

# **Programa R: Aplicações no Melhoramento de Plantas**

**Prof. Luiz Alexandre Peternelli  
2011**

Departamento de Estatística – UFV  
Pesquisador do PMGCA – UFV

# Objetivo do mini curso

- O objetivo deste curso não é tão somente ensinar um software ou rever conceitos de estatística e seus métodos, mas proporcionar um ponto de partida para pessoas que desejam começar a utilizar o R e suas ferramentas estatísticas

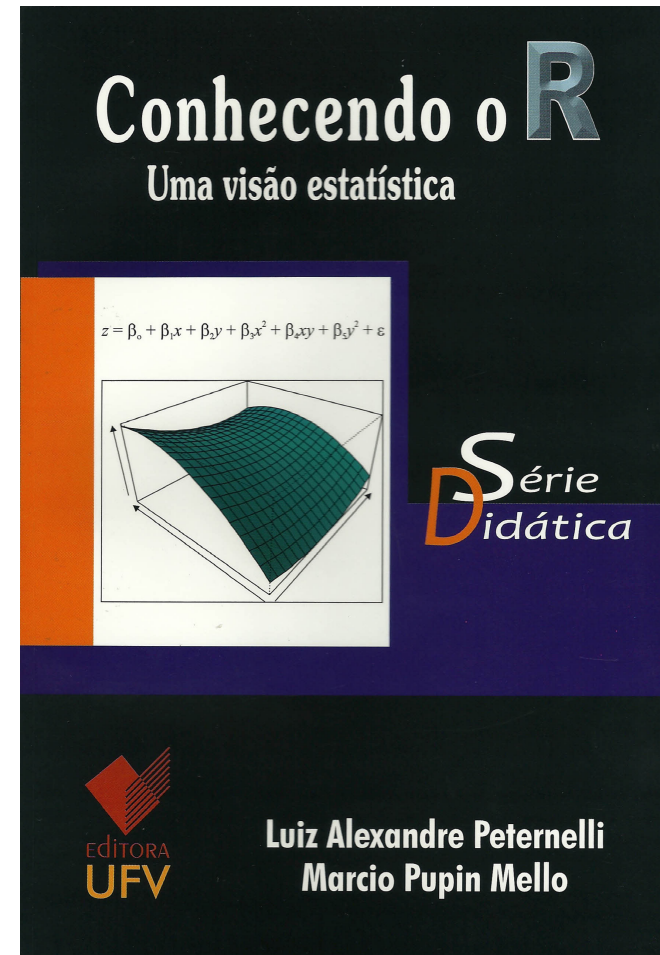
# Conteúdo

- Instalação
- entrada de dados; funções básicas; operações com matrizes
- análises descritivas
- testes de hipóteses
- delineamentos experimentais
- esquemas experimentais
- análise de regressão
- dicas de organização dos comandos e de elaboração dos relatórios de análise

# Material de consulta

- Internet
- Inúmeros [livros](#) (veja [www.amazon.com](http://www.amazon.com))
  - Estatística básica, gráficos, análises específicas etc
- Conhecendo o R: uma visão estatística

Peternelli e Pupin, 2011.  
disponível na Editora UFV  
([www.editoraufv.com.br](http://www.editoraufv.com.br))



# Editores para o R

- No caso do Windows
  - Descrição detalhada em:  
[www.sciviews.org/\\_rgui/projects/Editors.html](http://www.sciviews.org/_rgui/projects/Editors.html)
  - Word: uma opção. CUIDADO!
  - Script do R e notepad. Bons e simples!
  - Tinn-R ([www.sciviews.org/Tinn-R/](http://www.sciviews.org/Tinn-R/))
    - Colorido e interativo.
    - Problema: versões do Windows.
  - RWindEdt (pacote do R) - ??? (RStudio)

# Instalação

- Programa disponível em:

[www.r-project.org](http://www.r-project.org)

Etapas:

- Baixar programa
- Atualizar pacotes
- Baixar pacotes de interesse e usar suas funções, ou criá-las, conforme interesse específico.

# Informações Básicas do Programa R

- Entrando com dados
  - comandosR - [entrando com dados](#)

# Comandos R

```
#Entrando com dados
```

```
#scan(), edit(), read.table() e data().
```

```
x<-3
```

```
x<-c(1,2,3)
```

```
y<-edit(x)
```

```
x<-scan()
```

```
x<-read.table(argumentos) #veremos oportunamente
```

```
#Uso da função scan()
```

```
teste<-scan()      #usado para entrada de dados
```

```
#Uso da função edit()
```

```
teste<-c(10,20,30,40,50)
```

```
teste
```

```
teste2<-edit(teste)
```

```
teste2
```



# Informações Básicas do Programa R

- Operações com matrizes
  - comandosR – [operações com matrizes](#)

# Comandos R

```
#Operações com vetores e matrizes
coluna1<-c(2,1,0)
coluna2<-c(1,3,1)
coluna3<-c(1,1,2)
A<-cbind(coluna1,coluna2,coluna3)
```

```
#obtendo a transposta de A
t(A)
```

```
#cálculos interessantes
t(A)%*%A #A'A
```

```
#Produto de vetores
coluna1%*%coluna1
```

```
#cuidado!
coluna1*coluna1
```

```
#Inversa de uma matriz não singular
  quadrada:
solve(A) #inversa da matriz A
solve(A)%*%A      #verificação
round(solve(A)%*%A,5)
solve(t(A)%*%A)
```

```
#determinante da matriz A
det(A)
```

```
#dando nomes às linhas e colunas
dimnames(A)<-
  list(c("linha1","linha2","linha3"
),c("col1","col2","col3"))
```

```
row.names(A)<-c("l1","l2","l3")
```

```
#criando matriz de outra forma
B<-matrix(1:12,nrow=3,ncol=4)
```

```
B<-
  matrix(1:12,nrow=3,ncol=4,byrow=F
)
```

```
B<-
  matrix(1:12,nrow=3,ncol=4,byrow=T
)
```

```
#selecionando partes de uma matriz
B[1,3]; B[c(1,2),3];B[c(1,3),c(2,4)]
```

```
#Para outras funções associadas a
  matrizes, veja
```

```
??matrix
```

# Informações Básicas do Programa R

- Outras funções básicas
  - comandos R – funções básicas (tabelas a seguir)

# Algumas funções

Ação	Comando
Fazer com que o R ignore o que será digitado*	#
Sair do programa	q()
Salva o trabalho realizado	save.image()
Lista todos os objetos da área de trabalho atual	ls()
Remove o objeto x	rm(x)
Remove os objetos x e y	rm(x, y)
Dado ausente	NA
Verdadeiro se existir dados ausentes	is.na
Mostra todos os pacotes instalados**	library()
Carregar (por exemplo) o pacote nlme	library(nlme)

# Mais funções

Ação de ajuda	Comando
Obter ajuda sobre o comandoX	<code>help(comandoX)</code>
Iniciar ajuda no <i>browser</i> padrão instalado	<code>help.start()</code>
Obter ajuda sobre (p.ex.) o pacote mva	<code>help(package=mva)</code>
Procurar por <i>multivariate</i> em todos os pacotes instalados	<code>help.search("multivariate")</code>
Comando que procura objetos (p.ex.) pelo nome modelo	<code>apropos("modelo")</code>
Mostrar exemplos do "comandoX"	<code>example(comandoX)</code>
Listar as funções e operações contidas no pacote base do R	<code>ls("package:base")</code>

