

# **R para Ciência de Dados**

## **Modelos de Análise de Variância (Anova)**

Profa Carolina Paraíba e Prof Gilberto Sassi

Departamento de Estatística  
Instituto de Matemática e Estatística  
Universidade Federal da Bahia

Julho de 2024

# Curso R para Ciência de Dados: Modelos de Análise de Variância (Anova)

## Aula 2: Delineamento Aleatorizado em Blocos.

- Introdução
- Descrição e modelo para a análise de variância
- Análise de variância e teste-F para médias de tratamentos
- Estimação dos parâmetros do modelo
- Teste de Aditividade de Tukey
- Parcelas perdidas
- Exemplos
- Usando o R

## Introdução

Um fator de controle local (fator de bloqueamento) é um fator perturbador. Isto é, é um fator que afeta a variável resposta de interesse em um experimento, mas em cujo efeito não estamos interessados.

Se um fator perturbador é conhecido e controlável, podemos usar controle local (blocos) para eliminar seu efeito na resposta de interesse.

O uso de blocos (controle local) nos permite controlar alguma fonte de variação conhecida que afeta os resultados, diminuindo o erro experimental (e aumentando a sensibilidade do experimento).

## Introdução

Para eliminar ao máximo a variação natural e aumentar a sensibilidade de um experimento, é aconselhável que as unidades experimentais sejam tão homogêneas quanto possível. Desta forma, o termo de variância do erro diminui e o poder de detecção do efeito dos tratamentos aumenta. Por outro lado, queremos que o resultado do experimento tenha extensa aplicação.

**Exemplo 1.** Considere que uma pesquisadora deseja estudar o efeito de métodos de atividade física aeróbica em níveis de ansiedade e estresse de pessoas. ■

## Introdução

Sabe-se que existe uma grande variabilidade de níveis de ansiedade e estresse na população (conforme medições de teste padrões). Então, seria difícil observar a diferença entre os métodos de atividade a menos que os indivíduos em estudo tivessem uma classificação de nível de ansiedade e estresse similar (fossem homogêneos).

Não obstante, a pesquisadora deseja obter conclusões gerais a partir desse estudo para pessoas em todos os níveis de ansiedade e estresse.

Usando blocos, é possível alcançar os dois objetivos: obter grupos (blocos) de indivíduos homogêneos segundo nível de ansiedade e estresse e observar o efeito de diferentes métodos de atividade física aeróbica.

## Introdução

No delineamento em blocos ao acaso, utilizamos unidades experimentais heterogêneas para que as conclusões sejam mais gerais. Porém, essas unidades experimentais heterogêneas são agrupadas em subgrupos homogêneos antes de serem atribuídas aleatoriamente aos níveis de fator de tratamento de interesse.

O agrupamento das unidades experimentais em grupos homogêneos é chamado de controle local (bloqueamento).

Atribuir aleatoriamente níveis de fator de tratamento à unidades experimentais dentro de subgrupos homogêneos menores de unidades experimentais, ou blocos, tem o mesmo efeito que usar apenas unidades homogêneas, mas permite que as conclusões sejam generalizadas para toda a classe de unidades experimentais heterogêneas usadas no estudo.

## Introdução

**Exemplo 2.** Considere um experimento agrícola com três campos nos quais serão avaliados a qualidade de três variedades diferentes de cevada. Devido à forma como é feita a colheita da cevada, só é possível criar um máximo de três parcelas em cada campo. Neste exemplo, o fator de bloqueamento é o campo, pois pode haver diferenças no tipo de solo, drenagem, etc. de campo para campo. Em cada campo, são plantadas todas as três variedades para que seja possível distinguir entre as variedades sem que o efeito de bloco do campo confunda os resultados. Neste exemplo, as variedades estão aninhadas dentro dos campos. ■

# Introdução

**Exemplo 3.** Estamos interessados em saber como um ratos responde a cinco materiais diferentes inseridos no tecido subcutâneo para avaliar o uso dos materiais na medicina. Cada rato pode ter no máximo 3 inserções. Aqui vamos bloquear os ratos individuais porque até os ratos de laboratório têm variação individual. Observe que não estamos interessados em estimar o efeito dos ratos, mas o efeito de bloco dos ratos deve ser levado em consideração antes de fazermos qualquer inferência sobre os materiais. Se tivermos apenas uma inserção por rato, o efeito do rato será confundido com os materiais. ■



