
Editorial

DOI: 10.15689/ap.2016.1501.ed

Descobrir a conexão causal entre fatores latentes hipotéticos e um conjunto de itens ou indicadores é um dos propósitos dos modelos de análise fatorial (Markus & Borsboom, 2013). Nesses modelos, alguns dos parâmetros de interesse são as cargas fatoriais e as correlações entre os fatores, cujos valores são descobertos por procedimentos conhecidos como métodos de estimação ou simplesmente “estimadores”. O presente editorial ilustra o como a escolha fundamentada de um estimador pode fazer diferença em situações práticas de análise fatorial exploratória ou confirmatória.

Um pré-requisito de uma análise fatorial é a existência de correlações (comunalidade) entre as variáveis que são especificadas como indicadores no modelo (Edwards & Bagozzi, 2000). Não obstante, quando as variáveis da análise são assimétricas, as correlações empíricas entre elas podem resultar atenuadas se for empregado o tradicional coeficiente de Pearson, isto é, o valor produzido pelo programa estatístico pode acabar sendo menor do que o valor verdadeiro (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). Se isso ocorre durante uma análise fatorial, podem ser subestimadas as cargas fatoriais e as correlações entre os fatores, entre outros parâmetros (DiStefano, 2002). A área da Psicologia é particularmente suscetível a esse problema porque as variáveis de estudo tendem a ser assimétricas, e também porque os itens dos inventários e escalas são, com frequência, pontuados em escala Likert, o que produz variáveis categóricas ordenadas de natureza discreta.

Uma solução potencial é dar preferência a estimadores que empregam a matriz de correlações policóricas ou tetracóricas entre os itens durante a análise fatorial. Correlações dessa natureza tendem a ser, em comparação ao coeficiente de Pearson, uma estimativa mais consistente da verdadeira relação linear entre variáveis que possuem apenas algo entre dois a cinco valores possíveis. Por isso, análises fatoriais com o uso de estimadores que se valem de correlações policóricas tendem a acertar com mais frequência o número de fatores subjacentes aos dados, produzindo ainda estimativas paramétricas mais consistentes de cargas fatoriais e correlações entre fatores (Asún, Rdz-Navarro, & Alvarado, 2015; Holgado-Tello et al., 2010; Lara, 2014). No recente estudo de Asún et al. (2015), os autores recomendaram os estimadores *Unweighted Least Squares* (ULS ou Quadrados Mínimos não-Ponderados) e *Weighted Least Squares Mean- and Variance-adjusted* (WLSMV ou Quadrados Mínimos Ponderados Robustos) implementados com correlações policóricas. Ambos se mostraram superiores ao tradicional estimador Maximum Likelihood (ML ou Máxima Verossimilhança, que se vale de correlações de Pearson, e assume a normalidade dos dados) em 210 condições de simulação, variando a correlação entre os fatores, a qualidade dos indicadores, a assimetria dos indicadores e o tamanho amostral. Assim, há evidências que fundamentam a escolha pelos estimadores ULS e WLSMV no lugar do ML em uma diversidade de situações de pesquisa.

No presente editorial, dados simulados são empregados para ilustrar as vantagens dos estimadores ULS e WLSMV com correlações policóricas em comparação ao tradicional ML com correlações de Pearson quando os itens são categorias ordenadas. Com o uso do pacote *simsem* do programa R, foi simulado um banco de dados contendo 2.000 casos e, como descrito na Tabela 1¹, 10 itens explicados por dois fatores com correlação 0,30 entre eles. O primeiro fator (F1) explica os itens y1-y5, enquanto o segundo fator (F2) explica os itens y6-y10. Inicialmente, foram gerados itens contínuos com distribuição normal (média = 0, desvio-padrão = 1), que então foram transformados em variáveis com três categorias de resposta, com base nos pontos de corte 0,00 e 1,50. O objetivo foi tentar reproduzir uma situação, não rara na pesquisa em Psicologia, em que os itens possuem poucas categorias de resposta e uma distribuição moderadamente assimétrica. Análises fatoriais exploratórias dos dados simulados foram conduzidas com o pacote *psych* por meio dos estimadores ML e ULS, e análises confirmatórias foram conduzidas com o pacote *lavaan* usando ML e WLSMV.