# Mais funções

Função	Significado
<code>log(x)</code>	Log de base e de x
<code>exp(x)</code>	Antilog de x ( $e^x$ )
<code>log(x, n)</code>	Log de base n de x
<code>log10(x)</code>	Log de base 10 de x
<code>sqrt(x)</code>	Raiz quadrada de x
<code>choose(n, x)</code>	$n!/(x!(n-x)!)$
<code>cos(x)</code> , <code>sin(x)</code> , <code>tan(x)</code>	Funções trigonométricas de x em radianos
<code>acos(x)</code> , <code>asin(x)</code> , <code>atan(x)</code>	Funções trig. inversas de x em radianos
<code>abs(x)</code>	Valor absoluto de x

# Análises Descritivas

- Medidas de posição (OBS)

- Média (`mean()`)

- Mediana: (`median()`)

- $n$  ímpar  $\rightarrow X_{\text{md}} = X_{(n+1)/2}$  do rol

- $n$  par  $\rightarrow X_{\text{md}} = 1/2(X_{n/2} + X_{n/2+1})$  do rol

- Moda: (`table()`)

- $X_{\text{mo}}$ : é o valor mais frequente num conjunto de dados

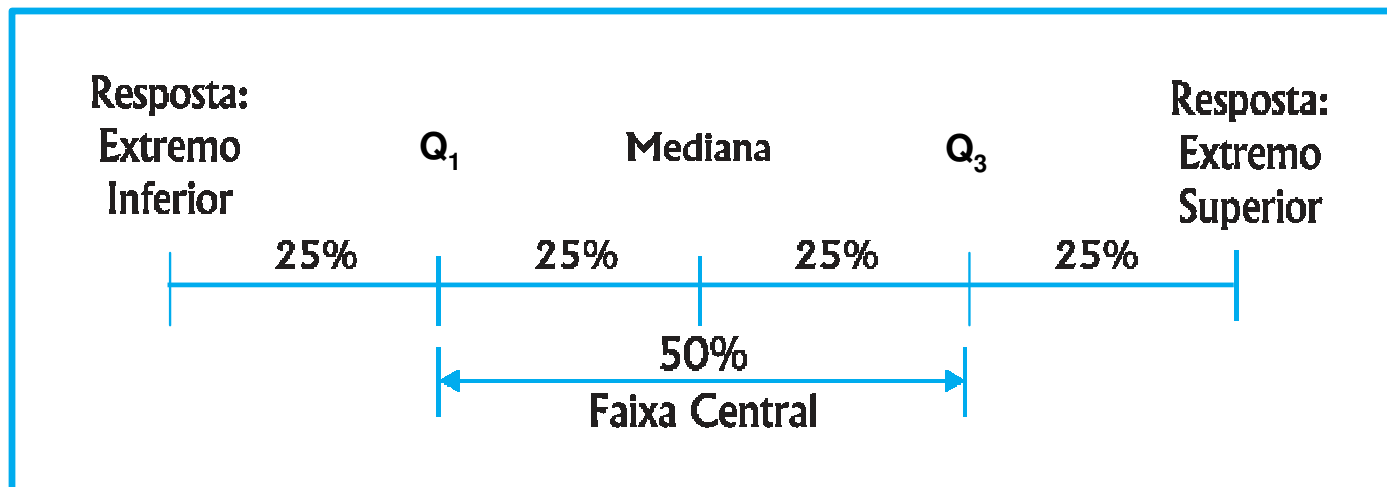
- Máximo (`max()`)

- Mínimo (`min()`)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

# Cont...

- Quantis (quartis, decis, percentis) (`quantile()`)





# Análises Descritivas

- Medidas de dispersão

- Amplitude total  $\rightarrow AT = \max(X) - \min(X)$

- Variância (`var()`)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{SQD_x}{n-1}$$

- Desvio padrão  $\rightarrow \sqrt{s^2}$

- Erro padrão da média  $\rightarrow \frac{s}{\sqrt{n}}$

- Coeficiente de variação  $\rightarrow \frac{s}{\bar{x}} \times 100$

[Usando o R 1](#)  
[Usando o R 2](#)

# Comandos R

```
#medidas descritivas - comandos  
simples
```

```
x<-sample(1:10,10,r=T) #entrando com  
um conjunto de dados qualquer.  
Nossa amostra.
```

```
n<-length(x) #obtendo o tamanho da  
amostra
```

```
mean(x) #média
```

```
median(x) #mediana
```

```
table(x) #dica para obter a moda (ou  
modas)
```

```
max(x) #máximo
```

```
min(x) #mínimo
```

```
quantile(x,c(.25,.5,.75)) #três  
quantis diferentes
```

```
#OBS: ver o que sai da função  
summary(x)
```

```
max(x)-min(x) #amplitude total
```

```
var(x) #variância
```

```
sqrt(var(x)) #ou sd(x). Desvio  
padrão
```

```
sd(x)/sqrt(n) #erro padrão da média
```

```
sd(x)/mean(x)*100 #coeficiente de  
variação
```

# Comandos R

#medidas descritivas - criando função

```
meu.estat.descr<-function(conj.de.dados,quantis=c(.25,.5,.75))
{
  x<-conj.de.dados
  n<-length(x)
  mx<-mean(x)
  medx<-median(x)
  maxx<-max(x)
  minx<-min(x)
  #qx<-quantile(x,c(.25,.5,.75))
  qx<-quantile(x,quantis)
  vx<-var(x)
  epmx<-sd(x)/sqrt(n)
  return(list(media=mx,mediana=medx,maximo=maxx,minimo=minx
    ,quantis=qx,var.de.x=vx,erro.padrao=epmx))
}
```

#testando

```
teste<-sample(1:10,10,r=T)
meu.estat.descr(teste)
meu.estat.descr(teste,c(.05,.95))
j<-meu.estat.descr(teste,c(.05,.95))
j$media
```

# Análises Descritivas

- Análises gráficas
  - Importância
  - Gráficos mais comuns:
    - Histograma
    - Boxplot
    - Diagrama de dispersão

[Usando o R](#)

# Comandos R

```
#comandos para graficos
```

```
x<-1:20
y<-x**3
plot(x,y)    #plota os pares (x,y).
```

```
plot(x,y,type="l")
```

```
plot(x,y,type="b")
plot(x,y,type="o")
plot(x,y,type="s")
plot(x,y,type="c")
plot(x,y,type="h")
```

```
par(mfrow=c(3,2))
```

```
#adicionando mais dados aos gráficos
plot(x,y)
points(rev(x),y) #adicionando pontos
lines(x,8000-y)
    #adicionando linhas
```

```
#mudando o padrão de pontos
```

```
plot(x,y)
points(rev(x),y,pch=3)    #adiciona
    cruces
points(x,8000-y,pch="$") #símbolo
    cifrão
```

```
plot(x,y)
plot(x,y,pch="@")
plot(x,y,pch=1:3)
plot(0:20,0:20,pch=0:20)
```

```
#mudando as linhas
```

```
plot(x,y)
lines(x,y,lwd=4)          #linha
    grossa
lines(rev(x),y,lty=2)     #linha
    tracejada
```

