
Editorial

DOI: <http://dx.doi.org/10.15689/ap.2018.1703.ed>

Como já destacado em outros editoriais, a modelagem de variáveis latentes envolve uma série de decisões nem sempre fáceis de tomar. Uma dessas decisões é teoricamente informada, e antecede outras etapas analíticas: saber se uma variável latente é de tipo dimensional ou categórico. O presente editorial aborda a questão de forma muito sucinta, indicando a análise taxométrica como uma possibilidade para decidir qual o caminho a seguir.

Variáveis latentes podem ocorrer de diversas maneiras, sendo dois dos tipos mais usuais o dimensional e o categórico (Masyn, Henderson, & Greenbaum, 2010). O primeiro caso ocorre quando a variável apresenta níveis ou gradações, enquanto o segundo ocorre quando ela apresenta grupos ou classes latentes (Borsboom, 2008; Meehl, 1992). Por exemplo, escalas de avaliação da depressão, em geral, partem da premissa de que os indivíduos se distinguem em variados graus de severidade de depressão, o que caracteriza a dimensionalidade (Baptista & Gomes, 2011; Coutinho, Hamdan, & Baptista, 2016). Isso significa que alguns indivíduos possuem níveis baixíssimos na variável, enquanto pacientes apresentam altos escores. Em contraste, o DSM-5 endossa uma abordagem categórica binária, segundo a qual a pessoa tem ou não tem um transtorno depressivo (American Psychiatric Association, 2013). Nessa abordagem, não existem níveis, mas apenas duas possibilidades, ou seja, ser ou não membro de um grupo de indivíduos com o transtorno.

O tipo de variável latente orienta a escolha pelo método de análise. Enquanto estudos que seguem uma modelagem dimensional tendem a utilizar análise fatorial (e.g., Gomez & McLaren, 2015), análise de classes latentes ou *clusters* são a escolha de investigações em que se testa a existência de grupos (e.g., Herman, Ostrander, Walkup, Silva, & March, 2007). A primeira abordagem produz fatores, e a segunda produz classes ou grupos latentes. O problema é que, havendo uma controvérsia sobre a natureza latente de uma determinada variável, o pesquisador pode ficar sem saber se deve conduzir um estudo fatorial ou de classes latentes. Por exemplo, no caso da depressão, qual deve ser a melhor estratégia: dimensional ou categórica?

A análise taxométrica é um método desenvolvido para oferecer uma resposta a essa questão preliminar (Meehl, 1995). Existem diferentes técnicas taxométricas, como MAMBAC (Meehl & Yonce, 1994), MAXEIG e Latent Mode (Waller & Meehl, 1998), todas elas elaboradas para informar ao pesquisador se um conjunto de (bons) indicadores avaliam uma variável latente contínuo ou categórica. Essas técnicas não informam quantos fatores ou classes latentes existem nos dados, mas, antes disso, indicam se o pesquisador deveria buscar extrair fatores ou grupos. Em outras palavras, direcionam o pesquisador para um de dois caminhos distintos: a modelagem dimensional ou categórica de variáveis latentes. No caso da depressão, embora existam resultados mistos, os estudos taxométricos têm indicado, em sua grande parte, uma estrutura dimensional (por exemplo, Liu, 2016).

O propósito desse editorial não é oferecer uma explicação detalhada sobre métodos taxométricos, algo que leitoras e leitores podem obter a partir de fontes melhores (Beauchaine, 2013; Hauck-Filho, Costa, & Cordeiro, 2017; Lenzenweger, 2004; Walters, 2012). Não obstante, alguns dos benefícios de usar a análise taxométrica podem ser facilmente demonstrados por meio de uma simulação de dados. A ideia é exemplificar o como a decisão entre análise fatorial ou classes latentes pode ser auxiliada pelas informações de uma análise taxométrica.

Os dados foram simulados da seguinte maneira. Especificou-se um fator como causa latente para as respostas a seis variáveis contínuas, com distribuição normal, média=0 e $DP=1$. Os três primeiros indicadores foram criados de modo a ter carga moderada no fator (0,40), enquanto os demais tiveram forte associação com o fator (0,70). A partir desse modelo, foram simuladas respostas de 2.000 indivíduos. A simulação foi conduzida utilizando o pacote *simsem* do programa R (os códigos podem ser obtidos mediante contato com o autor deste trabalho).

