

# Análise de Dados Experimentais e Verificação dos Pressupostos da ANOVA – Uma Abordagem Prática Utilizando o *Software R*



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Clima Temperado  
Ministério da Agricultura, Pecuária e Abastecimento*

**DOCUMENTOS 504**

**Análise de Dados Experimentais e Verificação dos Pressupostos  
da ANOVA: Abordagem Prática com o *Software R***

*Ricardo Alexandre Valgas  
Rosane Martinazzo  
Beatriz Marti Emygdio  
Rosa Lia Barbieri*

**Embrapa Clima Temperado**  
BR 392 km 78 - Caixa Postal 403  
CEP 96010-971, Pelotas, RS  
Fone: (53) 3275-8100  
[www.embrapa.br/clima-temperado](http://www.embrapa.br/clima-temperado)  
[www.embrapa.br/fale-conosco](http://www.embrapa.br/fale-conosco)

Comitê Local de Publicações

Presidente  
*Luis Antônio Suita de Castro*

Vice-Presidente  
*Walkyria Bueno Scivittaro*

Secretária-Executiva  
*Bárbara Chevallier Cosenza*

Membros  
*Ana Luiza B. Viegas, Fernando Jackson, Marilaine  
Schaun Pelufê, Sônia Desimon*

Revisão de texto  
*Bárbara Chevallier Cosenza*

Normalização bibliográfica  
*Marilaine Schaun Pelufê*

Editoração eletrônica  
*Fernando Jackson*

Foto de capa  
*Ricardo Alexandre Valgas*

**1ª edição**  
Obra digitalizada (2021)

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte,  
constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**  
Embrapa Clima Temperado

---

A532 Análise de dados experimentais e verificação dos  
pressupostos da ANOVA: abordagem prática com o  
*software R* / Ricardo Alexandre Valgas... [et al.]. –  
Pelotas: Embrapa Clima Temperado, 2019.  
35 p. (Documentos / Embrapa Clima Temperado,  
ISSN 1678-2518 ; 504)

1. Estatística. 2. Estatística agrícola. 3. Análise  
estatística. 4. Método estatístico. I. Valgas, Ricardo  
Alexandre. II. Série.

CDD 519.5

## Autores

### **Ricardo Alexandre Valgas**

Estatístico, mestre em Métodos Numéricos em Engenharia, pesquisador da Embrapa Clima Temperado, Pelotas, RS.

### **Rosane Martinazzo**

Engenheira-agrônoma, doutora em Ciência do Solo, pesquisadora da Embrapa Clima Temperado, Pelotas, RS.

### **Beatriz Marti Emygdio**

Bióloga, doutora em Ciência e Tecnologia de Sementes, pesquisadora da Embrapa Clima Temperado, Pelotas, RS.

### **Rosa Lia Barbieri**

Bióloga, doutora em Genética e Biologia Molecular, pesquisadora da Embrapa Clima Temperado, Pelotas, RS.



## Apresentação

A utilização de técnicas de estatística experimental vem sendo amplamente utilizada nos diversos campos da ciência. Há alguns anos, o conceito de experimentação era algo relativamente novo, porém passou a ganhar espaço cada vez maior nos setores produtivos, devido aos bons resultados da experimentação obtidos na agricultura, área em que seus princípios e prática já estavam mais difundidos entre os pesquisadores e agrônomos.

Com o avanço da informática, a realização de experimentos foi se tornando uma ferramenta cada vez mais comum para comprovar as hipóteses formuladas nos numerosos projetos de pesquisas agrícolas. Nessa linha, o programa estatístico **R** foi ocupando lugar de destaque pelo seu uso crescente no meio agrônômico e, atualmente, é um dos programas de análise de dados mais difundido no meio científico, fator levado em conta para compor sua escolha pelos autores.

Buscando qualificar a produção técnica de pesquisadores e estudantes, apresenta-se este trabalho, o qual contribui para o aprendizado e a aplicação da estatística experimental de forma simples e prática, proporcionando ao leitor realizar a verificação dos principais pressupostos da análise de dados experimentais utilizando o programa estatístico R.

Roberto Pedroso de Oliveira  
Chefe-Geral  
Embrapa Clima Temperado



## Sumário

Estatística experimental .....	9
Noções básicas da experimentação .....	9
Princípios básicos em experimentação .....	10
Os pressupostos para a ANOVA.....	10
Aditividade dos fatores do modelo .....	10
Normalidade dos resíduos .....	11
Homogeneidade de variâncias .....	11
Transformação Box-Cox.....	11
Delineamento completamente casualizado .....	11
Análise descritiva dos dados .....	13
Verificação dos pressupostos da análise experimental (DIC).....	14
Homogeneidade de variâncias .....	14
Normalidade dos resíduos .....	14
Transformação Box-Cox.....	16
Gráficos auxiliares .....	17
Análise de Variância - ANOVA.....	19
Teste de comparação de médias .....	19
Delineamento em blocos completos .....	20
Verificação dos pressupostos da análise experimental (DBC).....	22
Homogeneidade de variâncias .....	22
Normalidade dos resíduos .....	23
Aditividade de blocos.....	24
Transformação Box-Cox.....	25
Análise de variância - ANOVA.....	26
Delineamento em quadrado latino.....	28
Homogeneidade de variâncias .....	31
Normalidade dos resíduos .....	32
Teste de aditividade do controle local .....	32
Análise de variância - ANOVA.....	33
Considerações finais .....	34
Referências .....	35



## Estatística experimental

A estatística experimental é a parte da estatística que dedica-se ao estudo dos experimentos em todas suas etapas: planejamento, condução, coleta e análise de dados. A partir da metade do século XIX, começaram a surgir pelo mundo as primeiras estações experimentais e uma delas, a de Rothamstead, em Londres, ganhou destaque quando Ronald A. Fisher assumiu sua direção e começou a desenvolver modelos estatísticos para interpretar os fenômenos que estavam sendo estudados.

Antes mesmo dos trabalhos pioneiros de Fisher, Johan Baptista van Helmont, em Bruxelas, introduziu uma das primeiras tentativas de pesquisa na agricultura, ao realizar um experimento que durou cinco anos, no qual ele comparou uma planta de salgueiro plantada em um vaso regado com água, com outra planta de salgueiro cultivada no solo. A partir dos resultados encontrados, van Helmont pôde fazer inferências e tomar decisões a partir de sua interpretação.

Como tem se observado desde a época de Fisher, a estatística é a ciência que se insere no contexto da pesquisa científica mediante a experimentação e seu objetivo é fornecer a condição necessária para que o pesquisador possa interpretar os dados obtidos e, a partir deles, tomar decisões importantes. Os experimentos, portanto, são experiências controladas feitas em condições pré-determinadas a partir de um problema ou hipótese.

Todo experimento deve produzir dados os quais deverão ser analisados por meio de um método estatístico apropriado, conforme o delineamento experimental utilizado.

### Noções básicas da experimentação

A partir da formulação de uma hipótese, o pesquisador deve elaborar um experimento contendo os fatores que serão estudados a partir de variáveis associadas a eles. Toda variação conhecida deve ser controlada, porém, durante a condução de um experimento, pode-se observar interferências/variações nos valores das unidades experimentais, geradas por fontes não controláveis denominadas “erros”.

Para que os erros sejam minimizados, deve-se utilizar um plano adequado para designar os tratamentos às unidades experimentais, alinhado à análise dos dados. Esse plano denomina-se delineamento experimental e deve ser escolhido para que os valores encontrados sejam resultantes somente da ação dos tratamentos e não da variação ambiental presente.

Assim, pode-se elencar alguns conceitos necessários para prosseguir:

- **Experimentação:** pesquisa e experimentação são dois conceitos muito próximos. Denomina-se pesquisa quando se está investigando algo novo, já experimentação emprega-se para comprovar uma hipótese ou um conhecimento previamente adaptado.
- **Fator:** é o objeto de estudo em um experimento. Pode ser adubação, irrigação, reagente, cultivar, etc.
- **Nível:** são as diferentes maneiras que os fatores serão apresentados. Por exemplo: doses de um adubo, níveis de irrigação, diferentes elementos químicos em um reagente, cultivares diferentes, etc.
- **Tratamento:** é o nome dado para cada nível de um fator.
- **Delineamento experimental:** plano adotado para distribuir os tratamentos às unidades experimentais.
- **Unidade experimental:** objeto que recebe o tratamento. Também comumente chamado de parcela, pode ser um único objeto ou mais. Exemplo: uma planta, um animal, uma área de solo, etc.

## Princípios básicos em experimentação

Para se obter o máximo de efeito dos tratamentos, deve-se tomar muitos cuidados na condução dos experimentos e o primeiro deles é a escolha de um delineamento experimental apropriado para minimizar a chance de ocorrência de variações ao acaso ou não controláveis. Todos esses cuidados podem ser denominados de:

- **Repetição:** princípio que consiste em repetir mais de uma vez os tratamentos em diferentes unidades experimentais. Sua presença é de extrema importância para se obter uma estimativa representativa de um determinado tratamento com base na resposta de várias unidades experimentais e com isso encontrar conclusões válidas. A variabilidade dos dados é o principal determinante para se encontrar o número de repetições ideal, no entanto, o maior limitante é a disponibilidade de recursos.
- **Casualização:** consiste em atribuir os tratamentos de forma aleatória às unidades experimentais. Isso evita a distribuição sistemática dos tratamentos e ainda garante que todos eles tenham a mesma probabilidade de serem designados a uma unidade experimental. Para cada delineamento existe um procedimento adequado para se fazer tal alocação e o pesquisador não deve negligenciá-lo. Com a casualização garante-se a independência entre as observações, assim como estimativas não tendenciosas da média e do erro de cada tratamento.
- **Controle local:** faz-se o controle local ou bloqueamento em um experimento quando as unidades experimentais não são homogêneas. Nesse caso, divide-se as parcelas em subgrupos com menor variabilidade, chamados blocos, para reduzir o erro experimental e otimizar o experimento. Além disso, é desejável que cada bloco deva receber pelo menos uma repetição de cada tratamento e com isso obter um delineamento em blocos completos. Quando isso não for possível, tem-se um delineamento em blocos incompletos e, nos casos com duas ou mais repetições por blocos, tem-se o chamado delineamento em blocos aumentados. O número de blocos depende da variação ambiental do experimento, porém um controle local elevado reduz o número de graus de liberdade do resíduo em uma análise de variância. Assim, essa ferramenta deve ser utilizada pelo pesquisador somente quando necessário.

