

A estatística nos ensaios clínicos: Conceitos chave

Introdução

Os métodos estatísticos contabilizam formalmente as fontes de variabilidade nas respostas dos doentes ao tratamento. A utilização da estatística permite ao investigador clínico chegar a inferências razoáveis e precisas a partir da informação recolhida e tomar decisões sólidas na presença de incertezas. A estatística é fundamental para prevenir erros e desvios na investigação médica. Este artigo aborda alguns dos conceitos estatísticos chave e a sua aplicação em ensaios clínicos.

Teste de hipóteses

Uma hipótese é uma suposição ou um conjunto de suposições que a) afirma algo com uma base provisória para orientar a investigação científica; ou b) confirma algo como altamente provável em função dos fatos estabelecidos.

Para os nossos fins, interessa-nos a hipótese que afirma algo; por exemplo, que um novo tratamento para uma doença é melhor que o padrão de tratamento existente. Si o novo tratamento se denomina “B” e o padrão de tratamento se denomina “A”, então a hipótese afirma que “B” é melhor do que “A”.

Pode pensar que os cientistas se dedicarão a demonstrar esta hipótese, mas não é esse o caso. Em vez disso, a aproximação a este objetivo é indirecta. Em vez de intentar demonstrar a hipótese B, o método científico assume que, na verdade, A é verdadeira; que não existem diferenças entre o padrão de tratamento e o novo tratamento. Isto é conhecido como a hipótese “Nula”. Os cientistas tentam então demonstrar que A é

falso. Isto também é conhecido como demonstrar que a hipótese nula é falsa. Se for possível fazê-lo (demonstrar que a hipótese A é falsa e que o padrão de tratamento não é melhor que o novo tratamento), deduz-se que B é verdadeira e que o novo tratamento é melhor que o tratamento padrão.

Porque se faz isto?

Não existe uma resposta simples para esta pergunta. É um método amplamente aceite que evoluiu na ciência moderna, mas pode ajudar a utilizar uma analogia do âmbito jurídico. A hipótese nula abrange a nossa situação ou conhecimento atual (assim, utilizando uma analogia de tribunal, “o acusado é inocente”), na qual devemos confiar exceto se existirem evidências suficientes para não o fazer. Mas se pretendêssemos demonstrar a “hipótese alternativa” (conhecida como, contrária à “hipótese nula”), então, na verdade, “o acusado é culpado”.

Outra forma, talvez mais fácil, de chegar a este ponto é citar Albert Einstein:

“Nenhum número de experiências, por muitas que sejam, poderão provar que tenho razão, mas será suficiente uma só experiência para demonstrar que estou equivocado.”

Isto parece sugerir que tentar demonstrar que a hipótese nula é falsa ou está errada é um objetivo mais rigoroso e alcançável do que tentar demonstrar que a hipótese alternativa é correta. Tenha em consideração que isto NÃO explica devidamente porque é que a ciência adota esta abordagem; mas talvez nos possa ajudar a compreender e aceitar mais facilmente um conceito complicado.

Erros do Tipo I e do Tipo II

Se olhar para a tabela seguinte, poderá ver a diferença entre erros do Tipo I (falsos positivos) e erros do Tipo II (falsos negativos).

	A hipótese nula é verdadeira	A hipótese nula é falsa
Rejeitar a hipótese nula	Erro do Tipo I “Falso positivo”	Resultado correto “Verdadeiro Positivo”
Não se pode rejeitar a hipótese nula	Resultado correto “Verdadeiro negativo”	Erro do Tipo II “Falso negativo”

Isto continua a ser muito confuso, assim, de modo a expressá-lo de forma mais simples a seguir está um exemplo muito claro:

- Os erros do Tipo I poderão matar um doente. Imagine um estudo que determinasse incorretamente que o padrão de tratamento não era melhor que o novo tratamento, e que, como tal, levou a que fossem administrados novos tratamentos aos doentes com resultados catastróficos. Se forem cometidos erros do Tipo I, irá detetar-se incorretamente um efeito que não existe.
- Os erros do Tipo II significam que se desperdiça investigação potencialmente valiosa. Talvez esta investigação tivesse sido realmente útil, mas, uma vez que o estudo não é continuado, os doentes não sofrem nenhum dano. Se forem cometidos erros do Tipo II, não se irá detetar um efeito que existe.

