

Case 8 - Teste de Hipóteses: III

PRO3200 - Estatística

Camille Peixoto Almeida n°USP: 12702259

Isabela Belapetravicius n°USP: 11801971

11 de junho de 2023

O relatório a seguir se baseará em dois bancos de dados "acoes.rds" e "tratamento.rds" e com eles serão feitas análises estatísticas.

1 Teste de Hipótese e Teste de Equivariância

Para realizar um teste de hipótese de médias com variância desconhecida é possível usar a função `t.test()` da linguagem R.

1.1 Função `t.test()`

Essa função tem como parâmetros:

1. **x** e **y**: componentes de uma base de dados
2. **"Alternative"**: a hipótese alternativa (o padrão da linguagem R é se os estimadores são diferentes entre si, mas é possível testar para os casos em que o estimador 1 é maior ou menor que o estimador 2)
3. μ : valor verdadeiro da média (ou diferença de médias)
4. **"paired"**: indica se os dados são pareados ou não
5. **"var.equal"**: indica se as variâncias são admitidas iguais. Se `Var.equal = TRUE`, então a variância agrupada é usada para estimar a variância, caso contrário a aproximação de "Welch" para o cálculo dos graus de liberdade é usada.
6. **"conf.level"**: nível de confiança do intervalo.
7. **"na.action"**: função que indica o que deve ser feito quando a base de dados usada possui valores NA. O padrão usado é `getOption("na.action")`

1.2 P-valor

O valor p ou p-valor é uma medida estatística utilizada para avaliar a hipótese nula (H_0) em um teste de hipóteses. É a probabilidade de errar ao rejeitar a hipótese nula.

O teste de hipótese pode se apresentar em três configurações principais:

1.2.1 Configuração 1

$$\begin{aligned}H_0 : \mu &= A \\H_1 : \mu &\neq A\end{aligned}$$

A hipótese nula afirma que a média μ é igual a uma constante A, já a hipótese alternativa afirma que a média μ é diferente dessa mesma constante. Se for afirmado que o valor de μ é diferente de A, rejeita-se a hipótese nula com um nível de significância α escolhido e, conseqüentemente, o erro ao rejeitar H_0 é o p-valor.

1.2.2 Configuração 2

$$\begin{aligned}H_0 : \mu &= A \\H_1 : \mu &> A\end{aligned}$$

A hipótese nula afirma que a média μ é igual a uma constante A, já a hipótese alternativa afirma que a média μ é maior que essa mesma constante. Se for afirmado que o valor de μ é maior que A, rejeita-se a hipótese nula com um nível de significância α escolhido e, conseqüentemente, o erro ao rejeitar H_0 é o p-valor.

1.2.3 Configuração 3

$$\begin{aligned}H_0 : \mu &= A \\H_1 : \mu &< A\end{aligned}$$

A hipótese nula afirma que a média μ é igual a uma constante A, já a hipótese alternativa afirma que a média μ é menor que essa mesma constante. Se for afirmado que o valor de μ é menor que A, rejeita-se a hipótese nula com um nível de significância α escolhido e, conseqüentemente, o erro ao rejeitar H_0 é o p-valor.

Em resumo, o p-valor indica quão raro ou improvável seria obter os resultados observados se a hipótese nula fosse verdadeira ao rejeitá-la. Se o valor de p for muito pequeno (geralmente abaixo de um nível de significância predefinido α), considera-se evidência suficiente para rejeitar a hipótese nula em favor da hipótese alternativa.

Por outro lado, um valor de p alto indica que os dados não fornecem evidência suficiente para rejeitar a hipótese nula.

1.3 Teste de hipótese: médias ITUB3 e PRIO3

Deseja-se fazer um teste de hipótese para descobrir se a média da ação ITUB3 é igual a média dos valores da ação PRIO3, sendo assim:

$$\begin{aligned}H_0 : \mu_1 &= \mu_2 \\H_1 : \mu_1 &\neq \mu_2\end{aligned}$$

Em que:

1. μ_1 : média da ação ITUB3
2. μ_2 : média da ação PRIO3

Para as ações ITUB3 e PRIO3 do banco de dados "ações.rds" utilizou-se a função `t.test()` com variâncias desconhecidas e diferentes, ou seja, utilizou-se "var.equal = FALSE" e o cálculo de graus de liberdade por Welch foi usado, retornando:

Para um nível de significância de 5%, a função `t.test()` retorna um p-valor de 12.94%. Desse modo, como o p-valor é maior que o dobro do nível de significância escolhido, não existem evidências estatísticas suficientes para rejeitar a hipótese nula (H_0).

