TESTE DE HIPÓTESE E INTERVALO DE CONFIANÇA PARA O COEFICIENTE DE FIDEDIGNIDADE DE UM INSTRUMENTO DE MEDIDA

Maria Teresinha Albanese*

1 – INTRODUÇÃO

No dia-a-dia a palavra "medição" tem um sentido claro e conciso. Para medir em situações práticas, geralmente dispomos de instrumentos físicos que nos dão resultados precisos, em forma de escores. No entanto, a situação é diferente quando queremos medir variáveis psicológicas, em que nos deparamos com problemas de escalonamento muito complexos.

Para que um instrumento de medida ou teste possa ser usado, produzindo resultados dignos de confiança, ele deve ser válido e fidedigno.

A validade refere-se ao fato do instrumento medir realmente o que pretendemos, e a fidedignidade, se o teste mede de forma consistente e precisa, isto é, se ao retornarmos a aplicá-lo nas mesmas condições e às mesmas pessoas, obtemos os mesmos resultados.

Esses dois conceitos estão fortemente relacionados. Um coeficiente de fidedignidade alto é uma condição necessária, mas não suficiente, para um coeficiente de validade também alto. É inútil um teste ser altamente fidedigno, se a sua validade não puder ser considerada satisfatória.

Este artigo trata da fidedignidade de um teste, sob a suposição que sua validade já foi comprovada.

O coeficiente de fidedignidade é afetado em diferentes graus por fatores relativos ao teste (número de itens, amplitude de dificuldade e interdependência dos itens etc.) e aos respondentes (condições físicas, ambientais, velocidade na realização do teste etc.) Portanto, podemos obter, com um mesmo teste e método de estimação, diferentes estimativas para a fidedignidade, dependendo da heterogeneidade do grupo. Em conseqüência, é pouco significativo dispor somente do valor numérico deste coeficiente para descrever um teste como um instrumento de medida.

---

* Departamento de Estatística da Universidade Federal do Rio Grande do Sul.

Entretanto, se pudermos supor que os escores dos itens ou subtestes têm distribuição Normal Multivariada com iguais variâncias e iguais covariâncias, poderemos determinar um estimador de máxima verossimilhança da fidedignidade, e a partir de sua distribuição amostral, construir intervalos de confiança e realizar teste de hipóteses para este coeficiente, solucionando o problema anteriormente mencionado.

2 – MODELO LINEAR PARA TESTES DE TAMANHO FIXO

Suposições

Seja X uma variável aleatória definida sobre uma população P de pessoas, tomando valores x, correspondentes aos escores observados das diferentes pessoas. Seja T uma variável aleatória, assumindo valores τ , correspondentes aos escores verdadeiros, não observáveis destas pessoas.

Então a variável aleatória erro (ε) é definida pela relação linear

$$X = T + \varepsilon \tag{2.1}$$

sobre P.

Para uma determinada pessoa, T é uma constante, enquanto que X e E são variáveis aleatórias.

Sejam os símbolos E, σ^2 e ρ utilizados para representar esperanças, variâncias e correlações, respectivamente.

As suposições do modelo linear incluem a equação (2.1) e as apresentadas a seguir:

$$E(\varepsilon) = 0, \tag{2.2}$$

$$\rho(T, \varepsilon) = 0, \tag{2.3}$$

$$\rho(\varepsilon_i, \varepsilon_j) = 0 \qquad e \tag{2.4}$$

$$\rho(\varepsilon_i, T_j) = 0, \text{ para qualquer } \forall\ i \neq j, \tag{2.5}$$

que são válidas sob qualquer subpopulação não nula de P.

As conseqüências imediatas destas suposições são que

$$E(X) = E(T), \tag{2.6}$$

$$E(X/T = \tau) = \tau\ , \tag{2.7}$$

$$\sigma^2(X) = \sigma^2(T) + \sigma^2(\varepsilon) \qquad e \tag{2.8}$$

$$\rho^2(X, T) = \sigma^2(T)/\sigma^2(X). \tag{2.9}$$

Medidas Paralelas

Definição: Sejam as variáveis aleatórias X e X' tal que $X = T + \varepsilon$ e $X' = T + \varepsilon'$ e $\sigma^2(\varepsilon) = \sigma^2(\varepsilon')$, em qualquer subpopulação não nula de P.

