

*IBM SPSS Missing Values 25*

**IBM**

**Comunicado**

Antes de usar estas informações e o produto suportado por elas, leia as informações nos “Avisos” na página 21.

**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

<b>Valores omissos . . . . .</b>	<b>1</b>
Introdução a valores omissos . . . . .	1
Análise de valor ausente . . . . .	1
Exibindo padrões de valores omissos . . . . .	3
Exibindo estatísticas descritivas para valores omissos . . . . .	4
Estimando estatísticas e imputando valores omissos . . . . .	5
Recursos adicionais do comando MVA. . . . .	7
Imputação múltipla . . . . .	8
Analisar padrões . . . . .	9
Imputar valores de dados omissos. . . . .	10

Recursos adicionais do comando MULTIPLE IMPUTATION . . . . .	14
Trabalhando com dados de imputação múltipla . . . . .	14
Analisando Vários Dados de Entrada. . . . .	15
Opções de imputação múltipla . . . . .	18

<b>Avisos . . . . .</b>	<b>21</b>
Marcas comerciais . . . . .	23

<b>Índice Remissivo . . . . .</b>	<b>25</b>
-----------------------------------	-----------



---

## Valores omissos

Os recursos de valores omissos a seguir são incluídos na opção SPSS Statistics Edição Premium ou Valores Omissos.

---

### Introdução a valores omissos

Os casos com valores omissos representam um desafio importante, porque os procedimentos de modelagem típicos simplesmente descartam esses casos da análise. Quando houver poucos valores omissos (muito aproximadamente, menos de 5% do número total de casos) e esses valores puderem ser considerados como sendo omissos aleatoriamente; ou seja, se um valor é omissos e não depende de outros valores, o método típico de exclusão de listwise é relativamente "seguro". Os valores omissos podem ajudar a determinar se a exclusão listwise é suficiente e fornece métodos para manipulação de valores omissos quando ela não é.

### Análise de valor omissos versus procedimentos de Imputação múltipla

O tópico Valores Omissos fornece dois conjuntos de procedimentos para manipulação de valores omissos:

- Os procedimentos de Imputação múltipla fornecem a análise de padrões de dados omissos, orientada para uma eventual imputação múltipla de valores omissos. Ou seja, são produzidas várias versões do conjunto de dados, cada uma contendo seu próprio conjunto de valores imputados. Quando análises estatísticas são executadas, as estimativas paramétrica para todos os conjuntos de dados imputados são agrupadas, fornecendo estimativas que geralmente são mais precisas do que seriam com apenas uma imputação.
- Análise de valor omissos fornece um conjunto um pouco diferente de ferramentas descritivas para analisar dados omissos (mais especificamente o teste MCAR de Little) e inclui uma variedade de métodos de imputação única. Observe que a imputação múltipla geralmente é considerada superior à imputação única.

### Tarefas de valores omissos

É possível iniciar com a análise de valores omissos seguindo esses passos básicos:

1. **Examinar omissão.** Use Análise de valor omissos e Analisar padrões para explorar padrões de valores omissos em seus dados e determinar se a imputação múltipla é necessária.
2. **Imputar valores omissos.** Use Imputar valores de dados omissos para fazer a imputação múltipla de valores omissos.
3. **Analisar dados "completos".** Use qualquer procedimento que suporte dados de imputação múltipla. Consulte "Analisando Vários Dados de Entrada" na página 15 para obter informações sobre como analisar conjuntos de dados de imputação múltipla e uma lista de procedimentos que suportam esses dados.

---

### Análise de valor ausente

O procedimento Análise de valor omissos executa três funções principais:

- Descreve o padrão de dados omissos. Onde estão localizados os valores omissos? Qual a sua abrangência? Os pares de variáveis tendem a ter valores omissos em vários casos? Os valores de dados são extremos? Os valores são omissos aleatoriamente?
- Estima médias, desvios padrão, covariâncias e correlações para diferentes métodos de valor omissos: listwise, pairwise, regressão ou EM (expectativa-maximização). O método pairwise também exibe contagens de casos completos pairwise.

- Preenche (imputa) valores omissos com valores estimados usando métodos de regressão ou EM; no entanto, a imputação múltipla geralmente é considerada para fornecer resultados mais precisos.

A análise de valor omissos ajuda a direcionar várias preocupações causadas por dados incompletos. Se os casos com valores omissos forem sistematicamente diferentes dos casos sem valores omissos, os resultados podem ser enganosos. Além disso, os dados omissos podem reduzir a precisão de estatísticas calculadas, porque há menos informações do que o originalmente planejado. Outra preocupação é que as suposições atrás de muitos procedimentos estatísticos são baseadas em casos completos, e os valores omissos podem complicar a teoria necessária.

**Exemplo.** Na avaliação de um tratamento para leucemia, diversas variáveis são medidas. No entanto, nem todas as medições estão disponíveis para cada paciente. Os padrões de dados omissos são exibidos, tabulados e considerados aleatórios. Uma análise EM é usada para estimar as médias, correlações e covariâncias. Ela também é usada para determinar se os dados estão totalmente omissos de forma aleatória. Os valores omissos são então substituídos por valores imputados e salvos em um novo arquivo de dados para análise adicional.

**Estatísticas.** As estatísticas univariadas, incluindo o número de valores não omissos, média, desvio padrão, número de valores omissos e número de valores extremos. Médias estimadas, matriz de covariâncias e matriz de correlações, usando métodos de lista, de pares, EM ou de regressão. Teste MCAR de Little com resultados de EM. Sumarização de médias por vários métodos. Para grupos definidos por valores omissos versus não omissos: testes *t*. Para todas as variáveis: os padrões de valores omissos exibiram casos por variáveis.

Considerações de dados

**Dados.** Os dados podem ser categóricos ou quantitativos (de escala ou contínuos). No entanto, é possível estimar estatísticas e imputar dados omissos apenas para as variáveis quantitativas. Para cada variável, os valores omissos que não são codificados como omissos de sistema devem ser definidos como omissos de usuário. Por exemplo, se um item do questionário tiver a resposta *Não sei* codificada como 5 e você desejar tratá-lo como omissos, o item deverá ter 5 codificado como um valor omissos de usuário.

**Ponderações de frequência.** As ponderações de frequência (replicação) são consideradas por esse procedimento. Os casos com valor de ponderação de replicação negativo ou zero são ignorados. As ponderações de número não inteiro são truncadas.

**Suposições.** A estimação de lista, de pares e de regressão depende da suposição de que o padrão de valores omissos não depende dos valores dos dados. (Essa condição é conhecida como **totalmente omissos de forma aleatória**, ou MCAR.) Portanto, todos os métodos (incluindo o método EM) para estimação fornecem estimativas consistentes e imparciais das correlações e covariâncias quando os dados são MCAR. A violação da suposição MCAR pode conduzir a estimativas imparciais produzidas pelos métodos de lista, de pares e de regressão. Se os dados não forem MCAR, será necessário usar a estimação EM.

A estimação EM depende a suposição de que o padrão de dados omissos está relacionado apenas aos dados observados. (Essa condição é chamada **omissos de forma aleatória**, ou MAR.) Essa suposição permite que estimativas sejam ajustadas usando as informações disponíveis. Por exemplo, em um estudo de educação e receita, as pessoas com pouca educação podem ter mais valores de receita omissos. Nesse caso, os dados são MAR, não MCAR. Em outras palavras, para MAR, a probabilidade de que a receita seja registrada depende do nível de educação da pessoa. A probabilidade pode variar por educação, mas não por receita *nesse nível de educação*. Se a probabilidade de que a receita seja registrada também varia pelo valor de receita em cada nível de educação (por exemplo, pessoas com receitas altas não as relatam), os dados não serão MCAR nem MAR. Esta não é uma situação incomum e, se aplicável, nenhum dos métodos é apropriado.

