

Pressuposições da ANOVA

Renata Alcarde Sermarini

22 de setembro de 2016

Exemplo

Um pesquisador pretende comparar quatro variedades de pêsego quanto ao enraizamento de estacas. Para tanto, realizou um experimento de acordo com o delineamento inteiramente casualizado com cinco repetições, sendo cada parcela um vaso com vinte estacas. Passado o tempo necessário, o pesquisador anotou o número de estacas enraizadas, apresentado na Tabela a seguir.

Variedades	Repetições					Total
	1	2	3	4	5	
A	2	2	1	1	0	6
B	1	0	0	1	1	3
C	12	10	14	17	11	64
D	7	9	15	8	10	49

Exemplo

```
y<- c( 2,  2,  1,  1,  0,  
      1,  0,  0,  1,  1,  
      12, 10, 14, 17, 11,  
      7,  9, 15,  8, 10)  
trat<- rep(c("A","B","C","D"), each=5)  
dados<- data.frame(trat, y)  
head(dados)
```

```
##   trat y  
## 1    A 2  
## 2    A 2  
## 3    A 1  
## 4    A 1  
## 5    A 0  
## 6    B 1
```

Introdução

Modelo estatístico,

$$y_{ij} = \mu + \tau_i + e_{ij},$$

em que y_{ij} é o valor observado na j -ésima repetição do i -ésimo tratamento, μ é uma constante, comum a todas as observações, τ_i é o efeito do i -ésimo tratamento e e_{ij} é o efeito do acaso.

Quais são as pressuposições para a realização da Análise de Variância?

- ▶ Os erros devem seguir uma distribuição normal;
- ▶ Os erros devem ser independentes;
- ▶ Os erros devem apresentar variância constante, ou seja, homogeneidade de variâncias;
- ▶ O modelo deve ser aditivo.

Análise de Resíduos

Valores preditos para os erros:

$$y_{ij} = \hat{\mu} + \hat{\tau}_i + \tilde{e}_{ij} = \hat{\mu}_i + \tilde{e}_{ij},$$

em que μ_i representa a média do i -ésimo tratamento.

Logo,

$$\tilde{e}_{ij} = y_{ij} - \hat{\mu}_i.$$

Análise de Resíduos

Resíduos Padronizados \Rightarrow Observações discrepantes e aparente homogeneidade/heterogeneidade de variâncias

$$z_{ij} = \frac{e_{ij}}{\sqrt{QMRes}} \quad \text{ou} \quad d_{ij} = \frac{e_{ij}}{\sqrt{(1 - 1/J)QMRes}}.$$

Exemplo

- ▶ Análise de variância - ajuste do modelo

```
modelo=lm(y ~ trat, dados)
```

- ▶ Obtenção dos resíduos

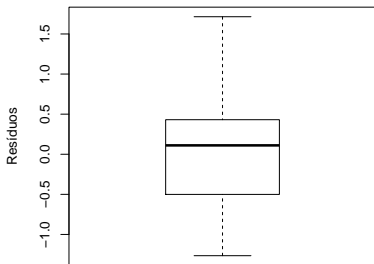
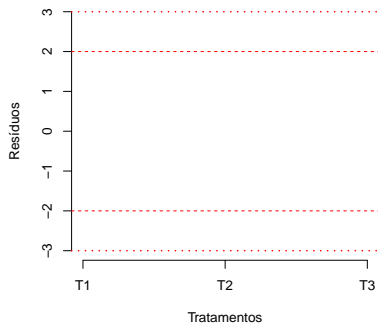
```
res <- residuals(modelo) # resíduos simples  
res_Stud <- rstandard(modelo) # resíduos Studentizados  
round(head(data.frame(res, res_Stud)), 5)
```

```
##      res res_Stud  
## 1  0.8  0.41703  
## 2  0.8  0.41703  
## 3 -0.2 -0.10426  
## 4 -0.2 -0.10426  
## 5 -1.2 -0.62554  
## 6  0.4  0.20851
```

Observações discrepantes

Avaliação gráfica

- ▶ Resíduos estudentizados (ou padronizados) *versus* Tratamentos
- ▶ Gráfico de caixas dos resíduos estudentizados (ou padronizados)



Observações discrepantes

Avaliação gráfica

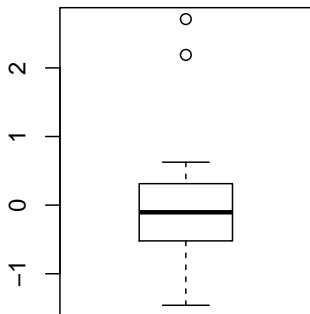
Observações:

- ▶ 95% dos resíduos devem pertencer ao intervalo $(-2,2)$;
- ▶ Valores não pertencentes ao intervalo $(-3,3)$ podem ser classificados como discrepantes;
- ▶ Não podemos excluir observações discrepantes antes de conversar com o pesquisador, pois pode ter ocorrido algum erro de tabulação ou a observação pode indicar uma característica importante associada ao respectivo tratamento.

Observações discrepantes: Exemplo

Avaliação gráfica

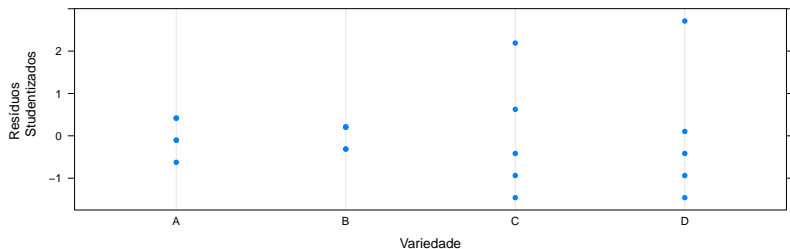
```
boxplot(res_Stud)
```



Observações discrepantes: Exemplo

Avaliação gráfica

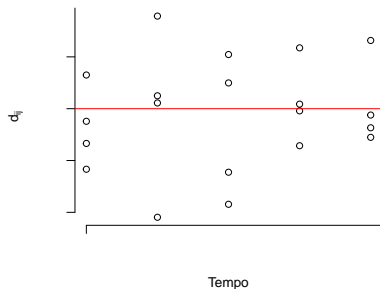
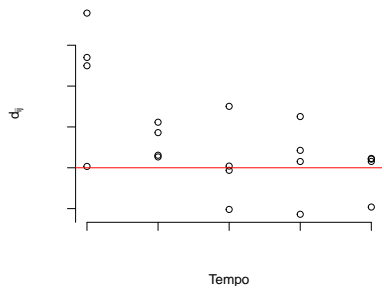
```
require(lattice)  
dotplot(res_Stud~trat, dados, ylab="Resíduos  
Studentizados", xlab="Variedade")
```



Independência dos Resíduos

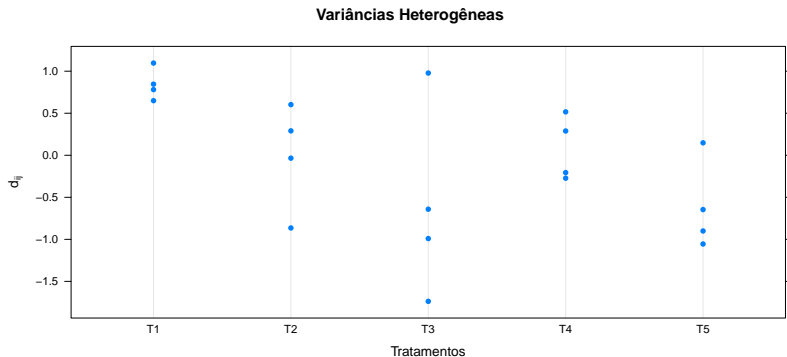
Até certo ponto é garantida pelo princípio da casualização, entretanto, devemos “verificar” a independência dos resíduos nos casos em que poderiam existir correlações entre as observações;

- ▶ observações no mesmo indivíduos/parcela ao longo do tempo;
- ▶ observações de indivíduos agrupados, como por exemplo cobaias em uma mesma gaiola.



