

CONTENIDO

Capítulo 1. Instalación y primeros pasos en SPSS 15	1
Instalación de SPSS 15.....	1
Comenzando con SPSS 15	7
El entorno de trabajo de SPSS 15.....	8
Opciones de la barra de menú principal	10
Estructura básica de los archivos de datos: casos y variables	16
Introducción de datos y tipos de datos	16
Etiquetas de valores y de variables	18
Copiar y pegar atributos y etiquetas de variables.....	20
Capítulo 2. Enfoque moderno del análisis de datos.....	23
Técnicas emergentes del análisis de datos	23
Modelado originado por la teoría (técnicas explicativas o de dependencia).....	28
Modelado originado por los datos (técnicas descriptivas o de interdependencia)	26
Extracción del conocimiento	27
Afinando concepto y clasificación de las técnicas de minería de datos	31
Capítulo 3. Fase de selección en el análisis de datos. Tratamiento con SPSS...	33
Fase de selección en el análisis de datos	33
Selección de datos por muestreo	34
Muestreo aleatorio simple	38
Muestreo estratificado	41
Muestreo sistemático.....	48
Muestreo unietápico de conglomerados	52
Muestreo bietápico de conglomerados	56
Muestreo polietápico de conglomerados.....	59
Diseños complejos: bietápico con estratificación en primera etapa.....	59
Técnicas de muestro con SPSS	60
Diseños complejos y el asistente de muestreo. Creación de un nuevo plan de muestreo.....	62
Asistente de muestreo: Modificar un plan.....	70
Asistente de muestreo: Ejecutar un plan	73
Preparación de una muestra compleja para su análisis:	
creación de un plan de análisis	73
Preparación de una muestra compleja para su análisis:	
modificar un plan de análisis.....	77
Cálculos en muestras complejas:	

VIII Contenido

frecuencias, descriptivos, tablas de contingencia y razones	77
Selección de casos con SPSS	83
Selección de casos con criterios condicionales	83
Selección de fechas, horas y filas	84
Selección de una muestra aleatoria	85
Semilla de aleatorización	85
Operadores de selección en SPSS	86
Operadores aritméticos	86
Operadores relacionales	87
Operadores aritméticos	86
Operadores lógicos	87
Funciones de generación de números aleatorios en SPSS	87
Capítulo 4. Fase de exploración en el análisis de datos.	
Tratamiento con SPSS	91
Exploración en el proceso de extracción del conocimiento	91
Análisis exploratorio	92
Herramientas de exploración visual	93
Histograma de frecuencias	93
Diagrama de tallo y hojas	95
Gráfico de caja y bigotes	96
Gráfico múltiple de caja y bigotes	98
Gráfico de simetría	100
Gráfico de dispersión	102
Gráficos para variables cualitativas	104
Herramientas de exploración formal	107
Contraste de la bondad de ajuste a una distribución: test de la Chi-cuadrado	107
Contraste de Kolmogorov-Smirnov Lilliefors de la bondad de ajuste a una distribución	109
Estadísticos robustos de centralización	110
Estadísticos robustos de dispersión	111
Estadísticos robustos de asimetría y curtosis	113
Contrastes de aleatoriedad	116
Transformaciones de variables	120
Supuestos subyacentes en las técnicas de minería de datos	121
Normalidad	121
Heteroscedasticidad	125
Multicolinealidad	127
Autocorrelación	127
Linealidad	128
Supuestos subyacentes en las técnicas de minería de datos	121
Un ejemplo de análisis exploratorio de datos con SPSS	129
Capítulo 5. Análisis exploratorio, gráfico y formal con SPSS	139

Análisis exploratorio de datos con SPSS. Procedimiento Explorar	139
Análisis exploratorio gráfico en SPSS	144
Generador de gráficos	144
Histogramas.....	145
Gráficos de normalidad	145
Gráficos de caja y bigotes	148
Gráfico de dispersión	150
Gráficos para variables cualitativas.....	104
Gráficos interactivos dinámicos de análisis exploratorio con SPSS	152
Creación interactiva de gráficos a partir de tablas	159
Gráficos interactivos de caja y bigotes.....	160
Histogramas interactivos	161
Diagramas interactivos de dispersión.....	163
Análisis exploratorio formal en SPSS. Contrastes de aleatoriedad y rachas	165
Contraste de ajuste a una distribución de frecuencias.	
Procedimiento Prueba de Kolmogorov-Smirnov	166
Capítulo 6. Fase de limpieza y transformación de datos.	
Tratamiento con SPSS	167
Limpieza y transformación de datos	167
Valores atípicos (<i>outliers</i>)	168
Información faltante (datos <i>missing</i>).....	172
Soluciones para los datos ausentes: supresión de datos o imputación de la información faltante.....	178
Transformación de datos	180
Transponer, fusionar, agregar, segmentar y ordenar archivos	181
Ponderar casos y categorizar y numerizar variables	182
Pareamiento o <i>matching</i>	182
SPSS y el análisis de datos missing. Imputación	184
Reemplazar valores perdidos	187
Detección de valores atípicos en SPSS	191
Detección de casos atípicos mediante gráficos de control	191
Detección de casos atípicos mediante gráficos de caja y bigotes.....	193
Transformación de datos con SPSS.....	195
Recodificación de variables	197
Ordenar casos	199
Transponer, fusionar, agregar y segmentar archivos. <i>Matching</i>	199
Ponderar casos.....	206
Categorizar variables: categorizador visual	207
Asignar rangos a casos y tipificar variables	211
Capítulo 7. Reducción de la dimensión con variables cuantitativas	213

X Contenido

Transformación de datos mediante técnicas de reducción de la dimensión	213
Componentes principales	214
Cálculo de las componentes principales.....	215
Puntuaciones o medición de las componentes	217
Número de componentes a retener	218
Matriz de cargas factoriales, comunalidad y círculos de correlación.....	219
Rotación de las componentes	220
Análisis factorial	221
Contrastes en el modelo factorial	224
Rotación de los factores	226
Interpretación gráfica de los factores	227
Puntuaciones o medición de los factores.....	228
Técnicas de reducción de la dimensión en SPSS	229
Componentes principales en SPSS.....	230
Análisis factorial en SPSS.....	241
<i>Capítulo 8. Reducción de la dimensión con variables cualitativas: correspondencias, análisis conjunto y escalamiento</i>	251
Categorías en SPSS y métodos de reducción de la dimensión.....	251
Análisis de correspondencias	252
Análisis de correspondencias simple.....	252
Análisis de correspondencias múltiple	253
Escalamiento óptimo	254
Escalamiento multidimensional	256
Modelos de escalamiento métrico	257
Modelos de escalamiento no métrico	257
Modelos de escalamiento de diferencias individuales	257
Modelos de escalamiento para datos de preferencia	258
Análisis conjunto.....	259
Análisis de correspondencias simple en SPSS	260
Análisis de correspondencias múltiple en SPSS	267
Componentes principales no lineales en SPSS.....	271
Correlación canónica no lineal en SPSS	275
Escalamiento multidimensional métrico en SPSS.....	279
Escalamiento multidimensional no métrico en SPSS.....	282
Análisis conjunto en SPSS	287
Escalamiento multidimensional métrico en SPSS.....	279

Capítulo 9. Técnicas predictivas. Regresión lineal, no lineal y categórica....	299
Técnicas predictivas para la modelización	299
Modelo de regresión lineal múltiple.....	304
Estimación del modelo, contrastes e intervalos de confianza	306
Análisis de la varianza.....	308
Predicciones	311
El análisis de los residuos.....	312
Autocorrelación, multicolinealidad y heteroscedasticidad	313
Modelos no lineales y ponderados	316
Regresión ordinal y categórica.....	316
Procedimiento Regresión lineal	317
Procedimiento Estimación ponderada y la homoscedasticidad.....	322
Procedimiento Mínimos cuadrados en dos fases. Variables instrumentales	324
Procedimiento Regresión ordinal	325
Procedimiento Regresión mediante escalamiento óptimo.....	328
Procedimiento Estimación curvilínea.....	331
Procedimiento Regresión no lineal	333
Capítulo 10. Modelos del análisis de la varianza y la covarianza.	
Modelo lineal general MLG.....	363
Modelos del análisis de la varianza ANOVA	363
Modelos de la covarianza ANCOVA.....	372
Análisis de la varianza múltiple MANOVA	374
Análisis de la covarianza múltiple MANCOVA	375
Modelo Lineal General (GLM)	375
SPSS y el procedimiento ANOVA de un factor.....	375
SPSS y el procedimiento MLG univariante	380
SPSS y el procedimiento MLG multivariante.....	386
SPSS y el procedimiento MLG medidas repetidas	391
SPSS y el procedimiento Componentes de la varianza.....	397
Capítulo 11. Modelos con censura, modelos de supervivencia, duración y Regresión de Cox.....	569
Análisis de la supervivencia	415
Tablas de mortalidad	416
Estimaciones no paramétricas de la función de supervivencia	417
Estimaciones paramétricas de la función de supervivencia	418
Procedimiento Tablas de mortalidad.....	419
Procedimiento Kaplan-Meier	422
Procedimiento Regresión de Cox y covariable dependiente del tiempo	426

Capítulo 12. Modelos mixtos y modelos con datos de panel.....	435
Introducción a los modelos mixtos.....	435
Modelos con datos de panel	436
Modelos de panel con coeficientes constantes	436
Modelos de panel con efectos fijos.....	437
Modelos de panel con coeficientes aleatorios	438
Procedimiento modelos mixtos de SPSS.....	439
Capítulo 13. Modelos en el contexto temporal.	
Análisis de series temporales	453
Predicción y alisado de series.....	453
Alisado Exponencial de Brown.....	454
Alisado Lineal de Holt	455
Alisado Estacional de Winters	455
Análisis de la tendencia.....	455
Modelos ARIMA de Box-Jenkins.....	456
Fases del modelado	456
Modelos AR(p), MA(q), ARMA(p,q) y ARIMA(p,d,q)	458
Modelos de la función de transferencia.....	460
Identificación del modelo	461
Modelos ARIMA(p,d,q)(P,D,Q) estacionales	462
Predicción.....	464
Procedimiento Suavizado exponencial.....	465
Procedimiento Autorregresión	467
Procedimiento ARIMA	468
Procedimiento Descomposición estacional	473
Capítulo 14. Modelos de clasificación. Modelos Logit y Probit	487
Modelos de elección discreta	487
El modelo de regresión logística	488
Estimación de los coeficientes	489
Estimación por intervalos y contrastes de hipótesis sobre los coeficientes ..	490
Modelos Probit	491
SPSS y la regresión logística binaria.....	492
SPSS y la regresión logística multinomial	500
SPSS y los modelos Probit	505

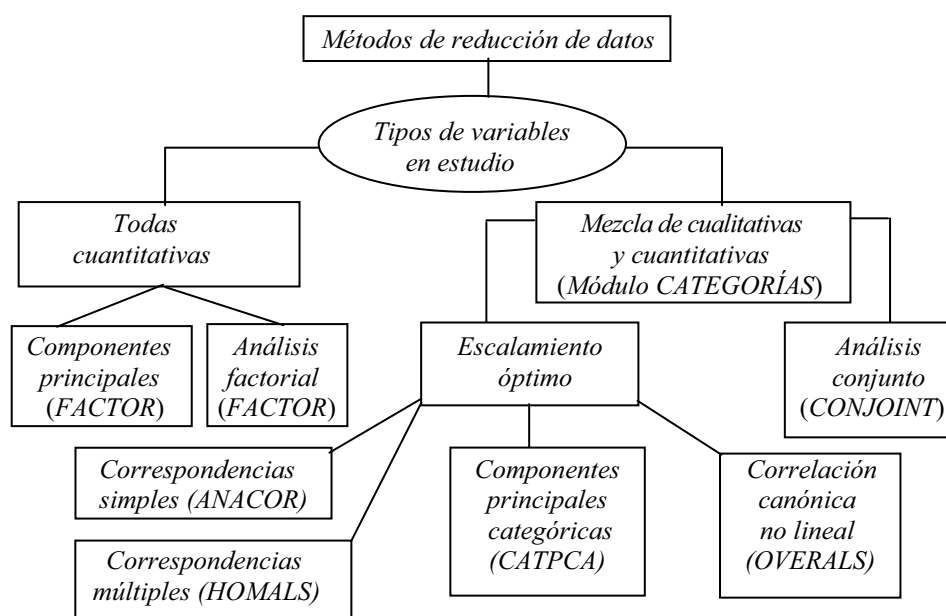
Capítulo 15. Modelos de clasificación ad-hoc. Análisis discriminante....	523
Modelos de análisis discriminante	523
Hipótesis en el modelo discriminante	524
Estimación en el modelo discriminante.....	525
Contrastes de significación en el modelo discriminante	527
Selección de variables discriminantes.....	531
Interpretación de la función discriminante.....	534
Clasificación de los individuos	537
Análisis discriminante canónico.....	540
SPSS y el análisis discriminante	542
Capítulo 16. Técnicas de clasificación post-hoc. Análisis clúster y árboles de decisión	557
El análisis clúster como técnica descriptiva de clasificación	557
Medidas de similitud	558
Técnicas en el análisis clúster	562
Clústers jerárquicos, secuenciales, aglomerativos y exclusivos (S.A.H.N.)...	564
El dendograma en el análisis cluster jerárquico	565
Análisis clúster no jerárquico	566
Los árboles de decisión como técnica predictiva de clasificación	570
Características de los árboles de decisión	571
Herramientas para el trabajo con los árboles de decisión	576
Árboles CHAID	577
Árboles CART.....	578
Árboles QUEST	580
Análisis de conglomerados y árboles de decisión como métodos de segmentación.....	581
Análisis clúster jerárquico con SPSS	582
Análisis clúster no jerárquico con SPSS	588
Análisis clúster en dos fases con SPSS	593
Árboles de decisión con SPSS	599
Creación de un árbol de decisión. Método CHAID	601
Métodos CR y QUEST. Poda de árboles	6078
Árboles QUEST	580
Capítulo 17. Modelos Log-lineales y técnicas de asociación.....	611
Introducción y conceptos	611
Tipos de modelos log-lineales.....	612
Modelo saturado.....	612
Modelo de independencia.....	613

XIV Contenido

Modelos jerárquicos	613
Fases en la elaboración de modelos log-lineales.....	613
Selección del modelo	614
Ajuste del modelo.....	614
Elección del modelo más adecuado.....	615
Estimación de parámetros	616
Significatividad de los efectos: Prueba de los efectos k y prueba de asociación parcial	616
Calidad de ajuste del modelo mediante residuales	617
Ajuste del modelo paso a paso (stepwise).....	617
Interpretación del modelo.....	618
Modelo Logit.....	618
Modelos log-lineales con datos ordinales	619
Tablas incompletas y ceros estructurales	619
SPSS y los modelo log-lineales.....	620
Selección del modelo	620
Análisis log-lineal general.....	620
Análisis logit	620
Asociación con variables cuantitativas. Independencia	636
Independencia y asociación con variables cualitativas. Coeficientes	639
El procedimiento Correlaciones bivariadas con SPSS	648
El procedimiento Correlaciones parciales con SPSS	650
El procedimiento Distancias de SPSS	652
El procedimiento Tablas de contingencia	654
 <i>Capítulo 18. Miscelánea de funciones</i>	 671
Funciones en SPSS 15.....	671
Funciones exponenciales y logarítmicas	671
Funciones trigonométricas	672
Funciones numéricas	672
Funciones de estadística descriptiva	672
Funciones de generación de números aleatorios	673
Funciones de distribución y sus inversas	675
Funciones de densidad	680
Funciones de cadena.....	682
Funciones lógicas	684
Funciones de valores desaparecidos.....	684
Funciones de conversión	685

8.1 Categorías en SPSS y métodos de reducción de la dimensión

SPSS habilita el módulo CATEGORÍAS para abordar los métodos de reducción de la dimensión que usen variables categóricas. Se presenta una clasificación de los procedimientos incluidos en el módulo CATEGORÍAS.



8.2 Análisis de correspondencias

El análisis factorial, al igual que el análisis en componentes principales, es una técnica multivariante que persigue *reducir la dimensión de una tabla de datos* formada por *variables cuantitativas*. Si las variables fuesen *variables cualitativas*, estaríamos ante el análisis de correspondencias.

Cuando se estudia conjuntamente el comportamiento de dos variables cualitativas estamos ante el *análisis de correspondencias simples*, pero este análisis puede ser generalizado para el caso en que se dispone de un número de variables cualitativas mayor que dos, en cuyo caso estamos ante el *análisis de correspondencias múltiples*. En el caso de correspondencias simples los datos de las dos variables cualitativas pueden representarse en una tabla de doble

entrada, denominada *tabla de contingencia*. En el caso de las correspondencias múltiples la tabla de contingencia de doble entrada pasa a ser una hipertabla en tres o más dimensiones, difícil de representar y que suele sintetizarse en la denominada *tabla de Burt*.

El objetivo del análisis de correspondencias es establecer relaciones entre variables no métricas enriqueciendo la información que ofrecen las tablas de contingencia, que sólo comprueban si existe alguna relación entre las variables (test de la chi-cuadrado, etc.) y la intensidad de dicha relación (test V de Cramer, etc.). El análisis de correspondencias revela además en qué grado contribuyen a esa relación detectada los distintos valores de las variables, información que suele ser proporcionada en modo gráfico (valores asociados próximos).

