

INTRODUÇÃO ÀS TÉCNICAS DE ANÁLISE FATORIAL E CONJUNTA DE DADOS

Henrique Martins Rocha

INTRODUÇÃO

Os estudos estatísticos ganham cada vez mais importância no mundo atual: análises financeiras, aspectos relacionados à concessão de crédito, análise da saúde financeira das empresas e análise de investimentos são algumas das áreas altamente dependentes da Estatística. Também nos estudos econômicos, com o uso dos modelos econométricos, e na análise de riscos de negócios em geral, a influência da Estatística é cada vez maior, fazendo parte do dia a dia dos profissionais.

Nos últimos anos, contudo, um novo fator veio à tona: a evolução das tecnologias da informação e comunicação, bem como a expansão da internet e das tecnologias de armazenagem de dados fizeram com que o volume de informações gerado e processado a cada momento alcançasse patamares nunca antes vistos. Por conta disso, o volume de dados nos estudos estatísticos também cresceu, e os métodos tradicionais se mostraram incapazes de lidar com tanta informação ao mesmo tempo. Surgiram então as técnicas de análise multivariada: ferramentas estatísticas que lidam com um grande volume de dados, oriundos de fontes diversas e com formato variado.

As informações a serem extraídas da massa de dados também passaram a ser mais complexas: hoje pequenas nuances ou perturbações nos dados podem significar o início de fenômenos de grande abrangência, exigindo rápida compreensão dos gestores para dar suporte às decisões de negócio a serem tomadas. A era da comunicação exige, portanto, que os profissionais conheçam os fundamentos das técnicas de análise multivariada, de forma a estarem habilitados a acompanhar o ritmo evolutivo e serem capazes de prover respostas rápidas e assertivas a novos desafios que possam surgir.

Considerando esse contexto, na apostila **Técnicas da análise multivariada de dados**, conheceremos as ferramentas de análise multivariada existentes, os seus desafios e implicações, bem como as suas formas de aplicação. Com isso, o nosso objetivo é oferecer reflexões e possibilidades de aplicação de técnicas de análise multivariada para suportar o processo decisório nas organizações, relacionando tais técnicas a informações e decisões no ambiente de negócios. Para tanto, esta apostila foi estruturada em dois módulos.

No módulo 1, trataremos da análise fatorial, uma técnica de análise multivariada que permite identificar o comportamento de grupos de variáveis ou de consumidores. Com isso, é possível extrair conclusões de um grande volume de informações e direcionar tanto os recursos e quanto os esforços das organizações aos aspectos de maior relevância para o negócio. Estudaremos ainda a análise conjunta, uma ferramenta de grande relevância no âmbito gerencial que permite prever a reação de consumidores a determinadas características de produtos e serviços.

SUMÁRIO

ANÁLISE FATORIAL E ANÁLISE CONJUNTA.....	7
INTRODUÇÃO	7
ANÁLISE FATORIAL.....	8
Objetivos da análise fatorial.....	9
Análises fatoriais R e Q	12
Tipos de saída geradas pela análise fatorial R	14
Planejamento e suposições em análise fatorial	15
Determinação de fatores, avaliação e validação	16
Métodos de extração de dados	17
Quantidade de fatores a ser selecionada	18
ANÁLISE CONJUNTA.....	19
Objetivos da análise conjunta	20
Projeto e suposições em análise conjunta	23
Autocorrelação entre fatores.....	32
Estimação e avaliação do modelo	39
BIBLIOGRAFIA	41
PROFESSOR-AUTOR.....	45



ANÁLISE FATORIAL E ANÁLISE CONJUNTA

Neste módulo, trataremos da análise fatorial, uma técnica de análise multivariada que permite identificar o comportamento de grupos de variáveis ou de consumidores. Com isso, é possível extrair conclusões de um grande volume de informações e direcionar tanto os recursos e quanto os esforços das organizações aos aspectos de maior relevância para o negócio. Estudaremos ainda a análise conjunta, uma ferramenta de grande relevância no âmbito gerencial que permite prever a reação de consumidores a determinadas características de produtos e serviços.

Introdução

As técnicas de análise multivariada têm como característica comum o fato de lidarem com inúmeras variáveis simultaneamente, possibilitando conclusões sobre elas e sobre os fenômenos a elas relacionados. Algumas técnicas têm como propósito esmiuçar os efeitos de cada uma das variáveis envolvidas, como a análise de regressão múltipla, em que são identificadas as variáveis independentes que afetam uma variável dependente bem como o tamanho, em termos quantitativos, dessa influência.

Outras ferramentas da análise multivariada seguem uma linha distinta: pela dificuldade de lidar com inúmeras variáveis simultaneamente, a análise fatorial, uma das ferramentas de análise multivariada mais conhecidas, identifica similaridades entre as variáveis, de forma que elas possam ser analisadas “em conjunto”. Isso permite que decisões referentes à análise e à gestão de negócios possam ocorrer com base no comportamento de grupos de variáveis: os denominados fatores, que servem de inspiração para o nome da técnica de análise fatorial.

A análise fatorial pode ser aplicada, por exemplo, por um prestador de serviços que coletou dados referentes à satisfação/insatisfação dos seus clientes. Devido à diversidade de escopo, as reclamações podem apresentar um extenso espectro, com dezenas (ou mesmo centenas) de queixas diferentes, o que tornaria a análise dos impactos de cada tipo de reclamação e, conseqüentemente, o processo de gestão e melhoria de desempenho muito difícil. Nesse caso, a análise fatorial permitiria identificar as reclamações semelhantes, de tal forma que cada grupo de reclamações fosse tratado como um único fator, facilitando a compreensão das relações de causa e efeito entre as reclamações e o desempenho do negócio, e permitindo tomadas de decisão mais assertivas por parte dos gestores.

Outra técnica multivariada interessante e de grande potencial para auxiliar tanto na compreensão de fenômenos quanto na tomada de decisão gerencial é a análise conjunta, que permite analisar efeitos de variáveis independentes sobre preferências e reações de consumidores quanto a produtos, serviços e ideias, mesmo que essas variáveis sejam de natureza qualitativa.

A seguir, vamos compreender o funcionamento e a aplicação de alguns desses instrumentos.

Análise fatorial

O desenvolvimento das análises multivariadas trouxe oportunidades nunca vislumbradas nos âmbitos da compreensão de fenômenos, das análises de negócios e do suporte à tomada de decisão: pela primeira vez, pudemos analisar inúmeras variáveis simultaneamente, ao invés de ficarmos restritos às análises *ceteris paribus*, em que só conseguimos analisar o comportamento de uma variável se, artificialmente, assumirmos que as demais permanecem inalteradas.

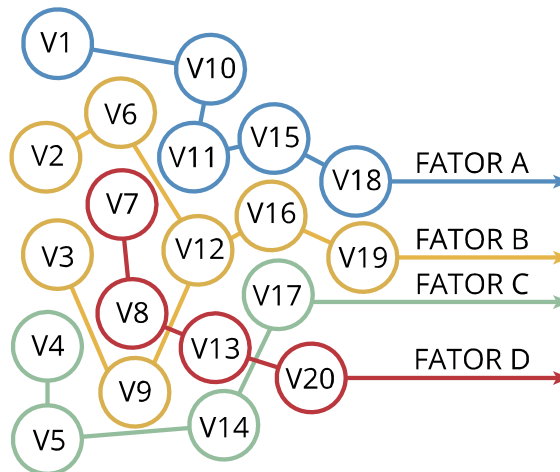
No entanto, tais avanços trouxeram também problemas: o grande volume de informações, que torna as análises cada vez mais difíceis e complexas. Por exemplo, a utilização de recursos visuais para identificar e analisar fenômenos diversos torna-se totalmente inefetiva quando lidamos com dezenas de variáveis. Além disso, a análise das relações entre as variáveis torna-se cada vez mais complexa: além da possibilidade de haver interações entre elas (fenômeno que não consegue ser detectado quando estudamos as variáveis aos pares), a possibilidade de haver autocorrelação entre variáveis que deveriam ser independentes cria, nas análises multivariadas, vieses que comprometem diversos métodos, como a análise de regressão múltipla.

Curiosamente, essa é característica de que a técnica da análise fatorial se aproveita para estudar fenômenos diversos e explicar os padrões de relações complexas multidimensionais entre fatores: a correlação entre variáveis. A técnica “pode ser utilizada para examinar os padrões ou relações latentes em um grande número de variáveis e determinar se a informação pode ser condensada ou resumida a um conjunto menor de fatores ou componentes” (HAIR *et al.*, 2009, p. 100). Para tanto, agrupa variáveis altamente correlacionadas, o que nos permite estudar as características e comportamento dos grupos.

Objetivos da análise fatorial

De acordo com Fávero e Belfiore (2015), a técnica de análise fatorial trabalha, na sua forma exploratória, com variáveis que apresentam entre si elevados coeficientes de correlação, de tal forma que seja possível prever o comportamento do conjunto de variáveis originais. Como afirmam Hair *et al.* (2009, p. 102), “esses grupos de variáveis (fatores) [...] são considerados representantes de dimensões dentro dos dados”, como demonstrado na figura a seguir.

Figura 1 – Agrupamento de variáveis e identificação de fatores



Fonte: ABG (2017a).

É possível, dessa forma, conciliar três aspectos conflitantes:

- a busca por significância estatística na identificação das variáveis relevantes, ou seja, variáveis que consigam explicar determinado fenômeno;
- a realização de previsões de valor sobre o fenômeno a partir de tais variáveis, com a menor perda possível de informação e
- a capacidade de lidar com tantas variáveis e os seus efeitos simultaneamente, por meio da identificação de uma quantidade relativamente pequena de fatores que representem o comportamento conjunto de variáveis originais interdependentes.

As implicações gerenciais de tal análise estariam na possibilidade de avaliar características que, de outra forma, estariam separadas, o que, por uma questão de tempo e custo, tornaria proibitivo o desenvolvimento de planos de ação para cada uma delas, seja no sentido de minimizar impactos negativos, seja para alavancar vantagens competitivas das organizações.

“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?” (HAIR *et al.*, 2009, p. 103).

