

IBM SPSS Custom Tables 22

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos" en la página 103.

Información sobre el producto

Esta edición se aplica a la versión 22, release 0, modificación 0 de IBM SPSS Statistics y a todos los releases y modificaciones posteriores hasta que se indique lo contrario en nuevas ediciones.

Contenido

Capítulo 1. Interfaz de generador de tablas. 1

Interfaz de generador de tablas	1
Generación de tablas.	1
Generación de una tabla	4
Apilado de variables.	4
Anidación de variables	4
Capas.	5
Cómo mostrar y ocultar nombres y/o etiquetas de variables.	5
Estadísticos de resumen	6
Categorías y totales.	11
Categorías calculadas	14
Tablas de variables con categorías compartidas (Tablas Comperimeter)	15
Personalización del generador de tablas	15
Tablas personalizadas: pestaña Opciones.	15
Tablas personalizadas: pestaña Títulos	16
Tablas personalizadas: pestaña Estadísticos de contraste	17

Capítulo 2. Tablas simples para variables categóricas 19

Tablas simples para variables categóricas	19
Una variable categórica individual.	19
Porcentajes	20
Totales	21
Tabulación cruzada.	21
Porcentajes en las tabulaciones cruzadas.	22
Control del formato de presentación	22
Totales marginales	23
Ordenación y exclusión de categorías.	24

Capítulo 3. Apilado, anidación y capas con variables categóricas 27

Variables categóricas de apilado	27
Apilado con tablas cruzadas.	27
Anidación de variables categóricas	28
Eliminación de las etiquetas de variable	29
Tabulación cruzada anidada	30
Capas	31
Dos variables de capa categóricas apiladas	32
Dos variables de capa categóricas anidadas.	32

Capítulo 4. Totales y subtotales de variables categóricas 35

Un total único para una sola variable.	35
Lo que ve es lo que se totaliza	35
Posición de presentación de los totales	36
Totales de tablas anidadas	36
Totales de la variable de capas	38
Subtotales	38
Lo que ve es lo que se subtotaliza	39
Ocultación de categorías de subtotal	39

Subtotales de la variable de capas	40
--	----

Capítulo 5. Categorías calculadas de variables categóricas 41

Categoría calculada simple	41
Ocultación de categorías en una categoría calculada	42
Referencia de subtotales en una categoría calculada	42
Uso de categorías calculadas para mostrar	
Subtotales no exhaustivos	44

Capítulo 6. Tablas para variables con categorías compartidas 47

Tabla de recuentos	47
Tabla de porcentajes	48
Control de categorías y totales	49
Anidación en tablas con categorías compartidas	50

Capítulo 7. Estadísticos de resumen . . . 51

Variable origen de los estadísticos de resumen.	51
Origen de los estadísticos de resumen para variables categóricas	52
Origen de estadísticos de resumen para variables de escala	52
Variables apiladas	53
Estadísticos de resumen del total personalizados para variables categóricas.	54
Presentación de los valores de categoría	55

Capítulo 8. Resumen de variables de escala 57

Resumen de variables de escala	57
Variables de escala apiladas	57
Múltiples estadísticos de resumen	58
Recuento, N válido y Valores perdidos	58
Diferentes resúmenes para distintas variables	59
Resúmenes por grupos en Categorías.	60
Variables de agrupación múltiples.	61
Anidación de variables categóricas dentro de las variables de escala	61

Capítulo 9. Estadísticos de contraste 63

Estadísticos de contraste	63
Comprobaciones de independencia (chi-cuadrado)	63
Efectos de anidación y apilado en comprobaciones de independencia.	64
Comparación de medias de columna	66
Efectos de anidación y apilado en comprobaciones de medias de columna	67
Comparación de proporciones de columna	69
Efectos de anidación y apilado en comprobaciones de proporciones de columna	72
Nota sobre ponderaciones y conjuntos de respuestas múltiples	73

Capítulo 10. Conjuntos de respuestas múltiples 75

Recuentos, respuestas, porcentajes y totales 75

Utilización de conjuntos de respuestas múltiples con otras variables 77

Variable origen de los estadísticos y estadísticos de resumen disponibles 78

Conjuntos de categorías múltiples y respuestas duplicadas. 78

Pruebas de significación con conjuntos de respuestas múltiples 80

Pruebas de independencia con Conjuntos de respuestas múltiples 80

Comparación de las medias de las columnas con conjuntos de respuestas múltiples 81

Capítulo 11. Valores perdidos 83

Tablas sin valores perdidos 83

Inclusión de valores perdidos en las tablas 84

Capítulo 12. Asignación de formato y personalización de tablas 87

Asignación de formato y personalización de tablas 87

Formato de presentación de los estadísticos de resumen 87

Etiquetas de presentación para los estadísticos de resumen 88

Ancho de columna 89

Valor de presentación para las casillas vacías . . . 90

Valor de presentación para los estadísticos perdidos 90

Capítulo 13. Archivos muestrales 93

Avisos 103

Marcas comerciales 105

Índice. 107

Capítulo 1. Interfaz de generador de tablas

Interfaz de generador de tablas

Tablas personalizadas utiliza una interfaz de generador de tablas mediante las acciones simples de arrastrar y soltar que le permite obtener una vista previa de su tabla cuando selecciona variables y opciones. También ofrece un nivel de flexibilidad que no se encuentra en un cuadro de diálogo típico, incluida la capacidad de cambiar el tamaño de la ventana y de los paneles de la ventana.

Generación de tablas



Figura 1. Cuadro de diálogo Tablas personalizadas, pestaña Tabla

Seleccione las variables y medidas de resumen que aparecerán en las tablas de la pestaña Tabla del generador de tablas.

Lista de variables. Las variables del archivo de datos se muestran en el panel superior izquierdo de la ventana. Tablas personalizadas distingue entre dos niveles de medición diferentes para las variables y las trata de forma diferente en función del nivel de medición:

Catógicas. Datos con un número limitado de valores o categorías distintas (por ejemplo, sexo o religión). Las variables categóricas pueden ser variables de cadena (alfanuméricas) o variables numéricas que utilizan códigos numéricos para representar a categorías (por ejemplo, 0 = *hombre* y 1 = *mujer*). También se hace referencia a estos datos como datos cualitativos. Las variables categóricas pueden ser **nominales** u **ordinales**.

- *Nominal.* Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen a una clasificación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado. Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.

- *Ordinal*. Una variable puede ser tratada como ordinal cuando sus valores representan categorías con alguna clasificación intrínseca. Por ejemplo, los niveles de satisfacción con un servicio, que abarquen desde muy insatisfecho hasta muy satisfecho. Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.

Escalas. Datos medidos en una escala de intervalo o de razón en los que los valores de los datos indican el orden de los valores y la distancia entre ellos. Por ejemplo, un salario de 72.195€ es superior a un salario de 52.398€ y la distancia entre ambos valores es 19.797€. También se hace referencia a estos datos como datos cuantitativos o continuos.

Las variables categóricas definen las categorías (filas, columnas y niveles) en la tabla y el estadístico de resumen predeterminado es el recuento (número de casos en cada categoría). Por ejemplo, una tabla predeterminada de una variable de género categórica simplemente mostrará el número de hombres y el de mujeres.

Las variables de escala se resumen normalmente dentro de categorías de variables categóricas y el estadístico de resumen predeterminado es la media. Por ejemplo, una tabla predeterminada de ingresos dentro de categorías de género mostrará los ingresos medios de hombres y los ingresos medios de mujeres.

También puede resumir las variables de escala sin utilizar una variable categórica para definir grupos. Es principalmente útil para los resúmenes de **apilado** de las variables de escala múltiple. Consulte el tema "Apilado de variables" en la página 4 para obtener más información.

Conjuntos de respuestas múltiples

Tablas personalizadas también admite un tipo especial de "variable" llamado **conjunto de respuestas múltiples**. Los conjuntos de respuestas múltiples no son variables realmente en el sentido normal. No puede verlos en el Editor de datos y otros procedimientos no los reconocen. Los conjuntos de respuestas múltiples utilizan variables para registrar respuestas a preguntas donde el encuestado puede ofrecer más de una respuesta. Los conjuntos de respuestas múltiples se tratan como variables categóricas y la mayoría de las acciones que puede realizar con las variables categóricas, también las puede realizar con los conjuntos de respuestas múltiples. Consulte el tema Capítulo 10, "Conjuntos de respuestas múltiples", en la página 75 para obtener más información.

Un icono junto a cada variable en la lista de variables identifica el tipo de variable.


Tabla 1. Iconos de nivel de medición

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

Tabla 2. Iconos de conjuntos de respuestas múltiples

Tipo de conjunto de respuestas múltiples	Icono
Conjunto de respuestas múltiples, categorías múltiples	

Tabla 2. Iconos de conjuntos de respuestas múltiples (continuación)

Tipo de conjunto de respuestas múltiples	Icono
Conjunto de respuestas múltiples, dicotomías múltiples	

Puede cambiar el nivel de medición de una variable en el generador de tablas pulsando con el botón derecho del ratón en la lista de variable y seleccionando **Categoría** o **Escala** del menú emergente. Puede cambiar de forma permanente el nivel de medición de una variable en la Vista de variable del Editor de datos. Las variables definidas como **nominal** u **ordinal** se tratan como categóricas en Tablas personalizadas.

Categorías. Cuando selecciona una variable categórica en la lista de variables, las categorías definidas por la variable se muestran en la lista Categorías. Estas categorías también se mostrarán en el panel de lienzo cuando utilice la variable en una tabla. Si la variable no tiene categorías definidas, la lista Categorías y el panel del lienzo mostrará dos categorías de marcador: *Categoría 1* y *Categoría 2*.

Las categorías definidas que se muestran en el generador de tablas se basan en **etiquetas de valores**, etiquetas descriptivas asignadas a valores de datos diferentes (por ejemplo, valores numéricos de 0 y 1, con etiquetas de valores *hombre* y *mujer*). Puede definir las etiquetas de valores en Vista de variable del Editor de datos o con Definir las propiedades de la variable en el menú Datos de la ventana del Editor de datos.

Panel de lienzo. Genera una tabla arrastrando y soltando las variables en las filas y columnas del panel de lienzo. El panel de lienzo muestra una vista previa de la tabla que se creará. El panel de lienzo no muestra los valores de datos reales en las casillas, pero debe proporcionar una vista bastante precisa del diseño de la tabla final. Para variables categóricas, la tabla real puede contener más categorías que la vista previa si el archivo de datos contiene valores exclusivos para los que no se han definido etiquetas de valores.

- La vista **Normal** muestra todas las filas y columnas que se mostrarán en la tabla, incluidas las filas y/o columnas de los estadísticos de resumen y las categorías de variables categóricas.
- La vista **Compacta** muestra sólo las variables que se incluirán en la tabla, sin una vista previa de las filas y columnas que contendrá la tabla.

Reglas básicas y limitaciones para la generación de una tabla

- Para las variables categóricas, los estadísticos de resumen se basan en la variable más interior de la dimensión de origen de estadísticos.
- La dimensión predeterminada de origen de estadísticos (fila o columna) para las variables categóricas se basa en el orden en el que arrastre y suelte las variables en el panel de lienzo. Por ejemplo, si primero arrastra una variable a la bandeja de filas, la dimensión de la fila es la dimensión predeterminada de origen de estadísticos.
- Las variables de escala se pueden resumir sólo dentro de las categorías de la variable más interior en la dimensión de la fila o la columna. (Puede colocar la variable de escala a cualquier nivel de la tabla, pero se resume al nivel más interior.)
- Las variables de escala no se pueden resumir dentro de otras variables de escala. Puede apilar resúmenes de variables de escala múltiple o resumir variables de escala dentro de categorías de variables categóricas. No puede anidar una variable de escala dentro de otra ni poner una variable de escala en la dimensión de la fila y otra variable de escala en la dimensión de la columna.
- Si cualquier variable del conjunto de datos activo contiene más de 12.000 etiquetas de valores definidos, no puede utilizar el generador de tablas para crear tablas. Si no necesita incluir variables que excedan este límite en sus tablas, puede definir y aplicar conjuntos de variables que excluyan dichas variables. Si necesita incluir cualquier variable con más de 12.000 etiquetas de valores definidos, no puede utilizar la sintaxis de comandos CTABLES para generar tablas.

Generación de una tabla

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Arrastre y suelte una o más variables en las áreas de fila y/o columna del panel de lienzo.
3. Pulse en **Aceptar** para crear la tabla.
4. Seleccione (pulse en) la variable en el panel de lienzo.
5. Arrastre la variable a cualquier lugar fuera del panel de lienzo o pulse la tecla Supr.
Para cambiar el nivel de medición de una variable:
6. Pulse con el botón derecho del ratón en la lista de variable (puede hacerlo sólo en la lista de variables, no en el lienzo).
7. Seleccione **Categorica** o **Escala** en el menú emergente.

Apilado de variables

El apilado se puede considerar como el procedimiento por el cual se toman tablas separadas y se pegan en la misma visualización. Por ejemplo, puede mostrar información en la categoría *Género* y *Edad* en secciones separadas de la misma tabla.

Apilado de variables

1. En la lista de variables, seleccione todas las variables que desea apilar, a continuación, arrástrelas y suéltelas en las filas o columnas del panel de lienzo.
o
2. Arrastre y suelte las variables de forma separada, soltando cada variable encima o debajo de las variables existentes en las filas o a la derecha o izquierda de las variables existentes en las columnas.

Tabla 3. Variables categóricas apiladas.

Variabes	Categorías	Estadísticos de resumen
Variable 1	Categoría 1	123
	Categoría 2	456
Variable 2	Categoría 1	123
	Categoría 2	456
	Categoría 3	789

Consulte el tema “Variables categóricas de apilado” en la página 27 para obtener más información.

Anidación de variables

La anidación, como tabulación de contingencia, puede mostrar la relación entre dos variables categóricas, excepto una variable anidada dentro de otra en la misma dimensión. Por ejemplo, puede anidar *Género* dentro de la categoría *Edad* en la dimensión de la fila, mostrando el número de hombres y mujeres de cada categoría de edad.

También puede anidar una variable de escala dentro de una variable categórica. Por ejemplo, podría anidar *Ingresos* dentro de *Género*, mostrando medias separadas (o medianas u otras medidas de resumen) de los valores de ingresos para hombres y mujeres.

Anidación de variables

1. Arrastre y suelte una variable categórica en el área de fila y/o columna del panel de lienzo.
2. Arrastre y suelte una variable categórica o de escala a la izquierda o derecha de la variable de la fila categórica o encima o debajo de la variable de columna categórica.

Tabla 4. Variables categóricas anidadas.

Variable 1	Variable 2	Estadísticos de resumen
Categoría 1	Categoría 1	12
	Categoría 2	34
	Categoría 3	56
Categoría 2	Categoría 1	12
	Categoría 2	34
	Categoría 3	56

Consulte el tema “Anidación de variables categóricas” en la página 28 para obtener más información.

Nota: las tablas personalizadas no tienen en cuenta el procesamiento del archivo segmentado en capas. Para lograr los mismos resultado que los archivos segmentados en capas, coloque las variables de archivo segmentado en las capas más exteriores de la tabla.

Capas

Puede utilizar capas para añadir una dimensión de profundidad a sus tablas, creando "cubos" tridimensionales. Las capas son similares a la anidación o el apilado; la diferencia principal es que sólo se puede mostrar una categoría cada vez. Por ejemplo, si utiliza la categoría *Edad* como la variable de la fila y *Género* como la variable de capa obtiene una tabla en la que se muestra la información para hombres y mujeres en capas diferentes de la tabla.

Creación de capas

1. Pulse en **Capas** en la pestaña Tabla del generador de tablas para mostrar la lista Capas.
2. Arrastre y suelte la escala o las variables categóricas que definirán las capas en la lista Capas.

No puede mezclar variables de escala y categóricas en la lista Capas. Todas las variables deben ser del mismo tipo. Los conjuntos de respuestas múltiples se tratan como categóricos en la lista Capas. Las variables de escala en las capas se apilan siempre.

Si tiene múltiples variables de capas categóricas, las capas se pueden apilar o anidar.

- **Mostrar cada categoría como una capa** es equivalente al apilado. Se mostrará una capa separada para cada categoría de cada variable de capa. El número total de capas es simplemente la *suma* del número de categorías para cada variable de capa. Por ejemplo, si tiene tres variables de capas, cada una con tres categorías, la tabla tendrá nueve capas.
- **Mostrar cada combinación de categorías como una capa** es equivalente a la anidación o tabulación cruzada de las capas. El número total de capas es el *producto* del número de categorías para cada variable de capa. Por ejemplo, si tiene tres variables, cada una con tres categorías, la tabla tendrá 27 capas.

Cómo mostrar y ocultar nombres y/o etiquetas de variables

Las siguientes opciones están disponibles para la visualización de los nombres y etiquetas de variables:

- Mostrar sólo etiquetas de variables. Para cualquier variable sin etiquetas de variables definidas, se mostrará el nombre de la variable. Ésta es el ajuste predeterminado.
- Mostrar sólo nombres de variables.
- Mostrar etiquetas de variables y nombres de variables.
- No mostrar nombres de variables ni etiquetas de variables. Aunque la columna/fila que contenga la etiqueta de variable o el nombre seguirá mostrándose en la vista previa del panel de lienzo, esta columna/fila no se mostrará en la tabla real.

Para mostrar u ocultar etiquetas de variables o nombres de variables:

1. Pulse con el botón derecho en la variable de la vista previa del panel de lienzo.
2. Seleccione **Mostrar etiqueta de variables** o **Mostrar nombre de variables** desde el menú emergente para activar o desactivar la visualización de etiquetas o nombres. Una marca de verificación junto a la selección indica que se mostrará.

Estadísticos de resumen

El cuadro de diálogo Estadísticos de resumen permite:

- Añadir y eliminar estadísticos de resumen de una tabla.
- Cambiar las etiquetas para los estadísticos.
- Cambiar el orden de los estadísticos.
- Cambiar el formato de los estadísticos, incluido el número de decimales.

Los estadísticos de resumen (y otras opciones) disponibles dependen del nivel de medición de la variable de origen de estadísticos de resumen, como se muestra en la parte superior del cuadro de diálogo. El origen de los estadísticos de resumen (la variable en la que se basan los estadísticos de resumen) se determina por:

- **Nivel de medición.** Si una tabla (o una sección de la tabla apilada) contiene una variable de escala, los estadísticos de resumen se basan en la variable de escala.
- **Orden de selección de variable.** La dimensión predeterminada de origen de estadísticos (fila o columna) para las variables categóricas se basa en el orden en el que arrastre y suelte las variables en el panel de lienzo. Por ejemplo, si primero arrastra una variable al área de filas, la dimensión de la fila es la dimensión predeterminada de origen de estadísticos.
- **Anidación.** Para las variables categóricas, los estadísticos de resumen se basan en la variable más interior de la dimensión de origen de estadísticos.

Una tabla anidada puede tener variables de origen de estadísticos de resumen múltiples (de escala y categóricas), pero cada sección de tabla tiene un único origen de estadísticos de resumen.

Cambio de la dimensión de origen de estadísticos de resumen

1. Seleccione la dimensión (filas, columnas o capas) desde la lista desplegable **Origen** en el grupo Estadísticos de resumen de la pestaña Tabla.

Control de los estadísticos de resumen mostrados en una tabla

1. Seleccione (pulse en) la variable de origen de estadísticos de resumen en el panel de lienzo de la pestaña Tabla.
2. En el grupo Definir de la pestaña Tabla, pulse en **Estadísticos de resumen**.
o
3. Pulse con el botón derecho del ratón en la variable de origen de estadísticos de resumen en el panel de lienzo y seleccione **Estadísticos de resumen** desde el menú emergente.
4. Seleccione los estadísticos de resumen que desea incluir en la tabla. Puede utilizar la flecha para mover los estadísticos seleccionados desde la lista Estadísticos a la lista Representación o puede arrastrar y soltar los estadísticos seleccionados desde la lista Estadísticos a la lista Representación.
5. Pulse en las flechas hacia arriba o hacia abajo para cambiar la posición de representación del estadístico de resumen seleccionado actualmente.
6. Seleccione un formato de presentación desde la lista desplegable Formato para el estadístico de resumen seleccionado.
7. Introduzca el número de decimales que desea mostrar en la casilla Decimales para el estadístico de resumen seleccionado.

8. Pulse en **Aplicar a la selección** para incluir los estadísticos de resumen seleccionados para las variables seleccionadas actualmente en el panel del lienzo.
9. Pulse en **Aplicar a todos** para incluir los estadísticos de resumen seleccionados para las variables seleccionadas actualmente en el panel del lienzo.

Nota: **Aplicar a todos** se diferencia de **Aplicar a la selección** sólo en las variables apiladas del mismo tipo que ya se incluyen en el panel del lienzo. En ambos casos, los estadísticos de resumen seleccionados se incluyen automáticamente para las variables apiladas adicionales del mismo tipo que haya añadido en la tabla.

Estadísticos de resumen para variables categóricas

Los estadísticos básicos disponibles para las variables categóricas son recuentos y porcentajes. También puede especificar estadísticos de resumen personalizados para los totales y subtotales. Estos estadísticos de resumen personalizados incluyen medidas de tendencia central (como media y mediana) y dispersión (como la desviación estándar) que puede ser adecuada para algunas variables categóricas ordinales. Consulte el tema “Estadísticos de resumen del total personalizados para variables categóricas” en la página 10 para obtener más información.

Recuento. Número de casos en cada casilla de la tabla o número de respuestas para conjuntos de respuestas múltiples.

Recuento no ponderado. Número de casos no ponderado en cada casilla de la tabla. Esto es distinto del recuento sólo si se realiza la ponderación

Porcentajes de Columna. Porcentajes dentro de cada columna. Los porcentajes de cada columna de una subtabla (para porcentajes simples) suman 100 %. Los porcentajes de columna son normalmente útiles sólo si tienen una variable de *fila* categórica.

Porcentajes de fila. Porcentajes dentro de cada fila. Los porcentajes de cada fila de una subtabla (para porcentajes simples) suman 100 %. Los porcentajes de fila son normalmente útiles sólo si tienen una variable de *columna* categórica.

Porcentajes de Fila de capa y Columna de capa. Los porcentajes de fila o columna (para porcentajes simples) suman 100 % en todas las subtablas de una tabla anidada. Si la tabla contiene capas, los porcentajes de fila o columna suman 100 % en todas las subtablas anidadas de cada capa.

Porcentajes de Capa. Porcentajes dentro de cada capa. En los porcentajes simples, los porcentajes de las casillas dentro de la capa visible actual suman 100 %. Si no tiene variables de capa, es equivalente a los porcentajes de tabla.

Porcentajes de Tabla. Los porcentajes de cada casilla se basan en toda la tabla. Todos los porcentajes de casillas se basan en el mismo número total de casos y suman 100 % (para porcentajes simples) de toda la tabla.

Porcentajes Subtabla. Los porcentajes de cada casilla se basan en la subtabla. Todos los porcentajes de casillas en la subtabla se basan en el mismo número total de casos y suman 100 % dentro de la subtabla. En las tablas anidadas, la variable que precede al nivel de anidación más interior define las subtablas. Por ejemplo, en una tabla de *Estado civil* dentro de *Género* dentro de la categoría *Edad*, *Género* define las subtablas.

Los conjuntos de respuestas múltiples pueden tener los porcentajes basados en casos, respuestas o recuentos. Consulte el tema “Estadísticos de resumen para conjuntos de respuestas múltiples” en la página 8 para obtener más información.

Tablas apiladas

Para los cálculos de porcentajes, cada sección de tabla definida por una variable de apilado se trata como una tabla separada. Los porcentajes de Fila de capa, Columna de capa y Tabla suman 100 % (para porcentajes simples) dentro de cada sección de tabla apilada. La base porcentual para cálculos de porcentajes diferentes se basa en los casos de cada sección de tabla apilada.

Base porcentual

Los porcentajes se pueden calcular de tres formas diferentes, determinadas por el tratamiento de los valores perdidos en la base computacional:

Porcentaje Simple. Los porcentajes se basan en el número de casos utilizado en la tabla y siempre suman 100 %. Si una categoría se excluye de la tabla, los casos en dicha categoría se excluyen de la base. Los casos con los valores perdidos del sistema siempre se excluyen de la base. Los casos con los valores perdidos del usuario se excluyen si las categorías perdidas del usuario se excluyen de la tabla (el valor predeterminado) o se incluyen si las categorías perdidas del usuario se han incluido en la tabla. Cualquier porcentaje que no contenga *N válido* o *N total* en su nombre es un porcentaje simple.