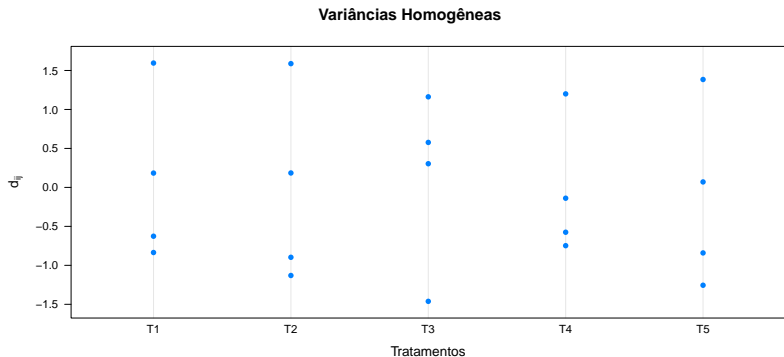
Homogeneidade de Variâncias

- ▶ A pressuposição de homogeneidade de variâncias é a mais importante a ser atendida.
- ▶ Verificação gráfica: Resíduos padronizados *versus* Tratamentos



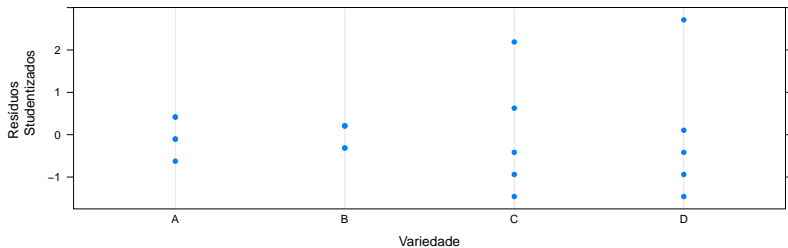
Homogeneidade de Variâncias

- ▶ A pressuposição de homogeneidade de variâncias é a mais importante a ser atendida.
- ▶ Verificação gráfica: Resíduos padronizados *versus* Tratamentos



Exemplo

```
dotplot(res_Stud~trat, dados, ylab="Resíduos  
Studentizados", xlab="Variedade")
```



Homogeneidade de Variâncias

Teste de hipóteses: Teste de Hartley

H_0 : Há homogeneidade de variâncias

H_a : Não há homogeneidade de variâncias

$$F_{max} = \frac{S_{max}^2}{S_{min}^2}.$$

Rejeita-se H_0 se $F_{max} \geq F_{max_{tab}}(\alpha, \nu_1, \nu_2)$

Homogeneidade de Variâncias

Teste de hipóteses: Teste de Breusch-Pagan (1979)

H_0 : Há homogeneidade de variâncias

H_a : Não há homogeneidade de variâncias

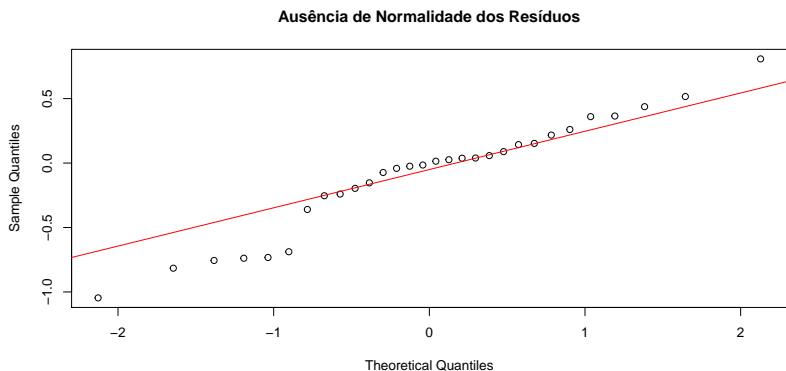
```
require(lmtest)
bptest(modelo)
```

```
##
## studentized Breusch-Pagan test
##
## data:  modelo
## BP = 4.8142, df = 3, p-value = 0.1859
```

Considerando-se o nível de 5% de significância não rejeitamos H_0 . Logo, não há evidências para afirmarmos que as variâncias são heterogêneas.