Um exemplo de utilização da análise fatorial pode ser encontrado no trabalho de Moraes e Abiko (2006), que pesquisaram a satisfação e/ou percepção de moradores quanto a unidades domiciliares de programas habitacionais do Governo e a moradias construídas pela própria população. De acordo com os autores, tal tipo de avaliação é fundamental para identificar pontos positivos e negativos de um projeto. No entanto, muitas análises têm sido restritas à verificação de frequências, diagramas de Pareto, tabelas de contingência e, quando aplicáveis, a medidas de posição e de variabilidade, que dão uma visão isolada de cada variável, “não contribuindo para o entendimento das relações de interdependência entre elas” (MORAES; ABIKO, 2006, p. 1.233).

Os autores analisaram, sob a ótica dos moradores, o nível de satisfação sobre as seguintes variáveis:

- tamanho da sala, dos quartos, da cozinha e do banheiro;
- qualidade do material do piso, das paredes, do teto, das portas e janelas, das instalações de água, de esgoto e elétricas;
- conforto, devido à temperatura no verão e no inverno, à ventilação na sala, nos quartos, na cozinha e nos banheiros, e à iluminação e ao ruído externo;
- segurança estrutural, no período de chuva e contra roubo, e
- resposta aos problemas de umidade, de infiltração e de odores.

Após o processamento e a análise dos dados, os pesquisadores detectaram que os seguintes fatores seriam importantes para um projeto de construção de imóvel domiciliar:

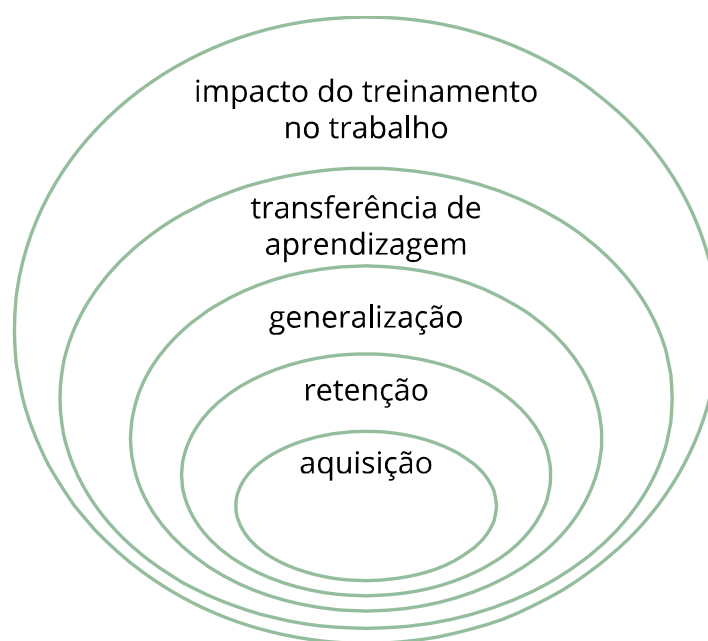
- segurança contra roubo;
- qualidade das instalações hidrossanitárias;
- cozinha;
- conforto no verão e iluminação;
- adequação e qualidade do piso;
- estrutura;
- cobertura;
- conforto em relação ao ruído externo e
- tamanho do banheiro.

Como implicações, além de identificarem os fatores críticos de sucesso para um projeto de construção de imóvel domiciliar, eles perceberam que ambientes como cozinha e banheiro são avaliados de forma isolada pelos moradores. Além disso, itens que poderiam ser vistos de forma isolada mostraram-se inter-relacionados, como o fator “segurança contra roubo e materiais das portas”.

É importante destacarmos também a existência da abordagem confirmatória da análise fatorial. Nesse caso, a existência de um referencial teórico ou de informações sobre pesquisas anteriores auxiliam-nos a identificar, previamente, quais variáveis determinam quais fatores. Dessa forma, buscamos confirmar o grau de ajuste dos dados observados à teoria existente.

Em tal abordagem, as relações entre variáveis observadas e fatores são modeladas na forma de equações estruturais, fazendo uso de uma série de regressões lineares. Essa abordagem foi utilizada por Pilati e Abbad (2005) para testar o modelo conceitual de impacto do treinamento no trabalho, mostrado na figura a seguir, em que cada elo de eventos depende de conjuntos diferentes de variáveis para ocorrer, de acordo com os padrões desejáveis.

Figura 2 – Modelo conceitual de Impacto do treinamento no trabalho e construtos correlatos



Fonte: Pilati e Abbad (2005, p. 44).

Os resultados das análises sugeriram reespecificações do modelo hipotético, ainda que a proposta conceitual de impacto do treinamento no trabalho tenha sido corroborada, relacionando “o resultado do treinamento no trabalho e questões substantivas do desempenho humano, auxiliando na compreensão do comportamento humano no trabalho e a função exercida pelos sistemas organizacionais” (PILATI; ABBAD, 2005, p. 50).

Análises fatoriais R e Q

Como pudemos perceber, o uso da análise fatorial tem como ponto de partida um problema de pesquisa claramente definido. Nesse contexto, de acordo com Hair *et al.* (2014), um aspecto relevante está na especificação da unidade de análise. Isso quer dizer que precisamos determinar se queremos identificar relações entre:

- **variáveis** – caso em que deve ser utilizado um tipo mais comum de análise fatorial, denominada análise fatorial R, na qual aplicamos uma matriz de correlação (uma matriz quadrada cujos elementos são as correlações entre as variáveis analisadas) ou
- **respondentes** – caso em que deve ser aplicada a análise fatorial Q (menos utilizada), que se baseia em algum tipo de análise de agrupamento, isto é, em técnicas para verificar a existência de comportamentos semelhantes entre indivíduos, empresas, municípios, países, etc. em relação a determinadas variáveis, identificando grupos homogêneos internamente e heterogêneos entre si.

Hair *et al.* (2014) apresentam um exemplo bastante elucidativo sobre as diferenças entre as análises fatoriais R e Q, considerando os resultados de quatro respondentes para três variáveis diferentes, como demonstrado a seguir.

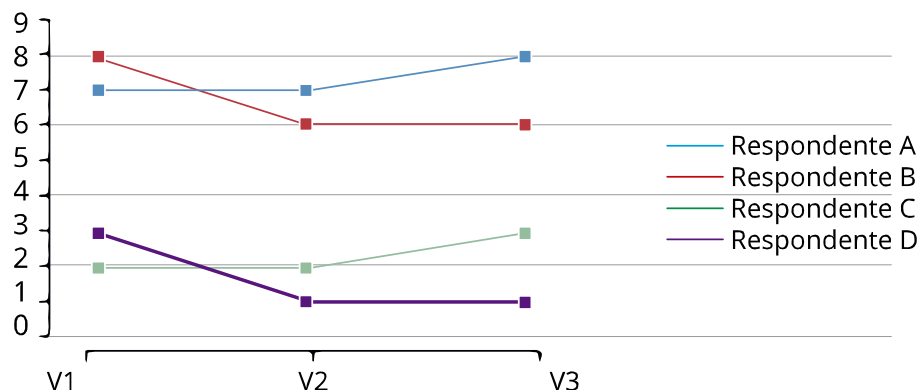
Tabela 1 – Resultado das respostas

respondente	V1	V2	V3
A	7	7	8
B	8	6	6
C	2	2	3
D	3	1	1

Fonte: Adaptado de Hair *et al.* (2014).

Colocando os resultados em um gráfico, para melhor visualização dos fenômenos, temos:

Figura 3 – Respostas



Agora, vejamos o *script* em R:

```
#Criando vetores com dados da tabela
```

```
respondenteA<-c(7,7,8)
```

```
respondenteB<-c(8,6,6)
```

```
respondenteC<-c(2,2,3)
```

```
respondenteD<-c(3,1,1)
```

```
#Criando gráfico em branco inicial com linhas de grade horizontais
```

```
plot(c(1,2,3), axes=FALSE, type='n', xlab = "", ylab = "", ylim = c(0,9), xlim = c(1,5),  
main="Respostas", panel.first=axis(side=2, at=seq(0,9,by=1), grid(0,NULL,lty=1), las=1))
```

```
#Adicionando linha do respondente A
```

```
par(new=TRUE) #Adicionar plotagem ao gráfico anterior
```

```
plot(c(1,2,3), respondenteA, type='b', pch=15, col="blue", axes=FALSE, xlab = "", ylab = "", ylim =  
c(0,9), xlim = c(1,5))
```

```
#Adicionando linha do respondente B
```

```
par(new=TRUE) #Adicionar plotagem ao gráfico anterior
```

```
plot(c(1,2,3), respondenteB, type='b', pch=15, col="red", axes=FALSE, xlab = "", ylab = "", ylim =  
c(0,9), xlim = c(1,5))
```

```

#Adicionando linha do respondente C
par(new=TRUE) #Adicionar plotagem ao gráfico anterior
plot(c(1,2,3), respondenteC, type='b', pch=15, col="green", axes=FALSE, xlab = "", ylab = "", ylim =
c(0,9), xlim = c(1,5))

#Adicionando linha do respondente D
par(new=TRUE) #Adicionar plotagem ao gráfico anterior
plot(c(1,2,3), respondenteD, type='b', pch=15, col="purple", axes=FALSE, xlab = "", ylab = "", ylim
= c(0,9), xlim = c(1,5))

#Definindo valores presentes na escala do eixo x
axis(side=1,at=c(1,2,3),labels=c("V1","V2","V3"))

#Criando legenda para o gráfico com uma coluna
legend("right", ncol=1, lwd=2, box.col = "white",pch=15, legend = c ("Respondente A",
"Respondente B", "Respondente C", "Respondente D"), col=c ("blue", "red", "green", "purple"))

```

Uma análise fatorial do tipo R, por considerar as distâncias entre os resultados (respostas) dos respondentes, alocaria os respondentes A e B – com valores mais altos – em um grupo, e os respondentes C e D em outro. Já uma análise fatorial do tipo Q detectaria dois grupos com covariância semelhantes: um com os respondentes A e C, e outro com os respondentes B e D.

