

Análise de variância e testes de médias: um estudo aplicado em experimentos com variedades de algodoeiro e seleções de citrumelo

Lyvia Gonzalez Pagotto, Josiane Rodrigues, Francisco Humberto Henrique, Jorgino Pompeu Junior e Silvia Blumer

Análise de variância e testes de médias: um estudo aplicado em experimentos com variedades de algodoeiro e seleções de citrumelo

Lyvia Gonzalez Pagotto¹, Josiane Rodrigues¹, Francisco Humberto Henrique², Jorgino Pompeu Junior³ e Silvia Blumer⁴

¹Centro de Ciências Agrárias – CCA, Universidade Federal de São Carlos – UFSCar, Rodovia Anhanguera, km 174, CEP 13600-970, Araras, SP, Brasil. E-mail: lygptto@gmail.com, josirodrigues@ufscar.br

²HNRQ AgTec, CEP 13500-190, Rio Claro, SP, Brasil. E-mail: franciscohenrique@gmail.com

³Centro de Citricultura Sylvio Moreira, Instituto Agrônomo, Rodovia Anhanguera, km 158, CEP 13490-970, Cordeirópolis, SP, Brasil. E-mail: jorginopompeu@gmail.com

⁴CiaCamp – da Ciência ao Campo, Rodovia Anhanguera, km 158, 13490-970, Cordeirópolis, SP, Brasil. E-mail: silviablumer@gmail.com

RESUMO

Um problema muito comum em experimentos agrícolas é a comparação de alguns tratamentos de interesse, de forma a determinar qual(is) o(s) melhor(es) tratamento(s), dentro dos objetivos da pesquisa. A maneira mais comum para tratar esse problema é por meio da análise de variância (ANOVA) dos dados do experimento, procedimento este responsável por decompor a variação total dos dados do ensaio dentre todas as fontes previstas de variação. O teste F global da ANOVA testa a hipótese nula de igualdade entre as médias populacionais dos tratamentos comparados, a qual também é chamada de hipótese de homogeneidade. Caso o teste F seja significativo, então rejeita-se a hipótese de homogeneidade e faz-se a aplicação de testes de comparação de médias, com o objetivo de investigar quais delas diferem entre si. Entretanto, um dos dilemas envolvidos com os testes de médias é justamente a sua aplicação apenas mediante um resultado significativo do teste F na ANOVA. Enquanto alguns pesquisadores defendem essa prática, outros acreditam que o teste de médias deve ser aplicado qualquer que seja o resultado do teste F. Nesse contexto, por meio da análise de um conjunto de dados referente ao peso de capulho de variedades de algodoeiro e da análise de um conjunto de dados referente à eficiência produtiva de seleções de citrumelos, foi confrontado o resultado do teste F global e os resultados dos seguintes procedimentos de comparação de médias de tratamentos, comumente aplicados na área agrícola: teste de Tukey e teste de Duncan. A partir da análise foi possível concluir que nem sempre há concordância entre o resultado do teste F e os resultados dos testes de médias. Para a variável peso de capulho das variedades de algodoeiro, enquanto o teste da ANOVA apresentou resultado não significativo, ao nível de significância de 5%, o teste de Duncan acusou diferenças entre as médias comparadas, ao mesmo nível de probabilidade. Já para a variável eficiência produtiva de citrumelos, enquanto o teste F global apresentou

resultado significativo, ao nível de 5%, o teste de Tukey não acusou diferenças entre as médias comparadas, ao mesmo nível de significância.

