2º Exercício Avaliativo
Planejamento de experimentos

Ariane Hayana Thomé de Farias
João Claudio da Silva Araujo Lobato

13/10/2021

Exercícios do Montgomery, D. C. (2017).
Design and Analysis of Experiments, Ninth Edition

4.21) Supondo que os tipos de produtos químicos e parafusos sejam fixos, estime os parâmetros do modelo τ_i e β_j no `Problema 4.8`.

Respostas:

$$Parafuso$$

Químico	1	2	3	4	5	$y_{i.}$	$\bar{y}_{i.}$
1	73	68	74	71	67	353	70,6
2	73	67	75	72	70	357	71,4
3	75	68	78	73	68	362	72,4
4	73	71	75	75	69	363	72,6
$y_{.j}$	294	274	302	291	274		
$\bar{y}_{.j}$	73,5	68,5	75,5	72,75	68,5		$\bar{y}_{..} =$ 71,75

Utilizando a tabela acima, temos que:

$\hat{\tau_1} = \bar{y_{1.}} - \bar{y_{..}} = 70,6 - 71,75 = -1,15$ ou $\frac{-23}{20}$

$\hat{\tau_2} = \bar{y_{2.}} - \bar{y_{..}} = 71,4 - 71,75 = -0,35$ ou $\frac{-7}{20}$

$\hat{\tau_3} = \bar{y_{3.}} - \bar{y_{..}} = 72,4 - 71,75 = 0,65$ ou $\frac{13}{20}$

$\hat{\tau_4} = \bar{y_{4.}} - \bar{y_{..}} = 72,6 - 71,75 = 0,85$ ou $\frac{17}{20}$

$\beta_1 = \bar{y_{.1}} - \bar{y_{..}} = 73,5 - 71,75 = 1,75$ ou $\frac{35}{20}$

$\beta_2 = \bar{y_{.2}} - \bar{y_{..}} = 68,5 - 71,75 = -3,25$ ou $\frac{-65}{20}$

$\beta_3 = \bar{y_{.3}} - \bar{y_{..}} = 75,5 - 71,75 = 3,75$ ou $\frac{75}{20}$

$\beta_4 = \bar{y_{.4}} - \bar{y_{..}} = 72,75 - 71,75 = 1$ ou $\frac{20}{20}$

$\beta_5 = \bar{y_{.5}} - \bar{y_{..}} = 68,5 - 71,75 = -3,25$ ou $\frac{-65}{20}$

4.26) Um engenheiro industrial está conduzindo um experimento sobre o tempo de foco do olho. Ele está interessado no efeito da distância do objeto do olho no tempo de foco. Quatro distâncias diferentes são de interesse. Ele tem cinco indivíduos disponíveis para o experimento. Como pode haver diferenças entre os indivíduos, ele decide conduzir o experimento em um delineamento de blocos ao acaso. Seguem os dados obtidos. Analise os dados deste experimento e tire as conclusões apropriadas (use $\alpha = 0,05$).

$$Indivíduo$$

Distância	1	2	3	4	5
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

Respostas:

Nesta questão utilizaremos o `delineamento de blocos ao acaso` e é possível verificar que existem 4 distâncias de tratamento (`t1 = 4`, `t2 = 6`, `t3 = 8` e `t4 = 10`) e 5 blocos, portanto, $a = 4$ e $b = 5$.

Para calcular as Somas de Quadrados (Total, Tratamento e Resíduos), é preciso conhecer os resultados das médias dos tratamentos, bem como a média geral de todas as distâncias e blocos. Nesta etapa, utilizou-se o software estatístico `R` e obteve-se os seguintes resultados:

> *Média dos tratamentos*

t1	t2	t3	t4
6,8	5,20	3,60	3,80

> *Média dos blocos*

I	II	III	IV	V
7,00	4,75	4,75	2,75	5,00

Antes de aplicarmos a ANOVA, checaremos alguns **pressupostos** importantes para a continuidade das nossas análises, tais como:

- 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(fit)
## W = 0.94976, p-value = 0.3634
```

- 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  3  0.0539 0.9829
##       16
```

- 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:  fit
## DW = 2.4353, p-value = 0.6247
## 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.

