

## CAPÍTULO

# Análise Fatorial



"Frequentemente, dentre as muitas variáveis que medimos, algumas estão mais relacionadas entre si do que com outras variáveis. A análise fatorial permite-nos estudar esses grupos de variáveis que tendem a se relacionar entre si e estimar quais são as razões subjacentes que fazem com que tais variáveis estejam mais fortemente correlacionadas."

Jeff Miller, vice-presidente de consultoria e análise, Burke, Inc.

### OBJETIVOS

Após a leitura deste capítulo, o aluno deverá estar apto a:

1. Descrever o conceito de análise fatorial e explicar em que essa análise difere da análise da variância, da regressão múltipla e da análise discriminante.
2. Discutir como fazer análise fatorial, inclusive a formulação do problema, a construção da matriz de correlação, a escolha de um método apropriado, a determinação do número de fatores, a rotação e a interpretação dos fatores.
3. Entender a distinção entre métodos de análise fatorial de componentes principais e de análise fatorial comum.
4. Explicar a seleção de variáveis substitutas e suas aplicações, enfatizando sua utilização em análises subsequentes.
5. Descrever o procedimento para verificar o ajuste de um modelo de análise fatorial, utilizando as correlações observadas e reproduzidas.

### Aspectos gerais

Em análise da variância (Capítulo 16), regressão (Capítulo 17) e análise discriminante (Capítulo 18), uma das variáveis é identificada claramente como a variável dependente. Vamos agora abordar um procedimento em que as variáveis não são classificadas como independentes ou dependentes; ao contrário, examina-se todo o conjunto de relações interdependentes entre variáveis. Neste capítulo, discute-se o conceito básico de análise fatorial, apresentando-se uma exposição do modelo fatorial. São descritos os diversos passos dessa análise, ilustrando-os no contexto da análise dos componentes principais. Apresenta-se, em seguida, uma aplicação da análise fatorial comum. Começamos com alguns exemplos para ilustrar a utilidade da análise fatorial.

### PESQUISA ATIVA

#### PROJETO DA LOJA DE DEPARTAMENTOS

##### *Análise fatorial*

No projeto da loja de departamentos, as classificações atribuídas pelos entrevistados a 21 diferentes estilos de vida foram analisados fatorialmente a fim de determinar os fatores fundamentais que determinam os modos de vida. Destacaram-se sete fatores: preferência pelo cartão bancário versus cartão da loja, propensão ao crédito,

aversão ao crédito, orientação do tempo de lazer, favoritismo do cartão de crédito, conveniência do crédito e conhecimento do custo do cartão de crédito. Esses fatores, juntamente com as características demográficas, foram utilizados para traçar um perfil dos segmentos formados como resultado de um *cluster*.

### PESQUISA REAL

#### *Análise fatorial ganha juros em bancos*

Como os clientes costumam avaliar os bancos? Em uma pesquisa, foi solicitado aos entrevistados que classificassem a importância de 15 atributos dos bancos. Utilizou-se uma escala de cinco pontos, de "não-importante" a "muito importante". Os dados foram estudados via análise dos componentes principais.

O estudo teve como resultado uma solução de quatro fatores, rotulados como serviços tradicionais, conveniência, visibilidade e competência. Os serviços tradicionais incluíram taxas de juro sobre empréstimos, reputação na comunidade, taxas baixas para contas correntes, atendimento cordial e personalizado, extratos mensais compreensíveis e acesso facilitado a empréstimos. A conveniência compreendeu: localização adequada da agência, localização conveniente de caixas automáticos, rapidez do serviço e horários de funcionamento. O fator visibilidade incluiu: recomendações de amigos e parentes, estrutura física agradável, envolvimento com a comunidade e facilidade de obtenção de empréstimos. A competência consistiu em: qualificação dos funcionários e disponibilidade de serviços bancários auxiliares. Concluiu-se que os clientes avaliavam os bancos com base nos quatro fatores básicos, serviços tradicionais, conveniência,

visibilidade e competência, e que os bancos devem apresentar excelência nesses fatores para projetar uma boa imagem.<sup>1</sup>

## CONCEITO BÁSICO

A **análise fatorial** é um nome genérico que denota uma classe de procedimentos utilizados essencialmente para redução e resumo dos dados. Na pesquisa de marketing, pode haver um grande número de variáveis, a maioria delas correlacionadas, que devem ser reduzidas a um nível gerenciável. Estudam-se as relações entre conjuntos de muitas variáveis inter-relacionadas representando-as em termos de alguns fatores fundamentais. Por exemplo, a imagem de um estabelecimento pode ser medida pedindo aos entrevistados que o avaliem segundo uma série de itens em uma escala diferencial semântica. Analisam-se então essas avaliações para determinar os fatores fundamentais da imagem do estabelecimento.

### análise fatorial

Um tipo de procedimento destinado essencialmente à redução e ao resumo dos dados.

Em análise da variância, regressão múltipla e análise discriminante, uma variável é considerada como dependente, ou variável critério, e as outras, como variáveis independentes ou previsoras. Em análise fatorial, entretanto, não se faz tal distinção. Ao contrário, a análise fatorial é uma **técnica de interdependência**, no sentido de que examinamos todo um conjunto de relações interdependentes.<sup>2</sup>

### técnica de interdependência

Técnica estatística multivariada em que se examina todo o conjunto de relações interdependentes.

Utiliza-se a análise fatorial nas seguintes circunstâncias:

1. Para identificar dimensões latentes ou **fatores** que expliquem as correlações entre um conjunto de variáveis. Por exemplo, pode-se utilizar um conjunto de afirmações sobre estilos de vida para avaliar os perfis psicográficos de consumidores. A seguir, essas afirmações podem ser analisadas fatorialmente para identificar os fatores psicográficos subjacentes, conforme ilustrado no exemplo da loja de departamentos.

### fator

Uma dimensão subjacente que explica as correlações entre um conjunto de variáveis.

<sup>1</sup> Barbara R. Lewis e Sotiris Spyropoulos, "Service Failures and Recovery in Retail Banking: The Customers' Perspective," *The International Journal of Bank Marketing*, 19(1) (2001): 37-48; e James M. Sinukula e Leanna Lavtor, "Positioning in the Financial Services Industry: A Look at the Decomposition of Image," em Jon M. Hawes e George B. Glisan, Eds., *Developments in Marketing Science*, Vol. 10 (Akron, OH: Academy of Marketing Science, 1987): 439-442.

<sup>2</sup> Para uma discussão detalhada da análise fatorial, ver Jacques Tacq, *Multivariate Analysis Techniques in Social Science Research* (Thousand Oaks, CA: Sage Publications, 1997); George H. Dunteman, *Principal Components Analysis* (Newbury Park, CA: Sage Publications, 1989).

2. Para identificar um conjunto novo, menor, de variáveis não-correlacionadas para substituir o conjunto original de variáveis correlacionadas na análise multivariada subsequente (regressão ou análise discriminante). Por exemplo, os fatores psicográficos identificados podem ser usados como variáveis independentes para explicar as diferenças entre clientes fiéis e eventuais.
3. Para identificar, em um conjunto maior, um conjunto menor de variáveis que se destacam para uso em uma análise multivariada subsequente. Por exemplo, algumas das afirmações originais sobre estilos de vida que se correlacionam fortemente com os fatores identificados podem ser usadas como variáveis independentes para explicar as diferenças entre os clientes fiéis e os eventuais.

A análise fatorial encontra numerosas aplicações na pesquisa de marketing. Por exemplo:

- Pode ser usada na segmentação de mercado para identificar as variáveis latentes segundo as quais se agrupam os consumidores. Os compradores de carros novos podem ser agrupados com base na ênfase relativa que dão à economia, à conveniência, ao desempenho, ao conforto e ao luxo. Isso pode resultar em cinco segmentos: os que buscam economia, os que buscam conveniência, os que procuram desempenho, os que procuram conforto e os que desejam luxo.
- Na pesquisa de um produto, pode-se empregar a análise fatorial para determinar os atributos de uma marca que influenciam a escolha do consumidor. As marcas de dentífricos podem ser avaliadas em termos de proteção contra cáries, branqueamento dos dentes, gosto, frescor do hálito e preço.
- Em estudos de propaganda, pode-se aplicar a análise fatorial para definir os hábitos de consumo do mercado-alvo. Os usuários de alimentos congelados podem ser espectadores constantes de TV a cabo, que assistem muitos filmes e preferem música *country*.
- Em estudos sobre preços, pode ser utilizada para identificar as características dos consumidores que são sensíveis a preços. Esses consumidores podem ser metódicos, ter mentalidade econômica e centrada no lar.

## MODELO DE ANÁLISE FATORIAL

Matematicamente, a análise fatorial é algo semelhante à análise de regressão múltipla, pelo fato de cada variável ser expressa como uma combinação linear de fatores subjacentes. A quantidade de variância que uma variável compartilha com todas as outras variáveis incluídas na análise é chamada de **comunalidade**. A covariação entre as variáveis é descrita em termos de um pequeno número de fatores comuns, mais um fator único (ou exclusivo) para cada variável. Esses fatores não são observados abertamente. Se as variáveis forem padronizadas, o modelo fatorial pode ser representado como:

$$X_i = A_{i1} F_1 + A_{i2} F_2 + A_{i3} F_3 + \dots + A_{im} F_m + V_i U_i$$

onde

$X_i$  = *i*ésima variável padronizada

$A_{ij}$  = coeficiente padronizado de regressão múltipla da variável *i* sobre o fator comum *j*

$F$  = fator comum

$V_i$  = coeficiente padronizado de regressão da variável *i* sobre o fator único *i*

$U_i$  = o fator único para a variável *i*

$m$  = número de fatores comuns

Os fatores únicos não são correlacionados uns com os outros e com os fatores comuns.<sup>3</sup> Os fatores comuns podem, eles próprios, ser expressos como combinações lineares de variáveis observáveis.

$$F_j = W_{j1} X_1 + W_{j2} X_2 + W_{j3} X_3 + \dots + W_{jk} X_k$$

onde

$F_j$  = estimativa do *j*ésimo fator

$W_j$  = peso ou coeficiente do escore fatorial

$k$  = número de variáveis

É possível escolher pesos ou coeficientes de escore fatorial de modo que o primeiro fator explique a maior parte da variância total. Em seguida, pode-se escolher um segundo conjunto de pesos, de modo que o segundo fator responda pela maior parte da variância residual, desde que não seja correlacionado com o primeiro fator. O mesmo princípio pode ser aplicado à escolha de pesos adicionais para os outros fatores. Assim, os fatores podem ser estimados de modo que seus escores, ao contrário dos valores das variáveis originais, não sejam correlacionados. Além disso, o primeiro fator responde pela maior variância nos dados, o segundo fator pela segunda variância mais alta, e assim por diante. No Apêndice 19.1 apresentamos um tratamento técnico do modelo de análise fatorial. Há várias estatísticas associadas à análise fatorial.

## ESTATÍSTICAS ASSOCIADAS À ANÁLISE FATORIAL

As estatísticas-chave associadas à análise fatorial são:

**Teste de esfericidade de Bartlett.** Uma estatística de teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas na população. Em outras palavras, a matriz de correlação da população é uma matriz de identidade; cada variável se correlaciona perfeitamente com ela própria ( $r = 1$ ) mas não apresenta correlação com as outras variáveis ( $r = 0$ ).

**Matriz de correlação.** É um triângulo inferior da matriz que exhibe as correlações simples,  $r$ , entre todos os pares possíveis de variáveis incluídas na análise. Os elementos da diagonal, que são todos iguais a 1, em geral são omitidos.

<sup>3</sup> Ver A. Adam Ding, "Prediction Intervals, Factor Analysis Models, and High-Dimensional Empirical Linear Prediction," *Journal of the American Statistical Association*, 94 (446) (June 1999): 446-455; e W. R. Dillon e M. Goldstein, *Multivariate Analysis: Methods and Applications* (New York: John Wiley, 1984): 23-99.

