

# Análise de dados com o ambiente





Governador do Estado  
Carlos Moisés da Silva

Secretário de Estado da Agricultura e da Pesca  
Ricardo de Gouvêa

Presidente da Epagri  
Edilene Steinwandter

Diretores

Célio Haverroth  
Desenvolvimento Institucional

Giovani Canola Teixeira  
Administração e Finanças

Humberto Bicca Neto  
Extensão Rural e Pesqueira

Vagner Miranda Portes  
Ciência, Tecnologia e Inovação



ISSN 1413/960X (Impresso)

ISSN 2674/9513 (*On-line*)

## BOLETIM TÉCNICO Nº 178

Análise de dados com o ambiente R

Cristiano Nunes Nesi  
Simone Silmara Werner



EMPRESA DE PESQUISA AGROPECUÁRIA E EXTENSÃO RURAL DE SANTA CATARINA

FLORIANÓPOLIS

2020

Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri)  
Rodovia Admar Gonzaga, 1347, Itacorubi, Caixa Postal 502  
88034-901 Florianópolis, SC, Brasil  
Fone: (48) 3665-5000, fax: (48) 3665-5010  
Site: [www.epagri.sc.gov.br](http://www.epagri.sc.gov.br)

Editado pelo Departamento Estadual de Marketing e Comunicação (DEMC)/Epagri

Editoração técnica: Lucia Morais Kinceler

Revisão textual: Laertes Rebelo

É permitida a reprodução parcial deste trabalho desde que a fonte seja citada.

### Ficha catalográfica

NESI, C.N.; WERNER, S.S. **Análise de dados com o ambiente R**. Florianópolis: Epagri, 2020.124p. (Epagri. Boletim Técnico, 178)

Análise de variância; Regressão; Pressuposições; Programa R

ISSN 1413/960X (Impresso)

ISSN 2674/9513 (*On-line*)

## AUTORES

**Cristiano Nunes Nesi**

Engenheiro-agrônomo, Dr. em Agronomia

Epagri, Centro de Pesquisa para Agricultura Familiar

Servidão Ferdinando Tuset, s/n

89801 970 - Chapecó, SC

cristiano@epagri.sc.gov.br

**Simone Silmara Werner**

Matemática, Dra. em Estatística e Experimentação Agronômica

Epagri, Estação Experimental de Lages

Rua João José Godinho, S/N - Morro do Posto

88502 970 - Lages, SC

simonewerner@epagri.sc.gov.br

## APRESENTAÇÃO

A pesquisa experimental é amplamente realizada em universidades e instituições de pesquisa do Brasil. Contudo, pesquisadores costumam utilizar inadequadamente testes estatísticos pela não verificação das pressuposições necessárias, acarretando dificuldades tanto na análise dos dados coletados, quanto na interpretação de resultados obtidos, podendo assim originar conclusões errôneas. A escolha do delineamento experimental adequado e o emprego correto do modelo estatístico são de fundamental importância para a pesquisa na estimação correta do erro experimental. Entretanto, observa-se que muitos resultados de experimentos são publicados sem que as pressuposições da análise de variância sejam mencionadas e, provavelmente, adequadamente verificadas.

Neste contexto, o objetivo deste material é disponibilizar instruções teóricas e um material que auxiliem na análise dos dados produzidos em pesquisas experimentais. O material aborda a análise de experimentos nos delineamentos inteiramente casualizados, blocos casualizados e quadrado latino, além dos esquemas fatoriais e em parcelas subdivididas. Em todos os casos, são apresentados os códigos em ambiente R para verificar a homogeneidade de variâncias, a distribuição normal dos resíduos e a aditividade de blocos quando aplicável. As pressuposições são verificadas com os testes clássicos disponíveis na literatura, além da inspeção visual em gráficos adequados, o que complementa e auxilia de forma esclarecedora o diagnóstico.

O Boletim Técnico é rico em exemplos e detalha os casos para auxiliar o leitor na aplicação prática dos exemplos e análises de forma que a inferência realizada seja válida.

A Diretoria Executiva

## SUMÁRIO

<b>Introdução</b>	<b>8</b>
<b>1 Delineamentos experimentais</b>	<b>9</b>
1.1 Princípios básicos da experimentação . . . . .	9
1.1.1 Repetição . . . . .	9
1.1.2 Casualização . . . . .	10
1.1.3 Controle local . . . . .	11
<b>2 O programa R</b>	<b>12</b>
2.1 O software RStudio . . . . .	12
2.2 O Tinn-R . . . . .	13
2.3 Pacotes no <b>R</b> . . . . .	13
2.4 Importação de dados . . . . .	13
2.5 Listar e remover objetos no R . . . . .	14
2.6 Como o R reconhece valores perdidos . . . . .	15
<b>3 Exemplos de aplicação</b>	<b>16</b>
3.1 Delineamento Inteiramente Casualizado . . . . .	16
3.1.1 Importando os dados . . . . .	16
3.1.2 Análise exploratória dos dados . . . . .	17
3.1.3 Pressuposições do modelo estatístico . . . . .	19
3.1.4 Análise de variância . . . . .	24
3.1.5 Testes de comparações múltiplas . . . . .	24
3.1.6 Análise com o pacote <b>ExpDes.pt</b> . . . . .	29
3.2 Delineamento em Blocos Casualizados . . . . .	31
3.3 Importando os dados . . . . .	31
3.3.1 Análise exploratória dos dados . . . . .	32
3.3.2 Verificação das pressuposições do modelo estatístico . . . . .	33
3.3.3 Análise de variância . . . . .	39
3.3.4 Teste de comparações múltiplas . . . . .	39

3.3.5	Análise de com o pacote <b>ExpDes.pt</b> . . . . .	44
3.4	Delineamento em Quadrado Latino . . . . .	46
3.4.1	Importando os dados . . . . .	46
3.4.2	Análise exploratória dos dados . . . . .	47
3.4.3	Verificação das pressuposições do modelo estatístico . . . . .	48
3.4.4	Análise de variância . . . . .	54
3.4.5	Teste de comparações múltiplas . . . . .	54
3.4.6	Análise de com o pacote <b>ExpDes.pt</b> . . . . .	57
3.5	Esquema Fatorial . . . . .	59
3.6	Importando os dados . . . . .	59
3.6.1	Análise exploratória dos dados . . . . .	60
3.6.2	Verificação das pressuposições do modelo estatístico . . . . .	62
3.6.3	Análise de variância . . . . .	67
3.6.4	Análise com o pacote <b>ExpDes.pt</b> . . . . .	71
3.7	Parcela Subdividida . . . . .	75
3.7.1	Importando os dados . . . . .	75
3.7.2	Análise exploratória dos dados . . . . .	75
3.7.3	Verificação das pressuposições do modelo estatístico . . . . .	77
3.7.4	Análise de variância . . . . .	90
3.7.5	Análise de variância com o pacote <b>ExpDes</b> . . . . .	92
3.8	Análise de Regressão . . . . .	96
3.8.1	Importando os dados . . . . .	96
3.8.2	Análise exploratória dos dados . . . . .	96
3.8.3	Pressuposições do modelo estatístico . . . . .	98
3.8.4	Análise de variância . . . . .	102
3.8.5	Modelos polinomiais . . . . .	103
3.8.6	Reparametrização do modelo polinomial de grau 2 . . . . .	110
3.9	Modelo de Mitscherlich . . . . .	117



## Introdução

Embora a importância da escolha do delineamento experimental e das análises estatísticas sejam reconhecidas pela comunidade científica, frequentemente os pesquisadores utilizam inadequadamente testes estatísticos devido à não verificação das pressuposições necessárias. Em consequência, surgem dificuldades tanto na análise dos dados coletados, quanto na interpretação de resultados obtidos, podendo assim originar conclusões errôneas. A escolha do delineamento experimental adequado, bem como o emprego correto do modelo estatístico, é de fundamental importância para a pesquisa (BERTOLDO et al., 2007).

Cardellino & Siewerdt (1992) realizaram um levantamento dos trabalhos publicados entre 1982 e 1989 na revista da sociedade brasileira de Zootecnia, quanto ao uso dos testes para comparação de médias e verificaram que, entre os trabalhos considerados, 64,2% utilizavam os testes de forma incorreta. Trabalho semelhante foi desenvolvido por Santos et al. (1998) para avaliar o emprego dos testes de comparação de médias na revista Pesquisa Agropecuária Brasileira (PAB) no período de 1980 a 1994, e os autores constataram que os testes foram utilizados de forma adequada em apenas 57% dos casos.

Bertoldo et al. (2008) revisaram 226 trabalhos científicos publicados na revista Ciência Rural de 2002 a 2006, na área de Fitotecnia, sendo utilizados 148 trabalhos para discussão. Os autores concluíram que a maioria dos trabalhos que estudaram mais de um fator foram classificados como incorretos (72%) devido ao abuso dos testes de comparações de médias.

Montanhini Neto e Ostrensky (2013) revisaram 307 artigos publicados entre 2000 e 2010 na Revista Archives of Veterinary Science e constataram que 34% dos artigos apresentavam conclusões sem nenhum apoio estatístico; 34% estavam embasados em métodos estatísticos inadequados à base de dados gerados; e apenas 32% apresentavam conclusões fundamentadas em métodos estatísticos condizentes com a estrutura dos dados analisados.

Além do conhecimento das técnicas de planejamento e análise de experimentos, é fundamental o domínio de um software que possibilite a utilização das técnicas adequadas. De acordo com Furtado (2013) o programa R é um dos melhores softwares de análise estatística existentes na atualidade, competindo com o SAS (SAS Institute) nas mesmas condições.

O objetivo deste texto é disponibilizar aos pesquisadores da Epagri e demais interessados um material que auxilie na análise dos dados produzidos em suas pesquisas experimentais com aplicações da análise de variância clássica após a verificação de todas as pressuposições do modelo estatístico de forma que a inferência realizada seja válida.

# 1 Delineamentos experimentais

O delineamento experimental é o plano utilizado na experimentação e implica a forma como os tratamentos serão designados às unidades experimentais e um amplo entendimento das análises a serem feitas.

Um delineamento experimental é composto por duas estruturas básicas:

- Estrutura de tratamentos: consiste em um conjunto de tratamentos de interesse e pode ser formado pelos níveis de um fator ou pela combinação dos níveis de dois ou mais fatores;
- Estrutura de unidades: está associada à classificação das unidades experimentais segundo características dessas unidades. Consiste em agrupar unidades experimentais homogêneas entre si ou formar blocos. Envolve o agrupamento de unidades experimentais tal que as condições sobre as quais os tratamentos são observados sejam as mais homogêneas possíveis.

Assim, a estrutura dos fatores de tratamentos, a estrutura dos fatores de unidade e a relação entre essas duas estruturas, determinadas pela casualização, constituem a estrutura do experimento ou o delineamento experimental.

Para que a metodologia estatística possa ser aplicada aos resultados de um experimento, é necessário que os princípios básicos da experimentação estejam presentes.

## 1.1 Princípios básicos da experimentação

### 1.1.1 Repetição

A ideia em experimentação é comparar grupos, não apenas unidades. A repetição do experimento básico tem por finalidade propiciar a estimativa do erro experimental e possibilitar a estimação das médias dos tratamentos (procurar confirmar a resposta que um indivíduo dá a um determinado tratamento). Deve existir um número mínimo de repetições para que se possam tirar conclusões válidas a respeito do fenômeno em estudo. Não existe um número adequado de repetições que possa servir indiferentemente a qualquer experimento. Há um número mais adequado de repetições para cada experimento – esse número varia de um experimento para outro e, portanto, precisa ser calculado para cada um deles.

É sempre desejável que os experimentos tenham um grande número de repetições, pois, quanto maior é o número de repetições, mais fidedigna será a estimativa do erro experimental. Portanto, haverá maior segurança na aplicação dos testes estatísticos e maior confiabilidade nos resultados.

A variabilidade dos dados experimentais rege a escolha do número de repetições mais adequado. Caso o experimento ainda não tenha sido realizado, a variação pode ser conhecida por meio da execução de um experimento piloto e da verificação preliminar da variabilidade dos dados experimentais.

O experimento piloto é aquele que se faz previamente à realização da pesquisa e visa testar o método de trabalho e os processos técnicos envolvidos na execução dos experimentos. Em geral o piloto segue o mesmo plano de trabalho que orienta a investigação como um todo, entretanto o número de repetições é pequeno, para não tornar o experimento piloto muito trabalhoso ou muito demorado. É bastante comum a escolha de 3 repetições como um bom número para começar, pois o número 3 evita a invariabilidade do número 1, foge do perigo da repetição coincidente representado pelo número 2, e já apresenta alguma variação, a qual no mais das vezes já é suficiente para estimar a variabilidade determinada pelos fatores de variação e pelas próprias repetições.

A verificação preliminar da variabilidade dos dados experimentais é feita após a execução do experimento piloto, sem qualquer preocupação com o tipo de distribuição dos dados, e seja lá qual for o número de repetições nele fixado. Realizada a análise de variância preliminar, faz-se o teste estatístico para determinar, especificamente para o experimento, qual seria o número mais adequado de repetições. Esse tipo de teste é iterativo, pois partindo do número inicial de repetições do experimento piloto, ele calcula um novo número de repetições, e volta a introduzir no teste esse número calculado, recalculando tudo e achando outro número de repetições. Assim sucessivamente, vai recalculando até que o número de entrada do teste seja igual ao número de saída. Quando essa igualdade ocorre, o teste é dado por terminado, e esse número de repetições constitui o número mais adequado de repetições para aquele experimento. O teste sugere que, para aquela variação detectada pelo experimento piloto, é necessário aquele número mínimo de repetições, para que se possam perceber diferenças estatisticamente significativas entre os fatores de variação estudados.

Na prática, o tamanho da amostra é determinado mais por considerações a respeito dos custos do que por técnicas estatísticas. Entretanto, deve-se considerar o que é usual na área, a experiência do pesquisador, a grandeza do erro desejado e da variabilidade do material. Experimentos muito pequenos não permitem conclusão e muito grandes, mas mal conduzidos, dão a ilusão de conter uma verdade que não contêm. Ao planejar um experimento, deve-se ter certeza de que se conseguirá detectar uma diferença real entre tratamentos da ordem de grandeza que se esteja interessado. Caso a probabilidade de conseguir esse objetivo com o número de repetições que se pode utilizar seja pequena, é preferível deixar o experimento para outra ocasião, em que se tenha recursos suficientes para realizá-lo com o número de repetições adequado.

### 1.1.2 Casualização

Os tratamentos devem ser designados às unidades experimentais por puro e simples sorteio, para evitar que um tratamento seja sistematicamente favorecido por qualquer fator externo, e tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer das unidades experimentais. A escolha da técnica de casualização fica a critério do pesquisador.

Deve-se utilizar a casualização mesmo que o pesquisador ache que todos os efeitos estão cuidadosamente controlados. O princípio da casualização é uma das maiores contribuições

dos estatísticos à ciência experimental, pois só a casualização garante que as unidades com características diferentes tenham igual probabilidade de serem designadas para os grupos.

Não existem alternativas válidas para a casualização. O pesquisador que “escolhe” as unidades por critério próprio introduz tendenciosidade aos resultados, embora muitas vezes seja possível fazer o sorteio mantendo as restrições que o pesquisador considera necessárias. Sua função é evitar possíveis induções nas conclusões da pesquisa e garantir a independência dos dados, tornando as estimativas das médias dos tratamentos e do erro não tendenciosas. Os princípios básicos da repetição e da casualização são obrigatórios em qualquer planejamento experimental. A metodologia estatística somente pode ser usada quando esses princípios foram considerados para que se tenha uma estimativa válida para o erro experimental.

### 1.1.3 Controle local

Refere-se a restringir a casualização dos tratamentos a grupos de unidades experimentais com pouca variabilidade entre si, tornando o delineamento experimental mais eficiente, pela redução do erro experimental. Nesse caso, pode existir uma variação acentuada de um grupo para outro de parcelas, no entanto, dentro de cada bloco as parcelas devem ser homogêneas. A cada grupo de parcelas homogêneas denomina-se BLOCO. Este princípio é aplicado quando se sabe antecipadamente da existência de variação no ambiente experimental, não podendo correr o risco de efetuar uma casualização completa dos tratamentos.

À medida que se aumenta o controle local diminui-se o número de graus de liberdade do resíduo de uma análise de variância. Como este número é um indicador da precisão da análise, só se deve fazer o controle local quando realmente for necessário. Às vezes o pesquisador não sabe se deve ou não organizar seu experimento em blocos. Por exemplo, para comparar a produtividade de diversas variedades de uma hortaliça o pesquisador pode ter dúvidas se deve ou não usar blocos. Ora, se os canteiros constituem uma causa de variação (há algum tipo de diferença entre os canteiros), a comparação de tratamentos melhora se o experimento for casualizado em blocos. No entanto, se os canteiros não constituem causa de variação e o experimento for casualizado em blocos, perderá em eficiência.

## 2 O programa R

O uso de softwares na análise de dados é de grande importância, desde o desenvolvimento e aplicação de métodos de análise até a interpretação dos resultados. Dentre os softwares de domínio público, o ambiente **R** vem ganhando destaque, adeptos e colaboradores. O **R** foi criado como um projeto de pesquisa pelo neozelandês **Ross Ihaka** e pelo canadense **Robert Gentleman**, e nos dias de hoje está sob constante desenvolvimento por um grupo chamado **R Core Team**. O **R** é uma linguagem e ambiente para computação estatística e elaboração de gráficos, com uma ampla variedade de métodos estatísticos e técnicas gráficas. O programa **R** é gratuito e de código aberto, o que propicia excelente ambiente para análises estatísticas e com recursos gráficos de alta qualidade. Detalhes sobre o projeto, colaboradores, a documentação e diversas outras informações podem ser encontradas na página oficial do projeto<sup>2</sup>.

A ideia deste material não é apresentar um material didático sobre o programa **R** ou sobre teoria e aplicação estatística, mas servir como guia no treinamento básico sobre algumas funcionalidades do ambiente **R**.

*"A pergunta certa sobre uma análise em **R** não é se é possível fazê-la, e sim como fazê-la."*

PhD. Paulo Justiniano Ribeiro Junior<sup>3</sup>.

### 2.1 O software RStudio

Para operar o **R**, na forma usual, é necessário conhecer e digitar comandos. Alguns usuários acostumados com outros programas notarão de início a falta de "menus". Na medida que utilizam o programa, os usuários (ou boa parte deles) tendem a preferir o mecanismo de comandos pois é mais flexível e com mais recursos. O **R** possui um *prompt* aguardando os comandos do usuário no Console do **R**. Alguns usuários não têm tanta facilidade em trabalhar com janelas do *prompt*, especialmente quando é necessário utilizar diversas linhas de comando como numa rotina de análises. Neste sentido, o RStudio<sup>4</sup> oferece uma interface de usuário mais completa para o **R**. Trata-se de um *software* que oferece a integração entre o **R** e um editor de textos voltado para a edição de comandos.

---

<sup>2</sup><http://www.r-project.org/contributors.html>

<sup>3</sup>Professor no departamento de Estatística na UFPR. Autor, em conjunto com Peter J. Diggle, do pacote **geoR**. Livre versão de *"This is R. There is no if. Only how."* - Simon 'Yoda' Blomberg, R-help (April 2005)

<sup>4</sup><http://www.rstudio.com/about/>

## 2.2 O Tinn-R

O Tinn-R<sup>4</sup> é um editor/processador de texto ASCII/UNICODE genérico para o sistema operacional Windows, muito bem integrado ao ambiente estatístico e computacional **R** para o qual possui características de Interface Gráfica (GUI) e Ambiente de Desenvolvimento Integrado (IDE). É um projeto registrado sob a Licença Pública Geral GPL, sendo portanto software livre, cujo objetivo é facilitar o aprendizado e o uso do potencial do **R**, que foi escrito e é mantido pela equipe do Tinn-R, coordenada pelo professor José Cláudio Faria<sup>5</sup>, professor da Universidade Estadual de Santa Cruz – UESC-BA, Brasil.

## 2.3 Pacotes no R

As capacidades do **R** são estendidas através de pacotes criados pelo usuário, que permitem técnicas estatísticas especializadas, dispositivos gráficos, capacidades de importação/exportação, ferramentas de relatórios, etc. Um conjunto básico de pacotes é incluído com a instalação do **R**, com mais de 9200 pacotes adicionais.

Para realização das análises apresentadas nesse material foram utilizados os pacotes **openxlsx** (Walker, 2015), **car** (Fox & Weisberg, 2011), **MASS** (Venables & Ripley, 2002), **agricolae** (Menduru, 2016), **ExpDes** (Ferreira et al., 2013), **factorplot** (Armstrong, 2016), **dae** (Brien, 2011).

O pacote **ExpDes** desenvolvido por Ferreira et al. (2013) é destinado a análise de delineamentos experimentais simples (DIC, DBC e DQL), experimentos em esquema de fatorial duplo (em DIC e DBC), experimentos em esquema de parcelas subdivididas (em DIC e DBC), experimentos em esquema de fatorial duplo com um tratamento adicional (em DIC e DBC), experimentos em esquema de fatorial triplo (em DIC e DBC) e experimentos em esquema de fatorial triplo com um tratamento adicional (em DIC e DBC); realizando a análise de variância e comparação de médias pelo ajuste de modelos de regressão até o terceiro grau (tratamentos quantitativos) ou por testes de comparação múltipla.

O **agricolae**, desenvolvido por Felipe de Mendiburu, contém funcionalidades para a análise estatística de dados provenientes de delineamentos experimentais, especialmente experimentos de campo em agricultura e melhoramento de plantas.

## 2.4 Importação de dados

O formato de tabela de dados para análise estatística geralmente é o seguinte: cada **linha é uma observação**, cada **coluna é uma variável** ou atributo que foi mensurado em cada observação.

O **R** pode importar arquivos em vários formatos. Os principais serão descritos a seguir:

1. Arquivos de texto (**.txt**)

---

<sup>4</sup><http://nbcgib.uesc.br/lec/software/editores/tinn-r/pt>

<sup>5</sup><http://nbcgib.uesc.br/lec/professores/jcfaria/pag-pessoal>

Para importar um arquivo que previamente foi salvo no formato **.txt** utiliza-se o comando **read.table**, especificando o nome do arquivo nesta extensão.

```
> read.table("dados.txt", dec=",", header=T)
```

As opções significam:

**dec:** informa o separador de decimais dos dados;

**header:** informa que primeira linha do arquivo de dados contém o nome das variáveis.

Existem outras opções que podem ser utilizadas devendo-se para isso consultar o help da função.

## 2. Arquivos separados por vírgula (**.csv**)

Para importar um arquivo salvo no formato **.csv** utiliza-se o comando **read.csv**, para arquivos separados por vírgula, e o comando **read.csv2**, para aqueles separados por ponto e vírgula.

```
> read.csv("dados.csv")
> read.csv2("dados.csv")
```

Existem outras opções que podem ser utilizadas devendo-se para isso consultar o help da função.

## 3. Arquivos em formato Excel (**.xlsx**)

Para importar um arquivo salvo no formato **.xlsx** é necessário instalar e requerer o pacote **openxlsx**. Após isso, utiliza-se o comando **read.xlsx**, especificando-se o nome e extensão do arquivo, além da aba da planilha que contém os dados. A aba da planilha pode ser especificada pelo seu nome ou pela ordem em que aparece na planilha.

```
> install.packages("openxlsx", dep=T)
> require(openxlsx)
> read.xlsx("dados.xlsx", sheet="ex1")
> read.xlsx("dados.xlsx", sheet=1)
```

Uma forma de encontrar o arquivo é utilizar a função **file.choose()**, que abrirá uma janela para a busca do arquivo desejado. Pode ser utilizado com qualquer uma das funções de importação anteriormente descritas.

## 2.5 Listar e remover objetos no R

Assim que os arquivos são carregados no ambiente R, eles ficam ativos e ocupam espaço na memória do computador. Para verificar quais objetos (arquivos) estão ativos, utiliza-se a função **objects()** ou a função equivalente **ls()**. Caso se deseje remover objetos ativos, utiliza-se a função **rm()**, com o nome do objeto dentro dos parênteses. Para remover todos os objetos, utiliza-se a função **rm(list=ls(all=TRUE))**.

## 2.6 Como o R reconhece valores perdidos

Para sinalizar ao **R** que um determinado valor perdido, utiliza-se o símbolo **NA** ("not available"). Existe ainda um segundo símbolo que pode ocorrer no caso de resultados de indeterminações matemáticas: o **NaN** ("not a number"). O símbolo **Inf** é utilizado para representar resultados cujo limite tende ao infinito.

Quando os dados são importados com os comandos **read.table**, **read.csv** e **read.csv2**, pode-se definir os valores perdidos no momento da importação dos dados, especificando-se o carácter utilizado para indentificar dado perdido, por meio da opção **na.strings=' '**, como no exemplo a seguir, em que utilizou-se (-):

```
> #read.table("dados.txt", dec=",", header=T, na.strings="-")
```

O pacote **openxlsx** utilizado para importar arquivos do tipo **.xlsx** converte **células vazias** para **NA**, assim na utilização desse pacote os dados perdidos devem ser digitados como células vazias.



## 3 Exemplos de aplicação

### 3.1 Delineamento Inteiramente Casualizado

Um pesquisador deseja comparar o rendimento de grãos e o peso de mil sementes de feijão em 10 variedades. Para tanto, utilizou uma área experimental homogênea e realizou um experimento no delineamento inteiramente casualizado (DIC) com 10 tratamentos e quatro repetições.

Tabela 1: Rendimento de grãos (kg/ha) e peso de mil sementes (g) de feijão em experimento de competição de 10 variedades, realizado em Chapecó, SC

Variedade	Repetições							
	1	2	3	4	1	2	3	4
	Rendimento (kg/ha)				Peso de 1000 sementes (g)			
Soberano	369	287	577	184	189	212	194	241
Campeiro	933	1091	1160	1074	203	221	208	237
Supremo	1014	957	1139	1563	207	217	210	239
Valente	944	942	649	887	205	209	198	251
Tuiuiu	1056	1006	845	1433	207	213	213	230
Tiziu	1252	1060	1164	1113	218	214	210	243
Uirapuru	1032	976	1214	1050	212	216	217	241
Diplomata	1249	1094	1188	1238	224	231	228	238
CHP97	1701	1839	1628	1526	228	234	241	234
CHP99	1966	1334	1796	1806	230	238	236	234

Neste experimento, o interesse do pesquisador é testar a hipótese:

$$\begin{cases} H_0 : \mu_{\text{Soberano}} = \mu_{\text{Campeiro}} = \dots = \mu_{\text{CHP99}} \\ H_1 : \mu_{\text{Soberano}} \neq \mu_{\text{Campeiro}} \neq \dots \neq \mu_{\text{CHP99}} \end{cases} \quad (1)$$

Para verificar se existe diferença entre as médias dos diferentes tratamentos, deve-se primeiramente verificar se existe efeito do tratamento, o que pode ser feito por meio da análise de variância clássica, mas antes disso deve-se verificar as pressuposições do modelo clássico. A seguir está descrito como seria realizada uma análise para esse conjunto de dados com o ambiente **R**, desde a importação dos dados até a realização do teste de comparação de médias.

#### 3.1.1 Importando os dados

```
> ex1 <- read.csv("ex1.csv", dec=",", header = T)
> head(ex1)
```

	tratamento	rep	cultivar	rendimento	P1000
1	1	1	Soberano	369	189
2	1	2	Soberano	287	212
3	1	3	Soberano	577	194
4	1	4	Soberano	184	241
5	2	1	Campeiro	933	203
6	2	2	Campeiro	1091	221

### 3.1.2 Análise exploratória dos dados

Antes de iniciar a análise propriamente dita, deve-se inspecionar os dados utilizando-se uma análise exploratória. O comando **str** apresenta a estrutura dos dados e o comando **summary** fornece um resumo dos dados, com mínimo e máximo, a média e as separatrizes.

```
> str(ex1)

'data.frame':      40 obs. of  5 variables:
 $ tratamento: int  1 1 1 1 2 2 2 2 3 3 ...
 $ rep       : int  1 2 3 4 1 2 3 4 1 2 ...
 $ cultivar  : Factor w/ 10 levels "Campeiro","CHP97",...: 5 5 5 5 1 1 1 1 6 6 ...
 $ rendimento: num  369 287 577 184 933 ...
 $ P1000     : num  189 212 194 241 203 221 208 237 207 217 ...
```

```
> summary(ex1)

      tratamento      rep      cultivar      rendimento      P1000
Min.   : 1.0   Min.   :1.00   Campeiro : 4   Min.   : 184.0   Min.   :189.0
1st Qu.: 3.0   1st Qu.:1.75   CHP97    : 4   1st Qu.: 953.8   1st Qu.:210.0
Median : 5.5   Median :2.50   CHP99    : 4   Median :1092.5   Median :219.5
Mean   : 5.5   Mean   :2.50   Diplomata: 4   Mean   :1133.4   Mean   :221.8
3rd Qu.: 8.0   3rd Qu.:3.25   Soberano : 4   3rd Qu.:1272.5   3rd Qu.:234.5
Max.   :10.0   Max.   :4.00   Supremo  : 4   Max.   :1966.0   Max.   :251.0
              (Other) :16
```

Para obter a descrição dos dados de acordo com os níveis de um determinado fator, pode-se utilizar a função **tapply**. Neste caso, deve-se colocar na sequência a variável resposta, o fator e a função a ser aplicada. O comando **with** especifica o arquivo de dados a ser trabalhado.

```
> ex1.nrep <- with(ex1, tapply(rendimento, cultivar, length))
> ex1.medias <- with(ex1, tapply(rendimento, cultivar, mean))
> ex1.var <- with(ex1, tapply(rendimento, cultivar, var))
> ex1.DP <- with(ex1, tapply(rendimento, cultivar, sd))
> (resumo <- data.frame(Rep = ex1.nrep, Medias = ex1.medias,
+                      Var = ex1.var, DP=ex1.DP))
```

	Rep	Medias	Var	DP
Campeiro	4	1064.50	9068.333	95.22780
CHP97	4	1673.50	17324.333	131.62193

CHP99	4	1725.50	74187.667	272.37413
Diplomata	4	1192.25	4994.917	70.67472
Soberano	4	354.25	27780.917	166.67608
Supremo	4	1168.25	75034.250	273.92380
Tiziu	4	1147.25	6679.583	81.72872
Tuiuiu	4	1085.00	61928.667	248.85471
Uirapuru	4	1068.00	10466.667	102.30673
Valente	4	855.50	19649.667	140.17727

Uma forma gráfica que auxilia na análise exploratória é o gráfico de caixa (*box plot*), obtido com a função **boxplot**, especificando-se a variável resposta e o fator de agrupamento.

```
> par(mfrow=c(1,2),mar=c(5,4,1,1))
> with(ex1, boxplot(rendimento ~ cultivar, xlab="Variedade",
+                  ylab="Rendimento de grãos (kg/ha)",
+                  las=2, cex.axis=0.6))
> with(ex1, boxplot(rendimento ~ cultivar, xlab="Variedade",
+                  ylab="Rendimento de grãos (kg/ha)",
+                  las=2, cex.axis=0.6,
+                  col=c("red", "blue", "yellow4", "tomato")))

```

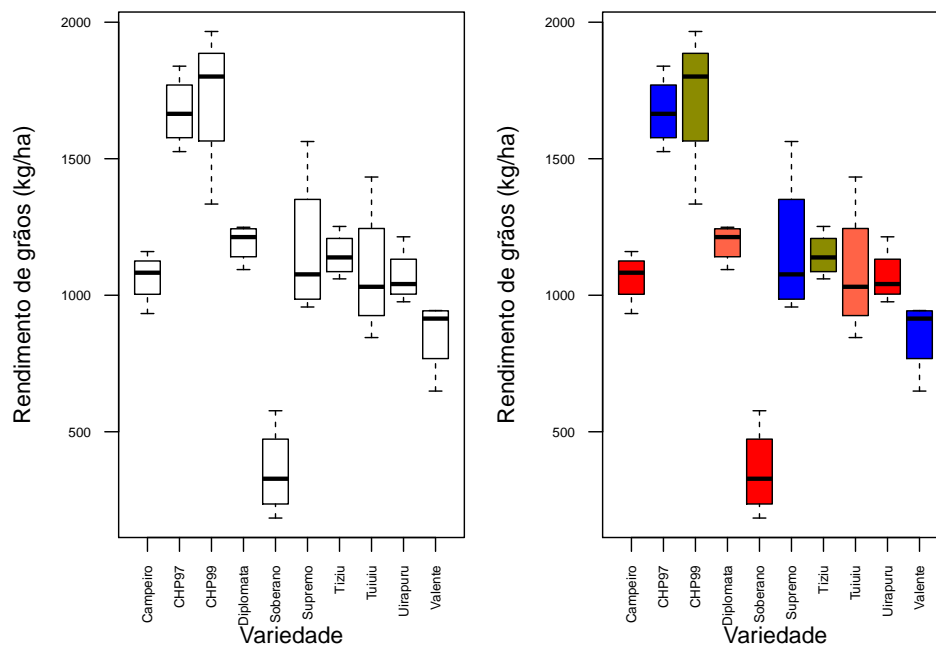


Figura 1: Gráficos de caixa (*boxplots*) dos dados de rendimento de grãos em função da variedade

As cores dos gráficos podem ser alteradas por meio do comando **col**, especificando-se as cores desejadas. Para visualizar as cores disponíveis no **R** utiliza-se a função **colors()**.

