

IBM SPSS Categories 23

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos" en la página 57.

Información del producto

Esta edición se aplica a la versión 23, release 0, modificación 0 de IBM SPSS Statistics y a todos los releases y modificaciones posteriores hasta que se indique lo contrario en ediciones nuevas.

Contenido

Capítulo 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 medición	2
Selección del nivel de escalamiento óptimo	2
Gráficos de transformación	3
Códigos de la categoría	3
¿Qué procedimiento es el mejor para la aplicación?	5
Regresión categórica	6
Análisis de componentes principales categórico	6
Análisis de correlación canónica no lineal	7
Análisis de correspondencias	8
Análisis de correspondencias múltiple	9
Escalamiento multidimensional	10
Despliegamiento multidimensional	10
Relación de aspecto en gráficos de escalamiento óptimo	11

Capítulo 2. Regresión categórica (CATREG) 13

Para definir la escala en regresión categórica	14
Discretización de regresión categórica	14
Valores perdidos de regresión categórica	15
Opciones de regresión categórica	15
Regularización de regresión categórica	16
Resultados de la regresión categórica	16
Regresión categórica: Guardar	17
Gráficos de transformación en regresión categórica	18
Características adicionales del comando CATREG	18

Capítulo 3. Análisis de componentes principales categórico (CATPCA) 19

Definir escala y ponderación en CATPCA	20
Análisis de componentes principales categórico: Discretización	21
Análisis de componentes principales categórico: Valores perdidos	21
Análisis de componentes principales categórico: Opciones	22
Análisis de componentes principales categórico: Resultados	23
Análisis de componentes principales categórico: Guardar	24
Análisis de componentes principales categórico: Gráficos de objetos	24
Análisis de componentes principales categórico: Gráficos de categorías	25
Análisis de componentes principales categórico: gráfico de cargas	25
Simulación de muestreo del análisis de componentes principales categórico	25
Características adicionales del comando CATPCA	26

Capítulo 4. Análisis de correlación canónica no lineal (OVERALS) 27

Definir rango y escala	28
Definir rango	29
Análisis de correlación canónica no lineal: Opciones	29
Características adicionales del comando OVERALS	30

Capítulo 5. Análisis de correspondencias 31

Definir rango de filas en Análisis de correspondencias	32
Definir rango de columnas en Análisis de correspondencias	32
Análisis de correspondencias: Modelo	33
Análisis de correspondencias: Estadísticos	34
Análisis de correspondencias: Gráficos	34
Características adicionales del comando CORRESPONDENCE	35

Capítulo 6. Análisis de correspondencias múltiple 37

Definir ponderación de la variable en el análisis de correspondencias múltiple	38
Discretización en el análisis de correspondencias múltiple	38
Valores perdidos en el análisis de correspondencias múltiple	38
Opciones de análisis de correspondencias múltiple	39
Resultados del análisis de correspondencias múltiple	40
Análisis de correspondencias múltiple: Guardar	40
Análisis de correspondencias múltiple: Gráficos de objetos	41
Análisis de correspondencias múltiple: Gráficos de variables	41
Características adicionales del comando MULTIPLE CORRESPONDENCE	41

Capítulo 7. Escalamiento multidimensional (PROXSCAL) 43

Proximidades en matrices a través de columnas	44
Proximidades en columnas	44
Proximidades en una columna	45
Crear proximidades de los datos	45
Crear la medida a partir de los datos	45
Definir un modelo de escalamiento multidimensional	46
Escalamiento multidimensional: Restricciones	47
Escalamiento multidimensional: Opciones	47
Escalamiento multidimensional: Gráficos, Versión 1	48
Escalamiento multidimensional: Gráficos, Versión 2	48
Escalamiento multidimensional: Resultados	48
Características adicionales del comando PROXSCAL	49

Capítulo 8. Desplegamiento multidimensional (PREFSCAL) 51

Definir un modelo de despliegamiento multidimensional 52
Restricciones del despliegamiento multidimensional 52
Opciones de despliegamiento multidimensional . . 53
Gráficos de despliegamiento multidimensional. . . 54
Resultados del despliegamiento multidimensional. . 55

Características adicionales del comando PREFSCAL 55

Avisos 57

Marcas comerciales 59

Índice 61

Capítulo 1. 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 medición

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 *escalán*. 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 mediciones:

- 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. 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 medición "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.

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 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 del 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 del sistema. Para el esquema C, también hay cuatro categorías que reciben indicadores de valores perdidos del sistema. Por el contrario, para el esquema A no hay ninguna cuantificación de valores perdidos del 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 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 del 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 del 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 **correlación 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 bidimensionales 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 desplegamiento 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 inclinación y una intersección 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 mediciones 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 estándar 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 estándar. 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 bidimensionales. 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 contingencia 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 bidimensional 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 bidimensional. 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 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 bidimensional 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 bidimensionales 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 cruzadas 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 bidimensionales 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 exclusivos 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 contingencia 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 varios factores. Las tablas de contingencia de varios factores también se pueden analizar con el procedimiento Tablas cruzadas, pero Tablas cruzadas 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 cargas 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.

Escalamiento multidimensional

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 estrés bruto normalizado, en vez del S-stress (también denominado **tensión**). El estrés 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 desplegado 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 similitudes (o disimilitudes) 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 Desplegado multidimensional (PREFSCAL) de categorías ofrece varias mejoras sobre la funcionalidad de desplegado 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 estrés penalizado 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.

Capítulo 2. 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 predicción 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 estándar 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.

