Scientia Agraria Paranaensis – Sci. Agrar. Parana.
ISSN: 1983-1471 – Online
DOI: https://doi.org/10.18188/sap.v21i2.29243

MULTIPLE COMPARISON AND CLUSTERING STATISTICAL TESTS IN THE SOFTWARE RBIO FOR LETTUCE AND MAIZE CROPS

Alex Oliveira Smaniotto[1*], Bruna Elaine de Almeida Silva[2], João Pedro Maia[3], Magno Gonçalves Braz[2], Juliano Henrique de Freitas Neto[2], Simério Carlos Silva Cruz[2]

SAP 29243 Received: 02/22/2021 Accepted: 05/19/2022
Sci. Agrar. Parana., Marechal Cândido Rondon, v. 21, n. 2, apr./jun., p. 126-130, 2022

ABSTRACT - The objective of this study was to evaluate the efficiency and uniformity of multiple comparison tests when compared to clustering test applied in the software RBio. The evaluations were carried out using data of agricultural experiments conducted by the authors, in the experimental field of the Federal University of Goiás. The data analyzed were from three experiments conducted for lettuce and maize crops: the first was conducted in a completely randomized design; the second in a randomized block design; and the third in a randomized block design with split-plot arrangement. The evaluation of the data collected in the lettuce and maize crops was carried out using the software Rbio. The data were subjected to analysis of variance by the F test at 5% probability. The means were compared by multiple comparison (Tukey, Duncan, and Student-Newman-Keuls), and clustering (Scott-Knott) tests. The lower rigor of the Tukey, Student-Newman-Keuls, and Duncan tests results in higher incidence of type I error, and the ambiguity allowed by them generates difficulties in the interpretation of results. Considering that the Scott-Knott test does not allow for a mean to belong to more than one group and it has higher rigor, which generates a lower incidence of type I error, it is the recommended test for the studies evaluated.
Keywords: *Lactuca sativa*, *Zea mays*, Statistical rigor, Type I error, Ambiguity.

TESTES ESTATÍSTICOS DE COMPARAÇÕES MÚLTIPLAS E AGLOMERATIVO UTILIZANDO O SOFTWARE RBIO PARA ALFACE E MILHO

RESUMO - O presente estudo teve como objetivo avaliar a eficiência e uniformidade dos testes de comparações múltiplas em relação ao teste aglomerativo aplicados no software RBio. Nas avaliações utilizaram-se dados de experimentos agrícolas conduzidos pelos próprios autores, no campo experimental da Universidade Federal de Goiás. Os dados analisados foram provenientes de três experimentos conduzidos com as culturas da alface e do milho: o primeiro foi disposto em delineamento inteiramente casualizado, o segundo em delineamento de blocos ao acaso e o terceiro em DBC, em esquema de parcelas subdivididas. Na avaliação dos dados (coletados nas culturas da alface e do milho) foi utilizado o software Rbio. Os dados foram submetidos à análise de variância pelo teste F a 5% de probabilidade. As médias foram comparadas por meio dos testes de comparações múltiplas (Teste Tukey, Duncan e Student-Newman-Keuls) e pelo teste aglomerativo (Scott-Knott). Com relação aos testes Tukey, Student-Newman-Keuls e Duncan, o menor rigor dos mesmos, aumenta a incidência de erro tipo I, e a ambiguidade permitida por eles gera dificuldades na interpretação dos resultados. Pelo o fato do teste de Scott-Knott não permitir que uma média pertença a mais de um grupo e em função do seu maior rigor, que gera menor incidência de erro tipo I, este é o teste mais indicado para os estudos avaliados.
Palavras-chave: *Lactuca sativa*, *Zea mays*, Rigor estatístico, Erro tipo I, Ambiguidade.

INTRODUCTION

In experimental researches, the formulated hypothesis is evaluated through the development of tests for the different treatments, and the data obtained are commonly subjected to analysis and statistical tests (CONAGIN; BARDIN, 2006). The F test is used to assess the hypothesis of equality of means of treatments; however, this test does not specifically indicate the differences (BANZATTO; KRONKA, 1989).

The application of regression equations is recommended for treatments whose levels are quantitative and the procedures usually used for those of qualitative type are multiple comparisons or clustering test (BORGES; FERREIRA, 2003). Tukey, Duncan, Dunnet, Skott-Knott, and Student-Newman-Keul (SKN) tests are among procedures that show which treatments are statistically different (SOUSA et al., 2012).