### 3.1.3 Pressuposições do modelo estatístico

Para que a análise de variância seja válida, no caso de um experimento em DIC, deve-se verificar se as pressuposições de homogeneidade de variâncias entre tratamentos e normalidade dos resíduos são satisfeitas.

#### 1. Homogeneidade de Variâncias

A hipótese a ser testada em um teste de homocedasticidade é dada por:

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

Num teste de hipóteses, sempre que o nível descritivo do teste (*p-value*) for menor que o nível de significância estabelecido ( $\alpha$ ), rejeita-se  $H_0$ . Caso contrário, não se rejeita e, neste caso, indica que não existem evidências contra a homocedasticidade.

Existem diferentes testes que podem ser utilizados para testar a hipótese de homogeneidade de variâncias. Na sequência estão apresentadas as funções que possibilitam a realização do teste de *Bartlett* (Snedecor & Cochran, 1983) e de *Levene* (Levene, 1960). Como os testes utilizam estatísticas diferentes e consequentemente distribuições diferentes os níveis descritivos de diferentes testes também serão diferentes.

```
> bartlett.test(rendimento ~ cultivar, data=ex1)
```

```
Bartlett test of homogeneity of variances
```

```
data: rendimento by cultivar
```

```
Bartlett's K-squared = 11.444, df = 9, p-value = 0.2465
```

No teste de Bartlett, como o p-valor (*p-value* = 0.2465) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que as variâncias não diferem, satisfazendo esta pressuposição.

```
> require(car)
```

```
> ex1$cultivar <- as.factor(ex1$cultivar)
```

```
> leveneTest(rendimento ~ cultivar, data=ex1)
```

```
Levene's Test for Homogeneity of Variance (center = median)
```

```
  Df F value Pr(>F)
```

```
group 9  0.5706  0.81
```

```
 30
```

No teste de Levene, como o p-valor (*p-value* = 0.8100) é maior que  $\alpha = 0.05$ , também não há evidências para rejeitar a hipótese de que as variâncias não diferem, satisfazendo esta pressuposição.

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, \sigma)$ )

A hipótese a ser testada em um teste de normalidade dos resíduos é dada por:

$$\begin{cases} H_0 : & \text{Os resíduos do modelo provêm de uma população com distribuição normal} \\ H_1 : & \text{Os resíduos do modelo não provêm de uma população com distribuição normal} \end{cases} \quad (3)$$

As três principais funções que possibilitam a extração dos resíduos de um modelo implementadas no **R** são:

- **residuals()**: Fornece os resíduos ordinários, obtidos pela diferença entre os valores observados e os ajustados.
- **rstandart()**: Fornece os resíduos padronizados por meio do quociente entre os resíduos ordinários e a estimativa do desvio padrão dos resíduos (raíz quadrada do quadrado médio do resíduo do modelo). Dessa forma espera-se que esse resíduo seja distribuído de acordo com a distribuição normal padrão.
- **rstudent()**: Pode ser interpretado de forma semelhante ao resíduo padronizado porém considera no cálculo de cada valor a variância residual obtida na ausência do mesmo (*leave-one-out measure of the error variance*).

A utilização dos resíduos padronizados ou *studentizados* facilita a identificação de possíveis pontos discrepantes, pois nesse caso valores dos resíduos maiores que 2 sugerem um desvio considerável (Kaps & Lamberson, 2004).

```
> anova <- aov(rendimento ~ cultivar, data=ex1)
> res <- rstudent(anova)
> shapiro.test(res)
```

```
Shapiro-Wilk normality test
```

```
data:  res
W = 0.96775, p-value = 0.3048
```

O mesmo comando pode ser agrupado em apenas uma linha:

```
> shapiro.test(rstudent(aov(rendimento ~ cultivar, data=ex1)))
```

```
Shapiro-Wilk normality test
```

```
data:  rstudent(aov(rendimento ~ cultivar, data = ex1))
W = 0.96775, p-value = 0.3048
```

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.6633$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, satisfazendo esta pressuposição.

```
> ks.test(rstudent(aov(rendimento ~ cultivar, data=ex1)), "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data: rstudent(aov(rendimento ~ cultivar, data = ex1))
D = 0.096972, p-value = 0.8114
alternative hypothesis: two-sided
```

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p\text{-value} = 0.8114$ ) é maior que  $\alpha = 0.05$ , também não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, satisfazendo esta pressuposição.

Uma outra forma de checar a adequação de distribuição normal é o gráfico *qqPlot* com envelope de simulação. Neste caso espera-se que todos os pontos estejam dentro dos limites definidos pelas linhas do envelope.

```
> require(car)
> qqPlot(rstudent(aov(rendimento ~ cultivar, data=ex1)),
+        pch=19, ylab="Resíduos")
```

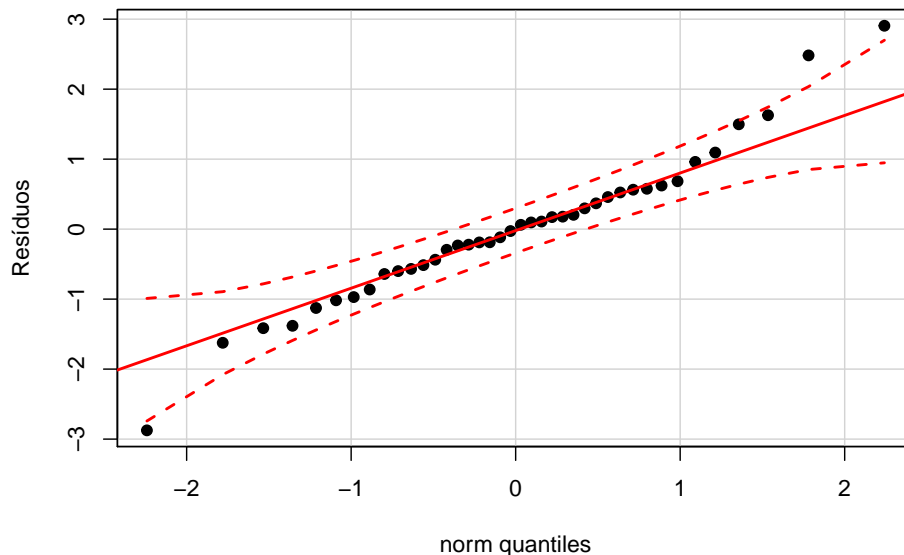


Figura 2: Gráfico qqplot para avaliar a distribuição normal dos resíduos

### 3. Transformação de Box-Cox

Quando uma ou mais pressuposições do modelo estatístico não é atendida a análise de variância clássica, com os dados na escala original, não é indicada. Entre as opções possíveis para trabalhar com os dados nesse caso está a realização de uma transformação de dados, sendo a inferência válida na escala dos dados transformados desde que esses atendam as pressuposições.

A família de transformações proposta por Box & Cox (1964) é dada por:

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

em que  $\lambda$  é o parâmetro da transformação e  $Y$  a variável resposta. Na ausência da necessidade de transformação  $\lambda = 1$ .

A função **boxcox** do pacote **MASS** (Venables e Ripley, 2002) pode ser utilizada para obtenção do perfil de verossimilhança para diferentes valores de lambda e auxilia na verificação das pressuposições do modelo:

```
> require(MASS)
> boxcox(aov(rendimento ~ cultivar, data=ex1), lambda = seq(0, 2, l = 101))
```

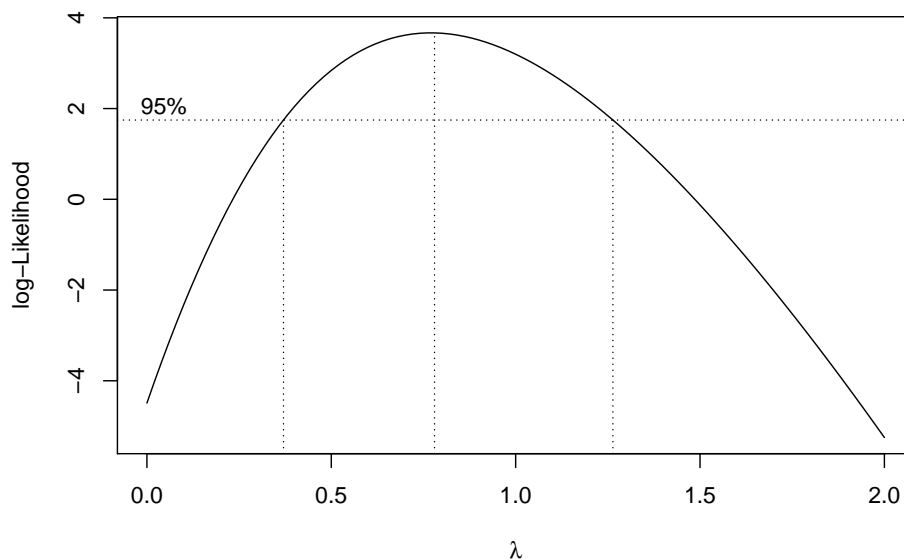


Figura 3: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Como o intervalo de confiança para o parâmetro ( $\lambda$ ) contém o valor 1, não há evidências de que os dados necessitam de transformação.

#### 4. Gráficos de diagnóstico

A seguir são apresentados comandos para construção de gráficos para o diagnóstico das pressuposições:

```
> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(rendimento ~ cultivar, data=ex1)),
+      rstudent(aov(rendimento ~ cultivar, data=ex1)),xlab="Valores ajustados",
+      ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(rendimento ~ cultivar, data=ex1)),prob=T,main=" ",
+      xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(rendimento ~ cultivar, data=ex1))), col="blue")
> require(car)
> qqPlot(rstudent(aov(rendimento ~ cultivar, data=ex1)),
+        ylab="Resíduos Studentizados", pch=19)
> boxplot(rstudent(aov(rendimento ~ cultivar, data=ex1)),
+         ylab="Resíduos Studentizados", pch=19, cex.lab=1.2,
+         cex.axis=1.2, col="turquoise")
```

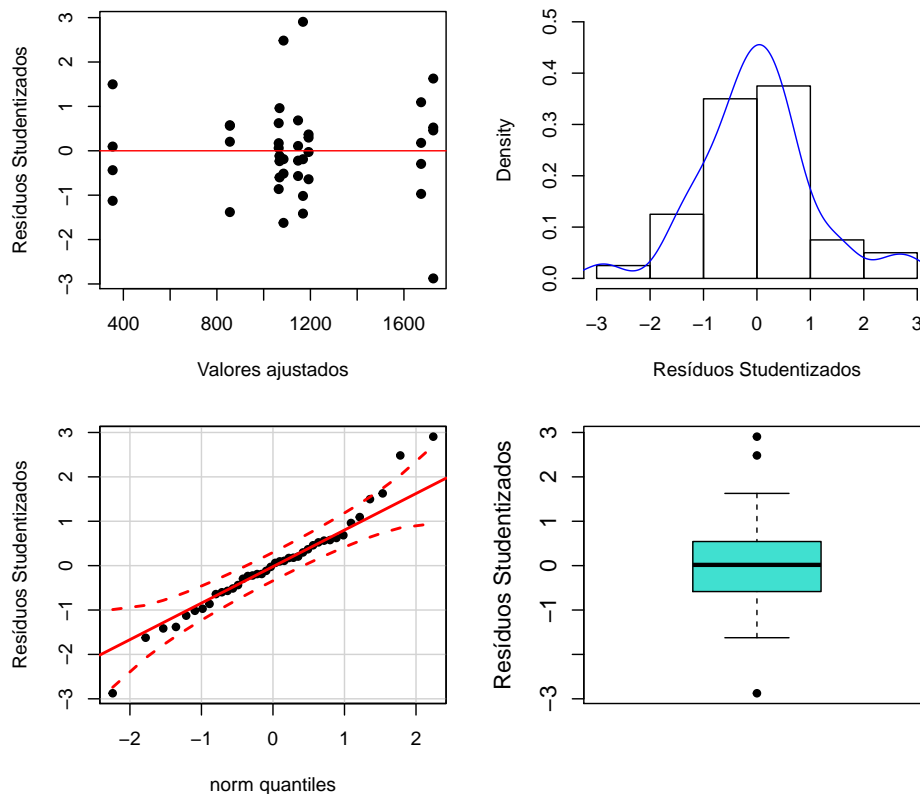


Figura 4: Gráficos de diagnósticos para verificar as pressuposições



Na figura, observa-se o atendimento as pressuposições pois nos erros não há evidências de formação de tendência, o histograma confirma uma distribuição próxima do normal, os pontos estão dentro do envoltório embora no gráfico Box-Plot há alguns *outliers* que devem ser averiguados.

Verificadas e atendidas as pressuposições do modelo, pode-se prosseguir a análise de variância.

### 3.1.4 Análise de variância

Para realizar uma análise de variância, após verificadas e atendidas as suas pressuposições, utiliza-se a função `aov` do **R** básico:

```
> ex1$cultivar<-as.factor(ex1$cultivar)
> mod1R <- aov(rendimento ~ cultivar, data=ex1)
> summary(mod1R)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cultivar	9	5371321	596813	19.43	3.31e-10 ***
Residuals	30	921345	30712		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Neste exemplo, o efeito de cultivar foi significativo ( $p\text{-valor} < 0.05$ ). Assim, rejeita-se  $H_0$  indicando que pelo menos dois cultivares diferem entre si quanto ao rendimento de grãos.

O pacote **agricolae** disponibiliza a função `cv.model` para o cálculo do coeficiente de variação experimental.

```
> require(agricolae)
> cv.model(mod1R)
```

[1] 15.46206

Como o teste F da análise de variância não identifica quais contrastes de médias de cultivares são significativos, deve-se complementar a análise com um teste de comparação de médias.

### 3.1.5 Testes de comparações múltiplas

A função **TukeyHSD** cria um conjunto de intervalos de confiança sobre a diferença entre as médias dos níveis de um fator, equivalente ao teste de Tukey.

```
> ex1.HSDR <- TukeyHSD(mod1R, ordered = F)
> ex1.HSDR
```

Tukey multiple comparisons of means  
95% family-wise confidence level

Fit: aov(formula = rendimento ~ cultivar, data = ex1)

\$cultivar

	diff	lwr	upr	p adj
CHP97-Campeiro	609.00	186.29193	1031.70807	0.0010671
CHP99-Campeiro	661.00	238.29193	1083.70807	0.0003392
Diplomata-Campeiro	127.75	-294.95807	550.45807	0.9876810
Soberano-Campeiro	-710.25	-1132.95807	-287.54193	0.0001138
Supremo-Campeiro	103.75	-318.95807	526.45807	0.9972476
Tiziu-Campeiro	82.75	-339.95807	505.45807	0.9995267
Tuiuiu-Campeiro	20.50	-402.20807	443.20807	1.0000000
Uirapuru-Campeiro	3.50	-419.20807	426.20807	1.0000000
Valente-Campeiro	-209.00	-631.70807	213.70807	0.7941167
CHP99-CHP97	52.00	-370.70807	474.70807	0.9999902
Diplomata-CHP97	-481.25	-903.95807	-58.54193	0.0160396
Soberano-CHP97	-1319.25	-1741.95807	-896.54193	0.0000000
Supremo-CHP97	-505.25	-927.95807	-82.54193	0.0098225
Tiziu-CHP97	-526.25	-948.95807	-103.54193	0.0063391
Tuiuiu-CHP97	-588.50	-1011.20807	-165.79193	0.0016697
Uirapuru-CHP97	-605.50	-1028.20807	-182.79193	0.0011521
Valente-CHP97	-818.00	-1240.70807	-395.29193	0.0000106
Diplomata-CHP99	-533.25	-955.95807	-110.54193	0.0054694
Soberano-CHP99	-1371.25	-1793.95807	-948.54193	0.0000000
Supremo-CHP99	-557.25	-979.95807	-134.54193	0.0032806
Tiziu-CHP99	-578.25	-1000.95807	-155.54193	0.0020859
Tuiuiu-CHP99	-640.50	-1063.20807	-217.79193	0.0005337
Uirapuru-CHP99	-657.50	-1080.20807	-234.79193	0.0003665
Valente-CHP99	-870.00	-1292.70807	-447.29193	0.0000034
Soberano-Diplomata	-838.00	-1260.70807	-415.29193	0.0000068
Supremo-Diplomata	-24.00	-446.70807	398.70807	1.0000000
Tiziu-Diplomata	-45.00	-467.70807	377.70807	0.9999972
Tuiuiu-Diplomata	-107.25	-529.95807	315.45807	0.9964735
Uirapuru-Diplomata	-124.25	-546.95807	298.45807	0.9898311
Valente-Diplomata	-336.75	-759.45807	85.95807	0.2118336
Supremo-Soberano	814.00	391.29193	1236.70807	0.0000115
Tiziu-Soberano	793.00	370.29193	1215.70807	0.0000183
Tuiuiu-Soberano	730.75	308.04193	1153.45807	0.0000723
Uirapuru-Soberano	713.75	291.04193	1136.45807	0.0001053
Valente-Soberano	501.25	78.54193	923.95807	0.0106675
Tiziu-Supremo	-21.00	-443.70807	401.70807	1.0000000
Tuiuiu-Supremo	-83.25	-505.95807	339.45807	0.9995033
Uirapuru-Supremo	-100.25	-522.95807	322.45807	0.9978766
Valente-Supremo	-312.75	-735.45807	109.95807	0.2964796
Tuiuiu-Tiziu	-62.25	-484.95807	360.45807	0.9999547
Uirapuru-Tiziu	-79.25	-501.95807	343.45807	0.9996659

Valente-Tiziu	-291.75	-714.45807	130.95807	0.3861152
Uirapuru-Tuiuiu	-17.00	-439.70807	405.70807	1.0000000
Valente-Tuiuiu	-229.50	-652.20807	193.20807	0.6987456
Valente-Uirapuru	-212.50	-635.20807	210.20807	0.7788194

Os intervalos de confiança para as diferenças podem ser mais bem visualizados de forma gráfica, conforme a seguir:

```
> par(mar=c(4,8,2.3,1))
> plot(ex1.HSDR, las=2, cex.axis=0.35)
```

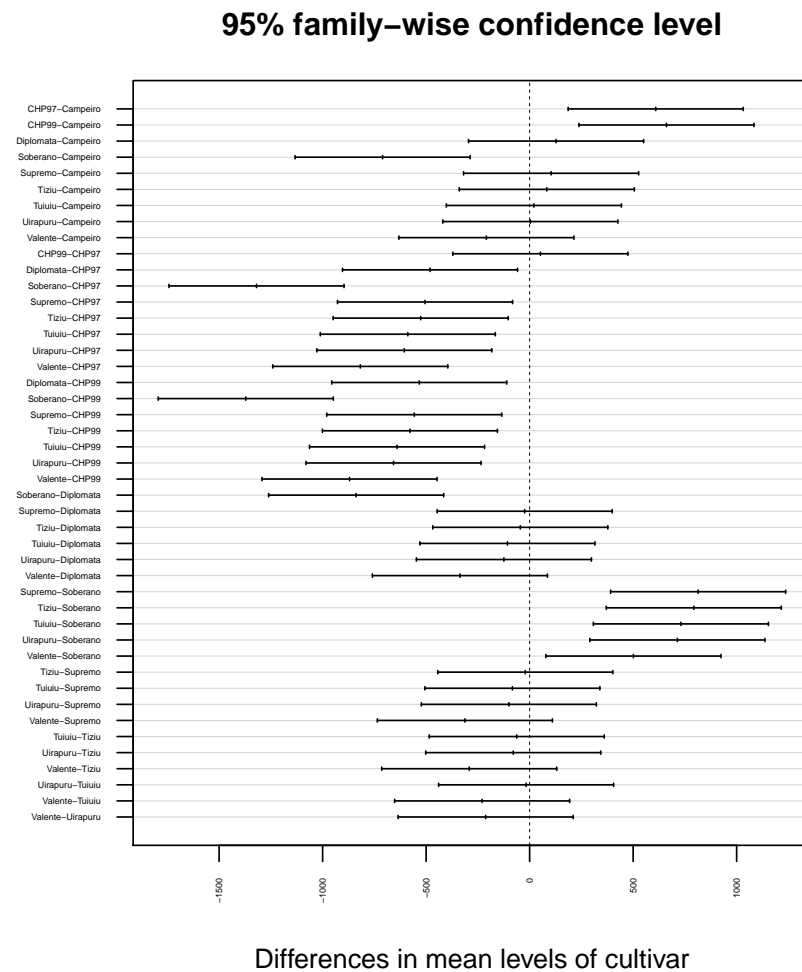


Figura 5: Intervalos de confiança para as diferenças entre as médias dos cultivares - teste de Tukey

O pacote **factorplot** tem funções para apresentar resultados de testes de comparações múltiplas de forma gráfica e numérica (Armstrong, 2014).

```
> require(factorplot)
> fpR <- factorplot(mod1R, adjust.method="none", factor.variable = "cultivar",
+                   pval = 0.05, two.sided = TRUE, order = "natural")
> print(fpR)
```

		Difference	SE	p.val
Campeiro -	CHP97	-609.000	123.918	0.000
Campeiro -	CHP99	-661.000	123.918	0.000
CHP97 -	CHP99	-52.000	123.918	0.678
Campeiro -	Diplomata	-127.750	123.918	0.311
CHP97 -	Diplomata	481.250	123.918	0.001
CHP99 -	Diplomata	533.250	123.918	0.000
Campeiro -	Soberano	710.250	123.918	0.000
CHP97 -	Soberano	1319.250	123.918	0.000
CHP99 -	Soberano	1371.250	123.918	0.000
Diplomata -	Soberano	838.000	123.918	0.000
Campeiro -	Supremo	-103.750	123.918	0.409
CHP97 -	Supremo	505.250	123.918	0.000
CHP99 -	Supremo	557.250	123.918	0.000
Diplomata -	Supremo	24.000	123.918	0.848
Soberano -	Supremo	-814.000	123.918	0.000
Campeiro -	Tiziu	-82.750	123.918	0.509
CHP97 -	Tiziu	526.250	123.918	0.000
CHP99 -	Tiziu	578.250	123.918	0.000
Diplomata -	Tiziu	45.000	123.918	0.719
Soberano -	Tiziu	-793.000	123.918	0.000
Supremo -	Tiziu	21.000	123.918	0.867
Campeiro -	Tuiuiu	-20.500	123.918	0.870
CHP97 -	Tuiuiu	588.500	123.918	0.000
CHP99 -	Tuiuiu	640.500	123.918	0.000
Diplomata -	Tuiuiu	107.250	123.918	0.394
Soberano -	Tuiuiu	-730.750	123.918	0.000
Supremo -	Tuiuiu	83.250	123.918	0.507
Tiziu -	Tuiuiu	62.250	123.918	0.619
Campeiro -	Uirapuru	-3.500	123.918	0.978
CHP97 -	Uirapuru	605.500	123.918	0.000
CHP99 -	Uirapuru	657.500	123.918	0.000
Diplomata -	Uirapuru	124.250	123.918	0.324
Soberano -	Uirapuru	-713.750	123.918	0.000
Supremo -	Uirapuru	100.250	123.918	0.425
Tiziu -	Uirapuru	79.250	123.918	0.527
Tuiuiu -	Uirapuru	17.000	123.918	0.892
Campeiro -	Valente	209.000	123.918	0.102
CHP97 -	Valente	818.000	123.918	0.000
CHP99 -	Valente	870.000	123.918	0.000
Diplomata -	Valente	336.750	123.918	0.011
Soberano -	Valente	-501.250	123.918	0.000
Supremo -	Valente	312.750	123.918	0.017
Tiziu -	Valente	291.750	123.918	0.025
Tuiuiu -	Valente	229.500	123.918	0.074
Uirapuru -	Valente	212.500	123.918	0.097

```
> par(mar=c(4,8,2.3,1))
> plot(fpR, abbrev.char = 10)
```

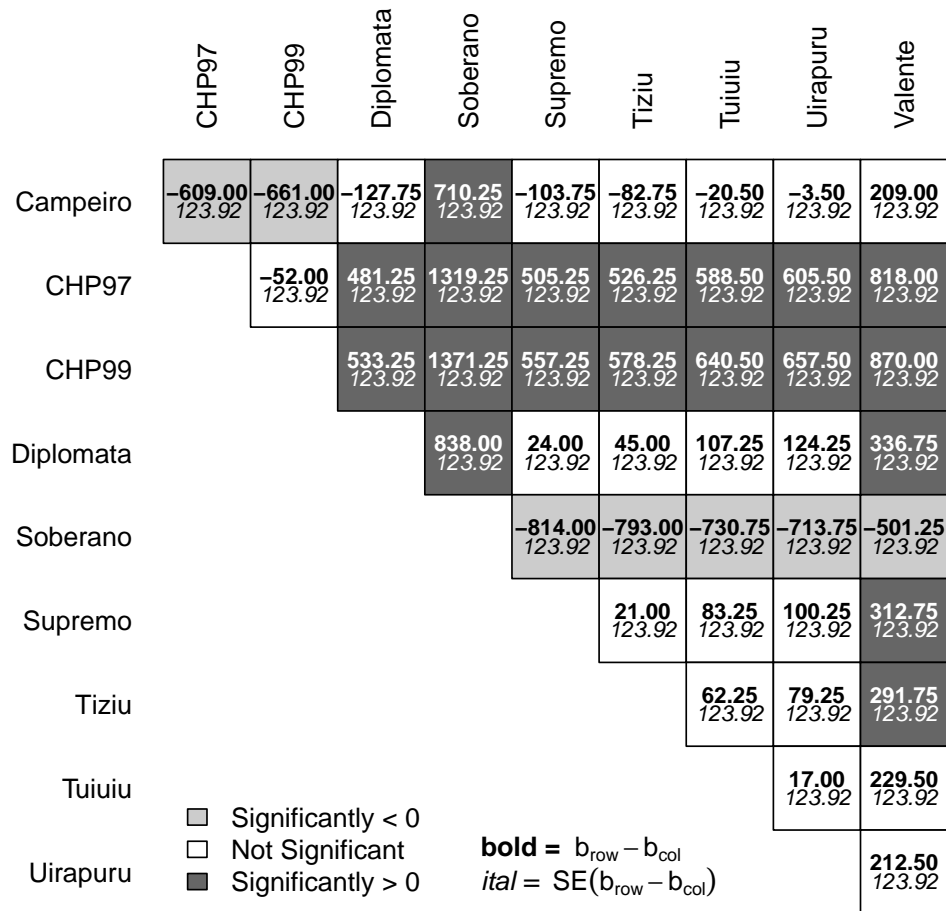


Figura 6: Comparações entre cultivares apresentadas em gráfico utilizando o pacote **factorplot**

O pacote **agricolae** disponibiliza a função **HSD.test** para realização do teste de Tukey, apresentando as médias seguidas por letras.

```
> require(agricolae)
> with(ex1, HSD.test(mod1R, "cultivar", alpha = 0.05, console=T))
```

Study: mod1R ~ "cultivar"

HSD Test for rendimento

Mean Square Error: 30711.5

cultivar, means

rendimento	std r	Min	Max
------------	-------	-----	-----

Campeiro	1064.50	95.22780	4	933	1160
CHP97	1673.50	131.62193	4	1526	1839
CHP99	1725.50	272.37413	4	1334	1966
Diplomata	1192.25	70.67472	4	1094	1249
Soberano	354.25	166.67608	4	184	577
Supremo	1168.25	273.92380	4	957	1563
Tiziu	1147.25	81.72872	4	1060	1252
Tuiuiu	1085.00	248.85471	4	845	1433
Uirapuru	1068.00	102.30673	4	976	1214
Valente	855.50	140.17727	4	649	944

Alpha: 0.05 ; DF Error: 30

Critical Value of Studentized Range: 4.824141

Minimun Significant Difference: 422.7081

Treatments with the same letter are not significantly different.

	rendimento	groups
CHP99	1725.50	a
CHP97	1673.50	a
Diplomata	1192.25	b
Supremo	1168.25	b
Tiziu	1147.25	b
Tuiuiu	1085.00	b
Uirapuru	1068.00	b
Campeiro	1064.50	b
Valente	855.50	b
Soberano	354.25	c

### 3.1.6 Análise com o pacote ExpDes.pt

Nos casos em que o experimento for balanceado e não ocorrerem parcelas perdidas, pode-se utilizar as funções do pacote **ExpDes.pt** que faz a análise de variância e o teste de comparação de médias escolhido. Para delineamentos inteiramente casualizados, a função a ser utilizada é **dic**, devendo-se especificar na sequência: o fator de tratamento, a variável resposta e o teste de comparação de médias (o *default* é o teste de Tukey). A função verifica automaticamente se os resíduos provêm de uma distribuição normal utilizando o teste de Shapiro-Wilk e a homogeneidade de variâncias pelo teste de Bartlett.

```
> require(ExpDes.pt)
> with(ex1, dic(cultivar, rendimento, quali = TRUE,
+             mcomp = "tukey", sigT = 0.05, sigF = 0.05))
```

-----  
Quadro da analise de variancia  
-----

GL	SQ	QM	Fc	Pr>Fc
----	----	----	----	-------

Tratamento 9 5371321 596813 19.433 3.3135e-10  
Residuo 30 921345 30712  
Total 39 6292666

---

CV = 15.46 %

---

Teste de normalidade dos residuos

Valor-p: 0.6633406

De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser con

---

Teste de homogeneidade de variancia

valor-p: 0.2465044

De acordo com o teste de bartlett a 5% de significancia, as variancias podem ser cons

---

Teste de Tukey

---

Grupos Tratamentos Medias

a	CHP99	1725.5
a	CHP97	1673.5
b	Diplomata	1192.25
b	Supremo	1168.25
b	Tiziu	1147.25
b	Tuiuiu	1085
b	Uirapuru	1068
b	Campeiro	1064.5
b	Valente	855.5
c	Soberano	354.25

---

### 3.2 Delineamento em Blocos Casualizados

Os dados a seguir se referem a um experimento de avaliação de variedades de milho de polinização aberta realizado em Chapecó, no delineamento em blocos casualizados com quatro repetições.

Tabela 2: Rendimento de grãos (t/ha) em variedades de milho de polinização aberta

Variedade	Bloco 1	Bloco 2	Bloco 3	Bloco 4
Rendimento de grãos (t/ha)				
BRS Missões	5,41	6,47	5,73	6,86
BRS Planalto	3,54	4,21	5,62	3,55
AM 4001	3,14	3,56	2,40	4,70
AM 4002	5,09	3,29	4,32	5,48
SCS 154 Fortuna	6,56	5,05	6,02	5,44
SCS 155 Catarina	7,40	5,98	6,04	6,86
SCS 156 Colorado	5,53	4,82	5,54	6,43

Neste experimento, o interesse do pesquisador é testar a hipótese:

$$\begin{cases} H_0 : \mu_{\text{BRS Missões}} = \mu_{\text{BRS Planalto}} = \dots = \mu_{\text{SCS 156 Colorado}} \\ H_1 : \mu_{\text{BRS Missões}} \neq \mu_{\text{BRS Planalto}} \neq \dots \neq \mu_{\text{SCS 156 Colorado}} \end{cases} \quad (4)$$

### 3.3 Importando os dados

```
> ex2 <- read.csv("ex2.csv", dec=",", header = T)
> head(ex2)
```

```
bloco  genotipo  espiga  rend
1      1  BRSMissões    135  5.41
2      2  BRSMissões    160  6.47
3      3  BRSMissões    110  5.73
4      4  BRSMissões    140  6.86
5      1  BRSPlanalto    130  3.54
6      2  BRSPlanalto    125  4.21
```

```
> ex2$genotipo <- as.factor(ex2$genotipo)
```



### 3.3.1 Análise exploratória dos dados

Assim como apresentado para o delineamento inteiramente casualizado, antes de iniciar a análise propriamente dita, deve-se realizar uma análise exploratória, o que pode ser feito por meio dos comandos **str**, que apresenta a estrutura dos dados e **summary** que fornece um resumo dos dados, com mínimo e máximo, a média e as separatrizes.

