



Métodos de Correção de Testes Estatísticos em Modelagem de Equações Estruturais

Marli Appel da Silva¹ , Irani Iracema de Lima Argimon 

Pontifícia Universidade Católica do Rio Grande do Sul – PUC/RS, Porto Alegre-RS, Brasil

Guilherme Welter Wendt 

Universidade Estadual do Oeste do Paraná – Unioeste, Francisco Beltrão-PR, Brasil

RESUMO

O objetivo deste estudo é comparar diferentes testes estatísticos mediante o aumento amostral em amostras que violem os pressupostos da normalidade na Modelagem de Equações Estruturais (*Structural Equation Modeling*, SEM). Os seguintes métodos foram comparados: teste de hipótese nula ($H_{0_{SB}}$) com qui-quadrado (χ^2_{SB}) corrigido pelo teste escalado de Satorra-Bentler, estratégia sugerida por Moshage e Erdfelder ($T_{M\&E}$) e teste de equivalência proposto por Yuan et al. (T_Y). Trata-se de um estudo simulado e exploratório. Os testes $H_{0_{SB}}$, $T_{M\&E}$ e T_Y revelaram comportamentos distintos, sugerindo que a $H_{0_{SB}}$ possa suportar amostras menores (100 casos) e rejeitar amostras maiores (≥ 500 casos), bem como os testes $T_{M\&E}$ e T_Y possam apoiar modelos com mais casos (≥ 500 casos). Recomenda-se o teste T_Y por apresentar mais recursos para a avaliação de modelos na SEM, embora o teste $T_{M\&E}$ mereça mais estudos sobre o seu desempenho, uma vez que possa ser uma ferramenta adicional importante.

Palavras-chave: método; análise estatística; modelagem de equações estruturais.

ABSTRACT – Correction methods for Structural Equation Modeling statistical tests

This study aimed to compare different statistical tests by increasing the sample size in samples that violate the assumptions of normality in Structural Equation Modeling (SEM). The following methods were compared: null hypothesis test ($H_{0_{SB}}$) – Satorra-Bentler scaled chi-square test (χ^2_{SB}); the strategy suggested by Moshage and Erdfelder ($T_{M\&E}$); and the equivalence test proposed by Yuan et al. (T_Y). This was a simulated, exploratory study. The $H_{0_{SB}}$, $T_{M\&E}$ and T_Y tests revealed distinct behaviors, suggesting that the $H_{0_{SB}}$ can support smaller samples (100 cases) and reject larger samples (≥ 500 cases), and that the $T_{M\&E}$ and the T_Y tests can support models with more cases (≥ 500 cases). The T_Y test is recommended because it presents more resources for evaluating models in SEM, although the $T_{M\&E}$ test deserves further studies on its performance, as it could be an important additional tool.

Keywords: methods; statistical analysis; Structural Equation Modeling.

RESUMEN – Métodos de corrección de pruebas estadísticas en el Modelo de Ecuaciones Estructurales

El objetivo de este estudio es comparar diferentes pruebas estadísticas a través del aumento de la muestra en muestras que violan los supuestos de normalidad en el Modelode Ecuaciones Estructurales (*Structural Equation Modeling* – SEM). Se compararon las siguientes estrategias: prueba de hipótesis nula ($H_{0_{SB}}$) – chi-cuadrado (χ^2_{SB}) corregido por la prueba escalada de Satorra-Bentler; estrategia sugerida por Moshage y Erdfelder ($T_{M\&E}$); y prueba de equivalencia propuesta por Yuan et al. (T_Y). Este es un estudio exploratorio y simulado. Las pruebas $H_{0_{SB}}$, $T_{M\&E}$ y T_Y revelaron comportamientos distintos, lo que sugiere que $H_{0_{SB}}$ puede admitir muestras más pequeñas (100 casos) y rechazar muestras más grandes (≥ 500 casos), así como las pruebas $T_{M\&E}$ y T_Y pueden admitir modelos con más casos (≥ 500 casos). Se recomienda la prueba T_Y porque presenta más recursos para evaluar modelos en SEM, aunque la prueba $T_{M\&E}$ merece más estudios sobre su desempeño, ya que puede ser una herramienta adicional importante.

Palabras clave: método; análisis estadístico; Modelos de Ecuaciones Estructurales.

O objetivo deste estudo é comparar diferentes testes estatísticos mediante o aumento amostral em amostras que violem os pressupostos da normalidade na Modelagem de Equações Estruturais (*Structural Equation Modeling*, SEM). A SEM congrega um conjunto de técnicas estatísticas com a finalidade de estabelecer se um modelo proveniente de uma determinada teoria está próximo de apresentar poder explicativo da realidade, tendo crescente aplicação no campo da Avaliação Psicológica e

contando, inclusive, com periódicos acadêmicos próprios para a disseminação do conhecimento na área da SEM (Kline, 2015; MacCallum & Austin, 2000; Schumacker & Lomax, 2016).

O tamanho amostral constitui fator importante na aceitação e rejeição de modelos na SEM. Amostras maiores (e.g., número amostral $[N] > 1.000$) podem propiciar com que o teste de hipótese nula (ou seja, a verificação de H_0) não possa ser corroborado, gerando um p -valor

¹ Endereço para correspondência: Rua Dr. Flores, 163, cj. 1101, 90020-120, Porto Alegre, RS. E-mail: mappel@uol.com.br

significativo (Yuan et al., 2016; Moshagen & Erdfelder, 2016). O teste de hipótese (H_0), baseado nas abordagens de Fisher e Neyman-Pearson (Perezgonzalez, 2015) é comumente efetuado com base no teste qui-quadrado (χ^2 , teste de hipótese não paramétrico). Usualmente, para que um modelo possa ser aceito como próximo de uma base populacional, seguindo o parâmetro da distribuição central, o p-valor (nível de significância) do teste χ^2 não deve ser significativo ($p \leq 0,05$; Kline, 2015; Schumacker & Lomax, 2016; Tarka, 2018).

Todavia, não existe consenso na literatura acerca de quando um modelo possa ser aceito ou rejeitado na pesquisa empírica em casos em que a hipótese nula (H_0) não possa ser suportada. Dessa maneira, a aceitação ou rejeição de modelos com base no teste de H_0 é tema controverso (Kline, 2015; Perezgonzalez, 2015; Satorra & Bentler, 2001; Yuan et al., 2016).

Logo, visando evitar especialmente os erros de Tipo I (alfa, α ; H_0 verdadeira e rejeitada) ou II (beta, β ; H_0 falsa e aceita), além de prover mais informações sobre os ajustes de modelos, especialmente em amostras maiores – quando a H_0 possa não ser corroborada – alguns autores sugeriram aplicar testes complementares ao teste H_0 (Moshagen & Erdfelder, 2016; Yuan et al., 2016). Nesse caso, a literatura indica duas propostas, a saber: a estratégia sugerida por Moshagen e Erdfelder ($T_{M\&E}$; 2016) e o teste de equivalência proposto por Yuan et al. (T_Y ; 2016). Ambas as propostas ($T_{M\&E}$ e T_Y) visam, com parcimônia, identificar se a extensão da violação do ajuste de um dado modelo é tolerável.

As estratégias citadas são de interesse indiscutível (MacCallum & Austin, 2000). Isso porque amostras que violam os pressupostos da normalidade são encontradas especialmente em estudos intrinsecamente psicossociais, devido à natureza relativamente atípica dos fenômenos estudados (Silva et al., 2018a), bem como, a forma de mensuração dos fenômenos examinados (e.g., escala Likert de autorrelato; Jebb et al., 2021). Em linhas gerais, uma distribuição normal univariada apresenta zero de curtose – que é o grau de “achatamento” de uma distribuição de frequência – e zero de assimetria, que designa o grau de afastamento da média amostral. Consequentemente, amostras não normais apresentam excesso de curtose e assimetria (> 1 ; Gao et al., 2019; Yang & Liang, 2013).

Os testes destinados a amostras normais (e.g., Máxima Verossimilhança – ML), quando aplicados indiscriminadamente em amostras com distribuições não normais, tendem a ocasionar valores inflados do qui-quadrado, aumentando a tendência de que os modelos avaliados sejam rejeitados (Kline, 2015; Lai & Green, 2016; Yang & Liang, 2013). Por conseguinte, torna-se necessária a adoção de testes que corrijam o valor do χ^2 em amostras não normais (Satorra & Bentler, 1994; 2001).

