

*IBM SPSS Direct Marketing 28*

**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 28, 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.

© **Copyright International Business Machines Corporation .**

---

# Índice

<b>Capítulo 1. Marketing Direto.....</b>	<b>1</b>
Marketing direto.....	1
Análise RFM.....	1
Escores de RFM de dados de transação.....	2
Escores de RFM para dados do cliente.....	2
Categorização RFM.....	3
Salvando escores de RFM de dados de transação.....	5
Salvando escores de RFM de dados do cliente.....	5
Saída de RFM.....	6
Análise de cluster.....	7
Configurações .....	8
Perfis de clientes em potencial.....	9
Configurações .....	10
Criando um campo de resposta categórica.....	11
Taxas de resposta do código de endereçamento postal.....	11
Configurações .....	13
Criando um campo de resposta categórica.....	14
Propensão para compra.....	14
Configurações .....	16
Criando um campo de resposta categórica.....	17
Teste de Pacote de Controle.....	18
<b>Avisos.....</b>	<b>21</b>
Marcas comerciais.....	22
<b>Índice Remissivo.....</b>	<b>25</b>



---

# Capítulo 1. Marketing Direto

Os seguintes recursos de marketing direto estão incluídos na opção SPSS Statistics Premium Edition ou Direct Marketing.

## Marketing direto

---

A opção Marketing direto fornece um conjunto de ferramentas designadas para melhorar os resultados de campanhas de marketing direto, identificando características demográficas, de compra e outras características que definem vários grupos de consumidores e têm como alvo grupos específicos para aumentar as taxas de respostas positivas.

**Análise de RFM.** Essa técnica identifica clientes existentes que têm maior probabilidade de responder a uma nova oferta.

**Análise de cluster.** Esta é uma ferramenta exploratória projetada para revelar agrupamentos naturais (ou clusters) em seus dados. Por exemplo, pode identificar diferentes grupos de clientes com base em diversas características demográficas e de compra.

**Perfis de clientes em potencial.** Essa técnica usa resultados de uma campanha anterior ou de teste para criar perfis descritivos. É possível usar os perfis para focar em grupos específicos de contatos em campanhas futuras. Consulte o tópico [“Perfis de clientes em potencial” na página 9](#) para obter mais informações

**Taxas de resposta de código de endereçamento postal.** Esta técnica usa resultados de uma campanha anterior para calcular as taxas de resposta de código postal. Essas taxas podem ser usadas para focar em códigos postais específicos em campanhas futuras. Consulte o tópico [“Taxas de resposta do código de endereçamento postal” na página 11](#) para obter mais informações

**Propensão para compra.** Essa técnica usa resultados de um teste de envio ou campanha anterior para gerar escores de propensão. As pontuações indicam quais contatos têm maior probabilidade de responder. Consulte o tópico [“Propensão para compra” na página 14](#) para obter mais informações

**Teste do pacote de controle.** Esta técnica compara campanhas de marketing para verificar se há uma diferença significativa em efetividade para diferentes pacotes ou ofertas. Consulte o tópico [“Teste de Pacote de Controle” na página 18](#) para obter mais informações

## Análise RFM

---

A análise de RFM é uma técnica usada para identificar clientes existentes que têm maior probabilidade de responder a uma nova oferta. Esta técnica é comumente usada em marketing direto. A análise de RFM é baseada na seguinte teoria simples:

- O fator mais importante na identificação de clientes que têm probabilidade de responder a uma nova oferta é **recência**. Os clientes que compraram mais recentemente têm maior probabilidade de comprar novamente do que os clientes que compraram mais no passado.
- O segundo fator mais importante é **frequência**. Os clientes que fizeram mais compras no passado têm maior probabilidade de responder do que os que fizeram menos compras.
- O terceiro fator mais importante é o valor total gasto, que é referido como **monetário**. Os clientes que gastaram mais (no total para todas as compras) no passado têm maior probabilidade de responder do que os que gastaram menos.

Como funciona a Análise de RFM

- Os clientes são designados a um escore de recência com base na data da compra mais recente ou no intervalo de tempo desde a compra mais recente. Esse escore é baseado em um posto simples de valores de recência em um pequeno número de categorias. Por exemplo, se você usar cinco categorias,

os clientes com as datas de compra mais recentes receberão um ranqueamento de recência de 5, e os com datas de compra mais distantes no passado receberão um ranqueamento de recência de 1.

- De forma semelhante, os clientes são então designados a um ranqueamento de frequência, com valores mais altos representando uma frequência mais alta de compras. Por exemplo, em um esquema de ranqueamento de categoria cinco, os clientes que compraram com mais frequência recebem um ranqueamento de frequência de 5.
- Por fim, os clientes são ranqueados por valor monetário, com os valores monetários mais altos recebendo o ranqueamento mais alto. Continuando o exemplo de categoria cinco, os clientes que gastaram mais recebem um ranqueamento monetário de 5.

O resultado são quatro escores para cada cliente: escore de recência, de frequência, monetário e de RFM combinado, que são simplesmente os três escores individuais concatenados em um único valor. Os "melhores" clientes (os que têm maior probabilidade de responder a uma oferta) são os que têm os escores de RFM combinados mais altos. Por exemplo, em um ranqueamento de categoria cinco, há um total de 125 possíveis escores de RFM combinados, e o escore de RFM combinado mais alto é 555.

Considerações de dados

- Se as linhas de dados representam transações (cada linha representa uma única transação e pode haver várias transações para cada cliente), use RFM de Transações. Consulte o tópico [“Escore de RFM de dados de transação”](#) na página 2 para obter mais informações.
- Se as linhas de dados representam clientes com informações de sumarização para todas as transações (com colunas que contêm valores para o valor total gasto, número total de transações e data de transação mais recente), use RFM de Dados do cliente. Consulte o tópico [“Escore de RFM para dados do cliente”](#) na página 2 para obter mais informações.

## Escores de RFM de dados de transação

Considerações de dados

O conjunto de dados deve conter variáveis que contêm as seguintes informações:

- Uma variável ou combinação de variáveis que identificam cada caso (cliente).
- Uma variável com a data de cada transação.
- Uma variável com o valor monetário de cada transação.

Criando escores de RFM a partir de dados de transação

1. Nos menus, escolha:

**Marketing direto > Escolher técnica**

2. Selecione **Ajudar a identificar meus melhores contatos (Análise de RFM)** e clique em **Continuar**.

3. Selecione **Dados de transação** e clique em **Continuar**.

4. Selecione a variável que contém datas de transação.

5. Selecione a variável que contém o valor monetário para cada transação.

6. Selecione o método para sumarizar valores de transação para cada cliente: Total (soma de todas as transações), média, mediana ou máximo (valor mais alto de transação).

7. Selecione a variável ou combinação de variáveis que identifica exclusivamente cada cliente. Por exemplo, os casos podem ser identificados por um código de ID exclusivo ou uma combinação do sobrenome e do nome.

