MODELAGEM POR EQUAÇÕES ESTRUTURAIS: CONCEITOS E APLICAÇÕES

MODELING BY STRUCTURAL EQUATIONS: CONCEPTS AND APPLICATIONS

Marlusa Gosling
Faculdade Batista/MG e Face-UFMG

Carlos Alberto Gonçalves
Face-UFMG

RESUMO

Buscando vantagem competitiva e melhor produtividade, pesquisadores e organizações estão usando mais e mais os resultados das pesquisas em seus processos de tomada de decisão. As técnicas quantitativas têm sido aplicadas para modelar a realidade e buscar a previsibilidade. Apesar das restrições acadêmicas de uma visão de ciência extremamente positivista, a Estatística deve ser considerada essencial às pesquisas de marketing. Nesse contexto, há o uso crescente de uma nova ferramenta que possibilita verificar se os dados coletados mostram evidências de que realmente se comportam como o modelo idealizado e testado. Tal ferramenta é denominada modelagem por equações estruturais. O objetivo principal deste artigo é apresentar uma revisão teórica dessa técnica. Apesar de seu fascínio, a modelagem por equações estruturais deve ser usada de acordo com o problema de pesquisa previamente definido.

PALAVRAS-CHAVE

Construtos; Modelagem; Equações estruturais; Marketing; Pesquisa quantitativa

ABSTRACT

Searching for competitive advantage and better productivity, researchers and organizations are using more and more the outcomes of researches in their decision making processes. The quantitative techniques have been applied to model reality and to look for predictability. Despite the academic restrictions of an extremely positive view of science, statistics must be classified as essential to marketing researches. In this context, there is a growing use of a new tool. It makes possible to verify if the collected data show evidences that they actually behave as the idealized and tested model. This tool is called Structural Equation Modeling (SEM). The principal aim of this article is to present a theoretical review of this technique. Despite its possible fascination, SEM should be used according to the previously defined research problem.

KEY WORDS

Constructs; Structural equation modeling; Marketing; Quantitative research.

Sabe-se que uma das fontes da construção do conhecimento são os resultados de pesquisas, principalmente no caso do saber científico. Entretanto, é interessante notar que, especialmente no contexto da Administração, cada vez mais as pesquisas deixam de se restringir à academia, já que também o meio empresarial se guia por elas. Outro aspecto evidente é que, dada a enorme complexidade do "mundo real", é imprescindível esboçar modelos que se aproximem o máximo possível da realidade. Assim, os modelos ajudam a entender se existem e como se dão as relações entre pessoas, fatos, idéias, conceitos e outros.

Os modelos são, portanto, uma "tentativa" de se explicar como a realidade se comporta. Cabe, no entanto, verificar se realmente o que se imagina (o modelo esboçado) traduz a realidade. Nesse contexto, surge a modelagem por equações estruturais. Structural Equation Modeling (SEM) é uma abordagem estatística para testar hipóteses a respeito de relações entre variáveis latentes e observadas.

OS MODELOS SÃO, PORTANTO, UMA "TENTATIVA" DE SE EXPLICAR COMO A REALIDADE SE COMPORTA. CABE, NO ENTANTO, VERIFICAR SE REALMENTE O QUE SE IMAGINA (O MODELO ESBOÇADO) TRADUZ A REALIDADE.

O presente artigo tem por objetivo fazer uma revisão teórica dos principais aspectos da modelagem de equações estruturais.

Hoyle (1995) explica que a modelagem de equações estruturais se inicia com a especificação do modelo a ser estimado. Entende-se modelo como uma proposição estatística de relações entre variáveis.[1] Sabe-se que a análise fatorial exploratória, por exemplo, inicia-se sem um modelo explicitado, mas com decisões a respeito de quantos fatores extrair, como extraí-los e qual método de rotação deve ser usado e que empregam, implicitamente, a especificação de um modelo.

Um modelo linear de equações estruturais é um caminho hipotético de relações lineares entre um conjunto de variáveis. A proposta de tal modelo é fornecer uma explicação que faça sentido e que seja parcimoniosa para as relações observadas (MacCallum, 1995).

Segundo o autor, na modelagem de equações estruturais a especificação do modelo envolve a formulação de proposição sobre um conjunto de parâmetros. Em tal contexto, os parâmetros que requerem especificação são constantes que indicam a natureza da relação entre duas variáveis. Apesar de a especificação poder ser muito particular, de acordo com a magnitude e o sinal dos parâmetros, estes geralmente são especificados como fixos ou livres. Parâmetros fixos não são estimados a partir dos dados e seus valores são tipicamente fixados em zero. Parâmetros livres são estimados a partir dos dados e são aqueles que o investigador acredita serem diferentes de zero. Os vários índices de adequação dos modelos, particularmente o teste qui-quadrado, indicam o grau no qual o caminho de parâmetros fixos ou livres especificados no modelo é consistente com o caminho das variâncias e covariâncias relativas a um conjunto de dados observados.

O caminho de parâmetros fixos ou livres em um modelo por equações estruturais define dois componentes de um modelo geral de equações estruturais: o modelo de mensuração e o modelo estrutural. O modelo de mensuração é o componente do modelo geral que contém as variáveis latentes. Variáveis latentes não são observáveis e são geradas pelas covariâncias entre dois ou mais indicadores. Geralmente são chamadas de fatores e são livres de erros randômicos, sendo exclusivamente associadas aos seus respectivos indicadores. O modelo estrutural é o componente do modelo geral que prescreve as relações entre variáveis latentes e observadas que não são indicadores das variáveis latentes (Hoyle, 1995; Hair et al., 1998; Jöreskog e Sörbom, 1996).

MacCallum (1995) explica que variáveis latentes são equivalentes aos fatores comuns da análise fatorial exploratória (AFE) e definidas a partir de um conjunto de indicadores, o que minimiza o erro de mensuração. Como na AFE um modelo não é explicitamente especificado, é interessante (embora não obrigatório) que a AFE preceda a análise fatorial confirmatória (AFC), para que o pesquisador possa descobrir as variáveis latentes e, suportado pela teoria, testar relações entre elas, através da AFC.