Para amostras não normais, Moshagen e Erdfelder (2016) sugeriram como alternativa o teste escalado de

Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001), considerado robusto e já consagrado na literatura da SEM. O teste T_{SB} corrige o valor do qui-quadrado (χ^2_{SB} , correção do valor do qui-quadrado [χ^2] estimado pela solução Máxima Verossimilhança [ML] pelo teste T_{SB}) de modo que tal valor se aproxime do retratado por uma amostra normal, indicado quando os pressupostos das Teorias da Normalidade, Elíptica (princípio da assimetria igual a zero com certo nível de curtose homogênea – positiva ou negativa) e da curtose heterogênea (distribuições com características de simetria e valores marginais heterogêneos de curtose) não sejam atendidos (Satorra & Bentler, 2001; Silva et al., 2018b; Yuan et al., 2016).

Além disso, Moshagen e Erdfelder (2016) indicaram que, em amostras não normais, a proposta $T_{M\&E}$ seja utilizada em combinação com outros procedimentos robustos, em especial o teste T_{SB} (Satorra & Bentler, 1994; 2001). Já Yuan et al. (2016) informaram que a alternativa T_Y é apropriada para a avaliação do limite tolerável de má especificação de modelos em amostras que violem os pressupostos da normalidade amostral. Assim, ambas as propostas ($T_{M\&E}$ e T_Y) podem ser adequadas para avaliar amostras não normais (Moshagen & Erdfelder, 2016; Yuan et al., 2016).

Portanto, neste estudo, serão comparadas as seguintes estratégias: teste de hipótese nula (H_{0SB}), com o qui-quadrado (χ^2_{SB}) corrigido pelo teste escalado de Satorra-Bentler (T_{SB}); estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016); e o teste T_Y (Yuan et al., 2016), mediante ao aumento amostral, em amostras que violem os pressupostos da normalidade. Será avaliada a aceitação ou rejeição de modelos conforme as estratégias comparadas (H_{0SB} , $T_{M\&E}$ e T_Y).

Todavia, serão inicialmente descritas noções sobre a hipótese nula (H_0), o teste escalado de Satorra-Bentler (T_{SB}), bem como sobre as estratégias $T_{M\&E}$ e T_Y , destacando-se as especificações e requisitos para o uso das técnicas. Posteriormente, por meio de um estudo simulado e exploratório, as estratégias H_{0SB} , $T_{M\&E}$ e T_Y serão comparadas para a verificação de seus desempenhos mediante ao crescimento amostral.

Teste de Hipótese Nula

A Modelagem de Equações Estruturais é uma ferramenta que permite testar simultaneamente relacionamentos complexos entre múltiplas variáveis. A SEM visa testar se um modelo estudado possui relações apropriadas entre as variáveis; para tanto, compara-se a matriz de covariância do modelo estudado (S) com a matriz de covariância de um modelo hipotetizado ou assintótico ($\Sigma(\theta)$; Peugh & Feldon, 2020). O pressuposto da hipótese nula rege que não há diferença entre essas duas matrizes ($H_0: S - \Sigma(\theta) = 0$). Logo, se $H_0 = 0$, então o valor do teste qui-quadrado do modelo estudado (χ^2_{mod}) será igual ao dos graus de liberdade do modelo examinado ($\chi^2_{mod} = GL_{mod}$), sendo GL igual à quantidade de "informação" dos dados

menos o número de parâmetros do modelo ($GL = p - q$; Schumacker & Lomax, 2016; Tarka, 2018).

Porém, na prática da pesquisa, a matriz de covariância do modelo estudado tende a diferir da matriz de covariância do modelo hipotetizado: $S \neq \Sigma(\theta)$. Dessa maneira, o valor do χ^2 expressa a diferença ou a variabilidade entre as matrizes estudada e hipotetizada. Consequentemente, dado um determinado valor de GL, para que o qui-quadrado do modelo estudado (χ^2_{mod}) apresente um p-valor não significativo, por convenção, o seu valor não deve ultrapassar ao do χ^2 crítico (χ^2_{crit} ; ponto sobre a distribuição populacional que permite rejeitar a hipótese nula) para o nível de erro do Tipo I (alfa, α), determinado como igual a 0,05. Portanto, a H_0 tenderá a ser aceita se: $\chi^2_{\text{mod}} < \chi^2_{\text{crit}}$, indicando que o modelo estudado se aproxima do modelo hipotetizado.

Em contrapartida, na SEM, o valor do qui-quadrado do modelo estudado (χ^2_{mod}) é oriundo de métodos de estimação (e.g., Máxima Verossimilhança – ML) que realizam uma função de minimização das matrizes de covariância. Sendo F a função de minimização dos métodos de estimação e N igual ao número amostral, o qui-quadrado do modelo estudado (χ^2_{mod}) pode ser expresso como: $\chi^2_{\text{mod}} = F(N - 1)$. Por conseguinte, a estimação do χ^2_{mod} é dependente do tamanho amostral (N), penalizando, especialmente, amostras maiores, que, conforme exposto, tendem a um p-valor significativo (MacCallum et al., 1996).

Ademais, de modo a corroborar com a hipótese nula (H_0), entre os vários índices avaliados para a aceitação de rejeição de modelos, destaca-se os usualmente utilizados na SEM: índice raiz do erro quadrático médio de aproximação (RMSEA, *Root Mean Square Error of Approximation*) e índice de ajuste comparativo (CFI, *Comparative Fit Index*). Ambos os índices (RMSEA e CFI) investigam a discrepância entre o modelo estudado e o hipotetizado. O índice RMSEA é estimado com base no número amostral (N), no parâmetro da não centralidade (λ) e nos graus de liberdade (GL). O parâmetro da não centralidade é uma medida de afastamento da hipótese nula, ou seja, o quanto a média da matriz de covariância do modelo estudado se afasta da média da matriz de covariância do modelo hipotetizado (Bhattacharya & Burman, 2016). Assim, a fórmula de RMSEA pode ser descrita como:

$$\text{RMSEA} = \sqrt{(\chi^2 - GL) / \sqrt{GL (N - 1)}}$$

O índice RMSEA estipula pontos de corte para a avaliação dos modelos em determinadas categorias, a saber: "excelente", "bom", "aceitável", "mediocre" e "pobre". Normalmente, valores iguais ou abaixo de 0,05 classificam o modelo avaliado como "excelente" (Browne & Cudeck, 1993; MacCallum et al., 1996).

Já o índice CFI compara o modelo estudado com o modelo independente. Quando Fm representa o valor da função de ajuste da população do modelo estudado e Fb

significa o valor da função de ajuste da população do modelo independente, o índice CFI pode ser descrito como: $\text{CFI} = 1 - Fm/Fb$. Os valores do índice CFI vão de 0 a 1; quanto mais o índice CFI se aproxime de 1, melhor pode ser considerado o ajuste do modelo (Kline, 2015).

Correção do Teste Escalado de Satorra-Bentler

Em amostras que violem os pressupostos da normalidade, o teste escalado de Satorra-Bentler (T_{SB}) efetua correção no qui-quadrado (χ^2), denominado neste estudo como χ^2_{SB} . Quando \bar{U} representa a média da estimação da matriz de covariância assintótica; $\bar{\Gamma}$ exprime a média da estimação da matriz de covariância do modelo estudado; GL retrata os graus de liberdade; e Z retrata a correção em escala da amostra baseada na estrutura do modelo estudado e no nível de não normalidade dos dados, ou seja, o fator de correção efetuado pelo teste T_{SB} , Z pode ser expresso como: $Z = \text{tr}(\bar{U} \bar{\Gamma} / GL)$. Então, o teste T_{SB} realiza a correção no qui-quadrado (χ^2_{SB}) ajustado pela média, estimado pelo teste T_{ML} (função de minimização efetuada pela solução ML), por meio da seguinte função: $T_{\text{SB}}: T_{\text{SB}} = T_{\text{ML}}/Z$. Quando $Z=1$, o valor do χ^2 é igual ao dos graus de liberdade ($\chi^2/GL=1$) e $T_{\text{SB}}=T_{\text{ML}}$. Quando $Z>1$, a correção realizada pelo teste T_{SB} será maior quanto maior for a diferença entre a média da matriz de covariância assintótica e a média do modelo estudado. Portanto, quando aplicada a correção do teste T_{SB} , tal teste gera um determinado p-valor (p_{SB}) relativo ao χ^2_{SB} que suporta a aceitação ou rejeição de um modelo pela H_0 ($H_{0\text{SB}}$; Satorra & Bentler, 2001).