## Escores de RFM para dados do cliente

Considerações de dados

O conjunto de dados deve conter variáveis que contêm as seguintes informações:

- A data de compra mais recente ou um intervalo de tempo desde a data de compra mais recente. Isso será usado para calcular os escores de recência.

- Número total de compras. Isso será usado para calcular os escores de frequência.
- Valor monetário de sumarização para todas as compras. Isso será usado para calcular escores monetários. Normalmente, essa é a soma (total) de todas as compras, mas poderia ser a média (média), máximo (maior valor) ou outra medida de sumarização.

Se desejar gravar escores de RFM em um novo conjunto de dados, o conjunto de dados ativo também deve conter uma variável ou combinação de variáveis que identificam cada caso (cliente).

Criando escores de RFM a partir de dados do cliente

1. Nos menus, escolha:

**Marketing direto > Escolher técnica**

2. Selecione **Ajudar a identificar meus melhores contatos (Análise de RFM)** e clique em **Continuar**.
3. Selecione **Dados do cliente** e clique em **Continuar**.
4. Selecione a variável que contém a data de transação mais recente ou um número que representa um intervalo de tempo desde a transação mais recente.
5. Selecione a variável que contém o número total de transações para cada cliente.
6. Selecione a variável que contém o valor monetário de sumarização para cada cliente.
7. Se desejar gravar escores de RFM em um novo conjunto de dados, selecione a variável ou combinação de variáveis que identifica exclusivamente cada cliente. Por exemplo, os casos podem ser identificados por um código de ID exclusivo ou uma combinação do sobrenome e do nome.

## Categorização RFM

O processo de agrupar uma grande quantidade de valores numéricos em uma pequena quantidade de categorias às vezes é referido como **categorização**. Na análise de RFM, as categorias são as categorias ranqueadas. É possível usar a guia Categorização para modificar o método usado para designar valores de recência, de frequência e monetários a essas categorias.

Método de discretização

**Aninhado.** Na categorização aninhada, um ranqueamento simples é designado a valores de recência. Em cada ranqueamento de recência, os clientes são então designados a um ranqueamento de frequência, e em cada ranqueamento de frequência, os clientes são designados a um ranqueamento monetário. Isso tende a fornecer uma distribuição mais uniforme de escores de RFM combinados, mas tem a desvantagem de tornar os escores de ranqueamento de frequência e monetário mais difíceis de interpretar. Por exemplo, um ranqueamento de frequência de 5 para um cliente com um ranqueamento de recência de 5 pode não significar a mesma coisa que um ranqueamento de frequência de 5 para um cliente com um ranqueamento de recência de 4, pois o ranqueamento de frequência é dependente do ranqueamento de recência.

**Independente.** Os ranqueamentos simples são designados aos valores de recência, de frequência e monetários. Os três ranqueamentos são designados de forma independente. Portanto, a interpretação de cada um dos três componentes de RFM é inequívoca; um escore de frequência de 5 para um cliente significa o mesmo que um escore de frequência de 5 para outro cliente, independentemente de seus escores de recência. Para amostras menores, isso tem a desvantagem de resultar em uma distribuição menos uniforme de escores de RFM combinados.

Número de depósitos

O número de categorias a serem usadas para cada componente para criar escores de RFM. O número total de possíveis escores de RFM combinados é o produto dos três valores. Por exemplo, 5 categorias de recência, 4 categorias de frequência e 3 categorias monetárias criariam um total de 60 possíveis escores de RFM combinados, variando de 111 a 543.

- O padrão é 5 para cada componente, o que criará 125 possíveis escores de RFM combinados, variando de 111 a 555.
- O número máximo de categorias permitido para cada componente de escore é nove.

## Vínculos

Um "empate" é simplesmente dois ou mais valores iguais de recência, de frequência ou monetários. De preferência, você deseja ter aproximadamente o mesmo número de clientes em cada categoria, mas um grande número de valores empatados pode afetar a distribuição de categorias. Existem duas alternativas para tratar empates:

- **Designar empates à mesma categoria.** Esse método sempre designa valores empatados à mesma categoria, independentemente de como isso afeta a distribuição da categoria. Isso fornece um método de categorização consistente: Se dois clientes tiverem o mesmo valor de recência, eles serão sempre designados ao mesmo escore de recência. No entanto, em um exemplo extremo, você pode ter 1.000 clientes, com 500 deles fazendo sua compra mais recente na mesma data. Em um ranqueamento de categoria 5, 50% dos clientes receberiam, portanto, um escore de recência de 5, em vez do valor ideal de 20%.

Observe que, com o método de categorização aninhado, "consistência" é algo mais complicado para escores de frequência e monetários, pois os escores de frequência são designados em categorias de escore de recência, e os escores monetários são designados em categorias de escore de frequência. Portanto, dois clientes com o mesmo valor de frequência não podem ter o mesmo escore de frequência se também não tiverem o mesmo escore de recência, independentemente de como os valores empatados são tratados.

- **Designar empates aleatoriamente.** Isso assegura uma distribuição de categoria uniforme, designando um fator de variância aleatório muito pequeno aos empates antes do ranqueamento; portanto, para o propósito de designar valores às categorias ranqueadas, não existem valores empatados. Este processo não tem efeito sobre os valores originais. Ele é usado apenas para desambiguar os empates. Embora isso produza uma distribuição de categoria uniforme (aproximadamente o mesmo número de clientes em cada categoria), ele pode resultar em resultados de escore completamente diferentes para clientes que parecem ter valores de recência, de frequência e/ou monetários semelhantes ou idênticos -- especificamente se o número total de clientes for relativamente pequeno e/ou o número de empates for relativamente alto.

<b>ID</b>	<b>Compra mais recente (Recência)</b>	<b>Designar empates à mesma categoria</b>	<b>Designar empates aleatoriamente</b>
1	29/10/2006	5	5
2	28/10/2006	4	4
3	28/10/2006	4	4
4	28/10/2006	4	5
5	28/10/2006	4	3
6	21/9/2006	3	3
7	21/9/2006	3	2
8	13/8/2006	2	2
9	13/8/2006	2	1
10	20/6/2006	1	1

- Nesse exemplo, designar empates à mesma categoria resulta em uma distribuição de categoria desigual: 5 (10%), 4 (40%), 3 (20%), 2 (20%), 1 (10%).
- Designar empates aleatoriamente resulta em 20% em cada categoria, mas para obter esse resultado, os quatro casos com um valor de data de 28/10/2006 são designados a 3 categorias diferentes, e os 2 casos com um valor de data de 13/8/2006 também são designados a categorias diferentes.

Observe que a maneira que os empates são designados a diferentes categorias é totalmente aleatória (nas restrições do resultado final sendo um número igual de casos em cada categoria). Se você calculou



um segundo conjunto de escores usando o mesmo método, o ranqueamento para qualquer caso específico com um valor empatado pode mudar. Por exemplo, os ranqueamentos de recência de 5 e 3 para os casos 4 e 5 respectivamente, podem ser alternados pela segunda vez.