Tipos de saída geradas pela análise fatorial R

A análise fatorial R pode gerar dois tipos de saída:

a) Resumo de dados:

Dimensões que, quando interpretadas e compreendidas, descrevem os dados com uma quantidade muito menor de elementos do que a contida nas variáveis individuais originais.

Nesse resumo, o pesquisador pode visualizar os dados em diferentes níveis: desde as variáveis individuais até os agrupamentos mais generalizados (representando, coletivamente, as variáveis), formados para explicar o conjunto inteiro de variáveis. De acordo com Hair *et al.* (2009, p. 106). “cada uma das variáveis observadas (originais) é uma variável dependente, [...] função de alguns conjuntos inerentes e latentes de fatores (dimensões), [...] compostos por todas as outras variáveis.”

b) Redução dos dados:

Deriva um valor empírico para cada fator substituindo o valor original por esse novo valor, criando um *ranking* de desempenho. Isso é feito pela identificação das variáveis representativas em um conjunto de variáveis ou pela criação de um conjunto novo de variáveis que substitui (parcial ou integralmente) o conjunto original de variáveis: o foco está, portanto, na identificação das contribuições de cada variável para os fatores.

Planejamento e suposições em análise fatorial

Independentemente de a análise fatorial ser usada para redução ou resumo dos dados, a seleção das variáveis que integrarão o modelo deve ser rigorosa e pautada no conhecimento dos fenômenos envolvidos por parte do pesquisador.

Por exemplo, na análise da atratividade de uma loja na percepção dos consumidores, a não inclusão de uma variável relacionada aos atendentes criará uma séria lacuna no estudo. Por outro lado, a inclusão indiscriminada de variáveis, na expectativa de que a técnica consiga “separar o joio do trigo”, fadará o estudo ao insucesso, visto que a análise fatorial sempre identificará fatores que podem estar contaminados por variáveis estranhas à população.

“A qualidade e o significado dos fatores obtidos refletem as bases conceituais das variáveis incluídas na análise” (HAIR *et al.*, 2009, p. 106).

Segundo Hair *et al.* (2014), o planejamento de uma análise fatorial envolve três passos:

1. Calcular os dados de entrada:

Isso deve ser feito na forma de uma matriz de correlação, de forma a atender aos objetivos de agrupamento, isto é, variáveis (análise tipo R) ou respondentes (análise tipo Q).

2. Estabelecer a quantidade de variáveis, as propriedades das medidas e os tipos de variável:

Para tanto, deve-se dar preferência a variáveis métricas, facilitando os cálculos de correlação. Caso alguma variável não métrica seja incluída, é aconselhável que seja uma variável dicotômica (codificada como 0 ou 1).

Em análises exploratórias, para análise de fatores, devem ser identificadas variáveis-chave que reflitam fatores latentes teoricamente previstos. Já em análises confirmatórias, devem ser incluídas cinco ou mais variáveis que possam representar cada fator proposto (essa técnica é de pouco uso na identificação de fatores compostos por uma única variável).

3. Estabelecer o tamanho da amostra em termos absolutos e como função da quantidade de variáveis da análise:

Nesse caso, como regra geral, é recomendado que o tamanho da mostra seja igual ou maior a 100. Com 50 ou menos observações, a capacidade de executar análises fica seriamente comprometida.

Recomenda-se também que a quantidade de observações seja, pelo menos, cinco vezes maior que a quantidade de variáveis (se possível, 10 vezes). No entanto, é importante observarmos que o aumento na quantidade de variáveis leva ao aumento na quantidade de correlações a serem calculadas. Consequentemente, a probabilidade de “falsas” correlações, oriundas do acaso, serem identificadas também aumenta. Hair *et al.* (2014) afirmam que uma análise com 30 variáveis e $\alpha = 0,05$ poderia apresentar cerca de 20 correlações falsas.

Quanto às suposições na análise fatorial, a compreensão dos fenômenos e das próprias variáveis por parte do pesquisador tem um grande peso. Hair *et al.* (2009, p.109) alertam que “a presença de variáveis correlacionadas e a subsequente definição de fatores não garante relevância, mesmo que elas satisfaçam as exigências estatísticas.”

Por exemplo, misturar variáveis sabidamente dependentes e independentes, e extrair pseudocorrelações entre as variáveis é algo inadequado. Da mesma forma, utilizar amostras não homogêneas (por exemplo, amostras com pessoas jovens e idosas para um conjunto de itens relacionados à força física) fará com que as correlações resultantes e a estrutura de fatores sejam questionáveis. Em situações como essa, é recomendável que sejam executadas análises separadas para cada grupo.

Sobre as suposições de cunho estatístico, aplicam-se as de normalidade, homocedasticidade e linearidade, visto que elas afetam as correlações, aspecto principal e necessário das análises fatoriais. Se todas as correlações forem baixas ou forem de mesma ordem de grandeza, isso significa que não há qualquer estrutura para agrupar as variáveis e, consequentemente, a análise fatorial não se aplica como técnica útil. Hair *et al.* (2014) estabelecem que uma proporção significativa de correlações seja superior a 0,30 para que a análise fatorial seja aplicada.

Determinação de fatores, avaliação e validação

Especificadas as variáveis e preparada a matriz de correlação, os próximos passos envolvem a decisão quanto:

- ao método de extração de dados – análise de componentes principais ou análise de fatores comuns) e
- à quantidade de fatores a ser selecionada para explicar a estrutura latente de dados.

Métodos de extração de dados

Vejamos, com detalhes, cada um dos métodos de extração de dados:

a) Análise de componentes principais:

A análise de componentes principais busca resumir a maior parte da informação original (a variância total) a um mínimo de fatores, derivando fatores com pequenas proporções de variância específica (também chamada de variância única, que é a associada a apenas uma variável, não sendo explicada pelas correlações com outras variáveis) para fins de previsão. Nela, de acordo com Manly e Alberto (2019), tomamos as variáveis e as combinamos para produzir índices não correlacionados que descrevem a variação dos dados.

Segundo Hair *et al.* (2009, p. 112), essa é a técnica mais adequada quando:

- a 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
- o conhecimento anterior sugere que a variância específica e de erro representam proporção pequena da variância total.

b) Análise de fatores comuns:

A análise de fatores comuns tem como foco a identificação de fatores ou dimensões latentes que refletem o que as variáveis têm em comum. Considera somente a denominada variância comum, aquela explicada pelas correlações com as demais variáveis analisadas.

Importante

O coeficiente de determinação (calculado como o quadrado do coeficiente de correlação entre duas variáveis) representa a proporção da variância comum entre as variáveis. Dessa forma, se a correlação entre duas variáveis é 0,60, cada variável compartilha 36% (0.60²) da sua variância com a outra. A estimativa da variância compartilhada ou comum é chamada de **comunalidade da variável**.

Segundo Hair *et al.* (2009, p. 112), a análise de fatores comuns é a técnica mais adequada quando:

- o objetivo prioritário é identificar as dimensões ou os constructos 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.

Quantidade de fatores a ser selecionada

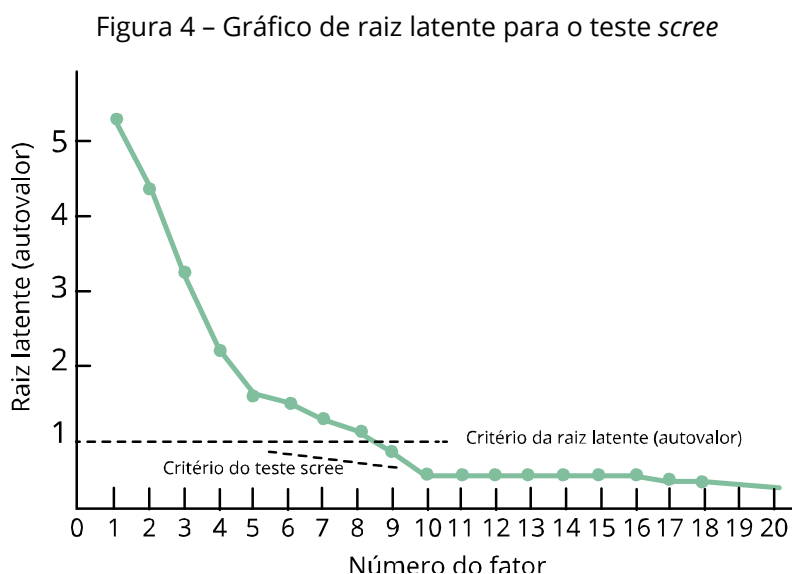
Para determinar a quantidade de fatores a ser selecionada para explicar a estrutura latente de dados, executa-se um processo sequencial: o primeiro fator é o que melhor resume as relações lineares entre os dados, o segundo explica a maior parte da variância após a remoção do primeiro fator, e o processo continua, com cada vez menos variância restando inexplicada.

Como os primeiros fatores explicam parte substancial da variância total, chega-se a um limite de quantidade de fatores de forma relativamente rápida. É importante salientarmos, contudo, que selecionar poucos fatores pode fazer com que a estrutura não seja revelada e dimensões importantes sejam ignoradas, ao passo que uma quantidade excessiva de fatores pode tornar a interpretação dos resultados difícil.

A decisão de parar a fatoraçoão pode-se basear na fundamentação teórica/conceitual ou em evidências empíricas. De acordo com Hair *et al.* (2014), também podem ser considerados outros critérios, como:

- **raiz latente** – cada fator deve ser mantido se ele explicar a variância de pelo menos uma variável;
- **percentagem da variância** – o processo não deve ser parado até que os fatores extraídos expliquem pelo menos 95% da variância (para ciências naturais) ou 60% (para ciências sociais), ou até o último fator explicar menos de 5% e
- **teste scree** – derivado dos critérios anteriores, prevê um gráfico das raízes latentes em função dos fatores por ordem de extração.

A seguir, apresentamos um exemplo de gráfico de raiz latente. Observe que o ângulo de inclinação decresce rapidamente, aproximando-se de uma reta horizontal. O ponto em que o gráfico começa a ficar horizontal é um indicativo de que não há mais justificativa para inclusão de fatores.