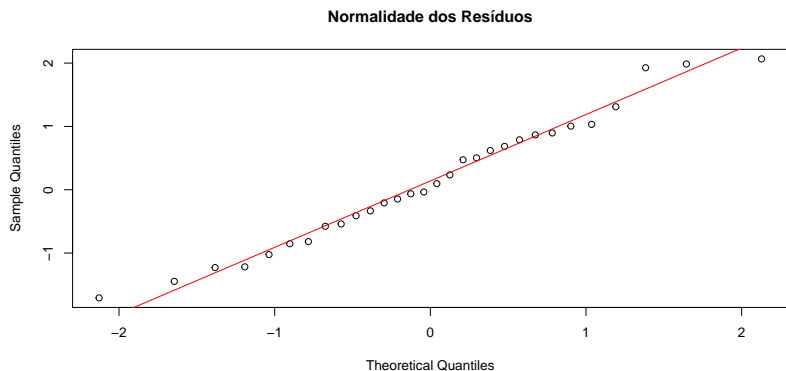
Normalidade dos Resíduos

Gráfico quantil-quantil



Normalidade dos Resíduos

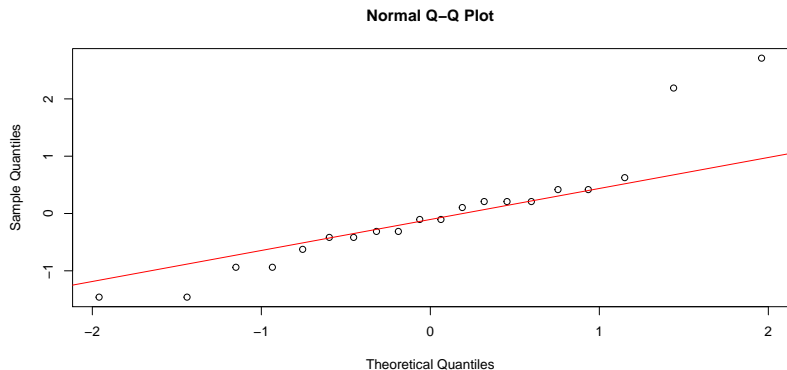
Gráfico quantil-quantil



Normalidade dos Resíduos: Exemplo

Gráfico quantil-quantil

```
qqnorm(res_Stud)  
qqline(res_Stud, col=2)
```



Normalidade dos Resíduos

Teste de hipóteses

H_0 : Os resíduos seguem uma distribuição normal

H_a : Os resíduos não seguem uma distribuição normal

- ▶ Shapiro-Wilks
- ▶ Kolmogorov-Smirnov
- ▶ entre outros.

Normalidade dos Resíduos

Teste de hipóteses: Shapiro-Wilks

H_0 : Os resíduos seguem uma distribuição normal

H_a : Os resíduos não seguem uma distribuição normal

```
shapiro.test(res_Stud)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  res_Stud  
## W = 0.8853, p-value = 0.02209
```

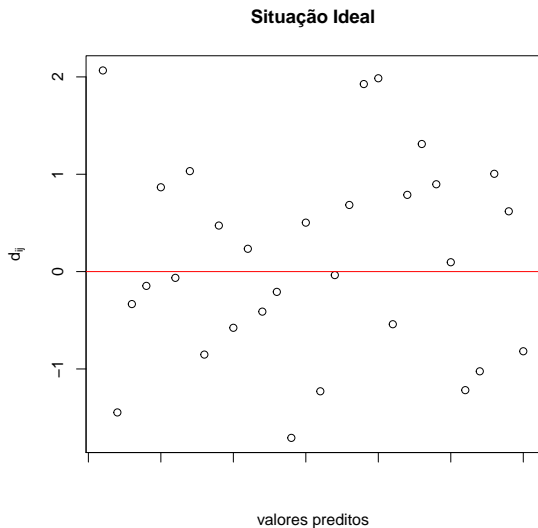
Considerando-se o nível de 5% de significância rejeitamos H_0 . Logo, há evidências para afirmarmos que os erros não seguem uma distribuição normal.