The Tukey test is a simple and commonly used procedure for comparing the contrast between two means of treatments with the same number of replications (OLIVEIRA, 2008). However, the Scott-Knott test is carried out for separating means of treatments in different groups, thus minimizing the existing variation within groups and maximizing the variation between groups (SCOTT; KNOTT, 1974). According to Gomes (2009), although the Duncan test is more laborious, it allows the

[1]Federal Institute Goiano (IF Goiano), *Campus* Rio Verde, Rio Verde, GO, Brazil. E-mail: aosmaniotto@gmail.com. *Corresponding author.
[2]Federal University of Jataí (UFJ), Jataí, GO, Brazil.
[3]Federal University of Goiás (UFG), *Campus* Samambaia, Goiânia, GO, Brazil.

obtaining of more detailed results and can be an alternative when the Tukey test does not enable to identify statistical differences. The Dunnett test is characterized by comparing treatments to a control treatment (BENZATO; KRONKA, 2006). Each test has its particularities and can be used for comparisons between pairs of treatments, between groups of treatments, and between a treatment and a control; the choice depends on the researcher, according to the hypothesis established (SILVA; AZEVEDO, 2016).

The ambiguity of results is a problem faced by researchers when working with statistical tests, as well as the control of type I and II errors (SILVA et al., 1999). According to Girardi et al. (2009), type I error is defined as the probability of rejecting a null hypothesis (H_0) when this is true and should be accepted, whereas type II error is characterized by the probability of accepting a null hypothesis when it is falsa and should be rejected.

Comparison tests may differ from each other regarding these errors. The Student-t and Duncan tests present high type I error rates and the Scheffé and Tukey tests have type I error rates lower than the significance level, whereas the SNK test is characterized by the control of this error (CARMER; SWANSON, 1973).

In this context, the objective of the present study was to evaluate the efficiency and uniformity of multiple comparison tests when compared to clustering test applied in the software RBio.

MATERIAL AND METHODS

Data from agricultural experiments conducted in the experimental field of the Federal University of Goiás (UFG), in Jataí, Jataí, GO, Brazil, were used in the evaluations.

The data analyzed were from three experiments: the first was conducted in completely randomized design (CRD), the second in a randomized block design (RBD), and the third in RBD with split-plot arrangement. The data used from the experiment conducted in CRD were from evaluations of 15 lettuce cultivars (treatments), namely: Gloriosa, Thainá, Angelina, Raider, Ludmila, Dora, Serena, Escarcha, Santa Celeste, Lucy Brown, Mauren, Bruma, Irene, Astra, and Aroeira (Table 1).

TABLE 1 - Data of fresh root weight (g plants^{-1}) of 15 lettuce cultivars from an experiment conducted in completely randomized design (CRD).

Treatments	Replications		
	1	2	3
Gloriosa	36.00	33.25	28.50
Thainá	38.50	33.75	33.75
Angelina	38.25	40.00	38.00
Raider	32.50	28.75	22.75
Ludmila	38.75	27.75	35.25
Dora	32.50	30.00	29.50
Serena	37.00	31.50	36.50
Escarcha	39.25	39.00	37.25
Santa Celeste	41.25	45.50	37.50
Lucy Brown	32.75	35.50	29.25
Mauren	34.00	37.00	32.75
Bruma	35.75	34.50	39.25
Irene	34.75	27.00	29.25
Astra	28.75	33.50	36.75
Aroeira	45.00	50.00	43.75

The data used from the experiment conducted in RBD were from evaluations of different forms for inoculation with *Azospirillum brasilense*, in which the response variable used was the gross protein contents in maize leaves (Table 2). Four blocks were used, testing the following treatments: without inoculation; inoculation with *A. brasilense* via seed; inoculation with *A. brasilense* via sowing furrow; one inoculation with *A. brasilense* via leaf; two inoculations with *A. brasilense* via leaf; and combination of the three methods (inoculation via seeds, sowing furrow, and via leaf.

