International Journal of Multidisciplinary Research and Development

IJMRD 2015; 2(1): 197-200
www.allsubjectjournal.com
Received: 10-12-2014
Accepted: 30-12-2014
e-ISSN: 2349-4182
p-ISSN: 2349-5979
Impact factor: 3.762

Siraj Osman Omer
Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan.

Shareif Hassan Hussin
Geneina research station, ARC, Sudan.

Correspondence:
Siraj Osman Omer
Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan.

An assessment of multiple comparisons for single variety trial in block designs

Siraj Osman Omer and Shareif Hassan Hussin

Abstract
Most of the agricultural researchers interested in statistical inferences that require multiple comparisons of means. The paper addresses the multiple comparisons procedures in crop variety trial, which is where multiple comparisons are a particular used in the Genstat statistical software. For multiple comparative assessment, grain yield (kg/ha) of millet genotypes was used. The results highlighted that Fisher's protected least significant difference (LSD) and Ryan-Einot-Gabriel Walsh were an appropriate procedure for multiple comparisons of means. Multiple comparison procedures in agricultural experiments are very useful for making the best decisions regarding to comparisons of treatment means in randomized complete block designs.

Keywords: Block design, multiple comparisons, Genstat software

1. Introduction
In the usual plant breeding program, standard varieties (or check cultivars) are included along with the new entries to provide a basis against which to measure progress in the breeding program (Finney, 1987). Thus, the comparison between the two varieties is on the basis of the gain, not the difference. Thus the breeder is normally not interested in significant differences between the two varieties, but he would like to be able to reject the null hypothesis $H_0{:}\mu_N{=}\mu_C$ in favor of alternative hypothesis $H_a{:}\mu_N{>}\mu_C$, where μ_N is the true mean of the new entry and μ_C is the true mean of the check or a control variety (Cohen and Sackrowitz, 2005). Analysis of variance (ANOVA) is most popular method for testing the equality of several means (Saville, 1990). It is probably the most useful technique in the field of statistical inferences (Gelman *et al.*, 2012). In general, many statisticians worries of making large numbers of correlated and unplanned pairwise comparisons between the means of the treatments included in an experiment (Saville, 2003). Multiple comparisons can be done using the chosen test, although it may be convenient (both practically and statistically) for researchers to assume that their samples are obtained from normally distributed populations, this assumption may rarely be accurate (Maxwell and Delaney, 1990). A number of authors (Camer and Walker, 1982; Plummer, 1982; Cribbie, 2003) have pointed out that multiple comparison procedures are frequently used in situations for when there are unstructured treatments. In the context of multiple comparison of means, the overall probability of falsely declaring some (at least one) pairs of means different, when in fact they are equal (i.e. Type –I error), is substantially larger than the specified alpha value (Bendera and Langeb, 2001). The appropriate procedure for testing the equality of several means is the analysis of variance (ANOVA). However, ANOVA has a wider application than the multiple comparison test (Perry, 1986). It is probably the most useful technique in the field of statistical inference (Day and Quinn, 1998). In this paper an example will used be used to illustrate to compare a few numbers in the crop verity trial. Applications of multiple comparison procedures are discussed in this study.

2. Multiple Comparison Test in GenStat
GenStat is the most popular software considered for multiple. GenStat software has been compared for availability of type of multiple comparison methods, and various components of statistical information expected in those methods they produce when analyzing data (Payne, 2013).

All pairwise procedure of GenStat performs a range of multiple comparison tests (Bechhofer *et al.* 1995). The means are inputted using the means parameter either a *variant* or in a one-way table. The replication (or number of observations for each mean) is supplied by the replications parameter, either in a scalar (if all the replications are equal) or in a structure of the same type as the means. The type of test to be performed is specified by the *method* option, with settings Tukey, SNK (Student-Newman-Keuls), REGWMR (Ryan/Einot-Gabriel/Welsch) multiple range test), Duncan, Scheffe, Fisher's Protected Least Significant Difference (FPLSD), Fisher's Unprotected Least Significant Difference (FULSD), Bonferroni and Sidak. The *direction* option allows the means to be arranged in ascending or descending order. The probability option allows the experiment wise significance level for the intervals to be changed from the default 0.05 to 0.01 (for further information see Hsu, 1996).