## Os pressupostos para a ANOVA

Toda vez que um experimento for realizado e seus dados forem analisados mediante a análise de variância (ANOVA), é desejável que algumas pressuposições a respeito dos tratamentos e dos resíduos sejam observadas. Quando isso ocorrer, pode-se assumir que os resultados obtidos na ANOVA são confiáveis e levam a uma boa tomada de decisão a partir do teste F.

Sobre os tratamentos, deve-se observar que seus efeitos sejam aditivos (assim como os blocos); e sobre os resíduos a distribuição observada deve ser aproximadamente normal, centralizada em zero, independente e com variação constante. Todas essas suposições deverão ser testadas e avaliadas quanto a sua aceitação ou rejeição.

### Aditividade dos fatores do modelo

O modelo a respeito da observação  $y_{ij}$  é composto pela soma de parâmetros que representam os efeitos avaliados no modelo. Nos casos contendo blocos, seu efeito não deve apresentar interação com o efeito de tratamentos.

Tukey propôs um teste de não aditividade mediante a obtenção da soma de quadrados do efeito da interação entre blocos e tratamentos, com 1 grau de liberdade, dividida pelo quadrado médio residual, resultando em um teste F (Snedecor; Cochran, 1967). A falta de aditividade é dada quando esse teste é significativo.

## Normalidade dos resíduos

Para cada observação  $y_{ij}$  pode-se calcular um erro  $e_{ij}$  ou um desvio associado. O formato da distribuição dos erros pode ser verificado através de testes de normalidade, como por exemplo o de Shapiro-Wilk, muito utilizado para um conjunto de dados relativamente pequeno, ou então o de Kolmogorov-Smirnov. Gráficos também ajudam na avaliação da distribuição dos erros e um deles é o Q-Q Plot, o qual divide os valores dos erros em quartis.

## Homogeneidade de variâncias

Esse pressuposto é fundamentado em uma ideia muito simples: os tratamentos que estão sendo avaliados devem apresentar variação constante o suficiente para serem comparados. Caso haja um tratamento com uma variação muito superior aos demais, significa que o tratamento não é estável o suficiente a ponto de ser comparado.

Seguindo tal lógica, deve-se observar se a variação do componente aleatório (ou erro  $e_{ij}$ ) é aproximadamente constante. Entre os diversos testes propostos para se testar a homogeneidade das variâncias estão os testes de Bartlett (Snedecor; Cochran, 1983) e o de Levene (Levene, 1960).

## Transformação Box-Cox

Box e Cox (1964) apresentaram em seu artigo uma família de transformações possível em busca de um conjunto de dados aproximadamente normal, com variação constante e linearidade. Os autores mostraram que o parâmetro  $\lambda$  pode ser estimado simultaneamente com os outros parâmetros do modelo pela expressão  $Y^* = Y^\lambda$  através do método da máxima verossimilhança. Tem-se então:

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases} \quad (1.1)$$

em que  $\lambda$  é o parâmetro de transformação e  $Y$  é a variável resposta. Quando o intervalo de confiança conter o valor 1, a transformação não é necessária, uma vez que  $\lambda$  assume esse valor.

## Delineamento completamente casualizado

Ao se iniciar a parte prática deste trabalho, optou-se pelo uso do programa estatístico **R** para realizar as análises porque este programa vem sendo utilizado cada mais vez por pesquisadores e estudantes de diversas áreas. Na última década, o número de usuários do **R** aumentou consideravelmente fazendo dele o principal programa de análise de dados de diversas instituições de ensino e de pesquisa no Brasil. Além de ser gratuito para download, também é open source e permite que qualquer pessoa possa colaborar para sua melhoria.

O primeiro delineamento apresentado corresponde a análise de um experimento mais simples. Sua aplicação requer que todas as unidades experimentais sejam consideradas homogêneas e possam responder da mesma forma à ação de um tratamento. Nesse tipo de delineamento não é necessária a utilização de qualquer controle local.

Considera-se como exemplo o experimento (adaptado) conduzido em 2013/ 2014 na Embrapa Clima Temperado - Pelotas/RS, no âmbito do projeto de avaliação de variedades crioulas de abóboras (*Curcubita moschata*). No experimento de campo foram comparados 10 acessos provenientes do Banco Genético de Cucurbitaceae da Embrapa Clima Temperado, com plantas espaçadas 1,5 metros entre as linhas e 3 metros entre as colunas, em uma área experimental considerada homogênea, considerando 10 repetições para cada acesso. A variável que será apresentada é a espessura da casca (mm).

Nesse experimento o objetivo principal é verificar a hipótese:

$$\begin{cases} H_0 : t_1 = t_2 = \dots = t_{10} \\ H_1 : t_1 \neq t_2 \neq \dots \neq t_{10} \end{cases} \quad (2.1)$$

Formuladas as hipóteses, tem-se o devido teste estatístico para que uma decisão seja tomada: rejeitar ou não rejeitar a hipótese nula. Porém, caso a tomada de decisão seja equivocada, é possível se cometer dois tipos de erro. O erro tipo I ocorre quando rejeita-se a hipótese nula a qual é considerada verdadeira, sendo a probabilidade de sua ocorrência conhecida como  $\alpha$  - exatamente o nível de significância definido para o teste. Já o erro tipo II ocorre quando uma hipótese nula é considerada falsa, porém não é rejeitada. A probabilidade do erro tipo II ocorrer é chamada  $\beta$  a qual vai depender do poder do teste.

Sobre os erros de decisão o importante é deixar registrado que nenhum teste de hipótese tem 100% de garantia pois está baseado em probabilidades, assim, sempre é possível de cometê-los. Deve-se, então, estabelecer níveis de confiança suficiente para minimizar ao máximo qualquer um desses erros.

Voltando à análise do experimento, tem-se que o modelo estatístico utilizado para um delineamento com estrutura completamente casualizado e balanceado é dado por:

$$y_{ij} = \mu + t_i + e_{ij}, i = 1, \dots, I \text{ e } j = 1, \dots, J \quad (2.2)$$

em que:

- $y_{ij}$  é o valor observado da variável resposta na parcela (repetição)  $j$  referente ao  $i$ -ésimo tratamento;
- $\mu$  é um parâmetro constante do modelo, comum a todas as observações, sendo a média geral quando os dados são balanceados;
- $t_i$  é o parâmetro que representa o efeito do  $i$ -ésimo tratamento;
- $e_{ij}$  é o erro experimental associado ao modelo. É aleatório, não observável e atribuído à observação  $y_{ij}$ .

Os valores obtidos no experimento constam na tabela a seguir:

**Tabela 1.** Espessura da casca (mm) de 10 acessos de abóbora.

Repetições	Acessos									
	1	2	3	4	5	6	7	8	9	10
R1	1,09	0,75	0,66	0,97	0,83	0,84	0,64	0,71	0,68	0,88
R2	1,05	0,49	0,84	0,64	0,65	0,68	0,83	0,93	0,72	0,79
R3	1,01	0,68	0,69	0,97	0,48	0,66	0,71	0,76	0,79	0,67
R4	1,06	0,91	0,57	0,68	0,7	0,67	0,58	0,57	0,72	0,64
R5	1,01	0,82	0,67	0,51	0,72	0,74	0,67	0,51	0,8	0,66
R6	1,18	0,58	0,77	0,82	0,63	0,73	0,84	0,71	0,51	0,79
R7	0,94	0,82	0,78	0,64	0,68	0,66	0,56	0,65	0,69	0,63
R8	0,99	0,89	0,91	0,63	0,59	0,88	0,74	0,9	0,61	0,73
R9	0,97	0,92	0,77	0,87	0,68	0,62	0,76	0,76	0,75	0,81
R10	0,98	0,71	0,67	0,67	0,72	0,69	0,67	0,73	0,9	0,7

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2014.