Cada associação entre as variáveis tem um valor numérico, que são os valores dos coeficientes de regressão (pesos aplicados às variáveis em equações de regressão linear), se os caminhos (setas) tiverem uma direção. Se forem bidirecionais, tais valores indicam as covariâncias (ou correlações, se as variáveis estiverem padronizadas) entre as variáveis. Esses pesos e covariâncias são os parâmetros do modelo. O principal objetivo do uso de SEM é estimá-los.

Cada variável pode ser chamada de endógena (se recebe uma seta de outra variável do sistema), ou exógena, em caso contrário. Uma característica importante de variáveis endógenas é o fato de que geralmente não se enxerga tal variável como perfeita e completamente explicada pelas variáveis que exercem influência sobre ela. Assim, define-se uma variável endógena também a partir de um erro, que representa a parte da variável endógena que não é levada em consideração pelas influências lineares das outras variáveis do sistema. Esses erros podem ser vistos como, em parte, randômicos e em parte sistemáticos e podem ser explicados como variáveis ou efeitos que não estão sendo considerados no modelo em questão. Fica claro que os erros são variáveis latentes, na medida em que não são diretamente observáveis. Além disso, em muitas aplicações os erros são exógenos, ou seja, não recebem influência de outras variáveis. Existem, no entanto, casos especiais em que se consideram influências direcionais entre os erros (MacCallum, 1995; Hair et al., 1998).

É interessante notar que não são os indicadores que influenciam as variáveis latentes e sim o contrário (MacCallum, 1995; Bagozzi, 1981). As relações entre as variáveis latentes e seus indicadores são usualmente definidas como direcionais, da latente ao indicador. Parâmetros associados com esses efeitos lineares são equivalentes às cargas dos fatores na AFE, ou seja, são coeficientes de regressão que representam a influência linear dos fatores comuns (variáveis latentes) nas variáveis observadas. Assim, cada indicador é uma variável endógena que recebe influência da variável latente que está medindo. Ou seja, convencionalmente, representa-se cada indicador como influenciado por um erro também, que seria o quanto a variável latente não representa o todo da influência sofrida pelo indicador. Fica claro que as variáveis observáveis podem ser parte do modelo estrutural e não somente indicadores de latentes. Nesse caso estão isentas de erros, pois medem exatamente o que se propõem. Justamente por isso, talvez seja melhor usar variáveis latentes com vários indicadores Uma variável latente endógena será geralmente especificada como influenciada, também, por um erro, representando a parte da latente que não é considerada quando da influência linear especificada no modelo. Todo erro no modelo pode ser visto como uma variável latente que exerce uma influência linear na variável a que está associada (MacCallum, 1995).

Autores como MacCallum (1995), Byrne (1995) e Hair et al. (1998) explicam que toda variável exógena do modelo (seja observável, latente ou erros) terá uma variância que é definida como parâmetro do modelo. Variáveis endógenas também têm variâncias, mas não são parâmetros. Contrariamente, as variâncias de variáveis endógenas são estimadas pelas outras variáveis e influências no modelo. Ou seja, a variância de qualquer variável endógena pode ser expressa algebricamente como uma função de variâncias de variáveis exógenas, incluindo os erros e parâmetros associados às influências lineares no modelo. Assim, as variâncias de variáveis endógenas não são parâmetros, mas funções de outros parâmetros do modelo.

Além disso, continuam os autores, qualquer covariância (associações entre variáveis exógenas de qualquer tipo - setas bidirecionais) serão parâmetros do modelo que só podem envolver variáveis exógenas. Não se permite especificar essas associações entre quaisquer variáveis endógenas, uma vez que todas as associações nas endógenas são estimadas por outras variáveis e influências do modelo. Assim como a variância de uma endógena pode ser expressa como uma função de outros parâmetros do modelo, também as covariâncias de endógenas (ou correlações, setas bidirecionais) podem ser expressas da mesma forma.

Chou e Bentler (1995) lembram que os parâmetros do modelo a serem estimados a partir dos dados são coeficientes de regressão e as variâncias e covariâncias das variáveis independentes. MacCallum (1995) também explica que qualquer efeito direcional especificado no modelo constitui uma outra categoria de parâmetros. Tais efeitos direcionais incluem efeitos de latentes em outras latentes, das latentes em seus respectivos indicadores, dos erros em variáveis associadas.

Segundo MacCallum (1995), cada um desses parâmetros[2] é designado como um parâmetro livre, o que significa que seu valor é desconhecido e deve ser estimado, ou como um parâmetro fixado, o que significa que lhe é dado um valor numérico no modelo original. Para parâmetros livres, também é possível definir restrições que envolvam a estimativa de parâmetros individuais ou combinações de parâmetros. Fica claro que tais aspectos da especificação de modelos devem estar cuidadosamente justificados na teoria ou no objetivo de pesquisa. Por exemplo, restrições de igualdade entre parâmetros são particularmente úteis em modelos longitudinais e em modelos que se ajustem simultaneamente a amostras múltiplas. Em modelos longitudinais, restrições e igualdade podem ser usadas para a formulação de hipóteses de não-variabilidade (invariance) de influências lineares entre as variáveis em pontos diferentes do tempo ou através de intervalos sucessivos de tempo. Em análises com amostras múltiplas, restrições de igualdade são geralmente usadas para testar igualdade de parâmetros de um modelo em grupos distintos de indivíduos. Os valores de parâmetros fixados são geralmente definidos com base em necessidades da especificação do modelo.

Em consonância com Rigdon (2001) e Hair et al. (1998), MacCallum (1995) nota que uma premissa crítica é que sejam estabelecidas escalas para cada variável latente nos modelos, incluindo-se os erros. Afinal, variáveis latentes são construtos não diretamente mensuráveis e, assim, não têm escala de medida.

