

Calcular e apresentar tamanhos do efeito em trabalhos científicos (2): Guia para reportar a força das relações

Calculating and reporting effect sizes on scientific papers (2): Guide to report the strength of relationships

Artigo de Revisão | Revision Article

Helena Espírito-Santo, PhD (1,2a), Fernanda Daniel, PhD (1a)

(1) Instituto Superior Miguel Torga, Coimbra, Portugal.

(2) Centro de Investigação do Núcleo de Estudos e Intervenções Cognitivo-comportamentais (CINEICC, FPCE da Universidade de Coimbra; Portugal.

(a) Elaboração do trabalho.

Autor para correspondência | Corresponding author: Helena Espírito-Santo; Largo de Celas, 1, 3000-132 Coimbra, Portugal; +351 910637946; helenum@gmail.com

RESUMO

Palavras-Chave

Tamanho do efeito
Coeficientes de correlação
Apresentação estatística
Interpretação estatística

No primeiro número da *Revista Portuguesa de Investigação Comportamental e Social* foi descrita a importância de calcular, indicar e interpretar os tamanhos do efeito para as diferenças de médias de dois grupos. A RPICS pretende continuar a alertar para a importância de reportar os tamanhos do efeito para outros testes estatísticos. A magnitude da força das relações não foi indicada no artigo prévio e talvez não se saiba que a correlação é um tamanho do efeito. Assim, este artigo pretende fornecer algumas diretrizes aos autores sobre os procedimentos de cálculo do coeficiente de correlação de Pearson e alguns coeficientes de correlação para dados especiais (Ró de Spearman, Tau de Kendall, Ponto-biserial e biserial, Fi, V de Cramér e Eta).

Com esse objetivo, serão apresentadas as fórmulas, os passos no SPSS (*Statistical Package for the Social Sciences*), pressupostos e precauções, classificação dos valores e sua interpretação. Uma vez que o SPSS não computa todos os coeficientes referidos, nos suplementos ao artigo são incluídas cinco folhas de cálculo (3 formas de comparar correlações, correlações ponto-biserial e biserial e correção de correlações para amostras < 60).

ABSTRACT

Keywords

Effect size
Correlation coefficients
Statistical reporting
Statistical interpretation

In the first issue of the Portuguese Journal of Behavioral and Social Research, it was described the importance of calculating, indicating and interpreting the effect sizes for the differences between means of two groups. The PJBSR intends to continue to remind of the importance of reporting effect sizes for other statistical tests. The magnitude of the strength of the relationships was not indicated in the previous paper, and it may not be known that correlation is an effect size. Thus, this paper intends to provide some guidelines to the authors on the procedures for calculating the Pearson correlation coefficient and some correlation coefficients for special data (Spearman Rho, Kendall's Tau, Point-biserial and biserial, Phi, Cramér's V, and Eta).

For this purpose, the formulas, steps in the SPSS (*Statistical Package for the Social Sciences*), assumptions and precautions, classification of values and their interpretation will be presented. Since SPSS does not compute all the mentioned coefficients, five spreadsheets (3 ways of comparing correlations, point-biserial and biserial, and correction of correlations for samples < 60) were included in the article supplements.

INTRODUÇÃO

Recordando e sintetizando a revisão efetuada no nosso trabalho prévio (Espírito-Santo e Daniel, 2015), indicámos que para todos os testes estatísticos é útil calcular e interpretar os *tamanhos do efeito* (TDE), havendo um índice diferente consoante o teste estatístico (J. Cohen, 1992; Ellis, 2010).

A apresentação dos TDE é muito importante na apresentação dos resultados das análises estatísticas, pois a) os TDE não dependem da dimensão da amostra (como o valor do p); b) ajudam a perceber o significado dos resultados e c) constituem um índice análogo para comparar resultados de estudos diferentes (Bezeau e Graves, 2001; Cumming, 2012; Ferguson, 2009; Kline, 2013; Lipsey et al., 2012; Olejnik e Algina, 2000; Snyder e Lawson, 1993).

Por todas estas razões, o cálculo dos TDE é também essencial para os estudos de metanálise (Berben et al., 2012; Cumming, 2012; Kline, 2013; Lipsey et al., 2012; Rosenthal, 1991).

Recorde-se que o TDE tanto indica a estimativa da diferença de uma medida entre duas ou mais observações (família d), como a magnitude da relação entre variáveis (família r) (Hedges, 1981; Rosenthal e DiMatteo, 2001).

A família d foi abordada na nossa revisão prévia, pelo que agora avançamos para a família r .

A família r inclui vários índices de associação relativos a duas ou mais variáveis: a) as medidas de *correlação* (p. ex., r , r_s , τ , r_{pb}) e b) as medidas de *proporção de variância* (p. ex., f^2 , η^2 , ε^2 , ω^2) (Hedges, 1981; Rosenthal e Rubin, 1982).

Outra diferenciação que importa fazer entre TDEs diz respeito à correção ou não do enviesamento (Ellis, 2010; Rosenthal, 1991; Thompson, 2007). Assim, vimos que o d de Cohen é um índice que pode sobrestimar o efeito, podendo ser corrigido através do g de Hedges (Espírito-Santo e Daniel, 2015). Na família r , a correção inclui vários índices alternativos, como o coeficiente de determinação múltipla (R^2_{ajust}), o ómega quadrado (ω^2) ou o épsilon quadrado (ε^2) (Snyder e Lawson, 1993).

Em síntese, os objetivos deste estudo de revisão são apresentar a primeira parte da família r , suas equações matemáticas, procedimentos no SPSS, pressupostos, formas de corrigir potenciais enviesamentos e indicações de interpretação. Os nossos objetivos, ainda assim, não são exaustivos, pretendendo descrever somente utilizações genéricas que o investigador pode encontrar nas áreas das ciências sociais e comportamentais.

A FAMÍLIA r (Parte 1)

Coeficientes de Correlação

Existem vários coeficientes de correlação, sendo o de Pearson o mais usado. A escolha do tipo de coeficiente está dependente da escala de medida usada (e.g., intervalar, de razão, ordinal ou nominal) ou da forma da relação (linear ou não-linear) (Berben, Sereika e Engberg, 2012; Breaugh, 2003; J. Cohen, Cohen, West e Aiken, 2003; Rosenthal e DiMatteo, 2001). Quando se usa um qualquer coeficiente de correlação, talvez não se saiba que é em si mesmo uma medida do tamanho do efeito e , tal como o d de Cohen, é uma medida padronizada (Rosnow e Rosenthal, 1996). Por esse motivo, o coeficiente de correlação é comparável entre estudos.

O valor de p que habitualmente acompanha a apresentação das correlações, e que é afetado pelo tamanho da amostra, indica se a relação é atribuível ao acaso ou não (Berben et al., 2012; Snyder e Lawson, 1993).¹ No entanto, é importante acrescentar que se pode ter uma correlação baixa e, no entanto, o valor do p ser estatisticamente significativo. Isto acontece, especialmente em estudos com grandes amostras ($n > 100$). Inversamente, em amostras pequenas ($n < 30$) podem encontrar-se correlações moderadas que não atingem significância estatística (Pallant, 2011).

Face ao uso extensivo da correlação de Pearson, esta revisão vai ser dedicada especialmente a este coeficiente. As equações matemáticas e as indicações para usar o coeficiente r de Pearson e para os restantes coeficientes de correlação são apresentadas previamente no Quadro 1.

Coeficiente de correlação r de Pearson. Este coeficiente quantifica a força e direção da relação linear entre duas variáveis e é representada pela letra “ r ”. Os valores de r podem variar entre -1 e 1 (Pearson, 1904). Quando duas variáveis não se correlacionam entre si, r é igual a zero, e quando se correlacionam totalmente, r é igual a 1 (ou $r = -1$) (Pearson, 1904).

Pressupostos e cautelas. Antes de se efetuar a análise correlacional há que verificar um conjunto de fatores que podem afetar a devida interpretação do coeficiente de correlação.

