

# Análise de Correlação Canônica

## Objetivos de aprendizagem

Ao concluir este capítulo, você deverá ser capaz de:

- Enunciar as semelhanças e diferenças entre regressão múltipla, análise fatorial, análise discriminante e correlação canônica.
- Resumir as condições que devem ser atendidas para a aplicação de análise de correlação canônica.
- Enunciar o que a raiz canônica mede e mostrar suas limitações.
- Dizer quantas funções canônicas independentes podem ser definidas entre os dois conjuntos de variáveis originais.
- Comparar as vantagens e desvantagens dos três métodos para interpretar a natureza de funções canônicas.
- Definir redundância e compará-la com o coeficiente  $R^2$  de regressão múltipla.

## Apresentação do capítulo

Até recentemente, a análise de correlação canônica era uma técnica estatística relativamente desconhecida. Como ocorre com quase todas as técnicas multivariadas, a disponibilidade de programas de computador facilitou sua crescente aplicação em problemas de pesquisa. Ela é particularmente útil em situações nas quais múltiplas medidas de resultados, tais como satisfação, volume de compra ou volume de vendas, estão disponíveis. Se as variáveis independentes fossem apenas categóricas, a análise multivariada de variância poderia ser usada. Mas e se as variáveis independentes forem métricas? A correlação canônica é a resposta, permitindo a avaliação da relação entre variáveis independentes métricas e medidas dependentes múltiplas. A correlação canônica é tida como o modelo geral no qual muitas outras técnicas multivariadas são baseadas, pois ela pode utilizar tanto dados métricos quanto não-métricos para variáveis dependentes ou independentes. Expressamos a forma geral da análise canônica como

$$\begin{matrix} Y_1 + Y_2 + Y_3 + \dots + Y_n & = & X_1 + X_2 + X_3 + \dots + X_n \\ \text{(métricas, não-métricas)} & & \text{(métricas, não-métricas)} \end{matrix}$$

Este capítulo apresenta ao pesquisador a técnica estatística multivariada de análise de correlação canônica. Especificamente, (1) descrevemos a natureza da análise de correlação canônica, (2) ilustramos sua aplicação e (3) discutimos suas potenciais vantagens e limitações.

## Termos-chave

Antes de começar o capítulo, leia os termos-chave para compreender os conceitos e a terminologia empregados. Ao longo do capítulo, os termos-chave aparecem em **negrito**. Outros pontos que merecem destaque, além das referências cruzadas nos termos-chave, estão em *italico*.

**Autovalores** Ver *raízes canônicas*.

**Cargas canônicas** Medida da correlação linear simples entre as variáveis independentes e suas respectivas *variáveis estatísticas canônicas*. Essas podem ser interpretadas como cargas fatoriais e também são conhecidas como correlações de estrutura canônica.

**Cargas cruzadas canônicas** Correlação de cada variável independente ou dependente observada com a *variável estatística canônica* oposta. Por exemplo, as variáveis independentes são correlacionadas com a variável estatística canônica dependente. Elas podem ser interpretadas como *cargas canônicas*, mas com a variável estatística canônica oposta.

**Composições lineares** Ver *variáveis estatísticas canônicas*.

**Correlação canônica** Medida da força das relações gerais entre as composições lineares (*variáveis estatísticas canônicas*) para as variáveis independentes e dependentes. Na verdade, ela representa a correlação bivariada entre as duas variáveis estatísticas canônicas.

**Função canônica** Relação (correlacional) entre duas composições lineares (*variáveis estatísticas canônicas*). Cada função canônica tem duas variáveis estatísticas canônicas, uma para o conjunto de variáveis dependentes e outra para o conjunto de variáveis independentes. A força da relação é dada pela *correlação canônica*.

**Índice de redundância** Quantidade de variância em uma *variável estatística canônica* (dependente ou independente) explicada pela outra variável estatística canônica na *função canônica*. Pode ser computado tanto para variáveis estatísticas canônicas dependentes quanto independentes em cada função canônica. Por exemplo, um índice de redundância da variável estatística dependente representa a quantidade de variância nas variáveis dependentes explicada pela variável estatística canônica independente.

**Ortogonal** Restrição matemática que especifica que as funções canônicas são independentes uma da outra. Em outras palavras, as funções canônicas são obtidas de modo que cada uma forma um ângulo reto com todas as outras funções, quando representadas em um espaço multivariado, garantindo assim independência estatística entre as funções canônicas.

**Raízes canônicas** *Correlações canônicas* ao quadrado, o que fornece uma estimativa da quantidade de variância compartilhada entre as respectivas *variáveis estatísticas canônicas* otimamente ponderadas de variáveis dependentes e independentes. Também são conhecidas como *autovalores*.

**Variáveis estatísticas canônicas** Combinações lineares que representam a soma ponderada de duas ou mais variáveis e podem ser definidas para variáveis dependentes ou independentes. Também chamadas de *composições lineares*, compostos lineares e combinações lineares.

## O QUE É CORRELAÇÃO CANÔNICA?

A análise de regressão múltipla é uma técnica multivariada que pode prever o valor de uma única variável dependente (métrica) a partir de uma função linear de um conjunto de variáveis independentes. Para alguns problemas de pesquisa, o interesse pode não se concentrar em uma única variável dependente; em vez disso, o pesquisador talvez esteja interessado em relações entre conjuntos de múltiplas variáveis dependentes e múltiplas variáveis independentes. A análise de **correlação canônica** é um modelo estatístico multivariado que facilita o estudo de

inter-relações entre conjuntos de múltiplas variáveis dependentes e múltiplas variáveis independentes [5, 6]. Ao contrário da regressão múltipla, que prevê uma única variável dependente a partir de um conjunto de variáveis independentes múltiplas, a correlação canônica simultaneamente prevê múltiplas variáveis dependentes a partir de múltiplas variáveis independentes.

A correlação canônica apresenta o menor número de restrições sobre os tipos de dados nos quais ela opera. Como as outras técnicas impõem restrições mais rígidas, em geral crê-se que a informação obtida a partir delas é de maior qualidade e pode ser apresentada de uma maneira melhor para a interpretação. Por essa razão, muitos pesquisadores consideram a correlação canônica como uma última alternativa, a ser usada quando todas as outras técnicas de nível mais alto forem descartadas. Mas em situações com múltiplas variáveis dependentes e independentes, a correlação canônica é a técnica multivariada mais adequada e poderosa. Ela obteve aceitação em muitas áreas e representa uma ferramenta útil para a análise multivariada, particularmente porque aumentou o interesse na consideração de múltiplas variáveis dependentes.

## EXEMPLO HIPOTÉTICO DE CORRELAÇÃO CANÔNICA

Para tornar mais clara a natureza da correlação canônica, consideremos uma extensão de um exemplo de análise de regressão múltipla. Considere que foi conduzida uma pesquisa para entender as relações entre tamanho de família e renda como preditores do número de cartões de crédito que uma família usaria. Esse problema envolveu o exame da relação entre duas variáveis independentes e uma única variável dependente.

Suponha que o pesquisador estivesse então interessado em um conceito mais amplo de uso de crédito. Para medir o uso de crédito, o pesquisador considerou não somente o número de cartões de crédito utilizados pela família, mas também a média mensal de despesas em dólares com todos os cartões. Essas duas medidas foram percebidas como oferecendo uma perspectiva muito melhor sobre o uso de cartões de crédito pela família. Os leitores interessados na abordagem do uso de múltiplos indicadores para representar um conceito devem consultar assuntos como análise fatorial e modelagem de equações estruturais. O problema agora envolve a previsão de duas medidas dependentes simultaneamente (número de cartões de crédito e despesas médias em dólares).

A regressão múltipla é capaz de lidar apenas com uma única variável dependente. A análise multivariada de variância poderia ser usada, mas somente se todas as variáveis independentes fossem não-métricas, o que não é o caso nesse problema. A correlação canônica representa a única técnica disponível para examinar a relação com múltiplas variáveis dependentes.

O problema de prever o uso de cartões de crédito é ilustrado na Tabela 1. As duas variáveis dependentes empregadas para medir o uso de crédito – número de cartões utilizados pela família e despesas mensais médias em dólares com todos os cartões – são listadas à esquerda. As duas variáveis independentes selecionadas para prever o uso de crédito – o tamanho da família e a renda familiar – são mostradas à direita. Usando a análise de correlação canônica, o pesquisador cria uma medida composta de uso de crédito que consiste em ambas as variáveis dependentes, ao invés de ter de computar uma equação de regressão separada para cada uma. O resultado da aplicação de correlação canônica é uma medida da força da relação entre dois conjuntos de múltiplas variáveis (variáveis estatísticas canônicas). A medida da força da relação entre as duas variáveis estatísticas é expressada como um coeficiente de correlação canônica ( $R_c$ ). O pesquisador agora tem dois resultados de interesse: as variáveis estatísticas canônicas que representam as combinações lineares ótimas de variáveis dependentes e independentes; e a correlação canônica que representa a relação entre elas.