# Introdução

## Vantagens de usar blocos:

- Útil para comparar tratamentos na presença de uma fonte de variação conhecida e controlável que não é de interesse.
- A análise estatística é simples.
- O delineamento é fácil de ser contruído.
- Se o agrupamento é efetivo, os resultados obtidos são mais precisos (comparando com delineamento completamente aleatorizado).
- Não precisamos usar tamanhos amostrais iguais. Por exemplo, se estamos comparando um controle com três tratamentos e o controle deve ter um tamanho amostral duas vezes maior que os tratamentos, então podemos usar blocos com cinco unidades: três unidades são alocadas para os tratamentos e duas para o controle.

# Introdução

## **Desvantagens de usar blocos:**

- Quando observações são perdidas dentro de um bloco, é necessária uma análise mais complexa.
- Os graus de liberdade do erro experimental não são grandes o suficiente (comparado com delineamento completamente aleatorizado). Um grau de liberdade é perdido para cada bloco após o primeiro.
- Mais suposições são necessárias para o modelo (ausência de interação entre os blocos e tratamentos, variância constante entre os blocos).

## Descrição e modelo para a análise de variância

Em um delineamento (completamente) aleatorizado em blocos, com um fator, quando o fator tem  $a$  níveis, o fator de bloqueio tem  $b$  subgrupos e cada um deles contém exatamente  $a$  unidades experimentais, formando um total de  $a \times b$  unidades experimentais.

As  $a$  unidades experimentais dentro de cada bloco são tão similares quanto possível, e os grupos de unidades experimentais variam suficientemente de bloco em bloco permitindo a generalização das conclusões.

## Descrição e modelo para a análise de variância

A aleatorização das unidades experimentais aos tratamentos é realizada dentro de cada bloco.

Esse processo resulta em maior homogeneidade entre as unidades experimentais, reduz o erro experimental e produz estimativas mais precisas dos efeitos dos tratamentos.

Em um delineamento completamente aleatorizado em blocos, cada bloco constitui uma replicação do experimento.

## Descrição e modelo para a análise de variância

O modelo para um delineamento completamente aleatorizado em bloco, com a suposição de ausência de efeito de interação, com  $b$  bloco e  $a$  tratamentos de efeitos fixos é definido por

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad (1)$$

onde  $\mu$  é uma média geral,  $\tau_i$  é o efeito do  $i$ -ésimo tratamento,  $\beta_j$  é o efeito do  $j$ -ésimo bloco e  $\epsilon_{ij} \sim N(0, \sigma^2)$ , para  $i = 1, \dots, a$  e  $j = 1, \dots, b$ .

O modelo (1) é sobre especificado. Considerando os  $\tau_i$ 's e  $\beta_j$ 's como desvios da média geral, impomos as restrições

$$\sum_{i=1}^a \tau_i = 0 \text{ e } \sum_{j=1}^b \beta_j = 0, \quad (2)$$

## Descrição e modelo para a análise de variância

As respostas  $Y_{ij}$  são independentes e normalmente distribuídas com média

$$E(Y_{ij}) = \mu + \tau_i + \beta_j, \quad (3)$$

e variância constante

$$V(Y_{ij}) = \sigma^2. \quad (4)$$

# Descrição e modelo para a análise de variância

## Notações:

Vamos definir:

- $y_{i.} = \sum_{j=1}^b y_{ij}; i = 1, \dots, a.$
- $y_{.j} = \sum_{i=1}^a y_{ij}; j = 1, \dots, b.$
- $y_{..} = \sum_{i=1}^a \sum_{j=1}^b y_{ij}; i = 1, \dots, a, j = 1, \dots, b.$

Assim, a média do  $i$ -ésimo tratamento, a média do  $j$ -ésimo bloco e a média geral são

$$\bar{y}_{i.} = \frac{y_{i.}}{b}, \bar{y}_{.j} = \frac{y_{.j}}{a} \text{ e } \bar{y}_{..} = \frac{y_{..}}{n}, \quad (5)$$

onde  $n = ab$  é o número total de observações.

