



Estatística Frequentista

Análise de Variância

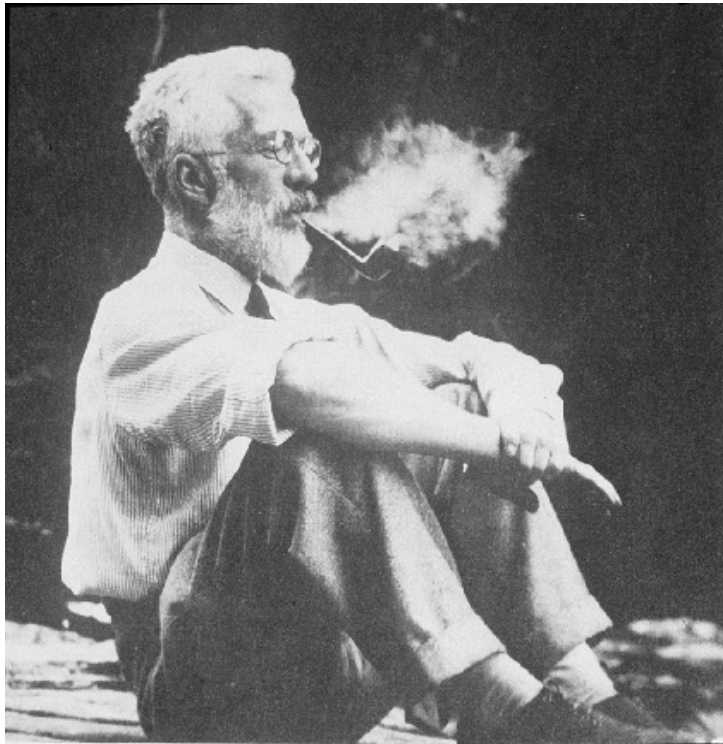
Alexandre Adalardo de Oliveira
PlanECO 2017

Conteúdo

- histórico
- conceituação básica
- exemplo

Testes Clássicos

Histórico



- desenho experimental
- teste associado

Testes Clássicos

Tipo de Variável		Estatística Clássica	
Resposta	Preditora	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência
Contínua	Categórica (2 níveis)	Teste t	$\mu_1 = \mu_2$
Contínua	Categórica	Anova	$\mu_1 = \mu_2 = \dots = \mu_n$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0; \beta_n = 0$
Contínua	Cont + Categ	Ancova	$\beta_1 = \beta_2; \alpha_1 = \alpha_2$
Proporção	Contínua	Reg. Logística	$\text{logit}(\beta_1) = 1$

Testes Clássicos

Tipo de Variável		Estatística Clássica	
Resposta	Preditora	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência
Contínua	Categórica (2 níveis)	Teste t	$\mu_1 = \mu_2$
Contínua	Categórica	Anova	$\mu_1 = \mu_2 = \mu_n$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0 ; \beta_n = 0$
Contínua	Cont + Categ	Ancova	$\beta_1 = \beta_2 ; \alpha_1 = \alpha_2$
Proporção	Contínua	Reg. Logística	$\text{logit}(\beta_1) = 1$

Anova: um exemplo

```
## 'data.frame':  30 obs. of  2 variables:  
## $ solo : chr  "are" "are" "are" "are" ...  
## $ colhe: int  6 10 8 6 14 17 9 11 7 11 ...
```