The data used from the experiment conducted in RBD with split-plot arrangement were from evaluations of agronomic performance of second crop maize as a function of previous crops and inoculation with *A. brasilense*. The response variable used was one-thousand grain weight (Table 3). The sources of variation were: a) plots - previous crops (common bean and soybean without inoculation, and *Brachiaria* sp. with and without inoculation) sown in the summer crop season and b) subplots - second crop maize with and without inoculation.

TABLE 2 - Gross protein content (%) in second crop maize as a function of forms of inoculation with *Azospirillum brasilense* (randomized block design - RBD).

Forms of inoculation with Azospirillum brasilense	Blocks			
	1	2	3	4
Without inoculation	13.03	13.03	12.78	12.93
Inoculation via seed	13.31	13.07	12.64	12.84
Inoculation via sowing furrow	13.46	13.27	13.28	13.37
One inoculation via leaf	12.93	12.69	12.99	12.90
Two inoculations via leaf	13.27	13.32	13.01	12.93
Inoculation via seeds + furrow + leaf	12.62	13.08	12.89	13.04

TABLE 3 - Data of one-thousand grain weight (g) as a function of previous crops and inoculation of second crop maize with *Azospirillum brasilense* (randomized block design - RBD with split-plot arrangement).

Plots (crops)	Subplots (maize)	Blocks			
		1	2	3	4
Common bean	With inoculation	274.18	247.60	260.88	249.73
Common bean	Without inoculation	280.29	250.13	249.09	243.93
Soybean	With inoculation	282.17	256.66	251.22	264.25
Soybean	Without inoculation	246.58	244.57	246.49	244.21
Brachiaria sp. without inoculation	With inoculation	247.01	227.63	252.99	226.01
Brachiaria sp. without inoculation	Without inoculation	262.57	231.49	247.53	244.93
Brachiaria sp. with inoculation	With inoculation	264.43	251.58	247.87	237.70
Brachiaria sp. with inoculation	Without inoculation	244.63	234.27	260.68	232.31

The data were evaluated using the software Rbio (BHERING, 2017). The data were subjected to analysis of variance by the F test at 5% probability. The means were compared through multiple comparison (Tukey, Duncan, and SNK) and clustering (Scott-Knott) tests.

RESULTS AND DISCUSSION

The analysis of variance (F test) for the experiments in CRD, RBD, and RBD with split-plot arrangement are shown in Table 4. Significant differences were found for the experiments in CRD and RBD. The interaction between the sources of variation (crops and inoculation with *Azospirillum brasilense*) was significant in the experiment in RBD with split-plot arrangement. Four mean tests, Tukey, Student-Newman-Keuls (SNK), Duncan, and Scott-Knott, were applied for a better evaluation of differences between the treatments to compare the experiments in CRD, RBD, and RBD with split-plot arrangement.

TABLE 4 - Analysis of variance for data of experiments in completely randomized design (CRD), randomized block design (RBD), and RBD with split-plot arrangement.

Experiments	Sources of variation	Degrees of freedom	Mean squares
Completely randomized design (CRD)	Treatments	14	64.937^{**}
	Residue	30	11.439
Randomized block design (RBD)	Treatments	5	0.128^{*}
	Blocks	3	0.035^{ns}
	Residue	15	0.033
Randomized block design (RBD) with split-plot arrangement	Crops	3	1.022^{ns}
	Blocks	3	5.499^{**}
	Error (a)	9	0.656
	Inoculation	1	2.147^{ns}
	Crops*Inoculation	3	2.295^{*}
	Error (b)	12	0.631

** = significant at 1% probability of error ($p<0.01$); * = significant at 5% probability of error ($0.01 \leq p < 0.05$); ^{ns} = not significant ($p \geq 0.05$).

Table 5 shows the comparison of means of treatments from the experiment conducted in CRD, using the Tukey, SNK, Duncan, and Scott-Knott tests. The evaluation of the data showed that the tests did not show equal classifications, except for the treatment comprising the aroeira cultivar. The Tukey, SNK, and Duncan tests showed a more detailed classification than the Scott-Knott test. The treatments were separated into three groups by the Scott-Knott test; into four levels by the Tukey and SNK tests; and into five levels by the Duncan test. The application of the clustering test (Scott-Knott test), as expected, did not overlap the treatments, i.e., the same treatment was not classified into more than one group, since the groups are auto-excluding.

