

# IBM SPSS Direct Marketing 20



*Nota:* Antes de utilizar esta información y el producto que admite, lea la información general en Avisos el p. 109.

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

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

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

Materiales bajo licencia: Propiedad de IBM

**© Copyright IBM Corporation 1989, 2011.**

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

---

# Prefacio

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

## ***Acerca de IBM Business Analytics***

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

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

## ***Asistencia técnica***

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

## ***Asistencia técnica para estudiantes:***

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

## ***Servicio de atención al cliente***

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

## ***Cursos de preparación***

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

## ***Publicaciones adicionales***

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

---

# Contenido

## **Parte I: Manual del usuario**

<b>1</b>	<b>Marketing directo</b>	<b>1</b>
<b>2</b>	<b>Análisis de RFM</b>	<b>2</b>
	Puntuaciones de RFM desde datos de transacciones . . . . .	3
	Puntuaciones de RFM desde datos de cliente . . . . .	5
	Agrupación de RFM . . . . .	6
	Almacenamiento de puntuaciones de RFM desde datos de transacciones . . . . .	9
	Almacenamiento de puntuaciones de RFM desde datos de cliente . . . . .	10
	Resultados de RFM . . . . .	12
<b>3</b>	<b>Análisis de conglomerados</b>	<b>15</b>
	Conglomerados . . . . .	18
<b>4</b>	<b>Perfiles de posibles clientes</b>	<b>20</b>
	Configuración . . . . .	24
	Creación de un campo de respuesta categórica . . . . .	25
<b>5</b>	<b>Índices de respuesta de código postal</b>	<b>26</b>
	Configuración . . . . .	31
	Creación de un campo de respuesta categórica . . . . .	33
<b>6</b>	<b>Propensión a la compra</b>	<b>34</b>
	Configuración . . . . .	38
	Creación de un campo de respuesta categórica . . . . .	40

**7 Prueba de paquete de control 42**

**Parte II: Ejemplos**

**8 Análisis de RFM a partir de datos de transacciones 46**

Datos de transacciones ..... 46  
Ejecución del análisis ..... 46  
Evaluación de los resultados ..... 48  
Fusión de datos de puntuación con datos de clientes ..... 50

**9 Análisis de conglomerados 53**

Ejecución del análisis ..... 53  
Resultados ..... 55  
Selección de registros en función de los conglomerados ..... 63  
    Creación de un filtro en el Visor de modelos de conglomerados ..... 64  
    Selección de registros en función de los valores de campo de conglomerado ..... 66  
Resumen ..... 69

**10 Perfiles de posibles clientes 70**

Consideraciones de los datos ..... 70  
Ejecución del análisis ..... 70  
Resultados ..... 73  
Resumen ..... 76

**11 Índices de respuesta de código postal 77**

Consideraciones de los datos ..... 77  
Ejecución del análisis ..... 77  
Resultados ..... 80  
Resumen ..... 83

<b>12 Propensión a la compra</b>	<b>84</b>
Consideraciones de los datos . . . . .	84
Generación de un modelo predictivo. . . . .	84
Evaluación del modelo. . . . .	88
Aplicación del modelo . . . . .	89
Resumen . . . . .	95
<b>13 Prueba de paquete de control</b>	<b>96</b>
Ejecución del análisis . . . . .	96
Resultados . . . . .	98
Resumen . . . . .	98
<b>Apéndices</b>	
<b>A Archivos muestrales</b>	<b>99</b>
<b>B Avisos</b>	<b>109</b>
<b>Índice</b>	<b>112</b>





***Parte I:***  
***Manual del usuario***



# ***Marketing directo***

La opción Marketing directo ofrece un conjunto de herramientas diseñadas para mejorar el resultado de campañas de marketing directo identificando y adquiriendo características y otras características que definen a diferentes grupos de consumidores y dirigiéndose a grupos concretos para aumentar al máximo los índices de respuesta positivos.

**Análisis de RFM.** Esta técnica identifica a clientes actuales con más posibilidades de responder a una nueva oferta. [Si desea obtener más información, consulte el tema Análisis de RFM en el capítulo 2 el p. 2.](#)

**Análisis de conglomerados.** Es una herramienta de exploración diseñada para revelar agrupaciones (o conglomerados) naturales dentro de sus datos. Por ejemplo, puede identificar diferentes grupos de clientes basándose en diversas características demográficas y de adquisición. [Si desea obtener más información, consulte el tema Análisis de conglomerados en el capítulo 3 el p. 15.](#)

**Perfiles de posibles clientes.** Esta técnica utiliza resultados de una prueba o campaña anterior para crear perfiles descriptivos. Puede utilizar los perfiles para dirigirse a grupos concretos de contactos en campañas futuras. [Si desea obtener más información, consulte el tema Perfiles de posibles clientes en el capítulo 4 el p. 20.](#)

**Índices de respuesta de código postal.** Esta técnica utiliza resultados de una campaña anterior para calcular índices de respuesta de código postal. Los índices de respuesta se pueden utilizar para dirigirse a códigos postales concretos en campañas futuras. [Si desea obtener más información, consulte el tema Índices de respuesta de código postal en el capítulo 5 el p. 26.](#)

**Propensión a la compra.** Esta técnica utiliza resultados de un envío de prueba o campañas anteriores para generar puntuaciones de propensión. Los resultados indican los contactos que tienen más posibilidades de responder. [Si desea obtener más información, consulte el tema Propensión a la compra en el capítulo 6 el p. 34.](#)

**Prueba de paquete de control.** Esta técnica compara campañas de marketing para ver si hay diferencias significativas en la eficacia entre los distintos paquetes u ofertas. [Si desea obtener más información, consulte el tema Prueba de paquete de control en el capítulo 7 el p. 42.](#)

## ***Análisis de RFM***

Análisis de RFM es una técnica utilizada para identificar a clientes actuales que tienen más posibilidades de responder a una nueva oferta. Esta técnica es muy común en el marketing directo. El análisis de RFM se basa en la siguiente teoría simple:

- El factor más importante para identificar a los clientes con más posibilidades de responder a una nueva oferta es la **actualidad**. Los clientes que han realizado adquisiciones recientemente tienen más posibilidades de volver a adquirir nuevos productos que aquellos clientes que han adquirido productos en el pasado.
- El segundo factor más importante es la **frecuencia**. Los clientes que han adquirido más productos tienen más posibilidades de responder que aquellos que han adquirido menos productos.
- El tercer factor más importante es la cantidad total invertida, a la que se le denomina **valor monetario**. Los clientes que han invertido más cantidad (en total en todas las compras) en el pasado tienen más posibilidades de responder que aquellos que han invertido menos.

### ***Funcionamiento del análisis RFM***

- A los clientes se les asigna una puntuación de actualidad en función de la fecha de compra más reciente o el intervalo de tiempo transcurrido desde la compra más reciente. Esta puntuación se basa en una clasificación simple de puntuación de actualidad en un pequeño número de categorías. Por ejemplo, si utiliza cinco categorías, los clientes con las fechas de compra más recientes reciben una puntuación de 5, y los clientes con las fechas de compra menos recientes, recibe una puntuación de 1.
- De la misma forma, a los clientes se les asigna una puntuación de frecuencia, con los valores más altos para las clientes con compras más recientes. Por ejemplo, en una escala del uno al cinco, los clientes que adquieran productos con más frecuencia reciben una puntuación de 5.
- Finalmente, los clientes se clasifican según un valor monetario, con la puntuación valor más alta para el cliente con el mayor valor monetario. Continuando con el ejemplo anterior, los clientes que han invertido mayores cantidades reciben un valor monetario de 5.

Al final, cada cliente recibe cuatro puntuaciones: actualidad, frecuencia, valor monetario y puntuación de RFM combinada, que combina simplemente las tres puntuaciones individuales en una única puntuación. Los “mejores” clientes (aquellos que tienen más posibilidades de responder a una oferta) son los que tengan una mayor puntuación de RFM combinada. Por ejemplo, en una escala del uno al cinco, existen un total de 125 posibles puntuaciones RFM combinadas y la mayor puntuación de RFM combinada es de 555.

**Consideraciones sobre los datos**

- Si las filas de datos representan transacciones (cada fila representa una transacción simple y cada cliente puede tener múltiples transacciones), utilice RFM desde datos de transacciones. Si desea obtener más información, consulte el tema [Puntuaciones de RFM desde datos de transacciones](#) el p. 3.
- Si las filas de datos representan clientes con información de resumen de todas las transacciones (con columnas que contienen puntuaciones de la cantidad total invertida, número total de transacciones y fecha de transacción más reciente), utilice RFM desde datos de cliente. Si desea obtener más información, consulte el tema [Puntuaciones de RFM desde datos de cliente](#) el p. 5.

Figura 2-1

*Datos de transacciones vs. datos de clientes*

Las filas representan transacciones			
ID	Gender	Date	Amount
1	Male	9/25/2005	21
2	Male	1/15/2006	297
4	Male	2/5/2006	249
4	Male	5/7/2005	172
6	Male	4/16/2005	164
6	Male	4/12/2005	286
7	Femal	7/12/2005	400
9	Male		
9	Male		
9	Male		
10	Femal		
10	Femal		

Las filas representan clientes con resúmenes de transacciones				
ID	Gender	Most Recent	Total Amount	Number of Purchases
1	Male	9/25/2005	21	1
2	Male	1/15/2006	297	1
4	Male	2/5/2006	421	2
6	Male	4/16/2005	450	2

**Puntuaciones de RFM desde datos de transacciones****Consideraciones sobre los datos**

El conjunto de datos debe contener variables con la siguiente información:

- Una variable o combinación de variables que identifiquen a cada caso (cliente).
- Una variable con la fecha de cada transacción.
- Una variable con el valor monetario de cada transacción.

Figura 2-2  
 Datos de transacciones de RFM

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

### Creación de puntuaciones de RFM desde datos de transacciones

- ▶ Elija en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Ayudarme a identificar los mejores contactos (Análisis RFM) y pulse en Continuar.
- ▶ Seleccione Datos de transacciones y, a continuación, pulse en Continuar.

Figura 2-3  
 Datos de transacciones, pestaña Variables

**Análisis de RFM a partir de datos de transacciones**

Variables    En intervalo    Guardar    Salida

Cada fila contiene datos para una transacción. Las transacciones serán agregadas por clientes para el análisis.

Variables:

- Línea de producto [ProductLine]
- Número de producto [ProductNumber]

Fecha de transacción:  
 Fecha de adquisición [Date]

Cantidad de transacción:  
 Cantidad de adquisición [Amount]

Método de resumen:  
 Total

Identificadores de cliente:  
 ID de cliente [ID]

Aceptar    Pegar    Restablecer    Cancelar    Ayuda

- ▶ Seleccione la variable que contiene las fechas de transacción.
- ▶ Seleccione la variable que contiene la cantidad monetaria de cada transacción.

- ▶ Seleccione el método para resumir las cantidades de transacción de cada cliente: Total (suma de las transacciones), media, mediana o máxima (cantidad máxima de transacciones).
- ▶ Seleccione la variable o combinación de variables que identifiquen de forma única cada caso del cliente. Por ejemplo, los casos se pueden identificar por un código de identificación único o una combinación del nombre y el apellido.

## ***Puntuaciones de RFM desde datos de cliente***

### ***Consideraciones sobre los datos***

El conjunto de datos debe contener variables con la siguiente información:

- Fecha de compra más reciente o intervalo de tiempo desde la fecha de compra más reciente. Se utilizará para calcular las puntuaciones de actualidad.
- Número total de adquisiciones. Se utilizará para calcular la puntuación de frecuencia.
- Valor monetario de resumen para todas las compras. Se utilizará para calcular valores monetarios. Normalmente, es la suma (total) de todas las compras, pero puede ser la media (promedio), valor máximo (mayor cantidad) u otras medidas de resumen.

Figura 2-4

*Datos de cliente RFM*

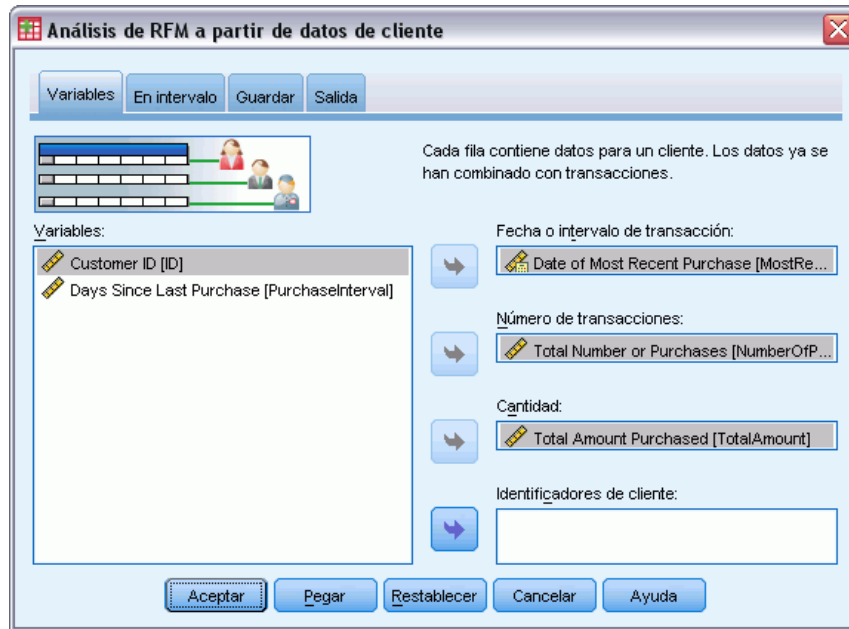
ID	TotalAmount	MostRecent	NumberOfPurchases
1	485.00	09/04/2006	5
2	350.00	11/10/2005	4
3	233.00	06/04/2005	2
4	936.00	08/18/2006	7
5	359.00	07/07/2006	3
6	249.00	07/16/2006	3
7	1089.00	02/15/2006	7
8	423.00	08/21/2006	4
9	689.00	08/31/2006	7
10	325.00	10/13/2005	3

Si desea escribir puntuaciones RFM en un nuevo conjunto de datos, el conjunto de datos activo también debe contener una variable o una combinación de variables que identifiquen a cada caso (cliente).

### ***Creación de puntuaciones de RFM desde datos de cliente***

- ▶ Elija en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Ayudarme a identificar los mejores contactos (Análisis RFM) y pulse en Continuar.
- ▶ Seleccione Datos de cliente y, a continuación, pulse en Continuar.

Figura 2-5  
 Datos de cliente, pestaña Variables



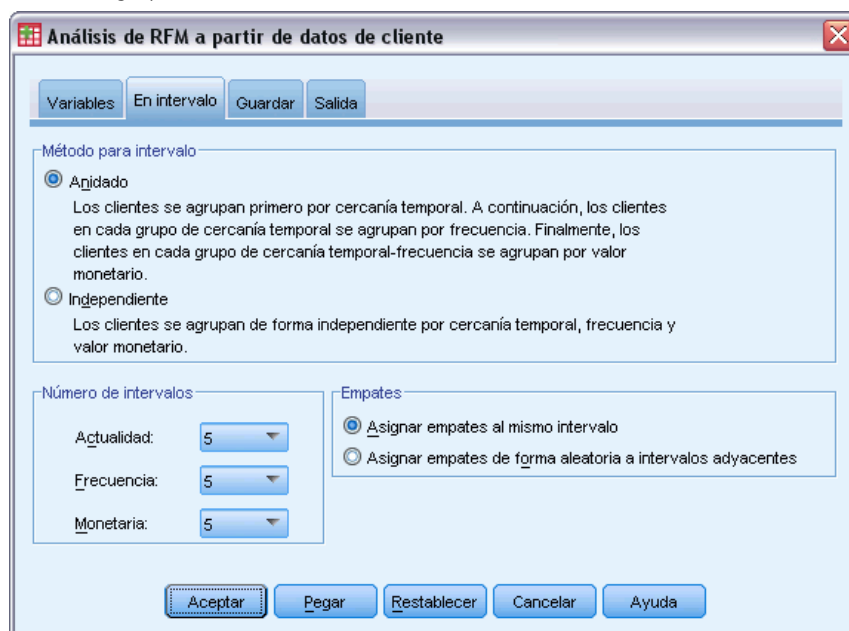
- ▶ Seleccione la variable que contiene la fecha de transacción más reciente o un número que represente un intervalo de tiempo desde la transacción más reciente.
- ▶ Seleccione la variable que contienen el número total de transacciones de cada cliente.
- ▶ Seleccione la variable que contiene la cantidad monetaria de resumen de cada cliente.
- ▶ Si desea escribir las puntuaciones RFM en un nuevo conjunto de datos, seleccione la variable o combinación de variables que identifican de forma única a cada cliente. Por ejemplo, los casos se pueden identificar por un código de identificación único o una combinación del nombre y el apellido.

## ***Agrupación de RFM***

El proceso de agrupar un gran número de valores numéricos en un pequeño número de categorías se denomina **agrupación**. En análisis de RFM, los grupos se clasifican en categorías. Puede utilizar la pestaña Agrupación para modificar el método utilizado para asignar los valores de actualidad, frecuencia y monetarios a esos intervalos.



Figura 2-6  
Pestaña Agrupación de RFM



### **Método para intervalo**

**Anidadas.** En la agrupación anidada, se asigna una puntuación simple a los valores de actualidad. En cada puntuación de actualidad, a los clientes se les asigna una puntuación de frecuencia y en cada puntuación de frecuencia a los clientes se les asigna un valor monetario. Este método tiende a proporcionar una distribución más uniforme de puntuaciones de RFM combinadas, pero tiene la desventaja de hacer que las puntuaciones de frecuencia y monetarias son más difíciles de interpretar. Por ejemplo, es posible que una puntuación de frecuencia de 5 para un cliente con una puntuación de actualidad de 5 no signifique lo mismo que una puntuación de frecuencia de 5 para un cliente con una puntuación de actualidad de 4, ya que la puntuación de frecuencia depende de la puntuación de actualidad.

**Independiente.** Las puntuaciones simples se asignan a actualidad, frecuencia y valores monetarios. Los tres rangos se asignan de forma independiente. La interpretación de cada uno de los tres componentes de RFM no es ambigua; una puntuación de frecuencia de 5 para un cliente significa lo mismo que una puntuación de frecuencia de 5 para otro cliente, con independencia de sus puntuaciones de actualidad. Para muestras menores tiene la desventaja de producir una distribución menos uniforme de las puntuaciones de RFM combinadas.

### **Número de intervalos**

El número de categorías (intervalos) que utiliza cada componente para crear puntuaciones RFM. El número total de puntuaciones de RFM combinadas es el producto de los tres valores. Por ejemplo, 5 grupos de actualidad, 4 grupos de frecuencia y 3 grupos monetarios crean un total de 60 puntuaciones de RFM combinadas posibles, entre 111 y 543.

- El valor por defecto es de 5 para cada componente, que crean un total de 125 puntuaciones de RFM combinadas posibles, entre 111 y 555.
- El número máximo de grupos permitidos para cada componente es nueve.

### **Empates**

Un “empate” son simplemente dos o más valores de actualidad, frecuencia o monetarios. Lo ideal es que desee tener aproximadamente el mismo número de clientes en cada grupo, pero un gran número de valores empatados que pueden afectar a la distribución de los grupos. Existen alternativas para gestionar los empates:

- **Asignación de empates al mismo grupo.** Este método asigna siempre valores empatados al mismo grupo, con independencia de cómo afecta a la distribución de grupos. Proporciona un método de agrupación coherente: Si dos clientes tienen el mismo valor de actualidad, siempre se les asignará la misma puntuación de actualidad. En un ejemplo extremo, sin embargo, puede tener 1.000 clientes y 500 de ellos realizando su adquisición más reciente en la misma fecha. En una clasificación del 1 al 5, el 50% de los clientes recibirían una puntuación de actualidad de 5, en lugar del 20% deseado.

Tenga en cuenta que con el método de agrupación anidado, “coherencia” es algo más complicado para las puntuaciones de frecuencia y de valor monetario, ya que las puntuaciones de frecuencia se asignan en grupos de puntuación de actualidad y los valores monetarios se asignan en grupos de puntuaciones de frecuencia. Es posible que dos clientes con el mismo valor de frecuencia no tengan la misma puntuación de frecuencia si tampoco tienen la misma puntuación de actualidad, con independencia de la forma en la que se gestionan los valores empatados.

- **Asignar aleatoriamente los empates.** De esta forma se asegura una distribución uniforme de grupos al asignar un factor de varianza muy pequeño a los empates antes de la puntuación; para que al asignar los valores a los grupos de puntuaciones, no existan valores empatados. Este proceso no afecta a los valores originales. Sólo se utiliza para deshacer los empates. Mientras que este método produce una distribución uniforme de los grupos (aproximadamente el mismo número de clientes en cada grupo), puede resultar en puntuaciones muy diferentes para clientes que parecen tener valores de actualidad, frecuencia y/o monetarios iguales: especialmente si el número total de clientes es relativamente pequeño y/o el número de empates es relativamente alto.

Tabla 2-1

*Asignación de empates al mismo grupo vs. Asignación de valores de forma aleatoria*

<b>ID</b>	<b>Compra más reciente (Actualidad)</b>	<b>Asignación de empates al mismo grupo</b>	<b>Asignación de empates aleatoriamente</b>
1	10/29/2006	5	5
2	10/28/2006	4	4
3	10/28/2006	4	4
4	10/28/2006	4	5
5	10/28/2006	4	3
6	9/21/2006	3	3
7	9/21/2006	3	2
8	8/13/2006	2	2

ID	Compra más reciente (Actualidad)	Asignación de empates al mismo grupo	Asignación de empates aleatoriamente
9	8/13/2006	2	1
10	6/20/2006	1	1

- En este ejemplo, la asignación de empates al mismo grupo resulta en una distribución de grupos no uniforme: 5 (10%), 4 (40%), 3 (20%), 2 (20%), 1 (10%).
- La asignación de empates de forma aleatoria da como resultado el 20% en cada grupo, pero para lograr este resultado, los cuatro casos con un valor de fecha de 10/28/2006 se asignan a 3 grupos diferentes y los 2 casos con un valor de fecha de 8/13/2006 también se asignan a grupos diferentes.

Tenga en cuenta que la manera en que se asignan los empates a grupos diferentes es totalmente aleatorio (con la limitación de que el resultado final es igual al número de casos de cada grupo). Si ha calculado un segundo conjunto de puntuaciones con el mismo método, la puntuación de cualquier caso con un valor empatado puede cambiar. Por ejemplo, las puntuaciones de actualidad de 5 y 3 para los casos 4 y 5 respectivamente se pueden cambiar la segunda vez.

## Almacenamiento de puntuaciones de RFM desde datos de transacciones

RFM desde datos de transacciones siempre crea un nuevo conjunto de datos agregado con una fila para cada cliente. Utilice la pestaña Guardar para especificar las puntuaciones y otras variables que desee guardar y la ubicación en la que desea guardarlas.

Figura 2-7  
Datos de transacciones, pestaña Guardar

Análisis de RFM a partir de datos de transacciones

Variables En intervalo Guardar Salida

Variables:

Guardar	Contenido	Nombre de variable guardada
<input checked="" type="checkbox"/>	Fecha de transacción más reciente	Fecha_más_reciente
<input checked="" type="checkbox"/>	Número de transacciones	Recuento_transacciones
<input checked="" type="checkbox"/>	Cantidad	Cantidad
<input checked="" type="checkbox"/>	Puntuación de actualidad	Puntuación_actualidad
<input checked="" type="checkbox"/>	Puntuación de frecuencia	Puntuación_frecuencia
<input checked="" type="checkbox"/>	Puntuación monetaria	Puntuación_monetaria
<input checked="" type="checkbox"/>	Puntuación de RFM	Puntuación_RFM

Ubicación

Crear un nuevo conjunto de datos Nombre: RFMData

Escribir un nuevo archivo de datos Examinar...

Aceptar Pegar Restablecer Cancelar Ayuda

### **Variables**

Las variables de ID que identifican de forma única a cada cliente se guardan automáticamente en el nuevo conjunto de datos. Las siguientes variables adicionales se pueden guardar en el nuevo conjunto de datos:

- **Fecha de la transacción más reciente de cada cliente.**
- **Número de transacciones.** El número total de filas de transacciones de cada cliente.
- **Cantidad.** La cantidad de resumen de cada cliente basada en el método de resumen que seleccione en la pestaña Variables.
- **Puntuación de actualidad.** La puntuación asignada a cada cliente basada en la fecha de transacción más reciente. Las puntuaciones más altas indican las fechas de transacción más recientes.
- **Puntuación de frecuencia.** La puntuación asignada a cada cliente basada en el número total de transacciones. Las puntuaciones más altas indican mayor cantidad de transacciones.
- **Valor monetario.** La puntuación asignada a cada cliente basada en la medida de resumen monetario seleccionada. Las puntuaciones más altas indican un valor monetario mayor para la medida de resumen.
- **Puntuación de RFM.** Las tres puntuaciones individuales combinadas en un único valor:  $(actualidad \times 100) + (frecuencia \times 10) + monetaria$ .

Por defecto, todas las variables disponibles en el nuevo conjunto de datos, Cancele la selección de las variables que no desea incluir. También puede especificar sus propios nombres de variable. Los nombres de las variables deben cumplir las normas de denominación de variables.

### **Ubicación**

RFM desde datos de transacciones siempre crea un nuevo conjunto de datos agregado con una fila para cada cliente. Puede crear un nuevo conjunto de datos en la sesión actual o guardar los datos de puntuación de RFM en un archivo de datos externo. Los nombres del conjunto de datos deben cumplir las normas de denominación de variables estándar. (Esta restricción no es aplicable a los nombres de archivos de datos externos.)

## **Almacenamiento de puntuaciones de RFM desde datos de cliente**

En datos del cliente puede añadir las variables de puntuación de RFM al conjunto de datos activo o crear un nuevo conjunto de datos que contiene las variables de puntuación seleccionadas. Utilice la pestaña Guardar para especificar las variables de puntuación y otras variables que desee guardar y la ubicación en la que desea guardarlas.

Figura 2-8  
Datos de cliente, pestaña Guardar

**Análisis de RFM a partir de datos de cliente**

Variables En intervalo Guardar Salida

Nombres de variable guardada

Generar automáticamente nombres exclusivos  
 Nombres personalizados

Variables:

Guardar	Contenido	Nombre de variable guardada
<input checked="" type="checkbox"/>	Puntuación de actualidad	Puntuación_actualidad
<input checked="" type="checkbox"/>	Puntuación de frecuencia	Puntuación_frecuencia
<input checked="" type="checkbox"/>	Puntuación monetaria	Puntuación_monetaria
<input type="checkbox"/>	Puntuación de RFM	Puntuación RFM

Ubicación

Escribir en el conjunto de datos activo  
 Crear un nuevo conjunto de datos Nombre:   
 Escribir un nuevo archivo de datos Examinar...

Aceptar Pegar Restablecer Cancelar Ayuda

### Nombres de las variables guardadas

- **Generar automáticamente nombres únicos.** Si añade variables de puntuación al conjunto de datos activo, se asegurará de que los nuevos nombres de variable son exclusivos. Es muy útil si desea añadir varios conjuntos diferentes de puntuaciones de RFM (basado en criterios diferentes) al conjunto de datos activo.
- **Nombres personalizados.** De esta forma puede asignar sus propios nombres de variable a las variables de puntuación. Los nombres de las variables deben cumplir las normas de denominación de variables.

### Variables

Seleccione las variables de puntuación que desee almacenar:

- **Puntuación de actualidad.** La puntuación asignada a cada cliente basada en el valor de la variable Fecha de transacción o Intervalo seleccionada en la pestaña Variables. Las puntuaciones más altas se asignan a las fechas más recientes o a valores de intervalos menores.
- **Puntuación de frecuencia.** La puntuación asignada a cada cliente basada en la variable Número de transacciones seleccionada en la pestaña Variables. Las puntuaciones más altas se asignan a los valores mayores.
- **Valor monetario.** La puntuación asignada a cada cliente basada en la variable Cantidades seleccionada en la pestaña Variables. Las puntuaciones más altas se asignan a los valores mayores.
- **Puntuación de RFM.** Las tres puntuaciones individuales combinadas en un único valor:  
 $(actualidad * 100) + (frecuencia * 10) + valor\ monetario$ .

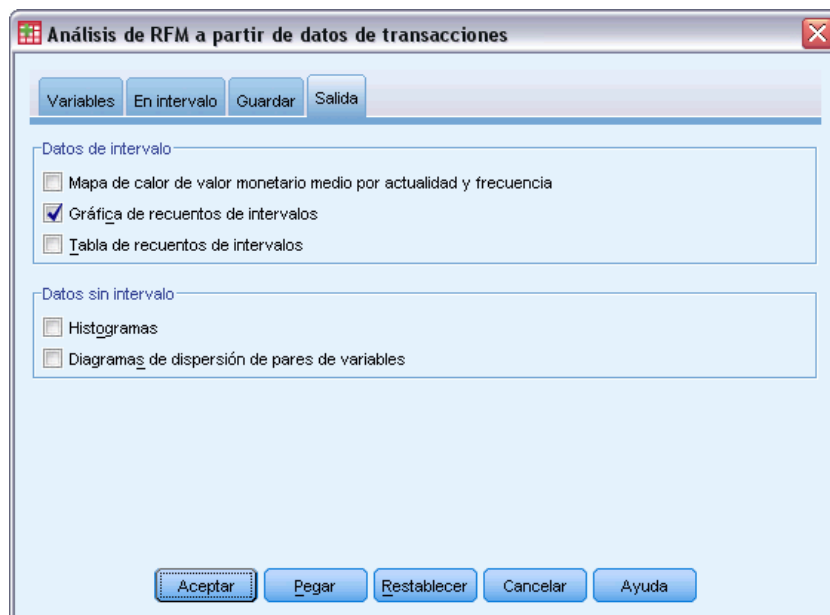
### Ubicación

Para datos de cliente, existen tres alternativas para las que puede guardar nuevas puntuaciones de RFM:

- **Conjunto de datos activo.** Las variables de puntuaciones de RFM se añaden al conjunto de datos activo.
- **Nuevo conjunto de datos.** Las variables de puntuación de RFM seleccionadas y las variables de ID que identifican de manera única a cada cliente (caso) se sobrescribirán en un nuevo conjunto de datos en la sesión actual. Los nombres del conjunto de datos deben cumplir las normas de denominación de variables estándar. Esta opción sólo está disponible si selecciona una o más variables identificadoras de cliente en la pestaña Variables.
- **Archivo.** Las puntuaciones de RFM seleccionadas y las variables de ID que identifican de manera única a cada cliente (caso) se guardarán en un archivo de datos externo. Esta opción sólo está disponible si selecciona una o más variables identificadoras de cliente en la pestaña Variables.

