

IBM SPSS Missing Values 23

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos" en la página 23.

Información de 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 valores perdidos | 1 |
|--|----------|

| | |
|---|----------|
| Capítulo 2. Análisis de valores perdidos 3 | 3 |
|---|----------|

| | |
|---|---|
| Visualización de los patrones de los valores perdidos | 4 |
|---|---|

| | |
|--|---|
| Visualización de los estadísticos descriptivos de los valores perdidos | 5 |
|--|---|

| | |
|---|---|
| Estimación de los estadísticos e imputación de los valores perdidos | 6 |
|---|---|

| | |
|-------------------------------------|---|
| Opciones de estimación EM | 7 |
|-------------------------------------|---|

| | |
|---|---|
| Opciones de estimación de regresión | 8 |
|---|---|

| | |
|---|---|
| Variables pronosticadas y predictoras | 8 |
|---|---|

| | |
|---|---|
| Características adicionales del comando MVA | 9 |
|---|---|

| | |
|--|-----------|
| Capítulo 3. Imputación múltiple | 11 |
|--|-----------|

| | |
|-----------------------------|----|
| Analizar patrones | 12 |
|-----------------------------|----|

| | |
|-----------------------------------|----|
| Imputar valores perdidos. | 13 |
|-----------------------------------|----|

| | |
|-----------------|----|
| Método. | 14 |
|-----------------|----|

| | |
|-------------------------|----|
| Restricciones | 15 |
|-------------------------|----|

| | |
|---------------------|----|
| Resultados. | 16 |
|---------------------|----|

| | |
|--|--|
| Características adicionales del comando MULTIPLE | |
|--|--|

| | |
|----------------------|----|
| IMPUTATION | 17 |
|----------------------|----|

| | |
|--|----|
| Trabajo con datos de imputación múltiple | 17 |
|--|----|

| | |
|--|----|
| Análisis de datos de imputación múltiple | 18 |
|--|----|

| | |
|---|----|
| Opciones de imputación múltiple | 21 |
|---|----|

| | |
|-------------------------|-----------|
| Avisos | 23 |
|-------------------------|-----------|

| | |
|------------------------------|----|
| Marcas comerciales | 25 |
|------------------------------|----|

| | |
|-------------------------|-----------|
| Índice | 27 |
|-------------------------|-----------|

Capítulo 1. Introducción a valores perdidos

Los casos con valores perdidos representan un reto importante, ya que los procedimientos de modelado tradicionales simplemente descartan estos casos para el análisis. Cuando hay pocos valores perdidos (aproximadamente, menos del 5 % del número total de casos) y dichos valores pueden considerarse perdidos de forma aleatoria (es decir, que la pérdida de un valor no depende de otros valores), entonces el método tradicional de eliminación según la lista es relativamente "seguro". La opción Valores perdidos puede ayudarle a determinar si la eliminación según la lista es suficiente; asimismo, proporciona métodos para gestionar los valores perdidos cuando no lo sea.

Análisis de valores perdidos frente a procedimientos de imputación múltiple

La opción Valores perdidos proporciona dos conjuntos de procedimientos para gestionar los valores perdidos:

- Los procedimientos de Imputación múltiple proporcionan un análisis de los patrones de datos perdidos, dirigidos a una imputación múltiple de valores perdidos. Esto es, se producen versiones múltiples del conjunto de datos, cada una con su propio conjunto de valores imputados. Cuando se realizan análisis estadísticos, se combinan las estimaciones de los parámetros de todos los conjuntos de datos imputados, con lo que se ofrecen estimaciones generalmente más precisas de lo que serían con sólo una imputación.
- Análisis de valores perdidos proporciona un conjunto ligeramente diferente de herramientas descriptivas para analizar los datos perdidos (en especial, la prueba MCAR de Little) e incluye una variedad de métodos de imputación individual. Tenga en cuenta que por lo general la imputación múltiple suele considerarse superior a la imputación individual.

Tareas de valores perdidos

Puede empezar con el análisis de valores perdidos siguiendo estos pasos básicos:

1. **Examinar la ausencia.** Utilice Análisis de valores perdidos y Analizar patrones para explorar patrones de valores perdidos en sus datos y determinar si es necesario recurrir a la imputación múltiple.
2. **Imputar valores perdidos.** Utilice Imputar valores perdidos para imputar de forma múltiple los valores perdidos.
3. **Analizar datos "completos".** Utilice cualquier procedimiento que admita datos de imputación múltiple. Consulte "Análisis de datos de imputación múltiple" en la página 18 para obtener más información sobre el análisis de conjuntos de datos de imputación múltiple y una lista de procedimientos que admiten dichos datos.

Capítulo 2. Análisis de valores perdidos

El procedimiento Análisis de valores perdidos realiza tres funciones principales:

- Describe el patrón de los datos perdidos. ¿Dónde se encuentran los valores perdidos? ¿Con qué frecuencia aparecen? ¿Hay pares de variables que tienden a tener valores perdidos en varios casos? ¿Son los valores de los datos extremos? ¿Están los valores perdidos de forma aleatoria?
- Estimar las medias, desviaciones estándar, covarianzas y correlaciones para los diferentes métodos de valores perdidos: por lista, por parejas, regresión o EM (maximización esperada). El método por parejas muestra, además, recuentos de los casos completos por parejas.
- Rellena (imputa) los valores perdidos con valores estimados utilizando el método EM o el de regresión; sin embargo, por lo general se considera que la imputación múltiple proporciona resultados más precisos.

El análisis de valores perdidos ayuda a resolver varios problemas ocasionados por los datos incompletos. Si los casos con valores perdidos son sistemáticamente diferentes de los casos sin valores perdidos, los resultados pueden ser equívocos. Además, los datos perdidos pueden reducir la precisión de los estadísticos calculados, porque no se dispone de tanta información como originalmente se pensaba. Otro problema radica en que los supuestos subyacentes a muchos procedimientos estadísticos se basan en casos completos y los valores perdidos pueden complicar la teoría exigida.

Ejemplo. En la evaluación de un tratamiento contra la leucemia se miden diversas variables. Sin embargo, no todas las mediciones se encuentran disponibles para todos los pacientes. Los patrones de los datos perdidos se inspeccionan, se tabulan y se consideran aleatorios. Se utiliza un análisis EM para estimar las medias, las correlaciones y las covarianzas. También se utiliza para determinar que los datos están perdidos completamente al azar. A continuación, los valores perdidos se reemplazan por los valores imputados y se guardan en un nuevo archivo de datos para análisis posteriores.

Estadísticos. Estadísticos univariados, incluido el número de valores no perdidos, media, desviación estándar, número de valores perdidos y número de valores extremos. Medias estimadas, matriz de covarianza y matriz de correlaciones, utilizando los métodos de regresión, EM, por lista o por parejas. Prueba MCAR de Little con resultados EM. Resumen de medias a través de varios métodos. Para los grupos definidos por valores perdidos frente a valores no perdidos: pruebas *t*. Para todas las variables: los patrones de valores perdidos representados como casos respecto a variables.

Consideraciones de los datos

Datos. Los datos pueden ser categóricos o cuantitativos (de escala o continuos). Sin embargo, puede estimar los estadísticos e imputar los datos perdidos únicamente en el caso de variables cuantitativas. Para cada variable, los valores perdidos que no están codificados como valores perdidos del sistema deben definirse como valores perdidos del usuario. Por ejemplo, si un elemento del cuestionario tiene la respuesta *No sabe* codificada como 5 y desea tratarlo como valor perdido, el elemento debería tener el 5 codificado como valor perdido del usuario.

Ponderaciones de frecuencia. Este procedimiento respeta las ponderaciones de frecuencia (réplica). Los casos con un valor negativo o cero de valor de ponderación de réplica se ignoran. Las ponderaciones no enteras se truncan.

