Laboratório de Psicologia, 4(1): 65-90 (2006)

Qual a fiabilidade do alfa de Cronbach? Questões antigas e soluções modernas?

João Maroco

Teresa Garcia-Marques

Instituto Superior de Psicologia Aplicada, Portugal

Resumo

A análise da consistência interna de uma medida psicológica é uma necessidade aceite na comunidade científica. Entre os diferentes métodos que nos fornecem estimativas do grau de consistência de uma medida salienta-se o índice de Cronbach sobre o qual acenta a confiança da maioria dos investigadores. Os utilizadores deste método têm-no sugerido como conservador especialmente para os casos em que os itens da escala são heterogéneos, são dicotómicos ou definem estruturas multi-factoriais: o alfa de Cronbach fornece uma sub-estimativa da verdadeira fiabilidade da medida. Neste artigo apresentamos e discutimos o método de Cronbach, com ênfase na inferência sobre este índice e nas propostas alternativas a este método de estudo da consistência interna. Por último faremos uma breve referência à discussão que emerge no campo no que concerne a interpretação deste índice feita pelas perspectivas psicométrica vs. *datamétrica.*

Palavras-chave: Alfa de Cronbach, Fiabilidade, Psicometria.

Abstract

The analysis and report of a psychological measure's internal consistency is a well established requirement in the scientific community. Among the several available methods to estimate internal consistency, Cronbach's α ranks high in most researchers preferences. However, Cronbach's α underestimates the true reliability specially when the scale's items are heterogeneous, dichotomous, or define multi-factorial structures. Thus, it is a conservative estimator of internal consistency. In this paper, we discuss Cronbach's α with emphasis on inference and on alternative proposals to estimate internal consistency. We also make reference to the emerging discussion in the psychometric vs. *datametric interpretations of Cronbach's α.*

Key words: Cronbach Alpha, Psychometric, Reliability.

A correspondência relativa a este artigo deverá ser enviada para: João Maroco, Instituto Superior de Psicologia Aplicada, Rua Jardim do Tabaco, 34, 1149-041 Lisboa; E-mail: jpmaroco@ispa.pt

Qualquer referencia a questões de fiabilidade[1] (*reliability*) de uma medida suscita referência ao índice alfa de Cronbach. A maioria dos investigadores, talvez com excepção daqueles que dedicam alguma atenção à área da psicometria, tende não apenas a considerá-lo o índice universalmente aconselhável para o estudo métrico de uma escala (qualquer que sejam as suas características) como tendem a percebê-lo como fornecendo "estimativas fiáveis" da "fiabilidade de uma escala".

Neste artigo pretendemos chamar atenção dos leitores para a diversidade de índices alternativos ao índice de Cronbach e para as características deste último. Queremos responder à questão da validade e fiabilidade das suas estimativas. Para podermos compreender a questão analisaremos em primeiro lugar o conceito de fiabilidade de uma medida, as propostas para a sua estimativa, para nos focarmos de seguida de forma mais aprofundada na proposta associada ao nome de Cronbach. Neste artigo adicionamos informação detalhada para aqueles que se interessam pela forma como as estatísticas são desenvolvidas e definidas e por isso assume-se que o leitor interessado nestas temáticas terá proficiência básica com a formulação estatística. Contudo, e procurando "instrumentalizar" esta revisão de literatura para o utilizador menos interessado nos pormenores técnicos, fornecemos informação sobre o modo como este índice se pode calcular com dois programas estatísticos de utilização ubíqua nas ciências sócias e humanas: o SPSS e o STATISTICA (Anexo), com ênfase nas limitações e interpretações da estimação da fiabilidade de um instrumento.

Apontaremos de seguida algumas limitações e más interpretações associadas ao uso do índice de Cronbach, apresentando as alternativas de cálculo actualmente aceites como melhores estimadores de fiabilidade de uma escala. Apresentamos ainda a formulação que permitirá ao leitor interessado utilizar estes estimadores que não se encontram ainda disponíveis nos softwares, mas que são já exigidos por algumas publicações das ciências sociais e humanas (como é por exemplo o caso dos intervalos de confiança para o alfa de Cronbach).

Por fim confrontaremos o leitor com duas perspectivas teóricas associadas ao significado de uma "estimativa de fiabilidade". Indica-nos ela que o instrumento utilizado para obter uma medida é fiável ou apenas que os dados com o instrumento são fiáveis?

O conceito de fiabilidade

A fiabilidade de uma medida refere a capacidade desta ser *consistente*. Se um instrumento de medida dá sempre os mesmos resultados (dados) quando aplicado a alvos estruturalmente iguais, podemos confiar no significado da medida e dizer que a *medida é fiável*. Dizemo-lo porém com maior ou menor grau de certeza porque toda a medida é sujeita a erro. Assim a fiabilidade que podemos observar nos nossos dados é uma estimativa, e não um "dado".

Qualquer medida, classificação X obtida por uma escala ou teste por um indivíduo, tem sempre duas componentes aditivas (ver e.g., Pasquali, 2003): o verdadeiro *score* (resultado), capacidade, classificação ou medida (τ) do objecto e o erro de medida do atributo ou capacidade do objecto (ε_x):

$$X = \tau + \varepsilon_x \quad (1)$$

"Erro" é a *variabilidade* observada no processo de mensuração de um mesmo objecto. Ausência de erro é "consistência". Consistência é assim o termo fundamental para definir o conceito de *fiabilidade*.

[2] Optamos pela tradução do termo "reliability" por fiabilidade. Em outros textos este termo tem sido traduzido por "precisão", "fidelidade", "fidedignidade".

Fiabilidade vs. validade

Mas o erro (ε_x) associado à variabilidade observada é um erro aleatório (o que é uma característica desejada mas que se pretende ser de magnitude reduzida). O erro pode porém ser sistemático. O erro sistemático traduz não uma questão de fiabilidade mas uma questão de Validade. O instrumento com erro sistemático é um instrumento com validade reduzida, é um instrumento que está a medir algo que não era suposto medir (mesmo que o faça de forma fiável). Qualquer medida para ser válida enquanto medida de um dado construto, tem necessariamente de ser fiável. Pelo que, a fiabilidade surge como condição necessária, mas não suficiente, para a validade. Note-se que os dados de uma medida não fiável, são dados aleatórios, logo dados sem significado. Nada nos dizem. Assim sendo, dados não fiáveis, não são, igualmente validos, visto não traduzirem o conceito que pretenderiam traduzir. Assim a fiabilidade de uma medida é o primeiro passo para saber da sua validade. No entanto se esta é condição necessária à validade ela não é suficiente. Após garantir fiabilidade é necessário pôr de lado a hipótese de existência de erro sistemático, para podermos garantir validade.

Fiabilidade e unidimensionalidade

É importante notar que se uma medida é unidimensional, ela apresenta de certo uma maior consistência. No entanto, a consistência de uma medida nada diz sobre a sua dimensionalidade. Na verdade a escala pode ter vários factores e ainda assim apresentar um nível de consistência interna elevado (ver e.g., Cortina, 1993). A consistência é uma condição necessária mas não suficiente para a unicidade da escala. Um conjunto de itens pode apresentar elevada consistência interna, i.e., apresentarem-se interrelacionados, mas ainda assim definir uma escala multidimensional (Green, Lissitz, & Mulaik, 1977; Cortina, 1993).

A elevada consistência na presença de multidimensionalidade indica que os itens que compõem as diferentes dimensões de uma medida estão fortemente correlacionados, apesar das dimensões em si, estabelecerem uma relação inferior àquela que é observada entre os itens que as compõem. Com um exemplo corriqueiro percebe-se facilmente o conceito. Imaginemos que queremos uma medida do tamanho do pé de uma pessoa. Medimos as suas meias, os seus ténis, as suas pantufas, etc. O grau de relação entre as diferentes medidas é elevado e a sua média pode fornecer-nos uma estimativa fiável do tamanho do pé do indivíduo. Nunca confundiríamos no entanto, o pé com o sapato, nem a meia com o sapato. Seriam dimensões distintas que nos informam sob o mesmo constructo subjacente. Contudo, e ao contrário deste exemplo, a ortogonalidade de factores nem sempre é aparente nas medidas psicológicas o que, associado, à tradição de se forçar a existência de uma estrutura factorial ortogonal, desaconselha a aplicação da fiabilidade como medida de dimensionalidade. E claro está, uma escala pode ser unidimensional e por falta de fiabilidade da sua medida ou elevado erro de medida, apresentar fraca consistência. Voltaremos adiante a este conceito.

O conceito estatístico de fiabilidade

Considerando o erro aleatório como variabilidade intra-sujeito, os dados associados a uma medida permitem-nos inferir a sua fiabilidade através da variância observada intra e inter-sujeitos/objectos.

Quanto maior a variância intersujeitos [$V(\tau)$] maior é a informação que essa medida transporta; pelo contrário, se esta variância for nula, a medida é constante, e a informação transportada é naturalmente, reduzida. Assumindo, teoricamente, que a característica mensurada (τ) é independente do erro de medida (ε_x), a variância geral dos dados [$V(X)$] é dada por:

$$V(X) = V(\tau) + V(\varepsilon_X) \quad (2)$$

i.e., a variância observada nos dados é a soma da variância intersujeitos e dos erros de medição (variancia intra-sujeitos). A fiabilidade de um instrumento, teste ou escala é então formalmente definida como sendo a fracção da variância (informação) do *score* verdadeiro (não medido directamente) que é retida pelo *score* observado:

$$\Phi = \frac{V(\tau)}{V(X)} = \frac{V(\tau)}{V(\tau) + V(\varepsilon_X)} \quad (3)$$

Por exemplo, um Φ=0.80 indica que 80% da variância observada nos *scores* do teste é devida ao facto de se estar a medir diferentes objectos (variância real) enquanto que o restante 20% é resultante do erro de medida (variabilidade associada à medida do mesmo objecto). Mas, como separar as componentes de variabilidade $V(\tau)$ e $V(\varepsilon)$?

Como 'estimar' a consistência de uma medida?