## Resultados de RFM

Figura 2-9  
Pestaña Resultados de RFM



### Datos de intervalo

Los gráficos y tablas de datos agrupados se basan en las puntuaciones de actualidad y frecuencia y valores monetarios calculados.

**Mapa de calor de valor monetario medio según actualidad y frecuencia.** El mapa de calor de la distribución monetaria media muestra el valor monetario medio de las categorías definidas según las puntuaciones de actualidad y frecuencia. Las zonas más oscuras indican un valor monetario medio más alto.

**Gráfico de recuentos de intervalos.** El gráfico de recuentos de intervalos muestra la distribución de intervalos para el método de agrupación seleccionado. Cada barra representa el número de casos que se asignará a cada puntuación de RFM combinada.

- Aunque normalmente desea una distribución bastante homogénea, con todas las barras (o la mayoría de ellas) aproximadamente con la misma altura, debe esperar cierta cantidad de varianza al utilizar el método de agrupación predeterminado que asigna valores empatados al mismo grupo.
- Las fluctuaciones extremas en la distribución de grupos y/o muchos grupos vacíos pueden indicar que debería probar con otro método de agrupación (menos grupos y/o la asignación aleatoria de empates) o reconsiderar la adecuación del análisis de RFM.

**Tabla de recuentos de intervalos.** La misma información que se encuentra en el gráfico de recuentos de intervalos, sólo que expresada en forma de tabla, con recuentos de intervalos en cada casilla.

### ***Datos sin intervalo***

Los gráficos y tablas de datos no agrupados se basan en las variables originales utilizadas para crear puntuaciones de actualidad y frecuencia y valores monetarios.

**Histogramas.** Los histogramas muestran la distribución relativa de valores para las tres variables utilizadas para calcular puntuaciones de actualidad y frecuencia y valores monetarios. No es raro que estos histogramas indiquen distribuciones algo asimétricas en lugar de una distribución normal o simétrica.

El eje horizontal de cada histograma siempre está ordenado desde los valores más bajos en la izquierda hasta los valores más altos a la derecha. Sin embargo, con la actualidad, la interpretación del gráfico depende del tipo de medida de actualidad: fecha o intervalo de tiempo. En el caso de fechas, las barras de la izquierda representan valores más antiguos (una fecha menos reciente tiene un valor inferior que una fecha más reciente). En el caso de intervalos de tiempo, las barras de la izquierda representan los valores más recientes (cuanto menor sea el intervalo de tiempo, más reciente será la transacción).

**Diagramas de dispersión de parejas de variables.** Estos diagramas de dispersión muestran las relaciones entre las tres variables utilizadas para calcular puntuaciones de actualidad y frecuencia y valores monetarios.

También es normal ver agrupaciones lineales apreciables de puntos sobre la escala de frecuencia, ya que a menudo la frecuencia representa un rango relativamente pequeño de valores discretos. Por ejemplo, si el número total de transacciones no supera 15, entonces sólo hay 15 valores de frecuencia posibles (a menos que cuente las transacciones fraccionarias), mientras que podría haber cientos de valores de actualidad y miles de valores monetarios posibles.

La interpretación del eje de actualidad depende del tipo de medida de actualidad: fecha o intervalo de tiempo. En el caso de fechas, los puntos más cercanos al origen representan fechas más antiguas. En el caso de intervalos de tiempo, los puntos más cercanos al origen representan los valores más recientes.



# *Análisis de conglomerados*

El análisis de conglomerados es una herramienta de exploración diseñada para revelar agrupaciones (o conglomerados) naturales dentro de sus datos. Por ejemplo, puede identificar diferentes grupos de clientes basándose en diversas características demográficas y de adquisición.

**Ejemplo.** Las empresas minoristas y de venta de productos para el consumidor suelen aplicar técnicas de conglomeración a los datos que describen los hábitos de consumo, sexo, edad, nivel de ingresos, etc. de los clientes. Estas empresas adaptan sus estrategias de desarrollo de productos y de marketing en función de cada grupo de consumidores para aumentar las ventas y el nivel de fidelidad a la marca.

## ***Análisis de conglomerados: Consideraciones sobre los datos***

**Datos.** Este procedimiento trabaja tanto con campos variables continuos como categóricos. Cada registro (fila) representa un cliente que se va a conglomerar y los campos (variables) representan los atributos en los que se va a basar la conglomeración.

**Orden de registro.** Tenga en cuenta que los resultados pueden depender del orden de los registros. Para minimizar los efectos del orden, es posible que desee considerar ordenar los registros de forma aleatoria. Puede que desee ejecutar el análisis varias veces, con los registros ordenados en distintos órdenes aleatorios para comprobar la estabilidad de una solución determinada.

**Nivel de medida.** La asignación del nivel correcto de medida es importante porque afecta al cálculo de los resultados.

- **Nominal.** Una variable se puede tratar como nominal si sus valores representan categorías que no obedecen a una ordenación intrínseca (por ejemplo, el departamento de la empresa en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- **Ordinal.** Una variable puede tratarse como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca (por ejemplo, los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- **Continuo.** Una variable puede tratarse como escala (continua) cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

Un icono junto a cada campo indica el nivel de medida actual.

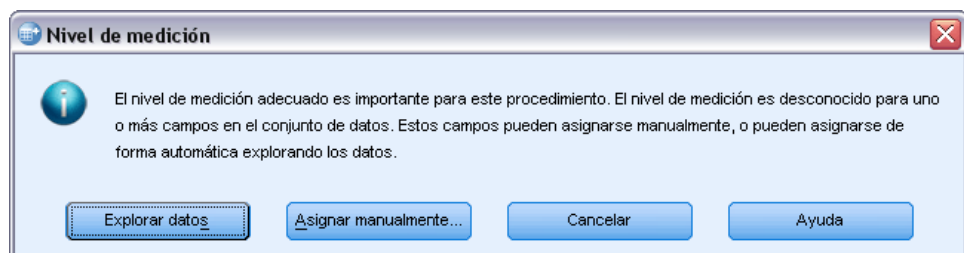
	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		
Ordinal				
Nominal				

Puede cambiar el nivel de medida en una vista de variables del Editor de datos o puede usar el cuadro de diálogo Definir propiedades de variables para sugerir un nivel de medición adecuado para cada campo .

### ***Campos con un nivel de medición desconocido***

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 3-1  
Alerta de nivel de medición



- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

### ***Para obtener un análisis de conglomerados***

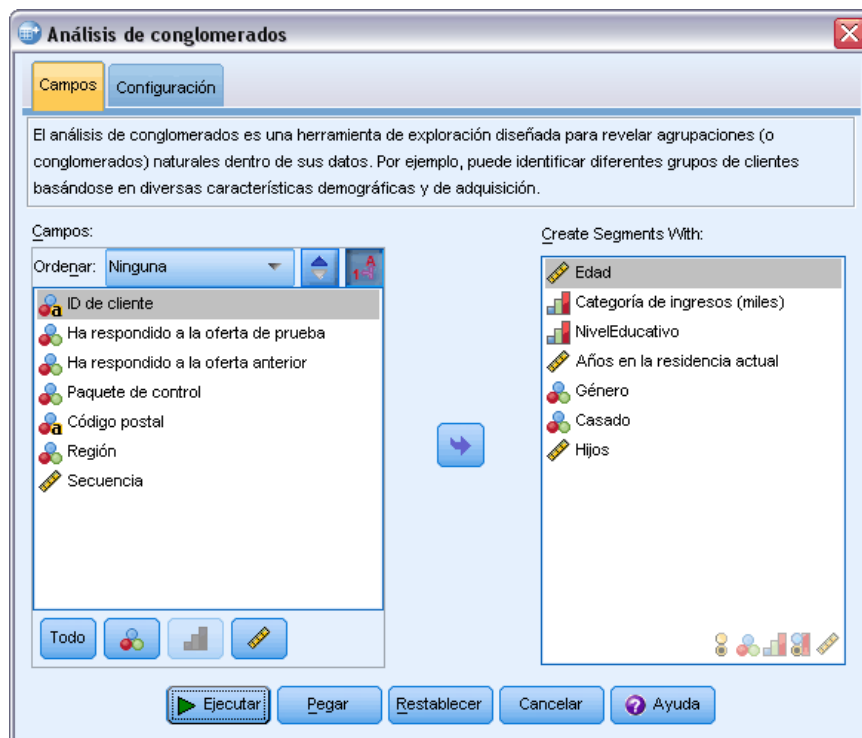
Seleccione en los menús:

Marketing directo > Seleccionar técnica

- Seleccione Segmentar mis contactos en conglomerados.

Figura 3-2

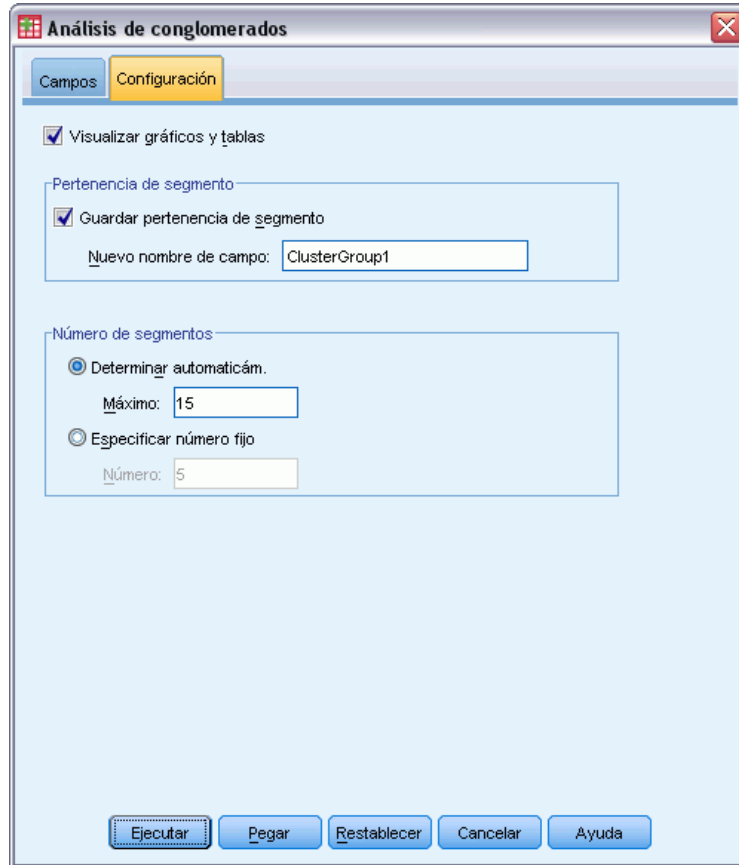
Pestaña Campos de análisis de conglomerados



- Seleccione los campos categóricos (nominal, ordinal) y continuo (escala) que desea utilizar para crear los segmentos.
- Pulse en Ejecutar para ejecutar el procedimiento.

## Conglomerados

Figura 3-3  
Pestaña Configuración de análisis de conglomerados



La pestaña Configuración permite mostrar o eliminar la visualización de gráficos y tablas que describen los segmentos, guardar un nuevo campo en el conjunto de datos que identifica el segmento (conglomerado) de cada registro en el conjunto de datos y especificar cuántos segmentos se incluirán en la solución de conglomerados.

**Visualizar gráficos y tablas.** Muestra tablas y gráficos que describen los segmentos.

**Pertenencia de segmento.** Guarda un nuevo campo (variable) que identifica el segmento al que pertenece cada registro.

- Los nombres de los campos deben cumplir las reglas de nombres de IBM® SPSS® Statistics.
- El campo de pertenencia de segmento no puede tener un nombre que ya exista en el conjunto de datos. Si ejecuta este procedimiento más de una vez en el conjunto de datos, necesitará especificar un nombre diferente cada vez.
- **Número de segmentos.** Controla cómo se determina el número de segmentos.
- **Determinar automáticamente.** El procedimiento determinará automáticamente el “mejor” número de segmentos, hasta el máximo especificado.

**Especificar número fijo.** El procedimiento producirá el número especificado de segmentos.

## ***Perfiles de posibles clientes***

Esta técnica utiliza resultados de una prueba o campaña anterior para crear perfiles descriptivos. Puede utilizar los perfiles para dirigirse a grupos concretos de contactos en campañas futuras. El campo Respuesta indica quién ha respondido a la campaña o prueba anterior. La lista Perfiles contiene las características que desea utilizar para crear el perfil.

**Ejemplo.** Basándose en los resultados de un correo de prueba, el departamento de marketing directo de una empresa desea generar perfiles de los tipos de clientes con más posibilidades de responder a la oferta, en función de la información demográfica.

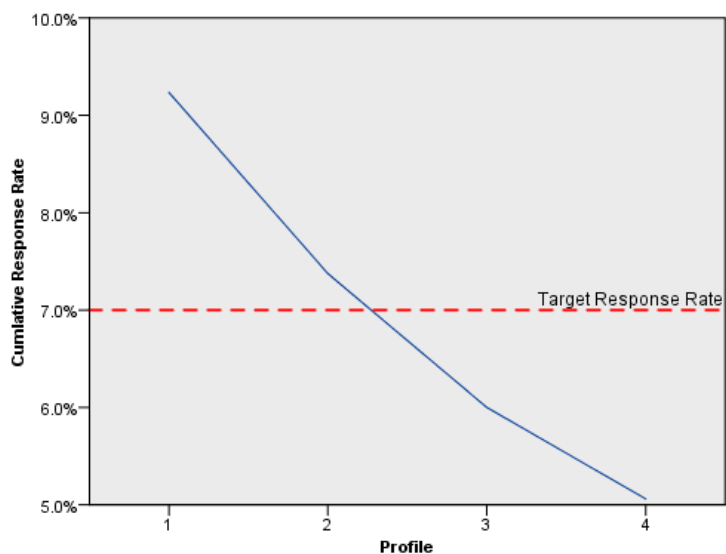
### ***Resultados***

El resultado incluye una tabla que proporciona una descripción de cada grupo de perfil y muestra índices de respuesta (porcentaje de respuestas positivas) e índices de respuesta acumulados y un gráfico de índices de respuesta acumulados. Si incluye un índice de respuesta de destino mínimo, la tabla tendrá una codificación de colores para mostrar los perfiles que cumplen el índice de respuesta acumulado mínimo y el gráfico incluye una línea de referencia en el valor de índice de respuesta mínimo especificado.

Figura 4-1  
Tabla y gráfico de índice de respuesta

Índice de respuesta				
Número	Descripción	Perfil		
		Tamaño de grupo	Índice de respuesta	Índice de respuesta acumulado
1	Región = "Oeste","Sur","Este" Género = "Mujer" Casado = "No"	379	9.2%	9.2%
2	Región = "Oeste","Sur","Este" Género = "Mujer" Casado = "Si"	299	5.0%	7.4%
3	Región = "Oeste","Sur","Este" Género = "Hombre"	722	4.7%	6.0%
4	Región = "Norte"	517	2.5%	5.1%

Verde: cumple el índice de respuesta de destino.  
Rojo: no cumple el índice de respuesta de destino.



### Consideraciones sobre los datos de perfiles de posibles clientes

**Campo respuesta.** El campo de respuesta puede ser nominal u ordinal. Puede ser numérico o de cadena. Si este campo contiene un valor que indica un número o cantidad de compras, necesitará crear un nuevo campo en el que un único valor represente todas las respuestas positivas. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica el p. 25.](#)

**Valor de respuesta positiva.** El valor de respuesta positiva identifica los clientes que han respondido de forma positiva (por ejemplo, los que han hecho una compra). Se asume que el resto de valores de respuesta no ausentes indican una respuesta negativa. Si hay alguna etiqueta de valor definida para el campo de respuesta, esas etiquetas se muestran en la lista desplegable.

**Crear perfiles con.** Estos campos pueden ser nominales, ordinales o continuos (escale). Pueden ser numéricos o de cadena.

**Nivel de medida.** La asignación del nivel correcto de medida es importante porque afecta al cálculo de los resultados.

- **Nominal.** Una variable se puede tratar como nominal si sus valores representan categorías que no obedecen a una ordenación intrínseca (por ejemplo, el departamento de la empresa en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- **Ordinal.** Una variable puede tratarse como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca (por ejemplo, los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- **Continuo.** Una variable puede tratarse como escala (continua) cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

Un icono junto a cada campo indica el nivel de medida actual.

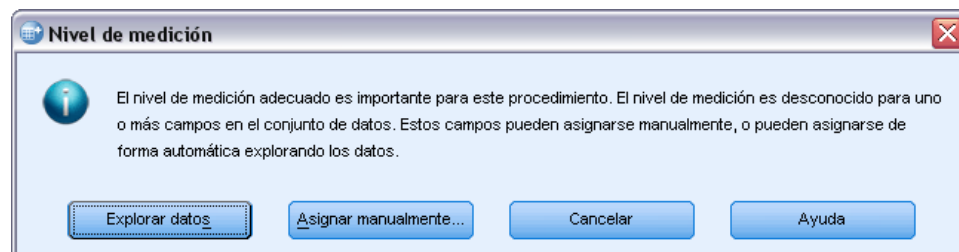
	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		
Ordinal				
Nominal				

Puede cambiar el nivel de medida en una vista de variables del Editor de datos o puede usar el cuadro de diálogo Definir propiedades de variables para sugerir un nivel de medición adecuado para cada campo .

### **Campos con un nivel de medición desconocido**

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 4-2  
Alerta de nivel de medición





- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

### **Para obtener perfiles de posibles clientes**

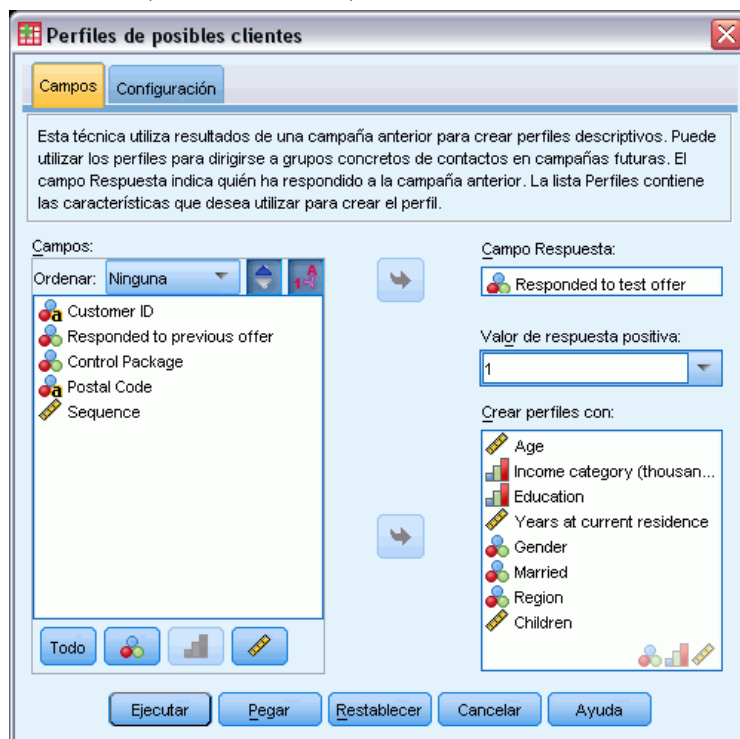
Seleccione en los menús:

Marketing directo > Seleccionar técnica

- Seleccione Generar perfiles de mis contactos que hayan respondido a una oferta.

Figura 4-3

*Pestaña Campos de Perfiles de posibles clientes*

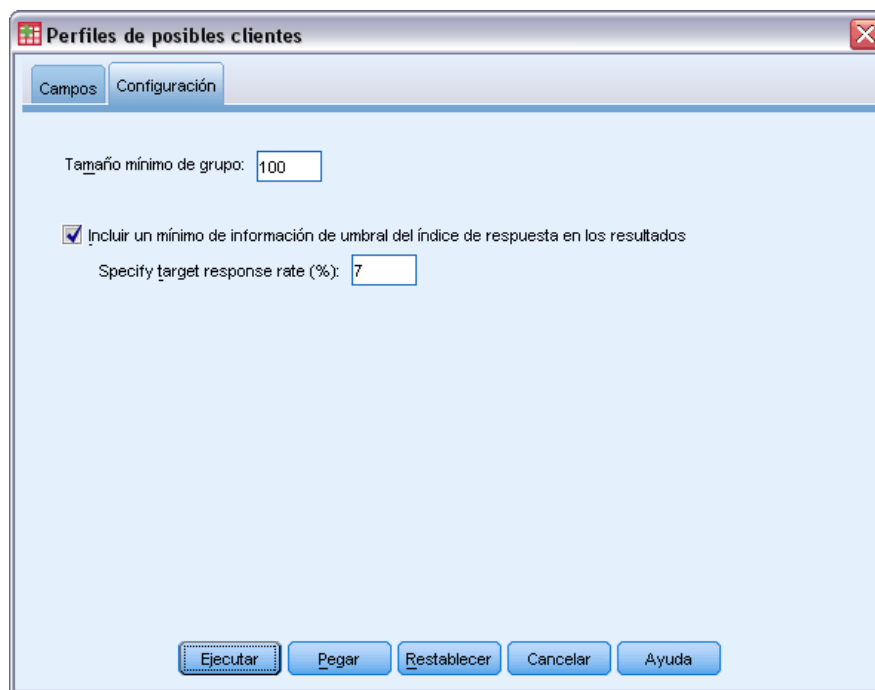


- Seleccione el campo que identifica los contactos que responderán a la oferta. El campo puede ser nominal u ordinal.

- ▶ Introduzca el valor que indica una respuesta positiva. Si algún valor ha definido etiquetas de valores, puede seleccionar la etiqueta de valor de la lista desplegable y se mostrará el valor correspondiente.
- ▶ Seleccione los campos que desea utilizar para crear los perfiles.
- ▶ Pulse en Ejecutar para ejecutar el procedimiento.

## Configuración

Figura 4-4  
Pestaña Configuración de Perfiles de posibles clientes



La pestaña Configuración permite controlar el tamaño del grupo de perfil mínimo e incluir un umbral mínimo de índice de respuesta en el resultado.

**Tamaño mínimo de grupo de perfil.** Cada perfil representa las características compartidas de un grupo de contactos en el conjunto de datos (por ejemplo, mujeres menores de 40 años que viven en la costa oeste). Por defecto, el menor tamaño de grupo de perfil es 100. Los tamaños de grupo menores pueden revelar más grupos, pero los tamaños de grupos más grandes proporcionan resultados más fiables. El valor debe ser un entero positivo.

**Incluir un mínimo de información de umbral del índice de respuesta en los resultados.** Los resultados incluyen una tabla que muestra índices de respuesta (porcentaje de respuestas positivas) e índices de respuesta acumulados y un gráfico de índices de respuesta acumulados. Si introduce un índice de respuesta de destino mínimo, la tabla tendrá una codificación de colores para mostrar los perfiles que cumplen el índice de respuesta acumulado mínimo y el gráfico incluye una línea de

referencia en el valor de índice de respuesta mínimo especificado. El valor debe ser mayor que 0 y menor que 100.

## ***Creación de un campo de respuesta categórica***

El campo de respuesta debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigna un valor de respuesta simple a todos los valores de respuesta distintos a cero.

- Si las respuestas negativas se registran como 0 (espacio cumplimentado, que se tratan como valores perdidos), se pueden calcular con la siguiente fórmula:

$$NewName=OldName>0$$

donde *NewName* es el nombre del nuevo campo y *OldName* es el nombre del campo original. Es una expresión lógica que asigna un valor de 1 a todos los valores no perdidos mayores que 0 y 0 a todos los valores no perdidos menores o iguales que 0.

- Si no se registra ningún valor para respuestas negativas, estos valores se consideran perdidos y la fórmula se complica un poco más:

$$NewName=NOT(MISSING(OldName))$$

En esta expresión lógica, a todos los valores de respuesta no perdidos se les asigna un valor de 1 y a todos los valores de respuesta no perdidos se les asigna un valor de 0.

- Si no puede distinguir entre valores de respuesta negativos (0) y valores perdidos, no se puede calcular un valor de respuesta preciso. Si hay relativamente pocos valores perdidos, es posible que no tenga un efecto significativo en los índices de respuesta calculados. Sin embargo, si hay demasiados valores perdidos, como cuando se registra información de respuesta para una pequeña prueba de muestra del conjunto de datos total, los índices de respuesta calculados no serán significativos, ya que son significativamente menores que los índices de respuesta verdaderos.

### ***Creación de un campo de respuesta categórica***

- ▶ Elija en los menús:  
Transformar > Calcular variable
- ▶ Para una variable de destino, introduzca el nuevo nombre de campo (variable).
- ▶ Si se registran respuestas negativas como 0, introduzca *OldName*>0 para Expresión numérica, donde *OldName* es el nombre del campo original.
- ▶ Si se registran respuestas negativas como perdidas (en blanco), introduzca NOT(MISSING(*OldName*)) para Expresión numérica, donde *OldName* es el nombre del campo original.

## *Índices de respuesta de código postal*

Esta técnica utiliza resultados de una campaña anterior para calcular índices de respuesta de código postal. Los índices de respuesta se pueden utilizar para dirigirse a códigos postales concretos en campañas futuras. El campo Respuesta indica quién ha respondido a la campaña anterior. El campo Respuesta indica quién ha respondido a la campaña anterior.

**Ejemplo.** Basándose en los resultados de un envío anterior, el departamento de marketing directo de una empresa genera índices de respuesta por códigos postales. Basándose en diferentes criterios como un índice de respuesta mínimo aceptable y/o el número máximo de contactos que incluirán en el envío, podrán dirigirse a códigos postales concretos.

### **Resultado**

El resultado de este procedimiento incluye un nuevo conjunto de datos que contiene índices de respuesta por código postal y una tabla y un gráfico que resumen los resultados por deciles (10% superior, 20% superior, etc.). La tabla puede tener una codificación de colores basada en un índice de respuesta acumulado mínimo de un usuario o un número máximo de contactos.

Figura 5-1  
Conjunto de datos con índices de respuesta por código postal

\*Sin título2 [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Aplicaciones Gráficos Utilidades Ventana Ayuda

1: CódigoPostal 932 Visible: 6 de 6 variables

	CódigoPostal	ResponseRate	Responses	Contacts	Index	Rank	
1	932	10.00%	4.00	40	3.60	1	
2	098	8.82%	6.00	68	5.47	1	
3	740	7.76%	9.00	116	8.30	1	
4	100	7.69%	7.00	91	6.46	1	
5	110	7.69%	5.00	65	4.62	1	
6	954	7.55%	4.00	53	3.70	1	
7	108	7.32%	6.00	82	5.56	1	
8	107	7.04%	5.00	71	4.65	1	
9	090	6.90%	4.00	58	3.72	1	
10	966	6.90%	4.00	58	3.72	1	
11	760	6.72%	8.00	119	7.46	1	
12	113	6.25%	5.00	80	4.69	1	
13	927	6.00%	3.00	50	2.82	1	
14	969	6.00%	3.00	50	2.82	1	
15	972	5.88%	3.00	51	2.82	2	
16	123	5.71%	2.00	35	1.89	2	
17	950	5.66%	3.00	53	2.83	2	
18	602	5.56%	6.00	108	5.67	2	
19	617	5.56%	5.00	90	4.72	2	
20	934	5.56%	3.00	54	2.83	2	
21	946	5.36%	3.00	56	2.84	2	

Vista de datos Vista de variables

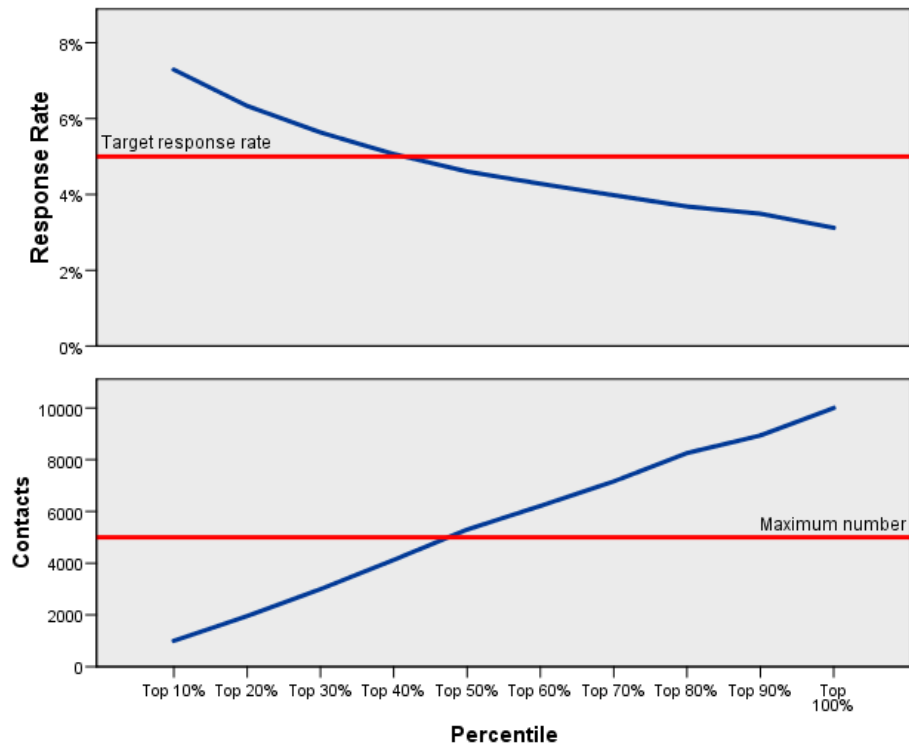
PASW Statistics Processor está listo

Figura 5-2  
Tabla de resumen y gráfico

**Índice de respuesta**

Percentile	Índice de respuesta	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

Green Red Caption



El nuevo conjunto de datos contiene los siguientes campos:

- **Código postal.** Si los grupos de códigos postales se basan sólo en una parte del valor completo, será el valor de esa parte del código postal. La etiqueta de la fila encabezado de esta columna en el archivo Excel es el nombre del campo de código postal en el conjunto de datos original.
- **Índice de respuesta.** El porcentaje de respuestas positivas en cada código postal.
- **Respuestas.** El número de respuestas positivas en cada código postal.

