INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.

8711623

Athey, Susan

A MENTOR SYSTEM INCORPORATING EXPERTISE TO GUIDE AND TEACH STATISTICAL DECISION MAKING

The University of Arizona PH.D. 1987

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages ____
2. Colored illustrations, paper or print ____
3. Photographs with dark background ____
4. Illustrations are poor copy ____
5. Pages with black marks, not original copy ____
6. Print shows through as there is text on both sides of page ____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements ____
9. Tightly bound copy with print lost in spine ____
10. Computer printout pages with indistinct print ____
11. Page(s) ____ lacking when material received, and not available from school or author.
12. Page(s) ____ seem to be missing in numbering only as text follows.
13. Two pages numbered 48. Text follows.
14. Curling and wrinkled pages ____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other ____

University
Microfilms
International

A MENTOR SYSTEM INCORPORATING EXPERTISE TO GUIDE AND TEACH STATISTICAL DECISION MAKING

by

Susan Athey

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 7

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read

the dissertation prepared by Susan Athey

entitled A Mentor System Incorporating Expertise to Guide and Teach Statistical Decision Making

and recommend that it be accepted as fulfilling the dissertation requirement

for the Degree of Doctor of Philosophy.

[signature]	Jan 30, 1987
	Date
[signature]	1-30-87
	Date
E. Sue Weber	1-30-87
	Date
[signature]	1-30-87
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

[signature]	Jan 30, 1987
Dissertation Director	Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED Susan Athey

ACKNOWLEDGEMENT

Numerous people deserve my thanks for the help they gave over the past three and a half years. First of all, thanks must go to the people who served on my committee: Jay Nunamaker, chairman, George Summers, Sue Weber, Sudha Ram, and Gary Wagner. Special thanks must go to Jay for his infinite supply of ideas, to George for his infinite supply of statistics books, and to Sue for her infinite supply of encouragement.

I also need to thank Pat Jones in the Education College for her work as an expert. She passed along excellent constructive ideas and criticisms about the knowledge base and its format.

This dissertation and degree would not have been completed without the cooperation and encouragement of the other doctoral students in the MIS program. I thank them for their time as subjects and for their numerous discussions on the whole idea of expert systems, decision support systems, and knowledge-based systems. I especially appreciate the support of Mari Heltne, Lynda Applegate, and Titus Purdin. I hope the completion of this dissertation means time for camping and expensive wine!

Finally, I have to acknowledge the contributions of the Ft. Collins gang in Colorado: Sharon, Gregg, Kathy, George, and John. I'd never have made it without being able to escape to their borrowed

houses and cars for vacations nor without all the movies, snowstorms, meals and phone calls for encouragement. <u>Now</u>, we can go skiing!

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF ILLUSTRATIONS viii

ABSTRACT x

1. INTRODUCTION 1

Importance of Research 6
Problem Boundaries 9
Research Process 11

2. REVIEW OF PREVIOUS WORK IN THE FIELD 14

Statistical Education Literature 14
Statistical Taxonomy Literature 19
Expert/Novice Literature 27

3. METHODOLOGY 33

Building and Validating The Knowledge Base 33
Building the Consultant System 40
Testing Procedure for the Consultant 41
Pilot Study Results 46

4. RESULTS OF THE KNOWLEDGE BASE AND SYSTEM DEVELOPMENT 48

The Knowledge Base 48
Knowledge Base Description 50
Rule Set Description 51
Pictorial Representation of the Knowledge Base 58
Representing the Knowledge Base in the Consultant 59

Consultant System Description 60
System Structure 60
Consultant System Logic 66

TABLE OF CONTENTS - continued

Page

5. EXPERIMENTAL TEST RESULTS 68

Undergraduate Test Group Results 68
Undergraduate Experiment Observations 75
Graduate Test Group Results 77

6. SUMMARY OF RESEARCH RESULTS 85

Contributions 85
Further Research Ideas 89

APPENDIX A: INITIAL FRAMEWORK 92

APPENDIX B: FRAMEWORK INCORPORATING FIRST EXPERT'S KNOWLEDGE 110

APPENDIX C: RULE BASE 122

APPENDIX D: INSTRUCTIONS TO USERS 141

APPENDIX E: PROBLEMS USED IN THE EXPERIMENT 145

LIST OF REFERENCES 151

LIST OF TABLES

Table Page

1. Data for Measuring the Proportion of Correct Answers - Undergraduates 69

2. Data for Measuring the Average Number of Correct Answers - Undergraduates 70

3. Chi-square, Measuring the Independence of Correct Answers and Method - Undergradute 72

4. Data for Measuring the Average Time to Solve Problems - Undergraduates 73

5. Summary of the Number of Statistics Courses Taken by Doctoral Students Who Participated in the Experiment 79

6. Data for Measuring the Proportion of Correct Answers - Doctoral 79

7. Data for Measuring the Average Number of Correct Answers - Doctoral 81

8. Data for Measuring the Average Time to Solve Problems - Doctoral 81

9. Chi-square, Measuring the Independence of Correct Answers and Method - Doctoral 83

LIST OF ILLUSTRATIONS

Figure Page

1. Types of Independent and Dependent Variables and Methods Appropriate to Their Analysis by Morris and Rolph 20

2. Decision Tree for the Degree of Relationship Among Variables by Tabachnick and Fidell 21

3. Decision Tree for Significance of Group Differences Among Variables by Tabachnick and Fidell 22

4. Decision Tree for Prediction of Group Membership by Tabachnick and Fidell 23

5. Decision Tree for Structure of Variables by Tabachnick and Fidell 24

6. Chart of Non-parametric Statistics by Siegel 26

7. Initial Framework Used for Classification of Statistics 35

8. Statistical Techniques Classified in Pass 1 of the Taxonomy 39

9. Diagram of Rule Construction in Table Format 49

10. Condition Types in Rule Set 1 (Relationships) 53

11. Condition Types in Rule Set 2 (Group Differences) 53

12. Condition Types in Rule Set 3 (Prediction) 55

13. Condition Types in Rule Set 4 (Inferences from a sample) 57

14. Condition Types in Rule Set 5 (Structure) 57

15. Consultant System Instruction Screen 63

16. Example of a Detailed Question Screen 64

ABSTRACT

The statistical mentor system incorporates a knowledge base into an educational tool for novices in statistical decision making to use in choosing a statistical technique. The novices are students in a business school curriculum who are expected to learn the basic statistical processes in business applications. The purpose of the system is to stimulate learning of the data analysis process on the part of the novice, usually a difficult task. The system acts as a consultant to the novice and approaches the task using a top-down problem solving strategy rather than the traditional bottom-up strategy used by novices.

The heart of the system is the rule base for differentiating between statistics. These rules were built by gathering expertise from two experts in statistical analysis. The rules are based on five questions which the data can answer, as well as the type of data, the number of variables, and any dependent/independent relationships which exist between the variables. The knowledge base consists of five rule sets and can be represented either by condition/conclusion rules or by a set of multi-dimensional tables. Twenty-nine statistics and the rules for choosing them are in the rules sets. The knowledge base was used to define the logic incorporated in the consultant system in order to aid the user in selecting a correct technique. A dialogue

mode is employed in the consultant to determine which conditions are true for the problem and data set. The rule sets are then checked to find the conclusion satisfying the conditions.

The computer mentor was tested against the usual textbook mentor method (search through a textbook until one finds a statistic that looks promising) with two different groups of subjects, 25 undergraduates and 19 doctoral students. The results were that the computer-assisted students in both samples correctly solved a larger proportion of problems and had a higher average number of problems correct than did the textbook assisted groups.

CHAPTER 1

INTRODUCTION

Students in an undergraduate business curriculum are required to include courses in statistics in their programs of study. These courses usually cover material in regression, analysis of variance, hypothesis testing, and making inferences about a population from a sample. Included in the discussion of each topic are the assumptions and theory underlying the model, the procedures needed, and the objectives. Having completed such a course, students often have a grasp of each individual technique but do not have a full understanding of which technique or techniques to choose for particular problems or experimental designs. They are novices in data analysis. Faced with problems to solve or the need to use statistics in another class, the students are lost. While they were enrolled in a statistics class, the students knew which section or chapter of the book a problem came from and used this as their frame of reference for choosing a technique. Without a frame of reference to apply to the problem, they usually have no idea what the appropriate technique is and, worse, no idea how to go about deciding upon a technique.

The student has acquired some declarative knowledge (facts and formulas) in class but has acquired little or no procedural knowledge. They do not know how to go about selecting a technique. They also do not have any conditional knowledge about when a technique is applicable.The main reason for this lack of conditional knowledge is because the conditions of applicability are not taught and emphasized in most statistics courses.

From observation and personal experience, students have several options when searching for a solution technique to a statistics problem outside of class. First, they may choose a reference book on statistics and compare the problem to be solved with solved problems in the book and in this way choose a method. The student could also talk to other students or professors if they are available. Alternatively, the student may use the data in the problem, choose a likely equation from a statistics book and apply the data to the variables in the equation until all of the data are accounted for and all of the variables have values. When all the variables have values, the student assumes that the correct technique has been chosen and goes on to perform the calculations called for in the problem. She fails to recognize that the same data can be used in different statistical formulas to test the different hypotheses. This is especially a problem when the student has extraneous data that is not relevant to the correct statistical technique.

Observing and questioning statistics students indicates that the last option is the approach they use most often when solving statistics problems. This is the same process described earlier by Chi, Glaser, and Rees (1982) for physics students. While this search technique used by statistics students can provide an acceptable solution to the statistics problems they encounter outside of the classroom, there is not a strong likelihood that the student will employ a process which guarantees the selection of a correct statistical technique.

The data analysis process is often a hit or miss proposition for a novice. The "goodness" of the data analysis is dependent upon the "goodness" of the expertise (books, old problems, fellow students, professors) used in each stage of the process. At best, the search for a problem solution is haphazard and inefficient.

The computer packages currently on the market do not help to solve this problem either. The packages, such as SPSS, SAS, BMDP, and MINITAB, can perform extensive calculations quickly and accurately but they do not offer any selection assistance to the student. These packages are primarily command driven. They require that the analyst learn and use a prescribed syntax in order to generate any statistics. The user must determine for herself what particular analysis she wishes to perform. The packages do not guide the user to the appropriate analytical techniques nor do they suggest a particular sequence of tests to be performed. The packages assume that the student possesses all of the knowledge needed. The packages do not attempt to impart new knowledge to the user.

When the student uses the common commercial statistical packages, she must use the accompanying manuals and references to decipher the meaning of the data generated by the packages. Little or no explanation of the results is offered by the package itself. If the analyst decides that incorrect procedures were performed on the data, she must begin the entire process over again: decide on new analysis methods, execute them via the commercial packages, and evaluate the results. The packages do not guide the analyst to more appropriate procedures based upon the results received in previous tests.

If the reasoning processes (the procedural and conditional knowledge) used by experts in choosing statistical techniques could be captured and systematized, then this knowledge could be used effectively to guide and teach students and to improve the "goodness" of their analysis. This expertise could be stored in a knowledge base and used to generate a consulting service to the student and to support the student's statistical problem-solving task. The first research question is suggested by the above discussion:

"What are the specific reasoning steps used by the experts to choose statistical techniques when faced with different types of data and different questions to be answered by that data?"