## Análise descritiva dos dados

É importante examinar os dados de forma exploratória antes de se iniciar a análise experimental em si. Obter informações básicas ajudam o pesquisador a verificar a natureza das informações contidas no conjunto de dados carregado.

```
dados_dic <- read.table("dados_dic.txt",header=T)
str(dados_dic)

## 'data.frame': 100 obs. of 3 variables:
## $ acesso : Factor w/ 10 levels "A01","A02","A03",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ rep : int 1 1 1 1 1 1 1 1 1 1 ...
## $ espessura: num 1.09 0.75 0.66 0.97 0.83 0.84 0.64 0.71 0.68 0.88 ...

attach(dados_dic)
```

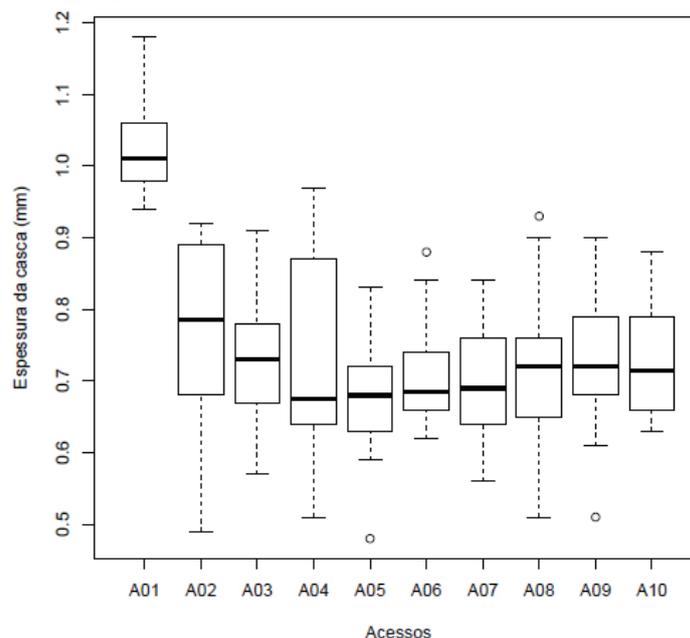
No *software* estatístico **R**, o comando *str* fornece informações sobre a natureza dos dados, identificando os fatores e seus respectivos níveis, as variáveis e sua natureza: se numérica ou inteira. A análise exploratória pode ser feita de diversas maneiras. Sugere-se o comando *summary* para obter informações básicas como o mínimo, o máximo, os quartis e a média dos dados, além do comando *boxplot* para fazer os respectivos *box-plots*.

```
summary(dados_dic)

##      acesso      rep      espessura
## A01   :10   Min.   : 1.0   Min.   :0.4800
## A02   :10  1st Qu.: 3.0   1st Qu.:0.6600
## A03   :10  Median : 5.5   Median :0.7200
## A04   :10  Mean   : 5.5   Mean   :0.7513
## A05   :10  3rd Qu.: 8.0   3rd Qu.:0.8325
## A06   :10  Max.   :10.0   Max.   :1.1800
## (Other):40

boxplot(espessura ~ acesso, xlab = "Acessos", ylab = "Espessura da casca (mm)")
```

Figura 1. Box-plot dos dados por tratamento (DIC).



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

Pode-se observar na Figura 1 alguns pontos fora dos limites das caixas e isso pode indicar um possível *outlier*. É muito importante investigá-los para saber se esses valores são considerados atípicos, devendo ou não ser excluídos do conjunto de dados. Nesse caso, optou-se por mantê-los no conjunto de dados uma vez que os valores foram corretamente registrados e não apresentam qualquer irregularidade.

## Verificação dos pressupostos da análise experimental (DIC)

Para proceder à análise de variância é fundamental realizar a conferência de dois pressupostos básicos para validação dos resultados da ANOVA: a homogeneidade das variâncias entre os tratamentos e a normalidade dos resíduos.

### Homogeneidade de variâncias

No caso desse delineamento experimental, a hipótese a ser testada é:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \dots = \sigma_{10}^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \dots \neq \sigma_{10}^2 \end{cases} \quad (2.3)$$

Nesse tipo de teste de hipótese, sempre se considera um teste bilateral. Caso o p-valor encontrado seja inferior ao nível de significância determinado ( $\alpha$ ), deve-se rejeitar  $H_0$ , assim as variâncias não podem ser consideradas homogêneas.

Um teste muito utilizado pelos agrônomos para verificar a igualdade de variâncias é o teste de Bartlett (Snedecor; Cochran, 1989). Esse teste está diretamente relacionado com a normalidade dos dados, porém, se tal condição não for atendida, é aconselhável optar pelo teste de Levene.

```
bartlett.test(espessura ~ acesso, data=dados_dic)

##
## Bartlett test of homogeneity of variances
##
## data:  espessura by acesso
## Bartlett's K-squared = 10.729, df = 9, p-value = 0.2947
```

Como o teste de Bartlett indicou um p-valor de 0,2947, não deve-se rejeitar a hipótese de igualdade das variâncias.

### Normalidade dos resíduos

Esse teste de hipótese pode ser descrito como:

$$\begin{cases} H_0 : \text{os resíduos apresentam distribuição normal} \\ H_1 : \text{os resíduos não apresentam distribuição normal} \end{cases} \quad (2.4)$$

Para se verificar a normalidade dos resíduos, é conveniente utilizar os resíduos padronizados ou *studentizados* (uma vez que resíduos maiores que 2 sugerem a presença de pontos discrepantes chamados *outliers*).

Para isso, utiliza-se o comando *rstudent*.

Entre os testes disponíveis, o mais utilizado é o teste de Shapiro-Wilk (1965) dado pelo comando:

```
shapiro.test(rstudent(aov(espessura ~ acesso, data=dados_dic)))

##
## Shapiro-Wilk normality test
##
## data:  rstudent(aov(espessura ~ acesso, data = dados_dic))
## W = 0.98689, p-value = 0.4301
```

De acordo com o p-valor (0,4301), existem evidências suficientes para não rejeitar a hipótese de normalidade dos resíduos. Outra opção é o teste proposto por Lilliefors (1967), o qual está contido no pacote **nortest**:

```
require(nortest)

## Loading required package: nortest
## Warning: package 'nortest' was built under R version 3.5.2

lillie.test(rstudent(aov(espessura ~ acesso, data=dados_dic)))

##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  rstudent(aov(espessura ~ acesso, data = dados_dic))
## D = 0.056121, p-value = 0.613
```

Mais uma vez, foi obtido um p-valor superior ao nível  $\alpha = 0,05$  estabelecido e, com isso, a decisão correta é não rejeitar a hipótese nula.

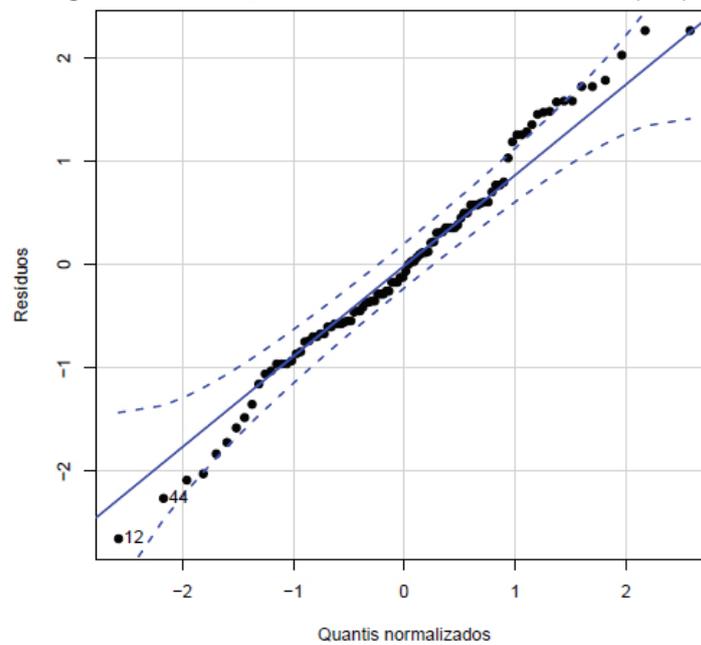
Além dos testes para a verificação da normalidade dos resíduos, também é possível fazer uma avaliação gráfica da sua distribuição através do chamado *Q-Q Plot* com um intervalo simulado. A indicação de normalidade é apresentada quando todos os pontos do gráfico encontram-se dentro dos limites estabelecidos (linhas pontilhadas). O comando *qqPlot*, usado para gerá-los, está contido no pacote **car**.

```
require(car)

## Loading required package: car
## Loading required package: carData

qqPlot(rstudent(aov(espessura ~ acesso, data = dados_dic))
       , pch=19, ylab="Resíduos", xlab="Quantis normalizados")
```

Figura 2 - Verificação da normalidade dos resíduos (DIC).



```
## [1] 12 44
```

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

Pontos fora dos limites ou com valores muito diferentes dos demais podem comprometer a normalidade dos resíduos. Esses pontos são destacados na Figura 2 para que o pesquisador investigue a sua causa e decida em mantê-los ou não nas análises. Nesse caso, os pontos 12 e 44 destacaram-se dos demais e foram confirmados nas planilhas de campo, logo permanecerão nas análises.

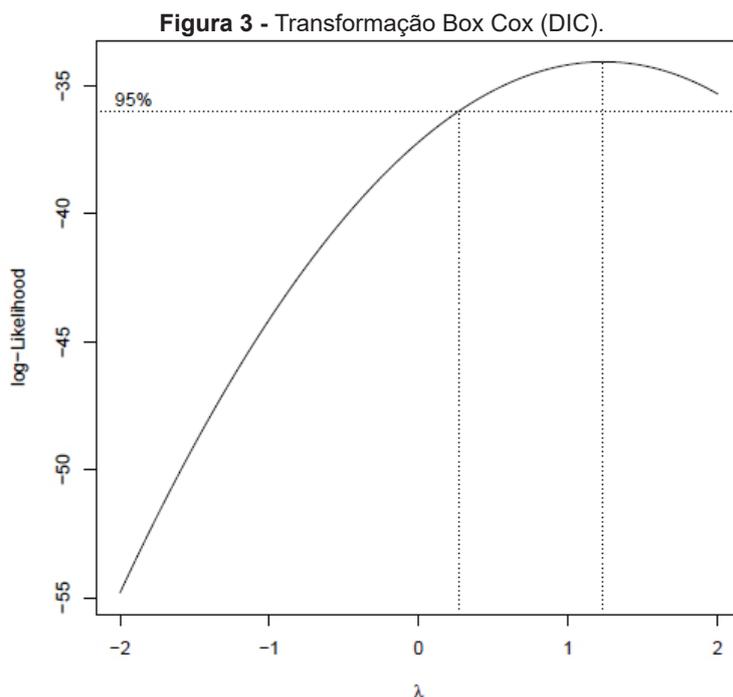
### Transformação Box-Cox

Para verificar se existe a indicação de alguma transformação dos dados, utiliza-se o comando `boxcox` do pacote **MASS**. Caso o intervalo de confiança para o parâmetro  $\lambda$  contenha o valor 1, não existem evidências suficientes para que os dados sejam transformados.

```
require(MASS)

## Loading required package: MASS
## Warning: package 'MASS' was built under R version 3.5.2

boxcox(aov(espessura ~ acesso, data=dados_dic))
```



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

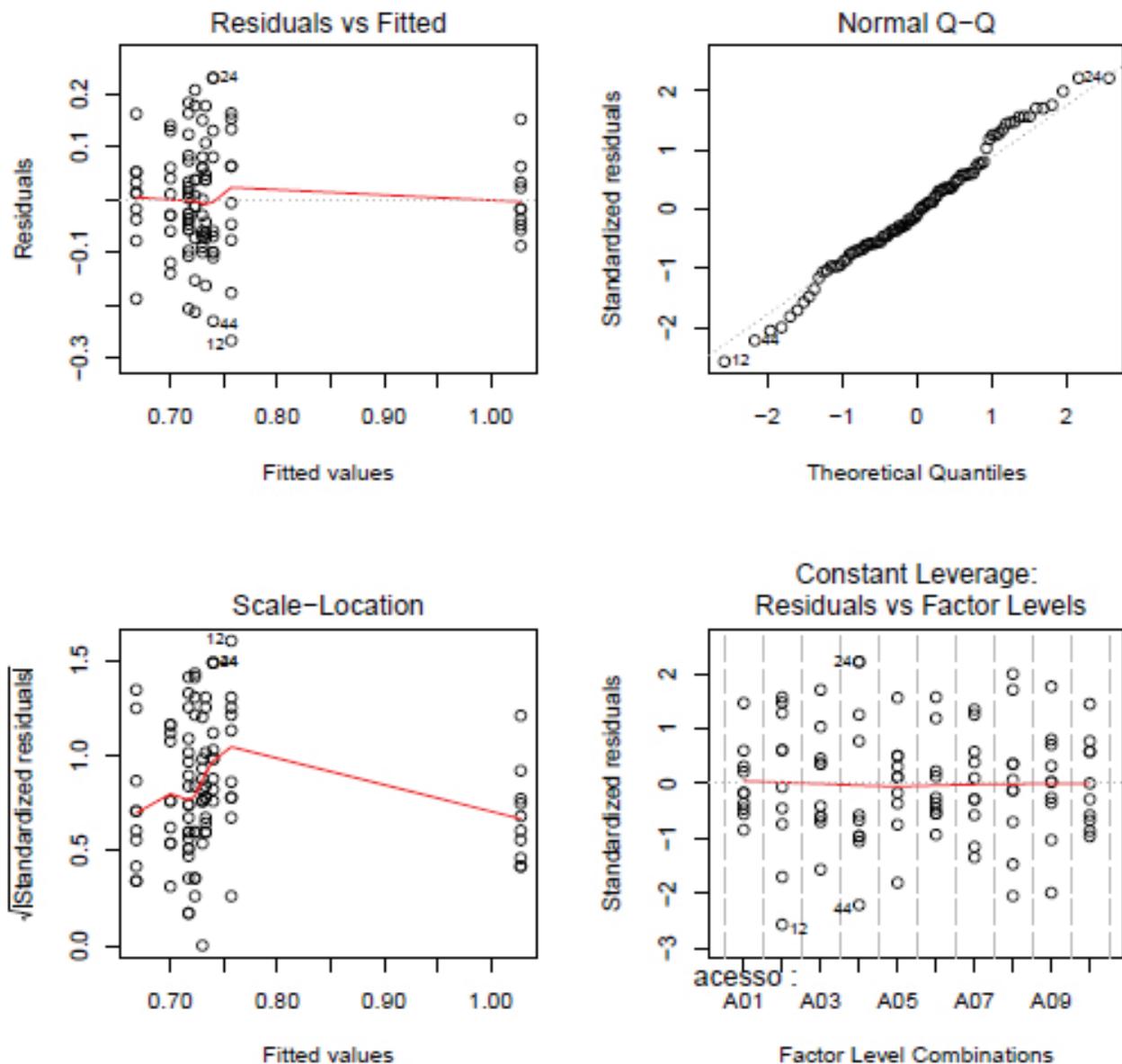
Quando o intervalo de confiança conter  $\lambda = 1$ , então não há necessidade de realizar qualquer transformação nos dados. Caso contrário, deve-se proceder a transformação dos dados e depois verificar se os pressupostos foram atendidos.

### Gráficos auxiliares

É possível a visualização de mais alguns gráficos que podem auxiliar na análise dos pressupostos da ANOVA. Utilizando o comando `plot` para o resultado da ANOVA obtido pelo comandos `aov`, pode-se observar quatro gráficos complementares.

```
par(mfrow=c(2,2))
plot(aov(espessura ~ acesso,data=dados_dic))
```

Figura 4 . Gráficos auxiliares da ANOVA (DIC).



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

Estes quatro gráficos nos fornecem informações a respeito da distribuição dos resíduos. Sua descrição ocorrerá esquerda para direita, de cima para baixo.

No primeiro gráfico, observa-se os valores dos resíduos em relação aos valores preditos e aqui verifica-se não haver qualquer tendência. Pode-se considerar a presença de tendência caso os valores não fiquem dispersos em torno do resíduo zero. O segundo gráfico é o *Normal Q-Q Plot*, no qual verifica-se a normalidade dos resíduos. Quando os pontos formam uma linha reta inclinada e concentram-se no intervalo de -2 a 2 no eixo X (quartis teóricos), há uma grande chance de se considerar a normalidade dos resíduos, porém a análise visual deve ser confirmada através de um teste específico. O terceiro gráfico fornece a dispersão da raiz quadrada dos resíduos padronizados - gráfico semelhante ao primeiro, apenas com uma pequena mudança na escala. Por fim, o último gráfico apresenta os resíduos padronizados em relação a cada tratamento, fornecendo uma noção da independência dos resíduos: valores distribuídos ao longo do resíduo zero sugerem a independência dos mesmos.

Verificados os pressupostos, inicia-se os procedimentos da ANOVA.

### Análise de Variância - ANOVA

Para executar a ANOVA, a sequência mais simples consiste em utilizar o comando `aov` e em seguida o comando `summary` no resultado anterior:

```
aovdic <- aov(espessura ~ acesso,data=dados_dic)
summary(aovdic)

##           Df Sum Sq Mean Sq F value    Pr(>F)
## acesso      9  0.9024  0.10026    8.351 5.94e-09 ***
## Residuals  90  1.0806  0.01201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Apresenta-se a tabela completa da ANOVA com todos os seus componentes. É possível notar que há um grande número de graus de liberdade total e nos resíduos, o que tornará a soma de quadrados residual baixa. Com isso, o experimento apresenta-se robusto para testar os efeitos dos tratamentos. Como o p-valor do teste F é aproximadamente igual a zero, deve-se rejeitar a hipótese nula.

Com o resultado da ANOVA pelo menos um dos acessos deve ser diferente dos demais, portanto é necessário aplicar um teste de comparação de médias. Um dos mais difundidos e utilizado pelos pesquisadores da área agrônômica é o teste de Tukey (1953).

### Teste de comparação de médias

Uma maneira simples de se encontrar a diferença mínima significativa (DMS) utilizada no teste de Tukey é dada pelo comando `HSD.test`, presente no pacote **agricolae**.

Esse comando é muito prático, pois fornece os valores das médias de cada tratamento, seguidos das letras provenientes do teste de Tukey. Como pode ser visto, somente o tratamento A01 diferiu dos demais, fato observado previamente no box-plot da Figura 1, gerado na análise exploratória dos dados.

```

require(agricolae)

## Loading required package: agricolae
## Warning: package 'agricolae' was built under R version 3.5.3

HSD.test(aovdic,"acesso",console=T)

