

IBM SPSS Complex Samples 23

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información de la sección "Avisos" en la página 59.

Información sobre el producto

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

Contenido

Capítulo 1. Introducción a los procedimientos de muestras complejas . 1

| | |
|---|---|
| Propiedades de las muestras complejas | 1 |
| Uso de los procedimientos de Muestras complejas | 2 |
| Archivos de plan | 2 |
| Lecturas adicionales | 3 |

Capítulo 2. Muestreo a partir de un diseño complejo 5

| | |
|---|----|
| Creación de un nuevo plan de muestreo | 5 |
| Asistente de muestreo: Variables del diseño | 5 |
| Controles de árbol para navegar por el Asistente de muestreo | 6 |
| Asistente de muestreo: Método de muestreo | 6 |
| Asistente de muestreo: Tamaño de la muestra | 7 |
| Definir tamaños desiguales | 8 |
| Asistente de muestreo: Variables de resultado | 8 |
| Asistente de muestreo: Resumen del plan | 9 |
| Asistente de muestreo: Extraer muestra, opciones de selección | 9 |
| Asistente de muestreo: Extraer muestra, archivos de resultado | 9 |
| Asistente de muestreo: Finalizar | 10 |
| Modificar un plan de muestreo existente | 10 |
| Asistente de muestreo: Resumen del plan | 10 |
| Ejecutar un plan de muestreo existente | 10 |
| Características adicionales de los comandos CSPLAN y CSSELECT | 11 |

Capítulo 3. Preparación de una muestra compleja para su análisis 13

| | |
|---|----|
| Creación de un nuevo plan de análisis | 13 |
| Asistente de preparación del análisis: Variables del diseño | 13 |
| Controles de árbol para desplazarse por el Asistente para el análisis | 14 |
| Asistente de preparación del análisis: Método de estimación | 14 |
| Asistente de preparación del análisis: Tamaño | 14 |
| Definir tamaños desiguales | 15 |
| Asistente de preparación del análisis: Resumen del plan | 15 |
| Asistente de preparación del análisis: Finalizar | 15 |
| Modificar un plan de análisis existente | 16 |
| Asistente de preparación del análisis: Resumen del plan | 16 |

Capítulo 4. Plan de muestras complejas 17

Capítulo 5. Frecuencias de muestras complejas 19

| | |
|---|----|
| Frecuencias de Muestras complejas: Estadísticos | 19 |
| Muestras complejas: Valores perdidos | 20 |
| Opciones de Muestras complejas | 20 |

Capítulo 6. Descriptivos de muestras complejas 23

| | |
|--|----|
| Descriptivos de Muestras complejas: Estadísticos | 23 |
| Valores perdidos en los descriptivos de Muestras complejas | 24 |
| Opciones de Muestras complejas | 24 |

Capítulo 7. Tablas cruzadas de muestras complejas 25

| | |
|---|----|
| Tablas cruzadas de Muestras complejas: Estadísticos | 25 |
| Muestras complejas: Valores perdidos | 26 |
| Opciones de Muestras complejas | 27 |

Capítulo 8. Razones de muestras complejas 29

| | |
|---|----|
| Razones de Muestras complejas. Estadísticos | 29 |
| Razones de Muestras complejas: Valores perdidos | 30 |
| Opciones de Muestras complejas | 30 |

Capítulo 9. Modelo lineal general de muestras complejas 31

| | |
|--|----|
| Modelo lineal general de muestras complejas | 31 |
| Estadísticos de Modelo lineal general de muestras complejas | 32 |
| Muestras complejas: Contrastes de hipótesis | 33 |
| Medias estimadas del Modelo lineal general de muestras complejas | 33 |
| Modelo lineal general de muestras complejas: Guardar | 34 |
| Modelo lineal general de muestras complejas: Opciones | 34 |
| Características adicionales del comando CSGLM | 35 |

Capítulo 10. Regresión logística de muestras complejas 37

| | |
|--|----|
| Regresión logística de muestras complejas: Categoría de referencia | 37 |
| Regresión logística de muestras complejas: Modelo | 38 |
| Regresión logística de muestras complejas: Estadísticos | 38 |
| Muestras complejas: Contrastes de hipótesis | 39 |
| Regresión logística de muestras complejas: Razones de las ventajas | 40 |
| Regresión logística de muestras complejas: Guardar | 40 |
| Regresión logística de muestras complejas: Opciones | 41 |
| Características adicionales del comando CSLOGISTIC | 41 |

Capítulo 11. Regresión ordinal de muestras complejas 43

| | |
|--|----|
| Regresión ordinal de muestras complejas: Probabilidades de respuesta | 44 |
| Regresión ordinal de muestras complejas: Modelo | 44 |

| | |
|--|----|
| Regresión ordinal de muestras complejas: | |
| Estadísticos | 44 |
| Muestras complejas: Contrastes de hipótesis | 45 |
| Regresión ordinal de muestras complejas: Razones de las ventajas | 46 |
| Regresión ordinal de muestras complejas: Guardar | 46 |
| Regresión ordinal de muestras complejas: Opciones | 47 |
| Características adicionales del comando CSORDINAL | 48 |

| | |
|--|-----------|
| Capítulo 12. Regresión de Cox de muestras complejas | 49 |
| Definir evento | 50 |
| Predictores | 51 |
| Definir predictor dependiente del tiempo | 51 |

| | |
|--|----|
| Subgrupos. | 51 |
| Modelo. | 52 |
| Estadísticas | 52 |
| Gráficos | 53 |
| Contrastes de hipótesis | 54 |
| Guardar | 54 |
| Exportar | 55 |
| Opciones | 57 |
| Características adicionales del comando CSCOXREG | 58 |

| | |
|------------------------------|-----------|
| Avisos | 59 |
| Marcas comerciales | 61 |

| | |
|-------------------------|-----------|
| Índice | 63 |
|-------------------------|-----------|

Capítulo 1. Introducción a los procedimientos de muestras complejas

Un supuesto inherente a los procedimientos de análisis en los paquetes de software tradicionales es que las observaciones de un archivo de datos representan una muestra aleatoria simple de la población de interés. Este supuesto es insostenible para un número cada vez mayor de empresas e investigadores que consideran más económico y cómodo obtener las muestras de una forma más estructurada.

La opción Muestras complejas permite seleccionar una muestra de acuerdo con un diseño complejo e incorporar las especificaciones del diseño al análisis de los datos, asegurando así que los resultados serán válidos.

Propiedades de las muestras complejas

Una muestra compleja puede ser distinta de una muestra aleatoria simple en muchos aspectos. En una muestra aleatoria simple, las unidades de muestreo individuales se seleccionan aleatoriamente con la misma probabilidad y sin reposición (SR) directamente a partir de la totalidad de la población. Por lo contrario, una muestra compleja determinada puede tener alguna o todas las características siguientes:

Estratificación. El muestreo estratificado implica seleccionar muestras independientemente dentro de los subgrupos de la población que no se solapan o estratos. Por ejemplo, los estratos pueden ser grupos socioeconómicos, categorías laborales, grupos de edad o grupos étnicos. Con la estratificación, puede asegurarse que los tamaños muestrales de los subgrupos de interés son adecuados, mejorar la precisión de las estimaciones globales y utilizar distintos métodos de muestreo entre los diferentes estratos.

Clústeres. El muestreo por clústeres implica la selección de grupos de unidades muestrales o clústeres. Por ejemplo, los clústeres pueden ser escuelas, hospitales o zonas geográficas y las unidades muestrales pueden ser alumnos, pacientes o ciudadanos. El clúster es común en los diseños polietápicos y en las muestras de zona (geográfica).

Múltiples etapas. En el muestreo polietápico, se selecciona una muestra de primera etapa basada en clústeres. A continuación, se crea una muestra de segunda etapa extrayendo submuestras a partir de los clústeres seleccionados. Si la muestra de segunda etapa está basada en subclústeres, entonces puede añadir una tercera etapa a la muestra. Por ejemplo, en la primera etapa de una encuesta, se podría extraer una muestra de ciudades. A continuación, y a partir de las ciudades seleccionadas, se podrían muestrear unidades familiares. Finalmente, a partir de las unidades familiares seleccionadas, se podría encuestar a individuos. Los Asistentes de muestreo y preparación del análisis permiten especificar tres etapas en un diseño.

Muestreo no aleatorio. Cuando es difícil obtener la muestra aleatoriamente, las unidades se pueden muestrear sistemáticamente (con un intervalo fijo) o secuencialmente.

Probabilidades de selección desiguales. Cuando se muestrean clústeres que contienen números de unidades desiguales, puede utilizarse el muestreo probabilístico proporcional al tamaño (PPS) para que la probabilidad de selección del clúster sea igual a la proporción de unidades que contiene. El muestreo PPS también puede utilizar esquemas de ponderación más generales para seleccionar unidades.

Muestreo no restringido. El muestreo no restringido selecciona las unidades con reposición (CR). Por lo tanto, se puede seleccionar más de una vez una unidad individual para la muestra.

Ponderaciones de muestreo. Las ponderaciones de muestreo se calculan automáticamente al extraer una muestra compleja y de forma ideal se corresponden con la "frecuencia" que cada unidad muestral

representa en la población objetivo. Por lo tanto, la suma de las ponderaciones muestrales debe estimar el tamaño de la población. Los procedimientos de análisis de muestras complejas requieren las ponderaciones de muestreo para poder analizar correctamente una muestra compleja. Tenga en cuenta que estas ponderaciones se deben utilizar exclusivamente dentro de la opción Muestras complejas y no con otros procesos analíticos a través del procedimiento Ponderar casos, el cual trata las ponderaciones como réplicas de casos.

Uso de los procedimientos de Muestras complejas

El uso de los procedimientos de Muestras complejas depende de las necesidades específicas. Los tipos fundamentales de usuarios son aquéllos que:

- Planifican y llevan a cabo encuestas de acuerdo con diseños complejos, analizando posiblemente la muestra más tarde. La herramienta principal de los encuestadores es el Asistente de muestreo.
- Analiza archivos de datos muestrales obtenidos previamente según diseños complejos. Antes de utilizar los procedimientos de análisis de muestras complejas puede que deba utilizar el Asistente de preparación del análisis.

Independientemente del tipo de usuario que sea, debe proporcionar información del diseño a los procedimientos de Muestras complejas. Esta información está almacenada en un **archivo de plan** para volver a utilizarla con mayor facilidad.

Archivos de plan

Los archivos de plan contienen especificaciones de la muestra compleja. Existen dos tipos de archivos de plan:

Plan de muestreo. Las especificaciones dadas en el Asistente de muestreo definen un diseño muestral que se utiliza para extraer una muestra compleja. El archivo del plan de muestreo contiene esas especificaciones. El archivo del plan de muestreo también contiene un plan de análisis predeterminado que utiliza métodos de estimación adecuados para el diseño muestral especificado.

Plan de análisis. Este archivo de plan contiene la información necesaria en los procedimientos de análisis de Muestras complejas para calcular correctamente las estimaciones de la varianza de una muestra compleja. El plan incluye la estructura de la muestra, los métodos de estimación de cada etapa y las referencias para variables necesarias como por ejemplo, las ponderaciones muestrales. El Asistente de preparación del análisis permite crear y editar los planes de análisis.

Existen distintas ventajas al guardar las especificaciones en un archivo de plan, por ejemplo:

- Un encuestador puede especificar la primera etapa de un plan de muestreo de varias etapas y extraer en el momento las unidades de la primera etapa, reunir información sobre las unidades muestrales para la segunda etapa y a continuación, modificar el plan de muestreo para incluir la segunda etapa.
- Un analista que no tenga acceso al archivo del plan de muestreo puede especificar un plan de análisis y hacer referencia a ese plan en cada procedimiento de análisis de Muestras complejas.
- Un diseñador de muestras a gran escala de uso público puede publicar el archivo del plan de muestreo, lo que simplifica las instrucciones para el analista y evita que cada analista deba especificar sus propios planes de análisis.

Lecturas adicionales

Si desea obtener más información sobre las técnicas de muestreo, consulte los siguientes textos:

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Capítulo 2. Muestreo a partir de un diseño complejo

El Asistente de muestreo le guía a través de los pasos necesarios para crear, modificar o ejecutar un archivo de plan de muestreo. Antes de utilizar el Asistente, debe tener en mente una población objetivo bien definida, una lista de las unidades muestrales y un diseño muestral adecuado.

Creación de un nuevo plan de muestreo

1. Seleccione en los menús:
Analizar > Muestras complejas > Seleccionar una muestra...
2. Seleccione **Diseñar una muestra** y elija un nombre de archivo de plan para guardar el plan de muestreo.
3. Pulse **Siguiente** para continuar usando el Asistente.
4. Si lo desea, en el paso Variables del diseño puede definir estratos, clústeres e introducir ponderaciones muestrales. Después de definirlos, pulse **Siguiente**.
5. Si lo desea, en el paso Método de muestreo, puede elegir un método para seleccionar los elementos. Si selecciona **Muestreo de Brewer proporcional al tamaño** o **Muestreo de Murthy proporcional al tamaño**, puede pulsar **Finalizar** para extraer la muestra. En caso contrario, pulse **Siguiente** y a continuación:
6. En el paso Tamaño de la muestra, especifique el número o proporción de unidades que muestrear.
7. Ahora puede pulsar **Finalizar** para extraer la muestra.

Si lo desea, en los siguientes pasos puede:

- Elegir las variables de resultado para guardar.
- Añadir una segunda o tercera etapa al diseño.
- Establecer varias opciones de selección, incluyendo las etapas a partir de las cuales se van a extraer las muestras, la semilla de aleatorización y si los valores perdidos del usuario se van a tratar como valores válidos de las variables del diseño.
- Elegir dónde guardar los datos de resultado.
- Pegar las selecciones como sintaxis de comandos.

Asistente de muestreo: Variables del diseño

Este paso permite seleccionar las variables de estratificación y de clúster y definir unas ponderaciones muestrales de entrada. También puede especificar una etiqueta para la etapa.

Estratificar por. La clasificación cruzada por las variables de estratificación define distintas subpoblaciones o estratos. Se obtienen muestras individuales para cada estrato. Para mejorar la precisión de las estimaciones, las unidades de los estratos deben ser tan homogéneas como sea posible respecto a las características de interés.

Clústeres. Las variables de clúster definen grupos de unidades de observación o clústeres. Los clústeres son útiles cuando es difícil o imposible realizar el muestreo de las unidades de observación directamente desde la población; en su lugar, se puede realizar el muestreo de los clústeres a partir de la población y a continuación, realizar el muestreo de las unidades de observación a partir de los clústeres seleccionados. Sin embargo, el uso de clústeres puede introducir correlaciones entre las unidades muestrales, con la consiguiente pérdida de precisión. Para minimizar este efecto, las unidades de los clústeres deben ser tan heterogéneas como sea posible respecto a las características de interés. Deberá definir una variable de

clúster como mínimo para planificar un diseño de varias etapas. Los clústeres también son necesarios al utilizar distintos métodos de muestreo. Consulte el tema “Asistente de muestreo: Método de muestreo” para obtener más información.

Introducir ponderación muestral. Si el diseño muestral actual forma parte de un diseño muestral mayor, puede disponer de ponderaciones muestrales de una etapa anterior del diseño mayor. Puede especificar una variable numérica que contenga estas ponderaciones en la primera etapa del diseño actual. Las ponderaciones muestrales se calculan automáticamente para las etapas posteriores del diseño actual.

Etiqueta de etapa. Puede especificar una etiqueta de cadena opcional para cada etapa. Esto se utiliza en los resultados para facilitar la identificación de la información por etapas.

Nota: la lista de variables origen tiene el mismo contenido a lo largo de los pasos del Asistente. En otras palabras, las variables de la lista de origen eliminadas en un paso determinado se borran de la lista en todos los pasos. Las variables devueltas a la lista de origen aparecen en la lista en todos los pasos.

Controles de árbol para navegar por el Asistente de muestreo

En la parte izquierda de cada paso del Asistente de muestreo se muestra un esquema de los titulares de todos los pasos. Puede navegar por el Asistente al pulsar el nombre de uno de los pasos activados en los titulares. Los pasos están activados cuando todos los pasos anteriores sean válidos, es decir, si cada uno de los pasos anteriores dispone de las especificaciones mínimas necesarias para ese paso. Consulte la ayuda de los pasos individuales para obtener más información sobre los motivos por los que un paso determinado puede no ser válido.

Asistente de muestreo: Método de muestreo

Este paso permite especificar cómo seleccionar los casos del conjunto de datos activo.