Palavras-chave: teste F global, Tukey, Duncan, algodoeiro, citrumelo

ABSTRACT

A very common problem in agricultural experiments is the comparison of some treatments of interest, in order to determine which is the best, within the objectives of the research. The most common way to deal with this problem is through the analysis of variance (ANOVA) of the experiment data, a procedure responsible for decomposing the total variation of the data among all the sources of variation. The overall F-test of ANOVA checks the null hypothesis of equality between the population means of the compared treatments, which is also called the homogeneity hypothesis. If the F-test is significant, then the homogeneity hypothesis is rejected and the means comparison tests are applied, in order to investigate which ones differ from each other. However, one of the dilemmas involved with means tests is precisely its application only through a significant result of the F-test of ANOVA. While some researchers defend this practice, others believe the means tests should be applied whatever the result of the F-test. In this context, through the analysis of a data set referring the weight of boll of cotton varieties, and the analysis of a data set referring to the productive efficiency of citrus selections, the result of the overall F-test was confronted with the results of the following procedures for comparing treatment means, commonly applied in the agricultural area: Tukey and Duncan tests. From the analysis it was possible to conclude that there is not always agreement between the result of the F-test and the results of the means tests. For the variable weight of boll of cotton varieties, while the ANOVA test showed a non-significant result, at the 5% level of significance, the Duncan test showed differences between the compared means, at the same level of probability. As for the variable productive efficiency of citrumelles, while the overall F-test showed a significant result, at the level of 5%, the Tukey test did not show differences between the compared means, at the same level of significance.

Keywords: overall F-test, Tukey, Duncan, cotton, citrumelo

1. INTRODUÇÃO

Um problema muito comum em experimentos agrícolas 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 dos objetivos da pesquisa.

De acordo com Girardi et al. (2009), a maneira mais usual para tratar esse problema é a análise de variância (ANOVA) dos dados do experimento, sendo este um procedimento estatístico que compara a variação devida aos tratamentos com a variação devida ao acaso. A hipótese testada na ANOVA, também chamada de hipótese de homogeneidade, é a de que as

médias populacionais 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 básico utilizado para comparar as médias dos tratamentos na ANOVA foi proposto por Ronald Aylmer Fisher, e é conhecido como teste z de Fisher. Atualmente, o teste foi substituído pelo seu equivalente teste F de Snedecor, ou simplesmente teste F . Caso o teste F seja significativo, sendo os tratamentos com mais de dois níveis e de natureza qualitativa, então a aplicação de testes de comparação de médias é feita, com o intuito de investigar eventuais diferenças entre pares de médias específicos ou combinações lineares dessas médias.

Os fundamentos dos testes de médias foram estabelecidos entre as décadas de 1940 e 1950, principalmente por David Duncan, Samarendra Nath Roy, Henry Scheffé e John Tukey, embora algumas das ideias sobre o assunto tenham aparecido muito mais cedo nos trabalhos de Ronald Aylmer Fisher, William Sealy Gosset (Student), dentre outros (HOCHBERG; TAMHANE, 1987).

Os testes de comparação de médias têm merecido a atenção de muitos pesquisadores e, embora a literatura sobre o assunto seja ampla, o campo ainda se encontra aberto para futuras pesquisas (NOGUEIRA, 2007). O que se tem hoje é um número elevado de procedimentos para a comparação de médias de tratamentos, cada um deles fundamentado num conjunto de suposições que os tornam específicos para determinados fins.

De acordo com um estudo realizado, o único método estatístico aplicado com mais frequência que os testes de médias é o teste F global da ANOVA (MEAD; PIKE, 1975 apud HSU, 1996). No entanto, enquanto testes de comparação de médias ocupam o segundo lugar em frequência de uso, eles talvez ocupem o primeiro lugar no que diz respeito a sua utilização de forma inadequada (HSU, 1996).

Nos levantamentos realizados por Petersen (1977), Cardellino e Siewerdt (1992), Santos et al. (1998), Bezerra Neto et al. (2002), Bertoldo et al. (2008a), Bertoldo et al. (2008b) e Souza et al. (2012), por exemplo, verificou-se que muitas vezes esses testes são utilizados de forma indiscriminada, o que pode levar a informações incompletas ou errôneas acerca dos tratamentos investigados.

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). Enquanto alguns pesquisadores defendem o uso condicional dos testes, outros defendem que eles devem ser aplicados qualquer que seja o resultado do teste F . 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.

Nesse contexto, a motivação para o desenvolvimento do presente estudo consiste em verificar, por meio da análise dois conjuntos de dados agrônômicos referentes a um experimento para comparação de variedades de algodoeiro e a um experimento para comparação de seleções de citrêmelos, 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 agrônômicos serão avaliados: teste de Tukey e teste de Duncan.

