

Análise da confiabilidade de instrumentos

Prof. Dr. Juliano Bortolini

Depto. de Estatística da UFMT

@profjulianobortolini



Fidedignidade (*ou confiabilidade ou precisão*)

Susana Urbina (2007):

- Fidedignidade refere-se à consistência e precisão dos resultados obtidos por um instrumento de mensuração.
- Se o teste for repetido, com os mesmos indivíduos ou grupos, espera-se que os resultados sejam próximos (pequena variabilidade).
- Capacidade de um teste produzir resultados estáveis e coerentes quando administrado em condições semelhantes.

Susana Urbina (2007). Fundamentos da testagem psicológica. Artmed.

Diferença entre Validade e Fidedignidade

...Lembrando:

- **Validade** é a capacidade do instrumento medir exatamente o que ele pretende avaliar. Um instrumento válido é aquele que cumpre seu propósito e mede o que é esperado.
 - *Exemplo: um teste de ansiedade deve medir o nível de ansiedade dos indivíduos, e não outros aspectos como autoestima ou satisfação.*
- **Fidedignidade**: indica a estabilidade e a consistência das medições de um instrumento. Um teste fidedigno gera resultados semelhantes em condições semelhantes, independentemente de medir com precisão o construto de interesse.
 - *Exemplo: Um termômetro que sempre mostra a mesma temperatura quando usado em condições iguais é confiável (fidedigno), mas pode não ser válido se estiver calibrado incorretamente.*

Fidedignidade na Teoria Clássica dos Testes

$$\text{Escore Observado} = \text{Escore Verdadeiro} + \text{Erro aleatório}$$

- *Escore Verdadeiro*: valor teórico, nível real de habilidade ou característica.
- *Erro aleatório*: variações nos escores que não refletem o verdadeiro nível do construto.
 - *Exemplo: cansaço, distração, condições ambientais, ...*

Fidedignidade como proporção de variâncias:

$$r = \frac{\sigma_{\mathbf{V}}^2}{\sigma_{\mathbf{O}}^2} = \frac{\sigma_{\mathbf{V}}^2}{\sigma_{\mathbf{V}}^2 + \sigma_{\mathbf{E}}^2}$$

Aceitável: $\hat{r} > 0,7$

Ideal: $\hat{r} > 0,8$

Calculando a Fidedignidade

1. Consistência Interna

- Avalia o grau de relacionamento **entre os itens** do instrumento.
- Itens que medem o mesmo construto devem estar correlacionados entre si.
 - *Exemplo (ansiedade): itens como "Sinto-me nervoso em situações sociais" e "Evito interações com pessoas desconhecidas" apresentam respostas semelhantes.*

Estatísticas:

- a) método das duas metades (Spearman-Brown)
- b) Kuder-Richardson-20 e coeficiente alfa (de Cronbach)
- c) Ômega de McDonald e fidedignidade composta

a) Método das duas metades (Spearman-Brown)

- Dividir o conjunto de itens de um teste em duas metades e avaliar a correlação entre essas metades.
- Se o teste for confiável, os escores obtidos em ambas as metades devem ser semelhantes.

Problema: o valor da correlação depende da divisão.

Solução: correção de Spearman-Brown (r_{SB})

$$r_{SB} = \frac{2r_{AB}}{1 + r_{AB}}$$

r_{AB} é a correlação entre as duas metades do teste.

Abordagem alternativa (?): realizar todas as divisões possíveis.

b) Kuder-Richardson-20 e coeficiente alfa (de Cronbach)

KR-20: é um índice de fidedignidade específico para instrumentos com itens **dicotômicos**.

- Correlação média de todas as divisões em duas metades do teste.

$$KR20 = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n p_i (1 - p_i)}{s^2} \right)$$

em que,

n é o número de itens do teste,

p_i é a proporção de acertos para o item i e

s^2 é a variância total dos escores dos participantes no teste.

b) Kuder-Richardson-20 e coeficiente alfa (de Cronbach)

Coeficiente alfa: é uma medida de fidedignidade para instrumentos com itens em escala Likert.

- Também é a correlação média de todas as divisões em duas metades do teste.

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n s_i^2}{s^2} \right)$$

em que,

n é o número de itens do teste,

s_i^2 é variância dos escores do item i e

s^2 é a variância total dos escores dos participantes no teste.

b) Kuder-Richardson-20 e coeficiente alfa (de Cronbach)

Limitações:

- Ambos os coeficientes consideram que todos os itens têm **peso igual** (pressuposto da tau-equivalência).
- Esses coeficientes podem subestimar a fidedignidade.
 - *Exemplo (habilidade matemática): itens: (1) adição e (2) cálculo diferencial.*

Solução: análise fatorial.

- A **Análise fatorial** é uma técnica estatística que ajuda a identificar **agrupamentos de itens** e a atribuir **pesos distintos** para cada item de acordo com a sua importância para o construto medido.

c) Ômega de McDonald (1999) (ou fidedignidade composta (Fornell e Larcker, 1981))

- Incorporam as **cargas fatoriais** dos itens no cálculo da fidedignidade.
- Ajusta-se melhor à realidade com itens que possuem relevâncias diferentes.

$$\Omega = \frac{(\sum_{i=1}^n \lambda_i)^2}{(\sum_{i=1}^n \lambda_i)^2 + \sum_{i=1}^n var(\epsilon_i)}$$

em que, λ_i é a carga fatorial do item i e $var(\epsilon_i)$ é a variância do erro do item i .

- O numerador representa a variância total explicada.
- O denominador é a variância total dos escores observados (variância explicada pelo fator mais a variância de erro).
- Esse coeficiente fornece uma estimativa de quão bem os itens do teste medem o mesmo construto, levando em conta o peso de cada item.

Calculando a Fidedignidade

2. Teste-Reteste: é uma técnica de fidedignidade que avalia a **estabilidade temporal**.

- Espera-se que os escores observados sejam semelhantes para os mesmos indivíduos em diferentes momentos, desde que as condições e o construto medido não tenham mudado.
- Não depende da correlação entre itens.
- **Aplicação em dois momentos distintos** para o mesmo grupo de indivíduos.
- **Correlação** entre os escores obtidos nas duas ocasiões. Espera-se uma alta correlação.

Limitação: tamanho do intervalo entre as aplicações. Nem muito curto e nem muito longo.

Estatísticas:

a) Correlação de Pearson

b) Coeficiente de Correlação Intraclass (ICC)

a) Correlação de Pearson

- Mede a força e a direção da relação linear entre os escores de dois momentos.
- Escore normalmente distribuído.
- valor de r próximo de 1 indica uma alta fidedignidade.

b) Coeficiente de Correlação Intraclassa (ICC)

- É uma medida de correlação que considera tanto a **variação entre os indivíduos** quanto a **variação dentro dos indivíduos**.
- **Cálculo do ICC:** Existem diferentes tipos de ICC, mas em geral, ele é calculado com base na decomposição da variância total em componentes de variância entre indivíduos e variância de erro.
- O ICC é mais conservador.
- ICC alto (próximo de 1) sugere boa confiabilidade teste-reteste.

Sendo prático...

Consistência Interna:

- Ômega de McDonald (ou Fidedignidade Composta)
- Coeficiente alfa

Teste-reteste:

- Coeficiente de Correlação Intraclassa
- Correlação de Pearson

Aceitável: $\hat{r} > 0,7$

Ideal: $\hat{r} > 0,8$

Referências

- **Urbina, S.** (2007). Fundamentos da testagem psicológica. Artmed.
- **Pasquali, L.** (2011). Psicometria: teoria dos testes na psicologia e na educação. Vozes.
- McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.