Regresión categórica: Consideraciones sobre los datos

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 estándar.

Para obtener una regresión categórica

1. Seleccione en los menús:
Analizar > Regresión > Escalamiento óptimo (CATREG)...
2. Seleccione la variable dependiente y la variable o variables independientes.
3. 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. De forma predeterminada, se escalan como líneasSp (ordinales) monotónicas de segundo orden con dos nudos interiores. Asimismo, se puede definir la ponderación para las variables del análisis.

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

- **Spline 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.
- **Spline 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.
- **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 spline 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 spline 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 estándar.

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 distintos 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. De forma predeterminada, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Método. Seleccione entre agrupación, clasificación y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.
- **Clasificación.** La variable se discretiza mediante la clasificación de 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.

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.

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 spline 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 spline 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. Consulte el tema “Regresión categórica: Guardar” en la página 17 para obtener más información.

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

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. Consulte el tema "Regresión categórica: Guardar" en la página 17 para obtener más información.

- **Regresión Ridge.** La regresión Ridge 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 Ridge; sin embargo, Lazo supone un procesamiento más extenso.
- **Red elástica.** La Red elástica simplemente combina las penalizaciones de Lazo y regresión Ridge, y buscará en la cuadrícula de valores especificada para encontrar los "mejores" coeficientes de penalización de Lazo y de regresión Ridge. En un par concreto de penalizaciones de Lazo y regresión Ridge, 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 Ridge. **Todos los gráficos posibles** utiliza todos los valores del rango determinados por los valores mínimo y máximo especificados de penalización de regresión Ridge. **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.

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 de 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.
- **Coefficientes.** Esta opción produce tres tablas: una tabla de coeficientes que incluye betas, error estándar de las betas, valores t y significación; una tabla de escalamiento óptimo de coeficientes con el error estándar 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.
- **Bootstrap .632.** Con la simulación de muestreo, las observaciones se trazan aleatoriamente a partir de los datos con reposición, repitiendo este proceso una serie de veces para obtener muestras de simulación de muestreo. Se ajusta un modelo para cada muestra de simulación de muestreo. El error de predicción de cada modelo se estima al aplicar el modelo ajustado a los casos que no se encuentran en la muestra de simulación de muestreo.

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.

Los coeficientes y modelos regularizados se guardan siempre que se selecciona un método de regularización en el cuadro de diálogo Regularización. De forma predeterminada, el procedimiento crea

un nuevo conjunto de datos con un nombre exclusivo, 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. De forma predeterminada, el procedimiento crea un nuevo conjunto de datos con un nombre exclusivo, 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.

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 los saltos 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.

Características 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. La sintaxis de comandos también le permite:

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

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 3. 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 *P*roincipal *C*omponents *A*nalysis.

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 estándar 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; cargas 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 trispaciales y gráficos de las cargas en componentes.

Consideraciones sobre los datos del Análisis de componentes principales categórico

Datos. Los valores de las variables de cadena se convierten en enteros positivos por orden alfabético ascendente. Los valores perdidos del 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 estándar. Hay características de representación

alternativas que están disponibles si se utilizan las variables transformadas en un análisis de componentes principales lineal estándar. 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

1. Seleccione en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...
2. Seleccione **Alguna variable no es nominal múltiple**.
3. Seleccione **Un conjunto**.
4. Pulse en **Definir**.
5. Seleccione al menos dos variables de análisis y especifique el número de dimensiones en la solución.
6. 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. De forma predeterminada, se escalan como líneasSp (ordinales) monotónicas de segundo orden con dos nudos interiores. Asimismo, se puede definir la ponderación para las variables del análisis.

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

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

- **Spline 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.
- **Spline 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 spline 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 spline 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 estándar.

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 distintos 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. De forma predeterminada, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Método. Seleccione entre agrupación, clasificación y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.
- **Clasificación.** La variable se discretiza mediante la clasificación de 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 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

Utilice el cuadro de diálogo Valores perdidos para elegir la estrategia para manejar los valores que faltan en las variables de análisis y las variables complementarias.

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.
 - **Moda.** Sustituya valores que faltan con el modo de la variable escalada de forma óptima.
 - **Categoría adicional.** Sustituya valores que faltan con la cuantificación de una categoría adicional. Este valor implica que los objetos con un valor que falta en esta variable se considera que pertenecen a la misma categoría (adicional).

- **Categoría aleatoria.** Impute cada valor que falta en una variable con el valor cuantificado de un número de categoría aleatoria diferente basándose en las frecuencias marginales de las categorías de la variable.
- **Imputar valores perdidos.** Los objetos con valores perdidos en la variable seleccionada tendrán dichos valores imputados. Se puede seleccionar el método de imputación.
 - **Moda.** Sustituya los valores que faltan con la categoría más frecuente. Cuando existen varias modas, se utiliza la que tiene el indicador de categoría más pequeño.
 - **Categoría adicional.** Sustituya los valores que falta con la misma cuantificación de una categoría adicional. Este valor implica que los objetos con un valor que falta en esta variable se considera que pertenecen a la misma categoría (adicional).
 - **Categoría aleatoria.** Sustituya cada valor que falta en una variable con un número de categoría aleatoria diferente basándose en las frecuencias marginales de las categorías.
- **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 Opciones proporciona controles para seleccionar la configuración inicial, especificar criterios de iteración y de convergencia, seleccionar un método de normalización, elegir el método para etiquetar gráficos y especificar objetos complementarios.

