CAPÍTULO 3

Análise Fatorial

Objetivos de aprendizagem

Ao concluir este capítulo, você deverá ser capaz de:

- Diferenciar as técnicas de análise fatorial de outras técnicas multivariadas.
- Distinguir entre usos exploratórios e confirmatórios das técnicas analíticas fatoriais.
- Entender os sete estágios da aplicação da análise fatorial.
- Distinguir entre as análises fatoriais R e Q.
- Identificar as diferenças entre modelos de análise de componentes e análise de fatores comuns.
- Dizer como determinar o número de fatores a serem extraídos.
- Explicar o conceito de rotação de fatores.
- Descrever como nomear um fator.
- Explicar os usos adicionais de análise fatorial.
- Estabelecer as principais limitações de técnicas de análise fatorial.

Apresentação do capítulo

Durante a década passada, o uso da técnica estatística multivariada de análise fatorial aumentou em todas as áreas de pesquisa relacionadas a negócios. À medida que o número de variáveis a serem consideradas em técnicas multivariadas aumenta, há uma necessidade proporcional de maior conhecimento da estrutura e das inter-relações das variáveis. Este capítulo descreve a análise fatorial, uma técnica particularmente adequada para analisar os padrões de relações complexas multidimensionais encontradas por pesquisadores. Este capítulo define e explica em termos conceituais amplos os aspectos fundamentais das técnicas analíticas fatoriais. A análise fatorial pode ser utilizada para examinar os padrões ou relações latentes para um grande número de variáveis e determinar se a informação pode ser condensada ou resumida a um conjunto menor de fatores ou componentes. Para melhor esclarecer os conceitos metodológicos, também foram incluídas orientações básicas para apresentar e interpretar os resultados dessas técnicas.

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 *itálico*. Exemplos ilustrativos estão em quadros.

Alfa de Cronbach Medida de *confiabilidade* que varia de 0 a 1, sendo os valores de 0,60 a 0,70 considerados o limite inferior de aceitabilidade.

Análise de agrupamentos Técnica multivariada com o objetivo de agrupar respondentes ou casos com perfis similares em um dado conjunto de características. Semelhante à *análise fatorial Q*.

Análise de componentes Modelo fatorial no qual os fatores são baseados na variância total. Na análise de componentes, unidades (1s) são usadas na diagonal da *matriz de correlação*; esse procedimento implica computacionalmente que toda a variância é *comum* ou compartilhada.

Análise de fatores comuns Modelo fatorial no qual os fatores são baseados em uma matriz de correlação reduzida. Ou seja, *comunalidades* são inseridas na diagonal da matriz de *correlação* e os fatores extraídos são baseados apenas na *variância comum*, com as *variâncias específicas* e *de erro* excluídas.

Análise fatorial *Q* Forma grupos de respondentes ou casos com base em sua similaridade em um conjunto de características (ver também a discussão sobre análise de agrupamentos no Capítulo 9).

Análise fatorial *R* Analisa relações entre variáveis para identificar grupos de variáveis que formam dimensões latentes (*fatores*).

Autovalor Soma em coluna de cargas fatoriais ao quadrado para um fator; também conhecido como *raiz latente*. Representa a quantia de variância explicada por um fator.

Carga cruzada Uma variável tem duas ou mais *cargas fatoriais* excedendo o valor de referência considerado necessário para inclusão no processo de interpretação do fator.

Cargas fatoriais Correlação entre as variáveis originais e os fatores, bem como a chave para o entendimento da natureza de um fator em particular. As cargas fatoriais ao quadrado indicam qual percentual da variância em uma variável original é explicado por um fator.

Comunalidade Quantia total de variância que uma variável original compartilha com todas as outras variáveis incluídas na análise.

Confiabilidade Grau em que uma variável ou conjunto de variáveis é consistente com o que se pretende medir. Se múltiplas medidas são realizadas, as medidas confiáveis serão muito consistentes em seus valores. É diferente de *validade*, no sentido de que não se relaciona com o que deveria ser medido, mas com o modo como é medido.

Definição conceitual Especificação da base teórica para um conceito representado por um fator.

EQUIMAX Um dos métodos de rotação fatorial ortogonal que é um “meio-termo” entre as técnicas VARIMAX e QUARTIMAX, mas não é amplamente usado.

Erro de medida Imprecisões ao se medirem os “verdadeiros” valores das variáveis, devido à falibilidade do instrumento de medida (ou seja, escalas de resposta inapropriadas), aos erros na entrada de dados, ou aos erros dos respondentes.

Escalas múltiplas Método de combinação de diversas variáveis que medem o mesmo conceito em uma única variável como tentativa de aumentar a *confiabilidade* da medida. Na maioria dos casos, as variáveis separadas são somadas e então seu total ou escore médio é usado na análise.

Escore fatorial Medida composta criada para cada observação de cada fator extraído na análise fatorial. Os pesos fatoriais são usados em conjunção com os valores da variável original para calcular o escore de cada observação. O escore fatorial pode então ser usado para representar o(s) fator(es) em análises subseqüentes. Os escores fatoriais são padronizados para que tenham uma média de 0 e um desvio-padrão de 1.

Escore reverso Processo de reversão dos escores de uma variável, embora mantenha as características de distribuição, para mudar as relações (correlações) entre duas variáveis. Usado na construção de *escala múltipla* para evitar um cancelamento entre variáveis com *cargas fatoriais* positivas e negativas no mesmo fator.

Fator Combinação linear (variável estatística) das variáveis originais. Os fatores também representam as dimensões latentes (construtos) que resumem ou explicam o conjunto original de variáveis observadas.

Indeterminância fatorial Característica de *análise de fatores comuns* tal que diversos escores fatoriais diferentes podem ser calculados para um respondente, cada um se adequando ao modelo fatorial estimado. Isso significa que os escores fatoriais não são únicos para cada indivíduo.

Indicador Variável simples usada em conjunção com uma ou mais variáveis distintas para formar uma *medida composta*.

Matriz de correlação anti-imagem Matriz das correlações parciais entre variáveis após a análise fatorial, e que representa o grau em que os fatores explicam um ao outro nos resultados. A diagonal contém as *medidas de adequação da amostra* para cada variável, e os demais valores são correlações parciais entre variáveis.

Matriz de correlação Tabela que mostra as intercorrelações entre todas as variáveis.

Matriz de estrutura fatorial Uma *matriz fatorial* obtida em uma *rotação oblíqua* que representa as correlações simples entre variáveis e fatores, incorporando a variância única e as correlações entre fatores. A maioria dos pesquisadores prefere usar a *matriz de padrão fatorial* no momento da interpretação de uma solução oblíqua.

Matriz de padrão fatorial Uma de duas *matrizes fatoriais* em uma *rotação oblíqua* que é mais comparável com a matriz fatorial em uma *rotação ortogonal*.

Matriz fatorial Tabela das *cargas fatoriais* de todas as variáveis sobre cada fator.

Medida composta Ver *escala múltipla*.

Medida de adequação da amostra (MSA) Medida calculada tanto para toda a matriz de correlação quanto para cada variável individual, e que permite avaliar o quão adequada é a aplicação da análise fatorial. Valores acima de 0,50 para a matriz toda ou para uma variável individual indicam tal adequação.

Multicolinearidade Grau em que uma variável pode ser explicada pelas outras variáveis na análise.

Ortogonal Independência matemática (sem correlação) de eixos fatoriais, um em relação ao outro (ou seja, em ângulos retos ou de 90 graus).

QUARTIMAX Um tipo de método de rotação fatorial ortogonal que foca a simplificação de colunas de uma matriz fatorial. Geralmente considerada menos efetiva do que a rotação VARIMAX.

Raiz latente Ver *autovalor*.

Rotação fatorial oblíqua *Rotação fatorial* computada de modo que os fatores extraídos são correlacionados. Ao invés de restringir arbitrariamente a rotação fatorial a uma solução *ortogonal*, a rotação oblíqua identifica o grau em que cada fator está correlacionado.

Rotação fatorial ortogonal Rotação fatorial na qual os fatores são extraídos de modo que seus eixos sejam mantidos em 90 graus. Cada fator é independente, ou *ortogonal*, em relação a todos os outros. A correlação entre os fatores é determinada como 0.

Rotação fatorial Processo de manipulação ou de ajuste dos eixos fatoriais para conseguir uma solução fatorial mais simples e pragmaticamente mais significativa.

Teste de esfericidade de Bartlett Teste estatístico da significância geral de todas as correlações em uma matriz de correlação.

Traço Representa a quantia total de variância na qual a solução fatorial é baseada. O traço é igual ao número de variáveis, baseado na suposição de que a variância em cada variável é igual a 1.

Validade Grau em que uma medida ou um conjunto de medidas corretamente representa o conceito de estudo – o grau em que se está livre de qualquer erro sistemático ou não-aleatório. A validade se refere a quão bem o conceito é definido pela(s) medida(s), ao passo que *confiabilidade* se refere à consistência da(s) medida(s).

Validade de conteúdo Avaliação do grau de correspondência entre os itens selecionados para constituir uma *escala múltipla* e sua *definição conceitual.*

Validade de expressão Ver *validade de conteúdo.*

Variância comum Variância compartilhada com outras variáveis na análise fatorial.

Variância do erro Variância de uma variável devido a erros na coleta de dados ou na medida.

Variância específica Variância de cada variável, única àquela variável e que não é explicada ou associada com outras variáveis na análise fatorial.

Variância única Ver *variância específica.*

Variável dicotômica Variável métrica binária usada para representar uma única categoria de uma variável não-métrica.

Variável estatística Combinação linear de variáveis formada ao se obter pesos empíricos aplicados a um conjunto de variáveis especificadas pelo pesquisador.

Variável substituta Seleção de uma única variável com a maior *carga fatorial* para representar um fator no estágio de redução de dados, em vez de usar uma *escala múltipla* ou um *escore fatorial.*

VARIMAX Os mais populares métodos de *rotação fatorial ortogonal*, concentrando-se na simplificação das colunas em uma *matriz fatorial.* Geralmente considerado superior a outros métodos de rotação fatorial ortogonal para conseguir uma estrutura fatorial simplificada.

O QUE É ANÁLISE FATORIAL?

Análise fatorial é uma técnica de interdependência, como definido no Capítulo 1, cujo *propósito principal é definir a estrutura inerente entre as variáveis na análise*. Obviamente, variáveis têm um papel chave em qualquer análise multivariada. Se estivermos fazendo uma previsão de vendas com regressão, prevendo sucesso ou fracasso de uma nova empresa com análise discriminante, ou usando qualquer uma das demais técnicas multivariadas discutidas no Capítulo 1, devemos ter um conjunto de variáveis sobre o qual deve-se formar relações (p. ex., quais são as variáveis que melhor prevêem vendas ou sucesso/fracasso?). Como tais, as variáveis são os alicerces fundamentais das relações.

À medida que empregamos técnicas multivariadas, por sua própria natureza, o número de variáveis aumenta. Técnicas univariadas são limitadas a uma única variável, mas técnicas multivariadas podem ter dezenas, centenas ou mesmo milhares de variáveis. Mas como descrevemos e representamos todas essas variáveis? Certamente, se temos apenas umas poucas variáveis, todas elas podem ser distintas e diferentes. À medida que acrescentamos mais e mais variáveis, cada vez mais a sobreposição (ou seja, correlação) acontece entre as mesmas. Em alguns casos, como aqueles nos quais estamos usando múltiplas medidas para superar erro de medida devido à medida multivariável (ver Capítulo 1 para uma discussão mais detalhada), o pesquisador ainda se esforça para uma correlação entre as variáveis. Quando as variáveis se tornam correlacionadas, o pesquisador precisa de caminhos para gerenciar essas variáveis – agrupando variáveis altamente correlacionadas, rotulando ou nomeando os grupos, e talvez até mesmo criando uma nova medida composta que possa representar cada grupo de variáveis.

Introduzimos a análise fatorial como nossa primeira técnica multivariada porque ela pode desempenhar um papel único na aplicação de outras técnicas multivariadas. Genericamente falando, a análise fatorial fornece as ferramentas para analisar a estrutura das inter-relações (correlações) em um grande número de variáveis (p. ex., escores de teste, itens de teste, respostas a questionários) definindo conjuntos de variáveis que são fortemente inter-relacionadas, conhecidos como **fatores**. Esses grupos de variáveis (fatores), que são por definição altamente intercorrelacionadas, são considerados como representantes de dimensões dentro dos dados. Se estamos preocupados apenas com a redução do número de variáveis, então as dimensões podem orientar a criação de novas medidas compostas. Por outro lado, se temos uma base conceitual para compreender as relações entre variáveis, então as dimensões podem realmente ter significado para aquilo que elas coletivamente representam. No último caso, essas dimensões podem corresponder a conceitos que não podem ser adequadamente descritos por uma única medida (p. ex., a atmosfera de uma loja é definida por muitos componentes sensoriais que devem ser medidos separadamente mas são todos relacionados entre si). Veremos que a análise fatorial apresenta diversas maneiras de representação desses grupos de variáveis para uso em outras técnicas multivariadas.

Devemos observar neste ponto que técnicas analíticas fatoriais podem atingir seus objetivos ou de uma perspectiva exploratória, ou de uma perspectiva confirmatória. Existe um debate contínuo sobre o papel apropriado da análise fatorial. Muitos pesquisadores consideram-na apenas exploratória, útil na busca da estrutura em um conjunto de variáveis ou como um método de redução de dados. Sob essa perspectiva, as técnicas analíticas fatoriais "consideram o que os dados oferecem" e não estabelecem restrições *a priori* sobre a estimação de componentes nem sobre o número de componentes a serem extraídos. Para muitas – talvez a maioria – das aplicações, esse uso da análise fatorial é adequado. No entanto, em outras situações, o pesquisador tem idéias preconcebidas sobre a real estrutura dos dados, baseado em suporte teórico ou em pesquisas anteriores. Ele pode desejar testar hipóteses envolvendo questões sobre, por exemplo, quais variáveis deveriam ser agrupadas em um fator, ou o número exato de fatores. Nesses casos, o pesquisador espera que a análise fatorial desempenhe um papel confirmatório – ou seja, avalie o grau em que os dados satisfazem a estrutura esperada. Os métodos que discutimos neste capítulo não fornecem diretamente a estrutura necessária para testes de hipóteses formalizadas. Abordamos explicitamente a perspectiva confirmatória da análise fatorial no Capítulo 11. Neste capítulo, porém, vemos as técnicas analíticas fatoriais principalmente de um ponto de vista exploratório ou não-confirmatório.

UM EXEMPLO HIPOTÉTICO DE ANÁLISE FATORIAL

Considere que, durante uma pesquisa qualitativa, uma empresa de varejo tenha identificado 80 características diferentes de lojas de varejo e seus serviços que, segundo os consumidores, afetaram sua preferência entre lojas. O varejista quer entender como os consumidores tomam decisões, mas sente que não pode avaliar 80 características separadas ou desenvolver planos de ação para todas essas variáveis, pois elas são específicas demais. Em vez disso, o varejista gostaria de saber se os consumidores pensam em dimensões de avaliação mais gerais, ao invés de itens específicos. Por exemplo, consumidores podem considerar vendedores como uma dimensão avaliativa mais geral que é composta de muitas outras características específicas, como conhecimento, cortesia, empatia, sensibilidade, simpatia, prontidão e assim por diante.

Para identificar essas dimensões mais amplas, o varejista poderia encomendar uma pesquisa que sondasse as avaliações dos consumidores sobre cada um dos 80 itens específicos. A análise fatorial seria então usada para identificar as dimensões de avaliação latentes. Itens específicos altamente correlacionados são considerados um elemento daquela dimensão mais ampla. Essas dimensões se tornam composições de variáveis específicas, as quais, por sua vez, permitem que as dimensões sejam interpretadas e descritas. Em nosso exemplo, a análise fatorial poderia identificar dimensões, como diversidade de produtos, qualidade de produtos, preços, profissionais da loja, serviço e atmosfera da loja como as dimensões de avaliação usadas pelos respondentes. Cada uma dessas dimensões contém itens específicos que são uma faceta da dimensão avaliativa mais ampla. A partir dessas descobertas, o varejista pode então usar as dimensões (fatores) para definir áreas amplas de planejamento e ação.

Este exemplo simples de análise fatorial demonstra seu objetivo básico de agrupar variáveis altamente correlacionadas em conjuntos distintos (fatores). Em muitas situações, esses fatores podem fornecer uma grande quantidade de informação sobre as inter-relações das variáveis. Neste exemplo, a análise fatorial identificou para gerenciamento de loja um conjunto menor de conceitos para se considerar em qualquer plano de marketing estratégico ou tático, enquanto ainda fornece uma visão sobre o que constitui cada área geral (i.e., as variáveis individuais definindo cada fator).

> Um exemplo ilustrativo de uma aplicação simples da análise fatorial é mostrado na Figura 3-1, a qual representa a matriz de correlação para nove elementos da imagem de uma loja. Incluídos nesse conjunto estão medidas da oferta de produtos, pessoal, níveis de preço, e serviços e experiências internos. A questão que um pesquisador pode querer levantar é: aqueles elementos todos são separados no que se refere às suas propriedades de avaliação, ou eles se "agrupam" em algumas áreas mais gerais de avaliação? Por exemplo, será que todos os elementos dos produtos se agrupam? Onde o nível de preço se encaixa ou está separado? Como as características internas (p.ex., pessoal, serviço e atmosfera) se relacionam umas com as outras? A inspeção visual da matriz de correlação original (Figura 3-1, parte 1) não revela facilmente qualquer padrão específico. Há correlações elevadas espalhadas, mas os agrupamentos de variáveis não são óbvios. A aplicação da análise fatorial resulta no agrupamento de variáveis, como se reflete na parte 2 da Figura 3-1. Aqui alguns padrões interessantes aparecem. Primeiro, quatro variáveis relacionadas com experiências internas de clientes são colocadas juntas. Em seguida, três variáveis que descrevem a diversidade e a disponibilidade de produtos são agrupadas. Finalmente, a qualidade de produto e os níveis de preço formam outro grupo. Cada grupo representa um conjunto de variáveis altamente inter-relacionadas que pode refletir uma dimensão avaliativa mais geral. Nesse caso, poderíamos rotular os três agrupamentos de variáveis pelos nomes experiência interna, oferta de produtos e valor.