Anova: gráfico padrão

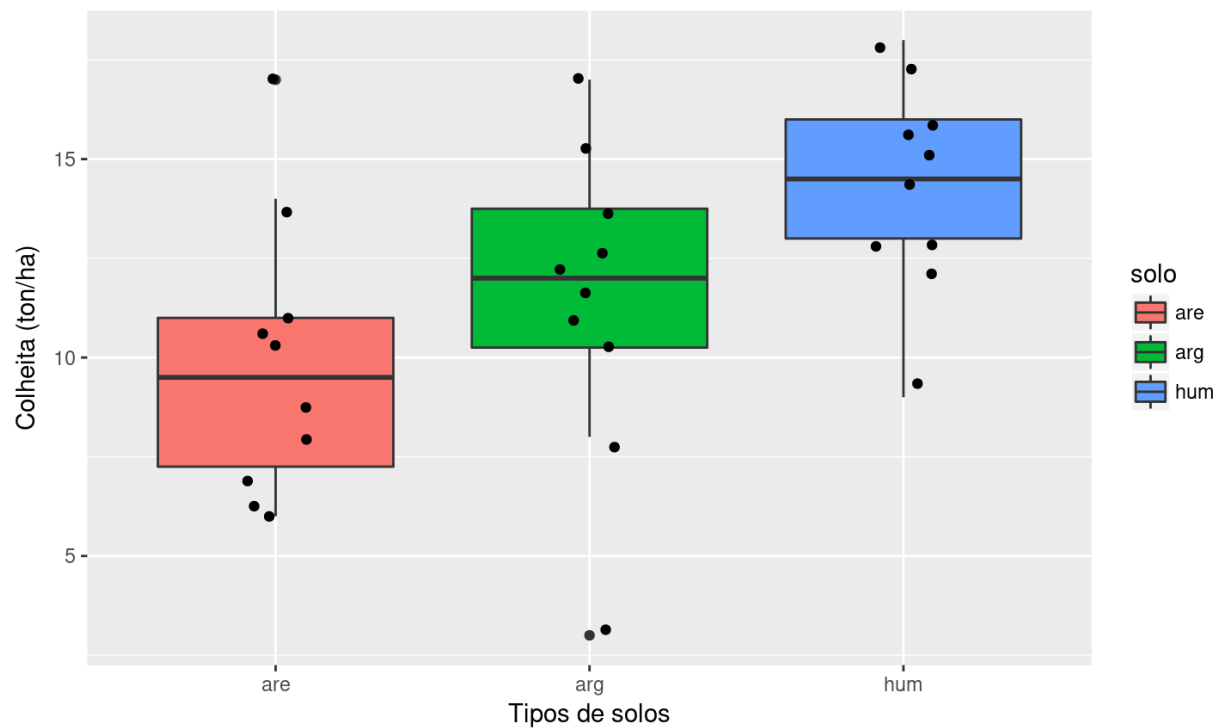
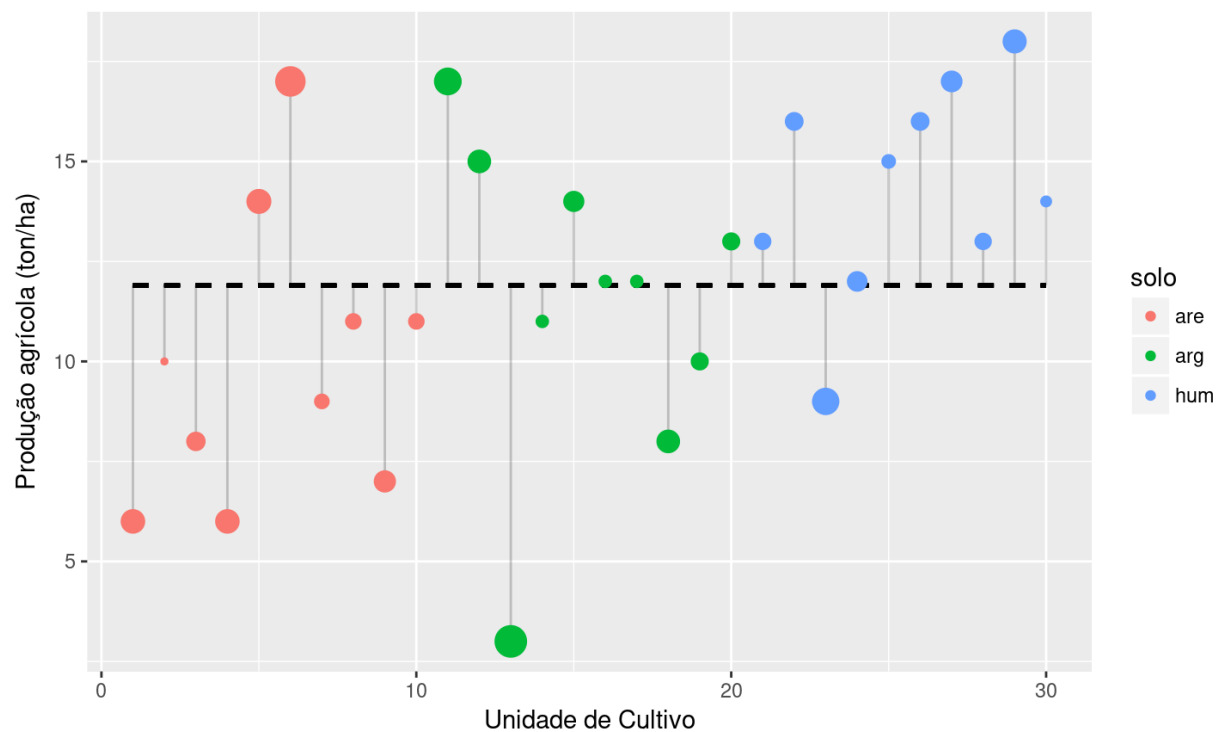