2. MATERIAL E MÉTODOS

2.1. Experimento com variedades de algodoeiro

O experimento foi conduzido de dezembro de 2008 a maio de 2009, no município de Uberaba, Minas Gerais, na Fazenda Escola das Faculdades Associadas de Uberaba-FAZU. O clima, segundo a classificação de Köppen, é AW (clima tropical quente e úmido com inverno frio e seco). O solo é do tipo Latossolo Vermelho distrófico, textura arenosa, com 74% de areia, 19% argila e 7% de silte (EMBRAPA, 2006). O objetivo do ensaio foi avaliar o comportamento morfológico e agrônômico de um novo genótipo e cinco cultivares de algodoeiro (*Gossypium hirsutum* L. r. *latifolium* Hutch.).

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). As parcelas foram constituídas de quatro linhas de cinco metros de comprimento com espaçamento entre linhas de 0,70 m, sendo que as duas linhas centrais foram consideradas, com a área útil de 3,5 m² e as outras duas linhas, uma de cada lado, foram as bordaduras.

Cada uma das variedades foi avaliada com relação a uma série de variáveis morfológicas e agrônômicas. Para o presente estudo, entretanto, foram utilizados apenas os dados referentes à variável “peso de capulho”, a qual está relacionada às características da fibra. Por ocasião da colheita, foram sorteados 10 capulhos de cada fileira da área útil, no terço médio da planta. Após seu descaroçamento, foram feitas várias medidas, dentre elas o peso de um capulho, em gramas.

Os dados foram submetidos à ANOVA e, posteriormente, à comparação múltipla das médias pelo teste de Tukey e pelo teste de Duncan, a fim de confrontar os resultados obtidos. As pressuposições de normalidade, homogeneidade de variâncias e aditividade entre tratamentos e blocos foram verificadas pelos testes de Shapiro-Wilk, Levene e teste de não-aditividade de Tukey, respectivamente. O nível de significância adotado em todos os casos foi o de 5%, e todas as análises estatísticas foram feitas usando o software R (R CORE TEAM, 2020).

2.2. Experimento com seleções de citrumelo

O experimento foi instalado em maio de 2003, no município de Comendador Gomes, Minas Gerais, sob clima CWA (clima subtropical de inverno seco), em um Latossolo Vermelho-amarelo, tendo como objetivo determinar a tolerância à doença MSC (morte súbita dos citros) de seleções de citrumelos das séries F.80 e F.81, bem como suas influências no crescimento e produção da laranjeira Valência.

O ensaio foi instalado em área anteriormente ocupada por pomares erradicados pela presença da MSC e vizinho de pomares portadores da doença. Foram avaliados os citrumelos (*C. paradisi* Macfaden x *P. trifoliata* Rafinesque) a seguir relacionados com seus números de registro no Banco Ativo de Germoplasma de Citros do Instituto Agrônomo de Campinas: Citrumelo Swingle, Citrumelo W-2 (1455), Citrumelo F.80.6 (1456), Citrumelo F.80.5 (1457), Citrumelo F.80.7 (1458), Citrumelo F.80.8 (1459), Citrumelo F.80.3 (1460), Citrumelo F.80.18 (1461), Citrumelo F.80.14 (1498), Citrumelo F.81.18 (1499) e Citrumelo 4481 (1704).

Como copa utilizou-se a laranjeira Valência (*C. sinensis* L. Osbeck), por ser suscetível a MSC quando enxertada nos limões Cravo e Volkameriano e pela ausência de relatos de sua incompatibilidade (formação de anel de goma na linha de enxertia) com outros porta-enxertos, exceto limão Rugoso (BRIDGES; YOUTSEY, 1968).

Utilizou-se o delineamento inteiramente casualizado com nove tratamentos (citrumelos) e cinco repetições. Cada um dos citrumelos foi então avaliado com relação à uma série de variáveis, do ano de 2006 ao ano de 2010, sendo uma delas a eficiência produtiva, a qual pode ser calculada pelo quociente da massa de frutos colhida num determinado período pelo respectivo volume das copas nesse mesmo período. Para fins de confronto entre o resultado do teste F da ANOVA e os testes de Tukey e de Duncan, utilizou-se aqui a eficiência produtiva referente ao ano de 2007.

