

# Algumas medidas típicas univariadas da magnitude do efeito (\*)

JOSEPH EDWARD CONBOY (\*\*)

As crescentes críticas ao uso de estatística inferencial para a análise de dados nos estudos em psicologia e outras ciências sociais, provocaram, ao longo do último quarto de século, um movimento para o desenvolvimento, e maior divulgação, de métodos alternativos (Conboy, 2002; Glass, 2000; Henkel, 1976; Hunt, 1997; Lipsey & Wilson, 2001). Uma das respostas às críticas foi o desenvolvimento (e a redescoberta) de medidas de *effect size*, ou seja, a magnitude do efeito. O uso destas métricas representa uma tentativa de substituir o conceito de significância estatística com noções mais úteis de significância prática específica (Kirk, 1996), por exemplo a significância clínica (Jacobson & Truax, 1991; Kendall, 1999) e a significância educacional. (Tallmadge, 1977). O uso da magnitude do efeito

como método de análise é cada vez mais aceite por investigadores de diversas disciplinas, quer nas ciências sociais, quer noutras áreas (nomeadamente, na medicina). A tendência é crescente: Por exemplo, em 1994, a American Psychological Association (APA, 1994) «encorajou» os autores de estudos de investigação a relatar medidas de magnitude do efeito. Cinco anos mais tarde, num reforço a este incitamento, Wilkinson & APA Task Force on Statistical Inference (1999) afirmaram que o relato de medidas de magnitude do efeito, e a sua interpretação face a efeitos previamente observados, eram «essenciais para a boa investigação». Ainda mais recentemente, na quinta edição do seu manual de publicação, a American Psychological Association (APA, 2001) descreve o relato de magnitudes de efeito como sendo «quase sempre necessário» (p. 25). Não é apenas a APA que se interessa pelo tópico: Outras associações profissionais de grande influência, preocupadas com a qualidade da investigação, também têm discutido os méritos dos métodos baseados na magnitude do efeito (Fidler, 2002; Thompson, 2002b).

Embora o relato de magnitudes de efeito já tenha sido aceite e esteja a ser considerado obrigatório por parte de um conjunto crescente de revistas da especialidade (Huberty, 2002; Roberts & Henson, 2002; Thompson, 2002a), é um facto que um grande número de investigadores e psicólogos não têm nem formação, nem experiência

---

(\*) Agradecimentos: Gostava de agradecer à Professora Jesuína M. B. Fonseca, da Universidade do Algarve, e A. K. Conboy da Universidade de Bowdoin que comentaram uma versão anterior do trabalho. Gostava igualmente de mencionar o meu reconhecimento aos responsáveis das Bibliotecas Guttman da Harvard Graduate School of Education e Milbank da Columbia University que possibilitaram o acesso a algumas das referências incluídas no artigo.

(\*\*) Instituto Superior Dom Afonso III, Loulé, Portugal.

directa, com essas métricas. Ao longo do século XX, os testes de hipótese têm constituído a ferramenta principal de investigação quantitativa nos campos da psicologia, educação e outras ciências sociais. A formação de investigadores foi, naturalmente, focada nesses métodos. É pois urgente proceder-se a modificação dos currículos nessa temática, sendo que a observação casual confirma que os conteúdos programáticos de disciplinas de estatística e investigação continuam a dar ênfase aos testes de hipótese e valores de  $p$  em detrimento de medidas de magnitude do efeito. Para além dessa falta de formação e prática específicas, existe igualmente o problema de muitos trabalhos sobre a magnitude do efeito terem como propósito a descrição de análises secundárias para efeitos de eventual meta-análise, ou seja, os referidos trabalhos não tratam da apresentação do problema de cálculo de valores de magnitude do efeito relativamente aos dados primários obtidos pelo investigador.

O propósito deste artigo é, então, o de apresentar aos leitores interessados algumas das formas mais comuns de medidas da magnitude do efeito e descrever a interpretação prática de cada uma. Tal conhecimento é necessário quer para o investigador que pretende relatar os seus próprios resultados, quer para o estudante cuidadoso que pretende compreender melhor os trabalhos publicados por outros investigadores.

A lógica subjacente ao uso de medidas de magnitudes de efeito, bem como as suas vantagens relativamente aos testes de hipótese, encontram-se descritas noutras publicações (Conboy, 2002; Glass, McGaw, & Smith, 1981; Henkel, 1976; Hunt, 1997). Sugere-se que o leitor recorra a tais fontes para informação mais detalhada, mas, de maneira resumida, podemos aqui referir o problema principal que existe com o uso de testes de hipótese: A natureza da informação que estes testes fornecem é sempre negativa e sempre do tipo «tudo ou nada». Nas ciências sociais muitos fenómenos existem no âmbito de um contínuo de valores possíveis. Os testes de significância, contudo, apenas permitem resultados dicotómicos: O investigador tem a opção de aceitar a independência presumível dos factores estudados, ou de rejeitar a probabilidade da independência; de aceitar que não existem diferenças entre grupos estudados, ou de rejeitar a probabilidade de uma diferença zero

entre eles. A análise das magnitudes de efeito, por outro lado, fornece informação positiva acerca da grandeza da relação observada entre factores. Interligado com a questão da natureza da informação fornecida está o problema do papel do factor  $N$ , a dimensão da amostra estudada, no processo de produzir um valor de  $p$  (a probabilidade de observar uma determinada diferença ou relação). O valor de  $p$  é sempre uma função de  $N$ , seja qual for a importância prática da diferença ou relação observada (Conboy, 2000; Hays, 1981). Deste modo, amostras grandes podem dar origem a valores reduzidos de  $p$ , exagerando assim a importância aparente de diferenças que podem ser, de facto, triviais. Para além disso, uma vez que a magnitude do efeito proporciona informação da intensidade de relações, também fornece ao investigador melhor informação acerca da generalização e eventual reprodutibilidade dos resultados observados.

Existem numerosas medidas de magnitude do efeito (Huberty, 2002). Têm sido classificadas frequentemente em duas grandes categorias: As diferenças padronizadas e as medidas de variância contabilizada (Thompson, 2002a). Cooper (1997), por outro lado, descreve três categorias da referida métrica, tendo-as denominado de tipo  $d$ , (equivalente às diferenças padronizadas), tipo  $r$  (equivalente às medidas de variância contabilizada) e uma terceira categoria, a *odds ratio*, ou seja, o rácio de probabilidades.