Estratégia de Moshagen e Erdfelder

A estratégia formulada por Moshagen e Erdfelder ($T_{\text{M\&E}}$, 2016) preconiza que, caso a hipótese alternativa (H_1) possa ser rejeitada, então, não há evidências para refutar H_0 . O índice alfa (α) compreende a probabilidade de recusar uma H_0 verdadeira (i.e., erro Tipo I). Normalmente, α tem um valor fixado em 0,05. Já o índice beta (β) representa a probabilidade de aceitar uma H_0 falsa ou erro Tipo II. Por convenção, β é igual a 0,80. Assim, segundo Moshagen e Erdfelder (2016), a diferença nas probabilidades entre α e β pode ocasionar mais erros β devido a sua maior chance de ocorrência em comparação ao erro α .

Com base nessa perspectiva, Moshagen e Erdfelder (2016) sugeriram instituir "um critério para equilibrar a probabilidade de rejeitar erroneamente uma H_0 correta e aceitar erroneamente uma H_0 incorreta" (pp. 3-4). Nesse caso, a proposta sugerida é a de considerar um balanceamento entre os tipos de erros (Tipo I e II), adotando o princípio que $\alpha=\beta$ (erro Tipo I = erro Tipo II). Ao se considerar que $\alpha=\beta$, a H_1 será avaliada em termos do nível da não violação do modelo estudado. Nesse contexto, a H_0 é comparada a H_1 ; sendo H_1 descrita como: $H_1: S - \Sigma(\theta) \neq 0$. Se a H_1 for verdadeira, ou melhor, se existir diferença entre a matriz de covariância do modelo

estudado e hipotetizado, o teste de χ^2 não seguirá mais uma distribuição central, estando em conformidade com o parâmetro da não centralidade (λ).

Sendo assim, na estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016), a primeira etapa consiste em encontrar o valor do parâmetro da não centralidade (λ). Quando F_o representa a discrepância entre S (matriz de covariância do modelo estudado), $\Sigma(\theta)$ (matriz de covariância do modelo hipotetizado) e N o número amostral, então, o parâmetro da não centralidade (λ) é representado por meio da seguinte equação: $\lambda = Fo(N - 1)$.

Na estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016), o índice raiz do erro quadrático médio de aproximação (RMSEA) é um dos sugeridos para encontrar o valor de λ . Quando $S \neq \Sigma(\theta)$, o parâmetro da não centralidade pode ser descrito como: $\lambda = RMSEA^2 GL(N - 1)$. No caso da proposição $T_{M\&E}$, o índice RMSEA é usualmente convençãoado com o valor de 0,05. Ao se encontrar o valor de λ , a próxima etapa visa determinar o valor crítico de λ (χ^2_{crit}) em relação a determinados graus de liberdade por meio do p-valor, igualando alfa e beta ($\alpha = \beta$). Então, na proposta $T_{M\&E}$, o valor do χ^2_{crit} é comparado com o valor do χ^2_{mod} (qui-quadrado do modelo estudado), sendo a possibilidade de preferir a H_1 e, assim, aceitar a H_0 quando: $\chi^2_{crit} > \chi^2_{mod}$; que representa que a diferença entre a matriz de covariância do modelo estudado (S) e a do modelo hipotetizado ($\Sigma(\theta)$) não foi suficiente para ocasionar a violação do modelo estudado.

Teste de Equivalência de Yuan et al.

O teste de equivalência proposto por Yuan et al. (T_Y ; 2016) visa identificar o limite tolerável da má especificação de um modelo denominado de *T-size* (*minimum tolerable size*). Adota a premissa que a hipótese nula seja definida como: $H_{0a}: F_{ML} > F_o$, quando F_{ML} é a função de minimização efetuada pela solução ML e F_o representa “um número positivo pequeno que possa tolerar o tamanho da má especificação” (Yuan et al., 2016, p. 321). Dessa forma, a H_{0a} é rejeitada quando F_o for maior ou igual a F_{ML} . Em termos do parâmetro da não centralidade, a hipótese $H_{0a}: F_{ML} > F_o$ pode ser expressa como: $H_{0a}: \lambda > \lambda_o$. Tendo em vista o parâmetro da não centralidade, o modelo estudado será suportado se o qui-quadrado crítico ($\chi^2_{crit}; \lambda$) for maior que a diferença entre a matriz de covariância do modelo estudado (S) e a matriz de covariância do modelo hipotetizado ou assintótico ($\Sigma(\theta)$); diferença representada como: λ_o .

Então, o teste de equivalência proposto por Yuan et al. (T_Y ; 2016) baseia-se, principalmente, em dois índices: RMSEA e índice de ajuste comparativo (CFI, *Comparative Fit Index*). Assim, o teste de equivalência estima intervalos de confiança (IC=95%) para os índices RMSEA (RMSEA *T-size* [*minimum tolerable size*] – RMSEA_i) e CFI (CFI *T-size* [*minimum tolerable size*] – CFI_i), que demonstram se a violação do modelo estudado possa ser tolerável (*T-size*; Marcoulides & Yuan, 2017; Yuan et al.,

2016). No contexto, o teste T_Y considera intervalo de confiança (IC) a partir do limite superior do índice RMSEA (RMSEA_i), bem como do limite inferior do índice CFI (CFI_i). Com base nos IC de ambos os índices (RMSEA_i e CFI_i), o teste de equivalência estima uma padronização com base nos preceitos de Browne e Cudeck (1993) e MacCallum et al. (1996), avaliando o modelo estudado nas seguintes categorias: “excelente”, “bom”, “aceitável”, “mediocre” e “pobre”; categorias respectivas os pontos de corte do índice CFI (0,99, 0,95, 0,92, 0,90) e RMSEA (0,01, 0,05, 0,08, 0,10; Marcoulides & Yuan, 2017; Yuan et al., 2016).

O Presente Estudo

É clara a importância dos estudos simulados na área da SEM; pois, por meio de simulações, é possível controlar diversas variáveis (e.g., tamanho amostral) para a verificação do desempenho de variáveis específicas. Tem-se, assim, algumas vantagens, como a possibilidade de ampliar o escopo de pesquisas para a testagem de hipóteses e investigações exploratórias (Chen et al., 2008; Nevitt & Hancock, 2000; Xia et al., 2016). Algumas simulações já foram realizadas na SEM em situações de amostras não normais (Chuang et al., 2015; Maydeu-Olivares et al., 2017; Savalei, 2010) e com diversos tamanhos amostrais (Chuang et al., 2015; Xia et al., 2016).

Ademais, pesquisas empíricas foram localizadas na literatura com o uso da estratégia T_Y (Yuan et al., 2016) para a avaliação da viabilidade de modelos (Fu et al., 2017; Rippe & Merkelbach, 2021; Silva et al., 2018b). A fiabilidade da estratégia T_Y também foi avaliada em estudos prévios (Counsell et al., 2020; Finch & French, 2018; Jiang et al., 2017; Peugh & Feldon, 2020). Porém, não foram detectadas investigações com o uso da estratégia de $T_{M\&E}$ (Moshagen & Erdfelder, 2016).

Todavia, investigações que comparassem os testes $T_{M\&E}$ e T_Y , especialmente em amostras que violem os pressupostos da normalidade amostral, são bastante raras. Uma busca em base de periódicos científicos (e.g., PubMed, EBESCO) não localizou estudos que comparassem ambas as abordagens ($T_{M\&E}$ e T_Y) em qualquer contexto.

Não obstante, o uso das propostas $T_{M\&E}$ e T_Y pode auxiliar os pesquisadores a verificar se o limite da violação dos seus modelos pode ser tolerável, podendo ser estratégias valiosas, especialmente na presença de amostras maiores (e.g., >1.000), quando o teste de significância da hipótese nula (H_0) possa tender a ser significativo (>0,05). Dessa forma, optou-se, neste estudo, por comparar a H_{0SB} (χ^2 corrigido pelo teste escalado de Satorra-Bentler – T_{SB}) com a estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016) e a T_Y (Yuan et al., 2016), bem como comparar a proposta $T_{M\&E}$ com a T_Y , pois são propostas encontradas na literatura para a verificar se a violação de um modelo estudado pode ser tolerável. Comumente, na SEM, vários índices (e.g., RMSEA) são utilizados para

corroborar a aceitação da hipótese nula (H_0). Contudo, os pontos de corte desses índices podem ser considerados controversos (Chuang et al., 2015; Xia et al., 2016), instigando a obtenção de mais informações sobre a viabilidade de um modelo estudado.