O que fazer quando alguma pressuposição não é atendida?

- ▶ Transformação de dados;
- ▶ Testes não paramétricos;
- ▶ Modelos lineares generalizados;
- ▶ Modelos mistos;
- ▶ Modelos lineares generalizados mistos. . .

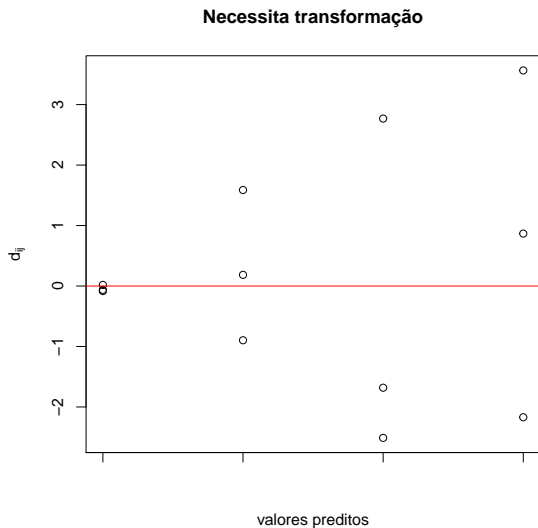
Necessidade de transformação de dados

Gráfico dos Resíduos *versus* Valores Preditos



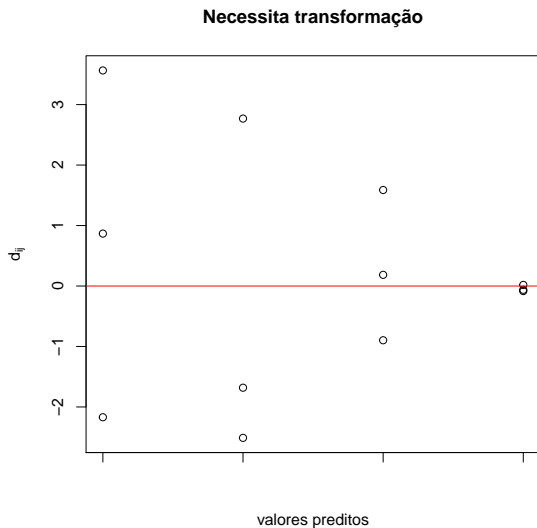
Necessidade de transformação de dados

Gráfico dos Resíduos *versus* Valores Preditos



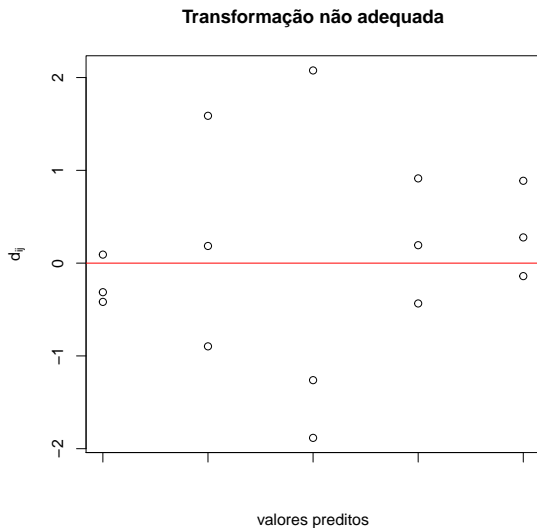
Necessidade de transformação de dados

Gráfico dos Resíduos *versus* Valores Preditos



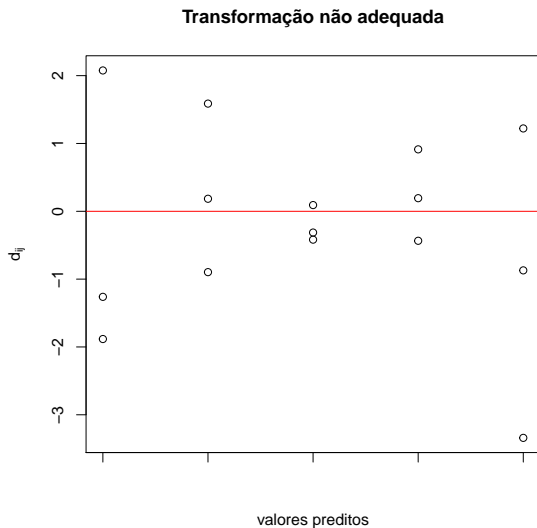
Necessidade de transformação de dados

Gráfico dos Resíduos *versus* Valores Preditos



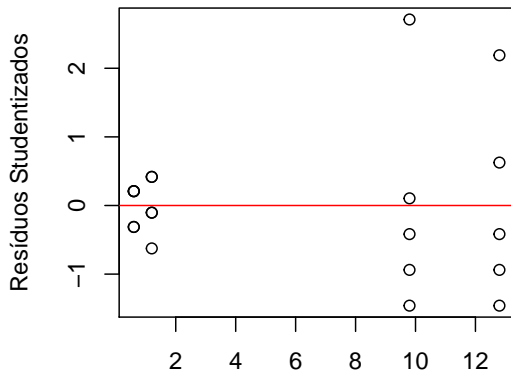
Necessidade de transformação de dados

Gráfico dos Resíduos *versus* Valores Preditos



Necessidade de transformação de dados: Exemplo

```
plot(res_Stud~fitted(modelo),ylab="Resíduos Studentizados",  
      xlab="Valores esperados (médias)")  
abline(h=0, col=2)
```



Necessidade de transformação de dados

Transformação Box-Cox (1964)

$$y^* = \begin{cases} \log y & \text{se } \lambda = 0 \\ y^\lambda & \text{se } \lambda \neq 0 \end{cases}$$