**Procedimentos relacionados.** Muitos procedimentos permitem usar a estimação de lista ou de pares. A regressão linear e a análise fatorial permitem a substituição de valores omissos pelos valores médios. No módulo complementar Previsão, vários métodos estão disponíveis para substituir valores omissos em séries temporais.

Para obter a análise de valor omissos

1. Nos menus, escolha:

**Analisar > Análise de valor omissos...**

2. Selecione pelo menos uma variável quantitativa (escala) para estimar estatísticas e, opcionalmente, imputar valores omissos.

Como opção, você pode:

- Selecione variáveis categóricas (numéricas ou de sequência de caracteres) e insira um limite no número de categorias (**Máximo de categorias**).
- Clique em **Padrões** para tabular os padrões de dados omissos. Consulte o tópico “Exibindo padrões de valores omissos” para obter mais informações
- Clique em **Descritivos** para exibir estatísticas descritivas de valores omissos. Consulte o tópico “Exibindo estatísticas descritivas para valores omissos” na página 4 para obter mais informações
- Selecione um método para estimar estatísticas (médias, covariâncias e correlações) e, possivelmente, imputar valores omissos. Consulte o tópico “Estimando estatísticas e imputando valores omissos” na página 5 para obter mais informações
- Se você selecionar EM ou Regressão, clique em **Variáveis** para especificar um subconjunto a ser usado para a estimação. Consulte o tópico “Variáveis preditas e preditoras” na página 7 para obter mais informações
- Selecione uma variável de rótulo de caso. Essa variável é usada para rotular casos em tabelas de padrões que exibem casos individuais.

## Exibindo padrões de valores omissos

É possível optar por exibir várias tabelas que mostram os padrões e extensão de dados omissos. Essas tabelas podem ajudar a identificar:

- Se os valores omissos foram localizados
- Se pares de variáveis tendem a ter valores omissos em casos individuais
- Se os valores de dados são extremos

Exibir

Três tipos de tabelas estão disponíveis para exibição de padrões de dados omissos.

**Casos tabulados.** Os padrões de valor omissos nas variáveis de análise são tabulados, com frequências mostradas para cada padrão. Use **Ordenar variáveis por padrão de valor omissos** para especificar se contagens e variáveis são ordenadas por similaridade de padrões. Use **Omitir padrões com menos de n % de casos** para eliminar padrões que não ocorrem com frequência.

**Casos com valores omissos.** Cada caso com um valor omissos ou extremo é tabulado para cada variável de análise. Use **Ordenar variáveis por padrão de valor omissos** para especificar se contagens e variáveis são ordenadas por similaridade de padrões.

**Todos os casos.** Cada caso é tabulado e valores omissos e extremos são indicados para cada variável. Os casos são listados na ordem em que aparecem no arquivo de dados, a menos que uma variável seja especificada em **Ordenar por**.

Nas tabelas que exibem casos individuais, são usados os seguintes símbolos:

- + . Valor extremamente alto
- . Valor extremamente baixo

S. Valor omissso do sistema

A. Primeiro tipo de valor omissso de usuário

B. Segundo tipo de valor omissso de usuário

C. Terceiro tipo de valor omissso de usuário

Variáveis

É possível exibir informações adicionais para as variáveis que estão incluídas na análise. As variáveis incluídas em **Informações adicionais para** são exibidas individualmente na tabela de padrões omissos. Para variáveis quantitativas (escala), a média é exibida; para variáveis categóricas, é exibido o número de casos que possuem o padrão em cada categoria.

- **Ordenar por.** Os casos são listados de acordo com a ordem crescente ou decrescente dos valores da variável especificada. Disponível apenas para **Todos os casos**.

Para exibir padrões de valor omissso

1. Na caixa de diálogo principal Análise de valor omissso, selecione as variáveis para as quais você deseja exibir padrões de valor omissso.
2. Clique em **Padrões**.
3. Selecione as tabelas de padrões que você deseja exibir.

## Exibindo estatísticas descritivas para valores omissos

Estatísticas univariadas

As estatísticas univariadas podem ajudar a identificar a extensão geral de dados omissos. Para cada variável, é exibido o seguinte:

- Número de valores não omissos
- Número e porcentagem de valores omissos

Para variáveis quantitativas (escala), o seguinte também é exibido:

- Média
- Desvio padrão
- Número de valores extremamente altos e baixos

Estatísticas de variável indicadora

Para cada variável, é criada uma variável indicadora. Esta variável categórica indica se a variável está presente ou omissa para um caso individual. As variáveis indicadoras são usadas para criar a incompatibilidade, o teste  $t$  e as tabelas de frequências.

**Percentual de incompatibilidade.** Para cada par de variáveis, exibe a porcentagem de casos nos quais uma variável possui um valor omissso e a outra variável possui um valor não omissso. Cada elemento da diagonal na tabela contém a porcentagem de valores omissos para uma única variável.

**Testes  $t$  com grupos formados por variáveis indicadoras.** As médias de dois grupos são comparadas para cada variável quantitativa, usando a estatística  $t$  de Student. Os grupos especificam se uma variável está presente ou omissa. A estatística  $t$ , os graus de liberdade, as contagens de valores omissos e não



omissos e as médias dos dois grupos são exibidos. Também é possível exibir probabilidades de duas partes associadas à estatística  $t$ . Se sua análise resultar em mais de um teste, não use essas probabilidades para o teste de significância. As probabilidades são apropriadas apenas quando um único teste é calculado.

**Tabulações cruzadas de variáveis categóricas e indicadoras.** É exibida uma tabela para cada variável categórica. Para cada categoria, a tabela mostra a frequência e a porcentagem de valores não omissos para as outras variáveis. As porcentagens de cada tipo de valor omissos também são exibidas.

**Omitir variáveis omissas em menos de n % dos casos.** Para reduzir o tamanho da tabela, é possível omitir estatísticas que são calculadas apenas para um pequeno número de casos.

Para exibir estatísticas descritivas

1. Na caixa de diálogo principal Análise de valor omissos, selecione as variáveis para as quais você deseja exibir estatísticas descritivas de valor omissos.
2. Clique em **Descritivos**.
3. Escolha as estatísticas descritivas que você deseja exibir.

## Estimando estatísticas e imputando valores omissos

É possível optar por estimar médias, desvios padrão, covariâncias e correlações usando métodos de lista (apenas casos completos), pairwise, EM (expectativa-maximização) e/ou de regressão. Também é possível optar por imputar os valores omissos (estimar valores de substituição). Observe que a Imputação múltipla geralmente é considerada superior à imputação única para resolver o problema de valores omissos. O teste MCAR de Little ainda é útil para determinar se a imputação é necessária.

Método listwise

Este método usa apenas casos completos. Se alguma das variáveis de análise tiver valores omissos, o caso será omitido dos cálculos.

Método pairwise

Esse método examina pares de variáveis de análise e usa um caso apenas se ele tiver valores não omissos para ambas as variáveis. Frequências, médias e desvios padrão são calculados separadamente para cada par. Como outros valores omissos no caso são ignorados, as correlações e covariâncias para duas variáveis não dependem de valores omissos em nenhuma outra variável.