# Análise de variância e teste-F para médias de tratamentos

## Hipóteses:

Nosso interesse é testar se as médias dos  $a$  tratamentos são iguais:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_a \\ H_1 : \mu_i \neq \mu_j \text{ para pelo menos um par } (i, j). \end{cases} \quad (6)$$

Note que, sob a restrição  $\sum_{j=1}^b \beta_j = 0$ , temos que

$$\mu_i = \frac{1}{b} \sum_{j=1}^b (\mu + \tau_i + \beta_j) = \mu + \tau_i, \quad i = 1, \dots, a, \quad (7)$$

Portanto, podemos escrever as hipóteses

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \tau_3 = \dots = \tau_a = 0 \\ H_1 : \tau_i \neq 0 \text{ para pelo menos um } i. \end{cases} \quad (8)$$



## Análise de variância e teste-F para médias de tratamentos

### Partição da variabilidade total:

- $SQT = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2$ : é a soma de quadrados total; com  $n - 1$  graus de liberdade.
- $SQTR = b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2$ : é a soma de quadrados dos tratamentos; com  $a - 1$  graus de liberdade.
- $SQB = a \sum_{j=1}^b (y_{.j} - \bar{y}_{..})^2$ : é a soma de quadrados de blocos; com  $b - 1$  graus de liberdade.
- $SQR = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ : é a soma de quadrados de resíduos; com  $(a - 1)(b - 1)$  graus de liberdade.

## Análise de variância e teste-F para médias de tratamentos

### Estatística teste:

A estatística teste para o teste de igualdade de médias (ou de efeito de tratamentos) é definida por

$$F_0 = \frac{QMTR}{QMR}. \quad (9)$$

Sob  $H_0$ , segue que  $F_0 \sim F(a-1, (a-1)(b-1))$ .

A um nível de significância  $\alpha \in (0, 1)$ , rejeitamos  $H_0$  se

$$F_0 > F_{\alpha; a-1; (a-1)(b-1)}. \quad (10)$$

## Análise de variância e teste-F para médias de tratamentos

### Tabela Anova:

**Tabela 1:** Tabela Anova: modelo com um fator (efeitos fixos).

Fonte de Variação	Soma de Quadrados	Graus de Liberdade	Quadrado Médio	$F_0$
Tratamento	$SQTR$	$a - 1$	$QMTR$	$\frac{QMTR}{QMR}$
Bloco	$SQB$	$b - a$	$QMR$	
Resíduo	$SQR$	$(a - 1)(b - 1)$	$QMR$	
Total	$SQT$	$n - 1$		

## Análise de variância e teste-F para médias de tratamentos

O modelo (1) é um modelo aditivo que não contém um termo de interação entre bloco e tratamento.

Se o termo de interação fosse incluído no modelo, como há apenas  $a \times b$  observações, o grau de liberdade para o termo de erro do modelo seria zero.

Porém, a interação entre bloco e tratamento será, de fato, o termo de erro correto para testar o efeito de tratamento.

Queremos generalizar as conclusões sobre os efeitos do tratamento sobre todas as unidades experimentais. Então os efeitos médios dos tratamentos devem ser maiores do que quaisquer diferenças nos efeitos do tratamento entre blocos de unidades experimentais.

A diferença nos efeitos dos tratamentos entre os blocos é exatamente o que a interação mede e, portanto, é o termo de erro correto. Ao deixar a interação fora do modelo, a soma de quadrado dos resíduos torna-se idêntica às somas dos quadrados da interação.

# Estimação dos parâmetros do modelo

## Estimação:

As estimativas de mínimos quadrados dos parâmetros do modelo são:

$$\hat{\mu} = \bar{y}_{..} \quad (11)$$

$$\hat{\mu}_i = \bar{y}_{i.} - \bar{y}_{..} \quad (12)$$

$$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \quad (13)$$

## Estimação dos parâmetros do modelo

**Valores ajustados:**

$$\begin{aligned}\hat{y}_{ij} &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}; \quad i = 1, \dots, a, j = 1, \dots, b\end{aligned}\quad (14)$$

**Resíduos:**

$$\begin{aligned}e_{ij} &= y_{ij} - \hat{y}_{ij} \\ &= y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}; \quad i = 1, \dots, a, j = 1, \dots, b.\end{aligned}\quad (15)$$

## Teste de Aditividade de Tukey

O modelo (1) de delineamento completamente aleatorizado em bloco faz a suposição de aditividade entre bloco e tratamento. Isto é, considera-se que não há interação entre os blocos e os tratamentos.

