# 4º Exercício Avaliativo

# Planejamento de experimentos

Ariane Hayana Thomé de Farias

João Claudio da Silva Araujo Lobato

17/11/2021

**Questão 14.1) Um fabricante de propelente de foguete está estudando a taxa de combustão do propelente de três processos de produção. Quatro lotes de propelente são selecionados aleatoriamente a partir da saída de cada processo, e três determinações de taxa de queima são feitas em cada lote. Os resultados seguem. Analise os dados e tire conclusões.**

| | Processo 1 | | | | Processo 2 | | | | Processo 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Lote** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** |
| | 25 | 19 | 15 | 15 | 19 | 23 | 18 | 35 | 14 | 35 | 38 | 25 |
| | 30 | 28 | 17 | 16 | 17 | 24 | 21 | 27 | 15 | 21 | 54 | 29 |
| | 26 | 20 | 14 | 13 | 14 | 21 | 17 | 25 | 20 | 24 | 50 | 33 |

**Respostas:**

> Estatística Descritiva

Para o entendimento das principais medidas descritivas do problema em estudo, faz-se uma breve análise descritiva dos dados:

Show 25 entries

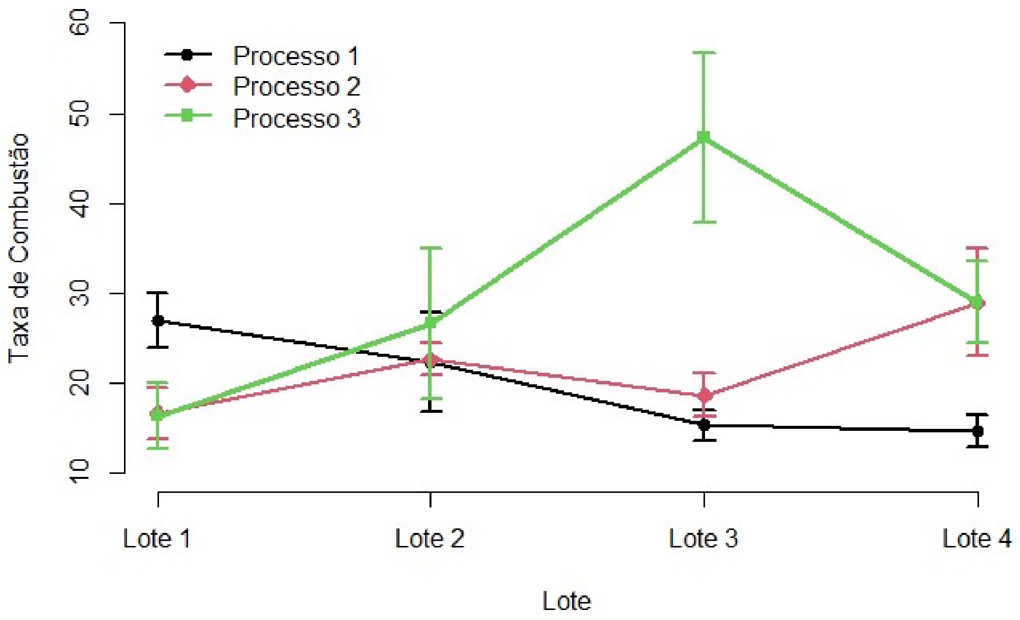
| Processo | Lote | Média | DP | Variância | Mínimo | Máximo | CV |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 27 | 2.65 | 7 | 25 | 30 | 9.8 |
| 1 | 2 | 22.33 | 4.93 | 24.33 | 19 | 28 | 22.09 |
| 1 | 3 | 15.33 | 1.53 | 2.33 | 14 | 17 | 9.96 |
| 1 | 4 | 14.67 | 1.53 | 2.33 | 13 | 16 | 10.41 |
| 2 | 1 | 16.67 | 2.52 | 6.33 | 14 | 19 | 15.1 |
| 2 | 2 | 22.67 | 1.53 | 2.33 | 21 | 24 | 6.74 |
| 2 | 3 | 18.67 | 2.08 | 4.33 | 17 | 21 | 11.15 |
| 2 | 4 | 29 | 5.29 | 28 | 25 | 35 | 18.25 |
| 3 | 1 | 16.33 | 3.21 | 10.33 | 14 | 20 | 19.68 |

| Processo | Lote | Média | DP | Variância | Mínimo | Máximo | CV |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 26.67 | 7.37 | 54.33 | 21 | 35 | 27.64 |
| 3 | 3 | 47.33 | 8.33 | 69.33 | 38 | 54 | 17.59 |
| 3 | 4 | 29 | 4 | 16 | 25 | 33 | 13.79 |