2 Teste de Hipótese F

Para populações que seguem uma distribuição normal, o teste de hipótese para a variância segue a **distribuição assimétrica de Fisher-Snedecor (distribuição F)** que é chamado de teste de hipótese F e é definido por:

$$F_{amostral} = F_{n_1-1, n_2-1} = \frac{S_1^2}{S_2^2} \quad (1)$$

e para $\alpha\%$ de significância:

$$F_{tabelado} = F_{crit} = F_{n_1-1, n_2-1, \alpha\%} \quad (2)$$

Como critério de decisão do teste de hipótese:

1. Se $F_{amostral} \geq F_{tabelado} = F_{crit}$, deve-se rejeitar a hipótese nula (H_0) com uma chance de erro de $\alpha\%$, ou também, se $p\text{-valor} < \alpha$.
2. Caso contrário, se $F_{amostral} \leq F_{tabelado} = F_{crit}$, não existem evidências estatísticas para afirmar que as variâncias são diferentes entre si, com um nível de significância de $\alpha\%$, ou também, se $p\text{-valor} > \alpha$.

3 Função `bartlett.test()`

A função `"bartlett.test"` é um teste estatístico da linguagem R para verificar a **homogeneidade de variâncias**, ou seja, testar se as variâncias de grupos independentes são iguais. Ele é utilizado em ANOVA, ou seja, em análises de variância. Como resultado, é apresentado o p-valor que é utilizado como critério de decisão do teste (rejeitar H_0 ou rejeitar H_1).

O teste de Bartlett para testar a hipótese de equi variância segue abaixo:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \dots = \sigma_n^2$$
$$H_1 : \sigma_i^2 \neq \sigma_n^2 \text{ (pelo menos uma das médias é diferente das outras)}$$

Em que:

1. σ_i^2 : variância do grupo i
2. σ_n^2 : variância do grupo n

A hipótese nula do teste consiste em afirmar que as variâncias são iguais em todos os grupos, enquanto a hipótese alternativa consiste em dizer que pelo menos um grupo tem uma variância diferente dos demais. Esse teste é importante para garantir que os resultados foram interpretados corretamente.

A função `bartlett.test()` recebe os seguintes parâmetros:

1. **x**: base de dados (vetor numérico de dados ou uma lista de vetores) que representa os objetos do modelo linear ajustados (herdados da classe `"lm"`)
2. **g**: vetor que fornece os elementos correspondentes a x, caso x não seja uma lista.
3. **"formula"**: é uma fórmula de `"lhs"` por `"rhs"`, em que `"lhs"` devolve os valores da base de dados e `"rhs"`, os valores correspondentes.
4. **"na.action"**: função que indica o que deve ser feito quando a base de dados usada possui valores NA. O padrão usado é `getOption("na.action")`

Utilizando o teste de Bartlett para os três grupos de ações (ITUB3, PRIO3 e LWSA3) tem-se como resultado o p-valor de $2.2 \cdot 10^{-16}$ que é realmente muito baixo em comparação com um nível de significância de 5% ou até mesmo 1%.

Isso significa que **se deve rejeitar a hipótese nula**, pois $p\text{-valor} < \alpha$ de significância.

De um ponto de vista "físico"/prático existe, de fato, uma forte evidência estatística pra aceitar a hipótese alternativa pois o p-valor é **extremamente** baixo, muito menor que qualquer nível de significância considerável na prática (5% ou 1%).

Em resumo, pode-se afirmar que pelo menos uma das ações possui uma variância diferente das demais.

Anteriormente, foi analisada o banco de dados "acoes.rds", a partir deste ponto será analisado o banco de dados "tratamento.rds" que contém um conjunto de dados sobre um ensaio realizado para verificar o crescimento de uma bactéria específica quando submetida a três meios de cultura diferentes (GF - gema de ovo fresca, GC - gema de ovo comercial e LJ - Low Jeinsen) e a 3 tratamentos diferentes (1,2,3).