A lógica de qualquer processo de estimativa é conhecida de forma intuitiva por todos nós. Tomemos como exemplo intuitivo, o uso de uma balança numa charcutaria. Encomendamos 100g de fiambre, o mostrador da balança marcou 101g. Por alguma razão o empregado repete o processo de mensuração. Se volta a marca 101g, não nos espantamos. A segunda medida estabeleceu uma relação perfeita com a segunda. E se marcar 105g? Percebemos que a balança "comete erros", porque induz variabilidade "*intra-fiambre"*. Se repetíssemos o processo e o resultado fosse 100g 101g 104g 101g, teríamos uma estimativa da grandeza dos erros cometidos pela balança. Estes parecem relativamente "insignificante". Mas, se o resultado fosse 80g, 106g, 85g, 119g essa estimativa sugeria um erro de elevada grandeza. Torna-se saliente neste exemplo que o processo básico para estimar a consistência de uma medida envolve a *repetição da mensuração* sob o mesmo objecto adicionada à *avaliação da relação* entre as diferentes medidas obtidas.

Estimação da magnitude da fiabilidade

É assim possível estimar a fiabilidade de um medida se tivermos, pelo menos duas medidas de um mesmo objecto. Esta estimativa de consistência entre as duas medidas vai depender da força da relação existente entre as duas medidas e da sua variabilidade.

Em termos estatísticos, $V(\tau)$ e $V(\varepsilon)$? são passíveis de serem estimados com base em, pelo menos, duas medidas X_1 e X_2 de um mesmo objecto:

$$\begin{aligned} X_1 &= \tau + \varepsilon_{X_1} \\ X_2 &= \tau + \varepsilon_{X_2} \end{aligned} \quad (4)$$

Estas duas medidas apresentam 3 propriedades importantes, de acordo com a teoria clássica da medida: a) a capacidade real latente (τ) mantém-se inalterada e não é afectada pelos erros [$Cov(\tau, \varepsilon_{x1})$=0]; b) a variação observada entre X_1 e X_2 é devida aos erros aleatórios (ε_{x1} e ε_{x2}) que são independentes [$Cov(\varepsilon_{x1}, \varepsilon_{x2})$=0] e de valor esperado nulo [$E(\varepsilon_{x1})$=0 e $E(\varepsilon_{x2})$=0] e c) X_1 e X_2 covariam pois partilham τ. Esta partilha, estimada pela covariância entre X_1 e X_2, é fulcral à estimação operacional da

fiabilidade já que é intuitivo que quanto maior a fracção da variância de X_1 e X_2 que é devida a τ, maior a correlação entre as duas medidas. A covariância entre X_1 e X_2, i.e., a variância comum de X_1 e X_2, é essencialmente, uma estimativa de $V(\tau)$ (sendo τ o elemento comum de X_1 e X_2). Estandardizando a covariância, i.e., dividindo a covariância de X_1 e X_2 pelos desvios-padrão de X_1 e X_2 obtém-se:

$$\frac{Cov(X_1, X_2)}{S_{X_1} \times S_{X_2}} = R \tag{5}$$

que é forma ubíqua do coeficiente de correlação de Pearson. Uma vez que X_1 e X_2 medem supostamente a mesma característica ou medida é expectável que $S'_{X1}=S'_{X2}=S'_X$ donde:

$$R = \frac{Cov(X_1, X_2)}{S_X^2} \approx \frac{V(\tau)}{V(X)} = \Phi \tag{6}$$

A fiabilidade pode assim ser 'estimada' pelo coeficiente de correlação de duas medições convergentes. De (6) resulta a definição operacional de fiabilidade: *correlação entre duas formas paralelas ou convergentes do teste ou instrumento de medida*. É sobre esta forma, usando medidas repetidas, que a fiabilidade é geralmente estimada. A forma mais intuitiva é a de utilizar o mesmo instrumento em momentos distintos e este procedimento designa-se *teste re-teste*. Se existir estabilidade na medida os resultados estarão fortemente relacionados. Quando falamos em medidas psicológicas sabemos porém que existem factores que podem induzir diferenças nas respostas apenas por estarmos a aceder a uma medida junto de um mesmo individuo duas vezes com o mesmo instrumento. Esta é a razão de ser do procedimento de comparação de *formas equivalentes*. Sabendo-as equivalentes pressupomos que os seus resultados estejam relacionados informando-nos igualmente da estabilidade ou consistência da medida.

Fiabilidade enquanto consistência interna de uma medida

Na Psicologia e nas Ciências Sociais, são geralmente usadas escalas multi-item (usualmente construídas segundo a metodologia proposta por Likert, 1932) para avaliar diferentes capacidades, caracteristicas de personalidade, ou outras dimensões psicológicas. Em muitas circunstâncias o *score* total da escala resulta da combinação dos itens (por exemplo, por soma) e os resultados são submetidos a análise de fiabilidade. Se a nossa medida é uma escala com vários itens, podemos estrategicamente pressupor que metade dos itens mede o mesmo que a outra metade, sendo formas equivalentes de medida. Mas o que mede exactamente a correlação entre os scores das duas metades do teste e o total da escala? O procedimento de *split-half*, proposto por C. Spearman (1910) e W. Brown (1910) refere que se a correlação entre as medidas fornecidas por duas metades da escala for elevada a escala tem coerência com o teste na sua globalidade, i.e., tem consistência interna: As duas metades do teste medem o mesmo constructo. Se a correlação for baixa, as duas metades medem constructos diferentes.

A correlação entre as duas metades de um teste é dada por:

$$r_{T_1T_2} = \frac{\frac{1}{2}\left(S_T^2 - S_{T_1}^2 - S_{T_2}^2\right)}{S_{T_1} S_{T_2}} \tag{7}$$

onde $S_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(T_i - \bar{T})^2$ é a variância dos resultados totais do teste (i.e., a variância dos scores totais do teste (i.e., a variância dos scores totais $T_i = \sum_{j=1}^{k} X_{ij}$ de cada indivíduo i nos k itens) e S^2_{T1} *e* S^2_{T1} são as variâncias dos resultados totais das metades 1 e 2 do teste. Assumindo, a homogeneidade de covariâncias e variâncias, e usando as correlações não redundantes entre os k itens, assumidas como

homogéneas ($\rho_{12}=\rho_{13}=...=\rho_{21}=\rho_{23}=...=\rho$) e estimadas pela correlação média entre os itens ($\bar{r}$), a expressão (7) pode ser reescrita como:

$$r_{SB2} = \frac{k \times \bar{r}}{1 + (k-1) \times \bar{r}} \tag{8}$$

É de referir porém (ver e.g., Laveault & Grégoire, 2002) que (i) a fiabilidade calculada deste modo fornece a precisão do resultado total a partir dos resultados parciais das metades do teste o que pode sub-estimar a verdadeira fiabilidade total (aquela que de facto interessa) e (ii) depende da forma de divisão dos itens pelas duas metades ou formas equivalentes.

A informação fornecida pelos diferentes procedimentos não é exactamente a mesma. Tal levou à consideração de três tipos de fiabilidade (ver por exemplo, Krathwohl, 1998). A "fiabilidade de estabilidade" avalia a consistência com que uma medida se perpetua ao longo do tempo; por outro lado, a "fiabilidade de equivalência" avalia a consistência com que diferentes formas de um teste ou instrumento medem um mesmo constructo latente. Finalmente, a "consistência interna" avalia a consistência com que um determinado conjunto de itens de medida estima um determinado constructo ou dimensão latente. Estudar a consistência interna de uma medida como uma estimativa da sua fiabilidade tem a vantagem de apenas implicar um processo de mensuração. Assim, são várias as propostas de índices que nos permitem aceder a essa estimativa. Spearman e Brown, propuseram uma correcção que permite, em termos práticos, corrigir a sub-estimação da consistência pelo método *split-half*.

Considerando a correlação entre as metades T_1 e T_2 do teste, a consistência corrigida de Spearman-Brown é estimada por:

$$\Phi_{SB} = \hat{\rho}_{T_1T_2} = \frac{2r_{T_1T_2}}{1 + r_{T_1T_2}} \tag{9}$$

Contudo, essa correcção só produz estimativas da verdadeira correlação entre as metades do teste, quando estas respeitam a definição de formas estritamente paralelas. Se as variâncias das duas metades forem muito diferentes, a estimativa da fiabilidade do teste na sua globalidade corre o risco de ser errónea (Laveault & Grégoire, 2002).

O segundo problema com a estimativa da fiabilidade resultante da forma de divisão dos itens é ainda mais sério. É possível conceber várias metades (por exemplo itens impares *vs.* itens pares como na proposta inicial de Spearman) e nada nos garante que os resultados fossem os mesmos (raramente são…). Os cálculos de consistência são, assim, afectados pela forma de divisão dos itens e qualquer coeficiente de fiabilidade calculado desta forma é, em certo, grau incorrecto (Cronbach & Shavelson, 2004). Um processo possível de ultrapassar este problema, seria então o de conceber todas as metades possíveis, e estabelecer as diferentes relações entre essas metades, computando a sua média como um índice de consistência interna. Kuder e Richardson (1937) tentaram clarificar a dispersão dos cálculos da consistência provocados pela multiplicidade de modos de divisão do teste em partes paralelas e propuseram dois índices que se distinguiram como medida de consistência interna: *KR 20* e *KR21*.

No caso dos itens serem dicotómicos (e.g., "Certo" e "Errado"; "Sim" e "Não" codificados respectivamente como $X_{ij}=1$e $X_{ij}=0$ onde $i=1, \ldots, n$ representa os n sujeitos avaliados nos j=1, … k itens da escala ou teste) a consistência interna é dada pela fórmula 20 de Kuder e Richardson:

$$KR20 = \frac{k}{k-1}\left[1 - \frac{\sum_{j=1}^{k} p_j q_j}{S_T^2}\right] \tag{10}$$

Onde p_j é a proporção de "1" do item j ($j=1, \ldots, k$) (se "1" indicar a resposta correcta, ou a presença atributo de interesse, p_j reflecte o coeficiente de dificuldade do item) e $q_j=1-p_j$. A expressão p_jq_j estima variância do item j e S^2_T é a variância do total da escala. Se os itens tiverem sensivelmente o mesmo coeficiente de dificuldade e a mesma variância, a fiabilidade pode ser estimada pela fórmula *KR21* de cálculos mais simples (e generalizável) uma vez que depende apenas da média do teste ($\bar{X}$) e da variância dos resultados totais:

$$KR21=\frac{k}{k-1}\left[1-\frac{\bar{X}(k-\bar{X})}{kS_T^2}\right] \tag{11}$$

Se os itens tiveram graus de dificuldade muito diferentes o KR21 dá tendencialmente resultados inferiores ao *KR20*.

