



Editorial

É possível identificar o nível de medida de variáveis latentes?

Víthor Rosa Franco 

Universidade São Francisco, Campinas-SP, Brasil

Neste editorial, apresentarei um breve estudo de simulação comparando procedimentos que, a princípio, nos auxiliam a responder à pergunta do título. Quando trabalhamos com psicometria, muitas vezes, nosso objetivo é criar estimativas de magnitude de uma variável latente. Dessa forma, é um pressuposto de procedimentos psicométricos convencionais como a Análise Fatorial (exploratória e confirmatória) e a Teoria de Resposta ao Item que o traço latente seja uma variável intervalar (McDonald, 2013). No entanto, existem procedimentos menos habituais em psicometria, mas bastante utilizados no contexto médico e psiquiátrico pela natureza mais dicotômica da área (ou seja, presença ou ausência de uma doença), que assumem que as variáveis latentes são categóricas. Entre os mais conhecidos, estão a Análise de Perfis Latentes e a Análise de Classes (ou Estruturas) Latentes (Gibson, 1959).

A discussão sobre o nível de medida das variantes latentes e como identificá-lo não é recente. O procedimento geralmente mais utilizado para tentar se identificar o nível de uma variável latente é a análise taxométrica proposta por Meehl (1995). Embora existam várias formas de se conduzir uma análise taxométrica (p.ex., Walters & Ruscio, 2010), o fundamento é identificar se a distribuição da variável latente é uni ou multimodal. Se ela for unimodal, a análise taxométrica irá dizer que existem mais evidências a favor de que o nível de medida seja intervalar. Se a distribuição da variável latente for multimodal, a análise taxométrica irá dizer que existem evidências a favor de que o nível de medida seja categórico.

Este fundamento da análise taxométrica, apesar de ser coerente à primeira vista, não apresenta robustez em termos de teoria estatística, dado que o modelo não é falseável. Mais especificamente, a estrutura de covariância observada nos dados sempre permitirá os ajustes idênticos de modelos com K classes latentes em comparação com modelos com $K - 1$ fatores latentes (Gibson, 1959). Em alguns casos, os testes de multimodalidade latente irão, de fato, identificar subgrupos latentes (Steinley & McDonald, 2007). No entanto, não existe consenso que tal resultado deva ser interpretado como evidência de existência de uma variável latente categórica. Por exemplo, nos modelos de mistura da TRI (Rost, 1990), que são similares em termos de procedimentos de estimação à análise taxométrica, a identificação de classes latentes é interpretada apenas como a presença de subgrupos nos dados, embora a principal variável de interesse continue sendo assumida como intervalar.

Há, no entanto, algumas limitações na interpretação dos resultados das análises taxométricas. Em termos de desempenho dos procedimentos taxométricos, Ruscio e Kaczetow (2009) demonstraram a partir de estudos com mais de 25 mil dados simulados, que o procedimento de índice de ajuste de comparação de curvas consegue, com uma acurácia de 93%, identificar o nível de medida da variável latente. No entanto, Haslam et al. (2012), a partir de uma meta-análise, identificaram que pesquisas fundamentadas em métodos mais robustos tendiam a mostrar menos evidências a favor de um nível de medida categórico do que de um nível de medida intervalar para a variável latente.

A partir dessa discussão, pode-se perguntar: será realmente possível identificar o nível de medida de variáveis latentes? Para além da literatura estatística e psicométrica, uma resposta plausível a esta pergunta pode ser encontrada na literatura de teoria de mensuração (Suppes, 1959) e na teoria de mensuração conjunta aditiva (*additive conjoint measurement theory*; ACM). A teoria de mensuração é uma área de estudo da psicologia matemática que busca estabelecer qual a forma mais adequada de se representar objetos numericamente. Mais especificamente, os problemas de representação e unicidade permitem inferir o nível de medida dadas condições suficientes e necessárias de relações empíricas. A ACM, neste contexto, é apenas uma entre várias teorias de mensuração (ver Krantz et al., 1971) que pode ser aplicada a fenômenos psicológicos e, apesar de ser raramente utilizada em psicometria (Michell, 2020), é provavelmente a teoria de mensuração mais conhecida (Heene, 2013).