Supuestos. La estimación por lista, por parejas y mediante regresión depende del supuesto de que el patrón de valores perdidos no depende de los valores de los datos (esta condición se conoce como perdidos completamente al azar o MCAR). (Esta condición se conoce como **perdida completamente al azar** o MCAR). Por tanto, todos los métodos (incluido el método EM) de estimación ofrecen estimaciones coherentes y no sesgadas de las correlaciones y las covarianzas cuando los datos son MCAR. El

incumplimiento del supuesto MCAR puede dar lugar a estimaciones sesgadas producidas por los métodos de regresión, por lista o por parejas. Si los datos no son MCAR, es necesario utilizar la estimación EM.

La estimación EM depende del supuesto de que el patrón de los datos perdidos está relacionado únicamente con los datos observados. (Esta condición se denomina **perdidos al azar** o MAR). Este supuesto permite ajustar las estimaciones utilizando la información disponible. Por ejemplo, en un estudio sobre la educación y los ingresos, los sujetos con un menor nivel educativo pueden tener más valores perdidos de ingresos. En este caso, los datos son MAR, no MCAR. Es decir, para MAR, la probabilidad de que se registren los ingresos depende del nivel educativo del sujeto. La probabilidad puede variar según el nivel educativo pero no según los ingresos *dentro de ese nivel educativo*. Si la probabilidad de que se registre el ingreso también depende del valor de los ingresos dentro de cada nivel educativo (por ejemplo, las personas con ingresos elevados no los declara), los datos no serán ni MCAR ni MAR. Se trata de una situación poco habitual y, si se produce, no hay ningún método adecuado.

Procedimientos relacionados. Muchos procedimientos permiten utilizar la estimación por lista o por parejas. Regresión lineal y Análisis factorial permiten reemplazar los valores perdidos por los valores de las medias. El módulo adicional Previsiones ofrece varios métodos para reemplazar los valores perdidos en las series temporales.

Para obtener un análisis de valores perdidos

1. Seleccione en los menús:
Analizar > Análisis de valores perdidos...
2. Seleccione al menos una variable cuantitativa (de escala) para estimar los estadísticos y, si lo desea, imputar los valores perdidos.

Si lo desea, puede:

- Seleccionar variables categóricas (numéricas o de cadena) y establecer un límite para el número de categorías (**N.º máximo de categorías**).
- Pulse en **Patrones** para tabular los patrones de los datos perdidos. Consulte el tema “Visualización de los patrones de los valores perdidos” para obtener más información.
- Pulse en **Descriptivos** para mostrar los estadísticos descriptivos de los valores perdidos. Consulte el tema “Visualización de los estadísticos descriptivos de los valores perdidos” en la página 5 para obtener más información.
- Seleccione un método para estimar los estadísticos (medias, covarianzas y correlaciones) y posiblemente imputar los valores perdidos. Consulte el tema “Estimación de los estadísticos e imputación de los valores perdidos” en la página 6 para obtener más información.
- Si selecciona EM o Regresión, pulsar en **Variables** para especificar el subconjunto que se va a utilizar para la estimación. Consulte el tema “Variables pronosticadas y predictoras” en la página 8 para obtener más información.
- Seleccione una variable de etiqueta de caso. Esta variable se utiliza para etiquetar los casos en las tablas de patrones que muestran los casos individuales.

Visualización de los patrones de los valores perdidos

Si lo desea, puede consultar varias tablas que muestran los patrones y el alcance de los datos perdidos. Estas tablas pueden ayudarle a identificar:

- Dónde se encuentran los valores perdidos
- Si hay pares de variables que tienden a tener valores perdidos en casos individuales
- Si los valores de los datos son extremos

Representación

Hay tres tipos de tablas disponibles para ver los patrones de los datos perdidos.

Casos tabulados. Se tabulan los patrones de los valores perdidos en las variables de análisis y se muestran las frecuencias de cada patrón. Utilice **Ordenar variables según patrón de valores perdidos** para especificar si los recuentos y las variables se ordenan según la similitud de los patrones. Utilice **Omitir patrones con menos del n % de los casos** para eliminar los patrones que aparecen con poca frecuencia.

Casos con valores perdidos. Cada caso con un valor perdido o extremo se tabula para cada variable de análisis. Utilice **Ordenar variables según patrón de valores perdidos** para especificar si los recuentos y las variables se ordenan según la similitud de los patrones.

Todos los casos. Se tabula cada caso y se indican los valores perdidos y extremos para cada variable. Los casos se enumeran en el orden en que aparecen en el archivo de datos, a menos que se especifique una variable en **Ordenar por**.

En las tablas que muestran los casos individuales, se utilizan los siguientes símbolos:

- + . Valor extremadamente alto
- . Valor extremadamente bajo
- S. Valor perdido del sistema
- A. Primer tipo de valor perdido del usuario
- B. Segundo tipo de valor perdido del usuario
- C. Tercer tipo de valor perdido del usuario

Variables

Puede mostrar información adicional acerca de las variables que se incluyen en el análisis. Las variables que se añadan a **Información adicional acerca de** aparecerán individualmente en la tabla de patrones perdidos. Para las variables cuantitativas (de escala), se muestra la media; para las variables categóricas, se muestra el número de casos que presentan el patrón en cada categoría.

- **Ordenar por.** Los casos se listan según el orden ascendente o descendente de los valores de la variable especificada. Esta opción está disponible sólo si se selecciona **Todos los casos**.

Para mostrar los patrones de los valores perdidos

1. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos patrones de valores perdidos desea ver.
2. Pulse en **Patrones**.
3. Seleccione las tablas de patrones que desea ver.

Visualización de los estadísticos descriptivos de los valores perdidos

Estadísticos univariantes

Los estadísticos univariados pueden ayudarle a identificar el alcance general de los datos perdidos. Para cada variable, se muestran los siguientes datos:

- Número de valores no perdidos
- Número y porcentaje de valores perdidos

Para las variables cuantitativas (de escala), también se muestran los siguientes datos:

- Media
- Desviación estándar
- Número de valores extremadamente altos o bajos

Estadísticos para las variables indicadoras

Para cada variable, se crea una variable de indicador. Esta variable categórica indica si la variable está presente o perdida en un determinado caso. Las variables de indicador se utilizan para crear la discordancia, la prueba t y las tablas de frecuencia.

Porcentaje de discordancia. Para cada par de variables muestra el porcentaje de casos en los que una variable tiene un valor perdido y la otra variable tiene un valor no perdido. Cada elemento diagonal de la tabla contiene el porcentaje de valores perdidos para una sola variable.

Pruebas t con los grupos formados por las variables de indicador. Se comparan las medias de los dos grupos para cada variable cuantitativa, utilizando el estadístico t de Student. Los grupos especifican si una variable está presente o perdida. Se muestra el estadístico t , los grados de libertad, los recuentos de valores perdidos y no perdidos y las medias de los dos grupos. También se pueden mostrar todas las probabilidades bilaterales asociadas con el estadístico t . Si el análisis genera más de una prueba, no utilice estas probabilidades para contrastar la significación. Estas probabilidades sólo son adecuadas cuando se calcula una única prueba.

Tabulaciones cruzadas de variables categóricas y de indicador. Para cada variable categórica se muestra una tabla. Para cada categoría, la tabla muestra la frecuencia y el porcentaje de los valores no perdidos para las demás variables. También se muestran los porcentajes de cada tipo de valor perdido.

Omitir variables con menos valores perdidos que el n % de los casos. Para reducir el tamaño de la tabla puede omitir los estadísticos que se calculen sólo para un pequeño número de casos.