λ	Transformação
-1	$1/(y + 0,5)$
-0,5	$1/\sqrt{y + 0,5}$
0	$\log(y + 0,5)$
0,5	$\sqrt{y + 0,5}$
1	Nenhuma

Necessidade de transformação de dados

Exemplo fictício

▶ Dados originais

Trat	Repetição			\bar{y}
	1	2	3	
A	1	100	10	37
B	10	0,1	1	3,7
C	10	10	100	40

▶ Dados transformados ($\log_{10}(y)$)

Trat	Repetição			\bar{y}^*
	1	2	3	
A	0	2	1	1
B	1	-1	0	0
C	1	1	2	$4/3=1,33$

Necessidade de transformação de dados

Após realizar a transformação dos dados deve-se verificar novamente as pressuposições da análise de variância. Caso sejam atendidas, pode-se prosseguir com a análise, lembrando de realizar a transformação inversa ao final.

Necessidade de transformação de dados

- ▶ Para o caso de Tratamentos ser um fator **Qualitativo**

Exemplo fictício

Trat	Repetição			\bar{y}^*	letras
	1	2	3		
A	0	2	1	1	a
B	1	-1	0	0	b
C	1	1	2	1,33	a

- ▶ Apresentação dos dados

Tratamento	y_{ti}	letras
A	$10^1 = 10$	a
B	$10^0 = 1$	b
C	$10^{4/3} = 21,54$	a

$$\log(y)_{10} = 1 \Rightarrow y = 10^1 = 10$$

Necessidade de transformação de dados

- ▶ Para o caso de Tratamentos ser um fator **Quantitativo**

Ajustado o modelo polinomial

$$y^* = \hat{\beta}_0 + \hat{\beta}_1 x + \dots + \hat{\beta}_p x^p,$$

deve-se realizar a transformação inversa.