## ANÁLISE DAS RELAÇÕES COM CORRELAÇÃO CANÔNICA

A análise de correlação canônica é o membro mais generalizado da família de técnicas estatísticas multivariadas. Está diretamente relacionada com diversos métodos de dependência. Semelhante à regressão, a meta da correlação canônica é quantificar a força da relação, nesse caso entre os dois conjuntos de variáveis (independentes e dependentes). Ela corresponde à análise fatorial na criação de composições de variáveis. Também se assemelha à análise discriminante por conta de sua habilidade de determinar dimensões independentes (semelhantes às funções discriminantes) para *cada* conjunto de variáveis nessa situação com o objetivo de produzir a máxima correlação entre as dimensões. Assim, a correlação canônica identifica a estrutura ou dimensionalidade ótima de cada conjunto de variáveis que maximiza a relação entre conjuntos de variáveis independentes e dependentes.

A análise de correlação canônica lida com a associação entre composições de conjuntos de múltiplas variáveis dependentes e independentes. Ao fazer isso, desenvolve diversas **funções canônicas** independentes que maximizam a correlação entre as **composições lineares**, também conhecidas como **variáveis estatísticas canônicas**, as quais são conjuntos de variáveis dependentes e independentes. Cada função canônica é realmente baseada na correlação entre duas variáveis estatísticas canônicas, uma para as variáveis dependentes e outra para as independentes. Uma outra característica singular da correlação canônica é que as variáveis estatísticas são obtidas para maximizar sua correlação. Além disso, a correlação canônica não termina com a derivação de uma única relação entre os

conjuntos de variáveis. Em vez disso, diversas funções canônicas (pares de variáveis estatísticas canônicas) podem ser obtidas.

A discussão de análise de correlação canônica a seguir é organizada em torno do processo de construção de modelo. Os passos neste processo incluem (1) a especificação dos objetivos da correlação canônica, (2) o desenvolvimento do plano de análise, (3) a avaliação das suposições inerentes à correlação canônica, (4) a estimação do modelo canônico e a avaliação do ajuste geral do modelo, (5) a interpretação das variáveis estatísticas canônicas e (6) a validação do modelo.

## ESTÁGIO 1: OBJETIVOS DA ANÁLISE DE CORRELAÇÃO CANÔNICA

Os dados apropriados à análise de correlação canônica são dois conjuntos de variáveis. Consideramos que a cada conjunto pode ser dado algum significado teórico, pelo menos a ponto de um conjunto poder ser definido como as variáveis independentes e o outro como as variáveis dependentes. Uma vez que essa distinção tenha sido feita, a correlação canônica lida com uma vasta gama de objetivos. Tais metas podem ser qualquer uma das seguintes, ou mesmo todas:

1. Determinar se dois conjuntos de variáveis (medidas feitas sobre os mesmos objetos) são independentes um do outro, ou, ao contrário, determinar a magnitude das relações que possam existir entre os dois conjuntos.
2. Determinar um conjunto de pesos para cada conjunto de variáveis dependentes e independentes de modo que as combinações lineares de cada conjunto sejam maximamente correlacionadas. Funções lineares adicionais que maximizem a correlação remanescente são independentes dos conjuntos precedentes de combinações lineares.
3. Explicar a natureza de quaisquer relações que venham a existir entre os conjuntos de variáveis dependentes e independentes, geralmente pela medida da contribuição relativa de cada variável às funções canônicas (relações) extraídas.

A flexibilidade inerente da correlação canônica em termos do número e dos tipos de variáveis com as quais se lida, sejam dependentes ou independentes, torna a mesma uma candidata lógica para muitos dos mais complexos problemas abordados com técnicas multivariadas.

## ESTÁGIO 2: PLANEJAMENTO DE UMA ANÁLISE DE CORRELAÇÃO CANÔNICA

Por ser a forma mais geral de análise multivariada, a análise de correlação canônica compartilha questões básicas de implementação comuns a todas as técnicas multivariadas. Discussões sobre o impacto de erro de medida, os tipos de variáveis e suas transformações que podem ser incluídas são igualmente relevantes para a análise de correlação canônica.

Os problemas referentes ao impacto do tamanho da amostra (pequena ou grande) e à necessidade de um número suficiente de observações por variável frequentemente são encontrados com correlação canônica. Os pesquisadores são tentados a incluir muitas variáveis nos dois conjuntos (dependentes e independentes), sem perceber as implicações para o tamanho da amostra. Tamanhos amostrais muito pequenos não irão representar bem as correlações, obscurecendo assim quaisquer relações significativas. Amostras muito grandes terão uma tendência a indicar significância estatística em todos os casos, mesmo onde significância prática não é indicada. O pesquisador também é encorajado a manter pelo menos 10 observações por variável para evitar “super-ajuste” dos dados.

A classificação de variáveis como dependentes ou independentes é de pouca importância para a estimação estatística das funções canônicas, pois a análise de correlação canônica pondera ambas as variáveis estatísticas para maximizar a correlação e não enfatiza qualquer variável estatística em particular. Além disso, como a técnica produz variáveis estatísticas para maximizar a correlação entre elas, uma variável em qualquer conjunto se relaciona com todas as outras variáveis em ambos os conjuntos. Isso permite que a adição ou eliminação de uma única variável afete a solução inteira, particularmente a outra variável estatística. A composição de cada variável estatística, seja dependente ou independente, se torna crítica. Um pesquisador deve ter conjuntos de variáveis conceitualmente ligados antes de aplicar a análise de correlação canônica. Isso torna a especificação de variáveis estatísticas dependentes versus independentes essencial para se estabelecer um forte fundamento conceitual para as variáveis.

### ESTÁGIO 3: SUPOSIÇÕES EM CORRELAÇÃO CANÔNICA

A generalidade da análise de correlação canônica também se estende às suas suposições estatísticas inerentes. A suposição de linearidade afeta dois aspectos dos resultados de correlação canônica. Primeiro, o coeficiente de correlação entre duas variáveis quaisquer é baseado em uma relação linear. Se a relação é não-linear, então uma ou ambas as variáveis devem ser transformadas, se possível. Segundo, a correlação canônica é a relação linear entre as variáveis estatísticas. Se as variáveis estatísticas se relacionarem de maneira não-linear, a relação não será capturada por correlação canônica. Logo, apesar de a análise de correlação canônica ser o método multivariado mais generalizado, ele ainda se limita a identificar relações lineares.

A análise de correlação canônica pode acomodar qualquer variável métrica sem a suposição estrita de normalidade. A normalidade é desejável porque padroniza uma distribuição para permitir uma maior correlação entre as variáveis. Mas no sentido mais estrito, a análise de corre-

lação canônica pode acomodar até mesmo variáveis não-normais se a forma de distribuição (por exemplo, altamente assimétrica) não diminuir a correlação com outras variáveis. Isso permite que dados não-métricos transformados (na forma de variáveis dicotômicas) também sejam usados. No entanto, a normalidade multivariada é exigida para o teste de inferência estatística da significância de cada função canônica. Como os testes de normalidade multivariada não estão prontamente disponíveis, a orientação predominante é garantir que cada variável tenha normalidade univariada. Assim, apesar de a normalidade não ser estritamente exigida, é altamente recomendado que todas as variáveis sejam avaliadas quanto a normalidade e transformadas se necessário.

A homocedasticidade, a partir do ponto em que ela diminui a correlação entre variáveis, também deve ser remediada. Finalmente, a multicolinearidade em qualquer conjunto de variáveis irá atrapalhar a habilidade da técnica em isolar o impacto de qualquer variável, tornando a interpretação menos confiável.

### ESTÁGIO 4: DETERMINAÇÃO DAS FUNÇÕES CANÔNICAS E AVALIAÇÃO DO AJUSTE GERAL

O primeiro passo da análise de correlação canônica é obter uma ou mais funções canônicas. Cada função consiste em um par de variáveis estatísticas, em que uma representa as variáveis independentes, e a outra, as dependentes. O número máximo de variáveis estatísticas (funções) canônicas que podem ser extraídas dos conjuntos de variáveis é igual ao número de variáveis no menor conjunto de dados, independentes ou dependentes. Por exemplo, quando o problema de pesquisa envolve cinco variáveis independentes e três variáveis dependentes, o número máximo de funções canônicas que podem ser extraídas é três.

#### Determinação de funções canônicas

A obtenção de variáveis estatísticas canônicas sucessivas é semelhante ao procedimento usado com a análise fatorial não-rotacionada. O primeiro fator explica a máxima quantia de variância no conjunto de variáveis. Em seguida, o segundo fator é computado de modo a explicar tanta variância (que não esteja explicada pelo primeiro fator) quanto possível e assim por diante, até que todos os fatores tenham sido extraídos. Portanto, fatores sucessivos são obtidos a partir da variância residual de fatores anteriores. A análise de correlação canônica segue um procedimento parecido, mas se concentra na explicação da máxima quantia de relação entre os dois conjuntos de variáveis, ao invés de considerar apenas um conjunto. O resultado é que o primeiro par de variáveis estatísticas canônicas é obtido de forma a ter a inter-correlação mais alta possível entre os dois conjuntos de variáveis. O segundo



par de variáveis estatísticas canônicas é então obtido de forma a exibir a relação máxima entre os dois conjuntos de variáveis (variáveis estatísticas) não explicada pelo primeiro par. Em resumo, pares sucessivos de variáveis estatísticas canônicas são baseados em variância residual, e suas respectivas correlações canônicas (as quais refletem as inter-relações entre as variáveis estatísticas) se tornam menores à medida que cada função adicional é extraída. Ou seja, o primeiro par de variáveis estatísticas canônicas exibe a maior inter-correlação, o próximo par exibe a segunda maior inter-correlação, e assim por diante. Um ponto adicional sobre a obtenção de variáveis estatísticas canônicas: como observado, pares sucessivos de variáveis estatísticas canônicas são baseados em variância residual. Portanto, cada par de variáveis estatísticas é **ortogonal** e independente de todas as outras variáveis estatísticas obtidas a partir do mesmo conjunto de dados.