Fonte: Hair *et al.* (2009, p. 115).

Quando há heterogeneidade entre os subgrupos, é recomendável ampliar a quantidade de fatores.

Por último, deve ser validado o grau de generalização dos resultados para a população e o grau de influência dos casos ou respondentes individuais sobre os resultados gerais. Para tanto, alguns métodos são recomendados:

- **aplicar a perspectiva confirmatória** – avaliar a repetitividade dos resultados em uma amostra particionada do conjunto de dados originais ou de uma amostra adicional;
- **avaliar a estabilidade da estrutura fatorial** – nos casos em que o tamanho da amostra permitir, particioná-la em dois subconjuntos e estimar os modelos fatoriais para cada um. Em seguida, comparar as matrizes fatoriais para avaliar a robustez da solução e
- **detecção de observações influentes** – estimar o modelo com e sem observações influentes, de forma a avaliar o impacto nos resultados.

Análise conjunta

A análise conjunta (*conjoint analysis*) é uma ferramenta que utiliza as bases do delineamento de experimentos (*design of experiments* – DoE), uma técnica que busca estabelecer níveis ótimos de parâmetros diversos de um produto ou processo por meio de matrizes de teste que combinam os diferentes níveis preestabelecidos de parâmetros às respostas dadas pelos consumidores quanto às características de tal produto ou processo.

De acordo com Sharpe, Veaux e Velleman (2011), um experimento é um estudo em que determinadas entradas de um processo são modificadas na intenção de provocar diferentes respostas nas suas saídas, permitindo a análise da interação existente entre os diversos fatores envolvidos e dos resultados por ela promovidos. Sendo assim, sob a ótica do DoE, um experimento é um processo previamente definido em que o nível de determinadas variáveis vai ser determinado de forma a gerar um resultado. Como diferentes níveis gerarão diferentes resultados finais, o experimento, usualmente na forma de delineamento fatorial fracionário, prevê a testagem de diferentes níveis em diferentes variáveis, de modo a gerar diferentes respostas nas suas saídas e a permitir a análise não só da interação entre os diversos fatores envolvidos mas também dos resultados por ela promovidos.

Navidi (2012, p. 439) comenta que “muitos experimentos envolvem a variação de vários fatores, cada um dos quais podendo afetar a resposta.” Por exemplo, suponhamos que o pesquisador de uma empresa queira determinar os parâmetros que gerarão a maior capacidade de limpeza de um detergente e, com base no seu conhecimento e nas suas experiências prévias, identifique que o percentual do componente A, do componente B e do componente C, bem como a temperatura, a pressão e o tempo de cozimento influenciarão a capacidade de limpeza. No entanto, como ele saberá que proporção, temperatura e tempo maximizarão a capacidade de remover sujeira do detergente?

Para responder a essa pergunta, ele realizará uma série de experimentos, na forma de matrizes previamente definidas, com diferentes níveis para cada um dos fatores (por exemplo, temperaturas de 60°, 70° e 80° C; tempos de seis, sete e oito minutos, etc.). Em seguida, medirá o resultado de cada experimento (teste) realizado, o que pode ser feito com base na quantidade de gordura residual que ficará nas lâminas em que o detergente foi aplicado.

Por meio de análises estatísticas (geralmente, utilizando ANOVA) é possível determinar os fatores que têm influência sobre o resultado e a combinação deles que levaria ao máximo desempenho do detergente. Por exemplo, os resultados podem apontar que a proporção ótima é de 20% do componente A, 50% do componente B e 30% do componente C, submetidos às temperaturas e pressões do processo durante nove minutos, e que os diferentes níveis de temperatura e pressão não afetam a capacidade de limpeza do detergente.

É importante observarmos que experimentos como esse não levam em consideração a percepção do cliente, mas somente aspectos técnicos. Nesse caso específico, surgiriam então alguns questionamentos, como:

- Os consumidores valorizam uma coloração mais clara do detergente, passando a noção de “limpeza”?
- A marca deve ser genérica ou atrelada a um nome específico?
- O mercado valoriza um apelo ecológico para detergentes?
- Que preço seria bem aceito pelo mercado?

Todos esses questionamentos são relevantes para a empresa que desenvolve o detergente, mas eles não foram cobertos pelo DoE.

É possível inserir os aspectos de percepção dos clientes no experimento de formato DoE, mas isso exigiria uma grande quantidade de respondentes bem como uma manipulação para a transformação dos dados. Mesmo assim, as respostas só apontariam para comparações entre grupos.

Podemos pensar também em fenômenos econômicos, como a demanda por determinado bem a partir de diferentes níveis de preço unitário, quantidade por unidade de embalagem, etc. “Como opção, a análise conjunta oferece ao pesquisador uma técnica que pode ser aplicada a um único indivíduo ou a um grupo de indivíduos e fornecer uma perspectiva não apenas sobre as preferências para cada atributo [...], mas também a quantia do mesmo” (HAIR *et al.*, 2009, p. 357).

Objetivos da análise conjunta

A análise conjunta, técnica multivariada usada para avaliar as preferências por produtos ou serviços dos consumidores, surgiu como uma ferramenta exclusiva das áreas de marketing e comportamento do consumidor, que buscavam compreender a mente dos consumidores. No entanto, ao longo do tempo, essa técnica vem ganhando espaço na área de desenvolvimento de novos produtos, pois auxilia na identificação e seleção de características a serem oferecidas em

produtos e serviços que ainda estejam no seu estágio conceitual, bem como em testes de aceitação e em análises estratégicas referentes à percepção de valor pelos clientes e pela sociedade. Engloba, portanto, fenômenos econômicos, modelos e métodos de precificação, previsão de demanda, receita e rentabilidade de novos projetos, análise competitiva, segmentação e participação de mercado, avaliação do apelo nas propagandas e no *design* de produtos, distribuição, etc.

Hair *et al.* (2014) defendem que a possibilidade oferecida pela análise conjunta de acomodar variáveis dependentes tanto métricas quanto não métricas faz da análise conjunta uma ferramenta única, de grande flexibilidade, permitindo que qualquer conjunto de características, como marca, posicionamento, benefícios, imagem, etc., seja avaliado como um conjunto de atributos. Na verdade, os autores afirmam que a percepção de valor dos clientes sobre os produtos e serviços surge como uma combinação dos valores percebidos para cada atributo. Em outras palavras, as estimativas de preferência podem ocorrer mediante o julgamento das combinações de atributos, sendo assim fundamentadas na utilidade, isto é, a valorização implícita dos elementos individuais que compõem o produto ou serviço pode ser determinada.

Ainda que a utilidade seja fundamentalmente subjetiva, torna-se uma medida quantitativa (numérica) da capacidade de satisfazer as necessidades dos consumidores ou da importância relativa percebida pelo mercado consumidor de um atributo em comparação a outro.

De acordo com Shimizu (2006), a noção de utilidade estaria relacionada ao fato de que as decisões ocorrem por parte dos consumidores no sentido de escolherem que o lhes provê o maior nível de satisfação entre as opções existentes: quanto maior a utilidade, maior a preferência.

Alguns aspectos devem ser considerados pertinentes à utilidade. Segundo Hair (*et al.*, 2009, p. 360), a utilidade:

- engloba todas as características do objeto, tangíveis ou intangíveis, e como tal é uma medida de preferência geral do indivíduo;
- é considerada baseada no valor colocado em cada um dos níveis dos atributos. Fazendo isso, respondentes reagem a várias combinações de níveis de atributos (ex.: diferentes preços, características ou marcas), com variados níveis de preferência e
- é expressa por uma relação que reflete a maneira como a utilidade é formulada para qualquer combinação de atributos. Por exemplo, poderíamos somar os valores de utilidade associados a cada característica de um produto ou serviço para chegar a uma utilidade geral. Dessa forma, assumiríamos que produtos ou serviços com maiores valores de utilidade são preferidos e têm uma maior chance de escolha.

Para que a análise conjunta nos permita compreender as reações de consumidores e avaliações de combinações predeterminadas de atributos que representam produtos ou serviços potenciais, devemos analisar:

a) Fator:

Variável manipulada pelo pesquisador, o fator representa um atributo específico ou outra característica do produto ou serviço. Para definir a utilidade com precisão, o pesquisador deve identificar todos os atributos importantes que poderiam afetar preferências e, desse modo, a utilidade.

b) Níveis:

Valores possíveis para cada fator. Cada fator é definido pelos seus níveis, que são os valores possíveis para aquele fator. Tais valores permitem ao pesquisador descrever um objeto em termos dos seus níveis no conjunto de fatores que o caracterizam. Por exemplo, nome da marca e preço poderiam ser dois fatores em uma análise conjunta. O nome da marca poderia ter dois níveis (marca X e marca Y), ao passo que o preço poderia ter quatro níveis (39, 49, 59 e 59 centavos).

c) Tratamento (estímulo):

Quando o pesquisador seleciona os fatores e os níveis para descrever um objeto de acordo com um plano específico, essa combinação é conhecida como um tratamento ou estímulo. O estímulo é o conjunto específico de fatores e níveis avaliado pelo entrevistado (MALHOTRA, 2001). No nosso exemplo, poderia ser a marca X a um preço de 49 centavos.

Um aspecto que diferencia a análise conjunta dos experimentos técnicos envolve o fato de que as combinações de atributos em níveis diversos são apresentadas a respondentes na forma de produtos ou serviços controlados, reais ou hipotéticos (ex.: protótipos, *mock-ups* ou imagens). Os respondentes devem então avaliar os conjuntos oferecidos e fazer a sua escolha. As unidades apresentadas devem ser semelhantes o suficiente para que os consumidores percebam que têm o mesmo uso, como concorrentes, mas distintas o suficiente para que eles possam determinar claramente uma preferência, e não uma indeterminação por indiferença.

Não é previsto que os respondentes desdobrem e detalhem as suas escolhas quanto, por exemplo, à importância de um atributo específico ou quanto ao desempenho de um atributo individual, exceto se lhes for solicitado, eventualmente, fazer avaliações individuais quanto ao ordenamento ou à preferência entre as combinações alternativas.