Em virtude da sua simplicidade e em particular da proposição de que a fiabilidade podia ser determinada pela aplicação singular do instrumento, a KR20 foi adaptada rapidamente pela maioria dos cientistas das ciências sociais no cálculo da fiabilidade. Contudo, apenas quando os pressupostos do método se verificam, esta formula estima de forma consistente a verdadeira fiabilidade (Φ). Em particular, se as covariâncias inter-itens não forem homogéneas, a *KR20* sub-estima de forma consistente a verdadeira fiabilidade e, contrariamente ao objectivo inicial, exige a aplicação do instrumento a pelo menos duas amostras independentes. Atento a este problema, e numa tentativa de unificar o conceito de fiabilidade em torno da ideia original de Sperman, L. Guttman (1945) derivou seis fórmulas que permitem estimar um limite inferior para a Φ a partir de uma única aplicação do instrumento de medida repetidamente a um único sujeito (Zimmerman, Williams, Zumbo, & Ross, 2005).

De entre as seis fórmulas propostas por Guttman, destacam-se os λ_2, λ_3 e λ_4.

$$\lambda_2=\lambda_1+\frac{\sqrt{\frac{2k}{k-1}\sum_{j=1}^{k}\sum_{i=1}^{k}(r_{ji}S_jS_i)^2}}{S_T^2} \tag{12}$$

Onde $\lambda_1=1-\frac{\sum_{j=1}^{k}S_j^2}{S_T^2}$ é um cálculo auxiliar na determinação de λ_2. O λ_3, é uma extensão da *KR20*, o que segundo Guttman (1945) é pura coincidência:

$$\lambda_3=\frac{k}{(k-1)}\left[1-\frac{\sum_{j=1}^{k}S_j^2}{S_T^2}\right] \tag{13}$$

O coeficiente λ_4 (Fiabilidade split-half de Guttman) é

$$\lambda_4=\frac{2(S_T^2-S_{T_1}^2-S_{T_2}^2)}{S_T^2} \tag{14}$$

Guttman recomenda que se experimente com a divisão em duas metades (1 e 2) do instrumento que maximize λ_4, usando depois o maior dos λ_2 e λ_3 como estimativa do limite inferior da fiabilidade. Num estudo posterior, e reconhecendo que o pressuposto de independência dos erros da teoria clássica de media é irrealista em muitas situações, Guttman (1953) expandiu as suas fórmulas de forma a considerarem erros de medida correlacionados apesar de estas fórmulas serem actualmente pouco usadas (Zimmerman et al., 2005).

O alfa de Cronbach: Uma estimativa estatística da consistência interna

L. J. Cronbach publica, em 1951, um artigo quase enciclopédico onde discute os problemas associados à estimação da consistência interna de uma escala ou teste e as propostas de outros autores para o seu cálculo. Neste artigo seminal, Cronbach considerando as derivações anteriores de Kuder-Richarson e Guttman, e assumindo os mesmos pressupostos mas sem limites no padrão de classificação dos itens, formaliza uma proposta de estimativa de consistência interna a partir das variâncias dos itens e dos totais do teste por sujeito, que ficou conhecida como o índice "alfa" de Cronbach.

A fórmula proposta por Cronbach é:

$$\alpha = \frac{k}{(k-1)} \times \left[1 - \frac{\sum_{j=1}^{k} S_j^2}{S_T^2} \right] \tag{15}$$

onde k é o número de itens do instrumento, $S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ é a variância do item j *(j= 1, ..., k)* e S^2_T é a variância dos totais da escala definida em (8).

Esta fórmula é uma aplicação particular do coeficiente de correlação intra-classes popularizado na década de 40 por R. A. Fisher em aplicações biométricas e é ubiquamente conhecida por α *de Cronbach* apesar de este coeficiente não ser mais do que uma generalização do *KR20* proposto alguns anos antes por Kuder e Richardson (1937) para itens dicotómicos[2]. Nos últimos 50 anos, o α de Cronbach, tem satisfeito a função que os psicometristas procuravam desde os primeiros trabalhos de Spearman e Brown, para uma medida valida de consistência interna e é a medida de consistência, compreendida ou não, usada por excelência. Curiosamente, como refere Cronbach e Shavelson (2004), a designação de "alfa" (inicialmente Alfa de Kuder-Richardson) pretendia apenas reflectir a convicção do autor de que esta fórmula é simplesmente a primeira de um conjunto de cálculos necessários para avaliar as propriedades de uma escala para além da fiabilidade. Usando a soma de variâncias, o α de Cronbach é algebricamente idêntico ao λ_3 de Guttman. Contudo, Guttman derivou os seus lambdas como uma estimativa do limite inferior da fiabilidade, impondo que, para que estas formas estimassem a verdadeira fiabilidade, era necessário que todas as variâncias-covariâncias inter-itens fossem homogéneas.

Porque a fórmula do α de Cronbach não faz assumpções relativas à homogeneidade das variâncias-covariâncias inter-itens, este índice subestima a verdadeira consistência de uma medida (veremos adiante outros factores que provocam a subestimação do α). Na realidade, raramente os itens de um instrumento apresentam a mesma variabilidade e/ou grau de dificuldade, pelo que o α de Cronbach tende a subavaliar a fiabilidade total de uma medida, estimando de forma conservadora a verdadeira fiabilidade. Uma forma de lidar com este problema é a de promover tal homogeneidade por estandardização dos itens antes de calcular o índice ou a de trabalhar directamente com coeficientes de correlação (covariância estandardizada), o que resulta num índice α *de Cronbach estandardizado*.

O α de Cronbach standardizado é calculado a partir da correlação média (r) dos $k(k\text{-}1)/2$ coeficientes de correlação não-redundantes inter-itens:

$$\alpha' = \frac{k \times \bar{r}}{1 + (k-1) \times \bar{r}} \tag{16}$$

que é a fórmula (8) de Sperman-Brown sobre o pressuposto de homogeneidade de correlações inter-itens. A fórmula (16) ilustra o facto de que o α, que deve variar entre 0 a 1, poder ser inferior 0, o que acontece quando a correlação média entre os itens é negativa.

---

[2] A variância do item dada por $S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ reduz-se a $S^2_j = p_j q_j$ se X for uma variável dicotómica com realizações "0" e "1" sendo p_j a proporção de "1" no item j. Substituindo $\sum_{j=1}^{k} p_j q_j$ por $\sum_{j=1}^{k} S_j^2$ por em (13) obtém-se (15).

O que indica um determinado valor de alfa de Cronbach?

O índice α *estima quão uniformemente os itens contribuem para a soma não ponderada do instrumento*, variando numa escala de 0 a 1. Esta propriedade é conhecida por *consistência interna da escala*, e assim, o α pode ser interpretado como coeficiente médio de todos as estimativas de consistência interna que se obteriam se todas as divisões possíveis da escala fossem feitas (Cronbach, 1951). Cortina (1993) descreve outras interpretações para o índice de Cronbach, referindo que o α é uma medida estável de fiabilidade pois não está sujeito à variabilidade resultante da forma como o instrumento ou teste é dividido para calcular a fiabilidade *split-half*. Do que foi apresentado até agora torna-se claro que quanto mais elevadas forem as covariâncias (ou correlações entre os itens) maior é a homogeneidade dos itens e maior é a consistência com que medem a mesma dimensão ou constructo teórico. Por outro lado a consistência interna estima a fiabilidade de um instrumento porque quanto menor é a variabilidade de um mesmo item numa amostra de sujeitos, menor é o erro de medida que este possui associado (ver e.g., Pasquali, 2003). Assim, quanto menor for a soma das variâncias dos itens [o numerador das fórmulas (10), (13) e (15)] relativamente à variância total dos sujeitos, mais o coefeciente se aproxima de 1, significando que mais consistente e, consequentemente, mais fiável é o instrumento. De acordo com esta definição, o α é, por vezes, interpretado como uma medida de saturação de um factor comum (ou constructo latente) de primeira ordem (i.e., uma medida do grau em que um único factor latente motiva a correlação entre todos os itens de uma escala). Contudo, como referimos anteriormente, ainda que um conjunto de itens com α elevado, defina a presença 'forte' de factor comum, um α elevado não demonstra a presença de uma escala uni-factorial. *Pelo que sendo o* α uma medida de fiabilidade ele *não nos informa sobre dimensionalidade.*

De um modo geral, um instrumento ou teste é classificado como tendo fiabilidade apropriada quando o α é pelo menos 0.70 (Nunnally, 1978). Contudo, em alguns cenários de investigação das ciências sociais, um α de 0.60 é considerado aceitável desde que os resultados obtidos com esse instrumento sejam interpretados com precaução e tenham em conta o contexto de computação do índice (DeVellis, 1991). Peterson (1994) numa meta-análise da utilização do α de Cronbach na literatura das ciências sociais e humanas, observou um α médio de 0.70 (na medição de valores) a 0.82 (na medição da satisfação com o trabalho). Com poucas excepções, este autor não observou nenhuma relação entre a magnitude do α e o design experimental das características investigadas. A Tabela 1, resume os níveis de fiabilidade recomendados por diversos autores, que pelo exposto anteriormente, deve servir como uma base de partida e não como critério definitivo de classificação.