Método EM

Esse método assume uma distribuição para os dados parcialmente omissos e baseia as inferências na probabilidade dessa distribuição. Cada iteração consiste em um passo E e em um passo M. O passo E localiza a expectativa condicional dos dados "omissos", dados os valores observados e as eparamétrica atuais dos parâmetros. Essas expectativas são então substituídas para os dados "omissos". No passo M, as eparamétrica de máxima verossimilhança dos parâmetros são calculadas como se os dados omissos tivessem sido preenchidos. "Omissos" está entre aspas porque os valores omissos não estão sendo diretamente preenchidos. Em vez disso, as funções deles são usadas no log da verossimilhança.

A estatística qui-quadrado de Roderick J. A. Little para testar se os valores são completamente omissos aleatoriamente (MCAR) é impressa como uma nota de rodapé nas matrizes EM. Para esse teste, a hipótese nula é que os dados são completamente omissos aleatoriamente, e o valor  $p$  é significativo no nível 0,05. Se o valor for menor que 0,05, os dados não serão completamente omissos aleatoriamente. Os dados podem ser omissos aleatoriamente (MAR) ou não omissos aleatoriamente (NMAR). Não é possível considerar um ou o outro e precisar analisar os dados para determinar como os dados são omissos.

Método de regressão

Esse método calcula eparamétrica de regressão linear e tem opções para aumentar as eparamétrica com componentes aleatórios. Para cada valor predito, o procedimento pode incluir um resíduo de um caso completo selecionado aleatoriamente, um desvio normal aleatório ou um desvio aleatório (escalado pela raiz quadrada do quadrado médio residual) da distribuição  $t$ .

### Opções de estimação EM

Usando um processo iterativo, o método EM estima as médias, a matriz de covariâncias e a correlação de variáveis quantitativas (escala) com valores omissos.

**Distribuição.** O EM faz inferências com base na probabilidade na distribuição especificada. Por padrão, uma distribuição normal é considerada. Se você souber que os rodapés da distribuição não maiores que os de uma distribuição normal, será possível solicitar que o procedimento construa a função de probabilidade a partir de uma distribuição  $t$  de Student com  $n$  graus de liberdade. A distribuição normal mista também fornece uma distribuição com rodapés maiores. Especifique a razão dos desvios padrão da distribuição normal mista e a proporção de mistura das duas distribuições. A distribuição normal mista considera que somente os desvios padrão das distribuições diferem. As médias devem ser iguais.

**Máximo de iterações.** Configura o número máximo de iterações para estimar a covariância verdadeira. O procedimento para quando esse número de iterações é atingido, mesmo que as estimativas não tenham convergido.

**Salvar dados completos.** É possível salvar o conjunto de dados com os valores imputados no lugar dos valores omissos. No entanto, saiba que as estatísticas baseadas em covariância usando os valores imputados irão subestimar seus respectivos valores paramétrica. O grau de subestimação é proporcional ao número de casos que são conjuntamente não observados.

Para especificar opções de EM

1. Na caixa de diálogo principal Análise de valor omissos, selecione as variáveis para as quais você deseja estimar valores omissos usando o método EM.
2. Selecione **EM** no grupo Estimação.
3. Para especificar variáveis preditas e predictoras, clique em **Variáveis**. Consulte o tópico “Variáveis preditas e predictoras” na página 7 para obter mais informações
4. Clique em **EM**.
5. Selecione as opções de EM desejadas.

### Opções de estimação de regressão

O método de regressão estima valores omissos usando a regressão linear múltipla. As médias, a matriz de covariâncias e a matriz de correlações das variáveis preditas são exibidas.

**Ajustamento de estimação.** O método de regressão pode incluir um componente aleatório para estimativas de regressão. É possível selecionar resíduos, variáveis normais, variáveis  $t$  de Student ou nenhum ajustamento.

- *Residuais.* Os termos do erro são escolhidos aleatoriamente a partir dos resíduos de casos completos observados a serem incluídos nas estimativas de regressão.
- *Variáveis Normais.* Os termos de erro são retirados aleatoriamente de uma distribuição com o valor esperado 0 e o desvio padrão igual à raiz quadrada do termo de erro quadrático médio da regressão.
- *Variáveis  $t$  de Student.* Termos de erro são retirados aleatoriamente de uma distribuição  $t$  com graus de liberdade especificados, e escalados pelo erro quadrático médio raiz (RMSE).

**Número máximo de preditores.** Configura um limite máximo para o número de variáveis predictoras (independentes) usadas no processo de estimação.

**Salvar dados completos.** Grava um conjunto de dados na sessão atual ou em um arquivo de dados externo do IBM® SPSS Statistics, com valores omissos substituídos por valores estimados pelo método de regressão.

Para especificar opções de regressão

1. Na caixa de diálogo principal Análise de valor omissos, selecione as variáveis para as quais você deseja estimar valores omissos usando o método de regressão.
2. Selecione **Regressão** no grupo Estimação.
3. Para especificar variáveis preditas e preditoras, clique em **Variáveis**. Consulte o tópico “Variáveis preditas e preditoras” para obter mais informações
4. Clique em **Regressão**.
5. Selecione as opções de regressão desejadas.

### **Variáveis preditas e preditoras**

Por padrão, todas as variáveis quantitativas são usadas para EM e estimação de regressão. Se necessário, é possível escolher variáveis específicas como variáveis preditas e preditoras nas estimações. Uma determinada variável pode estar em ambas as listas, mas há situações em que talvez você queira restringir o uso de uma variável. Por exemplo, alguns analistas não se sentem confortáveis ao estimar valores de variáveis de resultado. Talvez você também queira usar diferentes variáveis para diferentes estimações e executar o procedimento várias vezes. Por exemplo, se você tiver um conjunto de itens que são classificações de enfermeiros e outro conjunto que são classificações de médicos, talvez queira fazer uma execução usando o item de enfermeiros para estimar itens omissos de enfermeiros e outra execução para estimativa dos itens de médicos.

Outra consideração surge ao usar o método de regressão. Em regressão múltipla, o uso de um grande subconjunto de variáveis independentes pode produzir valores preditos mais simples do que um subconjunto menor. Portanto, uma variável deve atingir um limite *F-to-enter* de 4,0 a ser usado. Esse limite pode ser mudado com a sintaxe.

Para especificar variáveis preditas e preditoras

1. Na caixa de diálogo principal Análise de valor omissos, selecione as variáveis para as quais você deseja estimar valores omissos usando o método de regressão.
2. Selecione **EM** ou **Regressão** no grupo Estimação.
3. Clique em **Variáveis**.
4. Se desejar usar específicas em vez de todas as variáveis como variáveis preditas e preditoras, selecione **Selecionar variáveis** e mova as variáveis para a lista ou listas apropriadas.

### **Recursos adicionais do comando MVA**

O idioma da sintaxe de comando também permite:

- Especifique variáveis descritivas separadas para padrões de valor omissos, padrões de dados e padrões tabulados usando a palavra-chave DESCRIBE nos subcomandos MPATTERN, DPATTERN ou TPATTERN.
- Especifique mais de uma variável de ordenação para a tabela de padrões de dados usando o subcomando DPATTERN.
- Especifique mais de uma variável de ordenação para padrões de dados usando o subcomando DPATTERN.
- Especifique a tolerância e a convergência usando o subcomando EM.
- Especifique a tolerância e *F-to-enter*, usando o subcomando REGRESSION.
- Especifique listas de variáveis diferentes para EM e Regressão usando os subcomandos EM e REGRESSION.
- Especifique diferentes porcentagens para suprimir casos exibidos para cada um de TTESTS, TABULATE e MISMATCH.