Once the expertise has been captured, the next step in the consulting process is to disseminate the experts' knowledge. Since the computational steps in statistical analysis have been automated it seems reasonable to propose that an attempt be made to computerize the procedural knowledge. An appropriate framework should be designed to combine the computational and logical ability of the computer with the knowledge base provided by the experts to create a system which can serve as a consultant to the student. By incorporating the strategies and conditional knowledge used by the experts, in deciding what statistical techniques to use, this framework could also help increase the students' schematic and strategic knowledge.

The second research question that this suggests is

"What is an appropriate framework for disseminating the experts' knowledge to the novice in a consultative mode?"

In other words, the research must establish an appropriate design for disseminating the expert knowledge which has been captured in order to guide the student through the statistical reasoning process.

The next section of this paper will explore the importance of these two research questions in more depth.

IMPORTANCE OF RESEARCH

The first question addressed in the research,

> "What are the specific reasoning steps used by experts to choose statistical techniques when faced with different types of data and different questions to be answered by that data?",

is concerned with identifying specific reasoning processes used by experts in determining which methods to employ in data analysis. The elicitation of these expert processes is necessary if a consultant system is to be designed to aid students in developing new skills and new mental models of the problem analysis and statistical reasoning process. Human experts are capable of outstanding performance because they are knowledgeable and have well-developed reasoning processes in their field. Most importantly, the expert organizes the knowledge in a more useful fashion than does the novice and uses this organization to retrieve the procedures and information needed to solve problems. In this particular case, the expert has a thorough knowledge of the techniques as well as the theory behind the techniques. The expert also has extensive experience in actually performing the data analysis and evaluating the results. The expert's organization of statistical knowledge allows her to eliminate large groups of techniques at one time and to narrow down her choices. If this expert knowledge and experience can be captured, then private knowledge becomes public and can be used to teach novices the processes used by experts. This should help the novice develop procedural knowledge as well as conditional knowledge in statistics.

The second question addressed in the research is

"What is an appropriate framework for disseminating the expert's knowledge to the novice in a consultative mode?".

In other words, how can a novice begin to learn these thought processes and at the same time decide upon an appropriate statistical technique to solve her problem?

Traditionally, a bottom-up approach has been used to teach statistics to students and this bottom-up approach simply re-emphasizes the fill-in-the-blank approach that novices use naturally. (Larkin, 1981) They learn the statistical techniques and the arithmetic calculations in the techniques but most of the students do not learn any formal processes to employ when a choice must be made between statistical techniques. This failure to organize the knowledge they have about the techniques is not necessarily the novices' fault, however. In most cases, the rules needed for using the knowledge are not taught in statistics classes.

Using a top-down approach, classes of operations can be eliminated from the statistical reasoning model in several large pieces. The first cut is made by determining the information the student is trying to obtain from the data. By determining the statistical question being asked about the data, large numbers of statistics can be eliminated because they cannot be used to answer that particular statistical question. Additional classes of statistics can be eliminated by determining data type because the statistics only apply to certain types of data (ordinal, categorical, dichotomous, etc.).

An appropriate framework for a consultant system for students then is one which helps students select statistics by leading them through processes and questions used by experts in selecting an appropriate statistical technique in a top-down fashion. One approach for accomplishing this would be to use a strategy of applying questions from a rule base. The system would propose questions and then use the student's answers to these questions to eliminate possible techniques and to reduce the set of correct techniques to a small subset. This does not imply that the system will make all the decisions. Rather, the consultant system can guide the student through a correct selection strategy. The consultant can use a top down approach to mimic the organization of the knowledge employed by experts. The system should provide guidance on the assumptions underlying each technique and should provide ways to test whether the assumptions are met. Each of these tests, however, will have to conducted by the novice himself. After the technique is chosen, the novice could use any available statistical package to do the calculations on the data.

To summarize, the second objective of this research is the design of a statistical consultant which should accomplish three tasks:

1. Act as a consultant to the novice in the reasoning process in the data analysis problem;
2. Show the student appropriate models or schemata of the analysis process;

3. Lead the student to an appropriate statistical technique to answer her questions with her data.

Problem Boundaries

Obviously, hundreds of statistics are available to be used for analysis. Many of these are highly specialized statistics and apply to only one problem domain or one very special type of problem. In order to make this research tractable, problem boundaries must be set. The statistics to be included in the consultant will be limited to techniques generally included in statistics courses for undergraduate students in a business curriculum. These statistics are:

- o correlations
- o ANOVA
- o regression - simple, multiple, polynomial
- o statistics for making inferences about a population, such as t-test, z-test, chi-square, and F-test

Additionally, several techniques taught in advanced statistics courses will also be included in the consultant system. They are:

- o factor analysis
- o principal components
- o canonical correlation
- o discriminant analysis
- o MANOVA, ANCOVA, MANCOVA

The research must also be bounded by the intended audience or users of the consultant system. The system will be designed as a consultant for students who have had at least one introductory statistics course. No assumption is made that the user has had more than one statistics course.

Another boundary on the research is that the consultant system will not be designed to analyze the data nor is it intended to be an expert system, although it will use some ideas from expert systems, particularly the rule base. Rather, the system is intended to be a consultant to the student and will guide the student through the process of choosing a technique for herself, using a top-down approach. Included in the system should be the assumptions underlying the techniques. The consultant system should also suggest alternate approaches if the student's answers lead to dead ends, particularly alternate statistical questions the student may want to explore with her data. By leading the novice through the decision process, the system will be showing the student a top-down approach to the reasoning processes for selecting statistics by using problems selected by the student.

The final boundary placed on the research problem is that the consultant system must be able to run on the IBM PC family to be of maximum use to the students.

Research Process

The first step in the research will be to build a framework to be used as the core for the knowledge base of statistical techniques. An initial framework can be developed through the use of the current literature and textbooks (Andrews et al., 1974; Anderberg 1973; Harris 1975; Morris and Rolph 1981; Neter, Wasserman, and Kutner 1983; Siegel 1956; Tabachnick and Fidell 1983). This framework, including the statistical techniques, can be taken to experts to check for correctness and completeness. They are asked to make appropriate changes to both the structure and the classifications of statistics in order to make the framework more closely match their own decision processes. The experts can also make notes about limitations and assumptions of each method to clarify the boundaries on the subsets of methods. Several iterations will, obviously, be needed.

After the framework has been completed and the statistical techniques classified, this knowledge base will be transformed into a set of rules consisting of conditions and action or conclusions. These rules must be put in a form which allows them to be incorporated into a computerized consultant system.

Before the consultant system can be implemented, the components of the consultant must be determined. Obviously, the rule base must be included. However, other items are also necessary for the system to be of help to the students. These consist of such modules as information to support and explain each statistical question, definitions and examples of the various statistical terms, as well as definitions and examples of the different types of data a student could encounter in her problems. An appropriate form for the function of the system should be considered. This could include menu driven, function key driven, or question and answer format.

After the system has been designed, coded, implemented and debugged, it must be validated. Validation will be a three-step process: First, the consultant will be validated from an MIS standpoint. Does the system work functionally? Are the function keys correctly defined? Are the screens easy to read? This validation will be performed by MIS graduate students using sample problems.

The system will also be tested for effectiveness and efficiency using students. Two different groups of users will be tested, undergraduate students and doctoral students from the MIS department. The students will be given textbook problems unidentified by chapter or section number. One group of students will be asked to determine a statistical technique for solving the problem by using the consultant. The other group of students will choose a technique using the classical approach, textbook search. Effectiveness is measured by the number of correct techniques chosen. Efficiency is measured by the number of minutes it takes to find the solution. In this case, effectiveness of the system is the more important of the two measures because selecting a correct technique is more important than selecting a technique quickly, particularly if the quick selection process yields generally incorrect techniques. Only if the two methods of selection, textbook and computer, are equally effective will the efficiency be an important difference.

CHAPTER 2

REVIEW OF PREVIOUS WORK IN THE FIELD

This literature review consists of hree sections. The first section discusses previous research in the statistics education and omputing field. The second section reviews the statistics taxonomy literature and the third section reviews research about the differences between experts and novices.

Statistical Education Literature

The use of computers in statistics education began as early as the 1960's when Cramer and Cady (1969) proposed the use of a computer to generate unique data sets for statistics courses. They argued for the use of the data sets to make student analysis projects unique. Beaujon (1970) developed a display package to graphically display well-known discrete and continuous probability distributions. He used the power of the computer to pictorially demonstrate the concept of distributions and how they change when parameters are changed.

In 1971, Doane, Shapiro, and Young developed a package including different statistical applications for use in the classroom. These applications were built solely to ease the burden of computation on the student. (Young 1982) Minitab was introduced in 1973 at Pennsylvania State University and is command driven. If a user does not know what statistical technique to use, no help is available from the system for choosing a technique. The idea of giving advice is not built into the system.

In the late 1970's more work was presented in statistics education utilizing the computer for problem solving. Most of this work used the techniques of graphical displays, computational aids, simulations of distributions, drill-and-practice exercises, and computer-aided instruction. (Gentleman 1977; Novick, Hamer, and Chen 1979; Caldwell 1983; Dawes 1983; Southward, Urquhart, and Ortiz 1983) Meeker, Hahn, and Feder (1975) built a system, EXPLOR, whose purpose was to evaluate and compare the properties of proposed experimental designs. All of the input was in matrix form (0, 1, -1 input) and no tutorial help was offered for choosing a design. The output of the system consisted of such items as standard error of confidence intervals for each coefficient in the model, matrix of simple correlations between pairs of terms, and variance-covariance matrix.

Even in the early 1980's the computer systems presented at the statistics education meetings of the American Statistical Association simply aided computation. Southward and others (1983) developed a system to generate specialized data sets for each student based upon the student's area of interest. The computational module of the system restricts the menu selections to only those covered in class. No guidance is available from the system to help the novice choose a technique, however, after all of the modules have been covered in class.

CADA (Computer-Assisted Data Analysis) was designed by the University of Iowa as a monitoring system

> that does all the arithmetic, and, even further, guarantees that all of the steps in the analysis are performed correctly and in their proper sequence. This sequential guidance is useful in teaching students by directing their steps through valid statistical analyses even if they do not yet understand fully what they are doing. (Novick 1983, p.33)

CADA does not offer guidance in choosing the correct technique to use. The user must choose the technique before CADA will guarantee the correct analysis.

Emond (1982) developed interactive graphics programs to demonstrate the significance of regression parameter values. As the student changes the parameters the estimated line of best fit changes through the sample of points on the screen. The system will also display confidence bands for the true regression line.

The TREE system (Fertuck 1981) incorporates 'HELP' features but in no way leads a novice to the selection of a correct statistical technique. The 'HELP' feature simply explains the requested technique in textbook terms. It does not help the novice distinguish between techniques.

