Modelos Lineares Clássicos

Análise de Variância (ANOVA) de um fator

Fabio Cop (fabiocopf@gmail.com)
Instituto do Mar - UNIFESP
Última atualização em 13 de junho de 2023

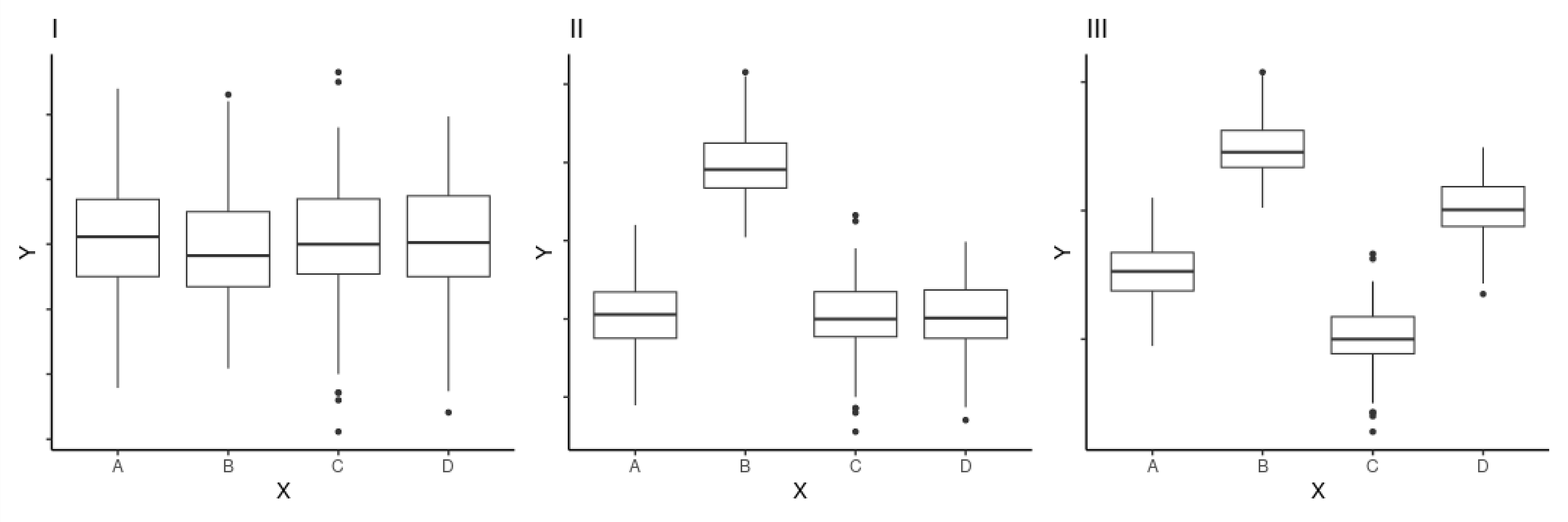
Análise de Variância (ANOVA) de um fator

