



**MINISTÉRIO DA EDUCAÇÃO - MEC
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
CURSO DE BACHARELADO EM AGRONOMIA**

**PRESSUPOSIÇÕES E A ANÁLISE DE VARIÂNCIA DE
EXPERIMENTOS AGROPECUÁRIOS EM SOFTWARE LIVRE**

RENATO DUSMON VIEIRA

Morrinhos, GO

2016

RENATO DUSMON VIEIRA

**PRESSUPOSIÇÕES E A ANÁLISE DE VARIÂNCIA DE
EXPERIMENTOS AGROPECUÁRIOS EM SOFTWARE LIVRE**

Trabalho de conclusão de curso apresentado ao Instituto Federal Goiano – Câmpus Morrinhos, como requisito parcial para a obtenção do Grau de Bacharel em Agronomia.

Orientador: Prof. Dr. Jeferson Corrêa Ribeiro

Morrinhos – GO

Junho, 2016

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - SIBI/IF Goiano Campus Morrinhos

V657p Vieira, Renato Dusmon.

Pressuposições e a análise de variância de experimentos agropecuários em software livre. / Renato Dusmon Vieira. – Morrinhos, GO: IF Goiano, 2016.
30 f. : il.

Orientador: Dr. Jeferson Corrêa Ribeiro.

Trabalho de conclusão de curso (graduação) – Instituto Federal Goiano Campus Morrinhos, Bacharelado em Agronomia, 2016.

1. Pressupostos. 2. Análise de variância. 3. Software R.
I. Ribeiro, Jeferson Corrêa. II. Instituto Federal Goiano.
Curso de Bacharelado em Agronomia. III. Título

CDU 631:004.457(043)

RENATO DUSMON VIEIRA

**PRESSUPOSIÇÕES E A ANÁLISE DE VARIÂNCIA DE
EXPERIMENTOS AGROPECUÁRIOS EM SOFTWARE LIVRE**

Trabalho de Conclusão de curso DEFENDIDO e APROVADO em 06 de Junho de 2016
pela Banca Examinadora constituída pelos membros:

Prof. Dr. Wallacy Barbacena Rosa dos Santos
Membro - Arguidor
IF Goiano – Campus Morrinhos

Prof. Esp. Renato Silva Vasconcelos
Membro - Arguidor
IF Goiano – Campus Morrinhos

Prof. Dr. Jeferson Corrêa Ribeiro
Presidente - Orientador
IF Goiano – Campus Morrinhos

Morrinhos – GO

Julho, 2016

AGRADECIMENTOS

A Deus nosso pai e criador.

Ao Instituto Federal Goiano Campus Morrinhos pela oportunidade e por todo conhecimento que adquiri desde a época de Centro Federal de Educação (CEFET).

Ao Prof. Dr. Jeferson Corrêa Ribeiro pela orientação e ao mesmo tempo contribuição para meus estudos e especializações.

Aos Professores Dr. Wallacy Barbacena Rosa dos Santos e Especialista Renato Silva Vasconcelos pela participação na banca e pelas contribuições, além da grande amizade é claro.

Ao Prof. Dr. Nadson de Carvalho Pontes pelos conhecimentos transmitidos na graduação e pela orientação no mestrado em Olericultura.

Aos Professores Dr Geovane Barbosa do Nascimento (Reitor do IFBaiano) e Dr. Aderbal Almeida Rocha (Prof. do IFSP). Ambos contribuíram muito para minha formação no curso técnico em agropecuária (2005/2007) no CEFET/IFGoiano de Morrinhos.

Aos professores do Curso de Matemática da Universidade Estadual de Goiás Campus Morrinhos (UEG) pelo companheirismo na prática de Lecionar com responsabilidade.

Muito obrigado!

SUMÁRIO

RESUMO.....	06
ABSTRACT.....	07
1 INTRODUÇÃO.....	08
2 REVISÃO BIBLIOGRÁFICA.....	09
2.1 Histórico e Parâmetros da Anova.....	09
2.2 Delineamentos e as Pressuposições.....	12
2.2.1 Delineamento Inteiramente Casualizado (DIC).....	12
2.2.2 Delineamento em Blocos ao Acaso (DBC).....	12
2.2.3 Pressuposições dos Modelos.....	13
2.2.3.1 Aditividade.....	13
2.2.3.2 Normalidade.....	14
2.2.3.2.1 Teste de Shapiro-Wilk (1965).....	14
2.2.3.3 Independência.....	15
2.2.3.4 Homogeneidade.....	15
3 MATERIAL E MÉTODOS.....	15
4 RESULTADOS E DISCUSSÕES.....	17
5 CONCLUSÃO.....	20
6 CONSIDERAÇÕES FINAIS.....	21
REFERENCIAS.....	22
ANEXOS.....	24

Pressuposições e a Análise de Variância de Experimentos

Agropecuários em Software Livre

Assumptions and Variance Analysis Experiments

Agricultural Free Software

VIEIRA¹, Renato Dusmon.

RESUMO

O presente trabalho traz uma breve discussão sobre a história da estatística moderna, levando em consideração as contribuições de Fisher (1945). Foca-se em, discutir sobre o surgimento da Análise de Variância (analysis of variance - ANOVA) como método de comparação múltipla, e ao mesmo tempo, falar sobre a verificação das pressuposições dos modelos matemáticos nos delineamentos inteiramente casualizados e em blocos ao acaso. A verificação dos pressupostos se dará por meio de comandos que acionam os recursos de pacotes adicionais do software livre “R versão 3.2.3”. Foram utilizados dados de um experimento “Produção de mudas de tomateiro (*Solanum lycopersicum*) em diferentes substratos” para a discussão. O mesmo foi conduzido no Instituto Federal Goiano - Campus Morrinhos, sendo que o delineamento adotado foi o Inteiramente Casualizado (DIC) com sete tratamentos e cinco repetições. Os dados foram resumidos em médias e dispostos numa tabela do Excel, sendo salvos em formato (txt) que é passível de leitura pelo R. Os comandos foram redigidos no editor bloco de notas e colados no prompt do R para verificação da normalidade, homogeneidade e independência dos erros. O software R versão 3.2.3 demonstrou ser muito eficiente para a análise das pressuposições e análise dos dados na ANOVA com posterior comparação múltipla pelo teste de tukey a 5% de probabilidade. Sendo assim, o software R pode ser adotado em outros trabalhos devido à qualidade dos seus resultados e a sua gratuidade.