Assim, com a finalidade de comparar o teste de significância da $H_{0_{SB}}$ (Satorra & Bentler, 2001), $T_{M\&E}$ (Moshagen & Erdfelder, 2016) e T_Y (Yuan et al., 2016) em amostras que violem os pressupostos da normalidade, será avaliada a aceitação ou rejeição de modelos pelas estratégias ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y) mediante o aumento amostral (N), controlando determinados parâmetros (desvio padrão, média, assimetria e curtose univariadas). Nesse contexto, este estudo apresenta a seguinte hipótese: os desempenhos dos testes $H_{0_{SB}}$, $T_{M\&E}$ e T_Y são similares na aceitação e rejeição de modelos mediante ao aumento amostral em amostras que violem os pressupostos da normalidade submetidas à correção do teste escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001).

Método

Delineamento

Trata-se de um estudo simulado. Para efetuar as simulações, obteve-se uma matriz de correlação de amostra normal hipotética com base no estudo de Chuang et al. (2015), manipulada de forma que se aproxime de estudos empíricos (Xia et al., 2016), quando a matriz de covariância do modelo estudado propende a diferir da matriz de covariância do modelo hipotetizado ($S \neq \Sigma(\theta)$; Satorra & Bentler, 2001).

Procedimentos

A partir da matriz de correlação gerada, os dados foram simulados no programa EQS, versão 6.3, com 100 replicações (Bentler, 2006; ver Anexo 1 – Exemplo de síntese utilizada no EQS). A transformação dos dados não normais multivariados baseou-se no método de Fleishman (1978) com extensão de Vale e Maurelli (1983). Os modelos foram estimados com base no teste Escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 2001).

Para a criação de amostras não normais, utilizou-se a transformação por meio da manipulação polinomial da assimetria (S) e curtose (K) (Fleishman, 1978; Vale & Maurelli, 1983). Aliás, estatisticamente, a não normalidade dos dados pode ser mensurada através da assimetria e curtose. Ademais, a existência de assimetria e curtose em conjunto afeta em maior medida os dados do que a presença apenas da assimetria ou curtose. Além disso, assimetria e curtose univariadas maiores que o valor de ± 1 podem ser caracterizadas como excessivas (Wang et al., 1996).

As amostras foram estipuladas em 100, 500, 1.500, 5.000 e 10.000 casos. O tamanho amostral de 100 casos pode ser mais comumente encontrado nas pesquisas

empíricas, enquanto as amostras com 500 casos ou mais podem ser relativamente grandes nas áreas das ciências sociais e humanas (Xia et al., 2016).

Utilizou-se um modelo da análise fatorial confirmatória (AFC) com 16 variáveis observáveis e dois latentes – cada variável latente contou com oito variáveis observáveis (Chuang et al., 2015), alcançando 103 graus de liberdade. O modelo com 103 graus de liberdade constituiu-se enquanto complexo. Na prática da SEM, modelos complexos são mais usualmente estimados (Herzog et al., 2007). Os modelos também contaram com dois fatores latentes por serem utilizados na pesquisa empírica (Damásio et al., 2011; Wagner et al., 2014) e experimental (Hayakawa, 2019; Yang & Liang, 2013). Utilizou-se variáveis observáveis contínuas, mais comuns nas pesquisas em Psicologia (e.g., escala de Likert de autorrelato; Jebb et al., 2021).

Ademais, os modelos foram selecionados de modo a se aproximarem das pesquisas empíricas (Damásio et al., 2011; Wagner et al., 2014); assim, apresentaram uma carga fatorial constricta em cada fator, bem como covariância e variância não constrictas entre os fatores latentes; além de não possuírem cargas cruzadas (Xia et al., 2016). Duas condições amostrais foram avaliadas: 1. DP (desvio padrão) = 2, M (média) = 2, S (assimetria) = 1,5 e K (curtose) = 7; 2. $DP=2$, $M=2$, $S=2$ e $K=21$. A condição 1 pode ser considerada como uma distribuição moderadamente não normal; bem como a condição 2 como uma distribuição extremamente não normal (Xia et al., 2016). O desvio padrão expressa o grau de dispersão de uma amostra. A média é uma medida de tendência central. Uma distribuição normal padrão, que segue o parâmetro da centralidade, apresenta o desvio padrão de 1 e a média igual a 0 (Salkind, 2007). Assim, dez modelos foram criados e codificados, conforme a Tabela 1.

A verificação da aceitação ou rejeição dos modelos foi efetuada com base no p -valor (p_{SB} ; $H_{0_{SB}}$) do teste T_{SB} (Satorra & Bentler, 2001) e por meio dos critérios utilizados pelas estratégias $T_{M\&E}$ e T_Y . Vale destacar que a proposta $T_{M\&E}$ preconiza que, para um modelo seja aceito, o valor do qui-quadrado do modelo estudado deve ser menor que o do qui-quadrado crítico do parâmetro da não centralidade quando $\alpha = \beta$; $\chi^2_{\text{mod}} < \chi^2_{\text{crit}}$; Moshagen & Erdfelder, 2016). Dessa maneira, para possibilitar a avaliação da proposta $T_{M\&E}$, o valor do qui-quadrado estimado pelo teste T_{SB} (χ^2_{SB}) também foi apresentado neste estudo. Os cálculos referentes à proposta de Moshagen e Erdfelder (2016) podem ser obtidos por meio do programa SEMPower, de distribuição livre (<http://psycho3.uni-mannheim.de/semppower>), na parte "compromise".

Com a finalidade de avaliar a estratégia T_Y (Marcoulides & Yuan, 2017; Yuan et al., 2016), foram auferidos os valores de $RMSEA_t$ e CFI_t , com seus intervalos de confiança, classificando os modelos com

ajuste “excelente/bom”, conforme os pontos de corte sugeridos por Yuan et al. (2016). De forma prática, para se obter o valor de $RMSEA_t$ e CFI_t , e seus intervalos de confiança, primeiramente, efetua-se a estimação do

modelo em um programa da SEM (e.g., EQS, Lisrel, Mplus, R). Posteriormente, sintaxes específicas são utilizadas e processadas no programa R (cf. Marcoulides & Yuan, 2017).

Tabela 1
Codificação dos Modelos

Condições Amostrais	Sigla	Número de casos	Desvio padrão	Média	Assimetria	Curtose
1	M*100	100	2	2	1,5	7
	M*500	500	2	2	1,5	7
	M*1500	500	2	2	1,5	7
	M*5000	5.000	2	2	1,5	7
	M*10000	10.000	2	2	1,5	7
2	MO_100	100	2	2	2	21
	MO_500	500	2	2	2	21
	MO_1500	500	2	2	2	21
	MO_5000	5.000	2	2	2	21
	MO_10000	10.000	2	2	2	21

Os valores do χ^2_{crit} foram auferidos com o programa SEMPower (Moshagen & Erdfelder, 2016). Já os valores do índice $RMSEA_t$ e CFI_t , bem como dos seus respectivos intervalos de confiança, foram obtidos por meio do programa R (Marcoulides & Yuan, 2017; Yuan et al., 2016). Portanto, as estratégias analisadas ($H0_{SB}$, $T_{M\&E}$ e T_Y) foram comparadas em termos da possibilidade de aceitação ou rejeição dos modelos com base nos critérios utilizados por essas estratégias, considerando o crescimento amostral.

Resultados

Este estudo comparou a hipótese nula ($H0_{SB}$), quando o qui-quadrado (χ^2) foi corrigido pelo teste Escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 2001); a estratégia de Moshagen e Erdfelder ($T_{M\&E}$; 2016); e o teste de equivalência de Yuan et al. (T_Y ; Yuan et al., 2016). Observou-se que os modelos com 100 casos (M*100 e MO_100) puderam ser suportados pela hipótese nula ($H0_{SB}$, $p_{SB} \geq 0,05$; Kline, 2015) em amostras que violaram os pressupostos da normalidade amostral ($DP=2$, $M=2$, $S=1,5$, $K=7$; $DP=2$, $M=2$, $S=2$, $K=21$). Além disso, a $H0_{SB}$ foi rejeitada ($p_{SB} < 0,05$; χ^2 do modelo estudado maior que o $\chi^2_{critico}$; $\chi^2_{mod} < \chi^2_{crit}$) nos modelos com 500, 1.500, 5.000 e 10.000 mil casos (M*500, M*1500, M*5000, MO_500, MO_1500, MO_5000 e MO_10000; Tabela 2).