- **Contactos.** El número total de contactos en cada código postal que contiene un valor no perdido en el campo de respuesta.
- **Índice.** La respuesta “ponderada” basada en la fórmula  $N \times P \times (I-P)$ , donde  $N$  es el número de contactos y  $P$  es el índice de respuesta expresado como proporción.
- **Rango.** Rango de decil (10% superior, 20% superior, etc.) de los índices de respuesta por código postal en orden descendente.

### ***Consideraciones sobre los datos de índices de respuesta de código postal***

**Campo respuesta.** El campo de respuesta puede ser de cadena o numérico. Si este campo contiene un valor que indica un número o valor monetario de compras, necesitará crear un nuevo campo en el que un único valor represente todas las respuestas positivas. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica el p. 33.](#)

**Valor de respuesta positiva.** El valor de respuesta positiva identifica los clientes que han respondido de forma positiva (por ejemplo, los que han hecho una compra). Se asume que el resto de valores de respuesta no ausentes indican una respuesta negativa. Si hay alguna etiqueta de valor definida para el campo de respuesta, esas etiquetas se muestran en la lista desplegable.

**Campo Código postal.** El campo código postal puede ser de cadena o numérico.

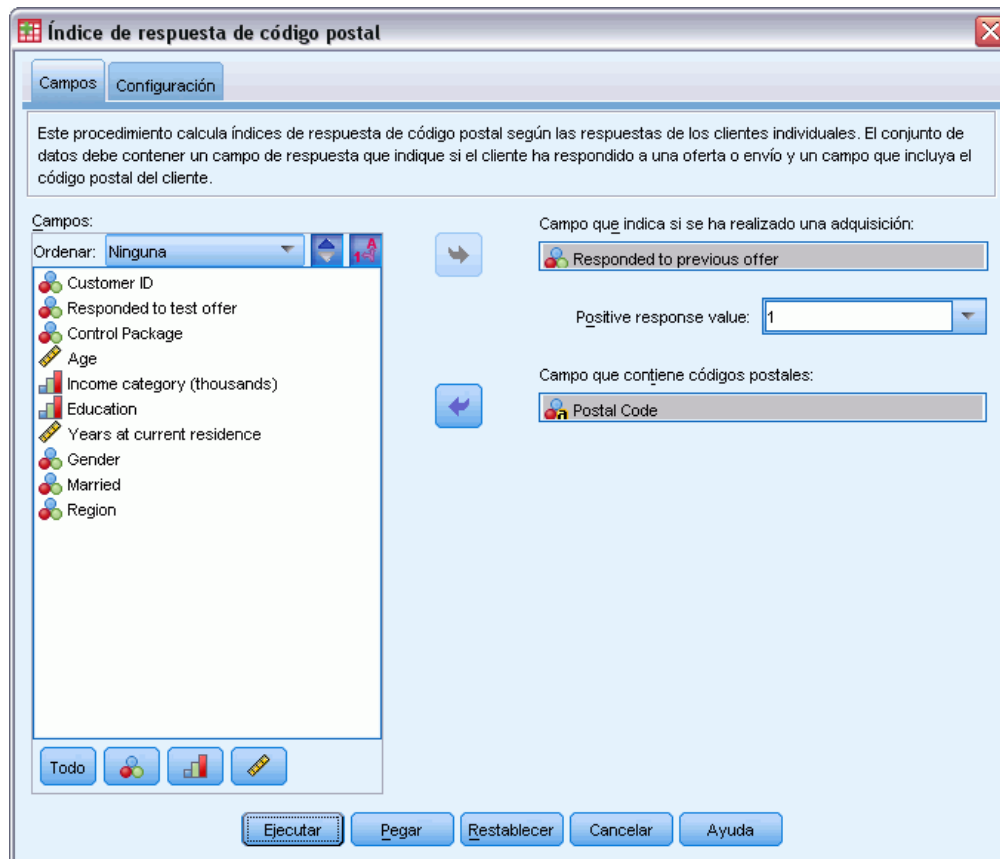
### ***Para obtener índices de respuesta de código postal***

Elija en los menús:

Marketing directo > Seleccionar técnica

- ▶ Seleccione Identificar principales códigos postales de respuesta.

Figura 5-3  
Pestaña Campos de Índices de respuesta de código postal



- ▶ Seleccione el campo que identifica los contactos que responderán a la oferta.
- ▶ Introduzca el valor que indica una respuesta positiva. Si algún valor ha definido etiquetas de valores, puede seleccionar la etiqueta de valor de la lista desplegable y se mostrará el valor correspondiente.
- ▶ Seleccione el campo que contiene el código postal.
- ▶ Pulse en Ejecutar para ejecutar el procedimiento.

Si lo desea, tiene la posibilidad de:

- Generar índices de respuesta basados en los primeros  $n$  caracteres o dígitos del código postal en lugar del valor completo
- Guardar automáticamente los resultados en un archivo Excel
- Controlar las opciones de visualización de los resultados



## Configuración

Figura 5-4  
Pestaña Configuración de Índices de respuesta de código postal

### **Agrupar códigos postales basados en**

Determina cómo se agrupan los registros para calcular los índices de respuesta. Por defecto, se utiliza el código postal completo y todos los registros con el mismo código postal se agrupan para calcular el índice de respuesta del grupo. De forma alternativa, puede agrupar los registros en función únicamente de una parte del código postal, utilizando los primeros  $n$  dígitos o caracteres. Por ejemplo, es posible que desee agrupar los registros por los 5 primeros caracteres de un código postal de 10 caracteres o por los 3 primeros dígitos de un código postal de 5 dígitos. El conjunto de datos de resultados contendrá un registro para cada grupo de código postal. Si introduce un valor, debe ser un entero positivo.

### **Formato numérico de código postal**

Si el campo de código postal es numérico y desea agrupar los códigos postales basados en los primeros  $n$  dígitos en lugar del valor completo, necesita especificar el número de dígitos en el valor original. El número de dígitos es el *máximo* número posible de dígitos en el código postal.

Por ejemplo, si el campo del código postal contiene una mezcla de códigos postales de 5 y 9 dígitos, deberá especificar 9 como el número de dígitos.

Nota: Dependiendo del formato de visualización, puede parecer que algunos códigos postales de 5 dígitos sólo contienen 4 dígitos, aunque se trata de un cero inicial.

### **Resultado**

Además del nuevo conjunto de datos que contiene índices de respuesta por código postal, puede mostrar una tabla y un gráfico que resumen los resultados por deciles (10% superior, 20% superior, etc.). La tabla muestra índices de respuesta, índices de respuesta acumulados, número de registros y el número acumulado de registros en cada decil. El gráfico muestra los índices de respuesta acumulados y número acumulado de registros de cada decil.

**Índice de respuesta mínimo aceptable.** Si introduce un índice de respuesta de destino mínimo o una fórmula de equilibrio, la tabla tendrá una codificación de colores para mostrar los deciles que cumplen el índice de respuesta acumulado mínimo y el gráfico incluye una línea de referencia en el valor de índice de respuesta mínimo especificado.

- **Índice de respuesta de destino.** Índice de respuesta expresado como un porcentaje (porcentaje de respuestas positivas en cada grupo de código postal). El valor debe ser mayor que 0 y menor que 100.
- **Calcular índice de equilibrio a partir de fórmula.** Calcula el índice de respuesta acumulado mínimo basado en la fórmula:  $(\text{Coste de envío de un paquete} / \text{Ingreso neto por respuesta}) \times 100$ . Ambos valores deben ser números enteros positivos. El resultado debe ser un valor mayor que 0 y menor que 100. Por ejemplo, si el coste de enviar un paquete es 0,75 dólares y el ingreso neto por respuesta es 56 dólares, el índice de respuesta mínimo es:  $(0,75/56) \times 100 = 1,34\%$ .

**Número máximo de contactos.** Si especifica un número máximo de contactos, la tabla tendrá una codificación de colores para mostrar los deciles que no superan el número máximo acumulado de contactos (registros) y el gráfico incluirá una línea de referencia en ese valor.

- **Porcentaje de contactos.** Máximo expresado como porcentaje. Por ejemplo, desea conocer los deciles con el mayor índice de respuesta que no contengan más del 50% de todos los contactos. El valor debe ser mayor que 0 y menor que 100.
- **Número de contactos.** Máximo expresado como número de contactos. Por ejemplo, si no va a enviar más de 10.000 paquetes, podrá definir el valor a 10000. El valor debe ser un entero positivo (sin símbolos de agrupamiento).

Si especifica un índice de respuesta mínimo aceptable y un número máximo de contactos, la codificación de colores de la tabla se basa en la condición que se cumpla primero.

**Exportar a Excel**

Este procedimiento crea automáticamente un nuevo conjunto de datos que contiene índices de respuesta por código postal. Cada registro (fila) del conjunto de datos representa un código postal. Puede guardar automáticamente la misma información en un archivo Excel. Este archivo se guarda en formato Excel 97-2003.

**Creación de un campo de respuesta categórica**

El campo de respuesta debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigna un valor de respuesta simple a todos los valores de respuesta distintos a cero.

- Si las respuestas negativas se registran como 0 (espacio cumplimentado, que se tratan como valores perdidos), se pueden calcular con la siguiente fórmula:

$$NewName=OldName>0$$

donde *NewName* es el nombre del nuevo campo y *OldName* es el nombre del campo original. Es una expresión lógica que asigna un valor de 1 a todos los valores no perdidos mayores que 0 y 0 a todos los valores no perdidos menores o iguales que 0.

- Si no se registra ningún valor para respuestas negativas, estos valores se consideran perdidos y la fórmula se complica un poco más:

$$NewName=NOT(MISSING(OldName))$$

En esta expresión lógica, a todos los valores de respuesta no perdidos se les asigna un valor de 1 y a todos los valores de respuesta no perdidos se les asigna un valor de 0.

- Si no puede distinguir entre valores de respuesta negativos (0) y valores perdidos, no se puede calcular un valor de respuesta preciso. Si hay relativamente pocos valores perdidos, es posible que no tenga un efecto significativo en los índices de respuesta calculados. Sin embargo, si hay demasiados valores perdidos, como cuando se registra información de respuesta para una pequeña prueba de muestra del conjunto de datos total, los índices de respuesta calculados no serán significativos, ya que son significativamente menores que los índices de respuesta verdaderos.

**Creación de un campo de respuesta categórica**

- ▶ Elija en los menús:  
Transformar > Calcular variable
- ▶ Para una variable de destino, introduzca el nuevo nombre de campo (variable).
- ▶ Si se registran respuestas negativas como 0, introduzca *OldName>0* para Expresión numérica, donde *OldName* es el nombre del campo original.
- ▶ Si se registran respuestas negativas como perdidas (en blanco), introduzca *NOT(MISSING(OldName))* para Expresión numérica, donde *OldName* es el nombre del campo original.

## ***Propensión a la compra***

La propensión a la compra utiliza resultados de un envío de prueba o campañas anteriores para generar puntuaciones. Los resultados indican los contactos que tienen más posibilidades de responder. El campo Respuesta indica quién ha respondido al correo de prueba o a la campaña anterior. Los campos de propensión son las características que quiere utilizar para pronosticar las probabilidades de que respondan contactos con características similares.

Esta técnica utiliza la regresión logística binaria para generar un modelo predictivo. El proceso de generar y aplicar un modelo predictivo tiene dos pasos básicos:

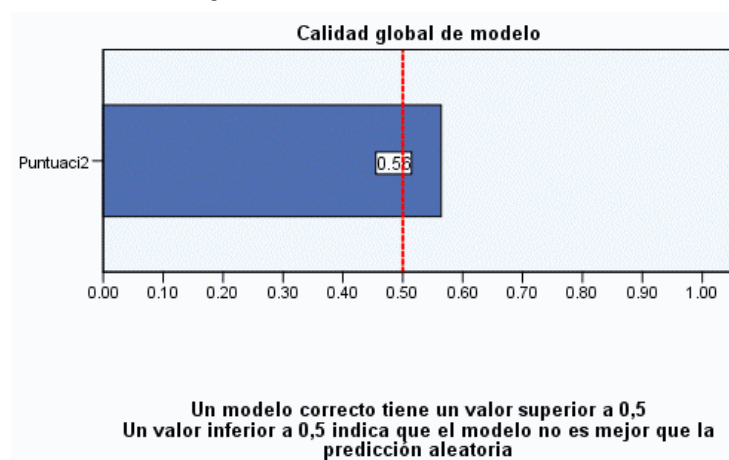
- ▶ Crear el modelo y guardar el archivo de modelo. Crea el modelo usando un conjunto de datos cuyo resultado deseado (a menudo conocido como **destino**) ya se conoce. Por ejemplo, si desea crear un modelo que prediga quién tiene más posibilidades de responder a una campaña de correo directo, deberá comenzar con un conjunto de datos que ya tenga información sobre quién respondió y quién no. Por ejemplo, esto puede tratarse de los resultados de un envío de prueba a un pequeño grupo de clientes o información sobre respuestas a una campaña similar en el pasado.
- ▶ Aplicar ese modelo a un conjunto de datos distinto (para el que no se conocen los resultados deseados) para obtener los resultados pronosticados.

**Ejemplo.** El departamento de marketing directo de una empresa utiliza los resultados de un envío de prueba para asignar puntuaciones de propensión al resto de su base de datos de contactos, utilizando diversas características demográficas para identificar los contactos con más posibilidades de responder y realizar una compra.

### ***Resultados***

Este procedimiento crea automáticamente un nuevo campo en el conjunto de datos que contienen puntuaciones de propensión para los datos de prueba y un archivo XML de modelo que se puede utilizar para puntuar otros conjuntos de datos. Los resultados estadísticos opcionales incluyen un gráfico de calidad global de modelo y una tabla de clasificación que compara las respuestas pronosticadas con respuestas actuales.

Figura 6-1  
Gráfico de calidad global de modelo



### Consideraciones de los datos de propensión a la compra

**Campo respuesta.** El campo de respuesta puede ser de cadena o numérico. Si este campo contiene un valor que indica un número o valor monetario de compras, necesitará crear un nuevo campo en el que un único valor represente todas las respuestas positivas. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica el p. 40.](#)

**Valor de respuesta positiva.** El valor de respuesta positiva identifica los clientes que han respondido de forma positiva (por ejemplo, los que han hecho una compra). Se asume que el resto de valores de respuesta no ausentes indican una respuesta negativa. Si hay alguna etiqueta de valor definida para el campo de respuesta, esas etiquetas se muestran en la lista desplegable.

**Pronosticar propensión con.** Los campos utilizados para pronosticar pueden ser numéricos o de cadena y pueden ser nominales, ordinales o continuos (escala), pero es importante asignar el nivel de medida adecuado a todos los campos predictores.

**Nivel de medida.** La asignación del nivel correcto de medida es importante porque afecta al cálculo de los resultados.

- **Nominal.** Una variable se puede tratar como nominal si sus valores representan categorías que no obedecen a una ordenación intrínseca (por ejemplo, el departamento de la empresa en el que trabaja un empleado). Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- **Ordinal.** Una variable puede tratarse como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca (por ejemplo, los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho). Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.
- **Continuo.** Una variable puede tratarse como escala (continua) cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

Un icono junto a cada campo indica el nivel de medida actual.

	Numérico	Cadena	Fecha	Hora
Escala (Continuo)		n/a		
Ordinal				
Nominal				

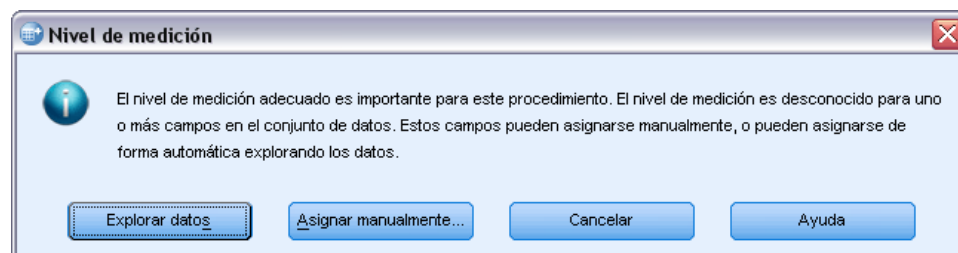
Puede cambiar el nivel de medida en una vista de variables del Editor de datos o puede usar el cuadro de diálogo Definir propiedades de variables para sugerir un nivel de medición adecuado para cada campo .

### ***Campos con un nivel de medición desconocido***

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 6-2

*Alerta de nivel de medición*



- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

### ***Para obtener puntuaciones de propensión a la compra***

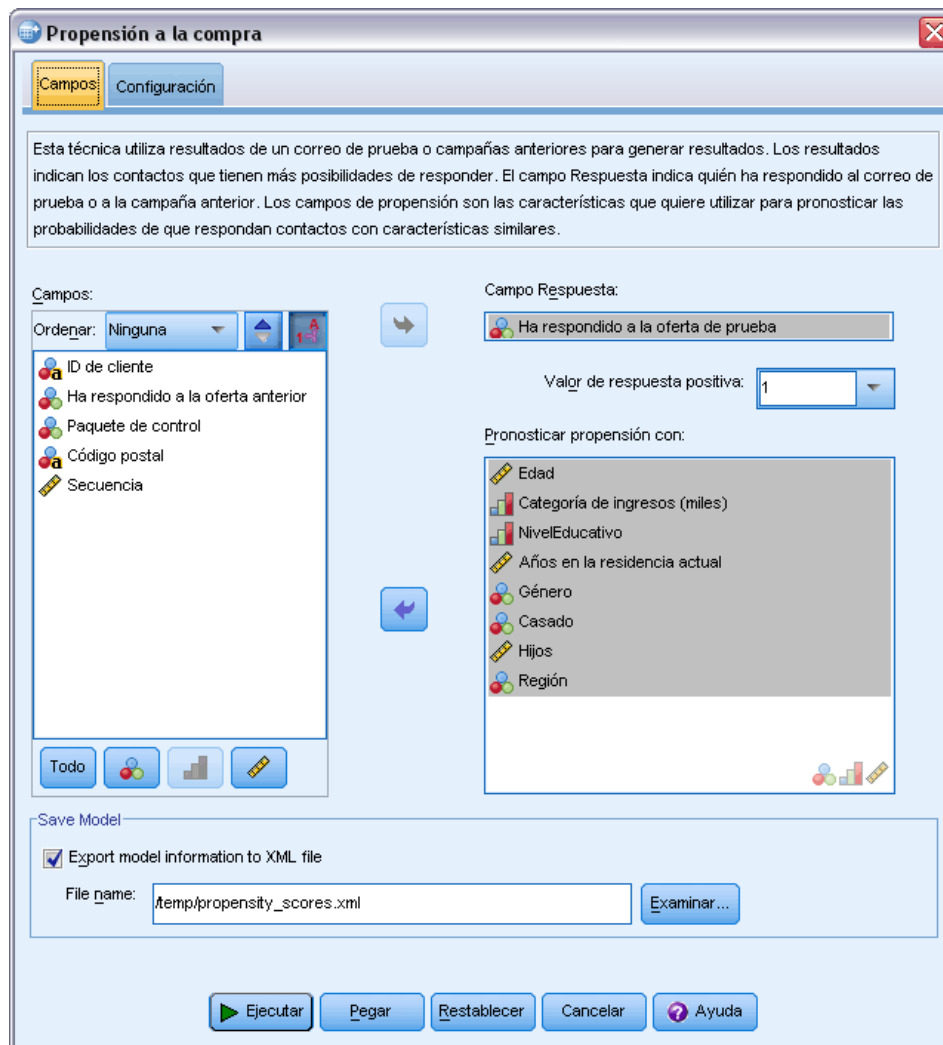
Seleccione en los menús:

Marketing directo > Seleccionar técnica

- Seleccione Seleccionar contactos con más probabilidades de realizar la compra.

Figura 6-3

Pestaña Campos de Propensión a la compra



- Seleccione el campo que identifica los contactos que responderán a la oferta.
- Introduzca el valor que indica una respuesta positiva. Si algún valor ha definido etiquetas de valores, puede seleccionar la etiqueta de valor de la lista desplegable y se mostrará el valor correspondiente.
- Seleccione los campos que desea utilizar para pronosticar la propensión.  
Para guardar un archivo XML modelo para puntuar otros archivos de datos:
- Seleccione (active) Exportar información del modelo a un archivo XML.
- Introduzca una ruta de directorio y el nombre del archivo o pulse en Examinar para desplazarse hasta la ubicación en la que desea guardar el archivo XML de modelo.

- ▶ Pulse en Ejecutar para ejecutar el procedimiento.  
Para utilizar el archivo modelo para puntuar otros conjuntos de datos:
- ▶ Abra el conjunto de datos que desee puntuar.
- ▶ Utilice el Asistente para puntuación para aplicar el modelo al conjunto de datos. Seleccione en los menús:  
Utilidades > Asistente para puntuación.

## Configuración

Figura 6-4  
Pestaña Configuración de Propensión a la compra

**Propensión a la compra**

**Campos Configuración**

**Validación de modelos**

Puede validar el modelo utilizado para generar resultados. Para validar el modelo, necesita dividir los datos en particiones. La partición de entrenamiento se utiliza para entrenar o crear el modelo. La partición de prueba se utiliza para validar el modelo. Si desea validar el modelo, esta técnica asignará automáticamente los registros a las particiones.

Valide el modelo

Tamaño de la partición de la muestra de entrenamiento (%):

Establecer semilla para replicar resultados

Semilla:

**Resultados diagnósticos**

Calidad global de modelo

Tabla de clasificación

Probabilidad mínima:

**Nombre y etiqueta del campo de respuesta recodificado**

Esta técnica automáticamente recodifica el campo de respuesta como un nuevo campo en el que 1 representa respuestas positivas y 0 representa respuestas negativas.

Nuevo nombre de campo:

Nueva etiqueta de campo:

**Guardar puntuaciones**

Esta técnica utiliza resultados de un correo de prueba o campañas anteriores para generar resultados. Las puntuaciones se guardan automáticamente para que pueda utilizarlas. El resto de controles de esta ficha proporcionan control adicional del contenido que se guarda.

Nuevo nombre de campo para puntuaciones:

**Ejecutar** **Pegar** **Restablecer** **Cancelar** **Ayuda**



### **Validación de modelos**

Validación de modelos crea grupos de entrenamiento y prueba para tareas de diagnóstico. Si selecciona la tabla de clasificación en la sección Resultados diagnósticos, la tabla se dividirá en entrenamiento (seleccionada) y prueba (sin seleccionar) para su comparación. No seleccione la validación de modelos salvo que también seleccione la tabla de clasificación. Los resultados se basan en el modelo generado a partir de la muestra de entrenamiento, que siempre contendrá menos resultados que el número total de registros disponibles. Por ejemplo, el tamaño por defecto de la muestra de entrenamiento es 50%, y es posible que un modelo creado con sólo la mitad de los registros disponibles no sea tan fiable como el modelo basado en todos los registros disponibles.

- **Tamaño de partición de muestra de entrenamiento (%).** Especifique el porcentaje de registros que se asignarán a la muestra de entrenamiento. El resto de registros con valores no perdidos para el campo de respuesta se asignan a la muestra de entrenamiento. El valor debe ser mayor que 0 y menor que 100.
- **Establecer semilla para replicar resultados.** Como los registros se asignan de forma aleatoria a las muestras de entrenamiento y prueba, cada vez que ejecute el procedimiento podrá obtener resultados diferentes, salvo que especifique siempre el mismo valor de semilla de número aleatorio de inicio.

### **Resultados diagnósticos**

**Calidad global de modelo.** Muestra un gráfico de la calidad global de modelo, expresada como un valor entre 0 y 1. Un buen modelo debe tener un valor superior a 0,5.

**Tabla de clasificación.** Muestra una tabla que compara las respuestas positivas y negativas pronosticadas con las respuestas positivas y negativas reales. El porcentaje de precisión global puede proporcionar algún tipo de indicación del funcionamiento del modelo, pero es posible que esté más interesado en el porcentaje de respuestas positivas correctas pronosticadas.

- **Probabilidad mínima.** Asigna registros con una puntuación mayor que el valor especificado en la categoría de respuesta positiva pronosticada en la tabla de clasificación. Las puntuaciones generadas por el procedimiento representan la probabilidad de que el contacto responderá de forma positiva (por ejemplo, realizar una compra). Como norma general, debe especificar un valor cercano a su tasa de respuesta de objetivo mínima, expresado como una proporción. Por ejemplo, si está interesado en una tasa de respuesta mínima del 5%, especifique 0,05. El valor debe ser mayor que 0 y menor que 1.

### **Nombre y etiqueta del campo de respuesta recodificado**

Este procedimiento recodifica automáticamente el campo de respuesta como un nuevo campo en el que 1 representa respuestas positivas y 0 representa respuestas negativas; y el análisis se realiza en el campo recodificado. Puede sustituir el nombre por defecto y asignar uno de su elección. Los nombres deben cumplir las reglas de nombres de IBM® SPSS® Statistics.

### **Guardar puntuaciones**

Un nuevo campo con puntuaciones de propensión se guarda automáticamente en el conjunto de datos original. Las puntuaciones representan la probabilidad de una respuesta positiva, expresada como una proporción.

- Los nombres de los campos deben cumplir las reglas de nombres de SPSS Statistics.
- El campo no puede tener un nombre que ya exista en el conjunto de datos. Si ejecuta este procedimiento más de una vez en el conjunto de datos, necesitará especificar un nombre diferente cada vez.

## **Creación de un campo de respuesta categórica**

El campo de respuesta debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigna un valor de respuesta simple a todos los valores de respuesta distintos a cero.

- Si las respuestas negativas se registran como 0 (espacio cumplimentado, que se tratan como valores perdidos), se pueden calcular con la siguiente fórmula:

$$NewName = OldName > 0$$

donde *NewName* es el nombre del nuevo campo y *OldName* es el nombre del campo original. Es una expresión lógica que asigna un valor de 1 a todos los valores no perdidos mayores que 0 y 0 a todos los valores no perdidos menores o iguales que 0.

- Si no se registra ningún valor para respuestas negativas, estos valores se consideran perdidos y la fórmula se complica un poco más:

$$NewName = NOT(MISSING(OldName))$$

En esta expresión lógica, a todos los valores de respuesta no perdidos se les asigna un valor de 1 y a todos los valores de respuesta no perdidos se les asigna un valor de 0.

- Si no puede distinguir entre valores de respuesta negativos (0) y valores perdidos, no se puede calcular un valor de respuesta preciso. Si hay relativamente pocos valores perdidos, es posible que no tenga un efecto significativo en los índices de respuesta calculados. Sin embargo, si hay demasiados valores perdidos, como cuando se registra información de respuesta para una pequeña prueba de muestra del conjunto de datos total, los índices de respuesta calculados no serán significativos, ya que son significativamente menores que los índices de respuesta verdaderos.

### **Creación de un campo de respuesta categórica**

- ▶ Elija en los menús:  
Transformar > Calcular variable
- ▶ Para una variable de destino, introduzca el nuevo nombre de campo (variable).
- ▶ Si se registran respuestas negativas como 0, introduzca *OldName*>0 para Expresión numérica, donde *OldName* es el nombre del campo original.

- ▶ Si se registran respuestas negativas como perdidas (en blanco), introduzca `NOT(MISSING(OldName))` para Expresión numérica, donde *OldName* es el nombre del campo original.

## Prueba de paquete de control

Esta técnica compara campañas de marketing para ver si hay diferencias significativas en la eficacia entre los distintos paquetes u ofertas. La eficacia de una campaña se mide por las respuestas. El campo Campaña identifica diferentes campañas, por ejemplo, Oferta A y Oferta B. El campo Respuesta indica si un contacto ha respondido a la campaña. Seleccione Cantidad de adquisición si la respuesta se registra como una cantidad de adquisición, por ejemplo “99,99”. Seleccione Respuesta si la respuesta indica simplemente si el contacto ha respondido positivamente o no, por ejemplo “Sí” o “No”.