Tabela 1

Critérios de recomendação de Fiabilidade estimada pelo α de Cronbach (adaptado de Peterson, 1994)

Autor	Condição	α considerado aceitável
Davis, 1964, p. 24	Previsão individual	Acima de 0.75
	Previsão para grupos de 25-50 indivíduos	Acima de 0.5
Kaplan & Sacuzzo, 1982, p. 106	Investigação fundamental	0.7-0.8
	Investigação aplicada	0.95
Murphy & Davidsholder, 1988, p. 89	Fiabilidade inaceitável	<0.6
	Fiabilidade baixa	0.7
	Fiabilidade moderada a elevada	0.8-0.9
	Fiabilidade Elevada	>0.9
Nunnally, 1978, p. 245-246	Investigação preliminar	0.7
	Investigação fundamental	0.8
	Investigação aplicada	0.9-0.95

Como referimos anteriormente o cálculo do α de Cronbach permite que este assuma valores negativos quando as correlações inter-itens são, elas próprias, negativas. *Um α' negativo reflecte normalmente um erro sério na codificação* dos pontos dos itens e a solução passa pela recodificação (inversão) dos pontos de forma a assegurar que todos os itens estão codificados na mesma direcção conceptual. Adicionalmente, *um α muito baixo pode reflectir a codificação errada de itens ou a mistura de itens de dimensões diferentes exigindo a reavaliação da base teórica que motivou a construção da escala.*

A fiabilidade do α de Cronbach: Computação de intervalos de confiança

Como já referimos, o índice α de Cronbach é uma estimativa "lower-bound" da fiabilidade de uma medida (ver exemplo, Crocker & Algina, 1986), pelo que, a verdadeira estimativa de fiabilidade dos dados actuais tem baixa probabilidade de ser mais pequena e elevada probabilidade de ser muito maior do que o valor reportado. Mas tal não significa que o índice associado a uma única medida não possa estar a sobre-estimar o que ocorrerá num segundo momento de mensuração. Apenas quer dizer que a distribuição da estimativa está centrada abaixo do verdadeiro valor de fiabilidade da medida. Qual poderá então ser esse valor?

Estando toda a estimativa estatística sujeita a erro – isto é qualquer a medida está impregnada de variância por explicar, a estimativa do índice de Cronbach não é nenhuma excepção. Quando o investigador possui ao seu dispor a inferência estatística, deixa de ser suficiente reportar a fiabilidade com base numa única estimativa pontual de α. O intervalo de confiança apresenta maior informação de diagnóstico da fiabilidade e por isso o seu cálculo é recomendado por diversas publicações periódicas (por exemplo a *Educational and Psychological Measurment*; Fan & Thompson, 2001).

Hoydt (1941) demonstrou que o α pode ser expresso como uma função simples dos quadrados médios dos sujeitos (QM_S)e dos quadrados médios dos itens x sujeitos (QM_{SxI}), obtidos de uma de ANOVA em blocos casualizados. Assim, o α pode estimar-se como:

$$\alpha = \frac{QM_S - QM_{SxI}}{QM_S} = 1 - \frac{QM_{SxI}}{QM_S} \tag{17}$$

A partir deste resultado, e sabendo que o rácio de quadrados médios apresenta distribuição F-Snedecor, Kristof (1963) e Feldt (1965) demonstram que $\hat{\alpha} \sim 1-(1-\alpha)F_{[(k-1)\,(n-1);\,(n-1)]}$ se os itens apresentarem distribuição normal multivariada com matriz de variâncias-covariâncias homogéneas (simetria composta) (Feldt, 1990). Um intervalo de confiança exacto para α a $(1-\gamma)x100\%$ pode então ser estimado por (para uma dedução mais recente deste intervalo ver Kistner & Muller, 2004):

$$\left]1-(1-\hat{\alpha})\times f_{1-\gamma/2;[(n-1),(n-1)(k-1)]};1-(1-\hat{\alpha})\times f_{\gamma/2;[(n-1),(n-1)(k-1)]}\right[\tag{18}$$

Onde $\hat{\alpha}$ é a estimativa amostral do α e $f_{\gamma/2;\,[(n-1),\,(n-1);\,(k-1)]}$ e $f_{1-\gamma/2;\,[(n-1),\,(n-1);\,(k-1)]}$ são os quantis da F-Snedecor nos percentis $\gamma/2$ e $1-\gamma/2$ com $(n-1)$ e $(k-1)\,(n-1)$ graus de liberdade respectivamente.

O estudo das características distribucionais do α de Cronback permite igualmente o desenvolvimento de estatística inferencial e o teste de hipóteses relativas à magnitude do valor α. É assim possível testar a probabilidade de erro associada à afirmação de que o teste tem um coeficiente de fiabilidade igual ou superior a, por exemplo, 0.70.

Como descrito em Feldt (1965) e mais recentemente em Charter e Feldt (1996), um teste de hipóteses a $H_0{:}\alpha=\alpha_0$ *vs.* $H_1{:}\alpha\neq\alpha_0$ pode fazer-se com a estatística de teste:

$$W = \frac{1-\alpha_0}{1-\hat{\alpha}} \sim F[(n-1),(k-1)(n-1)] \tag{19}$$

Se, para um nível de significância γ, $W \leq f_{\gamma/2;\,[(n-1),\,(n-1);\,(k-1)]}$ ou se $W \geq f_{1-\gamma/2;\,[(n-1),\,(n-1);\,(k-1)]}$ rejeita-se H_0. É contudo de referir que o teste bilateral tem interesse reduzido uma vez que de uma forma geral estamos interessados em que o nosso α seja superior a um valor limite (0.6 ou 0.7, ver Tabela 1) para aceitar o instrumento como fiável. O teste $H_0: \alpha \leq \alpha_0$ *vs.* $H_1: \alpha > \alpha_0$ onde $\alpha_0 = 0.7$ (ver e.g., Nunnaly & Bernstein, 1994; Fan & Thompson, 2001; Iacobucci & Duhacheck, 2003) é de aplicação mais generalizada. Naturalmente, rejeita-se a H_0 se $W \geq f_{1-\gamma;\,[(n-1),\,(n-1)\,(k-1)]}$.

Investigação mais recente sobre as propriedades distribucionais do α (van Zyl et al., 2000; Kistner & Muller, 2004) demonstraram que (13) é o estimador de máxima verosimilhança de α quando os itens apresentam distribuição normal multivariada e simetria composta. À medida que $n \to \infty$, a estatística $\sqrt{n}(\hat{\alpha}-\alpha) \sim N(0,\sqrt{Q})$ onde $\hat{\alpha}$ é o estimador de máxima verosimilhança de α e Q é a variância dada, em forma matricial, por:

$$Q = \left[\frac{2k^2}{(k-1)^2\,(\mathbf{1}'\Sigma\mathbf{1})^3}\right] \times \left[(\mathbf{1}'\Sigma\mathbf{1})\,tr\Sigma^2 + tr^2\Sigma - 2(tr\Sigma)(\mathbf{1}'\Sigma^2\right. \tag{20}$$

Onde $\mathbf{1}'_{1xk} = [1\ 1\ 1\ \ldots\ 1]$ é um vector de k *1's*, $\boldsymbol{\Sigma}$ é a matriz de covariâncias populacionais entre os itens (geralmente estimada pela matriz de covariâncias amostrais **S**) e *tr* é a função traço (soma dos elementos diagonal de uma matriz). Esta expressão, em forma algébrica pode exprimir-se como:

$$Q = \left[\frac{2k^2}{(k-1)^2\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}\right)^3}\right] \times \left[\left(\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}\right)\left(\sum_{i=1}^{k}\sum_{l=1}^{k}\sigma_{il}\sigma_{il} + \sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}\right)^2 - 2\sum_{i=1}^{k}\sigma_{il}\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma_{il}\sigma_{lj}\right] \tag{21}$$

Onde σ_{ij} é o elemento *ij* da matriz $\boldsymbol{\Sigma}$. No caso do α estandardizado (α') e sobre os mesmos pressupostos anteriores, a expressão 17 simplifica-se a (Duhachek et al. 2005):

$$Q' = \frac{2k(1-\bar{r})^2}{(p-1)[1+\bar{r}(p-1)]^2} \tag{22}$$

Intervalos de confiança assimptóticos a $(1-\gamma)x100\%$ para α (e α' substituindo Q por Q') podem então obter-se com a expressão[3].

$$]\hat{\alpha} - z_{1-\gamma/2} \times \sqrt{\frac{Q}{n}}\,;\hat{\alpha} + z_{1-\gamma/2} \times \sqrt{\frac{Q}{n}}\,[\tag{23}$$

Onde $\sqrt{Q/n}$ é o erro-padrão de. A estatística de teste para o teste de hipóteses a α é então

$$Z = \frac{(\hat{\alpha}-\alpha_0)}{\sqrt{\frac{Q}{n}}} \sim N\,(0,1) \tag{24}$$

A rejeição de H_0 ocorre quando o valor absoluto de Z for superior ou igual ao quantil da $N(0,1)$ no percentil $1-\gamma/2$ no caso do teste bilateral ou quando Z for superior ou igual ao quantil da $N(0,1)$ no percentil $1-\gamma$ no caso do teste unilateral à direita.

[3] Iacobucci e Duhacheck (2003) apresentam em Anexo Macros de SPSS e SAS para calcular o α, o erro-padrão do α e o intervalo de confiança.

Duhacheck e Iacobucci (2004) compararam propostas alternativas de outros autores para o cálculo do intervalo de confiança para o α assumindo a validade dos pressupostos descritos e, numa série de simulações de Monte-Carlo, apresentam resultados que demonstram a superioridade das fórmula (23) (em particular para amostras de grande dimensão) e (18) (em particular para amostras de dimensão moderada) relativamente a outras alternativas de cálculo.

Factores que afectam a magnitude do índice de fiabilidade: variabilidade e simetria da distribuição

Porque as características da variância observada nos dados é a base de inferência de uma estimativa de fiabilidade, depreende-se que as características dos participantes utilizados num estudo afectam a fiabilidade de uma dada medida:

> (…) A mesma medida, quando administrada a uma amostral de sujeitos mais homogéneos ou mais heterogéneos produzirá scores com diferentes fiabilidades (p. 839, Thompson, 2002).

Assim sendo, todas as características dos contextos de recolha dos dados que estejam directa ou indirectamente relacionadas com uma maior variabilidade observada nos dados (quer intra quer inter) afectam igualmente o valor do índice de Cronbach. De uma forma geral quanto menor a variabilidade das resposta intra-sujeitos e maior a variabilidade das respostas inter-sujeitos, maior o α. Por outro lado o α é, geralmente, maior quando existe homogeneidade de variâncias inter-itens do que quando não existe.

Sabendo que o número de observações são um factor que influencia a variabilidade observada (quanto menor a dimensão da amostra maior será a estimativa da sua variância) é assim de esperar que instrumentos de medida com um maior número de itens tenham valores de α superiores e de erro-padrão inferiores aos instrumentos com um menor número de itens (ver por exemplo, Brown, 2001).