##
## Study: aovdic - "acesso"
##
## HSD Test for espessura
##
## Mean Square Error: 0.01200633
##
## acesso, means
##
##      espessura      std  r  Min  Max
## A01      1.028 0.06988880 10 0.94 1.18
## A02      0.757 0.14391741 10 0.49 0.92
## A03      0.733 0.09957168 10 0.57 0.91
## A04      0.740 0.15726835 10 0.51 0.97
## A05      0.668 0.09199034 10 0.48 0.83
## A06      0.717 0.08340663 10 0.62 0.88
## A07      0.700 0.09498538 10 0.56 0.84
## A08      0.723 0.12970479 10 0.51 0.93
## A09      0.717 0.10708771 10 0.51 0.90
## A10      0.730 0.08406347 10 0.63 0.88
##
## Alpha: 0.05 ; DF Error: 90
## Critical Value of Studentized Range: 4.588313
##
## Minimum Significant Difference: 0.1589858
##
## Treatments with the same letter are not significantly different.
##
##      espessura groups
## A01      1.028      a
## A02      0.757      b
## A04      0.740      b
## A03      0.733      b
## A10      0.730      b
## A08      0.723      b
## A06      0.717      b
## A09      0.717      b
## A07      0.700      b
## A05      0.668      b

```

## Delineamento em blocos completos

Nesse tipo de delineamento o modelo aditivo apresenta o componente que representa o efeito de bloco  $b$  e sua expressão é dada por:

$$y_{ij} = \mu + t_i + b_j + e_{ij}, i = 1, \dots, I \text{ e } j = 1, \dots, J \quad (3.1)$$

em que:

- $y_{ij}$  é o valor observado da variável resposta na parcela com o tratamento  $i$  no bloco  $j$ ;
- $\mu$  é um parâmetro constante do modelo, comum a todas as observações, sendo a média geral quando os dados são balanceados;
- $t_i$  é o parâmetro que representa o efeito do  $i$ -ésimo tratamento;
- $b_j$  é o parâmetro que representa o efeito do  $j$ -ésimo bloco;
- $e_{ij}$  é um componente do erro aleatório associado ao  $i$ -ésimo tratamento do  $j$ -ésimo bloco.

O estudo dos pressupostos nesse delineamento será abordado a partir da adaptação de experimento conduzido no ano de 2013 na Embrapa Clima Temperado, em Pelotas/RS, a respeito da massa fresca de frutos (g) de 14 acessos de variedades crioulas de pimenta da espécie *Capsicum annuum* em um delineamento com três blocos.

**Tabela 2** - Dados de 14 acessos de pimenta

Acesso	Bloco	Massa fresca de frutos (g)
P7	1	715,0
P7	2	603,9
P22	3	587,4
P22	1	384,3
P22	2	159,2
P39	3	169,8
P39	1	333,4
P39	2	271,9
P58	3	257,5
P58	1	951,6
P58	2	378,9
P77	3	555,2
P77	1	346,4
P77	2	188,2
P111	3	264,5
P111	1	1064,1
P111	2	665,2
P119	3	800,6
P119	1	461,0
P119	2	330,3
P122	3	311,0
P122	1	295,0
P122	2	206,7
P143	3	63,3
P143	1	146,0
P143	2	298,2
P202	3	234,4
P202	1	993,4
P202	2	900,8
P258	3	441,1
P258	1	560,3
P258	2	324,6
P259	3	189,9
P259	1	1029,3
P259	2	551,0
P266	3	734,7
P266	1	1203,1
P266	2	949,1
P302	3	1180,2
P302	1	190,1
P302	2	84,7
P302	3	95,1

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2013.

Nesse delineamento experimental, o objetivo principal é verificar as hipóteses:

$$\text{Para os tratamentos: } \begin{cases} H_0 : t_1 = t_2 = \dots = t_{14} \\ H_1 : t_1 \neq t_2 \neq \dots \neq t_{14} \end{cases} \quad (3.2)$$

$$\text{Para os blocos: } \begin{cases} H_0 : b_1 = b_2 = \dots = b_{14} \\ H_1 : b_1 \neq b_2 \neq \dots \neq b_{14} \end{cases} \quad (3.3)$$

## Verificação dos pressupostos da análise experimental (DBC)

### Homogeneidade de variâncias

A verificação da homogeneidade da variância de um experimento em blocos casualizados possui as mesmas hipóteses que no delineamento completamente casualizado.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \dots = \sigma_{14}^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \dots \neq \sigma_{14}^2 \end{cases} \quad (3.4)$$

Da maneira como os blocos foram inseridos no arquivo de dados (números de 1 a 3), o programa **R** os considera uma informação numérica. Precisa-se, portanto, alterar sua classe para fator e, em seguida, realizar o teste de Bartlett para os dados residuais.

```
dados_dbc <- read.table("dados_dbc.txt",header=T)
str (dados_dbc)

## 'data.frame': 42 obs. of 3 variables:
## $ acesso : Factor w/ 14 levels "P111","P119",...: 13 13 13 6 6 6 11 11 11 12 ...
## $ bloco : int 1 2 3 1 2 3 1 2 3 1 ...
## $ mat_fre: num 715 604 587 384 159 ...

dados_dbc$bloco <- as.factor(dados_dbc$bloco)
attach (dados_dbc)

## The following object is masked from dados_dbc:
##
## acesso

bartlett.test(residuals(aov(mat_fre ~ bloco,
                          data=dados_dbc)) ~ acesso, data=dados_dbc)

##
## Bartlett test of homogeneity of variances
##
## data: residuals(aov(mat_fre ~ bloco, data = dados_dbc)) by acesso
## Bartlett's K-squared = 18.556, df = 13, p-value = 0.1375
```

De acordo com p-valor encontrado (0,1375), a decisão correta é não rejeitar a hipótese nula. Portanto, tal pressuposto é atendido. Apenas para constar, apresenta-se, também, o teste de Levene.

```
require(car)
leveneTest(residuals(aov(mat_fre ~ bloco,
                          data=dados_dbc)) ~ acesso, data=dados_dbc)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 13  0.7067 0.7413
##      28
```

O p-valor encontrado no teste de Levene (0,7413) é superior a 0,05, assim também não se rejeita  $H_0$ .

### Normalidade dos resíduos

As hipóteses consideradas são idênticas a um delineamento completamente casualizado:

$$\begin{cases} H_0 : \text{os resíduos apresentam distribuição normal} \\ H_1 : \text{os resíduos não apresentam distribuição normal} \end{cases} \quad (3.5)$$

Para a realização do teste de Shapiro-Wilk, é preciso incluir o fator bloco para a obtenção dos resíduos.

```
shapiro.test(residuals(aov(mat_fre ~ bloco + acesso,
                          data=dados_dbc)))

##
## Shapiro-Wilk normality test
##
## data: residuals(aov(mat_fre ~ bloco + acesso, data = dados_dbc))
## W = 0.97271, p-value = 0.4053
```

Como o p-valor do teste não é inferior ao nível de significância  $\alpha = 0,05$ , a decisão mais adequada é assumir que os resíduos do modelo seguem uma distribuição aproximadamente normal.

O teste de Kolmogorov-Smirnov também pode ser utilizado nesse tipo de delineamento, basta utilizar os comandos a partir dos resíduos studentizados:

```
ks.test(rstudent(aov(mat_fre ~ bloco + acesso,
                    data=dados_dbc)), "pnorm")

##
## One-sample Kolmogorov-Smirnov test
##
## data: rstudent(aov(mat_fre ~ bloco + acesso, data = dados_dbc))
## D = 0.089822, p-value = 0.857
## alternative hypothesis: two-sided
```

Pelo resultado encontrado (p-valor 0,857), deve-se tomar a mesma decisão de não rejeitar  $H_0$ . Ainda, uma última possibilidade é a utilização do teste de Lilliefors. A principal diferença entre eles se dá pelo fato de que o teste de Lilliefors considera a média e o desvio-padrão dos resíduos, enquanto o teste de Kolmogorov-Smirnov considera a média e os desvios encontrados.

```
require(nortest)
lillie.test(rstudent(aov(mat_fre ~ bloco + acesso, data=dados_dbc)))

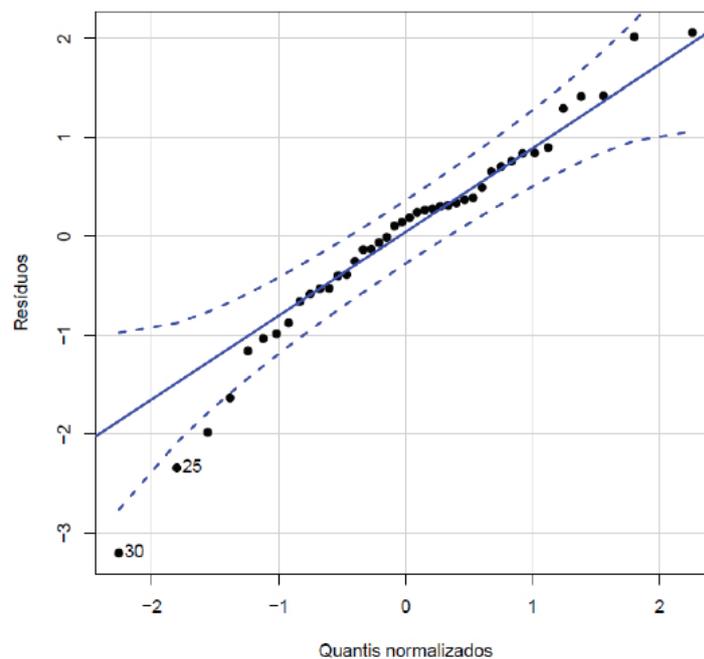
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: rstudent(aov(mat_fre ~ bloco + acesso, data = dados_dbc))
## D = 0.097413, p-value = 0.4042
```

Mais uma vez, a decisão correta é não rejeitar  $H_0$ .

Visualmente, a análise da normalidade dos resíduos pode ser feita mediante a construção do *Q-Q Plot* com o intervalo de simulação. Uma vez carregado o pacote **car**, pode-se gerar o gráfico executando o comando:

```
qqPlot(rstudent(aov(mat_fre ~ bloco + acesso, data=dados_dbc)),
       pch=19, ylab="Resíduos", xlab="Quantis normalizados")
```

Figura 5.- Verificação da normalidade dos resíduos (DBC).



```
## [1] 30 25
```

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

No gráfico da Figura 5 a grande maioria dos pontos estão distribuídos dentro dos limites estabelecidos, e isso é desejável, pois revela indícios da presença de normalidade. Visualiza-se apenas dois pontos fora dos limites: 25 e 30. Esses devem ser avaliados individualmente pois podem ser considerados *outliers* ou não. Em um grande volume de dados, é comum encontrar alguns pontos destacados, assim como ocorreu no experimento em questão.

Após verificação, constatou-se que tais pontos devem permanecer no conjunto de dados e nas análises. Havendo muitos pontos destacados, o pressuposto de normalidade dos resíduos pode não ser atendido.

### Aditividade de blocos

Nesse delineamento, é possível verificar a aditividade de blocos com a aplicação do teste de Tukey apropriado. Busca-se identificar a presença de aditividade para garantir que o efeito dos tratamentos não depende dos blocos. As hipóteses consideradas são:

$$\begin{cases} H_0 : \text{os blocos são aditivos} \\ H_1 : \text{os blocos não apresentam aditividade} \end{cases} \quad (3.6)$$

O teste é executado no **R** calculando-se primeiramente a ANOVA ajustada com os erros por bloco e utilizando-se posteriormente o comando *tukey.1df* do pacote **dae**:

No resultado do teste, encontramos o p-valor na saída *Tukey.p*, ou seja, 0,0843. Outra opção é dada pelo comando *tukey.add.test* do pacote **asbio**:

```
require(asbio)