**Ejemplo.** La división de marketing directo de una empresa quiere ver si un nuevo diseño de paquete generará respuestas más positivas que el paquete actual. Envían un correo de prueba para determinar si el nuevo paquete genera un porcentaje de respuesta significativamente más positivo. El correo de prueba consiste en un grupo de control que recibe el paquete actual y un grupo de prueba que recibe el nuevo diseño del paquete. Los resultados de los dos grupos se comparan para ver si hay una diferencia significativa.

### Resultado

Los resultados incluyen una tabla que muestra los recuentos y porcentajes de respuestas positivas y negativas de cada grupo definidos por el campo Campaña y una tabla que identifica los grupos que pueden ser muy diferentes entre sí.

Figura 7-1  
Resultado de prueba de paquete de control

		Paquete de control			
		Control		Prueba	
		Recuento	% del N de la columna	Recuento	% del N de la columna
Eficacia (1=S? 0=NO)	0	879	96,6%	984	97,7%
	1	31	3,4%	23	2,3%

No hay una diferencia significativa estadísticamente entre Control y Prueba.

### Consideraciones y supuestos de datos de prueba de paquete de control

**Campo Campaña.** El campo Campaña debe ser categórico (nominal u ordinal).

**Campo Respuesta de eficacia.** Si selecciona Cantidad de adquisición para el campo Eficacia, el campo debe ser numérico y el nivel de medida debe ser continuo (escala).

Si no puede distinguir entre valores de respuesta negativos (para la cantidad de adquisición, un valor de 0) y valores perdidos, no se puede calcular un valor de respuesta preciso. Si hay relativamente pocos valores perdidos, es posible que no tenga un efecto significativo en los

índices de respuesta calculados. Sin embargo, si hay demasiados valores perdidos, como cuando se registra información de respuesta para una pequeña prueba de muestra del conjunto de datos total, los índices de respuesta calculados no serán significativos, ya que son significativamente menores que los índices de respuesta verdaderos.

**Supuestos.** Este procedimiento supone que los contactos se han asignado aleatoriamente a cada grupo de campaña. En otras palabras, ningún tipo de información demográfica, historial de adquisiciones u otras características afectará a la asignación de grupos y todos los contactos tienen una probabilidad igual de ser asignados a cualquier grupo.

### Para obtener una prueba de paquete de control

Elija en los menús:

Marketing directo > Seleccionar técnica

- Seleccione Comparar la eficacia de las campañas.

Figura 7-2

Cuadro de diálogo Prueba de paquete de control

Prueba de paquete de control

Esta técnica compara campañas de marketing para ver si hay diferencias significativas entre los distintos paquetes u ofertas. El campo Campaña identifica las distintas campañas. El campo Eficacia indica si un cliente ha respondido a la campaña.

Campos:

Ordenar: Ninguna

- Customer ID
- Responded to previous offer
- Postal Code
- Age
- Income category (thousands)
- Education
- Years at current residence
- Gender
- Married
- Region

Campana:

Control Package

Eficacia:

Responded to test offer

Medida de la eficacia

Cantidad de adquisición

Respuesta positiva

Valor: 1

Name and Label for Recoded Effectiveness Response Field

Cuando la eficacia es una respuesta positiva, esta técnica crea automáticamente un campo de eficacia sí/no para su análisis.

Nuevo nombre de campo: Effectiveness\_new1

Nueva etiqueta de campo: Campo de prueba grabado (1=SÍ 0=N)

Ejecutar Pegar Restablecer Cancelar Ayuda

- ▶ Seleccione el campo que identifica el grupo de campaña al que pertenece cada contacto (por ejemplo, Oferta A y Oferta B, etc.). Este campo debe ser nominal u ordinal.
- ▶ Seleccione el campo que indica la respuesta de eficacia.

Si el campo de respuesta es una cantidad de adquisición, el campo debe ser numérico.

Si el campo de respuesta indica si el contacto han respondido positivamente o no (por ejemplo, “Sí” o “No”), seleccione Respuesta e introduzca el valor que representa una respuesta positiva. Si algún valor ha definido etiquetas de valores, puede seleccionar la etiqueta de valor de la lista desplegable y se mostrará el valor correspondiente.

Se crea automáticamente un nuevo campo, en el que 1 representa las respuestas positivas y 0 representa las respuestas negativas y el análisis se ejecuta en el nuevo campo. Puede sustituir el nombre por defecto y asignar uno de su elección. Los nombres deben cumplir las reglas de nombres de IBM® SPSS® Statistics.

- ▶ Pulse en Ejecutar para ejecutar el procedimiento.

## ***Parte II: Ejemplos***

# Análisis de RFM a partir de datos de transacciones

En un archivo de datos de transacciones, cada fila representa una transacción diferente, en lugar de un cliente diferente y cada cliente dispone de múltiples filas de transacciones. En este ejemplo se utiliza el archivo de datos *rfm\_transactions.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

## Datos de transacciones

El conjunto de datos debe contener variables con la siguiente información:

- Una variable o combinación de variables que identifiquen a cada caso (cliente).
- Una variable con la fecha de cada transacción.
- Una variable con el valor monetario de cada transacción.

Figura 8-1  
Datos de transacciones de RFM

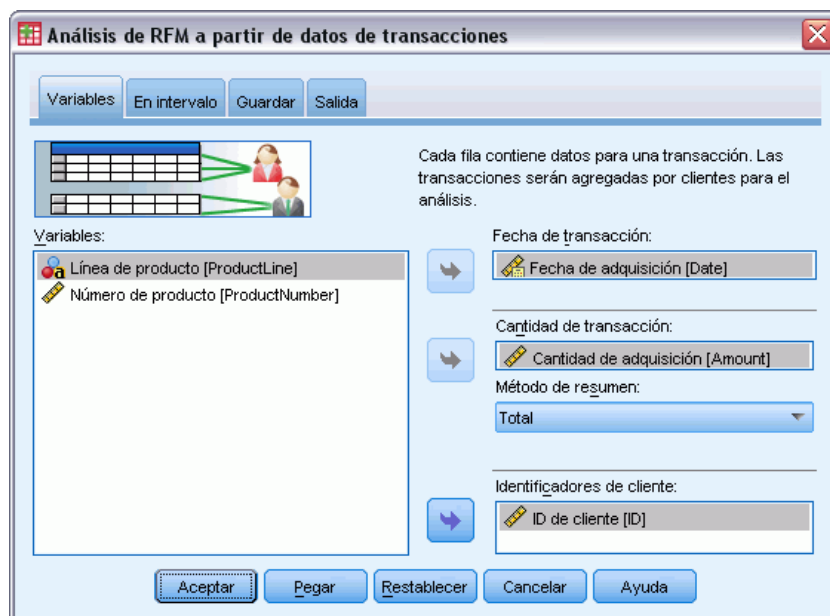
ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

## Ejecución del análisis

- ▶ Para calcular las puntuaciones de propensión, seleccione en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Ayudarme a identificar los mejores contactos (Análisis RFM) y pulse en Continuar.
- ▶ Pulse Datos de transacciones y, a continuación, pulse Continuar.

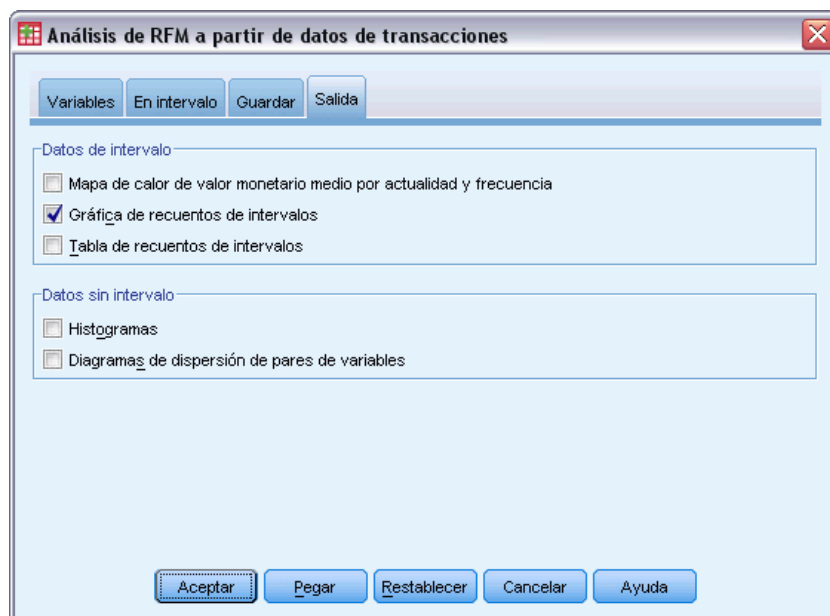


Figura 8-2  
RFM desde datos de transacciones, pestaña Variables



- ▶ Pulse en Restablecer para borrar los ajustes anteriores.
- ▶ En Fecha de transacción, seleccione *Fecha de compra [Fecha]*.
- ▶ En Importe de transacción, seleccione *Importe de compra [Importe]*.
- ▶ En Método de resumen, seleccione Total.
- ▶ En Identificadores de cliente, seleccione *ID cliente [ID]*.
- ▶ A continuación, pulse en la pestaña Resultados.

Figura 8-3  
RFM desde datos de transacciones, pestaña Resultados



- ▶ Seleccione Gráfico de recuentos de intervalos.
- ▶ Pulse en Aceptar para ejecutar el procedimiento.

## Evaluación de los resultados

Cuando calcule las puntuaciones de RFM a partir de los datos de transacciones, se creará un nuevo conjunto de datos que incluya las nuevas puntuaciones de RFM.

Figura 8-4  
Conjunto de datos RFM desde datos de transacciones

ID	Fecha_más_reciente	Recuento_transacciones	Cantidad	Puntuación_actualidad	Puntuación_frecuencia	Puntuación_monetaria	Puntuación_RFM
1	04-Sep-2006	5	485,00	4	3	4	434
2	10-Nov-2005	4	350,00	2	2	2	222
3	04-Jun-2005	2	233,00	1	2	4	124
4	18-Aug-2006	7	936,00	4	4	5	445
5	07-Jul-2006	3	359,00	4	1	5	415
6	16-Jul-2006	3	249,00	4	1	4	414
7	15-Feb-2006	7	1089,00	2	5	5	255

Por defecto, el conjunto de datos incluye la siguiente información para cada cliente:

- Variable(s) de ID de cliente
- Fecha de transacción más reciente
- Número total de transacciones

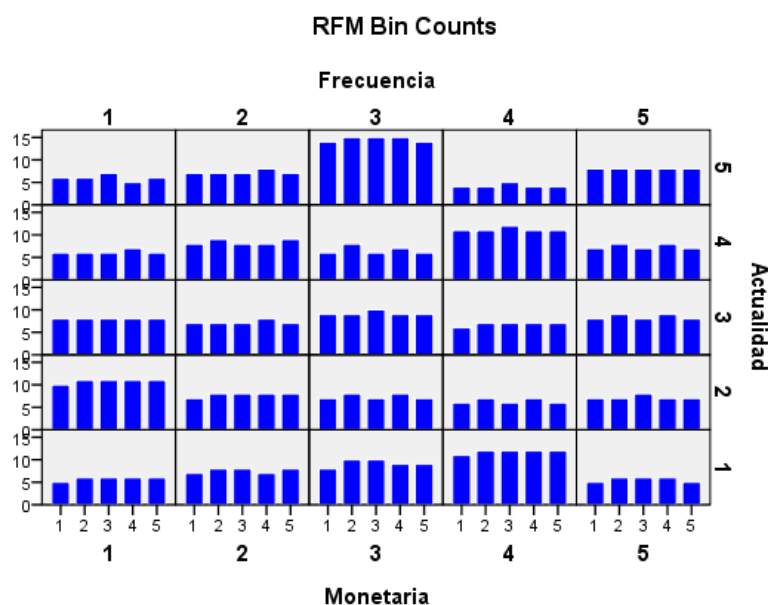
- Resumen de importe de transacción (el valor por defecto es total)
- Actualidad, frecuencia, valor monetario y puntuaciones de RFM combinadas

El nuevo conjunto de datos sólo contiene una fila (registro) para cada cliente. Los datos de transacciones originales se han agregado mediante los valores de las variables de identificación de los clientes. Las variables de identificación siempre se incluyen en el nuevo conjunto de datos, de lo contrario no habría forma de asignar las puntuaciones de RFM con los clientes.

La puntuación de RFM combinada de cada cliente es simplemente la concatenación de tres puntuaciones individuales, calculadas como:  $(\text{actualidad} \times 100) + (\text{frecuencia} \times 10) + \text{monetaria}$ .

El gráfico de recuentos de intervalos que se muestra en la ventana Visor muestra el número de clientes de cada categoría RFM.

Figura 8-5  
Gráfica de recuentos de intervalos



Si se utiliza el método por defecto de cinco categorías de puntuación para cada uno de los tres componentes RFM se obtiene como resultado 125 categorías de puntuación RFM. Cada barra del gráfico representa el número de clientes en cada categoría RFM.

Lo ideal es una distribución relativamente homogénea de clientes en todas las categorías de puntuación RFM. En la realidad, habrá cierto grado de variación, como el que puede ver en este ejemplo. Si hay demasiadas categorías vacías, puede considerar un cambio en el método de agrupación.

Hay una serie de estrategias para solucionar las distribuciones desiguales de las puntuaciones de RFM, incluyendo:

- Uso de agrupación anidada en lugar de anidación independiente.
- Reducción del número de categorías de puntuación posibles (intervalos).
- Si hay grandes números de valores empatados, asignación aleatoria de casos con las mismas puntuaciones a categorías diferentes.

Si desea obtener más información, consulte el tema Agrupación de RFM en el capítulo 2 el p. 6.

## ***Fusión de datos de puntuación con datos de clientes***

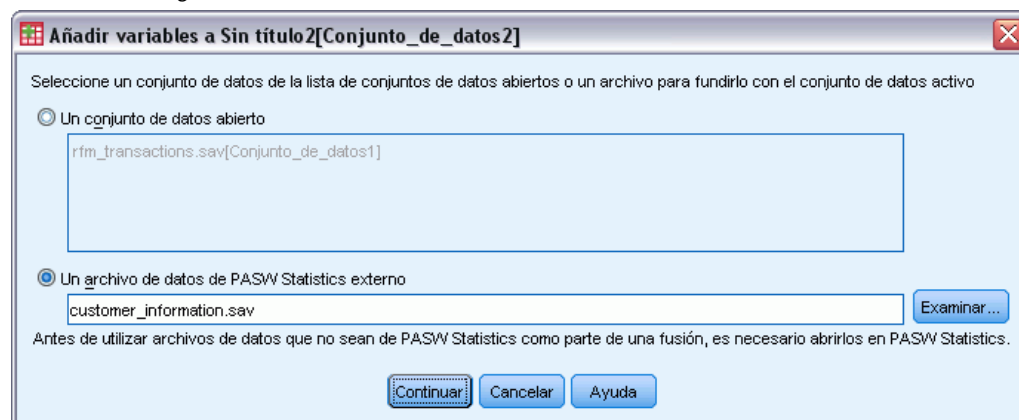
Ahora que tiene un conjunto de datos con puntuaciones RFM, deberá igualar estas puntuaciones con los clientes. Puede volver a fusionar las puntuaciones con el archivo de datos de transacción, pero lo más normal es que desee fusionar los datos de puntuaciones con un archivo de datos que, al igual que el conjunto de datos de puntuaciones RFM, contenga una fila (registro) para cada cliente, y contenga también información como el nombre y dirección del cliente.

**Figura 8-6**  
Conjunto de datos de puntuación RFM en la Vista de variables

Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores
ID	Numérico	5	0	Customer ID	Ninguna
Fecha_más_reciente	Fecha	10	0	Fecha de trans...	Ninguna
Recuento_transacciones	Numérico	7	0	Número de tran...	Ninguna
Cantidad	Numérico	8	2	Cantidad	Ninguna
Puntuación_actualidad	Numérico	3	0	Puntuación de ...	Ninguna
Puntuación_frecuencia	Numérico	3	0	Puntuación de f...	Ninguna
Puntuación_monetaria	Numérico	3	0	Puntuación mo...	Ninguna
Puntuación_RFM	Numérico	3	0	Puntuación de ...	Ninguna

- ▶ Defina que el conjunto de datos que contiene las puntuaciones RFM sea el conjunto de datos activo. (Pulse en cualquier punto de la ventana Editor de datos que contenga el conjunto de datos.)
- ▶ Seleccione en los menús:  
Datos > Fundir archivos > Añadir variables

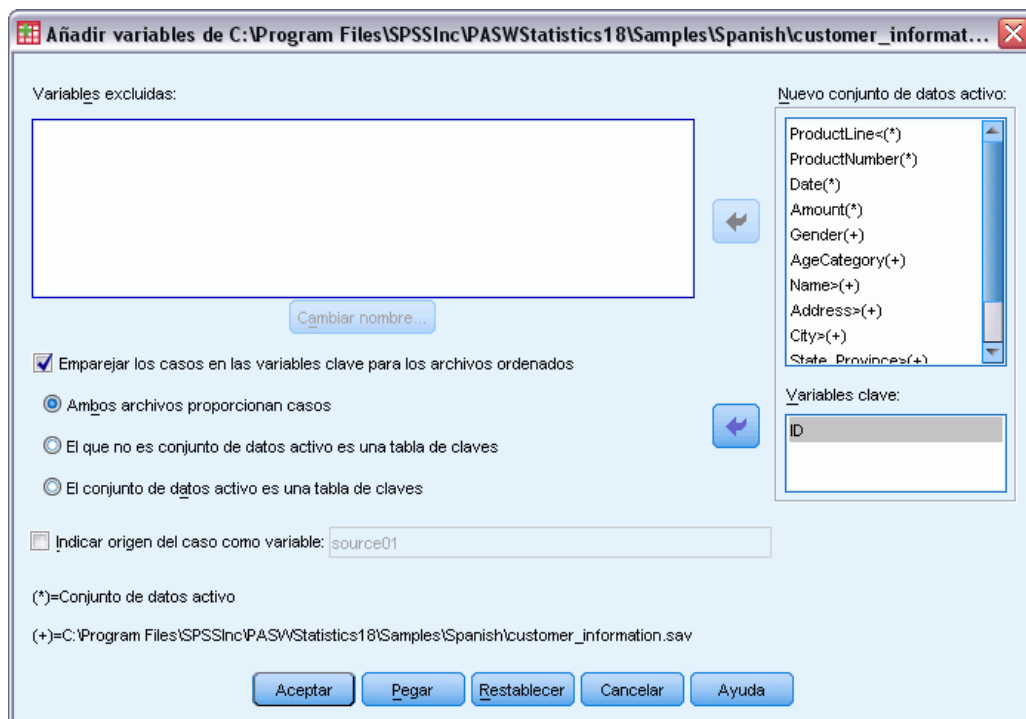
**Figura 8-7**  
Cuadro de diálogo Añadir variables, seleccionar archivos



- ▶ Seleccione Un archivo de datos externo.

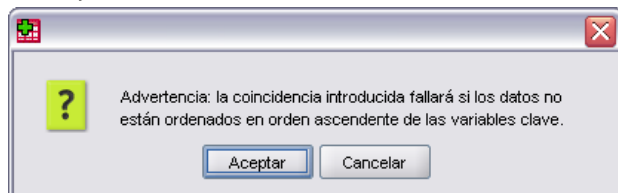
- ▶ Utilice el botón Examinar para desplazarse por la carpeta *Samples* y seleccionar *customer\_information.sav*. Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.
- ▶ Pulse en Continuar.

Figura 8-8  
Cuadro de diálogo Añadir variables, seleccionar variables



- ▶ Seleccione Emparejar los casos en las variables clave para los archivos clasificados.
- ▶ Seleccione Ambos archivos facilitan casos (Añadir variables).
- ▶ Seleccione *ID* en la lista Variables clave.
- ▶ Pulse en Aceptar.

Figura 8-9  
Mensaje de advertencia de Añadir variables



Observe el mensaje que le advierte de que ambos archivos deben clasificarse con las variables clave en orden ascendente. En este ejemplo, ambos archivos ya tienen la variable clave en orden ascendente, que es la variable de identificador de cliente seleccionada al calcular las puntuaciones RFM. Cuando calcule las puntuaciones RFM a partir de los datos de transacciones, el nuevo

conjunto de datos se clasificará automáticamente con las variables de identificación de clientes en orden ascendente. Si cambia el orden de clasificación del conjunto de datos de puntuaciones o del archivo de datos con el que desea fundir el conjunto de datos de puntuaciones no está clasificado en ese orden, en primer lugar deberá clasificar ambos archivos con las variable(s) de identificación de clientes en orden ascendente.

- Pulse en Aceptar para fundir los dos conjuntos de datos.

El conjunto de datos que contiene las puntuaciones RFM también contiene el nombre, dirección e información para cada cliente.

**Figura 8-10**  
*Conjuntos de datos fundidos*

ID	Numérico	5	0	Customer ID	Ninguna
Fecha_más_reciente	Fecha	10	0	Fecha de transacción...	Ninguna
Recuento_transacciones	Numérico	7	0	Número de transacci...	Ninguna
Cantidad	Numérico	8	2	Cantidad	Ninguna
Puntuación_actualidad	Numérico	3	0	Puntuación de actuali...	Ninguna
Puntuación_frecuencia	Numérico	3	0	Puntuación de frecue...	Ninguna
Puntuación_monetaria	Numérico	3	0	Puntuación monetaria	Ninguna
Puntuación_RFM	Numérico	3	0	Puntuación de RFM	Ninguna
Gender	Numérico	4	0		{0, Female}...
AgeCategory	Numérico	7	0	Age Category	{1, <25}...
Name	Cadena	4	0		Ninguna
Address	Cadena	14	0		Ninguna
City	Cadena	11	0		Ninguna
State_Province	Cadena	7	0	State/Province	Ninguna
PostalCode	Cadena	1	0	Postal Code	Ninguna
Country	Cadena	1	0		Ninguna

# Análisis de conglomerados

El análisis de conglomerados es una herramienta de exploración diseñada para revelar agrupaciones (o conglomerados) naturales dentro de sus datos. Por ejemplo, puede identificar diferentes grupos de clientes basándose en diversas características demográficas y de adquisición.

Por ejemplo, la división de marketing directo de una empresa desea identificar grupos demográficos de su base de datos de clientes para poder determinar estrategias de campañas de marketing y desarrollar nuevas ofertas de productos.

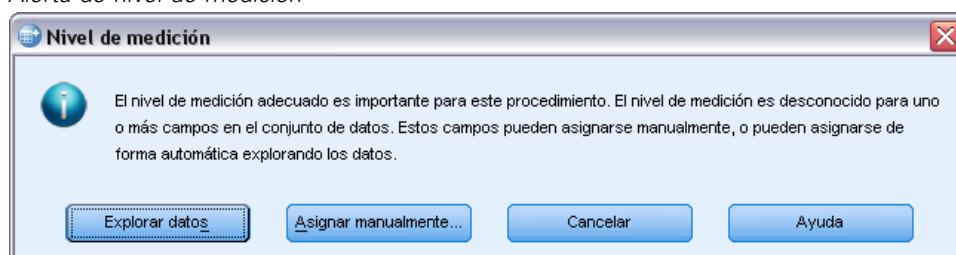
Esta información se recoge en el archivo *dmdata.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

## Ejecución del análisis

- ▶ Para ejecutar un análisis de conglomerados, elija en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Dividir mis contactos en conglomerados y pulse Continuar.

La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 9-1  
Alerta de nivel de medición

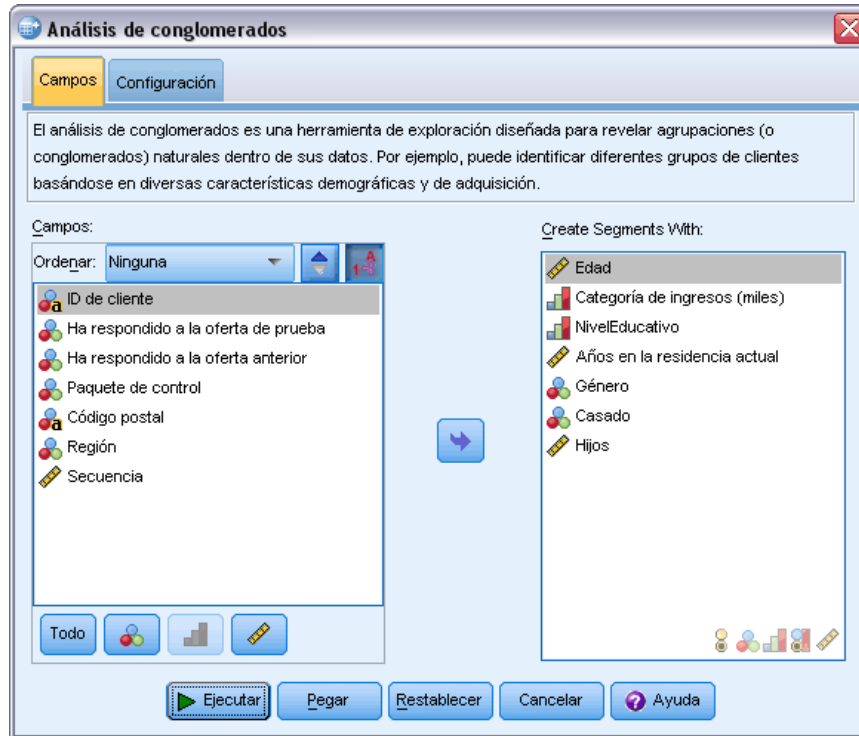


- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

En este archivo de ejemplo, no hay campos con un nivel de medición desconocido y todos los campos tienen el nivel de medición correcto; por lo que no aparecerá la alerta del nivel de medición.

Figura 9-2  
Pestaña Campos de Análisis de conglomerados

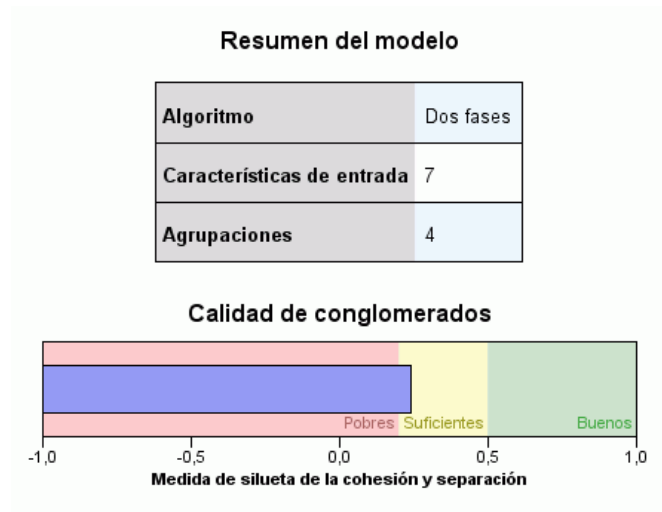


- ▶ Seleccione los siguientes campos para crear segmentos: *Edad*, *Categoría de ingresos*, *Nivel educativo*, *Años en la residencia actual*, *Género*, *Casado* e *Hijos*.
- ▶ Pulse en Ejecutar para ejecutar el procedimiento.



## Resultados

Figura 9-3  
Resumen del modelo de conglomerados



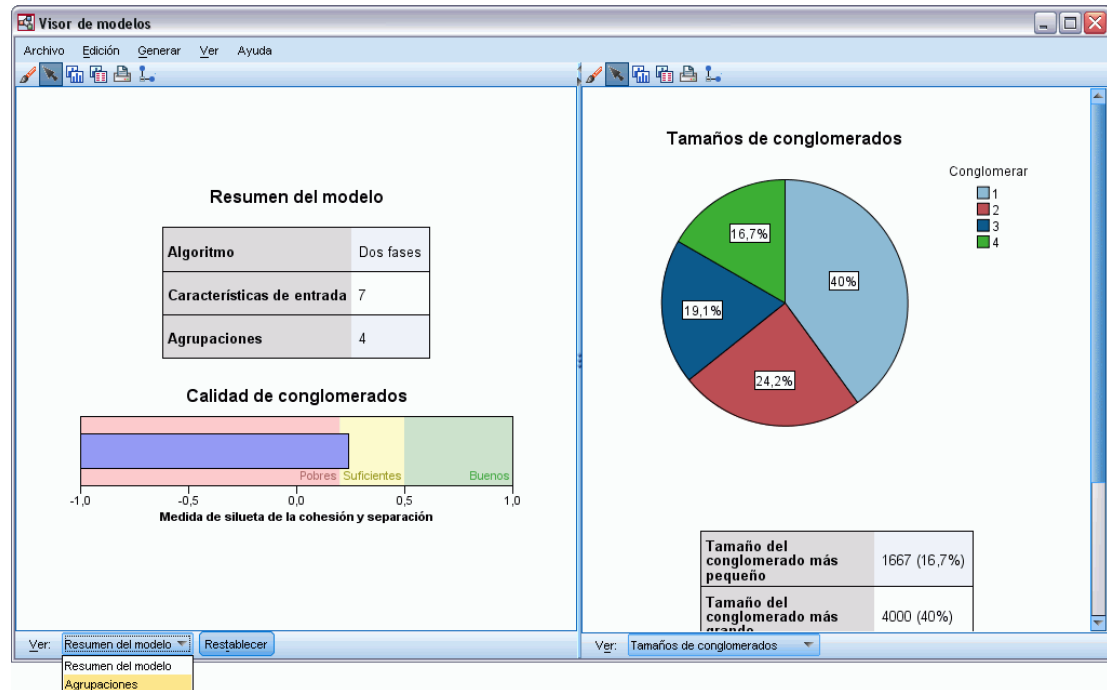
Los resultados se muestran en el Visor de modelos de conglomerados.