```
> str(ex2)

'data.frame':      28 obs. of  4 variables:
 $ bloco   : int  1 2 3 4 1 2 3 4 1 2 ...
 $ genotipo: Factor w/ 7 levels "AM4001","AM4002",...: 3 3 3 3 4 4 4 4 1 1 ...
 $ espiga  : int  135 160 110 140 130 125 150 144 105 125 ...
 $ rend    : num  5.41 6.47 5.73 6.86 3.54 4.21 5.62 3.55 3.14 3.56 ...

> summary(ex2)

      bloco          genotipo      espiga          rend
Min.   :1.00   AM4001           :4   Min.    :105.0   Min.    :2.400
1st Qu.:1.75   AM4002           :4   1st Qu.:120.0   1st Qu.:4.293
Median :2.50   BRSMissões        :4   Median :132.5   Median :5.460
Mean   :2.50   BRSPlanalto       :4   Mean   :131.2   Mean   :5.180
3rd Qu.:3.25   SCS154Fortuna     :4   3rd Qu.:142.0   3rd Qu.:6.025
Max.   :4.00   SCS155Catarina    :4   Max.   :160.0   Max.   :7.400
      SCS 156Colorado:4

> ex2$bloco <- as.factor(ex2$bloco)
```

Para realizar o cálculo das estatísticas descritivas por nível de um determinado fator de tratamento ou bloco, pode-se empregar a função **tapply**, especificando-se a variável resposta, o fator de agrupamento e a função a ser aplicada, conforme descrito para o caso do delineamento inteiramente casualizado.

```
> repet <- with(ex2, tapply(rend, list(genotipo, bloco), mean))
> ex2.nrep <- with(ex2, tapply(rend, genotipo, length))
> ex2.medias <- with(ex2, tapply(rend, genotipo, mean))
> ex2.var <- with(ex2, tapply(rend, genotipo, var))
> ex2.dp <- with(ex2, tapply(rend, genotipo, sd))
> (resumo <- data.frame(bloco=repet, Rep = ex2.nrep,
+                       Medias = ex2.medias, Var = ex2.var, DP=ex2.dp))
```

	bloco.1	bloco.2	bloco.3	bloco.4	Rep	Medias	Var	DP
AM4001	3.14	3.56	2.40	4.70	4	3.4500	0.9244000	0.9614572
AM4002	5.09	3.29	4.32	5.48	4	4.5450	0.9323000	0.9655568
BRSMissões	5.41	6.47	5.73	6.86	4	6.1175	0.4420917	0.6648997
BRSPlanalto	3.54	4.21	5.62	3.55	4	4.2300	0.9570000	0.9782638
SCS154Fortuna	6.56	5.05	6.02	5.44	4	5.7675	0.4379583	0.6617842
SCS155Catarina	7.40	5.98	6.04	6.86	4	6.5700	0.4673333	0.6836178
SCS 156Colorado	5.53	4.82	5.54	6.43	4	5.5800	0.4347333	0.6593431

O gráfico de caixa (*Box-plot*), obtido com a função **boxplot**, especificando-se a variável resposta e o fator de agrupamento, pode ser utilizado para auxiliar na análise exploratória. Nesse caso, pode-se obter gráficos tanto da variável resposta em função do fator de tratamento, como em função do fator de controle (blocos), conforme ilustrado a seguir:

```
> par(mfrow=c(1,2))
> mar=c(5,3,1,1)
> with(ex2, boxplot(rend ~ bloco, xlab="Bloco",
+                  ylab="Rendimento de grãos (t/ha)",
+                  ylim=c(1.5, 8.5), las=2, cex.axis=0.8,
+                  col=c("blue","green","red","violet")))
> mar=c(5,3,5,1)
> with(ex2, boxplot(rend ~ genotipo, xlab=" ",
+                  ylab="Rendimento de grãos (t/ha)",
+                  ylim=c(1.5, 8.5), las=2, cex.axis=0.8,
+                  col=rainbow(n=7)))
```

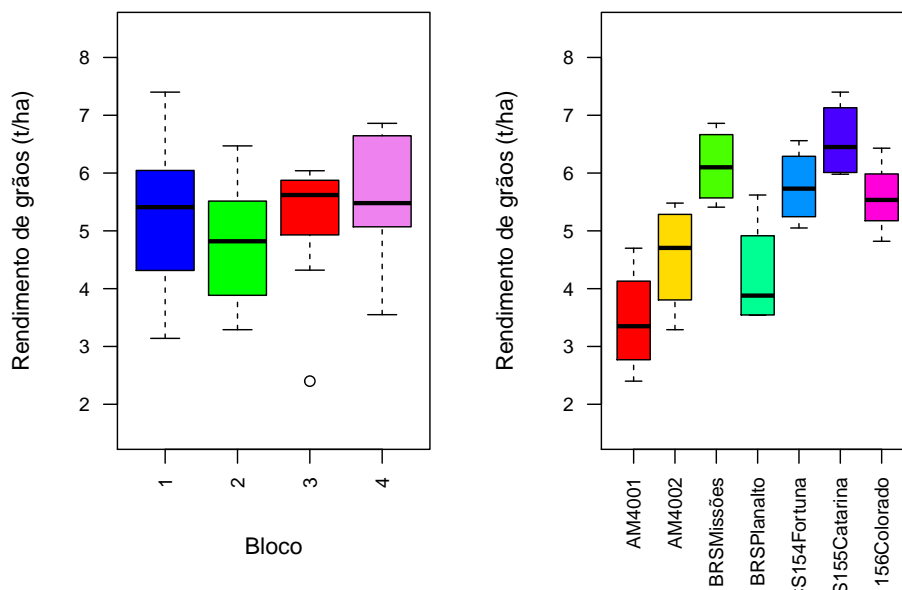


Figura 7: Gráficos de caixa (*boxplots*) dos dados de rendimento de milho em função dos blocos e dos genótipos

A opção `las=2`, dentro da função **Box-plot**, altera a orientação do texto dos eixos para a vertical, e a opção `cex.axis` altera o tamanho da letra a ser impressa nos eixos.

### 3.3.2 Verificação das pressuposições do modelo estatístico

Para que a análise de variância seja válida, no caso de um experimento em DBC, devem ser satisfeitas as pressuposições de homogeneidade de variâncias entre tratamentos, normalidade dos resíduos e aditividade de blocos.

## 1. Homogeneidade de Variâncias

Para verificar a homogeneidade de variâncias entre tratamentos em um experimento em DBC, deve-se utilizar os resíduos do modelo já ajustados para o efeito de blocos. A hipótese a ser testada e os testes a serem empregados são os mesmos apresentados para o delineamento inteiramente casualizado.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_t^2 \end{cases} \quad (5)$$

Sempre que o nível descritivo do teste ( $p$ -value) for menor que o nível de significância estabelecido ( $\alpha$ ), rejeita-se  $H_0$ . Caso contrário, não se rejeita indicando que não existem evidências contra a homocedasticidade.

```
> bartlett.test(residuals(aov(rend ~ bloco, data=ex2)) ~ genotipo, data=ex2)
```

```
Bartlett test of homogeneity of variances
```

```
data: residuals(aov(rend ~ bloco, data = ex2)) by genotipo
```

```
Bartlett's K-squared = 4.6235, df = 6, p-value = 0.5929
```

Neste caso, para o teste de Bartlett, como o nível descritivo do teste ( $p$  - value = 0.5929) é maior que o nível de significância considerado ( $\alpha = 0.05$ ), não se rejeita a hipótese de homogeneidade de variâncias.

```
> require(car)
```

```
> leveneTest(residuals(aov(rend ~ bloco, data=ex2)) ~ genotipo, data=ex2)
```

```
Levene's Test for Homogeneity of Variance (center = median)
```

```
  Df F value Pr(>F)
group 6  1.8306 0.1415
    21
```

Para o teste de Levene, também não se rejeita a hipótese de homogeneidade de variâncias, pois o nível descritivo do teste ( $p$  - value = 0.1415) é maior que o nível de significância  $\alpha = 0.05$ .

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, 1)$ )

A hipótese a ser testada neste caso é a mesma apresentada no caso do delineamento inteiramente casualizado:

$$\begin{cases} H_0: \text{ Os res\u00edduos do modelo prov\u00eam de uma popula\u00e7\u00e3o com distribui\u00e7\u00e3o normal} \\ H_1: \text{ Os res\u00edduos do modelo n\u00e3o prov\u00eam de uma popula\u00e7\u00e3o com distribui\u00e7\u00e3o normal} \end{cases} \quad (6)$$

Para realiza\u00e7\u00e3o dos testes, a \u00fanica diferen\u00e7a em rela\u00e7\u00e3o ao teste realizado para os dados de um experimento em DIC, \u00e9 que agora o fator de controle (blocos) tamb\u00e9m deve ser considerado no modelo para obten\u00e7\u00e3o dos res\u00edduos.

```
> shapiro.test(residuals(aov(rend ~ bloco+genotipo, data=ex2)))
```

Shapiro-Wilk normality test

```
data: residuals(aov(rend ~ bloco + genotipo, data = ex2))
W = 0.96824, p-value = 0.5342
```

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.5342$ ) \u00e9 maior que  $\alpha = 0.05$ , n\u00e3o h\u00e1 evid\u00eancias para rejeitar a hip\u00f3tese de que os res\u00edduos do modelo seguem a distribui\u00e7\u00e3o normal, sendo essa pressuposi\u00e7\u00e3o satisfeita.

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p\text{-value} = 0.3364$ ) \u00e9 maior que  $\alpha = 0.05$ , da mesma forma que observado para o teste de Shapiro-Wilk, n\u00e3o existem evid\u00eancias para rejeitar a hip\u00f3tese de que os res\u00edduos do modelo seguem a distribui\u00e7\u00e3o normal.

```
> ks.test(residuals(aov(rend ~ bloco+genotipo, data=ex2)), "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data: residuals(aov(rend ~ bloco + genotipo, data = ex2))
D = 0.17244, p-value = 0.3364
alternative hypothesis: two-sided
```

Para verificar a pressuposi\u00e7\u00e3o de normalidade dos res\u00edduos, pode-se ainda utilizar o gr\u00e1fico de quantis com envelope de simula\u00e7\u00e3o, por meio da fun\u00e7\u00e3o `qqPlot`, especificando os res\u00edduos do modelo. Nesse caso, espera-se que os pontos estejam dentro do intervalo formado pelo envelope simulado.

```
> qqPlot(rstudent(aov(rend ~ bloco+genotipo, data=ex2)),
+        pch=19, ylab="Resíduos")
```

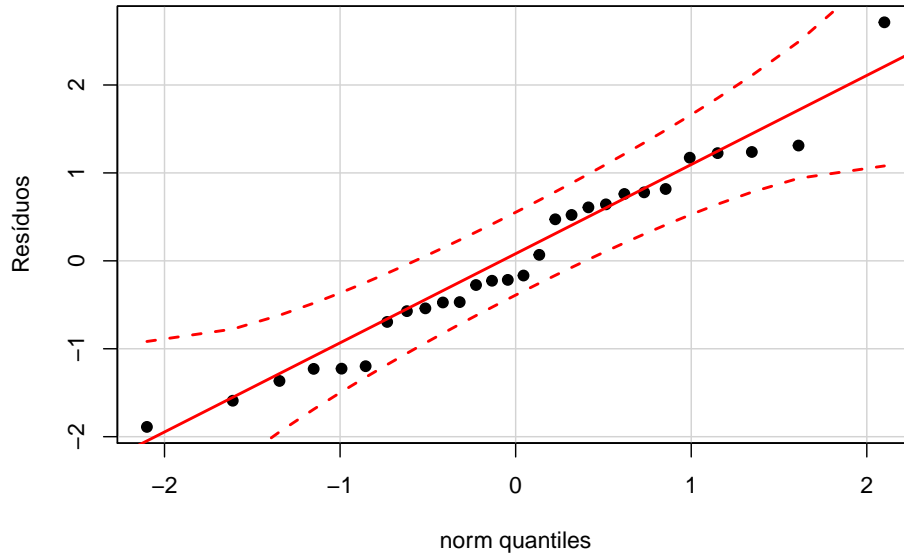


Figura 8: Gráfico qqplot para avaliar a distribuição normal dos resíduos

### 3. Bloco de efeito aditivo

A aditividade de blocos pode ser verificada pelo teste de não aditividade de Tukey utilizando a função `tukey.1df` do pacote `dae` desenvolvida por Brien (2011):

```
> require(dae)
> modad<-aov(rend ~ bloco + genotipo + Error(bloco), ex2)
> tukey.1df(modad, ex2, error.term= "Within")
```

```
$Tukey.SS
[1] 2.823803e-06
```

```
$Tukey.F
[1] 4.289525e-06
```

```
$Tukey.p
[1] 0.9983716
```

```
$Devn.SS
[1] 11.19113
```

Como o p-valor ( $Tukey.p = 0,9983$ ) é maior que o nível de significância estabelecido ( $\alpha = 5\%$ ) não se rejeita a hipótese de que os blocos têm efeito aditivo no modelo. Isto significa que o efeito do tratamento não depende do bloco.

#### 4. Transformação de Box-Cox

Equivalente ao apresentado para o delineamento inteiramente casualizado, pode-se utilizar a função `boxcox` para verificar a necessidade de transformação dos dados. Neste caso deve-se verificar se o valor 1 pertence ao intervalo de confiança para  $\lambda$ .

Para extrair o valor de  $\lambda$  com o máximo de verossimilhança utilizam-se os comandos a seguir:

```
> require(MASS)
> bc <- boxcox(aov(rend ~ bloco+genotipo, data=ex2),
+             lambda = seq(-2, 6, l = 101))
> bc$x[which.max(bc$y)]

[1] 1.6
```

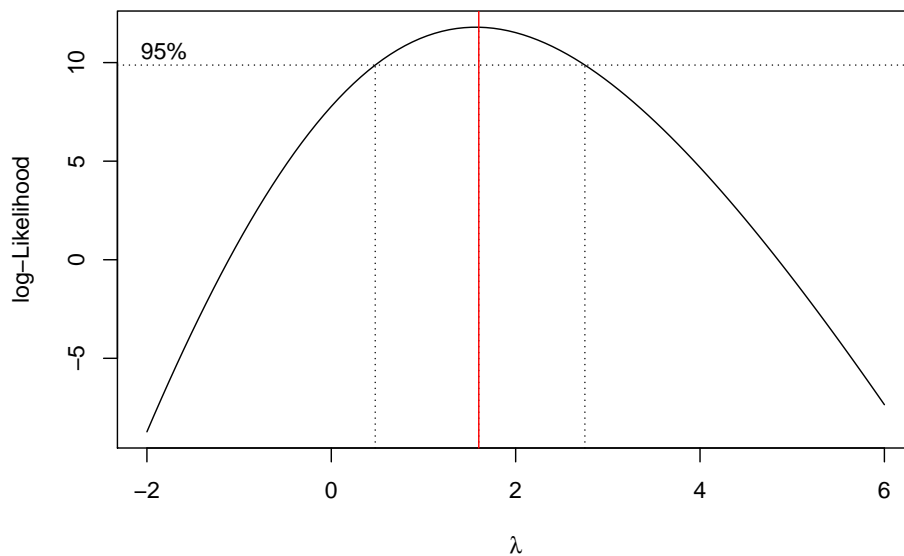


Figura 9: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Neste caso, não há evidências da necessidade de transformação.

#### 5. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(rend ~ bloco+genotipo, data=ex2)),
+      rstudent(aov(rend ~ bloco+genotipo, data=ex2)),
+      xlab="Valores ajustados",
+      ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(rend ~ bloco+genotipo, data=ex2)),prob=T,
+      main=" ",xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(rend ~ bloco+genotipo, data=ex2))),
+      col="blue")
> ##qqPlot precisa do pacote car
> qqPlot(rstudent(aov(rend ~ bloco+genotipo, data=ex2)),
+      ylab="Resíduos Studentizados", pch=19)
> ##Boxplot
> boxplot(rstudent(aov(rend ~ bloco+genotipo, data=ex2)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2,
+      cex.axis=1.2, col="turquoise")

```

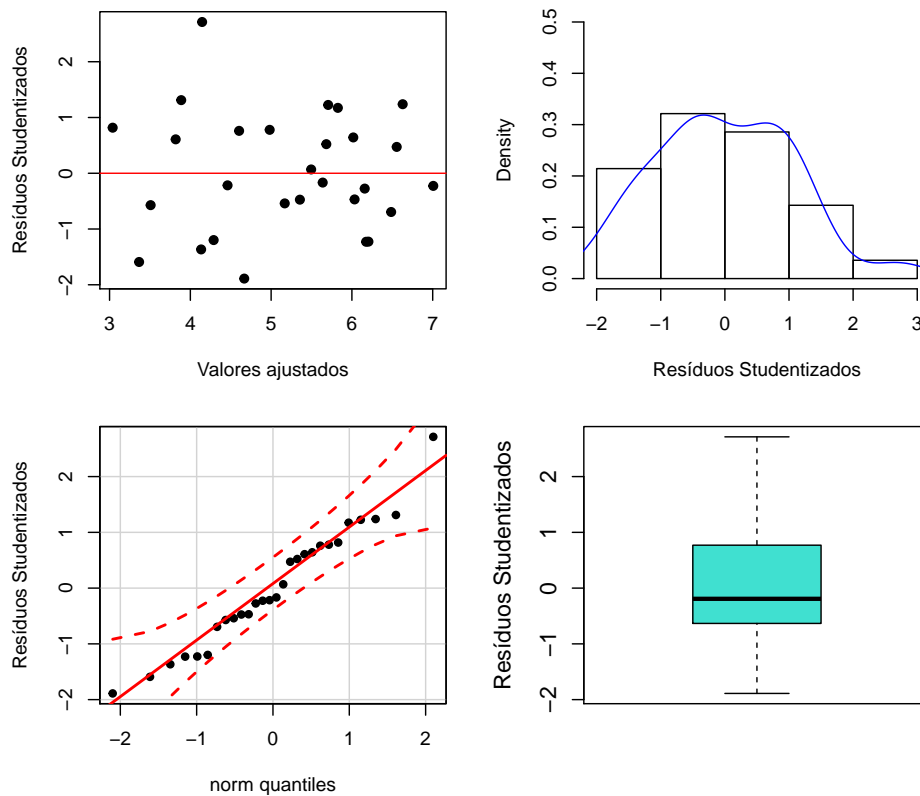


Figura 10: Gráficos de diagnósticos para verificar as pressuposições da análise de variância

Como todas as pressuposições da análise de variância foram satisfeitas, pode-se prosseguir com a análise.

### 3.3.3 Análise de variância

Pode-se utilizar a função `aov` especificando no modelo os efeitos de bloco e de tratamentos. É importante verificar se os blocos são considerados como um fator qualitativo.

```
> mod2 <- aov(rend ~ bloco+genotipo, data=ex2)
> summary(mod2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	3	2.596	0.865	1.392	0.277482
genotipo	6	30.459	5.077	8.165	0.000231 ***
Residuals	18	11.191	0.622		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Como  $p$ -valor para efeito de blocos não foi significativo ( $p - value = 0,2774$ ), não há evidências de efeito de bloco na variável testada. O efeito de genótipo foi significativo ( $p - value = 0,0002$ ), o que indica que pelo menos dois genótipos diferem entre si quanto ao seu efeito no rendimento de grãos. Neste caso, segue-se com um teste de comparação das médias.

### 3.3.4 Teste de comparações múltiplas

A função `TukeyHSD` pode ser utilizada para construir intervalos de confiança para as diferenças entre as médias dos genótipos.

```
> ex2.HSDM <- TukeyHSD(mod2, "genotipo", ordered = TRUE)
> ex2.HSDM
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
factor levels have been ordered
```

```
Fit: aov(formula = rend ~ bloco + genotipo, data = ex2)
```

```
$genotipo
```

	diff	lwr	upr	p adj
BRSPlanalto-AM4001	0.7800	-1.06237852	2.622379	0.7955020
AM4002-AM4001	1.0950	-0.74737852	2.937379	0.4678171
SCS 156Colorado-AM4001	2.1300	0.28762148	3.972379	0.0176057
SCS154Fortuna-AM4001	2.3175	0.47512148	4.159879	0.0087419
BRSMissões-AM4001	2.6675	0.82512148	4.509879	0.0023381
SCS155Catarina-AM4001	3.1200	1.27762148	4.962379	0.0004337
AM4002-BRSPlanalto	0.3150	-1.52737852	2.157379	0.9971131
SCS 156Colorado-BRSPlanalto	1.3500	-0.49237852	3.192379	0.2455832



SCS154Fortuna-BRSPlanalto	1.5375	-0.30487852	3.379879	0.1396732
BRSMissões-BRSPlanalto	1.8875	0.04512148	3.729879	0.0425998
SCS155Catarina-BRSPlanalto	2.3400	0.49762148	4.182379	0.0080331
SCS 156Colorado-AM4002	1.0350	-0.80737852	2.877379	0.5306124
SCS154Fortuna-AM4002	1.2225	-0.61987852	3.064879	0.3459547
BRSMissões-AM4002	1.5725	-0.26987852	3.414879	0.1248809
SCS155Catarina-AM4002	2.0250	0.18262148	3.867379	0.0259151
SCS154Fortuna-SCS 156Colorado	0.1875	-1.65487852	2.029879	0.9998477
BRSMissões-SCS 156Colorado	0.5375	-1.30487852	2.379879	0.9556054
SCS155Catarina-SCS 156Colorado	0.9900	-0.85237852	2.832379	0.5789421
BRSMissões-SCS154Fortuna	0.3500	-1.49237852	2.192379	0.9948869
SCS155Catarina-SCS154Fortuna	0.8025	-1.03987852	2.644879	0.7744635
SCS155Catarina-BRSMissões	0.4525	-1.38987852	2.294879	0.9806112

Observa-se, por exemplo, que os genótipos BRSPlanalto e AM4001 não diferem significativamente ( $p_{adj} = 0.7955$ ), enquanto o genótipo SCS155 Catarina produz em média 2,34 t/ha a mais que o genótipo BRS Planalto, sendo esta diferença significativa ( $p_{adj} = 0.0080$ ). Os intervalos de confiança podem ser visualizados de forma gráfica.

```
> par(mar=c(4,10,2,2))
> plot(ex2.HSDM, las=2, cex.axis=0.6)
```

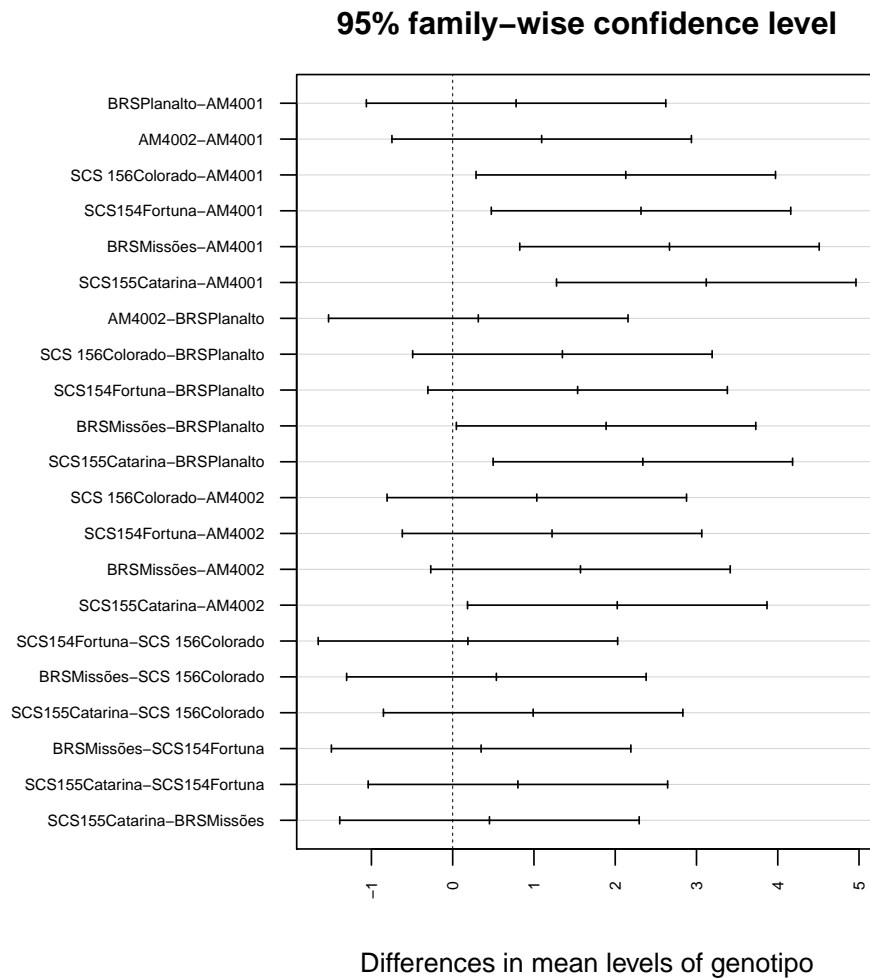


Figura 11: Intervalos de confiança para as diferenças entre as médias dos genótipos - teste de Tukey

No gráfico, intervalos de confiança que contêm o valor 0 (linha pontilhada) indicam que a diferença entre os genótipos não é significativa.

A seguir apresenta-se também o teste de comparação de médias utilizando-se o pacote `factorplot`.

```
> require(factorplot)
> fpM <- factorplot(mod2, p.adjust="hochberg",
+                   factor.variable = "genotipo", pval = 0.05,
+                   two.sided = TRUE, order = "natural")
> print(fpM)
```

		Difference	SE	p.val
AM4001 -	AM4002	-1.095	0.558	0.065
AM4001 -	BRSMissões	-2.667	0.558	0.000

AM4002 -	BRSMissões	-1.572	0.558	0.011
AM4001 -	BRSPlanalto	-0.780	0.558	0.179
AM4002 -	BRSPlanalto	0.315	0.558	0.579
BRSMissões -	BRSPlanalto	1.887	0.558	0.003
AM4001 -	SCS154Fortuna	-2.318	0.558	0.001
AM4002 -	SCS154Fortuna	-1.222	0.558	0.042
BRSMissões -	SCS154Fortuna	0.350	0.558	0.538
BRSPlanalto -	SCS154Fortuna	-1.538	0.558	0.013
AM4001 -	SCS155Catarina	-3.120	0.558	0.000
AM4002 -	SCS155Catarina	-2.025	0.558	0.002
BRSMissões -	SCS155Catarina	-0.453	0.558	0.428
BRSPlanalto -	SCS155Catarina	-2.340	0.558	0.001
SCS154Fortuna -	SCS155Catarina	-0.803	0.558	0.167
AM4001 -	SCS_156Colorado	-2.130	0.558	0.001
AM4002 -	SCS_156Colorado	-1.035	0.558	0.080
BRSMissões -	SCS_156Colorado	0.537	0.558	0.348
BRSPlanalto -	SCS_156Colorado	-1.350	0.558	0.026
SCS154Fortuna -	SCS_156Colorado	0.188	0.558	0.741
SCS155Catarina -	SCS_156Colorado	0.990	0.558	0.093

```
> plot(fpM, abbrev.char=12, cex=0.5)
```

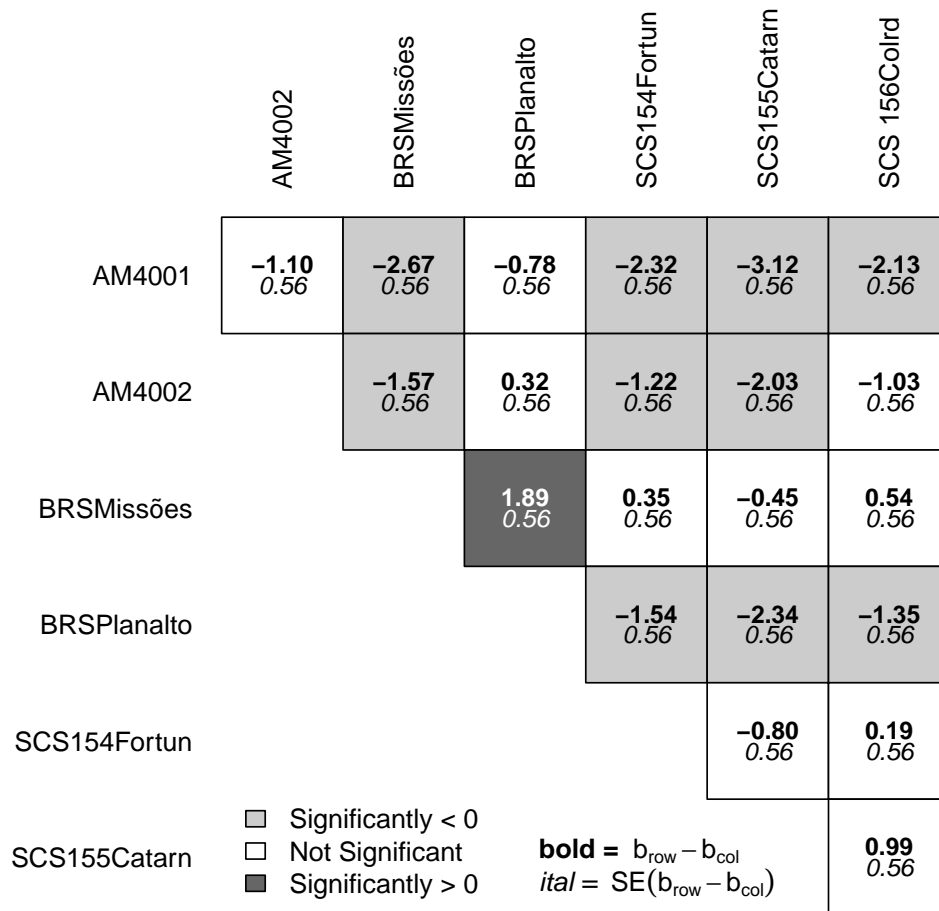


Figura 12: Comparações entre cultivares apresentadas em gráfico utilizando o pacote **factorplot**

A função `HSD.test`, do pacote `agricolae`, também realiza o teste de Tukey apresentando as médias seguidas por letras.

```
> require(agricolae)
> with(ex2, HSD.test(mod2, "genotipo", alpha = 0.05, group=T, console=T))
```

```
Study: mod2 ~ "genotipo"
```

```
HSD Test for rend
```

```
Mean Square Error: 0.6217298
```

```
genotipo, means
```

```

      rend      std r  Min  Max
AM4001 3.4500 0.9614572 4 2.40 4.70
```

AM4002	4.5450	0.9655568	4	3.29	5.48
BRSMissões	6.1175	0.6648997	4	5.41	6.86
BRSPlanalto	4.2300	0.9782638	4	3.54	5.62
SCS154Fortuna	5.7675	0.6617842	4	5.05	6.56
SCS155Catarina	6.5700	0.6836178	4	5.98	7.40
SCS 156Colorado	5.5800	0.6593431	4	4.82	6.43

Alpha: 0.05 ; DF Error: 18

Critical Value of Studentized Range: 4.673132

Minimum Significant Difference: 1.842379

Treatments with the same letter are not significantly different.

		rend groups
SCS155Catarina	6.5700	a
BRSMissões	6.1175	ab
SCS154Fortuna	5.7675	abc
SCS 156Colorado	5.5800	abc
AM4002	4.5450	bcd
BRSPlanalto	4.2300	cd
AM4001	3.4500	d

### 3.3.5 Análise de com o pacote ExpDes.pt

Assim como em um experimento em DIC, nos experimentos em DBC que forem balanceados, pode-se utilizar a função `dbc`, disponível no pacote `ExpDes.pt`, que executa a análise de variância e o teste de comparação de médias.