Já a estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016) rejeitou os modelos com 100 casos (M*100 e MO_100); tendo em vista que, nessas amostras, o valor do qui-quadrado crítico do parâmetro da centralidade no caso de alfa é igual a beta ($\alpha = \beta$) ($\chi^2_{\lambda_{crit}}$) foi menor que o valor do qui-quadrado (χ^2_{SB}) corrigido pelo teste de Santorra-Bentler

(T_{SB}). Assim, com base na estratégia $T_{M\&E}$, a hipótese nula ($H0$) foi rejeitada e a hipótese alternativa ($H1$) foi aceita nas amostras com 100 casos. Os outros modelos com 500, 1.500, 5.000 e 10.000 mil casos (M*500, M*1500, M*5000, M*10000, MO_500, MO_1500, MO_5000 e MO_10000) foram aceitos pelo teste $T_{M\&E}$ ($\chi^2_{\lambda_{crit}} < \chi^2_{mod}$), sendo que não houve evidências para a rejeição da $H1$.

Em relação ao teste T_Y (Yuan et al., 2016), considerando os modelos classificados como “excelente/bom” pelos índices $RMSEA_t$, todos os modelos puderam ser considerados como aceitáveis (M*100, M*500, M*1500, M*5000, MO_100, MO_500, MO_1500, MO_5000 e MO_10000). Por sua vez, os modelos com 100 casos (M*100 e MO_100) foram avaliados como “pobre” pelo índice CFI_t , já que os valores encontrados (M*100 = 0,78 e MO_100=0,73) foram menores que 0,88 (modelo “excelente/bom”), bem como, os modelos M*5000, M*10000, MO_5000 e MO_10000 foram classificados como “excelente/bom” pelo índice CFI_t .

Dessa forma, os testes p -valor (p_{SB} ; $H0_{SB}$), $T_{M\&E}$ e T_Y revelaram comportamentos diferenciados na aceitação e rejeição de modelos que violaram os pressupostos da normalidade amostral mediante ao seu crescimento. Logo, a hipótese do presente estudo foi rejeitada.

Porém, observou-se que o teste p -valor (p_{SB} ; $H0_{SB}$), $T_{M\&E}$ e T_Y revelaram-se sensíveis ao tamanho amostral. Por exemplo, o modelo M*100 e MO_100 foram aceitos pelo teste p_{SB} (0,20 e 0,14, respectivamente), mas rejeitados pelos testes $T_{M\&E}$ ($\chi^2_{\lambda_{crit}} < \chi^2_{SB}$) e T_Y ($CFI_t > 0,88$, modelo “pobre”). Além disso, os modelos acima de 500 casos foram rejeitados pelo teste p_{SB} ($> 0,05$; $\chi^2_{mod} > \chi^2_{crit}$), mas suportados pelas estratégias $T_{M\&E}$ e T_Y .

Tabela 2
Modelos Testados e Respectivos Valores de χ^2_{SB} , p_{SB} , Testes $T_{M\&E}$ e T_Y

Cond. Amost.	Modelo	χ^2_{SB}	p_{SB}	$T_{M\&E}$		T_Y		
				$\chi^2_{\lambda_{crit}}$	RAMSEA _t	Classificação Excel./Bom	CFI _t	Classificação Excel./Bom
1	M*100	114.818	0,20	114.209	0,06	≤0,08	0,78	≥0,88
	M*500	131.890	0,03	154.617	0,03	≤0,06	0,94	≥ 0,93
	M*1500	154.549	<0,001	239.598	0,02	≤0,06	0,97	≥ 0,94
	M*5000	184506	<0,001	499.166	0,02	≤0,05	0,99	≥ 0,95
	M*10000	337.497	<0,001	845.865	0,02	≤0,05	0,99	≥ 0,95
2	MO_100	118.490	0,14	114.209	0,07	≤0,08	0,73	≥0,88
	MO_500	147.482	<0,001	154.617	0,04	≤0,06	0,93	≥ 0,93
	MO_1500	159.744	<0,001	239.598	0,02	≤0,06	0,97	≥ 0,94
	MO_5000	174.627	<0,001	499.166	0,01	≤0,05	0,99	≥ 0,95
	MO_10000	318.981	<0,001	845.865	0,02	≤0,05	0,99	≥ 0,95

Notas. Os valores negritados no teste de hipótese são os referentes a aceitação do modelo avaliado ($p_{SB} \leq 0,05$); os valores negritados relativos à estratégia $T_{M\&E}$ são os respectivos a aceitação do modelo avaliado ($\chi^2_{\lambda_{crit}} > \chi^2_{SB}$); os valores negritados na estratégia T_Y demonstram quando o modelo foi classificado como “excelente/bom”. Excel., excelente; p_{SB} : p -valor do teste escalado de Satorra-Bentler; RMSEA_t: raiz do erro quadrático médio de aproximação T-size (*minimum tolerable size*); CFI_t: índice de ajuste comparativo T-size (*minimum tolerable size*); $T_{M\&E}$: teste proposto por Moshagen e Erdfelder (2016); T_Y : teste de equivalência proposto por Yuan et al. (2016); χ^2_{crit} : qui-quadrado crítico do parâmetro da não centralidade; χ^2_{SB} : qui-quadrado do teste Escalado de Satorra-Bentler.

Discussão

Este estudo simulado e exploratório comparou a aceitação e a rejeição de modelos pelo teste de hipótese nula ($H_{0_{SB}}$; p -valor do teste de Satorra-Bentler [p_{SB}]), a estratégia de Moshagen e Erdfelder ($T_{M\&E}$; Moshagen & Erdfelder, 2016) e o teste de equivalência de Yuan et al. (T_Y ; Yuan et al., 2016) em amostras que violaram os pressupostos da normalidade, considerando o aumento amostral. Foram avaliadas amostras de 100, 500, 1.500, 5.000 e 10.000 casos, com base em duas condições amostrais, que se diferenciaram pela assimetria ($S=1,5$ ou 2) e curtose ($K=7$ ou 21). As amostras foram corrigidas pelo teste escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001). Vale ressaltar que, embora a assimetria possua impactos nos valores do qui-quadrado corrigido pelo teste de Satorra-Bentler (χ^2_{SB}) em decorrência da estimação pela Máxima Verossimilhança (ML), o teste T_{SB} corrige especialmente a curtose (Xia et al., 2016; Yuan et al., 2016).

No contexto avaliado, a hipótese do presente estudo foi rejeitada, considerando que os testes $H_{0_{SB}}$ (p_{SB} ; Satorra & Bentler, 1994; 2001), $T_{M\&E}$ (Moshagen & Erdfelder, 2016) e T_Y (Yuan et al., 2016) revelaram comportamentos distintos. Destaca-se que os testes analisados ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y) utilizam estratégias diferenciadas para a aceitação ou rejeição dos modelos estudados.

O teste de hipótese nula (H_0) sugere que um modelo possa ser aceito quando o erro Tipo I (probabilidade de recusar uma H_0 verdadeira, α) não ultrapasse o valor de confiança de 95%; ou seja, um p -valor menor ou igual a 0,05 (Perezgonzalez, 2015). Para tanto, o teste

convencional da hipótese nula (H_0) compara a matriz de covariância do modelo estudado (S) com a matriz de covariância de um modelo hipotetizado ou assintótico ($\Sigma(\theta)$). Assim, o pressuposto da hipótese nula (H_0) preconiza que a diferença entre essas matrizes deve ser igual a zero ($H_0: S - \Sigma(\theta) = 0$; Tarka, 2018; Kline, 2015; Schumacker & Lomax, 2016).