Linearidade da relação. A correlação de Pearson mede a força da relação linear entre variáveis (J. Cohen et al., 2003; Tabachnick e Fidell, 2007). Quando a relação entre as variáveis é curvilínea ou não-linear (Figura 1), a correlação aparece com um valor muito baixo (J. Cohen et al., 2003; Goodwin e Leech, 2006).

¹ Para uma revisão alargada sobre as limitações e cuidados a ter com o valor do p veja-se a nossa revisão anterior (Espírito-Santo e Daniel, 2015).

Quadro 1

Equações Matemáticas e Indicações de Uso para as Várias Estimativas de Correlação da Família r dos Tamanhos do Efeito

| Equação | Estimativa | Fórmula | Uso | SPSS passo a passo |
|---------|---|---|---|--|
| 1.1 | Coefficiente de correlação de Pearson | $r = \frac{cov_{X,Y}}{DP_X DP_Y}$ | A relação é linear e as duas variáveis são contínuas (i.e., são medidas em escalas de intervalo ou de razão; e.g., peso e rendimento económico). Notas: $cov_{X,Y} = \frac{\sum (X_i - M_X)(Y_i - M_Y)}{N - 1}$ | Analisar [Analyse] > Correlacionar [Correlate] Bivariável [Bivariate] > inserir variáveis a analisar > Opções [Options] > Selecionar Pearson [Select Pearson] > Continuar [Continue] OK |
| 1.2 | Correlação por postos de Spearman | $r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{N^3 - N}$ | As duas variáveis devem apresentar-se pelo menos numa escala ordinal (e.g., níveis de escolaridade e nível de especialização). Notas: $d_i^2 = (r_{X_i} - r_{Y_i})^2$; r_{X_i} = ordens das observações de X; r_{Y_i} = ordens das observações de Y; N = número total. | Analisar [Analyse] > Correlacionar [Correlate] Bivariável [Bivariate] > Inserir variáveis a analisar > Opções [Options] > Selecionar Spearman [Select Spearman] > Continuar [Continue] OK |
| 1.3 | Tau de Kendall | $\tau = \frac{C - D}{C(n, 2)}$ | As duas variáveis são medidas no mínimo em uma escala de medida ordinal. Alternativa vantajosa relativamente ao Ró de Spearman porque pode ser generalizado para um coeficiente de correlação parcial. Notas: C = número de pares concordantes de X e Y; D = número de pares discordantes de X e Y; $C(n, 2)$ = número de possibilidades de selecionar pares distintos. | Analisar [Analyse] > Correlacionar [Correlate] Bivariável [Bivariate] > inserir variáveis a analisar > Opções [Options] > Selecionar Tau-b de Kendall [Select Tau-b de Kendall] > Continuar [continue] OK |
| 1.4 | Tau de Kendall a e b | $\tau_b = \frac{P-Q}{\sqrt{D_r D_c}} \quad \tau_c = \frac{q(P-Q)}{W^2(q-1)}$ | As variáveis utilizadas no cálculo do tau-b e o tau-c são ordinais. Estes coeficientes lidam com os empates de forma diferente. O tau-b contabiliza-os enquanto o tau-c ignora-os. No tau-c a variação no intervalo -1 e +1 é assegurada, independentemente da dimensão da tabela, enquanto no tau-b essa variação só é assegurada em tabelas quadradas. Notas: P = probabilidade da consequência binária; Q = probabilidade de um sujeito ao acaso estar no grupo de risco; $W = \sum_{j=1}^c c_j = \sum_{i=1}^R r_i$ | Analisar [Analyse] > Estatística Descritiva [Descriptive statistics] Tabela de referência cruzada... [Crosstabs...] > inserir as variáveis a analisar na linha(s) e coluna(s) > Opções [Options] > Selecionar Tau-b de Kendall Selecionar Tau-c de Kendall [Select Tau-b de Kendall Tau-c de Kendall] > Continuar [Continue] OK |
| 1.5 | Coefficiente de correlação ponto-(biserial) | $r_{pb} = \frac{cov_{Xd,Y}}{\sqrt{DP_{Xd} DP_Y}} ;$ $r_b = \frac{(m_1 - m_0)p_0 p_1}{DP_Y \times y}$ | Uma variável é dicotómica [e.g., pertença a um grupo (grupo em tratamento e grupo em lista de espera)] e a outra é contínua (i.e., de intervalo ou razão; e.g., pontuações numa escala de depressão). Deve ter-se o mesmo número de casos em cada categoria da variável dicotómica. Notas: Xd = variável X dicotomizada; n_0 = número de elementos em X que são zero; n_1 = número de elementos que são um; $n = n_0 + n_1$; $p_0 = \frac{n_0}{n}$; $p_1 = \frac{n_1}{n}$; m_0 = média de Y quando os valores de X são zero; m_1 a média de Y quando os valores de X são um; DP = desvio-padrão; y = distribuição normal da forma da função especificada com média 0 e desvio-padrão 1 e forma da densidade de probabilidade 0. | O mesmo processo utilizado no cálculo do Coeficiente de correlação produto momento de Pearson. Analisar [Analyse] > Correlacionar [Correlate] Bivariável [Bivariate] > inserir variáveis a analisar > Opções [Options] > Selecionar Pearson [Select Pearson] > Continuar [Continue] OK |
| 1.6 | Coefficiente Fi | $\phi = \sqrt{\frac{\chi_p^2}{W}}$ | As duas variáveis são qualitativas dicotómicas ordinais ou nominais (e.g., sexo e lateralidade), podendo ser apresentadas numa tabela de contingência de 2 x 2. Notas: χ_p^2 = Qui-quadrado de Pearson; W = soma dos pesos de todos os casos. | Analisar [Analyse] > Estatística Descritiva [Descriptive statistics] Tabela de referência cruzada... [Crosstabs...] > inserir as variáveis a analisar na linha(s) e coluna(s) > Opções [Options] > Selecionar V de Cramer e Fi [Select Phi and Cramer's V] > Continuar [continue] OK |
| 1.7 | V de Cramér | $V = \sqrt{\left(\frac{\chi_p^2}{N[\min(I, J) - 1]} \right)}$ | Versão ajustada de fi usada com variáveis nominais em tabelas acima de 2 x 2 (e.g., regiões do país e níveis de escolaridade). É uma medida de associação baseada no qui-quadrado. Notas: χ_p^2 = Qui-quadrado de Pearson; I = Linha e J = Coluna da tabela. | Analisar [Analyse] > Estatística Descritiva [Descriptive statistics] Tabela de referência cruzada... [Crosstabs...] > inserir as variáveis a analisar na linha(s) e coluna(s) > Opções [Options] > Selecionar V de Cramer e Fi [Select Phi and Cramer's V] > Continuar [continue] OK |
| 1.8 | Eta | $\eta Y = \sqrt{1 - \frac{DP_{Y W}}{DP(Y)}}$ ou $\eta Y = \sqrt{\frac{\text{Soma dos quadrados entre grupos}}{\text{Soma total dos quadrados}}}$ | A relação entre as duas variáveis é curvilínea (e.g., ansiedade e desempenho num teste). Aplica-se também quando uma das variáveis é nominal e a outra é contínua. Notas: $DP_{Y W} = \sum Y_j^2 f_{ij} - \sum_{i=1}^R \frac{1}{r_i} (\sum_{j=1}^C Y_j f_{ij})$; DP(Y) = desvio-padrão de Y. | Analisar [Analyse] > Estatística Descritiva [Descriptive statistics] Tabela de referência cruzada... [Crosstabs...] > inserir as variáveis a analisar na linha(s) e coluna(s) > Opções [Options] > Selecionar Eta [Select Eta] > Continuar [continue] OK Também se pode obter no comando da ANOVA. |

Notas gerais. Tabela baseada em Cohen et al. (2003), Ellis (2010), Glass e Hopkins (1995); IBM (2013), Pallant (2011). X = variável 1 (por vezes entendida como dependente); Y = variável 2 (por vezes entendida como independente).