A força da relação entre os pares de variáveis estatísticas é refletida pela correlação canônica. Quando elevada ao quadrado, a correlação canônica representa a quantia de variância em uma variável estatística canônica explicada pela outra variável estatística canônica. Isso também pode ser chamado de quantidade de variância compartilhada entre as duas variáveis estatísticas canônicas. Correlações canônicas ao quadrado são chamadas de **raízes** ou **autovalores canônicos**.

### Quais funções canônicas devem ser interpretadas?

Como ocorre com a pesquisa que usa outras técnicas estatísticas, a prática mais comum é analisar as funções cujos coeficientes de correlação canônica são estatisticamente significantes além de algum nível, normalmente 0,05 ou acima. Se outras funções independentes são consideradas insignificantes, essas relações entre as variáveis não são interpretadas. A interpretação das variáveis estatísticas canônicas em uma função significativa é baseada na premissa de que as variáveis em cada conjunto que contribuem fortemente para as variâncias compartilhadas para essas funções são consideradas relacionadas umas com as outras.

Os autores acreditam que o emprego de um único critério, como o nível de significância, é muito superficial. Ao invés disso, recomendam que três critérios sejam usados em conjunto uns com os outros para decidir quais funções canônicas devem ser interpretadas. Os três critérios são (1) nível de significância estatística da função, (2) magnitude da correlação canônica e (3) medida de redundância para o percentual de variância explicada a partir dos dois conjuntos de dados.

#### Nível de significância

O nível de significância de uma correlação canônica geralmente considerado como o mínimo aceitável para interpretação é de 0,05, o qual (juntamente com o nível 0,01)

tornou-se o nível geralmente aceito para considerar um coeficiente de correlação estatisticamente aceitável. Esse consenso se desenvolveu amplamente por conta da disponibilidade de tabelas para esses níveis. Tais níveis, porém, não são necessariamente exigidos em todas as situações, e pesquisadores de várias disciplinas freqüentemente devem confiar em resultados baseados em níveis menores de significância. O teste mais usado, normalmente fornecido em pacotes computacionais, é a estatística *F*, baseada na aproximação de Rao [3].

Além de testes separados de cada função canônica, um teste multivariado de todas as raízes canônicas também pode ser usado para avaliar a significância de raízes canônicas. Muitas das medidas para avaliar a significância de funções discriminantes, incluindo o  $\lambda$  de Wilks, o traço de Hotelling, o traço de Pillai, e o gcr de Roy, também são fornecidas.

#### Magnitude das relações canônicas

A significância prática das funções canônicas, representada pelo tamanho das correlações canônicas, também deve ser considerada na decisão sobre quais funções interpretar. Não há orientações geralmente aceitas referentes a tamanhos adequados para correlações canônicas. Ao invés disso, a decisão geralmente é baseada na contribuição das descobertas para melhor compreender o problema de pesquisa estudado. Parece lógico que as orientações sugeridas para cargas fatoriais significantes em análise fatorial poderiam ser úteis com correlações canônicas, particularmente quando se considera que estas se referem à variância explicada nas variáveis estatísticas canônicas (composições lineares), e não às variáveis originais.

#### Medida de redundância de variância compartilhada

Lembre-se que as correlações canônicas ao quadrado (raízes) fornecem uma estimativa da variância compartilhada entre as variáveis estatísticas canônicas. Apesar de ser uma medida simples e atraente da variância compartilhada, ela pode conduzir a algum erro de interpretação, pois as correlações canônicas ao quadrado representam a variância compartilhada pelas composições lineares dos conjuntos de variáveis dependentes e independentes, e não a variância extraída dos conjuntos de variáveis [1]. Assim, uma correlação canônica relativamente forte pode ser obtida entre duas composições lineares (variáveis estatísticas canônicas), mesmo que essas composições lineares não possam extrair partes significantes de variância a partir de seus respectivos conjuntos de variáveis.

Como é possível obter correlações canônicas consideravelmente maiores do que os coeficientes de correlação bivariada e múltipla previamente relatados, pode haver uma tentação de assumir que a análise canônica tenha descoberto relações substanciais de significância conceitual e prática. Antes que tais conclusões sejam garantidas, porém, uma análise mais detalhada, que envolva medidas

diferentes de correlações canônicas, deve ser realizada para determinar a quantidade de variância de variável dependente explicada ou compartilhada com as variáveis independentes [7].

Para superar o viés e a incerteza inerentes ao uso de raízes canônicas (correlações canônicas ao quadrado) como uma medida de variância compartilhada, um **índice de redundância** foi proposto [8]. Ele equivale a computar o coeficiente de correlação múltipla ao quadrado entre o conjunto total de variáveis independentes e cada variável no conjunto de variáveis dependentes, e então calcular a média desses coeficientes ao quadrado para chegar a um  $R^2$  médio. Esse índice fornece uma medida resumo da habilidade de um conjunto de variáveis independentes (assumido como um conjunto) para explicar a variação nas variáveis dependentes (consideradas uma por vez). Como tal, a medida de redundância é perfeitamente análoga à estatística  $R^2$  da regressão múltipla, e seu valor como um índice é semelhante.

O índice de Stewart-Love de redundância calcula a quantidade de variância em um conjunto de variáveis que pode ser explicada pela variância no outro conjunto. Esse índice serve como uma medida de variância explicada, semelhante ao cálculo de  $R^2$  usado em regressão múltipla. O  $R^2$  representa a quantidade de variância na variável dependente explicada pela função de regressão das variáveis independentes. Em regressão, a variância total na variável dependente é igual a 1, ou 100%. Lembre que a correlação canônica é diferente da regressão múltipla, no sentido de que não lida com uma única variável dependente, mas tem uma composição das variáveis dependentes, e essa composição tem apenas uma parte da variância total de cada variável dependente. Por essa razão, não podemos assumir que 100% da variância no conjunto de variáveis dependentes esteja disponível para ser explicada pelo conjunto de variáveis independentes. Pode-se esperar que o conjunto de variáveis independentes explique apenas a variância compartilhada na variável estatística canônica dependente. Por isso, o cálculo do índice de redundância é um processo de três passos. O primeiro envolve o cálculo da quantidade de variância compartilhada do conjunto de variáveis dependentes incluídas na variável estatística canônica dependente. O segundo passo envolve o cálculo da quantidade de variância na variável estatística canônica dependente que pode ser explicada pela variável estatística canônica independente. O último passo é calcular o índice de redundância, encontrado pela multiplicação dessas duas componentes.

**Passo 1: a quantidade de variância compartilhada.** Para calcular a quantia de variância compartilhada no conjunto de variáveis dependentes incluídas na variável estatística canônica dependente, consideremos primeiro o modo como a estatística  $R^2$  de regressão é calculada. A  $R^2$  é simplesmente o quadrado do coeficiente de correlação  $R$ , o qual representa a correlação entre a variável dependente

real e o valor previsto. No caso canônico, estamos preocupados com a correlação entre a variável estatística canônica dependente e cada uma das variáveis dependentes. Tal informação pode ser obtida a partir das cargas canônicas ( $L_i$ ), as quais representam a correlação entre cada variável de entrada e sua própria variável estatística canônica (discutida em maiores detalhes na seção seguinte). Elevando-se ao quadrado cada carga de variável dependente ( $L_i^2$ ), pode-se obter uma medida da quantidade de variação em cada variável dependente explicada pela variável estatística canônica dependente. Para calcular a quantidade de variância compartilhada explicada pela variável estatística canônica, é usada uma simples média das cargas ao quadrado.

**Passo 2: a quantidade de variância explicada.** O segundo passo do processo de redundância envolve o percentual de variância na variável estatística canônica dependente que pode ser explicada pela variável estatística canônica independente. É simplesmente a correlação ao quadrado entre a variável estatística canônica independente e a variável estatística canônica dependente, também conhecida como a correlação canônica. A correlação canônica ao quadrado é comumente chamada de  $R^2$  canônico.

