

IBM SPSS Bootstrapping 25

IBM

Comunicado

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 7.

Informações sobre o produto

Esta edição aplica-se à versão 25, liberação 0, modificação 0 do IBM® SPSS Statistics e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Autoinicialização	1
Introdução a autoinicialização	1
Autoinicialização	2
Procedimentos que suportam autoinicialização	3
Recursos adicionais do comando BOOTSTRAP	5

Avisos	7
Marcas comerciais	9
Índice Remissivo	11

Autoinicialização

Os seguintes recursos de autoinicialização são incluídos na SPSS Statistics Premium Edition ou na opção de autoinicialização.

Introdução a autoinicialização

Ao coletar dados, você está sempre interessado nas propriedades da população a partir da qual a amostra foi retirada. São feitas inferências sobre estes parâmetros da população com estimativas calculadas a partir da amostra. Por exemplo, se o conjunto de dados *Employee data.sav* que está incluído com o produto for uma amostra aleatória de uma população maior de funcionários, a média da amostra de \$34.419,57 para *Salário atual* será uma estimativa da média de salário atual para a população de funcionários. Além disso, esta estimativa tem um erro padrão de \$784,311 para uma amostra de tamanho 474, desse modo, um intervalo de confiança de 95% para a média de salário atual na população de funcionários é de \$32.878,40 a \$35.960,73. Mas, quão confiáveis são esses estimadores? Para determinadas populações e parâmetros bem-comportados "conhecidos", sabemos muito pouco sobre as propriedades das estimativas de amostra, e podem ser confiantes nesses resultados. A autoinicialização procura descobrir mais informações sobre as propriedades de estimadores para populações e parâmetros mal-comportados "desconhecidos".

Como funciona a autoinicialização

Na sua forma mais simples, para um conjunto de dados com um tamanho de amostra de N , são levadas amostras "bootstrap" B de tamanho N com substituição a partir do conjunto de dados original e é calculado o estimador para cada uma dessas amostras bootstrap B . Essas estimativas bootstrap B são uma amostra de tamanho B a partir das quais é possível fazer inferências sobre o estimador. Por exemplo, se você leva 1.000 amostras bootstrap do conjunto de dados *Employee data.sav*, o erro padrão do bootstrap estimado de \$776.91 para a média de amostra para *Salário atual* é uma alternativa para a estimativa de \$784.311.

Além disso, a autoinicialização fornece um erro padrão e intervalo de confiança para mediano, para os quais as estimativas paramétricas estão indisponíveis.

Suporte para autoinicialização no produto

A autoinicialização é incorporada como um subdiálogo em procedimentos que suportam a autoinicialização. Consulte "Procedimentos que suportam autoinicialização" na página 3 para obter informações sobre quais procedimentos suportam autoinicialização.

Quando a autoinicialização é solicitada nos diálogos, um comando `BOOTSTRAP` novo e separado é colado além da sintaxe comum gerada pelo diálogo. O comando `BOOTSTRAP` cria as amostras bootstrap de acordo com suas especificações. Internamente, o produto trata essas amostras bootstrap como divisões, embora não sejam explicitamente mostradas no Editor de Dados. Isso significa que, internamente, há efetivamente casos $B*N$, de modo que o contador de caso na barra de status contará de 1 para $B*N$ ao processar os dados durante a autoinicialização. O Output Management System (OSM) é usado para coletar os resultados da execução da análise em cada "divisão de autoinicialização". Esses resultados são agrupados em conjunto e os resultados de bootstrap agrupados em conjunto são exibidos no Visualizador com o resto da saída normal gerada pelo procedimento. Em determinados casos, uma referência a "divisão de autoinicialização 0" pode ser vista; este é o conjunto de dados original.