Para fins de organização deste artigo, referem-se alguns delineamentos experimentais comuns enquadrados, sempre que possível, dentro das grandes categorias identificadas. O rácio de probabilidades será tratado numa secção que se centra na análise de dados categóricos.

#### DIFERENÇAS PADRONIZADAS: MAGNITUDES DE EFEITO TIPO $d$

A experiência clássica submete um grupo de sujeitos a um tratamento experimental a fim de comparar o efeito observado na variável dependente com o efeito num grupo de controlo, que não sofre a exposição ao tratamento experimental. A magnitude do efeito, nessa situação experimental, é, geralmente, uma diferença padronizada entre os dois valores médios observados na

variável dependente. Assim, a magnitude do efeito é dada por:

$$(1) \quad d = (M_E - M_C) / DP$$

Onde:

$d$  é a magnitude do efeito;

$M_E$  é o valor médio observado no grupo experimental;

$M_C$  é o valor médio observado no grupo de controlo;

DP é uma medida de variabilidade, um desvio-padrão.

Deste modo, a magnitude do efeito, tipo  $d$ , é uma medida da diferença entre médias em termos de unidades de desvio-padrão. Mas qual a medida de variabilidade a utilizar? De facto, há várias possibilidades: podemos usar o desvio-padrão do grupo de controlo, o do grupo experimental, ou ainda alguma combinação dos dois. Para além disso, nos cálculos, podemos optar por utilizar o desvio-padrão da amostra observada (com  $N$  no denominador) ou a estimativa do desvio-padrão da população,  ${}_e\sigma$  (com  $N-1$  no denominador).

Nesta questão há essencialmente três posições. Segundo Glass (1976; 2000; Glass, & Hopkins, 1996; Glass, McGaw, & Smith, 1981; Smith & Glass, 1977) deve-se usar a estimativa não-enviesada do desvio-padrão da população do grupo de controlo. Então o valor a atribuir ao DP da fórmula (1) seria:

$$(2) \quad {}_e\sigma_c = [\sum(x_i - M_C)^2 / N - 1]^{1/2}$$

Onde:

${}_e\sigma_c$  é a estimativa não-enviesada do desvio-padrão da população da qual foi retirada a amostra do grupo de controlo;

$x_i$  é o valor observado do indivíduo  $i$  no grupo de controlo;

$M_C$  é o valor médio observado no grupo de controlo;

$N$  é o número de elementos no grupo de controlo.

Embora se pressuponha que, em princípio, existe uma homogeneidade da variância entre os dois grupos (experimental e de controlo) aleatoriamente seleccionados da mesma população,

Glass indica que a intervenção experimental pode alterar não só a tendência central da variável dependente no grupo experimental, mas também a sua variabilidade (Glass, McGaw, & Smith, 1981).

Numa posição contrária, Cohen (1988) defende que a melhor estimativa do desvio-padrão da população encontra-se na variabilidade combinada dos dois grupos,  ${}_e\sigma_w$  (desvio-padrão *pooled within groups*) dado por:

$$(3) \quad {}_e\sigma_w = (({}_e\sigma_E^2 + {}_e\sigma_C^2) / 2)^{1/2}$$

Onde:

${}_e\sigma_w$  é a estimativa não enviesada do desvio-padrão da população *pooled within groups*;

${}_e\sigma_E^2$  e  ${}_e\sigma_C^2$  são as estimativas não-enviesadas da variância populacional nos grupos experimental e de controlo (i.e. com  $N-1$  no denominador).

Hedges (1981; Hedges & Olkin, 1985) concorda com a posição de Cohen que o termo da padronização deverá incluir informação não só do grupo de controlo mas também do grupo experimental. Contudo, argumenta que o uso da estimativa não-enviesada da variabilidade da população inflaciona indevidamente a magnitude de efeito observada. Assim recomenda que no denominador da fórmula (1) se deve utilizar  $S_w$ , ou seja, o desvio-padrão *pooled within groups* :

$$(4) \quad S_w = [(S_E^2 + S_C^2) / 2]^{1/2}$$

Onde:

$S_w$  é o desvio-padrão *pooled within groups* e  $S_E^2$  e  $S_C^2$  são as variâncias das amostras observadas nos grupos experimental e de controlo (i.e. com  $N$  do denominador).

Resumindo, podemos definir as três medidas mais comuns da magnitude de efeito do tipo diferenças padronizadas, do seguinte modo:

$$(5) \quad d \text{ de Glass} = (M_E - M_C) / {}_e\sigma_C$$

$$(6) \quad d \text{ de Cohen} = (M_E - M_C) / {}_e\sigma_w$$

$$(7) \quad g \text{ de Hedges} = (M_E - M_C) / S_w$$

O uso de  ${}_e\sigma_w$  ou de  $S_w$  no denominador da fórmula (1) pressupõe que o tratamento experimental não altera a variabilidade do grupo expe-

rimental. Este pode ser um pressuposto demasiado forte e irrealista. Por outro lado,  $e\sigma_w$  tem a vantagem de ser uma estimativa não-enviesada da variabilidade verdadeira na população, quer a hipótese nula seja verdadeira, quer não (Glass, & Hopkins, 1996). Para além disso, com o desenvolvimento de métodos de meta-análise, foi demonstrado que o  $d$  de Cohen está menos enviesado que o  $d$  de Glass. Note-se também que o valor de  $g$  de Hedges será sempre menor que o  $d$  de Cohen. Por estas razões, o  $d$  de Cohen tem ganho maior aceitação entre os investigadores. De qualquer forma, ao relatar valores de magnitude do efeito do tipo diferenças padronizadas, o investigador tem a responsabilidade de informar os leitores de qual a fórmula que foi utilizada.