Então X e X' são denominadas medidas paralelas.

Segue-se imediatamente desta definição e das suposições do modelo linear que

$$E(X) = E(X'), \tag{2.10}$$

$$\sigma^2 (X) = \sigma^2 (X') \text{ e} \tag{2.11}$$

$$\sigma(X, X') = \sigma^2 (T), \tag{2.12}$$

em qualquer subpopulação não nula de P.

Das equações (2.10), (2.11) e (2.12), obtemos

$$\rho(X, X') = \sigma^2(T)/\sigma^2(X), \tag{2.13}$$

que é equivalente a $\rho^2 (X, T)$.

Medidas compostas por K componentes

A teoria apresentada anteriormente pode ser estendida facilmente para o modelo linear de testes compostos, isto é, testes formados por um conjunto de itens que, freqüentemente, estão agrupados em subtestes. Precisamos somente considerar que cada item, subteste ou teste gera uma medida e que a medida do subteste é determinada de forma aditiva a partir dos itens, e a do teste é determinada de forma semelhante a partir dos subtestes. Nestas condições, dizemos que a medida do teste é composta e que suas partes são componentes de medida. Esta extensão será representada a seguir.

Sejam as medidas $(Y_i, T_i, \varepsilon_i)$ assumindo valores (y_i, τ_i, e_i), para $i = 1, 2, \ldots, k$. Seja (X, T, ε) uma medida composta, assumindo o valor (x, τ, e), e definida por

$$X = \sum_{i=1}^{k} Y_i. \tag{2.14}$$

Então podemos dizer que

$$T = \sum_{i=1}^{k} T_i \qquad \text{e} \qquad \varepsilon = \sum_{i=1}^{k} \varepsilon_i.$$

Das equações (2.2), (2.4) e (2.14) segue-se que as médias de X, T e E são dadas, respectivamente, por

$$E(X) = E\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} E(Y_i),$$

$$E(T) = E\left(\sum_{i=1}^{k} T_i\right) = \sum_{i=1}^{k} E(T_i) \text{ e}$$

$$E(\varepsilon) = E\left(\sum_{i=1}^{k} \varepsilon_i\right) = \sum_{i=1}^{k} E(\varepsilon_i) = 0,$$

e as variâncias por

$$\sigma^2 (X) = \sum_{i=1}^{k} \sigma^2 (Y_i) + \sum_{i \neq j} \sum \rho(Y_i, Y_j)\, \sigma(Y_i)\, \sigma(Y_j), \tag{2.15}$$

$$\sigma^2(T) = \sum_{i=1}^{k} \sigma^2(T_i) + \sum_{i \neq j} \sum \rho(T_i, T_j)\,\sigma(T_i)\,\sigma(T_j) \quad \text{e}$$

$$\sigma^2(\varepsilon) = \sum_{i=1}^{k} \sigma^2(\varepsilon_i).$$

3 – COEFICIENTE DE FIDEDIGNIDADE

O coeficiente de fidedignidade de um teste é definido como o quadrado da correlação entre o escore verdadeiro e o observado, isto é,

$$\rho^2(X, T) = \frac{\sigma^2(T)}{\sigma^2(X)} = 1 - \frac{\sigma^2(\varepsilon)}{\sigma^2(X)}.$$

Portanto, a fidedignidade de um teste representa a proporção da variância do escore observado, que é explicada pela variância do escore verdadeiro.

Como $\sigma^2(X) \geqslant \sigma^2(\varepsilon)$, o coeficiente de fidedignidade assume valores no intervalo $[0, 1]$. Quanto maior a variância do erro, menor é a fidedignidade do teste.

