



Coordenação de Armindo Rodrigues

O problema dos falsos valores-p (*p-values*) na acumulação de evidência estatística

Autora:

Maria de Fátima Brilhante

Um princípio fundamental em Ciência é o da reprodutibilidade, que é a possibilidade de uma experiência poder ser replicada pelo mesmo investigador ou por outros de forma independente. Porém, num processo de replicação, e quando há amostras aleatórias em jogo, existe a possibilidade de sermos confrontados com conclusões diferentes ou contraditórias, isto no meio de um conjunto de experiências que foram realizadas com o objetivo de responder a questões semelhantes. Por isso, nos últimos anos vários autores têm chamado a atenção para a importância de acumular evidência estatística (científica), que se encontre dispersa por vários estudos similares, com o objetivo de construir o conhecimento científico.

Nas situações em que há informação contraditória ou dispersa sobre um problema, a Meta-Análise pode contribuir para combinar ou sintetizar toda a informação relevante disponível. Contudo, a Meta-Análise recorre essencialmente a métodos que se baseiam em modelos estatísticos que possuem um conjunto de pressupostos, que uma vez violados descredibilizam o uso destes. Uma forma de contornarmos a situação é recorrermos a uma abordagem não paramétrica, que pode consistir na com-



Gregor Mendel (1822-1884)

binção de valores-p (*p-values*), se o único objetivo for testar a veracidade de uma hipótese nula comum, pois estes contêm evidência estatística que poderá suportar a hipótese. Porém, a combinação de valores-p assenta no facto de estes serem observações independentes de um modelo uniforme padrão. A não uniformidade da amostra de valores-p traz problemas ao nível da combinação destes, pois os métodos de combinação conhecidos deixam de ter as propriedades que lhes são conhecidas. Mas existirão falsos valores-p, isto é, valores-p não uniformes? A explicação dada por dois estatísticos portugueses para acabar com a controvérsia de Mendel-Fisher aponta no sentido de sim.

Como nota, a controvérsia de Mendel-Fisher surgiu em 1936 com o artigo de R. Fisher "*Has Mendel's work been rediscovered?*". Nesse artigo, Fisher, um vulto da Estatística, levanta sérias suspeitas sobre a improbabilidade dos dados obtidos por G. Mendel, considerado o pai da Genética. Desde então, vários autores têm vindo a argumentar a favor de Fisher e outros a favor de Mendel, mas sem nunca se ter chegado a um veredicto final. Em 2010 surgiu uma explicação plausível para os dados "demasiado bons" de Mendel em "*A statistical model to explain the Mendel-Fisher controversy*" de Pires e Branco. Nesse artigo, o modelo com o melhor ajustamento aos dados é o que prevê a possibilidade de Mendel ter repetido as experiências, sempre que os resultados não eram o que esperava. Esta situação expõe o que muitos já suspeitavam, que é alguns investigadores decidirem pela repetição da experiência na esperança de obterem um "resultado melhor" que o primeiro obtido.

O nosso trabalho de investigação incidiu sobre o problema da estimação do número de falsos valores-p numa amostra. Foi necessário arranjar um modelo estatístico que espelhasse a realidade. Como a situação mais verosímil é um investigador repetir uma só vez a experiência, é expetável que este venha escolher como

melhor resultado aquele que conduzir ao valor mínimo ou ao valor máximo de dois valores-p uniformes. A escolha entre estes dois valores dependerá das expectativas inicialmente criadas pelo investigador em relação ao resultado da experiência. Assim, consideramos o que se chamou modelo de Mendel para descrevermos uma amostra de valores-p, a qual pode conter observações não uniformes. O modelo de Mendel tem um único parâmetro m ($-2 \leq m \leq 2$), que está relacionado com a proporção de falsos valores-p na amostra. Se $-2 \leq m < 0$, uma variável aleatória (v.a.) de Mendel é uma mistura entre uma v.a. uniforme padrão, e uma v.a. Beta(1,2); ou se $0 < m \leq 2$, é uma mistura entre uma v.a. uniforme padrão e uma v.a. Beta(2,1). Acontece que o mínimo entre duas uniformes padrão independentes é uma Beta(1,2), enquanto o máximo uma Beta(2,1). Logo, o coeficiente de mistura $|m|/2$ determina a proporção de valores-p não uniformes na amostra.

A estimação de m torna-se assim crucial por forma a decidirmos que tipo de metodologia estatística se deve usar, ou desenvolver, caso se conclua a não uniformidade da amostra ($m \neq 0$). Mas, quando se recorreu a estudos de simulação para atestarmos a eficácia de diferentes métodos de estimação do parâmetro m , todos eles demonstraram ser ineficazes na sua estimação, sobretudo se a proporção de falsos valores-p na amostra é infe-



Sir Ronald Fisher (1890-1962)

rior a 0.5. Isto levanta-nos um problema, porque queremos crer que seja apenas uma pequena fração de investigadores que decida replicar uma experiência. Este simples problema de estimação do número de falsos valores-p veio mostrar-nos o quão importante é os investigadores planearem com rigor as suas experiências, para que alguns não se sintam tentados a repetilas, caso os resultados os "surpreendam". Estar-se-á assim a contribuir para uma posterior acumulação fiável da evidência estatística disponível.

 III Encontro Luso-Galaico
de Biometria

III Encontro Luso-Galaico de Biometria, Aveiro 28 a 30 de junho

Maria de Fátima Brilhante é professora na Faculdade de Ciências e Tecnologia da Universidade dos Açores e membro do Centro de Estatística e Aplicações da Universidade de Lisboa. Participou

no III Encontro Luso-Galaico de Biometria, que decorreu de 28 a 30 de junho de 2018 em Aveiro, onde apresentou a comunicação oral "*Population Dynamics Equilibrium and Extreme Growth*".