No entanto, porque o que se deseja é estimar valores dos parâmetros que representam associações entre variáveis latentes e associações entre latentes e observáveis, é essencial que cada variável latente tenha uma escala definida. A escala de latentes é necessária para se saber o quanto um acréscimo de uma unidade de medida em uma variável exógena a uma variável endógena vai impactá-la. Segundo o autor, tal objetivo pode ser alcançado de duas formas. Na primeira, fixa-se a variância de cada latente em um valor numérico específico, tipicamente em 1,0. Cada variável latente que tenha esse parâmetro fixado em 1,0 é definida como variável padronizada, o que pode simplificar grandemente a interpretação da estimação de parâmetros subseqüentes. Estabelecer uma escala para cada construto dessa maneira permite ao pesquisador interpretar os coeficientes associados aos efeitos direcionais entre variáveis latentes como pesos de regressão padronizados e aqueles associados com relações não direcionais como correlações. Se variáveis observáveis também tiverem sido padronizadas (isto é, se o modelo é ajustado via matriz de correlação - a matriz observada Prelis 2 é de correlação), então os coeficientes associados às relações entre latentes e observáveis, ou entre observáveis, também podem ser interpretados como padronizados. Assim, tal procedimento de escala é recomendável para cada construto do modelo.

MacCallum (1995) mostra que uma segunda maneira de estabelecer uma escala para as latentes é fixar o valor de um parâmetro associado à influência direcional emitida pela latente em questão. Esse procedimento é recomendado para cada erro do modelo. Efetivamente, isso significa simplesmente que se dá o valor de 1,0 para a influência de cada erro nas sua variável endógena associada (é importante lembrar que tal variável endógena associada ao erro pode ser um indicador). Deve-se lembrar que, em modelos típicos, os erros são variáveis exógenas e, portanto, suas variâncias são parâmetros a serem estimados. Assim, a estimativa de cada parâmetro de variância do erro dirá quanta variância na variável endógena associada (que pode ser um indicador) não é considerada pelas outras influências no modelo. Esse caso pode ser exemplificado pela fixação do peso 1,0 na seta entre indicadores e seus respectivos erros.

Finalmente, deve-se estabelecer uma escala para uma variável latente endógena. A variância de uma variável endógena latente, conforme já mencionado, não é um parâmetro. É deduzida a partir de outras variáveis e influências no modelo. Para esse caso, têm-se também duas opções. Pode-se usar o segundo procedimento acima descrito e fixar em 1,0 um parâmetro que represente a influência da variável latente endógena em questão em outra variável. Isso é feito geralmente através da fixação em 1,0 do parâmetro que represente a influência da latente endógena em um de seus indicadores (em uma das setas da latente endógena para seus indicadores o peso 1,0 é fixado; isso é requerido, sob pena de o modelo não convergir). Uma segunda alternativa é fixar a variância deduzida da latente endógena em 1,0. No entanto, uma vez que a variância deduzida não é um parâmetro, esse procedimento na verdade representa introduzir uma restrição em um outro parâmetro a ser estimado. Assim, o segundo procedimento mostra-se mais desejável porque permite definir variáveis endógenas como padronizadas, o que simplifica a interpretação das estimativas dos parâmetros. No entanto, tal abordagem não é amplamente disponível nos softwares de SEM. Ou seja, resta usar o primeiro procedimento descrito para se estabelecer uma escala para uma variável latente endógena.

Hoyle (1995), Chou e Bentler (1995), Hair et al. (1998) e Rigdon (2001) clamam por especial atenção a um aspecto importante, mas difícil, no processo de especificação de um modelo: o da identificação.

Para se ter um entendimento básico a esse respeito é necessário considerar, primeiramente, os aspectos fundamentais do processo de estimação de parâmetros. Segundo MacCallum (1995), o quadro mostrado na SEM compõe-se simplesmente de uma relação matemática entre os parâmetros de um modelo, de um lado, e as variâncias/covariâncias das variáveis observáveis (dados coletados), de outro. Na prática, observa-se uma amostra da qual se retiram dados (os indicadores do modelo) para gerar a matriz de variâncias/covariâncias da amostra. Ou seja, dadas as variâncias/covariâncias para as variáveis observáveis (dados, indicadores) e dado o modelo especificado, deseja-se achar valores de parâmetros do modelo que reproduzam as variâncias/covariâncias observadas. Infelizmente, na prática, uma solução geralmente não pode ser encontrada de forma a conseguir um ajuste exato entre o modelo e os dados observados. Assim, os valores dos parâmetros são estimados a partir dos dados coletados da amostra para se obter uma solução cujas variâncias/covariâncias reconstruídas a partir dos parâmetros estimados do modelo especificado se aproximem ao máximo dos valores correspondentes da amostra. Essa é a função primária dos softwares de SEM.

Durante tais procedimentos, continua o autor, os parâmetros estimados do modelo são obtidos usando funções complexas da variância/covariância da amostra. Assim, para cada parâmetro livre (a ser estimado), é necessário que pelo menos uma solução algébrica possa ser obtida de forma a expressar que os parâmetros livres são função da variância/covariância da amostra. Parâmetros que satisfaçam tal condição são ditos identificados, e

parâmetros para os quais há mais de uma solução distinta são ditos superidentificados. Para essas classes de parâmetros é possível chegar a uma única solução. Na SEM, modelos com um ou mais parâmetros superidentificados são de particular interesse, pois é somente para tais modelos que o aspecto da correspondência entre o modelo e os dados faz sentido. Um modelo sem nenhum parâmetro superidentificado sempre irá se ajustar perfeitamente, tornando-se sem sentido descobrir a plausibilidade do modelo através da avaliação do ajuste (MacCallum, 1995; Hair et al., 1998; Rigdon, 2001). Modelos que contenham parâmetros superidentificados em geral não se ajustam exatamente aos dados e podem criar uma possibilidade criticamente importante, que é quando um modelo se acha pobremente ajustado aos dados. Fica claro que descobrir um ajuste melhor faz sentido somente quando tal situação existir.