Trata-se de uma grande vantagem da análise conjunta, visto que, conforme afirmam Aaker, Kumar e Day (2001), o aumento da quantidade de atributos torna o julgamento de características individuais muito mais complexo: mais de cinco ou seis atributos já seriam suficientes para reduzir a capacidade analítica dos respondentes, fazendo com que eles possam vir a ignorar as variações em outros fatores considerados de menor importância.

Lembre-se de que, à medida que a quantidade de combinações de atributos e níveis aumenta, o número de perfis também aumenta exponencialmente.

Esse processo aproxima-se bastante da situação real de compra, em que o cliente, ainda que inconscientemente, escolhe um produto pela melhor combinação de atributos. Usualmente, não há a explicitação, detalhamento ou mesmo hierarquização da importância de tais atributos e dos seus níveis: o cliente simplesmente toma a decisão e faz a compra.

Cabe então ao pesquisador entender a estrutura de preferência do respondente e identificar o conjunto de atributos do produto/serviço de acordo com a escolha do consumidor, analisando a influência de cada atributo e de cada nível no julgamento desse respondente quanto à utilidade. Em outras palavras, o pesquisador deve explicar a importância de cada fator na decisão do respondente e demonstrar como os níveis distintos dentro de um fator podem influenciar a sua preferência.

Projeto e suposições em análise conjunta

No projeto e planejamento de uma análise conjunta, o pesquisador deve estabelecer os atributos a serem utilizados, questionando se eles conferem utilidade ou valor ao produto/serviço estudado. Nesse sentido, Hair *et al.* (2014) destacam que todos os atributos que criam ou reduzem a utilidade geral do produto ou serviço (isto é, que tornam o objeto tanto atraente quanto desinteressante) devem ser incluídos: a omissão dos fatores negativos distorceria o julgamento dos respondentes, visto que, mesmo não fornecidos, podem ser inconscientemente considerados por eles.

É importante destacar também que os níveis devem ser especificados em termos precisos, de forma a evitar diferenças de percepção entre indivíduos. Dessa forma, classificações como alto, médio e baixo devem ser evitadas. Além disso, os intervalos entre os níveis não devem ser excessivos, fora do esperado, visto que isso pode comprometer a percepção dos respondentes. O pesquisador deve ainda procurar equilibrar a quantidade de níveis nos fatores, uma vez que a importância relativa estimada de uma variável tende a aumentar quando a quantidade de níveis cresce, o que poderia gerar distorções na análise.

Como regra geral para a quantidade de estímulos, Hair *et al.* (2014) recomendam que a quantidade mínima seja calculada pela seguinte fórmula:

$$\text{quantidade mínima de estímulos} = \text{quantidade total de níveis por todos os fatores} - \text{quantidade de fatores} + 1$$

Por exemplo, uma análise conjunta com cinco fatores e três níveis cada exigiria um mínimo de $5 \times 3 - 5 + 1 = 11$ estímulos. É importante destacar que se trata de uma recomendação de quantidade mínima: relações mais complexas podem requerer um aumento na quantidade de estímulos, mas isso pode tornar a escolha mais complexa e difícil para os respondentes.

Como exemplo, vamos tomar por base o desenvolvimento do detergente que vimos anteriormente. Para tanto, vamos adaptar o exemplo apresentado por Hair *et al.* (2014), que consideraram os fatores ingrediente, forma e nome da marca, como mostrado na tabela a seguir.

Tabela 2 – Fatores e níveis

fator	níveis	
ingredientes	sem fosfato	com fosfato
forma	líquido	pó
nome da marca	marca <i>top</i>	marca genérica

Fonte: Adaptado de Hair *et al.* (2014).

Nesse caso, há oito possibilidades de combinação:

1. sem fosfato, líquido, marca *top*;
2. sem fosfato, líquido, marca genérica;
3. sem fosfato, pó, marca *top*;
4. sem fosfato, pó, marca genérica;
5. com fosfato, líquido, marca *top*;
6. com fosfato, líquido, marca genérica;
7. com fosfato, pó, marca *top* e
8. com fosfato, pó, marca genérica.

Definidos os estímulos, eles devem ser submetidos aos respondentes para que estes possam expressar as suas preferências, o que pode ser feito na forma de escala de preferência para compra (de 1 = mais preferido a 8 = menos preferido). Apresentamos, a seguir, uma possível configuração de respostas.

Tabela 3 – Exemplo de possíveis respostas

estímulo #	forma	ingredientes	marca	respondente 1	respondente 2
1	líquido	sem fosfato	top	1	1
2	líquido	sem fosfato	genérica	2	2
3	líquido	com fosfato	top	5	3
4	líquido	com fosfato	genérica	6	4
5	pó	sem fosfato	top	3	7
6	pó	sem fosfato	genérica	4	5
7	pó	com fosfato	top	7	8
8	pó	com fosfato	genérica	8	6

Fonte: Adaptado de Hair *et al.* (2014).

Lembre-se de que a estrutura de preferência do detergente toma por base os três fatores e, dessa forma:

$$\text{utilidade} = \text{efeito da forma} + \text{efeito do ingrediente} + \text{efeito da marca}$$

Vamos, agora, ordenar os estímulos para cada respondente e para cada fator, e também calcular o desvio da média. Observe que, no exemplo, a média é dada por $(1 + 2 + \dots + 8)/8 = 4,5$.

Tabela 4 – Ordenação de estímulos e desvio da média – respondente 1

respondente 1		ordenação				média da ordem	desvio da média	desvio ao quadrado
forma	líquido	1	2	5	6	3,5	-1,0	1
	pó	3	4	7	8	5,5	1,0	1
ingredientes	sem fosfato	1	2	3	4	2,5	-2,0	4
	com fosfato	5	6	7	8	6,5	2,0	4
marca	top	1	3	5	7	4,0	-0,5	0,25
	genérica	2	4	6	8	5,0	0,5	0,25

Fonte: Adaptado de Hair *et al.* (2014).

Tabela 5 – Ordenação de estímulos e desvio da média – respondente 2

respondente 2		ordenação				média da ordem	desvio da média	desvio ao quadrado
forma	líquido	1	2	3	4	2,5	-2,0	4
	pó	5	6	7	8	6,5	2,0	4
ingredientes	sem fosfato	1	2	5	7	3,75	-0,75	0,5625
	com fosfato	3	4	6	8	5,25	0,75	0,5625
marca	top	1	3	7	8	4,75	0,25	0,0625
	genérica	2	4	5	6	4,25	-0,25	0,0625

Fonte: Adaptado de Hair *et al.* (2014).

Somamos, em seguida, os quadrados dos desvios (última coluna das tabelas), como mostrado a seguir.

Tabela 6 – Soma dos quadrados dos desvios

desvio ao quadrado/respondente		1	2
forma	líquido	1	4
	pó	1	4
ingredientes	sem fosfato	4	0,5625
	com fosfato	4	0,5625
marca	top	0,25	0,0625
	genérica	0,25	0,0625
total		10,5	9,25

Fonte: Adaptado de Hair *et al.* (2014).

Podemos, agora, calcular o valor de padronização para cada respondente, dividindo a quantidade de níveis (três fatores com dois níveis = seis níveis) pela soma dos quadrados. Temos então:

Tabela 7 – Valor da padronização para cada respondente

soma dos quadrados/respondente		1	2
forma	líquido	1	4
	pó	1	4
ingredientes	sem fosfato	4	0,5625
	com fosfato	4	0,5625
marca	top	0,25	0,0625
	genérica	0,25	0,0625
total		10,5	9,25
valor de padronização		0,571	0,649

Fonte: Adaptado de Hair *et al.* (2014).

Os valores de padronização são multiplicados então pelos quadrados dos desvios, e é calculada a raiz do produto, sendo que o sinal adotado deve ser o inverso do desvio (visto que estamos mensurando os afastamentos). Os resultados de tais operações são os seguintes:

Tabela 8 – Utilidades parciais – respondente 1

respondente 1		ordenação				média da ordem	desvio da média	desvio ao quadrado	x valor padroniz.	utilidade parcial
forma	líquido	1	2	5	6	3,5	-1,0	1	0,571	0,756
	pó	3	4	7	8	5,5	1,0	1	0,571	-0,756
ingredientes	sem fosfato	1	2	3	4	2,5	-2,0	4	2,286	1,512
	com fosfato	5	6	7	8	6,5	2,0	4	2,286	-1,512
marca	top	1	3	5	7	4,0	-0,5	0,25	0,143	0,378
	genérica	2	4	6	8	5,0	0,5	0,25	0,143	-0,378

Fonte: Adaptado de Hair *et al.* (2014).

Tabela 9 – Utilidades parciais – respondente 2

respondente 2		ordenação				média da ordem	desvio da média	desvio ao quadrado	x valor padroniz.	utilidade parcial
forma	líquido	1	2	3	4	2,5	-2,0	4	2,595	1,611
	pó	5	6	7	8	6,5	2,0	4	2,595	-1,611
ingredientes	sem fosfato	1	2	5	7	3,75	-0,75	0,5625	0,365	0,604
	com fosfato	3	4	6	8	5,25	0,75	0,5625	0,365	-0,604
marca	top	1	3	7	8	4,75	0,25	0,0625	0,041	-0,201
	genérica	2	4	5	6	4,25	-0,25	0,0625	0,041	0,201

Fonte: Adaptado de Hair *et al.* (2014).

Esse processo gera como resultados as utilidades parciais para cada nível e para cada respondente (última coluna das tabelas mostradas). No entanto, tal medida é de difícil interpretação: precisamos calcular a importância de cada fator.

Isso é feito quando calculamos a amplitude de utilidades parciais, totalizando-as e identificando a proporção correspondente de cada fator, como mostrado a seguir.

Tabela 10 – Importância de cada fator – respondente 1

respondente 1		utilidade parcial	amplitude da utilidade	importância do fator
forma	líquido	0,756	1,512	28,6%
	pó	-0,756		
ingredientes	sem fosfato	1,512	3,024	57,1%
	com fosfato	-1,512		
marca	top	0,378	0,756	14,3%
	genérica	-0,378		
total			5,292	

Fonte: Adaptado de Hair *et al.* (2014).

