

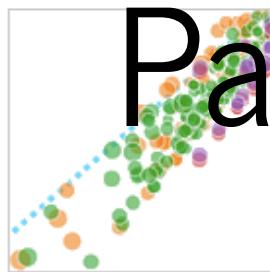
# Estatística Frequentista

## Análise de Variância

Alexandre Adalardo de Oliveira

PlanECO 2025

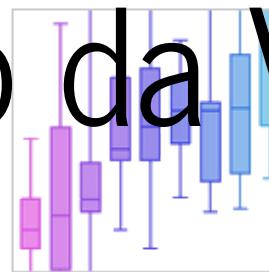
Line and Scatter Plots



Bar Charts



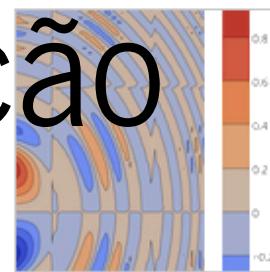
Box Plots



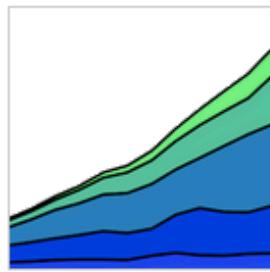
Bubble Charts



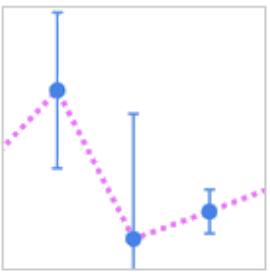
Contour Plots



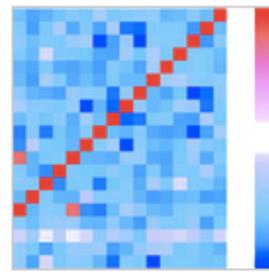
Filled Area Plots



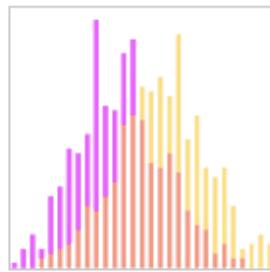
Error Bars



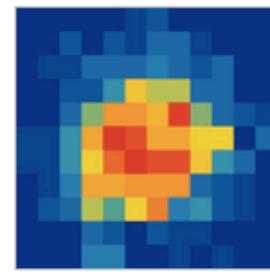
Heatmaps



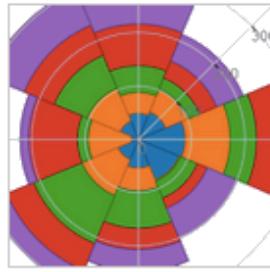
Histograms



2D Histograms



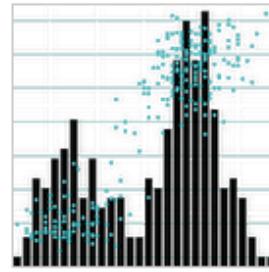
Polar Charts



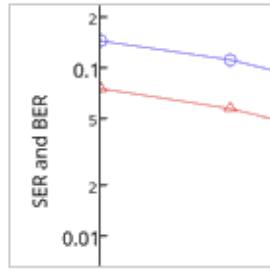
Time Series



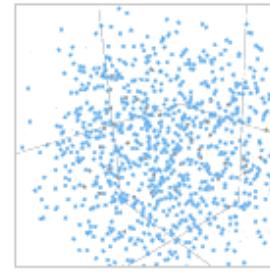
Multiple Chart Types



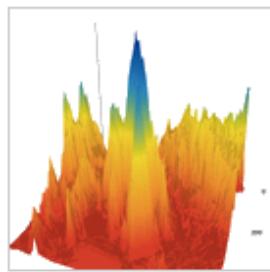
Log Plots



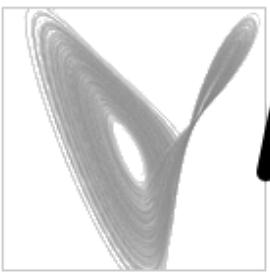
3D Scatter Plots



3D Surface Plots



3D Line Plots



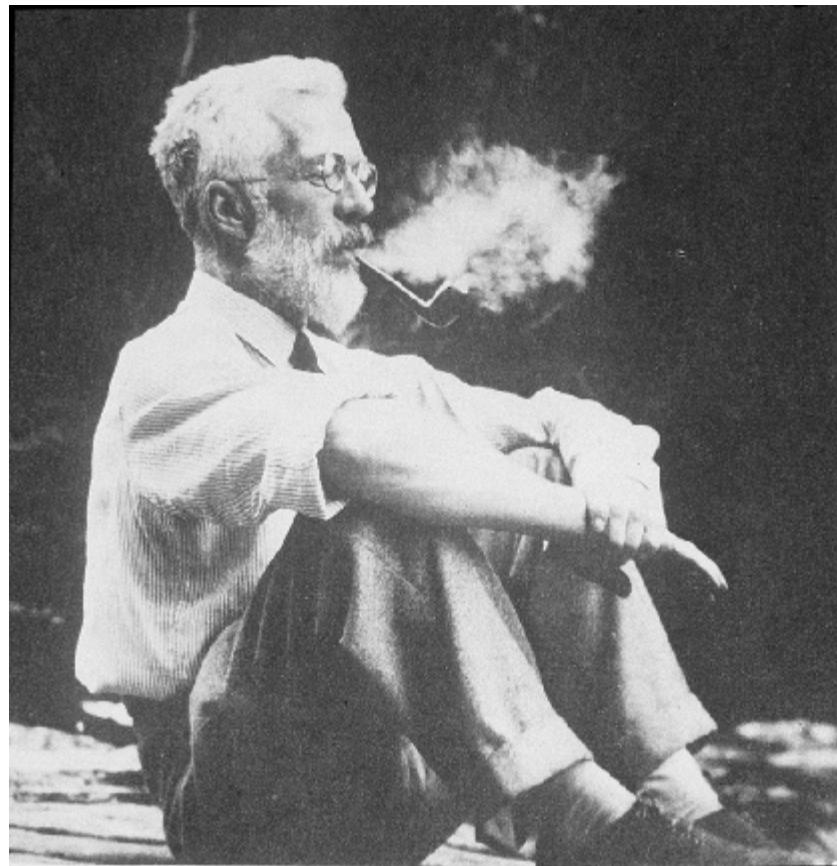
# PIAnEco

# Conteúdo

- partição da variabilidade
- razão de variâncias
- tabela de anova

# Histórico

## Ronald Fisher



- desenho experimental
- teste associado

# Testes Clássicos

Resposta	Preditoras	Teste	Hipótese
Categórica	Categórica	Qui-quadrado	independência
Contínua	Categórica(2)	Teste-t	$\mu_1 = \mu_2$
Contínua	Categórica (>2)	Anova	$\mu_1 = \mu_2 = \mu_3$
Contínua	1 Contínua	Regressão	$\beta_1 = 0$
Contínua	>1 Contínua	Reg. múltipla	$\beta_1 = 0; \beta_n = 0$
Contínua	Cont + Cat	Ancova	$\beta_1 = \beta_2; \alpha_1 = \alpha_2$
Proporção	Contínua	Reg. Logística	$\text{logit}(\beta_1) = 1$

# Anova: um exemplo

Tradução livre da descrição do livro “The R Book” ( [Crawley, 2007](#))



“... a melhor forma de entender o que está acontecendo é trabalharmos um exemplo. Temos um experimento em que a produção agrícola por unidade de área é medido em 10 campos de cultivo selecionadas aleatoriamente para cada um de tipos de três tipos de solo. Todos os campo foram semeados com a mesma variedade de semente e manejados com as mesmas técnicas (fertilizantes, controle de pragas). O objetivo é verificar se o tipo de solo afeta significativamente o rendimento de culturas, e caso afete, quanto.”<sup>2)</sup>

Roberto Crawley, autor “The R Book”

# Anova: um exemplo

arenoso	argiloso	húmico
6	17	13
10	15	16
8	3	9
6	11	12
14	14	15
17	12	16
9	12	17
11	8	13
7	10	18
11	13	14

Crawley, R. 2007. The R Book.

# Anova: um exemplo

arenoso	argiloso	húmico
6	17	13
10	15	16
8	3	9
6	11	12
14	14	15
17	12	16
9	12	17
11	8	13
7	10	18
11	13	14

Um exemplo **ERRADO!**

# Anova: um exemplo

	<b>solo</b>	<b>colhe</b>
1	are	6
2	are	10
3	are	8
4	are	6
5	are	14
11	arg	17
12	arg	15
13	arg	3
14	arg	11
15	arg	14
21	hum	13

	<b>solo</b>	<b>colhe</b>
22	hum	16
23	hum	9
24	hum	12
25	hum	15

# Dados Tabulares

THE AMERICAN STATISTICIAN  
2018, VOL. 72, NO. 1, 2–10  
<https://doi.org/10.1080/00031305.2017.1375989>



OPEN ACCESS 

## Data Organization in Spreadsheets

Karl W. Broman<sup>a</sup> and Kara H. Woo<sup>b</sup>

<sup>a</sup>Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI; <sup>b</sup>Information School, University of Washington, Seattle, WA

### ABSTRACT

Spreadsheets are widely used software tools for data entry, storage, analysis, and visualization. Focusing on the data entry and storage aspects, this article offers practical recommendations for organizing spreadsheet data to reduce errors and ease later analyses. The basic principles are: be consistent, write dates like YYYY-MM-DD, do not leave any cells empty, put just one thing in a cell, organize the data as a single rectangle (with subjects as rows and variables as columns, and with a single header row), create a data dictionary, do not include calculations in the raw data files, do not use font color or highlighting as data, choose good names for things, make backups, use data validation to avoid data entry errors, and save the data in plain text files.