Para verificar a suposição de aditividade, considere o modelo

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{ij}; i = 1, \dots, a, j = 1, \dots, b. \quad (16)$$

A soma de quadrados da interação  $\sum_i \sum_j \gamma^2 \tau_i^2 \beta_j^2$  precisa ser estimada.

## Teste de Aditividade de Tukey

Assumindo que os demais parâmetros são conhecidos, o estimador de mínimos quadrados de  $\gamma$  é

$$\hat{\gamma} = \frac{\sum_i \sum_j \tau_i \beta_j y_{ij}}{\sum_i \tau_i^2 \sum_j \beta_j^2}. \quad (17)$$

O estimador de  $\tau_i$  é  $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$  e o estimador de  $\beta_j$  é  $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$ .

Portanto,

$$\hat{\gamma} = \frac{\sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..}) (\bar{y}_{.j} - \bar{y}_{..}) y_{ij}}{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2}. \quad (18)$$

Substituindo em  $\sum_i \sum_j \gamma^2 \tau_i^2 \beta_j^2$ , obtemos

$$SQTR.B = \frac{\left[ \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..}) (\bar{y}_{.j} - \bar{y}_{..}) y_{ij} \right]^2}{\sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2}. \quad (19)$$



## Teste de Aditividade de Tukey

A decomposição da variabilidade do modelo (16) é

$$SQT = SQTR + SQB + SQTR.B + SQRe, \quad (20)$$

onde  $SQRe$  é a soma de quadrado restante:

$$SQRe = SQT - SQTR - SQB - SQTR.B. \quad (21)$$

Pode-se mostrar que, quando  $\gamma = 0$ , então  $SQTR.B$  e  $SQRe$  são independentes com distribuição  $\chi^2_1$  e  $\chi^2_{ab-a-b}$ , respectivamente.

## Teste de Aditividade de Tukey

Então, para o teste de hipóteses

$$\begin{cases} H_0 : \gamma = 0 \\ H_1 : \gamma \neq 0, \end{cases} \quad (22)$$

sob  $H_0$ , a estatística teste

$$F^* = \frac{SQTR.B/1}{SQRe/(ab - a - b)} \quad (23)$$

segue distribuição  $F_{1,ab-a-b}$ .

## Parcelas Perdidas

É comum que ocorra perda de observações em delineamentos completamente aleatorizados em blocos.

Quando há perda de observações, o balanceamento (ortogonalidade) do delineamento e torna os cálculos da Anova usual inapropriados.

Em caso de observações faltantes ou parcelas perdidas, podemos considerar duas abordagens:

- 1 Estimar o valor ausente.
- 2 Abordagem de regressão (modelo linear geral).

## Parcelas Perdidas

### Estimação do valor ausente:

No caso de uma observação perdida, uma opção é substituir seu valor por  $x$ , obter a SQR e encontrar o valor de  $x$  que minimiza a SQR. Como,

$$SQR = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2. \quad (24)$$

Segue que,

$$x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}. \quad (25)$$

onde  $y'_{i.}$  é o total das  $a - 1$  observações disponíveis no  $i$ -ésimo tratamento,  $y'_{.j}$  é o total das  $b - 1$  observações disponíveis no  $j$ -ésimo bloco e  $y'_{..}$  é o total das  $ab - 1$  observações disponíveis.

# Teste Linear Geral

## **Abordagem de regressão (modelo linear geral):**

Antes da computação de alta velocidade, a imputação de dados costumava ser feita porque os cálculos da Anova são feitos mais prontamente usando um projeto balanceado.

Em algumas situações, a imputação do(s) valor(es) ausente(s) ainda pode ser útil. Porém, na maioria das situações podemos considerar a abordagem de modelo linear geral e a abordagem de modelo completo e reduzido para fazer o teste apropriado (*teste linear geral*).