- El resumen de modelos indica que se han encontrado cuatro conglomerados en función de las siete características de entrada (campos) que ha seleccionado.
- El gráfico de calidad de conglomerados indica que la calidad general del modelo está a mitad del rango “Correcto”.

- Pulse dos veces el resultado del Visor de modelos de conglomerados para activar el Visor de modelos.

Figura 9-4

Visor de modelos de conglomerados activado



- En la lista desplegable Vista de la parte inferior de la ventana Visor de modelos de conglomerados, seleccione Conglomerados.

Figura 9-5  
vista conglomerado

Conglomerar	1	2	3	4
Etiqueta				
Descripción				
Tamaño	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
Funciones	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

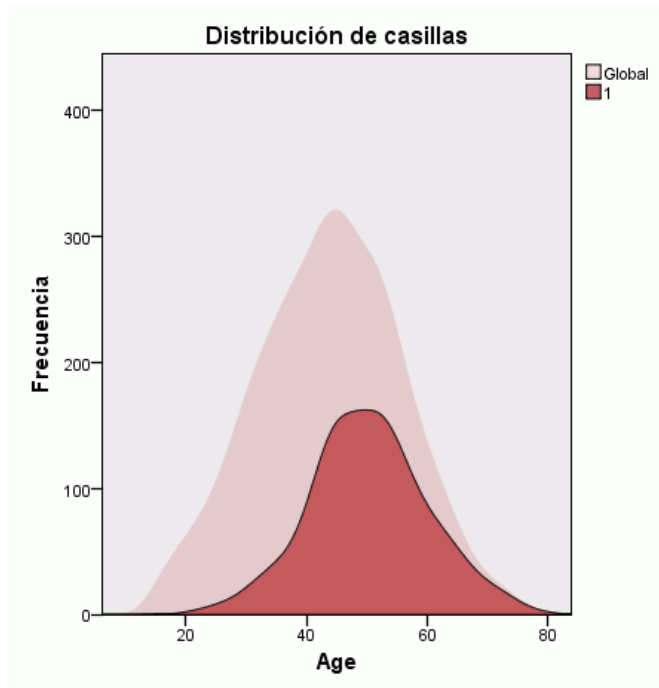
La vista Conglomerado muestra información sobre los atributos de cada conglomerado.

- En los campos continuos (escala), se muestra el valor de media.
- En los campos categóricos (nominales, ordinales), se muestra el modo. La moda es la categoría que contiene el mayor número de registros. En este ejemplo, cada registro es un cliente.
- Por defecto, los campos se muestran en el orden de su importancia general para el modelo. En este ejemplo, *Edad* tiene la mayor importancia general. También puede ordenar los campos por importancia dentro del conglomerado u orden alfabético.

Si selecciona (pulsar) cualquier casilla de la vista Conglomerado, podrá ver un gráfico que resume los valores de ese campo para ese conglomerado.

- Por ejemplo, seleccione la casilla *Edad* para el conglomerado 1.

Figura 9-6  
*Histograma Edad para el conglomerado 1*

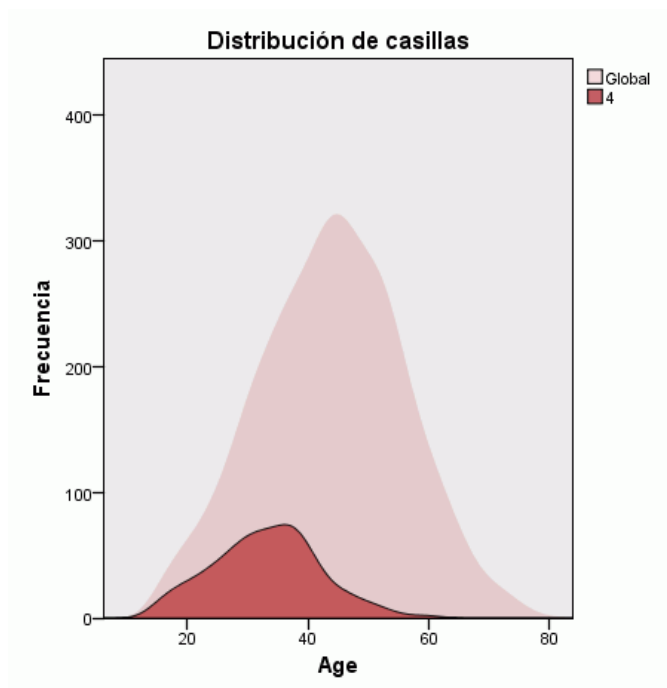


En los campos continuos se muestra un histograma. El histograma muestra tanto la distribución de valores para ese conglomerado como la distribución general de valores para el campo. El histograma indica que los clientes del conglomerado 1 tienden a ser mayores.

- Seleccione la casilla *Edad* para el conglomerado 4 en la vista Conglomerado.

Figura 9-7

*Histograma Edad para el conglomerado 4*

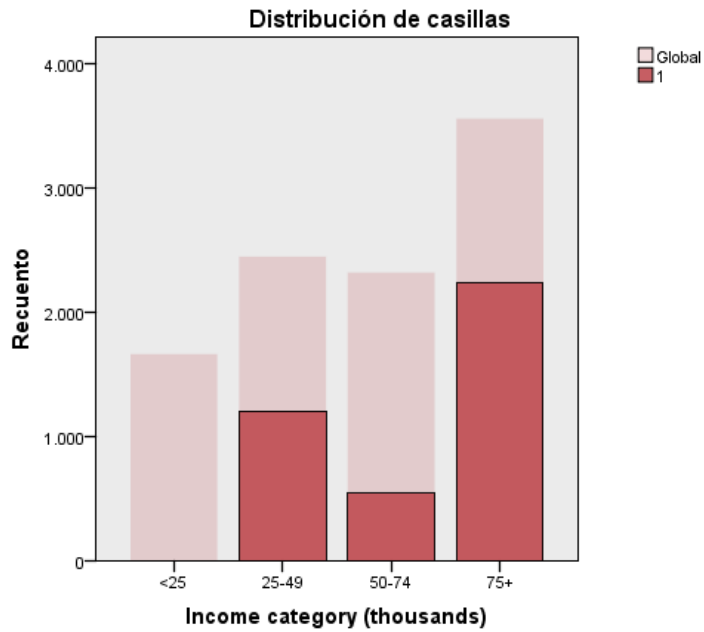


En oposición al conglomerado 1, los clientes del conglomerado 4 suelen ser más jóvenes que la media general.

- Seleccione la casilla *Categoría de ingresos* para el conglomerado 1 en la vista Conglomerado.

Figura 9-8

Gráfico de barras de categoría de ingresos para el conglomerado 1

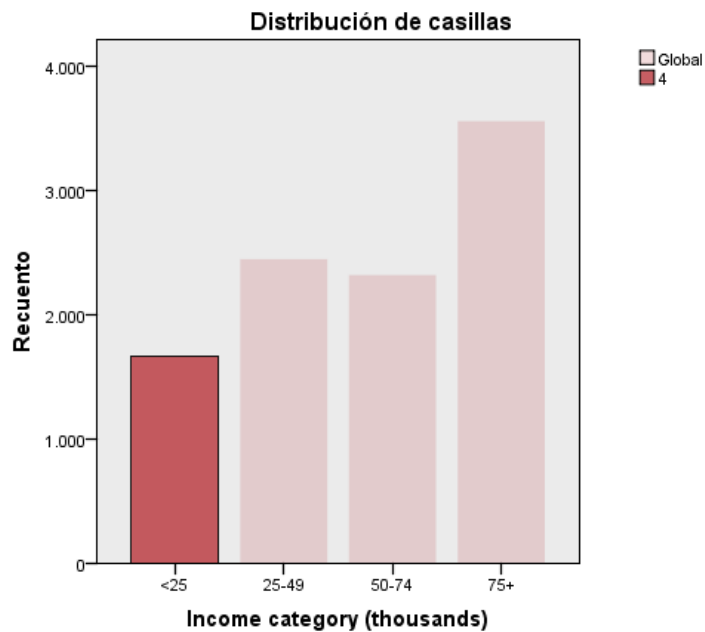


En los campos categóricos se muestra un gráfico de barras. La característica más notable del gráfico de barras de la categoría de ingresos de este conglomerado es la ausencia total de clientes en la categoría de ingresos más baja.

- Seleccione la casilla *Categoría de ingresos* para el conglomerado 4 en la vista Conglomerado.

Figura 9-9

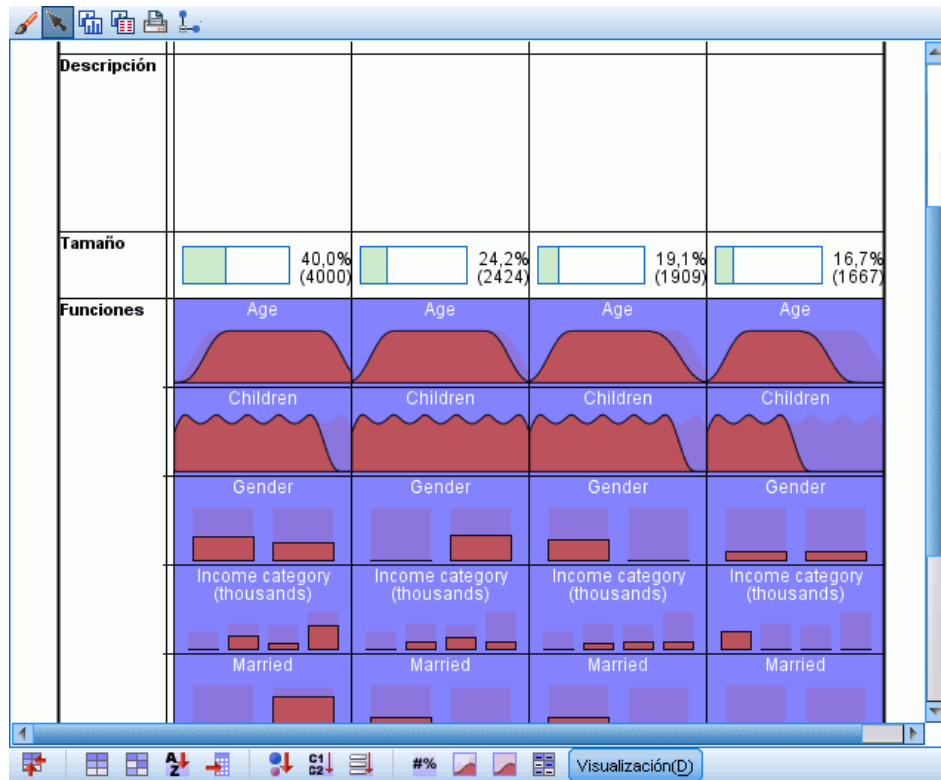
Gráfico de barras de categoría de ingresos para el conglomerado 4



En oposición al conglomerado 1, todos los clientes del conglomerado 4 se encuentran en la categoría de ingresos más baja.

También puede cambiar la vista Conglomerado para mostrar gráficos en las casillas, lo que facilita comparar rápidamente las distribuciones de los valores entre los conglomerados, usando la barra de herramientas de la parte inferior de la ventana Visor de modelos para cambiar la vista.

Figura 9-10  
Gráficos mostrados en el conglomerado



Si mira la vista Conglomerado y la información adicional provista en los gráficos de cada casilla, podrá ver algunas diferencias claras entre los conglomerados:





- Los clientes del conglomerado 1 suelen ser personas de mayor edad, casadas, con hijos y con mayores ingresos.
- Los clientes del conglomerado 2 suelen ser mujeres de mayor edad, solteras, con hijos y con ingresos moderados.
- Los clientes del conglomerado 3 suelen ser hombres más jóvenes, solteros y sin hijos.
- Los clientes del conglomerado 4 suelen ser mujeres más jóvenes, solteras, sin hijos y con ingresos más bajos.



Las casillas Descripción de la vista Conglomerado son campos de texto que puede editar para añadir descripciones de cada conglomerado.

Figura 9-11

Vista Conglomerado con descripciones de conglomerados

Conglomerar	1	2	3	4
<b>Etiqueta</b>				
<b>Descripción</b>	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
<b>Tamaño</b>	 40,0% (4000)	 24,2% (2424)	 19,1% (1909)	 16,7% (1667)
<b>Funciones</b>	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

## Selección de registros en función de los conglomerados

Hay dos formas de seleccionar registros en función de su pertenencia a conglomerados:

- Cree una condición de filtro interactivamente en el Visor de modelos de conglomerados.
- Utilice los valores del campo de conglomerado generado por el procedimiento para especificar el filtro o las condiciones de selección.

## Creación de un filtro en el Visor de modelos de conglomerados

Para crear una condición de filtro que seleccione registros de conglomerados específicos en el visor de modelos de conglomerados:

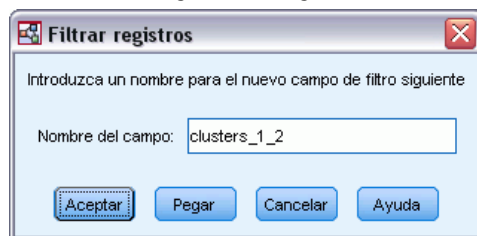
- ▶ Active (pulse dos veces en) el Visor de modelos de conglomerados.
- ▶ En la lista desplegable Vista de la parte inferior de la ventana Visor de modelos de conglomerados, seleccione Conglomerados.
- ▶ Pulse el número del conglomerado que desea que aparezca en la parte superior de la vista Conglomerado. Si desea seleccionar múltiples conglomerados, presione Ctrl mientras pulsa cada número de conglomerado que desea.

Figura 9-12  
Conglomerados seleccionados en la vista Conglomerado

Conglomerar	1	2	3	4
<b>Etiqueta</b>				
<b>Descripción</b>	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
<b>Tamaño</b>	40,0% (4000)	24,2% (2424)	19,1% (1909)	16,7% (1667)
<b>Funciones</b>	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

- ▶ En los menús del Visor de modelos de conglomerados, seleccione:  
Generar > Filtrar registros

Figura 9-13  
Cuadro de diálogo Filtrar registros



- Introduzca un nombre para el campo de filtro y pulse en Aceptar. Los nombres deben cumplir las reglas de nombres de IBM® SPSS® Statistics.

Figura 9-14  
Registros filtrados en el Editor de datos

	ID	Married	Children	Region	ClusterGroup1	clusters_1_2
<del>14</del>	03623	No	0	West	3	.00
<del>15</del>	01353	No	0	West	3	.00
<del>16</del>	07055	No	0	West	3	.00
17	04455	No	0	West	2	1.00
18	07210	No	1	West	2	1.00
<del>19</del>	08054	No	0	West	4	.00
<del>20</del>	06937	No	0	West	4	.00
<del>21</del>	06512	No	0	West	4	.00
<del>22</del>	08315	No	0	West	4	.00
23	09676	No	3	West	2	1.00
<del>24</del>	09636	No	0	West	4	.00
25	08579	No	1	West	2	1.00
26	01480	No	1	West	2	1.00

Crea un nuevo campo en los registros de conjuntos de datos y filtros de la base de datos en función de los valores de ese campo.

- Los registros que tengan el valor 1 en el campo de filtro se incluirán en los siguientes análisis, gráficos e informes.
- Los registros que tengan el valor 0 en el campo de filtro se excluirán.
- Los registros excluidos no se eliminan del conjunto de datos. Se conservan con un indicador de estado de filtro, que aparece como una barra transversal sobre el número de registro del Editor de datos.

### **Selección de registros en función de los valores de campo de conglomerado**

Por defecto, el análisis de conglomerado crea un nuevo campo que identifica el grupo de conglomerados de cada registro. El nombre predeterminado de este campo es *GrupoConglomeradosn*, donde *n* es un número entero que forma un nombre de campo único.

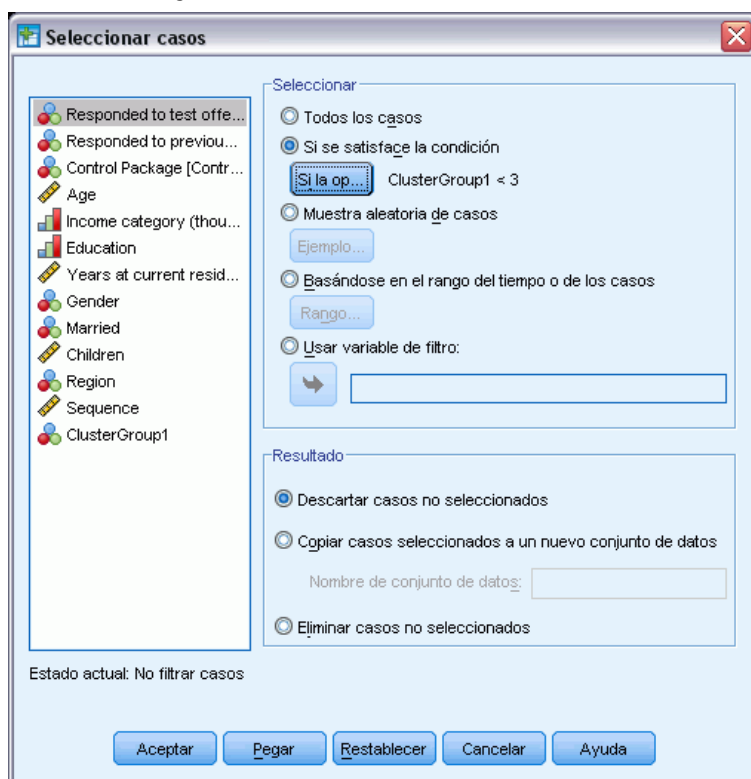
Figura 9-15  
Campo Conglomerado añadido al conjunto de datos

	ID	Gender	Married	Children	Region	ClusterGroup1
1	01359	Female	No	0	West	4
2	06262	Female	No	1	West	2
3	08031	Male	No	0	West	3
4	01971	Male	No	0	West	4
5	09689	Male	No	0	West	3
6	06108	Male	No	1	West	3
7	09853	Male	No	0	West	3
8	06802	Male	No	0	West	4
9	07597	Male	No	0	West	3
10	03692	Male	No	1	West	3
11	00071	Male	No	0	West	4
12	00769	Male	No	0	West	3

Para utilizar los valores del campo de conglomerado para seleccionar registros en conglomerados específicos:

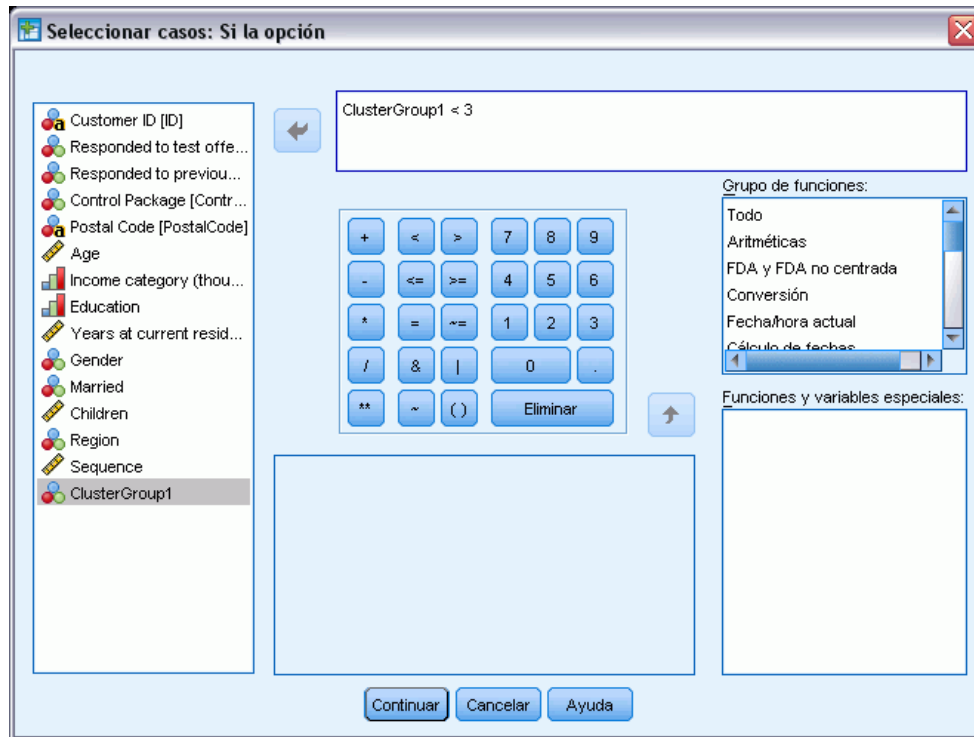
- Seleccione en los menús:  
Datos > Seleccionar casos

Figura 9-16  
Cuadro de diálogo Seleccionar casos



- En el cuadro de diálogo Seleccionar casos, seleccione Si se satisface la condición y pulse en Si.

Figura 9-17  
 Seleccionar casos: Cuadro de diálogo Si



- Introduzca la condición de selección.

Por ejemplo, `GrupoConglomerados1 < 3` seleccionará todos los registros en los conglomerados 1 y 2, y ejecutará registros en el conglomerado 3 y los superiores.

- Pulse en Continuar.

En el cuadro de diálogo Seleccionar casos hay varias opciones sobre qué hacer con los registros seleccionados y los no seleccionados:

**Descartar casos no seleccionados.** Crea un nuevo campo que especifica una condición de filtro. Los registros excluidos no se eliminan del conjunto de datos. Se conservan con un indicador de estado de filtro, que aparece como una barra transversal sobre el número de registro del Editor de datos. Esto equivale a seleccionar interactivamente conglomerados en el Visor de modelos de conglomerados.

**Copiar casos seleccionados a un nuevo conjunto de datos.** Esto crea un nuevo conjunto de datos en la sesión actual que contiene sólo los registros que cumplen las condiciones de filtro. El conjunto de datos original no se ve afectado.

**Eliminar casos no seleccionados.** Los registros no seleccionados se eliminarán del conjunto de datos. Sólo se pueden recuperar los registros eliminados saliendo del archivo sin guardar ningún cambio y abriéndolo de nuevo. La eliminación de los casos será permanente si se guardan los cambios en el archivo de datos.

El cuadro de diálogo Seleccionar casos también tiene una opción para utilizar una variable existente como una variable de filtro (campo). Si crea una condición de filtro interactivamente en el Visor de modelos de conglomerados y guarda el campo de filtro generado con el conjunto de datos podrá utilizar ese campo para filtrar registros en las siguientes sesiones.

## ***Resumen***

El análisis de conglomerados es una útil herramienta de exploración que permite revelar agrupaciones (o conglomerados) naturales dentro de sus datos. Puede utilizar la información de estos conglomerados para determinar estrategias de campañas de marketing y desarrollar nuevas ofertas de productos. Puede seleccionar registros en función de su pertenencia a conglomerados para analizarlos más detalladamente o usarlos en campañas de marketing dirigidas.

## ***Perfiles de posibles clientes***

Perfiles de posibles clientes utiliza resultados de una prueba o campaña anterior para crear perfiles descriptivos. Puede utilizar los perfiles para dirigirse a grupos concretos de contactos en campañas futuras. Por ejemplo, basándose en los resultados de un correo de prueba, el departamento de marketing directo de una empresa desea generar perfiles de los tipos de personas con más posibilidades de responder un cierto tipo de oferta, en función de la información demográfica. Basándose en esos resultados, pueden determinar los tipos de listas de correo que utilizarían para ofertas similares.

Por ejemplo, la división de marketing directo de una empresa envía un correo de prueba a aproximadamente el 20% del total de su base de datos de clientes. Los resultados de este correo de prueba se registran en un archivo de datos que también contiene características demográficas de cada cliente, incluyendo la edad, género, estado civil y región geográfica. Los resultados se registran en una forma simple de sí/no, indicando los clientes del correo de prueba que ha respondido (han realizado una compra) y los que no han respondido.

Esta información se recoge en el archivo *dmdata.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

### ***Consideraciones de los datos***

El campo de respuesta debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigna un valor de respuesta simple a todos los valores de respuesta distintos a cero. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica en el capítulo 4 el p. 25.](#)

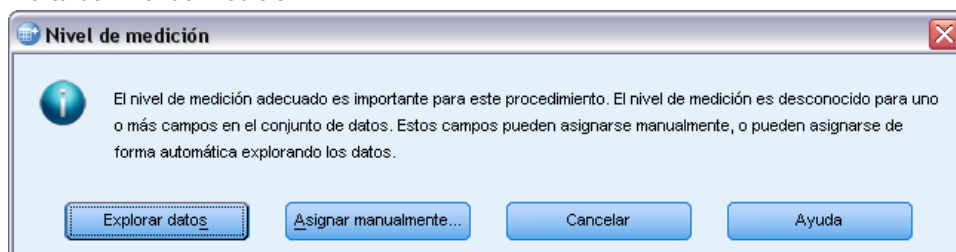
### ***Ejecución del análisis***

- ▶ Para ejecutar un análisis de Perfiles de posibles clientes, elija en los menús:  
Marketing directo > Seleccionar técnica
  
- ▶ Seleccione Generar perfiles de mis contactos que hayan respondido a una oferta y pulse Continuar.



La alerta de nivel de medición se muestra si el nivel de medición de una o más variables (campos) del conjunto de datos es desconocido. Como el nivel de medición afecta al cálculo de los resultados de este procedimiento, todas las variables deben tener un nivel de medición definido.

Figura 10-1  
Alerta de nivel de medición



- **Explorar datos.** Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.
- **Asignar manualmente.** Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

Como el nivel de medición es importante para este procedimiento, no puede acceder al cuadro de diálogo para ejecutar este procedimiento hasta que se hayan definido todos los campos en el nivel de medición.

En este archivo de ejemplo, no hay campos con un nivel de medición desconocido y todos los campos tienen el nivel de medición correcto; por lo que no aparecerá la alerta del nivel de medición.

Figura 10-2  
Pestaña Campos de Perfiles de posibles clientes



- ▶ En Campo Respuesta, seleccione *Ha respondido a la oferta de prueba*.
- ▶ Para un valor de respuesta positiva, seleccione *Sí* en la lista desplegable. Un valor de 1 se muestra en el campo de texto porque “Sí” es realmente una etiqueta de valor asociado con un valor registrado de 1. (Si el valor de respuesta positiva no tiene un valor de etiqueta de valor definida, puede introducir el valor en el campo texto.)
- ▶ En Crear perfiles con, seleccione *Edad*, *Categoría de ingresos*, *Nivel educativo*, *Años en la residencia actual*, *Género*, *Casado*, *Región* e *Hijos*.
- ▶ Pulse en la ficha Configuración.

Figura 10-3  
Pestaña Configuración de Perfiles de posibles clientes

- ▶ Seleccione (marque) Incluir un mínimo de información de umbral del índice de respuesta en los resultados.
- ▶ En índice de respuesta de destino, introduzca 7.
- ▶ A continuación, pulse en Ejecutar para ejecutar el procedimiento.

## Resultados

Figura 10-4  
Tabla de índice de respuesta

Número	Perfil			
	Descripción	Tamaño de grupo	Índice de respuesta	Índice de respuesta acumulado
1	Region = "West", "South", "East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West", "South", "East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West", "South", "East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

Verde: cumple el índice de respuesta de destino.  
Rojo: no cumple el índice de respuesta de destino.

La tabla de índice de respuesta muestra información acerca de cada grupo de perfiles identificados por el procedimiento.

- Los perfiles se muestran en orden descendente o por índice de respuesta.
- El índice de respuesta es el porcentaje de clientes que han respondido positivamente (han realizado una compra).
- El índice de respuesta acumulado es el índice de respuesta combinado de los grupos de perfiles actuales y anteriores. Ya que los perfiles se muestran en orden descendente del índice de respuesta, eso significa que el índice de respuesta acumulado es el índice de respuesta combinado para el grupo de perfiles actual junto con todos los grupos de perfiles con un índice de respuesta superior.
- La descripción del perfil incluye las características únicamente para aquellos campos que proporcionan una contribución significativa al modelo. En este ejemplo, la región, género y estado civil se incluyen en el modelo. El resto de campos, edad, ingresos, nivel educativo y años en la dirección actual, no se incluyen porque no han realizado una contribución significativa al modelo.
- El área verde de la tabla representa el conjunto de perfiles con un índice de respuesta acumulado es igual o mayor que el índice de respuesta de destino especificado, que en este ejemplo es 7%.
- El área roja de la tabla representa el conjunto de perfiles con un índice de respuesta acumulado menor que el índice de respuesta de destino especificado.
- El índice de respuesta acumulado en la última fila de la tabla es el índice de respuesta total o medio para todos los clientes incluidos en el correo de prueba, ya que es el índice de respuesta de todos los grupos de perfil.

Los resultados que se muestran en la tabla sugieren que si se dirige a mujeres en la zona oeste, sur y este, debería obtener un índice de respuesta ligeramente superior al índice de respuesta de destino.

Tenga en cuenta, sin embargo, que existe una diferencia sustancial entre los índices de respuesta de mujeres solteras (9,2%) y mujeres casadas (5,0%) en esas regiones. Aunque el índice de respuesta acumulado de ambos grupos es superior al índice de respuesta de destino, el índice de respuesta del último grupo solo es, de hecho, menor que el índice de respuesta de destino, que sugiere que es posible que desee buscar otras características que puedan mejorar el modelo.