Em termos gerais, os instrumentos cujos resultados se apresentam normalmente distribuídos (e.g., escalas construídas com a metodologia de Likert) têm valores de α superiores aos associados a distribuições assimétricas. Contudo, e no capítulo da inferência sobre o α, Yuan e Bentler (2002) demonstraram, na sua exploração extensiva dos efeitos do enviesamento e achatamento, que estes índices são razoavelmente robustos à violação do pressuposto da normalidade multivariada. A validade do pressuposto da simetria composta pode ser avaliada pelo teste M de Box (ver e.g., Maroco, 2003, pp. 157-158). Porém, e à semelhança de outros testes de ajustamento, este teste é sensível a desvios da normalidade e para amostras de grande dimensão, mesmo pequenos desvios entre as variâncias-covariâncias levam à rejeição do pressuposto de homogeneidade (acréscimo do erro de tipo I). Por outro lado, a presença de covariâncias heterogéneas não apresenta um efeito considerável sob a estimação do α mas aumenta o erro-padrão da estimativa. Finalmente, a heteroscedasticidade de variâncias provoca a redução do α com um aumento reduzido do erro-padrão da estimativa (Duhacheck & Iacobucci, 2004).

Assim sendo os *valores de α devem sempre ser interpretados à luz das características da medida a que se associa, e da população onde essa medida foi feita*. Contudo, e apesar das limitações à estimação da fiabilidade pelo α de Cronbach, este permanece a medida mais usada da fiabilidade de um instrumento.

Limitações do alfa de Cronbach e propostas alternativas

O facto do índice de Cronbach apresentar enviesamentos para estimativas inferiores à verdadeira fiabilidade de uma medida, faz com que outras propostas surjam no campo. Cronbach em colaboração com outros autores (Cronbach, Rajaratnam, & Gleser, 1963) rapidamente se aperceberam das limitações do α em particular face aos pressupostos restritivos que a sua aplicação exigia, e que, de um modo geral, são difíceis de realizar. Assim, um novo desenvolvimento da teoria da fiabilidade foi proposto com base na análise das propriedades aditivas dos modelos de análise de variância e do coeficiente de correlação inter-classes. Contudo, devido à complexidade desta nova teoria, designada por teoria da generabilidade, e à falta de procedimentos para estimar de forma eficiente muitos dos seus parâmetros, esta não tem assumido relevância prática e a sugestão do seu uso cauteloso continua em voga (Weiss & Davidson, 1981; Jones & Applebaum, 1989).

O α sub-estima a verdadeira fiabilidade principalmente no caso em que o instrumento define uma escala multifactorial (Cortina, 1993; Osbourn, 2000). Tal acontece uma vez que o α requer poder discriminante equivalente entre itens e unidimensionalidade da escala (representada por pesos factoriais iguais para todos os itens no modelo unifactorial analítico; Komaroff, 1997; McDonald, 1999)[4]. Osbourn (2000) e Kamata et al. (2003), numa série de estudos de simulação de Monte-Carlo, testaram vários estimadores alternativos de fiabilidade em diferentes escalas multidimensionais. Os seus resultados demonstraram que os métodos do alfa estratificado (25) e da máxima fiabilidade (26) são os melhores estimadores da verdadeira fiabilidade. Em particular, o alfa estratificado é o que apresenta melhor performance em condições de multidimensionalidade apesar das diferenças entre os dois estimadores não serem consideráveis (Kamata et al., 2003). É de referir porém, que num contexto real, Feldt e Qualls (1996), observaram que, em média, as duas formulas diferem em aproximadamente 1% nas suas estimativas de consistência interna de testes de aptidão matemática (conceitos e problemas).

O alfa estratificado foi proposto por Cronbach, Shonenman, e McKie (1965) para instrumentos cujos itens podem ser agrupados em *f* sub-testes ou factores de acordo com o seu conteúdo.

O índice alfa estratificado é definido como:

$$\alpha_{Estr} = 1 - \frac{\sum_{i=1}^{f} S_i^2 (1 - \alpha_i)}{S_T^2} \tag{25}$$

onde S^2_i é a variância dos itens que constituem o factor *i (i=1, ..., f)*, α_i é o α de Cronbach para o factor *i* e S^2_T é a variância do total do instrumento.

O estimador de máxima fiabilidade foi deduzido por Li et al. (1996) e assume que um instrumento ou escala é constituído por vários factores ou sub-escalas onde (*a*) os itens que constituem cada uma das sub-escalas são paralelos, i.e., apresentam a mesma fiabilidade e variância e (*b*) os itens nas diferentes sub-escalas podem apresentar diferentes fiabilidades e variâncias (Osbourn, 2000)[5].

O estimador de máxima fiabilidade é uma extensão da fiabilidade de Spearman-Brown para *K* factores onde o factor *i* (*i*=1, …, *K*) é constituído por k_i itens paralelos:

---

[4] Esta condição é conhecida por *tau*-equivalente. Por definição, a condição *tau*-equivalente é necessária, mas não suficiente, para que o α seja um estimador não enviesado da fiabilidade. Esta condição requer que os scores verdadeiros (τ) para duas aplicações do teste difiram apenas por uma constante como ilustrado pelas equações (14) (Lord & Novic, 1968).

[5] Se estas duas condições são válidas, os itens de todas as sub-escalas dizem-se congenéricos. As equações correspondentes são com $X_1=\beta_1\tau+\varepsilon_{x1}$ e $X_2=\beta_2\tau+\varepsilon_{x2}$ com $\beta_1 \neq \beta_2$ e $V(\varepsilon_{x1}) \neq V(\varepsilon_{x2})$.

$$F_M = \frac{\frac{k_1 r_1}{1-r_1} + \ldots + \frac{k_K r_K}{1-r_K}}{\frac{K}{1+(K-1)\rho} + \frac{k_1 r_1}{1-r_1} + \ldots + \frac{k_K r_K}{1-r_K}} \quad (26)$$

Onde r_i é a fiabilidade da sub-escala i e ρ é a correlação comum entre as sub-escalas. Para duas sub--escalas, $\rho = r_{12}/\sqrt{r_1 r_2}$ onde r_{12} é a correlação média entre os itens da sub-escala 1 com os itens da sub--escala 2. Para mais de duas sub-escalas, ρ é estimado pela média de todas os ρ's calculados para as sub-escalas duas-a-duas.

Partindo da generalização do modelo de medida em uso na Análise factorial, McDonald (1999) define um novo índice de fiabilidade ω como sendo o rácio da variância estimada e da variância total de um modelo de medida unifactorial.

No caso multidimensional, o modelo factorial de medida é

$$\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E} \quad (27)$$

Onde **X** é a matriz *nxk* dos scores observados dos *n* sujeitos nos *k* itens, **F** é matrix *nxp* dos scores factoriais dos *n* sujeitos nos *p* factores, $\boldsymbol{\Lambda}$ é a matriz *kxp* dos pesos factoriais dos *k* itens nos *p* factores e **E** é a *kxn* matriz dos erros. O ω multidimensional é então dado por:

$$\omega_{MD} = \frac{\mathbf{1}'\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} \quad (28)$$

Onde $\mathbf{1}'_{1xk}$ é um vector de k 1's, $\boldsymbol{\Phi}$=Cov(**F**) e $\boldsymbol{\Sigma}$ é a matriz de covariâncias estimada na amostra.

Kamata et al. (2003) demonstram que o ω_{MD} é ligeiramente superior ao α_{Estr} e ao F_M em particular quando a correlação entre os factores é reduzida (<0.5), chamando porém à atenção que o ω_{MD} pode sobre-estimar a verdadeira fiabilidade.

A subestimação do α de Cronbach é também severa quando os itens são dicotómicos (e.g., "Correcto" *vs.* "Incorrecto", ou "Sim" *vs.* "Não") porque o coeficiente de correlação entre itens dicotómicos (coeficiente Phi) tende a subestimar a correlação populacional. Sun, Chou, Stacy, Ma, Unger, e Gallaher 2006, propõem assim o cálculo do α' calculado para itens dicotómicos a partir dos coeficiente de correlação médio inter-itens utilizando o limite superior do coeficiente ϕ.

Este limite é dado por:

$$\phi_{MAX} = \sqrt{\frac{p_i(1-p_i)}{p_j(1-p_j)}} \quad (29)$$

Onde p_i e p_j são as proporções do sucesso (e.g., proporção da realização "sim") dos itens i e j respectivamente, organizados de forma que $p_i \geq p_j$.

Em estudos de simulação de Monte Carlo preliminares, Sun et al. (2006) demonstram que o α' calculado a partir de ϕ subestima seriamente a consistência interna enquanto que o α' calculado a partir de ϕ_{MAX} tem tendência a sobre-estimar a consistência interna. Provavelmente, o verdadeiro valor de consistência interna encontrar-se-á entre as duas estimativas.

Consequências de uso de dados com fraca fiabilidade

São essencialmente duas as consequências directas de uso de dados com fraca fiabilidade: (a) Existe elevada probabilidade da medida não ser valida – O resultado pode nada dizer sobre o constructo que se pretendia medir. Mas mesmo se a medida for válida, (b) O erro de medida é elevado, pelo que a variabilidade observada afecta o poder de qualquer teste estatístico realizado, aumentando a probabilidade de resultados não-significativos. [Ver Wilkinson e a APA Task force on Statistical Inference (1999) "interpreting the size of observed effects requires an assessment of the reliability of the scores" p. 596].

Considerações finais: Datametria *vs.* psicometria

Em 1999 um conjunto de autores liderados por Wilkinson autodesignaram-se de *APA Task Force for Statistical Inference* produzindo um documento que referencia algumas questões estatísticas que têm sido mal interpretados pelos investigadores em psicologia. Com o objectivo de introduzir alguma "ordem no caos" este documento (que pode ser consultado nas páginas do site da APA) chama a atenção para alguns cuidados a ter com o uso dos métodos estatísticos. Relativamente aos métodos de estimativa da fiabilidade de um instrumento de medida o documento refere:

> (...) é importante notar que um teste não é fiável ou não-fiável (...) assim sendo os autores devem fornecer os coeficientes de fiabilidade dos dados a serem analisados, mesmo quando o foco da sua pesquisa não é psicométrico (p. 570).

