**Statistics: An Introduction using R**

M.J. Crawley

Síntese - Josiane M Lopes

**Capítulo 1 – Fundamentos**

A parte mais difícil de um trabalho estatístico é o começo. E a escolha certa da análise estatística depende principalmente do tipo de dados que se possui e da pergunta a que se pretende responder, ou seja, conhecer quais os tipos de variáveis estão envolvidas.

A **variável resposta** é aquela do eixo y e a variável explanatória é aquela que explica o comportamento, ou seja, as do eixo x.

O método estatístico adequado irá variar de acordo com o tipo de variável que se deseja entender ou explicar.

**Tudo varia**

A estatística é uma ferramenta para isolar a variabilidade natural daquela causada por um fator cujo comportamento se deseja compreender. Ela ajuda a prever o que irá acontecer com um fator.

É importante salientar que existe diferença entre não-significância e sem importância, ou seja, nem sempre a falta de diferença estatística se traduz em um resultado ruim ou que não exista diferença.

**Significância**

A significância pode ter diferentes sentidos, na estatística se refere à incerteza de afirmar que um determinado fenômeno ocorreu ou não. Na estatística os eventos ditos improváveis são os que ocorrem em até 5 % das observações.

**Boa e má hipóteses**

A hipótese boa é aquela que pode ser rejeitada, já a má hipótese é aquela que não pode ser rejeitada. A ausência de evidência não é evidência de ausência.

**Hipótese nula**

É aquela que afirma que “nada está acontecendo”, ou seja, que não há efeito do fator estudado para a resposta observada. A hipótese nula afirma sempre que os tratamentos são iguais.

**Valor *p***

É a estimativa da probabilidade de um resultado ou um resultado mais extremo do que o observado possa ter ocorrido por acaso, caso a hipótese nula seja verdadeira. Em resumo o valor p é uma medida da credibilidade da hipótese nula. O valor *p* é interessante, mas não conta a história toda, pois os tamanhos da eficácia e das amostras são igualmente importantes para desenhar as conclusões.

**Interpretação**

Há dois tipos de erros na interpretação dos modelos estatísticos. Quando se rejeita a hipótese nula e ela é verdadeira é chamado de erro tipo I. Quando se aceita a hipótese nula e ela deve ser rejeitada, é o erro do tipo II.

**Modelagem estatística**

O ideal é que se encontre um modelo que se ajuste aos dados obtidos. O melhor modelo é aquele que terá menor variação não explicada (desvio residual mínimo). Pode haver mais de um modelo adequado, sendo necessário aplicar o princípio da parcimônia e escolher o modelo que explique de forma mais simples o conjunto dos dados.

**Probabilidade máxima**

É necessário escolher adequadamente o modelo, para que os dados estejam com a probabilidade máxima de estarem corretos.

**Delineamento experimental**

Baseado em dois conceitos importantes: a replicação e a casualidade. Quando se replica aumenta-se a confiabilidade, quando casualisa se reduz o comportamento tendencioso dos dados. O ideal é aplicar ambos os conceitos de casualidade e replicação. Se o delineamento experimental for ideal não importará quão avançada seja a análise estatística, será possível obter resultados confiáveis.

**O princípio da parcimônia (A navalha de Occam)**

Em síntese: a explicação mais simples é a mais correta.

**Observação, teoria e experimento**

Para se resolver problemas científicos são necessários ter uma mistura de observação, teoria e experimento. Porém, em algumas situações o pesquisador é obrigado a sacrificar um desses itens.

**Controle**

Sem controle não há conclusões.

**Replicação**

É necessário se ter repetições, pois se for feito a mesma coisa para indivíduos diferentes há uma possibilidade de se ter uma resposta diferente. Se fazem repetições para se aumentar a confiabilidade e tornar possível o cálculo da variabilidade do tratamento. Se forem feitas medidas repetidas de um mesmo animal ou de um mesmo local não é repetição, e sim uma pseudo-repetição.

**Quantas repetições?**

Quanto mais repetições, mais representativa a amostra e mais confiável a análise.

**Poder**

O poder de um teste é a probabilidade de rejeitar a hipótese nula quando ela é falsa.

**Aleatoriedade**

Pressupõe-se escolher as amostras ao acaso, mas o ideal é garantir a aleatoriedade na amostragem, a qual pode ser feita através de sorteio. Geralmente é algo mais complexo e que deve ser cuidadosamente planejado.

**Forte dedução**

Um das formas mais eficazes de se comprovar a precisão de uma idéia é a confirmação experimental de uma previsão que teve origem em uma hipótese bem formulada. Duas medidas são essenciais: formular uma hipótese clara e planejar um teste adequado. Sendo que as duas juntas garantem maior confiabilidade.

**Fraca dedução**

É uma expressão utilizada para interpretar estudos observacionais. O objetivo de uma estatística boa é obter o máximo de informações de um conjunto de dados. O conhecimento sobre as condições iniciais antes de o evento natural ocorrer e ser mensurado pela ciência é de fundamental importância ou qualquer conclusão não será satisfatória ao ponto de sustentar hipóteses futuras.

**Duração**

A duração deve ser determinada no início do planejamento do experimento.

**Pseudo-replicação**

Ocorre quando se analisa os dados e se obtém mais graus de liberdade do que realmente se dispõe. Existem dois tipos de pseudo-replicação:

1. Temporal: medidas repetidas de mesmo indivíduo;
2. Espacial: série de medidas da mesma vizinhança.

A pseudo-replicação pode ser um problema porque uma das principais pretensões de uma análise estatística é a independência de erros. Ao se perceber que esses erros foram cometidos, algumas medidas podem ser tomadas

**Condições iniciais**

Muitos trabalhos científicos podem ser prejudicados por falta de informação das condições iniciais, pois como saber se algo mudou quando não temos conhecimento de como era do estudo. Uma das razões para se conhecer as condições iniciais é que essas informações podem ser utilizadas para melhorar a resolução de uma análise final através da análise de covariância.

**Design ortogonal e dados observacionais não-ortogonais**

Os experimentos ortogonais são os controlados, nos quais não há valores perdidos ou não acessíveis. Nos experimentos observacionais não há controle sobre o número de indivíduos que se podem analisar. Isso significa que a ordem em que os fatores são adicionados ou removidos de um dado modelo, terá influencia diferente no caso dos dados não ortogonais mas não afetará o resultado para os dados ortogonais.