## Salvando escores de RFM de dados de transação

O RFM de dados de transação sempre cria um novo conjunto de dados agregado com uma linha para cada cliente. Use a guia Salvar para especificar quais escores e outras variáveis você deseja salvar e onde deseja salvá-los.

### Variáveis

As variáveis de ID que identificam exclusivamente cada cliente são automaticamente salvas no novo conjunto de dados. As seguintes variáveis adicionais podem ser salvas no novo conjunto de dados:

- **Data da transação mais recente para cada cliente.**
- **Número de transações.** O número total de linhas de transação para cada cliente.
- **Valor.** O valor de sumarização para cada cliente com base no método de sumarização selecionado na guia Variáveis.
- **Escore de recência.** O escore designado a cada cliente com base na data de transação mais recente. Escores mais altos indicam datas de transação mais recentes.
- **Escore de frequência.** O escore designado a cada cliente com base no número total de transações. Escores mais altos indicam mais transações.
- **Escore monetário.** O escore designado a cada cliente com base na medida de sumarização monetária selecionada. Escores mais altos indicam um valor mais alto para a medida de sumarização monetária.
- **Escore de RFM.** Os três escores individuais combinados em um único valor:  $(recência \times 100) + (frequência \times 10) + monetário$ .

Por padrão, todas as variáveis disponíveis são incluídas no novo conjunto de dados; portanto, cancele a seleção das que você não deseja incluir. Opcionalmente, é possível especificar seus próprios nomes de variáveis. Os nomes de variáveis devem estar em conformidade com as regras de nomenclatura de variável padrão.

### Localização

O RFM de dados de transação sempre cria um novo conjunto de dados agregado com uma linha para cada cliente. É possível criar um novo conjunto de dados na sessão atual ou salvar os dados de escore de RFM em um arquivo de dados externo. Os nomes de conjuntos de dados devem estar em conformidade com as regras de nomenclatura de variável padrão. (Essa restrição não se aplica a nomes de arquivos de dados externos.)

## Salvando escores de RFM de dados do cliente

Para dados do cliente, é possível incluir as variáveis de escore de RFM no conjunto de dados ativo ou criar um novo conjunto de dados que contenha as variáveis de escores selecionadas. Use a guia Salvar para especificar quais variáveis de escore você deseja salvar e onde deseja salvá-las.

### Nomes das variáveis salvas

- **Gerar nomes exclusivos automaticamente.** Ao incluir variáveis de escore no conjunto de dados ativo, isso assegura que os novos nomes de variáveis sejam exclusivos. Isso é útil, principalmente se você deseja incluir vários conjuntos diferentes de escores de RFM (com base em critérios diferentes) no conjunto de dados ativo.
- **Nomes customizados.** Isso permite designar seus próprios nomes de variáveis às variáveis de escore. Os nomes de variáveis devem estar em conformidade com as regras de nomenclatura de variável padrão.

### Variáveis

Selecione (marque) as variáveis de escore que você deseja salvar:

- **Escore de recência.** O escore designado a cada cliente com base no valor da variável Data de transação ou intervalo selecionada na guia Variáveis. Os escores mais altos são designados a valores de datas mais recentes ou de intervalos inferiores.
- **Escore de frequência.** O escore designado a cada cliente com base na variável Número de transações selecionada na guia Variáveis. Os escores mais altos são designados a valores mais altos.
- **Escore monetário.** O escore designado a cada cliente com base na variável Valor selecionada na guia Variáveis. Os escores mais altos são designados a valores mais altos.
- **Escore de RFM.** Os três escores individuais combinados em um valor único:  $(recency*100)+(frequency*10)+monetary$ .

#### Localização

Para dados do cliente, existem três alternativas de onde é possível salvar novos escores de RFM:

- **Conjunto de dados ativo.** As variáveis de escore de RFM selecionadas são incluídas no conjunto de dados ativo.
- **Novo conjunto de dados.** As variáveis de escore de RFM selecionadas e as variáveis de ID que identificam exclusivamente cada cliente (caso) serão gravadas em um novo conjunto de dados na sessão atual. Os nomes de conjuntos de dados devem estar em conformidade com as regras de nomenclatura de variável padrão. Essa opção estará disponível apenas se você selecionar uma ou mais variáveis de Identificador do cliente na guia Variáveis.
- **Arquivo.** As variáveis de escores de RFM selecionadas e de ID que identificam exclusivamente cada cliente (caso) serão salvas em um arquivo de dados externo. Essa opção estará disponível apenas se você selecionar uma ou mais variáveis de Identificador do cliente na guia Variáveis.

## Saída de RFM

#### Dados armazenados

Os gráficos e tabelas para dados categorizados são baseados na recência calculada, na frequência e em escores monetários.

**Heat map do valor médio monetário por recência e frequência.** O heat map de distribuição monetária média mostra a média de valor monetário para as categorias definidas por escores de recência e de frequência. Áreas mais escuras indicam a média de valor monetário mais alta.

**Gráfico de contagens de categorias.** O gráfico de contagens de bins exibe a distribuição de bin do método de discretização selecionado. Cada barra representa o número de casos aos quais será designado cada escore de RFM combinado.

- Embora, geralmente, você deseje uma distribuição bem uniforme, com todas (ou quase todas) as barras aproximadamente com a mesma altura, uma determinada quantia de variância deve ser esperada ao usar o método de categorização padrão que designa valores de empates para a mesma categoria.
- Flutuações extremas na distribuição de categoria e/ou muitas categorias vazias podem indicar que você deve tentar outro método de categorização (menos categorias e/ou designações aleatórias de empates) ou reconsiderar a adequação da análise de RFM.

**Tabela de contagens de categorias.** As mesmas informações que estão no gráfico de contagens de categorias, exceto que são expressas no formato de uma tabela, com contagens de categorias em cada célula.

#### Dados não armazenados

Os gráficos e tabelas para dados não categorizados são baseados nas variáveis originais usadas para criar recência, frequência e escores monetários.

**Histogramas.** Os histogramas mostram a distribuição relativa de valores para as três variáveis usadas para calcular recência, frequência e escores monetários. Não é incomum para esses histogramas indicar distribuições um pouco defasadas, em vez de uma distribuição normal ou simétrica.

O eixo horizontal de cada histograma é sempre ordenado de valores baixos à esquerda para valores altos à direita. Com a recência, no entanto, a interpretação do gráfico depende do tipo de medida de recência: intervalo de data ou de tempo. Para as datas, as barras à esquerda representam valores adicionais no passado (uma data menos recente possui um valor inferior a uma data mais recente). Para intervalos de tempo, as barras à esquerda representam valores mais recentes (o menor o intervalo de tempo, o mais recente a transação).

**Gráficos de dispersão de pares de variáveis.** Esses gráficos de dispersão mostram os relacionamentos entre as três variáveis usadas para calcular os escores de recência, de frequência e monetários.

