

Estudo de Caso 8: ANOVA

Analisaremos a planilha “acoes.rds”, que contém informações sobre o preço de três ações, ITUB3, PRIO3 e LWSA3 no decorrer de um ano. **Lembre-se de comentar e tirar conclusões dos exercícios realizados.**

1. Teste de Hipóteses e Teste de Equivariância

Após três cases, você já virou uma pessoa expert quando o assunto é fazer um teste de hipóteses no R. Iremos realizar esse teste novamente, porém utilizaremos um atalho muito útil para nossa vida profissional. Trata-se da função `t.test()`.¹

- a. Para realizar um teste de hipótese de médias (com variância desconhecida) podemos utilizar a função `t.test()`. Pesquise essa função no “help” do R e explique cada um dos parâmetros de entrada (atenção para o parâmetro `var.equal`).
- b. O que é o p-valor? O que ele representa?
- c. Agora que você já sabe tudo da função `t.test()`, faça um teste de hipóteses com o intuito de descobrir se a média dos valores da ação ITUB3 é igual a média dos valores da ação PRIO3 utilizando a função `t.test()`. O que podemos concluir através da ótica do p-value (p-valor)?

Podemos testar se a variância do preço de cada ação é igual. Uma maneira de cumprir essa tarefa é construir um teste F para cada par de ações. Outra forma é realizar um teste de equivariância, ou teste de Bartlett.

- d. O que é um teste F? Qual sua distribuição de probabilidades? É uma distribuição simétrica?
- e. Pesquise e descreva os argumentos da função `bartlett.test()`, responsável por realizar o teste de equivariância.
- f. Faça o teste de Bartlett ² para a base analisada. Quais conclusões podem ser tiradas, tanto do ponto de vista estatística quanto do ponto de vista físico?

¹ Existem outras funções similares aplicáveis a outros contextos, como `prop.test()` e `var.test()`. Recomenda-se a consulta desses atalhos para a praticidade de futuros projetos.

² Na última página deste documento há um trecho de código que pode ser útil para a aplicação do teste.

Agora analisaremos a planilha “tratamento.rds” que contém informações sobre um ensaio realizado para verificar como acontece o crescimento de determinada bactéria, quando submetida a 3 meios de cultura diferente (GF, GC e LJ) e a 3 tratamentos diferentes (1,2,3).
Lembre-se de comentar e tirar conclusões dos exercícios realizados.

2. ANOVA

Um determinado estudo está interessado em comparar o desenvolvimento da determinada bactéria. Procurou-se avaliar o melhor protocolo (tratamento e meio) que proporciona o melhor desenvolvimento. Inicialmente a análise será feita considerando o critério *meio de desenvolvimento*. Pretende-se entender se o meio em que é cultivada afeta o crescimento da bactéria. Os meios em análise são GF (gema de ovo fresca), GC (gema de ovo comercial) e LJ (Low Jeinsen). Considere um nível de significância de 5%.

- a. Construa o boxplot ³ de cada meio e informe também os valores de média, mediana, moda, desvio padrão e coeficiente de correlação.
- b. Escreva as hipóteses a serem testadas para esta primeira análise.
- c. Preencha a tabela da ANOVA para este caso.

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F teste	F tabela
Entre linhas					
Residual					
Total					

- d. Utilize as funções `aov()`, `summary(resultado_da_anova)` e `TukeyHSD(resultado_da_anova)` no R para realizar e analisar o teste ANOVA para a análise do crescimento médio entre os diferentes tratamentos utilizados. Explique os valores encontrados e conclua esta análise.

³ O resultado esperado é um plot contendo três boxplots representativos da distribuição dos valores de crescimento para cada meio analisado.

Em sequência, a análise será feita considerando o critério *tratamento*. Pretende-se entender se o tratamento dado afeta o crescimento da bactéria. Os tratamentos analisados são 1, 2 e 3. Considere um nível de significância de 5%.

- e. Construa o boxplot⁴ de cada tratamento e informe também os valores de média, mediana, moda, desvio padrão e coeficiente de correlação.
- f. Escreva as hipóteses a serem testadas para esta análise.
- g. Preencha a tabela da ANOVA para este caso.

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F teste	F tabela
Entre linhas					
Residual					
Total					

- h. Utilize as funções `aov()`, `summary(resultado_da_anova)` e `TukeyHSD(resultado_da_anova)` no R para realizar e analisar o teste ANOVA para a análise do crescimento médio entre os diferentes tratamentos utilizados. Explique os valores encontrados e conclua esta análise.

⁴ O resultado esperado é um plot contendo três boxplots representativos da distribuição dos valores de crescimento para cada tratamento analisado.

Por fim, a análise será feita considerando tanto o critério *meio de desenvolvimento* quanto o critério *tratamento*. Pretende-se agora verificar se o crescimento médio é igual quando se considera o tratamento e o meio. Considere um nível de significância de 5%.

- i. Quantos e quais são os fatores considerados nesta análise? Quantos e quais são os níveis de cada fator?
- j. Escreva as hipóteses a serem testadas.
- k. Preencha a tabela ANOVA para este caso.

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F teste	F tabela
Entre linhas					
Entre colunas					
Residual					
Total					

- l. Utilize as funções `aov()`, `summary(resultado_da_anova)` e `TukeyHSD(resultado_da_anova)` no R para realizar e analisar o teste ANOVA para a análise do crescimento médio entre os diferentes tratamentos utilizados. Explique os valores encontrados e conclua esta análise.

3. Entrega no Moodle.

Os cases devem ser enviados no e-disciplinas em um arquivo .pdf com o script do R anexo ao final do próprio PDF, de forma a possibilitar o Ctrl c, Ctrl v do mesmo para efeitos de correção.

Lembre-se de que dissertações e conclusões acerca dos resultados são mais importantes que a própria construção do código em R. Indique todos os resultados da maneira mais expositiva possível.

O prazo de entrega é domingo, 11/06, às 23h59.

OBS: Utilize o código abaixo, caso precise, para adequar seu banco de dados ao teste de Barlett. Explore cada resultado que esse código gera para entender a estrutura do banco de dados aplicável ao teste de Barlett. Esse código só funciona caso a biblioteca Tidyverse esteja ativa.

```
23 df_acoes3 <- df_acoes %>% select(PRI03)
24 df_acoes3$ação = "PRI03"
25 colnames(df_acoes3)[1]<-c("valor")
26
27 df_acoes4 <- df_acoes %>% select(LWSA3)
28 df_acoes4$ação = "LWSA3"
29 colnames(df_acoes4)[1]<-c("valor")
30
31 df_acoes_bartlett <- rbind(df_acoes2,df_acoes3,df_acoes4)
32
33 bartlett.test(df_acoes_bartlett$valor, df_acoes_bartlett$ação)
```