

Calcular e apresentar tamanhos do efeito em trabalhos científicos (3): Guia para reportar os tamanhos do efeito para análises de regressão e ANOVAs

Calculating and reporting effect sizes on scientific papers (3): Guide to report regression models and ANOVA effect sizes

Artigo de Revisão | Review Article

Helena Espírito-Santo, PhD (1 2 a), Fernanda Daniel, PhD (1 3 b)

(1) Instituto Superior Miguel Torga, Coimbra, Portugal

(2) Centro de Investigação do Núcleo de Estudos e Intervenção Cognitivo-Comportamental da Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Portugal

(3) Centro de Estudos e Investigação em Saúde da Universidade de Coimbra

(a) Redação do manuscrito, Análise estatística.

(b) Revisão do manuscrito, Elaboração das folhas de cálculo.

Autor para correspondência | Corresponding author: Helena Espírito Santo: helenum@gmail.com

RESUMO

Palavras-Chave

ANOVA

Tamanho do efeito

Valor p

No primeiro número da *Revista Portuguesa de Investigação Comportamental e Social* foi revista a importância de calcular, indicar e interpretar os tamanhos do efeito para as diferenças de médias de dois grupos (família d dos tamanhos do efeito). Os tamanhos do efeito são uma métrica comum que permite comparar os resultados das análises estatísticas de diferentes estudos, informando sobre o impacto de um fator na variável em estudo e sobre a associação entre variáveis.

Depois de rever os tamanhos do efeito para as diferenças de médias entre dois grupos (Espírito-Santo e Daniel, 2015) e a maior parte da família r (Espírito-Santo e Daniel, 2017), faltava rever os tamanhos do efeito para a análise da variância. A análise da variância pode ser compreendida como uma extensão da família d a mais de dois grupos (ANOVA) ou como uma subfamília r em que a proporção da variabilidade é imputável a um ou mais fatores. Na subfamília r revista neste estudo, analisa-se a mudança na variável dependente que decorre de uma ou mais variáveis independentes. Esta análise debruça-se sobre os modelos lineares gerais, onde se incluem os modelos de regressão e a ANOVA.

Este artigo fornece as fórmulas para calcular os tamanhos do efeito mais comuns, revendo os conceitos básicos sobre as estatísticas e facultando exemplos ilustrativos computados no *Statistical Package for the Social Sciences (SPSS)*. As orientações para a interpretação dos tamanhos do efeito são também apresentadas, assim como as cautelas no seu uso. Adicionalmente, o artigo acompanha-se de uma folha de cálculo em Excel para facilitar e agilizar os cálculos aos interessados.

ABSTRACT

Keywords

ANOVA

Effect size

p -value

Statistical significance

In the first issue of the Portuguese Journal of Behavioral and Social Research, the importance of calculating, indicating, and interpreting the effect sizes for the differences of means of two groups (d family of effect sizes) was reviewed. Effect sizes are standard metrics that allows the comparison of the results of statistical analyzes of different studies. Effect sizes also report on the impact of a factor on the variable under investigation and the association between variables.

After reviewing the effect sizes for the mean differences between two groups (Espírito-Santo & Daniel, 2015) and most of the r family (Espírito-Santo & Daniel, 2017), the review of effect sizes for analysis of variance was lacking. Analysis of variance can be understood as an extension of the d family to more than two groups (ANOVA) or as an r subfamily in which the proportion of variability is attributable to one or more factors. In the r subfamily reviewed in this study, we analyse the change in the dependent variable that results from one or more independent variables. This analysis is focused on general linear models, including regression models and ANOVA.

This article provides the formulas for calculating the most common effect sizes by reviewing the basic concepts of the statistics and providing illustrative examples computed in the Statistical Package for the Social Sciences (SPSS). The guidelines for the interpretation of effect sizes are also presented, as well as the cautions in their use. Also, the article is accompanied by an Excel spreadsheet to facilitate and expedite the calculations for interested readers.

INTRODUÇÃO

Quando se descrevem estatisticamente os resultados de um estudo importa apresentarem-se as medidas do *tamanho do efeito* (TDE) junto com os testes de significância estatística (Berben, Sereika e Engberg, 2012; Cohen, 1992a; Cumming, 2012; Ferguson, 2009; Keppel e Wickens, 2004; Kline, 2013; Lipsey et al., 2012; Olejnik e Algina, 2000; 2003; Snyder e Lawson, 1993).

A indicação dos TDE é fortemente recomendada pela *Task Force on Statistical Inference* da *American Psychological Association* (Wilkinson e Task Force on Statistical Inference, 1999), sendo também aconselhada pela política editorial da **Revista Portuguesa de Investigação Comportamental e Social**. A *American Psychological Association* (2010) recomenda inclusivamente que os TDE sejam apresentados, mesmo que os valores de p sejam não-significativos.

Um TDE é um indicador padronizado que, ao contrário do valor p , não depende do tamanho da amostra e possibilita a comparação entre resultados de diversos estudos (Berben et al., 2012; Bezeau e Graves, 2001; Cumming, 2012; Ferguson, 2009; Kline, 2013; Lipsey et al., 2012; Olejnik e Algina, 2000; Snyder e Lawson, 1993), sendo útil para os estudos de metanálise ou mera comparação dos resultados entre estudos (Berben et al., 2012; Cumming, 2012; Kline, 2013; Lipsey et al., 2012; Rosenthal, 1991).

Para cada teste estatístico existe um TDE apropriado (Cohen, 1992a; P. D. Ellis, 2010), podendo os vários TDE ser classificados em duas famílias. Uma delas, a família d , inclui os TDE relativos à magnitude da diferença. A outra, a família r , diz respeito à magnitude da relação entre variáveis ou da proporção da variância explicada (Hedges, 1981; Olejnik e Algina, 2003; Rosenthal, 1994).

No seguimento dos artigos anteriores para a **Revista Portuguesa de Investigação Comportamental e Social** (Espírito-Santo e Daniel, 2017; 2015), faltava completar os TDE no que diz respeito à análise da variância. A análise de variância pode ser entendida de duas formas: (1) como um alargamento da família d a mais de dois grupos ou (2) como uma subfamília r em que a proporção da variabilidade é atribuível a um efeito (Ialongo, 2016; Keppel e Wickens, 2004).

Na subfamília r considera-se que a mudança na variável dependente (Y) é determinada por ou uma ou

mais variáveis independentes (X). Assim, a relação entre essas variáveis, de forma simplificada, é uma função linear $Y = bX + e$ que se adequa aos modelos lineares gerais, em que se incluem a análise de variância (ANOVA), a regressão linear e outros modelos estatísticos que decorrem dessa função linear (Ialongo, 2016).

Os TDEs para esta família, originalmente, foram propostos por Fisher (1925) que considerava que o nível de significância na análise de variância deveria ser acompanhado pelo rácio de correlação ou pelo etaquadrado. Desde então, têm sido propostas várias medidas da magnitude do efeito. Neste artigo de revisão vamos dar conta das mais comuns, fornecendo as respetivas equações e exemplos ilustrativos.

EXTENSÃO À FAMÍLIA d : ANOVA

Análise da variância de um fator

Conceitos básicos

A **análise de variância** (ANOVA) **de um fator** ou de uma via consiste na comparação do efeito fixo¹ do fator sobre uma variável dependente Y .

O **fator** é a variável independente (X) que pode assumir vários níveis.

Os **níveis** de um fator correspondem a diferentes grupos ou a diferentes categorias. Por exemplo, o fator pode ser a variável *idade* categorizada em três níveis que equivalem a três grupos etários (14-24 anos; 25-64 anos; ≥ 65 anos).

Tome-se como exemplo a resposta à pergunta: qual o efeito da idade (*fator*) sobre o quociente de inteligência medido pelo teste das Matrizes Progressivas de Raven (MP-Raven) (*variável dependente*) no projeto *Estudos Normativos de Instrumentos Neuropsicológicos* a decorrer no Instituto Superior Miguel Torga ($N = 752$ pessoas)?

Neste exemplo o fator idade apresenta três níveis (e.g., 14-24 anos; 25-64 anos; ≥ 65 anos) e a variável dependente é representada pelas pontuações no MP-Raven (*variável contínua*).

Como se pode analisar o efeito de um fator sobre uma variável dependente? Uma forma seria pegar em cada dois pares de comparações como análises diferentes para dois grupos (A. Field, Miles e Field, 2012; Keppel e Wickens, 2004).

¹ Só discutimos aqui os fatores de efeito fixo em que as observações são realizadas em todos os níveis do fator de interesse. Ou seja, os níveis da variável independente são escolhidos especificamente pelo investigador (não-aleatórios). O fator aleatório assume teoricamente todos os níveis

possíveis da variável independente, pretendendo o investigador generalizar os dados a outros níveis da variável independente (Olejnik e Algina, 2000).

Retome-se o exemplo: comparava-se (1) o grupo dos 14-24 anos com o dos 25-64 anos, (2) o grupo dos 14-24 anos com o dos ≥ 65 anos e (3) o grupo dos 25-64 anos com o dos ≥ 65 anos.

O problema em proceder assim é que se está a usar o mesmo conjunto de dados uma e outra vez e cada conjunto aparece em duas comparações. Dessa forma, o Erro do Tipo I² aumenta (A. Field et al., 2012; Stevens, 2007). Na verdade, as comparações não são *independentes* (ainda assim, ver as secções mais à frente sobre os contrastes teóricos). Outra forma é considerar todas as categorias numa *única* avaliação e efetuar uma ANOVA com um fator.

Designa-se por **ANOVA** (*análise de variância*) porque compara a variabilidade de valores ou pontuações (variância) **entre** os grupos com a variabilidade **em** cada grupo.

A variância *entre* grupos é a soma dos quadrados das diferenças entre as médias dos grupos e a média global e é aqui designada por $SQ_{\text{entre grupos}}$ ou SQ_{fator} . A *média global* é a média de todos os dados considerando a amostra como um todo.

A variabilidade *em cada grupo* é designada por *termo de erro* (Keppel e Wickens, 2004; Pallant, 2011, p. 242), *variância de erro* (Cohen, 1973) ou *erro normal do acaso* (Ialongo, 2016). Assim, a variabilidade em cada grupo é a soma dos quadrados das diferenças entre os dados e as médias de cada grupo e é aqui representada por $SQ_{\text{nos grupos}}$ ou SQ_{erro} .