**Comunalidade.** Trata-se da porção da variância que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns.

**Autovalor (eigen value).** Representa a variância total explicada por cada fator.

**Cargas fatoriais.** São correlações simples entre as variáveis e os fatores.

**Gráfico das cargas fatoriais.** Gráfico das variáveis originais utilizando as cargas de fatores como coordenadas.

**Matriz de fatores.** Contém as cargas fatoriais de todas as variáveis em todos os fatores extraídos.

**Escore fatorial.** Escores compostos estimados para cada entrevistado nos fatores derivados.

**Medida de adequação da amostra de Kaiser-Meyer-Olkin (KMO).** Trata-se de um índice usado para avaliar a adequação da análise fatorial. Valores altos (entre 0,5 e 1,0) indicam que a análise fatorial é apropriada. Valores abaixo de 0,5 indicam que a análise fatorial pode ser inadequada.

**Porcentagem de variância.** É a porcentagem da variância total atribuída a cada fator.

**Resíduos.** Diferenças entre as correlações observadas, dadas na matriz de correlação de entrada e as correlações reproduzidas, conforme estimado pela matriz de fatores.

**Gráfico de declive (scree plot).** É um gráfico dos autovalores versus número de fatores por ordem de extração.

Na próxima seção, descreveremos as aplicações dessas estatísticas no contexto do procedimento da análise fatorial.

## COMO FAZER ANÁLISE FATORIAL

A Figura 19.1 ilustra os diversos passos de uma análise fatorial. O primeiro consiste em formular o problema de análise fatorial e identificar as variáveis a serem analisadas. A seguir, é construída uma matriz de correlação dessas variáveis, escolhendo-se um método de análise fatorial. O pesquisador decide quanto ao número de fatores a serem extraídos e quanto ao método de rotação. Em seguida, é preciso interpretar os fatores rotacionados. Dependendo dos objetivos, podem-se calcular os escores fatoriais ou selecionar variáveis substitutas para representar os fatores em uma análise multivariada subsequente. Finalmente, determina-se o ajuste do modelo de análise fatorial. Todos esses aspectos são abordados com maior detalhe nas seções seguintes.<sup>4</sup>

### Formular o problema

A formulação do problema inclui várias tarefas. Em primeiro lugar, devemos identificar os objetivos da análise fatorial. As variáveis a serem incluídas na análise fatorial devem ser especificadas com base em pesquisas anteriores, na teoria e no julgamento do pesquisador. É importante medir adequadamente as variáveis em

<sup>4</sup> Para uma aplicação recente da análise fatorial, ver Nick Johns e Szilvia Gyimothy, "Market Segmentation and the Prediction of Tourist Behavior: The Case of Bornholm, Denmark," *Journal of Travel Research*, 40 (3) (February 2002): 316-327.

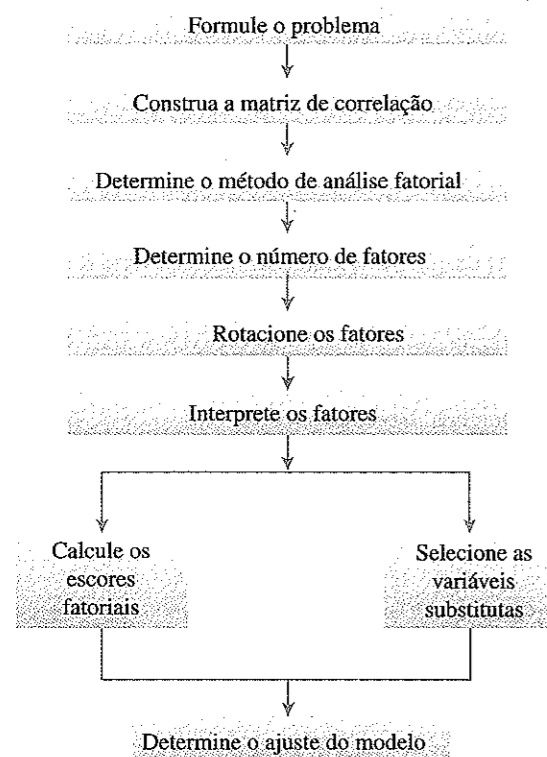


Figura 19.1 Como fazer análise fatorial.

uma escala intervalar ou razão. Deve-se utilizar um tamanho adequado de amostra. Como orientação geral, deve haver pelo menos quatro a cinco vezes mais observações (tamanho da amostra) do que variáveis.<sup>5</sup> Em muitas situações de pesquisa de marketing, o tamanho da amostra é pequeno e essa relação é consideravelmente menor. Em tais casos, os resultados devem ser interpretados com cautela.

A título de ilustração da análise fatorial, suponhamos que um pesquisador queira avaliar os benefícios que os consumidores esperam da compra de um dentífrico. Foi entrevistada em um supermercado uma amostra de 30 pessoas, para que indicassem seu grau de concordância com as seguintes afirmações, utilizando uma escala de sete pontos (1 = discordo totalmente, 7 = concordo totalmente):

- $V_1$ : É importante comprar um creme dental que evite cáries.
- $V_2$ : Gosto de um creme dental que clareie os dentes.
- $V_3$ : Um creme dental deve fortificar as gengivas.
- $V_4$ : Prefiro um creme dental que refresque o hálito.
- $V_5$ : Manter os dentes saudáveis não é uma vantagem importante de um creme dental.
- $V_6$ : O aspecto mais importante na compra de um creme dental é tornar os dentes bonitos.

<sup>5</sup> David J. Bartholomew e Martin Knott, *Latent Variable Models and Factor Analysis* (London, UK: Edward Arnold Publishers, 1999); Joseph F. Hair, Jr., Ralph E. Anderson, Ronald L. Tatham e William C. Black, *Multivariate Data Analysis with Readings*, 5th ed. (Upper Saddle River, NJ: Prentice-Hall, Inc., 1998); e Alexander Basilevsky, *Statistical Factor Analysis & Related Methods: Theory & Applications* (New York: John Wiley, 1994).

Os dados obtidos figuram na Tabela 19.1. Construiu-se uma matriz de correlação com base nesses dados de classificação.

### Construir a matriz de correlação

O procedimento analítico se baseia em uma matriz de correlações entre as variáveis. Um exame dessa matriz permite uma boa visualização. Para que a análise fatorial seja apropriada, as variáveis devem ser correlacionadas. Na prática, isso costuma ocorrer. Se as correlações entre as variáveis forem pequenas, a análise fatorial pode ser inadequada. Esperaríamos também que as variáveis altamente correlacionadas umas com as outras se correlacionem também com o(s) mesmo(s) fator(es).

Existem estatísticas formais para testar a conveniência do modelo fatorial. Pode-se aplicar o teste de esfericidade de Bartlett para testar a hipótese nula de que as variáveis não sejam correlacionadas na população. Em outras palavras, a matriz da correlação populacional é uma matriz identidade. Em uma matriz identidade, todos os termos da diagonal são 1, e todos os termos fora da diagonal são 0. A estatística de teste da esfericidade se baseia em uma transformação qui-quadrado do determinante da matriz de correlação. Um valor elevado dessa estatística de teste favorece a rejeição da hipótese nula. Se essa hipótese não puder ser rejeitada, então a conveniência da análise fatorial deve ser questionada. Outra estatística útil é a medida de adequação da amostra de Kaiser-Meyer-Olkin (KMO). Esse índice compara as magnitudes dos coeficientes de correlação observados com as magnitudes dos coeficientes de correlação parcial. Pequenos valores da estatística KMO indicam que as correlações entre pares de variáveis não podem ser explicadas por outras variáveis e que a análise fatorial pode ser inapropriada. Geralmente, um valor superior a 0,5 é desejável.

A Tabela 19.2 mostra a matriz de correlação, construída a partir dos dados obtidos para entender os benefícios de um dentífrico. Há correlações relativamente elevadas entre  $V_1$  (prevenção de cáries),  $V_3$  (gengivas fortes) e  $V_5$  (dentes saudáveis). É de esperar que essas variáveis se correlacionem com o mesmo conjunto de fatores. Da mesma forma, verificam-se correlações relativamente elevadas entre  $V_2$  (dentes claros),  $V_4$  (hálito puro) e  $V_6$  (boa aparência dos dentes). Essas variáveis também devem se correlacionar com os mesmos fatores.<sup>6</sup>

A Tabela 19.3 apresenta os resultados da análise fatorial. A hipótese nula de que a matriz de correlação da população seja uma matriz identidade é rejeitada pelo teste de esfericidade de Bartlett. A estatística qui-quadrado aproximada é 111,314, com 15 graus de liberdade, significativa ao nível de 0,05. O valor da estatística KMO (0,660) também é grande, (> 0,5). Assim, a análise fatorial pode ser considerada uma técnica apropriada para analisar a matriz de correlação da Tabela 19.2.

<sup>6</sup> A análise fatorial é influenciada pelo tamanho relativo das correlações, mais que pelo tamanho absoluto.

TABELA 19.1

### Classificação dos atributos de um creme dental

NÚMERO DO ENTREVISTADO	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$
1	7,00	3,00	6,00	4,00	2,00	4,00
2	1,00	3,00	2,00	4,00	5,00	4,00
3	6,00	2,00	7,00	4,00	1,00	3,00
4	4,00	5,00	4,00	6,00	2,00	5,00
5	1,00	2,00	2,00	3,00	6,00	2,00
6	6,00	3,00	6,00	4,00	2,00	4,00
7	5,00	3,00	6,00	3,00	4,00	3,00
8	6,00	4,00	7,00	4,00	1,00	4,00
9	3,00	4,00	2,00	3,00	6,00	3,00
10	2,00	6,00	2,00	6,00	7,00	6,00
11	6,00	4,00	7,00	3,00	2,00	3,00
12	2,00	3,00	1,00	4,00	5,00	4,00
13	7,00	2,00	6,00	4,00	1,00	3,00
14	4,00	6,00	4,00	5,00	3,00	6,00
15	1,00	3,00	2,00	2,00	6,00	4,00
16	6,00	4,00	6,00	3,00	3,00	4,00
17	5,00	3,00	6,00	3,00	3,00	4,00
18	7,00	3,00	7,00	4,00	1,00	4,00
19	2,00	4,00	3,00	3,00	6,00	3,00
20	3,00	5,00	3,00	6,00	4,00	6,00
21	1,00	3,00	2,00	3,00	5,00	3,00
22	5,00	4,00	5,00	4,00	2,00	4,00
23	2,00	2,00	1,00	5,00	4,00	4,00
24	4,00	6,00	4,00	6,00	4,00	7,00
25	6,00	5,00	4,00	2,00	1,00	4,00
26	3,00	5,00	4,00	6,00	4,00	7,00
27	4,00	4,00	7,00	2,00	2,00	5,00
28	3,00	7,00	2,00	6,00	4,00	3,00
29	4,00	6,00	3,00	7,00	2,00	7,00
30	2,00	3,00	2,00	4,00	7,00	2,00

TABELA 19.2

### Matriz de correlação

VARIÁVEIS	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$
$V_1$	1,00					
$V_2$	-0,053	1,00				
$V_3$	0,873	-0,155	1,00			
$V_4$	-0,086	0,572	-0,248	1,00		
$V_5$	-0,858	0,020	-0,778	-0,007	1,00	
$V_6$	0,004	0,640	-0,018	0,640	-0,136	1,00

TABELA 19.3

## Resultados da análise de componentes principais

Teste de esfericidade de Bartlett

Qui-quadrado aproximado = 111,314, gl = 15, significância = 0,00000

Medida de adequação da amostra de Kaiser-Meyer-Olkin = 0,660

## COMUNALIDADES

VARIÁVEL	INICIAL	EXTRAÇÃO
V <sub>1</sub>	1,000	0,926
V <sub>2</sub>	1,000	0,723
V <sub>3</sub>	1,000	0,894
V <sub>4</sub>	1,000	0,739
V <sub>5</sub>	1,000	0,878
V <sub>6</sub>	1,000	0,790

