

## Statistical Comparison using Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Test in a Study of Students Self-Concept and Mathematics Achievement in South-South, Nigeria

<sup>1</sup>Eyong Emmanuel Ikpi, <sup>2</sup>Umoh Augustus Jonny & <sup>3</sup>B. A. Bassey

<sup>1&3</sup>Department of Educational Foundations, University of Calabar

<sup>2</sup>Department of Curriculum and Teaching, Cross River University of Technology

### Abstract

The student was concerned on statistical comparison using turkey's HSD, Scheffe's, Fisher's LSD and John Sidak test in a study of students self-concept and Mathematics achievement in south-south, Nigeria. Literature for the study was reviewed empirical, conceptual and theoretical in line with the sub-variables of the study. The study utilized one research questions. The Ex-post facto research design was adopted for the study. The population of the study comprised of 23,109 respondents' drawn from 269 public secondary school in south-south, Nigeria using stratified and simple random sampling technique, a sample size of 866 respondents' were used in the study. The instrument for data collection was a 29 item structured Mathematics Performance Test Questionnaire (MPTQ) constructed by the researcher. The reliability was established with Cronbach alpha reliability estimate and the index was found to be .97 which implies that the instrument was highly reliable. Data collected were summarized and analyzed at .05 level of significance using One-way Analysis of variance. The findings revealed that Scheffe's, Post- hoc tests is more effective in spotting significant influence of students' class level on their academic performance in Mathematics than Turkey's HSD, Fisher's LSD and John Sidak, Scheffe's, Post- hoc tests was also more effective in spotting significant influence of students' self concept on their academic achievement in Mathematics than Turkey's HSD, Fisher's LSD and John Sidak. It was recommended among others that Student should be able to know which Post-hoc statistical test will be appropriately used for mean comparison especially in detecting significant difference. Keywords: Statistical comparison, Turkey's HSD, Scheffe's, Fisher's LSD, John Sidak test, self-concept, Mathematics achievement.

**Keywords:** Control environment, MSEs and Organizational performance

Corresponding Author: Y. M. Bashir

### **Background to the study**

In the design and analysis of experiments, post-hoc analysis (from Latin word post hoc, meaning "after this") consists of looking at the data after the experiment has been concluded for patterns that were not specified a priori. It is sometimes called by critics data dredging to evoke the sense that the more one looks the more likely something will be found. More subtly, each time a pattern in the data is considered, a statistical test is effectively performed. This greatly inflates the total number of statistical tests and necessitates the use of multiple testing procedures to compensate. However, this is difficult to do precisely and in fact most results of post-hoc analyses are reported as they are with unadjusted p-values. These p-values must be interpreted in light of the fact that they are a small and selected subset of a potentially large group of p-values. Results of post-hoc analyses should be explicitly labelled as such in reports and publications to avoid misleading readers (Howell, 2002).

In practice, post-hoc analyses are usually concerned with finding patterns and/or relationships between subgroups of sampled populations that would otherwise remain undetected and undiscovered were a scientific community to rely strictly upon a priori statistical methods (Fritz, 2009). Post-hoc tests also known as a posteriori tests greatly expand the range and capability of methods that can be applied in exploratory research. Post-hoc examination strengthens induction by limiting the probability that significant effects will seem to have been discovered between subgroups of a population when none actually exist. As it is, many scientific papers are published without adequate, preventative post-hoc control of the Type I Error Rate.

Post-hoc analysis with Turkey, Sheffe's, Fishers' test are specialized statistical tests designed to accompany ANOVA in order to adequately protect against making type 1 error ( false rejection). Multiple analyses are classified into two, namely the a posteriori and a priori (Hochberg & Tamhane, 1987). According to James, Carlson and keuls (1998), the first category, a posteriori, is also known and more popularly referred to as post hoc (ie after the fact) analysis. Post hoc or a posteriori analyses are those ones that the researcher decided to apply after the collection of data or after executing the ANOVA or after inspecting the various sample means. When post hoc-test is used to compare only two of the various means at a time, they are said to have taken the form of pair wise analysis (Best & Kahn, 2007). That is, in post hoc pair wise analyses, a pair (two) of the several sample means is compared at a time until all the possible pair wise analyses are done. Post-hoc multiple analyses that involve three or more of the sample means at a time are said to have taken the form of complex analyses. Complex post hoc analyses are used to compare the average of a set of sample means with another sample means.