Podríamos sintetizar diciendo que el análisis de correspondencias busca como objetivo el estudio de la asociación entre las categorías de múltiples variables no métricas, pudiendo obtenerse un mapa perceptual que ponga de manifiesto esta asociación en modo gráfico.

Análisis de correspondencias simple

Ya sabemos que el análisis factorial de correspondencias simple está particularmente adaptado para tratar tablas de contingencia, representando los efectivos existentes en las múltiples modalidades (categorías) combinadas de dos caracteres (variables cualitativas). Si cruzamos en una tabla de contingencia el carácter I con modalidades desde $i = 1$ hasta $i = n$ (en filas), con el carácter J con modalidades desde $j = 1$ hasta $j = p$ (en columnas), podemos representar el número de unidades estadísticas que pertenecen simultáneamente a la modalidad i del carácter I y a la modalidad j del carácter J mediante k_{ij} .

En este caso, la distinción entre observaciones y variables en el cuadro de doble entrada es artificial, pero, por similitud con componentes principales, suele hablarse a veces de individuos u observaciones cuando nos referimos al conjunto de las modalidades del carácter I (filas), y de variables cuando nos referimos al conjunto de las modalidades del carácter J (columnas), tal y como se observa en la tabla siguiente:

I	J					
	1	2	...	j	...	p
1						
2				⋮		
⋮						
i			...	k_{ij}	...	
⋮				⋮		
n						

De una forma general puede considerarse que los objetivos que se persiguen cuando se aplica el análisis factorial de correspondencias son similares a los perseguidos con la aplicación del análisis de componentes principales, y pueden resumirse en los dos puntos siguientes:

- Estudio de las relaciones existentes en el interior del conjunto de modalidades del carácter I y estudio de las relaciones existentes en el interior del conjunto de modalidades del carácter J .
- Estudio de las relaciones existentes entre las modalidades del carácter I y las modalidades del carácter J .

La tabla de datos (k_{ij}) es una matriz K de orden (n, p) donde k_{ij} representa la frecuencia absoluta de asociaciones entre los elementos i y j , es decir el número de veces que se presentan simultáneamente las modalidades i y j de los caracteres I y J .

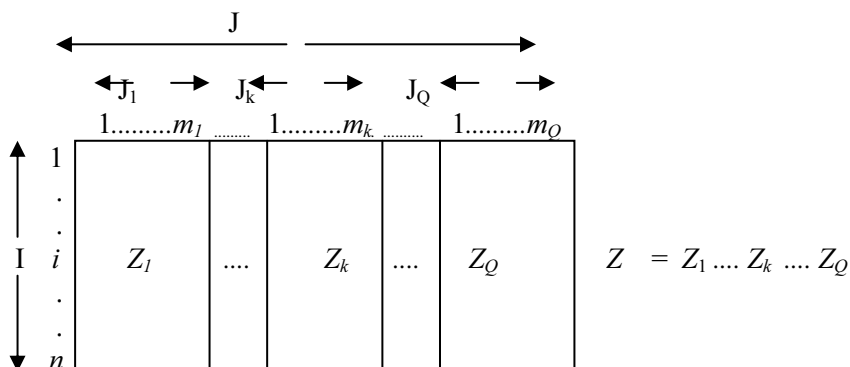
Análisis de correspondencias múltiple

Hemos visto que el análisis factorial de correspondencias es de aplicación con dos caracteres o variables cualitativas (*análisis de correspondencias simple* o sencillamente *análisis factorial de correspondencias*), cada una de las cuales puede presentar varias modalidades o categorías. Pero el método es generalizable al caso de un número de variables o caracteres cualitativos mayor de dos (*análisis de correspondencias múltiple*). Cuando el número de variables cualitativas es mayor que dos (en vez de tener sólo los caracteres I, J tenemos los caracteres J_1, J_2, \dots, J_Q) ya no se puede hablar de tabla de contingencia y la representación tabulada de los datos se complica. No obstante, el análisis en correspondencias múltiples permite estudiar las relaciones entre las modalidades de todas las características cualitativas consideradas.

En el análisis de correspondencias múltiples se ordenan los datos en una tabla Z denominada *tabla disyuntiva completa* que consta de un conjunto de individuos $I=1, \dots, i, \dots, n$ (en filas), un conjunto de variables o caracteres cualitativos $J_1, \dots, J_k, \dots, J_Q$ (en columnas) y un conjunto de modalidades excluyentes $1, \dots, m_k$ para

cada carácter cualitativo. El número total de modalidades será entonces $J = \sum_{k=1}^Q m_k$.

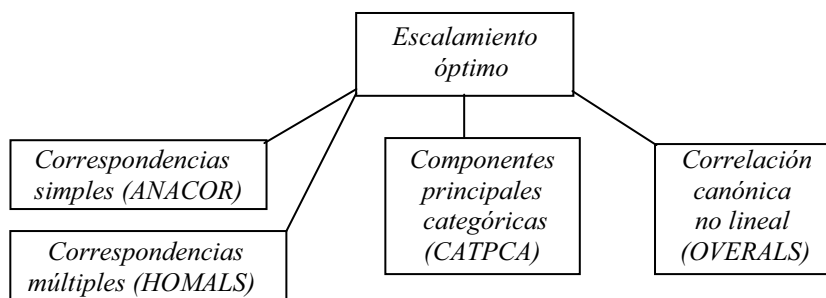
La tabla disyuntiva completa Z de dimensión $I \times J$ tiene el siguiente aspecto:



El elemento z_{ij} de la tabla toma el valor 0 o 1 según que el individuo i haya elegido (esté afectado por) la modalidad j o no. Por tanto, cada rectángulo de la tabla disyuntiva completa puede considerarse, aunque no lo sea, como una tabla de contingencia cuyos elementos son 0 o 1. La tabla disyuntiva completa Z consta entonces de Q subtablas yuxtapuestas, con la finalidad de obtener una representación simultánea de todas las modalidades (columnas) de todos los individuos (filas). Si las modalidades son excluyentes, cada subtabla tiene un único 1 en cada una de sus filas.

8.3 Escalamiento óptimo

El esquema de los procedimientos de escalamiento óptimo es el siguiente:



El escalamiento óptimo puede detectar relaciones no lineales y producir correlaciones máximas entre variables. Los cuatro procedimientos de escalamiento óptimo antes definidos amplían el ámbito de aplicación de las

técnicas estadísticas clásicas de Análisis de Componentes Principales (ACP) y de Análisis de Correlación Canónica (ACC), para acomodar variables de niveles mixtos de medida. Si todas las variables del análisis fuesen numéricas y las relaciones entre las variables lineales, entonces deberían emplearse los procedimientos estadísticos estándares basados en la correlación y no habría necesidad de utilizar los procedimientos de escalamiento óptimo. Sin embargo, si las variables de análisis tienen niveles mixtos de medida, o si se sospecha que existen relaciones no lineales entre algunos pares de variables, entonces debería utilizarse el procedimiento de escalamiento óptimo.

En el escalamiento óptimo, el usuario especifica el tipo de medida de cada variable, diferenciando el nivel de medida de cada una de las variables del análisis, permitiendo así la búsqueda de soluciones con el fin de que las variables elegidas por el modelo se ajusten bien a los datos. El escalamiento óptimo también revelará relaciones no lineales. Esto se hace de modo exploratorio, en contraposición con las pruebas de hipótesis estándar en el contexto de las suposiciones distributivas, tales como la normalidad y la linealidad de la regresión de las variables originales.

El escalamiento óptimo proporciona un conjunto de puntuaciones óptimas (o cuantificaciones de categorías), para las categorías de cada variable. Las puntuaciones óptimas se asignan a las categorías de cada variable, basadas en el criterio de optimización del procedimiento en uso. A diferencia de los valores originales de las variables nominales u ordinales del análisis, estas puntuaciones tienen propiedades métricas, por lo que éstas técnicas se describen frecuentemente como una forma de cuantificación de datos cualitativos, que también incluyen técnicas como el escalamiento no métrico multidimensional (disponible en el procedimiento ALSCAL). Las cuantificaciones de las categorías de cada variable pueden representarse sobre un plano bidimensional o, incluso, en un plano tridimensional, siendo su yuxtaposición en el mismo gráfico útil para revelar patrones de asociación entre variables.

En SPSS suelen utilizarse cuatro procedimientos relacionados con la ejecución del *Escalamiento Óptimo* que son los siguientes:

Análisis de Correspondencias Simples (ANACOR): analiza datos de tablas de contingencia de 2 dimensiones relativas a dos variables cualitativas.

Análisis de Correspondencias Múltiples u Homogeneidades (HOMALS): analiza datos de tablas de contingencia de múltiples dimensiones, donde todas las variables utilizadas son de nivel nominal y donde pueden ignorarse las interacciones de más dimensiones.

Análisis de Componentes Principales Categóricas (CATPCA): contabiliza los patrones de variación en un solo conjunto de variables de niveles de medición mixtos.

Análisis No Lineal de Correlación Canónica (OVERALS): contabiliza los patrones de variación cuando se correlacionan 2 o más conjuntos de variables de niveles de medición mixtos.

Estos procedimientos son técnicas de reducción de datos (dimensiones), que intentan representar las múltiples relaciones entre variables en un número de dimensiones reducido. Esto permite describir estructuras o patrones en las relaciones entre variables, difícilmente observables de otro modo. Estas técnicas pueden derivar en una forma de representación cartográfica perceptual (*perceptual mapping*). Una gran ventaja de estos procedimientos es que acomodan los datos a los diferentes niveles de medida.

8.4 Escalamiento multidimensional

El escalamiento multidimensional tiene como finalidad crear una representación gráfica (*mapa perceptual*) que permita conocer la situación de los individuos en un conjunto de objetos por posicionamiento de cada uno en relación a los demás. Dicha situación será producto de las percepciones y preferencias o similitudes entre los objetos apreciadas por los sujetos. Estas percepciones (preferencias o similitudes) son la entrada del análisis, y pueden ser variables métricas o no métricas. El escalamiento multidimensional transforma estas variables en distancias entre los objetos en un espacio de dimensiones múltiples, de modo que objetos que aparecen situados más próximos entre sí son percibidos como más similares por los sujetos.

Existe una diferencia clave entre el escalamiento multidimensional y el análisis cluster. En el escalamiento multidimensional se desconocen los elementos de juicio de los encuestados y no se conocen las variables que implícitamente están considerando éstos para realizar su evaluación de las preferencias por los objetos. En el análisis cluster las similitudes entre objetos se obtienen a partir de una combinación de variables estudiadas.

El escalamiento multidimensional es de más fácil aplicación que el análisis factorial, ya que no requiere supuestos de linealidad, ni que las variables sean métricas, ni un tamaño mínimo de muestra.

Resumiendo, podríamos definir el escalamiento multidimensional como una técnica cuyo fin es elaborar una representación gráfica que permita conocer la imagen que los individuos se crean de un conjunto de objetos por posicionamiento de cada uno en relación a los demás (mapa perceptual).

Modelos de escalamiento métrico

Los modelos de escalamiento parten de una función de representación de las proximidades estimadas δ_{ij} , $i, j = 1, \dots, n$ por los sujetos en forma de distancias entre n objetos: $d_{ij} \rightarrow f(\delta_{ij})$. En el caso del modelo métrico (también llamado *clásico*), la relación planteada generalmente entre proximidades y distancias es de tipo lineal: $d_{ij} = a + b \delta_{ij}$, aunque muchas variantes del modelo métrico admiten también transformaciones potenciales, logarítmicas o polinómicas de cualquier grado. El modelo de escalamiento métrico se aplica únicamente a datos medidos en escala de intervalo o razón. Toma como entrada, como hemos dicho, una matriz de proximidades entre n objetos y nos proporciona como salida las coordenadas de los n objetos en r dimensiones del espacio.

Modelos de escalamiento no métrico

Mientras que el modelo de MDS métrico plantea una relación lineal entre las proximidades de entrada y las distancias derivadas por el modelo, el modelo de escalamiento no-métrico plantea una relación de tipo monotónico, creciente entre ambas, es decir, una relación de tipo ordinal. En MDS no-métrico, por tanto, la relación entre proximidades y distancias es únicamente del tipo:

$$\text{si } \delta_{ij} > \delta_{kl}, \text{ entonces } d_{ij} \geq d_{kl}.$$

El procedimiento de MDS no métrico parte de una matriz de proximidades ordinal o de otro tipo, que es transformada en una matriz de proximidades en rangos, ordenados desde 1 hasta $(n^2 - n)/2$. Esta transformación se lleva a cabo simplemente asignando los rangos a las proximidades en función de su tamaño. A continuación, se calculan unos valores transformados, llamados *disparidades* (d_{ij}) que se ajustan monotónicamente a las proximidades. Generalmente se comienza con una configuración de distancias generada aleatoriamente o mediante algún otro método, y se va ajustando ésta hasta que los rangos de las disparidades coincidan en el sentido monotónico con los rangos de las proximidades.

Modelo de escalamiento de diferencias individuales

El modelo MDS de diferencias individuales, también conocido como *modelo ponderado*, es parte de una familia de procedimientos de análisis conocidos como *modelos euclídeos generalizados*, que tienen en común el hecho de que utilizan como entrada varias matrices de proximidad (una para cada fuente de datos) y que admiten ponderaciones diferentes de las dimensiones del espacio para cada fuente de datos.

Los distintos modelos difieren entre sí en el modo en que esta ponderación se lleva a cabo, en el uso de datos métricos o no-métricos, o en el permitir que las dimensiones sean, además de ponderadas, rotadas también de forma diferente para cada fuente de datos.

Lo interesante de estos modelos es que permiten tratar diferencias entre distintas fuentes de datos, como sujetos, grupos o momentos temporales. Los modelos métrico y no métrico también pueden utilizar como entrada varias matrices de proximidad, pero considerando a cada una de éstas como replicaciones de una misma fuente de datos, de tal modo que las diferencias existentes entre las distintas matrices se tratan como si fuesen errores. Sin embargo, es muy posible que estas diferencias no se deban a errores sino que, por el contrario, sean sistemáticas. Los modelos MDS de diferencias individuales permiten incorporar estas diferencias en la solución del análisis.

El más conocido y utilizado de estos modelos es el *modelo INDSCAL (Individual Differences SCALing)*.

Modelos de escalamiento para datos de preferencia

Normalmente, el MDS se aplica única y exclusivamente a datos de proximidad, bien sea obtenidos directamente, bien sea derivados a partir de datos multivariados. Sin embargo, existen modelos de MDS pensados para otro tipo de datos: los datos de dominancia. Para decirlo en pocas palabras, los datos de dominancia proporcionan información acerca del grado en que existen relaciones de precedencia o jerarquía entre éstos. Existen múltiples formas de recoger datos de dominancia, algunas de ellas increíblemente complejas y alambicadas, pero el ejemplo más habitual y sencillo de datos de dominancia son los datos de preferencia. Para obtener este tipo de datos a partir de una muestra de objetos es necesario únicamente solicitar a los sujetos que ordenen estos objetos en función de su preferencia. Esto nos proporcionará una matriz rectangular de preferencias P de dimensiones $n \times m$ (sujetos \times objetos), donde cada elemento p_{ij} de la matriz corresponderá a la preferencia del sujeto i por el objeto j .

Existen dos modelos de MDS muy utilizados con datos de preferencia: el *modelo desdoblado (unfolding)* y el *modelo vectorial*. En el modelo desdoblado tanto sujetos como objetos aparecen como puntos en un mismo espacio, y las preferencias de un sujeto deberían estar en correspondencia con la distancia a la que se hallen los objetos del punto que representa al sujeto, de tal modo que cuanto más preferido sea un objeto, más próximo debería encontrarse a ese punto. Interpretado así, el punto que representa al sujeto correspondería al objeto “ideal”, o de máxima preferencia. Por esta razón también se conoce al modelo desdoblado como modelo del “punto ideal”.

El modelo vectorial se diferencia del modelo desdoblado o del “punto ideal”, en que las filas de la matriz de preferencias (es decir, los sujetos) no se representan mediante puntos, sino mediante vectores de longitud unidad. Lo que pretende el modelo vectorial es encontrar una combinación lineal de los valores de coordenadas de los objetos, de modo que sus proyecciones sobre el vector que representa a un sujeto se correspondan lo más estrechamente posible con las preferencias manifestadas por ese sujeto.

8.5 Análisis conjunto

El análisis conjunto es una técnica estadística utilizada para analizar la relación lineal o no lineal entre una variable dependiente (o endógena) generalmente ordinal (aunque también puede ser métrica) y varias variables independientes (o exógenas) no métricas. La expresión funcional del análisis conjunto puede escribirse también como sigue:

$$y = F(x_1, x_2, \dots, x_n)$$

La variable dependiente recoge la preferencia (intención de compra, etc.) que el individuo exhibe hacia el producto (es decir, la utilidad global que el producto le aporta) y las variables dependientes son los atributos distintivos del producto. Es importante tener presente que sólo la variable dependiente recogerá información aportada por los individuos encuestados, ya que la información contenida en las variables independientes será especificada por el investigador en virtud de los productos que desee someter a evaluación por los encuestados.