Palavras-Chave: Pressupostos, Analise de Variância, Software R.

¹ Professor na Universidade Estadual de Goiás (UEG) e no Instituto Federal Goiano Campus Morrinhos. Mestrando em Agronomia IFGoiano Campus Morrinhos. renato-dv@hotmail.com.

ABSTRACT

This paper provides a brief discussion of the history of modern statistics, taking into account the contributions of Fisher (1945). Focuses on, discuss the emergence of ANOVA (analysis of variance - ANOVA) and multiple comparison method, and at the same time, talk about the verification of the assumptions of the mathematical models in a completely randomized design and in blocks. The verification of assumptions will be through commands that trigger additional package features free software "R version 3.2.3." data from one experiment were used "tomato seedling production (*Solanum lycopersicum*) on different substrates" for discussion. The same was conducted at the Federal Institute Goiano - Morrinhos Campus, and the design adopted was randomized (DIC) with seven treatments and five replications. Data were summarized as means and arranged in an Excel table, being saved in format (.txt) that is readable by R. The commands were written in the editor notepad and pasted into the R prompt to verify the normality and homogeneity independence of errors. The R version 3.2.3 software has shown to be very efficient for the analysis of assumptions and analysis in ANOVA with subsequent multiple comparisons by Tukey test at 5% probability. Thus, the R software can be adopted in other work because of the quality of its results and its gratuitousness.

Keywords: Assumptions, Analysis of Variance, Software R 3.2.2.

1 INTRODUÇÃO

As ciências experimentais sempre buscam novas tecnologias que possam ser implantadas no cotidiano, melhorando assim, a qualidade de vida, a eficiência produtiva e, no caso da área agrícola, aumentando produtividade dentre outros fatores. Uma grande evolução dentro dos trabalhos com experimento, agropecuários principalmente, se deu a partir das contribuições de Ronald Aylmer Fisher que foi o criador no método de Análise de Variância. Ele também criou os delineamentos mais conhecidos (Inteiramente Casualizados e em Blocos) e seus princípios (Casualização, Repetição e Controle Local).

A estatística é uma ciência grandiosa e se aplica a todos os estudos experimentais e descritivos de modo geral (BOLFARINE, 2010). Ao realizar uma pesquisa e tomar decisão, por mais simples que seja, deve-se apresentar um tratamento estatístico aos dados, dando a eles uma confiabilidade pouco questionável. Assim, antes de realizar uma ANOVA precedem-se os seus testes de pressuposições.

Para verificar se as pressuposições (aditividade, homogeneidade das variâncias, independência e normalidade dos erros) estão sendo satisfeitas, pode-se usar, por exemplo, o teste de não aditividade de Tukey, teste de Lilliefors para normalidade da distribuição dos erros, teste de Bartlett para verificação da homogeneidade e DurbinWatson para independência dos erros, dentre outros (CARVALHO et al., 2010). A verificação destas pressuposições em softwares estatísticos agiliza o processo e o torna mais dinâmico.

Em 1995, inicia-se um projeto de criação do programa de código aberto “R” e o mesmo é um dos mais utilizados atualmente, principalmente em decorrência da sua gratuidade e qualidade dos resultados. É adotado em todo o mundo (CARVALHO et al., 2010). O programa R se tornou muito dinâmico por possibilitar a implementação das ferramentas que o usuário deseja ao contrário dos demais programas que se limitam neste sentido (PETERNELLI & MELLO, 2007).

O objetivo deste trabalho foi relatar sucintamente o surgimento da ANOVA, além de demonstrar a utilização do programa estatístico “R 3.2.3” em delineamento Inteiramente Casualizado.

2 REVISÃO BIBLIOGRÁFICA

2.1 Histórico e Parâmetros da Anova

Muitos pesquisadores, estudiosos e professores não sabem da história, nem ao menos das obras, mais utilizam diariamente as idéias de Fisher nas análises de pesquisas. Apesar de ser desconhecido por muitos, Fisher é considerado o maior estatístico do século XX. Foi o criador dos delineamentos básicos da experimentação, dos seus princípios e de toda metodologia da análise de variância. Suas teorias revolucionaram tanto a estatística como também a genética (CECON et al., 2012). Ele tinha uma habilidade intensa em ligar duas grandes áreas da ciência (Exatas e Biológicas) e com isso ganha destaque até os dias atuais.

A contribuição de Fisher (1890-1962) para a Estatística Experimental é, sem dúvida muito importante (BUSSAB, 2013). Formado em Matemática e Astronomia pela Universidade de Cambridge em 1912, foi o fundador do Laboratório de Estudos em Estatística experimental de Rothamsted, onde fez muitas contribuições para a Experimentação e para a Genética também (MIRSHAWKA, 1980).

Por muitos anos foi, e ainda é até hoje, considerado o maior contribuidor para a estruturação da Estatística Moderna. Fisher que introduziu conceitos de aleatorização e dos princípios básicos da experimentação: Casualização, Repetição e Controle Local (BARBIN, 2013). Um crescimento contínuo da Estatística moderna e experimental começou a partir dele, no que tange todas as ciências experimentais.