É comum ver agrupamentos lineares notáveis de pontos na escala de frequência, pois a frequência geralmente representa uma amplitude relativamente pequena de valores distintos. Por exemplo, se o número total de transações não exceder 15, haverá somente 15 valores de frequência possíveis (a menos que você conte transações fracionárias), enquanto pode haver centenas de valores de recência possíveis e milhares de valores monetários.

A interpretação do eixo de recência depende do tipo de medida de recência: intervalo de data ou hora. Para as datas, os pontos mais próximos à origem representam datas adicionais no passado. Para intervalos de tempo, pontos mais próximos da origem representam valores mais recentes.

## Análise de cluster

---

A análise de clusters é uma ferramenta exploradora projetada para revelar agrupamentos naturais (ou clusters) dentro de seus dados. Por exemplo, pode identificar diferentes grupos de clientes com base em diversas características demográficas e de compra.

**Exemplo.** As empresas varejistas e de produtos de consumo aplicam regularmente técnicas de clusterização a dados que descrevem os hábitos de compra, gênero, idade, nível de renda, etc. de seus clientes. Essas empresas customizam suas estratégias de marketing e de desenvolvimento de produtos a cada grupo de consumidores para aumentar as vendas e construir uma fidelidade à marca.

Considerações de dados de Análise de cluster

**Dados.** Este procedimento funciona com campos contínuos e categóricos. Cada registro (linha) representa um cliente a ser armazenado em cluster e os campos (variáveis) representam atributos nos quais o clusterização é baseado.












**Ordem de registros.** Observe que os resultados podem depender da ordem de registros. Para minimizar os efeitos de ordem, talvez você queira considerar aleatoriamente a ordenação dos registros. Talvez você queira executar a análise várias vezes, com registros ordenados em diferentes ordens aleatórias para verificar a estabilidade de uma determinada solução.

**Nível de medição.** A designação correta do nível de medição é importante porque afeta o cálculo dos resultados.

- *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.
- *Contínuo.* 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.

Um ícone próximo a cada campo indica o nível de medição atual.

Tabela 2. Ícones do nível de medição

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

É possível mudar o nível de medição na Visualização de variável do Editor de dados ou é possível usar o diálogo Definir propriedades da variável para sugerir um nível de medição apropriado para cada campo.

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.

Para obter a Análise de cluster

Nos menus, escolha:

#### Marketing direto > Escolher técnica

1. Selecione **Segmentar meus contatos em clusters**.
2. Selecione os campos categóricos (nominais, ordinais) e contínuos (escala) que você deseja usar para criar segmentos.
3. Clique em **Executar** para executar o procedimento.

## Configurações

A guia Configurações permite mostrar ou suprimir a exibição de gráficos e tabelas que descrevem os segmentos, salvar um novo campo no conjunto de dados que identifica o segmento (cluster) para cada registro no conjunto de dados, e especificar quantos segmentos incluir na solução de cluster.

**Exibir gráficos e tabelas.** Exibe tabelas e gráficos que descrevem os segmentos.

**Associação de segmento.** Salva um novo campo (variável) que identifica o segmento ao qual cada registro pertence.

- Os nomes de campos devem estar de acordo com as regras de nomenclatura do IBM SPSS Statistics.
- O nome do campo de associação do segmento não pode duplicar um nome de campo que já existe no conjunto de dados. Se você executar este procedimento mais de uma vez no mesmo conjunto de dados, precisará especificar um nome diferente a cada vez.
- **Número de segmentos.** Controla como o número de segmentos é determinado.

- **Determinar automaticamente.** O procedimento irá determinar automaticamente o "melhor" número de segmentos, até o máximo especificado.

**Especificar fixo.** O procedimento produzirá o número especificado de segmentos.

## Perfis de clientes em potencial

---

Essa técnica usa resultados de uma campanha anterior ou de teste para criar perfis descritivos. É possível usar os perfis para focar em grupos específicos de contatos em campanhas futuras. O campo Resposta indica quem respondeu à campanha anterior ou de teste. A lista Perfis contém as características que você deseja usar para criar o perfil.

**Exemplo.** Com base nos resultados de um teste de envio, a divisão de marketing direto de uma empresa deseja gerar perfis dos tipos de clientes com maior probabilidade de responder a uma oferta, com base nas informações demográficas.

Saída

A saída inclui uma tabela que fornece uma descrição de cada grupo de perfis e exibe taxas de resposta (porcentagem de respostas positivas) e taxas de resposta acumulativa e um gráfico de taxas de resposta acumulativa. Se você incluir uma taxa de resposta mínima de destino, a tabela será codificada por cores para mostrar quais perfis atendem à taxa de resposta acumulativa mínima, e o gráfico incluirá uma linha de referência no valor de taxa de resposta mínima especificado.

Considerações de dados de perfis de clientes em potencial

**Campo de resposta.** O campo de resposta deve ser nominal ou ordinal. Ele pode ser de sequência de caracteres ou numérico. Se esse campo contiver um valor que indica o número ou valor de compras, será necessário criar um novo campo no qual um único valor representa todas as respostas positivas. Consulte o tópico [“Criando um campo de resposta categórica”](#) na página 11 para obter mais informações

**Valor de resposta positiva.** O valor de resposta positiva identifica os clientes que responderam positivamente (por exemplo, fizeram uma compra). Todos os outros valores de resposta não omissos são considerados para indicar uma resposta negativa. Se houver quaisquer rótulos de valor definidos para o campo de resposta, os rótulos serão exibidos na lista suspensa.












**Criar perfis com.** Esses campos podem ser nominais, ordinais ou contínuos (escala). Eles podem ser uma sequência de caracteres ou numéricos.

**Nível de medição.** A designação correta do nível de medição é importante porque afeta o cálculo dos resultados.

- *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.
- *Contínuo.* 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.

Um ícone próximo a cada campo indica o nível de medição atual.

Tabela 3. Ícones do nível de medição

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

É possível mudar o nível de medição na Visualização de variável do Editor de dados ou é possível usar o diálogo Definir propriedades da variável para sugerir um nível de medição apropriado para cada campo.

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.

Para obter perfis de clientes em potencial

Nos menus, escolha:

#### Marketing direto > Escolher técnica

1. Selecione **Gerar perfis de meus contatos que responderam a uma oferta.**
2. Selecione o campo que identifica quais contatos responderam à oferta. Este campo deve ser nominal ou ordinal.
3. Insira o valor que indica uma resposta positiva. Se quaisquer valores tiverem rótulos de valor definidos, será possível selecionar o rótulo de valor da lista suspensa e o valor correspondente será exibido.
4. Selecione os campos que você deseja usar para criar os perfis.
5. Clique em **Executar** para executar o procedimento.

#### Informações relacionadas

[Exemplos detalhados](#)

## Configurações

A guia Configurações permite controlar o tamanho mínimo do grupo de perfis e incluir um limite mínimo de taxa de resposta na saída.

