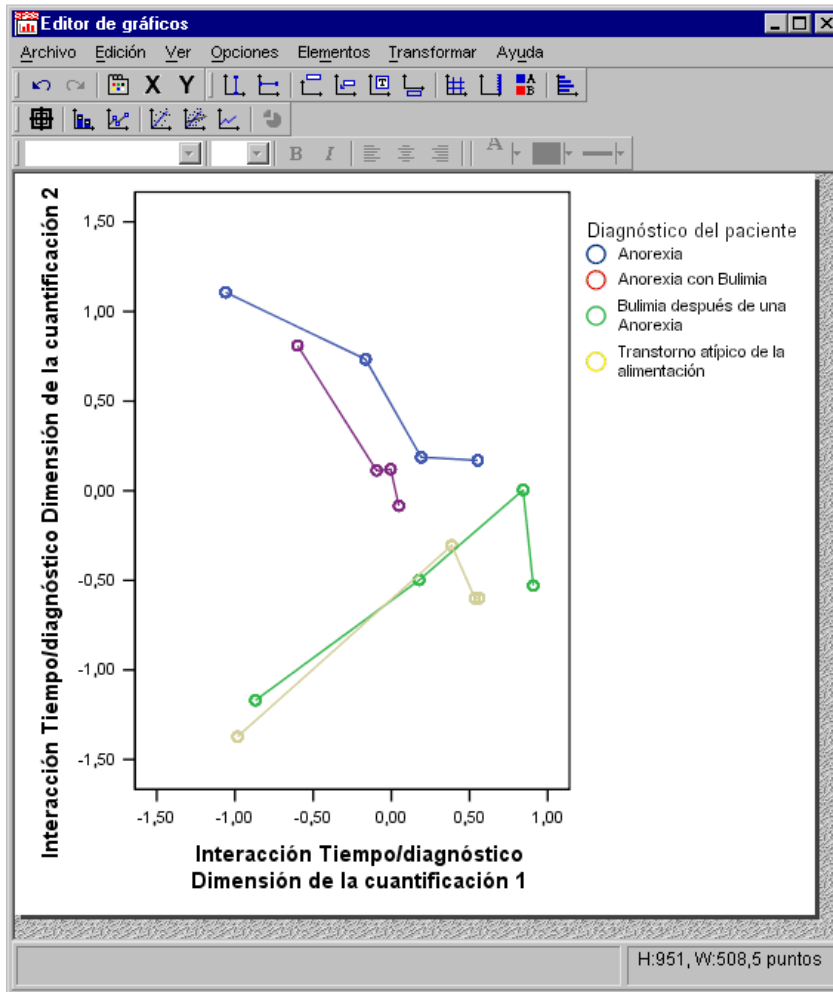
- Seleccione la galería Dispersión/Puntos y elija Dispersión Agrupado.

Figura 10-41
Generador de gráficos



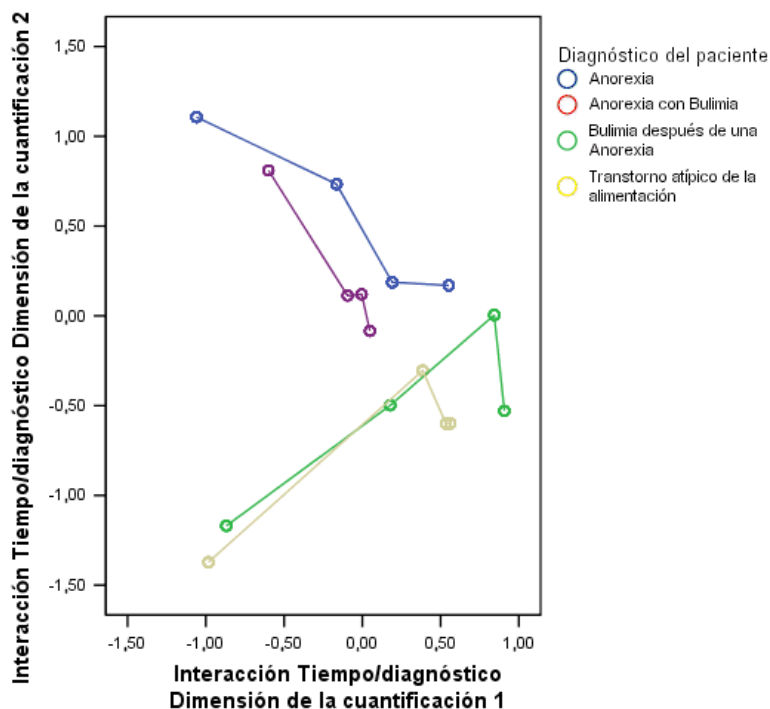
- ▶ Seleccione *Interacción tiempo/diagnóstico Cuantificación dimensión 2* como variable y e *Interacción tiempo/diagnóstico Cuantificación dimensión 1* como variable x .
- ▶ Seleccione establecer el color por *Diagnóstico del paciente*.
- ▶ Pulse en Aceptar.

Figura 10-42
Estructuras de los transcurso de las enfermedades



- ▶ A continuación, para conectar los puntos, pulse dos veces en el gráfico y pulse en la herramienta Añadir una línea de interpolación del Editor de gráficos.
- ▶ Cierre el Editor de gráficos.

Figura 10-43
Estructuras de los transcurso de las enfermedades



Al conectar los puntos de categoría de cada categoría de diagnóstico con el tiempo, los patrones sugieren automáticamente que la primera dimensión está relacionada con el tiempo y la segunda, con el diagnóstico, como se determinó previamente en los gráficos de las puntuaciones de los objetos.

No obstante, este gráfico muestra también que, con el tiempo, las enfermedades tienden a parecerse más. Además, para todos los grupos, el progreso es mayor entre los puntos temporales 1 y 2; los pacientes anoréxicos muestran algún progreso más de 2 a 3, pero los demás grupos muestran poco progreso.

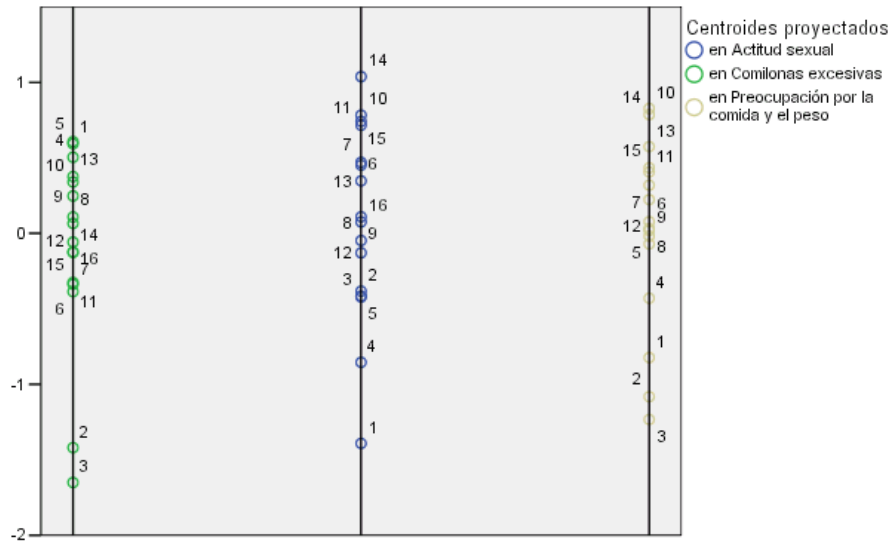
Desarrollo diferencial para las variables seleccionadas

Se seleccionó una variable de cada grupo de síntomas identificado por las saturaciones en componentes como “representativa” del grupo. La variable comilonas excesivas se seleccionó del grupo bulímico, la actitud sexual del grupo anoréxico/psicosocial y la preocupación por el cuerpo del tercer grupo.

Para examinar los posibles transcurso diferenciales de la enfermedad, las proyecciones de *Interacción tiempo/diagnóstico* sobre *Comilonas excesivas*, *Actitud sexual* y *Preocupación por la comida y el peso* se calcularon y representaron en la siguiente figura.

Figura 10-44

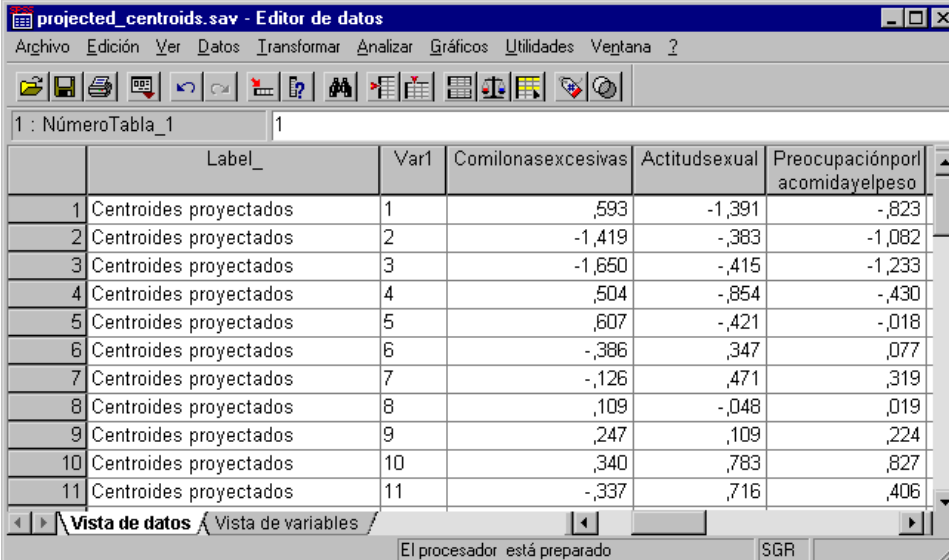
Centroides proyectados de Interacción tiempo/diagnóstico sobre Comilonas excesivas, Actitud sexual y Preocupación por la comida y el peso



Este gráfico muestra que el primer punto temporal, el síntoma de comilonas excesivas diferencia a los pacientes bulímicos (2 y 3) de los demás (1 y 4); y la actitud sexual diferencia a los pacientes anoréxicos y atípicos (1 y 4) de los demás (2 y 3); y la preocupación por el cuerpo no diferencia realmente a los pacientes. En muchas aplicaciones, este gráfico bastaría para describir la relación entre los síntomas y el diagnóstico, pero a causa de la complicación de los diferentes puntos temporales, el dibujo tiene un aspecto complicado.

Para ver estas proyecciones respecto al tiempo, es necesario poder representar el contenido de la tabla de centroides proyectados. Puede hacerse esto gracias a la solicitud del SGR que guardó esta información en *centroides_proyectados.sav*.

Figura 10-45
Centroides_proyectados.sav

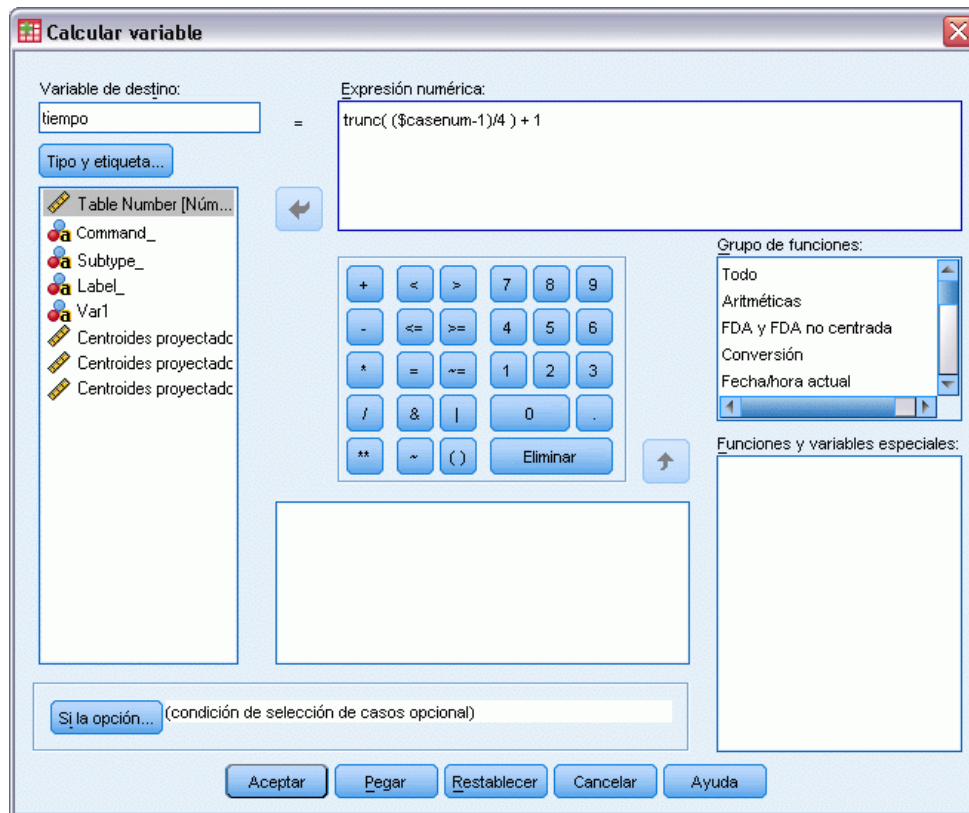


	Label_	Var1	Comilonasexcesivas	Actitudsexual	Preocupaciónporlacomidayelpeso
1	Centroides proyectados	1	,593	-1,391	-,823
2	Centroides proyectados	2	-1,419	-,383	-1,082
3	Centroides proyectados	3	-1,650	-,415	-1,233
4	Centroides proyectados	4	,504	-,854	-,430
5	Centroides proyectados	5	,607	-,421	-,018
6	Centroides proyectados	6	-,386	,347	,077
7	Centroides proyectados	7	-,126	,471	,319
8	Centroides proyectados	8	,109	-,048	,019
9	Centroides proyectados	9	,247	,109	,224
10	Centroides proyectados	10	,340	,783	,827
11	Centroides proyectados	11	-,337	,716	,406

Las variables *Comilonasexcesivas*, *Actitudsexual* y *Preocupaciónporlacomidayelpeso* contienen los valores de los centroides proyectados sobre cada uno de los síntomas de interés. El número de caso (de 1 a 16) corresponde a la interacción tiempo/diagnóstico. Deberá calcular nuevas variables que diferencien los valores Tiempo y Diagnóstico.

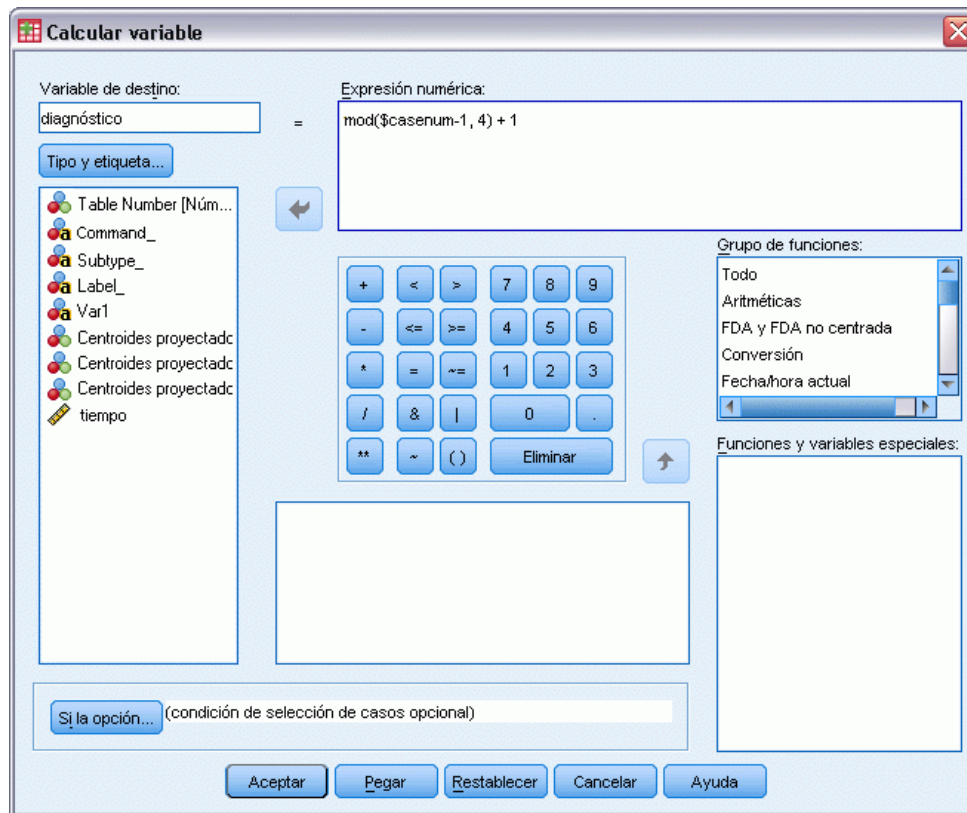
- Elija en los menús:
Transformar > Calcular variable...

Figura 10-46
Cuadro de diálogo *Calcular variable*



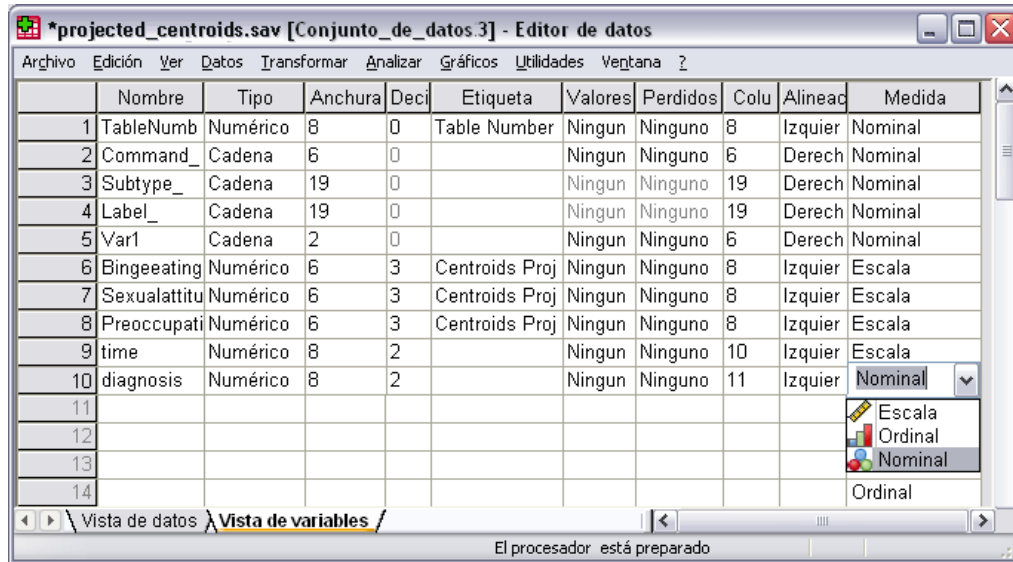
- ▶ Escriba *time* como variable de destino.
- ▶ Escriba $\text{trunc}(\$casenum-1)/4 + 1$ como expresión numérica.
- ▶ Pulse en Aceptar.

Figura 10-47
Cuadro de diálogo *Calcular variable*



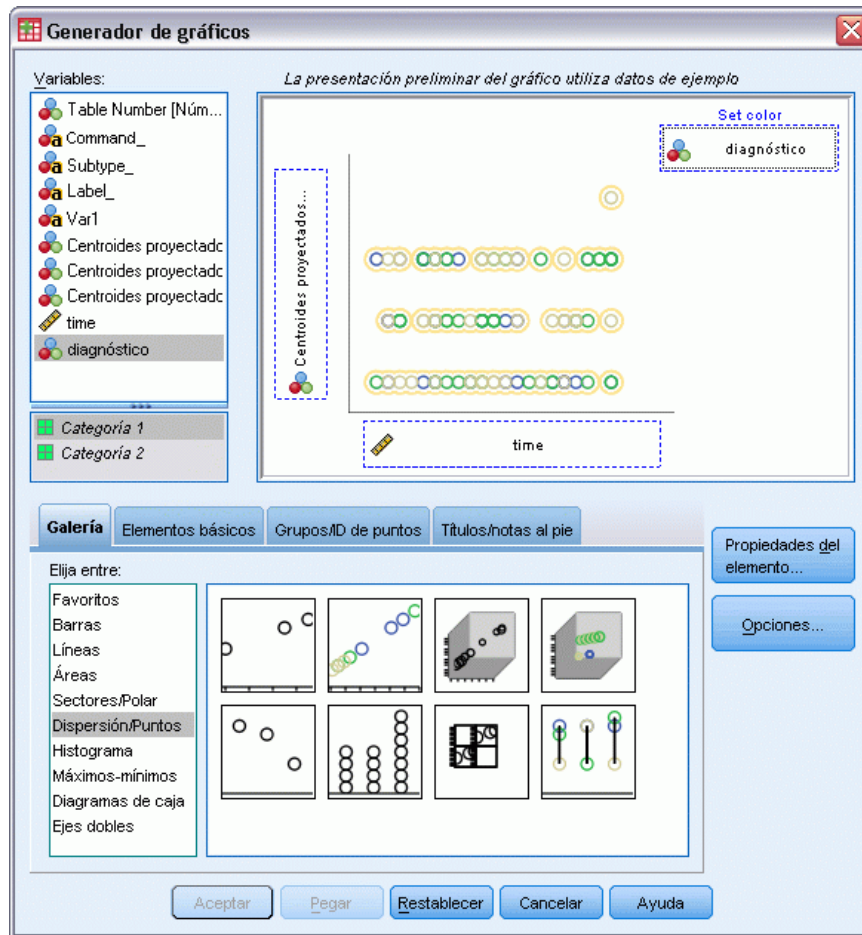
- ▶ Vuelva a abrir el cuadro de diálogo *Calcular variable*.
- ▶ Escriba *diagnóstico* como variable de destino.
- ▶ Escriba $\text{mod}(\$casenum-1, 4) + 1$ como expresión numérica.
- ▶ Pulse en *Aceptar*.

Figura 10-48
Centroides_proyectados.sav



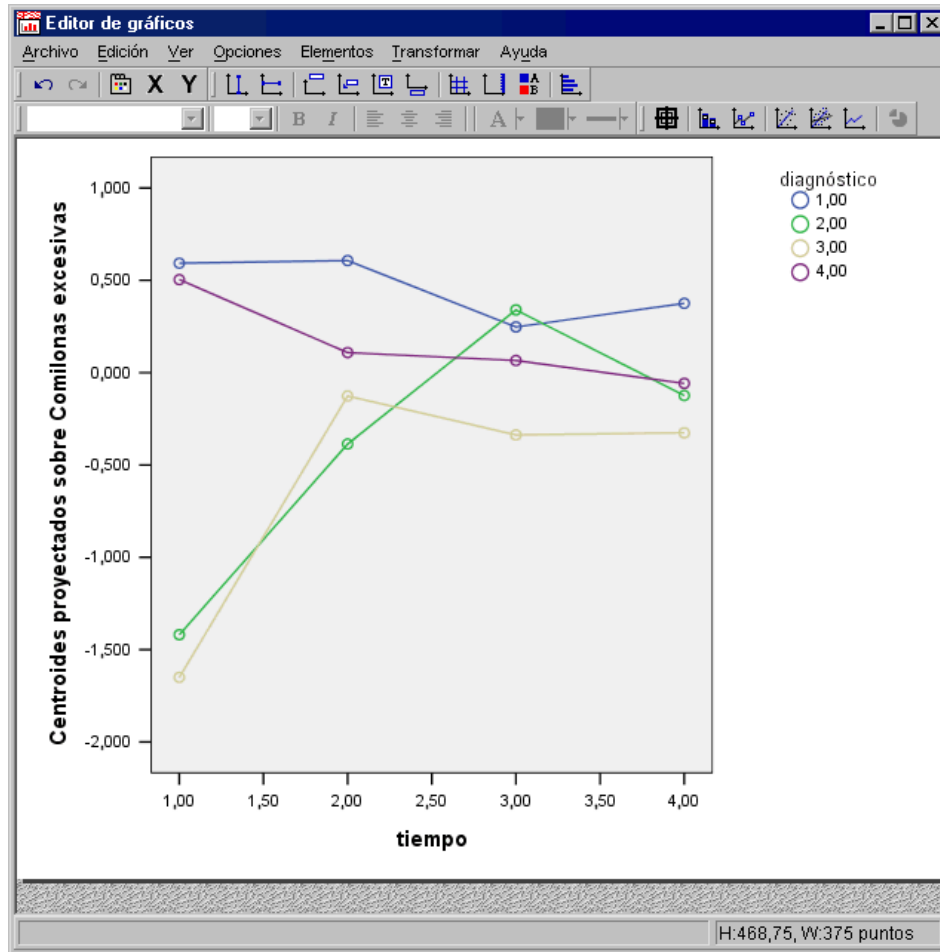
En la Vista de variables, cambie la medida de *diagnóstico* de Escala a Nominal.

Figura 10-49
Generador de gráficos



- ▶ Por último, para ver los centroides proyectados de tiempo de diagnóstico sobre la comida excesiva respecto al tiempo, vuelva a Generador de gráficos y pulse en Restablecer para borrar las selecciones anteriores.
- ▶ Seleccione la galería Dispersión/Puntos y elija Dispersión Agrupado.
- ▶ Seleccione *Centroides proyectados sobre Comilonas excesivas* como variable *y* y *time* como variable *x*.
- ▶ Seleccione establecer los colores por *diagnóstico*.
- ▶ Pulse en Aceptar.

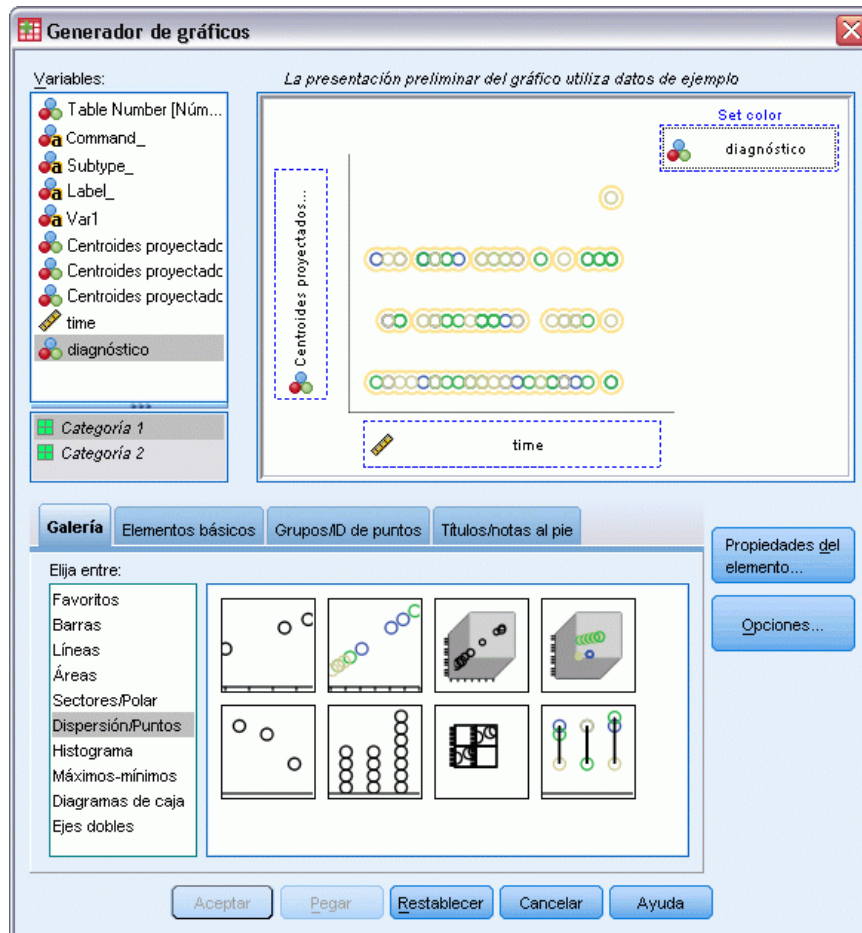
Figura 10-50
Centroides proyectados de tiempo de diagnóstico sobre las comilonas excesivas respecto al tiempo



- ▶ A continuación, para conectar los puntos, pulse dos veces en el gráfico y pulse en la herramienta Añadir una línea de interpolación del Editor de gráficos.
- ▶ Cierre el Editor de gráficos.

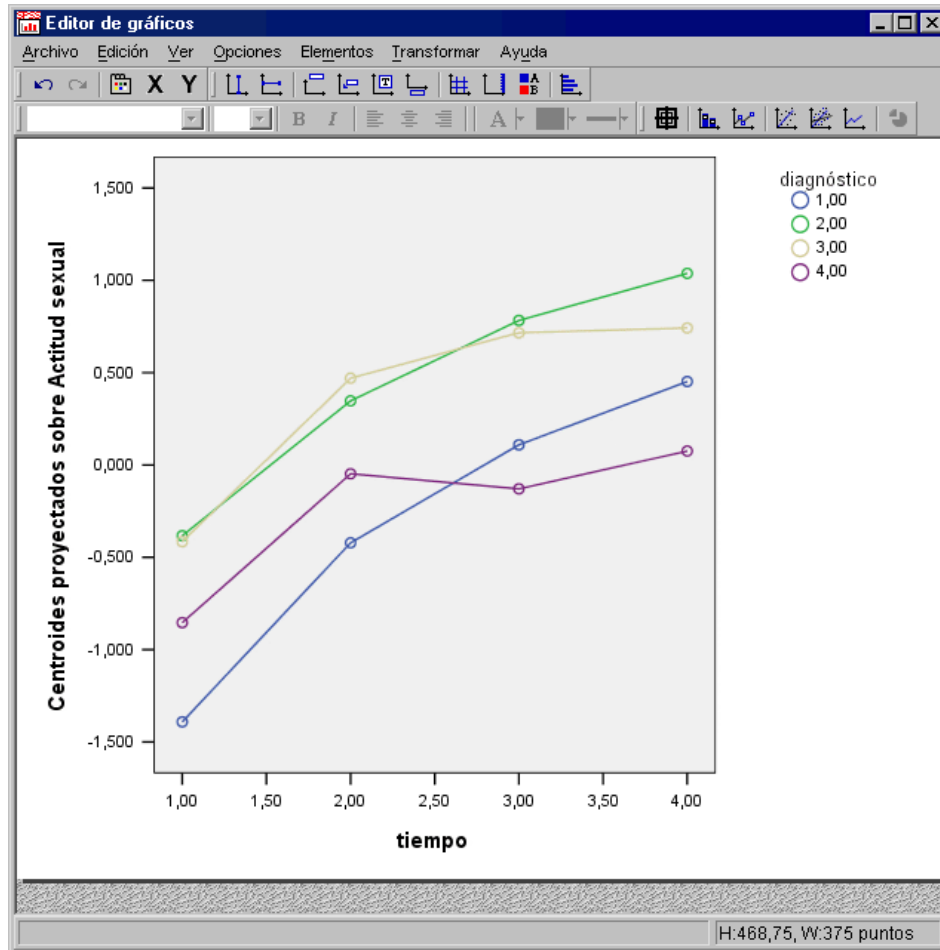
Respecto a las comilonas excesivas, resulta evidente que los grupos anoréxicos tienen diferentes valores de inicio que los grupos bulímicos. Esta diferencia se reduce con el tiempo, ya que los grupos anoréxicos apenas cambian, mientras que los grupos bulímicos muestran un progreso.

Figura 10-51
Generador de gráficos



- ▶ Recupere el Generador de gráficos.
- ▶ Anule la selección de *Centroides proyectados sobre Comilonas excesivas* como variable y seleccione *Centroides proyectados sobre Actitud sexual* como variable y.
- ▶ Pulse en Aceptar.

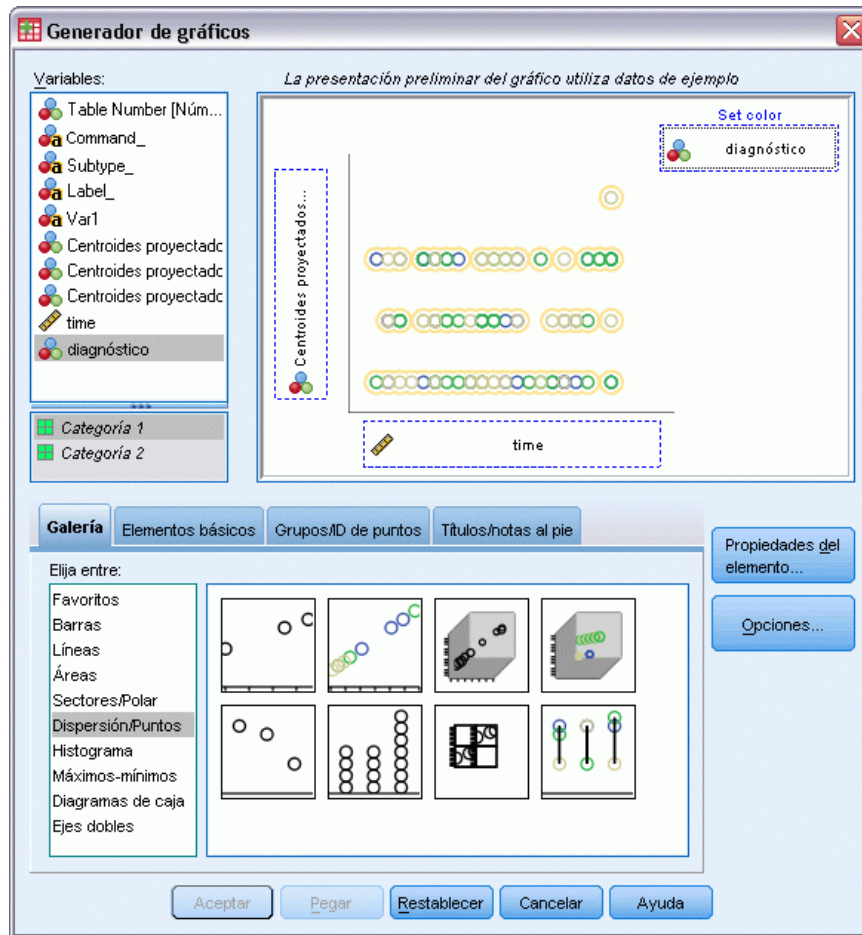
Figura 10-52
Centroides proyectados de tiempo de diagnóstico sobre la actitud sexual respecto al tiempo



- ▶ A continuación, para conectar los puntos, pulse dos veces en el gráfico y pulse en la herramienta Añadir una línea de interpolación del Editor de gráficos.
- ▶ Cierre el Editor de gráficos.

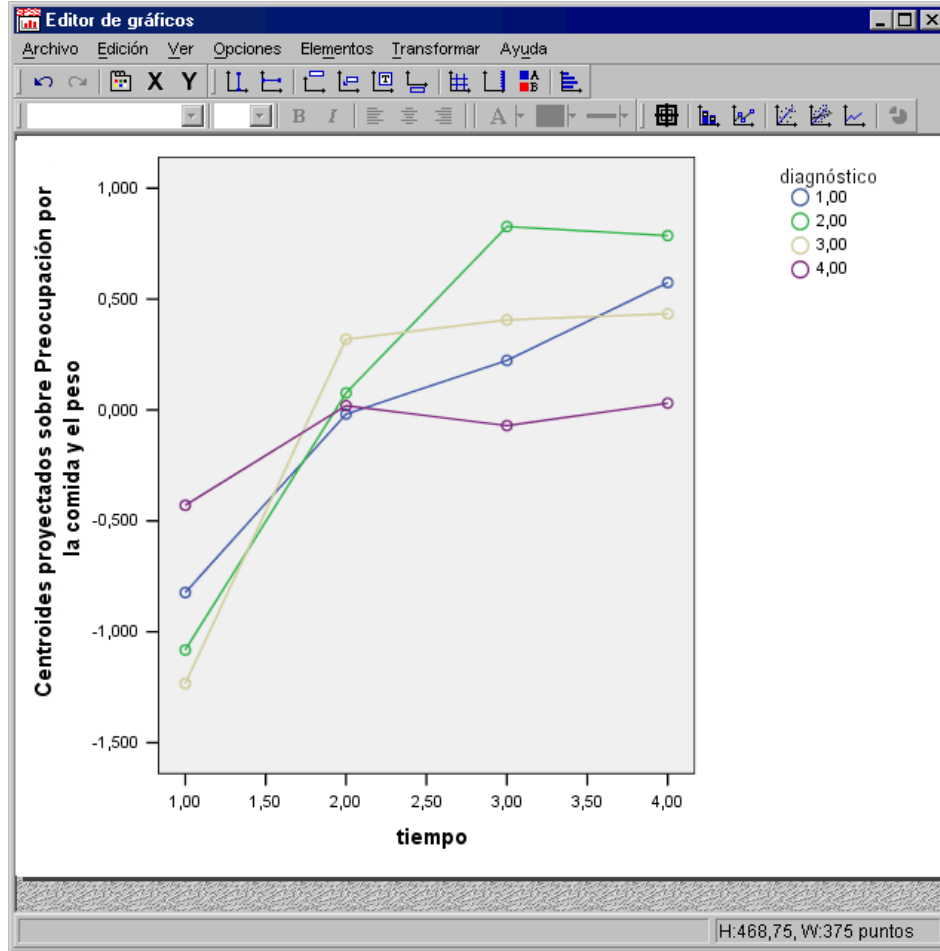
Respecto a la actitud sexual, las cuatro trayectorias son más o menos paralelas con el tiempo y todos los grupos muestran un progreso. Los grupos bulímicos, no obstante, tienen puntuaciones mayores (mejores) que el grupo anoréxico.

Figura 10-53
Generador de gráficos



- ▶ Recupere el Generador de gráficos.
- ▶ Anule la selección de *Centroides proyectados sobre Actitud sexual* como variable y seleccione *Centroides proyectados sobre Preocupación por la comida y el peso* como variable y.
- ▶ Pulse en Aceptar.

Figura 10-54
Centroides proyectados de tiempo de diagnóstico sobre la preocupación por el cuerpo respecto al tiempo



- ▶ A continuación, para conectar los puntos, pulse dos veces en el gráfico y pulse en la herramienta Añadir una línea de interpolación del Editor de gráficos.
- ▶ Cierre el Editor de gráficos.

La preocupación por el cuerpo es una variable que representa los síntomas básicos, los cuales los comparten los cuatro grupos diferentes. Además de los pacientes de trastornos de la alimentación atípicos, el grupo anoréxico y los dos grupos bulímicos tienen niveles muy similares tanto al comienzo como al final.

Lecturas recomendadas

Consulte los siguientes textos si desea obtener más información sobre el análisis de componentes principales categórico:

De Haas, M., J. A. Algera, H. F. J. M. Van Tuijl, y J. J. Meulman. 2000. Macro and micro goal setting: In search of coherence. *Applied Psychology*, 49, .

De Leeuw, J. 1982. Nonlinear principal components analysis. En: *COMPSTAT Proceedings in Computational Statistics*, Viena: Physica Verlag.

Eckart, C., y G. Young. 1936. The approximation of one matrix by another one of lower rank. *Psychometrika*, 1, .

Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal components analysis. *Biometrika*, 58, .

Gifi, A. 1985. *PRINCALS. Research Report UG-85-02*. Leiden: Department of Data Theory, University of Leiden.

Gower, J. C., y J. J. Meulman. 1993. The treatment of categorical information in physical anthropology. *International Journal of Anthropology*, 8, .

Heiser, W. J., y J. J. Meulman. 1994. Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. En: *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, M. Greenacre, y J. Blasius, eds. Nueva York: Academic Press.

Kruskal, J. B. 1978. Factor analysis and principal components analysis: Bilinear methods. En: *International Encyclopedia of Statistics*, W. H. Kruskal, y J. M. Tanur, eds. Nueva York: The Free Press.

Kruskal, J. B., y R. N. Shepard. 1974. A nonmetric variety of linear factor analysis. *Psychometrika*, 39, .

Meulman, J. J. 1993. Principal coordinates analysis with optimal transformations of the variables: Minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, 46, .

Meulman, J. J., y P. Verboon. 1993. Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, 58, .

Meulman, J. J., A. J. Van der Kooij, y A. Babinec. 2000. New features of categorical principal components analysis for complicated data sets, including data mining. En: *Classification, Automation and New Media*, W. Gaul, y G. Ritter, eds. Berlín: Springer-Verlag.

Meulman, J. J., A. J. Van der Kooij, y W. J. Heiser. 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. En: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..