Para mostrar los estadísticos descriptivos

1. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos estadísticos descriptivos de los valores perdidos desea ver.
2. Pulse en **Descriptivos**.
3. Elija los estadísticos descriptivos que desea que aparezcan.

Estimación de los estadísticos e imputación de los valores perdidos

Puede elegir que se estimen las medias, desviaciones estándar, covarianzas y correlaciones utilizando un método por lista (sólo casos completos), por parejas, EM (maximización esperada) y/o de regresión. También puede elegir imputar los valores perdidos (estimar los valores de sustitución). Tenga en cuenta que, por lo general, la Imputación múltiple suele considerarse superior a la imputación individual para solucionar el problema de los valores perdidos. La prueba MCAR de Little sigue siendo útil para determinar si la imputación es necesaria.

Método por lista

Este método únicamente utiliza los casos completos. Si alguna de las variables de análisis tiene valores perdidos, se omite dicho caso de los cálculos.

Método por parejas

Este método examina las parejas de variables del análisis y utiliza un caso únicamente si tiene valores no perdidos para ambas variables. Las frecuencias, medias y desviaciones estándar se calculan por separado

para cada pareja. Como se ignoran los demás valores perdidos del caso, las correlaciones y las covarianzas de las dos variables no dependen de los valores perdidos de ninguna otra variable.

Método EM

Este método supone que los datos parcialmente perdidos siguen una distribución determinada y basa las inferencias en la probabilidad según dicha distribución. Cada iteración se compone de un paso E y un paso M. El paso E determina la esperanza condicional de los datos "perdidos", teniendo en cuenta los valores observados y las estimaciones actuales de los parámetros. A continuación, se sustituyen estas esperanzas por los datos "perdidos". En el paso M, se calculan las estimaciones de máxima verosimilitud de los parámetros como si se hubieran rellenado los datos perdidos. Se especifica "perdidos" entre comillas ya que los valores perdidos no se rellenan directamente, sino que en su lugar se utilizan funciones de ellos en el log verosimilitud.

El estadístico de chi cuadrado de Roderick J. A. Little para contrastar si los valores están perdidos completamente al azar (MCAR) se imprime como nota al pie de las matrices de EM. Para este contraste, la hipótesis nula es que los datos están perdidos completamente al azar y el valor p es significativo al nivel 0,05. Si el valor es inferior a 0,05, los datos no están perdidos completamente al azar. Los datos pueden estar perdidos al azar (MAR) o no perdidos al azar (NMAR). No se puede suponer la situación en la que se encuentran los datos perdidos, por lo que es necesario analizar los datos para determinar de qué manera están perdidos.

Método de regresión

Este método calcula las estimaciones de regresión lineal múltiple y ofrece opciones que permiten incrementar las estimaciones con componentes aleatorios. Para cada valor predicho, el procedimiento puede añadir un residuo de un caso completo seleccionado de manera aleatoria, una desviación normal aleatoria o una desviación aleatoria (escalada por la raíz cuadrada del residuo cuadrático promedio) de la distribución t .

Opciones de estimación EM

Utilizando un proceso iterativo, el método EM estima las medias, la matriz de covarianzas y la correlación de las variables cuantitativas (de escala) con los valores perdidos.

Distribución. EM realiza las inferencias basándose en la verosimilitud según la distribución especificada. De forma predeterminada, se supone una distribución normal. Si sabe que las colas de la distribución son más largas que las de una distribución normal, puede solicitar que el procedimiento construya la función de verosimilitud a partir de una distribución t de Student con n grados de libertad. La distribución normal mixta también proporciona una distribución con colas más largas. Especifique la razón de las desviaciones estándar de la distribución normal mixta y la proporción de mezcla de las dos distribuciones. La distribución normal mixta supone que únicamente difieren las desviaciones estándar de las distribuciones. Las medias deben ser iguales.

Número máximo de iteraciones. Establece el número máximo de iteraciones para estimar la covarianza auténtica. El procedimiento se detiene cuando se alcanza el número de iteraciones, incluso si no han convergido las estimaciones.

Guardar datos completados. Puede guardar un conjunto de datos con los valores imputados en el lugar de los valores perdidos. No obstante, tenga en cuenta que los estadísticos basados en la covarianza que utilicen los valores imputados estimarán valores de los parámetros menores que los reales. El grado en que esta estimación es inferior a la real es proporcional al número de casos que no se observaron conjuntamente.

Para especificar las opciones de EM

1. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos valores perdidos desea estimar utilizando el método EM.
2. Seleccione **EM** en el grupo Estimación.
3. Para especificar las variables predictoras y pronosticadas, pulse en **Variables**. Consulte el tema “Variables pronosticadas y predictoras” para obtener más información.
4. Pulse en **EM**.
5. Seleccione las opciones de EM que desee.

Opciones de estimación de regresión

El método de regresión estima los valores perdidos utilizando la regresión lineal múltiple. Se muestran las medias, la matriz de covarianza y la matriz de correlaciones de las variables pronosticadas.

Corrección de la estimación. El método de regresión puede añadir un componente aleatorio a las estimaciones de regresión. Puede seleccionar residuos, variantes normales, variantes *t* de Student o sin corrección.

- *Residuos.* Los términos de error se eligen al azar de entre los residuos observados en los casos completos, para añadirlos a las estimaciones de regresión.
- *Variantes normales.* Los términos de error se escogen al azar de una distribución con valor esperado 0 y desviación estándar igual a la raíz cuadrada del término media cuadrática de los errores de la regresión.
- *Variantes *t* de Student.* Los términos de error se escogen al azar de una distribución *t* con los grados de libertad especificados y se escalan según la raíz de la media cuadrática de los errores (RMSE).

Número máximo de predictores. Establece un límite máximo para el número de variables predictoras (independientes) utilizadas en el proceso de estimación.

Guardar datos completados. Escribe un conjunto de datos en la sesión actual o en un archivo de datos externo con formato IBM® SPSS Statistics, reemplazando los valores perdidos por los valores estimados mediante el método de regresión.

Para especificar las opciones de regresión

1. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos valores perdidos desea estimar utilizando el método de regresión.
2. Seleccione **Regresión** en el grupo Estimación.
3. Para especificar las variables predictoras y pronosticadas, pulse en **Variables**. Consulte el tema “Variables pronosticadas y predictoras” para obtener más información.
4. Pulse en **Regresión**.
5. Seleccione las opciones de regresión que desee.

Variables pronosticadas y predictoras

De forma predeterminada, se utilizan todas las variables cuantitativas para la estimación de regresión y EM. Si es necesario, puede especificar que determinadas variables se utilicen como variables pronosticadas o variables predictoras en las estimaciones. Una determinada variable puede aparecer en ambas listas, pero hay situaciones en las que quizá quiera restringir el uso de una variable. Por ejemplo, a algunos analistas no les resulta cómodo estimar los valores de las variables de resultados. También es posible que quiera utilizar variables diferentes en estimaciones distintas y ejecutar el procedimiento varias veces. Por ejemplo, si tiene un conjunto de elementos que son valoraciones de enfermeras y otro conjunto que son valoraciones de médicos, tal vez quiera ejecutar el procedimiento una vez utilizando el elemento de las enfermeras para estimar los elementos de las enfermeras y otra vez para estimar los elementos de los médicos.

También hay que hacer otra consideración al utilizar el método de regresión. En la regresión múltiple, el uso de un subconjunto grande de variables independientes puede generar valores pronosticados de peor calidad que los que generaría un subconjunto más pequeño. Por tanto, para que se utilice una variable, debe alcanzar un límite de F para entrar de 4,0. Este límite se puede cambiar utilizando la sintaxis.