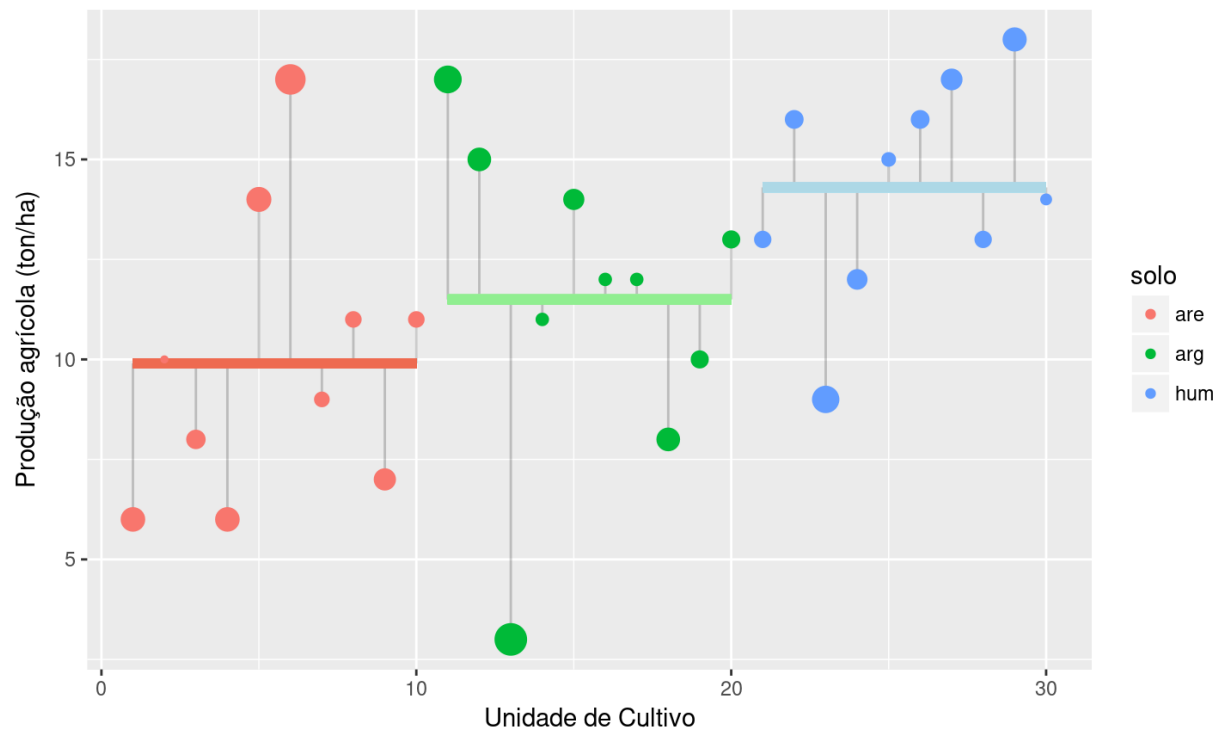
Gráfico dos desvios

A variação total dos dados (variável aleatória)



