

BIE5782

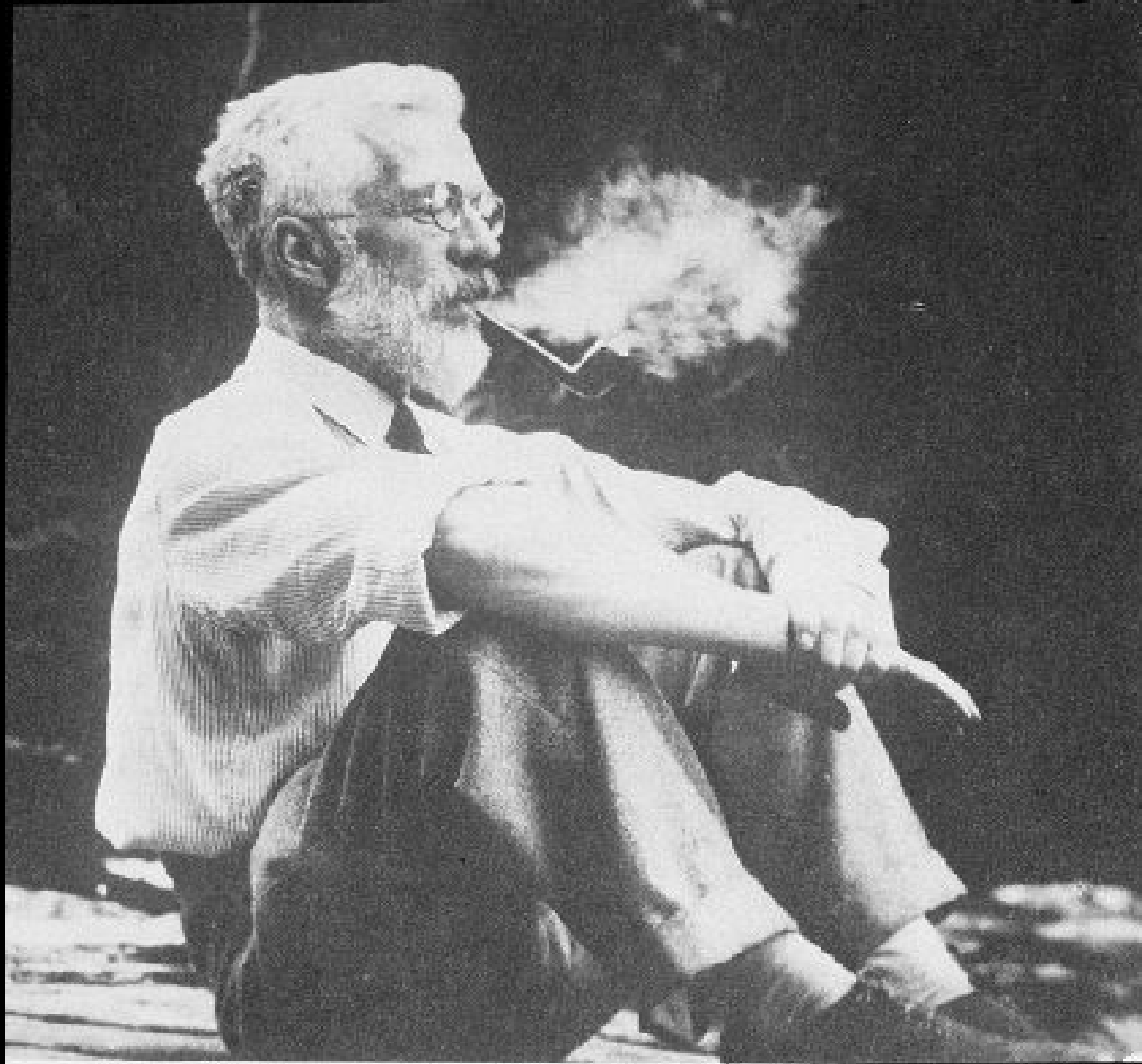
Aula 6:

Estadística Paramétrica e Simulações

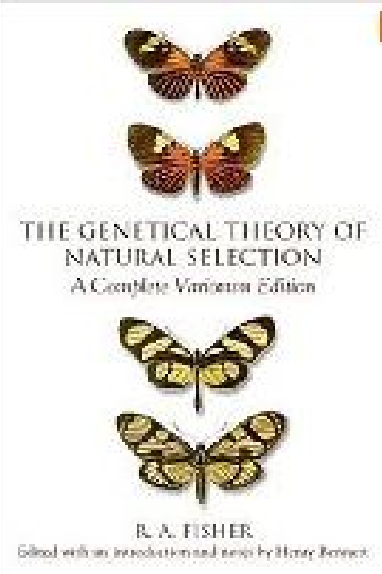
DESAFIOS

1. entender a lógica do teste
2. construir um teste
3. simular cenários nulos

A lógica do teste de significância!



Sir Ronald Fisher



THE GENETICAL THEORY OF
NATURAL SELECTION
A Complete Variorum Edition



R. A. FISHER

Edited with an introduction and notes by Hilary Kimmick

“...talvez o mais original cientista matemático do século 20...”

Bradley Efron (1976)

“Fisher era um gênio que criou praticamente sozinho todos os fundamentos da moderna ciência da estatística”

Anders Hald *A History of Mathematical Statistics* (1998)

“Sir Ronald Fisher ... pode ser reconhecido como o mais importante sucessor de Darwin no século 20”

Richard Dawkins *River out of Eden* (1995)

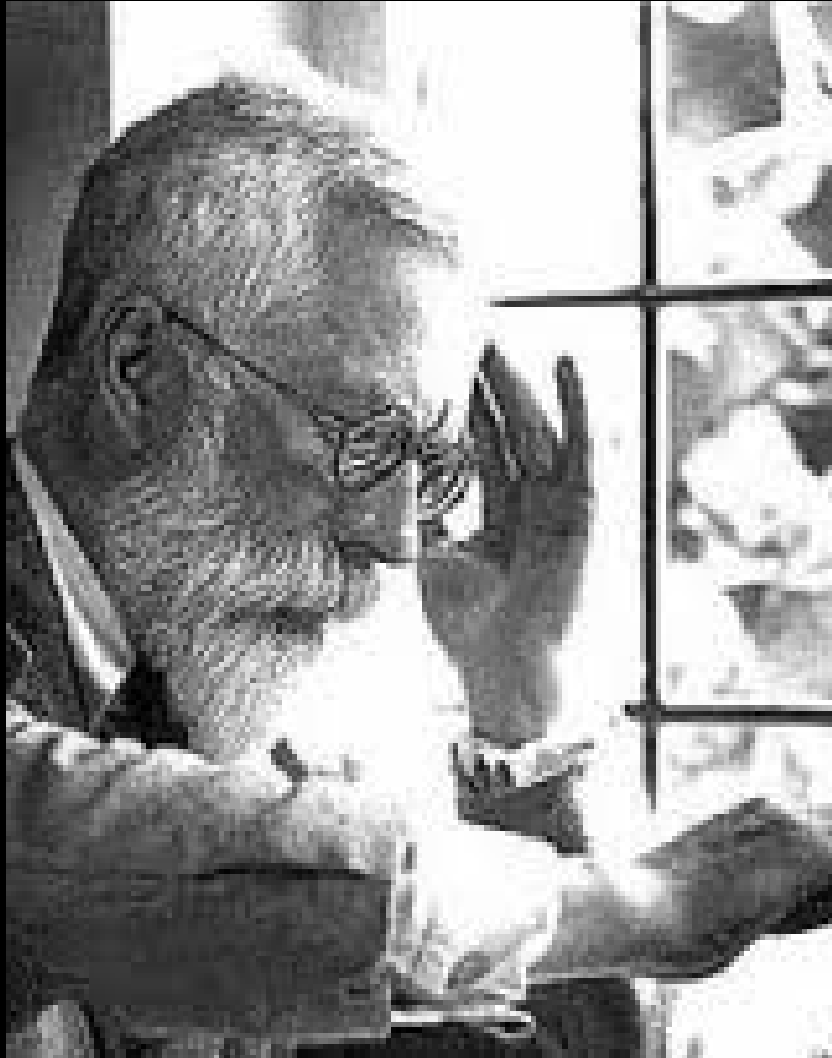
“ Eu frequentemente encontro geneticistas que me perguntam se é verdade que o grande geneticistas R. A. Fisher era também um grande estatístico”

Leonard J. Savage *Annals of Statistics* (1976)

“Kolmogorov certa vez em uma palestra referiu-se à obra *The genetical theory of natural selection* como —‘o maravilhoso livro de R. A. Fisher.’ Dois matemáticos dos EUA sentados ao meu lado sussurraram ‘Não pode ser o R. A Fisher que conhecemos’”

David Kendall *Bulletin of the London Mathematical Society* (1990).

Sir Ronald Fisher



- deriva genica
- eugenia
- relação entre tabaco e câncer de pulmão

Precisamos da Estatística?



Inferência Estatística

“...objetivo fazer afirmações sobre um universo a partir de um conjunto de valores representativo (amostra). Tal tipo de afirmação deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade.”

Inferência Estatística

“...objetivo fazer afirmações sobre um universo a partir de um conjunto de valores representativo (amostra). Tal tipo de afirmação deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade.”