São raros os exemplos de interseção entre a teoria de mensuração e a psicometria na literatura. Em um desses exemplos, Scheiblechner (1995; 1999; 2003) propôs três modelos não-paramétricos de TRI baseados nos axiomas da

ACM. Esses modelos estão em crescente nível de restrições sobre os pressupostos em relação ao nível de medida das variáveis observadas e latentes. No modelo probabilístico isotônico ordinal (*isotonic ordinal probabilistic model*, ISOP), tanto as variáveis latentes quanto as variáveis observadas são assumidas como ordinais. No modelo probabilístico isotônico aditivo (*additive isotonic probabilistic model*, ADISOP), as variáveis latentes são assumidas como intervalares, enquanto as variáveis observadas são assumidas como ordinais. Por fim, no modelo probabilístico isotônico completamente aditivo (*completely additive isotonic probabilistic model*, CADISOP), tanto as variáveis latentes quanto as variáveis observadas são assumidas como intervalares.

O modelo ISOP é similar a uma versão probabilística da escala de Guttman (Guttman, 1944). Já o modelo ADISOP é similar ao modelo de Rasch, mas sem exigir uma forma específica para a função de resposta ao item. Por fim, o modelo CADISOP é similar ao modelo politômico de Rasch, mas também sem exigir uma forma específica para a função de resposta ao item. Scheiblechner (1999) propôs métodos para estimação por mínimos quadrados ponderados que são equivalentes a métodos de estimação por máxima verossimilhança para modelos exponenciais generalizados. Apesar disso, esses modelos são raramente utilizados e, até onde foi possível identificar, apenas os modelos ISOP e ADISOP foram implementados, em apenas um software estatístico (Robitzsch, 2020).

Para ilustrar como estes métodos podem ser utilizados, foram simulados 190 bancos de dados a partir do modelo de Rasch, cada banco com um tamanho amostral entre 100 e 1.000 (com acréscimos progressivos de 50 casos) e um tamanho do teste (i.e., quantidade itens) entre 10 e 100 (com acréscimos progressivos de 10 itens). Na Figura 1, é apresentado o efeito do tamanho do teste no ajuste dos modelos (este sendo calculado pelo critério de informação Bayesiano – BIC – e normalizado em cada tamanho amostral). Foram ajustados três modelos: ISOP (representado pela linha sólida); ADISOP (representado pela linha descontinuada); e Rasch (representado pela linha pontilhada). É possível observar que, conforme o tamanho do teste aumenta, a evidência a favor do modelo ISOP diminui (dado que a linha sólida está aumentando), mas a evidência a favor dos modelos ADISOP e Rasch aumentam (dado que ambas as linhas mostram tendência de diminuição). No entanto, pelo menos nas condições testadas, o modelo de Rasch é sempre o modelo com pior ajuste. A partir de em torno de 35 itens, o modelo ADISOP tende a ser aquele que melhor se ajusta aos dados.

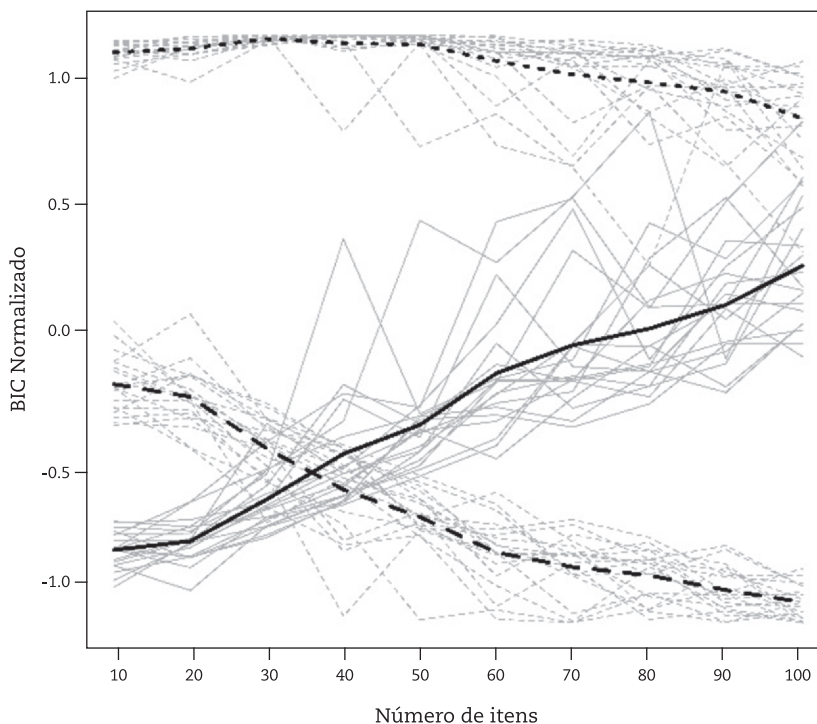


Figura 1. Efeitos do Tamanho do Teste no Ajuste Normalizado dos Modelos. A Linha Sólida Representa o Modelo ISOP, a Linha Descontínua Representa o Modelo ADISOP e a Linha Pontilhada Representa o Modelo de Rasch