Neste estudo, o p -valor (p_{SB}) foi respectivo ao qui-quadrado (χ^2_{SB}) corrigido pelo teste escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001). No caso, o teste T_{SB} corrige as amostras não normais estimadas pela solução Máxima Verossimilhança (ML) de modo que se aproximem de uma amostra normal. Assim, quanto maior for a diferença entre a média da matriz de covariância assintótica e a média do modelo estudado, maior será a correção do teste T_{SB} .

Porém, o p -valor (H_0) é sensível ao tamanho amostral (Kline, 2015). Neste estudo, a hipótese nula ($H_{0_{SB}}$; p_{SB}), considerando um qui-quadrado corrigido pelo teste de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001), foi corroborada unicamente nos modelos com 100 casos (M*100 e MO_100; $p_{SB} \leq 0,05$). Os modelos iguais ou acima de 500 casos foram rejeitados ($p_{SB} > 0,05$).

Devido a controvérsias em relação ao teste de hipótese nula (Perezgonzalez, 2015), Moshagen e Erdfelder (2016) e Yuan et al. (2016) propuseram estratégias complementares à H_0 , tendo em vista a possibilidade de rejeitar amostras maiores. A variabilidade do p -valor em função do aumento amostral foi apoiada por outros estudos (Brosseau-Liard, 2013; Chen et al., 2008; Kline, 2015; Nevitt & Hancock, 2000; Xia et al., 2016), o que

parece corroborar com a necessidade de testes complementares para a aceitação da H_0 (Chen et al., 2008; Kline, 2015; Yuan et al., 2016).

A estratégia de Moshagen e Erdfelder ($T_{M\&E}$; 2016) considera que existe uma disparidade entre a aceitação de modelos com base na hipótese nula (H_0) de 0,05 (erro Tipo I, α) e a rejeição da hipótese alternativa (H_1) de 0,80 (erro Tipo II, β); ou seja, existe uma chance de aceitar um modelo de 5% e de rejeitá-lo em 80%. Os autores, assim, propuseram que o erro Tipo I e II sejam iguais ($\alpha=\beta$); sugerindo que, se não existem evidências para rejeitar a H_1 , então, a H_0 pode ser suportada. Tendo em vista que o qui-quadrado crítico (χ^2_{crit}) representa um ponto sobre a distribuição populacional que permite rejeitar a hipótese nula (H_0), para que um modelo seja apoiado, a estratégia estipula que o qui-quadrado do modelo estudado (χ^2_{mod}) seja menor que o qui-quadrado crítico do parâmetro da não centralidade ($\chi^2_{\lambda,crit}$), quando alfa é igual a beta ($\alpha=\beta$).

No presente estudo, a estratégia de Moshagen e Erdfelder ($T_{M\&E}$; 2016) rejeitou os modelos com 100 casos (M^*100 e MO_{100}). Contudo, os autores (Moshagen & Erdfelder, 2016) sugeriram que, em amostras menores, possa ocorrer um maior risco de erros, recomendando que a estratégia ($T_{M\&E}$) seja aplicada a amostras maiores, reduzindo a possibilidade dos erros Tipo I e II.

Por sua vez, o teste de equivalência de Yuan et al. (T_Y ; Yuan et al., 2016) considerou os modelos com 100 casos (M^*100 e MO_{100}) como “pobre” em termos do índice CFI_t (índice de ajuste comparativo *T-size* [*minimum tolerable size*]); embora o índice $RMSEA_t$ (raiz do erro quadrático médio de aproximação *T-size* [*minimum tolerable size*]) avaliou todos os modelos como “excelente/bom”. Portanto, os modelos com 100 casos não foram suportados, considerando o índice CFI_t . Porém, Yuan et al. (2016) ressaltaram que a fiabilidade do teste de equivalência melhora em decorrência do aumento amostral.

Destaca-se, então, que os índices CFI_t e $RMSEA_t$, utilizados no teste de equivalência Yuan et al. (T_Y ; Yuan et al., 2016), são baseados nos índices CFI e $RMSEA$ convencionais, comumente utilizados nos Modelos de Equações Estruturais (SEM). Dessa forma, o teste T_Y estipula um IC de 95% para os índices $RMSEA$ ($RMSEA_t$) e CFI (CFI_t), que demonstra se a violação do modelo estudado possa ser suportada, condição denominada de *T-size* (*minimum tolerable size*; Marcoulides & Yuan, 2017; Yuan et al., 2016).

Logo, o fato do índice $RMSEA_t$ ter apontado para a aceitação do modelo com 100 casos (M^*100 e MO_{100}) e para rejeição pelo índice CFI_t pode ser explicado visto que os índices $RMSEA$ e CFI avaliam os modelos com premissas diferentes (Lai & Green, 2016). Por conseguinte, o índice $RMSEA_t$ e CFI_t baseiam-se no $RMSEA$ e CFI convencionais (Yuan et al., 2016). O

índice $RMSEA$ avalia o modelo estudado por meio da estimação do número amostral (N), do parâmetro da não centralidade (λ) e dos GL. O índice CFI compara o modelo estudado com o independente (Kline, 2015).

Acrescenta-se que estudos revelaram que o índice $RMSEA$ corrigido pelo teste T_{SB} pode apresentar bom desempenho mediante ao crescimento amostral (Brosseau-Liard, 2013; Brosseau-Liard et al., 2012; Xia et al., 2016). Porém, em estudo simulado, Brosseau-Liard et al. (2012) encontraram que a não normalidade amostral diminuiu o valor médio dos índices de ajuste a partir da correção do teste T_{SB} .

Aliás, Ainur et al. (2017) verificaram que índice $RMSEA$ foi menos afetado pela não normalidade dos dados que o índice CFI . Além disso, Sideridis et al. (2014) sugeriram que $RMSEA$ possa ser apropriado para amostras menores.

Ademais, Lai e Green (2016) sugeriram que, em amostras que violem os pressupostos da normalidade, não existe consenso nas definições dos pontos de cortes nos índices $RMSEA$ e CFI , já que as propriedades estatísticas dos estimadores para amostras não normais ainda não são inteiramente compreendidas. Desse modo, os desacordos entre os índices $RMSEA$ e CFI não podem ser considerados como desajustes no modelo avaliado. Contudo, como o índice CFI pode ser considerado como parcimonioso, costuma ser utilizado em adição ao $RMSEA$ para suportar análises de modelos em SEM (Kline, 2015).

Conforme verificado neste estudo, a hipótese nula (H_{0SB} ; p_{SB}), tendo em vista um qui-quadrado corrigido pelo teste de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001), apresentou uma tendência de rejeição de amostras maiores (≥ 500 casos; Kline, 2015). Ademais, ressalta-se que tanto o teste $T_{M\&E}$ (Moshagen & Erdfelder, 2016) e T_Y (Yuan et al., 2016) possuem a mesma premissa, avaliar se a violação do modelo estudado possa ser suportada quando a hipótese nula (H_0) possa ser rejeitada ($>0,05$) em decorrência do aumento amostral, embora ambas as estratégias ($T_{M\&E}$ e T_Y) utilizem métodos diferenciados para a aceitação ou rejeição de modelos. Aliás, tanto o $T_{M\&E}$ (Moshagen & Erdfelder, 2016) e T_Y (Yuan et al., 2016) podem ser mais adequados para avaliarem um número amostral maior (≥ 500 casos).

Todavia, enfatiza-se que, conforme o teste T_Y baseia-se em dois índices ($RMSEA_t$ e CFI_t) para a aceitação ou rejeição de modelos, esse teste pode ser preferível ao teste $T_{M\&E}$ para apoiar um modelo, pois pode apresentar mais fiabilidade para avaliação de modelos na SEM. Além disso, estudos prévios (Counsell et al., 2020; Finch & French, 2018; Fu et al., 2017; Jiang et al., 2017; Peugh & Feldon, 2020; Rippe & Merkelbach, 2021; Silva et al., 2018b) foram localizados na literatura utilizando o teste T_Y . Contudo, a partir de ampla pesquisa em bancos de dados científicos, nenhum estudo

empírico ou simulado foi localizado com a técnica $T_{M\&E}$. Assim, compreende-se que o teste T_Y está paulatinamente adquirindo escopo na literatura da SEM.

Contudo, Peugh e Feldon (2020) recomendaram cautela no uso do teste T_Y ; tendo em vista que o índice CFI depende da mensuração do modelo independente, os resultados do teste T_Y podem variar, já que os *softwares* de análise estatística na SEM podem especificar o modelo independente de maneiras diferenciadas. Também Montoya e Edwards (2021) sugeriram que o teste T_Y possa não ser apropriado para estudos exploratórios (e.g., Análise Fatorial Exploratória – AFE). Porém, o teste T_Y ; representa um avanço para avaliar os ajustes de modelos na SEM (Peugh & Feldon, 2020).