PROCESSO DE DECISÃO EM ANÁLISE FATORIAL

Centralizamos a discussão de análise fatorial sobre o paradigma da construção de modelo em seis estágios introduzido no Capítulo 1. A Figura 3-2 mostra os três primeiros estágios do tratamento estruturado para construção de modelo multivariado, e a Figura 3-4 detalha os três estágios finais, acrescidos de um estágio adicional (estágio 7), além da estimação, interpretação e validação dos modelos fatoriais, que ajuda a selecionar variáveis substitutas, computar escores fatoriais ou criar escalas múltiplas para uso em outras técnicas multivariadas. Uma discussão de cada estágio vem a seguir.

PARTE 1: MATRIZ ORIGINAL DE CORRELAÇÃO

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1 Nível de preço	1,000								
V_2 Pessoal da loja	0,427	1,000							
V_3 Política de devolução	0,302	0,771	1,000						
V_4 Disponibilidade do produto	0,470	0,497	0,427	1,000					
V_5 Qualidade do produto	0,765	0,406	0,307	0,427	1,000				
V_6 Profundidade de diversidade	0,281	0,445	0,423	0,713	0,325	1,000			
V_7 Amplidão da diversidade	0,345	0,490	0,471	0,719	0,378	0,724	1,000		
V_8 Serviço interno	0,242	0,719	0,733	0,428	0,240	0,311	0,435	1,000	
V_9 Atmosfera da loja	0,372	0,737	0,774	0,479	0,326	0,429	0,466	0,710	1,000

PARTE 2: MATRIZ DE CORRELAÇÃO DE VARIÁVEIS APÓS AGRUPAMENTO DE ACORDO COM ANÁLISE FATORIAL

	V_3	V_8	V_9	V_2	V_6	V_7	V_4	V_1	V_5
V_3 Política de retorno	1,000								
V_8 Serviço interno	0,773	1,000							
V_9 Atmosfera da loja	0,771	0,710	1,000						
V_2 Pessoal da loja	0,771	0,719	0,737	1,000					
V_6 Profundidade de diversidade	0,423	0,311	0,429	0,445	1,000				
V_7 Amplidão de diversidade	0,471	0,435	0,466	0,490	0,724	1,000			
V_4 Disponibilidade do produto	0,427	0,428	0,479	0,497	0,713	0,729	1,000		
V_1 Nível de preço	0,302	0,242	0,372	0,427	0,281	0,354	0,470	1,000	
V_5 Qualidade do produto	0,307	0,240	0,326	0,406	0,325	0,378	0,427	0,765	1,000

Nota: Áreas sombreadas representam variáveis agrupadas por análise fatorial.

FIGURA 3-1 Exemplo ilustrativo do uso de análise fatorial para identificar estrutura dentro de um grupo de variáveis.

Estágio 1: Objetivos da análise fatorial

O ponto de partida em análise fatorial, bem como em outras técnicas estatísticas, é o problema de pesquisa. O propósito geral de técnicas de análise fatorial é encontrar um modo de condensar (resumir) a informação contida em diversas variáveis originais em um conjunto menor de novas dimensões compostas ou variáveis estatísticas (fatores) com uma perda mínima de informação – ou seja, buscar e definir os construtos fundamentais ou dimensões assumidas como inerentes às variáveis originais [18,33]. Ao atingir seus objetivos, a análise fatorial é ajustada com quatro questões: especificação da unidade de análise; obtenção do resumo de dados e/ou redução dos mesmos; seleção de variáveis e uso de resultados da análise fatorial com outras técnicas multivariadas.

Especificação da unidade de análise

Até agora, definimos análise fatorial somente em termos da identificação de estrutura em um conjunto de variáveis. Análise fatorial é, na verdade, um modelo mais geral, no sentido de que ela pode identificar a estrutura de relações entre variáveis ou respondentes pelo exame ou de correlações entre as variáveis, ou de correlações entre os respondentes.

- Se o objetivo da pesquisa fosse resumir as características, a análise fatorial seria aplicada a uma **matriz de correlação** das variáveis. Esse é o tipo mais comum de análise fatorial e é chamado de **análise fatorial *R***, que analisa um conjunto de variáveis para identificar as dimensões latentes (que não são fáceis de observar).
- A análise fatorial também pode ser aplicada a uma matriz de correlação dos respondentes individuais baseada nas características dos mesmos. Chamado de **análise fatorial *Q***, este método combina ou condensa grandes números de pessoas em diferentes grupos de uma população maior. A análise fatorial *Q* não é utilizada muito freqüentemente por causa das dificuldades computacionais. Em vez disso, a maioria dos pesquisadores utiliza algum tipo de **análise de agrupamentos** (ver Capítulo 9) para agrupar respondentes individuais. Ver também Stewart [36] para outras possíveis combinações de grupos e tipos de variáveis.

Assim, o pesquisador deve primeiramente selecionar a unidade de análise para a análise fatorial: variáveis ou respondentes. Ainda que nos concentremos prioritariamente na estruturação de variáveis, a opção de empregar análise fatorial entre respondentes como uma alternativa para a análise de agrupamentos também está disponível. As implicações em termos da identificação de variáveis ou

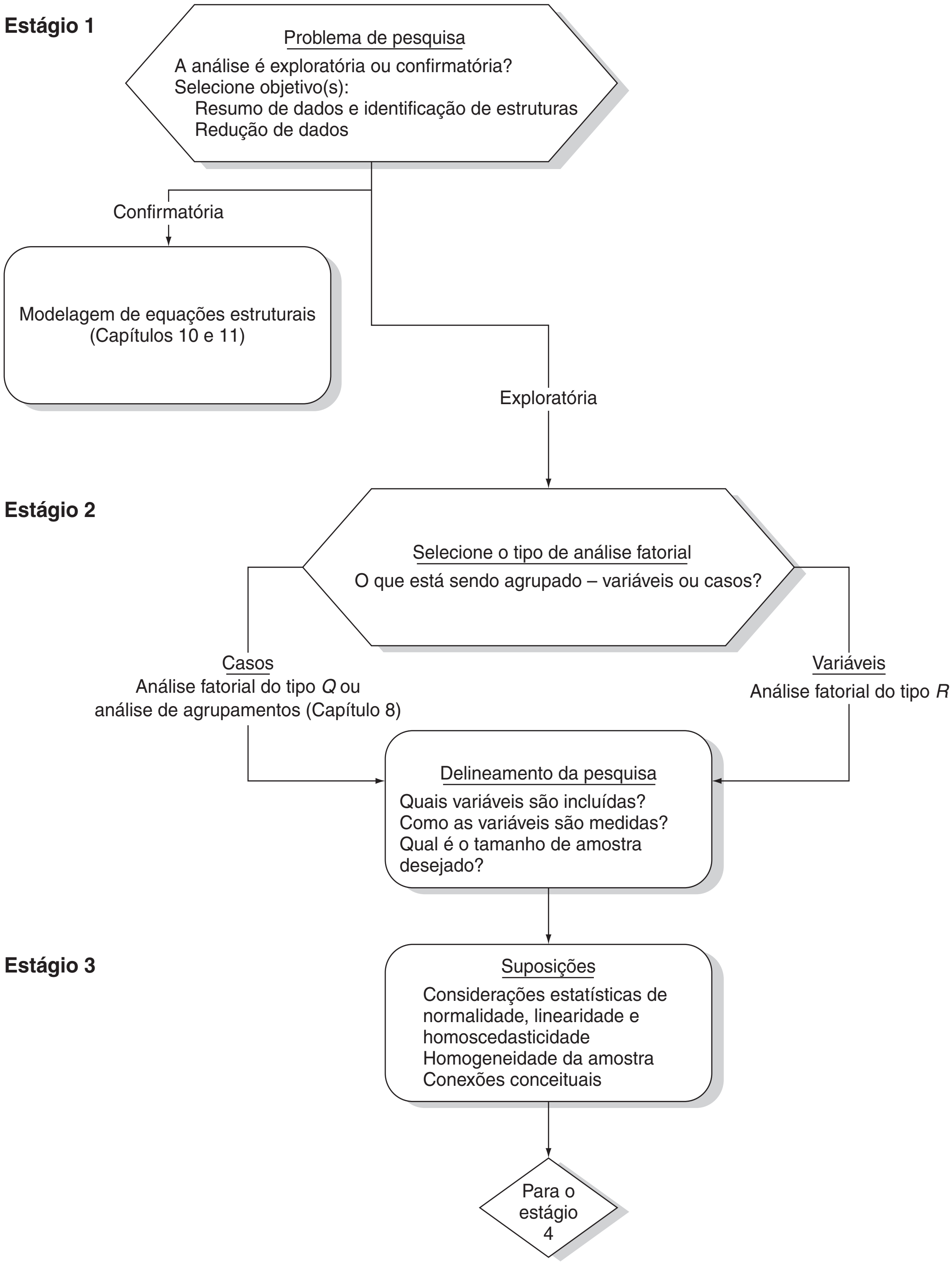

FIGURA 3-2 Estágios 1-3 no diagrama de decisão da análise fatorial.

respondentes similares serão discutidas no estágio 2 quando a matriz de correlação for definida.

Obtenção do resumo versus redução de dados

A análise fatorial fornece ao pesquisador duas saídas distintas mas relacionadas: resumo de dados e redução de dados. No resumo de dados, a análise fatorial obtém dimensões inerentes que, quando interpretadas e compreendidas, descrevem os dados em um número muito menor de conceitos do que as variáveis individuais originais. Redução de dados estende esse processo derivando um valor empírico (escore fatorial) para cada dimensão (fator) e então substituindo o valor original por esse novo valor.

Resumo de dados. O conceito fundamental envolvido no resumo de dados é a definição de estrutura. Através da estrutura, o pesquisador pode ver o conjunto de variáveis em diversos níveis de generalização, variando do nível mais detalhado (as próprias variáveis individuais) até o nível mais generalizado, onde variáveis individuais são agrupadas e então vistas não por aquilo que elas representam individualmente, mas por aquilo que representam coletivamente na expressão de um conceito.

> Por exemplo, variáveis no nível individual poderiam ser: "Compro coisas especiais", "Geralmente procuro os menores preços possíveis", "Compro produtos em promoções", "Marcas nacionais valem mais a pena do que marcas próprias". Coletivamente essas variáveis poderiam ser usadas para identificar consumidores que são "conscientes sobre preços" ou "caçadores de promoções".

A análise fatorial, enquanto técnica de interdependência, difere das técnicas de dependência discutidas na próxima seção (i.e., regressão múltipla, análise discriminante, análise multivariada de variância ou análise conjunta) onde uma ou mais variáveis são explicitamente consideradas o critério ou variáveis dependentes, e todas as outras são as variáveis preditoras ou independentes. Na análise fatorial, todas as variáveis são simultaneamente consideradas sem distinção quanto ao seu caráter de dependência ou independência. A análise fatorial ainda emprega o conceito de **variável estatística**, a composição linear de variáveis, mas em análise fatorial as variáveis estatísticas (fatores) são formadas para maximizar sua explicação do conjunto inteiro de variáveis, e não para prever uma ou mais variáveis dependentes. A meta do resumo de dados é atingida definindo-se um pequeno número de fatores que adequadamente representam o conjunto original de variáveis.

Se fizéssemos uma analogia com técnicas de dependência, seria no sentido de que cada uma das variáveis observadas (originais) é uma variável dependente, que é uma função de alguns conjuntos inerentes e latentes de fatores (dimensões), que são por sua vez compostos por todas as outras variáveis. Assim, cada variável é prevista por todos os fatores e, indiretamente, por todas as demais variáveis. Reciprocamente, pode-se olhar cada fator (variável estatística) como uma variável dependente, que é uma função do conjunto inteiro de variáveis observadas. Qualquer analogia ilustra as diferenças de meta entre técnicas de dependência (previsão) e interdependência (identificação de estrutura). Estrutura se define pelas relações entre variáveis, viabilizando a especificação de um número menor de dimensões (fatores) representando o conjunto original de variáveis.

Redução de dados. A análise fatorial também pode ser usada para conseguir redução de dados pela (1) identificação de variáveis representativas a partir de um conjunto muito maior de variáveis para uso em análises multivariadas subseqüentes, ou (2) pela criação de um conjunto inteiramente novo de variáveis, muito menor, para substituir parcial ou completamente o conjunto original de variáveis. Em ambos os casos, o propósito é manter a natureza e o caráter das variáveis originais, mas reduzir seu número para simplificar a análise multivariada a ser empregada a seguir. Ainda que as técnicas multivariadas tenham sido desenvolvidas para acomodar múltiplas variáveis, o pesquisador está sempre procurando o conjunto mais parcimonioso de variáveis para incluir na análise. Como discutido no Capítulo 1, tanto questões conceituais quanto empíricas apóiam a criação de medidas compostas. A análise fatorial fornece a base empírica para avaliar a estrutura de variáveis e o potencial para criar essas medidas compostas ou selecionar um subconjunto de variáveis representativas para análise posterior.

O resumo de dados faz da identificação das dimensões ou fatores latentes um fim em si próprio. Assim, as estimativas dos fatores e as contribuições de cada variável aos fatores (chamadas de *cargas*) são tudo o que requer a análise. A redução de dados também depende de cargas fatoriais, mas elas são usadas como a base para identificar variáveis para análises posteriores com outras técnicas ou para fazer estimativas dos próprios fatores (escores fatoriais ou escalas múltiplas), as quais substituem as variáveis originais em análises subseqüentes. O método de calcular e interpretar cargas fatoriais será apresentado posteriormente.

Seleção de variáveis

Seja a análise fatorial usada para redução e/ou resumo de dados, o pesquisador deve sempre considerar as bases conceituais das variáveis e julgar quanto à adequação das variáveis para a análise fatorial.

- Em ambos os usos da análise fatorial, o pesquisador implicitamente especifica as dimensões potenciais que podem ser identificadas por meio do caráter e da natureza das variáveis submetidas à análise fatorial. Por exemplo, ao avaliar as dimensões de imagem da loja, se nenhuma questão sobre pessoal da loja for incluída, a análise fatorial não será capaz de identificar tal dimensão.
- O pesquisador deve também lembrar que análise fatorial sempre produzirá fatores. Assim, a análise fatorial é sempre um candidato potencial para o fenômeno "lixo dentro, lixo fora". Se o pesquisador incluir indiscriminadamente um grande número de variáveis e esperar que a análise fatorial "arrume as coisas", então torna-se elevada a possibilidade de resultados pobres. A qualidade e o significado dos fatores obtidos reflete as bases conceituais das variáveis incluídas na análise.

Obviamente, o emprego da análise fatorial como uma técnica de resumo de dados baseia-se em ter uma base conceitual para qualquer variável analisada. Mas ainda que a análise fatorial seja usada apenas para fins de redução de dados, ela é mais eficiente quando dimensões conceitualmente definidas podem ser representadas pelos fatores obtidos.

Uso de análise fatorial com outras técnicas multivariadas

A análise fatorial, por fornecer uma visão muito direta das inter-relações entre variáveis e a estrutura subjacente dos dados, é um excelente ponto de partida para muitas outras técnicas multivariadas. Da perspectiva do resumo de dados, a análise fatorial fornece ao pesquisador uma clara compreensão sobre quais variáveis podem atuar juntas e quantas variáveis podem realmente ser consideradas como tendo impacto na análise.

- Variáveis determinadas como altamente correlacionadas e membros do mesmo fator devem ter perfis semelhantes de diferenças ao longo de grupos em análise multivariada de variância ou em análise discriminante.
- Variáveis altamente correlacionadas, como aquelas dentro de um fator, afetam os procedimentos por etapas de regressão múltipla e análise discriminante que seqüencialmente incluem variáveis com base em seu poder preditivo incremental sobre variáveis já presentes no modelo. Quando se inclui uma variável de um fator, torna-se menos provável que variáveis adicionais do mesmo fator sejam também incluídas, devido a suas elevadas correlações com variáveis já presentes no modelo, o que significa que elas têm pouco poder preditivo a ser acrescentado. Isso não significa que as outras variáveis do fator são menos importantes ou têm menor impacto, mas que seus efeitos já estão representados pela variável incluída do fator. Assim, o conhecimento em si da estrutura das variáveis daria ao pesquisador uma melhor compreensão da razão por trás da entrada de variáveis nesta técnica.

A visão dada pelo resumo de dados pode ser diretamente incorporada em outras técnicas multivariadas por meio de qualquer técnica de redução de dados. A análise fatorial fornece a base para a criação de um novo conjunto de variáveis que incorporam o caráter e a natureza das variáveis originais em um número muito menor de novas variáveis, usando variáveis representativas, escores fatoriais ou escalas múltiplas. Dessa maneira, problemas associados com grandes números de variáveis ou altas intercorrelações entre variáveis podem ser substancialmente reduzidos pela substituição das novas variáveis. O pesquisador pode se beneficiar com a estimação empírica de relações, bem como com a visão do fundamento conceitual e da interpretação dos resultados.