Palmitos – *Euterpe edulis*



Similar no



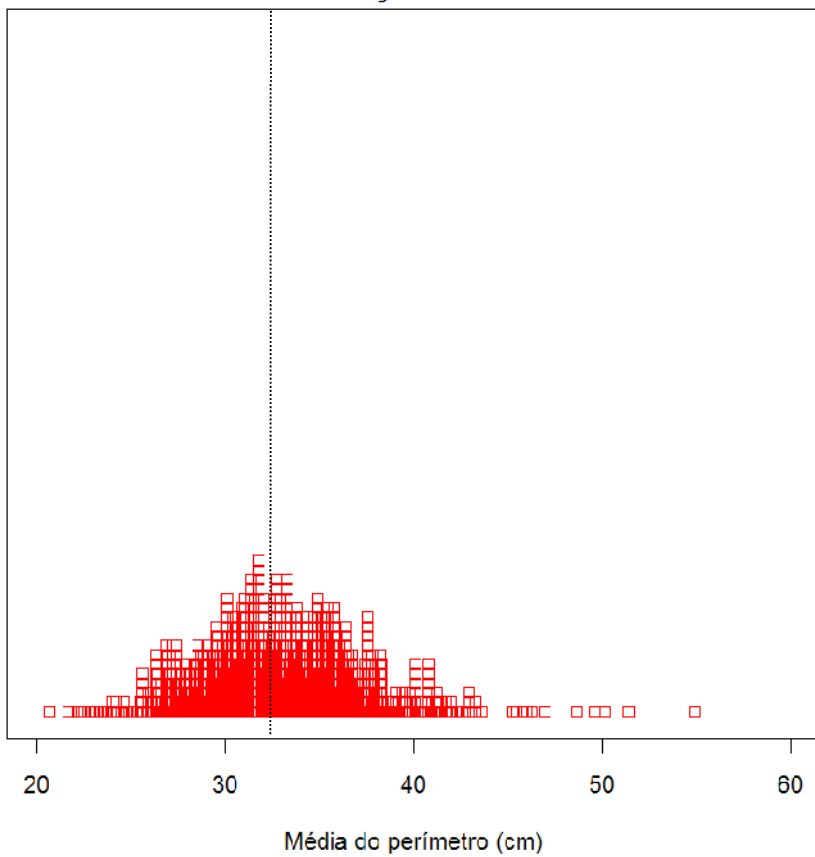


Palmitos PECB

Simulação do perímetro médio do Palmito no PECB

Amostra = 10

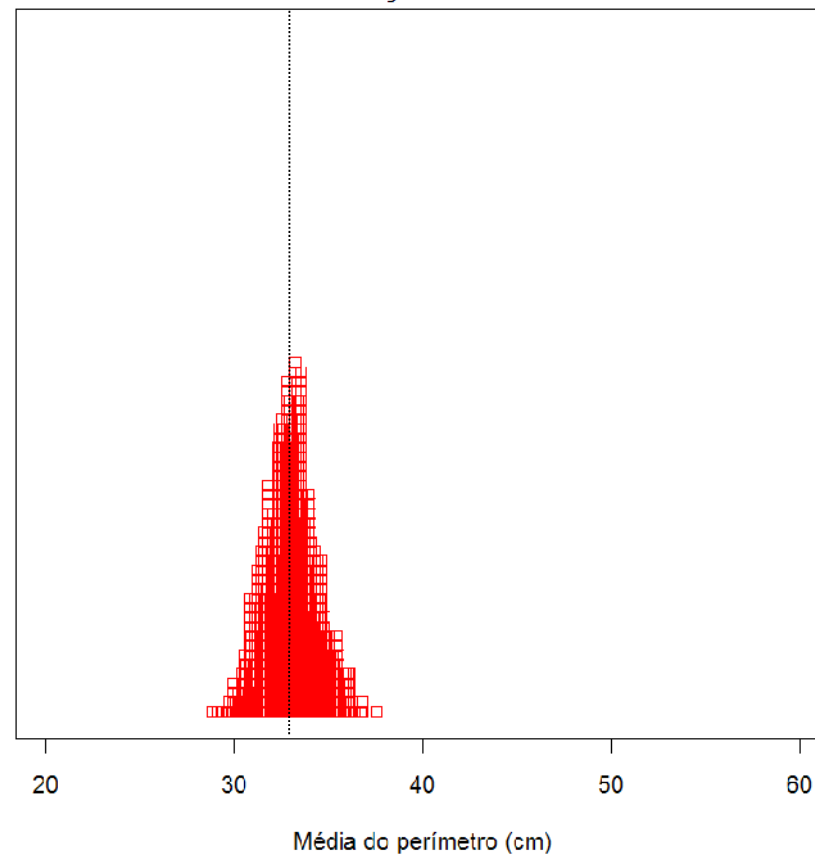
simulação no. 1000



Simulação do perímetro médio do Palmito no PECB

Amostra = 100

simulação no. 1000



Chacal Dourado (*Canis aureus*)



Higham et al. 1980

Dados

Tamanho da mandíbula em mm

```
> macho=c(120,107,110,116, 114, 111, 113,  
117, 114,112)
```

```
>femea=c(110,111,107,108,110,105,107,106,111  
, 111)
```

```
> media.m=mean(macho)
```

```
> media.m
```

```
[1] 113.4
```

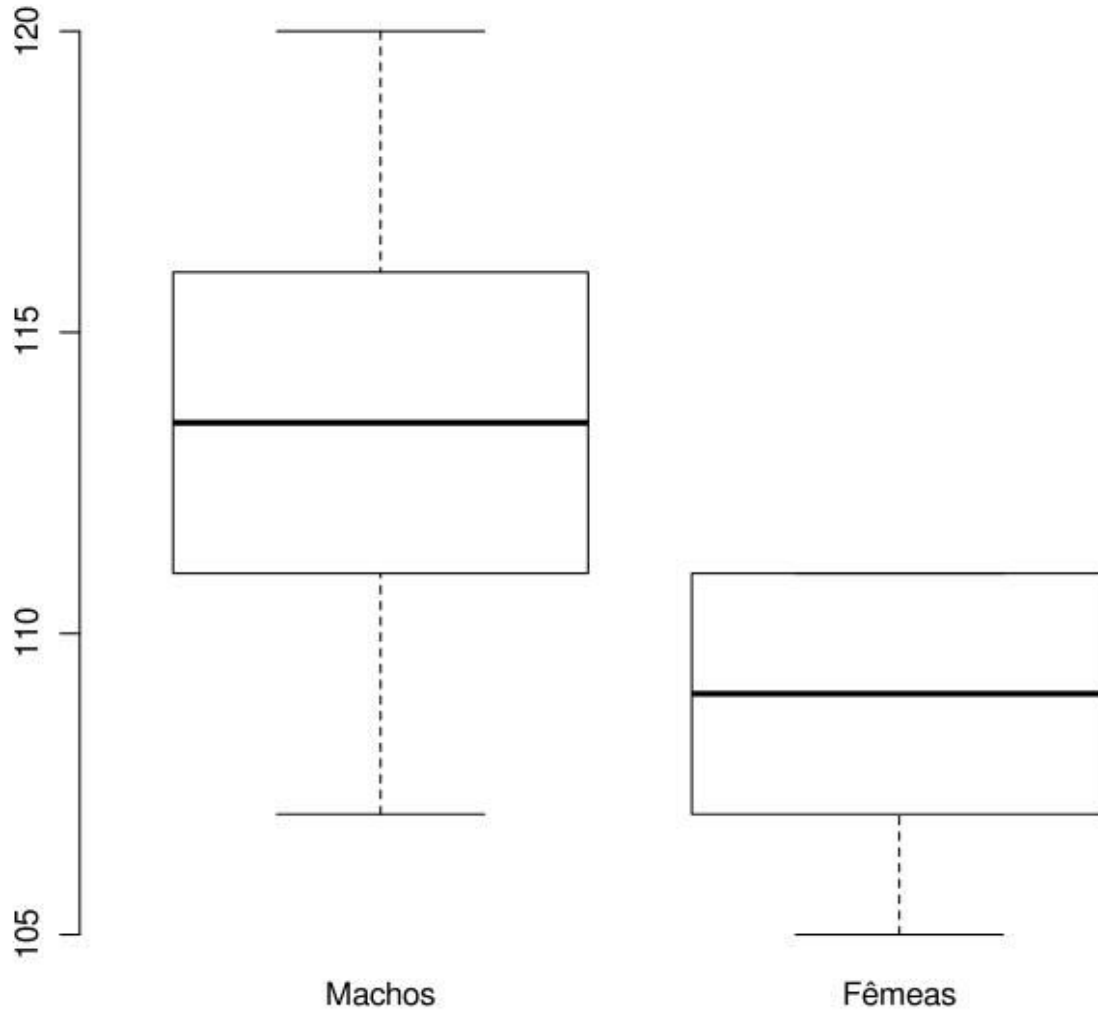
```
> media.f=mean(femea)
```

```
> mean(femea)
```