**Passo 3: o índice de redundância.** O índice de redundância de uma variável estatística é então conseguido multiplicando-se as duas componentes (variância compartilhada da variável estatística multiplicada pela correlação canônica ao quadrado) para encontrar a quantia de variância compartilhada que pode ser explicada por cada função canônica. Para se ter um alto índice de redundância, deve-se ter uma alta correlação canônica e um alto grau de variância compartilhada explicada pela variável estatística dependente. Uma alta correlação canônica sozinha não garante uma função canônica valiosa. Índices de redundância são calculados tanto para variáveis estatísticas dependentes quanto independentes, apesar de que, na maioria dos casos, o pesquisador está preocupado apenas com a variância extraída do conjunto de variáveis dependentes, a qual fornece uma medida muito mais realista da habilidade preditiva de relações canônicas. O pesquisador deve observar que, embora a correlação canônica seja a mesma para ambas as variáveis estatísticas na função canônica, é muito provável que o índice de redundância varie entre as duas variáveis estatísticas, uma vez que cada uma terá diferentes quantias de variância compartilhada. Qual é o índice de redundância mínimo aceitável necessário para justificar a interpretação de funções canônicas? Como ocorre com as correlações canônicas, nenhuma orientação generalizada foi estabelecida. O pesquisador deve julgar cada função canônica à luz de sua significância teórica e prática perante o problema de pesquisa investigado para determinar se o índice de redundância é suficiente para justificar a interpretação. Um teste para a significância do índice de redundância foi desenvolvido [2], apesar de ainda não ser amplamente utilizado.

## ESTÁGIO 5: INTERPRETAÇÃO DA VARIÁVEL ESTATÍSTICA CANÔNICA

Se a relação canônica é estatisticamente significativa e as magnitudes da raiz canônica e do índice de redundância são aceitáveis, o pesquisador ainda precisa fazer interpretações importantes dos resultados. Fazer essas interpretações envolve o exame das funções canônicas para determinar a importância relativa de cada variável original nas relações canônicas. Três métodos foram propostos: (1) pesos canônicos (coeficientes padronizados), (2) cargas canônicas (correlações estruturais) e (3) cargas cruzadas canônicas.

### Pesos canônicos

A abordagem tradicional para interpretar funções canônicas envolve o exame do sinal e da magnitude do peso canônico designado para cada variável em sua variável estatística canônica. Variáveis com pesos relativamente maiores contribuem mais para as variáveis estatísticas e vice-versa. Do mesmo modo, variáveis cujos pesos têm sinais opostos exibem uma relação inversa uma com a outra, e variáveis com pesos de mesmo sinal, uma relação direta. Contudo, interpretar a importância ou contribuição relativa de uma variável por seu peso canônico está sujeito às mesmas críticas associadas à interpretação dos pesos beta em técnicas de regressão. Por exemplo, um peso pequeno pode significar que sua variável correspondente é irrelevante para determinar uma relação ou que ela foi parcialmente eliminada do relacionamento por causa de um grau elevado de multicolinearidade. Um outro problema do uso de pesos canônicos é que esses pesos estão sujeitos a considerável instabilidade (variabilidade) de uma amostra para outra. Essa instabilidade ocorre porque o procedimento computacional para análise canônica conduz a pesos que maximizam as correlações canônicas para uma amostra em particular de variáveis dependentes e independentes [7]. Esses problemas sugerem um cuidado considerável no uso de pesos canônicos para interpretar os resultados de uma análise canônica.

### Cargas canônicas

As cargas canônicas têm sido cada vez mais empregadas como uma base para a interpretação por causa das deficiências inerentes aos pesos canônicos. As **cargas canônicas**, também conhecidas como correlações estruturais canônicas, medem a correlação linear simples entre uma variável observada original no conjunto dependente ou independente e a variável estatística canônica do conjunto. A carga canônica reflete a variância que a variável observada compartilha com a variável estatística canônica e pode ser interpretada como uma carga fatorial na avaliação da contribuição relativa de cada variável a cada função canônica. A metodologia considera cada função canônica independente separadamente e computa a correlação interna no conjunto de cada variável com variável

estatística. Quanto maior o coeficiente, mais importante é a variável para derivar a variável estatística canônica. Além disso, os critérios para determinar a significância de correlações estruturais canônicas são os mesmos utilizados para cargas fatoriais em análise fatorial.

Cargas canônicas, como pesos, podem estar sujeitas a considerável variabilidade de uma amostra para outra. Essa variabilidade sugere que as cargas, e portanto as relações atribuídas a elas, podem ser específicas da amostra, resultantes de fatores casuais extrínsecos [7]. Apesar de as cargas canônicas serem consideradas relativamente mais válidas do que os pesos como um meio de interpretar a natureza de relações canônicas, mesmo assim o pesquisador deve ser cuidadoso quando usar cargas para interpretar relações canônicas, particularmente com relação à validade externa das descobertas.

### Cargas cruzadas canônicas

A computação de **cargas cruzadas canônicas** foi sugerida como uma alternativa às cargas canônicas [4]. Esse procedimento envolve correlacionar cada variável dependente observada original diretamente com a variável estatística canônica independente e vice-versa. Lembre que as cargas convencionais correlacionam as variáveis observadas originais com suas respectivas variáveis estatísticas depois que as duas variáveis estatísticas canônicas (dependente e independente) estiverem maximamente correlacionadas uma com a outra. Isso também pode parecer semelhante à regressão múltipla, mas difere no sentido de que cada variável independente, por exemplo, está correlacionada com a variável estatística dependente, em vez de uma só variável dependente. Assim, as cargas cruzadas fornecem uma medida mais direta das relações de variáveis dependentes-independentes pela eliminação de um passo intermediário envolvido em cargas convencionais. Algumas análises canônicas não computam correlações entre as variáveis e as variáveis estatísticas. Em tais casos, os pesos canônicos são considerados comparáveis, mas não equivalentes para fins de nossa discussão.

### Qual abordagem de interpretação usar

Diversos métodos para interpretar a natureza de relações canônicas foram discutidos. A questão, porém, permanece: qual método o pesquisador deve usar? Como a maioria dos problemas canônicos exige um computador, o pesquisador frequentemente deve usar qualquer método disponível em pacotes estatísticos padrões. A abordagem de cargas cruzadas é preferida e fornecida por muitos programas de computador, mas se as cargas cruzadas não estiverem disponíveis, o pesquisador será forçado a computar as cargas cruzadas manualmente ou confiar nos outros métodos de interpretação. A abordagem de cargas canônicas é um pouco mais representativa do que o uso de pesos, assim como foi visto com a análise fatorial e a análise discriminante. Portanto, sempre que possível, a técnica



de cargas é recomendada como a melhor alternativa em relação ao método de cargas cruzadas canônicas.

## ESTÁGIO 6: VALIDAÇÃO E DIAGNÓSTICO

Como acontece com qualquer outra técnica multivariada, a análise de correlação canônica deve ser sujeita a métodos de validação para garantir que os resultados não sejam específicos apenas para os dados da amostra e possam ser generalizados para a população. O procedimento mais direto é criar duas sub-amostras dos dados (se o tamanho amostral permitir) e executar a análise em cada sub-amostra separadamente. Então os resultados podem ser comparados quanto a similaridade de funções canônicas, a cargas das variáveis estatísticas e semelhantes. Se são descobertas diferenças marcantes, o pesquisador deve pensar em uma investigação adicional para garantir que os resultados finais sejam representativos dos valores da população, e não apenas dos valores de uma única amostra.

Uma outra abordagem é avaliar a sensibilidade dos resultados quanto à remoção de uma variável dependente e/ou independente. Como o procedimento de correlação canônica maximiza a correlação e não otimiza a interpretabilidade, os pesos e cargas canônicos podem variar substancialmente se uma variável for removida de qualquer variável estatística. Para garantir a estabilidade dos pesos e das cargas canônicos, o pesquisador deve estimar múltiplas correlações canônicas, removendo uma variável diferente independente ou dependente a cada vez. Apesar de haver poucos procedimentos diagnósticos desenvolvidos especificamente para a análise de correlação canônica, o pesquisador deve ver os resultados dentro das limitações da técnica. Entre as limitações que podem ter o maior impacto sobre os resultados e sua interpretação, estão as seguintes:

1. A correlação canônica reflete a variância compartilhada pelas composições lineares dos conjuntos de variáveis, e não a variância extraída das variáveis.
2. Os pesos canônicos derivados no cálculo das funções canônicas estão sujeitos a uma alta instabilidade.
3. Os pesos canônicos são derivados para maximizar a correlação entre composições lineares, e não a variância extraída.
4. A interpretação das variáveis estatísticas canônicas pode ser difícil porque elas são calculadas para maximizar a relação, e não há qualquer auxílio na interpretação, como a rotação de variáveis estatísticas, conforme visto em análise fatorial.
5. É difícil identificar relações significativas entre os subconjuntos de variáveis independentes e dependentes porque ainda não foram desenvolvidas estatísticas precisas para interpretar análise canônica, e devemos confiar em medidas inadequadas, como cargas ou cargas cruzadas [7].

Essas limitações não devem desencorajar o uso da correlação canônica. Ao invés disso, são apontadas para

aumentar a efetividade da correlação canônica como instrumento de pesquisa.

## UM EXEMPLO ILUSTRATIVO

Para ilustrar a aplicação da correlação canônica, usamos variáveis obtidas de uma pequena pesquisa de uma firma conhecida como HATCO. Os dados consistem em uma série de medidas obtidas de uma amostra de 100 clientes da HATCO. As variáveis incluem avaliações da HATCO sobre sete atributos ( $X_1$  a  $X_7$ ) e duas medidas que refletem os efeitos dos esforços da HATCO ( $X_9$ , uso de produtos da HATCO, e  $X_{10}$ , satisfação do cliente com a HATCO). Uma descrição completa da pesquisa HATCO é fornecida no Apêndice 1.