Método. Los controles de este grupo se utilizan para elegir un método de selección. Algunos tipos de muestreo permiten elegir entre realizar un muestreo con reposición (CR) o sin reposición (SR). Si desea obtener más información, consulte las descripciones de los tipos. Tenga en cuenta que algunos tipos de probabilidad proporcional al tamaño (PPS) están disponibles sólo cuando se han definido clústeres y todos los tipos de PPS están disponibles sólo en la primera etapa de un diseño. Además, los métodos SR están disponibles sólo en la última etapa de un diseño.

- **Muestreo aleatorio simple:** Las unidades se seleccionan con probabilidad igual. Se pueden seleccionar con o sin reposición.
- **Sistemático simple.** Las unidades se seleccionan con un intervalo fijo en todo el marco muestral (o en los estratos, si se han especificado) y se extraen sin reposición. Se selecciona una unidad aleatoriamente dentro del primer intervalo como el punto inicial.
- **Secuencial simple.** Las unidades se seleccionan de forma secuencial con probabilidad igual y sin reposición.
- **Probabilidad proporcional al tamaño.** Método de primera etapa que selecciona unidades de forma aleatoria con probabilidad proporcional al tamaño. Se puede seleccionar cualquier unidad con reposición; sólo se puede realizar muestreo sin reposición de los clústeres.
- **Muestreo sistemático proporcional al tamaño.** Método de primera etapa que selecciona unidades de forma sistemática con probabilidad proporcional al tamaño. Se seleccionan sin reposición.
- **Muestreo secuencial proporcional al tamaño.** Método de primera etapa que selecciona unidades de forma secuencial con probabilidad proporcional al tamaño del clúster y sin reposición.
- **Muestreo de Brewer proporcional al tamaño.** Método de primera etapa que selecciona dos clústeres de cada estrato con probabilidad proporcional al tamaño del clúster y sin reposición. Se debe especificar una variable de clúster para utilizar este método.

- **Muestreo de Murthy proporcional al tamaño.** Método de primera etapa que selecciona dos clústeres de cada estrato con probabilidad proporcional al tamaño del clúster y sin reposición. Se debe especificar una variable de clúster para utilizar este método.
- **Muestreo de Sampford proporcional al tamaño.** Método de primera etapa que selecciona más de dos clústeres de cada estrato con probabilidad proporcional al tamaño del clúster y sin reposición. Es una extensión del método de Brewer. Se debe especificar una variable de clúster para utilizar este método.
- **Usar estimación CR para el análisis.** De forma predeterminada, el método de estimación se especifica en el archivo de plan de manera coherente con el método de muestreo seleccionado. Esta opción permite utilizar la estimación con reposición incluso si el método de muestreo implica la estimación SR. Esta opción solamente está disponible en la etapa 1.

Medida del tamaño (MDT). Si se selecciona un método PPS, deberá especificar una medida del tamaño que defina el tamaño de cada unidad. Estos tamaños pueden definirse explícitamente en una variable o se pueden calcular a partir de los datos. Opcionalmente, se pueden establecer los límites inferior y superior de la MDT, anulando cualquier valor encontrado en la variable MDT o calculado a partir de los datos. Estas opciones solamente están disponibles en la etapa 1.

Asistente de muestreo: Tamaño de la muestra

Este paso permite especificar el número o la proporción de unidades que se van a muestrear dentro de la etapa actual. El tamaño de la muestra puede ser fijo o variar entre estratos. Para el propósito de especificar el tamaño de la muestra, se pueden utilizar los clústeres elegidos en etapas anteriores para definir estratos.

Unidades. Puede especificar un tamaño de la muestra exacto o una proporción de unidades a muestrear.

- **Valor.** Se aplica un valor particular a todos los estratos. Si se selecciona **Recuentos** como la unidad métrica, deberá introducir un entero positivo. Si se selecciona **Proporciones**, deberá introducir un valor no negativo. A no ser que se realice una muestra con reposición, los valores de proporción no deberán ser mayores que 1.
- **Valores desiguales para estratos.** Permite introducir distintos valores de tamaño para cada estrato a través del cuadro de diálogo Definir tamaños desiguales.
- **Leer valores de la variable.** Permite seleccionar una variable numérica que contenga los valores de tamaño para los estratos.

Si se selecciona **Proporciones**, tiene la opción de establecer los límites inferior y superior para el número de unidades muestreadas.

Definir tamaños desiguales

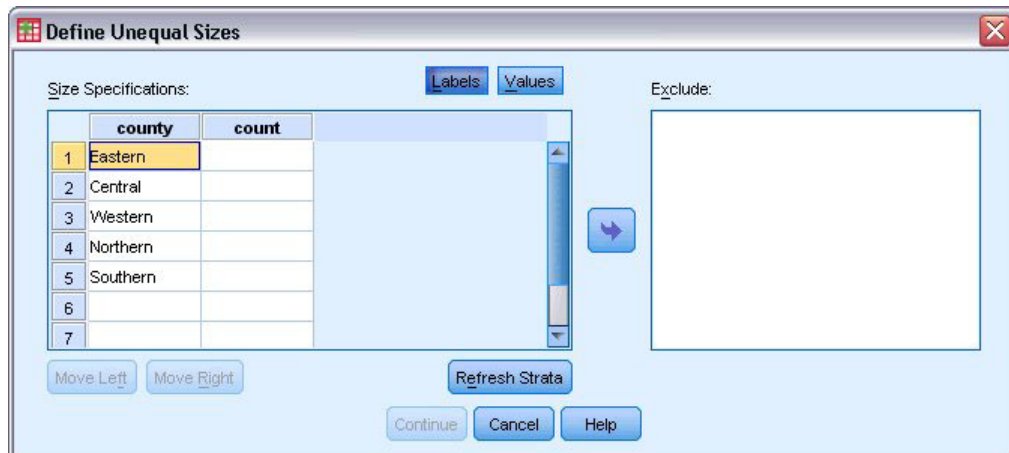


Figura 1. Cuadro de diálogo Definir tamaños desiguales

El cuadro de diálogo Definir tamaños desiguales permite introducir los tamaños para cada estrato.

Cuadrícula de especificaciones de tamaño. La cuadrícula muestra la clasificación cruzada de hasta cinco variables de clúster o estrato, con una combinación de estrato/clúster por fila. Las variables elegibles en la cuadrícula serán todas las variables de estratificación de las etapas anteriores y actuales además de todas las variables de clúster de las etapas anteriores. Las variables se pueden reordenar dentro de la cuadrícula o ser desplazadas a la lista Excluir. Introduzca los tamaños en la última columna de la derecha. Pulse en **Etiquetas** o **Valores** para conmutar entre la visualización de las etiquetas de valor y los valores de los datos para las variables de estratificación y de clúster de las casillas de la cuadrícula. Las casillas que contienen valores sin etiquetas siempre muestran valores. Pulse **Actualizar estratos** para volver a rellenar la cuadrícula con cada combinación de los valores de los datos etiquetados para las variables de la cuadrícula.

Excluir. Para especificar los tamaños de un subconjunto de combinaciones de estrato/clúster, desplace una o más variables a la lista Excluir. Estas variables no se utilizan para definir tamaños muestrales.

Asistente de muestreo: Variables de resultado

Este paso permite elegir las variables que desea guardar cuando se extraiga la muestra.

Tamaño poblacional. El número estimado de unidades en la población de una etapa dada. El nombre raíz de la variable guardada es *TamañoPoblación_*.

Proporción muestral. Tasa de la muestra en una etapa dada. El nombre raíz de la variable guardada es *TasaMuestreo_*.

Tamaño de la muestra. Número de unidades extraídas en una etapa dada. El nombre raíz de la variable guardada es *TamañoMuestral_*.

Ponderación muestral. La inversa de las probabilidades de inclusión. El nombre raíz de la variable guardada es *PonderaciónMuestral_*.

Algunas variables por etapa se generan automáticamente. Entre éstos se incluyen:

Probabilidades de inclusión. Proporción de unidades extraídas en una etapa dada. El nombre raíz de la variable guardada es *ProbabilidadInclusión_*.

Ponderación acumulada. Ponderación muestral acumulada a lo largo de las etapas anteriores a la actual e incluyendo esta última. El nombre raíz de la variable guardada es *PonderaciónMuestralAcumulada_*.

Índice. Identifica las unidades seleccionadas varias veces dentro de una etapa dada. El nombre raíz de la variable guardada es *Índice_*.

Nota: los nombres raíz de la variable guardada incluyen un sufijo entero que refleja el número de la etapa, por ejemplo, *TamañoPoblación_1_* para el tamaño de la población guardada de la etapa 1.

Asistente de muestreo: Resumen del plan

Último paso de cada etapa que proporciona un resumen de las especificaciones del diseño muestral hasta la etapa actual. A partir de aquí, puede pasar a la siguiente etapa (creándola si es necesario) o definir las opciones para extraer la muestra.

Asistente de muestreo: Extraer muestra, opciones de selección

Este paso permite elegir si desea extraer una muestra. También puede controlar otras opciones del muestreo, como la semilla aleatoria y el tratamiento de los valores perdidos.

Extraer muestra. Además de elegir si desea extraer una muestra, también puede elegir ejecutar parte del diseño del muestreo. Las etapas se deben extraer en orden (es decir, la etapa 2 no se puede extraer a menos que ya se haya extraído la etapa 1). Al editar o ejecutar un plan, no puede volver a muestrear etapas bloqueadas.

Semilla. Permite elegir un valor de semilla para la generación de números aleatorios.

Incluye los valores perdidos del usuario. Determina si los valores perdidos del usuario son tratados como válidos. Si es así, los valores perdidos del usuario se tratan como una categoría diferente.

Los datos ya están ordenados. Si el marco muestral está clasificado previamente por los valores de las variables de estratificación, esta opción permite acelerar el proceso de selección.

Asistente de muestreo: Extraer muestra, archivos de resultado

Este paso permite elegir dónde dirigir los casos muestreados, las variables de ponderación, las probabilidades conjuntas y las reglas de selección de casos.

Datos muestrales. Estas opciones permiten determinar dónde se escribe el resultado de la muestra. Se puede añadir a un conjunto de datos activo, escribir en un nuevo conjunto de datos o guardar en un archivo de datos con formato IBM® SPSS Statistics externo. Los conjuntos de datos están disponibles durante la sesión actual, pero no así en las sesiones posteriores, a menos que los haya guardado explícitamente como archivos de datos. El nombre de un conjunto de datos debe cumplir las normas de denominación de variables. Si se especifica un archivo externo o un nuevo conjunto de datos, se escribirán las variables de los resultados del muestreo y las variables del conjunto de datos activo para los casos seleccionados.

Probabilidades conjuntas. Estas opciones permiten determinar dónde se escriben las probabilidades conjuntas. Éstas se guardan en un archivo de datos con formato IBM SPSS Statistics externo. Las probabilidades conjuntas se producen si se seleccionan la probabilidad proporcional al tamaño sin reposición, el muestreo de Brewer proporcional al tamaño, el muestreo de Sampford proporcional al tamaño, o el muestreo de Murthy proporcional al tamaño y la estimación con reposición no se especifica.

Reglas de selección de casos. Si está construyendo la muestra por etapas, es posible que quiera guardar las reglas de selección de casos en un archivo de texto. Son útiles para construir el submarco de las etapas posteriores.

Asistente de muestreo: Finalizar

Este paso es el último. Puede guardar el archivo de plan y extraer la muestra ahora o pegar las selecciones en una ventana de sintaxis.

Al realizar cambios a las etapas del archivo de plan existente, puede guardar el plan editado en un archivo nuevo o sobrescribir el archivo existente. Al añadir etapas sin realizar cambios en las etapas existentes, el asistente sobrescribe de manera automática el archivo de planificación existente. Si desea guardar la planificación en un nuevo archivo, seleccione **Pegar la sintaxis generada por el asistente en una ventana de sintaxis** y cambie el nombre del archivo en los comandos de sintaxis.

Modificar un plan de muestreo existente

1. Seleccione en los menús:
Analizar > Muestras complejas > Seleccionar una muestra...
2. Seleccione **Editar un diseño muestral** y elegir un archivo de plan para editar.
3. Pulse **Siguiente** para continuar usando el Asistente.
4. Revise el plan de muestreo del paso Resumen del plan, y a continuación pulse **Siguiente**.
Los pasos posteriores son prácticamente iguales que los de un diseño nuevo. Si desea obtener más información sobre los pasos individuales, consulte la ayuda.
5. Vaya al paso final y especifique un nombre nuevo para el archivo de plan editado o sobrescriba el archivo de plan existente.

Si lo desea, puede:

- Especificar las etapas que ya se han muestreado.
- Eliminar etapas del plan.

Asistente de muestreo: Resumen del plan

Este paso permite revisar el plan de muestreo e indicar las etapas que ya se han muestreado. Al editar un plan, también puede eliminar etapas del plan.

Etapas muestreadas previamente. Si un marco de muestreo ampliado no está disponible, deberá ejecutar un diseño del muestreo polietápico etapa por etapa. Seleccione las etapas que ya se han muestreado en la lista desplegable. Las etapas que ya se hayan ejecutado estarán bloqueadas, por lo que no estarán disponibles en el paso Extraer muestra: Opciones de selección y no se podrán modificar al editar un plan.

Eliminar etapas. Puede eliminar las etapas 2 y 3 de un diseño polietápico.

Ejecutar un plan de muestreo existente

1. Seleccione en los menús:
Analizar > Muestras complejas > Seleccionar una muestra...
2. Seleccione **Extraer una muestra** y elija un archivo de plan para ejecutar.
3. Pulse **Siguiente** para continuar usando el Asistente.
4. Revise el plan de muestreo del paso Resumen del plan, y a continuación pulse **Siguiente**.
5. Cuando se ejecuta un plan de muestreo se omiten los pasos individuales que contienen información de la etapa. Ya puede pasar al paso de finalización.

Si lo desea, puede especificar las etapas que ya se han muestreado.

Características adicionales de los comandos CSPLAN y CSSELECT

La sintaxis de comandos también le permite:

- Especificar nombres personalizados para las variables de resultado.
- Controlar los resultados en el Visor. Por ejemplo, puede suprimir el resumen por etapas del plan que se muestra si se diseña o modifica una muestra, suprimir el resumen de la distribución de los casos muestreados por etapas que se muestra si el diseño muestral se ejecuta y solicitar un resumen del procesamiento de los casos.
- Elegir un subconjunto de las variables existentes en el conjunto de datos activo para escribirlo en un archivo muestral externo o en otro conjunto de datos.

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

Capítulo 3. Preparación de una muestra compleja para su análisis

El Asistente de preparación del análisis le guía a través de los pasos para crear o modificar un plan de análisis y utilizarlo con los distintos procedimientos de análisis de Muestras complejas. Antes de utilizar el Asistente, debe haber extraído la muestra de acuerdo con un diseño complejo.

Es más útil crear un plan nuevo cuando no se tiene acceso al archivo del plan de muestreo utilizado para extraer la muestra (recuerde que el plan de muestreo contiene un plan de análisis predeterminado). Si no tiene acceso al archivo del plan de muestreo utilizado para extraer la muestra, puede utilizar el plan de análisis contenido predeterminado en el archivo del plan de muestreo u omitir las especificaciones del análisis predeterminado y guardar los cambios en un archivo nuevo.

Creación de un nuevo plan de análisis

1. Seleccione en los menús:
Analizar > Muestras complejas > Preparar para el análisis...
2. Seleccione **Crear un archivo de plan**, y elija un nombre de archivo de plan para guardar el plan del análisis.
3. Pulse **Siguiente** para continuar usando el Asistente.
4. Especifique la variable que contiene las ponderaciones muestrales en el paso Variables del diseño, si lo desea puede definir estratos y clústeres.
5. Ahora puede pulsar **Finalizar** para guardar el plan.

Si lo desea, en los siguientes pasos puede:

- Seleccionar el método de estimación de los errores estándar en el paso Método de estimación.
- Especificar el número de unidades muestrales o la probabilidad de inclusión por unidad en el paso Tamaño.
- Añadir una segunda o tercera etapa al diseño.
- Pegar las selecciones como sintaxis de comandos.

Asistente de preparación del análisis: Variables del diseño

Este paso permite identificar las variables de estratificación y agrupación en clústeres y definir las ponderaciones muestrales. También puede proporcionar una etiqueta para la etapa.

Estratos. La clasificación cruzada por las variables de estratificación define distintas subpoblaciones o estratos. El total muestral representa la combinación de las muestras independientes pertenecientes a cada estrato.

Clústeres. Las variables de clúster definen grupos de unidades de observación o clústeres. Las muestras extraídas en varias etapas seleccionan clústeres en las etapas anteriores y, a continuación, unidades de submuestra dentro de los clústeres seleccionados. Al analizar un archivo de datos obtenido mediante el muestreo de clústeres con reposición, debe incluir el índice de duplicación como una variable de clúster.