Tabela 11 – Importância de cada fator – respondente 2

respondente 2		utilidade parcial	amplitude da utilidade	importância do fator
forma	líquido	1,611	3,222	66,7%
	pó	-1,611		
ingredientes	sem fosfato	0,604	1,208	25,0%
	com fosfato	-0,604		
marca	top	-0,201	0,403	8,3%
	genérica	0,201		
total			4,832	

Fonte: Adaptado de Hair *et al.* (2014).

Agora que já compreendemos a lógica da análise, podemos executá-la de forma mais rápida, utilizando a linguagem R. A seguir, apresentamos os *scripts*, relatórios e gráficos gerados.

```
#Carregando pacote de Análise conjunta
library(conjoint)

#Nomes dos níveis
níveis<-c("Líquido", "Pó", "Sem fosfato", "Com fosfato", "Marca Top", "Marca genérica")

#respostas
respondente1<-c(1,2,5,6,3,4,7,8)
respondente2<-c(1,2,3,4,7,5,8,6)

#Vetores com os parâmetros dos 3 fatores
Ingrediente<-c(1,1,2,2,1,1,2,2)
Forma<-c(1,1,1,1,2,2,2,2)
Marca<-c(1,2,1,2,1,2,1,2)

#Criando tabela com os 3 fatores
fatores<-data.frame(Forma, Ingrediente, Marca)
```

#Análise Conjunta do respondente 1

Conjoint(respondente1, fatores, níveis)

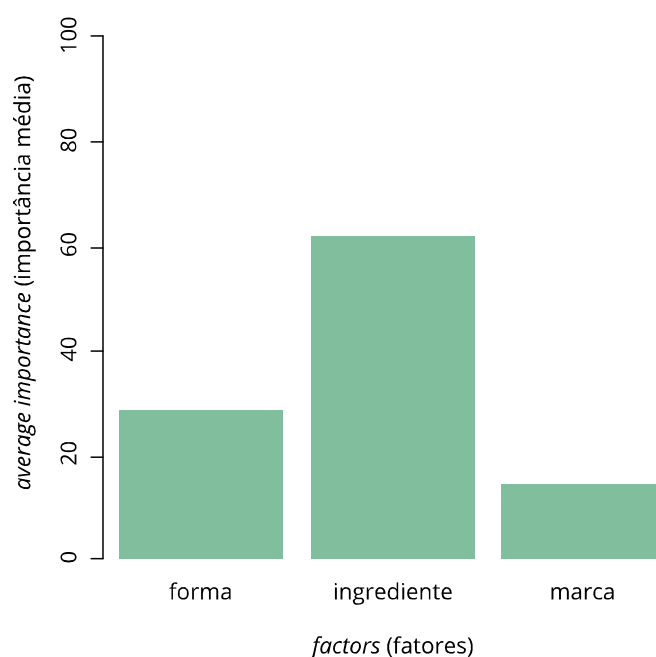
Tabela 12 – Níveis e utilidade – respondente 1

níveis	utilidade
líquido	-1
pó	1
sem fosfato	-2
com fosfato	2
marca <i>top</i>	-0,5
marca genérica	0,5

Tabela 13 – Importância dos fatores – respondente 1

importância dos fatores		
forma	ingrediente	marca
28%	57%	29%

Figura 5 – Importância média dos fatores – respondente 1



#Análise Conjunta do respondente 2

Conjoint(respondente2, fatores, níveis)

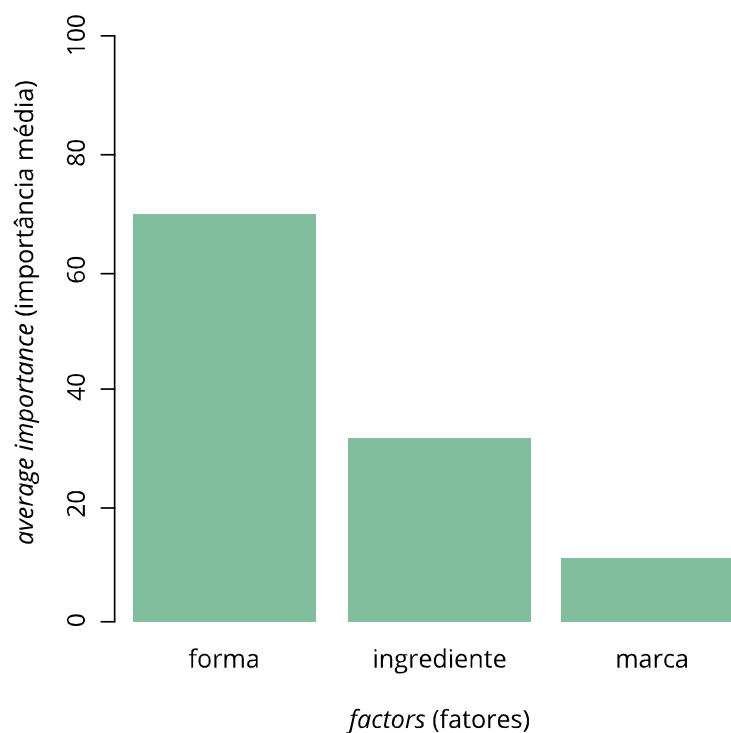
Tabela 14 – Níveis e utilidade – respondente 2

níveis	utilidade
líquido	-2
pó	2
sem fosfato	-0,75
com fosfato	0,75
marca <i>top</i>	0,25
marca genérica	0,25

Tabela 15 – Importância dos fatores – respondente 2

importância dos fatores		
forma	ingrediente	marca
67%	25%	8,33%

Figura 6 – Importância média dos fatores – respondente 2



Conforme Hair *et al.* (2014), após determinar as preferências na avaliação geral do consumidor, o pesquisador poderá:

- definir o objeto ou conceito com a combinação ótima de características;
- mostrar as contribuições relativas de cada atributo e cada nível para a avaliação geral do objeto;
- usar as estimativas de julgamentos de comprador ou cliente para prever preferências entre objetos com diferentes conjuntos de características;
- isolar grupos de clientes potenciais, segmentos com potenciais altos e baixos, e
- identificar oportunidades de marketing explorando o potencial de mercado.

Autocorrelação entre fatores

Um potencial problema na análise conjunta é a existência de autocorrelação entre fatores, caracterizando a falta de independência entre eles.

Vamos então praticar. Foi feito um levantamento de preferências dos consumidores de refrigerantes considerando o sabor (sabor cola *versus* sabor frutas cítricas), o tipo de recipiente em que o refrigerante é comercializado (latinha ou garrafa) e se o refrigerante é normal ou *light* (de baixa caloria), como mostrado a seguir.

Tabela 16 – Fatores e níveis – refrigerante

fator	níveis	
sabor	cola	cítrico
recipiente	lata	garrafa
tipo	normal	baixa caloria

Os resultados encontrados são mostrados na tabela que segue.

Tabela 17 – Resultados

estímulo #	sabor	recipiente	tipo	respondente 1	respondente 2	respondente 3
1	cola	lata	normal	1	3	3
2	cola	lata	baixa caloria	3	5	1
3	cola	garrafa	normal	2	4	4
4	cola	garrafa	baixa caloria	4	6	2
5	cítrico	lata	normal	5	2	7
6	cítrico	lata	baixa caloria	6	8	5
7	cítrico	garrafa	normal	7	1	8
8	cítrico	garrafa	baixa caloria	8	7	6

Agora, vejamos a análise em R:

```
#Carregando pacote de Análise conjunta
library(conjoint)

#Nomes dos níveis
níveis<-c("Cola","Crítico","lata","garrafa","normal","baixa caloria")

#respostas
respondente1<-c(1,3,2,4,5,6,7,8)
respondente2<-c(3,5,4,6,2,8,1,7)
respondente3<-c(3,1,4,2,7,5,8,6)

#Vetores com os parâmetros dos 3 fatores
Sabor<-c(1,1,1,1,2,2,2,2)
Recipiente<-c(1,1,2,2,1,1,2,2)
Tipo<-c(1,2,1,2,1,2,1,2)

#Criando tabela com os 3 fatores
fatores<-data.frame(Sabor,Recipiente,Tipo)
```

#Análise Conjunta do respondente 1
Conjoint(respondente1,fatores,níveis)

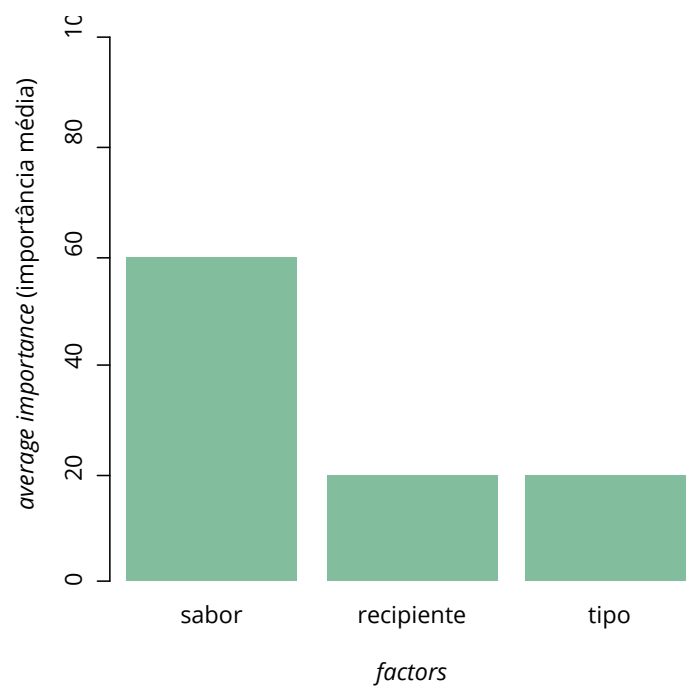
Tabela 18 – Níveis e utilidade – respondente 1

níveis	utilidade
cola	-2
crítico	2
lata	-0,75
garrafa	0,75
normal	-0,75
baixa caloria	0,75

