

IBM SPSS Bootstrapping 23

IBM

Nota

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos" en la página 9.

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 la simulación
de muestreo 1**

Capítulo 2. Simulación de muestreo. . . 3
Procedimientos que admiten simulación de muestreo 4
Características adicionales del comando BOOTSTRAP 7

Avisos 9
Marcas comerciales 11

Índice 13

Capítulo 1. Introducción a la simulación de muestreo

Cuando recopila datos suele estar interesado en las propiedades de la población de la que ha tomado la muestra. Hace inferencias acerca de los parámetros de la población con estimaciones calculadas de la muestra. Por ejemplo, si el conjunto de datos *Employee data.sav* que se incluye con el producto es una muestra aleatoria de una población mayor de empleados, la media muestral de 34.419,57 dólares como *Salario actual* es una estimación de la media del salario actual de la población de los empleados. Además, esta estimación tiene un error estándar de 784,311 dólares para una muestra de un tamaño de 474; y un intervalo de confianza del 95% para la media del salario actual de la población de los empleados es de 32.878,40 dólares a 35.960,73 dólares. Pero, ¿cuál es el nivel de fiabilidad de estos estimadores? Para algunas poblaciones "conocidas" y parámetros de buen comportamiento, sabemos algo acerca de las propiedades de las estimaciones de la muestra y podemos confiar en estos resultados. La simulación de muestreo busca más información acerca de las propiedades de los estimadores de poblaciones "desconocidas" y parámetros de mal comportamiento.

Funcionamiento de la simulación de muestreo

En su forma más simple, para un conjunto de datos con un tamaño de la muestra de N , tomará B muestras de "simulación de muestreo" de un tamaño N sustituyendo del conjunto de datos original y calcular el estimador de cada uno de estas B muestras de simulación de muestreo. Estas B estimaciones de simulación de muestreo son una muestra de un tamaño B de la que podrá realizar inferencias acerca del estimador. Por ejemplo, si toma 1.000 muestras de simulación de muestreo del conjunto de datos *Employee data.sav*, el error estándar de estimación de simulación de muestreo de 776,91 dólares para la media muestral de *Salario actual* es una alternativa a la estimación de 784,311 dólares.

Además, la simulación de muestreo proporciona un error estándar y un intervalo de confianza para la mediana, cuyas estimaciones paramétricas no están disponibles.

Compatibilidad de la simulación de muestreo en el producto

La simulación de muestreo se incorpora como un cuadro de diálogo subordinado en procedimientos que admiten simulación de muestreo. Consulte "Procedimientos que admiten simulación de muestreo" en la página 4 si desea obtener información acerca de los procedimientos que admiten simulación de muestreo.

Si se requiere simulación de muestreo en los cuadros de diálogo, se pega un nuevo comando `BOOTSTRAP` independiente, además de la sintaxis normal que genera el cuadro de diálogo. El comando `BOOTSTRAP` crea las muestras de simulación de muestreo en función de sus especificaciones. Internamente, el producto trata estas muestras de simulación de muestreo como segmentaciones, incluso si no se muestran de forma explícita en el Editor de datos. Significa que, de forma interna, son efectivamente $B*N$ casos, de forma que el recuento de casos en la barra de estado contará desde 1 a $B*N$ cuando se procesen los datos durante la simulación de muestreo. El Sistema de gestión de resultados (OMS) se utiliza para recopilar los resultados de la ejecución del análisis en cada "segmentación de simulación de muestreo". Estos resultados se combinan y los resultados de simulación de muestreo combinados se muestran en el Visor con el resto del resultado normal que genera el procedimiento. En algunos casos, podrá ver una referencia a "segmentación de simulación de muestreo 0"; es el conjunto de datos original.

Capítulo 2. Simulación de muestreo

La simulación de muestreo es un método para derivar estimaciones robustas de errores estándar e intervalos de confianza para estimaciones como la media, mediana, proporción, razón de las ventajas, coeficientes de correlación o coeficientes de regresión. También se puede utilizar para crear contrastes de hipótesis. La simulación de muestreo es más útil como alternativa a estimaciones paramétricas en caso de que los supuestos de esos métodos sean dudosos (como en el caso de modelos de regresión con residuos heteroscedástico se ajusten a muestras pequeñas), o si la inferencia paramétrica no es posible o requiere fórmulas muy complicadas para el cálculo de errores estándar (como en el caso de cálculo de intervalos de confianza de mediana, cuartiles y otros percentiles).