1. O modelo da ANOVA e as hipóteses estatísticas

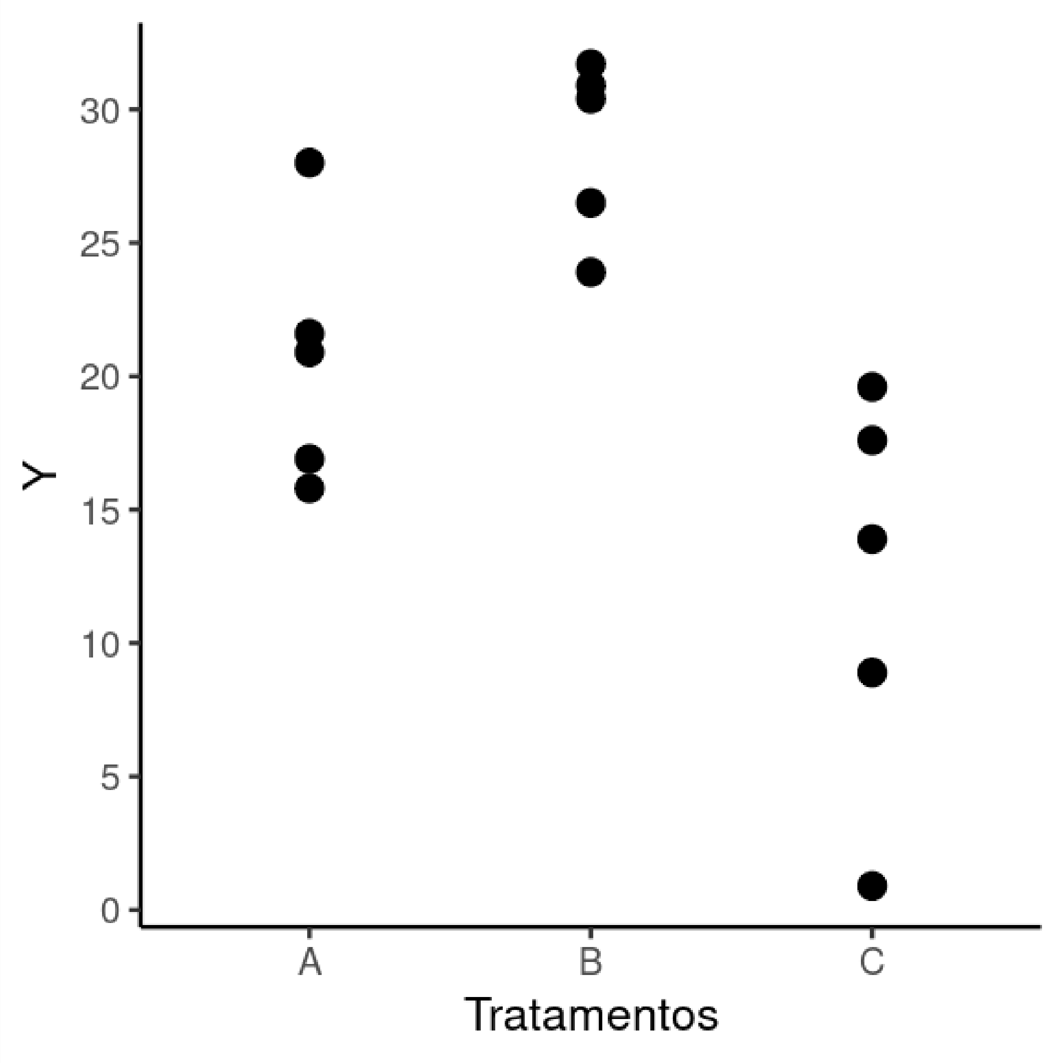
$$Y_{ij} = \mu + A_i + \epsilon_{ij}$$

$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ (HIPÓTESE NULA)

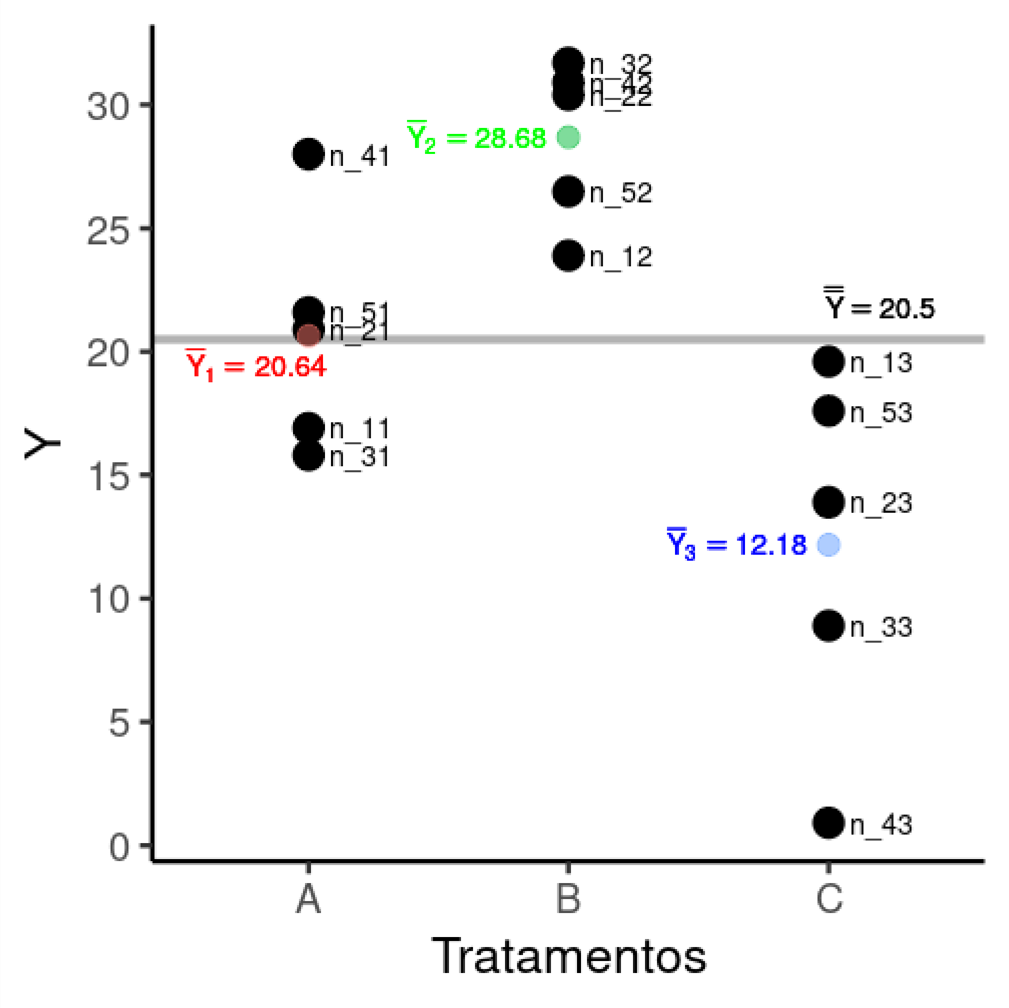
H_a: ao menos um par de médias diferen entre si (HIPÓTESE ALTERNATIVA)