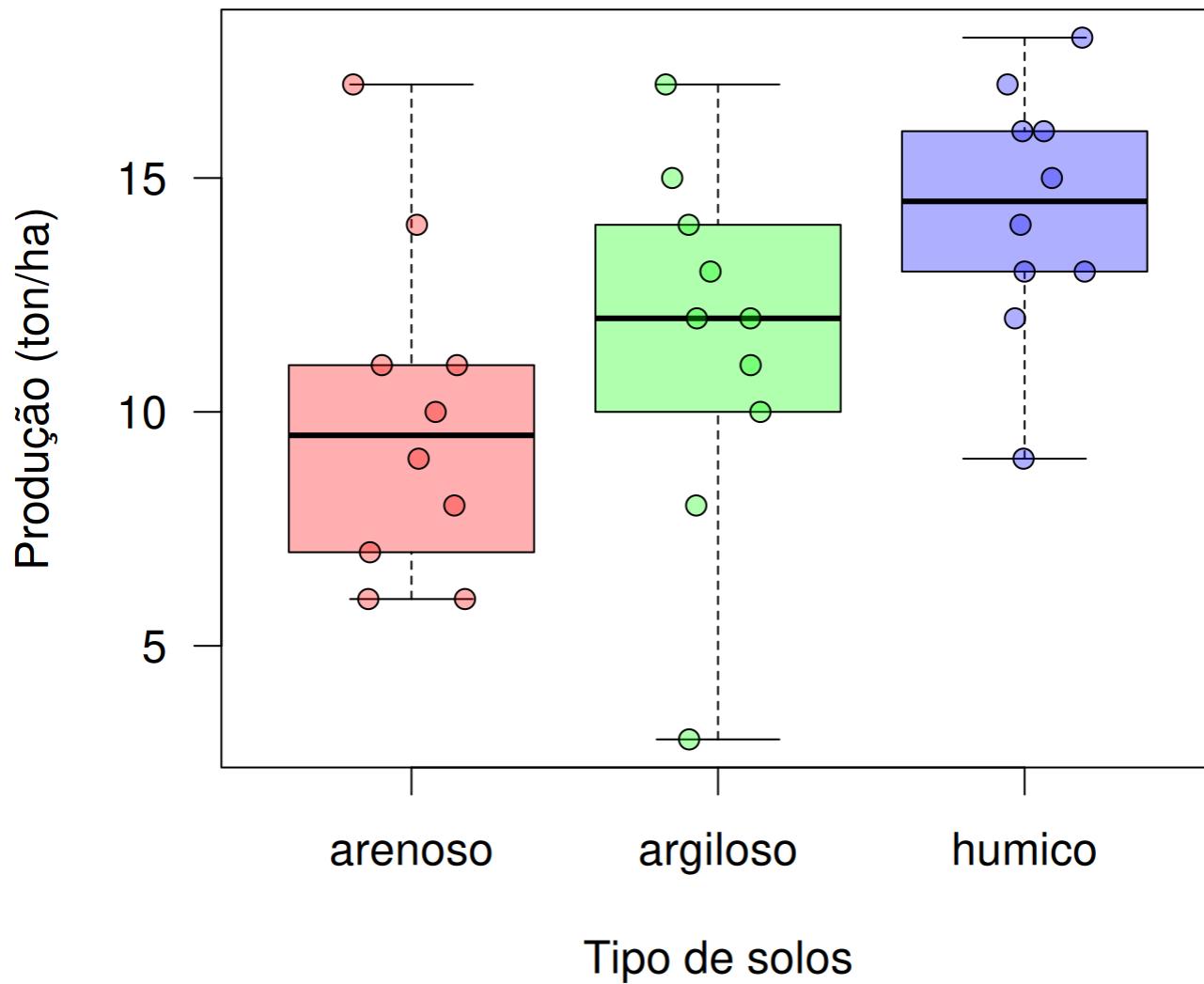
### ARTICLE HISTORY

Received June 2017  
Revised August 2017

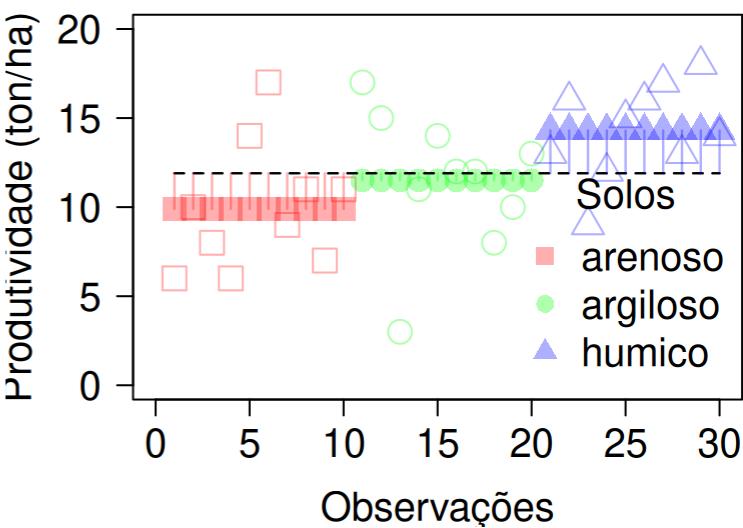
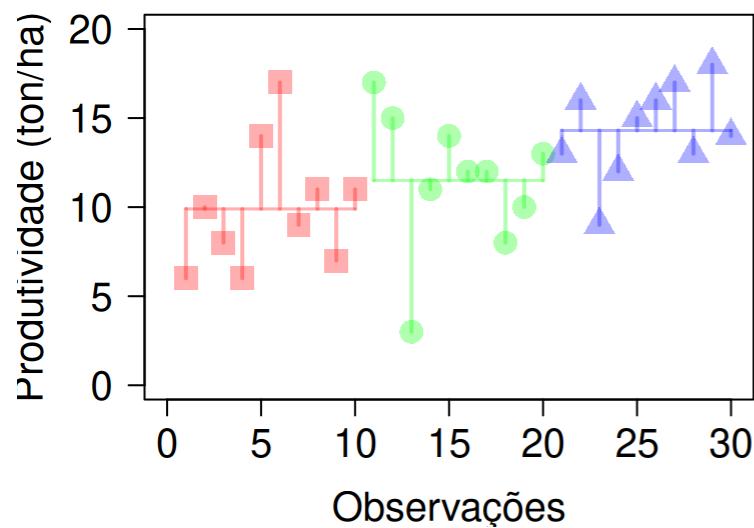
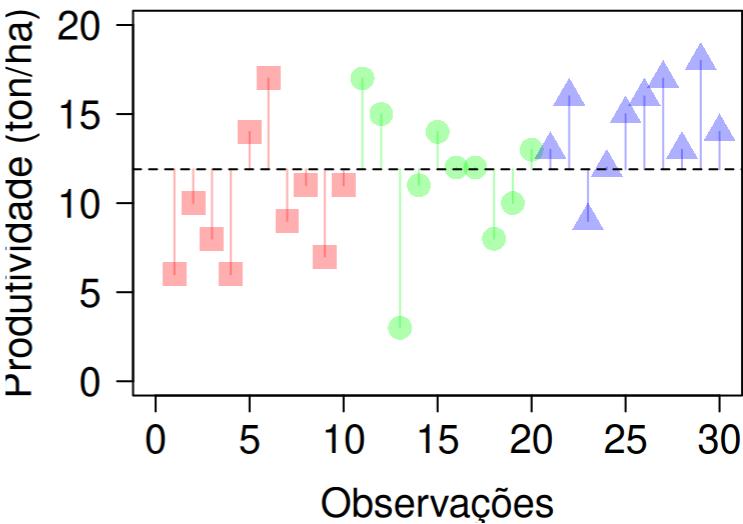
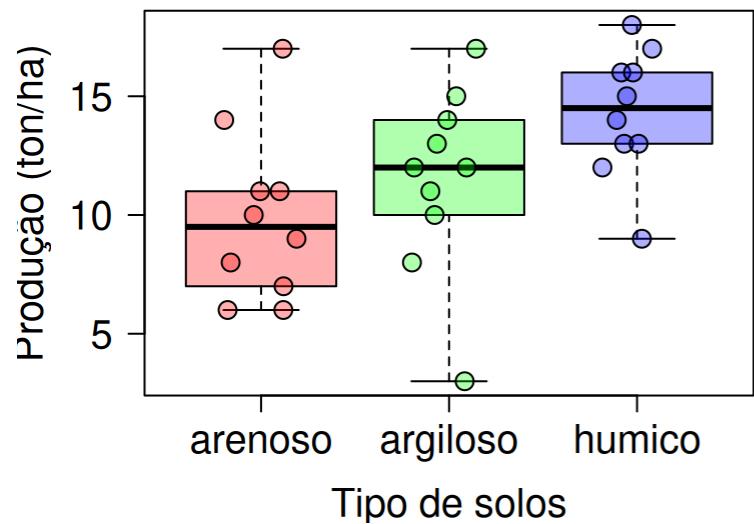
### KEYWORDS