El análisis conjunto permite generar un modelo individualizado por encuestado, de modo que el modelo general para toda la muestra resulte de la agregación de los modelos de todos los individuos que la componen. El análisis conjunto descompone las preferencias que el individuo manifiesta hacia el producto a fin de conocer qué valor le asigna a cada atributo (*técnica descomposicional*), mientras que en el análisis discriminante y en el análisis de la regresión las valoraciones de cada atributo que hace el sujeto se utilizan para componer su preferencia sobre el producto (*técnicas composicionales*).

La mayoría de software estadístico, y en concreto SPSS, utiliza la *aproximación de perfil completo (full profile)* para aplicar el análisis conjunto. En este caso, una vez planteado el problema, los sujetos que responden a la encuesta elaboran un rango de los perfiles o estímulos alternativos definidos por los niveles particulares de todos los atributos estudiados (o de un subconjunto definido por un diseño ortogonal como ORTHOPLAN) y recogen las preferencias en tarjetas de estímulo, elaboradas con procedimientos como PLANCARDS para ser luego analizadas y estimadas las utilidades por procedimientos como CONJOINT.

8.6 Análisis de correspondencias simple en SPSS

Mediante análisis de correspondencias simple analizaremos la relación entre las categorías de las variables origen y cilindros de los automóviles, características recogidas en el fichero COCHES.SAV.

Para realizar un análisis de correspondencias simple, elija en los menús *Analizar* → *Reducción de datos* → *Análisis de correspondencias* (Figura 8-1) y seleccione las variables y las especificaciones para el análisis (Figura 8-2). Previamente es necesario cargar en memoria el fichero de nombre COCHES mediante *Archivo* → *Abrir* → *Datos*. Este fichero contiene datos sobre automóviles y las variables a analizar son el origen de los coches (*origen*) y su cilindrada (*cilindrada*).

En cuanto a los 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.

En cuanto a los 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.

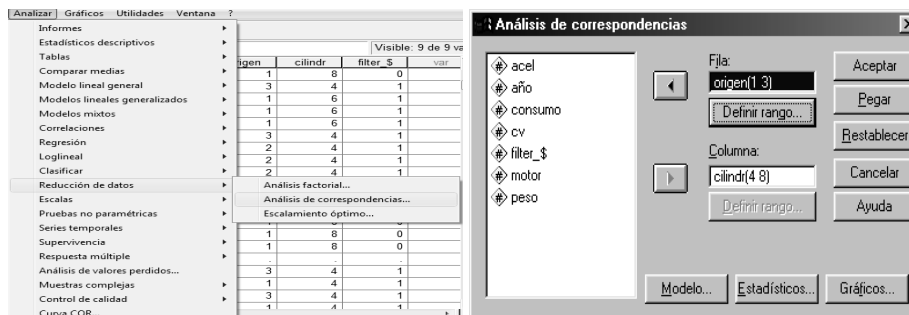


Figura 8-1

Figura 8-2

En los campos *Fila* y *Columna* de la Figura 8-2 se introducen las dos variables a cruzar en la tabla de contingencia. En los botones *Definir rango* debe definir un rango para las variables de filas (Figura 8-3) y columnas (Figura 8-4). 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 (campo *Restricciones para las categorías*) o puede definir cualquier categoría de fila como suplementaria. *Las categorías deben ser iguales* es una restricción que indica que 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 es una restricción que indica que 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.

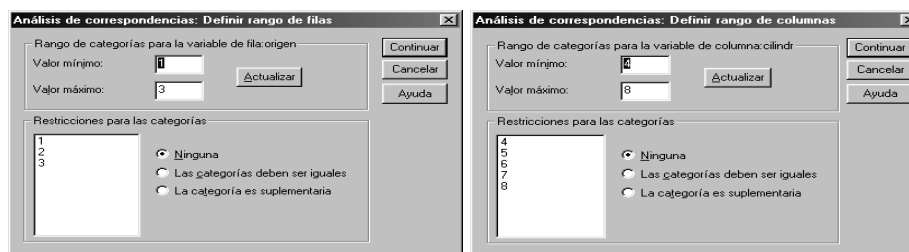


Figura 8-3

Figura 8-4

El cuadro de diálogo *Modelo* (Figura 8-5) permite especificar el número de dimensiones, la medida de distancia, el método de estandarización y el método de normalización. En la opción *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 y 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.

En el cuadro *Medida de distancia* puede seleccionar la medida de distancia entre las filas y columnas de la tabla de correspondencias. Seleccione *Chi-cuadrado* (utiliza una distancia ponderada entre los perfiles, donde la ponderación es la masa de las filas o de las columnas siendo una distancia necesaria para el análisis de correspondencias típico) o *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).

En el cuadro *Método de estandarización* seleccione la opción *Se eliminan las medias de filas y columnas* para centrar las filas y las columnas (este método es necesario para el análisis de correspondencias típico), seleccione *Se eliminan las medias de filas* sólo para centrar las filas, seleccione *Se eliminan las medias de columnas* sólo para centrar las columnas, seleccione *Se igualan los totales de fila y se eliminan las medias* para igualar los márgenes de fila antes de centrar las filas. Seleccione *Se igualan los totales de columna y se eliminan las medias* para igualar los márgenes de columna antes de centrar las columnas.

En el cuadro *Método de normalización* seleccione una de las siguientes opciones:

Simétrico: para cada dimensión, las puntuaciones de fila son la media ponderada de las puntuaciones de columna divididas por el valor propio coincidente y las puntuaciones de columna son la media ponderada de las puntuaciones de fila divididas por el valor propio coincidente. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de las dos variables.

Principal: las distancias entre los puntos de fila y los puntos de columna son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Utilice este método si desea examinar las diferencias entre las categorías de una o de ambas variables en lugar de las diferencias entre las dos variables.

Principal por fila: las distancias entre los puntos de fila son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Las puntuaciones de fila son la media ponderada de las puntuaciones de columna. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de la variable de filas.

Principal por columna: las distancias entre los puntos de columna son aproximaciones de las distancias en la tabla de correspondencias de acuerdo con la medida de distancia seleccionada. Las puntuaciones de columna son la media ponderada de las puntuaciones de fila. Utilice este método si desea examinar las diferencias o similitudes entre las categorías de la variable de columnas.

Personalizado: debe especificar un valor entre -1 y 1 . El valor -1 corresponde a *Principal por columna*. El valor 1 corresponde a *Principal por fila*. El valor 0 corresponde a *Simétrico*. Todos los demás valores dispersan la inercia entre las puntuaciones de columna y de fila en diferentes grados. Este método es útil para generar diagramas de dispersión biespaciales a medida.

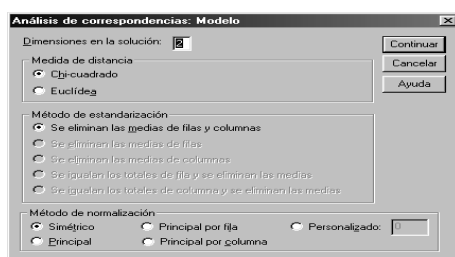


Figura 8-5

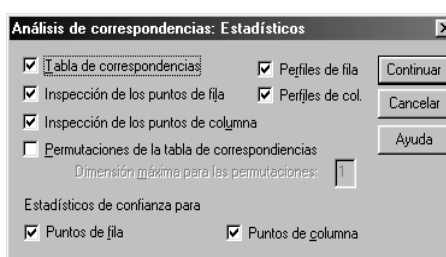


Figura 8-6

El botón *Estadísticos* de la Figura 8-2 nos lleva al cuadro de diálogo *Estadísticos* (Figura 8-6), que permite especificar los resultados numéricos producidos. Las opciones posibles son: *Tabla de correspondencias*, que ofrece la tabla de contingencia de las variables de entrada con los totales marginales de fila y columna; *Inspección de los puntos de fila*, que ofrece 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*, que ofrece 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*, que ofrece para cada categoría de fila la distribución a través de las categorías de la variable de columna; *Perfiles de col.*, que ofrece para cada categoría de columna la distribución a través de las categorías de la variable de fila y *Permutaciones de la tabla de correspondencias*, que ofrece 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. La opción *Estadísticos de confianza para puntos de fila* incluye la desviación típica y las correlaciones para todos los puntos de fila no suplementarios y la opción *Estadísticos de confianza para puntos de columna* incluye la desviación típica y las correlaciones para todos los puntos de columna no suplementarios.

El botón *Gráficos* de la Figura 8-2 nos lleva al cuadro de diálogo *Gráficos* de la Figura 8-7 que permite especificar qué gráficos se van a generar. La opción *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 y 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.

La opción *Gráfico de líneas* 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) y *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.

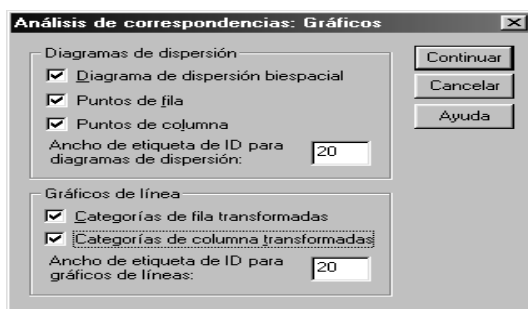


Figura 8-7

En todas las figuras, el botón *Restablecer* permite restablecer todas las opciones por defecto del sistema y elimina del cuadro de diálogo todas las asignaciones hechas con las variables.

Una vez elegidas las especificaciones, se pulsa el botón *Aceptar* en la Figura 8-2 para obtener los resultados del análisis de correspondencias según se muestra en la Figura 8-8. En la parte izquierda de la figura podemos ir seleccionando los distintos tipos de resultados haciendo clic sobre ellos. También se ven los resultados desplazándose a lo largo de la pantalla.

En la Figura 8-8 se presentan varias salidas tabulares de entre las múltiples que ofrece el procedimiento. Las Figuras 8-9 a 8-11 presentan salidas gráficas del procedimiento.

La Figura 8-8 muestra la sintaxis del procedimiento, la tabla de contingencia para las dos variables con sus marginales, los perfiles de fila y columna, que son las proporciones en cada fila y columna de cada celda basadas en los totales marginales. Los gráficos de puntos fila y columna de las Figuras 8-9 a 8-11 representan estas proporciones para la localización geométrica de los puntos. La Figura 8-8 también muestra un cuadro resumen con la solución que representa la relación entre las variables fila y columna en tan pocas dimensiones como es posible.

En nuestro caso tenemos dos dimensiones, mostrando la primera una cantidad mayor de inercia (el 95% de la inercia total). Los valores propios pueden interpretarse como la correlación entre las puntuaciones de filas y columnas. Para cada dimensión el cuadrado del valor propio es igual a la inercia y por tanto es otra medida de la importancia de esa dimensión.

En el examen de los puntos fila y columna (Figuras 8-8) se ofrecen las contribuciones a la inercia total de cada punto fila y columna. Los puntos fila y columna que contribuyen sustancialmente a la inercia de una dimensión son importantes para esa dimensión. Los puntos dominantes de la solución pueden detectarse fácilmente. Por ejemplo, Japón es un punto dominante de la segunda dimensión ya que su contribución a la inercia de esa dimensión es 0,635 y Estados Unidos en la primera dimensión pues su contribución es 0,609. Por otra parte, los coches de 8 cilindros (0,959) y 6 cilindros (0,651) contribuyen más que otros a la primera dimensión. A la segunda dimensión los que más contribuyen negativamente son los de 4 cilindros (0,096).

XXX Contenido

CORRESPONDENCE TABLE = origen(1 3) BY cilindr(4 8)
 /DIMENSIONS = 2 /MEASURE = CHISQ /STANDARDIZE = RCMEAN /NORMALIZATION =
 SYMMETRICAL /PRINT = TABLE RPOINTS CPOINTS RPROFILES CPROFILES RCONF CCONF
 /PLOT = NDI(1,MAX) BI(20) RPOINTS(20) CPOINTS(20) TRROWS(20) TRCOLUMNS(20)
) .

Tabla de correspondencias

País de origen	Número de cilindros					Margen activo
	4 cilindros	5 cilindros	6 cilindros	7	8 cilindros	
EE.UU.	72	0	74	0	107	253
Europa	66	3	4	0	0	73
Japón	69	0	6	0	0	75
Margen activo	207	3	84	0	107	401

Perfiles de fila

País de origen	Número de cilindros					Margen activo
	4 cilindros	5 cilindros	6 cilindros	7	8 cilindros	
EE.UU.	,285	,000	,292	,000	,423	1,000
Europa	,904	,041	,055	,000	,000	1,000
Japón	,920	,000	,080	,000	,000	1,000
Masa	,516	,007	,209	,000	,267	

Perfiles de columna

País de origen	Número de cilindros					Masa
	4 cilindros	5 cilindros	6 cilindros	7	8 cilindros	
EE.UU.	,348	,000	,881	,000	1,000	,631
Europa	,319	1,000	,048	,000	,000	,182
Japón	,333	,000	,071	,000	,000	,187
Margen activo	1,000	1,000	1,000	,000	1,000	

Resumen

Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de inercia		Confianza para el Valor propio	
					Explicada	Acumulada	Desviación típica	Correlación
1	,635	,404			,952	,952	,032	,088
2	,143	,020			,048	1,000	,040	
Total		,424	169,997	,000 ^a	1,000	1,000		

a. 8 grados de libertad

Examen de los puntos de fila^a

País de origen	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	Total
EE.UU.	,631	,609	-,013	,149	,368	,001	1,000	,000	1,000
Europa	,182	-1,102	-,607	,150	,348	,470	,936	,064	1,000
Japón	,187	-,982	,835	,125	,284	,529	,914	,086	1,000
Total activo	1,000			,424	1,000	1,000			

a. Normalización Simétrica

Examen de los puntos columna^a

Número de cilindros	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		
					1	2	1	2	Total
4 cilindros	,516	-,735	,096	,178	,439	,033	,996	,004	1,000
5 cilindros	,007	-1,734	-4,255	,034	,035	,949	,425	,575	1,000
6 cilindros	,209	,651	,034	,057	,140	,002	,999	,001	1,000
7	,000								
8 cilindros	,267	,959	-,092	,156	,386	,016	,998	,002	1,000
Total activo	1,000			,424	1,000	1,000			

a. Normalización Simétrica

Figura 8-8

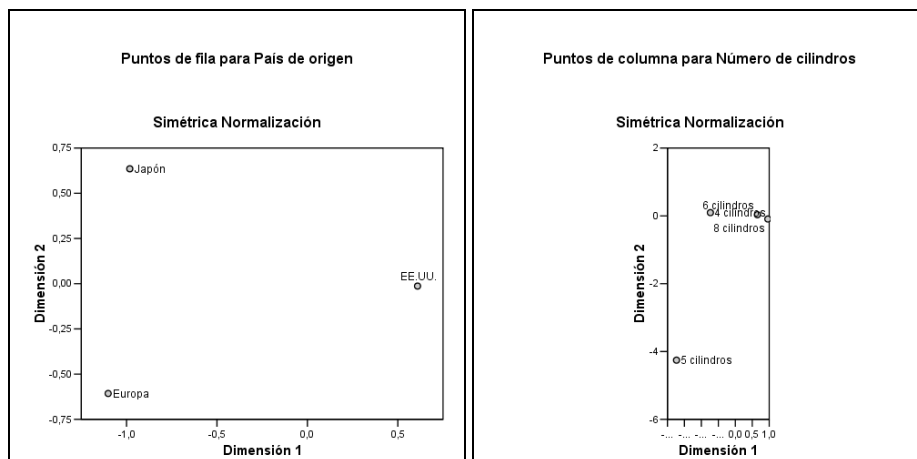


Figura 8-9

Figura 8-10

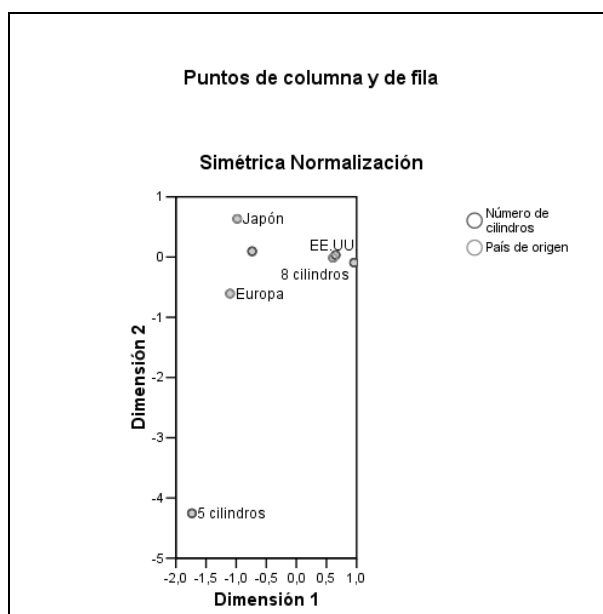


Figura 8-11

8.7 Análisis de correspondencias múltiple en SPSS

Utilizaremos análisis de correspondencias múltiple para observar la relación entre la categoría laboral (*catlab*), la clasificación étnica (*minoría*) y el género (*sexo*) de los empleados de una empresa. Los datos se recogen en el fichero *empleados.sav*

Para realizar un análisis de correspondencias múltiple, elija en los menús *Analizar* → *Reducción de datos* → *Escalamiento óptimo* (Figura 8-12). Previamente es necesario cargar en memoria el fichero de nombre EMPLEADOS mediante *Archivo* → *Abrir* → *Datos*. Este fichero contiene datos sobre los trabajadores de una empresa con las variables *catlab*, *minoría* y *sexo* antes descritas.