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

## Imputação múltipla

O propósito de imputação múltipla é gerar valores possíveis para valores omissos, criando, portanto, vários conjuntos de dados "completos". Os procedimentos analíticos que funcionam com conjuntos de dados de imputação múltipla produzem saída para cada conjunto de dados "completo", além da saída combinada que estima quais seriam os resultados se o conjunto de dados original não tivesse valores omissos. Esses resultados combinados geralmente são mais precisos do que os fornecidos por métodos de imputação única.

Considerações de dados de imputação múltipla












**Variáveis de análise.** As variáveis de análise podem ser:

- *Nominal.* Uma variável pode ser tratada como nominal quando seus valores representarem categorias sem ranqueamento intrínseco (por exemplo, o departamento da empresa na qual um funcionário trabalha). Exemplos de variáveis nominais incluem região, código de endereçamento postal e filiação religiosa.
- *Ordinal.* Uma variável pode ser tratada como ordinal quando seus valores representarem categorias com algum ranqueamento intrínseco (por exemplo, níveis de satisfação de serviço de muito insatisfeito para muito satisfeito). Exemplos de variáveis ordinais incluem escores de atitude que representam o grau de satisfação ou de confiança e os escores de classificação de preferência.
- *Escala.* Uma variável pode ser tratada como escala (contínua) quando os seus valores representarem categorias ordenadas com uma métrica significativa, de forma que as comparações de distância entre os valores sejam apropriadas. Exemplos de variáveis de escala incluem idade em anos e rendimento em milhares de dólares.

O procedimento supõe que o nível de medição apropriado foi designado a todas as variáveis; no entanto, é possível mudar temporariamente o nível de medição para uma variável, clicando com o botão direito na variável na lista de variáveis de origem e selecionando um nível de medição no menu pop-up. Para mudar permanentemente o nível de medição para uma variável,

Um ícone próximo a cada variável na lista de variáveis identifica o nível de medição e o tipo de dados:

Tabela 1. Ícones do nível de medição

	Numérico	Sequência de caracteres	Data	Horário
Escala (Contínua)		n/a		
Ordinal				
Nominal				

**Ponderações de frequência.** As ponderações de frequência (replicação) são consideradas por esse procedimento. Os casos com valor de ponderação de replicação negativo ou zero são ignorados. As ponderações de número não inteiro são arredondadas para o número inteiro mais próximo.

**Ponderação de Análise.** As ponderações de análise (regressão ou amostragem) são incorporadas em sumarizações de valores omissos e em modelos de imputação de ajuste. Os casos com uma ponderação de análise negativa ou zero são excluídos.

**Amostras complexas.** O procedimento Imputação múltipla não trata explicitamente estratos, clusters ou outras estruturas de amostragem complexa, embora possa aceitar ponderações de amostragem final em forma da variável de ponderação de análise. Observe também que os procedimentos de Amostragem complexa atualmente não analisam automaticamente múltiplos conjuntos de dados imputados. Para obter uma lista completa de procedimentos que suportam conjunto, consulte “Analisando Vários Dados de Entrada” na página 15.

**Valores omissos.** Os valores omissos de usuário e do sistema são tratados como valores inválidos; ou seja, os dois tipos de valores omissos são substituídos quando valores são imputados e ambos são tratados como valores inválidos de variáveis usadas como preditores de modelos de imputação. Os valores omissos de usuário e do sistema também são tratados como omissos em análises de valores omissos.

**Replicando resultados (Imputar valores de dados omissos).** Se desejar replicar seus resultados de imputação de forma exata, use o mesmo valor de inicialização para o gerador de números aleatórios, a mesma ordem de dados e a mesma ordem da variável, além de usar as mesmas configurações do procedimento.

- **Geração de número aleatório.** O procedimento usa a geração de números aleatórios durante o cálculo de valores imputados. Para reproduzir os mesmos resultados escolhidos aleatoriamente no futuro, use o mesmo valor de inicialização para o gerador de números aleatórios antes de cada execução do procedimento Imputar valores de dados omissos.
- **Ordem de casos.** Os valores são imputados em ordem de caso.
- **Ordem da variável.** O método de imputação de especificação totalmente condicional (FCS) imputa valores na ordem especificada na lista Variáveis de análise.

Há dois diálogos dedicados à imputação múltipla.

- Analisar padrões fornece medidas descritivas dos padrões de valores omissos nos dados e pode ser útil como um passo exploratório antes da imputação.
- Imputar valores de dados omissos é usado para gerar imputações múltiplas. Os conjuntos de dados completos podem ser analisados com procedimentos que suportam conjuntos de dados de imputação múltipla. Consulte “Analisando Vários Dados de Entrada” na página 15 para obter informações sobre como analisar conjuntos de dados de imputação múltipla e uma lista de procedimentos que suportam esses dados.

## **Analisar padrões**

Analisar padrões fornece medidas descritivas dos padrões de valores omissos nos dados e pode ser útil como um passo exploratório antes da imputação.

**Exemplo.** Um provedor de telecomunicações deseja entender melhor os padrões de uso de serviço em seu banco de dados de clientes. Eles possuem dados completos para serviços usados por seus clientes, mas as informações demográficas coletadas pela empresa possuem diversos valores omissos. A análise dos padrões de valores omissos pode ajudar a determinar os próximos passos para imputação.

Nos menus, escolha:

**Analisar > Imputação múltipla > Analisar padrões...**

1. Selecione pelo menos duas variáveis de análise. O procedimento analisa padrões de dados omissos para essas variáveis.

Configurações Opcionais

**Ponderação de Análise.** Essa variável contém ponderações de análise (regressão ou amostragem). O procedimento incorpora ponderações de análise em sumarizações de valores omissos. Os casos com uma ponderação de análise negativa ou zero são excluídos.

**Saída.** A seguinte saída opcional está disponível:

- **Sumarização de valores omissos.** Isso exibe um gráfico de pizza em painel que mostra o número e percentual de variáveis de análise, casos ou valores de dados individuais que possuem um ou mais valores omissos.
- **Padrões de valores omissos.** Isso exibe padrões tabulados de valores omissos. Cada padrão corresponde a um grupo de casos com o mesmo padrão de dados incompletos e completos em variáveis de análise. É possível usar essa saída para determinar se o método de imputação monótono pode ser usado para seus dados, ou se não puder, o quanto seus dados se aproximam de um padrão monótono. O procedimento ordena variáveis de análise para revelar ou aproximar um padrão monotônico. Se não existir nenhum padrão não monótono após a reordenação, será possível concluir que os dados possuem um padrão monotônico quando as variáveis de análise forem ordenadas dessa forma.
- **Variáveis com a frequência mais alta de valores omissos.** Isso exibe uma tabela de variáveis de análise ordenadas por percentual de valores omissos em ordem decrescente. A tabela inclui estatísticas descritivas (média e desvio padrão) para variáveis de escala.

É possível controlar o número máximo de variáveis a serem exibidas e a porcentagem mínima omissa para uma variável a ser incluída na exibição. É exibido o conjunto de variáveis que atendem aos dois critérios. Por exemplo, configurar o número máximo de variáveis como 50 e a porcentagem mínima omissa como 25 solicita que a tabela exiba até 50 variáveis que têm pelo menos 25% de valores omissos. Se houver 60 variáveis de análise mas somente 15 tiverem 25% ou mais de valores omissos, a saída incluirá apenas 15 variáveis.

