Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Cadernos de Matemática e Estatística
Série B: Trabalho de Apoio Didático

Coeficiente de Fidedignidade de um Instrumento de Medida

Maria Teresinha Albanese

Série B, nº 12,
Porto Alegre, outubro de 1992

COEFICIENTE DE FIDEDIGNIDADE DE UM INSTRUMENTO DE MEDIDA

Maria Teresinha Albanese

1992

ÍNDICE

P R E F Á C I O

Este texto reune os principais tópicos sobre o coeficiente de fidedignidade de um instrumento de medida, que encontram-se dispersos na literatura, tratando tanto dos aspectos teóricos, quanto práticos.

Nos capítulos 1, 2 e 3 encontra-se desenvolvida a fundamentação teórica dos métodos de estimação do coeficiente de fidedignidade que são apresentados no capítulo 4. Neste são considerados dois métodos: supondo que a distribuição dos escores dos indivíduos no instrumento é desconhecida, teste-reteste, formas paralelas, divisão em metades, coeficiente de Kuder-Richardon e α de Cronbach) e supondo distribuição Normal Multivariada (estimador de máxima verossimilhança). São realizadas várias comparações entre os diferentes estimadores, segundo alguns critérios. O primeiro método é desenvolvido de tal forma que se o leitor possui somente interesse nos aspectos práticos dos estimadores do coeficiente de fidedignidade, ele tem condições plenas de atingir este objetivo, sem consultar os três primeiros capítulos.

O capítulo 5 apresenta aplicações dos conteúdos desenvolvidos nos capítulos anteriores, em duas áreas de pesquisa.

Portanto a apresentação e desenvolvimento deste texto permite que ele seja útil, tanto para os leitores interessados somente na caracterização e obtenção dos principais estimadores do coeficiente de fidedignidade, quanto àqueles que buscam, além disto, a fundamentação teórica, apoiados em conhecimentos básicos da Teoria das Probabilidades e Inferência Estatística.

Porto Alegre, janeiro de 1984.

Maria Teresinha Albanese

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.

A situação, no entanto, é diferente quando queremos' medir variáveis psicológicas, como por exemplo, neurose, nível de satisfação, capacidade de pensamento lógico, em que nos deparamos com problemas de escalonamento muito complexos.

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

A validade se refere 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 aos mesmos sujeitos, obteremos os mesmos resultados.

Validade e fidedignidade são conceitos fortemente relacionados. Um coeficiente de fidedignidade alto é uma condição necessária, mas não suficiente, para um coeficiente de validade também alto.

Uma vez que não dispomos de um tempo ilimitado para aplicação de um teste, devemos usá-lo eficientemente para obter uma medida tão válida e fidedigna quanto possível, com o menor número de itens. Podemos atingir este objetivo através de uma análise de itens, baseada na contribuição de cada item para o aumento da validade e da fidedignidade.

Além destes coeficientes, na construção de um instrumento de medida, devemos considerar os seguintes aspectos práticos: É conveniente usá-lo? É econômico e prático aplicá-lo e interpretá-lo? Qual é o seu tempo de aplicação?

Neste trabalho vamos tratar somente do coeficiente de fidedignidade de um instrumento de medida, sob a suposição que a sua validade já foi constatada.

O coeficiente de fidedignidade é um parâmetro desconhecido, e sua estimação implica na necessidade de medidas repetidas de uma amostra de sujeitos. Estas múltiplas medidas podem ser obtidas de 2 maneiras: (a) usando basicamente o mesmo teste ou (b) usando partes comparáveis do mesmo 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. Nestas condições, podemos ser 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 nesta situação tem sido realizados, enquanto que na situação (b) os artigos são numerosos.

Na situação (a), o coeficiente de fidedignidade pode ser determinado pelo método do teste-reteste ou das formas paralelas.

Na situação (b), uma reformulação radical foi proposta por KUDER-RICHARDSON (1937). Eles propuseram várias formas alternativas quando um teste é composto por itens dicotômicos, e que foram amplamente adotadas, em especial, as fórmulas KR_{20} e KR_{21}. Um novo impulso foi dado por CRONBACH (1951) quando surgiu o coeficiente de fidedignidade α de Cronbach, uma generalização da fórmula KR_{20}. Ainda nesta situação, a fidedignidade pode ser determinada pelo método das metades.

A fidedignidade é afetada em diferentes graus por fatores relativos ao teste (número de itens, amplitude de dificuldade e interdependência dos itens, etc.), e por fatores relativos ao examinando (velocidade na realização do teste, motivação, cansaço, etc.). Estes fatores devem ser considerados na interpretação da fidedignidade.

Dispor somente do valor numérico do coeficiente de fidedignidade é pouco significativo para descrever um teste como um instrumento de medida, pois para um mesmo teste e mesmo método podemos obter diferentes graus de fidedignidade, conforme a heterogeneidade dos grupos examinados.

Entretanto este problema pode ser resolvido, se pudermos supor que os escores dos itens possuem distribuição Normal Multivariada, através da construção de intervalos de confiança e realização de testes de hipóteses para o coeficiente de fidedignidade populacional.

Provavelmente o primeiro artigo com ênfase na estimação da fidedignidade, baseado na distribuição de amostras pequenas, foi escrito por KRISTOF (1963). Outros artigos que se destacam na área de estimação e testes paramétricos são devido a KRISTOF (1970, 1972, 1974).

O grau mínimo exigido para a fidedignidade deve ser determinado de acordo com os objetivos e situações em que nos propomos usar o instrumento de medida. A sua magnitude deve corresponder a importância das decisões e dos efeitos que estas decisões venham ter sobre a população alvo.

Esse trabalho encontra-se dividido da seguinte forma:

Capítulo I - Modelo clássico linear para testes de tamanho fixo;

Capítulo II - Testes compostos;

Capítulo III - Fatores que afetam o coeficiente de fidedignidade de um teste;

Capítulo IV - Métodos de estimação do coeficiente de fidedignidade de um teste;

Capítulo V - Aplicações.

CAPÍTULO I

MODELO CLÁSSICO LINEAR PARA TESTES DE TAMANHO FIXO

Neste capítulo apresentaremos as suposições básicas do modelo clássico linear de testes com tamanho fixo. A partir destas suposições, deduziremos relações entre várias quantidades de interesse, que serão fundamentais para o desenvolvimento e compreensão de todo trabalho. Além disto, definiremos e interpretaremos estatisticamente os coeficientes de fidedignidade e de validade.

1.1 - *Suposições do Modelo Clássico*

Seja X uma variável aleatória definida sobre uma população $\mathcal{P}$ 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

$$X = T+\varepsilon \qquad (1.1.1)$$

sobre $\mathcal{P}$.

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

Sejam as variâncias de X, T e ε denotadas por σ_X^2, σ_T^2 e σ_ε^2, respectivamente. Denotemos as correlações entre X e ε, T e ε, e ε_1 e ε_2, onde ε_1 e ε_2 são os erros associados a duas medidas distintas, por $\rho_{X\varepsilon}$, $\rho_{T\varepsilon}$ e $\rho(\varepsilon_1,\varepsilon_2)$, respectivamente; e as correspondentes covariâncias por $\sigma_{X\varepsilon}$, $\sigma_{T\varepsilon}$ e $\sigma(\varepsilon_1,\varepsilon_2)$.

As suposições do modelo clássico linear incluem a equação (1.1.1) e as apresentadas a seguir:

$$E(\varepsilon) = 0, \quad (1.1.2)$$

$$\rho_{T\varepsilon} = 0, \quad (1.1.3)$$

$$\rho(\varepsilon_1,\varepsilon_2) = 0, \quad (1.1.4)$$

$$\rho(\varepsilon_1,T_2) = 0, \quad (1.1.5)$$

que são válidas sob qualquer subpopulação não nula de $\mathcal{P}$.

1.2 - *Conseqüências Imediatas das Suposições*

A seguir apresentaremos algumas conseqüências imediatas das suposições do modelo clássico linear para testes de tamanho fixo.

As variáveis aleatórias, escore verdadeiro e escore observado, possuem valores esperados iguais, isto é,

$$E(X) = E(T+\varepsilon) = E(T).$$

Como $E(\varepsilon) = 0$ em qualquer subpopulação não nula de $\mathcal{P}$, em particular, a esperança é nula na subpopulação de pessoas com qualquer escore verdadeiro específico τ. Segue-se então que

$$E(X/\tau) = \tau.$$

A variância de X, σ_X^2, é igual a $\sigma_T^2+\sigma_\varepsilon^2+2\sigma_{T\varepsilon}$. Da equação (1.1.3), $\sigma_{T\varepsilon} = \rho_{T\varepsilon}\sigma_\varepsilon\sigma_T = 0$. Portanto

$$\sigma_X^2 = \sigma_T^2+\sigma_\varepsilon^2 \qquad (1.2.1)$$

Por definição, $\sigma_{XT} = \sigma(T+\varepsilon,T) = \sigma_T^2+\sigma_{T\varepsilon}$. Como $\sigma_{T\varepsilon} = 0$, segue-se que

$$\sigma_{XT} = \sigma_T^2.$$

Por definição, $\rho_{XT}^2 = \dfrac{\sigma_{XT}^2}{\sigma_X^2\,\sigma_T^2}$.

Segue-se então que

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2}. \qquad (1.2.2)$$

A partir das equações (1.2.1) e (1.2.2) podemos escrever também

$$\rho_{XT}^2 = 1-\frac{\sigma_\varepsilon^2}{\sigma_X^2}. \qquad (1.2.3)$$

Analogamente a obtenção de (1.2.2), temos que $\rho_{X\varepsilon}^2 = \dfrac{\sigma_\varepsilon^2}{\sigma_X^2}$, a partir do qual vem a relação

$$\rho_{X\varepsilon}^2+\rho_{XT}^2 = 1.$$

A covariância entre os escores observados X_1 e X_2 é igual a covariância entre seus escores verdadeiros T_1 e T_2, isto é,

$$\begin{aligned}\sigma(X_1,X_2) &= \sigma[(T_1+\varepsilon_1),(T_2+\varepsilon_2)]\\ &= \sigma(T_1,T_2)+\sigma(T_1,\varepsilon_2)+\sigma(\varepsilon_1,T_2)+\sigma(\varepsilon_1,\varepsilon_2)\\ &= \sigma(T_1,T_2), \qquad (1.2.4)\end{aligned}$$

usando as equações (1.1.4) e (1.1.5).

1.3 - *Relações Baseadas em Medidas Paralelas*

1.3.1 - *Definição* - Sejam as variáveis aleatórias X e X' tal que $X = T+\varepsilon$, $X' = T+\varepsilon'$ e $\sigma^2_{\varepsilon} = \sigma^2_{\varepsilon'}$, em qualquer subpopulação não nula de $\mathcal{P}$.

Então X e X' são denominadas medidas paralelas.

Se X e X' são medidas paralelas, segue-se imediatamente da definição e das suposições do modelo clássico linear que

$$E(X) = E(X') \text{ e} \tag{1.3.1}$$

$$\sigma^2(X) = \sigma^2(X'), \tag{1.3.2}$$

em qualquer subpopulação não nula de $\mathcal{P}$.

A partir da equação (1.2.4), obtemos que a correlação entre qualquer par de medidas paralelas X e X' é

$$\sigma(X,X') = \sigma^2_T. \tag{1.3.3}$$

Das equações (1.2.4) e (1.3.2), segue-se então que

$$\rho_{XX'} = \frac{\sigma^2_T}{\sigma^2_X} \geq 0. \tag{1.3.4}$$

Comparando as equações (1.2.2) e (1.3.4) obtemos

$$\rho^2_{XT} = \rho_{XX'} \tag{1.3.5}$$

Logo supondo que no mínimo um par de medidas paralelas pode ser obtido, podemos expressar uma quantidade não observável ρ^2_{XT}, em termos de $\rho_{XX'}$ um parâmetro da distribuição bivariada do escore observado. Assim, a estimação de ρ^2_{XT} fica reduzida a estimação de $\rho_{XX'}$.

A partir de (1.3.4), temos

$$\sigma^2_T = \sigma^2_X \, \rho_{XX'}, \tag{1.3.6}$$

que expressa a variância do escore verdadeiro, uma quantidade não observável, como o produto de duas quantidades potencialmente observáveis, σ^2_X e $\rho_{XX'}$.

Combinando as equações (1.2.1) e (1.3.6) obtemos

$$\sigma_X^2 = \sigma_X^2\rho_{XX'}+\sigma_\varepsilon^2 \qquad \text{ou}$$

$$\sigma_\varepsilon^2 = \sigma_X^2(1-\rho_{XX'}), \qquad (1.3.7)$$

que novamente expressa uma quantidade não observável, a variância do erro, em termos de um produto de quantidades potencialmente observáveis.

A equação (1.3.7) pode ser escrita como

$$\sigma_\varepsilon = \sigma_X \sqrt{1-\rho_{XX'}},$$

e é denominada erro padrão de medida.

Finalmente a partir de (1.3.6), usando (1.2.1) e (1.3.4), obtemos que a correlação entre o escore observado e o erro é dada por

$$\rho_{X\varepsilon} = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_X^2}} = \sqrt{1-\frac{\sigma_T^2}{\sigma_X^2}} = \sqrt{1-\rho_{XX'}}$$

1.4 - Coeficiente de Fidedignidade

1.4.1 - *Definição* - O coeficiente de fidedignidade de um teste é definido como o quadrado da correlação entre o escore observado e o verdadeiro, isto é, ρ_{XT}^2.

Da relação $\rho_{XT}^2 = \sigma_T^2/\sigma_X^2$, equação (1.2.2), segue-se que a fidedignidade de um teste é uma medida do grau de variação do escore verdadeiro em relação ao escore observado, ou ainda, é a proporção da variância do escore observado que é devido a variância do escore verdadeiro.

Considerando a equação (1.2.3), a fidedignidade pode também ser expressada por

$$\rho_{XT}^2 = 1-\frac{\sigma_\varepsilon^2}{\sigma_X^2}.$$

Como $\sigma_X^2 \geq \sigma_\varepsilon^2$, a fidedignidade assume valores no intervalo [0,1]. Quanto maior a variância do erro, menor é a fidedignidade do teste. Quando a variância do erro se aproxima de zero, a fidedignidade se aproxima de um, assumindo este valor, somente quando σ_ε^2 é igual a zero.

Da igualdade (1.3.5), segue-se que quando dispomos de duas medidas paralelas X e X', a fidedignidade pode ser definida também como o coeficiente de correlação entre X e X', isto é, $\rho_{XX'}$.

1.5 - *Coeficiente de Validade*

Uma das medidas mais significativas para o valor de um teste é sua validade. Um teste é válido quando ele mede o que realmente se propõe medir.

De nada adianta um teste ser fidedigno, se ele não for válido.

Estatisticamente, a validade de um teste é verificada através do grau de relação linear do teste com uma variável de interesse.

1.5.1 - *Definição* - O coeficiente de validade de uma medida X em relação a outra medida Y é definido como o valor absoluto do coeficiente de correlação ρ_{XY}, que é dado por

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

É evidente que esta definição é simétrica. O coeficiente de validade de uma medida somente pode ser determinado em relação a uma segunda medida, portanto o coeficiente de validade é uma característica de um par de medidas.

LORD-NOVICK (1968, pág. 72) provaram que

$$\rho_{XY} \leqq \rho_{XT}.$$

Isto é, a validade de um teste em relação a qualquer outro teste não pode exceder a raiz quadrada da fidedignidade, denominada índice de fidedignidade. Portanto, um índice de fidedignidade alto é uma condição necessária, mas não suficiente, para um teste ter uma validade alta em relação a qualquer outro teste.

No entanto, a validade de um teste em relação a qualquer outro teste pode ser superior a sua fidedignidade, o que pode não parecer intuitivo. O paradoxo, entretanto, resulta de características não paralelas das definições do coeficiente de fidedignidade e de validade. Na prática, isto pode ocorrer quando o teste tem fidedignidade alta, mas na obtenção da validade ele é comparado com um teste com fidedignidade baixa.

Apresentamos a definição do coeficiente de validade porque ela é uma medida que deve ser considerada na avaliação de um teste como um instrumento de medida, e também, porque este coeficiente está relacionado com o coeficiente de fidedignidade, conforme mencionamos acima. Entretanto, neste trabalho, trataremos somente da fidedignidade de um teste, supondo que a sua validade já foi comprovada.

CAPÍTULO II

TESTES COMPOSTOS

2.1 - *Introdução*

Geralmente um teste é composto por um conjunto de itens, e freqüentemente, estes itens estão agrupados em subtestes. O escore que uma pessoa recebe num teste, resulta então da soma ou média(ponderada ou não) dos escores obtidos nos itens ou nos subtestes.

Neste capítulo trataremos da teoria de modelos de testes compostos, cujo escore final resulta da soma dos escores obtidos nos itens ou nos subtestes. Posteriormente aplicaremos a teoria desenvolvida nos dois primeiros capítulos.

A teoria de testes apresentada no capítulo anterior, pode ser extendida facilmente para o modelo de testes compostos. Precisamos somente considerar que cada item, subteste ou teste gera uma medida, que a medida do subteste é determinada de forma aditiva a partir das medidas dos itens, e que a medida do teste é determinada de forma semelhante a partir dos subtestes. Nestas condições, dizemos que a medida do teste é uma medida composta e que suas partes são componentes de medida.

As características de um teste são determinadas pelas características de suas componentes de medida.

2.2 - *Medidas Compostas por Duas Componentes*

Sejam as medidas $(Y_1, T_1, \varepsilon_1)$ e $(Y_2, T_2, \varepsilon_2)$ assumindo os valores (y_1, τ_1, e_1) e (y_2, τ_2, e_2), respectivamente. Seja (X, T, ε) uma medida composta, tomando o valor (x, τ, e), e definida por

$$X = Y_1 + Y_2. \qquad (2.2.1)$$

Podemos definir então

$$T = T_1 + T_2 \quad \text{e} \quad \varepsilon = \varepsilon_1 + \varepsilon_2, \qquad (2.2.2)$$

onde T e ε são o escore verdadeiro e o erro correspondentes ao escore observado X.

O valor esperado de X é dado por

$$E(X) = E(Y_1) + E(Y_2). \qquad (2.2.3)$$

As variâncias do escore observado X, do escore verdadeiro T e do erro ε são dadas, respectivamente, por

$$\sigma_X^2 = \sigma^2(Y_1+Y_2) = \sigma^2(Y_1) + \sigma^2(Y_2) + 2\rho(Y_1, Y_2)\sigma(Y_1)\sigma(Y_2), \qquad (2.2.4)$$

$$\sigma_T^2 = \sigma^2(T_1+T_2) = \sigma^2(T_1) + \sigma^2(T_2) + 2\rho(T_1, T_2)\sigma(T_1)\sigma(T_2) \text{ e} \qquad (2.2.5)$$

$$\sigma_\varepsilon^2 = \sigma^2(\varepsilon_1+\varepsilon_2) = \sigma^2(\varepsilon_1) + \sigma^2(\varepsilon_2). \qquad (2.2.6)$$

Se Y e Y' são medidas paralelas então as expressões para as variâncias de X, T e ε assumem as formas simples

$$\sigma_X^2 = 2\sigma^2(Y)(1+\rho_{YY'}), \qquad (2.2.7)$$

$$\sigma_T^2 = 4\sigma^2(T_1) \qquad \text{e} \qquad (2.2.8)$$

$$\sigma_\varepsilon^2 = 2\sigma^2(\varepsilon_1). \qquad (2.2.9)$$

A demonstração de (2.2.9) é trivial, pois $\sigma^2(\varepsilon_1) = \sigma^2(\varepsilon_2)$, enquanto que (2.2.8) segue do fato que $T_1 \equiv T_2$, e assim, $\sigma(T_1, T_2) = \sigma^2(T_1)$. A prova de (2.2.7) segue da igualdade $\sigma^2(Y_1) = \sigma^2(Y_2)$.