Post-hoc tests, also known as a posteriori tests, greatly expand the range and capability of methods that can be applied in exploratory research. Post-hoc examination strengthens induction by limiting the probability that significant effects will seem to have been discovered between subgroups of a population when none actually exist. As it is, many scientific papers are published without adequate, preventative post-hoc control of the Type I Error Rate (Fisher, 2001). Joe (2005) and Muth, (2006) opined that Post-hoc analysis

is an important procedure without which multivariate hypothesis testing would greatly suffer, rendering the chances of discovering false positives unacceptably high. In particular, it has been shown that post hoc power in its simplest form is a one-to-one function of the p-value attained. This has been extended to show that all post hoc power analysis suffer from what is called “power approach paradox” (PAP), in which a study with a null result is thought to show more evidence that the null hypothesis is actually true when the p-value is smaller, since the apparent power to detect an actual effect would be higher. In fact, a smaller p-value is properly understood to make the null hypothesis less likely to be true. The question now is which of these post hoc techniques is most effective in identifying significant effect of group means?

The object of an agricultural experimenter is generally to measure the effect of varying some factor, for example the level of protein in poultry 14 diets. It is logical to expect that if different levels of protein are applied to different birds, the variation in the weight gains observed would be due partly to the different levels of feeding and partly to the basic variation between birds fed at the same level. The first problem for the experimenter is to disentangle these two parts of the variation i.e. to carryout an analysis of variance (ANOVA) so as to obtain an estimate of the true difference caused by his treatments, i.e. the feeding levels. A significant F-value from the ANOVA indicates that there are differences in the treatment means. The second problem of the experimenter may be to draw some further conclusions. He may want to decide which pairs of treatments are different, or he may want to contrast one treatment effect with the average of some other treatments. To identify where the differences are, he could do a series of pair wise t-tests. The major set back here is that the significance levels can be misleading. If you have 6 groups for example, there will be 15 pair wise comparisons of means; it has long been recognized, however, that if several t-tests have been performed at 5% level of significance, say, the probability that at least one of these is apparently significant is greater than 0.05 (Cochran & Cox 1957). If the t-tests are independent, this probability is 0.23 for 5 tests, 0.4 for 10 tests and 0.64 for 20 tests. 15 Multiple Comparison Procedures (MCPs) give more detailed information about the differences among the treatment means, while controlling the probability of making an incorrect decision. Several multiple comparison procedures are available to researchers. Some notable ones are:

1. Fisher's least significant Difference;
2. Duncan's New Multiple range test;
3. The Student-Newman Keuls' Procedure;
4. Tukey's Honestly Significant Difference;
5. Scheffe's Method.

Which procedure should be used depends upon which type of error is more serious (Schirley & Wearden 1985). Where a type I error is not serious, a very powerful test like Fisher's Least Significant Difference (LSD) could be used, otherwise more conservative tests like Tukey's or Scheffe's are preferable. The Fisher's multiple comparison procedure is based on a t-test. If the treatment groups are all of equal size  $n$ , then two sample averages ( $\bar{y}_1$  and  $\bar{y}_2$ ) can be tested for a significant difference by a t-statistic. In order to protect the overall type I error rate for the experiment, Fisher's procedure requires a prior

significant F-test in the analysis of variance. With this condition, the overall error rate (comparison wise error Rate, CER) has been shown to be approximately the  $\alpha$  of the F test (Shirley & Stanley 1985). Duncan (1975) considers the error rate for each pairwise comparison and allows a higher rate for pairs of sample averages that are further apart when ordered by size. This method also is believed to control the CER. All the other procedures of multiple comparisons also aim at reducing the error rates so that valid conclusions can be drawn at the end of the analysis of variance.

Multiple comparisons with Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post-hoc tests with the influence of students' self-concept on their academic achievement in Mathematics has been a powerful tool for means comparison. These procedures are used to control for the family-wise error rate. For example, suppose that we have four groups and we want to carry out all pair-wise comparisons of the group means (Carma & Swanson, 2003). Steel and Torrie, (2000) carried out a comparative study in Canada. The researcher examined Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post-hoc tests on the influence of senior secondary school students using their self-concept (academic social and economic self concept) on their Mathematics achievement. Two hypotheses were tested at .05 level of significance with a sample of 467 Senior secondary one students from the state with the use of stratified and purposive sampling. A 31 achievement test on Mathematics was used constructed, validated and administered to the respondents. The findings revealed that scheffe's test seems to be the best post-hoc in multiple comparisons as it accept comparing groups irrespective of their category.