Por exemplo, se a transformação utilizada foi $\log_{10}(y)$, então,

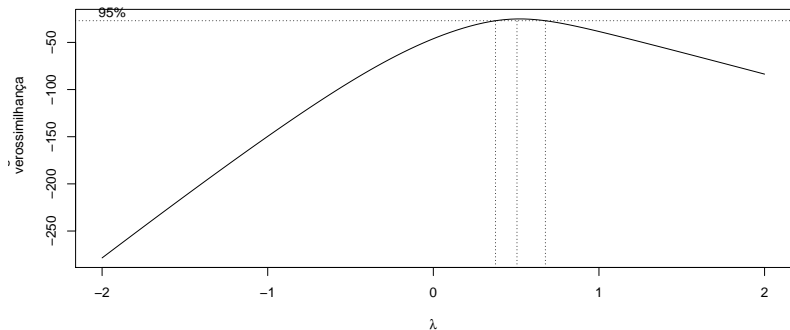
$$y = 10^{\hat{\beta}_0 + \hat{\beta}_1 x + \dots + \hat{\beta}_p x^p}.$$

Se a transformação utilizada foi y^λ , então,

$$y = (\hat{\beta}_0 + \hat{\beta}_1 x + \dots + \hat{\beta}_p x^p)^{1/\lambda}.$$

Necessidade de transformação de dados: Exemplo

```
library(MASS)
boxcox(dados$y+0.001 ~ dados$trata,ylab="logaritmo da
verossimilhança") #lambda=0,5.
```



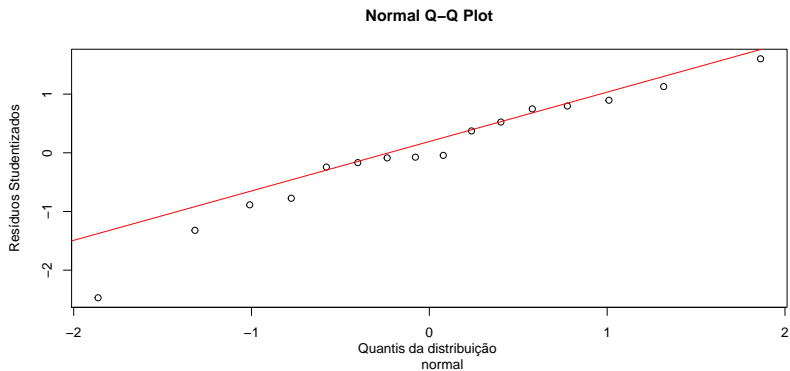
Análise dos Dados Transformados

```
dados$yt<- (y+0.01)^0.5  
modelot<- lm(yt ~ trat, dados)
```

Análise dos Dados Transformados

Normalidade dos Resíduos

```
qqnorm(rstandard(modelot), xlab="Quantis da distribuição  
normal", ylab="Resíduos Studentizados")  
qqline(rstandard(modelot), col=2)
```



Análise dos Dados Transformados

Normalidade dos Resíduos

- ▶ Teste de hipóteses

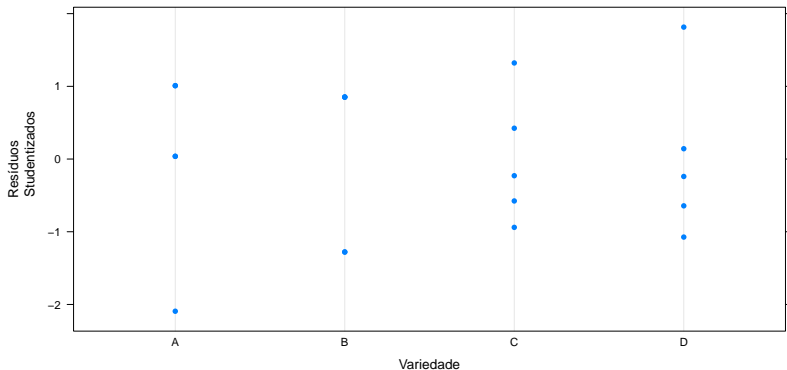
```
shapiro.test(rstandard(modelot))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  rstandard(modelot)  
## W = 0.9558, p-value = 0.5865
```