### Resultado inteligente

Figura 10-5  
Resultado inteligente

El índice de respuesta La tabla muestra información sobre cada grupo de perfiles identificados por el procedimiento. La descripción del perfil incluye las características únicamente para aquellos campos que proporcionan una contribución significativa al modelo. No se incluyen los campos que no realizan una contribución significativa.

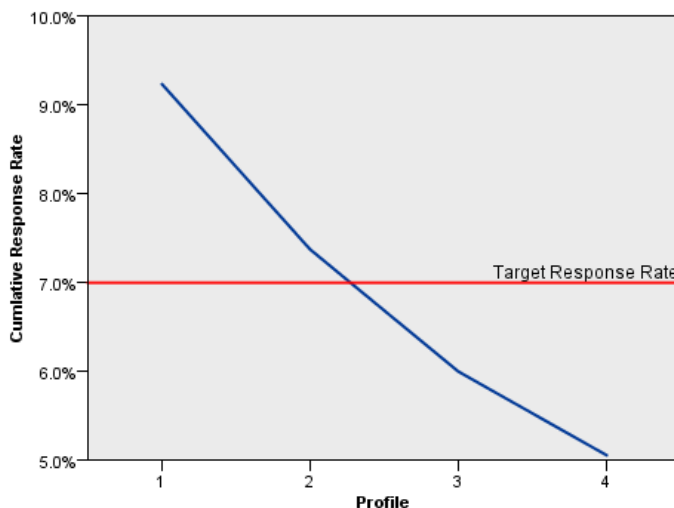
Los perfiles se muestran en orden descendente del índice de respuesta. El índice de respuesta es el porcentaje de clientes que han respondido de manera positiva (han realizado una compra).

El índice de respuesta acumulado es el índice de respuesta combinado para todos los grupos de perfiles anteriores. Ya que los perfiles se muestran en orden descendente del índice de respuesta, eso significa que el índice de respuesta acumulado es el índice de respuesta combinado para el grupo de perfiles actual junto con todos los grupos de perfiles con un índice de respuesta superior.

El índice de respuesta objetivo especificado es 7.00%. Las filas verdes tienen un índice de respuesta acumulado por encima de 7.00%, y las columnas rojas tienen un índice de respuesta acumulado por debajo de 7.00%. Aunque algunos grupos de perfiles en la zona verde pueden tener índices de respuestas individuales inferiores a 7.00%, el índice de respuesta acumulado en este punto es aún mayor a 7.00%.

La tabla está acompañada por “resultados inteligentes” que proporcionan información general sobre cómo interpretar la tabla e información específica sobre los resultados que se contienen en la tabla.

Figura 10-6  
Gráfico de índice de respuesta acumulado



El gráfico de índice de respuesta acumulado es básicamente una representación visual de los índices de respuesta acumulados que se muestran en la tabla. Como algunos perfiles se muestran en orden descendente según el índice de respuesta, la línea del índice de respuesta acumulado es siempre descendente para cada perfil. Al igual que la tabla, el gráfico muestra que el índice de respuesta acumulado es inferior al índice de respuesta de destino entre el grupo del perfil 2 y el grupo del perfil 3.

## **Resumen**

Para este correo de prueba concreto, se han identificado cuatro grupos de perfiles y los resultados indican que las únicas características demográficas significativas que parecen estar relacionadas con el hecho de que una persona haya respondido o no a la oferta son género, región y estado civil. El grupo con el mayor índice de respuesta está compuesto de mujeres solteras, que viven en el sur, este y oeste. Posteriormente, los índices de respuesta descienden rápidamente, aunque si incluye mujeres casadas de las mismas regiones obtendrá un índice de respuesta acumulada mayor que el índice de respuesta de destino.

# Índices de respuesta de código postal

Esta técnica utiliza resultados de una campaña anterior para calcular índices de respuesta de código postal. Los índices de respuesta se pueden utilizar para dirigirse a códigos postales concretos en campañas futuras.

Por ejemplo, basándose en los resultados de un envío anterior, el departamento de marketing directo de una empresa genera índices de respuesta por códigos postales. Basándose en diferentes criterios como un índice de respuesta mínimo aceptable y/o el número máximo de contactos que incluirán en el envío, podrán dirigirse a códigos postales concretos.

Esta información se recoge en el archivo *dmdata.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

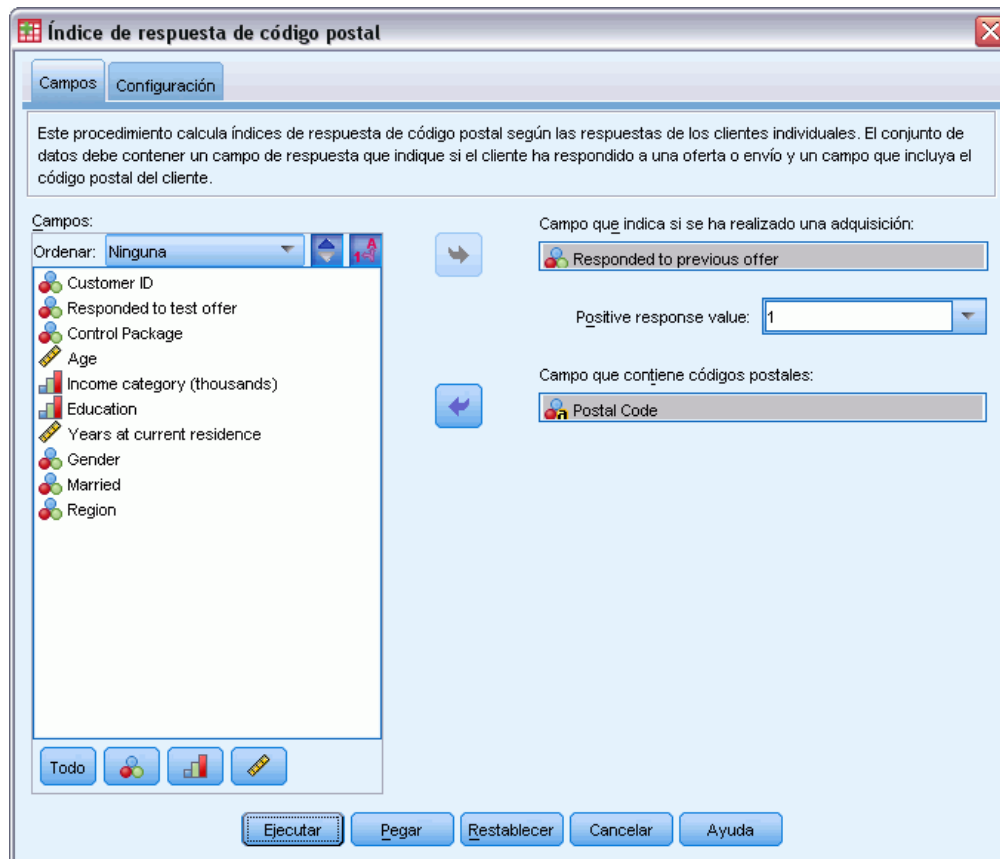
## Consideraciones de los datos

El campo de respuesta debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigne un valor de respuesta simple a todos los valores de respuesta distintos a cero. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica en el capítulo 5 el p. 33.](#)

## Ejecución del análisis

- ▶ Para calcular índices de respuesta de código postal desde los menús, seleccione:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Identificar principales códigos postales de respuesta y pulse Continuar.

Figura 11-1  
Pestaña Campos de Índices de respuesta de código postal



- ▶ En Campo Respuesta, seleccione *Ha respondido a la oferta anterior*.
- ▶ Para un valor de respuesta positiva, seleccione *Sí* en la lista desplegable. Un valor de 1 se muestra en el campo de texto porque “Sí” es realmente una etiqueta de valor asociado con un valor registrado de 1. (Si el valor de respuesta positiva no tiene un valor de etiqueta de valor definida, puede introducir el valor en el campo texto.)
- ▶ Para el Campo Código postal, seleccione *Código postal*.
- ▶ Pulse en la ficha Configuración.



Figura 11-2  
Pestaña Configuración de Índices de respuesta de código postal

- ▶ En el Agrupar códigos postales basados en grupos, seleccione Tres primeros dígitos o caracteres. Se calcularán los índices de respuesta combinados para todos los contactos con códigos postales que comiencen por los tres mismos primeros dígitos o caracteres. Por ejemplo, los tres primeros dígitos de un código postal de los Estados Unidos representan un área geográfica común mayor que el área geográfica definida por el código postal completo de 5 dígitos.
- ▶ En el grupo Resultados, seleccione (marque) Análisis de índice de respuesta y capacidad.
- ▶ Seleccione Índice de respuesta de destino e introduzca un valor de 5.
- ▶ Seleccione Número de contactos e introduzca un valor de 5000.
- ▶ A continuación, pulse en Ejecutar para ejecutar el procedimiento.

## Resultados

Figura 11-3  
Nuevo conjunto de datos con índices de respuesta por código postal

The screenshot shows the PASW Statistics Editor de datos interface. The main window displays a data table with 21 rows and 7 columns. The columns are labeled: CódigoPostal, ResponseRate, Responses, Contacts, Index, and Rank. The data is sorted in descending order of the Index column. The first row (row 1) is highlighted in yellow and shows a response rate of 10.00% for postal code 932. The status bar at the bottom indicates 'PASW Statistics Processor está listo'.

	CódigoPostal	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.00%	4.00	40	3.60	1
2	098	8.82%	6.00	68	5.47	1
3	740	7.76%	9.00	116	8.30	1
4	100	7.69%	7.00	91	6.46	1
5	110	7.69%	5.00	65	4.62	1
6	954	7.55%	4.00	53	3.70	1
7	108	7.32%	6.00	82	5.56	1
8	107	7.04%	5.00	71	4.65	1
9	090	6.90%	4.00	58	3.72	1
10	966	6.90%	4.00	58	3.72	1
11	760	6.72%	8.00	119	7.46	1
12	113	6.25%	5.00	80	4.69	1
13	927	6.00%	3.00	50	2.82	1
14	969	6.00%	3.00	50	2.82	1
15	972	5.88%	3.00	51	2.82	2
16	123	5.71%	2.00	35	1.89	2
17	950	5.66%	3.00	53	2.83	2
18	602	5.56%	6.00	108	5.67	2
19	617	5.56%	5.00	90	4.72	2
20	934	5.56%	3.00	54	2.83	2
21	946	5.36%	3.00	56	2.84	2

Se creará automáticamente un nuevo conjunto de datos. Este conjunto de datos contiene un único registro (fila) de cada código postal. En este ejemplo, cada fila contiene información de resumen de todos los códigos postales que comienzan con los tres mismos dígitos o caracteres.

Además de contener el código postal, el nuevo conjunto de datos contiene los siguientes campos:

- **Índice de respuesta.** El porcentaje de respuestas positivas en cada código postal. Los registros se clasifican automáticamente en orden descendente de índices de respuesta; de forma que los códigos postales con los mayores índices de respuesta aparecen en la parte superior del conjunto de datos.
- **Respuestas.** El número de respuestas positivas en cada código postal.
- **Contactos.** El número total de contactos en cada código postal que contiene un valor no perdido en el campo de respuesta.

- **Índice.** La respuesta “ponderada” basada en la fórmula  $N \times P \times (I-P)$ , donde  $N$  es el número de contactos y  $P$  es el índice de respuesta expresado como proporción. En el caso de dos códigos postales con el mismo índice de respuesta, esta fórmula asigna un valor de índice mayor al código postal con el mayor número de contactos.
- **Rango.** Rango de decil (10% superior, 20% superior, etc.) de los índices de respuesta por código postal en orden descendente.

Como hemos seleccionado un índice de respuesta y análisis de capacidad en la pestaña Configuración del cuadro de diálogo Índices de respuesta de código postal, se muestran en el Visor una tabla y un gráfico de índice de respuesta de resumen.

Figura 11-4

Tabla de índice de respuesta

Percentile	Índice de respuesta	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

Verde: cumple el índice de respuesta de destino.  
Rojo: no cumple el índice de respuesta de destino.

La tabla resume los resultados por deciles en orden descendente (10% superior, 20% superior, etc.).

- El índice de respuesta acumulado es el porcentaje combinado de respuestas positivas en las filas actual y anterior. Como los resultados se muestran por índices de respuesta en orden descendente, es el índice de respuestas combinado del decil actual y todos los deciles con un índice de respuestas superior.
- La tabla tiene una codificación de colores basada en los valores que haya introducido para un índice de respuesta de destino y un número máximo de contactos. Las filas con un índice de respuesta acumulado igual o mayor que el 5% y con 5.000 o menos contactos acumulados aparecen en verde. La codificación de colores se basa en el valor umbral que se alcance primero. En este ejemplo, ambos valores umbral se alcanzan en el mismo decil.

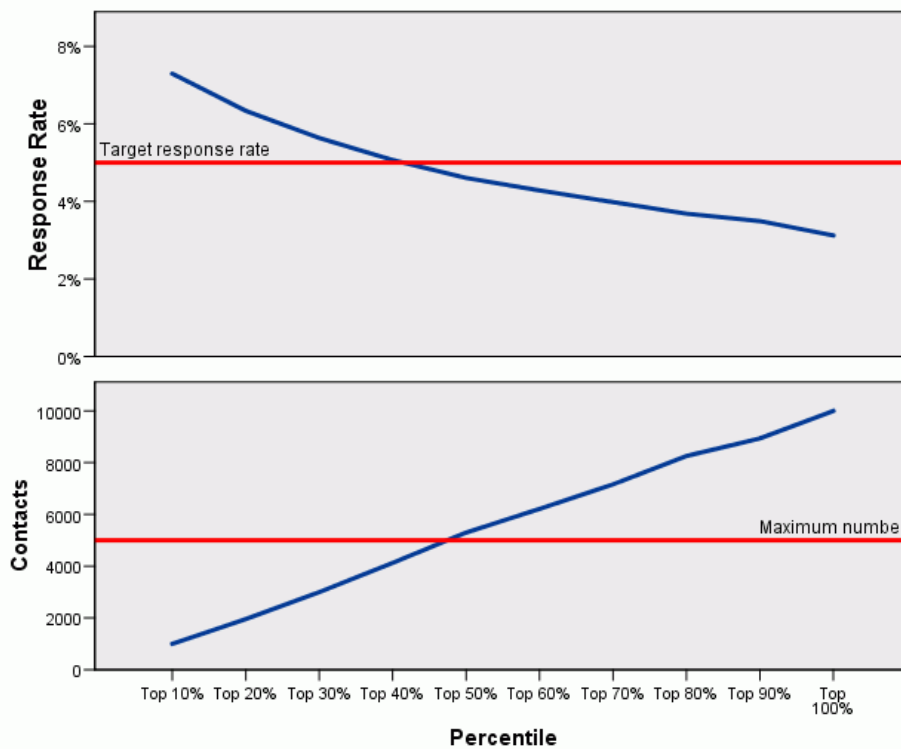
**Figura 11-5**  
*Resultado inteligente para la tabla de índice de respuestas*

La tabla de índice de respuesta resume los resultados por rango del decil en orden descendente (principales 10%, principales 20%, etc.). El índice de respuesta acumulado es el porcentaje combinado de respuestas positivas en las filas actuales y todas las precedentes. Como los resultados se muestran en orden descendente de índices de respuesta, éste es por lo tanto el índice de respuesta combinado para el decil actual y todos los deciles con un índice de respuesta más elevado. Como el rango del decil se incluye en el nuevo conjunto de datos, puede identificar fácilmente los códigos postales que cumplen cierto índice de respuesta acumulado. El campo del nuevo conjunto de datos que identifica el rango del decil se denomina Rango, donde 1 = principales 10%, 2 = principales 20%, etc.

El índice de respuesta mínimo especificado es 5.00%. El número máximo especificado de contactos es 5000. La codificación de colores de la tabla se basa en el valor de umbral que se alcance primero. Ambos umbrales se alcanzan en la misma categoría. Las filas verdes tienen un índice de respuesta mayor o igual al índice de respuesta mínimo especificado y un número acumulado de contactos inferior o igual al número máximo de contactos especificado. Las filas rojas tienen un índice de respuesta acumulado inferior al índice de respuesta mínimo especificado y un número acumulado de contactos superior al número máximo de contactos especificado.

La tabla está acompañada por texto que proporciona una descripción general sobre cómo leer la tabla. Si ha especificado un índice de respuesta mínimo o un número máximo de contactos, también incluye una sección que describe cómo se relacionan los resultados con los valores de umbral que especifique.

**Figura 11-6**  
*Gráfico de índice de respuesta acumulado*

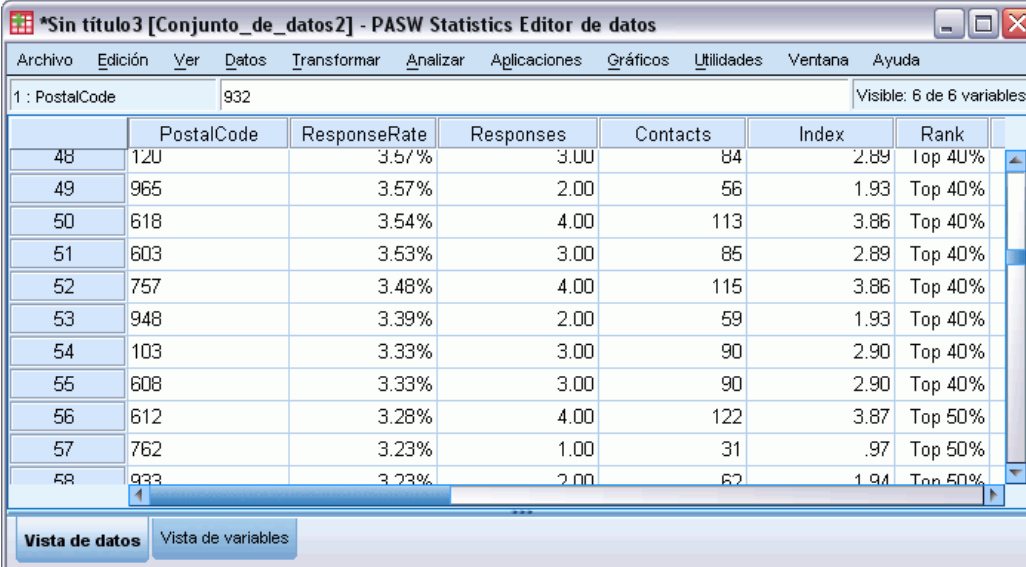


El gráfico del índice de respuesta acumulada y el número de contactos acumulados de cada decil es una representación visual de la misma información se muestra en la tabla de índices de respuesta. El umbral del índice de respuesta acumulado mínimo y el número acumulado máximo de contactos se alcanza entre el percentil 40 y 50.

- Como el gráfico muestra los índices de respuesta acumulados en orden descendente del decil del índice de respuesta, la línea del índice de respuesta acumulada siempre es inferior en cada decil siguiente.
- Como la línea del número de contactos representa el número acumulado de contactos, es siempre ascendente.

La información de la tabla y del gráfico indican que si desea lograr un índice de respuesta de al menos 5% pero no desea incluir más de 5.000 contactos en la campaña, debe centrarse en los códigos postales de los cuatro primeros deciles. Como el índice decil se incluye en el nuevo conjunto de datos, puede identificar fácilmente los códigos postales que cumplan el 40% principal de requisitos.

Figura 11-7  
Nuevo conjunto de datos



	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
48	120	3.57%	3.00	84	2.89	Top 40%
49	965	3.57%	2.00	56	1.93	Top 40%
50	618	3.54%	4.00	113	3.86	Top 40%
51	603	3.53%	3.00	85	2.89	Top 40%
52	757	3.48%	4.00	115	3.86	Top 40%
53	948	3.39%	2.00	59	1.93	Top 40%
54	103	3.33%	3.00	90	2.90	Top 40%
55	608	3.33%	3.00	90	2.90	Top 40%
56	612	3.28%	4.00	122	3.87	Top 50%
57	762	3.23%	1.00	31	.97	Top 50%
58	933	3.23%	2.00	67	1.94	Top 50%

Nota: *Índice* se registra como un valor entero de 1 a 10. El campo tiene etiquetas de valor definido, donde 1= 10% superior, 2= 20% superior, etc. Verá los valores de índices reales o las etiquetas de valor en el Visor de datos del editor de datos, dependiendo de su configuración de vistas.

## Resumen

El procedimiento de índices de respuesta de código postal utiliza resultados de una campaña anterior para calcular índices de respuesta de código postal. Los índices de respuesta se pueden utilizar para dirigirse a códigos postales concretos en campañas futuras. Este procedimiento crea automáticamente un nuevo conjunto de datos que contiene índices de respuesta por cada código postal. En función de la información de la tabla y el gráfico de índice de respuesta y la información del índice decil del nuevo conjunto de datos, puede identificar el conjunto de códigos postales que cumplan un índice de respuesta acumulado mínimo especificado y/o número máximo acumulado de contactos.

## ***Propensión a la compra***

La propensión a la compra utiliza resultados de un envío de prueba o campañas anteriores para generar puntuaciones de propensión. Las puntuaciones indican los contactos con más posibilidades de responder, en función de las diferentes características seleccionadas.

Esta técnica utiliza la regresión logística binaria para generar un modelo predictivo. El proceso de generar y aplicar un modelo predictivo tiene dos pasos básicos:

- ▶ Crear el modelo y guardar el archivo de modelo. Crea el modelo usando un conjunto de datos cuyo resultado deseado (a menudo conocido como **destino**) ya se conoce. Por ejemplo, si desea crear un modelo que prediga quién tiene más posibilidades de responder a una campaña de correo directo, deberá comenzar con un conjunto de datos que ya tenga información sobre quién respondió y quién no. Por ejemplo, esto puede tratarse de los resultados de un envío de prueba a un pequeño grupo de clientes o información sobre respuestas a una campaña similar en el pasado.
- ▶ Aplicar ese modelo a un conjunto de datos distinto (para el que no se conocen los resultados deseados) para obtener los resultados pronosticados.

Este ejemplo utiliza dos archivos de datos: *dmdata2.sav* se utiliza para generar el modelo y, a continuación, dicho modelo se aplica a *dmdata3.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

### ***Consideraciones de los datos***

El campo de respuesta (el resultado de destino deseado) debe ser categórico, con un valor que representa todas las respuestas positivas. Se asume que cualquier otro valor no perdido es una respuesta negativa. Si el campo de respuesta representa un valor continuo (escala), como el número de compras o las cantidades de compras, necesita crear un nuevo campo que asigna un valor de respuesta simple a todos los valores de respuesta distintos a cero. [Si desea obtener más información, consulte el tema Creación de un campo de respuesta categórica en el capítulo 6 el p. 40.](#)

### ***Generación de un modelo predictivo***

- ▶ Abra el archivo de datos *dmdata2.sav*.

Este archivo contiene diversas características demográficas de las personas que recibieron el envío de prueba, así como información sobre si respondieron o no al envío. Esta información se registra en el campo (variable) *Respondido*. El valor 1 indica que el contacto ha respondido al envío y el valor 0 indica que el contacto no ha respondido.

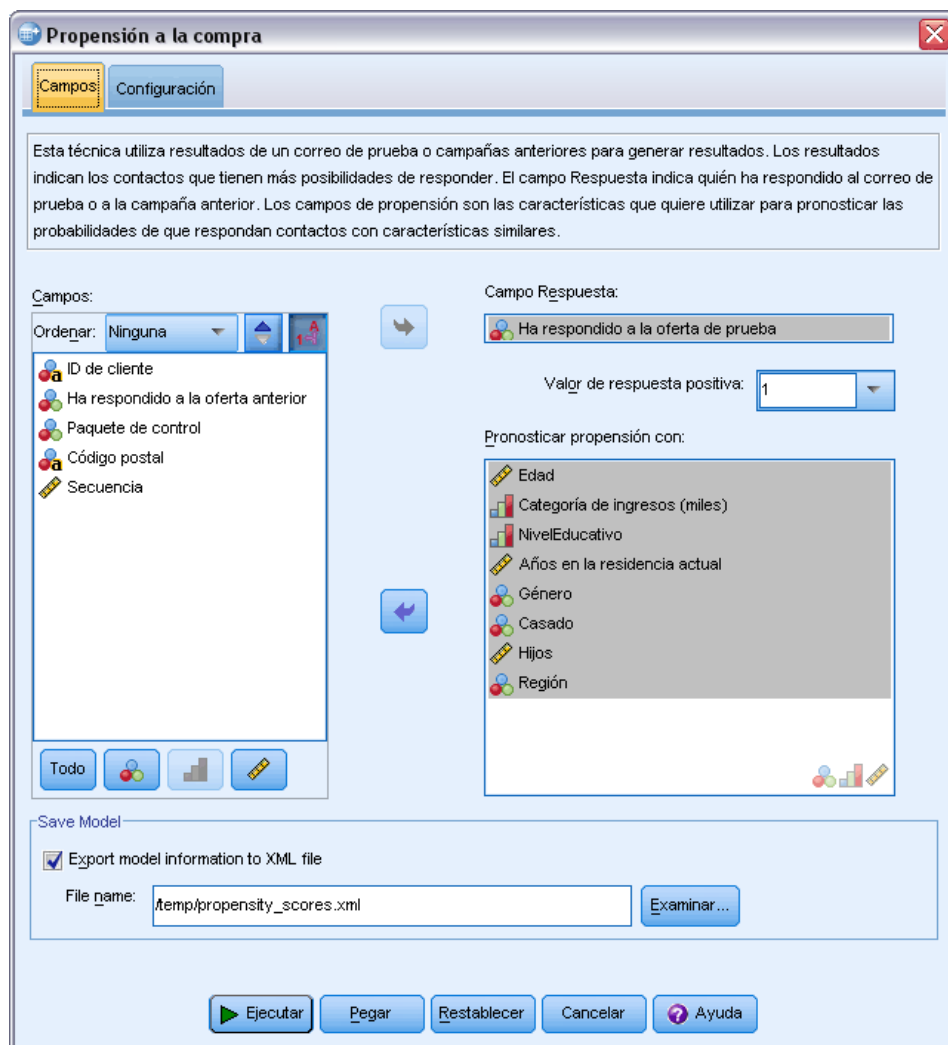
**Figura 12-1**

*Contenido de un archivo de datos en el Editor de datos*

ID	Respondido	Anterior	PaqueteDeControl	CódigoPostal	Edad	Ingresos	NivelEducativo	Residencia	Género
06262	0	0	1	96600	67	3	5	10	1
03692	0	0	0	95510	53	3	4	9	0
01480	0	0	1	92590	56	3	2	11	1
06118	0	0	1	92670	55	2	3	8	1
07378	0	0	1	92690	56	2	1	6	1
08467	0	0	0	93410	50	1	1	8	1
09621	0	0	1	93480	67	4	2	9	1
03029	0	0	0	93490	47	1	2	10	1
02660	0	0	1	93490	58	4	5	11	1

- ▶ Seleccione en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Seleccionar contactos con más probabilidades de realizar la compra y pulse en Continuar.

Figura 12-2  
Pestaña Campos de Propensión a la compra



- ▶ En Campo Respuesta, seleccione *Ha respondido a la oferta de prueba*.
- ▶ Para un valor de respuesta positiva, seleccione *Sí* en la lista desplegable. Un valor de 1 se muestra en el campo de texto porque “Sí” es realmente una etiqueta de valor asociado con un valor registrado de 1. (Si el valor de respuesta positiva no tiene un valor de etiqueta de valor definida, puede introducir el valor en el campo texto.)
- ▶ En Pronosticar propensión con, seleccione *Edad*, *Categoría de ingresos*, *Nivel educativo*, *Años en la residencia actual*, *Género*, *Casado*, *Región* e *Hijos*.
- ▶ Seleccione (active) Exportar información del modelo a un archivo XML.
- ▶ Pulse en Examinar para ir a donde desee guardar el archivo e introduzca un nombre para el mismo.
- ▶ En el cuadro de diálogo Propensión a la compra, pulse en la pestaña Configuración.



Figura 12-3  
Pestaña Configuración de Propensión a la compra

**Propensión a la compra**

**Campos Configuración**

**Validación de modelos**

Puede validar el modelo utilizado para generar resultados. Para validar el modelo, necesita dividir los datos en particiones. La partición de entrenamiento se utiliza para entrenar o crear el modelo. La partición de prueba se utiliza para validar el modelo. Si desea validar el modelo, esta técnica asignará automáticamente los registros a las particiones.

Valide el modelo

Tamaño de la partición de la muestra de entrenamiento (%):

Establecer semilla para replicar resultados

Semilla:

**Resultados diagnósticos**

Calidad global de modelo

Tabla de clasificación

Probabilidad mínima:

**Nombre y etiqueta del campo de respuesta recodificado**

Esta técnica automáticamente recodifica el campo de respuesta como un nuevo campo en el que 1 representa respuestas positivas y 0 representa respuestas negativas.

Nuevo nombre de campo:

Nueva etiqueta de campo:

**Guardar puntuaciones**

Esta técnica utiliza resultados de un correo de prueba o campañas anteriores para generar resultados. Las puntuaciones se guardan automáticamente para que pueda utilizarlas. El resto de controles de esta ficha proporcionan control adicional del contenido que se guarda.