## Loading required package: asbio
## Warning: package 'asbio' was built under R version 3.5.2
## Loading required package: tcltk

with(dados_dbc,tukey.add.test(mat_fre, acesso, bloco))

##
## Tukey's one df test for additivity
## F = 3.2303512 Denom df = 25 p-value = 0.0843764
```

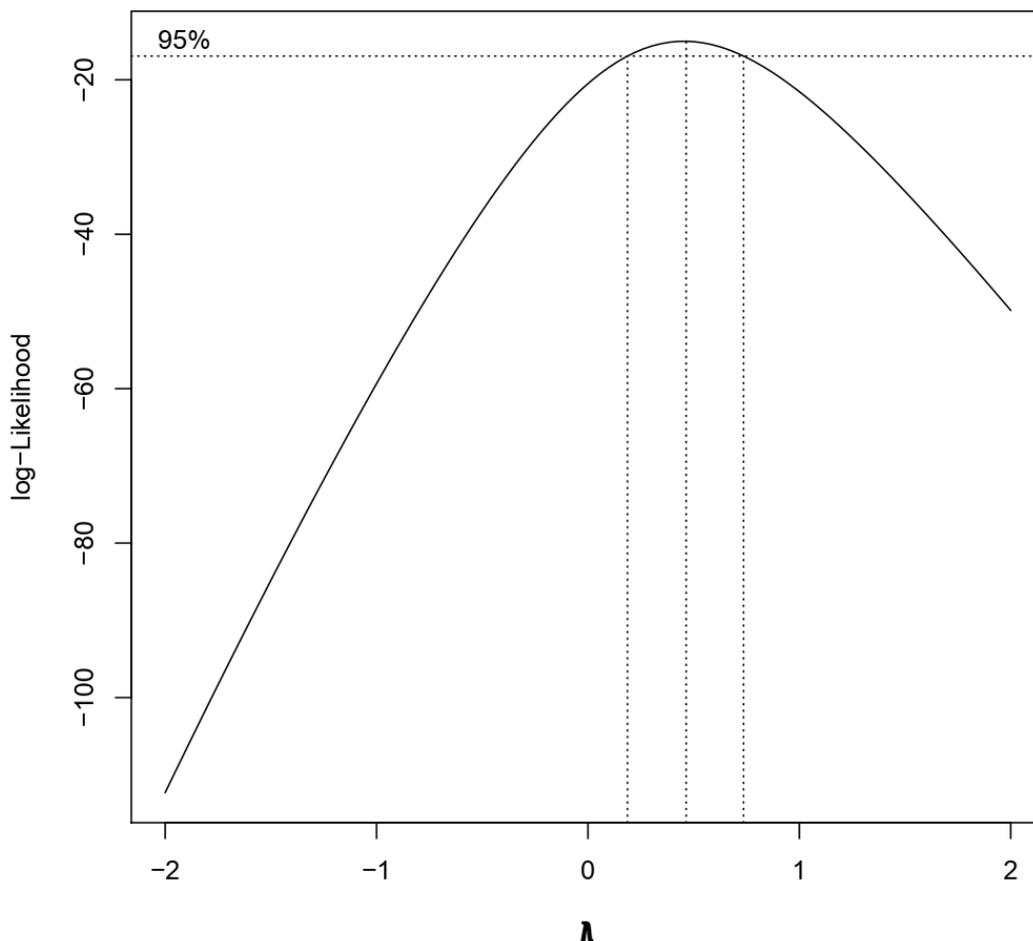
Nos dois casos, o p-valor encontrado é o mesmo, pois ambos utilizam as mesmas funções e cálculos disponíveis no R. Como seu valor é maior que o valor de  $\alpha = 0,05$ , não se deve rejeitar a hipótese de que os blocos têm efeito aditivo no modelo.

### Transformação Box-Cox

O pacote **MASS** proposto por Venables e Ripley (2002) apresenta o comando *boxcox* para traçar a função de verossimilhança, considerando diferentes valores de  $\lambda$  com um intervalo de 95% de confiança, e proporciona uma análise gráfica para a tomada de decisão em relação à transformação dos dados.

```
require(MASS)
boxcox(aov(mat_fre ~ bloco + acesso,data=dados_dbc))
```

Figura 6. Transformação Box Cox (DBC).



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

De acordo com o resultado encontrado, observa-se que o intervalo com 95% de confiança para  $\lambda$  não contempla o valor 1, logo uma transformação dos dados é sugerida. Entre as opções disponíveis, optou-se pela transformação raiz quadrado para proceder às análises.

```
mat_fre2 <- sqrt(mat_fre)
```

### Análise de variância - ANOVA

Embora o comando `av` já tenha sido utilizado para gerar os resíduos, esse mesmo comando será utilizado para obter a ANOVA a partir dos dados transformados e em seguida será aplicado o comando `summary`.

```
aovdbc <- aov(mat_fre2 ~ acesso + bloco, data=dados_dbc)
summary(aovdbc)

##           Df Sum Sq Mean Sq F value    Pr(>F)
## acesso     13 1841.7   141.67   18.85 5.39e-10 ***
## bloco       2  184.2    92.09   12.25 0.000178 ***
## Residuals  26  195.4     7.52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nessa situação contempla-se o fator bloco, porém sua significância não é o objeto principal de estudo. Apenas é desejável que seu efeito seja significativo para comprovar que o efeito local foi bem aplicado.

Verificando a significância dos tratamentos, observa-se que o p-valor encontrado é menor que 0,05, ou seja, tanto para blocos quanto para tratamentos os resultados foram significativos. Isso representa dizer que o controle local foi muito bem empregado e os acessos apresentam diferença significativa entre si. Deve-se, então, prosseguir com um teste de comparação de médias e, para isso, utilizar-se-á o comando `HSD.test`.

```
HSD.test(aovdbc, "acesso", console=T)

##
## Study: aovdbc ~ "acesso" ##
## HSD Test for mat_fre2 ##
## Mean Square Error: 7.515298 ##
## acesso, means
##

##           mat_fre2           std r           Min           Max
## P111  28.90230  3.454820 3  25.79147  32.62055
## P119  19.09342  2.076528 3  17.63519  21.47091
## P122  13.16959  4.726838 3   7.95613  17.17556
## P143  14.88721  2.618452 3  12.08305  17.26847
## P202  27.51132  5.686907 3  21.00238  31.51825
## P22   15.08392  3.919586 3  12.61745  19.60357
## P258  18.48925  4.962025 3  13.78042  23.67066
## P259  27.55381  4.322143 3  23.47339  32.08271
## P266  33.28241  2.149773 3  30.80747  34.68573
```

```
## P302 10.91429 2.503504 3 9.20326 13.78768
## P39 16.93181 1.170694 3 16.04681 18.25924
## P58 24.62535 5.765257 3 19.46535 30.84801
## P7 25.18340 1.358167 3 24.23634 26.73948
## P77 16.19796 2.447269 3 13.71860 18.61182
##
## Alpha: 0.05 ; DF Error: 26
## Critical Value of Studentized Range: 5.211004
##
## Minimum Significant Difference: 8.24772
##
## Treatments with the same letter are not significantly different.
##
##      mat_fre2 groups
## P266 33.28241    a
## P111 28.90230   ab
## P259 27.55381   ab
## P202 27.51132   ab
## P7   25.18340   abc
## P58  24.62535   bcd
## P119 19.09342   cde
## P258 18.48925   cde
## P39  16.93181   de
## P77  16.19796   e
## P22  15.08392   e
## P143 14.88721   e
## P122 13.16959   e
## P302 10.91429   e
```

Com 95% de confiança, o teste de Tukey mostrou as diferenças entre os acessos e, com isso, a formação dos diferentes grupos. Se, por algum motivo, esse experimento tivesse sido realizado sem o controle local, seria possível encontrar outro resultado no teste de Tukey e, certamente, uma outra conclusão.

## Delineamento em quadrado latino

O delineamento em quadrado latino considera a aplicação do controle local em duas direções distintas: linhas e colunas. Certamente, esse não é um dos delineamentos preferidos dos pesquisadores na área da agronomia devido a limitações práticas para sua implantação, porém sua aplicação ainda é válida.

Nesse tipo de delineamento, o número de linhas, de colunas e de repetições por tratamento deve ser o mesmo, formando, portanto, um quadrado, como o nome do delineamento sugere. Nesse contexto, experimentos com um número grande de tratamentos requerem um número muito grande de parcelas, e esse número cresce rapidamente, à medida que se aumenta o número de tratamentos.

Por isso, é muito comum encontrar experimentos em quadrado latino do tipo 2x2, 3x3, 4x4 ou, no máximo, 5x5. Casos com mais tratamentos raramente serão conduzidos dessa forma.

Ao se optar por esse delineamento, o croqui do experimento deverá conter um controle em linha e outro em coluna. Dessa forma, os tratamentos serão alocados às unidades experimentais de maneira que cada um esteja presente uma única vez em cada linha e cada coluna. O exemplo abaixo ilustra essa situação.

O modelo matemático de um delineamento em quadrado latino também é um modelo aditivo de forma que:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}, \quad j, i=1, 2, \dots, R \quad (4.1)$$

- $y_{ijk}$  é o valor observado da variável resposta na  $i$ -ésima linha e  $j$ -ésima coluna referente ao  $k$ -ésimo tratamento;
- $\mu$  é a média geral;
- $\alpha_i$  é o parâmetro que representa o efeito produzido pela  $i$ -ésima linha;
- $\beta_j$  é o parâmetro que representa o efeito da  $j$ -ésima coluna;
- $\tau_k$  é o parâmetro que representa o efeito pelo  $k$ -ésimo tratamento;
- $e_{ijk}$  é um componente do erro aleatório associado ao  $k$ -ésimo tratamento da  $i$ -ésima linha e  $j$ -ésima coluna.