Theunissen, N. C. M., J. J. Meulman, A. L. Den Ouden, H. M. Koopman, G. H. Verrips, S. P. Verloove-Vanhorick, y J. M. Wit. 2003. Changes can be studied when the measurement instrument is different at different time points. *Health Services and Outcomes Research Methodology*, 4, .

Tucker, L. R. 1960. Intra-individual and inter-individual multidimensionality. En: *Psychological Scaling: Theory & Applications*, H. Gulliksen, y S. Messick, eds. Nueva York: John Wiley and Sons.

Vlek, C., y P. J. Stallen. 1981. Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance*, 28, .

Wagenaar, W. A. 1988. *Paradoxes of gambling behaviour*. Londres: Lawrence Erlbaum Associates, Inc.

Young, F. W., Y. Takane, y J. De Leeuw. 1978. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, 43, .

Zeijl, E., Y. te Poel, M. du Bois-Reymond, J. Ravestloot, y J. J. Meulman. 2000. The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research*, 32, .

Análisis de correlación canónica no lineal

El propósito del análisis de correlación canónica no lineal es determinar el parecido entre dos o más conjuntos de variables. Al igual que en el análisis de correlación canónica no lineal, el propósito es explicar al máximo la varianza de las relaciones entre los conjuntos en un espacio de pocas dimensiones. Sin embargo, a diferencia del análisis de correlación canónica lineal, el análisis de correlación canónica no lineal no supone que haya un nivel de intervalo de medida o que las relaciones sean lineales. Otra diferencia importante es que el análisis de correlación canónica no lineal establece la similitud entre los conjuntos mediante la comparación simultánea de las combinaciones lineales de las variables en cada conjunto con un conjunto desconocido, es decir, las puntuaciones del objeto.

Ejemplo: Un análisis de los resultados de una encuesta

El ejemplo que se explica en este capítulo está tomado de una encuesta (Verdegaal, 1985). Se han registrado las respuestas de 15 sujetos a 8 variables. En la siguiente tabla se muestran las variables, las etiquetas de las variables y las etiquetas de valor (categorías) del conjunto de datos.

Tabla 11-1
Datos de la encuesta

Nombre de variable	Etiqueta de variable	Etiqueta de valor
<i>edad</i>	<i>Edad en años</i>	20–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, 66–70
<i>ecivil</i>	<i>Estado civil</i>	Soltero, casado, otro
<i>mascota</i>	<i>Mascotas que posee</i>	No, gato(s), perro(s), otro (distinto de gato o perro), varios animales domésticos
<i>noticia</i>	<i>Periódico leído con mayor frecuencia</i>	Ninguno, Telegraaf, Volkskrant, NRC, Otro
<i>música</i>	<i>Música preferida</i>	Clásica, moderna, popular, variada, no le interesa la música
<i>vivir</i>	<i>Vecindario preferido</i>	Ciudad, pueblo, campo
<i>matem</i>	<i>Puntuación prueba matemáticas</i>	0–5, 6–10, 11–15
<i>lengua</i>	<i>Puntuación prueba lengua</i>	0–5, 6–10, 11–15, 16–20

Este conjunto de datos se puede encontrar en *verd1985.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20](#). Las variables de interés son las primeras seis y se dividen en tres conjuntos. El conjunto 1 incluye *edad* y

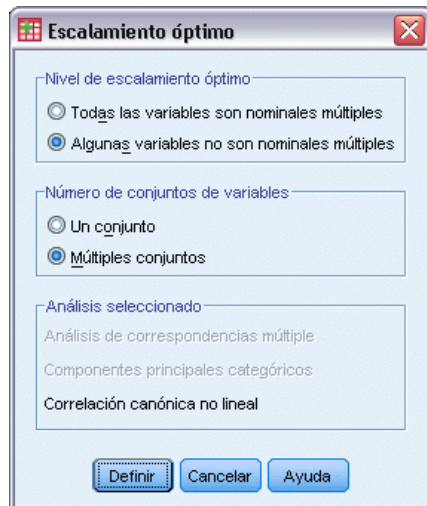
ecivil, el conjunto 2 incluye *mascota* y *noticia*, mientras que el conjunto 3 incluye *música* y *vivir*. Se escala *mascota* como nominal múltiple y *edad* como ordinal; todas las demás variables se escalan como nominal simple. Este tipo de análisis solicita una configuración inicial aleatoria. Por defecto, la configuración inicial es numérica. Sin embargo, cuando alguna de las variables se trata como nominal simple sin posibilidad de clasificación, lo mejor es escoger una configuración inicial aleatoria. Esto sucede con la mayoría de las variables de este estudio.

Examen de los datos

- Para conseguir un análisis de correlación canónica no lineal para este conjunto de datos, elija en los menús:

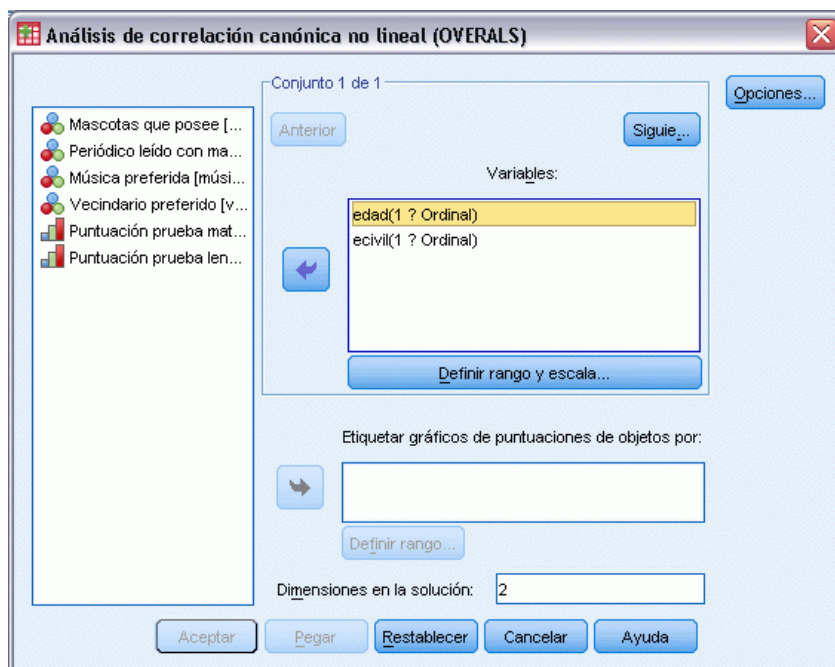
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 11-1
Cuadro de diálogo Escalamiento óptimo



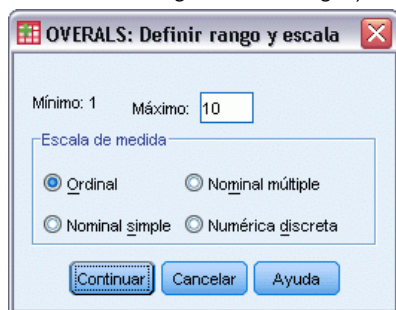
- Seleccione *Algunas variables no son nominales múltiples* en el grupo *Nivel para escalamiento óptimo*.
- Seleccione *Múltiples conjuntos* en el grupo *Número de conjuntos de variables*.
- Pulse en *Definir*.

Figura 11-2
Cuadro de diálogo Análisis de correlación canónica no lineal



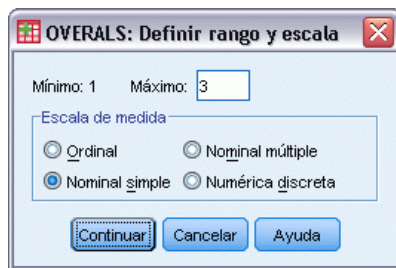
- ▶ Seleccione como variables *Edad en años* y *Estado civil* para el primer conjunto.
- ▶ Seleccione *edad* y pulse en Definir rango y escala.

Figura 11-3
Cuadro de diálogo Definir rango y escala



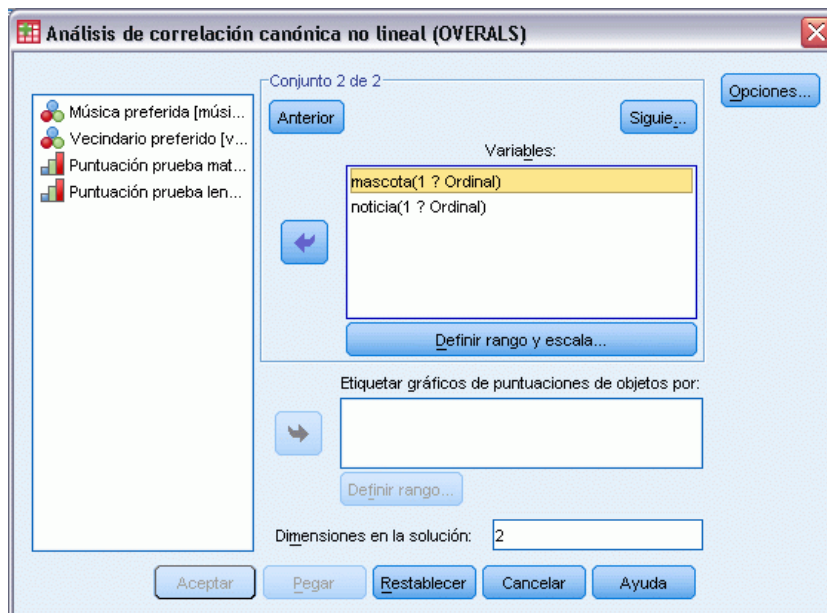
- ▶ Escriba 10 como valor máximo para esta variable.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, seleccione *ecivil* y pulse en Definir rango y escala.

Figura 11-4
Cuadro de diálogo Definir rango y escala



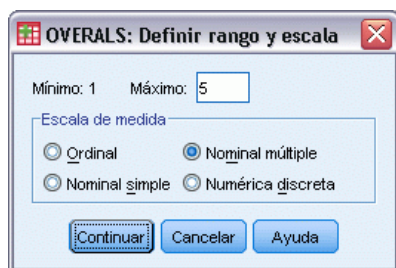
- ▶ Escriba 3 como valor máximo para esta variable.
- ▶ Seleccione como escala de la medida Nominal simple.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, pulse en Siguiete para definir el siguiente conjunto de variables.

Figura 11-5
Cuadro de diálogo Análisis de correlación canónica no lineal



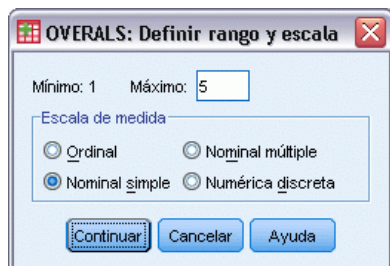
- ▶ Seleccione *Mascotas que posee* y *Periódico leído con mayor frecuencia* como variables para el segundo conjunto.
- ▶ Seleccione *mascota* y pulse en Definir rango y escala.

Figura 11-6
Cuadro de diálogo Definir rango y escala



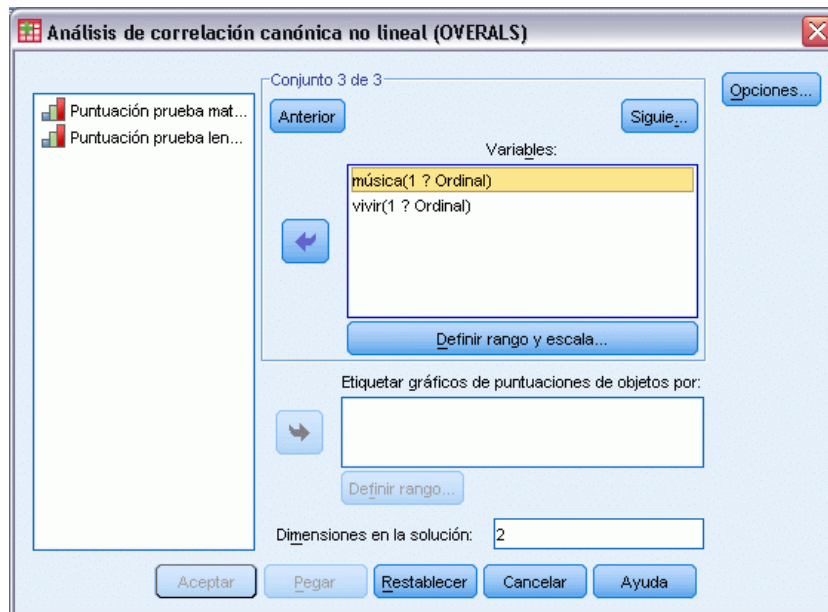
- ▶ Escriba 5 como valor máximo para esta variable.
- ▶ Seleccione como escala de la medida Nominal múltiple.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, seleccione *noticia* y pulse en Definir rango y escala.

Figura 11-7
Cuadro de diálogo Definir rango y escala



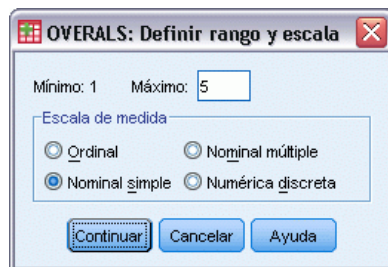
- ▶ Escriba 5 como valor máximo para esta variable.
- ▶ Seleccione como escala de la medida Nominal simple.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, pulse en Siguiete para definir el último conjunto de variables.

Figura 11-8
Cuadro de diálogo *Análisis de correlación canónica no lineal*



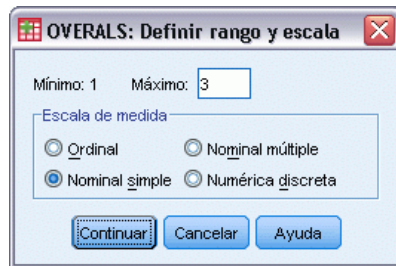
- ▶ Seleccione *Música preferida* y *Vecindario preferido* como variables para el tercer conjunto.
- ▶ Seleccione *música* y pulse en *Definir rango y escala*.

Figura 11-9
Cuadro de diálogo *Definir rango y escala*



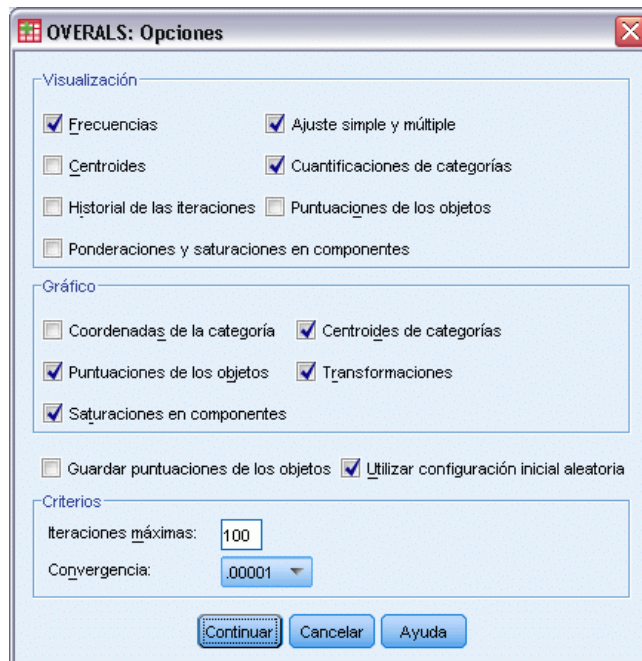
- ▶ Escriba 5 como valor máximo para esta variable.
- ▶ Seleccione como escala de la medida *Nominal simple*.
- ▶ Pulse en *Continuar*.
- ▶ En el cuadro de diálogo *Análisis de correlación canónica no lineal*, seleccione *vivir* y pulse en *Definir rango y escala*.

Figura 11-10
Cuadro de diálogo Definir rango y escala



- ▶ Escriba 3 como valor máximo para esta variable.
- ▶ Seleccione como escala de la medida Nominal simple.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, pulse en Opciones.

Figura 11-11
Cuadro de diálogo Opciones



- ▶ Anule la selección Centroides y seleccione Ponderaciones y saturaciones en componentes en el grupo Mostrar.
- ▶ Seleccione Centroides de categorías y Transformaciones en el grupo Gráfico.
- ▶ Seleccione Usar configuración inicial aleatoria.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, pulse en Aceptar.

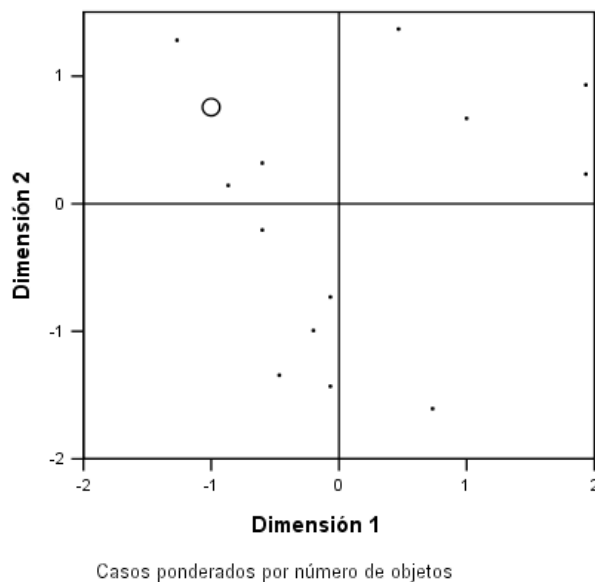
Después de generar una lista de variables con sus niveles de escalamiento óptimo, el análisis de correlación canónica categórico con escalamiento óptimo genera una tabla que muestra las frecuencias de objetos en las categorías. Esta tabla es especialmente importante si hay datos perdidos, ya que es más probable que las categorías casi vacías dominen la solución. En este ejemplo, no hay datos perdidos.

Se realiza una segunda comprobación para examinar el gráfico de puntuaciones de objeto para los valores atípicos. Los valores atípicos tienen una cantidad de cuantificaciones tan diferentes del resto de objetos que se encontrarán en los límites del gráfico, por lo que van a dominar una o varias dimensiones.

Si se encuentran valores atípicos, es posible gestionarlos de dos formas diferentes. Puede simplemente eliminarlos de los datos y ejecutar el análisis de correlación canónica no lineal de nuevo. También puede intentar recodificar las respuestas extremas de los objetos con valores atípicos si agrupa (fusiona) algunas categorías.

Como se muestra en el gráfico de puntuaciones de los objetos, no había valores atípicos para los datos de la encuesta.

Figura 11-12
Puntuaciones de los objetos



Explicación de la similitud entre los conjuntos

Hay varias formas de medir la asociación entre los conjuntos de un análisis de correlación canónica no lineal (cada uno de los cuales se detalla en una tabla diferente o conjunto de tablas).

Resumen de análisis

Los valores ajustados y perdidos indican la eficacia del ajuste de la solución del análisis de correlación canónica no lineal a los datos óptimos cuantificados con respecto a la asociación entre los conjuntos. El resumen de la tabla de análisis muestra el valor ajustado, los valores perdidos y los autovalores para el ejemplo de la encuesta.

Figura 11-13
Resumen del análisis

		Dimensión		Suma
		1	2	
Pérdida	Conjunto 1	,240	,183	,423
	Conjunto 2	,184	,408	,593
	Conjunto 3	,171	,205	,376
	Media	,199	,265	,464
Autovalores		,801	,735	
Ajuste				1,536

Se divide la pérdida por las dimensiones y conjuntos. Para cada dimensión y conjunto, la pérdida representa la proporción de variación en las puntuaciones de objeto que no puede explicarse mediante la combinación ponderada de las variables del conjunto. La pérdida media es la Mediana etiquetada. En este ejemplo, la pérdida media sobre los conjuntos es de 0,464. Observe que hay más pérdida en la segunda dimensión que en la primera dimensión.

El autovalor para cada dimensión es igual a 1 menos la pérdida media para la dimensión. Dicho autovalor indica la parte de relación que se muestra mediante cada dimensión. Los autovalores se añaden hasta el ajuste total. Para los datos de Verdegaal, $0,801/1,536 = 52\%$ del ajuste real se explica mediante la primera dimensión.

El máximo valor ajustado es igual al número de dimensiones y, si se logra, indica que la relación es perfecta. El valor de pérdida medio sobre los conjuntos y las dimensiones le indica la diferencia entre el ajuste máximo y el real. El ajuste más la pérdida media es igual al número de dimensiones. Es extraño que aparezca una perfecta similitud y normalmente saca provecho a los aspectos triviales en los datos.

Otro estadístico popular con dos conjuntos de variables es la correlación canónica. Ya que la correlación canónica está relacionada con el autovalor y no proporciona información adicional, no está incluida en el resultado del análisis de correlación canónica no lineal. Para dos conjuntos de variables, la correlación canónica por dimensión se obtiene mediante la siguiente fórmula:

$$\rho_d = 2 \times E_d - 1$$

donde d es el número de dimensión y E es el autovalor.

Puede generalizar la correlación canónica para más de dos conjuntos con la siguiente fórmula:

$$\rho_d = ((K \times E_d) - 1)/(K - 1)$$

donde d es el número de dimensión, K es el número de conjuntos y E es el autovalor. Para nuestro ejemplo:

$$\rho_1 = ((3 \times 0.801) - 1)/2 = 0.702$$

y

$$\rho_2 = ((3 \times 0.735) - 1)/2 = 0.603$$

Ponderaciones y saturaciones en componentes

Otra medida de asociación es la correlación múltiple entre las combinaciones lineales de cada conjunto y puntuaciones de los objetos. Si no hay variables nominales múltiples en un conjunto, puede calcular esta medida multiplicando la ponderación y la saturación en componentes de cada variable dentro del conjunto, añadiendo estos productos y calculando la raíz cuadrada de la suma.

Figura 11-14
Ponderaciones

Conjunto	Dimensión	
	1	2
1 Edad en años	,680	,789
Estado civil	,296	-1,016
2 Periódico leído conost mayor frecuencia	-,845	-,361
3 Música preferida	,631	-,749
Vecindario preferido	-,484	-,780

Figura 11-15
Saturaciones en componentes

Conjunto	Dimensión	
	1	2
1 Edad en años ^{a,b}	.834	.259
Estado civil ^{c,b}	.651	-.604
2 Mascotas que posee ^{d,e}	Dimensión 1	.397
	2	-.277
Periódico leído con mayor frecuencia ^{e,b}	-.667	-.391
3 Música preferida ^{c,b}	.786	-.500
Vecindario preferido ^{c,b}	-.687	-.540

- a. Nivel de escalamiento óptimo: Ordinal
- b. Proyecciones de las variables cuantificadas simples en el espacio de los objetos
- c. Nivel de escalamiento óptimo: Nominal simple
- d. Nivel de escalamiento óptimo: Nominal múltiple
- e. Proyecciones de las variables cuantificadas múltiple en el espacio de los objetos

Estas cifras dan las ponderaciones y saturaciones en componentes para las variables de este ejemplo. La correlación múltiple (R) es la siguiente para la primera suma ponderada de variables escaladas de manera óptima (*Edad en años* y *Estado civil*) con la primera dimensión de las puntuaciones de los objetos:

$$\begin{aligned} R &= \sqrt{(0.701 \times 0.841 + (-0.273 \times -0.631))} \\ &= \sqrt{(0.5895 + 0.1723)} \\ &= 0.873 \end{aligned}$$

Para cada dimensión, $1 - \text{pérdida} = R^2$. Por ejemplo, del resumen de la tabla de análisis, $1 - 0,238 = 0,762$, que es $0,873$ cuadrado (más el probable error de redondeo). En consecuencia, los pequeños valores de pérdida indican grandes correlaciones múltiples entre sumas ponderadas de variables escaladas de manera óptima y las dimensiones. Las ponderaciones no son únicas para las variables nominales múltiples. Para las variables nominales múltiples, utilice $1 - \text{pérdida}$ por conjunto.

Partición del ajuste y la pérdida

La pérdida de cada conjunto se divide por el análisis de correlación canónica no lineal de muchas formas. La tabla de ajuste presenta el ajuste múltiple, el ajuste simple y las tablas de pérdida simple generadas mediante el análisis de correlación canónica no lineal para el ejemplo de la encuesta. Observe que el ajuste múltiple menos el ajuste simple es igual a la pérdida simple.

Figura 11-16
Partición del ajuste y la pérdida

Conjunto	Ajuste múltiple			Ajuste simple			Pérdida simple		
	Dimensión		Suma	Dimensión		Suma	Dimensión		Suma
	1	2		1	2		1	2	
1 Edad en años ^a	,494	,676	1,170	,462	,622	1,085	,032	,054	,085
Estado civil ^b	,089	1,033	1,122	,088	1,033	1,120	,001	,000	,001
2 Mascotas que posee ^c	,402	,439	,841						
Periódico leído con mayor frecuencia ^b	,724	,187	,911	,714	,130	,844	,010	,057	,067
3 Música preferida ^b	,421	,577	,998	,398	,561	,960	,022	,016	,039
Vecindario preferido ^b	,234	,609	,843	,234	,608	,843	,000	,000	,000

a. Nivel de escalamiento óptimo: Ordinal

b. Nivel de escalamiento óptimo: Nominal simple

c. Nivel de escalamiento óptimo: Nominal múltiple

La pérdida simple indica la pérdida que se obtiene de restringir las variables a un conjunto de cuantificaciones (es decir, nominal simple, ordinal o nominal). Si una pérdida simple es grande, es mejor tratar las variables como nominales múltiples. Sin embargo, en este ejemplo, el ajuste simple y el múltiple son casi iguales, lo que significa que las coordenadas múltiples están casi en una línea recta en la dirección marcada por las ponderaciones.

El ajuste múltiple es igual a la varianza de las coordenadas de la categoría múltiple para cada variable. Estas medidas son análogas a las medidas de discriminación que se encuentran en el análisis de homogeneidad. Es posible examinar la tabla de ajuste múltiple para ver las variables que discriminan mejor. Por ejemplo, observe la tabla de ajuste múltiple para *Estado civil* y *Periódico leído con mayor frecuencia*. Los valores ajustados, sumados a lo largo de las

dos dimensiones, son 1,122 para *Estado civil* y 0,911 para *Periódico leído con mayor frecuencia*. Esta información nos indica que el estado civil de una persona proporciona mayor fuerza discriminatoria que el periódico al que está suscrita.

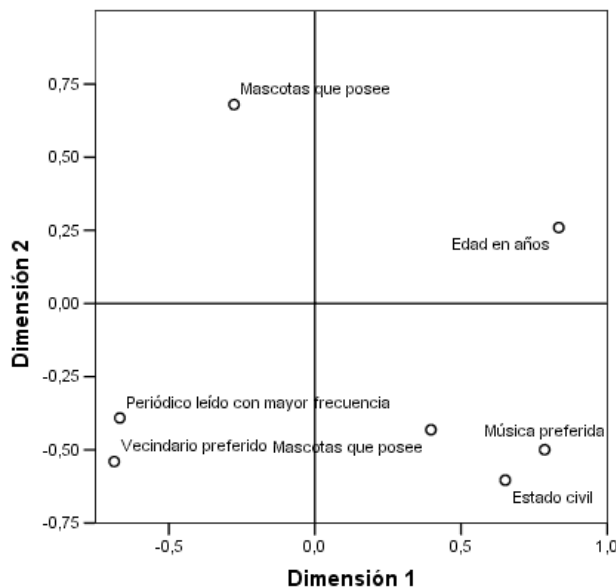
El ajuste simple se corresponde con la ponderación cuadrada para cada variable y es igual a la varianza de las coordenadas de las categorías simples. Como resultado, las ponderaciones son igual a las desviaciones típicas de las coordenadas de las categorías simples. Si se examina cómo se desglosa el ajuste simple a lo largo de las dimensiones, observamos que la variable *Periódico leído con mayor frecuencia* discrimina mayoritariamente en la primera dimensión y también vemos que la variable *Estado civil* discrimina casi totalmente en la segunda dimensión. Dicho de otro modo, las categorías de *Periódico leído con mayor frecuencia* se encuentran más alejadas en la primera dimensión que en la segunda, mientras que el patrón se invierte para *Estado civil*. Por el contrario, *Edad en años* discrimina tanto en la primera como en la segunda dimensión, por lo que la dispersión de las categorías es igual a lo largo de las dos dimensiones.

Saturaciones en componentes

La siguiente figura muestra el gráfico de las saturaciones en componentes para los datos obtenidos en la encuesta. Si no existen datos perdidos, las saturaciones en componentes equivalen a las correlaciones de Pearson entre las variables cuantificadas y las puntuaciones de los objetos.

La distancia desde el origen a cada punto de variable se aproxima a la importancia de esa variable. Las variables canónicas no están representadas, pero pueden representarse mediante líneas horizontales y verticales trazadas por el origen.

Figura 11-17
Saturaciones en componentes



Las relaciones entre las variables son aparentes. Hay dos direcciones que no coinciden con los ejes horizontales y verticales. Una dirección se determina mediante *Edad en años*, *Periódico leído con mayor frecuencia* y *Vecindario preferido*. La otra dirección se define mediante las variables

Estado civil, Música preferida y Mascotas que posee. La variable *Mascotas que posee* es una variable nominal múltiple, por lo que hay dos puntos representados para ella. Cada cuantificación se interpreta como una variable simple.

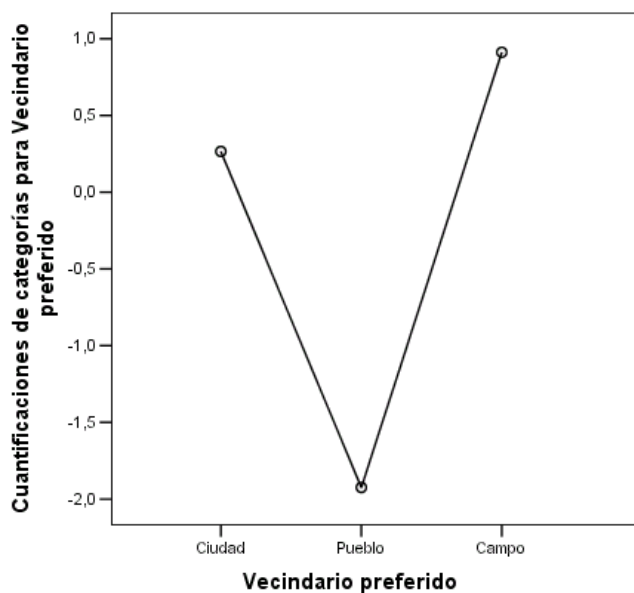
Gráficos de transformación

Los diferentes niveles en los que se puede escalar cada variable imponen restricciones sobre las cuantificaciones. Los gráficos de transformación ilustran la relación entre las cuantificaciones y las categorías originales que resultan del nivel de escalamiento óptimo seleccionado.

El gráfico de transformación para *Vecindario preferido*, que se había tratado como nominal, muestra un patrón en forma de U, en el que la categoría intermedia recibe la cuantificación menor y las categorías extremas reciben valores que son similares entre sí. Este patrón indica una relación cuadrática entre la variable original y la transformada. No se sugiere utilizar un nivel de escalamiento óptimo alternativo para *Vecindario preferido*.

Figura 11-18

Gráfico de transformación para la variable *Situación preferida de la vivienda (nominal)*



Las cuantificaciones para *Periódico leído con mayor frecuencia*, por el contrario, corresponden con una tendencia ascendente a lo largo de las tres categorías que tienen casos observados. La primera categoría recibe la menor cuantificación, la segunda recibe un mayor valor y la tercera

recibe el mayor de todos. A pesar de que la variable se escala como nominal, el orden de la categoría se recupera en las cuantificaciones.

Figura 11-19

El gráfico de transformación para Periódico leído con mayor frecuencia (nominal)

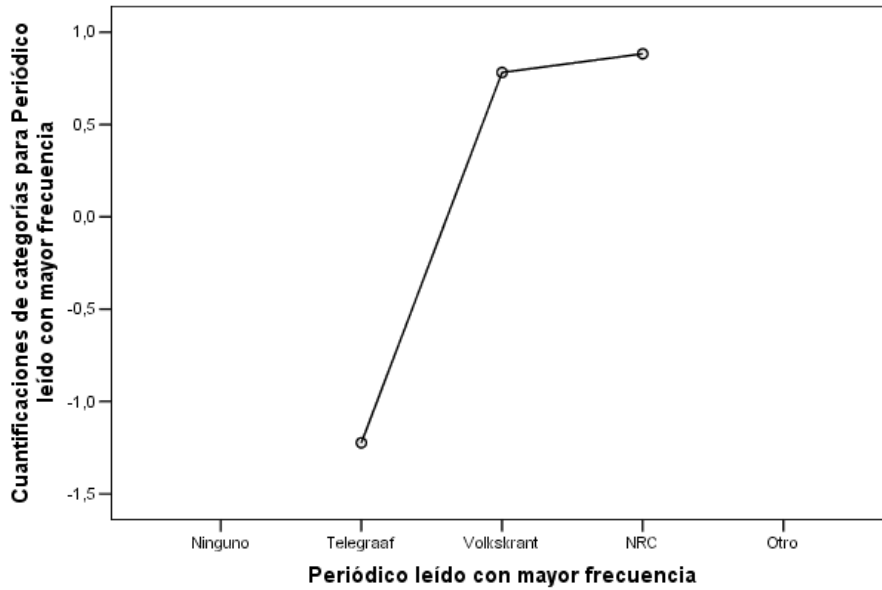
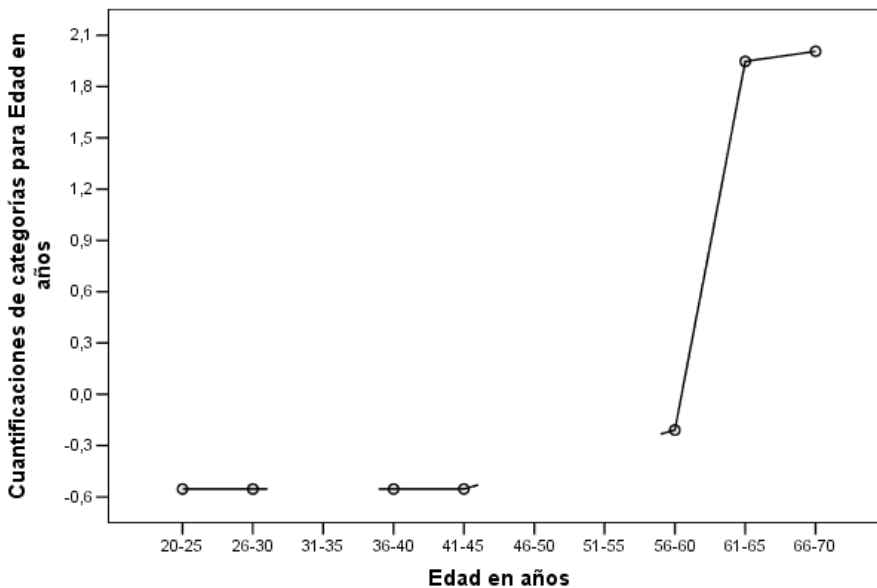


Figura 11-20

Gráfico de transformación para Edad en años (ordinal)



El gráfico de transformación para *Edad en años* muestra una curva en forma de S. Las cuatro categorías más jóvenes observadas reciben la misma cuantificación negativa, mientras que las dos de mayor edad reciben iguales valores positivos. Por consiguiente, puede intentar agrupar las edades más jóvenes en una categoría común (es decir, menos de 50) y agrupar las categorías de

mayor edad en otra categoría. Sin embargo, la exacta igualdad de las cuantificaciones para los grupos más jóvenes indica que puede no ser adecuado restringir el orden de las cuantificaciones al orden de las categorías originales. Ya que las cuantificaciones para los grupos 26–30, 36–40 y 41–45 no pueden ser menores que la cuantificación para el grupo 20–25, se establecen estos valores como iguales al valor para el límite. Es posible que el ajuste se mejore si permite que estos valores sean más pequeños que la cuantificación para el grupo de menor edad (es decir, tratando la edad como nominal). Por lo tanto, aunque la edad puede considerarse una variable ordinal, tratarla como tal no parece apropiado en este caso. Además, si se trata la edad como numérica, y, con ello, se mantienen las distancias entre las categorías, se reducirá el ajuste de manera sustancial.

Coordenadas de categoría simples frente a categorías múltiples

Para cada variable tratada como nominal simple, ordinal o numérica, se determinan las cuantificaciones, las coordenadas de las categorías simples y las coordenadas de categoría múltiples. Se presentan estos estadísticos para *Edad en años*.

Figura 11-21
Coordenadas para *Edad en años*

	Frecuencia marginal	Cuantificación	Coordenadas de categorías simples		Coordenadas de categorías múltiples	
			Dimensión		Dimensión	
			1	2	1	2
20-25	3	-,554	-,377	-,437	-,192	-,139
26-30	5	-,554	-,377	-,437	-,404	-,623
31-35	0	,000				
36-40	1	-,554	-,377	-,437	-,318	-,733
41-45	1	-,554	-,377	-,437	-,356	-,534
46-50	0	,000				
51-55	0	,000				
56-60	2	-,209	-,142	-,165	-,435	,087
61-65	1	1,947	1,324	1,536	1,710	1,204
66-70	2	2,006	1,364	1,583	1,215	1,711
Perdidos	0					

a. Nivel de escalamiento óptimo: Ordinal

Cada categoría para la que no se registraron casos recibe una cuantificación de 0. Para *Edad en años*, ésta incluye las categorías 31–35, 46–50 y 51–55. No se restringen estas categorías para que se ordenen con el resto de categorías y no afectan a los cálculos.

Para las variables nominales múltiples, cada categoría recibe una cuantificación diferente en cada dimensión. Para el resto de tipos de transformación, una categoría tiene sólo una cuantificación, sin importar la dimensionalidad de la solución. Cada conjunto de coordenadas de una categoría simple representa la ubicación de dicha categoría en una línea del espacio del objeto. Las coordenadas de una determinada categoría corresponden a la cuantificación multiplicada por las ponderaciones de dimensión de la variable. Por ejemplo, en la tabla para *Edad en años*, las coordenadas de las categorías simples para la categoría 56-60 (-0,142, -0,165) son la cuantificación (-0,209) multiplicada por las ponderaciones de dimensión (0,680, 0,789).

Las coordenadas de las categorías múltiples para las variables que se tratan como nominales simples, ordinales o numéricas representan las coordenadas de las categorías en el espacio del objeto antes de aplicar las restricciones ordinales o lineales. Estos valores son minimizadores sin

restringir de la pérdida. Para las variables nominales múltiples, estas coordenadas representan las cuantificaciones de las categorías.

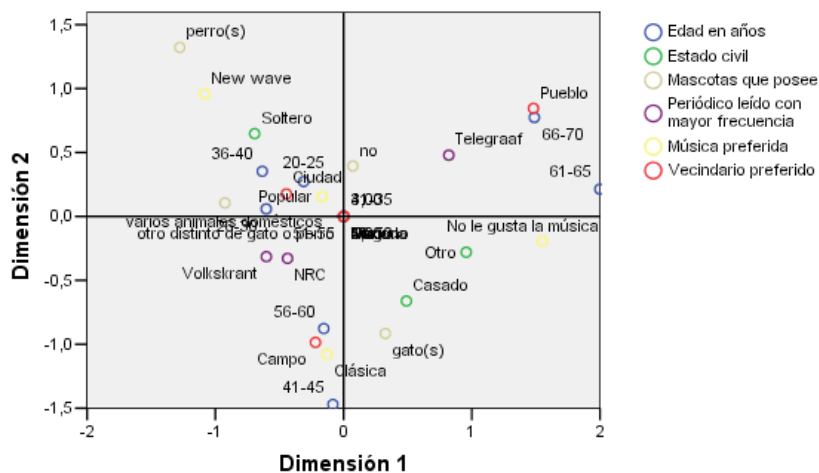
Se revelan los efectos de imponer restricciones en la relación entre las categorías y sus cuantificaciones mediante la comparación de las coordenadas de las categorías simples con las coordenadas de las categorías múltiples. En la primera dimensión, las coordenadas de las categorías múltiples para *Edad en años* disminuyen hasta la categoría 2 y permanecen relativamente en el mismo nivel hasta que la categoría 9, punto en el que se produce un incremento drástico. Se muestra un patrón similar para la segunda dimensión. Se eliminan estas relaciones de las coordenadas de las categorías simples, en las que se aplica la restricción ordinal. En ambas dimensiones, las coordenadas son ahora no decrecientes. La estructura diferente de los dos conjuntos de coordenadas sugiere que puede ser más adecuado un tratamiento nominal.