# Comandos R

```
#definindo os intervalos dos eixos
plot(c(0,20),c(-
      8000,8000),type='n')
lines(x,y)                #lembre-se:
      x<-1:20; y<-x**3
lines(x,-y)
```

```
#adicionando segmentos de reta
```

```
plot(c(0,20),c(0,30),type='n')#gráfi
co em branco
segments(5,3,15,20)        #um
segmento de reta
lines(c(12,15,7,20),
      c(3,10,8,15),        #duas
      pequenas linhas
      col="red",           #vermelhas
      lty=3)               #pontilhadas
abline(30,-2,lty=2,col=4)#linha
tracejada azul
```

```
#adicionando texto
```

```
plot(x,y,xlab="Eixo X
      aqui",ylab="Eixo Y aqui")
title("Título aqui \n (e repare a
      acentuação!!!)")
text(6,4000,"Texto em qualquer
      lugar")
```

```
#gráficos múltiplos
```

```
x<-1:20
y<-x**3
```

```
par(mfrow=c(2,2))
      #arranjo "2 por 2"
plot(x,y)
plot(rev(x),y)
plot(x,2*y)
plot(x,log(y))
```

# Comandos R

```
#personalizando o gráfico
x<-1:10
  #criando x
y<-c(2,5,9,6,7,8,4,1,3,10)
  #criando y
x;y
plot(x,y)          #plota x e y
plot(x, y,         #plota x e y
      xlab="Eixo x",#nomeia o eixo x
      ylab="Eixo y",#nomeia o eixo y
      main="Personalizando um
gráfico",#título
      xlim=c(0,10), #limites eixo x
      ylim=c(0,10), #limites eixo y
      col="red",     #cor dos pontos
      pch=22,        #formato dos pontos
      bg="blue",     #cor de preenchimento
      tcl=0.4,       #tamanho de traços dos
                    eixos
      las=1,         #orientação do texto em
                    y
      cex=1.5,       #tamanho do ponto
      bty="l")       #altera as
bordas
```

```
#histogramas
x<-
  c(2,2,2,2,2,3,3,3,4,4,5,5,6
  ) #vetor qualquer
hist(x)

table(x)

hist(x,          #histograma
     de x
     right=T,    #intervalos
                fechados à direita
     include.lowest=F, #não
                soma extremos do vetor
     breaks=c(1,2,3,4,5,6))
                #int. das classes
```

# Comandos R

```
x<-rchisq(1000,10); hist(x)
hist(x,                      #histograma de x
      main="Histograma\nQui-quadrado", #título
      xlab="Valores",             #texto do eixo das abscissas
      ylab="Prob",               #texto do eixo das ordenadas
      br=c(c(0,5),c(5,15),5*3:7), #int das classes
      xlim=c(0,30),             #limites do eixo de x
      ylim=c(0,0.1),            #limites do eixo y
      col="lightblue",          #cor das colunas
      border="white",           #cor das bordas das colunas
      prob=T,                   #mostrar probabilidades.
      right=T,                  #int fechados à direita
      adj=0,                    #alinhamento dos textos
      col.axis="red")           #cor do texto nos eixos

dados<-c(25,27,18,16,21,22,21,20,18,23,27,21,19,20,21,16)
hist(dados,                  #histograma de dados
      nc=6,                  #número de classes igual a 6
      right=F,               #int fechado à esquerda
      main="Histograma",     #título do histograma
      xlab="tempo (em minutos)", #texto do eixo x
      ylab="frequencia",     #texto do eixo y
      col=2)                 #usa a cor cinza nas barras
```



# Comandos R

```
#gráfico de barras  
barplot(table(c(  
  "a", "a", "a", "a", "a", "b", "b", "b", "c", "c", "v", "v")))
```

```
dados<-c("a"=4, "b"=7)  
barplot(dados)
```

```
barplot(table(c(  
  "a", "a", "a", "a", "a", "b", "b", "b", "c", "c", "v", "v")),  
  hor=T)#disposição horizontal das categorias
```

```
#boxplot  
x<-rnorm(100,10)  
y<-rchisq(100,10)  
boxplot(x,y)
```

```
#gráfico de setores: veja ?pie
```

```
demo(graphics)
```

# Testes de Hipóteses

- Entendendo os princípios básicos usando simulação
- Ideia geral

[Usando o R](#)

# Ideia geral

- regra decisória que nos permite rejeitar ou não rejeitar uma hipótese estatística com base nos resultados de uma amostra
- Parâmetro, estimador, estimativa
- Hipótese estatística:
  - É uma suposição quanto ao valor de um parâmetro populacional, ou uma afirmação quanto à natureza da população
- $H_0$  e  $H_a$

# Ideia geral – cont...

- Erros nas conclusões obtidas

Por causa das flutuações amostrais, ao testar certa hipótese e tomar uma decisão, pode-se tomar a decisão errada.

## Dois tipos de erro:

- ***Erro Tipo I:*** Rejeitar a hipótese nula quando na realidade ela é verdadeira.
- ***Erro Tipo II:*** Não rejeitar a hipótese nula quando na realidade ela é falsa.

## Ideia geral – cont...

- $P(\text{erro tipo I}) = \alpha$
- $P(\text{erro tipo II}) = \beta$
- Em geral podemos controlar apenas o erro tipo I, definindo o valor  $\alpha$  (nível de significância do teste)

# Passos usuais para a Construção de um Teste de Hipóteses

**Passo 1.** Definir uma hipótese  $H_0$  a ser testada e uma hipótese alternativa  $H_A$ ;

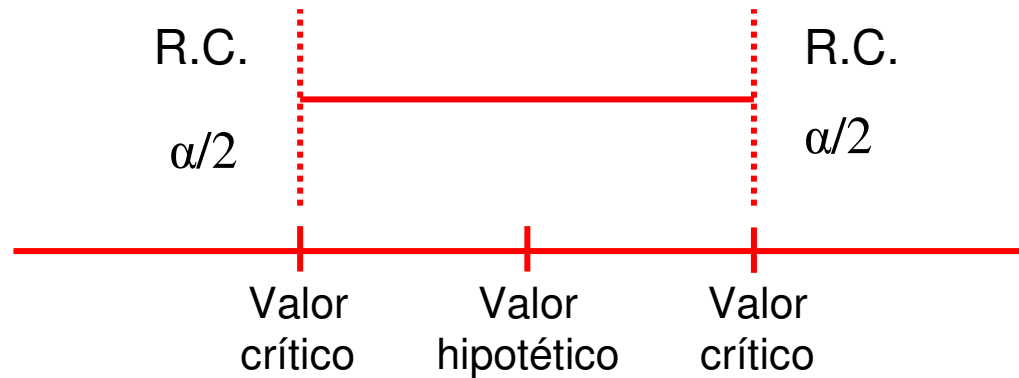
**Passo 2.** Usar a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para testar a hipótese  $H_0$ . Definir a distribuição amostral dessa estatística.