Porcentaje de N total. Los casos con los valores perdidos del sistema y del usuario se añaden a la base porcentual simple. Los porcentajes pueden sumar menos de 100 %.

Porcentaje de N válido. Los casos con los valores perdidos del usuario se eliminan de la base porcentual simple incluso si se incluyen en la tabla las categorías perdidas del usuario.

Nota: de la base siempre se excluyen los casos de categorías excluidas manualmente que no sean categorías perdidas del usuario.

Estadísticos de resumen para conjuntos de respuestas múltiples

Los siguientes estadísticos de resumen adicionales están disponibles en los conjuntos de respuestas múltiples.

Porcentaje de respuestas de columna/fila/capa. Porcentaje basado en respuestas.

Porcentaje de respuestas de columna/fila/capa (Base: recuento). Las respuestas son el numerador y el recuento total el denominador.

Porcentaje de recuento de columna/fila/capa (Base: respuestas). El recuento es el numerador y el total de respuestas es el denominador.

Porcentaje de respuestas de columna/fila de capa. Porcentajes en las subtablas. Porcentaje basado en respuestas.

Porcentaje de respuestas de columna/fila de capa (Base: recuento). Porcentajes en las subtablas. Las respuestas son el numerador y el recuento total el denominador.

Porcentaje de respuestas de columna/fila de capa (Base: respuestas). Porcentajes en las subtablas. El recuento es el numerador y el total de respuestas es el denominador.

Respuestas. Recuento de respuestas.

Subtable/Table Responses % [Porcentajes de respuestas de subtabla/tabla]. Porcentaje basado en respuestas.

Porcentaje de respuestas de subtabla/tabla (Base: recuento). Las respuestas son el numerador y el recuento total el denominador.

Porcentaje de recuentos de subtabla/tabla (Base: respuestas). El recuento es el numerador y el total de respuestas es el denominador.

Estadísticos de resumen para variables de escala y totales categóricos personalizados

Además de los recuentos y porcentajes disponibles para las variables categóricas, los siguientes estadísticos de resumen están disponibles para las variables de escala y como resúmenes de totales y subtotales personalizados para las variables categóricas. Estos estadísticos de resumen no están disponibles para conjuntos de respuestas múltiples o variables (alfanuméricas) de cadena.

Media. Media aritmética; la suma dividida por el número de casos.

Mediana. Los valores superior e inferior a los que corresponden la mitad de los casos; el percentil 50.

Modo. Valor más frecuente. Si hay un empate, se muestra el valor más pequeño.

Mínimo. El valor más pequeño (inferior).

Máximo. El valor más grande (superior).

Perdidos. Recuento de valores perdidos (tanto del usuario como del sistema).

Percentil. Puede incluir los percentiles 5, 25, 75, 95 y/o 99.

Rango. Diferencia entre los valores máximo y mínimo.

Error estándar de la media. Medida de cuánto puede variar el valor de la media de una muestra a otra, extraídas éstas de la misma distribución. Se puede utilizar para comparar en líneas generales la media observada con un valor hipotético (es decir, puede concluir que dos valores serán diferentes si la relación de la diferencia con el error estándar es inferior a -2 o superior a +2).

Desviación estándar. Una medida de dispersión sobre la media. En una distribución normal, el 68% de los casos entran dentro de una desviación estándar de la media y el 95% en dos desviaciones estándar. Por ejemplo, si la edad media es 45, con una desviación estándar del 10,95 % de los casos sería entre 25 y 65 en una distribución normal (la raíz cuadrada de la varianza).

Suma. Suma de valores.

Porcentaje de suma. Porcentajes basados en las sumas. Disponible para filas y columnas (dentro de subtablas), filas y columnas completas (en subtablas), capas, subtablas y tablas completas.

N total. Recuento de valores no perdidos, perdidos del usuario y perdidos del sistema. No incluye casos en categorías excluidas manualmente distintas de las categorías perdidas del usuario.

N válidos. Recuento de valores no perdidos. No incluye casos en categorías excluidas manualmente distintas de las categorías perdidas del usuario.

Varianza. Medida de dispersión sobre la media, igual a la suma de las desviaciones al cuadrado de la media dividida por el número de casos menos uno. La varianza se mide en unidades que son el cuadrado de las variables (el cuadrado de la desviación estándar).

Tablas apiladas

Cada sección de tabla definida por una variable de apilado se trata como una tabla separada y los estadísticos de resumen se calculan según ésta.

Estadísticos de resumen del total personalizados para variables categóricas

Para tablas de variables categóricas que contienen totales o subtotales, puede tener estadísticos de resumen diferentes de los resúmenes mostrados para cada categoría. Por ejemplo, podría visualizar los recuentos y los porcentajes de columna para una variable de fila categórica ordinal y mostrar la mediana de cada estadístico "total".

Para crear una tabla para una variable categórica con un estadístico de resumen personalizado total:

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
Se abrirá el generador de tablas.
2. Arrastre y suelte una variable categórica en el área de filas y/o columnas del lienzo.
3. Pulse con el botón derecho del ratón en el lienzo y seleccione **Categorías y totales** en el menú emergente.
4. Pulse en la casilla de verificación **Total** y, a continuación, seleccione **Aplicar**.
5. Vuelva a pulsar con el botón derecho del ratón en la variable del lienzo y seleccione **Estadísticos de resumen** desde el menú emergente.
6. Pulse en **Estadísticos de resumen personalizados para totales y subtotales** y, a continuación, seleccione los estadísticos de resumen personalizados que desee.

De forma predeterminada, todos los estadísticos de resumen, incluidos los resúmenes personalizados, se muestran en la dimensión opuesta de la que contiene la variable categórica. Por ejemplo, si tiene una variable de fila categórica, los estadísticos de resumen definen las columnas de la tabla como en:

Tabla 5. Categorías de variables ordinales en filas, recuento y media de estadísticos de resumen en columnas.

VARIABLES	CATEGORÍAS	RECuento	Media
Variable 1	1 De acuerdo	196	
	2 Neutral	936	
	3 En desacuerdo	744	
	Total	1876	2,29

Para visualizar los estadísticos de resumen en la misma dimensión que la variable categórica:

7. En la pestaña Tabla del generador de tablas, dentro del grupo Estadísticos de resumen, seleccione la dimensión de la lista desplegable Posición.

Por ejemplo, si la variable categórica se muestra en las filas, seleccione **Filas** en la lista desplegable.

Formatos de presentación de estadísticos de resumen

Están disponibles las siguientes opciones de formato de presentación:

nnnn. Numérico simple.

nnnn%. Signo de porcentaje añadido al final del valor.

Automático. Formato de presentación de variable definido, incluidos los números decimales.

N=nnnn. Muestra *N*= antes del valor. Puede ser útil para los recuentos, *N* válido y *N* total en las tablas donde no se muestran las etiquetas de estadísticos de resumen.

(nnnn). Todos los valores encerrados entre paréntesis.

(nnnn)(neg. value). Sólo los valores negativos encerrados entre paréntesis.

(**nnnn%**). Todos los valores encerrados entre paréntesis y un signo de porcentaje añadido al final de los valores.

n,nnn.n. Formato de coma. La coma se utiliza como separador de millares y el punto se utiliza como indicador decimal, independientemente de la configuración regional.

n.nnn,n. Formato de punto. El punto se utiliza como separador de millares y la coma se utiliza como indicador decimal, independientemente de la configuración regional.

\$n,nnn.n. Formato de dólar. El signo del dólar se muestra delante del valor; la coma se utiliza como separador de millares y el punto se utiliza como indicador decimal, independientemente de la configuración regional.

CCA, CCB, CCC, CCD, CCE. Formatos de moneda personalizados. El formato definido actualmente para cada moneda personalizada se muestra en la lista. Estos formatos se definen en la pestaña Moneda en el cuadro de diálogo Opciones (Menú Edición, Opciones).

Reglas generales y limitaciones

- Con la excepción de Automático, el número de decimales se determina por la configuración de la columna Decimales.
- Exceptuando la coma, el dólar y el formato de punto, el indicador de decimal utilizado es el definido para cada región del panel de control de su configuración regional en Windows.
- Aunque la coma, el dólar y el punto mostrarán una coma o un punto respectivamente como separador de millares, no hay un formato de presentación disponible en el tiempo de creación para visualizar un separador de millares en base a la configuración regional actual (definido en el panel de control de la configuración regional de Windows).

Categorías y totales

El cuadro de diálogo Categorías y totales permite:

- Reorganizar y excluir categorías.
- Insertar subtotales y totales.
- Insertar categorías calculadas.
- Incluir o excluir categorías vacías.
- Incluir o excluir categorías definidas como con valores perdidos.
- Incluir o excluir categorías que no tienen etiquetas de valores definidos.
- Este cuadro de diálogo está disponible sólo para las variables categóricas y los conjuntos de respuestas múltiples. No está disponible para las variables de escala.
- Si se seleccionan varias variables con distintas categorías, no se pueden insertar subtotales, insertar categorías calculadas, excluir categorías o reordenar las categorías manualmente. Esto sólo ocurre si selecciona varias variables en la vista previa del lienzo y accede a este cuadro de diálogo para todas las variables seleccionadas simultáneamente. También puede realizar estas acciones para cada variable de forma separada.
- Para las variables con etiquetas de valores no definidos, sólo puede clasificar las categorías e insertar totales.

Para acceder al cuadro de diálogo Categorías y totales

1. Arrastre y suelte una variable categórica o un conjunto de respuestas múltiples en el panel de lienzo.
2. Pulse con el botón derecho del ratón en la variable del panel de lienzo y seleccione **Categorías y totales** desde el menú emergente.

o

3. Seleccione (pulse en) la variable del panel de lienzo y pulse en **Categorías y totales** en el grupo Definir de la pestaña Tabla.
También puede seleccionar varias variables categóricas en la misma dimensión del panel de lienzo:
4. Pulse Ctrl y pulse en cada variable en el panel de lienzo.
o
5. Pulse fuera de la vista previa de la tabla en el panel de lienzo y, a continuación, pulse y arrastre para seleccionar el área que incluye las variables que desea seleccionar.
o
6. Pulse con el botón derecho del ratón en una dimensión y seleccione **Seleccionar todas las variables de la [dimensión]** para seleccionar todas las variables en esa dimensión.

Reorganización de categorías

Para reorganizar categorías manualmente:

1. Seleccione (pulse en) una categoría de la lista.
2. Pulse en la flecha hacia arriba o hacia abajo para mover la categoría hacia arriba o abajo en la lista.
o
3. Pulse en la columna Valor de cada categoría y arrástrela y suéltela en una posición diferente.

Exclusión de categorías

1. Seleccione (pulse en) una categoría de la lista.
2. Pulse en la flecha junto a la lista Excluir.
o
3. Pulse en la columna Valor de cada categoría y arrástrela y suéltela en cualquier lugar fuera de la lista.

Si excluye cualquier categoría, también se excluirá cualquier categoría sin etiquetas de valores definidos.

Clasificación de las categorías

Puede clasificar las categorías por el valor de datos, etiqueta de valor, recuento de casilla o estadístico de resumen en orden ascendente o descendente.

1. En el grupo Clasificar categorías, pulse en la lista desplegable Por y seleccione el criterio de clasificación que desee utilizar: valor, etiqueta, recuento o estadístico de resumen (como media, mediana o modo). Los estadísticos de resumen disponibles para la clasificación dependen de los estadísticos de resumen que haya seleccionado para mostrar en la tabla.
2. Pulse en la lista desplegable Orden para seleccionar el orden de clasificación (ascendente o descendente).

Las categorías de clasificación no están disponibles si ha excluido alguna categoría.

Subtotales

1. Seleccione (pulse en) la categoría de la lista que sea la última categoría en el rango de categorías que desee incluir en el subtotal.
2. Pulse en **Añadir subtotal...**
3. En el cuadro de diálogo Definir subtotal, modifique la etiqueta de texto de subtotal si lo desea.
4. Para mostrar sólo un subtotal y suprimir la presentación de las categorías que definen el subtotal, seleccione **Ocultar categorías subtotalizadas desde la tabla**.
5. Pulse en **Continuar** para añadir el subtotal.

Totales

1. Pulse en la casilla de verificación **Total**. También puede modificar el texto de la etiqueta total.

Si la variable seleccionada se anida dentro de otra variable, los totales se insertarán para cada subtabla.

Visualización de posición de totales y subtotales

Los totales y subtotales se pueden visualizar encima o debajo de las categorías incluidas en cada total.

- Si selecciona **Debajo** en el grupo de visualización de totales y subtotales, aparecerán los totales encima de cada subtabla y todas las categorías por encima e incluidas en la categoría seleccionada (pero debajo de cualquier subtotal precedente) se incluyen en cada subtotal.
- Si selecciona **Encima** en el grupo de visualización de totales y subtotales, aparecerán los totales debajo de cada subtabla y todas las categorías por debajo e incluidas en la categoría seleccionada (pero por encima de cualquier subtotal precedente) se incluyen en cada subtotal.

Importante: debe seleccionar la posición de visualización de subtotales antes de definir cualquier subtotal. El cambio de posición de visualización afecta a todos los subtotales (no sólo al subtotal seleccionado actualmente) y también *cambia las categorías incluidas en los subtotales*.

Categorías calculadas

Puede mostrar categorías calculadas de estadísticos de resumen, totales, subtotales y/o constantes. Consulte el tema "Categorías calculadas" en la página 14 para obtener más información.

Estadísticos de resumen de totales y subtotales personalizados

Puede visualizar los estadísticos distintos de "totales" en las áreas totales y subtotales de la tabla mediante el cuadro de diálogo Estadísticos de resumen. Consulte el tema "Estadísticos de resumen para variables categóricas" en la página 7 para obtener más información.

Nota: Si selecciona varios estadísticos de totales personalizados que también se incluyen en el cuerpo de la tabla y oculta las etiquetas de estadísticos, los totales se vuelven a clasificar en el mismo orden que en el cuerpo de la tabla y, dado que no se muestran las etiquetas, es posible que no sepa qué representa realmente cada estadístico total. En general, probablemente no sea una buena idea seleccionar varios estadísticos y ocultar las etiquetas de los estadísticos.

Totales, subtotales y categorías excluidas

Los casos de categorías excluidas no se incluyen en el cálculo de totales.

Valores perdidos, categorías vacías y valores sin etiquetas de valor.

Valores perdidos. Controla la visualización de valores **perdidos del usuario** o valores definidos como con valores perdidos (por ejemplo, un código de 99 para representar "no aplicable" al embarazo en hombres). De forma predeterminada, se excluyen los valores perdidos del usuario. Seleccione (marque) esta opción para incluir en las tablas las categorías perdidas del usuario. Aunque la variable puede contener más de una categoría de valor perdido, la vista previa de la tabla en el lienzo mostrará sólo una categoría de valor perdido genérica. Todas las categorías perdidas del usuario que se hayan definido se incluirán en la tabla. Los **valores perdidos del sistema** (casillas vacías para las variables numéricas en el Editor de datos) se excluyen siempre.

Categorías vacías. Las categorías vacías son categorías con etiquetas de valores definidos pero no casos en esa categoría para una tabla o subtabla determinada. De forma predeterminada, las categorías vacías se incluyen en las tablas. Anule la selección de esta opción para excluir de la tabla las categorías perdidas.

Se han encontrado otros valores al escanear los datos. De forma predeterminada, los valores de categoría del archivo de datos que no tienen etiquetas de valores definidos se incluyen automáticamente en las tablas. Anule la selección de esta opción para excluir de la tabla los valores sin etiquetas de valores definidos. Si excluye cualquier categoría con etiquetas de valores definidos, también se excluirán las categorías sin etiquetas de valores definidos.

Categorías calculadas

Además de mostrar los resultados agregados de estadísticos de resumen, una tabla puede mostrar una o más categorías calculadas a partir de estos resultados calculados, a partir de valores constantes, a partir de subtotales y totales o a partir de una combinación de todos ellos. Los resultados se conocen como categorías calculadas o cálculos posteriores. Una categoría calculada funciona como una categoría en una única variable con las siguientes similitudes y diferencias:

- Una categoría calculada se coloca como el resto de categorías.
- Una categoría calculada funciona en los mismos estadísticos que el resto de categorías.
- Las categorías calculadas no afectan a los subtotales, totales o pruebas de significación.
- De forma predeterminada, los valores de las categorías calculadas utilizan el mismo formato de estadísticos de resumen que el resto de categorías. Puede sustituir el formato cuando defina la categoría calculada.

Como las categorías calculadas se pueden utilizar para totalizar los resultados agregados, pueden ser similares a los subtotales. Sin embargo, las categorías calculadas tienen las siguientes ventajas con respecto a subtotales:

- Las categorías calculadas pueden calcularse a partir de los resultados de otros subtotales.
- Las categorías calculadas pueden solaparse entre sí, operando en las mismas categorías (o en parte de las mismas).
- Las categorías calculadas no tienen que incluir valores del resto de categorías superiores o inferiores a la categoría calculada. Es decir, las categorías calculadas no son exhaustivas.
- Las categorías calculadas pueden incluir valores de categorías no adyacentes.

Al contrario que los totales y subtotales, las categorías calculadas se calculan a partir de los datos agregados en lugar de los datos originales. Por lo tanto, es posible que los valores de las categorías calculadas no coincidan con los resultados de los totales y subtotales. Además, como tiene la opción de ocultar categorías de origen cuando se define una categoría calculada, puede ser difícil para interpretar los subtotales en la tabla resultante. Si utiliza categorías calculadas, es recomendable que especifique etiquetas personalizadas para los subtotales.

Para definir una categoría calculada

Las categorías calculadas se añaden desde el cuadro de diálogo Categorías y totales. Para obtener información sobre cómo acceder al cuadro de diálogo, consulte el tema “Categorías y totales” en la página 11.

1. En el cuadro de diálogo Categorías y totales, pulse en **Añadir categoría...**
2. En **Etiqueta para categoría calculada**, especifique una etiqueta para la categoría calculada. Puede arrastrar categorías desde la lista Categorías para incluir etiquetas para esas categorías.
3. Genere una expresión seleccionando categorías y/o totales y subtotales y utilizando operadores para definir las categorías calculadas. También puede escribir valores constantes (p. ej., 500) para incluirlos en la expresión.
4. Para mostrar sólo una categoría calculada y suprimir la presentación de las categorías que definen la categoría calculada, seleccione **Ocultar categorías empleadas en la expresión de la tabla**.
5. Pulse en la pestaña **Formatos de presentación** para cambiar el formato de presentación y el número de decimales de la categoría calculada. Consulte el tema “Formatos de presentación de categorías calculadas” en la página 15 para obtener más información.

6. Pulse en **Continuar** para añadir la categoría calculada.

Formatos de presentación de categorías calculadas

De forma predeterminada, una categoría calculada utiliza el mismo formato de presentación y número de decimales que las otras categorías de la variable. Puede sustituirlos en la pestaña Formatos de presentación del cuadro de diálogo Categoría calculada. La pestaña Formatos de presentación indica los estadísticos de resumen actuales en los que funciona la categoría calculada además de los formatos de presentación y número de decimales de dichos estadísticos.

Para cada estadístico de resumen, puede:

1. Seleccionar un formato de presentación para el estadístico de resumen en la lista desplegable Formato. Para obtener una lista completa de los formatos de presentación, consulte “Formatos de presentación de estadísticos de resumen” en la página 10.
2. Introducir el número de decimales que desea mostrar en la casilla Decimales para el estadístico de resumen seleccionado.

Tablas de variables con categorías compartidas (Tablas Comperimeter)

Las encuestas a menudo contienen muchas preguntas con un conjunto común de posibles respuestas. Puede utilizar el apilado para mostrar estas variables relacionadas en la misma tabla y puede mostrar las categorías de respuestas compartidas en las columnas de la tabla.

Creación de una tabla de múltiples variables con categorías compartidas

Tabla 6. Variables apiladas con categorías de respuestas compartidas en columnas

Variables	Categoría 1	Categoría 2	Categoría 3
Variable 1	12	34	56
Variable 2	56	12	34
Variable 3	34	56	12

Consulte el tema Capítulo 6, “Tablas para variables con categorías compartidas”, en la página 47 para obtener más información.

Personalización del generador de tablas

A diferencia de los cuadros de diálogo estándar, puede cambiar el tamaño del generador de tablas de la misma forma que puede cambiar el tamaño de una ventana estándar:

1. Pulse y arrastre la parte superior, inferior, ambos lados o cualquier esquina del generador de tablas para aumentar o disminuir su tamaño.
En la pestaña Tabla, también puede cambiar el tamaño de la lista de variables, la lista Categorías y el panel de lienzo.
2. Pulse y arrastre la barra horizontal entre la lista de variables y la lista Categorías para alargar o acortar las listas. Si la arrastra hacia abajo alarga la lista de variables y acorta la lista Categorías. Si la arrastra hacia arriba ocurre lo contrario.
3. Pulse y arrastre la barra vertical entre la lista de variables y la lista Categorías del panel de lienzo para ensanchar o estrechar las listas. El lienzo cambia de tamaño automáticamente para ajustarse al espacio restante.

Tablas personalizadas: pestaña Opciones

La pestaña Opciones permite:

- Especificar lo que se muestra en las casillas vacías y las casillas para las que no se pueden calcular estadísticos.
- Controlar cómo se tratan los valores perdidos en el cálculo de los estadísticos variables de escala.
- Definir anchos de columna de datos máximos o mínimos.
- Controlar el tratamiento de las respuestas duplicadas en conjuntos de categorías múltiples.

Aspecto de la casilla de datos. Controla lo que se muestra en las casillas vacías y las casillas para las que no se pueden calcular estadísticos.

- **Casillas vacías.** Para las casillas de las tablas que no contienen casos (recuento de casilla de 0) puede seleccionar una de las tres opciones de presentación: cero, en blanco o un valor de texto que especifique. El valor de texto puede contener hasta 255 caracteres.
- **Estadísticos que no se pueden calcular.** Texto mostrado si un estadístico no se puede calcular (por ejemplo, la media de una categoría sin casos). El valor de texto puede tener hasta 255 caracteres. El valor predeterminado es un punto (.).

Ancho de columnas de datos. Controla el ancho de columna máximo y mínimo de las columnas de datos. Este ajuste no afecta a los anchos de columna de las etiquetas de fila.

- **Configuración de TableLook.** Utiliza la especificación del ancho de columna de datos del TableLook predeterminado actual. Puede crear su propio TableLook personalizado predeterminado para utilizar cuando se crean nuevas tablas y puede controlar los anchos de columna de datos y columna de etiqueta de fila con TableLook.
- **Personalizado.** Sobrescribe la configuración predeterminada de TableLook del ancho de columna de datos. Especifique los anchos de columna de datos máximo y mínimo de la tabla y la unidad de medición: puntos, pulgadas o centímetros.

Valores perdidos de variables de escala. Para las tablas con dos o más variables de escala, controla el tratamiento de los datos perdidos de los estadísticos de variables de escala.

- **Maximizar el uso de datos disponibles (eliminación variable por variable).** Todos los casos con valores válidos para cada variable de escala se incluyen en los estadísticos de resumen para dicha variable de escala.
- **Utilizar una base de casos coherente en las variables de escala (igual que la eliminación).** Los casos con valores perdidos para cualquier variable de escala de la tabla se excluyen de los estadísticos de resumen para todas las variables de escala de la tabla.

Contar respuestas duplicadas de conjuntos de categorías múltiple. Una respuesta duplicada es la misma respuesta para dos o más variables en el conjunto de categorías múltiples. De forma predeterminada, las respuestas duplicadas no se cuentan, pero puede ser una condición perfectamente válida que desea incluir en el recuento (como un conjunto de categorías múltiples que representa el fabricante de los tres últimos coches comprados por el encuestado).

Ocultar recuentos pequeños. Puede optar por ocultar recuentos inferiores un número entero especificado. Los valores ocultos se mostrarán como $<N$, donde N es el número entero especificado. El número entero especificado debe ser mayor o igual a 2.