A Estatística Experimental é a ciência que tem como objetivo estudar experimentos (ensaios), englobando etapas como o planejamento, execução, coleta e análise dos dados experimentais e interpretação dos resultados obtidos (COSTA NETO, 2002). Um dos setores que mais se encontra trabalhos e publicações com aplicações da ANOVA é o agropecuário. A técnica é indispensável para a maioria das experiências agropecuárias de variáveis qualitativas e graças à interação com a regressão (para variáveis quantitativas) tem-se hoje um setor produtivo tão evoluído em publicações e que está em constante evolução, com diversos trabalhos de relevância. Além de tudo, ainda têm um grande potencial para continuar nesta direção.

A ANOVA fornece o embasamento teórico para que as médias dos tratamentos sejam comparadas dentro de uma margem de erro aceitável. Este erro é denominado de nível “ α (alfa)” de probabilidade, podendo ser aplicado nos valores de 1%, 5% ou 10%, onde, os dois primeiros são os mais utilizados na área agrícola e de ambos, o segundo (5%) é o mais abordado em vasta quantidade de trabalhos de pesquisas em todo país (BUSSAB, 2013).

O valor “ α ” representa o erro tolerável no experimento, sendo na maioria das vezes igual a 5% de probabilidade (PIMENTEL-GOMES, 2002). Seria o mesmo que afirmar que, em cem possibilidades têm-se cinco chances de estar errado, ou seja, tem-se 95 chances de acerto ao comparar os tratamentos com efeitos fixos.

Os tratamentos¹ podem ser comparados por meio de testes de hipóteses e com um fator² alfa³ fixado anteriormente. Podem ser feitas comparações com parâmetros de média, proporção, Variância ou Desvio Padrão. O mais comum é o de média, desde que represente fatores qualitativos para se enquadrar ao modelo de análise em questão (RESENDE, 2007).

Esta técnica de comparação age dentro de grupos e entre os tratamentos, servindo para verificar e identificar se existem diferenças significativas entre médias independentes (COSTA NETO, 2002). Em toda pesquisa simples ou experimental existem erros e no caso em questão, ele advém do chamado acaso. Assim, é necessária a avaliação previa (testes de pressuposições) dos dados para verificar se os valores advindos do experimento iram gerar uma conclusão confiável.

As pressuposições constituem em verificar se os dados experimentais seguem algum modelo probabilístico. Este descreve o comportamento dos dados e indicam se estão de acordo (ideais dentro do tolerável) para produzir uma inferência confiável. Caso contrário, as técnicas de comparações múltiplas serão verificações matemáticas que na prática podem gerar dúvidas. Uma das questões de utilizar softwares estatísticos na análise se deve a isto, além, é claro, de agilizar a análise e o tempo de execução dos cálculos.

Os programas computacionais utilizam várias casas decimais que em certos dados (contínuos) fazem as diferenças serem, ou não, significativas. Nesta revisão trabalha-se

¹ São valores que assume a variável ‘fator’;

² Variável cujo valor deseja-se conhecer;

³ Valor do erro tolerável em percentagem;

apenas com o Software Livre chamado “R versão 3.2.3”, onde se necessita ter certo conhecimento de códigos, visto que o mesmo só fornece o ambiente de programação. Têm-se, neste momento, a necessidade de conhecer uma vasta teoria em torno da condução de experimentos.

Em experimentação agropecuária três princípios básicos devem ser considerados: Casualização⁴, Repetição⁵ e Controle Local⁶ (VIEIRA, 1999). Este último é aplicado no Delineamento em Blocos Casualizados e em Quadrado Latino, sendo que neste segundo é aplicado duas vezes (linha e coluna). Os dois primeiros são necessários em todos os trabalhos de pesquisa, onde se deseja comparar médias em local homogêneo ou não. Agora, os três princípios juntos são necessários onde existe interesse de montar um experimento com heterogeneidade e variáveis não controláveis (BARBIN, 2013).

Segundo Pimentel-Gomes (2002), a aplicação dos princípios básicos da experimentação liga-se ao interesse do pesquisador em avaliar o efeito de algum tratamento. Dependendo do local onde se instala um experimento, podem existir diversas fontes de variação que interferem, devendo então tentar homogeneizar no máximo o local. É uma situação comum em campo aberto onde não se controla uma série de fatores, devendo então, instalar um experimento em Blocos ao Acaso (DBC).

No caso da Análise de Variância (em que a variável indicadora (x) dos tratamentos é qualitativa), os modelos podem incluir efeitos fixos e aleatórios (por exemplo, tratamentos com efeitos fixos e blocos de efeitos aleatórios, no delineamento em blocos ao acaso), mas as análises enfatizam os efeitos fixos, e os aleatórios são vistos como fatores de perturbação, não havendo interesse em testá-los (REDENDE, 2014).

Em um local protegido, como uma casa de vegetação, por exemplo, geralmente se adota uma montagem com apenas dois princípios: Casualização e Repetição. Porém, se existir heterogeneidades no ambiente (discutido acima), como, incidência de luz, irrigação e outros, adota-se os três princípios (também vistos anteriormente).

Em muitas situações do cotidiano há interesse em tomada de decisões e na Estatística Moderna e experimental não é diferente. Esta decisão se dá por rejeitar ou não determinada afirmação baseando-se em evidências (CECON et al., 2012). As

⁴ Disposição dos tratamentos ao acaso com a mesma probabilidade de todas as parcelas receberem tratamentos;

⁵ Refere-se ao número de vezes que o tratamento aparece no experimento;

⁶ Aplicado ao Delineamento em Blocos ao Acaso e Quadrado Latino;

hipóteses da estatística experimental, quando se tem interesse em fazer uma inferência são:

- H_0 (Hipótese Nula): que será atacada pelo teste estatístico;
- H_1 (Hipótese Alternativa): que representa o objetivo do trabalho;

2.2 Delineamentos e as Pressuposições

2.2.1 Delineamento Inteiramente Casualizado (DIC)

Este método de montagem de experimentos é considerado por muitos pesquisadores, o mais simples, isto se tratando de condução, análise e interpretação. Segundo BARBIN (2013) e PIMENTEL-GOMES (2002), os ensaios inteiramente ao acaso (também chamados assim) são muito utilizados em experimentos de laboratórios, casas de vegetação, viveiros etc. Assim sendo, o fato do ambiente ser homogêneo possibilita aplicação de dois princípios, casualidade e repetição para garantir a verificação dos testes e pressupostos.