Nuevo nombre de campo para puntuaciones:

**Ejecutar** **Pegar** **Restablecer** **Cancelar** **Ayuda**

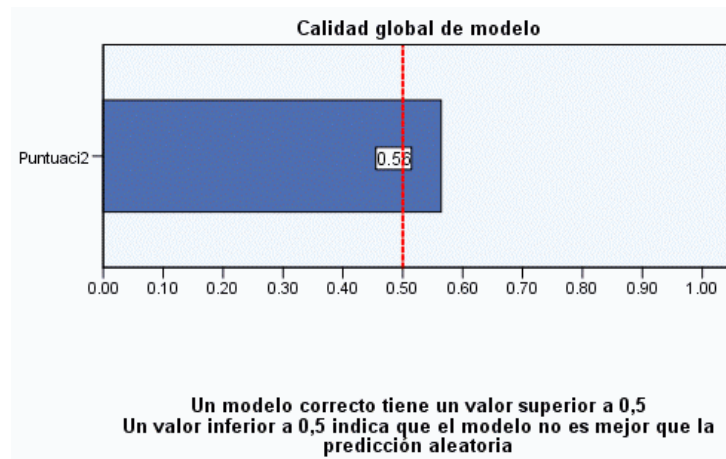
- ▶ En el grupo Validación de modelos, seleccione (marque) Validar modelo y Establecer semilla para replicar resultados.
- ▶ Utilice el tamaño de partición de muestra de formación predeterminado del 50% y el valor de semilla predeterminado de 2000000.
- ▶ En el grupo Resultados diagnósticos, seleccione (marque) Calidad global de modelo y Tabla de clasificación.
- ▶ Para Probabilidad mínima, introduzca 0,05. Como norma general, debe especificar un valor cercano a su tasa de respuesta de objetivo mínima, expresado como una proporción. Un valor de 0,05 representa un índice de respuesta del 5%.
- ▶ Pulse en **Ejecutar** para ejecutar el procedimiento y generar el modelo.

## Evaluación del modelo

Propensión a la compra produce un gráfico de calidad global de modelo y una tabla de clasificación que se pueden utilizar para evaluar el modelo.

El gráfico de calidad global de modelo proporciona una indicación visual rápida de la calidad del modelo. Como regla general, la calidad global de modelo debe estar por encima de 0,5.

Figura 12-4  
Gráfico de calidad global de modelo



Para confirmar que el modelo es adecuado para su puntuación, también debe examinar la tabla de clasificación.

Figura 12-5  
Tabla de clasificación

		Pronosticado					
		Muestra de entrenamiento			Muestra de prueba		
		Respuesta recodificada (1=Sí, 0=No)		Porcentaje correcto	Respuesta recodificada (1=Sí, 0=No)		Porcentaje correcto
Observado	No	Sí	No		Sí		
Respuesta recodificada (1=Sí, 0=No)	No	651	249	72.33	653	267	70.98
	Sí	19	20	51.28	36	22	37.93
Porcentaje global		2.84	7.43	71.46	5.22	7.61	69.02

La tabla de clasificación compara valores pronosticados del campo de destino con los valores reales del campo de destino. El porcentaje de precisión global puede proporcionar algún tipo de indicación del funcionamiento del modelo, pero puede que esté más interesado en el porcentaje de respuestas positivas correctas pronosticadas si el objetivo es generar un modelo que identifique el grupo de contactos que probablemente proporcionen un índice de respuesta positivo igual o mayor que el índice de respuesta positivo mínimo especificado.

En este ejemplo, la tabla de clasificación se divide en una **muestra de formación** y una **muestra de prueba**. La muestra de formación se utiliza para generar el modelo. A continuación, el modelo se aplica a la muestra de prueba para comprobar el funcionamiento del modelo.

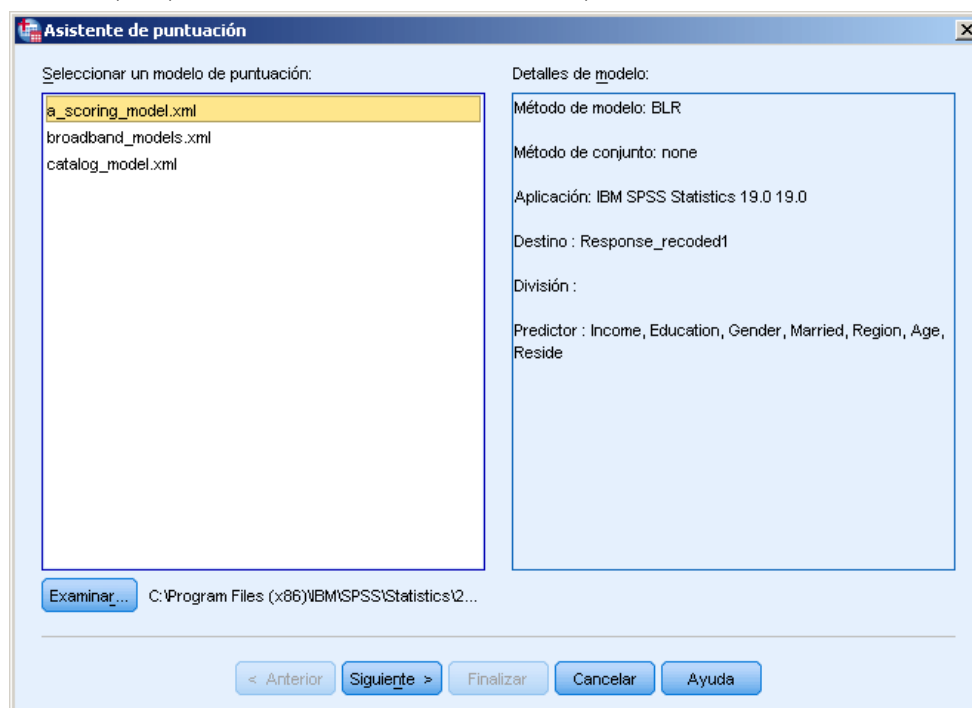
El índice de respuesta mínimo especificado fue de 0,05 o del 5%. La tabla de clasificación muestra que el índice de clasificación correcto para respuestas positivas es del 7,43% en la muestra de formación y del 7,61% en la muestra de prueba. Como el índice de respuesta de la muestra de prueba es mayor del 5%, este modelo debería poder identificar un grupo de contactos que probablemente proporcionen un índice de respuesta mayor del 5%.

## Aplicación del modelo

- ▶ Abra el archivo de datos *dmdata3.sav*. Este archivo de datos contiene información demográfica y de otro tipo de todos los contactos que no se incluyeron en el envío de prueba. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)
- ▶ Abra el Asistente para puntuación. Para abrir el Asistente para puntuación, elija en los menús: Utilidades > Asistente para puntuación

Figura 12-6

Asistente para puntuación, Seleccionar un modelo de puntuación



- ▶ Pulse en Examinar para ir a la ubicación donde guardó el archivo XML de modelo y pulse en Seleccionar en el cuadro de diálogo Examinar.

Todos los archivos con una extensión .xml o .zip se muestran en el Asistente para puntuación. Si el archivo seleccionado se reconoce como un archivo de modelo válido, aparecerá una descripción del modelo.

- ▶ Seleccione el archivo XML de modelo que ha creado y, a continuación, pulse en Siguiente.

Figura 12-7  
Asistente para puntuación, Comparar campos de modelo

Nombre de modelo: propensity\_scores  
Tipo de modelo: Regresión logística binaria

Igualar campos de modelo con el conjunto de datos:

Campos de conjunto de datos	Campos de modelo	Rol	Medida	Tipo
CategoríaIngresos	Income	Predictor	Nominal	Numérico
NivelEducativo	Education	Predictor	Nominal	Numérico
Género	Gender	Predictor	Nominal	Numérico
Casado	Married	Predictor	Nominal	Numérico
Región	Region	Predictor	Nominal	Numérico
Edad	Age	Predictor	Continuo	Numérico
Residencia	Reside	Predictor	Continuo	Numérico
ID	Sequence	Predictor	Continuo	Numérico

Valores perdidos

Utilizar sustitución de valores  
 Utilizar valores perdidos del sistema

< Anterior   Siguiente >   Finalizar   Cancelar   Ayuda

Para poder puntuar el conjunto de datos activo, éste debe contener campos (variables) que correspondan a todos los predictores del modelo. Si el modelo también contiene campos divididos, el conjunto de datos también contendrá campos que corresponderá con todos los campos divididos en el modelo.

- Por defecto, los campos del conjunto de datos activo que tengan el mismo nombre y tipo que los campos del modelo se emparejarán automáticamente.
- Use la lista desplegable para emparejar campos de conjuntos de datos con campos de modelo. El tipo de datos de cada campo debe ser igual tanto en el modelo como en el conjunto de datos para comparar campos.
- No podrá continuar con el asistente ni puntuar el conjunto de datos activo a no ser que todos los predictores (y campos divididos, si los hay) del modelo se emparejen con los campos del conjunto de datos activo.

El conjunto de datos activo no contiene ningún campo denominado *Ingresos*. Por lo tanto, la casilla de la columna Campos de conjunto de datos que corresponde al campo de modelo *Ingresos* está en blanco en un principio. Debe seleccionar un campo del conjunto de datos activo que sea equivalente a dicho campo de modelo.

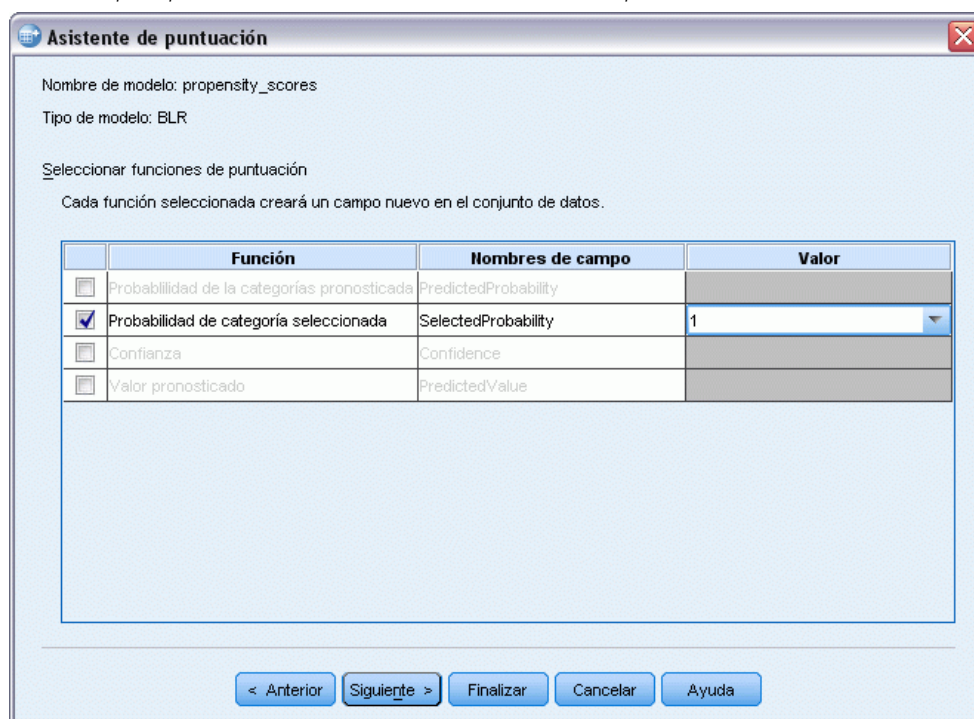
- En la lista desplegable de la columna Campos de conjunto de datos de la casilla en blanco para el campo de modelo *Ingresos*, seleccione *CategoríaIngresos*.

*Nota:* Además del nombre y tipo de campo, debería asegurarse de que los valores de datos reales del conjunto de datos que se está puntuando se registran del mismo modo que los valores de datos del conjunto de datos utilizado para generar el modelo. Por ejemplo, si el modelo se ha generado con un campo *Ingresos* que divide los ingresos en cuatro categorías y *CategoríaIngresos* del conjunto de datos activo divide los ingresos en seis categorías o cuatro categorías diferentes, esos campos no coinciden y las puntuaciones resultantes no serán fiables.

Pulse en **Siguiente** para ir al paso siguiente del Asistente para puntuación.

Figura 12-8

Asistente para puntuación: selección de las funciones de puntuación



Las funciones de puntuación son los tipos de “puntuaciones” disponibles en el modelo seleccionado. Las funciones de puntuación disponibles dependen del modelo. Para el modelo logístico binario utilizado en este ejemplo, las funciones disponibles son valor pronosticado, probabilidad del valor pronosticado, probabilidad de un valor seleccionado y confianza.

En este ejemplo, estamos interesados en la probabilidad pronosticada de una respuesta positiva al envío; por lo tanto, queremos la probabilidad de un valor seleccionado.

- ▶ Seleccione (marque) Probabilidad de la categoría seleccionada.
- ▶ En la columna Valor, seleccione 1 en la lista desplegable. La lista de posibles valores para el destino se define en el modelo, según los valores de destino del archivo de datos utilizado para generar el modelo.
- ▶ Anule la selección del resto de funciones de puntuación.

- ▶ También puede asignar un nombre más descriptivo al nuevo campo que contendrá los valores de puntuación del conjunto de datos activo. Por ejemplo, *Probabilidad\_de\_respuesta*.
- ▶ Pulse en Finalizar para aplicar el modelo al conjunto de datos activo.

El nuevo campo que contiene la probabilidad de una respuesta positiva se añade al final del conjunto de datos.

**Figura 12-9**

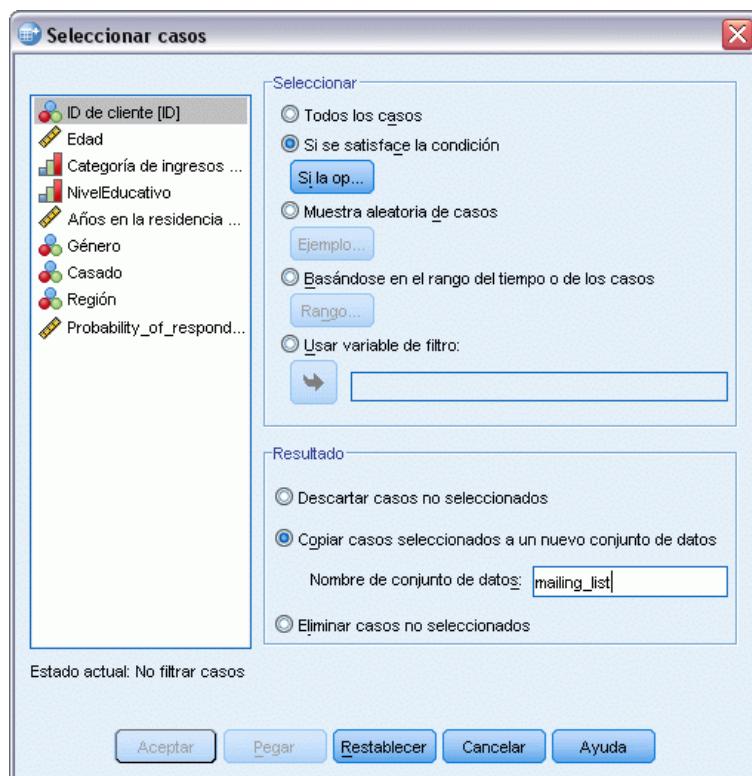
*Conjunto de datos con un nuevo campo de probabilidad*

Residencia	Género	Casado	Región	Probability_of_responding
10	1	0	4	.04
9	0	0	4	.03
11	1	0	4	.03
8	1	0	4	.04
6	1	0	4	.07
8	1	0	4	.04
9	1	0	4	.03
10	1	0	4	.05
11	1	0	4	.05
11	1	0	4	.02
14	1	0	4	.12

A continuación, puede utilizar ese campo para seleccionar el subconjunto de contactos que probablemente produzcan un índice de respuesta positivo a cierto nivel o por encima del mismo. Por ejemplo, puede crear un nuevo conjunto de datos que contenga el subconjunto de casos que probablemente produzcan un índice de respuesta positivo de como mínimo el 5%.

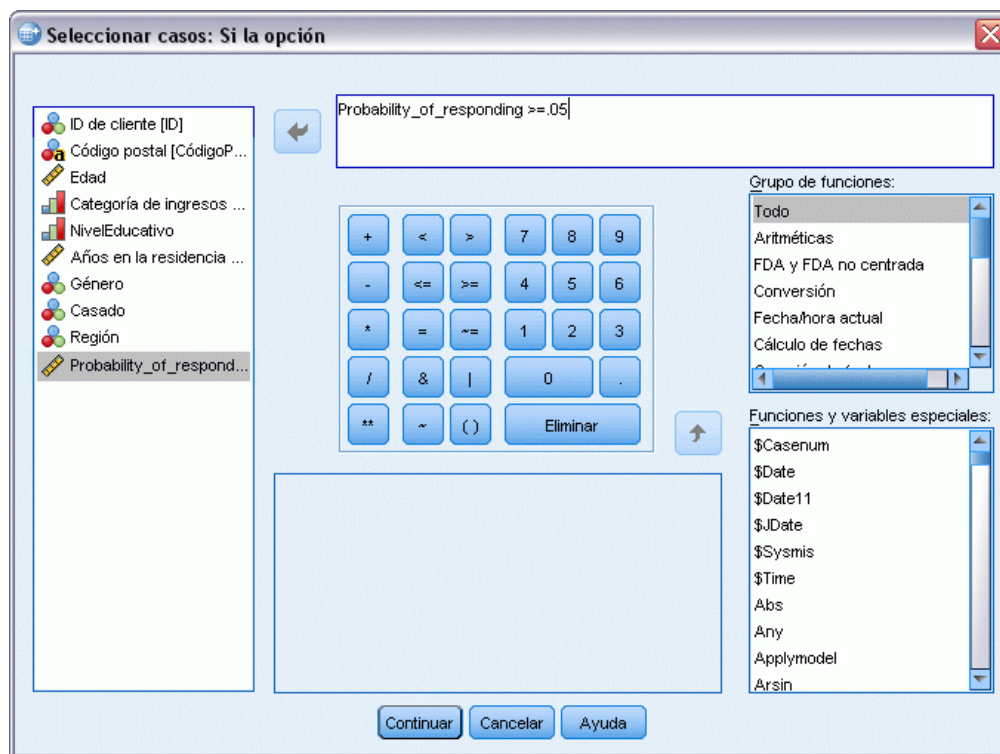
- Seleccione en los menús:  
Datos > Seleccionar casos

Figura 12-10  
Cuadro de diálogo Seleccionar casos



- En el cuadro de diálogo Seleccionar casos, seleccione Si se satisface la condición y pulse en Si.

Figura 12-11  
Seleccionar casos: Cuadro de diálogo Si



- ▶ En Seleccionar casos: Si hay cuadro de diálogo introduzca la siguiente expresión:

Probabilidad\_de\_respuesta >=0,05

*Nota:* Si ha utilizado un nombre diferente para el campo que contiene los valores de probabilidad, introduzca ese nombre en lugar de Probabilidad\_de\_respuesta. El nombre del campo por defecto es *ProbabilidadSeleccionada*.

- ▶ Pulse en Continuar.
- ▶ En el cuadro de diálogo Seleccionar casos, seleccione Copiar casos seleccionados en un nuevo conjunto de datos e introduzca un nombre para el nuevo conjunto de datos. El nombre de un conjunto de datos debe cumplir las normas de denominación de campos (variables).
- ▶ Pulse en Aceptar para crear el conjunto de datos con los contactos seleccionados.



El nuevo conjunto de datos sólo contiene los contactos que tengan una probabilidad pronosticada de una respuesta positiva de como mínimo el 5%.

**Figura 12-12**

*Nuevo conjunto de datos con los contactos seleccionados*

Residencia	Género	Casado	Región	Probability_of_responding
13	0	0	4	.07
15	0	0	4	.05
10	0	0	4	.05
5	0	0	4	.12
7	0	0	4	.08
10	0	0	4	.10
15	1	0	4	.05
11	1	0	4	.08
9	1	0	4	.08
9	1	0	4	.05

## **Resumen**

La propensión a la compra utiliza resultados de un envío de prueba o campañas anteriores para generar puntuaciones de propensión. Las puntuaciones indican los contactos con más posibilidades de responder, en función de las diferentes características seleccionadas. Esta técnica genera un modelo predictivo que luego puede aplicarse a un conjunto de datos para obtener puntuaciones de propensión.

## ***Prueba de paquete de control***

Esta técnica compara campañas de marketing para ver si hay diferencias significativas en la eficacia de distintos paquetes u ofertas. La eficacia de una campaña se mide por las respuestas.

Por ejemplo, la división de marketing directo de una empresa quiere ver si un nuevo diseño de paquete generará respuestas más positivas que el paquete actual. Envían un correo de prueba para determinar si el nuevo paquete genera un porcentaje de respuesta significativamente más positivo. El correo de prueba consiste en un grupo de control que recibe el paquete actual y un grupo de prueba que recibe el nuevo diseño del paquete. Los resultados de los dos grupos se comparan para ver si hay una diferencia significativa.

Esta información se recoge en el archivo *dmdata.sav*. [Si desea obtener más información, consulte el tema Archivos muestrales en el apéndice A el p. 99.](#)

### ***Ejecución del análisis***

- ▶ Para realizar una prueba de paquete de control, elija en los menús:  
Marketing directo > Seleccionar técnica
- ▶ Seleccione Comparar la eficacia de las campañas (Prueba de paquete de control) y pulse Continuar.

Figura 13-1  
Prueba de paquete de control, pestaña Campos

Esta técnica compara campañas de marketing para ver si hay diferencias significativas entre los distintos paquetes u ofertas. El campo Campaña identifica las distintas campañas. El campo Eficacia indica si un cliente ha respondido a la campaña.

Campos:

Ordenar: Ninguna

- Customer ID
- Responded to previous offer
- Postal Code
- Age
- Income category (thousands)
- Education
- Years at current residence
- Gender
- Married
- Region

Campaña: Control Package

Eficacia: Responded to test offer

Medida de la eficacia

Cantidad de adquisición

Respuesta positiva

Valor: 1

Nombre and Label for Recoded Effectiveness Response Field

Quando la eficacia es una respuesta positiva, esta técnica crea automáticamente un campo de eficacia sí/no para su análisis.

Nuevo nombre de campo: Effectiveness\_new1

Nueva etiqueta de campo: Campo de prueba grabado (1=SÍ 0=N)

Ejecutar Pegar Restablecer Cancelar Ayuda

- ▶ En el campo Campaña, seleccione *Paquete de control*.
- ▶ En el campo Respuesta de eficacia, seleccione *Ha respondido a la oferta de prueba*.
- ▶ Seleccione Respuesta.
- ▶ Para un valor de respuesta positiva, seleccione *Sí* en la lista desplegable. Un valor de 1 se muestra en el campo de texto porque “Sí” es realmente una etiqueta de valor asociado con un valor registrado de 1. (Si el valor de respuesta positiva no tiene un valor de etiqueta de valor definida, puede introducir el valor en el campo texto.)

Se crea automáticamente un nuevo campo, en el que 1 representa las respuestas positivas y 0 representa las respuestas negativas y el análisis se ejecuta en el nuevo campo. Puede sustituir el nombre por defecto y asignar uno de su elección. Para este ejemplo, utilizaremos el nombre de campo que ya se ha suministrado.

- ▶ Pulse en Ejecutar para ejecutar el procedimiento.

## Resultados

Figura 13-2  
Resultado de prueba de paquete de control

		Paquete de control			
		Control		Prueba	
		Recuento	% del N de la columna	Recuento	% del N de la columna
Eficacia (1=S? 0=NO)	0	879	96,6%	984	97,7%
	1	31	3,4%	23	2,3%

No hay una diferencia significativa estadísticamente entre Control y Prueba.

Los resultados del procedimiento incluyen una tabla que muestra los recuentos y porcentajes de respuestas positivas y negativas de cada grupo definidos por el campo Campaña y una tabla que identifica si los índices de respuesta de los grupos son muy diferentes entre sí.

- *Eficacia* es la versión recodificada del campo de respuesta, donde 1 representa las respuestas positivas y 0 representa las respuestas negativas.
- El índice de respuestas positivas del paquete de control es del 3,8%, mientras que el de respuestas positivas del paquete de prueba es del 6,2%.

La descripción de texto simple que hay bajo la tabla indica que la diferencia entre los grupos es considerablemente distinta, lo que significa que la proporción de respuestas mayor para el paquete de prueba probablemente no sea el resultado de una opción aleatoria. Esta tabla de texto contendrá una comparación para cada par posible grupos incluidos en el análisis. Como sólo hay dos grupos en estos ejemplos, sólo hay una comparación. Si hubiese más de cinco grupos, la tabla de descripción de texto se sustituye por la tabla Comparaciones de proporciones de columna.

## Resumen

La Prueba de paquete de control compara campañas de marketing para ver si hay diferencias significativas en la eficacia entre los distintos paquetes u ofertas. En este ejemplo, la respuesta positiva del 6,2% para el paquete de prueba era considerablemente superior al índice de respuestas positivas del 3,8% para el paquete de control. Esto sugiere que debe utilizar el nuevo diseño de paquete en vez del anterior, pero puede haber otros factores que debe tener en cuenta, como los costes adicionales asociados con el nuevo diseño del paquete.

## ***Archivos muestrales***

Los archivos muestrales instalados con el producto se encuentran en el subdirectorio *Samples* del directorio de instalación. Hay una carpeta independiente dentro del subdirectorio *Samples* para cada uno de los siguientes idiomas: Inglés, francés, alemán, italiano, japonés, coreano, polaco, ruso, chino simplificado, español y chino tradicional.

No todos los archivos muestrales están disponibles en todos los idiomas. Si un archivo muestral no está disponible en un idioma, esa carpeta de idioma contendrá una versión en inglés del archivo muestral.

### ***Descripciones***

A continuación, se describen brevemente los archivos muestrales usados en varios ejemplos que aparecen a lo largo de la documentación.

- **accidents.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo de edad y género que influyen en los accidentes de automóviles de una región determinada. Cada caso corresponde a una clasificación cruzada de categoría de edad y género.
- **adl.sav.** Archivo de datos hipotéticos relativo a los esfuerzos para determinar las ventajas de un tipo propuesto de tratamiento para pacientes que han sufrido un derrame cerebral. Los médicos dividieron de manera aleatoria a pacientes (mujeres) que habían sufrido un derrame cerebral en dos grupos. El primer grupo recibió el tratamiento físico estándar y el segundo recibió un tratamiento emocional adicional. Tres meses después de los tratamientos, se puntuaron las capacidades de cada paciente para realizar actividades cotidianas como variables ordinales.
- **advert.sav.** Archivo de datos hipotéticos sobre las iniciativas de un minorista para examinar la relación entre el dinero invertido en publicidad y las ventas resultantes. Para ello, se recopilaron las cifras de ventas anteriores y los costes de publicidad asociados.
- **aflatoxin.sav.** Archivo de datos hipotéticos sobre las pruebas realizadas en las cosechas de maíz con relación a la aflatoxina, un veneno cuya concentración varía ampliamente en los rendimientos de cultivo y entre los mismos. Un procesador de grano ha recibido 16 muestras de cada uno de los 8 rendimientos de cultivo y ha medido los niveles de aflatoxinas en partes por millón (PPM).
- **anorectic.sav.** Mientras trabajaban en una sintomatología estandarizada del comportamiento anoréxico/bulímico, los investigadores realizaron un estudio de 55 adolescentes con trastornos de la alimentación conocidos. Cada paciente fue examinado cuatro veces durante cuatro años, lo que representa un total de 220 observaciones. En cada observación, se puntuó a los pacientes por cada uno de los 16 síntomas. Faltan las puntuaciones de los síntomas para el paciente 71 en el tiempo 2, el paciente 76 en el tiempo 2 y el paciente 47 en el tiempo 3, lo que nos deja 217 observaciones válidas.

- **bankloan.sav.** Archivo de datos hipotéticos sobre las iniciativas de un banco para reducir la tasa de moras de créditos. El archivo contiene información financiera y demográfica de 850 clientes anteriores y posibles clientes. Los primeros 700 casos son clientes a los que anteriormente se les ha concedido un préstamo. Al menos 150 casos son posibles clientes cuyos riesgos de crédito el banco necesita clasificar como positivos o negativos.
- **bankloan\_binning.sav.** Archivo de datos hipotéticos que contiene información financiera y demográfica sobre 5.000 clientes anteriores.
- **behavior.sav.** En un ejemplo clásico, se pidió a 52 estudiantes que valoraran las combinaciones de 15 situaciones y 15 comportamientos en una escala de 10 puntos que oscilaba entre 0 =“extremadamente apropiado” y 9=“extremadamente inapropiado”. Los valores promediados respecto a los individuos se toman como disimilaridades.
- **behavior\_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución bidimensional de *behavior.sav*.
- **brakes.sav.** Archivo de datos hipotéticos sobre el control de calidad de una fábrica que produce frenos de disco para automóviles de alto rendimiento. El archivo de datos contiene las medidas del diámetro de 16 discos de cada una de las 8 máquinas de producción. El diámetro objetivo para los frenos es de 322 milímetros.
- **breakfast.sav.** En un estudio clásico, se pidió a 21 estudiantes de administración de empresas de la Wharton School y sus cónyuges que ordenaran 15 elementos de desayuno por orden de preferencia, de 1=“más preferido” a 15=“menos preferido”. Sus preferencias se registraron en seis escenarios distintos, de “Preferencia global” a “Aperitivo, con bebida sólo”.
- **breakfast-overall.sav.** Este archivo de datos sólo contiene las preferencias de elementos de desayuno para el primer escenario, “Preferencia global”.
- **broadband\_1.sav** Archivo de datos hipotéticos que contiene el número de suscriptores, por región, a un servicio de banda ancha nacional. El archivo de datos contiene números de suscriptores mensuales para 85 regiones durante un período de cuatro años.
- **broadband\_2.sav** Este archivo de datos es idéntico a *broadband\_1.sav* pero contiene datos para tres meses adicionales.
- **car\_insurance\_claims.sav.** Un conjunto de datos presentados y analizados en otro lugar estudia las reclamaciones por daños en vehículos. La cantidad de reclamaciones media se puede modelar como si tuviera una distribución Gamma, mediante una función de enlace inversa para relacionar la media de la variable dependiente con una combinación lineal de la edad del asegurado, el tipo de vehículo y la antigüedad del vehículo. El número de reclamaciones presentadas se puede utilizar como una ponderación de escalamiento.
- **car\_sales.sav.** Este archivo de datos contiene estimaciones de ventas, precios de lista y especificaciones físicas hipotéticas de varias marcas y modelos de vehículos. Los precios de lista y las especificaciones físicas se han obtenido de *edmunds.com* y de sitios de fabricantes.
- **car\_sales\_uprepared.sav.** Ésta es una versión modificada de *car\_sales.sav* que no incluye ninguna versión transformada de los campos.
- **carpet.sav** En un ejemplo muy conocido, una compañía interesada en sacar al mercado un nuevo limpiador de alfombras desea examinar la influencia de cinco factores sobre la preferencia del consumidor: diseño del producto, marca comercial, precio, sello de *buen producto para el hogar* y garantía de devolución del importe. Hay tres niveles de factores para el diseño del producto, cada uno con una diferente colocación del cepillo del aplicador;

tres nombres comerciales (*K2R*, *Glory* y *Bissell*); tres niveles de precios; y dos niveles (no o sí) para los dos últimos factores. Diez consumidores clasificaron 22 perfiles definidos por estos factores. La variable *Preferencia* contiene el rango de las clasificaciones medias de cada perfil. Las clasificaciones inferiores corresponden a preferencias elevadas. Esta variable refleja una medida global de la preferencia de cada perfil.