Data management; Data organization; Microsoft Excel; Spreadsheets

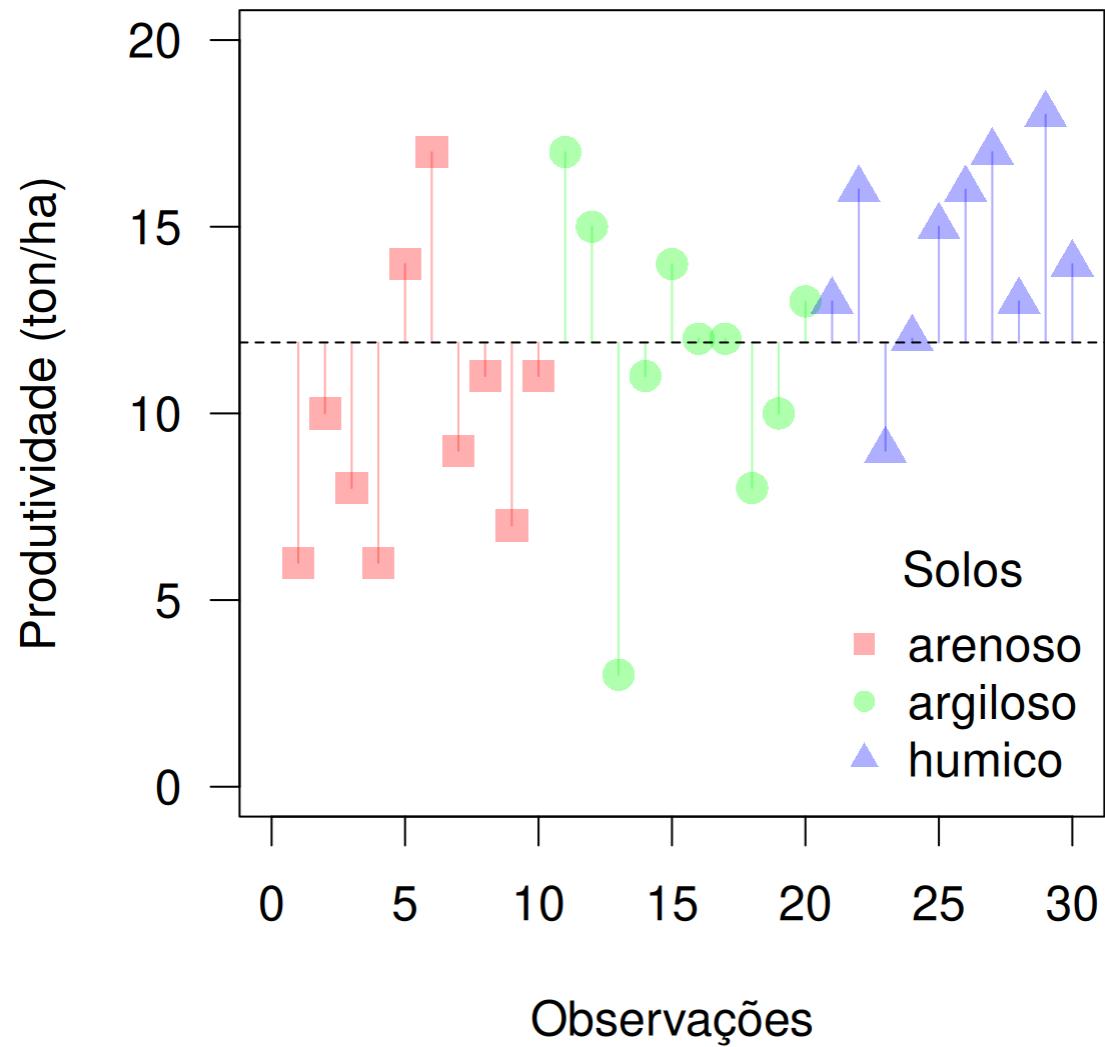
# Anova: um gráfico



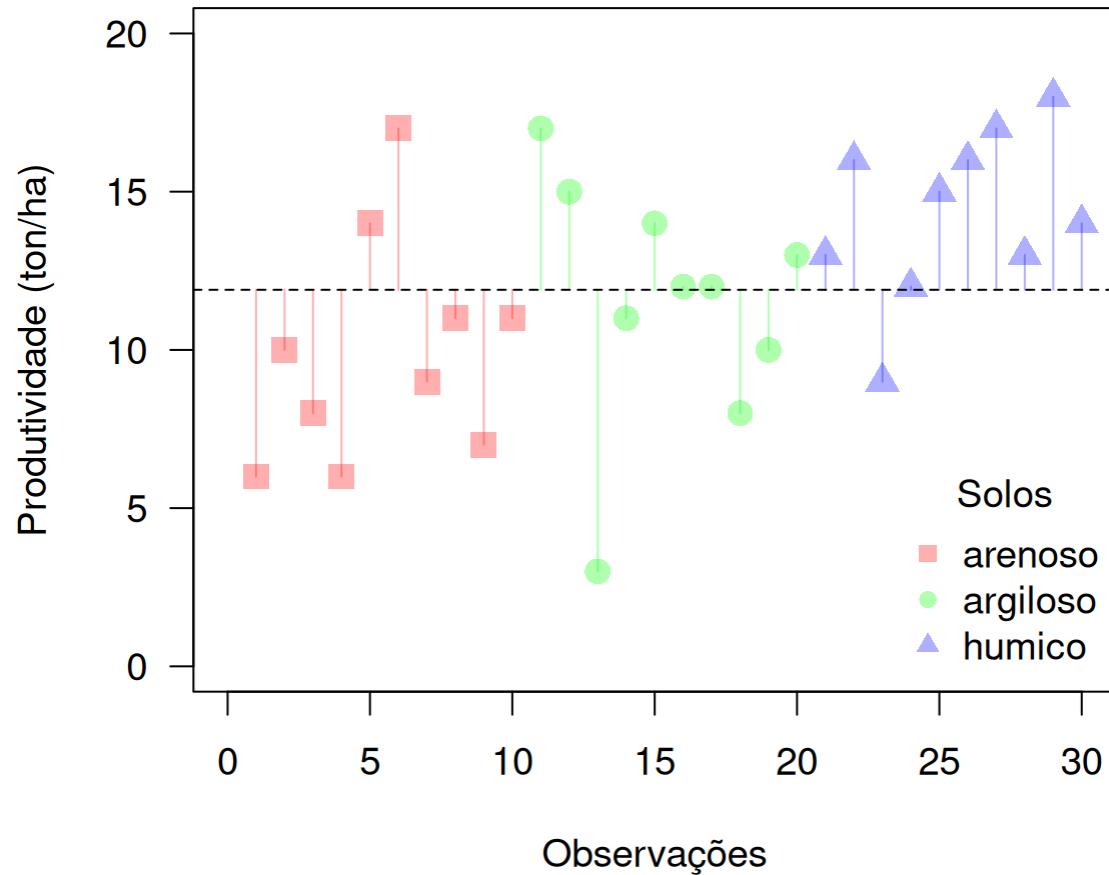
# Lógica da Anova



# Anova: variação total

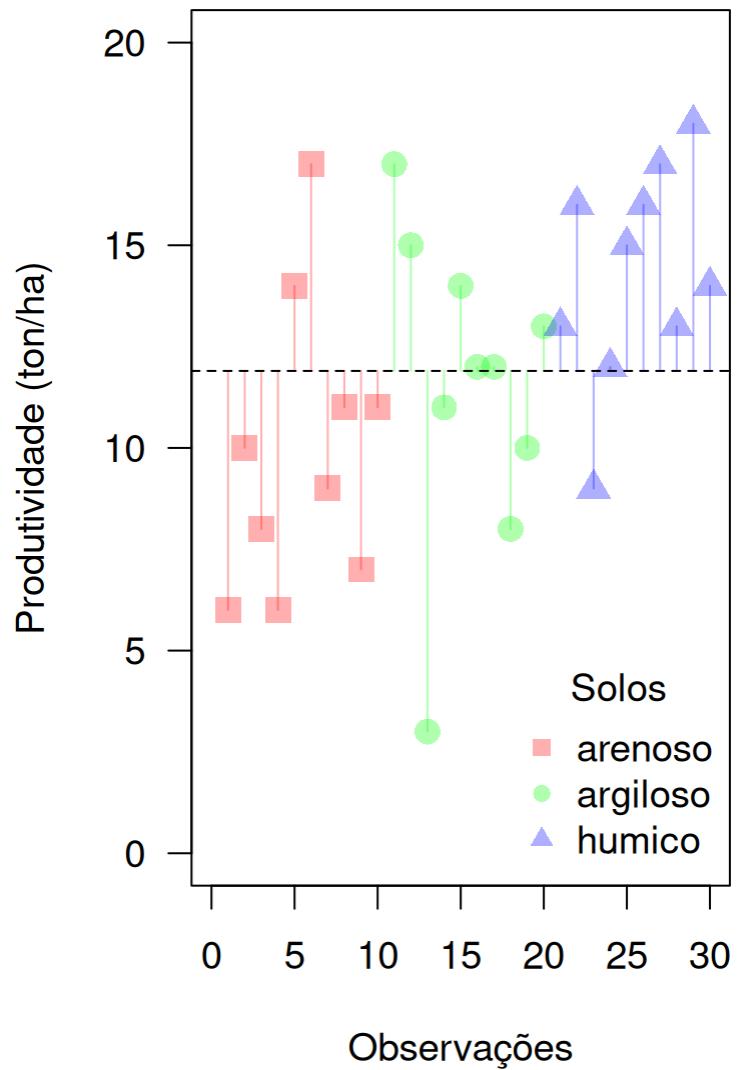


# Variação Total dos Dados



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

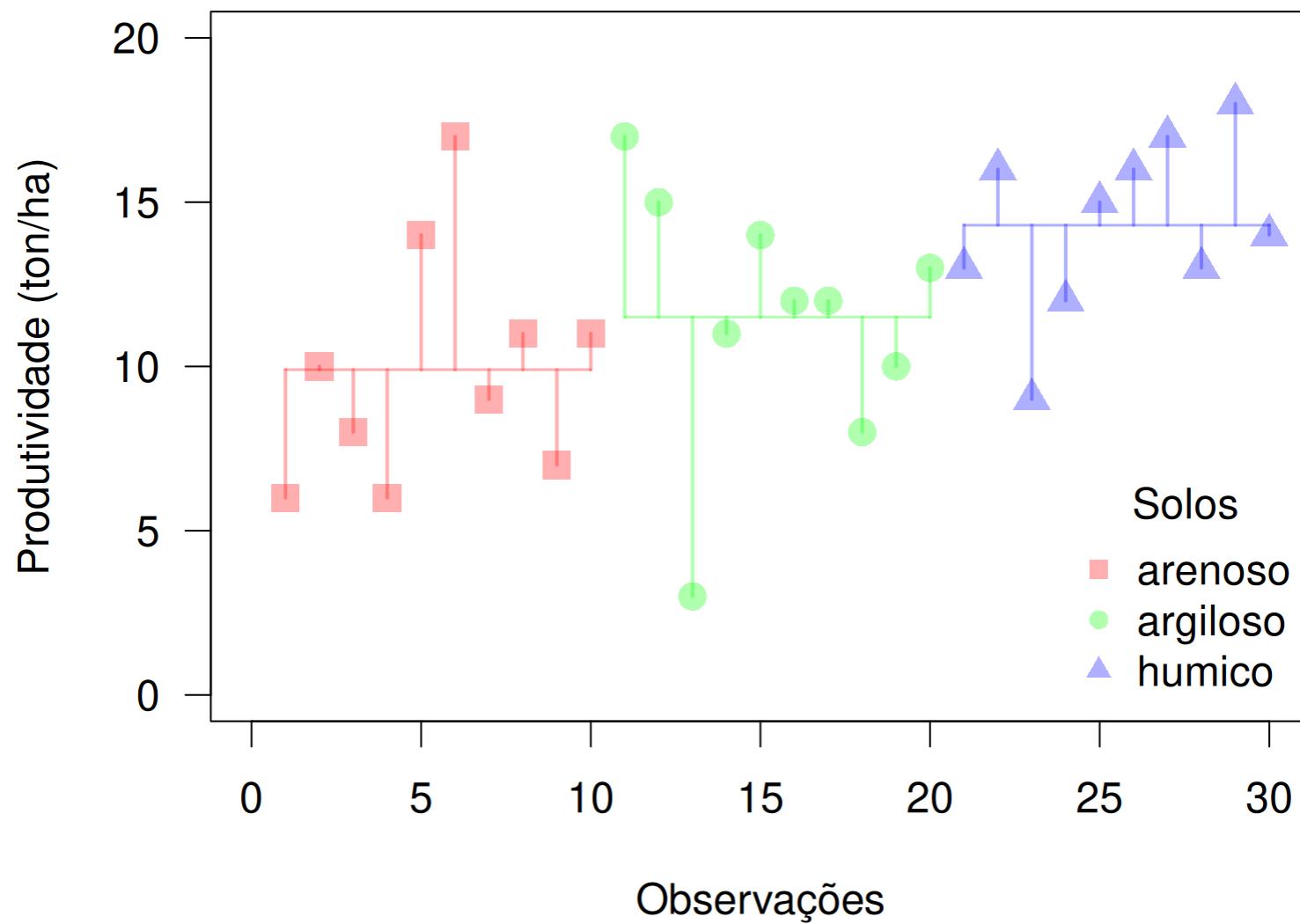
# Desvios Quadráticos: total



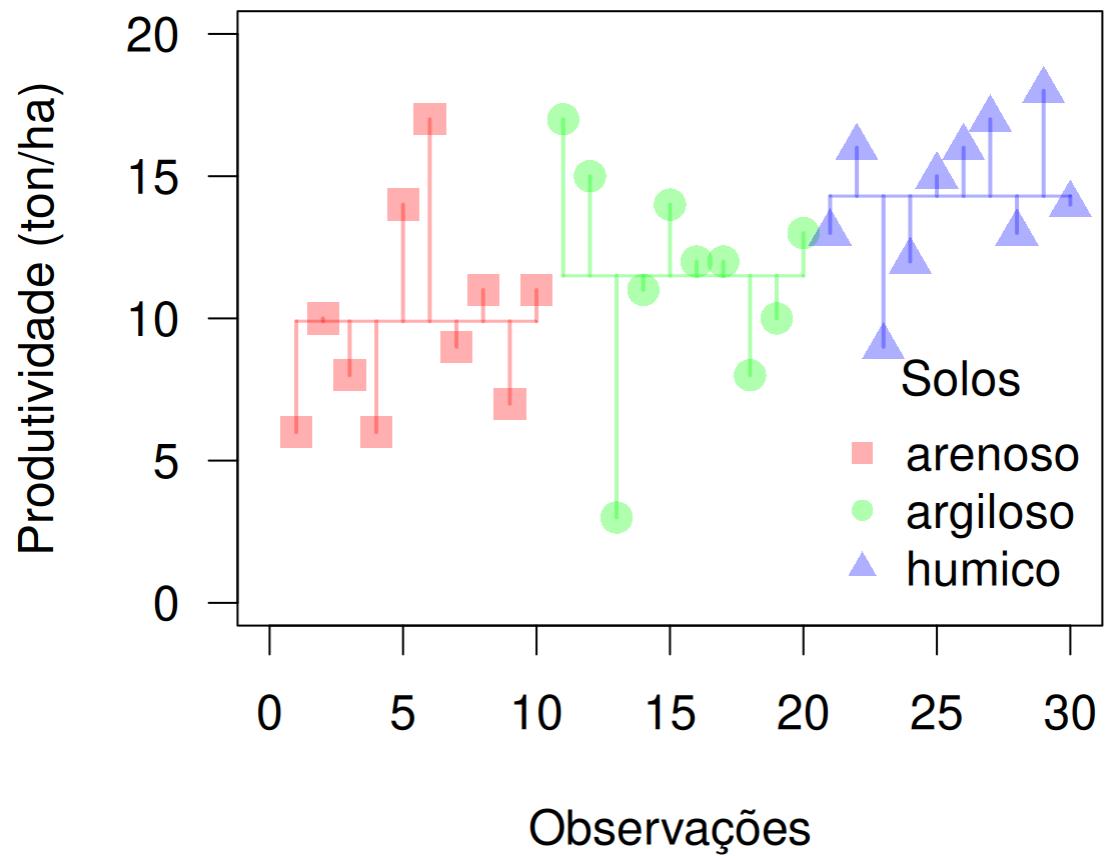
# Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos					
Intra Grupos	<b>X</b>				
<b>TOTAL</b>	<b>414.7</b>				