# Passos para a Construção de um Teste de Hipóteses

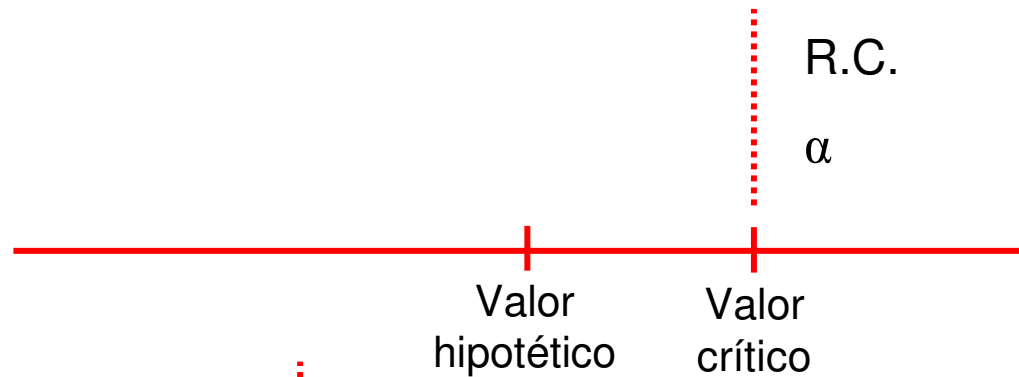
- Passo 3.** Fixar a probabilidade  $\alpha$  de cometer o erro tipo I e usar este valor para construir a região crítica (regra de decisão);
- Passo 4.** Usar as observações da amostra para calcular o valor da estatística de teste;
- Passo 5.** Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite  $H_0$ ; caso contrário, rejeite  $H_0$ .

# Esquemas baseados na $H_a$

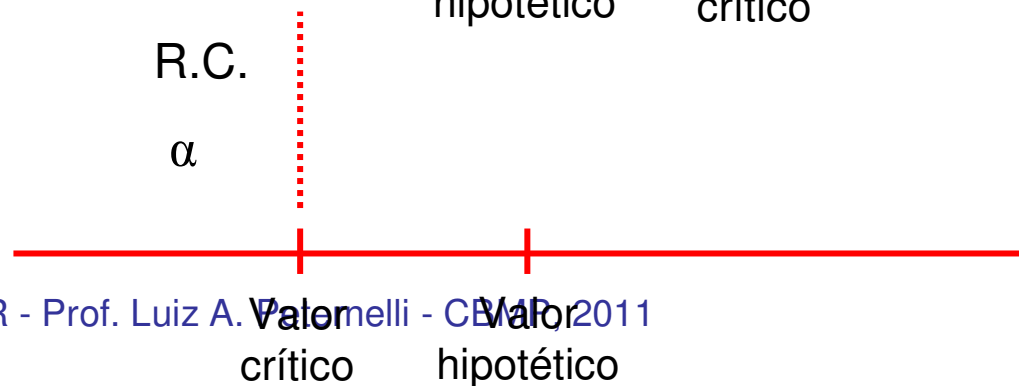
$$H_a : \theta \neq \theta_0$$



$$H_a : \theta > \theta_0$$



$$H_a : \theta < \theta_0$$





# Usando o p-valor

- O procedimento para o teste seria:
  - Formular  $H_0$  e  $H_a$  (e definir  $\alpha$ , se for de interesse);
  - Especificar a estatística do teste;
  - Determinar o valor da estatística do teste e o p-value correspondente baseado na amostra; e
  - Comparar p-value com  $\alpha$ 
    - Se  $\text{p-value} \leq \alpha \Rightarrow$  rejeição de  $H_0$ ;
    - Se  $\text{p-value} > \alpha \Rightarrow$  não-rejeição de  $H_0$ .
    - *Obs.: Ou permita que o julgamento seja particularizado.*

# Teste t

Hipóteses

$$H_0: \mu = \mu_0 \text{ vs } \begin{cases} H_A: \mu > \mu_0 \\ H_A: \mu < \mu_0 \\ H_A: \mu \neq \mu_0 \end{cases} \quad n \leq 30$$

Estatística de Teste:

$$t_{calc} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n - 1 \text{ g.l.})$$

Regra decisória: Se  $|t_{calc}| > t_{tab} \rightarrow$  rejeitamos  $H_0$

**Exemplo:** Foi afirmado que o crescimento de certa planta, num certo período, é de 140 cm (ou mais). Um pesquisador deseja testar a credibilidade da declaração. Para isso mediu a capacidade de uma amostra aleatória de 20 dessas plantas, durante aquele período declarado. Os resultados em cm, são os seguintes: (use  $\alpha = 5\%$ )

137,4   140,0   138,8   139,1   144,4   139,2  
141,8   137,3   133,5   138,2   141,1   139,7  
136,7   136,3   135,6   138,0   140,9   140,6   136,7  
134,1

ASSIM:  $H_0: \mu = 140$   
 $H_a: \mu < 140$

$n = 20$

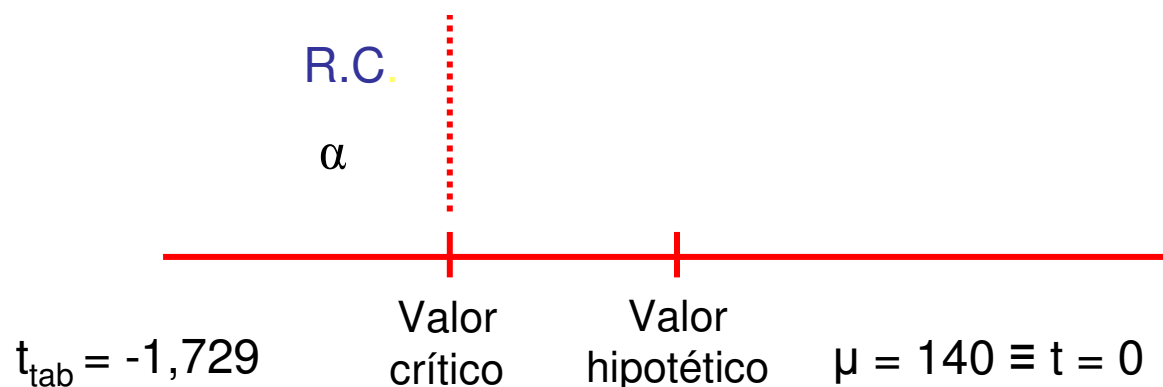
$\alpha = 0,05$

$s^2 = 7,0706$     $s = 2,66$

Estatística  
do teste:

$$t_{calc} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{138,47 - 140}{\frac{2,66}{\sqrt{20}}} = -2,57$$

Esquema



Conclusão:

Rejeita-se  $H_0$ , ou seja, há evidências de que a declaração é falsa, ao nível de significância de 5%.

# Comandos R

```
?t.test
#t.test(x, y = NULL,
#       alternative = c("two.sided", "less", "greater"),
#       mu = 0, paired = FALSE, var.equal = FALSE,
#       conf.level = 0.95, ...)
x<-scan(dec=",")
137,4    140,0    138,8    139,1    144,4    139,2
141,8    137,3    133,5    138,2    141,1    139,7
136,7    136,3    135,6    138,0    140,9    140,6    136,7    134,1
t.test(x,alternative="less",mu=140)
qt(.05,19)
?pt
pt(2.5734,19)
pt(-2.5734,19)
1-pt(2.5734,19)
qt(.05,19)
saida.t<-t.test(x,alternative="less",mu=140)
mode(saida.t)
names(saida.t)
```

# Outros testes

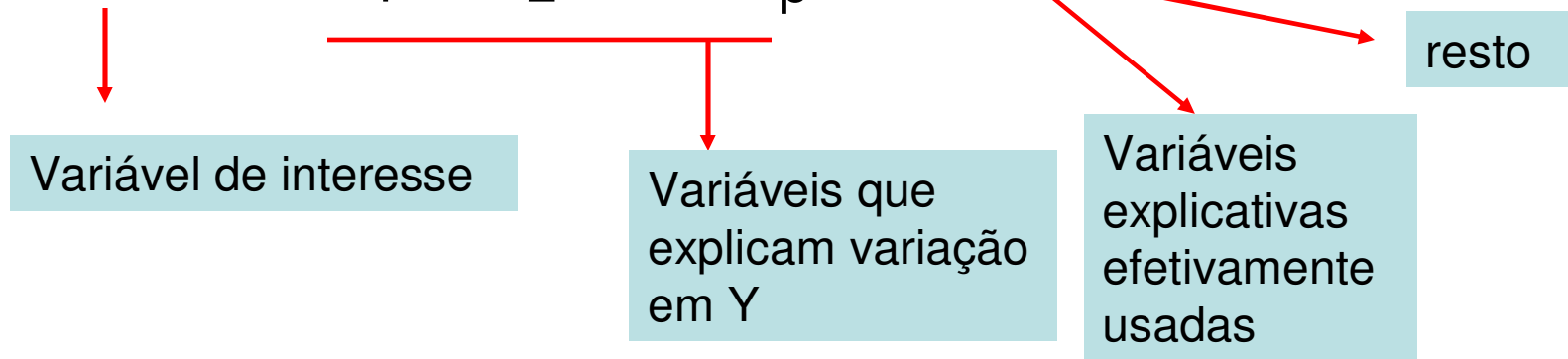
- Depende dos dados e/ou interesse do pesquisador
  - Teste F
  - Teste de Qui-quadrado
  - Teste de Kolmogorov-Smirnov
  - Teste de Shapiro-Wilks
- Veja “??test”

# Comandos R

```
#Kolmogorov-Smirnov
pesos<-scan()
46.88 47.17 64.46 67.84 85.76 65.41 60.10
75.84 61.21 61.65 63.87 53.95 63.66 69.06
76.41 75.56 69.04 35.18 66.42 58.78 73.02
51.69 90.88 53.01 64.31 61.91 79.42 57.78
62.73 60.63 63.29 46.53 84.64 61.76 85.08
59.66 54.89 94.18 59.89 68.56 75.66 72.06
62.00 43.43 73.38 73.31 66.37 73.72 66.15
67.79
hist(pesos)
ks.test(pesos,      #amostra a ser testada
        "pchisq",   #"p" e depois nome da distribuição
        49)         #graus de liberdade da amostra
curve(dchisq(x,49),30,100,add=T)
hist(pesos,prob=T)
curve(dchisq(x,49),30,100,add=T)
mean(pesos)
sd(pesos)
ks.test(pesos,"pnorm",65,12)
curve(dnorm(x,65,12),30,100,add=T,col="blue")
#####
#Shapiro-Wilks
shapiro.test(pesos)
```

# O MODELO ESTATÍSTICO

- $Y = f(X_1, X_2, \dots, X_p) = f(\mathbf{X}, \mathbf{e})$



- Utilizando a notação matricial

$$Y = X\theta + \varepsilon$$

- Pode-se usar representação algébrica



# Exemplo

- Usando representação algébrica

$$y_{ij} = m + t_i + e_{ij}$$

- Ver esse exemplo para  $i = 1$  a  $3$  e  $j = 1$  a  $2$

# Desenvolvimento e restrições do modelo

- Pressuposições básicas
  - Aditividade dos efeitos
  - Normalidade dos erros
  - Independência dos erros
  - Homogeneidade de variâncias dos tratamentos

# DELINEAMENTO INTEIRAMENTE CASUALIZADO

DIC

# Modelo

$$y_{ij} = m + t_i + e_{ij}$$

$$i = 1, 2, \dots, I$$

$$j = 1, 2, \dots, J$$

- $y_{ij}$  é o valor observado para a variável em estudo referente ao  $i$ -ésimo tratamento na  $j$ -ésima repetição
- $m$  é a constante inerente ao modelo
- $t_i$  é o efeito do tratamento  $i$  no valor observado
- $e_{ij}$  é o erro associado à observação  $y_{ij}$

# Construção de estimadores

- Montar quadro da ANOVA

FV	g.l.	SQ	QM	F
Tratamento	I – 1	SQTrat		
Resíduo	IJ – I	SQRes		
Total	IJ – 1	SQTotal		

$$SQTotal = \sum_{ij} y_{ij}^2 - \frac{\left( \sum_{ij} y_{ij} \right)^2}{IJ} \quad SQTrat = \frac{1}{J} \sum_i T_i^2 - \frac{\left( \sum_{ij} y_{ij} \right)^2}{IJ}$$

- $QM_{Trat} = \frac{SQ_{Trat}}{I - 1}$
- $QM_{Res} = \frac{SQ_{Res}}{I(J - 1)}$
- $F = QM_{Trat}/QM_{Res}$

FV	EQM	
	Fixo	Aleat
Trat	$\sigma^2 + J\Phi_t$	$\sigma^2 + J\sigma_t^2$
Res	$\sigma^2$	$\sigma^2$
Total		

$$\Phi_t = \frac{\sum_{i=1}^I t_i^2}{I - 1}$$

$$\hat{\sigma}_t^2 = \frac{QM_{Trat} - QM_{Res}}{J}$$

# DIC – Modelo fixo

- Hipóteses testadas

$$H_0 : m_1 = m_2 = \dots = m_I = m$$

$$H_a : \text{não } H_0$$

- Regra decisória

- se  $F_{\text{calculado}} \geq F_{\text{tabelado}} \rightarrow \text{rejeita-se } H_0$
- Ou avalia o “p-valor”

# Exemplo DIC 1

- GOMES (1984):  
experimento de  
competição de 4  
cultivares de cana-de-  
açúcar (A, B, C e D),  
utilizando o DIC com 6  
repetições

Repetição	Cultivares			
	A	B	C	D
1	54	60	59	45
2	40	55	47	33
3	51	66	44	34
4	36	61	49	48
5	50	54	62	42
6	48	61	60	44
Totais	279	357	321	246



# Comandos R

Arquivo: comandosR/comandosR DIC exemplo1.r

```
A<-scan()  
54 40 51 36 50 48  
B<-scan()  
60 55 66 61 54 61  
C<-scan()  
59 47 44 49 62 60  
D<-scan()  
45 33 34 48 42 44  
totais<-c(sum(A), sum(B), sum(C), sum(D))  
dados<-data.frame(repet=rep(1:6, 4)  
                  , cult=sort(rep(c("A", "B", "C", "D"), 6))  
                  , y=c(A, B, C, D))
```

$$\begin{aligned}
 SQ_{Total} &= \sum_{ij} y_{ij}^2 - \frac{\left( \sum_{ij} y_{ij} \right)^2}{IJ} = (54^2 + 40^2 + \dots + 44^2) - \frac{(54 + 40 + \dots + 44)^2}{4 \times 6} = \\
 &= 62301 - \frac{1203^2}{24} = 62301 - 60300,38 = 2000,62
 \end{aligned}$$

$$\begin{aligned}
 SQ_{Trat} &= \frac{1}{J} \sum_i T_i^2 - \frac{\left( \sum_{ij} y_{ij} \right)^2}{IJ} = \frac{1}{6} (279^2 + \dots + 246^2) - \frac{1203^2}{24} \\
 &= 61474,5 - 60300,38 = 1174,12
 \end{aligned}$$