Ao fazer esta proposta Wilkinson e colaboradores parecem adoptar a posição explicitamente tomada por Thompson (1994) e Vacha-Haase (1998) que referem que "nenhum teste ou instrumento é fiável" apenas os "dados são fiáveis ou não-fiáveis". Segundo estes autores apenas se pode falar de "score reliaibility" e é errado inferir a partir de uma estimativa de α de Cronbach que este é ou não fiável.

Em 2000, Shlomo S. Sawilowsky reage fortemente a esta posição, e num artigo intitulado "Psychometrics Versus Datametrics: Comment on Vacha-Haase's "Reliability Generalization" Method and Some EPM Editorial Policies", contra-argumenta a afirmação de que não podemos inferir a fiabilidade de um instrumento a partir da estimativa de um α de Cronbach.

Referindo, "Test reliability is psychometric terminology and score reliability is a score-centric termonilogy (datametrics)", Sawilowsky contrapõe a visão clássica psicométrica com a sugestão de Thompson e Vacha-Haase. O argumento básico de Sawilowsky é o comportamento daqueles que usam instrumentos de medida. Ele analisa um grande número de analises de fiabilidade de uma medida e que apenas 17,5% dos autores assumem uma posição de datametria, seguindo a maioria explicitamente uma visão psicométrica. Apenas assim, faz sentido o facto destes autores utilizarem a informação relativa á fiabilidade como critério de selecção de uma escala como melhor que outra.

Numa atitude sensata podemos ter em conta alguns dos argumentos da posição "psicometrica" e "datamétrica" no nosso comportamento face a estimativas de fiabilidade. Vejamos então os pressupostos em que nos podemos basear para sustentar qualquer posição:

a) Um α de Cronbach é apenas uma estimativa da fiabilidade dos dados obtidos com um dado instrumento (datametria);

b) A fiabilidade dos dados é afectada pela precisão do instrumento utilizado para medir (psicometria) (apesar de existirem outras variáveis capazes de afectarem a fiabilidade dos nosso dados, existem "balanças calibradas e não-calibradas");

c) A utilização de uma única estimativa de fiabilidade como base para concluir sobre um instrumento é sujeita a erro, visto que qualquer estimativa está igualmente sujeita a erro;

d) Podemos "estimar" o erro da inferência estatística associada ao uso da informação sobre a fiabilidade dos nossos dados para concluir sobre a fiabilidade do instrumento com base na estimação de intervalos de confiança;

e) Só o uso repetido do instrumento com diferentes amostras nos indica algo sobre a validade do processo inferencial: um instrumento que repetidamente gera dados fiáveis pode dizer-se, com maior confiança, fiável.

Assim sendo, defendemos a posição de Wilkinson e da *APA Task Force for Statistical Inference* (1999), de que qualquer utilização de um instrumento deve implicar a análise das suas características métricas, visto que nos diz algo sobre a validade dos dados a serem interpretados. Mas, consideramos que apesar de um instrumento fiável poder gerar dados não-fiáveis em dadas circunstancias (por exemplo, uma balança em circunstancias de fraca gravidade), a fiabilidade é uma característica, essencialmente associada ao instrumento de medida utilizado e à amostra onde este foi aplicado.

Sem dúvida alguma, o α de Cronbach é um instrumento útil para a investigação da fiabilidade de uma medida, e por tal permite o estudo da precisão de um instrumento. Contudo, é necessário ter em conta o que este instrumento é, e o que pretende medir, para que o seu uso seja eficaz e não induza a erros. O valor de fiabilidade estimado pelo α não é uma característica de um instrumento. É uma estimativa da fiabilidade dos dados obtidos que nos podem informar sobre a precisão do instrumento. Essa estimativa é sujeita a influencias várias, que devem ser tidas em conta na sua interpretação. Assim, a estimativa é sujeita a enviesamentos e erros vários. Não só se sabe ser a estimativa uma sub--avaliação (para o qual existem propostas de correcção), como é possível calcularmos os seus intervalos de confiança, e estes devem ser sempre reportados de forma a transmitir um nível de confiança à estimativa obtida e ao erro associado a esta estimativa. A extrapolação de uma estimativa de fiabilidade obtida com resultados associados a um dado estudo e dadas circunstâncias deve ser feita com a ponderação de um processo inferencial que se sabe sujeito a erro. Há que garantir a fiabilidade da estimativa de fiabilidade de uma medida.

Anexo 1

Uso do SPSS no calculo do alfa de Cronbach

Para ilustrar o cálculo do α de Cronbach com o SPSS, utilizaremos uma base de dados relativamente simples que envolve apenas 6 itens (Tabela 2). O instrumento utilizado é a escala de medida do "estado de espírito" (mood) desenvolvida por Garcia-Marques (ver Garcia-Marques, 2004) tendo sido avaliado em 24 sujeitos.

Tabela 2

Base de Dados usada no cálculo do α de Cronbach com o SPSS e com o Statistica

Sujeito	Negativo	Triste	Cansado	Aborrecido	Mal_bem	Tenso_rec
1	7	7	5	6	8	6
2	6	7	7	7	7	2
3	8	8	4	7	8	8
4	8	7	6	7	8	8
5	9	8	9	6	9	9
6	7	8	8	6	8	8
7	7	8	4	7	7	7
8	6	7	4	8	7	8
9	8	7	4	7	7	7
10	4	6	2	7	7	7
11	4	3	4	4	4	3
12	4	4	4	6	4	4
13	7	8	8	7	7	6
14	9	9	9	9	9	8
15	8	7	7	6	8	8
16	4	2	6	3	4	2
17	8	8	6	8	8	7
18	7	6	3	4	4	5
19	5	6	4	7	4	5
20	9	7	8	7	2	5
21	9	5	3	6	5	5
22	4	7	6	5	6	4
23	5	7	3	7	6	8
24	3	2	4	6	4	3

Para calcular o α é necessário calcular o total da escala por sujeito (i.e., a soma de todos os itens por sujeito) e a partir destes totais por sujeito calcular a variância total (S^2_T) e a variância de cada um dos j=1, ..., 6 itens (S^2_j). Estes cálculos resumem-se na tabela seguinte:

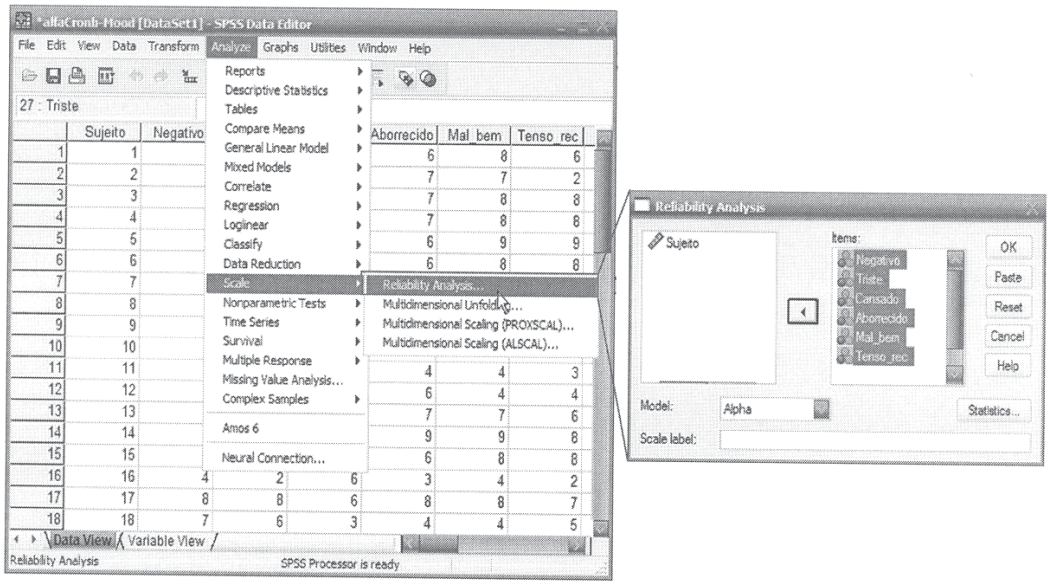
Sujeito	Negativo	Triste	Cansado	Aborrecido	Mal_bem	Tenso_rec	**Soma**
1	7	7	5	6	8	6	39
2	6	7	7	7	7	2	36
3	8	8	4	7	8	8	43
4	8	7	6	7	8	8	44
5	9	8	9	6	9	9	50
6	7	8	8	6	8	8	45
7	7	8	4	7	7	7	40
8	6	7	4	8	7	8	40
9	8	7	4	7	7	7	40
10	4	6	2	7	7	7	33
11	4	3	4	4	4	3	22
12	4	4	4	6	4	4	26
13	7	8	8	7	7	6	43
14	9	9	9	9	9	8	53
15	8	7	7	6	8	8	44
16	4	2	6	3	4	2	21
17	8	8	6	8	8	7	45
18	7	6	3	4	4	5	29
19	5	6	4	7	4	5	31
20	9	7	8	7	2	5	38
21	9	5	3	6	5	5	33
22	4	7	6	5	6	4	32
23	5	7	3	7	6	8	36
24	3	2	4	6	4	3	22
Variância	3.739	3.645	4.232	1.810	3.781	4.476	75.679

Usando agora (15):

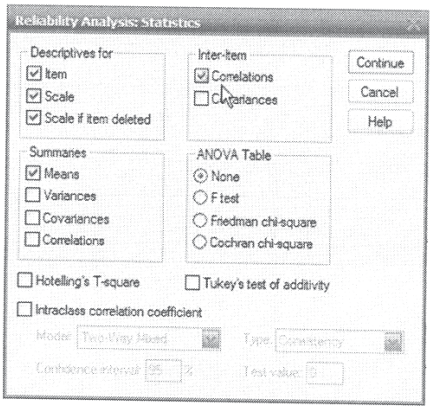
$$\iota = \frac{k}{(k-1)} \times \left[1 - \frac{\sum_{j=1}^{k} S_j^2}{S_T^2} \right] = \frac{6}{(6-1)} \times \left[1 - \frac{3.739 + 3.645 + 4.232 + 1.71 + 3.871 + 4.476}{75.679} \right] =$$

$$= 0.856$$

No SPSS, para calcular o α de Cronbach, recorra ao menu: **Analyze▶Scale▶Reliability analysis:**