Para especificar las variables pronosticadas y predictoras

1. En el cuadro de diálogo principal Análisis de valores perdidos, seleccione las variables cuyos valores perdidos desea estimar utilizando el método de regresión.
2. Seleccione **EM** o **Regresión** en el grupo Estimación.
3. Pulse en **Variables**.
4. Si desea utilizar determinadas variables, en vez de todas, como variables pronosticadas y variables predictoras, elija **Seleccionar variables** y mueva las variables a las listas adecuadas.

Características adicionales del comando MVA

La sintaxis de comandos también le permite:

- Especificar distintas variables descriptivas para los patrones de valores perdidos, los patrones de los datos y los patrones tabulados, mediante la palabra clave DESCRIBE en los subcomandos MPATTERN, DPATTERN o TPATTERN.
- Especificar más de una variable de ordenación para la tabla de patrones de los datos, utilizando el subcomando DPATTERN.
- Especificar más de una variable de ordenación para los patrones de los datos, utilizando el subcomando DPATTERN.
- Especificar la tolerancia y la convergencia mediante el subcomando EM.
- Especifique la tolerancia y la F para entrar mediante el subcomando REGRESSION.
- Especificar diferentes listas de variables para EM y para Regresión, con los subcomandos EM y REGRESSION.
- Especificar diferentes porcentajes para suprimir los casos mostrados para TTESTS, TABULATE y MISMATCH.

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

Capítulo 3. Imputación múltiple

El objetivo de la imputación múltiple es generar valores posibles para los valores perdidos, creando así varios conjuntos de datos "completos". Los procedimientos analíticos que trabajan con conjuntos de datos de imputación múltiple producen resultados para cada conjunto de datos "completo", además de resultados combinados que estiman cuáles habrían sido los resultados si el conjunto de datos original no tuviera valores perdidos. Estos resultados combinados suelen ser más precisos que los proporcionados por métodos de imputación individual.

Consideraciones sobre los datos de imputación múltiple

VARIABLES DE ANÁLISIS. Las variables de análisis pueden ser:

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

El procedimiento supone que se ha asignado el nivel de medición adecuado a todas las variables. No obstante, puede cambiar temporalmente el nivel de medición para una variable pulsando con el botón derecho en la variable en la lista de variables de origen y seleccionar un nivel de medición en el menú emergente. Para modificar permanentemente el nivel de medición de una variable,

Un icono situado junto a cada variable de la lista de variables identifica el nivel de medición y el tipo de datos.

Tabla 1. Iconos de nivel de medición

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

Ponderaciones de frecuencia. Este procedimiento respeta las ponderaciones de frecuencia (réplica). Los casos con un valor negativo o cero de valor de ponderación de réplica se ignoran. Las ponderaciones no enteras se redondean al número entero más cercano.

Ponderación de análisis. Las ponderaciones de análisis (regresión o muestreo) se incorporan en resúmenes de valores perdidos y en modelos de imputación que se ajusten. Los casos de ponderaciones de análisis con valor negativo o cero se excluirán.

Muestras complejas. El procedimiento de Imputación múltiple no trata explícitamente los estratos, agrupaciones u otras estructuras de muestreo complejas, aunque puede aceptar ponderaciones de muestreo finales en la forma del análisis de la variable de ponderación. Tenga también en cuenta que los procedimientos de muestreos complejos actualmente no analizan de forma automática varios conjuntos de datos imputados. Para obtener una lista completa de procedimientos que admiten la combinación, consulte “Análisis de datos de imputación múltiple” en la página 18.

Valores perdidos. Los valores perdidos tanto del usuario como del sistema se consideran valores no válidos; es decir, ambos tipos de valores perdidos se sustituyen cuando se imputan los valores y los dos se consideran valores no válidos de variables utilizadas como predictores de modelos de imputación. Los valores perdidos del usuario y del sistema también se consideran perdidos en los análisis de valores perdidos.

Replicación de los resultados (Imputar valores perdidos). Si desea replicar exactamente los resultados de imputación, utilice el mismo valor de inicialización para el generador de números aleatorios, el mismo orden de datos y el mismo orden de variables, además de utilizar la misma configuración del procedimiento.

- **Generación de números aleatorios.** El procedimiento utiliza la generación de números aleatorios durante el cálculo de valores imputados. Para reproducir los mismos resultados aleatorios en el futuro, utilice el mismo valor de inicialización para el generador de números aleatorios antes de cada ejecución del procedimiento Imputar valores perdidos.
- **Orden de casos.** Los valores se imputan en el orden de casos.
- **Orden de las variables.** El método de imputación de especificación totalmente condicional imputa los valores en el orden especificado en la lista Variables de análisis.

Existen dos cuadros de diálogo dedicados a la imputación múltiple.

- Analizar patrones proporciona medidas descriptivas de los patrones de valores perdidos en los datos y puede resultar útil como paso exploratorio antes de la imputación.
- Imputar valores perdidos se utiliza para generar imputaciones múltiples. Los conjuntos de datos completos pueden analizarse con procedimientos que admiten conjuntos de datos de imputación múltiple. Consulte “Análisis de datos de imputación múltiple” en la página 18 para obtener más información sobre el análisis de conjuntos de datos de imputación múltiple y una lista de procedimientos que admiten dichos datos.

Analizar patrones

Analizar patrones proporciona medidas descriptivas de los patrones de valores perdidos en los datos y puede resultar útil como paso exploratorio antes de la imputación.

Ejemplo. Un proveedor de telecomunicaciones desea comprender mejor los patrones de uso de servicio en su base de datos de clientes. Tienen datos completos de los servicios utilizados por sus clientes, pero la información demográfica recopilada por la empresa tiene diferentes valores perdidos. El análisis de patrones de valores perdidos puede ayudar a determinar los siguientes pasos que se imputarán. Consulte el tema para obtener más información.

Seleccione en los menús:

Analizar > Imputación múltiple > Analizar patrones...

1. Seleccione al menos dos variables de análisis. El procedimiento analiza patrones de datos perdidos en estas variables.

Configuración opcional

Ponderación de análisis. Esta variable contiene ponderaciones de análisis (regresión o muestra). El procedimiento incorpora ponderaciones de análisis en resúmenes de valores perdidos. Los casos de ponderaciones de análisis con valor negativo o cero se excluirán.

Resultado. Los siguientes resultados opcionales están disponibles:

- **Resumen de valores perdidos.** Esto muestra un gráfico circular con paneles que indica el número y el porcentaje de variables de análisis, casos o datos individuales que tengan uno o más valores perdidos.
- **Patrones de valores perdidos.** Esto muestra patrones tabulados de valores perdidos. Cada patrón se corresponde con un grupo de casos con el mismo patrón de datos completos e incompletos sobre variables de análisis. Puede utilizar este resultado para determinar si puede utilizar el método de imputación monotónica para sus datos o, si no, en qué medida se aproximan sus datos a un patrón monotónico. El procedimiento ordena las variables de análisis para revelar o aproximarse a un patrón monotónico. Si no hay patrones que no sean monotónicos después de la reordenación, puede llegar a la conclusión de que los datos tienen un patrón monotónico cuando las variables de análisis se ordenan de tal forma.
- **Variables con la mayor frecuencia de valores perdidos.** Esto muestra una tabla de variables de análisis ordenadas por el porcentaje de valores perdidos en orden descendente. La tabla incluye estadísticos descriptivos (media y desviación estándar) para variables de escala.

Puede controlar el número máximo de variables que se mostrará y el porcentaje de ausencia mínimo de una variable para que se incluya en la visualización. Se muestra el conjunto de variables que cumplen ambos criterios. Por ejemplo, si establece el número máximo de variables como 50 y el porcentaje de ausencia mínimo como 25, hará que la tabla muestre un máximo de 50 variables que tengan un mínimo del 25 % de valores perdidos. Si hay 60 variables de análisis pero sólo 15 tienen un porcentaje igual o mayor al 25 % de valores perdidos, el resultado sólo incluirá 15 variables.