A discussão dessa aplicação da análise de correlação canônica segue o processo de seis estágios discutido anteriormente no capítulo. Em cada estágio, os resultados que ilustram as decisões naquele estágio são examinados.

### Estágio 1: objetivos da análise de correlação canônica

Ao demonstrar a aplicação de correlação canônica, usamos todas as nove variáveis como dados de entrada. As avaliações da HATCO ( $X_1$  a  $X_7$ ) são designadas como o conjunto de variáveis independentes. As medidas de nível de uso e nível de satisfação (variáveis  $X_9$  e  $X_{10}$ ) são especificadas como o conjunto de variáveis dependentes. O problema estatístico envolve a identificação de quaisquer relações latentes (relações entre composições de variáveis no lugar de variáveis individuais propriamente) entre percepções de um cliente sobre a HATCO e o nível de uso e a satisfação do cliente.

### Estágios 2 e 3: planejamento de uma análise de correlação canônica e teste das suposições

A designação das variáveis inclui duas variáveis dependentes métricas e sete independentes métricas. A base conceitual de ambos os conjuntos é bem estabelecida, de forma que não há necessidade de formulações de modelos alternativos que testem diferentes conjuntos de variáveis. As sete variáveis resultaram em uma proporção de 13\* para 1 observações em relação a variáveis, excedendo a orientação de 10 observações por variável. O tamanho amostral de 100 não parece afetar fortemente as estimativas de erro amostral e, assim, não deve ter impacto sobre a significância estatística dos resultados. Finalmente, para fins deste exemplo, assumo que tanto variáveis dependentes quanto independentes foram avaliadas para atender às suposições básicas de distribuição inerentes a análises multivariadas e passaram em todos os testes estatísticos.

\* N. de R. T.: O correto é "...proporção de 14 para 1".



#### **Estágio 4: determinação das funções canônicas e avaliação do ajuste geral**

A análise de correlação canônica foi restrita à determinação de duas funções canônicas porque o conjunto de variáveis dependentes contém apenas duas variáveis. Para determinar o número de funções canônicas a incluir no estágio de interpretação, a análise se concentrou no nível de significância estatística, na significância prática da correlação canônica e nos índices de redundância para cada variável estatística.

##### ***Significância estatística e prática***

O primeiro teste de significância estatística é para as correlações canônicas de cada uma das duas funções canônicas. Nesse exemplo, ambas as correlações canônicas são estatisticamente significantes (ver Tabela 2). Além de testes de cada função canônica separadamente, testes multivariados das funções simultaneamente também são executados. As estatísticas de teste são o  $\lambda$  de Wilks, o critério de Pillai, o traço de Hotelling e o gcr de Roy. A Tabela 2 também detalha as estatísticas de teste multivariado, as quais indicam que as funções canônicas, tomadas coletivamente, são estatisticamente significantes no nível 0,01.

Além da significância estatística, as correlações canônicas tinham tamanho suficiente para serem consideradas com significância prática. O passo final foi executar análises de redundância sobre as duas funções canônicas.

##### ***Análise de redundância***

Um índice de redundância é calculado para as variáveis estatísticas independente e dependente da primeira função na Tabela 3. Como pode ser visto, o índice de redundância para a variável estatística dependente é substancial (0,751). A variável estatística independente, contudo, tem um índice de redundância consideravelmente menor (0,242), apesar de, nesse caso, como há uma clara fronteira entre as variáveis dependentes e independentes, esse valor menor não ser considerado inesperado ou problemático. A baixa redundância da variável estatística independente resulta da variância compartilhada relativamente baixa na variável estatística independente (0,276), e não do  $R^2$  canônico. A partir da análise de redundância e dos testes de significância estatística, a primeira função deve ser aceita.

A análise de redundância para a segunda função produz resultados bem diferentes (ver Tabela 4). Primeiro, o  $R^2$  canônico é muito menor (0,260). Além disso, os conjuntos de variáveis têm baixa variância compartilhada na segunda função (0,145 para a variável estatística dependente e 0,082 para a variável estatística independente). A combinação delas com a raiz canônica no índice de redundância produz valores de 0,038 para a variável estatística dependente e 0,021 para a variável estatística independente. Logo, apesar de a segunda função ser estatisticamen-

te significativa, ela tem pouca significância prática. Com um percentual tão pequeno, deve-se questionar o valor da função. Esse é um excelente exemplo de uma função canônica estatisticamente significativa que não tem significância prática pelo fato de não explicar uma grande parte da variância das variáveis dependentes. O pesquisador interessado deve considerar análise fatorial com atenção para a discussão sobre desenvolvimento de escala. A correlação canônica é, de certo modo, uma forma de desenvolvimento de escala, uma vez que variáveis estatísticas dependentes e independentes representam dimensões dos conjuntos de variáveis semelhantes às escalas desenvolvidas com análise fatorial. A principal diferença é que essas dimensões são desenvolvidas para maximizar a relação entre elas, enquanto a análise fatorial maximiza a explicação (variância compartilhada) do conjunto de variáveis.

#### **Estágio 5: interpretação das variáveis estatísticas canônicas**

Com a relação canônica considerada estatisticamente significativa e a magnitude da raiz canônica e o índice de redundância aceitáveis, o pesquisador prossegue fazendo interpretações importantes dos resultados. Apesar de a segunda função poder ser considerada praticamente não-significante, devido ao baixo valor de redundância, ela é incluída na fase de interpretação por motivos ilustrativos. Essas interpretações envolvem o exame das funções canônicas para determinar a importância relativa de cada variável original na obtenção das relações canônicas.

Os três métodos para interpretação são (1) pesos canônicos (coeficientes padronizados), (2) cargas canônicas (correlações estruturais) e (3) cargas cruzadas canônicas.

##### ***Pesos canônicos***

A Tabela 5 contém os pesos canônicos padronizados para cada variável estatística canônica para as variáveis dependentes e independentes. Como já discutido, a magnitude dos pesos representa sua contribuição relativa para com a variável estatística. Com base no tamanho dos pesos, a ordem de contribuição de variáveis independentes para a primeira variável estatística é  $X_3$ ,  $X_5$ ,  $X_4$ ,  $X_1$ ,  $X_2$ ,  $X_6$  e  $X_7$ , e a ordem de variáveis dependentes na primeira variável estatística é  $X_{10}$  seguida de  $X_9$ . Ordens semelhantes podem ser encontradas para as variáveis estatísticas da segunda função canônica. Como os pesos canônicos são geralmente instáveis, particularmente em casos de multicolinearidade, devido ao fato de que seu cálculo apenas otimiza a correlação canônica, a carga canônica e a carga cruzada são consideradas mais adequadas.

##### ***Cargas canônicas***

A Tabela 6 contém as cargas canônicas para as variáveis estatísticas dependente e independente para ambas as funções canônicas. O objetivo de maximizar a correlação entre as variáveis estatísticas resulta em variáveis estatísticas

“otimizadas” não para interpretação, mas para previsão. Isso torna a identificação de relações mais difícil. Na primeira variável estatística dependente, ambas as variáveis têm cargas que excedem 0,90, o que resulta em variância compartilhada elevada (0,855). Isso indica um alto grau de inter-correlação entre as duas variáveis e sugere que ambas as medidas, ou uma delas, são representativas dos efeitos de esforços da HATCO.

A primeira variável estatística independente tem um padrão muito diferente, em que as cargas variam de 0,061 a 0,765, e uma variável independente ( $X_7$ ) tendo uma carga negativa, apesar de ser muito pequena e sem muito interesse. As três variáveis com as mais altas cargas na variável estatística independente são  $X_5$  (serviço geral),  $X_1$  (velocidade de entrega) e  $X_3$  (flexibilidade de preço). Essa variável estatística não corresponde às dimensões extraídas em análise fatorial, mas isso não era de se esperar, pois as variáveis estatísticas em correlação canônica são extraídas apenas para maximizar objetivos preditivos. Como tal, ela deve corresponder mais aos resultados de outras técnicas de dependência. Existe uma forte correspondência com resultados de regressão múltipla tendo  $X_9$  como variável dependente. Duas dessas variáveis ( $X_3$  e  $X_5$ ) foram incluídas na análise de regressão *stepwise*, na qual  $X_9$  (uma das duas variáveis na variável estatística dependente) era a variável dependente. Assim, a primeira função canônica se aproxima dos resultados de regressão múltipla, e a variável estatística independente representa o conjunto de variáveis que melhor prevê as duas medidas dependentes. O pesquisador também deve realizar uma análise de sensibilidade da variável estatística independente nesse caso para ver se as cargas mudam quando uma variável independente é eliminada (ver estágio 6). Os baixos valores de redundância da segunda variável estatística são exibidos nas cargas substancialmente menores para ambas as variáveis estatísticas na segunda função. Logo, a interpretabilidade mais baixa refletida nas cargas menores, juntamente com os baixos valores de redundância, reforça a baixa significância prática da segunda função.