A partir da Figura 2, é possível observar que o tamanho amostral não parece ter tanto efeito quanto a quantidade de itens, embora alguma tendência possa ser observada. Para o modelo ISOP, conforme o tamanho amostral aumenta, melhor é o ajuste normalizado do modelo. Para o modelo ADISOP e para o modelo de Rasch, no entanto, o efeito é contrário: aumentar o tamanho amostral pior o ajuste destes modelos. Pode-se concluir, pela avaliação de ambas as figuras, que o modelo ADISOP é aquele que, de forma geral, melhor se ajusta aos dados. Este é um resultado intrigante, dado que os dados foram sempre simulados utilizando o modelo de Rasch e, portanto, era esperado que este fosse o melhor a se ajustar aos dados.

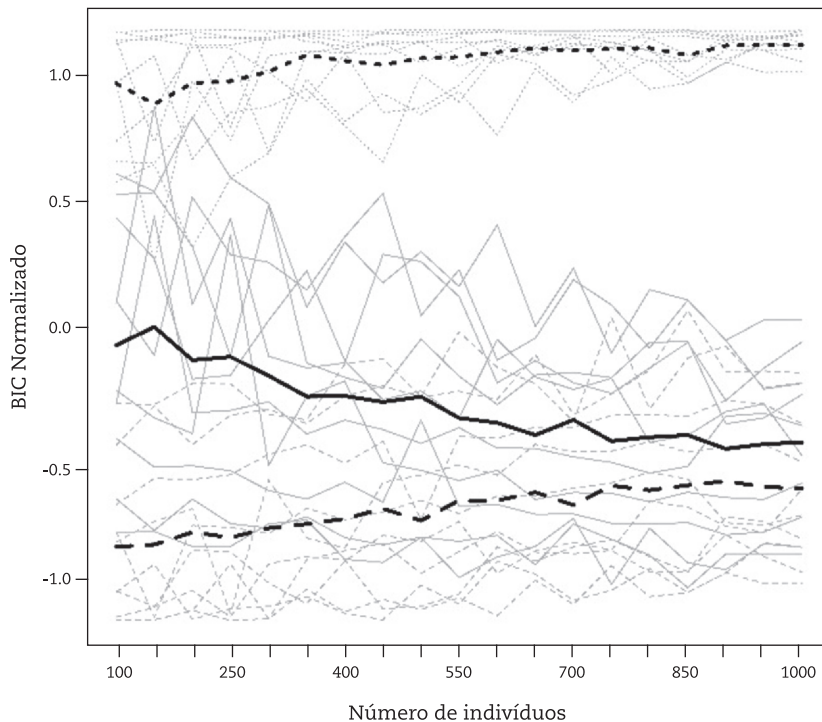


Figura 2. Efeitos do Tamanho Amostral no Ajuste Normalizado dos Modelos. A linha Sólida Representa o Modelo ISOP, a Linha Descontínua Representa o Modelo ADISOP e a Linha Pontilhada Representa o Modelo de Rasch

Os resultados da nossa simulação parecem apontar a três conclusões principais. Primeiro, é necessário aumentar a quantidade de itens para que possamos de fato alcançar medidas latentes intervalares. Em segundo lugar, aumentar o tamanho amostral não necessariamente irá melhorar a qualidade de ajuste do seu modelo. Em terceiro lugar, utilizar funções paramétricas para ajustar modelos de TRI, mesmo quando estas funções são as mesmas que as do processo gerador de dados, podem levar a um ajuste ruim. Essas conclusões podem ser diretamente derivadas da ACM e da escala de Guttman. De acordo com a escala de Guttman, em instrumentos com itens dicotômicos, caso o processo de resposta não apresentasse erros, só deveria ser possível ordenar os indivíduos de acordo com a quantidade de itens mais um. Por exemplo, com 10 itens, seria possível identificar apenas 11 diferentes padrões determinísticos de resposta. Dessa forma, aumentar o tamanho amostral pode aumentar a variância dos indivíduos, mas não aumenta a capacidade do instrumento de distingui-los e, portanto, isso não deveria, necessariamente, afetar a capacidade de se distinguirem modelos ordinais de intervalares. Além disso, de acordo com a ACM, quanto maior a quantidade de itens e de escores individuais possíveis, menor a chance de o modelo de medida se ajustar aleatoriamente (Arbuckle & Larimer, 1976).