Tablas personalizadas: pestaña Títulos

La pestaña Títulos controla la visualización de títulos, pies y etiquetas de esquina.

Título. Texto que se muestra sobre la tabla.

Pie. Texto que se muestra debajo de la tabla y encima de cualquier nota al pie.

Esquina. Texto que se muestra en la esquina superior izquierda de la tabla. El texto de esquina se muestra sólo si la tabla contiene variables de fila y si la propiedad de etiqueta de dimensión de fila de la tabla pivote esta definida como **Anidado**. Esta *no* es la configuración predeterminada de TableLook.

Puede incluir los siguientes valores generados automáticamente en el título, el pie o la etiqueta de la esquina de la tabla:

Fecha. Año, mes y día actual mostrados en un formato basado en su configuración regional de Windows.

Tiempo. Hora, minutos y segundos actuales mostrados en un formato basado en su configuración regional de Windows.

Expresión de tabla. Las variables utilizadas en la tabla y cómo se utilizan en la tabla. Si una variable tiene una etiqueta de variables definidas, se mostrará la etiqueta. En la tabla generada, los símbolos siguientes indican cómo se utilizan las variables en la tabla:

- + variables apiladas indicadas.
- > indica anidación.
- **BY** indica tabulación cruzada o capas.

Tablas personalizadas: pestaña Estadísticos de contraste

La pestaña Estadísticos de contraste permite solicitar diferentes comprobaciones de significación para las tablas personalizadas, entre las que se incluyen:

- Comprobación de independencia de chi-cuadrado.
- Comprobaciones de la igualdad de las medias de columna.
- Comprobaciones de la igualdad de las proporciones de columna.
- Comprobaciones de significación para conjuntos de respuestas múltiples y subtotales. (Si desea obtener información sobre las comprobaciones de significación de conjuntos de respuestas múltiples, consulte “Pruebas de significación con conjuntos de respuestas múltiples” en la página 80).

Estas comprobaciones no están disponibles para las tablas en las que las etiquetas de categoría cambian su dimensión de tabla predeterminada.

Comparar medias de columna (pruebas t). Esta opción realiza comprobaciones por parejas de la igualdad de las medias de columna en las tablas que contienen al menos una variable de categoría en las columnas y una variable de escala en las filas. La tabla debe incluir la media como estadístico de resumen. Puede seleccionar si los valores p de las comprobaciones se ajustan o no mediante el método de Bonferroni. También puede especificar el nivel alfa de la comprobación, que debe ser un valor superior a 0 e inferior a 1. Finalmente, mientras que la varianza de la comprobación de medias siempre se basa solamente en las categorías comparadas de comprobaciones de respuestas múltiples, las variables categóricas normales se pueden calcular a partir de las categorías comparadas o de todas las categorías.

Comparar proporciones de columnas (pruebas z). Esta opción realiza comprobaciones por parejas de la igualdad de proporciones de columna en las tablas que contienen al menos una variable de categoría en las filas y columnas. La tabla debe incluir recuentos o porcentajes simples de columna. Puede seleccionar si los valores p de las comprobaciones se ajustan o no mediante el método de Bonferroni. También puede especificar el nivel alfa de la comprobación, que debe tener un valor superior a 0 e inferior a 1.

Identificar diferencias significativas. Si selecciona **Comparar las medias de columna** o **Comparar las proporciones de columna**, puede seleccionar cómo indicar diferencias significativas.

- **En una tabla diferente.** Los resultados de las pruebas de significación se muestran en una tabla independiente. Si hay dos valores significativamente diferentes, la casilla del valor mayor muestra una clave que identifica la columna de menor valor. A continuación se muestra un ejemplo:

- **En la tabla principal utilizando subíndices de estilo APA.** La tabla principal identifica diferencias significativas con el formato de estilo APA utilizando subíndices de letras. Si hay dos valores significativamente diferentes, cuyos valores muestran subíndices de letras *diferentes*. Estos subíndices no son notas al pie. Si esta opción está activada, el estilo de nota al pie definido en el aspecto de tabla actual se sustituye y las notas al pie se muestran como números de superíndice. A continuación se muestra un ejemplo:

Para ver el conjunto completo que demuestra cómo crear e interpretar estas tablas, consulte “Comparación de proporciones de columna” en la página 69.

Nota: la tabla en estilo APA incluye un pie que explica cómo interpretar los subíndices de la tabla. Si especifica un pie en la pestaña Títulos, se mostrará el pie especificado en la pestaña en lugar del pie que explica los subíndices.

Comprobaciones de independencia (chi-cuadrado). Esta opción realiza una comprobación de independencia de chi-cuadrado en las tablas que contienen al menos una variable de categoría en las filas y columnas. También puede especificar el nivel alfa de la comprobación, que debe tener un valor superior a 0 e inferior a 1.

Utilizar subtotales en lugar de categorías con subtotales. Si se selecciona, cada subtotal sustituye sus categorías en la comprobación de significación. En caso contrario, sólo los subtotales cuyas categorías con subtotales están ocultas sustituirán sus categorías en la comprobación.

Incluir variables de respuestas múltiples. Si se selecciona, las comprobaciones se realizan utilizando categorías de conjuntos de respuestas múltiples. En caso contrario, los conjuntos de respuestas múltiples se ignoran durante la realización de comprobaciones.

Capítulo 2. Tablas simples para variables categóricas

Tablas simples para variables categóricas

La mayoría de las tablas que se crean probablemente incluirán como mínimo una **variable categórica**. Una variable categórica es una variable con un número limitado de valores distintos o categorías (por ejemplo, sexo o religión). Las variables categóricas pueden ser **nominales** u **ordinales**.

- *Nominal*. Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen a una clasificación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado. Algunos ejemplos de variables nominales son: región, código postal o confesión religiosa.
- *Ordinal*. Una variable puede ser tratada como ordinal cuando sus valores representan categorías con alguna clasificación intrínseca. Por ejemplo, los niveles de satisfacción con un servicio, que abarquen desde muy insatisfecho hasta muy satisfecho. Entre los ejemplos de variables ordinales se incluyen escalas de actitud que representan el grado de satisfacción o confianza y las puntuaciones de evaluación de las preferencias.

Un icono junto a cada variable en la lista de variables identifica el tipo de variable.

Tabla 7. Iconos de nivel de medición

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

Las Tablas personalizadas han sido optimizadas para el uso de variables categóricas para las que se hayan definido **etiquetas de valor**. Consulte el tema “Generación de tablas” en la página 1 para obtener más información.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Una variable categórica individual

Posiblemente una de las tablas más simples que se pueden crear sea una tabla con una única variable categórica, pero en la mayoría de las ocasiones puede que esto sea lo único que necesite o desee.

1. Seleccione en los menús:

Analizar > Tablas > Tablas personalizadas...

- En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
Se mostrará una vista previa de la tabla, en el panel de lienzo. Esta vista previa no muestra los valores reales de los datos, sino sólo marcadores de posición que indican dónde aparecerán los datos.
- Pulse en **Aceptar** para crear la tabla.
La tabla aparecerá en la ventana del Visor.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 2. Variable categórica única en filas

En esta tabla simple, el encabezado de columna *Recuento* no es en realidad necesario y se puede crear la tabla sin este encabezado.

- Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- En el grupo Estadísticos de resumen, seleccione (pulse) **Ocultar** para Posición.
- Pulse en **Aceptar** para crear la tabla.

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 3. Variable categórica única sin etiqueta de columna de estadísticos de resumen

Porcentajes

Además de los recuentos, también puede hacer que aparezcan los porcentajes. En una tabla simple de una única variable categórica, si la variable se muestra en las filas, seguramente desee consultar los porcentajes de columna. Por el contrario, si la variable se muestra en las columnas, probablemente desee consultar los porcentajes de fila.

- Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- En el grupo Estadísticos de resumen, deselectione **Ocultar** para Posición. Dado que esta tabla tendrá dos columnas, deseará que aparezcan las etiquetas de las columnas, para saber qué es lo que representa cada columna.
- Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
- En el cuadro de diálogo Estadísticos de resumen, seleccione **% del N de columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
- En la casilla Etiqueta de la lista Visualización, borre la etiqueta predeterminada y escriba Porcentaje.
- Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

Figura 4. Recuentos y porcentajes de columna

Totales

Los totales no se incluyen automáticamente en las tablas personalizadas, pero es fácil añadirlos a una tabla.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
3. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales.
4. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figura 5. Recuentos, porcentajes de columna y totales

Consulte el tema Capítulo 4, “Totales y subtotales de variables categóricas”, en la página 35 para obtener más información.

Tabulación cruzada

La tabulación cruzada es una técnica básica para el estudio de la relación existente entre dos variables categóricas. Por ejemplo, si se utiliza *Categoría de edad* como variable de fila y *Sexo* como variable de columna, se puede crear una tabulación cruzada de dos dimensiones que muestre el número de hombres y mujeres que pertenecen a cada categoría de edad.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para eliminar cualquier selección anterior en el generador de tablas.
3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y coloque *Sexo* desde la lista de variables y colóquela en el área Columnas. (Puede que tenga que desplazarse hacia abajo por la lista de variables para encontrarla.)
5. Pulse en **Aceptar** para crear la tabla.

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

Figura 6. Tabulación cruzadas de categoría de edad y sexo

Porcentajes en las tabulaciones cruzadas

En una tabulación cruzadas de dos dimensiones, los porcentajes de fila y los de columna pueden proporcionar información útil.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Sexo* en el panel de lienzo.
Puede observar que **Estadísticos de resumen** está desactivado en el menú emergente. Ello se debe a que es posible seleccionar estadísticos de resumen sólo para la variable más al interior de la dimensión origen de los estadísticos. La dimensión predeterminada de origen de estadísticos (fila o columna) para las variables categóricas se basa en el orden en el que arrastre y suelte las variables en el panel de lienzo. En este ejemplo, hemos arrastrado primero *Categoría de edad* a la dimensión de las filas y, teniendo en cuenta que no hay ninguna otra variable en la dimensión de las filas, *Categoría de edad* es la variable origen de los estadísticos. Puede cambiar la dimensión origen de los estadísticos, pero en este ejemplo, no es necesario. Consulte el tema “Estadísticos de resumen” en la página 6 para obtener más información.
3. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
4. En el cuadro de diálogo Estadísticos de resumen, seleccione **% del N de columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
5. Seleccione **% del N de la fila** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
6. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

Figura 7. Tabulación cruzada con porcentajes de fila y columna

Control del formato de presentación

Puede controlar el formato de presentación, incluyendo el número de decimales que aparecen en los estadísticos de resumen. Por ejemplo, los porcentajes se muestran de forma predeterminada con un decimal y el signo de porcentaje. Si desea que los valores de las casillas aparezcan con dos decimales y sin signo de porcentaje, sólo tiene que seguir estos pasos.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).

2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
3. Para los dos estadísticos de resumen de porcentaje seleccionados (**% de columna** y **% del N de la fila**), seleccione **nnnn.n** en la lista desplegable Formato y escriba 2 en la casilla Decimales de ambos.
4. Pulse en **Aceptar** para crear la tabla.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

Figura 8. Representación de casilla con formato para los porcentajes de fila y columna

Totales marginales

En las tabulaciones cruzadas es bastante común mostrar los **totales marginales** de cada fila y columna. Como de forma predeterminada no se incluyen estos valores en las tablas personalizadas, deberá añadirlos explícitamente a sus tablas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para eliminar cualquier selección anterior en el generador de tablas.
3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y coloque *Sexo* desde la lista de variables y colóquela en el área Columnas. (Puede que tenga que desplazarse hacia abajo por la lista de variables para encontrarla.)
5. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
6. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.
7. Pulse con el botón derecho del ratón *Género* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
8. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.
9. En el grupo Estadísticos de resumen, seleccione (pulse) **Ocultar** para Posición. (Como sólo se muestran los recuentos, no es necesario identificar el "estadístico" que se muestra en las casillas de datos de la tabla).
10. Pulse en **Aceptar** para crear la tabla.

		Gender		
		Male	Female	Total
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	Total	1228	1600	2828

Figura 9. Tabulación cruzadas con totales marginales

Ordenación y exclusión de categorías

De forma predeterminada, las categorías se muestran en el orden ascendente de los valores de los datos, representados mediante sus etiquetas de valor. Por ejemplo, aunque para la categoría de edad aparecen las etiquetas de valor *Menos de 25*, *25 a 34*, *35 a 44*, ..., etc., los valores reales de los datos subyacentes son 1, 2, 3, ..., etc., y son estos valores subyacentes los que controlan el orden de presentación predeterminado de las categorías.

Puede cambiar fácilmente el orden de las categorías, así como excluir aquellas categorías que no desea que aparezcan en la tabla.

Ordenación de las categorías

Puede organizar manualmente las categorías u ordenarlas en orden ascendente o descendente por:

- Valores de datos.
- Etiquetas de valor.
- Recuentos de las casillas.
- Estadísticos de resumen. Los estadísticos de resumen disponibles para la clasificación dependen de los estadísticos de resumen que haya seleccionado para mostrar en la tabla.

1. Abra el generador de tablas (menú **Analizar**, **Tablas**, **Tablas personalizadas**).
2. Si *Categoría de edad* todavía no aparece en el área Filas del panel de lienzo, arrástrela y colóquela aquí.
3. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.

Tanto los valores de los datos como las etiquetas de valor asociadas aparecerán siguiendo el orden actual, que en este caso es todavía el orden ascendente de los valores de los datos.

4. En el grupo Ordenar categorías, seleccione **Descendente** en la lista desplegable Orden.

Ahora se habrá invertido el orden.

5. Seleccione **Etiquetas** en la lista desplegable Por.

Las categorías aparecerán ahora ordenadas siguiendo el orden alfabético descendente de las etiquetas de valor.

Observe que la categoría con la etiqueta *Menos de 25* aparece en la parte superior de la lista. Según el orden alfabético, las letras van detrás de los números. Al ser ésta la única etiqueta que comienza con una letra y, dado que la lista aparece ordenada en orden descendente (inverso), esta categoría aparecerá en la parte superior de la lista.

Si desea que una determinada categoría aparezca en un lugar diferente de la lista, puede moverlo con facilidad.

6. Pulse en la categoría con la etiqueta *Menos de 25* en la lista Etiqueta.
7. Pulse en la flecha hacia abajo que hay a la derecha de la lista. La categoría se desplazará en la lista una fila hacia abajo.
8. Siga pulsando en la flecha hacia abajo hasta que la categoría llegue al final de la lista.

Exclusión de categorías

Si existen algunas categorías que no desea que aparezcan en la tabla, puede excluirlas.

1. Pulse en la categoría con la etiqueta *Menos de 25* en la lista Etiqueta.
2. Pulse en la flecha hacia abajo que hay a la izquierda de la lista Excluir.
3. Pulse en la categoría con la etiqueta *65 o mayor* en la lista Etiqueta.
4. Vuelva a pulsar en la flecha hacia abajo que hay a la izquierda de la lista Excluir.

Las dos categorías se desplazarán de la lista Visualización a la lista Excluir. Si cambia de idea, puede devolverlas con facilidad a la lista Visualización.

5. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Gender		
		Male	Female	Total
Age category	55 to 64	136	184	320
	45 to 54	221	260	481
	35 to 44	309	370	679
	25 to 34	276	351	627
	Total	942	1165	2107

Figura 10. Tabla ordenada por etiquetas de valor descendentes, con algunas categorías excluidas

Observe que los totales son menores de lo que eran antes de que se excluyesen las dos categorías. Se debe a que los totales se basan en las categorías que se incluyen en la tabla. Todas las categorías excluidas se excluyen del cálculo total. Consulte el tema Capítulo 4, “Totales y subtotales de variables categóricas”, en la página 35 para obtener más información.

Capítulo 3. Apilado, anidación y capas con variables categóricas

El apilado, la anidación y las capas son métodos para mostrar múltiples variables en la misma tabla. Este capítulo se centra en el uso de estas técnicas con variables categóricas, aunque también se pueden utilizar con variables de escala.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Variables categóricas de apilado

El apilado se puede considerar como el procedimiento por el cual se toman tablas separadas y se pegan en la misma visualización. Por ejemplo, puede mostrar información en la categoría *Género* y *Edad* en secciones separadas de la misma tabla.

1. Seleccione en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre y suelte la categoría *Género* de la lista de variables en el área Filas del panel de lienzo.
3. Arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas en *Género*.
Las dos variables están apiladas en la dimensión de fila.
4. Pulse en **Aceptar** para crear la tabla.

		Count
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figura 11. Tabla de variables categóricas de apilado en filas

También puede apilar variables en columnas de forma similar.

Apilado con tablas cruzadas

Una tabla apilada puede incluir otras variables en otras dimensiones. Por ejemplo, puede crear una tabulación cruzada con dos variables apiladas en las filas con una tercera variable en la dimensión de columna.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Si la categoría *Edad* y *Género* ya no están apiladas en las filas, siga las instrucciones anteriores antes de apilarlas.

3. Arrastre y suelte *Obtener noticias por Internet* de la lista de variables en el área Columnas del panel de lienzo.
4. Pulse en **Aceptar** para crear la tabla.

		Get news from internet	
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

Figura 12. Dos tablas cruzadas con variables de fila apiladas con una variable de columna

Nota: existen diferentes variables con etiquetas que comienzan por *Obtener noticias por ...*, de forma que puede ser difícil distinguirlas en la lista de variables (ya que las etiquetas pueden ser demasiado extensas para que aparezcan completamente en la lista de variables). Existen dos formas para visualizar la etiqueta de variable completa:

- Pase el puntero del ratón por encima de una variable de la lista para mostrar la etiqueta completa en una ayuda contextual emergente.
- Pulse y arrastre la barra vertical que separa la variable y la lista Categorías del panel de lienzo para ensanchar las listas.

Anidación de variables categóricas

La anidación, como tabulación de contingencia, puede mostrar la relación entre dos variables categóricas, excepto una variable anidada dentro de otra en la misma dimensión. Por ejemplo, puede anidar *Género* dentro de la categoría *Edad* en la dimensión de la fila, mostrando el número de hombres y mujeres de cada categoría de edad.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para eliminar cualquier selección anterior en el generador de tablas.
3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y suelte *Género* de la lista de variables a la derecha de la categoría *Edad* en el área Filas.

La vista previa del panel de lienzo muestra que la tabla anidada contendrá una columna simple de recuentos; cada casilla contiene el número de individuos masculinos o femeninos de la categoría de edad.

Puede observar que la etiqueta de la variable *Género* se muestra varias veces, una vez para cada categoría de edad. Puede reducir este tipo de repetición colocando la variable con el menor número de categorías en el nivel más exterior de la anidación.

5. Pulse en la etiqueta de la variable *Género* en el panel de lienzo.
6. Arrastre y suelte la variable en la parte izquierda del área Filas, tan cerca del borde izquierdo como le sea posible.

En lugar de repetirse *Género* seis veces, la categoría *Edad* se repite dos veces. Es una tabla menos poblada que esencialmente produce los mismos resultados.

7. Pulse en **Aceptar** para crear la tabla.

				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Figura 13. Tabla de la categoría Edad en Género

Nota: las tablas personalizadas no tienen en cuenta el procesamiento del archivo segmentado en capas. Para lograr los mismos resultado que los archivos segmentados en capas, coloque las variables de archivo segmentado en las capas más exteriores de la tabla.

Eliminación de las etiquetas de variable

Otra solución para evitar la repetición de etiquetas en las tablas anidadas consiste simplemente en suprimir la visualización de nombres de variables o etiquetas. Como las etiquetas de valores de *Género* y la categoría *Edad* son, con toda probabilidad, suficientemente descriptivas sin las etiquetas de variable, es posible eliminar las etiquetas de ambas variables.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel del lienzo y deseleccione **Mostrar etiqueta de variable** en el menú emergente.
3. Realice el mismo procedimiento en *Género*.

Las etiquetas de variable se siguen mostrando en la vista previa de la tabla, pero no se incluirán en la tabla.

4. Pulse en **Aceptar** para crear la tabla.

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 14. Tabla anidada sin etiquetas de variable

Si desea incluir las etiquetas de variable con la tabla en cualquier parte, sin que aparezcan varias veces en el cuerpo de la tabla, puede incluirlas en el título de la tabla o en la etiqueta de la esquina.

5. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
6. Pulse en la pestaña **Títulos**.
7. Pulse en cualquier parte del cuadro de texto Título.
8. Pulse en **Expresión de tabla**. Aparecerá el texto *&Expresión de tabla* en el cuadro de texto Título. Se generará un título de tabla que incluye las etiquetas de las variables de la tabla.
9. Pulse en **Aceptar** para crear la tabla.

Gender > Age category

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 15. Etiquetas de variable en el título de la tabla

El signo mayor que (>) del título indica que la categoría *Edad* está anidada en *Género*.

Tabulación cruzada anidada

Una tabla anidada puede contener otras variables en otras dimensiones. Por ejemplo, puede anidar la categoría *Edad* en *Género* en las filas y crear una tabulación cruzada con filas anidadas con una tercera variable en la dimensión de columna.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Si la categoría *Edad* todavía no está anidada en *Género* en las filas, siga las instrucciones anteriores para anidarlas.
3. Arrastre y suelte *Obtener noticias por Internet* de la lista de variables en el área Columnas del panel de lienzo.

Puede observar que la tabla es demasiado grande para aparecer completamente en el panel de lienzo. Puede desplazarse hacia cualquier dirección en el panel de lienzo para visualizar la vista previa de la tabla o bien puede:

- Pulsar en **Compacta** en el generador de tablas para abrir una vista compacta. Muestra únicamente las etiquetas de variable, sin información de las categorías o estadísticos de resumen incluidos en la tabla.
- Aumente el tamaño del generador de tablas pulsando y arrastrando cualquier lado o esquina del generador de tablas.

4. Pulse en **Aceptar** para crear la tabla.

				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
			65 or older	155	23
	Female	Age category	Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
			55 to 64	140	44
			65 or older	261	40

Figura 16. Tabulación cruzada anidada

Intercambio de filas y columnas

¿Qué debe hacer si pasa mucho tiempo definiendo una tabla compleja y decide que es absolutamente perfecta, salvo que desea cambiar la orientación, incluyendo todas las variables de fila en las columnas y viceversa? Por ejemplo, ha creado una tabulación cruzada anidada con la categoría *Edad* y *Género* anidadas en las filas, pero ahora sin embargo desea anidar estas dos variables demográficas en las columnas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en el panel de lienzo y seleccione **Intercambiar de filas y variables de columna** en el menú emergente.
Las variables de fila y columna se han intercambiado.
Antes de crear la tabla, realicemos algunas modificaciones para que la visualización esté menos poblada.
3. Seleccione **Ocultar** para suprimir la visualización de los estadísticos de resumen de la etiqueta de columnas.
4. Pulse con el botón derecho del ratón *Género* en el panel del lienzo y deselectione **Mostrar etiqueta de variable**.
5. Pulse en **Aceptar** para crear la tabla.

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Figura 17. Tabulación cruzada con variables demográficas anidadas en las columnas

Capas

Puede utilizar capas para añadir una dimensión de profundidad a sus tablas, creando "cubos" tridimensionales. Las capas son, de hecho, muy similares a la anidación o el apilado; la diferencia principal es que sólo se puede mostrar una categoría cada vez. Por ejemplo, si utiliza la categoría *Edad* como la variable de la fila y *Género* como la variable de capa obtiene una tabla en la que se muestra la información para hombres y mujeres en capas diferentes de la tabla.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para eliminar cualquier selección anterior en el generador de tablas.
3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
4. Pulse en **Capas** en la parte superior del generador de tablas para mostrar la lista Capas.
5. Arrastre y suelte *Género* de la lista de variables a la lista Capas.

En este punto, puede observar que si añade una variable de capa no afectará visiblemente a la vista previa del panel de lienzo. Las variables de capa no afectan a la vista previa del panel de lienzo salvo que la variable de capa sea la variable de origen de estadísticos y cambie los estadísticos de resumen.

6. Pulse en **Aceptar** para crear la tabla.

		Count
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

Figura 18. Tabla con capas simple

A primera vista, esta tabla no parece muy diferente que una tabla simple de una variable categórica simple. La única diferencia es la presencia de la etiqueta *Género masculino* en la parte superior de la tabla.