Ejemplos. Una empresa de telecomunicaciones pierde alrededor del 27% de sus clientes por abandono cada mes. Para reducir el porcentaje de abandono, los directivos quieren saber si este porcentaje varía en diferentes grupos de clientes predefinidos. Mediante la simulación de muestreo, puede determinar si un porcentaje concreto de abandonos describe de forma adecuada los cuatro tipos principales de clientes.

En una revisión de los registros de empleados, los directivos están interesados en las experiencias anteriores de los empleados. La experiencia laboral es asimétrica, lo que hace que la media sea una estimación menos deseable de la experiencia laboral "habitual" entre los empleados que la mediana. Sin embargo, los intervalos de confianza no están disponibles para la mediana en el producto.

Los directivos también están interesados en determinar los factores que están asociados con los aumentos de salarios de los empleados mediante la definición de un modelo lineal de la diferencia entre el salario inicial y el actual. Al realizar la simulación de muestreo de un modelo lineal, puede utilizar métodos de muestreo sucesivo especiales (muestreo de simulación de muestreo residual y wild) para obtener resultados más precisos.

Muchos procedimientos admiten simulación de muestreo y la combinación de resultados a partir del análisis de muestras de simulación de muestreo. Los controles para especificar análisis de simulación de muestreo se integran directamente como un cuadro de diálogo subordinado común en procedimientos que admiten simulación de muestreo. La configuración del cuadro de diálogo de simulación de muestreo permanece en los procedimientos de forma que, si ejecuta un análisis de frecuencias con simulación de muestreo en los cuadros de diálogo, la simulación de muestreo se activará de forma predeterminada para otros procedimientos que lo admitan.

Para obtener un análisis de simulación de muestreo

1. En los menús seleccione un procedimiento que admita la simulación de muestreo y pulse en **Simular muestreo**.
2. Seleccione **Ejecutar simulación de muestreo**.

También puede controlar las siguientes opciones:

Número de muestras. Para los intervalos de percentil y BCa producidos, se recomienda utilizar al menos 1000 muestras de simulación de muestreo. Especifique un número entero positivo.

Definir semilla para tornado de Mersenne. Si se establece una semilla es posible replicar análisis. El uso de este control es parecido a establecer el tornado de Mersenne como generador activo y especificar un punto de inicio fijo en el cuadro de diálogo Generadores de números aleatorios, con la importante diferencia de que la definición de la semilla de este cuadro de diálogo mantendrá el estado actual del generador de números aleatorios y restaurará dicho estado cuando haya terminado el análisis.

Intervalos de confianza. Especifique un nivel de confianza mayor que 50 y menor que 100. Los intervalos de percentiles sólo utilizan los valores de simulación de muestreo ordenados correspondientes a los percentiles de intervalo de confianza deseados. Por ejemplo, un intervalo de confianza de percentil del 95% utiliza los percentiles 2,5 y 97,5 de los valores de simulación de muestreo como los límites inferior y superior del intervalo (interpolando los valores de simulación de muestreo si es necesario). Los intervalos de sesgo corregidos y acelerados (BCa) son intervalos ajustados que son más precisos, pero necesitan más tiempo de cálculo.

Muestreo. El método **simple** consiste en volver a muestrear los casos reemplazándolos del conjunto de datos original. El método **estratificado** consiste en volver a muestrear los casos sustituyendo el conjunto de datos original, *en* los estratos definidos por las variables de estratos de clasificación cruzada. El muestreo de simulación de muestreo estratificado puede ser muy útil si las unidades de los estratos son relativamente homogéneas aunque las unidades para todos los estratos son muy diferentes.

Procedimientos que admiten simulación de muestreo

Los siguientes procedimientos admiten simulación de muestreo.

Nota:

- La simulación de muestreo no funciona con conjuntos de datos de imputación múltiple. Si hay una variable *Imputation_* en el conjunto de datos, el cuadro de diálogo Simular muestreo se desactiva.
- La simulación de muestreo no funciona si hay valores de ponderación no enteros.
- La simulación de muestreo utiliza eliminación por lista para determinar los casos; es decir, los casos con valores perdidos en cualquiera de las variables de análisis se eliminan del análisis, de forma que, cuando la simulación de muestreo está en efecto, eliminación por lista se activa incluso si el procedimiento de análisis especifica otra forma de manejo de valores perdidos.