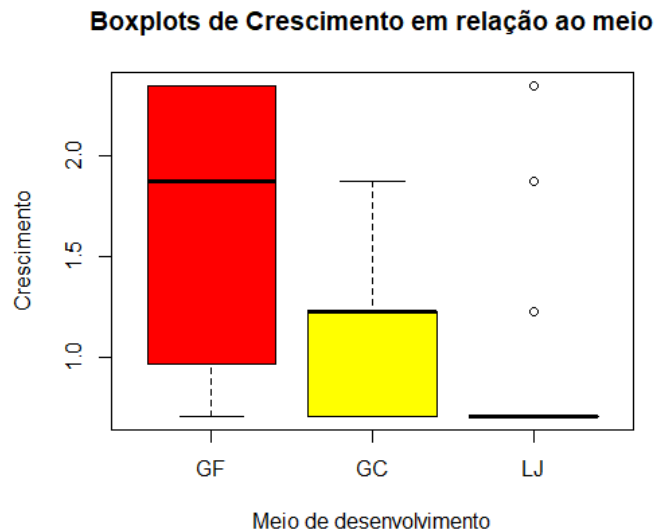
4 ANOVA - Análise de Variância

Com o objetivo de encontrar o melhor protocolo (tratamento e meio) que oferece o melhor desenvolvimento, utiliza-se do recurso estatístico de análise de variância (ANOVA). Desse modo, será possível comparar o desenvolvimento da bactéria para cada situação e entender se o meio em que ela é cultivada (GF, GC e LJ) afeta o seu crescimento. Para este estudo considera-se um nível de significância de 5%.

4.1 Meio

Em primeiro caso, para melhor visualização, é interessante a comparação do crescimento da bactéria para cada meio de cultura por meio da representação boxplot (figura 1):

Figura 1: Representação boxplot de cada meio



E as medidas estatísticas para cada meio estão abaixo na tabela:

	GF - gema de ovo fresca	GC - gema de ovo comercial	LJ - Low Jeinsen
Média	1.6550	1.1401	0.8087
Mediana	1.8708	1.2247	0.7071
Desvio padrão	0.6572	0.3350	0.3298
Coefficiente de Correlação			

Observação: não foi possível calcular o coeficiente de correlação de relação meio-crescimento, pois o banco de dados "tratamento.rds" possuía valores NA que não permitiram seu cálculo. Só foi possível calcular o coeficiente de correlação de GF x GC e GF x LJ:

$$R_{GF \times GC} = 0.05605 \text{ e } R_{GF \times LJ} = -0.01053$$

4.2 Teste de hipótese

O teste de hipótese utilizado na análise de variância (ANOVA) é usado para verificar se as médias de três ou mais grupos independentes são iguais (hipótese nula) ou não (hipótese alternativa):

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_i \neq \mu_j \text{ (pelo menos uma das três médias é diferente das demais)}$$

Em que:

1. μ_i : média do grupo i
2. μ_j : média do grupo j
3. Em que i e j podem ser 1, 2 ou 3

O teste F é usado para avaliar se há uma diferença significativa entre as médias dos grupos. Ele compara variância entre grupos com a variância dentro dos grupos.

Na prática, o teste F calcula a divisão entre a variância entre grupos e a variância dentro dos grupos, obtendo o valor $F_{calculado}$. Esse valor é comparado com o valor de $F_{tabelado}$ que é encontrado na tabela de Fisher-Snedecor para um nível de significância escolhido.

Se o valor de F calculado for maior que o valor de F tabelado, a hipótese nula deve ser rejeitada e, assim, pode-se afirmar que existe uma diferença significativa entre as médias dos grupos.

4.3 Tabela da ANOVA

Para os três meios foi analisado o crescimento da bactéria por meio da análise de variância que gerou a tabela abaixo:

Fonte de variação	Soma dos Quadrados	Graus de liberdade	Quadrado médio	F teste	F tabelado
Entre linhas	37.22	2	18.611	86.94	2.9957
Residual	65.08	304	0.214		
Total	102.3	306			

Da tabela acima, é possível observar que o $F_{teste}(F_{calculado})$ foi muito maior que o $F_{tabelado}$ para os graus de liberdade de $GL_{entregrupos} = 2$ e $GL_{residual} = 304$.

Isso significa que existem evidências estatísticas para rejeitar a hipótese nula, ou seja, existe uma diferença estatística entre as médias de crescimento da bactéria para pelo menos um dos meios de cultura.