2.2.1 - *Teorema* - Sejam as medidas paralelas Y_1, Y_2, Y_3, Y_4, e $X = Y_1+Y_2$, $X' = Y_3+Y_4$. Então a fidedignidade de X é dada por

$$\rho_{XX'} = \rho(Y_1+Y_2, Y_3+Y_4) = \frac{2\rho_{YY'}}{1+\rho_{YY'}} \tag{2.2.10}$$

onde

$$\rho_{YY'} = \rho(Y_1, Y_2).$$

Demonstração - A prova que X e X' são medidas paralelas é trivial. Vamos determinar a correlação entre X e X'.

$$\rho(Y_1+Y_2, Y_3+Y_4) = \frac{\sigma(Y_1+Y_2, Y_3+Y_4)}{\sigma(Y_1+Y_2)\sigma(Y_3+Y_4)}$$

$$= \frac{\sigma(Y_1,Y_3)+\sigma(Y_2,Y_3)+\sigma(Y_1,Y_4)+\sigma(Y_2,Y_4)}{\{[\sigma^2(Y_1)+\sigma^2(Y_2)+2\sigma(Y_1,Y_2)][\sigma^2(Y_3)+\sigma^2(Y_4)+2\sigma(Y_3,Y_4)]\}^{1/2}}$$

É fácil mostrar que como Y_1, Y_2, Y_3 e Y_4 são medidas paralelas, então as variâncias são todas idênticas, assim como todas as correlações entre estas medidas. Podemos então escrever $\rho(Y_1+Y_2, Y_3+Y_4)$ como a igualdade

$$\rho(Y_1+Y_2, Y_3+Y_4) = \frac{4\sigma^2(Y_1)\rho(Y_1,Y_2)}{2\sigma(Y_1)+2\rho(Y_1,Y_2)\sigma^2(Y_1)}$$

$$= \frac{2\rho(Y_1,Y_2)}{1+\rho(Y_1,Y_2)}$$

$$= \frac{2\rho_{YY'}}{1+\rho_{YY'}}, \text{ onde } \rho_{YY'} = \rho(Y_1,Y_2).$$

A equação (2.2.10) corresponde a fidedignidade $\rho_{XX'}$ de um teste composto por duas componentes paralelas, cada qual tendo fidedignidade $\rho_{YY'}$. Este resultado é conhecido como a fórmula de Spearman-Brown para um teste de duplo tamanho.

Se $0 < \rho_{YY'} < 1$ então $[2/(1+\rho_{YY'})] > 1$, e assim, $\rho_{XX'} > \rho_{YY'}$. Isto é, a fidedignidade de um teste composto por componentes paralelas é maior que a fidedignidade de suas componentes.

2.3 - *Medidas Compostas por* **k** *Componentes*

Como uma extensão natural da secção anterior, considere as medidas $(Y_i, T_i, \varepsilon_i)$, assumindo os valores (y_i, τ_i, e_i), para $i=1,2,\ldots,n$. Seja (X,T,ε) uma medida composta, assumindo o valor (x,τ,e), e definida por

$$X = \sum_{i=1}^{k} Y_i. \tag{2.3.1}$$

Então podemos dizer que

$$T = \sum_{i=1}^{k} T_i, \tag{2.3.2}$$

$$\varepsilon = \sum_{i=1}^{k} \varepsilon_i. \tag{2.3.3}$$

As médias de X, T e ε são dadas, respectivamente, por

$$E(X) = E(\sum_{i=1}^{k} Y_i) = \sum_{i=1}^{k} E(Y_i), \tag{2.3.4}$$

$$E(T) = E(\sum_{i=1}^{k} T_i) = \sum_{i=1}^{k} E(T_i) = \sum_{i=1}^{k} E(Y_i), \tag{2.3.5}$$

$$E(\varepsilon) = E(\sum_{i=1}^{k} \varepsilon_i) = \sum_{i=1}^{k} E(\varepsilon_i) = 0. \tag{2.3.6}$$

Da equação (2.3.1), obtemos ainda que as variâncias do escore composto observado, verdadeiro e erro, respectivamente, são iguais a

$$\sigma_X^2 = \sigma^2\left(\sum_{i=1}^{k} Y_i\right) = \sum_{i=1}^{k} \sigma^2(Y_i) + \sum_{i\neq j}\sum \rho(Y_i, Y_j)\sigma(Y_i)\sigma(Y_j), \qquad (2.3.7)$$

$$\sigma_T^2 = \sigma^2\left(\sum_{i=1}^{k} T_i\right) = \sum_{i=1}^{k} \sigma^2(T_i) + \sum_{i\neq j}\sum \rho(T_i, T_j)\sigma(T_i)\sigma(T_j) \quad \text{e} \qquad (2.3.8)$$

$$\sigma_\varepsilon^2 = \sum_{i=1}^{k} \sigma^2(\varepsilon_i) + \sum_{i\neq j}\sum \rho(\varepsilon_i, \varepsilon_j)\sigma(\varepsilon_i)\sigma(\varepsilon_j) = \sum_{i=1}^{k} \sigma^2(\varepsilon_i), \qquad (2.3.9)$$

pois $\rho(\varepsilon_i, \varepsilon_j) = 0$, para qualquer $i \neq j$.

Se as k medidas são paralelas, então as expressões das variâncias de X,T e E adquirem as formas simplificadas:

$$\sigma_X^2 = k\sigma_Y^2[1+(k-1)\rho_{YY'}], \qquad (2.3.10)$$

$$\sigma_T^2 = k^2\sigma^2(T_1), \qquad (2.3.11)$$

$$\sigma_\varepsilon^2 = k\,\sigma^2(\varepsilon_1). \qquad (2.3.12)$$

A equação (2.3.11) mostra que se o tamanho do teste é aumentado k vezes, adicionando medidas paralelas, então a variância do escore verdadeiro composto aumenta k^2 vezes, enquanto que em situação idêntica, a equação (2.3.12) mostra que a variância do erro composto aumenta somente k vezes. O fato da variância do escore verdadeiro aumentar mais rapidamente do que a variância do erro torna vantajoso aumentar o tamanho do teste, pois isto produz um aumento da fidedignidade do teste.

2.4 - *O Coeficiente α de Cronbach e a Fidedignidade de Medidas Compostas*

Derivação do Coeficiente α de Cronbach como um Limite Inferior para a Fidedignidade de um Teste

2.4.1 - *Teorema* - Sejam $Y_1, Y_2, \ldots, Y_k$, medidas com escores verdadeiros $T_1, T_2, \ldots, T_k$, respectivamente.

Se $X = \sum_{i=1}^{k} Y_i$ então

$$\rho^2_{XT} \geqq \frac{k}{k-1}\left[1-\frac{\sum_{i=1}^{k}\sigma^2(Y_i)}{\sigma^2_X}\right]. \tag{2.4.1}$$

Demonstração - A desigualdade (2.4.1), na forma dada e para casos especiais, foi objeto de estudo para vários pesquisadores em artigos como de KUDER-RICHARDSON(1937), GUTTMAN(1945), CROÑBACH(1951), CURETON(1958), NOVICK-LEWIS(1967), LORD-NOVICK (1968), entre outros. A demonstração dada a seguir baseia-se na prova dada por LORD-NOVICK(1968).

A partir das suposições do modelo clássico linear (secção 1.1), temos que $\sigma(Y_i,Y_j) = \sigma(T_i,T_j)$. Como

$$[\sigma(T_i)-\sigma(T_j)]^2 \geqq 0$$

obtemos

$$\sigma^2(T_i)+\sigma^2(T_j) \geqq 2\sigma(T_i)\sigma(T_j).$$

Por outro lado da desigualdade de Cauchy-Schawrtz vem que

$$\sigma(T_i)\sigma(T_j) \geqq |\sigma(T_i,T_j)|.$$

Segue-se então que

$$\sigma^2(T_i)+\sigma^2(T_j) \geqq 2|\sigma(T_i,T_j)| \geqq 2\sigma(T_i,T_j).$$

Somando para $i \neq j$, obtemos

$$\sum_{i \neq j}\sum [\sigma^2(T_i)+\sigma^2(T_j)] \geqq 2\sum_{i \neq j}\sum \sigma(T_i,T_j). \qquad (2.4.2)$$

Sejam as identidades matemáticas

$$\sum_{i=1}^{k}\sum_{j=1}^{k} [\sigma^2(T_i)+\sigma^2(T_j)] = \sum_{i=1}^{k} [n\sigma^2(T_i)+\sum_{j=1}^{k}\sigma^2(T_j)]$$

$$= k\sum_{i=1}^{k}\sigma^2(T_i)+n\sum_{j=1}^{k}\sigma^2(T_j)$$

$$= 2k\sum_{i=1}^{k}\sigma^2(T_i)$$

e

$$\sum_{i=1}^{k}\sum_{j=1}^{k} [\sigma^2(T_i)+\sigma^2(T_j)] = \sum_{i=j}\sum [\sigma^2(T_i)+\sigma^2(T_j)]+\sum_{i\neq j}\sum [\sigma^2(T_i)+\sigma^2(T_j)]$$

$$= 2\sum_{i=1}^{k}\sigma^2(T_i)+\sum_{i\neq j}\sum [\sigma^2(T_i)+\sigma^2(T_j)].$$

De acordo com estas identidades matemáticas, segue-se que

$$\sum_{i\neq j}\sum [\sigma^2(T_i)+\sigma^2(T_j)] = 2k\sum_{i=1}^{k}\sigma^2(T_i)-2\sum_{i=1}^{n}\sigma^2(T_i)$$

$$= 2(k-1)\sum_{i=1}^{k}\sigma^2(T_i). \qquad (2.4.3)$$

Substituindo (2.4.3) em (2.4.2), obtemos

$$\sum_{i=1}^{k}\sigma^2(T_i) \geqq \frac{\sum_{i\neq j}\sum \sigma(T_i,T_j)}{K-1}. \qquad (2.4.4)$$

Usando (2.4.4) na variância de T, que é dada por

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

obtemos

$$\sigma^2(T) \geq \sum_{i\neq j}\sum \frac{\sigma(T_i,T_j)}{k-1} + \sum_{i\neq j}\sum \sigma(T_i,T_j),$$

e combinando os termos da direita, segue-se então que

$$\sigma^2(T) \geq \frac{k}{k-1} \sum_{i\neq j}\sum \sigma(T_i,T_j). \qquad (2.4.5)$$

Para completar a demonstração, note que

$$\sigma^2(X) - \sum_{i=1}^{k} \sigma^2(Y_i) = \sum_{i\neq j}\sum \sigma(Y_i,Y_j) = \sum_{i\neq j}\sum \sigma(T_i,T_j).$$

Substituindo este resultado na desigualdade (2.4.5) e dividindo ambos os lados da desigualdade por σ_X^2, obtemos

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} \geq \frac{k}{k-1}[1 - \sum_{i=1}^{k} \frac{\sigma^2(Y_i)}{\sigma_X^2}].$$

O lado direito de (2.4.1),

$$\alpha \equiv \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma^2(Y_i)}{\sigma_X^2}\right) \qquad (2.4.6)$$

é denominado coeficiente α de CRONBACH(1951). Portanto, *o coeficiente* α, uma quantidade que pode ser calculada a partir dos resultados da aplicação de um teste, constitui um *limite inferior para a fidedignidade de um teste.*

Condições sob as quais o coeficiente α de Cronbach é igual a Fidedignidade de Medidas Compostas

A condição necessária e suficiente para a igualdade na Desigualdade de Cauchy-Schwartz, neste contexto, é que T_i seja uma função linear de T_j. Na demonstração do teorema (2.4.1) as outras únicas desigualdades usadas foram $[\sigma(T_i)-\sigma(T_j)]^2 \geqq 0$ e $|\sigma(T_i,T_j)| \geqq \sigma(T_i,T_j)$. Dado que $T_i = a_{ij}+b_{ij}T_j$, onde a_{ij} e b_{ij} são constantes quaisquer, a primeira destas desigualdades torna-se uma igualdade se e somente se $b_{ij} = \pm 1$ e a segunda se e somente se $|b_{ij}| = b_{ij}$. Conjuntamente, isto implica que $b_{ij} = 1$. Portanto, podemos concluir que

2.4.2 - *Corolário* - A condição necessária e suficiente para (2.4.1) valer como igualdade, e por conseguinte, o coeficiente α de Cronbach ser igual a fidedignidade de medidas compostas, é que $T_i = a_{ij}+T_j$, para qualquer i,j. Quando esta condição se verifica, dizemos que todas as componentes de medida são essencialmente τ-equivalentes.

A condição τ-equivalência essencial requer que o escore verdadeiro de qualquer respondente para uma dada componente de medida difere de seu escore verdadeiro em qualquer outra componente de medida do teste por somente uma constante. Isto implica que todas as covariâncias entre pares de componentes sejam iguais.

2.4.3 - *Corolário* - Se $Y_1, Y_2, \ldots, Y_k$, são k medidas paralelas, então a generalização da equação (2.2.10) é

$$\rho_{XX'} = \frac{k\rho_{YY'}}{1+(k-1)\rho_{YY'}} \qquad (2.4.7)$$

Demonstração - Por hipótese $Y_1, Y_2, \ldots, Y_K$ são medidas paralelas, implicando na igualdade das variâncias, isto é,

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

Utilizando este resultado e a equação (2.3.10). obtemos o membro da direita de (2.4.1), na forma dada em (2.4.7).

A igualdade se verifica, uma vez que medidas paralelas são essencialmente τ-equivalentes.

Se as k componentes de medida de um teste são paralelas, então a forma de coeficiente α de Cronbach, dada por (2.4.6) coincide com o resultado do corolário (2.4.3), que é conhecido como a fórmula de Spearman-Brown para a determinação do coeficiente de fidedignidade de um teste composto por K componentes paralelas.

Originalmente KUDER-RICHARDSON(1937) discutiram casos especiais deste coeficiente, isto é, situações em que as componentes são variáveis aleatórias binárias, assumindo valores zero e um, com probabilidades $1-p_i$ e p_i, respectivamente. Sob estas condições, o coeficiente α se reduz a Fórmula 20 de Kuder-Richardson:

$$KR_{20} = \frac{k}{k-1}\left(1 - \sum_{i=1}^{k} \frac{p_i q_i}{\sigma_X^2}\right) \qquad (2.4.8)$$

Se os itens tem valores p_i's idênticos, então (2.4.8) reduz-se a Fórmula 21 de Kuder-Richardson:

$$KR_{21} = \frac{k}{k-1}\left(1 - \frac{k\,\overline{p}\,\overline{q}}{\sigma_X^2}\right), \qquad (2.4.9)$$

onde $\overline{p} = \sum_{i=1}^{k} \frac{p_i}{k}$ e $\overline{q} = 1-\overline{p}$.

É fácil ver que $KR_{20} \geq KR_{21}$, com a igualdade ocorrendo se e somente se todos os p_i's **forem iguais.**

As duas fórmulas (2.4.8) e (2.4.9) representam a fidedignidade definida por KUDER-RICHARDSON(1937), mas derivadas sob uma suposição mais complicada daquela dada no corolário (2.4.2).

Condições ainda diferentes das estabelecidas no corolário (2.4.2) foram derivadas por GULLIKSEN(1950) e JACKSON-FERGUSON(1941). Vamos apresentar a seguir estas condições e provar que são equivalentes as do corolário (2.4.2), ou seja, que todas as componentes de medida devem ser essencialmente τ-equivalentes para que o coeficiente α de Cronbach, ou o caso particular KR_{20}, represente o coeficiente de fidedignidade de um teste.

Suposições de Gulliksen

GULLIKSEN(1950) considera dois testes, cada qual com n itens. Supõe que os itens entre os testes são paralelos dois a dois, e que a média das covariâncias entre itens paralelos é igual a média das covariâncias entre pares de itens dentro de cada teste. Se ρ_{ij} e σ_i denotam as correlações e os desvios padrões dos escores observados, a suposição de Gulliksen pode ser escrita como

$$\sum_{i=j}^{k} \rho_{ii'}\sigma_i^2 = \sum_{i\neq j}\sum \frac{\rho_{ij}\sigma_i\sigma_j}{k-1}, \qquad (2.4.10)$$

onde i e i' são itens paralelos e i,j são itens do mesmo teste.

Temos então n covariâncias entre itens paralelos (i,i') e n(n-1) covariâncias dentro de cada teste.

Mas

$$\rho_{ii'}\sigma_i^2 = \sigma(T_i,T_{i'}) = \sigma^2(T_i) \qquad \text{e}$$

$$\rho_{ij}\sigma_i\sigma_j = \sigma(Y_i,Y_j) = \sigma(T_i,T_j).$$

Logo (2.4.10) pode ser escrita como

$$\sum_{i=1}^{k} \sigma^2(T_i) = \sum\sum_{i\neq j} \frac{\sigma(T_i,T_j)}{k-1}.$$

A equivalência desta suposição com a de que todas as componentes de medida são essencialmente τ-equivalentes está contida na demonstração do corolário (2.4.3) (ver equação (2.4.4)).

Segue-se então que

$$\sum_{i=1}^{k} \rho_{ii'}\sigma_i^2 = \sum\sum_{i\neq j} \frac{\rho_{ij}\sigma_i\sigma_j}{k-1} \Leftrightarrow T_i = a_{ij}+T_j,\ \forall i,j.$$

Suposições de Jackson-Ferguson

O método e as suposições usadas por JACKSON-FERGUSON (1941) podem ser estabelecidas na seguinte derivação direta do coeficiente α. Sejam

$$X = \sum_{i=1}^{k} Y_i \quad \text{e} \quad X' = \sum_{j'=1}^{k} Y_{j'}$$

medidas compostas paralelas (as componentes de medida Y_i e $Y_{j'}$ correspondentes a X e X' não são necessariamente paralelas ou essencialmente τ-equivalentes).

Sejam σ_{ij} a covariância entre as componentes de medida Y_i e Y_j de X, e $\sigma_{ij'}$, a covariância entre as componentes de medida Y_i de X e $Y_{j'}$ de X'. Seja $\sigma_{ii} = \sigma_i^2$.

Da equação (1.3.5) temos que a fidedignidade ρ_{XT}^2 é igual a covariância $\rho_{XX'}$ entre as duas medidas paralelas do teste, isto é,

$$\rho_{XT}^2 = \rho_{XX'} = \frac{\sigma(X,X')}{\sigma^2(X)} = \frac{\sum_i\sum_{j'} \sigma(Y_i,Y_{j'})}{\sum_i\sum_j \sigma(Y_i,Y_j)}. \qquad (2.4.11)$$

Jackson-Ferguson supõe então que

$$\frac{\sum\limits_{i}\sum\limits_{j'} \sigma(Y_i, Y_{j'})}{k^2} = \frac{\sum\sum\limits_{i \neq j} \sigma(Y_i, Y_j)}{k(k-1)}, \qquad (2.4.12)$$

isto é, a média das covariâncias entre componentes de formas diferentes de um teste é igual a média das covariâncias entre componentes da mesma forma do teste. Então de (2.4.11) e (2.4.12) temos que

$$\rho^2_{XT} = \left(\frac{k}{k-1}\right)\left[\frac{\sum\sum\limits_{i \neq j} \sigma(Y_i, Y_j)}{\sum\limits_{i}\sum\limits_{j} \sigma(Y_i, Y_j)}\right]$$

$$= \left(\frac{k}{k-1}\right)\left[\frac{\sum\limits_{i}\sum\limits_{j} \sigma(Y_i, Y_j) - \sum\limits_{i} \sigma^2(Y_i)}{\sum\limits_{i}\sum\limits_{j} \sigma(Y_i, Y_j)}\right]$$

$$= \left(\frac{k}{k-1}\right)\left[1 - \frac{\sum\limits_{i=1}^{k} \sigma^2(Y_i)}{\sum\limits_{i}\sum\limits_{j} \sigma(Y_i, Y_j)}\right]. \qquad (2.4.13)$$

que é igual ao coeficiente α de Cronbach.

Uma proposição análoga aplica-se, naturalmente, a $\rho^2_{X'T'}$.

Assim as suposições de JACKSON-FERGUSON(1941) são suficientes para o coeficiente α ser igual a fidedignidade de um teste. Mas se α é igual a fidedignidade então todas as componentes de X devem ser essencialmente τ-equivalentes (condições do corolário (2.4.2)). Entretanto se X e X' são medidas compostas paralelas para as quais a fidedignidade é dada pelo coeficiente α então a igualdade (2.4.12) deve ocorrer. Isto pode ser mostrado facilmente igualando (2.4.11) a (2.4.13). Segue-se então que se X e X' são medidas compostas paralelas, uma condiçaõ necessária e suficiente para α ser igual a fidedignidade de X e X' é justamente a condição (2.4.12),e que é equivalente, a condição dada pelo corolário (2.4.2).

CAPÍTULO III

FATORES QUE AFETAM O COEFICIENTE DE FIDEDIGNIDADE DE UM TESTE

3.1 - *Heterogeneidade do Grupo*

É bastante conhecido o fato que o tamanho de um coeficiente de correlação depende da natureza da população na qual as medidas foram realizadas. Segundo LORD-NOVICK(1968,pág.129-131), este efeito sobre a fidedignidade pode ser mostrado através do teorema a seguir.