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**. Si se especifica un objeto como suplementario, se ignorarán las ponderaciones de casos para dicho objeto.

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

- **Principal por variable.** Esta opción optimiza la asociación entre las variables. Las coordenadas de las variables en el espacio de los objetos son las cargas en componentes (las correlaciones con componentes principales, como son las dimensiones y las puntuaciones de los objetos). Este método es útil si está interesado básicamente en la correlación entre las variables.
- **Principal por objeto.** Esta opción optimiza las distancias entre los objetos. Este método es útil si está interesado básicamente en las diferencias o 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 de 1 es igual al método principal del objeto. Un valor de 0 es igual al método simétrico. Un valor de -1 es igual al método principal de 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 triespaciales 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. Puede especificar si las variables y las etiquetas de valor o los nombres y valores de variable se utilizan en los gráficos. También se puede especificar una longitud máxima para las etiquetas.

Dimensiones del gráfico. Puede controlar las dimensiones que se visualizan en el resultado.

- **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.

Rotación. Puede seleccionar un método de rotación para obtener resultados rotados.

Nota: Estos métodos de rotación no están disponibles si selecciona **Realizar simulación de muestreo** en el diálogo **Simulación de muestreo**.

- **Varimax.** Un método de rotación ortogonal que minimiza el número de variables que tienen saturaciones altas en cada componente. Simplifica la interpretación de los componentes.
- **Quartimax.** Un método de rotación que minimiza el número de componentes que son necesarios para explicar cada variable. Simplifica la interpretación de las variables observadas.
- **Equamax.** Un método de rotación que es una combinación del método Varimax, que simplifica los componentes, y el método Quartimax, que simplifica las variables. El número de variables que imponen una alta saturación en un componente y el número de componentes que son necesarios para explicar una variable se minimizan.
- **Oblimin.** Un método para la rotación oblicua (no ortogonal). Cuando el delta es igual a 0, los componentes son los más oblicuos. Cuando delta se vuelve negativo, los componentes pasan a ser menos oblicuos. Los valores positivos permiten una correlación adicional de componentes. El valor de **Delta** debe ser menor o igual que 0,8.
- **Promax.** Una rotación oblicua (no ortogonal), que permite correlacionar los componentes. Se puede calcular más rápidamente que una rotación Oblimin directa, de forma que es útil para conjuntos de datos grandes. La cantidad de correlación (oblicuidad) que está permitida se controla mediante el parámetro kappa. El valor de **Kappa** debe ser mayor o igual que 1 y menor que 10.000.

Configuración. Puede leer datos en un archivo que contiene las coordenadas de una configuración. La primera variable del archivo contiene las coordenadas para la primera dimensión. La segunda variable contiene las coordenadas para la segunda dimensión y, así, sucesivamente.

- **Inicial.** La configuración en el archivo que se especifica se utiliza como el punto de partida del análisis.
- **Fija.** La configuración en el archivo que se especifica se utiliza 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 controla la visualización de 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.** En la lista de variables que se han especificado como variables de etiquetado, puede seleccionar una para etiquetar los objetos.

Saturaciones de componentes. Muestra las cargas en componentes para todas las variables que no recibieron niveles de escalamiento nominal múltiple. Puede ordenar las saturaciones de componentes por tamaño.

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 por. 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. Proporciona las cuantificaciones de categoría y las coordenadas para cada dimensión de las variables que se han seleccionado.

Estadísticos descriptivos. Muestra frecuencias, el número de valores que faltan y el modo de las variables que se han seleccionado.

Análisis de componentes principales categórico: Guardar

Los controles del recuadro de diálogo Guardar guardan datos discretizados, puntuaciones de objeto, valores transformados y otros resultados en el conjunto de datos activo, un nuevo conjunto de datos en la sesión actual, o un archivo externo.

- 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.
- Las opciones del grupo **Elipses de confianza de simulación de muestreo** solo están disponibles si selecciona **Realizar simulación de muestreo** en el diálogo Simulación de muestreo.

Análisis de componentes principales categórico: Gráficos de objetos

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

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 (cargas en componentes o centroides de las variables).

Objetos, cargas 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 cargas 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.

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 cargas

El cuadro de diálogo Gráficos de carga controla las variables que se incluyen en el gráfico, la visualización de centroides en el gráfico de cargas y la visualización de gráficos de la varianza explicada por.

Varianza explicada por. Para cada dimensión, muestra un gráfico de la varianza explicada por.

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

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

Incluir centroides. Las variables con un nivel de escalamiento nominal múltiple no tienen cargas 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.

Simulación de muestreo del análisis de componentes principales categórico

El diálogo Simulación de muestreo especifica los parámetros para el análisis de la simulación de muestreo.

Realizar simulación de muestreo. Realiza la repetición de la simulación de muestreo.

- Si se solicitan gráficos de cargas, categorías o puntuaciones de componente, se visualizan gráficos adicionales. Estos gráficos muestran los puntos para la muestra de datos y las estimaciones de simulación de muestreo. Los gráficos de transformación incluyen regiones de confianza. También se visualiza un gráfico para los autovalores.
- Si se especifica una solución bidimensional, se visualizan gráficos de elipses de confianza para los autovalores, las cargas de componente, los puntos de categoría y los puntos de objeto.
- La repetición de la simulación de muestreo no está disponible si especifica un método de rotación en el diálogo Opciones.

puede seleccionar una simulación de muestreo **Equilibrado** o **No equilibrado**.

