



Fernando Arrobas, médico dentista, Pós-graduado em Tratamento Estatístico de Dados Quantitativos

## A BAD SCIENCE EXISTE ( $p < 0.05$ )

A reprodutibilidade na investigação científica constitui um dos princípios fundamentais para se atestar a sua qualidade, isto é, um investigador que repita uma mesma experiência (sua ou de outro investigador) deve, em princípio, obter resultados equivalentes. Ora, em 2016, a revista Nature decidiu realizar um inquérito a mais de 1500 cientistas das mais diferentes áreas. A primeira pergunta do questionário era a seguinte: “Considera que existe uma crise de reprodutibilidade na produção científica atual?”. De forma surpreendente (ou não), a resposta dos investigadores foi clara: 90% dos investigadores afirmaram que “Sim”. Mais concretamente, 52% disseram que “Sim, uma crise significativa” e 38% que “Sim, uma crise ligeira”. Em seguida, quando convidados a responder se “Alguma vez tinham falhado em replicar outra experiência?”, 65% dos investigadores na área da medicina responderam que “Sim”, sendo que 55% deles não tinham conseguido replicar o seu próprio trabalho. Entre os fatores que contribuíam para esta crise, no *top-3* constavam “selective reporting” (preferência das revistas científicas e dos investigadores por resultados positivos em vez de negativos), “pressão para publicar” e “poucos conhecimentos de estatística”.

De facto, quem conhece os meandros da investigação, sabe que existe uma aversão geral dos investigadores para com a estatística e os métodos quantitativos. “Em cima da hora, contrata-se uma empresa ou um amigo das matemáticas que resolve o assunto...” – exclamam alguns com soberba. Na verdade, é boa prática contar na equipa de investigação com um membro que tenha esses conhecimentos especializados. Contudo, isso deve ser feito desde o início do projeto, sob pena de, no final, os dados recolhidos não espelharem a qualidade desejada. A título de exemplo, mesmo que o  $p$ -value seja  $< 0.05$ , encontram-se, com frequência, estudos com uma amostra não aleatória e de dimensão inadequada, pois a dimensão do efeito era pequena e, como tal, não se obtém a potência do teste recomendada ( $1 - \beta > 0.8$ ) .... Aliás, a leitura exclusiva do  $p$ -value é também uma conduta errónea comum. O valor de  $p < 0.05$  significa apenas que a hipótese nula ( $H_0$ ) apresenta uma reduzida probabilidade de ser verdadeira, isto é, que os testes estatísticos demonstraram que as diferenças observadas (por exemplo, entre o grupo experimental e o grupo controlo) são superiores àquelas que seriam esperadas encontrar por “mero aca-

so”. Na verdade, num trabalho rigoroso, a cruzada da investigação não termina quando se encontra o  $p$ -value.... Apenas começa! Sem a dimensão do efeito, não se entende qual é a significância na “prática” e é a potência do teste que permite rejeitar corretamente a hipótese nula quando ela é falsa. A tabela I demonstra como deve ser realizada a teoria da decisão em estatística.

A estatística é um mundo muito mais interessante do que pode parecer à primeira vista. Ainda por cima no século XXI, em que grande parte das fórmulas matemáticas foram substituídas por *softwares* de “alta cilindrada”. A sua missão é uma das mais nobres: ajudar a avaliar criticamente o que se lê, ouve e transmite aos outros. É importante lembrar que, hoje em dia, vivemos num mundo em que a informação se multiplica a cada milésimo de segundo. A internet, por exemplo, trouxe muitas vantagens e é hoje um atributo indispensável de uma sociedade desenvolvida. No entanto, também ajudou a criar muito “lixo” informativo. No caso da ciência, as plataformas online mudaram o modelo de negócio das publicações tradicionais e as soluções *open access* alargaram o “funil”, facilitando em alguns casos os critérios de seleção de artigos. Isto torna as leituras ainda mais exigentes, até porque o estudo dos *papers* publicados representa um passo fundamental da cadeia de valor em medicina. São os conhecimentos adquiridos através destas publicações que contribuem para as decisões clínicas. Com efeito, se a seleção não for feita de uma forma criteriosa, coloca-se em risco a saúde pública. Tal como as *(fake) news* influenciam eleições políticas, também os artigos científicos ou as teses podem influenciar a saúde e a qualidade de vida das pessoas. Não restam dúvidas: como demonstra o inquérito da Nature ou como se viu recentemente com o documentário *Root cause* da Netflix, a “bad science” existe ( $p < 0.05$ ,  $r = 0.5$ ,  $1 - \beta > 0.8$ ) e é importante que não tenha vindo para ficar. Para que tal aconteça, todos temos uma responsabilidade. Felizmente que, para cada problema, existe uma solução. Herbert George Wells, ilustre pensador e escritor inglês, escreveu em 1898: “Statistical thinking will be one day as necessary for efficient citizenship as the ability to read and write...”. A profecia estava certa. Só falta agora colocar em prática. ■

[fernando.arrobas@jornaldentistry.pt](mailto:fernando.arrobas@jornaldentistry.pt)

		Decisão do teste	
		Rejeitar $H_0$	Não rejeitar $H_0$
População	Efeito presente <b><math>H_0</math> é falsa</b> (exemplo: o novo tratamento tem efeito)	<b>Decisão correta</b> (Efeito está presente e foi detetado)  <b>Probabilidade = <math>1 - \beta</math></b> (Potência do teste)	<b>Erro de tipo II</b> (Efeito está presente, mas não foi detetado)  <b>Probabilidade [Erro tipo II] = <math>\beta</math></b>
	Sem efeito presente <b><math>H_0</math> é verdadeira</b> (exemplo: o novo tratamento não tem efeito)	<b>Erro de tipo I</b> (Efeito detetado, mas este não existe)  <b>Probabilidade [Erro tipo I] = <math>\alpha</math></b>	<b>Decisão correta</b> (Efeito não detetado e este não existe)  <b>Probabilidade = <math>1 - \alpha</math></b>

Tabela I – Teoria de decisão em estatística. Adaptado de: Marôco, J. (2018). Análise Estatística com o SPSS Statistics. 7ª edição. Pêro Pinheiro: ReportNumber, Lda. p. 57