The finding also revealed that Turkey's HSD, Fisher's LSD and John Sidak multiple range tests does not require a preliminary significant overall F-test and it uses a set of significant ranges each range depending upon the number of means in the comparison. However, it does not give real protection to the significance level. To Carmer and Swanson (2003) have claimed that Duncan's method is superior to Tukey's because of greater power without considering that the greater power is due to its higher type 1 error rate. John sidak test, different critical values are used depending on the span of the two ranked averages being compared. It is applicable only in situations where all k treatments are equally replicated n-times.

Duncan (2005) researched with 98 students in United States of America. The aim of the study was to Scheffe's, Fisher's LSD, Turkey's HSD and John Sidak Post-hoc tests on students self-concept in Mathematics. Self-concept was divided in to academic self concept, social self concept and religious self-concept on Mathematics achievement. The study utilized two research questions and hypotheses which were tested at .05 and .01 alpha level, stratified and simple random sampling was used in the study. A 22 item on Mathematics achievement test was used constructed and validated for data collection. The findings revealed that Scheffe's test has the advantage of giving the experimenter the flexibility to test any comparisons that appear interesting, with very high statistical power than other post hoc test compared. The finding also revealed that Scheffe's test was found to be more liberal and accommodation with less error rate than Turkey's HSD, Fisher's LSD and John Sidak as it was found that this method is relatively easy to apply, it

takes into account the number of treatments in the experiment and it protects the Type I error rate. The study also revealed that the problem with Turkey's HSD and John Sidak is that it cannot be used to construct confidence intervals for difference between means and it is less sensitive than the LSD and Scheffe's procedures in detecting real difference (Chew 2006).

In a survey study by Hayter (2004) on statistical comparison using turkey's HSD, Scheffe's, Fisher's LSD and John Sidak test in a study of students self-concept and Mathematics achievement in United State using questionnaire as the data collection procedure. The finding revealed that Scheffe's method is simple and it allows many comparison statements to be made (or confidence intervals to be constructed) while still assuring an overall confidence coefficient is maintained. The MCP applies to an ANOVA situation when the analyst has picked out a particular set of pairwise comparisons or contrasts or linear combinations in advance and it is valid for equal and unequal sample sizes. The scheffe and Fishers LSD method are a very flexible multiple comparison procedure and it compensates for the fact that multiple comparisons are being made. The finding spotted out major problems with Turkey's HSD and John Sidak that they are not an exact procedure as the groups compared may differ significantly in terms of their respective sizes. The Bonferroni adjusted p value is larger than the true p value The Sidak's method, like the Bonferroni method, also makes an adjustment for multiple comparisons. The adjustment computes the level of significance as:  $1 - (1 - \text{LSD significance})^{K(K-1)/2}$  The adjustment is aimed at reducing our overall chance of falsely rejecting each hypothesis to  $\alpha$  rather than letting it increase with each additional test. The Sidak's method makes adjustment for the comparison error rate as well as the level of significance and it is less conservative than the Bonferroni Fisher's LSD and Scheffe's methods (Einot & Gabriel, 2005).

### **Statement of the Problem**

Post-hoc analyses are specialized statistical analyses that often accompany ANOVA in data analysis. They are of great elegance, utility and flexibility and are the most effective method available for analyzing experimental data in which several treatments or factors are represented. In the simplest case of a one-way ANOVA, in which the experiment consists of a number of independent treatments or groups, the first stage of the analysis is to carry out a variance ratio test (F-test) to determine whether all group means are the same.

It has been observed that researchers out-rightly utilize post-hoc statistical test without full knowledge of the basic assumptions underlying such test. This has posed a worrisome problem in the analysis of educational data with Mathematics performance. In many circumstances, different post-hoc tests may lead to the same or different conclusions; and which of these tests is actually suitable has often been a major problem among research scholars. However, each test addresses the intended educational data in a unique way. It is against this background that this study seeks to analyze and compare post hoc statistics using Turkeys,' Scheffes' and Fishers' test on students' variables in mathematics achievement in south-south, Nigeria.

A good way of deciding which test to use is to consider the purpose of the experimental investigation, if the purpose is to decide which of a group of treatments is likely to have an effect, then it is better to use a more liberal test such as Fisher's Least Significant Difference and Fisher's Honest Significant Test (HSD and LSD). By contrast, if the objective is to be as certain as possible that a particular treatment does have an effect then a more conservative test such as the Scheffé's test would be appropriate. Turkey's HSD and the compromise method fall between the two extremes and the Student-Newman-Keuls (SNK) method is also a good choice.