Número de muestras. El número de muestras de la simulación de muestreo que se utilizan para calcular las estimaciones de la simulación de muestreo. El valor debe ser un entero positivo.

Nivel de confianza. El nivel de confianza de las estimaciones de la simulación de muestreo, expresado como porcentaje. El valor debe ser un número positivo menor que 100.

Método de búsqueda de coincidencias. Las dos alternativas son **Procusto** y **Reflejo**.

Elipses de confianza. Controla el área de umbral para las elipses de confianza en gráficos. Para áreas mayores que (> operador) el valor especificado, el número de elipses disminuye cuando aumenta el valor del umbral.

- El valor de este grupo solo está disponible si el número de dimensiones que se han especificado en el diálogo principal es 2.
- Las elipses de confianza para cargar gráficos solo están disponibles si está seleccionado **Mostrar cargas en componentes** en el diálogo Gráficos de cargas.
- Las elipses de confianza para los gráficos de objetos solo están disponibles si está seleccionado **Puntos de objeto** en el diálogo Gráficos de variables y objetos.
- Las elipses de confianza para los gráficos de categoría solo están disponibles si se ha especificado una o más variables en la lista **Gráficos de categorías** en el diálogo Gráficos de categorías.

Número de puntos de contorno de elipse. El número de puntos de gráfico utilizados para trazar cada elipse de confianza. Los valores más grandes generan elipses más suaves. El valor debe ser un entero positivo menor o igual que 100.

Características 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. La sintaxis de comandos también le permite:

- 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).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 4. 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 estándar 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, cargas en las componentes, ajuste simple y múltiple, gráficos de las puntuaciones de objeto, gráficos de las coordenadas de categoría, gráficos de las cargas en las componentes, gráficos de los centroides de categoría y gráficos de transformación.

Análisis de correlación canónica no lineal: Consideraciones sobre los datos

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

1. Seleccione en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...
2. Seleccione **Todas las variables son nominales múltiples** o **Alguna variable no es nominal múltiple**.
3. Seleccione **Múltiples conjuntos**.
4. Pulse en **Definir**.
5. 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 conjuntos adicionales. Pulse en **Anterior** para volver al conjunto de variables definido anteriormente.
6. Defina el rango de valores y la escala de medición (nivel de escalamiento óptimo) para cada variable seleccionada.
7. Pulse en **Aceptar**.
8. 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 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

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

Debe definir un rango para cada variable. El valor máximo especificado debe ser un entero. En el análisis, se truncan 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.

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

- *Centroides.* Las cualificaciones 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 casos que pertenecen a la misma categoría de la variable.
- *Ponderaciones y cargas 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.
- *Ajuste 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 cargas 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.

Características 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. La sintaxis de comandos también le permite:

- 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 `NOBSERVATIONS`).
- 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`).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 5. 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 cruzadas ofrece varias medidas y pruebas de asociación pero no puede representar gráficamente ninguna relación entre las variables.

El análisis factorial es una técnica estándar 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 singulares, 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.

Análisis de correspondencias: Consideraciones sobre los datos

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

1. Seleccione en los menús:
Analizar > Reducción de dimensiones > Análisis de correspondencias...
2. Seleccione una variable de filas.
3. Seleccione una variable de columnas.
4. Defina los rangos para las variables.
5. 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.

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.

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.

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 estándar.
- **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 estándar.
- **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 singular 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.

Tabla de correspondencias. Es la tabulación cruzada 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 estándar y las correlaciones para todos los puntos de fila no suplementarios.

Estadísticos de confianza para puntos de columna. Incluye la desviación estándar 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.

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.

Características 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. La sintaxis de comandos también le permite:

- 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 de matriz (mediante el subcomando OUTFILE).
- Escribir una matriz de estadísticos de confianza (varianzas y covarianzas) para los valores singulares y las puntuaciones en un archivo de datos de matriz (mediante el subcomando OUTFILE).
- Especificar varios conjuntos de categorías para igualar (mediante el subcomando EQUAL).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 6. 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.

Análisis de correspondencias múltiple: Consideraciones sobre los datos

Datos. Los valores de las variables de cadena se convierten en enteros positivos por orden alfabético ascendente. Los valores perdidos del 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

1. Seleccione en los menús:
Analizar > Reducción de dimensiones > Escalamiento óptimo...
2. Seleccione **Todas las variables son nominales múltiples.**
3. Seleccione **Un conjunto.**
4. Pulse en **Definir.**
5. Seleccione al menos dos variables de análisis y especifique el número de dimensiones en la solución.
6. 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.

Ponderación de la variable. Se puede definir una ponderación para cada variable. El valor especificado debe ser un entero positivo. El valor predeterminado 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 distintos 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. De forma predeterminada, las variables restantes se dejan inalteradas. A partir de ese momento, se utilizan en el análisis las variables discretizadas.

Método. Seleccione entre agrupación, clasificación y multiplicación.

- **Agrupación.** Se recodifica en un número especificado de categorías o se recodifica por intervalos.
- **Clasificación.** La variable se discretiza mediante la clasificación de 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.

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.

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 casos 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 cargas 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.

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 dinámica, mostrando las coordenadas de forma predeterminada. 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.

Análisis de correspondencias múltiple: Gráficos de objetos

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

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 que desee y las variables que se van a representar.

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; 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.

Características 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. La sintaxis de comandos también le permite:

- 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).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 7. 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 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.

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 similitud 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 estrés, descomposición del estrés, 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 estrés, 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 de residuos de Shepard y gráficos de transformación de las variables independientes.