**Tabela 3.** Croqui de um experimento com delineamento em quadrado latino 4x4.

		Colunas			
Linhas		1	2	3	4
1		C	D	A	B
2		B	C	D	A
3		A	B	C	D
4		D	A	B	C

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

Após o sorteio entre as linhas e as colunas do croqui, tem-se o delineamento devidamente esquematizado para receber os tratamentos. Assim, pode-se observar que esse delineamento não permite avaliar a presença de interação entre os controles locais considerados. Caso o pesquisador suspeite de interações significativas, sugere-se utilizar um experimento fatorial.

### Verificação dos pressupostos da análise experimental (DQL)

Considera-se o experimento (adaptado) desenvolvido na Embrapa Clima Temperado, em Pelotas/RS, no ano de 2016, para avaliar substratos na produção de mudas da variedade RB867515 de cana-de-açúcar. Esse experimento contém cinco tratamentos, dos quais quatro compõem diferentes proporções de lodo de esgoto e um representa o substrato comercial. A variável analisada foi o diâmetro do colmo da cana (mm) e os controles locais considerados foram: luminosidade (cinco faixas nas linhas) e maturidade das gemas da cana (cinco tempos nas colunas), as quais são representadas pela localização das gemas na plantas. O resultado do sorteio dos tratamentos nas linhas e nas colunas resultou no seguinte croqui:

**Tabela 4.** Croqui do delineamento em quadrado latino para os dados do diâmetro do colmo de cana-de-açúcar (mm).

Colunas					
Linhas	1	2	3	4	5
1	T4	T1	T3	T2	T5
2	T3	T5	T2	T1	T4
3	T2	T4	T1	T5	T3
4	T1	T3	T5	T4	T2
5	T5	T2	T4	T3	T1

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

Numericamente, tem-se os seguintes valores observados de cada tratamento:

**Tabela 5.** Dados do diâmetro do colmo (mm) do experimento para avaliar cinco substratos na produção de mudas de cana-de-açúcar.

Colunas					
Linhas	1	2	3	4	5
1	7,65	7,66	8,36	8,89	8,42
2	8,92	8,63	10,5	8,42	7,57
3	8,51	9,88	8,56	7,93	9,98
4	8,11	8,34	8,32	8,48	9,00
5	9,88	8,50	8,64	8,54	9,80

Fonte: Embrapa Clima Temperado - Pelotas/RS, 2016.

```
dados_dql <- read.table("dados_dql.txt", header=T) str
(dados_dql)

## data.frame: 25 obs. of 4 variables:
## $ trat : Factor w/ 5 levels "T1", "T2", "T3",...: 4 1 3 2 5 3 5 2 1 4 ...
## $ linha : int 1 1 1 1 1 2 2 2 2 2 ...
## $ coluna: int 1 2 3 4 5 1 2 3 4 5 ...
## $ diam : num 7.65 7.66 8.36 8.89 8.42 8.92 8.63 9.5 8.42 7.57 ...
```

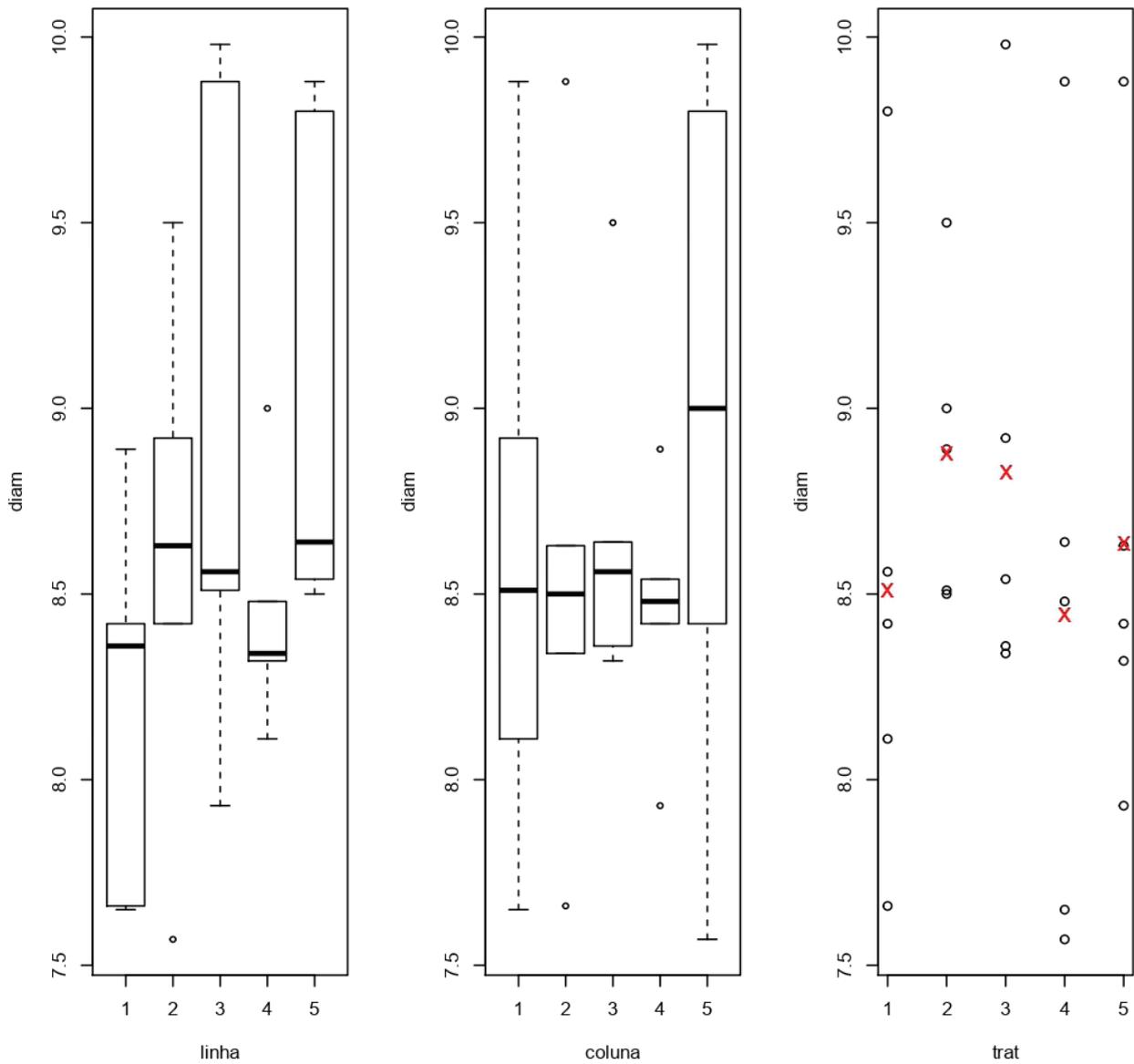
As linhas e colunas estão apresentadas como valores numéricos e é preciso defini-las como fatores. Para isso, utiliza-se o comando *as.factor*:

```
dados_dql$linha <- as.factor(dados_dql$linha)
dados_dql$coluna <- as.factor(dados_dql$coluna) attach
(dados_dql)
```

Pode-se proceder à análise exploratória gerando três gráficos para visualizar os valores obtidos em relação às linhas, colunas e aos tratamentos, observando a existência de interação entre eles, algo não desejável nesse tipo de delineamento.

```
par(mfrow=c(1,3))
plot(diam ~ linha + coluna)
med <- tapply(diam, trat, mean) plot.default(trat,
diam)
points(med, pch="x", col=2, cex=1.5)
```

Figura 7. Análise exploratória de dados (DQL).



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

No primeiro gráfico à esquerda observa-se a dispersão dos dados em relação às linhas; no segundo em relação às colunas e no terceiro a dispersão dos dados em relação a cada tratamento, em que o ponto "X" representa o valor médio de cada tratamento.