In many circumstances, different post-hoc tests may lead to the same conclusions and which of the above tests is actually used is often a major problem among budding researchers. Failing to understand that, the individual tests vary in how effectively they address a particular statistical problem and their sensitivity to violations of the assumptions of ANOVA has led to series of wrongful and outright application of post hoc test analysis. This has led to critical problem in data analysis one of which is the possibility of making a Type 1 error, i.e., rejecting the null hypothesis when it is true.

### **Research Question**

The following research question was formulated to guide this study;  
What is the level of effectiveness in Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post- hoc in spotting significant influence of students' self-concept on their academic achievement in Mathematics?

### **Methodology**

The research design that will be adopted for this study will be ex-post facto research design. According to Idaka and German (2012), Ex-post facto literally means 'after the fact'. It is a non-experimental design in which the phenomena of interest have already occurred and cannot be manipulated in any way. In other words, it is used in a situation in which variation in the independent variables has occurred naturally and no random assignment or manipulations are possible. The population of this study comprise all the Senior Secondary School students' in south-south, Nigeria which constitute a population 18,463 Senior Secondary School (SS1) students (Source; planning research and statistics, States Secondary Education Board February, 2014). The stratified and purposive sampling techniques were used for the study. The sample consisted of 866 respondents comprised of male and female students from south-south. A 29 items structured Mathematics Performance Test Questionnaire (MPTQ) was constructed by the researcher with the assistance of research experts and supervisors for data collection. Also, a table of specification covering all the course contents of SS 1 was constructed to appropriately show that the instrument covers the domain intended to measure.

### **Result and Discussions**

This section critically presents the findings of the study as follows:

#### **Research question one**

What is the level of effectiveness in Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post- hoc in spotting significant influence of students' self-concept on their academic achievement in Mathematics?

**Table 1**

Post-hoc comparison with Turkey's, Scheffe's and Fisher's LSD and John Sidak Post-hoc test with the influence of students' self-concept on their academic achievement in Mathematics.

POST-HOC	Self-concept	N	Academic	Social	Religious
TURKEY'S HSD	Academic	342	15.90	-.67*(Sig.= .040)	-.66* (Sig.= .043)
	Social	310	3.78	16.57	.01 (Sig.= 1.00)
	Religious	214	3.32	-.88	16.56
SCHEFFE	Academic	342	15.90	-.67* (Sig.= .001)	-.66* (Sig.= .003)
	Social	310	5.99	16.57	.01 (Sig.= .554)
	Religious	214	4.09	-.67	16.56
LSD	Academic	342	15.90	-.67* (Sig.= .015)	-.66* (Sig.= .016)
	Social	310	-2.11	16.57	.01 (Sig.= .997)
	Religious	214	-2.36	.04	16.56
JOHN SIDAK	Academic	342	15.90	-.67* (Sig.= .045)	-.66* (Sig.= .048)
	Social	310	-2.17	16.57	.01 (Sig.= 1.00)
	Religious	214	-2.00	.56	16.56

MSW = 7.64

From Table 1, it can be discerned that with Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post-hoc tests use in the experiment to determine which is more effective in spotting significant influence of students' self-concept on their academic achievement in Mathematics. The mean differences for academic and social self-concept has a significant mean differences as ( $\bar{X} = -.67, t=3.78; P=.040$ ), for academic and religious self concept there is also a significant mean difference as ( $\bar{X} = -.66, t=3.32; P=.043$ ), however, there is no significant mean difference for social and religious self concept as ( $\bar{X} = .01, t=-.88; P=.393$ ). When Scheffe's Post-hoc was employed in the statistical test with academic and social self-concept there is a significant mean difference as ( $\bar{X} = -.67, t=5.99; P=.001$ ), for academic and religious self-concept there is also a significant mean difference as ( $\bar{X} = -.66,$