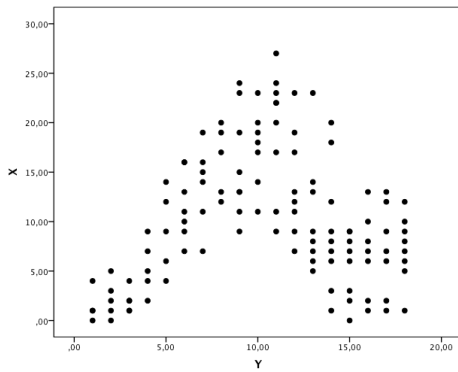


Figura 1. Diagrama de dispersão bivariado demonstrando uma relação curvilínea.

Nos casos em que a relação é curvilínea pode usar-se o Eta (ver em baixo no subcapítulo “Coeficientes de correlação para dados especiais”) ou pode tentar-se a transformação logarítmica, da raiz-quadrada ou da potência de uma ou de ambas variáveis (Abrami, Cholmsky e Gordon, 2001).

A melhor forma de verificar a linearidade da relação entre duas variáveis é analisar o diagrama de dispersão. Quanto mais os pontos se apresentarem próximos de uma linha reta no diagrama de dispersão², maior o grau de sobreposição/partilha da sua variância. Ou seja, mais alta será a correlação (J. Cohen et al., 2003; Pallant, 2011). Nestes casos, o diagrama de dispersão (Figura 2) tem uma forma oval (Tabachnick e Fidell, 2007).

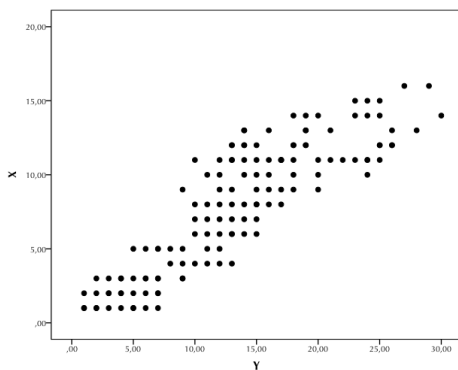


Figura 2. Diagrama de dispersão bivariado demonstrando uma relação linear.

Normalidade. As pontuações em cada variável contínua devem ter distribuição normal, aspeto que pode ser verificado através do teste de Kolmogorov-Smirnov, e análise da *curtose* (grau de achatamento da distribuição) e *obliquidade* (grau de simetria da distribuição) (Pallant, 2011; Tabachnick e Fidell, 2007). A estratégia mais segura é

² Ao usar o SPSS pode verificar-se o tipo de relação seguindo os passos seguintes: Gráficos [Graphs] > Diálogos anteriores [Legacy Dialogs] > Dispersão/Ponto [Scatter] > Dispersão simples [Simple Scatter] > Definir [Define] > Eixo Y [Y-axis] (inserir a Variável 1) e Eixo X [X-axis] (inserir a Variável 2). O gráfico de Dispersão/Ponto [scatterplot] permite também verificar se existem valores atípicos [outliers] (Pallant, 2011).

realizar a transformação das variáveis para melhorar a sua distribuição, a não ser que haja alguma razão forte para o não fazer. Alguns métodos comuns de transformação podem ser revistos em Tabachnick e Fidell (2007, pp. 86-88) ou pode seguir-se a abordagem dos dois passos sugerida por Templeton (2011).

Homoscedasticidade. A variância das duas variáveis deve ser homogénea (Pallant, 2011), e ainda que o incumprimento deste pressuposto não impeça a análise, esta fica enfraquecida (Tabachnick e Fidell, 2007). O diagrama de dispersão deve apresentar, assim, uma mancha de largura uniforme.

A falha na homoscedasticidade (heterocedasticidade) advém da não-normalidade da distribuição de uma das variáveis ou do facto de uma variável resultar de alguma transformação da outra. A heterocedasticidade pode resultar também de um maior erro de mensuração em alguns níveis da variável independente (Tabachnick e Fidell, 2007, p. 85). Por exemplo, a habilidade de discriminação visual relaciona-se com a velocidade de reação e tende a ser melhor entre os 20 e os 40 anos (Fortenbaugh et al., 2015). Assim, as medições dos tempos de reação nos mais novos e nos mais idosos serão menos fidedignas, aumentando a variância da velocidade de reação nestas idades.³

Valores atípicos. Os valores não-usuais ou outliers são valores que se afastam dos restantes dados, podendo ser encontrados quer em variáveis contínuas quer dicotómicas (Tabachnick e Fidell, 2007). Em alguns casos (e.g., amostras pequenas), mesmo a existência de poucos ou até mesmo de um só valor atípico pode prejudicar os resultados e sua interpretação (J. Cohen et al., 2003). Os valores atípicos podem aumentar a correlação artificialmente nuns casos, ou diminuir a verdadeira correlação noutras situações (Glass e Hopkins, 1995; J. Cohen et al., 2003). Estes valores podem ser detetados através da observação das caixas de bigodes (boxplots), das tabelas de frequência (mais limitado) ou dos diagramas de dispersão (J. Cohen et al., 2003). Os valores atípicos podem resultar de erros de introdução de dados que convém verificar (p. ex., escrever 11 em vez de 1) (Tabachnick e Fidell, 2007) ou do facto de os sujeitos não terem percebido as instruções (Cohen, 2001). Se não for esse o caso, são valores que podem ser removidos ou alterados para um valor menos extremo [para mais sugestões, consultar Tabachnick e Fidell (2007, p. 73)].

Valores omissos. Os dados em falta podem afetar as análises e invalidar as interpretações (Nakagawa e

³ Adaptação do exemplo de (Tabachnick e Fidell, 2007, p. 85) a um estudo real.

Freckleton, 2008) ou afetando a generalização dos resultados (Tabachnick e Fidell, 2007). No SPSS através do comando Descritivas [Descriptives] pode determinar-se qual a percentagem de dados omissos [missings] e através da Análise de Valor Omitido [Missing Values Analyses] determinar se ocorrem ao acaso ou se existe um padrão de casos omissos (Pallant, 2011). A seriedade dos casos omissos depende do padrão, do número de casos e da razão para a omissão, sendo o mais sério a existência de um padrão (Tabachnick e Fidell, 2007).

Existem vários métodos para lidar com os dados ausentes. Uma das formas é eliminar casos ou variáveis, mas suponha-se que a ausência de resposta se relaciona com uma das variáveis de interesse. Então a remoção dos casos irá distorcer a análise dessa variável (Tabachnick e Fidell, 2007). Outros métodos envolvem a imputação (i.e., preencher com) com valores substitutos razoáveis, baseando-se noutros valores caso a caso ou na distribuição da variável em análise. Estes métodos permitem manter o tamanho da amostra e evitar a perda de poder estatístico. O leitor interessado pode consultar alguns dos muitos trabalhos dedicados a este assunto (e.g., [the Alzheimer's Disease Neuroimaging Initiative, 2015](#); Bartlett, Seaman, White, Carpenter, Dempster, Laird e Rubin, 1977; He, Zaslavsky, Landrum, Harrington e Catalano, 2010; Horton e Kleinman, 2007; Rubin, 1996; Tabachnick e Fidell, 2007).

Independência das observações. As variáveis devem ser independentes e não ser influenciadas uma pela outra (Pituch e Stevens, 2015). Na análise correlacional é problemático usar as pontuações individuais de sujeitos cuja/cujo aprendizagem/intervenção/tratamento ocorreu no contexto de pequenos grupos. Neste contexto, o comportamento de cada elemento influencia o desempenho grupal, havendo dependência nas medições. Nas correlações, mesmo uma pequena quantidade de dependência nas observações aumenta o alfa várias vezes acima do nível de significância (Pituch e Stevens, 2015). Uma solução para este problema é restringir o nível de significância para 0,01 (Pituch e Stevens, 2015). Quando grupos pequenos estão envolvidos em algum tipo de tratamento/intervenção, outro recurso é considerar-se a média do grupo como unidade de análise (Pituch e Stevens, 2015).