Além disso, o p-valor calculado resultou na ordem de 10^{-6} que é muito menor que 0.05, logo, por outro modo de análise, se rejeita a hipótese inicial.

É importante ressaltar que a ANOVA apenas determina se há diferença entre as médias dos grupos, mas não indica especificamente qual grupo é diferente dos outros. Para identificar quais grupos diferem entre si, são realizados testes adicionais, como, por exemplo, o teste de Tukey (que na linguagem R, esse teste é feito por meio da função `tukeyHSD()`).

4.4 Teste de Tukey da Diferença Honestamente Significativa - Comparações múltiplas

Como foi verificado no cálculo da tabela de ANOVA, anteriormente, chegou-se a conclusão que pelo menos uma das médias foi diferente das demais, porém apenas com este teste F não foi possível esclarecer qual foi a média que se diferenciou das demais. Para isso se utiliza o teste de Tukey. Com ele será possível fazer comparações entre todos os pares.

Na linguagem R a função `TukeyHSD` mencionada retorna a matriz:

Tabela 1: Tukey

Meio	diff	lwr	upr	p adj
GF - GC	0.5150	0.3608	0.6691	0
LJ - GC	-0.3314	-0.4826	-0.1803	1.3E-06
LJ - GF	-0.8464	-0.9983	-0.6944	0

Em que:

1. "**diff**": diferença entre as médias observadas
2. "**lwr**": extremidade inferior do intervalo

3. "upr": extremidade superior do intervalo

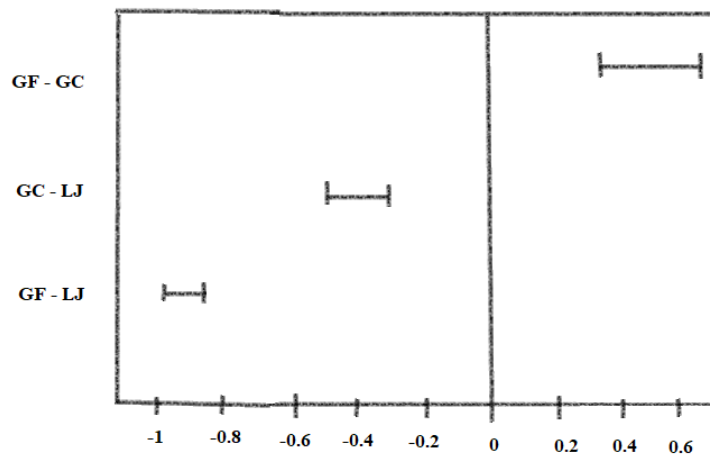
4. "p adj": p-valor depois do ajuste

Observação: os valores **lwr** e **upr** formam as extremidades do intervalo de confiança com 5% de significância.

Da tabela acima, vê-se que todos os valores de p ajustado são menores que o nível de significância 5% padrão da função `TukeyHSD()`. Portanto, é possível concluir que o crescimento médio da bactéria nos pares GF-GC, GF-LJ e GC-LJ são significativamente diferentes entre si.

Outra forma de analisar é que em nenhum dos intervalos de confiança o 0 está presente. A partir disso, novamente, conclui-se que o crescimento médio da bactéria nos pares GF-GC, GF-LJ e GC-LJ são significativamente diferentes entre si. Isso, na prática, significa que não existe nenhuma interseção entre os intervalos. Na figura a seguir, isso é melhor apresentado:

Figura 2: Intervalos de confiança dados pela função `TukeyHSD`

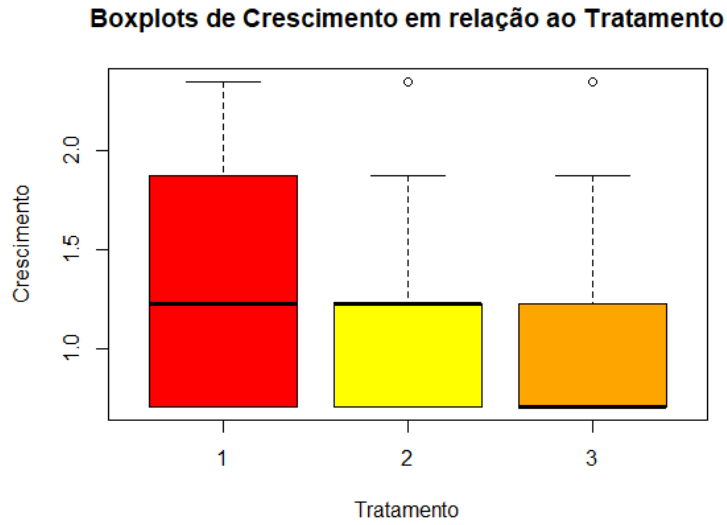


A partir deste ponto será analisada a resposta da bactéria a cada tipo de tratamento, novamente, por meio da análise de variância (ANOVA). Os tratamentos analisados são 1, 2 e 3 e considera-se um nível de significância de 5%.

4.5 Tratamento

Novamente, para melhor visualização, a representação boxplot para cada tratamento de bactéria segue abaixo na figura 2:

Figura 3: Representação boxplot de cada tratamento



E as medidas estatísticas para cada tratamento estão abaixo na tabela:

Tratamento	1	2	3
Média	1.2616	1.1740	1.1391
Mediana	1.2247	1.2247	0.7071
Desvio padrão	0.6062	0.5446	0.5801
Coefficiente de Correlação			

Os coeficientes de correlação são:

$$R_{trat.1Xtrat.3} = -0.3130 \text{ e } R_{trat.1Xtrat.2} = 0.5699$$

Novamente, o teste de hipótese para as médias de crescimento para cada tratamento será:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_1 : \mu_i \neq \mu_j \text{ (pelo menos uma das três médias é diferente das demais)}$$

Em que:

1. μ_i : média do grupo i
2. μ_j : média do grupo j
3. Em que i e j podem ser 1,2 ou 3

Para os três tratamentos foi analisado o crescimento da bactéria por meio da análise de variância que gerou a tabela abaixo:

Fonte de variação	Soma de Quadrados	Graus de liberdade	Quadrado médio	F teste	F tabelado
Entre linhas	0.77	1	0.7741	2.325	3.8415
Residual	101.52	305	0.3329		
Total	102.29	306			

Da tabela acima, é possível observar que o $F_{teste}(F_{calculado})$ menor que o $F_{tabelado}$ para os graus de liberdade de $GL_{entregrupos} = 1$ e $GL_{residual} = 305$.

Isso significa que não existem evidências estatísticas para rejeitar a hipótese nula, ou seja, existe não há uma diferença estatística entre as médias de crescimento da bactéria para pelo menos um dos tratamentos de cultura da bactéria.

Além disso, o p-valor calculado resultou em 0.128 que é maior que 0.005, desse modo, mais uma vez, não há evidência estatística para afirmar que existe diferença estatística entre as médias de crescimento da bactéria para pelo menos um dos tratamentos de cultura da bactéria.

Portanto, como se chegou à conclusão que não existe diferença no crescimento da bactéria quando submetida aos três tratamentos, não faz sentido utilizar o teste Tukey, pois não há média de grupos a comparar. Só para fins de testagem utilizou-se a função TukeyHSD() e como esperado o programa Rstudio imprimiu erro, uma vez que não existe diferença entre as médias com um nível de significância de 5%.

4.6 Desenvolvimento - tratamento e meio

Por fim, será analisado por meio de uma tabela de análise de variância (ANOVA) de **dois fatores** considerando o meio de desenvolvimento (fator 1: meio e fator 2: tratamento) com o objetivo de verificar se o crescimento médio é igual quando se consideram o meio e o tratamento juntos considerando 5% de nível de significância.

Para a tabela de ANOVA que será feita haverão 2 fatores: meio e tratamento. O meio possui três níveis: GF - gema de ovo fresca, GC - gema de ovo comercial e LJ - Low Jeinsen. Já o tratamento possui também outros três níveis: 1, 2 e 3.