## AUTOVALORES INICIAIS (INITIAL EIGEN VALUES)

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,731	45,520	45,520
2	2,218	36,969	82,488
3	0,442	7,360	89,848
4	0,341	5,688	95,536
5	0,183	3,044	98,580
6	0,085	1,420	100,000

## EXTRAÇÃO DA SOMA DO QUADRADO DAS CARGAS

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,731	45,520	45,520
2	2,218	36,969	82,488

## MATRIZ DE FATORES

	FATOR 1	FATOR 2
V <sub>1</sub>	0,928	0,253
V <sub>2</sub>	-0,301	0,795
V <sub>3</sub>	0,936	0,131
V <sub>4</sub>	-0,342	0,789
V <sub>5</sub>	-0,869	-0,351
V <sub>6</sub>	-0,177	0,871

## ROTAÇÃO DA SOMA DO QUADRADO DAS CARGAS

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,688	44,802	44,802
2	2,261	37,687	82,488

## MATRIZ DE FATORES ROTACIONADOS

	FATOR 1	FATOR 2
V <sub>1</sub>	0,962	-0,027
V <sub>2</sub>	-0,057	0,848
V <sub>3</sub>	0,934	-0,146
V <sub>4</sub>	-0,098	0,854
V <sub>5</sub>	-0,933	-0,084
V <sub>6</sub>	-0,083	0,885

(continua)

TABELA 19.3

## Resultados da análise de componentes principais

## MATRIZ DOS COEFICIENTES DE ESCORES FATORIAIS

	FATOR 1	FATOR 2
V <sub>1</sub>	0,358	0,011
V <sub>2</sub>	-0,001	0,375
V <sub>3</sub>	0,345	-0,043
V <sub>4</sub>	-0,017	0,377
V <sub>5</sub>	-0,350	-0,059
V <sub>6</sub>	0,052	0,395

## MATRIZ DE CORRELAÇÕES REPRODUZIDAS

	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>	V <sub>5</sub>	V <sub>6</sub>
V <sub>1</sub>	0,926*	0,024	-0,029	0,031	0,038	-0,053
V <sub>2</sub>	-0,078	0,723*	0,022	-0,158	0,038	-0,105
V <sub>3</sub>	0,902	-0,177	0,894*	-0,031	0,081	0,033
V <sub>4</sub>	-0,117	0,730	-0,217	0,739*	-0,027	-0,107
V <sub>5</sub>	-0,895	-0,018	-0,859	0,020	0,878*	0,016
V <sub>6</sub>	0,057	0,746	0,051	0,748	-0,152	0,790*

\* O triângulo inferior esquerdo contém a matriz de correlações reproduzidas; a diagonal contém as comunalidades; o triângulo superior direito, os resíduos entre as correlações observadas e as correlações reproduzidas.

## Determinar o método de análise fatorial

Uma vez definido que a análise fatorial é uma técnica adequada para analisar os dados, devemos selecionar um método apropriado. A abordagem usada para deduzir os pesos, ou coeficientes dos escores fatoriais, diferencia os diversos métodos de análise fatorial. As duas abordagens básicas são a análise dos componentes principais e a análise fatorial comum. Na *análise dos componentes principais*, leva-se em conta a variância total nos dados. A diagonal da matriz de correlação consiste em unidades, e a variância plena é introduzida na matriz de fatores. Recomenda-se a análise dos componentes principais quando a preocupação maior for determinar o número mínimo de fatores que respondem pela máxima variância nos dados para utilização em análises multivariadas subsequentes. Os fatores são chamados de componentes principais.

Na *análise fatorial comum*, os fatores são estimados com base apenas na variância comum. As comunalidades são inseridas na diagonal da matriz de correlação. Esse método é adequado quando a preocupação principal é identificar as dimensões subjacentes e a variância comum é um elemento de interesse. Esse método é conhecido também como *fatoração do eixo principal*.

## análise de componentes principais

Uma abordagem da análise fatorial que leva em conta a variância total nos dados.

## análise fatorial comum

Abordagem da análise fatorial que leva em conta apenas a variância comum nos dados.

Existem ainda outros métodos para estimar os fatores comuns, incluindo o método dos mínimos quadrados não-ponderados, o método dos mínimos quadrados generalizados, a máxima verossimilhança, o método alfa e a fatoração da imagem. Esses métodos são complexos e não são recomendados para usuários inexperientes.<sup>7</sup>

A Tabela 19.3 mostra a aplicação da análise de componentes principais ao exemplo do creme dental. Em "Comunalidades", coluna "Inicial", pode-se ver que a comunalidade para cada variável, de V<sub>1</sub> a V<sub>6</sub>, é 1,0, pois foram inseridas unidades na diagonal da matriz de correlação. A tabela rotulada "Autovalores iniciais" apresenta os autovalores. Como era de esperar, os autovalores se apresentam em ordem decrescente de magnitude, à medida que caminhamos do fator 1 para o fator 6. O autovalor de um fator indica a variância total atribuída àquele fator. A variância total atribuída a todos os seis fatores é 6,00, que é igual ao número de variáveis. O fator 1 responde por uma variância de 2,731, ou seja, (2,731/6), ou 45,52% da variância total. Da mesma forma, o segundo fator responde por (2,218/6) ou 36,97% da variância total e os dois primeiros fatores combinados respondem por 82,49% da variância total. Várias considerações entram em jogo na determinação do número de fatores que devem ser usados na análise.

<sup>7</sup> Ver Wagner A. Kamakura e Michel Wedel, "Factor Analysis and Missing Data," *Journal of Marketing Research*, 37 (4) (November 2000): 490-498; Sangit Chatterjee, Linda Jamieson e Frederick Wiseman, "Identifying Most Influential Observations in Factor Analysis," *Marketing Science* (Spring 1991): 145-160; e Frank Acito e Ronald D. Anderson, "A Monte Carlo Comparison of Factor Analytic Methods," *Journal of Marketing Research*, 17 (May 1980): 228-236.

### Determinar o número de fatores

As componentes principais a calcular são tantas quanto o número de variáveis, mas com isso nada se ganha na redução de variáveis. Para resumir as informações contidas nas variáveis originais, deve-se extrair um número menor de fatores. A questão é: quantos? Para determinar o número de fatores, foram sugeridos vários procedimentos. São eles: determinação *a priori* e abordagens baseadas em autovalores, gráfico de declive (*scree plot*), percentagem de variância em jogo, confiabilidade meio a meio e testes de significância.

**Determinação *a priori*.** Às vezes, em virtude de conhecimento prévio, o pesquisador sabe quantos fatores pode esperar, o que permite especificar o número de fatores a serem extraídos de antemão. A extração cessa quando se atinge o número desejado de fatores. A maioria dos programas de computador permite que o usuário especifique o número de fatores, possibilitando-lhe uma implementação fácil dessa abordagem.

**Determinação com base em autovalores.** Nessa abordagem, são retidos apenas os fatores com autovalores superiores a 1,0; os outros fatores não são incluídos no modelo. Um autovalor representa a quantidade da variância associada ao fator. Logo, só se incluem fatores com variância maior do que 1,0. Fatores com variância inferior a 1,0 não são melhores que uma variável isolada porque, devido à padronização, cada variável tem variância de 1,0. Se o número de variáveis for inferior a 20, essa abordagem resulta em um número conservador de fatores.

**Determinação com base em um gráfico de declive.** Um gráfico de declive (*scree plot*) é uma representação gráfica dos autovalores *versus* número de fatores pela ordem de extração. A forma do gráfico é usada para determinar o número de fatores. Tipicamente o gráfico apresenta uma acentuada interrupção entre o acentuado declive dos fatores com grandes autovalores e uma gradual redução relacionada com o restante dos fatores. Essa redução gradual é conhecida como declive. Evidências experimentais indicam que o ponto no qual começa o declive denota o verdadeiro número de fatores. Em geral, o número de fatores determinado por um gráfico de declive será superior em mais um ou mais alguns ao daquele determinado pelo critério dos autovalores.

**Determinação com base na percentagem da variância.** Nesta abordagem, determina-se o número de fatores extraídos de forma que a percentagem acumulada da variância extraída pelos fatores atinja um nível satisfatório. Recomenda-se, todavia, que os fatores extraídos respondam por, no mínimo, 60% da variância.

**Determinação com base em confiabilidade meio a meio.** A amostra é dividida ao meio, fazendo-se uma análise fatorial sobre cada metade. São retidos apenas os fatores com elevada correspondência de cargas fatoriais ao longo das duas subamostras.

**Determinação com base em testes de significância.** É possível determinar a significância estatística dos autovalores separados, retendo apenas os fatores que são estatisticamente significativos. Um empecilho é que, com grandes amostras (tamanho superior a 200), muitos fatores tendem a ser estatisticamente significativos, embora, do ponto de vista prático, muitos deles respondam apenas por uma pequena proporção da variância total.

Na Tabela 19.3, vemos que o autovalor maior do que 1,0 (opção padrão) resulta na extração de dois fatores. Nosso conheci-

mento *a priori* nos diz que o dentifrício é comprado por duas razões principais. A Figura 19.2 apresenta o gráfico de declive associado a essa análise; nele vemos por que ocorre uma quebra distinta em três fatores. Finalmente, pela percentagem acumulada da variância, vemos que os dois primeiros fatores respondem por 82,49% da variância e que o ganho obtido ao passarmos para três fatores é marginal. Além disso, a confiabilidade meio a meio indica também que dois fatores são suficientes. Assim, dois fatores se afiguram razoáveis nesta situação.

A segunda coluna em "Comunalidades" na Tabela 19.3 dá informações relevantes após extraído o número desejado de fatores. As comunalidades para as variáveis em "Extração" são diferentes das que aparecem sob "Inicial", porque nem todas as variâncias associadas às variáveis são explicadas, a menos que se retenham todos os fatores. A extração da soma do quadrado das cargas dão as variâncias associadas aos fatores retidos. Observe que estes são os mesmos que aparecem em "Autovalores Iniciais". Isso sempre ocorre no caso de análise de componentes principais. Determina-se a percentagem de variância de um fator dividindo o autovalor associado pelo número total de fatores (ou variáveis) e multiplicando o resultado por 100. Assim, o primeiro fator responde por  $(2,731/6) \times 100$ , ou 45,52% da variância das seis variáveis. Da mesma forma, o segundo fator responde por  $(2,218/6) \times 100$ , ou 36,96% da variância. A interpretação da solução em geral é melhorada por uma rotação dos fatores.

### Rotacionar os fatores

Um produto importante da análise fatorial é a matriz de fatores, também chamada de *matriz de fatores padronizados*. A matriz de fatores contém os coeficientes utilizados para expressar as variáveis padronizadas em termos dos fatores. Esses coeficientes, as cargas fatoriais, representam as correlações entre os fatores e as variáveis. Um coeficiente com valor absoluto grande indica que o fator e a variável estão estreitamente relacionados. Podemos utilizar os coeficientes da matriz de fatores para interpretar os fatores.