Parte-se da premissa que a variabilidade de valores *entre* os grupos se deve à variável independente e que a variabilidade *em* cada grupo se deve ao acaso (Pallant, 2011, p. 242).

No exemplo, a variabilidade das pontuações no MP-Raven entre as três faixas etárias dever-se-á à própria idade (14-24 anos; 25-64 anos; ≥ 65 anos) e a variabilidade das pontuações em cada faixa etária dever-se-á ao acaso.

A **estatística F** da ANOVA corresponde à divisão da variância *entre* grupos pela variância *nos* grupos (Cohen, 1973; 1992a; Ialongo, 2016; Keppel e Wickens, 2004; Pallant, 2011):

$$\frac{\text{Variabilidade entre diferentes grupos de sujeitos}}{\text{Variabilidade entre sujeitos dos mesmos grupos}}$$

Assim, um *F* alto significa que há mais variabilidade entre os grupos do que variabilidade *em cada grupo* (Pallant, 2011).

Continuando com o exemplo (Tabela 1), o *F* de 106,12 resultou da divisão entre o quadrado médio entre grupos ($SQ_{\text{entre grupos}} / gl_{\text{entre grupos}} = 9332,48$) pelo quadrado médio nos grupos ($SQ_{\text{nos grupos}} / gl_{\text{nos grupos}} = 87,94$).

Nota. Comandos no SPSS: Analisar > Comparar Médias > Análise de Variância Unidirecional > Lista de Variáveis Dependentes (MP-Raven) e Fator (Idade) > Opções > Estatísticas Descritivas.

Tabela 1

Exemplo de ANOVA das Pontuações do Teste das Matrizes Progressivas de Raven Entre os Grupos Definidos Pela Idade Calculado Através do Programa SPSS-25

Grupos	<i>n</i>	<i>M</i>	<i>DP</i>	<i>IC 95%</i>	
14-24 anos	231	47,73	7,07	46,82	48,65
25-64 anos	420	45,04	9,72	44,10	45,97
≥ 65 anos	101	31,81	12,15	29,41	34,31
Total	752	44,09	10,61	43,33	44,85

Componente da variância	<i>SQ</i>	<i>gl</i>	<i>QM</i>	<i>F</i>	<i>p</i>
Entre grupos	18664,96	2	9332,48	106,12	< 0,001
Nos grupos	65867,25	749	87,94		
Total	84532,21	751			

Nota. *SQ* = soma dos quadrados; *gl* = graus de liberdade; *QM* = quadrado médio; *F* = estatística *F* da ANOVA; *p* = nível de significância.

O **valor do *p***, que acompanha a estatística *F* abrangente ou global (*omnibus*), indica se se pode rejeitar ou não a hipótese nula de que as médias são semelhantes para todos os grupos (Pallant, 2011).

Assim, na Tabela 1 para o exemplo, pode afirmar-se que houve diferenças estatisticamente significativas nas pontuações do MP-Raven entre os grupos etários ($F = 106,12$; $p < 0,011$).

Para saber que grupos diferem entre si há que conduzir testes pós ANOVA — **testes post-hoc** — ou efetuar somente algumas comparações específicas — **contrastos teóricos**. Os testes *post-hoc* protegem a probabilidade do Erro do tipo I, mas dificultam a descoberta de diferenças significativas. No entanto, os contrastes teóricos só se devem usar quando há razões

² O erro do Tipo I significa rejeitar falsamente a hipótese nula (A. Field et al., 2012).

teóricas claras (Pallant, 2011) (ver em secção mais à frente sobre a ANOVA).

No exemplo, uma vez que o teste de homogeneidade de variâncias de Levene foi significativo ($W = 25,14$; $p < 0,001$) e os grupos tinham dimensão diferentes (Tabela 1), recorreu-se aos *post-hoc* de Games-Howell que revelaram diferenças estatisticamente significativas ($p < 0,001$) em todos os pares de comparações. Assim, a média de pontuações no grupo dos 14 aos 24 anos foi estatisticamente significativamente superior quando comparada com grupo dos 25 aos 64 anos e com a do grupo com idade superior aos 65 anos. A média do grupo dos 25 aos 64 anos foi também estatisticamente significativamente superior quando comparada com grupo com idade acima dos 65 anos.

Calculando o tamanho do efeito para a ANOVA

Considerando a família d e ampliando-a à diferença de mais de dois grupos, deve corrigir-se a diferença global (diferença de cada pontuação em relação à média de todas as pontuações) pelo número de grupos. Ou seja, tem-se em conta o efeito global (*omnibus*) da ANOVA a um fator de efeito fixo (Ialongo, 2016).

Mais, a estatística F é sensível ao tamanho da amostra e com amostras grandes tende a ser maior e o valor p tende a ser menor. Assim, é útil apresentar uma medida de TDE única que sumarie o efeito da comparação global ou efeito *omnibus* (Keppel e Wickens, 2004).

Para esse propósito, pode calcular-se o TDE Psi (Ψ) ou f^* de Steiger e Fouladi ou efeito padronizado quadrático médio [*root mean square standardized effect (RMSSE)*] (Steiger, 2004; Steiger e Fouladi, 2016). O índice $RMSSE$ representa a diferença coletiva de todo o modelo padronizado pela raiz quadrática média:

$$RMSSE = \sqrt{\frac{\sum(\bar{y} - MG)^2}{(k-1)QM_{erro}}} \quad \text{Equação 1}$$

Nesta equação, \bar{y} é a média de cada grupo; MG a média global; Σ é o somatório; k o número de grupos e QM_{erro} o quadrado médio de erro que corresponde à variância de erro dividida pelo N menos o número de grupos ($QM_{erro} = SQ_{erro} / (N - k)$).³

Continuando com o exemplo, e tendo em conta que a MG foi de 44,09, o cálculo fornece um TDE de 0,97 que corresponde a um TDE grande. O cálculo pode ser observado em detalhe a seguir:

$$QM_{erro} = 65867,25 / (752 - 3) = 87,94;$$

$$RMSSE = \sqrt{\frac{(47,73 - 44,09)^2 + (45,04 - 44,09)^2 + (31,81 - 44,09)^2}{(3-1) \times 87,94}}$$

Pelo facto de o índice $RMSSE$ ser afetado por dados assimétricos, longas caudas e/ou valores extremos, Zhang e Algina (2011) propõem uma formulação mais robusta com variações para a ANOVA de um fator e ANOVA de um fator com ns iguais. Os interessados podem consultar as fórmulas 1, 2, e 9, pp. 79-80 de Zhang e Algina (2011) ou rever a estimativa f de Cohen (1992b, p. 274-275).

EXTENSÃO À FAMÍLIA r : variabilidade explicada

Como enunciámos na introdução, na subfamília r considera-se que a variabilidade na variável dependente (Y) é uma função linear de uma ou mais variáveis independentes (X), representada por:

$$Y = bX + e \quad \text{Equação 2}$$

Nos modelos lineares gerais, X é nomeada por planeamento (*design*) e pode ser uma ou um conjunto de variáveis independentes, b é a ponderação, e o erro normal aleatório e Y varia com as mudanças em X (Bewick, Cheek e Ball, 2003; Ialongo, 2016).

Na subfamília r vamos considerar a regressão e a ANOVA de novo, pois este teste pode ser considerado como um caso especial de regressão (A. Field et al., 2012).

Regressão linear

Conceitos básicos

Na regressão linear pretende saber-se se Y — aqui designada por **variável de desfecho** — varia segundo a forma como X muda (seja uma variável ou um conjunto delas), usando a associação entre as variáveis para prever como é que isso acontece (Bewick et al., 2003; Ialongo, 2016; Pallant, 2011)⁴. Neste modelo estatístico, as variáveis do planeamento são contínuas, pelo que a predição é feita ponto-a-ponto entre X e Y (Ialongo, 2016).

No exemplo anterior, quer saber-se o efeito da idade (preditor ou variável X) sobre as pontuações do teste MP-Raven (variável desfecho ou Y). Ou seja, pretende estimar-se a relação linear subjacente por forma a que se consiga prever a pontuação do MP-Raven para uma determinada idade. A análise de regressão efetuada

³ O QM_{erro} também se designa por “erro quadrado médio”, aparecendo o seu valor no SPSS na interseção entre o *Quadrado médio* e *Nos grupos*.

⁴ A regressão é uma forma de correlação que se usa com a intenção de fazer a análise de predição (Tabachnick e Fidell, 2007).

através do SPSS-25 forneceu os dados para se construir a equação linear (Tabela 2):

$$\text{MPRaven} = (-0,27) \times \text{Idade} + 55,05$$

Assim, para cada ano que se avança na idade, prevê-se que as pontuações do MP-Raven desçam 0,27 pontos (a partir de um ponto médio de 55,05). O contributo da idade para a predição das pontuações no MP-Raven foi estatisticamente significativo ($p < 0,001$).

Nota. Comandos no SPSS: Analisar > Regressão > Linear > Dependente (MP-Raven); Independente (Idade).

Tabela 2

Exemplo de Regressão Linear Calculando a Capacidade da Idade Predizer as Pontuações do Teste das Matrizes Progressivas de Raven Através do Programa SPSS-25

Resumo do modelo						
Modelo	R	R ²	\bar{R}^2	EP		
1	0,49	0,244	0,243	9,231		
ANOVA						
Modelo	SQ	gl	QM	Z	p	
1 Regressão	20622,32	1	20622,32	242,01	< 0,001	
Resíduo	63909,89	750	85,21			
Total	84532,21	751				
Coeficientes						
Modelo	Não-padronizados		Padronizados		t	p
	B	Erro	Beta			
1 (Constante)	55,05	0,781			70,49	< 0,001
Idade	-0,271	0,017	-0,494		-15,56	< 0,001

Nota. R = correlação de Pearson; R = coeficiente de determinação; \bar{R}^2 = R ajustado; EP = erro padrão da estimativa; SQ = Soma dos quadrados; QM = Quadrado médio; Z = estatística F; B da constante = intercepto e da equação de regressão ou constante de Euler; B da idade = declive b da equação de regressão; t = valores t para os testes de H0: intercepto = 0 e H0: declive = 0; p = nível de significância.