Opción Estadísticas básicas

Frecuencias. Se da soporte a las siguientes características:

- La tabla Estadísticos admite estimaciones de simulación de muestreo de la media, la desviación estándar, la varianza, la mediana, la asimetría, la curtosis y los percentiles.
- La tabla Frecuencias admite estimaciones de simulación de muestreo del porcentaje.

Descriptivos. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media, la desviación estándar, la varianza, la asimetría y la curtosis.

Explorar. Se da soporte a las siguientes características:

- La tabla Descriptivos admite estimaciones de simulación de muestreo de la media, la media recortada al 5%, la desviación estándar, la varianza, la mediana, la asimetría, la curtosis y el rango intercuartil.
- La tabla Estimadores-M admite estimaciones de simulación de muestreo del estimador-M de Huber, el estimador bponderado de Tukey, el estimador-M de Hampel y la onda de Andrews.
- La tabla Percentiles admite estimaciones de simulación de muestreo de los percentiles.

Tablas cruzadas. Se da soporte a las siguientes características:

- La tabla Medidas direccionales admite estimaciones de simulación de muestreo de Lambda, Goodman y Kruskal Tau, coeficiente de incertidumbre y d de Somers.
- La tabla Medidas simétricas admite estimaciones de simulación de muestreo de Phi, V de Cramer, coeficiente de contingencia, tau-b de Kendall, tau-c de Kendall, Gamma, correlación de Spearman y r de Pearson.
- La tabla Estimación de riesgo admite estimaciones de simulación de muestreo de la razón de las ventajas.

- La tabla de razón de las ventajas común de Mantel-Haenszel admite estimaciones de simulación de muestreo y pruebas de significación de $\ln(\text{Estimación})$.

Medias. Se da soporte a las siguientes características:

- La tabla Informe admite estimaciones de simulación de muestreo de la media, la mediana, la mediana agrupada, la desviación estándar, la varianza, la curtosis, la asimetría, la media armónica y la media geométrica.

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

- La tabla Estadísticos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Prueba admite estimaciones de simulación de muestreo y pruebas de significación de la diferencia de medias.

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

- La tabla Estadísticos de grupo admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Prueba admite estimaciones de simulación de muestreo y pruebas de significación de la diferencia de medias.

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

- La tabla Estadísticos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Correlaciones admite estimaciones de simulación de muestreo de correlaciones.
- La tabla Prueba admite estimaciones de simulación de muestreo de la media.

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

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Comparaciones múltiples admite estimaciones de simulación de muestreo de la diferencia de medias.
- La tabla Pruebas de contraste admite estimaciones de simulación de muestreo y pruebas de significación de valor de contraste.

MLG Univariante. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Estimaciones de los parámetros admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.
- La tabla de resultados de contraste admite estimaciones de simulación de muestreo y pruebas de significación de la diferencia.
- Medias marginales estimadas: La tabla Estimaciones admite estimaciones de simulación de muestreo de la media.
- Medias marginales estimadas: La tabla Comparaciones por parejas admite estimaciones de simulación de muestreo de la diferencia de medias.
- La tabla Pruebas post hoc: Comparaciones múltiples admite estimaciones de simulación de muestreo de la diferencia de medias.

Correlaciones bivariadas. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media y la desviación estándar.

- La tabla Correlaciones admite estimaciones de simulación de muestreo y pruebas de significación de correlaciones.

Notas:

Si se requieren correlaciones no paramétricas (tau-b de Kendall o Spearman) además de las correlaciones de Pearson, el cuadro de diálogo pega los comandos CORRELATIONS y NONPAR CORR con un comando BOOTSTRAP diferente para cada una. Se utilizarán las mismas muestras de simulación de muestreo para calcular todas las correlaciones.

Antes de la combinación, la transformación Fisher Z se aplica a las correlaciones. Tras la combinación, se aplica la transformación Z inversa.

Correlaciones parciales. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Correlaciones admite estimaciones de simulación de muestreo de correlaciones.