Se não for possível expressar algebricamente um parâmetro livre como função das variâncias/covariâncias da amostra, tal parâmetro é chamado de subidentificado. Um modelo que contém tal situação na prática não pode ser usado. Infelizmente, a determinação de tal propriedade (identificação) para cada modelo em particular pode ser uma tarefa muito árdua. Não há nenhum conjunto simples de condições suficientes e necessárias que forneça meios de verificação de identificação dos parâmetros de um modelo. No entanto, MacCallum (1995) menciona que duas condições necessárias sempre devem ser checadas. Primeiro, como mencionado anteriormente, uma escala deve ser estabelecida para cada variável latente do modelo. Se tal condição não for satisfeita, um ou mais parâmetros não poderão ser identificados (especificamente, para uma variável latente exógena, sua variância e os coeficientes associados com todos os caminhos emitidos pela latente devem se mostrar não identificados; para uma latente endógena, a variância residual e os coeficientes associados com todos os caminhos que chegam ou que saem da latente endógena serão não identificados). Em segundo lugar, o número efetivo dos parâmetros do modelo não pode exceder o número de variâncias/covariâncias deduzidas das variáveis observáveis coletadas, que é $p(p+1)/2$. Se tal condição for violada, o pesquisador tem menos dados do que parâmetros a serem estimados, o que causará subidentificação.[3] Essas duas condições são necessárias, mas não suficientes. Problemas de identificação ainda podem acontecer, mesmo se nenhuma das duas condições for violada. Geralmente, quando um problema de identificação ocorre, o software aponta os parâmetros envolvidos.

MacCallum (1995) nota que, se um modelo consegue se ajustar a qualquer conjunto de variâncias/covariâncias observadas de maneira perfeita, então o modelo não é desconfirmável. Modelos assim geralmente têm tantos ou mais parâmetros do que variâncias/covariâncias observadas. Como já mencionado, tal modelo não é cientificamente interessante, uma vez que se mostra tão complexo como os dados observados e, portanto, não apresenta nenhuma proposta útil em termos de explicar a estrutura subjacente aos dados de maneira mais parcimoniosa. Para um modelo ser desconfirmável, o número de parâmetros efetivos deve ser menor que o número de variâncias/covariâncias observadas, significando que o modelo terá graus de liberdade (gl) positivos, uma vez que o número de graus de liberdade (gl) é igual ao número de variâncias/covariâncias menos o número efetivo de parâmetros. Modelos razoavelmente especificados com muitos parâmetros e, portanto, um número de graus de liberdade baixo tendem a se ajustar muito bem aos dados e, assim, não tendem a ser desconfirmáveis. Por outro lado, modelos com um número baixo de parâmetros relativos ao número de variâncias/covariâncias observadas tendem a ser altamente desconfirmáveis.

Para tais modelos, um mau ajuste para os dados observados é inteiramente possível. Assim, quando um bom ajuste acontece, conclusões de que o modelo é uma representação plausível dos dados podem ser feitas, de maneira mais confiável. Nota-

se, no entanto, que num modelo que não possa ser desconfirmado em algum grau razoável, o bom ajuste não faz sentido. Dessa forma, na especificação de modelos, os pesquisadores são encorajados a não se esquecerem do princípio da desconfirmação e a construírem modelos não muito parametrizados (MacCallum, 1995; Hair et al., 1998; Jöreskog e Sörbom, 1996). Adicionalmente, é essencial na caracterização do ajuste do modelo usar medidas de ajuste que levem em consideração o grau de desconfirmação do modelo. Caso esses índices não sejam usados, pode-se simplesmente (e erroneamente) concluir que, quanto mais altamente parametrizado o modelo, melhor, pois se ajustam melhor aos dados.[4] Segundo Browne e Cudeck (1989),[5] citados por Hu e Bentler (1995), os pesquisadores devem selecionar modelos menos complexos para amostras pequenas e somente aumentar a complexidade à medida que o tamanho da amostra também puder ser aumentado.

Já foi mencionado que os dados coletados (variáveis observáveis) são transformados em uma matriz de covariância (preserva as escalas) ou correlação (padroniza as escalas, sem as preservar). A partir do modelo especificado, uma outra matriz de covariância (ou correlação) será estimada e a seguir comparada à matriz de covariância dos dados coletados, a fim de se determinar se ocorre (e em que grau ocorre) a discrepância entre elas. Para a estimação da matriz de parâmetros, podem ser usados vários métodos de estimação fornecidos pelos softwares. Sob a premissa de normalidade multivariada, os métodos de estimação mais comumente usados são os interativos, como o de máxima verossimilhança (Maximum Likelihood - ML) e mínimos quadrados generalizados (Generalized Least Square - GLS). No caso de não-normalidade multivariada, recomenda-se o uso de mínimos quadrados ponderados generalizados (Weighted Least Square - WLS), também chamados de livres de distribuição assintótica (Assymptotic Distribution Free - ADF). Jöreskog e Sörbom (1996) afirmam que o método WLS requer que se estime uma matriz assintótica[6] de covariância de dados, o que pode ser feito através de comandos específicos no Prelis 2.

No entanto, nota-se que usar um método de estimação como o ADF tem limitações práticas, a saber: (i) o cálculo das estimativas é computacionalmente caro. Bentler (1989),[7] apud West, Finch e Curran (1995), mostra que se torna praticamente inviável para modelos com mais de 20 ou 25 variáveis, mesmo levando-se em conta a maior capacidade de processamento dos computadores atuais; (ii) o cálculo da matriz de momentos de quarta ordem, necessário para o método ADF, requer amostras grandes para produzir estimativas estáveis (West, Finch e Curran, 1995; Hair et al., 1998). Hu e Bentler (1995, p. 96) afirmam que "...ADF pode ser confiável apenas para tamanhos de amostra maiores que 5000"[8] (Traduzido pela autora).

A COMPARAÇÃO ENTRE A MATRIZ ESTIMADA E A OBSERVADA RESULTA EM UMA MATRIZ RESIDUAL. A MATRIZ RESIDUAL CONTÉM ELEMENTOS CUJOS VALORES SÃO AS DIFERENÇAS ENTRE OS VALORES CORRESPONDENTES DAS MATRIZES ESTIMADA E OBSERVADA.