- **carpet\_prefs.sav** Este archivo de datos se basa en el mismo ejemplo que el descrito para *carpet.sav*, pero contiene las clasificaciones reales recogidas de cada uno de los 10 consumidores. Se pidió a los consumidores que clasificaran los 22 perfiles de los productos empezando por el menos preferido. Las variables desde *PREF1* hasta *PREF22* contienen los ID de los perfiles asociados, como se definen en *carpet\_plan.sav*.
- **catalog.sav** Este archivo de datos contiene cifras de ventas mensuales hipotéticas de tres productos vendidos por una compañía de venta por catálogo. También se incluyen datos para cinco variables predictoras posibles.
- **catalog\_seasfac.sav** Este archivo de datos es igual que *catalog.sav*, con la excepción de que incluye un conjunto de factores estacionales calculados a partir del procedimiento Descomposición estacional junto con las variables de fecha que lo acompañan.
- **cellular.sav**. Archivo de datos hipotéticos sobre las iniciativas de una compañía de telefonía móvil para reducir el abandono de clientes. Las puntuaciones de propensión al abandono de clientes se aplican a las cuentas, oscilando de 0 a 100. Las cuentas con una puntuación de 50 o superior pueden estar buscando otros proveedores.
- **ceramics.sav**. Archivo de datos hipotéticos sobre las iniciativas de un fabricante para determinar si una nueva aleación de calidad tiene una mayor resistencia al calor que una aleación estándar. Cada caso representa una prueba independiente de una de las aleaciones; la temperatura a la que registró el fallo del rodamiento.
- **cereal.sav**. Archivo de datos hipotéticos sobre una encuesta realizada a 880 personas sobre sus preferencias en el desayuno, teniendo también en cuenta su edad, sexo, estado civil y si tienen un estilo de vida activo o no (en función de si practican ejercicio al menos dos veces a la semana). Cada caso representa un encuestado diferente.
- **clothing\_defects.sav**. Archivo de datos hipotéticos sobre el proceso de control de calidad en una fábrica de prendas. Los inspectores toman una muestra de prendas de cada lote producido en la fábrica, y cuentan el número de prendas que no son aceptables.
- **coffee.sav**. Este archivo de datos pertenece a las imágenes percibidas de seis marcas de café helado. Para cada uno de los 23 atributos de imagen de café helado, los encuestados seleccionaron todas las marcas que quedaban descritas por el atributo. Las seis marcas se denotan AA, BB, CC, DD, EE y FF para mantener la confidencialidad.
- **contacts.sav**. Archivo de datos hipotéticos sobre las listas de contactos de un grupo de representantes de ventas de ordenadores de empresa. Cada uno de los contactos está categorizado por el departamento de la compañía en el que trabaja y su categoría en la compañía. Además, también se registran los importes de la última venta realizada, el tiempo transcurrido desde la última venta y el tamaño de la compañía del contacto.
- **creditpromo.sav**. Archivo de datos hipotéticos sobre las iniciativas de unos almacenes para evaluar la eficacia de una promoción de tarjetas de crédito reciente. Para este fin, se seleccionaron aleatoriamente 500 titulares. La mitad recibieron un anuncio promocionando una tasa de interés reducida sobre las ventas realizadas en los siguientes tres meses. La otra mitad recibió un anuncio estacional estándar.

- **customer\_dbase.sav.** Archivo de datos hipotéticos sobre las iniciativas de una compañía para usar la información de su almacén de datos para realizar ofertas especiales a los clientes con más probabilidades de responder. Se seleccionó un subconjunto de la base de clientes aleatoriamente a quienes se ofrecieron las ofertas especiales y sus respuestas se registraron.
- **customer\_information.sav.** Archivo de datos hipotéticos que contiene la información de correo del cliente, como el nombre y la dirección.
- **customer\_subset.sav.** Un subconjunto de 80 casos de *customer\_dbase.sav*.
- **debate.sav.** Archivos de datos hipotéticos sobre las respuestas emparejadas de una encuesta realizada a los asistentes a un debate político antes y después del debate. Cada caso corresponde a un encuestado diferente.
- **debate\_aggregate.sav.** Archivo de datos hipotéticos que agrega las respuestas de *debate.sav*. Cada caso corresponde a una clasificación cruzada de preferencias antes y después del debate.
- **demo.sav.** Archivos de datos hipotéticos sobre una base de datos de clientes adquirida con el fin de enviar por correo ofertas mensuales. Se registra si el cliente respondió a la oferta, junto con información demográfica diversa.
- **demo\_cs\_1.sav.** Archivo de datos hipotéticos sobre el primer paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una ciudad diferente, y se registra la identificación de la ciudad, la región, la provincia y el distrito.
- **demo\_cs\_2.sav.** Archivo de datos hipotéticos sobre el segundo paso de las iniciativas de una compañía para recopilar una base de datos de información de encuestas. Cada caso corresponde a una unidad familiar diferente de las ciudades seleccionadas en el primer paso, y se registra la identificación de la unidad, la subdivisión, la ciudad, el distrito, la provincia y la región. También se incluye la información de muestreo de las primeras dos etapas del diseño.
- **demo\_cs.sav.** Archivo de datos hipotéticos que contiene información de encuestas recopilada mediante un diseño de muestreo complejo. Cada caso corresponde a una unidad familiar distinta, y se recopila información demográfica y de muestreo diversa.
- **dmdata.sav.** Éste es un archivo de datos hipotéticos que contiene información demográfica y de compras para una empresa de marketing directo. *dmdata2.sav* contiene información para un subconjunto de contactos que recibió un envío de prueba, y *dmdata3.sav* contiene información sobre el resto de contactos que no recibieron el envío de prueba.
- **dietstudy.sav.** Este archivo de datos hipotéticos contiene los resultados de un estudio sobre la “dieta Stillman”. Cada caso corresponde a un sujeto distinto y registra sus pesos antes y después de la dieta en libras y niveles de triglicéridos en mg/100 ml.
- **dvdplayer.sav.** Archivo de datos hipotéticos sobre el desarrollo de un nuevo reproductor de DVD. El equipo de marketing ha recopilado datos de grupo de enfoque mediante un prototipo. Cada caso corresponde a un usuario encuestado diferente y registra información demográfica sobre los encuestados y sus respuestas a preguntas acerca del prototipo.
- **german\_credit.sav.** Este archivo de datos se toma del conjunto de datos “German credit” de las Repository of Machine Learning Databases de la Universidad de California, Irvine.
- **grocery\_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *grocery\_coupons.sav* con las compras semanales “acumuladas” para que cada caso corresponda a un cliente diferente. Algunas de las variables que cambiaban semanalmente



desaparecen de los resultados, y la cantidad gastada registrada se convierte ahora en la suma de las cantidades gastadas durante las cuatro semanas del estudio.

- **grocery\_coupons.sav.** Archivo de datos hipotéticos que contiene datos de encuestas recopilados por una cadena de tiendas de alimentación interesada en los hábitos de compra de sus clientes. Se sigue a cada cliente durante cuatro semanas, y cada caso corresponde a un cliente-semana distinto y registra información sobre dónde y cómo compran los clientes, incluida la cantidad que invierten en comestibles durante esa semana.
- **guttman.sav.** Bell presentó una tabla para ilustrar posibles grupos sociales. Guttman utilizó parte de esta tabla, en la que se cruzaron cinco variables que describían elementos como la interacción social, sentimientos de pertenencia a un grupo, proximidad física de los miembros y grado de formalización de la relación con siete grupos sociales teóricos, incluidos multitudes (por ejemplo, las personas que acuden a un partido de fútbol), espectadores (por ejemplo, las personas que acuden a un teatro o de una conferencia), públicos (por ejemplo, los lectores de periódicos o los espectadores de televisión), muchedumbres (como una multitud pero con una interacción mucho más intensa), grupos primarios (íntimos), grupos secundarios (voluntarios) y la comunidad moderna (confederación débil que resulta de la proximidad cercana física y de la necesidad de servicios especializados).
- **health\_funding.sav.** Archivo de datos hipotéticos que contiene datos sobre inversión en sanidad (cantidad por 100 personas), tasas de enfermedad (índice por 10.000 personas) y visitas a centros de salud (índice por 10.000 personas). Cada caso representa una ciudad diferente.
- **hivassay.sav.** Archivo de datos hipotéticos sobre las iniciativas de un laboratorio farmacéutico para desarrollar un ensayo rápido para detectar la infección por VIH. Los resultados del ensayo son ocho tonos de rojo con diferentes intensidades, donde los tonos más oscuros indican una mayor probabilidad de infección. Se llevó a cabo una prueba de laboratorio de 2.000 muestras de sangre, de las cuales una mitad estaba infectada con el VIH y la otra estaba limpia.
- **hourlywagedata.sav.** Archivo de datos hipotéticos sobre los salarios por horas de enfermeras de puestos de oficina y hospitales y con niveles distintos de experiencia.
- **insurance\_claims.sav.** Éste es un archivo de datos hipotéticos sobre una compañía de seguros que desea generar un modelo para etiquetar las reclamaciones sospechosas y potencialmente fraudulentas. Cada caso representa una reclamación diferente.
- **insure.sav.** Archivo de datos hipotéticos sobre una compañía de seguros que estudia los factores de riesgo que indican si un cliente tendrá que hacer una reclamación a lo largo de un contrato de seguro de vida de 10 años. Cada caso del archivo de datos representa un par de contratos (de los que uno registró una reclamación y el otro no), agrupados por edad y sexo.
- **judges.sav.** Archivo de datos hipotéticos sobre las puntuaciones concedidas por jueces cualificados (y un aficionado) a 300 actuaciones gimnásticas. Cada fila representa una actuación diferente; los jueces vieron las mismas actuaciones.
- **kinship\_dat.sav.** Rosenberg y Kim comenzaron a analizar 15 términos de parentesco [tía, hermano, primos, hija, padre, nieta, abuelo, abuela, nieto, madre, sobrino, sobrina, hermana, hijo, tío]. Le pidieron a cuatro grupos de estudiantes universitarios (dos masculinos y dos femeninos) que ordenaran estos grupos según las similitudes. A dos grupos (uno masculino y otro femenino) se les pidió que realizaran la ordenación dos veces, pero que la segunda ordenación la hicieran según criterios distintos a los de la primera. Así, se obtuvo un total de seis “fuentes“. Cada fuente se corresponde con una matriz de proximidades de  $15 \times 15$  cuyas

casillas son iguales al número de personas de una fuente menos el número de veces que se particionaron los objetos en esa fuente.

- **kinship\_ini.sav.** Este archivo de datos contiene una configuración inicial para una solución tridimensional de *kinship\_dat.sav*.
- **kinship\_var.sav.** Este archivo de datos contiene variables independientes *sexo*, *gener(ación)*, y *grado* (de separación) que se pueden usar para interpretar las dimensiones de una solución para *kinship\_dat.sav*. Concretamente, se pueden usar para restringir el espacio de la solución a una combinación lineal de estas variables.
- **marketvalues.sav.** Archivo de datos sobre las ventas de casas en una nueva urbanización de Algonquin, Ill., durante los años 1999 y 2000. Los datos de estas ventas son públicos.
- **nhis2000\_subset.sav.** La National Health Interview Survey (NHIS, encuesta del Centro Nacional de Estadísticas de Salud de EE.UU.) es una encuesta detallada realizada entre la población civil de Estados Unidos. Las encuestas se realizaron en persona a una muestra representativa de las unidades familiares del país. Se recogió tanto la información demográfica como las observaciones acerca del estado y los hábitos de salud de los integrantes de cada unidad familiar. Este archivo de datos contiene un subconjunto de información de la encuesta de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Archivo de datos y documentación de uso público. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/). Fecha de acceso: 2003.
- **ozono.sav.** Los datos incluyen 330 observaciones de seis variables meteorológicas para pronosticar la concentración de ozono a partir del resto de variables. Los investigadores anteriores, han encontrado que no hay linealidad entre estas variables, lo que dificulta los métodos de regresión típica.
- **pain\_medication.sav.** Este archivo de datos hipotéticos contiene los resultados de una prueba clínica sobre medicación antiinflamatoria para tratar el dolor artrítico crónico. Resulta de particular interés el tiempo que tarda el fármaco en hacer efecto y cómo se compara con una medicación existente.
- **patient\_los.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de pacientes que fueron admitidos en el hospital ante la posibilidad de sufrir un infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.
- **patlos\_sample.sav.** Este archivo de datos hipotéticos contiene los registros de tratamiento de una muestra de pacientes que recibieron trombolíticos durante el tratamiento del infarto de miocardio (IM o “ataque al corazón”). Cada caso corresponde a un paciente distinto y registra diversas variables relacionadas con su estancia hospitalaria.
- **poll\_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de los encuestadores para determinar el nivel de apoyo público a una ley antes de una asamblea legislativa. Los casos corresponden a votantes registrados. Cada caso registra el condado, la población y el vecindario en el que vive el votante.
- **poll\_cs\_sample.sav.** Este archivo de datos hipotéticos contiene una muestra de los votantes enumerados en *poll\_cs.sav*. La muestra se tomó según el diseño especificado en el archivo de plan *poll\_csplan* y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. Sin embargo, tenga en cuenta que debido a que el plan muestral hace uso de un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*poll\_jointprob.sav*). Las

variables adicionales que corresponden a los datos demográficos de los votantes y sus opiniones sobre la propuesta de ley se recopilaron y añadieron al archivo de datos después de tomar la muestra.

- **property\_assess.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor del condado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a las propiedades vendidas en el condado el año anterior. Cada caso del archivo de datos registra la población en que se encuentra la propiedad, el último asesor que visitó la propiedad, el tiempo transcurrido desde la última evaluación, la valoración realizada en ese momento y el valor de venta de la propiedad.
- **property\_assess\_cs.sav.** Archivo de datos hipotéticos sobre las iniciativas de un asesor de un estado para mantener actualizada la evaluación de los valores de las propiedades utilizando recursos limitados. Los casos corresponden a propiedades del estado. Cada caso del archivo de datos registra el condado, la población y el vecindario en el que se encuentra la propiedad, el tiempo transcurrido desde la última evaluación y la valoración realizada en ese momento.
- **property\_assess\_cs\_sample.sav** Este archivo de datos hipotéticos contiene una muestra de las propiedades recogidas en *property\_assess\_cs.sav*. La muestra se tomó en función del diseño especificado en el archivo de plan *property\_assess\_csplan*, y este archivo de datos registra las probabilidades de inclusión y las ponderaciones muestrales. La variable adicional *Valor actual* se recopiló y añadió al archivo de datos después de tomar la muestra.
- **recidivism.sav.** Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un infractor anterior y registra su información demográfica, algunos detalles de su primer delito y, a continuación, el tiempo transcurrido desde su segundo arresto, si ocurrió en los dos años posteriores al primer arresto.
- **recidivism\_cs\_sample.sav.** Archivo de datos hipotéticos sobre las iniciativas de una agencia de orden público para comprender los índices de reincidencia en su área de jurisdicción. Cada caso corresponde a un delincuente anterior, puesto en libertad tras su primer arresto durante el mes de junio de 2003 y registra su información demográfica, algunos detalles de su primer delito y los datos de su segundo arresto, si se produjo antes de finales de junio de 2006. Los delincuentes se seleccionaron de una muestra de departamentos según el plan de muestreo especificado en *recidivism\_cs\_csplan*. Como este plan utiliza un método de probabilidad proporcional al tamaño (PPS), también existe un archivo que contiene las probabilidades de selección conjunta (*recidivism\_cs\_jointprob.sav*).
- **rfm\_transactions.sav.** Archivo de datos hipotéticos que contiene datos de transacciones de compra, incluida la fecha de compra, los artículos adquiridos y el importe de cada transacción.
- **salesperformance.sav.** Archivo de datos hipotéticos sobre la evaluación de dos nuevos cursos de formación de ventas. Sesenta empleados, divididos en tres grupos, reciben formación estándar. Además, el grupo 2 recibe formación técnica; el grupo 3, un tutorial práctico. Cada empleado se sometió a un examen al final del curso de formación y se registró su puntuación. Cada caso del archivo de datos representa a un alumno distinto y registra el grupo al que fue asignado y la puntuación que obtuvo en el examen.
- **satisf.sav.** Archivo de datos hipotéticos sobre una encuesta de satisfacción llevada a cabo por una empresa minorista en cuatro tiendas. Se encuestó a 582 clientes en total y cada caso representa las respuestas de un único cliente.

- **screws.sav** Este archivo de datos contiene información acerca de las características de tornillos, pernos, clavos y tacos .
- **shampoo\_ph.sav**. Archivo de datos hipotéticos sobre el control de calidad en una fábrica de productos para el cabello. Se midieron seis lotes de resultados distintos en intervalos regulares y se registró su pH. El intervalo objetivo es de 4,5 a 5,5.
- **ships.sav**. Un conjunto de datos presentados y analizados en otro lugar sobre los daños en los cargueros producidos por las olas. Los recuentos de incidentes se pueden modelar como si ocurrieran con una tasa de Poisson dado el tipo de barco, el período de construcción y el período de servicio. Los meses de servicio agregados para cada casilla de la tabla formados por la clasificación cruzada de factores proporcionan valores para la exposición al riesgo.
- **site.sav**. Archivo de datos hipotéticos sobre las iniciativas de una compañía para seleccionar sitios nuevos para sus negocios en expansión. Se ha contratado a dos consultores para evaluar los sitios de forma independiente, quienes, además de un informe completo, han resumido cada sitio como una posibilidad “buena”, “media” o “baja”.
- **smokers.sav**. Este archivo de datos es un resumen de la encuesta sobre toxicomanía 1998 National Household Survey of Drug Abuse y es una muestra de probabilidad de unidades familiares americanas. (<http://dx.doi.org/10.3886/ICPSR02934>) Así, el primer paso de un análisis de este archivo de datos debe ser ponderar los datos para reflejar las tendencias de población.
- **stocks.sav** Este archivo de datos hipotéticos contiene precios de acciones y volumen de un año.
- **stroke\_clean.sav**. Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haberla limpiado mediante los procedimientos de la opción Preparación de datos.
- **stroke\_invalid.sav**. Este archivo de datos hipotéticos contiene el estado inicial de una base de datos médica que incluye contiene varios errores de entrada de datos.
- **stroke\_survival**. Este archivo de datos hipotéticos registra los tiempos de supervivencia de los pacientes que finalizan un programa de rehabilitación tras un ataque isquémico. Tras el ataque, la ocurrencia de infarto de miocardio, ataque isquémico o ataque hemorrágico se anotan junto con el momento en el que se produce el evento registrado. La muestra está truncada a la izquierda ya que únicamente incluye a los pacientes que han sobrevivido al final del programa de rehabilitación administrado tras el ataque.
- **stroke\_valid.sav**. Este archivo de datos hipotéticos contiene el estado de una base de datos médica después de haber comprobado los valores mediante el procedimiento Validar datos. Sigue conteniendo casos potencialmente anómalos.
- **survey\_sample.sav**. Este archivo de datos contiene datos de encuestas, incluyendo datos demográficos y diferentes medidas de actitud. Se basa en un subconjunto de variables de NORC General Social Survey de 1998, aunque algunos valores de datos se han modificado y que existen variables ficticias adicionales se han añadido para demostraciones.
- **telco.sav**. Archivo de datos hipotéticos sobre las iniciativas de una compañía de telecomunicaciones para reducir el abandono de clientes en su base de clientes. Cada caso corresponde a un cliente distinto y registra diversa información demográfica y de uso del servicio.

- **telco\_extra.sav.** Este archivo de datos es similar al archivo de datos *telco.sav*, pero las variables de meses con servicio y gasto de clientes transformadas logarítmicamente se han eliminado y sustituido por variables de gasto del cliente transformadas logarítmicamente tipificadas.
- **telco\_missing.sav.** Este archivo de datos es un subconjunto del archivo de datos *telco.sav*, pero algunos valores de datos demográficos se han sustituido con valores perdidos.
- **testmarket.sav.** Archivo de datos hipotéticos sobre los planes de una cadena de comida rápida para añadir un nuevo artículo a su menú. Hay tres campañas posibles para promocionar el nuevo producto, por lo que el artículo se presenta en ubicaciones de varios mercados seleccionados aleatoriamente. Se utiliza una promoción diferente en cada ubicación y se registran las ventas semanales del nuevo artículo durante las primeras cuatro semanas. Cada caso corresponde a una ubicación semanal diferente.
- **testmarket\_1month.sav.** Este archivo de datos hipotéticos es el archivo de datos *testmarket.sav* con las ventas semanales “acumuladas” para que cada caso corresponda a una ubicación diferente. Como resultado, algunas de las variables que cambiaban semanalmente desaparecen y las ventas registradas se convierten en la suma de las ventas realizadas durante las cuatro semanas del estudio.
- **tree\_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.
- **tree\_credit.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios.
- **tree\_missing\_data.sav** Archivo de datos hipotéticos que contiene datos demográficos y de historial de créditos bancarios con un elevado número de valores perdidos.
- **tree\_score\_car.sav.** Archivo de datos hipotéticos que contiene datos demográficos y de precios de compra de vehículos.
- **tree\_textdata.sav.** Archivo de datos sencillos con dos variables diseñadas principalmente para mostrar el estado por defecto de las variables antes de realizar la asignación de nivel de medida y etiquetas de valor.
- **tv-survey.sav.** Archivo de datos hipotéticos sobre una encuesta dirigida por un estudio de TV que está considerando la posibilidad de ampliar la emisión de un programa de éxito. Se preguntó a 906 encuestados si verían el programa en distintas condiciones. Cada fila representa un encuestado diferente; cada columna es una condición diferente.
- **ulcer\_recurrence.sav.** Este archivo contiene información parcial de un estudio diseñado para comparar la eficacia de dos tratamientos para prevenir la reaparición de úlceras. Constituye un buen ejemplo de datos censurados por intervalos y se ha presentado y analizado en otro lugar .
- **ulcer\_recurrence\_recoded.sav.** Este archivo reorganiza la información de *ulcer\_recurrence.sav* para permitir modelar la probabilidad de eventos de cada intervalo del estudio en lugar de sólo la probabilidad de eventos al final del estudio. Se ha presentado y analizado en otro lugar .
- **verd1985.sav.** Archivo de datos sobre una encuesta . Se han registrado las respuestas de 15 sujetos a 8 variables. Se han dividido las variables de interés en tres grupos. El conjunto 1 incluye *edad* y *ecivil*, el conjunto 2 incluye *mascota* y *noticia*, mientras que el conjunto 3 incluye *música* y *vivir*. Se escala *mascota* como nominal múltiple y *edad* como ordinal; el resto de variables se escalan como nominal simple.

- **virus.sav.** Archivo de datos hipotéticos sobre las iniciativas de un proveedor de servicios de Internet (ISP) para determinar los efectos de un virus en sus redes. Se ha realizado un seguimiento (aproximado) del porcentaje de tráfico de correos electrónicos infectados en sus redes a lo largo del tiempo, desde el momento en que se descubre hasta que la amenaza se contiene.
- **wheeze\_steubenville.sav.** Subconjunto de un estudio longitudinal de los efectos sobre la salud de la polución del aire en los niños . Los datos contienen medidas binarias repetidas del estado de las sibilancias en niños de Steubenville, Ohio, con edades de 7, 8, 9 y 10 años, junto con un registro fijo de si la madre era fumadora durante el primer año del estudio.
- **workprog.sav.** Archivo de datos hipotéticos sobre un programa de obras del gobierno que intenta colocar a personas desfavorecidas en mejores trabajos. Se siguió una muestra de participantes potenciales del programa, algunos de los cuales se seleccionaron aleatoriamente para entrar en el programa, mientras que otros no siguieron esta selección aleatoria. Cada caso representa un participante del programa diferente.
- **worldsales.sav** Este archivo de datos hipotéticos contiene ingresos por ventas por continente y producto.

# Avisos

Esta información se ha desarrollado para los productos y servicios ofrecidos en todo el mundo.

Puede que IBM no ofrezca los productos, los servicios o las características de los que se habla en este documento en otros países. Consulte a su representante local de IBM para obtener información acerca de los productos y servicios que está disponibles actualmente en su zona. Toda referencia que se haga de un producto, programa o servicio de IBM no implica que sólo se deba utilizar ese producto, programa o servicio de IBM. En su lugar, puede utilizarse todo producto, programa o servicio con funcionalidades equivalentes que no infrinjan los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o aplicaciones de patentes pendientes que cubren el asunto descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, Estados Unidos*

Si tiene alguna pregunta sobre la licencia relacionada con la información del juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe sus preguntas por escrito a:

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

**El párrafo siguiente no se aplica a los Reino Unido o cualquier otro país donde tales disposiciones son incompatibles con la legislación local:** INTERNATIONAL BUSINESS MACHINES PROPORCIONA ESTA PUBLICACIÓN “TAL CUAL” SIN GARANTÍA DE NINGÚN TIPO, YA SEA EXPRESA O IMPLÍCITA, INCLUYENDO, PERO NO LIMITADA A, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN PROPÓSITO PARTICULAR. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones , por lo tanto , esta declaración no se aplique a usted.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente, se efectúan cambios en la información aquí y estos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en los productos y/o los programas descritos en esta publicación en cualquier momento sin previo aviso.

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales en esos sitios Web no forman parte del material de este producto de IBM y el uso de estos sitios Web es bajo su propio riesgo.

IBM puede utilizar cualquier información que le suministre en cualquier forma que considere adecuada, sin incurrir en ninguna obligación para usted.

Los licenciatarios de este programa que deseen tener información sobre el mismo con el objetivo de habilitar: (i) el intercambio de información entre programas creados independientemente y otros programas (incluyendo este) y (ii) el uso común de la información que se ha intercambiado, deben ponerse en contacto con:

*IBM Software Group, a la atención de: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.*

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

IBM proporciona el programa bajo licencia que se describe en este documento y todo el material bajo licencia disponible para el mismo bajo los términos de IBM Customer Agreement (Acuerdo de cliente de IBM), IBM International Program License Agreement (Acuerdo de licencia de programa internacional de IBM) o cualquier acuerdo equivalente entre las partes.

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otras fuentes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión de su rendimiento, su compatibilidad o cualquier otra reclamación relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

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

Si está viendo esta información en copia electrónica, es posible que las fotografías y las ilustraciones en color no aparezcan.

### ***Marcas registradas***

IBM, el logotipo de IBM, [ibm.com](http://www.ibm.com) y SPSS son marcas comerciales de IBM Corporation, registradas en muchas jurisdicciones de todo el mundo. Existe una lista actualizada de marcas comerciales de IBM en Internet en <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, el logotipo Adobe, PostScript y el logotipo PostScript son marcas registradas o marcas comerciales de Adobe Systems Incorporated en Estados Unidos y/o otros países.

Intel, el logotipo de Intel, Intel Inside, el logotipo de Intel Inside, Intel Centrino, el logotipo de Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium y Pentium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en Estados Unidos y otros países.

Java y todas las marcas comerciales y los logotipos basados en Java son marcas comerciales de Sun Microsystems, Inc. en Estados Unidos, otros países o ambos.

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

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

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

Este producto utiliza WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.



Otros productos y nombres de servicio pueden ser marcas comerciales de IBM u otras empresas.

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

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



---

# Índice

análisis de conglomerados, 15, 53  
archivos de ejemplo  
  ubicación, 99  
avisos legales, 109

conglomerado, 15

índices de respuesta de código postal, 26, 77

marcas registradas, 110

perfiles de posibles clientes, 20  
Perfiles de posibles clientes, 70  
propensión a la compra, 34, 84  
prueba de paquete de control, 42, 96

regresión logística , 34, 84  
RFM, 2, 9–10, 12, 46  
  agrupación, 6  
  datos de cliente, 5  
  datos de transacciones, 3, 46