Análise dos Dados Transformados

Homogeneidade de Variâncias

```
dotplot(rstandard(modelot)~trat,ylab="Resíduos  
Studentizados",xlab="Variedade")
```



Análise dos Dados Transformados

Homogeneidade de Variâncias

- ▶ Teste de hipóteses

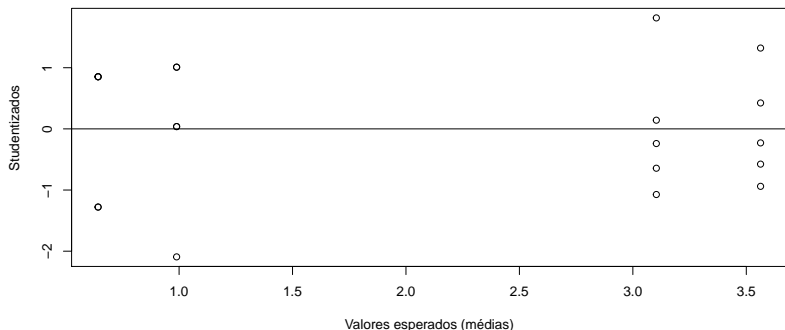
```
bptest(modelot)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: modelot  
## BP = 0.895, df = 3, p-value = 0.8266
```


Análise dos Dados Transformados

Necessidade de Transformação de Dados

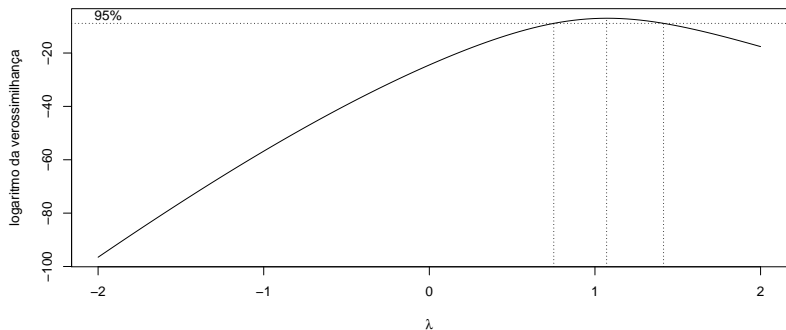
```
plot(rstandard(modelot)~fitted(modelot),ylab="Resíduos  
Studentizados",xlab="Valores esperados (médias)")  
abline(h=0)
```



Análise dos Dados Transformados

Necessidade de Transformação de Dados

```
boxcox(modelot,ylab="logaritmo da verossimilhança")
```



Análise de Variância

Atendidas as pressuposições para a realização da análise de variância considerando-se os dados transformados, temos:

```
anova(modelot)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: yt
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## trat         3  32.509  10.8363   48.053 3.162e-08 ***
```

```
## Residuals  16   3.608   0.2255
```


```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

Análise de Variância e Teste de Tukey

```
require(ExpDes.pt)
dic(dados$strat, dados$yt)
```

```
## -----
## Quadro da analise de variancia
## -----
##           GL      SQ      QM      Fc      Pr>Fc
## Tratamento  3 32.509 10.8363 48.053 3.1617e-08
## Residuo      16  3.608  0.2255
## Total        19 36.117
## -----
## CV = 22.89 %
##
## -----
## Teste de normalidade dos residuos (Shapiro-Wilk)
## p-valor:  0.848062
## De acordo com o teste de Shapiro-Wilk a 5% de significancia
## -----
```



Análise de Variância e Teste de Tukey

Teste de Tukey

Grupos	Tratamentos	Medias
a	C	3.562969
a	D	3.103527
b	A	0.9890929
b	B	0.6429925

Análise de Variância e Teste de Tukey

Apresentação das médias

```
round((tapply(dados$yt, dados$trat, mean))^2-0.01,4)
```

```
##          A          B          C          D
## 0.9683 0.4034 12.6848 9.6219
```