```
> require(ExpDes.pt)
> with(ex2, dbc(genotipo, bloco, rend, quali = TRUE, mcomp = "tukey", sigT = 0.05, s
```

-----  
Quadro da análise de variancia

	GL	SQ	QM	Fc	Pr>Fc
Tratamento	6	30.459	5.0765	8.1652	0.000231
Bloco	3	2.596	0.8654	1.3920	0.277482
Residuo	18	11.191	0.6217		
Total	27	44.247			

-----  
CV = 15.22 %

-----  
Teste de normalidade dos residuos

valor-p: 0.5341963

De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser con

-----  
-----  
Teste de homogeneidade de variancia

valor-p: 0.2879556

De acordo com o teste de oneillmathews a 5% de significancia, as variancias podem ser

-----  
Teste de Tukey

-----  
Grupos Tratamentos Medias

a	SCS155Catarina	6.57
ab	BRSMissões	6.1175
abc	SCS154Fortuna	5.7675
abc	SCS 156Colorado	5.58
bcd	AM4002	4.545
cd	BRSPlanalto	4.23
d	AM4001	3.45

-----

A função realiza automaticamente o teste de Shapiro-Wilk para verificar se os resíduos provêm de uma distribuição normal. Para verificar a homogeneidade de variâncias os autores do pacote definiram com *default* o teste de "oneillmathews" (O'NEILL & MATHEWS, 2000). Contudo, fornecem outras opções como o teste de Han ("han") e o teste de Ascombe e Tukey ("anscombetukey").

### 3.4 Delineamento em Quadrado Latino

Sejam os resultados de um experimento no delineamento quadrado latino, em que foram avaliados o rendimento (t/ha) de quatro cultivares (Caçador, Contestado, Chonan e Quitéria) de alho. Na escolha do delineamento, o bloqueamento de linhas foi em razão da heterogeneidade da fertilidade entre as curvas de nível (cada curva igual a uma linha) e o bloqueamento de colunas foi devido à heterogeneidade entre os tratamentos dos bulbos de alho (classes: florão, graúdo, médio e miúdo). Os resultados estão apresentados na tabela a seguir, onde o valor está escrito ao lado do nome de cada tratamento.

Tabela 3: Rendimento de bulbos de alho (t/ha) em quatro cultivares

Linha	Bulbo (Coluna)			
	Florão	Graúdo	Médio	Miúdo
1	11 <sup>Quitéria</sup>	4 <sup>Contestado</sup>	7 <sup>Caçador</sup>	15 <sup>Chonan</sup>
2	12 <sup>Caçador</sup>	9 <sup>Chonan</sup>	7 <sup>Contestado</sup>	9 <sup>Quitéria</sup>
3	8 <sup>Contestado</sup>	7 <sup>Quitéria</sup>	12 <sup>Chonan</sup>	5 <sup>Caçador</sup>
4	17 <sup>Chonan</sup>	7 <sup>Caçador</sup>	12 <sup>Quitéria</sup>	10 <sup>Contestado</sup>

Neste experimento, o interesse do pesquisador é testar a hipótese:

$$\begin{cases} H_0 : \mu_{\text{Quitéria}} = \mu_{\text{Caçador}} = \dots = \mu_{\text{Chonan}} \\ H_1 : \mu_{\text{Quitéria}} \neq \mu_{\text{Caçador}} \neq \dots \neq \mu_{\text{Chonan}} \end{cases} \quad (7)$$

#### 3.4.1 Importando os dados

```
> #ex3 <- read.xlsx("dados.xlsx", sheet="ex3")
> ex3 <- read.csv("ex3.csv", dec=",", header = T)
> head(ex3)
```

```
linha bulbo variedade rendimento
1     1 florão  Quitéria          11
2     2 florão  Caçador           12
3     3 florão  Contestado          8
4     4 florão  Chonan            17
5     1 graúdo  Contestado          4
6     2 graúdo  Chonan             9
```

### 3.4.2 Análise exploratória dos dados

A análise exploratória dos dados pode ser realizada como nos delineamentos anteriores.

```
> str(ex3)

'data.frame':      16 obs. of  4 variables:
 $ linha      : int  1 2 3 4 1 2 3 4 1 2 ...
 $ bulbo      : Factor w/ 4 levels "florão","graúdo",...: 1 1 1 1 2 2 2 2 3 3 ...
 $ variedade  : Factor w/ 4 levels "Caçador","Chonan",...: 4 1 3 2 3 2 4 1 1 3 ...
 $ rendimento: int  11 12 8 17 4 9 7 7 7 7 ...

> summary(ex3)

      linha      bulbo      variedade      rendimento
Min.   :1.00   florão:4   Caçador   :4   Min.    : 4.0
1st Qu.:1.75   graúdo:4   Chonan   :4   1st Qu.: 7.0
Median :2.50   médio :4   Contestado:4   Median  : 9.0
Mean   :2.50   miúdo :4   Quitéria  :4   Mean    : 9.5
3rd Qu.:3.25                                     3rd Qu.:12.0
Max.   :4.00                                     Max.    :17.0

> ex3$linha <- as.factor(ex3$linha)
```

A aplicação da função `tapply` segue como nos exemplos anteriores.

```
> ex3.nrep <- with(ex3, tapply(rendimento, variedade, length))
> ex3.medias <- with(ex3, tapply(rendimento, variedade, mean))
> ex3.var <- with(ex3, tapply(rendimento, variedade, var))
> ex3.dp <- with(ex3, tapply(rendimento, variedade, sd))
> (resumo <- data.frame( Rep = ex3.nrep, Medias = ex3.medias,
+                       Var = ex3.var, DP=ex3.dp))
```

	Rep	Medias	Var	DP
Caçador	4	7.75	8.916667	2.986079
Chonan	4	13.25	12.250000	3.500000
Contestado	4	7.25	6.250000	2.500000
Quitéria	4	9.75	4.916667	2.217356

O gráfico box-plot pode ser construído da mesma forma como nos delineamentos anteriores. Neste caso, especificando a variável resposta em função dos fatores de agrupamento (variedade, bulbo e linha).



```

> par(mfrow=c(1,3),mar=c(6,5.5,1,1),mgp=c(4.5,1,0), mar=c(6,4,1,1))
> with(ex3, boxplot(rendimento ~ variedade, xlab="Variedade",
+                  ylab="Rendimento (t/ha)", ylim=c(4, 18),
+                  las=2, cex.axis=0.6, horizontal = F))
> with(ex3, boxplot(rendimento ~ bulbo, xlab="Bulbo",
+                  ylab="Rendimento (t/ha)", ylim=c(4, 18),
+                  las=2, cex.axis=0.6, horizontal = F))
> with(ex3, boxplot(rendimento ~ linha, xlab="Linha",
+                  ylab="Rendimento (t/ha)", ylim=c(4, 18),
+                  las=2, cex.axis=0.6, horizontal = F))

```

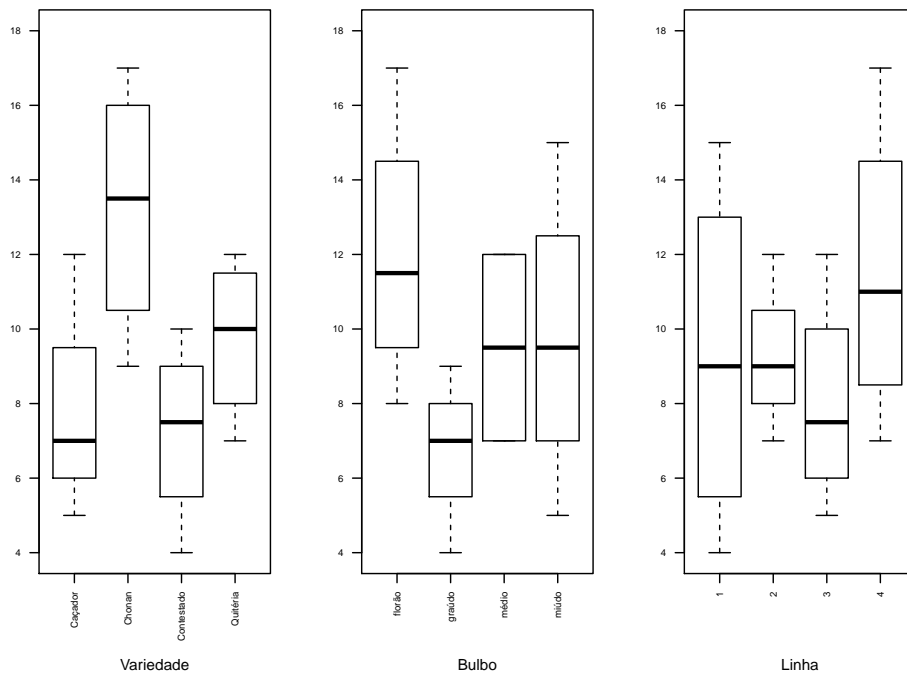


Figura 13: Gráficos de caixa (*boxplots*) dos dados de rendimento em função da variedade, do tamanho do bulbo e da linha

### 3.4.3 Verificação das pressuposições do modelo estatístico

Para que a análise de variância seja válida, no caso de um experimento em DQL, devem ser satisfeitas as pressuposições de homogeneidade de variâncias entre tratamentos, normalidade dos resíduos e aditividade de linhas e colunas.

#### 1. Homogeneidade de Variâncias

Para verificar a homogeneidade de variâncias entre tratamentos em um experimento em DQL, deve-se utilizar os resíduos do modelo já ajustado para os efeitos de linhas

e colunas. A hipótese a ser testada e os testes a serem empregados são os mesmos apresentados para os delineamentos anteriores.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_t^2 \end{cases} \quad (8)$$

```
> bartlett.test(residuals(aov(rendimento ~ linha+bulbo, data=ex3))
+ ~ variedade, data=ex3)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(rendimento ~ linha + bulbo, data = ex3)) by variedade
Bartlett's K-squared = 4.3238, df = 3, p-value = 0.2286
```

Neste caso, para o teste de Bartlett, como o nível descritivo do teste ( $p - value = 0.2286$ ) é maior que o nível de significância considerado para o teste ( $\alpha = 0.05$ ) não se rejeita a hipótese de homogeneidade de variâncias.

```
> require(car)
> leveneTest(residuals(aov(rendimento ~ linha+bulbo,
+ data=ex3)) ~ variedade, data=ex3)
```

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 3    1.125 0.3778
  12
```

Para o teste de Levene, também não se rejeita a hipótese de homogeneidade de variâncias, pois o nível descritivo do teste ( $p - value = 0.3778$ ).

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, 1)$ )

A hipótese a ser testada neste caso é a mesma apresentada para os casos dos delineamentos anteriores:

$$\begin{cases} H_0 : \text{Os resíduos do modelo provêm de uma população com distribuição normal;} \\ H_1 : \text{Os resíduos do modelo não provêm de uma população com distribuição normal.} \end{cases} \quad (9)$$

```
> shapiro.test(residuals(aov(rendimento ~ linha+bulbo+variedade, data=ex3)))
```

Shapiro-Wilk normality test

```
data: residuals(aov(rendimento ~ linha + bulbo + variedade, data = ex3))
W = 0.93557, p-value = 0.2982
```

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.2982$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

```
> ks.test(residuals(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),  
+         "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data: residuals(aov(rendimento ~ linha + bulbo + variedade, data = ex3))  
D = 0.15129, p-value = 0.8059  
alternative hypothesis: two-sided
```

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p\text{-value} = 0.8059$ ) é maior que  $\alpha = 0.05$ , da mesma forma que observado para o teste de Shapiro-Wilk, não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal.

Para verificar a pressuposição de normalidade dos resíduos, pode-se ainda utilizar o gráfico de quantis (qqPlot) com envelope de simulação, por meio da função qqPlot, especificando os resíduos do modelo. Nesse caso, espera-se que os pontos estejam dentro do intervalo formado pelo envelope simulado.

```
> qqPlot(rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),  
+        pch=19, ylab="Resíduos")
```

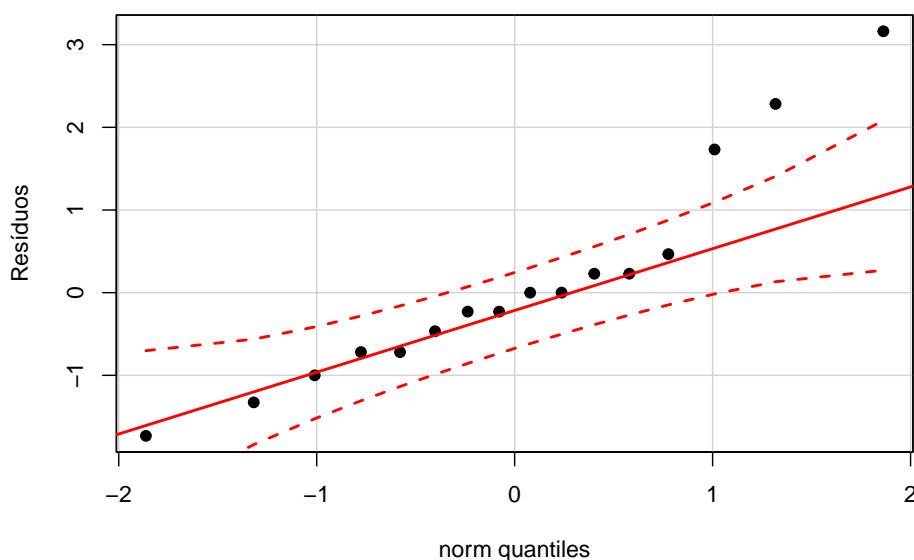


Figura 14: Gráfico qqplot para avaliar a distribuição normal dos resíduos

Observam-se alguns pontos fora do intervalo, indicando que há observações potencialmente discrepantes, embora a pressuposição de normalidade tenha sido satisfeita nos testes aplicados.

### 3. Efeito aditivo de linhas e colunas

A aditividade de linhas e colunas pode ser verificada pelo teste de não aditividade de Tukey:

```
> require(dae)
> modad<-aov(rendimento ~ linha+bulbo+variedade + Error(linha*bulbo), ex3)
> tukey.1df(modad, ex3, error.term= "linha:bulbo")
```

```
$Tukey.SS
[1] 0.3586648
```

```
$Tukey.F
[1] 0.1146529
```

```
$Tukey.p
[1] 0.748664
```

```
$Devn.SS
[1] 15.64134
```

Como o p-valor ( $Tukey.p = 0.7486$ ) é maior que o nível de significância estabelecido ( $\alpha = 5\%$ ) não se rejeita a hipótese de que linhas e colunas têm efeito aditivo no modelo. Isto significa que o efeito do tratamento não depende das linhas e das colunas.

### 4. Transformação de Box-Cox

Equivalente ao apresentado para os delineamentos anteriores, pode-se utilizar a função `boxcox` para verificar a necessidade de transformação dos dados. Neste caso deve-se verificar se o valor 1 pertence ao intervalo de confiança para  $\lambda$ .

Para extrair o valor de  $\lambda$  com o máximo de verossimilhança utilizam-se os comandos a seguir:

```
> require(MASS)
> bc <-boxcox(aov(rendimento ~ linha+bulbo+variedade, data=ex3),
+           lambda = seq(-2, 6, l = 101))
> bc$x[which.max(bc$y)]
```

```
[1] 0.96
```

```
> require(MASS)
> boxcox(aov(rendimento ~ linha+bulbo+variedade, data=ex3),
+        lambda = seq(-2, 6, l = 101))
```

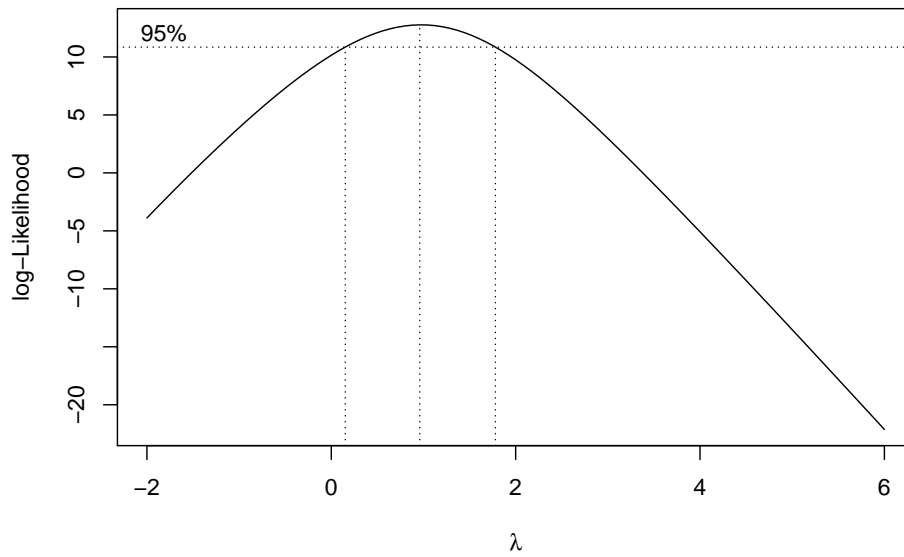


Figura 15: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Neste caso, não há evidências da necessidade de transformação.

## 5. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),
+      rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),
+      xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),
+      prob=T,main=" ", xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3))),
+      col="blue")
> ##qqPlot precisa do pacote car
> qqPlot(rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),
+      ylab="Resíduos Studentizados", pch=19)
> ##Boxplot
> boxplot(rstudent(aov(rendimento ~ linha+bulbo+variedade, data=ex3)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2,
+      cex.axis=1.2, col="turquoise")

```

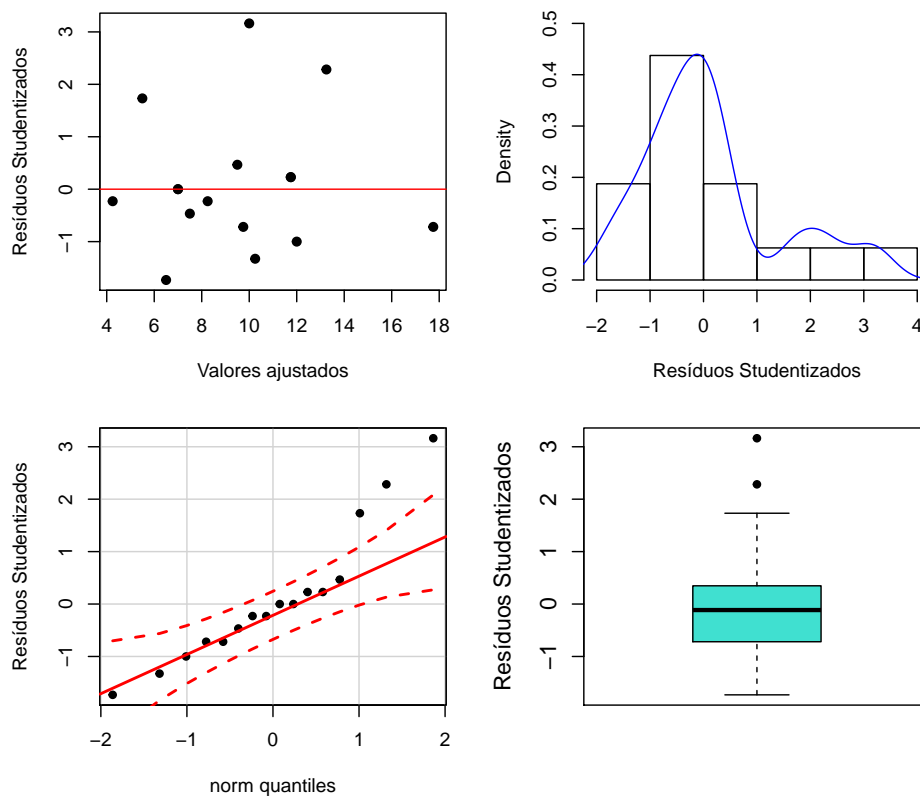


Figura 16: Gráficos de diagnósticos para verificar as pressuposições da análise de variância no delineamento em quadrado latino

Como todas as pressuposições da análise de variância foram satisfeitas, pode-se prosseguir com a análise.

### 3.4.4 Análise de variância

Pode-se utilizar a função `aov` especificando no modelo os efeitos de linhas, colunas e de variedade. É importante verificar se linhas e colunas são consideradas como fatores qualitativos.

```
> mod3 <- aov(rendimento ~ linha+bulbo+variedade, data=ex3)
> summary(mod3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
linha	3	25.5	8.500	3.188	0.10555
bulbo	3	55.5	18.500	6.938	0.02235 *
variedade	3	89.0	29.667	11.125	0.00728 **
Residuals	6	16.0	2.667		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

O p-valor para efeito de linhas não foi significativo ( $p - value = 0,10555$ ), não há evidências de efeito de linhas na variável testada. O efeito de bulbos (colunas) foi significativo ( $p - value = 0,0223$ ), o que indica que pelo menos dois tamanhos de bulbos diferem entre si quanto ao seu efeito no rendimento. O efeito de variedade foi significativo ( $p - value = 0,0072$ ), o que indica que pelo menos duas variedades diferem entre si quanto ao seu efeito no rendimento de bulbos. Neste caso, segue-se com um teste de comparação das médias.

### 3.4.5 Teste de comparações múltiplas

A função `TukeyHSD` pode ser utilizada para construir intervalos de confiança para as diferenças entre as médias das variedades.

```
> ex3.HSDM <- TukeyHSD(mod3, "variedade", ordered = TRUE)
> ex3.HSDM
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
factor levels have been ordered
```

```
Fit: aov(formula = rendimento ~ linha + bulbo + variedade, data = ex3)
```

```
$variedade
```

	diff	lwr	upr	p adj
Caçador-Contestado	0.5	-3.49724	4.49724	0.9705511
Quitéria-Contestado	2.5	-1.49724	6.49724	0.2348039
Chonan-Contestado	6.0	2.00276	9.99724	0.0080742
Quitéria-Caçador	2.0	-1.99724	5.99724	0.3855911
Chonan-Caçador	5.5	1.50276	9.49724	0.0123008
Chonan-Quitéria	3.5	-0.49724	7.49724	0.0829038

Observa-se, por exemplo, que as variedades Caçador e Contestado não diferem significativamente ( $p_{adj} = 0.9705$ ), enquanto a variedade Chonan produz em média 6,0 t/ha a mais que a variedade Contestado, sendo esta diferença significativa ( $p_{adj} = 0.0080$ ).

Os intervalos de confiança podem ser visualizados de forma gráfica.

```
> par(mar=c(5.1,10,4.1,2.1))
> plot(ex3.HSDM, las=2, cex.axis=0.6)
> par(mar=c(5.1,10,4.1,2.1))
```

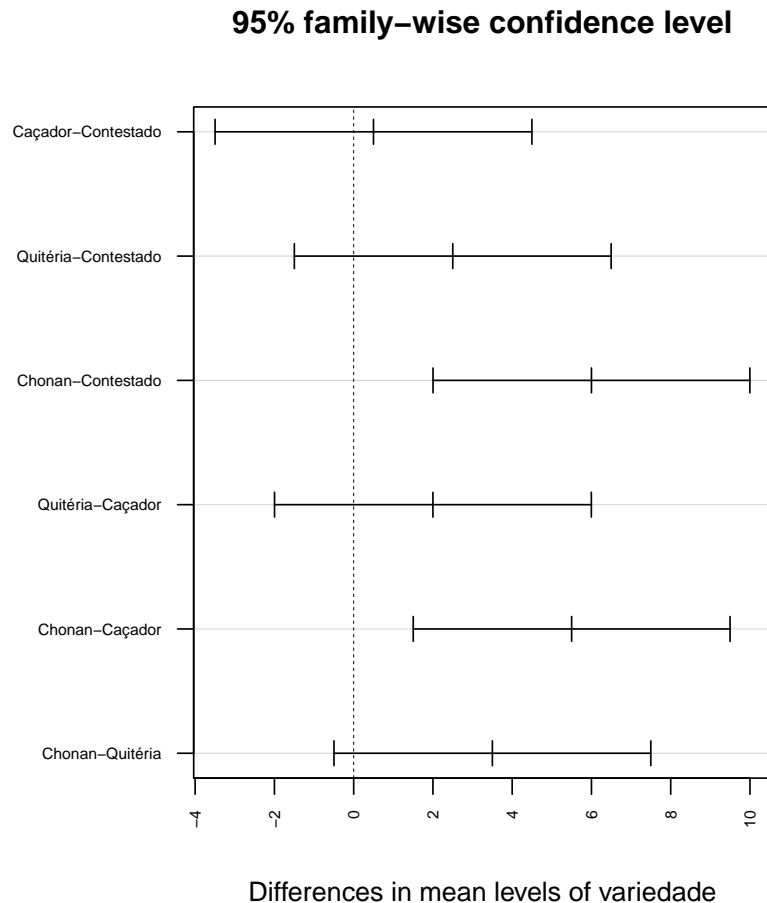


Figura 17: Intervalos de confiança para as diferenças entre as médias das variedades - teste de Tukey

No gráfico, intervalos de confiança que contêm o valor 0 (linha pontilhada), indicam que a diferença entre os genótipos não é significativa.

A função `HSD.test`, do pacote `agricolae`, também realiza o teste de Tukey apresentando as médias seguidas por letras.

```
> with(ex3, HSD.test(mod3, "variedade", alpha = 0.05, group=T, console=T))
```



Study: mod3 ~ "variedade"

HSD Test for rendimento

Mean Square Error: 2.666667

variedade, means

	rendimento	std	r	Min	Max
Caçador	7.75	2.986079	4	5	12
Chonan	13.25	3.500000	4	9	17
Contestado	7.25	2.500000	4	4	10
Quitéria	9.75	2.217356	4	7	12

Alpha: 0.05 ; DF Error: 6

Critical Value of Studentized Range: 4.895599

Minimum Significant Difference: 3.99724

Treatments with the same letter are not significantly different.

	rendimento	groups
Chonan	13.25	a
Quitéria	9.75	ab
Caçador	7.75	b
Contestado	7.25	b

A análise com o pacote factorplot segue a seguir:

```
> require(factorplot)
> fpM <- factorplot(mod3, factor.variable = "variedade", pval = 0.05,
+                   two.sided = TRUE, order = "natural",
+                   adjust.method = "holm")
> print(fpM)
```

		Difference	SE	p.val
Caçador -	Chonan	-5.500	1.155	0.016
Caçador -	Contestado	0.500	1.155	0.680
Chonan -	Contestado	6.000	1.155	0.012
Caçador -	Quitéria	-2.000	1.155	0.268
Chonan -	Quitéria	3.500	1.155	0.092
Contestado -	Quitéria	-2.500	1.155	0.221

```
> plot(fpM, abbrev.char=12, cex=0.5)
```

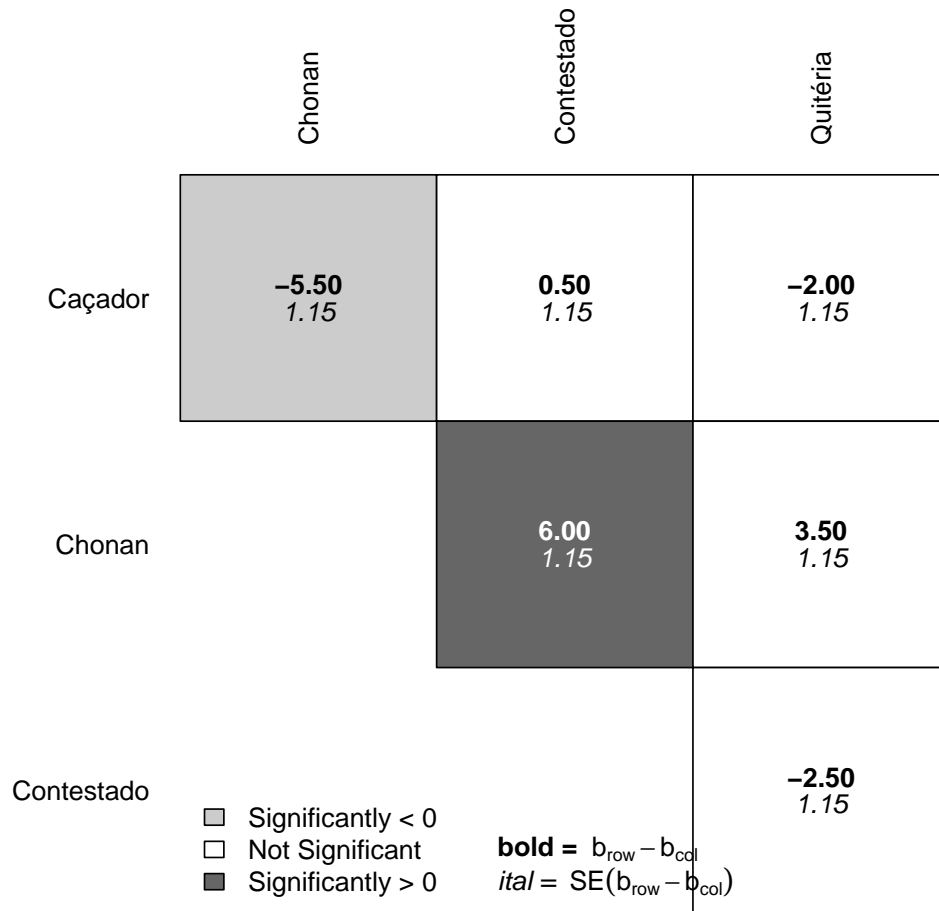


Figura 18: Comparações entre cultivares apresentadas em gráfico utilizando o pacote `factorplot`

### 3.4.6 Análise de com o pacote `ExpDes.pt`

Assim como nos demais delineamentos balanceados pode-se utilizar a função `dql`, disponível no pacote `ExpDes.pt`, que executa a análise de variância e o teste de comparação de médias.

```
> require(ExpDes.pt)
> with(ex3, dql(variedade, linha, bulbo, rendimento, quali = TRUE,
+             mcomp = "tukey", sigT = 0.05, sigF = 0.05))
```

-----  
 Quadro da análise de variancia  
 -----

	GL	SQ	QM	Fc	Pr>Fc
Tratamento	3	89.0	29.6667	11.1250	0.007283

Linha	3	25.5	8.5000	3.1875	0.105554
Coluna	3	55.5	18.5000	6.9375	0.022354
Residuo	6	16.0	2.6667		
Total	15	186.0			

-----

CV = 17.19 %

-----

Teste de normalidade dos residuos (Shapiro-Wilk)

valor-p: 0.2982307

De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser con

-----

Teste de Tukey

-----

Grupos Tratamentos Medias

a	Chonan	13.25
ab	Quitéria	9.75
b	Caçador	7.75
b	Contestado	7.25

-----

A função realiza automaticamente o teste de Shapiro-Wilk para verificar se os resíduos provêm de uma distribuição normal.

### 3.5 Esquema Fatorial

A seguir são apresentados os resultados de um experimento conduzido no delineamento em blocos casualizados com tratamentos no esquema fatorial, com objetivo de verificar o efeito da adubação de base e de cobertura na produtividade de laranjas.