## Imputar valores de dados omissos

Imputar valores de dados omissos é usado para gerar imputações múltiplas. Os conjuntos de dados completos podem ser analisados com procedimentos que suportam conjuntos de dados de imputação múltipla. Consulte “Analisando Vários Dados de Entrada” na página 15 para obter informações sobre como analisar conjuntos de dados de imputação múltipla e uma lista de procedimentos que suportam esses dados.

**Exemplo.** Um provedor de telecomunicações deseja entender melhor os padrões de uso de serviço em seu banco de dados de clientes. Eles possuem dados completos para serviços usados por seus clientes, mas as informações demográficas coletadas pela empresa possuem diversos valores omissos. Além disso, esses valores não são omissos de forma completamente aleatória, portanto, a imputação múltipla será usada para concluir o conjunto de dados.

Nos menus, escolha:

### **Analisar > Imputação múltipla > Imputar valores de dados omissos...**

1. Selecione pelo menos duas variáveis no modelo de imputação. O procedimento imputa valores múltiplos para dados omissos para essas variáveis.
2. Especifique o número de imputações a serem calculadas. Por padrão, esse valor é 5.
3. Especifique um conjunto de dados ou arquivo de dados de formato do IBM SPSS Statistics no qual os dados imputados devem ser gravados.

O conjunto de dados de saída consiste nos dados do caso original com dados omissos, além de um conjunto de casos com valores imputados para cada imputação. Por exemplo, se o conjunto de dados original tiver 100 casos e você tiver cinco imputações, o conjunto de dados de saída terá 600 casos. Todas as variáveis no conjunto de dados de entrada estão incluídas no conjunto de dados de saída. As propriedades de dicionário (nomes, rótulos, etc.) de variáveis existentes são copiadas para o novo

conjunto de dados. O arquivo também contém uma nova variável, *Imputation\_*, uma variável numérica que indica a imputação (0 para dados originais ou 1..n para os casos que possuem valores imputados).

O procedimento define automaticamente a variável *Imputation\_* como uma variável de divisão quando o conjunto de dados de saída é criado. Se as divisões estiverem em vigor quando o procedimento for executado, o conjunto de dados de saída incluirá um conjunto de imputações para cada combinação de valores de variáveis de divisão.

## Configurações Opcionais

**Ponderação de Análise.** Essa variável contém ponderações de análise (regressão ou amostragem). O procedimento incorpora ponderações de análise em modelos de regressão e de classificação usados para imputar valores omissos. As ponderações de análise também são usadas em sumarizações de valores imputados; por exemplo, média, desvio padrão e erro padrão. Os casos com uma ponderação de análise negativa ou zero são excluídos.

## Campos com nível de medição desconhecido

O alerta de Nível de Medição é exibido quando o nível de medição para uma ou mais variáveis (campos) no conjunto de dados é desconhecido. Como o nível de medição afeta o cálculo de resultados para este procedimento, todas as variáveis devem ter um nível de medição definido.

**Dados de varredura.** Lê os dados no conjunto de dados ativo e designa o nível de medição padrão para quaisquer campos com um nível de medição desconhecido atualmente. Se o conjunto de dados for grande, isso poderá demorar algum tempo.

**Designar Manualmente.** Abre um diálogo que lista todos os campos com um nível de medição desconhecido. É possível utilizar este diálogo para designar o nível de medição para esses campos. Também é possível designar o nível de medição na Visualização de Variável do Editor de Dados.

Como o nível de medição é importante para este procedimento, não é possível acessar o diálogo para executar este procedimento até que todos os campos possuam um nível de medição definido.

## Método

A guia Método especifica como os valores omissos serão imputados, incluindo os tipos de modelos usados. Os preditores categóricos são codificados por indicador (simulado).

**Método de imputação.** O método **Automático** varre os dados e usa o método monótono se os dados mostrarem um padrão monótono de valores omissos; caso contrário, será usada a especificação totalmente condicional. Se você tem certeza de qual método deseja usar, será possível especificá-lo como um método **Customizado**.

- **Especificação totalmente condicional.** Esse é um método Markov chain Monte Carlo (MCMC) iterativo que pode ser usado quando o padrão de dados omissos for arbitrário (monótono ou não monótono).

Para cada iteração e para cada variável na ordem especificada na lista de variáveis, o método de especificação totalmente condicional (FCS) ajusta um modelo (variável dependente única) univariado usando todas as outras variáveis disponíveis no modelo como preditores, em seguida, imputa Valores omissos para a variável que está sendo ajustada. O método continua até que o número máximo de iterações seja atingido, e os valores imputados na iteração máxima sejam salvos no conjunto de dados imputados.

**Máximo de iterações.** Isso especifica o número de iterações, ou "passos", executados pela cadeia Markov usada pelo método FCS. Se o método FCS foi escolhido automaticamente, ele usará o número padrão de 10 iterações. Ao escolher explicitamente o FCS, é possível especificar um número customizado de iterações. Pode ser necessário aumentar o número de iterações se a cadeia Markov não

tiver convergido. Na guia Saída, é possível salvar os dados do histórico de iteração FCS e representá-los no gráfico para avaliar a convergência.

- **Monótono.** Esse é um método não iterativo que pode ser usado apenas quando os dados tiverem um padrão monótono de valores omisso. Existe um padrão monótono quando é possível ordenar as variáveis de forma que, se uma variável tiver um valor não omisso, todas as variáveis precedentes também terão valores não omisso. Ao especificar isso como um método **Customizado**, certifique-se de especificar as variáveis na lista em uma ordem que mostre um padrão monótono.

Para cada variável na ordem monótona, o método monótono ajusta um modelo (variável dependente única) univariado usando todas as variáveis anteriores no modelo como preditores, em seguida, imputa valores omisso para a variável que está sendo ajustada. Esses valores imputados são salvos no conjunto de dados imputados.

**Incluir interações de dois fatores.** Quando o método de imputação é escolhido automaticamente, o modelo de imputação para cada variável inclui um termo constante e os principais efeitos para variáveis preditoras. Ao escolher um método específico, opcionalmente, é possível incluir todas as possíveis interações de dois fatores entre variáveis preditoras categóricas.

**Tipo de modelo para variáveis de escala.** Quando o método de imputação é escolhido automaticamente, a regressão linear é usada como o modelo univariado para variáveis de escala. Ao escolher um método específico, como alternativa, é possível escolher a correspondente de média preditiva (PMM) como o modelo para variáveis de escala. PMM é uma variante de regressão linear que corresponde valores imputados calculados pelo modelo de regressão ao valor observado mais próximo.

A regressão logística é sempre usada como o modelo univariado para variáveis categóricas. Independentemente do tipo de modelo, os preditores categóricos são tratados usando a codificação de indicador (simulado).

**Tolerância à singularidade.** Matrizes singulares (ou não inversíveis) possuem colunas linearmente dependentes que podem causar sérios problemas para o algoritmo de estimação. Como até mesmo matrizes quase singulares podem levar a resultados ruins, o procedimento tratará como singular uma matriz cuja determinante for menor que a tolerância. Especifique um valor positivo.

## Restrições

A guia Restrições permite restringir a função de uma variável durante a imputação e restringir o intervalo de valores imputados de uma variável de escala para que eles são plausíveis. Além disso, é possível restringir a análise a variáveis com uma porcentagem menor que o máximo de valores omisso.

