Use and Misuse of multiple Comparisons Procedures of Means in Factorials Experiments

Siraj Osman Omer[1] & AbdelGadir Mohamed Abdellah[2]

1. Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan, E-mail: sirajstat@yahoo.com
2. National Insect Collection, Crop Protection Research Center, ARC, Wad Medani,

Abstract

Multiple comparison procedures of means are frequently misused and such misuse may result in incorrect scientific conclusions. The objectives of this study was to identify the most common errors made in the use of multiple comparison procedures on means in factorial experiments and present correct method. The results highlighted that only 20% could be considered to use pair-wise test and multiple comparison test (MCT) completely correct. A planned contrast was also found misused in comparison of levels of a quantitative factor and comparison of treatment means. In some cases, totally incorrect Duncan multiple range test were made. In conclusion, factorial arrangement is needed but with due statistical reasoning for evaluating appropriate multiple comparison procedures in factorial experiments for qualitative and quantitative levels and in that way to appraise the right statistical differences.

Key words: Factorial experiments, multiple comparisons, mean comparisons

1. Introduction

The goal of agricultural experiments generally is to detect meaningful relationships among treatments and associated responses (Chew, 1980). Types of comparisons of means include pair wise multiple comparisons, planned orthogonal or no orthogonal contrasts, and orthogonal polynomials (Petersen, 1977). Some procedures are appropriate only for specific types of treatment designs and specific types of objectives (Nelson et al., 1983). Pair-wise multiple comparison are appropriate only for

2. Comparison of Factorial Treatment Means

A simple factorial experiment has two factors to be evaluated in an experimental design. Examples include two-factor factorial combinations in a randomized complete blocks, a split-plot experiment in complete blocks, or a strip-plot experiment in complete blocks (Hinkelmann and Kempthorne, 2008). In the analysis of data from a factorial experiment, one normally tests significance of main effects of each factor and the interaction between them, and estimates the effects with associated standard errors (Saville, 2014). For the designs, with both factors assumed to have fixed effects and qualitative in nature (e.g. varieties of a crop, or a set of cropping systems), this study will make appropriate multiple comparisons for each factor.

In majority of agricultural experiments, Finney (1987) noted inappropriateness of multiple comparisons because of their symmetric relation to all comparison for example

- In factorial experiment, main effect and various types of comparing unstructured qualitative treatments (Lowry, 1992).

Always, pair-wise multiple comparison tests should be used only when the treatment structure is not well understood (Carmer, 1984). Plan and design the experiments with structured treatments or factorial sets of treatments, so that the analysis assesses how factors jointly and independently affect response (Day and Quinn, 1989). Multiple comparison tests has many limitations, when little information exists on the structure of the treatments and that orthogonal contrasts must be used when the treatments have a logical structure among treatments application to factorials (Atil and Unver, 2001). Orthogonal polynomial procedures assess relationships between quantitative treatments and response when a full range of responses or an optimal dose is of interest (Little, 1978). The goal of this study is to identify the most common miss-use of multiple comparison procedures (MCP) on means in factorial experiments and to present the correct methods.

comparisons. Scheffe method was most powerful and suitable method, if number of comparison is large relative to number of treatment means and Dennett test is recommended for comparison treatments with control. Some researchers prefer using Duncan test for large numbers of treatment, while REWGQ test is the most appropriate methods for multiple comparisons.

When treatment were factorial, the regression between dependent and independent variables was more appropriate than multiple comparison tests. Sudan Journal of Agricultural Research (SJAR), showed that, out of 15% of articles that used multiple comparison published in volumes 23 and 54 (2005-2010), only 20% of papers correctly used multiple comparisons and while 80% incorrectly used multiple comparisons. Most of the researchers use LSD and Duncan multiple range tests for means comparison in their analysis, among other procedures. For scientific published papers and postgraduate studies, LSD is commonly used method in comparing means (Siraj and

interaction are not of equal interest, but they constitute a sensible way of examining results, e.g., comparing levels of one factor for a given level of the other in case of substantial/significant interaction.

- If treatments are several levels of the same applied material (e.g. amount of irrigation water, concentration of a chemical or several duration of exposure to a treatment) the interest would be to study any trend along the factor level and not to individual differences (Westfall, 1999).

3. Appropriate and Inappropriate Uses of Mean Comparison Procedures

Siraj and Singh (2012) presented comparisons of means and their interpretations. Accordinraly, the least significant difference (LSD) was very simple for comparing mean, but it is not recommended for all pair-wise comparisons due to increase in false positive. Tukey method was the more appropriate than LSD test for all pair

split plot design and factorial design (combination factors).