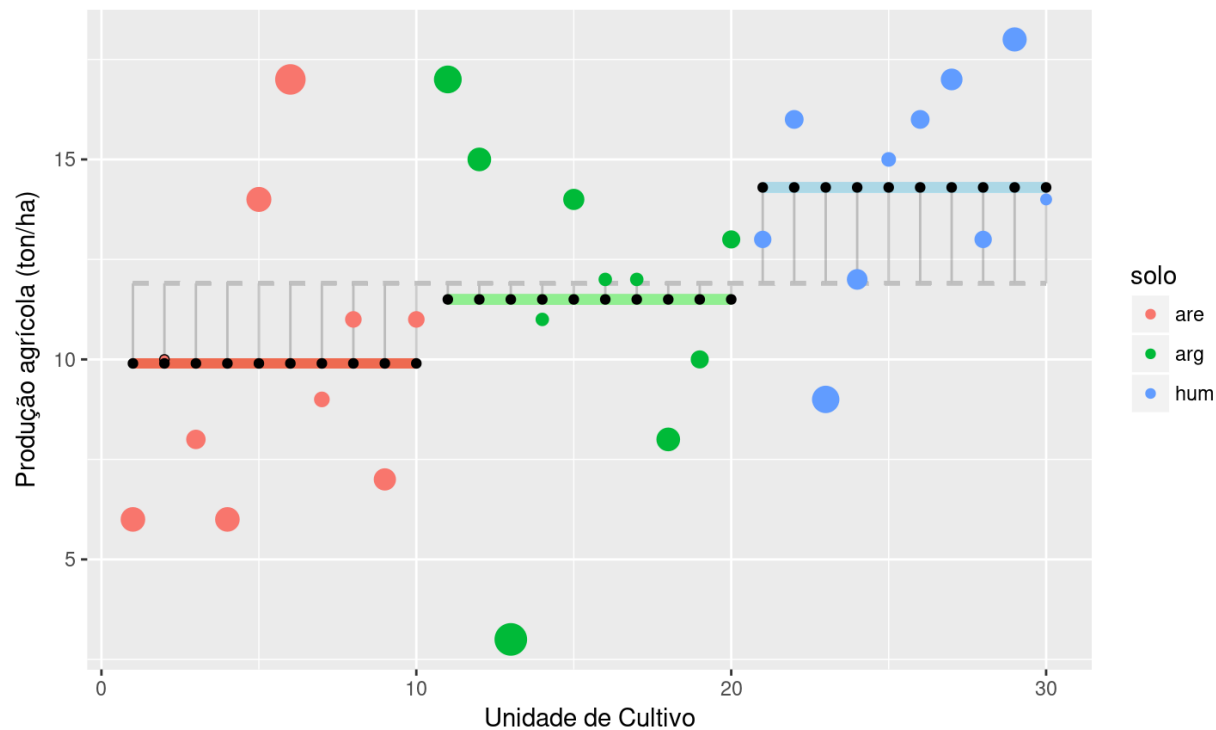
A variação interna aos grupos

O quanto cada observação difere da média do grupo

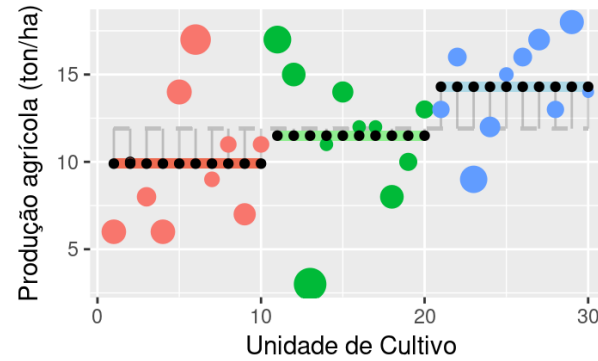
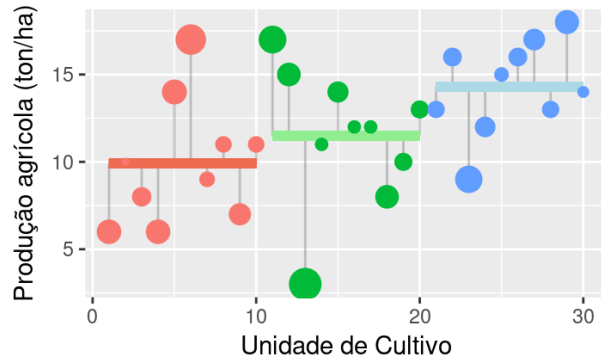
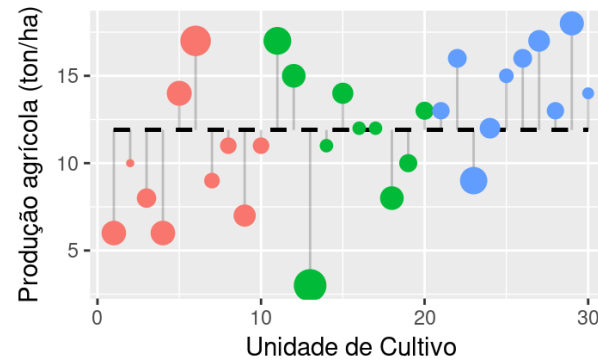
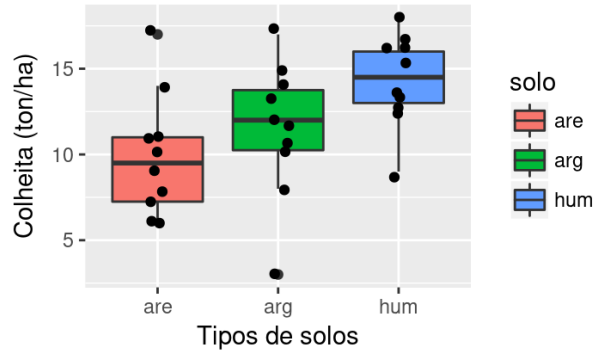