Embora a matriz de fatores inicial (não-rotacionada) indique a relação entre os fatores e as variáveis individuais, ela raramente

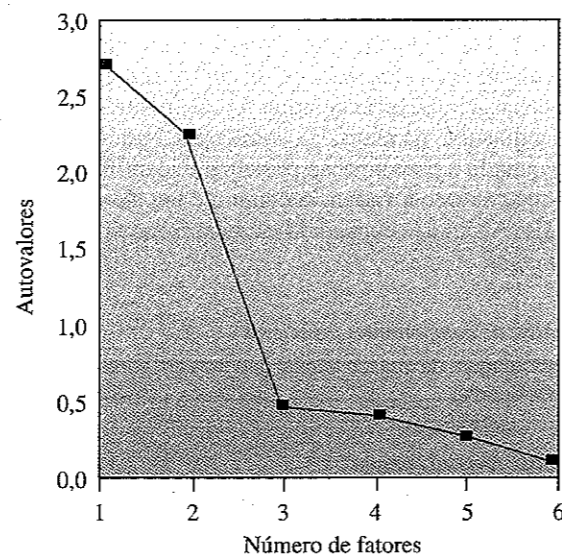


Figura 19.2 Gráfico de declive.

resulta em fatores que possam ser interpretados, porque os fatores são correlacionados com muitas variáveis. Por exemplo, na Tabela 19.3, o fator 1 tem alguma correlação com cinco das seis variáveis (valor absoluto da carga fatorial superior a 0,3). Como devemos interpretar esse fator? Em uma matriz complexa como essa, é difícil interpretar os fatores. Assim, por meio de uma rotação, a matriz de fatores é transformada em uma matriz mais simples, mais fácil de interpretar.

Ao rotacionar os fatores, seria interessante que cada fator tivesse cargas ou coeficientes não-zero para apenas algumas das variáveis. Seria igualmente interessante que cada variável tivesse carregamentos não-zero ou significativos com uns poucos fatores apenas – se possível, com apenas um. Se vários fatores tiverem altas cargas com a mesma variável, torna-se difícil interpretá-los. A rotação não afeta as comunalidades e a percentagem explicada da variância total. Entretanto, a percentagem da variância explicada por cada fator varia, conforme se pode ver na Tabela 19.3. A variância explicada pelos fatores individuais é redistribuída por rotação. Logo, diferentes métodos de rotação podem resultar na identificação de diferentes fatores. A rotação é chamada de *ortogonal* se os eixos são mantidos em ângulo reto. O método de rotação mais comumente utilizado é o *procedimento varimax*. Trata-se de um método ortogonal de rotação que minimiza o número de variáveis, com altas cargas sobre um fator, reforçando, assim, a interpretabilidade dos fatores.<sup>8</sup> A rotação ortogonal tem como resultado fatores não-correlacionados. A rotação se diz *oblíqua* quando os eixos não se mantêm em ângulo reto e os fatos são correlacionados. Por vezes, o fato de admitirmos correlações entre fatores pode simplificar a matriz de fatores padronizados. Deve-se utilizar a rotação oblíqua quando os fatores na população tendem a ser fortemente correlacionados.

#### rotação ortogonal

Rotação de fatores em que os eixos são mantidos em ângulo reto.

#### procedimento varimax

Método ortogonal de rotação de fatores que minimiza o número de variáveis com altas cargas sobre um fator, reforçando, assim, a interpretabilidade dos fatores.

#### rotação oblíqua

Rotação de fatores quando os eixos não são mantidos em ângulo reto.

Na Tabela 19.3, comparando a matriz de fatores rotacionados varimax com a matriz não-rotacionada (chamada matriz de fatores), podemos ver como a rotação acarreta simplicidade e reforça a interpretabilidade. Enquanto cinco variáveis se correlacionam com o fator 1 na matriz não-rotacionada, apenas as variáveis  $V_1$ ,  $V_3$  e  $V_5$  se correlacionam com o fator 1 após a rotação. As restantes variáveis  $V_2$ ,  $V_4$  e  $V_6$  se correlacionam fortemente com o fator 2. Além disso, nenhuma variável apresenta

forte correlação com ambos os fatores. A matriz de fatores rotacionados constitui a base para a interpretação dos fatores.

### Interpretar os fatores

A interpretação é facilitada pela identificação das variáveis que apresentam grandes cargas sobre o mesmo fator. O fator pode então ser interpretado em termos das variáveis que o oneram fortemente. Outro recurso que ajuda na interpretação é o gráfico das variáveis, utilizando as cargas fatoriais como coordenadas. As variáveis no final de um eixo são as que têm altas cargas somente sobre aquele fator e, por conseguinte, o descrevem. As variáveis próximas da origem têm pequenas cargas sobre os fatores. As variáveis que não estão próximas de nenhum dos eixos estão relacionadas a ambos os fatores. Se um fator não pode ser definido claramente em termos das variáveis originais, deve ser considerado como indefinido, ou geral.

Na matriz de fatores rotacionados da Tabela 19.3, o fator 1 tem altos coeficientes para as variáveis  $V_1$  (prevenção de cáries),  $V_3$  (gengivas saudáveis) e um coeficiente negativo para  $V_5$  (evitar cáries não é importante). Por conseguinte, esse fator pode ser rotulado como um fator de benefício para a saúde. Observe que um coeficiente negativo para uma variável negativa ( $V_5$ ) conduz a uma interpretação positiva de que dentes saudáveis são importantes. O fator 2 relaciona-se fortemente com as variáveis  $V_2$  (dentes claros),  $V_4$  (hálito puro) e  $V_6$  (boa aparência dos dentes). Por isso, o fator 2 pode ser rotulado como fator de benefício social. Um gráfico de cargas fatoriais, apresentado na Figura 19.3, confirma essa interpretação. As variáveis  $V_1$ ,  $V_3$  e  $V_5$  (denotadas por 1, 3 e 5, respectivamente) ficam no fim do eixo horizontal (fator 1), com  $V_5$  na extremidade oposta de  $V_1$  e  $V_3$ , enquanto as variáveis  $V_2$ ,  $V_4$  e  $V_6$  (denotadas por 2, 4 e 6) estão na extremidade do eixo vertical (fator 2). Seria possível resumir os dados com a afirmação de que os consumidores tendem a buscar dois grandes conjuntos de proveitos de um creme dental: benefícios de saúde e benefícios sociais.

### Calcular os escores fatoriais

Depois da interpretação, podem-se calcular os escores fatoriais, se necessário. A análise fatorial tem valor por si só. Todavia, se

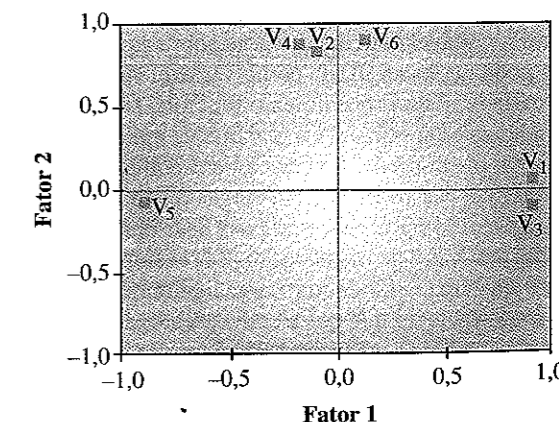


Figura 19.3 Gráfico de cargas fatoriais.

<sup>8</sup> Outros métodos de rotação ortogonal também estão disponíveis. O método quartimax minimiza o número de fatores necessários para explicar uma variável. O método equamax é uma combinação do varimax e quartimax.

o objetivo da análise fatorial for reduzir o conjunto de variáveis originais a um conjunto menor de variáveis compostas (fatores) para uso em uma análise multivariada subsequente, é conveniente calcular escores fatoriais para cada entrevistado. Um fator nada mais é do que uma combinação linear das variáveis originais. O *escore fatorial* para o *i*-ésimo fator pode ser estimado como segue:

$$F_i = W_{i1} X_1 + W_{i2} X_2 + W_{i3} X_3 + \dots + W_{ik} X_k$$

Esses símbolos já foram definidos anteriormente neste capítulo.

#### escores fatoriais

Escores compostos estimados para cada entrevistado nos fatores derivados.

Os pesos, ou coeficientes de escores fatoriais, usados para combinar as variáveis padronizadas podem ser obtidos da matriz dos coeficientes de escores fatoriais. A maioria dos programas de computador permite requisitar escores fatoriais. Somente no caso da análise dos componentes principais é que podemos calcular exatamente os escores fatoriais. Além disso, nessa análise, esses escores não são correlacionados. Na análise fatorial comum, obtêm-se estimativas desses escores, mas não há garantia de que os fatores sejam não-correlacionados uns com os outros. Os escores fatoriais podem ser usados em lugar das variáveis originais na análise multivariada subsequente. Por exemplo, com a matriz dos coeficientes de escores fatoriais da Tabela 19.3, podemos calcular dois escores fatoriais para cada entrevistado. Os valores da variável padronizada seriam multiplicados pelos coeficientes do escore fatorial correspondente, gerando os escores fatoriais.

#### Selecionar as variáveis substitutas

Às vezes, em lugar de calcular escores fatoriais, o pesquisador deseja selecionar *variáveis substitutas*. A escolha de variáveis substitutas envolve a seleção de algumas das variáveis originais para serem usadas na análise subsequente. Isso possibilita que o pesquisador faça a análise subsequente e interprete os resultados em termos das variáveis originais, e não em escores fatoriais. Examinando a matriz de fatores, podemos escolher para cada fator a variável com maior carga sobre aquele fator. Essa variável pode então ser usada como variável substituta para o fator associado. Esse procedimento funciona bem se uma carga fatorial para uma variável for claramente maior do que todas as outras cargas fatoriais. Entretanto, a escolha não é tão fácil se duas ou mais variáveis tiverem cargas igualmente altas. Em tal caso, a escolha entre essas variáveis deve basear-se em considerações teóricas e de mensuração. Por exemplo, a teoria pode sugerir que uma variável com carga ligeiramente inferior seja mais importante do que outra com carga ligeiramente superior. Da mesma forma, se uma variável tiver carga ligeiramente inferior, mas tiver sido medida com maior precisão, ela é que deve ser escolhida como variável substituta. Na Tabela 19.3, as variáveis  $V_1$ ,  $V_3$  e  $V_5$  têm elevadas cargas sobre o fator 1, e to-

das estão razoavelmente próximas em magnitude, embora  $V_1$  tenha relativamente a mais alta carga, devendo, por conseguinte, ser uma candidata provável. Todavia, se o conhecimento anterior sugerir a denteção sadia como um benefício muito importante, deve-se escolher  $V_5$  como substituta para o fator 1. Por outro lado, a escolha de uma substituta para o fator 2 não é tão direta. As variáveis  $V_2$ ,  $V_4$  e  $V_6$  têm cargas elevadas comparáveis sobre esse fator. Se o conhecimento anterior sugerir a boa aparência dos dentes como o benefício social mais importante que se pode esperar de um creme dental, o pesquisador deve selecionar  $V_6$ .

#### Determinar o ajuste do modelo

O passo final na análise fatorial consiste em determinar o ajuste do modelo. Uma suposição básica fundamental da análise fatorial é que a correlação observada entre as variáveis pode ser atribuída a fatores comuns. Logo, as correlações entre as variáveis podem ser deduzidas ou reproduzidas das correlações estimadas entre as variáveis e os fatores. Pode-se examinar as diferenças entre as correlações observadas (dadas na matriz de correlação de entrada) e as correlações reproduzidas (estimadas com base na matriz de fatores) a fim de se determinar o ajuste do modelo. Essas diferenças são chamadas de *resíduos*. Se houver muitos resíduos grandes, o modelo fatorial não proporcionará um bom ajuste aos dados, e deve ser reconsiderado. Na Tabela 19.3, vemos que apenas cinco resíduos são maiores do que 0,05, o que indica um ajuste aceitável do modelo.

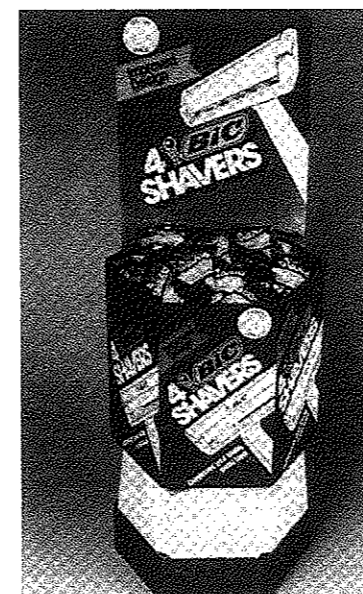
#### PESQUISA REAL

##### Componentes de uma promoção do fabricante

O objetivo deste estudo foi desenvolver um inventário abrangente das variáveis de promoção controladas pelo fabricante e demonstrar que existe uma associação entre essas variáveis e a decisão do varejista de apoiar a promoção. O apoio do varejista ou do comércio foi definido operacionalmente como a atitude do comprador para com a promoção.