Tabela 19 – Importância dos fatores – respondente 1

importância dos fatores		
sabor	recipiente	tipo
57%	21%	21%

Figura 7 – Importância média dos fatores – respondente 1



#Análise Conjunta do respondente 2

Conjoint(respondente2,fatores,níveis)

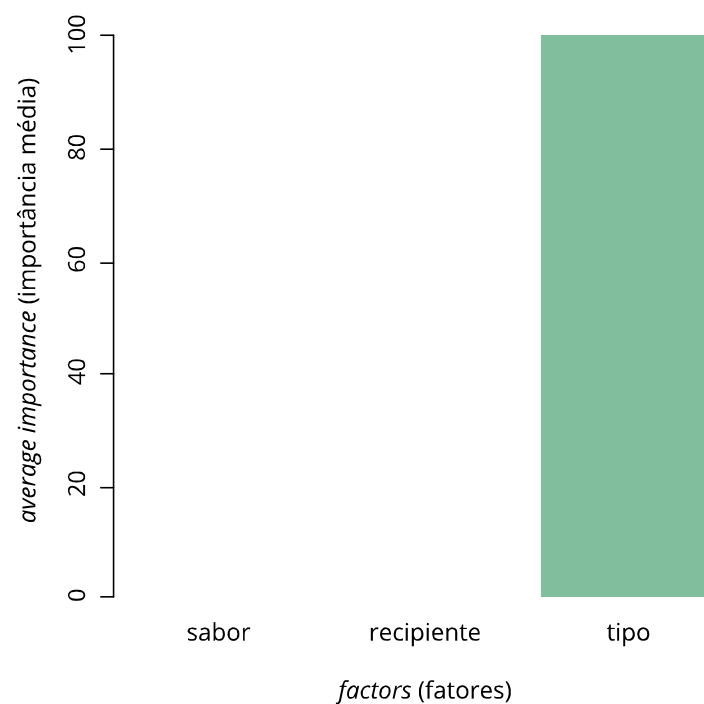
Tabela 20 – Níveis e utilidade – respondente 2

níveis	utilidade
cola	0
crítico	0
lata	0
garrafa	0
normal	-2
baixa caloria	2

Tabela 21 – Importância dos fatores – respondente 2

importância dos fatores		
sabor	recipiente	tipo
0%	0%	100%

Figura 8: Importância média dos fatores – respondente 2



#Análise Conjunta do respondente 3

Conjoint(respondente3,fatores,níveis)

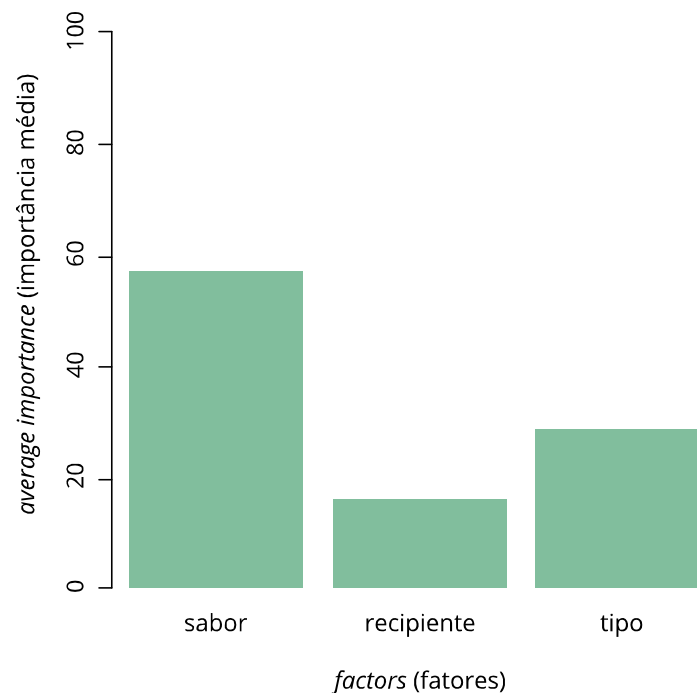
Tabela 22 – Níveis e utilidade – respondente 3

níveis	utilidade
cola	-2
crítico	2
lata	-0,5
garrafa	0,5
normal	1
baixa caloria	-1

Tabela 23 – Importância dos fatores – respondente 3

importância dos fatores		
sabor	recipiente	tipo
57%	14%	28%

Figura 8 – Importância média dos fatores – respondente 3



Os resultados referentes ao respondente 2 parecem estranhos, não é mesmo? Principalmente se observarmos o gráfico, chama-nos a atenção o fato de o fator “tipo” ter 100% de importância, ao passo que os outros dois têm importância zero. Será que houve algum erro?

Vamos rearranjar as respostas apresentadas e analisá-las de forma mais clara. Vejamos:

Tabela 24 – Respostas rearranjadas

estímulo #	sabor	recipiente	tipo	respostas
7	cítrico	garrafa	normal	1
5	cítrico	lata	normal	2
1	cola	lata	normal	3
3	cola	garrafa	normal	4
2	cola	lata	baixa caloria	5
4	cola	garrafa	baixa caloria	6
8	cítrico	garrafa	baixa caloria	7
6	cítrico	lata	baixa caloria	8

Na tabela apresentada, podemos observar que a sequência de preferência deixa clara a importância do fator “tipo”, pois todas as alternativas do nível “normal” tiveram prioridade superior às de baixa caloria. Mas, e os demais fatores? Não têm qualquer importância? Repare que a sequência do fator “sabor” para refrigerantes do tipo normal é cítrico-cítrico-cola-cola, ao passo que, para os de baixa caloria, é exatamente o oposto. Para o fator “recipiente”, a mesma coisa acontece.

Em outras palavras, não há qualquer lógica identificável na preferência de tais fatores. Podemos então inferir que há aleatoriedade, isto é, incerteza quanto a qualquer preferência por parte do respondente 2 quanto a tais fatores.

Mas, por que não podemos fazer a mesma inferência quanto às preferências dos respondentes 1 e 3, visto que ambos mostraram clara preferência pelo sabor cola, como podemos perceber nas tabelas a seguir?

Tabela 25 – Preferências do respondente 1

estímulo #	sabor	recipiente	tipo	respondente 1
1	cola	lata	normal	1
3	cola	garrafa	normal	2
2	cola	lata	baixa caloria	3
4	cola	garrafa	baixa caloria	4
5	cítrico	lata	normal	5
6	cítrico	lata	baixa caloria	6
7	cítrico	garrafa	normal	7
8	cítrico	garrafa	baixa caloria	8

Tabela 26 – Preferências do respondente 3

estímulo #	sabor	recipiente	tipo	respondente 3
2	cola	lata	baixa caloria	1
4	cola	garrafa	baixa caloria	2
1	cola	lata	normal	3
3	cola	garrafa	normal	4
6	cítrico	lata	baixa caloria	5
8	cítrico	garrafa	baixa caloria	6
5	cítrico	lata	normal	7
7	cítrico	garrafa	normal	8

Observe com mais atenção: para ambos os respondentes, a sequência de preferência parece indicar certa preferência da lata a garrafa. Além disso, o respondente 1 parece ter certa preferência pelo refrigerante normal, enquanto o respondente 3 parece preferir o de baixa caloria. Assim sendo, diferentemente do que vimos no respondente 2, os fatores “recipiente” e “tipo” têm algum peso nas escolhas dos clientes. No entanto, obviamente, tais pesos, ainda que não nulos, são pequenos quando comparados com o do fator “sabor”, e isso é o que vemos nos relatórios e gráficos gerados.

Estimação e avaliação do modelo

A estimação da análise conjunta pode ocorrer sob as seguintes óticas:

a) Regressão linear múltipla:

Quando a variável dependente é métrica, a técnica mais utilizada na estimação do modelo é a regressão linear múltipla, analisando as utilidades parciais de cada nível dos atributos com os coeficientes estimados da reta de regressão.

b) Programação linear:

Quando a variável dependente é ordinal, aplicam-se técnicas de programação linear, estimando os parâmetros de modo a que a concordância entre a ordenação observada e estimada seja máxima.

c) Modelos probabilísticos de escolha:

Trata-se de modelos com origem microeconômica em que se estima um modelo agregado: cada respondente seleciona um estímulo e explica a sua escolha em função das características dos estímulos.

Quanto à interpretação dos resultados, ela pode ser feita pela análise agregada ou desagregada. A análise desagregada é a mais comum. Nela cada respondente é modelado separadamente, e os resultados são então examinados.

A validação dos resultados da análise conjunta engloba:

- a validação interna – confirmação de que o modelo adotado é apropriado e
- a validação externa – relacionada à representatividade da amostra.

A validação do modelo é um dos aspectos centrais da aplicação da análise conjunta.

Na validação do modelo, devem-se considerar os seguintes aspectos:

a) Confiabilidade:

A confiabilidade deve ser considerada sob diferentes óticas, tais como:

- temporal – os resultados seriam os mesmos em outro período de tempo?
- dos atributos – os resultados seriam os mesmos caso fossem adicionados outros atributos?
- dos estímulos – os resultados seriam os mesmos caso fosse adotado outro tipo de arranjo?
- do método de coleta de dados – os resultados seriam os mesmos caso fosse utilizado outro método de abordagem, apresentação ou escala da variável dependente?

b) Precisão preditiva:

Como os arranjos fatoriais são fracionários, é recomendado medir a capacidade de o modelo conjunto estimado prever a avaliação dos estímulos. Níveis elevados de precisão preditiva para os estímulos de estimação e de validação entre os respondentes confirmam as regras de composição.

c) Validade preditiva:

A validade preditiva consiste em comparar resultados conjuntos com os resultados de mercado, podendo traduzir-se em medidas como a proporção de respondentes cuja escolha foi corretamente prevista.

BIBLIOGRAFIA

AAKER, David A.; KUMAR, V.; DAY, George S. *Marketing research*. New York: John Wiley & Sons, 2001.

ALBUQUERQUE, Eduardo Rui Viana Barbas. *A representação gráfica de dados multivariados como instrumento de apoio à análise econômica e financeira: uma adaptação dos rostos de Chernoff na representação gráfica da informação contabilística*. 2016. 400f. Tese (Doutorado em Gestão) – Universidade Lusíada de Lisboa, Faculdade de Ciências da Economia e da Empresa, Lisboa, 2016.