### **Cargas cruzadas canônicas**

A Tabela 6 também inclui as cargas cruzadas para as duas funções canônicas. Ao estudar a primeira função canônica, percebemos que ambas as variáveis independentes\* ( $X_9$  e  $X_{10}$ ) exibem altas correlações com a variável estatística canônica independente (função 1): 0,855 e 0,877, respectivamente. Isso reflete a elevada variância compartilhada entre essas duas variáveis. Elevando ao quadrado esses termos, descobrimos o percentual da variância para cada variável explicada pela função 1. Os resultados mostram que 73% da variância em  $X_9$  e 77% da variância em  $X_{10}$  são explicados pela função 1. Olhando as cargas cruzadas das variáveis independentes, vemos que as variáveis  $X_1$  e  $X_5$  têm altas correlações de aproximadamente 0,72 com a variável estatística canônica dependente. A partir dessa informação, aproximadamente 52% da variância em

cada uma dessas duas variáveis é explicada pela variável estatística dependente (os 52% são obtidos elevando-se ao quadrado o coeficiente de correlação, 0,72). A correlação de  $X_3$  (0,584) pode parecer alta, mas depois de se elevar ao quadrado essa correlação, apenas 34% da variação é incluída na variável estatística canônica. A última questão de interpretação é examinar os sinais das cargas cruzadas. Todas as variáveis independentes, exceto  $X_7$  (qualidade do produto), têm uma relação positiva, direta. Para a segunda função, duas variáveis independentes ( $X_4$  e  $X_6$ ) e uma variável dependente ( $X_{10}$ ) são negativas. As três cargas cruzadas mais elevadas da primeira variável estatística independente correspondem às variáveis que também têm as cargas canônicas mais altas. Assim, todas as relações são diretas, exceto por uma relação inversa na primeira função.

### **Estágio 6: validação e diagnóstico**

O último estágio deve envolver uma validação das análises de correlação canônica por meio de um entre diversos procedimentos. Entre as abordagens disponíveis estão (1) a partição da amostra em amostras de estimação e de validação, ou (2) a análise de sensibilidade do conjunto de variáveis independentes. A Tabela 7 contém o resultado de tal análise de sensibilidade, na qual as cargas canônicas são examinadas quanto a estabilidade quando variáveis individuais independentes são eliminadas da análise. Como percebido, as cargas canônicas em nosso exemplo são notavelmente estáveis e consistentes nos três casos onde uma variável independente ( $X_1$ ,  $X_2$  ou  $X_7$ ) é eliminada. As correlações canônicas gerais também permanecem estáveis. Mas o pesquisador que examinasse os pesos canônicos (não apresentados na tabela) encontraria resultados bastante variados, dependendo de qual variável fosse eliminada. Isso reforça o procedimento de usar a carga canônica e a carga cruzada para fins de interpretação.

## **UMA VISÃO GERENCIAL**

A análise de correlação canônica tem dois objetivos principais: (1) a identificação de dimensões entre as variáveis dependentes e independentes que (2) maximizam a relação entre as dimensões. De uma perspectiva administrativa, isso fornece ao pesquisador uma visão da estrutura dos diferentes conjuntos de variáveis e de como eles correspondem a uma relação de dependência. Primeiro, os resultados indicam que apenas uma relação existe, sustentada pela baixa significância prática da segunda função canônica. Ao examinar essa relação, primeiro vemos que as duas variáveis dependentes são fortemente relacionadas e criam uma dimensão bem definida para representar os resultados dos esforços da HATCO. Segundo, essa dimensão de resultado é bem prevista pelo conjunto de variáveis independentes quando estas agem como um conjunto. O valor de redundância de 0,750 seria um  $R^2$  bastante aceitável para uma regressão múltipla comparável. Quando

\* N. de R. T.: A frase correta seria “... variáveis dependentes”.

interpretamos a variável estatística independente, vemos que três variáveis,  $X_5$  (serviço geral),  $X_1$  (velocidade de entrega) e  $X_3$  (flexibilidade de preço), fornecem as contribuições importantes e assim são os preditores-chave da dimensão de resultado. Esses deveriam ser os pontos centrais no desenvolvimento de qualquer estratégia dirigida ao impacto de resultados da HATCO.

## Resumo

A análise de correlação canônica é uma técnica útil e poderosa para explorar as relações entre múltiplas variáveis dependentes e independentes. A técnica é, antes de tudo, descritiva, apesar de poder ser usada para fins preditivos. Resultados obtidos de uma análise canônica devem sugerir respostas a questões referentes ao número de maneiras nas quais os dois conjuntos de variáveis múltiplas estão relacionados, às forças das relações e à natureza das relações definidas. A análise canônica permite ao pesquisador combinar na forma de uma medida composta o que de outra maneira poderia ser um grande número intratável de correlações bivariadas entre conjuntos de variáveis. É útil para identificar relações gerais entre múltiplas variáveis independentes e dependentes, particularmente quando o pesquisador tem pouco conhecimento *a priori* sobre as relações entre os conjuntos de variáveis. Essencialmente, o pesquisador pode aplicar a análise de correlação canônica a um conjunto de variáveis, selecionar as variáveis (tanto independentes quanto dependentes) que parecem estar significativamente relacionadas, e executar correlações canônicas subseqüentes com as variáveis restantes mais significantes, ou realizar regressões individuais com essas variáveis.

## Questões

1. Sob quais circunstâncias você selecionaria a análise de correlação canônica em vez da regressão múltipla como a técnica estatística apropriada?
2. Quais são os três critérios que você deve usar para decidir quais funções canônicas devem ser interpretadas? Explique o papel de cada uma.
3. Como você interpretaria uma análise de correlação canônica?
4. Qual é a relação entre a raiz canônica, o índice de redundância e o  $R^2$  de regressão múltipla?
5. Quais são as limitações associadas à análise de correlação canônica?
6. Por que a análise de correlação canônica tem sido usada com menos frequência do que as outras técnicas multivariadas?

## Referências

1. Alpert, Mark I., and Robert A. Peterson (1972), "On the Interpretation of Canonical Analysis." *Journal of Marketing Research* 9 (May): 187.
2. Alpert, Mark I., Robert A. Peterson, and Warren S. Marti (1975), "Testing the Significance of Canonical

Correlations." *Proceedings, American Marketing Association* 37: 117-19.

3. Bartlett M. S. (1941), "The Statistical Significance of Canonical Correlations." *Biometrika* 32: 29.
4. Dillon, W. R., and M. Goldstein (1984), *Multivariate Analysis: Methods and Applications*. New York: Wiley.
5. Green, P. E. (1978), *Analyzing Multivariate Data*. Hinsdale, Ill.: Holt, Rinehart, & Winston.
6. Green, P. E., and J. Douglas Carroll (1978), *Mathematical Tools for Applied Multivariate Analysis*. New York: Academic Press.
7. Lambert, Z., and R. Durand (1975), "Some Precautions in Using Canonical Analysis." *Journal of Marketing Research* 12 (November): 468-75.
8. Stewart, Douglas, and William Love (1968), "A General Canonical Correlation Index." *Psychological Bulletin* 70: 160-63.

## Artigos complementares

Os seguintes artigos complementares são fornecidos como ilustrações da aplicação da análise de correlação canônica a questões de pesquisa importantes de natureza conceitual e administrativa. O leitor é encorajado a ler o artigo completo para mais detalhes sobre questões específicas referentes à metodologia ou às descobertas.

Schul, Patrick L., William M. Pride, and Taylor L. Little (1983), "The Impact of Channel Leadership Behavior on Intrachannel Conflict." *Journal of Marketing* 47 (3): 21-34.

Esse artigo usa a análise de correlação canônica, uma técnica que permite a investigação de efeitos de múltiplas variáveis independentes sobre múltiplas variáveis dependentes. Nesse artigo, a técnica multivariada é aplicada para determinar os efeitos de estilo de liderança sobre as percepções de conflito dentro de um canal. Examinando a força e direção da relação entre estilo de liderança e conflito geral em um canal, a análise de correlação canônica fornece ao pesquisador informação para melhorar as transações de canal de distribuição. Descobertas dessa natureza também fornecem a administradores de organizações de canais um meio de estruturar sua conduta com outros membros de canais. Três estilos de liderança – participativa, diretiva e de apoio – que, segundo a teoria, exibiam uma relação inversa com duas formas de conflito interno de canais – administrativa e produto-serviço – são definidos e medidos. É por meio dessa técnica que os pesquisadores são capazes de explorar a relação entre construtos multifacetados como liderança e conflito.

Dados de uma amostra de 349 corretores imobiliários franqueados foram analisados para testar uma associação entre liderança de canal (como as variáveis preditoras ou independentes) e conflito interno de canal (como as variáveis de critério ou dependentes). Os autores conseguiram confirmar que existe uma forte relação entre o tipo de liderança de canal e conflito interno de canal pelo exame do índice de redundância (que indica a quantia de variância



explicada em uma variável estatística canônica pela outra variável estatística canônica) e da correlação entre as duas variáveis estatísticas. Por meio de análise individual, todos os três estilos de liderança reduzem o conflito. Técnicas participativas e de apoio facilitam o entendimento e a aceitação de políticas e procedimentos, enquanto um estilo diretivo reduz a ambigüidade de papéis desempenhados. A partir dos grandes valores de cargas cruzadas, os resultados indicam que a liderança de apoio tem uma relação com a redução de conflito de canal mais forte do que a liderança participativa ou diretiva. Para fins de validação, os autores fazem uma análise de correlação canônica em uma amostra de análise e outra de validação. Pesos canônicos são comparados nas duas amostras, fornecendo assim uma indicação da estabilidade dessa medida para a amostra combinada. Esses resultados indicam que um líder de canal deve implementar um estilo que melhor se ajuste às necessidades do franqueado a fim de minimizar o conflito de canal.