**Varredura de dados para sumarização de variável.** Clicar em **Varrer dados** faz a lista mostrar variáveis de análise e o percentual omisso, mínimo e máximo observado para cada uma. As sumarizações podem ser baseadas em todos os casos ou limitadas a uma varredura dos primeiros  $n$  casos, conforme especificado na caixa de texto Casos. Clicar em **Varrer dados novamente** atualiza as sumarizações de distribuição.

Definir restrições

- **Papel.** Isso permite customizar o conjunto de variáveis a serem imputadas e/ou tratadas como preditoras. Geralmente, cada variável de análise é considerada como um dependente e preditor no modelo de imputação. A **Função** pode ser usada para desativar a imputação para variáveis que você deseja **Usar apenas como preditor** ou excluir variáveis de serem usadas como preditores (**Apenas imputar**) e, portanto, tornar o modelo de predição mais compacto. Essa é a única restrição que pode ser especificada para variáveis categóricas, ou para variáveis que são usadas apenas como preditores.
- **Mín. e Máx.** Essas colunas permitem especificar os valores imputados mínimo e máximo permitidos para variáveis de escala. Se um valor imputado estiver fora desse intervalo, o procedimento exibirá outro valor até encontrar um no intervalo ou até o número máximo de exibições ser atingido (consulte **Máximo de exibições** abaixo). Essas colunas estarão disponíveis apenas se **Regressão linear** estiver selecionado como o tipo de modelo de variável de escala na guia Método.



- **Arredondamento.** Algumas variáveis podem ser usadas como escala, mas possuem valores que são naturalmente mais restritos; por exemplo, o número de pessoas em uma família deve ser um número inteiro e a quantia gasta durante uma visita ao supermercado não pode ter centavos fracionários. Essa coluna permite especificar o menor valor a ser aceito. Por exemplo, para obter valores de número inteiro, especifique 1 como o valor de arredondamento; para obter valores arredondados para o centavo mais próximo, especifique 0,01. Em geral, os valores são arredondados para o múltiplo de número inteiro mais próximo do valor de arredondamento. A tabela a seguir mostra como diferentes valores de arredondamento agem em um valor imputado de 6,64823 (antes do arredondamento).

Tabela 2. Arredondando resultados.

Valor de arredondamento	Valor para o qual 6,64832 é arredondado
10	10
1	7
0.25	6,75
0,1	6,6
0,01	6,65

**Excluir variáveis com grandes quantidades de dados omissos.** Geralmente, as variáveis de análise são imputadas e usadas como preditores, sem considerar quantos valores omissos elas possuem, desde que tenham dados suficientes para estimar um modelo de imputação. É possível optar por excluir variáveis que têm uma alta porcentagem de valores omissos. Por exemplo, se você especificar 50 como a **Porcentagem máxima omissa**, as variáveis de análise que tiverem mais de 50% de valores omissos não serão imputadas, nem serão usadas como preditores em modelos de imputação.

**Máximo de exibições.** Se os valores mínimo ou máximo forem especificados para valores imputados de variáveis de escala (consulte **Mín. e Máx.** acima), o procedimento tentará exibir valores para um caso até encontrar um conjunto de valores que estão nos intervalos especificados. Se um conjunto de valores não for obtido dentro do número especificado de exibições por caso, o procedimento exibirá outro conjunto paramétricas de modelo e repetirá o processo de exibição de casos. Ocorrerá um erro se um conjunto de valores nos intervalos não for obtido dentro do número especificado de exibições de casos e paramétricas.

Observe que aumentar esses valores pode aumentar o tempo de processamento. Se o procedimento estiver demorando muito ou se não puder localizar exibições apropriadas, verifique os valores mínimo e máximo especificados para assegurar que eles sejam apropriados.

## Saída

**Exibição.** Controla a exibição de saída. Uma sumarização de imputação geral é sempre exibida, que inclui tabelas relacionadas a especificações de imputação, iterações (para o método de especificação totalmente condicional), variáveis dependentes imputadas, variáveis dependentes excluídas da imputação e sequência de imputação. Se especificado, as restrições para variáveis de análise também serão mostradas.

- **Modelo de imputação.** Isso exibe o modelo de imputação para variáveis dependentes e preditores, e inclui o tipo de modelo univariado, efeitos do modelo e número de valores imputados.
- **Estatísticas descritivas.** Isso exibe estatísticas descritivas para variáveis dependentes para as quais os valores são imputados. Para variáveis de escala, as estatísticas descritivas incluem média, contagem, desvio padrão, mínimo e máximo para os dados de entrada originais (antes da imputação), valores imputados (por imputação) e dados completos (valores originais e imputados juntos — por imputação). Para variáveis categóricas, as estatísticas descritivas incluem contagem e percentual por categoria para os dados de entrada originais (antes da imputação), valores imputados (por imputação) e dados completos (valores originais e imputados juntos — por imputação).

**Histórico de iteração.** Quando o método de imputação de especificação totalmente condicional é usado, é possível solicitar um conjunto de dados que contém os dados do histórico de iteração para imputação

FCS. O conjunto de dados contém médias e desvios padrão por iteração e imputação para cada variável dependente de escala para a qual os valores são imputados. É possível plotar os dados para ajudar a avaliar a convergência de modelo.

## Recursos adicionais do comando MULTIPLE IMPUTATION

O idioma da sintaxe de comando também permite:

- Especifique um subconjunto de variáveis para as quais as estatísticas descritivas são mostradas (subcomando IMPUTATIONSUMMARIES).
- Especifique uma análise de padrões omissos e imputação em uma única execução do procedimento.
- Especifique o número máximo paramétricas de modelo permitidos ao imputar qualquer variável (palavra-chave MAXMODELPARAM ).

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

## Trabalhando com dados de imputação múltipla

Quando um conjunto de dados de imputação múltipla (MI) é criado, é incluída uma variável chamada *Imputation\_*, com o rótulo de variáveis *Imputation Number*, e o conjunto de dados é ordenado em ordem crescente. Os casos do conjunto de dados original possuem um valor 0. Os casos para valores imputados são numerados de 1 a *M*, em que *M* é o número de imputações.

Ao abrir um conjunto de dados, a presença de *Imputation\_* identifica o conjunto de dados como um possível conjunto de dados de MI.

Ativando um conjunto de dados de imputação múltipla para análise

O conjunto de dados deve ser dividido usando a opção **Comparar grupos**, com *Imputation\_* como uma variável de agrupamento, para ser tratada como um conjunto de dados de MI na análise. Também é possível definir divisões em outras variáveis.

Nos menus, escolha:

### Dados > Arquivo dividido...

1. Selecione **Comparar grupos**.
2. Selecione *Número de imputação [Imputation\_]* como uma variável na qual agrupar casos.

Como alternativa, ao ativar marcações (consulte abaixo), o arquivo é dividido em *Número de imputação [Imputation\_]*.

Distinguindo valores imputados de valores observados

É possível distinguir valores imputados de valores observados por cor do plano de fundo da célula, a fonte e negrito (para valores imputados). Ao criar um novo conjunto de dados na sessão atual com Imputar valores omissos, as marcações são ativadas por padrão. Ao abrir um arquivo de dados salvo que inclui imputações, as marcações são desativadas.

Para ativar marcações, nos menus do Editor de dados, escolha:

### Visualizar > Marcar dados imputados...

Como alternativa, é possível ativar marcações clicando no botão de marcação de imputação na borda direita da barra de edição na Visualização de dados do Editor de dados.

Movendo-se entre imputações

1. Nos menus, escolha:  
**Editar > Acessar imputação...**
2. Selecione a imputação (ou dados originais) da lista suspensa.