Estágio 2: Planejamento de uma análise fatorial

O planejamento de uma análise fatorial envolve três decisões básicas: (1) cálculo dos dados de entrada (uma matriz de correlação) para atender os objetivos especificados de agrupamento de variáveis ou respondentes; (2) planejamento do estudo em termos de número de variáveis, propriedades de medida das variáveis e tipos de variáveis admissíveis; e (3) o tamanho necessário para a amostra em termos absolutos e como função do número de variáveis na análise.

Correlações entre variáveis ou respondentes

A primeira decisão no planejamento de uma análise fatorial focaliza o cálculo dos dados de entrada para a análise. Discutimos anteriormente as duas formas de análise fatorial: análise fatorial do tipo *R* versus tipo *Q*. Ambos os tipos utilizam uma matriz de correlação como os dados de entrada básicos. Com a análise fatorial do tipo *R*, o pesquisador usaria uma matriz tradicional de correlação (correlações entre variáveis) como entrada. Mas o pesquisador poderia também escolher a opção de obter a matriz de correlação a partir das correlações entre os respondentes individuais. Nessa análise fatorial de tipo *Q*, os resultados seriam uma matriz fatorial que identificaria indivíduos semelhantes.

> Por exemplo, se os respondentes individuais são identificados por números, o padrão fatorial resultante pode nos dizer que os indivíduos 1, 5, 6 e 7 são semelhantes. Do mesmo modo, os respondentes 2, 3, 4 e 8 talvez ocupassem juntos um outro fator, e então os rotularíamos como semelhantes.

Dos resultados de uma análise fatorial *Q*, poderíamos identificar grupos ou agrupamentos de indivíduos que demonstrassem um padrão parecido nas variáveis incluídas na análise.

Uma questão lógica neste ponto seria: de que forma a análise fatorial do tipo *Q* difere da análise de agrupamentos, uma vez que ambas comparam o padrão de respostas ao longo de várias variáveis e estabelecem os respondentes em grupos? A resposta é que a análise fatorial *Q* é baseada nas intercorrelações entre os respondentes, enquanto a análise de agrupamentos forma grupos com base em uma medida de similaridade dada em termos de distância entre os escores dos respondentes para as variáveis que são analisadas.

> Para ilustrar essa diferença, considere a Figura 3-3, que contém os escores de quatro respondentes para três variáveis diferentes. Uma análise fatorial do tipo *Q* desses quatro respondentes produziria dois grupos com estruturas de covariância semelhantes, que consistiria nos respondentes A e C versus B e D. Em contrapartida, a análise de agrupamentos seria sensível às distâncias reais entre os escores dos respondentes e conduziria a um agrupamento dos pares mais próximos. Logo, com uma análise de agrupamentos, os respondentes A e B seriam colocados em um grupo, e C e D, em outro.

	Variáveis		
Respondente	V_1	V_2	V_3
A	7	7	8
B	8	6	6
C	2	2	3
D	3	1	1

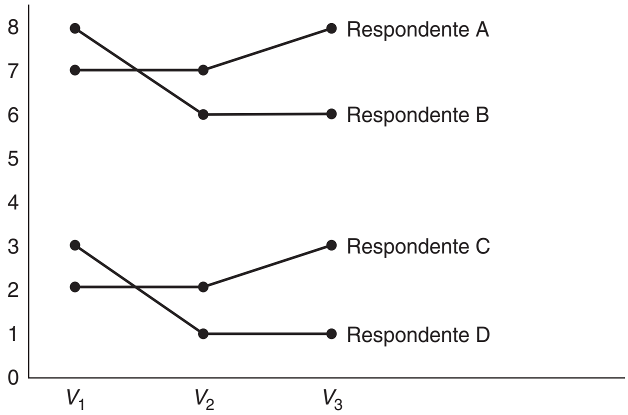

FIGURA 3-3 Comparações de perfis de escore para análise fatorial do tipo *Q* e análise de agrupamentos.

Se o pesquisador decidir empregar análise fatorial do tipo *Q*, essas diferenças em relação às técnicas tradicionais de análise de agrupamentos deverão ser notadas. Com a disponibilidade de outras técnicas de agrupamento e o amplo uso de análise fatorial para redução e resumo de dados, o restante da discussão neste capítulo focaliza a análise fatorial do tipo *R*, o agrupamento de variáveis e não de respondentes.

Questões sobre seleção de variáveis e medidas

Duas questões específicas devem ser respondidas neste ponto: (1) quais tipos de variáveis podem ser usadas em análise fatorial? e (2) quantas variáveis devem ser incluídas? Em termos dos tipos de variáveis incluídas, o requisito principal é que um valor de correlação possa ser calculado entre todas as variáveis. Variáveis métricas são facilmente medidas por vários tipos de correlações. Variáveis não-métricas, contudo, são mais problemáticas por não poderem usar os mesmos tipos de medida de correlação empregados em variáveis métricas. Apesar de alguns métodos especializados calcularem correlações entre variáveis não-métricas, a abordagem mais prudente é evitá-las. Se uma variável não-métrica deve ser incluída, um método é definir **variáveis dicotômicas** (codificadas como 0 e 1) para representarem categorias de variáveis não-métricas. Se todas as variáveis são dicotômicas, então formas especializadas de análise fatorial, como análise fatorial booleana, são mais adequadas [5].

O pesquisador também deve tentar minimizar o número de variáveis incluídas, mas manter um número razoável de variáveis por fator. Se um estudo está sendo planejado para avaliar uma estrutura proposta, o pesquisador deve certificar-se de incluir diversas variáveis (cinco ou mais) que possam representar cada fator proposto. A força da análise fatorial reside em encontrar padrões entre grupos de variáveis, e é de pouco uso na identificação de fatores compostos por uma única variável. Finalmente, quando se planeja um estudo para ser analisado por fatores, o pesquisador deve, se possível, identificar diversas variáveis-chave (algumas vezes chamadas de indicadores-chave ou variáveis de marcação) que intimamente reflitam os fatores latentes que foram previstos hipoteticamente. Isso ajudará na validação dos fatores determinados e na avaliação da significância prática dos resultados.

Tamanho da amostra

No que se refere à questão do tamanho da amostra, o pesquisador dificilmente realiza uma análise fatorial com uma amostra com menos de 50 observações, e de preferência o tamanho da amostra deve ser maior ou igual a 100. Como regra geral, o mínimo é ter pelo menos cinco vezes mais observações do que o número de variáveis a serem analisadas, e o tamanho mais aceitável teria uma proporção de dez para um. Alguns pesquisadores chegam a propor um mínimo de 20 casos para cada variável. Deve-se lembrar, contudo, que com 30 variáveis, por exemplo, há 435 correlações a serem calculadas na análise fatorial. Em um nível de significância de 0,05, talvez até mesmo 20 dessas correlações fossem consideradas significantes e aparecessem na análise fatorial somente por sorte. O pesquisador sempre deve tentar obter a maior proporção de casos-por-variável para minimizar as chances de superajustar os dados (ou seja, determinar fatores específicos da amostra, com pouca generalidade). O pesquisador pode fazer isso empregando o conjunto de variáveis mais parcimonioso, guiado por considerações conceituais e práticas, e então obtendo um tamanho adequado da amostra para o número de variáveis examinadas. Quando se lida com amostras menores e/ou com uma proporção menor de casos-por-variáveis, o pesquisador sempre deve interpretar qualquer descoberta com precaução. A questão do tamanho da amostra também será abordada em uma seção adiante ao se interpretarem cargas fatoriais.

Resumo

Questões no planejamento de uma análise fatorial são de importância igualmente crítica se uma perspectiva exploratória ou confirmatória é assumida. Sob qualquer ponto de vista, o pesquisador está confiando na técnica para fornecer uma visão sobre a estrutura dos dados, mas a estrutura revelada na análise depende de decisões do pesquisador em áreas como variáveis incluídas, tamanho da amostra e assim por diante. Desse modo, diversas considerações-chave são listadas nas Regras Práticas 3-1.

Estágio 3: Suposições na análise fatorial

As suposições críticas na análise fatorial são mais conceituais do que estatísticas. O pesquisador está sempre preocupado em atender a exigência estatística para qualquer técnica multivariada, mas em análise fatorial as preocupações que se impõem se centram muito mais no caráter e na composição das variáveis incluídas na análise do que em suas qualidades estatísticas.

Questões conceituais

As premissas conceituais subjacentes à análise fatorial se referem ao conjunto de variáveis selecionadas e à amostra escolhida. Uma suposição básica da análise fatorial é que *existe alguma estrutura subjacente* no conjunto de variáveis escolhidas. A presença de variáveis correlacionadas e a subseqüente definição de fatores não garantem relevância, mesmo que elas satisfaçam as exigências estatísticas. É responsabilidade do pesquisador garantir que os padrões observados sejam conceitualmente válidos e adequados para se estudar com análise fatorial, pois a técnica não dispõe de meios para determinar adequação além das correlações entre variáveis. Por exemplo, misturar variáveis dependentes e independentes em uma análise fatorial e então usar os fatores obtidos para apoiar relações de dependência é inadequado.

REGRAS PRÁTICAS 3-1

Planejamento de análise fatorial

- Análise fatorial é executada geralmente apenas sobre variáveis métricas, apesar de existirem métodos especializados para o emprego de variáveis dicotômicas; um número pequeno de "variáveis dicotômicas" pode ser incluído em um conjunto de variáveis métricas que são analisadas por fatores.
- Se um estudo está sendo planejado para revelar estrutura fatorial, esforce-se para ter pelo menos cinco variáveis para cada fator proposto.
- Para tamanho de amostra:
  - A amostra deve ter mais observações do que variáveis.
  - O menor tamanho absoluto de amostra deve ser de 50 observações.
- Maximize o número de observações por variável, com um mínimo de 5 e, com sorte, com pelo menos 10 observações por variável.

O pesquisador deve também garantir que a amostra é homogênea com relação à estrutura fatorial inerente. É inadequado aplicar análise fatorial em uma amostra de homens e mulheres para um conjunto de itens conhecidos por diferirem por conta de sexo. Quando as duas subamostras (homens e mulheres) são combinadas, as correlações resultantes e a estrutura de fatores serão uma representação pobre da estrutura exclusiva de cada grupo. Logo, sempre que grupos diferentes são esperados na amostra, análises fatoriais separadas devem ser executadas, e os resultados devem ser comparados para identificar diferenças não refletidas nos resultados da amostra combinada.

Questões estatísticas

De um ponto de vista estatístico, os desvios da normalidade, da homocedasticidade e da linearidade aplicam-se apenas porque eles diminuem as correlações observadas. Apenas a normalidade é necessária se um teste estatístico é aplicado para a significância dos fatores, mas esses testes raramente são usados. Na verdade, um pouco de **multicolinearidade** é desejável, pois o objetivo é identificar conjuntos de variáveis inter-relacionadas.

Assumindo que o pesquisador atende as exigências conceituais para as variáveis incluídas na análise, o próximo passo é garantir que as variáveis são suficientemente correlacionadas umas com as outras para produzir fatores representativos. Como veremos, podemos avaliar esse grau de relacionamento a partir de pontos de vista geral ou individual. A seguir há diversas medidas empíricas para ajudar no diagnóstico da fatorabilidade da matriz de correlação.

Medidas gerais de intercorrelação. Além das bases estatísticas para as correlações da matriz de dados, o pesquisador também deve garantir que a matriz de dados tenha correlações suficientes para justificar a aplicação da análise fatorial. Se se descobrir que todas as correlações são pequenas, ou que todas as correlações são iguais (mostrando que não existe qualquer estrutura para agrupar variáveis), então o pesquisador deve questionar a aplicação de análise fatorial. Para esse propósito, diversas abordagens estão disponíveis:

- Se a inspeção visual não revela um número substancial de correlações maiores que 0,30, então a análise fatorial provavelmente é inapropriada. As correlações entre variáveis também podem ser analisadas computando-se as correlações parciais entre variáveis. Uma correlação parcial é aquela que não é explicada quando os efeitos de outras variáveis são levados em consideração. Se existem fatores "verdadeiros" nos dados, a correlação parcial deverá ser pequena, pois a variável pode ser explicada pelas variáveis que compõem os fatores. Se as correlações parciais são altas, indicando ausência de fatores inerentes, então a análise fatorial é inadequada. O pesquisador está procurando um padrão de altas correlações parciais, denotando uma variável não correlacionada com um grande número de outras variáveis na análise.

A única exceção referente a elevadas correlações como indicativas de uma matriz de correlação pobre acontece quando duas variáveis estão altamente correlacionadas e têm cargas substancialmente maiores do que outras variáveis naquele fator. Logo, a correlação parcial delas pode ser elevada porque elas não são explicadas em grande parte pelas outras variáveis, mas explicam umas a outras. Essa exceção espera-se também quando um fator tem somente duas variáveis com cargas elevadas.

Uma elevada correlação parcial é aquela com significância prática e estatística, e uma regra prática seria considerar correlações parciais acima de 0,7 como elevadas. O SPSS e o SAS fornecem a **matriz de correlação anti-imagem**, que é simplesmente o valor negativo da correlação parcial, enquanto o BMDP nos dá as correlações parciais diretamente. Em cada caso, as correlações parciais ou correlações antiimagem maiores são indicativas de uma matriz de dados que talvez não seja adequada para análise fatorial.

- Outro modo de determinar a adequação da análise fatorial examina a matriz de correlação inteira. O **teste de esfericidade de Bartlett**, um teste estatístico para a presença de correlações entre as variáveis, é uma medida dessa natureza. Ele fornece a significância estatística de que a matriz de correlação tem correlações significantes entre pelo menos algumas das variáveis. O pesquisador deve perceber, porém, que aumentar o tamanho da amostra faz com que o teste Bartlett se torne mais sensível na detecção de correlações entre as variáveis.
- Uma terceira medida para quantificar o grau de intercorrelações entre as variáveis e a adequação da análise fatorial é a **medida de adequação da amostra (MSA)**. Esse índice varia de 0 a 1, alcançando 1 quando cada variável é perfeitamente prevista sem erro pelas outras variáveis. A medida pode ser interpretada com as seguintes orientações: 0,80 ou acima, admirável; 0,70 ou acima, mediano; 0,60 ou acima, medíocre; 0,50 ou acima, ruim; e abaixo de 0,50, inaceitável [22,23]*. O MSA aumenta quando (1) o tamanho da amostra aumenta, (2) as correlações médias aumentam, (3) o número de variáveis aumenta, ou (4) o número de fatores diminui [23]. O pesquisador sempre deve ter um valor MSA geral acima de 0,50 antes de proceder com a análise fatorial. Se o valor MSA ficar abaixo de 0,50, então os valores específicos MSA (ver a discussão que se segue) podem identificar variáveis para eliminação para atingir um valor geral de 0,50.

Medidas específicas de intercorrelação de variáveis. Além de um exame visual das correlações de uma variável com outras na análise, as orientações MSA podem ser estendidas para variáveis individuais. O pesquisador deve examinar os valores MSA para cada variável e excluir aquelas que estão no domínio inaceitável. No processo de eliminação de variáveis, o pesquisador deve primeiro eliminar a variável com o menor MSA e então recalcular a análise fatorial. Continue esse processo de eliminar a variável com o menor valor MSA abaixo de 0,50 até que todas as variáveis tenham um valor aceitável. Uma vez que variáveis individuais atinjam um nível aceitável, então o MSA geral pode ser calculado e uma decisão pode ser tomada sobre a continuidade da análise fatorial.

* N. de R. T.: A frase correta seria "0,80 ou acima, admirável; maior ou igual a 0,70 e abaixo de 0,80, mediano; maior ou igual a 0,60 e abaixo de 0,70, medíocre; maior ou igual a 0,50 e abaixo de 0,60, ruim; e abaixo de 0,50, inaceitável".

Resumo

A análise fatorial, enquanto técnica de interdependência, é de várias formas mais afetada se não atender suas premissas conceituais inerentes do que pelas suposições estatísticas. O pesquisador deve se certificar de compreender completamente as implicações não apenas de garantir que os dados atendem as exigências estatísticas para uma estimação apropriada da estrutura fatorial, mas de que o conjunto de variáveis tem a fundamentação conceitual para embasar os resultados. Fazendo isso, o pesquisador deve considerar várias orientações importantes, como listadas nas Regras Práticas 3-2.

Estágio 4: Determinação de fatores e avaliação do ajuste geral

Uma vez que as variáveis sejam especificadas e a matriz de correlação seja preparada, o pesquisador está pronto para aplicar a análise fatorial para identificar a estrutura latente de relações (ver Figura 3-4). Nisso, as decisões devem ser tomadas com relação (1) ao método de extração dos fatores (análise de fatores comuns versus análise de componentes) e (2) ao número de fatores selecionados para explicar a estrutura latente dos dados.

Seleção do método de extração de fatores

O pesquisador pode escolher a partir de dois métodos semelhantes, ainda que únicos, para definir (extrair) os fatores que representem a estrutura das variáveis na análise. Essa decisão sobre o método a ser usado deve combinar os objetivos da análise fatorial com o conhecimento sobre algumas características básicas das relações entre variáveis. Antes de discutirmos sobre os dois métodos disponíveis para extração de fatores, apresentamos uma breve introdução à partição da variância de uma variável.

REGRAS PRÁTICAS 3-2

Teste das suposições da análise fatorial

- Uma forte fundamentação conceitual é necessária para embasar a suposição de que existe uma estrutura antes que a análise fatorial seja realizada.
- Um teste de esfericidade de Bartlett estatisticamente significante (sign. < 0,05) indica que correlações suficientes existem entre as variáveis para se continuar a análise.
- Medidas de valores de adequação da amostra (MSA) devem exceder 0,50 tanto para o teste geral quanto para cada variável individual; variáveis com valores inferiores a 0,50 devem ser omitidas da análise fatorial uma por vez, sendo aquela com menor valor eliminada a cada vez.