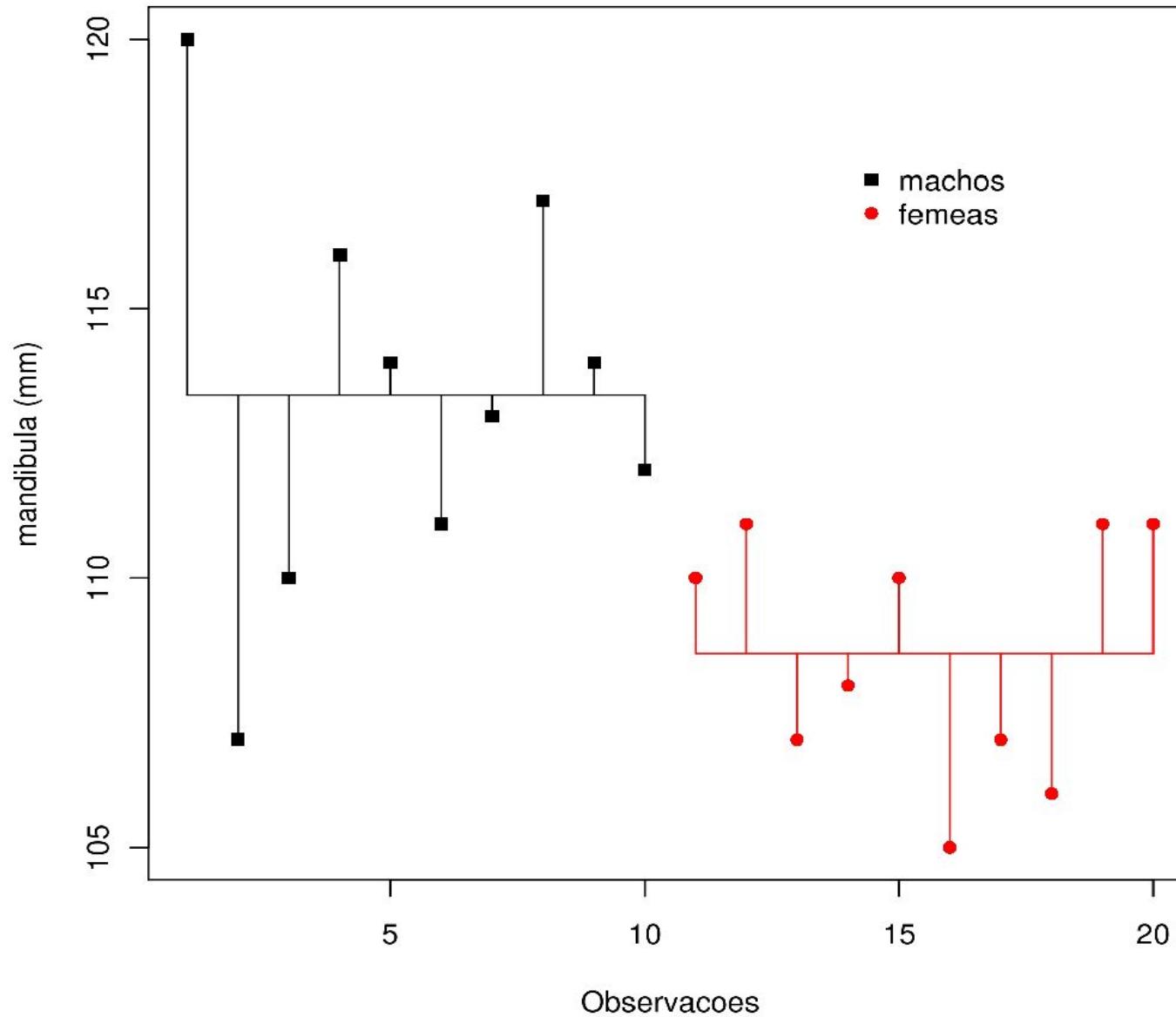
```
[1] 108.6
```

boxplot ()

Tamanho da mandibula de chagal



Variação nos dados



Vamos ao



Dados

```
> media.m
```

```
[1] 113.4
```

```
> media.f
```

```
[1] 108.6
```

```
>mean (macho) -mean (femea)
```

```
[1] 4.8
```

```
>mean (femea) -mean (macho)
```

```
[1] -4.8
```

```
>round (abs (mean (femea) -mean (macho) ) , 1)
```

```
[1] 4.8
```

INFERÊNCIA

“..fazer afirmações sobre um universo a partir de um conjunto de valores representativo..”

Machos e fêmeas de chacal apresentam tamanhos de mandíbulas diferentes, em média.

INFERÊNCIA

“..fazer afirmações sobre um universo a partir de um conjunto de valores representativo..”

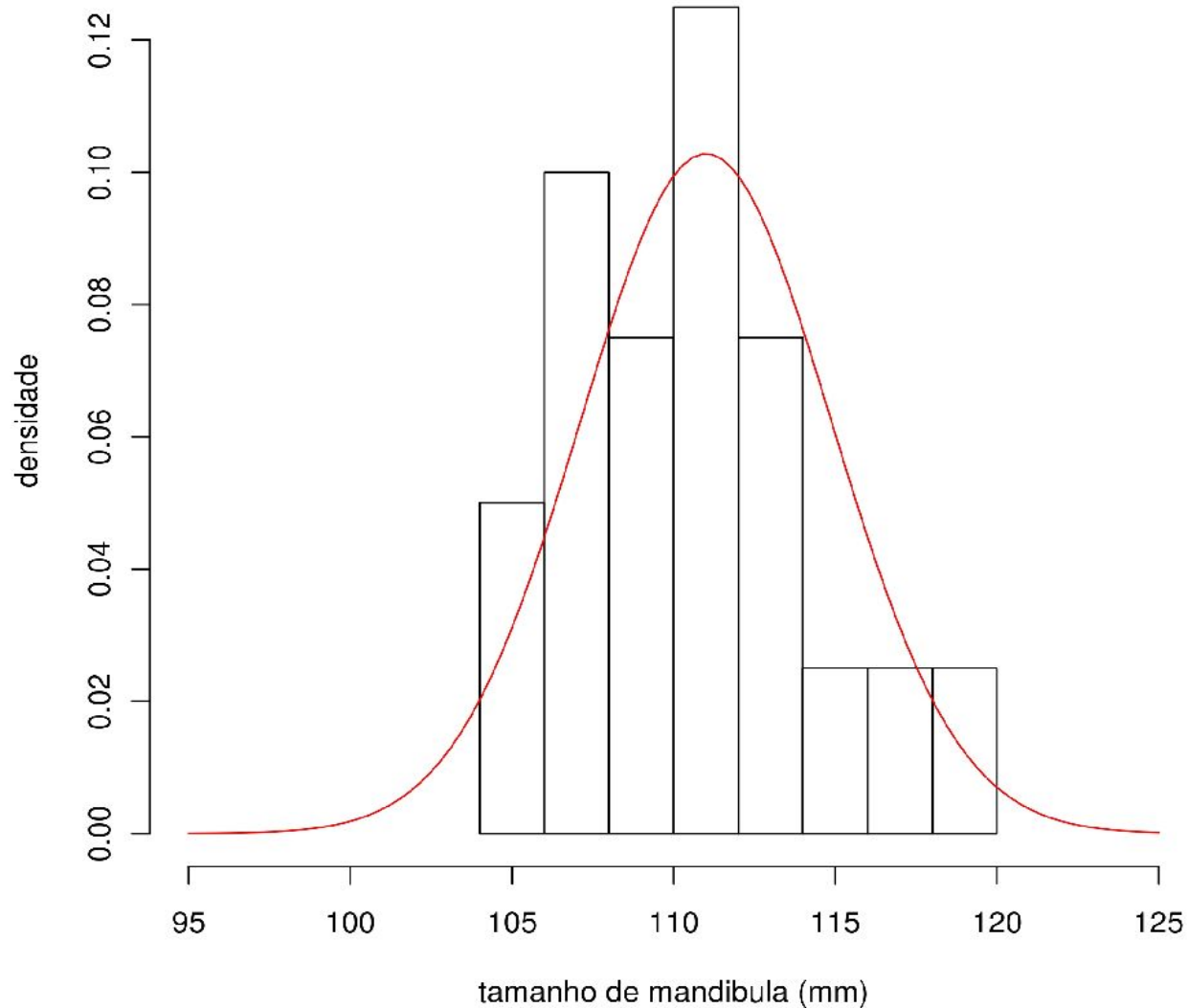
Machos e fêmeas de chacal apresentam tamanhos de mandíbulas diferentes, em média.

4,8 mm

“..deve sempre vir acompanhada de uma medida de precisão sobre sua veracidade. ”

Pressupostos

Distribuição de tamanho de mandibulas de Chacal



```
>curve(exp=dnorm(x, mean=mean(chacal$tam.mand),sd=sd(chacal$tam.mand)),  
+ from=95,to=125, col="red", add=T)
```

Simulando um cenário

```
> rnorm(10, mean=mean(chacal), sd=sd(chacal))
```

```
[1] 107.8909 113.4513 109.2128 114.2411 114.0735  
113.5735 112.7718
```

```
> round(rnorm(10, mean=mean(chacal), sd=sd(chacal)))
```

```
[1] 105 115 109 115 115 110 114 108 108 110
```

```
> abs(round(rnorm(10, mean=mean(chacal),  
+ sd=sd(chacal))))
```

```
[1] 108 117 115 109 109 110 112 116 110 116
```

Simulando um cenário

```
>abs (round (mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) )  
- mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) ) ) )  
[1] 1
```

```
>abs (round (mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) )  
-mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) ) ) )  
[1] 2
```

```
>abs (round (mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) )  
- mean (rnorm (10 ,mean=mean (chacal) , sd=sd (chacal) ) ) ) )  
[1] 3
```

for() {}

```
> for(i in 1:10)
  {
    cat("\n\t", i)
  }

> resulta=rep(NA,10)

> for (i in 1:10)
  {
    + resulta[i]=
    + abs(round
(mean(rnorm(10,mean=mean(chacal),sd=sd(chacal)))-
mean(rnorm(10,mean=mean(chacal),sd=sd(chacal))))))
  }

> resulta
[1] 0 1 1 3 1 1 3 2 1 2
```

Perguntas

Os tamanhos de mandíbulas de machos e fêmeas de chacal são diferentes?