Ponderación muestral. Debe proporcionar ponderaciones muestrales en la primera etapa. Las ponderaciones muestrales se calculan automáticamente para las etapas posteriores del diseño actual.

Etiqueta de etapa. Puede especificar una etiqueta de cadena opcional para cada etapa. Esto se utiliza en los resultados para facilitar la identificación de la información por etapas.

Nota: la lista de variables origen tiene el mismo contenido a lo largo de los pasos del Asistente. En otras palabras, las variables de la lista de origen eliminadas en un paso determinado se borran de la lista en todos los pasos. Las variables devueltas a la lista de origen aparecen en todos los pasos.

Controles de árbol para desplazarse por el Asistente para el análisis

En la parte izquierda de cada paso del Asistente para el análisis se muestra un esquema con los titulares de todos los pasos. Puede navegar por el Asistente al pulsar el nombre de uno de los pasos activados en los titulares. Los pasos están activados mientras todos los pasos anteriores sean válidos, es decir, mientras cada uno de los pasos anteriores tenga las especificaciones mínimas necesarias para ese paso. Consulte la ayuda de los pasos individuales para obtener más información sobre los motivos por los que un paso dado puede no ser válido.

Asistente de preparación del análisis: Método de estimación

Este paso permite especificar un método de estimación para la etapa.

CR (muestreo con reposición). La estimación CR no incluye una corrección de muestreo para poblaciones finitas (FPC) al estimar la varianza bajo el diseño del muestreo complejo. Puede incluir o excluir la FPC al estimar la varianza bajo muestreo aleatorio simple (SRS).

Se recomienda no incluir la FPC para la estimación de varianza SRS cuando las ponderaciones de análisis se hayan escalado de forma que no se agreguen al tamaño de la población. La estimación de varianza SRS se utiliza para calcular estadísticos como el efecto de diseño. La estimación CR sólo se puede especificar en la etapa final de un diseño; el Asistente no permitirá añadir otra etapa si se selecciona la estimación CR.

Igual SR (muestreo de igual probabilidad sin reposición). La estimación Igual SR incluye la corrección para poblaciones finitas y supone que las unidades se muestrearon con la misma probabilidad. El método Igual SR se puede especificar en cualquiera de las etapas de un diseño.

Desigual SR (muestreo de probabilidad desigual sin reposición). Además de utilizar la corrección para poblaciones finitas, el método Desigual SR tiene en cuenta las unidades muestrales (normalmente clústeres) que han sido seleccionadas con probabilidades desiguales. Este método de estimación sólo está disponible en la primera etapa.

Asistente de preparación del análisis: Tamaño

Este paso se utiliza para especificar las probabilidades de inclusión o los tamaños poblacionales para la etapa actual. Los tamaños pueden ser fijos o variar entre estratos. Para especificar los tamaños, los clústeres especificados en las etapas anteriores se pueden utilizar para definir estratos. Tenga en cuenta que este paso sólo es necesario cuando se elige el método Igual SR como método de estimación.

Unidades. Puede especificar los tamaños poblacionales exactos o las probabilidades con las que se ha realizado el muestreo de las unidades.

- **Valor.** Se aplica un valor particular a todos los estratos. Si se selecciona **Tamaños poblacionales** como la unidad métrica, se deberá introducir un entero no negativo. Si se selecciona **Probabilidades de inclusión**, se deberá introducir un valor entre 0 y 1, ambos incluidos.
- **Valores desiguales para estratos.** Permite introducir distintos valores de tamaño para cada estrato a través del cuadro de diálogo Definir tamaños desiguales.
- **Leer valores de la variable.** Permite seleccionar una variable numérica que contenga los valores de tamaño para los estratos.

Definir tamaños desiguales

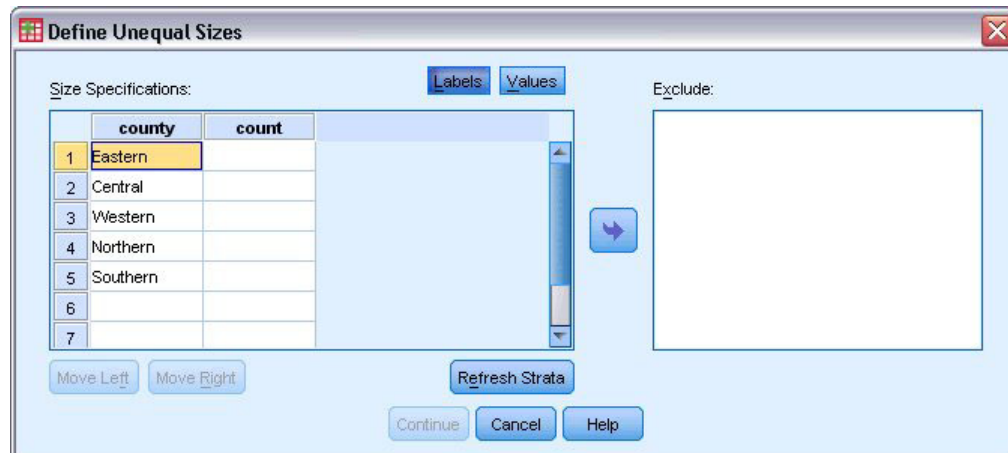


Figura 2. Cuadro de diálogo Definir tamaños desiguales

El cuadro de diálogo Definir tamaños desiguales permite introducir los tamaños para cada estrato.

Cuadrícula de especificaciones de tamaño. La cuadrícula muestra la clasificación cruzada de hasta cinco variables de clúster o estrato, con una combinación de estrato/clúster por fila. Las variables elegibles en la cuadrícula serán todas las variables de estratificación de las etapas anteriores y actuales además de todas las variables de clúster de las etapas anteriores. Las variables se pueden reordenar dentro de la cuadrícula o ser desplazadas a la lista Excluir. Introduzca los tamaños en la última columna de la derecha. Pulse en **Etiquetas** o **Valores** para conmutar entre la visualización de las etiquetas de valor y los valores de los datos para las variables de estratificación y de clúster de las casillas de la cuadrícula. Las casillas que contienen valores sin etiquetas siempre muestran valores. Pulse **Actualizar estratos** para volver a rellenar la cuadrícula con cada combinación de los valores de los datos etiquetados para las variables de la cuadrícula.

Excluir. Para especificar los tamaños de un subconjunto de combinaciones de estrato/clúster, desplace una o más variables a la lista Excluir. Estas variables no se utilizan para definir tamaños muestrales.

Asistente de preparación del análisis: Resumen del plan

Este paso es el último de cada etapa y proporciona un resumen de las especificaciones del diseño del análisis hasta la etapa actual, ésta incluida. A partir de aquí, puede pasar a la siguiente etapa (creándola si fuera necesario) o guardar las especificaciones del análisis.

Si no puede añadir otra etapa, esto puede deberse a:

- No se especificó ninguna variable de clúster en el paso Variables del diseño.
- Seleccionó la estimación CR en el paso Método de estimación.
- Este paso es el tercero del análisis; el Asistente admite un máximo de tres etapas.

Asistente de preparación del análisis: Finalizar

Este paso es el último. Puede guardar el archivo del plan ahora o pegar las selecciones en una ventana de sintaxis.

Al realizar cambios a las etapas del archivo de plan existente, puede guardar el plan editado en un archivo nuevo o sobrescribir el archivo existente. Al añadir etapas sin realizar cambios en las etapas existentes, el asistente sobrescribe de manera automática el archivo de planificación existente. Si desea

guardar la planificación en un nuevo archivo, elija **Pegar la sintaxis generada por el asistente en una ventana de sintaxis** y cambie el nombre del archivo en los comandos de sintaxis.

Modificar un plan de análisis existente

1. Seleccione en los menús:
Analizar > Muestras complejas > Preparar para el análisis...
2. Seleccione **Editar un archivo de plan** y elija un nombre de archivo de plan en el que se guardará el plan del análisis.
3. Pulse **Siguiente** para continuar usando el Asistente.
4. Revise el plan de análisis en el paso Resumen del plan y, a continuación, pulse **Siguiente**.
Los pasos posteriores son prácticamente iguales que los de un diseño nuevo. Si desea obtener más información, consulte la ayuda sobre los pasos individuales.
5. Vaya al paso final y especifique un nombre nuevo para el archivo de plan editado o sobrescriba el archivo de plan existente.

Si lo desea, puede eliminar algunas etapas del plan.

Asistente de preparación del análisis: Resumen del plan

Este paso permite revisar el plan de análisis y eliminar etapas del plan.

Eliminar etapas. Puede eliminar las etapas 2 y 3 de un diseño polietápico. Debido a que los planes deben tener al menos una etapa, puede editar la etapa 1 pero no eliminarla del diseño.

Capítulo 4. Plan de muestras complejas

Los procedimientos de análisis de Muestras complejas requieren las especificaciones de análisis de un archivo de plan de muestreo o un plan de análisis para poder proporcionar resultados válidos.

Plan. Especifique la ruta de un archivo de plan de muestreo o análisis.

Probabilidades conjuntas. Para utilizar una estimación Desigual SR para los clústeres extraídos utilizando un método PPS SR, debe especificar un archivo independiente o un conjunto de datos abierto que contenga las probabilidades conjuntas. El archivo o conjunto de datos se crea mediante el Asistente de muestreo durante el muestreo.

Capítulo 5. Frecuencias de muestras complejas

El procedimiento Frecuencias de Muestras complejas genera tablas de frecuencias para las variables seleccionadas y muestra estadísticos univariantes. Si lo desea, puede solicitar estadísticos por subgrupos, definidos por una o más variables categóricas.

Ejemplo. Mediante el procedimiento Frecuencias de Muestras complejas, puede obtener estadísticos tabulares univariantes para el consumo de vitaminas entre los ciudadanos de EE.UU., basados en los resultados del National Health Interview Survey (NHIS, Centro Nacional de Estadísticas de Salud) y con un plan de análisis adecuado para estos datos de uso público.

Estadísticos. El procedimiento genera estimaciones de los tamaños poblacionales de las casillas, además de errores estándar, intervalos de confianza, coeficientes de variación, efectos de diseño, raíz cuadrada de los efectos de diseño, valores acumulados y recuentos no ponderados para cada estimación. Además, se calculan los estadísticos de chi-cuadrado y la razón de verosimilitud para el contraste de proporciones de casilla iguales.

Frecuencias de Muestras complejas: Consideraciones sobre los datos

Datos. Variables para las que se generan las tablas de frecuencias deben ser categóricas. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de Frecuencias de Muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Frecuencias...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Frecuencias de muestras complejas, seleccione al menos una variable de frecuencia.

Si lo desea, puede especificar variables para definir subpoblaciones. Los estadísticos se calculan por separado para cada subpoblación.

Frecuencias de Muestras complejas: Estadísticos

Casillas. Este grupo permite solicitar estimaciones de los tamaños poblacionales de las casillas así como porcentajes de tabla.

Estadísticos. Este grupo genera estadísticos asociados con el tamaño poblacional o los porcentajes de tabla.

- **Error estándar.** El error estándar de la estimación.
- **Intervalo de confianza.** Intervalo de confianza para la estimación, utilizando el nivel especificado.
- **Coficiente de variación.** Cociente del error estándar de la estimación dividida por la estimación.
- **Recuento no ponderado.** Número de unidades utilizadas para calcular la estimación.

- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Valores acumulados.** La estimación acumulada a través de los valores de la variable.

Contraste sobre proporciones de casilla iguales. Esto genera los contrastes de chi-cuadrado y la razón de verosimilitud sobre la hipótesis de que las categorías de una variable tienen la misma frecuencia. Se realizan contrastes por separado para cada variable.

Muestras complejas: Valores perdidos

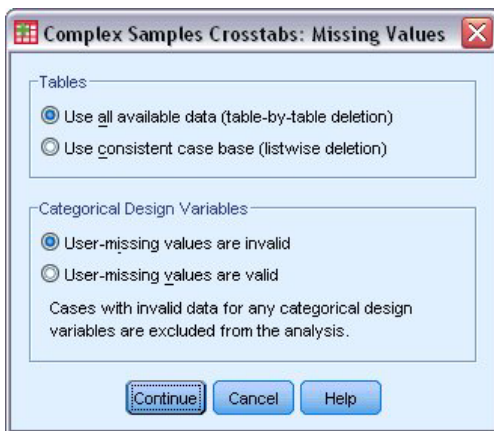


Figura 3. Cuadro de diálogo Valores perdidos

Tablas. Este grupo determina los casos que se utilizan en el análisis.

- **Utilizar todos los datos disponibles.** Los valores perdidos se determinan en base a tabla por tabla. Así, los casos utilizados para calcular los estadísticos pueden variar a través de la frecuencia o tablas de tabulación cruzada.
- **Utilizar una base coherente para los casos.** Los valores perdidos se determinan a través de todas las variables. Por lo tanto, los casos utilizados para calcular los estadísticos son coherentes con las tablas.

Variables categóricas del diseño. Este grupo determina si los valores perdidos del usuario son considerados válidos o inválidos.

Opciones de Muestras complejas

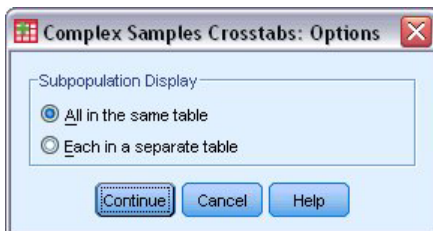


Figura 4. Cuadro de diálogo Opciones

Mostrar subpoblación. Puede elegir entre mostrar las subpoblaciones en la misma tabla o en tablas separadas.

Capítulo 6. Descriptivos de muestras complejas

El procedimiento Descriptivos de Muestras complejas muestra estadísticos de resumen univariantes para distintas variables. Si lo desea, puede solicitar estadísticos por subgrupos, definidos por una o más variables categóricas.

Ejemplo. Mediante el procedimiento Descriptivos de Muestras complejas, puede obtener estadísticos descriptivos univariantes de los niveles de actividad de los ciudadanos de EE.UU., basados en los resultados de la National Health Interview Survey (NHIS, Centro Nacional de Estadísticas de Salud) y con un plan de análisis adecuado para estos datos de uso público.

Estadísticos. El procedimiento genera medias y sumas, además de pruebas t , errores estándar, intervalos de confianza, coeficientes de variación, recuentos no ponderados, efectos de diseño y la raíz cuadrada del efecto de diseño de cada estimación.

Consideraciones sobre los datos de Descriptivos de Muestras complejas

Datos. Las medidas deben ser variables de escala. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de Descriptivos de Muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Descriptivos...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Descriptivos de muestras complejas, seleccione al menos una variable de medición.

Si lo desea, puede especificar variables para definir subpoblaciones. Los estadísticos se calculan por separado para cada subpoblación.

Descriptivos de Muestras complejas: Estadísticos

Resúmenes. Este grupo permite solicitar estimaciones de las medias y sumas de las variables de medida. Además, puede solicitar pruebas t de las estimaciones con respecto a un valor especificado.

Estadísticos. Este grupo genera estadísticos asociados con la media o la suma.

- **Error estándar.** El error estándar de la estimación.
- **Intervalo de confianza.** Intervalo de confianza para la estimación, utilizando el nivel especificado.
- **Coficiente de variación.** Cociente del error estándar de la estimación dividida por la estimación.
- **Recuento no ponderado.** Número de unidades utilizadas para calcular la estimación.
- **Tamaño poblacional.** Número estimado de unidades en la población.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

Valores perdidos en los descriptivos de Muestras complejas

Estadísticos para variables de medida. Este grupo determina los casos que se utilizan en el análisis.

- **Utilizar todos los datos disponibles.** Los valores perdidos se determinan variable por variable; por ello los casos utilizados para calcular los estadísticos pueden variar entre las variables de medida.
- **Asegurar una base coherente para los casos.** Los valores perdidos se determinan a partir de todas las variables, así, los casos utilizados para calcular los estadísticos son coherentes.

Variables categóricas del diseño. Este grupo determina si los valores perdidos del usuario son considerados válidos o inválidos.

Opciones de Muestras complejas

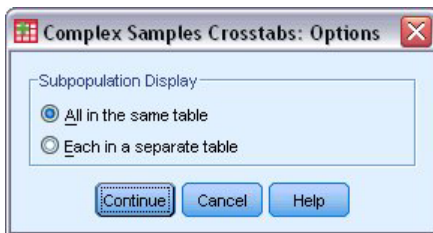


Figura 5. Cuadro de diálogo Opciones

Mostrar subpoblación. Puede elegir entre mostrar las subpoblaciones en la misma tabla o en tablas separadas.

Capítulo 7. Tablas cruzadas de muestras complejas

El procedimiento Tablas cruzadas de Muestras complejas genera tablas de tabulación cruzada para los pares de variables seleccionadas y muestra estadísticos bidimensionales. Si lo desea, puede solicitar estadísticos por subgrupos, definidos por una o más variables categóricas.

