



Análise de variância e testes de comparação de médias: um estudo de caso

Lyvia Gonzalez Pagotto ¹
Josiane Rodrigues ²

Centro de Ciências Agrárias (CCA) da Universidade Federal de São Carlos (UFSCar)

Resumo

Um problema muito comum na área agrônômica é a comparação das médias de alguns tratamentos de interesse, de forma a verificar se existem ou não evidências de diferença entre elas, além de encontrar também, no caso de diferença significativa, qual(is) o(s) melhor(es) tratamento(s), dentro das expectativas da pesquisa. A maneira mais comum para tratar esse problema é a análise de variância (ANOVA). O teste F global da ANOVA testa a hipótese de igualdade entre as médias dos tratamentos comparados. Caso o teste F seja significativo, então a aplicação de testes de comparação de médias é feita, com o objetivo de investigar diferenças entre as médias dos tratamentos comparados. Um dos dilemas envolvidos com os testes de médias é justamente a sua aplicação condicional a um resultado preliminar significativo do teste F global na ANOVA. Nesse contexto, o objetivo do presente trabalho é avaliar, por meio da análise de um experimento real, se é possível encontrar divergências entre o resultado do teste F global e os seguintes testes de comparações múltiplas de médias comumente utilizados na análise de experimentos agrícolas: teste de Tukey e teste de Duncan, visando assim verificar se a aplicação dos testes de médias deve ser feita, de fato, apenas mediante um resultado significativo do teste F.

Palavras-chave: testes de comparações múltiplas; análise de variância; teste F global.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônômica. josi Rodrigues@ufscar.br



Introdução

Um problema muito comum em experimentos agronômicos é a comparação das médias de alguns tratamentos de interesse, de forma a verificar se existem ou não evidências de diferença entre eles, além de encontrar também, no caso dessa diferença ser significativa, qual(is) o(s) melhor(es) tratamento(s), dentro das expectativas da pesquisa.

De acordo com Girardi, Cargnelutti Filho e Storck (2009), a maneira mais usual para tratar esse problema é a análise de variância (do inglês, *Analysis of Variance* - ANOVA), um procedimento estatístico que compara, dentro de um experimento, a variação devida aos tratamentos com a variação devida ao acaso. A hipótese testada na ANOVA é a de que as médias dos tratamentos não diferem entre si, a determinado nível de significância, a qual pode ser representada da forma:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a,$$

em que a representa o número total de tratamentos do ensaio e μ_i representa a média populacional do i -ésimo tratamento, $i = 1, \dots, a$.

O teste estatístico utilizado para comparar as médias dos tratamentos na ANOVA foi proposto por Ronald Aylmer Fisher, e é chamado teste z de Fisher. Atualmente, o teste foi suprido pelo seu equivalente teste F de Snedecor, ou simplesmente teste F . Caso o teste F seja significativo, sendo os tratamentos qualitativos, então a aplicação de testes de médias é feita, com o objetivo de investigar eventuais diferenças entre pares de médias específicos ou combinações lineares dessas médias.

Um dos dilemas envolvidos com os testes de médias é a sua aplicação condicional a um resultado preliminar significativo do teste F global na ANOVA (HSU, 1996). De acordo com Cardellino e Siewerdt (1992), essa é uma questão polêmica e responsável por dividir a opinião de muitos pesquisadores, e que deve, portanto, ser melhor investigada.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agronômica. josi Rodrigues@ufscar.br



Nesse contexto, a motivação para o desenvolvimento do presente projeto consiste em verificar, por meio da análise de um conjunto de dados real, se é possível encontrar eventuais divergências entre o resultado do teste F global da ANOVA e o resultado de procedimentos de comparação de médias, ou seja, verificar se mesmo diante de um resultado não significativo do teste F global ainda assim um teste de médias pode acusar diferenças significativas entre as médias dos tratamentos estudados ou, ao contrário, tendo-se obtido um teste F significativo na ANOVA, algum teste de médias não encontra diferença significativa entre elas.

Dentro deste cenário, dois testes de comparações múltiplas de médias (comparações das médias de tratamentos entre si duas a duas) comumente utilizados na análise de experimentos agronômicos serão avaliados: teste de Tukey e teste de Duncan.