Cautela com leque restrito de observações. O valor de r será sempre maior quando maior for a variabilidade das observações (Goodwin e Leech, 2006). Assim, é necessário ser-se cauteloso ao interpretar as pontuações obtidas em populações especiais que podem restringir a amplitude da variação (Pallant, 2011; Tabachnick e Fidell, 2007). Por exemplo, frequentemente em pessoas idosas

institucionalizadas a variação obtida nas pontuações do *Mini-Mental State Examination* é mais limitada (Costa et al., 2013) do que a obtida em pessoas idosas avaliadas no contexto hospitalar (Fermino et al., 2012). Para outro exemplo muito interessante, veja-se a descrição de Sprinthall (2003) relativamente ao estudo de Fancher (1985) sobre a descida do QI com a idade, explicável pela descida da variabilidade das pontuações do QI (Goodwin e Leech, 2006).

Por este motivo, ou se procura amostras diversificadas por forma a obter um coeficiente de correlação preciso e fidedigno ou não se pode generalizar a correlação obtida para amostras de outras proveniências (Pallant, 2011). Quando uma correlação é muito baixa por causa da restrição da gama de pontuações, pode computar-se a sua magnitude através da equação (1), se se conseguir estimar o desvio-padrão a partir de dados prévios ou do conhecimento sobre a distribuição na população (Tabachnick e Fidell, 2007).

$$r_{ajust} = \frac{r_{T(XY)} \frac{DP_X}{DP_{T(X)}}}{\sqrt{1 + r_{T(XY)}^2 \left[\frac{DP_X}{DP_{T(X)}} \right] - r_{T(XY)}^2}} \quad (1)$$

Onde r_{ajust} = correlação ajustada; $r_{T(XY)}$ = correlação entre as variáveis X e Y na gama das médias truncadas ou aparadas; DP_X = desvio-padrão não restringido de X; $DP_{T(X)}$ = desvio-padrão truncado de X.

Cautela com a natureza da amostra. A correlação pode ser diferente de um grupo para o outro por causa das características dos sujeitos (Goodwin e Leech, 2006). Goodwin e Leech, (2006) dão como exemplo o estudo de Glenberg (1996) em que a correlação entre viúvos/as e o desejo de recasar era positivo para os dois grupos separados (novos vs. velhos), mas no grupo total a relação era negativa.

Cautela com variáveis compósitas. Se as variáveis em análise incluem elementos ou itens em comum as correlações ficam inflacionadas. Por exemplo, no uso de questionários cujos índices resultam da soma de vários itens, se se correlacionarem índices do mesmo instrumento e esses índices contiverem itens em comum, então as correlações aumentam artificialmente. Assim, devem ser usadas somente uma das variáveis compósitas na análise (Tabachnick e Fidell, 2007).

Cautela com o formato da distribuição de X e Y. O formato da distribuição das pontuações das variáveis é um fator a ter em consideração, pois afeta a força da correlação (Goodwin e Leech, 2006). Assim, quanto mais idêntico for o formato das distribuições de X e Y, maior o valor da correlação (Glass e Hopkins, 1995; Goodwin e

Leech; 2006). Carroll (1961) mostrou que o valor máximo que a correlação pode atingir depende da similitude da simetria e curtose entre as duas distribuições. Portanto, o valor máximo da correlação desce quando a simetria e a curtose das duas distribuições são diferentes. Isto é especialmente verdade em correlações que já de si seriam altas. Goodwin e Leech (2006) mostraram que uma correlação de 0,90 entre duas variáveis com formato de distribuição semelhante desce para 0,80 ou 0,70 ao alterar-se esse formato numa das distribuições.

Sinal da correlação. O sinal da correlação indica a direção da relação. Assim, as correlações com sinal positivo são sugestivas de uma relação linear positiva (o aumento numa variável acompanha-se do aumento na outra ou a descida de uma acompanha-se da descida da outra). As correlações de sinal negativo indicam uma relação linear inversa (a subida de pontuações de uma variável acompanha-se da descida de pontuações da outra variável) (Shieh, 2010; Rosenthal, 1991; Tabachnick e Fidell, 2007).

Nunca é demais destacar que uma correlação negativa se refere à direção da relação e não à força da relação (Pallant, 2011). Assim, quando no estudo de Simões, Ferreira, Braga e Vicente (2015) se encontrou uma correlação negativa ($r = -0,36$; $p < 0,05$) entre a comunicação com o pai e a conduta agressiva na escola em 50 adolescentes portugueses, isso significa que quanto mais o pai comunica com o seu filho adolescente menor a conduta agressiva na escola.

Força da correlação. Os valores da força de uma relação, entre os extremos -1 e 1, foram definidos por Karl Pearson (1904) como “altos” quando se situavam entre 0,75 e 1; “consideráveis” quando se localizavam entre 0,50 e 0,75; “moderados” entre 0,25 e 0,50 e “baixos” entre 0 e 0,25.

Com Cohen (1988) ficou estabelecida uma das classificações mais usadas: “grande” 0,50 a 1,00; “moderada” de 0,30 a 0,49 e “pequena” de 0,10 a 0,29.

Se se preferir uma classificação mais minuciada, pode recorrer-se à de Hinkle, Wiersma e Jurs (2003): 0,90 a 1,00 “Muito alta”; 0,70 a 0,90 “Alta”; 0,50 a 0,70 “Moderada”; 0,30 a 0,50 “Baixa”; 0,10 a 0,30 “Pequena”.

Todas as diretrizes aplicam-se quer a correlações positivas, quer as correlações negativas.

Tome-se o sistema de classificação que se tomar, os valores entre 0 e 1 não fazem parte de métricas a que se esteja habituado e o seu significado não é intuitivo. Entra aqui a propósito o **coeficiente de determinação** (r^2) que se refere à proporção de variância partilhada entre duas variáveis. Apesar do fraseado matemático poder assustar alguns, ele obtém-se muito simplesmente calculando o

quadrado do valor r e multiplicando-o por 100 (J. Cohen e Cohen, 1983). Assim, o valor de -0,36 atrás referido do estudo de Simões et al. (2015) converte-se em $-0,36 \times -0,36 = 0,1296$, ou seja, 13% da variância da comunicação com o pai é partilhada com a variância da conduta agressiva na escola. Dito ainda de outro modo, 13% da variância conduta agressiva na escola por parte de adolescentes pode ser atribuída à comunicação com o pai.

Interpretação e significado prático. Através do valor da correlação, do seu sinal e do coeficiente de determinação, já se sabe qual a força e direção de uma relação entre variáveis. Mas como se interpreta esse valor na área de estudo em que a relação se insere? Na interpretação dos valores de correlação e respetivos valores do p há que refletir sobre a questão da causalidade e do contexto do estudo e comparar com outras pesquisas na mesma área de investigação. Vamos então dividir esta secção em três aspetos.

Relação é causalidade? Comece-se já por negar esta ideia: a atribuição de causalidade não é uma inferência que se possa fazer a partir da análise correlacional, ainda que possa parecer uma inferência lógica (Cohen, 1988). Acrescente-se ainda que nem mesmo de correlações muito elevadas se pode estabelecer causalidade (Shieh, 2010).

Assim, a correlação entre as variáveis X e Y , tanto pode decorrer de X causar Y ou de Y causar X . Por vezes, X e Y podem ser causa e efeito ao mesmo tempo. O exemplo de Huff (1993) insere-se nessa categoria: a correlação entre rendimentos e a posse de ações cotadas na bolsa pode resultar do facto de se conseguir comprar mais ações por se ter mais dinheiro e de se fazer mais dinheiro por se ter mais ações.

Pode ainda acontecer que X e Y são causadas por uma terceira variável W (habitualmente designada por variável espúria ou de confusão). Por exemplo, a relação entre a demência e a depressão na população idosa está bem estabelecida. No entanto, a doença vascular pode ser uma variável que explica a relação entre depressão e a demência, sendo um fator causal para as duas variáveis⁴ (Butters et al., 2008). O ideal é avaliar a terceira variável na mesma altura e depois, pode usar-se a correlação parcial para controlar o papel da variável adicional (Pallant, 2011).