1. O modelo da ANOVA e as hipóteses estatísticas



Tratamentos		
A	B	C
16.90	23.90	19.60
20.90	30.40	13.90
15.80	31.70	8.90
28.00	30.90	0.90
21.60	26.50	17.60
20.64	**28.68**	**12.18**



- $k = 3$ grupos: `A`, `B` ou `C`
- $n_1 = n_2 = n_3 = n = 5$ observações por grupo. Denotamos por n_{ij} o número de
- $N = k \times n = n_1 + n_2 + n_3 = 15$
- $\overline{Y}_A = 20.64$; $\overline{Y}_B = 28.68$, $\overline{Y}_D = 12.18$ - **estimam** μ_1, μ_2 e μ_3
- $\overline{\overline{Y}}$: a **Grande Média** - **estima** μ.

$$\overline{\overline{Y}} = \sum_{j=1}^{k} \sum_{i=1}^{n} \frac{Y_{ij}}{N} = \frac{\overline{Y_1} + \overline{Y_2} + \overline{Y_3}}{3} = 20.5$$

2. Partição da Soma dos Quadrados (SQ)

i. **Soma dos Quadrados Totais** - SQ_{Total}

$$SQ_{Total} = \sum_{j=1}^{k} \sum_{i=1}^{n} (Y_{ij} - \overline{\overline{Y}})^2$$

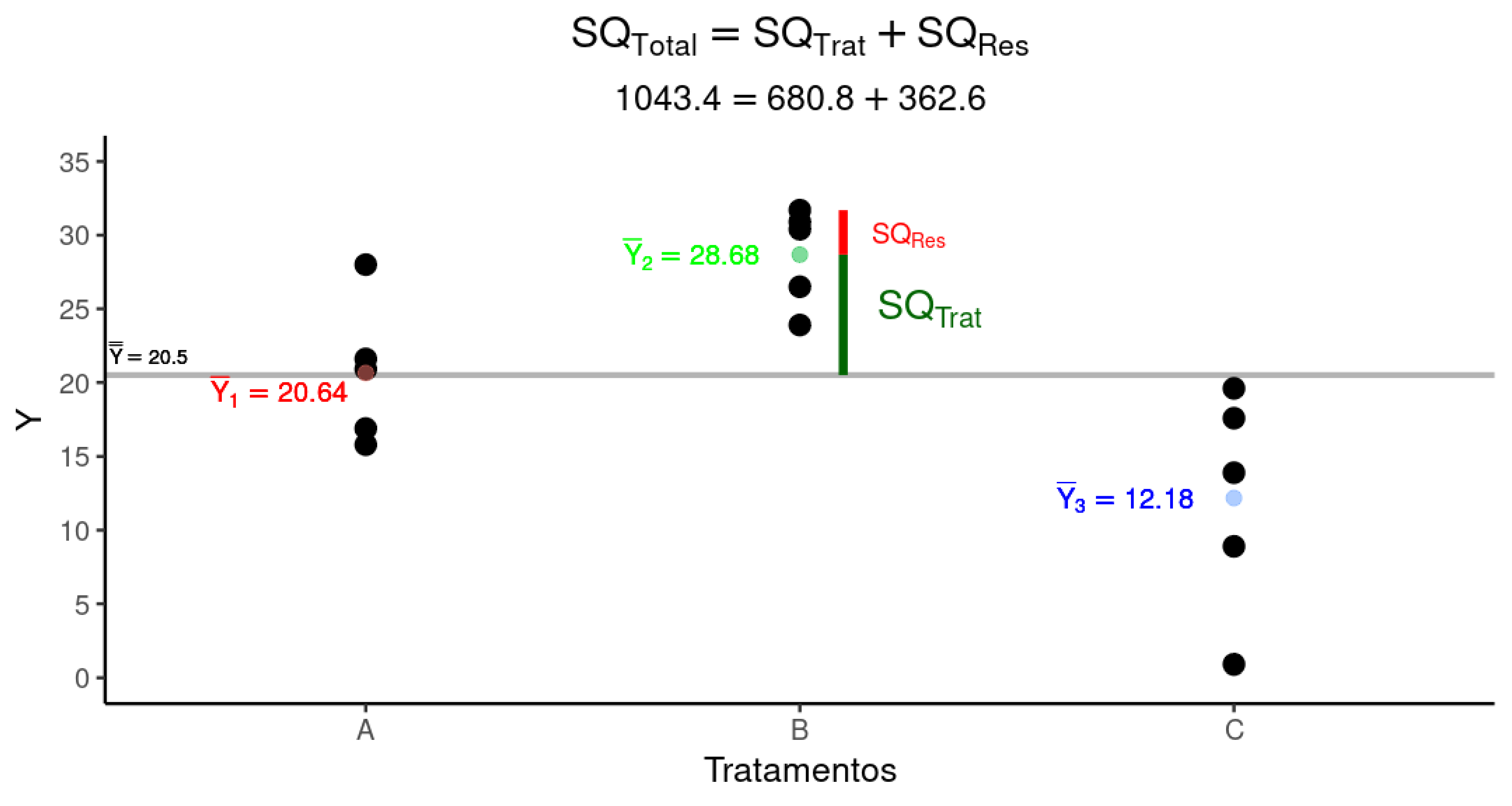
ii. **Soma dos Quadrados dos Tratamentos** - SQ_{Trat}:

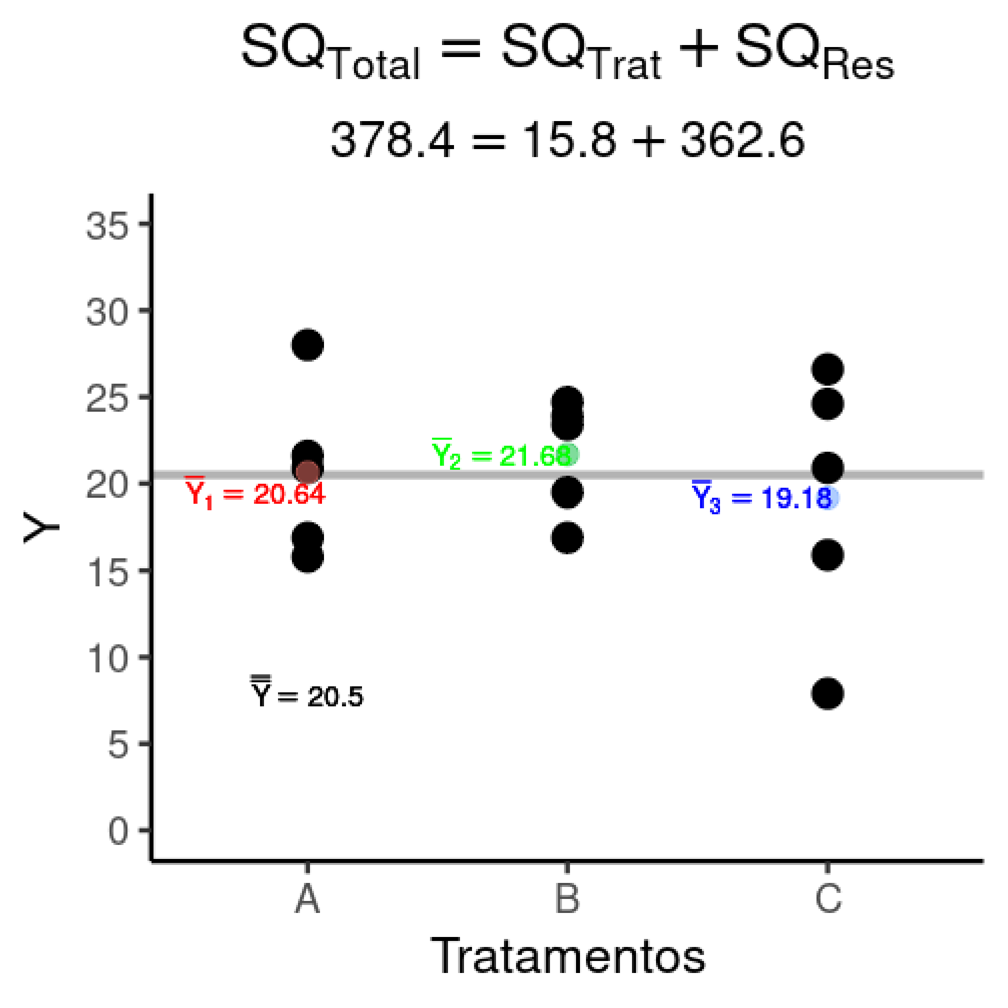
$$SQ_{Trat} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{Y}_j - \overline{\overline{Y}})^2 = \sum_{j=1}^{k} n_j (\overline{Y}_j - \overline{\overline{Y}})^2$$