# Teste Linear Geral

## Abordagem de regressão (modelo linear geral):

- Usa a SQ ajustada tipo II: em modelo linear geral, o ajustamento da soma de quadrados é realizada para levar em conta a falta de ortogonalidade (desbalanceamento) do delineamento. Uma SQ ajustada tipo II para um termo envolve ajustar para todos os outros termos no modelo que não contém o termo em questão (é invariante à ordem dos termos no modelo).
  - Para testar a hipótese  $H_0 : \tau_1 = \dots = \tau_a = 0$ , iremos usar a soma de quadrados ajustada:  $R(\tau|\mu, \beta)$ .

# Teste Linear Geral

## Teste linear geral:

Teste linear geral (teste-F linear geral) envolve três passos básicos:

- 1 Definir o modelo completo (modelo maior: o modelo com mais parâmetros).
- 2 Definir o modelo reduzido (modelo menor: modelo com menos parâmetros).
- 3 Usar a estatística  $F^*$  para decidir se o modelo menor deve ser rejeitado em favor do modelo maior.

A hipótese nula sempre afirma o modelo reduzido, enquanto a hipótese alternativa afirma o modelo completo.

## Teste Linear Geral

### Teste linear geral:

O teste linear geral envolve a comparação entre a soma de quadrados de resíduos do modelo reduzido,  $SQR(R)$  e a soma de quadrados de resíduos do modelo completo,  $SQR(F)$ .

O valor de  $SQR(R)$  não pode ser menor do que  $SQR(F)$  (pode ser igual).

Se  $SQR(R)$  está próximo de  $SQR(F)$ , então a variação em torno do modelo completo estimado é quase tão grande quanto a variação em torno do modelo reduzido estimado. Nesse caso, faz sentido usar o modelo reduzido.

Se  $SQR(R)$  está distante de  $SQR(F)$ , então os parâmetros adicionais do modelo completo reduzem substancialmente a variação em torno do modelo estimado. Nesse caso, faz sentido usar o modelo completo.



## Teste Linear Geral

### Teste linear geral:

As hipóteses do teste são

$$\begin{cases} H_0 : \text{Modelo Reduzido} \\ H_1 : \text{Modelo Completo.} \end{cases} \quad (26)$$

A estatística teste é dada por

$$F^* = \left( \frac{SQR(R) - SQR(F)}{g.l.R - g.l.F} \right) \div \left( \frac{SQR(F)}{g.l.F} \right). \quad (27)$$

A um nível de significância  $\alpha \in (0, 1)$ , rejeita-se a hipótese nula se  $F^* > F_{\alpha; g.l.R - g.l.F, g.l.F}$ .

## Exemplos

**Exemplo 4.** O conjunto de dados *oatvar* no arquivo *dados\_aula2.xlsx* contém informações sobre um experimento em oito variedades diferentes de aveia. A área em que o experimento foi feito tinha alguma variabilidade sistemática e a pesquisadora dividiu a área em cinco blocos diferentes nos quais a área era uniforme (embora reconhecendo que alguns blocos são provavelmente superiores a outros para o cultivo). Dentro de cada bloco, a pesquisadora criou oito parcelas e atribuiu aleatoriamente uma variedade a cada parcela. ■

## Exemplos

**Exemplo 5.** O conjunto de dados *sementes* no arquivo *dados\_aula2.xlsx* contém os resultados de um experimento para comparar os efeitos de três pesticidas diferentes em uma variedade de vagens. Para obter uma quantidade suficiente de dados, foi necessário usar quatro lotes de terra diferentes. A resposta de interesse foi o número de mudas que emergiram. ■

## Exemplos

**Exemplo 6.** O conjunto de dados *droga* no arquivo *dados\_aula2.xlsx* traz os resultados de um experimentos onde o interesse é a taxa na qual ratos privados de água pressionavam uma alavanca para obter água. Os níveis do fator de tratamento foram cinco dosagens diferentes de anfetamina em miligramas por quilograma de peso corporal, incluindo uma dosagem de controle que consiste em solução salina. Como havia grande variabilidade na taxa de pressão à barra entre os ratos, o fator de bloqueamento é o rato. Cada rato recebeu todas as cinco doses em ordem aleatória com um período apropriado entre elas. ■