En el cuadro de diálogo *Escalamiento óptimo* de la Figura 8-13, seleccione *Todas las variables son nominales múltiples*. A continuación seleccione *Un conjunto*, pulse en *Definir*, y en la Figura 8-14 seleccione dos o más variables para el análisis. Defina el número de categorías para las variables y su agrupación con el botón *Discretizar* (figura 8-15). Si lo desea, tiene la posibilidad de seleccionar una o más variables para proporcionar etiquetas de punto en los gráficos de las puntuaciones de objeto (campo *Variables de etiquetado*). Cada variable genera un gráfico diferente, con los puntos etiquetados mediante los valores de dicha variable. Mediante el cuadro de diálogo, no se puede utilizar una misma variable en el análisis y como variable de etiquetado. Si se desea etiquetar el gráfico de las puntuaciones de objeto con una variable utilizada ya en el análisis, utilice la función *Calcular* en el menú *Transformar* para crear una copia de dicha variable. Utilice la nueva variable para etiquetar el gráfico. También se puede utilizar la sintaxis de comandos. En el botón *Dimensiones* en la solución 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, SPSS genera gráficos tridimensionales de las tres primeras dimensiones. Si se edita el gráfico, se pueden representar otras dimensiones.

El botón *Opciones* permite seleccionar métodos de normalización, criterios de iteración y de convergencia y otras características del algoritmo. El botón *Gráficos* → *Opciones* (Figura 8-16) permite seleccionar opciones gráficas muy importantes para la interpretación de los resultados como el diagrama de representación biespacial con sus variables y etiquetas. El botón *Gráficos* → *Variables* (Figura 8-17) permite seleccionar categorías a graficar y medidas discriminantes a obtener.

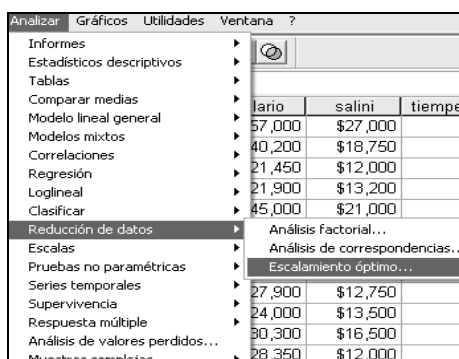


Figura 8-12

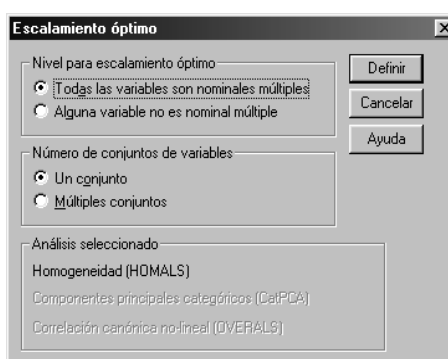


Figura 8-13

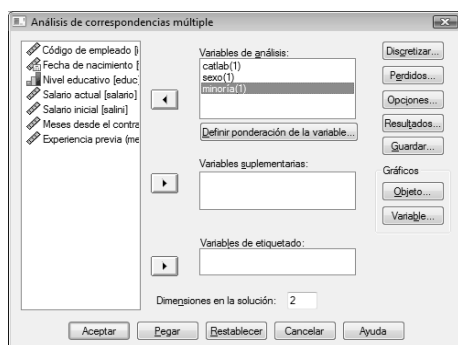


Figura 8-14

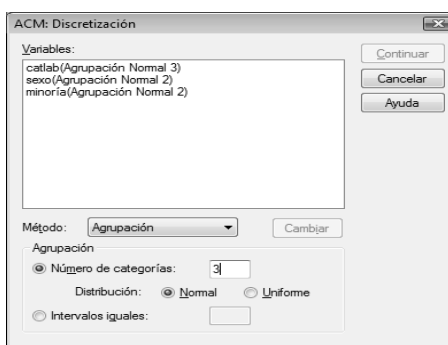


Figura 8-15

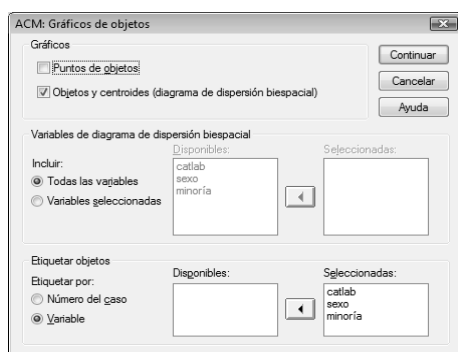


Figura 8-16

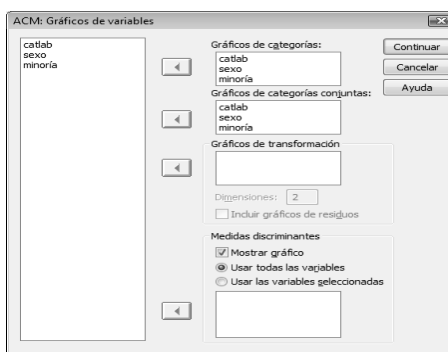


Figura 8-17

En cuanto a los datos, todas las variables son nominales múltiples y tienen cuantificaciones de categorías que pueden diferir para cada dimensión. Una vez elegidas las especificaciones (que se aceptan con el botón *Continuar*), se pulsa el botón *Aceptar* en la Figura 8-14 para obtener los resultados del análisis de correspondencias múltiples. En las Figuras 17-Figuras 8-18 a 8-21 se presentan varias salidas tabulares y gráficas de entre las múltiples que ofrece el procedimiento.

En la Figura 8-18 aparece la historia del proceso de homogeneización a través de las distintas iteraciones que el procedimiento considera necesarias para llegar a una solución de convergencia que refleje el ajuste total, así como la tabla de autovalores para cada dimensión del análisis. Como el análisis se realiza sobre los dos primeros ejes o dimensiones, se muestra en cada una de ellas la medida de la varianza explicada por cada dimensión. La magnitud de esta varianza es una muestra del grado de importancia de dicha dimensión en la solución global. Se observa que las dos dimensiones son casi igual de importantes ya que los dos valores propios están muy próximos.

En la figura 8-19 se muestra el diagrama conjunto de puntos categorías que muestra que ser administrativo y mujer está muy asociado. También se observa buena asociación entre ser hombre y directivo.

El pertenecer a una minoría étnica (*Si*) también está bien asociado con ser administrativo, pero mal asociado con ser directivo. Pertenecer o no a una minoría étnica está igualmente asociado con ser hombre que con ser mujer. De esta forma se miden las relaciones entre todas las categorías de las tres variables cualitativas. Las figuras 8-20 y 8-21 muestran los diagramas de dispersión biespacial por pares de variables.

Historial de iteraciones			
Número de iteraciones	Varianza explicada		
	Total	Incremento	Pérdida
12 ^a	1,321629	,000007	1,678371

a. Se ha detenido el proceso de iteración debido a que se ha alcanzado el valor de la prueba para la convergencia.

Resumen del modelo				
Dimensión	Alfa de Cronbach	Varianza explicada		
		Total (Autovalores)	Inercia	% de la varianza
1	,436	1,409	,470	46,978
2	,284	1,234	,411	41,131
Total		2,643	,881	
Media	,365 ^a	1,322	,441	44,054

a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

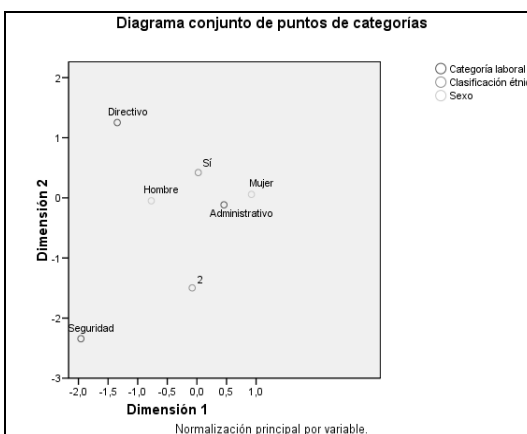


Figura 8-18

Figura 8-19

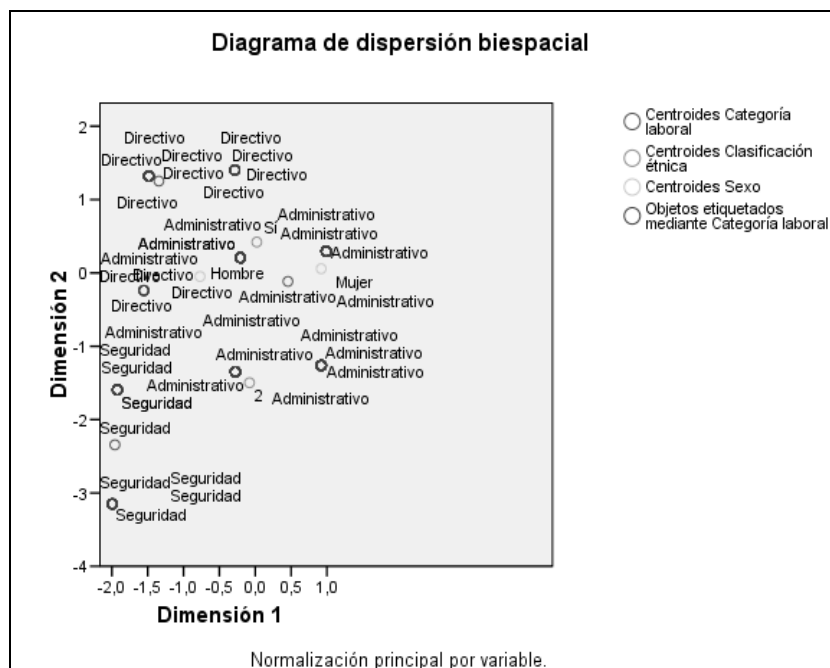


Figura 8-20

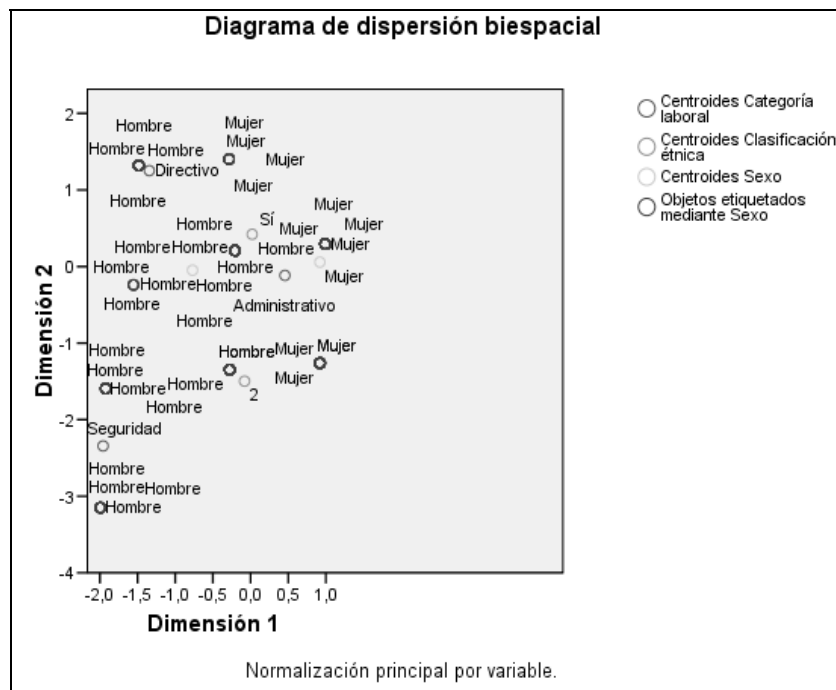


Figura 8-21

8.8 Componentes principales no lineales en SPSS

Partimos del fichero *8-3.sav* que contiene los resultados de una encuesta en la que a los individuos encuestados se les pedía manifestar el grado de acuerdo con nueve afirmaciones. Las respuestas se codifican en las nueve variables *ítem1* a *ítem9* y adicionalmente se clasifican según la variable *sexo*. Realizar un análisis de no lineal de componentes principales que permita reducir la dimensión de la información original de forma coherente.

Comenzamos cargando en el editor de SPSS los datos del fichero *8-3.sav* mediante *Abrir* → *Datos* y a continuación se selecciona *Analizar* → *Reducción de datos* → *Escalamiento óptimo* (Figura 8-22). Se obtiene la pantalla de selección del tipo de escalamiento óptimo que se rellena como se indica en la Figura 8-23 seleccionando *CatPCA*. Al pulsar en *Definir* se obtiene la pantalla de *Componentes principales categóricas* (Figura 8-24). Con el botón *Resultados* se elige la salida que se desea (Figura 8-25) y con los botones del campo *Gráficos* se elige la salida gráfica (Figura 8-26).

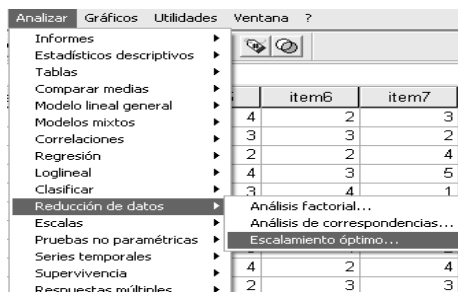


Figura 8-22

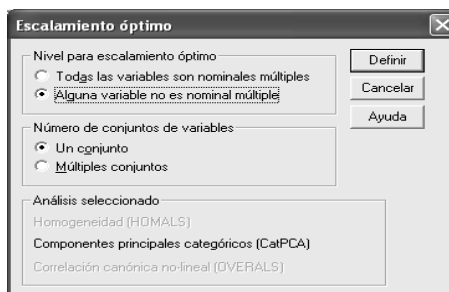


Figura 8-23

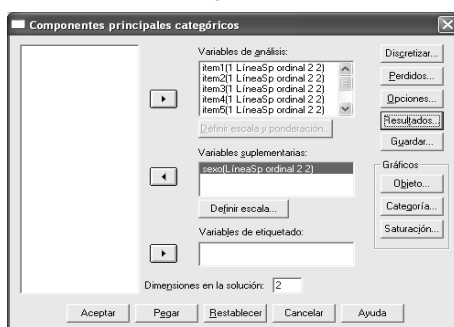


Figura 8-24

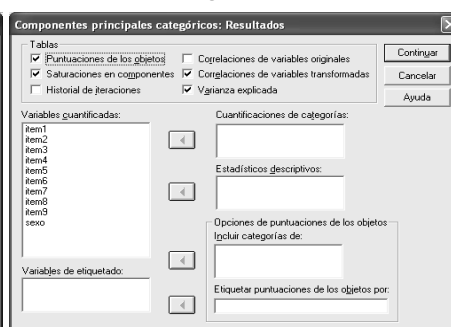


Figura 8-25

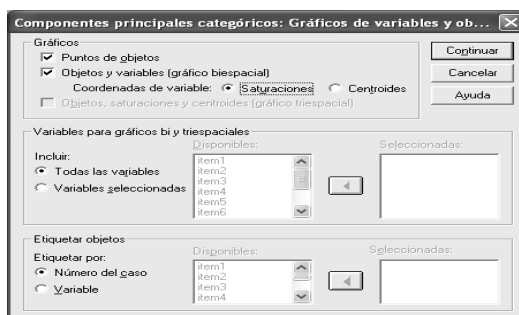


Figura 8-26

Al hacer clic en *Continuar* y *Aceptar* se obtiene la salida del procedimiento de componentes principales categóricas CATPCA (Figura 8-27). En la Figura 8-28 se obtiene la salida resumen del modelo que selecciona dos componentes principales que recogen el 40,538% de la varianza total del modelo. En la Figura 8-29 se ofrece el historial de iteraciones hasta llegar a la solución. En la Figura 8-30 se ve el tanto por ciento de la varianza asociada a cada variable en cada dimensión. En la Figura 8-31 se recogen las cargas o saturaciones de cada una de las variables sobre cada una de las dimensiones del modelo factorial, que representan las proyecciones de cada variable cuantificada en el espacio de los objetos. Se trata del coeficiente de correlación entre cada una de las variables intervinientes en el modelo con cada una de las dos dimensiones.

Créditos	
CATPCA	
Version 1.1	
by	
Data Theory Scaling System Group (DTSS)	
Faculty of Social and Behavioral Sciences	
Leiden University, The Netherlands	

Figura 8-27

Resumen del modelo			
Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	% de la varianza
1	,557	1,981	22,007
2	,450	1,668	18,531
Total	,817 ^a	3,648	40,538

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

Figura 8-28

Historial de iteraciones				
Número de iteración	Varianza explicada		Pérdida	
	Total	Incremento	Total	Restricción del centroide a las coordenadas del vector
24 ^a	3,648388	,000008	14,351612	13,736081 ,615531

a. Se ha detenido el proceso de iteración debido a que se ha alcanzado el valor de la prueba para la convergencia.