#### INTERPRETAÇÃO DA MAGNITUDE DO EFEITO TIPO $d$

Vimos que a magnitude do efeito, tipo  $d$ , é a quantificação padronizada do aumento, incremento, melhoria ou benefício que observamos, devido à intervenção estudada (pressupondo, claro, um valor positivo de  $d$ ). Se aceitarmos o pressuposto da normalidade da distribuição, este benefício pode ser descrito em termos de uma proporção de sujeitos do grupo experimental que podemos esperar exceder o valor típico do grupo de controlo em intervenções futuras. O valor da magnitude do efeito, tipo  $d$ , permite-nos chegar a uma estimativa da proporção de sujeitos num grupo experimental que terá melhor cotação, em estudos futuros, do que o valor médio do grupo de controlo. O valor desta proporção é facilmente derivado através do pressuposto de normalidade da distribuição da variável dependente e de uma análise de uma tabela de valores de  $z$  que fornece as proporções de área localizada por baixo de secções da curva normal padronizada (para ilustração de um exemplo trabalhado, consultar Conboy, 2002). Por exemplo, uma magnitude de efeito de  $d=0.4$  implica que, nos estudos futuros, podemos esperar que 65.54% dos sujeitos num grupo experimental excederá o valor médio de um grupo de controlo;  $d=1.0$  implica 84.13%;  $d=1.4$  implica 91.92%.

#### A MAGNITUDE DO EFEITO E A SIGNIFICÂNCIA PRÁTICA

Uma análise dos valores acima citados leva-nos à conclusão de que um  $d=1.4$  representa um efeito superior a um  $d=0.4$ . Mas será assim mesmo? E «quão grande é suficientemente grande»? Esta é a questão fundamental da estatística aplicada e a pergunta a que os testes de hipótese, e os seus valores associados de  $p$ , respondem presumivelmente com clareza. Ao adoptar uma metodologia que enfatiza a magnitude do efeito, quer como complemento quer em substituição dos testes de hipótese, não podemos fugir a esta questão primordial.

Tem havido várias tentativas de estabelecer um guião geral que caracterize os níveis de magnitude do efeito segundo as categorias de «pequeno» ou «grande». Talvez o mais conhecido seja o esquema de Cohen (1988) que sugeriu  $d=0.2$  como sendo um efeito «pequeno»,  $d=0.5$  como «intermédio» e  $d=0.8$  como «grande». Para alguns autores (e.g. Barnette & McLean, 2002b), o estabelecimento dessas categorias foi um erro grave. Mesmo assim, muitos investigadores agarraram-se a estes termos a fim de facilitar a interpretação dos efeitos observados. Mas quem fizer isso está por certo a simplificar demasiado as suas conclusões. O próprio Cohen avisa-nos contra o uso arbitrário destes termos indicando que

Os termos «pequeno», «intermédio» e «grande» são relativos, não só uns perante os outros, mas em relação à área científica em que o estudo se insere, ou mais particularmente, relativamente ao conteúdo e método de investigação usados num determinado estudo (...) O uso de um quadro de referência comum (...) é recomendado somente no caso de não existir nenhuma base de comparação melhor (1988, p. 25).

Relativamente à mesma problemática, Glass afirma,

Não faz sentido tentar associar regiões da medida de magnitude do efeito com adjectivos descritivos como «pequeno», «moderado», «grande» etc. Desligado de um contexto de decisão e valor comparativo, não existe valor inerente a uma

magnitude do efeito quer ela seja de 3.5 ou 0.2. Dependendo dos benefícios que podem ser alcançados a determinado custo, uma magnitude do efeito de 2.0 pode ser «fraca» e uma de valor 0.1 pode ser «boa» (Glass, McGaw, & Smith, 1981, p. 104).

Nas duas opiniões supracitadas temos exemplos de duas abordagens à interpretação de valores de magnitude do efeito e duas noções do conceito de «significância prática». Cohen prefere que o efeito observado seja comparado com efeitos previamente estabelecidos dentro de uma determinada área de investigação. Deste modo, encontrar um efeito de  $d=0.3$  quando valores tipicamente observados são na ordem de 0.2, pode indicar um efeito importante. Mas a observação de um efeito de  $d=0.6$  por outro lado, pode não ser considerado como importante se os valores tipicamente observados no âmbito da área científica de interesse forem de ordem de  $d=0.8$ . Glass, por sua vez, advoga que seja efectuada uma análise de custos e benefícios. Assim, se os custos de uma determinada intervenção forem muito baixos, e os benefícios grandes, um efeito observado de  $d=0.2$  pode ter enorme significância prática. Por outro lado, a demonstração de um efeito de  $d=0.6$  pode não ter significância prática se os custos forem demasiado elevados.

De qualquer modo, não existe nenhum nível de magnitude do efeito que se traduza, automaticamente, em significância prática. A significância prática terá que ser estabelecida caso a caso, usando um destes métodos: uma análise de custos versus os benefícios e/ou através de uma comparação com os efeitos previamente observados na mesma área temática.

#### MAGNITUDE DO EFEITO NO ESTUDO PRÉ-TESTE / PÓS-TESTE

No caso de um estudo feito com um delineamento pré-teste / pós-teste, a magnitude do efeito é dada por:

$$(8) \quad d = (M_{\text{pós}} - M_{\text{pré}}) / DP$$

Onde:

$d$  é a magnitude do efeito;  
 $M_{\text{pós}}$  é o valor médio observado no pós-teste;  
 $M_{\text{pré}}$  é o valor médio observado no pré-teste;  
 $DP$  é uma medida de variabilidade, por exemplo o desvio-padrão.

Permanece a polémica da escolha do valor de  $DP$  nesta fórmula. Uma vez que há fortes razões para pressupor uma alteração na variabilidade numa amostra de sujeitos do pré- para o pós-teste, o argumento de Glass (no sentido de utilizar o desvio-padrão «controlo» – aqui o  $\sigma$  do pré-teste) parece particularmente forte. Por outro lado, as vantagens já referidas do uso de  $\sigma_w$  ou  $S_w$  no denominador também permanecem na situação pré-pós. O investigador cuidadoso informará os seus leitores de decisões tomadas nesse sentido e acerca das respectivas justificações.

Deve-se mencionar que a fórmula (8) é indicada para a análise de dados primários. É prática corrente, no âmbito de estudos de meta-análise, utilizar fórmulas para transformar valores estatísticos (por exemplo  $t$ ) em valores de  $d$ . Se for essa a intenção do investigador, é imprescindível lembrar que as fórmulas de transformação são diferentes para os delineamentos envolvendo grupos independentes e os de medidas repetidas. Cortina e Nouri (2000) apresentam as respectivas fórmulas de transformação. Realce-se que, no delineamento de medidas repetidas, a correlação entre os valores pré e pós é um elemento imprescindível no cálculo da magnitude de efeito obtido à base de valores estatísticos. Por esta razão, o investigador deverá sempre relatar o valor da correlação entre valores de pré-teste e pós-teste.