Imputar valores perdidos

Imputar valores perdidos se utiliza para generar imputaciones múltiples. Los conjuntos de datos completos pueden analizarse con procedimientos que admiten conjuntos de datos de imputación múltiple. Consulte “Análisis de datos de imputación múltiple” en la página 18 para obtener más información sobre el análisis de conjuntos de datos de imputación múltiple y una lista de procedimientos que admiten dichos datos.

Ejemplo. Un proveedor de telecomunicaciones desea comprender mejor los patrones de uso de servicio en su base de datos de clientes. Tienen datos completos de los servicios utilizados por sus clientes, pero la información demográfica recopilada por la empresa tiene diferentes valores perdidos. Además, estos valores no están perdidos de forma aleatoria, por lo que se utilizará la imputación múltiple para completar el conjunto de datos. Consulte el tema para obtener más información.

Seleccione en los menús:

Analizar > Imputación múltiple > Imputar valores de datos perdidos...

1. Elija al menos dos variables en el modelo de imputación. El procedimiento imputa valores múltiples para los datos perdidos de estas variables.
2. Especifique el número de imputaciones que deben calcularse. Este valor es 5 de forma predeterminada.
3. Especifique un conjunto de datos o archivo de datos con formato IBM SPSS Statistics en el que se escribirán los datos imputados.

El conjunto de datos de resultados consiste en los datos de casos originales con datos perdidos más un conjunto de casos con valores imputados para cada imputación. Por ejemplo, si el conjunto de datos original tiene 100 casos y usted tiene cinco imputaciones, el conjunto de datos de resultados contendrá 600 casos. Todas las variables del conjunto de datos de entrada se incluyen en el conjunto de datos de salida. Las propiedades de diccionario (nombres, etiquetas, etc.) de las variables existentes

se copian en el nuevo conjunto de datos. El archivo también contiene una nueva variable, *Imputation_*, una variable numérica que indica la imputación (0 para datos originales o 1..*n* para casos con valores imputados).

El procedimiento define automáticamente la variable *Imputation_* como una variable de segmentación cuando se crea el conjunto de datos de resultados. Si las divisiones están activadas cuando se ejecuta el procedimiento, el conjunto de datos de resultados incluye un conjunto de imputaciones por cada combinación de valores de variables de segmentación.

Configuración opcional

Ponderación de análisis. Esta variable contiene ponderaciones de análisis (regresión o muestra). El procedimiento incorpora ponderaciones de análisis en modelos de regresión y clasificación utilizados para imputar valores perdidos. Las ponderaciones de análisis también se utilizan en resúmenes de valores imputados; por ejemplo, media, desviación estándar y error estándar. Los casos de ponderaciones de análisis con valor negativo o cero se excluirán.

Campos con un nivel de medición desconocido

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

Explorar datos. Lee los datos del conjunto de datos activo y asigna el nivel de medición predefinido en cualquier campo con un nivel de medición desconocido. Si el conjunto de datos es grande, puede llevar algún tiempo.

Asignar manualmente. Abre un cuadro de diálogo que contiene todos los campos con un nivel de medición desconocido. Puede utilizar este cuadro de diálogo para asignar el nivel de medición a esos campos. También puede asignar un nivel de medición en la Vista de variables del Editor de datos.

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

Método

La pestaña Método especifica la forma en la que se imputarán los valores perdidos, incluidos los tipos de modelos utilizados. Los predictores categóricos están codificados con indicadores (dummy).

Método de imputación. El método **Automático** explora los datos y utiliza el método monotónico si los datos muestran un patrón monotónico de valores perdidos; de lo contrario, se utiliza la especificación totalmente condicional. Si está seguro de qué método desea utilizar, puede especificarlo como un método **Personalizado**.

- **Especificación totalmente condicional.** Éste es un método de Monte Carlo y cadenas de Markov (MCMC) iterativo que puede utilizarse cuando el patrón de datos perdidos es arbitrario (monotónico o no monotónico).

El método de especificación totalmente condicional (FCS) ajusta un modelo univariante (variable dependiente simple) para cada iteración y variable en el orden especificado en la lista de variables utilizando como predictores todas las demás variables disponibles en el modelo para luego imputar los valores perdidos de las variables que se están ajustando. El método continúa hasta que se alcanza el número máximo de iteraciones y los valores imputados en la máxima iteración se guardan en el conjunto de datos imputado.

Número máximo de iteraciones. Esto especifica el número de iteraciones, o "pasos", realizadas por la cadena de Markov utilizada por el método de especificación totalmente condicional. Si el método de especificación totalmente condicional se seleccionó automáticamente, utilizará el número predeterminado de 10 iteraciones. Cuando selecciona la especificación totalmente condicional de

manera explícita, puede especificar un número personalizado de iteraciones. Puede que deba aumentar el número de iteraciones si la cadena de Markov no ha convergido. En la pestaña Resultados, puede guardar los datos de historial de iteraciones de especificación totalmente condicional y realizar un gráfico de los mismos para evaluar la convergencia.

- **Monotónico.** Éste es un método no iterativo que sólo puede utilizarse cuando los datos tienen un patrón monotónico de valores perdidos. Existe un patrón monotónico cuando puede ordenar las variables de tal forma que, si una variable tiene un valor no perdidos, todas las variables precedentes también tienen valores no perdidos. Al especificar que se trata de un método **Personalizado**, asegúrese de especificar las variables en la lista y ordenar que muestre un patrón monotónico.

El método monotónico ajusta un modelo univariante (variable dependiente simple) para cada variable del orden monotónico utilizando como predictores todas las variables anteriores, para luego imputar los valores perdidos de las variables que se están ajustando. Estos valores imputados se guardan en el conjunto de datos imputado.

Incluir interacciones bidimensionales. Cuando el método de imputación se selecciona automáticamente, el modelo de imputación de cada variable incluye un término constante y los efectos principales de las variables predictoras. Al seleccionar un método específico, puede incluir opcionalmente todas las interacciones bidimensionales posibles entre las variables predictoras categóricas.

Tipo de modelo para variables de escala. Cuando el método de imputación se selecciona automáticamente, la regresión lineal se utiliza como modelo univariante para variables de escala. Al seleccionar un método específico, también puede seleccionar alternativamente equivalencia de media predictiva como modelo para variables de escala. La equivalencia de media predictiva es una variante de la regresión lineal que iguala los valores imputados calculados por el modelo de regresión con el valor observado más cercano.

La regresión logística siempre se utiliza como modelo univariante para variables categóricas. Los predictores categóricos están codificados con indicadores (dummy), independientemente del tipo de modelo.

Tolerancia para la singularidad. Las matrices singulares (que no se pueden invertir) tienen columnas linealmente dependientes, lo que causará graves problemas al algoritmo de estimación. Incluso las matrices casi singulares pueden generar resultados deficientes, por lo que el procedimiento tratará una matriz cuyo determinante es menor que la tolerancia como singular. Especifique un valor positivo.

Restricciones

La pestaña Restricciones le permite restringir el papel de una variable durante la imputación y restringir el rango de valores imputados de una variable de escala de modo que sean convincentes. Además, puede restringir el análisis a variables con menos de un porcentaje máximo de valores perdidos.

Exploración de datos para resumen de variables. Al pulsar en **Explorar datos** la lista muestra variables de análisis y el porcentaje observado de ausencia, mínimo y máximo para cada una. Los resúmenes pueden basarse en todos los casos o limitarse a una exploración de los primeros n casos, como aparece especificado en el cuadro de texto Casos. Puede actualizar los resúmenes de distribución al pulsar en **Volver a explorar datos**.