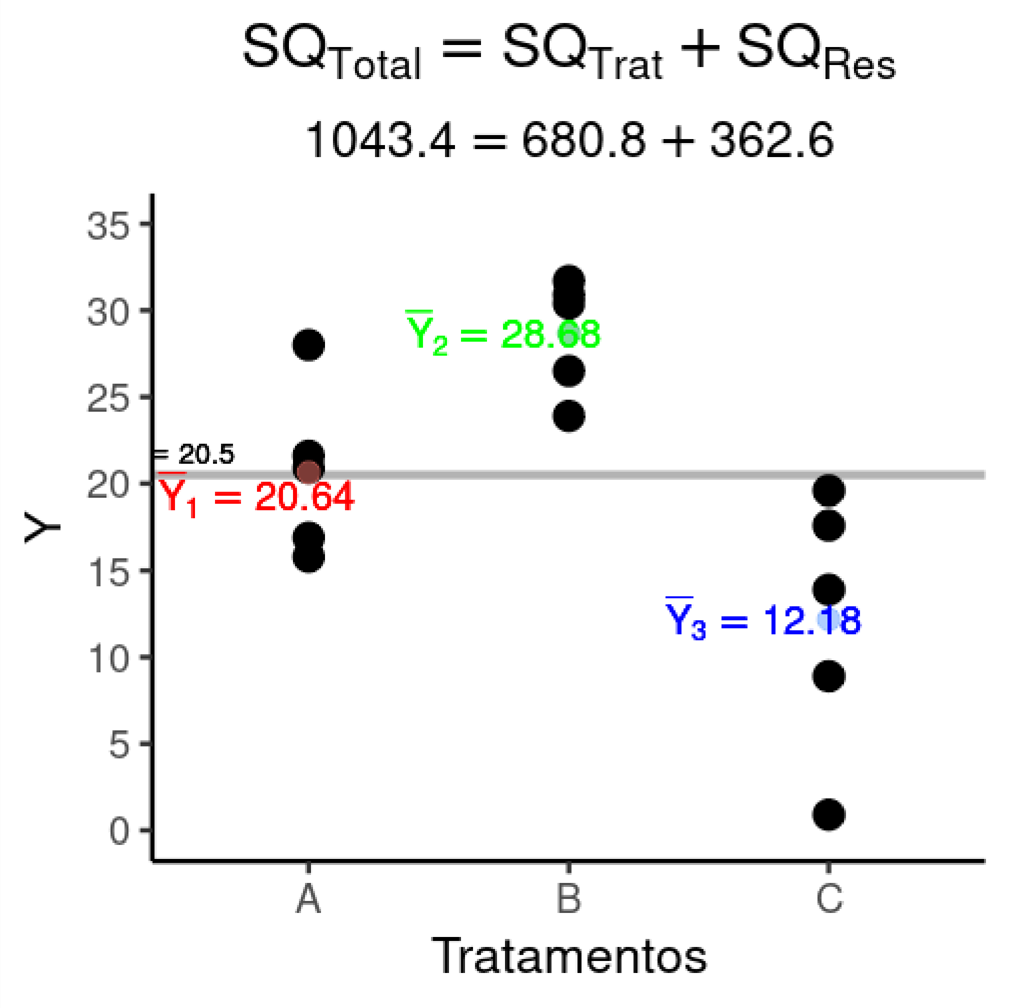
iii. **Soma dos Quadrados dos Resíduos** - SQ_{Res}

$$SQ_{Res} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_j)^2$$

2. Partição da Soma dos Quadrados (SQ)




$SQ_{Total} = SQ_{Trat} + SQ_{Res}$
$378.4 = 15.8 + 362.6$
$\bar{Y}_1 = 20.64$
$\bar{Y}_2 = 21.68$
$\bar{Y}_3 = 19.18$
$\bar{\bar{Y}} = 20.5$
Y
Tratamentos
A
B
C


$SQ_{Total} = SQ_{Trat} + SQ_{Res}$
$1043.4 = 680.8 + 362.6$
$\bar{Y}_1 = 20.64$
$\bar{Y}_2 = 28.68$
$\bar{Y}_3 = 12.18$
= 20.5
Y
Tratamentos
A
B
C