O modelo estatístico adotado na análise de experimento nesta conformação tem a relação: $Y_{ij} = m + t_i + e_{ij}$, onde “ Y_{ij} ” representa os comprimentos da parte aérea e do sistema radicular, “ m ” a média geral dos dados, “ t_i ” é o efeito dos tratamentos e “ e_{ij} ” é o erro experimental ocorrido. Este modelo tem as suas pressuposições, ou seja, verificações para validade da ANOVA de um fator.

2.2.2 Delineamento em Blocos ao Acaso (DBC)

Adotado em situações onde não se tem controle de fatores que interferem no experimento. Em relação ao DIC, têm a adição dos blocos para distribuir os tratamentos em faixas mais homogêneas. Segue três princípios, casualização, repetição e controle local. Segundo Resende (2014), o modelo matemático é: $Y_{ij} = m + t_i + b_i + e_{ij}$, com adição do efeito de bloco em relação ao modelo anterior.

Constitui o método mais implantado em campo aberto em pesquisas agrícolas (PIMENTEL-GOMES, 2002). Têm os mesmos pressupostos para rodar a ANOVA,

porém, se adiciona uma linha de blocos no quadro proposto por Fisher em 1945, devendo assim, testar a aditividade dos dados. Analogamente, será adotado neste trabalho apenas um exemplo em DIC. Os comandos do R serão os mesmos para ambos os delineamentos devendo ser adotados na íntegra para um experimento em DBC, ao se verificar as pressuposições, adicionando a aditividade.

2.2.3 Pressuposições dos Modelos

Segundo RESENDE (2007), deve-se verificar antecipadamente os pressupostos de Normalidade de erros, Independência de erros e Homogeneidade de variâncias dos erros em um experimento em DIC. Fica assim, explícita a importância da análise de resíduos no contexto da análise de variância. Erros ou resíduos referem-se à discrepância entre os valores observados e preditos pelo modelo.

Estas pressuposições tem por objetivo estatístico, facilitar a interpretação dos resultados e testar a significância das hipóteses formuladas a priori. Enfim, o que pode ocorrer é a validade aproximada, que, no caso da normalidade, por exemplo, se garante por meio do teorema do limite central (TORMA et al., 2012). Para todas elas deve-se fixar o teste de hipóteses.

Quando algumas das pressuposições da análise não se verificam, existem alternativas que podem ser usadas. Dentre elas a transformação de dados com a posterior análise de variância destes dados transformados, tendo o mesmo efeito prático e estatístico na análise (BUSSAB, 2013).

2.2.3.1 Aditividade

Segundo BARBIN (2013), os modelos matemáticos e os dados experimentais (que depende o delineamento adotado) devem ser aditivos, ou seja, os efeitos devem se somar não havendo interação entre os mesmos. A aditividade é um dos pressupostos mais importantes, pois sua violação pode levar a inadequação do modelo matemático aos dados e vice-versa. Caso este pressuposto for não significativo no teste de hipóteses, deve-se aplicar a transformação logarítmica (RESENDE, 2014).

Este pressuposto será verificado pelo teste de aditividade de tukey, tendo as seguintes hipóteses:

- H_0 : Os dados são aditivos;
- H_1 : Os dados não são aditivos.

2.2.3.2 Normalidade

A falta de normalidade de dados em estatística paramétrica e qualitativa é verificada por meio dos erros. “Uma das exigências do modelo matemático é, portanto, da validade da análise de variância, é que os erros e_{ij} tenham distribuição normal” (BARBIN, 2013). Quando os erros são aproximadamente normais, pelo Teorema do Limite Central (TLC), que é válido ao se aumentar o tamanho da amostra, deduz-se assim, que os dados também são normais (RESENDE, 2014). O teste F (em questão) é robusto em relação à normalidade total, fornecendo assim conclusões válidas dentro de uma margem de erro tolerável.

A verificação destas pressuposições melhora a qualidade da análise dos experimentos, devendo ser aplicado antes de qualquer análise (RESENDE, 2007). Muitos softwares e aplicativos facilitam a Anova, porém, são poucos os que verificam pressupostos. Uma análise de variância com dados não confiáveis sempre roda nestes programas, o problema é a não adequação ao modelo matemático, devendo sempre verificar a satisfação. Isto leva a apenas uma verificação matemática que não tem efeito decisivo, ou seja, é erro. No caso de falsas conclusões podem-se citar os erros tipo I e erro tipo II (PIMENTEL-GOMES, 2002).

Existem vários procedimentos para avaliar a normalidade de dados. Dentre eles, se destacam os testes de Kolmogorov-Smirnov, de Shapiro-Wilk e de aderência do qui-quadrado (χ^2), que são os mais utilizados. Os testes de assimetria e curtose também permitem inferir sobre a normalidade (RESENDE, 2014).

2.2.3.2.1 Teste de Shapiro-Wilk (1965)

O objetivo do teste é verificar se as amostras de tamanho menores que 50 (cinquenta dados) têm ou não uma distribuição normal, dentro de certo nível de erro

tolerável no modelo. As hipóteses do modelo são formuladas para posterior verificação no Software “R 2.3.2”.