Defina las restricciones

- **Papel.** Esto le permite personalizar el conjunto de variables que deben imputarse y/o tratarse como predictores. Normalmente, cada variable de análisis se considera tanto dependiente como predictora en el modelo de imputación. El **Papel** puede utilizarse para desactivar la imputación de variables que desee **Utilizar sólo como predictor** o para excluir el uso de variables como predictores (**Sólo imputar**) y, por lo tanto, hacer que el modelo de predicción sea más compacto. Ésta es la única restricción que puede especificarse para variables categóricas o para variables que se utilizan sólo como predictores.

- **Mín. y Máx.** Estas columnas le permiten especificar los valores mínimos y máximos imputados que se permiten para las variables de escala. Si un valor imputado se sale del rango, el procedimiento extrae otro valor hasta que encuentra uno dentro del rango o se alcanza al número máximo de extracciones (consulte **Máximo de extracciones** a continuación). Estas columnas sólo están disponibles si se selecciona **Regresión lineal** como el tipo de modelo de variable de escala en la pestaña Método.
- **Redondeo.** Algunas variables se pueden utilizar como escala, pero tienen valores que están restringidos de forma natural, por ejemplo, el número de miembros de una familia deben ser un número entero y la cantidad gastada durante una visita a una tienda de alimentación no puede tener decimales. Esta columna le permite especificar la menor denominación que se puede aceptar. Por ejemplo, para obtener valores enteros, especifique 1 como la denominación de redondeo; para obtener valores redondeados hacia el decimal siguiente, especifique 0,01. En general, los valores se redondean hacia el múltiplo entero más cercano a la denominación de redondeo. La siguiente tabla muestra cómo actúan los diferentes valores de redondeo sobre un valor imputado de 6.64823 (antes del redondeo).

Tabla 2. Redondeo de resultados.

| Denominación de redondeo | Valor al que se redondea 6,64832 |
|--------------------------|----------------------------------|
| 10 | 10 |
| 1 | 7 |
| 0,25 | 6,75 |
| 0,1 | 6,6 |
| 0,01 | 6,65 |

Exclusión de variables con grandes cantidades de datos perdidos. Normalmente, las variables de análisis se imputan y utilizan como predictores independientemente de cuántos valores perdidos tengan, siempre que tengan los suficientes datos para calcular un modelo de imputación. Puede decidir excluir variables con un alto porcentaje de valores perdidos. Por ejemplo, si especifica 50 como **Porcentaje máximo de valores perdidos**, las variables de análisis con más del 50 % de valores perdidos no se imputarán, ni se utilizarán como predictores en modelos de imputación.

Máximo de extracciones. Si se especifican valores mínimos o máximos para valores imputados de variables de escala (consulte **Mín. y Máx.** anteriormente), el procedimiento intentará extraer valores para un caso hasta que encuentre un conjunto de valores que se encuentren dentro de los rangos especificados. Si no se obtiene un conjunto de valores dentro del número especificado de extracciones por caso, el procedimiento extraerá otro conjunto de parámetros de modelo y repetirá el proceso de extracción de casos. Se producirá un error si no se obtiene un conjunto de valores que se halle entre los rangos dentro del número especificado de extracciones de casos y parámetros.

Tenga en cuenta que el incremento de estos valores puede incrementar el tiempo de procesamiento. Si el procedimiento tarda demasiado tiempo o si no puede encontrar extracciones adecuadas, compruebe los valores mínimos y máximos especificados para asegurarse de que sean correctos.

Resultados

Representación. Controla la visualización de resultados. Siempre se muestra un resumen de imputación general, que incluye tablas que relacionan las especificaciones de imputación, iteraciones (para el método de especificación totalmente condicional), variables dependientes imputadas, variables dependientes excluidas de la imputación y la secuencia de imputación. Si se especifica, también se muestran las restricciones para variables de análisis.

- **Modelo de imputación.** Esto muestra el modelo de imputación para las variables dependientes y los predictores e incluye el tipo de modelo univariante, efectos de modelo y el número de valores imputados.
- **Estadísticos descriptivos.** Esto muestra estadísticos descriptivos para variables dependientes para los que se imputan valores. En el caso de las variables de escala, los estadísticos descriptivos incluyen

media, recuento, desviación estándar, mín. y máx. de los datos de entrada originales (antes de la imputación), valores imputados (mediante imputación) y datos completos (valores originales e imputados juntos mediante imputación). En el caso de las variables categóricas, los estadísticos descriptivos incluyen recuento y porcentaje por categoría de los datos de entrada originales (antes de la imputación), valores imputados (mediante imputación) y datos completos (valores originales e imputados juntos mediante imputación).

Historial de iteraciones. Cuando se utiliza el método de imputación de especificación totalmente condicional, puede solicitar un conjunto de datos que contenga datos del historial de iteraciones para la imputación de especificación totalmente condicional. El conjunto de datos contiene medias y desviaciones estándar mediante iteración e imputación por cada variable dependiente de escala para la que se imputan valores. Puede realizar un gráfico de los datos para ayudar a evaluar la convergencia de modelo. Consulte el tema para obtener más información.

Características adicionales del comando MULTIPLE IMPUTATION

La sintaxis de comandos también le permite:

- Especificar un subconjunto de variables para los que se muestran estadísticos descriptivos (subcomando IMPUTATIONSUMMARIES).
- Especificar tanto un análisis de patrones perdidos como la imputación en una única ejecución del procedimiento.
- Especifique el número máximo de parámetros de modelo permitido al imputar cualquier variable (palabra clave MAXMODELPARAM).

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

Trabajo con datos de imputación múltiple

Cuando se crea un conjunto de datos de imputación múltiple, se añade una variable llamada, *Imputation_* con etiqueta variable *Número de imputación*, y el conjunto de datos se ordena según el mismo en orden ascendente. Los casos del conjunto de datos original tienen el valor 0. Los casos de valores imputados se numeran del 1 al *M*, donde *M* es el número de imputaciones.

Cuando abre un conjunto de datos, la presencia de *Imputation_* identifica el conjunto de datos como un posible conjunto de datos de imputación múltiple.

Activación de un conjunto de datos de imputación múltiple para su análisis

El conjunto de datos debe dividirse utilizando la opción **Comparar los grupos**, con *Imputation_* como variable de agrupación para que se considere un conjunto de datos de imputación múltiple en los análisis. También puede definir divisiones en otras variables.

Seleccione en los menús:

Datos > Segmentar archivo...

1. Seleccione **Comparar los grupos**.
2. Seleccione *Número de imputación* [*Imputation_*] como variable en la que agrupar los casos.

Asimismo, cuando activa las marcas (consulte a continuación), el archivo se divide en el *Número de imputación* [*Imputation_*].

Distinción entre valores imputados y valores observados

Puede distinguir entre los valores imputados y los observados según el color de fondo de las casillas, la fuente y la negrita (en el caso de valores imputados). Cuando cree un conjunto de datos nuevo en la

sesión actual con el procedimiento Imputar valores perdidos, las marcas se activan de forma predeterminada. Cuando abra un archivo de datos guardado que incluye imputaciones, las marcas se desactivan.

Para activar las marcas, elija en los menús del Editor de datos:

Ver > Marcar datos imputados...

También puede activar las marcas pulsando el botón de activación/desactivación de marcas de imputación situado en el borde derecho de la barra de edición en Vista de datos del Editor de datos.

Desplazamiento entre imputaciones

1. Seleccione en los menús:

Editar > Ir a la imputación...

2. Seleccione la imputación (o datos originales) en la lista desplegable.

También puede seleccionar la imputación en la lista desplegable de la barra de edición en Vista de datos del Editor de datos.