3.1.1 - *Teorema* - Sejam A e A' populações de respondentes tal que $A' \subset A$. Suponhamos que para a medida X definida em A, os erros são homocedásticos, isto é, $\sigma^2(\varepsilon/\tau) = \sigma^2_{\varepsilon}$.

Seja $\tilde{X}$ a medida X, restrita a A'.

Suponhamos que a restrição de A em A' seja feita eliminando aleatoriamente respondentes com escores verdadeiros em algum conjunto específico de valores τ.

Então o coeficiente de fidedignidade na população restrita á dado por

$$\rho^2_{\tilde{X}\tilde{T}} = 1- \frac{\sigma^2_X}{\sigma^2_{\tilde{X}}}(1- \rho^2_{XT}) \qquad (3.1.1)$$

Demonstração - De acordo com a igualdade (1.3.7) temos que $\sigma^2_{\varepsilon} = \sigma^2_X(1-\rho^2_{XT})$.

Como a seleção de respondentes é feita baseada apenas no conjunto de valores τ, e que $\sigma^2(\varepsilon/\tau)$ é constante para todo τ, segue-se que $\sigma^2(\tilde{\varepsilon}) = \sigma^2(\varepsilon)$ para qualquer subpopulação.

Portanto

$$\sigma^2_{\varepsilon} = \sigma^2_{\tilde{\varepsilon}} = \sigma^2_{\tilde{X}}(1-\rho^2_{\tilde{X}\tilde{T}}) \quad \text{e} \quad \sigma^2_X(1-\rho^2_{XT}) = \sigma^2_{\tilde{X}}(1-\rho^2_{\tilde{X}\tilde{T}}).$$

Segue-se então que

$$\rho^2_{\tilde{X}\tilde{T}} = 1- \frac{\sigma^2_X}{\sigma^2_{\tilde{X}}}(1-\rho^2_{XT}).$$

A partir do teorema (3.1.1), temos que à fidedignidade $\rho^2_{\tilde{X}\tilde{T}}$ é uma função estritamente crescente de $\sigma^2_X/\sigma^2_{\tilde{X}}$, e que $\rho^2_{XT} = \rho^2_{\tilde{X}\tilde{T}}$, quando $\sigma^2_X = \sigma^2_{\tilde{X}}$. Portanto $\rho^2_{\tilde{X}\tilde{T}}$ é maior (menor) que ρ^2_{XT} quando $\sigma^2_{\tilde{X}}$ é maior (menor) que σ^2_X. Isto é, quando restringimos a população A para A', e assim, reduzimos a variância do escore observado, também diminuímos a fidedignidade associada ao instrumento de medida, mostrando que a fidedignidade e a heterogeneidade dos escores observados correlacionam-se positivamente.

Além disto, podemos notar que a derivação do resultado deste teorema não depende da suposição que A' está contido em A. Realmente poderíamos ter A contido em A' e $\sigma^2_{\tilde{X}} > \sigma^2_X$, e o teorema (3.1.1) continuaria sendo válido.

Uma vez que $0 < \rho^2_{XT} < 1$, $\sigma^2_X = \sigma^2_T+\sigma^2_{\varepsilon}$ e $\sigma^2_{\tilde{X}} = \sigma^2_{\tilde{T}}+\sigma^2_{\varepsilon}$, podemos mostrar facilmente que

$$0 < (\sigma^2_X/\sigma^2_{\tilde{X}})(1-\rho^2_{XT}) < 1.$$

Portanto, quando substituímos σ^2_X, $\sigma^2_{\tilde{X}}$ e ρ^2_{XT} pelos seus respectivos valores, a fórmula (3.1.1) produz $0 < \rho^2_{\tilde{X}\tilde{T}} < 1$. Para σ^2_X e ρ^2_{XT} fixados, $\rho^2_{\tilde{X}\tilde{T}}$, é uma função estritamente crescente de $\sigma^2_{\tilde{X}}$.

Convém salientar que em muitas aplicações os dados são tais que

$$\sigma^2_{\tilde{\varepsilon}} \leqq \sigma^2_{\varepsilon}, \qquad (3.1.2)$$

isto é, a variância do erro na população restrita é menor ou igual a variância do erro na população não restrita. Isto pode ocorrer, se os erros tiverem distribuição Binomial e, somente 10 ou 15% dos respondentes que obtiverem escores mais elevados estiverem incluídos no grupo restrito. Para tais modelos temos freqüentemente $\sigma^2_{\tilde{\varepsilon}} \leqq \sigma^2_{\varepsilon}$ ou

$$\rho^2_{XT} \leqslant 1 - \frac{\sigma^2_{\tilde{X}}}{\sigma^2_X}(1-\rho^2_{\tilde{X}\tilde{T}}), \qquad (3.1.3)$$

ocorrendo a igualdade se e somente se $\sigma^2_{\tilde{\varepsilon}} = \sigma^2_{\varepsilon}$.

Prosseguindo a análise do teorema (3.1.1), temos que se ele for aplicado sob a condição (3.1.2), o valor da fidedignidade de um teste na população não restrita será tanto maior, quanto maior for a variância do erro desta população em relação a variância do erro da população restrita. Por outro lado, em muitas aplicações, a determinação da fidedignidade num grupo restrito, a partir da fidedignidade num grupo não restrito, inverte a desigualdade em (3.1.3) e fornece um limite inferior para a fidedignidade neste grupo. No caso do erro se distribuir Binomialmente, isto ocorreria se apenas 10 ou 15% dos respondentes com escores mais baixos fossem excluídos do grupo restrito.

Embora para aplicar o teorema (3.1.1) deve-se ter $\sigma^2(\varepsilon/\tau) = \sigma^2_{\varepsilon}$, para qualquer par de populações, em aplicações simples pode-se desprezar esta suposição e exigir somente que as variâncias dos erros sejam iguais.

A seguir apresentaremos um exemplo da aplicação do teorema (3.1.1).

Exemplo - Suponha que o coeficiente de fidedignidade do teste A é 0,5 na população restrita, e do teste B é 0,7, na população não restrita. Considere também que $\sigma_X^2/\sigma_{\tilde{X}}^2$ é igual a 2, em ambos os testes.

Se $\sigma_{\varepsilon}^2 = \sigma_{\tilde{\varepsilon}}^2$ então a fidedignidade do teste A na população não restrita é

$$\rho_{XT}^2 = 1 - \frac{\sigma_{\tilde{X}}^2}{\sigma_X^2}(1-\rho_{\tilde{X}\tilde{T}}^2)$$

$$= 1 - \frac{1}{2}(1-0,5) = 0,75.$$

Portanto o teste A é mais fidedigno que o B.

No entanto se $\sigma_{\varepsilon}^2 = 1,5\sigma_{\tilde{\varepsilon}}^2$ (que em algumas situações poderia ser uma situação mais plausível), a fidedignidade do teste A seria 0,63, e portanto, B seria mais fidedigno que A.

De acordo com este exemplo, e com a dificuldade de estabelecer com segurança a relação entre σ_{ε}^2 e $\sigma_{\tilde{\varepsilon}}^2$, o resultado do teorema (3.1.1) deve ser aplicado com muita cautela e nível de confiança elevado sobre as suposições.

De qualquer maneira estes resultados mostram que um coeficiente de fidedignidade alto, pode freqüentemente ser obtido , aplicando-se um teste a um grupo de sujeitos suficientemente heterogêneo. Assim, por exemplo, um coeficiente de fidedignidade de 0,91($\sigma_{\varepsilon}^2 = 16$ e $\sigma_X^2 = 185$) e 0,50($\sigma_{\varepsilon}^2 = 16$ e $\sigma_X^2 = 32$) podem ser encontrados no mesmo teste, dependendo do grupo em que ele é aplicado.

SYMONDS(vol XIX,pág.73-87) indica vários fatores que podem afetar um coeficiente de fidedignidade. Em particular, é pos sível garantir altos coeficientes de fidedignidade, construindo testes com grande número de itens, ou aplicando testes pequenos para um grupo grande e extremamente variado de indivíduos.

Quando a heterogeneidade do grupo aumenta, tanto o desvio padrão, quanto a fidedignidade, também aumentam. No entanto, o erro padrão de medida (fórmula (1.3.7)), que é uma função do desvio padrão dos escores do teste e da fidedignidade, permanece constante. Por exemplo, um erro padrão igual a 6 pode ser obtido, tanto a partir de σ_X^2 igual a 10 e ρ_{XT}^2 igual a 0,64, quanto de σ_X^2 igual a 15 e ρ_{XT}^2 igual a 0,84.

Portanto para determinar se um coeficiente de fidedignidade é suficientemente alto para produzir uma precisão satisfatória, é conveniente ter alguma informação sobre a heterogeneidade do grupo, a partir do qual foi obtido este coeficiente.

Por estas razões, o STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS AND MANUAL do American Psychological Association, American Educational Research Association, e o National Council on Measurement in Education (1966), recomenda que a variância do escore observado sempre acompanhe o coeficiente de fidedignidade. A proposição é que no mínimo dois momentos de segunda ordem ou seus equivalentes (por exemplo: ρ_{XT}^2 e σ_X^2, ou ρ_{XT}^2 e σ_ε^2), são necessários para descrever a precisão de um teste como um instrumento de medida.

3.2 - *Tamanho do Teste*

Suponhamos que o tamanho de um teste é definido como o número de componentes paralelos que o compõem. Se K componentes paralelos realmente podem ser obtidos, então exceto para flutuações amostrais, a fórmula de Spearman-Brown dada por

$$\rho_k = \frac{K\rho}{1+(K-1)\rho} \qquad (3.2.1)$$

onde K é o número de vezes que o teste é aumentado,

ρ é a fidedignidade do teste inicial,

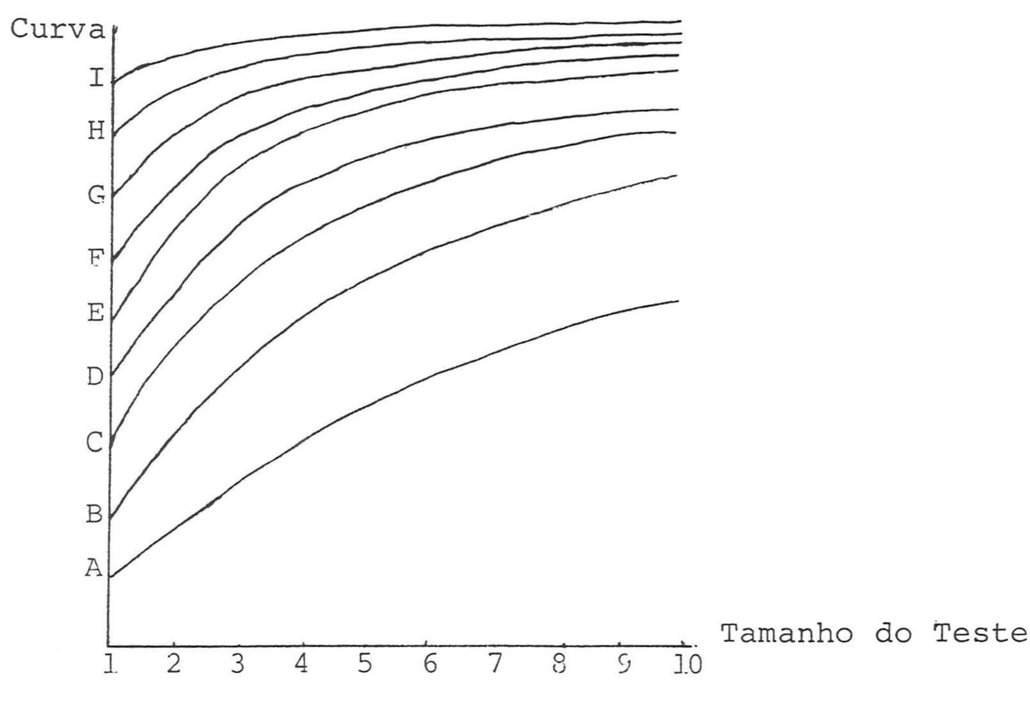
ρ_k é a fidedignidade do teste aumentado,

prevê corretamente a fidedignidade de um teste aumentado K vezes, a partir da fidedignidade do teste original. Se as componentes são essencialmente τ-equivalentes, mas não são paralelas, então a fórmula (2.4.1) deve ser utilizada.

A seguir apresentaremos uma tabela e um gráfico, que ilustram a relação entre a fidedignidade e o tamanho de um teste, de acordo com a fórmula de Spearman-Brown.

Relação entre o coeficiente de fidedignidade e o tamanho de um teste

Curva	Tamanho do Teste									
	1	2	3	4	5	6	7	8	9	10
	1/10	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	1
A	.100	.182	.250	.308	.357	.400	.430	.471	.500	.526
B	.200	.333	.429	.500	.556	.600	.636	.667	.692	.714
C	.300	.461	.562	.631	.682	.720	.750	.774	.794	.811
D	.400	.571	.667	.727	.769	.800	.824	.842	.857	.870
E	.500	.667	.750	.800	.833	.857	.875	.889	.900	.909
F	.600	.750	.818	.857	.882	.900	.913	.923	.931	.938
G	.700	.824	.875	.903	.921	.933	.942	.949	.955	.959
H	.800	.889	.923	.941	.952	.960	.966	.970	.973	.976
I	.900	.947	.964	.973	.978	.982	.984	.986	.988	.989



Para determinar o efeito do tamanho de um teste na fidedignidade, use a primeira linha da escala de tamanho dos testes, na qual o valor um corresponde a um teste com tamanho unitário e fidedignidade dada à esquerda. Para determinar o efeito da diminuição do tamanho de um teste na fidedignidade, use a segunda linha, na qual o valor de referência á dado à direita.

Por exemplo, a curva tem uma fidedignidade de 0,857 no valor escalar 6 e fidedignidade 0,800 e 0,667 com 2/3 e 1/3 do tamanho do teste, respectivamente, para uma fidedigniadde inicial de 0,500.

É possível determinar também o tamanho que um teste deve ter para atingir uma fidedignidade específica. Assim, por exemplo a curva F tem uma fidedignidade de 0,750 no valor escalar 2 e 0,938 para o valor escalar 10. Portanto, é necessário um teste 5 vezes maior para a fidedigniadde ser aumentada de 0,750 para 0,938.

No valor escalar 6, temos 0,800 para uma fidedignidade inicial de 0,400 e 0,982 para uma fidedignidade inicial de 0,900, indicando que para um mesmo aumento no tamanho de um teste, no primeiro caso temos um aumento na fidedignidade de 100%, enquanto que no segundo caso, de apenas 9,11%.

Na prática as componentes paralelas exigidas na fórmula de Spearman-Brown são muito improváveis de serem encontradas. De qualquer maneira ela é útil, no sentido que ajuda decidir pela rejeição ou aceitação de um teste, de acordo com o tamanho que ele deve ter para atingir a fidedignidade desejada. Por exemplo, não é conveniente aumentar um teste de 10 a 15 vezes. Não somente é difícil encontrar o material adequado, como também o cansaço e outros fatores, fazem os resultados não serem tão precisos quanto se desejaria. Se o conteúdo a ser adicionado ao teste estiver estritamente relacionado com o original (mediante uma cuidadosa seleção e equiparação de itens), e se a motivação permanecer substancialmente constante, as experiências indicam que um teste pode ser aumentado 5 ou 6 vezes em relação ao tamanho inicial, e ainda assim a fórmula (3.2.1) produzirá uma estimação bastante satisfatória dos resultados. Se o tamanho do teste deve aumentar ainda mais do que isto, a fórmula de predição tenderá superestimar o coeficiente de fidedignidade. Contudo este fato não é demasiadamente sério, pois um teste que necessita de semelhante aumento, provavelmente deveria ser modificado radicalmente ou ser abandonado em favor de outro.

Particularmente, se somente a suposição de paralelismo for violada, para qualquer aumento do tamanho de um teste, LORD-NOVICK(1968,pág.139) afirmam que a fórmula de Spearman-Brown subestima a fidedignidade do teste aumentado. Além disto, a possível violação de experimentos independentes pode subestimar a variância do erro do teste aumentado,e conseqüentemente, superestimar a fidedigniadde.

Não há dúvidas, portanto, que se a fórmula (3.2.1) for usada imprudentemente, a estimativa resultante terá pouca relação com a realidade. Entretanto se a aplicação da fórmula estiver restrita às situações nas quais as suposições são razoavelmente satisfeitas, pode-se obter estimativas totalmente satisfatórias.

Comparação entre Coeficientes de Fidedignidade de Testes com Tamanhos Diferentes

Suponhamos que temos um conjunto de testes de diferentes tamanhos e desejamos comparar suas fidedignidades. Um procedimento simples para atingir este objetivo é usar a fórmula de Spearman-Brown no sentido inverso, para determinar qual a fidedignidade que cada teste teria com tamanho unitário. Facilmente obtemos

$$\rho = \frac{\rho_k}{K-(K-1)\rho_k} \qquad (3.2.2)$$

onde ρ e ρ_k são as fidedignidades dos testes com tamanho unitário e K, respectivamente.

Uma vez que a fórmula de Spearman-Brown determina a fidedignidade para um teste aumentado K vezes, como uma função estritamente crescente de sua fidedignidade no tamanho unitário, segue-se que a ordenação das fidedignidades é a mesma tanto para K, quanto para o tamanho unitário. Portanto o teste que apresentar maior ρ, será o teste mais fidedigno.

Exemplo - Suponhamos que 4 testes de habilidades mentais estão disponíveis cujos tamanhos (número de itens) e correspondentes fidedignidades são (20;0,80), (30;0,85), (40;0,93), e (60;0,95), e que desejamos determinar qual deles é o mais fidedigno.

Aplicando a fórmula (3.2.2), determinamos que a fidedignidade do primeiro teste com tamanho unitário é

$$\rho = \frac{0,80}{20-(20-1)0,80} = 0,167.$$

Analogamente encontramos as fidedignidades dos outros testes no tamanho unitário, e que correspondem a 0,159; 0,249 e 0,240, respectivamente. Segue-se então, que para qualquer tamanho, o terceiro teste é o mais fidedigno.

Podemos ainda concluir da fórmula de Spearman-Brown que a fidedignidade de um teste não é diretamente proporcional ao seu tamanho, mas que há uma função monótona de ρ, isto é, $\rho/(1-\rho)$, que é proporcional ao tamanho do teste. Verifiquemos isto através do teorema a seguir.

3.2.1 - *Teorema* - Sejam ρ_k e $\rho_{k'}$, fidedignidades de testes com tamanhos K e K', respectivamente. Então

$$\frac{\rho_k/(1-\rho_k)}{\rho_{k'}/(1-\rho_{k'})} = \frac{K}{K'} . \qquad (3.2.3)$$

Demonstração - Da fórmula (3.2.2) podemos escrever

$$\rho = \frac{\rho_{k'}}{K'-(K'-1)\rho_{k'}} . \qquad (3.2.4)$$

Usando a fórmula (3.2.4) em (3.2.1), obtemos facilmente que

$$\rho_k = \frac{K\rho_{k'}}{K'+(K-K')\rho_{k'}} .$$

Segue-se então que

$$\frac{\rho_k/(1-\rho_k)}{\rho_{k'}/(1-\rho_{k'})} = \frac{K}{K'} .$$

De acordo com este teorema podemos, a partir de um teste com tamanho K' e fidedignidade $\rho_{k'}$, determinar qual deverá ser seu tamanho para possuir uma fidedignidade específica ρ_k. Isto é,

$$K = K'\frac{\rho_k/(1-\rho_k)}{\rho_{k'}/(1-\rho_{k'})} . \qquad (3.2.5)$$

Exemplo - Um teste com 25 itens possui fidedignidade igual a 0,70 e pretendemos que sua fidedignidade seja de 0,80. Então pela fórmula (3.2.5) temos que

$$K = 25 \times \frac{0,80/(1-0,80)}{0,70/(1-0,70)} = 42,85$$

é o número de itens necessários para que o teste tenha uma fidedignidade de 0,80.

CAPÍTULO IV

MÉTODOS DE ESTIMAÇÃO DO COEFICIENTE DE FIDEDIGNIDADE DE UM TESTE

4.1 - *Introdução*

O coeficiente de fidedignidade refere-se a precisão (consistência interna, estabilidade e equivalência) de um instrumento de medida ou teste. A precisão é um termo geral, sinônimo de fidedignidade; a consistência interna refere-se a forma da distribuição dos escores; a estabilidade diz respeito a constância da característica que está sendo medida, no decorrer do tempo, enquanto que a equivalência trata do grau de dependência das medidas, em relação a um conjunto particular de itens ou conteúdo.