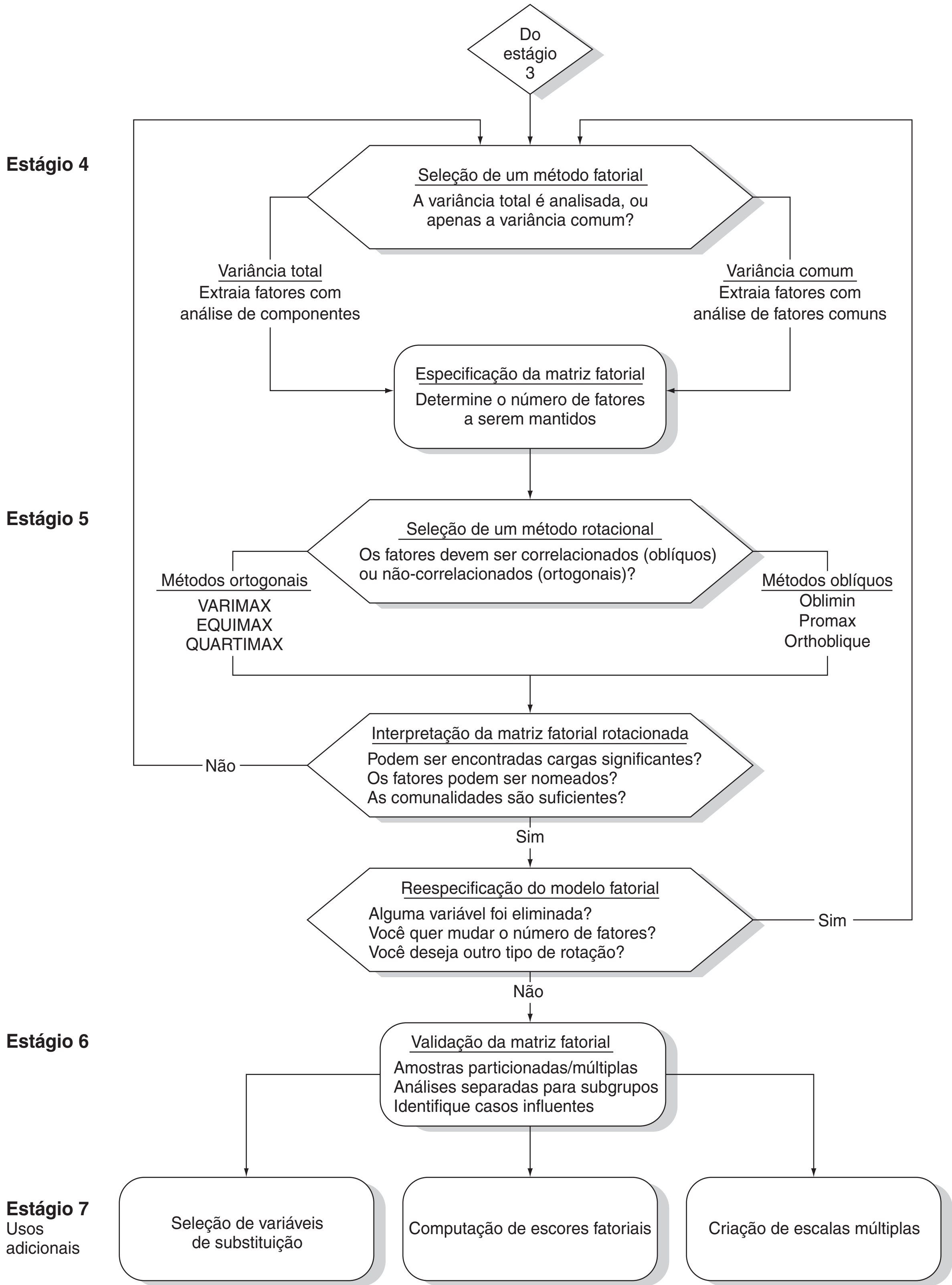

FIGURA 3-4 Estágios 4-7 no diagrama de decisão da análise fatorial.

Partição da variância de uma variável. Para escolher entre os dois métodos de extração de fatores, o pesquisador deve primeiro ter certa compreensão da variância para uma variável e como ela é dividida ou particionada. Primeiro, lembre que variância é um valor (i.e., o quadrado do desvio padrão) que representa a quantia total de dispersão de valores para uma única variável em torno de sua média. Quando uma variável é correlacionada com outra, dizemos muitas vezes que ela compartilha variância com a outra variável, e essa quantia de partilha entre apenas duas variáveis é simplesmente a correlação ao quadrado. Por exemplo, se duas variáveis têm uma correlação de 0,50, cada variável compartilha 25% ($0{,}50^2$) de sua variância com a outra.

Em análise fatorial, agrupamos variáveis por suas correlações, de modo que variáveis em um grupo (fator) têm elevadas correlações umas com as outras. Assim, para os propósitos da análise fatorial, é importante entender o quanto da variância de uma variável é compartilhado com outras variáveis naquele fator versus o que não pode ser compartilhado (p.ex., inexplicado). A variância total de qualquer variável pode ser dividida (particionada) em três tipos de variância:

1. **Variância comum** é definida como aquela variância em uma variável que é compartilhada com todas as outras variáveis na análise. Essa variância é explicada (compartilhada) com base nas correlações de uma variável com as demais na análise. A **comunalidade** de uma variável é a estimativa de sua variância compartilhada, ou em comum, entre as variáveis como representadas pelos fatores obtidos.
2. **Variância específica** (também conhecida como **variância única**) é aquela associada com apenas uma variável específica. Essa variância não pode ser explicada pelas correlações com as outras variáveis, mas ainda é associada unicamente com uma variável.
3. **Variância de erro** é também variância que não pode ser explicada por correlações com outras variáveis, mas resulta da não confiabilidade no processo de coleta de dados, de erro de medida ou de componente aleatório no fenômeno medido.

Assim, a variância total de uma variável é composta de suas variâncias comum, única e de erro. Quando uma variável é mais correlacionada com uma ou mais variáveis, a variância comum (comunalidade) aumenta. Por outro lado, se medidas não-confiáveis ou outras fontes de variância de erros externos são introduzidas, então a quantia de variância comum possível é reduzida, bem como a habilidade de relacionar a variável com qualquer outra.

Análise de fatores comuns versus análise de componentes. Com uma compreensão básica sobre como a variância pode ser particionada, o pesquisador está pronto para abordar as diferenças entre os dois métodos, conhecidos como **análise de fatores comuns e análise de componentes**. A escolha de um método em vez do outro é baseada em dois critérios: (1) os objetivos da análise fatorial e (2) o montante de conhecimento prévio sobre a variância nas variáveis. A análise de componentes é usada quando o objetivo é resumir a maior parte da informação original (variância) a um número mínimo de fatores para fins de previsão. Em contraste, análise de fatores comuns é usada prioritariamente para identificar fatores ou dimensões latentes que refletem o que as variáveis têm em comum. A comparação mais direta entre os dois métodos é pelo seu uso da variância explicada versus não-explicada:

- Análise de componente, também conhecida como análise de componentes principais, *considera a variância total e deriva fatores que contêm pequenas proporções de variância única e, em alguns casos, variância de erro*. Não obstante, os primeiro poucos fatores não contêm variância de erro ou única o suficiente para distorcer a estrutura fatorial geral. Especificamente, com análise de componentes, unidades (valores de 1,0) são inseridas na diagonal da matriz de correlação, de modo que a variância completa é trazida à matriz fatorial. A Figura 3-5 retrata o emprego da variância total em análise de componentes e as diferenças quando comparada com análise de fatores comuns.
- Análise de fator comum, em contraste, considera apenas variância em comum ou compartilhada, *assumindo que tanto a variância de erro quanto a única não são de interesse na definição da estrutura das variáveis*. Para empregar apenas variância comum na estimação dos fatores, comunalidades (ao invés de unidades) são inseridas na diagonal. Assim, fatores resultantes da análise de fator comum se baseiam somente na variância comum. Como mostrado na Figura 3-5, a análise de fator comum exclui uma porção da variância incluída em uma análise de componentes.

Como o pesquisador escolherá entre os dois métodos? Primeiro, tanto o modelo de fator comum quanto o de análise de componente são amplamente usados. Em termos práticos, o modelo por componentes é o método padrão típico da maioria dos programas estatísticos, quando se executa uma análise fatorial. Além do padrão em programas, casos distintos indicam qual dos dois métodos é o mais adequado:

A análise fatorial de componentes é a mais adequada quando:

- *redução de dados é uma preocupação prioritária*, focando o número mínimo de fatores necessários para explicar a porção máxima da *variância total* representada no conjunto original de variáveis, e
- conhecimento anterior sugere que variância específica e de erro representam uma *proporção relativamente pequena* da variância total.

Análise de fatores comuns é mais apropriada quando:

- *o objetivo prioritário é identificar as dimensões ou construtos latentes* representados nas variáveis originais, e
- o pesquisador tem *pouco conhecimento sobre a quantia de variância específica e de erro*, e, portanto, deseja eliminar essa variância.

A análise de fatores comuns, com suas suposições mais restritivas e uso apenas de dimensões latentes (variância compartilhada), muitas vezes é vista como algo teoricamente mais fundamentado. No entanto, apesar de teoricamente válida, ela tem vários problemas. Primeiro, a análise de fa-

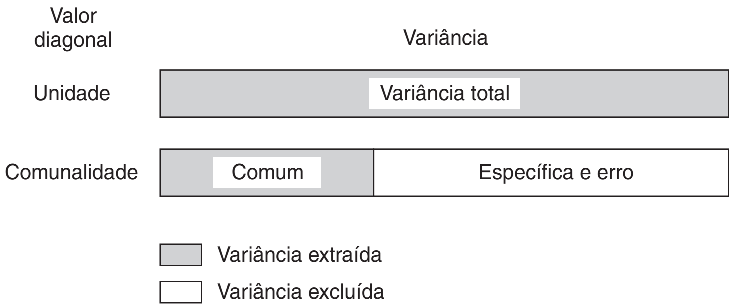

FIGURA 3-5 Tipos de variância considerados na matriz fatorial.

tores comuns sofre de **indeterminância fatorial**, o que significa que para qualquer respondente individual, diversos escores fatoriais diferentes podem ser calculados a partir dos resultados de um único modelo fatorial [26]. Não há solução única, como ocorre em análise de componentes, mas na maioria dos casos as diferenças não são substanciais. A segunda questão envolve o cálculo de comunalidades estimadas usadas para representar a variância compartilhada. Às vezes as comunalidades não são estimáveis ou podem ser inválidas (p.ex., valores maiores que 1 ou menores que 0), exigindo a eliminação da variável da análise.

A escolha de um modelo ou de outro realmente afeta os resultados? As complicações da análise de fatores comuns têm contribuído para o amplo uso de análise de componentes. Mas a base de proponentes para o modelo de fator comum é forte. Cliff [13] caracteriza a disputa entre os dois lados como se segue:

> Algumas autoridades insistem que análise de componentes é a única abordagem adequada, e que os métodos de fatores comuns apenas impõem terminologia confusa, lidando com coisas fundamentalmente não-mensuráveis, os fatores comuns. Os sentimentos são, em certo sentido, ainda mais fortes no outro lado. Partidários da análise de fatores comuns insistem que a análise de componentes é, na melhor das hipóteses, uma análise de fatores comuns com algum erro acrescentado, e, na pior das hipóteses, uma mistura confusa e inaceitável de coisas a partir das quais nada pode ser determinado. Alguns chegam a insistir que o termo "análise fatorial" não deve ser usado quando a análise de componentes é executada.

Apesar de ainda haver muito debate sobre qual modelo fatorial é o mais apropriado [6, 19, 25, 35], a pesquisa empírica tem demonstrado resultados análogos em muitos casos [37]. Na maioria das aplicações, *tanto a análise de componentes quanto a análise de fatores comuns chegam a resultados essencialmente idênticos se o número de variáveis exceder 30 [18], ou se as comunalidades excederem 0,60 para a maioria das variáveis*. Se o pesquisador estiver preocupado com as suposições da análise de componentes, então a análise de fatores comuns também deve ser aplicada para avaliar sua representação da estrutura.

Quando uma decisão foi tomada no modelo fatorial, o pesquisador está pronto para extrair os fatores iniciais não-rotacionados. Examinando a matriz fatorial não-rotacionada, ele pode explorar as possibilidades de redução de dados e obter uma estimativa preliminar do número de fatores a extrair. A determinação final do número de fatores, porém, deve esperar até o momento em que os resultados sejam rotacionados e os fatores sejam interpretados.

Critérios para o número de fatores a extrair

Como decidimos sobre o número de fatores a serem extraídos? Ambos os métodos de análise fatorial estão interessados na melhor combinação linear de variáveis – melhor no sentido de que a combinação particular de variáveis originais explica a maior parte da variância nos dados como um todo comparada a qualquer outra combinação linear de variáveis. Logo, o primeiro fator pode ser visto como o melhor resumo de relações lineares exibidas nos dados. O segundo fator é definido como a segunda melhor combinação linear das variáveis, sujeita à restrição de que é ortogonal ao primeiro fator. Para ser **ortogonal** ao primeiro fator, o segundo fator deve ser obtido da variância remanescente depois que o primeiro fator foi extraído. Assim, o segundo fator pode ser definido como a combinação linear de variáveis que explica a maior parte da variância que ainda é inexplicada após o efeito da remoção do primeiro fator dos dados. O processo continua extraindo fatores que explicam quantias cada vez menores de variância até que toda a variância seja explicada. Por exemplo, o método de componentes realmente extrai n fatores, onde n é o número de variáveis na análise. Assim, se 30 variáveis estão na análise, 30 fatores são extraídos.

Assim, o que se ganha com análise fatorial? Apesar de nosso exemplo conter 30 fatores, alguns dos primeiros fatores podem explicar uma porção substancial da variância total ao longo de todas as variáveis. Espera-se que o pesquisador possa reter ou usar apenas um pequeno número de variáveis* e ainda representar adequadamente o conjunto inteiro de variáveis. Assim, a questão-chave é: *quantos fatores devem ser extraídos ou retidos*?

* N. de R. T.: A palavra correta seria "fatores".

Ao decidir quando parar a fatoração (i.e., quantos fatores devem ser extraídos), o pesquisador deve combinar uma fundamentação conceitual (quantos fatores devem estar na estrutura?) com alguma evidência empírica (quantos fatores podem ser razoavelmente sustentados?). O pesquisador geralmente começa com alguns critérios pré-determinados, como o número geral de fatores mais algumas referências gerais de relevância prática (p.ex., percentual exigido de variância explicada). Esses critérios são combinados com medidas empíricas da estrutura fatorial. Uma base quantitativa exata para decidir o número de fatores a extrair ainda não foi desenvolvida. No entanto, os seguintes critérios de parada têm sido utilizados.

Critério da raiz latente. A técnica mais comumente usada é o critério da raiz latente. Esta técnica é simples de aplicar na análise de componentes, bem como na análise de fatores comuns. O raciocínio para o critério da raiz latente é que qualquer fator individual deve explicar a variância de pelo menos uma variável se o mesmo há de ser mantido para interpretação. Com a análise de componentes, cada variável contribui com um valor 1 do autovalor total. Logo, apenas os fatores que têm **raízes latentes** ou **autovalores** maiores que 1 são considerados significantes; todos os fatores com raízes latentes menores que 1 são considerados insignificantes e são descartados. Usar o autovalor para estabelecer um corte é mais confiável quando o número de variáveis está entre 20 e 50. Se o número de variáveis é menor que 20, há uma tendência para que esse método extraia um número conservador (muito pouco) de fatores; ao passo que, quando mais de 50 variáveis estão envolvidas, não é raro que muitos fatores sejam extraídos.

Critério *a priori*. O critério *a priori* é um critério simples, ainda que razoável sob certas circunstâncias. Quando aplicado, o pesquisador já sabe quantos fatores extrair antes de empreender a análise fatorial. O pesquisador simplesmente instrui o computador a parar a análise quando o número desejado de fatores tiver sido extraído. Este tratamento é útil quando se testa uma teoria ou hipótese sobre o número de fatores a serem extraídos. Também pode ser justificado na tentativa de repetir o trabalho de outro pesquisador e extrair o mesmo número de fatores anteriormente encontrado.

Critério de percentagem de variância O critério de percentagem de variância é uma abordagem baseada na conquista de um percentual cumulativo especificado da variância total extraída por fatores sucessivos. O objetivo é garantir significância prática para os fatores determinados, garantindo que expliquem pelo menos um montante especificado de variância. Nenhuma base absoluta foi adotada para todas as aplicações. No entanto, em ciências naturais, o procedimento de obtenção de fatores geralmente não deveria ser parado até os fatores extraídos explicarem pelo menos 95% da variância, ou até o último fator explicar apenas uma pequena parcela (menos que 5%). Em contraste, em ciências sociais, nas quais as informações geralmente são menos precisas, não é raro considerar uma solução que explique 60% da variância total (e em alguns casos até menos) como satisfatória.

Uma variante deste critério envolve a seleção de fatores suficientes para atingir uma comunalidade pré-especificada para cada variável. Se razões teóricas ou práticas requerem uma certa comunalidade para cada variável, então o pesquisador incluirá tantos fatores quanto necessários para representar adequadamente cada uma das variáveis originais. Isso difere de focalizar somente o montante total de variância explicada, o que negligencia o grau de explicação para as variáveis individuais.

Critério do teste *scree*. Lembre que, no modelo fatorial de análise de componentes, os últimos fatores extraídos contêm tanto a variância comum quanto a única. Apesar de todos os fatores conterem pelo menos alguma variância única, a proporção de variância única é substancialmente maior nos últimos fatores. O teste *scree* é usado para identificar o número ótimo de fatores que podem ser extraídos antes que a quantia de variância única comece a dominar a estrutura de variância comum [9].