Hoyle (1995) explica que a iteração começa com um conjunto de valores de início (start), valores de tentativa para os parâmetros livres dos quais uma matriz de covariância estimada pudesse ser computada e comparada à matriz de covariância dos dados observados. Os valores de início também são fornecidos pelo pesquisador ou, mais comumente, pelo software. Depois de cada iteração, a matriz estimada resultante é comparada à matriz observada. A comparação entre a matriz estimada e a observada resulta em uma matriz residual. A matriz residual contém elementos cujos valores são as diferenças entre os valores correspondentes das matrizes estimada e observada. As interações continuam até não ser mais possível melhorar as estimativas dos

parâmetros e produzem uma matriz de covariância sugerida cujos elementos estão o mais próximo possível em termos de magnitude e de direção dos elementos correspondentes da matriz de covariância observada. Assim, as interações continuam até os valores dos elementos da matriz residual não poderem mais ser minimizados. Nesse ponto, o procedimento de estimação converge. Nota-se que problemas de não-convergência podem ser comuns. Segundo Hoyle (1995, p. 6),

> *Quando a convergência acontece, um único número é produzido, que resume o grau de correspondência entre as matrizes de covariância estimada e observada. Esse número, às vezes chamado de valor da função de ajuste, se aproxima de zero na medida em que a matriz de covariância estimada se parece com a observada. Um ajuste perfeito entre as duas matrizes produz um valor da função de ajuste que é o ponto de partida para a construção de índices de ajustamento do modelo. (Traduzido pelos autores)*[9]

O autor explica que um modelo se ajusta aos dados na medida em que a matriz de covariância estimada pelo modelo é equivalente à matriz de covariância obtida através dos dados coletados, ou seja, que os elementos da matriz residual são próximos de zero. A questão de ajuste é, obviamente, uma questão estatística em que se deve levar em consideração características da matriz observada advinda dos dados coletados, o modelo e o método de estimação utilizado. Por exemplo, se a matriz de covariância observada é tratada como representante da população, a matriz sofrerá de erros de amostragem que aumentam à medida que o tamanho da amostra diminui. Ainda, conforme já mencionado, quanto maior o número de parâmetros livres no modelo, mais provavelmente o modelo se ajustará aos dados, pois a estimativa dos parâmetros é derivada dos dados. Para complicar ainda mais, os diferentes métodos de estimação variam em termos de efetividade à medida que o tamanho da amostra e a complexidade do modelo aumentam.

Segundo Jöreskog e Sörbom (1996), o teste de ajuste global do modelo, chamado qui-quadrado, é calculado como sendo (N-1) vezes o valor mínimo encontrado para a função de ajuste (F), sendo N o tamanho da amostra. Assim, o qui-quadrado é, na verdade, uma medida de má qualidade do ajuste, já que valores menores do qui-quadrado mostram um ajuste melhor, sendo zero o ajuste perfeito (a matriz residual será composta somente de zeros).

Segundo Hu e Bentler (1995), a estatística T (usualmente referida como teste c+) segue uma distribuição c+ assintótica. Uma estatística T com um valor alto relativo ao graus de liberdade associados ao modelo indica que o modelo pode não ser uma boa representação do processo que gerou os dados da população.

No entanto, os autores mostram que várias pesquisas descobriram rapidamente problemas associados ao teste c+ para ajustamento de modelos. Uma das preocupações diz respeito ao tamanho da amostra. A teoria estatística para T é assintótica, o que torna o teste sensível ao tamanho da amostra. Assim, T pode não seguir a distribuição c+ nos casos de amostras pequenas e, assim, pode não se mostrar correto para avaliação de modelos em situações práticas (em que o tamanho da amostra é geralmente pequeno). Por outro lado, com o poder estatístico do teste aumentado pelo tamanho maior da amostra, uma diferença trivial entre a matriz de covariância dos dados amostrais e a matriz estimada pode resultar na rejeição do modelo especificado[10] (Hu e Bentler, 1995). Além disso, T pode também não seguir a distribuição c+ em casos de violação da premissa de normalidade multivariada. Hoyle (1995, p. 6) menciona que, dadas as condições da pesquisa em ciências sociais e comportamentais, em que raramente se conseguem dados normais, e dada a dificuldade de se obter uma amostra grande, além da dificuldade de se especificar um modelo totalmente isento de erros de especificação (que

são comuns, como evidenciam Hair et al., 1998), pelo menos uma dessas premissas é violada na maioria dos estudos que usam a modelagem de equações estruturais.

O teste estatístico para o ajustamento segue o qui-quadrado, que tem como hipótese nula (H0) a igualdade entre a matriz de covariância dos dados e a matriz de covariância estimada. A hipótese alternativa (H1) é que a matriz de covariância dos dados e a matriz de covariância estimada a partir do modelo proposto são diferentes. Fica claro que o que se pretende é que H0 não seja rejeitada. Como exemplo, para um nível de confiança igual a 5% (a=0,05), tem-se que, se o p-valor for menor do que a (ou seja, se p-valor<0,05), rejeita-se a hipótese nula. Por outro lado, sendo p-valor maior que a (p-valor > 0,05), não há como rejeitar a hipótese nula em favor da hipótese alternativa. Nesse caso, ao aceitar a hipótese nula (p-valor>a), confirma-se o ajustamento do modelo proposto.

Para Hoyle (1995), a insatisfação crescente com o teste qui-quadrado tem conduzido à criação de um número crescente de índices de ajuste adicionais, que são índices descritivos de ajuste freqüentemente interpretados de maneira intuitiva. Em vez de compararem as matrizes de covariância observada e estimada, esses índices partem da comparação entre o ajuste de um modelo especificado e o ajuste de um modelo independente ou nulo. O modelo nulo é aquele em que nenhuma relação entre as variáveis é especificada. Ou seja, todos os caminhos relacionais são fixados em zero e somente as variâncias são estimadas. Assim, a maioria dos índices de ajuste adicionais reflete uma melhoria no ajuste de um modelo especificado, que inclui parâmetros estruturais fixos e livres, em relação a um modelo nulo, em que todos os parâmetros estruturais são fixados em zero. Nota-se que os índices de ajuste adicionais não são estatísticos e, portanto, não podem ser usados para conduzir testes estatísticos formais de ajuste de modelo. Em vez disso, são tratados como índices globais de adequação do modelo. A maioria deles varia entre zero e 1,0, sendo que 0,90 é largamente aceito como um valor que tais índices devem exceder antes de um modelo ser visto como consistente com os dados observados. No entanto, o autor reconhece que, como não são testes estatísticos, não existe valor de corte específico para tais índices.