3. Methodology

The data used in the current paper are from randomized complete block design (RCBD). The experiment was used to investigate nine genotypes of millet with four replications established in 2010; the experiment was located at the Nyala Research Station, Agricultural Research Corporation (ARC), Sudan. Grain yield data was used and analysis of variance (ANOVA) was performed to determine the effect of genotypes.

4. Results and Discussion

Following Table 2 and Table 3 show the critical values information on the pairwise multiple comparisons (i.e. for the Scheffe, Fisher's protected least significant difference, Fisher's unprotected least significant difference, Bonferroni and Sidak) and multiple test range, i.e. for the Tukey's , (Student - Newman - Keuls, Ryan - Einot – Gabriel - Welch, Duncan). Results of the least significant difference and maximum significant difference tests are summarized below for ANOVA with different level k compares-wise and critical value at alpha=0.05 for various multiple comparison tests.

Table 1: Analysis of variance of grain yield for pearl milt genotypes.

Source of variation	df	ss	ms	vr	Pr>
Rep	3	3547344	1182448	14.83	
Genotypes	8	29876400	3734550	46.84	<.001
Residual	24	1913356	79723		
Total	35	35337100			

Where df =Degree of Freedom, ss= Sum of Square, ms= Mean Square, vr = (variance ratio) instead of F-value, and Pr> = (F probability) instead of P-value.

The ANOVA for grain yield is presented. The overall F-test is significant (F=46.84 p<0.001), indicating that there was significant differences among the genotypes at 0.05 level of significant, the standard errors of differences of means were 199.7, with coefficients of variation was 11.6. Thus; a multiple comparison procedure on means will be used to examine the differences among the genotypes.

Table 2: Critical values of Scheffe, FPLSD, FUPLSD, Bonferroni, Sidak of all pairwise multiple comparison tests at significant levels 5%, 1% and 10%.

Significant level (alpha)	Scheffe	FPLSD	FUPLSD	Bonferroni	Sidak
5.00%	4.341	2.064	2.064	3.614	3.604
1%	5.187	2.797	2.797	4.253	4.251
0.10%	6.319	3.745	3.745	5.159	5.159

Where FPLSD = Fisher's protected least significant difference test, FUPLSD =Fisher's unprotected least significant difference test

Table 2 indicates that the Fisher's protected and unprotected LSD method controls the Type I comparison-wise error rate, but not the experiment-wise error. For example, the critical value at 5% for the difference of genotype means is 2.06 for least significant difference (LSD) and 4.34 for Scheffe, which uses the minimum significant difference. It varies from LSD, Bonferroni and Sidak. The critical values in Bonferroni and Sidak, procedures are the same at 5% (3.6) or at 10% (5.2). Generally, Scheffe procedures for multiple comparisons of means have a higher Type II error rate than LSD for all pairwise comparisons with critical value using Minimum Significant Difference.

Table 3: Critical values at alpha=0.05 of multiple range tests for Tukey's, SNK, REGWQ, Duncan's for comparison-wise error rates.

Range	Tukey's		SNK		REGWQ		Duncan's	
	Prob	C	Prob	C	Prob	C	Prob	C
2	0.95	2.064	0.99	3.879	0.95	2.064	0.95	2.064
3	0.95	2.497	0.98	4.225	0.90	2.168	0.95	2.497
4	0.95	2.759	0.98	4.412	0.86	2.234	0.95	2.759
5	0.95	2.946	0.97	4.537	0.81	2.281	0.95	2.946
6	0.95	3.092	0.97	4.629	0.77	2.317	0.95	3.092
7	0.95	3.211	0.96	4.701	0.74	2.344	0.95	3.211
8	0.95	3.312	0.95	4.684	0.70	2.365	0.95	3.312
9	0.95	3.399	0.95	4.807	0.66	2.383	0.95	3.399