Da equação (2.13) temos que, quando dispomos de duas medidas paralelas X e X', então o coeficiente de fidedignidade é definido pela correlação entre estas medidas. Portanto, supondo que, no mínimo, um par de medidas paralelas pode ser obtido, podemos expressar uma quantidade não observável $\rho^2(X, T)$, em termos de $\rho(X, X')$, um parâmetro da distribuição bivariada do escore observado. Assim, a estimação de $\rho^2(X, T)$ fica reduzida à estimação de $\rho(X, X')$.

Uma extensão para o caso em que dispomos de k medidas paralelas é dada pelo teorema a seguir, também conhecido como a fórmula de Spearman-Brown para a determinação do coeficiente de fidedignidade de um teste composto por k componentes paralelas.

Teorema – Se $Y_1, Y_2, \ldots, Y_k$ são k medidas paralelas e $X = \sum_{i=1}^{k} Y_i$ então

$$\rho(X, X') = \frac{k\,\rho(Y, Y')}{1 + (k-1)\,\rho(Y, Y')} \tag{3.1}$$

onde $\rho(Y, Y') = \rho(Y_i, Y_j),\ \forall\ i \neq j$.

Demonstração

Lord-Novick (1968) provaram que

$$\rho^2(X, T) \geqslant \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k} \sigma^2(Y_i)}{\sigma^2(X)}\right]$$

ocorrendo a igualdade quando as medidas são paralelas.

Da equação (2.11),

$$\sigma^2(Y_1) = \sigma^2(Y_2) = \ldots = \sigma^2(Y_k).$$

A partir desse resultado, e da extensão da equação (2.15) para o caso de k medidas paralelas, segue-se que

$$\sigma^2 (X) = k \ \sigma^2 (Y) [1 + (k-1) \rho (Y, Y')]$$

Substituindo estes dois últimos resultados na expressão dada por Lord-Novick, obtemos o membro da direita da equação (3.1).

A partir da definição do coeficiente de fidedignidade, temos que ele é um parâmetro desconhecido e, portanto, a sua determinação fica reduzida à estimação desse parâmetro, que implica na necessidade de medidas repetidas de uma amostra de pessoas. Estas múltiplas medidas podem ser obtidas de duas maneiras: (a) usando basicamente o mesmo teste ou (b) usando partes comparáveis do teste.

Na segunda situação, a fidedignidade de um teste é obtida a partir da fidedignidade das partes que o compõem, enquanto que na primeira situação sua obtenção é direta. Nessas condições, somos levados a concluir que a situação (b) exige uma teoria estatística mais complexa do que a situação (a). No entanto, o oposto é verdadeiro. Poucos trabalhos com ênfase nessa situação têm sido realizados, enquanto que na situação (b) os artigos são numerosos.

Na estimação desse coeficiente temos que, além de considerar estas duas situações, supor que os escores dos itens ou dos subtestes possuem determinada distribuição de probabilidade ou, então, não considerar esta suposição.

Conforme já nos referimos na introdução, analisar somente o valor numérico (estimativa pontual) do coeficiente de fidedignidade é pouco significativo para considerar um teste como um instrumento de medida, uma vez que ele é influenciado fortemente por fatores relativos ao teste e aos respondentes. A solução que propomos considera estes fatores e desenvolve-se usando a situação (b), bem como a distribuição de probabilidade dos escores dos itens ou subtestes, e permite, sob condições que serão apresentadas a seguir, determinar intervalos de confiança e teste de hipóteses para o coeficiente de fidedignidade.

Estimador de Máxima Verossimilhança

Suposições

Consideremos um teste dividido em k partes ou subtestes, $k \geqslant 2$, com mesma estrutura fatorial, tal que os escores destas partes tem distribuição Normal Multivariada, com iguais variâncias e iguais covariâncias. Suponhamos, além disso, que tomamos uma amostra aleatória de tamanho n.