Figura 8-29

	Varianza explicada					
	Coordenadas de centroide			Total (coordenadas del vector)		
	Dimensión		Media	Dimensión		Total
1	2	1		2		
item1	,065	,475	,270	,014	,456	,469
item2	,286	,293	,289	,266	,235	,501
item3	,056	,588	,322	,003	,573	,575
item4	,272	,075	,173	,254	,022	,276
item5	,381	,076	,228	,357	,007	,364
item6	,400	,124	,262	,346	,112	,458
item7	,398	,092	,245	,375	,067	,441
item8	,270	,037	,153	,264	,015	,279
item9	,171	,205	,188	,103	,181	,284
sexo ^a	,010	,001	,005	,010	,001	,011
Total activo	2,299	1,965	2,132	1,981	1,668	3,648
% de la varianza	25,542	21,834	23,688	22,007	18,531	40,538

a. Variable suplementaria.

Figura 8-30

	Saturaciones en componentes	
	Dimensión	
	1	2
item1	-,117	,675
item2	,515	-,485
item3	,055	,757
item4	-,504	-,149
item5	-,597	-,086
item6	,588	-,335
item7	,612	,258
item8	,514	,124
item9	,321	,425
sexo ^a	,100	,025

Normalización principal por variable.
a. Variable suplementaria.

Figura 8-31

En cuanto a las salidas gráficas del procedimiento, en la Figura 8-32 se presenta el gráfico de saturaciones en las componentes que se utiliza para agrupar nuestras variables en las dos componentes. Está claro que *ítem2* e *ítem6* se asocian con una primera componente e *ítem4* e *ítem5* con la segunda componente. Pero ya no está tan claro con qué componente principal asociar el resto de las variables. Según la figura podría ser lógico asociarlas todas con la primera componente. También podrían asociarse *ítem7* e *ítem8* con la primera componente e *ítem1*, *ítem3* e *ítem9* con la segunda.

No obstante, esta clasificación de las variables en componentes también puede realizarse observando la tabla de saturaciones en las componentes de la Figura 8-31. Se observa en esta tabla que para la componente 2, las saturaciones más altas las presentan las variables *ítem1*, *ítem3* e *ítem9*. Para la componente 1 las saturaciones más altas las presentan *ítem2*, *ítem6*, *ítem4*, *ítem5* e *ítem8* (*ítem4*

e *ítem5* con valor negativo, por eso aparecen a la izquierda del gráfico). Luego la forma definitiva de agrupar las variables en componentes sería asociar las variables *ítem4*, *ítem5*, *ítem2*, *ítem6*, *ítem7* e *ítem 8* en una componente y las variables *ítem1*, *ítem3* e *ítem9* en la otra componente, siendo las asociaciones más indefinidas las de las variables *ítem7* e *ítem 8*. Se observa que la mejor forma de asociar las variables a las componentes principales es analizar simultáneamente la tabla de las saturaciones en las componentes de la Figura 8-31 y el gráfico de las saturaciones en las componentes de la Figura 8-32. La Figura 8-33 presenta la gráfica de puntuaciones de los objetos etiquetadas por el número de caso y en la Figura 8-34 se observa el gráfico de dispersión biespacial, que muestra sobre el mismo gráfico las puntuaciones de los objetos etiquetadas por el número de caso y las saturaciones en las componentes.

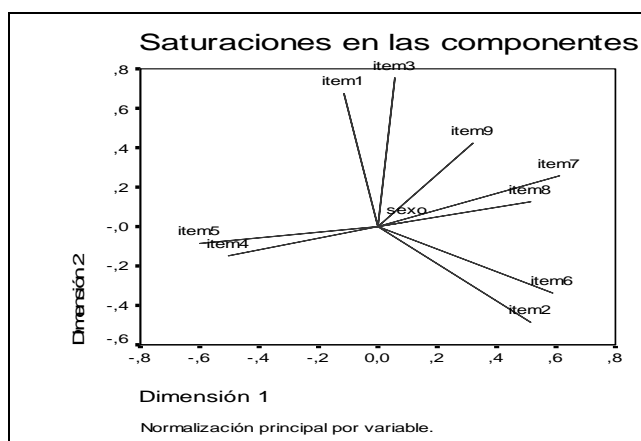


Figura 8-32

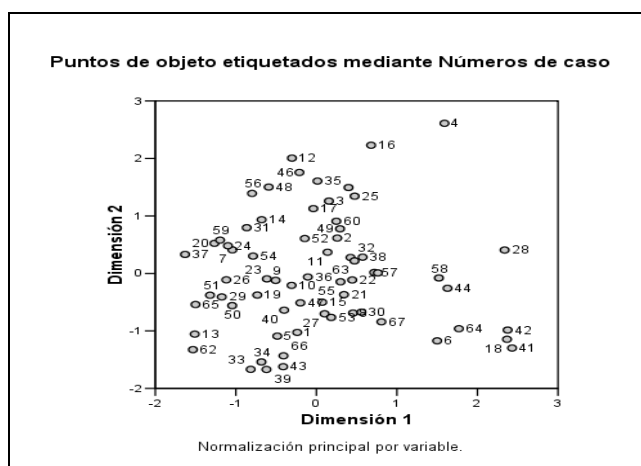


Figura 8-33

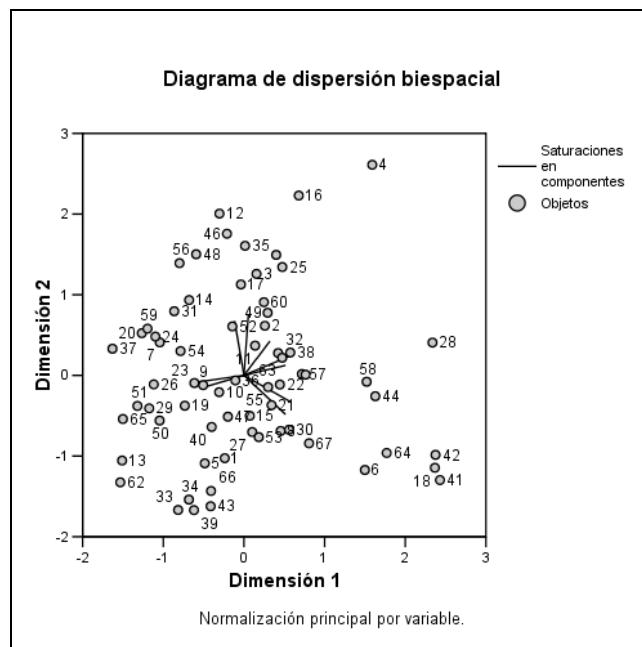


Figura 8-34

8.9 Correlación canónica no lineal en SPSS

Partimos del Fichero *8-4.sav* que contiene los resultados de una encuesta en la que a los individuos encuestados se les pedía manifestar el grado de acuerdo con nueve afirmaciones. Las respuestas se codifican en las nueve variables *ítem1* a *ítem9* y adicionalmente se clasifican según la variable *sexo*. Realizar un análisis no lineal de correlación canónica tomando como primer conjunto de variables *ítem1*, *ítem4* e *ítem6*, y como segundo conjunto de variables *ítem2*, *ítem3* e *ítem5*.

Comenzamos cargando en el editor de SPSS los datos del fichero *8-4.sav* mediante *Abrir* → *Datos* y a continuación se selecciona *Analizar* → *Reducción de datos* → *Escalamiento óptimo*. Se obtiene la pantalla de selección del tipo de escalamiento óptimo que se rellena como se indica en la Figura 8-35 seleccionando *OVERALS (Múltiples conjuntos)*. Al pulsar en *Definir* se obtiene la pantalla de *Análisis de correlación canónica no lineal* (Figura 8-36) en cuyo campo *Variables* se introducen el primer conjunto de variables para el análisis. Con el botón *Definir rango y escala* se declara el máximo y el mínimo de la escala de medida (Figura 8-37). Se pulsa en *Continuar* y ya se tiene definido el primer conjunto de variables (Figura 8-38). Se pulsa en *Siguiente* y se introduce el segundo conjunto de variables

definiendo también su rango y escala (Figura 8-39). Con el botón *Opciones* se elige la salida que se desea para el análisis, tanto tabular como gráfica (Figura 8-40). Se pulsa en *Continuar* y en *Aceptar*, con lo que ya tenemos la salida del procedimiento OVERALS.

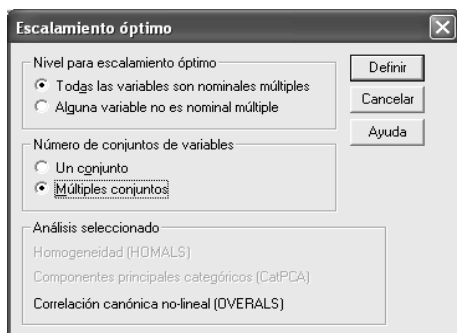


Figura 8-35

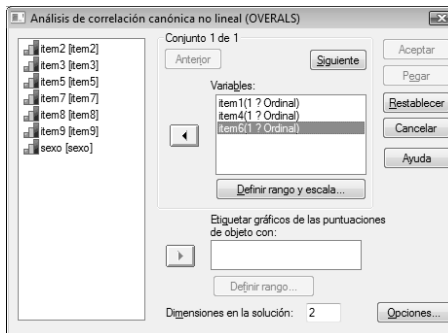


Figura 8-36



Figura 8-37

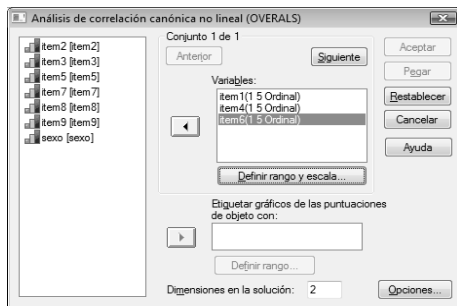


Figura 8-38

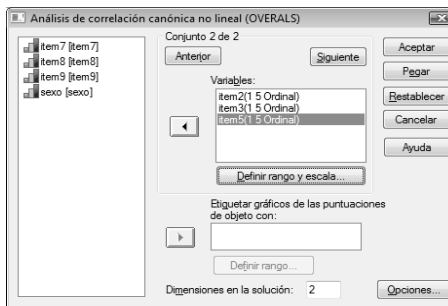


Figura 8-39

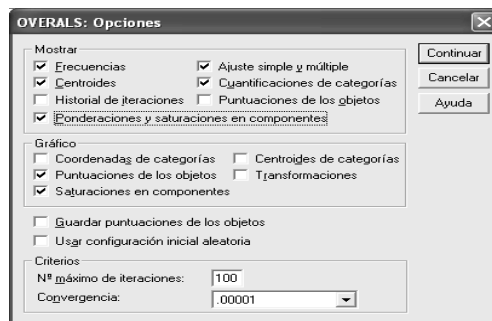


Figura 8-40

La salida tabular comienza ofreciendo listado de las variables con los dos grupos de variables que intervienen en el análisis junto a su número de categorías (Figura 8-41), el historial de iteraciones y el resumen del análisis (Figura 8-42). El historial de iteraciones presenta un informe sobre las iteraciones sucesivas que se llevan a cabo para establecer una relación canónica entre los dos conjuntos. Este proceso de búsqueda de una solución que satisfaga el valor de convergencia (llamado valor del test de la convergencia) desemboca en el cálculo de un valor de pérdida y otro de ajuste para la iteración 0 y la iteración en la que se produce la convergencia (la 77 en nuestro caso). También se presenta la diferencia entre las dos últimas iteraciones (0,000007). En el resumen del análisis, OVERALS muestra la pérdida por cada conjunto en cada dimensión. La suma de las pérdidas del conjunto 1 y del conjunto 2 deben coincidir. La pérdida media por dimensiones indica una pérdida moderada (0,580). El ajuste de la prueba representa un valor alto (1,420) y los autovalores (0,751 y 0,670) muestran una distribución de cargas de explicación de la varianza del modelo algo superior en la dimensión 1 que en la 2. La Figura 8-43 presenta la tabla de ponderaciones y la de saturaciones en las componentes.

Créditos		
OVERALS Version 1.0 by Data Theory Scaling System Group (DTSS) Faculty of Social and Behavioral Sciences Leiden University, The Netherlands		
Resumen del procesamiento de los casos		
Casos usados en el análisis	67	
Lista de variables		
Conjunto	Número de categorías	Nivel de escalamiento óptimo
1	item1	5 Ordinal
	item4	5 Ordinal
	item6	5 Ordinal
2	item2	5 Ordinal
	item3	5 Ordinal
	item5	5 Ordinal

Figura 8-41

Historial de iteraciones			
	Pérdida	Ajuste	Diferencia desde la iteración anterior
0 ^a	,798666	1,201334	
77 ^b	,579106	1,420894	,000007

a. La pérdida de la iteración 0 es la pérdida de la solución con todas las variables simples tratadas como numéricas (con una diferencia de pérdida de 0,0001 y un número máximo de 50 iteraciones).

b. Se ha detenido el proceso de iteración debido a que se ha alcanzado el valor de la prueba para la convergencia.

Resumen del análisis				
		Dimensión		Suma
		1	2	
Pérdida	Conjunto 1	,250	,331	,581
	Conjunto 2	,249	,329	,579
	Media	,249	,330	,580
Autovalores		,751	,670	
Ajuste				1,420

Figura 8-42

Ponderaciones			
		Dimensión	
		1	2
1	item1	,514	,671
	item4	-,370	-,380
	item6	,584	-,351
2	item2	,182	-,838
	item3	-,206	-,384
	item5	-,777	-,359

Saturaciones en componentes			
		Dimensión	
		1	2
1	item1 a.b	,530	,638
	item4 a.b	-,436	-,273
	item6 a.b	,542	-,389
2	item2 a.b	,436	-,665
	item3 a.b	-,190	-,201
	item5 a.b	-,814	-,102

a. Nivel de escalamiento óptimo: Ordinal
b. Proyecciones de las variables cuantificadas simples en el espacio de los objetos

Figura 8-43

XLII Contenido

La tabla de ponderaciones muestra los pesos por cada dimensión desglosados por un grupo de ítems del primer análisis y por sus respectivos elementos. Se puede observar la elevada fuerza explicativa del *ítem5* dentro de la dimensión 1 y de la carga del *ítem2* en la dimensión 2. Estas ponderaciones o pesos representan los coeficientes de correlación de cada dimensión para todas las variables cuantificadas de un conjunto, donde las puntuaciones de los objetos efectúan un análisis de la regresión sobre las variables cuantificadas.

La tabla de saturaciones en las componentes contempla las cargas de las componentes por variables simples, es decir las proyecciones de las variables cuantificadas en el espacio de los objetos. Estas cargas son una indicación de la contribución de cada variable a la dimensión dentro de cada conjunto. Se aprecia la elevada fuerza explicativa del *ítem3*, así como las de los *ítems1* e *ítem2*. El gráfico de saturaciones en componentes (Figura 8-45) representa en el plano de las dos dimensiones las cargas de las componentes para variables simples. Como ya hemos dicho, se observa la elevada fuerza explicativa del *ítem3* en la dimensión 1, así como la de los *ítems1* e *ítem2* en la dimensión 2. La tabla de ajuste de la Figura 8-44 resume datos de ajuste múltiple, simple y pérdida simple por dimensiones para cada variable de cada uno de los conjuntos del análisis.

Conjunto		Ajuste								
		Ajuste múltiple			Ajuste simple			Pérdida simple		
		Dimensión		Suma	Dimensión		Suma	Dimensión		Suma
1	2	1	2		1	2				
1	ítem1 ^a	,283	,456	,739	,265	,451	,715	,018	,005	,023
	ítem4 ^a	,163	,147	,310	,137	,144	,281	,026	,003	,029
	ítem6 ^a	,355	,140	,495	,341	,123	,464	,014	,017	,031
2	ítem2 ^a	,064	,705	,770	,033	,703	,736	,031	,002	,034
	ítem3 ^a	,053	,160	,213	,043	,147	,190	,011	,012	,023
	ítem5 ^a	,605	,137	,742	,604	,129	,733	,001	,008	,009

a. Nivel de escalamiento óptimo: Ordinal

Figura 8-44

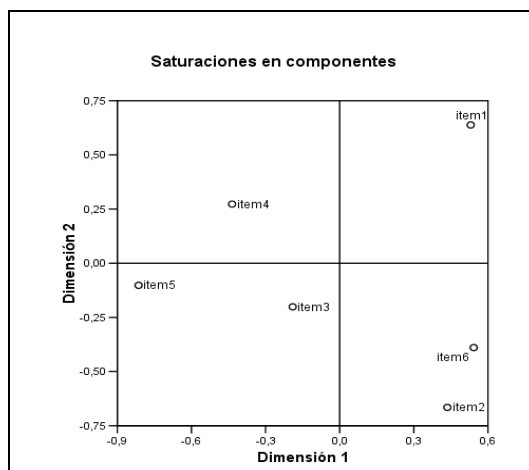


Figura 8-45

8.10 Escalamiento multidimensional métrico en SPSS

Consideramos la matriz de distancias entre 10 ciudades europeas siguiente:

Ciudad	Atenas	Berlín	Estocolmo	Londres	Madrid	Moscú	París	Roma	Varsovia	Viena
Atenas	0
Berlín	1 774	0
Estoco	2 371	806	0
Londre	2 355	9 19	1 387	0
Madrid	2 387	1 855	2 548	1 258	0
Moscú	2 177	1 565	1 210	2 419	3 371	0
París	2 065	871	1 516	339	1 048	2 419	0	.	.	.
Roma	1 048	1 177	1 952	1 419	1 371	2 323	1 097	0	.	.
Varsov	1 581	484	790	1 403	2 258	1 129	1 323	1 290	0	.
Viena	1 274	516	1 226	1 210	1 806	1 613	1 016	758	548	0

A partir de estas distancias, realizar un escalamiento métrico que sitúe estas ciudades sobre un mapa perceptual que emule el continente europeo.