Tabela 4: Avaliação de fontes de adubação nitrogenada na base e de adubos químicos em cobertura aplicados e em laranjeiras cultivar 'Açúcar'

Adubação		Peso de frutos (kg/planta)			Número de frutos/planta		
Base	Cobertura	Bloco 1	Bloco 2	Bloco 3	Bloco 1	Bloco 2	Bloco 3
Adubo químico	Nada	205	192	225	1508	1263	1573
	Somente N	278	255	338	1872	1826	2377
	1/2 dose NPK	275	270	243	1954	1766	1643
	1 dose NPK	267	259	315	1752	1848	2202
Esterco de aves	Nada	316	312	308	2000	2076	2142
	Somente N	304	291	302	1909	1914	1957
	1/2 dose NPK	307	324	318	1975	2198	2087
	1 dose NPK	329	302	308	2231	1926	2100
Esterco de suínos	Nada	269	282	266	1739	1834	1759
	Somente N	300	309	364	2013	2176	2210
	1/2 dose NPK	324	293	312	2150	2002	2113
	1 dose NPK	320	324	342	2105	2158	2207

### 3.6 Importando os dados

```
> #ex4 <- read.xlsx("dados.xlsx", sheet="ex4")
> ex4 <- read.csv("ex4.csv", dec=",", header = T)
> head(ex4)
```

```
      base cobertura bloco FRUTOS PESO
1 Adubo químico      nada      1  1508  205
2 Adubo químico      nada      2  1263  192
3 Adubo químico      nada      3  1573  225
4 Adubo químico Somente N      1  1872  278
5 Adubo químico Somente N      2  1826  255
6 Adubo químico Somente N      3  2377  338
```

```
> ex4$bloco <- as.factor(ex4$bloco)
> ex4$base <- as.factor(ex4$base)
> ex4$cobertura <- as.factor(ex4$cobertura)
```

### 3.6.1 Análise exploratória dos dados

A análise exploratória dos dados pode ser realizada como os exemplos anteriores.

```
> str(ex4)
```

```
'data.frame':      36 obs. of  5 variables:
 $ base      : Factor w/ 3 levels "Adubo químico",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ cobertura: Factor w/ 4 levels "Meia dose NPK",...: 2 2 2 3 3 3 1 1 1 4 ...
 $ bloco     : Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
 $ FRUTOS    : int  1508 1263 1573 1872 1826 2377 1954 1766 1643 1752 ...
 $ PESO      : int  205 192 225 278 255 338 275 270 243 267 ...
```

```
> summary(ex4)
```

	base	cobertura	bloco	FRUTOS
Adubo químico	:12	Meia dose NPK:9	1:12	Min. :1263
Esterco de aves	:12	nada :9	2:12	1st Qu.:1832
Esterco de suínos	:12	Somente N :9	3:12	Median :1988
		Uma dose NPK :9		Mean :1960
				3rd Qu.:2144
				Max. :2377
PESO				
Min.	:192.0			
1st Qu.	:269.8			
Median	:303.0			
Mean	:293.0			
3rd Qu.	:316.5			
Max.	:364.0			

O gráfico box-plot pode ser utilizado da mesma forma apresentada anteriormente. Neste caso, especificando a variável resposta em função dos fatores de agrupamento (bloco, cobertura e base).

```

> par(mfrow=c(2,2),mar=c(8,4,0,0))
> mar=c(4,3,1,1)
> with(ex4, boxplot(PESO ~ bloco, xlab="Blocos",
+                   ylab="Peso de frutos (kg/planta)",
+                   ylim=c(150,400), las=1, cex.axis=0.6, horizontal = F))
> mar=c(4,3,1,1)
> with(ex4, boxplot(PESO ~ cobertura, xlab="Cobertura",
+                   ylab="Peso de frutos (kg/planta)",
+                   ylim=c(150,400), las=1, cex.axis=0.6, horizontal = F))
> mar=c(4,3,1,1)
> with(ex4, boxplot(PESO ~ base, xlab="Base",
+                   ylab="Peso de frutos (kg/planta)",
+                   ylim=c(150,400), las=1, cex.axis=0.6, horizontal = F))
> mar=c(8,3,1,1)
> with(ex4, boxplot(PESO ~ cobertura*base, xlab="Interação",
+                   ylab="Peso de frutos (kg/planta)",
+                   ylim=c(150,400), las=2, cex.axis=0.6, horizontal = F))

```

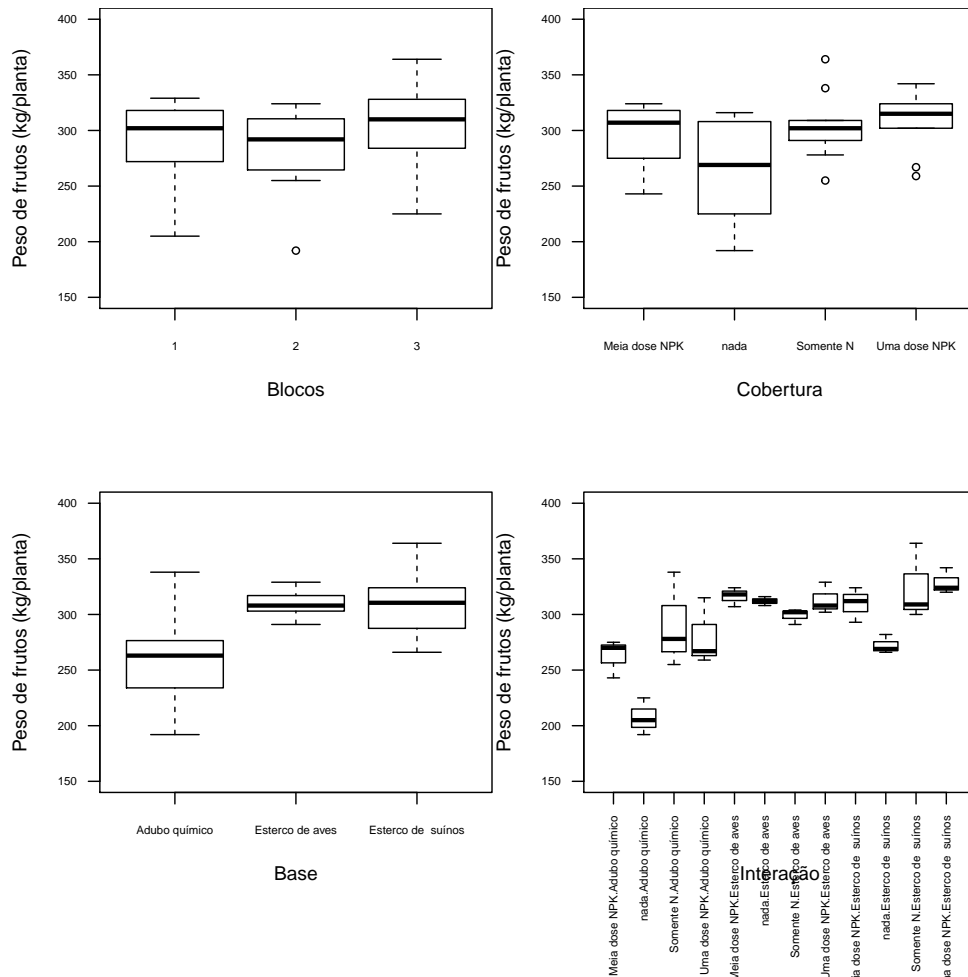


Figura 19: Gráficos de caixa (*boxplots*) dos dados de peso de frutos (kg/planta) em função dos blocos, adubação de base e adubação de cobertura

### 3.6.2 Verificação das pressuposições do modelo estatístico

Para que a análise de variância seja válida, devem ser satisfeitas as pressuposições de homogeneidade de variâncias entre tratamentos, normalidade dos resíduos e aditividade de blocos.

#### 1. Homogeneidade de Variâncias

Para verificar a homogeneidade de variâncias entre tratamentos, a hipótese a ser testada e os testes a serem empregados são os mesmos apresentados para os exemplos anteriores.

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

Sempre que o nível descritivo do teste (*p-value*) for menor que o nível de significância estabelecido ( $\alpha$ ), rejeita-se  $H_0$ . Caso contrário, não se rejeita indicando que não existem evidências contra a homocedasticidade.

```
> bartlett.test(residuals(aov(PESO ~ bloco, data=ex4))
+               ~ cobertura , data=ex4)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(PESO ~ bloco, data = ex4)) by cobertura
Bartlett's K-squared = 4.1219, df = 3, p-value = 0.2486
```

```
> bartlett.test(residuals(aov(PESO ~ bloco+ cobertura, data=ex4))
+               ~ base, data=ex4)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(PESO ~ bloco + cobertura, data = ex4)) by base
Bartlett's K-squared = 2.5307, df = 2, p-value = 0.2821
```

```
> bartlett.test(residuals(aov(PESO ~ bloco+ cobertura+base, data=ex4))
+               ~ interaction(cobertura,base), data=ex4)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(PESO ~ bloco + cobertura + base, data = ex4)) by interaction
Bartlett's K-squared = 9.745, df = 11, p-value = 0.5535
```

No caso do esquema fatorial, deve-se verificar a homogeneidade de variâncias dos fatores principais e das combinações. Para o teste de Bartlett, como o nível descritivo do teste é maior que o nível de significância considerado ( $\alpha = 0.05$ ) em todos os casos testados não se rejeita a hipótese de homogeneidade de variâncias.

```

> require(car)
> leveneTest(residuals(aov(PESO ~ bloco, data=ex4))
+             ~ cobertura , data=ex4)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  2.0794 0.1225
      32

> leveneTest(residuals(aov(PESO ~ bloco+ cobertura, data=ex4))
+             ~ base, data=ex4)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.9501  0.397
      33

> leveneTest(residuals(aov(PESO ~ bloco+ cobertura+base, data=ex4))
+             ~ interaction(cobertura,base), data=ex4)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 11   0.39  0.947
      24

```

Para o teste de Levene, também não se rejeita a hipótese de homogeneidade de variâncias, pois o nível descritivo do teste é maior que o nível de significância ( $\alpha = 0.05$ ) em todos os casos.

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, \sigma)$ )

A hipótese a ser testada neste caso é a mesma apresentada para os casos dos delineamentos anteriores, dada por:

$$\begin{cases} H_0 : & \text{Os resíduos do modelo provêm de uma população com distribuição normal} \\ H_1 : & \text{Os resíduos do modelo não provêm de uma população com distribuição normal} \end{cases} \quad (11)$$

```

> shapiro.test(rstudent(aov(PESO ~ bloco+cobertura*base,
+                          data=ex4)))

```

Shapiro-Wilk normality test

```

data:  rstudent(aov(PESO ~ bloco + cobertura * base, data = ex4))
W = 0.97767, p-value = 0.6663

```



No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.6663$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

```
> ks.test(rstudent(aov(PESO ~ bloco+cobertura*base,  
+ data=ex4)), "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data: rstudent(aov(PESO ~ bloco + cobertura * base, data = ex4))  
D = 0.13981, p-value = 0.4424  
alternative hypothesis: two-sided
```

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p\text{-value} = 0.4424$ ) é maior que  $\alpha = 0.05$ , da mesma forma que observado para o teste de Shapiro-Wilk, não existem evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

Para verificar a pressuposição de normalidade dos resíduos, pode-se ainda utilizar o gráfico de quantis (qqPlot) com envelope de simulação, por meio da função `qqPlot`, especificando os resíduos do modelo. Nesse caso, espera-se que os pontos estejam dentro do intervalo formado pelo envelope simulado.

```
> qqPlot(rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+         pch=19, ylab="Resíduos")
```

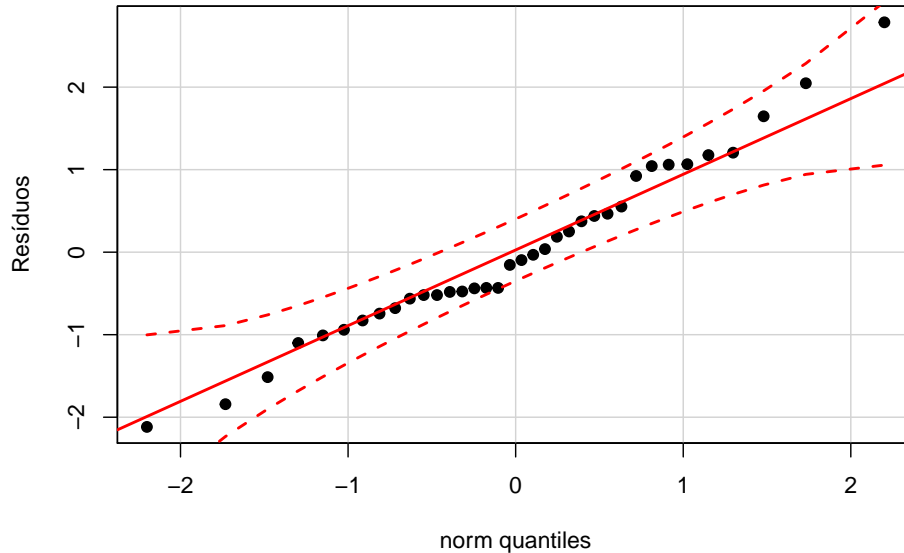


Figura 20: Gráfico qqplot para avaliar a distribuição normal dos resíduos num esquema fatorial em DBC

### 3. Bloco de efeito aditivo

A aditividade de blocos pode ser verificada pelo teste de não aditividade de Tukey:

```
> require(dae)
> modad<-aov(PESO ~ bloco+cobertura*base + Error(bloco), ex4)
> tukey.1df(modad, ex4, error.term= "Within")
```

```
$Tukey.SS
[1] 1.480905
```

```
$Tukey.F
[1] 0.003696093
```

```
$Tukey.p
[1] 0.952097
```

```
$Devn.SS
[1] 8414.019
```

Como o p-valor ( $Tukey.p = 0.9520$ ) é maior que o nível de significância estabelecido ( $\alpha = 0.05$ ) não se rejeita a hipótese de que os blocos têm efeito aditivo no modelo. Isto significa que o efeito do tratamento não depende do bloco.

#### 4. Transformação de Box-Cox

Equivalente ao apresentado anteriormente, pode-se utilizar a função `boxcox` para verificar a necessidade de transformação dos dados. Neste caso deve-se verificar se o valor 1 pertence ao intervalo de confiança para  $\lambda$ .

```
> require(MASS)
> boxcox(aov(PESO ~ bloco+cobertura*base, data=ex4),
+       lambda = seq(-4, 4, l = 101))
```

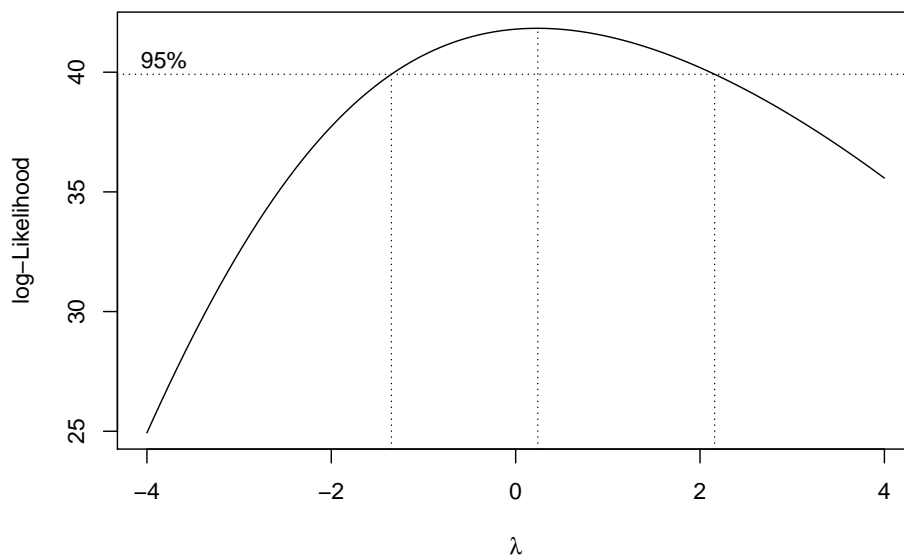


Figura 21: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Neste caso, não há evidências da necessidade de transformação.

#### 5. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+      rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+      xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+      prob=T,main=" ", xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4))),
+      col="blue")
> ##qqPlot precisa do pacote car
> qqPlot(rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+      ylab="Resíduos Studentizados", pch=19)
> ##Boxplot
> boxplot(rstudent(aov(PESO ~ bloco+cobertura*base, data=ex4)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2,
+      cex.axis=1.2, col="turquoise")

```

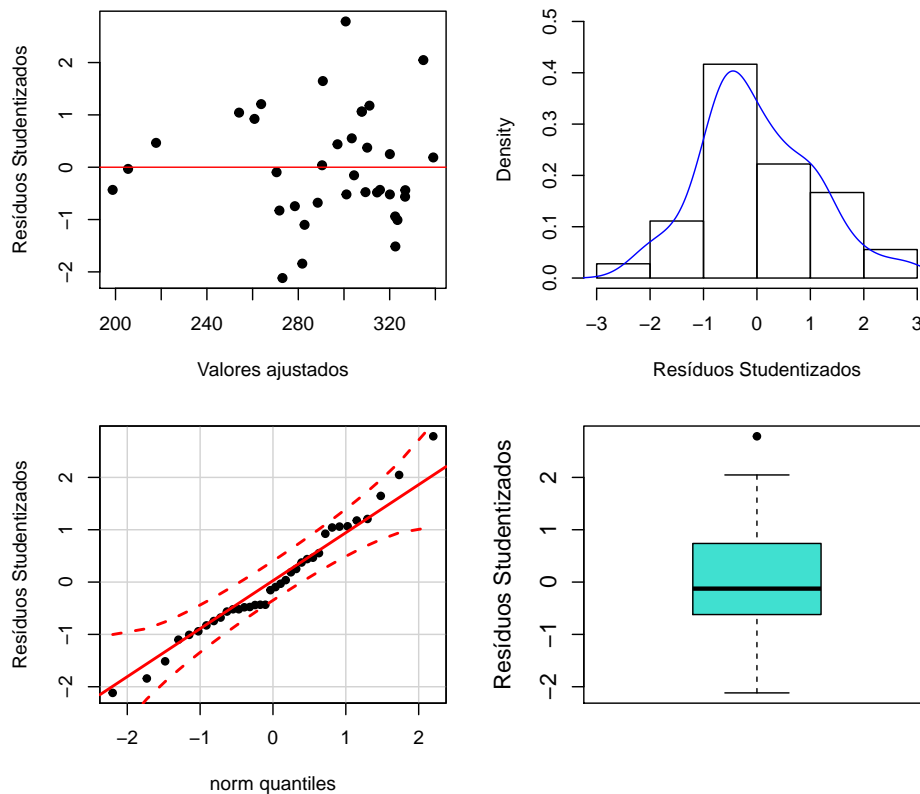


Figura 22: Gráficos de diagnósticos para verificar as pressuposições

Como todas as pressuposições da análise de variância foram satisfeitas, pode-se prosseguir com a análise.

### 3.6.3 Análise de variância

Pode-se utilizar a função `aov` especificando no modelo os efeitos de blocos, cobertura e adubo de base.

```
> mod4 <- aov(PESO ~ bloco+ cobertura*base, data=ex4)
> summary(mod4)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
cobertura	3	10771	3590	9.386	0.000346 ***
base	2	19415	9708	25.378	1.93e-06 ***
cobertura:base	6	7994	1332	3.483	0.014217 *
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

O p-valor para efeito de blocos não foi significativo ( $p - value = 0.0756$ ), não há evidências de efeito de blocos na variável testada. Como o efeito da interação entre adubação de cobertura e adubação de base foi significativo ( $p - value = 0.0142$ ), não se deve considerar os efeitos individuais, pois isto indica que o efeito de um fator depende dos níveis do outro fator.

Para realizar o desdobramento de cobertura dentro de cada nível da base ou base dentro de cada nível de cobertura, utiliza-se um artifício de criar uma variável auxiliar que fornece todas as combinações entre cobertura e base, isso é possível utilizando o co-mando `interaction`. No modelo de análise de variância será utilizado o fator que não será desdobrado e a combinação dos fatores. Posteriormente, para visualizar os testes do desdobramento, será utilizado o `summary` com a opção `split`, na qual devem ser designados os graus de liberdade associados a cada nível, segundo o observado nos coeficientes do modelo. Neste caso, o desdobramento da interação será feito da seguinte forma:

```
> cb=with(ex4,interaction(cobertura, base))
> cob_d_base=aov(PESO ~ bloco + base + cb , ex4)
> summary(cob_d_base) ### modelo que soma cobertura
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
base	2	19415	9708	25.378	1.93e-06 ***
cb	9	18765	2085	5.451	0.000565 ***
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
> # e interação base*cobertura
> coef(cob_d_base)
```

	(Intercept)		bloco2
	260.833333		-6.750000
		bloco3	baseEsterco de aves
	12.250000		50.333333
	baseEsterco de suínos		cbnada.Adubo químico

```

66.000000 -55.333333
cbSomente N.Adubo químico cbUma dose NPK.Adubo químico
27.666667 17.666667
cbMeia dose NPK.Esterco de aves cbnada.Esterco de aves
3.333333 -1.000000
cbSomente N.Esterco de aves cbMeia dose NPK.Esterco de suínos
-14.000000 -19.000000
cbnada.Esterco de suínos cbSomente N.Esterco de suínos
-56.333333 -4.333333

```

```

> summary(cob_d_base, split = list( cb = list(cob_d_AQ = 1:3,
+                                           cob_d_EA = 4:6,
+                                           cob_d_ES = 7:9)))

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
base	2	19415	9708	25.378	1.93e-06 ***
cb	9	18765	2085	5.451	0.000565 ***
cb: cob_d_AQ	3	12343	4114	10.756	0.000149 ***
cb: cob_d_EA	3	522	174	0.455	0.716355
cb: cob_d_ES	3	5900	1967	5.141	0.007605 **
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

> base_d_cob=aov(PESO ~ bloco+ cobertura + cb , ex4)
> summary(base_d_cob)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
cobertura	3	10771	3590	9.386	0.000346 ***
cb	8	27409	3426	8.957	2.12e-05 ***
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

> coef(base_d_cob)

```

(Intercept)	bloco2
260.833333	-6.750000
bloco3	coberturanada
12.250000	9.666667
coberturaSomente N	coberturaUma dose NPK
61.666667	66.000000
cbnada.Adubo químico	cbSomente N.Adubo químico
-65.000000	-34.000000
cbUma dose NPK.Adubo químico	cbMeia dose NPK.Esterco de aves
-48.333333	53.666667
cbnada.Esterco de aves	cbSomente N.Esterco de aves

```

39.666667 -25.333333
cbUma dose NPK.Esterco de aves cbMeia dose NPK.Esterco de suínos
-15.666667 47.000000

```

```

> summary(base_d_cob, split = list( cb = list(base_d_Nada = c(1,5),
+                                           base_d_N =c(2,6) ,
+                                           base_d_meioNPK = c(4,8) ,
+                                           base_d_NPK=c(3,7))))

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
cobertura	3	10771	3590	9.386	0.000346 ***
cb	8	27409	3426	8.957	2.12e-05 ***
cb: base_d_Nada	2	16754	8377	21.899	5.84e-06 ***
cb: base_d_N	2	1873	936	2.448	0.109662
cb: base_d_meioNPK	2	5134	2567	6.710	0.005307 **
cb: base_d_NPK	2	3649	1824	4.769	0.019029 *
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```

> summary(mod4);summary(base_d_cob);summary(cob_d_base)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
cobertura	3	10771	3590	9.386	0.000346 ***
base	2	19415	9708	25.378	1.93e-06 ***
cobertura:base	6	7994	1332	3.483	0.014217 *
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
cobertura	3	10771	3590	9.386	0.000346 ***
cb	8	27409	3426	8.957	2.12e-05 ***
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	2	2227	1113	2.910	0.075618 .
base	2	19415	9708	25.378	1.93e-06 ***
cb	9	18765	2085	5.451	0.000565 ***
Residuals	22	8416	383		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Verifica-se que não existe efeito significativo da adubação em cobertura, quando se utiliza esterco de aves como adubação de base. Entretanto, quando a adubação de base é feita com adubo químico ou esterco de suínos, há diferenças entre as adubações de cobertura.

### 3.6.4 Análise com o pacote ExpDes.pt

Pode-se utilizar a função `fat2.dbc`, disponível no pacote `ExpDes.pt`, que executa a análise de variância, desdobramento da interação e o teste de comparação de médias.

```
> require(ExpDes.pt)
> with(ex4, fat2.dbc(cobertura, base, bloco, PESO, quali = c(TRUE, TRUE),
+                  mcomp = "tukey", fac.names = c("Cobertura", "Base"),
+                  sigT = 0.05, sigF = 0.05))
```

---

Legenda:

FATOR 1: Cobertura

FATOR 2: Base

---

Quadro da análise de variancia

---

	GL	SQ	QM	Fc	Pr>Fc
Bloco	2	2227	1113.2	2.9103	0.075618
Cobertura	3	10771	3590.4	9.3862	0.000346
Base	2	19415	9707.6	25.3778	0.000002
Cobertura*Base	6	7994	1332.2	3.4828	0.014217
Residuo	22	8416	382.5		
Total	35	48822			

---

CV = 6.68 %

---

Teste de normalidade dos residuos (Shapiro-Wilk)

valor-p: 0.7081692

De acordo com o teste de Shapiro-Wilk a 5% de significancia, os residuos podem ser con

---

Interacao significativa: desdobrando a interacao

---

Desdobrando Cobertura dentro de cada nivel de Base



-----  
-----  
Quadro da análise de variancia  
-----

	GL	SQ	QM	Fc	Pr.Fc
Bloco	2	2226.500	1113.2500	2.9103	0.0756
Base	2	19415.167	9707.5833	25.3778	0
Cobertura:Base Adubo químico	3	12343.000	4114.3333	10.7558	1e-04
Cobertura:Base Esterco de aves	3	522.250	174.0833	0.4551	0.7164
Cobertura:Base Esterco de suínos	3	5899.583	1966.5278	5.1409	0.0076
Residuo	22	8415.500	382.5227		
Total	35	48822.000	1394.9143		

-----

Cobertura dentro do nivel Adubo químico de Base  
-----

Teste de Tukey  
-----

Grupos Tratamentos Medias

a	3	290.3333
a	4	280.3333
a	1	262.6667
b	2	207.3333

-----

Cobertura dentro do nivel Esterco de aves de Base

De acordo com o teste F, as medias desse fator sao estatisticamente iguais.  
-----

	Niveis	Medias
1	1	316.3333
2	2	312.0000
3	3	299.0000
4	4	313.0000

-----

Cobertura dentro do nivel Esterco de suínos de Base  
-----

Teste de Tukey  
-----

Grupos Tratamentos Medias

a	4	328.6667
a	3	324.3333
ab	1	309.6667

b                    2                    272.3333

---

Desdobrando Base dentro de cada nivel de Cobertura

---

---

Quadro da analise de variancia

---

	GL	SQ	QM	Fc	Pr.Fc
Bloco	2	2226.500	1113.2500	2.9103	0.0756
Cobertura	3	10771.333	3590.4444	9.3862	3e-04
Base:Cobertura Meia dose NPK	2	5133.556	2566.7778	6.7101	0.0053
Base:Cobertura nada	2	16753.556	8376.7778	21.8988	0
Base:Cobertura Somente N	2	1872.889	936.4444	2.4481	0.1097
Base:Cobertura Uma dose NPK	2	3648.667	1824.3333	4.7692	0.019
Residuo	22	8415.500	382.5227		
Total	35	48822.000	1394.9143		

---

Base dentro do nivel Meia dose NPK de Cobertura

---

Teste de Tukey

---

Grupos Tratamentos Medias

a	2	316.3333
a	3	309.6667
b	1	262.6667

---

Base dentro do nivel nada de Cobertura

---

Teste de Tukey

---

Grupos Tratamentos Medias

a	2	312
a	3	272.3333
b	1	207.3333

---

Base dentro do nivel Somente N de Cobertura

De acordo com o teste F, as medias desse fator sao estatisticamente iguais.

---

	Níveis	Medias
1	1	290.3333
2	2	299.0000
3	3	324.3333

---

Base dentro do nível Uma dose NPK de Cobertura

---

Teste de Tukey

---

Grupos	Tratamentos	Medias
a	3	328.6667
ab	2	313
b	1	280.3333

---

É usual em experimentos no esquema fatorial apresentar as médias em uma tabela de dupla entrada, tendo um fator na linha e o outro na coluna, de forma que as diferenças significativas entre os níveis dos fatores são representadas por letras minúsculas/maiúsculas nas linhas/colunas.

Tabela 5: Peso médio dos frutos de laranja (kg/planta) em função da adubação de base e da adubação de cobertura. Médias seguidas por letras maiúsculas iguais na coluna e minúsculas iguais na linha não diferem entre si

	Meia dose NPK	nada	Somente N	Uma dose NPK
Adubo químico	262,67aB	207,33bB	290,33aA	280,33aB
Esterco de aves	316,33aA	312,00aA	299,00aA	313,00aAB
Esterco de suínos	309,67abA	272,33aA	324,33aA	328,67aA

### 3.7 Parcela Subdividida

Realizou-se um experimento para avaliar o efeito de fontes de adubo nitrogenado em dois sistemas de plantio na produtividade de milho. O delineamento experimental foi em blocos casualizados com quatro repetições, e os tratamentos dispostos em esquema de parcelas subdivididas, com sistema de plantio na parcela e fonte de nitrogênio na subparcela.

Tabela 6: Produtividade de milho (kg/ha) em função do sistema de plantio e da fonte de nitrogênio

Sistema de Plantio	Fonte de Nitrogênio	Blocos			
		1	2	3	4
Convencional	Esterco de Aves	9317	9617	10889	10753
	Esterco de Suínos	8881	10347	12204	9373
	Testemunha	5146	5801	6003	6991
	Nitrato	9308	11200	9720	11305
Direto	Esterco de Aves	8045	8012	8121	8656
	Esterco de Suínos	9171	12312	10373	11752
	Testemunha	5222	6087	5536	6474
	Nitrato	11233	9709	11086	10845

#### 3.7.1 Importando os dados

```
> #ex5 <- read.xlsx("dados.xlsx", sheet="ex5")
> ex5 <- read.csv("ex5.csv", dec=",", header = T)
> head(ex5)
```

```
      sistema      fonte bloco produtividade
1 convencional      Eaves      1           9317
2 convencional      Esuinos     1           8881
3 convencional Testemunha     1           5146
4 convencional      Nitrato     1           9308
5 convencional      Eaves      2           9617
6 convencional      Esuinos     2          10347
```

#### 3.7.2 Análise exploratória dos dados

A análise exploratória dos dados pode ser realizada como os exemplos anteriores.

```
> str(ex5)

'data.frame':      32 obs. of  4 variables:
 $ sistema      : Factor w/ 2 levels "convencional",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ fonte        : Factor w/ 4 levels "Eaves","Esuinos",...: 1 2 4 3 1 2 4 3 1 2 ...
 $ bloco        : int  1 1 1 1 2 2 2 2 3 3 ...
 $ produtividade: int  9317 8881 5146 9308 9617 10347 5801 11200 10889 12204 ...
```

```
> summary(ex5)
```

	sistema	fonte	bloco	produtividade
convencional:	16	Eaves :8	Min. :1.00	Min. : 5146
direto :	16	Esuinos :8	1st Qu.:1.75	1st Qu.: 7757
		Nitrato :8	Median :2.50	Median : 9345
		Testemunha:8	Mean :2.50	Mean : 9047
			3rd Qu.:3.25	3rd Qu.:10856
			Max. :4.00	Max. :12312

O bloco não é considerado como um fator, o que precisa ser corrigido antes de se prosseguir as análises.

```
> ex5$bloco <- as.factor(ex5$bloco)
> str(ex5)
```

```
'data.frame':      32 obs. of  4 variables:
 $ sistema      : Factor w/ 2 levels "convencional",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ fonte        : Factor w/ 4 levels "Eaves","Esuinos",...: 1 2 4 3 1 2 4 3 1 2 ...
 $ bloco        : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 2 2 2 2 3 3 ...
 $ produtividade: int  9317 8881 5146 9308 9617 10347 5801 11200 10889 12204 ...
```

Observa-se que agora o bloco foi transformado em um fator.

O gráfico box-plot pode ser utilizado da mesma forma apresentada anteriormente. Neste caso, especificando a variável resposta em função dos fatores de agrupamento (bloco, sistema e fonte).

```

> par(mfrow=c(2,2),mar=c(4,4,1,1))
> mar=c(4,3,1,1)
> with(ex5, boxplot(produtividade ~ bloco, xlab="Blocos",
+                  ylab="Rendimento de milho (kg/ha)",
+                  ylim=c(5000,13000), las=1, cex.axis=0.6, horizontal = F))
> mar=c(4,3,1,1)
> with(ex5, boxplot(produtividade ~ sistema, xlab="Sistema",
+                  ylab="Rendimento de milho (kg/ha)", ylim=c(5000,13000),
+                  las=2, cex.axis=0.6, horizontal = F))
> mar=c(4,3,1,1)
> with(ex5, boxplot(produtividade ~ fonte, xlab="Fonte",
+                  ylab="Rendimento de milho (kg/ha)", ylim=c(5000,13000),
+                  las=2, cex.axis=0.6, horizontal = F))
> mar=c(4,3,1,1)
> with(ex5, boxplot(produtividade ~ sistema*fonte, xlab="Interação",
+                  ylab="Rendimento de milho (kg/ha)", ylim=c(5000,13000),
+                  las=1, cex.axis=0.6, horizontal = F))

```

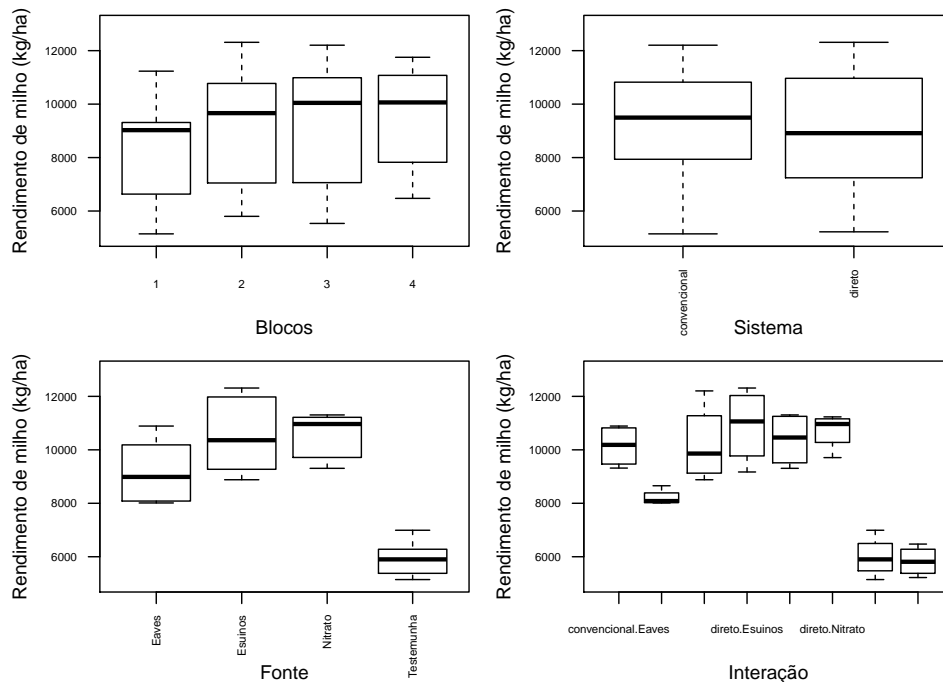


Figura 23: Gráficos de caixa (*boxplots*) dos dados de rendimento de milho (kg/ha) em função dos blocos, sistema de plantio, fonte de nitrogênio e interação entre os fatores

### 3.7.3 Verificação das pressuposições do modelo estatístico

Para que a análise de variância seja válida, devem também ser satisfeitas as pressuposições de homogeneidade de variâncias entre tratamentos, normalidade dos resíduos e aditividade de blocos.

## 1. Homogeneidade de Variâncias

Para verificar a homogeneidade de variâncias entre tratamentos, a hipótese a ser testada e os testes a serem empregados são os mesmos apresentados para os exemplos anteriores.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_t^2 \end{cases} \quad (12)$$

Sempre que o nível descritivo do teste (*p-value*) for menor que o nível de significância estabelecido ( $\alpha$ ), rejeita-se  $H_0$ . Caso contrário, não se rejeita indicando que não existem evidências contra a homocedasticidade. No caso do esquema em parcela subdividida, deve-se verificar a homogeneidade de variâncias dos fatores e das combinações.