La posición relativa de caso se mantiene al seleccionar imputaciones. Por ejemplo, si hay 1.000 casos en el conjunto de datos original, el caso 1.034, el 34º caso de la primera imputación, aparece en la parte superior de la cuadrícula. Si selecciona la imputación 2 en la lista desplegable, el caso 2034, el 34º caso de la segunda imputación, aparecerá en la parte superior de la cuadrícula. Si selecciona **Datos originales** en la lista desplegable, el caso 34 aparecerá en la parte superior de la cuadrícula. La posición de columna también se mantiene al desplazarse entre imputaciones, de modo que es fácil comparar valores entre imputaciones.

Transformación y edición de valores imputados

A veces deberá realizar transformaciones en datos imputados. Por ejemplo, puede que desee tomar el registro de todos los valores de una variable de salario y guardar el resultado en una nueva variable. Un valor calculado mediante datos imputados se considerará imputado si difiere del valor calculado utilizando los datos originales.

Si edita un valor imputado en una casilla del Editor de datos, dicha casilla se seguirá considerando imputada. No se recomienda editar valores imputados de esta forma.

Análisis de datos de imputación múltiple

Muchos procedimientos admiten la combinación de resultados a partir del análisis de conjuntos de datos de imputación múltiple. Cuando las marcas de imputación están activadas, aparece un icono especial junto a los procedimientos que admiten la combinación. Por ejemplo, en el submenú Estadísticos descriptivos del menú Analizar, Frecuencias, Descriptivos, Explorar y Tablas cruzadas admiten la combinación, mientras que Cociente, Gráficos P-P y Gráficos Q-Q no lo hacen.

Tanto los resultados tabulares como el modelo PMML pueden combinarse. No hay ningún procedimiento nuevo para solicitar resultados combinados; en su lugar, una nueva pestaña del cuadro de diálogo Opciones le permite tener un control global sobre los resultados de imputación múltiple.

- **Combinación de resultados tabulares.** De manera predeterminada, cuando ejecuta un procedimiento compatible en un conjunto de datos de imputación múltiple, se producen resultados automáticamente para cada imputación, los datos originales (no imputados) y los resultados combinados (finales) que tienen en cuenta la variación entre las imputaciones. Los estadísticos combinados varían según el procedimiento.

- **Combinación de PMML.** También puede obtener la combinación de PMML a partir de procedimientos compatibles que exporten PMML. El modelo PMML combinado se solicita de la misma forma que el no combinado y se guarda en su lugar.

Los procedimientos incompatibles no producen resultados combinados ni archivos PMML combinados.

Niveles de combinación

Los resultados se combinan utilizando uno de los dos niveles siguientes:

- **Combinación naïve.** Sólo está disponible el parámetro combinado.
- **Combinación univariante.** El parámetro combinado, su error estándar, el estadístico de contraste y los grados de libertad eficaces, el valor p , el intervalo de confianza y los diagnósticos de combinación (fracción de información perdida, eficacia relativa, aumento relativo de la varianza) se mostrarán cuando estén disponibles.

Los coeficientes (regresión y correlación), quieren decir (diferencias) y los recuentos se combinan típicamente. Si el error estándar del estadístico está disponible, se utiliza la combinación univariante; en caso contrario se utiliza la combinación naïve.

Procedimientos que admiten combinación

Los siguientes procedimientos admiten conjuntos de datos de imputación múltiple a los niveles de combinación especificados para cada resultado.

Frecuencias. Se da soporte a las siguientes características:

- La tabla Estadísticos admite Medias en la combinación Univariante (si también se pide E. T. de la media) y N válido y N perdido en la combinación Naïve.
- La tabla Frecuencias admite Frecuencia en la combinación Naïve.

Descriptivos. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite Medias en la combinación Univariante (si también se pide E. T. de la media) y N en la combinación Naïve.

Tablas cruzadas. Se da soporte a las siguientes características:

- La tabla de tabulación cruzada admite Recuento en la combinación Naïve.

Medias. Se da soporte a las siguientes características:

- La tabla Informe admite Medias en la combinación Univariante (si también se pide E. T. de la media) y N en la combinación Naïve.

Prueba T para una muestra. Se da soporte a las siguientes características:

- La tabla Estadísticos admite Media en la combinación Univariante y N en la combinación Naïve.
- La tabla Prueba admite Diferencia de medias en la combinación Univariante.

Prueba T para muestras independientes. Se da soporte a las siguientes características:

- La tabla Estadísticos de grupo admite Medias en la combinación Univariante y N en la combinación Naïve.
- La tabla Prueba admite Diferencia de medias en la combinación Univariante.

Prueba T para muestras relacionadas. Se da soporte a las siguientes características:

- La tabla Estadísticos admite Medias en la combinación Univariante y N en la combinación Naïve.
- La tabla Correlaciones admite correlaciones y N en la combinación Naïve.

- La tabla Prueba admite Media en la combinación Univariante.

ANOVA de un factor. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite Media en la combinación Univariante y N en la combinación Naïve.
- La tabla Pruebas de contraste admite Valor de contraste en la combinación Univariante.

Modelos lineales mixtos. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite Media y N en la combinación Naïve.
- La tabla Estimaciones de efectos fijos admite Estimación en la combinación Univariante.
- La tabla Estimaciones de parámetros de covarianzas admite Estimación en la combinación Univariante.
- Medias marginales estimadas: la tabla Estimaciones admite Media en la combinación Univariante.
- Medias marginales estimadas: la tabla Comparaciones por parejas admite Diferencia de medias en la combinación Univariante.

Modelos lineales generalizados y Ecuaciones de estimación generalizadas. Estos procedimientos admiten la combinación de PMML.

- La tabla Información sobre la variable categórica admite N y Porcentajes en la combinación Naïve.
- La tabla Información sobre la variable continua admite N y Media en la combinación Naïve.
- La tabla Estimaciones de los parámetros admite el coeficiente B en la combinación Univariante.
- Medias marginales estimadas: la tabla Coeficientes de estimación admite Correlaciones en la combinación Naïve.
- Medias marginales estimadas: la tabla Estimaciones admite Media en la combinación Univariante.
- Medias marginales estimadas: la tabla Comparaciones por parejas admite Diferencia de medias en la combinación Univariante.

Correlaciones bivariadas. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite Media y N en la combinación Naïve.
- La tabla Correlaciones admite correlaciones y N en la combinación Univariante. Tenga en cuenta que las correlaciones se transforman utilizando la transformación z de Fisher antes de realizar la combinación, y luego se transforman retrospectivamente tras la combinación.

Correlaciones parciales. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite Media y N en la combinación Naïve.
- La tabla Correlaciones admite correlaciones en la combinación Naïve.

Regresión lineal. Este procedimiento admite la combinación de PMML.

- La tabla Estadísticos descriptivos admite Media y N en la combinación Naïve.
- La tabla Correlaciones admite correlaciones y N en la combinación Naïve.
- La tabla Coeficientes admite B en la combinación Univariante y Correlaciones en la combinación Naïve.
- La tabla Coeficientes de correlación admite correlaciones en la combinación Naïve.
- La tabla Estadísticas de residuos admite Media y N en la combinación Naïve.

Regresión logística binaria. Este procedimiento admite la combinación de PMML.

- La tabla Variables en la ecuación admite B en la combinación Univariante.

Regresión logística multinomial. Este procedimiento admite la combinación de PMML.

- La tabla Estimaciones de los parámetros admite el coeficiente B en la combinación Univariante.

Regresión ordinal. Se da soporte a las siguientes características:

- La tabla Estimaciones de los parámetros admite el coeficiente B en la combinación Univariante.

Análisis discriminante. Este procedimiento admite la combinación del modelo XML.