Luthans, Fred, Dianne H. B. Welsh, and Lewis A. Taylor III (1988), "A Descriptive Model of Managerial Effectiveness." *Group and Organization Studies* 13(2): 148-62.

Conduzindo uma análise de correlação canônica, esse estudo busca determinar quais atividades administrativas específicas se relacionam com a efetividade organizacional. Os autores identificam nove atividades administrativas e oito itens sobre efetividade de sub-unidade organizacional. As atividades de 78 administradores são observadas e registradas para medir o compromisso nos comportamentos identificados. Para eliminar vieses de mesma fonte que podem ser introduzidos se os administradores avaliam a efetividade de sub-unidade, seus subordinados (278 ao todo) avaliam a efetividade de sub-unidade. Enquanto outros estudos examinam as atividades de administradores de sucesso (isto é, aqueles a caminho de rápida promoção), esse procura identificar os comportamentos que contribuem para a efetividade organizacional. A análise de correlação canônica entre a frequência das atividades administrativas e a efetividade de sub-unidade relatada por subordinados é usada para revelar a presença e força da relação entre os dois conjuntos de variáveis.

Resultados indicam uma variável estatística canônica significativa (correlação canônica = 0,44), porém a força da relação não é avaliada (ou seja, o índice de redundância, que é uma medida melhor da habilidade das variáveis preditoras na explicação da variação nas variáveis critério, não é relatado). A interpretação dos resultados sugere um contínuo de orientação administrativa a partir de recursos humanos orientados para a quantidade até a tradicional orientação para a qualidade. Administradores de recursos humanos orientados para a quantidade são concentrados em colocação de pessoal e atividades de motivação ou reforço, e são percebidos como tendo desempenho de quantidade em suas unidades. Esses administradores, porém,

têm limitada interação externa, limitado compromisso com atividades de controle e planejamento, ou limitada percepção do desempenho de qualidade em suas unidades. Ainda assim, os administradores tradicionais orientados para a qualidade são vistos como tendo desempenho de qualidade em suas unidades, interagindo com o meio externo e com comprometimento com atividades de controle e planejamento. Apesar de as descobertas não serem validadas, os autores sustentam que os resultados devem ajudar planejadores organizacionais a identificar as habilidades administrativas necessárias para o resultado organizacional desejado (ou seja, atividades de recursos humanos podem ajudar na realização de mais resultados, ao passo que atividades administrativas tradicionais podem melhorar a qualidade).

Van Auken, Howard E., B. Michael Doran, and Kil-Jhin Yoon (1993), "A Financial Comparison Between Korean and U.S. Firms: A Cross-Balance Sheet Canonical Correlation Analysis." *Journal of Small Business Management* 31(3): 73-83.

Nesse artigo, os autores procuram examinar as relações entre folhas de balanço e estratégias financeiras gerais de empresas coreanas de pequeno a médio porte por meio do uso de análise de correlação canônica. De uma amostra aleatória de 45 empresas coreanas, declarações de posição financeira de 1988 foram obtidas de várias contas de ativos, passivo e proventos. As relações entre ativos (dinheiro, contas a receber, inventários e ativos de longo prazo) e passivos e proventos (contas pagáveis, outros passivos correntes, débitos de longo prazo e proventos) foram exploradas usando a análise de correlação canônica. O delineamento do estudo permitiu aos autores comparar os resultados de trabalhos previamente publicados sobre empresas de pequeno e médio porte do México e dos Estados Unidos. Esse e estudos anteriores demonstraram que as estratégias financeiras de pequenas empresas são influenciadas por condições econômicas e elementos culturais. Os resultados do estudo devem ajudar na tomada de decisões de pequenos empresários que estão desenvolvendo estratégias de financiamento em economias semelhantes.

A análise resultou na significância das quatro funções canônicas. Como uma avaliação complementar das funções canônicas, os autores calculam um índice de redundância. A proporção da variância de ativos explicada pela variância de passivos é de 0,60. A variância de passivos compartilhada com a variância de ativos é de 0,24. Como as relações canônicas eram aceitáveis, os autores fizeram as seguintes interpretações sobre empresas coreanas de pequeno a médio porte: (1) elas praticam compensação com outras transações para limitar riscos, (2) elas usam a propriedade como garantia para empréstimos, (3) seus inventários são associados com contas pagáveis e (4) elas administram o risco com o uso simultâneo de saldo de



menor crédito e maior liquidez. Comparadas com as empresas norte-americanas, as empresas coreanas confiam muito no uso de dívida corrente. As descobertas ampliam estudos anteriores, os quais sugerem que estratégias financeiras de pequenas empresas em outros países dependem das restrições de marketing do país onde a empresa opera.

*Mahmood, Mo Adam, and Gary J. Mann (1993), "Measuring the Organizational Impact of Information Technology Investment: An Exploratory Study." **Journal of Management Information Systems** 10(1): 97-122.*

Esse artigo busca determinar se existe uma relação entre investimento em tecnologia de informação (IT) e o desempenho estratégico e econômico da empresa. A partir de uma pesquisa prévia que mediu o impacto de IT sobre a organização, os autores determinam quais medidas incluir no estudo. As variáveis preditoras consistem em cinco medidas de investimento em IT, e as variáveis de critério incluem seis medidas de desempenho organizacional estratégico e econômico. A análise canônica é usada para fins exploratórios, sem hipóteses específicas fornecidas. A técnica permite que os pesquisadores determinem

a presença e magnitude da associação entre IT múltipla e medidas de desempenho organizacional. A partir dos resultados, os autores oferecem hipóteses e um modelo que descreve as inter-relações entre investimento em IT e desempenho.

A análise de correlação canônica é realizada com uma amostra de 100 empresas. Os resultados indicam uma relação significativa, com 10,4% da variação nas medidas de desempenho organizacional explicada por investimento em IT. Apesar de apenas uma das cinco funções canônicas ser significativa, os autores interpretam as duas funções que explicam cerca de 86% da variação explicada total (ou seja, dos 10,4%). Pelo exame das cargas canônicas das duas funções, os autores são capazes de determinar a importância relativa de cada variável. Todas juntas, as descobertas são interpretadas como um indicativo de que o investimento em IT contribui para o desempenho organizacional quando a empresa investe em equipamentos e em treinamento de pessoal em IT. Essas conclusões levaram os autores a realizar mais pesquisas para testar as interdependências entre investimento em IT e desempenho da empresa, usando diferentes métodos e amostras.

## Apêndice 1

# Descrição da pesquisa HATCO

A pesquisa de HATCO foi um estudo conduzido com 100 clientes sobre 14 variáveis separadas em um estudo de segmentação para uma situação de *business-to-business*, especificamente uma pesquisa de clientes existentes da HATCO. Três tipos de informação foram coletados. O primeiro é a percepção da HATCO quanto a sete atributos identificados em estudos anteriores como os mais influentes na escolha de fornecedores. Os respondentes, gerentes de compras de empresas que negociam com a HATCO, avaliaram a HATCO quanto a cada atributo. O segundo tipo de informação se refere a resultados reais de compras, como as avaliações da satisfação de cada respondente com a HATCO ou o percentual das compras do produto daquele respondente da HATCO. O terceiro tipo de informação contém características gerais das companhias de compra (p.ex., tamanho da firma, tipo de indústria). Uma descrição breve das variáveis é dada abaixo.

### Percepções de HATCO

Cada uma das variáveis foi medida em uma escala gráfica de avaliação, na qual uma linha de 10 centímetros foi desenhada entre os pontos extremos rotulados como “Ruim” e “Excelente”.

Respondentes indicaram suas percepções fazendo uma marca em algum ponto da linha. A marca é então medida e a distância a partir de 0 (em centímetros) é armazenada. O resultado é uma escala que varia de 0 a 10, arredondada na primeira casa após a vírgula.

Os sete atributos HATCO avaliados pelos respondentes são:

- $X_1$  Velocidade de entrega – o tempo que demora para entregar o produto uma vez que a encomenda tenha sido confirmada
- $X_2$  Nível de preço – nível percebido de preço cobrado por fornecedores
- $X_3$  Flexibilidade de preço – disposição percebida dos representantes da HATCO em negociar preço em todos os tipos de compras
- $X_4$  Imagem do fabricante – imagem geral do fabricante ou fornecedor
- $X_5$  Serviço geral – nível geral de serviço necessário para manutenção de uma relação satisfatória entre fornecedor e comprador

- $X_6$  Imagem da equipe de venda – imagem geral da equipe de venda de um fabricante
- $X_7$  Qualidade de produto – nível percebido da qualidade de um produto em particular (p.ex., desempenho ou benefícios)

### Resultados de compra

Duas medidas específicas foram obtidas refletindo os resultados das relações de compra do respondente com a HATCO.