Metodologia

A fim de confrontar o resultado do teste F global da ANOVA com o resultado de procedimentos para a comparação de médias de tratamentos, em particular aqui o teste de Tukey (1953) e o teste de Duncan (1955), utilizou-se um conjunto de dados obtido de um experimento referente à caracterização morfológica e agronômica de genótipos de algodoeiro (*Gossypium hirsutum* L. r. *latifolium* Hutch.) no município de Uberaba, Minas Gerais, na região do Triângulo Mineiro (HENRIQUE; LACA-BUENDÍA, 2010). O experimento foi conduzido de Dezembro de 2008 a Maio de 2009 na fazenda escola das Faculdades Associadas de Uberaba – FAZU, localizada à longitude 47°57'22" WGR, latitude 19°44'6,82" S e à altitude de 775 m.

No experimento foi utilizado o delineamento casualizado em blocos com seis tratamentos e quatro repetições, em que foram avaliadas cinco variedades: Delta Opal (Delta and Pine), Delta Penta (Delta and Pine), BRS Cedro (EMBRAPA), IAC-25 (IAC), EPAMIG Precoce I e uma progênie IAC-06/191 (IAC). Cada uma das variedades foi avaliada com relação a uma série de variáveis morfológicas e agronômicas. Para o presente estudo, entretanto, foram utilizados apenas os dados referentes às variáveis “peso do capulho” e “índice de fibra”.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agronômica. josirodurigues@ufscar.br



Os dados referentes a cada uma das variáveis foram submetidos à ANOVA e, posteriormente, a aplicação dos testes de Tukey e Duncan, para a comparação das médias de tratamentos entre si duas a duas. Além disso, as pressuposições da ANOVA – normalidade, homogeneidade de variância e aditividade do modelo – foram verificadas por meio dos testes de Shapiro-Wilk (SHAPIRO; WILK, 1965), Levene (1960) e teste de não-aditividade de Tukey (1949), respectivamente. O nível de significância adotado em todos os casos foi o de 5%. Para realizar todas as análises estatísticas uma rotina foi desenvolvida usando o software R (2019).

Resultados e Discussão

Índice de Fibra

Os dados referentes à variável índice de fibra, em gramas, estão apresentados na Tabela 1.

Tabela 1 – Índice de fibra (g) das variedades de algodoeiro do experimento.

Variedade	Bloco			
	I	II	III	IV
Delta Opal	7,818	7,418	7,914	7,926
Delta Penta	7,778	7,497	8,495	8,442
BRS-Cedro	8,047	9,144	8,718	8,616
IAC-25	8,419	8,643	8,360	8,270
EPAMIG Precoce I	7,188	6,579	7,226	7,337
IAC-06/191	8,600	8,514	8,412	8,351

Fonte: Henrique e Laca-Buendía (2010).

Com relação às pressuposições do modelo, foi observado que os dados possuem distribuição normal (Teste de Shapiro-Wilk apresentou p-valor igual a 0,6739145), que há homogeneidade de variâncias (Teste de Levene apresentou p-valor igual a 0,21609) e que o modelo é aditivo (Teste de não-aditividade de Tukey apresentou p-valor igual a 0,1666278), ou seja, as pressuposições foram atendidas.

A partir da tabela da ANOVA referente à variável índice de fibra (Tabela 2), verifica-se que os tratamentos apresentaram resultado significativo (p-valor = 0,000103), ou seja, temos evidências de diferenças entre as médias das variedades comparadas.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônômica. josirodrigues@ufscar.br



Tabela 2 – Análise de variância dos dados referentes à variável índice de fibra (g).

Conjunto de Variáveis	Graus de liberdade	Somas de quadrados	Quadrados médios	F calculado	Pr(>F)
Tratamentos	5	6660	1,3320	11553	0,000103*
Blocos	3	247	0,0825	715	0,557937
Resíduos	15	1729	0,1153		

*significativo ao nível de 5%.

Fonte: Software R (2019).

Após a realização da ANOVA, foram aplicados os testes de Tukey e de Duncan, a fim de investigar quais médias de tratamentos diferiam entre si. Os resultados obtidos com cada um deles estão apresentados nas Tabelas 3 e 4, respectivamente. É possível notar que a variedade BRS-Cedro foi a que apresentou maior média para a variável índice de fibra. De acordo com o Teste de Tukey, entretanto, ela não foi considerada diferente das variedades IAC-06/191, IAC-25 e Delta Penta. Já pelo teste de Duncan, ela não foi considerada estatisticamente diferente das variedades IAC-06/191 e IAC-25 apenas.