Como alternativa, é possível selecionar a imputação da lista suspensa na barra de edição na Visualização de dados do Editor de dados.

A posição relativa do caso é preservada durante a seleção de imputações. Por exemplo, se houver 1000 casos no conjunto de dados original, o caso 1034, o 34º caso na primeira imputação, será exibido na parte superior da grade. Se você selecionar imputação 2 na lista suspensa, o caso 2034, o 34º caso na imputação 2, será exibido na parte superior da grade. Se você selecionar **Dados originais** na lista suspensa, o caso 34 será exibido na parte superior da grade. A posição da coluna também é preservada durante a navegação entre imputações, para que fique fácil comparar valores entre imputações.

Transformando e editando valores imputados

Às vezes, será necessário executar transformações nos dados imputados. Por exemplo, talvez você queira obter o log de todos os valores de uma variável de salário e salvar o resultado em uma nova variável. Um valor calculado usando dados imputados será tratado como imputado se diferir do valor calculado usando os dados originais.

Se você editar um valor imputado em uma célula no Editor de dados, essa célula ainda será tratada como imputada. Não é recomendado editar valores imputados desta maneira.

## Analizando Vários Dados de Entrada

Muitos procedimentos suportam a definição de conjunto de resultados da análise de multiplicação de conjuntos de dados imputados. Quando as marcações de imputação são ativadas, é exibido um ícone especial próximo aos procedimentos que suportam a definição de conjunto. No submenu Estatística descritiva do menu Analisar, por exemplo, Frequências, Descritivos, Exploração e Tabulações cruzadas suportam a definição de conjunto, enquanto Razão, Gráficos P-P e Gráficos Q-Q não suportam.

O resultado tabular e o PMML de modelo podem ser agrupados. Não há nenhum novo procedimento para solicitar o resultado agrupado; em vez disso, uma nova guia no diálogo Opções oferece controle global sobre o resultado de imputação múltipla.

- **Definição de conjunto de resultado tabular.** Por padrão, ao executar um procedimento suportado em um conjunto de dados de imputação múltipla (MI), os resultados são automaticamente produzidos para cada imputação, os dados originais (não imputados) e os resultados agrupados (finais) que consideram a variação entre as imputações. As estatísticas agrupadas variam por procedimento.
- **Definição de conjunto de PMML.** Também é possível obter o PMML agrupado de procedimentos suportados que exportam PMML. O PMML agrupado é solicitado da mesma forma que um PMML não agrupado e, em vez disso, é salvo.

Os procedimentos não suportados não produzem resultado agrupado nem arquivos PMML agrupados.

## Níveis de conjunto

O resultado é agrupado usando um de dois níveis:

- **Combinação Naïve.** Apenas o parâmetro agrupado está disponível.
- **Combinação Univariada.** O parâmetro agrupado, seu erro padrão, estatística do teste e graus de liberdade efetivos, valor  $p$ , intervalo de confiança e diagnósticos de definição de conjunto (fração de informações omissas, eficiência relativa, aumento relativo na variação) são mostrados quando disponíveis.

Coefficientes (regressão e correlação), médias (e diferenças médias) e contagens geralmente são agrupados. Quando o erro padrão da estatística está disponível, a definição de conjunto univariada é usada; caso contrário, é usada a definição de conjunto naïve.

## **Procedimentos que suportam conjunto**

Os procedimentos a seguir suportam conjuntos de dados MI, nos níveis de definição de conjunto especificada para cada parte do resultado.

**Frequências.** Os seguintes recursos são suportados:

- A tabela Estatísticas suporta definição de conjunto Médias em Univariado (se a média S.E. também for solicitada) e definição de conjunto N Válido e N Omisso em Naïve.
- A tabela Frequências suporta definição de conjunto Frequência em Naïve.

**Descritivos.** Os seguintes recursos são suportados:

- A tabela Estatística descritiva suporta definição de conjunto Médias em Univariado (se a média S.E. também for solicitada) e definição de conjunto N em Naïve.

**Tabulações cruzadas.** Os seguintes recursos são suportados:

- A tabela Tabulação cruzada suporta definição de conjunto Contagem em Naïve.

**Médias.** Os seguintes recursos são suportados:

- A tabela Relatório suporta definição de conjunto Média em Univariado (se a média S.E. também for solicitada) e definição de conjunto N em Naïve.

**Teste T de uma amostra.** Os seguintes recursos são suportados:

- A tabela Estatísticas suporta definição de conjunto Média em Univariado e definição de conjunto N em Naïve.
- A tabela Teste suporta definição de conjunto Diferença média em Univariado.

**Teste T de amostras independentes.** Os seguintes recursos são suportados:

- A tabela Estatísticas do grupo suporta definição de conjunto Médias em Univariado e definição de conjunto N em Naïve.
- A tabela Teste suporta definição de conjunto Diferença média em Univariado.

**Teste T de amostras pairwise.** Os seguintes recursos são suportados:

- A tabela Estatísticas suporta definição de conjunto Médias em Univariado e definição de conjunto N em Naïve.
- A tabela Correlações suporta definição de conjunto Correlações e N em Naïve.
- A tabela Teste suporta definição de conjunto Média em Univariado.

**ANOVA unidirecional.** Os seguintes recursos são suportados:

- A tabela Estatística descritiva suporta definição de conjunto Média em Univariado e definição de conjunto N em Naïve.
- A tabela Testes de contraste suporta definição de conjunto Valor de contraste em Univariado.

**UNIANOVA.** Os seguintes recursos são suportados:

- Descritivos em conjunto Naïve.
- Fatores entre sujeitos em conjunto Naïve.
- Estimativas de parâmetro em conjunto univariado.
- Médias marginais estimadas em conjunto univariado.

- Comparações de médias marginais estimadas em conjunto univariado.

**GLM Univariados.** Os seguintes recursos são suportados:

- A tabela Estimativas paramétrica suporta definição de conjunto o coeficiente,  $B$ , em Univariado.

**Modelos Lineares Mistos.** Os seguintes recursos são suportados:

- A tabela Estatística descritiva suporta definição de conjunto Média e  $N$  em Naïve.
- A tabela Estimativas de efeitos fixos suporta definição de conjunto Estimativa em Univariado.
- A tabela Estimativas paramétricas de covariância suporta definição de conjunto Estimativa em Univariado.
- A tabela Médias marginais estimadas: Estimativas suporta definição de conjunto Média em Univariado.
- A tabela Médias marginais estimadas: Comparações entre pares suporta definição de conjunto Diferença média em Univariado.

**Modelos lineares generalizados e Equações de estimativa generalizada.** Esses procedimentos suportam PMML agrupado.

- A tabela Informações de variável categórica suporta definição de conjunto  $N$  e Percentuais em Naïve.
- A tabela Informações de variável contínua suporta definição de conjunto  $N$  e Média em Naïve.
- A tabela Estimativas paramétrica suporta definição de conjunto o coeficiente,  $B$ , em Univariado.
- A tabela Médias marginais estimadas: Coeficientes de estimação suporta definição de conjunto Média em Naïve.
- A tabela Médias marginais estimadas: Estimativas suporta definição de conjunto Média em Univariado.
- A tabela Médias marginais estimadas: Comparações entre pares suporta definição de conjunto Diferença média em Univariado.

**Correlações bivariadas.** Os seguintes recursos são suportados:

- A tabela Estatística descritiva suporta definição de conjunto Média e  $N$  em Naïve.
- A tabela Correlações suporta definição de conjunto Correlações e  $N$  em Univariado. Observe que as correlações são transformadas usando a transformação  $z$  de Fisher antes da definição de conjunto e, em seguida, são transformadas de volta após a definição de conjunto.