5. Results and Discussion

Among two-hundred articles, published between the years of 2005 and 2010, corresponding to volumes 23 to 54 of the Sudan Journal of Agricultural Research (SJAR), 70% of papers were based on use of randomized complete block designs while 30% were based on factorial designs which include a split plot designs and full factorial combination. Cardellino and Siewerdt (1995) critically reviewed evaluation of the use of tests for comparison of treatment means were tabulated into one of three categories: correct, partially correct and incorrect.

Singh, 2009). The criteria used to consider a correct use of the test were decided according to: (a) the objective of the research, (b) the structure of the treatments used (which proved in many occasions to be inconsistent with the objectives of the research) and (c) if the test fits the information in (a) and (b). The statistical merit of the test was not taken into account.

4. Methodology

To evaluate the quantity and quality of the use of multiple comparisons in Sudan Journal of Agricultural Research (SJAR) from 2005 to 2010 of volume number 23 up to volume number 54. The use and mis-use of appropriate procedures of comparison mean has been investigated based on factorial experimental evaluated in

Table1. Summary of use and misuse of multiple comparisons in factorial designs.

No	Items	MCP	Use of MCP	Type of Deign	Other	Standard error
1	Misuse	---	---	Split-plot	Interaction effect	----
2	Misuse	---	---	Split-plot	Interaction effect	SE±
3	Use	Duncan	Incorrect	Split-plot	----	----
4	Misuse	Duncan	Incorrect	Split-plot	----	----
5	Use	LSD	Correct	Full factorial	----	----
6	Use	Duncan	Incorrect	Full factorial	----	----
7	Use	Duncan	Incorrect	Full factorial	----	----
8	Misuse	---	---	Spilt-plot	Interaction effect	SE±
9	Use	Duncan	Incorrect	Full factorial	----	----
10	Use	Duncan	Incorrect	Split-plot	Interaction effect	SE±

11	Use	Duncan	Incorrect	Split-plot	----	----
12	Use	Duncan	Incorrect	Split-plot	----	-----
13	Misuse	----	----	Split-plot	Interaction effect	SE±
14	Use	Duncan	Incorrect	Split-plot	----	----
15	Use	LSD	Correct	Full factorial	----	----

Where LSD: Least significant difference, Duncan: Dunant multiple rang test and MCP: Multiple comparison procedures.

and other cases never used any kind of multiple comparisons. 18% of researchers reported standard error of mean in case of factorial design into interaction table of means, which have provided definitive, useful suggestions on how to integrate statistical methods including appropriate use of multiple comparisons into scientific publications.

In this study two kinds of experimental design were used; split-plot design and factorial (combination) by 67% and 33% respectively. Among the 150 subjects only in 15 cases multiple comparisons were used, 2 cases applied pair wise multiple comparisons by using LSD for quantitative data, 9 cases applied multiple range test by using Duncan in case of combinations in factorial arrangement of treatments,

Table 2. Classification % of LSD and Duncan use for comparison of treatment means.

Test	Correct use	Misuse
LSD	100	0
Duncan	30	70

those papers that didn't fit into either of the previous cases with respect to the choice of multiple comparisons in factorial arrangements where all treatment means are compared pairwise in situations where, fitting a regression equation, a pairwise comparison test is additionally used.

The most common used methods were LSD and Duncan multiple rang test, in

A survey in recent issues of SJAR showed an abuse of pair-wise comparison methods by using Duncan's tests in such situations where their uses are not adequate, when other tests would provide better quality information. The table shows that LSD totally correct use while Duncan was incorrect; partially correct is amount to be at 10% and 20% respectively. The partially correct category included

(Brow and Feng, 1999). If the factor is quantitative, an alternative method is to use the regression analysis, to compare the different levels of the quantitative factor with a control level. A better approach is to fit a regression equation for quantitative factor in addition to perform multiple comparison of qualitative treatment. If the factor is quantitative and the use of a multiple comparison test such as Duncan's multiple-range test is in appropriate. The correct procedure is to adjust a regression equation that is biologically the interpretable. The relationships among qualitative and quantitative factors comprise a pair-wise comparison and contrasts comparison (or use of regression analysis).