A correlação entre duas variáveis pode ainda decorrer sem a presença de uma variável de confusão óbvia. O exemplo real fornecido por Matthews (2000) ilustra bem esta ideia. O autor encontrou uma correlação alta ($r = 0,62$; $p = 0,008$) entre o número de cegonhas

⁴ O modelo proposto pelos autores é mais complexo do que aqui exposto, mas que serve a finalidade da ilustração do argumento.

brancas (*Ciconia ciconia*) de várias partes da Europa e o número de nascimentos de bebés. Esta correlação não permite (à partida) tirar conclusões importantes e, muito menos, inferir que as cegonhas entregam os bebés (Matthews, 2000). Uma potencial variável de confusão será a dimensão da área geográfica onde os registos de cegonhas e de nascimentos teve lugar (Matthews, 2000). Mas mais importante que isso, é reforçar o que já sublinhámos no artigo anterior (Espírito-Santo e Daniel, 2015) e que Matthews (2000) também conclui: um valor do p significativo só permite rejeitar a hipótese nula, e por muito grande que seja não permite afirmar que a hipótese substantiva é correta.

Contexto dos dados. O contexto do estudo pode transformar uma correlação baixa numa relação relevante, e transformar uma correlação alta numa relação irrelevante. Por exemplo, uma correlação classificada como baixa ($r = 0,26$) no domínio da ideação suicidária (correlação entre “desespero” e o “desejo de cometer uma tentativa ativa de suicídio”) (Mendonca e Holden, 1996) pode, ainda assim, ser clinicamente relevante. Em contraste, a correlação acima referida entre o número de cegonhas brancas e o número de nascimentos de bebés não terá interesse de maior.

No contexto do tratamento pode ser útil acrescentar a leitura da reflexão de Rosenthal e Rubin (1982) e Rosenthal e DiMatteo (2001), onde é mostrado como correlações baixas podem corresponder a taxas de sucesso altas e as vantagens de mostrar o tamanho do efeito binomial.

Comparação com outras pesquisas. A comparação com outros estudos na mesma área de investigação deve ser tida sempre em conta para determinar o significado prático de um coeficiente de correlação. Assim, por exemplo, no estudo de Chiu, Hong e Chiu (2013) a correlação entre o tempo de uso do telemóvel e da internet variou entre 0,34 e 0,35 (11,6-12,3%), mostrando-se pouco importante quando comparada com a obtida por Beranuy, Oberst, Carbonell e Chamarro (2009) ($r = 0,52$; 27,0%).

Cautela com séries que aumentam ao longo do tempo. Se as variáveis consistirem em séries de valores que aumentam ao longo do tempo, então a interpretação do coeficiente de correlação deve ser prudente. Neste tipo de situação, a correlação tende a ser inflacionada (Granger e Newbold, 1974). Um exemplo clássico relatado por Huff (1993) é o da correlação entre o salário de padres metodistas na Nova Inglaterra entre 1860 e 1940 e a quantidade de rum vendido na Jamaica. A única verdadeira ligação aqui é que ambos os valores mudaram ao longo do tempo. Para além de que havia uma ou duas variáveis espúrias ou de confusão e que

foram o aumento da população entre 1860 e 1940 e o aumento do preço de quase tudo. A forma apropriada para testar relações deste tipo é usar um procedimento que se designa por cointegração (Enders, 2015). Há ainda que ter em atenção à inferência que não se pode fazer ao observarem-se dois fenómenos a crescer ao mesmo tempo e que estão correlacionados. A existência dessa correlação não significa que se vá manter no tempo e que não possa mesmo inverter-se. Mais uma vez, tome-se o exemplo de Huff (1993). Se é verdade que quanto mais chove mais milho há (correlação positiva), se continuar a chover sem parar o milho pode ser destruído (correlação negativa).

Por fim, a questão da interpretação não se esgota no âmbito deste artigo e é importante saber que uma correlação pode ser interpretada de várias formas: Rodgers e Nicewander (1988) indicaram treze formas de interpretar uma correlação e Rovine e von Eye (1997) acrescentaram uma décima quarta.

Correlações e outros testes estatísticos. Outro aspeto útil a reter é que a correlação r pode ser calculada através de várias outras estatísticas. Tal como, através do teste t de Student (Durlak, 2009):

$$r = \sqrt{\frac{t^2}{t^2 + gl}} \quad (2)$$

Pode ainda ser calculada através do d de Cohen (Durlak, 2009):

$$r = \sqrt{\frac{d}{\sqrt{d^2 + 4}}} \quad (3)$$

Esta conversão pode ser útil, pois o r enquanto tamanho do efeito apresenta vantagens face ao dicotómico d de Cohen. De facto, o r , na sua forma ponto-bisserial (ver à frente), representa a relação entre os dois níveis da variável independente e as pontuações da variável dependente (Rosenthal e DiMatteo, 2001, p. 71). Para outras vantagens ler o artigo de Rosenthal e DiMatteo (2001), com destaque para a importância prática do r face ao d .

Ou do g de Hedges (Durlak, 2009):

$$r = \sqrt{\frac{g^2 n_1 n_2}{g^2 n_1 n_2 + (n_1 + n_2) gl}} \quad (4)$$

Enviesamentos e correções das correlações. Note-se que apesar da facilidade que os programas informáticos trouxeram à computação das correlações e não obstante o uso generalizado, o coeficiente de correlação de Pearson é uma estimativa que pode ser enviesada (Shieh, 2010; Wang e Thompson, 2007).

A distribuição de r não é normal. De facto, o coeficiente de correlação r subestima ligeiramente o valor numa

população (ρ), pois a distribuição de r é assimétrica, tendendo a enviesar para cima (Fisher, 1921). Ou seja, quanto mais altas as correlações, mais difícil é aumentá-las. Assim, a diferença entre uma correlação de 0,90 e outra de 0,80 é muito maior do que a diferença entre uma correlação de 0,20 e uma de 0,10. Para melhor compreender isto, basta recordar o coeficiente de determinação r^2 . Então, a diferença entre 0,90 ($r^2 = 0,81$; 81%) e 0,80 ($r^2 = 0,64$; 64%) é de 17 pontos percentuais e entre 0,20 ($r^2 = 0,04$; 4%) e 0,10 ($r^2 = 0,01$; 1%) é de 3 pontos percentuais.

Fisher (1915; 1924), ao considerar este problema, criou duas equações para corrigir o r . A primeira é tão complexa que raramente é usada (Wang e Thompson, 2007). A segunda, que o próprio considerou mais útil (Fisher, 1921), ficou designada por transformação z de Fisher, apresentando-se como se segue:

$$z = \log_{10}(1 + r) - \log_{10}(1 - r) \quad (5)$$

Pela sua importância, nos suplementos incluímos uma folha de cálculo em Excel que faz essa transformação (Suplemento 1). Destaque-se que quando r é inferior a 0,25, a diferença entre z e r é mínima, mas aumenta com valores de r maiores. Por exemplo, uma correlação de 0,45 corresponde a um valor z de 0,49 e uma correlação de 0,87 a um valor z de 1,33.

Enviesamento relacionado com o tamanho da amostra. Outro enviesamento decorre de os efeitos não-corrigidos do tamanho da amostra tenderem a sofrer de enviesamento positivo. Muito especialmente, quando as amostras são pequenas, os valores do r de Pearson podem sofrer um enviesamento positivo bastante considerável, mesmo com dados normalmente distribuídos (Thompson, 2006).

Sempre que um estudo contenha 60 participantes ou menos, sugere-se então que se use uma fórmula de correção. A utilização de uma fórmula de correção é desnecessária quando o estudo for efetuado numa área onde tradicionalmente os tamanhos do efeito são grandes (Wang e Thompson, 2007).

Uma das equações de correção mais adequadas é a de Ezekiel (1929) que tem como vantagem ser calculada no SPSS, ao computar-se o r^2 no comando da regressão (Wang e Thompson, 2007). Outra das equações de correção com provas dadas (Wang e Thompson, 2007), é a de Smith (1923). Para ambas as equações, nos suplementos é apresentada uma folha de cálculo em Excel que facilitará a sua computação.

