

Scientific Electronic Archives

Issue ID: Sci. Elec. Arch. Vol. 11 (6)

December 2018

Article link

<http://www.seasinop.com.br/revista/index.php?journal=SEA&page=article&op=view&path%5B%5D=560&path%5B%5D=pdf>

Included in DOAJ, AGRIS, Latindex, Journal TOCs, CORE, Discoursio Open Science, Science Gate, GFAR, CIARDRING, Academic Journals Database and NTHRYS Technologies, Portal de Periódicos CAPES.



Pressuposições e a análise de variância de experimentos agropecuários em software livre

Assumptions and variance analysis experiments agricultural free software

R. D. Vieira, J. P. E. Bianchini, W. D. Vieira, P. M. Silva, R. S. Vasconcelos, W. B. R. Santos, J. C. Ribeiro

Universidade Estadual de Goiás
Instituto Federal Goiano

Author for correspondence: renato-dv@hotmail.com

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* L) 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, Análise de Variância, Software R.

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.

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.

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.

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.

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: Arquivo pessoal do autor principal. rep: repetição; test: testemunha; Ep: Esterco Peneirado.

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

(`(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.

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 CECOM, 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)”.

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 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.

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 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.

Conclusões

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%).

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: Mecenas, 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 16/07/2017;

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_estatistica.pdf>. Acessado em 10/07/17.
- 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. 3ª 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 14/08/17.
- VIEIRA, S.. Estatística para a qualidade: como avaliar com precisão a qualidade em produtos e serviços. 7ª tiragem. Rio de Janeiro: Elsevier, 1999.