Centroides y centroides proyectados

El gráfico de los centroides etiquetados mediante las variables deberían interpretarse de la misma forma que el gráfico de las cuantificaciones de las categorías en el análisis de homogeneidad o coordenadas de las categorías múltiples en el análisis de componentes principales no lineal. Por si mismo, dicho gráfico muestra la eficacia con la que las variables separan grupos de objetos (los centroides están en el centro de gravedad de los objetos).

Observe que las categorías para *Edad en años* no están separadas de manera muy clara. Las categorías de menor edad se agrupan juntas en la parte izquierda del gráfico. Como se ha sugerido con anterioridad, el ordinal puede ser un nivel de escalamiento demasiado estricto para imponerse a *Edad en años*.

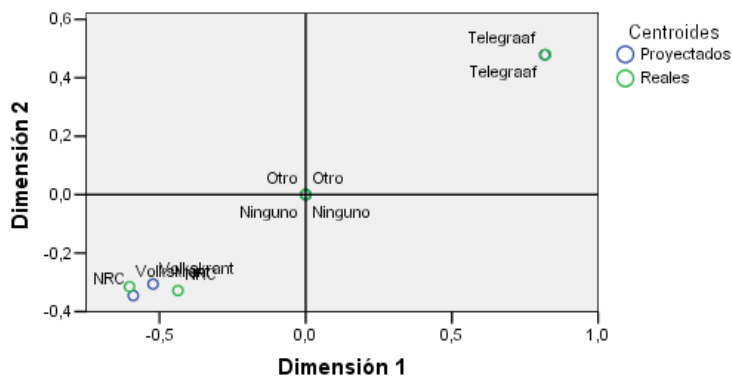
Figura 11-22
Centroides etiquetados por las variables



Al solicitar los gráficos de los centroides, también se generan el centroide individual y los gráficos de centroides proyectados para cada variable que se etiqueta mediante las etiquetas de valor. Los centroides proyectados se encuentran sobre una línea en el espacio del objeto.

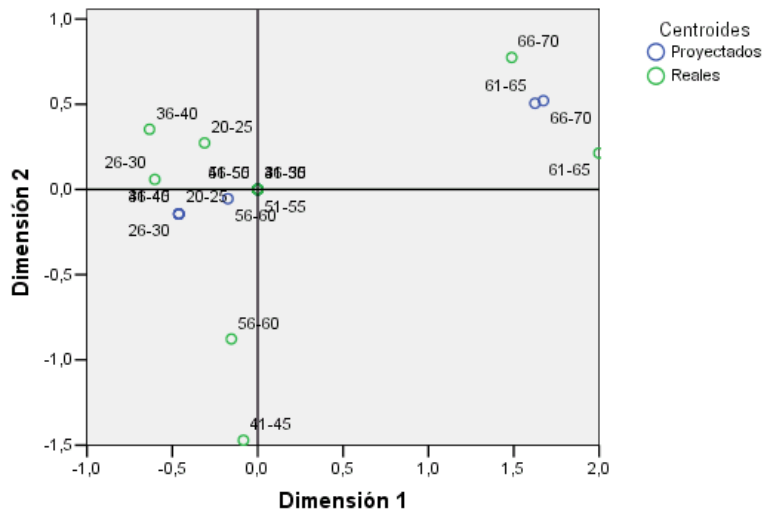
Figura 11-23

Centroides y centroides proyectados para Periódico que lee generalmente



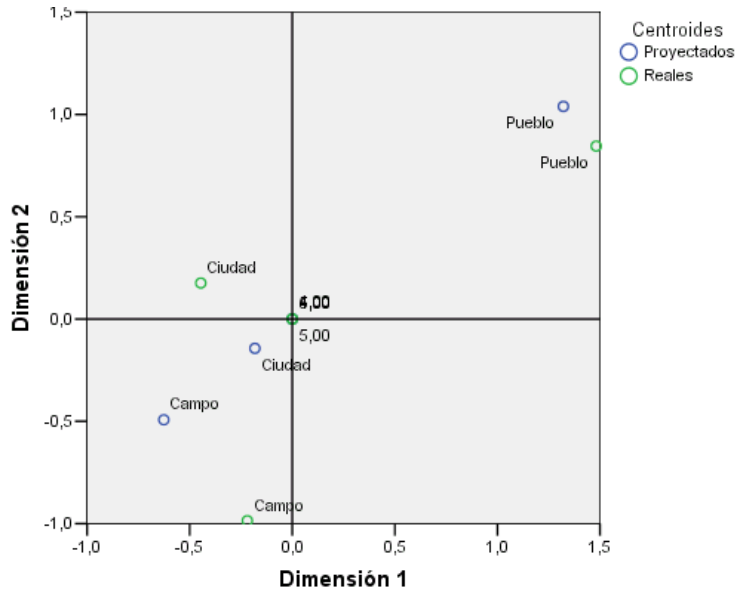
Los centroides reales se proyectan en los vectores que se definen mediante las saturaciones en componentes. Estos vectores se han añadido a los gráficos de los centroides para ayudar a distinguir los centroides proyectados de los centroides reales. Los centroides proyectados se encuentran dentro de uno de cuatro cuadrantes formados por la extensión de dos líneas de referencia perpendiculares que pasan por el origen. Se obtiene la interpretación de la dirección de las variables nominales simples, ordinales o numéricas desde la posición de los centroides proyectados. Por ejemplo, la variable *Periódico leído con mayor frecuencia* se especifica como una variable nominal simple. Los centroides proyectados muestran que *Volkskrant* y *NRC* se contrastan con *Telegraaf*.

Figura 11-24
 Los centroides y los centroides proyectados para la variable *Edad en años*



El problema con *Edad en años* es evidente de los centroides proyectados. Tratar *Edad en años* como ordinal implica que ha de conservarse el orden de los grupos de edad. Para satisfacer esta restricción, se proyectan todos los grupos de edad por debajo de 45 en el mismo punto. A lo largo de la dirección definida mediante *Edad en años*, *Periódico leído con mayor frecuencia* y *Vecindario preferido*, no hay separación de los grupos de edad menores. Este descubrimiento sugiere tratar la variable como nominal.

Figura 11-25
 Los centroides y los centroides proyectados para *Situación preferida de la vivienda*



Para comprender las relaciones entre las variables, averigüe lo que son las categorías específicas (valores) para los conglomerados de categorías en los gráficos de los centroides. Las relaciones entre *Edad en años*, *Periódico leído con mayor frecuencia* y *Vecindario preferido* pueden

describirse si miramos la parte superior derecha e inferior izquierda de los gráficos. En la parte superior derecha, los grupos de edad son los encuestados de mayor edad que leen el Telegraaf y prefieren vivir en un pueblo. Si observa la esquina inferior izquierda de cada gráfico, verá que los encuestados jóvenes y personas de mediana edad leen Volkskrant o NRC y quieren vivir en el campo o en la ciudad. Sin embargo, separar los grupos más jóvenes es muy difícil.

Pueden realizarse los mismos tipos de interpretaciones sobre otra dirección (*Música preferida*, *Estado civil* y *Mascotas que posee*) si nos centramos en la parte superior izquierda y la parte inferior derecha de los gráficos de los centroides. En la esquina superior izquierda, observamos que los solteros tienen perros y les gusta la música moderna. Los casados y otras categorías para ecivil tienen gatos; el grupo anterior prefiere la música clásica y al último grupo no le gusta la música.

Un análisis alternativo

Los resultados del análisis sugieren que tratar *Edad en años* como ordinal no parece adecuado. Aunque *Edad en años* se mide en un nivel ordinal, sus relaciones con otras variables no son monótonicas. Para investigar los efectos de cambiar el nivel de escalamiento óptimo por nominal simple, puede ejecutar de nuevo el análisis.

Para ejecutar el análisis

- ▶ Vuelva al cuadro de diálogo Análisis de correlación canónica no lineal y vaya al primer conjunto.
- ▶ Seleccione *edad* y pulse en Definir rango y escala.
- ▶ En el cuadro de diálogo Definir rango y escala, seleccione Nominal simple como rango del escalamiento.
- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Análisis de correlación canónica no lineal, pulse en Aceptar.

Los autovalores para una solución bidimensional son 0,806 y 0,757, respectivamente, con un ajuste total de 1,564.

Figura 11-26

Los autovalores para la solución bidimensional

		Dimensión		Suma
		1	2	
Pérdida	Conjunto 1	,249	,115	,363
	Conjunto 2	,176	,408	,584
	Conjunto 3	,157	,205	,363
	Media	,194	,243	,436
Autovalores		,806	,757	
Ajuste				1,564

Las tablas de ajuste múltiple y ajuste simple muestran que *Edad en años* sigue siendo una variable de gran potencia discriminadora, como se evidencia mediante la suma de los valores ajustados múltiple. Sin embargo, en oposición a los resultados anteriores, un examen de los

valores ajustados simple revela que la discriminación se encuentra casi enteramente a lo largo de la segunda dimensión.

Figura 11-27
Partición del ajuste y la pérdida

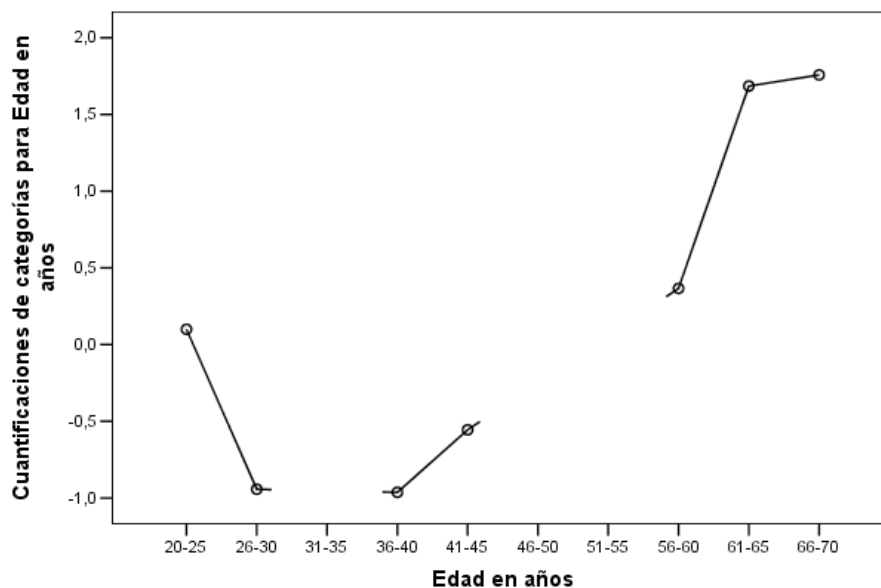
Conjunto	Ajuste múltiple			Ajuste simple			Pérdida simple		
	Dimensión		Suma	Dimensión		Suma	Dimensión		Suma
	1	2		1	2		1	2	
1 Edad en años ^a	,246	1,197	1,443	,195	1,188	1,384	,051	,008	,059
Estado civil ^a	,273	1,136	1,409	,272	1,135	1,407	,001	,000	,002
2 Mascotas que posee ^b	,530	,392	,921						
Períodico leído con mayor frecuencia ^a	,639	,185	,824	,631	,149	,780	,008	,036	,044
3 Música preferida ^a	,604	,438	1,041	,603	,437	1,040	,000	,001	,001
Vecindario preferido ^a	,075	,822	,897	,075	,822	,897	,000	,000	,000

a. Nivel de escalamiento óptimo: Nominal simple

b. Nivel de escalamiento óptimo: Nominal múltiple

Convierta el gráfico de transformación para *Edad en años*. Las cuantificaciones para una variable nominal no están restringidas, por lo que la tendencia no decreciente que se ha mostrado cuando *Edad en años* se ha tratado de forma ordinal ya no está presente. Hay una tendencia decreciente hasta la edad de 40 y una ascendente de ahí en adelante, que se corresponde con una relación en forma de U (cuadrática). Las dos categorías de mayor edad siguen recibiendo puntuaciones similares, y el análisis siguiente puede implicar combinar estas categorías.

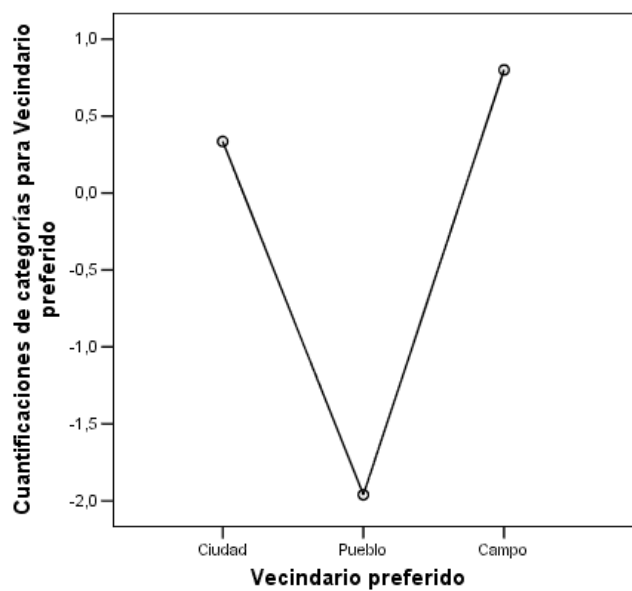
Figura 11-28
Gráfico de transformación para *Edad en años* (nominal)



El gráfico de transformación para *Vecindario preferido* se muestra aquí. Tratar *Edad en años* como nominal no afecta a las cuantificaciones para *Vecindario preferido* para cualquier grado significativo. La categoría intermedia recibe la cuantificación más pequeña, mientras que las categorías extremas reciben valores positivos grandes.

Figura 11-29

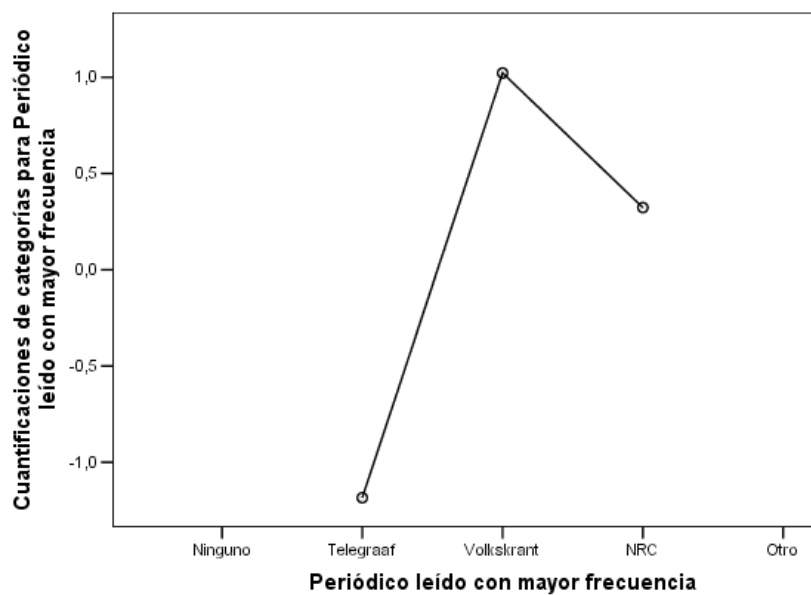
Gráfico de transformación para la variable *Vecindario preferido* (edad nominal)



Se encontró un cambio en el gráfico de transformación para *Periódico leído con mayor frecuencia*. Anteriormente, una tendencia ascendente estaba presente en las cuantificaciones, lo que posiblemente sugiera un tratamiento ordinal para esta variable. Sin embargo, tratar *Edad en años* como nominal elimina esta tendencia de las cuantificaciones de noticia.

Figura 11-30

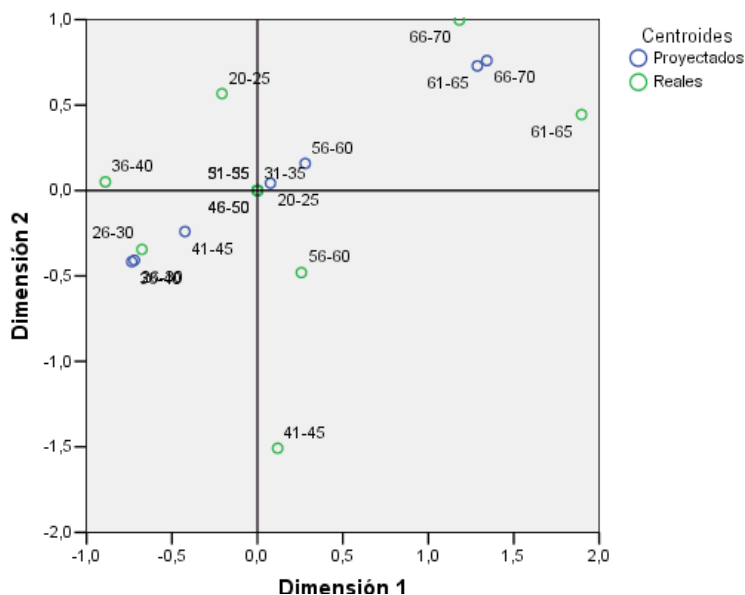
El gráfico de transformación para *Periódico que lee generalmente* (edad nominal)



Este gráfico es el gráfico del centroide para *Edad en años*. Observe que las categorías no se disponen en orden cronológico a lo largo de la línea que une los centroides proyectados. El grupo 20–25 está situado en medio en lugar de al final. La dispersión de las categorías mejora de manera notable en comparación con el ordinal correspondiente que se presentó anteriormente.

Figura 11-31

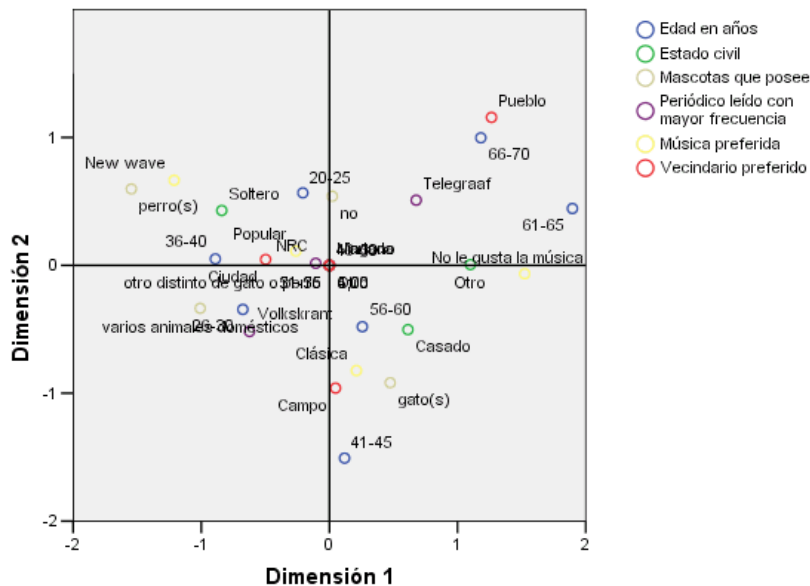
Los centroides y los centroides proyectados para la variable *Edad en años (nominal)*



La interpretación de los grupos de menor de edad es actualmente posible gracias al gráfico del centroide. Las categorías *Volkskrant* y *NRC* se encuentran también más alejadas que en el análisis anterior, lo que permite interpretaciones diferentes de cada uno de ellos. Los grupos entre las edades de 26 y 45 leen el *Volkskrant* y prefieren vivir en el campo. Los grupos de edad de 20–25 y 56–60 leen el *NRC*. El primer grupo prefiere vivir en la ciudad, mientras que el último prefiere vivir en el campo. Los grupos de mayor edad leen el *Telegraaf* y prefieren vivir en un pueblo.

La interpretación de la otra dirección (*Música preferida, Estado civil y Mascotas que posee*) no ha cambiado sustancialmente con respecto al análisis anterior. La única diferencia obvia es que la gente con un estado civil de los incluidos en el grupo *Otro* tienen gatos o no poseen mascotas en absoluto.

Figura 11-32
Centroides etiquetados por las variables (edad nominal)



Sugerencias generales

Tras examinar los resultados iniciales, es posible que desee ajustar el análisis cambiando algunas especificaciones para el análisis de correlación canónica no lineal. A continuación mostramos algunos consejos para estructurar el análisis:

- Cree tantos conjuntos como sea posible. Coloque una variable importante que desee predecir en un conjunto diferente que contenga únicamente dicha variable.
- Ponga las variables que considere predictores juntas en un único conjunto. Si hay muchos predictores, intente dividirlos en varios conjuntos.
- Coloque la variable nominal múltiple en un conjunto diferente que la contenga sólo a ella.
- Si las variables están muy correlacionadas entre sí y no desea que esta relación domine la solución, coloque dichas variables juntas en el mismo conjunto.

Lecturas recomendadas

Si desea obtener más información sobre el análisis de correlación canónica no lineal:

Carroll, J. D. 1968. Generalization of canonical correlation analysis to three or more sets of variables. En: *Proceedings of the 76th Annual Convention of the American Psychological Association*, 3, Washington, D.C.: American Psychological Association.

De Leeuw, J. 1984. *Canonical analysis of categorical data*, 2nd ed. Leiden: DSWO Press.

Horst, P. 1961. Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology*, 17, .

Horst, P. 1961. Relations among m sets of measures. *Psychometrika*, 26, .

Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58, .

Van der Burg, E. 1988. *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.

Van der Burg, E., y J. De Leeuw. 1983. Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, 36, .

Van der Burg, E., J. De Leeuw, y R. Verdegaal. 1988. Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, .

Verboon, P., y I. A. Van der Lans. 1994. Robust canonical discriminant analysis. *Psychometrika*, 59, .

Análisis de correspondencias

Correspondence analysis

Una **tabla de correspondencias** es una tabla de doble clasificación cuyas casillas contienen alguna medida de correspondencia entre las filas y las columnas. La medida de correspondencia puede ser cualquier indicación de la similaridad, afinidad, confusión, asociación o interacción entre las variables de fila y de columna. Un tipo muy habitual de tabla de correspondencias es una tabla de contingencia, en la que las casillas contienen las frecuencias.

Estas tablas se pueden obtener con facilidad mediante el procedimiento Tablas de contingencia. Sin embargo, una tabla de contingencia no proporciona siempre una imagen clara de la naturaleza de la relación entre las dos variables. Así ocurre especialmente si las variables de interés son nominales (sin ningún orden o rango inherente) y contienen numerosas categorías. Las tablas de contingencia pueden indicarle que las frecuencias observadas de las casillas difieren considerablemente de los valores esperados en una tabla de contingencia de 10×9 de *ocupación y desayuna cereales*, pero puede ser difícil determinar qué grupos de ocupaciones tienen gustos similares o cuáles son estos gustos.

El análisis de correspondencias permite examinar la relación entre dos variables nominales de manera gráfica en un espacio multidimensional. Se calculan las puntuaciones de fila y de columna y se generan los gráficos basados en las puntuaciones. Las categorías que son similares entre sí aparecen juntas en los gráficos. De esta manera, es fácil ver las categorías de una variable que son similares entre sí o las categorías de las dos variables que están relacionadas. El procedimiento Análisis de correspondencias también permite ajustar puntos suplementarios en el espacio definido por los puntos activos.

Si el orden de las categorías de acuerdo con sus puntuaciones no es deseable o se opone a la intuición, se pueden imponer restricciones de orden imponiendo que sean iguales las puntuaciones de algunas categorías. Por ejemplo, supongamos que espera que la variable *consumo de tabaco* cuyas categorías son *ninguno, bajo, medio y alto* tengan puntuaciones que correspondan a este orden. Sin embargo, si el análisis ordena las categorías como *ninguno, bajo, alto y medio*, si obliga a que las puntuaciones de *alto y medio* sean iguales, se conserva el orden de las categorías en sus puntuaciones.

La interpretación del análisis de correspondencias en términos de distancias depende del método de normalización utilizado. El procedimiento Análisis de correspondencias se puede utilizar para analizar tanto las diferencias entre las categorías de una variable como las diferencias entre las variables. Con la normalización por defecto, se analizan las diferencias entre las variables de fila y de columna.

El algoritmo de análisis de correspondencias puede realizar muchos tipos de análisis. El centrado de las filas y las columnas y el uso de distancias chi-cuadrado corresponden al análisis de correspondencias típico. Sin embargo, el uso de las opciones de centrado alternativo combinado con las distancias euclídeas permite obtener una representación alternativa de una matriz en un espacio de pocas dimensiones.

A continuación, veremos tres ejemplos. El primero utiliza una tabla de correspondencias relativamente pequeña para ilustrar los conceptos inherentes al análisis de correspondencias. El segundo ejemplo muestra una aplicación práctica de marketing. El último ejemplo utiliza una tabla de distancias en una aproximación con escalamiento multidimensional.

Normalización

La normalización se utiliza para distribuir la inercia sobre las puntuaciones de fila y de columna. Algunos aspectos de la solución de análisis de correspondencias, como los valores propios, la inercia por dimensión y las contribuciones, no cambian con las diferentes normalizaciones. Las puntuaciones de fila y de columna y sus varianzas si se ven afectadas. El análisis de correspondencias tiene varias maneras de distribuir la inercia. Las tres más habituales incluyen la distribución de la inercia únicamente sobre las puntuaciones de fila, la distribución de la inercia únicamente sobre las puntuaciones de columna y la distribución de la inercia simétricamente sobre las puntuaciones de fila como de columna.

Principal por fila. En la normalización principal por fila, las distancias euclídeas entre los puntos de fila aproximan las distancias chi-cuadrado entre las filas de la tabla de correspondencias. Las puntuaciones de fila son la media ponderada de las puntuaciones de columna. Las puntuaciones de columna se tipifican para tener una suma ponderada de los cuadrados de las distancias al centroide de 1. Como este método maximiza las distancias entre las categorías de fila, debe utilizar la normalización principal por fila si está interesado principalmente en ver cómo difieren entre sí las categorías de la variable de fila.

Principal por columna. Por otra parte, es posible que quiera aproximar las distancias chi-cuadrado entre las columnas de la tabla de correspondencias. En este caso, las puntuaciones de columna deben ser la media ponderada de las puntuaciones de fila. Las puntuaciones de fila se tipifican para tener una suma ponderada de cuadrados de las distancias al centroide de 1. Este método maximiza las distancias entre las categorías de columna y se debe utilizar si está interesado principalmente en ver cómo difieren entre sí las categorías de la variable de columna.

Simétrico. También puede tratar a las filas y las columnas de manera simétrica. Esta normalización distribuye la inercia de manera idéntica sobre las puntuaciones de fila y de columna. Observe que ni las distancias entre los puntos de fila ni las distancias entre los puntos de columna son aproximaciones de las distancias chi-cuadrado en este caso. Utilice este método si está interesado principalmente en las diferencias y las similitudes entre las dos variables. Normalmente, éste es el método preferido para hacer los diagramas de dispersión biespaciales.

Principal. Una cuarta opción se denomina normalización principal, en la que la inercia se distribuye dos veces sobre la solución, una vez sobre las puntuaciones de fila y una vez sobre las puntuaciones de columna. Debe utilizar este método si le interesan las distancias entre los puntos de fila y las distancias entre los puntos de columna por separado, pero no en cómo están relacionados entre sí los puntos de fila y de columna. Los diagramas de dispersión biespacial no son apropiados para esta opción de normalización y, por tanto, no están disponibles si se ha especificado el método de normalización principal.

Ejemplo: Percepciones de marcas de café

El ejemplo anterior implicaba una tabla pequeña de datos hipotéticos. Las aplicaciones reales con frecuencia implican tablas de tamaño mucho mayor. En este ejemplo, utilizaremos los datos pertenecientes a las imágenes percibidas de seis marcas de café helado (Kennedy, Riquier, y Sharp, 1996). Este conjunto de datos se puede encontrar en *coffee.sav*. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en *IBM SPSS Categories 20*.

Para cada uno de los 23 atributos de imagen de café helado, los encuestados seleccionaron todas las marcas que quedaban descritas por el atributo. Las seis marcas se denotan *AA*, *BB*, *CC*, *DD*, *EE* y *FF* para mantener la confidencialidad.

Tabla 12-1
Atributos de café helado

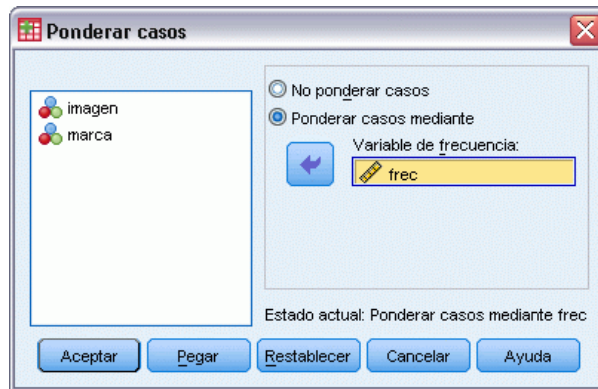
Atributo de imagen	Etiqueta	Atributo de imagen	Etiqueta
buen remedio para la resaca	<i>remedio</i>	marca que engorda	<i>engorda</i>
marca baja en calorías	<i>baja en calorías</i>	atractiva para los hombres	<i>hombres</i>
marca para niños	<i>children</i>	marca del sur de Australia	<i>sur de Australia</i>
marca para la clase trabajadora	<i>trabajadora</i>	marca tradicional/pasada de moda	<i>tradicional</i>
marca apetitosa/dulce	<i>dulce</i>	marca de primera calidad	<i>calidad</i>
marca poco popular	<i>poco popular</i>	marca sana	<i>sana</i>
marca para personas gordas/feas	<i>feas</i>	marca alta en cafeína	<i>cafeína</i>
gran frescor	<i>frescor</i>	marca nueva	<i>nuevo</i>
marca para yupis	<i>yupis</i>	marca para personas atractivas	<i>atractivas</i>
marca nutritiva	<i>nutritiva</i>	marca severa	<i>severa</i>
marca para mujeres	<i>mujeres</i>	marca popular	<i>popular</i>
marca menor	<i>menor</i>		

En principio, nos centraremos en cómo están relacionados los atributos entre sí y cómo están relacionadas las marcas entre sí. El uso de la normalización principal distribuye la inercia total una vez sobre las filas y una vez sobre las columnas. Aunque esto impide la interpretación del diagrama de dispersión biespacial, es posible examinar las distancias entre las categorías de cada variable.

Ejecución del análisis

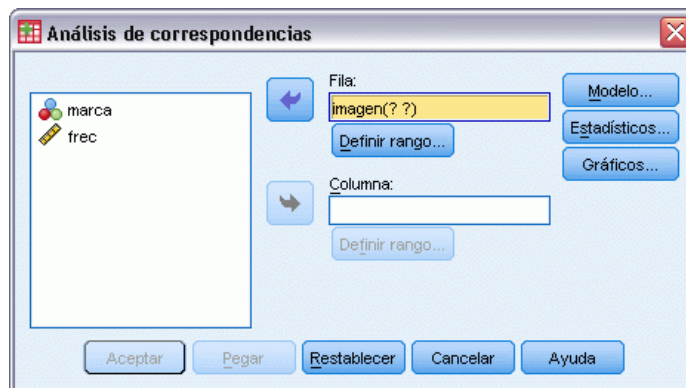
- La configuración de los datos requiere que los casos se ponderen mediante la variable *frec*. Para ello, elija en los menús:
Datos > Ponderar casos...

Figura 12-1
Cuadro de diálogo Ponderar casos



- ▶ Ponderar casos por *frec*.
- ▶ Pulse en Aceptar.
- ▶ Para obtener una solución inicial en cinco dimensiones con normalización principal, elija en los menús:
Analizar > Reducción de dimensiones > Análisis de correspondencias...

Figura 12-2
Cuadro de diálogo Análisis de correspondencias



- ▶ Seleccione *imagen* como variable de fila.
- ▶ Pulse en Definir rango.

Figura 12-3
Cuadro de diálogo Definir rango de filas

Análisis de correspondencias: Definir rango de filas

Rango de categorías para la variable de filas: imagen

Valor mínimo: 1

Valor máximo: 23

Actualizar

Restricciones para las categorías

1
2
3
4
5
6
7

Ninguna
 Las categorías deben ser iguales
 La categoría es suplementaria

Continuar Cancelar Ayuda

- ▶ Escriba 1 como valor mínimo.
- ▶ Escriba 23 como valor máximo.
- ▶ Pulse en Actualizar.
- ▶ Pulse en Continuar.
- ▶ Seleccione *marca* como variable de columna.
- ▶ Pulse en Definir rango en el cuadro de diálogo Análisis de correspondencias.

Figura 12-4
Cuadro de diálogo Definir rango de columnas

Análisis de correspondencias: Definir rango de columnas

Rango de categorías para la variable de columnas: marca

Valor mínimo: 1

Valor máximo: 6

Actualizar

Restricciones para las categorías

1
2
3
4
5
6

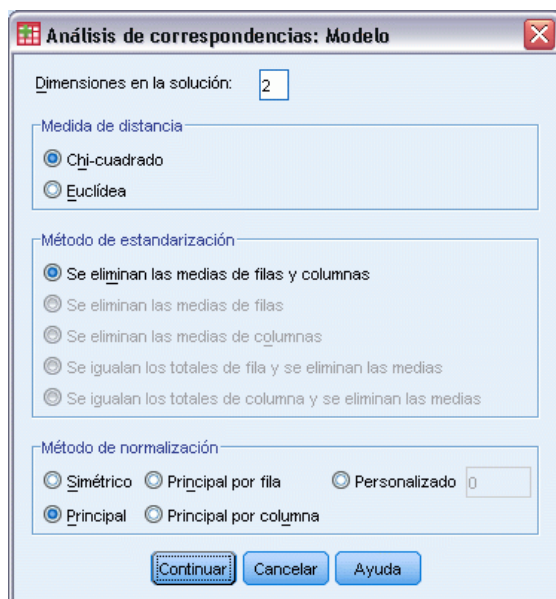
Ninguna
 Las categorías deben ser iguales
 La categoría es suplementaria

Continuar Cancelar Ayuda

- ▶ Escriba 1 como valor mínimo.
- ▶ Escriba 6 como valor máximo.

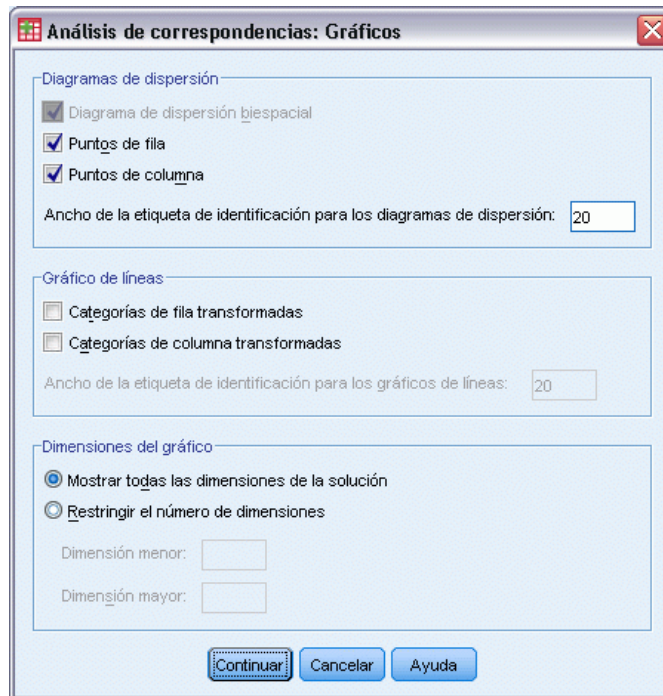
- ▶ Pulse en Actualizar.
- ▶ Pulse en Continuar.
- ▶ Pulse en Modelo en el cuadro de diálogo Análisis de correspondencias.

Figura 12-5
Cuadro de diálogo Modelo



- ▶ Seleccione Principal como método de normalización.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Análisis de correspondencias.

Figura 12-6
Cuadro de diálogo Gráficos



- ▶ Seleccione Puntos de fila y Puntos de columna en el grupo Diagramas de dispersión.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Análisis de correspondencias.

Dimensionalidad

La inercia por dimensión muestra la descomposición de la inercia total a lo largo de cada dimensión. Dos dimensiones explican el 83% de la inercia total. Si se añade una tercera dimensión sólo se añade un 8,6% a la inercia explicada. Por tanto, puede elegir utilizar una representación en dos dimensiones.

Figura 12-7
Inercia por dimensión

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación 2
1	,711	,506			,629	,629	,009	,132
2	,399	,159			,198	,827	,014	
3	,263	,069			,086	,913		
4	,234	,055			,068	,982		
5	,121	,015			,018	1,000		
Total		,804	3746,968	,000 ^a	1,000	1,000		

a. 110 grados de libertad

Contribuciones

Una visión general de los puntos de fila muestra las contribuciones de los puntos de fila a la inercia de las dimensiones y las contribuciones de las dimensiones a la inercia de los puntos de fila. Si todos los puntos contribuyen de igual manera a la inercia, las contribuciones serían 0,043. *Sana y baja en calorías* contribuyen en una parte importante a la inercia de la primera dimensión. *Hombres y severa* contribuyen con las mayores cantidades a la inercia de la segunda dimensión. *Feas y frescor* contribuyen muy poco a ambas dimensiones.

Figura 12-8
Contribuciones de los atributos

imagen	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
engorda	,080	-,514	-,265	,033	,042	,035	,652	,173	,825
hombres	,051	-,852	,825	,072	,073	,219	,512	,480	,992
SudAustraliano	,057	-,303	-,350	,046	,010	,044	,114	,152	,266
tradicional	,040	-,703	-,532	,043	,039	,071	,454	,260	,715
superior	,042	-,444	-,582	,028	,016	,090	,296	,509	,805
sano	,053	1,200	,174	,081	,152	,010	,953	,020	,973
cafeína	,047	-,452	,124	,014	,019	,005	,702	,053	,755
nuevo	,047	,960	,147	,048	,086	,006	,893	,021	,914
atractivo	,041	,657	-,056	,019	,035	,001	,911	,007	,918
duro	,039	-,850	1,002	,070	,056	,246	,404	,560	,964
popular	,060	-,697	-,042	,038	,058	,001	,771	,003	,774
curativo	,026	-,389	,266	,009	,008	,011	,446	,209	,655
bajo en grasa	,052	1,305	,196	,094	,175	,013	,941	,021	,962
niños	,024	-,352	-,513	,017	,006	,041	,179	,380	,559
trabajando	,045	-,785	,477	,040	,055	,064	,693	,255	,948
dulce	,038	-,519	-,683	,048	,020	,112	,212	,368	,580
impopular	,024	,489	,186	,010	,011	,005	,585	,085	,670
feo	,030	,006	-,109	,003	,000	,002	,000	,131	,131
fresco	,036	-,096	-,100	,002	,001	,002	,196	,214	,410
yupis	,034	,380	-,301	,012	,010	,019	,392	,246	,637
nutritivo	,040	,722	,055	,022	,041	,001	,946	,006	,951
mujeres	,054	,758	-,063	,032	,062	,001	,965	,007	,972
secundario	,040	,579	,063	,023	,027	,001	,593	,007	,600
Total activo	1,000			,804	1,000	1,000			

Dos dimensiones contribuyen con una gran cantidad a la inercia para la mayoría de los puntos de fila. Las contribuciones grandes a la primera dimensión de *sana, nueva, atractivas, baja en calorías, nutritiva y mujeres* indican que estos puntos aparecen muy bien representados en una dimensión. Por consiguiente, las dimensiones superiores contribuyen poco a la inercia de estos puntos, que estarán situados muy cerca del eje horizontal. La segunda dimensión contribuye sobre todo a *hombres, calidad y severa*. Ambas dimensiones contribuyen muy poco a la inercia de *sur de Australia y feas*, por lo que estos puntos aparecen pobremente representados.

La visión general de los puntos de columna muestra las contribuciones que implican a los puntos de columna. Las marcas *CC* y *DD* contribuyen sobre todo a la primera dimensión, mientras que *EE* y *FF* explican una gran cantidad de la inercia para la segunda dimensión. *AA* y *BB* contribuyen muy poco a ambas dimensiones.