Está então claro que os erros do Tipo I são mais graves que os erros do Tipo II no que respeita aos doentes.

Nível de significância

O nível de significância é a probabilidade de cometer um erro do Tipo I. Este nível é afetado pelo tamanho da amostra e pelo “poder estatístico” do teste.

Poder estatístico

O “poder” de um teste estatístico é a probabilidade de levar corretamente à rejeição de uma hipótese nula. Ou, por outras palavras, a capacidade de um teste para detetar um efeito se este existir. Outra forma de o descrever é dizer que o “poder” de um teste é a probabilidade de NÃO cometer um erro do Tipo II.

Valores de p

Os valores de p ou valores de “probabilidade”, ponderam a força da evidência numa escala de 0 a 1. Um valor p baixo (normalmente inferior a 0,05 ou 5%) indica que existem evidências sólidas *contra* a hipótese nula, o que poderia levar à sua rejeição, enquanto que um valor p alto (superior a 0,05) indica o contrário.

Correlação versus causalidade

Ao analisar os resultados de um ensaio, é importante recordar que correlação não é o mesmo que causalidade. A correlação é quando duas variáveis estão relacionadas de alguma forma, no entanto, isto não significa que uma cause a outra (existe uma associação entre as duas variáveis). Um exemplo disto envolve a terapia hormonal de substituição (THS) e a doença coronária (DC), em que as mulheres às quais se administrava THS apresentavam um menor risco de DC. No entanto, isto não era devido ao processo de THS em si, mas sim ao fato de que o grupo de mulheres que recebiam THS costumavam pertencer a um grupo socioeconómico superior, com dietas e planos de exercício melhores que a média.

A causalidade pode ser observada quando um fator causa um resultado. Um fator causal é frequentemente a causa parcial de um resultado. Para diferenciar entre correlação e causalidade, é importante registar tanta informação quanto possível sobre

os participantes em ensaios. Também é necessário aplicar com cuidado a metodologia científica no desenho dos ensaios clínicos e avaliar o possível desvio no ensaio.

Manipulação de dados

A manipulação de dados é a prática de informar dados incorretamente (de forma seletiva) ou de criar resultados falsos. Um exemplo disto é quando se descartam intencionalmente dados que não estão de acordo com o resultado esperado, para aumentar a proporção de resultados que confirmarão a hipótese formulada. Quando um investigador elimina os valores aberrantes (um resultado que é muito inferior ou superior ao resultado seguinte mais próximo) dos resultados, é importante verificar que se tratam realmente de valores aberrantes e não simplesmente de resultados que são diferentes dos resultados esperados ou pretendidos. Outro exemplo de manipulação de dados é quando um recompilador de dados gera aleatoriamente um conjunto completo de dados a partir da medição realizada num único doente.

Transformação de dados

A transformação de dados é a aplicação de uma fórmula matemática a alguns dados obtidos num ensaio. É frequentemente utilizada para que a apresentação dos dados seja mais clara ou fácil de entender. Por exemplo, se se estiver a medir a eficiência do combustível dos automóveis, é normal medir a eficiência na forma de “quilómetros por litro”. No entanto, se se estiver a avaliar quanto combustível adicional seria necessário para aumentar a distância percorrida, seria expressa em “litros por quilómetro”. A aplicação de uma fórmula incorreta para obter a nova apresentação destes dados neste caso, iria afetar os resultados globais do ensaio.

Fusão de dados

A fusão de dados é a ação de combinar os dados de múltiplos estudos de modo a poder compreender melhor uma situação. Um dos seus formatos mais frequentes é a meta-análise, na qual se reúnem os resultados de vários ensaios publicados para os agregar e comparar. Ao realizar uma meta-análise, é importante verificar cuidadosamente que as metodologias dos ensaios são as mesmas ou comparáveis. Deverá ter-se em consideração qualquer diferença no desenho, de modo a que não existam variáveis diferentes subjacentes (variáveis confundentes). Um exemplo de fusão incorreta de dados é a agregação de dados de vários ensaios com diferentes espécies de ratinhos como animal do ensaio.

A2-4.33.3-V1.1