- H_0 : Os dados são normais;
- H_1 : Os dados não são normais.

2.2.3.3 Independência

A independência dos erros é, até certo ponto, garantida pelo princípio da casualização (BARBIN, 2013). Porém, é analisada matematicamente pelo teste de Durbin e Watson (cujas metodologia e descrição estão no anexo). Assim como nos demais testes fixam-se as hipóteses para este pressuposto:

- H_0 : Os dados são independentes;
- H_1 : Os dados não são independentes.

2.2.3.4 Homogeneidade

Este pressuposto é necessário para analisar se os erros de tratamentos apresentam variâncias similares (RESENDE, 2014). Os testes mais comuns são os de Levene e Bartlett. Este último, segundo esse mesmo autor, só é indicado em dados com normalidade, seja ela com os dados originais ou por meio de transformação raiz quadrada ou logarítmica. Em caso de normalidade significativa pode-se aplicar o teste de Levene que analisa a homogeneidade pela média e pela mediana dos dados, sendo em ambos os casos, necessário as hipóteses:

- H_0 : Os dados são Homogêneos;
- H_1 : Os dados não são Homogêneos.

3 MATERIAL E MÉTODOS

Para a verificação das pressuposições, análise de dados (ANOVA) e aplicação do teste de Tukey (5%) por meio de comandos do software livre R 3.2.3 foram utilizados os dados coletados em um experimento com diferentes substratos. O mesmo foi conduzido no ano de 2015 em uma casa de vegetação do setor de olericultura no Instituto Federal Goiano Campus Morrinhos.

O experimento em questão foi conduzido no Delineamento Inteiramente Casualizado (DIC). Foram realizadas cinco repetições em bandejas de polietileno de 200 células. Cada tratamento era composto por 30 células que receberam uma semente em cada uma. Foram coletados dados de quinze mudas centrais, sendo eles, comprimento de sistema radicular e parte aérea, conforme apresentados na tabela 1.

Tabela 1. Dados médios de comprimentos da parte aérea e de raízes (cm).

	test	test	test	test	test	Ep	Ep	Ep	Ep	Ep	...
rep	1	2	3	4	5	1	2	3	4	5	...
raiz	5	5,5	6	4	4,5	6,5	8	8	7	7,5	...
aerea	7,5	8	9,5	6,5	6	8	10,5	10,5	9,5	9	...

Fonte: Adaptado de de Arquivo pessoal.

Foi utilizado o software estatístico “R 3.2.3” e para organização dos dados foi utilizado o software Microsoft Excel (2010). Na tabela 1 são apresentados os dados do experimento (Anexo 01 e ex2) em formato txt. O “R” tem capacidade de importar os dados diretamente de tabelas organizadas em formato “.xls ou .txt”, conservando sua organização original. Toda vez que se abre o software “R” é necessário informar em qual diretório (pasta) esta salva a tabela de dados no computador.

Para que o mesmo software capture os dados, adotou-se o procedimento de importação de arquivos de blocos de notas por meio do comando “*dados2<-read.table("ex2.txt",h=T)*”. Logo após a instalação foram adicionados quatro pacotes incorporam os recursos necessários para a verificação dos pressupostos (*Car*: Serve para verificar os pressupostos de Independência e Homogeneidade; *Asbio*: A aditividade; *NorTeste*: Analisa a Normalidade).

Depois de instalados, os pacotes foram carregados, ou seja, acionados por meio do comando “*require (nome do pacote)*”. Os testes de pressuposições exigem os comandos do teste de Shapiro-Wilk ((*shapiro.test(residuals(lm(raiz+aerea~subs),data=dados2)*) que avalia a normalidade dos resíduos), teste de Levene e de Bartlett ((*leveneTest(raiz+aerea,subs,center=median)*) e (*bartlett.test(raiz+aerea~subs,data=dados2)*) para homogeneidade de variâncias em Delineamento Inteiramente Casualizado) e o teste de Durbin-Watson

rep – Repetição;
test – Testemunha;
Ep – Esterco Peneirado.

((`durbinWatsonTest(lm(raiz+aerea~subs),data=dados2)`) para independência. Este último nem precisa ser aplicado, pois, os dados foram obtidos independentemente (RESENDE, 2007).

A Análise de Variância se deu pela instalação do pacote “*ExpDes.pt*” que disponibiliza os testes de comparações múltiplas, também. O comando aplicado foi: (`dic(subs, raiz+aerea, quali = TRUE, mcomp = "tukey", sigT = 0.05, sigF = 0.05)` que além de fazer a ANOVA ainda faz a aplicação do testes de tukey com 5% de probabilidade)

4 RESULTADOS E DISCUSSÕES

Antes de proceder a ANOVA, assim como já foi discutido anteriormente, foi viabilizada a verificação das pressuposições. Ao rodar os pressupostos com a tabela 1 observa-se a existência de normalidade, homogeneidade e independência se $p > 0,05$ (PETERNELLI e MELLO, 2007). Isto indica a não rejeição de H_0 que afirma a existência destes requisitos iniciais aos dados.

De acordo com os resultados liberados pelo R (tabela 2) verifica-se que, pelo teste de normalidade de Shapiro-Wilk, se têm os dados normalmente distribuídos quando as médias originais (tabela 1) forem transformadas pelos métodos: Raiz quadrada e/ou logarítmica.

Os dados da tabela 1 não tiveram comportamento normalmente distribuído, visto que o valor “p-valor” foi menor que 5%. Isto indica a não rejeição da hipótese nula no tópico 2.2.3.2.1. Analogamente têm-se o mesmo resultado na pressuposição de homogeneidade pelo teste de Bartlett que não rejeita a hipótese nula do tópico 2.2.3.4.