Figura 12-9
Contribuciones a la marca

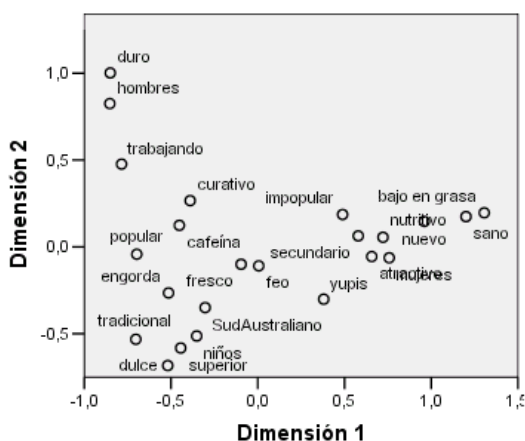
marca	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
AA	,217	-,659	,046	,127	,187	,003	,744	,004	,748
BB	,131	-,284	-,404	,078	,021	,134	,135	,272	,407
CC	,185	,996	,076	,193	,362	,007	,951	,006	,957
DD	,162	,915	,101	,146	,267	,010	,928	,011	,939
EE	,152	-,651	,706	,153	,127	,477	,420	,494	,914
FF	,153	-,343	-,618	,107	,036	,369	,169	,550	,718
Total activo	1,000			,804	1,000	1,000			

En dos dimensiones, todas las marcas salvo *BB* están bien representadas. *CC* y *DD* están bien representadas en una dimensión. La segunda dimensión contribuye a *EE* y *FF* con sus mayores cuantías. Observe que *AA* está bien representada en la primera dimensión, pero no tiene una contribución muy alta a dicha dimensión.

Gráficos

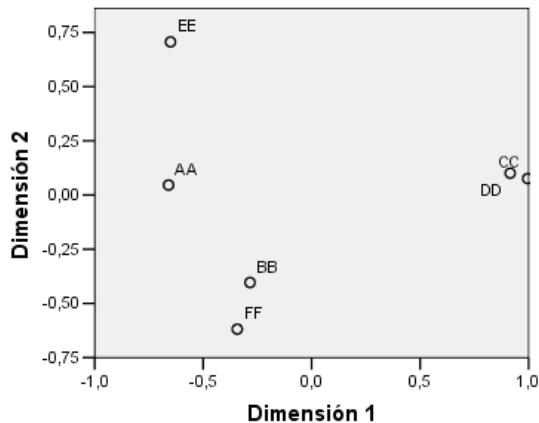
El gráfico de puntos de fila muestra que *frescor* y *feas* están muy cerca del origen, lo que indica que difieren muy poco del perfil de fila medio. Surgen así tres clasificaciones generales. Situado en la parte izquierda superior del gráfico, *severa, hombres y trabajadora* son similares entre sí. La parte inferior izquierda contiene *dulce, engorda, niños y calidad*. Por el contrario, *sana, baja en calorías, nutritiva y nueva* se agrupan en la parte izquierda del gráfico.

Figura 12-10
Gráfico de los atributos de imagen (normalización principal)



Observe en el gráfico de los puntos de columna que todas las marcas están lejos del origen, por lo que no hay ninguna marca que sea similar al centroide global. Las marcas *CC* y *DD* se agrupan juntas a la derecha, mientras que las marcas *BB* y *FF* se agrupan en la mitad inferior del gráfico. Las marcas *AA* y *EE* no son similares a ninguna otra marca.

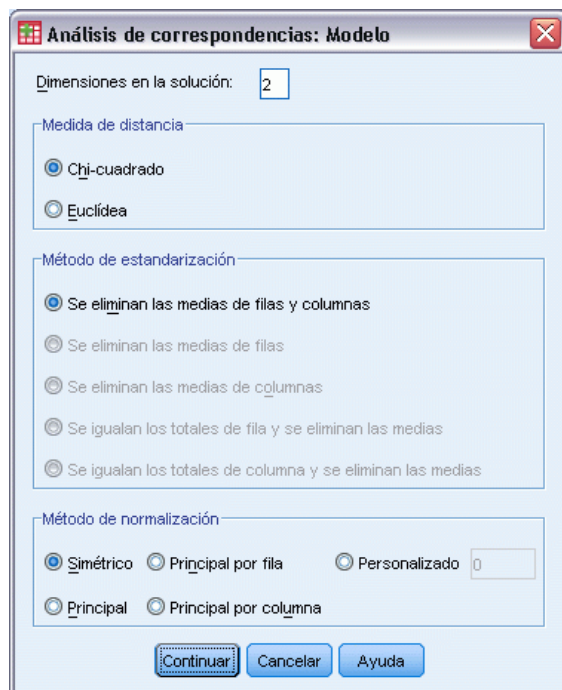
Figura 12-11
Gráfico de marcas (normalización principal)



Normalización simétrica

¿Cómo están relacionadas las marcas con los atributos de imagen? La normalización principal no puede tratar estas relaciones. Para centrarnos en cómo están relacionadas las variables entre sí, utilizaremos la normalización simétrica. En vez de distribuir la inercia dos veces (como ocurre en la normalización principal), la normalización simétrica divide la inercia de idéntica manera sobre las filas y las columnas. Las distancias entre categorías para una única variable no se pueden interpretar, pero las distancias entre las categorías de diferentes variables son significativas.

Figura 12-12
Cuadro de diálogo Modelo



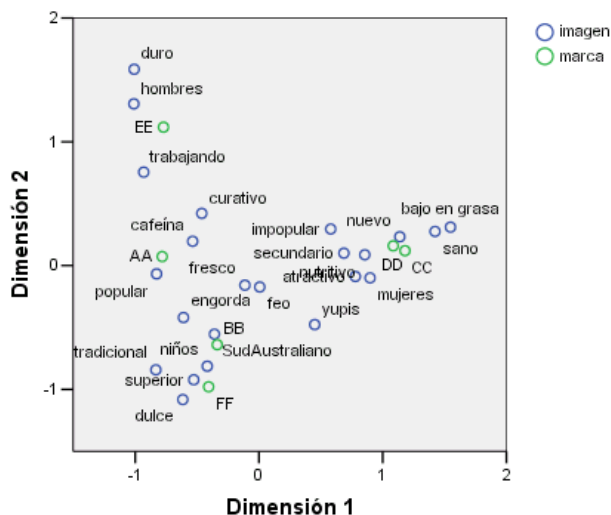
- ▶ Para generar la siguiente solución con la normalización simétrica, vuelva al cuadro de diálogo Análisis de correspondencias y pulse en Modelo.
- ▶ Seleccione Simétrica como método de normalización.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Análisis de correspondencias.

En la parte superior del diagrama de dispersión biespacial resultante, la marca *EE* es la única severa y de la clase trabajadora que resulta atractiva a los hombres. La marca *AA* es la más popular y también se considera la que tiene más caféina. Las marcas dulces y que engordan

incluyen *BB* y *FF*. Las marcas *CC* y *DD*, aunque se consideran como nuevas y sanas, también son las menos populares.

Figura 12-13

Diagrama de dispersión biespacial de las marcas y los atributos (normalización simétrica)



Para continuar con la interpretación, puede dibujar una línea a través del origen y los dos atributos de imagen *hombres* y *yupis*, y proyectar las marcas sobre esta línea. Los dos atributos están opuestos el uno al otro, lo que indica que el patrón de asociación de las marcas para *hombres* está invertido en comparación con el patrón de *yupis*. Es decir, los hombres son los que están asociados con mayor frecuencia con la marca *EE* y con menor frecuencia con la marca *CC*, mientras que los yupis se asocian con mayor frecuencia con la marca *CC* y con menor frecuencia con la marca *EE*.

Lecturas recomendadas

Si desea obtener más información sobre el análisis de correspondencias:

Fisher, R. A. 1938. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Fisher, R. A. 1940. The precision of discriminant functions. *Annals of Eugenics*, 10, .

Gilula, Z., y S. J. Haberman. 1988. The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83, .

Análisis de correspondencias múltiple

El propósito del análisis de correspondencias múltiple, también conocido como análisis de homogeneidad, es encontrar las cuantificaciones óptimas, es decir, que separen al máximo las categorías entre sí. Esto implica que los objetos de la misma categoría se representan cercanos entre sí, mientras que los objetos de categorías diferentes se representan lo más alejados posible. El término **homogeneidad** también se refiere al hecho de que se efectuará el análisis más adecuado cuando las variables sean homogéneas, es decir, cuando dividan los objetos en conglomerados con las mismas o similares categorías.

Ejemplo: Características de la ferretería

Para explorar cómo funciona el análisis de correspondencias múltiple, se utilizarán los datos de Hartigan (Hartigan, 1975), que pueden encontrarse en *screws.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20](#). Este conjunto de datos contiene información acerca de las características de los tornillos, los pernos, las tuercas y las tachuelas. La siguiente tabla muestra las variables (junto con las etiquetas de variable) y las etiquetas de valor que se han asignado a las categorías de cada variable en el conjunto de datos de la ferretería Hartigan.

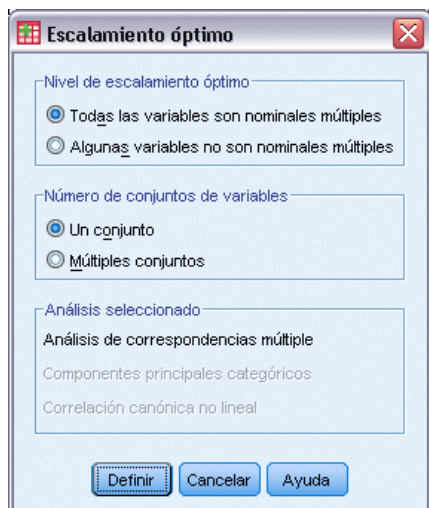
Tabla 13-1
Conjunto de datos de la ferretería Hartigan

Nombre de variable	Etiqueta de variable	Etiqueta de valor
<i>rosca</i>	<i>Rosca</i>	<i>Sí_Rosca, No_Rosca</i>
<i>cabeza</i>	<i>Forma de la cabeza</i>	<i>Plana, ahuecada, cónica, redonda, cilíndrica</i>
<i>muesca</i>	<i>muesca de la cabeza</i>	<i>Ninguno, en estrella, rallado</i>
<i>punta</i>	<i>Forma de la punta</i>	<i>afilada, plana</i>
<i>long</i>	<i>Longitud en mm</i>	<i>1/2 mm, 1 mm, 1 1/2 mm, 2 mm, 2 1/2 mm</i>
<i>cobre</i>	<i>Cobre</i>	<i>Sí_Co, No_Co</i>
<i>objeto</i>	<i>Objeto</i>	<i>tachuela, clavo1, clavo2, clavo3, clavo4, clavo5, clavo6, clavo7, clavo8, tornillo1, tornillo2, tornillo3, tornillo4, tornillo5, perno1, perno2, perno3, perno4, perno5, perno6, tachuela1, tachuela2, clavo1, tornillo1</i>

Ejecución del análisis

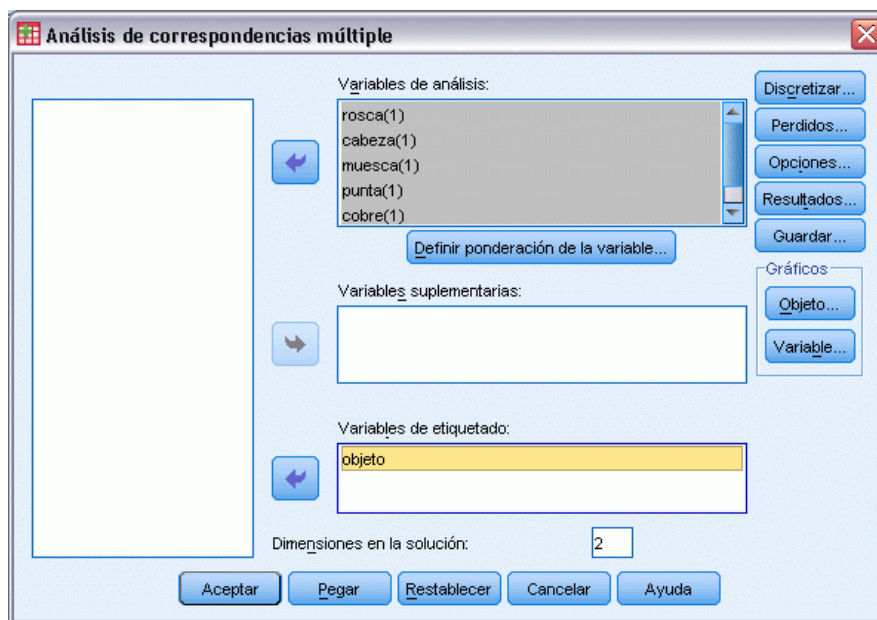
- Para obtener un análisis de correspondencias múltiple, elija en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...

Figura 13-1
Cuadro de diálogo Escalamiento óptimo



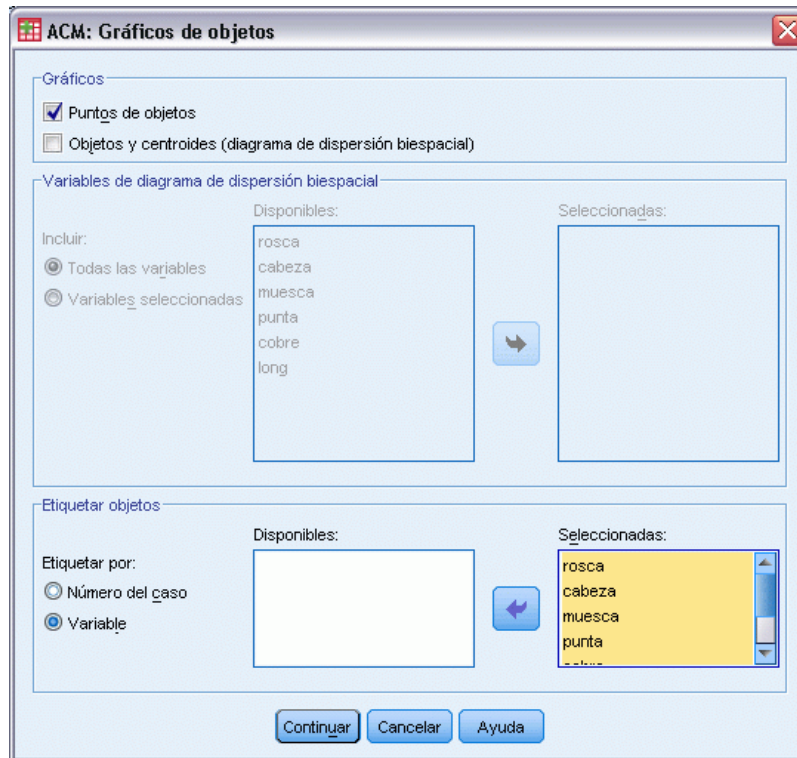
- Asegúrese de que están seleccionadas Todas las variables son nominales múltiples y Un conjunto y pulse en Definir.

Figura 13-2
Cuadro de diálogo Análisis de correspondencias múltiple



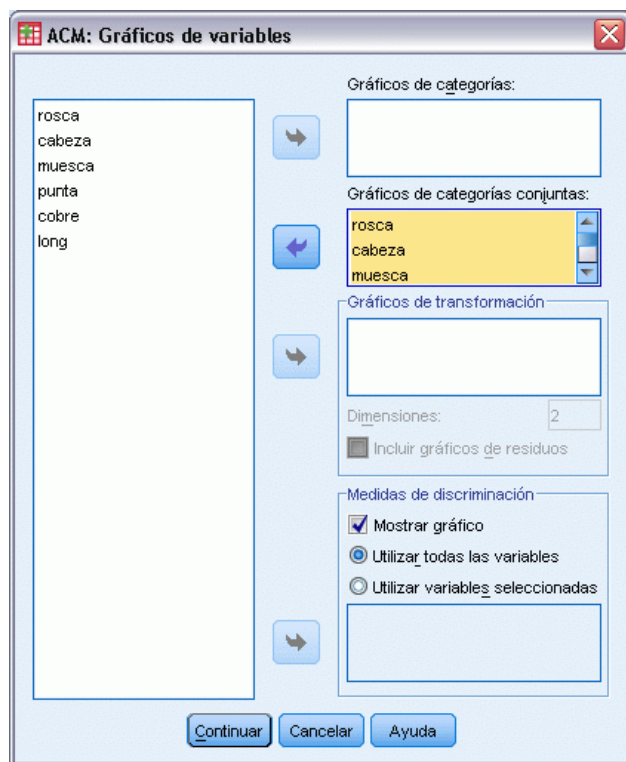
- Seleccione como variables de análisis desde *Rosca* hasta *Longitud en mm*.
- Seleccione como variable de etiquetado *objeto*.
- Pulse en Objeto en el grupo Gráficos.

Figura 13-3
Cuadro de diálogo Gráficos de objetos



- ▶ Seleccione etiquetar los objetos por Variable.
- ▶ Seleccione como variables de etiquetado desde *rosca* hasta *objeto*.
- ▶ Pulse en Continuar y, a continuación, pulse en Variable en el grupo Gráficos del cuadro de diálogo Análisis de correspondencias múltiple.

Figura 13-4
Cuadro de diálogo Gráficos de variables



- ▶ Elija generar un gráfico de categorías conjuntas desde *rosca* hasta *long*.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Análisis de correspondencias múltiple.

Resumen del modelo

El análisis de homogeneidad puede calcular una solución para varias dimensiones. El número máximo de dimensiones es igual al número de categorías menos el número de variables sin datos perdidos o menos el número de observaciones menos uno, el que sea menor. Sin embargo, en raras ocasiones se debería utilizar el número máximo de dimensiones. Un número menor de dimensiones es más fácil de interpretar y, después de un cierto número de dimensiones, la cantidad de asociación adicional explicada se vuelve despreciable. Una solución de una, dos o tres dimensiones en el análisis de homogeneidad es muy común.

Figura 13-5
Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada		
		Total (Autovalores)	Inercia	% de la varianza
1	,878	3,727	,621	62,123
2	,657	2,209	,368	36,809
Total		5,936	,989	
Media	,796 ^a	2,968	,495	49,466

^a El Alfa de Cronbach Promedio está basado en los autovalores promedio.

Casi la totalidad de la varianza de los datos se explica mediante la solución, 62,1% mediante la primera y el 36,8% gracias a la segunda dimensión.

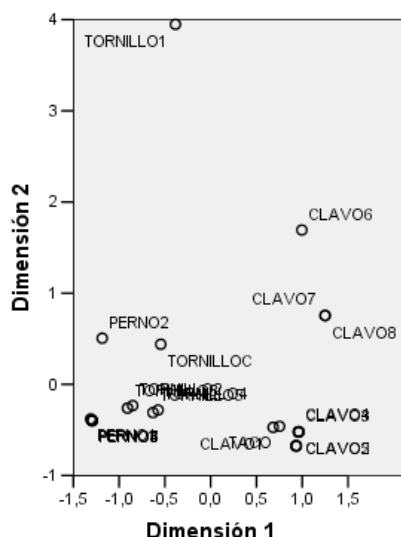
Ambas dimensiones juntas proporcionan una interpretación según las distancias. Si una variable discrimina correctamente, los objetos se encontrarán cerca de las categorías a las que pertenecen. Lo ideal es que los objetos de la misma categoría estén cercanos entre sí (es decir, tengan puntuaciones similares), y las categorías de variables diferentes estén cercanas si pertenecen a los mismos objetos (es decir, dos objetos con puntuaciones similares para una variable deben también puntuar cerca uno del otro para el resto de variables de la solución).

Puntuaciones de los objetos

Después de examinar el resumen del modelo, debe observar las puntuaciones de los objetos. Es posible especificar una o varias variables para etiquetar el gráfico de las puntuaciones del objeto. Cada variable de etiquetado genera un gráfico diferente etiquetado con los valores de esa variable. Vamos a echar un vistazo al gráfico de las puntuaciones del objeto etiquetadas por el objeto de la variable. Ésta es simplemente una variable de identificación de casos y no se ha utilizado para realizar los cálculos.

La distancia desde un objeto hasta el origen refleja la variación desde el patrón de respuesta “medio”. Este patrón se corresponde con la categoría más frecuente para cada variable. Los objetos con muchas características que se corresponden con las categorías más frecuentes deben estar cerca del origen. Por el contrario, los objetos con características únicas están localizados lejos del origen.

Figura 13-6
Gráficos con las puntuaciones de los objetos etiquetados por el objeto



Si examina el gráfico, observará que la primera dimensión (el eje horizontal) diferencia entre los tornillos y los pernos (que tienen roscas) y los clavos y las tachuelas (que no tienen rosca). Es fácil observar esto en el gráfico ya que los tornillos y los pernos se encuentran en uno de los extremos del eje horizontal, y las tachuelas y los clavos están en el contrario. En menor grado, la primera dimensión también separa los pernos (que tienen las puntas planas) del resto (con puntas afiladas).

La segunda dimensión (el eje vertical) parece separar *TORNILLO1* y *CLAVO6* del resto de objetos. Lo que *TORNILLO1* y *CLAVO6* tienen en común son los valores en la longitud de la variable: son los objetos más largos de los datos. Además, *TORNILLO1* está mucho más lejos del origen que los otros objetos, lo que sugiere que, si lo tomamos como un todo, no se comparten muchas de las características de este objeto con otros objetos.

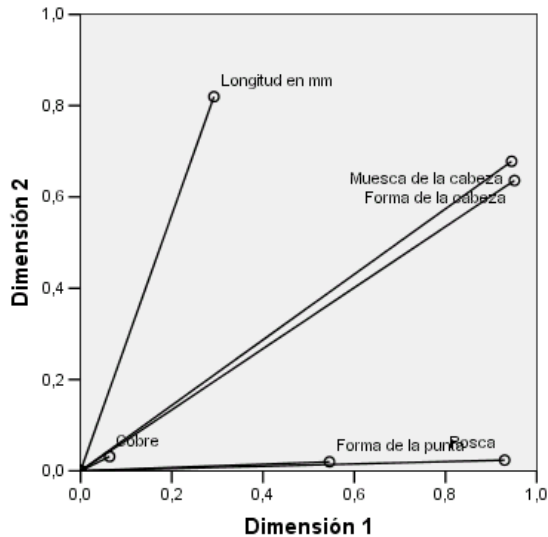
El gráfico de puntuaciones del objeto es especialmente útil para observar los valores atípicos. *TORNILLO1* puede considerarse un valor atípico. Más adelante, veremos qué pasa si se hace descender este objeto.

Medidas de discriminación

Antes de examinar el resto de gráficos de puntuaciones de objeto, vemos a ver si las medidas de discriminación concuerdan con lo que hemos dicho hasta ahora. Para cada variable, se calcula una medida de discriminación para cada dimensión, que puede ser tratada como una saturación del componente cuadrado. Esta medida es también la varianza de la variable cuantificada en esa dimensión. Tiene un valor máximo de 1, que se logra si ambas puntuaciones del objeto disminuyen hacia grupos exclusivos y todas las puntuaciones de los objetos de una categoría son idénticas. (Nota: Esta medida puede tener un valor mayor que 1 si existen datos perdidos.) Las medidas de discriminación grandes se corresponden con una gran dispersión entre las categorías de la variable y, por consiguiente, indican un alto grado de discriminación entre las categorías de una variable a lo largo de esa dimensión.

La media de las medidas de discriminación para cualquier dimensión es igual al porcentaje de varianza explicado para esa dimensión. Por consiguiente, las dimensiones se ordenan según la discriminación de la media. La primera dimensión tiene la discriminación media mayor, la segunda dimensión tiene la segunda discriminación media mayor, y así sucesivamente para todas las dimensiones de la solución.

Figura 13-7
Gráfico de medidas de discriminación



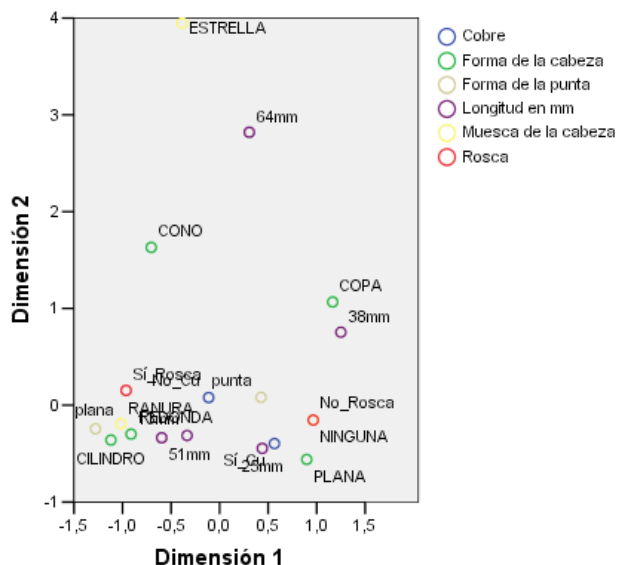
Como se ha registrado en el gráfico de puntuaciones del objeto, el gráfico de las medidas de discriminación muestra que la primera dimensión está relacionada con las variables *Rosca y Forma de la punta*. Estas variables tienen medidas de discriminación grandes en la primera dimensión y medidas de discriminación pequeñas en la segunda dimensión. Así, para estas dos variables, las categorías se dispersan sólo lejos de la primera dimensión. *Longitud (en medias pulgadas)* tiene un valor grande en la segunda dimensión, pero un valor pequeño en la primera. Como resultado, *long* está más cerca de la segunda dimensión, de acuerdo con la observación del gráfico de puntuaciones de objetos de que la segunda dimensión parece separar los objetos más largos del resto. *Muesca de la cabeza y Forma de la cabeza* tienen valores relativamente grandes en ambas dimensiones, lo que indica discriminación en la primera y segunda dimensiones. La variable *Latón*, situada muy cerca del origen, no discrimina en absoluto en las dos primeras dimensiones. Esto tiene sentido ya que todos los objetos pueden estar hechos o no de latón.

Quantificaciones de categorías

Recuerde que una medida de discriminación es la varianza de la variable cuantificada junto con una dimensión particular. El gráfico de medidas de discriminación contiene estas varianzas, lo que indica qué variables discriminan junto con qué dimensión. Sin embargo, la misma varianza podría corresponderse con todas las categorías que se hayan dispersado medianamente lejos, o con la mayoría de las categorías juntas (con unas cuantas categorías diferentes de este grupo). El gráfico de discriminación no puede diferenciar entre estas dos condiciones.

Los gráficos de cuantificación de las categorías proporcionan un método alternativo de mostrar la discriminación de las variables que pueden identificar las relaciones de las categorías. En este gráfico, se muestran las coordenadas de cada categoría en cada dimensión. De este modo, es posible determinar qué categorías son similares para cada variable.

Figura 13-8
Cuantificaciones de categorías



Longitud en mm tiene cinco categorías, tres de las cuales se agrupan en la parte superior del gráfico. Las dos categorías restantes se encuentran en la mitad inferior del gráfico, con la categoría *2_1/2_mm* realmente lejos del grupo. La discriminación grande para la longitud a lo largo de la dimensión 2 es el resultado de que esta categoría sea muy diferente a las otras categorías de longitud. De igual modo, para *Forma de la cabeza*, la categoría *EN ESTRELLA* se encuentra muy lejos del resto de categorías y provoca una medida de discriminación grande a lo largo de la segunda dimensión. Estos patrones no pueden ilustrarse en un gráfico de medidas de discriminación.

La dispersión de las cuantificaciones de la categoría para una variable refleja la varianza y, por lo tanto, indica la calidad de discriminación de esa variable en cada dimensión. Si nos centramos en la dimensión 1, las categorías para *Rosca* están lejos. Sin embargo, a lo largo de la dimensión 2, las categorías para esta variable están muy cerca. Así, *Rosca* discrimina mejor en dimensión 1 que en dimensión 2. Por el contrario, las categorías para *Forma de la cabeza* están dispersadas y alejadas a lo largo de ambas dimensiones, lo que sugiere que esta variable discrimina bien en ambas dimensiones.

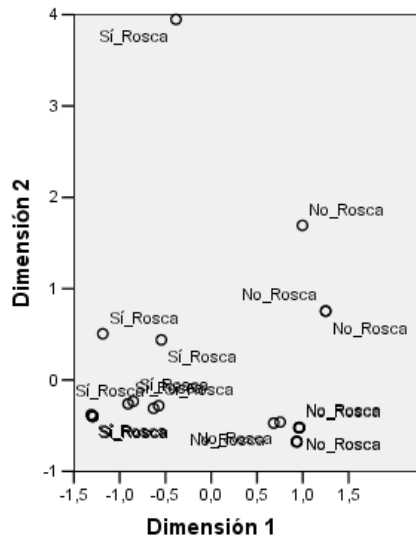
Además de determinar las dimensiones a lo largo de las que una variable discrimina, así como la forma de discriminar de dicha variable, el gráfico de cuantificación de la categoría también compara la discriminación de la variable. Una variable con categorías que están alejadas entre sí discrimina mejor que una variable con categorías que se encuentran muy juntas. Por ejemplo, a lo largo de dimensión 1, las dos categorías de *Latón* están mucho más cerca entre sí que las dos categorías de *Rosca*, lo que indica que *Rosca* discrimina mejor que *Latón* a lo largo de esta dimensión. Sin embargo, a lo largo de la dimensión 2, las distancias son muy similares, lo que sugiere que estas variables discriminan al mismo nivel a lo largo de esta dimensión. El gráfico de

medidas de discriminación que se ha discutido anteriormente identifica estas mismas relaciones mediante el uso de las varianzas para reflejar la dispersión de las categorías.

Una observación más detallada de las puntuaciones de los objetos

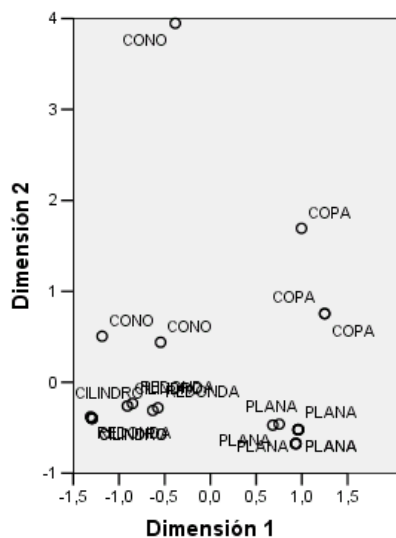
Se puede obtener información más detallada acerca de los datos si se examinan los gráficos de puntuaciones de objetos etiquetados por cada variable. Lo ideal es que los objetos similares deberían formar grupos exclusivos, que deberían encontrarse alejados entre sí.

Figura 13-9
Puntuaciones de objetos etiquetados con Rosca



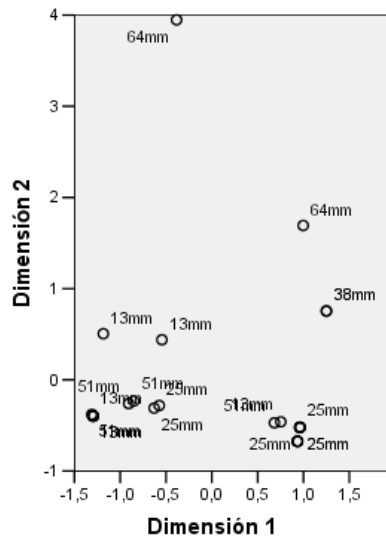
El gráfico etiquetado con *Rosca* muestra que la primera dimensión separa *Sí_Rosca* y *No_Rosca* perfectamente. Todos los objetos con rosca tienen puntuaciones de objeto negativas, mientras que todos los objetos sin rosca tienen puntuaciones positivas. Aunque las dos categorías no formen grupos compactos, la perfecta diferenciación entre las categorías se considera generalmente un buen resultado.

Figura 13-10
Puntuaciones de objetos etiquetados con *Forma de la cabeza*



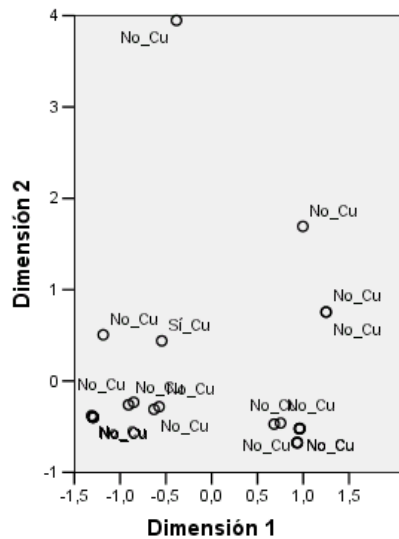
El gráfico etiquetado con *Forma de la cabeza* muestra que esta variable discrimina en ambas dimensiones. El grupo de objetos *PLANA* se encuentran agrupados en la esquina inferior derecha del gráfico, mientras que el grupo de objetos *AHUECADA* lo hacen en la esquina superior derecha. Los objetos *CÓNICA* se encuentran en la esquina superior izquierda. Sin embargo, estos objetos están más dispersados que el resto de grupos y, por lo tanto, no son tan homogéneos. Finalmente, los objetos de la categoría *CILÍNDRICA* no pueden separarse de los objetos de la categoría *REDONDA*, los cuales se encuentran en la esquina inferior izquierda del gráfico.

Figura 13-11
Las puntuaciones de los objetos etiquetadas con *Longitud* (en medias pulgadas)



El gráfico etiquetado con *Longitud en mm* muestra que esta variable no discrimina en la primera dimensión. Sus categorías no muestran agrupación cuando se proyectan sobre una línea horizontal. Sin embargo, *Longitud en mm* sí discrimina en la segunda dimensión. Los objetos más cortos se corresponden con puntuaciones positivas, mientras que los más largos lo hacen con puntuaciones negativas.

Figura 13-12
Puntuaciones de objetos etiquetados con *Latón*



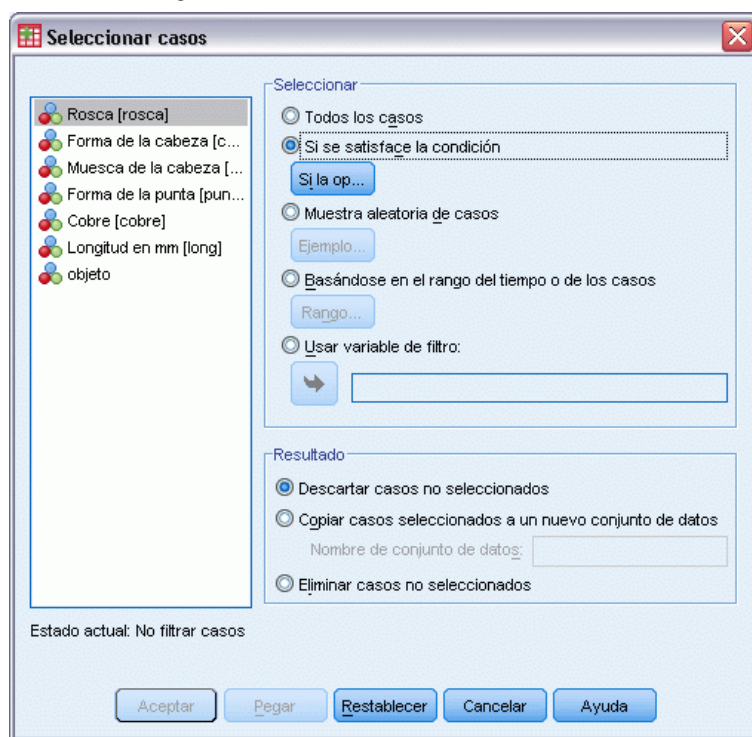
El gráfico etiquetado con *Cobre* muestra que esta variable tiene categorías que no pueden separarse de manera correcta en la primera o segunda dimensiones. Las puntuaciones de los objetos están muy dispersados a lo largo del espacio. Los objetos de cobre no pueden diferenciarse de los objetos que no son de cobre.

Omisión de valores atípicos

En el análisis de homogeneidad, los valores atípicos son objetos que tienen demasiadas funciones únicas. Como dijimos anteriormente, *TORNILLO1* puede considerarse un valor atípico.

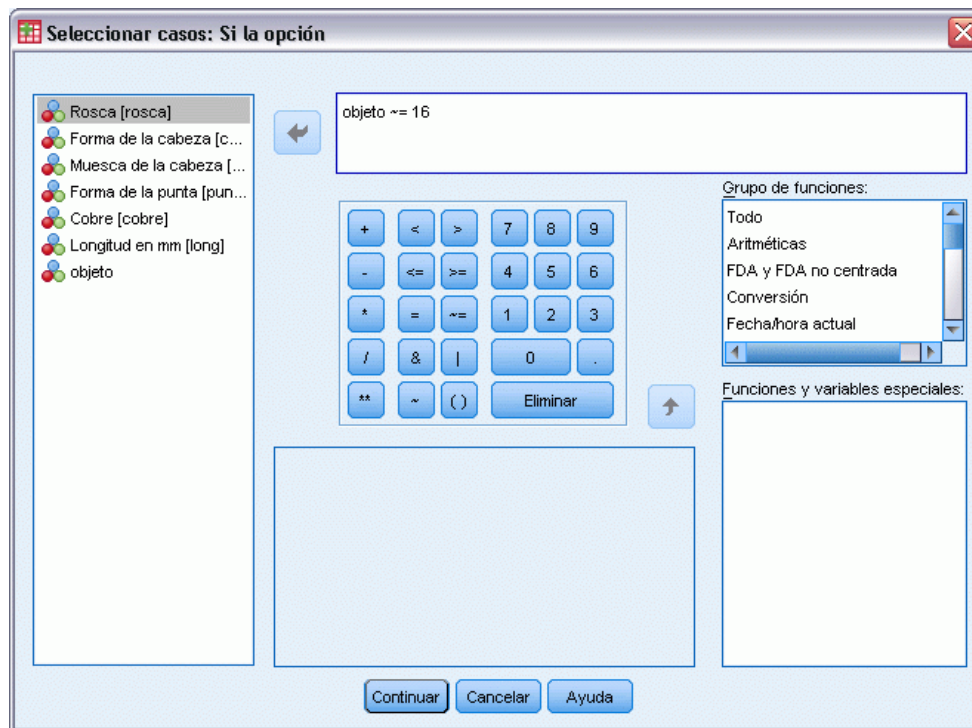
Para eliminar este objeto y ejecutar de nuevo el análisis, elija en los menús:
Datos > Seleccionar casos...

Figura 13-13
Cuadro de diálogo Seleccionar casos



- ▶ Seleccione Si se satisface la condición.
- ▶ Pulse en Si.

Figura 13-14
Si el cuadro de diálogo



- ▶ Introduzca objeto ~= 16 como la condición.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Seleccionar casos.
- ▶ Finalmente, vuelva al cuadro de diálogo Análisis de correspondencias múltiple y pulse en Aceptar.

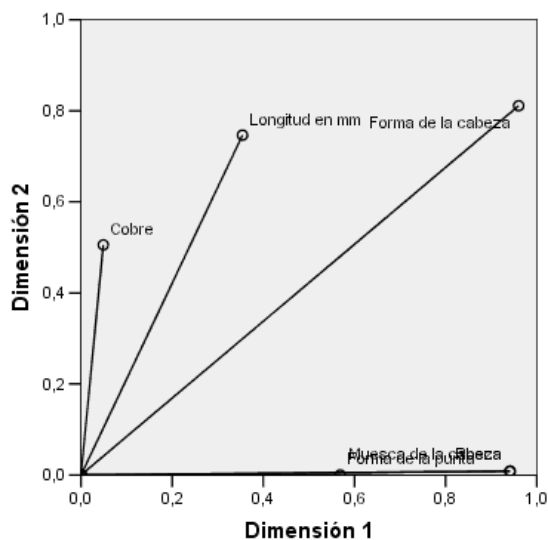
Figura 13-15
Resumen del modelo (valor atípico eliminado)

Dimensión	Alfa de Cronbach	Varianza explicada		
		Total (Autovalores)	Inercia	% de la varianza
1	,885	3,815	,636	63,591
2	,623	2,081	,347	34,676
Total		5,896	,983	
Media	,793 ^a	2,948	,491	49,133

^a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

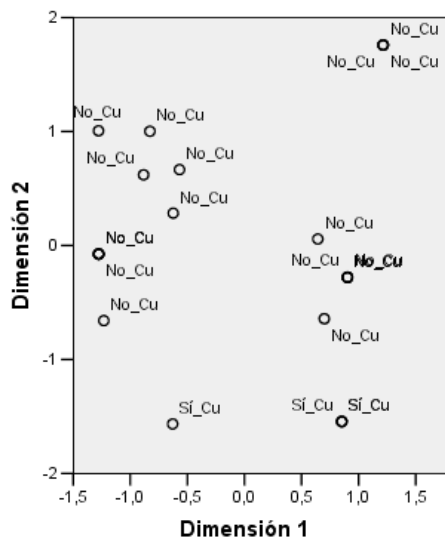
Los autovalores cambian ligeramente. La primera dimensión ahora explica un poco más de la varianza.