No caso de análise de variância de dois fatores sem réplica duas hipóteses nulas podem ser testadas:

Teste de hipótese 1 - meio (A)

$$H_{0A} : \mu_{A_1} = \mu_{A_2} = \mu_{A_3} = \dots = \mu_{A_n} =$$

$$H_{1A} : \text{existe pelo menos um par } \mu_{A_i} \neq \mu_{A_j} \text{ diferente dos demais}$$

Teste de hipótese 2 - tratamento (B)

$$H_{0_B} : \mu_{B_1} = \mu_{B_2} = \mu_{B_3} = \dots = \mu_{B_n} =$$

$$H_{1_B} : \text{existe pelo menos um par } \mu_{B_i} \neq \mu_{B_j} \text{ diferente dos demais}$$

Desse modo, a tabela ANOVA com 2 fatores é gerada:

Fonte de variação	Soma de Quadrados	Graus de liberdade	Quadrado médio	F teste	F tabelado
Entre linhas	37.22	2	18.611	87.077	3.6889
Entre colunas	0.57	1	0.573	2.681	5.0239
Residual	64.33	301	0.214		
Total	102.12	304			

Escolhido um nível de significância α de 5%, o critério de decisão é dado por:

1. Rejeita-se H_{0_A} se $F_{A_{calculado}} > F_{A_{tabelado}}$
2. Rejeita-se H_{0_B} se $F_{B_{calculado}} > F_{B_{tabelado}}$

Da tabela acima, verifica-se que, "entre linhas- meio, o $F_{A_{calculado}}$ foi muito maior que o $F_{A_{tabelado}}$ para os graus de liberdade $GL_{entrelinhas} = 2$ e $GL_{residual} = 301$. Consequentemente, é possível afirmar que existem evidências estatísticas para rejeitar a hipótese nula (H_{0_A}), ou seja, existe uma diferença relativa entre as médias de crescimento da bactéria para pelo menos um dos meios de cultura.

Por outro lado, verifica-se que, "entre colunas- tratamento, o $F_{B_{calculado}}$ foi menor que o $F_{B_{tabelado}}$ para graus de liberdade $GL_{entrecolunas} = 1$ e $GL_{residual} = 301$. Dessa forma, não existem evidências estatísticas para rejeitar a hipótese nula (H_{0_B}), ou seja, não é possível afirmar com 5% de significância que existe uma diferença relativa entre as médias de crescimento da bactéria quando comparados os três tratamentos.

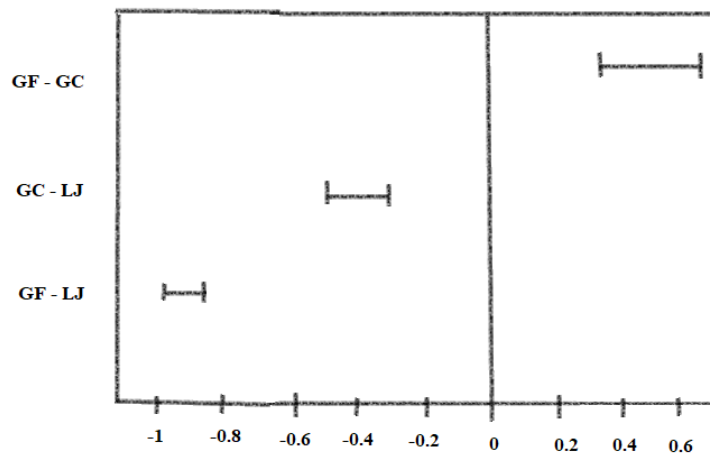
Logo, como foi concluído que houve diferença significativa entre as médias para o meio, faz sentido utilizar, novamente, o teste de Tukey:

Tabela 2: Tukey

Meio	diff	lwr	upr	p adj
GF - GC	0.5150	0.3608	0.6691	0
LJ - GC	-0.3314	-0.4826	-0.1803	1.3E-06
LJ - GF	-0.8464	-0.9983	-0.6944	0

E, novamente, a representação dos intervalos apresenta-se:

Figura 4: Intervalos de confiança dados pela função TukeyHSD



Novamente, em nenhum dos intervalos de confiança o 0 está presente. A partir disso, conclui-se que o crescimento médio da bactéria nos pares GF-GC, GF-LJ e GC-LJ são significativamente diferentes entre si. Isso, na prática, significa que não existe nenhuma interseção entre os intervalos.

Finalmente, o tipo de meio influencia no crescimento da bactéria, porém o tratamento não diferencia esse crescimento.