É pertinente frisar que o presente estudo possui limitações. Assim, ressalta-se que apenas duas técnicas ($T_{M\&E}$ e T_Y) foram comparadas entre si e com o teste de hipótese ($H_{0_{SB}}$), com correção do χ^2 pelo teste escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001). Contudo, outras propostas para a verificação da má especificação de modelos em SEM podem ser encontradas na literatura (e.g., aplicações Bayesianas; Dienes, 2014). Ademais, neste estudo, não foram verificados outros modelos (e.g., multinível) além da análise fatorial confirmatória (AFC) clássica com um nível de fator latente e 16 variáveis; porém, os modelos da AFC clássica são os mais utilizados na literatura (Damásio et al., 2011; Hayakawa, 2019; Wagner et al., 2014; Yang & Liang, 2013). Igualmente, apenas o teste robusto escalado de Satorra-Bentler (Satorra & Bentler, 2001) foi utilizado para as análises por ser um dos mais encontrados na literatura para a estimação de amostras que violem a normalidade amostral (Moshagen & Erdfelder, 2016). Além disso, utilizou-se apenas uma modalidade de matriz de correlação para as simulações; embora a matriz de correlação utilizada tenha sido manipulada para se aproximar de estudos empíricos ($S \neq \sum(\theta)$; Yuan et al., 2016), tal matriz de correlação pode não retratar a realidade da pesquisa aplicada. Além disso, apenas dois níveis de assimetria (S ; 1,5 e 2) e de curtose (K ; 7 e 21) foram manipuladas, contudo, buscou-se uma distribuição moderadamente não normal ($S=1,5$ e $K=7$) e outra extremamente não normal ($S=2$, $K=21$; Xia et al., 2016). Não obstante, este estudo tenha buscado se aproximar da pesquisa aplicada na seleção dos aspectos metodológicos, vale lembrar que estudos simulados apenas se aproximam das pesquisas empíricas, embora, possam contribuir com a geração de conhecimentos (Chen et al., 2008; Nevitt & Hancock, 2000; Xia et al., 2016).

Considerações Finais

Este estudo visou comparar as seguintes estratégias: o teste de hipótese nula ($H_{0_{SB}}$), com um qui-quadrado corrigido pelo teste de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001); a proposta de Moshagen

e Erdfelder ($T_{M\&E}$; 2016) e o teste de equivalência sugerido por Yuan et al. (T_Y ; 2016). Por meio de uma investigação simulada e exploratória, em amostras que violaram os pressupostos da normalidade, as estratégias foram avaliadas quanto a aceitação ou rejeição de modelos com base nas premissas das respectivas propostas ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y).

O estudo permite concluir que os testes ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y) revelam desempenhos diferenciados. Embora tanto o teste $H_{0_{SB}}$ como as estratégias $T_{M\&E}$ e T_Y tenham o objetivo de avaliar se a violação do modelo possa ser aceita, esses testes ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y) recorrem a métodos distintos, incorrendo em resultados dessemelhantes. Os resultados das simulações sugeriram que a $H_{0_{SB}}$ possa suportar ($p_{SB} \leq 0,05$) amostras menores (100 casos) e rejeitar ($p_{SB} > 0,05$) amostras maiores (≥ 500 casos). Por sua vez, os resultados também indicaram que os testes $T_{M\&E}$ e T_Y são mais apropriados para a avaliação de amostras maiores (≥ 500 casos).

Como vantagem, os testes ($H_{0_{SB}}$, $T_{M\&E}$ e T_Y) são de fácil execução, favorecendo o uso por parte dos pesquisadores. A H_0 ($H_{0_{SB}}$) corrigida pelo teste escalado de Satorra-Bentler (T_{SB} ; Satorra & Bentler, 1994; 2001) está disponível em *softwares* da SEM (EQS, Mplus, R) com a saída de dados por meio do p-valor (p_{SB}). A estratégia $T_{M\&E}$ (Moshagen & Erdfelder, 2016) conta com software específico (SemPower) e o teste T_Y (Yuan et al., 2016) disponibiliza sintaxes a serem processadas no programa R.

Todavia, como recomendação, sugere-se a utilização do teste T_Y em detrimento do teste $T_{M\&E}$ quando a H_0 ($H_{0_{SB}}$) não possa ser aceita em decorrência do aumento amostral. Considerou-se o fato de que o teste T_Y valer-se de mais recursos que a estratégia $T_{M\&E}$, que se baseia em dois índices (RMSEA_t e CFI_t) para a avaliação dos modelos. A estratégia $T_{M\&E}$ merece mais estudos a respeito de seu desempenho, pois pode ser valiosa para a avaliação de modelos. Possivelmente, a ausência de estudos com essa estratégia ($T_{M\&E}$) possa ser uma consequência da técnica se fundamentar no qui-quadrado crítico do parâmetro da centralidade ($\chi^2_{\text{crít}}$) a partir de igualar o erro Tipo I e II ($\alpha = \beta$; Moshagen & Erdfelder, 2016).

Quanto à possibilidade de replicabilidade deste estudo, destaca-se que se buscou condições amostrais que se aproximassem de amostras utilizadas nas pesquisas psicossociais, tais como a não normalidade dos dados, modelos comuns na SEM (e.g., CFA) e tamanhos amostrais (e.g., 100 casos). Contudo, como limitação à replicabilidade deste estudo, ressalta-se que, por se tratar de um estudo simulado, utilizou-se dados fixos, que compreendem a assimetria e curtose pré-determinadas.

Assim, com a crescente utilização das técnicas da SEM nas ciências aplicadas, recomenda-se a replicação deste estudo com diferentes números amostrais, graus de liberdade, níveis de má especificação dos modelos e

outros testes robustos. Também, investigações devem ser efetuadas comparando as estratégias $T_{M\&E}$ e T_Y em amostras em consonância com os pressupostos da normalidade ou não normalidade amostral, tendo em vista a ausência de investigações a esse respeito. Além disso, o teste T_Y deve ser avaliado com mais critério em estudos simulados, verificando o seu desempenho em situações em que os índices $RMSEA_t$ e CFI_t possam estar em desacordo.

Agradecimentos

Não há menções.

Financiamento

A presente pesquisa não recebeu nenhuma fonte de financiamento sendo custeada com recursos dos próprios autores.

Contribuições dos autores

Declaramos que todos os autores participaram da

elaboração do manuscrito. Especificamente, os autores Marli Appel da Silva e Guilherme Welter Wendt participaram da redação inicial do estudo (conceitualização, investigação e investigação), bem como da análise dos dados, e a autora Irani Iracema de Lima Argimon participou da redação final do trabalho – revisão e edição.

Disponibilidade de dados e materiais

Todos os dados e sintaxes gerados e analisados durante esta pesquisa serão tratados com total sigilo devido às exigências do Comitê de Ética em Pesquisa com Seres Humanos. Porém, o conjunto de dados e sintaxes que apoiam as conclusões deste artigo estão disponíveis mediante razoável solicitação ao autor principal do estudo.

Conflitos de interesses

Os autores declaram que não há conflitos de interesses.