Segundo Hair et al. (1998), as medidas de ajuste são de três tipos:

a) medidas de ajuste absoluto: para verificar o ajuste global do modelo (Quadro 1);

b) medidas de ajuste incremental: para a comparação entre o modelo proposto e um outro, especificado pelo pesquisador (Quadro 2);

c) medidas de ajuste parcimonioso: para adequar as medidas de ajuste de forma a fornecer uma comparação entre modelos com número diferente de parâmetros estimados, tendo como proposta determinar a magnitude do ajuste resultante de cada parâmetro estimado. O objetivo básico dessas medidas é verificar se o ajuste do modelo foi obtido através de um superajuste dos dados devido à grande quantidade de parâmetros a serem estimados. Assim, tais medidas relacionam a qualidade do ajuste do modelo ao número necessário de parâmetros a serem estimados para obter esse nível de ajuste. Depreende-se disso que níveis maiores de parcimônia são desejados (Quadro 3).

Finalmente, explica-se que existem outras medidas, como o qui-quadrado normalizado, sugerido por Jöreskog (1970)[12] apud Hair et al. (1998). Para calculá-la, divide-se o qui-quadrado pelos graus de liberdade (c+/gl), normalizando-o. Tal medida deve ser maior que 1,0 e menor que 2,0 ou 3,0. Valores que excedam 2,0 ou 3,0 mostram um ajuste inadequado dos valores estimados em relação aos dados.

Hu e Bentler (1995) e Hair et al. (1998, p. 615) explicam que uma das formas de se verificar a necessidade de reespecificação de um modelo é através da análise da matriz de resíduos gerada, referente aos resíduos, que representam as diferenças entre a matriz de covariância (ou de correlação) de dados e a matriz de covariância (ou de correlação) estimada. Os resíduos padronizados

QUADRO 1 - Medidas de ajuste absoluto

Medidas de ajuste absoluto	Definição	Valores esperados
Qui-quadrado (c^2)	É a principal medida para o grau de ajuste global do modelo, sendo a única medida estatística. Valores altos do qui-quadrado em relação aos graus de liberdade significam que as matrizes de dados observados e estimada diferem consideravelmente. Assim, o pesquisador está à procura de níveis estatísticos não significativos, em termos de diferença de matrizes, sendo importante enfatizar que tal interpretação difere do desejo habitual de se encontrar significância estatística. Mesmo se for encontrada a não-significância estatística, isso não garante que o modelo correto foi encontrado, apenas que o modelo proposto se ajusta às covariâncias dos dados observados. Note-se, também, que o teste qui-quadrado é muito sensível ao tamanho da amostra, especialmente para amostras superiores a 200, o que pode fazer com que os resultados sejam distorcidos.	Valores baixos do qui-quadrado resultam em um nível de significância (p) maior do que 0,05, por exemplo, o que impossibilita a rejeição da hipótese nula (as matrizes de dados e estimada são estatisticamente iguais) e é exatamente isso que é desejável no caso de modelagem de equações estruturais indicando que existe ajuste adequado.
Parâmetros de não-centralidade (NCP)[1]	Resulta da tentativa de estatísticos em encontrar uma medida alternativa ao qui-quadrado que fosse menos sensível ao tamanho da amostra. Recomenda-se seu uso na comparação de modelos NCP= c^2- gl	Quanto menor seu valor, melhor, pois isso indica que as matrizes de dados e estimada não diferem consideravelmente.
Parâmetros de não-centralidade ajustados (SNCP)[2]	É a medida anterior (NCP) padronizada pelos graus de liberdade. Assim SNCP= (c^2- gl)/N, sendo N o tamanho da amostra.	Quanto menor seu valor, melhor, pois isso indica que as matrizes de dados e estimada não diferem consideravelmente.
Raiz quadrada da média dos resíduos ao quadrado (RMSR)[4]	É uma média dos resíduos entre as matrizes de dados coletados e a matriz estimada. É uma medida mais útil quando todas as variáveis observadas estão padronizadas, por isso recomenda-se o uso de matrizes de correlação.	Não há níveis aceitáveis preestabelecidos, cabendo ao pesquisador fixá-los de acordo com os objetivos da pesquisa. Nota-se, no entanto, que quanto menores os valores, melhor.
Raiz quadrada da média dos quadrados dos erros de aproximação (RMSEA)[5]	Medida semelhante à RMSR, diferindo no sentido em que a discrepância das matrizes é medida em relação à população e não à amostra. Portanto, o valor é representativo da qualidade de ajuste esperada se o modelo fosse estimado na população.	Valores que variem de 0,05 a 0,08 são considerados aceitáveis.
Índice esperado de validação cruzada (ECVI)[6]	É uma aproximação da qualidade de ajuste que o modelo estimado apresentaria em uma outra amostra, de igual tamanho.	Não há intervalos especificados para valores aceitáveis, sendo mais bem utilizada na comparação de modelos alternativos.
Índice de validação cruzada (CVI)[7]	Verifica a qualidade de ajuste quando uma validação cruzada é elaborada de fato.	Não há intervalos especificados para valores aceitáveis, sendo mais bem utilizada na comparação de modelos alternativos.

FONTE: Hair *et al.*, 1998, p. 654-656.
NOTAS: [1] *Noncentrality parameters*, no original em inglês.
[2] *Scaled noncentrality parameter*, no original em inglês.
[3] *Goodness-of-fit index*, no original em inglês.
[4] *Root mean square residual*, no original em inglês.
[5] *Root mean square error of approximation*, no original em inglês.
[6] *Expected cross validation index*, no original em inglês.
[7] *Cross validation index*, no original em inglês.

QUADRO 2 - Medidas de ajuste incremental

Medidas de ajuste absoluto	Definição	Valores esperados [11]
Índice ajustado de qualidade de ajuste (AGFI)[1]	É uma extensão do GFI, ajustada através da razão dos graus de liberdade do modelo proposto pelos graus de liberdade do modelo nulo.	Valores maiores ou iguais a 0,90 indicam níveis aceitáveis de ajuste
Índice de Tucker-Lewis (TLI ou NNFI)[2]	Combina uma medida de parcimônia em um índice comparativo entre os modelos proposto e nulo.	Valores maiores ou iguais a 0,90 indicam níveis aceitáveis de ajuste.
Índice de ajuste normalizado (NFI)[3]	Compara o modelo proposto ao modelo nulo. Hu e Bentler (1995) mostram que o NFI não é um bom índice para N pequenos, sendo N o tamanho da amostra.	Valores maiores ou iguais a 0,90 indicam níveis aceitáveis de ajuste.
Índice comparativo de ajuste (CFI)[4]	Também compara o modelo proposto com o modelo nulo.	Valores mais próximos da unidade indicam melhor ajuste.