- La tabla Estadísticos de grupo admite Media y N válido en la combinación Naïve.
- La tabla Matrices intra-grupos combinadas admite Correlaciones en la combinación Naïve.
- La tabla Coeficientes de funciones discriminantes canónicas admite Coeficientes no tipificados en la combinación Naïve.
- La tabla Funciones en centroides de grupo admite Coeficientes no tipificados en la combinación Naïve.
- La tabla Coeficientes de función de clasificación admite Coeficientes en la combinación Naïve.

Prueba de chi-cuadrado. Se da soporte a las siguientes características:

- La tabla Descriptivos admite Media y N en la combinación Naïve.
- La tabla Frecuencias admite N Observado en la combinación Naïve.

Prueba binomial. Se da soporte a las siguientes características:

- La tabla Descriptivos admite Medias y N en la combinación Naïve.
- La tabla Prueba admite N, Proporción observada y Proporción de prueba en la combinación Naïve.

Prueba de rachas. Se da soporte a las siguientes características:

- La tabla Descriptivos admite Medias y N en la combinación Naïve.

Prueba de Kolmogorov-Smirnov para una muestra. Se da soporte a las siguientes características:

- La tabla Descriptivos admite Medias y N en la combinación Naïve.

Pruebas para dos muestras independientes. Se da soporte a las siguientes características:

- La tabla Rangos admite Rango promedio y N en la combinación Naïve.
- La tabla Frecuencias admite N en la combinación Naïve.

Pruebas para varias muestras independientes. Se da soporte a las siguientes características:

- La tabla Rangos admite Rango promedio y N en la combinación Naïve.
- La tabla Frecuencias admite Recuentos en la combinación Naïve.

Pruebas para dos muestras relacionadas. Se da soporte a las siguientes características:

- La tabla Rangos admite Rango promedio y N en la combinación Naïve.
- La tabla Frecuencias admite N en la combinación Naïve.

Pruebas para varias muestras relacionadas. Se da soporte a las siguientes características:

- La tabla Rangos admite Rango promedio en la combinación Naïve.

Regresión de Cox. Este procedimiento admite la combinación de PMML.

- La tabla Variables en la ecuación admite B en la combinación Univariante.
- La tabla Medias de covariables admite media en la combinación Naïve.

Opciones de imputación múltiple

La pestaña Imputaciones múltiples controla dos tipos de preferencias relacionadas con las imputaciones múltiples:

Aspecto de datos imputados. De manera predeterminada, las casillas que contienen datos imputados tendrán un color de fondo diferente que las casillas con datos no imputados. El aspecto distintivo de los

datos imputados debería facilitarle el desplazamiento por un conjunto de datos y la localización de estas casillas. Puede cambiar el color de fondo predeterminado de las casillas, la fuente y hacer que los datos imputados aparezcan en negrita.

Resultados. Este grupo controla el tipo de resultados del Visor producidos cuando se analiza un conjunto de datos imputado de forma múltiple. De manera predeterminada, se producirán resultados para el conjunto de datos originales (de antes de la imputación) y para cada uno de los conjuntos de datos imputados. Además, se generarán resultados combinados finales para los procedimientos que sean compatibles con la combinación de datos imputados. Los diagnósticos de combinación también aparecerán cuando se realice una combinación univariante. Sin embargo, puede suprimir los resultados que no desee ver.

Para establecer opciones de imputación múltiple

Elija en los menús:

Editar > Opciones

Pulse la pestaña Imputaciones múltiples.

Avisos

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

Es posible que IBM no ofrezca los productos, servicios o características que se tratan en este documento en otros países. Consulte al representante local de IBM para obtener información sobre los productos y servicios disponibles actualmente en su zona. Las referencias a programas, productos o servicios de IBM no pretenden establecer ni implicar que sólo puedan utilizarse dichos productos, programas o servicios de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio 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 el asunto descrito en este documento. Este documento no le otorga ninguna licencia para estas patentes. Puede enviar preguntas acerca de las licencias, por escrito, a:

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

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

Intellectual Property Licensing
Ley de propiedad legal e intelectual
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

El párrafo siguiente no se aplica al Reino Unido ni a ningún país donde tales determinaciones son incoherentes con las leyes locales: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍA DE CUALQUIER ESPECIE, YA SEA EXPRESA O IMPLÍCITA, INCLUIDAS, PERO SIN ESTAR LIMITADO A, LAS GARANTÍAS IMPLÍCITAS DE NO-INFRACCIÓN, COMERCIALIZACIÓN O ADECUACIÓN PARA UNA FINALIDAD ESPECÍFICA. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se aplica 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 cualquier información que proporcione en la forma que crea adecuada sin incurrir por ello en ninguna obligación con el cliente.

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

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

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

El programa bajo licencia descrito en este documento y todo el material bajo licencia disponible para el mismo lo proporciona IBM de acuerdo con los términos del Acuerdo de Cliente de IBM, el Acuerdo Internacional de Programa bajo Licencia de IBM o cualquier otro 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 comprobado estos productos y no puede confirmar la precisión de su rendimiento, 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 COPYRIGHT:

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

Cada una de las copias, totales o parciales, de estos programas de ejemplo o cualquier trabajo derivado de ellos, debe incluir el siguiente aviso de copyright:

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

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 producto y servicio 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 con base Java son marcas comerciales o son marcas registradas de Oracle y/o sus filiales.

Índice

A

- Análisis de valores perdidos 3
 - Características adicionales del comando 9
 - EM 7
 - estadísticos descriptivos 5
 - estimación de los estadísticos 6
 - expectation-maximization (maximización esperada) 8
 - imputación de valores perdidos 6
 - métodos 6
 - patrones 4
 - prueba MCAR 6
 - regression 8
- Analizar patrones 12

C

- correlaciones
 - en Análisis de valores perdidos 7, 8
- covarianza
 - en Análisis de valores perdidos 7, 8

D

- datos incompletos
 - consultar Análisis de valores perdidos 3
- desviación estándar
 - en Análisis de valores perdidos 5
- discordancia
 - en Análisis de valores perdidos 5

E

- eliminación por lista
 - en Análisis de valores perdidos 3
- eliminación por parejas
 - en Análisis de valores perdidos 3
- EM
 - en Análisis de valores perdidos 7
- especificación totalmente condicional
 - en imputación múltiple 14

H

- historial de iteraciones
 - en imputación múltiple 16

I

- imputación monotónica
 - en imputación múltiple 14
- imputación múltiple 11
 - analizar patrones 12
 - imputar valores perdidos 13
- Imputación múltiple 17, 18
- Imputar valores de datos perdidos 13

- Imputar valores de datos perdidos (*continuación*)
 - método de imputación 14
 - restricciones 15
 - salida 16

M

- media
 - en Análisis de valores perdidos 5, 7, 8

O

- ordenación de casos
 - en Análisis de valores perdidos 4

P

- prueba MCAR
 - en Análisis de valores perdidos 3
- prueba MCAR de Little 6
 - en Análisis de valores perdidos 3
- Prueba t
 - en Análisis de valores perdidos 5
- prueba t de Student
 - en Análisis de valores perdidos 8

R

- recuentos de valores extremos
 - en Análisis de valores perdidos 5
- regression
 - en Análisis de valores perdidos 8
- residuos
 - en Análisis de valores perdidos 8

T

- tablas de frecuencias
 - en Análisis de valores perdidos 5
- tabulación de casos
 - en Análisis de valores perdidos 4
- tabulación de categorías
 - en Análisis de valores perdidos 5

V

- valores perdidos
 - estadísticos univariados 5
- variables de indicador
 - en Análisis de valores perdidos 5
- variables de indicador de valores perdidos
 - en Análisis de valores perdidos 5
- variantes normales
 - en Análisis de valores perdidos 8



Impreso en España