Fez-se uma análise fatorial sobre as variáveis explanatórias, visando principalmente à redução de dados. O método dos componentes principais, utilizando a rotação varimax, reduziu as 30 variáveis explanatórias a 8 fatores com autovalores superiores a 1,0. Para fins de interpretação, cada fator compreende variáveis com carga de 0,40 ou mais sobre aquele fator. Em duas instâncias, em que as variáveis apresentaram carga de 0,40 ou mais sobre dois fatores, cada variável foi associada ao fator onde tinha maior carga. Apenas uma variável, "facilidade de manuseio/estocagem no varejo", não apresentou carga mínima de 0,40 sobre qualquer fator. Ao todo, os 8 fatores explicam 62% da variância total. A interpretação da matriz de cargas fatoriais foi direta. A Tabela 1 relaciona os fatores na ordem em que foram extraídos.

Foi feita uma análise discriminante passo a passo para determinar quais dos oito fatores (ou nenhum) previram o apoio do comércio com um grau estatisticamente significativo. Os escores para os oito fatores foram as variáveis independentes. A variável dependente consistiu na classificação global da oferta atribuída pelo comprador, que se reduziu a uma medida do apoio do co-



A análise dos componentes principais revela que o item importância é o fator de maior relevância dentre os que influenciam a decisão do varejista de apoiar a promoção.

mércio em três grupos (baixo, médio e alto). A Tabela 2 dá os resultados da análise discriminante. Todos os oito fatores entraram nas funções discriminantes. As medidas de aderência indicam que, como um grupo, os oito fatores discriminaram entre apoio do comércio alto, médio e baixo. As razões  $F$  multivariadas, indicando o grau de discriminação entre cada par de grupos, se revelaram significativas ao nível  $p < 0,001$ . Obteve-se a classificação correta nas categorias alta, média e baixa para 65% dos casos. Utilizou-se a ordem de entrada na análise discriminante para determinar a importância relativa de fatores como influenciadores do apoio do comércio, conforme mostra a Tabela 3.<sup>9</sup>

A análise dos componentes principais revela que o item importância é o fator de maior relevância dentre os que influenciam a decisão do varejista de apoiar a promoção.

<sup>9</sup> Jorge M. Silva-Risso, Randolph E. Bucklin e Donald G. Morrison, "A Decision Support System for Planning Manufacturers' Sales Promotion Calendars," *Marketing Science*, 18 (3) (1999): 274; Ronald C. Curhan e Robert J. Kopp, "Obtaining Retailer Support for Trade Deals: Key Success Factors," *Journal of Advertising Research* (December 1987-January 1988): 51-60.

TABELA 1 – FATORES QUE INFLUENCIAM O APOIO PROMOCIONAL DO COMÉRCIO

Fator	Interpretação do fator (% da variância explicada)	Carga	Variáveis incluídas no fator
$F_1$	Importância do item ou artigo (16,3%)	0,77	O item é suficientemente significativo para justificar a promoção
		0,75	A categoria responde bem à promoção
		0,66	É provável que o concorrente comercial mais próximo faça uma promoção do artigo
		0,64	Importância da categoria do produto promovido
		0,59	Volume de vendas regulares do artigo
$F_2$	Elasticidade da promoção (9,3%)	0,57	A oferta se mistura com os requisitos promocionais do comércio
		0,86	Estimativa do comprador quanto ao aumento de vendas com base em:
		0,82	Redução de preço e mostruário
		0,80	Somente mostruário
		0,70	Somente redução de preço
$F_3$	Apoio do fabricante à marca (8,2%)	0,85	Redução de preço, mostruário e propaganda
		0,81	Apoio do fabricante à marca, sob a forma de:
		0,80	Cupons
		0,75	Propaganda no rádio e na televisão
		0,72	Anúncio em jornal
$F_4$	Reputação do fabricante (7,3%)	0,72	Promoção no ato da compra (p.ex., mostruário)
		0,72	Reputação geral do fabricante
		0,64	O fabricante coopera no atendimento das necessidades promocionais do comércio
		0,64	O fabricante coopera em pedidos de emergência, etc.
		0,55	Qualidade da apresentação de vendas
$F_5$	Desgaste da promoção (6,4%)	0,51	Qualidade global do produto
		0,93	A categoria do produto é promovida em excesso
		0,93	O artigo é excessivamente promovido

(continua)

TABELA 1 – FATORES QUE INFLUENCIAM O APOIO PROMOCIONAL DO COMÉRCIO

Fator	Interpretação do fator (% da variância explicada)	Carga	Variáveis incluídas no fator
F <sub>6</sub>	Velocidade das vendas (5,4%)	-0,81	Fatia de participação da marca no mercado <sup>a</sup>
		0,69	Volume regular de vendas do artigo <sup>a</sup>
		0,46	Volume regular de vendas do artigo
F <sub>7</sub>	Lucratividade do item (4,5%)	0,79	Margem bruta regular do artigo
		0,72	Margem bruta regular do artigo <sup>a</sup>
		0,49	As exigências quanto ao desempenho do negócio são razoáveis
F <sub>8</sub>	Montante de incentivos (4,2%)	0,83	Montante absoluto de abatimentos
		0,81	Abatimentos como percentagem do custo comercial regular <sup>a</sup>
		0,49	Montante absoluto de abatimentos <sup>a</sup>

<sup>a</sup> Denota medida objetiva.

TABELA 2 – RESULTADOS DA ANÁLISE DISCRIMINANTE: ANÁLISE DA CLASSIFICAÇÃO (N = 564)

Fator	Coeficientes discriminantes padronizados Análise da classificação		
	Função 1	Função 2	
F <sub>1</sub>	Importância do item	0,861	-0,253
F <sub>2</sub>	Elasticidade da promoção	0,081	0,398
F <sub>3</sub>	Apoio do fabricante à marca	0,127	-0,036
F <sub>4</sub>	Reputação do fabricante	0,394	0,014
F <sub>5</sub>	Desgaste da promoção	-0,207	0,380
F <sub>6</sub>	Velocidade das vendas	0,033	-0,665
F <sub>7</sub>	Lucratividade do item	0,614	0,357
F <sub>8</sub>	Montante de incentivos	0,461	0,254
	λ de Wilks para cada fator	Todos significativos ao nível $p < 0,001$	
	Razão F multivariada	Todos significativos ao nível $p < 0,001$	
	Percentagem de casos classificados corretamente	65% corretos	

TABELA 3 – IMPORTÂNCIA RELATIVA DOS INFLUENCIADORES DO APOIO DO COMÉRCIO (CONFORME INDICADO PELA ORDEM DE ENTRADA NA ANÁLISE DISCRIMINANTE)

Análise da classificação	
Ordem de entrada	Nome do fator
1	Importância do item
2	Lucratividade do item
3	Montante de incentivos
4	Reputação do fabricante
5	Desgaste da promoção
6	Velocidade das vendas
7	Elasticidade da promoção
8	Apoio do fabricante à marca ■

### APLICAÇÕES DA ANÁLISE FATORIAL COMUM

Os dados da Tabela 19.1 foram analisados utilizando o modelo de análise fatorial comum. Em lugar de utilizar unidades na diagonal, inseriram-se as comunalidades. A saída apresentada na Tabela 19.4 é similar à saída da análise de componentes principais constante da Tabela 19.3. Em "Comunalidades", na

coluna "Inicial", as comunalidades para as variáveis não são mais 1,0. Com base no critério dos autovalores, extraem-se novamente dois fatores. As variâncias, após a extração dos fatores, são diferentes dos autovalores iniciais. O primeiro fator responde por 42,84% da variância, enquanto o segundo responde por 31,13% – em cada caso um pouco menos do que o que se observou na análise de componentes principais.

TABELA 19.4

### Resultado da análise fatorial comum

Teste de esfericidade de Bartlett  
Qui-quadrado aproximado = 111,314,  $gl = 15$ , significância = 0,00000  
Medida de adequação da amostra Kaiser-Meyer-Olkin = 0,660

#### COMUNALIDADES

VARIÁVEL	INICIAL	EXTRAÇÃO
V1	0,859	0,928
V2	0,480	0,562
V3	0,814	0,836
V4	0,543	0,600
V5	0,763	0,789
V6	0,587	0,723

#### AUTOVALORES INICIAIS (INITIAL EIGEN VALUES)

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,731	45,520	45,520
2	2,218	36,969	82,488
3	0,442	7,360	89,848
4	0,341	5,688	95,536
5	0,183	3,044	98,580
6	0,085	1,420	100,000

#### EXTRAÇÃO DA SOMA DO QUADRADO DAS CARGAS

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,570	42,837	42,837
2	1,868	31,126	73,964

#### MATRIZ DE FATORES

	FATOR 1	FATOR 2
V <sub>1</sub>	0,949	0,168
V <sub>2</sub>	-0,206	0,720
V <sub>3</sub>	0,914	0,038
V <sub>4</sub>	-0,246	0,734
V <sub>5</sub>	-0,850	-0,259
V <sub>6</sub>	-0,101	0,844

#### ROTAÇÃO DA SOMA DO QUADRADO DAS CARGAS

FATOR	AUTOVALOR (EIGEN VALUE)	PERCENTAGEM DE VARIÂNCIA	PERCENTAGEM ACUMULADA
1	2,541	42,343	42,343
2	1,897	31,621	73,964

#### MATRIZ DE FATORES ROTACIONADOS

	FATOR 1	FATOR 2
V <sub>1</sub>	0,963	-0,030
V <sub>2</sub>	-0,054	0,747
V <sub>3</sub>	0,902	-0,150
V <sub>4</sub>	-0,090	0,769
V <sub>5</sub>	-0,885	-0,079
V <sub>6</sub>	0,075	0,847

(continua)

TABELA 19.4

## Resultado da análise fatorial comum

## MATRIZ DOS COEFICIENTES DE ESCORES FATORIAIS

	FATOR 1	FATOR 2
V <sub>1</sub>	0,628	0,101
V <sub>2</sub>	-0,024	0,253
V <sub>3</sub>	0,217	-0,169
V <sub>4</sub>	-0,023	0,271
V <sub>5</sub>	-0,166	-0,059
V <sub>6</sub>	0,083	0,500

## MATRIZ DE CORRELAÇÕES REPRODUZIDAS

	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>	V <sub>5</sub>	V <sub>6</sub>
V <sub>1</sub>	0,928*	0,022	-0,000	0,024	0,008	-0,042
V <sub>2</sub>	-0,075	0,562*	0,006	-0,008	0,031	-0,012
V <sub>3</sub>	0,873	-0,161	0,836*	-0,051	0,008	0,042
V <sub>4</sub>	-0,110	0,580	-0,197	0,600*	-0,025	-0,004
V <sub>5</sub>	-0,850	-0,012	-0,786	0,019	0,789*	-0,003
V <sub>6</sub>	0,046	0,629	-0,060	0,645	-0,133	0,723*

\* O triângulo inferior esquerdo contém a matriz de correlações reproduzidas; a diagonal contém as comunalidades; o triângulo superior direito, os resíduos entre as correlações observadas e as correlações reproduzidas.

Os valores na matriz não-rotacionada de fatores padronizados da Tabela 19.4 são um pouco diferentes dos constantes da Tabela 19.3, embora o padrão dos coeficientes seja similar. Por vezes, entretanto, o padrão de cargas para a análise de fatores comuns é diferente do padrão para análise das componentes principais, com algumas variáveis apresentando cargas sobre diferentes fatores. A matriz de fatores rotacionados tem o mesmo padrão que o da Tabela 19.3, levando a uma interpretação análoga dos fatores.