Tabela 3 – Teste de Tukey aplicado às médias das variedades de algodoeiro com relação à variável índice de fibra.

Variedade	Média	Grupo*
BRS-Cedro	8,63125	a
IAC-06/191	8,46925	ab
IAC-25	8,42300	ab
Delta Penta	8,05300	ab
Delta Opal	7,76900	bc
EPAMIG Precoce I	7,08250	c

*Médias com letras iguais não diferem entre si pelo Teste de Tukey, ao nível de 5% de significância.

Fonte: Software R (2019).

Tabela 4 – Teste de Duncan aplicado às médias das variedades de algodoeiro com relação à variável índice de fibra.

Variedade	Média	Grupo*
BRS-Cedro	8,63125	a
IAC-06/191	8,46925	ab

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônômica. josirodurigues@ufscar.br



IAC-25	8,42300	ab
Delta Penta	8,05300	bc
Delta Opal	7,76900	c
EPAMIG Precoce I	7,08250	d

*Médias com letras iguais não diferem entre si pelo Teste de Duncan, ao nível de 5% de significância.

Fonte: Software R (2019).

Ao fazer o confronto entre o resultado do teste F global e os resultados dos testes de médias para os dados referentes ao índice de fibra das variedades de algodoeiro, percebe-se que houve concordância entre eles. O teste F da ANOVA rejeitou a hipótese nula de igualdade entre as médias dos tratamentos, dando evidências de que ao menos uma das médias diferia das demais. O teste de Tukey e o teste de Duncan, por sua vez, também encontraram diferenças entre as médias dos tratamentos comparados.

Peso de Capulho

Os dados referentes à variável peso de capulho, em gramas, estão apresentados na Tabela 5.

Tabela 5 – Peso de capulho (g) das variedades de algodoeiro.

Variedade	Bloco			
	I	II	III	IV
Delta Opal	5,5	5,2	5,8	4,7
Delta Penta	5,3	5,1	5,3	4,6
BRS-Cedro	5,1	6,2	5,9	5,6
IAC-25	5,6	5,3	5,6	5,1
EPAMIG Precoce I	5,1	4,9	5,2	5,6
IAC-06/191	4,9	5,9	5,8	5,2

Fonte: Henrique e Laca-Buendía (2010).

Com relação às pressuposições, foi observado que os dados possuem distribuição normal (Teste de Shapiro-Wilk apresentou p-valor igual a 0,2304574), que há homogeneidade de variâncias (Teste de Levene apresentou p-valor igual a 0,5871203) e que o modelo é aditivo, ou seja, não há interação entre tratamentos e blocos (Teste de não-aditividade de Tukey apresentou p-valor igual a 0,4391976).

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônômica. josi Rodrigues@ufscar.br



Pela tabela da ANOVA referente à variável peso de capulho (Tabela 6), verifica-se que os tratamentos não apresentaram resultado significativo (p -valor = 0,280), ou seja, não há evidências de diferença significativa entre as médias das variedades de algodoeiro quanto a essa variável.

Tabela 6 – Análise de variância dos dados referentes à variável peso de capulho (g).

Conjunto de Variáveis	Graus de liberdade	Somas de quadrados	Quadrados médios	F calculado	Pr(>F)
Tratamentos	5	0,9421	0,1884	1399	0,280 ^{ns}
Blocos	3	0,7579	0,2526	1876	0,177
Resíduos	15	2,0196	0,1346		

^{ns} não significativo ao nível de 5% de significância.

Fonte: Software R (2019).

Mesmo o teste F da ANOVA não apresentando evidências de diferença entre os tratamentos, os testes de Tukey e Duncan foram aplicados com a finalidade de verificar se eles também não apresentariam diferenças entre as variedades de algodoeiro. Os resultados obtidos com cada um deles estão apresentados nas Tabelas 7 e 8, respectivamente. É possível observar que dentre as variedades estudadas, nenhuma se destaca quanto ao peso do capulho, de acordo com o teste de Tukey. Já pelo teste de Duncan, é possível notar que a variedade BRS-Cedro foi considerada estatisticamente diferente da variedade EPAMIG Precoce I.