O teste *scree* é determinado fazendo-se o gráfico das raízes latentes em relação ao número de fatores em sua ordem de extração, e a forma da curva resultante é usada para avaliar o ponto de corte. A Figura 3-6 exibe os primeiros 18 fatores extraídos em um estudo. Começando com o primeiro fator, os ângulos de inclinação rapidamente decrescem no início e então lentamente se aproximam de uma reta horizontal. O ponto no qual o gráfico começa a ficar horizontal é considerado indicativo do número máximo de fatores a serem extraídos. No presente caso, os primeiros 10 fatores se qualificariam. Além de 10, uma grande proporção de variância única seria incluída; logo, esses fatores não seriam aceitáveis. Observe que, ao se usar o critério da raiz latente, apenas 8 fatores teriam sido considerados. Entretanto, usar o teste *scree* nos dá 2 fatores a mais. Como regra geral, o teste *scree* resulta em pelo menos um e às vezes dois ou três fatores a mais sendo considerados para inclusão em relação ao critério da raiz latente [9].

Heterogeneidade dos respondentes. A variância compartilhada entre variáveis é a base para ambos os modelos fatoriais, de fator comum e de componentes. Uma suposição inerente é que a variância compartilhada se estende ao longo de toda a amostra. Se esta é heterogênea em relação a pelo menos um subconjunto das variáveis, então os primeiros fatores representam aquelas variáveis mais homogêneas em toda a amostra. As variáveis que são melhores discriminadoras entre os subgrupos da amostra carregam nos últimos fatores, muitas vezes aqueles não-selecionados pelos critérios recém discutidos [17]. Quan-

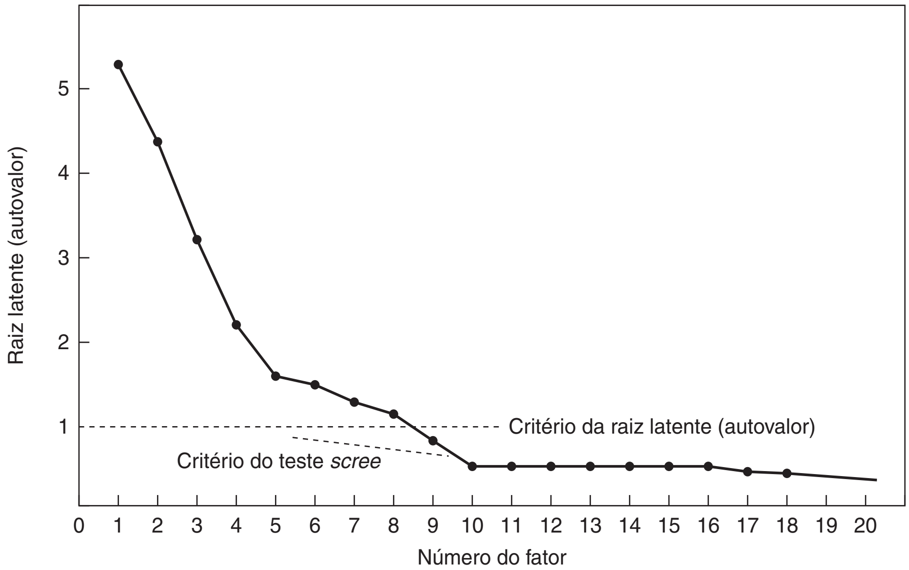

FIGURA 3-6 Gráfico de autovalor para o critério de teste *scree*.

do o objetivo é identificar fatores que discriminam entre os subgrupos de uma amostra, o pesquisador deve extrair fatores adicionais, além dos indicados pelos métodos citados, e examinar a habilidade dos fatores adicionais de discriminar os grupos. Se eles demonstrarem serem menos benéficos na discriminação, a solução poderá ser processada novamente e esses últimos fatores, eliminados.

Resumo dos critérios de seleção de fatores. Na prática, a maioria dos pesquisadores raramente usa um único critério para determinar quantos fatores devem ser extraídos. Inicialmente eles usam um critério como o da raiz latente como uma orientação para a primeira tentativa de interpretação. Depois que os fatores foram interpretados, como discutido nas seções seguintes, a sua praticabilidade é avaliada. Os fatores identificados por outros critérios também são interpretados. A seleção do número de fatores é inter-relacionada com uma avaliação da estrutura, a qual é revelada na fase de interpretação. Assim, diversas soluções fatoriais com diferentes números de fatores são examinadas antes que a estrutura seja bem definida. Ao tomar a decisão final sobre a solução fatorial para representar a estrutura das variáveis, o pesquisador deve lembrar as considerações listadas nas Regras Práticas 3-3.

Uma advertência quanto à seleção do conjunto final de fatores: há conseqüências negativas na seleção de fatores em excesso ou a menos para representar os dados. Se pouquíssimos fatores são selecionados, a estrutura correta não é revelada e dimensões importantes podem ser omitidas. Se muitos fatores são mantidos, a interpretação se torna mais difícil quando os resultados são rotacionados (como discutido na próxima seção). Apesar de os fatores serem independentes, você pode ter fatores a mais ou a menos, sem dificuldades. Por analogia, escolher o número

REGRAS PRÁTICAS 3-3

Escolha de modelos fatoriais e número de fatores

- Apesar de os modelos de análise de fatores comuns e de análise de componentes levarem a resultados similares em ambientes comuns de pesquisa (30 variáveis ou mais, ou comunalidades de 0,60 para a maioria das variáveis):
  - O modelo de análise de componentes é mais adequado quando a redução de dados é soberana
  - O modelo de fatores comuns é melhor em aplicações teóricas bem especificadas
- Qualquer decisão sobre o número de fatores a serem mantidos deve se basear em diversas considerações:
  - Uso de diversos critérios de parada para determinar o número inicial de fatores a serem mantidos:
    - Fatores com autovalores maiores do que 1,0
    - Um número pré-determinado de fatores baseado em objetivos da pesquisa e/ou pesquisa anterior
    - Fatores suficientes para atender um percentual especificado de variância explicada, geralmente 60% ou mais
    - Fatores apontados pelo teste *scree* como tendo quantias substanciais de variância comum (i.e., fatores antes do ponto de inflexão)
    - Mais fatores quando heterogeneidade está presente entre subgrupos da amostra
  - Consideração de várias soluções alternativas (um fator a mais e um a menos em relação à solução inicial) para garantir que a melhor estrutura seja identificada

de fatores é algo como focar um microscópio. Ajuste muito alto ou muito baixo irá obscurecer uma estrutura que é óbvia quando o ajuste está simplesmente correto. Logo, pelo exame de uma certa quantia de diferentes estruturas fatoriais obtidas a partir de várias tentativas de soluções, o pesquisador pode comparar e contrastar para chegar à melhor representação dos dados.

Assim como outros aspectos de modelos multivariados, a parcimônia é importante. A exceção notável é quando a análise fatorial é usada exclusivamente para redução de dados e um nível estabelecido de variância a ser extraído é especificado. O pesquisador sempre deve se empenhar em ter o conjunto de fatores mais representativo e parcimonioso possível.

Estágio 5: Interpretação dos fatores

Apesar de não existirem processos ou orientações inequívocas para determinar a interpretação de fatores, o pesquisador com forte fundamentação conceitual para a estrutura antecipada e sua justificativa tem a maior chance de sucesso. Não podemos estabelecer de maneira suficientemente impactante a importância de uma forte fundamentação conceitual, seja ela vinda de pesquisa anterior, paradigmas teóricos ou princípios comumente aceitos. Como veremos, o pesquisador deve repetidamente fazer julgamentos subjetivos em decisões, como o número de fatores, quais são as relações suficientes para garantir variáveis que discriminam grupos, e como podem ser identificados esses grupos. Como pode atestar o pesquisador experiente, praticamente qualquer coisa pode ser descoberta se houver empenho suficientemente insistente (p.ex., usando diferentes modelos fatoriais, extraindo diferentes números de fatores, usando várias formas de rotação). Portanto, deixa-se para o pesquisador o papel de juiz de última instância quanto à forma e à adequação de uma solução fatorial, e tais decisões são melhor guiadas por bases conceituais do que por bases empíricas.

Para auxiliar no processo de interpretação de uma estrutura fatorial e escolher a solução final, três processos fundamentais são descritos. Dentro de cada processo, diversas questões importantes (rotação fatorial, significância de carga fatorial e interpretação) são encontradas. Assim, após a breve descrição de cada processo, os mesmos serão discutidos mais detalhadamente.

Os três processos de interpretação fatorial

A interpretação fatorial é circular por natureza. O pesquisador primeiramente avalia os resultados iniciais, em seguida faz alguns julgamentos vendo e refinando tais resultados, com a evidente possibilidade de que a análise seja reespecificada, exigindo-se uma volta ao passo avaliativo. Assim, o pesquisador não deve se surpreender se executar diversas iterações até que uma solução final seja obtida.

Estimativa da matriz fatorial. Primeiro, a **matriz fatorial** inicial não-rotacionada é computada, contendo as cargas fatoriais para cada variável sobre cada fator. **Cargas fatoriais** são a correlação de cada variável com o fator. Cargas indicam o grau de correspondência entre a variável e o fator, com cargas maiores tornando a variável representativa do fator. Cargas fatoriais são o meio de interpretar o papel que cada variável tem na definição de cada fator.

Rotação de fatores. Soluções fatoriais não-rotacionadas atingem a meta de redução de dados, mas o pesquisador deve perguntar se a solução fatorial não-rotacionada (que preenche as exigências matemáticas desejáveis) fornecerá informação que oferece interpretação a mais adequada das variáveis sob exame. Na maioria dos casos, a resposta a essa questão é negativa, pois rotação fatorial (uma discussão mais detalhada segue na próxima seção) deve simplificar a estrutura fatorial. Portanto, o pesquisador a seguir emprega um método rotacional para conseguir soluções mais simples e teoricamente mais significativas. Na maioria das vezes, a rotação de fatores melhora a interpretação pela redução de algumas das ambigüidades que freqüentemente acompanham as soluções fatoriais não-rotacionadas.

Interpretação e reespecificação de fatores. Como um processo final, o pesquisador avalia as cargas fatoriais (rotacionadas) para cada variável a fim de determinar o papel da mesma e sua contribuição na determinação da estrutura fatorial. No curso deste processo de avaliação, pode surgir a necessidade de reespecificar o modelo fatorial devido (1) à eliminação de uma variável(is) da análise, (2) ao desejo de empregar um método rotacional diferente para interpretação, (3) à necessidade de extrair um número diferente de fatores, ou (4) ao desejo de mudar de um método de extração para outro. A reespecificação de um modelo fatorial é realizada retornando-se ao estágio de extração (estágio 4), extraindo fatores e interpretando-os novamente.

Rotação de fatores

Talvez a ferramenta mais importante na interpretação de fatores seja a **rotação fatorial**. O termo **rotação** significa exatamente o que sugere. Especificamente, os eixos de referência dos fatores são rotacionados em torno da origem até que alguma outra posição seja alcançada. Como anteriormente indicado, as soluções de fatores não-rotacionados extraem fatores na ordem de sua variância extraída. O primeiro fator tende a ser um fator geral com quase toda variável com carga significante, e explica a quantia maior de variância. O segundo fator e os seguintes são então baseados na quantia residual de variância. Cada fator explica porções sucessivamente menores de variância. O efeito final de rotacionar a matriz fatorial é redistribuir a variância dos primeiros fatores para os últimos com o objetivo de atingir um padrão fatorial mais simples e teoricamente mais significativo.

O caso mais simples de rotação é uma **rotação ortogonal**, na qual os eixos são mantidos a 90 graus. Também é possível rotacionar os eixos sem manter o ângulo de 90 graus entre os eixos de referência. Quando não há a restrição de ser ortogonal, o procedimento de rotação se chama **rotação oblíqua**. Rotações fatoriais ortogonais e oblíquas são demonstradas nas Figuras 3-7 e 3-8, respectivamente.

A Figura 3-7, na qual cinco variáveis são retratadas em um diagrama fatorial bidimensional, ilustra a rotação fatorial. O eixo vertical representa o fator II não-rotacionado, e o horizontal corresponde ao fator I não-rotacionado. Os eixos são rotulados com 0 na origem e prolongados para +1,0 ou –1,0. Os números nos eixos representam as cargas fatoriais. As cinco variáveis são rotuladas por V_1, V_2, V_3, V_4 e V_5. A carga fatorial para a variável 2 (V_2) no fator II não-rotacionado é determinada desenhando-se uma linha tracejada horizontalmente a partir do ponto do dado até o eixo vertical do fator II. De modo similar, uma linha vertical é tracejada a partir da variável 2 até o eixo horizontal do fator não-rotacionado I para determinar a carga da variável 2 no fator I. Um procedimento semelhante adotado para as outras variáveis determina as cargas fatoriais para as soluções não-rotacionadas e rotacionadas, como exibido na Tabela 3-1 para fins de comparação. No primeiro fator não-rotacionado, todas as variáveis têm carga alta. No segundo, as variáveis 1 e 2 são muito altas na direção positiva. A variável 5 é moderadamente alta na direção negativa, e as variáveis 3 e 4 têm cargas consideravelmente menores na direção negativa.

A partir da inspeção visual da Figura 3-7, é óbvio que há dois agrupamentos de variáveis. As variáveis 1 e 2 estão juntas, assim como as variáveis 3, 4 e 5. No entanto, tal padrão de variáveis não é tão óbvio a partir das cargas fatoriais não-rotacionadas. Rotacionando os eixos originais no sentido horário, como indicado na Figura 3-7, obtemos um padrão de cargas fatoriais completamente diferente. Observe que, rotacionando os fatores, os eixos são mantidos a 90 graus. Esse procedimento significa que os fatores são matematicamente independentes e que a rotação foi ortogonal. Após rotacionar os eixos fatoriais, as variáveis 3, 4 e 5 têm cargas altas no fator I, e as variáveis 1 e 2 têm cargas elevadas no fator II. Logo, o agrupamento ou padrão dessas variáveis em dois grupos é mais óbvio após a rotação, ainda que a posição ou configuração relativa das variáveis permaneça a mesma.

Os mesmos princípios gerais de rotações ortogonais são aplicáveis a rotações oblíquas. No entanto, o método de rotação oblíqua é mais flexível, pois os eixos fatoriais não precisam ser ortogonais. Além disso, é mais realista, porque as dimensões inerentes que são teoricamente importantes não são supostas sem correlações entre si. Na Figura 3-8, os dois métodos rotacionais são comparados. Note que a rotação fatorial oblíqua representa o agrupamento de variáveis com maior precisão. Essa precisão é um resultado do fato de que cada eixo fatorial rotacionado agora está mais próximo do respectivo grupo de variáveis. Além disso, a solução oblíqua fornece informações sobre o grau em que os fatores realmente estão correlacionados um com o outro.

Muitos pesquisadores concordam que a maioria das soluções diretas não-rotacionadas não é suficiente. Ou seja, na maioria dos casos, a rotação melhora a interpretação reduzindo algumas das ambigüidades que freqüentemente acompanham a análise preliminar. A principal opção disponível é escolher um método de rotação ortogonal ou oblíqua. A meta final de qualquer rotação é obter alguns fatores teoricamente significativos e, se possível, a estrutura fatorial mais simples. As rotações ortogonais são mais amplamente usadas porque todos os pacotes computacionais com análise fatorial contêm opções de rotação ortogonal, enquanto os métodos oblíquos não são tão difundidos. As rotações ortogonais também são utilizadas mais freqüentemente porque os procedimentos analíticos para rotações oblíquas não são tão bem desenvolvidos e ainda estão sujeitos a considerável controvérsia. Várias abordagens diferentes estão à disposição para a execução de rotações ortogonais ou oblíquas. Contudo, apenas um número limitado de procedimentos de rotação oblíqua está disponível na maioria dos pacotes estatísticos. Logo, o pesquisador provavelmente deverá aceitar o que lhe é fornecido.

Métodos rotacionais ortogonais. Na prática, o objetivo de todos os métodos de rotação é simplificar as linhas e colunas da matriz fatorial para facilitar a interpretação. Em uma matriz fatorial, as colunas representam fatores, e cada linha corresponde às cargas de uma variável ao longo dos fatores. Por simplificação das linhas, queremos dizer tornar o máximo de valores em cada linha tão próximos de zero quanto possível (isto é, maximizar a carga de uma variável em um único fator). Simplificação das colunas significa tornar o máximo de valores em cada coluna tão próximos de zero quanto possível (ou seja, tornar o número de cargas "elevadas" o menor possível). Três abordagens ortogonais principais foram desenvolvidas:

1. A meta final de uma rotação **QUARTIMAX** é simplificar as linhas de uma matriz fatorial; ou seja, QUARTIMAX se concentra em rotacionar o fator inicial de modo que uma variável tenha carga alta em um fator e cargas tão baixas quanto possível em todos os outros fatores. Nessas rotações, muitas variáveis podem ter carga alta no mesmo fator, pois a técnica se concentra em simplificar as linhas. O método QUARTIMAX não tem se mostrado bem-sucedido na produção de estruturas mais simples. Sua dificuldade é que ele tende a produzir um fator geral como o primeiro fator, no qual a maioria das variáveis, se não todas, tem cargas altas. Independentemente de qualquer conceito do que é uma estrutura "mais simples", ela inevitavelmente envolve lidar com agrupamentos de variáveis; um método que tende

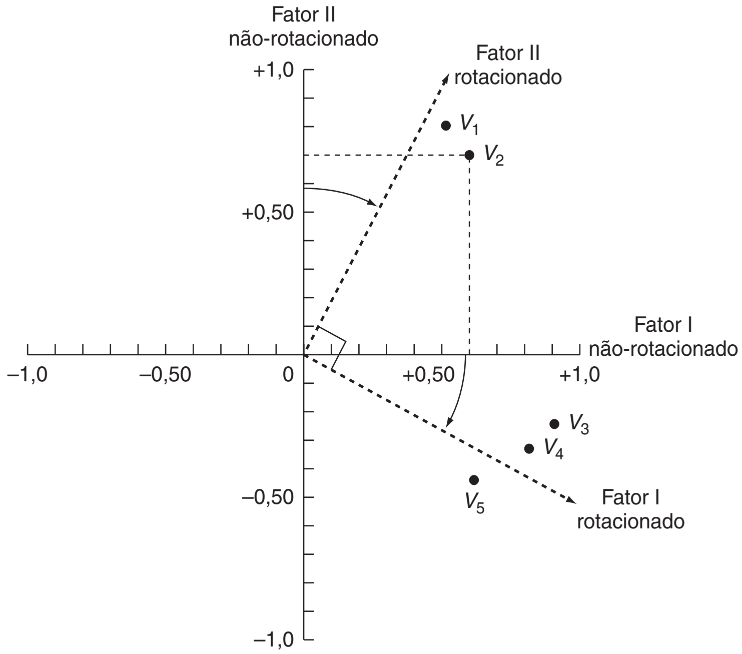

FIGURA 3-7 Rotação fatorial ortogonal.

a criar um grande fator geral (isto é, QUARTIMAX) não está de acordo com os propósitos de rotação.