Da definição (1.4) temos que o valor numérico da fidedignidade de um teste é obtido através do cálculo do quadrado da correlação entre o escore verdadeiro T e o seu correspondente escore observado X, ou então, se dispomos de duas medidas paralelas X e X', através do coeficiente de correlação entre elas. Tanto ρ^2_{XT}, quanto $\rho_{XX'}$, são parâmetros desconhecidos, e portanto, a determinação do coeficiente de fidedignidade fica reduzida a estimação destes parâmetros. Esta estimação implica na necessidade de medidas repetidas de uma amostra aleatória de sujeitos.

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. Nestas condições podemos ser 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 nesta situação tem sido realizados, enquanto que na situação (b) os artigos são numerosos.

Quando pretendemos estimar a fidedignidade temos que, além de considerar essas duas situações, supor que os escores dos itens do teste ou dos subtestes possuem uma determinada distribuição de probabilidade, ou então, não considerar esta suposição.

Apresentaremos a estimação deste coeficiente, tratando inicialmente com as situações (a) e (b) e não considerando a distribuição dos escores do teste ou dos subtestes. Concluído este tratamento, consideraremos a situação (b), sob a suposição que os escores dos itens ou subtestes que compõem o teste tem distribuição Normal Multivariada.

4.2 - *Estimação do Coeficiente de Fidedignidade, Omitindo a Distribuição de Probabilidade dos Escores do Teste ou dos Subtestes*

Na prática a estimação do coeficiente de fidedignidade, sem considerar a distribuição dos escores do teste ou de suas partes (subtestes), é realizada através de 3 métodos gerais: teste-reteste, formas paralelas e consistência interna.

Estes métodos partem de suposições diferentes, de tal forma que, a partir do conhecimento da estimativa da fidedignidade por um dos métodos, não podemos obter qualquer informação a respeito da estimativa por qualquer outro método. A principal diferença está na fonte de variação do erro e do escore verdadeiro das medidas do teste. Fatores que contribuem para a variância do erro em um dos métodos, podem contribuir para a variância do escore verdadeiro em outro método.

A seguir definiremos e analisaremos os principais procedimentos usados para estimar o coeficiente de fidedignidade de um teste.

4.2.1 - *Coeficiente de Estabilidade*

Teste-Reteste

Neste método o mesmo teste é respondido duas vezes pelos mesmos sujeitos, e o coeficiente de correlação linear entre as duas medidas é considerado um estimador do coeficiente de fidedignidade. Os ensaios são supostos independentes, mas isto pode não ocorrer, produzindo uma subestimativa ou superestimativa da fidedignidade do teste.

O coeficiente obtido pelo método de teste-reteste produz informações sobre a estabilidade dos sujeitos, no período de tempo considerado, por esta razão também denominado coeficiente de estabilidade. Neste caso, a variância do erro é devido às flutuações aleatórias do desempenho dos sujeitos, entre as duas aplicações do teste.

Um coeficiente de estabilidade alto pode indicar que ocorreram poucas mudanças no comportamento dos sujeitos da primeira para a segunda aplicação, como também que o teste mede a mesma função em ambas as ocasiões. Um coeficiente baixo pode indicar que ocorreram alterações nos comportamentos em diferentes direções ou proporções. A análise das variações nas médias e desvios padrões pode ajudar a interpretar os tipos de mudanças que ocorreram.

Há vários problemas relacionados com a utilização do método teste-reteste para estimar a fidedignidade. Primeiramente, há provavelmente o efeito da prática ou memória na segunda aplicação, tal que os erros entre as duas aplicações tendem estar correlacio nente no comportamento do sujeito, frente à característica que está sendo medida.

Quando utilizamos este método é difícil saber se estes efeitos inflacionam ou deflacionam a correlação entre as medidas repetidas. De acordo com GULLIKSEN(1950,pág.197-98), estes efeitos, geralmente, tendem inflacionar a correlação, produzindo uma superestimativa do coeficiente de estabilidade, principalmente quando o período entre as aplicações for pequeno e os efeitos da fadiga forem minimizados.

Certamente a medida em que o tempo entre as aplicações aumenta, os efeitos da memória tornam-se menos importantes, dando lugar aos efeitos da alteração de comportamento dos sujeitos, implicando num decréscimo na correlação entre as medidas do teste e reteste, e portanto, produzindo uma subestimativa do coeficiente de fidedignidade. Entretanto, se o intervalo entre a primeira e a segunda aplicação for suficiente para neutralizar, pelo menos em parte, os efeitos da memória, fadiga, mudanças de comportamento , então a correlação entre as medidas das duas aplicações produzirá uma estimativa aproximada da estabilidade dos sujeitos no teste. Sob estas condições, GUTTMAN(1945) é favorável a utilização deste método.

4.2.2 - *Coeficiente de Equivalência*

Formas Paralelas

A estimação do coeficiente de fidedignidade através deste método exige, inicialmente, a construção de duas formas paralelas do mesmo teste, cada qual composta por itens que fornecem medidas paralelas quando aplicadas.

Duas formas de um teste são paralelas quando a média das correlações entre os itens dentro de cada forma é igual a média das correlações entre os itens das duas formas. Esta condição é razoavelmente satisfeita quando os itens são retirados de uma população, em que todos medem o mesmo fator.

Na elaboração, de formas paralelas, deve-se ter o cuidado de equiparar os instrumentos de medida, quanto ao conteúdo, dificuldade e forma, além de tomar precauções para que os itens de ambas as formas não sejam demasiadamente semelhantes.

O coeficiente de correlação linear entre as medidas das duas formas paralelas, aplicadas ao mesmo tempo, é um estimador da fidedignidade do teste, e é denominado coeficiente de equivalência.

Este coeficiente produz informações sobre a estabilidade dos sujeitos nas variações do mesmo teste ou dos diferentes itens que foram construídos para medir a(s) mesma(s) função(funções). Ele indica o grau de dependência das medidas em relação a um particular conjunto de itens ou do conteúdo usado.

O coeficiente de equivalência é uma característica de um par de formas paralelas de um mesmo teste, e varia de acordo com o tipo de similaridade estabelecido no equacionamento das formas.

Se as formas paralelas forem aplicadas com um intervalo de tempo entre elas então o coeficiente obtido é de equivalência e de estabilidade.

Este método considera como erro tanto a imprecisão do instrumento e variações entre as formas, quanto as mudanças de comportamento (alterações nos escores verdadeiros) entre as duas aplicações.

Se as formas paralelas forem demasiadamente semelhantes, o coeficiente obtido superestimará a fidedignidade, enquanto que subestimará se elas forem muito diferentes.

Divisão de um Teste pela Metade (Método das Metades)

Entre os artigos consultados, encontrou-se divergências quanto a considerar o estimador obtido pelo método da divisão em metades, como um estimador de equivalência ou de consistência interna. Por exemplo, CRONBACH(1947,1951), o considera de equivalência, enquanto que LORD-NOVICK(1968, pág. 135) o consideram de consistência interna. Em função do tratamento que estamos dando à teoria da fidedignidade, optamos por considerá-lo como um coeficiente de equivalência.

Um único teste é aplicado somente uma vez. A seguir, divide-se o teste em duas partes de mesmo tamanho, somam-se os escores de cada sujeito em cada metade. Define-se então, o coeficiente de correlação linear entre as medidas das metades como um estimador da fidedignidade de um teste com a metade do tamanho do teste inicial. Aplica-se então a fórmula de Spearman-Brown (equação (2.2.10)), obtendo-se o estimador da fidedignidade do teste total.

A aplicação da fórmula de Spearman-Brown exige que as metades sejam paralelas, o que não implica que os itens que as formam também apresentem esta característica. Observa-se, porém, que esta condição geralmente não é observada pelos pesquisadores, em vista dos procedimentos de divisão em metades propostos e usados na prática, e também porque não é fácil obter metades paralelas. Os procedimentos propostos correspondem, por exemplo, em designar os itens aleatoriamente para cada metade; numerar os itens e separá-los em pares e ímpares; agrupar os itens de tal forma que as metades tenham iguais médias e iguais desvios padrões (metades paralelas).

GUTTMAN (1945) mostrou que a estimativa da fidedignidade obtida, usando qualquer critério para dividir o teste pela meta de, é um limite inferior para o verdadeiro coeficiente de fidedignidade.

CRONBACH(1947) analisou o efeito das várias divisões em relação a estimação da fidedignidade e concluiu que o melhor critério é aquele em que as metades são paralelas ou aproximadamente paralelas, uma vez que neste caso, a estimativa é maior ou igual que qualquer outra.

A carência de uma única estimativa para a fidedignidade é um dos motivos pelo qual este método tem recebido muitas críticas. Ao invés de se obter um único coeficiente, resultam diferentes coeficientes, dependendo de como os itens foram agrupados nas metades. Segue-se então que a estimativa obtida é uma característica da divisão realizada e não do teste.

O método das metades deve ser aplicado com cautela. Se os itens estiverem dispostos em ordem crescente de dificuldade, o que é comum, por exemplo, em testes de escolaridade, então a primeira metade estará correlacionada com a segunda, contradizendo as suposições do modelo clássico linear de testes. No entanto, se os itens não estiverem assim ordenados, mas existir um limite de tempo para o teste ser respondido, este limite atuará de forma mais efetiva sobre a segunda metade. Na prática, esta situação tende transformar a primeira metade em um teste de capacidade, e a segunda num teste, no mínimo, parcialmente de velocidade (LORD-NOVICK(1968-pág.131-33)).

A divisão em metades constituídas, respectivamente por itens pares e ímpares, também inspira cuidados, principalmente em testes de velocidade ou parcialmente de velocidade. Qualquer indivíduo, deixando de responder os últimos 2r itens, terá um escore nulo em cada um destes itens, e conseqüentemente, eles estarão perfeitamente correlacionados. Quando estes itens forem agrupados com os respondidos, para produzir a estimativa da fidedignidade, pode-se obter um valor substancialmente distante do verdadeiro, pela violação de uma das suposições da teoria de testes em consideração.

4.2.3 - *Coeficiente de Consistência Interna*

O terceiro método para estimar a fidedignidade envolve uma análise interna das variâncias e covariâncias dos escores observados nos itens ou subtestes, em uma única aplicação de um teste.

Se um teste tem substancial consistência interna então ele é psicologicamente interpretável. Dois testes nestas condições, compostos por diferentes itens, geralmente registram o mesmo resultado.

Entretanto, se um teste é composto por um conjunto de itens, cada qual medindo um fator diferente, torna-se duvidoso que fator invocar para explicar o significado de um escore simples.

Para que um teste seja interpretável, não é necessário que todos os itens sejam fatorialmente iguais, o que é indispensável é que uma grande proporção da variância do teste seja atribuída ao fator principal distribuído no teste.

A consistência interna de um teste será tanto maior , quanto mais fortemente os itens se correlacionarem positivamente, e apresentarem o mesmo nível de dificuldade.

O coeficiente de consistência interna pode ser obtido através das fórmulas 20 e 21 de Kuder-Richardson e do coeficiente α de Cronbach.

Fórmulas de Kuder-Richardson

Uma reformulação radical nos problemas de fidedignidade foi proporcionada por KUDER-RICHARDSON(1937). Eles propuseram várias fórmulas alternativas quando um teste é composto por itens dicotômicos, e que foram amplamente adotadas, em especial as fórmulas KR_{20} e KR_{21}. Recordando (equações (2.4.8) e (2.4.9)), estas fórmulas são definidas por

$$KR_{20} = \frac{k}{k-1}\{1- \frac{\sum_{i=1}^{k} p_i q_i}{\sigma_X^2}\} \quad \text{e}$$

$$KR_{21} = \frac{k}{k-1}\{1- \frac{k\,\bar{p}\,\bar{q}}{\sigma_X^2}\}, \quad \bar{p} = \sum_{i=1}^{k} \frac{p_i}{k}$$

tal que K é o número de itens do teste, (itens essencialmente τ - equivalentes)

p_i é a proporção de sucessos no item i e

σ_X^2 é a variância do teste.

A derivação original de KR_{20} e KR_{21} foi criticada por causa das numerosas suposições, que transformadas para a linguagem usada neste trabalho, equivale a condição de que todos os itens que formam o teste devem ser essencialmente τ-equivalentes.

A medida que os itens não são essencialmente τ-equivalentes, estas fórmulas tendem subestimar o coeficiente de fidedignidade, porém não gravemente, a menos que as medidas se afastam radicalmente da τ-equivalência.

Conforme já provamos no capítulo II, secção 4, os coeficientes obtidos pelas fórmulas de Kuder-Richardson, KR_{20} e KR_{21}, são casos particulares do coeficiente α de Cronbach.

Coeficiente α de Cronbach

Dispensamos a definição do coeficiente uma vez que já o apresentamos no capítulo II, a vamos que ele é um limite inferior para a fidedignida teste, e coincide com esta, quando todas as medidas que compo o teste são essencialmente τ-equivalentes.

Entre todos os métodos de estimação do coeficiente de fidedignidade, o coeficiente α é o mais utilizado.

O Coeficiente α de Cronbach Aplicado a uma Bateria de Testes ou Subtestes

Ao invés de considerarmos α como um índice de consistência interna dos itens de um teste, podemos aplicá-lo também nas questões de consistência interna dos subtestes.

Se cada subteste é considerado como um item de um teste, então a fórmula (2.4.6) torna-se

$$\alpha = \frac{K}{K-1}\left\{1-\frac{\sum_{i=1}^{k}\sigma^2(Y_i)}{\sigma_X^2}\right\}, \qquad (4.2.1)$$

onde K é o número de subtestes,

Y_i é o escore observado no subteste i e

$$\sigma_X^2 = \sum_{i=1}^{k}\sigma^2(Y_i)+\sum_{i\neq j}\sum\sigma(Y_i,Y_j).$$

Se a fórmula (4.2.1) é aplicada num teste composto por subtestes, então ela produz informações úteis sobre a interpretabilidade desta composição.

Sob a suposição de que a média das variâncias, devido aos fatores comuns, dentro dos subtestes é igual a média das covariâncias entre os subtestes, α indica qual a proporção da variância da composição é devido aos fatores comuns entre os sub-testes (CRONBACH(1951)).

Em algumas ocasiões a variância do teste não é conhecida, mas conhecemos as covariâncias entre os subtestes, e conseqüentemente, podemos fazer uma inferência a partir destas informações. Neste caso então, sob a hipótese de que todos os subtestes tem a mesma variância, o coeficiente α é dado por

$$\alpha = \frac{K}{K-1}\left\{\frac{\sum\limits_{i\neq j}\sum \sigma(Y_i,Y_j)}{n+\sum\limits_{i\neq j}\sum \sigma(Y_i,Y_j)}\right\}, \tag{4.2.2}$$

onde Y_i e Y_j são os escores observados nos subtestes i e j.

ROZEBOOM(1966-pág.451-55) afirma que a melhor maneira de maximizar α, quando um teste é composto por K subtestes, é agrupar os itens em dois subtestes de mesmo tamanho. Em particular, α_k pode exceder o máxα_2, somente se K for ímpar e os itens puderem ser alocados em subtestes, de tal maneira que, as covariâncias entre os escores dos subtestes sejam aproximada - mente iguais; e no caso de K ser par, α_k excede α_2 no máximo em $(K^2-1)^{-1}$.

CRONBACH(1951) estabeleceu uma propriedade importante para o coeficiente α, e que será demonstrada no teorema (4.4.1) a seguir.

Suponha que temos um teste com 2k itens. Então, este teste pode ser dividido pela metade de $(2k)!/[2(k!)^2]$ maneiras. Para cada uma destas divisões, podemos considerar o valor de k, no membro à direita de (2.4.1), igual a 2 e calcular o coeficiente α para um teste composto por duas partes de mesmo tamanho e representar por α_2. Podemos também calcular o coeficiente α para os 2k ítens do teste, e analogamente, denotar por α_{2k}. Então podemos estabelecer uma relação entre os coeficientes α_2 e α_{2n} através do

4.2.3.1 - *Teorema* - $E(\alpha_2) = \alpha_{2k}$

Isto é, o valor esperado de α_2 sobre todas as possíveis divisões de um teste pela metade é igual a α_{2k}.

Este resultado pode ser interpretado de duas maneiras: primeiro ele mostra que se nós calcularmos todos os $(2k!)/[2(k!)^2]$ possíveis valores de α_2, e tomarmos a sua média aritmética, obteremos um valor igual a α_{2k}. Ele mostra também que, se os itens forem designados aleatoriamente para as duas metades então o valor esperado de α_2 é α_{2k}.

Demonstração: Seja

$$\frac{k}{k-1}\left[1-\frac{\sum\limits_{i=1}^{k}\sigma^2(Y_i)}{\sigma_X^2}\right] = \frac{k}{k-1}\left[\sum_{i\neq j}\sum\frac{\sigma(Y_i,Y_j)}{\sigma_X^2}\right] = \frac{1}{k(k-1)}\left[\frac{\sum\sum_{i\neq j}\sigma(Y_i,Y_j)}{k^{-2}\sigma_X^2}\right] \tag{4.2.3}$$

Sejam Z_1 e Z_2 metades arbitrárias do teste. Sem perda de generalidade, podemos supor que

$$Z_1 = \sum_{i=1}^{k} Y_i \qquad \text{e} \qquad Z_2 = \sum_{i=k+1}^{2k} Y_i,$$

para um teste com $2K$ itens.

Para esta divisão o coeficiente α, que denotamos por α_2 é

$$\alpha_2 = 2\left[1-\frac{\sigma^2(Z_1)+\sigma^2(Z_2)}{\sigma_X^2}\right] = \frac{4\sigma(Z_1,Z_2)}{\sigma_X^2}.$$

Expandindo $\sigma(Z_1,Z_2)$, como a covariância de suas componentes, temos que

$$\alpha_2 = \frac{4}{\sigma_X^2}\sum_{i=1}^{k}\sum_{j=k+1}^{2}\Sigma(Y_i,Y_j).$$

Tomando o valor esperado sobre todas as divisões (supondo as divisões equiprováveis devido a amostragem aleatória), obtemos

$$E(\alpha_2) = \frac{4}{\sigma_X^2} \sum_{i=1}^{k} \sum_{j=k+1}^{2k} E[\sigma(Y_i, Y_j)].$$

Para uma divisão arbitrária, todos os pares distintos de componentes tem a mesma chance (probabilidade) de serem designados por Y_i e Y_j, para $i=1,2,\ldots,k$ e $j=k+1,k+2,\ldots,2k$. Assim, todos os termos no somatório duplo são iguais, e portanto, todas as médias de covariâncias entre as componentes das diferentes partes do teste também são iguais. Como há k^2 termos no somatório duplo, podemos escrever $\varepsilon(\alpha_2)$ como

$$E(\alpha_2) = \frac{4k^2}{\sigma_X^2} \sum_{i \neq j}^{2k} \sum \frac{\sigma(Y_i, Y_j)}{2k(2k-1)} .$$

De (4.2.3) a fórmula geral para α_{2k}, o coeficiente α para um teste com 2n itens, é dada por

$$\alpha_{2k} = \frac{1}{2k(2k-1)} \sum_{i \neq j}^{2k} \sum \frac{\sigma(Y_i, Y_j)}{(2k)^{-2}\sigma_X^2} .$$

Segue-se então que

$$E(\alpha_2) = \alpha_{2k}.$$

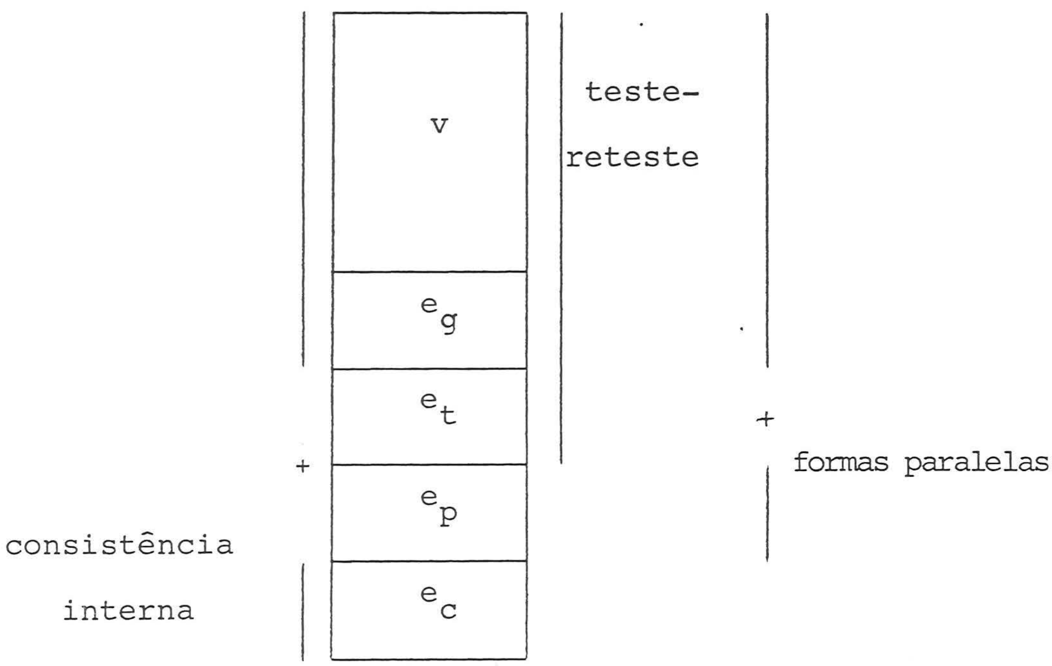
4.2.3.2 - *Corolário* - Se os 2k itens do teste são essencialmente τ-equivalentes então $\alpha_2 = \alpha_{2k}$.