FONTES: Hair *et al.*, 1998, p. 657.
Hu e Bentler, 1995, p. 76-99.
NOTAS: [1] *Adjusted goodness-of-fit*, no original em inglês
[2] *Tucker Lewis index*, no original em inglês.
[3] *Normed Fit index*, no original em inglês.
[4] *Comparative Fit Index*, no original em inglês.

QUADRO 3 - Medidas de parcimônia de ajuste

Medidas de ajuste absoluto	Definição	Valores esperados
Índice de parcimônia de ajuste normalizado (PNFI)[1]	Esse índice utiliza em seu cálculo o número de graus de liberdade necessário para se obter um certo nível de ajuste. É uma medida usada para comparar modelos alternativos com diferentes graus de liberdade.	Não há recomendações de níveis aceitáveis de ajuste. No entanto, diferenças de 0,06 a 0,09 podem indicar diferenças substanciais entre os modelos alternativos.
Índice de qualidade de ajuste parcimonioso (PGFI)[2]	Esse índice modifica o GFI, ajustando-o sob o aspecto de parcimônia do modelo.	Valores mais altos indicam maior parcimônia do modelo.

FONTE: Hair *et al.*, 1998, p. 658.
NOTAS: [1] *Parsimonious normed fit index*, no original em inglês.
[2] *Parsimonious goodness-of-fit index*, no original em inglês.

são divididos pelos respectivos erros-padrão estimados e tornam-se, portanto, independentes das unidades de medidas das variáveis. Como seguem aproximadamente uma distribuição normal ou uma distribuição "z", valores maiores ou menores do que |2,58| são significativos, podendo indicar erros de especificação de modelos, não-normalidade das variáveis ou a presença de relações não lineares entre variáveis (Jöreskog e Sörbom, 1996; Hair et al., 1998; Rigdon, 2001).

Os autores sugerem ainda que seja observado o gráfico Q-plot dos resíduos (presente na saída do Lisrel). Se os pontos forem plotados de forma a não se aproximarem de uma reta cujo coeficiente angular é 1,0 (45°), isso é um indicativo de muitos resíduos significativos e de um ajuste não muito adequado.

O último aspecto a ser mencionado sobre a modelagem de equações estruturais refere-se aos índices de modificação. Quando um modelo não apresenta ajustes adequados, os softwares sugerem modificações no modelo de forma a aumentar os parâmetros e, conseqüentemente, melhorar o ajuste através da diminuição de valores do qui-quadrado. É consenso entre os autores que tais índices de modificação devem ser usados somente se apoiados pela teoria. Além disso, após reespecificado o modelo, índices de ajuste apresentados nada significam,[13] na medida em que, para serem calculados, os mesmos dados foram aproveitados. O procedimento, então, é feito com base nos dados, impedindo, assim, a generalização e a replicação do modelo. Para ter validade, o modelo reespecificado deve ser testado com outros (novos) dados através da validação cruzada.

CONSIDERAÇÕES FINAIS

De tudo o que foi exposto pode-se perceber que a modelagem de equações estruturais é, sem dúvida, uma técnica avançada de tratamento e análise estatística de dados. Suas características próprias permitem pesquisas refinadas. Mas os pesquisadores não se podem deixar seduzir pelo uso dessa técnica. Como em qualquer estudo científico bem estruturado, é imprescindível que o problema a ser investigado preceda e promova a escolha da técnica a ser utilizada. Além disso, ao considerar o uso da modelagem de equações estruturais, o pesquisador deve, ainda, ponderar a respeito de outros aspectos peculiares, como, por exemplo, a necessidade de se ter amostras grandes e cada vez maiores à medida que os modelos se tornam complexos. Tudo isso, juntamente com um software adequado, possibilita, sem dúvida, que trabalhos inovadores sejam desenvolvidos e contribuam para que o conhecimento, em suas diversas facetas, seja aprimorado.

Recebido em: mai./2003. Aprovado em: jun./2003

Marlusa Gosling
M.Sc., doutoranda em Administração - Cepead/UFMG
Profª e pesquisadora da Faculdade Batista de Minas Gerais
Telefone: (31) 3422.4487
E-mail: marlusa@uai.com.br

Carlos Alberto Gonçalves
Doutor em Administração na USP
Prof. e Pesquisador do Cepead/UFMG
Telefone: (31) 3279.9040
E-mail: carlos@face.ufmg.br

NOTAS

[1] Jöreskog e Sörbom (1996, ajuda do Lisrel 8.3 *on line*, via *software*) explicam que, usualmente, assume-se que as relações estruturais são lineares, mas alguns modelos não lineares também têm sido propostos.

[2] O número efetivo de parâmetros de um modelo pode ser definido como o número de parâmetros livres menos o número de restrições impostas a esses parâmetros. Além disso, sabe-se que para P variáveis observáveis de um modelo, o número de variâncias e covariâncias existente é p (p + 1)/2.

[3] Hair *et al.* (1998) explicam que, muitas vezes, ao se determinar que cada construto deve ter no mínimo três indicadores, os problemas de subidentificação são corrigidos.

[4] MacCallum (1995) encoraja, então, os pesquisadores a considerarem um índice como o RMSEA, que é essencialmente uma medida de ausência de ajuste por grau de liberdade

[5] Browne, M. W., Cudeck, R. Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, v. 24, p. 445-455, 1989

[6] Diz-se que uma distribuição é assintótica quando depende do tamanho da amostra.

Bentler, P. M. *EQS structural equations program manual*. Los Angeles: BMDP Statistical Software, 1989.

[7] Original em inglês.

[8] Original em inglês.