Regresión lineal. Se da soporte a las siguientes características:

- La tabla Estadísticos descriptivos admite estimaciones de simulación de muestreo de la media y la desviación estándar.
- La tabla Correlaciones admite estimaciones de simulación de muestreo de correlaciones.
- La tabla Resumen de modelo admite estimaciones de simulación de muestreo de Durbin-Watson.
- La tabla Coeficientes admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.
- La tabla Coeficientes de correlación admite estimaciones de simulación de muestreo de correlaciones.
- La tabla Estadísticos residuales admite estimaciones de bootstrap de la media y la desviación estándar.

Regresión ordinal. Se da soporte a las siguientes características:

- La tabla Estimaciones de los parámetros admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

Análisis discriminante. Se da soporte a las siguientes características:

- La tabla Coeficientes de funciones discriminantes canónicas tipificados admite estimaciones de simulación de muestreo de coeficientes tipificados.
- La tabla Coeficientes de funciones discriminantes canónicas admite estimaciones de simulación de muestreo de coeficientes no tipificados.
- La tabla Coeficientes de función de clasificación admite estimaciones de simulación de muestreo de coeficientes.

Opción Estadísticas avanzadas

MLG Multivariante. Se da soporte a las siguientes características:

- La tabla Estimaciones de los parámetros admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

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

- La tabla Estimaciones de efectos fijos admite estimaciones de simulación de muestreo y pruebas de significación de estimación.
- La tabla Estimaciones de parámetros de covarianzas admite estimaciones de simulación de muestreo y pruebas de significación de estimación.

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

- La tabla Estimaciones de los parámetros admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

Regresión de Cox. Se da soporte a las siguientes características:

- La tabla Variables en la ecuación admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

Opción Regresión

Regresión logística binaria. Se da soporte a las siguientes características:

- La tabla Variables en la ecuación admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

Regresión logística multinomial. Se da soporte a las siguientes características:

- La tabla Estimaciones de los parámetros admite estimaciones de simulación de muestreo y pruebas de significación de coeficiente B.

Características adicionales del comando BOOTSTRAP

La sintaxis de comandos también le permite:

- Realice muestreos de simulación de muestreo residuales y wild (subcomando SAMPLING)

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 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. Cualquier referencia a un producto, programa o servicio de IBM no pretende afirmar ni implicar que solamente se pueda utilizar ese producto, programa o servicio 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.

Puede que IBM tenga patentes o solicitudes de patente pendientes que cubran la materia descrita 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
EE.UU.

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

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

El párrafo siguiente no se aplica al Reino Unido ni a ningún otro país donde estas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL" SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUIDAS, PERO NO LIMITÁNDOSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO. Algunos estados no permiten la renuncia a expresar o a garantías implícitas en determinadas transacciones, por lo tanto, esta declaración no se 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 se reserva el derecho a realizar en cualquier momento y sin notificación previa, mejoras o modificaciones en los productos y programas que se describen en el presente manual.

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 será responsabilidad del usuario.

IBM puede utilizar o distribuir cualquier información que se le proporcione en la forma que considere adecuada, sin incurrir por ello en ninguna obligación para con el remitente.

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
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 los proporciona IBM bajo los términos del Acuerdo de cliente de IBM, el 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 esos productos y no puede confirmar la exactitud del rendimiento, de la compatibilidad ni de ninguna otra declaración relacionada con productos que no sean de IBM. Las preguntas acerca de las aptitudes de productos que no sean de IBM deben dirigirse a los proveedores de dichos productos.

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 escritos en lenguaje fuente, los cuales muestran técnicas de programación para diversas plataformas operativas. Puede copiar, modificar y distribuir estos programas de ejemplo 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 ejemplo. Los ejemplos no se han probado minuciosamente 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 alguna. IBM no será responsable de los daños que surjan por el uso de los programas de muestra.

Cada copia o parte de estos programas de ejemplo, así como todo trabajo derivado, 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.

Marcas comerciales

IBM, el logotipo de IBM e ibm.com son marcas registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios 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 registradas y logotipos basados en Java son marcas registradas de Oracle o sus filiales.

Índice

S

- simulación de muestreo 3
- procedimientos admitidos 4



Impreso en España