$$SQ_{Res} = SQ_{Total} - SQ_{Trat} = 2000,62 - 1174,12 = 826,50$$

# Comandos R

```
> attach(dados)
> SQy<-sum(y^2)
> Sy<-sum(y)
> SQTrat<-sum(totais^2)/6
> #obtendo as somas de quadrados
> SQTotal<-SQy-Sy^2/(4*6)
> SQTotal
[1] 2000.625
> SQTrat<-SQTrat-Sy^2/(4*6)
> SQTrat
[1] 1174.125
> SQRes<-SQTotal-SQTrat
> SQRes
[1] 826.5
> detach(dados)
```

$$QM_{Trat} = \frac{SQ_{Trat}}{GL_{Trat}} = \frac{1174,12}{3} = 391,37$$

$$QM_{Res} = \frac{SQ_{Res}}{GL_{Res}} = \frac{826,50}{20} = 41,32$$

$$F_{calc} = \frac{QM_{Trat}}{QM_{Res}} = \frac{391,37}{41,32} = 9,47$$

$$F_{tab} = F_{5\%}(3, 20) = 3,10$$

FV	GL	SQ	QM	F
Trat	3	1174,12	391,37	9,47 *
Res	20	826,50	41,32	
Total	23	2000,62		

$$C.V. (\%) = \frac{\sqrt{QM_{Res}}}{\hat{m}} \cdot 100 = \frac{\sqrt{41,32}}{50,125} \cdot 100 = 12,82 \%$$

# Comandos R

```
> QMTrat<-SQTrat/(4-1)
> QMTrat [1] 391.375
> QMRes<-SQRes/(4*(6-1))
> QMRes
[1] 41.325
> Fcalc<-QMTrat/QMRes
> Fcalc
[1] 9.470659
> Ftab<-qf(0.95,3,20)
> Ftab
[1] 3.098391
> CV<-sqrt(QMRes)/mean(dados$y)*100
> CV
[1] 12.82484
> #Fazendo o quadro da ANOVA diretamente:
> anova(aov(y~cult,data=dados))
> #obtendo o pvalor
> 1-pf(Fcalc,3,20)
[1] 0.0004232293
```

## Exemplo DIC 2

Um pesquisador avaliou a altura de colmos (média do sulco, em dm) de quatro famílias de cana de açúcar organizados segundo o DIC, obtendo os seguintes dados numa escala apropriada

						Total	Média
A	25	26	20	23	21	115	23
B	31	25	28	27	24	135	27
C	22	26	28	25	29	130	26
D	22	28	27	23	20	120	24

Realizar a ANOVA para testar a hipótese de igualdade dos efeitos de tratamentos

# Comandos R

```
#anova DIC - referente a arquivo na  
pasta de dados
```

```
trat<-scan()  
#copiar e colar os valores da coluna  
trat diretamente no console do R
```

```
y<-scan()  
#copiar e colar os valores da coluna  
y diretamente no console do R
```

```
dados<-cbind(trat,y)  
#para evitar erros futuros  
rm(y); rm(trat)
```

```
#ver outras formas de entrar com  
dados  
d1<-read.table("curso R slides/dados  
curso R/DIC 1.txt",h=T)  
d2<-read.csv("curso R slides/dados  
curso R/DIC 1.csv")
```

```
#cuidado!  
aov(y~trat,dados)
```

```
saida<-aov(y~factor(trat),dados)
```

```
is.data.frame(dados)
```

```
dados<-data.frame(dados)  
#ou dados<-as.data.frame(dados)
```

```
saida<-aov(y~factor(trat),dados)  
saida
```

```
summary (saida)  
anova(saida)  
#Não significativo ... para por  
aqui. Voltar para os slides  
#model.tables(saida,"means")
```

# PROCEDIMENTOS PARA COMPARAÇÃO DE MÉDIAS

- Ou testes para comparação de médias, ou testes de comparações múltiplas
  - DMS, Bonferroni, Scheffé, Tukey, Duncan, Dunnett
- Complemento ao teste F
- detectar diferenças entre tratamentos
- Situações ou uso específico para cada teste



# Contrastes

- Funções do tipo

$$Y = f(x) = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

- Onde

$$\sum_{i=1}^n a_i = 0$$

- Exemplo:  $Y = 2x_1 - x_2 - x_3$

$$\sum_{i=1}^4 a_i = 0, \quad a_1 = 2, \quad a_2 = -1, \quad a_3 = -1$$

- Se  $X \equiv \text{média} \rightarrow \text{contraste de média}$

- para um contraste de médias em sua forma geral

$$Y = c_1 m_1 + c_2 m_2 + c_3 m_3 + \dots + c_n m_n \qquad \sum_{i=1}^n c_i = 0$$

- Podemos obter a estimativa

$$\hat{Y} = c_1 \hat{m}_1 + c_2 \hat{m}_2 + \dots + c_n \hat{m}_n$$

- Para aplicar testes precisamos conhecer o estimador da variância do estimador do contraste

$$\hat{Y} = c_1 \hat{m}_1 + c_2 \hat{m}_2 + \dots + c_n \hat{m}_n$$

$$\hat{V}(\hat{Y}) = c_1^2 \hat{V}(\hat{m}_1) + c_2^2 \hat{V}(\hat{m}_2) + \dots + c_n^2 \hat{V}(\hat{m}_n)$$

Se “médias” independentes.

Se variâncias “iguais”, e mesmo nº repetições:

$$\hat{V}(\hat{Y}) = (c_1^2 + c_2^2 + c_3^2 + \dots + c_n^2) \frac{s^2}{r} = \sum_{i=1}^n c_i^2 \frac{s^2}{r}$$

# Teste da Diferença Mínima Significativa (DMS)

- A DMS ao nível  $\alpha$  de significância para comparar  $m_i$  com  $m_j$  é:

- $$\text{DMS} = t_{\alpha/2, v} \sqrt{QM \text{ Res} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

- Concluir que  $m_i \neq m_j$  se  $|\hat{m}_i - \hat{m}_j| > \text{DMS}$

# Generalizando

- ao nível  $\alpha$  de probabilidade, concluir que

$$\sum c_i m_i \neq 0 \text{ se } \left| \sum c_i \hat{m}_i \right| > t_{\alpha/2, v} \sqrt{QM Res \sum_{i=1}^I \frac{c_i^2}{n_i}}$$

Número de g.l. resíduo

Número de trat

Número de repetições do trat i

# Teste de Bonferroni

- bom quando se quer fazer um pequeno número de comparações
- mantém o nível de significância conjunto igual ou menor que  $\alpha$ , quando aplicado a um grupo de contrastes de interesse
- Se  $p$  contrastes estão sendo comparados:
  - para cada contraste  $\rightarrow$  usar  $\alpha' = \alpha/p$ .