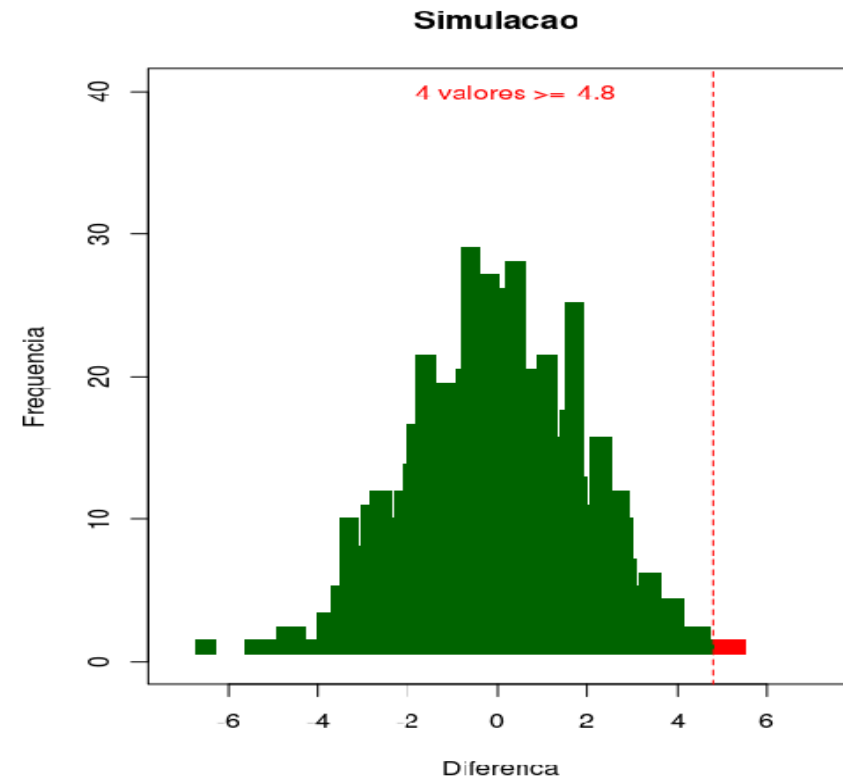
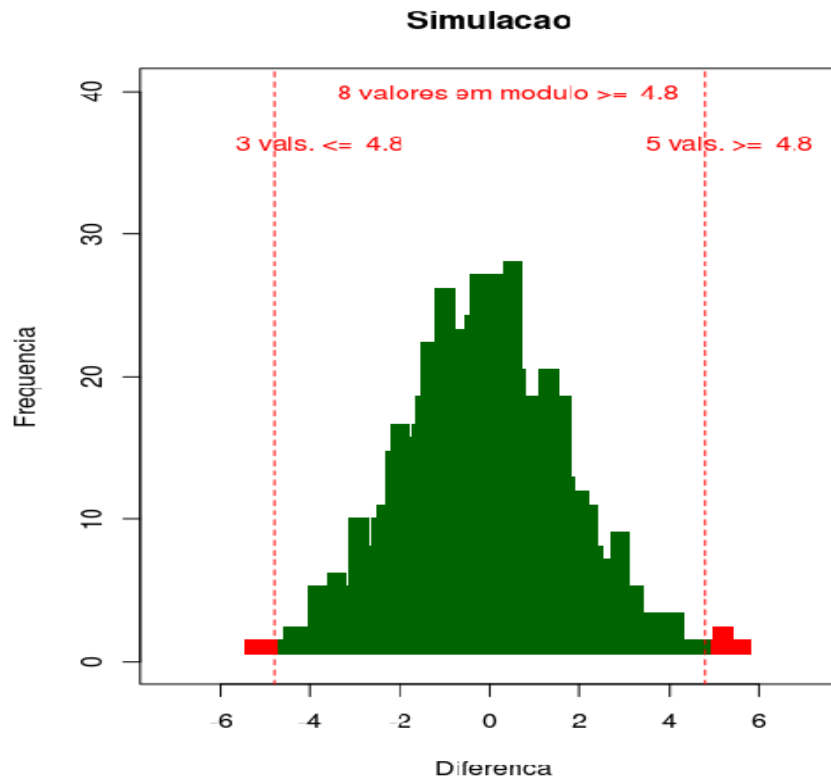
Chacal machos tem mandíbulas maiores que as fêmeas?

Qual a probabilidade de erro ao afirmar que existe diferença ?

Vamos ao

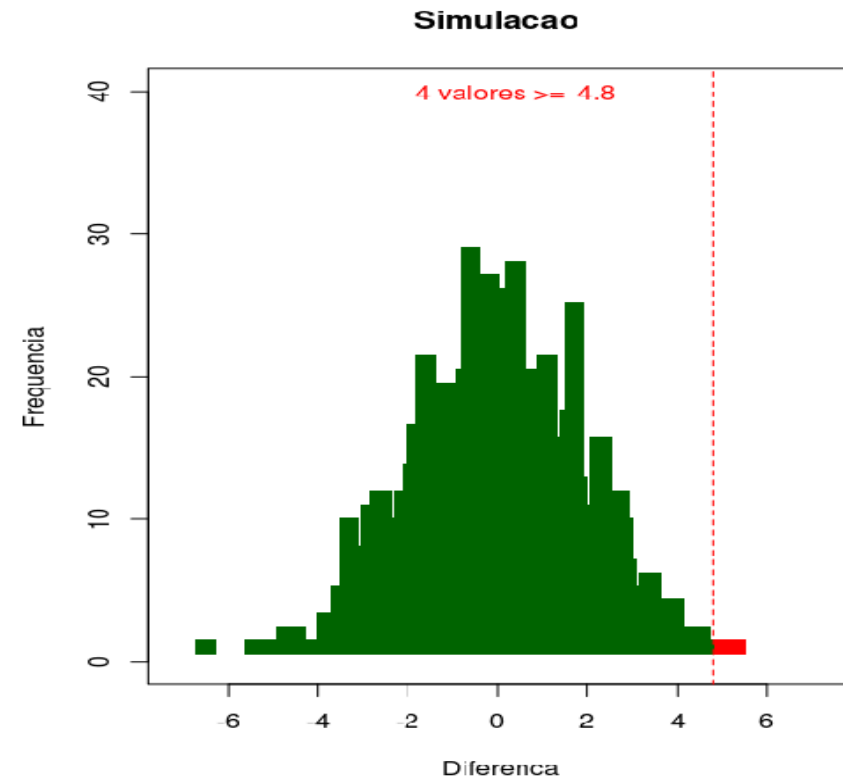
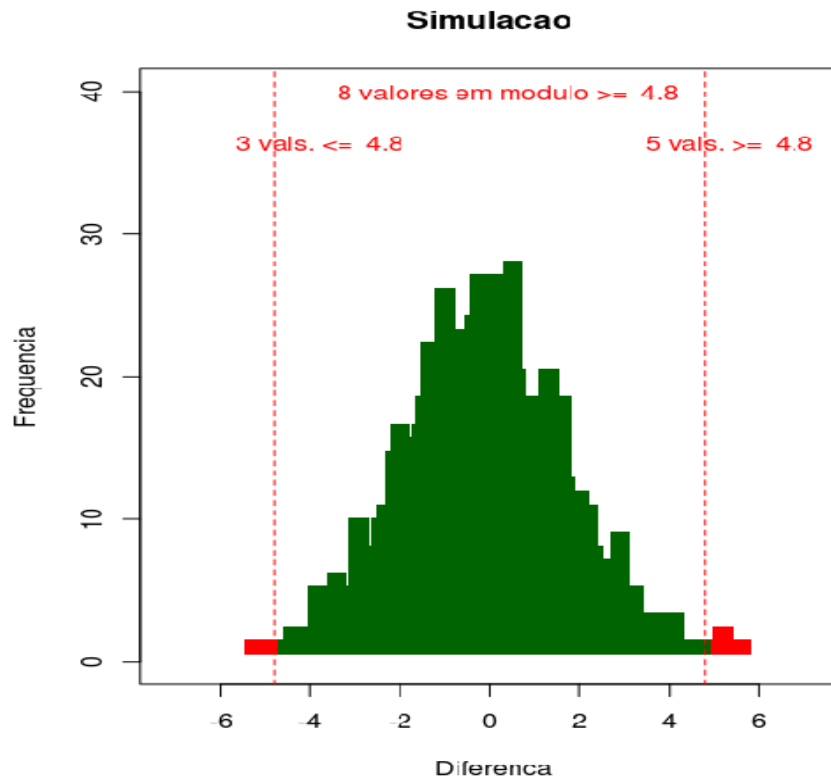


Os tamanhos de mandibulas de machos e fêmeas de chacal são diferentes?



Qual a probabilidade de erro ao afirmar que existe diferença ?

Chacal machos tem mandíbulas maiores que as fêmeas?



Qual a probabilidade de erro ao afirmar que existe diferenca ?

Análise de Variância



“The best way to see what is happening is to work through a simple example. We have an experiment in which crop yields per unit area were measured from 10 randomly selected fields on each of three soil types. All fields were sown with the same variety of seed and provided with the same fertilizer and pest control inputs. The question is whether soil type significantly affects crop yield, and if so, to what extent.”