## PESQUISA REAL

## Percepções "comuns" de descontos

Os descontos contribuem para a conquista de novos usuários, mudança de preferência por uma marca e repetição de compras entre usuários. A Microsoft implantou um programa de desconto como forma de atrair novos usuários a seu serviço de Internet MSN. A intenção da Microsoft por trás desse plano de desconto era conquistar novos usuários de concorrentes como o serviço American Online da AOL Time Warner, que tinha 33 milhões de assinantes em 2002. Com o plano de desconto, a Microsoft ofereceu uma opção de reembolso para novos usuários que se comprometessem com dois anos de serviços MSN. O que torna o desconto tão eficaz?

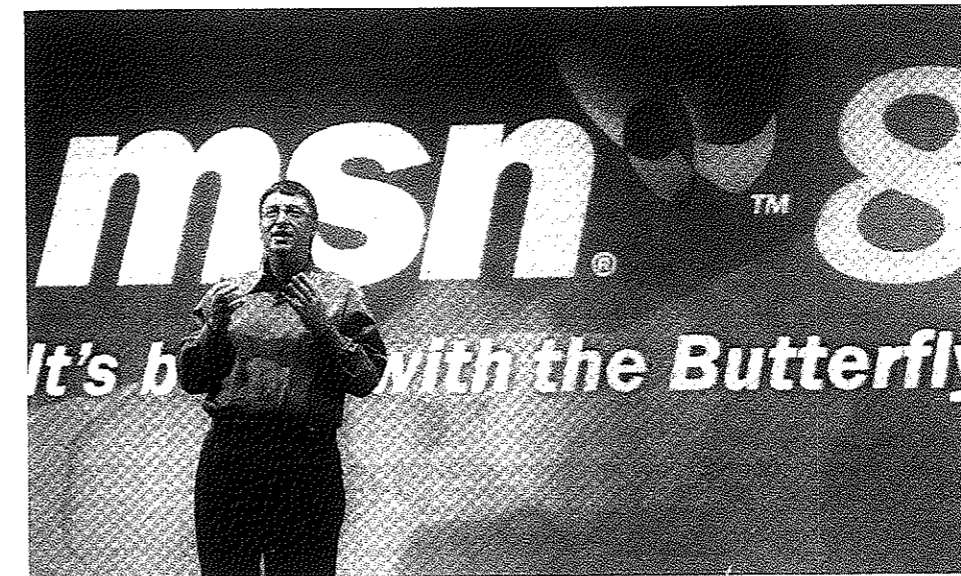
Foi feito um estudo para determinar os fatores subjacentes à percepção do consumidor em relação aos descontos. Construiu-se um conjunto de 24 itens avaliadores da percepção do consumidor em relação aos descontos. Os entrevistados deviam indicar seu grau de concordância com esses itens em uma escala Likert de cinco pontos. Os dados foram coletados em

pesquisa telefônica feita na área metropolitana de Memphis, obtendo-se um total de 303 questionários válidos.

Os 24 itens que avaliavam as percepções em relação aos descontos foram estudados por meio de uma análise fatorial comum. A solução fatorial inicial não revelou uma estrutura simples de percepções fundamentais em relação a descontos. Por conseguinte, foram excluídos itens que apresentavam cargas baixas, fazendo a análise fatorial sobre os itens restantes. Essa segunda solução resultou em três fatores interpretáveis. As cargas fatoriais são apresentadas na tabela a seguir. Os três fatores contêm quatro, quatro e três itens, respectivamente, conforme mostram as cargas fatoriais sublinhadas. O Fator 1 parece captar as percepções dos consumidores quanto aos esforços e dificuldades relacionadas com a efetivação dos descontos (esforços). O Fator 2 foi definido como a confiança dos consumidores no sistema de descontos (confiança). O Fator 3 representa as percepções dos consumidores em relação aos motivos que o fabricante tem para oferecer descontos (motivos). As cargas dos itens sobre seus respectivos fatores variaram de 0,527 a 0,744.

Portanto, empresas como a Microsoft que empregam descontos devem assegurar-se de que sejam minimizados os esforços e as dificuldades dos consumidores ao aproveitar os descontos. Elas também devem tentar conquistar a confiança do consumidor em relação ao sistema de descontos e apresentar motivos honestos para oferecer tais descontos.<sup>10</sup>

<sup>10</sup> Anônimo, "Microsoft to Discontinue MSN Rebate Promotion," *The Los Angeles Times* (February 3, 2001): C1; e Peter Tat, William A. Cunningham III e Emin Babakus, "Consumer Perceptions of Rebates," *Journal of Advertising Research* (August/September 1988): 45-50.



A análise fatorial comum revela que três fatores que caracterizam a percepção dos consumidores em relação aos descontos são os esforços relativos à efetivação dos descontos, a confiança no sistema de descontos e os motivos que os fabricantes têm para oferecer tais descontos.

## ANÁLISE FATORIAL DA PERCEPÇÃO SOBRE DESCONTOS

Itens da escala <sup>a</sup>	Carga fatorial		
	Fator 1	Fator 2	Fator 3
Os fabricantes complicam demasiadamente o procedimento de desconto.	0,194	<u>0,671</u>	-0,127
Os descontos enviados por correio não valem o trabalho que acarretam.	-0,031	<u>0,612</u>	0,352
Leva muito tempo para receber o cheque do fabricante relativo ao reembolso do desconto.	0,013	<u>0,718</u>	0,051
Os fabricantes poderiam fazer mais para facilitar a utilização dos descontos.	0,205	<u>0,616</u>	0,173
Os fabricantes oferecem descontos porque os usuários os desejam. <sup>b</sup>	<u>0,660</u>	0,172	0,101
Os fabricantes de hoje têm interesse real pelo bem-estar do consumidor. <sup>b</sup>	<u>0,569</u>	0,203	0,334
O benefício do consumidor é usualmente a principal consideração nas ofertas de desconto. <sup>b</sup>	<u>0,660</u>	0,002	0,318
Em geral os fabricantes são sinceros em suas ofertas de desconto aos consumidores. <sup>b</sup>	<u>0,716</u>	0,047	-0,033
Os fabricantes oferecem descontos para levar os consumidores a adquirirem algo que, na verdade, não lhes faz falta.	0,099	0,156	<u>0,744</u>
Os fabricantes utilizam ofertas de desconto para induzir os consumidores a comprarem artigos de pouca saída.	0,090	0,027	<u>0,702</u>
As ofertas de desconto exigem que o usuário compre mais do que realmente precisa.	0,230	0,066	<u>0,527</u>
Autovalores	2,030	1,334	1,062
Porcentagem de variância explicada	27,500	12,200	9,700

<sup>a</sup>As categorias de resposta para todos os itens foram: concordo totalmente (1), concordo (2), não concordo nem discordo (3), discordo (4), discordo totalmente (5), não sei (6). As respostas "não sei" foram excluídas da análise de dados.

<sup>b</sup>Os escores desses itens foram invertidos. ■

Observe que, nesse exemplo, quando a solução do fator inicial não podia ser interpretada, itens que tinham pouca carga foram eliminados e a análise fatorial foi realizada com os itens restantes. Se o número de variáveis for grande (maior que 15), a análise dos componentes principais e a análise fatorial comum resultam em soluções similares. Entretanto, a

análise dos componentes principais é menos propensa a interpretações equivocadas e é recomendada para usuários não-experientes. O exemplo a seguir ilustra uma aplicação da análise dos componentes principais na pesquisa de marketing internacional, e o exemplo posterior apresenta uma aplicação na área da ética.



## PESQUISA REAL

### Loucos pelo fusca

Geralmente, com o passar do tempo, as necessidades e os gostos dos consumidores se alteram. As preferências dos consumidores por automóveis precisam ser rastreadas continuamente para identificar as diferentes demandas e especificações. Entretanto, existe um carro que é praticamente uma exceção – o Fusca, da Volkswagen (conhecido em inglês por seu apelido Beetle, isto é, “besouro”). Mais de 22 milhões foram construídos desde seu lançamento em 1938. Foram realizadas pesquisas em diferentes países para determinar as razões pe-

las quais as pessoas compram Fuscas. As análises dos principais componentes das variáveis que medem as razões para possuir Fuscas revelou consistentemente um fator dominante – uma fidelidade fanática. A empresa desejou por muito tempo sua morte natural, mas sem efeito algum. Esse “besouro” barulhento e apertado inspirou devoção nos motoristas. Agora, os velhos Fuscas são procurados por todas partes. “Os japoneses estão simplesmente enlouquecendo pelos Beetles”, diz Jack Finn, um reciclador de Fusquinhas antigos em West Palm Beach, na Flórida. Por causa da fidelidade leal do “besouro”, a Volkswagen relançou-o em 1998 como o New Beetle (“novo besouro”). O New Beetle provou ser muito mais

### ESCALAS PARA ANÁLISE FATORIAL DE PROBLEMAS ÉTICOS E AÇÃO DA ALTA GERÊNCIA EMPRESARIAL

	Extensão dos problemas éticos dentro da organização (Fator 1)	Ações da alta gerência quanto à ética (Fator 2)
1. Os executivos bem-sucedidos em minha empresa fazem com que seus rivais sejam mal vistos perante as pessoas importante da empresa.	0,66	
2. Colegas executivos em minha empresa com frequência adotam comportamentos que considero antiéticos.	0,68	
3. Há muitas oportunidades para que os executivos em minha empresa se envolvam em comportamentos antiéticos.	0,43	
4. Os executivos bem-sucedidos em minha empresa assumem o mérito de idéias e realizações de outros.	0,81	
5. Para se ter sucesso em minha empresa, muitas vezes é preciso comprometer nossa ética.	0,66	
6. Os executivos bem-sucedidos em minha empresa em geral se revelam mais antiéticos do que aqueles que não são tão bem-sucedidos.	0,64	
7. Os executivos bem-sucedidos em minha empresa procuram um “bode expiatório” quando acham que podem estar associados a um fracasso.	0,78	
8. Os executivos bem-sucedidos em minha empresa costumam reter informações que possam prejudicar seus próprios interesses.	0,68	
9. A direção superior de minha empresa deixou claro que não tolerará qualquer comportamento antiético.		0,73
10. Se for descoberto que um executivo em minha empresa se envolveu em algum comportamento antiético que resulte em ganho pessoal (em lugar de ganho geral), ele receberá imediatamente uma advertência.		0,80
11. Se for descoberto que um executivo em minha empresa se envolveu em algum comportamento antiético que resulte em ganho geral (em lugar de ganho pessoal), ele receberá imediatamente uma advertência.		0,78
Autovalor	5,06	1,17
Porcentagem da variância explicada	46%	11%
Coefficiente alfa	0,87	0,75

Para simplificar a tabela, foram relatadas apenas cargas rotacionadas por varimax de 0,40 ou mais. Cada uma foi avaliada em uma escala de cinco pontos, com 1 = concordo totalmente a 5 = discordo totalmente.

que uma seqüência de seu homônimo legendário. Venceu vários prêmios automotivos ilustres, incluindo o Best Pick da revista *Money* na categoria de carros pequenos em 2001. Devido ao sucesso do relançamento do Beetle, a Volkswagen decidiu revelar uma nova versão da clássica van “romântica” chamada de “Shaggin’ Wagon II” no Auto Show de Detroit. Planejava-se lançar esse novo microônibus em 2003 por uma estimativa de US\$ 30.000. Se o microônibus mostrar seu sucesso, como ocorreu com o New Beetle, a Volkswagen pode lançar até mesmo um veículo tipo minicape para agrandar a turma das 4x4 para todo terreno.<sup>11</sup>

## PESQUISA REAL

### Fatores que prevêm práticas de pesquisa de marketing antiéticas

Em 2001, o Departamento de Comércio dos EUA estimou que o comportamento antiético de funcionários custa à Corporate America cerca de US\$ 60 bilhões. Se as empresas querem funcionários éticos, elas mesmas devem amoldar-se a altos padrões éticos. Isso também se aplica ao segmento de pesquisa de marketing. Para identificar as variáveis organizacionais que são determinantes da incidência de práticas de pesquisa de marketing antiéticas, foi feita uma amostra de 420 profissionais de marketing. Pediu-se a esses profissionais que fornecessem respostas em várias escalas e avaliações de incidência de 15 práticas de pesquisa que se constatou suscitarem problemas éticos de pesquisa.