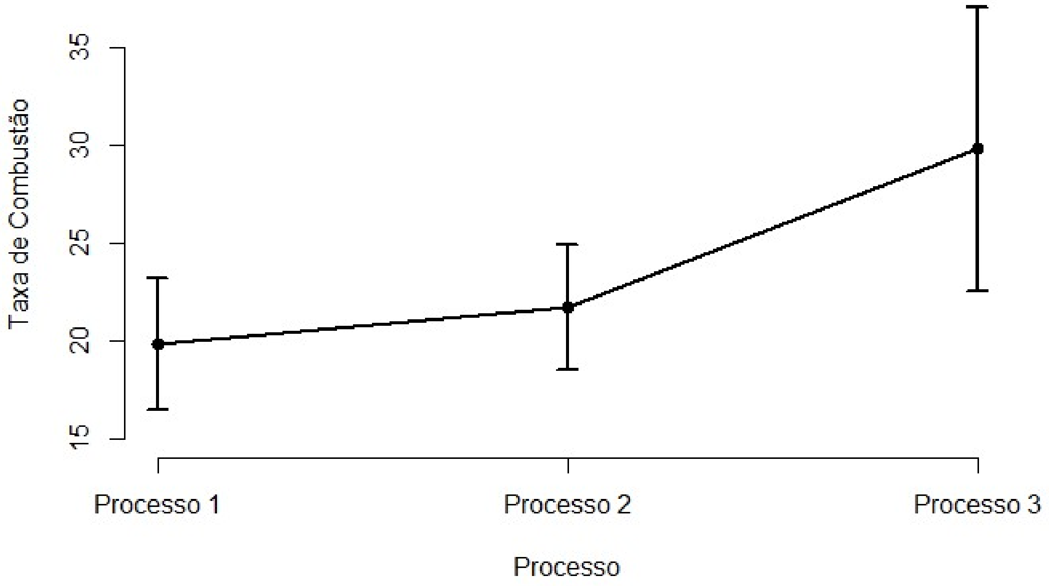
Showing 1 to 12 of 12 entries

Previous 1 Next

Neste cenário, objetiva-se um estudo relacionado a taxa de combustão do propelente de `3 processos de produção`, os quais `4 lotes` do propelente são selecionados de forma aleatória e `3 determinações de taxa de queima` são realizadas. Com base nestas informações, é possível verificar que a média em cada processo e lote varia entre $14,67$ e $47,33$. No que se refere ao desvio padrão, nota-se que o valor mais alto é encontrado no `Processo 3` e `Lote 3`, em que o mesmo corresponde a $8,33$ e variância de $69,33$.
Por outro lado, temos na tabela acima o coeficiente de variação (CV), que é a razão entre o desvio padrão e a média amostral expresso em porcentagem. Assim, quanto menor for o coeficiente de variação, mais homogêneo será o grupo analisado, o que implica em uma menor dispersão em torno da média. Portanto, considerando as informações da tabela é possível observar que no `Processo 2` e `Lote 2` é onde encontra-se o menor CV, que é classificado como `Baixo`, tendo em vista que seu CV é menor que $10\%$. A maioria dos processos e lotes se enquadram na classificação de `Médio` CV, com resultados entre $10\% <= CV < 20\%$ e o valor mais alto corresponde ao CV do `Processo 3` e `Lote 2`, com $27,64\%$.




Acima temos o gráfico com intervalos de confiança para a média de cada Lote por Processo e através dos intervalos vemos que as médias não possuem um comportamento linear e nota-se indícios de uma variabilidade nas médias de todos os lotes, mas de forma mais atenuada nos lotes `1` e `3`.




Com o gráfico acima podemos ver a média da taxa de combustão para cada Processo, juntamente com o seu intervalo de confiança. É interessante vermos que o valor médio parece aumentar conforme o número do Processo. No `Processo 1` temos uma taxa média de combustão de $19,83$ e já no `Processo 3` temos uma média de $29,83$.

## Teste de hipóteses

Agora, testaremos a hipótese de que, ao nível de significância de $5\%$:

$$H_0 : \beta_{j(i)} = 0$$

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

$$H_1 : \text{Pelo menos um} \beta_{j(i)0} \neq 0$$

$$H_1 : \text{Pelo menos um } \tau_i \neq 0$$

## Anova Efeito Fixo de forma operacional

Aqui iremos realizar os cálculos da tabela ANOVA de modo operacional. Para tanto, os cálculos foram elaborados utilizando o software `R`, através do algoritmo a seguir:

```
# ---------------------- SQ_Total

soma_1 <- sum(resp^2);soma_1
soma_2 <- (sum(resp))^2;soma_2
a <- 3
b <- 4
n <- 3

SQ_total <- soma_1 - (soma_2/(a*b*n));SQ_total
```

```r
# --------------------- SQ_Processo

proc1 <- matrix(c(25,19,15,15,30,28,17,16,26,20,14,13), nrow = 3, ncol = 4, byrow = T)
proc2 <- matrix(c(19,23,18,35,17,24,21,27,14,21,17,25), nrow = 3, ncol = 4, byrow = T)
proc3 <- matrix(c(14,35,38,25,15,21,54,29,20,24,50,33), nrow = 3, ncol = 4, byrow = T)

Totais_lote_proc1 <- apply(proc1,2,sum)
Totais_lote_proc2 <- apply(proc2,2,sum)
Totais_lote_proc3 <- apply(proc3,2,sum)

Totais_fornecedor_proc1 <- sum(Totais_lote_proc1)
Totais_fornecedor_proc2 <- sum(Totais_lote_proc2)
Totais_fornecedor_proc3 <- sum(Totais_lote_proc3)

soma_3 <- ((Totais_fornecedor_proc1^2)+
             (Totais_fornecedor_proc2^2) + (Totais_fornecedor_proc3^2))/(b*n)

SQ_processo <- soma_3 - (soma_2/(a*b*n));SQ_processo

gl_proc <- a-1;gl_proc

QM_proc <- SQ_processo/gl_proc;QM_proc


# --------------------- SQ_Lotes

soma_4 <- sum((Totais_lote_proc1^2)+(Totais_lote_proc2^2)+ (Totais_lote_proc3^2))/n

SQ_lotes <- soma_4 - soma_3;SQ_lotes

gl_lotes <- a*(b-1);gl_lotes

QM_lotes <- SQ_lotes/gl_lotes;QM_lotes

# --------------------- SQ_res

SQ_res <- (soma_1 - soma_4);SQ_res

gl_res <- a*b*(n-1);gl_res

QM_res <- SQ_res/gl_res;QM_res
```