**Correlações parciais.** Os seguintes recursos são suportados:

- A tabela Estatística descritiva suporta definição de conjunto Média e  $N$  em Naïve.
- A tabela Correlações suporta definição de conjunto Correlações em Naïve.

**Regressão linear.** Esse procedimento suporta PMML agrupado.

- A tabela Estatística descritiva suporta definição de conjunto Média e  $N$  em Naïve.
- A tabela Correlações suporta definição de conjunto Correlações e  $N$  em Naïve.
- A tabela Coeficientes suporta definição de conjunto  $B$  em Univariado e definição de conjunto Correlações em Naïve.
- A tabela Coeficientes de correlação suporta definição de conjunto Correlações em Naïve.
- A tabela Estatísticas de resíduos suporta definição de conjunto Média e  $N$  em Naïve.

**Regressão logística binária.** Esse procedimento suporta PMML agrupado.

- A tabela Variáveis na equação suporta definição de conjunto  $B$  em Univariado.

**Regressão logística multinomial.** Esse procedimento suporta PMML agrupado.

- A tabela Estimativas paramétrica suporta definição de conjunto o coeficiente,  $B$ , em Univariado.

**Regressão ordinal.** Os seguintes recursos são suportados:

- A tabela Estimativas paramétrica suporta definição de conjunto o coeficiente, B, em Univariado.

**Análise discriminante.** Esse procedimento suporta XML de modelo agrupado.

- A tabela Estatísticas de grupo suporta definição de conjunto Média e N Válido em Naïve.
- A tabela Matrizes agrupadas em grupos suporta definição de conjunto Correlações em Naïve.
- A tabela Coeficientes de função discriminante canônica suporta definição de conjunto Coeficientes não padronizados em Naïve.
- A tabela Funções em centroides de grupo suporta definição de conjunto Coeficientes não padronizados em Naïve.
- A tabela Coeficientes de função de classificação suporta definição de conjunto Coeficientes em Naïve.

**Teste qui-quadrado.** Os seguintes recursos são suportados:

- A tabela Descritivos suporta definição de conjunto Média e N em Naïve.
- A tabela Frequências suporta definição de conjunto N Observado em Naïve.

**Teste binômio.** Os seguintes recursos são suportados:

- A tabela Descritivos suporta definição de conjunto Média e N em Naïve.
- A tabela Teste suporta definição de conjunto N, Proporção observada e Proporção de teste em Naïve.

**Teste de execuções.** Os seguintes recursos são suportados:

- A tabela Descritivos suporta definição de conjunto Média e N em Naïve.

**Teste Kolmogorov-Smirnov de uma amostra.** Os seguintes recursos são suportados:

- A tabela Descritivos suporta definição de conjunto Média e N em Naïve.

**Testes de duas amostras independentes.** Os seguintes recursos são suportados:

- A tabela Ranqueamentos suporta definição de conjunto Ranqueamento médio e N em Naïve.
- A tabela Frequências suporta definição de conjunto N em Naïve.

**Testes para várias amostras independentes.** Os seguintes recursos são suportados:

- A tabela Ranqueamentos suporta definição de conjunto Ranqueamento médio e N em Naïve.
- A tabela Frequências suporta definição de conjunto Contagens em Naïve.

**Testes de duas amostras relacionadas.** Os seguintes recursos são suportados:

- A tabela Ranqueamentos suporta definição de conjunto Ranqueamento médio e N em Naïve.
- A tabela Frequências suporta definição de conjunto N em Naïve.

**Testes para várias amostras relacionadas.** Os seguintes recursos são suportados:

- A tabela Ranqueamentos suporta definição de conjunto Ranqueamento médio em Naïve.

**Regressão de Cox.** Esse procedimento suporta PMML agrupado.

- A tabela Variáveis na equação suporta definição de conjunto B em Univariado.
- A tabela Médias covariáveis suporta definição de conjunto Média em Naïve.

## Opções de imputação múltipla

A guia Imputação múltipla controla dois tipos de preferências relacionadas a Imputações múltiplas:

**Aparência de dados imputados.** Por padrão, as células contendo dados imputados terão uma cor do plano de fundo diferente das células contendo dados não imputados. A aparência diferenciada dos dados

imputados deve facilitar a rolagem por um conjunto de dados e a localização dessas células. É possível mudar a cor do plano de fundo da célula padrão, a fonte, e fazer os dados imputados serem exibidos em negrito.

**Saída de análise.** Esse grupo controla o tipo de saída do Visualizador produzida sempre que um conjunto de dados imputados múltiplas vezes é analisado. Por padrão, a saída será produzida para o conjunto de dados original (pré-imputação) e para cada um dos conjuntos de dados imputados. Além disso, para os procedimentos que suportam a definição de conjunto de dados imputados, os resultados combinados finais serão gerados. Quando a definição de conjunto univariada for executada, os diagnósticos de definição de conjunto também serão exibidos. No entanto, é possível suprimir qualquer saída que você não deseja ver.

Para configurar opções de imputação múltipla

Nos menus, escolha:

**Editar > Opções**

Clique na guia Imputação múltipla.





---

## 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](http://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](http://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

- Analisar padrões 9
- Análise de valor ausente 1
  - EM 6
  - estatísticas descritivas 4
  - estimando estatísticas 5
  - expectativa-maximização 7
  - imputando valores omissos 5
  - métodos 5
  - padrões 3
  - recursos adicionais do comando 7
  - regressão 6
  - teste MCAR 5

## C

- categorias de tabulação
  - na Análise de valor omissos 4
- contagens de valor extremo
  - na Análise de valor omissos 4
- correlações
  - na Análise de valor omissos 6
- covariância
  - na Análise de valor omissos 6

## D

- dados incompletos
  - consulte Análise de valor omissos 1
- desvio padrão
  - na Análise de valor omissos 4

## E

- EM
  - na Análise de valor omissos 6
- especificação totalmente condicional em Imputação múltipla 11
- exclusão de listwise
  - na Análise de valor omissos 1
- exclusão dos pares
  - na Análise de valor omissos 1

## H

- histórico de iteração
  - em Imputação múltipla 13

## I

- imputação monótona
  - em Imputação múltipla 11
- imputação múltipla 8
  - analisar padrões 9
  - imputar valores de dados omissos 10
- Imputação múltipla 14, 15
- Imputar valores de dados faltantes 10
  - método de imputação 11

- Imputar valores de dados faltantes (*continuação*)
  - restrições 12
  - saída 13
- incompatibilidade
  - na Análise de valor omissos 4

## M

- média
  - na Análise de valor omissos 4, 6

## O

- ordenando casos
  - na Análise de valor omissos 3

## R

- regressão
  - na Análise de valor omissos 6
- resíduos
  - na Análise de valor omissos 6

## T

- tabelas de frequências
  - na Análise de valor omissos 4
- tabulando casos
  - na Análise de valor omissos 3
- teste MCAR
  - na Análise de valor omissos 1
- teste MCAR de Little 5
  - na Análise de valor omissos 1
- teste t
  - na Análise de valor omissos 4
- teste t de Student
  - na Análise de valor omissos 6

## V

- valores omissos
  - estatísticas univariadas 4
- variáveis de indicadores
  - na Análise de valor omissos 4
- variáveis indicadoras omissas
  - na Análise de valor omissos 4
- variáveis normais
  - na Análise de valor omissos 6







Impresso no Brasil