Figura 13-16
Medidas discriminantes



Como se muestra en el gráfico de discriminación, *Muesca de la cabeza* ya no discrimina en la segunda dimensión, mientras que *Cobre* varía de no discriminación en ninguna de las dimensiones a discriminación en la segunda dimensión. La discriminación para el resto de variables no cambia en su mayor parte.

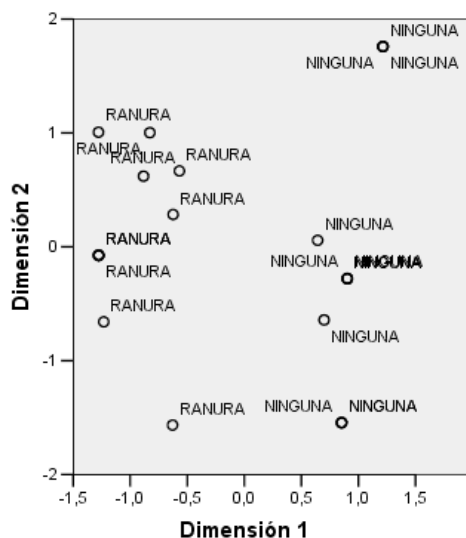
Figura 13-17
Las puntuaciones del objeto etiquetadas con Latón (valor atípico eliminado)



El gráfico de puntuaciones del objeto etiquetado por *Latón* muestra que los cuatro objetos de latón aparecen cerca de la parte inferior del gráfico (tres objetos ocupan idénticas localizaciones), lo que indica alta discriminación a lo largo de la segunda dimensión. Como era el caso para *Rosca* en el análisis anterior, los objetos no forman grupos compactos, pero la diferenciación de objetos mediante las categorías es perfecta.

Figura 13-18

Las puntuaciones del objeto etiquetadas con *Muesca de la cabeza* (valor atípico eliminado)



El gráfico de puntuaciones del objeto etiquetado mediante *Muesca de la cabeza* muestra que la primera dimensión discrimina perfectamente entre los objetos con y sin muesca, como en el análisis anterior. Sin embargo, al contrario que en el análisis anterior, la segunda dimensión no puede distinguir ahora las dos categorías.

Por ello, la omisión de *TORNILLO1*, que es el único objeto con una cabeza en forma de estrella, afecta drásticamente a la interpretación de la segunda dimensión. Esta dimensión diferencia ahora los objetos basados en *Cobre*, *Forma de la cabeza* y *Longitud en mm*.

Lecturas recomendadas

Si desea obtener más información sobre el análisis de correspondencias múltiple:

Benzécri, J. P. 1992. *Correspondence analysis handbook*. Nueva York: Marcel Dekker.

Guttman, L. 1941. The quantification of a class of attributes: A theory and method of scale construction. En: *The Prediction of Personal Adjustment*, P. Horst, ed. Nueva York: Social Science Research Council.

Meulman, J. J. 1982. *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.

Meulman, J. J. 1996. Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification*, 13, .

Meulman, J. J., y W. J. Heiser. 1997. Graphical display of interaction in multiway contingency tables by use of homogeneity analysis. En: *Visual Display of Categorical Data*, M. Greenacre, y J. Blasius, eds. Nueva York: Academic Press.

Nishisato, S. 1984. Forced classification: A simple application of a quantification method. *Psychometrika*, 49, .

Tenenhaus, M., y F. W. Young. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, .

Van Rijckevorsel, J. 1987. *The application of fuzzy coding and horseshoes in multiple correspondence analysis*. Leiden: DSWO Press.

Multidimensional Scaling

Dado un conjunto de objetos, el propósito del escalamiento multidimensional es encontrar una representación de los objetos en un espacio de pocas dimensiones. La solución más sencilla es utilizar **proximidades** entre los objetos. El procedimiento minimiza las desviaciones cuadradas entre las proximidades de los objetos, originales, posiblemente transformadas, y sus distancias euclídeas en el espacio de pocas dimensiones.

El propósito del espacio de pocas dimensiones es descubrir las relaciones entre los objetos. Si se restringe la combinación para que sea una combinación lineal de variables independientes, es posible que sea capaz de interpretar las dimensiones de la solución en función de estas variables. En el siguiente ejemplo, observará cómo 15 términos diferentes de parentesco pueden representarse en tres dimensiones y cómo ese espacio puede interpretarse en función del sexo, la generación y el grado de separación de cada uno de los términos.

Ejemplo: Un examen de los términos de parentesco

Rosenberg y Kim (Rosenberg y Kim, 1975) comenzaron a analizar 15 términos de parentesco [tía, hermano, primos, hija, padre, nieta, abuelo, abuela, nieto, madre, sobrino, sobrina, hermana, hijo, tío]. Le pidieron a cuatro grupos de estudiantes universitarios (dos masculinos y dos femeninos) que ordenaran estos grupos según las similitudes. A dos grupos (uno masculino y otro femenino) se les pidió que realizaran la ordenación dos veces, pero que la segunda ordenación la hicieran según criterios distintos a los de la primera. Así, se obtuvieron un total de seis “fuentes”, como se explica en la siguiente tabla.

Tabla 14-1
Estructura de la fuente de los datos de parentesco

Origen	Sexo	Condición	Tamaño muestral
1	Mujer	Ordenación única	85
2	Hombre	Ordenación única	85
3	Mujer	Primera ordenación	80
4	Mujer	Segunda ordenación	80
5	Hombre	Primera ordenación	80
6	Hombre	Segunda ordenación	80

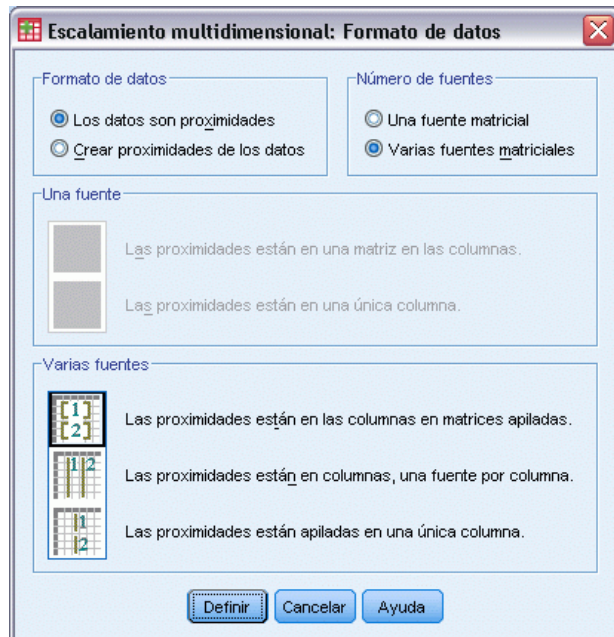
Cada fuente se corresponde con una matriz de proximidades de 15×15 cuyas casillas son iguales al número de personas de una fuente menos el número de veces que se partitionaron los objetos en esa fuente. Este conjunto de datos se puede encontrar en *kinship_dat.sav*. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en *IBM SPSS Categories 20*.

Elección del número de dimensiones

Depende de usted decidir cuántas dimensiones debe tener la solución. El diagrama de sedimentación puede ayudarle a tomar esta decisión.

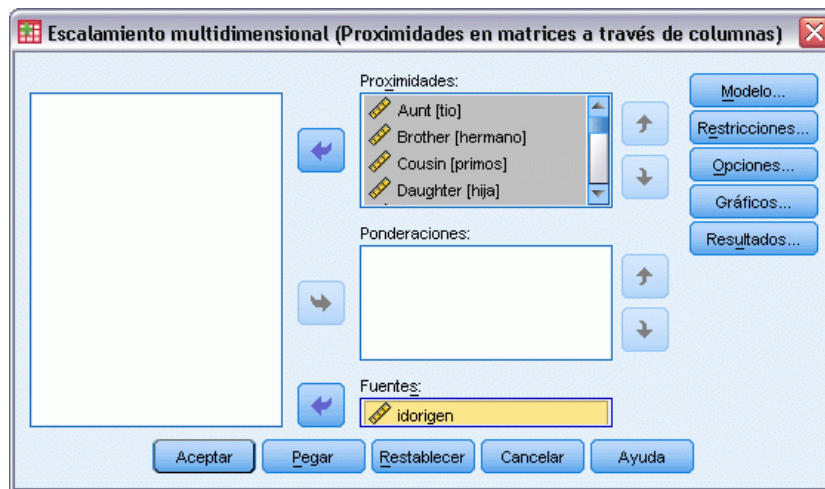
- Para crear un diagrama de sedimentación, elija en los menús:
Analizar > Escala > Escalamiento multidimensional (PROXSCAL)...

Figura 14-1
Cuadro de diálogo Formato de datos



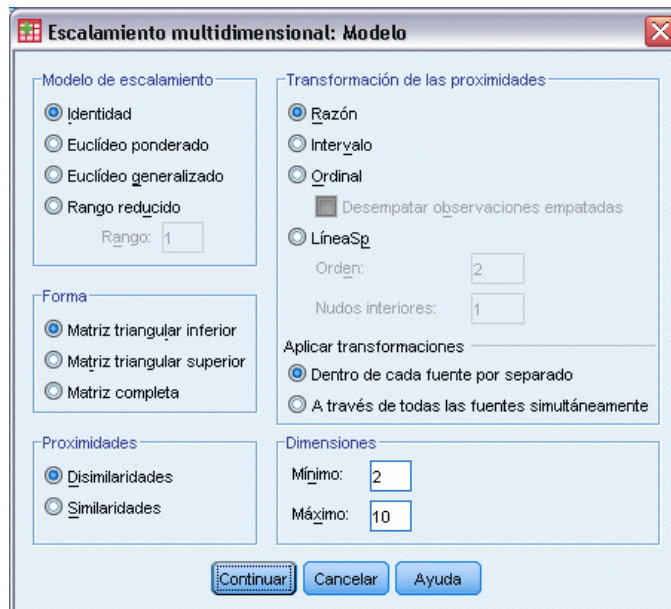
- Seleccione Varias fuentes matriciales en el grupo Número de fuentes.
- Pulse en Definir.

Figura 14-2
Cuadro de diálogo Escalamiento multidimensional



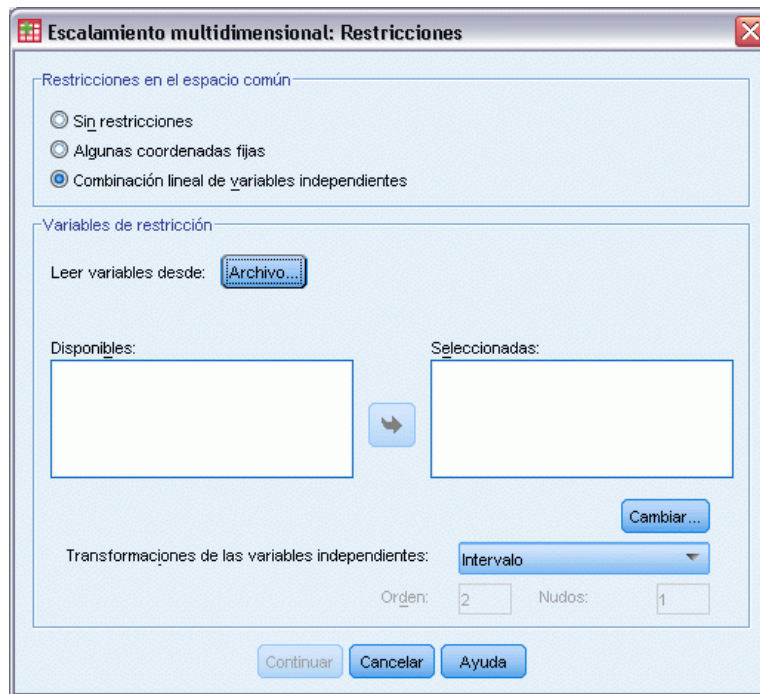
- ▶ Seleccione *Tía* hasta *Tío* como variables de proximidad.
- ▶ Seleccione *idorigen* como la variable que identifica la fuente.
- ▶ Pulse en *Modelo*.

Figura 14-3
Cuadro de diálogo *Modelo*



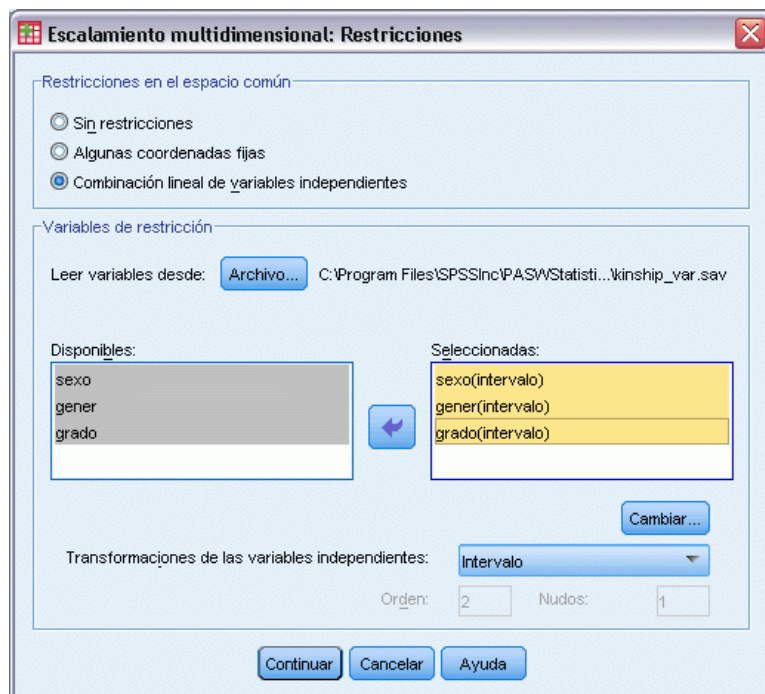
- ▶ Introduzca 10 como el número máximo de dimensiones.
- ▶ Pulse en *Continuar*.
- ▶ Pulse en *Restricciones* en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-4
Cuadro de diálogo Restricciones



- ▶ Seleccione *Combinación lineal de variables independientes*.
- ▶ Pulse *Archivo* para seleccionar la fuente de variables independientes.
- ▶ Seleccione *kinship_var.sav*.

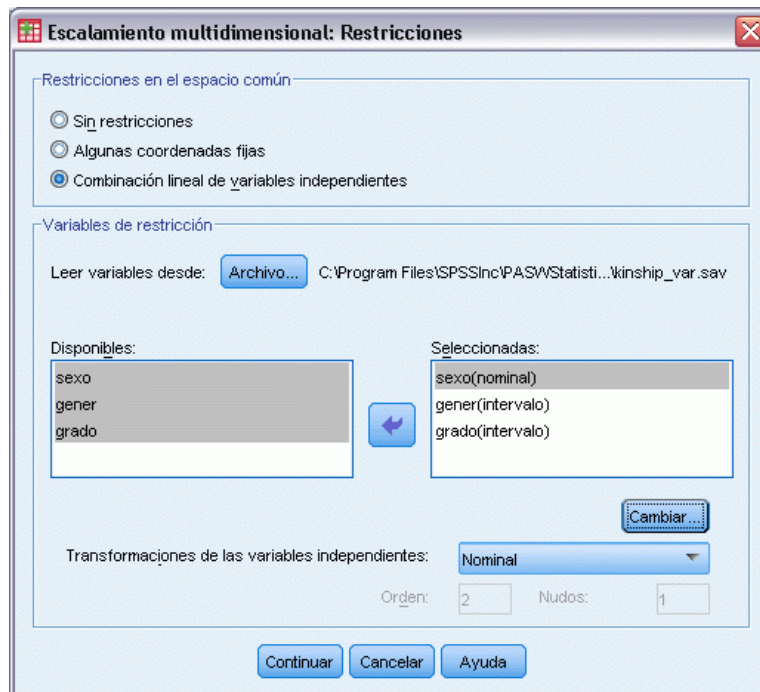
Figura 14-5
Cuadro de diálogo Restricciones



- Seleccione *sexo*, *gener* y *grado* como variables de restricción.

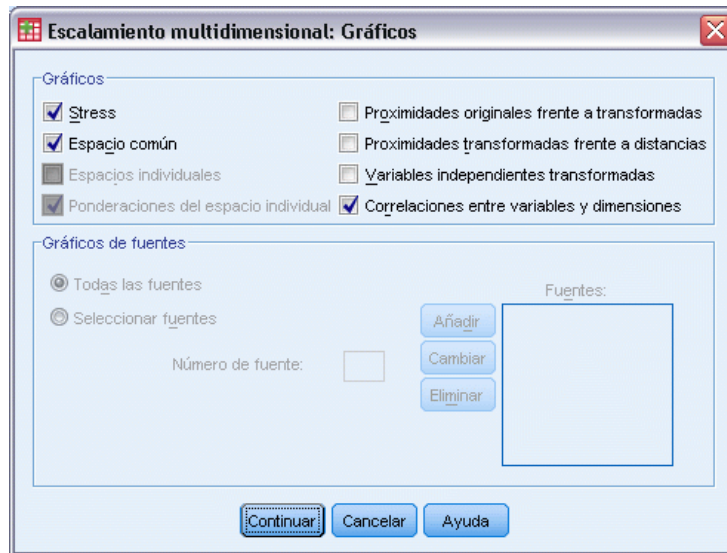
Observe que la variable *sexo* tiene un valor definido como perdido por el usuario, —9 = perdido (para primos). El procedimiento la trata como una categoría válida. Por lo tanto, no es probable que la transformación lineal por defecto sea adecuada. En su lugar, utilice una transformación nominal.

Figura 14-6
Cuadro de diálogo Restricciones



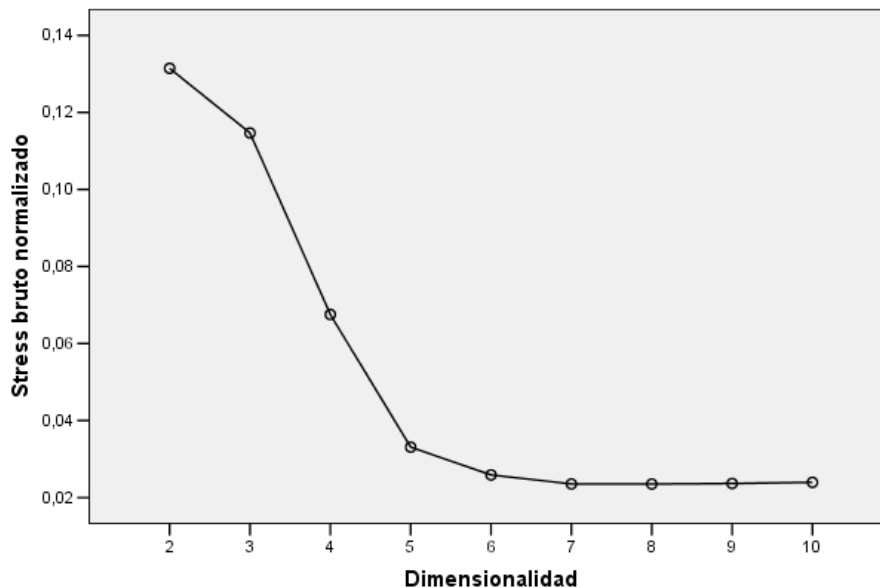
- ▶ Seleccione *sexo*.
- ▶ Seleccione Nominal de la lista desplegable de transformaciones de variables independientes.
- ▶ Pulse en Cambiar.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-7
Cuadro de diálogo Gráficos



- ▶ Seleccione Stress en el grupo Gráficos.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-8
Gráfico de sedimentación



El procedimiento comienza con una solución de 10 dimensiones y disminuye hasta una solución bidimensional. El gráfico de sedimentación muestra el stress bruto normalizado de la solución en cada dimensión. En el gráfico es posible observar que el incremento de la dimensionalidad de 2 a 3 y de 3 a 4 ofrece grandes mejoras en el stress. Después de 4, las mejoras son bastante pequeñas.

Elija el análisis de los datos mediante una solución tridimensional, ya que los resultados son más fáciles de interpretar.

Una solución tridimensional

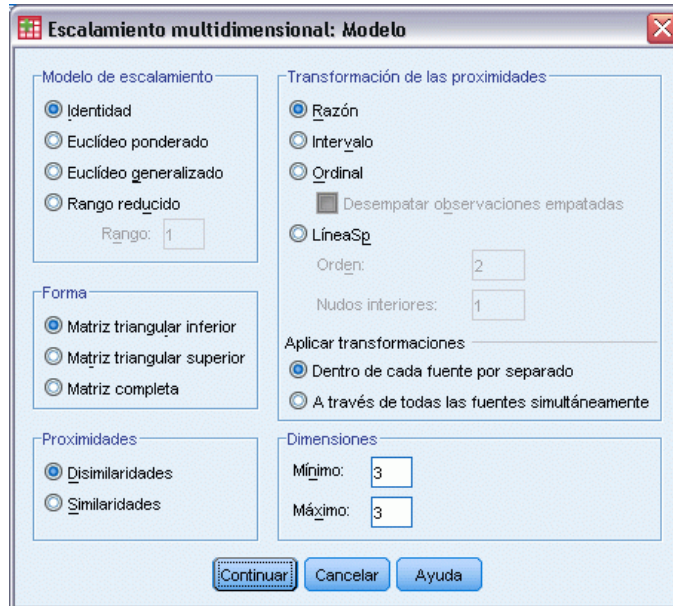
Las variables independientes *sexo*, *gener* (generación) y *grado* (de separación) se construyeron con la intención de ser utilizadas para interpretar las dimensiones de la solución. Las variables independientes se han construido de la siguiente manera:

<i>sexo</i>	1 = hombre, 2 = mujer, 9 = perdido (para primos)
<i>gener</i>	El número de generaciones de su familia, en caso de que el término se refiera a sus familiares, con los números menores correspondientes a las generaciones mayores. Por eso, los abuelos son -2, los nietos son 2 y los hermanos, 0.
<i>grado</i>	El número de grados de separación en su árbol familiar. Por eso, sus padres se encuentran 1 nodo hacia arriba, mientras que sus hijos están 1 nodo hacia abajo. Entre sus hermanos y usted existen 2 grados de separación, 1 nodo hacia arriba hacia sus padres y otro hacia abajo hacia sus hermanos. Sus primos se encuentran a 4 grados de distancia: 2 hacia arriba hasta llegar a sus abuelos y 2 hacia abajo pasando por sus tíos para llegar a sus primos.

Los variables externas se pueden encontrar en *kinship_var.sav*. Además, una configuración inicial procedente de un análisis anterior está disponible en *kinship_ini.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20.](#)

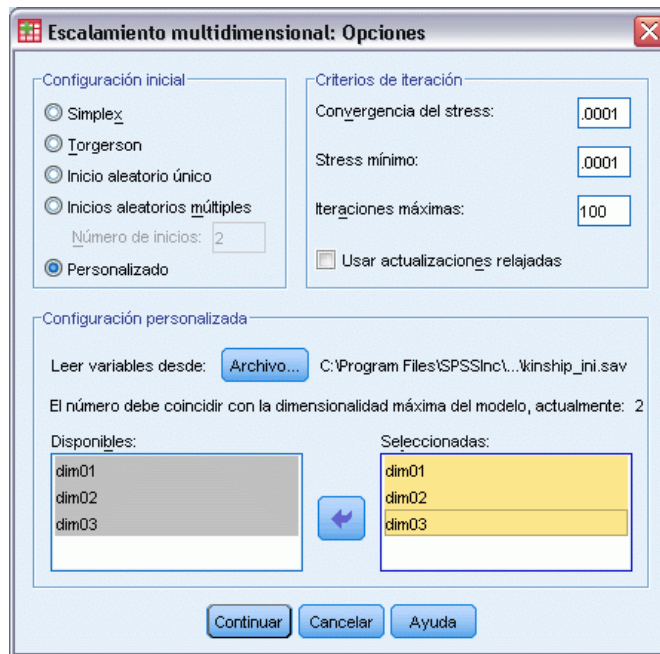
Ejecución del análisis

Figura 14-9
Cuadro de diálogo Modelo



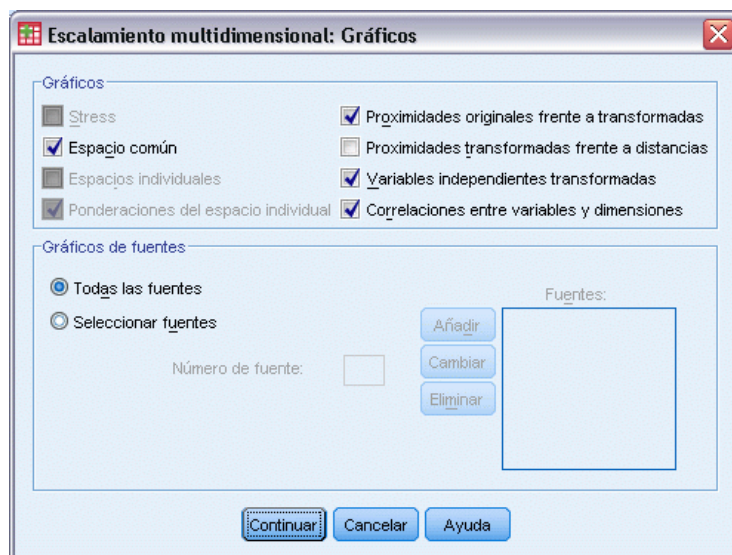
- ▶ Para obtener una solución tridimensional, vuelva al cuadro de diálogo Escalamiento multidimensional y pulse en Modelo.
- ▶ Introduzca 3 como el número máximo y mínimo de dimensiones.
- ▶ Pulse en Continuar.
- ▶ Pulse en Opciones en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-10
Cuadro de diálogo Opciones



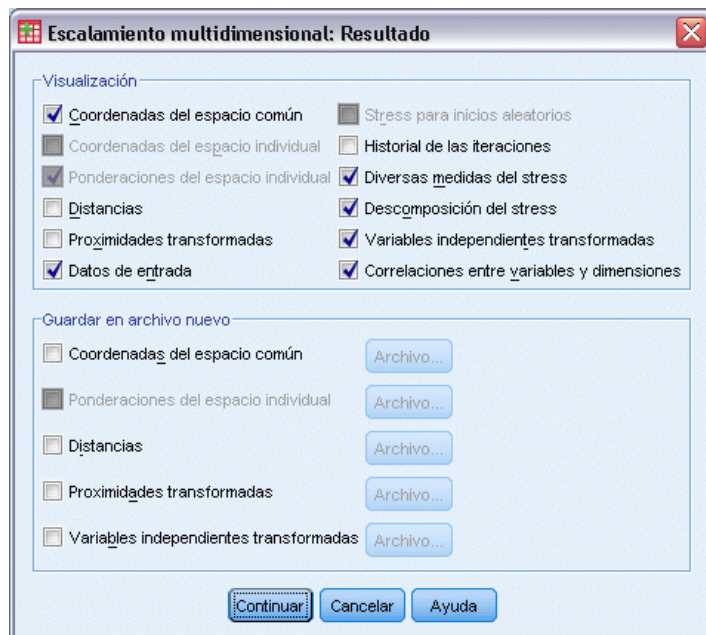
- ▶ Seleccione personalizada como la configuración inicial.
- ▶ Seleccione *kinship_ini.sav* como el archivo desde el que leer las variables.
- ▶ Seleccione *dim01*, *dim02* y *dim03* como variables.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-11
Cuadro de diálogo Gráficos



- ▶ Seleccione Proximidades originales frente a transformadas y Variables independientes transformadas.
- ▶ Pulse en Continuar.
- ▶ Pulse en Resultados en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-12
Cuadro de diálogo Resultados



- ▶ Seleccione Datos de entrada, Descomposición del stress y Correlaciones entre variables y dimensiones.

- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Escalamiento multidimensional.

Medidas del stress

El stress y las medidas de ajuste indican la eficacia con la que se aproximan las distancias de la solución a las distancias originales.

Figura 14-13
Stress y medidas de ajuste

Stress bruto normalizado	,06234
Stress-I	,24968 ^a
Stress-II	,87849 ^a
S-Stress	,14716 ^b
Dispersión explicada (D. A.F.)	,93766
Coefficiente de congruencia de Tucker	,96833

PROXSCAL minimiza el stress bruto normalizado.

a. Factor para escalamiento óptimo = 1,066.

b. Factor para escalamiento óptimo = ,984.

Cada uno de los cuatro estadísticos stress mide el desajuste de los datos, mientras que la dispersión explicada y el coeficiente de congruencia de Tucker miden el ajuste. Las medidas de stress inferior (hasta un mínimo de 0) y las mayores medidas de ajuste (hasta un máximo de 1) indican las mejores soluciones posibles.

Figura 14-14
Descomposición de stress bruto normalizado

		Fuente						Media
		SRC_1	SRC_2	SRC_3	SRC_4	SRC_5	SRC_6	
Objeto	Tía	,0991	,0754	,0629	,0468	,0391	,0489	,0620
	Hermano	,1351	,0974	,0496	,0813	,0613	,0597	,0807
	Primos	,0325	,0336	,0480	,0290	,0327	,0463	,0370
	Hija	,0700	,0370	,0516	,0229	,0326	,0207	,0391
	Padre	,0751	,0482	,0521	,0225	,0272	,0298	,0425
	Nieta	,1410	,0736	,0801	,0707	,0790	,0366	,0802
	Abuelo	,1549	,1057	,0858	,0821	,0851	,0576	,0952
	Abuela	,1550	,0979	,0858	,0844	,0816	,0627	,0946
	Nieto	,1374	,0772	,0793	,0719	,0791	,0382	,0805
	Madre	,0813	,0482	,0526	,0229	,0260	,0227	,0423
	Sobrino	,0843	,0619	,0580	,0375	,0317	,0273	,0501
	Sobrina	,0850	,0577	,0503	,0353	,0337	,0260	,0480
	Hermana	,1361	,0946	,0496	,0816	,0629	,0588	,0806
	Hijo	,0689	,0373	,0456	,0242	,0337	,0253	,0392
	Tío	,0977	,0761	,0678	,0489	,0383	,0498	,0631
Media		,1035	,0681	,0613	,0508	,0496	,0407	,0623

La descomposición del stress le ayuda a identificar qué fuentes y objetos contribuyen más al stress global de la solución. En este caso, la mayoría del stress entre las fuentes se puede atribuir a las fuentes 1 y 2, mientras que en los objetos, la mayoría del stress es atribuible a *hermano*, *nieta*, *abuelo*, *abuela*, *nieto* y *hermana*.

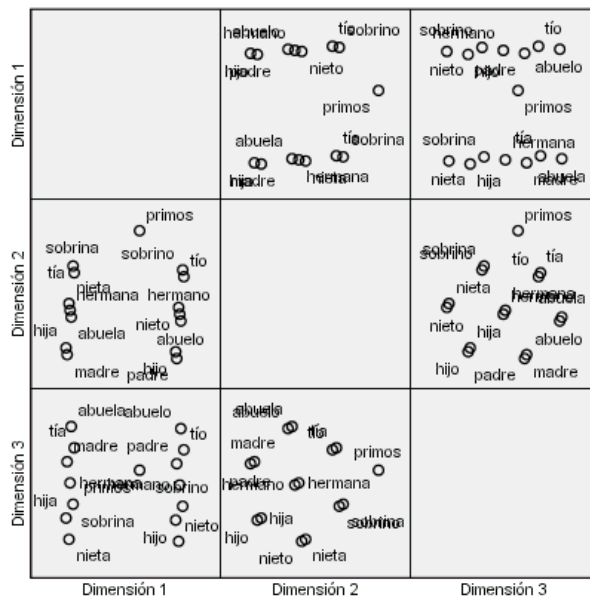
Las dos fuentes responsables de la mayoría del stress son los dos grupos que ordenaron los términos sólo una vez. Esta información sugiere que los estudiantes consideraron varios factores a la hora de clasificar los términos y aquéllos a los que se les permitió clasificar los términos dos veces se centraron en una parte de factores durante la primera clasificación y, a continuación, consideraron los factores restantes durante la segunda.

Los objetos responsables de la mayoría del stress son aquellos objetos con un *grado* de 2. Estas personas constituyen relaciones que no son parte de la familia “nuclear” (*madre, padre, hija, hijo*), pero que, sin embargo, están más cercanas que otras relaciones. Esta posición media podría causar fácilmente cierta clasificación diferencial de estos términos.

Coordenadas finales del espacio común

El gráfico del espacio común proporciona una representación visual de las relaciones entre los objetos.

Figura 14-15
Coordenadas del espacio común



Observe las coordenadas finales de los objetos en las dimensiones 1 y 3, es decir, el gráfico situado en la esquina inferior izquierda del diagrama de dispersión matricial. Este gráfico muestra que la dimensión 1 (en el eje *X*) está correlacionada con el *sexo* de la variable y la dimensión 3 (en el eje *y*) está correlacionada con *generación*. De izquierda a derecha, se observa que la dimensión 1 separa los términos masculinos y los femeninos, y el término *Primos* sin sexo en el centro. En el gráfico, desde abajo hacia arriba, los valores que aumentan a lo largo del eje se corresponden con los términos que son más antiguos.

Observe las coordenadas finales de los objetos en las dimensiones 2 y 3, es decir, este gráfico es el gráfico situado en la parte central derecha del diagrama de dispersión matricial. Desde este gráfico, se puede ver que la segunda dimensión (a lo largo del eje *Y*) se corresponde con el

grado de variable, con los mayores valores a lo largo del eje que corresponde con los términos más alejados de la llamada familia “nuclear”.

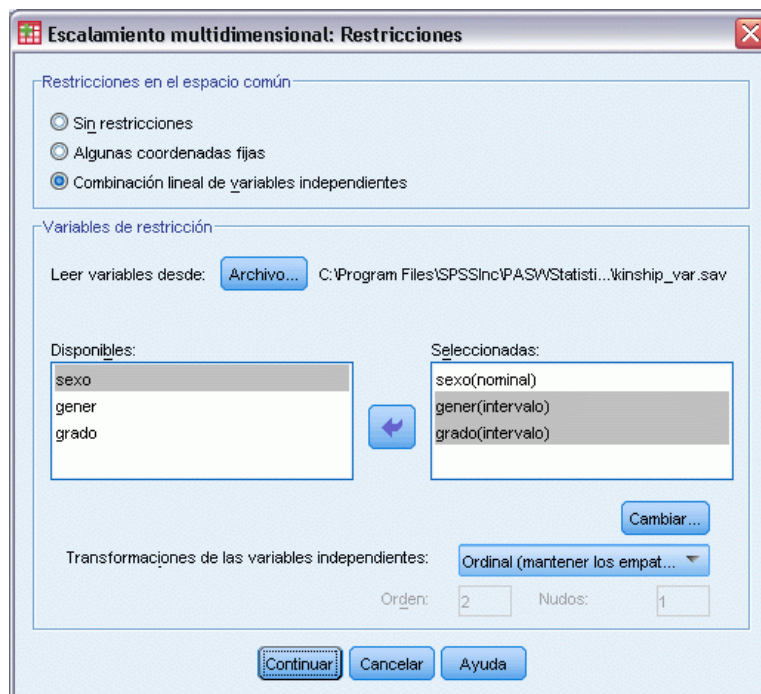
Una solución tridimensional con transformaciones que no son por defecto

La solución anterior se calculó mediante la transformación de la razón por defecto para las proximidades y las transformaciones del intervalo para las variables independientes *gener* y *grado*. Los resultados son bastante buenos, pero puede ser capaz de mejorarlos si utiliza otras transformaciones. Por ejemplo, las proximidades, la *gener* y el *grado* tienen clasificaciones naturales, pero es posible crear modelos mejores mediante una transformación ordinal, en lugar de una lineal.

Figura 14-16
Cuadro de diálogo Modelo

- ▶ Para ejecutar de nuevo el análisis, debe escalar las proximidades, la *gener* y el *grado* en un nivel ordinal (manteniendo los empates), vuelva al cuadro de diálogo Escalamiento multidimensional y pulse en Modelo:
- ▶ Seleccione Ordinal como la transformación de proximidad.
- ▶ Pulse en Continuar.
- ▶ Pulse en Restricciones en el cuadro de diálogo Escalamiento multidimensional.

Figura 14-17
Cuadro de diálogo Restricciones



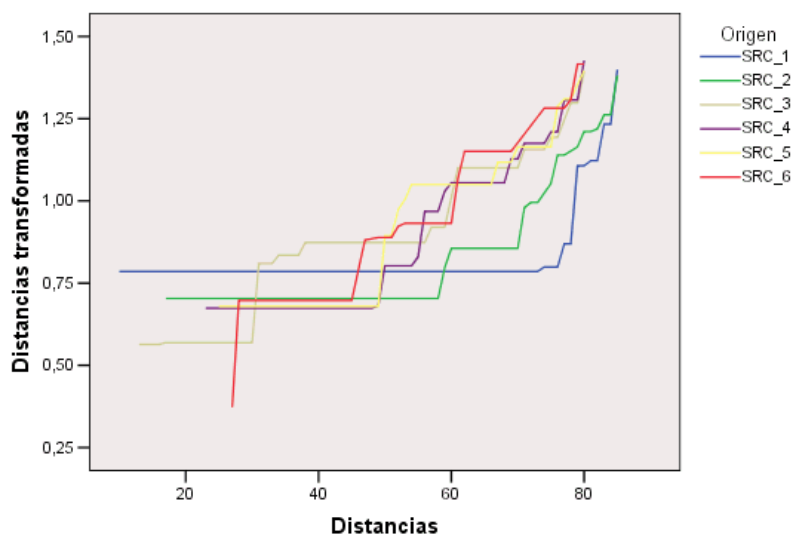
- ▶ Seleccione *gener* y *grado*.
- ▶ Seleccione Ordinal (mantenga los empates) de la lista desplegable de transformaciones de variables independientes.
- ▶ Pulse en Cambiar.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Escalamiento multidimensional.

Gráficos de transformación

Los gráficos de transformación son una buena herramienta para comprobar si las transformaciones originales eran apropiadas. Si los gráficos son aproximadamente lineales, el supuesto lineal es el adecuado. Si no es así, compruebe las medidas de stress para ver si hay una mejora en el ajuste y, asimismo, compruebe el gráfico del espacio común para ver si la interpretación es más útil.

Cada una de las variables independientes consiguen transformaciones prácticamente lineales, por lo que puede ser apropiado interpretarlas como numéricas. Sin embargo, las proximidades no consiguen una transformación lineal, por lo que es posible que la transformación ordinal sea más adecuada para las proximidades.

Figura 14-18
Proximidades transformadas



Medidas del stress

El stress para la solución actual apoya el argumento para el escalamiento de las proximidades en un nivel ordinal.

Figura 14-19
Stress y medidas de ajuste

Stress bruto normalizado	,03137
Stress-I	,17712 ^a
Stress-II	,61987 ^a
S-Stress	,07953 ^b
Dispersión explicada (D. A.F.)	,96863
Coefficiente de congruencia de Tucker	,98419

PROXSICAL minimiza el stress bruto normalizado.

a. Factor para escalamiento óptimo = 1,032.

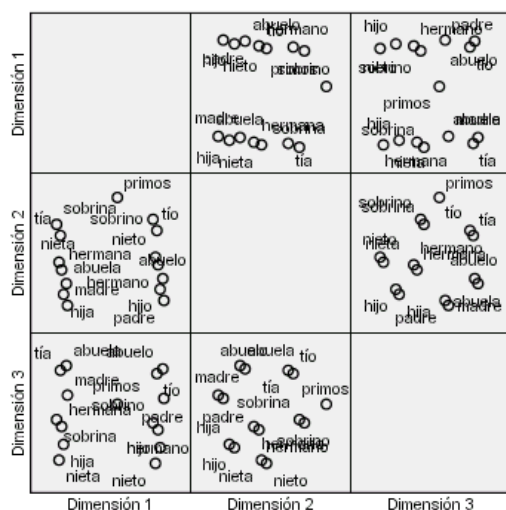
b. Factor para escalamiento óptimo = ,980.

El stress bruto normalizado para la solución anterior es 0,06234. Escalar las variables mediante transformaciones diferentes de las realizadas por defecto reduce el stress a 0,03137.