7. Pulse dos veces en la tabla en la ventana Visor para activarla.
8. Ahora puede ver que la etiqueta *Género masculino* es una opción de la lista desplegable.
9. Pulse en la flecha hacia abajo en la lista desplegable para mostrar la lista completa de capas.
En esta tabla sólo hay una opción en la lista.
10. Seleccione *Género femenino* en la lista desplegable.

		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figura 19. Tabla con capas simple con diferentes capas

Dos variables de capa categóricas apiladas

Si tiene más de una variable categórica en las capas, puede apilar o anidar las variables de capa. De forma predeterminada, las variables de capa están apiladas. (Nota: si tiene una variable de capa de escala, las variables de capa sólo se pueden apilar).

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Si no tiene una categoría *Edad* en las filas y *Género* en las capas, siga las instrucciones anteriores para crear una tabla con capas.
3. Arrastre y suelte *Grado más alto* de la lista de variables en el lista Filas en *Género*.
Los dos botones de selección de la lista Capa en el grupo Capa de salida se activan. La selección predeterminada es **Mostrar cada categoría como una capa**. Es una operación equivalente a apilar.
4. Pulse en **Aceptar** para crear la tabla.
5. Pulse dos veces en la tabla en la ventana Visor para activarla.
6. Pulse en la flecha hacia abajo en la lista desplegable para mostrar la lista completa de capas.

Hay siete capas en la tabla: dos capas de las dos categorías *Género* y cinco capas de las cinco categorías *Grado más alto*. En las capas apiladas, el número total de capas es la suma del número de categorías de las variables de capa (incluyendo el total o subtotal de categorías que haya solicitado para las variables de capa).

Dos variables de capa categóricas anidadas

La anidación de variables de capas categóricas crea una capa diferente para cada combinación de categorías de variables de capa.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).

2. Si aún no lo ha hecho, siga las instrucciones anteriores para crear una tabla de capas apiladas.
3. En el grupo Capa de salida, seleccione **Mostrar cada combinación de categorías como una capa**. Es una operación equivalente a anidar.
4. Pulse en **Aceptar** para crear la tabla.
5. Pulse dos veces en la tabla en la ventana Visor para activarla.
6. Pulse en la flecha hacia abajo en la lista desplegable para mostrar la lista completa de capas.

Hay 10 capas en la tabla (tiene que desplazarse por la lista para poder verlas todas), una para cada combinación de *Género* y *Grado más alto*. En las capas anidadas, el número total de capas es el *producto* del número de categorías para cada variable de capa (en este caso, $5 \times 2 = 10$).

Impresión de tablas con capas

De forma predeterminada, sólo se imprime la capa visible actualmente. Para imprimir todas las capas de una tabla:

1. Pulse dos veces en la tabla en la ventana Visor para activarla.
2. En los menús de la ventana Visor, seleccione:
Formato > Propiedades de tabla...
3. Pulse en la pestaña **Impresión**.
4. Seleccione **Imprimir todas las capas**.

También puede guardar esta configuración como parte de TableLook, incluyendo el TableLook predeterminado.

Capítulo 4. Totales y subtotales de variables categóricas

En las tablas personalizadas puede incluir tanto los totales como los subtotales. Los totales y subtotales se pueden aplicar a las variables categóricas en cualquier nivel de anidación de la dimensión: fila, columna o capa.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Un total único para una sola variable

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
3. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
4. En el cuadro de diálogo Estadísticos de resumen, seleccione **% del N de columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
5. En la casilla Etiqueta de la lista Visualización, borre la etiqueta predeterminada y escriba Porcentaje.
6. Pulse en **Aplicar a la selección**.
7. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
8. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales.
9. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figura 20. Total único para una sola variable categórica

Lo que ve es lo que se totaliza

Los totales se basan en las categorías que se muestran en la tabla. Si decide excluir categorías de una tabla, los casos pertenecientes a estas categorías no se incluirán en los cálculos totales.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.

3. Pulse en la categoría con la etiqueta *Menos de 25* en la lista Etiqueta.
4. Pulse en la flecha hacia abajo que hay a la izquierda de la lista Excluir.
5. Pulse en la categoría con la etiqueta *65 o mayor* en la lista Etiqueta.
6. Vuelva a pulsar en la flecha hacia abajo que hay a la izquierda de la lista Excluir.
Las dos categorías se desplazarán de la lista Visualización a la lista Excluir.
7. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Percent
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	Total	2107	100.0%

Figura 21. Total en una tabla con categorías excluidas

El recuento total de esta tabla sólo asciende a 2.107, que se puede comparar con el total de 2.828 cuando se incluyen todas las categorías. Únicamente las categorías utilizadas en la tabla se han tenido en cuenta para el total. (El porcentaje total se mantiene en 100%, dado que todos los porcentajes se basan en el número total de casos que se han utilizado en la tabla y no en el número total de casos incluidos en el archivo de datos).

Posición de presentación de los totales

De forma predeterminada, los totales se presentan debajo de las categorías totalizadas. Puede cambiar la posición de presentación de los totales para mostrarlos antes de las categorías totalizadas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
3. En el grupo Totales y subgrupos aparecen, seleccione **Por encima de las categorías a las que se aplican**.
4. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Percent
Age category	Total	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

Figura 22. Total presentado antes de las categorías totalizadas

Totales de tablas anidadas

Dado que los totales se pueden aplicar a las variables categóricas en cualquier nivel de anidación, puede crear tablas que contengan totales de grupo en distintos niveles de anidación.

Totales de grupo

Los totales de las variables categóricas que están anidadas dentro de otras variables categóricas representan totales de grupo.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre y coloque *Sexo* a la izquierda de *Categoría de edad* en el panel de lienzo.

- Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
Antes de crear esta tabla, volveremos a colocar los totales debajo de las categorías totalizadas.
- En el grupo Totales y subgrupos aparecen, seleccione **Por debajo de las categorías a las que se aplican**.
- Pulse en **Aplicar** para guardar el ajuste y volver al generador de tablas.
- Pulse en **Aceptar** para crear la tabla.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%

Figura 23. Totales de categoría de edad dentro de categorías de Sexo

La tabla muestra dos totales de grupo: uno para hombres y otro para mujeres.

Totales finales

Los totales que se aplican a las variables anidadas siempre son totales de grupo y no totales finales. Para obtener totales de una tabla completa, aplique los totales a la variable en el nivel de anidación más exterior.

- Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- Pulse con el botón derecho del ratón *Género* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
- Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales.
- Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%
	Total	Age category	25 to 34	627	29.8%
			35 to 44	679	32.2%
			45 to 54	481	22.8%
			55 to 64	320	15.2%
			Total	2107	100.0%

Figura 24. Totales finales de tablas anidadas

Observe que el total final asciende ahora a 2.107 y no a 2.828. Dos categorías de edad siguen excluidas de la tabla, de modo que los casos que se incluyen en estas categorías se han excluido de todos los totales.

Totales de la variable de capas

Los totales de las variables de capas se presentan en la tabla como capas diferenciadas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. En el generador de tablas, pulse en **Capas** para que se muestre la lista Capas.
3. Arrastre y coloque *Sexo* desde el área de filas del panel de lienzo en la lista Capas.
Nota: no es necesario que especifique ahora los totales de *Sexo*, puesto que ya se han especificado anteriormente. La variable se puede mover entre dimensiones sin que su configuración se vea afectada.
4. Pulse en **Aceptar** para crear la tabla.
5. Pulse dos veces en la tabla del Visor para activarla.
6. Pulse en la flecha hacia abajo de la lista desplegable Capa para ver una lista de todas las capas de la tabla.

Hay tres capas en la tabla: *Sexo masculino*, *Sexo femenino* y *Total de sexo*.

Posición de presentación de los totales de capa

Respecto a los totales de la variable de capas, la posición de presentación (encima o debajo) de los totales determinará la posición de la capa para los totales. Por ejemplo, si especifica **Por encima de las categorías a las que se aplican** para el total de una variable de capas, la primera capa que se mostrará será la capa de los totales.

Subtotales

Puede incluir los subtotales para los subconjuntos de categorías de una variable. Por ejemplo, puede incluir los subtotales para las categorías de edad que representan a todos los encuestados en la encuesta de ejemplo menores y mayores de 45 años.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
 2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
 3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
 4. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
 5. Seleccione **3.00** en la lista Valore(s).
 6. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
 7. En el campo de texto Etiqueta, introduzca Subtotal < 45.
 8. Pulse en **Continuar**.
- Al hacer esto, se insertará una fila con el subtotal de las tres primeras categorías de edad.
9. Seleccione **6,00** en la lista Valor(es).
 10. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
 11. En el campo de texto Etiqueta, introduzca Subtotal 45+.
 12. Pulse en **Continuar**.

Nota importante: debe seleccionar la posición de presentación de los totales y los subtotales (**Por encima de las categorías a las que se aplican** o **Por debajo de las categorías a las que se aplican**) antes de definir cualquier subtotal. El cambio de posición de visualización afecta a todos los subtotales (no sólo al subtotal seleccionado actualmente) y también *cambia las categorías incluidas en los subtotales*.

13. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal < 45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

Figura 25. Subtotales para Categoría de edad

Lo que ve es lo que se subtotaliza

Al igual que los totales, los subtotales se basan en las categorías que se incluyen en la tabla.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.

Nota: el valor (y no la etiqueta) que se muestra para el primer subtotal es **1,00...3,00**, lo que indica que el subtotal incluye todos los valores de la lista comprendidos entre 1 y 3.

3. Seleccione **1,00** en la lista de valores (o pulse la etiqueta *Menos de 25*).
4. Pulse en la flecha hacia abajo que hay a la izquierda de la lista Excluir.

En este momento se encuentra excluida la primera categoría de edad y el valor que se muestra para el primer subtotal cambia a **2,00...3,00**, lo que indica que la categoría excluida no se incluirá en el subtotal debido a que los subtotales se basan en las categorías que sí están incluidas en la tabla. Cuando se excluye una categoría, se excluye automáticamente de los subtotales, de modo que no es posible, por ejemplo, mostrar únicamente los subtotales sin las categorías en las que están basados los subtotales.

Ocultación de categorías de subtotal

Si lo desea, puede suprimir la presentación de las categorías que definen un subtotal y hacer que se muestre sólo el subtotal; para ello, "agrupe" las categorías de forma efectiva, sin afectar a los datos subyacentes.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
3. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
4. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
5. Seleccione **3.00** en la lista Valore(s).
6. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
7. En el campo de texto Etiqueta, introduzca Menor que 45.
8. Seleccione (marque) **Ocultar categorías subtotalizadas desde la tabla**.
9. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las tres primeras categorías de edad.
10. Seleccione **6,00** en la lista Valor(es).
11. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
12. En el campo de texto Etiqueta, introduzca 45 o mayor.
13. Seleccione (marque) **Ocultar categorías**.
14. Pulse en **Continuar**.

15. Para incluir un total con los subtotales, seleccione (marque) **Total** en la parte inferior del cuadro de diálogo.
16. Pulse en **Aplicar**.
En el lienzo se refleja el hecho de que se mostrarán los subtotales pero no las categorías que los definen.
17. Pulse en **Aceptar** para generar la tabla.

		Count
Age category	Less than 45	1548
	45 or older	1280
	Total	2828

Figura 26. Tabla que muestra sólo los valores totales y subtotales

Subtotales de la variable de capas

Al igual que los totales, los subtotales de las variables de capas se presentan en la tabla como capas diferenciadas. Básicamente, los subtotales se tratan como categorías. Cada categoría constituye una capa diferente de la tabla y el orden de presentación de las categorías de capa queda determinado por el orden de categorías especificado en el cuadro de diálogo Categorías y totales, incluida la posición de presentación de las categorías de subtotal.

Capítulo 5. Categorías calculadas de variables categóricas

Puede incluir categorías calculadas en tablas personalizadas. Hay nuevas categorías que se calculan a partir de otras categorías de la misma variable en cualquier nivel de anidación en cualquier dimensión: fila, columna o capa. Por ejemplo, puede incluir una categoría calculada que muestre la diferencia entre dos categorías.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Categoría calculada simple

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre y suelte la categoría *Edad* de la lista de variables en el área Filas del panel de lienzo.
3. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
4. Seleccione **3,00** en la lista Valore(s).
5. Pulse en **Añadir categoría** para mostrar el cuadro de diálogo Definir categoría calculada.
6. En el campo de texto Etiqueta de Categoría calculada, introduzca Menor que 45.
7. Seleccione **Menos de 25 (1,00)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada. [1] aparecerá en la expresión.
8. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
9. Seleccione **De 25 a 34 (2,00)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
10. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
11. Seleccione **De 35 a 44 (3,00)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
12. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las tres primeras categorías de edad.
13. Seleccione **5,00** en la lista Valor(es).
14. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
15. En el campo de texto Etiqueta, introduzca Menor que 65.
16. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las cinco primeras categorías.
17. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Less than 45	1548
	45 to 54	481
	55 to 64	320
	Less than 65	2349
	65 or older	479

Figura 27. Categoría calculada con subtotal

La tabla incluye una categoría calculada (*Menor que 45*) y un subtotal (*Menor que 65*). El subtotal incluye categorías también incluidas en la categoría calculada. No ha podido crear la misma tabla con subtotales únicamente, porque los subtotales no pueden compartir las mismas categorías.

Ocultación de categorías en una categoría calculada

Al igual que con los subtotales, puede eliminar la visualización de las categorías que se utilizan en una expresión de categoría calculada y mostrar únicamente la categoría calculada en sí. El siguiente ejemplo se basa en el anterior.

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
3. Seleccione la categoría calculada *Menor que 45* en la lista Valor(es).
4. Pulse en **Edición** para mostrar el cuadro de diálogo Definir categoría calculada.
5. Seleccione **Ocultar categorías empleadas en la expresión de la tabla**.
6. Pulse en **Continuar**.
7. Seleccione el subtotal *Menor que 65* en la lista Valor(es).
8. Pulse en **Edición** para mostrar el cuadro de diálogo Definir subtotal.
9. Seleccione **Ocultar categorías subtotalizadas desde la tabla**.
10. Pulse en **Continuar**.
11. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count
Age category	Less than 45	1548
	Less than 65	2349
	65 or older	479

Figura 28. Categoría calculada con subtotal y categorías ocultas

Como en el ejemplo anterior, la tabla incluye una categoría calculada y un subtotal. Pero en este caso, cada categoría están ocultas, de forma que sólo se muestran estos totales.

Referencia de subtotales en una categoría calculada

Puede incluir subtotales en una expresión de categoría calculada.

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.

3. En el generador de tablas, arrastre y suelte *Cargo profesional* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y suelte *Estado civil* de la lista de variables en el área Columnas.
5. Pulse con el botón derecho del ratón *Cargo profesional* en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
6. Seleccione **2** en la lista Valor(es).
7. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
8. En el campo de texto Etiqueta, introduzca Trabajo.
9. Seleccione **Ocultar categorías subtotalizadas desde la tabla**.
10. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las dos primeras categorías de trabajo.
11. Seleccione **8** en la lista Valor(es).
12. Pulse en **Añadir subtotal** para mostrar el cuadro de diálogo Definir subtotal.
13. En el campo de texto Etiqueta, introduzca Sin trabajo.
14. Seleccione **Ocultar categorías subtotalizadas**.
15. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal del resto de las categorías de trabajo.
16. Seleccione el subtotal *Sin trabajo* en la lista Valor(es).
17. Pulse en **Añadir categoría** para mostrar el cuadro de diálogo Definir categoría calculada.
18. En el campo de texto Etiqueta de Categoría calculada, introduzca Trabajo / Sin trabajo.
19. Seleccione **Trabajo (Trabajo 1)** en la lista de totales y subtotales y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
20. Pulse en el botón de operador de división (/) en el cuadro de diálogo (o pulse la tecla / del teclado).
21. Seleccione **Sin trabajo (Sin trabajo 2)** en la lista de totales y subtotales y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
De forma predeterminada, la categoría calculada utiliza el mismo formato que el estadístico de variable, que en este caso es Recuento. Como queremos mostrar los decimales como resultado de dividir la expresión de la categoría calculada y el formato predefinido de Recuento no incluye decimales, es necesario modificar el formato.
22. Pulse en la pestaña Mostrar formatos.
23. Cambie el ajuste de decimales de Recuento a **2**.
24. Pulse en **Continuar**.
25. Pulse en **Aplicar** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Count	Count	Count	Count	Count
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Working / Not Working	2.14	.29	2.84	2.58	2.92

Figura 29. Categoría calculada mostrando el porcentaje de subtotales

La tabla incluye dos subtotales y una categoría calculada. La categoría calculada muestra el porcentaje de subtotales para que pueda comparar fácilmente los grupos que representa cada subtotal. Hay una proporción mucho menor de encuestados viudos con y sin trabajo en comparación con el resto de grupos. Además, hay un porcentaje ligeramente inferior de encuestados casados, seguramente como resultado de que las esposas dejan de trabajar para quedarse en casa cuidando a los hijos.

Uso de categorías calculadas para mostrar Subtotales no exhaustivos

Los subtotales son exhaustivos. Es decir, todos los subtotales de la tabla incluyen todos los valores superiores o inferiores a sus posiciones en la tabla. Las categorías calculadas, por otra parte, no son exhaustivas y le permiten sumar una mezcla de categorías en una tabla.

1. Seleccione en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
3. En el generador de tablas, arrastre *Se considera liberal o conservador*, desde la lista de variables, y colóquela en el área Filas del panel de lienzo.
4. Pulse con el botón derecho del ratón *Se considera liberal o conservador* en el panel de lienzo y elija **Categorías y totales** en el menú contextual emergente.
5. Seleccione **3** en la lista Valor(es).
6. Pulse en **Añadir categoría** para mostrar el cuadro de diálogo Definir categoría calculada.
7. En el campo de texto Etiqueta de Categoría calculada, introduzca Subtotal de liberales. Tenga en cuenta que hay cuatro espacios antes del texto. Estos espacios se utilizan como sangría en la tabla resultante.
8. Seleccione **Extremadamente liberal (1)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
9. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
10. Seleccione **Liberal (2)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
11. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
12. Seleccione **Ligeramente liberal (3)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
13. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las tres primeras categorías de liberal.
14. Seleccione **7** en la lista Valor(es).
15. Pulse en **Añadir categoría** para mostrar el cuadro de diálogo Definir categoría calculada.
16. En el campo de texto Etiqueta de Categoría calculada, introduzca Subtotal de conservadores. Tenga en cuenta que hay cuatro espacios antes del texto. Estos espacios se utilizan como sangría en la tabla resultante.
17. Seleccione **Ligeramente conservador (5)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
18. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
19. Seleccione **Conservador (6)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
20. Pulse en el botón de operador de suma (+) en el cuadro de diálogo (o pulse la tecla + del teclado).
21. Seleccione **Extremadamente conservador (7)** en la lista de categorías y pulse el botón de flecha para copiarlo en el cuadro de texto Expresión para categoría calculada.
22. Pulse en **Continuar**.
Al hacer esto, se insertará una fila con el subtotal de las categorías de conservadores.
23. Pulse en **Aplicar y**, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

Figura 30. Categorías calculadas mostrando subtotales no exhaustivos

La tabla incluye dos categorías calculadas que no incluyen todas las categorías mostradas en la tabla. La categoría *Moderado* no se incluye en ninguna categoría calculada. No puede crear la misma tabla con subtotales porque los subtotales son exhaustivos.

Capítulo 6. Tablas para variables con categorías compartidas

Las encuestas a menudo contienen muchas preguntas con un conjunto común de posibles respuestas. Por ejemplo, nuestra encuesta de ejemplo contiene un determinado número de variables acerca de la confianza en varias instituciones y servicios públicos y privados, todas ellas con el mismo conjunto de categorías de respuesta: 1 = *Mucha*, 2 = *Alguna*, and 3 = *Poca*. Puede utilizar la apilación para mostrar estas variables relacionadas en la misma tabla y puede hacer que aparezcan las categorías de respuesta compartidas en las columnas de la tabla. Estas características también están disponibles si utiliza categorías calculadas, con la provisión de que cualquier etiqueta y expresión de categoría calculada son las mismas en todas las variables.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figura 31. Tabla de variables que comparten categorías

Nota: en la versión anterior de Tablas personalizadas, se las conocía con el nombre de "tabla de frecuencias".

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Tabla de recuentos

1. Elija en los menús:

Analizar > Tablas > Tablas personalizadas...

2. En la lista de variables del generador de tablas, pulse en *Confianza en los bancos...* y, a continuación, mientras mantiene pulsada la tecla Mayús, pulse en *Confianza en la televisión* para seleccionar todas las variables de "confianza". (*Nota:* para realizar este paso, se supone que las etiquetas de variable aparecen ordenadas por orden alfabético en la lista de variables y no por orden de archivo).

3. Arrastre y coloque las seis variables de confianza al área Filas del panel de lienzo.

Esto apilará las variables en la dimensión de las filas. De forma predeterminada, también se muestran en las filas las etiquetas de las categorías para cada variable, dando como resultado una tabla muy larga y estrecha (6 variables x 3 categorías = 18 filas); pero como las seis variables comparten la definición de las mismas etiquetas de categoría (etiquetas de valor), se pueden colocar las etiquetas de las categorías en la dimensión de las columnas.

4. En la lista desplegable Posición de categoría, seleccione **Etiquetas de fila en las columnas**.

Ahora la tabla sólo tiene seis filas, una para cada una de las variables apiladas, y las categorías que tienen definidas se convierten en las columnas de la tabla.

- Antes de crear la tabla, seleccione (marque) **Ocult.** para Posición en el grupo Estadísticos de resumen, ya que la etiqueta de estadísticos de resumen *Recuento* no es realmente necesaria.
- Pulse en **Aceptar** para crear la tabla.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figura 32. Tabla de variables de filas apiladas con las etiquetas de categorías compartidas en las columnas

En lugar de representar las variables en las filas y las categorías en las columnas, se puede crear una tabla con las variables apiladas en las columnas y mostrando las categorías en las filas. Esta sería una mejor opción si hubiera más categorías que variables, mientras que en nuestro ejemplo hay más variables que categorías.

Tabla de porcentajes

Para una tabla con las variables apiladas en las filas y las categorías mostradas en las columnas, el porcentaje más interpretable es el que representa los porcentajes de fila (o al menos el más fácil de entender). (En una tabla con las variables acodadas en las columnas y mostrando las categorías en las filas, probablemente los más deseables serían los porcentajes de columna).

- Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- Pulse con el botón derecho del ratón cualquiera de las variables de confianza de la vista previa de la tabla del panel de lienzo y elija **Estadísticos de resumen** en el menú emergente.
- Seleccione **% del N de la fila** en la lista Estadísticos y pulse el botón de flecha para añadirlo a la lista Visualización.
- Pulse cualquier casilla de la fila *Recuento* en la lista Visualización y pulse el botón de flecha para devolverlo a la lista Estadísticos y suprimirlo de la lista Visualización.
- Pulse en **Aplicar a todo** para aplicar la modificación del estadístico de resumen a todas las variables apiladas en la tabla.

Nota: si la vista previa de la tabla no se parece a esta imagen, probablemente haya pulsado en **Aplicar a selección**, que sólo aplica el nuevo estadístico de resumen a la variable seleccionada, en lugar de pulsar **Aplicar a todo**. En este ejemplo, esto daría como resultado dos columnas para cada categoría: una con los marcadores de posición de los recuentos mostrados para todas las demás variables y una con un marcador de posición de porcentaje de fila mostrado para la variable seleccionada. Ésta es exactamente la tabla que se generaría pero *no* la que deseamos crear en este ejemplo.

- Pulse en **Aceptar** para crear la tabla.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

Figura 33. Tabla de porcentajes de fila para las variables apiladas en filas y las categorías mostradas en las columnas

Nota: puede incluir cualquier número de estadísticos de resumen en una tabla de variables con categorías compartidas. Nuestros ejemplos sólo muestran uno cada vez para simplificarlos.