3. Quadrados médios e graus de liberdade

i. **Quadrado médio total** - QM_{Total}

$$QM_{Total} = \frac{SQ_{Total}}{gl_{Total}}$$

ii. **Quadrado médio entre tratamentos** - QM_{Trat}

$$QM_{Trat} = \frac{SQ_{Trat}}{gl_{Trat}}$$

iii. **Quadrado médio dentro dos tratamentos** - QM_{Res}

$$QM_{Res} = \frac{SQ_{Res}}{gl_{Res}}$$

$$gl_{Total} = N - 1$$

$$gl_{Trat} = k - 1$$

$$gl_{Res} = N - k$$

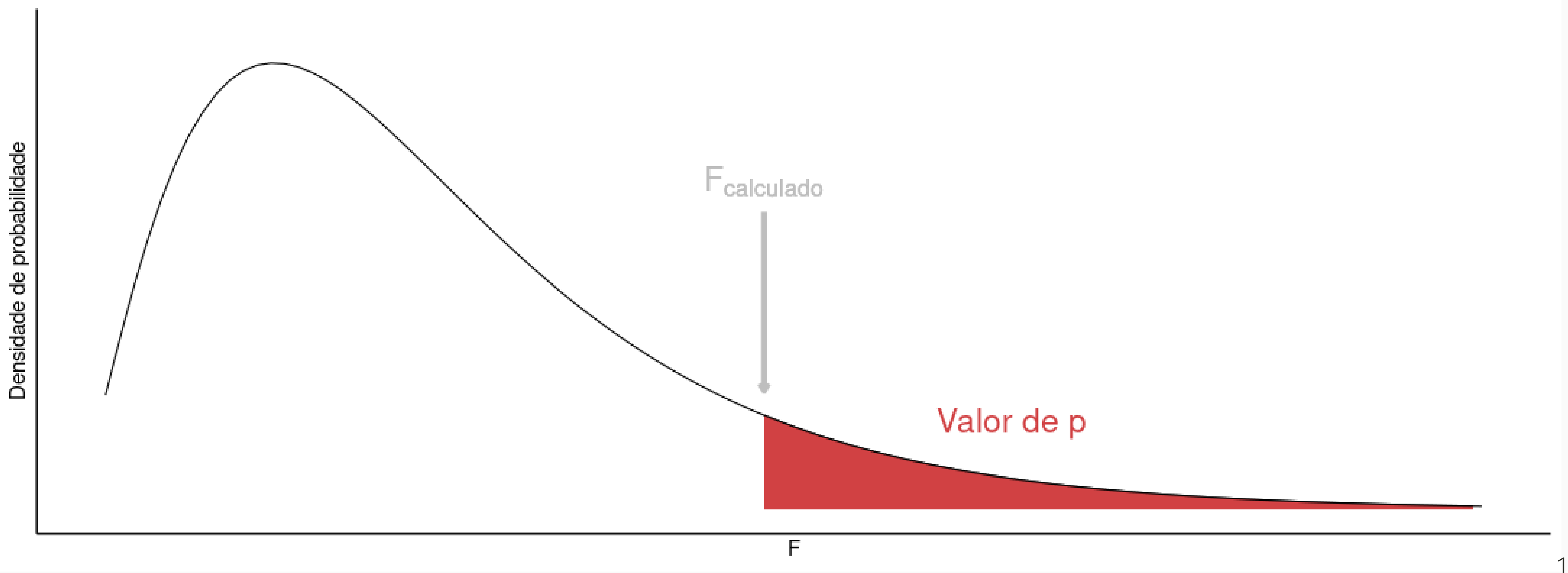
$$gl_{Total} = gl_{Trat} + gl_{Res} = (k-1) + (N-K) = N - 1$$

4. Estatística F e teste de hipóteses

$$F_{calculado} = \frac{QM_{Trat}}{QM_{Res}}$$

$F_{calculado}$ é comparado ao nível de significância α

> Se $p > \alpha$ --> **ACEITAMOS** H_0
>
> Se $p \leq \alpha$ --> **REJEITAMOS** H_0 (e assumimos H_a como verdadeira)