Os dados foram submetidos à ANOVA e, posteriormente, à comparação múltipla das médias por cada um dos testes de médias. As pressuposições de normalidade e homogeneidade de variâncias foram verificadas pelos testes de Shapiro-Wilk e Bartlett, respectivamente. O nível de significância adotado em todos os casos foi o de 5%, e todas as análises estatísticas foram feitas usando o software R (R CORE TEAM, 2020).

3. RESULTADOS E DISCUSSÃO

3.1. Experimento com variedades de algodoeiro

Antes de proceder com a ANOVA dos dados, verificou-se inicialmente as pressuposições de normalidade (valor de $p = 0,2305$), homogeneidade de variâncias (valor de $p = 0,5871$) e aditividade entre tratamentos e blocos (valor de $p = 0,4392$). Em seguida, fez-se a ANOVA dos dados do experimento, onde verificou-se que os tratamentos não apresentaram resultado significativo, ou seja, não houve evidências de diferença entre as médias das variedades de algodoeiro quanto à variável peso de capulho (Tabela 1).

Tabela 1. Análise de variância dos dados referentes à variável peso de capulho (g) das variedades de algodoeiro.

Causas de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F	Valor de p
Tratamentos	5	0,9421	0,1884	1,3990	0,2800 ^{ns}
Blocos	3	0,7579	0,2526	1,8760	0,1770
Resíduo	15	2,0196	0,1346		
Total	23	3,7196			

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

Mesmo o teste F da ANOVA não apresentando resultado significativo, os testes de Tukey e de Duncan foram aplicados, com a finalidade de verificar se eles também não apresentariam diferenças entre as variedades comparadas (Tabela 2). Foi possível observar, pela aplicação do teste de Tukey, que dentre as variedades estudadas, a BRS-Cedro apresentou maior média para a variável peso de capulho, mas essa média não foi considerada estatisticamente diferente das demais. Já pelo teste de Duncan, foi possível notar que a variedade BRS-Cedro foi considerada estatisticamente diferente da variedade EPAMIG Precoce I, ou seja, o teste de Duncan acusou diferença entre as médias comparadas.

Tabela 2. Teste de Tukey e teste de Duncan aplicados às médias das variedades de algodoeiro com relação à variável peso de capulho (g).

Variedade	Média	
	Tukey	Duncan
BRS-Cedro	5,700 ^a	5,700 ^a
IAC-06/191	5,450 ^a	5,450 ^{ab}
IAC-25	5,400 ^a	5,400 ^{ab}
Delta Penta	5,300 ^a	5,300 ^{ab}
Delta Opal	5,200 ^a	5,200 ^{ab}
EPAMIG Precoce I	5,075 ^a	5,075 ^b

Médias seguidas pela mesma letra não diferem entre si, ao nível de 5%.

Dessa forma, ao fazer o confronto entre o teste F da ANOVA e cada um dos testes de comparação de médias aplicados, 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 acusou diferenças entre elas.

3.2. Experimento com seleções de citrumelo

Inicialmente foram verificadas as pressuposições de normalidade (valor de $p = 0,4599$) e homogeneidade de variâncias (valor de $p = 0,9101$) dos dados do experimento. Em seguida,

fez-se a ANOVA dos dados, onde verificou-se que os tratamentos apresentaram resultado significativo, ou seja, houve evidências de diferença entre as médias dos citrumelos quanto à variável eficiência produtiva (Tabela 3).

Tabela 3. Análise de variância dos dados referentes à eficiência produtiva dos citrumelos.

Causas de Variação	Graus de Liberdade	Somas de Quadrados	Quadrados Médios	F	Valor de p
Tratamentos	8	605,64	75,7050	2,3114	0,0410*
Resíduo	36	1179,13	32,7540		
Total	44	1784,77			

* significativo, ao nível de 5%.