> *Teste de hipóteses*

Agora, testaremos a hipótese de que, ao nível de significância de 0,05, os tratamentos para as 4 distâncias são iguais ou se existe uma diferença entre eles. Isto implica que queremos testar se:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

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

O mesmo é válido para os blocos, portanto, o teste será:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

O passo seguinte consiste em realizar a análise de variância e, para tanto, será necessário inicialmente encontrar os valores da Soma de Quadrado de Tratamento (SQTrat), Soma de Quadrado de Resíduo (SQRes) e Soma de Quadrado Total (SQTot) e seus respectivos graus de liberdade. Posteriormente, calcula-se o Quadrado Médio dos Tratamentos (QMTrat), Quadrado Médio dos Resíduos (QMR) e o Quadrado Médio Total (QMT). Na etapa final, calcula-se o valor da estatística F e o valor-p para que possamos testar as hipóteses.

Para esta etapa optou-se por utilizar a função `aov` do `R`, obtemos os seguintes resultados para a análise de variância (ANOVA):

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## trat         3  32.95  10.983   8.614 0.00254 **
## bloco        4  36.30   9.075   7.118 0.00355 **
## Residuals   12  15.30   1.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Portanto, considerando um $\alpha = 0,05$ e o valor-p obtido de $0,00254$ para os `tratamentos`, podemos concluir que $valor - p < \alpha$, portanto, **rejeitamos a hipótese nula**, o que implica dizer que há evidências de que pelo menos um tratamento difere dos demais nas diferentes distâncias. O mesmo ocorre com o teste para `blocos`, ou seja, com um valor-p de $0,00355 < \alpha$, também existe uma diferença significativa, o que implica dizer que há a necessidade de todas as vezes que se fizer o experimento da mesma natureza que o atual, deve-se utilizar o delineamento em blocos ao acaso.

Com base nos resultados apresentados anteriormente, os quais vimos que há evidências de diferença significativas tanto nos tratamentos quanto nos blocos, abordaremos agora quais diferem entre si utilizando o `teste de Tukey`.

Para o teste supracitado são consideradas as seguintes hipóteses:

$H_0 : \mu_i = \mu_k$

H_1 : Pelo menos um par das médias é desigual

A mesma ideia é válida para os blocos, portanto, o resultado do teste utilizando a função `TukeyHSD()` é dado abaixo:

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = resp ~ trat + bloco)
##
## $trat
##       diff       lwr        upr     p adj
## t2-t1 -1.6 -3.720219  0.5202195 0.1675242
## t3-t1 -3.2 -5.320219 -1.0797805 0.0036156
## t4-t1 -3.0 -5.120219 -0.8797805 0.0058370
## t3-t2 -1.6 -3.720219  0.5202195 0.1675242
## t4-t2 -1.4 -3.520219  0.7202195 0.2554374
## t4-t3  0.2 -1.920219  2.3202195 0.9919294
##
## $bloco
##      diff        lwr        upr     p adj
## 2-1 -2.25 -4.7949606  0.2949606 0.0928513
## 3-1 -2.25 -4.7949606  0.2949606 0.0928513
## 4-1 -4.25 -6.7949606 -1.7050394 0.0013855
## 5-1 -2.00 -4.5449606  0.5449606 0.1536619
## 3-2  0.00 -2.5449606  2.5449606 1.0000000
## 4-2 -2.00 -4.5449606  0.5449606 0.1536619
## 5-2  0.25 -2.2949606  2.7949606 0.9976084
## 4-3 -2.00 -4.5449606  0.5449606 0.1536619
## 5-3  0.25 -2.2949606  2.7949606 0.9976084
## 5-4  2.25 -0.2949606  4.7949606 0.0928513
```

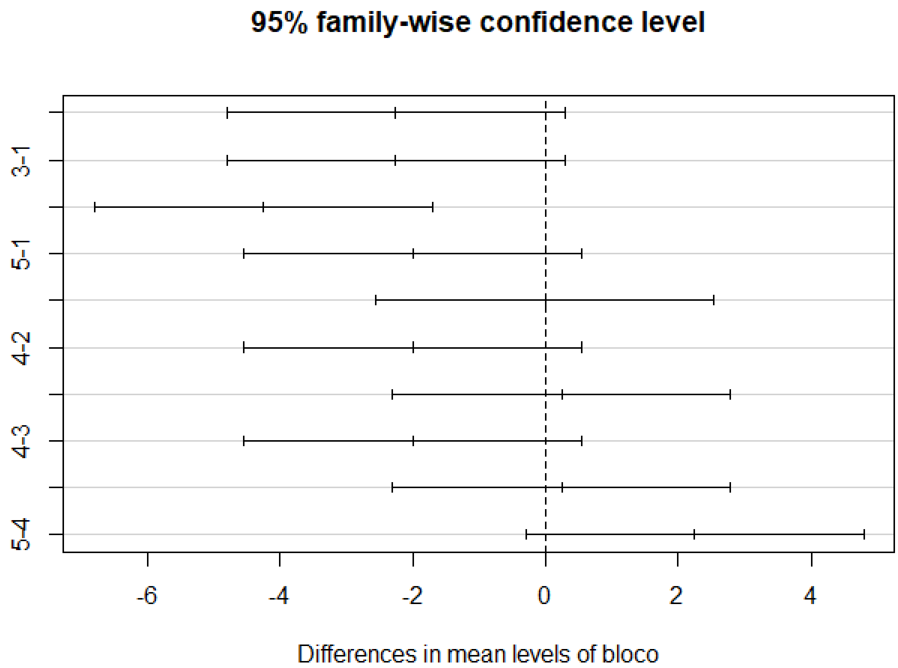
Através do Teste de Tukey e do valor-p fornecido para a comparação de cada um dos tratamentos, podemos destacar as diferenças significativas entre:

- `Tratamento 3` $\times$ `Tratamento 1`
- `Tratamento 4` $\times$ `Tratamento 1`

Para os blocos, temos:

- `Bloco 4` $\times$ `Bloco 1`

Graficamente, podemos observar os resultados a seguir:



4.27) O efeito de cinco ingredientes diferentes (A, B, C, D, E) no tempo de reação de um processo químico está sendo estudado. Cada lote de novo material é grande o suficiente para permitir que cinco execuções sejam feitas. Além disso, cada execução requer aproximadamente $1\frac{1}{2}$ horas, portanto, apenas cinco execuções podem ser feitas em um dia. O experimentador decide executar o experimento como um quadrado latino para que os efeitos do dia e do lote possam ser sistematicamente controlados. Ele obtém os dados que se seguem. Analise os dados deste experimento (use $\alpha = 0,05$) e tire conclusões.

$$Dia$$

Lote	1	2	3	4	5
1	A = 8	B = 7	D = 1	C = 7	E = 3
2	C = 11	E = 2	A = 7	D = 3	B = 8

Lote	1	2	3	4	5
3	B = 4	A = 9	C = 10	E = 1	D = 5
4	D = 6	C = 8	E = 6	B = 6	A = 10
5	E = 4	D = 2	B = 3	A = 8	C = 8