## Exemplos

**Exemplo 7.** O conjunto de dados *enxerto* no arquivo *dados\_aula2.xlsx* contém os resultados de um experimento para investigar um procedimento para criar artérias artificiais usando uma resina. A resina é prensada através de uma abertura que transforma a resina em um tubo. Para conduzir este experimento, precisamos atribuir todas as 4 pressões aleatoriamente a cada um dos 6 lotes de resina. Cada batelada de resina é chamada de "bloco", já que uma batelada é um conjunto mais homogêneo de unidades experimentais para testar as pressões de extrusão. A resposta de interesse é a porcentagem de produtos dentro de especificação. ■

## Exemplos

**Exemplo 8.** Considere um experimento para comparar quatro processos, A, B, C e D, de produção de penicilina. Sabe-se que uma matéria prima importante, o xarope de milho, é muito variável e a quantidade de mistura produzida desse ingrediente somente é suficiente para quatro rodadas do experimento. Portanto, um delineamento completamente aleatorizado em bloco parece ser adequado. Os dados estão na planilha *penicilina* do arquivo *dados\_aula2.xlsx*. ■

## Exemplos

**Exemplo 9.** Um experimento foi conduzido para estudar a performance de quatro tipos diferentes de detergentes na lavagem de roupas. A resposta de interesse é uma medida quantitativa do índice de limpeza da roupa. Como pode haver variação na resposta devida ao tipo de mancha, este fator foi controlado localmente. Os dados são mostrados na planilha *detergente* do arquivo *dados\_aula2.xlsx*. ■

## Exemplos

**Exemplo 10.** Um experimento foi realizado para investigar o efeito de estrogênio no ganho de peso de ovelhas. Os tratamentos são combinações de sexo (masculino, feminino) e níveis de estrogênio (Est0, Est1). As ovelhas foram agrupadas (bloquadas) por fazenda com uma replicação de cada tratamento por fazenda. A variável resposta de interesse é o peso ganho em libras. Os dados estão na planilha *ovelha* do arquivo *dados\_aula2.xlsx*. ■



## Usando o R

```
library(readxl)  
library(ggthemes)  
library(ggpubr)  
library(MASS)  
library(car)  
library(DescTools)  
library(agricolae)  
library(tidymodels)  
library(tidyverse)
```

```
dados <- read_xlsx("../dados/dados_aula2.xlsx",  
                  sheet = "detergente")
```

```
dados$detergente <- as.factor(dados$detergente)  
dados$mancha <- as.factor(dados$mancha)
```

```
glimpse(dados)
```

```
## Rows: 12
```

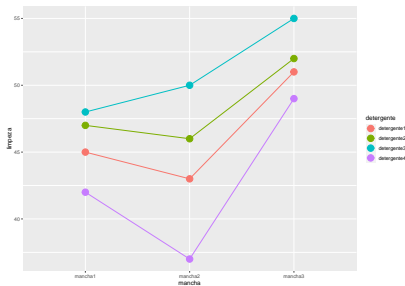
```
## Columns: 3
```

```
## $ detergente <fct> detergente1, detergente2, det
```

```
## $ mancha <fct> mancha1, mancha1, mancha1, ma
```

```
## $ limpeza <dbl> 45, 47, 48, 42, 43, 46, 50, 3
```

```
ggplot(dados, aes(y = limpeza,
                   x = mancha, color = detergente)) +
  geom_point(size = 5) +
  geom_line(aes(x = as.integer(mancha)))
```



```
fit <- aov(limpeza ~ detergente + mancha, data = dados)
```

```
summary.aov(fit)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## detergente    3 110.92   36.97    11.78 0.00631 **
## mancha        2 135.17   67.58    21.53 0.00183 **
## Residuals     6  18.83    3.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

```
model.tables(fit, "means")
```

```
## Tables of means
## Grand mean
##
## 47.08333
##
## detergente
## detergente
## detergente1 detergente2 detergente3 detergente4
##          46.33          48.33          51.00          42.67
##
## mancha
## mancha
## mancha1 mancha2 mancha3
##      45.50      44.00      51.75
```

```
comp_bon <- LSD.test(fit, "detergente", alpha = 0.05, p.adj = "bonferroni",
                     console = FALSE)
```