$t=4.09$ ;  $P=.003$ ), there is no significant mean difference for social and religious self-concept as ( $\bar{X}=.01$ ,  $t=-.67$ ;  $P=.554$ ), with LSD Post-hoc for academic and social self-concept there is a significant mean difference as ( $\bar{X}=-.67$ ,  $t=-2.11$ ;  $P=.040$ ), for academic and religious self-concept there is also a significant mean difference as ( $\bar{X}=-.66$ ,  $t=-2.35$ ;  $P=.043$ ), there is no significant mean difference for social and religious self-concept as ( $\bar{X}=.01$ ,  $t=-.04$ ;  $P=.393$ ), with John Sidak Post-hoc test was carried out for academic and social self-concept there is a significant mean difference as ( $\bar{X}=-.67$ ,  $t=-2.17$ ;  $P=.044$ ), for academic and religious self-concept there is also a significant mean difference as ( $\bar{X}=-.66$ ,  $t=-2.00$ ;  $P=.040$ ), there is no significant mean difference for social and religious self-concept as ( $\bar{X}=.01$ ,  $t=-.56$ ;  $P=1.00$ ). The result implies that that Scheffe's post-hoc test is better in sporting lesser significance than Fisher's LSD, Turkey's and John Sidak test. This is evident in their respective p-values as indicated in Table 1.

### Discussion of Findings

The result emanating from the findings of the study was discussed hypothesis by hypothesis as presented below: Compare Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post-hoc tests with the influence of students' class level on their academic performance in Mathematics.

The result revealed that with Multiple comparisons with turkey Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak Post-hoc tests with the influence of students' class level on their academic performance in Mathematics. the result showed that sheffe's test was detct significant than other post hoc test as seen from the analysis. The result agrees with that by teel and Torrie, (2000) whose findings revealed that scheffe's test seems to be the best post-hoc in multiple comparison as it accept comparing groups irrespective of their category. The finding also revealed that Turkey's HSD, Fisher's LSD and John Sidak multiple range tests does not require a preliminary significant overall F-test and it uses a set of significant ranges each range depending upon the number of means in the comparison. However, it does not give real protection to the significance level. The result is also in consonance with that of Duncan (2005) who found that that Scheffe's test has the advantage of giving the experimenter the flexibility to test any comparisons that appear interesting, with very high statistical power than other post hoc test compared. The finding also revealed that Scheffe's test was found to be more liberal and accommodation with less error rate than Turkey's HSD, Fisher's LSD and John Sidak as it was found that this method is relatively easy to apply, it takes into account the number of treatments in the experiment and it protects the Type I error rate. The study also revealed that the problem with Turkey's HSD and John Sidak is that it cannot be used to construct confidence intervals for difference between means and it is less sensitive than the LSD and Scheffe's procedures in detecting real difference. In the same vein Edwards & Berry (2007) findings revealed that Scheffe and Fishers LSD method as being easier and liberal to apply since it requires a single value for judging the significance of all differences, that is, it is a fixed range. It can also be used to compute a set of confidence intervals for differences and it takes into account the number of treatments in the experiment. Studies by Hayter (2004) found that Scheffe's method is simple and it allows many comparison statements to be made (or confidence intervals to be constructed) while still assuring an overall confidence coefficient is maintained. The



finding spotted out major problems with Turkey's HSD and John Sidak that they are not an exact procedure as the groups compared may differ significantly in terms of their respective sizes.

### **Summary, Conclusion and Recommendations**

The purpose of the study was to carry out a comparative Post-Hoc Comparison using Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak in a study on the influence of Students self-concept on academic achievement in Mathematics in south-south, Nigeria. In executing the purpose of this study, one research questions was answered to guide the study. Empirical, conceptual and theoretical literature was reviewed in line with the present study. The study utilized the ex-post facto research design with a population of 23,109 students from the 269 public secondary schools were used for the study. A sample size of 866 students was drawn using stratified and purposive sampling technique. Data was gathered through a researcher's developed instrument titled structured Mathematics Performance Test Questionnaire (MPTQ). To establish the reliability of the instrument, Cronbach Alpha reliability method was utilized. The questionnaire was adequately administered to the respondents. The statistical technique utilized for data analysis was One-way analysis of variance. Since the hypothesis was rejected the post hoc multiple comparison with Turkey's HSD, Scheffe's, Fisher's LSD and John Sidak was adopted to detect the post-hoc with least significant difference. The result of the statistical analysis revealed that Scheffe test is the best Post-hoc in comparing when compared with Turkey's HSD, Fisher's LSD and John Sidak Post- hoc tests with the influence of students' self-concept on their academic achievement in Mathematics.