Roberto Crawley – The R Book

Análise de Variância

produção em diferentes solos

```
> are=c(6,10,8,6,14,17,9,11,7,11)
> are
[1] 6 10 8 6 14 17 9 11 7 11
> arg=c(17,15,3,11,14,12,12,8,10,13)
> arg
[1] 17 15 3 11 14 12 12 8 10 13
> hum=c(13,16,9,12,15,16,17,13,18,14)
> hum
[1] 13 16 9 12 15 16 17 13 18 14
```

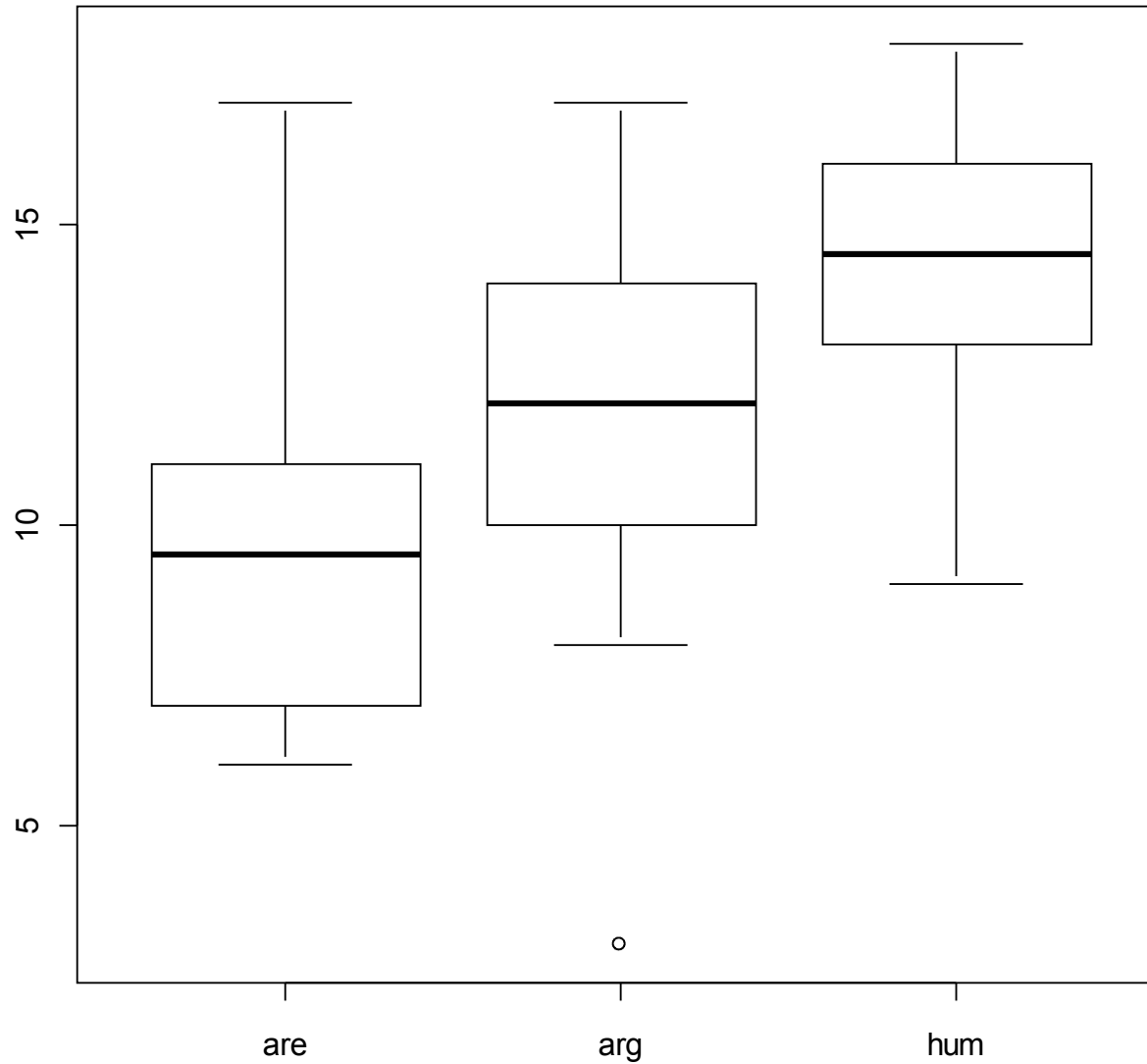
Análise de Variância

```
> solos=data.frame(are, arg, hum)
```

```
> solos
```

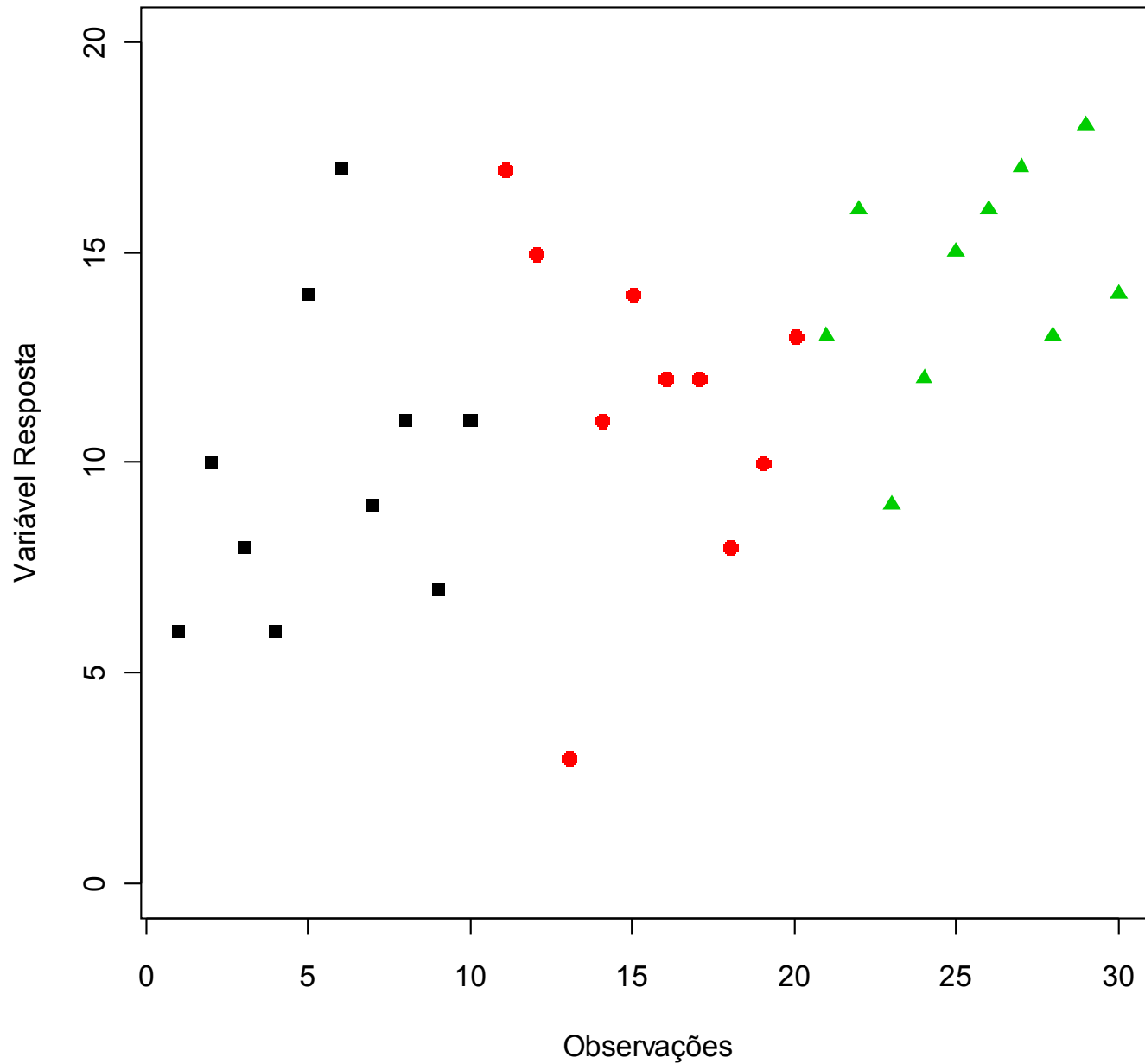
	are	arg	hum
1	6	17	13
2	10	15	16
3	8	3	9
4	6	11	12
5	14	14	15
6	17	12	16
7	9	12	17
8	11	8	13
9	7	10	18
10	11	13	14