Comparações entre correlações. Como já se referiu em cima, a distribuição do r não é normal, e as diferenças entre

correlações altas (p. ex., 0,85 e 0,90) é maior do que entre correlações baixas (p. ex., 0,15 e 0,10). Recorde-se o coeficiente de determinação também já referido atrás para entender o porquê. Afortunadamente, a transformação em z desenvolvida por Fisher (1921), que permite que a distribuição se aproxime da normal, já possibilita a comparação entre correlações, determinando se a diferença entre elas é significativamente diferente.

No entanto, existem três processos consoante se quer comparar correlações de amostras independentes, de amostras dependentes ou de amostras dependentes com uma variável em comum. As equações são as que a seguir se descrevem.

Para a comparação entre correlações de amostras independentes, primeiro cada coeficiente de correlação é convertido em pontuação z através da equação (5). Depois, as duas pontuações z são confrontadas através da equação seguinte (J. Cohen e Cohen, 1983):

$$Z_{\text{obs}} = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \quad (6)$$

Com z_1 a corresponder à transformação z da correlação 1; z_2 a corresponder à transformação z da correlação 2; n_1 ao tamanho da amostra para a correlação 1 e n_2 ao tamanho da amostra para a correlação 2. Os valores de Z_{obs} acima de |1,96| são considerados significativos e acima de |2,58| muito significativos (J. Cohen e Cohen, 1983).

Quando se quer comparar correlações em amostras dependentes, ou seja, quer verificar-se se as correlações entre X e Y e entre W e V da mesma amostra diferem em magnitude. Ora, uma vez que a significância depende das intercorrelações par a par entre todas as variáveis envolvidas (X , Y , W e V), então tem de se entrar com parâmetros adicionais. Assim a primeira etapa é a mesma: conversão da correlação em valor z através da equação (5). De seguida, usam-se as Equações 7 e 8 de Steiger (1980, pp. 245, 247) para determinar a covariância assintótica das estimativas (Quadro 2). Estes valores são usados num teste z assintótico.

Em amostras dependentes com uma variável em comum, i.e., quando se quer comparar as correlações entre X e Y e entre W e Y e verificar se apresentam uma magnitude diferente, o procedimento é o seguinte: a primeira etapa é ainda a mesma (Equação 5), seguindo-se o cálculo por meio das Equações 9 e 10 de Steiger (1980, pp. 245, 247) para determinar a covariância assintótica das estimativas e usá-las num teste z assintótico (Quadro 2).

Quadro 2

Equações 7 a 9 de Steiger (1980)

$$k = (r_{WY} - r_{WX}r_{XY})(r_{XV} - r_{XY}r_{YV}) + (r_{WV} - r_{WY}r_{YV})(r_{XY} - r_{WX}r_{WY}) + (r_{WY} - r_{WV}r_{YV})(r_{XV} - r_{WX}r_{WY}) + (r_{WV} - r_{WX}r_{WY})(r_{XY} - r_{XV}r_{YV}) \quad (7)$$

$$c = \frac{\frac{1}{2} \bar{r}^2 (r_{WY}^2 + r_{WV}^2 + r_{XY}^2 + r_{XV}^2) + r_{WY}r_{XV} + r_{WV}r_{XY} - (\bar{r}r_{WY}r_{WV} + \bar{r}r_{XY}r_{XV} + r_{WY}r_{XY}\bar{r} + r_{WV}r_{XV}\bar{r})}{(1 - \bar{r}^2)^2} \quad (8)$$

$$k = r_{XY} (1 - r_{WX}^2 - r_{WY}^2) - \frac{1}{2} (r_{WX}r_{WY})(1 - r_{WX}^2 - r_{WY}^2 - r_{XY}^2) \quad (9)$$

$$c = \frac{r_{XY} (1 - 2\bar{r}^2) - \frac{1}{2} \bar{r}^2 (1 - 2\bar{r}^2 - r_{XY}^2)}{(1 - \bar{r}^2)^2} \quad (10)$$

Para agilizar os cálculos, no Suplemento 1, existem três folhas de cálculo que permitem fazer essas comparações. Nas folhas de cálculo são apresentados todos os passos e a interpretação do tamanho do efeito q que consiste na comparação entre os valores z . Na mesma folha é também indicado se a diferença entre os valores z das correlações é significativa ou não.

Assim, como exemplo da comparação entre correlações de amostras independentes, repare-se nos dados do estudo de (Napoleão, Monteiro e Espírito-Santo, 2016). Neste estudo, verificou-se que a correlação entre o Índice de Qualidade Subjetiva do Sono e os sintomas depressivos medidos pela *Geriatric Depression Scale* foi de 0,39 em 70 pessoas idosas institucionalizadas e de 0,35 em 70 não-institucionalizadas. Os valores z respetivos são de 0,41 e 0,37, sendo a diferença não significativa ($Z_{obs} = 0,27$) e com um tamanho do efeito (q) trivial de 0,05.

Coefficientes de correlação para dados especiais.

Quando os pressupostos do coeficiente de Pearson não são cumpridos ou quando a natureza das variáveis assim o exige, existe a possibilidade de optar por outros coeficientes. Vamos desenvolver aqui os mais usados.

Correlação por postos de Spearman.

O Ró de Spearman é um coeficiente de correlação não-paramétrico a escolher quando os dados não apresentam distribuição normal ou quando a presença de valores atípicos distorce a associação entre as variáveis X e Y (Pallant, 2011). Da mesma forma que a correlação Pearson, varia entre -1 e 1 e consiste na correlação entre os postos de dois conjuntos de observações (J. Cohen e Cohen, 1983). Por postos entende-se a posição ordinal da observação ou caso na variável. Ver o Quadro 1 para consultar a Equação 1.2 e comandos no SPSS. As cautelas, força da correlação, direção e interpretação seguem as diretrizes das enunciadas para as correlações de Pearson.

Tau de Kendall.

O tau de Kendall é igualmente uma medida de associação por postos (Kendall, 1938). Este tipo de correlação usa-se quando as duas variáveis são então medidas em uma escala de medida ordinal. É semelhante

um medida não-paramétrica que varia entre -1 e 1. O Tau de Kendall é uma opção favorável comparativamente com o Ró de Spearman porque pode ser generalizado para um coeficiente de correlação parcial. Existem fórmulas de cálculo do tau que entram em linha de conta com a existência ou não de empates e com o formato da tabela (Nelsen, 2012). Ver Quadro 1...

Correlações ponto-biserial e biserial.

Este tipo de correlação é uma versão curta da correlação de Pearson (Nunnally, 1994) que se utiliza quando uma variável é contínua e a outra é dicotómica, assumindo os valores 0 e 1 (Gupta, 1960; Tabachnick e Fidell, 2007). A designação biserial deriva do facto de haver duas séries de pessoas observadas em Y (aqueles que pontuam 0 e os que pontuam 1 na variável X) (Glass e Hopkins, 1995). O valor que se obtém com a correlação ponto-biserial é o mesmo da correlação de Pearson (e daí o comando ser o mesmo no SPSS).

Registe-se que quanto mais a distribuição de Y se afastar de 50/50, mais restritos se tornam os valores da correlação (Glass e Hopkins, 1995). O valor fica ainda muito baixo se a maior parte (> 90%) das respostas na variável dicotómica caírem numa categoria (Tabachnick e Fidell, 2007). Veja-se um exemplo menos dramático da descida da correlação quando se dicotomiza uma variável. No Projeto *Trajétórias do Envelhecimento* do Instituto Superior Miguel Torga (PTE) foram avaliadas 1040 pessoas idosas através do *Geriatric Depression Scale* (GDS) e do *Geriatric Anxiety Inventory* (GAI). A correlação das pontuações do GDS com o GAI foi de 0,80 ($p < 0,001$; $r^2 = 64,0\%$). através do ponto de corte de 13, a variável GAI foi dicotomizada na binária GAI₂, onde 1 corresponde a “poucos sintomas ansiosos” e 2 a “muitos sintomas ansiosos”. A correlação entre o GDS e o GAI₂ foi menor ($r = 0,69$; $p < 0,001$; $r^2 = 47,6\%$). A dicotomização resultou numa menor proporção da variância e é por isso que Cohen (1983) adverte contra este procedimento.