```
comp_bon
```

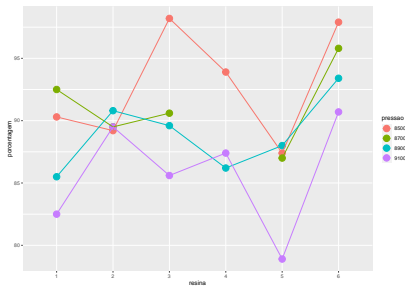
```
## $statistics
##      MSerror Df      Mean      CV  t.value      MSD
##      3.138889  6 47.08333 3.762883 3.862991 5.588123
##
## $parameters
##      test  p.adjusted      name.t ntr alpha
##      Fisher-LSD bonferroni detergente  4  0.05
##
## $means
##      limpeza      std r      LCL      UCL Min Max  Q25 Q50  Q75
## detergente1 46.33333 4.163332 3 43.83042 48.83625 43 51 44.0 45 48.0
## detergente2 48.33333 3.214550 3 45.83042 50.83625 46 52 46.5 47 49.5
## detergente3 51.00000 3.605551 3 48.49709 53.50291 48 55 49.0 50 52.5
## detergente4 42.66667 6.027714 3 40.16375 45.16958 37 49 39.5 42 45.5
##
## $comparison
##      NULL
##
## $groups
##      limpeza groups
## detergente3 51.00000      a
## detergente2 48.33333      a
## detergente1 46.33333     ab
## detergente4 42.66667      b
##
## attr(,"class")
## [1] "group"
```

## Usando o R

```
dados <- read_xlsx("../dados/dados_aula2.xlsx",  
                  sheet = "enxerto_miss")
```

```
dados$resina <- as.factor(dados$resina)  
dados$pressao <- as.factor(dados$pressao)
```

```
ggplot(dados, aes(y = porcentagem,
                   x = resina, color = pressao)) +
  geom_point(size = 5) +
  geom_line(aes(x = as.integer(resina)))
```





```
fit <- lm(porcentagem ~ pressao + resina, data = dados)
```

```
Anova(fit, type = "II")
```

```
## Anova Table (Type II tests)
```

```
##
```

```
## Response: porcentagem
```

##		Sum Sq	Df	F value	Pr(>F)	
##	pressao	163.40	3	7.4981	0.003130	**
##	resina	189.52	5	5.2181	0.006533	**
##	Residuals	101.70	14			

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
```

```
library(lsmeans)
```

```
lsmeans(fit, "pressao")
```

```
##   pressao lsmean    SE df lower.CL upper.CL  
##   8500      92.8 1.10 14     90.5     95.2  
##   8700      91.1 1.24 14     88.4     93.7  
##   8900      88.9 1.10 14     86.6     91.3  
##   9100      85.8 1.10 14     83.4     88.1  
##
```

```
## Results are averaged over the levels of: resina
```

```
## Confidence level used: 0.95
```

```
comp_bon <- LSD.test(fit, "pressao", alpha = 0.05, p.adj = "bonferroni",  
                     console = FALSE)
```

```
comp_bon
```

```
## $statistics  
##      MSerror Df      Mean      CV  
##      7.264 14 89.58261 3.008598  
##  
## $parameters  
##      test  p.adjusted name.t ntr alpha  
## Fisher-LSD bonferroni pressao 4 0.05  
##  
## $means  
##      percentagem      std r      LCL      UCL  Min  Max    Q25  Q50    Q75  
## 8500      92.81667 4.577081 6 90.45675 95.17658 87.4 98.2 89.475 92.1 96.900  
## 8700      91.08000 3.304088 5 88.49484 93.66516 87.0 95.8 89.500 90.6 92.500  
## 8900      88.91667 2.966760 6 86.55675 91.27658 85.5 93.4 86.650 88.8 90.500  
## 9100      85.76667 4.445072 6 83.40675 88.12658 78.9 90.7 83.275 86.5 88.975  
##  
## $comparison  
## NULL  
##  
## $groups  
##      percentagem groups  
## 8500      92.81667      a  
## 8700      91.08000      a  
## 8900      88.91667     ab  
## 9100      85.76667      b  
##  
## attr(,"class")  
## [1] "group"
```