O teste de Levene foi significativo apenas nos dados transformados, o que indicou a utilização dos mesmos neste formato, levando em consideração os demais pressupostos. Estes resultados dispostos logo abaixo na tabela dois se referem aos dados da tabela 1 quando se utiliza a transformação Raiz quadrada dos valores originais em centímetros. Estes dados estão no anexo 2 (ex2.txt), assim como os originais (anexo 1) e os dados de transformação Logarítmica (anexo 3 – ex3.txt). Os comando completos (rotina) estão todos no anexo 4, onde todas as linhas são alto explicativas.

Todos os pressupostos foram significativos na tabela 2, que se referem à análise dos dados do anexo 2 (transformação raiz quadrada). Os testes liberados pelo software R são significativos onde o P-Valor for maior que 0,05 (5%).

Tabela 2. Resultados dos pressupostos na transformação raiz quadrada.

```
> shapiro.test(residuals(lm(raiz+aerea~subs),data=dados2)) #amostras menores q
50.
```

Shapiro-Wilk normality test

```
data: residuals(lm(raiz + aerea ~ subs), data = dados2)
```

W = 0.66206, p-value = 0.6076 (60% é significativo para normalidade)

```
> lillie.test(residuals(lm(raiz+aerea~subs),data=dados2)) #Amostras maiores que
50.
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: residuals(lm(raiz + aerea ~ subs), data = dados2)
```

D = 0.31271, p-value = 0.689 (68,9% é significativo para normalidade)

```
> leveneTest(raiz+aerea,subs,center=mean) #homogeneidade pela média. Serve pra
DIC.
```

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

	Df	F value	Pr(>F)
group 6	1.7901	0.05573	(5,57% é significativo)

```
> bartlett.test(raiz+aerea~subs,data=dados2) #homogeneidade.
```

Bartlett test of homogeneity of variances

```
data: raiz + aerea by subs
```

Bartlett's K-squared = 8.7866, df = 6, p-value = 0.1859 (18,6% é significativo)

```
> durbinWatsonTest(lm(raiz+aerea~subs),data=dados2) #independência.
```

lag Autocorrelation D-W Statistic p-value

1	-0.01936092	2.025376	0.33 (0,33 é significativo)
----------	--------------------	-----------------	------------------------------------

Fonte: Arquivo pessoal. Liberado pelo R.

De acordo com o teste de Shapiro-Wilk a 5% de significância, os resíduos podem ser considerados normais. Assim como para o teste de *Kolmogorov-Smirnov* e

(ou) *Lilliefors* (que são os mesmos testes onde o segundo é uma adaptação do primeiro) também aponta normalidade para os dados transformados pela raiz quadrada e pelo logaritmo.

Observa-se que todos os resultados em que o p-valor foi maior do que 0,05 (5%) os testes foram significativos, ou seja, não se pode rejeitar a hipótese inicial (H_0). O teste de Bartlett é mais eficiente que o de Levene quando não se rejeita a hipótese de normalidade, e em casos (assim como os dados da tabela1) onde se verificam poucas variáveis (BARBIN, 2013 e CECON, 2012).

O pressuposto de normalidade se torna o mais importante que os demais em alguns casos e admitem-se certa margem de tolerância para falha destes outros (CARVALHO et al., 2010). Para dados transformados, o teste F que é feito no quadro da ANOVA se demonstrou muito robusto em relação a falhas de homogeneidade e independência, devendo assim, nesses casos, aplicá-la normalmente nos dados transformados, diminuindo assim, as distâncias e as variações entre os dados originais.

A ANOVA do sistema teve o seguinte resultado, para os dados do anexo 2, disposto na tabela 3: Em resposta ao comando: “*dic(subs, raiz+aerea, quali = TRUE, mcomp = "tukey", sigT = 0.05, sigF = 0.05)*”.

Tabela 3. Resultado da ANOVA

Quadro da análise de variância					
	GL	SQ	QM	Fc	Pr>Fc
Substratos	6	8.8057	1.46762	22.218	1.8996e-09
Resíduos/Erro	28	1.8495	0.06605		
Total	34	10.6552			

CV = 4.65%

Fonte: Arquivo pessoal. Resultado do R.

Teste de normalidade dos resíduos (Shapiro-Wilk)

p-valor: 0.607638

De acordo com o teste de Shapiro-Wilk a 5% de significância, os resíduos podem ser considerados normais.

De acordo com PIMENTEL-GOMES (2002), uma vez que o teste F da ANOVA for significativo, ou seja, verificar pelo menos uma diferença entre os tratamentos em questão deve-se proceder a um teste de Comparação múltipla. Aplica-se então o teste de

Tukey através do mesmo comando “*dic(subs, raiz+aerea, quali = TRUE, mcomp = "tukey", sigT = 0.05, sigF = 0.05)*” e o sistema libera a tabela 4. Nesta mesma tabela a conclusão se dá por meio de letras minúsculas. Letras iguais nas colunas indicam que não existe diferença entre os substratos na produção de mudas de tomateiro.

Tabela 4. Teste de comparação múltipla de Tukey com 5% de probabilidade

Grupos	Tratamentos	Médias
a	TEs (Terra + Esteco 3:1)	6.1496
a	TE (Terra + Entulho 1:1)	5.8646
a	Ep (Esterco peneirado)	5.7744
a	TC (Terra + Cinza 2:8)	5.6612
a	EP (Entulho peneirado)	5.6150
b	Test (Substrato Carolina®)	4.9260
b	CA (Cinza + Areia 2:8)	4.6303

Fonte: Arquivo do R. Médias seguidas de mesma letra não de diferem pelo teste de Tukey a 5% de Probabilidade.

De acordo com com o teste de comparação múltipla de tukey e a tabela 4, verifica-se que apenas os tratamentos testemunha (substrato comercial carolina®) e Cinza + Areia foram piores, ambos não se diferiram entre si, porém, foram menos eficientes para a produção de mudas de tomateiro cereja (tiveram menor média de tamanhos de mudas), dentro das condições do experimento. Os demais tratamentos tiveram comportamentos similares (dentro da margem de 5%) se destacando o Terra + Esterco com maior média geral.