Autoinicialização

A autoinicialização é um método para derivação de estimativas robustas de erros padrão e intervalos de confiança para estimativas, como média, mediana, proporção, razão de chances, coeficiente de correlação ou coeficiente de regressão. Ela também pode ser usada para construir testes de hipótese. A autoinicialização é mais útil como uma alternativa para estimativas paramétricas quando as suposições desses métodos estão em dúvida (como no caso de modelos de regressão com ajuste de resíduos heteroscedásticos para pequenas amostras), ou onde a inferência paramétrica é impossível ou requer fórmulas muito complicadas para o cálculo dos erros padrão (como no caso de calcular intervalos de confiança para a mediana, quartis e outras percentis).

Exemplos. Uma firma de telecomunicações perde cerca de 27% de seus clientes para a migração para o concorrente a cada mês. A fim de concentrar os esforços na redução de migração adequada, o gerenciamento quer saber se esta porcentagem varia entre grupos de clientes predefinidos. Usando a autoinicialização, é possível determinar se uma única taxa de migração descreve adequadamente os quatro principais tipos de clientes.

Em uma revisão de registros de funcionários, o gerenciamento está interessado na experiência de trabalho anterior dos funcionários. A experiência de trabalho está inclunada para a direita, o que torna a média uma estimativa menos desejável da experiência de trabalho anterior "típica" entre os funcionários do que a mediana. No entanto, os intervalos de confiança dos parâmetros não estão disponíveis para a mediana no produto.

O gerenciamento está também interessado em determinar quais fatores estão associados a aumentos de salário dos funcionários, ajustando um modelo linear à diferença entre os salários atual e inicial. Ao efetuar a autoinicialização de um modelo linear, é possível usar métodos complementares especiais (bootstrap residuais e brutos) para obter resultados mais precisos.

Muitos procedimentos suportam amostragem bootstrap e o agrupamento de resultados a partir da análise de amostras de bootstrap. Os controles para especificar as análises de bootstrap são integradas diretamente como um subdiálogo comum em procedimentos que suportam a autoinicialização. As configurações no diálogo de bootstrap persistem nos procedimentos de modo que, se você executar uma análise Frequências com autoinicialização por meio dos diálogos, a autoinicialização será ativada por padrão para outros procedimentos que a suportam.

Para obter uma análise de bootstrap

1. A partir dos menus, escolha um procedimento que suporta autoinicialização e clique em **Autoinicialização**.
2. Selecione **Executar autoinicialização**.

Opcionalmente, é possível controlar as opções a seguir:

Número de amostras. Para os intervalos de percentil e BCa produzidos, é recomendável usar pelo menos 1.000 amostras bootstrap. Especifique um número inteiro positivo.

Configurar semente para Mersenne Twister. Configurar uma semente permite replicar análises. O uso desse controle é semelhante a configurar o Mersenne Twister como o gerador ativo e especificar um ponto inicial fixo no diálogo Geradores de Número Aleatório, com a diferença importante de que configurar a semente neste diálogo irá preservar o estado atual do gerador de número aleatório e restaurar esse estado após a análise ser concluída.

Intervalos de confiança. Especifique um nível de confiança maior que 50 e menor que 100. Os intervalos de percentil simplesmente usam os valores de bootstrap ordenados que correspondem aos percentis do intervalo de confiança. Por exemplo, um intervalo de confiança de percentil 95% usa os percentis 2.5° e 97.5° dos valores de bootstrap como os limites inferior e superior do intervalo (interpolando os valores de

bootstrap, se necessário). Os intervalos de viés corrigido e acelerado (BCa) são intervalos ajustados que são mais precisos quanto ao custo de precisar de mais tempo para calcular.

Amostragem. O método **Simple**s é a reamostragem de caso com substituição do conjunto de dados original. O método **Estratificado** é a reamostragem de caso com substituição do conjunto de dados original *dentro* da camada definida pela classificação cruzada de variáveis de camadas. A amostragem estratificada de bootstrap pode ser útil quando as unidades dentro das camadas são relativamente homogêneas enquanto as unidades entre as camadas são muito diferentes.

Procedimentos que suportam autoinicialização

Os procedimentos a seguir suportam autoinicialização.