A variação entre os grupos

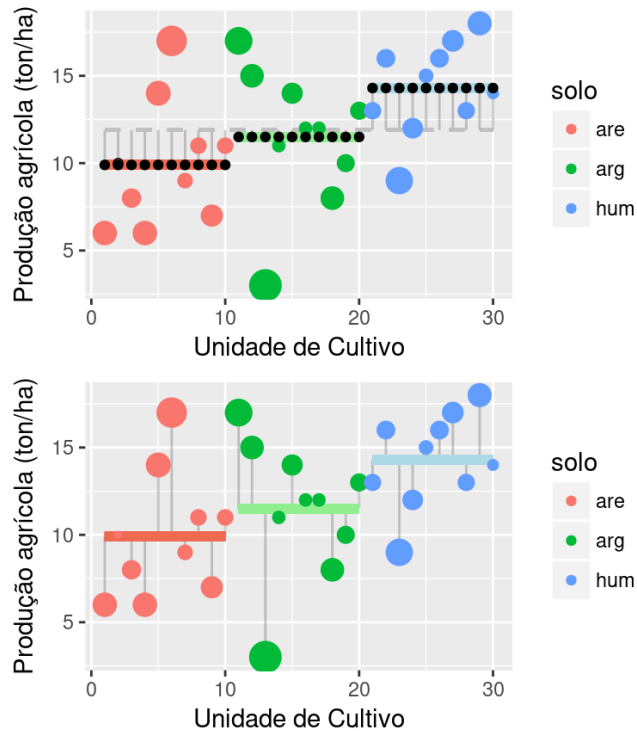
O quanto a média dos grupos difere da geral