### **Conclusion**

Once an Analysis of Variance (ANOVA) test has been completed, the researcher may still need to understand subgroup differences among the different experimental and control groups. The subgroup differences are called "pairwise" differences. ANOVA does not provide tests of pairwise differences. When the researcher needs to test pairwise differences, follow-up tests called Post-hoc tests are required. ANOVA output does not provide any analysis of pairwise differences, so the first approach that comes to mind is to perform a number of t-tests between each of the pairs of interest. The researcher may wish to test differences between one or more study groups and a set of combined study groups. Pairwise t-tests cannot perform that kind of analysis. However, there are a set of multivariate statistics that overcome all the limitations of the pairwise t-test approach. This category of statistics is called multiple comparison analysis. One of the multiple comparison analysis statistics should be used to examine pairwise and subgroup differences after the full ANOVA has found significance. The key tests of pairwise differences include: John Sidak, Sheffè, Tukey, Fisher LSD, Turkey HSD etc.

Each of the multiple comparison analysis (MCA) tests has its own particular strengths and limitations. Some will automatically test all of the pairwise comparisons; others allow the researcher to limit the tests to only pairs or subgroups of interest. Each approach has implications for alpha inflation and for the kind of answers the researcher can derive from the test. Therefore, the choice of an MCA statistic, as all choices about which statistic to use

should be based on the specific research questions. For example, the researcher may have one experimental group of particular interest that should be compared separately against each of the control groups. Alternatively, the researcher may want to compare one experimental group against a combination of all the control groups, or against only some of the control groups, or even against one or more of the other experimental groups.

### Recommendation

Based on the findings of the study, the following recommendations are made:

1. Student should be able to know which post-hoc comparison will be appropriately used for mean comparison, they should only be run when you have shown an overall significant difference in group means (i.e., a significant one-way ANOVA result).
2. Post-hoc tests attempt to control the experiment wise error rate (usually  $\alpha = 0.05$ ) in the same manner that the one-way ANOVA is used, research scholars should be well equipped on the need to acquire a better understanding of family wise error rate in multiple comparison.
3. Tukey's honestly significant difference (HSD) or Scheffé post hoc tests, often, Tukey's HSD test is recommended by this research study because it is not as conservative as the Scheffé test (which means that you are more likely to detect differences if they exist with Tukey's HSD test).

### References

- Best, O. & Kahn, W. (2007). *Keeping America competitive: five strategies to improve mathematics and science education*. Report by Education Commission of the States, <http://www.ecs.org/clearinghouse>.
- Carma, S.G & Swanson, M.R. (2003). An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association* 68 (341), 66-74.
- Chew, V. (2006). Comparing treatment means: A compendium. Hortscience. New York: John Comparisons. *Journal of Science*, 43 (12), 913-928.
- Cochran, G. & Cox, O.O. (1957). *Statistical methods for psychology, 5th ed.*. Belmont, CA: Duxbury Press.
- Duncan D.B. (2005). T-test and intervals for comparisons suggested by data. *Euphytica*: (6), 112-122
- Edwards, D. & Berry, J.J. (2007). The efficiency of simulation-based multiple comparisons. *Journal of Science*, 43 (12), 913-928.
- Einot, I. & Gabriel, K.R. (2005). A study of the powers of several methods of multiple comparisons. *Journal of Humanities*, 70 (7), 35-37.

- Fisher, U. (2001). The *distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation*. *Biometrika* 31 (1), 20-30.
- Fritz, P. F. (2009). Scheffé's more powerful F-Protected post Hoc procedure. *Journal of Educational and Behavioral Statistics* 25 (1), 13-19.
- Hayter, A.J. (2004). A proof of the conjecture that the Tukey-Kramer method is 87 conservative. *The Annals of Statistics*, 12, 61-75.
- Hochberg, R.T. & Tamhane, W. Y. (1987). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.
- James, C. & keuls, B. (2008). The use of the studentized range in connection with an analysis of variance. *Journal of Science* 1 (9), 112-122.
- Joe, U. (2005). *Comparing treatment means: a Compendium*. Hortscience. New York: John.
- Muth, O.P.(2006). *Understanding statistics in the behavioral sciences* (7th ed.). Belmont, CA: Thomson/Wadsworth.
- Schirley, I.Q. & Wearden, C. (1985). *SPSS for introductory statistics: Use and Interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shirley & Stanley (1985). *Multiple comparisons: theory and methods*. London: Chapman and Hall.
- Steel, G. & Torrie,D.I. (2000). *Using SPSS for windows and Macintosh: analyzing and understanding data* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.