Da hipótese que todos os 2k itens são essencialmente τ-equivalentes, segue-se que todas as possíveis divisões do teste pela metade também são essencialmente τ-equivalentes. Portanto pelo corolário (2.4.2), cada valor possível de α_2 é igual a fidedignidade do teste composto, α_{2k}.

4.2.4 - *Relação entre os Métodos de Estimação do Coeficiente de Fidedignidade quanto aos Fatores que Contribuem para a Variância Total*

Seja v a proporção da variância total que é devido a variância verdadeira, independente do método empregado para estimar a fidedignidade. Seja e_g as fontes de erro que afetam os 3 métodos, e que são as flutuações de atenção, memória ou motivação, que ocorrem de momento para momento ou de item para item . Sejam e_t, e_p, e_c as fontes de erro quando é aplicado o método de teste-reteste, algum método de formas paralelas e de consistência interna, respectivamente.

O diagrama a seguir apresenta a variância total na obtenção dos escores de um teste pelos 3 métodos de estimação da fidedignidade.



Em alguns testes, as suposições das fontes de erro enunciadas anteriormente, e relacionadas, no momento da apresentação de cada um dos métodos, são contribuições importantes para a variância do erro.

Se um teste é difícil, a distribuição de freqüências do escore total apresenta uma variância grande, implicando em baixa fidedignidade. Por outro lado, se um teste é fácil, tanto a dispersão do escore total quanto a fidedignidade são pequenos.

Testes com 2 alternativas são, em geral, menos fidedignos do que os testes com 4 ou 5 alternativas. Deve-se evitar testes com pequeno número de alternativas, mas se isto ocorrer, deve-se aumentar o número de itens.

4.2.5 - *Relação entre o Coeficiente α de Cronbach e o Coeficiente de Estabilidade*

GUILFORD(1950-pág.485) afirma que, é provavelmente falso que um teste possa ter alta consistência interna e baixa fidedignidade pelo método do teste-reteste, exceto depois de um longo período de tempo. Se isto ocorre, podemos inferir com alguma confiança que o teste é heterogêneo.

Ele afirma também que, se um teste for construído, de tal forma que, cada item está correlacionado com um critério , mas não está correlacionado com qualquer outro item, então a consistência interna é nula, mas o teste pode ter um coeficiente de estabilidade alto. Portanto a consistência interna e a estabilidade de um teste não concordam necessariamente.

As afirmações de GUILFORD são aceitáveis somente se vistas como um resumo de sua experiência, pois não existem relações matemáticas que comprovam estas afirmações, uma vez que as contribuições para a variância do erro e do escore verdadeiro diferem de método para método, conforme já mostramos na secção anterior.

4.2.6 - *O Método mais adequado para Estimar o Coeficiente de Fidedignidade*

A escolha do método a ser usado para estimar o coeficiente de fidedignidade depende da quantidade de informação disponível, grau de precisão e objetivo do teste.

Entretanto, na avaliação de um teste como instrumento de medida , todos os estimadores da fidedignidade são de interesse, uma vez que um mesmo teste pode ser usado para atingir diferentes objetivos. Por exemplo, podemos estar interessados na medição de características estáveis, e neste caso, o coeficiente de estabilidade se impõe. Porém, se o propósito for de pesquisa, um teste com alto coeficiente de consistência interna e baixa estabilidade pode ser satisfatório.

Existem situações em que um dos métodos é mais indicado do que os outros. Assim, por exemplo, podemos considerar o teste-reteste adequado em alguns testes de velocidade, nos quais os efeitos da aprendizagem e memorização não estão presentes, devido ao intervalo de tempo entre as duas aplicações.

No entanto, é o método das formas paralelas, geralmente, o mais indicado para estimar a fidedignidade quando os testes são de velocidade, incluindo também aqueles nos quais um número apreciável de examinados atingem o último item. Neste caso, um bom procedimento é preparar um teste com duas metades paralelas e aplicá-las um após a outra como dois testes com tempos independentes. A correlação entre as duas metades, independentemente aplicadas, pode ser utilizada na fórmula de Spearman-Brown para a determinação do estimador da fidedignidade do teste total.

Se um teste é de capacidade, o coeficiente mais indicado para estimar a fidedigniadde é o de equivalência ou de consistência interna. A preferência por um deles depende da informação disponível, dos objetivos do teste, etc.

4.2.7 - *Coeficiente de Fidedignidade Mínimo para um Teste ser considerado um Instrumento de Medida*

KELLEY(1927,pág.210-11) sugeriu que o coeficiente de fidedignidade deve ser maior ou igual a 0,5 quando o teste tem como objetivo medir algum fator (tratamento) relativo a um grupo, enquanto que deve ser pelo menos igual a 0,94 se o fator (tratamento) for individual.

Geralmente, um coeficiente de fidedignidade maior ou igual a 0,70 é considerado satisfatório.

Entretanto o grau mínimo exigido para a fidedignidade deve ser determinado de acordo com os objetivos e situações em que nos propomos usar o teste como um instrumento de medida. A sua magnitude deve corresponder a importância das decisões e dos efeitos que estas decisões venham ter sobre a população em questão.

4.3 - *Estimação do Coeficiente de Fidedignidade, considerando a Distribuição de Probabilidade dos Escores do Teste ou dos Subtestes*

Provavelmente o primeiro artigo com ênfase na estimação da fidedigniadde, baseado na distribuição de amostras pequenas, foi escrito por KRISTOF(1963). FELDT(1965) determinou a distribuição amostral do coeficiente de fidedignidade KR_{20}. Outros artigos que se destacam na área de estimação e testes paramétricos são devido a KRISTOF(1970,1972,1974).

Os resultados desta secção, baseiam-se nos artigos referidos acima.

Suposições:

Consideremos um teste dividido em K partes (sub-testes), $K \geq 2$, com mesma estrutura fatorial, cujos escores são utilizados como base para a estimação do coeficiente de fidedignidade.

Vamos supor duas situações:

Caso *1* - as K partes do teste possuem iguais variâncias e iguais covariâncias, mas nenhuma suposição é feita sobre as médias.

Caso 2 - as K partes do teste possuem iguais variâncias, iguais covariâncias e médias também iguais.

Em ambos os casos, suponhamos que os escores das k partes tem distribuição Normal Multivariada e que tomamos uma amostra aleatória de tamanho n.

Denotaremos os parâmetros populacionais por letras gregas e as correspondentes estatísticas amostrais por letras romanas.

A partir destas suposições determinaremos os estimadores de máxima verossimilhança dos parâmetros desconhecidos.

4.3.1 - *Estimadores de Máxima Verossimilhança*

Parâmetros Básicos

Seja $\underset{\sim}{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)$ o vetor aleatório dos escores das K partes de um teste com distribuição Normal Multivariada, com média $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)$ e matriz de variância-covariância R. Então a função densidade de $\underset{\sim}{\tau}$ é dada por

$$f(\underset{\sim}{\tau}) = (2\pi)^{-K/2} |R|^{-1/2} \exp\{-\frac{1}{2}(\underset{\sim}{\tau}-\underset{\sim}{\mu})R^{-1}(\underset{\sim}{\tau}-\underset{\sim}{\mu})'\}. \qquad (4.3.1)$$

Inicialmente determinemos R^{-1} e $|R|$ explicitamente.

Seja σ^2 a variância de cada parte e ρ, a correlação entre duas partes quaisquer. Então

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

Vamos escrever R^{-1} em termos de 2 parâmetros desconhecidos γ e β.

$$R^{-1} = \begin{bmatrix} \gamma & & \beta & \cdots & \beta \\ & \ddots & & \ddots & \vdots \\ \beta & & \ddots & & \beta \\ \vdots & \ddots & & \ddots & \\ \beta & \cdots & \beta & & \gamma \end{bmatrix}$$

Da condição $RR^{-1}=I$, onde I é uma matriz identidade com dimensão K, vem que

$$\gamma = \frac{1}{\sigma^2} \frac{1+(K-2)\rho}{1+(K-2)\rho-(K-1)\rho^2} \quad \text{e} \tag{4.3.2}$$

$$\beta = \frac{-1}{\sigma^2} \frac{\rho}{1+(K-2)\rho-(K-1)\rho^2} \tag{4.3.3}$$

Segue-se então que

$$|R| = \sigma^{2K} \begin{bmatrix} 1 & & \rho & \cdots & \rho \\ & \ddots & & \ddots & \vdots \\ \rho & & \ddots & & \rho \\ \vdots & \ddots & & \ddots & \\ \rho & \cdots & \rho & & 1 \end{bmatrix}$$

Após alguns cálculos obtemos que

$$|R| = \sigma^{2K}(1-\rho)^{K-1}[1+(K-1)\rho] \qquad (4.3.4)$$

e que

$$(\underset{\sim}{\tau}-\underset{\sim}{\mu})R^{-1}(\underset{\sim}{\tau}-\underset{\sim}{\mu})' = \gamma \sum_{j=1}^{K} (\tau_i-\mu_j)^2+\beta \sum_{j\neq K} (\tau_j-\mu_j)(\tau_k-\mu_k)$$

A equação (4.3.1) pode ser reescrita como

$$f(\underset{\sim}{\tau}) = (2\pi)^{-K/2}|R|^{-1/2} \exp\{-\frac{1}{2}[\gamma \sum_{j=1}^{K} (\tau_i-\mu_i)^2+\beta \sum_{j\neq K} (\tau_j-\mu_j)(\tau_k-\mu_k)]\}.$$

Suponhamos que dispomos dos escores dos n indivíduos da amostra nas K partes do teste, isto é, $Y^{(i)}=(Y_{1i},Y_{2i},\ldots,Y_{ki})$, $i=1,2,\ldots,n$. Obviamente, a função de verossimilhança L é dada por

$$L = (2\pi)^{-nK/2}|R|^{-n/2} \exp\{-\frac{1}{2}[\gamma \sum_{j=1}^{K} \sum_{i=1}^{n} (Y_{ji}-\mu_j)^2 +$$

$$+ \beta \sum_{j\neq K} \sum_{i=1}^{n} (Y_{ji}-\mu_j)(Y_{ki}-\mu_k)]\},$$

ou usando as abreviações

$$A = \sum_{j=1}^{K} \sum_{i=1}^{n} (Y_{ji}-\mu_j)^2 \qquad \text{e}$$

$$B = \sum_{j\neq K} \sum_{i=1}^{n} (Y_{ji}-\mu_j)(Y_{ki}-u_k),$$

podemos escrever

$$L = (2\pi)^{-nK/2}|R|^{-n/2} \exp\{-\frac{1}{2}[\gamma A+\beta B]\}.$$

Para a obtenção dos estimadores de máxima verossimilhança dos parâmetros, calculamos as derivadas parciais do logarítmo natural de L

$$\ln L = \frac{-nK}{2} \ln 2\pi - \frac{n}{2} \ln|R| - \frac{\gamma A}{2} - \frac{\beta B}{2}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{2} \frac{1}{|R|} \frac{\partial |R|}{\partial \sigma^2} - \frac{A}{2} \frac{\partial \gamma}{\partial \sigma^2} - \frac{B}{2} \frac{\partial \beta}{\partial \sigma^2}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{-n}{2} \frac{1}{|R|} \frac{\partial |R|}{\partial \rho} - \frac{A}{2} \frac{\partial \gamma}{\partial \rho} - \frac{B}{2} \frac{\partial \beta}{\partial \rho}.$$

Para o caso 1, onde nenhuma suposição é feita sobre μ_j, segue-se a seqüência de equações

$$\frac{\partial \ln L}{\partial \mu_j} = \frac{-\gamma}{2} \frac{\partial A}{\partial \mu_j} - \frac{\beta}{2} \frac{\partial B}{\partial \mu_j}, \; j=1,2,\ldots,K. \qquad (4.3.5)$$

No caso 2, como todas as médias são iguais, podemos substituí-las por μ_0, produzindo ao invés de (4.2.5), a equação

$$\frac{\partial \ln L}{\partial \mu_0} = \frac{-\gamma}{2} \frac{\partial A}{\partial \mu_0} - \frac{\beta}{2} \frac{\partial B}{\partial \mu_0} .$$

Tornando todas estas derivações iguais a zero, segue-se o sistema de equações de máxima verossimilhança

$$\frac{n}{|R|} \frac{\partial |R|}{\partial \sigma^2} + A \frac{\partial \gamma}{\partial \sigma^2} + B \frac{\partial \beta}{\partial \sigma^2} = 0,$$

$$\frac{n}{|R|} \frac{\partial |R|}{\partial \rho} + A \frac{\partial \gamma}{\partial \rho} + B \frac{\partial \beta}{\partial \rho} = 0.$$

No caso 1, obtemos as K equações adicionais

$$\gamma \frac{\partial A}{\partial u_j} + \beta \frac{\partial B}{\partial \mu_j} = 0, \; j=1,2,\ldots,K , \qquad (4.3.6)$$

enquanto que no caso 2, obtemos uma única equação adicional

$$\gamma \frac{\partial A}{\partial u_0} + \beta \frac{\partial B}{\partial \mu_0} = 0 \qquad (4.3.7)$$

As soluções destes sistemas de equações de máxima verossimilhança são os estimadores de máxima verossimilhança, que denotaremos por $(\wedge)$.

Vamos determinar inicialmente os estimadores de máxima verossimilhança para as médias.

No caso 1, temos

$$\frac{\partial A}{\partial \mu_j} = -2 \sum_{i=1}^{n} (Y_{ji}-\mu_j)$$

$$\frac{\partial B}{\partial \mu_j} = -2 \sum_{K\neq j} \sum_{i=1}^{n} (Y_{ki}-\mu_k).$$

As equações (4.2.6) são equivalentes ao sistema de equações

$$\hat{\gamma} \sum_i (Y_{1i}-\hat{\mu}_1)+\hat{\beta} \sum_i (Y_{2i}-\hat{\mu}_2)+\ldots+\hat{\beta} \sum_i (Y_{ki}-\hat{\mu}_k) = 0,$$

$$\hat{\beta} \sum_i (Y_{1i}-\hat{\mu}_1)+\hat{\gamma} \sum_i (Y_{2i}-\hat{\mu}_2)+ \hat{\beta} \sum_i (Y_{3i}-\hat{\mu}_3)+\ldots+\hat{\beta} \sum_i (Y_{ki}-\hat{\mu}_k) = 0,$$

$$\cdots$$

$$\hat{\beta} \sum_i (Y_{1i}-\hat{\mu}_1)+\ldots+\hat{\beta} \sum_i (Y_{k-1,i}-\hat{\mu}_{k-1})+\hat{\gamma} \sum_i (Y_{ki}-\hat{\mu}_k) = 0 .$$

Este sistema de K equações homogêneas é de posto K e, por conseguinte, possuidor da solução trivial

$$\sum_i (Y_{1i}-\hat{\mu}_1) = \ldots. = \sum_i (Y_{ki}-\hat{\mu}_k) = 0,$$

da qual se infere imediatamente que

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} Y_{ji} = \overline{X}_j \quad , \qquad j = 1,2,\ldots,K.$$

No caso 2, temos

$$\frac{\partial A}{\partial \mu_0} = -2 \sum_{j=1}^{K} \sum_{i=1}^{n} (Y_{ji}-\mu_0),$$

$$\frac{\partial B}{\partial \mu_0} = -2(K-1) \sum_{j=1}^{K} \sum_{i=1}^{n} (Y_{ji}-\mu_0).$$

A equação (4.3.7) produz então

$$[\hat{\gamma}+(K-1)\hat{\beta}] \sum_{j=1}^{K} \sum_{i=1}^{n} (Y_{ji}-\hat{\mu}_0) = 0,$$

para o qual

$$\hat{\mu}_0 = \frac{1}{nK} \sum_{j=1}^{K} \sum_{i=1}^{n} Y_{ji}$$

é inferido imediatamente.

Portanto, no caso 1, o estimador de máxima verossimilhança da média de cada parte do teste é a média aritmética dos escores desta parte, enquanto que no caso 2, o estimador da média de qualquer parte é a média aritmética de todos os escores do teste.

Analogamente a obtenção dos estimadores das médias, obtemos que o estimador de máxima verossimilhança para ρ é dado por

$$\hat{\rho}' = \frac{B}{(K-1)A}, \tag{4.3.8}$$

tal que os estimadores para μ_1, $\mu_2, \ldots,$ μ_k e μ_0 estão inseridos em A e B.

Mais precisamente, no caso 1, temos que

$$\hat{\rho}' = \frac{1}{K-1} \frac{\sum\sum_{j\neq \ell} S_{j\ell}}{\sum^{K} S^2}, \tag{4.3.9}$$

onde

$$S_{j\ell} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_{ji}-\overline{Y}_j)(Y_{\ell i}-\overline{Y}_\ell). \qquad (4.3.10)$$

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_{ji}-\overline{Y}_j)^2. \qquad (4.3.11)$$

E no caso 2 obtemos

$$\hat{\rho}' = \frac{1}{K-1} \frac{\sum_{j\neq\ell} \sum_{i=1}^{n} Y_{ji}Y_{\ell i} - [(K-1)/Kn]\left(\sum_{j=1}^{K} \sum_{i=1}^{n} Y_{ji}\right)^2}{\sum_{i=1}^{K} \sum_{j=1}^{n} Y_{ij}^2 - (1/Kn)\left(\sum_{j=1}^{K} \sum_{i=1}^{n} Y_{ji}\right)^2}. \qquad (4.3.12)$$

O estimador da variância $\hat{\sigma}^2$ não será determinado, pois não será usado neste trabalho.

O Coeficiente de Fidedignidade do Teste

O estimador do coeficiente de fidedignidade do teste completo será denotado por $\hat{\rho}$ e seu valor verdadeiro por ρ. Distinguiremos os estimadores para os dois casos considerados, utilizando os índices 1 e 2.

Usando a propriedade da invariância dos estimadores de máxima verossimilhança e a fórmula de Spearman-Brown, segue-se que o estimador de máxima verossimilhança da fidedignidade de um teste formado por K partes é dado por

$$\hat{\rho} = \frac{K\hat{\rho}'}{1+(K-1)\hat{\rho}'}. \qquad (4.3.13)$$

De (4.3.9) e algumas manipulações algébricas, segue-se para o caso 1 que

$$\hat{\rho}_1 = \frac{K}{K-1}\left[1-\frac{\sum_{j=1}^{K} S_j^2}{S_X^2}\right], \qquad (4.3.14)$$

onde

$$S_X^2 = \sum_{j=1}^{K} S_j^2 + \sum\sum_{j\neq \ell} S_{j\ell}. \qquad (4.3.15)$$

Naturalmente que S_X^2 é a variância dos escores totais $\sum_{j=1}^{K} Y_{ji}$, i=1,2,...,n. A equação (4.3.14) coincide com a definição do coeficiente de fidedignidade α de CRONBACH(1951), que a partir de agora passa a ser o estimador de máxima verossimilhança de fidedignidade de um teste, sob as suposições do caso 1.

Combinando (4.3.12) e (4.3.13), obtemos depois de alguns cálculos, o resultado para o caso 2,

$$\hat{\rho}_2 = \frac{K}{K-1}\left[1 - \frac{Kn-1}{n-1}\,\frac{S_X'^2}{S_X^2}\right], \qquad (4.3.16)$$

onde $S_X'^2$ é definido por

$$S_X'^2 = \frac{1}{Kn-1}\left[\sum_{j=1}^{K}\sum_{i=1}^{n}(Y_{ji})^2 - \frac{1}{Kn}\left(\sum_{j=1}^{K}\sum_{i=1}^{n}Y_{ji}\right)^2\right] \qquad (4.3.17)$$

Portanto $S_X'^2$ é a variância global sobre todas as partes do teste.