Comenzamos introduciendo los datos de las distancias entre capitales europeas en el editor de SPSS (archivo *8-5.sav*) y a continuación se selecciona *Analizar* → *Escalas* → *Escalamiento multidimensional* (Figura 8-46). Se obtiene la pantalla de entrada del procedimiento de la Figura 8-47. Con los botones *Opciones* y *Modelo* se obtienen pantallas que se rellenan como se indica en la Figuras 19-48 y 19-49 (se observa *Razón* en *Nivel de medida*).

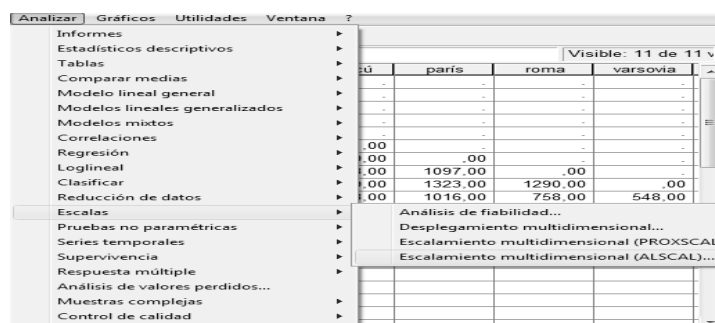


Figura 8-46

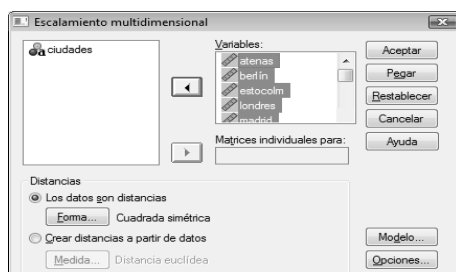


Figura 8-47

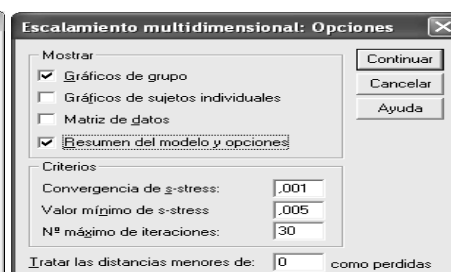


Figura 8-48

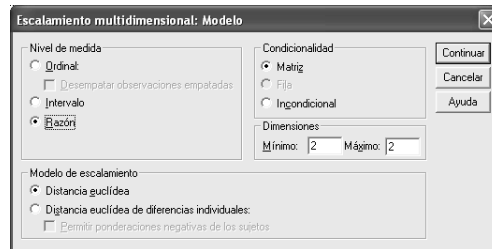


Figura 8-49

Al pulsar en *Continuar* y *Aceptar*, se obtiene la sintaxis del procedimiento y la salida textual del procedimiento ALSCAL que expresa las opciones de datos, de modelo, de salida y de algoritmo, así como el historial de iteraciones y la matriz de coordenadas normalizadas o coordenadas estímulos. El mapa perceptual de la Figura 8-50 muestra la situación de las capitales europeas (sería necesario un giro adecuado para obtener la situación real).

ALSCAL

```
VARIABLES= atenas berlin estocolm londres madrid moscú parís
roma varsovia vie
na
/SHAPE=SYMMETRIC
/LEVEL=RATIO
/CONDITION=MATRIX
/MODEL=EUCLID
/CRITERIA=CONVERGE(.001) STRESSMIN(.005) ITER(30) CUTOFF(0)
DIMENS(2,2)
/PLOT=DEFAULT
/PRINT=HEADER .
```

Alscal Procedure Options

Data Options-

```
Number of Rows (Observations/Matrix). 10
Number of Columns (Variables) . . . 10
Number of Matrices . . . . . 1
Measurement Level . . . . . Ratio
Data Matrix Shape . . . . . Symmetric
Type . . . . . Dissimilarity
Approach to Ties . . . . . Leave Tied
Conditionality . . . . . Matrix
Data Cutoff at . . . . . ,000000
```

Model Options-

```
Model . . . . . Euclid
Maximum Dimensionality . . . . . 2
Minimum Dimensionality . . . . . 2
Negative Weights . . . . . Not Permitted
```

Output Options-

Job Option Header Printed
 Data Matrices Not Printed
 ConFigurations and Transformations Plotted
 Output Dataset Not Created
 Initial Stimulus Coordinates Computed

Algorithmic Options-

Maximum Iterations 30
 Convergence Criterion ,00100
 Minimum S-stress ,00500
 Missing Data Estimated by Ulbounds

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	,00373	

Iterations stopped because
 S-stress is less than ,005000

Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities) in the partition (row, matrix, or entire data) which is accounted for by their corresponding distances. Stress values are Kruskal's stress formula 1.

For matrix
 Stress = ,00352 RSQ = ,99994

— Configuration derived in 2 dimensions

Stimulus Coordinates

Stimulus Number	Stimulus Name	Dimension	
		1	2
1	ATENAS	-,1860	1,9206
2	BERLÍN	-,2171	-,3693
3	ESTOCOLM	-,9986	-1,0381
4	LONDRES	,7928	-,9659
5	MADRID	2,1610	-,0867
6	MOSCÚ	-2,2021	-,0454
7	PARÍS	,8881	-,5269
8	ROMA	,6604	,8665
9	VARSOVIA	-,7562	-,0390
10	VIENA	-,1423	,2844

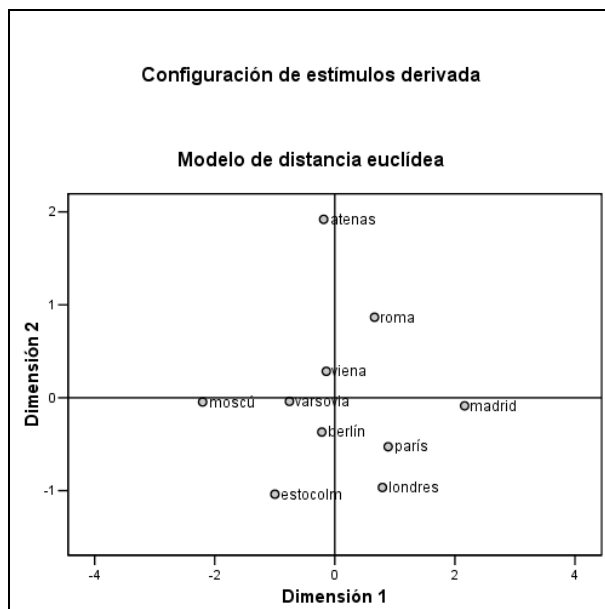


Figura 8-50

8.11 Escalamiento multidimensional no métrico en SPSS

En este caso se trata de estudiar las relaciones que existen entre 10 tipos diferentes de delitos. Para ello se han formado todos los pares posibles de delitos y se han ordenado estos pares en función de su similitud como sigue:

Delito	Homic	Atraco	Robo	Violación	Agresi	Desfal	Chant	Secues	Contra	Terr
Homic	0
Atraco	21	0
Robo	11	2	0
Violaci	3	7	9	0
Agresi	6	4	12	5	0
Desfal	45	26	13	40	36	0
Chantaj	29	28	25	20	22	37	0	.	.	.
Secues	18	23	16	15	14	41	10	0	.	.
Contrab	34	31	24	30	27	43	42	38	0	.
Terroris	8	35	33	32	17	44	19	1	39	0

A partir de esta matriz de similitudes entre delitos (contenida en el fichero 8-6.sav, realizar un escalamiento no métrico que sitúe estos delitos sobre un mapa perceptual que aclare la clasificación y las relaciones convenientemente.

Comenzamos introduciendo los datos de las similitudes entre delitos en el editor de SPSS y a continuación se selecciona *Analizar* → *Escalas* → *Escalamiento*

multidimensional (Figura 8-51). Se obtiene la pantalla de entrada del procedimiento de la Figura 8-52. Con los botones *Opciones* y *Modelo* se obtienen pantallas que se rellenan como se indica en la Figuras 8-53 y 8-54 (se observa *Ordinal* en *Nivel de medida*).

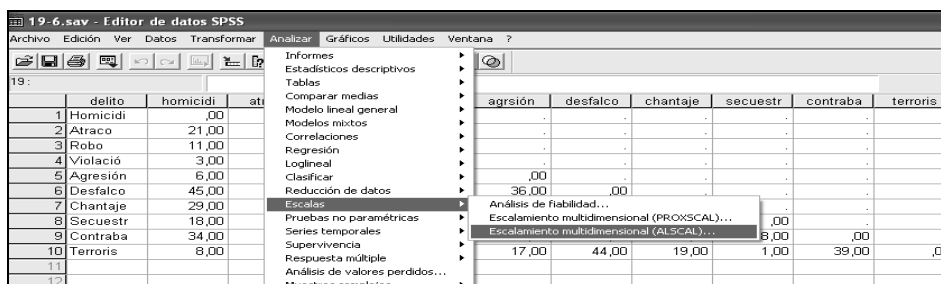


Figura 8-51

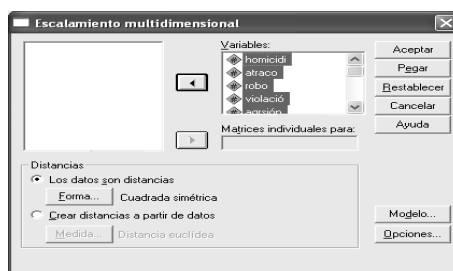


Figura 8-52

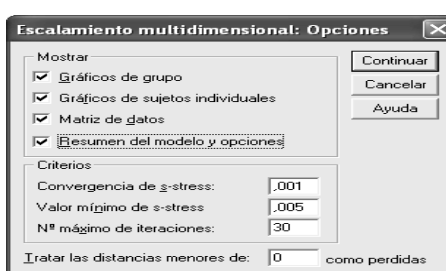


Figura 8-53

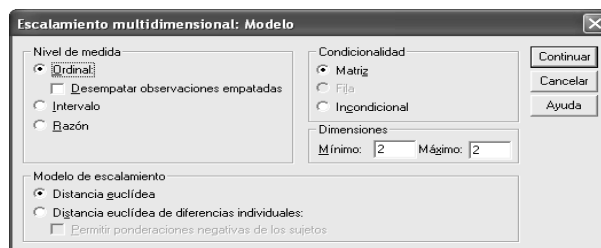


Figura 8-54

El análisis MDS muestra que la solución en dos dimensiones proporciona un buen ajuste (buena convergencia y buenos valores de Stress y RSQ), proporcionando la matriz X de coordenadas en dos dimensiones. La interpretación gráfica de esta matriz se observa en la Figura 8-55. La salida es la siguiente:

ALSCAL

```
VARIABLES= homicidi atraco robo violació agrsi3n desfalco
chantaje secuestr co
ntraba terroris
/SHAPE=SYMMETRIC
/LEVEL=ORDINAL
```

XLV Contenido

```
/CONDITION=MATRIX
/MODEL=EUCLID
/CRITERIA=CONVERGE(.001) STRESSMIN(.005) ITER(30) CUTOFF(0)
DIMENS(2,2)
/PLOT=DEFAULT ALL
/PRINT=DATA HEADER .
Alscal Procedure Options
```

Data Options-

```
Number of Rows (Observations/Matrix).    10
Number of Columns (Variables) . . .    10
Number of Matrices . . . . .    1
Measurement Level . . . . .    Ordinal
Data Matrix Shape . . . . .    Symmetric
Type . . . . .    Dissimilarity
Approach to Ties . . . . .    Leave Tied
Conditionality . . . . .    Matrix
Data Cutoff at . . . . .    ,000000
```

Model Options-

```
Model . . . . .    Euclid
Maximum Dimensionality . . . . .    2
Minimum Dimensionality . . . . .    2
Negative Weights . . . . .    Not Permitted
```

Output Options-

```
Job Option Header . . . . .    Printed
Data Matrices . . . . .    Not Printed
ConFigurations and Transformations .    Plotted
Output Dataset . . . . .    Not Created
Initial Stimulus Coordinates . . .    Computed
```

Algorithmic Options-

```
Maximum Iterations . . . . .    30
Convergence Criterion . . . . .    ,00100
Minimum S-stress . . . . .    ,00500
Missing Data Estimated by . . . .    Ulbounds
Tiestore . . . . .    45
```

—

>Number of parameters is 20. Number of data values is 45

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	,15675	
2	,09231	,06444
3	,08345	,00886
4	,08168	,00177
5	,08118	,00050

Iterations stopped because
S-stress improvement is less than ,001000

Stress = ,09688 RSQ = ,95104

ConFiguRation derived in 2 dimensions

Stimulus Coordinates

Stimulus Number	Stimulus Name	Dimension	
		1	2
1	HOMICIDI	,9716	,7170
2	ATRACO	-,7309	-,0354
3	ROBO	-,7067	-,1541
4	VIOLACIÓ	,2849	,6861
5	AGRSIÓN	,1973	,5506
6	DESFALCO	-2,2252	-1,3093
7	CHANTAJE	,7028	-1,1680
8	SECUESTR	,9582	-,7533
9	CONTRABA	-1,0016	1,7579
10	TERRORIS	1,5496	-,2915

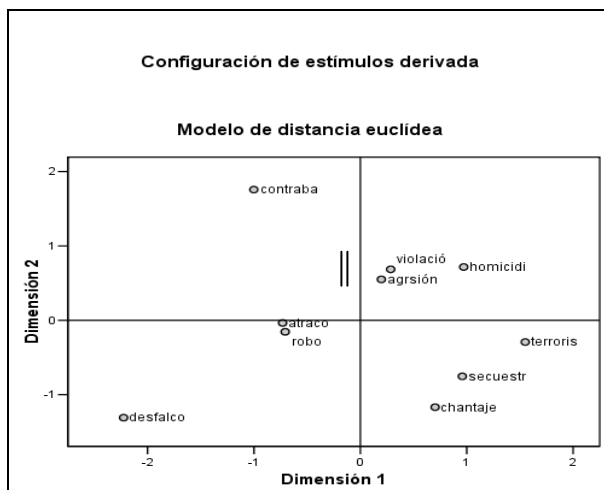


Figura 8-55

L Contenido

Para interpretar el gráfico se han situado dos flechas sobre él. La primera, etiquetada con una barra vertical, muestra una ordenación de los delitos en función de lo personal e impersonal de su naturaleza. Los delitos contra personas (terrorismo, secuestro y chantaje) aparecen en la zona inferior derecha, y a medida que nos desplazamos hacia la zona superior izquierda, encontramos delitos cada vez más impersonales, siendo el más impersonal de todos, el contrabando. La segunda flecha, etiquetada con dos barras verticales, muestra una ordenación de los delitos según su gravedad. Los delitos más graves (homicidio, terrorismo y contrabando) aparecen en la parte más alejada hacia la derecha, y la gravedad descende a medida que nos desplazamos hacia la izquierda de la gráfica, siendo el delito menos grave el desfalco.

En la Figura 8-56 se presenta el gráfico de transformación de proximidades en rangos originales (de 1 a 45) en disparidades. Cuando el gráfico escalonado es muy brusco (escalones muy diferentes en anchura y separación), el ajuste de disparidades en proximidades es malo, mientras que si los puntos ascienden suavemente hacia la derecha, el ajuste es bueno. En nuestro caso observamos un par de escalones demasiado grandes, lo que indica algún problema en el ajuste, pero tampoco demasiado fuerte como para invalidarlo.