Coordenadas finales del espacio común

Los gráficos del espacio común ofrecen básicamente la misma interpretación de las dimensiones como la solución anterior.

Figura 14-20
Coordenadas del espacio común



Discusión

Lo mejor es tratar las proximidades como variables ordinales, ya que hay una gran mejora en las medidas del stress. Es posible que, a continuación, desee “desempatar” las variables ordinales, es decir, permitir los valores equivalentes de las variables originales para obtener valores transformados diferentes. Por ejemplo, en la primera fuente, las proximidades entre *tía* e *hijo*, así como *tía* y *nieta* son 85. El enfoque “empatado” de las variables ordinales obliga a los valores transformados de estas proximidades a ser equivalentes, pero no hay existe razón alguna para que usted suponga que deberían serlo. En este caso, el hecho de permitir que las proximidades desempaten le libra de restricciones innecesarias.

Lecturas recomendadas

Si desea obtener más información sobre el escalamiento multidimensional:

Commandeur, J. J. F., y W. J. Heiser. 1993. *Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices*. Leiden: Department of Data Theory, University of Leiden.

De Leeuw, J., y W. J. Heiser. 1980. Multidimensional scaling with restrictions on the configuration. En: *Multivariate Analysis, Vol. V*, P. R. Krishnaiah, ed. Ámsterdam: North-Holland.

Heiser, W. J. 1981. *Unfolding analysis of proximity data*. Leiden: Department of Data Theory, University of Leiden.

Heiser, W. J., y F. M. T. A. Busing. 2004. Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. En: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, .

Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, .

Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function I. *Psychometrika*, 27, .

Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika*, 27, .

desplegamiento multidimensional

El procedimiento Desplegamiento multidimensional intenta buscar una escala cuantitativa común que permita examinar visualmente las relaciones entre dos conjuntos de objetos.

Ejemplo: Preferencias del elemento de desayuno

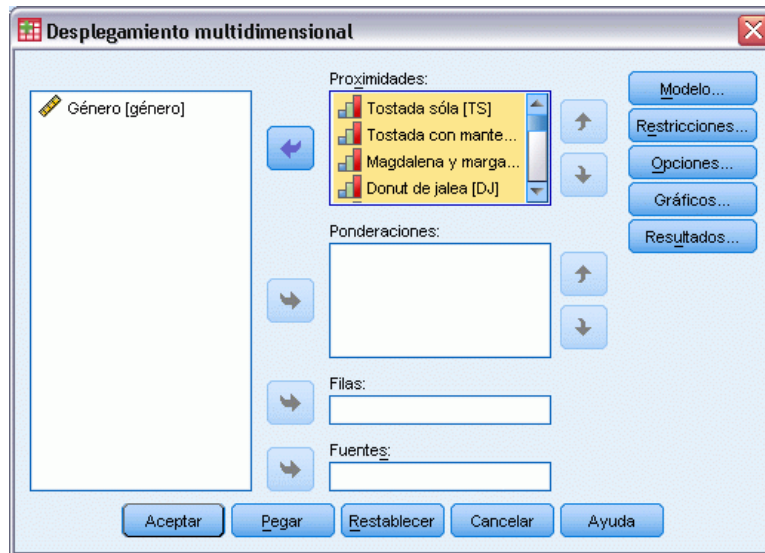
En un estudio clásico (Green y Rao, 1972), se pidió a 21 estudiantes de administración de empresas de la Wharton School y sus cónyuges que ordenaran 15 elementos de desayuno por orden de preferencia, de 1 = “más preferido” a 15 = “menos preferido”. Esta información se recoge en el archivo *breakfast_overall.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20.](#)

Los resultados del estudio proporcionan un ejemplo típico del problema de degeneración inherente a la mayoría de los algoritmos de desplegado multidimensional que se resuelven al penalizar el coeficiente de variación de las proximidades transformadas (Busing, Groenen, y Heiser, 2005). Verá una solución degenerada y cómo solucionar el problema mediante el procedimiento Desplegado multidimensional, lo que permite determinar la manera en la que los individuos discriminan entre los elementos de desayuno. Puede encontrar la sintaxis para reproducir estos análisis en *prefscal_breakfast-overall.sps*.

Creación de una solución degenerada

- ▶ Para ejecutar un análisis de desplegado multidimensional, elija en los menús:
Analizar > Escala > Desplegado multidimensional (PREFSCAL)...

Figura 15-1
Cuadro de diálogo principal Despliegamiento multidimensional



- ▶ Seleccione desde *Tostada sola* hasta *Bollo de maíz y mantequilla* como variables de proximidad.
- ▶ Pulse en Opciones.

Figura 15-2
Cuadro de diálogo Opciones

The dialog box 'Desplegamiento multidimensional: Opciones' contains the following elements:

- Configuración inicial:** Radio buttons for 'Clásica' (selected), 'Ross-Cliff', 'Correspondencia', 'Centroides', 'Inicios aleatorios', and 'Personalizado'. A dropdown menu for 'Imputación mediante:' is set to 'Spearman'. Input fields for 'Opciones:' and 'Número de inicios:' are both set to '1'.
- Criterios de iteración:** Input fields for 'Convergencia del stress:' (0.000001), 'Stress mínimo:' (0.0001), and 'Iteraciones máximas:' (5000).
- Término de penalización:** Input fields for 'Magnitud:' (1.0) and 'Rango:' (0.0).
- Configuración personalizada:** A button 'Leer variables desde: Archivo...', a note 'El número debe coincidir con la dimensionalidad máxima del modelo, actualmente: 2', and a note 'Las variables que contienen coordenadas de fila deben preceder a las que contienen coordenadas de columna.' Below these are two lists: 'Disponibles:' and 'Seleccionadas:', with an arrow button between them.
- Buttons:** 'Continuar', 'Cancelar', and 'Ayuda' at the bottom.

- ▶ Seleccione Spearman como método de imputación para el inicio clásico.
- ▶ En el grupo Término de penalización, escriba 1,0 como el valor del parámetro de magnitud y 0,0 como el valor del parámetro de rango. Esto desactiva el término de penalización.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Desplegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TMd BTJ TMn CB DP GD CC CMB
/INITIAL=CLASSICAL (SPEARMAN)
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(1.0) OMEGA(0.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON .
```

- Esta sintaxis especifica un análisis de las variables *ts* (*Tostada sola*) hasta *mb* (*Bollo de maíz y mantequilla*).

- El subcomando `INITIAL` especifica que los valores iniciales se imputan mediante distancias de Spearman.
- Básicamente, los valores especificados en el subcomando `PENALTY` desactivan el término de penalización y, como resultado, el procedimiento minimiza el Stress-I de Kruskal, lo que generará una solución degenerada.
- El subcomando `PLOT` requiere gráficos del espacio común.
- El resto de parámetros vuelven a sus valores por defecto.

Medidas

Figura 15-3
Medidas para la solución degenerada

Iteraciones		154
Valor de función final		,0000990
Partes del valor de función	Parte de Stress	,0000990
	Parte de penalización	1,0000000
Maldad de ajuste	Stress normalizado	,0000000
	Stress-I de Kruskal	,0000990
	Stress-II de Kruskal	,6129749
	S-Stress-I de Young	,0001980
	S-Stress-II de Young	,7703817
Bondad de ajuste	Dispersión explicada	1,0000000
	Varianza explicada	,6230788
	Órdenes de preferencia recuperados	,7074830
	Rho de Spearman	,7450748
	Tau-b de Kendall	,6218729
Coeficientes de variación	Variación de las Proximidades	,5590170
	Variación de las Proximidades transformadas	,0000924
	Variación de las Distancias	,1808765
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	117,3115413
	Índice de no-degeneración	,0000000
	aproximada de Shepard	

El algoritmo converge a una solución después de 154 iteraciones, con una tensión penalizada (valor de función final marcado) de 0,0000990. Como el término de penalización se ha desactivado, la tensión penalizada será igual al Stress-I de Kruskal (la parte de tensión del valor de la función es equivalente a la medida de la maldad de ajuste de Kruskal). Los valores de stress

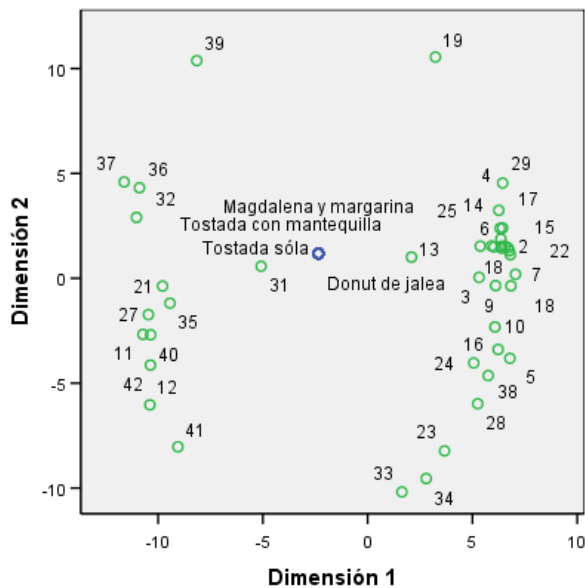
bajo suelen indicar que la solución se ajusta adecuadamente a los datos, pero existen algunos signos de advertencia de una solución degenerada:

- El coeficiente de variación de las proximidades transformadas es muy pequeño respecto al coeficiente de variación de las proximidades originales. Esto sugiere que las proximidades transformadas para cada fila son aproximadamente constantes y que, por lo tanto, la solución no proporcionará ninguna discriminación entre objetos.
- La suma de cuadrados de los índices de entremezclado de DeSarbo es una medida de lo bien que se entremezclan los puntos de los distintos conjuntos. Si no se entremezclan, es señal de que la solución puede ser degenerada. Cuanto más cerca de 0, más entremezclada es la solución. El valor obtenido es muy alto, lo cual indica que la solución no es entremezclada.
- El índice de no degeneración aproximada de Shepard, que se expresa como un porcentaje de distintas distancias, es igual a 0. Ésta es una indicación numérica clara de que no existen suficientes distancias distintas y de que la solución probablemente sea degenerada.

Espacio común

Figura 15-4

Gráfico conjunto del espacio común para la solución degenerada



La confirmación visual de que la solución es degenerada se encuentra en el gráfico conjunto del espacio común de objetos de fila y columna. Los objetos de fila (individuos) se sitúan en la circunferencia de un círculo centrado en los objetos de columna (elementos de desayuno), cuyas coordenadas se han colapsado en un único punto.

Ejecución de un análisis no degenerado

Figura 15-5
Cuadro de diálogo Opciones

- ▶ Para generar una solución no degenerada, pulse en la herramienta Recuperar cuadros de diálogo y seleccione Desplegamiento multidimensional.
- ▶ Pulse en Opciones en el cuadro de diálogo Desplegamiento multidimensional.
- ▶ En el grupo Término de penalización, escriba 0,5 como el valor del parámetro de magnitud y 1,0 como el valor del parámetro de rango. Esto desactiva el término de penalización.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Desplegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TmD BTJ TMn CB DP GD CC CMB
/INITIAL=CLASSICAL (SPEARMAN)
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON .
```

- El único cambio se encuentra en el subcomando `PENALTY.LAMBDA` se ha establecido en 0,5, y `OMEGA` se ha establecido en 1,0, sus valores por defecto.

Medidas

Figura 15-6
Medidas para la solución no degenerada

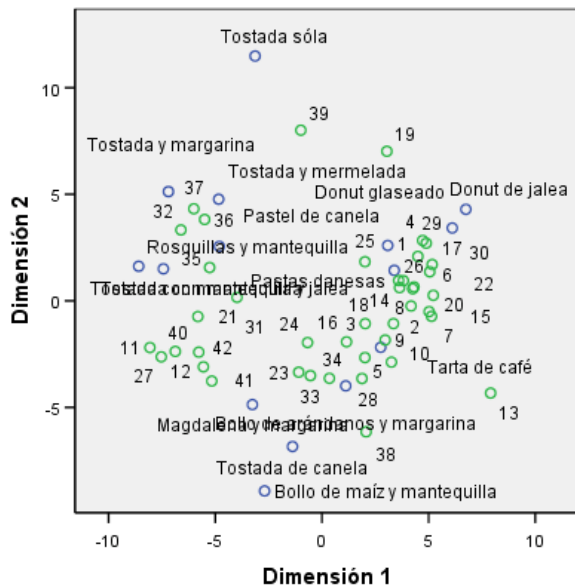
Iteraciones		157
Valor de función final		,6848930
Partes del valor de función	Parte de Stress	,2428268
	Parte de penalización	1,9317409
Maldad de ajuste	Stress normalizado	,0583589
	Stress-I de Kruskal	,2415758
	Stress-II de Kruskal	,5875599
	S-Stress-I de Young	,3446361
	S-Stress-II de Young	,5030127
Bondad de ajuste	Dispersión explicada	,9416411
	Varianza explicada	,7651552
	Órdenes de preferencia recuperados	,7818594
	Rho de Spearman	,8179181
	Tau-b de Kendall	,6916725
Coeficientes de variación	Variación de las Proximidades	,5590170
	Variación de las Proximidades transformadas	,6006156
	Variación de las Distancias	,4833617
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	,1590979
	Índice de no-degeneración aproximada de Shepard	,7895692

Los problemas detectados en las medidas para la solución degenerada se han corregido aquí.

- El stress normalizado ya no es 0.
- El coeficiente de variación de las proximidades transformadas ahora tiene un valor similar al coeficiente de variación de las proximidades originales.
- Los índices de entremezclado de DeSarbo están mucho más cerca del 0, lo que indica que la solución está mucho mejor entremezclada.
- El índice de no-degeneración aproximada de Shepard, que se expresa como un porcentaje de distintas distancias, ahora es casi del 80%. Existen suficientes distancias distintas y es probable que la solución sea no degenerada.

Espacio común

Figura 15-7
Gráfico conjunto del espacio común para la solución no degenerada



El gráfico conjunto del espacio común permite interpretar de las dimensiones. La dimensión horizontal parece discriminar entre pan o tostada blando o duro, con elementos más blandos a medida que se desplaza hacia la derecha del eje. La dimensión vertical no tiene una interpretación clara, aunque tal vez discrimine en función de la comodidad, con elementos más “formales” a medida que desciende por el eje.

Esto crea varios conglomerados de elementos de desayuno. Por ejemplo, los donuts, los pasteles de canela y las pastas danesas forman un conglomerado de elementos blandos y algo informales. Las magdalenas y la tostada de canela forman un conglomerado de elementos más duros pero más formales. El resto de tostadas y rosquillas forman un conglomerado de elementos duros y algo informales. La tostada sola es un elemento duro extremadamente informal.

Los individuos representados por los objetos de fila se dividen claramente en conglomerados según la preferencia por los elementos duros o blandos, con una variación considerable dentro de los conglomerados en la dimensión vertical.

Ejemplo: Desplegamiento triple de las preferencias de los elementos del desayuno

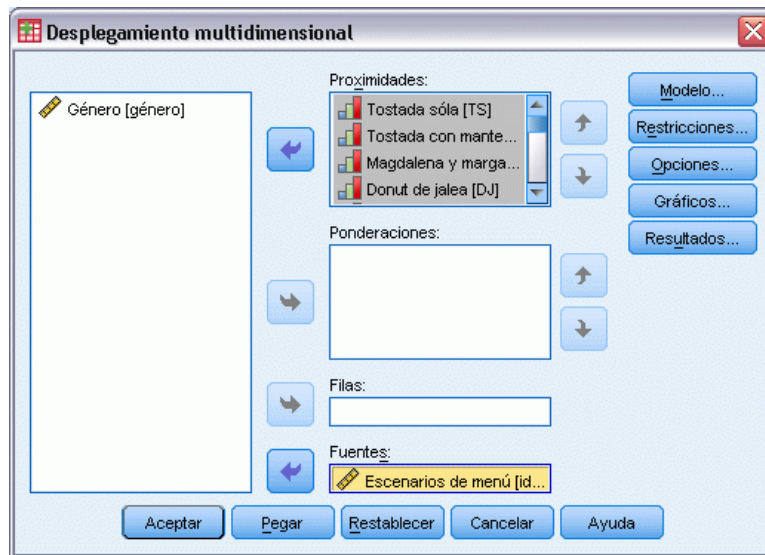
En un estudio clásico (Green et al., 1972), se pidió a 21 estudiantes de administración de empresas de la Wharton School y sus cónyuges que ordenaran 15 elementos de desayuno por orden de preferencia, de 1 = “más preferido” a 15 = “menos preferido”. Sus preferencias se registraron en seis escenarios distintos, de “Preferencia global” a “Aperitivo, con bebida sólo”. Esta información se recoge en el archivo *breakfast.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en IBM SPSS Categories 20.](#)

Los seis escenarios pueden tratarse como orígenes independientes. Utilice PREFSCAL para realizar un despliegamiento triple de las filas, columnas y orígenes. Puede encontrar la sintaxis para reproducir estos análisis en *prefscal_breakfast.sps*.

Ejecución del análisis

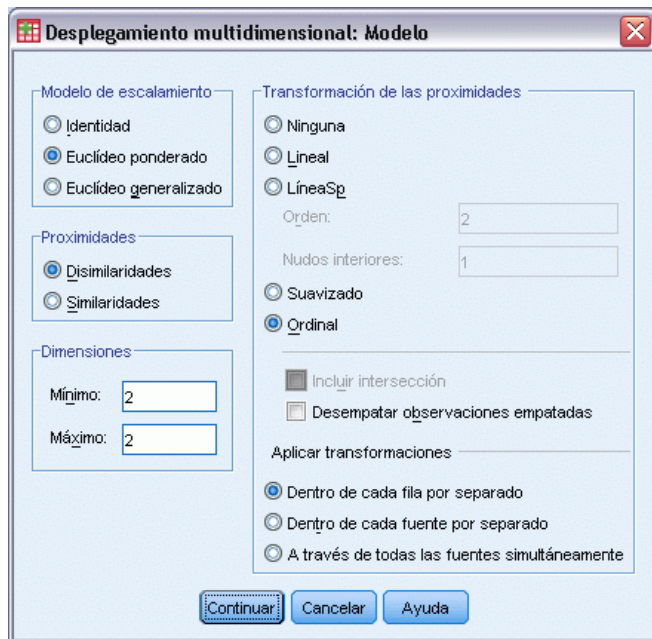
- ▶ Para ejecutar un análisis de despliegamiento multidimensional, elija en los menús: Analizar > Escala > Despliegamiento multidimensional (PREFSCAL)...

Figura 15-8
Cuadro de diálogo principal Despliegamiento multidimensional



- ▶ Seleccione desde *Tostada sola* hasta *Bollo de maíz y mantequilla* como variables de proximidad.
- ▶ Seleccione *Escenarios de menú* como la variable de fuente.
- ▶ Pulse en *Modelo*.

Figura 15-9
Cuadro de diálogo Modelo



- ▶ Seleccione Euclídeo ponderado como el modelo de escalamiento.
- ▶ Pulse en Continuar.
- ▶ Pulse en Opciones en el cuadro de diálogo Despliegamiento multidimensional.

Figura 15-10
Cuadro de diálogo Opciones

Desplegamiento multidimensional: Opciones

Configuración inicial

Clásica
Imputación mediante: Spearman

Ross-Cliff

Correspondencia

Centroides
Opciones: 1

Inicios aleatorios
Número de inicios: 1

Personalizado

Criterios de iteración

Convergencia del stress: .000001

Stress mínimo: .0001

Iteraciones máximas: 5000

Término de penalización

Magnitud: 0.5

Rango: 1.0

Configuración personalizada

Leer variables desde: Archivo...

El número debe coincidir con la dimensionalidad máxima del modelo, actualmente: 2

Las variables que contienen coordenadas de fila deben preceder a las que contienen coordenadas de columna.

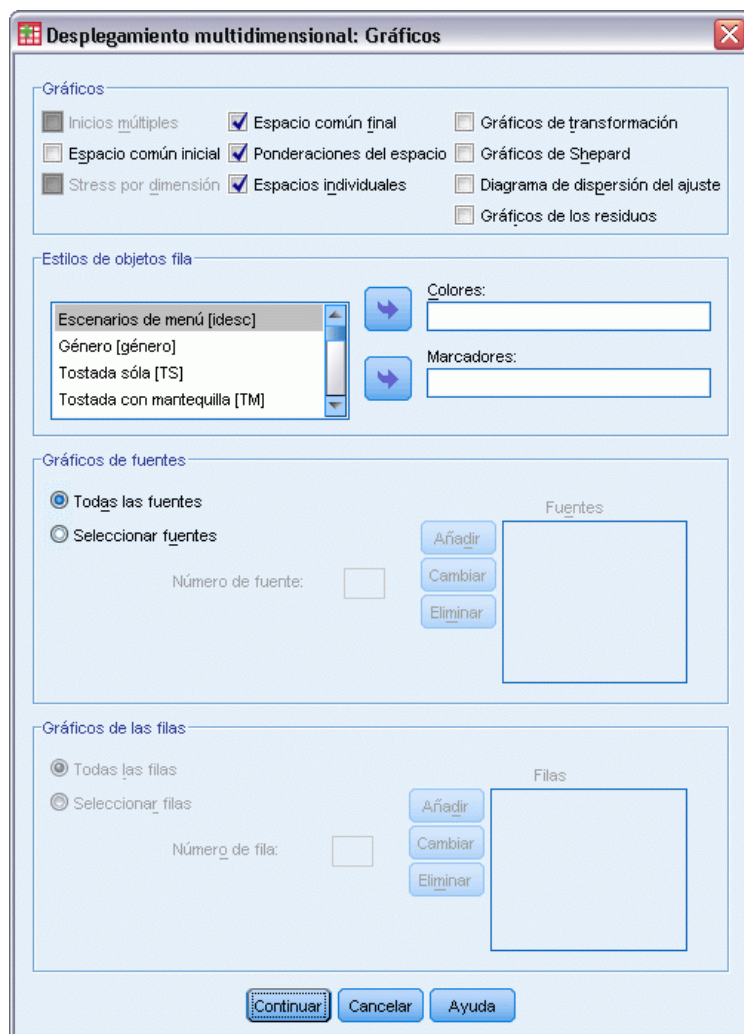
Disponibles: []

Seleccionadas: []

[Continuar] [Cancelar] [Ayuda]

- ▶ Seleccione Spearman como método de imputación para el inicio clásico.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Desplegamiento multidimensional.

Figura 15-11
Cuadro de diálogo Gráficos



- ▶ Seleccione Espacios individuales en el grupo Gráficos.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Despliegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TmD BTJ TMn CB DP GD CC CMB
/INPUT=SOURCES (srcid )
/INITIAL=CLASSICAL (SPEARMAN)
/CONDITION=ROW
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/MODEL=WEIGHTED
/CRITERIA=DIMENSIONS (2,2) DIFFSTRESS (.000001) MINSTRESS (.0001)
MAXITER (5000)
```

```

/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON WEIGHTS INDIVIDUAL ( ALL ) .

```

- Esta sintaxis especifica un análisis de las variables *ts* (*Tostada sola*) hasta *mb* (*Bollo de maíz y mantequilla*). La variable *idsc* se utiliza para identificar los orígenes.
- El subcomando `INITIAL` especifica que los valores iniciales se imputan mediante distancias de Spearman.
- El subcomando `MODEL` especifica un modelo euclídeo ponderado, que permite que cada espacio individual pondere las dimensiones del espacio común de manera diferente.
- El subcomando `PLOT` solicita gráficos del espacio común, los espacios individuales y las ponderaciones de los espacios individuales.
- El resto de parámetros vuelven a sus valores por defecto.

Medidas

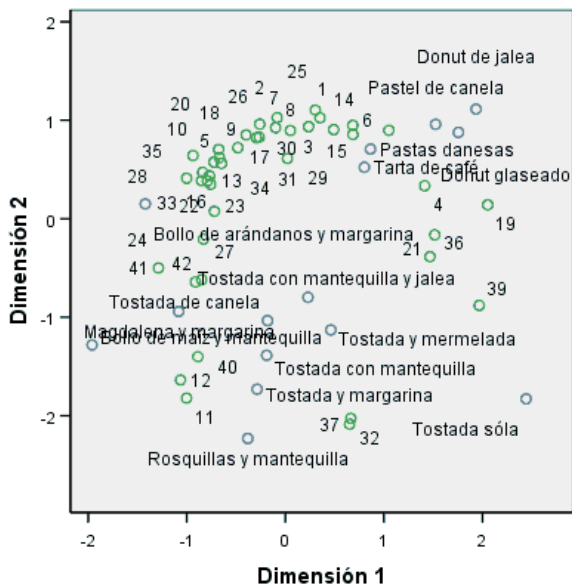
Figura 15-12
Medidas

Iteraciones		481
Valor de función final		,8199642
Partes del valor de función	Parte de Stress	,3680994
	Parte de penalización	1,8265211
Maldad de ajuste	Stress normalizado	,1335343
	Stress-I de Kruskal	,3654234
	Stress-II de Kruskal	,9780824
	S-Stress-I de Young	,4938016
	S-Stress-II de Young	,6912352
Bondad de ajuste	Dispersión explicada	,8664657
	Varianza explicada	,5024853
	Órdenes de preferencia recuperados	,7025321
	Rho de Spearman	,6271702
	Tau-b de Kendall	,4991188
Coeficientes de variación	Variación de las Proximidades	,5590170
	Variación de las Proximidades transformadas	,6378878
	Variación de las Distancias	,4484515
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	,2199287
	Índice de no-degeneración aproximada de Shepard	,7643613

El algoritmo converge después de 481 iteraciones, con una tensión final penalizada de 0,8199642. Los coeficientes de variación y el índice de Shepard son suficientemente altos y los índices de DeSarbo son suficientemente bajos como para sugerir que no existen problemas de degeneración.

Espacio común

Figura 15-13
Gráfico conjunto del espacio común



El gráfico conjunto del espacio común muestra una configuración final muy similar al análisis de dos factores en las preferencias generales, con la solución volteada sobre la línea de 45 grados. Así, la dimensión vertical parece discriminar entre pan o tostada blando o duro, con elementos más blandos a medida que sube por el eje. La dimensión horizontal ahora no tiene una interpretación clara, aunque tal vez discrimine en función de la comodidad, con elementos más “formales” a medida que se desplaza hacia la izquierda del eje.

Los individuos representados por los objetos de fila siguen estando claramente divididos en conglomerados según la preferencia por los elementos “duros” o “blandos”, con una variación considerable dentro de los conglomerados en la dimensión horizontal.

Espacios individuales

Figura 15-14
Ponderaciones de dimensión

		Dimensión		Especificidad ^a
		1	2	
Origen	Preferencia global	3,235	4,297	,186
	Desayuno, con zumo, huevos con bacón, y bebida	4,883	2,193	,457
	Desayuno, con zumo, cereales fríos, y bebida	4,131	3,438	,109
	Desayuno, con zumo, tortitas, salchichas, y bebida	4,291	3,267	,164
	Desayuno, con bebida sólo	3,124	4,413	,223
	Aperitivo, con bebida sólo	2,750	4,541	,313
Importancia ^b		,504	,496	

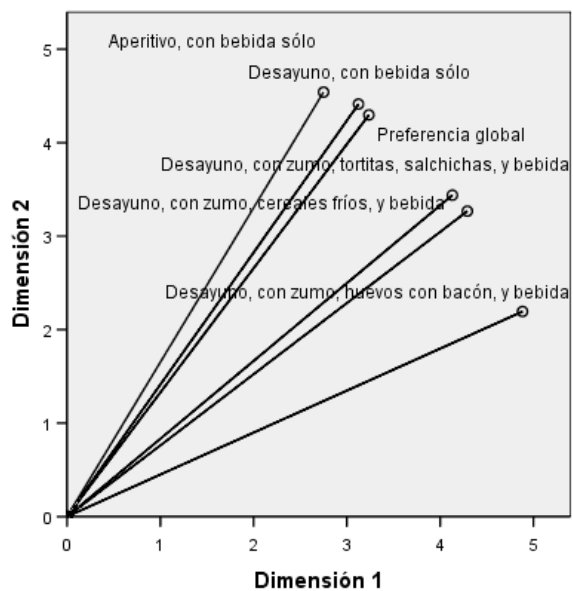
a. La especificidad indica lo habitual de un origen. El intervalo de especificidad oscila entre cero y uno, donde un cero indica un origen medio con idénticas ponderaciones de dimensión, y uno corresponde a un origen muy específico con una ponderación de dimensión grande y excepcional y las demás ponderaciones son prácticamente iguales a cero.

b. Importancia relativa de cada dimensión, indicada como la razón de la suma de cuadrados de una dimensión entre la suma de cuadrados total.

Se calcula un espacio individual para cada fuente. Las ponderaciones de dimensión muestran cómo se cargan los espacios individuales en las dimensiones del espacio común. Una ponderación mayor indica una distancia mayor en el espacio individual y, por lo tanto, una mayor discriminación entre los objetos de esa dimensión para ese espacio individual.

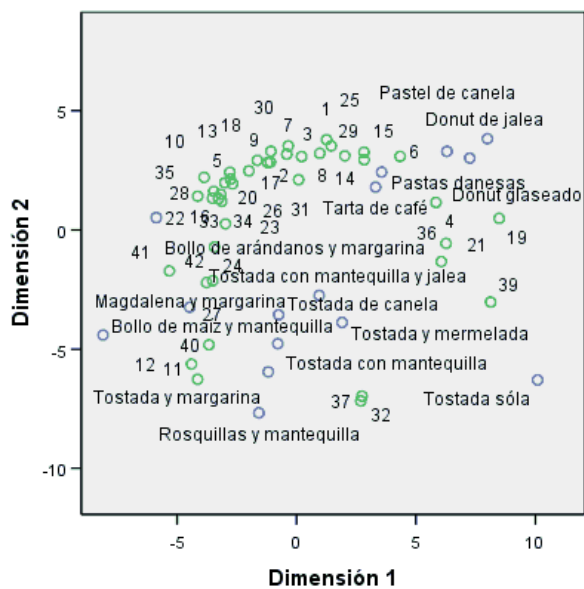
- La **especificidad** es una medida que indica hasta qué punto un espacio individual es distinto del espacio común. Un espacio individual que fuera idéntico al espacio común tendría ponderaciones de dimensión idénticas y una especificidad de 0, mientras que un espacio individual que fuera específico de una dimensión determinada tendría una única ponderación de dimensión grande y una especificidad de 1. En este caso, los orígenes más divergentes son *Desayuno, con zumo, huevos con bacón, y bebida* y *Aperitivo, con bebida sólo*.
- La **importancia** es una medida de la contribución relativa de cada dimensión a la solución. En este caso, las dimensiones son igual de importantes.

Figura 15-15
Ponderaciones de dimensión



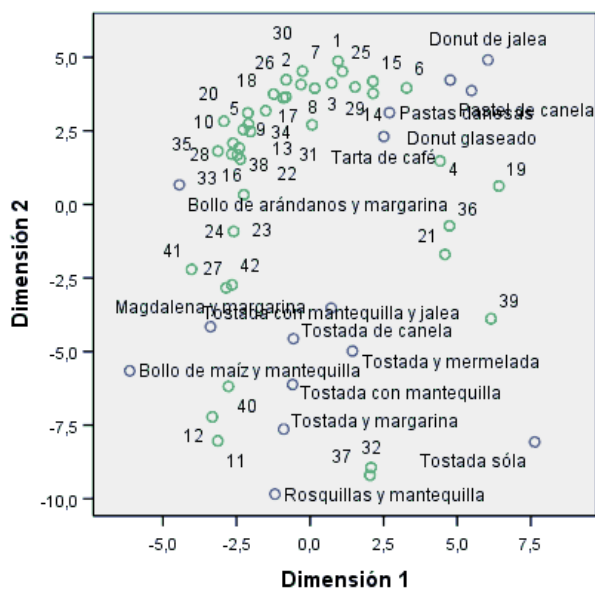
El gráfico de ponderaciones de dimensión proporciona una visualización de la tabla de ponderaciones. *Desayuno, con zumo, huevos con bacón, y bebida* y *Aperitivo, con bebida sólo* están más cerca de los ejes de dimensión, pero ninguno es especialmente específico de una dimensión determinada.

Figura 15-16
Gráfico conjunto del espacio individual "Desayuno, con zumo, huevos con bacón, y bebida"



El gráfico conjunto del espacio individual *Desayuno, con zumo, huevos con bacón, y bebida* muestra el efecto de este escenario sobre las preferencias. Esta fuente se carga sobre todo en la primera dimensión, por lo que la diferenciación entre los elementos se debe fundamentalmente a la primera dimensión.

Figura 15-17
Gráfico conjunto del espacio individual "Aperitivo, con bebida sólo"



El gráfico conjunto del espacio individual *Aperitivo, con bebida sólo* muestra el efecto de este escenario sobre las preferencias. Esta fuente se carga sobre todo en la segunda dimensión, por lo que la diferenciación entre los elementos se debe fundamentalmente a la segunda dimensión. Sin embargo, existe todavía una ligera diferenciación en la primera dimensión debido a la especificidad bastante baja de esta fuente.

Uso de una configuración inicial diferente

La configuración final puede depender de los puntos de inicio dados al algoritmo. Lo ideal es que la estructura general de la solución permanezca intacta; de lo contrario, puede ser difícil determinar cuál es la correcta. Sin embargo, es posible centrarse más en los detalles probando distintas configuraciones iniciales, como utilizar un inicio de correspondencia en el análisis triple de los datos de desayuno.

- Para generar una solución con un inicio de correspondencia, pulse en la herramienta Recuperar cuadros de diálogo y seleccione Despliegamiento multidimensional.

- Pulse en Opciones en el cuadro de diálogo Despliegamiento multidimensional.

Figura 15-18
Cuadro de diálogo Opciones

Despliegamiento multidimensional: Opciones

Configuración inicial

Clásica
Imputación mediante: Spearman

Ross-Cliff

Correspondencia

Centroides
Opciones: 1

Inicios aleatorios
Número de inicios: 1

Personalizado

Criterios de iteración

Convergencia del stress: .000001

Stress mínimo: .0001

Iteraciones máximas: 5000

Término de penalización

Magnitud: 0.5

Rango: 1.0

Configuración personalizada

Leer variables desde: Archivo...

El número debe coincidir con la dimensionalidad máxima del modelo, actualmente: 2

Las variables que contienen coordenadas de fila deben preceder a las que contienen coordenadas de columna.

Disponibles: []

Seleccionadas: []

Continuar Cancelar Ayuda

- Seleccione Correspondencia en el grupo Configuración inicial.
- Pulse en Continuar.
- Pulse en Aceptar en el cuadro de diálogo Despliegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=TP BT EMM JD CT BMM HRB TMd BTJ TMn CB DP GD CC CMB
/INPUT=SOURCES(srcid )
/INITIAL=CORRESPONDENCE
/TRANSFORMATION=NONE
/PROXIMITIES=DISSIMILARITIES
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)
MAXITER(5000)
/PENALTY=LAMBDA(0.5) OMEGA(1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON WEIGHTS INDIVIDUAL ( ALL ) .
```

- El único cambio se encuentra en el subcomando `INITIAL`. La configuración inicial se ha establecido en `CORRESPONDENCE`, que utiliza los resultados de un análisis de correspondencias de los datos reservados (similitudes en lugar de disimilaridades), con una normalización simétrica de puntuaciones de fila y de columna.

Medidas

Figura 15-19
Medidas para la configuración inicial de la correspondencia

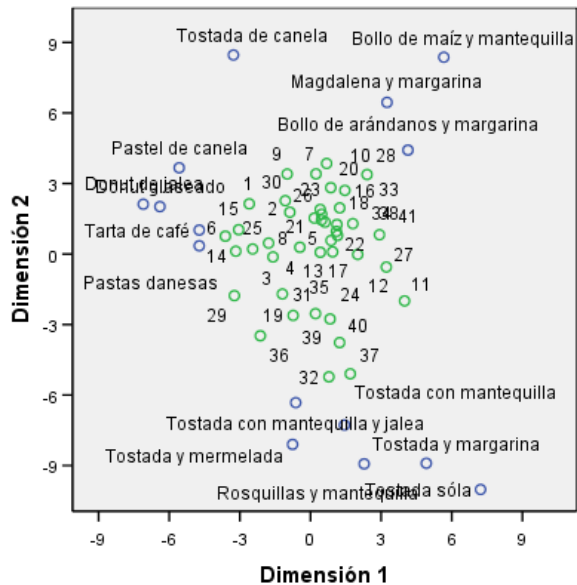
Iteraciones		385
Valor de función final		,8140741
Partes del valor de función	Parte de Stress	,3493640
	Parte de penalización	1,8969229
Maldad de ajuste	Stress normalizado	,1212145
	Stress-I de Kruskal	,3481587
	Stress-II de Kruskal	1,0770522
	S-Stress-I de Young	,4812632
	S-Stress-II de Young	,6871733
Bondad de ajuste	Dispersión explicada	,8787855
	Varianza explicada	,5183498
	Órdenes de preferencia recuperados	,7174981
	Rho de Spearman	,6446272
Coeficientes de variación	Tau-b de Kendall	,5165230
	Variación de las Proximidades	,5590170
	Variación de las Proximidades transformadas	,6122308
	Variación de las Distancias	,4043695
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	1,7571887
	Índice de no-degeneración aproximada de Shepard	,7532124

El algoritmo converge después de 385 iteraciones, con una tensión final penalizada de 0,8140741. Este estadístico, la maldad de ajuste, la bondad de ajuste, los coeficientes de variación y el índice de Shepard son muy similares a los de la solución que utiliza el inicio clásico de Spearman. Los índices de DeSarbo son algo diferentes, con un valor de 1,7571887 en lugar de 0,2199287, lo cual sugiere que la solución que utiliza el inicio de correspondencia no está igual de bien mezclada. Para ver cómo afecta esto a la solución, consulte el gráfico conjunto del espacio común.

Espacio común

Figura 15-20

Gráfico conjunto del espacio común para la configuración inicial de la correspondencia



El gráfico conjunto del espacio común muestra una configuración final similar al análisis con la configuración inicial de Spearman clásica; sin embargo, los objetos de columna (elementos de desayuno) se sitúan alrededor de los objetos de fila (individuos) en lugar de entremezclarse con ellos.

Espacios individuales

Figura 15-21

Ponderaciones de dimensión para la configuración inicial de la correspondencia

		Dimensión		Especificidad ^a
		1	2	
Origen	Preferencia global	2,836	3,877	,279
	Desayuno, con zumo, huevos con bacón, y bebida	4,727	1,207	,636
	Desayuno, con zumo, cereales fríos, y bebida	4,183	2,377	,263
	Desayuno, con zumo, tortitas, salchichas, y bebida	4,412	1,993	,389
	Desayuno, con bebida sólo	2,605	4,050	,351
	Aperitivo, con bebida sólo	1,864	4,415	,552
Importancia ^b		,556	,444	

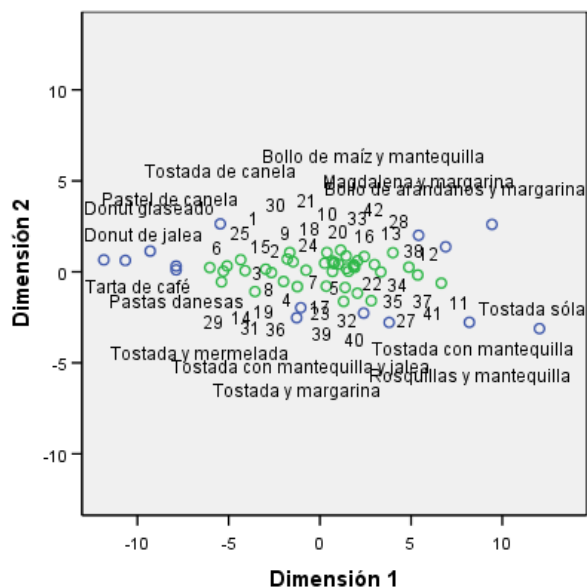
a. La especificidad indica lo habitual de un origen. El intervalo de especificidad oscila entre cero y uno, donde un cero indica un origen medio con idénticas ponderaciones de dimensión, y uno corresponde a un origen muy específico con una ponderación de dimensión grande y excepcional y las demás ponderaciones son prácticamente iguales a cero.

b. Importancia relativa de cada dimensión, indicada como la razón de la suma de cuadrados de una dimensión entre la suma de cuadrados total.