2. Diferentemente de QUARTIMAX, o critério **VARIMAX** se concentra na simplificação das colunas da matriz fatorial. Com a abordagem rotacional VARIMAX, a simplificação máxima possível é conseguida se houver apenas 1s e 0s em uma coluna. Ou seja, o método VARIMAX maximiza a soma de variâncias de cargas exigidas da matriz fatorial. Lembre-se que, nas abordagens QUARTIMAX, muitas variáveis podem ter cargas altas ou próximas de altas no mesmo fator, pois a técnica se concentra em simplificar as linhas. Com a abordagem rotacional VARIMAX, há uma tendência para algumas cargas altas (isto é, próximas de –1 ou +1) e algumas cargas próximas de 0 em cada coluna da matriz. A lógica é que a interpretação é mais fácil quando as correlações variável-fator são (1) próximas de +1 ou –1, indicando assim uma clara associação positiva ou negativa entre a variável e o fator; ou (2) próximas de 0, apontando para uma clara falta de associa-

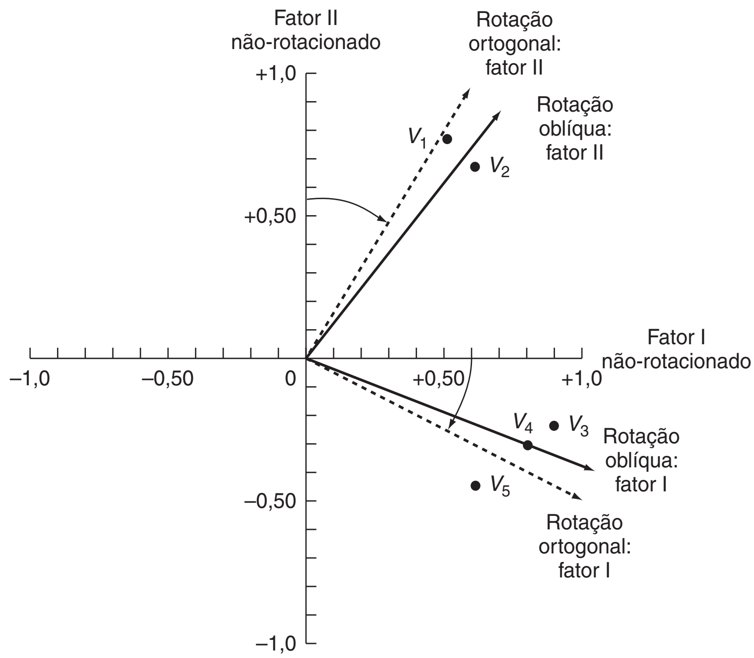

FIGURA 3-8 Rotação fatorial oblíqua.

TABELA 3-1 Comparação entre cargas fatoriais rotacionadas e não-rotacionadas

	Cargas fatoriais não-rotacionadas		*Cargas fatoriais rotacionadas*	
Variáveis	**I**	**II**	**I**	**II**
V_1	0,50	0,80	0,03	0,94
V_2	0,60	0,70	0,16	0,90
V_3	0,90	–0,25	0,95	0,24
V_4	0,80	–0,30	0,84	0,15
V_5	0,60	–0,50	0,76	–0,13

ção. Essa estrutura é fundamentalmente simples. Apesar de a solução QUARTIMAX ser analiticamente mais simples do que a VARIMAX, esta parece fornecer uma separação mais clara dos fatores. Em geral, o experimento de Kaiser [22, 23] indica que o padrão fatorial obtido por rotação VARIMAX tende a ser mais invariante do que o obtido pelo método QUARTIMAX quando diferentes subconjuntos de variáveis são analisados. O método VARIMAX tem sido muito bem-sucedido como uma abordagem analítica para a obtenção de uma rotação ortogonal de fatores.

3. O método **EQUIMAX** é uma espécie de meio-termo entre QUARTIMAX e VARIMAX. Em vez de se concentrar na simplificação de linhas ou de colunas, ele tenta atingir um pouco de cada. EQUIMAX não tem obtido ampla aceitação e é pouco usado.

Métodos de rotação oblíqua. As rotações oblíquas são semelhantes às ortogonais, porém as oblíquas permitem fatores correlacionados em vez de manterem independência entre os fatores rotacionados. Porém, enquanto há várias escolhas entre abordagens ortogonais, há apenas escolhas limitadas na maioria dos pacotes estatísticos para rotações oblíquas. Por exemplo, SPSS fornece OBLIMIN; SAS tem PROMAX e ORTHOBLIQUE; e BMDP fornece DQUART, DOBLIMIN e ORTHOBLIQUE. Os objetivos de simplificação são comparáveis aos métodos ortogonais, com a característica extra de fatores correlacionados. Com a possibilidade de fatores correlacionados, o pesquisador deve ter o cuidado extra de validar fatores rotacionados obliquamente, uma vez que eles têm uma maneira adicional (não-ortogonalidade) de se tornarem específicos à amostra e não-generalizáveis, particularmente com pequenas amostras ou pequenas proporções de casos por variáveis.

Seleção entre métodos rotacionais. Nenhuma regra específica foi desenvolvida para guiar o pesquisador na seleção de uma técnica rotacional ortogonal ou oblíqua em particular. Na maioria dos casos, o pesquisador simplesmente utiliza a técnica rotacional dada pelo programa de computador. A maioria dos programas tem como padrão de rotação o VARIMAX, mas todos os métodos rotacionais mais importantes estão amplamente disponíveis. No entanto, não há razão analítica para favorecer um método rotacional em detrimento de outro. A escolha de uma rotação ortogonal ou oblíqua deveria ser feita com base nas necessidades particulares de um dado problema de pesquisa. Para essa finalidade, diversas considerações (nas Regras Práticas 3-4) devem orientar a seleção do método rotacional.

Julgamento da significância de cargas fatoriais

Ao interpretar fatores, é preciso tomar uma decisão sobre quais cargas fatoriais vale a pena considerar. A discussão a seguir detalha questões relativas à significância prática e estatística, bem como ao número de variáveis, que afetam a interpretação de cargas fatoriais.

Garantia de significância prática. A primeira orientação não é baseada em qualquer proposição matemática, mas se refere mais à significância prática ao fazer um exame preliminar da matriz fatorial em termos das cargas fatoriais. Como uma carga fatorial é a correlação da variável e do fator, a carga ao quadrado é a quantia de variância total da variável explicada pelo fator. Assim, uma carga de 0,30 reflete aproximadamente 10% de explicação, e uma carga de 0,50 denota que 25% da variância é explicada pelo fator. A carga deve exceder 0,70 para que o fator explique 50% da variância de uma variável. Logo, quanto maior o valor absoluto da carga fatorial, mais importante a carga na interpretação da matriz fatorial. Usando significância prática como critério, podemos avaliar as cargas como se segue:

- Cargas fatoriais na faixa de ± 0,30 a ± 0,40 são consideradas como atendendo o nível mínimo para interpretação de estrutura.
- Cargas de ± 0,50 ou maiores são tidas como praticamente significantes.
- Cargas excedendo + 0,70* são consideradas indicativas de estrutura bem definida e são a meta de qualquer análise fatorial.

O pesquisador deve perceber que cargas extremamente altas (0,80 ou superiores) não são comuns e que a significância prática das cargas é um critério importante. Essas orientações são aplicáveis quando o tamanho da amostra é de 100 ou maior e onde a ênfase é a significância prática, e não estatística.

Avaliação da significância estatística. Como anteriormente observado, uma carga fatorial representa a correlação entre uma variável original e seu fator. Ao determinar um nível de significância para a interpretação de cargas, uma abordagem semelhante à determinação da significância

* N. de R. T.: O texto correto seria "±0,70".

REGRAS PRÁTICAS 3-4

Escolha de métodos de rotação fatorial

- Métodos de rotação ortogonal
  - São os mais empregados
  - São os métodos preferidos quando o objetivo da pesquisa é redução de dados a um número menor de variáveis ou a um conjunto de medidas não-correlacionadas para uso subseqüente em outras técnicas multivariadas
- Métodos de rotação oblíqua
  - São mais adequados ao objetivo de se obter diversos fatores ou construtos teoricamente relevantes, pois, realisticamente falando, poucos construtos no mundo são não-correlacionados

estatística de coeficientes de correlação poderia ser usada. Entretanto, pesquisas [14] têm demonstrado que as cargas fatoriais têm erros-padrão substancialmente maiores do que as correlações normais. Assim, as cargas fatoriais devem ser avaliadas em níveis consideravelmente mais restritos. O pesquisador pode empregar o conceito de poder estatístico discutido no Capítulo 1 para especificar cargas fatoriais consideradas significantes para diferentes tamanhos de amostra. Com o objetivo estabelecido de conseguir um nível de poder de 80%, o uso de um nível de significância de 0,05 e a inflação proposta dos erros padrão de cargas fatoriais, a Tabela 3-2 contém os tamanhos de amostra necessários para cada valor de carga fatorial ser considerado significante.

Por exemplo, em uma amostra de 100 respondentes, as cargas fatoriais de 0,55 ou mais são significantes. No entanto, em uma amostra de 50, é exigida uma carga fatorial de 0,75 para significância. Em comparação com a regra prática anterior, que denotava todas as cargas de 0,30 como tendo significância prática, essa abordagem consideraria cargas de 0,30 como significantes somente para amostras de 350 ou maiores.

Essas são orientações muito conservadoras quando comparadas com as da seção anterior ou mesmo com níveis estatísticos associados aos coeficientes de correlação convencionais. Assim, essas orientações devem ser usadas como ponto de partida na interpretação de cargas fatoriais, sendo as cargas menores consideradas significantes e acrescentadas à interpretação com base em outras considerações. A seção a seguir detalha o processo de interpretação, bem como o papel de outras considerações.

Ajustes baseados no número de variáveis. Uma desvantagem das duas abordagens anteriores é que o número de variáveis analisadas e o fator específico em exame não são considerados. Foi mostrado que quando o pesquisador se move do primeiro fator para fatores posteriores,

TABELA 3-2 Diretrizes para identificação de cargas fatoriais significantes com base em tamanho de amostra

Carga fatorial	Tamanho da amostra necessário para significância[a]
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

[a]Significância se baseia em um nível de significância (α) de 0,05, um nível de poder de 80%, e erros-padrão considerados como o dobro daqueles de coeficientes de correlação convencionais

Fonte: Cálculos feitos com SOLO Power Analysis, BMDP Statistical Software, Inc., 1993.

o nível aceitável para que uma carga seja julgada significante deve aumentar. O fato de que a variância única e a variância do erro começam a surgir em fatores posteriores significa que algum ajuste para cima no nível de significância deve ser incluído [22]. O número de variáveis em análise também é importante na decisão sobre quais cargas são significantes. À medida que o número de variáveis em análise aumenta, o nível aceitável para considerar uma carga significante diminui. O ajuste para o número de variáveis é cada vez mais importante à medida que se vai do primeiro fator extraído para fatores posteriores.

As Regras Práticas 3-5 resumem os critérios para significância prática ou estatística de cargas fatoriais.

REGRAS PRÁTICAS 3-5

Avaliação de cargas fatoriais

- Apesar de cargas fatoriais de ± 0,30 a ± 0,40 serem minimamente aceitáveis, valores maiores que ± 0,50 são geralmente considerados necessários para significância prática
- A ser considerado significante:
  - Uma carga menor com uma amostra maior ou um número maior de variáveis sob análise
  - Uma carga maior faz-se necessária com uma solução fatorial com um número maior de fatores, especialmente na avaliação de cargas em fatores posteriores
- Testes estatísticos de significância para cargas fatoriais são geralmente conservadores e devem ser considerados apenas como pontos de partida necessários para inclusão de uma variável para futura consideração

Interpretação de uma matriz fatorial

A tarefa de interpretar uma matriz de cargas fatoriais para identificar a estrutura entre as variáveis pode parecer à primeira vista muito complicada. O pesquisador deve classificar todas as cargas fatoriais (lembre-se, cada variável tem uma carga sobre cada fator) para identificar as mais indicativas da estrutura latente. Mesmo uma análise relativamente simples de 15 variáveis sobre 4 fatores precisa de avaliação e interpretação de 60 cargas fatoriais. Usando os critérios para interpretação de cargas descritos na seção anterior, o pesquisador descobre aquelas variáveis distintas para cada fator e procura uma correspondência com a fundamentação conceitual ou as expectativas administrativas depositadas na pesquisa para avaliar significância prática. Logo, interpretar as complexas relações representadas em uma matriz fatorial exige uma combinação da aplicação de critérios objetivos com julgamento gerencial. Seguindo-se o procedimento de cinco etapas delineado a seguir, o processo pode ser simplificado consideravelmente. Depois da discussão sobre o processo, um breve exemplo será usado para ilustrá-lo.

Etapa 1: Examine a matriz fatorial de cargas. A matriz de cargas fatoriais contém a carga fatorial de cada variável em cada fator. Elas podem ser cargas rotacionadas ou não-rotacionadas, mas, como anteriormente discutido, cargas rotacionadas são geralmente empregadas na interpretação fatorial a menos que a redução de dados seja o único objetivo. Tipicamente, os fatores são dispostos como colunas; assim, cada coluna de números representa as cargas de um único fator. Se uma rotação oblíqua foi usada, duas matrizes fatoriais de cargas são fornecidas. A primeira é a **matriz de padrão fatorial**, a qual tem cargas que representam a contribuição única de cada variável ao fator. A segunda é a **matriz de estrutura fatorial**, a qual tem correlações simples entre variáveis e fatores, mas essas cargas contêm tanto a variância única entre variáveis e fatores quanto a correlação entre fatores. À medida que a correlação entre fatores se torna maior, fica mais difícil distinguir quais variáveis têm cargas únicas em cada fator na matriz de estrutura fatorial. Logo, a maioria dos pesquisadores relata os resultados da matriz de padrão fatorial.

Etapa 2: Identifique a(s) carga(s) significante(s) para cada variável. A interpretação deve começar com a primeira variável no primeiro fator e se mover horizontalmente da esquerda para a direita, procurando a carga mais alta para aquela variável em qualquer fator. Quando a maior carga (em valor absoluto) é identificada, deve ser sublinhada se for significante como determinado pelos critérios anteriormente discutidos. A atenção agora se dirige para a segunda variável, e, novamente movendo-se horizontalmente da esquerda para a direita, procura-se a maior carga para aquela variável em qualquer fator, e a mesma deve ser sublinhada. Esse procedimento deve continuar para cada variável até que todas as variáveis tenham sido revistas quanto às suas maiores cargas em um fator.

Entretanto, a maioria das soluções fatoriais não resulta em uma estrutura simples (uma única carga alta para cada variável em um único fator). Logo, o pesquisador continuará, depois de sublinhar a carga mais alta de uma variável, a avaliar a matriz fatorial, sublinhando todas as cargas significantes para uma carga em todos os fatores. O processo de interpretação seria extremamente simplificado se cada variável tivesse apenas uma variável* significante. Na prática, no entanto, o pesquisador pode descobrir que uma ou mais variáveis tem cargas de tamanho moderado sobre diversos fatores, todas significantes, e o trabalho de interpretar fatores torna-se muito mais difícil. Quando uma variável demonstra ter mais de uma carga significante, ela é chamada de **carga cruzada**.

A dificuldade surge porque uma variável com diversas cargas significantes (uma carga cruzada) deve ser usada na rotulação de todos os fatores nos quais ela tem uma carga significante. No entanto, como os fatores podem ser distintos e potencialmente representar conceitos separados quando eles "compartilham" variáveis? Em última análise, o objetivo é minimizar o número de cargas significantes sobre cada linha da matriz fatorial (i.e., fazer com que cada variável se associe com um único fator). O pesquisador pode descobrir que diferentes métodos de rotação eliminam cargas cruzadas e, portanto, definem uma estrutura simples. Se uma variável persiste em ter cargas cruzadas, ela se torna candidata à eliminação.