**Tamanho mínimo do grupo de perfis.** Cada perfil representa as características compartilhadas de um grupo de contatos no conjunto de dados (por exemplo, mulheres com menos de 40 que moram na região oeste). Por padrão, o menor tamanho do grupo de perfis é 100. Tamanhos menores de grupos podem

revelar mais grupos, mas tamanhos maiores de grupos podem fornecer resultados mais confiáveis. O valor deve ser um número inteiro positivo.

**Incluir informações do limite mínimo de taxa de resposta nos resultados.** Os resultados incluem uma tabela que exibe taxas de resposta (porcentagem de respostas positivas) e taxas de resposta acumulativa e um gráfico de taxas de resposta acumulativa. Se você inserir uma taxa de resposta mínima de destino, a tabela será codificada por cores para mostrar quais perfis atendem à taxa de resposta acumulativa mínima e o gráfico incluirá uma linha de referência no valor da taxa de resposta mínima especificada. O valor deve ser maior que 0 e menor que 100.

## Criando um campo de resposta categórica

O campo de resposta deve ser categórico, com um valor representando todas as respostas positivas. Qualquer outro valor não omitido é considerado como sendo uma resposta negativa. Se o campo de resposta representar um valor contínuo (escala), como o número de compras ou valor monetário de compras, será necessário criar um novo campo que designa um único valor de resposta positiva a todos os valores de resposta diferentes de zero.

- Se respostas negativas forem registradas como 0 (não em branco, que é tratado como omitido), isso pode ser calculado com a seguinte fórmula:

$$\text{NewName} = \text{OldName} > 0$$

em que *NewName* é o nome do novo campo e *OldName* é o nome do campo original. Se essa for uma expressão lógica que designa um valor 1 a todos os valores não omitidos maiores que 0, e 0 a todos os valores não omitidos menores ou iguais a 0.

- Se nenhum valor for registrado para respostas negativas, esses valores serão tratados como omitidos, e a fórmula será um pouco mais complicada:

$$\text{NewName} = \text{NOT}(\text{MISSING}(\text{OldName}))$$

Nessa expressão lógica, todos os valores de resposta não omitidos são designados a um valor 1 e todos os valores de resposta omitidos são designados a um valor 0.

- Se não for possível distinguir entre valores de resposta negativa (0) e valores omitidos, um valor de resposta exato não poderá ser calculado. Se houver relativamente alguns valores realmente omitidos, isso pode não ter um efeito significativo nas taxas de resposta calculadas. No entanto, se houver muitos valores omitidos -- como quando as informações de resposta são registradas apenas para uma pequena amostra de teste do conjunto de dados total -- as taxas de resposta calculadas não serão significativas, pois serão significativamente menores que as taxas de resposta verdadeiras.

Para criar um campo de resposta categórica

1. Nos menus, escolha:

**Transformar > Calcular variável**

2. Para Variável de destino, insira o nome do novo campo (variável).
3. Se as respostas negativas forem registradas como 0, para Expressão numérica, insira  $\text{OldName} > 0$ , em que *OldName* é o nome do campo original.
4. Se as respostas negativas forem registradas como omitidas (em branco), para a Expressão numérica, insira  $\text{NOT}(\text{MISSING}(\text{OldName}))$ , em que *OldName* é o nome do campo original.

## Taxas de resposta do código de endereçamento postal

Esta técnica usa resultados de uma campanha anterior para calcular as taxas de resposta de código postal. Essas taxas podem ser usadas para focar em códigos postais específicos em campanhas futuras. O campo Resposta indica quem respondeu à campanha anterior. O campo Código Postal identifica o campo que contém os códigos postais.

**Exemplo.** Com base nos resultados de uma correspondência anterior, a divisão de marketing direto de uma empresa gera taxas de resposta por códigos de endereçamento postal. Com base em vários critérios,

como uma taxa de resposta mínima aceitável e/ou número máximo de contatos a serem incluídos na correspondência, eles podem então direcionar códigos de endereçamento postal específicos.

Saída

A saída desse procedimento inclui um novo conjunto de dados que contém taxas de resposta por código de endereçamento postal, e uma tabela e gráfico que sumarizam os resultados por ranqueamento de decil (superior a 10%, superior a 20%, etc.). A tabela pode ser codificada por cores com base em uma taxa de resposta acumulativa mínima especificada pelo usuário ou no número máximo de contatos.

O novo conjunto de dados contém os seguintes campos:

- **Código de endereçamento postal.** Se os grupos de códigos de endereçamento postal forem baseados apenas em uma parte do valor completo, esse será o valor dessa parte do código de endereçamento postal. O rótulo da linha de cabeçalho para essa coluna no arquivo do Excel é o nome do campo do código de endereçamento postal no conjunto de dados original.
- **ResponseRate.** A porcentagem de respostas positivas em cada código de endereçamento postal.
- **Respostas.** O número de respostas positivas em cada código de endereçamento postal.
- **Contatos.** O número total de contatos em cada código de endereçamento postal que contém um valor não omissos para o campo de resposta.
- **Índice.** A resposta "ponderada" baseada na fórmula  $N \times P \times (1-P)$ , em que  $N$  é o número de contatos e  $P$  é a taxa de resposta expressa como uma proporção.
- **Posição.** Ranqueamento de decil (superior a 10%, superior a 20%, etc.) das taxas de resposta acumulativa de código de endereçamento postal em ordem decrescente.

Considerações de dados de taxas de resposta de código de endereçamento postal

**Campo de resposta.** O campo de resposta pode ser de sequência de caracteres ou numérico. Se esse campo contiver um valor que indica o número ou valor monetário de compras, será necessário criar um novo campo no qual um valor único representa todas as respostas positivas. Consulte o tópico [“Criando um campo de resposta categórica”](#) na página 14 para obter mais informações

**Valor de resposta positiva.** O valor de resposta positiva identifica os clientes que responderam positivamente (por exemplo, fizeram uma compra). Todos os outros valores de resposta não omissos são considerados para indicar uma resposta negativa. Se houver quaisquer rótulos de valor definidos para o campo de resposta, os rótulos serão exibidos na lista suspensa.

**Campo de código de endereçamento postal.** O campo de código de endereçamento postal pode ser de sequência de caracteres ou numérico.

Para obter taxas de resposta do código de endereçamento postal

Nos menus, escolha:

#### Marketing direto > Escolher técnica

1. Selecione **Identificar principais códigos de endereçamento postal de resposta**.
2. Selecione o campo que identifica quais contatos responderam à oferta.
3. Insira o valor que indica uma resposta positiva. Se quaisquer valores tiverem rótulos de valor definidos, será possível selecionar o rótulo de valor da lista suspensa e o valor correspondente será exibido.
4. Selecione o campo que contém o código de endereçamento postal.
5. Clique em **Executar** para executar o procedimento.