Comparando as expressões (4.3.14) e (4.3.16), segue-se que $\hat{\rho}_1 \geqq \hat{\rho}_2$. Portanto o valor do estimador para a fidedignidade é maior quando não supomos que as médias das K partes são iguais.

Quando K=2, isto é, quando um teste é dividido em 2 partes, os estimadores de máxima verossimilhança são dados por

$$\hat{\rho}_1 = 2\left[1 - \frac{\sum_{j=1}^{2} S_j^2}{S_X^2}\right],$$

tal que $S_X^2 = \sum_{j=1}^{2} S_j^2 + 2S_{1,2}$, e

$$\hat{\rho}_2 = 2[1 - \frac{2n-1}{n-1}\frac{S'^2_X}{S^2_X}],$$

tal que $S'^2_X = \frac{1}{2n-1}[\sum_{j=1}^{2}\sum_{i=1}^{n}(Y_{ji})^2 - \frac{1}{2n}(\sum_{j=1}^{2}\sum_{i=1}^{n}Y_{ji})^2]$.

4.3.2 - *Distribuições Amostrais dos Estimadores de Máxima Verossimilhança*

KRISTOF(1963) derivou as distribuições amostrais dos estimadores de máxima verossimilhança para a fidedignidade de um teste dividido em K partes. Devido a complexidade matemática envolvida nesta derivação, optamos pela apresentação somente dos principais resultados, tanto para o caso 1 como para o caso 2.

Caso 1

De KRISTOF(1963), as quantidades

$$\chi^2_{n-1} = \frac{n-1}{K}\frac{S^2_X}{\sigma^2[1+(K-1)\rho]}, \tag{4.3.18}$$

onde S^2_X é definida por (4.3.15), tem distribuição χ^2 com n-1 graus de liberdade, e

$$\chi^2_{(K-1)(n-1)} = \frac{n-1}{\sigma^2(1-\rho)}(S^2_1 + \dots + S^2_k - \frac{1}{K}S^2_X), \tag{4.3.19}$$

onde S^2_j é definida por (4.3.11), tem distribuição χ^2 com (K-1)(n-1) graus de liberdade.

Uma vez que (4.3.18) e (4.3.19) são independentes, segue-se que a razão

$$F_{(n-1),(K-1)(n-1)} = \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{(K-1)(n-1)}/(K-1)(n-1)}$$

tem distribuição F com n-1 e (K-1)(n-1) graus de liberdade. Esta expressão é equivalente a

$$F_{(n-1),(K-1)(n-1)} = \frac{1-\rho}{1-\hat{\rho}_1}, \tag{4.3.20}$$

onde ρ é o coeficiente de fidedignidade do teste completo e $\hat{\rho}_1$ é o estimador de máxima verossimilhança de ρ, definido por (4.3.14). Portanto a distribuição amostral de $\hat{\rho}_1$, isto é, do coeficiente α de Cronbach é conhecida.

Evidentemente, a inversa de (4.3.20),

$$F_{(K-1)(n-1),(n-1)} = \frac{1-\hat{\rho}_1}{1-\rho}$$

tem distribuição F com (K-1)(n-1) e (n-1) graus de liberdade.

Caso 2

Analogamente ao caso 1, KRISTOF(1963) obteve que a quantidade

$$\chi^2_{n-1} = \frac{n-1}{K}\,\frac{S_X^2}{\sigma^2[1+(K-1)\rho]}, \tag{4.3.21}$$

tem distribuição χ^2 com n-1 graus de liberdade.

Obteve também que

$$\chi^2_{n(K-1)} = \frac{1}{\sigma^2(1-\rho)}\left[(Kn-1)S_X'^2 - \frac{n-1}{K}S_X^2\right], \tag{4.3.22}$$

com S_X^2 e $S_X'^2$ definidos por (4.3.15) e (4.3.17), tem distribuição χ^2 com n(K-1) graus de liberdade.

Da independência entre (4.3.21) e (4.3.22) segue-se que a razão

$$F_{(n-1),n(K-1)} = \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{n(K-1)}/n(K-1)}$$

tem distribuição F com n-1 e n(K-1) graus de liberdade. Esta expressão é equivalente a

$$F_{(n-1),n(K-1)} = \frac{n}{n-1}\,\frac{1-\rho}{1-\hat{\rho}_2}, \tag{4.3.23}$$

onde ρ é o coeficiente de fidedignidade do teste completo e $\hat{\rho}_2$ é o teste estimador de máxima verossimilhança de ρ, definido por (4.3.16).

Evidentemente, a inversa de (4.3.23),

$$F_{n(K-1),(n-1)} = \frac{1-\hat{\rho}_2}{1-\rho}$$

tem distribuição F com n(K-1) e n-1 graus de liberdade.

4.3.3 - *Estimadores de Máxima Verossimilhança Não Viciados*

Verifiquemos inicialmente se os estimadores de máxima verossimilhança, $\hat{\rho}_1$ e $\hat{\rho}_2$, são não-viciados (não-tendenciosos).

Caso 1

De (4.3.21) segue-se que

$$\hat{\rho}_1 = 1-(1-\rho)F_{(K-1)(n-1),(n-1)}.$$

Portanto o valor esperado de $\hat{\rho}_1$,

$$E\hat{\rho}_1) = 1-(1-\rho)E(F_{(K-1)(n-1),(n-1)}).$$

De MOOD-GRAYBILL-BOES(1974-pág.542), $E(F_{(K-1)(n-1),(n-1)}) = \frac{n-1}{n-3}$.

Segue-se então que

$$E(\hat{\rho}_1) = \frac{-2}{n-3} + \frac{n-1}{n-3}\rho, \qquad (4.3.24)$$

e portanto, exceto para $\rho=1$, $\hat{\rho}_1$ é um estimador viciado para ρ.

O vício de $\hat{\rho}_1$ é dado por

$$E(\hat{\rho}_1)-\rho = \frac{-2}{n-3}(1-\rho) \leq 0. \qquad (4.3.25)$$

Assim $\hat{\rho}_1$ tende subestimar ρ, independente de K.

O vício de $\hat{\rho}_1$ nos leva a determinar um estimador não viciado $\hat{\hat{\rho}}_1$ para ρ. Podemos definí-lo como uma função linear de $\hat{\rho}_1$.

Sejam as constantes a_1 e a_2, então podemos escrever

$$\hat{\hat{\rho}}_1 = a_1+a_2\hat{\rho}_1$$

sob a condição: $E(\hat{\hat{\rho}}_1) = a_1+a_2E(\hat{\rho}_1) = \rho$.

Usando (4.3.24), obtemos que

$$a_1(n-3)-2a_2+[a_2(n-1)-(n-3)]\rho = 0.$$

Uma vez que o membro da direita deve ser nulo, concluímos que

$$a_1(n-3)-2a_2 = 0 \quad \text{e}$$

$$a_2(n-1)-(n-3) = 0.$$

Resolvendo estas equações, vem que

$$a_1 = \frac{2}{n-1} \quad \text{e} \quad a_2 = \frac{n-3}{n-1}.$$

Segue-se então que o estimador de máxima verossimilhança não viciado para ρ é definido por

$$\hat{\hat{\rho}}_1 = \frac{2}{n-1} + \frac{n-3}{n-1}\hat{\rho}_1, \tag{4.3.26}$$

ou substituindo $\hat{\rho}_1$,

$$\hat{\hat{\rho}}_1 = \frac{2}{n-1} + \frac{n-3}{n-1}\,\frac{K}{K-1}\left[1-\frac{\sum_{j=1}^{K} S_j^2}{S_X^2}\right]. \tag{4.3.27}$$

Caso 2

De (4.3.23), obtemos que

$$\hat{\rho}_2 = 1-\frac{n}{n-1}(1-\rho)\frac{1}{F}.$$

Analogamente ao caso 1, segue-se que

$$E(\hat{\rho}_2) = \frac{-3}{n-3} + \frac{n}{n-3}\rho. \tag{4.3.28}$$

Exceto para $\rho=1$, $\hat{\rho}_2$ é um estimador viciado para ρ, cujo vício é dado por

$$E(\hat{\rho}_2)-\rho = \frac{-3}{n-3}(1-\rho) \leqq 0. \qquad (4.3.29)$$

Portanto, $\hat{\rho}_2$ tende subestimar ρ, independente de K. Comparando (4.3.25) e (4.3.29), conclui-se que o vício no caso 2 é um pouco maior que no caso 1.

Analogamente ao caso 1, determinamos com o emprego de (4.3.28) que o estimador de máxima verossimilhança não viciado para ρ é dado por

$$\hat{\hat{\rho}}_2 = \frac{3}{n} + \frac{n-3}{n}\hat{\rho}_2. \qquad (4.3.30)$$

Substituindo $\hat{\rho}_2$, segue-se

$$\hat{\hat{\rho}}_2 = \frac{3}{n} + \frac{n-3}{n}\frac{K}{K-1}\left[1-\frac{Kn-1}{n-1}\frac{S_X'^2}{S_X^2}\right]. \qquad (4.3.31)$$

4.3.4 - *Comparação entre os Estimadores de Máxima Verossimilhança em Termos dos Erros Quadráticos Médios*

Obtivemos, na secção anterior, os estimadores de máxima verossimilhança não viciados para a fidedignidade de um teste. Esta propriedade (ser não viciado) é um dos critérios para a escolha de um estimador para um parâmetro. No entanto, o fato de um estimador ser não viciado, por si só não representa uma qualidade apreciável,pois se ele apresentar uma variância grande, a probabilidade de observarmos um valor $\hat{\hat{\rho}}$ que difira substancialmente de ρ, também é grande. Se ao contrário, a variância de $\hat{\hat{\rho}}$ for pequena, então os seus valores estarão concentrados em torno de ρ, que é uma qualidade desejável.

Portanto torna-se indispensável considerarmos conjunta mente o valor esperado e a variância dos estimadores de máxima verossimilhança para ρ, a fim de determinarmos o melhor estimador, de acordo com este critério. Este objetivo será atingido através da comparação dos estimadores $\hat{\rho}$ e $\hat{\hat{\rho}}$ de ρ, em termos de seus erros quadráticos médios.

O erro quadrático médio de um estimador, por exemplo, $\hat{\rho}$ é definido como a média dos desvios ao quadrado de $\hat{\rho}$ em relação a ρ, isto é,

$$EQM(\hat{\rho}) = E(\hat{\rho}-\rho)^2 = \sigma^2(\hat{\rho}) + (E(\hat{\rho})-\rho)^2.$$

A seguir, vamos determinar os erros quadráticos médios dos estimadores $\hat{\rho}$ e $\hat{\hat{\rho}}$ de ρ, para ambos os casos, e relacioná-los em função desta medida.

Caso 1: A partir de (4.3.20), obtemos

$$\sigma^2(F_{(K-1)(n-1),(n-1)}) = \frac{1}{(1-\rho)^2}\,\sigma^2(\hat{\rho}_1).$$

De MOOD-GRAYBILL-BOES(1974,pág542), segue-se que

$$\sigma^2(\hat{\rho}_1) = \frac{2(1-\rho)^2(n-1)[K(n-1)-2]}{(K-1)(n-3)^2(n-5)}, \tag{4.3.32}$$

para $n > 5$.

Da expressão (4.3.26), obtemos

$$\sigma^2(\hat{\hat{\rho}}_1) = \left(\frac{n-3}{n-1}\right)^2 \sigma^2(\hat{\rho}_1). \tag{4.3.33}$$

Deste resultado e de (4.3.32), segue-se então que

$$\sigma^2(\hat{\hat{\rho}}_1) = \frac{2(1-\rho)^2[K(n-1)-2]}{(K-1)(n-1)(n-5)}, \tag{4.3.34}$$

para $n > 5$.

Uma vez que $\hat{\hat{\rho}}_1$ é não viciado para ρ, o $EQM(\hat{\hat{\rho}}_1)$ é igual a $\sigma^2(\hat{\hat{\rho}}_1)$, Além disto, $\sigma^2(\hat{\hat{\rho}}_1) < \sigma^2(\hat{\rho}_1)$, pois $n-3 < n-1$. Segue-se então que

$$EQM(\hat{\hat{\rho}}_1) < EQM(\hat{\rho}_1). \qquad (4.3.35)$$

Este é um resultado favorável para $\hat{\hat{\rho}}_1$, pois além dele ser não viciado, sua variância também é menor que a de $\hat{\rho}_1$. Logo de acordo com o critério do EQM, $\hat{\hat{\rho}}_1$ é melhor do que $\hat{\rho}_1$.

Caso 2: A partir de (4.3.23) temos

$$\sigma^2(F_{n(K-1),(n-1)}) = (\frac{n-1}{n})^2 \frac{1}{(1-\rho)^2} \sigma^2(\hat{\rho}_2).$$

De MOOD-GRAYBILL-BOES(1974,pág.542), segue-se que

$$\sigma^2(\hat{\rho}_2) = \frac{2(1-\rho)^2 n(nK-3)}{(K-1)(n-3)^2(n-5)}, \qquad (4.3.36)$$

para $n > 5$.

Da expressão (4.3.30), obtemos

$$\sigma^2(\hat{\hat{\rho}}_2) = (\frac{n-3}{n})^2 \sigma^2(\hat{\rho}_2). \qquad (4.3.37)$$

A partir deste resultado e de (4.3.36), segue-se então que

$$\sigma^2(\hat{\hat{\rho}}_2) = \frac{2(1-\rho)^2(nK-3)}{n(K-1)(n-5)}, \qquad (4.3.38)$$

para $n > 5$.

Uma vez que $n-3$ é menor que n, de (4.3.37) segue-se que $\sigma^2(\hat{\hat{\rho}}_2) < \sigma^2(\hat{\rho}_2)$. Além disto, $\hat{\hat{\rho}}_2$ é não viciado para ρ, portanto

$$EQM(\hat{\hat{\rho}}_2) < EQM(\hat{\rho}_2). \qquad (4.3.39)$$

A desigualdade (4.3.39) reflete que devemos escolher $\hat{\hat{\rho}}_2$ para estimar ρ, ao invés de $\hat{\rho}_2$, de acordo com o critério de EQM.

Comparação entre $\hat{\rho}_1$ *e* $\hat{\rho}_2$

Utilizando (4.3.25) e (4.3.29) para compararmos os vícios ao quadrado de $\hat{\rho}_1$ e $\hat{\rho}_2$, obtemos

$$[E(\hat{\rho}_1)-\rho]^2 < [E(\hat{\rho}_2)-\rho]^2. \qquad (4.3.40)$$

De (4.3.32) e (4.3.36), obtemos

$$\frac{\sigma^2(\hat{\rho}_1)}{\sigma^2(\hat{\rho}_2)} = \frac{(n-1)[K(n-1)-2]}{n(nK-3)} < 1,$$

isto é,

$$\sigma^2(\hat{\rho}_1) < \sigma^2(\hat{\rho}_2).$$

Desta desigualdade e de (4.3.40), segue-se então que

$$EQM(\hat{\rho}_1) < EQM(\hat{\rho}_2). \qquad (4.3.41)$$

Estes resultados indicam que não somente o vício de $\hat{\rho}_1$ é menor que $\hat{\rho}_2$, mas que $\hat{\rho}_1$ tem também uma variância menor que $\hat{\rho}_2$. Isto implica que é vantajoso não considerar a informação que as K partes de um teste tem médias iguais, sempre que o caso 2 ocorrer, visando a obtenção de uma estimativa mais precisa.

Comparação entre $\hat{\hat{\rho}}_1$ *e* $\hat{\hat{\rho}}_2$

A comparação entre $\hat{\hat{\rho}}_1$ e $\hat{\hat{\rho}}_2$ em termos do erro quadrático médio é equivalente a comparação em termos da variância, uma vez que ambos são não viciados. Em função disto, consideremos a razão

$$\frac{\sigma^2(\hat{\hat{\rho}}_1)}{\sigma^2(\hat{\hat{\rho}}_2)} = \frac{n[K(n-1)-2]}{(n-1)(nK-3)} > 1.$$

Portanto

$$\sigma^2(\hat{\hat{\rho}}_1) > \sigma^2(\hat{\hat{\rho}}_2)$$

que é equivalente a

$$EQM(\hat{\hat{\rho}}_1) > EQM(\hat{\hat{\rho}}_2). \qquad (4.3.42)$$

Logo quando pudermos supor que as K partes de um teste tem médias iguais, $\hat{\rho}_2$ deve ser usado ao invés de $\hat{\rho}_1$, a fim de obtermos um estimador mais preciso.

Conclusão:

A partir das expressões (4.3.35), (4.3.39), (4.3.41) e (4.3.42) obtemos a seguinte relação

$$EQM(\hat{\hat{\rho}}_2) < EQM(\hat{\hat{\rho}}_1) < EQM(\hat{\rho}_1) < EQM(\hat{\rho}_2),$$

para todo $n > 5$ e $\rho < 1$.

Este resultado implica que se o critério para a escolha do estimador para a fidedignidade é o erro quadrático médio e não pudermos supor que as médias das K partes são iguais, devemos preferir $\hat{\hat{\rho}}_1$ ao invés de $\hat{\rho}_1$; porém, se pudermos supor que as médias são iguais deve-se usar $\hat{\hat{\rho}}_2$ para garantir maior precisão.

Vamos verificar a magnitude da superioridade de $\hat{\hat{\rho}}_2$ em relação a $\hat{\hat{\rho}}_1$, através da tabela a seguir.

Razões $EQM(\hat{\hat{\rho}}_1)/EQM(\hat{\hat{\rho}}_2)$

K \ n	6	10	20	50	100
2	1,0667	1,0458	1,0242	1,0099	1,0050
3	1,0400	1,0288	1,0157	1,0065	1,0033
5	1,0222	1,0165	1,0092	1,0039	1,0020
10	1,0105	1,0080	1,0045	1,0019	1,0010
20	1,0051	1,0039	1,0022	1,0010	1,0005

Esta tabela reflete que a medida que K aumenta, fixando n, a diferença entre $\hat{\hat{\rho}}_1$ e $\hat{\hat{\rho}}_2$ em termos de precisão (EQM) diminui, o mesmo ocorrendo quando fixamos K e aumentamos o tamanho da amostra.

Analisando as razões $EQM(\hat{\hat{\rho}}_1)/EQM(\hat{\hat{\rho}}_2)$, para qualquer K e n, pode-se desprezar a superioridade de $\hat{\hat{\rho}}_2$ em relação a $\hat{\hat{\rho}}_1$, e em vista disto, utilizar $\hat{\hat{\rho}}_1$ para estimar ρ, em função da vantagem de não termos que ter um teste com as médias das K partes obrigatoriamente iguais.

Verifiquemos agora a magnitude da superioridade da precisão da $\hat{\hat{\rho}}_1$ em relção a $\hat{\rho}_1$(o coeficiente α de Cronbach).

Usando (4.3.24) e (4.3.32), obtemos

$$EQM(\hat{\rho}_1) = \frac{2(1-\rho)^2[K(n^2-9)-4n+12]}{(K-1)(n-3)^2(n-5)} .$$

Da expressão (4.3.34), temos

$$EQM(\hat{\hat{\rho}}_1) = \frac{2(1-\rho)^2[K(n-1)-2]}{(K-1)(n-1)(n-5)} .$$

Portanto a razão

$$\frac{EQM(\hat{\rho}_1)}{EQM(\hat{\hat{\rho}}_1)} = \frac{(n-1)[K(n^2-9)-4n+12]}{(n-3)^2[K(n-1)-2]} > 1.$$

A seguir apresentaremos uma tabela das razões $EQM(\hat{\rho}_1)/$ $/EQM(\hat{\hat{\rho}}_2)$, a fim de analisarmos o seu comportamento para vários valores de K e n.

Razões: $EQM(\hat{\rho}_1)/EQM(\hat{\hat{\rho}}_1)$

K \ n	6	10	20	50	100
2	2,9166	1,7678	1,3039	1,1077	1,0519
3	2,9487	1,8000	1,3208	1,1144	1,0552
5	2,9710	1,8239	1,3340	1,1198	1,0579
10	2,9861	1,8409	1,3436	1,1237	1,0599
20	2,9932	1,8449	1,3483	1,1257	1,0609

Estes resultados permitem concluir que para amostras pequenas, $\hat{\rho}_1$ é geralmente muito inferior a $\hat{\hat{\rho}}_1$, inferioridade que aumenta quando um teste é dividido num maior número de partes. No entanto, esta diferença diminui a medida que o tamanho da amostra aumenta, e para $n \geqq 50$, a precisão de $\hat{\rho}_1$ é praticamente igual a $\hat{\hat{\rho}}_1$, para qualquer K.