# Variação Não Explicada

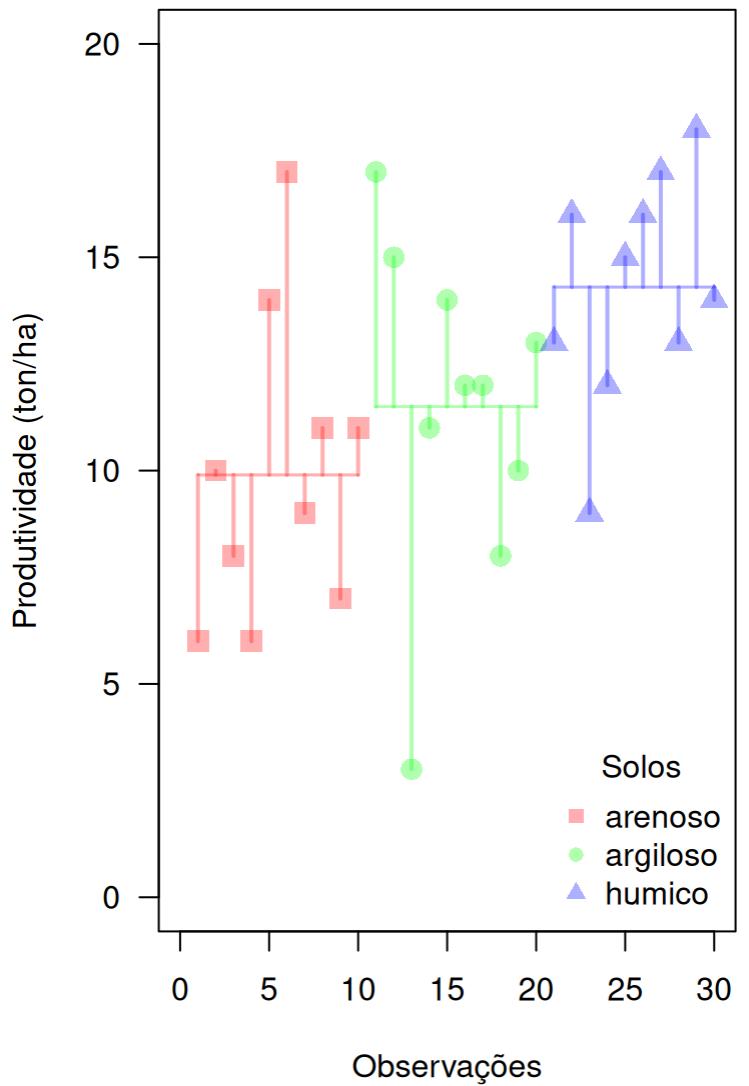


# Soma Quadratica: intra



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

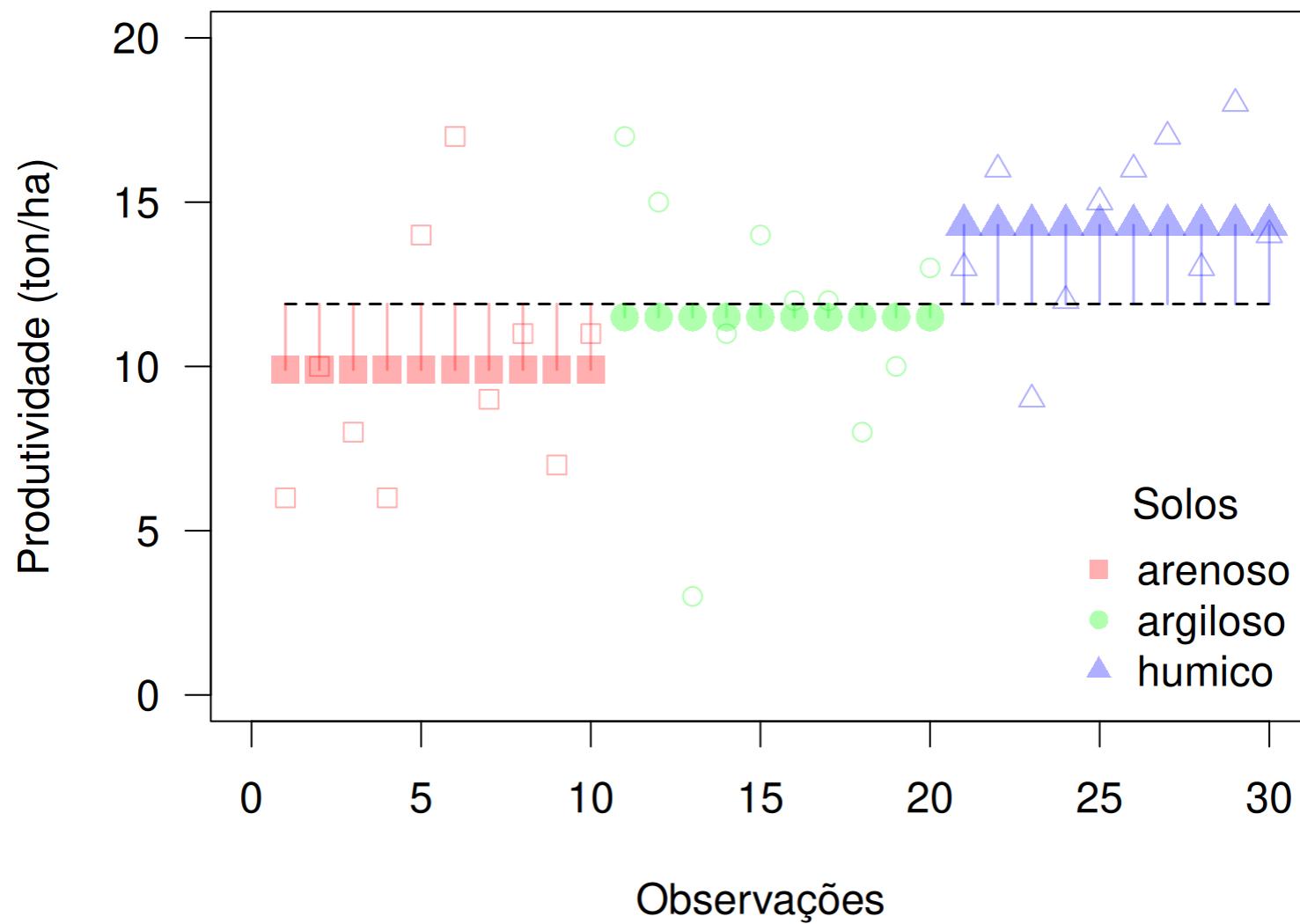
# Desvios Quadráticos: intra



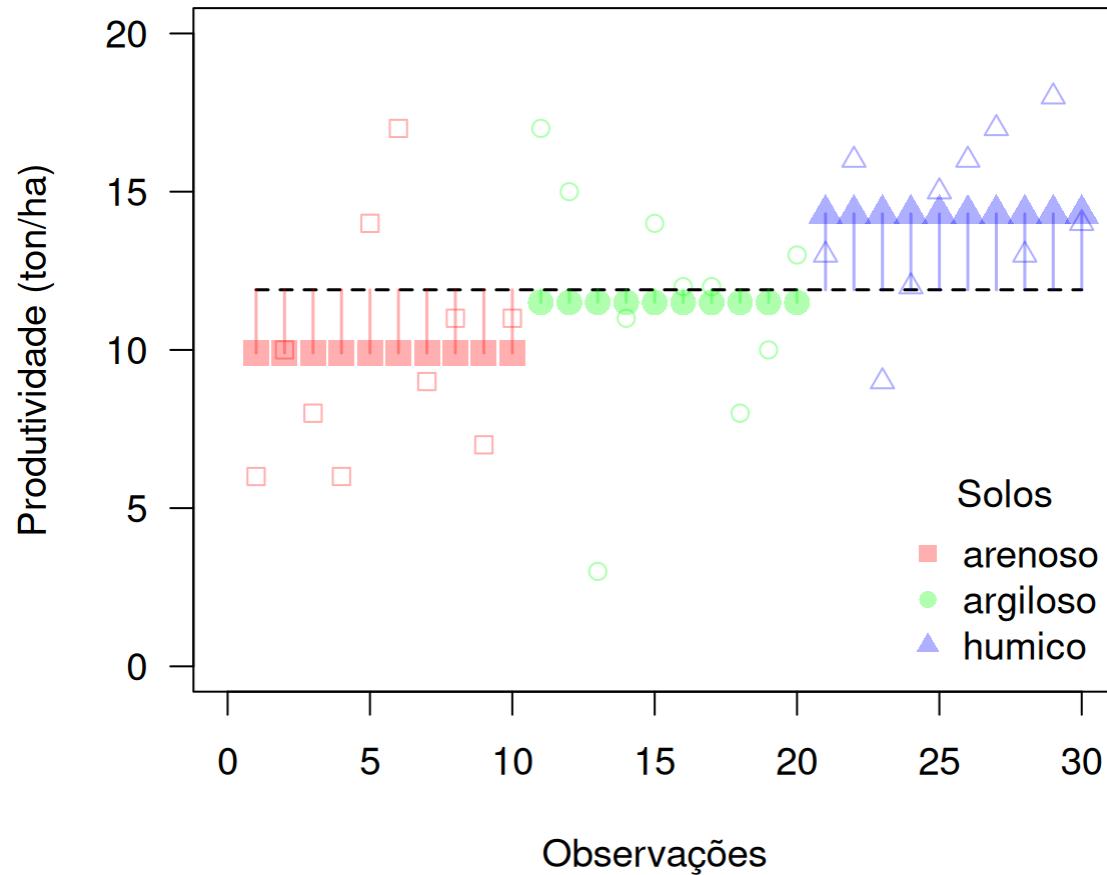
# Construindo a tabela de Anova

Fonte	Desvio Quadrático	Graus de Liberdade	Desvio Médio	Razão das Variâncias	P-valor
Entre Grupos	<b>X</b>				