A correlação que está patente no exemplo de cima é, na verdade, uma correlação biserial. Esta correlação é, na essência a mesma que a ponto-biserial, mas o

pressuposto é que existe uma variável contínua que subjaz a medida dicotômica (Glass e Hopkins, 1995). A bisserial fornece uma melhor estimativa e usa-se quando a variável contínua subjacente tem distribuição normal (Glass e Hopkins, 1995; MacCallum, Zhang, Preacher e Rucker, 2002). A folha de cálculo no Suplemento 3 fornece a computação dos dois coeficientes.

No Quadro 1 pode examinar-se a Equação 1.3 e os comandos no SPSS. As precauções, força da correlação, direção e interpretação acompanham as diretrizes das expressas para as correlações de Pearson.

Note-se, no entanto, que **nem sempre a correlação ponto-bisserial tem a possibilidade de chegar a 1,00, pois o seu máximo é determinado pela distribuição da variável dicotômica**. Assim, para um número igual de pessoas em cada grupo, o máximo é de 0,80; para uma divisão de 90% a 10% já é somente de 0,59 e para 99% a 1% é de 0,27. Por isso há que interpretar à luz do máximo possível (Breugh, 2003, p. 85).

→ Coeficiente Fi. Este coeficiente é outro caso especial da correlação de Pearson e consiste na **correlação entre duas variáveis dicotômicas** (Glass e Hopkins, 1995; MacCallum et al., 2002). Cohen (1988) designou-o ainda por **coeficiente de correlação ponto quádrupla**.

Por exemplo, no PTE poder-se-ia usar para o coeficiente Fi a medição da relação entre o sexo e a prática ou não de atividade física ($\phi = 0,08$), revelando-se um valor muito pequeno segundo os critérios de Cohen (1988). De acordo com este autor, este coeficiente varia entre 0 e 1, com os valores mais altos a indicarem uma relação mais forte: entre 0,10 e 0,29 é **pequeno; médio** entre 0,30 e 0,49 e **grande** acima de 0,50.

Se as **duas variáveis dicotômicas tiverem subjacentes variáveis contínuas, então usam-se as correlações tetracóricas** (para o efeito consultar, Glass e Hopkins, 1995; MacCallum, Zhang, Preacher e Rucker, 2002).

Como o Fi é uma medida de correlação, ele é ainda uma **medida do tamanho do efeito para o teste do qui-quadrado da independência** (Gravetter e Wallnau, 2013).

Finalmente, para as limitações no uso do Fi, veja-se a revisão de Breugh (2003, pp. 82-83)

→ Coeficiente V de Cramér. Consiste numa modificação do Fi e utiliza-se quando as **duas variáveis são nominais** (Gravetter e Wallnau, 2013). Para a classificação dos valores do V de Cramér praticam-se os mesmos critérios do coeficiente Fi (Cohen, 1988).

Ainda, tal como o coeficiente Fi, o V de Cramér usa-se para representar a força da associação na análise do qui-quadrado (Ferguson, 2009).

Eta. O rácio de correlação ou Eta foi definido por Pearson (1905) como a medida de associação entre

variáveis independentes (categorial ou ordinal) e dependentes (contínua ou de intervalo/razão) e varia entre 0 e 1. Vogt (1999, p. 99) definiu o Eta como um **coeficiente de correlação que não pressupõe que a relação entre as duas variáveis seja linear**. Deste modo, aplica-se especialmente quando a relação entre as duas variáveis é curvilínea e nos casos em que não é adequado usar o coeficiente de correlação de Pearson (e.g., ansiedade e desempenho num teste⁵; ver Figura 1) (Brogden, 1949).

O Eta tanto é uma estimativa de correlação como um critério preditor para uma dada pontuação num teste, no entanto não é desejável selecionar um grupo com uma determinada pontuação, mas um grupo acima de uma determinada pontuação (Brogden, 1949). Aplica-se também quando uma das variáveis é nominal e a outra é contínua.

CONCLUSÃO

Concentrámo-nos neste artigo nos tamanhos do efeito referentes a uma das partes da família r — as magnitudes da correlação —, reservando para um próximo número a proporção da variância que também permite estimar a força da relação entre as duas variáveis.

Um desses índices foi já discutido neste artigo, o *coeficiente de determinação*, mas falta então rever os tamanhos do efeito relativos à regressão múltipla, regressão logística, ANOVA, ANCOVA e MANOVA.

Face à revisão efetuada, sublinhamos de novo aqui a importância de não tomar decisões dicotômicas de aceitar ou rejeitar um resultado com base no p , mas também de não o fazer com base no valor de um coeficiente de correlação e seus pontos de corte.

Recorde-se, todos os resultados devem ser interpretados à luz do contexto do estudo e ser contrastado com os valores de outras investigações.

Em conclusão, as diretrizes que fomos enunciando ao longo do artigo pretendem estimular a apresentação dos tamanhos do efeito, sem terem de se considerar regras absolutas. Esta prática é essencial para a comunicação e contraste de resultados dentro da comunidade científica, desde que devidamente enquadrada e refletida.

Conflito de interesses | Conflict of interest: nenhum | none.

Fontes de financiamento | Funding sources: nenhuma | none.

⁵ Abrami et al. (2001) descrevem esta relação como um dos exemplos mais famosos de uma relação curvilínea nas ciências sociais, recordando que os níveis moderados de ansiedade otimizam o desempenho num teste.