Análise de Variância



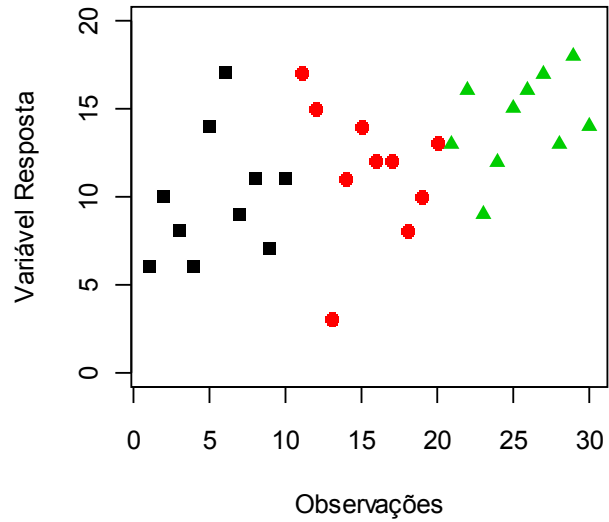
Análise de Variância

Efeito do Solo

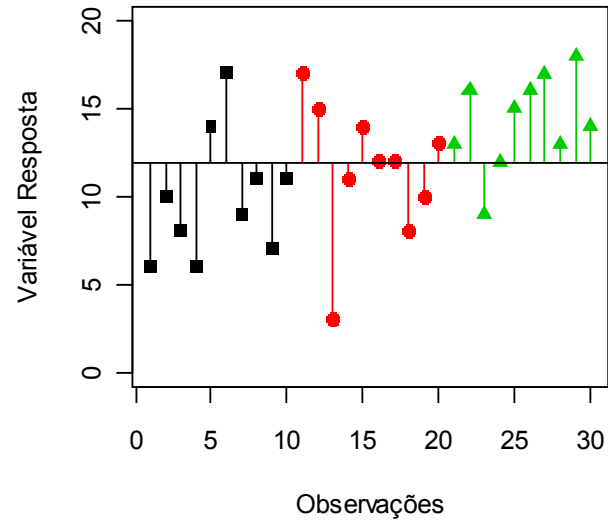


Análise de Variância

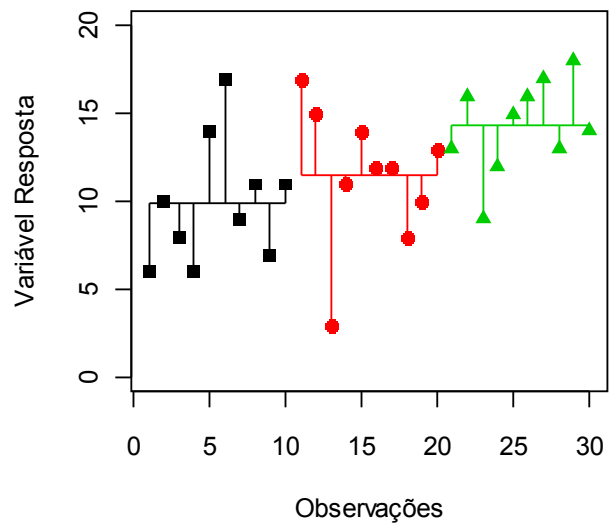
Efeito do Solo



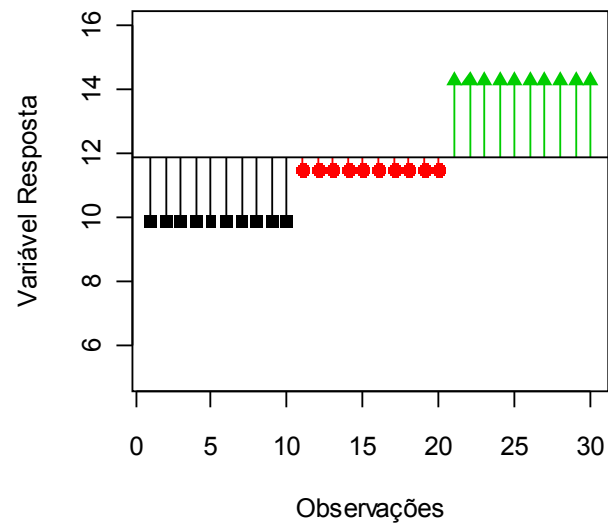
Variação Entre Grupos



Variação Intra Grupos



Variação Entre Grupos



Análise de Variância

$$F = \frac{\text{Variância entre níveis}}{\text{Variância dentro dos níveis}}$$

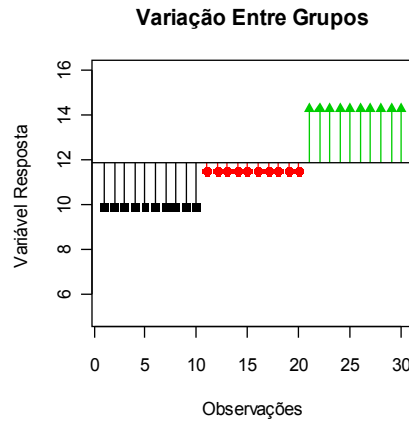
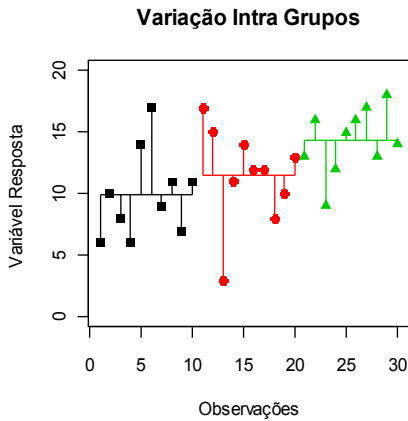
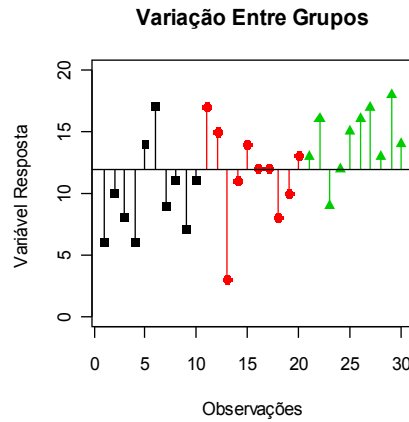
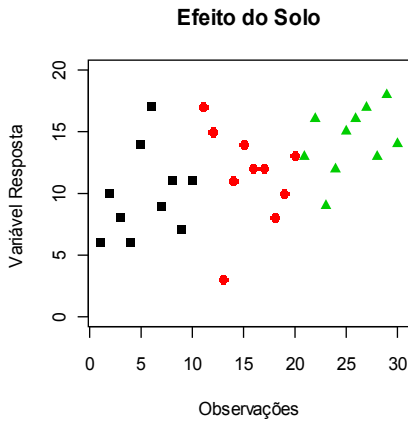


Tabela Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	Probabilidade
Entre Grupos					
Intra Grupos					
TOTAL	X				