Intra Grupos	<b>315.5</b>				
<b>TOTAL</b>	<b>414.7</b>				

# Variação Explicada

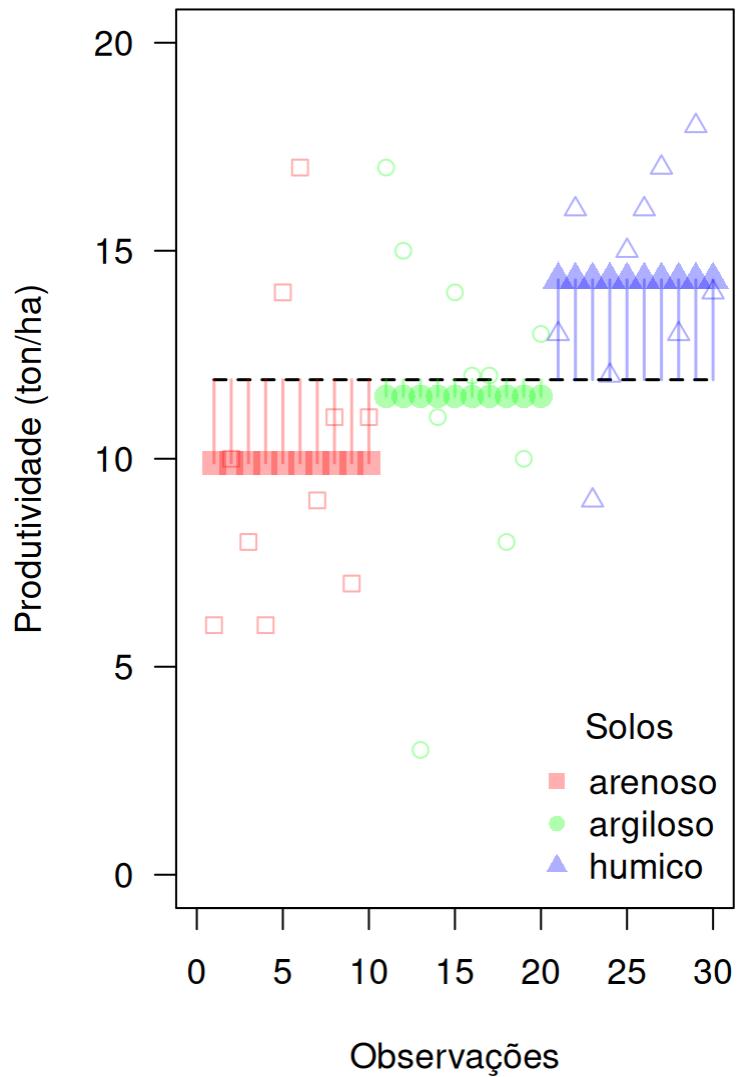


# Soma Quadráticos: entre



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

# Desvios Quadráticos: entre



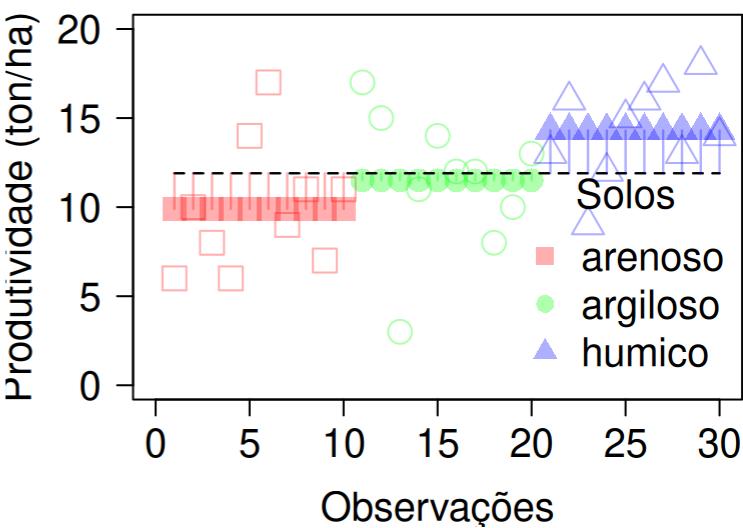
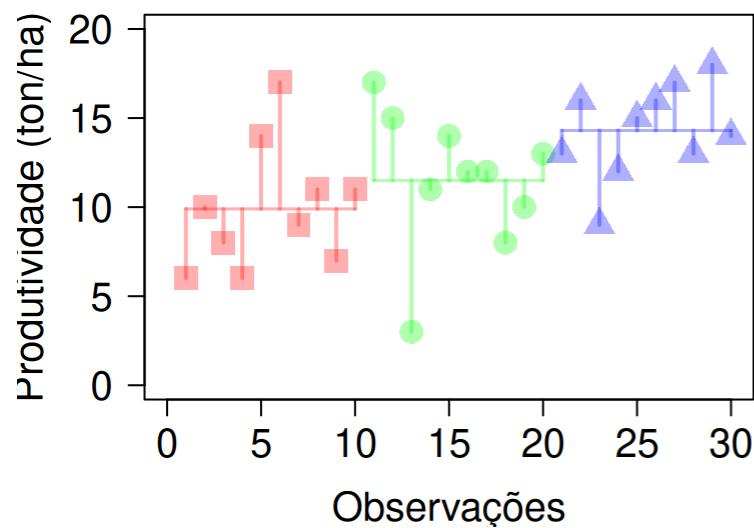
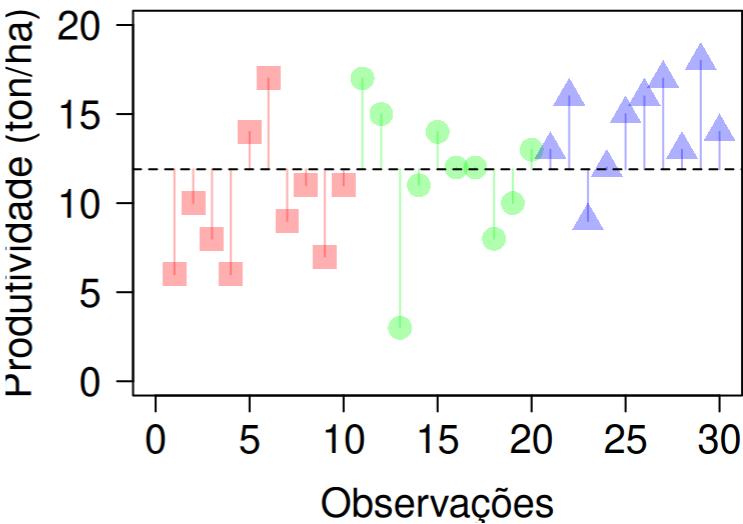
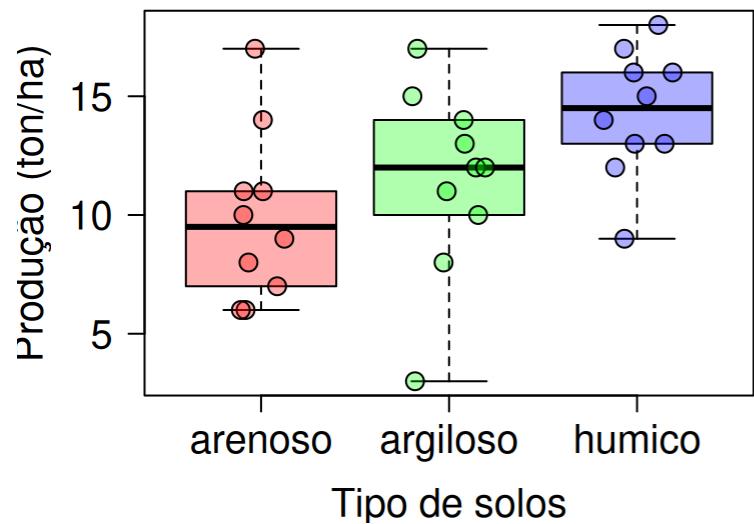
# Construindo a tabela de Anova