Etapa 3: Avalie as comunalidades das variáveis. Uma vez que todas as cargas significantes tenham sido identificadas, o pesquisador deve procurar por variáveis que não sejam adequadamente explicadas pela solução fatorial. Uma abordagem simples é identificar variáveis nas quais faltam pelo menos uma carga significante. Outro método é examinar a comunalidade de cada variável, representando a quantia de variância explicada pela solução fatorial para cada variável. O pesquisador deve ver as comunalidades para avaliar se as variáveis atendem níveis aceitáveis de explicação. Por exemplo, um pesquisador pode especificar que pelo menos metade da variância de cada variável deve ser levada em conta. Usando essa diretriz, o pesquisador identificaria todas as variáveis com comunalidades menores que 0,50 como não tendo explicação suficiente.

Etapa 4: Reespecifique o modelo fatorial se neces-sário. Uma vez que todas as cargas significantes tenham sido identificadas e as comunalidades, examinadas, o pesquisador pode encontrar diversos problemas: (a) uma variável não tem cargas significantes; (b) mesmo com uma carga significante, a comunalidade de uma variável é considerada muito baixa, ou (c) uma variável tem uma carga cruzada. Nesta situação, o pesquisador pode executar qualquer combinação das seguintes ações corretivas, listadas da menos para a mais extrema:

* N. de R. T.: A frase correta seria "se cada variável tivesse apenas uma carga significante".

- *Ignorar aquelas variáveis problemáticas* e interpretar a solução como ela é, o que é apropriado se o objetivo é somente redução de dados, mas o pesquisador deve ainda observar que as variáveis em questão são pobremente representadas na solução fatorial.
- Avaliar cada uma daquelas variáveis para *possível eliminação*, dependendo da contribuição geral da variável para a pesquisa, bem como de seu índice de comunalidade. Se a variável é de menor importância para o objetivo do estudo ou tem um valor inaceitável de comunalidade, ela pode ser eliminada, e o modelo pode então ser reespecificado pela derivação de uma nova solução fatorial com aquelas variáveis eliminadas.
- Empregar um método alternativo de rotação, particularmente um método oblíquo, caso apenas métodos ortogonais tenham sido usados.
- Diminuir/aumentar o número de fatores mantidos para ver se uma estrutura fatorial menor/maior representará aquelas variáveis problemáticas.
- Modificar o tipo de modelo fatorial usado (componentes versus fatores comuns) para avaliar se mudanças do tipo de variância considerada afetam a estrutura fatorial.

Quaisquer que sejam as opções escolhidas pelo pesquisador, o objetivo final deve sempre ser a obtenção de uma estrutura fatorial com apoio tanto empírico quanto conceitual. Como vimos, muitos "truques" podem ser utilizados para melhorar a estrutura, mas a responsabilidade final está com o pesquisador e com a fundamentação conceitual subjacente à análise.

Etapa 5: Rotule os fatores. Quando é obtida uma solução fatorial aceitável na qual todas as variáveis têm uma carga significante em um fator, o pesquisador tenta designar algum significado para o padrão de cargas fatoriais. As variáveis com cargas mais altas são consideradas mais importantes e têm maior influência sobre o nome ou rótulo selecionado para representar um fator. Assim, o pesquisador examina todas as variáveis significantes para um fator particular e, enfatizando aquelas variáveis com maiores cargas, tenta designar um nome ou rótulo para um fator que reflita com precisão as variáveis com carga naquele fator. Os sinais são interpretados simplesmente como quaisquer outros coeficientes de correlação. Em cada fator, sinais concordantes significam que as variáveis estão positivamente relacionadas, e sinais opostos significam que as variáveis estão negativamente relacionadas. Em soluções ortogonais, os fatores são independentes uns dos outros. Portanto, os sinais para cargas fatoriais relacionam-se apenas com o fator no qual elas aparecem, e não com outros fatores na solução.

Esse rótulo não é determinado ou designado pelo programa computacional que realiza a análise fatorial; em vez disso, o rótulo é desenvolvido intuitivamente pelo pesquisador com base em sua adequação para representar as dimensões latentes de um fator particular. Segue-se esse procedimento para cada fator extraído. O resultado final será um nome ou rótulo que represente cada fator determinado da melhor maneira possível.

Como discutido anteriormente, a seleção de um número específico de fatores e o método de rotação são inter-relacionados. Várias tentativas adicionais de rotações podem ser executadas, e, comparando as interpretações fatoriais para diversas tentativas de rotações diferentes, o pesquisador pode selecionar o número de fatores a extrair. Em resumo, a habilidade de designar algum significado aos fatores, ou de interpretar a natureza das variáveis, se torna uma consideração extremamente importante ao se determinar o número de fatores a serem extraídos.

Um exemplo de interpretação fatorial. Para servir como ilustração de interpretação fatorial, nove medidas foram obtidas em um teste piloto baseado em uma amostra de 202 respondentes. Após a estimação dos resultados iniciais, análises posteriores indicaram que uma solução com três fatores era adequada. Logo, agora o pesquisador tem a tarefa de interpretar as cargas fatoriais das nove variáveis.

A Tabela 3-3 contém uma série de matrizes de cargas fatoriais. A primeira a ser considerada é a matriz fatorial não-rotacionada (parte a). Examinaremos as matrizes de cargas fatoriais não-rotacionadas e rotacionadas através do processo de cinco etapas anteriormente descrito.

Etapas 1 e 2: Examinar a matriz de cargas fatoriais e identificar cargas significantes. Dado o tamanho da amostra de 202, cargas fatoriais de 0,40 ou mais serão consideradas significantes para fins de interpretação. Usando esse padrão para as cargas fatoriais, podemos ver que a matriz não-rotacionada contribui pouco para se identificar qualquer forma de estrutura simples. Cinco das nove variáveis têm cargas cruzadas, e para muitas das outras variáveis as cargas significantes são relativamente baixas. Nesta situação, rotação pode melhorar nossa compreensão da relação entre as variáveis.

Como mostrado na Tabela 3-3b, a rotação VARIMAX melhora consideravelmente a estrutura de duas maneiras notáveis. Primeiro, as cargas são melhoradas para quase todas as variáveis, com as mesmas mais proximamente alinhadas ao objetivo de se ter uma elevada carga sobre um único fator. Segundo, agora somente uma variável (V_1) tem uma carga cruzada.

Etapa 3: Avaliar comunalidades. Apenas V_3 tem uma comunalidade que é baixa (0,299). Para nossos propósitos V_3 será mantida, mas um pesquisador pode considerar a eliminação de tais variáveis em outros contextos de pesquisa. Isso ilustra o caso em que uma variável tem uma carga significante, mas pode ainda ser pobremente explicada pela solução fatorial.

Etapa 4: Reespecificar o modelo fatorial se necessário. Se estabelecemos um valor de referência de 0,40 para significância de carga e novamente arranjamos as variáveis de acordo com cargas, emerge o padrão exibido na

(*Continua*)

(*Continuação*)

Tabela 3-3c. As variáveis V_7, V_9 e V_8 têm cargas elevadas sobre o fator 1, o fator 2 é caracterizado pelas variáveis V_5, V_2 e V_3, e o fator 3 tem duas características distintas (V_4 e V_6). Somente V_1 é problemática, com cargas significantes sobre os fatores 1 e 3. Sabendo que pelo menos duas variáveis são dadas sobre esses dois fatores, V_1 é eliminada da análise e as cargas são novamente calculadas.

Etapa 5: Rotular os fatores. Como mostrado na Tabela 3-3d, a estrutura fatorial para as oito variáveis remanescentes é agora muito bem definida, representando três grupos distintos de variáveis que o pesquisador pode agora utilizar em pesquisas posteriores.

Como se mostra no exemplo anterior, o processo de interpretação de fatores envolve julgamentos tanto objetivos quanto subjetivos. O pesquisador deve considerar uma vasta gama de questões o tempo todo, nunca perdendo de vista a meta final de definir a melhor estrutura do conjunto de variáveis. Apesar de muitos detalhes estarem envolvidos, alguns dos princípios gerais são encontrados nas Regras Práticas 3-6.

REGRAS PRÁTICAS 3-6

Interpretação dos fatores

- Existe uma estrutura ótima quando todas as variáveis têm cargas altas em um único fator
- Variáveis com carga cruzada (cargas elevadas sobre dois ou mais fatores) são geralmente eliminadas a menos que sejam teoricamente justificadas ou o objetivo seja apenas redução de dados.
- Variáveis em geral deveriam ter comunalidades maiores que 0,50 para serem mantidas na análise.
- Reespecificação de uma análise fatorial pode incluir opções como as que se seguem:
  - Eliminar uma ou mais variáveis
  - Mudar os métodos de rotação
  - Aumentar ou diminuir o número de fatores

Estágio 6: Validação da análise fatorial

O sexto estágio envolve a avaliação do grau de generalidade dos resultados para a população e da influência potencial de casos ou respondentes individuais sobre os resultados gerais. A questão da generalidade é crítica para todo método multivariado, mas é especialmente relevante nos métodos de interdependência, pois eles descrevem uma estrutura de dados que também deve ser representativa da população. No processo de validação, o pesquisador deve abordar várias questões na área de delineamento de pesquisa e características de dados, como discutido a seguir.

Uso de uma perspectiva confirmatória

O método mais direto para validar os resultados é partir para uma perspectiva confirmatória e avaliar a repetitividade dos resultados, seja com uma amostra particionada no conjunto de dados originais, seja com uma amostra separada. A comparação de dois ou mais resultados de um modelo fatorial sempre é problemática. No entanto, existem várias opções para realizar uma comparação objetiva. A emergência da análise fatorial confirmatória (CFA) por meio da modelagem de equações estruturais tem fornecido uma opção, mas geralmente é mais complicada e exige pacotes computacionais adicionais, como LISREL ou EQS [4,21]. Os Capítulos 10 e 11 discutem a análise fatorial confirmatória de forma mais detalhada. Além da CFA, diversos outros métodos têm sido propostos, variando de um simples índice de emparelhamento [10] a programas (FMATCH) projetados especificamente para avaliar a correspondência entre matrizes fatoriais [34]. Esses métodos têm tido uso esporádico, devido em parte (1) à sua percebida falta de sofisticação e (2) à indisponibilidade de softwares ou programas analíticos para automatizar as comparações. Assim, quando a CFA não é adequada, esses métodos fornecem alguma base objetiva para a comparação.

Avaliação da estabilidade da estrutura fatorial

Um outro aspecto da generalidade é a estabilidade dos resultados do modelo fatorial. A estabilidade fatorial depende principalmente do tamanho da amostra e do número de casos por variável. O pesquisador sempre é encorajado a obter a maior amostra possível e a desenvolver modelos parcimoniosos para aumentar a proporção casos-por-variáveis. Se o tamanho da amostra permite, o pesquisador pode querer particionar aleatoriamente a amostra em dois subconjuntos e estimar modelos fatoriais para cada um. A comparação das duas matrizes fatoriais resultantes fornecerá uma avaliação da robustez da solução ao longo das amostras.

Detecção de observações influentes

Além da generabilidade, uma outra questão importante para a validação da análise fatorial é a detecção de observações influentes. Discussões no Capítulo 2 sobre a identificação de observações atípicas, bem como no Capítulo 4 sobre as observações influentes em regressão, encontram aplicabilidade em análise fatorial. O pesquisador é encorajado a estimar o modelo com e sem observações identificadas como atípicas para avaliar seu impacto nos resultados. Se a omissão das observações atípicas é justificada, os resultados deveriam ter maior generalidade. Além disso, como discutido no Capítulo 4, diversas medidas de influência que refletem a posição de uma observação relativa a todas as outras (por exemplo, proporção de covariância) são igualmente aplicáveis à análise fatorial. Finalmente, a complexidade dos métodos propostos para identificação de observações influentes específicas à análise fatorial [11] limita a aplicação dos mesmos.

TABELA 3-3 Interpretação de uma matriz hipotética de cargas fatoriais

(a) Matriz de cargas fatoriais não-rotacionada

	Fator		
	1	2	3
V_1	0,611	0,250	–0,204
V_2	0,614	–0,446	0,264
V_3	0,295	–0,447	0,107
V_4	0,561	–0,176	–0,550
V_5	0,589	–0,467	0,314
V_6	0,630	–0,102	–0,285
V_7	0,498	0,611	0,160
V_8	0,310	0,300	0,649
V_9	0,492	0,597	–0,094

(b) Matriz VARIMAX de cargas fatoriais rotacionada

	Fator			
	1	2	3	*Comunalidade*
V_1	0,462	0,099	0,505	0,477
V_2	0,101	0,778	0,173	0,644
V_3	–0,134	0,517	0,114	0,299
V_4	–0,005	0,184	0,784	0,648
V_5	0,087	0,801	0,119	0,664
V_6	0,180	0,302	0,605	0,489
V_7	0,795	–0,032	0,120	0,647
V_8	0,623	0,293	–0,366	0,608
V_9	0,694	–0,147	0,323	0,608

(c) Matriz simplificada de cargas fatoriais rotacionada[1]

	Componente		
	1	2	3
V_7	0,795		
V_9	0,694		
V_8	0,623		
V_5		0,801	
V_2		0,778	
V_3		0,517	
V_4			0,784
V_6			0,605
V_1	0,462		0,505

[1] *Cargas menores que 0,40 não são exibidas, e variáveis são ordenadas pelas maiores cargas*

(d) Matriz de cargas fatoriais rotacionada com V_1 eliminada[2]

	Fator		
	1	2	3
V_2	0,807		
V_5	0,803		
V_3	0,524		
V_7		0,802	
V_9		0,686	
V_8		0,655	
V_4			0,851
V_6			0,717

[2] *V_1 eliminada da análise, cargas menores que 0,40 não são exibidas, e variáveis são ordenadas pelas maiores cargas*

Estágio 7: Usos adicionais dos resultados da análise fatorial

Dependendo dos objetivos da aplicação da análise fatorial, o pesquisador pode parar com a interpretação fatorial ou utilizar-se de um dos métodos para redução de dados. Se o objetivo é simplesmente identificar combinações lógicas de variáveis e entender melhor as inter-relações entre variáveis, então a interpretação fatorial basta. Isso fornece uma base empírica para julgar a estrutura das variáveis e o impacto dessa estrutura quando se interpretam os resultados a partir de outras técnicas multivariadas. Se o objetivo, porém, é identificar variáveis apropriadas para a aplicação subseqüente em outras técnicas estatísticas, então alguma forma de redução de dados será empregada. As duas opções incluem o seguinte:

- *Selecionar a variável com a maior carga fatorial* como uma representativa substituta para uma dimensão fatorial particular
- *Substituir o conjunto original de variáveis* por um conjunto menor e inteiramente novo, criado a partir de *escalas múltiplas* ou *escores fatoriais*.

Qualquer opção fornecerá novas variáveis para uso, por exemplo, como variáveis independentes em uma análise de regressão ou discriminante, variáveis dependentes em análise multivariada de variância, ou mesmo as variáveis de agrupamento em análise de agrupamentos. Discutimos cada uma dessas opções para redução de dados nas seções seguintes.

Seleção de variáveis substitutas para análise subseqüente

Se a meta do pesquisador é simplesmente identificar variáveis apropriadas para a aplicação subseqüente com outras técnicas estatísticas, o pesquisador tem a opção de examinar a matriz fatorial e selecionar a variável com a maior carga fatorial em cada fator para atuar como uma **variável**

substituta representativa daquele fator. Essa é uma abordagem simples e direta somente quando uma variável tem uma carga fatorial bem maior do que todas as demais. Em muitos casos, porém, o processo de seleção é mais difícil porque duas ou mais variáveis têm cargas significantes e bastante próximas umas das outras, ainda que apenas uma seja escolhida como representativa de uma dimensão em particular. Essa decisão deve ser baseada no conhecimento *a priori* que o pesquisador tem da teoria, que pode sugerir que uma variável, mais que as outras, seria logicamente representativa da dimensão. Além disso, o pesquisador pode ter conhecimento sugerindo que uma variável com carga levemente inferior é de fato mais confiável do que a variável com carga fatorial maior. Nesses casos, o pesquisador pode escolher a variável com carga ligeiramente menor como a melhor variável para representar um fator particular.

O método de selecionar uma única variável substituta como representativa do fator – apesar de ser simples e manter a variável original – tem várias desvantagens potenciais.

- *Não aborda a questão do erro de medida* encontrada quando se usam medidas únicas (ver a seção seguinte para uma discussão mais detalhada).
- Também corre-se o *risco de resultados potencialmente enganadores pela seleção de somente uma variável para representar um resultado que talvez seja mais complexo*. Por exemplo, suponha que variáveis que representem a competitividade de preço, a qualidade de produto e o valor fossem encontradas com elevadas cargas em um único fator. A seleção de uma dessas variáveis em separado criaria interpretações muito diferentes em qualquer análise posterior, ainda que todas possam estar tão intimamente relacionadas a ponto de tornar qualquer distinção definitiva impossível.

Em casos nos quais diversas cargas elevadas complicam a seleção de uma única variável, o pesquisador pode não ter escolha a não ser empregar a análise fatorial como a base para calcular uma escala múltipla ou escores fatoriais para uso como uma variável substituta. O objetivo, como no caso da seleção de uma única variável, é representar melhor a natureza básica do fator ou do componente.

Criação de escalas múltiplas

O Capítulo 1 introduziu o conceito de uma **escala múltipla**, a qual é formada pela combinação de diversas variáveis individuais em uma única **medida composta**. Em termos simples, todas as variáveis com cargas elevadas em um fator são combinadas, e o total – ou, mais comumente, o escore médio das variáveis – é usado como uma variável de substituição. Uma escala múltipla apresenta dois benefícios específicos.