```
> bartlett.test(residuals(aov(produtividade ~ bloco, data=ex5))
+               ~ sistema , data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(produtividade ~ bloco, data = ex5)) by sistema
Bartlett's K-squared = 0.14015, df = 1, p-value = 0.7081
```

```
> bartlett.test(residuals(aov(produtividade ~ bloco+ sistema,
+                           data=ex5))
+               ~ fonte, data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(produtividade ~ bloco + sistema, data = ex5)) by fonte
Bartlett's K-squared = 9.3929, df = 3, p-value = 0.0245
```

```
> bartlett.test(residuals(aov(produtividade ~ bloco+ sistema+fonte,
+                           data=ex5))
+               ~ interaction(sistema,fonte), data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(produtividade ~ bloco + sistema + fonte, data = ex5)) by int
Bartlett's K-squared = 10.016, df = 7, p-value = 0.1877
```

Para o teste de Bartlett, o nível descritivo do teste é menor que o nível de significância considerado ( $\alpha = 0.05$ ) para fonte, rejeitando-se a hipótese de homocedasticidade entre as fontes. Para os demais casos, não se rejeita a hipótese de homogeneidade.

```

> require(car)
> leveneTest(residuals(aov(produtividade ~ bloco, data=ex5))
+             ~ sistema , data=ex5)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  0.7383 0.397
      30

> leveneTest(residuals(aov(produtividade ~ bloco+ sistema, data=ex5))
+             ~ fonte, data=ex5)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  2.4603 0.08341 .
      28

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> leveneTest(residuals(aov(produtividade ~ bloco+ sistema+fonte,
+                         data=ex5))
+             ~ interaction(sistema,fonte), data=ex5)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  7  1.7475 0.1453
      24

```

Para o teste de Levene, não se rejeitou a hipótese de homogeneidade de variâncias em nenhuma situação, pois o nível descritivo do teste é maior que o nível de significância  $\alpha = 0.05$ .

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, 1)$ )

A hipótese a ser testada neste caso é a mesma apresentada para os casos anteriores, dada por:

$$\begin{cases} H_0 : & \text{Os resíduos do modelo provêm de uma população com distribuição normal;} \\ H_1 : & \text{Os resíduos do modelo não provêm de uma população com distribuição normal.} \end{cases} \quad (13)$$

```

> shapiro.test(rstudent(aov(produtividade ~
+                       bloco+sistema*fonte+bloco*sistema,
+                       data=ex5)))

```



### Shapiro-Wilk normality test

```
data: rstudent(aov(produtividade ~ bloco + sistema * fonte + bloco *  
W = 0.95601, p-value = 0.213
```

sistema

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.2130$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

```
> ks.test(rstudent(aov(produtividade ~  
+ bloco+sistema*fonte+bloco*sistema,  
+ data=ex5)),  
+ "pnorm")
```

### One-sample Kolmogorov-Smirnov test

```
data: rstudent(aov(produtividade ~ bloco + sistema * fonte + bloco *  
D = 0.091241, p-value = 0.9306  
alternative hypothesis: two-sided
```

sistema

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p\text{-value} = 0.9306$ ) é maior que  $\alpha = 0.05$ , da mesma forma que observado para o teste de Shapiro-Wilk, não existem evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

Para verificar a pressuposição de normalidade dos resíduos, pode-se ainda utilizar o gráfico de quantis (qqPlot) com envelope de simulação, por meio da função `qqPlot`, especificando os resíduos do modelo. Nesse caso, espera-se que os pontos estejam dentro do intervalo formado pelo envelope simulado.

```

> #par(mfrow=c(2,2),
> mar=c(4,4,1,1)
> qqPlot(rstudent(aov(produtividade ~ bloco+sistema*fonte+(bloco*sistema),
+ data=ex5)), pch=19, ylab="Resíduos")

```

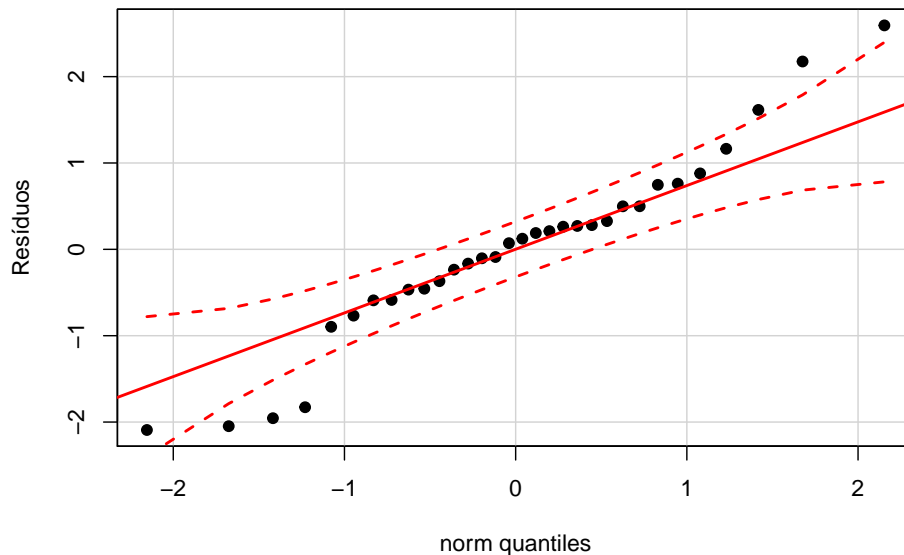


Figura 24: Gráfico *qqplot* para avaliar a distribuição normal dos resíduos

Observam-se alguns pontos fora do envelope, o que neste caso sugere a necessidade de uma inspeção nos dados.

### 3. Bloco de efeito aditivo

A aditividade de blocos pode ser verificada pelo teste de não aditividade de Tukey:

```

> require(dae)
> modad<-aov(produtividade ~ bloco+sistema*fonte +
+ Error(bloco/sistema), ex5)
> tukey.1df(modad, ex5, error.term= "Within")

```

```

$Tukey.SS
[1] 3452.05

```

```

$Tukey.F
[1] 0.004289369

```

```

$Tukey.p
[1] 0.9485453

```

```
$Devn.SS  
[1] 13681464
```

Como o p-valor ( $Tukey.p = 0.9485$ ) é maior que o nível de significância estabelecido ( $\alpha = 0.05$ ) não se rejeita a hipótese de que os blocos têm efeito aditivo no modelo. Isto significa que o efeito do tratamento não depende do bloco.

#### 4. Transformação de Box-Cox

Equivalente ao apresentado anteriormente, pode-se utilizar a função `boxcox` para verificar a necessidade de transformação dos dados. Neste caso deve-se verificar se o valor 1 pertence ao intervalo de confiança para  $\lambda$ .

```
> #par(mfrow=c(2,2),  
> mar=c(4,4,1,1)  
> require(MASS)  
> boxcox(aov(produtividade ~ bloco+sistema*fonte+bloco*sistema,  
+          data=ex5), lambda = seq(-3, 3, l = 101))
```

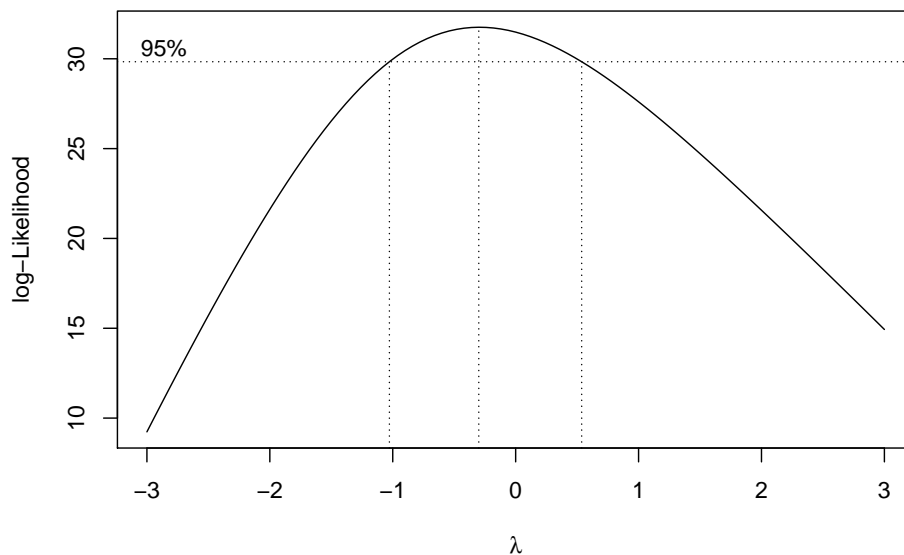


Figura 25: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Neste caso, o intervalo de confiança para  $\lambda$  não contém o valor 1, o que indica a necessidade de uma transformação nos dados. Como o valor zero pertence ao intervalo de confiança para  $\lambda$ , indica uma transformação logarítmica.

## 5. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(produtividade ~ bloco+sistema*fonte+(bloco/sistema),
+             data=ex5)),rstudent(aov(produtividade ~
+                                 bloco+sistema*fonte+(bloco/sistema),
+                                 data=ex5)),xlab="Valores ajustados",
+    ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(produtividade ~ bloco+sistema*fonte+(bloco/sistema),
+                 data=ex5)),prob=T,main=" ", xlab="Resíduos Studentizados",
+    ylim=c(0,0.5))
> lines(density(rstudent(aov(produtividade ~ bloco+sistema*fonte+(bloco/sistema),
+                           data=ex5))), col="blue")
> ##qqPlot precisa do pacote car
> qqPlot(rstudent(aov(produtividade ~ bloco+sistema*fonte+(bloco/sistema),
+                   data=ex5)), ylab="Resíduos Studentizados", pch=19)
> ##Boxplot
> boxplot(rstudent(aov(produtividade ~ bloco+sistema*fonte+(bloco/sistema),
+                   data=ex5)), ylab="Resíduos Studentizados", pch=19,
+    cex.lab=1.2, cex.axis=1.2, col="turquoise")

```

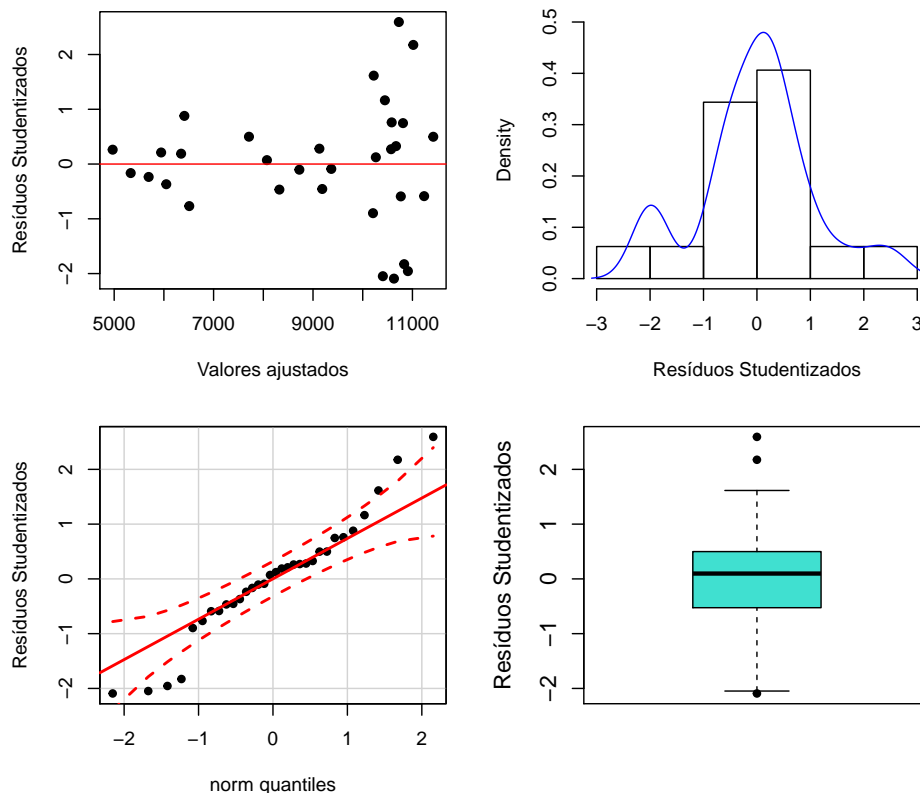


Figura 26: Gráficos de diagnósticos para verificar as pressuposições

Como a homocedasticidade não foi satisfeita e há pontos fora do envelope de confiança, há necessidade de transformação dos dados. A partir de agora serão novamente verificadas as pressuposições da análise de variância com os dados transformados, simplesmente extraindo-se o logaritmo da variável resposta (produtividade).

```
> bartlett.test(residuals(aov(log(produtividade)
+ ~ bloco, data=ex5))
+ ~ sistema , data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(log(produtividade) ~ bloco, data = ex5)) by sistema
Bartlett's K-squared = 0.072046, df = 1, p-value = 0.7884
```

```
> bartlett.test(residuals(aov(log(produtividade)
+ ~ bloco+ sistema, data=ex5))
+ ~ fonte, data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(log(produtividade) ~ bloco + sistema, data = ex5)) by fonte
Bartlett's K-squared = 3.2863, df = 3, p-value = 0.3496
```

```
> bartlett.test(residuals(aov(log(produtividade)
+ ~ bloco+ sistema+fonte, data=ex5))
+ ~ interaction(sistema,fonte), data=ex5)
```

Bartlett test of homogeneity of variances

```
data: residuals(aov(log(produtividade) ~ bloco + sistema + fonte, data = ex5)) b
Bartlett's K-squared = 5.4163, df = 7, p-value = 0.6093
```

Para o teste de Bartlett, como o nível descritivo do teste é maior que o nível de significância considerado ( $\alpha = 0.05$ ), em todos os casos testados, não se rejeita a hipótese de homogeneidade de variâncias.

```
> require(car)
> leveneTest(residuals(aov(log(produtividade)
+ ~ bloco, data=ex5)) ~ sistema , data=ex5)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group 1	1	0.5072	0.4818
	30		

```
> leveneTest(residuals(aov(log(produtividade)
+ ~ bloco+ sistema, data=ex5)) ~ fonte, data=ex5)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.2282 0.3179
      28
```

```
> leveneTest(residuals(aov(log(produktividade)
+               ~ bloco+ sistema+fonte, data=ex5))
+             ~ interaction(sistema,fonte), data=ex5)
```

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  7   0.874 0.5408
      24
```

Para o teste de Levene, também não se rejeita a hipótese de homogeneidade de variâncias, pois o nível descritivo do teste é maior que o nível de significância  $\alpha = 0.05$  em todos os casos.

## 6. Normalidade dos resíduos ( $\epsilon \sim N(0, 1)$ )

```
> shapiro.test(rstudent(aov(log(produktividade)
+                   ~ bloco+sistema*fonte+bloco*sistema,
+                   data=ex5)))
```

Shapiro-Wilk normality test

```
data:  rstudent(aov(log(produktividade) ~ bloco + sistema * fonte + bloco *
W = 0.976, p-value = 0.6778
```

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.6778$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

```
> ks.test(rstudent(aov(log(produktividade)
+                   ~ bloco+sistema*fonte+bloco*sistema,
+                   data=ex5)), "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data:  rstudent(aov(log(produktividade) ~ bloco + sistema * fonte + bloco *
D = 0.077752, p-value = 0.9821
alternative hypothesis: two-sided
```

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p - value = 0.9821$ ) é maior que  $\alpha = 0.05$ , da mesma forma que observado para o teste de Shapiro-Wilk, não existem evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, sendo essa pressuposição satisfeita.

A pressuposição de normalidade dos resíduos, observando o gráfico de quantis (qq-Plot) com envelope de simulação (função `qqPlot`), aproximou-se do desejado. Nesse caso, espera-se que os pontos estejam dentro do intervalo formado pelo envelope simulado.

```
> qqPlot(rstudent(aov(log(produtividade) ~ bloco+sistema*fonte+(bloco*sistema),
+ data=ex5)), pch=19, ylab="Resíduos")
```

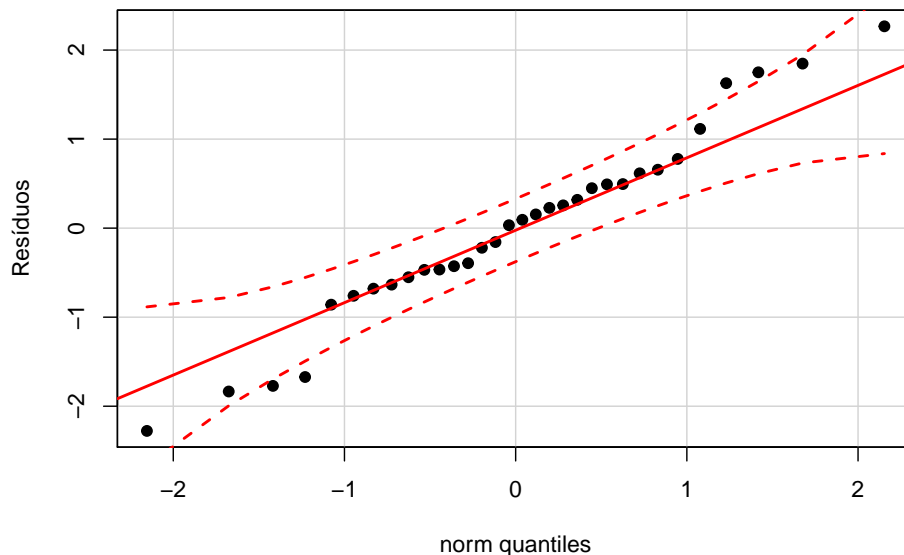


Figura 27: Gráfico *qqplot* para avaliar a distribuição normal dos resíduos

## 7. Bloco de efeito aditivo

A aditividade de blocos pode ser verificada pelo teste de não aditividade de Tukey. Antes disso, os dados de produtividade foram transformados. Para que a função realize os testes corretamente, incluiu-se uma coluna no arquivo original com a variável transformada:

```
> require(dae)
> ex5$lp <-log(ex5$produtividade)
```

```
> modad<-aov(lp ~ bloco+sistema*fonte + Error(bloco/sistema), ex5)
> tukey.1df(modad, ex5, error.term= "Within")
```

```
$Tukey.SS
```

```
[1] 0.008085765
```

```
$Tukey.F
```

```
[1] 1.043855
```

```
$Tukey.p
```

```
[1] 0.3212468
```

```
$Devn.SS
```

```
[1] 0.1316831
```

Como o p-valor ( $Tukey.p = 0.3212$ ) é maior que o nível de significância estabelecido ( $\alpha = 0.05$ ) não se rejeita a hipótese de que os blocos têm efeito aditivo no modelo. Isto significa que o efeito do tratamento não depende do bloco.

## 8. Transformação de Box-Cox

Equivalente ao apresentado anteriormente, pode-se utilizar a função `boxcox` para verificar a necessidade de transformação dos dados. Neste caso deve-se verificar se o valor 1 pertence ao intervalo de confiança para  $\lambda$ .



```

> require(MASS)
> boxcox(aov(log(produtividade) ~ bloco+sistema*fonte+bloco*sistema,
+           data=ex5), lambda = seq(-10, 10, l = 101))
> abline(v=1, col="red")

```

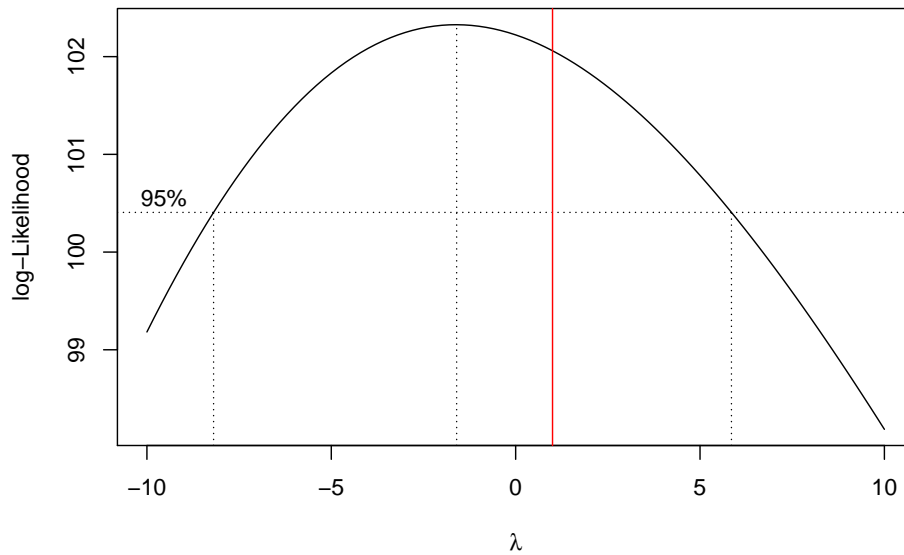


Figura 28: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox. Ao descrever o modelo, omite-se a expressão "Error". A linha sólida incica que o valor "1" está no intervalo

Observa-se que agora o intervalo com 95% contém o valor 1, o que indica que a transformação foi efetiva.

## 9. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(log(produtividade) ~ bloco+sistema*fonte+(bloco/sistema),
+             data=ex5)),rstudent(aov(log(produtividade)
+                                 ~ bloco+sistema*fonte+
+                                 (bloco/sistema),
+                                 data=ex5)),
+      xlab="Valores ajustados",
+      ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(log(produtividade) ~ bloco+sistema*fonte+
+                 (bloco/sistema), data=ex5)),
+      prob=T,main=" ", xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(log(produtividade) ~
+                         bloco+sistema*fonte+(bloco/sistema),
+                         data=ex5))), col="blue")
> ##qqPlot precisa do pacote car
> qqPlot(rstudent(aov(log(produtividade) ~
+                 bloco+sistema*fonte+(bloco/sistema),
+                 data=ex5)),
+        ylab="Resíduos Studentizados", pch=19)
> ##Boxplot
> boxplot(rstudent(aov(log(produtividade) ~
+                   bloco+sistema*fonte+(bloco/sistema),
+                   data=ex5)), ylab="Resíduos Studentizados",
+        pch=19, cex.lab=1.2,
+        cex.axis=1.2, col="turquoise")

```

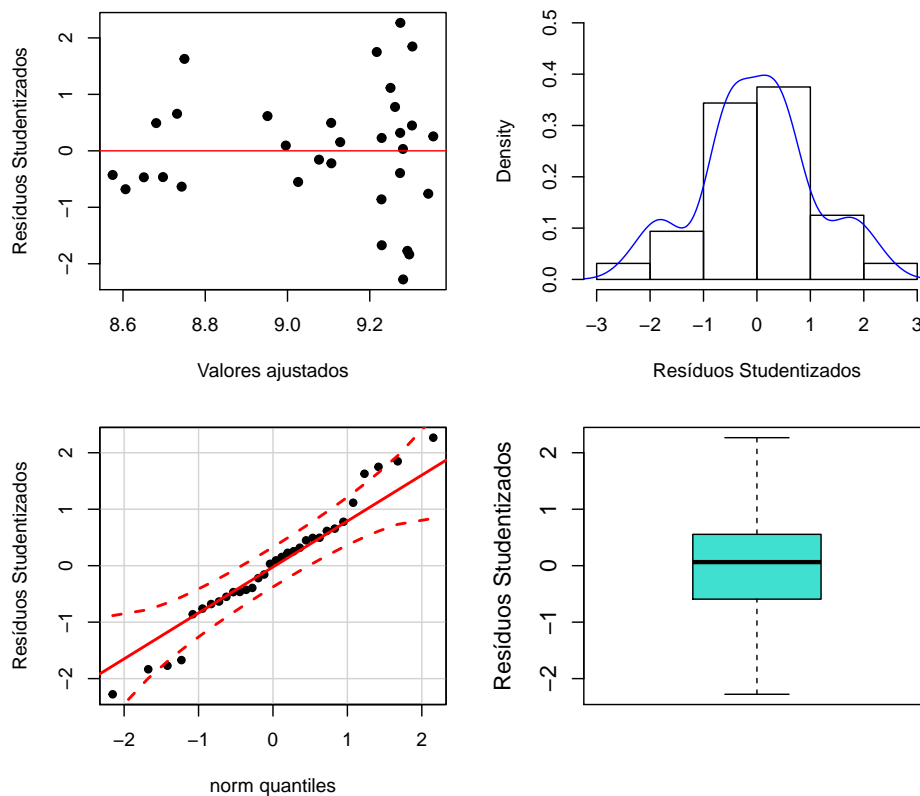


Figura 29: Gráficos de diagnósticos para verificar as pressuposições

Como todas as pressuposições da análise de variância foram satisfeitas, pode-se prosseguir com a análise.

### 3.7.4 Análise de variância

Pode-se utilizar a função `aov` especificando no modelo os efeitos de blocos, sistema e fonte.

```
> mod5 <- aov(log(produtividade) ~
+             bloco+sistema*fonte+Error(bloco/sistema), data=ex5)
> summary(mod5)
```

Error: bloco

	Df	Sum Sq	Mean Sq
bloco	3	0.09606	0.03202

Error: bloco:sistema

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sistema	1	0.008776	0.008776	1.692	0.284
Residuals	3	0.015556	0.005185		

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fonte	3	1.8084	0.6028	77.631	1.72e-10 ***
sistema:fonte	3	0.0918	0.0306	3.943	0.0252 *
Residuals	18	0.1398	0.0078		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

O pacote `agricolae` disponibiliza a função `sp.plot` que facilita a elaboração do modelo para a análise de variância, além de fornecer informações como os erros associados às parcelas e subparcelas, além dos coeficientes de variação e a média geral do experimento.

```
> with(ex5, sp.plot(bloco, sistema, fonte, log(produtividade)))
```

ANALYSIS SPLIT PLOT: log(produtividade)

Class level information

```
sistema      : convencional direto
fonte        : Eaves Esuinos Testemunha Nitrato
bloco        : 1 2 3 4
```

Number of observations: 32

Analysis of Variance Table

Response: log(produtividade)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bloco	3	0.09606	0.03202	6.1748	0.08455 .
sistema	1	0.00878	0.00878	1.6924	0.28419
Ea	3	0.01556	0.00519		
fonte	3	1.80840	0.60280	77.6311	1.716e-10 ***
sistema:fonte	3	0.09184	0.03061	3.9426	0.02524 *
Eb	18	0.13977	0.00776		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

cv(a) = 0.8 %, cv(b) = 1 %, Mean = 9.078562

O p-valor para efeito de blocos não foi significativo ( $p - value = 0.08455$ ), não há evidências de efeito de blocos na variável testada. Como o efeito da interação entre sistema e fonte foi significativo ( $p - value = 0.0252$ ), não se deve considerar os efeitos individuais, pois isto indica que o efeito de um fator depende dos níveis do outro fator. Neste caso, segue-se com o desdobramento da interação.

```
> ex5$sf=with(ex5,interaction(fonte,sistema))
> modsub_desd=aov(log(produtividade) ~ bloco+sistema+sf
+ Error(bloco/sistema), data=ex5)
> summary(modsub_desd)
```

Error: bloco

	Df	Sum Sq	Mean Sq
bloco	3	0.09606	0.03202

Error: bloco:sistema

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sistema	1	0.008776	0.008776	1.692	0.284
Residuals	3	0.015556	0.005185		

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sf	6	1.9002	0.3167	40.79	1.61e-09 ***
Residuals	18	0.1398	0.0078		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

```
> coef(modsub_desd)
```

(Intercept) :

(Intercept)  
9.078562

bloco :

	bloco2	bloco3	bloco4
	0.09840178	0.10611180	0.14987236

```
bloco:sistema :
sistemadireto
-0.03312026
```

```
Within :
```

```
sfEsuinos.convencional    sfNitrato.convencional  sfTestemunha.convencional
      0.0004742059                0.0219769507                -0.5312944907
sfEaves.direto            sfEsuinos.direto        sfNitrato.direto
      0.3451487565                0.6229370625                0.6107985417
```

```
> summary(modsub_desd, split=list(sf=list(fonte_d_Convencional=c(1,2,3),
+                                       fonte_d_Direto=4:6)) )
```

```
Error: bloco
```

```
      Df Sum Sq Mean Sq
bloco  3 0.09606 0.03202
```

```
Error: bloco:sistema
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
sistema  1 0.008776 0.008776  1.692  0.284
Residuals  3 0.015556 0.005185
```

```
Error: Within
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
sf      6 1.9002  0.3167  40.79 1.61e-09 ***
sf: fonte_d_Convencional  3 0.8721  0.2907  37.44 6.04e-08 ***
sf: fonte_d_Direto        3 1.0281  0.3427  44.14 1.68e-08 ***
Residuals                18 0.1398  0.0078
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observa-se que há diferença significativa entre pelo menos duas fontes testadas em cada um dos sistemas avaliados.

### 3.7.5 Análise de variância com o pacote ExpDes

```
> require(ExpDes.pt)
> with(ex5, psub2.dbc(sistema, fonte, bloco, log(produtividade),
+                   quali = c(TRUE, TRUE), mcomp = "tukey",
+                   fac.names = c("Sistema", "Fonte"),
+                   sigT = 0.05, sigF = 0.05))
```

---

```
Legenda:
```

```
FATOR 1 (parcela): Sistema
FATOR 2 (subparcela): Fonte
```

-----  
 \$`Quadro da analise de variancia\n-----

	GL	SQ	QM	Fc	Pr(>Fc)
Sistema	1	0.00878	0.00878	1.692	0.28419
Bloco	3	0.09606	0.03202	6.175	0.08454 .
Erro a	3	0.01556	0.00518		
Fonte	3	1.80840	0.60280	77.631	< 2e-16 ***
Sistema*Fonte	3	0.09184	0.03061	3.943	0.02524 *
Erro b	18	0.13977	0.00776		
Total	31	2.16040			

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

-----  
 CV 1 = 0.7931851 %  
 CV 2 = 0.9706258 %

Interacao significativa: desdobrando a interacao

-----  
 Desdobrando Sistema dentro de cada nivel de Fonte

	GL	SQ	QM	Fc	valor.p
Sistema : Fonte Eaves	1.00000	0.0881190	0.088119	12.37625	0.002073
Sistema : Fonte Esuinos	1.00000	0.0090880	0.009088	1.27646	0.271456
Sistema : Fonte Nitrate	1.00000	0.0022810	0.002281	0.320329	0.577483
Sistema : Fonte Testemunha	1.00000	0.0011290	0.001129	0.158547	0.694567
Erro combinado	20.73973	0.1476669	0.007120		

-----  
 Sistema dentro de Fonte Eaves

-----  
 Teste de Tukey

-----  
 Grupos Tratamentos Medias

a	convencional	9.222333
b	direto	9.01243

-----  
 Sistema dentro de Fonte Esuinos

-----  
 De acordo com o teste F, as medias desse fator sao estatisticamente iguais.  
 -----

	Niveis	Medias
1	convencional	9.222807
2	direto	9.290218

---

Sistema dentro de Fonte Nitrato

---

De acordo com o teste F, as medias desse fator sao estatisticamente iguais.

---

	Niveis	Medias
1	convencional	9.244310
2	direto	9.278079

---

Sistema dentro de Fonte Testemunha

---

De acordo com o teste F, as medias desse fator sao estatisticamente iguais.

---

	Niveis	Medias
1	convencional	8.691039
2	direto	8.667281

---

Desdobrando Fonte dentro de cada nivel de Sistema

---

	GL	SQ	QM	Fc	valor.p
Fonte : Sistema convencional	3	0.872107	0.290702	37.437504	0
Fonte : Sistema direto	3	1.028137	0.342712	44.135506	0
Erro b	18	0.139769	0.007765		

---

Fonte dentro de Sistema convencional

---

Teste de Tukey

---

Grupos	Tratamentos	Medias
a	Nitrato	9.24431
a	Esuinos	9.222807
a	Eaves	9.222333
b	Testemunha	8.691039

---

Fonte dentro de Sistema direto

---

## Teste de Tukey

---

### Grupos Tratamentos Medias

a	Esuinos	9.290218
a	Nitrato	9.278079
b	Eaves	9.01243
c	Testemunha	8.667281

---

---

Observe que as conclusões para a análise de variância com o procedimento do **Exp-Des.pt** são as mesmas que as observadas nos procedimentos anteriores. Entretanto, este pacote não verifica as pressuposições da análise de variância.