Calculando o tamanho do efeito para a regressão linear

Regressão linear simples. O TDE para a regressão linear corresponde ao coeficiente de determinação ou *rácio de correlação*, dado pela seguinte equação (Ialongo, 2016; Keppel e Wickens, 2004; Keren e Lewis, 1969; Richardson, 2011):

$$R^2 = \frac{[\sum(x_i - \bar{x})(y_i - \bar{y})]^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} = \frac{SQ_{regr}}{SQ_{total}} \quad \text{Equação 3}$$

Em que x_i , y_i são os valores das variáveis X e Y; \bar{y} ; \bar{x} são as médias do grupo/amostra e Σ é o somatório. Ou,

em que SQ_{regr} é a soma dos quadrados relativa à regressão e SQ_{total} é a soma dos quadrados total.

O R-quadrado representa então o ajustamento entre a variável dependente e a variáveis independente preditora que designa o efeito (Keren e Lewis, 1969). Dito ainda de outro modo, o R-quadrado é a proporção da variância de Y que pode ser predita ou explicada com base na sua relação linear com X (Hays, 1963).

Quanto mais forte for o coeficiente de determinação, mais forte é a predição ou a força da associação entre a variável dependente Y e os múltiplos preditores X_i em termos lineares (Ferguson, 2009). O valor teórico mínimo de R-quadrado é de 0% e o máximo de 100%, ainda que o máximo não seja expectável nas ciências sociais⁵.

No exemplo, para a contribuição da idade para as pontuações do MP-Raven o TDE foi de 0,244; ou seja, a idade contribuiu em 24,4% para a variância das pontuações do MP-Raven. Este valor pode ser obtido facilmente no *output* do SPSS (ver Tabela 2) ou calculado através da Equação 3:

$$R^2 = \frac{20622,32}{84532,21} = 0,244$$

Uma vez que o R-quadrado tende a ser sobrestimado em amostras pequenas (Pallant, 2011; Tabachnick e Fidell, 2007) e aumenta quando se adiciona uma nova variável ao modelo (mesmo que não se associe à variável-desfecho), há que corrigir o efeito.

Regressões lineares múltiplas. Assim, nestas regressões em que pode haver mais do que uma variável preditora (X_1 , X_2 , etc.) pode usar-se o R-quadrado ajustado (Tabela 2) (P. D. Ellis, 2010; Ialongo, 2016; Pallant, 2011; Vacha-Haase e Thompson, 2004)⁶.

Nesta correção, é feito o ajustamento para o número de variáveis preditoras e de casos (Ialongo, 2016), sendo apropriado para amostras pequenas (Tabachnick e Fidell, 2007). Deste modo, o R-quadrado ajustado aumenta somente se o novo preditor melhorar o modelo mais do que seria esperado pelo acaso. A equação é a que se segue (Wherry, 1931):

$$\bar{R}^2 = \left(\frac{(N-1)R^2 - (p-1)}{N-p} \right) \quad \text{Equação 4}$$

Na equação, R^2 é o coeficiente de determinação, N o tamanho da amostra e p o número de preditores (VI).

⁵ Segundo a revisão de Ellis (2010), alguns não consideram útil o coeficiente de determinação, pois combina os efeitos de vários preditores. A contribuição individual de cada preditor, ou seja, a alteração em Y quando X_1 é alterado por uma unidade enquanto se controlam todos os outros preditores (X_2 , X_3 ... etc.), deve ser examinada

através dos coeficientes de correlação parcial (Hair, Black, Babin e Anderson, 2009).

⁶ O R-quadrado ajustado adequa-se também às análises em que as amostras são pequenas (P. D. Ellis, 2010; Ialongo, 2016; Pallant, 2011).

O *R*-quadrado ajustado pode também obter-se na análise de regressão de linear realizada através do SPSS (ver a Tabela 2 do exemplo). No entanto, como Ialongo (2016) faz notar, o ajustamento do *R*-quadrado não é uma medida do TDE, é, ao invés, uma medida da adequação do conjunto de preditores para o modelo.

Outras fórmulas do *R*-quadrado ajustado e sua discussão podem ser consultadas em Herzberg (1969), Snyder e Lawson (1993) e Stevens (2007).

Existem ainda outras fórmulas de correção para amostras com menos de 60 sujeitos e vários preditores [para uma revisão ver Tabachnick e Fidell (2007)].

Note-se que o *R*-quadrado pode ser negativo e, ainda que matematicamente possível, indica que o modelo tem problemas sérios (Vacha-Haase e Thompson, 2004).

Em alternativa ao *R*-quadrado ajustado para regressões lineares múltiplas com mais do que uma variável preditora (X_1, X_2), pode calcular-se o valor f^2 de Cohen⁷ que indica o efeito incrementador de se adicionar X_2 ao modelo (Cohen, 1992b). A fórmula é a seguinte (Ialongo, 2016):

$$f^2 = \frac{R^2}{1-R^2} \quad \text{Equação 5}$$

De novo, na equação R^2 é o coeficiente de determinação.

No exemplo anterior, quer-se agora saber o efeito da idade e do grau de escolaridade (preditoras) sobre as pontuações do teste MP-Raven (Y) (Tabela 3; Coluna Correlações Parte). Comparando o contributo de cada variável preditora (coluna dos *Beta*), o coeficiente maior foi o referente à escolaridade ($Beta = 0,49$; $p < 0,001$), sendo o contributo da idade menor ($Beta = -0,13$; $p < 0,01$), com ambos os contributos a serem estatisticamente significativos.

Neste modelo, a idade contribuiu em 0,79% ($0,089^2 \times 100$) e a escolaridade em 11,2% ($0,334^2 \times 100$) para a variância das pontuações do MP-Raven com um f^2 de 0,54 que corresponde a um alto valor preditivo.

Nota. Comandos no SPSS: Analisar > Regressão > Linear > Dependente (MP-Raven); Independent (Idade; Escolaridade).

⁷ Deve-se notar que muitas destas estimativas de teste exigem distribuição normal das observações e homocedasticidade (variações iguais para os grupos que estão sendo comparados). Caso isso não aconteça, o poder estatístico dos testes cai, e a detecção de efeitos torna-se mais difícil (P. D. Ellis, 2010). Existem métodos estatísticos alternativos para a violação daqueles pressupostos (Erceg-Hurn e Mirosevich, 2008;

Tabela 3

Exemplo de Regressão Linear Calculando a Capacidade da Idade e Escolaridade Predizerem as Pontuações do Teste das Matrizes Progressivas de Raven Através do Programa SPSS-25

Resumo do modelo					
Modelo	<i>R</i>	R^2	\bar{R}^2	<i>EP</i>	
1	0,59	0,35	0,35	8,56	
Coefficientes					
Modelo	Não-padronizados		Padronizados	<i>t</i>	<i>p</i>
	B	Erro			
1 (Constante)	27,69	2,58		10,73	<0,001
Idade	-0,07	0,02	-0,130	-2,96	0,003
Escolaridade	3,39	0,31	0,489	11,11	<0,001
Correlações					
	Ordem zero	Parciais	Parte		
1 (Constante)					
Idade	-0,49	-0,11	-0,09		
Escolaridade	0,59	0,38	0,33		

Nota. *R* = correlação de Pearson; R^2 = coeficiente de determinação; \bar{R}^2 = *R* ajustado; *EP* = erro padrão da estimativa; B da constante = intercepto e da equação de regressão ou constante de Euler; B da idade = declive *b* da equação de regressão; *t* = valores *t* para os testes de H_0 : intercepto = 0 e H_0 : declive = 0; *p* = nível de significância; Parte = coeficiente de correlação semiparcial. Por economia de espaço, não se apresenta o modelo da ANOVA.

Uso da regressão para comparar médias. Pode recorrer-se à regressão linear para comparar as médias, e, nesse sentido, um TDE útil pode ser o expresso na equação seguinte (Fritz, Morris e Richler, 2012):

$$r = \sqrt{\frac{t^2}{t^2 + gl}} \quad \text{Equação 6}$$

Nesta equação, *t* é estatística *t* de Student e *gl* os graus de liberdade (número de grupos – 1). Esta equação é também designada por correlação ponto biserial (Fritz et al., 2012).

Retomemos o exemplo e use-se a variável postíça (*dummy*)⁸ “sexo” na regressão linear do teste MP-Raven (Tabela 4).

Nota. Comandos no SPSS: Analisar > Regressão > Linear > Dependente (MP-Raven); Independent (Sexo).

Keselman, Algina, Lix, Wilcox e Deering, 2008) e TDEs igualmente robustos, como por exemplo, a probabilidade de superioridade ou TDE em Linguagem Comum (ver revisão em Espírito-Santo e Daniel (2015).

⁸ As variáveis postíças ou *dummy* são variáveis nominais (e.g., sexo, religião) que são transformadas em numéricas (e.g., sexo masculino = 0 e sexo feminino = 1) (Cohen, 1992).

Tabela 4

Exemplo de Regressão Linear Calculando a Capacidade do Sexo Predizer as Pontuações do Teste das Matrizes Progressivas de Raven Através do Programa SPSS-25

Resumo do modelo					
Modelo	R	R ²	\bar{R}^2	EP	
1	0,10	0,01	0,009	10,563	
ANOVA					
Modelo	SQ	gl	QM	F	p
Regressão	852,70	1	852,703	7,64	0,006
Residual	83679,50	750	111,57	111,57	
Total	84532,21	751			
Coeficientes					
Modelo	Não-padronizados		Padronizados	t	p
	B	Erro	Beta		
1 (Constante)	47,41	1,26		37,597	<0,001
Sexo	-2,141	0,774	-0,100	-2,765	0,006

Nota. R = correlação de Pearson; R = coeficiente de determinação; \bar{R}^2 = R ajustado; EP = erro padrão da estimativa; B da constante = intercepto e da equação de regressão ou constante de Euler; B da idade = declive b da equação de regressão; t = valores t para os testes de H0: intercepto = 0 e H0: declive = 0; p = nível de significância.

Compare-se com os obtidos através da análise do teste t de Student: $t_{(750)} = -2,765$ e $p = 0,006$ e note-se que $F = 7,64$ na regressão é o quadrado do valor de t. O R-quadrado é igual a $t^2 / t^2 + gl$ ($7,64 / 7,64 + 750$).

De acordo com a Equação 6, r é igual a 0,100, o que significa que o impacto do sexo sobre as pontuações das MP-Raven é pequeno⁹.

Um caso especial: regressão logística. A regressão logística é uma forma especial de regressão a que se recorre quando a variável dependente é dicotómica (Hair et al., 2009; Pallant, 2011)¹⁰. Devido à sua relativa simplicidade e propriedades tem-se tornado popular (Pampel, 2000).

O TDE associado é o coeficiente de logit¹¹ ou a razão de chances (*odds ratio*) (P. D. Ellis, 2010).

Ora os logits, não são facilmente interpretáveis, pelo que é usual transformar os logits em métricas mais intuitivas, como as chances (*odds* ou razão de probabilidades) e a razão de chances (*odds ratio* ou rácio das razões de probabilidade) (P. D. Ellis, 2010; Pampel, 2000).

⁹ As variáveis independentes foram selecionadas com propósitos pedagógicos, não existindo preocupação em comprovar hipóteses previamente estabelecidas.

¹⁰ A regressão logística é uma combinação de regressão múltipla (uma ou mais variáveis X predizem uma única Y) e de análise discriminante (variável Y não-métrica). Distingue-se da análise discriminante porque pode incluir variáveis X métricas e não métricas e não exige normalidade multivariada. Se a variável Y tiver mais de dois níveis, a análise discriminante pode ser mais indicada (Hair et al., 2009).

Continuando com o exemplo, depois de recategorizar a variável MP-Raven (Y) em “capacidades intelectuais abaixo da mediana” (valor 0) e “capacidades intelectuais acima da mediana” (valor 1), computou-se a regressão logística para verificar se $Y_{\text{categorizada}}$ variava em função do sexo (X). O logit foi de $0,503 - 0,375 * X$ e o rácio de chances de 0,689 (indicando que as mulheres tenderam menos a ter capacidades intelectuais acima da mediana do que os homens).

Transformando para métricas interpretáveis, 53,3% dos homens e 44,0% das mulheres apresentaram capacidades intelectuais acima da mediana. O que significa que as chances de os homens apresentarem capacidades intelectuais acima da mediana foram de 1,14 (0,533 / 0,467), o que indica que cerca de 11 homens estão acima da mediana para cada 10 que estão abaixo. As chances para as mulheres apresentarem capacidades intelectuais acima da mediana foram de 0,79 (0,440 / 0,560), o que indica que cerca de 8 mulheres estão acima da mediana para cada 10 que estão abaixo. A razão de chances dos homens para as mulheres foi de 1,44 (1,14 / 0,79), o que significa as chances de ter capacidades intelectuais acima da mediana são cerca de 1,44 vezes maiores para os homens do que para as mulheres (chances baixas).

Nota. Comandos no SPSS: Analisar > Regressão Logística Binária > Dependente (MP-Raven recategorizada); Covariáveis (Sexo).

ANOVA de novo: 1 fator ou n-fatores

Conceitos básicos

À semelhança da regressão linear, na ANOVA também se intenta perceber se Y varia quando X muda, mas agora X é uma variável discreta ou nominal variável (fator) e as suas mudanças são discretas (i.e., muda de uma categoria ou nível para outra/o).

Neste modelo estatístico, a predição é feita ponto-a-nível, sendo o cálculo um pouco diferente, mas conceptualmente semelhante (Ialongo, 2016). O efeito pode ser entendido como uma mudança no agrupamento de Y em relação ao valor de X (Ialongo, 2016).

Seguindo o mesmo exemplo, pretende descrever-se quanto variam as pontuações no teste MP-Raven (Y) com as mudanças nas faixas etárias (X).

¹¹ O logit é o logaritmo das chances (Hair et al., 2009). Considere-se que cada sujeito tem a probabilidade P_i de ter uma característica. Assim, a variável dependente assume os valores de 0 e de 1 e as chances são o rácio de P_i com $1-P_i$. A transformação logit consiste no logaritmo das chances de ter a característica [$L_i = \log [P_i / (1 - P_i)]$] (Pampel, 2000). No SPSS, o logit L_i é igual ao beta da constante mais o beta da variável independente vezes X. Os *odds ratio* ou razão das chances encontram-se em $\text{Exp}(B)$.

ANOVA com um fator. Já vimos anteriormente, na ANOVA de uma via compara-se a variabilidade de pontuações *entre* grupos (soma dos quadrados das diferenças entre as médias dos grupos e a média global = SQ_{fator}) com a variabilidade *em* cada grupo (soma dos quadrados das diferenças entre cada observação e a média do seu grupo = SQ_{erro}).

A variável independente (X) ou fator contém um conjunto de níveis diferentes (grupos ou categorias) (e.g., Pallant, 2011).

ANOVA com dois ou mais fatores entre grupos (independente). Nesta ANOVA significa que existem duas ou mais variáveis independentes e *entre grupos* significa que os sujeitos em cada grupo são diferentes.

Este tipo de ANOVA permite ainda testar o **efeito principal** para cada variável independente e verificar ainda se existe um **efeito de interação**. Se o efeito de uma variável independente sobre a variável dependente decorrer do nível de uma segunda variável independente, então houve efeito de interação (e.g., Pallant, 2011).

Com a introdução de mais uma variável independente, recorremos à Figura 1 para resumir os tipos de variância.

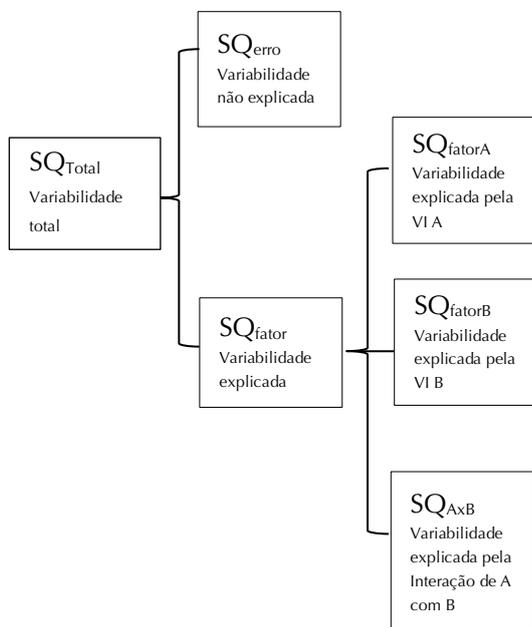


Figura 1. Síntese gráfica dos tipos de variabilidade envolvidos na ANOVA com dois ou mais fatores (VI = variáveis independentes). **Nota.** SQ = Soma dos quadrados ou variância.

Mantendo o estudo do efeito da idade, verificámos que os mais velhos tiveram piores pontuações do que os

dois grupos mais novos no MP-Raven. Agora, podemos questionar: isso continua a acontecer para os dois sexos?

Assim, consideraram-se a idade e o sexo como fatores fixos e as pontuações do MP-Raven como variável dependente. Os resultados são apresentados na Tabela 5.

Primeiramente, verificou-se que não houve efeito de interação ($p = 0,164$) — se houvesse, seria mais difícil interpretar os efeitos principais —, significando que não ocorreram diferenças significativas no efeito da idade sobre as pontuações do MP-Raven para o sexo feminino e masculino. Depois, o efeito principal da idade foi significativo ($p < 0,001$), assim como foi significativo o efeito do sexo ($p = 0,045$).

Nota. Comandos no SPSS: Analisar > Modelo Linear Geral > Univariado > Dependente (MP-Raven); Fatores Fixos (Idade; Sexo).

Tabela 5

Exemplo de ANOVA das Pontuações do Teste das Matrizes Progressivas de Raven Entre os Grupos Definidos Pela Idade e pela Escolaridade Calculado Através do Programa SPSS-25

Grupos		n	M	DP
14-24 anos	Masculino	88	47,52	7,87
	Feminino	143	47,86	6,55
25-64 anos	Masculino	215	46,22	9,23
	Feminino	205	43,79	10,09
≥ 65 anos	Masculino	35	33,71	10,83
	Feminino	66	30,80	12,77

Testes de efeitos entre sujeitos						
Origem	SQ	gl	QM	F	p	η_p^2
MC	19486,20	5	3897,24	44,697	<0,001	0,23
Intercepto	872195,18	1	872195,18	10003,04	<0,001	0,93
Idade	15974,80	2	7987,40	91,606	<0,001	0,20
Sexo	350,08	1	350,08	4,015	0,045	0,005
Idade x Sexo	315,85	2	157,93	1,811	0,164	0,005
Erro	65046,00					
Total	1546218,00					
Total corrigido	84532,21					

Nota. MC = Modelo corrigido; SQ = soma dos quadrados; gl = graus de liberdade; QM = quadrado médio; F = estatística F da ANOVA; p = nível de significância; η_p^2 = eta-quadrado parcial.

^a R = 0,231 (R-ajustado = 0,225).

Outros tipos de ANOVA. Pode ser confuso entender as ANOVAS pois há vários tipos (Tabela 6), mas todas elas têm dois aspetos em comum: (1) envolvem um determinado número de variáveis independentes ou fatores, (2) as variáveis independentes podem ser medidas usando participantes diferentes ou usando os mesmos participantes.

Se se usarem sujeitos diferentes, a ANOVA diz-se *independente* (todas as ANOVAS das secções anteriores),

se se usarem os mesmos sujeitos, designa-se por ANOVA de *medidas repetidas*.

Quando há duas ou mais variáveis independentes, é possível que algumas incluam os mesmos participantes e outras envolvam sujeitos diferentes. Neste caso a ANOVA designa-se por *mista* (A. Field et al., 2012).

Tabela 6

Tipos de ANOVAS Consoante o Número de Variáveis Independentes (VI) e Consoante os Sujeitos são Diferentes (≠) ou os Mesmos (=)

ANOVA Tipo	# VI	Sujeitos
De um fator independente (Entre sujeitos [between-subjects])	1	≠
De um fator medidas repetidas [within-subjects]	1	=
De n-fatores independente (Entre sujeitos [between-subjects]) (Fatorial)	≥ 2	≠
De n-fatores medidas repetidas	≥ 2	=
Mista [split-plot]	≥ 1	=
	≥ 1	≠

Nota. Entre parêntesis são indicadas as várias designações, incluindo as de língua inglesa.

Calculando o tamanho do efeito para a ANOVA

O TDE para a ANOVA pode não ser muito interessante de apresentar, pois está a testar-se somente a hipótese geral (A. Field et al., 2012) e, raramente, responde ao que se quer saber (Rosnow, Rosenthal e Rubin, 2000). Consoante o objetivo do estudo, pode ser mais relevante obter os TDE para as diferenças entre pares de grupos (A. Field et al., 2012).

Ainda assim, o TDE relativo à ANOVA deve mostrar quanto é que o fator de agrupamento explica a variabilidade da variável dependente Y, comparando com a variabilidade global observada de Y — a dispersão de todas as pontuações de Y. A equação (Ialongo, 2016; Keppel e Wickens, 2004) pode ser transcrita como:

$$TDE_{\text{variância}} = \frac{\text{Variabilidade}_{\text{explicada}}}{\text{Variabilidade}_{\text{total}}} = \frac{\text{Variabilidade}_{\text{total}} - \text{Variabilidade}_{\text{não-explicada}}}{\text{Variabilidade}_{\text{total}}}$$

Equação 7

TDE para ANOVA a um fator. Para medir a proporção da variância em Y associada aos diferentes grupos definidos por X pode usar-se o eta-quadrado (η^2) (Fritz et al., 2012; Ialongo, 2016; Keppel e Wickens, 2004), que corresponde à proporção da variância total

associada ao fator (revisão de Cohen, 1973, p. 107): Pearson (1905) designou-o por rácio de correlação:

$$\eta^2 = \frac{SQ_{\text{fator}}}{SQ_{\text{total}}} \quad \text{Equação 8}$$

Em que SQ_{fator} é a soma dos quadrados *entre* grupos para o fator (variância entre grupos) e SQ_{total} é a soma dos quadrados total (variância total).

O eta-quadrado é zero quando não há diferenças entre os grupos e aproxima-se de um quando as diferenças entre grupos ultrapassam a variabilidade entre grupos (Keppel e Wickens, 2004; Pearson, 1905). É preciso ter particular atenção ao facto de muitos artigos publicados que indicam ter computado o eta-quadrado, na verdade calcularam o eta-quadrado parcial (ver à frente). O erro deteta-se algumas vezes, pois o valor apresentado ultrapassa a unidade, o que só pode acontecer com a segunda estimativa (Levine e Hullett, 2002; Pierce, Block e Aguinis, 2004).

Como se pode verificar, a fórmula do eta-quadrado é similar à do coeficiente de determinação (R^2) (Ialongo, 2016), é, de facto, “literalmente uma generalização da correlação ponto-biserial” e é uma função da estimativa f (Cohen, 1992b, pp. 282-283). No entanto, diferentemente do R -quadrado que só mede relações lineares (a não ser que os preditores sejam elevados a expoentes), o eta-quadrado é sensível a vários tipos de relação, incluindo os não-lineares, daí a designação ser diferente (Vacha-Haase e Thompson, 2004).

Voltando à Tabela 1 da ANOVA no mesmo exemplo, o eta-quadrado fica da seguinte forma:

$$\eta^2 = \frac{18664,96}{84532,21} = 0,221$$

Este valor (22,1%), como se pode observar, é muito semelhante ao coeficiente de determinação ($R^2 = 0,244$).

Se se confrontar o valor do eta-quadrado com o psi de Steiger ($\psi = 0,97$), sobressai a grande diferença entre os dois. O que acontece é que o psi (RMSSE) de Steiger pode ser influenciado por um único grande desvio da média entre os grupos, pelo que o efeito *omnibus* deve ser considerado a título meramente indicativo (Ialongo, 2016).

Uma fórmula equivalente (Keppel e Wickens, 2004) baseia-se na estatística F e é muito útil para obter o TDE em resultados publicados que somente reportam esta estatística (raramente são publicadas as somas dos quadrados):

$$\eta^2 = \frac{(k-1)F}{(k-1)F + k(n-1)} \quad \text{Equação 9}$$

Nesta fórmula, k é o número de grupos e n é o número médio de sujeitos nos grupos.

Prosseguindo com a Tabela 1 da ANOVA no mesmo exemplo, o eta-quadrado resulta igual:

$$\eta^2 = \frac{(3-1) \times 106,12}{(3-1) \times 106,12 + 3(250,67-1)} = 0,2208$$

O eta-quadrado, deve notar-se, tende a inflacionar a variabilidade explicada e deve ser considerado como uma estimativa para uma amostra e não para a população e, por isso, foram sugeridas duas correções (Keppel e Wickens, 2004; Olejnik e Algina, 2000)¹².

Segundo vários autores (e.g., A. Field et al., 2012; Keppel e Wickens, 2004; Lakens, 2013; Levine e Hullett, 2002; Olejnik e Algina, 2000; 2003), o ómega-quadrado é uma medida corrigida do TDE (menos enviesada e menos inflacionada). A fórmula é de Hays (1963):

$$\omega^2 = \frac{SQ_{fator} - (k-1)QM_{erro}}{SQ_{total} + QM_{erro}} \quad \text{Equação 10}$$

ou

$$\omega^2 = \frac{(k-1)(QM_{fator} - QM_{erro})}{SQ_{total} + QM_{erro}} \quad \text{Equação 11}$$

Em que SQ_{fator} é a soma dos quadrados entre grupos para o fator (variância entre grupos), k é o número de grupos, SQ_{total} é a soma dos quadrados total (variância total) e QM_{erro} o quadrado médio de erro.

Deve evitar usar-se o ómega-quadrado quando os tamanhos das amostras não são iguais e são geralmente desproporcionais (Carroll e Nordholm, 1975; Vaughan e Corballis, 1969) e, simultaneamente, as variâncias são desiguais e as amostras são pequenas.

O ómega-quadrado pode também ser calculado quando se dispõe somente do parâmetro F (sendo vantajoso para calcular o TDE em resultados publicados noutros estudos):

$$\omega^2 = \frac{(k-1)(F-1)}{(k-1)(F-1) + kn} \quad \text{Equação 12}$$

Nesta fórmula (Keselman, 1975), k é o número de grupos, F o parâmetro da ANOVA, e n é o número médio de sujeitos nos grupos.

Assim, o ómega-quadrado para o nosso exemplo resulta um pouco menos inflacionado.

Através da Fórmula 10:

$$\omega^2 = \frac{18664,96 - (3-1) 87,94}{84532,21 + 87,94} = 0,2185$$

Através da Fórmula 11:

$$\omega^2 = \frac{(3-1) \times (106,12 - 1)}{(3-1) \times (106,12 - 1) + 3 \times 250,67} = 0,2185$$

Outra alternativa ao enviesamento positivo do eta-quadrado é o épsilon-quadrado (ϵ^2) (T. L. Kelley, 1935; Kirk, 1996; Olejnik e Algina, 2000; Snyder e Lawson, 1993). Esta é uma estimativa raramente usada, mas será ainda menos enviesada do que o ómega-quadrado (Okada, 2013). A fórmula que apresentamos é a de Glass e Hakstian (1969):

$$\epsilon^2 = \frac{SQ_{fator} - (k-1)QM_{erro}}{SQ_{total}} \quad \text{Equação 13}$$

Mais uma vez, SQ_{fator} é a soma dos quadrados entre grupos para o fator, k é o número de grupos, SQ_{total} é a soma dos quadrados total (variância total) e QM_{erro} o quadrado médio de erro.

No nosso exemplo, o épsilon-quadrado resulta num TDE ligeiramente superior:

$$\epsilon^2 = \frac{18664,96 - 2 \times 87,94}{84532,21} = 0,2187$$

As precauções indicadas para o ómega-quadrado aplicam-se também ao épsilon-quadrado (Carroll e Nordholm, 1975; Vaughan e Corballis, 1969).

Como também se pode constatar pelos exemplos, o épsilon-quadrado tende a ser um pouco maior do que o ómega-quadrado e ambos são menores do que o eta-quadrado (Ferguson, 2009; Jalongo, 2016; Olejnik e Algina, 2000; Richardson, 2011). O ómega-quadrado e o épsilon-quadrado podem assumir valores negativos, no entanto, habitualmente o que se apresenta é o número inteiro não positivo zero.

Estas estimativas não-enviesadas ou corrigidas são geralmente menores do que as suas contrapartes não corrigidas e estão mais próximas dos TDE populacionais (Snyder e Lawson, 1993). A diferença entre medidas corrigidas e não corrigidas é referida como *contração*

¹² Essa inflação decorre do princípio da maximização matemática usado no cálculo dos modelos lineares gerais que explica que qualquer variação nos dados — devida aos efeitos naturais na população ou peculiaridades da amostra — será considerada quando se estimarem os efeitos (P. D. Ellis, 2010). Cada amostra é única e a sua unicidade impede a replicação. A unicidade das amostras ou erro de amostragem

correlaciona-se positivamente com o número de variáveis independentes e negativamente com o tamanho da amostra (Thompson, 2002). De onde resulta que a inflação atribuível ao erro de amostragem é maior quando o tamanho e os efeitos das amostras são pequenos e quando o número de variáveis no modelo é alto (Vacha-Haase e Thompson, 2004).

(Vacha-Haase e Thompson, 2004). A *contração* tende a diminuir à medida que o tamanho das amostras aumenta e o número de preditores no modelo cai. No entanto, a *contração* tende a ser muito pequena se os efeitos forem grandes, independentemente do tamanho da amostra (e.g., os R^2 maiores tendem a convergir para as suas contrapartes ajustadas) (P. D. Ellis, 2010).

TDE para ANOVAs a n-fatores. Referiu-se atrás que os TDEs tendem a diminuir em modelos com mais do que um fator $X (X_1, X_2, \dots X_i)$, reduzindo-se tanto mais quanto mais fatores forem incluídos (Ferguson, 2009). Presentemente vamos apenas considerar os modelos a dois fatores, mas as equações podem generalizar-se para integrar mais fatores.

Assim, para planeamentos com mais do que um fator, o TDE a usar é o eta-quadrado parcial (Ialongo, 2016; Keppel e Wickens, 2004; Olejnik e Algina, 2000; Richardson, 2011; Tabachnick e Fidell, 2007):

$$\eta_p^2 = \frac{SQ_{fator}}{SQ_{fator} + SQ_{erro}} \quad \text{Equação 14}$$

Nesta fórmula, se considerarmos somente dois fatores, SQ_{fator} corresponde à soma dos quadrados entre grupos para o fator X_1 ou X_2 ou para a interação de X_1 com X_2 ; SQ_{total} à soma dos quadrados total e SQ_{erro} à soma dos quadrados para o erro.

Se um fator for significativo num dos níveis do outro fator, usualmente reporta-se também o TDE para a interação. Assim, devem indicar-se os etas-quadrados parciais para os efeitos simples e quaisquer contrastes derivados (Keppel e Wickens, 2004).

Retomando o exemplo agora com os fatores idade e sexo (Tabela 5), podemos verificar essa redução, ficando o eta-quadrado parcial para a idade em:

$$\eta_p^2 = \frac{15974,80}{15974,80 + 65046,004} = 0,197$$

Como se pode verificar pela Tabela 4, o SPSS-25 fornece esse mesmo valor. Quanto ao sexo acontece o mesmo:

$$\eta_p^2 = \frac{350,077}{350,077 + 65046,004} = 0,0054$$

Como se pode verificar pela Tabela 4, o SPSS-25 fornece também o eta-quadrado parcial para a interação idade com o sexo. Apesar de o fator idade não ser

significativo nos níveis do fator sexo, apresenta-se o cálculo para ilustrar o processo.

$$\eta_p^2 = \frac{315,85}{315,85 + 65046,004} = 0,0048$$

Note-se que a Equação 14 é genérica e a precisão dos seus termos depende do planeamento (Olejnik e Algina, 2003). Refira-se ainda que o eta-quadrado e o eta-quadrado parcial coincidem na ANOVA com um fator (Levine e Hullett, 2002; Pierce et al., 2004), especialmente com grandes amostras (Ferguson, 2009). O eta-quadrado parcial pode variar entre 0 e acima de 1, por isso não deve ser interpretado como percentagem da variância (Levine e Hullett, 2002; Pierce et al., 2004; Richardson, 2011). Finalmente, há que ter em consideração que o eta-quadrado varia consoante a natureza e o número de variáveis que se juntam ao modelo, pelo que pode não ser comparável entre estudos diferentes (Kennedy, 1970). Mais, nem sequer deve ser usado para comparar os efeitos dos diferentes fatores na mesma análise, uma vez que os denominadores são diferentes (Olejnik e Algina, 2000; Sechrest e Yeaton, 2016).

Uma vez que o eta-quadrado parcial só pode ser usado para comparar efeitos entre estudos com planeamentos experimentais semelhantes, Olejnik e Algina (2003) propuseram um eta-quadrado geral (η_G^2) que retira a variação dos outros fatores do cálculo do TDE, mas inclui a variância devida a diferenças individuais. Este eta-quadrado geral permite a comparação entre estudos em que os fatores não foram manipulados¹³ (nos estudos com fatores manipulados η_G^2 e η_p^2 são idênticos). A fórmula é a seguinte:

$$\eta_G^2 = \frac{SQ_{fatorX_1}}{SQ_{Total} - SQ_{fatorX_2}} \quad \text{Equação 15}$$

Nesta fórmula, se consideramos somente dois fatores, SQ_{fatorX_1} corresponde à soma dos quadrados entre grupos para o fator X_1 , SQ_{total} à soma dos quadrados total e SQ_{fatorX_2} à soma dos quadrados para o fator X_2 .

Para a inclusão de mais fatores veja-se a revisão de Fritz et al. (2012).

Assim, no nosso exemplo seria:

$$\eta_G^2 = \frac{15974,80}{1546218,00 - 350,08} = 0,010$$

que irá obter um determinado efeito (e.g., efeito de uma determinada intervenção terapêutica sobre o funcionamento cognitivo). Tipicamente, as variáveis são manipuladas em estudos experimentais ou quase-experimentais.

¹³ Os fatores não-manipulados são as variáveis que o investigador não controla (e.g., pertença ou não a grupo de fumadores para determinar o efeito do fumar na saúde pulmonar). Os fatores não-manipulados são típicos dos estudos não experimentais. Fatores ou variáveis manipuladas são aqueles que o investigador controla, estando relativamente seguro

Tal como no anterior, o eta-quadrado parcial pode ser corrigido através do ómega ou do épsilon quadrados (sendo indicados para amostras pequenas), cujas equações são, respetivamente, as seguintes (Bakeman, 2005; Olejnik e Algina, 2000):

$$\omega_p^2 = \frac{(k-1)(QM_{fator} - QM_{erro})}{(k-1)QM_{fator} + [N - (k-1)]QM_{erro}} \quad \text{Equação 16}$$

$$\varepsilon_p^2 = \frac{gl_{fator}(SQ_{fator} - SQ_{erro})}{SQ_{fator} + SQ_{erro}} \quad \text{Equação 17}$$

SQ_{fator} é a soma dos quadrados entre grupos para o fator, k é o número de grupos, gl são os graus de liberdade, QM_{fator} é o quadrado médio do fator e QM_{erro} o quadrado médio de erro.

O cálculo do ómega e do épsilon quadrados pode tornar-se complexo em planeamentos com muitos fatores e níveis, pelo que, dada a pequena diferença, se recomenda o uso do eta-quadrado (Lakens, 2013).

O f^2 de Cohen, cuja fórmula se descreve a seguir, é também apropriado para a ANOVA com n -fatores (Cohen, 1992b; Jalongo, 2016):

$$f^2 = \frac{\eta_p^2}{1 - \eta_p^2} \quad \text{Equação 18}$$

No exemplo com os fatores idade e sexo (Tabela 4), o f^2 de Cohen para a idade foi de:

$$f^2 = \frac{0,0048}{1 - 0,0048} = 0,0048$$

ANOVAs: Outros tipos. Os modelos de medidas repetidas e outros mais complexos não serão descritos neste artigo, mas remetemos o leitor interessado para bibliografia a seguir apontada. Assim, para a ANOVA de medidas repetidas, ANOVA mista (usa-se o η_p^2) e análise de variância multivariada (MANOVA: usa-se o η^2 ou o η_p^2) recomendamos o livro de Keppel e Wickens (2004) e o artigo de Steyn e Ellis (2009).

Para a análise de covariância (ANCOVA; usa-se o η^2) e de análise de covariância multivariada (MANCOVA; usa-se o f^2 de Cohen), sugerem-se os livros de Keppel e Wickens (2004) e de Tabachnick e Fidell (Tabachnick e Fidell, 2007). Existem também fórmulas de correção para a análise multivariada (Vacha-Haase e Thompson, 2004).

INTERPRETAÇÃO DA MAGNITUDE DOS TDE

Quando se reporta um TDE, o objetivo é permitir ao leitor a interpretação da importância dos resultados. Mantidas inalteradas todas as outras coisas, quanto mais alto for um TDE, maior o impacto da variável

independente ou fator na variável dependente (Fritz et al., 2012). O estabelecimento de diretrizes para qualificar a magnitude de um efeito em pequeno, médio e grande foi marcado essencialmente por Cohen (1988).

No entanto, os TDEs foram definidos com base no tipo de fenómenos que Cohen observou na sua área (psicologia), de modo que eles dificilmente podem ser traduzíveis em domínios fora das ciências comportamentais.

Outro problema, como refere o próprio Cohen (1988), é que muita da variância na variável dependente nas ciências comportamentais não é facilmente explicável. Quando se considera um TDE como grande nas ciências comportamentais não é a mesma coisa que, por exemplo, na física em que cerca de 99% da variância é explicável. Assim, nas ciências comportamentais, os valores têm de ser entendidos de forma relativa e não absoluta, reconhecendo que, frequentemente, são muito pequenos (Cohen, 1988).

Na verdade, não se tem conseguido fornecer uma escala universal e os valores de referência atuais são apenas uma herança derivada da forma como se iniciou o estudo dos TDE.

Ainda assim, observando as cautelas devidas apresentadas na secção seguinte, a Tabela 7 pode ser usada como guia de orientação na interpretação dos TDE, especialmente quando o estudo for inovador e não houver um contexto de efeitos anteriores.

NOTAS DE CAUTELA

O contexto dos tamanhos do efeito

Os TDE são estimativas importantes e são estatísticas descritivas úteis, mas, tal como referimos anteriormente, é preciso não cair no erro que se tem caído com o valor do p — tomar os TDE como se fossem entidades divinas que ditam dogmaticamente o destino do valor de um estudo. Portanto, as indicações relativamente às diretrizes sumarizados neste artigo são arbitrários, pelo que a adesão rígida a essas indicações não é aconselhada, devendo ser usados meramente como guias de orientação (A. Field, 2005). Assim, na interpretação dos TDEs, é importante compará-los com os reportados na literatura (A. Field, 2005; Fritz et al., 2012; Lakens, 2013) e, na medida do possível, convertê-los em tamanho do efeito em linguagem comum (Fritz et al., 2012; Lakens, 2013).

Acrescente-se que a *significância substantiva* ou importância de um TDE depende em grande parte da área em que o estudo é efetuado (Rosnow e Rosenthal, 1989). Assim, a mesma magnitude do TDE num estudo sobre o efeito de um tratamento com consequências potencialmente mortais será de grande significado

prático, mesmo que o valor seja pequeno (Ferguson, 2009). Se, por exemplo, num estudo for obtido uma razão de chances (OR) de 2,5, que é “médio” segundo as indicações de Wilson (2010), mas no contexto da área a média dos OR for de 1,3, então o valor não deve ser considerado médio (Ialongo, 2016).

Desta forma, um ponto chave a reter é que os TDE devem ser interpretados no contexto ou área da investigação. Quando um processo de decisão se baseia em critérios relativamente arbitrários e em comparações com outros estudos da mesma área, ele passa a incluir um elemento de subjetividade. Inevitavelmente, o processo de decisão passa então a envolver o sistema de valores,

preocupações sociais, preocupações com custos e benefícios por parte do investigador (Snyder e Lawson, 1993). No entanto, como Kirk (1996) aponta, é uma obrigação do investigador fazer esse tipo de julgamento.

Intervalos de confiança

Os TDE devem ser sempre acompanhados dos respetivos intervalos de confiança com o argumento de que é fornecido uma margem de incerteza associada à magnitude do efeito. É, assim, facultado um conjunto de valores plausíveis para a estimativa em análise, com um risco tradicionalmente fixado em 5% de que o verdadeiro valor seja excluído (P. D. Ellis, 2010).

Tabela 7

Interpretação dos Tamanhos do Efeito (TDE)

Referência	TDE Estatística	Valores	Interpretação	Comandos no SPSS
Steiger e Fouladi, 1997/2016	RMSSE ou f^* ANOVA omnibus	0,10	Pequeno	Analisar, Compare Means, Comparar Médias, Análise de variância unidirecional
		0,25	Médio	
		0,40	Grande	
Cohen, 1988	r^2 Regressão bivariada	0,0196	Pequeno	Analisar, Regressão, Linear
		0,1300	Médio	
		0,2600	Grande	
McGrath e Meyer, 2006	r Correlação ponto biserial Regressão linear para contrastes teóricos	0,10	Pequeno	Analisar, Regressão, Linear
		0,24	Médio	
		0,37	Grande	
Cohen, 1988	R^2 Regressão múltipla	0,0196	Pequeno	Analisar, Regressão, Linear
		0,1300	Médio	
		0,2600	Grande	
Cohen, 1988	f^2 de Cohen Regressão múltipla	0,02	Pequeno	Analisar, Regressão, Linear
		0,15	Médio	
		0,35	Grande	
Ferguson, 2009	R^2_{ajust} Regressão múltipla	0,04**	Pequeno	Analisar, Regressão, Linear
		0,25	Médio	
		0,64	Grande	
Wilson, 2010	Razão de chances (Logit) Regressão logística	1,50	Pequeno	Analisar, Regressão, Logística binária Analisar, Regressão, Logística binária Antilog do logit exponenciando o coeficiente
		2,50	Médio	
		4,30	Grande	
Cohen, 1988	η^2 ANOVA	0,02	Pequeno	Analisar, Modelo Linear Geral, Univariado, Opções, Estimativas de tamanho de efeito
		0,13	Médio	
		0,26	Grande	
Cohen 1988	ω^2 Alternativa corrigida ANOVA	0,01	Pequeno	Analisar, Modelo Linear Geral, Univariado
		0,06	Médio	
		0,14	Grande	
Ferguson, 2009	ϵ^2 Alternativa corrigida ANOVA	0,04**	Pequeno	Analisar, Modelo Linear Geral, Univariado
		0,25	Médio	
		0,64	Grande	
Cohen, 1988	η^2_p ANOVA n-fatores	0,0099***	Pequeno	Analisar, Modelo Linear Geral, Univariado
		0,0588	Médio	
		0,1379	Grande	

Nota. Os valores de corte para a consideração dos tamanhos do efeito devem ser considerados de forma não-dogmática, somente como linhas de orientação, e sempre no contexto dos tamanhos do efeito reportados ou calculados noutras investigações. A conversão em tamanhos do efeito em linguagem comum (ver Espírito-Santo e Daniel, 2015) para o r , R^2 , e eta-quadrado podem ser consultadas em Fritz, Morris e Richler (2012).

* Estes valores não estão ancorados ao r pelo que devem ser interpretados cautelosamente.

** Estes valores devem ser usados com cautela, pois não encontramos outro suporte e Ferguson (2009) atribui os mesmos valores para diferentes estimativas.

*** Haase, Waechter e Solomon (Haase, Waechter e Solomon, 1982) calcularam os valores do eta-parcial em 11.044 estatísticas reportadas no *Journal of Counseling Psychology* entre 1970 e 1979 e encontraram um valor mediano de 0,0830 que é um pouco maior que o de Cohen.

Fontes: Ferguson (2009), Keppel e Wickens (2004), Kirk (1996), Lipsey e Wilson (2001), Thompson (2007), Wilson (2010) .

A maior parte dos intervalos de confiança para os TDE tem de ser computada com *software* apropriado (e.g., Smithson, 2003; Steiger, 2004) e, talvez por isso, raramente são indicados (Fritz et al., 2012).

Não cabendo nesta revisão o seu cálculo e exemplificação, existem vários artigos que recomendamos ao leitor interessado (e.g., K. Kelley, 2007; Keselman et al., 2008; Bird:2016cl Thompson, 2007). Adicionalmente, para os intervalos de confiança associados à ANOVA e à regressão, Smithson (2003) fornece instruções e sintaxes para o SPSS, SAS, SPlus e R. Outros detalhes são proporcionados por Steiger (2004).

Fatores de enviesamento

Apesar do valor e utilidade, os TDEs refletem as propriedades dos dados e as condições em que os dados foram recolhidos (Fritz et al., 2012). Assim, estas estimativas podem ser inflacionadas ou deflacionadas pela qualidade da medição (Ferguson, 2009). Um TDE obtido com um questionário fracamente validado, provavelmente no “mundo real” não terá significado prático. A falha em conter as ameaças à validade interna (e.g., não-controlo do efeito placebo; não-controlo do efeito de aprendizagem nos testes), a fraca fidedignidade e validade das respostas aos instrumentos e a heterogeneidade da população, ao invés, podem reduzir a magnitude dos TDE (Ferguson, 2009; Fritz et al., 2012; Olejnik e Algina, 2000).

Outro aspeto a considerar são as estratégias de amostragem (Ferguson, 2009). Por exemplo, se a amostra incluir somente sujeitos de um grupo restrito (e.g., meio urbano), a variabilidade de erro é menor (inflacionando o TDE) do que numa amostra que inclua sujeitos de grupos diversos (e.g., meios rural, semiurbano e urbano) (Fritz et al., 2012).

Amostras pequenas e não aleatórias são outro detalhe que deve remeter para a prudência na interpretação dos TDE (Ferguson, 2009).

Depois há que considerar o planeamento do estudo. Quando se interpreta um TDE deve ter-se em consideração a falha em controlar outros preditores relevantes, pois podem amplificar os TDE, especialmente se estiverem correlacionados com as variáveis independentes (Olejnik e Algina, 2003). Por exemplo, se um estudo considerar a variabilidade associada ao grupo económico e incluir como covariável o salário, então vai reduzir a variabilidade de erro e o TDE tende a aumentar (Fritz et al., 2012). No entanto, a análise através do eta-quadrado geral poderá controlar o enviesamento (Olejnik e Algina, 2003).

Outro aspeto a considerar no planeamento é que em modelos com fatores fixos, os TDE para os efeitos *omnibus* dependem dos níveis específicos. Logo, os TDE não são comparáveis para diferentes níveis do fator (Olejnik e Algina, 2000). A amplitude dos níveis de um fator definidos de forma concentrada (e.g., 20, 25, 30, 35 anos de idade) tende a reduzir o TDE, e, inversamente, um maior espaçamento entre os níveis (e.g., 20, 50, 80 anos de idade) tende a ampliar o TDE (Olejnik e Algina, 2000).

A própria análise estatística pode ter impacto sobre o valor de um TDE. Assim, o uso de métodos estatísticos que reduzam a variância (e.g., usar falsas dicotomias ou categorizações de variáveis com base em pontos de corte) podem também reduzir os TDE (Ferguson, 2009).

RECOMENDAÇÕES

Como reportar os tamanhos do efeito

Antes de se reportarem os TDEs, primeiramente, descrevam-se os dados, incluindo as médias ou outras medidas apropriadas de tendência central, pelo menos uma medida associada de variabilidade (e.g., desvio-padrão, amplitude, amplitude interquartílica) e o quadrado médio de erro para a ANOVA (Fritz et al., 2012).

Depois, apresentem-se as estimativas dos efeitos, indicando explicitamente que TDEs estão a ser reportados, os seus valores, incluindo os intervalos de confiança para uso potencial em estudos de metanálise (Fritz et al., 2012; Vacha-Haase e Thompson, 2004; Wilkinson e Task Force on Statistical Inference, 1999).

Selecione-se o eta-quadrado apropriado entre os disponíveis; se os efeitos ou os *Ns* forem pequenos, indique-se a possível inflação do eta-quadrado, apresentando o ómega-quadrado adequado ao modelo (Fritz et al., 2012). Neste respeito, são preferidos os TDE corrigidos aos não-corrigidos, pois tendem a ser estimativas mais precisas para a população alvo (Ferguson, 2009). No entanto, Vacha-Haase e Thompson (2004) tendem para as estimativas não corrigidas quando as amostras são grandes e o número de variáveis pequeno. Adicionalmente, Roberts e Henson (2002) consideram que têm havido excesso de “correção” nas estimativas, pelo o mais correto será que se reportem ambas (P. D. Ellis, 2010).

Por vezes, vale a pena reportar mais do que um TDE para interpretar convenientemente os dados. Usando o exemplo dado por Fritz et al. (2012), pode ser adequado apresentar o eta-quadrado parcial para revelar a proporção da variabilidade relativa a um fator quando se controlam os restantes fatores, e também reportar o eta-

quadrado geral para espelhar a contribuição do fator para o desempenho global.

Em análises mais complexas (ANOVA de n-fatores ou regressão múltipla), reportem-se todos os efeitos: a estimativa F , os graus de liberdade e o quadrado médio de erro, para que o leitor possa computar outros TDE de interesse (Fritz et al., 2012). É também útil reportar o TDE para o efeito global e para cada um dos contrastes planejados. Tal como dissemos noutra secção, os TDE para testes planejados (determinadas comparações par-a-par) podem até ser mais interessantes de reportar que os TDEs do efeito global. Por vezes, mais do que “interessantes”, uma única comparação entre dois níveis é a responsável pelo efeito *omnibus*. Nesse sentido, o melhor TDE a reportar para as comparações par-a-par será a estatística g de Hedges (Ferguson, 2009). As considerações sobre o g de Hedges podem ser consultadas no nosso artigo de 2015 (Espírito-Santo e Daniel, 2015).

Depois, indique-se a interpretação dos TDE, tendo em consideração os seus pressupostos (e.g., normalidade das distribuições ou homogeneidade de variâncias) e as suas limitações. O manual da *American Psychological Association* (2010) sugere que se discutam os TDEs quer nas suas unidades originais — por ser mais claro e fácil de discutir no contexto de um estudo único —, quer em termos padronizados — por facilitar a comparação de estudos e as metanálises.

Compare-se, de seguida, os TDEs com os obtidos noutros estudos, não deixando de contemplar as diferenças nos planeamentos (Fritz et al., 2012; Vacha-Haase e Thompson, 2004). Tenha-se em atenção que as generalizações devem ser limitadas aos delineamentos que envolvam os mesmos níveis da variável independente e números e tipos semelhantes de sujeitos (Snyder e Lawson, 1993). Em delineamentos com fatores fixos será mais preciso indicar quais k níveis da variável independente explicaram determinada percentagem da variância na variável dependente quando n sujeitos do tipo p foram designados para cada nível (Snyder e Lawson, 1993).

Finalize-se, discutindo as implicações práticas, clínicas ou teóricas dos TDE mais relevantes (Fritz et al., 2012; Vacha-Haase e Thompson, 2004).

CÁLCULO DOS TDE

Muitos dos programas informáticos de estatística não calculam os TDE. Por isso, este artigo acompanha-se por uma folha Excel para as equações relevantes indicadas para os vários testes estatísticos.

Existem também online várias calculadoras compreensivas (e.g., (W. Lenhard e Lenhard, 2016; Lyons e Morris, 2018; Wilson, 2018).

As fórmulas apresentadas ao longo do texto podem também, obviamente, ser usadas para o cálculo dos vários TDE.

CONCLUSÕES

Neste artigo, mais uma vez, procurámos defender a importância de apresentar as magnitudes do efeito. Abordámos, assim, os tamanhos do efeito relativos às proporções de variância, referindo questões metodológicas, teóricas e substantivas quanto à interpretação dos mesmos.

Para determinar a significância ou importância substantivas, não é suficiente apresentar a significância estatística, nem a dimensão relativa de um tamanho de efeito.

Todos os tamanhos de efeito devem ser interpretados no contexto específico da investigação, do nível de desenvolvimento teórico na área da investigação e das aplicações práticas no mundo real.

Pode ser complexo e intrincado comparar os tamanhos de efeito entre estudos devido ao número de fatores previstos e não-previstos e suas possíveis combinações. Apesar disso, mantém-se a recomendação de se indicar os tamanhos do efeito a par com o nível de significância estatística.

Desta forma, confiamos mais uma vez que os autores da RPICS usem as diretrizes e sugestões fornecidas neste artigo para decidir pelos tamanhos de efeito adequados e interpretação adequada nas análises de variância.

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

Fontes de financiamento | Funding sources: nenhuma | none.

REFERÊNCIAS

- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6^a ed.). Washington, DC: American Psychological Association.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379–384.
- 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

- Bewick, V., Cheek, L. e Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care*, 7(6), 451–459. doi:10.1186/cc2401
- 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.
- Carroll, R. M. e Nordholm, L. A. (1975). Sampling characteristics of Kelley's ϵ^2 and Hays' ω^2 . *Educational and Psychological Measurement*, 35(3), 541–554. doi:10.1177/001316447503500304
- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33, 107–112. doi:10.1177/001316447303300111
- Cohen, J. (1992a). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cohen, J. (1992b). Statistical power analysis. *Current Directions in Psychological Science*, 1(3), 98–101.
- Cumming, G. (2012). *Understanding the new statistics*. New York: Routledge.
- Ellis, P. D. (2010). *The essential guide to effect sizes*. Cambridge: Cambridge University Press.
- Erceg-Hurn, D. M. e Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *The American Psychologist*, 63(7), 591–601. doi:10.1037/0003-066X.63.7.591
- 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. doi: 10.7342/ismt.rpics.2015.1.1.14
- Espirito-Santo, H. e Daniel, F. (2017). Calcular e apresentar tamanhos do efeito em trabalhos científicos (2): Guia para reportar a força das relações. *Revista Portuguesa de Investigação Comportamental e Social*, 3(1), 53–64. doi:10.7342/ismt.rpics.2017.3.1.48
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538.
- Field, A. (2005, May 5). *Effect sizes*. Recuperado de <http://www.statisticshell.com/docs/effectsizes.pdf>
- Field, A., Miles, J. e Field, Z. (2012). *Discovering statistics using R*. Londres: Sage.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.
- Fritz, C. O., Morris, P. E. e Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. doi:10.1037/a0024338
- Glass, G. V. e Hakstian, A. R. (1969). Measures of association in comparative experiments: Their development and interpretation. *American Educational Research Journal*, 6(3), 403–414. doi:10.3102/00028312006003403
- Haase, R. F., Waechter, D. M. e Solomon, G. S. (1982). How significant is a significant difference? Average effect size of research in counseling psychology. *Journal of Counseling Psychology*, 29(1), 58–65. doi:10.1037//0022-0167.29.1.58
- Hair, J., Black, B., Babin, B. e Anderson, R. (2009). *Multivariate data analysis* (7ª ed.). Upper Saddle River: Pearson Higher Ed.
- Hays, W. L. (1963). *Statistics for psychologists*. New York: Holt, Rinehart and Winston.
- 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
- Herzberg, P. A. (1969). *The parameters of cross-validation*. Richmond, VA: William Byrd Press.
- Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica*, 26(2), 150–163. doi:10.11613/BM.2016.015
- Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software*, 20(8), 1–24. doi:10.18637/jss.v020.i08
- Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences of the United States of America*, 21(9), 554–559.
- Kennedy, J. J. (1970). The eta coefficient in complex ANOVA designs. *Educational and Psychological Measurement*, 30(4), 885–889. doi:10.1177/001316447003000409
- Keppel, G. e Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4ª ed.). New Jersey: Pearson
- Keren, G. e Lewis, C. (1969). Partial omega squared for Anova designs. *Educational and Psychological Measurement*, 39(1), 119–128. doi:10.1177/001316447903900116
- Keselman, H. J. (1975). A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Canadian Psychological Review/Psychologie Canadienne*, 16(1), 44–48. doi:10.1037/h0081789
- Keselman, H. J., Algina, J., Lix, L. M., Wilcox, R. R. e Deering, K. N. (2008). A generally robust approach for testing hypotheses and setting confidence intervals for effect sizes. *Psychological Methods*, 13(2), 110–129. doi:10.1037/1082-989X.13.2.110
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological*

- Measurement*, 56(5), 746–759.
doi:10.1177/0013164496056005002
- Kline, R. B. (2013). *Beyond significance testing: Reforming data analysis methods in behavioral research* (2ª ed.). Washington, DC: American Psychological Association.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4(863), 1–12. doi:10.3389/fpsyg.2013.00863
- Lenhard, W. e Lenhard, A. (2016). Calculation of effect sizes. doi:10.13140/RG.2.1.3478.4245. Recuperado de https://www.psychometrica.de/effect_size.html.
- Levine, T. R. e Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research*, 28(4), 612–625.
- Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinkamp, K., ... Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research, Institute of Education Sciences.
- Lipsey, M. W. e Wilson, D. B. (2001). *Practical meta-analysis*. New York: Sage Publications, Inc.
- Lyons, L. C. e Morris, W. A. (2018). *The meta analysis calculator*. Recuperado de <http://www.lyonsmorris.com/ma1/>
- Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way anova. *Behaviormetrika*, 40(2), 129–147. doi:10.2333/bhmk.40.129
- 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
- Olejnik, S. e Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434–447. doi:10.1037/1082-989X.8.4.434
- Pallant, J. (2011). *SPSS: Survival manual* (4ª ed.). Crows Nest: Allen & Unwin.
- Pampel, F. C. (2000). *Logistic regression*. Thousand Oaks: SAGE Publications.
- Pearson, K. (1905). *Mathematical contributions to the theory of evolution. XIV. On the general theory of skew correlation and non-linear regression*. Londres: Dulau.
- Pierce, C. A., Block, R. A. e Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, 64(6), 916–924.
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135–147.
- Roberts, J. K. e Henson, R. K. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement*, 62(2), 241–253. doi:10.1177/0013164402062002003
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Revised). Newbury Park: Sage.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5(3), 127–134. doi:10.1111/j.1467-9280.1994.tb00646.x
- Rosnow, R. L. e Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *The American Psychologist*, 44(10), 1276–1284.
- Rosnow, R. L., Rosenthal, R. e Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science*, 11(6), 446–453.
- Sechrest, L. e Yeaton, W. H. (2016). Magnitudes of experimental effects in social science research. *Evaluation Review*, 6(5), 579–600. doi:10.1177/0193841X8200600501
- Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage.
- Snyder, P. e Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *The Journal of Experimental Education*, 61(4), 334–349.
- Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9(2), 164–182.
- Steiger, J. H. e Fouladi, R. T. (2016). Noncentrality interval estimation and the evaluation of statistical models. Em L. L. Harlow, S. A. Mulaik e J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 197–229). Routledge.
- Stevens, J. P. (2007). *Intermediate statistics* (3ª ed.). New York: Lawrence Erlbaum Associates.
- Steyn, H. S., Jr. e Ellis, S. M. (2009). Estimating an effect size in one-way multivariate analysis of variance (MANOVA). *Multivariate Behavioral Research*, 44(1), 106–129. doi:10.1080/00273170802620238
- Tabachnick, B. G. e Fidell, L. S. (2007). *Using multivariate statistics* (5ª ed.). Boston: Pearson Education.
- Thompson, B. (2002). “Statistical,” “practical,” and “clinical”: How many kinds of significance do counselors need to consider? *Journal of Counseling & Development*, 80(1), 64–71. doi:10.1002/j.1556-6678.2002.tb00167.x
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44(5), 423–432. doi:10.1002/pits.20234

- Vacha-Haase, T. e Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, 51(4), 473–481. doi:10.1037/0022-0167.51.4.473
- Vaughan, G. M. e Corballis, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, 72(3), 204–213. doi:10.1037/h0027878
- Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *The Annals of Mathematical Statistics*, 2(4), 440–457. doi:10.1214/aoms/1177732951
- Wilkinson, L. e Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, 54(8), 594–604.
- Wilson, D. B. (2010). *Meta-analysis stuff*. Recuperado de <http://mason.gmu.edu/~dwilsonb/ma.html>
- Wilson, D. B. (2018). *Practical meta-analysis effect size calculator*. Recuperado de <http://cebcp.org/practical-meta-analysis-effect-size-calculator/>
- Zhang, G. e Algina, J. (2011). A robust Root Mean Square Standardized Effect Size in one-way fixed-effects ANOVA. *Journal of Modern Applied Statistical Methods*, 10(1), 77–96. doi:10.22237/jmasm/1304222880