Little attempt is being made to incorporate the ideas of expertise and artificial intelligence from the MIS field into statistics' education packages. AI-based tutorials have been developed in areas such as geography, electronics, medicine, and mathematics. Commonly known as Intelligent Tutoring Systems, these systems in other fields "... attempt to combine the problem-solving experience and motivation of 'discovery' learning with the effective guidance of tutorial interactions." (Sleeman and Brown 1982, p.1)

In most of these tutorial systems, the student is given information only after mistakes are made in answering questions. (Stevens, Collins, and Goldin 1982) Some systems even wait for patterns of mistakes to evolve. These programs may even generate new problems for the student to solve in order to discover the pattern of mistakes. (Burton 1982) Other research in this area of artificial intelligence is concerned with developing models of the students and their knowledge structures. In this work, a network of nodes and links is used to store skills of students and relationships between those skills. (Goldstein 1982; Sleeman 1982) None of these tutorials, however, help a student solve a student-selected problem. Instead, these tutorials present the student with problems to be solved and allow them to explore different solutions in order to learn. If a student has a problem he wishes to solve, he cannot use the intelligence of these tutors to find a solution.

Some work has been started in the area of combining artificial intelligence and statistics applications. Bell Laboratories built REX, a regression expert system. REX employs a frame-based system to check assumptions in the initial model by examining the data and assessing the adequacy of the model. REX does little tutoring for the novice user, however. (Gale and Pregibon 1982) O'Keefe built ASA which helps a user analyze designed experiments not been performed. Other systems have been developed to assist users in developing hypotheses and models. Roos' system guides economists in the construction of econometric models. Hayed and Havranek are working in the area of automatic hypothesis generation. None of these systems yet act as an intelligent interface to a statistical package for a novice user. (Gale and Pregibon 1985)

Some specialized statistical applications have been developed using artificial intelligence techniques. RX is a system for "discovery, confirmation, and incorporation of causal relationships from a large time-oriented clinical data base". (Blum 1982, p.31) Other systems help a designer choose sample sizes easily or automatically build a Box-Jenkins model and forecast. (Hahn 1985)

Statistical Taxonomy Literature

This section includes a review of taxonomies developed for classifying statistical techniques and the conditions under which the techniques should be used.

Harris (1975) developed a table relating the number and types of predictor variables and the number of outcome variables with multivariate statistical techniques. The taxonomy does not relate a technique to the question which is answered. The techniques included in the taxonomy are t-test, ANOVA, correlation, bivariate regression, multiple regression, ANCOVA, Hotelling's T-squared, discriminant analysis, MANOVA, MANCOVA, canonical correlation, principal components, and factor analysis.

Tabachnick and Fidell (1983) present multivariate statistics in relationship to the types of research questions that can be answered by the statistics. Their classification uses four questions. The questions are:

1. strength of relationship between two variables;
2. measuring the significance of group differences;
3. predicting group membership;
4. determining structure. (Tabachnick and Fidell 1983, pp.13-15)

A decision tree using the four research questions, the number of dependent variables, and the number of independent variables to choose an analytic technique is developed [Figures 2,3,4,5] The techniques included in the tree are bivariate-r, multiple-r, canonical correlation, ANOVA, ANCOVA, MANOVA, MANCOVA, Hotelling's T-squared, t-test, discriminant analysis, principal components, and factor analysis.

Morris and Rolph (1981) develop a table of techniques relating the type of independent variable with either continuous or binary dependent variables. The techniques included in the table are discriminant analysis, maximum likelihood logit, minimum logit, and contingency tables. [Figure 1]

	Independent Variables		
	Continuous	Mixed	Categorical
Continuous Dependent Variable	Regression	ANCOVA	ANOVA
Binary Dependent Variable	Discriminant Analysis Max. Likelihood Logit	Maximum likelihood Logit	Minimum Logit Max. Likelihood Contingency Table

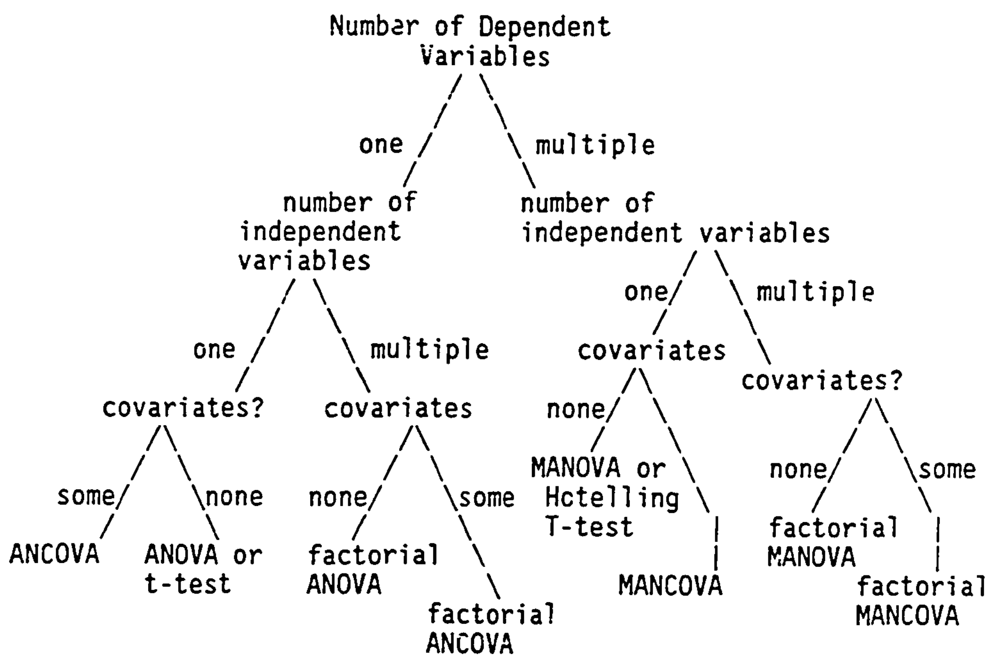
Types of Independent & Dependent Variables
and Methods Appropriate to Their Analysis
by Morris and Rolph (1981, p.8)

Figure 1

```
                    Number of Dependent
                         Variables
                            /\
                           /  \
                   one    /    \  multiple
                         /      \
                 Number of       Number of Independent
               Independent          Variables?
             Variables?                 \
                 /\                      \
                /  \                      \ multiple
          one  /    \ multiple             \
              /      \                 CANONICAL R
     BIVARIATE R   Covariates?
                       /\
                      /  \
              none   /    \  some
                    /      \
          MULTIPLE R      HIERARCHICAL
                          MULTIPLE R
```

Decision Tree for the Degree of Relationship Among
Variables by Tabachnick and Fidell
(1983, p.63)

Figure 2



Decision Tree for Significance of Group
Differences Among Variables
by Tabachnick and Fidell
(1983, p. 63)

Figure 3

```
                    Number of Dependent
                         Variables
                          /    \
                         /      \
                 one    /        \  multiple
                       /          \
            number of          number of
            independent        independent
            variables          variables
                |                  |
                |                  |
     multiple   |                  |  multiple
                |                  |
          covariates?        covariates?
            /      \            /     \
           /        \          /       \
     none/           \ some  none/      \some
        /             \        /         \
Discriminant           \  Factorial       \
Analysis                \ Discriminant    Hierarchical
                         \Analysis        Factorial
                          \               Discriminant
           Hierarchical                   Analysis
           Discriminant
           Analysis
```

Decision Tree for Prediction of Group Membership
by Tabachnick and Fidell
(1983, p.63)

Figure 4

```
        Number of Dependent
           Variables
               |
               |
multiple       |
               |
     Number of Independent
           Variables
             /   \
            /     \
  multiple /       \ multiple
          /         \
PRINCIPAL          FACTOR
COMPONENTS         ANALYSIS
(empirical)        (theoretical)
```

Decision Tree for Structure of
Variables by Tabachnick and Fidell
(1983, p.63)

Figure 5

Siegel (1956) develops a taxonomy of techniques for non-parametric statistical techniques in terms of the number and types of variables. The table does not include any information about the questions answered by each technique. [Figure 6]

Andrews and others (1974) develop a decision tree as the basic architecture in their "attempt to record the sequential decisions a social scientist might make in selecting a particular statistic or statistical technique appropriate for a given analysis." (Andrews, et al. 1974, pp.vi) The decision tree includes number of variables, types of variables and some information about the questions answered by the statistics.

None of the manuals accompanying the major statistical packages offer any guidance as to the proper techniques. In fact, the SAS Introductory Guide states "If you don't understand when a given statistical test should be used or how to interpret the results, a good statistics text is a necessity." (SAS Institute 1983, p.1) Other researchers agree with this observation. "... for the novice, current packages are lacking important features. They provide only numbers, not interpretations. They provide no guidance on what to do next or what should have been done before. Moreover, they provide no instruction." (Gale and Pregibon 1985, p.74)

	Nominal	Ordinal	Interval
ONE SAMPLE CASE	Binomial Test Chi-square one sample test	Kolmogorov-Smirnov one-sample test One-sample Runs Test	
TWO SAMPLE CASE			
Related samples	McNemar test for significance of changes	Sign Test Wilcoxon matched pairs signed rank test	Walsh Test Randomization test for matched pairs
Independent	Fisher Exact Probability Test Chi-square test for 2 Independ. samples	Median Test Mann-Whitney U-test K-S 2 Sample Test Wald-Wolfowitz Runs Test Moses Test for Extreme Reactions	Randomization Test for 2 independent samples
K-SAMPLE CASE			
Related samples	Cochran Q-test	Friedman 2-way ANOVA	
Independent samples	Chi-square test for independent samples	Extension of median test Kruskal-Wallis One-way ANOVA	
NON-PARAMETRIC MEASURES OF CORRELATION	Contingency Coefficient	Spearman Rank Coefficient Kendall Rank Correlation Kendall Partial Rank Correlation Kendall Coefficent of Concordance	

Chart of Non-parametric Statistics
by Siegel (1956, inside cover)

Figure 6

Expert/Novice Literature

Because of training and experience, experts are able to do things the rest of us cannot. They are not only proficient but also smooth and efficient in the actions they take. (Johnson 1983) Experts appear to solve problems and make decisions by recognizing situations as instances of things with which they are familiar. (Jeffries, Turner, Polson, and Atwood (1980); Malhotra, Thomas, Carroll, and Miller (1980); and Weber (1985)) In other words, experts solve problems by analogy. Experts also differ from novices in the quality of the organization of their knowledge. The expert possesses more domain specific knowledge in her field of expertise (Simon and Simon, 1979) and has better organized knowledge, allowing a more efficient search for a solution. In a problem-solving situation, experts know of more problem types than do novices and can categorize new problems by their problem type (Mayer, Larkin, and Kadane, 1984)

Feltovich (1981) found that expert heart specialists could keep more than one hypothesis in mind at a time while novices were able to entertain only one active hypothesis at a time. In addition, the experts used all of the information they received in developing these hypotheses but novices discarded the old information as they received new data. Feltovich used these findings as an indicator that experts' knowledge is both interconnected and organized in a hierarchical structure.