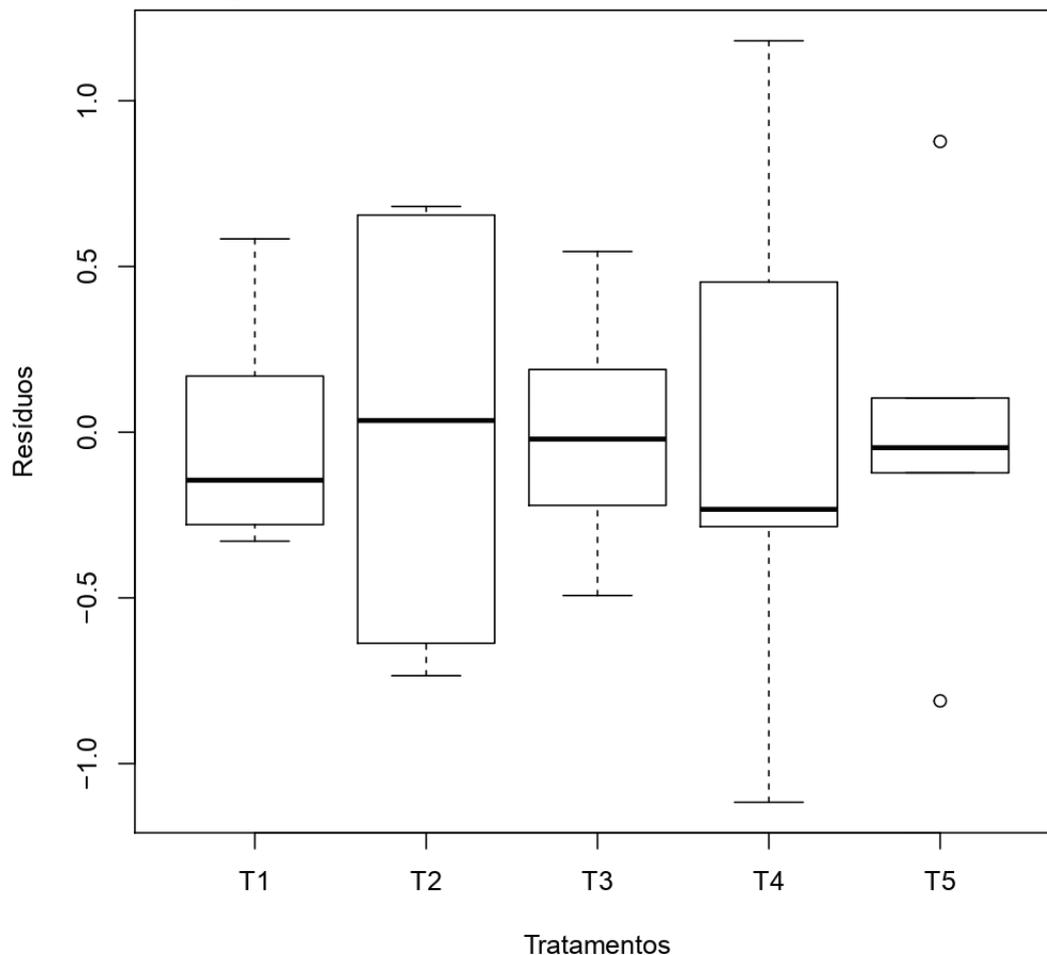
## Homogeneidade de variâncias

Para verificar a homogeneidade das variâncias, precisa-se dos resíduos da ANOVA e, a partir deles, fazer a análise gráfica e aplicar um teste estatístico.

Graficamente visualiza-se os resíduos da ANOVA versus os tratamentos. Os resíduos devem distribuir-se ao redor do zero com variação constante ao longo dos tratamentos, porém aceitar ou rejeitar a igualdade das variâncias pode ser um pouco arriscado.

```
aovdql <- aov(diam ~ linha + coluna + trat)
plot(trat, aovdql$residuals, ylab="Resíduos", xlab="Tratamentos")
```

Figura 8. Box-plot dos resíduos da ANOVA por tratamento (DQL).



Fonte: Embrapa Clima Temperado - Pelotas/RS, 2021.

O melhor é proceder ao teste estatístico de Bartlett considerando as mesmas hipóteses dos delineamentos anteriores para a comparação das variâncias:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \dots = \sigma_5^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \dots \neq \sigma_5^2 \end{cases} \quad (4.2)$$

```
bartlett.test(residuals(aov(diam ~ linha + coluna, data=dados_dql))
             ~trat, data=dados_dql)

##
## Bartlett test of homogeneity of variances
##
## data: residuals(aov(diam ~ linha + coluna, data = dados_dql)) by trat ##
Bartlett's K-squared = 3.4826, df = 4, p-value = 0.4805
```

Como o p-valor encontrado (0,4805) é bem superior ao nível  $\alpha$  de 0,05, a decisão que se deve tomar é não rejeitar a hipótese de igualdade entre as variâncias.

### Normalidade dos resíduos

As hipóteses consideradas nesse delineamento são as mesmas dos delineamentos anteriores:

$$\begin{cases} H_0 : \text{os resíduos apresentam distribuição normal} \\ H_1 : \text{os resíduos não apresentam distribuição normal} \end{cases} \quad (4.3)$$

Para realizar o teste de Shapiro-Wilk diretamente, pode-se utilizar os resíduos que já foram calculados na análise de variância através do comando `aov` anteriormente mencionado. Nesse exemplo, associa-se o nome `aovdql` à ANOVA, para então utilizar o comando `shapiro.test`:

```
shapiro.test(aovdql$res)

##
## Shapiro-Wilk normality test
##
## data:  aovdql$res
## W = 0.98603, p-value = 0.9735
```

De acordo com o p-valor encontrado de 0,9735, a decisão a tomar é de não rejeitar a hipótese nula. Caso o comando `aov` não tivesse sido executado anteriormente, outra maneira de realizar o teste de normalidade seria calcular diretamente os resíduos da seguinte maneira:

```
shapiro.test(residuals(aov(diam ~ linha + coluna + trat,
                          data=dados_dql)))

##
## Shapiro-Wilk normality test
##
## data:  residuals(aov(diam ~ linha + coluna + trat, data = dados_dql))
## W = 0.98603, p-value = 0.9735
```

### Teste de aditividade do controle local

Foi visto anteriormente no delineamento em blocos completos o teste de aditividade de blocos. Naquela situação, buscava-se identificar a presença de aditividade para garantir que o efeito dos tratamentos não fosse influenciado pelo efeito do controle local. Agora a situação é semelhante: busca-se comprovar que os tratamentos não são influenciados pelo controle local formado pelas linhas e pelas colunas do delineamento em quadrado latino.

Para verificar a presença de aditividade no controle local, aplica-se a ANOVA considerando as três fontes de variação: tratamentos, linhas e colunas, e assim obter os valores quadráticos dos preditos com base no modelo obtido.

```
mod1 <- aov(diam ~ linha + coluna + trat)
pred2 <- (predict(mod1))^2
```

Em seguida ajusta-se um novo modelo (`mod2`) contendo esses valores e faz-se a comparação dos dois modelos pelo teste F da ANOVA, utilizando o comando `anova`.

```

mod2 <- update(mod1, .~. + pred2)
summary(mod2)

##              Df Sum Sq Mean Sq F value Pr(>F)
## linha         4  2.646  0.6615   0.981  0.457
## coluna        4  0.677  0.1693   0.251  0.903
## trat          4  0.732  0.1829   0.271  0.890
## pred2         1  0.064  0.0637   0.095  0.764
## Residuals    11  7.415  0.6741

summary(mod1)

##              Df Sum Sq Mean Sq F value Pr(>F)
## linha         4  2.646  0.6615   1.061  0.417
## coluna        4  0.677  0.1693   0.272  0.891
## trat          4  0.732  0.1829   0.294  0.877
## Residuals    12  7.479  0.6232

anova(mod1, mod2, test="F")

## Analysis of Variance Table
##
## Model 1: diam ~ linha + coluna + trat
## Model 2: diam ~ linha + coluna + trat + pred2
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      12 7.4786
## 2      11 7.4149  1  0.063728 0.0945 0.7642

```

De acordo o teste realizado, o p-valor encontrado é igual a 0,7642 (maior que o nível de significância de 0,05), portanto não se rejeita a hipótese nula, ou seja, os tratamentos não são influenciados pelo controle local.

### Análise de variância - ANOVA

Depois que os pressupostos para o delineamento proposto foram atendidos, prossegue-se com a análise dos dados através da ANOVA. Como o comando `aov` já foi utilizado durante a verificação dos pressupostos, seus resultados estão armazenados em `aovdql`. Assim, basta utilizá-lo para obter a análise de variância aplicando o comando `summary`.

```

summary(aovdql)

##              Df Sum Sq Mean Sq F value Pr(>F)
## linha         4  2.646  0.6615   1.061  0.417
## coluna        4  0.677  0.1693   0.272  0.891
## trat          4  0.732  0.1829   0.294  0.877
## Residuals    12  7.479  0.6232

```

Nesse caso, a ANOVA possui três fontes de variação: linhas, colunas e tratamentos. As duas primeiras referem-se ao controle local, e sua significância no teste não são o objeto principal de estudo, logo não devem ser abordadas na discussão. Já a fonte de variação referente aos tratamentos deve ser avaliada e discutida com base no resultado do teste F e do respectivo p-valor.

Em relação aos tratamentos, encontra-se um p-valor maior que 0,05, o que sugere não haver diferença significativa entre eles, já que a hipótese nula não deve ser rejeitada. Caso esse efeito fosse significativo, dever-se-ia utilizar o comando *HSD.test* para proceder ao teste de Tukey visando verificar quais tratamentos diferem entre si, de forma análoga aos delineamentos anteriores.

## Considerações finais

Como pode ser observado nesse trabalho, é possível realizar o estudo dos pressupostos da ANOVA utilizando-se comandos simples, disponíveis em pacotes do *software R*, sem exigir um conhecimento aprofundado do seu funcionamento.

Fica evidente que, com a análise dos pressupostos da ANOVA, o pesquisador reúne informações essenciais para proceder à análise de variância de forma segura e eficiente, tornando a tomada de decisão muito mais confiável.

Também destaca-se a importância de se escolher corretamente o delineamento experimental a ser utilizado pois, de acordo com as condições de cada experimento, pode-se utilizar controles locais para otimizar o resultado dos tratamentos em estudo.

Por fim, observa-se que, para cada delineamento experimental, existe uma maneira adequada de verificar os pressupostos da ANOVA e proceder à análise de forma simples e rápida, sem a necessidade da utilização de comandos complexos no *software R*.

## Referências

BOX, G. E. P. and COX, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*. Vol. 26, No. 2, pp. 211-252.

LEVENE, H. (1960). Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

LILLIEFORS, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, Vol. 62, pp. 399-402.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

SHAPIRO, S. S. and WILK, M. B. (1965). An analysis test for normality (complete samples). *Biometrika*. Vol. 52, No. 3 and 4. Oxford University Press, pp. 591-611.

SNEDECOR, G.W. and COCHRAN, W. G. (1967). *Statistical methods*. 6th Edition, Ames, Iowa. The Iowa State University.

SNEDECOR, G.W. and COCHRAN, W. G. (1983). *Statistical Methods*. 6th Edition, Oxford and IBH, New Delhi.

SNEDECOR, G. W. and COCHRAN, W. G. (1989). *Statistical Methods*, 8th Edition, Iowa State University.

TUKEY, J.W. (1953). The problem of multiple comparisons. Mimeographs Princeton University, Princeton, N.J.

VENABLES, W. N. and RIPLEY, B. D. (2002). *Modern Applied Statistics with S*. 4th Edition. Springer, New York.

**Embrapa**  

---

**Clima Temperado**