O Efeito da Divisão de um Teste na
Precisão da Estimação da Fidedignidade ρ.

Investiguemos agora se existe vantagem em dividir um teste em muitas partes, quando o estimador de máxima verossimilhança não viciado $\hat{\hat{\rho}}$ for utilizado.

Convém salientar que a quantidade de vício do estimador $\hat{\rho}$ não depende do número de divisões K de um teste (ver(4.3.25) e (4.3.27)), e assim não pode diminuir aumentando K.

Caso 1 - A expressão (4.3.34) indica que

$$\sigma^2(\hat{\hat{\rho}}_1) = \frac{2(1-\rho)^2}{(n-1)(n-5)} \frac{K(n-1)-2}{K-1} .$$

Derivando $\sigma^2(\hat{\hat{\rho}}_1)$ em relação a K, obtemos

$$\frac{\partial\sigma^2(\hat{\hat{\rho}}_1)}{\partial K} = \frac{2(1-\rho)^2}{(n-1)(n-5)} \frac{3-n}{(K-1)^2} < 0$$

para $n > 5$. Como o valor da derivada é negativo, concluimos que $\sigma^2(\hat{\hat{\rho}}_1)$ diminui a medida que K aumenta. Portanto é conveniente dividir um teste em várias partes, quando $\hat{\hat{\rho}}_1$ é usado para estimar o coeficiente de fidedignidade ρ.

Ilustremos este resultado com um exemplo do ganho que ocorre em precisão, quando dividimos um teste em 4 partes ao invés de em 2 partes.

Temos

$$\frac{\sigma^2(\hat{\hat{\rho}}_1) \text{ para } K=4}{\sigma^2(\hat{\hat{\rho}}_1) \text{ para } K=2} = \frac{2}{3}.$$

Portanto, $\frac{2}{3}$ é o fator que a variância de $\hat{\hat{\rho}}_1$ decresce quando um teste é dividido em 4 partes ao invés de em 2.

Caso 2 - A expressão (4.3.38) indica que

$$\sigma^2(\hat{\hat{\rho}}_2) = \frac{2(1-\rho)^2}{n(n-5)} \frac{(nK-3)}{K-1}.$$

Derivando $\sigma^2(\hat{\hat{\rho}}_2)$ em relação a K, obtemos

$$\frac{\partial\sigma^2(\hat{\hat{\rho}}_2)}{\partial K} = \frac{2(1-\rho)^2}{n(n-5)} \frac{3-n}{(K-1)^2} < 0$$

para n > 5. Concluímos que $\sigma^2(\hat{\hat{\rho}}_2)$ decresce a medida que dividimos um teste num maior número de partes.

Ilustramos este resultado, verificando o aumento que ocorre na precisão, quando dividimos um teste em 4 partes ao invés de 2, e $\hat{\hat{\rho}}_2$ é usado para estimar ρ.

$$\frac{\sigma^2(\hat{\hat{\rho}}_2) \text{ para } K=4}{\sigma^2(\hat{\hat{\rho}}_2) \text{ para } K=2} = \frac{1}{3}\left(\frac{4n-3}{2n-3}\right).$$

Portanto, quando n é grande, 2/3 é o fator para o qual a variância de $\hat{\hat{\rho}}_2$ decresce quando o número de partes vai de 2 para 4, obtendo-se um ganho em precisão neste aumento.

4.3.5 - *Testes de Hipóteses e Intervalos de Confiança*

A partir da análise das suposições estabelecidas nas derivações dos estimadores de máxima verossimilhança e da viabilidade de obtê-las na prática, das propriedades destes estimado -

res, dos resultados obtidos nas comparações realizadas na secção anterior e da familiaridade que os pesquisadores que usam testes como um instrumento de medida tem com o coeficiente α de Cronbach decidimos apresentar a dedução dos testes de hipóteses e intervalos de confiança para o coeficiente de fidedignidade de um teste quando usamos os estimadores $\hat{\rho}_1$ (o coeficiente α de Cronbach) e $\hat{\hat{\rho}}_1$ (o estimador de máxima verossimilhança não viciado).

As suposições sob as quais realizaremos os testes de hipóteses e intervalos de confiança continuam sendo as apresentadas na secção (4.3.1).

Testes de Hipóteses para o Coeficiente de Fidedignidade usando o Coeficiente α de Cronbach

Teste da Hipótese H_0: $\rho \leqq \rho_0$ versus H_1: $\rho > \rho_0$

A partir da teoria de testes de hipóteses, temos que a região crítica deste teste é dada por

$$C = \{A \in R^k \times R^n \text{ tal que } \hat{\rho}_1 > \rho_c\} \quad ,$$

onde A é a matriz dos dados amostrais e ρ_c é determinado pela condição:

$$P(\hat{\rho}_1 > \rho_c / H_0 \text{ é verdadeiro}) = \alpha$$

sendo α o nível de significância do teste.

Supondo que H_0 é verdadeiro, obtemos

$$P(\hat{\rho}_1 > \rho_c) = P\left(\frac{1-\hat{\rho}_1}{1-\rho_0} < \frac{1-\rho_c}{1-\rho_0}\right) = \alpha \quad .$$

De (4.3.20), temos

$$F_{(K-1)(n-1),(n-1)} = \frac{1-\hat{\rho}_1}{1-\rho_0} \quad , \text{ quando } \rho = \rho_0 \quad .$$

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

Segue-se então que a quantidade $\frac{1-\rho_c}{1-\rho_0}$ é igual ao correspondente valor tabelado da

$$F_{(k-1)(n-1),(n-1);\alpha'}$$

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 ao valor ρ_0 com nível de significância α, se

$$\hat{\rho}_1 > 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);}. \qquad (4.3.43)$$

Teste da Hipótese H_0: $\rho \geqq \rho_0$ versus H_1: $\rho < \rho_0$

Usando a teoria de testes de hipóteses, temos que a região crítica deste teste é dada por

$$C = \{A \in R^k \times R^n \text{ tal que } \hat{\rho}_1 < \rho_0\}$$

onde A é a matriz dos dados amostrais e ρ_c é determinado pela condição:

$$P(\hat{\rho}_1 < \rho_c/H_0 \text{ é verdadeiro}) = \alpha \quad .$$

Supondo que H_0 é verdadeiro, obtemos

$$P(\hat{\rho}_1 < \rho_c) = P(\frac{1-\hat{\rho}_1}{1-\rho_0} > \frac{1-\rho_c}{1-\rho_0}) = \alpha \quad .$$

Analogamente ao desenvolvimento do teste de hipótese anterior, obtemos que a quantidade

$\frac{1-\rho_c}{1-\rho_0}$ é igual ao correspondente valor tabelado da $F_{(K-1)(n-1),(n-1);1-\alpha}$, isto é, que

$$\rho_c = 1-(1-\rho_0)F_{(K-1)(n-1);1-\alpha}.$$

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

$$\hat{\rho}_1 < 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);1-\alpha}. \quad (4.3.44)$$

Teste da Hipótese H_0: $\rho = \rho_0$ versus H_1: $\rho \neq \rho_0$

Analogamente as deduções dos dois testes de hipóteses anteriores, obtemos que a região crítica C, neste caso, é dada por

$$C = \{A \in R^k \times R^n \text{ tal que } \hat{\rho}_1 < \rho_{c1} \text{ ou } \hat{\rho}_1 > \rho_{c2}\}$$

onde A é a matriz dos dados amostrais e ρ_{c1} e ρ_{c2} são obtidos a partir da igualdade, $P(\hat{\rho}_1 < \rho_{c1} \mid H_0 \text{ é V}) = P(\hat{\rho}_1 > \rho_{c2} \mid H_0 \text{ é V}) = \frac{\alpha}{2}$.

Novamente, usando os resultados anteriores, concluímos facilmente que rejeitamos H_0, isto é, que $\rho = \rho_0$, ao nível de significância α se

$$\hat{\rho}_1 < 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);1-\frac{\alpha}{2}} \text{ ou}$$
$$\hat{\rho}_1 < 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);\frac{\alpha}{2}} . \quad (4.3.45)$$

Intervalo de Confiança para o Coeficiente de Fidedignidade, usando o Coeficiente α de Cronbach

Baseados na dualidade intrínseca entre testes de hipóteses e intervalos de confiança, todos os resultados obtidos anteriormente podem ser utilizados para produzir intervalos de confiança para ρ, com qualquer coeficiente de confiança, vamos determinar um dos possíveis intervalos de confiança.

De (4.3.20), obtemos que

$$P(F_{(n-1),(K-1)(n-1),\frac{\alpha}{2}} < \frac{1-\rho}{1-\hat{\rho}_1} < F_{(n-1),(K-1)(n-1);1-\frac{\alpha}{2}}) = 1-\alpha.$$

Após alguns cálculos obtemos que um intervalo de confianca para a fidedignidade de um teste com coeficiente de confiança $1-\alpha$ é dado por

$$(1-(1-\hat{\rho}_1)F_{(n-1),(K-1)(n-1);1-\frac{\alpha}{2}}\ ;\ 1-(1-\hat{\rho}_1)F_{(n-1),(K-1)(n-1);\frac{\alpha}{2}}). \qquad (4.3.46)$$

Testes de Hipóteses para o Coeficiente de Fidedignidade, usando o Estimador de Máxima Verossimilhança $\hat{\hat{\rho}}_1$.

Apresentaremos somente os principais resultados relativos ao teste de hipóteses e intervalos de confiança para o coeficiente de fidedignidade, usando o estimador $\hat{\hat{\rho}}_1$, pois a obtenção de todos os resultados são análogos aos apresentados anteriormente, quando consideramos o estimador $\hat{\rho}_1$ (ou coeficiente α de Cronbach).

Teste da Hipótese H_0: $\rho \leqq \rho_0$ versus H_1: $\rho > \rho_0$

Determinemos o valor crítico ρ_c tal que se $\hat{\hat{\rho}}_1 > \rho_c$, H_0 é rejeitada com nível de significância α.

A partir de (4.3.20) e de (4.3.26), após alguns cálculos obtemos que, quando $\rho = \rho_0$,

$$F_{(K-1)(n-),(n-1)} = \frac{n-1}{n-3}\,\frac{1-\hat{\hat{\rho}}_1}{1-\rho_0}$$

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

Segue-se então que

$\dfrac{n-1}{n-3}\,\dfrac{1-\rho_c}{1-\rho_0}$ é igual ao valor tabelado da $F_{(K-1)(n-1),(n-1);\alpha}$

isto é, que

$$\rho_c = 1-(1-\rho_0)\frac{n-3}{n-1}F_{(K-1)(n-1),(n-1);\alpha}.$$

Portanto, rejeitamos a hipótese de que $\rho \leq \rho_0$ com nível de significância α se

$$\hat{\rho}_1 > 1-(1-\rho_0)\frac{n-3}{n-1} F_{(K-1)(n-1),(n-1);\alpha} . \tag{4.3.47}$$

Teste da Hipótese H_0: $\rho \geq \rho_0$ versus H_1: $\rho < \rho_0$

Utilizando (4.3.47), da teoria de testes segue-se que rejeitamos a hipótese de que $\rho \geq \rho_0$ com nível de significância α, se

$$\hat{\rho}_1 < 1-(1-\rho_0)\frac{n-3}{n-1} F_{(K-1)(n-1),(n-1);1-\alpha} . \tag{4.3.48}$$

Teste da Hipótese H_0: $\rho = \rho_0$ versus H_1: $\rho \neq \rho_0$

Usando os resultados obtidos para testar as duas hipóteses anteriores, da teoria de testes, segue-se que rejeitamos a hipótese que ρ assume um valor específico ρ_0 com nível de significância α se

$$\begin{aligned} &\hat{\rho}_1 > 1-(1-\rho_0)\frac{n-3}{n-1} F_{(K-1)(n-1),(n-1);\frac{\alpha}{2}} \quad \text{ou} \\ &\hat{\rho}_1 < 1-(1-\rho_0)\frac{n-3}{n-1} F_{(K-1)(n-1),(n-1);1-\frac{\alpha}{2}} \end{aligned} \tag{4.3.49}$$

Intervalo de Confiança para o Coeficiente de Fidedignidade, usando o Estimador de Máxima Verossimilhança $\hat{\rho}_1$.

A partir de (4.3.20) e de (4.3.26), após alguns cálculos obtemos que

$$F_{(n-1),(K-1)(n-1)} = \frac{n-3}{n-1}\,\frac{1-\rho}{1-\hat{\rho}_1}$$

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

Segue-se então que

$$P[F_{(n-1),(K-1)(n-1);\frac{\alpha}{2}} < \frac{n-3}{n-1}\frac{1-\rho}{1-\hat{\hat{\rho}}_1} < F_{(n-1),(K-1)(n-1);1-\frac{\alpha}{2}}] = 1-\alpha.$$

Portanto um intervalo de confiança para a fidedignidade de um teste com coeficiente de confiança $1-\alpha$ é dado por

$$(1-(1-\hat{\hat{\rho}}_1)\frac{n-1}{n-3}F_{(n-1),(K-1)(n-1);1-\frac{\alpha}{2}}\ ;\ 1-(1-\hat{\hat{\rho}}_1)\frac{n-1}{n-3}F_{(n-1),(K-1)(n-1);\frac{\alpha}{2}}) \qquad (4.3.50)$$

Substituindo $\hat{\hat{\rho}}_1$ por $\frac{2}{n-1} + \frac{n-3}{n-1}\hat{\rho}_1$ (fórmula 4.3.26), segue-se que o intervalo acima é equivalente ao obtido para ρ quando usamos $\hat{\rho}_1$ (fórmula (4.3.46)).

CAPÍTULO V

APLICAÇÕES

5.1 - *Introdução*

Neste capítulo vamos apresentar exemplos de aplicações de alguns resultados obtidos anteriormente.

Queremos salientar que estes exemplos foram obtidos juntos a pesquisadores das áreas de Psicologia e Enfermagem, e que em nenhum momento pretendemos avaliar o instrumento, em função do que ele pretendia medir, da objetividade, clareza, número e disposição dos itens, etc. O nosso objetivo é, a partir dos resultados amostrais que nos foram gentilmente fornecidos, mostrar como se obtém alguns dos resultados apresentados nos capítulos anteriores.

Dos métodos de estimação do coeficiente de fidedignidade de um teste, vamos exemplificar o método das metades, o coeficiente α de Cronbach e os estimadores de máxima verossimilhança. Achamos desnecessário apresentar exemplos do método de teste-reteste e das formas paralelas.

Para a obtenção das estatísticas básicas, utilizamos o pacote SPSS (STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES), versão 8, por se encontar a disposição do Centro de Processamento de Dados da UFRGS, e ser manualmente muito trabalhoso.

5.2 - *Aplicações em Enfermagem*

Inicialmente apresentaremos um teste que foi aplicado numa amostra aleatória de 30 alunos do Curso de Enfermagem, e a seguir, os resultados obtidos para alguns estimadores do coeficiente de fidedignidade.

Apresentação do Instrumento de Medida

Responda a cada um dos 10 itens a seguir, expressando em que medida foi significativo para a aprendizagem da disciplina de Enfermagem Médica, o estágio realizado por você no Hospital A, de acordo com o seguinte critério:

1. Nada significativo
2. Pouco significativo
3. Medianamente significativo
4. Muito significativo
5. Muitíssimo significativo

1. Seu crescimento como pessoa..........................( )
2. Seu crescimento como futuro profissional...............( )
3. Uma maior integração com a equipe multidisciplinar......( )
4. O desenvolvimento de habilidades na assistência ao cliente hospitalizado......................................( )
5. O desenvolvimento de habilidades na tomada de decisões...( )
6. A integração da teoria e prática.......................( )
7. O sentir-se satisfeito(a) com o processo do cuidado.....( )
8. O sentir-se motivado(a) para prestar cuidado ao cliente.( )
9. Um trabalho com o cliente dentro da proposição de assistência centrada.......................................( )
10. O desenvolvimento da criatividade......................( )

Estimação do Coeficiente de Fidedignidade

Divisão de um Teste pela Metade (Método da Metades)

O teste foi dividido de tal forma que os 5 primeiros itens formavam uma parte e os demais a outra parte. Ao utilizarmos este procedimento, não estávamos preocupados em obter metades paralelas, que segundo a teoria apresentada no capítulo IV, secção (4.2.2), é o procedimento que fornece a estimativa mais próxima do verdadeiro valor da fidedignidade.

A partir dos dados amostrais, obtivemos um coeficiente de correlação linear entre as metades igual a 0,75522. Portanto pela fórmula de Spearman-Brown, que é dada por

$$\rho_{XX'} = \frac{2\rho_{YY'}}{1+\rho_{YY'}} ,$$

a estimativa do coeficiente de fidedignidade do teste é

$$\rho_{XX'} = \frac{2 \times 0{,}75522}{1 + 0{,}75522} = 0{,}86054.$$

Coeficiente α de Cronbach

Sob a suposição de que todas as componentes de medida do teste são essencialmente τ-equivalentes, vamos determinar a estimativa da fidedignidade do teste através do coeficiente α de Cronbach (fórmula(2.4.6)). Para obtermos esta estimativa, precisamos da matriz de variâncias e covariâncias entre os itens, que apresentamos a seguir.

	1	2	3	4	5	6	7	8	9	10
1	.32299									
2	.07816	.25747								
3	.12299	.20230	.92989							
4	.11034	.08966	.08276	.38621						
5	.16782	.21609	.17471	.28276	1,02989					
6	.15517	.13793	.18966	.06897	.17241	.39655				
7	.13678	.17471	.26782	.13103	.25747	.18966	.46092			
8	.08621	.13793	.18966	.20690	.31034	.08621	.25862	.32759		
9	.10460	.10115	.18046	.04138	.19080	.15517	.12529	.08621	.32299	
10	.18161	.11954	.16782	.05517	.28506	.10345	.10575	.13793	.10115	.60320

De acordo com esta matriz e com a notação utilizada no capítulo 2, secção 2.4, temos que $\sum_{i=1}^{10} \sigma^2(Y_i) = 5,0369$ e $\sum\sum_{i \neq j} \sigma(Y_i, Y_j) = 13,8552$, portanto

$$\sigma_X^2 = 18,8921.$$

Segue-se então que a estimativa de fidedignidade do teste, usando o coeficiente α é

$$\alpha = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{10} \sigma^2(Y_i)}{\sigma_X^2}\right], \text{ onde K é o número de itens}$$

$$= \frac{10}{9}\left[1 - \frac{5,0369}{18,8921}\right]$$

$$= 0,81487.$$

Estimadores de Máxima Verossimilhança

Analisando a distribuição dos escores atribuídos aos 10 itens, concluímos que podemos supor que eles provém de uma população com distribuição Normal Multivariada.

Antes de determinarmos os estimadores de máxima verossimilhança da fidedignidade. precisamos verificar se os dados a mostrais satisfazem as condições especificadas nos casos 1 e 2.

Caso 1 - Os 10 itens do teste possuem iguais variâncias e iguais covariâncias.

De acordo com a teoria apresentada em WILKS(1948-pág.275 -278), obteve-se os seguintes resultados:

$\chi^2 = 66,48204$ e 53 graus de liberdade,

que implica em aceitar a igualdade das variâncias e a igualdade das covariâncias com nível de significância menor ou igual a 0.10097.

Considerando os resultados satisfatórios para a validade das suposições, prosseguimos na determinação das estimativas de $\hat{\rho}_1$ e de $\hat{\hat{\rho}}_1$.

De acordo com a fórmula (4.3.14),

$$\hat{\rho}_1 = \frac{K}{K-1}\left[1-\frac{\sum_{i=1}^{K} S_i^2}{S_X^2}\right],$$

e com a matriz de variâncias e covariâncias entre os itens, segue-se que

$$\hat{\rho}_1 = \frac{10}{9}\left[1-\frac{13,8552}{18.8921}\right] = 0,81487,$$

que coincide com o coeficiente α de Crcnbach.

Utilizando a fórmula (4.3.26). isto é,

$$\hat{\hat{\rho}}_1 = \frac{2}{n-1} + \frac{n-3}{n-1}\,\hat{\rho}_1,$$

segue-se que o estimador de máxima verossimilhança não viciado, assume um valor igual a

$$\hat{\hat{\rho}}_1 = \frac{2}{29} + \frac{27}{29}\,0,81487 = 0,82764.$$

Caso 2 - Os 10 itens do teste possuem iguais variâncias, iguais covariâncias e médias também iguais.

Para a verificação da suposição do caso 2, utilizamos a teoria apresentada em WILKS(1946-pág.268-275).