ALMEIDA, José Renato Monteiro Nascimento de. *Metodologia para análise da confiabilidade de um conjunto de alta criticidade de uma planta siderúrgica integrada*. 2008. 180f. Dissertação (Mestrado em Engenharia de Produção) – PPGEPP / Centro de Tecnologia / Campus I / Universidade Federal da Paraíba – UFPB, 2008.

ANÁLISE fatorial, uma importante técnica multivariada. *ABG Consultoria*, 7 ago. 2017a. Disponível em: <http://www.abgconsultoria.com.br/blog/analise-fatorial/>. Acesso em: 1 nov. 2018.

BARROS, Bruna Soares Xavier de; BARROS, Zacarias Xavier de; POLLO, Ronaldo Alberto. Estatística multivariada e análise de contrastes ortogonais aplicadas em variáveis dimensionais de bacias hidrográficas. *Irriga*, Botucatu, Edição Especial, 20 anos Irriga + 50 anos FCA, p. 221-233, 2015.

BATTESINI, Marcelo. Método de análise conjunta com estimulação em duas etapas. 2002. 134f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul. Escola de Engenharia, Porto Alegre, 2002.

BOMTEMPO, Maurício Scagliante. Representação de dados multivariados através das faces de Chernoff. *Estação Científica*, Juiz de Fora, n.14, dez. 2015.

CAMPANA, Ângela Nogueira; TAVARES, Maria da Consolação Gomes Cunha Fernandes; SILVA, Dirceu da. Modelagem de equações estruturais: apresentação de uma abordagem estatística multivariada para pesquisas em Educação Física. *Motri*, v. 5, n. 4, p. 59-80, Vila Real, dez. 2009.

CAMPOS, Livia Maria Almendra Lemos Moerbeck de. *Abordagem da análise conjunta no processo decisório de aquisição do serviço de academia de ginástica*. 2006. 89f. Dissertação (Mestrado em Administração) – Fundação Getulio Vargas. Rio de Janeiro, 2006.

CATEN, Carla S.; BATTESINI, Marcelo. Análise conjunta: uma abordagem para obtenção de maiores informações sobre o comportamento do consumidor. *In: ENCONTRO NACIONAL DE ENGENHARIA DA PRODUÇÃO(Enegep). Anais [...]*. Salvador: Abepro, 2001.

CHRISTOFOLETTI, Sergio Ricardo; THOMAZELLA, Helber Roberto; MORENO, Maria Margarida Torres; MASSON, Marcos Roberto. Utilização da análise estatística multivariada no tratamento de dados aplicados a matérias-primas cerâmicas. *Revista do Instituto Geológico*, São Paulo, v. 26 (1/2), p. 19-29, 2005.

DELLA LUCIA, Suzana Maria *et al.* Análise conjunta de fatores baseada em escolhas no estudo da embalagem de iogurte light sabor morango. *Brazil Journal Food Technology*, 6º Sensiber, p. 11-18, ago. 2010.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. *Análise de dados: técnicas multivariadas exploratórias com SPSS e Stata*. Rio de Janeiro: Elsevier, 2015.

FOGLIATTO, Flavio Sanson. Estratégias para modelagem de dados multivariados na presença de correlação. *Gestão e Produção*, São Carlos , v. 7, n. 1, p. 17-28, abr. 2000.

GUEDES, Hugo A. S.; SILVA, Demetrius D. da; ELESBON, Abrahão A. A. ROBEIRO, Celso B. M.; MATOS, Antonio T. de; SOARES, José H. P. Aplicação da análise estatística multivariada no estudo da qualidade da água do Rio Pomba, MG. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 16, n. 5, p. 558-563, 2012.

GUIMARÃES, Francisco Roberto F.; FREITAS, Ana Augusta Ferreira de; ARRUDA, João Bosco Furtado; LEMENHE, Flávio. Análise de demanda por computador popular através da técnica de análise conjunta com dados de preferência declarada. *In: XXVI ENCONTRO NACIONAL DE ENGENHARIA DA PRODUÇÃO(Enegep). Anais [...]*. Fortaleza, CE: Abrepro, 2006.

GOMES, Vânia Sofia Pires Simões. *Análise estatística multivariada aplicada a dados hidrogeológicos*. 2013. 128f. Dissertação (Mestrado em Estatística) – Universidade de Lisboa, Faculdade de Ciências, Departamento de Estatística e Investigação Operacional. Lisboa, 2013.

HAIR JR., Joseph F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E. THATAM, Ronald. *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009.

HAIR JR., Joseph F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E. *Multivariate data analysis. seventh edition*. Essex: Pearson Education, 2014.

HENRIQUE, Jorge Luiz; SOUZA, Rosana Vieira de. O uso da técnica de análise conjunta na pesquisa em marketing: uma avaliação das publicações brasileiras. *In: 30º ENCONTRO DA ANPAD (EnAPAD). Anais [...]. Anpad: Salvador. 2006.*

MALHOTRA, Naresh K. *Pesquisa de marketing: uma orientação aplicada*. 3. ed. Porto Alegre: Bookman, 2001.

MANLY, Bryan F. J.; ALBERTO, Jorge A. Navarro. *Métodos estatísticos multivariados: uma introdução*. 4. ed. Porto Alegre: Bookman, 2019.

MARQUES, Jair Mendes; LIMA, José Donizetti de. A estatística multivariada na análise econômico-financeira de empresas. *Revista FAE*, Curitiba, v. 5, n. 3, p. 51-59, set./dez. 2002.

MORAES, Odair Barbosa de; ABIKO, Alex Kenya. Utilização da análise fatorial para a identificação de estruturas de interdependência de variáveis em estudos de avaliação pós-ocupação. *In: XI ENCONTRO NACIONAL DE TECNOLOGIA NA AMBIENTE CONSTRUÍDO. Anais [...]. Antac: Florianópolis, 2006. p. 1233-1242. Disponível em: http://www.infohab.org.br/entac2014/2006/artigos/ENTAC2006_1233_1242.pdf. Acesso em: 1 nov. 2018.*

MOREIRA, Daniel Augusto. *Administração da produção e operações*. 2. ed. São Paulo: Cengage Learning, 2008.

MOSKOWITZ, Howard; KRIEGER, Bert; RABINO, Samuel. Element category importance in conjoint analysis: evidence for segment differences. *Journal of Targeting, Measurement and Analysis for Marketing*, v. 10, n. 4, 2002.

MUCCI, Daniel Magalhães; FREZATTI, Fábio; DIENG, Mamadou. As múltiplas funções do orçamento empresarial. *RAC*, Rio de Janeiro, v. 20, n. 3, art. 2, p. 283-304, maio/jun. 2016.

NAVIDI, Willian. *Probabilidade e estatística para ciências exatas*. Porto Alegre: AMGH, 2012.

NEVES, Jorge Alexandre Barbosa. *Modelo de equações estruturais: uma introdução aplicada*. Brasília: Enap, 2018.

O QUE é modelagem de equações estruturais?. *ABG Consultoria*, 29 maio 2017b. Disponível em: <http://www.abgconsultoria.com.br/blog/o-que-e-modelagem-de-equacoes-estruturais/>. Acesso em: 1 nov. 2018.

PILATI, Ronaldo; ABBAD, Gardênia. Análise fatorial confirmatória da escala de impacto do treinamento no trabalho. *Psicologia: Teoria e Pesquisa*, v. 21 n. 1, p. 43-51, jan./abr. 2005.

PORTELA, Domingas Maria Pina. *Contributo das técnicas de análise fatorial para o estudo do programa “Ocupação Científica de Jovens nas Férias”*. 2012. 169f. Dissertação (Mestrado em Estatística, Matemática e Computação) – Universidade Aberta. Lisboa, Portugal, 2012.

SANCHEZ-PEDRAZA, Ricardo; CÁCERES, Heidi Alexandra. Análisis de los datos mediante herramientas gráficas. *Rev. Fac. Med. UN Col.*, v. 48, n. 2, 2000.

SANTOS, Levi Alá Neves dos. *Contribuição da mineração de dados e da otimização heurística para a interpretação dos dados da produção científica brasileira*. 2011. 114f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal da Bahia. Instituto de Ciência da Informação. Salvador, 2011.

SILVA NETO, Marco Aurélio. *Mineração visual de dados: extração do conhecimento a partir das técnicas de visualização da informação e mineração de dados*. 2008. 172f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, UFPR, Curitiba, 2008.

SHARPE, Noreen R.; VEAUX, Richard D. De; VELLEMAN, Paul F. *Estatística aplicada: administração, economia e negócios*. Porto Alegre: Bookman, 2011.

SHIMIZU, Tamio. *Decisão nas organizações*. 2. ed. São Paulo: Atlas, 2006.

PROFESSOR-AUTOR

Henrique Martins Rocha é D.Sc. em Engenharia Mecânica pela Universidade Estadual Paulista, com estudos pós-doutorais em projetos e desenvolvimento de novos produtos pela mesma instituição. Atuou durante 27 anos no Brasil, nos USA e no Canadá, em funções executivas e técnicas de empresas como Xerox, White Martins, Flextronics, Remington e CBV, em áreas como: análise de negócios e estratégia empresarial, projetos, programas, desenvolvimento e lançamento de produtos, processos, planejamento e controle de produção. Atua também na área acadêmica desde 2001, como professor e coordenador de cursos de graduação e pós-graduação, bem como professor-tutor de EAD e MOOC. Recebeu diversos prêmios, títulos e homenagens na área acadêmica e profissional. Orientou quase 300 trabalhos de conclusão (doutorado, mestrado, especialização/MBA e graduação) e de iniciação científica. Publicou quase uma centena de artigos em periódicos e anais de eventos científicos. É autor de três livros, coautor de seis e organizador de um. É ainda autor de 29 capítulos de livros e foi responsável pela revisão técnica de cerca de 250 obras nas áreas de engenharia, gestão de operações e educação superior.