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

# Construindo a tabela de Anova

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

# Graus de Liberdade



# Construindo a tabela de Anova

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

# Construindo a tabela de Anova

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

# Estatística F

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

ou

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

(99.2/2) / (315.5/27)

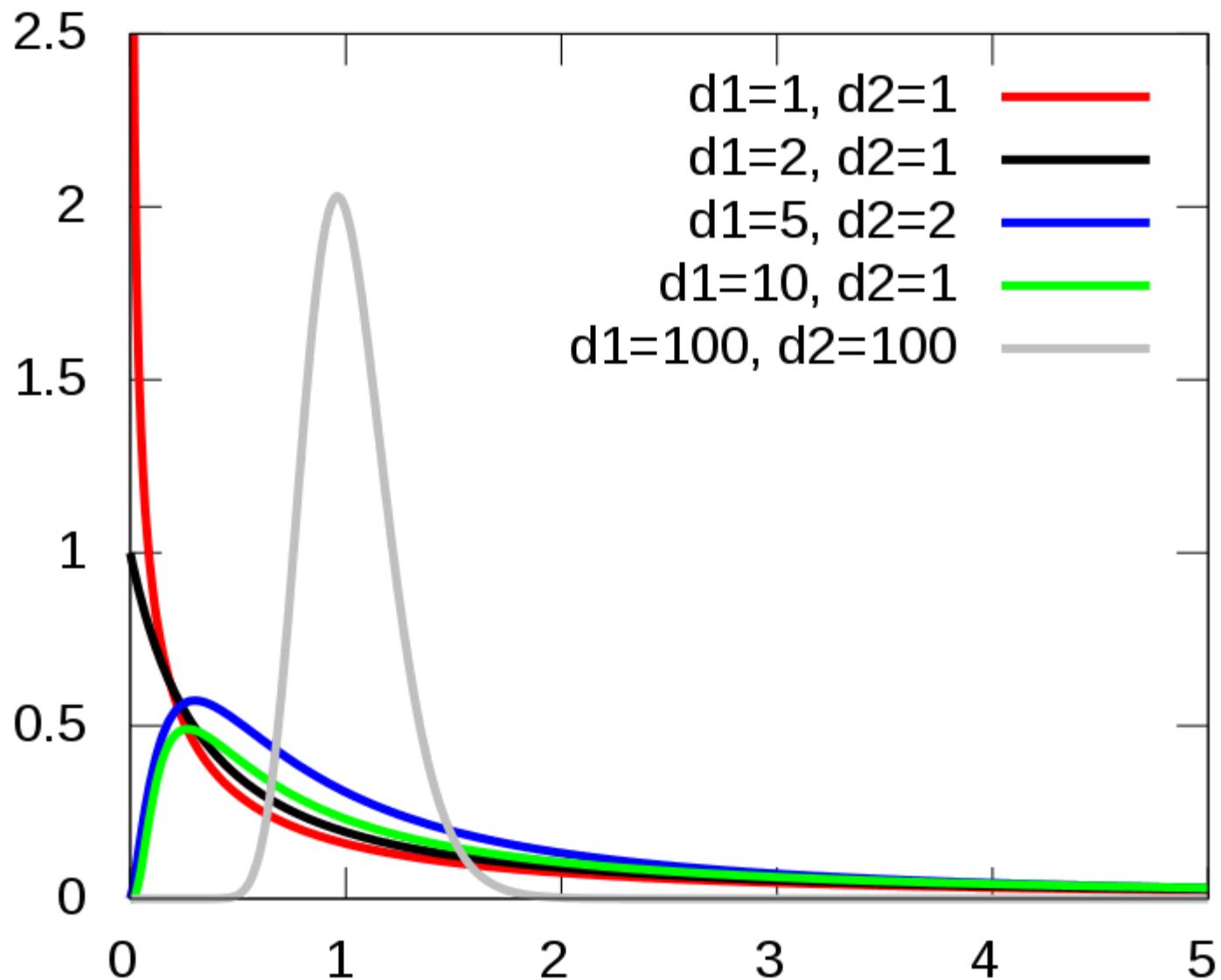