The multiple comparison tests (Tukey, SNK, and Duncan) were those that presented higher differentiation of the treatments, separating them into more levels than the clustering test (Scott-Knott). According to Sousa et al. (2012), the higher the distinction between treatments, the less rigorous is the test, and the higher the incidence of type I errors, i.e., differences between treatments are found when these differences do not exist.

The comparison of means of treatments from the experiment conducted in RBD using the Tukey, SNK, Duncan, and Scott-Knott tests are shown in Table 6. The SNK and Duncan tests showed equal classification; however, the Tukey test was less sensitive to differences, despite presenting the same number of levels than the others. It is shown by the comparison between treatments without inoculation, inoculation via seeds, and inoculation via sowing furrow, which did not present differences from each other when compared by the Tukey test. However, when using SNK and Duncan tests for the same comparison, the treatments without inoculation and inoculation via seeds were different from the treatment with inoculation via sowing furrow. The Scott-Knott test separated the treatments into two different groups, partially following the same pattern of the multiple comparison tests, which presented two levels.

TABLE 5 - Comparison of lettuce cultivars (treatments) in an experiment in completely randomized design (CRD) through four mean tests at 5% probability of error.

Cultivars of lettuce	Means	Tests			
		Tukey	SNK	Duncan	Scott-Knott
Aroeira	46.25	a	a	a	a
Santa Celeste	41.42	ab	ab	ab	b
Angelina	38.75	abc	bc	bc	b
Escarcha	38.50	abc	bc	bc	b
Bruma	36.50	abcd	bcd	bcd	c
Thainá	35.33	bcd	bcd	bcd	c
Serena	35.00	bcd	bcd	cd	c
Mauren	34.58	bcd	bcd	cd	c
Ludmila	33.92	bcd	bcd	cde	c
Astra	33.00	bcd	bcd	cde	c
Gloriosa	32.58	bcd	bcd	cde	c
Lucy Brown	32.50	bcd	bcd	cde	c
Dora	30.67	cd	cd	de	c
Irene	30.33	cd	cd	de	c
Raider	28.00	d	d	e	c

Means followed by the same letter in the columns are not significantly different from each other.

TABLE 6 - Comparison of treatments of the experiment (previous crops and inoculation of second crop maize with *Azospirillum brasilense*) in randomized block design (RBD) through four mean tests at 5% probability of error.

Forms of inoculation of AB	Means	Test			
		Tukey	SNK	Duncan	Scott-Knott
Inoculation via sowing furrow	13.35	a	a	a	a
Two inoculations via leaf	13.14	ab	ab	ab	a
Inoculation via seeds	12.97	ab	b	b	b
Without inoculation	12.95	ab	b	b	b
Inoculation via seeds + furrow + leaf	12.91	b	b	b	b
One inoculation leaf	12.88	b	b	b	b

Means followed by the same letter in the columns are not significantly different from each other.

The comparison of means of treatments in RBD with split-plot arrangement, using the tests Tukey, SNK, Duncan, and Scott-Knott test are shown in Table 7. Considering the statistical breakdown of the sources of variation (inoculation within previous crops), the classification was similar for all tests applied. Considering the previous crops within the inoculation, the dynamics were equal for all tests, except for the Scott-Knott test. The Tukey, SNK, and Duncan tests detected differences between common bean and soybean, as previous crops, within treatments without inoculation, whereas the Scott-Knott test ranked the treatments in only one group. The power of the Scott-Knott test to detect differences decreased, when compared to the Tukey, SNK, and Duncan, as the magnitude between means decreased, showing that it is a more rigorous test, with lower incidence of type I error.

Caierão (2006) evaluated the application of tests for comparison of means in barley, using 258 scientific articles published in the annals of the Brazilian Barley Research Committee meetings from 2001 to 2005; they found that the Tukey test was the most common, found in 103 articles, representing more than 80% of the articles.

The Duncan test was applied in approximately 15% articles, followed by the Scott-Knott test (2.3%), t test (2.3%), and those that were not possible to identify (approximately 1%). Despite the Tukey test is less rigorous and have higher incidence of type I error compared to the Scott-Knott test, Caierão (2006) report that its high use may indicate the rigor level desired by researchers at the time of choice of procedures or lack of familiarity with other test types.

TABLE 7 - Comparison of treatments of the experiment in randomized block design (RBD) with split-plot arrangement through four mean tests at 5% probability of error.