Escalamiento multidimensional: Consideraciones sobre los datos

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 spline.

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 un origen, todos los modelos son equivalentes al modelo de identidad, por consiguiente, el análisis toma de forma predeterminada el modelo de identidad.

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

Para obtener un escalamiento multidimensional

1. Seleccione en los menús:

Analizar > Escala > Escalamiento multidimensional (PROXSCAL)...

Accederá al cuadro de diálogo Formato de datos.

2. 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 orígenes. Si los datos son proximidades, debe especificar si dispone de un único o de diversos orígenes de medidas de proximidad.

Un origen. Si hay un solo origen 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 entre 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.

Diversos orígenes. Si hay diversos orígenes 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 un origen por 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 orígenes. Lo que conduce al cuadro de diálogo Proximidades en matrices a través de columnas.
- *Las proximidades están en columnas, un origen por columna..* Las matrices de proximidad 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 el origen de cada casilla. Lo que conduce al cuadro de diálogo Proximidades en una columna.

3. Pulse en **Definir**.

Proximidades en matrices a través de columnas

Si selecciona el modelo para los datos con las proximidades en matrices para uno o varios orígenes en el cuadro de diálogo Formato de datos, realice lo siguiente:

1. 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.)
2. 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.)
3. Como alternativa, si existen diversos orígenes, seleccione una variable de orígenes. (El número de casos en cada variable de proximidades debe ser igual al número de variables de proximidades multiplicado por el número de orígenes.)

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 diversos orígenes en el cuadro de diálogo Formato de datos, realice lo siguiente:

1. Seleccione dos o más variables de proximidades. (Se asume que cada variable es una matriz de proximidades de un origen diferente.)
2. Seleccione una variable de fila para definir las ubicaciones de fila para las proximidades en cada variable de proximidades.
3. 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 perdidas.)
4. 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 varios orígenes en el cuadro de diálogo Formato de datos, realice lo siguiente:

1. Seleccione una variable de proximidades. (Se asume que es una o más matrices de proximidades.)
2. Seleccione una variable de fila para definir las ubicaciones de fila para las proximidades en la variable de proximidades.
3. Seleccione una variable de columna para definir las ubicaciones de columna para las proximidades en la variable de proximidades.
4. Si existen diversos orígenes, seleccione una variable de orígenes. (Para cada origen, las casillas de la matriz de proximidades que no reciban una designación fila/columna se tratarán como perdidas.)
5. 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, realice lo siguiente:

1. 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 diversos orígenes). Si crea las distancias entre los casos, sólo será necesaria una variable.
2. Si existen diversos orígenes, seleccione una variable de orígenes.
3. 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

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.
- **Binario.** Distancia euclídea, Distancia euclídea al cuadrado, Diferencia de tamaño, Diferencia de configuración, Varianza o Lance y Williams.

Crear matriz de proximidades. 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 origen por separado o de forma incondicional en el origen.

Modelo de escalamiento. Seleccione entre las siguientes alternativas:

- **Identidad.** Todos los orígenes tienen la misma configuración.
- **Euclídeo ponderado.** Éste es un modelo de diferencias individuales. Cada origen 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 origen 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.
- **Spline.** 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 origen se comparan entre sí o si las comparaciones son incondicionales respecto al origen.

Dimensiones. De forma predeterminada, 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.

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 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.
- **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 spline 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.

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 estrés 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 estrés.** El algoritmo detendrá la iteración cuando la diferencia entre valores consecutivos del estrés bruto normalizado sea menor que el número aquí especificado, que debe estar entre 0,0 y 1,0.
- **Estrés mínimo.** El algoritmo detendrá la iteración cuando el estrés bruto normalizado sea menor que el número aquí especificado, que debe estar entre 0,0 y 1,0.
- **Número máximo de iteraciones.** 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. Este tema describe el cuadro de diálogo Gráficos si dispone del formato de datos Proximidades en columnas. Para los gráficos **Ponderaciones del espacio individual**, **Proximidades originales frente a transformadas** y **Proximidades transformadas frente a distancias**, puede especificar los orígenes para los que se generarán los gráficos. La lista de orígenes disponibles es la lista de variables de proximidades en el cuadro de diálogo principal.

Estrés. Se genera un gráfico del estrés 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 origen, 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. Este tema describe el cuadro de diálogo Gráficos si dispone de un formato de datos distinto a Proximidades en columnas. Para los gráficos **Ponderaciones del espacio individual**, **Proximidades originales frente a transformadas** y **Proximidades transformadas frente a distancias**, puede especificar los orígenes para los que se generarán los gráficos. Los números de orígenes introducidos deben ser valores de la variable de orígenes especificada en el cuadro de diálogo principal y el rango debe estar entre 1 y el número de orígenes.

Escalamiento multidimensional: Resultados

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

Representación. 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 de 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.
- **Estrés para inicios aleatorios.** Muestra la semilla de aleatorización y el valor del estrés bruto normalizado para cada inicio aleatorio.
- **Historial de iteraciones.** Muestra el historial de iteraciones del algoritmo principal.
- **Diversas medidas del estrés.** Muestra distintos valores del estrés. La tabla contiene valores para el estrés bruto normalizado, Stress-I, Stress-II, S-Stress, la Dispersión explicada (DAF),
- **Descomposición del estrés.** Muestra la descomposición por orígenes y por objetos del estrés 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.