Parâmetros básicos

Seja $\underset{\sim}{y} = (y_1, y_2, \ldots, y_k)$ o vetor aleatório dos escores das k partes de um teste com distribuição Normal Multivariada, média $\underset{\sim}{\mu} (\mu_1, \mu_2, \ldots, \mu_k)$ e matriz de variância-covariância R definida por

$$R = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho \\ \sigma^2\rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma^2\rho \\ \sigma^2\rho & \cdots & \sigma^2\rho & \sigma^2 \end{bmatrix}$$

σ^2: variância de cada parte,
ρ: correlação entre duas partes quaisquer.

Então, a função densidade de $\underset{\sim}{y}$ é dada por

$$f(\underset{\sim}{y}) = (2\pi)^{-k/2} \, |R|^{-1/2} \exp\left[-1/2(\underset{\sim}{y} - \underset{\sim}{\mu}) R^{-1} (\underset{\sim}{y} - \underset{\sim}{\mu})^{+}\right]$$

Além disso, supondo que dispomos dos escores dos n indivíduos da amostra nas k partes do teste, isto é,

$Y_{(i)} = (Y_{1i}, Y_{2i}, \ldots, Y_{Ki})$, $i = 1, 2, \ldots, n$, Kristof (1963) provou que o estimador de máxima verossimilhança para ρ é dado por

$$\hat{\rho}' = \frac{1}{k-1} \; \frac{\sum\sum_{j \neq \ell} S_{j\ell}}{\sum_j S_j^2}, \tag{3.2}$$

$$\text{onde } S_{j\ell} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_{ji} - \bar{Y}_j)(Y_{\ell_i} - \bar{Y}_\ell) \text{ e} \tag{3.3}$$

$$S_j^2 = \frac{1}{n-1} \sum_{i=1} (Y_{ji} - \bar{Y}_j)^2, \; j = 1, \ldots, k. \tag{3.4}$$

Usando a propriedade da invariância e a fórmula de Spearman-Brown (equação (3.1), decorrente da condição de igualdade das variâncias e das covariâncias), segue-se então que o estimador de máxima verossimilhança da fidedignidade de um teste formado por k partes é dado por

$$\hat{\rho} = k\,\hat{\rho}' / [1 + (k-1)\,\hat{\rho}'],$$

que é equivalente a

$$\hat{\rho} = \frac{k}{k-1} \left[1 - \frac{\sum_{j=1}^{k} S_j^2}{S_X^2}\right], \tag{3.5}$$

onde $S_X^2 = \sum_{j=1}^{k} S_j^2 + \sum_{j \neq \ell} \sum S_{j\ell}$. (3.6)

De (2.14) temos que S_X^2 é a variância dos escores totais $\sum_{j=1}^{k} Y_{ij}$, para $i = 1,2,\ldots,n$.

De (3.5) segue-se que o estimador de máxima verossimilhança da fidedignidade coincide com o coeficiente de fidedignidade α de Cronbach (1951).

Distribuição Amostral de $\hat{\rho}$

Devido à complexidade matemática envolvida na derivação da distribuição amostral de $\hat{\rho}$, apresentaremos a seguir somente os principais resultados.

De Kristof (1963), as quantidades

$$X^2_{n-1} = \frac{n-1}{k} \left[\frac{S_X^2}{\sigma^2 \left[1 + (k-1)\rho \right]} \right], \tag{3.7}$$

onde S_X^2 é definido por (3.6), tem distribuição X^2 com n-1 graus de liberdade, e

$$X^2_{(k-1)(n-1)} = \frac{n-1}{\sigma^2 (1-\rho)} \left(S_1^2 + \ldots + S_k^2 \; \frac{1}{k} \; S_X^2 \right), \tag{3.8}$$

onde S_j^2 é definido por (3.4), tem distribuição X^2 com (k - 1) (n - 1) graus de liberdade.