Prob= probability C=Critical values

Table 3 shows multiple range tests for Tukey's, SNK, REGWQ, Duncan's for comparison-wise error rates. The critical value of Scheffe, FPLSD, FUPLSD, Bonferroni and Sidak were 4.341, 2.064, 2.064, 3.614 and 3.604 respectively. The critical value of Tukey's, REGWQ and Duncan multiple ranges for group of 2 means are similar for FPLSDand FUPLSD, but the critical value of SNK for a group of 2 means is 3.879 which are approximately equal to the Scheffe method (4.34). The critical values for Bonferroni and Sidak test were 3.614 and 3.604 which are approximately equal (Table 1). The probability of critical values for Tukey's, SNK and Duncan's are same at 0.95 conference interval at 0.05 level of significance, while the probability of REGWQ test is different to each other. The difference between the maximum and minimum value for Tukey's, SNK, REGWQ and Duncan procedures are 1.34, 0.93, 0.32 and 1.34 respectively. This indicates that, the Duncan procedure is similar to Tukey's. The REGWQ difference is less than the difference for each other. We observed that REGWQ test controls experiment-wise error rate for multiple range test.

Table 4: Genotype mean of Scheffe FPLSD, FUPLSD, Bonferroni, Sidak of pairwise multiple comparison tests at significant levels 5% denoted by letters A to E.

Genotypes	Mean	Scheffe	Fplsd	Fuplsd	Bonferroni	Sidak
1	3870	A	A	A	A	A
7	3792	A	A	A	A	A
4	2672	B	B	B	B	B
8	2620	BC	B	B	B	B
2	2615	BC	B	B	B	B
5	2055	BCD	C	C	C	BC
3	1770	CDE	C	C	C	CD
9	1358	DE	D	D	D	CD
6	1132	E	D	D	D	D

Where FPLSD = Fisher's protected least significant difference test, FUPLSD =Fisher's unprotected least significant difference test.

Table 5: Genotype mean of Turkey's, SNK, REGWG and Duncan's of multiple comparison range tests at significant levels 5% denoted by letters A to E.

Genotypes	means	Tukey's	SNK	REGWG	Duncan's
1	3870	A	A	A	A
7	3792	A	A	A	A
4	2672	B	B	B	B
8	2620	B	B	B	B
2	2615	B	B	BC	B
5	2055	BC	BC	BD	C
3	1770	CD	CD	DE	C
9	1358	D	D	EF	D
6	1132	D	D	F	D

Where Tukey's= Tukey's 95% confidence intervals, SNK=Student-Newman-Keuls test, REGWQ= Ryan/Einot-Gabriel/Welsch multiple range test, Duncan's =Duncan's multiple range test.

The number of significant comparisons between genotype pairs where a genotype is under the same letter from A to G in is shown in Table 2 and Table 3 for each of the procedures, all methods labeled A, B, C, D, E and F giving 8 homogeneous groups, can be presented in the order A > B> C> D> E> F> G is using the maximums of genotypes means in the group generated from the letters. Genotype1 and Genotype 7 (mean= 3870 and 3792) is shown by the letter A for all the methods, Genotype4, Genotype8, Genotype 2 and Genotype5 (mean =3792, 2672, 2620 and 2615) with the letter B for all methods, expect in from FPLSD, FUPLSD, Bonferroni and Duncan's, where letter C is taken. Genotype3 (1770) is shown letter D for Scheffe, Sidak, Tukey's, SNK and REGWQ methods, in from expected FPLSD, FUPLSD and Bonferroni Duncan and SNK procedures, are shown letter (C). G8 (1358) is shown letter D for all the methods, expected Scheffe, Sidak and REGWG procedures are shown letter (DE, CD and EF) respectively. G6 (mean =1132), the lowest yielding genotype, has been found to belong to one of the three genotypes shown by letters (D) for Sidak, Tukey's and SNK methods. While Scheffe and REGWQ shown letter E and F. From table (3) above it are clear that the FPLSD, FUPLSD, Bonferroni and Duncan's gave the same the results, also Tukey's and SNK given the similar results, while Scheffe procedures are different compared to each other procedures. The above results indicated there were differences among the nine multiple comparison procedures.