Na tabela a seguir são apresentadas as médias dos dados originais seguidas das letras do teste de comparação de médias.

Tabela 7: Médias de produtividade de milho (kg/ha) em função do sistema de plantio e da fonte de nitrogênio. Médias seguidas por letras maiúsculas iguais na coluna e minúsculas iguais na linha não diferem entre si pelo teste de Tukey

	Eaves	Esuinos	Nitrato	Testemunha
convencional	10144,00 Aa	10201,25 Aa	10383,25 Aa	5985,25 Ab
direto	8208,50 Ba	10902,00 Ba	10718,25 Bb	5829,75 Ac

Com base na análise, para sistema plantio convencional não há diferença significativa entre as fontes de nitrogênio, mas todas diferem do tratamento testemunha sem adubação. No sistema plantio direto, esterco de suíno e nitrato não diferem significativamente, mas ambas diferem de esterco de aves. Assim como no plantio convencional, a testemunha sem adubação confere os piores resultados.



## 3.8 Análise de Regressão

O experimento a seguir avaliou o efeito de doses crescentes de um composto orgânico no rendimento de grãos de milho em Sistema Plantio Direto. Utilizou-se o delineamento inteiramente casualizado com quatro repetições

Tabela 8: Produtividade de milho (kg/ha) em função da dose de composto orgânico

Repetição	Dose de Composto (t/ha)				
	0	2	4	8	16
	Produtividade (t/ha)				
1	1,910	5,596	7,134	6,918	6,006
2	2,256	7,187	7,229	6,883	6,349
3	2,939	7,535	6,455	7,841	7,575
4	3,369	5,725	5,995	7,845	7,861

### 3.8.1 Importando os dados

```
> #ex6 <- read.xlsx("dados.xlsx", sheet="ex6")
> ex6 <- read.csv("ex6.csv", dec=",", header = T)
> head(ex6)
```

```
  dose prod
1    0 1.910
2    0 2.256
3    0 2.939
4    0 3.369
5    2 5.596
6    2 7.187
```

### 3.8.2 Análise exploratória dos dados

A análise exploratória dos dados pode ser realizada como os exemplos anteriores.

```
> str(ex6)

'data.frame':      20 obs. of  2 variables:
 $ dose: int  0 0 0 0 2 2 2 2 4 4 ...
 $ prod: num  1.91 2.26 2.94 3.37 5.6 ...
```

```
> summary(ex6)
```

dose	prod
Min. : 0	Min. :1.910
1st Qu.: 2	1st Qu.:5.693
Median : 4	Median :6.669
Mean : 6	Mean :6.030
3rd Qu.: 8	3rd Qu.:7.306
Max. :16	Max. :7.861

```
> mar=c(4,3,1,1)
```

```
> with(ex6, boxplot(prod ~ dose, xlab="Dose de esterco (t/ha)",  
+ ylab="Rendimento de milho (t/ha)", ylim=c(0,9),  
+ las=1, cex.axis=0.6, horizontal = F))
```

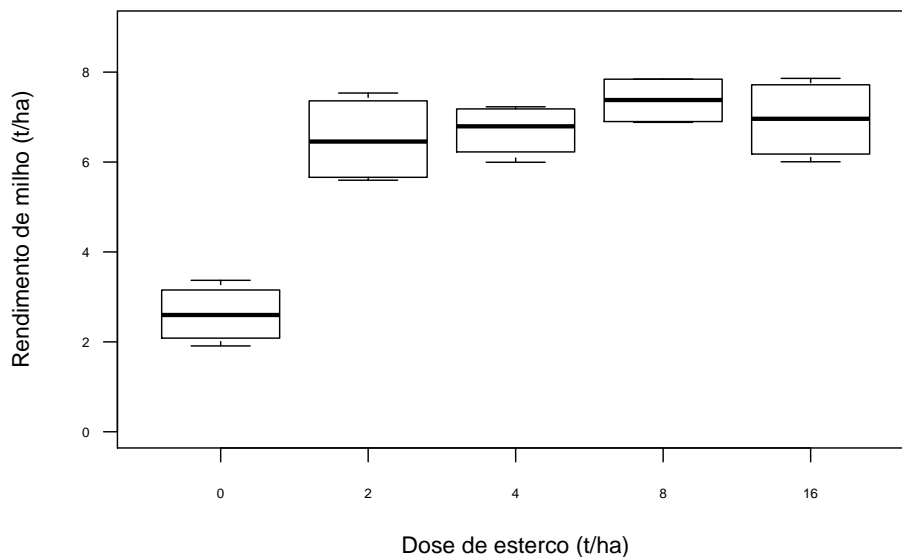


Figura 30: Gráficos de caixa (*boxplots*) dos dados de rendimento de milho (t/ha) em função da dose de esterco aplicada

Um gráfico de dispersão é uma ferramenta útil para observar o comportamento dos dados em função da dose de composto aplicada.

```
> plot(prod ~ dose, ylim=c(0, 9), pch=19, xlab="Dose de composto (t/ha)",
+       ylab="Produtividade (t/ha)", data=ex6)
```

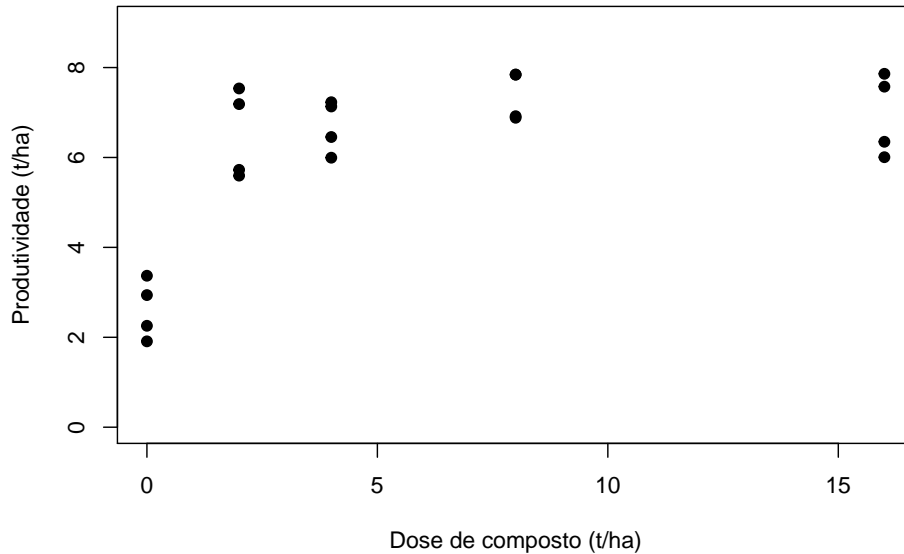


Figura 31: Produtividade de grãos em função da dose de composto

### 3.8.3 Pressuposições do modelo estatístico

Para que a análise de variância seja válida, no caso de um experimento em DIC, deve-se verificar se as pressuposições de homogeneidade de variâncias entre tratamentos e normalidade dos resíduos são satisfeitas.

#### 1. Homogeneidade de Variâncias

A hipótese a ser testada em um teste de homocedasticidade é dada por:

$$\begin{cases} H_0 : \sigma_{Dose0}^2 = \sigma_{Dose2}^2 = \dots = \sigma_{Dose16}^2 \\ H_1 : \sigma_{Dose0}^2 \neq \sigma_{Dose2}^2 \neq \dots \neq \sigma_{Dose16}^2 \end{cases} \quad (14)$$

Sempre que o nível descritivo do teste (*p-value*) for menor que o nível de significância estabelecido ( $\alpha$ ), rejeita-se  $H_0$ . Caso contrário, não se rejeita indicando que não existem evidências contra a homocedasticidade.

```
> bartlett.test(prod ~ as.factor(dose), data=ex6)
```

```
Bartlett test of homogeneity of variances
```

```
data: prod by as.factor(dose)
```

```
Bartlett's K-squared = 1.5328, df = 4, p-value = 0.8208
```

No teste de Bartlett, como o p-valor ( $p\text{-value} = 0.8208$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que as variâncias não diferem, satisfazendo esta pressuposição.

```
> require(car)
> leveneTest(prod ~ as.factor(dose), data=ex6)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group 4  3.9049 0.02287 *
      15
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

No teste de Levene, como o p-valor ( $p\text{-value} = 0.0228$ ) é menor que  $\alpha = 0.05$ , rejeita-se a hipótese de que as variâncias são homogêneas.

## 2. Normalidade dos resíduos ( $\epsilon \sim N(0, \sigma)$ )

A hipótese a ser testada em um teste de normalidade dos resíduos é dada por:

$$\begin{cases} H_0: \text{Os resíduos do modelo provêm de uma população com distribuição normal;} \\ H_1: \text{Os resíduos do modelo não provêm de uma população com distribuição normal.} \end{cases} \quad (15)$$

```
> shapiro.test(rstudent(aov(prod ~ as.factor(dose), data=ex6)))
```

Shapiro-Wilk normality test

```
data:  rstudent(aov(prod ~ as.factor(dose), data = ex6))
W = 0.90711, p-value = 0.05617
```

No teste de Shapiro-Wilk, como o p-valor ( $p\text{-value} = 0.0561$ ) é maior que  $\alpha = 0.05$ , não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, satisfazendo esta pressuposição.

```
> ks.test(rstudent(aov(prod ~ as.factor(dose), data=ex6)), "pnorm")
```

One-sample Kolmogorov-Smirnov test

```
data:  rstudent(aov(prod ~ as.factor(dose), data = ex6))
D = 0.18961, p-value = 0.4166
alternative hypothesis: two-sided
```

Para o teste de Kolmogorov-Smirnov, como o p-valor ( $p - value = 0.4166$ ) é maior que  $\alpha = 0.05$ , também não há evidências para rejeitar a hipótese de que os resíduos do modelo seguem a distribuição normal, satisfazendo esta pressuposição.

```
> qqPlot(rstudent(aov(prod~as.factor(dose), data=ex6)),  
+         pch=19, ylab="Resíduos")
```

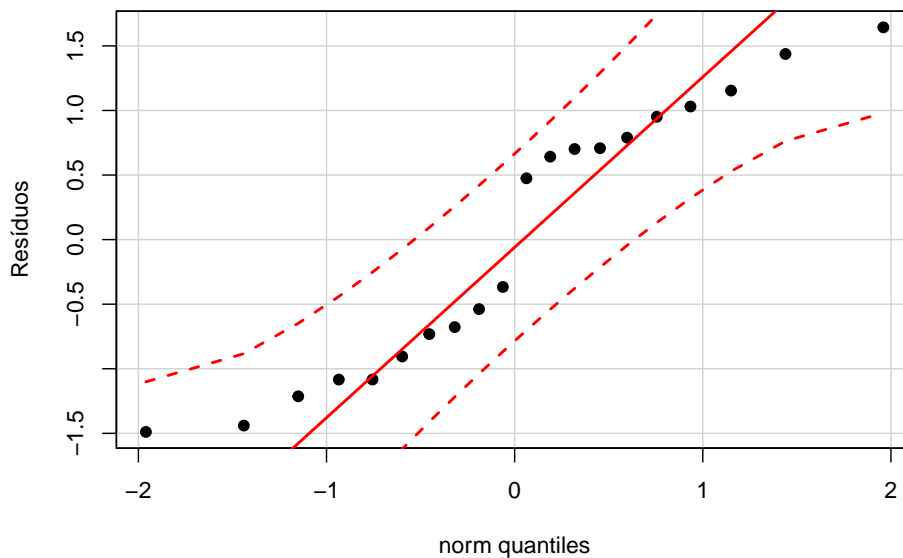


Figura 32: Gráfico *qqplot* para avaliar a distribuição normal dos resíduos

Uma outra forma de checar a adequação de distribuição é o gráfico *qqPlot* com envelope de simulação. Neste caso espera-se que todos os pontos estejam dentro dos limites definidos pelas linhas do envelope.

### 3. Transformação de Box-Cox

A função **boxcox** do pacote **MASS** (Venables & Ripley, 2002) auxilia na verificação das pressuposições do modelo:

```

> require(MASS)
> boxcox(aov(prod~as.factor(dose),
+         data=ex6), lambda = seq(-2, 4, l = 101))

```

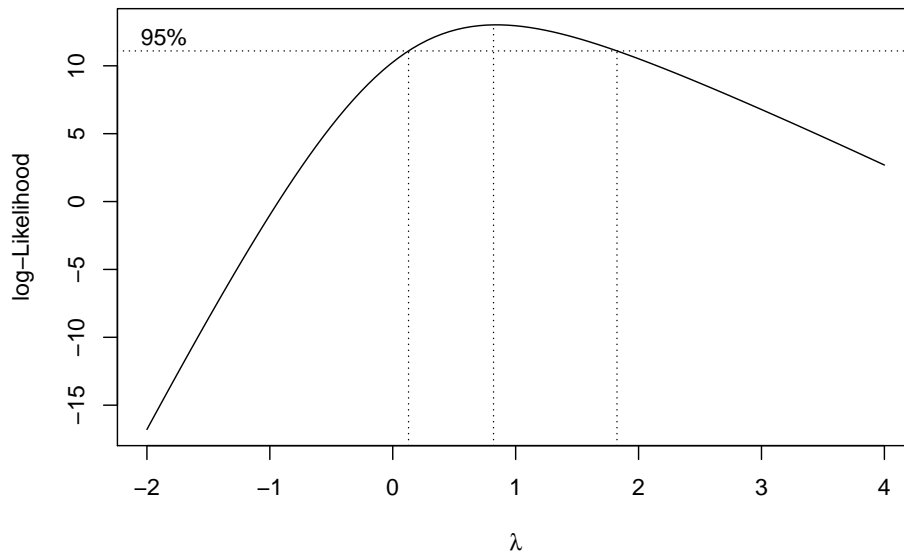


Figura 33: Gráfico da verossimilhança perfilhada para  $\lambda$  da transformação de Box-Cox

Como o intervalo de confiança para o parâmetro contém o valor 1, não há evidências de que os dados necessitam de transformação. Verificadas e atendidas as pressuposições do modelo, pode-se prosseguir a análise de variância.

#### 4. Gráficos de diagnóstico

```

> par(mfrow=c(2,2), mar=c(4,4,2,2))
> plot(fitted(aov(prod ~ as.factor(dose), data=ex6)),
+      rstudent(aov(prod ~ as.factor(dose), data=ex6)),
+      xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> hist(rstudent(aov(prod ~ as.factor(dose), data=ex6)),prob=T,
+      main=" ", xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(aov(prod ~ as.factor(dose), data=ex6))),
+      col="blue")
> qqPlot(rstudent(aov(prod ~ as.factor(dose), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19)
> boxplot(rstudent(aov(prod ~ as.factor(dose), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2,
+      cex.axis=1.2, col="turquoise")

```

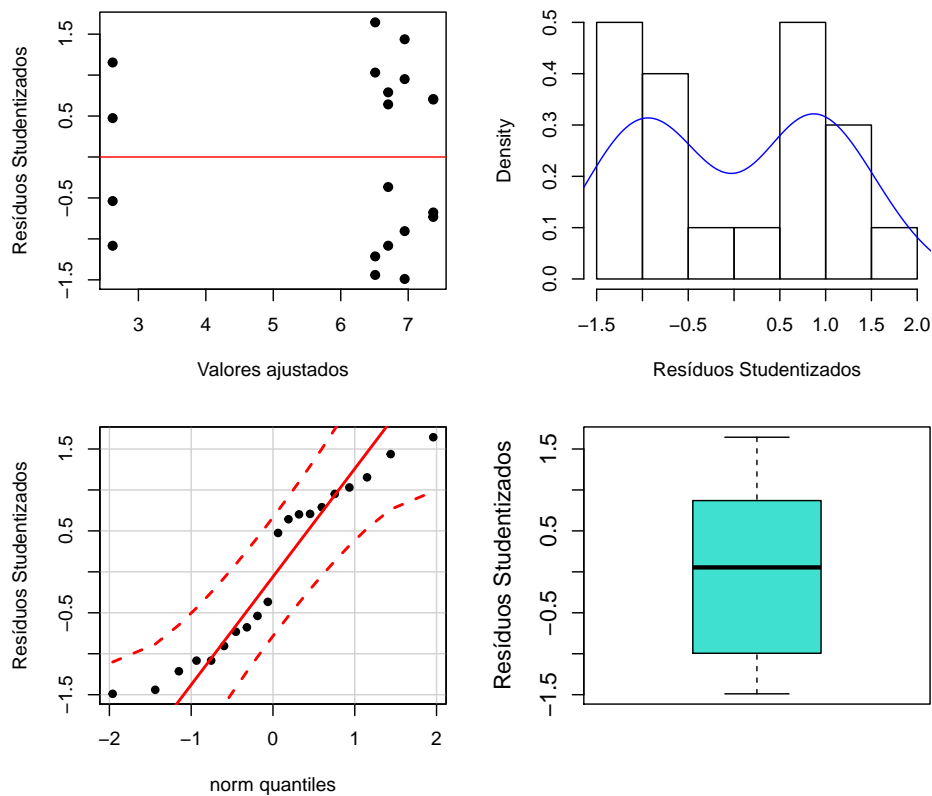


Figura 34: Gráficos de diagnósticos para verificar as pressuposições

### 3.8.4 Análise de variância

A análise de variância deve ser realizada considerando-se dose como um fator qualitativo.

```

> mod6<- aov(prod ~ as.factor(dose), data=ex6)
> summary(mod6)

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(dose)	4	59.86	14.965	25.96	1.37e-06 ***
Residuals	15	8.65	0.576		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Neste caso, há efeito significativo de dose  $p - valor < 0,05$ .

Como dose é um fator quantitativo, o correto é ajustar um modelo de regressão e basear as inferências na interpretação dos parâmetros. A ideia é propor uma série de modelos e selecionar aquele que apresentar o melhor ajuste aos dados, além de proporcionar parâmetros interpretáveis.

### 3.8.5 Modelos polinomiais

#### 1. Polinômio de grau 1

A equação da reta é dada por:

$$y = b_1 + b_2x$$

$b_1$ : intercepto;

$b_2$ : coeficiente angular;

A função `lm` é utilizada para o ajuste de modelos lineares. Para especificar um modelo polinomial de grau 1, basta especificar a variável resposta em função do fator quantitativo que será utilizado para descrever seu comportamento.

```
> mod6.1 <- lm(prod ~ dose, data=ex6)
> summary(mod6.1)
```

Call:

```
lm(formula = prod ~ dose, data = ex6)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.0306	-1.5162	0.3745	1.4483	2.2311

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.94063	0.52893	9.341	2.52e-08 ***
dose	0.18163	0.06414	2.832	0.0111 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



Residual standard error: 1.623 on 18 degrees of freedom  
Multiple R-squared: 0.3082, Adjusted R-squared: 0.2697  
F-statistic: 8.018 on 1 and 18 DF, p-value: 0.01106

Avaliando a qualidade do ajuste:

```

> par(mfrow=c(2,3))
> plot(fitted(lm(prod ~ dose, data=ex6)),rstudent(lm(prod ~ dose,
+
+       xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> plot(fitted(lm(prod ~ dose, data=ex6)),ex6$prod,xlab="Valores ajustados",
+
+       ylab="Valores observados",pch=19)
> abline(a=0,b=1,col="red")
> plot(fitted(lm(prod ~ dose, data=ex6)),sqrt(rstudent(lm(prod ~
+
+       dose, data=ex6))^2),
+
+       xlab="Valores ajustados",
+       ylab="Raíz quadrada dos resíduos Studentizados",pch=19)
> lines(with(ex6,lowess(fitted(lm(prod ~ dose, data=ex6)),
+
+       sqrt(rstudent(lm(prod ~ dose, data=ex6))^2))),
+
+       col="red")
> hist(rstudent(lm(prod ~ dose, data=ex6)),prob=T,main=" ",
+
+       xlab="Resíduos Studentizados", ylim=c(0,0.5))
> lines(density(rstudent(lm(prod ~ dose, data=ex6))), col="blue")
> qqPlot(rstudent(lm(prod ~ dose, data=ex6)),
+
+       ylab="Resíduos Studentizados",
+       pch=19)
> boxplot(rstudent(lm(prod ~ dose, data=ex6)),
+
+       ylab="Resíduos Studentizados",
+       pch=19, cex.lab=1.2, cex.axis=1.2, col="turquoise")

```

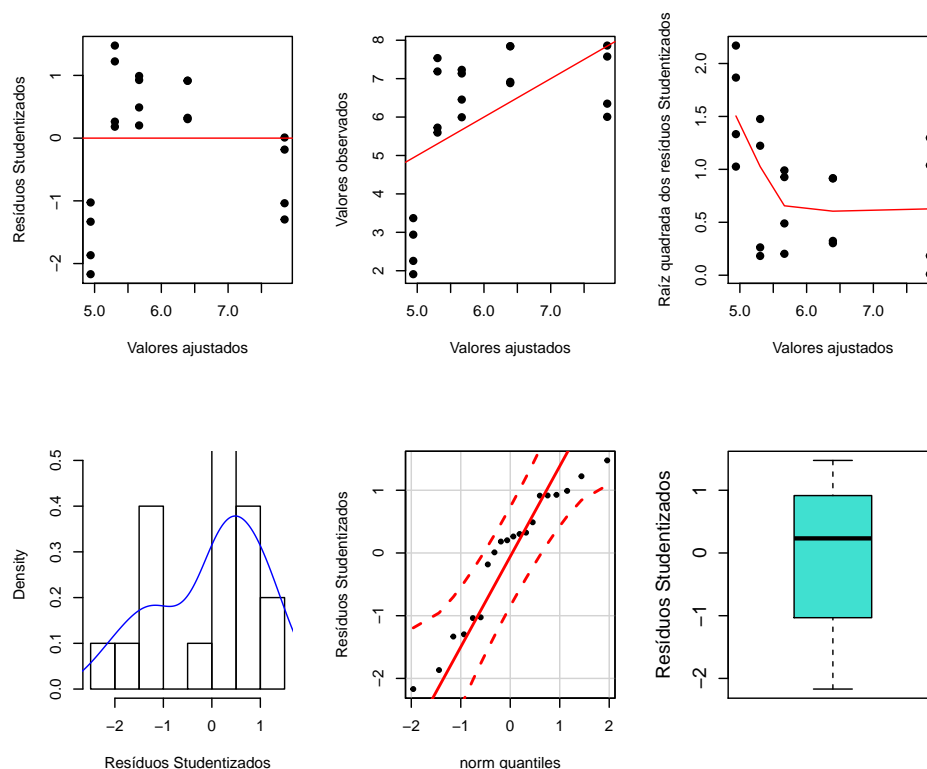


Figura 35: Gráficos de diagnósticos para verificar as pressuposições

Para construção do gráfico com a linha ajustada é necessário escrever o modelo que será utilizado para plotar a curva, o que pode ser feito utilizando o comando `function`.

```
> funcL <- function (x, beta){
+   return (beta[1]+beta[2]*(x))
+ }
```

O gráfico com a linha ajustada pode ser obtido da seguinte forma:

```
> plot(prod ~ dose, ylim=c(0, 10), xlab="Dose de composto (t/ha)",
+   ylab="Produtividade (t/ha)", data=ex6)
> curve(funcL (x, beta=coef(mod6.1)), from=0, to=16, type = "l",
+   col = "red", add = T, lty = 1, lwd=3)
```

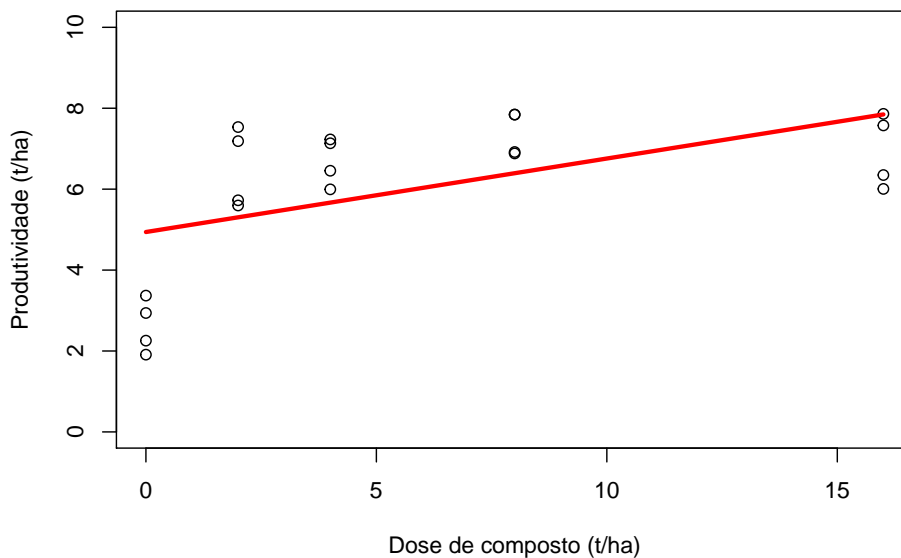


Figura 36: Modelo polinomial de 1º grau ajustado aos dados de produtividade em função da dose do composto

## 2. Polinômio de grau 2

Um modelo polinomial de grau 2 é dado pela seguinte expressão:

$$y = b_1 + b_2x + b_3x^2$$

$b_1$ : intercepto;

$b_2$ : é a taxa de variação na origem;

$b_3$ : mede o grau de curvatura e orientação da concavidade da parábola;

Para especificar um modelo polinomial de grau 2, deve-se especificar a variável resposta em função do fator quantitativo que será utilizado para descrever seu comportamento, tanto na forma linear como na forma quadrática. A função `I()` permite que sejam realizadas operações aritméticas dentro do comando `lm`.

```
> mod6.2 <- lm(prod ~ dose + I(dose^2), data=ex6)
> summary(mod6.2)
```

Call:

```
lm(formula = prod ~ dose + I(dose^2), data = ex6)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.6256	-0.6470	-0.1961	0.6295	2.3218

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.535600	0.471467	7.499	8.70e-07	***
dose	0.929467	0.166975	5.567	3.41e-05	***
I(dose^2)	-0.045324	0.009765	-4.642	0.000233	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.109 on 17 degrees of freedom

Multiple R-squared: 0.6949, Adjusted R-squared: 0.659

F-statistic: 19.36 on 2 and 17 DF, p-value: 4.151e-05

Avaliando a qualidade do ajuste:

```

> par(mfrow=c(2,3))
> plot(fitted(lm(prod ~ dose + I(dose^2), data=ex6)),
+      rstudent(lm(prod ~ dose + I(dose^2), data=ex6)),
+      xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> plot(fitted(lm(prod ~ dose + I(dose^2), data=ex6)),ex6$prod,
+      xlab="Valores ajustados", ylab="Valores observados",pch=19)
> abline(a=0,b=1,col="red")
> plot(fitted(lm(prod ~ dose + I(dose^2), data=ex6)),
+      sqrt(rstudent(lm(prod ~ dose + I(dose^2), data=ex6))^2),
+      xlab="Valores ajustados",
+      ylab="Raíz quadrada dos resíduos Studentizados",pch=19)
> lines(with(ex6,lowess(fitted(lm(prod ~ dose + I(dose^2), data=ex6)),
+                        sqrt(rstudent(lm(prod ~ dose + I(dose^2),
+                        data=ex6))^2))), col="red")
> hist(rstudent(lm(prod ~ dose + I(dose^2), data=ex6)),prob=T,main=" ",
+      xlab="Resíduos Studentizados", ylim=c(0,0.7))
> lines(density(rstudent(lm(prod ~ dose + I(dose^2), data=ex6))), col="blue")
> qqPlot(rstudent(lm(prod ~ dose + I(dose^2), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19)
> boxplot(rstudent(lm(prod ~ dose + I(dose^2), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2, cex.axis=1.2,
+      col="turquoise")

```

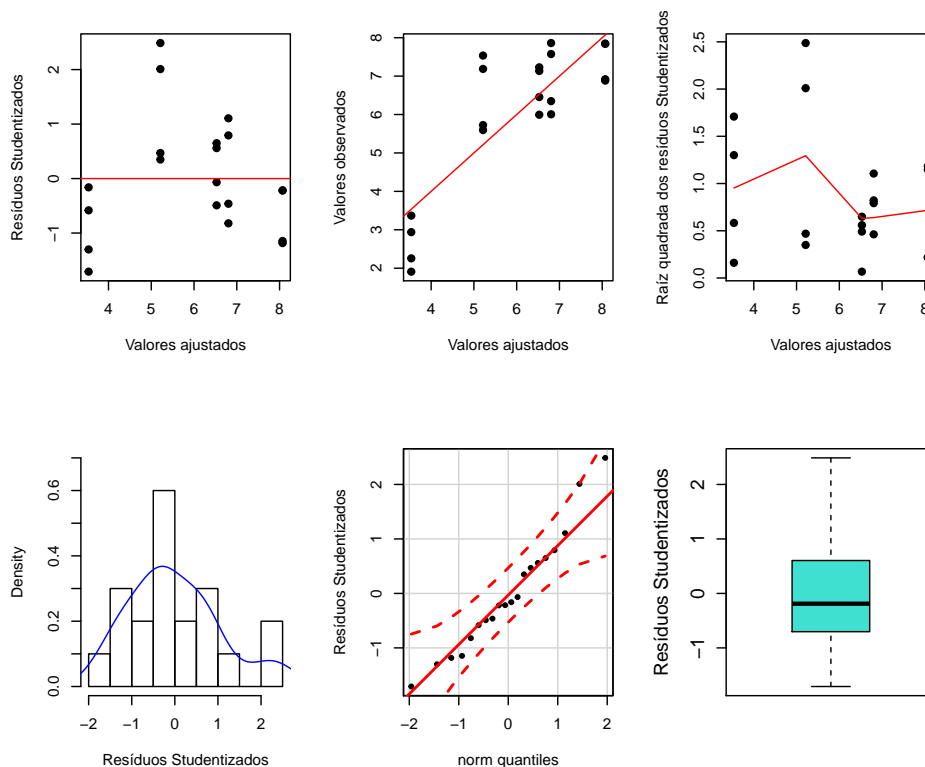


Figura 37: Gráficos de diagnósticos para verificar as pressuposições

Para construção do gráfico de dispersão com a linha ajustada é necessário escrever o modelo que será utilizado para plotar a curva:

```
> funcQ <- function (x, beta){
+   return (beta[1]+beta[2]*(x)+ beta[3]*(x)^2)
+ }
```

O gráfico com a curva ajustada pode ser obtido da seguinte forma:

```
> plot(prod ~ dose, ylim=c(0, 10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6)
> curve(funcQ (x, beta=coef(mod6.2)), from=0, to=16, type = "l",
+      col = "red", add = T, lty = 1, lwd=3)
```

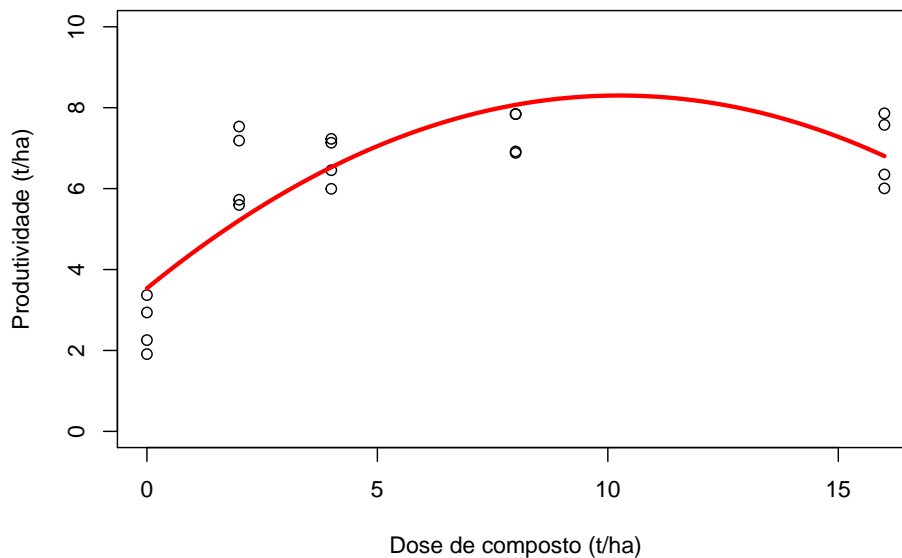


Figura 38: Modelo polinomial de 2º grau ajustado aos dados de produtividade em função da dose de composto

Utilizando o modelo polinomial de grau 2, pode-se calcular a máxima eficiência técnica, ou seja, a dose de adubo que leva à maior produtividade, utilizando o vértice da parábola:

```
> coef1 <- coef(mod6.2)
> (Xv <- -(coef1[2])/(2*coef1[3]))

      dose
10.25368
```

```

> Delta <- ((coef1[2])^2)-4*coef1[3]*coef1[1]
> (Yv <- -(Delta)/(4*coef1[3]) )

      dose
8.300831

```

O ponto de máximo pode ser ilustrado de forma gráfica:

```

> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+       ylab="Produtividade (t/ha)", data=ex6)
> curve(funcQ(x, beta=coef(mod6.2)), from=0, to=16, type = "l",
+       col = "red", add = T, lty = 1, lwd=3)
> abline(v=Xv, lty=2, lwd=2)
> abline(h=Yv, lty=2, lwd=2)

```

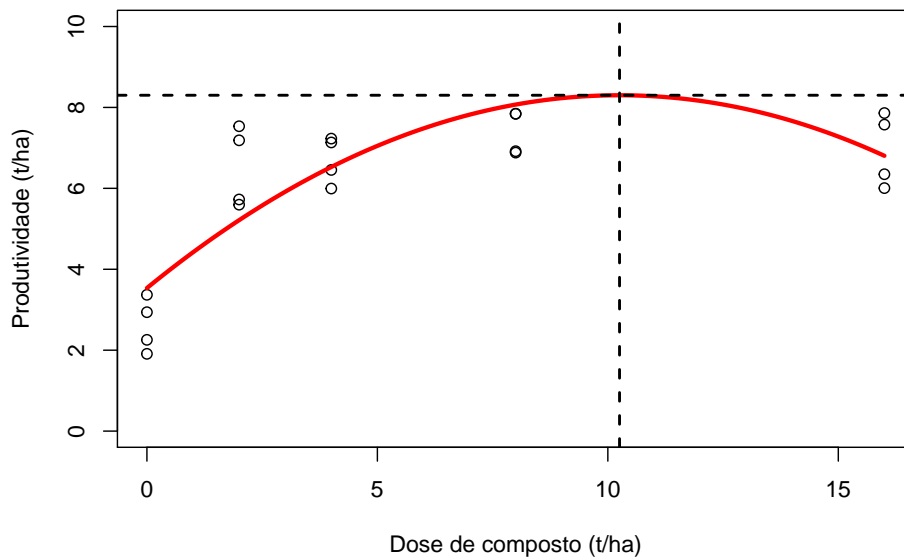


Figura 39: Modelo quadrático ajustado aos dados de produtividade em função da dose de composto, destacando-se a dose  $X_v$  que leva à máxima produtividade  $Y_v$

### 3.8.6 Reparametrização do modelo polinomial de grau 2

Com o objetivo de se calcular a máxima eficiência técnica, pode-se ajustar uma equação de regressão quadrática, parametrizada para  $X_{mx}$   $Y_{mx}$  por Zeviani (2013):

$$y = y_m + c(x + x_m)^2$$

$y_m$ : valor máximo;

$c$ : parâmetro sem interpretação que controla a forma da curva;

$x_m$ : valor de  $x$  para o máximo da função;

Como o modelo agora é não linear, utiliza-se a função `nls` e fazem-se as estimativas dos valores iniciais.

```
> mod6.2R <- nls(prod ~ ym+c*(dose-xm)^2,
+               start=list(ym=6.3, c=-0.024, xm=10.76), data=ex6)
> coef6.2R <- coef(mod6.2R)
> summary(mod6.2R)
```

Formula: `prod ~ ym + c * (dose - xm)^2`

Parameters:

	Estimate	Std. Error	t value	Pr(> t )	
ym	8.300831	0.447642	18.543	1.03e-12	***
c	-0.045324	0.009765	-4.642	0.000233	***
xm	10.253682	0.648209	15.818	1.33e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.109 on 17 degrees of freedom

Number of iterations to convergence: 3

Achieved convergence tolerance: 1.395e-08

```
> confint.default(mod6.2R)
```

	2.5 %	97.5 %
ym	7.42346917	9.17819343
c	-0.06446216	-0.02618501
xm	8.98321514	11.52414902

Com este modelo, além de estimar diretamente os parâmetros, pode-se estimar os limites do intervalo de confiança para os mesmos:

```
> confint.default(mod6.2R)
```



	2.5 %	97.5 %
ym	7.42346917	9.17819343
c	-0.06446216	-0.02618501
xm	8.98321514	11.52414902

Neste caso, a dose de composto que maximiza a produtividade fica entre 8,98 t/ha e 11,52 t/ha para produtividade entre 7,42 t/ha e 9,18t/ha de milho.