```
## [1] 4.244691
```

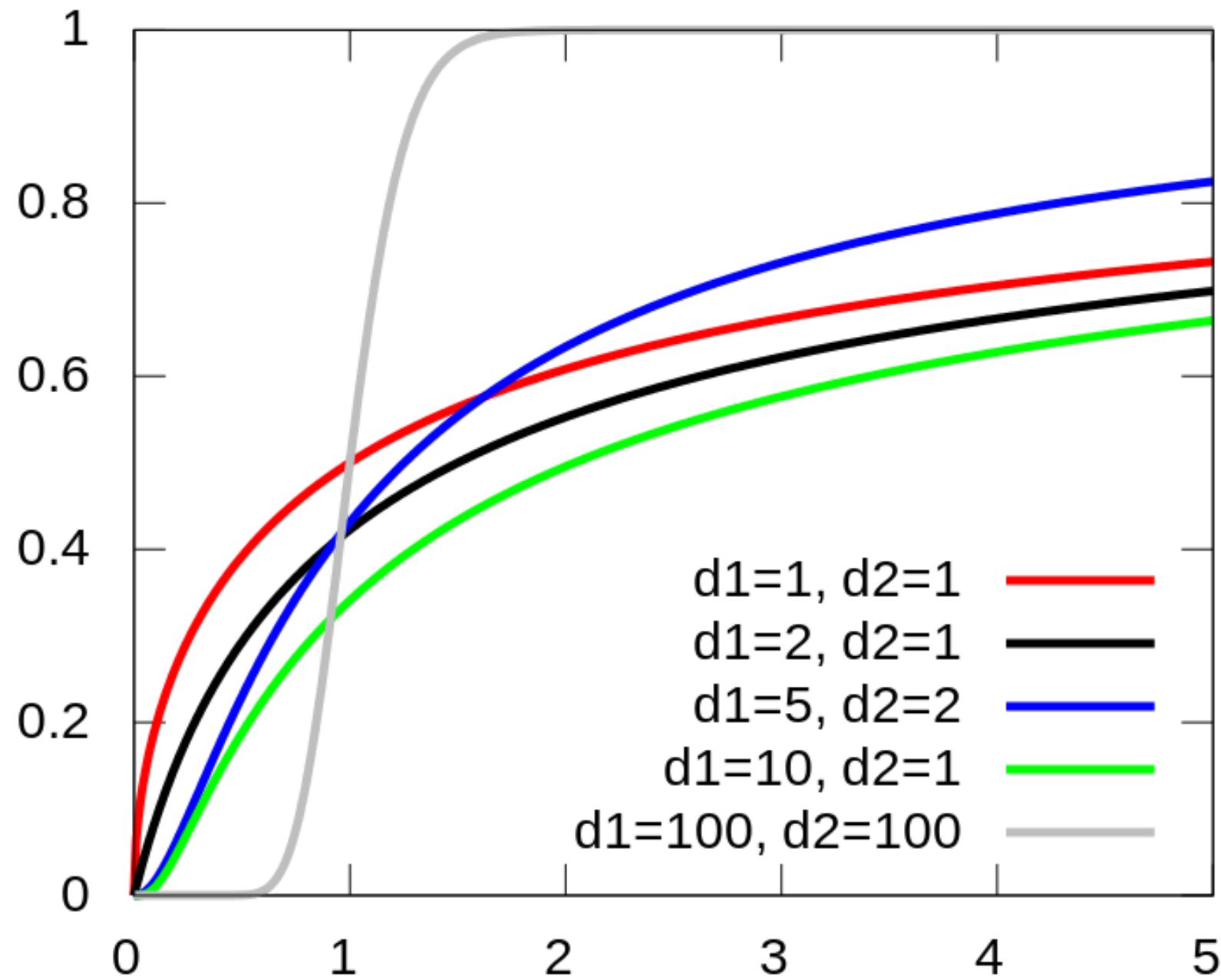
# Finalizando a tabela de Anova

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

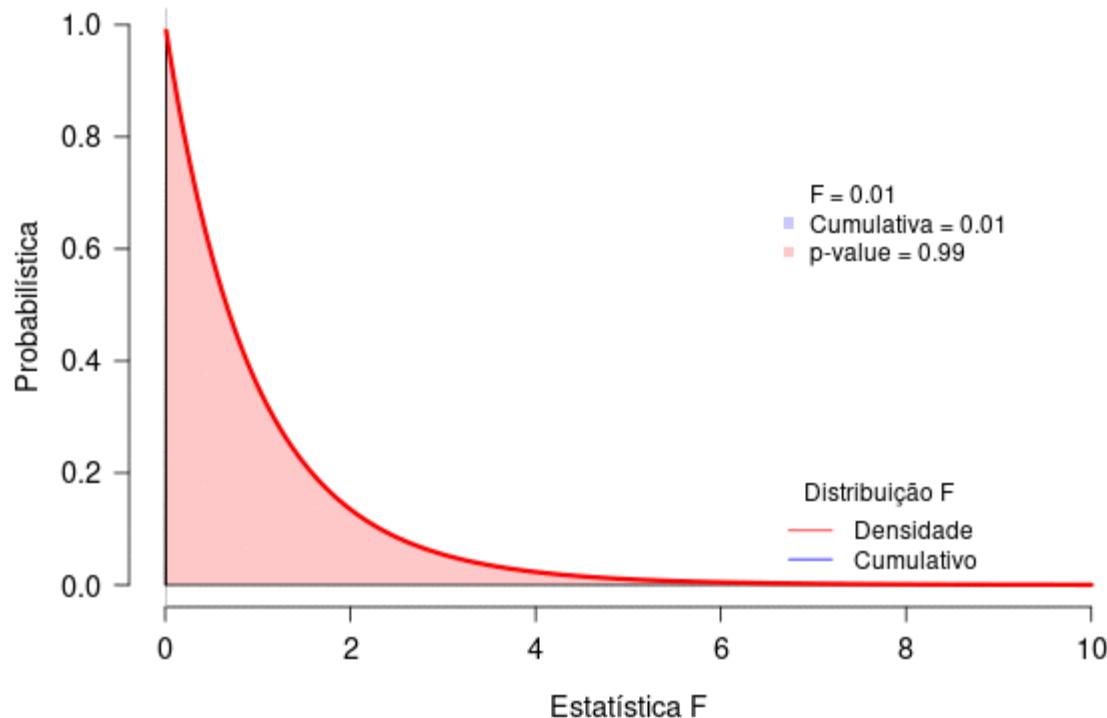
# Distribuição F

## Densidade Probabilística

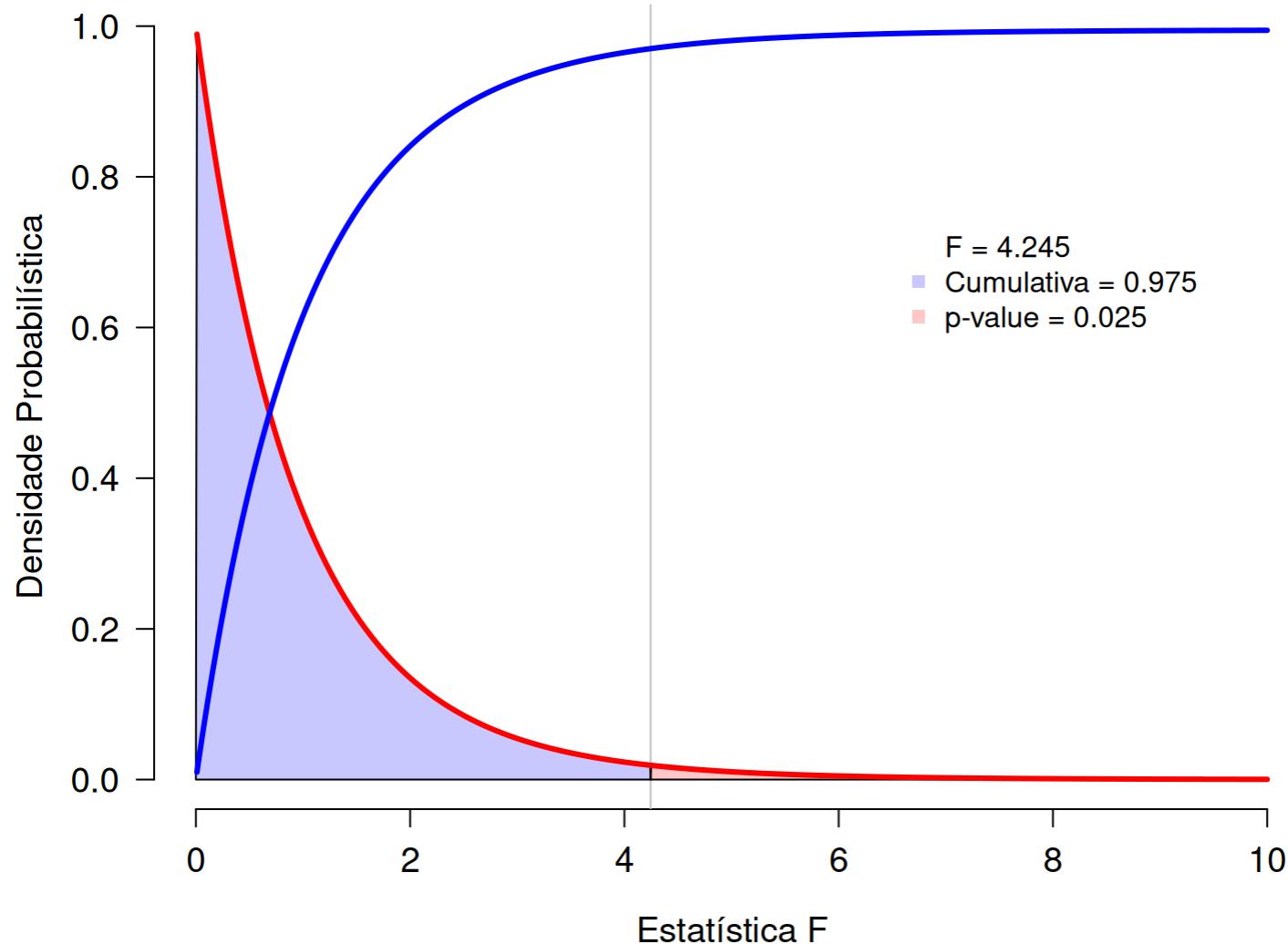




# Distribuição F(2,27)



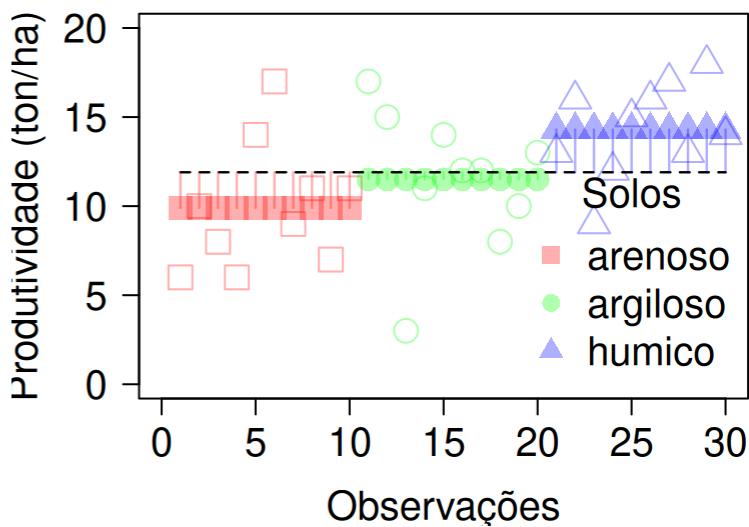
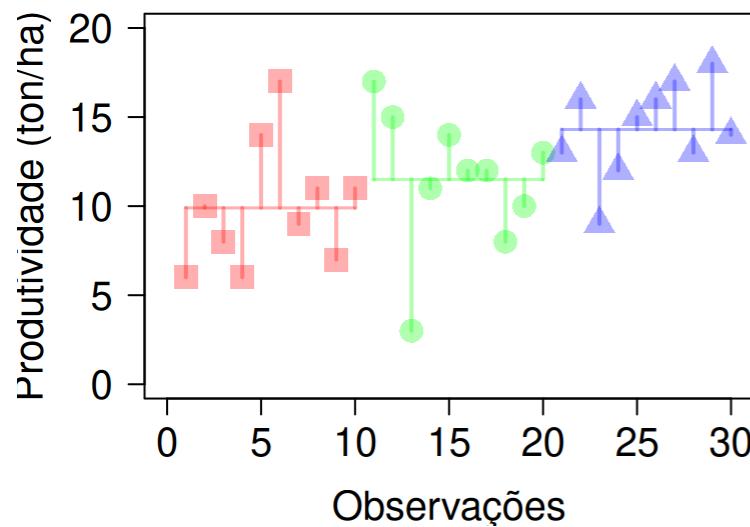
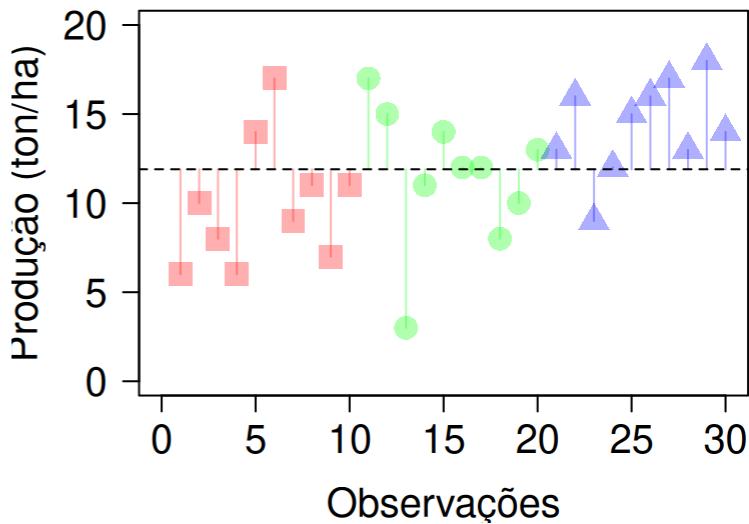
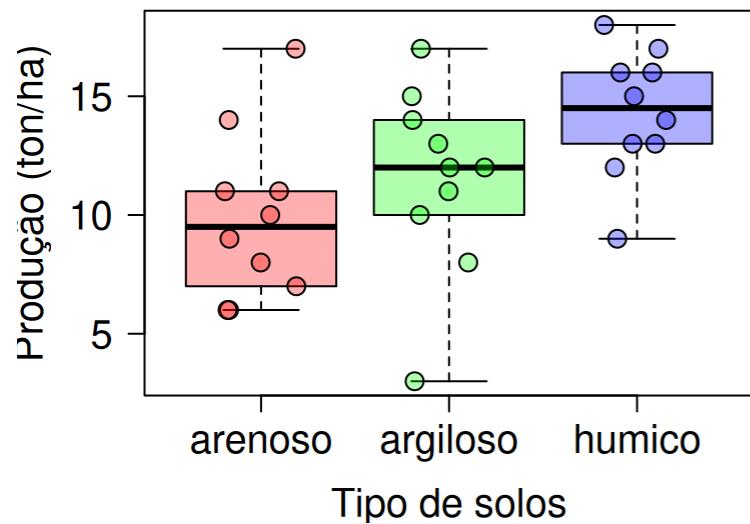
Distribuição F(2,27) = 4.24