Ejemplo. Mediante el procedimiento Tablas cruzadas de Muestras complejas, se pueden obtener estadísticos de clasificaciones cruzadas de la frecuencia de consumo de tabaco por el consumo de vitaminas en los ciudadanos de EE.UU, basado en los resultados del National Health Interview Survey (NHIS, Centro Nacional de Estadísticas de Salud) y con un plan de análisis adecuado para estos datos de uso público.

Estadísticos. El procedimiento genera estimaciones de los tamaños poblacionales de las casillas, así como porcentajes de tabla, columna y fila, además de errores estándar, intervalos de confianza, coeficientes de variación, valores esperados, efectos de diseño, raíz cuadrada de los efectos de diseño, residuos, residuos corregidos y frecuencias no ponderadas para cada estimación. Para las tablas 2 por 2, se calculan la razón de las ventajas, el riesgo relativo y la diferencia de riesgos. Además, para el contraste de independencia de las variables de las filas y las variables de las columnas, se calculan los estadísticos de Pearson y de la razón de verosimilitud.

Tablas cruzadas de Muestras complejas: Consideraciones sobre los datos

Datos. Las variables de fila y columna deben ser categóricas. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de Tablas cruzadas de Muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Tablas cruzadas...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Tablas cruzadas de muestras complejas, seleccione al menos una variable de fila y una variable de columna.

Si lo desea, puede especificar variables para definir subpoblaciones. Los estadísticos se calculan por separado para cada subpoblación.

Tablas cruzadas de Muestras complejas: Estadísticos

Casillas. Este grupo permite solicitar estimaciones del tamaño poblacional de las casillas así como porcentajes de columna, fila y de tabla.

Estadísticos. Este grupo genera estadísticos asociados con el tamaño de la población y los porcentajes de tabla, columna y fila.

- **Error estándar.** El error estándar de la estimación.
- **Intervalo de confianza.** Intervalo de confianza para la estimación, utilizando el nivel especificado.
- **Coficiente de variación.** Cociente del error estándar de la estimación dividida por la estimación.

- **Valores esperados.** Valor esperado de la estimación, bajo la hipótesis de independencia de las variables de fila y columna.
- **Recuento no ponderado.** Número de unidades utilizadas para calcular la estimación.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Residuos.** El valor esperado es el número de casos que se esperaría encontrar en la casilla si no hubiera relación entre las dos variables. Un residuo positivo indica que hay más casos en la casilla de los que habría en ella si las variables de fila y columna fueran independientes.
- **Residuos corregidos.** El residuo de una casilla (el valor observado menos el valor esperado) dividido por una estimación de su error estándar. El residuo estandarizado resultante viene expresado en unidades de desviación estándar, por encima o por debajo de la media.

Resúmenes para las tablas 2 por 2. Este grupo genera estadísticos para las tablas en las que la variable de fila y la de columna tienen dos categorías. Cada una es una medida de la fuerza de la asociación entre la presencia de un factor y la aparición de un evento.

- **Razón de las ventajas.** Cuando la ocurrencia del factor es poco común, se puede utilizar la razón de las ventajas como estimación del riesgo relativo.
- **Riesgo relativo.** La razón del riesgo de un evento en presencia del factor respecto al riesgo del evento en ausencia del factor.
- **Diferencia de riesgos.** La diferencia entre el riesgo de un evento en presencia del factor y el riesgo del evento en ausencia del factor.

Contraste sobre la independencia de filas y columnas. Esta opción genera los contrastes de chi-cuadrado y la razón de verosimilitud sobre la hipótesis de que las variables de fila y columna son independientes. Se realizan contrastes por separado para cada pareja de variables.

Muestras complejas: Valores perdidos

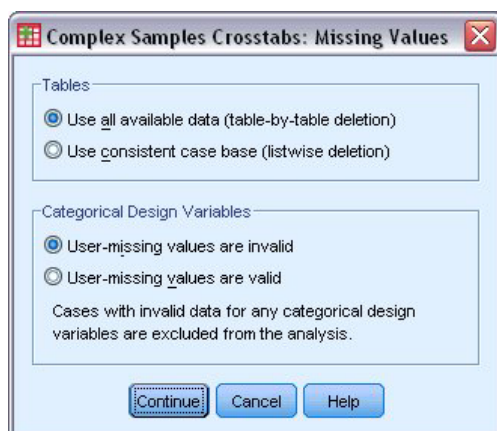


Figura 6. Cuadro de diálogo Valores perdidos

Tablas. Este grupo determina los casos que se utilizan en el análisis.

- **Utilizar todos los datos disponibles.** Los valores perdidos se determinan en base a tabla por tabla. Así, los casos utilizados para calcular los estadísticos pueden variar a través de la frecuencia o tablas de tabulación cruzada.

- **Utilizar una base coherente para los casos.** Los valores perdidos se determinan a través de todas las variables. Por lo tanto, los casos utilizados para calcular los estadísticos son coherentes con las tablas.

VARIABLES CATEGÓRICAS DEL DISEÑO. Este grupo determina si los valores perdidos del usuario son considerados válidos o inválidos.

Opciones de Muestras complejas

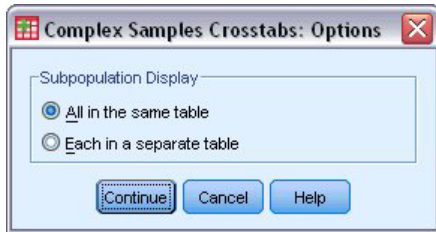


Figura 7. Cuadro de diálogo Opciones

Mostrar subpoblación. Puede elegir entre mostrar las subpoblaciones en la misma tabla o en tablas separadas.

Capítulo 8. Razones de muestras complejas

El procedimiento Razones de Muestras complejas muestra estadísticos de resumen univariantes para razones de variables. Si lo desea, puede solicitar estadísticos por subgrupos, definidos por una o más variables categóricas.

Ejemplo. Mediante el procedimiento Razones de Muestras complejas, puede obtener estadísticos descriptivos para el cociente del valor de la propiedad actual sobre el último valor certificado, basado en los resultados de una encuesta a nivel estatal llevada a cabo según un diseño complejo y con un plan de análisis adecuado para los datos.

Estadísticos. El procedimiento genera estimaciones de razón, pruebas t , errores estándar, intervalos de confianza, coeficientes de variación, recuentos no ponderados, tamaños poblacionales, efectos de diseño y raíz cuadrada del efecto de diseño.

Consideraciones sobre los datos en las Razones de Muestras complejas

Datos. Los numeradores y los denominadores deben ser variables de escala con valores positivos. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de razones de Muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Razones...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Razones de muestras complejas, seleccione al menos una variable de numerador y una variable de denominador.

Si lo desea, puede especificar variables para definir subgrupos para los que se desea generar estadísticos.

Razones de Muestras complejas. Estadísticos

Estadísticos. Este grupo genera estadísticos asociados con la estimación de la razón.

- **Error estándar.** El error estándar de la estimación.
- **Intervalo de confianza.** Intervalo de confianza para la estimación, utilizando el nivel especificado.
- **Coficiente de variación.** Cociente del error estándar de la estimación dividida por la estimación.
- **Recuento no ponderado.** Número de unidades utilizadas para calcular la estimación.
- **Tamaño poblacional.** Número estimado de unidades en la población.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

Prueba t . Puede solicitar pruebas t de las estimaciones con respecto a un valor especificado.

Razones de Muestras complejas: Valores perdidos

Razones. Este grupo determina los casos que se utilizan en el análisis.

- **Utilizar todos los datos disponibles.** Los valores perdidos se determinan en base a razón por razón. Así, los casos utilizados para calcular los estadísticos pueden variar a través de pares numerador-denominador.
- **Asegurar una base coherente para los casos.** Los valores perdidos se determinan a través de todas las variables. Por lo tanto, los casos utilizados para calcular los estadísticos son coherentes con las tablas.

Variables categóricas del diseño. Este grupo determina si los valores perdidos del usuario son considerados válidos o inválidos.

Opciones de Muestras complejas

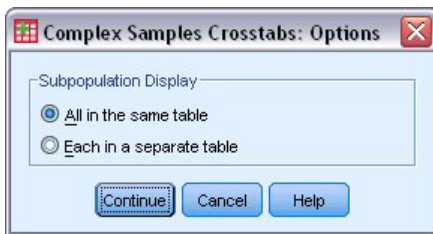


Figura 8. Cuadro de diálogo Opciones

Mostrar subpoblación. Puede elegir entre mostrar las subpoblaciones en la misma tabla o en tablas separadas.

Capítulo 9. Modelo lineal general de muestras complejas

El procedimiento Modelo lineal general de muestras complejas (CSGLM) realiza análisis de regresión lineal y análisis de varianza y covarianza de muestras extraídas mediante métodos de muestreo complejo. Si lo desea, puede solicitar análisis de una subpoblación.

Ejemplo. Una cadena de tiendas de alimentos realiza una encuesta sobre los hábitos de compra de una serie de clientes basándose en un diseño complejo. Una vez obtenidos los resultados de la encuesta y la cantidad que cada cliente gastó el mes anterior, la cadena desea averiguar si la frecuencia con que los clientes hacen la compra está relacionada con la cantidad mensual que gastan, controlando el sexo del cliente e incorporando el diseño del muestreo.

Estadísticos. El procedimiento genera estimaciones, errores estándar, pruebas *t*, efectos de diseño, raíz cuadrada de los efectos de diseño para parámetros de modelo y las correlaciones y covarianzas entre las estimaciones de los parámetros. Las medidas de ajuste del modelo y los estadísticos descriptivos de las variables dependientes e independientes también están disponibles. Además, se pueden solicitar medias marginales estimadas para los niveles de factores de modelado u las interacciones de los factores.

Consideraciones sobre los datos del Modelo lineal general de muestras complejas

Datos. La variable dependiente es cuantitativa. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Para obtener un Modelo lineal general de muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Modelo lineal general...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Modelo lineal general de muestras complejas, seleccione una variable dependiente.

Si lo desea, puede:

- Seleccione variables para factores y covariables, según corresponda a los datos.
- Especifique una variable para definir una subpoblación. El análisis se lleva a cabo únicamente en la categoría seleccionada de la variable de subpoblación.

Modelo lineal general de muestras complejas

Especificar efectos del modelo. De forma predeterminada, el procedimiento crea un modelo de efectos principales utilizando los factores y las covariables especificadas en el cuadro de diálogo principal. Si lo desea, también puede crear un modelo personalizado que contenga los efectos de la interacción y los términos anidados.

Términos no anidados

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel para todas las variables seleccionadas.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4. Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quintuples posibles de las variables seleccionadas.

Términos anidados

En este procedimiento, puede generar términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

Limitaciones. Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A(A)$.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si A es un factor y X es una covariable, no es válido especificar $A(X)$.

Intersección. La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección. Incluso aunque incluya la intersección en el modelo, puede suprimir los estadísticos relacionados con ella.

Estadísticos de Modelo lineal general de muestras complejas

Parámetros del modelo. Este grupo permite controlar la presentación de estadísticos relacionados con los parámetros del modelo.

- **Estimación.** Muestra estimaciones de los coeficientes.
- **Error estándar.** Muestra el error estándar de cada estimación de los coeficientes.
- **Intervalo de confianza.** Muestra un intervalo de confianza para cada estimación de los coeficientes. El nivel de confianza de los intervalos se configura en el cuadro de diálogo Opciones.
- **Prueba t.** Muestra una prueba t de cada estimación de coeficientes. La hipótesis nula de cada prueba es que el valor del coeficiente sea 0.
- **Covarianzas de las estimaciones de los parámetros.** Muestra una estimación de la matriz de covarianzas de los coeficientes del modelo.
- **Correlaciones de las estimaciones de los parámetros.** Muestra una estimación de la matriz de correlaciones de los coeficientes del modelo.

- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

Ajuste del modelo. Muestra R^2 y estadísticas de la raíz de la media cuadrática de los errores.

Medias de población de covariables y variables dependientes. Muestra información resumida acerca de los factores, las covariables y las variables dependientes.

Información del diseño muestral. Muestra información resumida acerca de la muestra, incluidos un recuento no ponderado y el tamaño de la población.

Muestras complejas: Contrastes de hipótesis

Estadístico de contraste. Este grupo le permite seleccionar el tipo de estadístico utilizado para contrastar las hipótesis. Es posible elegir entre F , F corregida, chi-cuadrado y chi-cuadrado corregido.

Muestreo de grados de libertad. Este grupo permite controlar los grados de libertad en el diseño del muestreo usados para calcular los valores p de todos los estadísticos de prueba. Si se basa en el diseño del muestreo, el valor es la diferencia entre el número de unidades de muestra primarias y el número de estratos de la primera etapa del muestreo. Si lo desea, puede especificar los grados de libertad que desee introduciendo un número entero positivo.

Corrección para comparaciones múltiples. Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- *Sidak secuencial.* Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Bonferroni secuencial.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Sidak.* Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- *Bonferroni.* Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.

Medias estimadas del Modelo lineal general de muestras complejas

En el cuadro de diálogo Medias estimadas se pueden ver las medias marginales estimadas por el modelo para los niveles de factores y las interacciones de factores especificadas en el subcuadro de diálogo Modelo. También se puede solicitar que se muestre la media de población global.

Término. Se calculan las medias estimadas de los factores seleccionados y las interacciones de los factores.

Contraste. El contraste determina como se configuran los contrastes de hipótesis para comparar las medias estimadas.

- *Simple.* Compara la media de cada nivel con la media de un nivel especificado. Este tipo de contraste resulta útil cuando existe un grupo de control.

- *Desviación.* Compara la media de cada nivel (excepto una categoría de referencia) con la media de todos los niveles (media global). Los niveles del factor pueden colocarse en cualquier orden.
- *Diferencia.* Compara la media de cada nivel (excepto el primero) con la media de los niveles anteriores (a veces también se denominan contrastes de Helmert inversos). En ocasiones se les denomina contrastes de Helmert invertidos.
- *Helmert.* Compara la media de cada nivel del factor (excepto el último) con la media de los niveles siguientes.
- *Repetido.* Compara la media de cada nivel (excepto el último) con la media del nivel siguiente.
- *Polinómico.* Compara el efecto lineal, cuadrático, cúbico, etc. El primer grado de libertad contiene el efecto lineal a través de todas las categorías; el segundo grado de libertad, el efecto cuadrático, y así sucesivamente. Estos contrastes se utilizan a menudo para estimar las tendencias polinómicas.

Categoría de referencia. Los contrastes simple y de desviación requieren una categoría de referencia o un factor de nivel con que comparar los demás.

Modelo lineal general de muestras complejas: Guardar

Guardar variables. Este grupo permite guardar los valores pronosticados para el modelo y los residuos como nuevas variables en el archivo de trabajo.

Exportar modelo como datos de IBM SPSS Statistics. Escribe un conjunto de datos de formato IBM SPSS Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores estándar, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **rowtype_.** Toma los valores (y las etiquetas de valor), COV (covarianzas), CORR (correlaciones), EST (estimaciones de los parámetros), SE (errores estándar), SIG (niveles de significación) y DF (grados de libertad del diseño muestral). Hay un caso diferente con el tipo de fila COV (o CORR) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.
- **varname_.** Toma los valores P1, P2, ..., correspondientes a una lista ordenada de todos los parámetros del modelo para los tipos de fila COV o CORR, con las etiquetas de valor correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.
- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila. Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores estándar, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

Nota: este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan.

Exportar modelo como XML. Guarda las estimaciones de los parámetros y la matriz de covarianzas de los parámetros (si se selecciona) en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Modelo lineal general de muestras complejas: Opciones

Valores perdidos del usuario. Todas las variables de diseño, así como la variable dependiente y cualquier covariable, deben contener datos válidos. Los casos con datos no válidos de cualquiera de estas variables se excluyen del análisis. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre las variables de estratificación, de clúster, de subpoblación y de factor.

Intervalo de confianza. Se trata del nivel de intervalo de confianza para las estimaciones de coeficiente y las medias marginales estimadas. Especifique un valor mayor o igual a 50 e inferior a 100.

Características adicionales del comando CSGLM

La sintaxis de comandos también le permite:

- Especificar contrastes personalizados de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando CUSTOM).
- Fijar covariables en valores distintos los de sus medias al calcular las medias marginales estimadas (utilizando el subcomando EMMEANS).
- Especificar una métrica para los contrastes polinómicos (utilizando el subcomando EMMEANS).
- Especificar un valor de tolerancia para la comprobación de la singularidad (utilizando el subcomando CRITERIA).
- Crear nombres especificados por el usuario para las variables almacenadas (utilizando el subcomando SAVE).
- Generar una tabla de función estimable general (utilizando el subcomando PRINT).