Características 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. La sintaxis de comandos también le permite:

- Especificar una lista de variables distinta para los gráficos de transformación y los de residuos (con el subcomando PLOT).
- Especificar listas de orígenes 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).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Capítulo 8. 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 despliegamiento 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 similaridades. Utilice el despliegamiento multidimensional para encontrar clústeres 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 estrés, descomposición del estrés, coordenadas del espacio común, distancias entre objetos dentro de la configuración final, ponderaciones del espacio individual, espacios individuales, proximidades transformadas, gráficos del estrés, 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 de residuos de Shepard.

Desplegamiento multidimensional: Consideraciones sobre los datos

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 diversos orígenes 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 un origen, todos los modelos son equivalentes al modelo de identidad, por consiguiente, el análisis toma de forma predeterminada el modelo de identidad.

Para obtener un despliegamiento multidimensional

1. Seleccione en los menús:
Analizar > Escala > Desplegamiento multidimensional (PREFSCAL)...
2. Seleccione dos o más variables que identifiquen las columnas en la matriz de proximidades rectangular. Cada variable representa un objeto de columna distinto.
3. 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.
4. 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.
5. Si existen diversos orígenes, puede seleccionar una variable de orígenes, 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 orígenes.

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 el origen o de forma incondicional en la fila.

Modelo de escalamiento. Seleccione entre las siguientes alternativas:

- **Identidad.** Todos los orígenes tienen la misma configuración.
- **Euclídeo ponderado.** Éste es un modelo de diferencias individuales. Cada origen 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 origen 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 similaridad o disimilaridad.

Dimensiones. De forma predeterminada, 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:

- **Ninguna.** 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 inclinación 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.
- **Spline.** 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 origen, o las comparaciones será incondicionales en la fila o en el origen, es decir, si las transformaciones se realizarán por fila, por origen o sobre todas las proximidades a la vez.

Restricciones del despliegamiento multidimensional

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

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 estrés.

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 en valores singulares 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 el estrés penalizado 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 estrés.** El algoritmo detendrá la iteración cuando la diferencia relativa entre valores consecutivos de estrés penalizado sea menor que el número aquí especificado, que debe ser no negativo.
- **Estrés mínimo.** El algoritmo detendrá la iteración cuando el estrés penalizado sea menor que el número aquí especificado, que debe ser no negativo.
- **Número máximo de iteraciones.** 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 el estrés penalizado, 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.

Diagramas. Se encuentran disponibles los siguientes gráficos:

- **Inicios múltiples.** Muestra un histograma apilado de estrés penalizado que indica tanto el estrés como la penalización.
- **Espacio común inicial.** Muestra un diagrama de dispersión matricial de las coordenadas del espacio común inicial.
- **Estrés por dimensión.** Genera un gráfico lineal de estrés penalizado 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 todos los orígenes 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 origen.
- **Espacios individuales.** Se muestra un diagrama de dispersión matricial de las coordenadas del espacio individual de cada origen. 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 u origen. 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 u origen. 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 diversos orígenes, se asignará un color distinto a cada uno.
- **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 diversos orígenes, se asignará un color distinto a cada uno.

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 orígenes. 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 los orígenes para los que se deben generar los gráficos. Los números de orígenes introducidos deben ser valores de la variable de orígenes especificada en el cuadro de diálogo principal y el rango debe estar entre 1 y el número de orígenes.

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.

Representación. 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 del estrés penalizado 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 estrés.** Muestra la descomposición de objetos, filas y orígenes de estrés penalizado, incluidas las desviaciones estándar 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.

Características 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. La sintaxis de comandos también le permite:

- Especifique varias listas de origen 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 orígenes en lugar de una variable de ID de fuente (con el subcomando INPUT).

Consulte la *Referencia de sintaxis de comandos* para obtener información completa de la sintaxis.

Avisos

Esta información se ha desarrollado para productos y servicios ofrecidos en los EE.UU.

Es posible que IBM no ofrezca los productos, servicios o características que se tratan en este documento en otros países. El representante local de IBM le puede informar sobre los productos y servicios que están actualmente disponibles en su localidad. Cualquier referencia a un producto, programa o servicio de IBM no pretende afirmar ni implicar que solamente se pueda utilizar ese producto, programa o servicio de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja 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 solicitudes de patente en tramitación que cubran la materia descrita en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
EE.UU.

Para consultas sobre licencias relacionadas con información de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus consultas, por escrito, a:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502, Japón

El siguiente párrafo no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles con la legislación vigente: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO DETERMINADO. 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 en cualquier momento mejoras o cambios en los productos o programas descritos en esta publicación sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales de estos sitios Web no forman parte de los materiales destinados a este producto de IBM, y el usuario será responsable del uso que se haga de estos sitios Web.

IBM puede utilizar o distribuir la información que se le proporcione de la forma que considere adecuada, sin incurrir en ninguna obligación con el cliente.

Los propietarios de licencia de este programa que deseen tener información sobre el mismo con el fin de: (i) intercambiar información entre programas creados de forma independiente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deberán ponerse en contacto con:

Tel. 901 100 400
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
EE.UU.

Esta información estará disponible, bajo las condiciones adecuadas, incluyendo en algunos casos el pago de una cuota.

El programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible los proporciona IBM bajo los términos de las Condiciones Generales de IBM, Acuerdo Internacional de Programas Bajo Licencia de IBM o cualquier acuerdo equivalente entre las partes.