- Fornece um *meio de superar consideravelmente o erro de medida* inerente em todas as variáveis medidas. **Erro de medida** é o grau em que os valores observados não são representativos dos valores "reais" devido a diversas razões, que variam de erros reais (p.ex., erros na entrada de dados) à falta de habilidade de indivíduos fornecerem informações precisas. O impacto do erro de medida é mascarar parcialmente relações (p.ex., correlações ou comparações de médias de grupos) e dificultar a estimação de modelos multivariados. A escala múltipla reduz o erro de medida usando **indicadores** (variáveis) múltiplos para reduzir a dependência de uma única resposta. Usando a resposta média ou típica de um conjunto de variáveis relacionadas, o erro de medida que poderia ocorrer em uma única questão será reduzido.
- Um segundo benefício da escala múltipla é sua *habilidade de representar os múltiplos aspectos de um conceito com uma medida única*. Muitas vezes, empregamos mais variáveis em nossos modelos multivariados como uma tentativa de representar as muitas facetas de um conceito que sabemos ser muito complexo. Entretanto, ao fazer isso, complicamos a interpretação dos resultados por causa da redundância nos itens associados ao conceito. Logo, gostaríamos de não apenas acomodar as descrições mais ricas de conceitos usando múltiplas variáveis, mas também de manter a parcimônia no número de variáveis em nossos modelos multivariados. A escala múltipla, quando corretamente construída, combina os múltiplos indicadores em uma só medida que representa o que acontece em comum no conjunto de medidas.

O processo de construção de escala tem fundamentos teóricos e empíricos em diversas disciplinas, incluindo a teoria psicométrica, a sociologia e o marketing. Apesar de um tratamento completo das técnicas e questões envolvidas estarem além do escopo deste livro, existem várias fontes excelentes para leitura complementar sobre esse assunto [2,12,20,30,31]. Além disso, há uma série de compilações de escalas existentes que podem ser aplicadas em várias situações [3,7,32]. Discutimos aqui, porém, quatro questões básicas para a construção de qualquer escala múltipla: definição conceitual, dimensionalidade, confiabilidade e validade.

Definição conceitual. O ponto de partida para criar qualquer escala múltipla é sua **definição conceitual**. A definição conceitual especifica a base teórica para a escala múltipla definindo o conceito a ser representado em termos aplicáveis ao contexto de pesquisa. Na pesquisa acadêmica, as definições teóricas são baseadas em pesquisa prévia que define o caráter e a natureza de um conceito. Em um contexto gerencial, conceitos específicos podem ser definidos de modo que se relacionem a objetivos propostos, como imagem, valor ou satisfação. Em qualquer caso, a criação de uma escala múltipla sempre é orientada pela definição conceitual, especificando o tipo e o caráter dos itens que são candidatos à inclusão na escala.

A **validade de conteúdo** é a avaliação da correspondência das variáveis a serem incluídas em uma escala múltipla e sua definição conceitual. Essa forma de validade, também conhecida como **validade de expressão**, avalia subjetivamente a correspondência entre os itens individuais e o conceito por meio de avaliações de especialistas, pré-testes com múltiplas subpopulações ou outros meios. O objetivo é garantir que a seleção de itens de escala aborde não apenas questões empíricas, mas também inclua considerações práticas e teóricas [12,31].

Dimensionalidade. Uma suposição inerente e exigência essencial para a criação de uma escala múltipla é que os itens sejam unidimensionais, significando que eles estão fortemente associados um com o outro e representam um só conceito [20,24]. A análise fatorial tem um papel essencial na realização de uma avaliação empírica da dimensionalidade de um conjunto de itens, pela determinação do número de fatores e das cargas de cada variável nos mesmos. O teste de unidimensionalidade significa que cada escala múltipla deve consistir de itens com cargas altas em um único fator [1,20,24,28]. Se uma escala múltipla é proposta como tendo múltiplas dimensões, cada dimensão deve ser refletida por um fator separado. O pesquisador pode avaliar unidimensionalidade com análise fatorial exploratória, como discutido neste capítulo, ou com análise fatorial confirmatória, como descrito nos Capítulos 10 e 11.

Confiabilidade. **Confiabilidade** é uma avaliação do grau de consistência entre múltiplas medidas de uma variável. Uma forma de confiabilidade é teste/reteste, pelo qual a consistência é medida entre as respostas para um indivíduo em dois pontos no tempo. O objetivo é garantir que as respostas não sejam muito variadas durante períodos de tempo, de modo que uma medida tomada em qualquer instante seja confiável. Uma segunda medida de confiabilidade, mais comumente usada, é a consistência interna, a qual avalia a consistência entre as variáveis em uma escala múltipla. A idéia da consistência interna é que os itens ou indicadores individuais da escala devem medir o mesmo construto, e assim serem altamente intercorrelacionados [12,28].

Como nenhum item isolado é uma medida perfeita de um conceito, devemos confiar em várias medidas diagnósticas para avaliar consistência interna.

- As primeiras *medidas que consideramos se relacionam a cada item separado*, incluindo a correlação item-com-total (a correlação do item com o escore da escala múltipla) e a correlação inter-itens (a correlação entre itens). Regras práticas sugerem que as correlações item-com-total excedam 0,50 e que as correlações inter-itens excedam 0,30 [31].
- O segundo tipo de medida diagnóstica é o *coeficiente de confiabilidade* que avalia a consistência da escala inteira, sendo o **alfa de Cronbach** [15,28,29] a medida mais amplamente usada. O limite inferior para o alfa de Cronbach geralmente aceito é de 0,70 [31,32], apesar de poder diminuir para 0,60 em pesquisa exploratória [31]. Uma questão na avaliação do alfa de Cronbach é sua relação positiva com o número de itens na escala. Como o aumento do número de itens, mesmo com grau igual de intercorrelação, aumenta o valor de confiabilidade, os pesquisadores devem fazer exigências mais severas para escalas com muitos itens.
- Também estão disponíveis as *medidas de confiabilidade determinadas a partir da análise fatorial confirmatória*. Incluídas nessas medidas estão a confiabilidade composta e a variância média extraída, discutidas mais detalhadamente no Capítulo 11.

Cada um dos principais programas estatísticos agora tem módulos ou programas de avaliação de confiabilidade, de modo que o pesquisador dispõe de uma análise completa tanto das medidas específicas de itens quanto de medidas gerais de confiabilidade. Qualquer escala múltipla deve ter sua confiabilidade analisada para garantir sua adequação antes de se proceder a uma avaliação de sua validade.

Validade. Após garantir que uma escala (1) está de acordo com sua definição conceitual, (2) é unidimensional e (3) atende aos níveis necessários de confiabilidade, o pesquisador deve fazer uma avaliação final: validade da escala. **Validade** é o grau em que uma escala ou um conjunto de medidas representa com precisão o conceito de interesse. Já vimos uma forma de validade – validade de conteúdo ou expressão – na discussão sobre definições conceituais. Outras formas de validade são medidas empiricamente pela correlação entre conjuntos de variáveis teoricamente definidos. As três formas mais amplamente aceitas de validade são a convergente, a discriminante e a nomológica [8,30].

- A validade convergente avalia o *grau em que duas medidas do mesmo conceito estão correlacionadas*. Neste ponto, o pesquisador pode procurar medidas alternativas de um conceito e então correlacioná-las com a escala múltipla. Correlações altas indicam que a escala está medindo seu conceito pretendido.
- A validade discriminante é o *grau em que dois conceitos similares são distintos*. O teste empírico é novamente a correlação entre medidas, mas dessa vez a escala múltipla está correlacionada com uma medida semelhante, mas conceitualmente distinta. Agora, a correlação deve ser baixa, demonstrando que a escala múltipla é suficientemente diferente do outro conceito semelhante.
- Finalmente, a validade nomológica refere-se ao *grau em que a escala múltipla faz previsões precisas de outros conceitos em um modelo teórico*. O pesquisador deve identificar relações teóricas a partir de pesquisa anterior ou de princípios aceitos e então avaliar se a escala tem relações correspondentes. Em resumo, a validade convergente confirma que a escala está correlacionada com outras medidas conhecidas do conceito; a validade discriminante garante que a escala é suficientemente diferente de outros conceitos semelhantes para ser distinta; e a validade nomológica determina se a escala demonstra as relações mostradas como existentes, com base em teoria ou pesquisa prévia.

Vários métodos para avaliar a validade estão disponíveis, variando de matrizes multitraço, multimétodo (MTMM) a abordagens de equações estruturais. Apesar de estar além do escopo deste texto, diversas fontes disponíveis abordam vários métodos e as questões envolvidas nas técnicas específicas [8,21,30].

Cálculo de escalas múltiplas. O cálculo de escalas múltiplas é um processo direto no qual os itens compreendendo a escala múltipla (i.e., os itens com cargas altas da análise fatorial) são somados ou têm suas médias calculadas. A abordagem mais comum é considerar a média dos itens na escala, o que fornece ao pesquisador um controle completo sobre o cálculo e facilita o uso em análises posteriores.

Sempre que variáveis têm cargas positivas e negativas dentro do mesmo fator, ou as variáveis com cargas positivas, ou aquelas com cargas negativas devem ter seus dados revertidos. Tipicamente, as variáveis com as cargas negativas são revertidas no escore, de modo que correlações e cargas são agora todas positivas no mesmo fator. **Escore reverso** é o processo pelo qual os valores dos dados para uma variável são revertidos de forma que suas correlações com outras variáveis são revertidas (i.e., passam de negativas para positivas). Por exemplo, em nossa escala de 0 a 10, revertemos o escore de uma variável subtraindo o valor original de 10 (ou seja, escore reverso = 10 – valor original). Desse modo, escores originais de 10 e 0 agora têm os valores revertidos de 0 e 10. Todas as características de distribuição são mantidas; apenas a distribuição é revertida.

A meta do escore reverso é prevenir um anulamento de variáveis com cargas positivas e negativas. Usemos um exemplo de duas variáveis com correlação negativa.

> Estamos interessados em combinar V_1 e V_2, com V_1 tendo carga positiva, e V_2, negativa. Se 10 é o escore máximo em V_1, o máximo em V_2 seria 0. Agora considere dois casos. No caso 1, V_1 tem um valor igual a 10 e V_2 tem valor 0 (o melhor caso). No segundo caso, V_1 tem um valor 0 e V_2 tem valor 10 (o pior caso). Se V_2 não é escore revertido, então o escore calculado pela soma das duas variáveis para ambos os casos é 10, mostrando nenhuma diferença, apesar de sabermos que o caso 1 é melhor e o 2 é o pior. Não obstante, se revertemos o escore V_2, a situação muda. Agora o caso 1 tem valores 10 e 10 em V_1 e V_2, respectivamente, e o caso 2 tem valores 0 e 0. Os escores de escala múltipla são agora 20 para o caso 1 e 0 para o caso 2, o que os distingue como a melhor e a pior situação.

Resumo. As escalas múltiplas, um dos desenvolvimentos recentes em pesquisa acadêmica, estão encontrando aplicação crescente em pesquisa aplicada e gerencial também. A habilidade da escala múltipla de representar conceitos complexos em uma única medida e ainda reduzir erros de medida a torna uma valiosa ferramenta em qualquer análise multivariada. A análise fatorial fornece ao pesquisador uma avaliação empírica das inter-relações entre variáveis, essencial na formação do fundamento conceitual e empírico de uma escala múltipla por meio da avaliação da validade de conteúdo e da dimensionalidade da escala (ver Regras Práticas 3-7).

Cálculo de escores fatoriais

A terceira opção para criar um conjunto menor de variáveis para substituir o conjunto original é o cálculo de **escores fatoriais**. Escores fatoriais também são medidas compostas de cada fator computadas para cada indivíduo. Conceitualmente, o escore fatorial representa o grau em que cada indivíduo tem escore elevado no grupo de itens que têm cargas elevadas em um fator. Assim, valores mais altos nas variáveis com cargas elevadas em um fator resultam em um escore fatorial superior. A característica-chave que diferencia um escore fatorial de uma escala múltipla é que o escore fatorial é computado com base nas cargas fatoriais de todas as variáveis no fator, enquanto a escala múltipla é calculada combinando-se apenas variáveis selecionadas. Portanto, apesar de o pesquisador ser capaz de caracterizar um fator pelas variáveis com as maiores cargas, ele também deve considerar as cargas das outras variáveis, embora menores, e sua influência no escore fatorial.

A maioria dos programas estatísticos computa facilmente escores fatoriais para cada respondente. Selecionando-se a opção de escore fatorial, esses escores são salvos para uso em análises posteriores. A desvantagem dos escores fatoriais é que eles não são facilmente repetidos em estudos, pois são baseados na matriz fatorial, a qual é determinada separadamente em cada estudo. A repetição da mesma matriz fatorial em estudos requer substancial programação computacional.

REGRAS PRÁTICAS 3-7

Escalas múltiplas

- Uma escala múltipla é apenas tão boa quanto os itens usados para representar o construto; ainda que possa passar em todos os testes empíricos, é inútil sem justificativa teórica
- Nunca crie uma escala múltipla sem primeiro avaliar sua unidimensionalidade com análise fatorial exploratória ou confirmatória
- Uma vez que uma escala é considerada unidimensional, seu escore de confiabilidade, medido pelo alfa de Cronbach:
  - Deve exceder uma referência de 0,70, apesar de um nível de 0,60 poder ser utilizado em pesquisa exploratória
  - Deve ter seu valor de referência aumentado à medida que o número de itens aumenta, especialmente quando o número de itens se aproxima de 10 ou mais
- Com a confiabilidade estabelecida, a validade deve ser avaliada em termos de:
  - Validade convergente – a escala se correlaciona com outras escalas semelhantes
  - Validade discriminante – a escala é suficientemente diferente de outras escalas relacionadas
  - Validade nomológica – a escala "prevê" como teoricamente sugerido

Seleção entre os três métodos

Para escolher entre as três opções de redução de dados, o pesquisador deve tomar várias decisões, ponderando as vantagens e desvantagens de cada abordagem com os objetivos da pesquisa. As diretrizes nas Regras Práticas 3-8 abordam as condições fundamentais associadas com cada método.

A regra de decisão é, portanto, a seguinte:

- Se dados são usados somente na amostra original ou se ortogonalidade deve ser mantida, escores fatoriais são adequados.
- Se generalidade ou capacidade de transferência são desejáveis, então escalas múltiplas ou variáveis substitutas são mais apropriadas. Se a escala múltipla é um instrumento bem construído, válido e confiável, então é provavelmente a melhor alternativa.
- Se a escala múltipla não é testada e é exploratória, com pouca ou nenhuma evidência de confiabilidade ou validade, variáveis substitutas deverão ser consideradas caso uma análise adicional não seja possível para melhorar a escala múltipla.

REGRAS PRÁTICAS 3-8

Representação da análise fatorial em outras análises

- **A variável substituta única**
  Vantagens
  - Simples de administrar e interpretar

  Desvantagens
  - Não representa todas as "facetas" de um fator
  - Suscetível a erro de medida
- **Escores fatoriais**
  Vantagens
  - Representam todas as variáveis com cargas naquele fator
  - Melhor método para completa redução de dados
  - São naturalmente ortogonais e podem evitar complicações provocadas por multicolinearidade

  Desvantagens
  - Interpretação mais difícil, pois todas as variáveis contribuem com as cargas
  - Difícil de repetir em estudos
- **Escalas múltiplas**
  Vantagens
  - Conciliação entre a variável substituta e opções de escore fatorial
  - Reduzem erro de medida
  - Representam múltiplas facetas de um conceito
  - Facilmente replicáveis em estudos

  Desvantagens
  - Incluem apenas as variáveis com cargas elevadas sobre o fator e excluem aquelas com impacto pequeno ou periférico
  - Não há necessariamente ortogonalidade
  - Exigem análise extensiva de questões de confiabilidade e validade

UM EXEMPLO ILUSTRATIVO

Nas seções anteriores, as questões mais importantes referentes à aplicação de análise fatorial foram discutidas dentro da estrutura de construção de modelos introduzida no Capítulo 1. Para melhor esclarecer esses tópicos, usamos um exemplo ilustrativo da aplicação de análise fatorial baseado em informações da base de dados apresentada no Capítulo 1. Nossa discussão do exemplo empírico também segue o processo de construção de modelo em seis estágios. Os três primeiros estágios, comuns à análise de componentes ou à análise de fatores comuns, são discutidos primeiramente. Em seguida, os estágios 4 a 6, de análise de componentes, serão discutidos, juntamente com exemplos do uso adicional de resultados fatoriais. Concluímos com um exame das diferenças em relação à análise de fatores comuns nos estágios 4 e 5.

Estágio 1: Objetivos da análise fatorial

A análise fatorial pode identificar a estrutura de um conjunto de variáveis, bem como fornecer um processo para a redução de dados. Em nosso exemplo, as percepções da HBAT sobre 13 atributos (X_6 a X_{18}) são examinadas pelos seguintes motivos:

- *Entender se essas percepções podem ser "agrupadas"*. Mesmo o número relativamente pequeno de percepções examinadas aqui apresenta um complexo quadro de 78 correlações distintas. Agrupando as percepções, a HBAT será capaz de exibir o quadro geral em termos de compreensão de seus clientes e o que os mesmos pensam sobre a HBAT.
- *Reduzir as 13 variáveis a um número menor*. Se as 13 variáveis podem ser representadas em um número menor de variáveis compostas, então as outras técnicas multivariadas podem se tornar mais parcimoniosas. É claro que essa abordagem considera que exista algum arranjo latente nos dados em análise.

Qualquer um ou ambos os objetivos podem ser encontrados em uma questão de pesquisa, tornando a análise fatorial aplicável a uma vasta gama de questões. Além disso, como a base para o desenvolvimento de escalas múltiplas, ela tem conquistado cada vez maior uso nos últimos anos.

Estágio 2: Planejamento de uma análise fatorial

Compreender a estrutura das percepções de variáveis requer análise fatorial do tipo *R* e uma matriz de correlações entre variáveis, não respondentes. Todas as variáveis são métricas e constituem um conjunto homogêneo de percepções adequado à análise fatorial.