Após gerados, os dados foram usados para testar dois tipos de modelos: modelos fatoriais e modelos de classes latentes (perfis latentes). Foram testados modelos de um e de dois fatores (o máximo possível com seis indicadores) e modelos com um crescente número de classes latentes. Para possibilitar a comparação entre modelos, o estimador MLR foi utilizado em todas as análises. Os índices eleitos para comparação foram o valor *loglikelihood* (LL; quanto maior, melhor), *Akaike Information Criterion* (AIC; quanto menor, melhor) e *Bayesian Information Criterion* (BIC; quanto menor, melhor).

Adicionalmente, para ajudar na determinação do número de classes latentes, utilizou-se o teste Vuong-Lo-Mendel-Rubin (Lo, Mendell, & Rubin, 2001). Esse teste é comumente utilizado em estudos com classes latentes (Nylund, Asparouhov, & Muthén, 2007), e permite saber se há uma melhora significativa no ajuste do modelo ao acrescentar uma classe. Por exemplo, se o p valor para o modelo de três classes é significativo, isso sugere que esse modelo se ajusta melhor aos dados quando comparado a um modelo de duas classes apenas. Essas análises foram realizadas com o programa Mplus.

Os resultados do teste dos modelos estão descritos na Tabela 1. Como se pode ver, quando comparadas as soluções fatoriais, percebe-se que os índices de ajuste indicaram o modelo de um fator como aquele mais representativo da estrutura dos dados. Por outro lado, entre os modelos de classes, as recomendações foram o modelo de duas classes, a partir dos índices LL, AIC e BIC, e o modelo de três classes, a partir do teste VLMR. Se fosse dado mais crédito ao teste VLMR, que tende a apresentar um bom desempenho (Nylund et al., 2007), a escolha seria por reter três classes. Essa decisão poderia ter também amparo teórico: alguém poderia estar interessado em inspecionar um maior número de classes, pois elas poderiam ajudar a identificar pequenos grupos-alvo de indivíduos, como aqueles com determinada psicopatologia ou condição.

Tabela 1
Resultados das análises fatoriais e de classes latentes

Modelos	LL	AIC	BIC	VLMRT
Modelos fatoriais				
1 fator	-7981,25	15998,50	16086,84	-
2 fatores	-7979,06	16004,12	16117,00	-
Modelos de classes				
2 classes	-8083,90	16205,81	16299,06	<0,001
3 classes	-8000,61	16053,21	16180,82	<0,001
4 classes	-7980,81	16027,61	16189,57	0,110

As implicações dessas análises são importantes. Vamos supor que um pesquisador A tivesse escolhido conduzir um estudo de análise fatorial, enquanto outro pesquisador B estivesse interessado em uma modelagem de classes latentes desses mesmos dados. A partir dos resultados da Tabela 1, talvez ambos chegassem a conclusões marcadamente diversas: A diria ter encontrado um fator, mas B relataria a existência de três classes latentes. A questão é: como saber qual o caminho a seguir?

Uma possível resposta pode estar na análise taxométrica. Ao analisar esses mesmos dados com os métodos taxométricos MAMBAC, MAXEIG e LMode, obtêm-se os seguintes índices de ajuste: $CCFI_{MAMBAC}=0,421$, $CCFI_{MAXEIG}=0,383$, $CCFI_{Lmode}=0,221$, além de um $CCFI_{Médio}=0,342$. Ruscio, Walters, Marcus e Kacetow (2010) recomendaram considerar valores CCFI médios de 0 a 0,40 como evidência de dimensionalidade, e de 0,60 a 1 como evidência da presença de categorias latentes não-arbitrárias. Nesse caso, as evidências seriam sugestivas de dimensionalidade latente, e o caminho indicado de modelagem dos dados seria a análise fatorial. Seguindo esse conselho, a análise fatorial (como visto na Tabela 1) levaria ao modelo de um fator, que é o modelo verdadeiro utilizado para produzir os dados. Ou seja, a análise taxométrica, indiretamente, conduziria o pesquisador a descobrir a verdadeira estrutura dos dados.

O presente exercício é bastante simples, mas ilustra a utilidade de conduzir um estudo taxométrico quando não está claro se a estrutura latente de um fenômeno é dimensional ou categórica (Lenzenweger, 2004). Como demonstrado, a análise fatorial vai sempre produzir fatores, e a análise de classes latentes vai produzir grupos, independentemente de qual for a verdadeira estrutura dos dados. Tomar a decisão correta depende, portanto, de informação teórica externa a esses métodos. Pesquisadoras e pesquisadores podem se valer de resultados taxométricos—seja conduzindo um estudo ou revisando estudos publicados—para decidir entre ambas as abordagens.