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

Capítulo 10. Regresión logística de muestras complejas

El procedimiento Regresión logística de muestras complejas lleva a cabo análisis de regresión logística sobre una variable binaria o una variable dependiente multinomial para muestras extraídas mediante métodos de muestreo complejo. Si lo desea, puede solicitar análisis de una subpoblación.

Ejemplo. Un encargado de préstamos ha recopilado registros antiguos de préstamos concedidos a clientes en diversas ramas, de acuerdo con un diseño complejo. Al incorporar el diseño muestral, el encargado desea comprobar si la probabilidad con que las moras de un cliente se asocian a su edad, historial de empleo y cantidad de crédito adeudado; posteriormente.

Estadísticos. El procedimiento genera estimaciones, estimaciones exponenciadas, errores estándar, intervalos de confianza, pruebas t , efectos de diseño, raíz cuadrada de los efectos de diseño para parámetros de modelo y las correlaciones y covarianzas entre las estimaciones de los parámetros. También hay disponibles estadísticos pseudo R^2 , tablas de clasificación y estadísticos descriptivos para las variables dependientes e independientes.

Regresión logística de muestras complejas: Consideraciones sobre los datos

Datos. La variable dependiente es categórica. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de Regresión logística de muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Regresión logística...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Regresión logística de muestras complejas, seleccione una variable dependiente.

Si lo desea, puede:

- Seleccione variables para factores y covariables, según corresponda a los datos.
- Especifique una variable para definir una subpoblación. El análisis se lleva a cabo únicamente en la categoría seleccionada de la variable de subpoblación.

Regresión logística de muestras complejas: Categoría de referencia

De forma predeterminada, el procedimiento Regresión logística de muestras complejas hace de la categoría con el valor más alto la categoría de referencia. Este cuadro de diálogo permite especificar el valor más alto, el valor más bajo o una categoría personalizada como la categoría de referencia.

Regresión logística de muestras complejas: Modelo

Especificar efectos del modelo. De forma predeterminada, el procedimiento crea un modelo de efectos principales utilizando los factores y las covariables especificadas en el cuadro de diálogo principal. Si lo desea, también puede crear un modelo personalizado que contenga los efectos de la interacción y los términos anidados.

Términos no anidados

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel para todas las variables seleccionadas.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4. Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quintuples posibles de las variables seleccionadas.

Términos anidados

En este procedimiento, puede generar términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

Limitaciones. Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar A^*A .
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A(A)$.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si A es un factor y X es una covariable, no es válido especificar $A(X)$.

Intersección. La intersección se incluye normalmente en el modelo. Si asume que los datos pasan por el origen, puede excluir la intersección. Incluso aunque incluya la intersección en el modelo, puede suprimir los estadísticos relacionados con ella.

Regresión logística de muestras complejas: Estadísticos

Ajuste del modelo. Controla la presentación de estadísticos que miden el rendimiento global del proceso.

- **Pseudo R cuadrado.** El estadístico R^2 de regresión lineal no cuenta con un análogo exacto entre los modelos de regresión logística. En su lugar existen varias medidas que tratan de imitar las propiedades del estadístico R^2 .
- **Tabla de clasificación.** Muestra las clasificaciones cruzadas tabuladas de la categoría observada por la categoría pronosticada por el modelo en la variable dependiente.

Parámetros. Este grupo permite controlar la presentación de estadísticos relacionados con los parámetros del modelo.

- **Estimación.** Muestra estimaciones de los coeficientes.
- **Estimación exponenciada.** Muestra la base del logaritmo natural elevada a la potencia de las estimaciones de los coeficientes. Mientras que las estimaciones tienen propiedades agradables para la comprobación estadística, la estimación exponenciada (o $\exp[B]$) es más sencilla de interpretar.
- **Error estándar.** Muestra el error estándar de cada estimación de los coeficientes.
- **Intervalo de confianza.** Muestra un intervalo de confianza para cada estimación de los coeficientes. El nivel de confianza de los intervalos se configura en el cuadro de diálogo Opciones.
- **Prueba t.** Muestra una prueba t de cada estimación de coeficientes. La hipótesis nula de cada prueba es que el valor del coeficiente sea 0.
- **Covarianzas de las estimaciones de los parámetros.** Muestra una estimación de la matriz de covarianzas de los coeficientes del modelo.
- **Correlaciones de las estimaciones de los parámetros.** Muestra una estimación de la matriz de correlaciones de los coeficientes del modelo.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

Estadísticos de resumen para las variables del modelo. Muestra información resumida acerca de los factores, las covariables y las variables dependientes.

Información del diseño muestral. Muestra información resumida acerca de la muestra, incluidos un recuento no ponderado y el tamaño de la población.

Muestras complejas: Contrastes de hipótesis

Estadístico de contraste. Este grupo le permite seleccionar el tipo de estadístico utilizado para contrastar las hipótesis. Es posible elegir entre F , F corregida, chi-cuadrado y chi-cuadrado corregido.

Muestreo de grados de libertad. Este grupo permite controlar los grados de libertad en el diseño del muestreo usados para calcular los valores p de todos los estadísticos de prueba. Si se basa en el diseño del muestreo, el valor es la diferencia entre el número de unidades de muestra primarias y el número de estratos de la primera etapa del muestreo. Si lo desea, puede especificar los grados de libertad que desee introduciendo un número entero positivo.

Corrección para comparaciones múltiples. Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- *Sidak secuencial.* Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Bonferroni secuencial.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Sidak.* Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- *Bonferroni.* Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.

Regresión logística de muestras complejas: Razones de las ventajas

El cuadro de diálogo Razones de las ventajas permite mostrar las razones de las ventajas estimadas por el modelo para los factores y las covariables que se especifican. Se calcula un conjunto independiente de razones de las ventajas para cada categoría de la variable dependiente excepto para el caso de la categoría de referencia.

Factores. En cada factor seleccionado, muestra la razón de las probabilidades de cada categoría del factor hasta la probabilidad en la categoría de referencia especificada.

Covariables. En cada covariable seleccionada, muestra la razón de las probabilidades en el valor medio de la covariable más las unidades de cambio especificadas para la probabilidad de la media.

Al calcular las razones de las ventajas de un factor o una covariable, el procedimiento fija todos los demás factores en sus niveles más altos y el resto de covariables, en sus niveles medios. Si un factor o una covariable interactúan con otros predictores en el modelo, las razones de las ventajas dependerán no sólo de la modificación en la variable especificada, sino también de los valores de las variables con las que interactúe. Si una covariable especificada interactúa consigo misma en el modelo (por ejemplo, *edad*edad*), las razones de las ventajas dependerán entonces tanto del cambio en la covariable como del valor de ésta.

Regresión logística de muestras complejas: Guardar

Guardar variables. Este grupo permite guardar la categoría pronosticada para el modelo y las probabilidades pronosticadas como nuevas variables en el conjunto de datos activo.

Exportar modelo como datos de IBM SPSS Statistics. Escribe un conjunto de datos de formato IBM SPSS Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores estándar, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **rowtype_.** Toma los valores (y las etiquetas de valor), COV (covarianzas), CORR (correlaciones), EST (estimaciones de los parámetros), SE (errores estándar), SIG (niveles de significación) y DF (grados de libertad del diseño muestral). Hay un caso diferente con el tipo de fila COV (o CORR) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.
- **varname_.** Toma los valores P1, P2, ..., correspondientes a una lista ordenada de todos los parámetros del modelo para los tipos de fila COV o CORR, con las etiquetas de valor correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.
- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila. Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores estándar, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

Nota: este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan.

Exportar modelo como XML. Guarda las estimaciones de los parámetros y la matriz de covarianzas de los parámetros (si se selecciona) en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Regresión logística de muestras complejas: Opciones

Estimación. Este grupo otorga el control sobre varios criterios utilizados en la estimación del modelo.

- **Nº máximo de iteraciones.** Número máximo de iteraciones que se ejecutará el algoritmo. Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Limitar las iteraciones en función del cambio en las estimaciones de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros sean inferiores que el valor especificado, que debe ser no negativo.
- **Limitar las iteraciones en función del cambio en la log-verosimilitud.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en la función de log-verosimilitud sean inferiores que el valor especificado, que debe ser no negativo.
- **Comprobar si hay separación completa de los puntos de los datos.** Si se activa, el algoritmo realiza una prueba para garantizar que las estimaciones de los parámetros tienen valores exclusivos. Se produce una separación cuando el procedimiento pueda generar un modelo que clasifique cada caso de forma correcta.
- **Mostrar historial de iteraciones.** Muestra los estadísticos y las estimaciones de los parámetros cada n iteraciones, comenzando por la iteración 0 (estimaciones iniciales). Si decide imprimir el historial de iteraciones, la última iteración se imprimirá siempre independientemente del valor de n .

Valores perdidos del usuario. Todas las variables de diseño, así como la variable dependiente y cualquier covariable, deben contener datos válidos. Los casos con datos no válidos de cualquiera de estas variables se excluyen del análisis. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre las variables de estratificación, de clúster, de subpoblación y de factor.

Intervalo de confianza. Se trata del nivel de intervalo de confianza para las estimaciones de coeficiente, las estimaciones de coeficiente exponenciadas y las razones de las ventajas. Especifique un valor mayor o igual a 50 e inferior a 100.

Características adicionales del comando CSLOGISTIC

La sintaxis de comandos también le permite:

- Especificar contrastes personalizados de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando CUSTOM).
- Fijar valores de otras variables de modelo al calcular las razones de las ventajas para factores y covariables (utilizando el subcomando ODDS RATIOS).
- Especificar un valor de tolerancia para la comprobación de la singularidad (utilizando el subcomando CRITERIA).
- Crear nombres especificados por el usuario para las variables almacenadas (utilizando el subcomando SAVE).
- Generar una tabla de función estimable general (utilizando el subcomando PRINT).

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

Capítulo 11. Regresión ordinal de muestras complejas

El procedimiento Regresión ordinal de muestras complejas realiza análisis de regresión sobre una variable dependiente ordinal o binaria para muestras extraídas mediante métodos de muestreo complejo. Si lo desea, puede solicitar análisis de una subpoblación.

Ejemplo. Los diputados que estudian un proyecto de ley antes de una asamblea legislativa se interesan por conocer si la opinión pública apoya dicho proyecto de ley y qué relación guarda dicho apoyo con los datos demográficos de los votantes. Los encuestadores diseñan entrevistas y las realizan siguiendo un diseño del muestreo complejo. Utilice la regresión ordinal de muestras complejas para ajustar un modelo acerca del nivel de apoyo a la ley de acuerdo en los datos demográficos de los votantes.

Regresión ordinal de muestras complejas: Consideraciones sobre los datos

Datos. La variable dependiente es ordinal. Los factores son categóricos; pueden tener valores numéricos o valores de cadena de hasta ocho caracteres. Las covariables son variables cuantitativas que están relacionadas con la variable dependiente. Las variables que definen las subpoblaciones pueden ser numéricas o de cadena, pero siempre deben ser categóricas.

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Obtención de regresión ordinal de muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Regresión ordinal...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. En el cuadro de diálogo Regresión ordinal de muestras complejas, seleccione una variable dependiente.

Si lo desea, puede:

- Seleccione variables para factores y covariables, según corresponda a los datos.
- Especifique una variable para definir una subpoblación. El análisis se realiza únicamente para la categoría seleccionada de la variable de subpoblación, aunque para la estimación correcta de las varianzas sigue siendo necesario basarse en el conjunto de datos completo.
- Seleccione una función de enlace.

Función de enlace. La función de enlace es una transformación de las probabilidades acumuladas que permiten la estimación del modelo. Se encuentran disponibles las cinco funciones siguientes.

- **Logit.** $f(x)=\log(x/(1-x))$. Se utiliza típicamente para categorías uniformemente distribuidas.
- **Log-log complementario.** $f(x)=\log(-\log(1-x))$. Se utiliza normalmente cuando las categorías más altas son más probables.
- **Log-log negativo.** $f(x)=-\log(-\log(x))$. Se utiliza normalmente cuando las categorías más bajas son más probables.
- **Probit.** $f(x)=\Phi^{-1}(x)$. Se utiliza normalmente cuando la variable latente sigue una distribución normal.
- **Cauchit (Cauchy inversa).** $f(x)=\tan(\pi(x-0.5))$. Se utiliza normalmente cuando la variable latente tiene muchos valores extremos.

Regresión ordinal de muestras complejas: Probabilidades de respuesta

El cuadro de diálogo Probabilidades de respuesta permite especificar si la probabilidad acumulada de una respuesta (es decir, la probabilidad de pertenecer hasta una determinada categoría, incluida la propia categoría, de la variable dependiente) aumenta con valores de que aumentan o disminuyen de la variable dependiente.

Regresión ordinal de muestras complejas: Modelo

Especificar efectos del modelo. De forma predeterminada, el procedimiento crea un modelo de efectos principales utilizando los factores y las covariables especificadas en el cuadro de diálogo principal. Si lo desea, también puede crear un modelo personalizado que contenga los efectos de la interacción y los términos anidados.

Términos no anidados

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel para todas las variables seleccionadas.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4. Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quántuples posibles de las variables seleccionadas.

Términos anidados

En este procedimiento, puede generar términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

Limitaciones. Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A(A)$.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si A es un factor y X es una covariable, no es válido especificar $A(X)$.

Regresión ordinal de muestras complejas: Estadísticos

Ajuste del modelo. Controla la presentación de estadísticos que miden el rendimiento global del proceso.

- **Pseudo R cuadrado.** El estadístico R^2 de regresión lineal no cuenta con un análogo exacto entre los modelos de regresión ordinal. En su lugar existen varias medidas que tratan de imitar las propiedades del estadístico R^2 .
- **Tabla de clasificación.** Muestra las clasificaciones cruzadas tabuladas de la categoría observada por la categoría pronosticada por el modelo en la variable dependiente.

Parámetros. Este grupo permite controlar la presentación de estadísticos relacionados con los parámetros del modelo.

- **Estimación.** Muestra estimaciones de los coeficientes.
- **Estimación exponenciada.** Muestra la base del logaritmo natural elevada a la potencia de las estimaciones de los coeficientes. Mientras que las estimaciones tienen propiedades agradables para la comprobación estadística, la estimación exponenciada (o $\exp[B]$) es más sencilla de interpretar.
- **Error estándar.** Muestra el error estándar de cada estimación de los coeficientes.
- **Intervalo de confianza.** Muestra un intervalo de confianza para cada estimación de los coeficientes. El nivel de confianza de los intervalos se configura en el cuadro de diálogo Opciones.
- **Prueba t.** Muestra una prueba t de cada estimación de coeficientes. La hipótesis nula de cada prueba es que el valor del coeficiente sea 0.
- **Covarianzas de las estimaciones de los parámetros.** Muestra una estimación de la matriz de covarianzas de los coeficientes del modelo.
- **Correlaciones de las estimaciones de los parámetros.** Muestra una estimación de la matriz de correlaciones de los coeficientes del modelo.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida, expresada en unidades y comparable a las de los errores estándar, resultado de especificar un diseño complejo, donde los valores más distantes de 1 indican mayores efectos.

Líneas paralelas. Este grupo permite solicitar estadísticos asociados a un modelo con líneas no paralelas, donde se ajusta una línea de regresión distinta para cada categoría de respuesta (excepto la última).

- **Prueba de Wald.** Produce una prueba de la hipótesis nula de que los parámetros de regresión son iguales para todas las respuestas acumuladas. Se estima el modelo con líneas no paralelas y se aplica la prueba de Wald de parámetros iguales.
- **Estimaciones de los parámetros.** Muestra las estimaciones de los coeficientes y errores estándar para el modelo con líneas no paralelas.
- **Covarianzas de las estimaciones de los parámetros.** Muestra una estimación de la matriz de covarianza para los coeficientes del modelo con líneas no paralelas.

Estadísticos de resumen para las variables del modelo. Muestra información resumida acerca de los factores, las covariables y las variables dependientes.

Información del diseño muestral. Muestra información resumida acerca de la muestra, incluidos un recuento no ponderado y el tamaño de la población.

Muestras complejas: Contrastes de hipótesis

Estadístico de contraste. Este grupo le permite seleccionar el tipo de estadístico utilizado para contrastar las hipótesis. Es posible elegir entre F , F corregida, chi-cuadrado y chi-cuadrado corregido.

Muestreo de grados de libertad. Este grupo permite controlar los grados de libertad en el diseño del muestreo usados para calcular los valores p de todos los estadísticos de prueba. Si se basa en el diseño

del muestreo, el valor es la diferencia entre el número de unidades de muestra primarias y el número de estratos de la primera etapa del muestreo. Si lo desea, puede especificar los grados de libertad que desee introduciendo un número entero positivo.