Opcionalmente, também é possível:

- Gerar taxas de resposta baseadas nos primeiros  $n$  caracteres ou dígitos do código de endereçamento postal em vez do valor completo
- Salvar automaticamente os resultados em um arquivo do Excel
- Controlar opções de exibição de saída



## Configurações

### Agrupar Códigos Postais com Base em

Isso determina como os registros são agrupados para calcular taxas de resposta. Por padrão, é usado o código de endereçamento postal inteiro, e todos os registros com o mesmo código de endereçamento postal são agrupados para calcular a taxa de resposta do grupo. Como alternativa, é possível agrupar registros com base apenas em uma parte do código de endereçamento postal completo, que consiste nos primeiros  $n$  dígitos ou caracteres. Por exemplo, talvez você queira agrupar registros com base apenas nos primeiros 5 caracteres de um código de endereçamento postal de 10 caracteres ou nos três primeiros dígitos de um código de endereçamento postal de 5 dígitos. O conjunto de dados de saída conterá um registro para cada grupo de códigos de endereçamento postal. Se você inserir um valor, ele deve ser um número inteiro positivo.

### Formato do Código Postal Numérico

Se o campo de código de endereçamento postal for numérico e você desejar agrupar códigos de endereçamento postal com base nos primeiros  $n$  dígitos em vez do valor inteiro, será necessário especificar o número de dígitos no valor original. O número de dígitos é o número *máximo* possível de dígitos no código de endereçamento postal. Por exemplo, se o campo de código de endereçamento postal contiver uma combinação de CEPs de 5 dígitos e de 9 dígitos, será necessário especificar 9 como o número de dígitos.

Nota: Dependendo do formato de exibição, apenas alguns CEPs de 5 dígitos podem parecer conter somente 4 dígitos, mas há um zero à esquerda implícito.

### Saída

Além do novo conjunto de dados que contém taxas de resposta por código de endereçamento postal, é possível exibir uma tabela e um gráfico que resumem os resultados por ranqueamento de decil (superior a 10%, superior a 20%, etc.). A tabela exibe taxas de resposta, taxas de resposta acumulativa, número de registros e número de registros acumulativos em cada decil. O gráfico exibe taxas de resposta acumulativa e o número de registros acumulativos em cada decil.

**Taxa de resposta mínima aceitável.** Se você inserir uma taxa de resposta mínima de destino ou uma fórmula de ponto de equilíbrio, a tabela será codificada por cores para mostrar quais decis atendem à taxa de resposta acumulativa mínima e o gráfico incluirá uma linha de referência no valor da taxa de resposta mínima especificada.

- **Taxa de resposta de destino.** Taxa de resposta expressa como uma porcentagem (porcentagem de respostas positivas em cada grupo de códigos de endereçamento postal). O valor deve ser maior que 0 e menor que 100.
- **Calcular taxa de ponto de equilíbrio da fórmula.** Calcule a taxa de resposta acumulativa mínima com base na fórmula:  $(\text{Custo de envio de um pacote} / \text{Renda líquida por resposta}) \times 100$ . Ambos os valores devem ser números positivos. O resultado deve ser um valor maior que 0 e menor que 100. Por exemplo, se o custo de envio de um pacote for R\$ 0,75 e a renda líquida por resposta for R\$ 56, a taxa de resposta mínima será:  $(0,75/56) \times 100 = 1,34\%$ .

**Número máximo de contatos.** Se você especificar um número máximo de contatos, a tabela será codificada por cores para mostrar quais decis não excedem o número máximo de contatos (registros) acumulativos e o gráfico incluirá uma linha de referência nesse valor.

- **Porcentagem de contatos.** Máximo expresso como porcentagem. Por exemplo, talvez você queira saber os decis com as taxas de resposta mais altas que contêm um máximo de 50% de todos os contatos. O valor deve ser maior que 0 e menor que 100.
- **Número de contatos.** Máximo expresso como um número de contatos. Por exemplo, se você não pretende enviar mais de 10.000 pacotes, será possível configurar o valor em 10000. O valor deve ser um número inteiro positivo (sem símbolos de agrupamento).

Se você especificar uma taxa de resposta mínima aceitável e um número máximo de contatos, a codificação por cores da tabela será baseada na condição que for atendida primeiro.

### Exportar para Excel

Esse procedimento cria automaticamente um novo conjunto de dados que contém taxas de resposta por código de endereçamento postal. Cada registro (linha) no conjunto de dados representa um código de endereçamento postal. É possível salvar automaticamente as mesmas informações em um arquivo do Excel. Este arquivo é salvo em formato Excel 97-2003.

## Criando um campo de resposta categórica

O campo de resposta deve ser categórico, com um valor representando todas as respostas positivas. Qualquer outro valor não omissos é considerado como sendo uma resposta negativa. Se o campo de resposta representar um valor contínuo (escala), como o número de compras ou valor monetário de compras, será necessário criar um novo campo que designa um único valor de resposta positiva a todos os valores de resposta diferentes de zero.

- Se respostas negativas forem registradas como 0 (não em branco, que é tratado como omissos), isso pode ser calculado com a seguinte fórmula:

`NewName=OldName>0`

em que *NewName* é o nome do novo campo e *OldName* é o nome do campo original. Se essa for uma expressão lógica que designa um valor 1 a todos os valores não omissos maiores que 0, e 0 a todos os valores não omissos menores ou iguais a 0.

- Se nenhum valor for registrado para respostas negativas, esses valores serão tratados como omissos, e a fórmula será um pouco mais complicada:

`NewName=NOT(MISSING(OldName))`

Nessa expressão lógica, todos os valores de resposta não omissos são designados a um valor 1 e todos os valores de resposta omissos são designados a um valor 0.

- Se não for possível distinguir entre valores de resposta negativa (0) e valores omissos, um valor de resposta exato não poderá ser calculado. Se houver relativamente alguns valores realmente omissos, isso pode não ter um efeito significativo nas taxas de resposta calculadas. No entanto, se houver muitos valores omissos -- como quando as informações de resposta são registradas apenas para uma pequena amostra de teste do conjunto de dados total -- as taxas de resposta calculadas não serão significativas, pois serão significativamente menores que as taxas de resposta verdadeiras.

Para criar um campo de resposta categórica

1. Nos menus, escolha:

**Transformar > Calcular variável**

2. Para Variável de destino, insira o nome do novo campo (variável).
3. Se as respostas negativas forem registradas como 0, para Expressão numérica, insira `OldName>0`, em que *OldName* é o nome do campo original.
4. Se as respostas negativas forem registradas como omissas (em branco), para a Expressão numérica, insira `NOT(MISSING(OldName))`, em que *OldName* é o nome do campo original.

## Propensão para compra

---

A Propensão para compra usa resultados de um teste de envio ou campanha anterior para gerar escores. As pontuações indicam quais contatos têm maior probabilidade de responder. O campo Resposta indica quem respondeu ao teste de envio ou campanha anterior. Os campos de Propensão são as características que você deseja usar para prever a probabilidade de que contatos com características similares responderão.