Observação:

- A autoinicialização não funciona com conjuntos de dados imputados de multiplicação. Se houver uma variável *Imputation_* no conjunto de dados, o diálogo Bootstrap será desativado.
- A autoinicialização não funcionará se houver valores de ponderação de números não inteiros.
- A autoinicialização usa exclusão de listwise para determinar a base de caso; isto é, casos com valores omissos em qualquer uma das variáveis de análise são excluídos da análise, de modo que quando a autoinicialização está em vigor, a exclusão de listwise está em vigor, mesmo se o procedimento de análise especificar outra forma de tratamento de valor omissos.

Statistics Base Edition

Frequências. Os seguintes recursos são suportados:

- A tabela Estatísticas suporta estimativa bootstrap para a média, desvio padrão, variância, mediana, assimetria, curtose e percentis.
- A tabela de frequências suporta estimativas bootstrap para percentual.

Descritivos. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suportam estimativas bootstrap para a média, desvio padrão, variância, assimetria e curtose.

Explorar. Os recursos a seguir são suportados:

- A tabela Descritivos suporta estimativas bootstrap para a média, 5% de média cortada, desvio padrão, variância, mediana, assimetria, curtose e amplitude interquartil.
- A tabela Estimadores M suporta estimativa bootstrap para Estimador M de Huber, Biweight de Tukey, Estimador M de Hampel e Wave de Andrew.
- A tabela Percentis suporta estimativas bootstrap para percentis.

Tabulações cruzadas. Os recursos a seguir são suportados:

- A tabela Medidas Direcionais suporta estimativas bootstrap para Lambda, Goodman e Kruskal Tau, coeficiente de incerteza e d de Somers
- A tabela Medidas Simétricas suporta estimativas bootstrap para Phi, V de Cramer, Coeficiente de Contingência, tau-b de Kendall, tau-c de Kendall, Gama, Correlação de Spearman e R de Pearson.
- A tabela Estimativa de Risco suporta estimativas bootstrap para razão de chances.
- A tabela Razão de Chances Comum de Mantel-Haenszel suporta as estimativas bootstrap e testes de significância para ln(Estimativa).

Médias. Os recursos a seguir são suportados:

- A tabela Relatório suporta estimativas bootstrap para a média, mediana, mediana agrupada, desvio padrão, variância, curtose, assimetria, média harmônica e média geométrica.

Teste T de uma amostra. Os recursos a seguir são suportados:

- A tabela Estatísticas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Teste suporta estimativas bootstrap e testes de significância para a diferença média.

Teste T de amostras independentes. Os recursos a seguir são suportados:

- A tabela Estatísticas do Grupo suporta as estimativas bootstrap para a média e o desvio padrão.
- A tabela Teste suporta estimativas bootstrap e testes de significância para a diferença média.

Teste T de amostras pairwise. Os recursos a seguir são suportados:

- A tabela Estatísticas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Correlações suporta estimativas bootstrap para correlações.
- A tabela Teste suporta estimativas bootstrap para a média.

Análise de variância unidirecional. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Comparações Múltiplas suporta estimativas bootstrap para a diferença média.
- A tabela Testes de Contraste suporta estimativa bootstrap e testes de significância para o valor de contraste.

GLM Univariados. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Estimativas de Parâmetro suporta estimativas bootstrap e testes de significância para o coeficiente, B.
- A tabela Resultados de Contraste suporta estimativas bootstrap e testes de significância para a diferença.
- As Médias Marginais Estimadas: a tabela Estimativas suporta estimativas bootstrap para a média.
- A Médias Marginais Estimadas: a tabela Comparação entre pares suporta estimativas bootstrap para a diferença média.
- Os Testes posteriori: a tabela Comparações Múltiplas suporta estimativas bootstrap para a Diferença Média.

Correlações bivariadas. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Correlações suporta estimativas bootstrap e testes de significância para correlações.

Notas:

Se as correlações não paramétricas (tau-b de Kendall ou Spearman) forem solicitadas além das correlações de Pearson, o diálogo cola os comandos CORRELATIONS e NONPAR CORR com um comando BOOTSTRAP separado para cada um. As mesmas amostras bootstrap serão usadas para calcular todas as correlações.