Other researchers have found evidence of expert knowledge organizations. Chi, Glaser, and Rees (1982) found that experts use large chunks of information while novices use much smaller chunks. This has also been demonstrated by expert chess players (Chase and Simon, 1973) and electronic designers (Egan and Schwartz, 1979). Experts also solve problems faster. Chi, Glaser, and Rees (1982) found a 4:1 difference between physicists and physics students. The time spent thinking between retrieval of equations and chunks of equations was shorter for experts. They also found that experts have several different methods available to solve problems so they are able to double check their solutions. Interestingly, experts employ a forward strategy for problem solving while novices use a backward strategy. The novice will pick a likely equation and try to make the variables fit the equation. The expert will use the variables to build the equation. Larkin's (1981) work agreed with this. Novices possess a list of fragmented principles and work from the givens to the goal. Experts on the other hand worked from goals to the given with larger units of knowledge.

Rumelhart and Norman (1981) define three stages of learning. The first stage of learning is accretion in which new information is encoded in terms of existing schemata or models. In the structuring stage new schemata are created. This occurs only when an old schema or model is found to be incorrect or new insights are gained into the way the knowledge should be organized. In the third stage, tuning, schemata are modified and refined, depending upon the ways the schemata are applied. This can only happen with experience.

Norman (1982) uses this model of accretion, structuring, and tuning to explain the shift from novice to expert. In accretion, the novice adds new knowledge to existing knowledge structures. After the novice works with these old structures long enough, she may gain new insights into the ways that the knowledge should be organized and move to the structuring phase in which new knowledge structures or schemata are created. Finally, through fine adjustments to these knowledge structures, a novice moves closer to the expert level. This is the tuning stage, the slowest mode of learning but a necessary process for changing knowledge into expert performance.

Rumelhart and Norman (1981) showed that experts possess and choose an appropriate model for a situation while novices rely on the 'best' model they have if given no other guidance, even if this model may not be completely applicable. They also found that the "different kinds of mental models that students bring to the situation clearly plays a critical role in the kinds of analogies students will employ. It is a far more important role than that of the formal instruction received." (Rumelhart and Norman 1981, p.355) They found this to be true with students learning a new computer language. If a student did not possess models from previous experience on which to draw, she created her own models, often erroneously.

Three types of knowledge have been defined: declarative, conditional, and procedural (Gagne, 1985). Declarative knowledge consists of facts and theories in a subject area. It is the knowledge <u>that</u> something is the case. Conditional knowledge is the understanding of the conditions of applicability of the declarative knowledge. Procedural knowledge consists of the strategies and heuristics of <u>how</u> to do something.

Another major difference between experts and novices is that experts possess more procedural and conditional knowledge. Chi, Glaser, and Rees (1982) reported that in experiments with physics students only the experts mentioned any conditions for applicability of different solutions in their protocols. Without the knowledge of <u>when</u> a particular solution is applicable, the novices' problem solving skills will not reach the expert level. By developing the conditional knowledge, persons move from novice to expert. Larkin, McDermott, Simon, and Simon (1980), again working with physics students, agree that experts possess solution strategies (procedural knowledge) that can be applied to their domain specific knowledge.

As Larkin (1983) found when working with physics students and physicists, the experts used a physical representation of the problems they were solving while novices used a naive representation. In the physical representation, experts used entities which have meaning only in the context of formal physics (eg. torque). The naive representation, however, uses entities that can be seen in some visible sense. (eg. screwdriver). As Gagne (1985, p.139) explains. "The experts' representation goes quickly to fundamental principles, whereas the novice represents the problem in terms of superficial, but perceptually salient, attributes."

Wiser and Carey also experimented with physics problems in order to try to define the shift from novice to expert. Their definition involves three shifts: "a shift from one system of beliefs about the physical world to another, one set of concepts to another, one set of problem solving capabilities to another". (Wiser and Carey 1983, p.267)

These theories can be applied to statistics students. Novices in a statistics class add data and information about statistical techniques to existing schemata (accretion). They are adding to their declarative knowledge. As they solve homework problems, they may begin to slowly develop new structures in which to store the statistical techniques. However, in most classes, students do not form new structures because they do not possess the conditional or procedural knowledge needed to form new schemata. The students are not taught what sort of data is appropriate for each technique nor are they taught any strategies for choosing between techniques. They also have difficulty developing any new structures because of a lack of organization in their knowledge.

In summary, the differences between novices and experts have been found to be:

1. Experts use different strategies for solving problems (forward instead of backward searches).
2. Experts bring larger amounts of knowledge (chunks) into short term memory.

3. Experts possess a better knowledge organization, many times in a hierarchical structure.

4. Experts possess more domain specific knowledge.

5. Experts are thought to store knowledge in knowledge structures or schemata containing procedural knowledge with explicit conditions for applicability. Novices lack the procedural knowledge.

6. Experts are able to infer principles from the cues in the problem statement.

CHAPTER 3

METHODOLOGY

The first section of the methodology is concerned with identifying a specific taxonomy of statistical techniques used in analysis of problems and decision-making. The second section explains building the consultant system and the third section discusses testing and validating the consultant system.

Building and Validating The Knowledge Base

Statistical techniques used in research are numerous, many of them specific to problems in particular subject areas. In order to make the design of a taxonomy into a tractable problem, the types of statistical techniques must be narrowed down. The subset of techniques employed in this research is comprised mainly of parametric (versus non-parametric), cross-sectional (versus time-series) statistical techniques where the assumption of normality is needed or is of no consequence by reason of the design of the technique. The initial subset is limited to those statistical techniques of particular use to the analyst in business and the social sciences and which, consequently, should be learned by students in a business curriculum.

Constructing the knowledge base was an iterative process. Two persons recognized as skillful statistical analysts acted as experts. The first step in the construction was a search of the statistical literature. Using information gathered from published sources, an initial decision tree was built. (Siegel 1956; Andrews, and others 1974; Harris 1975; Tabachnick and Fidell 1983; Morris and Rolph 1981) However, the decision points chosen for the tree did not allow the entire initial subset of statistics to be classified, including factor analysis and principal components analysis.

The second taxonomy [Figure 7] that was developed used a 4-dimensional table. This taxonomy differed from the first by adding a dimension for the questions which form the basis of the statistics used in business and, consequently, should be learned by students. The questions were narrowed down to five general ones:

1. What is the strength or degree of relationship, if any, among the variables?

2. Are there any differences between groups which are subjected to different treatments?

3. Can predictions of individual values or group membership be made?

4. Is there any pattern or structure to a large data set?

5. Are inferences about a population based on a sample needed?

After these questions were determined, the statistics used to find answers to these questions were needed. An initial cut at determining these statistics was made by again consulting multiple

QUESTION: Strength of relationship among variables

No distinction made between independent (predictor) and dependent variables	Quantitative	
	Qualitative	
	Mixed types	
Single Variable		
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Initial Framework Used for Classification of Statistics

Figure 7

Independent Variable (Predictor)	Dependent	
>1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Figure 7 - continued

Independent Variable (Predictor)	Dependent	
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Figure 7 - continued

references. (Andrews and others 1974; Anderberg 1971) In some cases a table of rules was found for selecting between statistics. (Siegel 1956; Tabachnick and Fidell 1983; Morris and Rolpn 1981) In each of these tables, the decision factors were types of data to be analyzed and the number of variables being analyzed. Twenty-six different statistical techniques were initially classified. [Figure 8] The statistical techniques were categorized by the types of data for which they were applicable: independent or dependent data, qualitative, quantitative, or mixed (both quantitative and qualitative) groups of data; and the statistical questions they could be used to answer. The three dimensions (question, independent vs. dependent, and data type) were used to build a knowledge base in the form of a table. The statistics were filled in the diagram [Appendix A] and it was taken to the first expert.

The first expert filled in many of the missing statistics in the table. She observed that the original dimensions and categories in the table were not appropriate for all questions. A finer breakdown is needed for some questions while a less fine-grained breakdown is useful for other questions. In particular, when examining relationships between variables, differentiating between independent and dependent variables is not always necessary. However, the types of data do need to be well-defined. They should be categorized as interval, ordinal, dichotomous, dichotomous with an underlying normal, and categorical with more than 2 categories rather than simply quantitative or qualitative.

- ANOVA - single factor and multi-factor
- MANOVA - single factor and multi-factor
- simple linear regression
- multiple linear regression
- polynomial regression
- maximum likelihood logit
- minimum logit chi-squared
- r-squared - simple and multiple
- correlation ratio (eta)
- tetrachoric correlation
- contingency correlation
- Phi coefficient
- biserial r
- point biserial r
- r-squared using logit approximation
- canonical correlation
- factor analysis
- principal components analysis
- t-test
- z-test
- binomial test
- chi-squared
- Hotelling's T-test
- discriminant analysis
- cluster analysis
- non-metric multi-dimensional scaling

Statistical Techniques Classified in
Pass 1 of the Taxonomy

Figure 8

The second iteration of the knowledge base construction used the input from the first expert and reconstructed the tables. Obviously, one table format was not appropriate for all five questions. Individual tables were built for each of the five questions and extensive modifications were made. Some new categories were constructed. Others were eliminated. The tables were fine-tuned to match each question. [Appendix B]

The second knowledge base was examined by another expert. Only minor modifications were made to this version. One missing category and three new statistics (Bartlett's, Hartley's, and Kolmogorov-Smirnov tests) were added to the tables.

Building the Consultant System

The knowledge base table was converted into a set of rules before the knowledge base was incorporated into a computerized mentor system. [Appendix C] The rules are all in the condition/action form. More specifically they are all in the (IF condition is true, THEN action) format. The rules provided the foundation for the logic of leading the student user through the necessary steps to determine a correct statistical technique or set of techniques, based upon the statistical question she is trying to answer about the data.

Standard system design methodology was followed in the construction of the mentor computer system. Screens were designed for presenting each of the statistics in the knowledge base. Other screens were designed to handle question presentation, data entry and data correction, definitions, and help facilities. The logic to tie these screens together was formulated and the system was programmed and tested in pieces. As each section was tested and debugged, another section was added to the system.

Testing Procedure for the Consultant

Several preliminary steps for validation were taken. Three persons experienced in statistical analysis and in MIS used the system and critiqued the structure, content, and layout of the program and screens. One person from the same population as the formal test group (business school senior, one statistics course completed, at least one micro-computer course completed) also tested the system. He solved the same set of problems later used in the formal validation and testing procedure. This helped to assure that the questions were clear and understandable before being assigned to the students. One person inexperienced in statistical analysis (one introductory statistics course) also used the system to critique it for the clarity of explanations, definitions, and questions. This preliminary validation was an iterative process. Based upon each set of comments and suggestions, revisions were made to the consultant system.

The formal validation procedure consists of a test for effectiveness of the system using novice analysts. The purpose of the effectiveness test is to compare the effectiveness of the computerized statistical analysis consultant with the effectiveness of the traditional textbook approach in choosing correct statistical techniques to analyze data and answer questions about the data. At the same time, the efficiency of each approach (textbook versus consultant) in terms of the time required to select a correct technique was also measured.

Two groups were used to test the effectiveness of the consultant system against the traditional textbook approach to statistical problem solving. The first group was comprised of junior and senior business administration majors at the University of Arizona who had passed at least one statistics course with a grade of A, B, or C and were currently enrolled in a second statistics course in the College of Business. Each of these students had also taken at least one MIS course in which they used micro-computers. This experience should help to insure that fear of the computer was not a factor in the students' performance and choice of a statistical technique. Twenty-five subjects were chosen and randomly assigned to one of two groups, computer or textbook mentor groups. Twelve persons were assigned to the computer assisted group and thirteen students were randomly assigned to the textbook assisted group. The undergraduate group of students were offered the equivalent of two homework assignments for extra credit in their statistics course to participate. They were also told that a $10 reward would be given to the person in each group (computer or textbook) who found the most correct solutions (techniques) in the shortest amount of time.

The second sample was comprised of nineteen doctoral students enrolled in the MIS Ph.D. program at the University of Arizona. Each of the students had taken at least one statistics course but the time since these courses had been taken ranged from ten years to one semester. These students were also randomly assigned to each of two groups, ten to the computer assisted group and nine to the textbook assisted group. The small sample size was due to the limited population of Ph.D. students.

Two distinct sets of experiments were run, one with the undergraduate sample and one with the graduate student sample. The students in the computer mentor group were divided into smaller groups of three or four in order to be accommodated more easily in the computer lab. Each student was assigned an IBM PC microcomputer. They were positioned such that they could not see each other's terminals. Directions for the use of the computer and the mentor system were given. [Appendix D] The students were directed to use the computer mentor to choose a statistical technique to analyze each of the problems they would be given. They were not to do any calculations or find actual numerical solutions to the problems.

The students in the textbook mentor group were also divided into smaller groups of three or four. Each student had the introductory statistics textbook used in her introductory statistics class available as a reference. The students were seated in different parts of the room so they could not see each other's work. Directions were given to the students. [Appendix D] They were asked to determine the statistical technique which could be used to solve the problems they would be given. They were told to use the textbook as a reference as much as needed. No actual calculations were to be done.

All students were given the same five problems of types covered in introductory statistics classes (correlation, regression, t-test) and a fifth problem (ANOVA) typically covered in their second statistics class [Appendix E]. The problems came from a standard textbook except that data such as the mean, standard deviation, and variance were removed from each problem. The problems indicated that the data had been collected but the actual values were not given to the students. The choice of questions was not biased toward the computer assisted group. The techniques in the computer consultant system are not a subset of the textbook techniques which would force the textbook people to choose from a larger set of statistics and, therefore, have a more difficult task. In fact, the computer system contains a larger number of techniques which should, theoretically, give the computer group the more difficult task.

The problems were handed out and solved one at a time using the respective computer consultant or textbook and the time to find a solution was measured for each problem for each student. When a student knew which technique she wanted to use, she wrote the answer down on a supplied form. The students were also asked to mark which of the five statistical questions they were trying to answer about the problem and the number of variables involved in the problem.

The data used in the analysis was problem number, technique used to solve the problem, and the time required to solve each problem for each student. These data were analyzed in four ways. For each of the five problems, a chi-square test was used to determine whether there was significant difference in the number of correct and incorrect computer and textbook mentor solutions. The hypothesis for effectiveness is:

> H0: Choosing a correct solution is independent of the type of mentor (textbook or computer) used.
>
> H1: Type of mentor and a correct solution are not independent.

Another hypothesis for effectiveness is based on the average number of correct solutions each group obtained.

> H0: The average number of correct solutions per student is the same for both the computer consultant and textbook.
>
> H1: The average number of correct solutions per student is different for the computer assisted group than for the textbook assisted group.

This was tested using a two-sample t-test for small samples.

The third hypothesis was whether the proportion of problems solved correctly was the same for both the computer assisted and textbook assisted group.

> H0: The same proportion of problems was solved correctly for both the computer and textbook assistance.
>
> H1: The computer assisted group solved a different proportion of problems correctly than the textbook assisted group.

The hypothesis can be tested using a chi-squared test.

The time for solving each problem was measured and the average times compared for each method. This was tested using a medians test. The hypothesis is:

> H0: The average time for finding a solution is the same or less for the computer consultant than for the textbook assistance method.
>
> H1: The average time is greater for the computer method than for the textbook method.

The test for a longer time for the computer consultant was made due to the amount of reading required on the consultant method. A learning curve is also involved when using a computer tool for the first time.

Pilot Study Results

A pilot study was conducted with eight students from the target population of undergraduate students to check the procedures for conducting the tests. Four students were assigned to each of two groups, computer-assisted or the traditional textbook-assisted approach. The subjects were all juniors or seniors in MIS who had received either an A or a B in the introductory statistics course required of all business students. Five problems were given to each student in each group. Each student received the same problems in the same order.

The computer assisted group chose 17 out of 20 correct techniques. The textbook assisted group chose 2 out of 20 correct techniques. In the computer assisted group, the first problem took an average of 20.2 minutes to solve. No other problem took the computer-assisted group longer than an average of 12 minutes to reach a decision about a technique. The large difference in times occurred because the students in the computer assisted group used the first problem as a warm-up problem to learn how the consultant system operates and to become comfortable with it. They also used the first problem to become acquainted with the keyboard. All of the subjects had previous experience with a microcomputer but may have never seen the particular keyboard configuration.

The pilot exposed the need to make a few changes in the procedures for the computer assisted treatment. For example, more explicit instructions were needed about the task to be performed. However, five problems seemed to be an appropriate number for the students to solve because the first problem acts as a warm-up problem but even with the extended time for this problem the total time to solve all of the problems was no greater than 60 minutes for the computer assisted group and no greater than 43 minutes for the textbook group. This appeared to be a satisfactory amount of time for the students to work on statistics problems. The most efficient people used 50 minutes and 37 minutes, respectively. When questioned after the pilot, the computer subjects explained that they were comfortable with the five problems and were not bored or tired at the end of the problem solving session. The five problems also allowed five different statistical questions to be included in the experiment. This precluded the possibility that students would be given all problems which they had recently studied in class. The recent classroom experience could act as a reference point and produce correct solutions in both groups. The mixture of problem types helped to assure that techniques would be needed for which the students did not have a recent reference point.

CHAPTER 4

RESULTS OF THE KNOWLEDGE BASE AND SYSTEM DEVELOPMENT

This chapter discusses the results of the research on educational consultant systems for statistics. The first section discusses the knowledge base itself which is the foundation of the consultant system. The second section explains the rule base in detail while the third section describes the actual consultant system.

The Knowledge Base

The knowledge base in the consultant system is essentially a rule-based system. The knowledge is represented by a set of rules, relating five statistical questions, the number of variables in the problem, and the types of data to appropriate statistics. Each rule is composed of a set of conditions (IF clauses) and a conclusion or action based upon these conditions (THEN clause). A rule-based or production system is an appropriate design in this case because the knowledge domain (statistics) can be regarded as a case of a taxonomy. Historically, taxonomies have been represented as rules.

The rules in the knowledge base can be represented as a set of tables, one for each of the statistical questions to be answered.

	a	b	c	d
1	1a	1b	1c	1d
2	2a	2b	2c	2d
3	3a	3b	3c	3d

{a,b,c,d} and {1,2,3} are conditions.

{1a, 1b, 1c, ... 3c, 3d} are the conclusions which satisfy the conditions.

Diagram of Rule Construction in Table Format

Figure 9

Along the perimeter of the tables are the conditions of the rule base. The interior of the tables contains the conclusions. [Figure 9]

The use of tables to represent the rules rather than the IF/THEN rule format proved to be most useful when working with experts. The experts had no problems verifying or understanding the contents of the knowledge base when using a diagram of the rules. The tables allowed missing conclusions and missing conditions to be spotted readily from blank cells in the table. Multiple statistical techniques as conclusions were also easy to detect and suggested that additional conditions were needed to differentiate between techniques.

Knowledge Base Description

The knowledge base is composed of five independent sets of rules which use condition/conclusion or IF/THEN statements. Most of the rules are composed of compound conditions. The first rule always determines which question is to be answered about the data. For example,

IF trying to measure the relationships between variables
THEN ...

IF trying to predict individual values from a set of variables
THEN ...

The five questions in the rule base are:

- o trying to measure relationships;
- o trying to measure differences between groups;
- o trying to predict individual or group values from a set of variables;
- o trying to make inferences about a population based on a sample;
- o trying to identify structure and patterns in the data.

The conclusion to these first rules designates which of the five rule sets is to be checked. The size of the rule sets, including the number and type of the rules to be checked, is dependent upon the statistical question to be answered. Question 5 (structure) needs only five additional rules in order to choose a technique to study structure. Question 1 on the other hand invokes up to 16 additional rules before a statistical technique can be selected. Twelve rules are composed of only two conditions each. For example,

IF variable 1 is ordinal
AND variable 2 is interval
THEN .....

The rule set for statistical question 2, selecting statistical techniques to measure group differences, also consists of 16 rules but each rule has up to four different conditions. For example,

IF number of independent variables > 1
AND number of dependent variables > 1
AND independent variables are both quantitative and qualitative data types
AND dependent variables are both quantitative and qualitative

The condition portions of the rules of the knowledge base belong to three different categories: number of variables, data types of variables, and dependent/independent relationship between variables. Many times the conditions are interconnected. In other words, the conditions could include number and type of dependent variables, or number and type of independent variables. In other rules, only the number of variables and the data types of the variables are important. No dependent/independent relationship condition is necessary.

Rule Set Description

The rules for determining a technique to measure relationships between variables in rule set 1 include a condition about the number of variables being studied. If only one variable is available then no other rules are triggered. Two variables trigger fifteen more rules to be checked with the determining condition being the data type of the two variables in the data set (ordinal, interval, dichotomous, dichotomous with an underlying normal distribution, other categorical). More than two variables triggers only three additional rules to determine a statistical technique. The difference between the number of rules occurs because more well-defined statistics have been developed to measure relationships between two variables than have been developed for more than two variables. For three or more variables the deciding condition simply checks a gross data type (all quantitative, all qualitative, or both qualitative and quantitative) to select a statistic. [Figure 10]

Rule set 2 for determining differences between groups, first checks rules containing compound conditions about the variable type (independent or dependent) and the number of these variables. For example,

```
IF there is one dependent variable
AND there is one independent variable
THEN ...
```

If no distinction can be made between dependent and independent variables, then no more rules are triggered. The conclusion is that the question is inappropriate for the data. With a dependency relationship concluded, more rules are checked based upon the data types (qualitative or quantitative). This piece of the knowledge base is constructed of 25 different rules. [Figure 11]

<u>Two Variables</u>

```
CONDITIONS:        Variable 1 DATA TYPE
                             |
                             |
   Variable 2                V
   DATA TYPE  --------> 25 rules
```

<u>More than two variables</u>

```
DATA TYPE ----------> 3 rules
```

Condition Types in Rule Set 1
(Relationships)

Figure 10

```
                              DEPENDENT Variables
                                   DATA TYPES
                           NUMBER of Dependent Variables
 CONDITIONS:                              |
                                          |
INDEPENDENT Variables,                    |
   DATA TYPES,                            V
NUMBER of Independent  ------------>  25 rules
   Variables
```

Condition Types in Rule Set 2
(Group Differences)

Figure 11

Rule set 3, determining techniques to build predictive models, starts by checking a condition for the type of predictive model, individual or group predictions. The second rule checked in both cases is for a dependency relationship. If no distinction is made between dependent and independent variables, no further rules are triggered. The conclusion is that the data in the problem cannot be used to build a prediction model and the student is asked to reconsider the problem and the statistical question again.

For individual value predictions, a condition is checked for the number of predictor and outcome variables in the data set. The two conditions triggering other rules are [Figure 12]:

```
IF number of independent variables = 1
AND number of dependent variables = 1
THEN ...    (9 rules)

IF number of independent variables > 1
AND number of dependent variables = 1
THEN ...    (12 rules)
```

Two other rules reach conclusions immediately based upon the initial conditions of the number of variables.

```
IF number of independent variables > 1
AND number of dependent variables > 1
THEN ...

IF number of independent variables = 1
AND number of dependent variables > 1
THEN ...
```

Group Predictions

OUTCOME Variables
DATA TYPES
NUMBER of Variables

	1 Variable	>1 Variable	
PREDICTOR Variables, DATA TYPES, NUMBER of Variables	9 rules	12 rules	1 Variable
	12 rules	16 rules	>1 Variable

Individual Predictions

OUTCOME Variables
DATA TYPES
NUMBER of Variables

	1 Variable	>1 Variable	
PREDICTOR Variables, DATA TYPES, NUMBER of Variables	9 rules	3 rules	1 Variable
	12 rules	4 rules	>1 Variable

Condition Types in Rule Set 3
(Predictions)

Figure 12

The same four conditions are checked to determine group prediction methods. Forty-nine additional rules are in the rule set for selecting a model. [Figure 12]

The selection of techniques to make inferences about a population based on a sample (rule set 4) is made by using a relatively simple rule set. The first rule checks the number of populations sampled. For one population, six additional rules are checked. Two populations trigger eight rules to be checked. For more than two populations, three rules need to be checked. Each of these rules is compound with one condition being the type of data. Another condition is the inference being made about the population. Sometimes, one other condition is needed to differentiate between sample sizes. [Figure 13] For example,

```
IF data in the sample is quantitative
AND the inference is population mean = x
AND sample size > 40
THEN ...
```

The final rule set in the knowledge base, rules for dealing with structure, is the simplest and the smallest. The first rule checks that more than one variable is in the problem. If there is more than one variable then a conclusion about a technique can be reached using only an additional five rules, each of which checks only data type (interval, ordinal, categorical, ordinal and interval, categorical and quantitative). The simplicity indicates how new and undefined this area of statistics is. Very few techniques exist to

	NUMBER of Populations		
	<u>1 population</u>	<u>2 populations</u>	<u>>2 populations</u>
DATA TYPES, INFERENCE to be made	7 rules	10 rules	3 rules

Condition Types in Rule Set 4
(Inferences from a sample)

Figure 13

DATA TYPE ------------> 5 rules

Condition Types in Rule Set 5
(Structure)

Figure 14

study data and determine structure or to help reduce the data set in size by combining variables. As more theoretical work is completed in this area and the distinctions between techniques increase, the rule base should grow in size. [Figure 14]

It is not always possible to select an appropriate statistic. This can occur if no statistical technique exists for a particular type of data. When no technique is available, the conclusion clause in the rule suggests another approach to the problem if one is available. In some cases, the best conclusion is that the analyst re-examine the question under consideration. It may be that re-thinking the question and re-examining the data is the most beneficial and correct approach to the problem. This idea is used as the conclusion to some rules.

Pictorial Representation of the Knowledge Base

Although the knowledge base is constructed as a set of conditions and conclusions, it is possible to represent the rules in the knowledge base pictorially. This is possible because the conclusions are based upon a relatively small number of different conditions. In this knowledge base, the conclusions are based on a fixed set of conditions. They are not based on processes which cause the conclusions to change as the conditions change. This knowledge base represents a taxonomy.

When the knowledge base was validated by the experts, they used the pictorial model of the rule sets. Seeing a diagram of the rules helped the experts examine the knowledge base for validity and completeness. As mentioned, missing conclusions were easy to spot, showing up as empty cells in the diagram. Multiple statistics used as the conclusion to the same set of conditions can also be easily spotted in a box. Multiple statistics highlight the need to examine the conditions carefully. An additional condition is probably necessary to differentiate between the multiple statistics and to determine under which circumstances each statistic is most appropriate. For example, Hartley's test and Bartlett's test can both be used to test for equal variances between multiple populations of interval type data. However, Hartley's test is only appropriate if the sample sizes of the populations are equivalent. One more condition (IF SAMPLE SIZES ARE EQUAL) must be added to the rule base to differentiate between the two statistics.

Representing the Knowledge Base in the Consultant

The decision was made to code the consultant system in Pascal. This decision was made for several reasons, one was the ability to have easy access to the consultant by many students in University of Arizona computer labs. The ability to design and implement programmer-designed screens was also an important consideration. Further discussion of this is found in the next section, System Description.

The easier way to represent the knowledge base in the consultant system is to use the rule-based knowledge description rather than the table format. The rules were used to design the logic which leads the user through the steps of choosing an appropriate statistic. First, sufficient information is obtained about the data in the problem by asking questions to determine the necessary conditions to check. The conditions are stored in either CASE statements or IF/THEN statements. Rule sets 4 and 5 are represented strictly by CASE statements. The rule sets 1, 2, and 3 are represented by both IF/THEN and CASE statements. The rule sets are coded in nine different procedures.

Consultant System Description

This section contains a description of the physical structure of the program used to implement the mentor system as well as a description of the logic used internally in the program.

System Structure

The consultant system was built using the Pascal language. The decision was made to write the system in a conventional programming language for several reasons. The ability to use programmer designed screens in the system was highly desirable. The conclusion to a rule is usually a statistic which has to be presented to the user, including the formula and an explanation of the variables in the formula. Sometimes, additional text is needed to explain the statistic. Even though this system can be thought of as a rule-based system, the decision was made not to use a commercial rule-based expert system package because most of these do not provide the capability for extensive programmer screen design. Those packages that do allow screen design have a limited capability. The ability to use the system in a university computer lab was also necessary. By compiling the code into .COM files, the system is made widely available without requiring additional software.

The heart of the consultant system is the five rule sets that form the knowledge base. The five rule sets are modeled in nine procedures in the system, one each for group differences, individual value predictions, group predictions, and finding structure in the data. The rule set for analyzing relationships requires two procedures and the rule set for making inferences about a population requires three procedures to model the rules. These procedures use two different forms, either CASE statements or IF/THEN/ELSE statements. The most appropriate form was chosen for each rule set. The rule sets for relationships between two variables, finding structure and making inferences are built with CASE statements. The rule sets for individual predictions and relationships with more than two variables are represented by IF/THEN statements. The other rule sets incorporate both of the program structures in the procedures which implement them.

The consultant system begins with an information screen on how to use the system in the most effective manner. The instructions shown in Figure 15 tell the student to compare the problem being solved with the five detailed statistical question screens. Each of the five detail screens describes the general statistical question being answered and then gives specific examples of statistical questions of that type. [Figure 16] Since choosing the question which represents the problem which the novice is trying to analyze is the most difficult aspect of the task, a careful study of each question screen and a comparison with the problem will help the novice make a better choice of statistical question.

The user controls her movement through the consultant system in two ways. In most of the system, control is maintained by using one of the function keys, F1 through F10, or by pressing any key on the keyboard. The function key definitions are listed across the bottom of the screen. Flow of control is also guided by the dialogue carried on between the user and the system through a series of questions and answers. The questions are presented in a logical order to lead the user through the correct process for determining a statistical technique for answering a question about the data. This logical order is important because it allows the student to observe a model of the reasoning processes and procedural knowledge used to determine a statistical technique. The answers to the questions determine the state of the conditions in the rule sets. By asking only necessary questions, the consultant system shows the user which conditions are important. The condition values determine the conclusion to the rules. The conclusion is the statistical technique and when a conclusion is determined, this technique is displayed on the screen. Either the formula for the technique or the technique's description along with the definitions of each of the symbols in the formula are on the screen. Assumptions and references to sources for the lesser known techniques may be included on the screen.

The consultant system includes a condition value editing procedure. If the student sets the value of a condition incorrectly, then a procedure is available to let her modify the value. For example, if a student describes the variable AGE as being a categorical variable when it is actually a qualitative variable, the condition editor allows the student to correct the mistakes. Additions and deletions can also be made to the variable descriptions.

****
****

IMPORTANT - BE SURE TO READ THIS

The next 6 screens will present information for you to use to make more intelligent choices of statistical questions. There are 5 screens which will describe different questions that statistics can answer. These screens will also give examples of these questions. You should study the questions carefully and compare them to your problem. This will help you make better use of this tool.

The 5 question screens will be followed by a MAIN QUESTION MENU to allow you to choose a question. Depending upon which question you choose to answer about your problem, you will be asked a different sequence of questions. When you are finished with one statistical question and want to work on a new one, find a MAIN QUESTION MENU function key and you will be returned to the main menu. This will be your starting point for every question. You can always REVIEW QUESTIONS from that point.

Press any key AFTER you read this screen.

****
****

Consultant System Instruction Screen

Figure 15

*** Question 1 ***

Is there any relationship between 2 or more variables?
What is that relationship?
How strong is that relationship?

EXAMPLES ---

Is there any relationship between marital status and drop out rates for college students?

Is there any relationship between the time spent studying statistics and the ability to pass tests?

Is the shelf position for boxes of crackers related to the amount of sales of the crackers?

Example of a Detailed Question Screen

Figure 16

A help facility with instructions on how to use the system is also built into the consultant.

Two separate sets of definitions are available, one for the definition of statistical terms and one specifically for the explanation of data types. Since understanding the types of data being analyzed is crucial to the process of choosing a statistic, this second set of definitions is most important to the user. Definitions and examples of quantitative and qualitative data are given. These broad data types are broken down into sub-categories. Ordinal data, interval data, dichotomous data, as well as data which is dichotomous with an underlying distribution or categorical with more than two categories are described along with examples. Separate examples and definitions of independent/dependent or predictor/outcome variables are also available. The distinction between these two roles of the variables is important in some of the rules and a thorough understanding is necessary.

Physically, the mentor system is composed of six Pascal program files, CHOOSE.PAS is the main program file and includes the other files at compilation time. The entire program is built from 122 procedures. The screens are all stored in separate screen files on a separate floppy disk. When needed, the screens are read and displayed from the files. Although this is a fairly slow procedure, it is not a hinderance to the program due to the large amount of reading required on each screen. The users are not ready to move to the next screen any more quickly than the screens are displayed. Sixty-three different screens are defined. Two disk drives (any combination of hard and floppy drives) are needed to run the program.

Consultant System Logic

The logic of the consultant system follows logic patterns of the experts as stored in the knowledge base in the rule sets. This aids the user in learning these thought processes since one of the goals of the system is to help the novice learn the procedural and conditional knowledge employed by experts in choosing a technique. The user must first decide what question she is trying to answer about the data. Descriptions and examples of each question are presented on different screens for the user to flip through, study, and compare to her own data and problem. The question being answered about the data determines the content and order of the next sequence of questions.

Each of the questions has its own rule set of conditions whose values must be known before a conclusion can be reached. Only the questions necessary to find the condition values for a particular rule set are asked of the user. Explanations are included as to why some values are not necessary and, therefore, not asked for. For example, to determine correct conclusions in the rule set about relationships between variables, the user is asked for the number of variables and the types of data being analyzed. To determine techniques for making inferences, however, questions about the number of populations, the inference to be made, and the types of data must be answered. If mistakes are made when answering the questions, the user can employ an editing function to change the values of the conditions.

The values of the conditions determine the conclusion, the technique appropriate for analyzing the statistical question and the data under consideration. When a technique is chosen, the technique, the formula for the statistic, and an explanation of each variable in the formula are displayed on the screen for the user. If the technique is a lesser known one, then a reference is also given. A user has the choice to quit at any time or to choose another question and begin the decision process again.

CHAPTER 5

EXPERIMENTAL TEST RESULTS

This section discusses the results of two sets of experiments which were run on the consultant system to compare the success of students using the standard textbook approach to problem solving. The two independent experiments were conducted using one group of undergraduates and one group of graduate students.

Undergraduate Test Group Results

The undergraduate group consisted of 25 students from the second business statistics course in the University of Arizona College of Business. Twelve students were randomly assigned to the computer-assisted group and 13 were randomly assigned to the textbook assisted group. Each person in each group solved the same five problems, a total of 60 problems for the computer assisted group and 65 for the other group. The computer group selected 38 correct techniques (63% correct). The textbook-assisted group had 27 correct solutions or 41%. A chi-squared statistic was used to test the hypothesis:

H0: The proportion of problems solved correctly is the same for both computer-assisted and textbook-assisted methods.

H1: The proportion of problems solved correctly is different for the computer group than for the textbook-assisted group.

chi-squared = 5.937 (p = 0.015) [Table 1]

Table 1

Data for Measuring the Proportion of Correct Answers - Undergraduates

	Textbook	Computer
Number of Problems	65	60
Number Correct	27	38
Proportion Correct	41%	63%

chi-squared = 5.937 (p = 0.015)

This statistic is significant at approximately 0.02. The conclusion can be drawn that the computer group did solve a larger proportion of the questions correctly than did the textbook assisted group.

The second hypotheses compared the average number of correct solutions per student in each group. The hypotheses is:

H0: The average number of correct solutions per person is the same for both groups or less for the textbook group than for the computer group.

H1: The average number of correct solutions is greater for the computer group than for then textbook group.

t = 2.005 with 23 d.f. (P = 0.028, one-tailed) [Table 2]

Table 2

Data for Measuring the Average Number of Correct Answers - Undergraduate

	Textbook	Computer
Number of subjects	13	12
Number of correct problems	27	38
Average Correct	2.07	3.08
S-squared	1.90	1.23

Sp = .5035

t = 2.005 with 23 d.f. (p = 0.028, one tailed)

This statistic is significant at approximately the 0.03 level with 23 degrees of freedom. This evidence strongly suggest that the students in the computer group solved a larger average number of questions correctly than did the textbook assisted group.

To examine the data in more detail, chi-square tests were run on each question. The chi-squared tests measured the hypotheses for each question: [Table 3]

> H0: Choosing a correct solution is independent of the type of assistance used.
>
> H1: The type of assistance and the number of correct solutions are not independent.

Three problems were found to have non-independent solutions. [Table 3] In problem 2, both groups solved the problem incorrectly approximately the same number of times. The consultant group did not reach correct solutions because they failed to recognize the statistical questions which was to be answered. The problem was to test whether the variances of two populations were the same and both groups failed to recognize this. This is one indication that the consultant should include more extensive help in recognizing key words and in deciphering problems.

In the third problem, large portions of both groups reached the correct solution. This can be explained by the fact that the problem was an application of ANOVA and this subject wa being covered in their statistics class at the time the experiment was run. This acted as a frame of reference for the students to use. They can easily recognize the problem as being the same as the ones in class. In fact, several students made such a strong connection with ANOVA that they tried to match non-related problem types to the ANOVA solution.

In problem 5, chi-square is significant at the 0.01 level and in problem 1 the p-value for chi-squared is 0.14. In both of these problems, the difference can be attributed to the assistance method used. [Table 3]

Table 3

Chi-square, Measuring the Independence of Correct Answers and Method - Undergraduates

	Question 1	2	3	4	5
Computer assisted					
Correct	9	3	9	7	10
Incorrect	3	9	3	5	2
Textbook assisted					
Correct	6	3	9	5	4
Incorrect	7	10	4	8	9
Chi-squared	2.163	.013	.103	.986	6.99

The final hypotheses that was tested measured the differences between the average time taken to solve problems using each of the two methods of assistance. A medians test was used because of the non-homogeneity of the variances of the two samples. [Table 4]

H0: There is no difference between the average time to solve questions using a computer assistant or a textbook.

H1: There is a difference between the average time to solve the problems using the computer and the average time using the textbook.

chi-squared = 36.00 with 1 d.f. ($p < 10^{-6}$)

The conclusion is that there is a highly significant difference in the average time to solve the problems using the two methods of assistance. In fact, the computer assisted group took much longer than the textbook assisted group (an average of 10.5 minutes versus an average of 5.8 minutes). Just as in the pilot test, the computer group used the first question as a warm-up question to get used to the system and how it operates. They took approximately twice as long to solve the first problem as the textbook assisted group did. The average time to solve the first problem was 14.8 minutes but the individual times ranged from 10.2 minutes to 21.8 minutes. The times to solve problem 1 for the textbook group ranged from 3.5 minutes to 15.0 minutes with an average of 7.6 minutes.

Table 4

Data For Measuring the Average Time to Solve Problems - Undergraduates

	Textbook	Computer
Number of Problems	65	60
Average Time (minutes)	5.8	10.5
Variance	7.78	24.15
Times above median (7.0)	16	47
Times below median	49	13

chi-squared = 36.0 with 1 d.f. ($p < 10^{-6}$)

An additional medians test was run on the times the students used to solve problems two through five. The chi-squared value was reduced to 25.089 but this still indicates that the computer group takes significantly longer to solve the problems than does the textbook group. The learning curve for the computer consultant certainly requires more than one problem to overcome. Even after solving five problems, each of which explored a different statistical question, the computer group has probably not overcome the effect of the learning curve.

Another reason for the difference in the average time is that the computer system imposed the additional task of reading large amounts of material. With textbook assistance, the students could skim the book looking for equations which could be of use to them. On the other hand, the computer assistant required that screens of information be read and digested in order to move along in the process. If the students were slow readers, the time to solve a problem would be increased. The reading time would be reduced, however, with repeated use of the system as students become familiar with the contents of each screen.

In an attempt to decide why the computer assisted group chose incorrect solutions, the incorrect solutions were classified into two categories: incorrect with the wrong question chosen or incorrect with the correct question chosen but an incorrect variable description used. For those who chose the wrong question, the type of variables involved is not really relevant. Eighteen of the 22 incorrect solutions fell in the wrong question category. The other four incorrect solutions were in the incorrect variable description category. This classification points out the need for a possible enhancement to the consultant system to increase its ability to help the student choose an appropriate statistical question to describe the problem at hand. A knowledge base a key words and the questions they trigger could be added to the front of the consultant to assist the student in determining what question is being asked by the problem.

The other classification of correct question, incorrect variable description possibly shows the need for more examples and explanation about variables and they different roles they play.

Undergraduate Experiment Observations

When questioned, some of the students in the computer-assisted group expressed feelings that the system made them uncomfortable because it forced them to solve statistical problems using a process that they were not used to. The students explained that they usually opened a book, hunted for a likely equation, and then tried to plug numbers into the equation. This is the same behavior that Chi, Glaser, and Rees (1982) observed in their work with novice and expert physics students.

The subjects in this study also did not like the fact that they could not see any equations until the end of the decision process. This was true even though the problems they were given had no numbers in them. The problems had only expression such as "mean" and "standard deviation" which could take on numerical values after the data was collected. The subjects felt that if they used they system enough they would become more comfortable with "backwards" process and learn this alternate way of evaluating statistical problems.

The students had not been taught in their statistics classes the conditions under which different statistical techniques are to be used. They were missing this component needed to become skilled at solving statistics problems. The textbook assisted subjects were observed as they worked through the problems and they followed the same techniques that the computer assisted group had explained they would use if given a textbook instead of the computer mentor. Most of the subjects flipped through the textbook from front to back trying to find an appropriate procedure. After the experiment was over, the subjects explained that they were trying to find some equation or technique to jog their memory or fit the data in the problem. One subject was observed using the index to look up a particular technique. He explained that he read through the index until he recognized the name of a statistic and then checked to see if the equation fit the data. None of the subjects seemed to have a procedure or plan of attack to help them choose a technique.

The subjects in both groups said they were uncomfortable because these problems were more realistic than the ones they encountered in their statistics classes and texts. THey also said that without a frame of reference of a section or chapter in a book, they had no idea where to begin to decide upon a technique. The students were unable to transfer the knowledge they had been given in a standard class to real life problems.

The use of an instructor in conjunction with the consultant system would probably improve the results achieved by the students. This became apparent from observing students who had difficulty with the vocabulary in the consultant, especially the definitions of the data types. Even though definitions and examples of the various data types were available from the consultant system and were used by the students, they still had difficulty connecting the definitions, examples, and the data in the problem. An instructor could be helpful to ask leading questions when the student is staring at the computer screen for an extended amount of time.

Graduate Test Group Results

Nineteen doctoral students in MIS at the University of Arizona comprised the second experimental group to test the consultant system. Ten students were randomly assigned to the computer assisted group and nine students were randomly assigned to the textbook assisted group. This small sample size was used due to the limited size of the population of doctoral students. These students had all taken at least one statistics course but had taken them at a variety of times during bachelor's, master's, or doctoral degree programs. Only three persons had taken statistics courses during their doctoral programs. Nine persons had studied statistics at a master's degree level and 13 had taken statistics courses while pursuing a bachelor's degree. Of these students, five had taken courses at both the bachelor's and master's level and one student took statistics courses at all three levels. [Table 5] The most courses any of the students took was six (one student). Three persons had taken only one statistics course. The <u>number</u> of statistics courses did not have any relationship to the number of problems answered correctly by a student. However, the students who had the most problems correct were the ones who had used statistics most recently. For example, the only textbook group member to answer all five problems correctly had just finished analyzing data for two months.

The computer assisted group answered 37 out of 50 problems correctly or 74%. The textbook assisted group had 26 out of 45 problems or 58% correct. The hypotheses to test the proportion of correct answers is:

> H0: The proportion of problems solved correctly is the same for both computer-assisted and textbook-assisted methods.
>
> H1: The proportion of problems solved correctly is different for the computer group than for the textbook-assisted group.
>
> chi-squared = 2.729 (p = 0.10) [Table 6]

This test suggests that the computer group did solve a larger proportion of the questions correctly than did the textbook assisted group.