7. Conclusion

The results of this study shows that the use and the misuse of multiple comparisons in experimental design especially in factorial designs are complicated issues due to nature of the factors. Conducting field trials with different factors include crop varieties, a number of treatments systems of land and water management, fertilizers etc. the Sudan Journal of Agricultural Research (SJAR) as proper ways to interpret the results of agricultural experiments. More details on these tests, their uses and limitations, can be found in Steel and Torrie (1980) and the articles, Perecin and Malheiros (1989), and Siraj and Murari (2012).

6. Factorial Experiments Consideration

In factorial experiments, the response is obtained at combinations of levels of the different factors, every observation provides information on all the factors includes in the experiments (Jain et al., 2001). When treatments are the levels of qualitative factors the most appropriate is use to pair-wise comparison. When treatments are levels of a quantitative factor, the most appropriate is to partition the degrees of freedom for treatment into polynomial components or, if relevant, to fit a suitable regression model like the logistic curve, Gompertz curve or Mitscherlish's Law. In specific cases, where the goal is to compare many levels of the quantitative factor with a control, Williams' test should be used

Analysis of Variance in Ecology Source". Ecological Monographs, 59(4):433-
463.
[4] Carmer, S.G and Walker, W.M. (1985). "Pairwise multiple comparisons of
treatment means in agronomic research". J. Agron. Educ. 14:19–26.
[5] Chew, V. 1976. Comparing treatment means: A compendium. Hort-Science
11:348– 357.
[6] Chew, V. 1980. Testing differences among means: Correct interpretation and
somealternatives. HortScience 15:467–470.
[7] Cardellino, R.A. and F, Siewerd. (1992). Use and misuse of statistical tests for
Comparison of means, Brazil, Journal of the Brazilian Society of Animal Science
21:985-995.
[8] Hinkelmann, K and Kempthorne, O. (2008)." Design and Analysis of
Experiments". Volume 1, Introduction to Experimental Design, 2nd Edition
ISBN: 978-0-471-72756-9.
[9] Finney, H. (1987). "Statistical method in biological assay".Charles Griffin and Cy.
[10] Jain, R.C and Srivastava, R. (2001). "Factorial experiments some variation".
IASR Library Avenue, PUSA, New Delhi - 110 012 (INDIA).
[11] Little, T.M. (1978). If Galileo published in HortScience.HortScience13:504–506.
[12] Lowry, S. R.(1992). "Use and Misuse of Multiple Comparisons in Animal".
J. Anim. Sci.70 (6):1971-7.

The cause of misuse may be the lack of knowledge of alternative procedures to the pair-wise comparison; this may lead to inability to interpret the results in the right way. More statistical investigations are needed with practical suggestions to determine the appropriate procedures of means comparison in factorial experiments.

Acknowledgments

The author would like to thank Dr. Murari Singh, senior biometrician, international center for agricultural research in the dry area (ICARDA) Amman, Jordan for his guidance, insightful and valuable recommendations.

References

[1] Atil, H and Unver, Y. (2001)." Multiple Comparison". Asian Network for
scientific Information". 1(8)723-727.
[2] Brown, M.B and Feng, S Feng (1999). Williams' test comparing multiple dose
groupsto a control. Bulletin of the International Statistical Institute, Contributed
Papers,Tome 58, Book 1, pp. 129-130.
[3] Day, R. W and Quinn, G. P. (1989). "Comparisons of Treatments after an

[13] Nelson, L.A., and Rawlings, J.O. (1983). "Ten common misuses of statistics in agronomic research and reporting". J. Agron. Educ. 12:100–105.

[14] Pearce, S. C. (2005). "Data Analysis in Agricultural Experimentation. III. Multiple Comparison". Applied Statistics Research Unit, University of Kent at Canterbury,England

[15] Petersen, R.G. (1977). "Use and misuse of multiple comparison procedures". Agron. J.69(2):205–208.

[16] Saville, D.J. (2014). "Multiple Comparison Procedures—Cutting the Gordian Knot, doi:10.2134/agronj2012.0394.

[17] Siraj, O. O and Singh, M, (2012), "Appropriate Multiple Comparisons of Means in Statistical Analysis". SJAR 19 .

[18] Steel, R. G. D. and J. H. Torrie. (1997), "Principles and Procedures of Statistics a Biometrical Approach 3rd edition". McGraw-Hill. Co.Ini., New York.

[19] Westfall, P.H. (1999). Multiple Comparisons and Multiple Tests using the SAS® System, Book by users, SAS Institute.