5 CONCLUSÃO

O software R versão 3.2.3 se demonstrou eficiente para a análise de pressuposições do modelo matemático, para viabilizar a ANOVA (5%) em DIC e aplicar o teste de Comparação de Tukey (5%).

6 CONSIDERAÇÕES FINAIS

O software livre “R 3.2.3” possibilitou uma análise rápida e fácil das pressuposições do modelo matemático para um delineamento inteiramente casualizado, além de rodar a análise de variância por completo e disponibilizar a rejeição da hipótese nula através do valor p , ainda fez a comparação pelo teste de tukey a 0,05 de probabilidade. O teste de tukey é mais viável em casos onde se tenha um coeficiente de variação (CV) menor que 10%, o que foi verídico na tabela 3. Aplicar o teste de tukey com $CV > 10\%$ gera muita sobreposição e posteriores confusões nos resultados (BARBIN, 2013).

Sendo assim, o programa em questão é uma alternativa que deve ser considerada na prática, pois, uma vez se adaptando ao mesmo, evita-se assim, as dependências dos softwares sem licença e de propriedade privada, algo que aumenta os custos com pesquisas em geral.

REFERÊNCIAS

- BARBIN, D.. **Planejamento e análise estatística de experimentos agrônômicos**. 2 ed. Londrina: Mecenaz, 2013;
- BOLFARINE, H.. **Introdução à inferência estatística**. 2 ed. Rio de Janeiro: SBM, 2010;
- BUSSAB, W. O. **ESTATÍSTICA BÁSICA**. São Paulo: Saraiva, 2013;
- CARVALHO, R. C.; ALVES, S. D. F.; ALVES, L.; SILVEIRA, A. C. P. **Análise das Pressuposições do Modelo Matemático para Análise de Variância em Experimentos Agrícolas**. Anais do VIII Seminário de Iniciação Científica e V Jornada de Pesquisa e Pós-Graduação UNIVERSIDADE ESTADUAL DE GOIÁS 10 a 12 de novembro de 2010. Disponível em: <http://www.prp2.ueg.br/sic2010/apresentacao/trabalhos/pdf/agrarias/seminario/analise_das_pressuposicoes.pdf> Acessado em 16/06/2016;
- CECON, P. R. [et al]. **Métodos Estatísticos**. Viçosa: UFV, 2012;
- COSTA NETO, P. L.O.. **Estatística**. 2º edição. São Paulo: Editora Edgar Blucher, 2002.
- MIRSHAWKA, **Probabilidades e estatísticas para engenharia**. São Paulo: Nobel. 1980;
- MORETTIN, L. G.. **Estatística básica: probabilidade e inferência**. São Paulo: Pearson Prentice Hall, 2010;
- MEMÓRIA, J. M. P.. **Breve história da estatística**. Brasília, DF: Embrapa Informação Tecnológica, 2004. 111 p. ; 21 cm. – (Texto para discussão, ISSN 1677-5473 ; 21). Disponível em: <http://www.im.ufrj.br/~lpbraga/prob1/historia_estadistica.pdf>. Acessado em 10/06/16.
- PETERNELLI, L.A.; MELLO, M.P. **Conhecendo o R: uma visão estatística**. Viçosa: Ed UFV, 2007. 118 p.
- PIMENTEL-GOMES, F.. **Estatística aplicada a experimentos agrônômicos e florestais: exposição com exemplos e orientações para o uso de aplicativos**. Piracicaba: FEALQ, 2002;
- RESENDE, M. D. V. **ESTATÍSTICA MATEMÁTICA, BIOMÉTRICA E COMPUTACIONAL: Modelos Mistos, Multivariados, Categóricos e**

Generalizados, Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Espacial e Temporal, Competição, Sobrevivência. Viçosa: Suprema, 2014;

_____. **MATEMÁTICA E ESTATÍSTICA NA ANÁLISE DE EXPERIMENTOS: e no Melhoramento Genético.** Colombo: Embrapa Florestas, 2007.

SPIEGEL, M. R.. **Estatística.** Coleção Schaum. Tradução e Revisão Técnica Pedro Consentino. 3º edição. São Paulo: Pearson Makron Books, 1993

TORMA, V. B. L. T. [et al]. **Normalidade de variáveis: métodos de verificação e comparação de alguns testes não paramétricos por simulação.** Revista HCPA. 2012;32(2):227-234. 2012. Disponível em: < <http://seer.ufrgs.br/hcpa> >. Acessado em 14/01/16.

VIEIRA, S.. **Estatística para a qualidade: como avaliar com precisão a qualidade em produtos e serviços.** 7º tiragem. Rio de Janeiro: Elsevier, 1999.

ANEXOS

Anexo 1. Tabela 1. Dados médios originais das medidas de comprimento de raízes e parte aéreas de mudas de tomateiro cereja em DIC (cm)

subs	rep	raiz	aerea
test	1	5.0	7.5
test	2	5.5	8.0
test	3	6.0	8.5
test	4	4.0	6.5
test	5	4.5	6.0
Ep	1	6.5	8.0
Ep	2	8.0	9.8
Ep	3	8.0	9.8
Ep	4	7.0	9.5
Ep	5	7.5	9.7
CA	1	5.0	7.5
CA	2	3.5	5.0
CA	3	4.0	6.5
CA	4	5.0	7.5
CA	5	4.0	6.5
EP	1	7.0	9.2
EP	2	7.0	9.2
EP	3	6.5	8.0
EP	4	7.0	9.3
EP	5	8.0	9.0
TE	1	8.0	9.0
TE	2	9.0	9.5
TE	3	8.5	9.0
TE	4	7.5	9.2
TE	5	7.0	9.5
TC	1	6.0	8.5
TC	2	8.0	9.7
TC	3	8.0	9.7
TC	4	6.0	8.5
TC	5	7.0	9.3
TEs	1	9.8	9.9
TEs	2	9.5	9.8
TEs	3	9.9	9.0
TEs	4	9.7	9.0
TEs	5	9.0	9.0