#### O ÍNDICE DE MUDANÇA FIEL E A SIGNIFICÂNCIA CLÍNICA

Uma outra abordagem à descrição da magnitude do efeito em estudos que usam um delineamento pré- pós é o índice de mudança fiel (*Reliable Change Index-RCI*) sugerido por Jacobson e Truax (1991).

Em questões relacionadas com a eficácia de uma dada terapia, um objectivo comum é descrever se um determinado cliente/paciente demonstrou melhoria em relação à sua situação

clínica inicial. Sendo assim, não é suficiente focar a mudança de médias de grupos. Para além disso, não é suficiente descrever melhoria apenas em relação à posição clínica inicial do cliente, uma vez que um cliente pode ficar melhor sem ficar bom. Assim, na situação de terapia, o terapeuta tem que responder a duas perguntas: (a) Será que a mudança observada no cliente é verdadeira? Por outras palavras, será que a mudança observada não se deve apenas a erro de medida?; e, (b) Será que a mudança observada coloca o cliente dentro das normas de grupos funcionais?

Com o objectivo de responder a essas questões, Jacobson e Truax (1991) sugerem que a diferença observada entre o pós-teste e o pré-teste seja dividida pelo erro padrão da diferença ( $S_{dif}$ ) a fim de criar um índice de mudança fiel. O erro padrão da diferença é o desvio padrão de uma distribuição teórica de mudanças (diferenças) que irão ser observadas numa população, se de facto, não houver mudança na população (i.e. erros de diferença que serão observáveis se a diferença verdadeira for de zero). Assim, a mudança observada no cliente é posta à prova relativamente ao erro de medida e à fidelidade do instrumento usado. Após a verificação de que a mudança observada é verdadeira, é então comparada com as normas previamente definidas.

O índice de mudança fiel é dado por

$$(9) \quad RCI = (Pós - Pré) / S_{dif}$$

e o erro padrão de diferença é dado por:

$$(10) \quad S_{dif} = (2 (S_{em})^2)^{1/2}$$

onde  $S_{em}$  é o erro padrão de medida. Finalmente, o  $S_{em}$  é dado por

$$(11) \quad S_{em} = DP_x (1 - r_{xx'})^{1/2}$$

Onde

$S_{em}$  é o erro padrão de medida,

$DP$  é o desvio-padrão do instrumento

$r_{xx'}$  é uma estimativa da fidelidade do instrumento.

Assim, tendo disponível o desvio-padrão do instrumento utilizado e uma estimativa da sua fidelidade, podemos calcular o  $S_{dif}$  e o  $RCI$ . Deste modo, o índice  $RCI$  relata a mudança

observada num indivíduo, ao longo do período de terapia, em termos de unidades de  $S_{dif}$  de uma população em que a mudança verdadeira é zero.

Uma vez que esta métrica,  $RCI$ , não é nem muito conhecida nem disponível através de programas como o SPSS (2001), vamos dar um exemplo específico do seu cálculo e interpretação. Como o índice  $RCI$  é especialmente indicado para situações em que se pretende estabelecer significância clínica, imaginemos um grupo de sujeitos depressivos que completa um percurso de terapia. Antes de iniciar o tratamento todos os sujeitos respondem a uma escala de sintomas depressivos. A escala tem fidelidade de  $\alpha = .85$  e um desvio-padrão de  $DP = 10$ . Neste caso, o erro padrão de medida é

$$S_{em} = 10 \times (1 - .85)^{1/2} = 3.87$$

e o erro padrão de diferença é:

$$S_{dif} = (2 \times (3.87)^2)^{1/2} = 5.48$$

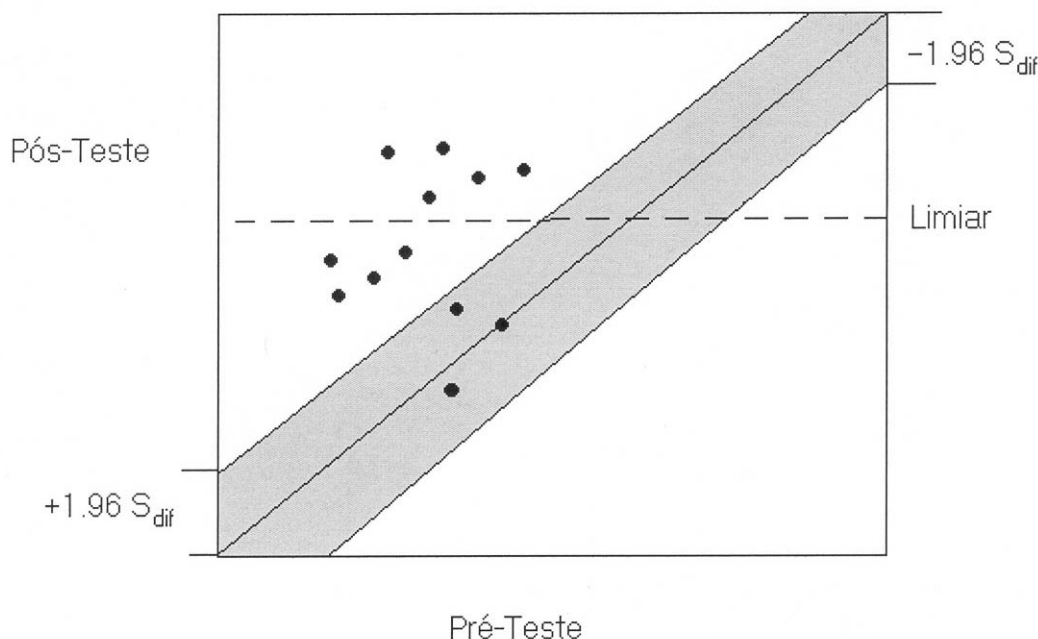
Agora, imaginemos que um dos clientes depressivos teve um valor pré-terapia de 30 pontos e um valor pós-terapia de 45 na escala utilizada. Assim o índice de mudança fiel para esse cliente é dado por

$$RCI = (45 - 30) / 5.48 = 2.74$$

O índice de mudança fiel ensaia, essencialmente, a hipótese de que a mudança observada no cliente foi de zero. Se um cliente mostrar mudança maior que 1.96 erros padrão de diferença, então rejeitamos essa hipótese de «mudança zero» (com confiança de .95) e aceitamos que a mudança observada é verdadeira. No exemplo acima, uma vez que o valor do  $RCI$  atinge um valor de 2.74 ( $RCI > 1.96$ ), a mudança observada será considerada fiel. Tendo concluído que a mudança é fiel, o terapeuta agora compara o resultado do cliente com um valor limiar que indica se o cliente fica, ou não, dentro das normas de grupos funcionais.