Corrección para comparaciones múltiples. Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- *Sidak secuencial.* Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Bonferroni secuencial.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Sidak.* Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- *Bonferroni.* Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.

Regresión ordinal de muestras complejas: Razones de las ventajas

El cuadro de diálogo Razones de las ventajas permite mostrar las razones de las ventajas acumuladas estimadas por el modelo para los factores y las covariables que se especifican. Esta característica está disponible únicamente para modelos que utilizan la función de enlace Logit. Se calcula una sola razón de las ventajas acumulada para todas las categorías de la variante dependiente, excepto la última; el modelo de razones proporcionales postula que son todas iguales.

Factores. En cada factor seleccionado, muestra la odds ratio acumulada de cada categoría del factor hasta la probabilidad en la categoría de referencia especificada.

Covariables. En cada covariable seleccionada, muestra la odds ratio acumulada en el valor medio de la covariable más las unidades de cambio especificadas para la probabilidad de la media.

Al calcular las razones de las ventajas de un factor o una covariable, el procedimiento fija todos los demás factores en sus niveles más altos y el resto de covariables, en sus niveles medios. Si un factor o una covariable interactúan con otros predictores en el modelo, las razones de las ventajas dependerán no sólo de la modificación en la variable especificada, sino también de los valores de las variables con las que interactúe. Si una covariable especificada interactúa consigo misma en el modelo (por ejemplo, *edad*edad*), las razones de las ventajas dependerán entonces tanto del cambio en la covariable como del valor de ésta.

Regresión ordinal de muestras complejas: Guardar

Guardar variables. Este grupo permite guardar la categoría pronosticada para el modelo, la probabilidad de la categoría pronosticada, la probabilidad de la categoría observada, las probabilidades acumuladas y las probabilidades pronosticadas como nuevas variables en el conjunto de datos activo.

Exportar modelo como datos de IBM SPSS Statistics. Escribe un conjunto de datos de formato IBM SPSS Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores estándar, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **rowtype_.** Toma los valores (y las etiquetas de valor), COV (covarianzas), CORR (correlaciones), EST (estimaciones de los parámetros), SE (errores estándar), SIG (niveles de significación) y DF (grados de

libertad del diseño muestral). Hay un caso diferente con el tipo de fila COV (o CORR) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.

- **varname_.** Toma los valores P1, P2, ..., correspondientes a una lista ordenada de todos los parámetros del modelo para los tipos de fila COV o CORR, con las etiquetas de valor correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.
- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila. Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores estándar, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

Nota: este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan.

Exportar modelo como XML. Guarda las estimaciones de los parámetros y la matriz de covarianzas de los parámetros (si se selecciona) en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Regresión ordinal de muestras complejas: Opciones

Método de estimación. Puede seleccionar un método de estimación de parámetros. Los métodos disponibles son Newton-Raphson, Puntuación de Fisher o un método híbrido en el que las iteraciones de Puntuación de Fisher se realizan antes de cambiar al método de Newton-Raphson. Si se logra la convergencia durante la fase de Puntuación de Fisher del método híbrido antes de que se lleven a cabo el número máximo de iteraciones de Fisher, el algoritmo continúa con el método de Newton-Raphson.

Estimación. Este grupo otorga el control sobre varios criterios utilizados en la estimación del modelo.

- **Nº máximo de iteraciones.** Número máximo de iteraciones que se ejecutará el algoritmo. Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Limitar las iteraciones en función del cambio en las estimaciones de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros sean inferiores que el valor especificado, que debe ser no negativo.
- **Limitar las iteraciones en función del cambio en la log-verosimilitud.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en la función de log-verosimilitud sean inferiores que el valor especificado, que debe ser no negativo.
- **Comprobar si hay separación completa de los puntos de los datos.** Si se activa, el algoritmo realiza una prueba para garantizar que las estimaciones de los parámetros tienen valores exclusivos. Se produce una separación cuando el procedimiento pueda generar un modelo que clasifique cada caso de forma correcta.
- **Mostrar historial de iteraciones.** Muestra los estadísticos y las estimaciones de los parámetros cada n iteraciones, comenzando por la iteración 0 (estimaciones iniciales). Si decide imprimir el historial de iteraciones, la última iteración se imprimirá siempre independientemente del valor de n .

Valores perdidos del usuario. Las variables de diseño de escala, así como la variable dependiente y cualquier covariable, deben contener datos válidos. Los casos con datos no válidos de cualquiera de estas variables se excluyen del análisis. Estos controles permiten decidir si los valores perdidos del usuario se deben tratar como válidos entre las variables de estratificación, de clúster, de subpoblación y de factor.

Intervalo de confianza. Se trata del nivel de intervalo de confianza para las estimaciones de coeficiente, las estimaciones de coeficiente exponenciadas y las razones de las ventajas. Especifique un valor mayor o igual a 50 e inferior a 100.

Características adicionales del comando CSORDINAL

La sintaxis de comandos también le permite:

- Especificar contrastes personalizados de los efectos respecto a una combinación lineal de efectos o un valor (utilizando el subcomando CUSTOM).
- Fijar valores de otras variables de modelo en valores distintos de sus medias al calcular las razones de las ventajas para factores y covariables (utilizando el subcomando ODDS RATIOS).
- Utilice valores sin etiquetar como categorías de referencia personalizadas para los factores cuando se soliciten razones de las ventajas (usando el subcomando ODDS RATIOS).
- Especificar un valor de tolerancia para la comprobación de la singularidad (utilizando el subcomando CRITERIA).
- Generar una tabla de función estimable general (utilizando el subcomando PRINT).
- Guarde más de 25 variables de probabilidad (usando el subcomando SAVE).

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

Capítulo 12. Regresión de Cox de muestras complejas

El procedimiento Regresión de Cox de muestras complejas realiza análisis de supervivencias para muestras extraídas mediante métodos de muestreo complejo. Si lo desea, puede solicitar análisis de una subpoblación.

Ejemplos. Un organismo de orden público está preocupado por los índices de reincidencia en su área de jurisdicción. Una de las medidas de la reincidencia es el tiempo que transcurre antes del segundo arresto de los delincuentes. El organismo desea crear un modelo que refleje el tiempo que transcurre antes de un nuevo arresto utilizando la regresión de Cox, pero les preocupa que el supuesto de proporcionalidad de los riesgos no sea válido en todas las diferentes categorías de edad.

Un grupo de investigadores estudia los tiempos de supervivencia de los pacientes que finalizan un programa de rehabilitación tras un ataque isquémico. Pueden existir varios casos por sujeto, ya que los historiales de los pacientes cambian cuando se registra la ocurrencia de eventos diferentes de la muerte y el momento en que ocurren dichos eventos. La muestra también está truncada a la izquierda en el sentido de que los tiempos de supervivencia observados se han "inflado" con la duración de la rehabilitación, ya que el comienzo del riesgo comienza en el momento del ataque isquémico, dado que la muestra incluye únicamente a los pacientes que han sobrevivido al programa de rehabilitación.

Regresión de Cox de muestras complejas: Consideraciones sobre los datos

Tiempo de supervivencia. El procedimiento aplica la regresión de Cox al análisis de los tiempos de supervivencia, es decir, el tiempo que transcurre hasta que se produzca un evento. Hay dos maneras de especificar el tiempo de supervivencia, dependiendo del momento de inicio del intervalo:

- **Tiempo=0.** Normalmente, dispondrá de información completa acerca del inicio del intervalo para cada sujeto y sencillamente tendrá una variable que contenga los momentos de finalización (o creará una única variable con los momentos de finalización a partir de variables de fecha y hora, como verá a continuación).
- **Varía según el sujeto.** Esta opción es adecuada cuando tiene **truncación a la izquierda**, también denominada **entrada retardada**; por ejemplo, si desea analizar los tiempos de supervivencia de los pacientes que finalizan un programa de rehabilitación tras un ataque, tal vez prefiera considerar que el comienzo del riesgo comienza en el momento en que se produjo el ataque. No obstante, si la muestra incluye únicamente a aquellos pacientes que han sobrevivido al programa de rehabilitación, la muestra está truncada a la izquierda en el sentido de que los tiempos de supervivencia observados se han "inflado" con la duración de la rehabilitación. Puede tener este hecho en cuenta si especifica el momento en el que abandonaron la rehabilitación como el momento de entrada en el estudio.

Variabes de fecha y hora. No se pueden utilizar directamente variables de fecha y hora para definir el inicio y la finalización del intervalo. Si tiene variables de fecha y hora, deberá utilizarlas para crear variables que contengan tiempos de supervivencia. Si no hay ninguna truncación a la izquierda, basta con crear una variable que contenga las horas de finalización basadas en la diferencia entre la fecha de entrada en el estudio y la fecha de la observación. Si existe truncación a la izquierda, puede crear una variable que contenga las horas de inicio, basadas en la diferencia entre la fecha de inicio del estudio y la fecha de entrada, y otra variable que contenga las horas de finalización, basadas en la diferencia entre la fecha de inicio del estudio y la fecha de la observación.

Estado del evento. Necesita una variable que registre si el sujeto ha experimentado el evento de interés dentro del intervalo. Los sujetos para los que no se ha producido el evento se censuran a la derecha.

Identificador de sujetos. Puede incorporar fácilmente predictores dependientes del tiempo y constantes por tramos dividiendo las observaciones de un único sujeto en varios casos. Por ejemplo, si desea

analizar los tiempos de supervivencia de los pacientes tras un ataque, las variables que representan su historial médico probablemente sean útiles como predictores. Con el tiempo, pueden sufrir eventos médicos importantes que alteren su historial médico. La siguiente tabla muestra cómo estructurar un conjunto de datos de este tipo: *ID de paciente* es el identificador de los sujetos, *Hora de finalización* define los intervalos observados, *Estado* registra los eventos médicos importantes; *Historial anterior de ataques al corazón* e *Historial previo de hemorragias* son predictores dependientes del tiempo y constantes por tramos.

Tabla 1. Estructura de datos para la incorporación de predictores dependientes del tiempo constantes por tramos.

| <i>ID de paciente</i> | <i>Hora de finalización</i> | <i>Estado</i> | <i>Historial anterior de ataques al corazón</i> | <i>Historial previo de hemorragias</i> |
|-----------------------|-----------------------------|-------------------|---|--|
| 1 | 5 | Ataque al corazón | No | No |
| 1 | 7 | Hemorragia | Sí | No |
| 1 | 8 | Fallecimiento | Sí | Sí |
| 2 | 24 | Fallecimiento | No | No |
| 3 | 8 | Ataque al corazón | No | No |
| 3 | 15 | Fallecimiento | Sí | No |

Supuestos. Los casos del archivo de datos representan una muestra de un diseño complejo que se debe analizar según las especificaciones del archivo seleccionado en el Cuadro de diálogo Plan de muestras complejas.

Normalmente, los modelos de regresión de Cox suponen que los riesgos son proporcionales, es decir, que la proporción de riesgos de un caso a otro no cambia con el tiempo. Si no se cumple este supuesto, tal vez sea necesario añadir predictores dependientes del tiempo al modelo.

Análisis Kaplan-Meier. Si no selecciona ningún predictor (o no introduce ninguno de los predictores seleccionados en el modelo) y elige el método de límite de producto para calcular la curva de supervivencia basal en la pestaña Opciones, el procedimiento realizará un análisis de supervivencia de tipo Kaplan-Meier.

Para obtener la regresión de Cox de muestras complejas

1. Seleccione en los menús:
Analizar > Muestras complejas > Regresión de Cox...
2. Seleccione un archivo de plan. Si lo desea, elija un archivo de probabilidades conjuntas personalizado.
3. Pulse en **Continuar**.
4. Especifique el tiempo de supervivencia seleccionando los momentos de entrada y salida del estudio.
5. Seleccione una variable de estado del evento.
6. Pulse Definir evento y defina al menos un valor de evento.

Si lo desea, puede seleccionar un identificador de sujetos.

Definir evento

Especifique los valores que indican que se ha producido un evento terminal.

- **Valores individuales.** Especifique uno o más valores introduciéndolos en la cuadrícula o seleccionándolos en una lista de valores con etiquetas de valor definidas.
- **Rango de valores.** Especifique un rango de valores introduciendo los valores mínimo y máximo o seleccionando los valores en una lista con etiquetas de valor definidas.

Predictores

La pestaña Predictores permite especificar los factores y las covariables que se utilizarán para crear los efectos del modelo.

Factores. Los factores son predictores categóricos y pueden ser numéricos o de cadena.

Covariables. Las covariables son predictores de escala y deben ser numéricas.

Predictores dependientes del tiempo. Hay ciertas situaciones en las que no se cumple el supuesto de proporcionalidad de los riesgos. Es decir, las tasas de riesgo cambian con el tiempo y los valores de uno (o más) de los predictores son diferentes en los distintos puntos temporales. En tales casos, es necesario especificar predictores dependientes del tiempo. Consulte el tema "Definir predictor dependiente del tiempo" para obtener más información. Los predictores dependientes del tiempo se pueden seleccionar como factores o como covariables.

Definir predictor dependiente del tiempo

El cuadro de diálogo Definir predictor dependiente del tiempo le permite crear un predictor que dependa de la variable de tiempo preincorporada, T_* . Puede utilizar esta variable para definir covariables dependientes del tiempo empleando dos métodos generales:

- Si desea estimar un modelo de regresión de Cox extendido que permita riesgos no proporcionales, defina el predictor dependiente del tiempo como una función de la variable de tiempo T_* y la covariable en cuestión. Un ejemplo habitual sería el simple producto de la variable de tiempo y el predictor, pero también se pueden especificar funciones más complejas.
- Algunas variables pueden tener valores distintos en períodos diferentes del tiempo, pero no están sistemáticamente relacionadas con el tiempo. En tales casos es necesario definir un **predictor dependiente del tiempo segmentado**, lo cual puede llevarse a cabo usando expresiones lógicas. Las expresiones lógicas toman el valor 1 cuando son verdaderas y el valor 0 cuando son falsas. Es posible crear un predictor dependiente del tiempo a partir de un conjunto de mediciones, usando una serie de expresiones lógicas. Por ejemplo, si se toma la tensión una vez a la semana durante cuatro semanas (identificadas como $BP1$ a $BP4$), puede definir el predictor dependiente del tiempo como $(T_* < 1) * BP1 + (T_* \geq 1 \ \& \ T_* < 2) * BP2 + (T_* \geq 2 \ \& \ T_* < 3) * BP3 + (T_* \geq 3 \ \& \ T_* < 4) * BP4$. Tenga en cuenta que exactamente uno de los términos entre paréntesis será igual a uno para cualquier caso dado y el resto serán todos 0. En otras palabras, esta función se puede interpretar diciendo que "Si el tiempo es inferior a una semana, use $BP1$; si es más de una semana pero menos de dos, utilice $BP2$; y así sucesivamente".

Nota: si el predictor dependiente del tiempo segmentado es constante en los segmentos, como ocurre en el ejemplo de la tensión arterial explicado anteriormente, es posible que sea más sencillo especificar el predictor dependiente del tiempo y constante por tramos dividiendo los sujetos en varios casos.

Consulte la explicación acerca de los identificadores de sujetos en Capítulo 12, "Regresión de Cox de muestras complejas", en la página 49 para obtener más información.

En el cuadro de diálogo Definir predictor dependiente del tiempo, puede utilizar los controles de generación de funciones para crear la expresión para la covariable dependiente del tiempo o bien introducirla directamente en el área de texto Expresión numérica. Tenga en cuenta que las constantes de cadena deben ir entre comillas o apóstrofes y que las constantes numéricas se deben escribir en formato americano, con el punto como separador de la parte decimal. A la variable resultante se le asignará el nombre especificado y deberá incluirse como factor o covariable en la pestaña Predictores.

Subgrupos

Estratos de referencia. Se calcula una función de supervivencia y de riesgo basal diferente para cada valor de esta variable, a la vez que se estima un único conjunto de coeficientes del modelo en todos los estratos.

Variable de subpoblación. Especifique una variable para definir una subpoblación. El análisis se lleva a cabo únicamente en la categoría seleccionada de la variable de subpoblación.

Modelo

Especificar efectos del modelo. De forma predeterminada, el procedimiento crea un modelo de efectos principales utilizando los factores y las covariables especificadas en el cuadro de diálogo principal. Si lo desea, también puede crear un modelo personalizado que contenga los efectos de la interacción y los términos anidados.

Términos no anidados

Para las covariables y los factores seleccionados:

Interacción. Crea el término de interacción de mayor nivel para todas las variables seleccionadas.

Efectos principales. Crea un término de efectos principales para cada variable seleccionada.

Todas de 2. Crea todas las interacciones bidimensionales posibles de las variables seleccionadas.

Todas de 3. Crea todas las interacciones tridimensionales posibles de las variables seleccionadas.

Todas de 4. Crea todas las interacciones tetradimensionales posibles de las variables seleccionadas.

Todas de 5. Crea todas las interacciones quintuples posibles de las variables seleccionadas.

Términos anidados

En este procedimiento, puede generar términos anidados para el modelo. Los términos anidados resultan útiles para modelar el efecto de un factor o covariable cuyos valores no interactúan con los niveles de otro factor. Por ejemplo, una cadena de tiendas de comestibles desea realizar un seguimiento de los hábitos de gasto de los clientes en las diversas ubicaciones de sus tiendas. Dado que cada cliente frecuenta tan sólo una de estas ubicaciones, se puede decir que el efecto de *Cliente* está **anidado dentro** del efecto de *Ubicación de la tienda*.

Además, puede incluir efectos de interacción, como términos polinómicos que implican a la misma covariable, o añadir varios niveles de anidación al término anidado.

Limitaciones. Existen las siguientes restricciones para los términos anidados:

- Todos los factores incluidos en una interacción deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A*A$.
- Todos los factores incluidos en un efecto anidado deben ser exclusivos entre sí. Por consiguiente, si A es un factor, no es válido especificar $A(A)$.
- No se puede anidar ningún efecto dentro de una covariable. Por consiguiente, si A es un factor y X es una covariable, no es válido especificar $A(X)$.

Estadísticas

Información del diseño muestral. Muestra información resumida acerca de la muestra, incluidos un recuento no ponderado y el tamaño de la población.

Resumen de censura y eventos. Muestra información resumida acerca del número y el porcentaje de los casos censurados.

Riesgo establecido en las horas de los eventos. Muestra el número de eventos y el número bajo riesgo para cada momento de evento en cada estrato de referencia.

Parámetros. Este grupo permite controlar la presentación de estadísticos relacionados con los parámetros del modelo.

- **Estimación.** Muestra estimaciones de los coeficientes.
- **Estimación exponenciada.** Muestra la base del logaritmo natural elevada a la potencia de las estimaciones de los coeficientes. Mientras que las estimaciones tienen propiedades agradables para la comprobación estadística, la estimación exponenciada (o $\exp[B]$) es más sencilla de interpretar.
- **Error estándar.** Muestra el error estándar de cada estimación de los coeficientes.
- **Intervalo de confianza.** Muestra un intervalo de confianza para cada estimación de los coeficientes. El nivel de confianza de los intervalos se configura en el cuadro de diálogo Opciones.
- **Prueba t.** Muestra una prueba t de cada estimación de coeficientes. La hipótesis nula de cada prueba es que el valor del coeficiente sea 0.
- **Covarianzas de las estimaciones de los parámetros.** Muestra una estimación de la matriz de covarianzas de los coeficientes del modelo.
- **Correlaciones de las estimaciones de los parámetros.** Muestra una estimación de la matriz de correlaciones de los coeficientes del modelo.
- **Efecto de diseño.** Cociente de la variación de la estimación entre la variación obtenida al suponer que la muestra es una muestra aleatoria simple. Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.
- **Raíz cuadrada del efecto de diseño.** Es una medida del efecto de especificar un diseño complejo donde los valores más distantes de 1 indican efectos mayores.

Supuestos del modelo. Este grupo le permite generar una prueba del supuesto de proporcionalidad de los riesgos. El contraste compara el modelo ajustado con un modelo alternativo que incluye los predictores dependientes del tiempo x^*_{TF} para cada predictor x , donde $_{TF}$ es la función de tiempo especificada.

- **Función de tiempo.** Especifica la forma de $_{TF}$ para el modelo alternativo. Para la función **identidad**, $_{TF}=T_$. Para la función **log**, $_{TF}=\log(T_)$. Para la función **Kaplan-Meier**, $_{TF}=1-S_{KM}(T_)$, donde $S_{KM}(\cdot)$ es la estimación de Kaplan-Meier de la función de supervivencia. Para **rango**, $_{TF}$ es el orden de rango de $T_$ entre los momentos de finalización observados.
- **Estimaciones de los parámetros para el modelo alternativo.** Muestra la estimación, el error estándar y el intervalo de confianza de cada parámetro del modelo alternativo.
- **Matriz de covarianzas del modelo alternativo.** Muestra la matriz de las covarianzas estimadas entre los parámetros del modelo alternativo.

Funciones de supervivencia de referencia y riesgo acumulado. Muestra la función de supervivencia de referencia y la función de riesgos acumulados de línea base junto con sus errores estándar.

Nota: si se han incluido en el modelo los predictores dependientes del tiempo definidos en la pestaña Predictores, esta opción no está disponible.

Gráficos

La pestaña Gráficos le permite solicitar gráficos de la función de riesgo, la función de supervivencia, la función log menos log de la supervivencia y la función uno menos la supervivencia. También puede solicitar que se representen los intervalos de confianza junto con las funciones especificadas; el nivel de confianza se establece en la pestaña Opciones.

Patrones de predictores. Puede especificar que se utilice un patrón de valores de predictores para los gráficos solicitados y el archivo de supervivencia exportado en la pestaña Exportar. Tenga en cuenta que estas opciones no están disponibles si se han incluido en el modelo los predictores dependientes del tiempo definidos en la pestaña Predictores.

- **Representar factores respecto a.** De forma predeterminada, cada factor se evalúa respecto a su nivel superior. Si lo desea, introduzca o seleccione otro nivel. También puede solicitar que se representen líneas distintas para cada nivel de un factor individual activando la casilla de verificación correspondiente a dicho factor.
- **Representar covariables respecto a.** Cada covariable se evalúa respecto a su media. Si lo desea, introduzca o seleccione otro valor.

Contrastes de hipótesis

Estadístico de contraste. Este grupo le permite seleccionar el tipo de estadístico utilizado para contrastar las hipótesis. Es posible elegir entre F , F corregida, chi-cuadrado y chi-cuadrado corregido.

Muestreo de grados de libertad. Este grupo permite controlar los grados de libertad en el diseño del muestreo usados para calcular los valores p de todos los estadísticos de prueba. Si se basa en el diseño del muestreo, el valor es la diferencia entre el número de unidades de muestra primarias y el número de estratos de la primera etapa del muestreo. Si lo desea, puede especificar los grados de libertad que desee introduciendo un número entero positivo.

Corrección para comparaciones múltiples. Al realizar contrastes de hipótesis con varios contrastes, el nivel de significación global se puede ajustar utilizando los niveles de significación de los contrastes incluidos. Este grupo permite elegir el método de ajuste.

- **Diferencia menos significativa.** Este método no controla la probabilidad general de rechazar las hipótesis de que algunos contrastes lineales son diferentes a los valores de hipótesis nula.
- *Sidak secuencial.* Este es un procedimiento de Sidak de rechazo secuencial decreciente que es mucho menos conservador en términos de rechazar las hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Bonferroni secuencial.* Éste es un procedimiento de Bonferroni de rechazo secuencial decreciente que es mucho menos conservador en cuanto al rechazo de hipótesis individuales pero que mantiene el mismo nivel de significación global.
- *Sidak.* Este método ofrece límites más estrechos que los de la aproximación de Bonferroni.
- *Bonferroni.* Este método corrige el nivel de significación observado por el hecho de que se están poniendo a prueba múltiples contrastes.

Guardar

Guardar variables. Este grupo permite guardar las variables relacionadas con el modelo en el conjunto de datos activo para utilizarlas posteriormente en otros diagnósticos y los informes sobre los resultados. Ninguna de estas opciones estará disponible si se incluyen en el modelo predictores dependientes del tiempo.

- **Función de supervivencia.** Guarda la probabilidad de supervivencia (el valor de la función de supervivencia) en el momento observado y los valores de los predictores para cada caso.
- **Límite inferior del intervalo de confianza para la función de supervivencia.** Guarda el límite inferior del intervalo de confianza de la función de supervivencia en el momento observado y los valores de los predictores para cada caso.
- **Límite superior del intervalo de confianza para la función de supervivencia.** Guarda el límite superior de intervalo de confianza de la función de supervivencia en el momento observado y los valores de los predictores para cada caso.
- **Función de riesgo acumulado.** Guarda el riesgo acumulado, o $-\ln(\text{supervivencia})$, en el momento observado y los valores de los predictores para cada caso.

- **Límite inferior del intervalo de confianza para la función de riesgo acumulado.** Guarda el límite superior de intervalo de confianza de la función de riesgo acumulado en el momento observado y los valores de los predictores para cada caso.
- **Límite superior del intervalo de confianza para la función de riesgo acumulado.** Guarda el límite inferior del intervalo de confianza de la función de riesgo acumulado en el momento observado y los valores de los predictores para cada caso.
- **Valor predicho del predictor lineal.** Guarda la combinación lineal de los predictores corregidos de los valores de referencia por los coeficientes de regresión. El predictor lineal es la proporción de la función de riesgo respecto al riesgo de línea base. En el modelo de riesgos proporcionales, este valor es constante respecto al tiempo.
- **Residuo de Schoenfeld.** Para cada caso no censurado y cada parámetro no redundante del modelo, el residuo de Schoenfeld es la diferencia entre el valor observado del predictor asociado al parámetro del modelo y el valor esperado del predictor para los casos con el riesgo establecido en el momento del evento observado. Los residuos de Schoenfeld se pueden utilizar para evaluar el supuesto de proporcionalidad de los riesgos; por ejemplo, para un predictor x , los gráficos de los residuos de Schoenfeld de un predictor dependiente del tiempo $x \cdot \ln(T_)$ respecto al tiempo mostrarían una línea horizontal en 0 si se cumple el supuesto de proporcionalidad de los riesgos. Se guarda una variable diferente para cada parámetro no redundante del modelo. Sólo se calculan los residuos de Schoenfeld para los casos no censurados.
- **Residuo de Martingale.** Para cada caso, el residuo de Martingale es la diferencia entre la censura observada (0 si está censurado, 1 si no lo está) y la esperanza de un evento durante el tiempo de observación.
- **Residuo de desviación.** Los residuos de desviación son residuos de Martingale "corregidos" para aparecer más simétricos alrededor de 0. Los gráficos de los residuos de desviación respecto a los predictores no deben mostrar ningún patrón.
- **Residuo de Cox-Snell.** Para cada caso, el residuo de Cox-Snell es la esperanza de que se produzca un evento durante el tiempo de observación o la censura observada menos el residuo de Martingale.
- **Residuo de puntuación.** Para cada caso y cada parámetro no redundante del modelo, el residuo de puntuación es la contribución del caso a la primera derivada de la pseudoverosimilitud. Se guarda una variable diferente para cada parámetro no redundante del modelo.
- **Residuo de DFBeta.** Para cada caso y cada parámetro no redundante del modelo, el residuo de DFBeta es una aproximación del cambio en el valor de la estimación del parámetro cuando se elimina el caso del modelo. Los casos con residuos de DFBeta relativamente grandes pueden estar ejerciendo una influencia indebida sobre el análisis. Se guarda una variable diferente para cada parámetro no redundante del modelo.
- **Residuos agregados.** Cuando varios casos representan a un único sujeto, el residuo agregado de un sujeto es sencillamente la suma de los residuos de todos los casos que corresponden al mismo sujeto. Para el residuo de Schoenfeld, la versión agregada es la misma que la versión no agregada, ya que el residuo de Schoenfeld sólo se define para los casos no censurados. Estos residuos sólo están disponibles cuando se ha especificado un identificador de sujetos en la pestaña Momento y evento.

Nombres de las variables guardadas. La generación automática de nombres garantiza que conserva todo su trabajo. Los nombres personalizados le permiten descartar/reemplazar los resultados de las ejecuciones anteriores sin eliminar antes las variables guardadas en el Editor de datos.

Exportar

Exportar modelo como datos de IBM SPSS Statistics. Escribe un conjunto de datos de formato IBM SPSS Statistics que contiene la matriz de covarianzas o correlaciones de los parámetros con las estimaciones de los parámetros, errores estándar, valores de significación y grados de libertad. El orden de las variables en el archivo matricial es el siguiente.

- **rowtype_.** Toma los valores (y las etiquetas de valor), COV (covarianzas), CORR (correlaciones), EST (estimaciones de los parámetros), SE (errores estándar), SIG (niveles de significación) y DF (grados de

libertad del diseño muestral). Hay un caso diferente con el tipo de fila COV (o CORR) para cada parámetro del modelo, además de un caso diferente para cada uno de los otros tipos de filas.

- **varname_.** Toma los valores P1, P2, ..., correspondientes a una lista ordenada de todos los parámetros del modelo para los tipos de fila COV o CORR, con las etiquetas de valor correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros. Las casillas están vacías para los demás tipos de filas.
- **P1, P2, ...** Estas variables corresponden a una lista ordenada de todos los parámetros del modelo con las etiquetas de variable correspondientes a las cadenas de parámetros mostradas en la tabla de estimaciones de los parámetros y toman valores según el tipo de fila. Para los parámetros redundantes, todas las covarianzas se establecen en cero, las correlaciones se establecen en el valor perdido del sistema; todas las estimaciones de los parámetros se establecen en cero; y todos los errores estándar, niveles de significación y los grados de libertad residuales se establecen en el valor perdido del sistema.

Nota: este archivo no se puede utilizar directamente para realizar otros análisis en otros procedimientos que lean un archivo matricial a menos que dichos procedimientos acepten todos los tipos de filas que aquí se exportan.

Exportar función de supervivencia como datos de IBM SPSS Statistics. Escribe un conjunto de datos de formato IBM SPSS Statistics que contiene la función de supervivencia; el error estándar de la función de supervivencia; los límites superior e inferior del intervalo de confianza de la función de supervivencia; y la función de riesgos acumulados de cada momento de evento o fallo, evaluada en la línea base y en los patrones de predictores especificados en la pestaña Gráfico. El orden de las variables en el archivo matricial es el siguiente.

- **Variable de estratos de referencia.** Se generan una tabla de supervivencia diferente para cada valor de la variable de los estratos.
- **Variable de tiempo de supervivencia.** Hora del evento; se crea un caso distinto para cada hora de evento exclusiva.
- **Sur_0, LCL_Sur_0, UCL_Sur_0.** Función de supervivencia de referencia y límites inferior y superior de su intervalo de confianza.
- **Sur_R, LCL_Sur_R, UCL_Sur_R.** Función de supervivencia evaluada en el patrón de "referencia" (consulte la tabla de valores de patrones que aparece en los resultados) y los límites superior e inferior de su intervalo de confianza.
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...** La función de supervivencia evaluada en cada uno de los patrones de predictor especificados en la pestaña Gráficos y los límites superior e inferior de sus intervalos de confianza. Consulte la tabla de valores de los patrones que aparece en los resultados para ver la correspondencia entre los patrones y el número ##.
- **Haz_0, LCL_Haz_0, UCL_Haz_0.** Función de riesgo acumulado de línea base y límites inferior y superior de su intervalo de confianza.
- **Haz_R, LCL_Haz_R, UCL_Haz_R.** Función de riesgo acumulado evaluada en el patrón de "referencia" (consulte la tabla de valores de patrones que aparece en los resultados) y los límites superior e inferior de su intervalo de confianza.
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...** La función de riesgo acumulado evaluada en cada uno de los patrones de predictor especificados en la pestaña Gráficos y los límites superior e inferior de sus intervalos de confianza. Consulte la tabla de valores de los patrones que aparece en los resultados para ver la correspondencia entre los patrones y el número ##.

Exportar modelo como XML. Guarda toda la información necesaria para pronosticar la función de supervivencia, incluidas las estimaciones de los parámetros y la función de supervivencia de referencia, en formato XML (PMML). Puede utilizar este archivo de modelo para aplicar la información del modelo a otros archivos de datos para puntuarlo.

Opciones

Estimación. Estos controles especifican los criterios que se utilizan para estimar los coeficientes de regresión.

- **Nº máximo de iteraciones.** Número máximo de iteraciones que se ejecutará el algoritmo. Especifique un número entero no negativo.
- **Máxima subdivisión por pasos.** En cada iteración, se reduce el tamaño del paso mediante un factor de 0,5 hasta que aumenta el logaritmo de la verosimilitud o se alcanza la máxima subdivisión por pasos. Especifique un número entero positivo.
- **Limitar las iteraciones en función del cambio en las estimaciones de los parámetros.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en las estimaciones de los parámetros son inferiores al valor especificado, que debe ser positivo.
- **Limitar las iteraciones en función del cambio en la log-verosimilitud.** Si se activa, el algoritmo se detiene tras una iteración en la que las modificaciones absolutas o relativas en la función de log-verosimilitud sean inferiores que el valor especificado, que debe ser positivo.
- **Mostrar historial de iteraciones.** Muestra el historial de iteraciones de las estimaciones de los parámetros y la pseudo log-verosimilitud. Además, imprime la última evaluación del cambio en las estimaciones de los parámetros y el pseudo log-verosimilitud. La tabla del historial de iteraciones se imprime cada n iteraciones a partir de la iteración 0 (las estimaciones iniciales), donde n es el valor del incremento. Si se solicita el historial de iteraciones, la última iteración siempre se muestra independientemente de n .
- **Método de ruptura de empates para la estimación de los parámetros.** Cuando hay momentos de fallo observados empatados, se utiliza uno de los siguientes métodos para romper los empates. El método de Efron requiere un proceso de cálculo más extenso.