A partir dos dados amostrais obtivemos

χ^2 = 157,87598 e 62 graus de liberdade,

que implica em rejeitar a hipótese de igualdade de variâncias, igualdade de covariâncias e também igualdade de médias com nível de significância 0,00000.

Segue-se então que na prática o caso 2 não pode ser utilizado para estimar o coeficiente de fidedignidade do teste. Entretanto, como estamos apresentando estes exemplos somente para melhor compreensão da obtenção dos resultados apresentados em capítulos anteriores, decidimos determinar assim mesmo as estimativas para $\hat{\rho}_2$ e $\hat{\hat{\rho}}_2$.

Conforme a fórmula (4.3.16),

$$\hat{\rho}_2 = \frac{K}{K-1}\left[1-\frac{Kn-1}{n-1}\,\frac{S_X'^2}{S_X^2}\right],$$

onde K é o número de partes do teste,

n é o número de itens do teste,

$$S_X^2 = \sum_{j=1}^{K} S_j^2 + \sum_{i\neq j}\sum S_{ij} \qquad \text{e}$$

$$S_X'^2 = \frac{1}{Kn-1}\left[\sum_{j=1}^{K}\sum_{i=1}^{n}(X_{ij})^2 - \frac{1}{Kn}\left(\sum_{j=1}^{K}\sum_{i=1}^{n}X_{ij}\right)^2\right],$$

a partir dos dados amostrais temos que n=30, K=10, S_X^2 = 18,8921 e

$$S_X'^2 = \frac{1}{10 \times 30-1}\left[5500-\frac{1}{10 \times 30}(1262)^2\right] = 0,63942.$$

Portanto a estimativa de $\hat{\rho}_2$ é

$$\hat{\rho}_2 = \frac{10}{10-1}\left[1- \frac{10 \times 30-1}{30-1}\ \frac{0,63942}{18,8921}\right] = 0,72337.$$

De acordo com a fórmula (4.3.30),

$$\hat{\hat{\rho}}_2 = \frac{3}{n} + \frac{n-3}{n}\ \hat{\rho}_2.$$

Segue-se então que a estimativa de $\hat{\hat{\rho}}_2$ é

$$\hat{\hat{\rho}}_2 = \frac{3}{30} + \frac{30-3}{30}\ 0,72337 = 0,75103$$

Teste de Hipótese para o Coeficiente de Fidedignidade

Suponhamos que desejamos testar a hipótese

H_0: $\rho \leqq 0,70$ versus H_1: $\rho > 0,70$ com $\alpha = 0,05$.

Se $\hat{\rho}_1$ é usado para estimar ρ, temos da desigualdade (4.3.43) que rejeitamos H_0, com nível de significância α se

$$\hat{\rho}_1 > 1-(1-\rho_0)F_{(K-1)(n-1),(n-1):\alpha},$$

isto é, se

$$\hat{\rho}_1 > 1-(1-0,70)F_{261,29;0,05} = 0,80000.$$

Uma vez que $\hat{\rho}_1$ é igual 0,81487, rejeitamos a hipótese que $\rho \leqq 0,70$ com nível de significância 0,05.

Se $\hat{\hat{\rho}}_1$ é usado para estimar ρ, temos da desigualdade (4.3.47), que rejeitamos H_0 com nível de significância α se

$$\hat{\hat{\rho}}_1 > 1-(1-\rho_0)\frac{n-3}{n-1}\ F_{(K-1)(n-1),(n-1);\alpha},$$

isto é, se

$$\hat{\hat{\rho}}_1 > 1-(1-0,70)\frac{30-3}{30-1}\ F_{261,29;0,05} = 0,84483.$$

Como $\hat{\rho}_1$ é igual a 0,82764, não rejeitamos a hipótese que $\rho \leqq 0,70$ com nível de significância 0,05.

Intervalo de Confiança para o Coeficiente de Fidedignidade

Se $\hat{\rho}_1$ é usado para estimar ρ, então da fórmula (4.3.46), temos que um intervalo de confiança para a fidedignidade do teste com coeficiente de confiança $1-\alpha$ é dado por

$$(1-(1-\hat{\rho}_1)F_{(n-1),(K-1)(n-1);1-\frac{\alpha}{2}}\ ;\ 1-(1-\hat{\rho}_1)F_{(n-1),(K-1)(n-1);\frac{\alpha}{2}}).$$

Para $\alpha = 0,05$ temos

$$(1-(1-0,81487\ F_{29,261;0,975};1-(1-0,81487)F_{29,261;0,025}),$$

que é equivalente a afirmar que o coeficiente de fidedignidade ρ pertence ao intervalo (0,70008;0,89993) com coeficiente de confiança 0,95.

O intervalo de confiança para ρ, usando o estimador $\hat{\hat{\rho}}_1$ é idêntico ao que acabamos de obter, conforme já mostramos na secção (4.3.5).

5.3 - *Aplicações em Psicologia*

Descrição do Instrumento de Medida

Vamos exemplificar uma aplicação em Psicologia, usando um teste que foi administrado em uma amostra aleatória de 36 funcionários da Empresa FG. Este teste era composto por 80 itens, divididos em 8 subtestes com 10 itens cada, e visava medir o nível de satisfação dos funcionários desta empresa, considerando conjuntamente e isoladamente os seguintes aspectos (subtestes): normas administrativas, estilos de chefias, salários, relações interpessoais, promoções, sentido de realização, reconhecimento pelo trabalho e conteúdo do trabalho.

Os 80 itens que constituiam o teste eram afirmativas que os funcionários amostrados deveriam responder, marcando com um círculo o escore que representava a situação atual, de acordo com a seguinte escala

0	1	2	3	4	5	6	7	8	9	10

Os itens de cada subteste se encontravam distribuídos ao longo de todo o teste, e foi informado apenas que o objetivo era utilizar os dados a fim de tornar a situação de trabalho mais agradável, satisfatória e produtiva.

Como este teste é muito extenso para ser apresentado aqui integralmente, vamos considerar somente o subteste que visava medir o grau de satisfação dos funcionários da Empresa FG em relação às promoções.

Apresentação dos Itens Relativos às Promoções

No Departamento em que trabalho

1. O sistema de promoções visa aproveitar os funcionários do próprio Departamento.
2. Ser promovido significa poder enfrentar desafios maiores.
3. As promoções são efetivadas com justiça.
4. As possibilidades de promoção dependem da capacidade do funcionário.
5. É bom ser promovido.
6. Eu conheço os critérios de promoção.
7. Os funcionários são treinados para serem promovidos.
8. Qualquer funcionário que tenha capacidade pode ser promovido.
9. Pode-se falar francamente sobre promoções.
10. Pode-se sentir o interesse na promoção dos funcionários.

Estimação do Coeficiente de Fidedignidade

Analogamente ao exemplo da área de Enfermagem, vamos determinar a seguir, algumas estimativas da fidedignidade do teste que visava medir o grau de satisfação dos funcionários em relação às promoções.

Divisão do Teste pela Metade

Dividimos o teste (promoção) de tal forma que os 5 primeiros itens formavam uma parte e os demais a outra parte.

A partir dos dados amostrais resultou um coeficiente de correlação linear entre as metades igual a 0,63733, e aplicando a fórmula de Spearman-Brown (fórmula 2.2.10), obtivemos que a estimativa da fidedignidade do teste é igual a

$$\rho_{XX'} = \frac{2 \times 0,63733}{1 + 0,63733} = 0,77850.$$

Coeficiente α de Cronbach

Sob a suposição de que todas as componentes de medida do teste são essencialmente τ-equivalentes, vamos determinar a estimativa da fidedignidade através do coeficiente α de Cronbach.

Usando os dados amostrais, segue-se que a matriz de variâncias e covariâncias entre os itens é dada por

	1	2	3	4	5	6	7	8	9	10
1	4,54902									
2	2,32085	6,16488								
3	2,69519	1,90018	4,83422							
4	3,03922	1,84759	2,64706	6,38592						
5	2,79323	3,02674	2,67914	2,39216	5,26560					
6	2,74153	2,22282	3,68984	3,06595	2,90018	7,81462				
7	3,50446	1,97683	3,06595	3,57932	3,43850	3,56326	6,94118			
8	3,04456	1,96524	3,90196	3,80838	3,20321	3,32977	5,83601	9,34492		
9	1,11765	0,58824	0,85027	2,77897	1,20677	1,44029	2,46524	3,31907	4.30660	
10	1,10517	-0,85294	0,47772	0,59626	0,23351	1,93583	2,13904	3,64795	1,16756	6,79947

A partir desta matriz temos que

$$\sum_{i=1}^{10} \sigma^2(Y_i) = 62{,}40643 \quad e \quad \sum_{i \neq j}\sum \sigma(Y_i . Y_j) = 216{,}79152,$$

e portanto, $\sigma_X^2 = 279{,}19795$.

Segue-se então que a estimativa do coeficiente de fidedignidade é

$$\alpha = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{10} \sigma^2(Y_i)}{\sigma_X^2}\right]$$

$$= \frac{10}{9}\left[1 - \frac{62{,}40643}{279{,}19795}\right]$$

$$= 0{,}86276.$$

Estimadores de Máxima Verossimilhança

Após análise dos dados amostrais concluímos que podemos considerá-los provindos de uma população com distribuição Normal Multivariada. Além disto, deve-se verificar se as condições dos casos 1 e 2 estão satisfeitos, para então calcularmos as estimativas de máxima verossimilhança para ρ.

Caso 1 - Os 10 itens do teste possuem iguais variâncias e iguais covariâncias.

Esta condição foi testada de acordo com WILKS (1948, pág. 275-278), obtendo-se os seguintes resultados

$\chi^2 = 72{,}19233$ e 53 graus de liberdade,

que implica em aceitar a igualdade de médias e igualdade de variâncias com nível de significância menor ou igual a 0,04090.

Considerando a validade da suposição do caso 1, prosseguimos determinando as estimativas de $\hat{\rho}_1$ e de $\hat{\hat{\rho}}_1$.

Como já mostramos no exemplo anterior e no capítulo IV, $\hat{\rho}_1$ coincide com o coeficiente α de Cronbach, cujo valor foi igual a 0,86276.

Utilizando a fórmula (4.3.26), isto é,

$$\hat{\hat{\rho}}_1 = \frac{2}{n-1} + \frac{n-3}{n-1}\,\hat{\rho}_1,$$

segue-se que a estimativa de $\hat{\hat{\rho}}_1$, o estimador de máxima verossimilhança não-viciado, é igual a

$$\hat{\hat{\rho}}_1 = \frac{2}{34-1} + \frac{34-3}{34-1}\;0{,}86276 = 0{,}87108.$$

Neste caso, o tamanho da amostra passou de 36 para 34, pois 2 funcionários deixaram de responder pelo menos um item.

Caso 2 - Os 10 itens do teste possuem iguais variâncias, iguais covariâncias e médias também iguais.

A verificação desta condição foi realizada usando a teoria dada por WILKS(1946,pág.268-275).

A partir dos dados amostrais, obtivemos

$\chi^2 = 90{,}12919$ e 62 graus de liberdade,

que implica em aceitar a condição imposta no caso 2 com nível de significância menor ou igual a 0,01132.

Considerando que este nível de significância pode ser aceito, prosseguimos determinando as estimativas de $\hat{\rho}_2$ e $\hat{\hat{\rho}}_2$.

A partir da fórmula (4.3.16) e dos dados amostrais, segue-se que

$K = 10,\quad n=34,\quad S_X^2 = 279{,}19795,$

$$S_X'^2 = \frac{1}{10 \times 34-1}[22050 - \frac{1}{10 \times 34}(2612)^2] = 5{,}85161$$

e

$$\hat{\rho}_2 = \frac{K}{K-1}[1 - \frac{Kn-1}{n-1}\,\frac{S_X'^2}{S_X^2}]$$

$$= \frac{10}{10-1}\left[1- \frac{10 \times 34-1}{34-1}\ \frac{5,85161}{279,19795}\right] =$$

$$= 0,87189.$$

Usando a estimativa de $\hat{\rho}_2$ e a fórmula(4.3.30), segue-se que a estimativa da fidedignidade, de acordo com $\hat{\hat{\rho}}_2$ é

$$\hat{\hat{\rho}}_2 = \frac{3}{n} + \frac{n-3}{n}\ \hat{\rho}_2$$

$$= \frac{3}{34} + \frac{34-3}{34}\ 0,87189$$

$$= 0,88319.$$

Teste de Hipótese para o Coeficiente de Fidedignidade

Suponhamos que desejamos testar a hipótese H_0: $\rho = 0,80$ versus H_1: $\rho \neq 0,80$ com nível de significância 0,10.

Se $\hat{\rho}_1$ é usado para estimar ρ, temos da desigualdade (4.3.45) que rejeitamos H_0, com nível de significância α se

$$\hat{\rho}_1 < 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);1-\frac{\alpha}{2}} \quad \text{ou}$$

$$\hat{\rho}_1 > 1-(1-\rho_0)F_{(K-1)(n-1),(n-1);\frac{\alpha}{2}}\ ,$$

isto é, se

$$\hat{\rho}_1 < 1-(1-0,80)F_{297,33;0,95} = 0,67800 \quad \text{ou}$$

$$\hat{\rho}_1 > 1-(1-0,80)F_{297,33;0,05} = 0,86667.$$

Mas $\hat{\rho}_1$ é igual a 0,86276, e portanto, não rejeitamos a hipótese de que o coeficiente de fidedignidade do teste é igual a 0,80, com nível de significância 0,10.

Se $\hat{\hat{\rho}}_1$ é usado para estimar ρ, a desigualdade (4.3.49) expressa que rejeitamos H_0 com nível de significância α se

$$\hat{\hat{\rho}}_1 < 1-(1-\rho_0)\ \frac{n-3}{n-1}\ F_{(K-1)(n-1),(n-1);1-\frac{\alpha}{2}} \quad \text{ou}$$

$$\hat{\hat{\rho}}_1 > 1-(1-\rho_0)\ \frac{n-3}{n-1}\ F_{(K-1)(n-1),(n-1);\frac{\alpha}{2}}\ ,$$

isto é, se

$$\hat{\hat{\rho}}_1 < 1-(1-0,80)\frac{34-3}{34-1}\ F_{279,33;0,95} = 0,69752 \quad \text{ou}$$

$$\hat{\hat{\rho}}_1 > 1-(1-0,80)\frac{34-3}{34-1}\ F_{297,33;0,05} = 0.87475$$

Como $\hat{\hat{\rho}}_1$ é igual a 0,87108 não rejeitamos a hipótese de que o coeficiente de fidedignidade do teste é igual a 0,80 com nível de significância 0,10.

Intervalo de Confianca para o Coeficiente de Fidedignidade

Se usamos $\hat{\hat{\rho}}_1$ para estimar ρ, então de acordo com a fórmula (4.3.50) temos que um intervalo de confiança para a fidedignidade do teste com coeficiente de confiança $1-\alpha$ é dado por

$$\left(1-(1-\hat{\hat{\rho}}_1)\frac{n-1}{n-3}\ F_{(n-1),(K-1)(n-1):1-\frac{\alpha}{2}}\ ;\ 1-(1-\hat{\hat{\rho}}_1)\frac{n-1}{n-3}\ F_{(n-1),(K-1)(n-1);\frac{\alpha}{2}}\right).$$

Para $\alpha = 0,10$ temos

$$\left(1-(1-0,87108)\frac{34-1}{34-3}\ F_{33,297;0,95};1-(1-0,87108)\frac{34-1}{34-3}\ F_{33,297;0,05}\right),$$

que é equivalente a afirmar que o coeficiente de fidedignidade do teste pertence ao intervalo (0.79414;0,91476) com coeficiente de confianca 0.90, usando $\hat{\rho}_1$ ou $\hat{\hat{\rho}}_1$.

BIBLIOGRAFIA

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Standards for Educational and Psychological Tests and Manuals. Washington, D.C.: American Psicological Association, 1966.

CRONBACH, L.J. A case study of the split-half reliability coefficient. The Journal of Educational Psychology, 37 : 473-80, 1946.

CRONBACH, L.J. Test Reliability: its meaning and determination. Psychometrika, 12(1): 1-15, 1947.

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

CURETON, E.E. The definition and estimation of test reliability. Educational and Psychological Measurement, 18(4): 715-38, 1958.

FELDT, L.S. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. Psychometrika, 30(3): 357-70, 1965.

HULL, C.H. & NIE, N.H. Statistical Package for the Social Sciences -SPSS UPDATE: New Procedures and Facilities for Releases 7 and 8 - New York: Mc Graw-Hill Boock Company, 1979.

GUILFORD, J.P. Fundamental Statistics in Psychology and Educational. New York: Mc Graw-Hill Boock Co., 1950.

GULLIKSEN, H. Theory of Mental Tests. New York: John Wiley, 1950.

GUTTMAN, L. A basis for analyzing test-retest reliability. Psychometrika, 10(4): 255-82, 1945.

JACKSON, R.W. & FERGUSON, G.A. Studies on the Reliability of Tests. Toronto, Departament of Educational Research Bulletin, 12, 1941.

KELLEY, T.L. Interpretation of Educational Measurement. World Book Company, 1927.

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

KRISTOF, W. On the sampling theory of reliability estimation. Journal of Mathematical Psychology, 7: 371-77, 1970.

KRISTOF, W. On a statistic arising in testing correlation. Psychometrika, 37: 377-84, 1972.

KRISTOF, W. On accuracy in reliability estimation. Psychometrika, 39(1): 23-9, 1974.

KUDER, G.F. & RICHARDSON, M.W. The theory of the estimation of test reliability. Psichometrika, 2(3): 151-60, 1937.

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

MOOD, A.M., GRAYBILL, F.A. & BOES, D.C. Introduction to the theory of statistics. New York, McGrawn-Hill, 1974.

NOVICK, M.R. & LEWIS, C. Coefficient alpha and the reliability of composite measurements. Psychometrika, 32(1): 1-13, 1967.

ROZEBOOM, W.W. Foundations of the theory of prediction. Homewood, Hlinois: Dorsev Press, 1966.

SYMONDS, P.M. Factors influencing test reliability. Journal of E-ducacional Psychology, 19:73-87.

WILKS, S.S. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a Normal Multivariate distribution. The Annalls of Mathematical Statistical, 17: 251-81, 1946.

Publicações do Instituto de Matemática da UFRGS
Cadernos de Matemática e Estatística

Série B: Trabalho de Apoio Didático

1. Elsa Mundstock - Curso Básico Sobre Wordstar 3.45 - MAR/89.
2. Jaime B. Ripoll - Introdução ao Cálculo Diferencial Via Funções de Uma Variável Real - OUT/89.
3. Edmund R. Puczylowski - Dimension of Modular Lattices - JUN/90
4. Marcos Sebastiani - Geometrias Não Euclidianas - JUL/90
5. Sandra R. C. Pizzato - Cálculo Numérico - AGO/91
6. Vera Clotilde G. Carneiro - Elementos de Cálculo para Biologia - AGO/91
7. Elsa Mundstock - Iniciaç ao ao SPSS/PC - SET/91
8. Elisa Haag, Loiva C. de Zeni, Maria Alice Gravina e Vera Clotilde - Notas da 1a. Oficina de Matemática da UFRGS - JAN/92
9. Paulo Werlang de Oliveira, Elisabete Rambo, Suzana Lima dos Santos. Coordenaç ao: Profa. Maria Alice Gravina - A Tartaruga no Espaço Tridimensional - FEV/92
10. Silvio Possoli - Análise Multivariada - JUL/92
11. Dinara Westphalen Fernandez - Números Índices - OUT/92
12. Maria Teresinha Albanese - Coeficiente de Fidedignidade de um instrumento de Medida - OUT/92

Universidade Federal do Rio Grande Sul
Reitor: Professor Hélgio Trindade

Instituto de Matemática
Diretor: Professor Aron Taitelbaum
Núcleo de Atividades Extra Curriculares
Coordenador: Professor Aron Taitelbaum
Secretária: Faraildes Beatriz da Silva

Os Cadernos de Matemática e Estatística publicam as seguintes séries:
Série A: Trabalho de Pesquisa
Série B: Trabalho de Apoio Didático
Série C: Colóquio de Matemática SBM/UFRGS
Série D: Trabalho de Graduação
Série E: Dissertações de Mestrado
Série F: Trabalho de Divulgação
Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:
NAEC - Núcleo de Atividades Extra Curriculares
Instituto de Matemática - UFRGS
Av. Bento Gonçalves, 9500
91.540-000 - Agronomia - POA/RS
Telefone: 336.98.22 ou 339.13.55 Ramal: 6176