Uma dessas escalas incluía 11 itens referentes às proporções que os problemas éticos tinham tomado na organização, incluindo também que ações da alta gerência se centravam às situações éticas. Uma análise dos componentes principais com rotação varimax indicou que os dados poderiam ser representados por dois fatores.

Esses dois fatores foram usados, então, em uma regressão múltipla junto a quatro outras variáveis previsoras. Descobriu-

<sup>11</sup> Bret Begun, Susannah Meadows e Katherine Stroup, “Now Playing: ‘Dude, Where’s My Microbus?’” *Newsweek*, 137 (4) (January 22, 2001): 9; e “Return of the Beetle,” *The Economist*, 346 (8050) (January 10, 1998).

se que eram as duas melhores previsões de práticas de pesquisa de marketing antiéticas.<sup>12</sup>

## APLICAÇÕES NA INTERNET E EM COMPUTADORES

Existem programas de computador para implementar ambas as abordagens: análise dos componentes principais e análise fatorial comum. A estrutura básica e os programas de microcomputador são semelhantes para SPSS e SAS. Nos pacotes SPSS, pode-se utilizar o programa FACTOR não só para análise de componentes principais como para análise fatorial comum. Existem também outros métodos de análise fatorial, com escores fatoriais calculados.

No sistema SAS, o programa PRINCOMP faz análise de componentes principais e calcula escores desses componentes. Para fazer a análise fatorial comum, pode-se usar o programa FACTOR, que faz também análise de componentes principais.

No MINITAB, acessa-se a análise fatorial utilizando Multivariate>Factor analysis. Pode-se utilizar a análise de componentes principais ou a máxima probabilidade para determinar a extração do fator inicial. No caso de usarmos a máxima verossimilhança, é preciso especificar o número de fatores a serem extraídos. Se não for especificado um número em uma extração de componentes principais, o programa o fará igual a um número de variáveis no conjunto de dados. A análise fatorial não está incluída no EXCEL.

### SPSS Windows

Para selecionar esse procedimento usando o SPSS para Windows, clique em:

Analyze>Data Reduction>Factor...

<sup>12</sup> Erin Stout, “Are Your Salespeople Ripping You Off?” *Sales and Marketing Management*, 153 (2) (February 2001): 56-62; David J. Fritzsche, “Ethical Climates and the Ethical Dimension of Decision Making,” *Journal of Business Ethics*, 24 (2) (March 1, 2000): 125-140; e Ishmael P. Akaah e Edward A. Riordan, “The Incidence of Unethical Practices in Marketing Research: An Empirical Investigation,” *Journal of the Academy of Marketing Science*, 18(1990): 143-152.

### FOCO NA BURKE

A Burke define em uma palavra seu objetivo ao aplicar a análise fatorial: simplificar! Em uma entrevista típica, a Burke formula muitas perguntas, e grupos dessas questões estão imensamente correlacionados uns aos outros. Independentemente de a análise fatorial ser utilizada para esclarecer que perguntas tendem a medir os mesmos conceitos básicos ou para auxiliar no uso das respostas como previsoires em uma regressão, a Burke deve ser cuidadosa ao interpretar os resultados.

Um projeto recente desenvolvido pela Burke envolveu 16 questões (variáveis). A matriz de correlações foi submetida a uma análise de componentes principais. Retiveram-se cinco fatores ou componentes. A solução varimax rotacionada mostra que o fator 5 é essencialmente uma componente de “uma questão”, isto é, está ligada apenas a uma variável (omitimos as cargas inferiores a 0,40 para tornar o quadro mais claro). Observe as questões que oneram a primeira componente rotacionada.

(continua)

MATRIZ DE FATORES ROTACIONADOS<sup>a</sup>

	Fatores/Componentes				
	1	2	3	4	5
V01		-0,649			
V02	0,460			0,720	
V03					0,873
V04	0,553	0,675			
V05			0,840		
V06			0,683		
V07			0,857		
V08				0,881	
V09	0,898				
V10	0,472				
V11		0,697			
V12	0,826				
V13	0,764				
V14	0,701				
V15		0,860			
V16	0,578	0,617			

Método de extração: análise de componentes principais.

Método de rotação: varimax com normalização Kaiser.

<sup>a</sup> A rotação converge em 6 iterações.

MATRIZ DE FATORES ROTACIONADOS<sup>a</sup>

	Fatores/Componentes			
	1	2	3	4
V01		-0,651		
V02				0,813
V03				
V04	0,516	0,662		
V05			0,837	
V06			0,672	
V07			0,843	
V08				0,817
V09	0,870			
V10				0,436
V11		0,679		
V12	0,734			
V13	0,810			
V14	0,589			0,438
V15		0,860		
V16	0,550	0,604		

Método de extração: análise de componentes principais.

Método de rotação: varimax com normalização de Kaiser.

<sup>a</sup> A rotação converge em 5 iterações.

Por estar ligado a apenas uma variável, o fator 5 não ajuda a levar à melhor opção; por isso determinaremos uma solução de quatro componentes.

Observe que V10 parece agora estar relacionada com as variáveis 2, 8 e 14, embora anteriormente estivesse mais re-

lacionada com as questões na componente 1. A variável I4 parece agora menos relacionada com a componente 1 do que aparentava na primeira solução (de cinco componentes). A estrutura dos fatores 2 e 3 permanece muito estável nas duas soluções. Isso leva a várias considerações de ordem prática:

1. Nem todas as questões estarão necessariamente correlacionadas com outras questões de modo a formar uma componente. Quando isso ocorre, em geral a questão é ambígua ou então é uma questão para a qual todos parecem dar a mesma resposta, com uma variação exclusivamente aleatória entre as respostas. Leia a questão e veja se realmente é uma "boa questão".
2. As cargas das questões agrupam-se de maneira diversa quando rotacionamos um número diferente de componentes. Procure as mais estáveis para sua interpretação dos resultados. Mesmo uma elevada carga sobre uma componente não significa que se trata de uma relação que vá permanecer válida sob uma rota-

ção diferente (atente para V14 na primeira solução de cinco componentes).

3. Por último, as questões que conjuntamente apresentam carga são o resultado do tamanho relativo, e não do tamanho absoluto, das correlações entre as questões. Podemos dividir por 10 todas as correlações nesta matriz, e as mesmas questões apresentariam carga nas mesmas componentes após a rotação... apenas os carregamentos seriam menores. Não devemos supor que as questões têm forte relacionamento apenas por se concentrarem em uma componente. A intensidade da relação é indicada pela quantidade da variância na variável original que é captada pelo fator (a carga ao quadrado).

## RESUMO

A análise fatorial é um conjunto de procedimentos utilizados para reduzir e resumir dados. Cada variável se expressa como uma combinação linear dos fatores subjacentes. Da mesma forma, os próprios fatores podem ser expressos como combinações lineares das variáveis observadas. Os fatores são extraídos de tal forma que o primeiro fator responda pela mais alta variância nos dados, o segundo pela segunda variância mais alta, e assim por diante. Além disso, é possível extrair os fatores de modo que eles sejam não-correlacionados, como na análise de componentes principais.

Na formulação de um problema de análise fatorial, as variáveis a serem incluídas na análise devem ser especificadas com base em pesquisa anterior, na teoria e no julgamento do pesquisador. Essas variáveis devem ser medidas em escala intervalares ou razão. A análise fatorial se baseia em uma matriz de correlação entre as variáveis. Pode-se testar estatisticamente a conveniência da matriz de correlação para a análise fatorial.

As duas abordagens básicas da análise fatorial são a análise de componentes principais e a análise fatorial comum. Na análise de componentes principais, leva-se em conta a variância total. Recomenda-se a análise de componentes principais quando o objetivo principal do pesquisador é determinar o número mínimo de fatores que responderão pela variância máxima nos dados a serem usados em análises multivariadas subsequentes. Na análise fatorial comum, os fatores são estimados

apenas com base na variância comum. Esse método é apropriado quando a preocupação é identificar as dimensões latentes e quando há interesse na variância comum. Esse método também é conhecido como fatoração do eixo principal.

O número de fatores a serem extraídos pode ser determinado *a priori* ou com base em autovalores, gráficos de declive, percentagem de variância, confiabilidade meio a meio ou testes de significância. Embora a matriz inicial de fatores (ou não-rotacionada) indique a relação entre os fatores e variáveis individuais, raramente resulta em fatores que possam ser interpretados, porque os fatores são correlacionados com muitas variáveis. Portanto, utiliza-se a rotação para transformar a matriz de fatores em uma matriz mais simples e mais fácil de interpretar. O método de rotação mais comumente usado é o procedimento varimax, que resulta em fatores ortogonais. Se os fatores forem altamente correlacionados na população, pode-se usar a rotação oblíqua. A matriz de fatores rotacionados constitui a base para interpretação dos fatores.

Podem-se calcular escores fatoriais para cada entrevistado. Alternativamente, podem-se escolher variáveis substitutas examinando a matriz de fatores e selecionando para cada fator uma variável com a carga mais alta, ou a segunda mais alta. Podem-se examinar as diferenças entre as correlações observadas e as correlações reproduzidas, como as estimadas pela matriz de fatores, para determinar o ajuste do modelo.

## TERMOS E CONCEITOS FUNDAMENTAIS

análise fatorial, 548

técnica de interdependência, 548

fator, 548

teste de esfericidade de Bartlett, 549

matriz de correlação, 549

comunalidade, 549

autovalor, (*eigen value*) 549

cargas fatoriais, 549

gráfico das cargas fatoriais, 549

matriz de fatores, 549

escores fatoriais, 549

medida de adequação da amostra de

Kaiser-Meyer-Olkin (KMO), 549

percentagem de variância, 549

resíduos, 549

gráfico de declive, (*scree plot*) 549

análise de componentes principais, 553

análise fatorial comum, 553

rotação ortogonal, 555

procedimento varimax, 555

rotação oblíqua, 555

escores fatoriais, 556

(continua)

**EXERCÍCIOS**

**Perguntas**

1. Em que a análise fatorial difere da regressão múltipla e da análise discriminante?
2. Quais são as principais aplicações da análise fatorial?
3. Descreva o modelo de análise fatorial.
4. Que hipótese é examinada pelo teste de esfericidade de Bartlett? Com que finalidade se aplica esse teste?
5. O que significa a expressão “comunalidade de uma variável”?
6. Defina sucintamente autovalor, carga fatorial, matriz de fatores e escores fatoriais.
7. Com que objetivo é utilizada a medida de adequação da amostra de Kaiser-Meyer-Olkin?
8. Qual é a principal diferença entre análise de componentes principais e análise fatorial comum?
9. Explique como são utilizados os autovalores para determinar o número de fatores.
10. Que é um gráfico de declive? Para que é usado?
11. Por que é útil rotacionar fatores? Qual o método mais comum de rotação?
12. Quais as diretrizes existentes para a interpretação de fatores?
13. Quando é conveniente calcular escores fatoriais?

14. Que são variáveis substitutas? Como são determinadas?
15. Como se examina o ajuste do modelo de análise fatorial?

**Problemas**

1. Calcule a parte faltante dos dados abaixo referentes ao resultado de uma análise de componentes principais:

Variável	Comunalidade	Fator	Autovalor	Porcentagem de variância
V <sub>1</sub>	1,0	1	3,25	
V <sub>2</sub>	1,0	2	1,78	
V <sub>3</sub>	1,0	3	1,23	
V <sub>4</sub>	1,0	4	0,78	
V <sub>5</sub>	1,0	5	0,35	
V <sub>6</sub>	1,0	6	0,30	
V <sub>7</sub>	1,0	7	0,19	
V <sub>8</sub>	1,0	8	0,12	

2. Trace um gráfico de declive baseado nos dados do problema 1.
3. Quantos fatores devem ser extraídos no problema 1? Explique seu raciocínio.

**EXERCÍCIOS PARA INTERNET E COMPUTADORES**

1. Em um estudo sobre a relação entre o comportamento doméstico e o comportamento de compras, obtiveram-se dados das seguintes afirmações referentes a modos de vida (com base em uma escala de sete pontos: 1 = discordo, 7 = concordo):  
 V<sub>1</sub> Prefiro passar uma noite tranqüila em casa a sair para uma festa.  
 V<sub>2</sub> Sempre verifico os preços, mesmo em artigos baratos.  
 V<sub>3</sub> As lojas são mais interessantes do que os filmes.  
 V<sub>4</sub> Não compro produtos anunciados em cartazes.  
 V<sub>5</sub> Sou uma pessoa caseira.  
 V<sub>6</sub> Economizo e troco cupons.  
 V<sub>7</sub> As companhias gastam demais com propaganda.  
 A tabela ao lado apresenta os dados obtidos de uma amostra pré-teste de 25 entrevistados:  
 a. Analise esses dados por meio da análise de componentes principais usando procedimento de rotação varimax.  
 b. Interprete os fatores extraídos.  
 c. Calcule os escores fatoriais para cada entrevistado.  
 d. Se fosse necessário escolher variáveis substitutas, quais delas você selecionaria?  
 e. Examine o ajuste do modelo.  
 f. Analise os dados utilizando a análise fatorial comum e responda as questões de b a e.
2. Realize a seguinte análise sobre os dados da Nike apresentados no exercício para Internet e computadores 1 do Capítulo 15. Analise apenas as seguintes variáveis: consciência, atitude, preferência, intenção e lealdade à Nike.

Nº	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>	V <sub>5</sub>	V <sub>6</sub>	V <sub>7</sub>
1	6	2	7	6	5	3	5
2	5	7	5	6	6	6	4
3	5	3	4	5	6	6	7
4	3	2	2	5	1	3	2
5	4	2	3	2	2	1	3
6	2	6	2	4	3	7	5
7	1	3	3	6	2	5	7
8	3	5	1	4	2	5	6
9	7	3	6	3	5	2	4
10	6	3	3	4	4	6	5
11	6	6	2	6	4	4	7
12	3	2	2	7	6	1	6
13	5	7	6	2	2	6	1
14	6	3	5	5	7	2	3
15	3	2	4	3	2	6	5
16	2	7	5	1	4	5	2
17	3	2	2	7	2	4	6
18	6	4	5	4	7	3	3
19	7	2	6	2	5	2	1
20	5	6	6	3	4	5	3
21	2	3	3	2	1	2	6
22	3	4	2	1	4	3	6
23	2	6	3	2	1	5	3
24	6	5	7	4	5	7	2
25	7	6	5	4	6	5	3

- a. Analise esses dados por meio da análise de componentes principais usando o procedimento de rotação varimax.
  - b. Interprete os fatores extraídos.
  - c. Calcule os escores fatoriais para cada entrevistado.
  - d. Se fosse necessário escolher variáveis substitutas, quais delas você selecionaria?
  - e. Examine o ajuste do modelo.
  - f. Analise os dados utilizando a análise fatorial comum e responda as questões de b a e.
3. Realize a seguinte análise sobre os dados de estilo de vida fora de casa apresentados no exercício para Internet e computadores 2 do Capítulo 15. Analise apenas as seguintes variáveis: a importância conferida a desfrutar da natureza, relacionar-se com o clima, viver em harmonia com o meio ambiente, exercitar-se regularmente e encontrar-se com outras pessoas (de V<sub>2</sub> a V<sub>6</sub>).  
 a. Analise esses dados por meio da análise de componentes principais usando o procedimento de rotação varimax.  
 b. Interprete os fatores extraídos.  
 c. Calcule os escores fatoriais para cada entrevistado.  
 d. Se fosse necessário escolher variáveis substitutas, quais delas você selecionaria?  
 e. Examine o ajuste do modelo.  
 f. Analise os dados utilizando a análise fatorial comum e responda as questões de b a e.

4. Realize a seguinte análise sobre os dados de calçados esportivos apresentados no exercício para Internet e computadores 3 do Capítulo 17. Analise apenas as seguintes variáveis: avaliação dos tênis em conforto (V<sub>2</sub>), estilo (V<sub>3</sub>) e durabilidade (V<sub>4</sub>).  
 a. Analise esses dados por meio da análise de componentes principais usando o procedimento de rotação varimax.  
 b. Interprete os fatores extraídos.  
 c. Calcule os escores fatoriais para cada entrevistado.  
 d. Se fosse necessário escolher variáveis substitutas, quais delas você selecionaria?  
 e. Examine o ajuste do modelo.  
 f. Analise os dados utilizando a análise fatorial comum e responda as questões de b a e.
5. Faça uma análise fatorial dos dados de estilo de vida e psicográficos a respeito de vestimentas coletados no exercício de trabalho de campo 1 usando a análise dos componentes principais. Use os programas para main-frame e microcomputador do SPSS, SAS ou MINITAB.
6. Faça uma análise fatorial dos dados sobre tempo de lazer coletados no exercício de trabalho de campo 2 usando a análise fatorial comum. Use SPSS, SAS ou MINITAB.

**ATIVIDADES**

**Trabalho de campo**

1. Você é um analista de pesquisa de marketing que trabalha para uma fabricante de roupas casuais. Foi-lhe pedido que desenvolvesse um conjunto de 10 declarações para medir as características psicográficas e o estilo de vida de alunos, porque podem estar relacionados com o uso de roupas casuais. Será pedido aos entrevistados que indiquem sua concordância ou discordância com as afirmações usando uma escala de sete pontos (1 = discordo totalmente, 7 = concordo totalmente). Obtenha dados de 40 alunos de seu campus.
2. Você foi encarregado por uma fabricante de produtos esportivos de determinar as atitudes de alunos em relação a seu comportamento de lazer. Construa uma escala de oito itens para esse propósito. Aplique essa escala a 35 alunos de seu campus.

**Discussão em grupo**

1. Em um pequeno grupo, identifique os usos da análise fatorial em cada uma das áreas de principais decisões em marketing:  
 a. Segmentação de marketing  
 b. Decisões de produtos  
 c. Decisões de promoção  
 d. Decisões de preço  
 e. Decisões de distribuição

## APÊNDICE 19.1

## Equações fundamentais da análise fatorial

No modelo de análise fatorial, deduzem-se componentes hipotéticas que respondem pela relação linear entre variáveis observadas.<sup>1</sup> O modelo de análise fatorial exige que as relações entre as variáveis observadas sejam lineares e que as variáveis tenham correlações diferentes de zero entre elas. As componentes hipotéticas deduzidas apresentam as seguintes propriedades:

1. Formam um conjunto linearmente independente de variáveis. Nenhuma componente hipotética é dedutível das outras componentes hipotéticas como combinação linear delas.
2. As variáveis das componentes hipotéticas podem ser divididas em dois tipos básicos de componentes: fatores comuns e fatores únicos. Essas duas componentes podem se distinguir em termos dos padrões de pesos nas equações lineares, que deduzem as variáveis observadas das variáveis das componentes hipotéticas. Um fator comum tem mais de uma variável com peso ou carga não-nula associada a ele. Um fator único tem apenas uma variável com peso diferente de zero associada a ela. Logo, há apenas uma variável dependente de um fator único.
3. Supõe-se sempre que os fatores comuns sejam não-correlacionados com os fatores únicos. Em geral, supõe-se também que os fatores únicos sejam mutuamente não-correlacionados, mas os fatores comuns podem ser correlacionados entre eles, ou não.
4. Em geral, admite-se que haja menos fatores comuns do que variáveis observadas. Todavia, costuma-se admitir que o número de fatores únicos seja igual ao número de variáveis observadas.

Emprega-se a notação seguinte:

$X$  = vetor aleatório  $n \times 1$  de variáveis aleatórias observadas  $X_1, X_2, X_3, \dots, X_n$

Supõe-se que

$$E(X) = 0$$

$E(XX')$  =  $R_{xx}$  uma matriz de correlação com unidades na diagonal principal.

$F$  = um vetor  $m \times 1$  de  $m$  fatores comuns  $F_1, F_2, \dots, F_m$

Supõe-se que

$$E(F) = 0$$

$E(FF')$  =  $R_{ff}$  uma matriz de correlação.

$U$  = um vetor aleatório  $n \times 1$  das variáveis de fatores únicos  $U_1, U_2, \dots, U_n$

Admite-se que

$$E(U) = 0, e$$

$$E(UU') = I$$

Os fatores únicos são normalizados a fim de terem variâncias unitárias e serem mutuamente não-correlacionados.

$A$  = uma matriz diagonal  $n \times m$  de coeficientes, chamada matriz do padrão de fatores.

$V$  = uma matriz diagonal  $n \times n$  de coeficientes para os fatores únicos

As variáveis observadas, que são as coordenadas de  $X$ , são combinações ponderadas dos fatores comuns e dos fatores únicos. A equação fundamental da análise fatorial pode escrever-se, portanto, como:

$$X = AF + VU$$

As correlações entre variáveis, em termos dos fatores, podem ser estabelecidas como segue:

$$\begin{aligned} R_{xx} &= E(XX') \\ &= E\{(AF + VU)(AF + VU)'\} \\ &= E\{(AF + VU)(F'A' + U'V')\} \\ &= E(AFF'A' + AFU'V' + VUF'A' + VUU'V') \\ &= AR_{ff}A' + AR_{ff}V' + V_{ff}A' + V^2 \end{aligned}$$

Como os fatores comuns são não-correlacionados com os fatores únicos, temos:

$$R_{fu} = R'_{uf} = 0$$

Logo,

$$R_{xx} = AR_{ff}A' + V^2$$

Suponha que subtraímos de ambos os membros a matriz de variância de fator único,  $V^2$ . Obteremos:

$$R_{xx} - V^2 = AR_{ff}A'$$

$R_{xx}$  depende apenas das variáveis de fator comum, e as correlações entre as variáveis estão relacionadas apenas com os fatores comuns. Seja  $R_c = R_{xx} - V^2$  a matriz reduzida de correlação.

Já definimos a matriz de fatores padronizados  $A_s$ . Os coeficientes dessa matriz são pesos atribuídos aos fatores comuns quando as variáveis observadas são expressas como combinações lineares dos fatores comuns e únicos. Definimos a seguir a matriz de estrutura de fatores. Os coeficientes da matriz de estrutura de fatores são as covariâncias entre as variáveis observadas e os fatores. A matriz de estrutura de fatores é útil na interpretação dos fatores, pois mostra quais variáveis são semelhantes a uma variável de fator comum. Define-se a matriz de estrutura de fatores,  $A_s$ , como

$$\begin{aligned} A_s &= E(XF') \\ &= E[(AF + VU)F'] \\ &= AR_{ff} + VR_{uf} \\ &= AR_{ff} \end{aligned}$$

variâncias entre os fatores  $R_{ff}$ . Substituindo  $A_s$  por  $R_{ff}$  a matriz reduzida de correlação se torna o produto das matrizes de estrutura de fatores e de fatores padronizados.

$$\begin{aligned} R_c &= AR_{ff}A' \\ &= A_sA' \end{aligned}$$

Assim, a matriz de estrutura de fatores é equivalente à matriz de fatores padronizados  $A$  multiplicada pela matriz de co-

<sup>1</sup> O material deste apêndice foi extraído de Stanley A. Muliak, *The Foundations of Factor Analysis* (New York: McGraw-Hill, 1972).