Anova: partição da variância



Anova: partição da variância



Estatística F

Razão entre variâncias

$$F = \frac{\sigma_{entre}^2}{\sigma_{intra}^2}$$

Médias Quadráticas: intra

$$SS_{intra} = \sum_{i=1}^k \sum_{j=1}^n (y_{i,j} - \bar{y}_i)^2$$

$$MS_{intra} = \frac{SS_{intra}}{df_{intra}}$$

Médias Quadráticas: entre

$$SS_{entre} = \sum_{i=1}^k \sum_{j=1}^n (\bar{y}_i - \bar{\bar{y}})^2$$

$$MS_{entre} = \frac{SS_{entre}}{df_{intra}}$$

Estatística F

$$F = \frac{MS_{entre}}{MS_{intra}}$$

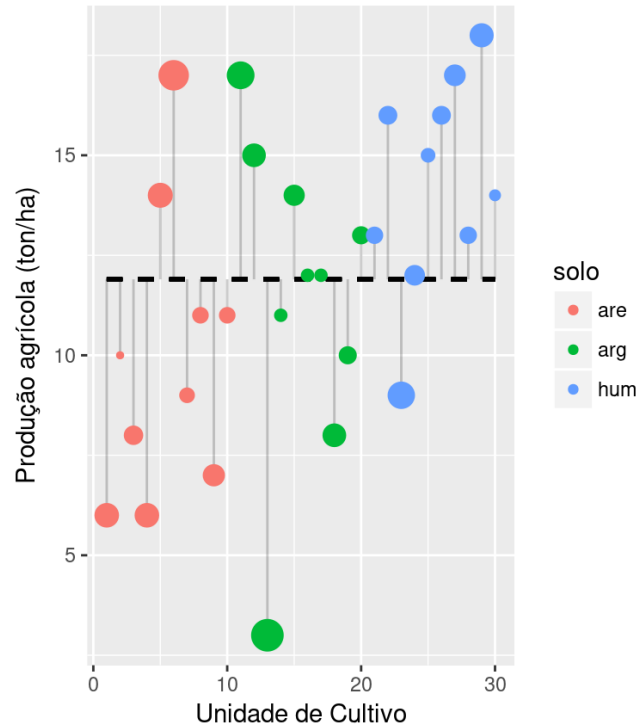
Mesmo que:

$$F = \frac{\sigma_{entre}^2}{\sigma_{intra}^2}$$

Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos					
Intra Grupos					
TOTAL	X				

Desvios Quadráticos: total



Desvios quadráticos totais

$$SS_{total} = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2$$

Desvios quadráticos: totais

```
crop$colhe
```

```
## [1] 6 10 8 6 14 17 9 11 7 11 17 15 3 11 14 12 12 8 10 13 13 16 9  
## [24] 12 15 16 17 13 18 14
```

```
mean(crop$colhe)
```

```
## [1] 11.9
```

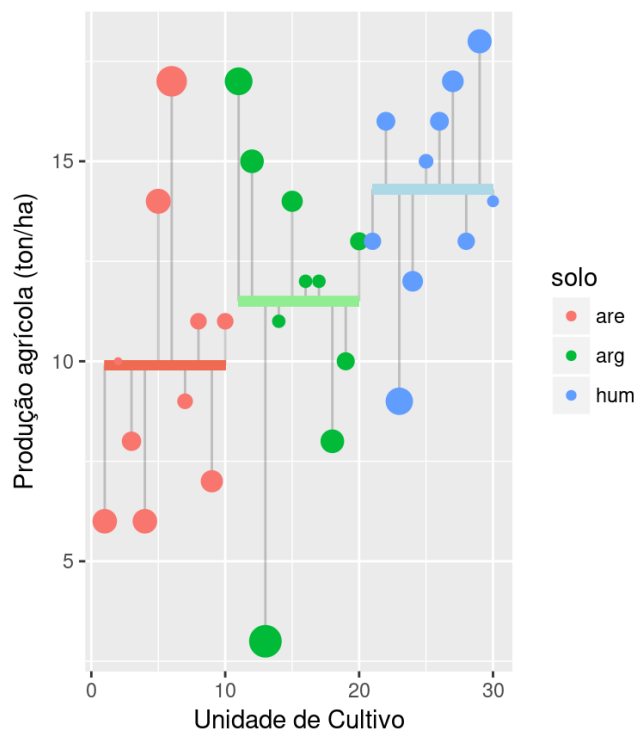
```
(ss_total <- sum((crop$colhe - mean(crop$colhe))^2 ))
```

```
## [1] 414.7
```

Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos					
Intra Grupos	x				
TOTAL	414.7				

Desvios Quadráticos: intra



Desvios quadráticos internos ao grupo

$$SS_{in} = \sum_{i=1}^k \sum_{j=1}^n (y_{i,j} - \bar{y}_i)^2$$

Desvios Quadráticos: intra grupos

```
crop$colhe
```

```
## [1] 6 10 8 6 14 17 9 11 7 11 17 15 3 11 14 12 12 8 10 13 13 16 9  
## [24] 12 15 16 17 13 18 14
```

```
(mcrop_vetor <- rep(mcrop, each=10))
```

```
## are are are are are are are are are are arg arg arg arg arg  
## 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 11.5 11.5 11.5 11.5 11.5  
## arg arg arg arg arg hum hum hum hum hum hum hum hum hum hum  
## 11.5 11.5 11.5 11.5 11.5 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3
```