Referências

- Ainur, A. K., Sayang, M. D., Jannoo, Z., & Yap, B. W. (2017). Sample size and non-normality effects on goodness of fit measures in Structural Equation Models. *Pertanika Journal of Science & Technology*, 25(2), 575-586.
- Bhattacharya, P., & Burman, P. (2016). *Theory and methods of statistics*. London, United Kingdom: Academic Press. doi: 10.1016/c2014-0-02379-9
- Brosseau-Liard, P. E. (2013). Abstract: RMSEA with nonnormal data: A comparison of two robust corrections. *Multivariate Behavioral Research*, 48(1), 150-151. <https://doi.org/10.1080/00273171.2013.751295>
- Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, 47(6), 904-930. <https://doi.org/10.1080/00273171.2012.715252>
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. Em K. A. Bollen & J. S. Long (Orgs.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in Structural Equation Models. *Sociological Methods & Research*, 36(4), 462-494. <https://doi.org/10.1177/0049124108314720>
- Chuang, J., Savalei, V., & Falk, C. F. (2015). Investigation of type I error rates of three versions of robust chi-square difference tests. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(4), 517-530. <https://doi.org/10.1080/10705511.2014.938713>
- Counsell, A., Cribbie, R., & Flora, D. (2020). Evaluating equivalence testing methods for measurement invariance. *Multivariate Behavioral Research*, 55(2), 312-328. <https://doi.org/10.1080/00273171.2019.1633617>
- Damásio, B. F., Machado, W. L., & Silva, J. P. (2011). Estrutura fatorial do Questionário de Saúde Geral (QSG-12) em uma amostra de professores escolares. *Avaliação Psicológica*, 10(1), 99-105. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712011000100011&lng=pt&tlng=pt
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers In Psychology*, 5. Recuperado de <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781/full>. <https://doi.org/10.3389/fpsyg.2014.00781>
- Finch, W., & French, B. (2018). A simulation investigation of the performance of invariance assessment using Equivalence Testing Procedures. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(5), 673-686. <https://doi.org/10.1080/10705511.2018.1431781>
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532. <https://doi.org/10.1007/BF02293811>
- Fu, Y., Wen, Z., & Wang, Y. (2017). The total score with maximal reliability and maximal criterion validity: An illustration using a Career Satisfaction Measure. *Educational and Psychological Measurement*, 78(6), 1108-1122. <https://doi.org/10.1177/0013164417738564>
- Gao, C., Shi, D., & Maydeu-Olivares, A. (2019). Estimating the Maximum Likelihood Root Mean Square Error of Approximation (RMSEA) with non-normal data: A Monte-Carlo Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(2), 192-201. <https://doi.org/10.1080/10705511.2019.1637741>
- Hayakawa, K. (2019). Corrected goodness-of-fit test in covariance structure analysis. *Psychological Methods*, 24(3), 371-389. <https://doi.org/10.1037/met0000180>
- Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 361-390. <https://doi.org/10.1080/10705510701301602>

- Jebb, A., Ng, V., & Tay, L. (2021). A review of key Likert Scale Development Advances: 1995-2019. *Frontiers In Psychology*, 12. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8129175/>. <https://doi.org/10.3389/fpsyg.2021.637547>
- Jiang, G., Mai, Y., & Yuan, K. (2017). Advances in measurement invariance and mean comparison of latent variables: Equivalence Testing and a projection-based approach. *Frontiers In Psychology*, 8. Recuperado de <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01823/full>. <https://doi.org/10.3389/fpsyg.2017.01823>
- Kline, R. (2015). *Principles and practice of Structural Equation Modeling*. (4th ed.). New York: Guilford Press.
- Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, 51(2-3), 220-239. <https://doi.org/10.1080/00273171.2015.1134306>
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology*, 51(1), 201-226. <https://doi.org/10.1146/annurev.psych.51.1.201>
- Maccallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149. <https://doi.org/10.1037//1082-989x.1.2.130>
- Marcoulides, K. M., & Yuan, K. (2017). New ways to evaluate goodness of fit: A note on using Equivalence Testing to assess Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 148-153. <https://doi.org/10.1080/10705511.2016.1225260>
- Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2017). Assessing fit in Structural Equation Models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 389-402. <https://doi.org/10.1080/10705511.2017.1389611>
- Montoya, A. K., & Edwards, M. C. (2021). The poor fit of model fit for selecting number of factors in Exploratory Factor Analysis for Scale Evaluation. *Educational and Psychological Measurement*, 81(3), 413-440. <https://doi.org/10.1177/0013164420942899>
- Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 54-60. <https://doi.org/10.1080/10705511.2014.950896> <https://doi.org/10.1080/10705511.2017.1389611>
- Nevitt, J., & Hancock, G. R. (2000). Improving the Root Mean Square Error of Approximation for nonnormal conditions in Structural Equation Modeling. *The Journal of Experimental Education*, 68(3), 251-268. <https://doi.org/10.1080/00220970009600095>
- Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology*, 6. Recuperado de <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00223/full>. <https://doi.org/10.3389/fpsyg.2015.00223>
- Peugh, J., & Feldon, D. (2020). "How well does your Structural Equation Model fit your data?": Is Marcoulides and Yuan's Equivalence Test the answer? *CBE – Life Sciences Education*, 19(3), es5. <https://doi.org/10.1187/cbe.20-01-0016>
- Rippe, R., & Merckelbach, I. (2021). Planned missing data in early literacy interventions: A replication study with an additional gold standard. *PLOS ONE*, 16(3). Recuperado de <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0249175>. <https://doi.org/10.1371/journal.pone.0249175>
- Salkind, N. J. (2007). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage Publications, Inc. <https://doi.org/10.4135/9781412952644>
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. Em A. Von Eye, & C. Clogg (Eds.), *Latent variables analysis: applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507-514. doi: <https://doi.org/10.1007/BF02296192>
- Savalei, V. (2010). Small sample statistics for incomplete nonnormal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2), 241-264. <https://doi.org/10.1080/10705511003659375>
- Schumacker, R. E., & Lomax, R. G. (2016). *A beginner's guide to Structural Equation Modeling* (4th ed.). London, UK: Routledge.
- Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using Structural Equation Modeling to Assess Functional Connectivity in the Brain. *Educational And Psychological Measurement*, 74(5), 733-758. <https://doi.org/10.1177/0013164414525397>
- Silva, M. A., Wendt, G. W., & Argimon, I. L. (2018a). Inventário de Depressão de Beck II: Análises pela Teoria do Traço Latente. *Avaliação Psicológica*, 17(3), 339-350. <https://doi.org/10.15689/ap.2018.1703.14651.07>
- Silva, M. A., Wendt, G. W., Argimon, I. L., & Fernandes-Lopes, R. (2018b). Técnicas de correção do teste qui-quadrado para amostras não normais. *Avaliação Psicológica*, 17(4), 407-416. <https://dx.doi.org/10.15689/ap.2018.1704.13238.01>
- Tarka, P. (2018). An overview of Structural Equation Modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity*, 52(1), 313-354. Recuperado de <https://link.springer.com/article/10.1007/s11135-017-0469-8>
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465-471. <https://doi.org/10.1007/BF02293687>
- Wagner, F., Camey, S. A., & Trentini, C. M. (2014). Análise fatorial confirmatória da escala de inteligência Wechsler abreviada: Versão português brasileiro. *Avaliação Psicológica*, 13(3), 383-389. Recuperado de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712014000300010&lng=pt&tlng=pt
- Wang, L., Fan, X., & Willson, V. L. (1996). Effects of nonnormal data on parameter estimates and fit indices for a model with latent and manifest variables: An empirical study. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(3), 228-247. <https://doi.org/10.1080/10705519609540042>
- Xia, Y., Yung, Y., & Zhang, W. (2016). Evaluating the selection of Normal-Theory weight matrices in the Satorra-Bentler Correction of chi-square and standard errors. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4), 585-584. <https://doi.org/10.1080/10705511.2016.1141354>
- Yang, Y., & Liang, X. (2013). Confirmatory Factor Analysis under violations of distributional and structural assumptions. *International Journal of Quantitative Research in Education*, 1(1), 61-84. <https://doi.org/10.1504/ijqre.2013.055642>

Yuan, K. H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing Structural Equation Models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(3), 319-330. <https://doi.org/10.1080/10705511.2015.1065414>

recebido em abril de 2019
aprovado em março de 2024

Sobre os autores

Marli Appel da Silva é Psicóloga, Doutora em Psicologia pela Pontifícia Universidade Católica do Rio Grande do Sul (PUC/RS).

Irani Iracema de Lima Argimon é Coordenadora do Grupo de Pesquisa Avaliação e Intervenção no Ciclo Vital do Programa de Pós-Graduação em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sul (PUC/RS).

Guilherme Welter Wendt é Coordenador do Laboratório de Personalidade e Diferenças Individuais (LAPEDI) da Universidade Estadual do Oeste do Paraná. Docente permanente do Programa de Pós-Graduação em Ciências Aplicadas à Saúde.

Como citar este artigo

da Silva, M. A., Argimon, I. I. L., & Wendt, G. W. (2024). Métodos de Correção de Testes Estatísticos em Modelagem de Equações Estruturais. *Avaliação Psicológica*, 23(1), 109-120. <http://dx.doi.org/10.15689/ap.2024.2301.18031.11>