Essa técnica usa a regressão logística binária para construir um modelo preditivo. O processo de construção e aplicação de um modelo preditivo tem dois passos básicos:

1. Construir o modelo e salvar o arquivo de modelo. Você constrói o modelo usando um conjunto de dados para o resultado de interesse (geralmente referido como o **destino**) é conhecido. Por exemplo, se desejar construir um modelo que irá prever quem tem probabilidade de responder a uma campanha de mala direta, é necessário iniciar com um conjunto de dados que já contenha informações sobre

quem respondeu e quem não respondeu. Por exemplo, isso pode ser os resultados de um teste de envio para um pequeno grupo de clientes ou informações sobre respostas a uma campanha semelhante no passado.

2. Aplicar esse modelo a um conjunto de dados diferente (para o qual o resultado de interesse não é conhecido) para obter os resultados preditos.

**Exemplo.** A divisão de marketing direto de uma empresa usa resultados de um teste de envio para designar escores de propensão ao restante de seu banco de dados de contatos, usando várias características demográficas para identificar contatos com maior probabilidade de responder e fazer uma compra.

Saída

Esse procedimento cria automaticamente um novo campo no conjunto de dados que contém escores de propensão para os dados de teste e um arquivo de modelo XML que pode ser usado para escorar outros conjuntos de dados. A saída de diagnóstico opcional inclui um gráfico de qualidade de modelo global e uma tabela de classificação que compara respostas preditas com respostas reais.

Considerações de dados de propensão para compra

**Campo de resposta.** O campo de resposta pode ser de sequência de caracteres ou numérico. Se esse campo contiver um valor que indica o número ou valor monetário de compras, será necessário criar um novo campo no qual um valor único representa todas as respostas positivas. Consulte o tópico [“Criando um campo de resposta categórica”](#) na página 17 para obter mais informações

**Valor de resposta positiva.** O valor de resposta positiva identifica os clientes que responderam positivamente (por exemplo, fizeram uma compra). Todos os outros valores de resposta não omissos são considerados para indicar uma resposta negativa. Se houver quaisquer rótulos de valor definidos para o campo de resposta, os rótulos serão exibidos na lista suspensa.





**Prever propensão com.** Os campos usados para prever a propensão podem ser uma sequência de caracteres ou numéricos, e podem ser nominais, ordinais ou contínuos (escala) -- mas é importante designar o nível de medição apropriado a todos os campos preditores.

**Nível de medição.** A designação correta do nível de medição é importante porque afeta o cálculo dos resultados.

- *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.
- *Contínuo.* 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.

Um ícone próximo a cada campo indica o nível de medição atual.

Tabela 4. Ícones do nível de medição				
	Numérico	Sequência de caracteres	Data	Time
Escala (Contínua)		n/a		
Ordinal				

	Numérico	Sequência de caracteres	Data	Time
Nominal				

É possível mudar o nível de medição na Visualização de variável do Editor de dados ou é possível usar o diálogo Definir propriedades da variável para sugerir um nível de medição apropriado para cada campo.

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.

Para obter escores de propensão para compra

Nos menus, escolha:

#### Marketing direto > Escolher técnica

1. Selecione **Selecionar contatos com maior probabilidade de compra.**
2. Selecione o campo que identifica quais contatos responderam à oferta.
3. Insira o valor que indica uma resposta positiva. Se quaisquer valores tiverem rótulos de valor definidos, será possível selecionar o rótulo de valor da lista suspensa e o valor correspondente será exibido.
4. Selecione os campos que você deseja usar para prever a propensão.

Para salvar um arquivo XML de modelo para escorar outros arquivos de dados:

5. Selecione (marque) **Exportar informações de modelo para arquivo XML.**
6. Insira um caminho do diretório e nome do arquivo ou clique em **Procurar** para navegar para o local onde você deseja salvar o arquivo XML de modelo.
7. Clique em **Executar** para executar o procedimento.

Para usar o arquivo de modelo para escorar outros conjuntos de dados:

8. Abra o conjunto de dados que você deseja escorar.
9. Use o Assistente de escoragem para aplicar o modelo ao conjunto de dados. Nos menus, escolha:

**Utilitários > Assistente de escoragem.**

## Configurações

Validação do Modelo

A validação de modelo cria grupos de treinamento e de teste para propósitos de diagnóstico. Se você selecionar a tabela de classificação na seção Saída de diagnóstico, a tabela será dividida em seções de treinamento (selecionado) e teste (desmarcado) para propósitos de comparação. Não selecione validação de modelo, a menos que também selecione a tabela de classificação. Os escores são baseados no

modelo gerado da amostra de treinamento, que sempre conterá menos registros que o número total de registros disponíveis. Por exemplo, o tamanho da amostra de treinamento padrão é 50%, e um modelo construído em apenas metade dos registros disponíveis pode não ser tão confiável quanto um modelo construído em todos os registros disponíveis.

- **Tamanho da partição da amostra de treinamento (%).** Especifique a porcentagem de registros a serem designados à amostra de treinamento. O restante dos registros com valores não omissos para o campo de resposta são designados à amostra de teste. O valor deve ser maior que 0 e menor que 100.
- **Configurar semente para replicar resultados.** Como os registros são designados aleatoriamente às amostras de treinamento e de teste, sempre que você executar o procedimento, pode obter resultados diferentes, a menos que sempre especifique o mesmo valor de valor semente de número aleatório inicial.

Saída de Diagnóstico

**Qualidade geral do modelo.** Exibe um gráfico de barras de qualidade geral do modelo, expressa como um valor entre 0 e 1. Um modelo válido deve ter um valor maior que 0,5.

**Tabela de classificação.** Exibe uma tabela que compara respostas positivas e negativas preditas com respostas positivas e negativas reais. A taxa de precisão geral pode fornecer alguma indicação de como o modelo funciona, mas você pode estar mais interessado na porcentagem de respostas positivas preditas corretas.

- **Probabilidade mínima.** Designa registros com um valor de escore maior que o valor especificado à categoria de resposta positiva predita na tabela de classificação. Os escores gerados pelo procedimento representam a probabilidade de que o contato responderá positivamente (por exemplo, fazer uma compra). Como regra geral, você deve especificar um valor próximo de sua taxa de resposta de destino mínima, expressa como uma proporção. Por exemplo, se estiver interessado em uma taxa de resposta de pelo menos 5%, especifique 0,05. O valor deve ser maior que 0 e menor que 1.

Nome e Rótulo do Campo de Resposta Recodificada

Esse procedimento recodifica automaticamente o campo de resposta em um novo campo no qual 1 representa respostas positivas e 0 representa respostas negativas, e a análise é executada no campo recodificado. É possível substituir o nome e o rótulo padrão e fornecer os seus próprios. Os nomes devem estar de acordo com as regras de nomenclatura do IBM SPSS Statistics.