En la configuración inicial de la correspondencia, cada uno de los espacios individuales tiene una especificidad más alta; es decir, cada situación bajo la que los participantes ordenaron los elementos de desayuno está más asociada a una dimensión específica. Los orígenes más divergentes siguen siendo *Desayuno, con zumo, huevos con bacón, y bebida* y *Aperitivo, con bebida sólo*.

Figura 15-22

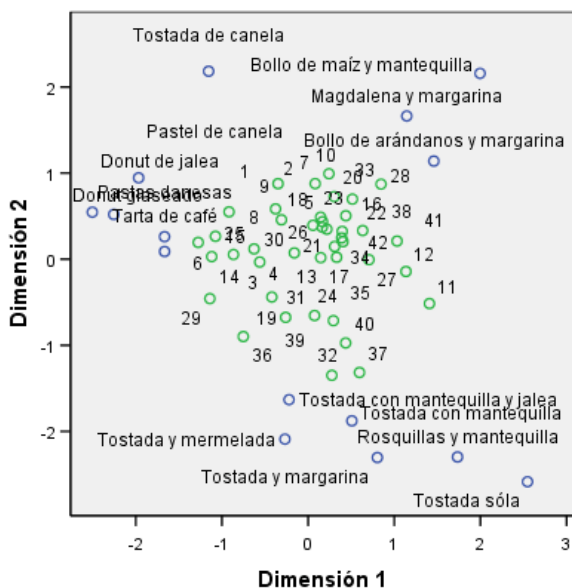
Gráfico conjunto del espacio individual "Desayuno, con zumo, huevos con bacón, y bebida" para la configuración inicial de la correspondencia.



La mayor especificidad es evidente en el gráfico conjunto del espacio individual *Desayuno, con zumo, huevos con bacón, y bebida*. La fuente se carga incluso más en la primera dimensión que bajo el inicio de Spearman clásico, por lo que los objetos de fila y de columna muestran una variación algo menor en el eje vertical y algo mayor en el eje horizontal.

Figura 15-23

Gráfico conjunto del espacio individual "Aperitivo, con bebida sólo" para la configuración inicial de la correspondencia.



El gráfico conjunto del espacio individual *Aperitivo, con bebida sólo* muestra que los objetos de fila y de columna se encuentran más cerca de una línea vertical que bajo el inicio de Spearman clásico.

Ejemplo: Examen de lo apropiado de conducta y situación

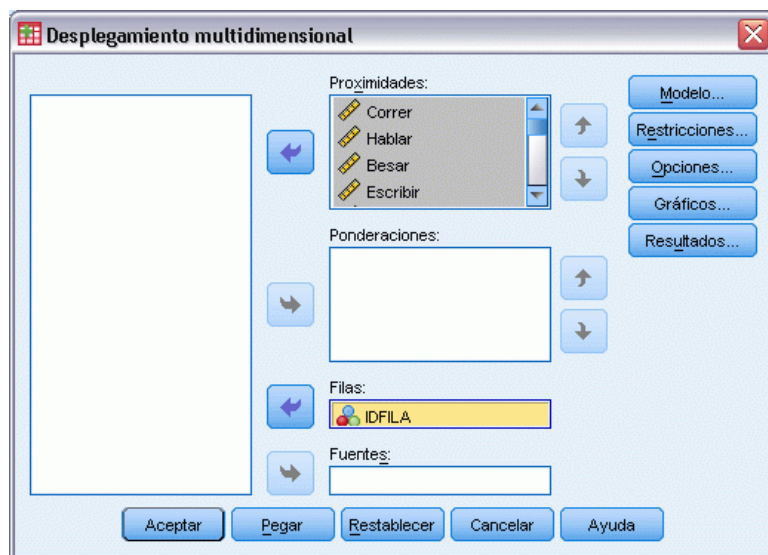
En un ejemplo clásico (Price y Bouffard, 1974), se pidió a 52 estudiantes que valoraran las combinaciones de 15 situaciones y 15 comportamientos en una escala de 10 puntos que oscila entre 0 = "extremadamente apropiado" y 9 = "extremadamente inapropiado". Los valores promediados respecto a los individuos se toman como disimilaridades.

Esta información se recoge en el archivo *behavior.sav*. Si desea obtener más información, consulte el tema [Archivos muestrales en el apéndice A en IBM SPSS Categories 20](#). Utilice el desplegado multidimensional para encontrar conglomerados de situaciones similares y los comportamientos con los que más se asocian. Puede encontrar la sintaxis para reproducir estos análisis en *prefscal_behavior.sps*.

Ejecución del análisis

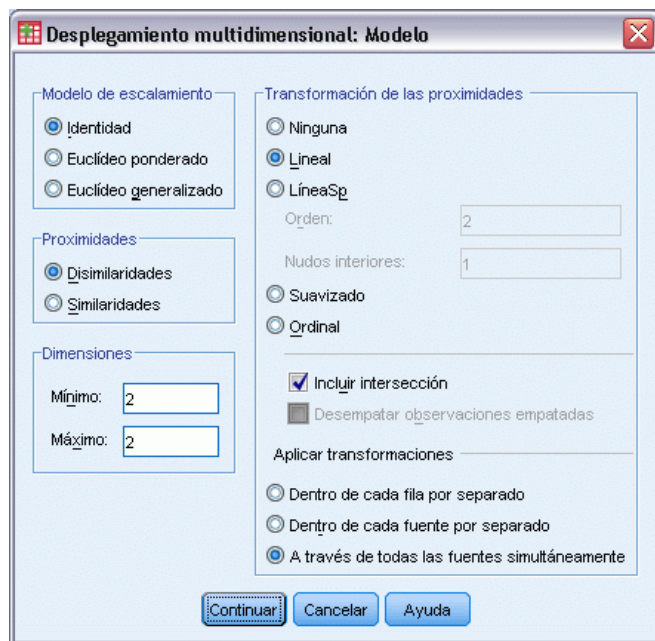
- Para ejecutar un análisis de desplegado multidimensional, elija en los menús: Analizar > Escala > Desplegado multidimensional (PREFSCAL)...

Figura 15-24
Cuadro de diálogo principal Despliegamiento multidimensional



- ▶ Seleccione desde *Correr* hasta *Gritar* como variables de proximidad.
- ▶ Seleccione *IDFILA* como la variable de fila.
- ▶ Pulse en *Modelo*.

Figura 15-25
Cuadro de diálogo *Modelo*



- ▶ Seleccione *Lineal* como la transformación de proximidad y seleccione *Incluir intersección*.
- ▶ Seleccione aplicar transformaciones *A través de todas las fuentes simultáneamente*.

- ▶ Pulse en Continuar.
- ▶ Pulse en Opciones en el cuadro de diálogo Despliegamiento multidimensional.

Figura 15-26
Cuadro de diálogo Opciones

Despliegamiento multidimensional: Opciones

Configuración inicial

Clásica
Imputación mediante: Triángulo

Roggs-Cliff

Correspondencia

Centroides
Opciones: 1

Inicios aleatorios
Número de inicios: 1

Personalizado

Criterios de iteración

Convergencia del stress: .000001

Stress mínimo: .0001

Iteraciones máximas: 5000

Término de penalización

Magnitud: 0.5

Rango: 1.0

Configuración personalizada

Leer variables desde: Archivo... C:\Program Files\SPSS\Inc\PASWStatistics18\Samples\...\behavior_ini.sav

El número debe coincidir con la dimensionalidad máxima del modelo, actualmente: 2

Las variables que contienen coordenadas de fila deben preceder a las que contienen coordenadas de columna.

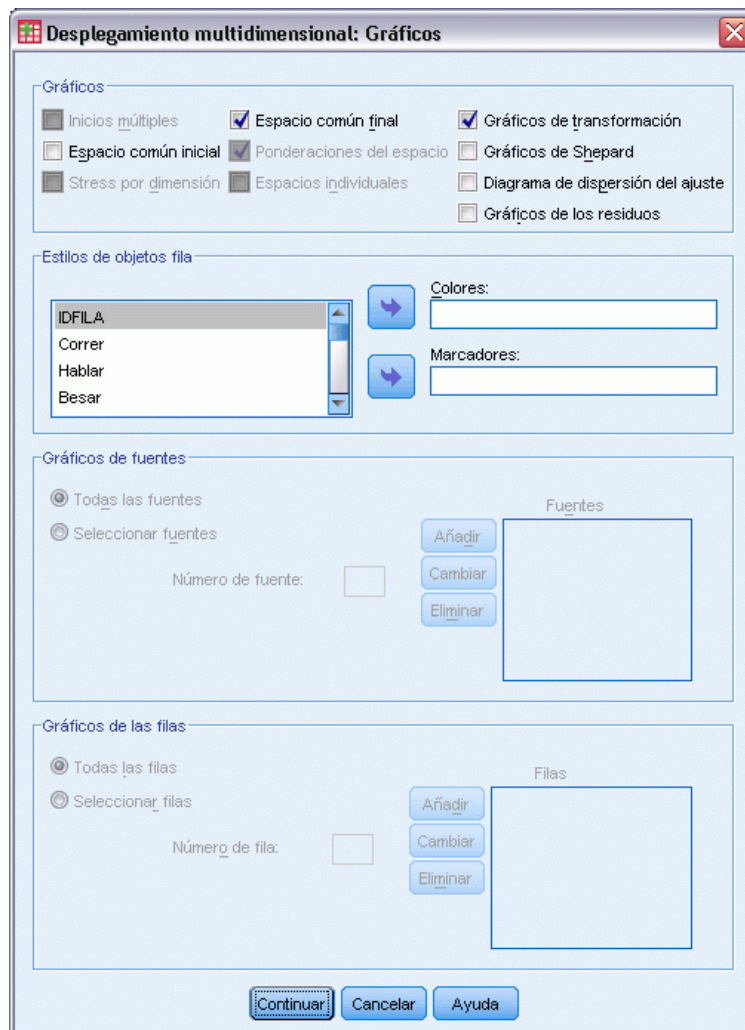
Disponibles: dim1, dim2

Seleccionadas: dim1, dim2

Continuar Cancelar Ayuda

- ▶ Seleccione Personalizado en el grupo Configuración inicial.
- ▶ Acceda al archivo *behavior_ini.sav* y selecciónelo como el archivo que contiene la configuración personalizada inicial. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A en *IBM SPSS Categories 20*.
- ▶ Seleccione *dim1* y *dim2* como las variables que especifican la configuración inicial.
- ▶ Pulse en Continuar.
- ▶ Pulse en Gráficos en el cuadro de diálogo Despliegamiento multidimensional.

Figura 15-27
Cuadro de diálogo Gráficos



- ▶ Seleccione Gráficos de transformación en el grupo Gráficos.
- ▶ Pulse en Continuar.
- ▶ Pulse en Aceptar en el cuadro de diálogo Desplegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=Correr Hablar Besar Escribir Comer Dormir Mascullar Leer Pelear Eructar
Discutir Saltar
Llorar Reir Gritar
/INPUT=ROWS (ROWID )
/INITIAL=( 'samplesDirectory/behavior_ini.sav' )
dim1 dim2
/CONDITION=UNCONDITIONAL
/TRANSFORMATION=LINEAR (INTERCEPT)
/PROXIMITIES=DISSIMILARITIES
```

```
/MODEL=IDENTITY  
/CRITERIA=DIMENSIONS(2,2) DIFFSTRESS(.000001) MINSTRESS(.0001)  
MAXITER(5000)  
/PENALTY=LAMBDA(0.5) OMEGA(1.0)  
/PRINT=MEASURES COMMON  
/PLOT=COMMON TRANSFORMATIONS .
```

- Esta sintaxis especifica un análisis de las variables *correr* hasta *gritar*. La variable *idfila* se utiliza para identificar las filas.
- El subcomando `INITIAL` especifica que los valores iniciales se toman del archivo *behavior_ini.sav*. Las coordenadas de la fila y la columna están apiladas; las coordenadas de la columna siguen a las coordenadas de la fila.
- El subcomando `CONDITION` especifica que todas las proximidades se pueden comparar entre ellas. Esto se cumple en este análisis, ya que debería poder comparar las proximidades para correr en un parque y correr en una iglesia y ver que un comportamiento se considera menos apropiado que el otro.
- El subcomando `TRANSFORMATION` especifica una transformación lineal de las proximidades, con intersección. Esto es adecuado si una diferencia de 1 punto en las proximidades es equivalente en el rango de la escala de 10 puntos. Es decir, si los estudiantes asignan sus puntuaciones para que la diferencia entre 0 y 1 sea la misma que la diferencia entre 5 y 6, la transformación lineal será adecuada.
- El subcomando `PLOT` requiere gráficos del espacio común y gráficos de transformación.
- El resto de parámetros vuelven a sus valores por defecto.

Medidas

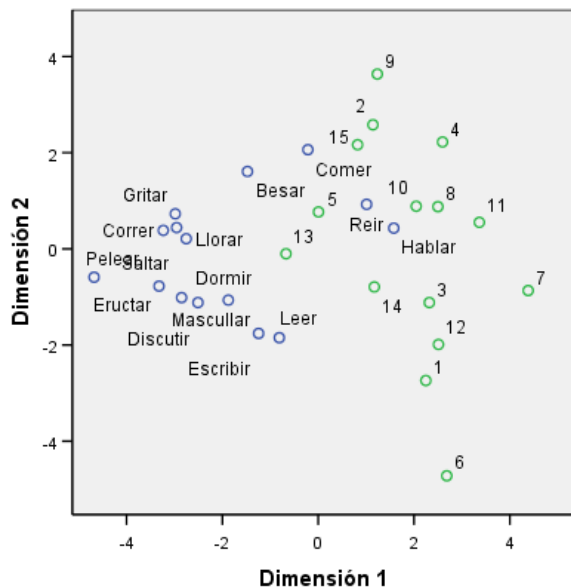
Figura 15-28
Medidas

Iteraciones		169
Valor de función final		,6427725
Partes del valor de función	Parte de Stress	,1900001
	Parte de penalización	2,1745069
Maldad de ajuste	Stress normalizado	,0361000
	Stress-I de Kruskal	,1900001
	Stress-II de Kruskal	,5224668
	S-Stress-I de Young	,2760971
	S-Stress-II de Young	,4525933
Bondad de ajuste	Dispersión explicada	,9639000
	Varianza explicada	,8082862
	Órdenes de preferencia recuperados	,8608333
	Rho de Spearman	,8981120
Coeficientes de variación	Tau-b de Kendall	,7202452
	Variación de las Proximidades	,5138436
	Variación de las Proximidades transformadas	,4751934
	Variación de las Distancias	,3912592
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	,4957969
	Índice de no-degeneración aproximada de Shepard	,7173810

El algoritmo converge después de 169 iteraciones, con una tensión final penalizada de 0.6427725. Los coeficientes de variación y el índice de Shepard son suficientemente altos y los índices de DeSarbo son suficientemente bajos como para sugerir que no existen problemas de degeneración.

Espacio común

Figura 15-29
Gráfico conjunto del espacio común

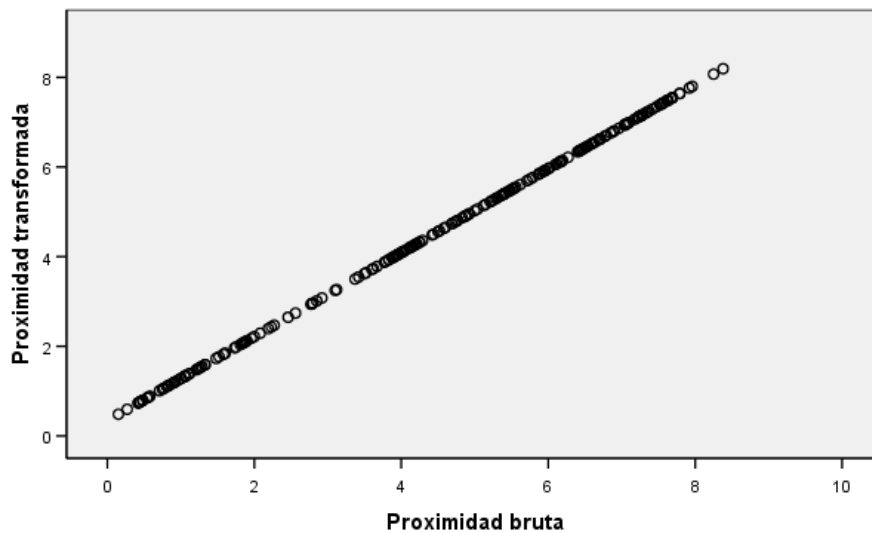


La dimensión horizontal parece estar más asociada a los objetos de columna (comportamientos) y discrimina entre comportamientos inapropiados (pelear, eructar) y comportamientos más apropiados. La dimensión vertical parece estar más asociada a los objetos de fila (situaciones) y define distintas restricciones situacionales o de comportamiento.

- Hacia la parte inferior de la dimensión vertical hay situaciones (iglesia, clase) que restringen el comportamiento a los tipos de comportamiento más tranquilos o introspectivos (leer, escribir). Por lo tanto, estos comportamientos descienden en el eje vertical.
- Hacia la parte superior de la dimensión vertical hay situaciones (películas, juego, cita) que restringen el comportamiento a los tipos de comportamiento más sociales o extrovertidos (comer, besar, reír). Por lo tanto, estos comportamientos ascienden en el eje vertical.
- En el centro de la dimensión vertical, las situaciones se separan en la dimensión horizontal según lo restrictivo que resulta en general la situación. Las más alejadas de los comportamientos (entrevista) son las más restrictivas, mientras que las más cercanas a los comportamientos (dormitorio, parque) suelen ser las menos restrictivas.

Transformaciones de proximidades

Figura 15-30
Gráfico de transformación



Incondicional lineal transformación con intersección

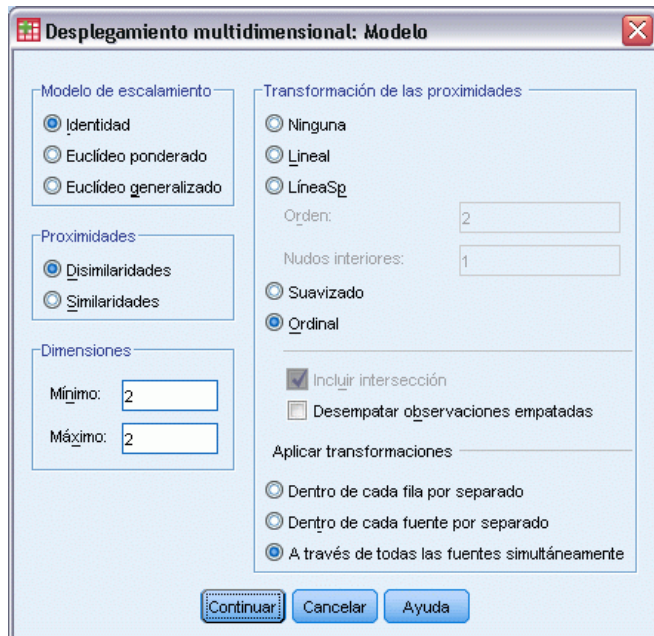
Las proximidades se trataron como lineales en este análisis, de manera que el gráfico de los valores transformados respecto a las proximidades originales presenta una línea recta. El ajuste de esta solución es bueno, aunque es posible lograr un ajuste mejor con una transformación diferente de las proximidades.

Cambio de la transformación de proximidades (ordinal)

- Para generar una solución con una transformación ordinal de las proximidades, pulse en la herramienta Recuperar cuadros de diálogo y seleccione Despliegamiento multidimensional.

- Pulse en Modelo en el cuadro de diálogo Despliegamiento multidimensional.

Figura 15-31
Cuadro de diálogo Modelo



- Seleccione Ordinal como la transformación de proximidad.
- Pulse en Continuar.
- Pulse en Aceptar en el cuadro de diálogo Despliegamiento multidimensional.

A continuación se muestra la sintaxis de comandos generada por estas selecciones:

```
PREFSCAL
VARIABLES=Correr Hablar Besar Escribir Comer Dormir Mascullar Leer Pelear Eructar Discutir Saltar
Llorar Reir Gritar
/INPUT=ROWS (ROWID )
/INITIAL=( 'samplesDirectory/behavior_ini.sav' )
dim1 dim2
/CONDITION=UNCONDITIONAL
/TRANSFORMATION=ORDINAL (KEEPTIES)
/PROXIMITIES=DISSIMILARITIES
/MODEL=IDENTITY
/CRITERIA=DIMENSIONS (2,2) DIFFSTRESS (.000001) MINSTRESS (.0001)
MAXITER (5000)
/PENALTY=LAMBDA (0.5) OMEGA (1.0)
/PRINT=MEASURES COMMON
/PLOT=COMMON TRANSFORMATIONS .
```

- El único cambio se encuentra en el subcomando TRANSFORMATION. La transformación se ha establecido en ORDINAL, que conserva el orden de proximidades pero no requiere que los valores transformados sean proporcionales a los valores originales.

Medidas

Figura 15-32
 Medidas para la solución con transformación ordinal

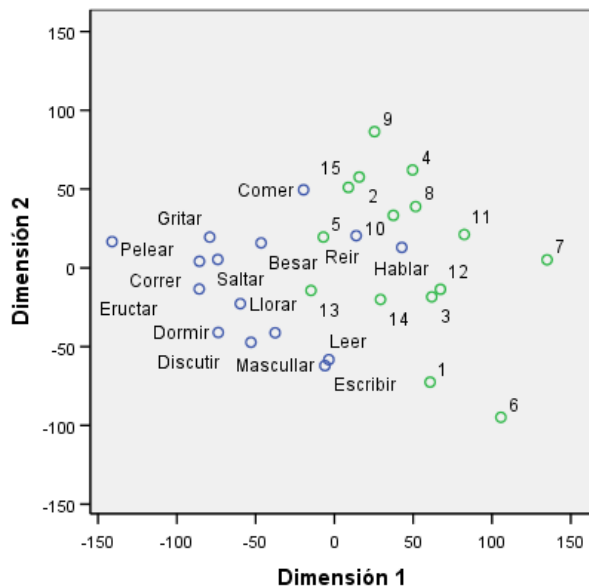
Iteraciones		268
Valor de función final		,6044671
Partes del valor de función	Parte de Stress	,1747239
	Parte de penalización	2,0911875
Maldad de ajuste	Stress normalizado	,0305285
	Stress-I de Kruskal	,1747239
	Stress-II de Kruskal	,4444641
	S-Stress-I de Young	,2707147
	S-Stress-II de Young	,3978003
Bondad de ajuste	Dispersión explicada	,9694715
	Varianza explicada	,8454488
	Órdenes de preferencia recuperados	,8574206
	Rho de Spearman	,9032676
	Tau-b de Kendall	,7532788
Coeficientes de variación	Variación de las Proximidades	,5138436
	Variación de las Proximidades transformadas	,4930018
	Variación de las Distancias	,4284849
Índices de degeneración	Suma de cuadrados de los índices de entremezclado de DeSarbo	,3610680
	Índice de no-degeneración aproximada de Shepard	,7469048

El algoritmo converge después de 268 iteraciones, con una tensión final penalizada de 0,6044671. Este estadístico y las otras medidas son ligeramente mejores para esta solución que la solución obtenida con una transformación lineal de las proximidades.

Espacio común

Figura 15-33

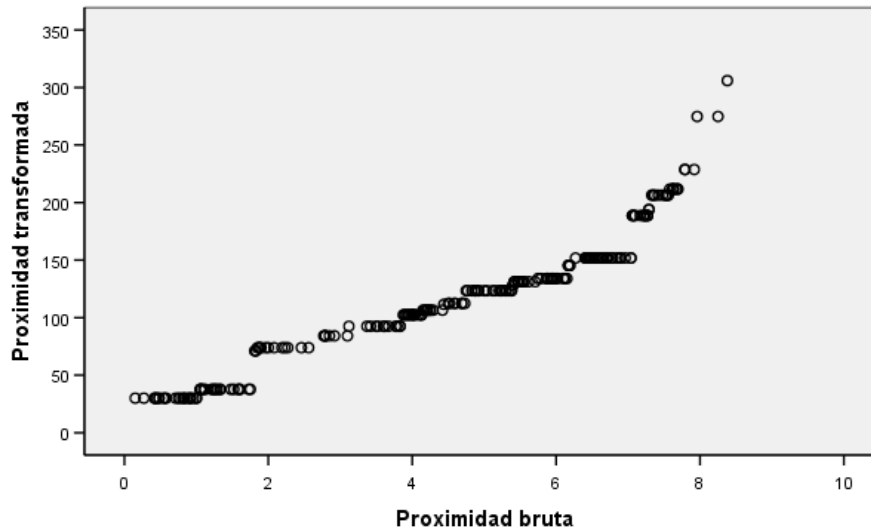
Gráfico conjunto del espacio común para la solución con transformación ordinal



La interpretación del espacio común es la misma en ambas soluciones. Quizás esta solución (con la transformación ordinal) tiene una variación relativamente menor en la dimensión vertical que en la dimensión horizontal de lo que es evidente en la solución con la transformación lineal.

Transformaciones de proximidades

Figura 15-34
Gráfico de transformación para la solución con transformación ordinal



Incondicional ordinal transformación con empates que se dejan empatados

Aparte de los valores con las proximidades mayores, que se alejan del resto de los valores, la transformación ordinal de las proximidades es bastante lineal. Estas proximidades suelen explicar la mayoría de las diferencias entre las soluciones ordinal y lineal; sin embargo, no hay suficiente información aquí como para determinar si esta tendencia no lineal de los valores más altos es verdadera o se trata de una anomalía.

Lecturas recomendadas

Consulte los siguientes textos si desea obtener más información:

Busing, F. M. T. A., P. J. F. Groenen, y W. J. Heiser. 2005. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70, .

Green, P. E., y V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

Price, R. H., y D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .

Archivos muestrales

Los archivos muestrales instalados con el producto se encuentran en el subdirectorio *Samples* del directorio de instalación. Hay una carpeta independiente dentro del subdirectorio *Samples* para cada uno de los siguientes idiomas: Inglés, francés, alemán, italiano, japonés, coreano, polaco, ruso, chino simplificado, español y chino tradicional.

No todos los archivos muestrales están disponibles en todos los idiomas. Si un archivo muestral no está disponible en un idioma, esa carpeta de idioma contendrá una versión en inglés del archivo muestral.

Descripciones

A continuación, se describen brevemente los archivos muestrales usados en varios ejemplos que aparecen a lo largo de la documentación.

- **accidents.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo de edad y género que influyen en los accidentes de automóviles de una región determinada. Cada caso corresponde a una clasificación cruzada de categoría de edad y género.
- **adl.sav.** Archivo de datos hipotéticos relativo a los esfuerzos para determinar las ventajas de un tipo propuesto de tratamiento para pacientes que han sufrido un derrame cerebral. Los médicos dividieron de manera aleatoria a pacientes (mujeres) que habían sufrido un derrame cerebral en dos grupos. El primer grupo recibió el tratamiento físico estándar y el segundo recibió un tratamiento emocional adicional. Tres meses después de los tratamientos, se puntuaron las capacidades de cada paciente para realizar actividades cotidianas como variables ordinales.
- **advert.sav.** Archivo de datos hipotéticos sobre las iniciativas de un minorista para examinar la relación entre el dinero invertido en publicidad y las ventas resultantes. Para ello, se recopilaron las cifras de ventas anteriores y los costes de publicidad asociados.
- **aflatoxin.sav.** Archivo de datos hipotéticos sobre las pruebas realizadas en las cosechas de maíz con relación a la aflatoxina, un veneno cuya concentración varía ampliamente en los rendimientos de cultivo y entre los mismos. Un procesador de grano ha recibido 16 muestras de cada uno de los 8 rendimientos de cultivo y ha medido los niveles de aflatoxinas en partes por millón (PPM).
- **anorectic.sav.** Mientras trabajaban en una sintomatología estandarizada del comportamiento anoréxico/bulímico, los investigadores (Van der Ham, Meulman, Van Strien, y Van Engeland, 1997) realizaron un estudio de 55 adolescentes con trastornos de la alimentación conocidos. Cada paciente fue examinado cuatro veces durante cuatro años, lo que representa un total de 220 observaciones. En cada observación, se puntuó a los pacientes por cada uno de los 16 síntomas. Faltan las puntuaciones de los síntomas para el paciente 71 en el tiempo 2, el paciente 76 en el tiempo 2 y el paciente 47 en el tiempo 3, lo que nos deja 217 observaciones válidas.

- **bankloan.sav.** Archivo de datos hipotéticos sobre las iniciativas de un banco para reducir la tasa de moras de créditos. El archivo contiene información financiera y demográfica de 850 clientes anteriores y posibles clientes. Los primeros 700 casos son clientes a los que anteriormente se les ha concedido un préstamo. Al menos 150 casos son posibles clientes cuyos riesgos de crédito el banco necesita clasificar como positivos o negativos.
- **bankloan_binning.sav.** Archivo de datos hipotéticos que contiene información financiera y demográfica sobre 5.000 clientes anteriores.
- **behavior.sav.** En un ejemplo clásico (Price y Bouffard, 1974), se pidió a 52 estudiantes que valoraran las combinaciones de 15 situaciones y 15 comportamientos en una escala de 10 puntos que oscilaba entre 0 = “extremadamente apropiado” y 9 = “extremadamente inapropiado”. Los valores promediados respecto a los individuos se toman como disimilaridades.
- **behavior_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución bidimensional de *behavior.sav*.
- **brakes.sav.** Archivo de datos hipotéticos sobre el control de calidad de una fábrica que produce frenos de disco para automóviles de alto rendimiento. El archivo de datos contiene las medidas del diámetro de 16 discos de cada una de las 8 máquinas de producción. El diámetro objetivo para los frenos es de 322 milímetros.
- **breakfast.sav.** En un estudio clásico (Green y Rao, 1972), se pidió a 21 estudiantes de administración de empresas de la Wharton School y sus cónyuges que ordenaran 15 elementos de desayuno por orden de preferencia, de 1 = “más preferido” a 15 = “menos preferido”. Sus preferencias se registraron en seis escenarios distintos, de “Preferencia global” a “Aperitivo, con bebida sólo”.
- **breakfast-overall.sav.** Este archivo de datos sólo contiene las preferencias de elementos de desayuno para el primer escenario, “Preferencia global”.
- **broadband_1.sav** Archivo de datos hipotéticos que contiene el número de suscriptores, por región, a un servicio de banda ancha nacional. El archivo de datos contiene números de suscriptores mensuales para 85 regiones durante un período de cuatro años.
- **broadband_2.sav** Este archivo de datos es idéntico a *broadband_1.sav* pero contiene datos para tres meses adicionales.
- **car_insurance_claims.sav.** Un conjunto de datos presentados y analizados en otro lugar (McCullagh y Nelder, 1989) estudia las reclamaciones por daños en vehículos. La cantidad de reclamaciones media se puede modelar como si tuviera una distribución Gamma, mediante una función de enlace inversa para relacionar la media de la variable dependiente con una combinación lineal de la edad del asegurado, el tipo de vehículo y la antigüedad del vehículo. El número de reclamaciones presentadas se puede utilizar como una ponderación de escalamiento.
- **car_sales.sav.** Este archivo de datos contiene estimaciones de ventas, precios de lista y especificaciones físicas hipotéticas de varias marcas y modelos de vehículos. Los precios de lista y las especificaciones físicas se han obtenido de *edmunds.com* y de sitios de fabricantes.
- **car_sales_uprepared.sav.** Ésta es una versión modificada de *car_sales.sav* que no incluye ninguna versión transformada de los campos.

- **carpet.sav** En un ejemplo muy conocido (Green y Wind, 1973), una compañía interesada en sacar al mercado un nuevo limpiador de alfombras desea examinar la influencia de cinco factores sobre la preferencia del consumidor: diseño del producto, marca comercial, precio, sello de *buen producto para el hogar* y garantía de devolución del importe. Hay tres niveles de factores para el diseño del producto, cada uno con una diferente colocación del cepillo del aplicador; tres nombres comerciales (*K2R*, *Glory* y *Bissell*); tres niveles de precios; y dos niveles (no o sí) para los dos últimos factores. Diez consumidores clasificaron 22 perfiles definidos por estos factores. La variable *Preferencia* contiene el rango de las clasificaciones medias de cada perfil. Las clasificaciones inferiores corresponden a preferencias elevadas. Esta variable refleja una medida global de la preferencia de cada perfil.
- **carpet_prefs.sav** Este archivo de datos se basa en el mismo ejemplo que el descrito para *carpet.sav*, pero contiene las clasificaciones reales recogidas de cada uno de los 10 consumidores. Se pidió a los consumidores que clasificaran los 22 perfiles de los productos empezando por el menos preferido. Las variables desde *PREF1* hasta *PREF22* contienen los ID de los perfiles asociados, como se definen en *carpet_plan.sav*.
- **catalog.sav** Este archivo de datos contiene cifras de ventas mensuales hipotéticas de tres productos vendidos por una compañía de venta por catálogo. También se incluyen datos para cinco variables predictoras posibles.
- **catalog_seasfac.sav** Este archivo de datos es igual que *catalog.sav*, con la excepción de que incluye un conjunto de factores estacionales calculados a partir del procedimiento Descomposición estacional junto con las variables de fecha que lo acompañan.
- **cellular.sav**. Archivo de datos hipotéticos sobre las iniciativas de una compañía de telefonía móvil para reducir el abandono de clientes. Las puntuaciones de propensión al abandono de clientes se aplican a las cuentas, oscilando de 0 a 100. Las cuentas con una puntuación de 50 o superior pueden estar buscando otros proveedores.
- **ceramics.sav**. Archivo de datos hipotéticos sobre las iniciativas de un fabricante para determinar si una nueva aleación de calidad tiene una mayor resistencia al calor que una aleación estándar. Cada caso representa una prueba independiente de una de las aleaciones; la temperatura a la que registró el fallo del rodamiento.
- **cereal.sav**. Archivo de datos hipotéticos sobre una encuesta realizada a 880 personas sobre sus preferencias en el desayuno, teniendo también en cuenta su edad, sexo, estado civil y si tienen un estilo de vida activo o no (en función de si practican ejercicio al menos dos veces a la semana). Cada caso representa un encuestado diferente.
- **clothing_defects.sav**. Archivo de datos hipotéticos sobre el proceso de control de calidad en una fábrica de prendas. Los inspectores toman una muestra de prendas de cada lote producido en la fábrica, y cuentan el número de prendas que no son aceptables.
- **coffee.sav**. Este archivo de datos pertenece a las imágenes percibidas de seis marcas de café helado (Kennedy, Riquier, y Sharp, 1996). Para cada uno de los 23 atributos de imagen de café helado, los encuestados seleccionaron todas las marcas que quedaban descritas por el atributo. Las seis marcas se denotan AA, BB, CC, DD, EE y FF para mantener la confidencialidad.
- **contacts.sav**. Archivo de datos hipotéticos sobre las listas de contactos de un grupo de representantes de ventas de ordenadores de empresa. Cada uno de los contactos está categorizado por el departamento de la compañía en el que trabaja y su categoría en la compañía. Además, también se registran los importes de la última venta realizada, el tiempo transcurrido desde la última venta y el tamaño de la compañía del contacto.

- **creditpromo.sav.** Archivo de datos hipotéticos sobre las iniciativas de unos almacenes para evaluar la eficacia de una promoción de tarjetas de crédito reciente. Para este fin, se seleccionaron aleatoriamente 500 titulares. La mitad recibieron un anuncio promocionando una tasa de interés reducida sobre las ventas realizadas en los siguientes tres meses. La otra mitad recibió un anuncio estacional estándar.
- **customer_dbase.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía para usar la información de su almacén de datos para realizar ofertas especiales a los clientes con más probabilidades de responder. Se seleccionó un subconjunto de la base de clientes aleatoriamente a quienes se ofrecieron las ofertas especiales y sus respuestas se registraron.
- **customer_information.sav.** Archivo de datos hipotéticos que contiene la información de correo del cliente, como el nombre y la dirección.
- **customer_subset.sav.** Un subconjunto de 80 casos de *customer_dbase.sav*.
- **debate.sav.** Archivos de datos hipotéticos sobre las respuestas emparejadas de una encuesta realizada a los asistentes a un debate político antes y después del debate. Cada caso corresponde a un encuestado diferente.
- **debate_aggregate.sav.** Archivo de datos hipotéticos que agrega las respuestas de *debate.sav*. Cada caso corresponde a una clasificación cruzada de preferencias antes y después del debate.
- **demo.sav.** Archivos de datos hipotéticos sobre una base de datos de clientes adquirida con el fin de enviar por correo ofertas mensuales. Se registra si el cliente respondió a la oferta, junto con información demográfica diversa.
- **demo_cs_1.sav.** Archivo de datos hipotéticos sobre el primer paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una ciudad diferente, y se registra la identificación de la ciudad, la región, la provincia y el distrito.
- **demo_cs_2.sav.** Archivo de datos hipotéticos sobre el segundo paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una unidad familiar diferente de las ciudades seleccionadas en el primer paso, y se registra la identificación de la unidad, la subdivisión, la ciudad, el distrito, la provincia y la región. También se incluye la información de muestreo de las primeras dos etapas del diseño.
- **demo_cs.sav.** Archivo de datos hipotéticos que contiene información de encuestas recopilada mediante un diseño de muestreo complejo. Cada caso corresponde a una unidad familiar distinta, y se recopila información demográfica y de muestreo diversa.
- **dmdata.sav.** Éste es un archivo de datos hipotéticos que contiene información demográfica y de compras para una empresa de marketing directo. *dmdata2.sav* contiene información para un subconjunto de contactos que recibió un envío de prueba, y *dmdata3.sav* contiene información sobre el resto de contactos que no recibieron el envío de prueba.
- **dietstudy.sav.** Este archivo de datos hipotéticos contiene los resultados de un estudio sobre la “dieta Stillman” (Rickman, Mitchell, Dingman, y Dalen, 1974). Cada caso corresponde a un sujeto distinto y registra sus pesos antes y después de la dieta en libras y niveles de triglicéridos en mg/100 ml.
- **dvdplayer.sav.** Archivo de datos hipotéticos sobre el desarrollo de un nuevo reproductor de DVD. El equipo de marketing ha recopilado datos de grupo de enfoque mediante un prototipo. Cada caso corresponde a un usuario encuestado diferente y registra información demográfica sobre los encuestados y sus respuestas a preguntas acerca del prototipo.

- **german_credit.sav.** Este archivo de datos se toma del conjunto de datos “German credit” de las Repository of Machine Learning Databases (Blake y Merz, 1998) de la Universidad de California, Irvine.
- **grocery_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *grocery_coupons.sav* con las compras semanales “acumuladas” para que cada caso corresponda a un cliente diferente. Algunas de las variables que cambiaban semanalmente desaparecen de los resultados, y la cantidad gastada registrada se convierte ahora en la suma de las cantidades gastadas durante las cuatro semanas del estudio.
- **grocery_coupons.sav.** Archivo de datos hipotéticos que contiene datos de encuestas recopilados por una cadena de tiendas de alimentación interesada en los hábitos de compra de sus clientes. Se sigue a cada cliente durante cuatro semanas, y cada caso corresponde a un cliente-semana distinto y registra información sobre dónde y cómo compran los clientes, incluida la cantidad que invierten en comestibles durante esa semana.
- **guttman.sav.** Bell (Bell, 1961) presentó una tabla para ilustrar posibles grupos sociales. Guttman (Guttman, 1968) utilizó parte de esta tabla, en la que se cruzaron cinco variables que describían elementos como la interacción social, sentimientos de pertenencia a un grupo, proximidad física de los miembros y grado de formalización de la relación con siete grupos sociales teóricos, incluidos multitudes (por ejemplo, las personas que acuden a un partido de fútbol), espectadores (por ejemplo, las personas que acuden a un teatro o de una conferencia), públicos (por ejemplo, los lectores de periódicos o los espectadores de televisión), muchedumbres (como una multitud pero con una interacción mucho más intensa), grupos primarios (íntimos), grupos secundarios (voluntarios) y la comunidad moderna (confederación débil que resulta de la proximidad cercana física y de la necesidad de servicios especializados).
- **health_funding.sav.** Archivo de datos hipotéticos que contiene datos sobre inversión en sanidad (cantidad por 100 personas), tasas de enfermedad (índice por 10.000 personas) y visitas a centros de salud (índice por 10.000 personas). Cada caso representa una ciudad diferente.
- **hivassay.sav.** Archivo de datos hipotéticos sobre las iniciativas de un laboratorio farmacéutico para desarrollar un ensayo rápido para detectar la infección por VIH. Los resultados del ensayo son ocho tonos de rojo con diferentes intensidades, donde los tonos más oscuros indican una mayor probabilidad de infección. Se llevó a cabo una prueba de laboratorio de 2.000 muestras de sangre, de las cuales una mitad estaba infectada con el VIH y la otra estaba limpia.
- **hourlywagedata.sav.** Archivo de datos hipotéticos sobre los salarios por horas de enfermeras de puestos de oficina y hospitales y con niveles distintos de experiencia.
- **insurance_claims.sav.** Éste es un archivo de datos hipotéticos sobre una compañía de seguros que desea generar un modelo para etiquetar las reclamaciones sospechosas y potencialmente fraudulentas. Cada caso representa una reclamación diferente.
- **insure.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo que indican si un cliente tendrá que hacer una reclamación a lo largo de un contrato de seguro de vida de 10 años. Cada caso del archivo de datos representa un par de contratos (de los que uno registró una reclamación y el otro no), agrupados por edad y sexo.
- **judges.sav.** Archivo de datos hipotéticos sobre las puntuaciones concedidas por jueces cualificados (y un aficionado) a 300 actuaciones gimnásticas. Cada fila representa una actuación diferente; los jueces vieron las mismas actuaciones.

- **kinship_dat.sav.** Rosenberg y Kim (Rosenberg y Kim, 1975) comenzaron a analizar 15 términos de parentesco [tía, hermano, primos, hija, padre, nieta, abuelo, abuela, nieto, madre, sobrino, sobrina, hermana, hijo, tío]. Le pidieron a cuatro grupos de estudiantes universitarios (dos masculinos y dos femeninos) que ordenaran estos grupos según las similitudes. A dos grupos (uno masculino y otro femenino) se les pidió que realizaran la ordenación dos veces, pero que la segunda ordenación la hicieran según criterios distintos a los de la primera. Así, se obtuvo un total de seis “fuentes“. Cada fuente se corresponde con una matriz de proximidades de 15×15 cuyas casillas son iguales al número de personas de una fuente menos el número de veces que se partitionaron los objetos en esa fuente.
- **kinship_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución tridimensional de *kinship_dat.sav*.
- **kinship_var.sav.** Este archivo de datos contiene variables independientes *sexo*, *gener(ación)*, y *grado* (de separación) que se pueden usar para interpretar las dimensiones de una solución para *kinship_dat.sav*. Concretamente, se pueden usar para restringir el espacio de la solución a una combinación lineal de estas variables.
- **marketvalues.sav.** Archivo de datos sobre las ventas de casas en una nueva urbanización de Algonquin, Ill., durante los años 1999 y 2000. Los datos de estas ventas son públicos.
- **nhis2000_subset.sav.** La National Health Interview Survey (NHIS, encuesta del Centro Nacional de Estadísticas de Salud de EE.UU.) es una encuesta detallada realizada entre la población civil de Estados Unidos. Las encuestas se realizaron en persona a una muestra representativa de las unidades familiares del país. Se recogió tanto la información demográfica como las observaciones acerca del estado y los hábitos de salud de los integrantes de cada unidad familiar. Este archivo de datos contiene un subconjunto de información de la encuesta de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Archivo de datos y documentación de uso público. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Fecha de acceso: 2003.
- **ozono.sav.** Los datos incluyen 330 observaciones de seis variables meteorológicas para pronosticar la concentración de ozono a partir del resto de variables. Los investigadores anteriores (Breiman y Friedman, 1985), (Hastie y Tibshirani, 1990) han encontrado que no hay linealidad entre estas variables, lo que dificulta los métodos de regresión típica.
- **pain_medication.sav.** Este archivo de datos hipotéticos contiene los resultados de una prueba clínica sobre medicación antiinflamatoria para tratar el dolor artrítico crónico. Resulta de particular interés el tiempo que tarda el fármaco en hacer efecto y cómo se compara con una medicación existente.
- **patient_los.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de pacientes que fueron admitidos en el hospital ante la posibilidad de sufrir un infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.
- **patlos_sample.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de una muestra de pacientes que recibieron trombolíticos durante el tratamiento del infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.

- **poll_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de los encuestadores para determinar el nivel de apoyo público a una ley antes de una asamblea legislativa. Los casos corresponden a votantes registrados. Cada caso registra el condado, la población y el vecindario en el que vive el votante.
- **poll_cs_sample.sav.** Este archivo de datos hipotéticos contiene una muestra de los votantes enumerados en *poll_cs.sav*. La muestra se tomó según el diseño especificado en el archivo de plan *poll_csplan* y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. Sin embargo, tenga en cuenta que debido a que el plan muestral hace uso de un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*poll_jointprob.sav*). Las variables adicionales que corresponden a los datos demográficos de los votantes y sus opiniones sobre la propuesta de ley se recopilaron y añadieron al archivo de datos después de tomar la muestra.
- **property_assess.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor del condado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a las propiedades vendidas en el condado el año anterior. Cada caso del archivo de datos registra la población en que se encuentra la propiedad, el último asesor que visitó la propiedad, el tiempo transcurrido desde la última evaluación, la valoración realizada en ese momento y el valor de venta de la propiedad.
- **property_assess_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor de un estado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a propiedades del estado. Cada caso del archivo de datos registra el condado, la población y el vecindario en el que se encuentra la propiedad, el tiempo transcurrido desde la última evaluación y la valoración realizada en ese momento.
- **property_assess_cs_sample.sav.** Este archivo de datos hipotéticos contiene una muestra de las propiedades recogidas en *property_assess_cs.sav*. La muestra se tomó en función del diseño especificado en el archivo de plan *property_assess_csplan*, y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. La variable adicional *Valor actual* se recopiló y añadió al archivo de datos después de tomar la muestra.
- **recidivism.sav.** Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un infractor anterior y registra su información demográfica, algunos detalles de su primer delito y, a continuación, el tiempo transcurrido desde su segundo arresto, si ocurrió en los dos años posteriores al primer arresto.
- **recidivism_cs_sample.sav.** Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un delincuente anterior, puesto en libertad tras su primer arresto durante el mes de junio de 2003 y registra su información demográfica, algunos detalles de su primer delito y los datos de su segundo arresto, si se produjo antes de finales de junio de 2006. Los delincuentes se seleccionaron de una muestra de departamentos según el plan de muestreo especificado en *recidivism_cs_csplan*. Como este plan utiliza un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Archivo de datos hipotéticos que contiene datos de transacciones de compra, incluida la fecha de compra, los artículos adquiridos y el importe de cada transacción.

- **salesperformance.sav.** Archivo de datos hipotéticos sobre la evaluación de dos nuevos cursos de formación de ventas. Sesenta empleados, divididos en tres grupos, reciben formación estándar. Además, el grupo 2 recibe formación técnica; el grupo 3, un tutorial práctico. Cada empleado se sometió a un examen al final del curso de formación y se registró su puntuación. Cada caso del archivo de datos representa a un alumno distinto y registra el grupo al que fue asignado y la puntuación que obtuvo en el examen.
- **satisf.sav.** Archivo de datos hipotéticos sobre una encuesta de satisfacción llevada a cabo por una empresa minorista en cuatro tiendas. Se encuestó a 582 clientes en total y cada caso representa las respuestas de un único cliente.
- **screws.sav** Este archivo de datos contiene información acerca de las características de tornillos, pernos, clavos y tacos (Hartigan, 1975).
- **shampoo_ph.sav.** Archivo de datos hipotéticos sobre el control de calidad en una fábrica de productos para el cabello. Se midieron seis lotes de resultados distintos en intervalos regulares y se registró su pH. El intervalo objetivo es de 4,5 a 5,5.
- **ships.sav.** Un conjunto de datos presentados y analizados en otro lugar (McCullagh et al., 1989) sobre los daños en los cargueros producidos por las olas. Los recuentos de incidentes se pueden modelar como si ocurrieran con una tasa de Poisson dado el tipo de barco, el período de construcción y el período de servicio. Los meses de servicio agregados para cada casilla de la tabla formados por la clasificación cruzada de factores proporcionan valores para la exposición al riesgo.
- **site.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía para seleccionar sitios nuevos para sus negocios en expansión. Se ha contratado a dos consultores para evaluar los sitios de forma independiente, quienes, además de un informe completo, han resumido cada sitio como una posibilidad “buena”, “media” o “baja”.
- **smokers.sav.** Este archivo de datos es un resumen de la encuesta sobre toxicomanía 1998 National Household Survey of Drug Abuse y es una muestra de probabilidad de unidades familiares americanas. (<http://dx.doi.org/10.3886/ICPSR02934>) Así, el primer paso de un análisis de este archivo de datos debe ser ponderar los datos para reflejar las tendencias de población.
- **stocks.sav** Este archivo de datos hipotéticos contiene precios de acciones y volumen de un año.
- **stroke_clean.sav.** Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haberla limpiado mediante los procedimientos de la opción Preparación de datos.
- **stroke_invalid.sav.** Este archivo de datos hipotéticos contiene el estado inicial de una base de datos médica que incluye contiene varios errores de entrada de datos.
- **stroke_survival.** Este archivo de datos hipotéticos registra los tiempos de supervivencia de los pacientes que finalizan un programa de rehabilitación tras un ataque isquémico. Tras el ataque, la ocurrencia de infarto de miocardio, ataque isquémico o ataque hemorrágico se anotan junto con el momento en el que se produce el evento registrado. La muestra está truncada a la izquierda ya que únicamente incluye a los pacientes que han sobrevivido al final del programa de rehabilitación administrado tras el ataque.
- **stroke_valid.sav.** Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haber comprobado los valores mediante el procedimiento Validar datos. Sigue conteniendo casos potencialmente anómalos.

- **survey_sample.sav.** Este archivo de datos contiene datos de encuestas, incluyendo datos demográficos y diferentes medidas de actitud. Se basa en un subconjunto de variables de NORC General Social Survey de 1998, aunque algunos valores de datos se han modificado y que existen variables ficticias adicionales se han añadido para demostraciones.
- **telco.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía de telecomunicaciones para reducir el abandono de clientes en su base de clientes. Cada caso corresponde a un cliente distinto y registra diversa información demográfica y de uso del servicio.
- **telco_extra.sav.** Este archivo de datos es similar al archivo de datos *telco.sav*, pero las variables de meses con servicio y gasto de clientes transformadas logarítmicamente se han eliminado y sustituido por variables de gasto del cliente transformadas logarítmicamente tipificadas.
- **telco_missing.sav.** Este archivo de datos es un subconjunto del archivo de datos *telco.sav*, pero algunos valores de datos demográficos se han sustituido con valores perdidos.
- **testmarket.sav.** Archivo de datos hipotéticos sobre los planes de una cadena de comida rápida para añadir un nuevo artículo a su menú. Hay tres campañas posibles para promocionar el nuevo producto, por lo que el artículo se presenta en ubicaciones de varios mercados seleccionados aleatoriamente. Se utiliza una promoción diferente en cada ubicación y se registran las ventas semanales del nuevo artículo durante las primeras cuatro semanas. Cada caso corresponde a una ubicación semanal diferente.
- **testmarket_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *testmarket.sav* con las ventas semanales “acumuladas” para que cada caso corresponda a una ubicación diferente. Como resultado, algunas de las variables que cambiaban semanalmente desaparecen y las ventas registradas se convierten en la suma de las ventas realizadas durante las cuatro semanas del estudio.
- **tree_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.
- **tree_credit.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios.
- **tree_missing_data.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios con un elevado número de valores perdidos.
- **tree_score_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.
- **tree_textdata.sav.** Archivo de datos sencillos con dos variables diseñadas principalmente para mostrar el estado por defecto de las variables antes de realizar la asignación de nivel de medida y etiquetas de valor.
- **tv-survey.sav.** Archivo de datos hipotéticos sobre una encuesta dirigida por un estudio de TV que está considerando la posibilidad de ampliar la emisión de un programa de éxito. Se preguntó a 906 encuestados si verían el programa en distintas condiciones. Cada fila representa un encuestado diferente; cada columna es una condición diferente.
- **ulcer_recurrence.sav.** Este archivo contiene información parcial de un estudio diseñado para comparar la eficacia de dos tratamientos para prevenir la reaparición de úlceras. Constituye un buen ejemplo de datos censurados por intervalos y se ha presentado y analizado en otro lugar (Collett, 2003).

- **ulcer_recurrence_recoded.sav.** Este archivo reorganiza la información de *ulcer_recurrence.sav* para permitir modelar la probabilidad de eventos de cada intervalo del estudio en lugar de sólo la probabilidad de eventos al final del estudio. Se ha presentado y analizado en otro lugar (Collett et al., 2003).
- **verd1985.sav.** Archivo de datos sobre una encuesta (Verdegaal, 1985). Se han registrado las respuestas de 15 sujetos a 8 variables. Se han dividido las variables de interés en tres grupos. El conjunto 1 incluye *edad* y *ecivil*, el conjunto 2 incluye *mascota* y *noticia*, mientras que el conjunto 3 incluye *música* y *vivir*. Se escala *mascota* como nominal múltiple y *edad* como ordinal; el resto de variables se escalan como nominal simple.
- **virus.sav.** Archivo de datos hipotéticos sobre las iniciativas de un proveedor de servicios de Internet (ISP) para determinar los efectos de un virus en sus redes. Se ha realizado un seguimiento (aproximado) del porcentaje de tráfico de correos electrónicos infectados en sus redes a lo largo del tiempo, desde el momento en que se descubre hasta que la amenaza se contiene.
- **wheeze_steubenville.sav.** Subconjunto de un estudio longitudinal de los efectos sobre la salud de la polución del aire en los niños (Ware, Dockery, Spiro III, Speizer, y Ferris Jr., 1984). Los datos contienen medidas binarias repetidas del estado de las sibilancias en niños de Steubenville, Ohio, con edades de 7, 8, 9 y 10 años, junto con un registro fijo de si la madre era fumadora durante el primer año del estudio.
- **workprog.sav.** Archivo de datos hipotéticos sobre un programa de obras del gobierno que intenta colocar a personas desfavorecidas en mejores trabajos. Se siguió una muestra de participantes potenciales del programa, algunos de los cuales se seleccionaron aleatoriamente para entrar en el programa, mientras que otros no siguieron esta selección aleatoria. Cada caso representa un participante del programa diferente.
- **worldsales.sav** Este archivo de datos hipotéticos contiene ingresos por ventas por continente y producto.

Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Puede que IBM no ofrezca los productos, los servicios o las características de los que se habla en este documento en otros países. Consulte a su representante local de IBM para obtener información acerca de los productos y servicios que está disponibles actualmente en su zona. Toda referencia que se haga de un producto, programa o servicio de IBM no implica que sólo se deba utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse todo producto, programa o servicio con funcionalidades equivalentes que no infrinjan los derechos 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.

IBM puede tener patentes o aplicaciones de patentes pendientes que cubren el asunto descrito 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, Estados Unidos

Si tiene alguna pregunta sobre la licencia relacionada con la información del juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus preguntas por escrito a:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

El párrafo siguiente no se aplica a los Reino Unido o cualquier otro país donde tales disposiciones son incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES PROPORCIONA ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, PERO NO LIMITADA A, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones , por lo tanto , esta declaración no se aplique 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 mejoras y/o cambios en los productos y/o los programas descritos en esta publicación en cualquier momento sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no forman parte del material de este producto de IBM y el uso de estos sitios Web es bajo su propio riesgo.

IBM puede utilizar cualquier información que le suministre en cualquier forma que considere adecuada, sin incurrir en ninguna obligación para usted.

Los licenciatarios de este programa que deseen tener información sobre el mismo con el objetivo de habilitar: (i) el intercambio de información entre programas creados independientemente y otros programas (incluyendo este) y (ii) el uso común de la información que se ha intercambiado, deben ponerse en contacto con:

IBM Software Group, a la atención de: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

IBM proporciona el programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible para el mismo bajo los términos de IBM Customer Agreement (Acuerdo de cliente de IBM), IBM International Program License Agreement (Acuerdo de licencia de programa internacional de IBM) o cualquier acuerdo equivalente entre las partes.

Se ha obtenido información acerca de productos que no son de IBM 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.

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 registradas

IBM, el logotipo de IBM, [ibm.com](http://www.ibm.com) y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/o otros países.

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.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales de Sun Microsystems, Inc. en Estados Unidos, otros países o ambos.

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.

Este producto utiliza WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

Capturas de pantalla de productos de Adobe reimpresas con permiso de Adobe Systems Incorporated.

Capturas de pantalla de productos de Microsoft reimpresas con permiso de Microsoft Corporation.



Bibliografía

- Barlow, R. E., D. J. Bartholomew, D. J. Bremner, y H. D. Brunk. 1972. *Statistical inference under order restrictions*. Nueva York: John Wiley and Sons.
- Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. Nueva York: Harper & Row.
- Benzécri, J. P. 1969. Statistical analysis as a tool to make patterns emerge from data. En: *Methodologies of Pattern Recognition*, S. Watanabe, ed. Nueva York: Academic Press.
- Benzécri, J. P. 1992. *Correspondence analysis handbook*. Nueva York: Marcel Dekker.
- Bishop, Y. M., S. E. Feinberg, y P. W. Holland. 1975. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass.: MIT Press.
- Blake, C. L., y C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., y J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, .
- Buja, A. 1990. Remarks on functional canonical variates, alternating least squares methods and ACE. *Annals of Statistics*, 18, .
- Busing, F. M. T. A., P. J. F. Groenen, y W. J. Heiser. 2005. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70, .
- Carroll, J. D. 1968. Generalization of canonical correlation analysis to three or more sets of variables. En: *Proceedings of the 76th Annual Convention of the American Psychological Association*, 3, Washington, D.C.: American Psychological Association.
- Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Commandeur, J. J. F., y W. J. Heiser. 1993. *Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices*. Leiden: Department of Data Theory, University of Leiden.
- De Haas, M., J. A. Algera, H. F. J. M. Van Tuijl, y J. J. Meulman. 2000. Macro and micro goal setting: In search of coherence. *Applied Psychology*, 49, .
- De Leeuw, J. 1982. Nonlinear principal components analysis. En: *COMPSTAT Proceedings in Computational Statistics*, Viena: Physica Verlag.
- De Leeuw, J. 1984. *Canonical analysis of categorical data*, 2nd ed. Leiden: DSWO Press.
- De Leeuw, J. 1984. The Gifi system of nonlinear multivariate analysis. En: *Data Analysis and Informatics III*, E. Diday, et al., ed..
- De Leeuw, J., y W. J. Heiser. 1980. Multidimensional scaling with restrictions on the configuration. En: *Multivariate Analysis, Vol. V*, P. R. Krishnaiah, ed. Amsterdam: North-Holland.
- De Leeuw, J., y J. Van Rijkevorsel. 1980. HOMALS and PRINCALS—Some generalizations of principal components analysis. En: *Data Analysis and Informatics*, E. Diday, et al., ed. Amsterdam: North-Holland.
- De Leeuw, J., F. W. Young, y Y. Takane. 1976. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .
- De Leeuw, J. 1990. Multivariate analysis with optimal scaling. En: *Progress in Multivariate Analysis*, S. Das Gupta, y J. Sethuraman, eds. Calcuta: Indian Statistical Institute.

- Eckart, C., y G. Young. 1936. The approximation of one matrix by another one of lower rank. *Psychometrika*, 1, .
- Fisher, R. A. 1938. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, R. A. 1940. The precision of discriminant functions. *Annals of Eugenics*, 10, .
- Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal components analysis. *Biometrika*, 58, .
- Gifi, A. 1985. *PRINCALS. Research Report UG-85-02*. Leiden: Department of Data Theory, University of Leiden.
- Gifi, A. 1990. *Nonlinear multivariate analysis*. Chichester: John Wiley and Sons.
- Gilula, Z., y S. J. Haberman. 1988. The analysis of multivariate contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association*, 83, .
- Gower, J. C., y J. J. Meulman. 1993. The treatment of categorical information in physical anthropology. *International Journal of Anthropology*, 8, .
- Green, P. E., y V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.
- Green, P. E., y Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1941. The quantification of a class of attributes: A theory and method of scale construction. En: *The Prediction of Personal Adjustment*, P. Horst, ed. Nueva York: Social Science Research Council.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, .
- Hartigan, J. A. 1975. *Clustering algorithms*. Nueva York: John Wiley and Sons.
- Hastie, T., y R. Tibshirani. 1990. *Generalized additive models*. Londres: Chapman and Hall.
- Hastie, T., R. Tibshirani, y A. Buja. 1994. Flexible discriminant analysis. *Journal of the American Statistical Association*, 89, .
- Hayashi, C. 1952. On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 2, .
- Heiser, W. J. 1981. *Unfolding analysis of proximity data*. Leiden: Department of Data Theory, University of Leiden.
- Heiser, W. J., y F. M. T. A. Busing. 2004. Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. En: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..
- Heiser, W. J., y J. J. Meulman. 1994. Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. En: *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, M. Greenacre, y J. Blasius, eds. Nueva York: Academic Press.
- Heiser, W. J., y J. J. Meulman. 1995. Nonlinear methods for the analysis of homogeneity and heterogeneity. En: *Recent Advances in Descriptive Multivariate Analysis*, W. J. Krzanowski, ed. Oxford: Oxford University Press.
- Horst, P. 1961. Generalized canonical correlations and their applications to experimental data. *Journal of Clinical Psychology*, 17, .

- Horst, P. 1961. Relations among m sets of measures. *Psychometrika*, 26, .
- Israëls, A. 1987. *Eigenvalue techniques for qualitative data*. Leiden: DSWO Press.
- Kennedy, R., C. Riquier, y B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, .
- Kettenring, J. R. 1971. Canonical analysis of several sets of variables. *Biometrika*, 58, .
- Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, .
- Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, .
- Kruskal, J. B. 1965. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society Series B*, 27, .
- Kruskal, J. B. 1978. Factor analysis and principal components analysis: Bilinear methods. En: *International Encyclopedia of Statistics*, W. H. Kruskal, y J. M. Tanur, eds. Nueva York: The Free Press.
- Kruskal, J. B., y R. N. Shepard. 1974. A nonmetric variety of linear factor analysis. *Psychometrika*, 39, .
- Krzanowski, W. J., y F. H. C. Marriott. 1994. *Multivariate analysis: Part I, distributions, ordination and inference*. Londres: Edward Arnold.
- Lebart, L., A. Morineau, y K. M. Warwick. 1984. *Multivariate descriptive statistical analysis*. Nueva York: John Wiley and Sons.
- Lingoes, J. C. 1968. The multivariate analysis of qualitative data. *Multivariate Behavioral Research*, 3, .
- Máx, J. 1960. Quantizing for minimum distortion. *Proceedings IEEE (Information Theory)*, 6, .
- McCullagh, P., y J. A. Nelder. 1989. *Modelos lineales generalizados*, 2nd ed. Londres: Chapman & Hall.
- Meulman, J. J. 1982. *Homogeneity analysis of incomplete data*. Leiden: DSWO Press.
- Meulman, J. J. 1986. *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
- Meulman, J. J. 1992. The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika*, 57, .
- Meulman, J. J. 1993. Principal coordinates analysis with optimal transformations of the variables: Minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, 46, .
- Meulman, J. J. 1996. Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification*, 13, .
- Meulman, J. J. 2003. Prediction and classification in nonlinear data analysis: Something old, something new, something borrowed, something blue. *Psychometrika*, 4, .
- Meulman, J. J., y W. J. Heiser. 1997. Graphical display of interaction in multiway contingency tables by use of homogeneity analysis. En: *Visual Display of Categorical Data*, M. Greenacre, y J. Blasius, eds. Nueva York: Academic Press.

- Meulman, J. J., y P. Verboon. 1993. Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, 58, .
- Meulman, J. J., A. J. Van der Kooij, y A. Babinec. 2000. New features of categorical principal components analysis for complicated data sets, including data mining. En: *Classification, Automation and New Media*, W. Gaul, y G. Ritter, eds. Berlín: Springer-Verlag.
- Meulman, J. J., A. J. Van der Kooij, y W. J. Heiser. 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. En: *Handbook of Quantitative Methodology for the Social Sciences*, D. Kaplan, ed. Thousand Oaks, Calif.: Sage Publications, Inc..
- Nishisato, S. 1980. *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Nishisato, S. 1984. Forced classification: A simple application of a quantification method. *Psychometrika*, 49, .
- Nishisato, S. 1994. *Elements of dual scaling: An introduction to practical data analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.
- Pratt, J. W. 1987. Dividing the indivisible: Using simple symmetry to partition variance explained. En: *Proceedings of the Second International Conference in Statistics*, T. Pukkila, y S. Puntanen, eds. Tampere (Finlandia): Universidad de Tampere.
- Price, R. H., y D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Ramsay, J. O. 1989. Monotone regression splines in action. *Statistical Science*, 4, .
- Rao, C. R. 1973. *Linear statistical inference and its applications*, 2nd ed. Nueva York: John Wiley and Sons.
- Rao, C. R. 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. En: *Multivariate Analysis, Vol. 5*, P. R. Krishnaiah, ed. Ámsterdam: North-Holland.
- Rickman, R., N. Mitchell, J. Dingman, y J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .
- Rosenberg, S., y M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Roskam, E. E. 1968. *Metric analysis of ordinal data in psychology*. Voorschoten: VAM.
- Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function I. *Psychometrika*, 27, .
- Shepard, R. N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function II. *Psychometrika*, 27, .
- Shepard, R. N. 1966. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 3, .
- Tenenhaus, M., y F. W. Young. 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical multivariate data. *Psychometrika*, 50, .
- Theunissen, N. C. M., J. J. Meulman, A. L. Den Ouden, H. M. Koopman, G. H. Verrips, S. P. Verloove-Vanhorick, y J. M. Wit. 2003. Changes can be studied when the measurement instrument is different at different time points. *Health Services and Outcomes Research Methodology*, 4, .

- Tucker, L. R. 1960. Intra-individual and inter-individual multidimensionality. En: *Psychological Scaling: Theory & Applications*, H. Gulliksen, y S. Messick, eds. Nueva York: John Wiley and Sons.
- Van der Burg, E. 1988. *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.
- Van der Burg, E., y J. De Leeuw. 1983. Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, 36, .
- Van der Burg, E., J. De Leeuw, y R. Verdegaal. 1988. Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 53, .
- Van der Ham, T., J. J. Meulman, D. C. Van Strien, y H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .
- Van der Kooij, A. J., y J. J. Meulman. 1997. MURALS: Multiple regression and optimal scaling using alternating least squares. En: *Softstat '97*, F. Faulbaum, y W. Bandilla, eds. Stuttgart: Gustav Fisher.
- Van Rijckevorsel, J. 1987. *The application of fuzzy coding and horseshoes in multiple correspondence analysis*. Leiden: DSWO Press.
- Verboon, P., y I. A. Van der Lans. 1994. Robust canonical discriminant analysis. *Psychometrika*, 59, .
- Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (en neerlandés)*. Leiden: Department of Data Theory, University of Leiden.
- Vlek, C., y P. J. Stallen. 1981. Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance*, 28, .
- Wagenaar, W. A. 1988. *Paradoxes of gambling behaviour*. Londres: Lawrence Erlbaum Associates, Inc.
- Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, y B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .
- Winsberg, S., y J. O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika*, 67, .
- Winsberg, S., y J. O. Ramsay. 1983. Monotone spline transformations for dimension reduction. *Psychometrika*, 48, .
- Wolter, K. M. 1985. *Introduction to variance estimation*. Berlín: Springer-Verlag.
- Young, F. W. 1981. Quantitative analysis of qualitative data. *Psychometrika*, 46, .
- Young, F. W., J. De Leeuw, y Y. Takane. 1976. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, 41, .
- Young, F. W., Y. Takane, y J. De Leeuw. 1978. The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, 43, .
- Zeijl, E., Y. te Poel, M. du Bois-Reymond, J. Ravesloot, y J. J. Meulman. 2000. The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research*, 32, .

- actualizaciones relajadas
 - en Escalamiento multidimensional, 79
- ajuste
 - en Análisis de correlación canónica no lineal, 46
- alfa de Cronbach
 - en Análisis de componentes principales categórico, 146
- Análisis de componentes principales categórico, 27, 33, 141, 154
 - almacenamiento de variables, 37
 - cuantificaciones, 147, 167
 - funciones adicionales del comando, 41
 - historial de iteraciones, 146
 - nivel de escalamiento óptimo, 29
 - puntos de categorías, 173
 - puntuaciones de los objetos, 149, 152, 171
 - resumen del modelo, 146, 152, 169
 - saturaciones en componentes, 150, 154, 170
- Análisis de correlación canónica no lineal, 42, 45, 191
 - centroides, 206
 - coordenadas de la categoría, 205
 - cuantificaciones, 203
 - estadísticos, 42
 - funciones adicionales del comando, 47
 - gráficos, 42
 - ponderaciones, 200
 - resumen de análisis, 199
 - saturaciones en componentes, 200, 202
- Análisis de correspondencias, 49–52, 54–55, 216, 218
 - contribuciones, 223
 - dimensiones, 222
 - estadísticos, 49
 - funciones adicionales del comando, 57
 - gráficos, 49
 - gráficos de puntuaciones de columna, 224
 - gráficos de puntuaciones de fila, 224
 - normalización, 217
- Análisis de correspondencias múltiple, 58, 63, 228
 - almacenamiento de variables, 66
 - cuantificaciones de categorías, 234
 - funciones adicionales del comando, 68
 - medidas de discriminación, 233
 - nivel de escalamiento óptimo, 60
 - puntuaciones de los objetos, 232, 236
 - resumen del modelo, 231
 - valores atípicos, 239
- ANOVA
 - en Regresión categórica, 22
- archivos de ejemplo
 - ubicación, 297
- autovalores
 - en Análisis de componentes principales categórico, 146, 152, 169
 - en Análisis de correlación canónica no lineal, 199
- avisos legales, 307
- centroides
 - en Análisis de correlación canónica no lineal, 46, 206
- centroides proyectados
 - en Análisis de correlación canónica no lineal, 206
- coeficiente de variación
 - en despliegamiento multidimensional, 265, 268, 274, 281, 290
- coeficientes
 - en Regresión categórica, 109
- coeficientes de regresión
 - en Regresión categórica, 22
- configuración inicial
 - en Análisis de correlación canónica no lineal, 46
 - en despliegamiento multidimensional, 89
 - en Escalamiento multidimensional, 79
 - en Regresión categórica, 20
- contribuciones
 - en Análisis de correspondencias, 223
- coordenadas de la categoría
 - en Análisis de correlación canónica no lineal, 205
- coordenadas del espacio común
 - en despliegamiento multidimensional, 92
 - en Escalamiento multidimensional, 82
- coordenadas del espacio individual
 - en despliegamiento multidimensional, 92
- correlaciones
 - en Escalamiento multidimensional, 82
- correlaciones de orden cero
 - en Regresión categórica, 110
- correlaciones parciales
 - en Regresión categórica, 110
- correlaciones semiparciales
 - en Regresión categórica, 110
- criterios de iteración
 - en despliegamiento multidimensional, 89
 - en Escalamiento multidimensional, 79
- cuantificaciones
 - en Análisis de componentes principales categórico, 147, 167
 - en Análisis de correlación canónica no lineal, 203
- cuantificaciones de categorías
 - en Análisis de componentes principales categórico, 35
 - en Análisis de correlación canónica no lineal, 46
 - en el análisis de correspondencias múltiple, 65, 234
 - en Regresión categórica, 22
- despliegamiento multidimensional, 85, 262, 285
 - despliegamiento triple, 269
 - espacio común, 266, 269, 275, 282, 291, 295
 - espacios individuales, 276, 283
 - estadísticos, 85
 - funciones adicionales del comando, 94
 - gráficos, 85, 91
 - medidas, 265, 268, 274, 281, 290, 294
 - modelo, 86

- opciones, 89
- restricciones en el espacio común., 88
- resultados, 92
- soluciones degeneradas, 262
- transformaciones de proximidades, 292, 296
- desplegamiento triple
 - en desplegado multidimensional, 269
- diagrama de dispersión del ajuste
 - en desplegado multidimensional, 91
- diagramas de dispersión biespacial
 - en Análisis de componentes principales categórico, 38
 - en Análisis de correspondencias, 55
 - en el análisis de correspondencias múltiple, 67
- diagramas de dispersión trispacial
 - en Análisis de componentes principales categórico, 38
- dimensiones
 - en Análisis de correspondencias, 52, 222
- discretización
 - en Análisis de componentes principales categórico, 31
 - en el análisis de correspondencias múltiple, 60
 - en Regresión categórica, 18
- distancias
 - en desplegado multidimensional, 92
 - en Escalamiento multidimensional, 82
- espacio común
 - en desplegado multidimensional, 266, 269, 275, 282, 291, 295
 - en Escalamiento multidimensional, 256, 259
- espacios individuales
 - en desplegado multidimensional, 276, 283
- estadísticos de confianza
 - en Análisis de correspondencias, 54
- estadísticos descriptivos
 - en Regresión categórica, 22
- gráfico conjunto de espacios individuales
 - en desplegado multidimensional, 276, 283
- gráfico conjunto del espacio común
 - en desplegado multidimensional, 266, 269, 275, 282, 291, 295
- gráficos
 - en Análisis de correlación canónica no lineal, 46
 - en Análisis de correspondencias, 55
 - en Escalamiento multidimensional, 81–82
 - en Regresión categórica, 25
- gráficos de categorías
 - en Análisis de componentes principales categórico, 39
 - en el análisis de correspondencias múltiple, 67
- gráficos de categorías conjuntas
 - en Análisis de componentes principales categórico, 39
 - en el análisis de correspondencias múltiple, 67
- gráficos de centroides proyectados
 - en Análisis de componentes principales categórico, 39
- gráficos de correlaciones
 - en Escalamiento multidimensional, 81
- gráficos de espacio común inicial
 - en desplegado multidimensional, 91
- gráficos de espacios individuales
 - en desplegado multidimensional, 91
 - en Escalamiento multidimensional, 81
- gráficos de inicio múltiple
 - en desplegado multidimensional, 91
- gráficos de medidas de discriminación
 - en el análisis de correspondencias múltiple, 67
- gráficos de ponderaciones del espacio
 - en desplegado multidimensional, 91
- gráficos de ponderaciones del espacio individual
 - en desplegado multidimensional, 91
 - en Escalamiento multidimensional, 81
- gráficos de puntos de objetos
 - en Análisis de componentes principales categórico, 38
 - en el análisis de correspondencias múltiple, 67
- gráficos de puntuaciones de columna
 - en Análisis de correspondencias, 224
- gráficos de puntuaciones de fila
 - en Análisis de correspondencias, 224
- gráficos de residuos
 - en desplegado multidimensional, 91
- gráficos de saturaciones en componentes
 - en Análisis de componentes principales categórico, 40
- gráficos de Shepard
 - en desplegado multidimensional, 91
- gráficos de stress
 - en desplegado multidimensional, 91
 - en Escalamiento multidimensional, 81
- gráficos de transformación
 - en Análisis de componentes principales categórico, 39
 - en desplegado multidimensional, 91, 292, 296
 - en el análisis de correspondencias múltiple, 67
 - en Escalamiento multidimensional, 81, 258
 - en Regresión categórica, 112
- gráficos del espacio común
 - en desplegado multidimensional, 91
 - en Escalamiento multidimensional, 81
- gráficos del espacio común final
 - en desplegado multidimensional, 91
- historial de iteraciones
 - en Análisis de componentes principales categórico, 35, 146
 - en desplegado multidimensional, 92
 - en el análisis de correspondencias múltiple, 65
 - en Escalamiento multidimensional, 82
- importancia
 - en Regresión categórica, 110
- Índice de no-degeneración aproximada de Shepard
 - en desplegado multidimensional, 265, 268, 274, 281, 290
- Índices de entremezclado de DeSarbo
 - en desplegado multidimensional, 265, 268, 274, 281, 290

- inercia
 - en Análisis de correspondencias, 54
- intercorrelaciones
 - en Regresión categórica, 108
- lazo
 - en Regresión categórica, 21
- marcas registradas, 308
- matriz de correlaciones
 - en Análisis de componentes principales categórico, 35
 - en el análisis de correspondencias múltiple, 65
- medidas de discriminación
 - en el análisis de correspondencias múltiple, 65, 233
- medidas de distancia
 - en Análisis de correspondencias, 52
- medidas del stress
 - en despliegamiento multidimensional, 92
 - en Escalamiento multidimensional, 82, 255, 259
- modelo de escalamiento
 - en despliegamiento multidimensional, 86
- modelo de identidad
 - en despliegamiento multidimensional, 86
- modelo Euclídeo generalizado
 - en despliegamiento multidimensional, 86
- modelo Euclídeo ponderado
 - en despliegamiento multidimensional, 86
- Multidimensional Scaling, 70, 72–76, 244
 - espacio común, 256, 259
 - estadísticos, 70
 - funciones adicionales del comando, 84
 - gráficos, 70, 81–82
 - gráficos de transformación, 258
 - medidas del stress, 255, 259
 - modelo, 77
 - opciones, 79
 - restricciones, 78
 - resultados, 82
- nivel de escalamiento óptimo
 - en Análisis de componentes principales categórico, 29
 - en el análisis de correspondencias múltiple, 60
- normalización
 - en Análisis de correspondencias, 52, 217
- normalización principal
 - en Análisis de correspondencias, 217
- normalización principal por columna
 - en Análisis de correspondencias, 217
- normalización principal por fila
 - en Análisis de correspondencias, 217
- normalización simétrica
 - en Análisis de correspondencias, 217
- objetos suplementarios
 - en Regresión categórica, 20
- ponderación de la variable
 - en Análisis de componentes principales categórico, 29
 - en el análisis de correspondencias múltiple, 60
- ponderaciones
 - en Análisis de correlación canónica no lineal, 46, 200
- ponderaciones de dimensión
 - en despliegamiento multidimensional, 276, 283
- ponderaciones del espacio individual.
 - en despliegamiento multidimensional, 92
 - en Escalamiento multidimensional, 82
- PREFSCAL, 85
- proximidades transformadas
 - en despliegamiento multidimensional, 92
 - en Escalamiento multidimensional, 82
- puntos de categorías
 - en Análisis de componentes principales categórico, 173
- puntuaciones de los objetos
 - en Análisis de componentes principales categórico, 35, 149, 152, 171
 - en Análisis de correlación canónica no lineal, 46
 - en el análisis de correspondencias múltiple, 65, 232, 236
- R múltiple
 - en Regresión categórica, 22
- R^2
 - en Regresión categórica, 109
- red elástica
 - en Regresión categórica, 21
- Regresión categórica, 15, 96
 - ajuste del modelo, 109
 - correlaciones, 109–110
 - estadísticos, 15
 - funciones adicionales del comando, 26
 - gráficos, 15
 - gráficos de transformación, 112
 - guardar, 24
 - importancia, 110
 - intercorrelaciones, 108
 - nivel de escalamiento óptimo, 16
 - regularización, 21
 - residuos, 114
- regresión contraída
 - en Regresión categórica, 21
- residuos
 - en Regresión categórica, 114
- restricciones
 - en Escalamiento multidimensional, 78
- restricciones en el espacio común.
 - en despliegamiento multidimensional, 88
- resumen del modelo
 - en el análisis de correspondencias múltiple, 231
- saturaciones en componentes
 - en Análisis de componentes principales categórico, 35, 150, 154, 170
 - en Análisis de correlación canónica no lineal, 46, 202

- tensión penalizada
 - en despliegamiento multidimensional, 265, 274, 281, 290, 294
- término de penalización
 - en despliegamiento multidimensional, 89
- tipificación
 - en Análisis de correspondencias, 52
- transformaciones de proximidades
 - en despliegamiento multidimensional, 86

- valores ajustados
 - en Análisis de correlación canónica no lineal, 199
- valores atípicos
 - en el análisis de correspondencias múltiple, 239
- valores perdidos
 - en Análisis de componentes principales categórico, 32
 - en Análisis de correlación canónica no lineal, 199
 - en el análisis de correspondencias múltiple, 61
 - en Regresión categórica, 19
- variables independientes transformadas
 - en Escalamiento multidimensional, 82
- varianza explicada
 - en Análisis de componentes principales categórico, 35, 146, 169