REFERÊNCIAS

- Abrami, P. C., Cholsky, P. e Gordon, R. (2001). *Statistical analysis for the social sciences: An interactive approach*. Needham Heights, MA: Allyn & Bacon. ↗
- Bartlett, J. W., Seaman, S. R., White, I. R., Carpenter, J. R., for the Alzheimer's Disease Neuroimaging Initiative*. (2015). Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Statistical Methods in Medical Research*, 24(4), 462–487. doi:10.1177/0962280214521348
- Beranuy, M., Oberst, U., Carbonell, X. e Chamorro, A. (2009). Problematic Internet and mobile phone use and clinical symptoms in college students: The role of emotional intelligence. *Computers in Human Behavior*, 25(5), 1182–1187. doi:10.1016/j.chb.2009.03.001
- Berben, L., Sereika, S. M. e Engberg, S. (2012). Effect size estimation: methods and examples. *International Journal of Nursing Studies*, 49(8), 1039–1047. doi:10.1016/j.ijnurstu.2012.01.015
- Bezeau, S. e Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychology research. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, 23(3), 399–406.
- Breaugh, J. A. (2003). Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management*, 29(1), 79–97.
- Brogden, H. E. (1949). A new coefficient; application to biserial correlation and to estimation of selective efficiency. *Psychometrika*, 14(3), 169–182.
- Butters, M. A., Young, J. B., Lopez, O., Aizenstein, H. J., Mulsant, B. H., Reynolds, C. F. ... Becker, J. T. (2008). Pathways linking late-life depression to persistent cognitive impairment and dementia. *Dialogues in Clinical Neuroscience*, 10(3), 345–357.
- Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, 26, 247–272.
- Chiu, S.-I., Hong, F.-Y. e Chiu, S.-L. (2013). An analysis on the correlation and gender difference between college students' internet addiction and mobile phone Addiction in Taiwan. *ISRN Addiction*, 2013, 360607. doi:10.1155/2013/360607
- Cohen, B. H. (2001). *Explaining psychological statistics* (2ª ed.). New York: Wiley.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7(3), 249–253.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2ª ed.). New York: Lawrence Erlbaum Pub.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cohen, J. e Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2ª ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publ.
- Cohen, J., Cohen, P., West, S. G. e Aiken, L. S. (2003). *Applied multiple correlation/regression analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Costa, M., Espirito-Santo, H., Simões, S., Conde, Â., Correia, A., Almeida, R. ... Lemos, L. (2013, April 1). Correlates of elderly loneliness [Resumo]. 21st European Congress of Psychiatry. Nice. ↗
- 1549 – Correlates of elderly loneliness [Resumo]. *European Psychiatry*, 28(1), 1–6. doi:10.1016/S0924-9338(13)76559-3
- Cumming, G. (2012). *Understanding the new statistics*. New York: Routledge.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–38. doi:10.2307/2984875
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34(9), 917–928.
- Ellis, P. D. (2010). *The essential guide to effect sizes*. Cambridge: Cambridge University Press.
- Enders, W. (2015). *Applied econometric time series* (4ª ed.). Hoboken, NJ: Wiley Global Education.
- Espirito-Santo, H. e Daniel, F. B. (2015). Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): As limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos. *Revista Portuguesa De Investigação Comportamental e Social*, 1(1), 3–16.
- Ezekiel, M. (1929). The Application of the Theory of Error to Multiple and Curvilinear Correlation. *Journal of the American Statistical Association*, 24(165A), 99–104. doi:10.1080/01621459.1929.10506278
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538.
- Fermino, S., Espirito-Santo, H., Matreno, J., Daniel, F., Maia, S., Gonçalves, D. ... Gaspar, A. (2012). Diferenças sintomáticas, neuropsicológicas e sociodemográficas em idosos com doença de Alzheimer versus idosos com depressão. Em A. E. de Psicologia Conductual (Ed.), (p. 551). *Libro de Resúmenes de los Trabajos Aceptados en el V Congreso Internacional Y X Nacional de Psicología Clínica*. Santander.
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4), 507. doi:10.2307/2331838
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32. doi:10.1234/12345678
- Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron*, 3, 329-332.
- Fortenbaugh, F. C., DeGutis, J., Gemine, L., Wilmer, J. B., Grosso, M., Russo, K. e Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000. *Psychological Science*, 26(9), 1497–1510. doi:10.1177/0956797615594896
- Glass, G. V. e Hopkins, K. D. (1995). *Statistical methods in education and psychology* (3ª ed.). Needham Heights, MA: Allyn & Bacon.
- Goodwin, L. D. e Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r . *The Journal of Experimental Education*, 74(3), 251–266. doi: 10.3200/JEXE.74.3.249-266
- Granger, C. e Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, 2(2), 111–120. doi:10.1016/0304-4076(74)90034-7 ↗
- Gravetter, F. J. e Wallnau, L. B. (2013). *Statistics for the behavioral sciences* (9 ed.). Belmont, CA: Cengage Learning.
- Gupta, D. S. (1960). Point biserial correlation coefficient and its generalization. *Psychometrika*, 25(4), 393–408. doi:10.1007/BF02289756
- He, Y., Zaslavsky, A. M., Landrum, M. B., Harrington, D. P. e Catalano, P. (2010). Multiple imputation in a large-scale complex survey: a practical guide. *Statistical Methods in Medical Research*, 19(6), 653–670. doi:10.1177/0962280208101273
- Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics*, 6(2), 107–128. doi:10.3102/10769986006002107
- Hinkle, D. E., Wiersma, W. e Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5 ed.). Boston: Houghton Mifflin.
- Horton, N. J. e Kleinman, K. P. (2007). Much ado about nothing. *The American Statistician*, 61(1), 79–90. doi:10.1198/000313007X172556
- Huff, D. (1993). *How to Lie with Statistics*. New York: W. W. Norton & Company. ↗
- IBM (2013). *IBM SPSS modeler 16 algorithms guide*. IBM Corporation
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81. doi:10.2307/2332226
- Kline, R. B. (2013). *Beyond significance testing: Reforming data analysis methods in behavioral research* (2nd ed.). Washington, DC: American Psychological Association. ↗
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., et al. (2012). *Translating the statistical representation of the effects* ↗

- of education interventions into more readily interpretable forms. *National Center for Special Education Research*. National Center for Special Education Research, Institute of Education Sciences.
- MacCallum, R. C., Zhang, S., Preacher, K. J. e Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19–40.
- Matthews, R. (2000). Storks deliver babies ($p=0.008$). *Teaching Statistics*, 22(2), 36–38. doi:10.1111/1467-9639.00013
- Mendonça, J. D. e Holden, R. R. (1996). Are all suicidal ideas closely linked to hopelessness? *Acta Psychiatrica Scandinavica*, 93(4), 246–251. doi:10.1111/j.1600-0447.1996.tb10642.x
- Nakagawa, S. e Freckleton, R. P. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*, 23(11), 592–596. doi:10.1016/j.tree.2008.06.014
- Napoleão, M., Monteiro, B. e Espírito-Santo, H. (2016). Qualidade subjetiva do sono, sintomas depressivos, sentimentos de solidão e institucionalização em pessoas idosas. *Revista Portuguesa De Investigação Comportamental e Social*, 2(2), 12–24. ←
- Nelsen, R. B. (2012). Kendall tau metric. Em M. Hazewinkel (Ed.), *Encyclopedia of mathematics*. Retrieved from URL: http://www.encyclopediaofmath.org/index.php?title=Kendall_tau_metric&oldid=12869
- Nunnally, J. C. (1994). *Psychometric theory* (3ª ed.). New York: Tata McGraw-Hill Education.
- Olejnik, S. e Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25(3), 241–286. doi:10.1006/ceps.2000.1040
- Pallant, J. (2011). *SPSS Survival Manual* (4ª ed.). Crows Nest: Allen & Unwin.
- Pearson, K. (1905). Mathematical contributions to the theory of evolution XIV: On the general theory of skew correlation and non-linear regression. *Draper's Company Research Memoirs, Biometric Series II*. Londres: Dulau & Co.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, 2(2288), 1243–1246. doi:10.1136/bmj.2.2288.1243
- Pituch, K. A. e Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences* (6 ed.). New York: Routledge.
- Rodgers, J. L. e Nicewander, W. L. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59–66.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revised). Newbury Park: Sage.
- Rosenthal, R. e DiMatteo, M. R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59–82. doi:10.1146/annurev.psych.52.1.59
- Rosenthal, R. e Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74(2), 166.
- Rosnow, R. L. e Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1(4), 331–340. doi:10.1037/1082-989X.1.4.331
- Rovine, M. J. e von Eye, A. (1997). A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. *The American Statistician*, 51, 42–46.
- Rubin, D. B. (1996). Multiple Imputation after 18+ Years. *Journal of the American Statistical Association*, 91(434), 473–489. doi:10.1080/01621459.1996.10476908
- Shieh, G. (2010). Estimation of the simple correlation coefficient. *Behavior Research Methods*, 42(4), 906–917. doi:10.3758/BRM.42.4.906
- Simões, S., Ferreira, J. J., Braga, S. e Vicente, H. T. (2015). Bullying, vinculação e estilos educativos parentais em adolescentes do 3º ciclo do ensino básico. *Revista Portuguesa De Investigação Comportamental e Social*, 1(1), 30–41. ←
- Smith, B. B. (1923). *Handbook of statistical terms and methods*. Bureau of Agricultural Economics.
- Snyder, P. e Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *The Journal of Experimental Education*, 61(4), 334–349.
- Sprinthall, R. C. (2003). *Basic statistical analysis* (7ª ed.). Boston: Allyn & Bacon.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251. doi:10.1037/0033-2909.87.2.245
- Tabachnick, B. G. e Fidell, L. S. (2007). *Using multivariate statistics* (5 ed.). Boston: Pearson Education.
- Templeton, G. F. (2011). *A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research* (Vol. 28, pp. 41–58). Presented at the Communications of the Association for Information. ↙
- Thompson, B. (2006). Research synthesis: Effect sizes. In J. Green, G. Camilli e P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 583–603). Washington, DC: Routledge.
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44(5), 423–432.
- Vogt, W. P. (1999). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (2ª ed.). Thousand Oaks, CA: Sage.
- Wang, Z. e Thompson, B. (2007). Is the Pearson r^2 biased, and if so, what is the best correction formula? *The Journal of Experimental Education*, 75(2), 109–125. doi:10.3200/JEXE.75.2.109-125