Plots (crops)	Subplots (Inoculation)									
	Means		Tukey		SNK		Duncan		Scott-Knott	
	With	Without	With	Without	With	Without	With	Without	With	Without
Soybean	25.75*	26.18	aA	aA	aA	aA	aA	aA	aA	aA
Brachiaria sp. without inoculation	25.51	25.29	aA	abA	aA	abA	aA	abA	aA	aA
Brachiaria sp. with inoculation	25.21	25.00	aA	abA	aA	abA	aA	abA	aA	aA
Common bean	26.46	24.40	aA	bB	aA	bB	aA	bB	aA	aB

Means followed by the same lowercase letter in the columns, or uppercase letter in the rows, are not significantly different from each other.

Silva and Azevedo (2016) evaluated the comparison of means of agricultural experiment data through different tests using the software Assistat and reported that experiments with number of treatments lower than or equal to 8 and well-defined differences between them tend to present the same results when using the SNK, Scott-Knott, and Duncan tests; and that the Tukey test tend to show results that partially agree with the other three, but with a more detailed classification.

According to Borges and Ferreira (2003), the main procedures of multiple comparisons used are: Tukey, SNK, and Duncan tests. However, these tests present difficulties in the interpretation of their results. All of them present characteristics of ambiguity in the results. This ambiguity is due to the possibility that two levels of treatments are considered equal to a third, but different from each other, making it difficult the interpretation of results and the determination of the best treatment.

CONCLUSIONS

The Tukey, Student-Newman-Keuls, and Duncan tests showed the lowest rigor, which increases the incidence of type I error, and the ambiguity presented by them generates difficulties in the interpretation of results.

The Scott-Knott test was the most appropriate for the studies evaluated, as it does not allow for a mean to belong to more than one group, and generates a lower incidence of type I error due to its higher rigor.

REFERENCES

BANZATTO, D.A.; KRONKA, S.N. **Experimentação Agrícola.** 4a. ed. Jaboticabal: FUNEP editora. 2006, 250p.

BANZATTO, D.A.; KRONKA. S.N. **Experimentação Agrícola.** 1ª. ed. Jaboticabal: FUNEP editora. 1989. 247p.

BHERING, L.L. Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v.17, n.2, p.187-190, 2017.

BORGES, L.C.; FERREIRA, D.F. Poder e taxas de erro tipo i dos testes scott-knott, tukey e student-newman- keuls sob distribuições normal e não normais dos resíduos. **Revista Matemática e Estatística**, v.21, n.1, p.67-83, 2003.

CAIERÃO, E. Aplicação dos testes de comparação de médias em ensaios de cevada. **Pesquisa Agropecuária Gaúcha**, v.12, n.1-2, p.51-55, 2006.

CARMER, S.G.; SWANSON, M.R. An evaluation of ten pairwise multiple comparison procedures by Monte Carlo Methods. **Jornal of the American Statistical Association**, v.68, n.341, p.66-74, 1973.

CONAGIN, A.; BARDIN, D. Poder e eficiência dos diferentes testes estatísticos para comparações multiplas. **Revista de Agricultura**, v.81, n.1, p.118-137, 2006.

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

GOMES, F.P. **Curso de estatística Experimental.** 15a. ed. Piracicaba: FEALQ editora. 2009. 451p.

OLIVEIRA, A.F.G. Testes estatísticos para comparação de médias. **Revista Eletrônica Nutritime**, v.5, n.6, p.777-788, 2008.

SCOTT, A.J.; KNOTT, M. A. Cluster analysis method for grouping means in the analysis of variance. **Biometrics**, v.30, n.3, p.507-512, 1974.

SILVA, E.C.; FERREIRA, D.F.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de scott-knott por meio do Método de Monte Carlo. **Ciência e Agrotecnologia**, v.23, n.3, p.687-696, 1999.

SILVA, F.A.S.E.; AZEVEDO, C.A.V. Comparison of means of agricultural experimentation data through different tests using the software Assistat. **African Journal Agricultural Research**, v.11, n.37, p.3527-3531, 2016.

SOUSA, C.A.; LIRA JUNIOR, M.A.; FERREIRA, R.L.C. Avaliação de testes estatísticos de comparações múltiplas de médias. **Revista Ceres**, v.5, n.3, p.350-354, 2012.