Após realizadas todas as operações, os resultados obtidos foram dispostos na tabela abaixo:

| Causas de Variação | SQ | GL | QM |
|---|---|---|---|
| Processo | 676,06 | 2 | 338,03 |
| Processo:Lote | 2077,58 | 9 | 230,84 |
| Resíduos | 454 | 24 | 18,92 |

Para validar os resultados, fez-se também a tabela ANOVA utilizando a função do `R`, conforme próximo tópico.

### Anova Efeito Fixo através da função

Na etapa seguinte, utilizando a função `aov` do `R`, teremos os seguintes resultados para a análise de variância (ANOVA):

```
## Analysis of Variance Table
##
## Response: resp
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## processo       2  676.06  338.03  17.869 1.768e-05 ***
## processo:Lote  9 2077.58  230.84  12.203 5.477e-07 ***
## Residuals     24  454.00   18.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nesta etapa foram realizadas algumas alterações para corrigir o valor de $F_0$ e do $valor - p$, que nos resultados acima aparecem incorretos. Assim, após correções, a ANOVA corrigida será:

```
## Analysis of Variance Table
##
## Response: resp
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## processo       2  676.06  338.03  1.4643    0.2815
## processo:Lote  9 2077.58  230.84 12.2031 5.477e-07 ***
## Residuals     24  454.00   18.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Portanto, considerando o nosso $\alpha$ de 5%, não observamos um efeito significativo na taxa média de combustão entre os 3 diferentes tipos de Processos, sendo o seu valor p obtido de 0,2815, então **rejeitamos a hipótese nula**. Mas quando analisamos os 4 diferentes Lotes do mesmo Processo, vemos que há sim uma diferença significativa em pelo menos 1 dos Lotes, apresentando uma combustão média diferente dos demais com seu respectivo valor p de <0,000, ou seja, **rejeitamos** $H_0$.

### Estimativas

Abaixo temos as estimativas do efeito fixo:

```
##
## Call:
## lm(formula = resp ~ processo/Lote)
##
## Coefficients:
##     (Intercept)        processo2        processo3  processo1:Lote2
##          27.000          -10.333          -10.667           -4.667
## processo2:Lote2  processo3:Lote2  processo1:Lote3  processo2:Lote3
##           6.000           10.333          -11.667            2.000
## processo3:Lote3  processo1:Lote4  processo2:Lote4  processo3:Lote4
##          31.000          -12.333           12.333           12.667
```

> Análise dos Resíduos

Agora, testamos 3 pressupostos necessários para a validação da tabela da ANOVA que vimos acima, são eles:

- Normalidade dos Resíduos

Para o teste de *Normalidade dos Resíduos*, utilizamos a função `shapiro.test()`, a qual testaremos a hipótese de normalidade dos resíduos. Caso o valor-p seja maior que 0,05, aceitaremos a hipótese nula de que os resíduos seguem uma distribuição normal.

```
## 
##  Shapiro-Wilk normality test
## 
## data:  resid(mod)
## W = 0.9765, p-value = 0.6265
```

- Homogeneidade de Variância

Para o teste de *Homogeneidade de Variâncias*, utilizamos a função `leveneTest()`. Caso o valor-p seja maior que 0,05, aceitaremos a hipótese nula de que os dados em estudo apresentam homogeneidade de variância.

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group 11   0.585 0.8222
##       24
```

- Independência

Para verificar se existe algum tipo de correlação (autocorrelação) nos resíduos utiliza-se o `Teste de Durbin-Watson`. Caso o valor-p seja maior que 0,05, aceitaremos a hipótese nula de que nos dados não existe autocorrelação, portanto, podemos afirmar que os resíduos não são autocorrelacionados.

```
## 
##  Durbin-Watson test
## 
## data:  mod
## DW = 2.3578, p-value = 0.1747
## alternative hypothesis: true autocorrelation is greater than 0
```

Note que, em todos os testes apresentados *aceitamos* $H_0$, ou seja, os nossos dados apresentam homogeneidade de variância, os resíduos não são autocorrelacionados e possuem distribuição normal.