VARIAÇÃO TOTAL

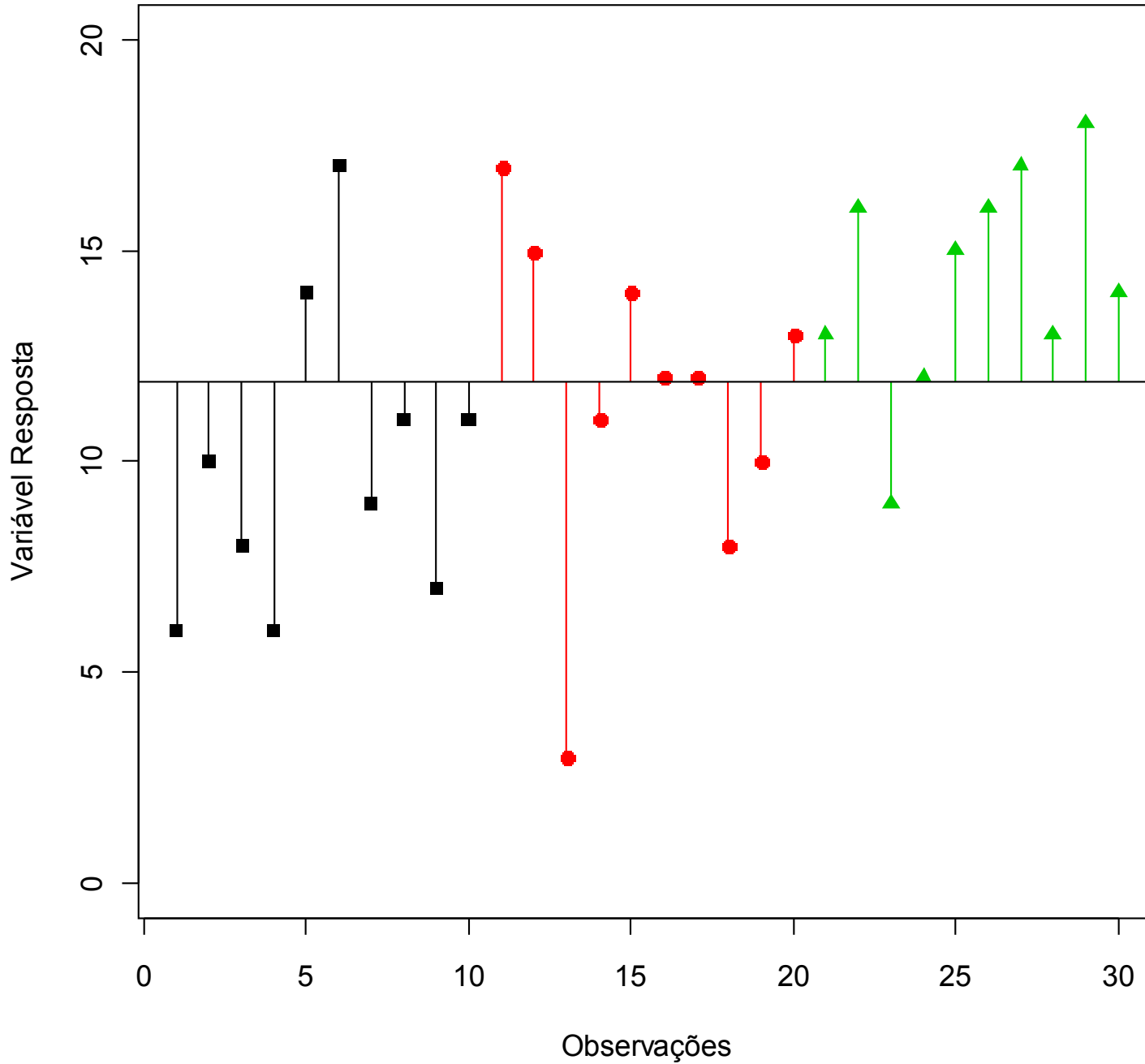


Tabela Anova

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

Varição Intra Grupos

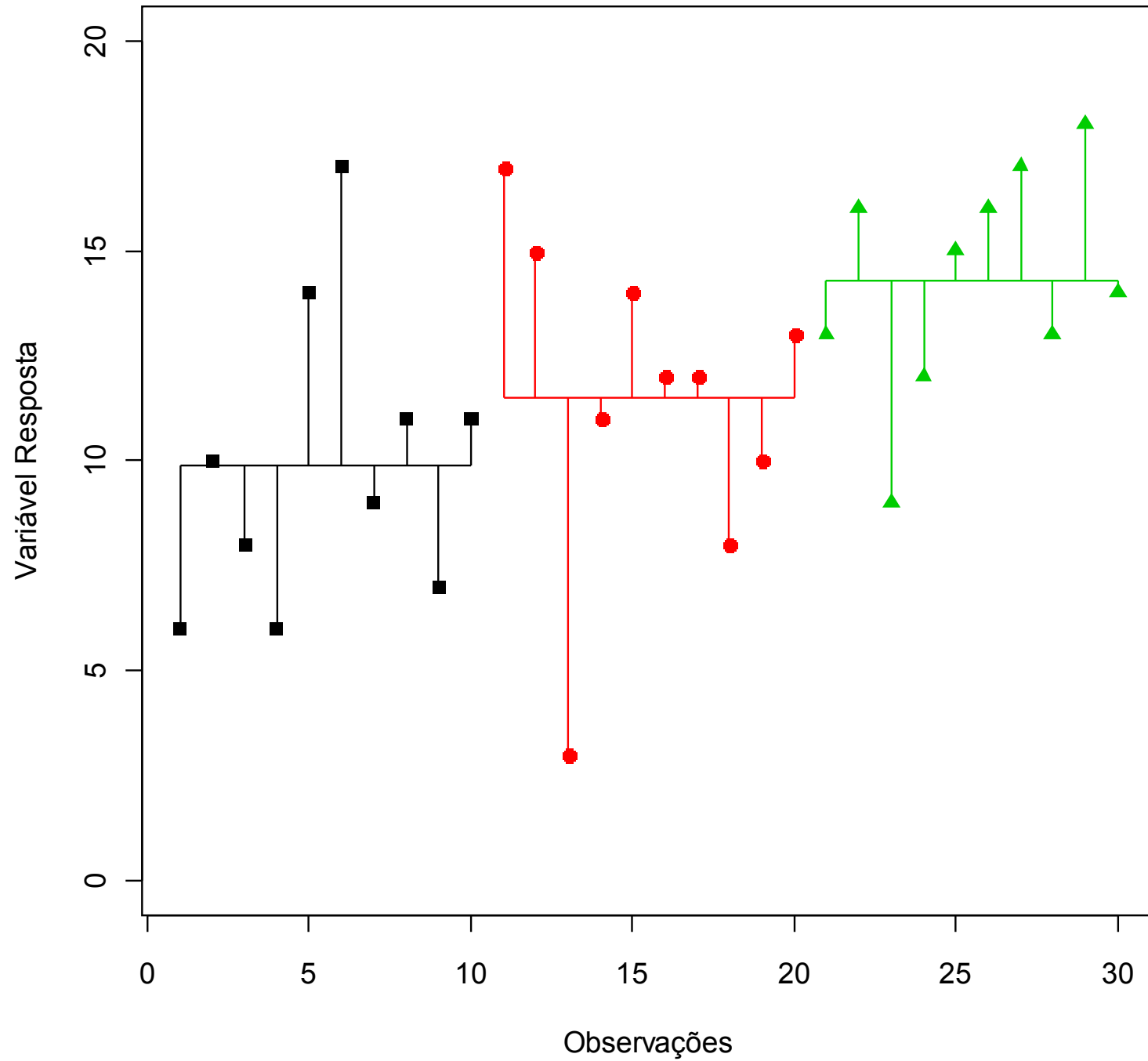


Tabela Anova

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

Variação Entre Grupos

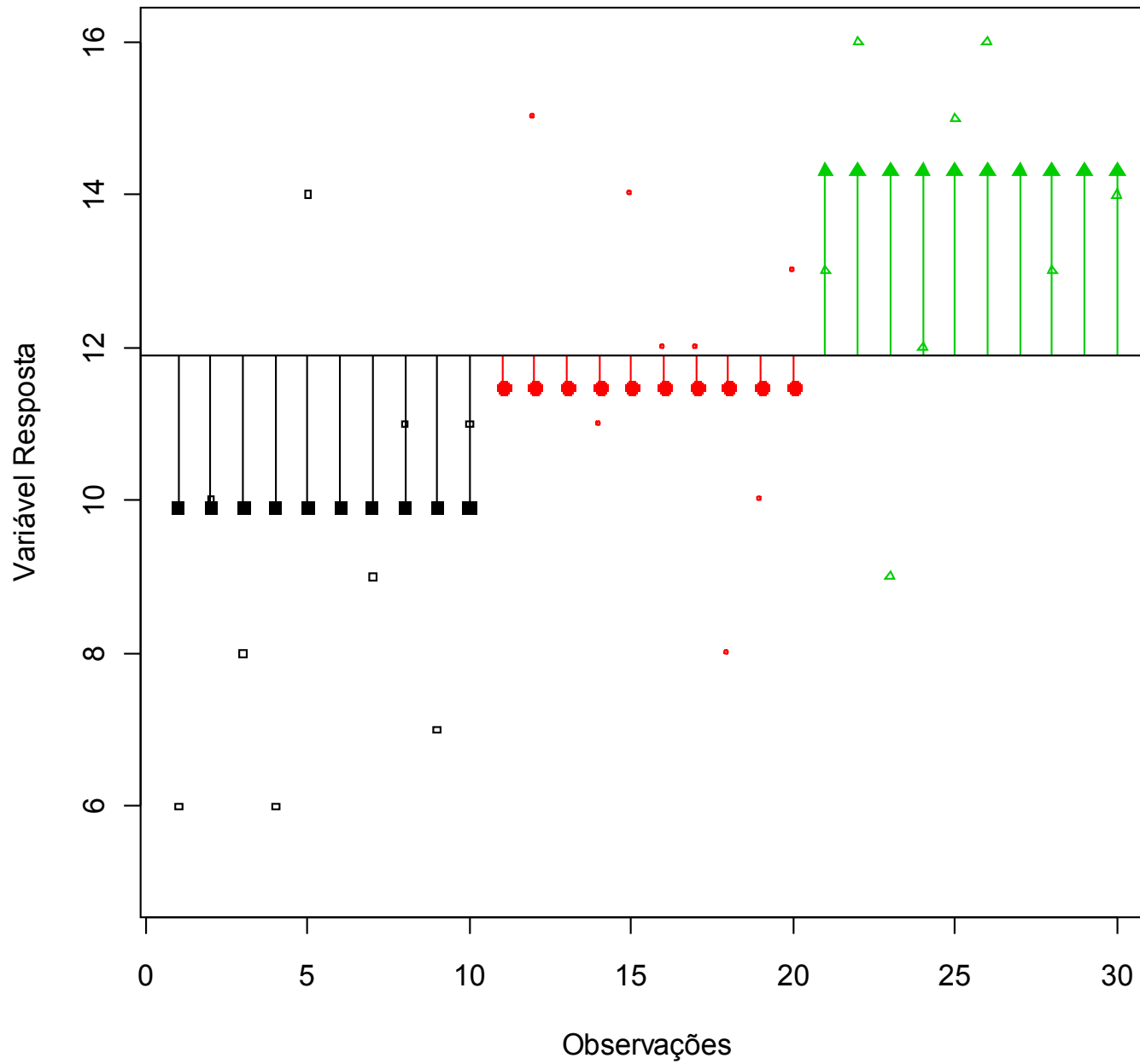


Tabela Anova

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

Tabela Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	Probabilidade
Entre Grupos	99.2				
Intra Grupos	315.5				
TOTAL	414.7	X			

Tabela Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	Probabilidade
Entre Grupos	99.2				
Intra Grupos	315.5	X			
TOTAL	414.7	29			

Tabela Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	Probabilidade
Entre Grupos	99.2	X			
Intra Grupos	315.5	27			
TOTAL	414.7	29			

Tabela Anova

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

Tabela Anova

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

Tabela Anova

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

Distribuição F de Fisher (df=2,27)

