



---

## A Prototype Expert System for Selecting a Multiple Comparison Test

**Adel M. Aladwani**

Department of QM & IS  
College of Administrative Sciences  
Kuwait University  
P. O. Box 34927  
Edailiyah, Kuwait 73251  
E-mail : [adwani@kuc01.kuniv.edu.kw](mailto:adwani@kuc01.kuniv.edu.kw)

---

- [Abstract](#)
  - [Introduction](#)
  - [ES and Statistical Analysis](#)
  - [The MCP Choice Problem](#)
  - [Detail of the System](#)
  - [Evaluation of the System](#)
  - [Result and Discussion](#)
  - [Conclusion](#)
  - [Acknowledgment](#)
- 

### **Abstract**

Modeling the choice of a Multiple Comparison (MC) test is one important area of expertise that did not attract enough attention from researchers interested in the applications of Expert Systems (ES) technology. Probably this is so because of the highly demanding nature of the MCP selection problem. Modeling the MC procedure choice problem and building a prototype Expert System to help researchers as well as students is one important aim of this article. Another goal of the paper is to examine the appropriateness of ES as a supporting development paradigm for modeling such a perplexing process. A prototype expert system was developed and described. The results of an experimental test are encouraging.

**Keywords:** Expert Systems; Multiple Comparison Tests.



### **Introduction**

Since it was first commercially introduced in the beginning of the past decade, Expert Systems (ES) technology has been applied to numerous problem domains. The applications of ES to Operation Research/Management Science (1), in general, and to the area of statistical analysis (2), in particular, are two areas that have been the focus of close attention from researchers. The expertise of one or more expert analysts in a specific statistical analysis problem domain can be captured in a computer program to achieve several critical goals such as providing consistent and timely statistical knowledge to users. Actually, several attempts have been made towards that general end (e.g., 3 and 4).

Selecting a Multiple Comparison Procedure (MCP) is one important problem domain, however, that did not attract enough attention. Although there are many analytical models that have been developed to help in conducting multiple comparisons among the means, no attempt has been made to develop a decision aid to help researchers in selecting the appropriate procedure given a research design. The importance of such a problem stems from the fact that some times to avoid confusion, a researcher (or student) may choose the test that (s)he knows best. Hence, some possible interesting findings are neglected because of this rather rudimentary choice. Modeling the MCP choice problem and building a prototype Expert System to help experienced and novice researchers as well as students with the selection process is one important aim of this article. Another goal of the paper is to examine the appropriateness of ES as a supporting development paradigm for modeling such a choice.

The rest of the paper is developed as follows. Section two and three serve as a review of ES technology and application in the area of statistical analysis. Section 4 characterizes the MCP problem domain and models its choice process. Section 5 describes the structure of the prototype. Section 6 reports the results of an experimental test. The final section offers concluding remarks and implications for future research.



## **ES and Statistical Analysis**

There has been a wide spread use of ES in the area of statistical analysis. These applications can be classified according to two different levels of abstraction: the general problem domain discussed and the particular statistical technique investigated.

Under the generic category, ES have been used for diagnosis, selection, interpretation, control, prediction as well as many other domains. For example, Kumar and Cheng (5) gave an illustration of a system that detects trends from historical data and selects an appropriate forecasting model. Mellichamp (6) developed an expert system that interprets the results of a simulation experiment. Remus and Kotteman (4) described a hypothetical statistical Expert System for predicting the appropriate statistical technique for the specific problem at hand. White (7) describes a prototype expert system for choosing statistical tests. The system asks the user a set of questions and builds up a picture until it is able to suggest the appropriate statistical test.

In the vein of the specific technique category, ES has been used to model techniques like regression, MANOVA, to name a few. For instance, Pregibon and Gale (8) developed a knowledge-based system that advises users in the analysis of regression problems. In their ES, the authors provide guidelines as to when such type of a statistical analysis technique is best used and what the regression coefficients do mean. In addition, Hand (9) described a hypothetical system that depicts the statistical strategy for Multivariate Analysis of Variance

(MANOVA). A Detailed description of the decision process of such research designs was outlined. Theoretical and practical assumptions are considered.



## The MCP Choice Problem

One of the recurrent problems when conducting behavioral research is the question of choosing a multiple comparison method that best serves the purpose of the researcher. Pedhazur (10) notes that the need to choose among the different MCPs arises when a researcher is faced with results that indicate the rejection of the null hypothesis that all the means are equal given a pre-designated alpha level. He asserts that this rejection, however, does not provide enough information as to which means are significantly different. Thus, multiple comparison tests are used to further examine the various relationships between the means.

It is so often that a researcher encounters such a situation where (s)he has to choose from a number of multiple comparison procedures. Sometimes to avoid the confusion, a researcher (or student) may choose the procedure that (s)he knows best. Consequently, several important comparisons may be overlooked; and thus some possible interesting findings are neglected.

To date many MC procedures have been developed but only eight have become prevalent in use. These are Trend Analysis, Planned Orthogonal Contrasts, Dunnett's method, Dunn's method, Scheffe's method, Tukey's method, Marascuilo's method, and Newman-Keuls's method. These techniques are classified according to six dimensions (11): orthogonal versus non-orthogonal; planned versus non-planned; stepwise versus simultaneous test procedures; pairwise versus non-pairwise; number of comparisons; type of statistics; and type of error rates. Other evaluation criteria include the assumptions concerning the homogeneity of variance (12). The following is an attempt to explain how some of the considered factors would determine the decision as to which MCP should be used for a particular research design.

To select an appropriate MCP that would fit a given comparison problem, the researcher should take into account a variety of fundamental considerations. First is the type of statistics under examination. The researcher should ask him- or herself this fundamental question: am I interested in a comparison among means or some other statistic? Of the previously mentioned MCP methods, only Marascuilo method, for example, could define a technique that contrasts any number of independent statistics, e.g., Pearson's  $r$  correlation coefficient (13). Therefore, if a researcher is interested in comparing statistics other than groups' means, Marascuilo method is the appropriate choice.

The question of whether there is a continuum underlying the levels of the independent variable is another major evaluation criteria. Trend Analysis provides superior information regarding the mean differences between the different levels of the independent variable when there is a continuum (10). In other words, the choice is highly constrained by the research question. Although the decision as to whether there is a continuum underlying the variable is a "manual" process, as explained in the above statistical ES examples, it can be "automated" by incorporating special capabilities in the ES to detect such an attribute. Hence, this concern could be relieved somewhat.

Moreover, whether the researcher wishes to conduct a planned or non-planned comparison is another important criteria. Planned refers to the fact that a researcher has a predefined set of hypotheses (s)he wants to answer. A non-planned test refers to situations where according to

whether there is at least one significant result, further investigations are carried out to pinpoint which contrast(s) is significant. Dunn, Dunnett, and Planned Orthogonal Contrasts, for example, are a priori techniques. However, Tukey, Scheffe, Marascuilo, and Newman-Keuls are post-hoc contrasts. Therefore, the type of the research hypothesis, whether it is a priori considered or otherwise, determines the most suitable MCP technique.

The MCP methods, also, can be distinguished according to the reasoning behind the alpha level, i.e., the probability of making a type I error. All MCP techniques except Newman-Keuls technique and Planned Orthogonal Comparisons use family-based alpha. The error rate per family refers to the error rate used in testing the null hypotheses of all contrasts of interest that are related to a certain treatment or interaction (14). The other type of error rate of interest is the contrast-based alpha, which is the probability that a contrast will be incorrectly declared significant (14). A researcher's decision as to which contrast-based alpha is preferred and the level of precision as opposed to the level of power desired should be taken into account when selecting a multiple comparison method.

The influences of the above-discussed parsimonious dimensions affect a researcher's decision as to what method should be employed for a specific analysis. Although not all the relevant factors are taken into account in our discussion, the addressed dimensions tell most of the story. Most of the introductory statistical classes that I am informed about consider mainly the influences of the above reviewed determinants while the influences of the remaining factors are usually left for more specialized courses.



## **Details of the System**

A prototype ES was developed using EXSYS PRO (a microcomputer shell) to model the selection process as discussed in the previous section. The final prototype is a menu driven system. The expert system operates in a dialogue mode. The process starts by asking the end user to answer a set of various questions that compose the pre-constructed knowledge base. Based on the replies of the user, a choice is made.

The system was developed using IF-THEN-ELSE rules. There are nine goals in the prototyped system including: Trend Analysis, Planned Orthogonal Contrasts, Dunnett's method, Dunn's method, Scheffe's method, Tukey's method, Marascuilo's method, Newman-Keuls's method, and the fact that there are no differences among the means.

The proposed prototype has rules for the different goals in the goal box. For example, there is a rule that states the condition for selecting the Marascuilo method as follows:

**IF:**

The comparison in the research problem is among {parameters other than means }

**THEN:**

Marascuilo method

The above rule indicates that the Marascuilo method is the appropriate MCP choice if the researcher is interested in comparing statistics other than means.

Another rule in the rule base describes the conditions for selecting Trend analysis as follows:

*IF:*

*The comparison in the research problem is among {means} and: Underlying the independent variable is {a continuum}*

*THEN:*

*Trend analysis*

That is, Trend analysis is the appropriate procedure when the researcher is interested in comparing groups' means and there is a continuum underlying the independent variable.

Another rule states:

*IF:*

*The comparison in the research problem is among {means} and: Underlying the independent variable is {no continuum} and: The variances are {non-homogenous}*

*THEN:*

*Marascuilo method*

Under certain circumstances, Marascuilo method can be used to compare groups' means. The above rule summarizes these conditions. It indicates that when the comparison is among groups' means, when there is no continuum underlying the independent variable, and when the variances are non-homogenous, Marascuilo method could be an appropriate MCP choice.

EXSYS PRO uses backward-chaining inferencing. Backward chaining is a goal-driven mechanism, which starts with the goal and works backward searching for arguments that satisfy the specified objective. To illustrate, in order to reach "RECOMMENDATION-1", which could be "Trend analysis", the system will search the THEN part of the knowledge base to find the first "value" (goal). Once the system locates the first rule that satisfies this recommendation, the system will try to prove the rule's conditions and fire the rule. The system will repeat this process until there are no more goals in the goal box.



## **Evaluation of the System**

### **Method**

#### *Subjects*

Subjects were twenty-two graduate students enrolled in a statistical analysis class at a large university located in the Midwestern side of the United States of America. The students were randomly selected and evenly assigned to two groups (each group with  $n=11$ ); a group that uses the expert system in making MCP selection decisions and a group without the aid of the developed expert system. No significant background differences were found between the two groups in terms of computer experience, age, and GPA.

#### *Instrument*

Three statistical human-experts were offered nine MCP problems. Each problem is carefully designed along the guidelines of the previously described choice model to represent a corresponding decision output. The human experts were asked to make what is thought of as an

appropriate MCP choice for each of the nine short scenarios. The consensus among the experts on a MCP problem is considered the "correct" choice for that particular problem.

### *Procedure*

The students in both groups were asked to make a choice regarding each of the MCP problems that were introduced earlier to the human-experts. The students are expected to consider the different assumptions underlying a particular test and consequently come up with the most appropriate choice.

The same TA guided the two groups to reduce experimenter's effect. The role of the TA is to assure the smooth operation of the experiment by guiding, not answering, confused students. The ES group was allowed fifteen minutes training on the prototype before the beginning of the experiment.

No time limits were imposed, but all students completed testing within forty-five minutes.

### *Model and Data Analysis*

The dependent variable in the experiment is the number of correct decisions a student makes out of nine. The research model can be specified as follows:

Number of correct answers =  $\mu + \epsilon$  (type of group)

The associated research hypothesis can be stated as follows:

**H<sub>1</sub> : The group that uses the prototype will show higher mean than that of the non-use group**

A t-test was used to test the difference between the two groups.



## **Results and Discussion**

The results of a descriptive analysis show that little more than 50 percent of the students that used the ES made the "correct" selection for all nine problems compared to a 27 percent complete "correct" answers for the students without the aid of the developed ES. The two groups also differed in the time required to finish the test. The ES group took, on average, six minutes to complete the test, while the control group needed thirty-two minutes on average.

Furthermore, the "use" group has a mean score of 7.72 and the "non-use" group has a mean score of 5.18 correct answers (Table 1 summarizes the descriptive statistics). The result of a t-test shows that there is a significant difference between the means of the two groups ( $t_{20}=2.67$ ) at 0.05 alpha level. This leads to the rejection of the null hypothesis that the two means are equal. More specifically, the rejection supports the hypothesis that using the prototype ES does make a difference in performance. Table 2 presents t-test output.

### **Table 1: Descriptive statistics**

Group	N	Mean	STD	Std. Err.	Min.	Max.
1 (ES)	11	7.72	1.79	0.54	4	9
2 (NES)	11	5.18	2.60	0.78	2	9

**Table 2: t-test Output**

t-test score	df	Prob T
2.67	20	0.0146

Although the effect of automating the process is one explanation for the difference, it is clearly possible that the problem solving guidance that the system provides for the users has its share (e.g. 15, 16, and 17). A programmed and interactive system such as the one we have here, helps users add some sort of discipline to their solution process when attacking the depicted problem. The by product of this situation is more quality decisions in a shorter period. In general, the developed decision aid proved to have a significant effect on user performance.

An interesting finding of the analysis is the high score of correct answers the control group has made. One explanation of this result is that the professor teaching this course made a noticeable job in explaining the rather perplexing MCP subject.



## Conclusion

There are two goals in this research. The first is that the proposed model and prototype expert system seeks to provide consistent and up-to-date MCP selection knowledge both for researchers and students. The designed system, however, is not proposed to take the place of good intuitive judgement of these user groups. Given the test results, the system is recommended to be taken as both a training game for students and a decision-aiding tool for researchers. The second goal is to examine whether the ES development paradigm can be used for the MCP choice problem. The successful completion and adequate testing results of the system demonstrate that it is quite adequate to use Expert Systems for such an important decision process.

There are three possible future research directions. First is expanding the knowledge base to incorporate deep knowledge as opposed to the relatively shallow knowledge that form the basis for our prototype. This requires the consideration of the specific antecedents of each criterion. Second is enhancing the system by integrating it with an existing statistical program so that the input to the system is downloaded directly without human intervention. Blackboarding could also be used. This will achieve two viable objectives: 1) to reduce input errors, and 2) to alert the end-user as to what (s)he might have overlooked in terms of possible alternatives (e.g., to flag that additional comparisons are needed). Finally, although system testing was not the main purpose of this study, future research directions should consider the limitations of the experiment. Sample size, sample constituents, and technology type should be considered in future replications of the study.

## Acknowledgment

The author would like to thank Professors J. Cox, J. Mouw, T. Ravichandran, and M. Troutt for

their assistance.



---

## References

- 1.O'keefe, R. "Expert Systems and Operational Research--Mutual Benefits." J. of Operation Research Society, 36:2, 1985, pp. 125-129.
- 2.Gottinger, H. "Statistical Expert Systems." Expert Systems, 5:3, 1988, pp.186-195.
- 3.Marcoulides, G. "An Expert System for Statistical Consulting." 12th Decision Sciences Institute Proceedings, 1981.
- 4.Remus, W. and J. Kotteman "Toward Intelligent Decision Support Systems: an Artificially Intelligent Statistician." MIS Quarterly, 10:4, 1986, pp. 403-418.
- 5.Kumar, S. and H. Cheng "An Expert System Framework for Forecasting Method Selection." Proceedings of the Hawaii International Conference on Systems Science, 1988, pp. 86-95.
- 6.Mellichamp, J. "An Expert System for FMS Design." Simulation, 48:5, 1987, pp. 201-209.
- 7.White, A. "An Expert System for Choosing Statistical Tests." New Review of Applied Expert Systems, 1:1, 1995, pp. 111-122.
- 8.Pregibon, D. and W. Gale "REX: An Expert System for Regression Analysis." COMPSTAT, 1984, pp. 242-248.
- 9.Hand, D. "Patterns in Statistical Strategy." In W. Gale (editor), Artificial Intelligence in Statistics. Addison- Wesley, Reading, MA, 1986, pp. 355-387.
- 10.Pedhazur, E. Multiple Regression in Behavioral Research. 2nd edition. Dryden Press, Fort Worth, TX, 1982.
- 11.Toothaker, L. Multiple Comparison Procedure. Sage Publications, Newbury Park, 1993.
- 12.Petrinovich, L. and C. Hardyck "Error Rates for Multiple Comparison Methods: Some Evidence Concerning the Frequency of Erroneous Conclusions." Psychological Bulletin, 71:1, 1969, pp. 43-54.
- 13.Marascuilo, L. "Large-Sample Multiple Comparisons." Psychological Bulletin, 65:4, 1966, pp. 289-299.
- 14.Kirk, R. Experimental Design: Procedures for the Behavioral Sciences. 2nd edition. Brooks/Cole Publishing Co., Pacific Grove, CA, 1982.
- 15.Benbasat, I. and A. Dexter "Individual Differences in the Use of Decision Support Aids." J. of Accounting Research, 20:1, Spring 1982, pp. 1-11.

16.Mcintyre, S. "An Experimental Study of the Impact of Judgement-Based Marketing Models." Management Science, 28:1, January 1982, pp. 17-23.

17.Sharda, R., S. Barr, and J. McDonnell "Decision Support System Effectiveness: A Review and an Empirical Test." Management Science, 34:2, February 1988, pp. 139-159.



---

## APPENDIX

### 1. Example problem

Given the following situation, please recommend the appropriate Multiple Comparison test. You should select one and only one test.

#### Situation:

A researcher is interested in examining the difference in algae growth between the tanks that are treated with chemicals (tank 1 with chemical X and tank 2 with chemical Y) and the tank that was not treated. The researcher has a chance to obtain three tanks that are equal in capacity.

### 2. Example run

\*\*\* Welcome to ESSMCT: The Expert System for Selecting an MC Test \*\*\*

This interactive system is designed to help you select an appropriate multiple comparison procedure for your research problem. The system will ask you to answer a number of questions. At the end of the session, a recommendation as to which test you should use will be provided ... Good Luck.

The comparison in the research problem is AMONG MEANS  
Underlying the independent variable is NO CONTINUUM  
The variances are HOMOGENOUS  
The contrasts are PLANNED  
The contrasts are INDEPENDENT

Based on what you have described, we suggest that you use:  
PLANNED ORTHOGONAL CONTRASTS

You do not have to perform an overall test of the null hypothesis. You seem to have determined a priori a set of hypotheses to test. The receptive tests as you described are orthogonal, hence, the tests should be independent. This helps to cause the comparisonwise error rate and the experimentwise error rate to be equal to alpha.