This study widely investigated appropriate and applicable single-step and step-wise multiple comparison procedures. All necessary tests to perform the comparison of genotypes means were conducted and total numbers of significant results for each test were observed. Accordingly, could be concluded that, least significant difference (LSD) was very simple for comparing treatment means for all pairwise comparisons, while REGWQ method was appropriate test for multiple range test than Tukey's, SNK, and Duncan method treatments. Acceding to Saville, 2003 and Omer and Murari 2009, an appropriate multiple comparison could be only when there no prior hypotheses, and for comparing all treatments with all other treatments.

5. Conclusion

A Genstat statistical package is easy to use, where only a few instructions are needed to do a standard analysis. GenStat program is very useful for all types of agricultural experiments. Using Genstat statistical packages can offer multiple comparisons of means as an option under ANOVA. FPLSD, FUPLSD, Bonferroni and Duncan's methods were very close in detecting the significance on pairs of means.

We highlighted that FPLSD and REGWQ results were appropriate procedures. Depend on the critical values for the t-statistic for pairwise multiple tests (i.e. for the Bonferroni and Sidak methods, as well as the FPLSD and FUPLSD methods provided all the comparisons have the same number of residual degrees of freedom. The choice of the best method to carry out MCPs is determined by the nature of the number of treatments.

6. References

1. Bendera R, Langeb S. Adjusting for multiple testing-when and how. Journal of Clinical Epidemiology 2001; 54:343–349.
2. Carmer SG, Walker WM. Baby bear‟s dilemma: a statistical tale. Agronomy Journal 1982; 74:122-124.
3. Cohen A, Sackrowitz HB. Decision theory results for one-sided multiple comparison procedures. The Annals of Statistics 2005; 33(1):126-144.
4. Cribbie RA. Pairwise multiple comparisons new yardstick, new results. The journal of experimental education 2003; 71(3):251-265.
5. Day RW, Quinn GP. Comparisons of treatments after an analysis of variance in ecology. Ecological Monographs 1998; 51(4):433-463.
6. Gelman A, Hill J, Yajima M. Why we (Usually) Don't Have to Worry about Multiple Comparisons, Journal of Research on Educational Effectiveness 2012; 5:189–211.
7. Finney H. Statistical method in biological assay. Charles Griffin and Cy, 1987.
8. Hsu JC. Multiple Comparisons Theory and Methods. Chapman & Hall, London, 1996.
9. Omer SO, Singh M. Appropriate Multiple Comparisons of Means in Statistical Analysis. Sudan Journal of agricultural research, 2012, 19.
10. Maxwell SE, Delaney HD. Designing experiments and analyzing data: A model comparison perspective. Belmont, CA: Wadsworth, 1990.
11. Payne R. A Guide to Anova and Design in GenStat® (16th Edition). GenStat Reference Manual (Release 16), VSN International, Hemel, Hempstead, UK, 2013.
12. Perry JN. Multiple-comparison procedures: a dissenting view. J. Econ. Entomol 1986; 79(11):49-1155.
13. Plummer JL. Multiple comparison procedures in the analysis of designed experiments. Anesthesiology 1985, 62(1):97-98.
14. Saville DJ. Multiple comparison procedures: the practical solution. The American Statistician, 1990, 44(2):174-180.
15. Saville DJ. Basic Statistics and the Inconsistency of Multiple Comparison Procedures, AgResearch, New Zealand. Canadian Journal of Experimental Psychology 2003; 57(3):167-175.
16. Wilcox RR. Three multiple comparison procedures for trimmed means, 1997.