Passe os 6 itens para a caixa "Itens" e seleccione o modelo "Model: Alfa". Clique no botão Statistics... para definir as estatísticas a calcular. Seleccione as opções "Descriptives" (para produzir as estatísticas descritivas para cada um dos itens), "Scale" (para produzir a estatística descritiva para o total da escala" e "Scale if item deleted" (para calcular a evolução do a se cada um dos itens for eliminado da análise). Seleccione ainda a opção "Correlations" na área "Inter-item" para calcular a correlação inter-itens e o R^2 (coeficiente de determinação linear) entre cada um dos itens e os restantes itens:



Clique no botão Continue e OK para obter os *Outputs*:

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.856	.861	6

Item Statistics

	Mean	Std. Deviation	N
Negativo	6.50	1.934	24
Triste	6.42	1.909	24
Cansado	5.33	2.057	24
Aborrecido	6.38	1.345	24
Mal_bem	6.29	1.944	24
Tenso_rec	5.96	2.116	24

Inter-Item Correlation Matrix

	Negativo	Triste	Cansado	Aborrecido	Mal_bem	Tenso_rec
Negativo	1.000	.683	.459	.426	.457	.590
Triste	.683	1.000	.428	.664	.704	.736
Cansado	.459	.428	1.000	.173	.355	.163
Aborrecido	.426	.664	.173	1.000	.488	.571
Mal_bem	.457	.704	.355	.488	1.000	.711
Tenso_rec	.590	.736	.163	.571	.711	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	6.146	5.333	6.500	1.167	1.219	.194	6

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
36.88	75.679	8.699	6

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Negativo	30.38	52.766	.683	.554	.825
Triste	30.46	49.042	.860	.759	.790
Cansado	31.54	59.216	.386	.362	.882
Aborrecido	30.50	61.478	.587	.466	.846
Mal_bem	30.58	51.993	.710	.612	.820
Tenso_rec	30.92	49.906	.712	.693	.819

Figura 1. α de Cronbach e estatística descritiva dos itens com o SPSS.

O quadro "Reliabity Statistics" apresente as estimativas do α de Cronbach e do α estandardizado que são, neste exemplo, 0.856 e 0.861 respectivamente. O quadro "Item Statistics" apresenta a média e o desvio-padrão de cada um dos 6 itens e o quadro Inter-item correlation matrix" apresente as correlações inter-intens. O sumário de todos os itens é apresentado no quadro "Summary Itens Statistics" enquanto que a estatística descritiva da escala (i.e., da soma dos itens para cada sujeito) é apresentada no quadro "Scale Statistics". Finalmente, o quadro Item-Total Statistics" apresenta o efeito da remoção de cada um dos itens no total da escala. Por exemplo, se o item "Negativo" fosse removido, a média da escala passaria a ser 30.38 e a variância 52.766. De maior interesse são as colunas com a correlação entre os *scores* do item e o total da escala (3ª coluna), o coeficiente de determinação múltipla (R^2) entre o item e os restantes itens da escala (4ª coluna), e o valor do α de Cronbach da escala se esse item fosse eliminado da escala (5ª coluna). No nosso exemplo, o item que está pior correlacionado quer com o total da escala quer com os outros itens é o item "Cansado". Se este item

for eliminado, o a de Cronbach passaria a ser 0.882 (em vez do 0.856 actual). Pelo contrário se o iten "Triste" fosse eliminado o novo α seria 0.790. A análise conjunta do R^2 e dos valores do "α se o item for eliminado" permite perceber qual a qualidade dos itens e o seu contributo para a consistência interna da escala. Naturalmente, podem eliminar-se os itens cuja remoção aumente consideravelmente o α da escala. É contudo de referir que outros critérios, nomeadamente a relevância do item, podem contrapor à sua remoção.

Finalmente, um Intervalo de confiança a 95% para α é dado por (18) uma vez que a amostra é de dimensão reduzida:

$$\left]1-(1-\hat{\alpha})\times f_{1-\gamma/2;[(n-1),(n-1)(k-1)]};1-(1-\hat{\alpha})\times f_{\gamma/2;[(n-1),(n-1)(k-1)]}\right[$$

$$\left]1-(1-0.856)\times f_{0.975;[23,115]};1-(1-0.856)\times f_{0.025;[23,115]}\right[$$

Consultando uma tabela da distribuição F e sabendo que $f_{0.025;[23,115]}=1/f_{0.975;[115,23]}$ vem

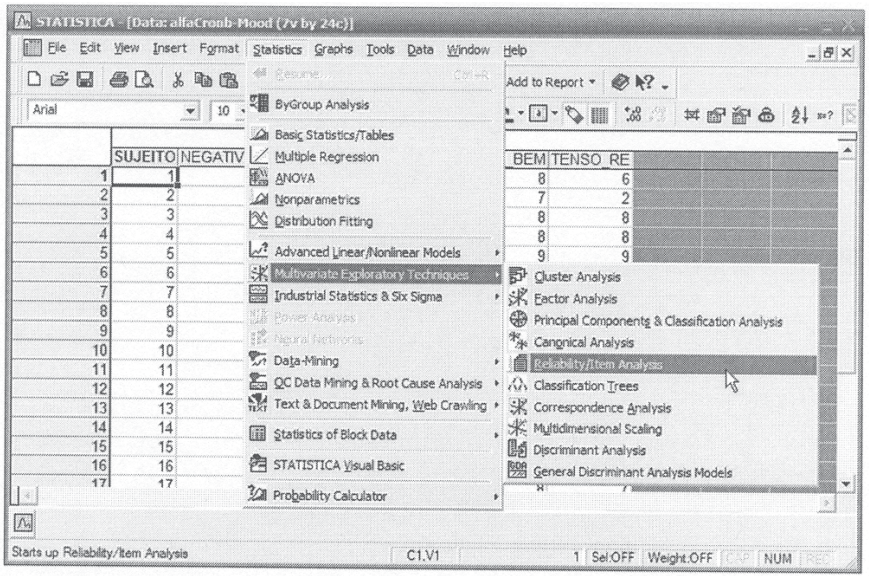
$$\left]1-(1-0.856)\times 1.779;1-(1-0.856)\times 0.489\right[$$

Um Intervalo de confiança a 95% para α é]0.744;0.930[. Nenhum dos softwares vulgarmente utilizados nas ciências sociais e humanas, estima um intervalo de confiança para o α, pelo que os cálculos tem de ser feitos de forma manual.

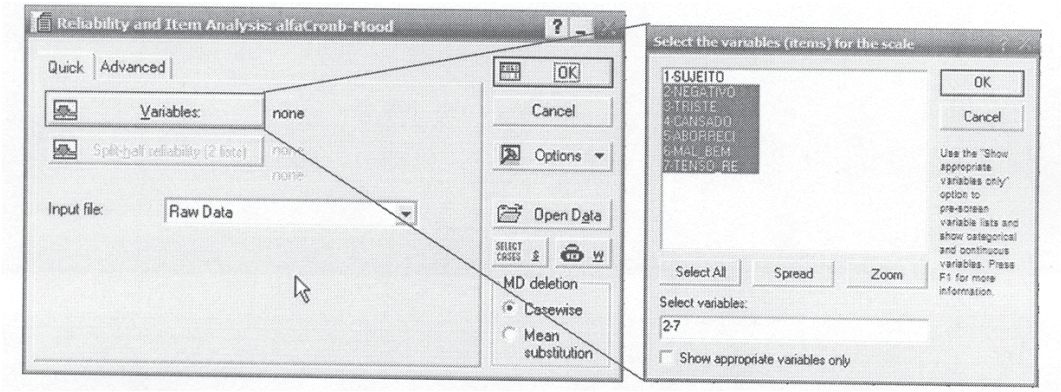
Anexo 2

Uso do Statistica 7 no cálculo de Índice de Cronbach

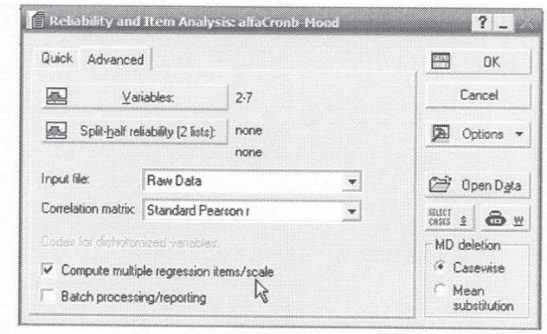
O cálculo do α no Statistica faz-se no menu **Statistics▶Multivariate Exploratory Techniques▶ Reliability/Item analysis:**



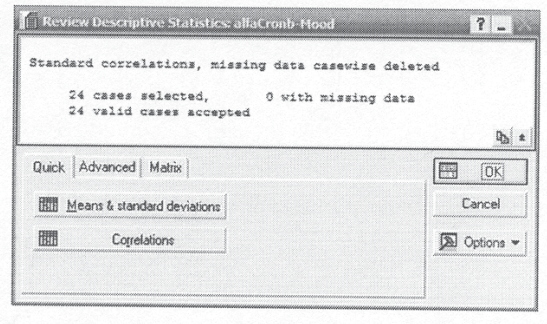
Depois clique no botão Variables: definir as variáveis (itens) a analisar. Selecione as variáveis 2 a 7:



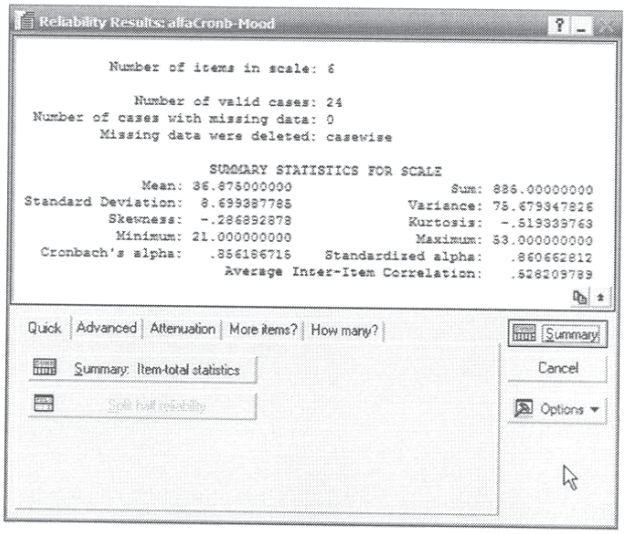
Clique agora no botão OK seleccione na patilha "Advanced" a opção "Compute multiple regression items/scale":



Clique agora no botão OK para obter o quadro dos resultados.