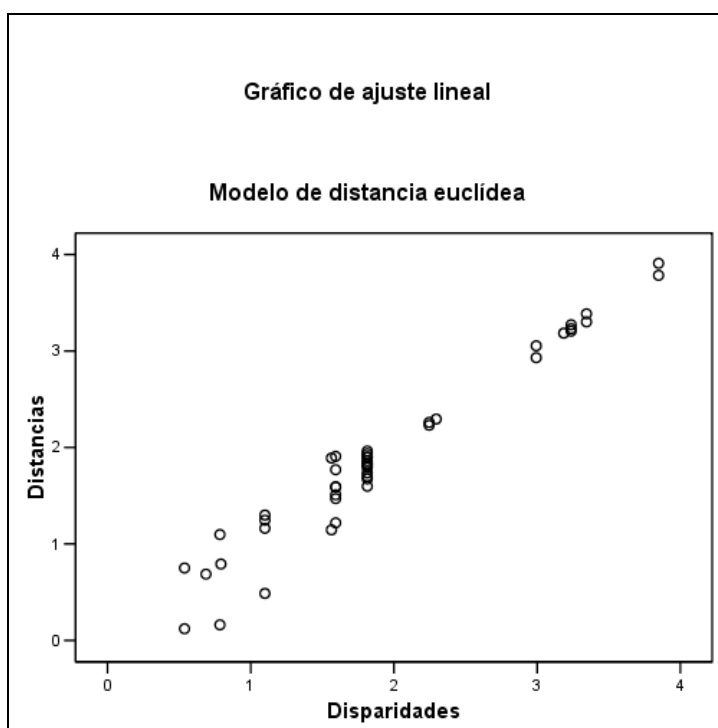



Figura 8-56

8.12 Análisis conjunto en SPSS

Consideremos una compañía que desea lanzar una campaña de mercado para un nuevo limpiador de moquetas y quiere examinar la influencia de los siguientes factores sobre las preferencias del consumidor de artículos de limpieza de moquetas: diseño del paquete (paquete), nombre de la marca (marca), precio del producto (precio), sello de calidad (sello) y garantía de devolución del dinero (dinero). Como niveles factoriales para el diseño del paquete, cada uno de los cuales difiere en la localización del cepillo aplicador del producto, se consideran A*, B*, C*. Como nombres de marca se consideran K2R, Glory y Bisseli. Como niveles de precios para el producto se consideran \$1.19, \$1.39 y \$1.59 y también se consideran 2 niveles (Sí o No) para cada uno de los últimos 2 factores (sello y dinero). Generar un fichero de datos que contenga un diseño o efectos principales que permite el contraste estadístico de varios factores sin tener que comprobar todas las combinaciones de niveles factoriales. Generar también las tarjetas de estímulo para la recogida de datos y análisis posterior de las preferencias mediante análisis conjunto

El procedimiento *Generar diseño ortogonal* genera un fichero de datos que contiene un diseño (efectos principales) que permite el contraste estadístico de varios factores sin tener que comprobar todas las combinaciones de niveles factoriales. Para poder realizar las operaciones de este tipo es muy importante que SPSS muestre etiquetas de valores en vez de valores de las variables. Por ello hay que comprobar que esté seleccionada la opción *Etiquetas de valor* en el menú *Ver* (Figura 8-57), o pulsado el botón *Etiquetas de valor*  en la barra de herramientas estándar del Editor de datos.

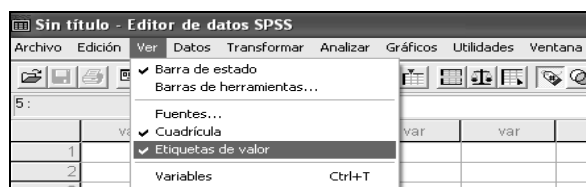


Figura 8-57

A continuación se elige *Datos* → *Diseño ortogonal* → *Generar* (Figura 8-58) para obtener la pantalla de entrada del procedimiento *Generar diseño ortogonal* de la Figura 8-59. Introducimos el nombre del primer factor y su etiqueta en la Figura 8-60, pulsamos en el botón *Añadir* y el factor se incorpora al diseño (Figura 8-61). Se selecciona con el ratón su nombre sobre la pantalla *Generar diseño ortogonal* (Figura 8-62), se hace clic en *Definir valores* y se rellena la pantalla resultante como se indica en la Figura 8-63. Se pulsa en *Continuar* y ya aparece la pantalla *Generar diseño* con el nuevo factor y sus valores incorporado (Figura 8-64). A continuación se introduce el nombre y la etiqueta de un nuevo factor en

la pantalla *Generar diseño* (Figura 8-65) y se pulsa *Añadir*. Se selecciona el nuevo factor, se pulsa en *Definir valores* y se rellena la pantalla resultante como se indica en la Figura 8-66.

Se pulsa en *Continuar* y ya aparece la pantalla *Generar diseño* con los dos factores definidos hasta ahora y sus valores incorporados (Figura 8-67). Se repite el proceso hasta generar los 5 factores con sus etiquetas y valores (Figura 8-68). Pulsando en el botón *Archivo* se puede guardar el diseño con el nombre por defecto (*orto.sav*) o con cualquier otro a especificar en la Figura 8-69.

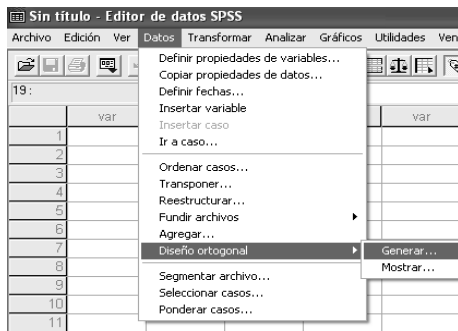


Figura 8-58

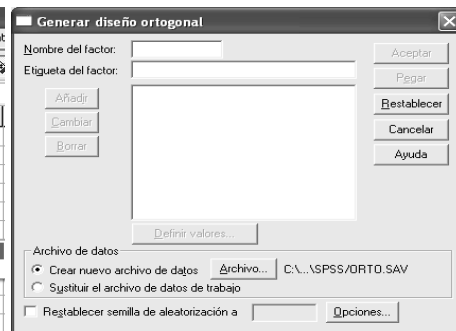


Figura 8-59



Figura 8-60

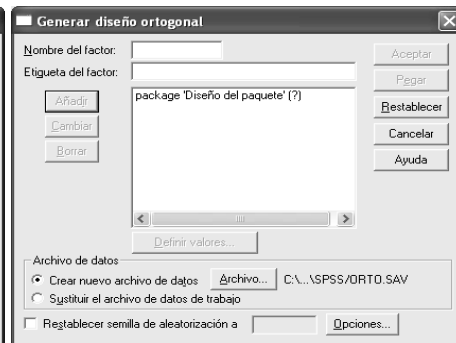


Figura 8-61

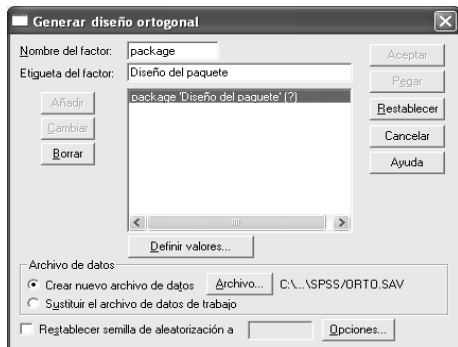


Figura 8-62

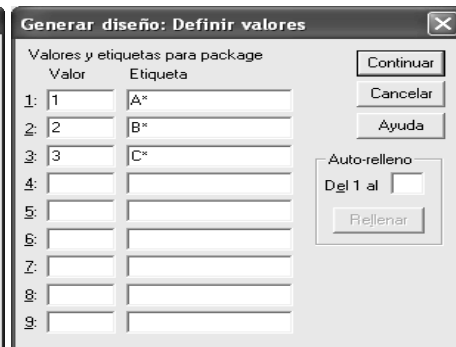


Figura 8-63

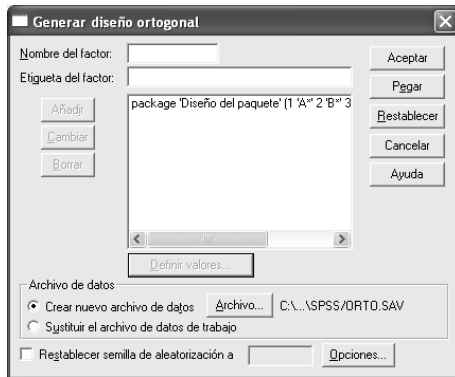


Figura 8-64



Figura 8-65

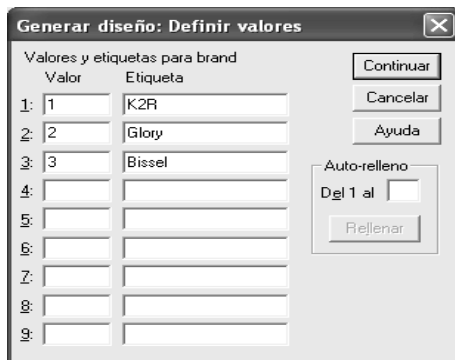


Figura 8-66

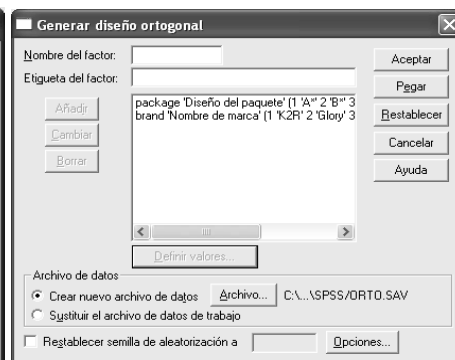


Figura 8-67

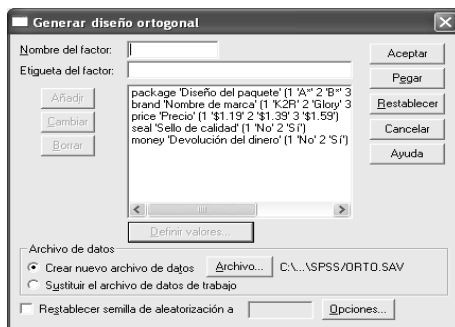


Figura 8-68

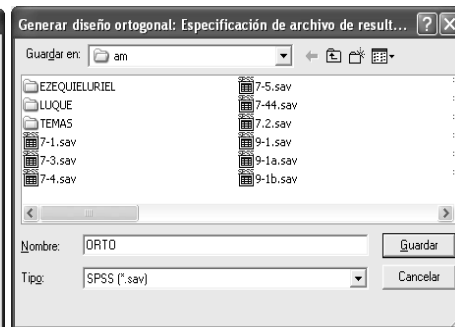


Figura 8-69

En la pantalla *Generar diseño ortogonal* se puede utilizar la casilla *Restablecer semilla de aleatorización* para controlar la generación de los números aleatorios para la creación del diseño ortogonal (Figura 8-70). Además, mediante el botón *Opciones* la pantalla *Generar diseño ortogonal* se pueden especificar un número mínimo de casos a incluir en el diseño ortogonal y definir el número de casos de reserva prorateados por los sujetos pero no incluidos por el análisis conjunto (Figura 8-71).

Los casos de reserva se utilizan en la encuesta, pero el procedimiento CONJOINT no los utiliza al estimar las utilidades. Los casos de reserva se generan a partir de otro plan aleatorio, no a partir del plan experimental de efectos principales y no duplican los perfiles experimentales. La opción *Combinar al azar con los otros casos* permite mezclar aleatoriamente los casos de reserva con los casos experimentales. Pulsando *Continuar* y *Aceptar* se genera el diseño ortogonal (Figura 8-72).

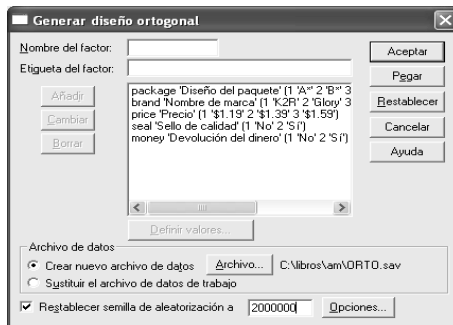


Figura 8-70

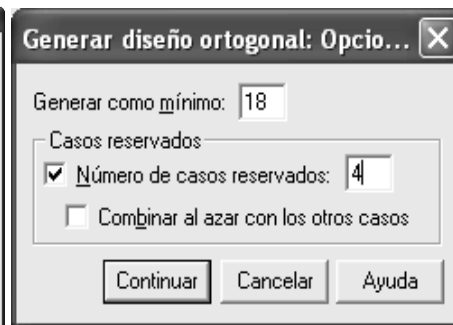


Figura 8-71

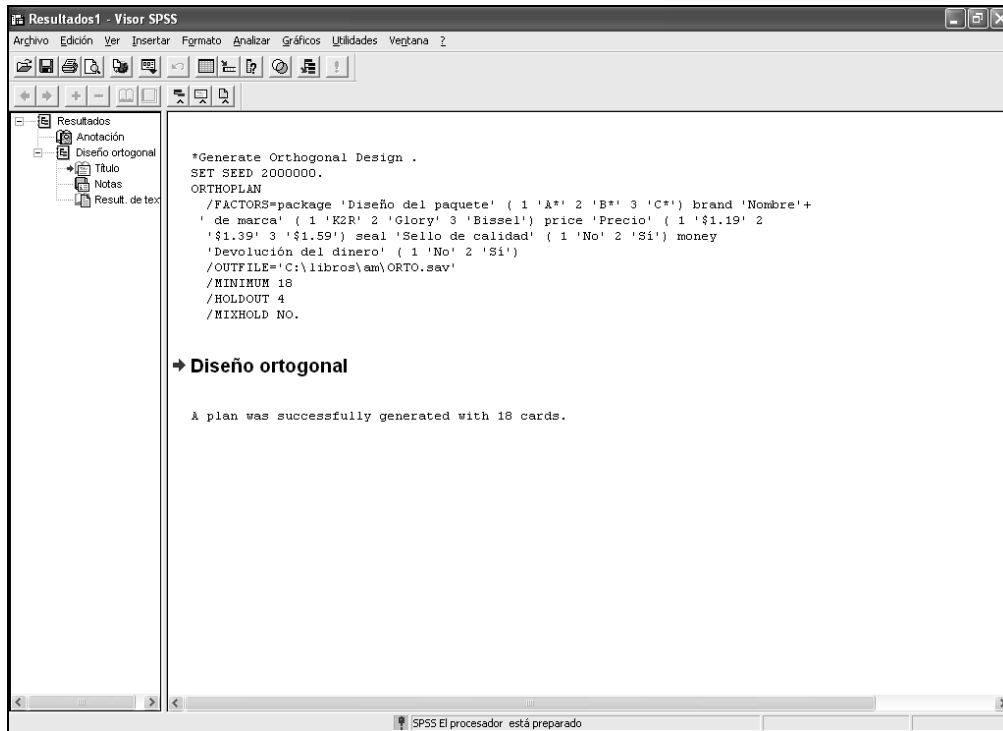


Figura 8-72

Ya hemos realizado el diseño del plan y ahora debemos situar cada concepto completo en un perfil separado con el objeto de presentárselo a los encuestados en forma de tarjeta (cada caso del diseño ortogonal se muestra como un perfil). Los perfiles pueden visualizarse y personalizarse, siendo posible producir cada concepto como una página separada, añadir títulos y notas a pie de página, controlar el espaciado, etc. Mediante el procedimiento *Visualizar diseño experimental* es posible mostrar el diseño generado por el procedimiento *Generar diseño ortogonal* (o cualquier otro diseño recogido en un fichero de datos de trabajo) en formato de listado de borrador o como perfiles a mostrar a los sujetos en un análisis conjunto.

Para visualizar el diseño cargamos el fichero de datos con el diseño ortogonal *orto.sav* recién generado (Figura 8-73). A continuación elegimos *Datos* → *Diseño ortogonal* → *Mostrar* (Figura 8-74) para obtener la pantalla *Mostrar el diseño* (Figura 8-75). La opción *Listado para el experimentador* permite mostrar el diseño en formato de borrador diferenciando los perfiles de reserva de los perfiles experimentales y listando los posibles perfiles de simulación (*tarjetas de estímulo*) de modo separado a continuación de los perfiles experimentales y de reserva. La opción *Perfiles para sujetos* produce perfiles que pueden presentarse a los sujetos y la opción *Saltos de página después de cada perfil* muestra cada perfil en una página nueva. Si se pulsa en el botón *Títulos* de la pantalla *Mostrar el diseño* se puede situar un título y un pie para el perfil (Figura 8-76) que aparecerán en el encabezado y en el pie de cada nuevo perfil.



Figura 8-73

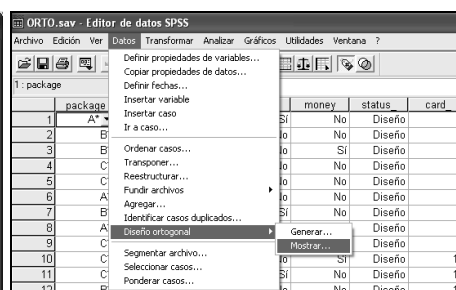


Figura 8-74

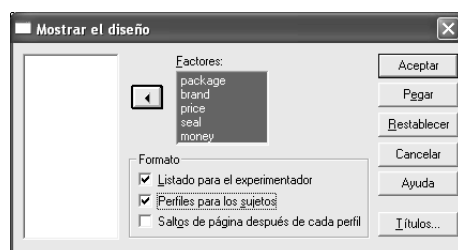


Figura 8-75

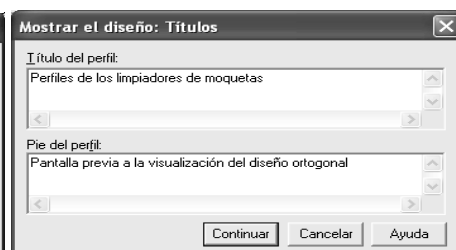


Figura 8-76

LVI Contenido

Al pulsar en *Continuar* y en *Aceptar* se muestran las tarjetas del diseño otogonal generado.