Da independência entre (3.7) e (3.8), segue-se que a razão

$$F_{(n-1),(k-1)(n-1)} = \frac{(k-1)\,(n-1)}{n-1} \left[\frac{X^2{}_{n-1}}{X^2{}_{(k-1)(n-1)}} \right]$$

tem distribuição F com (n - 1) e (k - 1) (n - 1) graus de liberdade, que é equivalente a

$$F_{(n-1),(k-1)(n-1)} = \frac{1-\rho}{1-\hat{\rho}} \tag{3.9}$$

onde $\hat{\rho}$ é o estimador de máxima verossimilhança de ρ, o coeficiente de fidedignidade do teste.

Teste de Hipótese

Suponhamos que desejamos testar a hipótese $H_0: \rho \leqslant \rho_0$ versus $H_1: \rho > \rho_0$, com nível de significância α, onde ρ_0 é uma constante.

Da teoria de teste de hipótese temos que a região crítica do teste proposto é definida por

$$C = \{ A \in R^k X R^n \;/\; \hat{\rho} > \rho_c \},$$

onde A é a matriz dos dados amostrais e ρ_c é determinado pela condição:
$P(\hat{\rho} > \rho_c / H_0 \text{ é verdadeiro}) = \alpha$

Supondo que H_0 é verdadeiro, obtemos

$$P(\hat{\rho} > \rho_c) = P\left(\frac{1 - \hat{\rho}}{1 - \rho_0} < \frac{1 - \rho_c}{1 - \rho_0}\right) = \alpha.$$

A partir de (3.9) temos que, quando $\rho = \rho_0$,

$$F_{(k-1)(n-1),(n-1)} = \frac{1 - \hat{\rho}}{1 - \rho_0}$$

tem uma distribuição F com $(k - 1)(n - 1)$ e $(n - 1)$ graus de liberdade.

Segue-se então que a quantidade $(1 - \rho_c) / (1 - \rho_0)$ é igual ao correspondente valor tabelado da distribuição $F_{(k-1)(n-1),(n-1);\alpha}$, isto é, que $\rho_c = 1 - (1 - \rho_0) F_{(k-1)(n-1),(n-1);\alpha}$.

Portanto, rejeitamos a hipótese de que o coeficiente de fidedignidade ρ é menor ou igual a ρ_0 com nível de significância α se

$$\hat{\rho} > 1 - (1 - \rho_0) F_{(k-1)(n-1),(n-1);\alpha} \qquad (3.10)$$

onde $\hat{\rho}$ é o coeficiente α de Cronbach.

Intervalo de confiança

Baseados na dualidade intrínseca entre teste de hipótese e intervalos de confiança, os resultados obtidos anteriormente podem ser usados para produzir intervalos de confiança para o coeficiente de fidedignidade, com qualquer coeficiente da confiança. Vamos determinar um dos possíveis intervalos de confiança.

A partir de (3.9), obtemos

$$P\left(F_{(n-1),(k-1)(n-1);\frac{\alpha}{2}} < \frac{1 - \rho}{1 - \hat{\rho}} < F_{(n-1),(k-1)(n-1);1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Após alguns cálculos, segue-se que um intervalo de confiança para o coeficiente de fidedignidade com nível de confiança $1 - \alpha$ é dado por

$$\left(1 - (1 - \hat{\rho}) F_{(n-1),(k-1)(n-1);1-\frac{\alpha}{2}},\ 1 - (1 - \hat{\rho}) F_{(n-1),(k-1)(n-1);\frac{\alpha}{2}}\right) \qquad (3.11)$$

onde $\hat{\rho}$ é o coeficiente α de Cronbach.

4 – REFERÊNCIAS BIBLIOGRÁFICAS

ALBANESE, M.T. (1984) – Coeficiente de Fidedignidade de um Instrumento de Medida – UFRGS, Porto Alegre. (em fase de publicação).

CRONBACH, L. J. (1951) – Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3): 297-334.

KRISTOF, W. (1963). The statistical theory of stepped-up reability coefficients when a test has been divided into several equivalent parts. *Psychometrika, 28* (3): 221-38.

LORD, F.M. & NOVICK, M.R. *(1968) – Statistical theories of mental test scores.* Reading, Massachussets: Addison-Wesley.