Neste quadro clique no botão Means & standard deviations para obter as estatísticas descritivas dos itens, ou no botão Correlations para obter as correlações inter-itens. Clique na patilha Review Descriptive S... para continuar com a análise. No quadro seguinte clique na patilha "Advanced" para obter mais estatísticas descritivas ou clique no botão OK para obter o quadro final:



O α de Cronbach é 0.856 e o α' é 0.861. A correlação média inter-itens é 0.528, a média da escala é 36.875, etc... Cada uma das patilhas/botões do quadro acima permite obter diferentes análises adicionais à consistência interna da escala. Por exemplo o botão 

permite obter as correlações entre os itens e o total da escala e a análise do a se cada um dos itens for eliminado:

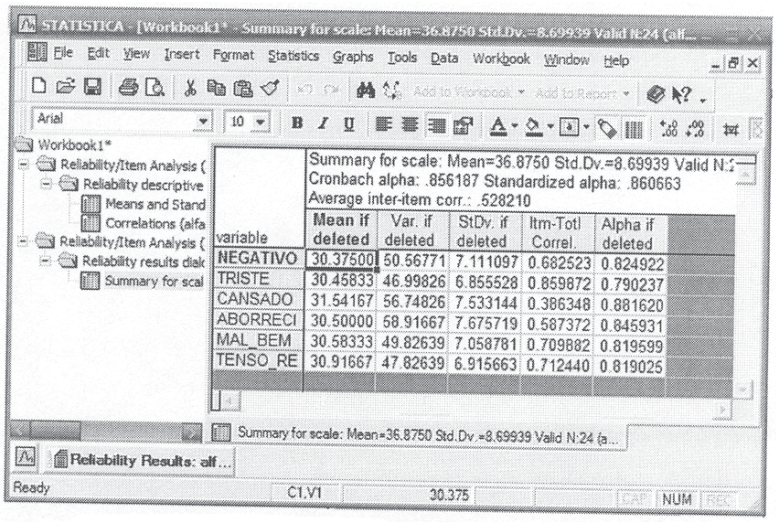
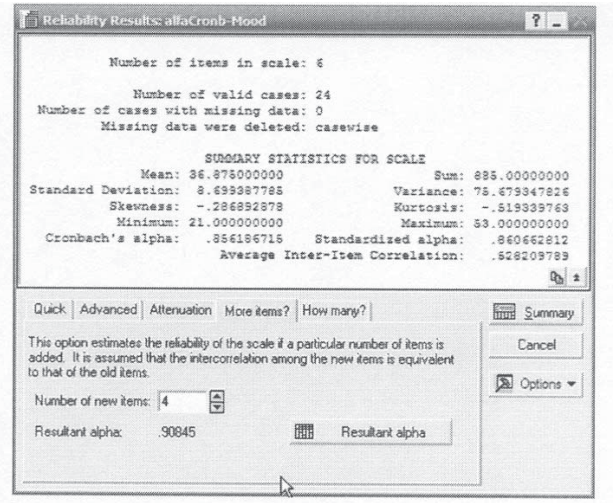
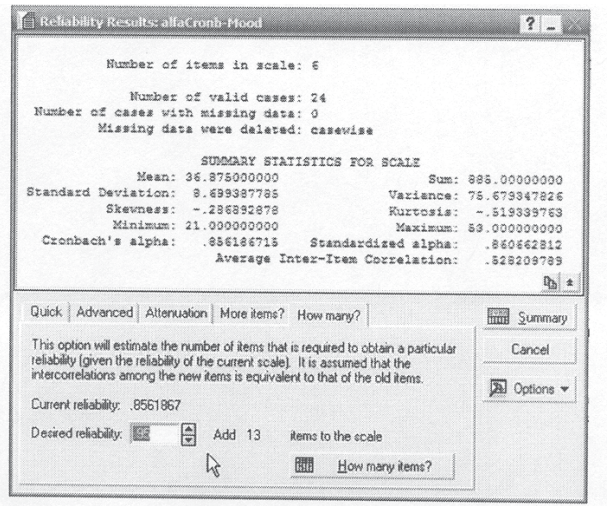


Figura 2. α de Cronbach com o Statistica

Uma análise interessante que não se encontra no SPSS, é a previsão de em quanto variaria o α se fossem adicionados mais itens à escala (assumindo que a correlação média inter-itens após a adição dos novos itens se mantinha inalterada). Clique na patilha "More items?" e digite 4 na caixa "Number of new items", Repare que a adição de 4 novos itens à escala permitiria aumentar o α para 0.908:



Um outro tipo de análise com interesse é a do número de itens que é preciso adicionar para que o α assuma um determinado valor. Clique na patilha "How many?" e digite 0.95 na caixa "Desired reliability":



Note que, assumindo que todas as correlações inter-itens se mantêm idêntica às correlações inter-itens da escala antiga, seria necessário adicionar 13 itens à escala para que o a passasse a 0.95.

Referências

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296-322.

Brown, J. D. (2001). Statistics Corner. Questions and answers about language testing statistics: Can we use the Spearman-Brown prophecy formula to defend low reliability? *Shiken: JALT Testing & Evaluation SIG Newsletter, 4*(3), 7-9.

Charter, R. A., & Feldt, L. S. (1996). Testing the equality of two alpha coefficients. *Perceptual and Motor Skills, 82*, 763-768.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and application. J*ournal of Applied Psychology, 78*, 98-104.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theor*y. New York: CBS College Publishing.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 1*6, 297-37.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3)*,* 391-418.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *The British Journal of Statistical Psychology, 16*, 137-163.

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratifiedparallel tests. *Educational and Psychological Measurement, 25*, 291-312.

Davis, F. B. (1964). *Educational measurements and their interpretation*. Wadsworth Publishing Co.: Belmont, California.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: SAGE Publications.

Duhacheck, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science, 24*(2), 294-301.

Duhachek, A., & Iacobucci, D. (2004). Alpha's Standard Error (ASE): An accurate and precise confidence interval estimate. J*ournal of Applied Psychology, 89*(5), 792-808.

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*, 517-531.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30,* 357-370.

Feldt, L. S. (1990). The sampling theory for the intraclass reliability coefficient. *Applied Measurement in Education, 3*, 361-367.

Garcia-Marques, T. (2004). A mensuração da variável "Estado de Espírito" na população portuguesa. *Laboratório de Psicologia, 2*(1), 77-94.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37, 8*27-838.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255-282.

Guttman, L. (1953). Reliability formulas that do not assume experimental independence. *Psychometrika, 18*, 225-239.

Hoyt, C. J. (1941). Test reliability estimated by analysis of variance. *Psychometrika*, *6*, 153-160.

Iacobucci, D., & Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology, 13*(4), 478-487.

Jones, L. V., & Appelbaum, M. I. (1989). Psychometric methods. *Annual Review of Psychology, 40,* 23-43.

Kamata, A., Turhan, A., & Darandari, E. (2003). *Estimating reliability for multidimensional composite scale scores*. Annual meeting of American Educational Research Association, Chicago, April 2003.

Kaplan, R., & Saccuzzo, D. (1982). *Psychological testing: Principles, applications and issues*. Monterey, CA: Brooks/Cole Publishing Company.

Kistner, E. O., & Muller, K. E. (2004). Exact distributions of intraclass correlation and Cronbach's alpha with Gaussian data and general covariance, *Psychometrika, 69*(3), 459-474.

Komaroff, E. (1997). Effect of simultaneous violations of essential tau-equivalence and uncorrelated error on coefficient alpha. *Applied Psychological Measurement, 21*, 337-348.

Krathwohl, D. R. (1998). *Methods of educational and social science research: An integrated approach.* (2nd ed.). New York: Addison-Wesley.

Kristof, W. (1963). The statistical theory of stepped-up reliability when a test has been divided into several equivalent parts. *Psychometrika*, *28*, 221-228.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151-160.

Laveault, D., & Grégoire, J. (2002). *Introduçao às teorias dos testes em ciências humanas.* Porto: Porto Editora.

Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. P*sychological Methods, 1,* 98-107.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140.*

Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.

Maroco, J. (2003). *Análise estatística com utilização do SPSS*. Lisboa: Edições Sílabo.

McDonald, R. P. (1999). T*est Theory: Unified treatment*. Lawrence Erlbaum Associates.

Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing: Principles and applications*. Englewood Cliffs, New Jersey: Prentice Hall.

Nunnally, J. C. (1978). P*sychometric theory.* New York: McGraw-Hill Inc.

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory.* New York: McGraw-Hill.

Osbourn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*, 343-355.

Pasquali, L. (2003). *Psicometria teoria dos testes na psicologia e na educação*. Petrópolis: Ed. Vozes.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. J*ournal of. Consumer Research, 21*(2), 381-391.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271-295.

Sawilowsky, S. S. (2000). Psychometrics versus datametrics: Comment on Vacha-Haase's "Reliability generalization" method and some EPM Editorial Policies. *Educational and Psychological Measurement, 60*(2), 157-173.

Sun, W., Chou, C-P., Stacy, A. W., Ma, H., Unger, J., & Gallaher, P. (2006). SAS and SPSS Macros to calculate standardized Cronbach's alpha using upper bound phi coefficient for dichotomous items. *Behavior Research Methods* (In press).

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, *54*, 837-847.

Thompson, B. (Ed.). (2002). *Contemporary thinking on reliability issues*. Newbury Park, CA: Sage.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*, 174-195.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 5*8, 6-20.

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65*, 271-280.

Weiss, D. J., & Davison, M. L. (1981). Test theory and methods. *Annual Review of Psychology, 32,* 629-658.

Wilkinson, L., & Task Force on Statistical Inference, APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594-604.

Yuan, K., & Bentler, P. M. (2002). On robustness of the normal-theory cased asymptotic distributions of three reliability coefficient estimates. *Psychometrika, 67*, 251-259.

Zimmerman, D. W., Williams, R. H., Zumbo, B. D., & Ross, D. (2005). Louis Guttman's Contributions to Classical Test Theory. *International Journal of Testing, 5*, 81-95.

Submissão: 18/05/2006 *Aceitação:* 03/07/2006