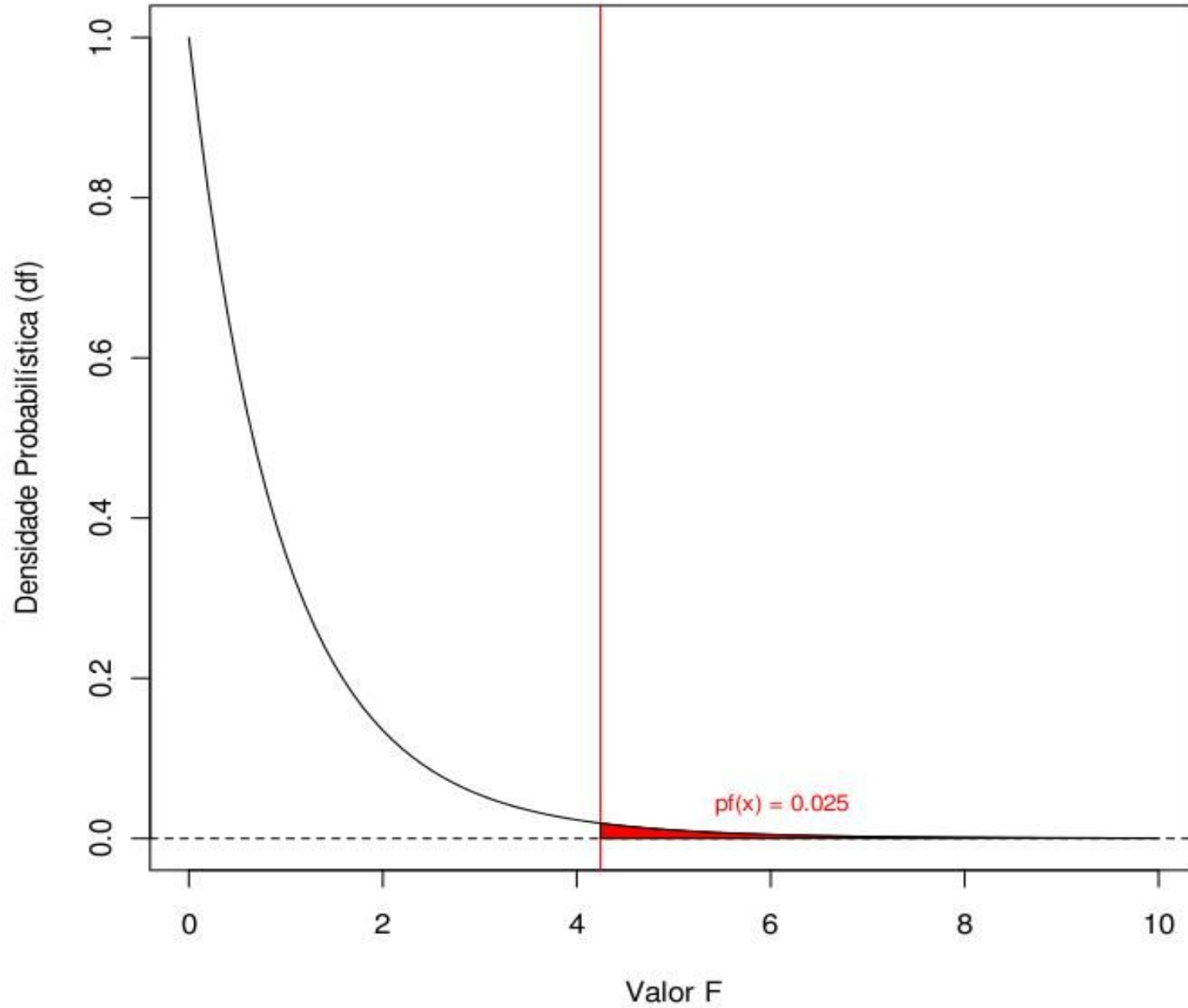
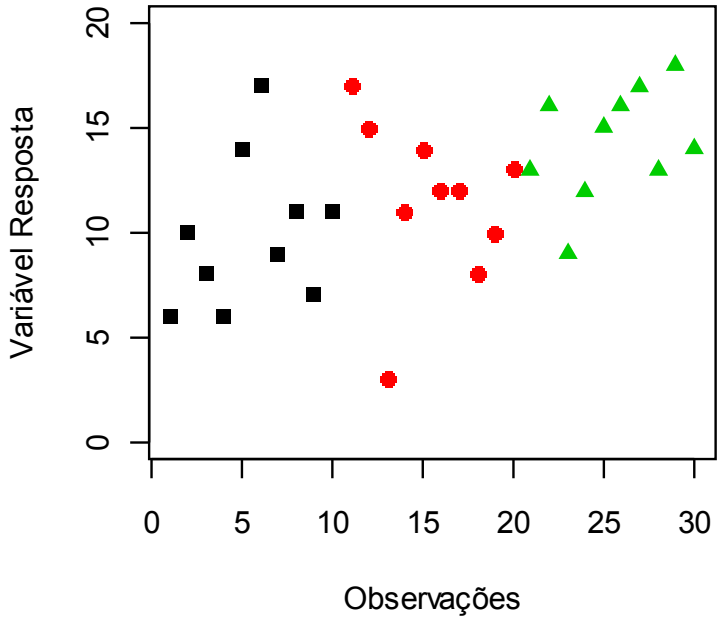


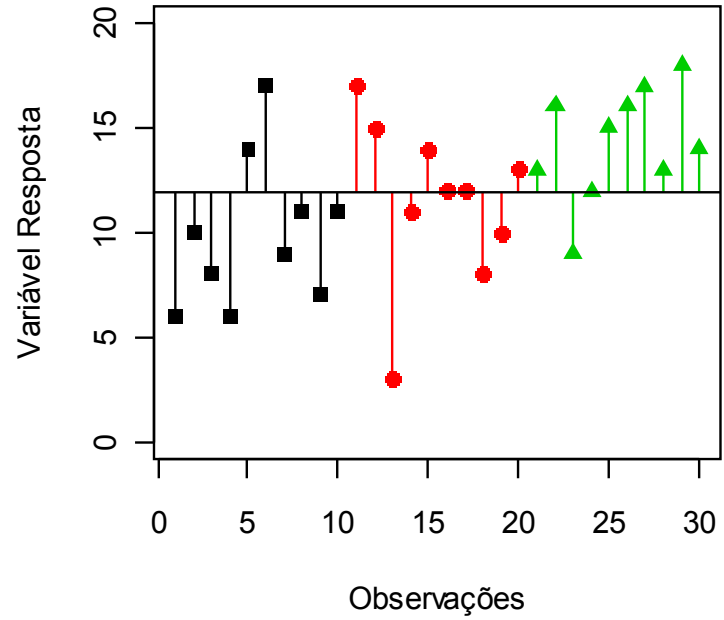
Tabela Anova

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

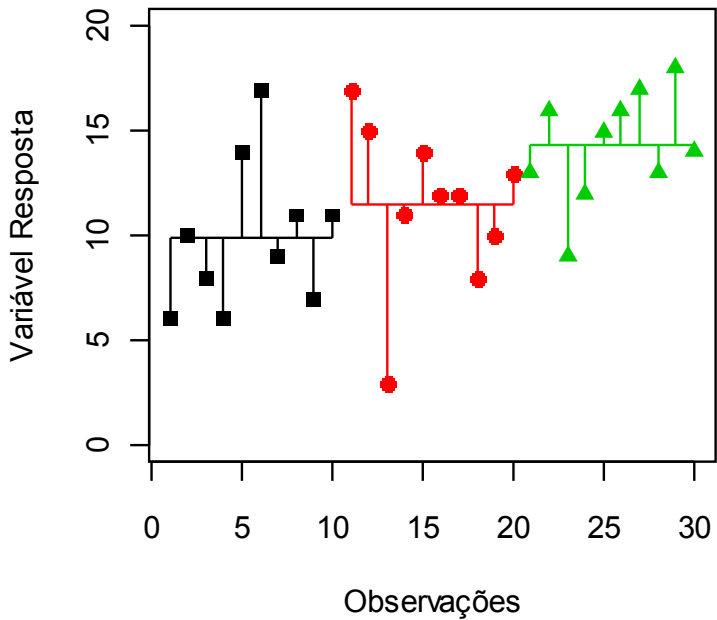
Efeito do Solo



Variação Total



Variação Intra Grupos



Variação Entre Grupos

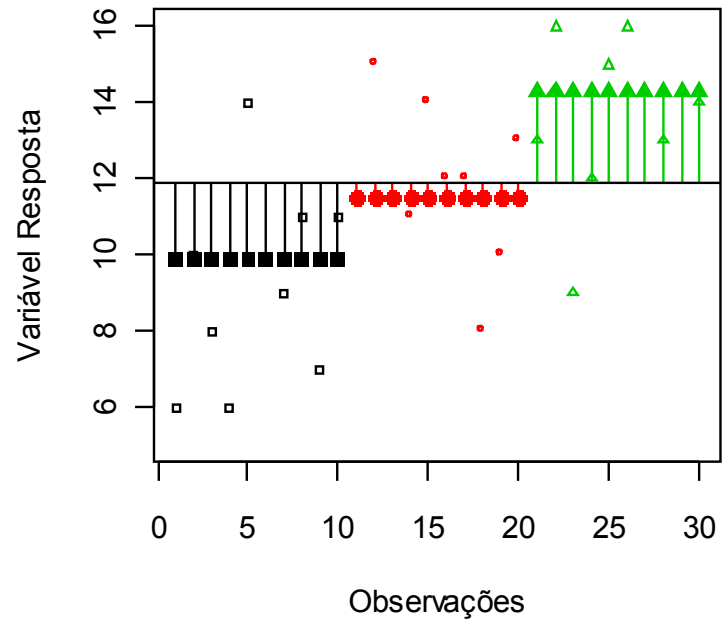
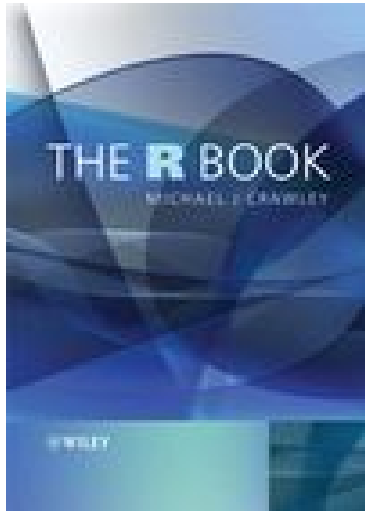


Tabela Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias (F)	Probabilidade de
Entre Grupos	99.2 (24%)	2	49.6	4.24	0.025
Intra Grupos	315.5	27	11.7		
TOTAL	414.7	29			

Anova no





FIM DA UNIDADE 6

Para a tarde:

Plantão Tutoriais e exercícios Unidade 6

Até sexta:

Lista 6 de Exercícios: