Estatística Aplicada a Recursos Hídricos

Docente: Rachid Muleia

(rachid.muleia@uem.mz)

Mestrado em Gestão de Recursos Hídricos - DGEO/UEM

Tema: Inferência Estatística: Análise de Variância- ANOVA

Ano lectivo: 2023

ANOVA- Motivação

- Considera duas variáveis: uma quantitativa e uma qualitativa
- A variável quantitativa é designada de variável resposta ou variável de interesse
- A variável qualitativa é designada de factor, que é uma variável que se julga ter uma influência sobre a variável resposta
- A ANOVA é um procedimento de teste de hipótese usado para avaliar as diferenças entre as médias de dois ou mais tratamentos ou grupos (populações). ANOVA usa dados de amostra para fazer inferências sobre populações.

ANOVA e Test *t*

ANOVA é uma versão mais geral do teste t de duas maneiras:

- Ambos os testes usam dados de amostra para testar hipóteses sobre médias populacionais. ANOVA, no entanto, pode testar hipóteses sobre duas ou mais médias populacionais. O teste *t* só pode testar hipóteses sobre duas médias populacionais.
- O teste *t* só pode ser usado com uma variável independente (classificação), enquanto o ANOVA pode ser usado com qualquer número de variáveis independentes (classificação), isto é , ANOVA com um, dois ou mais factores

ANOVA- exemplo

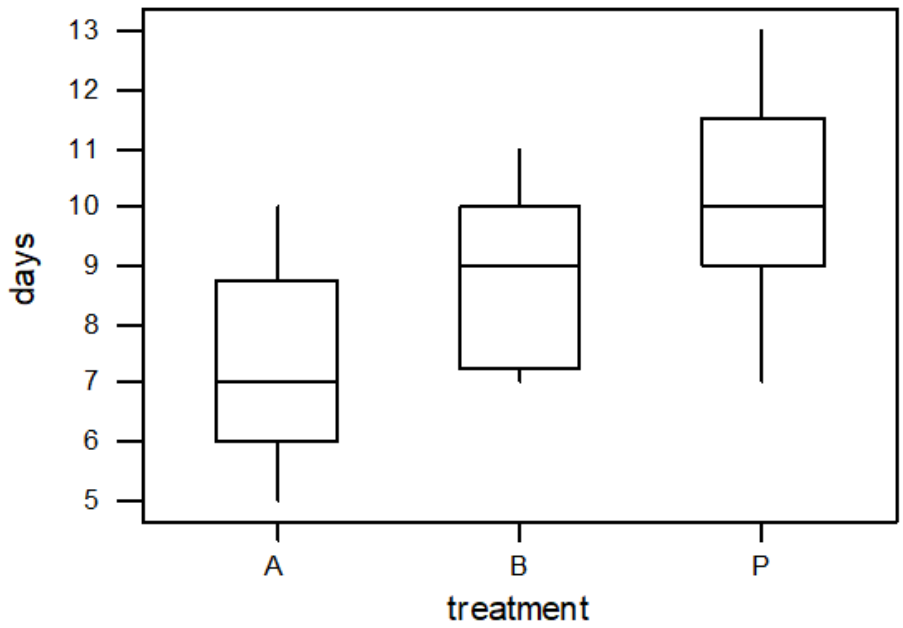
Considere 25 pacientes com bolhas na pele e são submetidos a três tratamentos, nomeadamente, tratamento A, tratamento B e Placebo. Em cada um dos tratametos mede-se o número de dias até a cura das bolhas.

	Número de dias até a cura								$\bar{x}$
Tratamento A	5	6	6	7	7	8	9	10	7.25
Tratamento B	7	7	8	9	9	10	10	11	8.875
Placebo	7	9	9	10	10	11	12	13	10.11

As diferenças observadas nas médias são estatisticamente significativas?

ANOVA- Teste informal

Inspeção visual: box-plot para cada categoria, lado a lado ou multiplos histogramas



A ANOVA determina a quantidade de variabilidade em cada grupo de dados e verificar se a variabilidade entre os grupos é maior do que a variabilidade dentro dos grupos.

O que a ANOVA faz?

Na sua forma mais simples (existem extensões), a ANOVA testa as seguintes hipóteses:

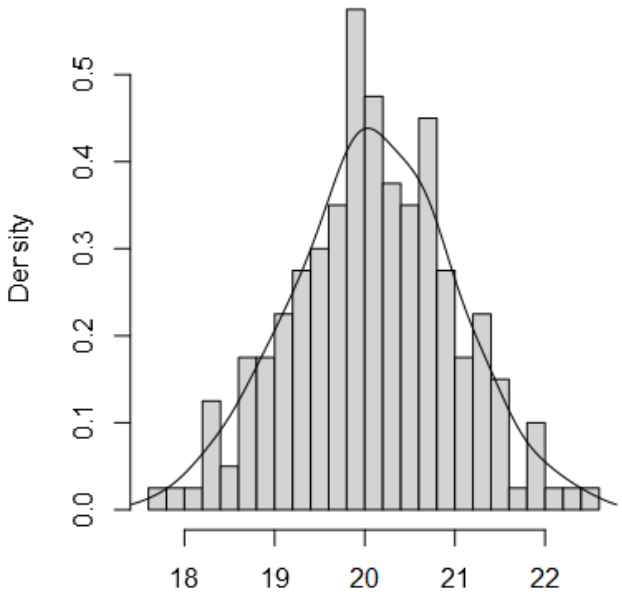
H_0 :As médias dos grupos são todas iguais

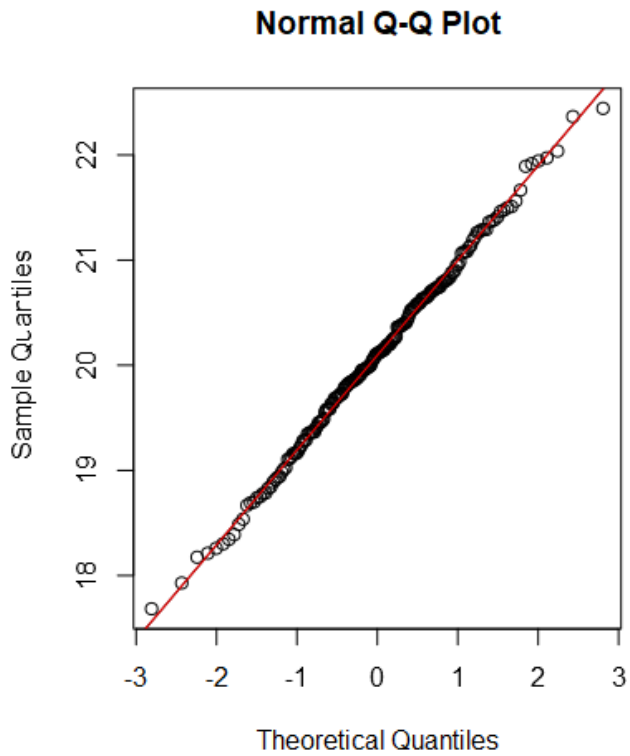
H_1 :Nem todas as médias são iguais

- A hipótese alternativa não diz que pares de média são diferentes
- Pode-se achar os pares de médias que diferem através de comparações múltiplas

Pressupostos da ANOVA

- Os dados em cada sub-população ou grupo seguem distribuição normal: este pressuposto pode ser verificado usando o histograma ou o qq-plot





- A ANOVA é robusta a desvios não serveros da normalidade
- As variâncias de cada grupo devem ser iguais

Como Conduzir a ANOVA

- Considere o exemplo da cura das bolhas

$$H_0 : \mu_A = \mu_B = \mu_P$$
$$H_1 : \text{pelo menos um par de médias difere} \quad \mu_i \neq \mu_j$$

- Uma maneira de fazer isso seria usar testes t em todos os pares possíveis de testes (aqui são apenas três). No entanto, se tivermos mais grupos, isso torna-se complicado. Por exemplo, com 10 grupos, necessário fazer ${}^{10}C_2 = 45$.
- Este procedimento é designado de comparações múltiplas
- No entanto, além de ser demorado, realizar vários testes, pode inflacionar o erro tipo I, colocando em causa a validade dos resultados

Comparações múltiplas

- Suponha que todas as médias sejam realmente iguais (H_0 é verdadeira) e conduzamos todos os três testes aos pares
- Suponha também que os testes sejam independentes e feitos a $0,05$ nível de significância
- Então a probabilidade de não rejeitar H_0 em todos os três testes é de $(1 - 0,05)^3 = 0,953 = 0,857$ e, portanto, probabilidade de rejeitar pelo menos uma das hipoteses nula , chamada de taxa de erro familiar, é $1 - 0,857 = 0,143 > 0,05$
- Com 45 testes, a probabilidade de rejeitar pelo menos 1 deles (incorretamente!) é superior a 90%!

Como conduzir a ANOVA

- Na ANOVA, temos duas fontes de variação: variabilidade entre os grupos e variabilidade dentro dos grupos
- Variabilidade entre os grupos: verifica-se a diferençã entre a média de cada grupo e a média total

$$\left(\bar{x}_i - \bar{X}\right)^2$$

- Variabilidade dentro dos grupos: verifica-se a diferença entre os valores de cada cada grupo e a sua respectiva média.

$$\left(x_{ij} - \bar{x}_i\right)^2$$

- Se as médias específicas do grupo variarem em torno da média geral mais do que as observações individuais variam em torno de suas médias amostrais específicas do grupo, então temos evidências de que as médias populacionais correspondentes são de facto diferentes.

Teste F

- Considere o exemplo da cura das bolhas

```
            Df Sum Sq Mean Sq F value  Pr(>F)
tr           2  33.25  16.625   5.892 0.00931 **
Residuals   21  59.25   2.821
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- $F = \dfrac{\text{Variabilidade entre}}{\text{Variabilidade dentro}} = \dfrac{MST}{MSE} = \dfrac{16,625}{2,821} = 5.892$
- Um valor alto de F indica relativamente mais diferença entre os grupos do que dentro dos grupos (evidência contra H_0). Valores de F próximo de 1 indicam evidências a favor da alternativa, visto que $MST \approx MSE$

Teste F

- Como em qualquer teste estatístico, precisamos definir uma regra de decisão: quão grande o valor observado para F, F_{obs}, precisa ser para rejeitar H_0?
- Se a hipótese nula for verdadeira, então a nossa estatística de teste F tem distribuição $F - Snedecor$, isto é, $F \sim F_{(r-1,n_T-r)}$
- Rejeitamos H_0 quando F_{obs} assume valores altos, o que significa que nossa região de rejeição R está na cauda direita da distribuição. Para um nível fixo de significância, tem-se $R = \{F : F_{obs} > F_{(1-\alpha,r-1,n_T-r)}\}$
- O teste F para ANOVA é inerentemente unilateral, rejeitando H0 somente se F é consideravelmente maior do que um
- Contudo, isto não quer dizer que a ANOVA é um teste bilateral