Tabela 7 – Teste de Tukey aplicado às médias das variedades de algodoeiro com relação à variável peso de capulho.

Variedade	Média	Grupo*
BRS-Cedro	5,700	a
IAC-06/191	5,450	a
IAC-25	5,400	a
Delta Penta	5,300	a
Delta Opal	5,200	a
EPAMIG Precoce I	5,075	a

*Médias com letras iguais não diferem entre si pelo Teste de Tukey, ao nível de 5% de significância.

Fonte: Software R (2019).

1 Graduanda em Biotecnologia. lygppto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônômica. josi Rodrigues@ufscar.br



Tabela 8 – Teste de Duncan aplicado às médias das variedades de algodoeiro com relação à variável peso de capulho.

Variedade	Média	Grupo*
BRS-Cedro	5,700	a
IAC-06/191	5,450	ab
IAC-25	5,400	ab
Delta Penta	5,300	ab
Delta Opal	5,200	ab
EPAMIG Precoce I	5,075	b

*Médias com letras iguais não diferem entre si pelo Teste de Duncan, ao nível de 5% de significância.

Fonte: Software R (2019).

Dessa forma, ao fazer o confronto entre o teste F da ANOVA e cada um dos testes de comparação de médias aplicado, verificou-se uma contradição entre o teste F e o teste de Duncan. Enquanto o primeiro não apontou diferenças significativas entre as médias dos tratamentos, o segundo detectou diferenças entre elas.

Considerações finais

A partir da realização do presente trabalho podemos concluir que nem sempre há concordância entre o resultado do teste F global da análise de variância (ANOVA) e os resultados obtidos com procedimentos de comparação de médias de tratamentos, mesmo em situações em que as pressuposições da análise estão todas sendo satisfeitas. Com relação à variável peso do capulho, por exemplo, verificamos que o teste F apresentou resultado não significativo, mas, ainda assim, o teste de Duncan acusou diferenças entre as médias de tratamentos comparadas.

Referências

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.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônoma. josirodrigues@ufscar.br



DUNCAN, D. B. Multiple Range and Multiple F Tests. **Biometrics**, Raleigh, v.11. n.1, p.1-42, 1955.

GIRARDI, L. H.; CARGNELUTTI FILHO, A.; STORCK, L. Erro tipo I e poder de cinco testes de comparação múltipla de médias. **Revista Brasileira de Biometria**, São Paulo, v.27, n.1, p.23-36, 2009.

HENRIQUE, F.H.; LACA-BUENDÍA, J.P. **Comportamento Morfológico e Agrônomo de Genótipos de Algodoeiro no Município de Uberaba – MG**. FAZU em Revista, Uberaba, n.7, p. 32- 36, 2010.

HSU, J. C. **Multiple Comparisons: theory and methods**. Great Britain: Chapman & Hall, 1996. 277p.

LEVENE, H. Robust tests for equality of variances. In: OLKIN, I.; GHURYE, S. G.; HOEFFDING, W.; MADOW, W. G.; MANN, H. B. (Ed.) **Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling**. Menlo Park: Stanford University Press, 1960, p.278-292.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2019. Disponível em: <http://www.R-project.org/>.

SHAPIRO, S. S.; WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, Oxford, v.52, n.3-4, p.591-611, 1965. Disponível em: <http://sci2s.ugr.es/keel/pdf/algorithm/articulo/shapiro1965.pdf>. Acesso em: 17 de abr. de 2020.

TUKEY, J. W. The problem of multiple comparisons. In: BRAUN, H. I. (Ed.) **The collected works of John W. Tukey, Volume VIII, Multiple comparisons: 1948-1983**. New York: Chapman & Hall, 1953. p.1-300.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agrônomo. josi Rodrigues@ufscar.br



TUKEY, J. W. One Degree of Freedom for Non-Additivity. **Biometrics**. Raleigh, v.5, n.3, p.232-242, 1949. Disponível em: <<https://www.jstor.org/stable/3001938?seq=1>>. Acesso em: 12 de abr. de 2020.

1 Graduanda em Biotecnologia. lygptto@gmail.com

2 Graduada em Matemática, Mestre e Doutora em Ciências com ênfase em Estatística e Experimentação Agronômica. josi Rodrigues@ufscar.br