Control de categorías y totales

Se pueden crear tablas con categorías situadas en la dimensión opuesta a las variables sólo si todas las variables de la tabla tienen las mismas categorías, mostradas en el mismo orden. Incluyendo totales, subtotales y cualquier otro ajuste de categorías que realice. Esto significa que cualquier modificación que realice en el cuadro de diálogo Categorías y totales se debe realizar para todas las variables que comparten categorías en la tabla.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón la primera variable de confianza en la vista previa de la tabla en el panel de lienzo y elija **Categorías y totales** en el menú emergente.
3. Seleccione (marque) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.
Lo primero que probablemente observará es que las etiquetas de las categorías se han desplazado de las columnas de nuevo a las filas. Puede que también haya observado que el control Posición de categoría está ahora desactivado. Esto se debe a que las variables ya no comparten el mismo conjunto exacto de "categorías". Una de las variables tiene ahora una categoría de totales.
4. Pulse con el botón derecho del ratón en cualquiera de las variables de confianza del panel de lienzo y seleccione **Seleccionar todas las variables de fila** del menú emergente; o mantenga pulsada la tecla Ctrl mientras pulsa con el ratón en cada variable apilada en el panel de lienzo hasta que estén todas seleccionadas (es posible que deba desplazar el panel hacia abajo o expandir la ventana del generador de tablas).
5. En el grupo Definir, pulse en **Categorías y totales**.
6. Si **Total** aún no está seleccionado (marcado) en el cuadro de diálogo Categorías y totales, hágalo ahora y, a continuación, pulse en **Aplicar**.
7. La lista desplegable Posición de categoría debe estar activada de nuevo, ya que ahora todas las variables cuentan con la categoría de total adicional, por lo que se puede seleccionar **Etiquetas de fila en las columnas**.
8. Pulse en **Aceptar** para crear la tabla.

	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

Figura 34. Tabla de porcentajes de fila para variables apiladas en las filas, con las categorías y totales mostrados en las columnas

Anidación en tablas con categorías compartidas

En las tablas anidadas, si desea mostrar las etiquetas de las categorías en la dimensión opuesta, las variables apiladas con las categorías compartidas deben estar en el nivel de anidación más al interior de su dimensión.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre *Sexo* desde la lista de variables y colóquela en el lado izquierdo del área Filas.
En la vista previa de la tabla, las variables apiladas con las categorías compartidas aparecen ahora anidadas dentro de las categorías de sexo.
3. Ahora arrastre y coloque *Sexo* a la derecha de una de las variables de confianza apiladas en la vista previa de la tabla.

Una vez más, las etiquetas de las categorías se han vuelto a situar en la dimensión de las filas y el control Posición de categoría vuelve a aparecer desactivado. En este momento, dispone de una variable apilada que también tiene *Sexo* anidada en su interior, mientras que las demás variables apiladas no contienen ninguna variable anidada. Puede añadir *Sexo* como una variable anidada a cada una de las variables apiladas, pero, a continuación, el desplazamiento de las etiquetas de filas a las columnas dará como resultado que las etiquetas de categoría para *Sexo* se muestren en las columnas, y no las etiquetas de categorías para las variables apiladas con las categorías compartidas. Esto se debe a que *Sexo* será ahora la variable anidada más al interior y el cambio de la posición de categoría siempre afecta a la variable anidada más al interior.

Capítulo 7. Estadísticos de resumen

Los **estadísticos de resumen** incluyen desde simples recuentos para las variables categóricas hasta medidas de dispersión, como el error estándar de la media para las variables de escala. *No* incluye las pruebas de significación disponibles en la pestaña Estadísticos de contraste del cuadro de diálogo Tablas personalizadas. Consulte el tema “Estadísticos de contraste” en la página 63 para obtener más información.

Los estadísticos de resumen disponibles para variables categóricas y conjuntos de respuestas múltiples incluyen recuentos y diversos tipos de porcentajes, entre ellos:

- Porcentajes de fila
- Porcentajes de columna
- Porcentajes de subtabla
- Porcentajes de tabla
- Porcentaje del N válido de la fila

Además de los estadísticos de resumen disponibles para las variables categóricas, los estadísticos de resumen para las variables de escala y los resúmenes de los totales personalizados de las variables categóricas incluyen:

- Media
- Mediana
- Percentiles
- Suma
- Desviación estándar
- Rango
- Valores mínimo y máximo

También están disponibles estadísticos de resumen adicionales para los conjuntos de respuestas múltiples. También está disponible una lista completa de estadísticos de resumen. Consulte el tema “Estadísticos de resumen” en la página 6 para obtener más información.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Variable origen de los estadísticos de resumen

Los estadísticos de resumen disponibles dependen del nivel de medición de la variable origen de los estadísticos de resumen. El origen de los estadísticos de resumen (la variable en la que se basan los estadísticos de resumen) se determina por:

- **Nivel de medición.** Si una tabla (o una sección de la tabla apilada) contiene una variable de escala, los estadísticos de resumen se basan en la variable de escala.
- **Orden de selección de variable.** La dimensión predeterminada de origen de estadísticos (fila o columna) para las variables categóricas se basa en el orden en el que arrastre y suelte las variables en

el panel de lienzo. Por ejemplo, si primero arrastra una variable al área de filas, la dimensión de la fila es la dimensión predeterminada de origen de estadísticos.

- **Anidación.** Para las variables categóricas, los estadísticos de resumen se basan en la variable más interior de la dimensión de origen de estadísticos.

Una tabla anidada puede tener variables de origen de estadísticos de resumen múltiples (de escala y categóricas), pero cada sección de tabla tiene un único origen de estadísticos de resumen.

Origen de los estadísticos de resumen para variables categóricas

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre *Categoría de edad* desde la lista de variables y colóquela en el área Filas del panel de lienzo.
3. Pulse con el botón derecho del ratón *Categoría de edad* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente. (Al ser ésta la única variable de la tabla, será la variable origen de los estadísticos).
4. En el cuadro de diálogo Estadísticos de resumen, seleccione *% del N de columna* en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
5. Pulse en **Aplicar a la selección**.
6. En el generador de tablas, arrastre *Obtener noticias de Internet* y colóquela a la derecha de *Categoría de edad* en el panel de lienzo.
7. Vuelva a pulsar con el botón derecho del ratón sobre *Categoría de edad* en el panel del lienzo. El elemento **Estadísticos de resumen** del menú emergente aparecerá ahora desactivado, ya que *Categoría de edad* no es la variable anidada más al interior en la dimensión origen de los estadísticos.
8. Pulse con el botón derecho del ratón en *Obtener noticias de Internet* en el panel de lienzo. El elemento **Estadísticos de resumen** aparecerá ahora activado porque ahora es ésta la variable origen de los estadísticos, al ser la variable anidada más al interior en la dimensión origen de los estadísticos. (Al tener la tabla una única dimensión (las filas), ésta será la dimensión origen de los estadísticos).
9. Arrastre *Obtener noticias de Internet* desde el área Filas y colóquela en el panel de lienzo del área Columnas.
10. Vuelva a pulsar con el botón derecho del ratón en *Obtener noticias de Internet* en el panel de lienzo. El elemento **Estadísticos de resumen** del menú emergente aparecerá ahora desactivado, ya que la variable habrá dejado de estar en la dimensión origen de los estadísticos.

Categoría de edad vuelve a ser la variable origen de los estadísticos, ya que en el caso de variables categóricas, la dimensión predeterminada de origen de los estadísticos es la primera dimensión en la que se colocan las variables al crear la tabla. En este ejemplo, lo primero que se hizo fue colocar variables en la dimensión de las filas. Por tanto, la dimensión de las filas es la dimensión predeterminada de origen de los estadísticos y al ser ahora *Categoría de edad* la única variable en esta dimensión, será la variable origen de los estadísticos.

Origen de estadísticos de resumen para variables de escala

1. Arrastre la variable de escala *Horas diarias viendo TV* y colóquela a la izquierda de *Categoría de edad* en el área Filas del panel de lienzo.
Lo primero que puede observar es que los resúmenes *Recuento* y *% del N de columna* se han reemplazado por *Media*, y si pulsa con el botón derecho del ratón en *Horas diarias viendo TV* en el panel de lienzo, comprobará que ahora es la variable origen de los estadísticos de resumen. En el caso de una tabla con una variable de escala, la variable de escala es siempre la variable origen de los estadísticos, independientemente de su nivel de anidación o su dimensión de ubicación. Además la media es el estadístico de resumen predeterminado para las variables de escala.
2. Arrastre *Horas diarias viendo TV* desde el área Filas y colóquela en el área Columnas que hay encima de *Obtener noticias de Internet*.

- Pulse con el botón derecho del ratón en *Horas diarias viendo TV* y seleccione **Estadísticos de resumen** en el menú contextual emergente. (Seguirá siendo la variable origen de los estadísticos, aunque se desplace a una dimensión diferente).
- En el cuadro de diálogo Estadísticos de resumen, pulse en la casilla **Formato** correspondiente a la media en la lista Visualización y seleccione **nnnn** en la lista desplegable Formato. (Puede que tenga que desplazarse hacia arriba por la lista para llegar hasta este elemento).
- En la casilla Decimales, introduzca 2.
- Pulse en **Aplicar a la selección**.
La vista previa de la tabla que aparece en el panel de lienzo indica ahora que los valores de la media aparecerán con dos decimales.
- Pulse en **Aceptar** para crear la tabla.

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Figura 35. Variable de escala resumida dentro de una tabulación cruzadas de variables categóricas

Variables apiladas

Teniendo en cuenta que una tabla apilada puede incluir varias variables origen de los estadísticos y que se pueden especificar estadísticos de resumen diferentes para cada una de estas variables origen, hay que señalar algunas consideraciones especiales sobre la especificación de los estadísticos de resumen en las tablas apiladas.

- Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
- Pulse en *Obtener noticias de Internet* en la lista de variables y, a continuación, manteniendo pulsada la tecla Mayús pulse en *Obtener noticias de televisión* en la lista de variables para seleccionar todas las variables de "noticias". (Nota: para realizar este paso, se supone que las etiquetas de variable aparecen ordenadas por orden alfabético en la lista de variables y no por orden de archivo).
- Arrastre las cinco variables de noticias y colóquelas en el área Filas del panel de lienzo.
Las cinco variables de noticias aparecen apiladas en la dimensión de las filas.
- Pulse en *Obtener noticias de Internet* en el panel de lienzo, para que sólo aparezca seleccionada esa variable.
- Ahora pulse con el botón derecho del ratón en *Obtener noticias de Internet* y seleccione **Estadísticos de resumen** en el menú emergente.
- En el cuadro de diálogo Estadísticos de resumen, seleccione *% del N de columna* en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización. (Puede utilizar la flecha para desplazar los estadísticos seleccionados de la lista Estadísticos a la lista Visualización o bien puede arrastrar los estadísticos seleccionados de la lista Estadísticos y colocarlos en la lista Visualización).
- A continuación, pulse en **Aplicar a selección**.

Se ha añadido una columna para los porcentajes de columna, pero la vista previa de la tabla que aparece en el panel de lienzo indica que se mostrarán los porcentajes de columna correspondientes a una única variable, ya que en una tabla apilada hay varias variables origen de estadísticos y cada

una de ellas puede tener diferentes estadísticos de resumen. En este ejemplo, sin embargo, queremos mostrar los mismos estadísticos de resumen para todas las variables.

9. Pulse con el botón derecho del ratón en *Obtener noticias de periódicos* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
10. En el cuadro de diálogo Estadísticos de resumen, seleccione *% del N de columna* en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
11. A continuación, pulse en **Aplicar a todo**.

Ahora la vista previa de la tabla indica que se mostrarán los porcentajes de columna de todas las variables apiladas.

Estadísticos de resumen del total personalizados para variables categóricas

Para las variables categóricas origen de estadísticos, puede incluir estadísticos de resumen del total personalizados que sean diferentes de los estadísticos que aparecen para las categorías de la variable. Por ejemplo, para una variable ordinal, puede hacer que aparezcan los porcentajes de cada categoría y la media o la mediana para el estadístico de resumen del total personalizado.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
3. Pulse en *Confianza en la prensa* en la lista de variables y, a continuación, manteniendo pulsada la tecla Ctrl pulse en *Confianza en la TV* para seleccionar ambas variables.
4. Arrastre las dos variables y colóquelas en el área Filas del panel de lienzo. Esto apilará las dos variables en la dimensión de las filas.
5. Pulse con el botón derecho del ratón en cualquiera de estas dos variables en el panel de lienzo y seleccione **Seleccionar todas las variables de fila** en el menú emergente. (Probablemente ambas ya estén seleccionadas, pero es preferible asegurarse).
6. Vuelva a pulsar con el botón derecho del ratón en la variable y seleccione **Categorías y totales** en el menú emergente.
7. En el cuadro de diálogo Categorías y totales, pulse (seleccione) **Total** y, a continuación, **Aplicar**.
La vista previa de la tabla del panel de lienzo ahora mostrará una fila total para ambas variables. Para mostrar los estadísticos de resumen del total personalizados, es necesario especificar los totales o los subtotales de la tabla.
8. Pulse con el botón derecho del ratón en cualquiera de las dos variables, en el panel de lienzo, y seleccione **Estadísticos de resumen** en el menú emergente.
9. En el cuadro de diálogo Estadísticos de resumen, pulse en *Recuento* en la lista Visualización y pulse en la flecha para devolverlo a la lista Estadísticos, suprimiéndolo de la lista Visualización.
10. Pulse en *% del N de columna* en la lista Estadísticos y pulse el botón de flecha para añadirlo a la lista Visualización.
11. Pulse en (marque) **Estadísticos de resumen personalizados para totales y subtotales**.
12. Pulse en *Recuento* en la lista Visualización de resumen personalizado y pulse en la flecha para devolverlo a la lista de estadísticos de resumen personalizados y suprimirlo de la lista Visualización.
13. Pulse en *Media* en la lista de estadísticos de resumen personalizados y pulse en la flecha para moverlo a la lista de visualización.
14. Pulse en la casilla **Formato** para la media en la lista Visualización y seleccione **nnnn** en la lista desplegable de formatos. (Puede que tenga que desplazarse hacia arriba por la lista para llegar hasta este elemento).
15. En la casilla Decimales, introduzca 2.
16. Pulse en **Aplicar a todo** para aplicar esta configuración a ambas variables de la tabla.

Se ha añadido una nueva columna para el estadístico de resumen del total personalizado, que tal vez no sea exactamente lo que quiere, ya que la vista previa que aparece en el panel de lienzo indica claramente que se obtendrá una tabla con numerosas casillas vacías.

- En el generador de tablas, en el grupo Estadísticos de resumen, seleccione **Filas** en la lista desplegable Posición.

Con esto se moverán todos los estadísticos de resumen a la dimensión de las filas, mostrando todos los estadísticos de resumen en una sola columna de la tabla.

- Pulse en **Aceptar** para crear la tabla.

Confidence in press	A great deal	Column N %	9.5%
	Only some	Column N %	47.2%
	Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	A great deal	Column N %	10.4%
	Only some	Column N %	49.9%
	Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figura 36. Variables categóricas con estadísticos de resumen del total personalizados

Presentación de los valores de categoría

Tan sólo hay un pequeño problema con la tabla anterior, puede resultar difícil interpretar el valor medio sin conocer los valores de las categorías subyacentes en los que se basan. ¿La media de 2,34 es un valor entre *Mucha* y *Alguna* o un valor entre *Alguna* y *Poca*?

Aunque no podemos tratar este problema directamente en las tablas personalizadas, podemos tratarlo de una manera más general.

- Seleccione en los menús:

Editar > Opciones...

- En el cuadro de diálogo Opciones, pulse en la pestaña **Etiquetas de los resultados**.
- En el grupo Etiquetado de tablas pivote, seleccione **Valores y etiquetas** en la lista desplegable **Los valores de las variables se muestran en las etiquetas como**.
- Pulse en **Aceptar** para guardar este ajuste.
- Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas) y pulse en **Aceptar** para volver a crear la tabla.

Confidence in press	1 A great deal	Column N %	9.5%
	2 Only some	Column N %	47.2%
	3 Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	1 A great deal	Column N %	10.4%
	2 Only some	Column N %	49.9%
	3 Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figura 37. Valores y etiquetas mostrados para las categorías de la variable

Los valores de categoría indican claramente que una media de 2,34 representa un valor entre *Alguna* y *Poca*. Al aparecer los valores de categoría en la tabla, resulta mucho más fácil interpretar el valor de los estadísticos de resumen del total personalizados, como por ejemplo la media.

Este ajuste de presentación es un ajuste global que afecta a todos los resultados de tablas pivote de todos los procedimientos y se mantendrá en todas las sesiones hasta que se cambie. Para cambiar el ajuste de manera que se vuelvan a mostrar sólo las etiquetas de valor:

- Seleccione en los menús:

Editar > Opciones...

7. En el cuadro de diálogo Opciones, pulse en la pestaña **Etiquetas de los resultados**.
8. En el grupo Etiquetado de tablas pivote, seleccione **Etiquetas** en la lista desplegable **Los valores de las variables se muestran en las etiquetas como**.
9. Pulse en **Aceptar** para guardar este ajuste.

Capítulo 8. Resumen de variables de escala

Resumen de variables de escala

Para las variables de escala, se dispone de una amplia gama de estadísticos de resumen. Además de las frecuencias y los porcentajes, disponibles para las variables categóricas, para las variables de escala también se incluyen los siguientes estadísticos de resumen:

- Media
- Mediana
- Percentiles
- Suma
- Desviación estándar
- Rango
- Valores mínimo y máximo

Consulte el tema “Estadísticos de resumen para variables de escala y totales categóricos personalizados” en la página 9 para obtener más información.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de presentación de la lista de variables se especifican en la pestaña General del cuadro de diálogo Opciones (menú Edición, Opciones).

VARIABLES DE ESCALA APILADAS

Puede resumir varias variables de escala en una única tabla, apilándolas dentro de la propia tabla.

1. Seleccione en los menús:

Analizar > Tablas > Tablas personalizadas...

2. En el generador de tablas, pulse en *Edad del encuestado*, en la lista de variables, y pulse Ctrl y el botón del ratón para seleccionar *Años de escolarización* y *Horas diarias viendo TV* para elegir las tres variables.

3. Arrastre y coloque las tres variables seleccionadas en el área Filas del panel de lienzo.

Las tres variables aparecen apiladas en la dimensión de las filas. Como las tres variables son de escala, no se mostrará ninguna categoría y la media será el estadístico de resumen predeterminado.

4. Pulse en **Aceptar** para crear la tabla.

	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Figura 38. Tabla de los valores promedio de las variables de escala apiladas

Múltiples estadísticos de resumen

En las variables de escala, de forma predeterminada se muestra la media; sin embargo, puede elegir otros estadísticos de resumen para las variables de escala y puede representar más de un estadístico de resumen.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en una de las tres variables de escala de la vista previa de la tabla del panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
3. En el cuadro de diálogo Estadísticos de resumen, seleccione *Mediana* en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización. (Puede utilizar la flecha para desplazar los estadísticos seleccionados de la lista Estadísticos a la lista Visualización o arrastrar y colocar en la lista Visualización los estadísticos seleccionados en la lista Estadísticos).
4. Pulse en la casilla **Formato** para la mediana en la lista Visualización y seleccione **nnnn** en la lista desplegable de formatos.
5. En la casilla Decimales, escriba 1.
6. Realice los mismos cambios en la lista Visualización para la media.
7. Pulse en **Aplicar a todo** para aplicar estos cambios a las tres variables.
8. Pulse en **Aceptar** en el generador de tablas para crear la tabla.

	Mean	Median
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

Figura 39. Media y mediana visualizadas en la tabla de variables de escala apiladas

Recuento, N válido y Valores perdidos

A menudo es útil mostrar el número de casos que ha sido utilizado para calcular los estadísticos, como la media, y puede suponer (razonablemente) que el estadístico *Recuento* le puede proporcionar dicha información. Sin embargo, cuando hay valores perdidos, no va a obtener con ello una base precisa de referencia respecto a los casos utilizados. Utilice el *N válido* para obtener una base de cálculo de los casos que sea precisa.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en una de las tres variables de escala de la vista previa de la tabla del panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
3. En el cuadro de diálogo Estadísticos de resumen, seleccione **Recuento** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
4. A continuación, seleccione **N válido** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
5. Pulse en **Aplicar a todo** para aplicar estos cambios a las tres variables.
6. Pulse en **Aceptar** en el generador de tablas para crear la tabla.

	Mean	Median	Count	Valid N
Age of respondent	45.6	42.0	2832	2828
Highest year of school completed	13.3	13.0	2832	2820
Hours per day watching TV	2.9	2.0	2832	2337

Figura 40. Recuento frente a N válido

Para las tres variables, el valor de *Recuento* es el mismo: 2.832. Es el número total de casos en el archivo de datos, y no es una casualidad. Como las variables de escala no están anidadas dentro de ninguna variable categórica, *Recuento* representa simplemente el número total de casos del archivo de datos.

Sin embargo, el valor del *N válido* es diferente para cada variable y para la variable *Horas diarias viendo TV* el valor es bastante diferente del *Recuento*. Esto se debe a que esta variable contiene una gran cantidad de **valores perdidos**; esto es, casos cuyo valor no está registrado en la variable o valores definidos de forma que representen datos perdidos (como el código 99 para representar *No procede*, en el caso de los hombres, al embarazo).

7. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
8. Pulse con el botón derecho del ratón en una de las tres variables de escala de la vista previa de la tabla del panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
9. En el cuadro de diálogo Estadísticos de resumen, seleccione **N válido**, en la lista Visualización, y pulse el botón de flecha para devolverlo a la lista Estadísticos, eliminándolo de la lista Visualización.
10. Seleccione **Recuento**, en la lista Visualización, y pulse el botón de flecha para devolverlo a la lista Estadísticos, eliminándolo de la lista Visualización.
11. Seleccione **Perdidos**, en la lista Estadísticos, y pulse el botón de flecha para añadirlo a la lista Visualización.
12. Pulse en **Aplicar a todo** para aplicar estos cambios a las tres variables.
13. Pulse en **Aceptar** en el generador de tablas para crear la tabla.

	Mean	Median	Missing
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

Figura 41. Número de valores perdidos, mostrados en la tabla de estadísticos de resumen de escala

La tabla muestra ahora el número de valores perdidos para cada variable de escala. De esta forma, es bastante obvio que *Horas diarias viendo TV* tiene una gran cantidad de valores perdidos, mientras que las otras dos variables tienen muy pocos. Este hecho puede ser un factor a tener en cuenta antes de confiar demasiado en los valores de resumen de esa variable.

Diferentes resúmenes para distintas variables

Además de mostrar varios estadísticos de resumen, en una tabla apilada, también puede mostrar diferentes estadísticos de resumen para las distintas variables de escala. Por ejemplo, la tabla anterior revelaba que sólo una de las tres variables tiene una gran cantidad de valores perdidos; por tanto, puede mostrar el número de valores perdidos solamente para esa variable.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en *Edad del encuestado* en la vista previa de la tabla del panel de lienzo y, a continuación pulse Ctrl y el botón del ratón para seleccionar *Años de escolarización* y elegir las dos variables.
3. Pulse con el botón derecho del ratón en una de las dos variables y seleccione **Estadísticos de resumen** en el menú emergente.
4. En el cuadro de diálogo Estadísticos de resumen, seleccione **Perdidos** en la lista Visualización y pulse el botón de flecha para devolverlo a la lista Estadísticos y suprimirlo de la lista Visualización.
5. Pulse en **Aplicar a selección** para aplicar el cambio sólo a las dos variables seleccionadas.

Los marcadores de posición de las casillas de datos de la tabla indican que sólo se visualizará el número de valores perdidos para *Horas diarias viendo TV*.

6. Pulse en **Aceptar** para crear la tabla.

	Mean	Median	Missing
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

Figura 42. Tabla con estadísticos de resumen diferentes para las distintas variables

Aunque esta tabla proporciona la información necesaria, el diseño puede dificultar la interpretación de la tabla. Cualquiera que lea la tabla puede pensar que las casillas en blanco de la columna *Perdidos* indican cero valores perdidos para esas variables.

- Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- En el grupo Estadísticos de resumen del generador de tablas, seleccione **Filas** en la lista desplegable Posición.
- Pulse en **Aceptar** para crear la tabla.

Age of respondent	Mean	45.6
	Median	42.0
Highest year of school completed	Mean	13.3
	Median	13.0
Hours per day watching TV	Mean	2.9
	Median	2.0
	Missing	495

Figura 43. Estadísticos de resumen y variables mostrados en la dimensión de fila

Ahora está claro que la tabla muestra el número de valores perdidos solamente para una variable.