Respostas:

Neste exercício, utilizaremos o `Delineamento Quadrado Latino`, cujo o objetivo, inicialmente, é testar as médias dos tratamentos, de modo que:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1 : Pelo menos uma média é desigual

Para esta etapa optou-se por utilizar a função `aov` do `R`, obtemos os seguintes resultados para a análise de variância (ANOVA):

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## trat         4 141.44   35.36  11.309 0.000488 ***
## lote         4  15.44    3.86   1.235 0.347618    
## dia          4  12.24    3.06   0.979 0.455014    
## Residuals   12  37.52    3.13                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Portanto, considerando um $\alpha = 0,05$ e o valor-p obtido de 0,000488 para os tratamentos, podemos concluir que $valor - p < \alpha$, portanto, rejeitamos a hipótese nula, o que implica dizer que há evidências de que pelo menos um ingrediente químico tem uma diferença significativa no tempo de reação em comparação aos demais. Por outro lado, também podemos concluir que os efeitos do lote e do dia não são significativos no tempo de reação, tendo em vista que seus resultados no valor-p foram superiores ao nível de significância de 0,05 (0,347618 e 0,455014, respectivamente).

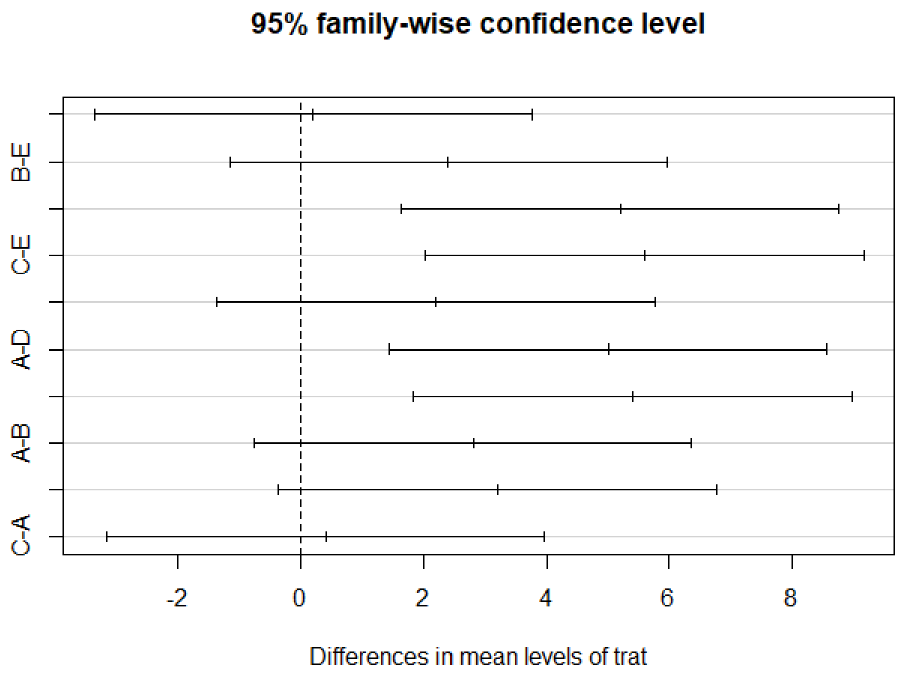
Para o Teste de Tukey são consideradas as seguintes Hipóteses:

$$H_0 : \mu_i = \mu_k$$

$$H_1 : \text{Pelo menos um par das médias é desigual}$$

O resultado do Teste utilizando a função `TukeyHSD()` é dado abaixo:

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
## 
## Fit: aov(formula = resp ~ trat + lote + dia)
## 
## $trat
##     diff        lwr      upr     p adj
## D-E  0.2 -3.3646078 3.764608 0.9997349
## B-E  2.4 -1.1646078 5.964608 0.2631551
## A-E  5.2  1.6353922 8.764608 0.0041431
## C-E  5.6  2.0353922 9.164608 0.0023007
## B-D  2.2 -1.3646078 5.764608 0.3365811
## A-D  5.0  1.4353922 8.564608 0.0055862
## C-D  5.4  1.8353922 8.964608 0.0030822
## A-B  2.8 -0.7646078 6.364608 0.1539433
## C-B  3.2 -0.3646078 6.764608 0.0864353
## C-A  0.4 -3.1646078 3.964608 0.9960012
```