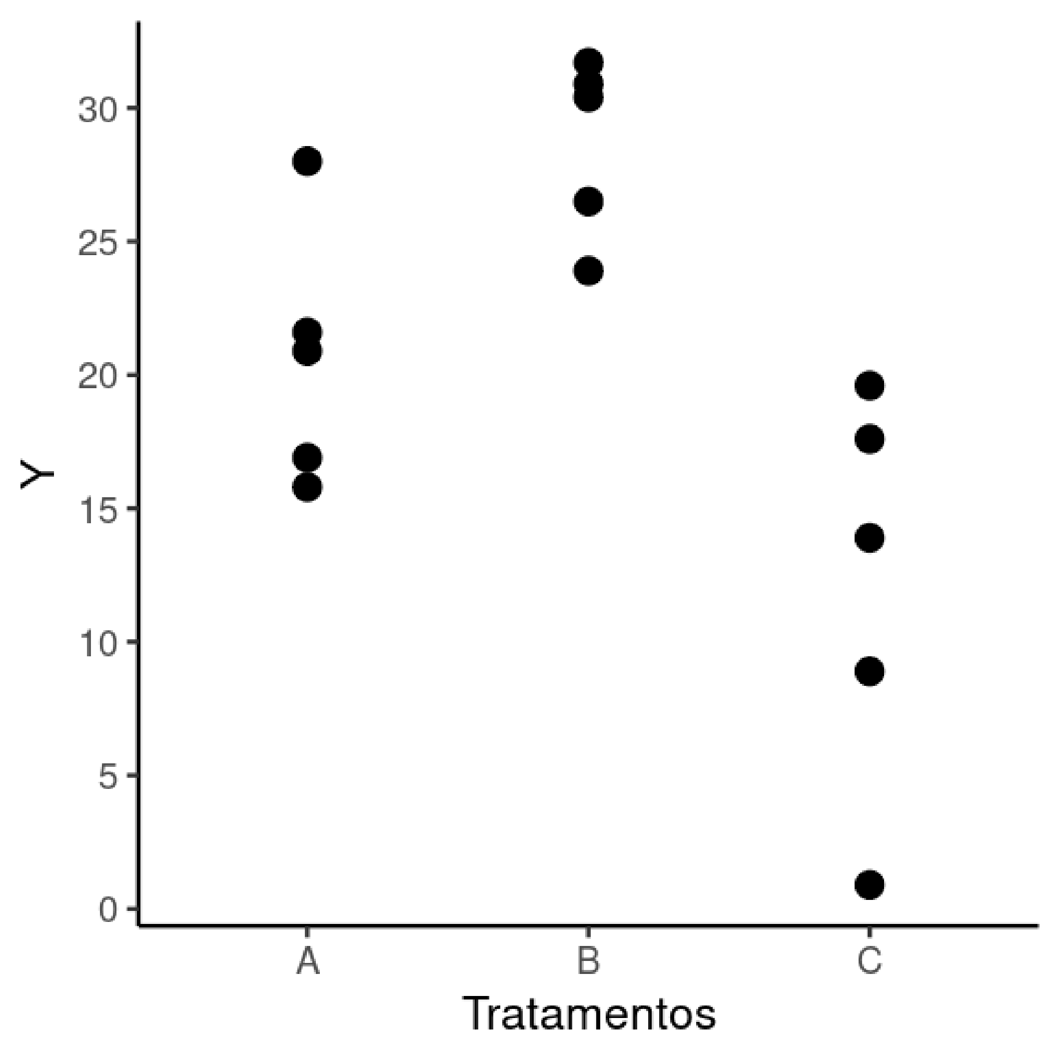
5. Um exemplo de ANOVA

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

H_a: ao menos um μ é diferente

$\alpha = 0.05$

Tratamentos		
A	B	C
16.90	23.90	19.60
20.90	30.40	13.90
15.80	31.70	8.90
28.00	30.90	0.90
21.60	26.50	17.60
20.64	**28.68**	**12.18**



1. Somatórios dos quadrados

$$SQ_{Trat} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (\overline{Y}_j - \overline{\overline{Y}})^2 = 680.772$$

$$SQ_{Res} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_j)^2 = 362.568$$

2. Graus de liberdade

$$gl_{Trat} = k - 1 = 2$$

$$gl_{Res} = N - k = 12$$

3. Quadrados médios

$$QM_{Trat} = \frac{SQ_{Trat}}{gl_{Trat}} = 340.386$$

$$QM_{Res} = \frac{SQ_{Res}}{gl_{Res}} = 30.214$$

4. Estatística F

$$F_{calculado} = \frac{QM_{Trat}}{QM_{Res}} = 11.266$$

6. A tabela da ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X	2	680.772	340.386	11.26584	0.0017611
Residuals	12	362.568	30.214	NA	NA

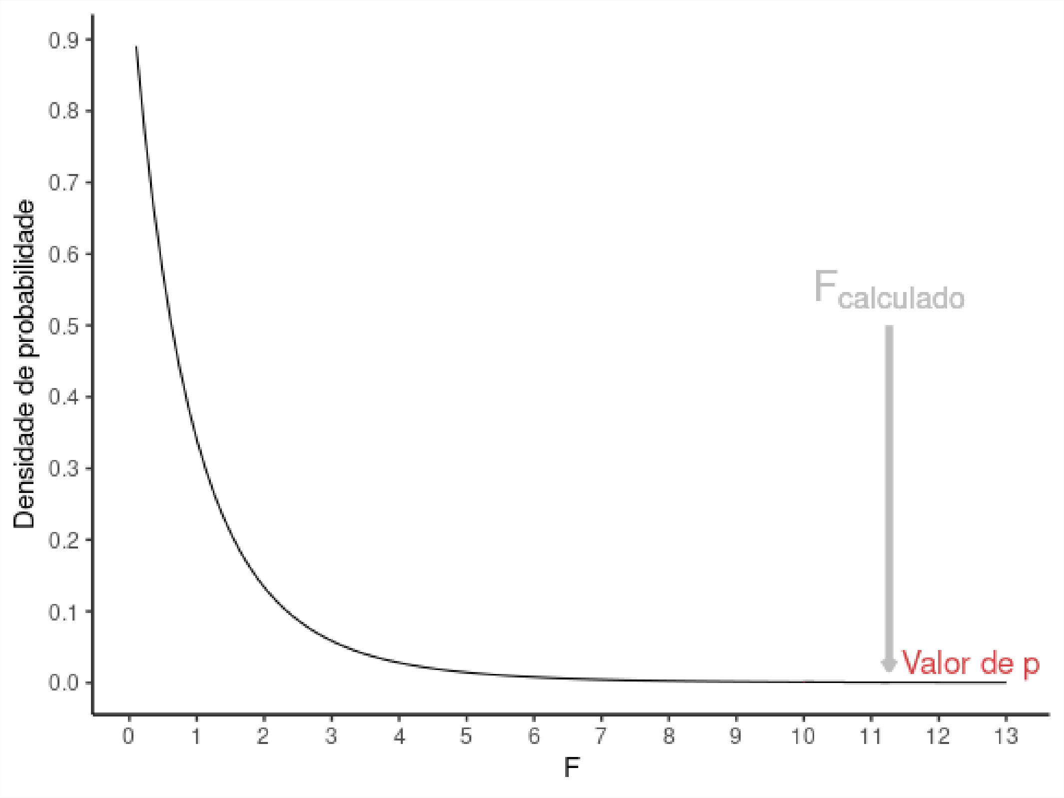
`Df`: graus de liberdade

`Sum Sq`: soma dos quadrados

`Mean Sq`: quadrados médios

`F value`: valor de $F_{calculado}$

`Pr(>F)`: valor de p



7. Testes a *posteriori* de comparação de médias

Quais pares são estatisticamente diferentes?