¹ Os dados e os códigos estão disponíveis mediante contato pelo e-mail hauck.nf@gmail.com

É possível ver que, em todos os casos, os estimadores ULS e WLSMV baseados em correlações policóricas produziram cargas fatoriais e correlações entre fatores mais próximas dos valores verdadeiros usados na simulação dos dados. Em outras palavras, mesmo que de forma sutil em alguns casos, esses estimadores obtiveram melhor desempenho na tentativa de recuperar os verdadeiros parâmetros do modelo em comparação ao estimador ML. Um aspecto chama ainda a atenção para as possíveis implicações do emprego desses estimadores distintos. Um dos pontos de corte adotados com mais frequência por pesquisadores para considerar um item como discriminativo de um fator é uma carga fatorial com valor mínimo de 0,30. Os resultados mostram que um pesquisador poderia, inadvertidamente, concluir que os indicadores y5 e y10 (carga fatorial verdadeira = 0,30) não são minimamente discriminativos com base nos resultados do estimador ML, o que o levaria a excluir do instrumento dois itens potencialmente úteis. O desfecho poderia ser diferente no caso de usar os estimadores ULS ou WLSMV (única exceção foi a carga 0,29 do item y5 com ULS exploratório).

Tabela 1
Resultados das análises dos dados simulados

	Cargas fatoriais verdadeiras		Assimetria empírica	Análise exploratória				Análise confirmatória			
				ML		ULS		ML		WLSMV	
	F1	F2		F1	F2	F1	F2	F1	F2	F1	F2
y1	0,70	0,00	0,58	0,65	-0,02	0,74	-0,02	0,64	-	0,73	-
y2	0,70	0,00	0,62	0,67	-0,01	0,76	-0,01	0,67	-	0,75	-
y3	0,70	0,00	0,65	0,61	0,01	0,70	0,02	0,62	-	0,71	-
y4	0,70	0,00	0,70	0,60	0,02	0,69	0,02	0,61	-	0,70	-
y5	0,30	0,00	0,59	0,25	0,02	0,29	0,03	0,26	-	0,31	-
y6	0,00	0,70	0,62	0,00	0,56	0,00	0,65	-	0,56	-	0,65
y7	0,00	0,70	0,57	0,02	0,61	0,01	0,70	-	0,61	-	0,71
y8	0,00	0,70	0,62	0,00	0,62	-0,01	0,71	-	0,62	-	0,70
y9	0,00	0,70	0,61	-0,01	0,61	-0,01	0,69	-	0,61	-	0,69
y10	0,00	0,30	0,66	0,03	0,26	0,03	0,30	-	0,27	-	0,32
Correlação entre os fatores				0,26		0,27		0,27		0,28	

Nota. Em negrito, as cargas fatoriais empíricas acima de 0,30

Não obstante, duas ressalvas devem ser feitas. A primeira é que, apesar de subestimar cargas fatoriais e correlações, o estimador ML se mostrou robusto às violações da normalidade dos dados na análise confirmatória, revelando um bom ajuste do modelo aos dados, comparável àquele produzido pelo estimador WLSMV [$\chi^2_{ML}(34)=22,16, p=0,941, CFI_{ML}=1,00, RMSEA_{ML}=0,00; \chi^2_{WLSMV}(34)=14,68, p=0,998, CFI_{WLSMV}=1,00, RMSEA_{WLSMV}=0,00$]. Esse resultado também foi encontrado por Rhemtulla et al. (2012). A segunda ressalva é que o raciocínio aqui apresentado somente se aplica à situação em que o modelo testado pelo pesquisador é o modelo verdadeiro, uma vez que as estimativas paramétricas jamais serão consistentes se o modelo hipotetizado não reproduzir a estrutura verdadeira que produziu os dados. Ambos os pontos colocam a ênfase da modelagem na necessidade de uma compreensão causal prévia de como as variáveis estão relacionadas, mais do que em meramente buscar um bom ajuste do modelo aos dados (Hayduk, 2014).

O presente editorial está longe de querer apresentar um estudo sério de simulação de dados, uma vez que as análises se valem de apenas um banco de dados com 10 indicadores todos com a mesma dificuldade de endosso de suas categorias. Ainda assim, defende-se que a análise ilustrativa é o suficiente para chamar a atenção para os benefícios da escolha empiricamente fundamentada do estimador em uma análise fatorial.

Nelson Hauck Filho
Editor Associado
Universidade São Francisco

Referências

- Asún, R. A., Rdz-Navarro, K., & Alvarado, J. M. (2015). Developing multidimensional likert scales using item factor analysis: The case of four-point items. *Sociological Methods & Research*, 0049124114566716-. doi:10.1177/0049124114566716
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(3), 327-346. doi:10.1207/S15328007SEM0903_2
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5(2), 155-174.
- Hayduk, L. A. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, 74(6), 905-926. doi:10.1177/0013164414527449
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153-166. doi:10.1007/s11135-008-9190-y
- Lara, S. A. D. (2014). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. *Revista Argentina de Ciencias del Comportamiento*, 6(1), 39-48.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning (Multivariate Applications Series)*. New York: Routledge.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354-373. doi:10.1037/a0029315