Cualquier dato de rendimiento mencionado aquí ha sido determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en otros entornos operativos pueden variar de forma significativa. Es posible que algunas mediciones se hayan realizado en sistemas en desarrollo y no existe ninguna garantía de que estas mediciones sean las mismas en los sistemas comerciales. Además, es posible que algunas mediciones hayan sido estimadas a través de extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben consultar los datos que corresponden a su entorno específico.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado esos productos y no puede confirmar la exactitud del rendimiento, de la compatibilidad ni de ninguna otra declaración 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.

Todas las declaraciones sobre el futuro del rumbo y la intención de IBM están sujetas a cambio o retirada sin previo aviso y representan únicamente metas y objetivos.

Esta información contiene ejemplos de datos e informes utilizados en operaciones comerciales diarias. Para ilustrarlos lo máximo posible, los ejemplos incluyen los nombres de las personas, empresas, marcas y productos. Todos esos nombres son ficticios y cualquier parecido con los nombres y direcciones utilizados por una empresa real es pura coincidencia.

LICENCIA DE DERECHOS DE AUTOR:

Esta información contiene programas de aplicación de muestra escritos en lenguaje fuente, los cuales muestran técnicas de programación en diversas plataformas operativas. Puede copiar, modificar y distribuir estos programas de muestra de cualquier modo sin realizar ningún pago a IBM, con el fin de desarrollar, utilizar, comercializar o distribuir programas de aplicación que se ajusten a la interfaz de programación de aplicaciones para la plataforma operativa para la que se han escrito los programas de muestra. Estos ejemplos no se han probado exhaustivamente en todas las condiciones. Por lo tanto, IBM no puede garantizar ni dar por supuesta la fiabilidad, la capacidad de servicio ni la funcionalidad de estos programas. Los programas de muestra se proporcionan "TAL CUAL" sin garantía de ningún tipo. IBM no será responsable de ningún daño derivado del uso de los programas de muestra.

Cada copia o fragmento de estos programas de ejemplo o de cualquier trabajo derivado de ellos, debe incluir el siguiente aviso de copyright:

© (nombre de la compañía) (año). Algunas partes de este código procede de los programas de ejemplo de IBM Corp.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas o marcas comerciales registradas de International Business Machines Corp., registrada en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios podrían ser marcas registradas de IBM u otras compañías. Encontrará una lista actual de marcas registradas de IBM en la Web en "Información de copyright y marca registrada" en www.ibm.com/legal/copytrade.shtml.

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.

Linux es una marca registrada de Linus Torvalds en Estados Unidos, otros países o ambos.

Microsoft, Windows, Windows NT, y el logotipo de Windows son marcas comerciales de Microsoft Corporation en Estados Unidos, otros países o ambos.

UNIX es una marca registrada de The Open Group en Estados Unidos y otros países.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales o registradas de Oracle y/o sus afiliados.

Índice

A

- actualizaciones relajadas
 - en Escalamiento multidimensional 47
- ajuste
 - en Análisis de correlación canónica no lineal 29
- Análisis de componentes principales
 - categorico 19, 22
 - almacenamiento de variables 24
 - Características adicionales del comando 26
 - nivel de escalamiento óptimo 20
 - simulación de muestreo 25
- Análisis de correlación canónica no lineal 27, 28, 29
 - Características adicionales del comando 30
 - estadísticos 27
 - gráficos 27
- Análisis de correspondencias 31, 32, 33, 34
 - Características adicionales del comando 35
 - estadísticos 31
 - gráficos 31
- Análisis de correspondencias múltiple 37, 39
 - almacenamiento de variables 40
 - Características adicionales del comando 41
 - nivel de escalamiento óptimo 38
- ANOVA
 - en Regresión categórica 16

C

- cargas en componentes
 - en Análisis de componentes principales categorico 23
 - en Análisis de correlación canónica no lineal 29
- centroides
 - en Análisis de correlación canónica no lineal 29
- coeficientes de regresión
 - en Regresión categórica 16
- configuración inicial
 - en Análisis de correlación canónica no lineal 29
 - en despliegamiento multidimensional 53
 - en Escalamiento multidimensional 47
 - en Regresión categórica 15
- coordenadas del espacio común
 - en despliegamiento multidimensional 55
 - en Escalamiento multidimensional 48
- coordenadas del espacio individual
 - en despliegamiento multidimensional 55

- correlaciones
 - en Escalamiento multidimensional 48
- criterios de iteración
 - en despliegamiento multidimensional 53
 - en Escalamiento multidimensional 47
- cuantificaciones de categorías
 - en Análisis de componentes principales categorico 23
 - en Análisis de correlación canónica no lineal 29
 - en el análisis de correspondencias múltiple 40
 - en Regresión categórica 16

D

- Despliegamiento multidimensional 51
 - Características adicionales del comando 55
 - estadísticos 51
 - gráficos 51, 54
 - modelo 52
 - opciones 53
 - restricciones en el espacio común 52
 - salida 55
- diagrama de dispersión del ajuste
 - en despliegamiento multidimensional 54
- diagramas de dispersión biespacial
 - en Análisis de componentes principales categorico 24
 - en Análisis de correspondencias 34
 - en el análisis de correspondencias múltiple 41
- diagramas de dispersión trispacial
 - en Análisis de componentes principales categorico 24
- dimensiones
 - en Análisis de correspondencias 33
- discretización
 - en Análisis de componentes principales categorico 21
 - en el análisis de correspondencias múltiple 38
 - en Regresión categórica 14
- distancias
 - en despliegamiento multidimensional 55
 - en Escalamiento multidimensional 48