$$q = \frac{\overline{Y}_1 - \overline{Y}_2}{SE}$$

em que:

$$SE = \sqrt{\frac{QM_{Res}}{2}(\frac{1}{n_1} + \frac{1}{n_2})}$$

q: é e estatística do teste

$\overline{Y}_1$: é a maior das médias do par consideraddo;

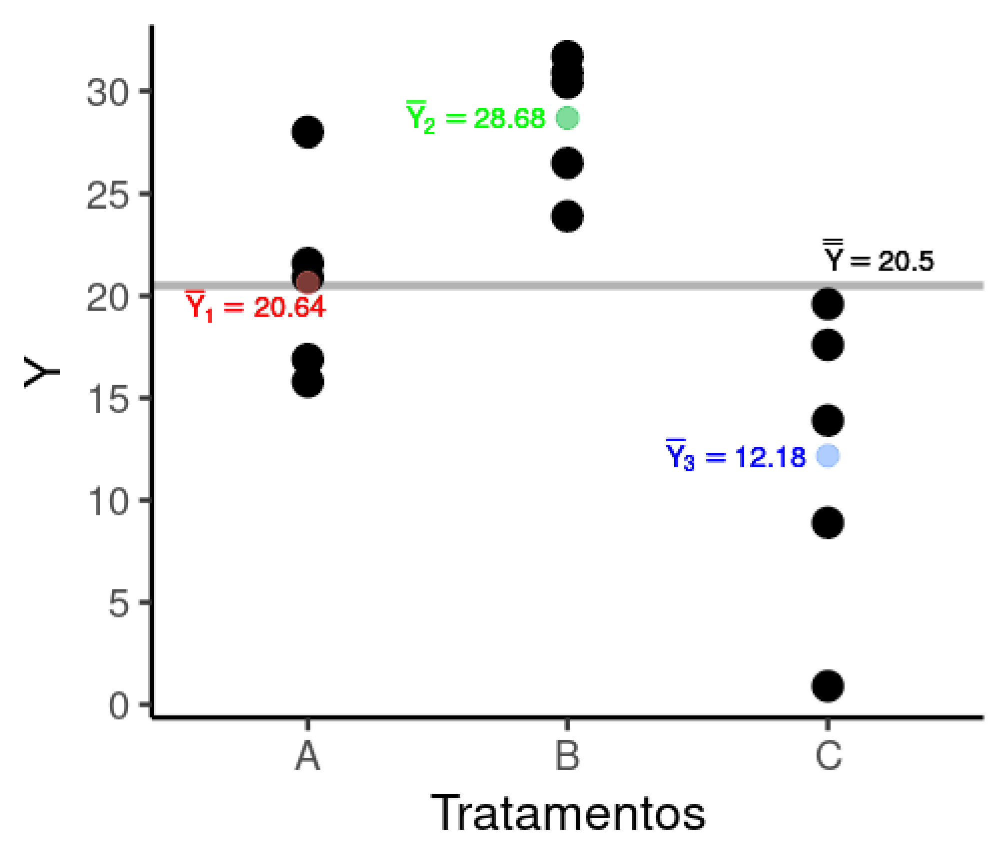
$\overline{Y}_2$: é a menor das médias do par consideraddo

QM_{Res}: é quadrado médio do resíduo obtido na ANOVA, e;

n_1, n_2: os tamanhos amostrais de cada grupo envolvido na comparação.

7. Testes a *posteriori* de comparação de médias

$$SE = \sqrt{\frac{30.214}{2}(\frac{1}{5} + \frac{1}{5})} = 2.458$$



$$q = \frac{\overline{Y}_1 - \overline{Y}_2}{SE}$$

combinacoes	diff	n1	n2	se	q	H0
B-A	8.04	5	5	2.458	3.271	Aceita H0
C-A	8.46	5	5	2.458	3.442	Aceita H0
C-B	16.50	5	5	2.458	6.712	Rejeita H0

O valor de **q** *crítico* pode ser procurado na tabela Studentized Range q Table, em que:

k: número de grupos envolvido na ANOVA

df: graus de liberdade dos resíduos da ANOVA

8. Ajustando a ANOVA no R

Ao fazer no R, temos diretamente os valores de p (`p adj`) resultantes do teste de Tukey para cada par de médias.

```r
ajuste = aov(Y ~ X, data  = Tab)
anova(ajuste)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## X          2 680.77  340.39  11.266 0.001761 **
## Residuals 12 362.57   30.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
alfa = 0.05
TukeyHSD(ajuste, conf.level = 1-alfa)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Y ~ X, data = Tab)
##
## $X
##       diff        lwr        upr     p adj
## B-A   8.04  -1.234654 17.3146545 0.0923564
## C-A  -8.46 -17.734654  0.8146545 0.0751622
## C-B -16.50 -25.774654 -7.2253455 0.0012751
```