Funciones de supervivencia. Estos controles especifican los criterios de los cálculos que implican a la función de supervivencia.

- **Método de estimación de las funciones de supervivencia de referencia.** El método de **Breslow** (o de Nelson-Aalan o empírico) estima el riesgo acumulado de línea base mediante una función de paso no decreciente con pasos en los momentos de fallo observados. A continuación, calcula la supervivencia de referencia mediante la relación $\text{supervivencia} = \exp(-\text{riesgo acumulado})$. El método de **Efron** requiere un proceso de cálculo más extenso y se reduce al método de Breslow cuando no hay ningún empate. El método de **límite del producto** estima la supervivencia de referencia mediante una función continua a la derecha no creciente; cuando no hay ningún predictor en el modelo, este método se reduce a la estimación de Kaplan-Meier.
- **Intervalos de confianza de las funciones de supervivencia.** El intervalo de confianza se puede calcular de tres maneras: en unidades originales, mediante una transformación logarítmica o mediante una transformación log menos log. Sólo la transformación log menos log garantiza que los límites del intervalo de confianza estarán comprendidos entre 0 y 1, pero la transformación logarítmica suele funcionar "mejor".

Valores perdidos del usuario. Para que un caso se incluya en el análisis, todas las variables deben tener valores válidos para dicho caso. Estos controles permiten decidir si los valores perdidos del usuario se tratan como válidos en los modelos categóricos (incluidas las variables de subpoblación, estratos, evento y factores) y las variables del diseño del muestreo.

Intervalo de confianza (%). Nivel del intervalo de confianza utilizado para las estimaciones de los coeficientes, estimaciones de los coeficientes exponenciadas, estimaciones de la función de supervivencia y estimaciones de la función de riesgo acumulado. Especifique un valor mayor o igual a 0 y menor que 100.

Características adicionales del comando CSCOXREG

Con el lenguaje de comandos también podrá:

- Realizar contrastes de hipótesis personalizados (utilizando el subcomando CUSTOM y /PRINT LMATRIX).
- Especificar la tolerancia (utilizando /CRITERIA SINGULAR).
- Tabla de función estimable general (utilizando /PRINT GEF).
- Varios patrones de predictores (utilizando varios subcomandos PATTERN).
- Número máximo de variables guardadas cuando se especifica un nombre de raíz (utilizando el subcomando SAVE). El cuadro de diálogo respeta el valor predeterminado de CSCOXREG de 25 variables.

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

Avisos

Esta información se ha desarrollado para productos y servicios que se ofrecen en Estados Unidos.

Puede que en otros países IBM no ofrezca los productos, servicios ni características que se describen en esta información. Póngase en contacto con el representante de IBM de su localidad para obtener información acerca de los productos y servicios que actualmente están disponibles en su zona. Las referencias a programas, productos o servicios de IBM no pretenden establecer ni implicar que sólo se puedan utilizar productos, programas o servicios de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja ninguno de los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y comprobar el funcionamiento de todo producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de 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
EE.UU.

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:

Licencia de propiedad intelectual
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

El siguiente párrafo no se aplica al Reino Unido ni a ningún otro país donde dichas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN NINGÚN TIPO DE GARANTÍA, YA SEA EXPLÍCITA O IMPLÍCITA, INCLUIDAS, PERO SIN LIMITARSE A, 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. El material de esos sitios web no forman parte del material de este producto de IBM y el uso de dichos sitios web es responsabilidad del usuario.

IBM puede utilizar o distribuir la información que el usuario le suministre en el modo que considere apropiado sin incurrir en ninguna obligación con el usuario.

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

Tel. 901 100 400
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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

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

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

Se ha obtenido información acerca de productos que no son de IBM de los proveedores de esos productos, de sus publicaciones anunciadas o de otros orígenes disponibles públicamente. IBM no ha probado dichos productos y no puede confirmar la precisión de su rendimiento, la compatibilidad ni contemplar ninguna otra reclamación relacionada 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.

LICENCIA DE DERECHOS DE AUTOR:

Esta información contiene programas de aplicación de ejemplo escritos en lenguaje fuente, que muestran técnicas de programación en diversas plataformas operativas. Puede copiar, modificar y distribuir estos programas de muestra de cualquier forma, sin pagar nada a IBM, con los fines de desarrollar, utilizar, comercializar o distribuir programas de aplicación de acuerdo con la interfaz de programación de aplicaciones para la plataforma operativa para la cual se han escrito los programas de muestra. Estos programas no se han probado de forma exhaustiva en todas las circunstancias. IBM, por tanto, no puede garantizar ni presuponer la fiabilidad, el servicio o el funcionamiento de dichos programas. Los programas de ejemplo se proporcionan "TAL CUAL", sin garantía de ningún tipo. IBM no será responsable de los daños que surjan por el uso de los programas de muestra.

Cada copia o porción de estos programas de ejemplo, o cualquier trabajo derivado de los mismos, debe incluir un aviso de copyright como el siguiente:

© el nombre de la empresa) (año). Algunas partes de este código proceden de IBM Corp. Sample Programs.

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

Marcas comerciales

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

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

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

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

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

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

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

Índice

A

archivo de plan 2

B

Bonferroni
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54

C

categoría de referencia
en Modelo lineal general de muestras complejas 33
en Regresión logística de muestras complejas 37
categorías pronosticadas
en la regresión ordinal de muestras complejas 46
en Regresión logística de muestras complejas 40
chi-cuadrado
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
chi-cuadrado corregido
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
clústeres
en asistente de muestreo 5
en asistente de preparación del análisis 13
coeficiente de variación (CDV)
en Descriptivos de Muestras complejas 23
en Frecuencias de Muestras complejas 19
en Razones de Muestras complejas 29
en tablas cruzadas de Muestras complejas 25
contrastes
en Modelo lineal general de muestras complejas 33
contrastes de desviación
en Modelo lineal general de muestras complejas 33
contrastes de diferencia
en Modelo lineal general de muestras complejas 33
Contrastes de Helmert
en Modelo lineal general de muestras complejas 33
contrastes polinómicos
en Modelo lineal general de muestras complejas 33

contrastes repetidos
en Modelo lineal general de muestras complejas 33
contrastes simples
en Modelo lineal general de muestras complejas 33
Convergencia de la verosimilitud
en la regresión ordinal de muestras complejas 47
en Regresión logística de muestras complejas 41
convergencia de los parámetros
en la regresión ordinal de muestras complejas 47
en Regresión logística de muestras complejas 41
corrección de Bonferroni secuencial
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
corrección de Sidak
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
corrección de Sidak secuencial
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
correlaciones de estimaciones de parámetros
en la regresión ordinal de muestras complejas 44
en Modelo lineal general de muestras complejas 32
en Regresión logística de muestras complejas 38
covarianzas de estimaciones de parámetros
en la regresión ordinal de muestras complejas 44
en Modelo lineal general de muestras complejas 32
en Regresión logística de muestras complejas 38

D

Descriptivos de muestras complejas 23
estadísticos 23
valores perdidos 24
diferencia de riesgos
en tablas cruzadas de Muestras complejas 25
diferencia menos significativa
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54

E

efecto de diseño
en Descriptivos de Muestras complejas 23
en Frecuencias de Muestras complejas 19
en la regresión ordinal de muestras complejas 44
en Modelo lineal general de muestras complejas 32
en Razones de Muestras complejas 29
en regresión de Cox de muestras complejas 52
en Regresión logística de muestras complejas 38
en tablas cruzadas de Muestras complejas 25
error estándar
en Descriptivos de Muestras complejas 23
en Frecuencias de Muestras complejas 19
en la regresión ordinal de muestras complejas 44
en Modelo lineal general de muestras complejas 32
en Razones de Muestras complejas 29
en Regresión logística de muestras complejas 38
en tablas cruzadas de Muestras complejas 25
estadístico F
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
estadístico F corregido
en muestras complejas 33, 39, 45
en regresión de Cox de muestras complejas 54
estadísticos pseudo R2
en la regresión ordinal de muestras complejas 44
en Regresión logística de muestras complejas 38
estadísticos R2
en Modelo lineal general de muestras complejas 32
estimación de la muestra
en asistente de preparación del análisis 14
estimaciones de los parámetros
en la regresión ordinal de muestras complejas 44
en Modelo lineal general de muestras complejas 32
en regresión de Cox de muestras complejas 52
en Regresión logística de muestras complejas 38

- estratificación
 - en asistente de muestreo 5
 - en asistente de preparación del análisis 13
- estratos de referencia
 - en regresión de Cox de muestras complejas 51

F

- Frecuencias de muestras complejas 19
 - estadísticos 19

G

- grados de libertad
 - en muestras complejas 33, 39, 45
 - en regresión de Cox de muestras complejas 54

H

- historial de iteraciones
 - en la regresión ordinal de muestras complejas 47
 - en Regresión logística de muestras complejas 41

I

- información del diseño muestral
 - en regresión de Cox de muestras complejas 52
- intervalos de confianza
 - en Descriptivos de Muestras complejas 23
 - en Frecuencias de Muestras complejas 19
 - en la regresión ordinal de muestras complejas 44
 - en Modelo lineal general de muestras complejas 32, 34
 - en Razones de Muestras complejas 29
 - en Regresión logística de muestras complejas 38
 - en tablas cruzadas de Muestras complejas 25
- introducir ponderaciones muestrales
 - en asistente de muestreo 5
- iteraciones
 - en la regresión ordinal de muestras complejas 47
 - en Regresión logística de muestras complejas 41

M

- media
 - en Descriptivos de Muestras complejas 23
- medias marginales estimadas
 - en Modelo lineal general de muestras complejas 33

- medida del tamaño
 - en asistente de muestreo 6
- método de estimación de Breslow
 - en regresión de Cox de muestras complejas 57
- método de estimación de Efron
 - en regresión de Cox de muestras complejas 57
- método de muestreo
 - en asistente de muestreo 6
- método de muestreo de Brewer
 - en asistente de muestreo 6
- método de muestreo de Murthy
 - en asistente de muestreo 6
- método de muestreo de Sampford
 - en asistente de muestreo 6
- método de Newton-Raphson
 - en la regresión ordinal de muestras complejas 47
- Modelo lineal general de muestras complejas 31
 - almacenamiento de variables 34
 - Características adicionales del comando 35
 - estadísticos 32
 - medias estimadas 33
 - model 31
 - opciones 34
- Muestras complejas
 - contrastes de hipótesis 33, 39, 45
 - opciones 20, 24, 27, 30
 - valores perdidos 20, 26
- muestreo
 - diseño complejo 5
 - muestreo aleatorio simple
 - en asistente de muestreo 6
 - muestreo complejo
 - plan de análisis 13
 - plan de muestreo 5
 - muestreo de PPS
 - en asistente de muestreo 6
 - muestreo secuencial
 - en asistente de muestreo 6
 - muestreo sistemático
 - en asistente de muestreo 6

N

- nivel de confianza
 - en la regresión ordinal de muestras complejas 47
 - en Regresión logística de muestras complejas 41

P

- plan de análisis 13
- plan de muestreo 5
- ponderaciones muestrales
 - en asistente de muestreo 8
 - en asistente de preparación del análisis 13
- porcentajes de fila
 - en tablas cruzadas de Muestras complejas 25

- porcentajes de la columna
 - en tablas cruzadas de Muestras complejas 25
- porcentajes de tabla
 - en Frecuencias de Muestras complejas 19
 - en tablas cruzadas de Muestras complejas 25
- predictor dependiente del tiempo
 - en regresión de Cox de muestras complejas 51
- probabilidad pronosticada
 - en la regresión ordinal de muestras complejas 46
 - en Regresión logística de muestras complejas 40
- probabilidades acumuladas
 - en la regresión ordinal de muestras complejas 46
- probabilidades de inclusión
 - en asistente de muestreo 8
- probabilidades de respuesta
 - en la regresión ordinal de muestras complejas 44
- proporción muestral
 - en asistente de muestreo 8
- prueba de líneas paralelas
 - en la regresión ordinal de muestras complejas 44
- prueba de riesgos proporcionales
 - en regresión de Cox de muestras complejas 52
- Prueba t
 - en la regresión ordinal de muestras complejas 44
 - en Modelo lineal general de muestras complejas 32
 - en Regresión logística de muestras complejas 38
- Puntuación de Fisher
 - en la regresión ordinal de muestras complejas 47

R

- raíz cuadrada del efecto de diseño
 - en Descriptivos de Muestras complejas 23
 - en Frecuencias de Muestras complejas 19
 - en la regresión ordinal de muestras complejas 44
 - en Modelo lineal general de muestras complejas 32
 - en Razones de Muestras complejas 29
 - en regresión de Cox de muestras complejas 52
 - en Regresión logística de muestras complejas 38
 - en tablas cruzadas de Muestras complejas 25
- razones de las ventajas
 - en la regresión ordinal de muestras complejas 46
 - en Regresión logística de muestras complejas 40

- razones de las ventajas (*continuación*)
 - en tablas cruzadas de Muestras complejas 25
- Razones de muestras complejas 29
 - estadísticos 29
 - valores perdidos 30
- recuento no ponderado
 - en Descriptivos de Muestras complejas 23
 - en Frecuencias de Muestras complejas 19
 - en Razones de Muestras complejas 29
 - en tablas cruzadas de Muestras complejas 25
- Regresión de Cox de muestras complejas
 - almacenamiento de variables 54
 - análisis Kaplan-Meier 49
 - contrastes de hipótesis 54
 - definición de eventos 50
 - estadísticos 52
 - exportación del modelo 55
 - gráficos 53
 - model 52
 - opciones 57
 - predictor dependiente del tiempo 51
 - predictores 51
 - subgrupos 51
 - variables de fecha y hora 49
- Regresión logística de muestras complejas 37
 - almacenamiento de variables 40
 - Características adicionales del comando 41
 - categoría de referencia 37
 - estadísticos 38
 - model 38
 - opciones 41
 - razones de las ventajas 40
- Regresión ordinal de muestras complejas 43
 - almacenamiento de variables 46
 - estadísticos 44
 - model 44
 - opciones 47
 - probabilidades de respuesta 44
 - razones de las ventajas 46
- residuos
 - en Modelo lineal general de muestras complejas 34
 - en tablas cruzadas de Muestras complejas 25
- residuos agregados
 - en regresión de Cox de muestras complejas 54
- residuos corregidos
 - en tablas cruzadas de Muestras complejas 25
- residuos de Cox-Snell
 - en regresión de Cox de muestras complejas 54
- residuos de desviación
 - en regresión de Cox de muestras complejas 54
- residuos de Martingale
 - en regresión de Cox de muestras complejas 54

- residuos de puntuación
 - en regresión de Cox de muestras complejas 54
- residuos parciales de Schoenfeld
 - en regresión de Cox de muestras complejas 54
- riesgo relativo
 - en tablas cruzadas de Muestras complejas 25

S

- separación
 - en la regresión ordinal de muestras complejas 47
 - en Regresión logística de muestras complejas 41
- subdivisión por pasos
 - en la regresión ordinal de muestras complejas 47
 - en Regresión logística de muestras complejas 41
- subpoblación
 - en regresión de Cox de muestras complejas 51
- suma
 - en Descriptivos de Muestras complejas 23

T

- Tablas cruzadas de muestras complejas 25
 - estadísticos 25
- tablas de clasificación
 - en la regresión ordinal de muestras complejas 44
 - en Regresión logística de muestras complejas 38
- tamaño de la muestra
 - en asistente de muestreo 7, 8
- tamaño de la población
 - en asistente de muestreo 8
 - en Descriptivos de Muestras complejas 23
 - en Frecuencias de Muestras complejas 19
 - en Razones de Muestras complejas 29
 - en tablas cruzadas de Muestras complejas 25

V

- valores acumulados
 - en Frecuencias de Muestras complejas 19
- valores esperados
 - en tablas cruzadas de Muestras complejas 25
- valores perdidos
 - en Descriptivos de Muestras complejas 24
 - en la regresión ordinal de muestras complejas 47

- valores perdidos (*continuación*)
 - en Modelo lineal general de muestras complejas 34
 - en muestras complejas 20, 26
 - en Razones de Muestras complejas 30
 - en Regresión logística de muestras complejas 41
 - valores pronosticados
 - en Modelo lineal general de muestras complejas 34



Impreso en España