Plancards:

Title: Perfiles de los limpiadores de moquetas

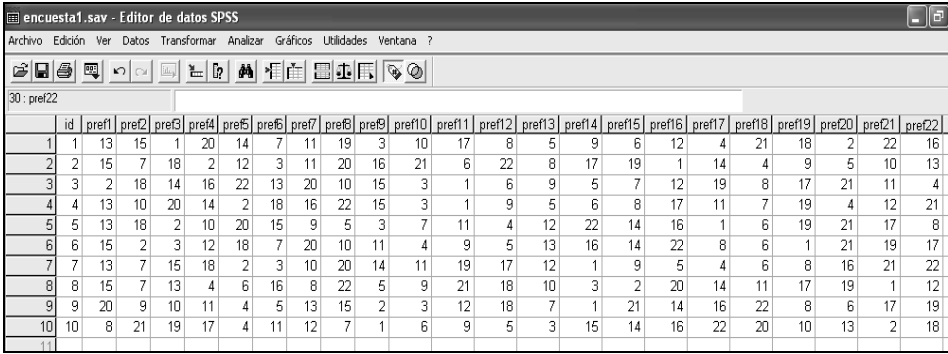
Card 1

Diseño del paquete A*
Nombre de marca Glory
Precio \$1.39
Sello de calidad Sí
Devolución del dinero No

Card 2

Diseño del paquete B*
Nombre de marca K2R
Precio \$1.19
Sello de calidad No
Devolución del dinero No

Una vez generado el diseño ortogonal y preparadas las tarjetas de estímulos, se recogen los datos de preferencias de 10 sujetos (fichero ENCUESTA.SAV según la Figura 8-77) que ordenan los perfiles del más al menos preferido (cada sujeto asigna un número entre 1 y 22 a cada perfil). Realizar un análisis de las preferencias mediante la técnica del análisis conjunto.



30 : pref22

	id	pref1	pref2	pref3	pref4	pref5	pref6	pref7	pref8	pref9	pref10	pref11	pref12	pref13	pref14	pref15	pref16	pref17	pref18	pref19	pref20	pref21	pref22
1	1	13	15	1	20	14	7	11	19	3	10	17	8	5	9	6	12	4	21	18	2	22	16
2	2	15	7	18	2	12	3	11	20	16	21	6	22	8	17	19	1	14	4	9	5	10	13
3	3	2	18	14	16	22	13	20	10	15	3	1	6	9	5	7	12	19	8	17	21	11	4
4	4	13	10	20	14	2	18	16	22	15	3	1	9	5	6	8	17	11	7	19	4	12	21
5	5	13	18	2	10	20	15	9	5	3	7	11	4	12	22	14	16	1	6	19	21	17	8
6	6	15	2	3	12	18	7	20	10	11	4	9	5	13	16	14	22	8	6	1	21	19	17
7	7	13	7	15	18	2	3	10	20	14	11	19	17	12	1	9	5	4	6	8	16	21	22
8	8	15	7	13	4	6	16	8	22	5	9	21	18	10	3	2	20	14	11	17	19	1	12
9	9	20	9	10	11	4	5	13	15	2	3	12	18	7	1	21	14	16	22	8	6	17	19
10	10	8	21	19	17	4	11	12	7	1	6	9	5	3	15	14	16	22	20	10	13	2	18

Figura 8-77

Una vez generado el diseño ortogonal (recogido en el fichero *orto.sav* y recogidos los datos sobre las preferencias en las tarjetas de estímulos provenientes de los sujetos (recogidos en el fichero *encuesta.sav*), sólo resta analizar los datos utilizando el procedimiento CONJOINT. Para ejecutar este procedimiento se utilizará la sintaxis de SPSS, abriendo un fichero de sintaxis mediante *Nuevo* → *Sintaxis* (Figura 8-78) y escribiendo la sintaxis de la Figura 8-79.

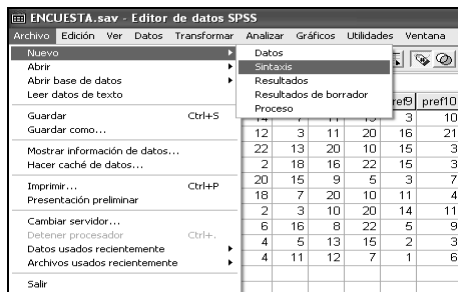


Figura 8-78

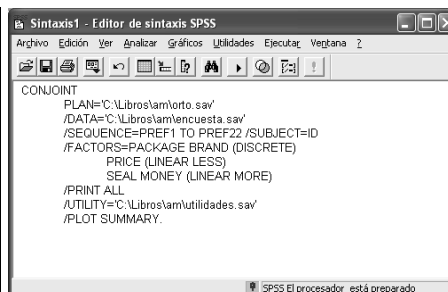


Figura 8-79

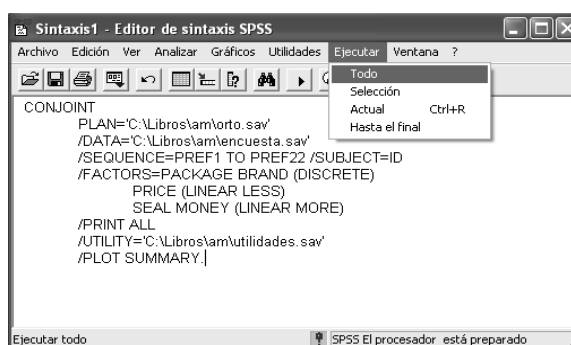


Figura 8-80

La primera línea de la sintaxis es la llamada al procedimiento CONJOINT. El subcomando PLAN identifica el fichero que contiene el diseño ortogonal. El subcomando DATA identifica el fichero que contiene los resultados codificados de la encuesta. El subcomando SEQUENCE indica que los resultados de la encuesta recogidos en el fichero ENCUESTA.SAV han sido codificados en orden secuencial, empezando con la tarjeta más preferida 'pref1' y terminando con la menos preferida 'pref22', siendo 22 el número de tarjetas generadas. El subcomando SUBJECT identifica la variable que contiene el número del sujeto encuestado. El subcomando FACTORS especifica los factores (variables) definidos en el fichero que contiene el diseño ortogonal identificado por el subcomando PLAN. Se observa que los factores *package* y *brand* se definen como discretos (variables categóricas) y no se hace ninguna asunción sobre la relación entre los niveles y los datos. El factor *price* se define como menos lineal (variable lineal para la que los consumidores prefieren los precios más bajos). Los factores *seal* y *money* se definen como más lineales (variables lineales para las que se supone que los consumidores prefieren aquella para la que el producto tenga sello de calidad y se garantice la devolución del dinero. El subcomando PRINT permite controlar la salidas de texto y ALL especifica que se presenten tanto los resultados de los datos experimentales, como los de simulación. El subcomando UTILITY identifica el fichero en el que CONJOINT guardará las utilidades calculadas generándose un caso por cada sujeto encuestado. El subcomando PLOT solicita las salidas gráficas.

LVII Contenido

La palabra clave SUMMARY produce un diagrama de barras para cada variable, mostrando las puntuaciones de la utilidad para cada categoría de esa variable y un gráfico que muestra las puntuaciones de importancia de resumen por sujetos con la palabra clave SUBJECT. Con *Ejecutar* → *Todo* (Figura 8-80) se tiene la salida del procedimiento CONJOINT, que empieza con los factores del diseño ortogonal (Figura 8-81).

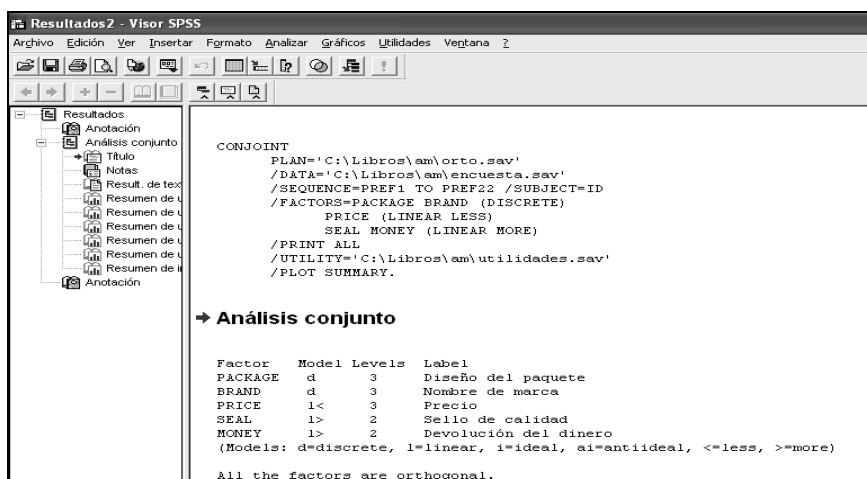


Figura 8-81

Los resultados del procedimiento CONJOINT se ofrecen ordenadamente por sujetos. A continuación se muestra la salida para el primer sujeto

```

SUBJECT NAME:          1

Importance  Utility(s.e.)  Factor  ** Reversed ( 1 reversal )

                PACKAGE      Diseño del paquete
7,21          ,0000 ( ,6303)
              -,6667 ( ,6303)
              ,6667 ( ,6303)
                A*
                B*
                C*

                BRAND      Nombre de marca
12,61        -1,3333 ( ,6303)
              1,0000 ( ,6303)
              ,3333 ( ,6303)
                K2R
                Glory
                Bissel

                PRICE      ** Precio
4,50          ,4167 ( ,5458)
              ,8333 (1,0916)
              1,2500 (1,6375)
              $1.19
              $1.39
              $1.59
              B = ,4167 ( ,5458)

                SEAL      Sello de calidad
48,65         9,0000 ( ,9454)
              18,0000 (1,8908)
              --      No
              ----     Sí
              B = 9,0000 ( ,9454)
  
```

```

                MONEY          Devolución del dinero
27,03   5,0000 ( ,9454)      -          No
        10,0000 (1,8908)     --         Sí
        B = 5,0000 ( ,9454)

        -10,000 (2,1373) CONSTANT

Pearson's R   = ,962                Significance = ,0000

Kendall's tau = ,869                Significance = ,0000
Kendall's tau = ,667 for 4 holdouts Significance = ,0871

```

En la salida se observan las puntuaciones de la utilidad y su error estándar para cada nivel factorial. La utilidad total de una combinación específica se halla sumando los valores de sus puntuaciones correspondientes. Por ejemplo, la utilidad total de un limpiador con diseño de paquete (*package*) B*, marca (*brand*) K2R, precio (*price*) \$1,19, no sellado (*no seal*) y sin garantía de devolución del dinero (*no money back*) sería:

```
utility(package B*) + utility(K2R) + utility($1.19) + utility(no
seal) + utility(no money-back) + constante
```

Realizando la valoración de la utilidad total del limpiador anterior tenemos:

$$(-0,6667) + (-1,3333) + 2,4792 + 9,0000 + 5,0000 + (-12,0620) = 2,4172$$

Las utilidades totales distan algo de los datos observados (aunque teóricamente deberían de coincidir). El error estándar de cada utilidad indica el grado de ajuste del modelo a los datos del sujeto particular considerado. En la salida para el sujeto 1 se observan errores estándar altos para *price*, con lo que puede ser que el modelo lineal no sea el más adecuado para este factor en el caso de este sujeto.

La columna más a la izquierda de la salida anterior presenta las puntuaciones de la importancia de cada factor, junto con un gráfico de barras para dar una idea de cómo se comparan los factores. Las puntuaciones de la importancia se calculan tomando el rango de la utilidad para el factor particular y dividiéndolo por la suma de todos los rangos de las utilidades. Los estadísticos R de Pearson y Tau de Kendall indican también el grado de ajuste de los datos al modelo y representan las correlaciones entre las preferencias observadas y estimadas y, por tanto, deberían ser siempre muy altas. La salida de CONJOINT ofrece este mismo análisis para los 10 sujetos.

Cuando se utiliza el subcomando SUBJECT con CONJOINT, se consiguen, además de los resultados para cada sujeto, unos resultados medios para todo el grupo denominados resultados agrupados del comando CONJOINT y etiquetados SUBFILE SUMMARY *Averaged Importante* (resumen del subfichero en importancia media) y que se presentarán a continuación. Al final de la salida se observa el resumen de reversiones y de simulaciones, que ofrece las

LX Contenido

probabilidades de elegir los perfiles de simulación particulares como perfiles más preferidos, bajo el modelos de probabilidad de elección de la *Máxima Utilidad* (probabilidad de elegir un perfil como el más preferido), bajo el modelo BTL (*Bradley-Terry-Luce*) que calcula la probabilidad de elegir un perfil como el más preferido dividiendo la utilidad del perfil entre la suma de todas las utilidades totales de la simulación, y bajo el modelo logit, que es similar al modelo BTL, pero que utiliza el logaritmo de las utilidades en vez de las utilidades mismas. Se observa que para los diez sujetos de estudio, los tres modelos indican que el perfil de simulación 2 sería el más preferido.

SUBFILE SUMMARY

Averaged Importance	Utility	Factor	
		PACKAGE	Diseño del paquete
35,63	-2,2333	--	A*
	1,8667	--	B*
	,3667		C*
		BRAND	Nombre de marca
14,91	,3667		K2R
	-,3500		Glory
	-,0167		Bissel
		PRICE	Precio
29,41	-1,1083	-	\$1.19
	-2,2167	--	\$1.39
	-3,3250	---	\$1.59
	B = -1,1083		
		SEAL	Sello de calidad
11,17	2,0000	--	No
	4,0000	----	Sí
	B = 2,0000		
		MONEY	Devolución del dinero
8,87	1,2500	-	No
	2,5000	---	Sí
	B = 1,2500		
	7,3833	CONSTANT	
Pearson's R	= ,982		Significance = ,0000
Kendall's tau	= ,892		Significance = ,0000
Kendall's tau	= ,667 for 4 holdouts		Significance = ,0871

—

SUBFILE SUMMARY

Reversal Summary:

2 subjects had 2 reversals
3 subjects had 1 reversals

Reversals by factor:

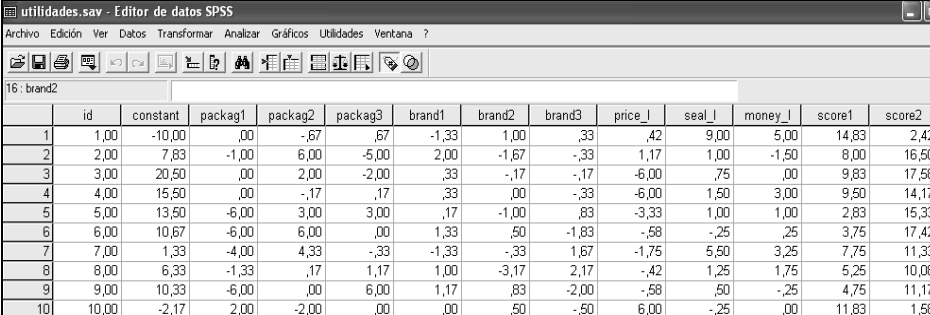
```
PRICE      3
MONEY     2
SEAL      2
BRAND     0
PACKAGE   0
```

Reversal index:

Page	Reversals	Subject
1	1	1
2	2	2
3	0	3
4	0	4
5	0	5
6	1	6
7	0	7
8	0	8
9	1	9
10	2	10

El subcomando PLOT de CONJOINT aporta un modo gráfico de observar los resultados del grupo. La palabra clave SUMMARY produce un diagrama de barras para cada variable mostrando las puntuaciones de la utilidad para cada categoría de esa variable y un gráfico que muestra las puntuaciones de importancia del resumen por sujetos con la palabra clave SUBJECT.

El subcomando UTILITY crea un fichero de datos de SPSS (Figura 8-82) que contiene para cada sujeto las utilidades para los factores DISCRETE, la pendiente y las funciones cuadráticas para los factores LINEAL, DEAL y ANTIDEAL (etiquetas B y C en la salida), la constante de regresión y las puntuaciones estimadas de las preferencias. Estos valores se pueden utilizar en análisis posteriores para realizar gráficos adicionales y gráficos con otros procedimientos.



	id	constant	packag1	packag2	packag3	brand1	brand2	brand3	price_l	seal_l	money_l	score1	score2
1	1,00	-10,00	,00	-,67	-,67	-1,33	1,00	,33	-,42	9,00	5,00	14,83	2,42
2	2,00	7,83	-1,00	6,00	-5,00	2,00	-1,67	-,33	1,17	1,00	-1,50	8,00	16,50
3	3,00	20,50	,00	2,00	-2,00	,33	-,17	-,17	-6,00	,75	,00	9,83	17,58
4	4,00	15,50	,00	-,17	,17	,33	,00	-,33	-6,00	1,50	3,00	9,50	14,17
5	5,00	13,50	-6,00	3,00	3,00	,17	-1,00	,83	-3,33	1,00	1,00	2,83	15,33
6	6,00	10,67	-6,00	6,00	,00	1,33	,50	-1,83	-,58	-,25	,25	3,75	17,42
7	7,00	1,33	-4,00	4,33	-,33	-1,33	-,33	1,67	-1,75	5,50	3,25	7,75	11,33
8	8,00	6,33	-1,33	,17	1,17	1,00	-3,17	2,17	-,42	1,25	1,75	5,25	10,08
9	9,00	10,33	-6,00	,00	6,00	1,17	,83	-2,00	-,58	,50	-,25	4,75	11,17
10	10,00	-2,17	2,00	-2,00	,00	,00	,50	-,50	6,00	-,25	,00	11,83	1,58

Figura 8-82