Anexo 2. ex2. Dados transformados pela raiz quadrada das medidas de comprimento de raízes e parte aéreas de mudas de tomateiro cereja em DIC (cm)

subs	rep	raiz	aerea
test	1	2.236	2.738
test	2	2.345	2.828
test	3	2.449	2.915
test	4	2.000	2.549
test	5	2.121	2.449
Ep	1	2.549	2.828
Ep	2	2.828	3.130
Ep	3	2.828	3.130
Ep	4	2.645	3.082
Ep	5	2.738	3.114
CA	1	2.236	2.738
CA	2	1.870	2.236
CA	3	2.000	2.549
CA	4	2.236	2.738
CA	5	2.000	2.549
EP	1	2.645	3.033
EP	2	2.645	3.033
EP	3	2.549	2.828
EP	4	2.645	3.049
EP	5	2.828	3.000
TE	1	2.828	3.000
TE	2	3.000	3.082
TE	3	2.915	3.000
TE	4	2.738	3.033
TE	5	2.645	3.082
TC	1	2.449	2.915
TC	2	2.828	3.114
TC	3	2.828	3.114
TC	4	2.449	2.915
TC	5	2.645	3.049
TEs	1	3.130	3.146
TEs	2	3.082	3.130
TEs	3	3.146	3.000
TEs	4	3.114	3.000
TEs	5	3.000	3.000

Anexo 3. ex3. Dados transformados pelo logaritmo das medidas de comprimento de raízes e parte aéreas de mudas de tomateiro cereja em DIC (cm)

subs	rep	raiz	aerea
test	1	0.690	0.875
test	2	0.740	0.903
test	3	0.778	0.929
test	4	0.602	0.813
test	5	0.653	0.778
Ep	1	0.813	0.903
Ep	2	0.903	0.991
Ep	3	0.903	0.991
Ep	4	0.845	0.977
Ep	5	0.875	0.986
CA	1	0.690	0.875
CA	2	0.544	0.698
CA	3	0.602	0.813
CA	4	0.690	0.875
CA	5	0.602	0.813
EP	1	0.845	0.963
EP	2	0.845	0.963
EP	3	0.813	0.903
EP	4	0.845	0.968
EP	5	0.903	0.954
TE	1	0.903	0.954
TE	2	0.954	0.977
TE	3	0.929	0.954
TE	4	0.875	0.963
TE	5	0.845	0.977
TC	1	0.778	0.929
TC	2	0.903	0.986
TC	3	0.903	0.986
TC	4	0.778	0.929
TC	5	0.845	0.968
TEs	1	0.991	0.995
TEs	2	0.977	0.991
TEs	3	0.995	0.954
TEs	4	0.986	0.954
TEs	5	0.954	0.954

Anexo 4. Dicas para análise no “R 3.2.3”

1 BAIXAR OS ARQUIVOS DO R:

CAR: Serve para verificar os pressupostos de Independência e Homogeneidade.

ASBIO: Verifica a aditividade.

NORTESTE: Normalidade.

ExpDes.pt: DIC e DBC.

2 DEPOIS DE INSTALADOS, DEVE-SE ABILITAR OS PACOTES: Usar o comando `require` (nome do pacote):

```
require (ExpDes.pt)
```

```
require (car)
```

```
require (asbio)
```

```
require (nortest)
```

```
#####
```

3 INFORME AO SISTEMA ONDE ESTÁ OS DADOS E SUAS CARACTERÍSTICAS, PELOS COMANDOS:

```
dados2<-read.table("ex2.txt",h=T)
```

```
subs<-as.factor(dados2[,1]) # substrato
```

```
rep<-as.factor(dados2[,2]) #repetição é um fator
```

```
raiz<-dados2[,3] #diâmetro e variável resposta
```

```
aerea<-dados2[,4]
```

```
dados2<-data.frame(subs,rep,raiz,aerea) #construindo uma nova tabela de dados.
```

```
attach(dados2)
```

```
#####
```

4 VERIFIQUEM AS PRESSUPOSIÇÕES PELOS SEGUINTE COMANDOS:

```
tukey.add.test(raiz,aerea,subs) # testes para aditividade em DBC
```

```
shapiro.test(residuals(lm(raiz+aerea~subs),data=dados2)) #amostras menores q  
50. Teste de normalidade nos resíduos.
```

```
lillie.test(residuals(lm(raiz+aerea~subs),data=dados2)) #amostras maiores q 50.  
Nos resíduos.
```

-ANOVA e seus Pressupostos no software livre “R 3.2.3”-

leveneTest(raiz+aerea,subs, center=median) #homogenidade de variâncias pela mediana...OBS: só serve pra DIC.

leveneTest(raiz+aerea,subs,center=mean) #homogenidade pela média...OBS: só serve pra DIC.

bartlett.test(raiz+aerea~subs,data=dados2) #homogenidade...OBS: só serve pra DIC.Quando existe normalidade o teste de bartlett é o melhor para indicar homogeneidade.

durbinWatsonTest(lm(raiz+aerea~subs),data=dados2) #independência. OBS: Só usar este teste se não existir dependência que surge ao repetir medições em mesmo local.

#####

6 RODAR A ANOVA E O TESTE DE TUKEY PELOS COMANDOS:

dic(subs, raiz+aerea, quali = TRUE, mcomp = "tukey", sigT = 0.05, sigF = 0.05)
#dentro do ExpDes.pt