5 Script - Case 8

```
library(dplyr)

library("tidyverse")

df_acoes <- readRDS("H:/Meu Drive/USP/semestres_passados/1ºQuadri2023/reof_estat/
Estudo de Caso 8 - ANOVA-20230531/Case8/acoes.rds")

## TESTE DE MÉDIAS

help("t.test")
# parâmetros
# x e y: bases de dados
# mu: valor verdadeiro da média (ou da diferença e médias)
# paired: indica se os dados são pareados ou não
# var.equal: indica se as variâncias são admitidas iguais
# conf.level: nível de significância
```

```

t.test(df_acoes$ITUB3,df_acoes$PRI03)

# o p-valor é a área à direita a partir do valor amostral na distribuição t-student e
# representa a probabilidade de errar ao rejeitar H0

# H0: médias iguais
# H1: médias diferentes

# p-valor = 12,94%
# para 5% de sigficância => NÃO REJEITO H0

## TESTE DE VARI NCIAS

# o teste F é utilizado em testes de variâncias e sua distribuição é a F-Snedecor,
# que não é simétrica

help("bartlett.test")
# parâmetros
# x: base de dados
# g: indica os elementos correspondentes a x, caso este não seja uma lista
# formula: uma fórmula de "lhs" por "rhs", em que "lhs" devolve os valores da base
# de dados e "rhs", os valores correspondentes

# H0: variâncias homogêneas (iguais)
# H1: variâncias não homogêneas (diferentes)

df_acoes2 <- df_acoes %>% select(ITUB3)
df_acoes2$ação = "ITUB3"
colnames(df_acoes2) [1]<-c("valor")

df_acoes3 <- df_acoes %>% select(PRI03)
df_acoes3$ação = "PRI03"
colnames(df_acoes3) [1]<-c("valor")

df_acoes4 <- df_acoes %>% select(LWSA3)
df_acoes4$ação = "LWSA3"
colnames(df_acoes4) [1]<-c("valor")

df_acoes_bartlett<-rbind(df_acoes2,df_acoes3,df_acoes4)

bartlett.test(df_acoes_bartlett$valor,df_acoes_bartlett$ação)

# K-squared = 686.95
# p-value < 2.2e-16
# para 5% de sigficância => REJEITO H0, AFIRMO H1