Antes de agrupar, a conversão de Fischer Z é aplicada às correlações. Após o agrupamento, a conversão de Z inversa é aplicada.

Correlações parciais. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suporta estimativas bootstrap para a média e o desvio padrão.
- A tabela Correlações suporta estimativas bootstrap para correlações.

Regressão linear. Os recursos a seguir são suportados:

- A tabela Estatísticas Descritivas suporta estimativas bootstrap para a média e o desvio padrão.

- A tabela Correlações suporta estimativas bootstrap para correlações.
- A tabela Sumarização do Modelo suporta estimativas bootstrap para Durbin-Watson.
- A tabela Coeficientes suporta estimativas de bootstrap e testes de significância para o coeficiente, B.
- A tabela Coeficientes de Correlação suporta estimativas bootstrap para correlações.
- A tabela de Estatísticas de Resíduos suporta estimativas bootstrap para a média e o desvio padrão.

Regressão ordinal. Os recursos a seguir são suportados:

- A tabela Estimativas de Parâmetro suporta estimativas bootstrap e testes de significância para o coeficiente, B.

Análise discriminante. Os recursos a seguir são suportados:

- A tabela Coeficientes de Função Discriminante Canônica Padronizados suporta estimativas bootstrap para coeficientes padronizados.
- A tabela Coeficientes de Função Discriminante Canônica suporta estimativas bootstrap para coeficientes não padronizados.
- A tabela Coeficientes de Função de Classificação suporta estimativas bootstrap para coeficientes.

SPSS Statistics Premium Edition e opção de estatísticas avançadas

GLM Multivariado. Os recursos a seguir são suportados:

- A tabela Estimativas de Parâmetro suporta estimativas bootstrap e testes de significância para o coeficiente, B.

Modelos Lineares Mistos. Os recursos a seguir são suportados:

- A tabela Estimativas de Efeitos Fixos suporta estimativas bootstrap e testes de significância para a estimativa.
- A tabela Estimativas de Parâmetros de Covariância suporta estimativas bootstrap e testes de significância para a estimativa.

Modelos lineares generalizados. Os recursos a seguir são suportados:

- A tabela Estimativas de Parâmetro suporta estimativas bootstrap e testes de significância para o coeficiente, B.

Regressão de Cox. Os recursos a seguir são suportados:

- A tabela Variáveis na Equation suporta estimativas bootstrap e testes de significância para o coeficiente, B.

SPSS Statistics Standard Edition e opção de regressão

Regressão logística binária. Os recursos a seguir são suportados:

- A tabela Variáveis na Equation suporta estimativas bootstrap e testes de significância para o coeficiente, B.

Regressão logística multinomial. Os recursos a seguir são suportados:

- A tabela Estimativas de Parâmetro suporta estimativas bootstrap e testes de significância para o coeficiente, B.

Recursos adicionais do comando BOOTSTRAP

O idioma da sintaxe de comando também permite:

- Execute amostragens bootstrap residuais e naturais (subcomando AMOSTRAGEM)

Consulte a *Referência da sintaxe de comando* para obter informações de sintaxe completa.

Avisos

Essas informações foram desenvolvidas para produtos e serviços oferecidos nos Estados Unidos. Esse material pode estar disponível a partir da IBM em outros idiomas. No entanto, pode ser necessário possuir uma cópia do produto ou da versão do produto nesse idioma para acessá-lo.

É possível que a IBM não ofereça produtos, serviços ou recursos discutidos neste documento em outros países. Consulte um representante IBM local para obter informações sobre produtos e serviços disponíveis atualmente em sua área. Qualquer referência a produtos, programas ou serviços IBM não significa que apenas produtos, programas ou serviços IBM possam ser utilizados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser utilizado em substituição a este produto, programa ou serviço. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do Cliente.