Em seguida, procedeu-se com a aplicação dos testes de médias (Tabela 4). Foi possível observar, pela aplicação do teste de Tukey, que todos os citrumelos foram considerados iguais entre si, com relação à eficiência produtiva. Já pelo teste de Duncan, foi possível notar diferença significativa entre os citrumelos F80.18, F80.5, F80.3 e o citrumelo Swingle.

Tabela 4. Teste de Tukey e teste de Duncan aplicados às médias dos citrumelos com relação à variável eficiência produtiva.

Citrumelo	Média	
	Tukey	Duncan
F80.18	21,96 ^a	21,96 ^a
F80.5	21,54 ^a	21,54 ^a
F80.3	14,94 ^a	14,94 ^a
F80.6	14,08 ^a	14,08 ^{ab}
F80.7	13,88 ^a	13,88 ^{ab}
W2	12,84 ^a	12,84 ^{ab}
F80.8	12,44 ^a	12,44 ^{ab}
4481	12,34 ^a	12,34 ^{ab}
Swingle	11,90 ^a	11,90 ^b

Médias seguidas pela mesma letra não diferem entre si, ao nível de 5%.

Assim, ao fazer o confronto entre o resultado teste F da ANOVA e os resultados de cada um dos testes de médias, verificou-se uma contradição entre o teste F e o teste de Tukey. Enquanto o primeiro apontou diferenças significativas entre as médias dos tratamentos, o segundo não encontrou nenhuma diferença entre elas.

4. CONCLUSÕES

1. A partir da realização do presente trabalho foi possível verificar que nem sempre há concordância entre o resultado do teste F global da 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.
2. Com relação à variável peso de capulho das variedades de algodoeiro, verificou-se que o teste F apresentou resultado não significativo, mas, ainda assim, o teste de Duncan acusou diferenças entre as médias de peso das variedades.
3. Com relação à variável eficiência produtiva de seleções de citrumelos, verificou-se que o teste F apresentou resultado significativo e, ainda assim, o teste de Tukey não encontrou diferenças entre as médias dos citrumelos.

5. REFERÊNCIAS

- BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MANTOVANI, A.; VALE, N. M. do. 2008. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. **Revista Biotemas**, Florianópolis, v.21, n.2, p.145-153.
- BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MIQUELOTO, A.; TOALDO, D. 2008. Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico? **Ciência Rural**, Santa Maria, v.38, n.4, p.1145-1148.
- BEZERRA NETO, F.; NUNES, G. H. S.; NEGREIROS, M. Z. de. 2002. Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1983 a 2000. **Horticultura Brasileira**, Brasília, v.20, n.1, p.5-9.
- BRIDGES, G. D.; YOUTSEY, C. O. 1968. Further studies of the bud-union abnormality of rough lemon rootstocks with sweet orange scions. In: **Conference of the International Organization of Citrus Virologists**, 4., 1966, Catania. Proceedings. Gainesville: International Organization of Citrus Virologists, p. 236-239.
- CARDELLINO, R. A.; SIEWERDT, F. 1992. Utilização correta e incorreta dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v.21, n.6, p.985-995.
- GIRARDI, L. H.; CARGNELUTTI FILHO, A.; STORCK, L. 2009. 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.
- HOCHBERG, Y.; TAMHANE, A. C. 1987. **Multiple comparisons procedures**, New York: John Wiley & Sons. 450p.

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

NOGUEIRA, M. C. S. 2007. **Experimentação Agrônômica I – conceitos, planejamentos e análise estatística**, Piracicaba: ESALQ. 479p.

PETERSEN, G. R. 1977. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v.69, n.2, p.205-208.

R CORE TEAM (2020). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

SANTOS, J. W. dos; MOREIRA, J. A. N.; BELTRÃO, N. E. de M. 1998. Avaliação do emprego dos testes de comparação de médias na revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994. **Pesquisa Agropecuária Brasileira, Brasília**, v.33, n.3, p.225-230.

SOUZA, C. A.; LIRA JUNIOR, M. A.; FERREIRA, R. L. C. 2012. Avaliação de testes estatísticos de comparações múltiplas de médias. **Revista Ceres, Viçosa**, v.59, n.3, p.350–354.