Resúmenes por grupos en Categorías

Puede utilizar variables categóricas como variables de agrupación para visualizar resúmenes de variables de escala dentro de los grupos definidos por las categorías de la variable categórica.

- Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- Arrastre y coloque *Sexo*, desde la lista de variables, y colóquela en el área Columnas.
Pulse con el botón derecho del ratón en *Sexo* en la vista previa de la tabla del panel de lienzo; verá que **Estadísticos de resumen** aparece desactivado en el menú emergente. Esto se debe a que en una tabla con variables de escala, éstas son siempre las variables origen de los estadísticos.
- Pulse en **Aceptar** para crear la tabla.

		Gender	
		Male	Female
Age of respondent	Mean	44.6	46.3
	Median	42.0	43.0
Highest year of school completed	Mean	13.4	13.2
	Median	13.0	13.0
Hours per day watching TV	Mean	2.8	2.9
	Median	2.0	2.0
	Missing	213	282

Figura 44. Resúmenes de escala agrupados que utilizan una variable de columna categórica

Esta tabla facilita la comparación de los promedios (media y mediana) para hombres y mujeres, y muestra claramente que no existe mucha diferencia entre ellos; puede que no sea demasiado interesante, pero sí puede ser una información útil.

VARIABLES DE AGRUPACIÓN MÚLTIPLES

Puede subdividir más los grupos, anidando y/o utilizando variables categóricas para la agrupación tanto de las filas como de las columnas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre de la lista de variables *Obtener noticias de Internet* y colóquela en la parte más a la izquierda del área Filas del panel de lienzo. Asegúrese de colocarla de manera que las tres variables de escala, no sólo una de ellas, queden anidadas bajo ella.

Aunque es posible que a veces desee obtener algo parecido al segundo ejemplo anterior, no es lo que se busca en este caso.

3. Pulse en **Aceptar** para crear la tabla.

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	Mean	47.0	48.8
			Median	45.0	46.0
		Highest year of school completed	Mean	13.4	13.1
			Median	13.0	12.0
		Hours per day watching TV	Mean	3.2	3.4
			Median	2.0	3.0
	Yes	Age of respondent	Missing	213	282
			Mean	38.7	41.1
			Median	35.0	38.0
		Highest year of school completed	Mean	13.2	13.3
			Median	13.0	13.0
		Hours per day watching TV	Mean	2.1	2.1
			Median	2.0	2.0
				Missing	0

Figura 45. Resúmenes de escala agrupados por variables categóricas de fila y columna

ANIDACIÓN DE VARIABLES CATEGÓRICAS DENTRO DE LAS VARIABLES DE ESCALA

Aunque la tabla anterior puede proporcionar la información que desea, es posible que no la proporcione en el formato más fácil de interpretar. Por ejemplo, puede comparar la edad media de los hombres que utilizan Internet para obtener noticias y de los que no, pero sería más fácil si los valores estuvieran juntos en lugar de separados. El intercambio de posiciones de las dos variables de fila y la anidación de la variable de agrupación categórica dentro de las tres variables de escala podría mejorar la tabla. Con las variables de escala, el nivel de anidación no afecta a la variable origen de los estadísticos. La variable de escala es siempre la variable origen de los estadísticos, independientemente del nivel de anidación.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en *Edad del encuestado* en la vista previa de la tabla del panel de lienzo, pulse Ctrl y el botón del ratón para seleccionar *Años de escolarización* y en *Horas diarias viendo TV* y elija las tres variables de escala.
3. Arrastre y coloque las tres variables de escala en la parte más a la izquierda del área Filas, anidando la variable categórica *Obtener noticias de Internet* dentro de cada una de las tres variables de escala.
4. Pulse en **Aceptar** para crear la tabla.

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	Mean	47.0	48.8
			Median	45.0	46.0
		Yes	Mean	38.7	41.1
			Median	35.0	38.0
Highest year of school completed	Get news from internet	No	Mean	13.4	13.1
			Median	13.0	12.0
		Yes	Mean	13.2	13.3
			Median	13.0	13.0
Hours per day watching TV	Get news from internet	No	Mean	3.2	3.4
			Median	2.0	3.0
			Missing	213	282
		Yes	Mean	2.1	2.1
			Median	2.0	2.0
			Missing	0	0

Figura 46. Variable de fila categórica anidada dentro de las variables de escala apiladas

La elección del orden de anidación depende de las relaciones o comparaciones que desee destacar en la tabla. Al cambiar el orden de anidación de las variables de escala no cambian los valores de los estadísticos de resumen; sólo cambian sus posiciones relativas en la tabla.

Capítulo 9. Estadísticos de contraste

Estadísticos de contraste

Hay tres comprobaciones diferentes de significación disponibles para estudiar la relación entre variables de fila y columna. Este capítulo analiza los resultados de cada comprobación, con atención especial a los efectos de anidación y apilado. Consulte el tema Capítulo 3, “Apilado, anidación y capas con variables categóricas”, en la página 27 para obtener más información.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Comprobaciones de independencia (chi-cuadrado)

La comprobación de independencia chi-cuadrado se utiliza para determinar si hay una relación entre dos variables categóricas. Por ejemplo, es posible que desee determinar si *Cargo profesional* está relacionado con *Estado civil*.

1. Seleccione en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre y suelte *Cargo profesional* de la lista de variables en el área Filas del panel de lienzo.
3. Arrastre y suelte *Estado civil* de la lista de variables en el área Columnas.
4. Seleccione **Filas** como la posición de los estadísticos de resumen.
5. Seleccione *Cargo profesional* y pulse en **Estadísticos de resumen** en el grupo Definir.
6. Seleccione **% de Columna N** en la lista Estadísticos y añádalo a la lista Representación.
7. Pulse en **Aplicar a la selección**.
8. En el cuadro de diálogo Tablas personalizadas, pulse en la pestaña **Estadísticos de contraste**.
9. Seleccione **Comprobaciones de independencia (chi-cuadrado)**.
10. Pulse en **Aceptar** para crear la tabla y obtener la comprobación chi-cuadrado.

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	Count	778	44	295	58	392
		Column %	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	Count	138	20	35	9	102
		Column %	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	Count	23	2	9	1	11
		Column %	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	Count	13	3	10	0	32
		Column %	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	Count	168	150	53	6	17
		Column %	12.5%	53.0%	11.9%	6.5%	2.6%
	School	Count	9	1	7	2	60
		Column %	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	Count	200	55	25	13	35
		Column %	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	Count	16	8	12	4	14
		Column %	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 47. Cargo profesional por estado civil

Esta tabla es una tabulación cruzada de *Cargo profesional* y *Estado civil*, con recuentos y proporciones de columnas mostrados como estadísticos de resumen. Las proporciones de columna se calculan de manera que sumen 100 % debajo de cada columna. Si estas dos variables no están relacionadas, en cada fila las proporciones serán similares en todas las columnas. Parece haber diferencias en las proporciones, pero puede realizar la comprobación chi-cuadrado para asegurarse.

		Marital status
Labor force status	Chi-square	729.242
	df	28
	Sig.	.000*

*. The Chi-square statistic is significant at the 0.05 level.

Figura 48. Comprobación chi-cuadrado de Pearson

La comprobación de independencia hace hipótesis de que *Cargo profesional* y *Estado civil* no están relacionados, es decir, que las proporciones de columnas son las mismas en todas las columnas y que cualquier discrepancia observada se debe a la variación atribuible al azar. El estadístico de chi-cuadrado mide la discrepancia general entre los recuentos de casillas observados y los recuentos que esperaría si las proporciones de columna fuesen las mismas entre columnas. Un estadístico de chi-cuadrado mayor indica una discrepancia mayor entre los recuentos de casillas observados y esperados; lo que prueba aún más que las proporciones de columna no son iguales, que la hipótesis de independencia no es correcta y, por tanto, que *Cargo profesional* y *Estado civil* están relacionados.

El estadístico de chi-cuadrado calculado tiene un valor de 729.242. Para determinar si sirve como prueba suficiente para reflejar la hipótesis de independencia, se calcula el valor de significación del estadístico. El valor de significación es la probabilidad de que una variable aleatoria dibujada a partir de una distribución chi-cuadrado con 28 grados de libertad sea mayor que 729.242. Puesto que este valor es inferior al nivel alfa especificado en la pestaña Estadísticos de contraste, puede rechazar la hipótesis de independencia en el nivel 0,05. Así, *Cargo profesional* y *Estado civil* quedan relacionados.

Efectos de anidación y apilado en comprobaciones de independencia

La regla para las comprobaciones de independencia es la siguiente: se realiza una comprobación distinta para cada subtabla más interior. Para ver cómo afecta la anidación a las comprobaciones, considere el ejemplo anterior, pero con *Estado civil* anidado con los niveles de *Género*.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).

2. Arrastre y suelte *Género* de la lista de variables en el área Columnas del panel de lienzo sobre *Estado civil*.
3. Pulse en **Aceptar** para crear la tabla.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*,1,2}	.000 ^{*,1,2}

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figura 49. Comprobación chi-cuadrado de Pearson

Con *Estado civil* anidado dentro de los niveles de *Género*, se han realizado dos comprobaciones, una por cada nivel de *Género*. El valor de significación para cada comprobación indica que puede rechazar la hipótesis de independencia entre *Estado civil* y *Cargo profesional* para hombres y mujeres. No obstante, la tabla indica que más del 20 % de las casillas de cada tabla han esperado recuentos inferiores a 5 y el recuento de casilla mínimo esperado es inferior a 1. Estas notas indican que los supuestos de la comprobación de chi-cuadrado pueden no coincidir con estas tablas, de manera que los resultados de las comprobaciones son sospechosos.

Nota: las notas al pie pueden cortarse en la vista por los límites de las casillas. Puede visualizarlas cambiando la alineación de dichas casillas en el cuadro de diálogo Propiedades de casilla.

Para ver cómo afecta el apilado a las comprobaciones:

4. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
5. Arrastre y suelte *Grado más alto* de la lista de variables en el área Filas debajo de *Cargo profesional*.
6. Pulse en **Aceptar** para crear la tabla.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*,1,2}	.000 ^{*,1,2}
Highest degree	Chi-square	43.844	105.506
	df	16	16
	Sig.	.000 [*]	.000 [*]

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figura 50. Comprobación chi-cuadrado de Pearson

Con *Grado más alto* apilado con *Cargo profesional*, se realizan cuatro comprobaciones: una comprobación de independencia de *Estado civil* y *Cargo profesional* y una comprobación de *Estado civil* y *Grado más alto* para cada nivel de *Género*. Los resultados de las comprobaciones de *Estado civil* y *Cargo profesional* son los mismos que antes. Los resultados de las comprobaciones de *Estado civil* y *Grado más alto* indican que estas variables no son independientes.

Comparación de medias de columna

Las comprobaciones de medias de columna se utilizan para determinar si existe una relación entre una variable categórica en Columnas y una variable continua en Filas. Además, puede utilizar los resultados de las comprobaciones para determinar el orden relativo de categorías de la variable categórica en términos del valor medio de la variable continua. Por ejemplo, es posible que desee determinar si *Horas al día de televisión* está relacionado con *Obtener noticias de periódicos*.

1. Seleccione en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Pulse en **Restablecer** para restaurar la configuración predeterminada de todas las tablas.
3. En el generador de tablas, arrastre y suelte *Horas al día de televisión* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y suelte *Obtener noticias de periódicos* de la lista de variables en el área Columnas.
5. Seleccione *Horas al día de televisión* y pulse en **Estadísticos de resumen** en el grupo Definir.
6. Seleccione **nnnn** como formato.
7. Seleccione **2** como el número de decimales que desea visualizar. Tenga en cuenta que ahora el formato será **nnnn.nn**.
8. Pulse en **Aplicar a la selección**.
9. En el cuadro de diálogo Tablas personalizadas, pulse en la pestaña **Estadísticos de contraste**.
10. Seleccione **Comparar medias de columna (pruebas t)**.
11. Pulse en **Aceptar** para crear la tabla y obtener las comprobaciones de medias de columna.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

Figura 51. *Obtener noticias de periódicos y horas al día de televisión*

Esta tabla muestra la media de *Horas al día de televisión* para las personas que obtienen las noticias de los periódicos y las que no. La diferencia observada en estas medias sugiere que las personas que no obtienen las noticias de los periódicos pasan aproximadamente 0,18 horas más viendo la televisión que las personas que obtienen las noticias de los periódicos. Para ver si esta diferencia se debe a una variación atribuible al azar, consulte las comprobaciones de medias de columna.

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

Figura 52. *Comparaciones de medias de columna*

La tabla de comprobaciones de medias de columnas asigna una clave de letra a cada categoría de la variable de columna. Para *Obtener noticias de periódicos*, la categoría *No* se asigna a la letra A y *Sí* se asigna a la letra B. Para cada par de columnas, las medias de columna se comparan utilizando una prueba *t*. Dado que sólo hay dos columnas, sólo se realiza una comprobación. Para cada par significativo, la clave de la categoría con la media más baja se coloca debajo de la categoría con la media más alta. Dado que las claves se revelan en las casillas de la tabla, quiere decir que las medias de columna no son estadísticamente diferentes.

Resultados de significación en notación de estilo APA

Si no quiere que los resultados de significación en una tabla diferente, puede seleccionar visualizarlos en la tabla principal. Los resultados de significación se identifican utilizando una notación de estilo APA con subíndice. Complete los pasos anteriores para comparar medias de columnas, pero realice los siguientes cambios en la pestaña Estadísticos de contraste:

1. En el área Identificar diferencias significativas, seleccione **En la tabla principal utilizando subíndices de estilo APA**.
2. Pulse en **Aceptar** para crear la tabla y obtener las pruebas sobre las medias de las columnas utilizando notación de estilo APA.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92 _a	2.74 _a

Figura 53. Comparaciones de medias de columnas utilizando notación de estilo APA

La tabla de las pruebas de las medias de las columnas asigna un subíndice a las categorías de la variable de las columnas. Las medias de las columnas se comparan utilizando una prueba t , para cada pareja de columnas. Si un par de valores es significativamente diferente, los valores tienen *diferentes* subíndices asignados. Dado que sólo hay dos columnas, sólo se realiza una comprobación. Como las medias de columnas de este ejemplo comparten el mismo subíndice, las medias de columnas no son estadísticamente diferentes.

Efectos de anidación y apilado en comprobaciones de medias de columna

La regla para las comprobaciones de media de columna es la siguiente: se realiza un conjunto distinto de comprobación por parejas para cada subtabla más interior. Para ver cómo afecta la anidación a las comprobaciones, considere el ejemplo anterior, pero con *Horas al día de televisión* anidado con los niveles de *Cargo profesional*.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre y suelte *Cargo profesional* de la lista de variables en el área Filas del panel de lienzo.
3. Pulse en **Aceptar** para crear la tabla.

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

Figura 54. Comparaciones de medias de columna

Con *Horas al día de televisión* anidado dentro de los niveles de *Cargo profesional*, se realizan siete conjuntos de comprobaciones de medias de columna: una para cada nivel de *Cargo profesional*. Las mismas claves de letras se asignan a las categorías de *Obtener noticias de periódicos*. Para los encuestados de *trabajo de jornada completa*, la clave B aparece en la columna A. Lo que significa que para los empleados de jornada completa, el valor medio de *Horas al día de televisión* es inferior al de personas que obtienen las noticias de los periódicos. No aparecen otras claves en las columnas, por lo que puede concluir que no hay otras diferencias estadísticamente significativas en las medias de columna.

Correcciones de Bonferroni Cuando se realizan varias comprobaciones, la corrección de Bonferroni se aplica a las comprobaciones de medias de columna para asegurar que el nivel alfa (o tasa de falsos positivos) especificado en la pestaña Estadísticos de contraste se aplica a cada *conjunto* de comprobaciones. Así, en esta tabla, no se aplicaron correcciones de Bonferroni porque aunque se realizaron siete conjuntos de comprobaciones, dentro de cada conjunto sólo se comparan un par de columnas.

Para ver cómo afecta el apilado a las comprobaciones:

4. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
5. Arrastre y suelte *Obtener noticias por Internet* de la lista de variables en el área Columnas a la izquierda de *Obtener noticias de periódicos*.
6. Pulse en **Aceptar** para crear la tabla.

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

Figura 55. Comparaciones de medias de columna

Con *Obtener noticias por Internet* apilado con *Obtener noticias de periódicos*, se realizan 14 conjuntos de comprobaciones de medias de columna, uno para cada nivel de *Cargo profesional* para *Obtener noticias por Internet* y *Obtener noticias de periódicos*. De nuevo, no se aplican correcciones de Bonferroni dado que dentro de un conjunto, sólo se comparan un par de columnas. Las comprobaciones de *Obtener noticias de periódicos* son las mismas que antes. Para *Obtener noticias por Internet*, la categoría *No* se asigna a la letra A y *Sí* se asigna a la letra B. La clave B se incluye en la columna A para cada conjunto de comprobaciones de medias de columna excepto para los encuestados que no trabajan temporalmente. Lo que significa que el valor medio de *Horas al día de televisión* es inferior para las personas que obtienen las noticias por Internet que para las personas que no obtienen las noticias de periódicos. No se especifican claves para el conjunto *sin trabajo temporalmente*, así que las medias de columna no son estadísticamente diferentes para estos encuestados.

Comparación de proporciones de columna

Las comprobaciones de proporciones de columna se utilizan para determinar el orden relativo de categorías de la variable categórica Columnas en términos de las proporciones de categoría de la variable categórica Filas. Por ejemplo, después de utilizar una comprobación de chi-cuadrado para descubrir que *cargo profesional* y *Estado civil* no son independientes, es posible que desee ver qué filas y columnas son las responsables de esta relación.

1. Seleccione en los menús:
Analizar > Tablas > Tablas personalizadas...
2. Pulse en **Restablecer** para restaurar la configuración predeterminada de todas las tablas.
3. En el generador de tablas, arrastre y suelte *Cargo profesional* de la lista de variables en el área Filas del panel de lienzo.
4. Arrastre y suelte *Estado civil* de la lista de variables en el área Columnas.
5. Seleccione *Cargo profesional* y pulse en **Estadísticos de resumen** en el grupo Definir.
6. Seleccione **% de Columna N** en la lista Estadísticos y añádalo a la lista Representación.
7. Deseleccionar **Recuento** de la lista Representación.
8. Pulse en **Aplicar a la selección**.
9. En el cuadro de diálogo Tablas personalizadas, pulse en la pestaña **Estadísticos de contraste**.
10. Seleccione **Comparar proporciones de columnas (pruebas z)**.

11. Pulse en **Aceptar** para crear la tabla y obtener las comprobaciones de proporciones de columna.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figura 56. Cargo profesional por estado civil

Esta tabla es una tabulación cruzada de *Cargo profesional* y *Estado civil*, con proporciones de columnas mostradas como estadísticos de resumen.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		A B	B	B
	Working part-time					A B C
	Temporarily not working					A B
	Unemployed, laid off					A B
	Retired	E	A C D E	E		
	School					A B C
	Keeping house	C E	C E		C E	
	Other					

Figura 57. Comparaciones de proporciones de columna

La tabla de comprobaciones de proporciones de columnas asigna una clave de letra a cada categoría de las variables de columna. Para *Estado civil*, la categoría *Casado* se asigna a la letra A, *Viudo* se asigna a la letra B y así sucesivamente hasta la categoría *Soltero* que se asigna a la letra E. Para cada par de columnas, las proporciones de columna se comparan utilizando una prueba z. Se realizan siete conjuntos de comprobaciones de proporciones de columna, uno para cada nivel de *Cargo profesional*. Dado que hay cinco niveles de *Estado civil*, se comparan $(5 \cdot 4) / 2 = 10$ pares de columnas en cada conjunto de comprobaciones y las correcciones de Bonferroni se utilizan para ajustar los valores de significación. Para cada par significativo, la clave de la categoría menor se coloca debajo de la categoría con la proporción más alta.

Para el conjunto de comprobaciones asociadas a *trabajo de jornada completa*, la clave B aparece en cada una de las demás columnas. Asimismo, la clave A aparece en la columna C. No se incluyen otras claves en las demás columnas. De manera que puede concluir que la proporción de divorciados que trabajan jornada completa es mayor que la proporción de casados que trabajan jornada completa, que a su vez es mayor que la proporción de viudos que trabajan jornada completa. Las proporciones de personas separadas o solteras y que trabajan jornada completa no se pueden diferenciar de los divorciados o casados que trabajan jornada completa, pero estas proporciones son mayores que la proporción de viudos que trabajan jornada completa.

Para las comprobaciones asociadas a *Trabajo de media jornada* o *Estudio*, las claves A, B y C aparecen en la columna E. No se incluyen otras claves en las demás columnas. Así, las proporciones de solteros que están estudiando o trabajan media jornada son mayores que las proporciones de casados, viudos o divorciados que están estudiando o trabajan media jornada.

Para las comprobaciones asociadas a *Sin trabajo temporalmente* u *Otro*, no se incluyen claves en ninguna columna. De manera que no existe una diferencia apreciable en las proporciones de casados, viudos, divorciados, separados o solteros que no trabajan temporalmente o se encuentran en otra situación laboral con otra clasificación.

Las comprobaciones asociadas a *Jubilado* muestran que la proporción de viudos jubilados es mayor que las proporciones de otros estados civiles jubilados. Además, las proporciones de casados o divorciados y jubilados son mayores que la proporción de solteros jubilados.

La proporción de casados, viudos o separados y amos de casa es menor que la proporción de divorciados o casados y amos de casa.

La proporción de solteros y *Desempleados o despedidos* es superior a las proporciones de casados o viudos y desempleados. Asimismo, tenga en cuenta que la columna *Separado* está marcada con un ".", que indica que la proporción observada de separados en la fila *Desempleados, despedidos* es 0 ó 1 y, por consiguiente, no se pueden realizar comparaciones utilizando esa columna para los encuestados desempleados.

Resultados de significación en notación de estilo APA

Si no quiere que los resultados de significación en una tabla diferente, puede seleccionar visualizarlos en la tabla principal. Los resultados de significación se identifican utilizando una notación de estilo APA con subíndice. Complete los pasos anteriores para comparar proporciones de columnas, pero realice los siguientes cambios en la pestaña Estadísticos de contraste:

1. En el área Identificar diferencias significativas, seleccione **En la tabla principal utilizando subíndices de estilo APA**.
2. Pulse en **Aceptar** para crear la tabla y obtener las pruebas sobre las medias de las columnas utilizando notación de estilo APA.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% _a	15.5% _b	66.1% _c	62.4% _{a,c}	59.1% _{a,c}
	Working part-time	10.3% _a	7.1% _a	7.8% _a	9.7% _{a,b}	15.4% _b
	Temporarily not working	1.7% _a	.7% _a	2.0% _a	1.1% _a	1.7% _a
	Unemployed, laid off	1.0% _a	1.1% _a	2.2% _{a,b}	.0%	4.8% _b
	Retired	12.5% _a	53.0% _b	11.9% _a	6.5% _{a,c}	2.6% _c
	School	.7% _a	.4% _a	1.6% _a	2.2% _{a,b}	9.0% _b
	Keeping house	14.9% _a	19.4% _a	5.6% _b	14.0% _a	5.3% _b
	Other	1.2% _a	2.8% _a	2.7% _a	4.3% _a	2.1% _a

Figura 58. Comparaciones de proporciones de columnas utilizando notación de estilo APA

La tabla de las pruebas de las proporciones de las columnas asigna un subíndice a las categorías de la variable de las columnas. Las proporciones de las columnas se comparan utilizando una prueba z, para cada pareja de columnas. Si un par de valores es significativamente diferente, los valores tienen *diferentes* subíndices asignados.

Para el conjunto de pruebas asociado a *A tiempo completo*, la categoría Viudo tiene un subíndice que no se utiliza en las otras columnas, y las categorías Separado y Nunca casado comparten los mismos dos subíndices. Considerando estos subíndices y las proporciones reales de la tabla, puede realizar el mismo análisis que se demuestra en el ejemplo anterior con tablas diferentes. De manera que puede concluir que

la proporción de divorciados que trabajan jornada completa es mayor que la proporción de casados que trabajan jornada completa, que a su vez es mayor que la proporción de viudos que trabajan jornada completa. Las proporciones de personas separadas o solteras y que trabajan jornada completa no se pueden diferenciar de los divorciados o casados que trabajan jornada completa, pero estas proporciones son mayores que la proporción de viudos que trabajan jornada completa. Se aplica el resto de análisis del ejemplo anterior.