Salvar Pontuações

Um novo campo contendo escores de propensão é salvo automaticamente no conjunto de dados original. Os escores representam a probabilidade de uma resposta positiva, expressa como uma proporção.

- Os nomes de campos devem estar de acordo com as regras de nomenclatura do IBM SPSS Statistics.
- O nome do campo não pode duplicar um nome de campo que já existe no conjunto de dados. Se você executar este procedimento mais de uma vez no mesmo conjunto de dados, precisará especificar um nome diferente a cada vez.

## Criando um campo de resposta categórica

O campo de resposta deve ser categórico, com um valor representando todas as respostas positivas. Qualquer outro valor não omissos é considerado como sendo uma resposta negativa. Se o campo de resposta representar um valor contínuo (escala), como o número de compras ou valor monetário de compras, será necessário criar um novo campo que designa um único valor de resposta positiva a todos os valores de resposta diferentes de zero.

- Se respostas negativas forem registradas como 0 (não em branco, que é tratado como omissos), isso pode ser calculado com a seguinte fórmula:

$$\text{NewName} = \text{OldName} > 0$$

em que *NewName* é o nome do novo campo e *OldName* é o nome do campo original. Se essa for uma expressão lógica que designa um valor 1 a todos os valores não omissos maiores que 0, e 0 a todos os valores não omissos menores ou iguais a 0.

- Se nenhum valor for registrado para respostas negativas, esses valores serão tratados como omissos, e a fórmula será um pouco mais complicada:

`NewName=NOT(MISSING(OldName))`

Nessa expressão lógica, todos os valores de resposta não omissos são designados a um valor 1 e todos os valores de resposta omissos são designados a um valor 0.

- Se não for possível distinguir entre valores de resposta negativa (0) e valores omissos, um valor de resposta exato não poderá ser calculado. Se houver relativamente alguns valores realmente omissos, isso pode não ter um efeito significativo nas taxas de resposta calculadas. No entanto, se houver muitos valores omissos -- como quando as informações de resposta são registradas apenas para uma pequena amostra de teste do conjunto de dados total -- as taxas de resposta calculadas não serão significativas, pois serão significativamente menores que as taxas de resposta verdadeiras.

Para criar um campo de resposta categórica

1. Nos menus, escolha:

**Transformar > Calcular variável**

2. Para Variável de destino, insira o nome do novo campo (variável).
3. Se as respostas negativas forem registradas como 0, para Expressão numérica, insira `OldName>0`, em que *OldName* é o nome do campo original.
4. Se as respostas negativas forem registradas como omissas (em branco), para a Expressão numérica, insira `NOT (MISSING (OldName) )`, em que *OldName* é o nome do campo original.

## Teste de Pacote de Controle

---

Esta técnica compara campanhas de marketing para verificar se há uma diferença significativa em efetividade para diferentes pacotes ou ofertas. A efetividade da campanha é medida por respostas. O Campo de Campanha identifica diferentes campanhas, por exemplo, Oferta A e Oferta B. O Campo de Resposta indica se um contato respondeu à campanha. Selecione Valor de compra quando a resposta for registrada como um valor de compra, por exemplo, "99,99". Selecione Responder quando a resposta simplesmente indicar se o contato respondeu positivamente ou não, por exemplo, "Sim" ou "Não".

**Exemplo.** A divisão de marketing direto de uma empresa deseja verificar se um novo design de pacote gerará mais respostas positivas do que o pacote existente. Portanto, eles enviam um teste de envio para determinar se o novo pacote gera uma taxa de resposta positiva significativamente mais alta. O teste de envio consiste em um grupo de controle que recebe o pacote existente e um grupo de testes que recebe o novo design do pacote. Os resultados para os dois grupos são então comparados para ver se há uma diferença significativa.

Saída

A saída inclui uma tabela que exibe contagens e porcentagens de respostas positivas e negativas para cada grupo definido pelo Campo de campanha e uma tabela que identifica quais grupos diferem significativamente uns dos outros.

Considerações e suposições de dados de teste do pacote de controle

**Campo de campanha.** O Campo de campanha deve ser categórico (nominal ou ordinal).

**Campo de resposta de eficácia.** Se você selecionar Valor de compra para o Campo de eficácia, o campo deverá ser numérico e o nível de medição deverá ser contínuo (escala).

Se não for possível distinguir entre valores de resposta negativa (para valor de compra, um valor 0) e valores omissos, uma taxa de resposta precisa não poderá ser calculada. Se houver relativamente alguns valores realmente omissos, isso pode não ter um efeito significativo nas taxas de resposta calculadas. No entanto, se houver muitos valores omissos -- como quando as informações de resposta são registradas apenas para uma pequena amostra de teste do conjunto de dados total -- as taxas de resposta calculadas não serão significativas, pois serão significativamente menores que as taxas de resposta verdadeiras.

**Suposições.** Este procedimento supõe que os contatos tenham sido designados aleatoriamente a cada grupo de campanha. Em outras palavras, nenhum demográfico específico, histórico de compra ou outras características afetam a designação de grupo, e todos os contatos têm uma probabilidade igual de serem designados a qualquer grupo.

Para obter um teste do pacote de controle

Nos menus, escolha:

### **Marketing direto > Escolher técnica**

1. Selecione **Comparar eficácia de campanhas**.
2. Selecione o campo que identifica a qual grupo de campanha cada contato pertence (por exemplo, oferta A, oferta B, etc.) Este campo deve ser nominal ou ordinal.
3. Selecione o campo que indica a eficácia de resposta.

Se o campo de resposta for um valor de compra, o campo deve ser numérico.

Se o campo de resposta simplesmente indicar se o contato respondeu positivamente ou não (por exemplo, "Sim" ou "Não"), selecione **Responder** e insira o valor que representa uma resposta positiva. Se quaisquer valores tiverem rótulos de valor definidos, será possível selecionar o rótulo de valor da lista suspensa e o valor correspondente será exibido.

Um novo campo é criado automaticamente, em que 1 representa respostas positivas e 0 representa respostas negativas, e a análise é executada no novo campo. É possível substituir o nome e o rótulo padrão e fornecer os seus próprios. Os nomes devem estar de acordo com as regras de nomenclatura do IBM SPSS Statistics.

4. Clique em **Executar** para executar o procedimento.





## 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:

© Copyright IBM Corp. 2021. Partes deste código são derivadas dos Programas de Amostra da IBM Corp.

© Copyright IBM Corp. 1989 - 2021. Todos os direitos reservados.

## 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

análise de cluster [7](#)  
análise de cluster (opção Marketing direto) [7](#)

## C

cluster [7](#)

## P

perfis de clientes em potencial (opção Marketing direto) [9](#)  
propensão para compra [14](#)

## R

regressão logística (Marketing direto) [14](#)  
RFM  
  categorização [3](#)  
  dados de transação [2](#)  
  dados do cliente [2](#)

## T

taxas de resposta do CEP [11](#)  
taxas de resposta do código de endereçamento postal [11](#)  
teste do pacote de controle [18](#)