Essas medidas incluem:

- $X_9$  Nível de uso – quanto do produto total da firma é comprado da HATCO, medido em uma escala de 100 pontos percentuais, variando de 0 a 100
- $X_{10}$  Nível de satisfação – o quão satisfeito o comprador está com negócios anteriores com a HATCO, medido na mesma escala gráfica usada nas percepções de  $X_1$  a  $X_7$

### Características do comprador

As cinco características das firmas respondentes usadas no estudo, algumas métricas e outras não, são as seguintes:

- $X_8$  Tamanho da firma – tamanho da firma relativamente a outras neste mercado. Esta variável tem duas categorias: 1 = grande, 0 = pequeno
- $X_{11}$  Compra por especificação – extensão em que um comprador em particular avalia cada compra separadamente (análise de valor total) versus o uso de compra por especificação, que detalha precisamente as características desejadas do produto. Esta variável tem duas categorias: 1 = emprega a abordagem de análise de valor total, avaliando separadamente cada compra; 0 = usa compra por especificação
- $X_{12}$  Estrutura de procura – método para procurar ou comprar produtos em uma companhia em especial. Esta variável apresenta duas categorias: 1 = procura centralizada, 0 = procura descentralizada
- $X_{13}$  Tipo de indústria – classificação da indústria à qual um comprador pertence. Esta variável admite duas categorias: 1 = indústria A, 0 = outras indústrias
- $X_{14}$  Tipo de situação de compra – tipo de situação encarada pelo comprador. Esta variável tem três categorias: 1 = tarefa nova, 2 = recompra modificada, 3 = recompra simples

**TABELA 1** Correlação canônica de uso de crédito (número de cartões e taxa de uso) com características do usuário (tamanho da família e renda familiar)

Variáveis da pesquisa		
Medidas de uso de crédito		Medidas de características do usuário
<ul style="list-style-type: none"> <li>• Número de cartões de crédito usados</li> <li>• Despesas mensais médias em dólares com cartões de crédito</li> </ul>		<ul style="list-style-type: none"> <li>• Tamanho da família</li> <li>• Renda familiar</li> </ul>
Elementos de análise canônica		
<b>Composição de variáveis dependentes</b>	<b>Correlação canônica</b>	<b>Composição de variáveis independentes</b>
Variável estatística canônica dependente	$R_c$	Variável estatística canônica independente

**TABELA 2** Análise de correlação canônica que relaciona níveis de uso e satisfação com a HATCO e percepções da HATCO

Medidas de ajuste geral do modelo para análise de correlação canônica				
Função canônica	Correlação canônica	$R^2$ canônico	Estatística $F$	Probabilidade
1	0,937	0,878	30,235	0,000
2	0,510	0,260	5,391	0,000
Testes multivariados de significância				
Estatística	Valor	Estatística $F$ aproximada	Probabilidade	
Lambda de Wilks	0,090	30,235	0,000	
Traço de Pillai	1,138	17,348	0,000	
Traço de Hotelling	7,535	48,441	0,000	
gcr de Roy	0,878			

**TABELA 3** Cálculo dos índices de redundância para a primeira função canônica

Variável estatística/variáveis	Carga canônica	Carga canônica quadrada	Carga canônica quadrada média	$R^2$ canônico	Índice de redundância <sup>a</sup>
Variáveis dependentes					
$X_9$ Nível de uso	0,913	0,834			
$X_{10}$ Nível de satisfação	0,936	0,876			
Variável estatística dependente		1,710	0,855	0,878	0,751
Variáveis independentes					
$X_1$ Velocidade de entrega	0,764	0,584			
$X_2$ Nível de preço	0,061	0,004			
$X_3$ Flexibilidade de preço	0,624	0,389			
$X_4$ Imagem do fabricante	0,414	0,171			
$X_5$ Serviço geral	0,765	0,585			
$X_6$ Imagem da equipe de vendas	0,348	0,121			
$X_7$ Qualidade do produto	-0,278	0,077			
Variável estatística independente		1,931	0,276	0,878	0,242

<sup>a</sup> O índice de redundância é calculado como a carga quadrada média vezes o  $R^2$  canônico.

**TABELA 4** Análise de redundância de variáveis estatísticas dependentes e independentes para ambas as funções canônicas

Variância padronizada das variáveis dependentes explicada					
Função canônica	Por sua própria variável estatística canônica (variância compartilhada)		$R^2$ canônico	Pela variável estatística canônica oposta (redundância)	
	Percentual	Percentual cumulativo		Percentual	Percentual cumulativo
1	0,855	0,855	0,878	0,751	0,751
2	0,145	1,000	0,260	0,038	0,789
Variância padronizada das variáveis independentes explicada					
Função canônica	Por sua própria variável estatística canônica (variância compartilhada)		$R^2$ canônico	Pela variável estatística canônica oposta (redundância)	
	Percentual	Percentual cumulativo		Percentual	Percentual cumulativo
1	0,276	0,276	0,878	0,242	0,242
2	0,082	0,358	0,260	0,021	0,263

**TABELA 5** Pesos canônicos para as duas funções canônicas

	Pesos canônicos	
	Função 1	Função 2
Coeficientes canônicos padronizados para as variáveis independentes		
$X_1$ Velocidade de entrega	0,225	2,965
$X_2$ Nível de preço	0,103	2,868
$X_3$ Flexibilidade de preço	0,569	0,160
$X_4$ Imagem do fabricante	0,348	-1,456
$X_5$ Serviço geral	0,445	1,530
$X_6$ Imagem da equipe de vendas	-0,051	0,736
$X_7$ Qualidade do produto	0,001	0,478
Coeficientes canônicos padronizados para as variáveis dependentes		
$X_9$ Nível de uso	0,501	1,330
$X_{10}$ Nível de satisfação	0,580	-1,298



**TABELA 6** Estrutura canônica para as duas funções canônicas

	Cargas canônicas	
	Função 1	Função 2
Correlações entre as variáveis independentes e suas variáveis estatísticas canônicas		
$X_1$ Velocidade de entrega	0,764	0,109
$X_2$ Nível de preço	0,061	0,141
$X_3$ Flexibilidade de preço	0,624	0,123
$X_4$ Imagem do fabricante	0,414	-0,626
$X_5$ Serviço geral	0,765	0,222
$X_6$ Imagem da equipe de vendas	0,348	-0,199
$X_7$ Qualidade do produto	-0,278	0,219
Correlações entre as variáveis dependentes e suas variáveis estatísticas canônicas		
$X_9$ Nível de uso	0,913	0,408
$X_{10}$ Nível de satisfação	0,936	-0,352
	Cargas cruzadas canônicas <sup>a</sup>	
	Função 1	Função 2
Correlações entre as variáveis independentes e variáveis estatísticas canônicas dependentes		
$X_1$ Velocidade de entrega	0,716	0,056
$X_2$ Nível de preço	0,058	0,072
$X_3$ Flexibilidade de preço	0,584	0,063
$X_4$ Imagem do fabricante	0,388	-0,319
$X_5$ Serviço geral	0,717	0,113
$X_6$ Imagem da equipe de vendas	0,326	-0,102
$X_7$ Qualidade do produto	-0,261	0,112
Correlações entre as variáveis dependentes e variáveis estatísticas canônicas independentes		
$X_9$ Nível de uso	0,855	0,208
$X_{10}$ Nível de satisfação	0,877	-0,180

<sup>a</sup> As cargas cruzadas canônicas são fornecidas pelo SAS porque SPSS não apresenta as cargas cruzadas\*.

\* N. de R. T.: É possível obter as cargas cruzadas com o SPSS por meio de sintaxe de comandos da macro CANCECORR.

**TABELA 7** Análise de sensibilidade dos resultados de correlação canônica conseqüentes da remoção de uma variável independente

		Resultados após eliminação de		
	Variável estatística completa	$X_1$	$X_2$	$X_7$
Correlação canônica ( $R$ )	0,937	0,936	0,937	0,937
Raiz canônica ( $R^2$ )	0,878	0,876	0,878	0,878
VARIÁVEL ESTATÍSTICA INDEPENDENTE				
<i>Cargas canônicas</i>				
$X_1$ Velocidade de entrega	0,764	omitida	0,765	0,764
$X_2$ Nível de preço	0,061	0,062	omitida	0,061
$X_3$ Flexibilidade de preço	0,624	0,624	0,624	0,624
$X_4$ Imagem do fabricante	0,414	0,413	0,414	0,415
$X_5$ Serviço geral	0,765	0,766	0,766	0,765
$X_6$ Imagem da equipe de vendas	0,348	0,348	0,348	0,348
$X_7$ Qualidade do produto	-0,278	-0,278	-0,278	omitida
Variância compartilhada	0,276	0,225	0,322	0,309
Redundância	0,242	0,197	0,282	0,271
VARIÁVEL ESTATÍSTICA DEPENDENTE				
<i>Cargas canônicas</i>				
$X_9$ Nível de uso	0,913	0,915	0,914	0,913
$X_{10}$ Nível de satisfação	0,936	0,934	0,935	0,936
Variância compartilhada	0,855	0,855	0,855	0,855
Redundância	0,750	0,749	0,750	0,750