E

- Escalamiento multidimensional 43, 44, 45
 - Características adicionales del comando 49
 - estadísticos 43
 - gráficos 43, 48
 - modelo 46

Escalamiento multidimensional (continuación)

- opciones 47
 - restricciones 47
 - salida 48
- estadísticos de confianza
 - en Análisis de correspondencias 34
 - estadísticos descriptivos
 - en Regresión categórica 16

G

- gráficos
 - en Análisis de correlación canónica no lineal 29
 - en Análisis de correspondencias 34
 - en Escalamiento multidimensional 48
 - en Regresión categórica 18
- gráficos de cargas en componentes
 - en Análisis de componentes principales categorico 25
- gráficos de categorías
 - en Análisis de componentes principales categorico 25
 - en el análisis de correspondencias múltiple 41
- gráficos de categorías conjuntas
 - en Análisis de componentes principales categorico 25
 - en el análisis de correspondencias múltiple 41
- gráficos de centroides proyectados
 - en Análisis de componentes principales categorico 25
- gráficos de correlaciones
 - en Escalamiento multidimensional 48
- gráficos de espacio común inicial
 - en despliegamiento multidimensional 54
- gráficos de espacios individuales
 - en despliegamiento multidimensional 54
 - en Escalamiento multidimensional 48
- gráficos de estrés
 - en despliegamiento multidimensional 54
 - en Escalamiento multidimensional 48
- gráficos de inicio múltiple
 - en despliegamiento multidimensional 54
- gráficos de medidas de discriminación
 - en el análisis de correspondencias múltiple 41
- gráficos de ponderaciones del espacio
 - en despliegamiento multidimensional 54
- gráficos de ponderaciones del espacio individual
 - en despliegamiento multidimensional 54
 - en Escalamiento multidimensional 48

- gráficos de puntos de objetos
 - en Análisis de componentes principales categórico 24
 - en el análisis de correspondencias múltiple 41
- gráficos de residuos
 - en desplegamiento multidimensional 54
- Gráficos de Shepard
 - en desplegamiento multidimensional 54
- gráficos de transformación
 - en Análisis de componentes principales categórico 25
 - en desplegamiento multidimensional 54
 - en el análisis de correspondencias múltiple 41
 - en Escalamiento multidimensional 48
- gráficos del espacio común
 - en desplegamiento multidimensional 54
 - en Escalamiento multidimensional 48
- gráficos del espacio común final
 - en desplegamiento multidimensional 54

H

- historial de iteraciones
 - en Análisis de componentes principales categórico 23
 - en desplegamiento multidimensional 55
 - en el análisis de correspondencias múltiple 40
 - en Escalamiento multidimensional 48

I

- inercia
 - en Análisis de correspondencias 34

L

- lazo
 - en Regresión categórica 16

M

- matriz de correlaciones
 - en Análisis de componentes principales categórico 23
 - en el análisis de correspondencias múltiple 40
- medidas de discriminación
 - en el análisis de correspondencias múltiple 40
- medidas de distancia
 - en Análisis de correspondencias 33
- medidas del estrés
 - en desplegamiento multidimensional 55
 - en Escalamiento multidimensional 48

- modelo de escalamiento
 - en desplegamiento multidimensional 52
- modelo de identidad
 - en desplegamiento multidimensional 52
- modelo Euclídeo generalizado
 - en desplegamiento multidimensional 52
- modelo Euclídeo ponderado
 - en desplegamiento multidimensional 52

N

- nivel de escalamiento óptimo
 - en Análisis de componentes principales categórico 20
 - en el análisis de correspondencias múltiple 38
- normalización
 - en Análisis de correspondencias 33

O

- objetos suplementarios
 - en Regresión categórica 15

P

- ponderación de la variable
 - en Análisis de componentes principales categórico 20
 - en el análisis de correspondencias múltiple 38
- ponderaciones
 - en Análisis de correlación canónica no lineal 29
- ponderaciones del espacio individual.
 - en desplegamiento multidimensional 55
 - en Escalamiento multidimensional 48
- PREFSCAL 51
- proximidades transformadas
 - en desplegamiento multidimensional 55
 - en Escalamiento multidimensional 48
- puntuaciones de los objetos
 - en Análisis de componentes principales categórico 23
 - en Análisis de correlación canónica no lineal 29
 - en el análisis de correspondencias múltiple 40

R

- R múltiple
 - en Regresión categórica 16
- red elástica
 - en Regresión categórica 16
- Regresión categórica 13
 - Características adicionales del comando 18
 - estadísticos 13

- Regresión categórica (*continuación*)
 - gráficos 13
 - guardar 17
 - nivel de escalamiento óptimo 14
 - regularización 16
- regresión Ridge
 - en Regresión categórica 16
- restricciones
 - en Escalamiento multidimensional 47
- restricciones en el espacio común
 - en desplegamiento multidimensional 52

S

- simulación de muestreo
 - Análisis de componentes principales categórico 25

T

- término de penalización
 - en desplegamiento multidimensional 53
- tipificación
 - en Análisis de correspondencias 33
- transformaciones de proximidades
 - en desplegamiento multidimensional 52

V

- valores perdidos
 - en Análisis de componentes principales categórico 21
 - en el análisis de correspondencias múltiple 38
 - en Regresión categórica 15
- variables independientes transformadas
 - en Escalamiento multidimensional 48
- varianza explicada
 - en Análisis de componentes principales categórico 23



Impreso en España