[9] Rigdon (2001), em mensagem enviada em 28 de agosto de 2001 ao grupo de discussão da internet denominado SEMNET, afirma que: " ...um exemplo recente em que um tamanho de amostra N bastante grande não levou à rejeição de um modelo. [...] com N = 1115 e três modelos separados e pequenos, para diferentes modos de transporte, os autores mostram qui-quadrado de 3,17; 14,81 e 1,32, com graus de liberdade igual a 3, para cada modelo" (original em inglês, traduzido pela autora, grifo nosso). A referência do artigo é a seguinte: Fan Yang-Wallentin, Peter Schmidt and Sebastian Bamberg. Testing interactions with three different methods in the theory of planned behavior: analysis of traffic behavior data. In: Robert Cudeck, Stephen du Toit and Dag Sorbom (eds.), *Structural equation modeling*: present and future: a festschrift in honor of Karl Jöreskog. Chicago: Scientific Software, 2001, p. 405-423.

[10] Hu e Bentler (1995), através de várias simulações Monte Carlo, questionam tais valores de corte (0,90), salientando que eles dependem do método de estimação usado, do tamanho da amostra, da complexidade do modelo e das características dos dados coletados.

[12] Jöreskog, K. A general method for analysis of covariance structures. *Biometrika*, v. 57, p. 239-251, 1970.

[13] Buscou-se simplesmente um melhor ajuste estatístico e que pode não ter nenhum sentido teoricamente. Existe, inclusive, um *software* que tenta chegar ao melhor ajuste possível, o Tetrads.

REFERÊNCIAS BIBLIOGRÁFICAS

AAKER, David A.; BAGOZZI, Richard P. Unobservable variables in Structural Equations Models with an application in industrial selling. Journal of Marketing Research, v. 16, p.147-158, May, 1979.

ARBUCKLE, James L.; WOTHKE, Werner. Amos 4.0 user's guide. Chicago: SmallWaters Corp., 1995.

BABAKUS, Emin; FERGUSON Jr., Carl E.; JÖRESKOG, Karl G. The sensivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. Journal of Marketing Research, v. 23, p. 222-228, May, 1987.

BAGOZZI, Richard P. Evaluating Structural Equation Models with unobservable variables and measurement error: a comment. Journal of Marketing Research, v. 18, p. 375-381, Aug., 1981.

BAGOZZI, Richard P.; YI, Youjae; PHILLIPS, Lynn W. Assessing construct validity in organizational research. Administrative Science Quaterly, v. 36, n. 3, p. 421-458, Sept., 1991.

BYRNE, Barbara M. One application of Structural Equation modeling from two perspectives: exploring the EQS and Lisrel strategies. In: HOYLE, Rick H (ed.). Structural Equation Modeling: concepts, issues and applications. London: Sage Publications Inc., 1995. cap. 8, p. 138-157.

CHOU, Chih-Ping; BENTLER, Peter M. Estimates and tests in Structural Equation Modeling. In: HOYLE, Rick H (ed.). Structural Equation Modeling: concepts, issues and applications. London: Sage Publications Inc., 1995. cap. 3, p. 37-55.

CHURCHILL, Jr. Gilbert A. A paradigm for developing better measures of marketing constructs. Journal of Marketing Research, v. 16, p. 64-73, Feb., 1979.

DeCARLO, Lawrence T. On the meaning and use of kurtosis. Psychological Methods, v. 2, n. 3, p. 292-307, 1997.

FORNELL, Claes; LARCKER, David F. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, v. 18, p. 39-50, Feb., 1981.

FORNELL, Claes; LARCKER, David F. Structural Equation Models with unobservable variables and measurement error: algebra and statistics. Journal of Marketing Research, v. 18, p. 382-388, Aug., 1981.

HAIR Jr., Joseph F. et al. Multivariate data analysis. 5 ed. New Jersey: Prentice Hall, 1998.

HOYLE, Rick H. (ed). Structural equation modeling: concepts, issues and applications. London: Sage Publications, 1995.

HOYLE, Rick H.; PANTER, Abigail T. Writing about structural equations models. In: HOYLE, Rick H (ed.). Structural equation modeling: concepts, issues and applications. London: Sage Publications Inc., 1995. cap. 9, p. 158-176.

HU-Li-tze; BENTLER, Peter M. Evaluating model fit. In: HOYLE, Rick H (ed.). Structural equation modeling: concepts, issues and applications. London: Sage Publications Inc., 1995. cap. 5, p. 76-99.

JOHNSON, Richard A.; WICHERN, Dean W. Applied multivariate statistical analysis. 4. ed. New Jersey: Prentice Hall, 1998, 816p.

JÖRESKOG, Karl; SÖRBOM, Dag. Lisrel 8: structural equation modeling with the Simplis command language. Chicago: SSI, Inc., 1993.

JÖRESKOG, Karl; SÖRBOM, Dag. Lisrel 8: user's reference guide. Chicago: SSI, Inc., 1996.

JÖRESKOG, Karl; SÖRBOM, Dag. Prelis 2: user's reference guide. Chicago: SSI, Inc., 1996.

McDONALD, Roderick P.; MARSH, Herbert W. Choosing a multivariate model: noncentrality and goodness of fit. Psychological Bulletin, v. 107, n. 2, p. 247-255, 1990.

MUELLER, Ralph O. Basic principles of structural equation modeling: an introduction to Lisrel and EQS. New York: Springer, 1996. 229p.

OLSSON, Ulf Henning et al. The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. Structural Equation Modeling, v. 7, n. 4, p. 557-595, 2000.

RIGDON, Edward E. Structural equation modeling: a guide for consultants. [mensagem pessoal]. Mensagem recebida por <marlusa@ig.com.br <mailto:marlusa@ig.com.br>>, em 7 de maio 2001.

RIGDON, Edward E.; FERGUSON Jr., Carl E. The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. Journal of Marketing Research, v. 28, p. 491-497, 1994.

SCHUMACKER, Randall E.; BEYERLEIN, Susan T. Confirmatory factor analysis with different correlation types and estimation methods. Structural Equation Modeling, v. 7, n. 4, p. 629-636, 2000.

WEST, Stephen G.; FINCH, John F.; CURRAN, Patrick J. Structural equation models with nonnormal variables: problems and remedies. In: HOYLE, Rick H (ed.). Structural equation modeling: concepts, issues and applications. London: Sage Publications Inc., 1995. cap. 4, p. 56-75.