```

```
## ANOVA UM FATOR
```

```
df_tratamento <- readRDS("H:/Meu Drive/USP/semestres_passados/1ºQuadri2023/reof_estat/  
Estudo de Caso 8 - ANOVA-20230531/Case8/tratamento.rds")  
df_tratamento <- na.omit(df_tratamento)  
# meio
```

```
meio_gf <- subset(df_tratamento, df_tratamento$MEIO == "GF")  
meio_gc <- subset(df_tratamento, df_tratamento$MEIO == "GC")  
meio_lj <- subset(df_tratamento, df_tratamento$MEIO == "LJ")
```

```
meio_gf <- na.omit(meio_gf)  
meio_gc <- na.omit(meio_gc)  
meio_lj <- na.omit(meio_lj)
```

```
boxplot(meio_gf$CRESCIMENTO,  
        meio_gc$CRESCIMENTO,  
        meio_lj$CRESCIMENTO,  
        names=c("GF", "GC", "LJ"),  
        xlab="Meio de desenvolvimento",  
        ylab="Crescimento",  
        main="Boxplots de Crescimento em relação ao meio",  
        col= c("red", "yellow", "orange"))
```

```
mode <- function(x) {  
  u <- unique(x)  
  tab <- tabulate(match(x, u))  
  u[tab == max(tab)]  
}
```

```
mediaGF <- mean(meio_gf$CRESCIMENTO) # média = 1.655039  
medianaGF <- median(meio_gf$CRESCIMENTO) # mediana = 1.870829  
desvioGF <- sd(meio_gf$CRESCIMENTO) # desvio = 0.6571736  
modaGF <- mode(meio_gf$CRESCIMENTO) # moda = 2.345208
```

```
mediaGC <- mean(meio_gc$CRESCIMENTO) # média = 1.40085  
medianaGC <- median(meio_gc$CRESCIMENTO) # mediana = 1.224745  
desvioGC <- sd(meio_gc$CRESCIMENTO) # desvio = 0.3349757  
modaGC <- mode(meio_gc$CRESCIMENTO) # moda = 1.224745
```

```
mediaLJ <- mean(meio_lj$CRESCIMENTO) # média = 0.808651  
medianaLJ <- median(meio_lj$CRESCIMENTO) # mediana = 0.707107  
desvioLJ <- sd(meio_lj$CRESCIMENTO) # desvio = 0.3298223  
modaLJ <- mode(meio_lj$CRESCIMENTO) # moda = 0.707107
```

```

cor(df_tratamento$GF, df_tratamento$GC)
cor(df_tratamento$GF, df_tratamento$LJ)

help(aov)

resultado_anova <- aov(formula = CRESCIMENTO ~ MEIO,data = df_tratamento)

summary(resultado_anova)

help("TukeyHSD")

TukeyHSD(resultado_anova)

#####
# Análise de variância - ANOVA - Tratamento

tratamento_1 <- subset(df_tratamento, df_tratamento$TRAT == 1)
tratamento_2 <- subset(df_tratamento, df_tratamento$TRAT == 2)
tratamento_3 <- subset(df_tratamento, df_tratamento$TRAT == 3)

boxplot(tratamento_1$CRESCIMENTO, tratamento_2$CRESCIMENTO, tratamento_3$CRESCIMENTO,
names=c("1","2","3"), xlab="Tratamento", ylab="Crescimento", main="Boxplots de
Crescimento em relação ao Tratamento", col=c("red","yellow","orange"))

media_trat_1 <- mean(tratamento_1$CRESCIMENTO) # média = 1.261598
mediana_trat_1 <- median(tratamento_1$CRESCIMENTO) # mediana = 1.224745
desvio_trat_1 <- sd(tratamento_1$CRESCIMENTO) # desvio = 0.6061808
moda_trat_1 <- mode(tratamento_1$CRESCIMENTO) # moda = 0.707107

media_trat_2 <- mean(tratamento_2$CRESCIMENTO) # média = 1.173958
mediana_trat_2 <- median(tratamento_2$CRESCIMENTO) # mediana = 1.224745
desvio_trat_2 <- sd(tratamento_2$CRESCIMENTO) # desvio = 0.5445513
moda_trat_2 <- mode(tratamento_2$CRESCIMENTO) # moda = 0.707107

media_trat_3 <- mean(tratamento_3$CRESCIMENTO) # média = 1.139064
mediana_trat_3 <- median(tratamento_3$CRESCIMENTO) # mediana = 0.707107
desvio_trat_3 <- sd(tratamento_3$CRESCIMENTO) # desvio = 0.5801029
moda_trat_3 <- mode(tratamento_3$CRESCIMENTO) # moda = 0.707107

cor(tratamento_1$CRESCIMENTO, tratamento_3$CRESCIMENTO)

tratamento_1_novo <- slice(tratamento_1, -c(34, 32))
cor(tratamento_1_novo$CRESCIMENTO, tratamento_2$CRESCIMENTO)

```

```

resultado_anova_tratamento_1fator <- aov(CRESCIMENTO ~ TRAT, df_tratamento)
summary(resultado_anova_tratamento_1fator)
TukeyHSD(resultado_anova_tratamento_1fator)

## ANOVA DOIS FATORES

# meio e tratamento

resultado_anova_tratamento_2fatores <- aov(CRESCIMENTO ~ MEIO*TRAT, df_tratamento)
summary(resultado_anova_tratamento_2fatores)
TukeyHSD(resultado_anova_tratamento_2fatores)

```

6 Referências

1. RAMOS, Alberto. Apostila de Estatística-PRO3200. Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Produção, São Paulo.2021
2. HO, Linda Lee; RIBEIRO, Celma de Oliveira. Intervalo de confiança.PRO3200 - Estatística, Departamento de Engenharia e Produção, Universidade de São Paulo.2022.
3. Teste de Tukey para Comparações Múltiplas. Statplace.
Disponível em: <https://statplace.com.br/blog/comparacoes-multiplas-teste-de-tukey/>
Acesso em: 07/06/2023
4. F Distribution Tables.
Disponível em: https://socr.umich.edu/Applets/F_Table.html
Acesso em: 07/06/2023
5. Tutorial ANOVA. RPubS by RSstudio.
Disponível em: https://rpubs.com/dangs12/tutorial_anova
Acesso em: 07/06/2023
6. PERES, Fernanda, ANOVA de duas vias no R (Parte 2).
Disponível em: <https://youtu.be/16xKsd0eo-Y>
Acesso em: 07/06/2023
7. PERES, Fernanda, ANOVA de duas vias no R (Parte 1).
Disponível em: <https://www.youtube.com/watch?v=f4gaMVxJVpc>
Acesso em: 07/06/2023