Por fim, pode-se retomar a pergunta que motivou este editorial: será realmente possível identificar o nível de medida de variáveis latentes? A resposta mais adequada parece ser “depende”. No exemplo, foi ilustrado como se poderia identificar se a variável latente é mensurada num nível ordinal ou intervalar e, pelo menos a princípio, existem condições que permitem concluir, corretamente, em qual nível de medida as variáveis latentes estão de fato. No entanto, é importante salientar que, assim como o modelo de Rasch, os modelos ISOP e ADISOP também irão se ajustar aos dados mesmo quando os dados forem simulados utilizando lançamentos aleatórios de uma moeda como o modelo gerador (Wood, 1978). Mesmo que os dados sejam gerados por um processo latente categórico, os modelos ajustados

necessariamente atribuiriam pelo menos um nível ordinal a esse processo latente. Por essa razão, é necessário se definir melhores condições e modelos para se realizar esses tipos de teste. Espera-se que mais pesquisas na interseção entre a teoria de mensuração e a psicometria sejam realizadas e ajudem a melhor responder à pergunta apresentada neste Editorial.

Referências

- Arbuckle, J., & Larimer, J. (1976). The number of two-way tables satisfying certain additivity axioms. *Journal of Mathematical Psychology*, *12*, 89-100. [https://doi.org/10.1016/0022-2496\(76\)90036-5](https://doi.org/10.1016/0022-2496(76)90036-5)
- Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, *24*, 229-252. <https://doi.org/10.1007/BF02289845>
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139-150. <https://doi.org/10.2307/2086306>
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine*, *42*(5), 903-920. <https://doi.org/10.1017/S0033291711001966>
- Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in Psychology: Quantitative Psychology and Measurement*, *4*, Article 246. <https://doi.org/10.3389/fpsyg.2013.00246>
- Krantz, D. H., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Vol. I: Additive and polynomial representations*. Academic Press.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press.
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266-275. <https://doi.org/10.1037/0003-066X.50.4.266>
- Michell, J. (2020). Thorndike's credo: Metaphysics in psychometrics. *Theory & Psychology*, *30*(3), 309-328. <https://doi.org/10.1177/0959354320916251>
- Robitzsch, A. (2020). *sirt: Supplementary Item Response Theory Models*. R package version 3.9-4, <https://CRAN.R-project.org/package=sirt>.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271-282. <https://doi.org/10.1177/014662169001400305>
- Ruscio, J., & Kaczetow, W. (2009). Differentiating categories and dimensions: Evaluating the robustness of taxometric analyses. *Multivariate Behavioral Research*, *44*, 259-280. <https://doi.org/10.1080/00273170902794248>
- Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika*, *60*(2), 281-304. <https://doi.org/10.1007/BF02301417>
- Scheiblechner, H. (1999). Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika*, *64*(3), 295-316. <https://doi.org/10.1007/BF02294297>
- Scheiblechner, H. (2003). Nonparametric IRT: Testing the bi-isotonicity of isotonic probabilistic models (ISOP). *Psychometrika*, *68*(1), 79-96. <https://doi.org/10.1007/BF02296654>
- Steinley, D., & McDonald, R. P. (2007). Examining factor score distributions to determine the nature of latent spaces. *Multivariate Behavioral Research*, *42*(1), 133-156. <https://doi.org/10.1080/00273170701341217>
- Suppes, P. (1959). Measurement, empirical meaningfulness and three-valued logic. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 129-143). Wiley.
- Walters, G. D., & Ruscio, J. (2010). Where do we draw the line? Assigning cases to subsamples for MAMBAC, MAXCOV, and MAXEIG taxometric analyses. *Assessment*, *17*(3), 321-333. <https://doi.org/10.1177/1073191109356539>
- Wood, R. (1978). Fitting the Rasch model – A heady tale. *British Journal of Mathematical and Statistical Psychology*, *31*(1), 27-32. <https://doi.org/10.1111/j.2044-8317.1978.tb00569.x>

Como citar este artigo

Franco, V. R. (2021). É possível identificar o nível de medida de variáveis latentes? [Editorial]. *Avaliação Psicológica*, *20*(2), A-D. <http://dx.doi.org/10.15689/ap.2021.2002.e2>