Através do Teste de Tukey e do valor-p fornecido para a comparação de cada um dos tratamentos, podemos destacar as diferenças significativas entre:

- `Ingrediente A` $\times$ `Ingrediente E`
- `Ingrediente C` $\times$ `Ingrediente E`
- `Ingrediente A` $\times$ `Ingrediente D`
- `Ingrediente C` $\times$ `Ingrediente D`

4.40) O rendimento de um processo químico foi medido usando cinco lotes de matéria-prima, cinco concentrações de ácido, cinco tempos de espera (A, B, C, D, E), e cinco concentrações de catalisador (α, β, γ, δ, ϵ). Foi utilizada o quadrado greco-latina que se segue. Analise os dados deste experimento (use $\alpha = 0,05$) e tire conclusões.

Concentração de ácido

Lote	1	2	3
1	Aα = 26	Bβ = 16	Cγ = 19
2	Bγ = 18	Cδ = 21	Dϵ = 18
3	Cϵ = 20	Dα = 12	Eβ = 16
4	Dβ = 15	Eγ = 15	Aδ = 22
5	Eδ = 10	Aϵ = 24	Bα = 17

Respostas:

> *Delineamento Quadrado Greco Latino*

Neste exercício, utilizaremos o `Delineamento Quadrado Greco Latino`, cujo o objetivo, inicialmente, é testar as médias dos conjuntos, de modo que:

$H_0 : \omega_k = 0$ para todo k

$H_1 : \omega_k \neq 0$ para pelo menos um k

Para esta etapa optou-se por utilizar a função `aov` do `R`, obtemos os seguintes resultados para a análise de variância (ANOVA):

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## Tempo        4  342.8   85.70  14.650 0.000941 ***
## Catalyst     4   12.0    3.00   0.513 0.728900    
## Batch        4   10.0    2.50   0.427 0.785447    
## Acido        4   24.4    6.10   1.043 0.442543    
## Residuals    8   46.8    5.85                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A um nível de significância de 5%, temos evidências estatísticas suficientes que apenas o `Tempo de espera` aparece como efeito significativo.

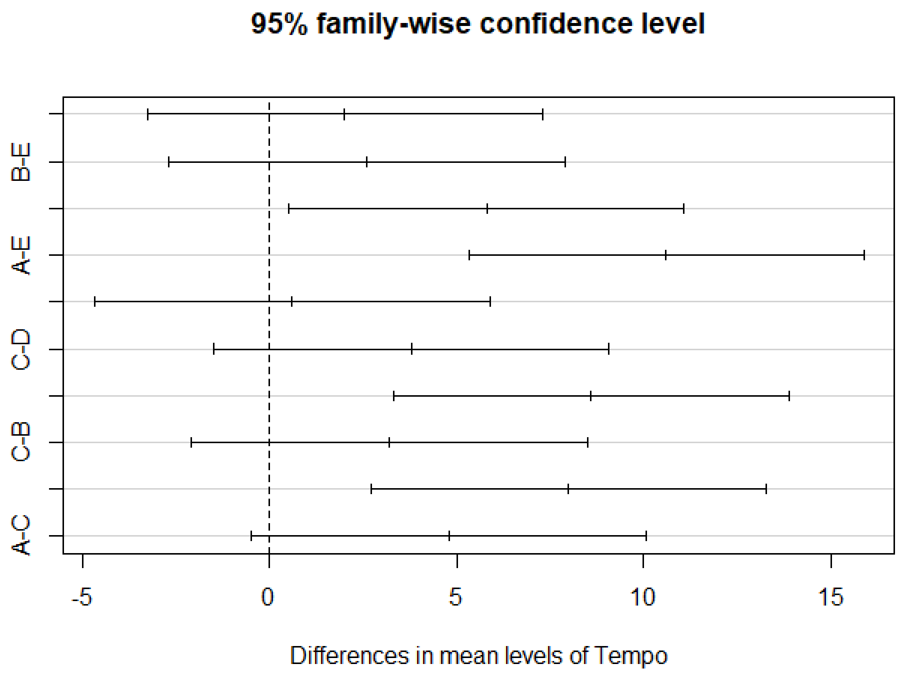
Para o Teste de Tukey são consideradas as seguintes hipóteses:

$$H_0 : \mu_i = \mu_k$$

$$H_1 : \text{Pelo menos um par das médias é desigual}$$

O resultado do Teste utilizando a função `TukeyHSD()` é dado abaixo:

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
## 
## Fit: aov(formula = Campo ~ Tempo + Catalyst + Batch + Acido, data = dados)
## 
## $Tempo
##     diff        lwr       upr     p adj
## D-E  2.0 -3.2847512  7.284751 0.6948188
## B-E  2.6 -2.6847512  7.884751 0.4837165
## C-E  5.8  0.5152488 11.084751 0.0317351
## A-E 10.6  5.3152488 15.884751 0.0008219
## B-D  0.6 -4.6847512  5.884751 0.9939694
## C-D  3.8 -1.4847512  9.084751 0.1869031
## A-D  8.6  3.3152488 13.884751 0.0032815
## C-B  3.2 -2.0847512  8.484751 0.3087034
## A-B  8.0  2.7152488 13.284751 0.0051639
## A-C  4.8 -0.4847512 10.084751 0.0770797
```



Tempo:	A	B	C	D	E
Rendimento médio	23,6	15,6	18,8	15,0	13,0

Através do Teste de Tukey e do valor-p fornecido para a comparação de cada um dos rendimentos, podemos destacar as diferenças significativas entre:

- `Tempo C` $\times$ `Tempo E`
- `Tempo A` $\times$ `Tempo E`
- `Tempo A` $\times$ `Tempo D`
- `Tempo A` $\times$ `Tempo B`

4.45) Um engenheiro está estudando as características de desempenho de quilometragem de cinco tipos de aditivos para gasolina. No teste de estrada, ele deseja usar carros como blocos; no entanto, devido a uma restrição de tempo, ele deve usar um projeto de bloco incompleto. Ele executa o projeto equilibrado com os cinco blocos que se seguem. Analise os dados deste experimento (use $\alpha = 0,05$) e tire conclusões.

$$Carro$$

Aditivo	1	2	3	4	5
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

Respostas:

Utilizando o `Delineamento por Blocos Incompletos`:

Para a nossa análise, temos: 5 tipos de aditivos, 5 blocos, com 4 aditivos por bloco, 4 repetições por aditivo e 20 observações.

O número de vezes que cada par de aditivo aparece junto no mesmo bloco é dado por:

$$\lambda = \frac{r(k-1)}{a-1}$$

substituindo os valores, temos:

$$\lambda = \frac{4(4-1)}{5-1} = 3$$

A seguir, testaremos as seguintes hipóteses a cerca do desempenho para os 5 tipos de aditivos:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1 : Pelo menos uma média é desigual

Através do cálculo utilizando o R, obtemos o seguinte valor-p:

```
##             Df Sum Sq Mean Sq F value  Pr(>F)   
## trat         4  31.70   7.925   8.703 0.00203 **
## blocos       4  35.23   8.808   9.673 0.00132 **
## Residuals   11  10.02   0.911                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como podemos observar, o nosso valor-p é menor que nosso nível de significância pré-estabelecido ($\alpha = 0,05$), portanto, temos evidências estatísticas suficientes para rejeitar a hipótese nula e aceitar a hipótese alternativa, ou seja, podemos concluir que o desempenho de quilometragem não é o mesmo para os 5 tipos de aditivos para gasolina.

Com base nos resultados anteriores, onde vimos que há uma diferença no desempenho de quilometragem para os 5 tipos de aditivos de gasolina, veremos, efetivamente, quais os aditivos que diferem entre si através do Teste de Tukey.

Para o Teste de Tukey são consideradas as seguintes hipóteses:

$$H_0 : \mu_i = \mu_k$$

$$H_1 : \text{Pelo menos um par das médias é desigual}$$

O resultado do Teste utilizando a função `TukeyHSD()` é dado abaixo:

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
## 
## Fit: aov(formula = resp ~ trat + blocos, data = bibd.df)
## 
## $trat
##     diff         lwr      upr     p adj
## C-E 1.25 -0.93219269 3.432193 0.3939985
## D-E 1.50 -0.68219269 3.682193 0.2406188
## B-E 2.50  0.31780731 4.682193 0.0231360
## A-E 3.75  1.56780731 5.932193 0.0012777
## D-C 0.25 -1.93219269 2.432193 0.9953712
## B-C 1.25 -0.93219269 3.432193 0.3939985
## A-C 2.50  0.31780731 4.682193 0.0231360
## B-D 1.00 -1.18219269 3.182193 0.5930409
## A-D 2.25  0.06780731 4.432193 0.0424345
## A-B 1.25 -0.93219269 3.432193 0.3939985
```

Aditivo:	1	2	3	4	5
Desempenho médio(km):	14,00	12,75	11,50	11,75	10,25

Através do Teste de Tukey e do valor-p fornecido para a comparação de cada um dos aditivos, podemos destacar as diferenças significativas entre:

- `Aditivo B` × `Aditivo E`, com o aditivo B oferecendo maior desempenho.
- `Aditivo A` × `Aditivo E`, com o aditivo A oferecendo maior desempenho.
- `Aditivo A` × `Aditivo C`, com o aditivo A oferecendo maior desempenho.
- `Aditivo A` × `Aditivo D`, com o aditivo A oferecendo maior desempenho.