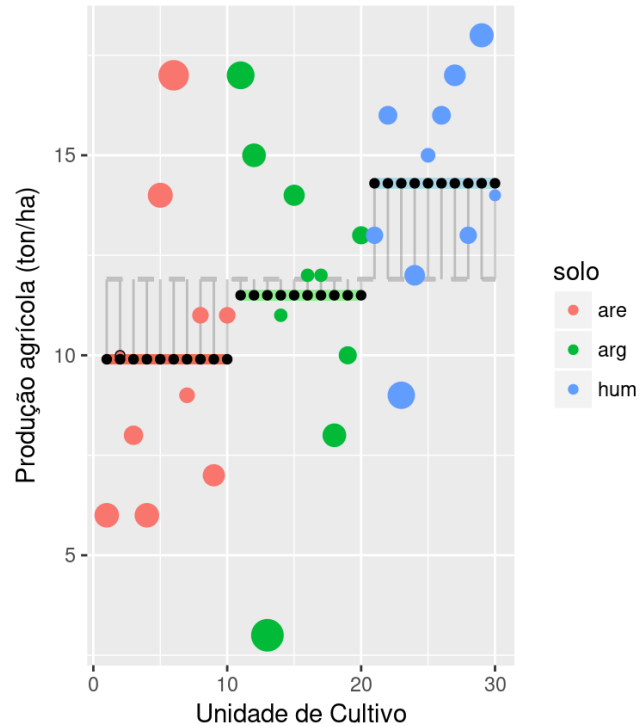
```
(ss_intra <- sum((crop$colhe- mcrop_vetor)^2))
```

```
## [1] 315.5
```

Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos	x				
Intra Grupos	315.5				
TOTAL	414.7				

Desvios Quadráticos: entre



Desvios quadráticos entre grupo

$$SS_{en} = \sum_{i=1}^k \sum_{j=1}^n (\bar{y}_i - \bar{y})^2$$

Desvios Quadráticos: intra grupos

```
mcrop_vetor
```

```
## are are are are are are are are are are arg arg arg arg arg  
## 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 9.9 11.5 11.5 11.5 11.5 11.5  
## arg arg arg arg arg hum hum hum hum hum hum hum hum hum hum  
## 11.5 11.5 11.5 11.5 11.5 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3 14.3
```

```
(mean(crop$colhe))
```

```
## [1] 11.9
```

```
(ss_entre <- sum((mcrop_vetor- mean(crop$colhe))^2))
```

```
## [1] 99.2
```


Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos	99.2				
Intra Grupos	315.5				
TOTAL	414.7				

Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias (F)	P-valor
Entre Grupos	99.2	2	X	X	
Intra Grupos	315.5	27	X		
TOTAL	414.7	29			

Estatística F

$$F = \frac{\sigma_{entre}^2}{\sigma_{intra}^2}$$

O mesmo que:

$$F = \frac{MS_{entre}}{MS_{intra}}$$

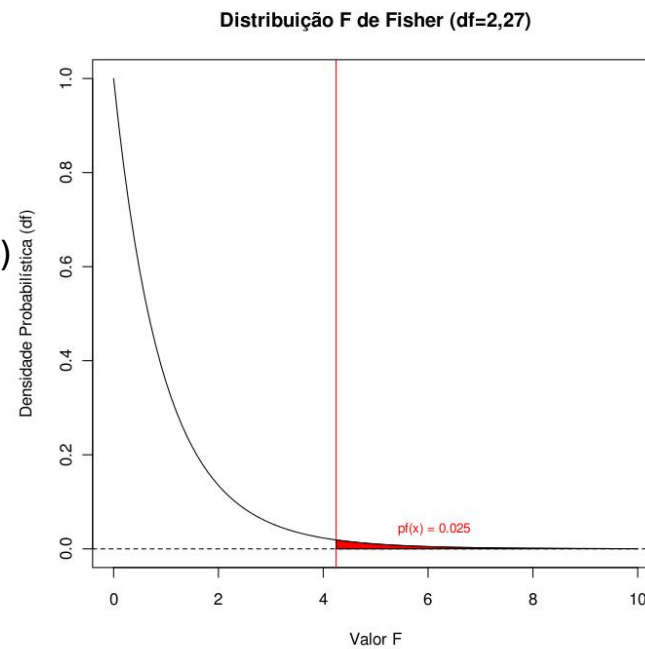
Estatística F

```
(ss_entre/2) / (ss_intra/27)
```

```
## [1] 4.244691
```

```
pf(fanova, df1=2, df2=27, lower.tail = FALSE)
```

```
## [1] 0.02495065
```