A análise taxométrica tem sido indicada como uma etapa preliminar à modelagem dos dados (Walters, 2012). Ela não é capaz de dizer quantos fatores ou quantas classes latentes existem, mas permite ter uma noção sobre qual das duas é a abordagem mais apropriada de modelagem dos dados. A taxometria pode, portanto, aprofundar modelos teóricos sobre fenômenos psicológicos e também ajudar a refinar instrumentos e métodos de avaliação psicométrica.

Nelson Hauck Filho
Editor-chefe
hauck.nf@gmail.com

Referências

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*. Washington: American Psychiatric Publishing.
- Baptista, M. N., & Gomes, J. O. (2011). Escala Baptista de Depressão (Versão Adulto) - EBADEP-A: evidências de validade de construto e de critério. *Psico-USF (Impresso)*, *16*(2), 151-161. doi: 10.1590/S1413-82712011000200004
- Beauchaine, T. (2013). Taxometrics. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis* (pp. 612-634). New York: Oxford University Press.
- Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089-1108. doi: 10.1002/jclp.20503
- Coutinho, F. L., Hamdan, A. C., & Baptista, M. N. (2016). Escala Baptista de Depressão para Idosos – EBADEP-ID: evidências de validade. *Perspectivas En Psicología*, *13*(2), 1-9. doi: 10.1590/S1413-82712011000200004
- Gomez, R., & McLaren, S. (2015). The Center for Epidemiologic Studies Depression Scale: Support for a Bifactor Model With a Dominant General Factor and a Specific Factor for Positive Affect. *Assessment*, *22*(3), 351-360. doi: 10.1590/S1413-82712011000200004
- Hauck-Filho, N., Costa, A. R. L., & Cordeiro, N. C. A. (2017). Análise taxométrica: decidindo entre a modelagem dimensional ou categórica de variáveis latentes. In B. F. Damásio & J. C. Borsa (Eds.), *Manual de desenvolvimento de instrumentos psicológicos* (pp. 417-436). São Paulo: Vetor.
- Herman, K. C., Ostrander, R., Walkup, J. T., Silva, S. G., & March, J. S. (2007). Empirically derived subtypes of adolescent depression: Latent profile analysis of co-occurring symptoms in the Treatment for Adolescents with Depression Study (TADS). *Journal of Consulting and Clinical Psychology*, *75*(5), 716-728. doi: 10.1037/0022-006X.75.5.716
- Lenzenweger, M. F. (2004). Consideration of the challenges, complications, and pitfalls of taxometric analysis. *Journal of Abnormal Psychology*, *113*(1), 10-23. doi: 10.1037/0021-843X.113.1.10
- Liu, R. T. (2016). Taxometric evidence of a dimensional latent structure for depression in an epidemiological sample of children and adolescents. *Psychological Medicine*, *46*(6), 1265. doi: 10.1017/S0033291715002792
- Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778. doi: 10.1093/biomet/88.3.767
- Masyn, K. E., Henderson, C. E., & Greenbaum, P. E. (2010). Exploring the Latent Structures of Psychological Constructs in Social Development Using the Dimensional-Categorical Spectrum. *Social Development*, *19*(3), 470-493. doi: 10.1111/j.1467-9507.2009.00573.x
- Meehl, P. E. (1992). Factors and Taxa, Traits and Types, Differences of Degree and Differences in Kind. *Journal of Personality*, *60*(1), 117-174. doi: 10.1111/j.1467-6494.1992.tb00269.x
- Meehl, P. E. (1995). Bootstraps taxometrics. Solving the classification problem in psychopathology. *The American Psychologist*, *50*(4), 266-275. doi: 10.1037/0003-066X.50.4.266
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059-1274.
- Nylund, K., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535-569. doi: 10.1080/10705510701575396
- Ruscio, J., Walters, G. D., Marcus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment*, *22*(1), 5-21. doi: 10.1037/a0018259
- Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.
- Walters, G. D. (2012). Taxometrics and Criminal Justice: Assessing the Latent Structure of Crime-Related Constructs. *Journal of Criminal Justice*, *40*(1), 10-20. doi: 10.1016/j.jcrimjus.2011.11.003