Efectos de anidación y apilado en comprobaciones de proporciones de columna

La regla para las comprobaciones de proporciones de columna es la siguiente: se realiza un conjunto distinto de comprobación por parejas para cada subtabla más interior. Para ver cómo afecta la anidación a las comprobaciones, considere el ejemplo anterior, pero con *Cargo profesional* anidado con los niveles de *Género*.

1. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre y suelte *Género* de la lista de variables en el área Filas del panel de lienzo.
3. Pulse en **Aceptar** para crear la tabla.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								

Figura 59. Comparaciones de proporciones de columna

Con *Cargo profesional* anidado dentro de los niveles de *Género*, se han realizado 14 conjuntos de comprobaciones de proporciones de columna, uno por cada nivel de *Cargo profesional* para cada nivel de *Género*. Las mismas claves de letras se asignan a las categorías de *Estado civil*.

Debemos señalar un par de puntos en los resultados de la tabla:

- Con más comprobaciones, hay más columnas con proporción de columna cero. Son más comunes en encuestados separados y hombres viudos.
- Parece que las diferencias de columna apreciadas anteriormente entre los encuestados *amos de casa* se deben completamente a las mujeres.

Para ver cómo afecta el apilado a las comprobaciones:

4. Vuelva a abrir el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).

5. Arrastre y suelte *Grado más alto* de la lista de variables en el área Filas debajo de *Género*.
6. Pulse en **Aceptar** para crear la tabla.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
			Keeping house	C E	C E		C	
Highest degree	LT High school			A C E				
	High school							
	Junior college			B		B		
	Bachelor			B		B		
	Graduate			B				

Figura 60. Comparaciones de proporciones de columna

Con *Grado más alto* apilado con *Género*, se realizan 19 conjuntos de comprobaciones de medias de columna, los 14 tratados anteriormente y uno para cada nivel de *Grado más alto*. Las mismas claves de letras se asignan a las categorías de *Estado civil*.

Debemos señalar los siguientes puntos en los resultados de la tabla:

- Los resultados de comprobaciones de los 14 conjuntos de comprobaciones realizados anteriormente son los mismos.
- Las personas que tienen un nivel inferior a la enseñanza secundaria son más comunes en los viudos que en los casados, divorciados o solteros.
- Las personas con un nivel educativo superior a la enseñanza secundaria son más comunes en los casados, divorciados y solteros que en los viudos.

Nota sobre ponderaciones y conjuntos de respuestas múltiples

Las ponderaciones de casos se basan siempre en recuentos, no en respuestas, incluso cuando una de las variables es de respuestas múltiples.

Capítulo 10. Conjuntos de respuestas múltiples

Las Tablas personalizadas y el Generador de gráficos admiten un tipo especial de "variable" al que se denomina **conjunto de respuestas múltiples**. En realidad, los conjuntos de respuestas múltiples no son, en sentido estricto, "variables". No aparecen en el Editor de datos y los demás procedimientos no los reconocen. Los conjuntos de respuestas múltiples utilizan variables para registrar respuestas a preguntas donde el encuestado puede ofrecer más de una respuesta. Los conjuntos de respuestas múltiples se tratan como variables categóricas y la mayoría de las acciones que puede realizar con las variables categóricas, también las puede realizar con los conjuntos de respuestas múltiples.

Los conjuntos de respuestas múltiples se crean a partir de múltiples variables del archivo de datos. Un conjunto de respuestas múltiples es un constructo especial perteneciente a un archivo de datos. Se pueden definir y guardar varios conjuntos en un archivo de datos IBM® SPSS Statistics, pero no se pueden importar o exportar conjuntos de respuestas múltiples desde o a otros formatos de archivo. (Se pueden copiar conjuntos de respuestas múltiples de otros archivos de datos IBM SPSS Statistics mediante Copiar propiedades de datos en el menú Datos en la ventana Editor de datos.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de presentación de la lista de variables se especifican en la pestaña General del cuadro de diálogo Opciones (menú Edición, Opciones).

Recuentos, respuestas, porcentajes y totales

Todos los estadísticos de resumen disponibles para las variables categóricas también están disponibles para los conjuntos de respuestas múltiples. Asimismo, hay disponibles estadísticos de resumen adicionales para los conjuntos de respuestas múltiples.

1. Seleccione en los menús:

Analizar > Tablas > Tablas personalizadas...

2. Arrastre y coloque *Fuentes de noticias* (esta es la etiqueta descriptiva para el conjunto de respuestas múltiples \$*mltnews*) desde la lista de variables hasta el área Filas del panel de lienzo.

El icono junto a la "variable" en la lista de variables lo identifica como un conjunto de dicotomías múltiples.



Figura 61. Icono de conjunto de dicotomías múltiples

Para un conjunto de dicotomías múltiples, cada "categoría" es, de hecho, una variable independiente, y las etiquetas de las categorías son las etiquetas de las variables (o los nombres de variable para las variables que no tengan definida la etiqueta de variable). En este ejemplo, los recuentos que se mostrarán representan el número de casos con la respuesta *Sí* para cada variable del conjunto.

3. Pulse con el botón derecho del ratón en *Fuentes de noticias* en la vista previa de la tabla en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.

4. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.

5. Vuelva a pulsar con el botón derecho del ratón en *Fuentes de noticias* y seleccione **Estadísticos de resumen** en el menú emergente.
6. En el cuadro de diálogo Estadísticos de resumen, seleccione **% del N de columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
7. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** para crear la tabla.

		Count	Column N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
	Total	2081	100.0%

Figura 62. Porcentajes de columna y recuentos de dicotomías múltiples

Totales que no suman correspondientemente

Si observa los números en la tabla, notará que hay una enorme discrepancia entre los "totales" y los valores que se supone que están totalizando; es decir, los totales parecen ser muy inferiores a lo que deberían ser. Esto se debe a que el recuento para cada "categoría" en la tabla es el número de casos con un valor de 1 (una respuesta *Sí*) para dicha variable y el número total de respuestas *Sí* para las cinco variables del conjunto de dicotomías múltiples podría exceder fácilmente el número total de casos en el archivo de datos.

No obstante, el "recuento" total es el número total de casos con una respuesta *Sí* para al menos una variable del conjunto, que nunca puede superar el número total de casos del archivo de datos. En este ejemplo, el recuento total de 2.081 es inferior en aproximadamente 800 casos al número total de casos en el archivo de datos. Si ninguna de dichas variables tiene valores perdidos, significa que al menos 800 encuestados respondieron que no obtienen las noticias de ninguna de esas fuentes. El recuento total es la base para los porcentajes de columna; por consiguiente, los porcentajes de columna de este ejemplo suman más del 100% mostrado para el porcentaje de columna total.

Totales que suman correspondientemente

Aunque "recuento" es un término habitualmente poco ambiguo, el ejemplo anterior demuestra lo confuso que puede resultar en el contexto de los totales de los conjuntos de respuestas múltiples, en los que, a menudo, *respuestas* es el estadístico de resumen que se desea realmente.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en *Fuentes de noticias* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
3. En el cuadro de diálogo Estadísticos de resumen, seleccione **Respuestas** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
4. Seleccione **% de respuestas de la columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
5. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** para crear la tabla.

		Count	Column N %	Responses	Column Responses %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
	Total	2081	100.0%	3594	100.0%

Figura 63. Respuestas de dicotomías múltiples y porcentajes de respuestas de las columnas

Para cada "categoría" del conjunto de dicotomías múltiples, *Respuestas* es idéntico a *Recuento*, y este será siempre el caso con los conjuntos de dicotomías múltiples. Los totales, sin embargo, son muy diferentes. El número total de respuestas es de 3.594, unas 1.500 más que el recuento total y unas 700 más que el número total de casos existentes en el archivo de datos.

En los porcentajes, los totales de % del N de columna y % de respuestas de la columna son ambos 100%; sin embargo, los porcentajes de cada categoría del conjunto de dicotomías múltiples son muy inferiores a los porcentajes de respuesta de columna. Esto es debido a que la base porcentual utilizada en los porcentajes de respuesta de columna es el número total de respuestas, que en este caso es 3.594, lo que da como resultado porcentajes muy inferiores a los del porcentaje de columna, cuya base de cálculo es 2.081.

Porcentajes totales superiores a 100%

Tanto los porcentajes de columna como los porcentajes de respuesta de la columna generan totales porcentuales de 100% aun cuando, en nuestro ejemplo, la suma de los valores individuales de la columna % del N de columna supera claramente el 100%. Por lo tanto, ¿qué sucede si se desea mostrar porcentajes basados en el recuento total, en lugar del total de respuestas, pero también se desea que el porcentaje "total" refleje con exactitud la suma de los porcentajes de las categorías individuales?

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en *Fuentes de noticias* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
3. En el cuadro de diálogo Estadísticos de resumen, seleccione **% del respuestas de columna (Base: recuento)** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
4. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** para crear la tabla.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
	Total	2081	100.0%	3594	100.0%	172.7%

Figura 64. Porcentajes de respuestas de columna con el recuento como base porcentual

Utilización de conjuntos de respuestas múltiples con otras variables

En general, puede utilizar los conjuntos de respuestas múltiples, de manera similar a las variables categóricas. Por ejemplo, puede cruzar en una tabulación cruzadas un conjunto de respuestas múltiples respecto a una variable categórica o anidar un conjunto de respuestas múltiples dentro de una variable categórica.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).

2. Arrastre la variable *Sexo* desde la lista de variables y colóquela en la parte izquierda del panel de vista previa, anidando el conjunto de respuestas múltiples *Fuentes de noticias* dentro de las categorías de sexo.
3. Pulse con el botón derecho del ratón sobre *Género* en la vista previa de la tabla en el panel del lienzo y deselectione **Mostrar etiqueta de variable** en el menú emergente.
4. Proceda de la misma manera para *Fuentes de noticias*.

De esta manera, se eliminarán de la tabla las columnas correspondientes a las etiquetas de las variables (ya que en este caso no son verdaderamente necesarias).

5. Pulse en **Aceptar** para crear la tabla.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	Total	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	Total	1185	100.0%	2055	100.0%	173.4%

Figura 65. Conjunto de respuestas múltiples anidado dentro de una variable categórica

Variable origen de los estadísticos y estadísticos de resumen disponibles

En ausencia de una variable de escala en la tabla, las variables categóricas y los conjuntos de respuestas múltiples reciben el mismo trato, en cuanto a cuál es la variable origen de los estadísticos: la variable origen de los estadísticos es la variable anidada a mayor profundidad en la dimensión origen de los estadísticos. Debido a que algunos estadísticos de resumen sólo pueden ser asignados a los conjuntos de respuestas múltiples, si desea obtener alguno de los estadísticos especiales de respuestas múltiples, el conjunto de respuestas múltiples debe ser la variable anidada a mayor profundidad, en la dimensión origen de los estadísticos.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. En la vista previa de la tabla del panel de lienzo, arrastre *Fuentes de noticias* y colóquela a la izquierda de *Sexo*, cambiando el orden de anidación.

Todos los estadísticos especiales de respuestas múltiples (respuestas, porcentaje de respuestas de las columnas) han sido eliminados de la vista previa de la tabla, ya que la variable categórica *Sexo* es ahora la variable anidada más profunda y, por tanto, es la variable origen de los estadísticos.

Afortunadamente, el generador de tablas “recuerda” esta configuración. Si vuelve a colocar *Fuentes de noticias* en su posición anterior, anidada dentro de *Sexo*, se restaurarán en la vista previa de la tabla todos los estadísticos de resumen correspondientes a las respuestas.

Conjuntos de categorías múltiples y respuestas duplicadas

Los conjuntos de categorías múltiples proporcionan una característica que no está disponible en los conjuntos de dicotomías múltiples: la capacidad de contar las respuestas duplicadas. En los conjuntos de categorías múltiples, las respuestas duplicadas probablemente representan errores de codificación en la mayoría de los casos. Por ejemplo, en la pregunta de una encuesta “¿Cuáles son los tres países que fabrican los mejores coches?”, es muy probable que la respuesta *Suecia, Alemania y Suecia* no sea válida.

No obstante, en otros casos, es posible que las respuestas duplicadas sean totalmente válidas. Por ejemplo, si la pregunta fuera “¿Dónde se han fabricado sus tres últimos coches?”, la respuesta *Suecia, Alemania y Suecia* tendría pleno sentido.

Tablas personalizadas ofrece la opción de proporcionar varias respuestas duplicadas en conjuntos de categorías múltiples. De forma predeterminada, no se cuentan las respuestas duplicadas, pero puede solicitar que sean incluidas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar los ajustes anteriores.
3. Arrastre *Fabricante de coches, coches más recientes*, desde la lista de variables, y colóquela en el área Filas del panel de lienzo.

El icono que aparece junto a la “variable” en la lista de variables lo identifica como un conjunto de categorías múltiples.



Figura 66. Icono de conjunto de categorías múltiples

Para los conjuntos de categorías múltiples, las categorías mostradas representan el conjunto común de etiquetas de valor definidas para todas las variables del conjunto (mientras que para los conjuntos de dicotomías múltiples, las “categorías” son realmente las etiquetas de variable de cada variable del conjunto).

4. Pulse con el botón derecho del ratón en *Fabricante de coches, coches más recientes* en la vista previa de la tabla del panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
5. Seleccione (pulse) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.
6. Vuelva a pulsar con el botón derecho del ratón en *Fabricante de coches, coches más recientes* en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
7. En el cuadro de diálogo Estadísticos de resumen, seleccione **Respuestas** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
8. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** para crear la tabla.

		Count	Responses
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	Total	2832	5356

Figura 67. Conjunto de categorías múltiples: recuentos y respuestas sin duplicados

De forma predeterminada, no se cuentan las respuestas duplicadas; por tanto, en esta tabla, los valores para cada categoría en las columnas *Recuento* y *Respuestas* son idénticos. Sólo los totales son diferentes.

9. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
10. Pulse en la pestaña **Opciones**.
11. Pulse en (marque) **Contar las respuestas duplicadas en los conjuntos de categorías múltiples**.
12. Pulse en **Aceptar** para crear la tabla.

		Count	Responses
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	383
	Other	343	359
	Total	2832	6770

Figura 68. Conjunto de categorías múltiples con las respuestas duplicadas incluidas

En esta tabla, se puede apreciar la diferencia entre los valores de las columnas *Recuento* y *Respuestas*, especialmente en lo que respecta a los coches americanos, indicando que muchos encuestados han tenido varios coches americanos.

Pruebas de significación con conjuntos de respuestas múltiples

Puede utilizar conjuntos de respuestas múltiples en pruebas de significación prácticamente de la misma forma que utilizaría las variables categóricas.

- Para las pruebas de independencia (chi-cuadrado) o para comparar proporciones de columna (pruebas z), se realizan pruebas en los recuentos y uno de los estadísticos de resumen que aparecen en la tabla debe ser Recuento.
- Para los conjuntos de categorías múltiple, no se realizan pruebas de comparación de las proporciones de columna o las medias de columna (pruebas t) si **Contar las respuestas duplicadas en los conjuntos de categorías múltiples** está seleccionado en la pestaña Opciones. Consulte el tema “Tablas personalizadas: pestaña Opciones” en la página 15 para obtener más información.

Pruebas de independencia con Conjuntos de respuestas múltiples

Este ejemplo crea una tabulación cruzada de una variable categórica y un conjunto de respuestas múltiples y ejecuta una prueba de chi-cuadrado de independencia en la tabulación cruzada.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar los ajustes anteriores.
3. Arrastre y coloque *Fuentes de noticias* (esta es la etiqueta descriptiva para el conjunto de dicotomías múltiples \$mltnews) de la lista de variables en el área Columnas del panel de lienzo.
4. Arrastre y suelte *Género* de la lista de variables en el área Filas del panel de lienzo.
5. Pulse en la pestaña **Estadísticos de contraste**.
6. Seleccione (marque) **Pruebas de independencia (chi-cuadrado)**.
7. Si aún no está seleccionado, seleccione **Incluir variables de respuesta múltiple en las pruebas**.
8. Pulse en **Aceptar** para iniciar el procedimiento.

		News sources
Gender	Chi-square	10.266
	df	5
	Sig.	.068

Figura 69. Resultados de chi-cuadrado

El nivel de significación de 0,068 para la prueba de chi-cuadrado indica que no hay una diferencia significativa en las elecciones de fuentes de noticias entre hombres y mujeres (suponiendo que se utilice un valor de significación de 0,05 o menor como criterio para determinar la significación estadística).

Comparación de las medias de las columnas con conjuntos de respuestas múltiples

Este ejemplo calcula las medias de una variable de escala en categorías definidas por un conjunto de respuestas múltiples y compara cada media de categoría con todas las demás para encontrar diferencias significativas.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar los ajustes anteriores.
3. Arrastre y coloque *Fuentes de noticias* (esta es la etiqueta descriptiva para el conjunto de dicotomías múltiples \$mltnews) de la lista de variables en el área Columnas del panel de lienzo.
4. Arrastre y coloque *Edad del encuestado* en el área Filas del panel de lienzo.
5. Pulse en la pestaña **Estadísticos de contraste**.
6. Seleccione (marque) **Comparar las medias de columna (pruebas t)**.
7. Si aún no está seleccionado, seleccione **Incluir variables de respuesta múltiple en las pruebas**.
8. Pulse en **Aceptar** para iniciar el procedimiento.

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	Mean	Mean	Mean	Mean	Mean
Age of respondent	52	40	48	40	40

Comparisons of Column Means

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

Results are based on two-sided tests assuming equal variances with significance level 0.05. For each significant pair, the key of the smaller category appears under the category with larger mean.

Figura 70. Resultados de pruebas de significación

- Cada categoría del conjunto de respuestas múltiples se identifica con una letra (A, B, C, D, E) y para cada categoría en la que la media de otra es menor o difiere significativamente, se muestra la letra que representa la categoría con la media menor.
- *Obtener noticias de periódicos (A)* tiene la edad media mayor y todas las demás medias de categorías difieren de ella significativamente.
- *Obtener noticias de periódicos (A)* tiene la siguiente edad media mayor y todas las demás medias de categorías restantes difieren de ella significativamente. (C también difiere significativamente de A, como ya se indicó anteriormente.)
- Las edades medias para *Obtener noticias de periódicos (B)*, *Obtener noticias de radio (D)* y *Obtener noticias de Internet (E)* no difieren significativamente entre ellas.

Capítulo 11. Valores perdidos

Muchos archivos de datos contienen una cierta cantidad de datos perdidos. Una amplia variedad de factores puede dar como resultado la aparición de datos perdidos. Por ejemplo, es posible que los encuestados no respondan a todas las preguntas, puede que algunas variables no sean aplicables a ciertos casos y los errores de codificación pueden dar como resultado que algunos valores sean desechados.

En IBM SPSS Statistics existen dos clases de valores perdidos:

- **Perdido del usuario.** Valores que por definición contienen datos perdidos. Se pueden asignar etiquetas de valor a dichos valores, para identificar por qué son datos perdidos (como, por ejemplo, un código de 99 y una etiqueta de valor *No procede* para el embarazo en el caso de los hombres).
- **Perdido del sistema.** Si no hay ningún valor presente en una variable numérica, se le asignará el valor perdido del sistema. Esta situación se indica mediante un punto en la Vista de datos del Editor de datos.

Hay una serie de utilidades que ayudan a compensar los efectos de los datos perdidos e incluso analizar los patrones existentes en los datos perdidos. No obstante, este capítulo, tiene un objetivo mucho más simple: describir cómo gestionan los valores perdidos las Tablas personalizadas y cómo dichos valores afectan al cálculo de los estadísticos de resumen.

Archivo de datos de muestra

En los ejemplos de esta sección se utiliza el archivo de datos *missing_values.sav*. Consulte el tema archivos de datos para obtener más información. Se trata de un archivo muy simple que contiene datos completamente artificiales, con una única variable y diez casos, diseñado para ilustrar diversos conceptos básicos sobre los valores perdidos.

Tablas sin valores perdidos

De forma predeterminada, las categorías perdidas del usuario no se muestran en las tablas personalizadas (y los valores perdidos del sistema no se muestran nunca).

1. Seleccione en los menús:

Analizar > Tablas > Tablas personalizadas...

2. En el generador de tablas, arrastre *Variable con valores perdidos* (la única variable del archivo), desde la lista de variables, y colóquela en el área Filas del panel de lienzo.
3. Pulse con el botón derecho del ratón en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
4. Pulse en (marque) **Total** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.
5. Vuelva a pulsar con el botón derecho del ratón en *Variable con valores perdidos* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
6. En el cuadro de diálogo Estadísticos de resumen, seleccione **% del N de columna** en la lista Estadísticos y pulse en la flecha para añadirlo a la lista Visualización.
7. Pulse en **Aplicar a la selección**.

Observará que existe una pequeña discrepancia entre las categorías que se muestran en la vista previa de la tabla en el panel de lienzo y las categorías que aparecen en la lista Categorías (situada debajo de la lista de variables, en la parte izquierda del generador de tablas). La lista Categorías contiene una categoría con la etiqueta *Valores perdidos* que no está incluida en la vista previa de la tabla ya que, de

forma predeterminada, se excluyen las categorías correspondientes a valores perdidos. El hecho de que "valores" aparezca en plural en la etiqueta indica que la variable tiene dos o más categorías de valores perdidos del usuario.

8. Pulse en **Aceptar** para crear la tabla.

		Count	Column N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Figura 71. Tabla sin valores perdidos

Todo lo que aparece en esta tabla es totalmente correcto. Los valores de las categorías suman los totales y los porcentajes reflejan con precisión los valores obtenidos utilizando el recuento total como base porcentual (por ejemplo, $3/7 = 0,429$ o 42,9%). No obstante, el recuento total no es el número total de casos del archivo de datos; sino que es el número total de casos con valores **no perdidos** o los casos que no tienen valores perdidos del usuario ni del sistema para dicha variable.

Inclusión de valores perdidos en las tablas

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse con el botón derecho del ratón en *Variable con valores perdidos* en la vista previa de la tabla en el panel de lienzo y seleccione **Categorías y totales** en el menú emergente.
3. Pulse en (marque) **Valores perdidos** en el cuadro de diálogo Categorías y totales y, a continuación, pulse en **Aplicar**.

Ahora la vista previa de la tabla incluye una categoría *Valores perdidos*. Aunque la vista previa de la tabla sólo muestra una categoría para los valores perdidos, en la tabla se mostrarán todas las categorías con valores perdidos del usuario.

4. Vuelva a pulsar con el botón derecho del ratón en *Variable con valores perdidos* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
5. En el cuadro de diálogo Estadísticos de resumen, pulse en (marque) **Estadísticos de resumen personalizados para totales y subtotales**.
6. Seleccione **N válido** en la lista Estadísticos de resumen personalizados y pulse el botón de flecha para añadirlo a la lista Visualización.
7. Proceda de la misma manera para **N total**.
8. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** en el generador para crear la tabla.

		Count	Column N %	Valid N	Total N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	Total	9	100.0%	7	10

Figura 72. Tabla con valores perdidos

Las dos categorías de valores perdidos del usuario ya definidas, *No sabe* y *No procede*, se muestran ahora en la tabla y el recuento total es ahora 9, en lugar de 7, hecho que refleja la adición de los dos casos con valores perdidos del usuario (uno en cada categoría de valores perdidos del usuario). Los porcentajes de columna también son diferentes ahora, ya que se basan en el número de valores no perdidos y en los valores perdidos del usuario. Quedan excluidos del cálculo de los porcentajes únicamente los valores perdidos del sistema.

N válido muestra el número total de casos no perdidos (7) y *N total* muestra el número total de casos, incluyendo tanto los valores perdidos del usuario como los perdidos del sistema. El número total de casos es 10, uno más que el recuento de valores no perdidos y perdidos del usuario que aparece como el total en la columna *Recuento*. Esto se debe a que hay un caso con un valor perdido del sistema.

9. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
10. Pulse con el botón derecho del ratón en *Variable con valores perdidos* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
11. Seleccione **% del N válido de columna** en la lista Estadísticos de la parte superior (no los resúmenes personalizados para los totales y subtotales) y pulse el botón de flecha para añadirlo a la lista Visualización.
12. Proceda de la misma manera para **% del N total de columna**.
13. También se pueden añadir ambos a la lista de estadísticos de resumen personalizados para los totales y subtotales.
14. Pulse en **Aplicar a selección** y, a continuación, pulse en **Aceptar** para crear la tabla.

		Count	Column N %	Column Valid N %	Column Total N %	Valid N	Total N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	Total	9	100.0%	100.0%	100.0%	7	10

Figura 73. Tabla con valores perdidos y porcentajes válidos y totales

- *% del N de columna* indica el porcentaje en cada categoría en función del número de valores no perdidos y perdidos del usuario (ya que se han incluido en la tabla de forma explícita los valores perdidos del usuario).
- *% del N válido de columna* indica el porcentaje en cada categoría en función únicamente de los casos válidos, no perdidos. Estos valores son los mismos que tenían los porcentajes de columna en la tabla original, en la que no se incluían los valores perdidos del usuario.
- *% del N total de columna* indica el porcentaje en cada categoría en función de todos los casos, incluidos los valores perdidos del usuario y del sistema. Si suma los porcentajes de las categorías individuales en esta categoría, verá que la suma sólo alcanza el 90%, ya que uno del total de los 10 casos (10%) presenta el valor perdido del sistema. Si bien este caso se incluye en la base para el cálculo de los porcentajes, en la tabla no se proporciona ninguna categoría para los casos con valores perdidos del sistema.

Capítulo 12. Asignación de formato y personalización de tablas

Asignación de formato y personalización de tablas

La opción Tablas personalizadas ofrece la posibilidad de controlar ciertas propiedades del formato de las tablas como parte del proceso de creación de las mismas, entre las que se incluye:

- Formato de presentación y etiquetas para los estadísticos de resumen
- Ancho mínimo y máximo de la columna de datos
- Texto o valor mostrado en las casillas vacías

Esta configuración se conservará en la interfaz del generador de tablas (hasta que la modifique, restablezca la configuración del generador de tablas o abra un archivo de datos distinto), permitiéndole crear varias tablas con las mismas propiedades de formato, sin necesidad de modificar manualmente las tablas tras su creación. También puede guardar esta configuración del formato, junto con todos los demás parámetros de la tabla, utilizando el botón Pegar en la interfaz del generador de tablas para pegar la sintaxis de comandos en una ventana de sintaxis, que podrá almacenar a continuación como un archivo.

Asimismo, después de la creación de las tablas, podrá cambiar muchas de sus propiedades de formato, utilizando todas las funciones de formato disponibles en el Visor, para las tablas pivote. No obstante, este capítulo se centrará en controlar las propiedades de asignación de formato a las tablas antes de su creación. Si desea obtener más información acerca de las tablas pivote, utilice la pestaña Índice en el sistema de ayuda y escriba tablas pivote como palabra clave.

Archivo de datos de muestra

Los ejemplos de este capítulo utilizan el archivo de datos *survey_sample.sav*. Consulte el tema archivos de datos para obtener más información.

Todos los ejemplos que se incluyen muestran etiquetas variables en cuadros de diálogo, clasificadas en orden alfabético. Las propiedades de representación de la lista de variables se definen en la pestaña General del cuadro de diálogo Opciones (Menú Edición, Opciones).

Formato de presentación de los estadísticos de resumen

De forma predeterminada, la opción Tablas personalizadas intenta aplicar a los estadísticos formatos relativamente inteligentes, pero puede que en alguna ocasión desee omitir dichos valores predeterminados.

1. Elija en los menús:
Analizar > Tablas > Tablas personalizadas...
2. En el generador de tablas, arrastre *Categoría de edad* desde la lista de variables y colóquela en el área Filas del panel de lienzo.
3. Arrastre *Confianza en la televisión* y colóquela debajo de *Categoría de edad* en el área Filas, apilando las dos variables en la dimensión de las filas.
4. Pulse con el botón derecho del ratón en *Categoría de edad* en la vista previa de la tabla, en el panel de lienzo, y seleccione **Seleccionar todas las variables de fila** en el menú emergente.
5. Vuelva a pulsar con el botón derecho del ratón *Categoría de edad* y seleccione **Categorías y totales** en el menú emergente.
6. En el cuadro de diálogo Categorías y totales, seleccione (marque) **Total** y, a continuación, pulse en **Aplicar**.

7. Pulse con el botón derecho del ratón en una variable en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
8. Seleccione **% del N de columna** en la lista Estadísticos y pulse el botón de flecha para añadirlo a la lista Visualización.
9. Seleccione (marque) **Estadísticos de resumen personalizados para totales y subtotales**.
10. En la lista Estadísticos para los estadísticos de resumen personalizados, seleccione **% del N de columna** y pulse en la flecha para añadirlo a la lista Visualización.
11. Proceda de la misma manera para la **Media**.
12. A continuación, pulse en **Aplicar a todo**.

Los valores de los marcadores de la vista previa de la tabla reflejan el formato predeterminado para cada estadístico de resumen.

- Para los recuentos, el formato predeterminado de presentación es **nnnn**, correspondiente a valores enteros sin posiciones decimales.
- Para los porcentajes, el formato predeterminado de presentación es **nnnn.n%**, correspondiente a números con una única posición decimal y con el signo de porcentaje situado tras el valor.
- Para la media, el formato predeterminado de presentación es *distinto* para las dos variables.

Para los estadísticos de resumen que no sean de alguna forma un tipo de recuento (incluyendo N válido y N total) o porcentaje, el formato predeterminado de presentación será el formato definido para la variable, en el Editor de datos. Si observa las variables en la Vista de variables en el Editor de datos, verá que *Categoría de edad* (variable *edadcat*) aparece definida con dos posiciones decimales, mientras que *Confianza en la televisión* (variable *contv*) está definida con cero posiciones decimales.

Este es uno de esos casos en los que posiblemente el formato predeterminado no sea el formato que desee, ya que probablemente sería mejor que ambos valores de las medias mostraran el mismo número de decimales.

13. Pulse con el botón derecho del ratón en una variable en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú emergente.
Para la media, la casilla Formato de la lista Visualización indica que el formato es *Automático*, que significa que se utilizará el formato de presentación definido para la variable, y la casilla Decimales aparece desactivada. Para especificar el número de decimales, es necesario seleccionar primero un formato distinto.
14. En la lista Visualización de los estadísticos de resumen personalizados, pulse en la casilla Formato para la media y seleccione **nnnn** en la lista desplegable de formatos.
15. En la casilla Decimales, escriba el valor 1.
16. A continuación, pulse en **Aplicar a todo** para aplicar este ajuste a ambas variables.

Ahora la vista previa de la tabla indica que ambos valores de las medias se mostrarán con una posición decimal. (En este momento podría decidir finalizar y crear esta tabla; pero es posible que el valor "media" para *Categoría de edad* le resultara difícil de interpretar, debido a que los códigos numéricos reales para esta variable sólo van de 1 a 6.)

Etiquetas de presentación para los estadísticos de resumen

Además de los formatos de presentación para los estadísticos de resumen, también se pueden controlar las etiquetas descriptivas para cada estadístico de resumen.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en **Restablecer** para borrar cualquier configuración anteriormente realizada en el generador de tablas.
3. En el generador de tablas, arrastre *Categoría de edad* desde la lista de variables y colóquela en el área Filas del panel de lienzo.
4. Arrastre *Cómo le pagaron la semana pasada* desde la lista de variables y colóquela en el área Columnas del panel de lienzo.

- Pulse con el botón derecho del ratón en *Categoría de edad* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú contextual emergente.
- Seleccione **% del N de columna** en la lista Estadísticos y pulse el botón de flecha para añadirlo a la lista Visualización.
- Pulse con el botón derecho del ratón en cualquier lugar de la palabra *Columna* en la casilla Etiqueta de la lista Visualización para modificar el contenido de la casilla. Elimine la palabra *Columna* de la etiqueta, cambiando la etiqueta a simplemente %.
- Modifique la casilla Etiqueta para *Recuento* de la misma manera, cambiando la etiqueta a simplemente *N*.
Sin movernos de aquí, vamos a cambiar el formato del estadístico % del N de columna para eliminar el signo de porcentaje innecesario (ya que la etiqueta de la columna indica que la columna contiene porcentajes).
- Pulse en la casilla Formato para *% del N de columna* y seleccione **nnnn.n** de la lista desplegable de formatos.
- A continuación, pulse en **Aplicar a selección**.
La vista previa de la tabla muestra la modificación de las etiquetas y del formato de presentación.
- Pulse en **Aceptar** para crear la tabla.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figura 74. Tabla con las etiquetas de los estadísticos de resumen modificadas

Ancho de columna

Es posible que haya observado que la tabla del ejemplo anterior es bastante ancha. Una posible solución a este problema podría ser simplemente intercambiar las variables de filas y columnas. Otra solución sería estrechar las columnas, ya que parece que son más anchas de lo necesario. (De hecho, la razón por la que acortamos las etiquetas de los estadísticos de resumen era para poder estrechar ahora las columnas.)

- Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
- Pulse en la pestaña **Opciones**.
- En el grupo Ancho de columnas de datos, seleccione **Personalizado**.
- Para el máximo, indique 36. (Asegúrese de que el ajuste de la unidad de medida sea **Puntos**.)
- Pulse en **Aceptar** para crear la tabla.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figura 75. Tabla con anchos de columna reducidos

Ahora la tabla es mucho más compacta.

Valor de presentación para las casillas vacías

De forma predeterminada, se muestra un 0 en las casillas vacías (las casillas que no contienen ningún caso). En su lugar, puede no mostrar nada en dichas casillas (dejándolas en blanco) o especificar una cadena de texto para que se muestre en las casillas vacías.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Pulse en la pestaña **Opciones**.
3. En el grupo Aspecto de las casillas de datos, para Casillas vacías seleccione **Texto** y escriba Ninguno.
4. Pulse en **Aceptar** para crear la tabla.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	None	None	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figura 76. Tabla con "Ninguno" mostrado en las casillas vacías

Ahora las cuatro casillas vacías de la tabla muestran el texto *Ninguno* en lugar del valor 0.

Valor de presentación para los estadísticos perdidos

Si no se puede calcular un estadístico, el valor de visualización predeterminado es un punto (.), que es el símbolo que se utiliza para indicar el valor perdido del sistema. Este caso es distinto del de una casilla "vacía" y, por consiguiente, el valor de presentación para los estadísticos perdidos se controla de forma independiente del valor de presentación para las casillas que no contienen ningún caso.

1. Abra el generador de tablas (menú Analizar, Tablas, Tablas personalizadas).
2. Arrastre y coloque *Horas diarias viendo TV* desde la lista de variables hasta la parte superior del área Columnas del panel de lienzo, encima de *Cómo le pagaron la semana pasada*.
Como *Horas diarias viendo TV* es una variable de escala, automáticamente se convertirá en la variable origen de los estadísticos y el estadístico de resumen cambiará a la media.
3. Pulse con el botón derecho del ratón en *Horas diarias viendo TV* en la vista previa de la tabla en el panel de lienzo y seleccione **Estadísticos de resumen** en el menú contextual emergente.
4. Seleccione **N válido** en la lista Estadísticos y pulse el botón de flecha para añadirlo a la lista Visualización.
5. Pulse en **Aplicar a la selección**.
6. Pulse en la pestaña **Opciones**.
7. En el campo de texto para Estadísticos que no se pueden calcular, escriba ND.
8. Pulse en **Aceptar** para crear la tabla.

		Hours per day watching TV											
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N
Age category	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34
	55 to 64	3	40	NA	None	3	7	2	15	2	23	3	15
	65 or older	3	18	2	2	1	1	NA	0	1	2	3	11

Figura 77. Tabla con "ND" mostrado para los estadísticos perdidos

El texto *ND* se muestra para la media en tres casillas de la tabla. En cada caso, el correspondiente valor *N* *válido* explica por qué: no hay casos con los que calcular la media.

Sin embargo, es posible que observe lo que parece una ligera discrepancia: uno de los tres valores *N* *válido* aparece como 0, en lugar de la etiqueta *Ninguno* que se suponía que debía aparecer en las casillas que no tuvieran casos. Esto se debe a que, si bien no hay casos *válidos* para calcular la media, la categoría no está realmente vacía. Si regresa a la tabla original con sólo las dos variables categóricas, verá que en realidad hay tres casos en esta categoría de la tabulación cruzadas. No obstante, no hay ningún caso *válido* ya que los tres tienen valores perdidos para la variable de escala *Horas diarias viendo TV*.

Capítulo 13. 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 han recopilado 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 ¹ 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.

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

- **behavior.sav.** En un ejemplo clásico ², se pidió a 52 estudiantes que valoraran las combinaciones de 15 situaciones y 15 comportamientos en una escala de 10 puntos que oscila entre 0 = "extremadamente apropiado" y 9 = "extremadamente inapropiado". Los valores promediados respecto a los individuos se toman como disimilaridades.
- **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 mediciones 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 ³, 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 ⁴ 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 peso 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, ⁵, 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*.

2. Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

3. Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

4. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

5. Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **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 ⁶. 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.

6. Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **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 del 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 recibieron un envío de correos 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"⁷. 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 Databases⁸ 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.** Bell⁹ presentó una tabla para ilustrar posibles grupos sociales. Guttman¹⁰ 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).

7. Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

- **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 señalar 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 ¹¹ comenzaron a analizar 15 términos de parentesco (tía, hermano, primo, 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 “orígenes”. Cada origen se corresponde con una matriz de proximidades de 15 x 15 cuyas casillas son iguales al número de personas de un origen menos el número de veces que se partitionaron los objetos en ese origen.
- **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 1999–2000. Los datos de estas ventas son públicos.
- **nhis2000_subset.sav.** La National Health Interview Survey (NHIS, encuesta del Centro Nacional de Estadísticas de Salud de EE.UU.) es una encuesta detallada realizada entre la población civil de Estados Unidos. Las encuestas se realizaron en persona a una muestra representativa de las unidades familiares del país. Se recogió tanto la información demográfica como las observaciones acerca del estado y los hábitos de salud de los integrantes de cada unidad familiar. Este archivo de datos contiene un subconjunto de información de la encuesta de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Archivo de datos y documentación de uso público. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Fecha de acceso: 2003.

11. Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

- **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^{12, 13} han encontrado que no hay linealidad entre estas variables, lo que dificulta los métodos de regresión estándar.
- **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

12. Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

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 delinquentes 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 ¹⁴.
- **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 ¹⁵ 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.

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **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 logarímicamente se han eliminado y sustituido por variables de gasto del cliente transformadas logarímicamente 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 predeterminado de las variables antes de realizar la asignación de nivel de medición 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 ¹⁶.
- **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 ¹⁷.
- **verd1985.sav.** Archivo de datos sobre una encuesta ¹⁸. 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 *masкота* y *noticia*, mientras que el conjunto 3 incluye *música* y *vivir*. Se escala *masкота* como nominal múltiple y *edad* como ordinal; el resto de variables se escalan como nominal simple.

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

- **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 ¹⁹. 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.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Avisos

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

Es posible que IBM no ofrezca en otros países los productos, servicios o características que se describen en este documento. Consulte al representante local de IBM para obtener información sobre los productos y servicios disponibles actualmente en su zona. Las referencias a productos, programas o servicios de IBM no pretenden establecer ni implicar que sólo puedan utilizarse productos, programas o servicios de IBM. En su lugar se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no vulnere ningún derecho de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes que cubran la materia descrita en esta información. 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
U.S.A.

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

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

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones , por lo tanto , esta declaración no se aplique a usted.

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

Cualquier referencia a sitios Web que no sean de IBM en esta información sólo es ofrecida por comodidad y de ningún modo sirve como aprobación de esos sitios Web. Los materiales de esos sitios web no forman parte de los materiales de este producto de IBM y la utilización de esos sitios web será responsabilidad del usuario.

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

Los titulares de licencias de este programa que deseen obtener información sobre el mismo con el fin de permitir: (i) el intercambio de información entre programas creados independientemente y otros programas (incluido éste) y (ii) el uso mutuo de información que se haya intercambiado, deben ponerse en contacto con:

Tel.: 900 100 400
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
EE.UU.

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

El programa bajo licencia descrito en este documento y todo el material bajo licencia disponible se proporcionan bajo los términos de IBM Customer Agreement, IBM International Program License Agreement o cualquier otro acuerdo equivalente entre IBM y el cliente.

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

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado estos productos y no puede confirmar la precisión del rendimiento, la compatibilidad u otros aspectos relacionados con los productos que no son de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

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

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

Cada copia o cada parte de los programas de ejemplo o de los trabajos que se deriven de ellos debe incluir un aviso de copyright como se indica a continuación:

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.

Cada copia o cada parte de los programas de ejemplo o de los trabajos que se deriven de ellos debe incluir un aviso de copyright como se indica a continuación:

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

© Copyright IBM Corp. _escriba el año o años_. Reservados todos los derechos.

Marcas comerciales

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

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

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

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

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

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

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

Índice

A

- ancho de columna
 - control en tablas personalizadas 15, 89
- anidación de variables
 - Tablas personalizadas 28, 30
 - variables de escala 61
- apilado de variables
 - diferentes estadísticos de resumen para distintas variables 59
 - múltiples variables origen de los estadísticos de resumen 53
 - Tablas personalizadas 27
 - variables de capa apiladas 32
 - variables de escala 57
- archivos de ejemplo
 - ubicación 93

C

- casillas vacías
 - valor visualizado en tablas personalizadas 15, 90
- categorías calculadas
 - a partir de subtotales 42
 - formatos de presentación 15
 - ocultación de categorías en una expresión 42
 - Tablas personalizadas 14, 41
- categorías calculadas posteriormente
 - Tablas personalizadas 14, 41
- chi-cuadrado
 - Tablas personalizadas 63
- comprobaciones de la significación
 - conjuntos de respuestas múltiples 80, 81
- conjuntos de respuestas múltiples 75
 - comprobación de la significación 73, 80, 81
 - duplicar respuestas en conjuntos de categorías múltiples 15
 - porcentajes 8
- contracción de categorías
 - Tablas personalizadas 39
- control de números decimales mostrados 22

D

- desviación estándar
 - Tablas personalizadas 9
- diferentes estadísticos de resumen para distintas variables
 - tablas apiladas 59

E

- eliminación de categorías
 - Tablas personalizadas 11, 24

- estadísticos
 - estadísticos de resumen 51
 - estadísticos de resumen del total personalizados 54
 - tablas apiladas 53
- estadísticos de contraste
 - Tablas personalizadas 17, 63
- estadísticos de medias de columna
 - tablas personalizadas 66
- estadísticos de proporciones de columna
 - tablas personalizadas 69
- estadísticos de resumen 51
 - cambio del texto de las etiquetas 88
 - diferentes resúmenes para distintas variables en tablas apiladas 59
 - dimensión origen 51
 - estadísticos de resumen del total personalizados 54
 - formato de presentación 87
 - tablas apiladas 53
 - variable origen 51
- estadísticos de resumen del total personalizados 54
- etiquetas
 - cambio del texto de las etiquetas para los estadísticos de resumen 88
- etiquetas de esquina
 - Tablas personalizadas 16
- etiquetas de valor y valores 55
- etiquetas de variable
 - eliminación de visualización en las tablas personalizadas 5
- exclusión de categorías
 - Tablas personalizadas 11, 24

F

- fecha
 - adición de la fecha actual en las tablas personalizadas 16
- formatos de presentación 22
 - estadísticos de resumen en tablas personalizadas 10, 87

H

- hora
 - adición de la hora actual en las tablas personalizadas 16

I

- impresión de tablas con capas 33
- intervalo
 - Tablas personalizadas 9

M

- máximo
 - Tablas personalizadas 9
- media 57
 - Tablas personalizadas 9
- mediana 58
 - Tablas personalizadas 9
- mínimo
 - Tablas personalizadas 9
- moda
 - Tablas personalizadas 9

N

- N total 84
- N válido 58, 84
 - Tablas personalizadas 9
- nivel de medición
 - cambio en las tablas personalizadas 1

O

- ocultación de las etiquetas de los estadísticos en las tablas personalizadas 19
- omisión de categorías
 - Tablas personalizadas 24
- ordenación de categorías
 - Tablas personalizadas 24

P

- pies
 - Tablas personalizadas 16
- porcentajes
 - conjuntos de respuestas múltiples 8
 - en las tablas personalizadas 7, 8, 20, 22
 - valores perdidos 84
- posiciones decimales
 - control del número de decimales que aparecen en las tablas personalizadas 6, 22, 87
- presentación de los valores de categoría 55
- procesamiento de segmentación de archivos
 - tablas personalizadas 4
- pruebas de significación
 - Tablas personalizadas 17

R

- recuento
 - frente a N válido 58
- reordenación de categorías
 - Tablas personalizadas 11

resúmenes agrupados
variables de escala 60

S

subtotales 38
ocultación de categorías de
subtotal 39
Tablas personalizadas 11, 35
suma
Tablas personalizadas 9

T

tabla de frecuencias
Tablas personalizadas 15, 47
tablas
Tablas personalizadas 1
tablas comperimeter 15, 47
tablas de frecuencia media 10, 54
tablas personalizadas
ancho de columna 15
cambio de dimensión de estadísticos
de resumen 10
cambio de nivel de medición 1
casillas vacías 15
categorías calculadas 11
cómo generar una tabla 4
cómo mostrar y ocultar nombres y
etiquetas de variables 5
conjuntos de categorías múltiples 15
conjuntos de respuestas múltiples 1
control de números decimales
mostrados 6
estadísticos de contraste 17
estadísticos de resumen 7, 8, 9
etiquetas de esquina 16
etiquetas de valores para las variables
categóricas 1
exclusión de categorías 11
exclusión de valores perdidos para
resúmenes de escala 15
formatos de presentación 6
Formatos de presentación de
estadísticos de resumen 10
pies 16
porcentajes 7, 8
porcentajes de conjuntos de respuestas
múltiples 8
procesamiento de segmentación de
archivos 4
reordenación de categorías 11
subtotales 11
tabla de frecuencias 15, 47
tablas comperimeter 15, 47
tablas de frecuencia media 10
tablas de variables con categorías
compartidas 15, 47
títulos 16
totales 11
totales personalizados 10
variables categóricas 1
variables de escala 1
Tablas personalizadas
anidación de variables 28, 30
anidación de variables de capa 32

Tablas personalizadas (*continuación*)
apilado de variables 27
cambio de las etiquetas de los
estadísticos de resumen 20
categorías calculadas 14, 41
categorías calculadas
posteriormente 14, 41
comprobación de significación y
respuesta múltiple 73
conjuntos de respuestas múltiples 75
contracción de categorías 39
dimensión origen de estadísticos 22
estadísticos de contraste 63
exclusión de categorías 24
impresión de tablas con capas 33
intercambio de variables de fila y de
columna 31
ocultación de categorías de
subtotal 39
ocultación de las etiquetas de los
estadísticos 19
ordenación de categorías 24
porcentajes 20, 22
porcentajes de fila frente a los de
columna 20
subtotales 35
tablas simples para variables
categóricas 19
tabulación cruzada 21
totales 21, 35
totales en tablas con categorías
excluidas 24
totales marginales 23
variables de capas 31, 32
vista compacta 30
tabulación cruzada
Tablas personalizadas 21
títulos
Tablas personalizadas 16
totales
capas 38
categorías excluidas 35
posición de presentación 36
tablas anidadas 36
Tablas personalizadas 11, 21, 35
totales de grupo 36
totales marginales para tablas
personalizadas 23
totales de grupo 36
totales de subgrupo 36

V

valores
presentación de los valores y las
etiquetas de categoría 55
valores perdidos 58, 83
efecto en el cálculo de los
porcentajes 84
inclusión en tablas personalizadas 84
valores perdidos del sistema 83
valores perdidos del usuario 83
variable origen de los estadísticos de
resumen
variables de escala 61
variables de capas
anidación de variables de capa 32

variables de capas (*continuación*)
impresión de tablas con capas 33
Tablas personalizadas 31, 32
variables de capa apiladas 32
variables de escala
anidación 61
apilación 57
estadísticos de resumen 57
múltiples estadísticos de resumen 58
resúmenes agrupados 60
resúmenes agrupados por variables
categóricas de filas y columnas 61
varianza
Tablas personalizadas 9



Impreso en España