Finalizando a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias (F)	P-valor
Entre Grupos	99.2	2	49.6	4.24	0.025
Intra Grupos	315.5	27	11.7		
TOTAL	414.7	29			

Lógica da Anova

$$SS_{total} = SS_{entre} + SS_{intra}$$

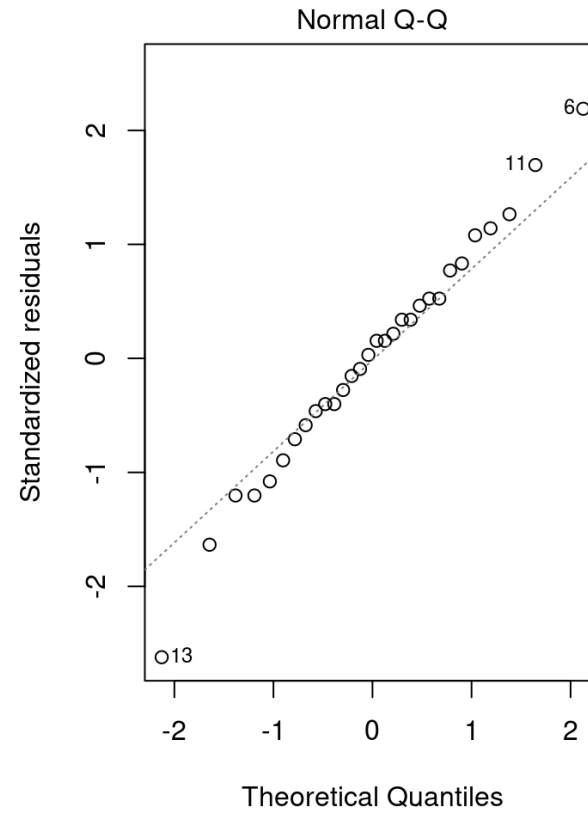
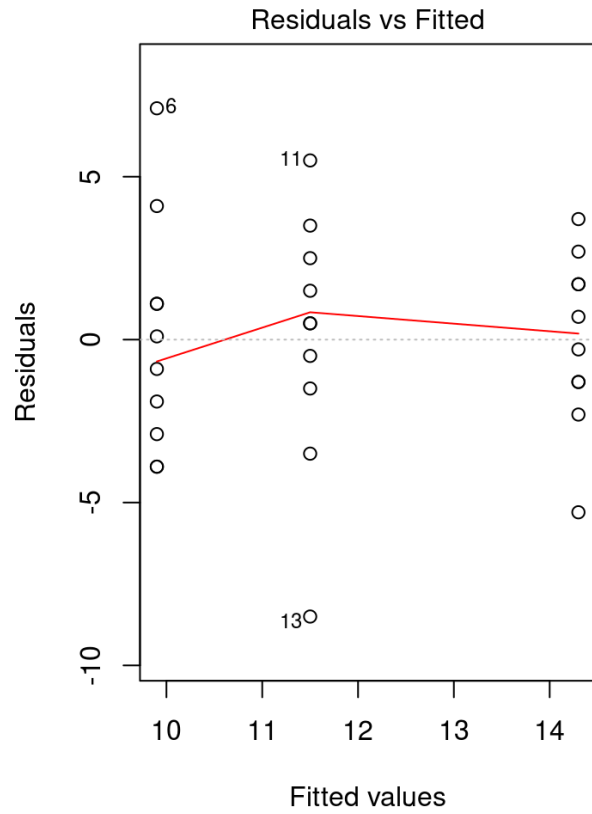
```
(varExpl = ss_entre/ss_total)
```

```
## [1] 0.2392091
```

```
anovaCrop <- aov(colhe~solo, data=crop)
summary(anovaCrop)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## solo           2    99.2   49.60   4.245  0.025 *
## Residuals     27   315.5   11.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova diagnóstico



Constant Leverage:

Atividade