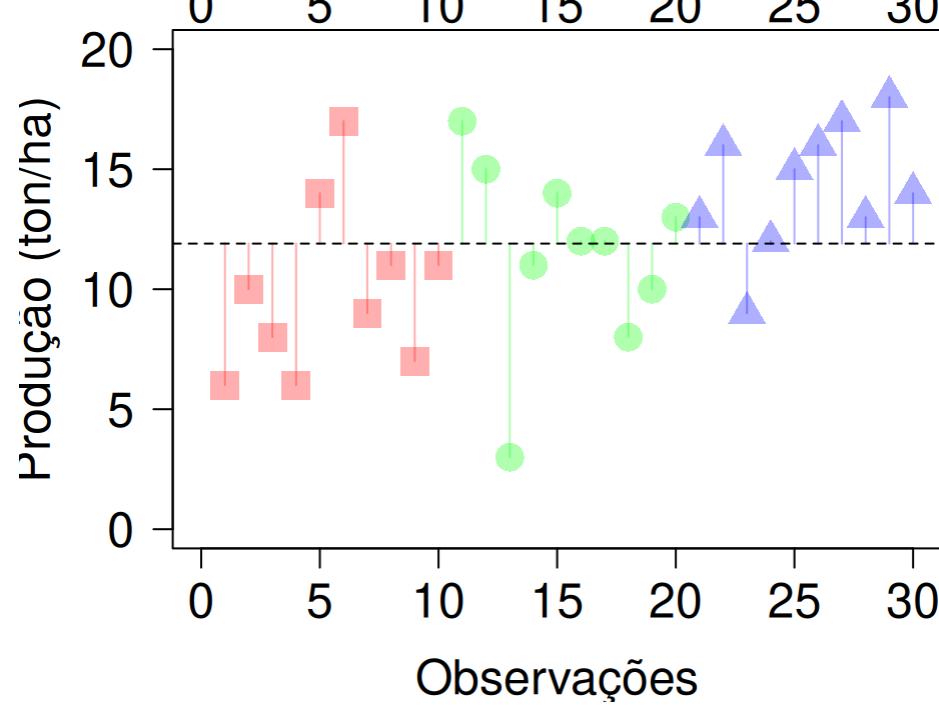
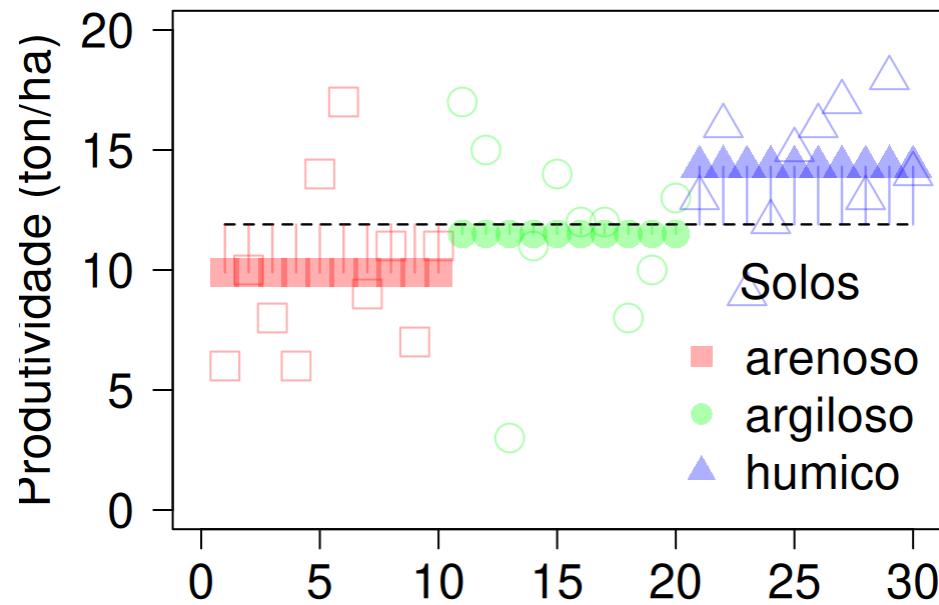
# Finalizando a tabela de Anova

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

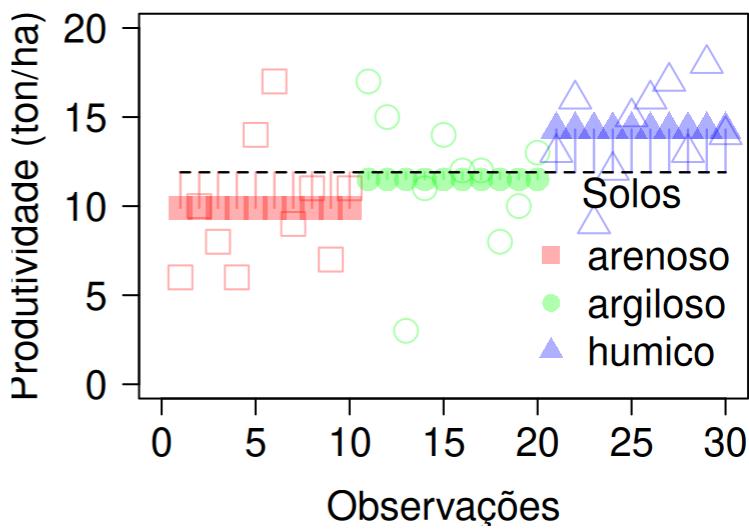
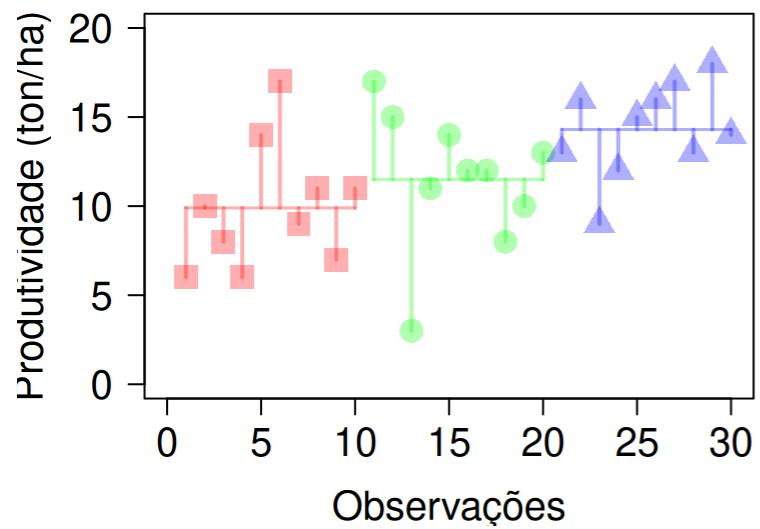
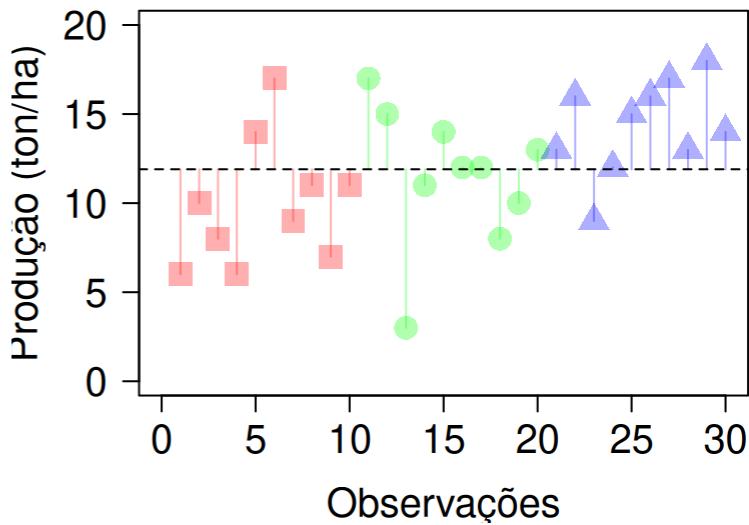
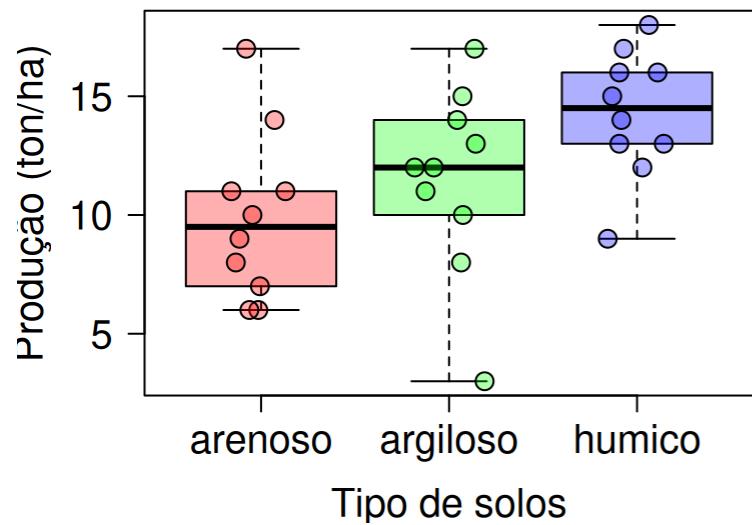
# Lógica da Anova



# Lógica da Anova



# Análise de Variância



# Atividade:

# ANOVA: Análise de Variância

Na aula sobre **teste de hipótese** utilizamos técnicas de Monte Carlo para testar a hipótese de que duas médias são distintas, ou que uma é maior/menor que outra, tanto no exemplo do tutorial **Tutorial Árvores do Mangue**, quanto no exercício **altura dos alunos**. Em ambos os casos estávamos comparando médias de dois grupos distintos, por exemplo dois tipos de solos no mangue ou gênero dos alunos. O nosso procedimento simulou o teste frequentista **t** de Student, que utiliza uma distribuição estatística **t**<sup>6)</sup> e dessa forma podemos comparar o valor observado em nossos dados com a distribuição estatística e testar a hipótese das médias serem diferentes, sem a necessidade de simular o cenário nulo, como apresentamos na aula **teste de hipótese**.



Caso não esteja confortável com o procedimento de simulação do cenário nulo e consequente obtenção do **p-valor**, refaça o tutorial **teste de hipótese**. No procedimento apresentado está a lógica básica por trás de todos os testes de hipótese clássicos.

A *Análise de Variância* ( **ANOVA** ) é uma generalização do **teste-t**, desenvolvida por  **Ronald Fisher** 100 anos atrás (1918). Apesar de idoso, é um teste muito popular, talvez o mais utilizado em ciências naturais. A hipótese subjacente da ANOVA é de diferença entre as média de 2 ou mais grupos. O procedimento para o cálculo da estatística da ANOVA, chamada de **F**, está associado à partição da variância dos dados, por isso o nome. Uma maneira clássica de apresentar o resultado do teste de **ANOVA** é a a chamada **tabela de ANOVA**. Essa tabela será utilizada para avaliarmos outros modelos também, por isso é importante entender o que ela nos diz.

## Tabela de ANOVA