Jacobson e Truax sugerem a apresentação dos dados num gráfico de dispersão, acompanhado de a) uma linha diagonal que representa «mudança zero»; b) ponto normativo limiar que indica funcionamento normativo (representado por uma linha tracejada horizontal); e c) o inter-

FIGURA 1  
**Índice de mudança fiel: Dados fictícios apresentados no formato recomendado por Jacobson & Truax (1991)**



valo de confiança ( $1.96 S_{dif}$ ) à volta da linha de «mudança zero». Apresenta-se na Figura 1 um exemplo em que três indivíduos ficam situados na zona que indica que as mudanças observadas não são fiéis (um caso mostra deterioração, um fica precisamente no ponto de «mudança zero», e um terceiro apresenta-se com mudança positiva mas não fiel); quatro casos situam-se na zona representando mudança fiel sem terem alcançado o valor limiar para serem considerados parte do grupo funcional; e cinco clientes mostram mudança fiel bem como a passagem para o grupo funcional.

**MEDIDAS DE VARIÂNCIA CONTABILIZADA:  
 MAGNITUDES DE EFEITO TIPO  $r$**

As medidas mais comuns do tipo variância contabilizada são  $r^2$  e  $R^2$ ,  $\eta^2$  ( $\eta^2$ ), e  $\omega^2$  ( $\omega^2$ ). A primeira,  $r^2$ , também conhecida como o

coeficiente de determinação, diz-nos qual a proporção de variabilidade contabilizada («explicada») por uma relação linear entre duas variáveis. Assim, a correlação  $r = .50$  dá origem a um valor de  $r^2 = .25$ , ou seja 25% da variabilidade observada «explica-se» através de uma possível relação linear entre os dois factores. O coeficiente de correlação múltipla,  $R$ , quando elevado ao quadrado, dá-nos a proporção da variância explicada por uma relação linear entre uma variável prevista e um conjunto de variáveis de previsão. Os coeficientes  $r^2$  e  $R^2$  são as medidas de magnitude do efeito mais comuns na investigação efectuada com uma análise correlacional e de regressão. No caso de se utilizar, como é vulgar hoje em dia, um software estatístico, pode constatar-se que este tipicamente relata não só o  $R^2$  mas também o  $R^2$  ajustado (uma vez que o  $R^2$  baseado numa amostra sobre-estima, em geral, o valor na população). Se utilizar um valor de  $R^2$  como uma magnitude do efeito, o

investigador deverá ter o cuidado de distinguir qual o valor que relata,  $R^2$  ou  $R^2$  ajustado.

Na situação de Análise da Variância Oneway, as medidas frequentemente relatadas são de  $\eta^2$  e  $\omega^2$ . Essas medidas de magnitude do efeito são apenas válidas, contudo, se os níveis da variável independente forem ordenados: por exemplo, se a variável independente fosse a administração de uma nova droga anti-depressiva e os níveis fossem constituídos por grupos de sujeitos que receberam 0.1 mg, 0.2 mg, 0.3 mg e 0.4 mg. Neste caso, a ordenação dos níveis da variável independente é evidente e invariante. Por outro lado, se os níveis da variável independente fossem três diferentes drogas anti-depressivas e um placebo, então o uso de  $\eta^2$  ou  $\omega^2$  seria incorrecto uma vez que não existe nenhuma seriação evidente dos valores da variável independente. Nesse caso, a solução mais típica é a de relatar magnitudes de efeito de contrastes planeados: Droga 1 vs. Placebo; Droga 2 vs. Placebo, Droga 1 vs. Droga 2, e assim por diante.

Se existirem apenas dois níveis da variável independente, quer o  $\eta^2$  quer o  $\omega^2$  podem ser utilizados como uma medida de magnitude do efeito, mas nesta situação é mais tipicamente relatado um valor de uma das versões de magnitude do efeito do tipo  $d$ .

O cálculo de  $\eta^2$  é dado por

$$(12) \quad \eta^2 = (SS_{\text{total}} - SS_{\text{erro}}) / SS_{\text{total}}$$

Vários autores notaram que  $\eta^2$  se encontra enviesado positivamente (Hays, 1981; Huberty, 1994). Hays propôs então  $\omega^2$  como uma alternativa não enviesada: O cálculo de  $\omega^2$  é dado por

$$(13) \quad \omega^2 = (SS_{\text{trat}} - (k - 1) MS_{\text{erro}}) / (SS_{\text{total}} + MS_{\text{erro}})$$

Barnette e McLean (2002a) descreveram a relação entre  $d$ ,  $\eta^2$  e  $\omega^2$ . Estes autores realizaram um estudo tipo Monte Carlo e concluíram que a relação entre  $\eta^2$  e  $\omega^2$  é linear e previsível seja qual for a dimensão da amostra. Contudo, a relação entre estas medidas e  $d$  é mais variável: Quando o número de grupos analisados é de  $k=2$ , a relação é essencialmente linear; mas nas situações de  $k > 2$  essa relação desmantela-se rapidamente.

Ainda não se encontram bem desenvolvidos

métodos de cálculo da magnitude do efeito em modelos de ANOVA mais complexos. O leitor deverá consultar os trabalhos de Cortina e Nouri (2000) e Cooper e Hedges (1994) para uma análise mais profunda sobre este assunto. Nos delineamentos experimentais analisados por tais modelos mais complexos, a abordagem mais simples é a de efectuar uma série de contrastes específicos sempre entre dois grupos, que resulta no relato de um dos valores do tipo  $d$ .

#### MAGNITUDES DE EFEITO PARA DADOS CATEGÓRICOS

Depois de efectuar uma análise do cruzamento de duas variáveis nominais e a análise de frequências por células, enfrentamos o mesmo problema que surge relativamente aos testes paramétricos: nem o valor da medida estatística, nem o valor associado de  $p$ , nos dizem alguma coisa sobre a intensidade da relação entre as duas variáveis. O  $\chi^2$  significativo, por exemplo, implica apenas que o investigador rejeita a probabilidade de independência total das duas distribuições. Assim, após ter efectuado uma análise do tipo  $\chi^2$ , é aconselhável calcular e relatar uma estimativa da magnitude do efeito. Dependendo da situação experimental, existe uma série de medidas úteis e ao dispor do investigador. Algumas (por exemplo  $\phi$ ,  $\lambda$ , e  $V$  de Cramér) já foram descritas por Siegel (1975) e Hays (1972; 1981) e encontram-se, geralmente, disponíveis como opções em SPSS versão 11.0, (Analyze>Descriptive Statistics>Crosstabs/Statistics). Outros, como BESD e o Rácio de Probabilidades, são menos conhecidos.

$\phi$  e  $\phi^2$

O uso do coeficiente phi ( $\phi$ ) é mais indicado na situação em que a tabela de contingência tem apenas duas colunas e duas filas. Após ter efectuado uma análise de  $\chi^2$ , o valor desse coeficiente é dado por

$$(14) \quad \phi = (\chi^2 / N)^{1/2}$$

Tal como o  $r$  de Pearson, o coeficiente  $\phi$  pode variar entre  $-1$  e  $+1$ . Se os valores das duas variáveis puderem ser considerados como sendo dicotómicos e ordenados (codificados «1» ou «0»),



então pode-se demonstrar que  $\phi = r$  (Hays, 1972). Sendo assim, o valor de  $\phi^2$  (dado simplesmente por  $\chi^2 / N$ ) pode ser tratado como uma medida de variância contabilizada tal como  $r^2$ . Embora o  $\phi^2$  possa ser relatado para tabelas com mais de duas colunas ou filas, essa métrica perde a sua utilidade em tal circunstância uma vez que o limite superior poderá então exceder 1.0.

*BESD*

Após o cálculo de um valor  $\phi$ , ou de um valor  $r$ , é possível construir uma representação teórica que ilustre o aumento revelado pelos dados a respeito da taxa de sucesso. Esta representação constitui-se na forma de um quadro de dupla entrada no qual a informação proveniente dos dados é colocada em categorias dicotômicas. Rosenthal e Rubin (1982) referem-se a este tipo de representação como um *Binomial Effect Size Display* (BESD). É especialmente útil na situação de meta-análise quando o investigador quer interpretar o significado de um valor médio de  $r$  calculado à base de uma série de estudos independentes. O valor de  $r$  é transformado e apresentado numa tabela de contingência, 2 x 2, com a atribuição de valores iguais a todos os valores marginais. As linhas correspondem a uma variável independente ou de previsão, e as colunas a uma variável dependente ou prevista. Como exemplo, pensemos numa série de estudos genéricos que envolvem a aplicação de um

Tratamento versus um Placebo. Suponhamos que se observa, neste estudo, uma correlação média entre as duas variáveis de  $r = .40$ . Qual seria o significado de tal correlação se os resultados fossem simplificados em «Melhoria Observada» e «Melhoria Não Observada»? Constrói-se a tabela BESD (Quadro 1), atribuindo à partida valores iguais nas margens. A taxa de sucesso, calcula-se por

$$(15) \quad .50 + r / 2$$

e a taxa de insucesso (ou seja, sucesso associado com o placebo) calcula-se por

$$(16) \quad .50 - r / 2$$

A interpretação do BESD é simples e directa. Se não existisse nenhum efeito do tratamento (i. e., no caso de  $r = 0$ ), os valores calculados seriam todos .50. No entanto, o que se verificou, com base na correlação de  $r = .40$ , foi que a aplicação do tratamento neste estudo aumentou a taxa de sucesso de .30 para .70. (É importante lembrar que estas proporções não representam os dados brutos observados, mas sim, uma representação teórica do que se pode esperar em futuros estudos com base na correlação observada). A grande utilidade do BESD é que nos permite responder a questões como: Se, numa determinada população, à metade dos elementos for admi-

QUADRO 1  
*Binomial Effect Size Display (BESD) que tem por base  $r = .4$*

	Resultado		Total
	Melhoria Observada	Melhoria Não Observada	
Tratamento	.70	.30	$N_{\text{tratamento}}$
Placebo	.30	.70	$N_{\text{placebo}}$
Total	$N_{\text{melhoria}}$	$N_{\text{não melhoria}}$	

Nota:  $N_{\text{tratamento}} = N_{\text{placebo}} = N_{\text{melhoria}} = N_{\text{não melhoria}}$

nistrado um tratamento e à outra metade um placebo, e se metade da população melhora e a outra metade não, qual é o aumento na taxa de sucesso que podemos esperar observar devido ao efeito do tratamento? Wolf (1986, p. 33) fornece uma tabela indicando o aumento na taxa de sucesso para vários valores de  $r$ .

*V de Cramér*

O coeficiente  $\phi$  é um caso específico da medida mais generalizada  $V$  de Cramér. Nomeadamente,  $V = \phi$  se o número de colunas e linhas na tabela de contingência for de dois (Iversen & Gergen, 1997). Sendo assim, o uso de  $V$  como uma medida de magnitude do efeito é especialmente indicado quando o número de colunas ou linhas na tabela de contingência for maior que dois. O cálculo desta métrica é dado por

$$(17) \quad V = [(\chi^2 / N (L-1))]^{1/2}$$

onde  $L$  representa o mais pequeno dos dois números, o número de linhas e o número de colunas. Diferentemente do coeficiente  $\phi$ ,  $V$  assume valores entre 0 e +1 (os valores perto de 0 indicam a independência dos factores, enquanto valores perto de 1 indicam dependência). Assim, os valores de  $V$  não devem ser interpretados como indicando «variância contabilizada».

*$\lambda$  de Goodman e Kruskal*

Até que ponto será possível prever a preferência de um aluno em relação à apresentação de matéria curricular, se soubermos qual o estilo de aprendizagem do aluno? Para responder a esse

tipo de pergunta podemos recorrer ao índice  $\lambda$  de Goodman e Kruskal.  $\lambda$  é um índice de «associação previsível» com valores possíveis entre 0 e +1. Um determinado valor de  $\lambda$  diz-nos até que ponto a informação numa variável nos ajuda a prever o valor de uma segunda variável. O valor de 0 implica que o conhecimento relativamente à categoria a que pertence um sujeito, numa variável, não nos fornece informação nenhuma que ajude a prever a categoria do sujeito numa segunda variável. Um valor de +1 implica que tal conhecimento fornece toda a informação necessária para prever com precisão a categoria da segunda variável.

No referido exemplo, para calcular  $\lambda$ , realiza-se primeiro o somatório das frequências celulares máximas observadas para cada estilo de aprendizagem (sublinhadas no Quadro 2), depois subtrai-se o valor máximo observado na coluna de frequências totais para as diferentes categorias (igualmente sublinhado). Divide-se o resultado obtido pelo resultado da subtração entre o número total de observações e o valor máximo observado na coluna de frequências totais. Assim, para os dados do Quadro 2:

$$\lambda = (150 + 90 + 90 - 250) / (410 - 250) = .50$$

A interpretação de  $\lambda$  é muito fácil de compreender: A disponibilidade de informação sobre uma variável reduz a probabilidade de erro na previsão do valor da segunda variável por  $\lambda \times 100\%$ . Seguindo o exemplo, se tivermos disponível informação acerca do estilo de aprendizagem de um aluno (A, B, ou C), reduz-se em 50%

QUADRO 2  
*Dados fictícios para o cálculo de  $\lambda$  de Goodman e Kruskal*

Preferência	Estilo A	Estilo B	Estilo C	Total
Verbal	<u>150</u>	<u>90</u>	10	<u>250</u>
Visual	30	40	<u>90</u>	160
Total	180	130	100	<u>410</u>

a probabilidade de errar ao prever a posição daquele aluno relativamente à preferência por apresentação verbal ou visual.

*O Rácio de Probabilidades*

*Odds ratio* (Haddock, Rindskopf, & Shadish, 1998; Rudas, 1998) pode ser traduzido como rácio de probabilidades, rácio de diferenças, disparidades de proporção ou até rácio de partido ou aposta. A ideia é simples para quem já trabalhou com tabelas de contingências e lembra-nos a situação que requer uma análise do tipo qui-quadrado. O rácio de probabilidades é muito comum nos estudos médicos de sobrevivência uma vez que associa a presença de um factor (por exemplo, Tratamento versus Placebo) com um acontecimento (por exemplo, Sobrevivência versus Morte). Imagine um estudo no âmbito da reinserção social. Um grande número de reclusos, prestes a sair da prisão, participa em dois programas de intervenção com vista à sua reinserção na sociedade. Realiza-se um estudo longitudinal a fim de avaliar a eficácia dos dois programas, P<sub>1</sub> e P<sub>2</sub>. Pergunta-se, após um período de cinco anos, qual a taxa de reincidência nos dois grupos? Apresentam-se no Quadro 3 dados fictícios para um tal estudo.

À primeira vista parece que o programa de intervenção P<sub>1</sub> foi mais eficaz que o programa P<sub>2</sub> e podemos efectuar uma análise do tipo qui-quadrado para ensaiar a hipótese de independência de programas de intervenção e o resultado em termos de reincidência. A fim de descrever a magnitude de efeito, contudo, determinamos o rácio de probabilidades (RP). O rácio dentro do programa P<sub>1</sub> é de 75:25, ou 3:1 a favor da situa-

ção «recluso consegue manter uma vida social sem voltar à prisão». Dentro do programa P<sub>2</sub> o rácio é de 60:40 ou 3:2. Se o tipo de programa não constituir um factor em termos de reincidência, esperamos um RP igual a 1.0. No caso do exemplo, o rácio das probabilidades (RP= 3:1 / 3:2) é igual a 2. O recluso que participa no programa P<sub>1</sub> tem duas vezes mais a probabilidade de ficar fora da prisão após cinco anos que o recluso que participa no programa P<sub>2</sub> – um resultado não necessariamente evidente quando se comparam os valores sucesso de 75% versus 60%. Assim o rácio de probabilidades fornece uma medida mais concreta da magnitude do efeito do que um mero valor de qui-quadrado ou do valor de *p* que lhe está associado.

Como indicaram Haddock, Rindskopf, e Shadish (1998), o investigador pode preferir relatar o logaritmo natural do RP em vez do valor bruto do mesmo. Uma vez que o logaritmo natural de 1 é zero, este procedimento tem a vantagem de transformar a escala de RP de tal modo que um valor de zero represente a ausência total de associação entre o tratamento e o resultado.

A QUESTÃO DOS INTERVALOS DE CONFIANÇA

Uma das recomendações do Task Force da APA (Wilkinson et al., 1999) secundada por Thompson (2002b), é a de que investigadores deverão relatar os intervalos de confiança à volta de medidas de magnitude do efeito. Segundo Cortina e Nouri (2000), os procedimentos para o cálculo de intervalos de confiança à volta de me-

QUADRO 3

*Rácio de probabilidades: Dados fictícios para um estudo de taxa de reincidência após cinco anos*

	Programa de Reinserção Social	
	P <sub>1</sub>	P <sub>2</sub>
Participantes fora da prisão após cinco anos	75%	60%
Participantes em prisão após cinco anos	25%	40%

didadas de magnitude do efeito são iguais aos procedimentos para calcular um intervalo à volta de outras estatísticas, por exemplo, à volta de um valor médio: Primeiro, calcula-se o erro padrão de  $d$  e seguidamente multiplica-se esse valor por um valor crítico da escala de  $z$  (por exemplo, 1.96, a fim de criar um intervalo de confiança de 95%). O produto desses factores é então somado e subtraído do valor observado de  $d$ , produzindo os limites superiores e inferiores do intervalo.

Contudo, este procedimento foi questionado por alguns autores (Cumming & Finch, 2001; Steiger & Fouladi, 1997; Thompson, 2002b) que defendem que a determinação dos limites superior e inferior dos intervalos à volta de magnitudes de efeito deverá ser efectuada usando distribuições não-centrais. Uma análise mais aprofundada desta polémica extravasa o presente trabalho. A temática é apenas mencionada de forma cursiva no sentido de alertar o leitor para o facto de que, embora a inclusão de intervalos de confiança à volta de magnitudes de efeito seja um objectivo desejável, não é necessariamente uma tarefa simples. No futuro, face a um maior desenvolvimento das metodologias de medidas estatísticas de magnitude do efeito, por certo esta questão virá a ser clarificada.

### CONCLUSÃO

As medidas de magnitude do efeito são ferramentas estatísticas genericamente úteis quer em estudos isolados quer na situação de meta-análise (Lipsey, & Wilson, 2001). Estas métricas facultam informação respeitante à intensidade de relações entre variáveis, ou seja, ao grau de divergência entre a hipótese nula e os resultados observados.

Neste artigo foram tratadas as medidas mais comuns da magnitude do efeito: a da experiência clássica que inclui um grupo de controlo e um grupo experimental, a situação pré-teste/pós-teste, os estudos de correlação e regressão, a análise de variância, e o caso da análise de dados categóricos. Seguindo as numerosas recomendações recentes (APA, 2001; Barnette, & McLean, 2002b; Thompson, 2002b; Wilkinson et al., 1999), os investigadores em ciências sociais, comportamentais, médicas e educacionais deverão ter o cuidado de relatar magnitudes de

efeito adequadas aos seus dados. Os agentes de formação em estatística e metodologias de investigação deverão incluir estes tópicos nos seus planos curriculares. Os métodos de aplicação da magnitude de efeito estão actualmente bem desenvolvidos para as situações de delineamentos experimentais mais simples e comuns.

Questões ainda em aberto sobre o tópico incluem aquelas relacionadas com modelos de análise mais complexos: os procedimentos para o cálculo de magnitude do efeito ficam progressivamente mais intrincados com delineamentos mais complexos. Para além dessa questão, o uso de intervalos de confiança à volta de valores de magnitude de efeito é recomendado mas, por enquanto, o assunto está insuficientemente desenvolvido para o investigador avançar no seu uso com a desejada confiança.

### REFERÊNCIAS

- APA (1994). *Publication manual of the American Psychological Association* (4ª edição). Washington, DC: Autor.
- APA (2001). *Publication manual of the American Psychological Association* (5ª edição). Washington, DC: Autor.
- Barnette, J., & McLean, J. (2002a, Abril). *Shedding light on the eta-square and omega-square relationships with the standardized effect size*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.
- Barnette, J., & McLean, J. (2002b, Abril). *The need to abolish the arbitrary effect size standards*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans: LA.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Conboy, J. (2002). A magnitude de efeito como forma de análise em ciências sociais. *Lusiada Psicologia*, 1 (1), 187-210.
- Conboy, J. (2000). Estimando a dimensão do N de uma amostra. *INUAF Studia*, 1 (2), 227-274.
- Cooper, H. (1997). Some finer points in meta-analysis. In M. Hunt (Ed.), *How science takes stock: The story of meta-analysis* (pp. 169-181). New York: Russell Sage Foundation.
- Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Cortina, J., & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage.

- Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and non-central distributions. *Educational and Psychological Measurement, 61*, 532-575.
- Fidler, F. (2002). The 5th edition of the APA Publication Manual: Why its statistics recommendations are so controversial. *Educational and Psychological Measurement, 62* (5), 749-770.
- Glass, G. (2000). Meta-analysis at 25. Retirado do World Wide Web em 1 de Fevereiro de 2002: <http://glass.ed.asu.edu/gene/papers/meta25.html>
- Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 10*, 3-8.
- Glass, G., McGaw, B., & Smith, M. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Glass, G., & Hopkins, K. (1996). *Statistical methods in education and psychology*, Third Edition. Needham Heights, MA: Allyn & Bacon.
- Haddock, C., Rindskopf, D., & Shadish, W. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3* (3), 339-353.
- Hays, W. L. (1981). *Statistics for the social sciences* (3rd ed.). New York: Holt, Rinehart and Winston.
- Hays, W. L. (1972). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart and Winston.
- Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Henkel, R. (1976). *Tests of significance*. Beverly Hills, CA: Sage.
- Huberty, C. (2002). A history of effect size indices. *Educational and Psychological Measurement, 62* (2), 227-240.
- Huberty, C. (1994). A note on interpreting an R-squared value. *Journal of Educational and Behavioral Statistics, 19*, 351-356.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.
- Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Kendall, P. (Ed.) (1999). Clinical significance [Special section]. *Journal of Consulting and Clinical Psychology, 67*, 283-339.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Applied Social Research Methods Series (Vol. 49). Thousand Oaks, CA: Sage.
- Roberts, J. K., & Henson, R. (2002). Correction for bias in estimating effect sizes. *Educational and Psychological Measurement, 62* (2), 241-253.
- Rosenthal, R., & Rubin, D. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166-169.
- Rudas, T. (1998). *Odds ratios in the analysis of contingency tables*. Thousand Oaks, CA: Sage.
- Siegel, S. (1975). *Estatística não-paramétrica*. São Paulo: McGraw-Hill.
- Smith, M., & Glass, G. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32* (9), 752-760.
- SPSS. (2001). *Statistical package for the social sciences* (Versão 11.0). [Computer Software]. Chicago: IL: SPSS.
- Steiger, J., & Fouladi, R. (1997). Non-centrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.
- Tallmadge, G. (1977). *The joint dissemination review panel ideabook*. Washington, DC: National Institute of Education and the US Office of Education.
- Thompson, B. (2002a). «Statistical», «practical», and «clinical»: How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80* (1), 64-71.
- Thompson, B. (2002b). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31* (3), 25-32.
- Wilkinson, L., & APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

## RESUMO

As medidas de magnitude do efeito representam uma resposta eficaz às diversas críticas às metodologias de investigação tradicionalmente utilizadas nas ciências sociais. Tais medidas evitam algumas das dificuldades mais evidentes dos testes de hipótese; são de interpretação simples, sem o recurso a valores de  $p$ , que podem ser enganadores; e fornecem uma base para a síntese de grande número de estudos. Este artigo descreve o cálculo e interpretação das formas mais comuns desta abordagem alternativa, ou complementar, aos testes de hipótese. Pressupõe-se que o leitor dispõe de um conhecimento básico de estatística aplicada sem ter, necessariamente, uma forte formação em estatística matemática.

*Palavras-chave:* Magnitude do efeito, Métodos de Investigação, Análise de Dados, Ciências Sociais.

## ABSTRACT

Measures of effect size represent an alternative to traditional methods of data analysis in the social sciences. These metrics avoid some of the most evident problems associated with tests of hypotheses, are more simply interpreted than sometimes misleading  $p$  values, and provide the basis for the synthesis of large

numbers of research studies. This text assumes that the reader has a basic understanding of inferential statistics and describes the calculation and interpretation of common forms of effect size that can be used either as alternatives, or complements, to classical hypothesis testing.

*Key words:* Effect size, Research methods, Data analysis, Social Sciences.