- se quisermos fazer  $p$  comparações, concluiremos que a  $q$ -ésima comparação

$$\sum c_{iq} m_i \neq 0$$

- se

$$\left| \sum c_{iq} \hat{m}_i \right| > t_{(\alpha/2p),v} \sqrt{QM \text{ Res} \sum_{i=1}^I \frac{c_{iq}^2}{n_i}}$$

# Teste de Scheffé

- Pode testar todo e qualquer contraste de médias, mesmo quando sugerido pelos dados
- muito utilizado para testar contrastes que envolvem grupos de médias
- Não exige a ortogonalidade e nem que os contrastes sejam estabelecidos a “priori”
- exige apenas que, na ANOVA, o teste F para tratamentos seja significativo



## O teste

- Estatística do teste

$$S = \sqrt{(I - 1) F \hat{V}(\hat{Y})}$$

- Regra decisória:

Se  $|\hat{Y}| \geq S$  contraste é significativo ao nível  $\alpha$

# Teste de Tukey

- Testa todo e qualquer contraste entre 2 médias
- Não permite comparar grupos de médias
- Estatística do teste:

$$\Delta = q \sqrt{\frac{1}{2} \hat{V}(\hat{Y})}$$




[Tabela de Tukey](#)

- Se  $|\hat{Y}| \geq \Delta$  contraste é significativo ao nível  $\alpha$

# Teste de Duncan

- fornece resultados mais discriminados que os do teste de Tukey
- de aplicação mais trabalhosa
- exige que as médias sejam colocadas em ordem decrescente
- Teste é exato quando todos os tratamentos possuírem o mesmo número de repetições

# O teste

- Estatística do teste 
$$D_i = Z_i \sqrt{\frac{1}{2} \hat{V}(\hat{Y})}$$
  
[Tabela de Duncan](#)
- Se  $\hat{Y} < D_i \rightarrow$  contraste não significativo
- Se  $\hat{Y} \geq D_i \rightarrow$  médias diferem  $\rightarrow$  repetir procedimento para abrangência  $i - 1$

# Teste de Dunnett

- Comparar um tratamento padrão com os demais tratamentos
- Estatística do teste:

$$d' = t_d \sqrt{V(\hat{Y})}$$



Tabela de Dunnett  
f(g.l. trat.; g.l. res.)

- Se  $|\hat{Y}| \geq d' \rightarrow$  contraste é significativo

# Exemplos

- Exemplos diversos usando
  - Função `aov()` e `anova()`
  - Pacote “agricolae”

# Comandos R

```
#Alterando os dados agora
dados2<-edit(dados) #acrescentando 6
    a cada valor do trat 4

saida2<-aov(y~factor(trat),dados2)

summary (saida2)
anova(saida2)

res<-resid(saida2)

sum(res)
sum(res*res)

#PARA REALIZAR AS ANÁLISES A MÃO
#Os totais de tratamentos podem ser
    obtidos por
tapply(dados2$y,dados2$trat,sum)
#ou
rowsum(dados2$y,dados2$trat)
#ou
rowsum(dados2,dados2$trat)
```

```
#As médias de tratamentos podem ser
    obtidas por
tapply(dados2$y,dados2$trat,mean)
#ou dividindo os totais pelo número de
    repetições
rowsum(dados2$y,dados2$trat)/5

#comparando as médias
TukeyHSD(saida2,"factor(trat)",ordered=T)
plot(TukeyHSD(saida2,"factor(trat)",order
    ed=T))

library(agricolae)
args(HSD.test)
HSD.test(dados2$y,dados2$trat,16,8.25)
args(duncan.test)
duncan.test(dados2$y,dados2$trat,16,8.25)
args(bar.err)
comparacao<-
    HSD.test(dados2$y,dados2$trat,16,8.25)
bar.err(comparacao)
bar.group(comparacao)
```

# EXPERIMENTOS EM BLOCOS CASUALIZADOS

## DBC



## modelo estatístico

$$Y_{ij} = m + t_i + b_j + e_{ij}$$

- A variação total é dividida em:  
Tratamentos + Blocos + Resíduo
- Obtem-se:
  - G.I. de cada F.V.
  - Somas de Quadrados
  - Quadrados médios
  - Valor F para a F.V. tratamentos

# DBC – Modelo fixo

- Hipóteses testadas

$$H_0 : m_1 = m_2 = \dots = m_I = m$$

$$H_a : \text{não } H_0$$

- Regra decisória

- se  $F_{\text{calculado}} \geq F_{\text{tabelado}} \rightarrow \text{rejeita-se } H_0$
- Ou avalia o “p-valor”

# Comandos R

```
#SE USAR O R DE DENTRO DA PASTA DE TRABALHO DO "PROJETO"
dir()
dados<-read.table("curso R slides/dados curso R/DBC 1.txt",h=T)
dados

saida<-aov(y~trat+rep,data=dados)
anova(saida)
saida<-aov(y~factor(trat)+factor(rep),data=dados)
anova(saida)

library(agricolae)
args(duncan.test)
#function (y, trt, DError, MSError, alpha = 0.05, group = TRUE,
#      main = NULL)
duncan.test(dados$y,dados$trat,9,0.00889)

HSD.test(dados$y,dados$trat,saida$df.residual,anova(saida)[3,3])

#fazendo automaticamente
glres<-saida$df.residual
qmres<-deviance(saida)/glres
duncan.test(dados$y,dados$trat,glres,qmres)
```

# Esquemas Experimentais

- Experimentos Fatoriais

# Conceituação

- Dois ou mais fatores são estudados simultaneamente
- Cada Fator apresenta dois ou mais níveis
- É um tipo de esquema e não um delineamento
  - Ex.: podemos ter experimento fatorial no DIC ou no DBC
- Os tratamentos são obtidos pelas combinações dos níveis dos fatores

# Aplicações

- Sejam dois fatores
  - fator A com I níveis e o fator B com J níveis, com k repetições
- instalado segundo o DIC

$$Y_{ijk} = m + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

- Instalado segundo o DBC

$$Y_{ijk} = m + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \omega_k + e_{ijk}$$

# Quadros da ANOVA

- Fatorial no DIC

FV	G.I.	SQ	QM	F
A	$I - 1$			
B	$J - 1$			
AxB	$(I - 1)(J - 1)$			
(Trat)	$IJ - 1$			
Resíduo	Por dif.			
Total	$IJK - 1$			

# Quadros da ANOVA

- Fatorial no DBC

FV	G.I.	SQ	QM	F
Bloco	$K - 1$			
A	$I - 1$			
B	$J - 1$			
AxB	$(I - 1)(J - 1)$			
(Trat)	$IJ - 1$			
Resíduo	$(IJ - 1)(K - 1)$			
Total	$IJK - 1$			



# Exemplo – Exp. Fatorial

```
dados<-read.table("curso R
  slides/dados curso
  R/fatorial.txt",h=T)

Af<-factor(dados$A)
Bf<-factor(dados$B)
#agora solicitando a ANOVA
saida<-aov(y~Af+Bf+Af:Bf,data=dados)
#ou simplesmente: saida<-
  aov(y~Af*Bf,data=dados)

#pedindo o quadro informativo da
  ANOVA
summary(saida)

interaction.plot(Bf,Af,dados$y)
#interaction.plot(Af,Bf,dados$y)

#dados2<-edit(dados)
```

```
#no DBC
blocof<-factor(dados$rep)
#agora solicitando a ANOVA
saida<-
  aov(y~blocof+Af+Bf+Af:Bf,data=dados
  )
#ou simplesmente: saida<-
  aov(y~blocof+Af*Bf,data=dados)
#quadro informativo da ANOVA
summary(saida); anova(saida)
interaction.plot(Bf,Af,dados$y)

model.tables(saida) # efeitos
model.tables(saida,"means")
dados.A1<-dados[which(dados$A==1),]
dados.A2<-dados[which(dados$A==2),]

#totais.A<-
  rowsum(dados.A1$y,dados.A1$B)
HSD.test(dados.A1$y,dados.A1$B,7,1.92)
HSD.test(dados.A2$y,dados.A2$B,7,1.92)
```

# Outros experimentos

- Parcela subdividida
  - O termo “parcelas subdivididas”, assim como o “fatorial”, refere-se à maneira como os tratamentos são organizados e atribuídos às unidades experimentais.
  - Definimos, aqui, os fatores primários e secundários, ou seja, parcelas e subparcelas respectivamente.

# Exemplo

---

	A1						A2					
Bloco	B1		B2		B3		B1		B2		B3	
1	12	12	15	14	15	16	21	19	22	20	16	19
2	15	16	16	17	12	12	18	19	19	21	21	20
3	17	16	13	15	12	11	17	19	20	18	19	21
4	14	13	16	15	14	17	16	17	17	20	18	18

---

- Entrar com os comandos

# Comandos R

```
> #cria o fator das parcelas
> A<-gl(2,24,label=paste("A",1:2,sep=""))
> #fator das subparcelas
> B<-
  rep(gl(3,8,label=paste("B",1:3,sep="")
    ),2)
> #blocos
> bl<-
  rep(gl(4,2,label=paste("bl",1:4,sep="")
    ),6)
> #variável resposta
> dados<-c(12,12,15,16,17,16,14,13,
  15,14,16,17,13,15,16,15,
  15,16,12,12,12,11,14,17,
  21,19,18,19,17,19,16,17,
  22,20,19,21,20,18,17,20,
  16,19,21,20,19,21,18,18)
> #juntando tudo num data.frame
> tabela<-
  data.frame(A=A,B=B,bloco=bl,dados=dados)
> #fazendo a anova
```

```
> #fazendo a anova
> saida<-aov(
dados~bloco+A+B+A*B>Error(bloco/A),table)
> summary(saida)
> model.tables(saida,"means")
```

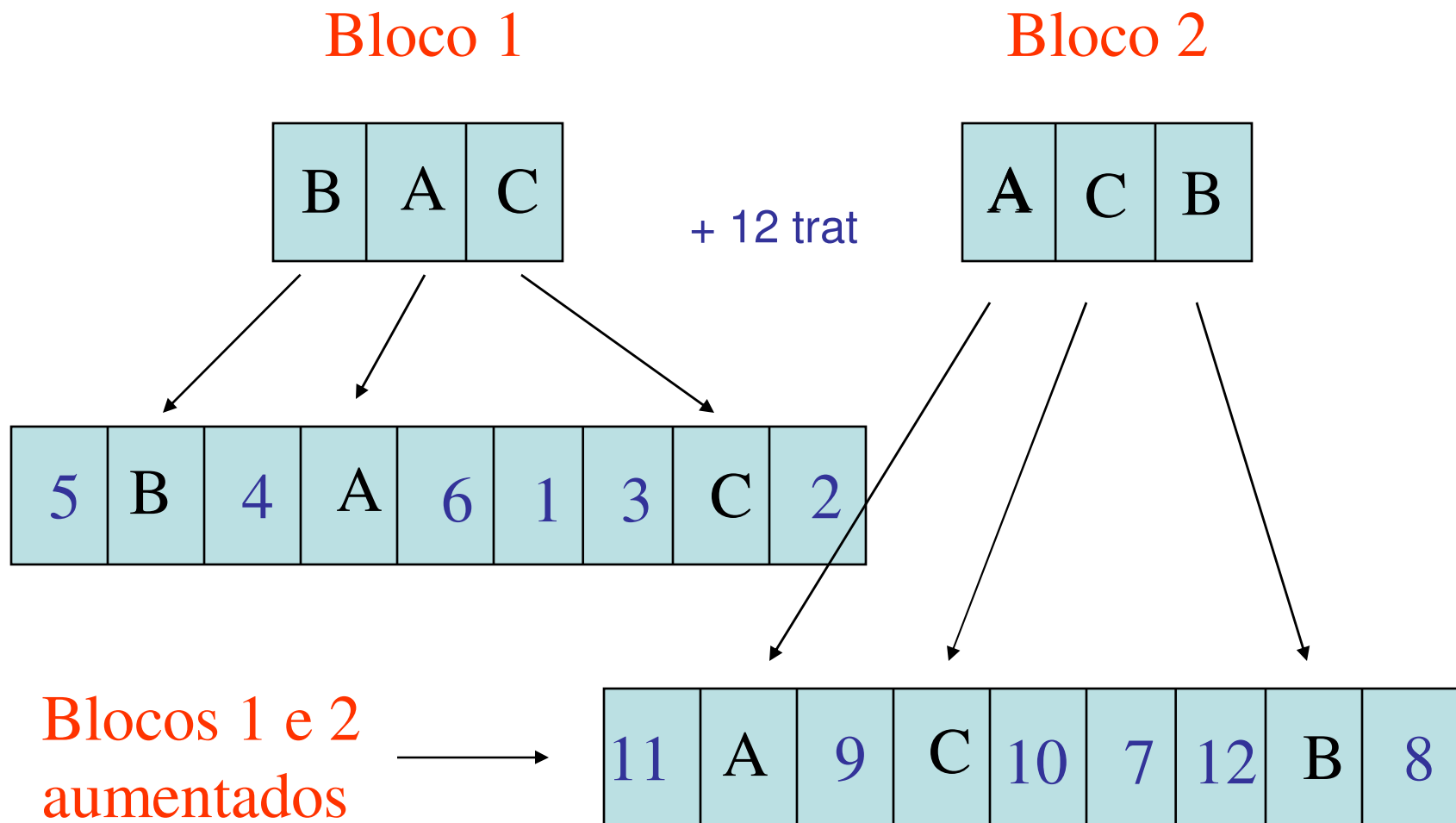
FV	GL
Bloco	3
A	1
Resíduo(a)	3
B	2
A x B	2
Resíduo	36

[Usando o R](#)

# Blocos incompletos

- Blocos Aumentados de Federer
- Blocos Incompletos Balanceados
- Blocos Incompletos não Balanceados
- Veja possibilidades no pacote agricolae:  
`help(package="agricolae")`

# Esquema ilustrativo - Blocos Aumentados



# Blocos Aumentados

```
#Criando o delineamento (design.dau())
library(agricolae)
# 3 tratamentos and 2 blocos
T1<-c("A","B","C") ; T2<-letters[1:12]
dau <-design.dau(T1,T2, r=2)
# livro de campo: dau by(dau,dau[2],function(x)
  paste(x[,1],"-",as.character(x[,3])))
# escrevendo no HD:
# write.table(dau,"dau.txt", row.names=FALSE, sep="\t")
# file.show("dau.txt")
# Delineamento aumentado no DIC:
trt<-c(T1,T2); r<-c(4,4,4,1,1,1,1,1,1,1,1,1,1,1)
livro <- design.crd(trt,r)
```

# Gerando e analisando dados

```
> #Exemplo DBA
> #Gerar y = rnorm(18,30,2)
> args(DAU.test):
  function (block, trt, y, method = c("lsd", "tukey"))
> cbind(dau,y)
> DAU.test(dau$block,dau$trt,y, "lsd")
> saida<- DAU.test(dau$block,dau$trt,y, "lsd")
> saida
```



# Comparando Pacotes

- Comparando o agricolae com o ExpDes

# Modelos mistos

- Para a análise de modelos mistos, o R apresenta uma biblioteca muito versátil e extremamente poderosa chamada *nlme* – *nonlinear mixed effects model*, que permite a avaliação de modelos mistos lineares e não-lineares.
- Para acessá-la basta entrar com o seguinte comando:
  - `> library(nlme)`
- Uma vez carregada essa biblioteca, digite `help(package=nlme)` para maiores informações e exemplo de aplicação.
- Veja, por exemplo:
  - `> ?lme`