Vale ressaltar que o diagnóstico do modelo não necessita ser feito, pois a reparametrização não altera o ajuste.

A função a seguir pode ser utilizada para o cálculo do  $R^2$  para modelos não lineares.

```
> R2 <- function(nls.obj){
+ da <- eval(nls.obj$data)
+ resp.name <- all.vars(summary(nls.obj)$formula)[1]
+ names(da)[which(names(da)==resp.name)] <- "y"
+ sqn <- deviance(nls.obj)
+ sqe <- deviance(lm(y~1, da))
+ 1-(sqn/sqe)
+ }
```

Para obtenção do  $R^2$  para o modelo ajustado, basta aplicar esta função no modelo de interesse:

```
> R2(mod6.2R)
[1] 0.6948696
```

Na nova parametrização a função para construção do gráfico será escrita da seguinte forma:

```
> funcQW <- function (x, beta){
+ return (beta[1]+beta[2]*(x-beta[3])^2)
+ }
```

A seguir é apresentado o gráfico da produtividade em função da dose com a linha do modelo ajustado. Observa-se também que é possível apresentar o intervalo de confiança para os parâmetros.

```

> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+       ylab="Produtividade (t/ha)", data=ex6)
> curve(funcQW(x, beta=coef(mod6.2R)), from=0, to=16, type="l",
+       col="red", add=T, lty=1, lwd=3)
> abline(h=coef6.2R[1], lty=2, lwd=2)
> abline(v=coef6.2R[3], lty=2, lwd=2)
> abline(v=confint.default(mod6.2R)[3,], col="blue", lty=2)

```

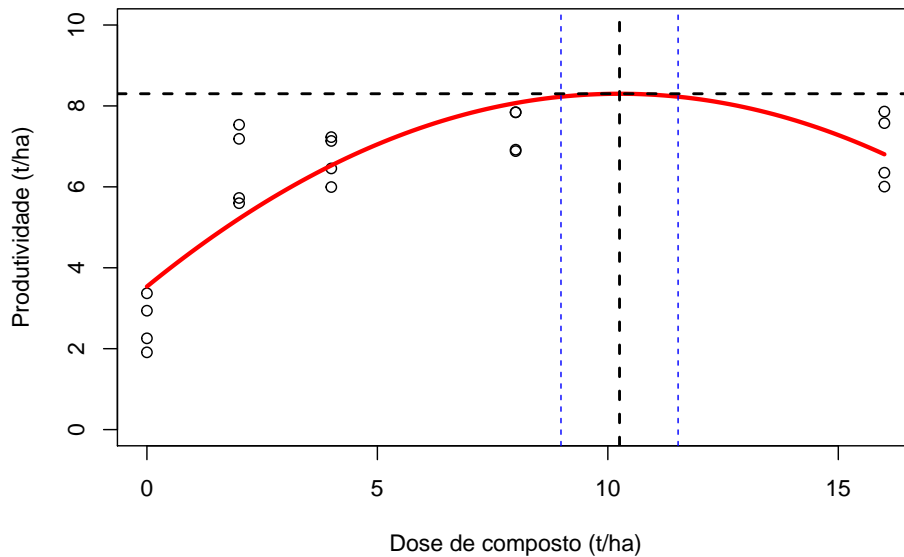


Figura 40: Modelo quadrático parametrizado por Zeviani (2013) ajustado aos dados de produtividade em função da dose de composto

### 3. Polinômio de grau 3

Um polinômio de terceiro grau é dado pela seguinte expressão:

$$y = b_1 + b_2x + b_3x^2 + b_4x^3$$

O ajuste é realizado da mesma forma apresentada para o polinômio de grau 2.

```

> mod6.3 <- lm(prod ~ dose + I(dose^2) + I(dose^3),
+             data=ex6)
> summary(mod6.3)

```

Call:

```
lm(formula = prod ~ dose + I(dose^2) + I(dose^3), data = ex6)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.2958	-0.5894	-0.1355	0.6163	1.6958

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.838838	0.422673	6.716	4.95e-06	***
dose	1.957302	0.328840	5.952	2.03e-05	***
I(dose^2)	-0.246022	0.059391	-4.142	0.000765	***
I(dose^3)	0.008736	0.002564	3.408	0.003601	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8701 on 16 degrees of freedom

Multiple R-squared: 0.8232, Adjusted R-squared: 0.79

F-statistic: 24.83 on 3 and 16 DF, p-value: 2.926e-06

Avaliando a qualidade do ajuste:

```

> par(mfrow=c(2,3))
> plot(fitted(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      xlab="Valores ajustados", ylab="Resíduos Studentizados",pch=19)
> abline(h=0,col="red")
> plot(fitted(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      ex6$prod,xlab="Valores ajustados", ylab="Valores observados",pch=19)
> abline(a=0,b=1,col="red")
> plot(fitted(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      sqrt(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6))^2),
+      xlab="Valores ajustados",
+      ylab="Raíz quadrada dos resíduos Studentizados",pch=19)
> lines(with(ex6,lowess(fitted(lm(prod ~ dose + I(dose^2) + I(dose^3),
+                          data=ex6)),
+                          sqrt(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3),
+                          data=ex6))^2))),col="red")
> hist(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      prob=T,main=" ",
+      xlab="Resíduos Studentizados", ylim=c(0,0.7))
> lines(density(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6))),
+      col="blue")
> qqPlot(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19)
> boxplot(rstudent(lm(prod ~ dose + I(dose^2) + I(dose^3), data=ex6)),
+      ylab="Resíduos Studentizados", pch=19, cex.lab=1.2, cex.axis=1.2,
+      col="turquoise")

```

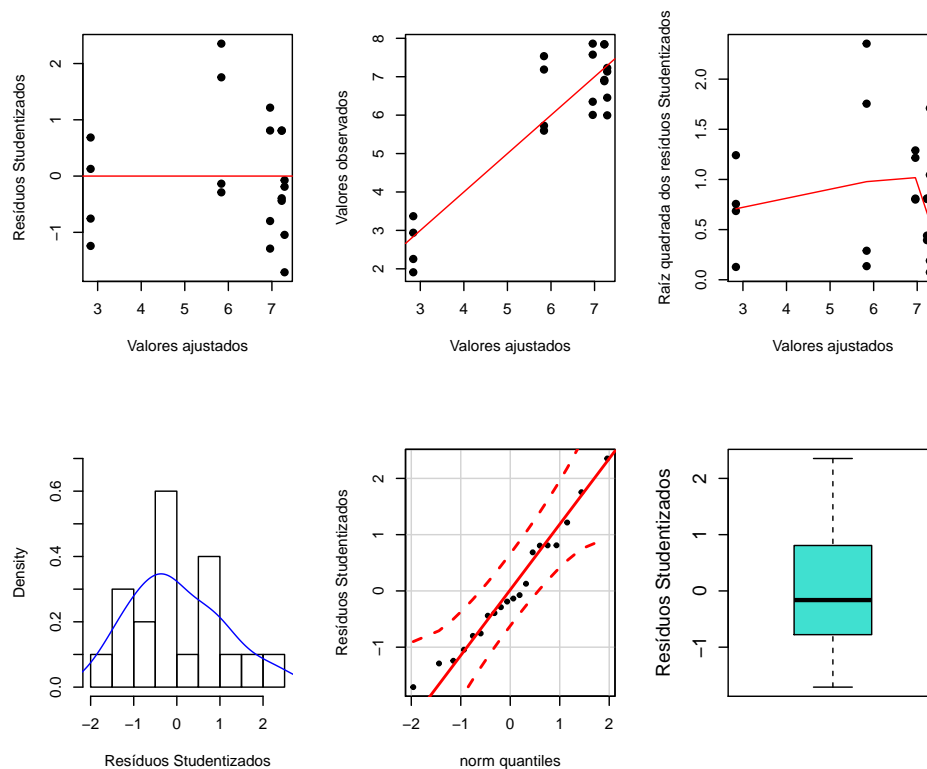


Figura 41: Gráficos de diagnósticos para verificar as pressuposições do polinômio de grau

Para construção do gráfico, é necessário escrever o modelo que será utilizado para plotar a curva, o que pode ser feito utilizando o comando `function`.

```
> funcC <- function (x, beta){
+   return (beta[1]+beta[2]*(x)+beta[3]*(x)^2+beta[4]*(x)^3)
+ }
```

O gráfico com a curva ajustada pode ser obtido da seguinte forma:

```
> plot(prod ~ dose, ylim=c(0, 10), xlab="Dose de composto (t/ha)",
+   ylab="Produtividade (t/ha)", data=ex6)
> curve(funcC(x, beta=coef(mod6.3)), from=0, to=16, type = "l",
+   col = "red", add = T, lty = 1, lwd=3)
```

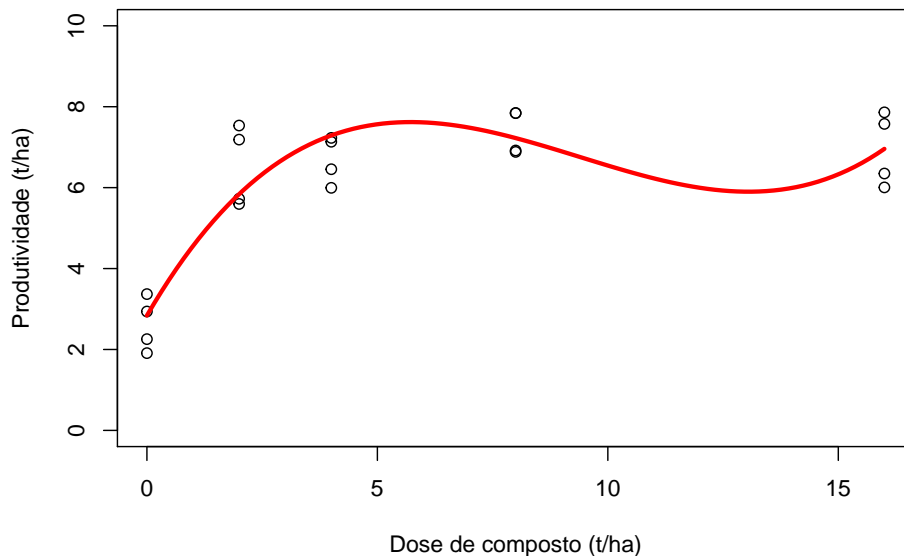


Figura 42: Modelo polinomial de 3º grau ajustado aos dados de produtividade em função da dose

Para verificar se existe a necessidade de utilizar um polinômio de grau maior que três para descrever o comportamento da produtividade em função da dose, pode-se realizar um teste de falta de ajuste. Neste caso, pode-se realizar o teste da razão de verossimilhanças, com o ajuste sequencial dos diferentes polinômios e o modelo maximal, obtido utilizando-se dose como fator qualitativo na descrição do modelo (Anova), da seguinte forma:

```
> anova(mod6.1,mod6.2,mod6.3,mod6, test="Chisq")
```

Analysis of Variance Table

```

Model 1: prod ~ dose
Model 2: prod ~ dose + I(dose^2)
Model 3: prod ~ dose + I(dose^2) + I(dose^3)
Model 4: prod ~ as.factor(dose)

```

```

  Res.Df    RSS Df Sum of Sq  Pr(>Chi)
1      18 47.395
2      17 20.904  1   26.4913 1.210e-11 ***
3      16 12.112  1    8.7914 9.415e-05 ***
4      15  8.647  1    3.4656  0.01421 *

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mesmo utilizando um polinômio de grau 3, observa-se que a variabilidade não explicada em relação ao modelo maximal ainda é significativa, evidenciando que estes modelos não são adequados para estes dados. Uma alternativa para análise dos dados é utilizar outros modelos não lineares.

### 3.9 Modelo de Mitscherlich

Uma extensão do modelo exponencial assintótico também conhecido como modelo Mitscherlich reparametrizado é apresentado por:

$$y = b_1 * (1 - e^{\log(1-0.99) * (\frac{dose-b_1}{b_2-b_1})})$$

$b_1$ : assíntota máxima;  
 $b_2$ : dose que confere 99% da assíntota máxima;  
 $b_3$ : parâmetro que controla a forma da função;

Para atribuição dos valores iniciais é necessário observar a dispersão dos dados e a interpretação dos parâmetros da função.

```

> mod6.nl <- nls(prod ~ b1*(1-exp(log(1-0.99)*((dose-b3)/(b2-b3)))),
+               start=list(b1=7.05, b2=4.2, b3=0.5), data=ex6)
> summary(mod6.nl)

```

```
Formula: prod ~ b1 * (1 - exp(log(1 - 0.99) * ((dose - b3)/(b2 - b3))))
```

Parameters:

```

  Estimate Std. Error t value Pr(>|t|)
b1  7.0594     0.2360  29.919 3.82e-16 ***
b2  4.2104     1.5222   2.766  0.0132 *
b3 -0.4724     0.1903  -2.483  0.0238 *

```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.7437 on 17 degrees of freedom
```

Number of iterations to convergence: 7  
Achieved convergence tolerance: 5.033e-06

```
> confint.default(mod6.nl)
```

```
          2.5 %      97.5 %  
b1  6.5969663  7.52188847  
b2  1.2269292  7.19386252  
b3 -0.8454348 -0.09945084
```

```
> coefR41<- coef(mod6.nl)
```

```
> R2(mod6.nl)
```

```
[1] 0.8627571
```

Avaliando a qualidade do ajuste:

```

> par(mfrow=c(2,3))
> plot(fitted(mod6.nl),residuals(mod6.nl),xlab="Valores ajustados",
+      ylab="Resíduos",pch=19)
> abline(h=0,col="red")
> plot(fitted(mod6.nl),ex6$prod,xlab="Valores ajustados",
+      ylab="Valores observados",pch=19)
> abline(a=0,b=1,col="red")
> plot(fitted(mod6.nl),sqrt(residuals(mod6.nl)^2),
+      xlab="Valores ajustados", ylab="Raiz quadrada dos resíduos",pch=19)
> lines(with(ex6,lowess(fitted(mod6.nl),sqrt(residuals(mod6.nl)^2))),col="red")
> hist(residuals(mod6.nl),prob=T,main=" ", xlab="Resíduos", ylim=c(0,0.7))
> lines(density(residuals(mod6.nl)), col="blue")
> qqPlot(residuals(mod6.nl), ylab="Resíduos", pch=19)
> boxplot(residuals(mod6.nl), ylab="Resíduos", pch=19, cex.lab=1.2,
+         cex.axis=1.2, col="turquoise")

```

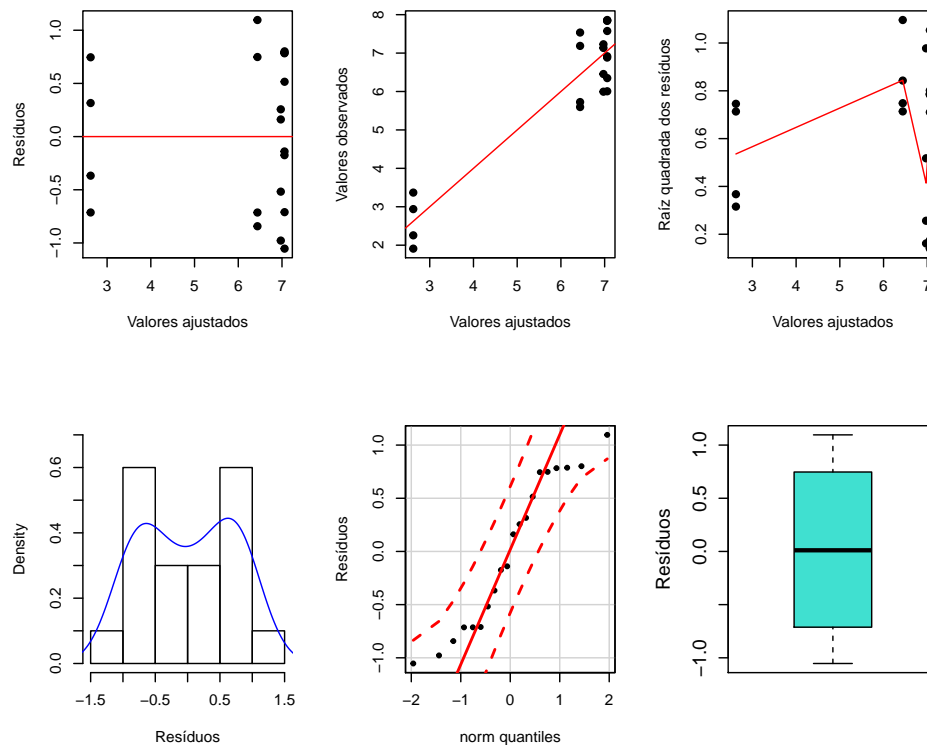


Figura 43: Gráficos de diagnósticos para verificar o ajuste

A função para plotar a curva no gráfico é escrita da seguinte forma:

```

> funcMI <- function (x, beta){
+ return(beta[1]*(1-exp(log(1-0.99)*((x-beta[3])/(beta[2]-beta[3])))))
+ }

```

O gráfico com a curva ajustada pode ser obtido da seguinte forma:



```

> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6)
> curve(funcMI(x, beta=coef(mod6.nl)), from=0, to=16, type="l",
+      col="red", add=T, lty=1, lwd=3)
> abline(h=coef(mod6.nl)[1], lty=2, lwd=2)
> abline(v=coef(mod6.nl)[2], lty=2, lwd=2)
> abline(v=confint.default(mod6.nl)[2,], col="blue", lty=2)

```

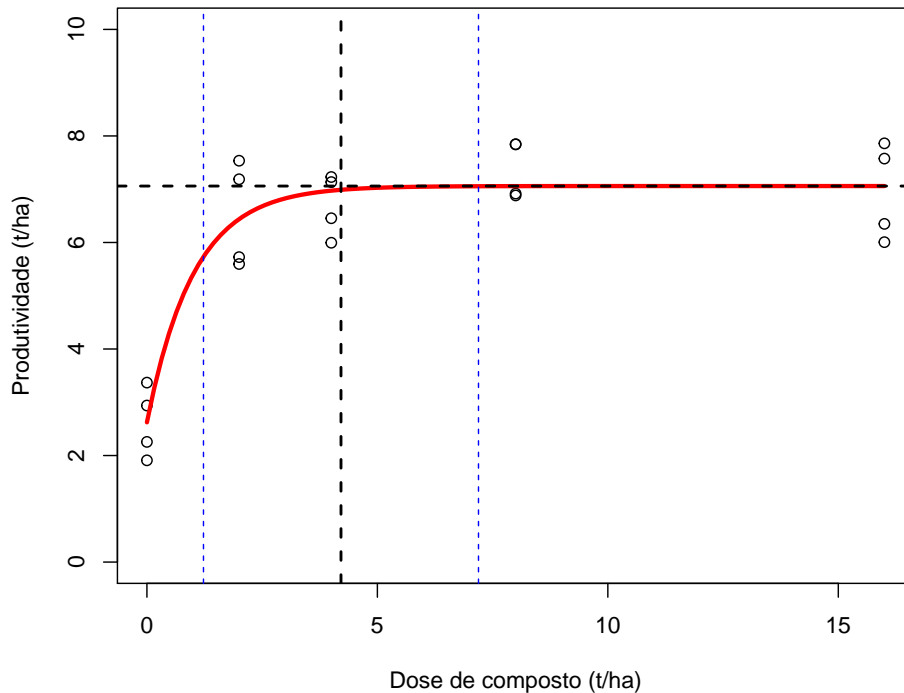


Figura 44: Modelo de Mitscherlich ajustado aos dados de produtividade em função da dose

Pode-se novamente realizar o teste da razão de verossimilhanças para verificar se a variabilidade não explicada pelo modelo é significativa:

```

> anova(mod6.nl, mod6, test="Chisq")

```

Analysis of Variance Table

Model 1:  $\text{prod} \sim b1 * (1 - \exp(\log(1 - 0.99) * ((\text{dose} - b3)/(b2 - b3))))$

Model 2:  $\text{prod} \sim \text{as.factor}(\text{dose})$

	Res.Df	Res.Sum Sq	Df	Sum Sq	F value	Pr(>F)
1	17	9.4022				
2	15	8.6469	2	0.75532	0.6551	0.5336

Como o teste da razão de verossimilhanças resultou em não significativo ( $p - \text{value} = 0.5336$ ), conclui-se que o modelo de Mitscherlich se ajusta bem aos dados, tanto quanto o modelo maximal.

O critério de informação de Akaike também pode ser utilizado para a comparação dos modelos. Nesse caso, quanto menor o valor de AIC, melhor o ajuste.

```
> AIC(mod6);
```

```
[1] 51.98695
```

```
> AIC(mod6.nl)
```

```
[1] 49.66185
```

Para obtenção do coeficiente de determinação do modelo não linear, pode-se utilizar a função R2 descrita anteriormente.

```
> R2(mod6.nl)
```

```
[1] 0.8627571
```

As diferentes curvas ajustadas podem ser observadas em um mesmo gráfico na figura a seguir:

```

> par(mfrow=c(2,2),mar=c(5,4,1,1))
> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6,pch=19)
> curve(funcL(x, beta=coef(mod6.1)), from=0, to=16, type = "l",
+       col = "blue", add = T, lty = 1, lwd=3);legend("topright", bty="n", "a")
> legend("topleft", bty="n", expression(R^2=='26,97'))
> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6,pch=19)
> curve(funcQ(x, beta=coef(mod6.2)), from=0, to=16, type = "l",
+       col = "darkgreen", add = T, lty = 1, lwd=3);legend("topright", bty="n", "b")
> legend("topleft", bty="n", expression(R^2=='64,90'))
> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6,pch=19)
> curve(funcC(x, beta=coef(mod6.3)), from=0, to=16,
+       type = "l", col = "brown", add = T, lty = 1, lwd=3);legend("topright", bty="n", "c")
> legend("topleft", bty="n", expression(R^2=='79,00'))
> plot(prod ~ dose, ylim=c(0,10), xlab="Dose de composto (t/ha)",
+      ylab="Produtividade (t/ha)", data=ex6,pch=19)
> curve(funcMI(x, beta=coef(mod6.nl)), from=0, to=16,
+       type = "l", col = "red", add = T, lty = 1, lwd=3);legend("topright", bty="n", "d")
> legend("topleft", bty="n", expression(R^2=='86,27'))
> legend(x=2,y=4,
+       legend=expression(y==b[1]*(1-e^(log(1-0.99)*
+       frac(dose-b[3],b[2]-b[3])))), bty="n")

```

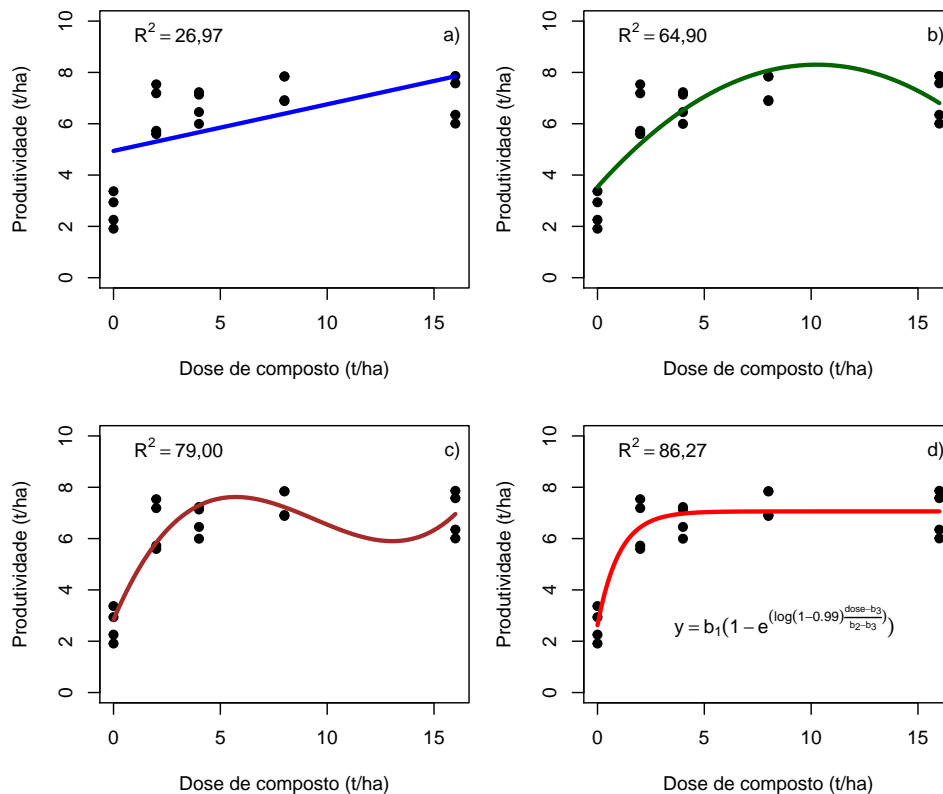


Figura 45: Modelos ajustados aos dados de produtividade de milho em função da dose de composto

## Referências

- ARMSTRONG, D. **factorplot: factorplot. R package version 1.1-1.** <http://CRAN.R-project.org/package=factorplot>, 2016.
- BERTOLDO, J.G.; ROCHA, F.; COIMBRA, J.L.M.; ZITTERELL, D.; GRAH, V.F. Teste de comparação de médias: dificuldades e acertos em artigos científicos. **R. Bras. Agrobiência**, Pelotas, v.13, n.4, p.441-447, out-dez, 2007.
- BERTOLDO, J.G.; COIMBRA, J.L.M.; GUIDOLIN, A.F.; MIQUELOTO, A.; TOALDO, D. Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico? **Ciência Rural**, Santa Maria v.38, n.4, p.1145-1148, 2008.
- BRIEN, C. **dae: Functions useful in the design and ANOVA of experiments.** R package version 2.1-7. <http://CRAN.R-project.org/package=dae>. 2011.
- CARDELLINO, R.A.; SIEWERDT, F. Utilização correta e incorreta dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v. 21, n. 6, p. 985-995, 1992.
- FERREIRA, E.B.; CAVALCANTI, P.P., NOGUEIRA, D.A. **ExpDes.pt: Experimental Designs package (Portuguese).** R package version 1.1.2, 2013.
- FOX, J.; WEISBERG, S. **An R Companion to Applied Regression**, Second Edition. Thousand Oaks CA: Sage. <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>, 2011.
- FURTADO, D.F. **Recursos Computacionais utilizando R.** Lavras, 2013. Disponível em: <http://www.dex.ufla.br/danielff/RRC0.pdf>. Acesso em: 10/05/2014
- KAPS, M.; LAMBERSON, W.R. **Biostatistics for Animal Science.** CABI Publishing Series, 445p, 2004.
- MENDIBURU, F. **agricolae: Statistical Procedures for Agricultural Research.** R package version 1.2-1. <http://CRAN.R-project.org/package=agricolae>, 2014.
- MONTANHINI NETO, R.; OSTRENSKY, A. Assessment of the use of statistical methods in articles published in a journal of veterinary science from 2000 to 2010. **Acta Scientiarum**. Technology, Maringá, v. 35, n. 1, p. 97-102, Jan.-Mar., 2013
- O'NEILL, M. E.; MATHEWS, K. L. A weighted least squares approach to levene test of homogeneity of variance. **Australian e New Zealand Journal Statistical**, 42(1):81-100, 2000.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria, 2015.

VENABLES, W.N.; RIPLEY, B.D. **Modern Applied Statistics with S**. Fourth Edition. Springer, New York. ISBN 0-387-95457-0, 2002.

WALKER, A. openxlsx: Read, Write and Edit XLSX Files. R package version 3.0.0. 2015. <https://CRAN.R-project.org/package=openxlsx>

ZEVIANE, W.M. Parametrizações interpretáveis em modelos não lineares. Tese (Doutorado em Estatística e Experimentação Agropecuária). 146 p. Lavras: UFLA, 2013.



[www.epagri.sc.gov.br](http://www.epagri.sc.gov.br)



[www.youtube.com/epagritv](http://www.youtube.com/epagritv)



[www.facebook.com/epagri](http://www.facebook.com/epagri)



[www.twitter.com/epagrioficial](http://www.twitter.com/epagrioficial)



[www.instagram.com/epagri](http://www.instagram.com/epagri)



[linkedin.com/company/epagri](http://linkedin.com/company/epagri)



<http://publicacoes.epagri.sc.gov.br>