A IBM pode ter patentes ou solicitações de patentes pendentes relativas a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença podem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
CEP 22290-240
Rio de Janeiro, RJ
Brasil

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO ÀS GARANTIAS IMPLÍCITAS DE NÃO-VIOLAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias explícitas ou implícitas em certas transações; portanto, esta instrução pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar o(s) produto(s) e/ou programa(s) descritos nesta publicação, sem aviso prévio.

Qualquer referência nestas informações a websites não IBM são fornecidas apenas por conveniência e não representam de forma alguma um endosso a esses websites. Os materiais contidos nesses websites não fazem parte dos materiais para esse produto IBM e o uso desses websites é de inteira responsabilidade do Cliente.

A IBM por usar ou distribuir as informações fornecidas da forma que julgar apropriada sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações sobre o mesmo com o objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) o uso mútuo de informações trocadas, devem entrar em contato com:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
CEP 22290-240
Rio de Janeiro, RJ
Brasil

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriadas, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito nesta publicação e todo o material licenciado disponível são fornecidos pela IBM sob os termos do Contrato com o Cliente IBM, do Contrato Internacional de Licença do Programa IBM ou de qualquer outro contrato equivalente.

Os exemplos de dados de desempenho e do Cliente citados são apresentados apenas para propósitos ilustrativos. Resultados de desempenho reais podem variar dependendo das configurações específicas e das condições operacionais.

Informações relativas a produtos não IBM foram obtidas junto aos fornecedores dos respectivos produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou esses produtos e não pode confirmar a precisão de desempenho, compatibilidade nem qualquer outra reivindicação relacionada a produtos não IBM. Perguntas sobre os recursos de produtos não IBM devem ser endereçadas aos fornecedores desses produtos.

Instruções relativas à direção futura ou intento da IBM estão sujeitas a mudança ou retirada sem aviso e representam metas e objetivos apenas.

Estas informações contêm exemplos de dados e relatórios utilizados nas operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de assuntos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com pessoas ou empresas reais é mera coincidência.

LICENÇA DE COPYRIGHT:

Estas informações contêm programas de aplicativos de amostra na linguagem fonte, ilustrando as técnicas de programação em diversas plataformas operacionais. O Cliente pode copiar, modificar e distribuir estes programas de amostra sem a necessidade de pagar à IBM, com objetivos de desenvolvimento, utilização, marketing ou distribuição de programas aplicativos em conformidade com a interface de programação de aplicativo para a plataforma operacional para a qual os programas de amostra são criados. Esses exemplos não foram testados completamente em todas as condições. Portanto, a IBM não pode garantir ou implicar a confiabilidade, manutenção ou função destes programas. Os programas de amostra são fornecidos "NO ESTADO EM QUE SE ENCONTRAM", sem garantia de qualquer tipo. A IBM não será responsabilizada por quaisquer danos decorrentes do uso dos programas de amostra.

Cada cópia ou parte destes programas de amostra ou qualquer trabalho derivado deve incluir um aviso de copyright com os dizeres:

© nome de sua empresa) (ano). Partes deste código são derivadas dos Programas de Amostra da IBM Corp.

Marcas comerciais

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em muitos países no mundo todo. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. A lista atual de marcas comerciais da IBM está disponível na web em "Copyright and trademark information" em www.ibm.com/legal/copytrade.shtml.

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas registradas ou marcas comerciais da Adobe Systems Incorporated nos Estados Unidos e/ou em outros países.

Intel, o logotipo Intel, Intel Inside, o logotipo Intel Inside, Intel Centrino, o logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas comerciais ou marcas registradas da Intel Corporation ou de suas subsidiárias nos Estados Unidos e em outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos, e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas comerciais da Microsoft Corporation nos Estados Unidos e/ou em outros países.

UNIX é uma marca registrada da The Open Group nos Estados Unidos e em outros países.

Java e todas as marcas comerciais e logotipos baseados em Java são marcas comerciais ou marcas registradas da Oracle e/ou suas afiliadas.

Índice Remissivo

A

autoinicialização 2
procedimentos suportados 3



Impresso no Brasil