Table 5

Summary of the Number of Statistics Courses Taken by Doctoral Students Who Participated in the Experiment

Number of Courses	Level of Courses: Bachelor's	Master's	Doctoral
0	5	9	16
1	4	6	0
2	7	2	2
3	2	1	1

Table 6

Data for Measuring the Proportion of Correct Answers - Doctoral

	Textbook	Computer
Number of Problems	45	50
Number Correct	26	37
Proportion Correct	58%	74%

chi-squared = 2.729 ($p = 0.10$)

The second hypotheses compared the average number of correct solutions per student in each group. The hypotheses is:

H0: The average number of correct solutions per person is the same for both groups or less for the textbook group than for the computer group.

H1: The average number of correct solutions is greater for the computer group than for then textbook group.

t = 1.367 with 17 d.f. (P = 0.09, one-tailed) [Table 7]

The test suggests that the computer assisted group had a larger average number of problems correct than did the textbook group.

The average amount of time each group used to find solutions was also compared. A medians test was used again because of the non-homogeneity of the variances of the two samples. [Table 8]

H0: There is no difference between the average time to solve questions using a computer assistant or a textbook.

H1: There is a difference between the average time to solve the problems using the computer and the average time using the textbook.

chi-squared = 4.678 with 1 d.f. (p =.03)

The statistic is significant st the 0.03 level. The conclusion can be drawn that the computer assisted group takes more time than the textbook group assisted group to answer each problem. This follows the same pattern as the undergraduate and pilot groups, although the differences in the average times is much smaller for the doctoral students. In fact, if an additional medians test is conducted in which the times for the first question are removed from

Table 7

Data for Measuring the Average Number of Correct Answers - Doctoral

	Textbook	Computer
Number of subjects	9	10
Number of correct problems	26	37
Average Correct	2.88	3.7
S-squared	1.34	2.11

Sp = .599

t = 1.367 with 17 d.f. (p = 0.09, one tailed)

Table 8

Data For Measuring the Average Time to Solve Problems - Doctoral

	Textbook	Computer
Number of Problems	45	50
Average Time (minutes)	7.8	12.5
Variance	14.00	61.48
Times above median (7.0)	17	30
Times below median	28	20

chi-squared = 4.678 with 1 d.f. (p = 0.03) the data, then the chi-squared statistic is only 1.90 for which the p-value is 0.17. In all cases, the first problem was used as a warm-up problem for the computer assisted group to learn the mechanics of operating the system.. The group of graduate students using the consultant took approximately two and a half time as long as the textbook group to solve the first problem. This was a bigger difference than the undergraduate group exhibited but the reason for this difference, other than the learning curve, is that the graduate students also used the first problem to check everything the consultant could do. Being inquisitive MIS students, they were as anxious to learn the capabilites of the system as they were to find a correct solution in a timely manner. Once they explored the capabilities of the system, however, they worked through the other problems at approximately the same rate as the textbook group did. No other problem using the computer took more than 1.7 times as long to solve as the equivalent textbook problem.

Individual chi-squared test were performed on each problem for this group of students. The hypotheses for each questions was [Table 9]:

> H0: Choosing a correct solution is independent of the type of assistance used.
>
> H1: The type of assistance and the number of correct solutions are not independent.

When analyzed individually, only the third problem showed a dependence between correctness and assistance type. This problem is

Table 9

Chi-square, Measuring the Independence of Correct Answers and Method - Doctoral

	Question 1	2	3	4	5
Computer assisted					
Correct	7	6	10	5	9
Incorrect	3	4	0	5	1
Textbook assisted					
Correct	5	5	4	6	6
Incorrect	4	4	5	3	3
Chi-squared	.613	.22	7.59	.539	1.55

the one which the undergraduates solved correctly, regardless of assistance type. Part of the reason for the independence exhibited by the other problems is that one subject in the textbook group correctly solved every problem. This was due to extensive recent statistics analysis experience which could remove him from the novice category. This seems to agree with the idea that skill in problem solving is gained through experience.

Several of the graduate student users expressed the notion that the consultant has a role as a method of reviewing statistical techniques that had once been learned but were forgotten over time. While they could use the consultant as a tool to choose a technique, this group saw the possibility of giving many different combinations of answers to the consultant as a way to explore different techniques and their appropriate uses.

CHAPTER 6

SUMMARY OF RESEARCH RESULTS

This chapter first presents a summary of the results of this research project. The second section discusses extensions to the research and further questions which have come out of the work.

Contributions

Four significant results have resulted from this work with an educational consultant system for statistics. The first result was the development of a knowledge base for 29 common statistics for business applications. In this knowledge base, the statistics were categorized using four classifications or combinations of these classes. The first classification is the statistical question one is trying to answer about the data in the problem. Five general statistical questions are included in the knowledge base:

1. What is the strength or degree of the relationship, if any, among variables?
2. Are there any differences between groups which are subjected to different treatments?
3. Can predictions of individual values or group membership be made?
4. Are inferences about a population based on sample needed?

5. Is there any pattern or structure to a large data set?

The other categories used to classify the statistics are type of data being analyzed such as qualitative (dichotomous, dichotomous with an underlying distribution, non-dichotomous) or quantitative (interval, ordinal); the number of variables being analyzed; and whether a dependency relationship exists or must exist between the data. While using these classes to categorize the statistics within a question, the discovery was made that one categorization schema was not sufficient for all five questions. Each question required its own categorization schema for the knowledge base. Therefore, five different schemas were developed.

Tables were chosen as an appropriate format for representing the knowledge because they allowed the experts who were consulted to easily spot missing statistics as well as to spot categories with multiple statistics. If multiple statistics were in a cell in the table, this indicated the need to find a condition to distinguish between these statistics. For example, sample size may be a factor which tells the analyst which of two statistics to use. The tables were easier for the experts to use than a listing of the knowledge base as a set of rules.

After the statistics were categorized, however, the knowledge base in table format was transformed into a rule base. The rules are all in a condition/action format. The conditions are the categories from the knowledge base (represented by IF clauses) and the actions are the statistics to be used by the analyst (represented by THEN clauses).

The second result of this research was the development of the educational consultant system incorporating the rule base, instructional screens, and dialogue with the user. The consultant system uses a top-down approach to analyzing statistical data beginning with the statistical question to be answered about the data and working toward a particular statistical technique. This is exactly the opposite approach taken by most students when analyzing problems. They traditionally uses a bottom up approach and search for equations in a statistics book. They fit the data from the problem into the equation and, when all the data are gone and the variables in the equation all have values, the students decide they have chosen the appropriate technique. They are not applying any logical procedures to help them choose a technique. The consultant system, however, allows the students to be exposed to procedural and conditional knowledge used by the experts.

This consultant system was tested using two different groups, undergraduate business students currently enrolled in a second level statistics course and doctoral students in MIS. Both groups were split in half, randomly assigned to either a textbook or consultant system group, and given five problems to solve. In both cases, the group using the computer consultant achieved a significantly higher proportion of correct answers and a significantly higher average number of correct answers per person. This difference occurred even though none of the users of the consultant system had ever been exposed to a top-down approach to statistical analysis and non had used the consultant system before. This indicates that the consultant system is a definite step forward in helping students choose statistical techniques to solve problems.

Another result came from observing the students as they used the consultant system and from analyzing the reasons that incorrect conclusions were reached by the system. The main reason for incorrect solutions was that an inappropriate statistical question to be answered about the data was chosen. This suggests the need for additional logic for the student to consult while selecting a question. This could consist of some sort of key word/question association or some logic for analyzing the problem definition. This logic could be incorporated into the consultant system at the front end to help users and, ultimately, the consultant achieve a higher success rate.

To summarize, the results of this work are:

1. Development of a knowledge base of common business statistics and the categorization of these statistics by question being answered, types of data being analyzed, number of variables being analyzed, and dependency relationship.

2. Construction of a rule base from this pictorial knowledge base in a condition/action format.

3. Design and implementation of an educational consultant system to help students choose a correct statistical technique,

incorporating the rule base, educational screens, and a dialogue with the user.

4. The discovery of the need for an additional rule base or logic base for deciding on the appropriate question to be answered about the data. This discovery was made through the use of an experimental approach to software validation. If the experiments had not been run on the software, the need for additional problem analysis logic would probably not have been discovered.

Further Research Ideas

The most obvious additional research needed is the enlargement of the statistical technique set handled by the consultant system. A logical extension to the set would be to include a subset of non-parametric techniques in the rule base. This would serve to make the consultant useful to more than the first two levels of business statistics students. Additional rules would also be needed to help the student decide on the appropriate times for using non-parametric versus parametric statistics.

This consultant system has only been testes on a short exposure basis by students being taught statistics in the traditional bottom-up approach. They learn the techniques but do not learn the conditions for using each of the techniques. An interesting research question would be: "What is the long-term effect on the students' statistical skills if the consultant system were used in conjunction with an instructor?". The students could use the consultant to become aware of the conditions under which each technique should be used as well as to learn the process of choosing statistics. An instructor who wished to teach using a top-down approach could also use the consultant but in this case the consultant system would be used to reinforce the ideas learned in class. In both cases, the question could be asked: "Would the students' performance with only a textbook as an assistant improve after using the consultant system for an entire semester?" In other words, would the students have developed some procedural and conditional knowledge in statistics as a result of using the consultant?

A two-part enhancement to the consultant system which could be both beneficial and interesting was indicated by the analysis of the reasons for incorrect solutions. Some sort of problem analysis module is needed to help the student analyze the problem being solved before any questions are asked by the consultant to determine values for the conditions. The students had difficulty deciding what statistical question they were trying to answer in the problem. If a set of rule could be developed based on key words in a problem for the student to consult, then the system's (and the student's) performance possibly could be improved. This module also needs logic for the student to use when deciding how many variables they are dealing with in the problem. This was the second reason identified for selecting incorrect statistical techniques, incorrect description of the variables in the problem. Although the consultant presented the definitions and examples of both problem and variable types, the students had some difficulty making a connection between these example and the actual problems.

A further module which could be incorporated into the consultant would be rules about assumptions. After a technique is chosen for the student, the student could be instructed to perform certain tests on the data to check whether the assumptions underlying the technique are met by the data. The rules for applying the results of these tests could be included in the consultant. Alternative techniques could be included as part of the rule base to be triggered if the data fail to meet the assumptions.

APPENDIX A

INITIAL FRAMEWORK

QUESTION: Strength of relationship among variables

No distinction made between independent (predictor) and dependent variables	Quantitative	Canonical Correlation Factor Analysis
	Qualitative	
	Mixed types	
Single Variable	--- not applicable ---	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	simple R-squared Biserial r (y reduces to 2 classes)
	>1 Dependent Quantitative	not applicable
	1 Dependent Qualitative	R-squared using logit approximation Biserial r Point Biserial r
	>1 Dependent Qualitative	not applicable

Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	multiple R-squared correlation ratio (eta) curved regression
	>1 Dependent Quantitative	canonical correlation
	1 Dependent Qualitative	R-squared logit
	>1 Dependent Qualitative	canonical correlation
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	multiple R-squared point biserial r (x is exactly 2 categories)
	>1 Dependent Quantitative	--- not applicable ---
	1 Dependent Qualitative	R-squared logit tetachoric correlation (2 categories each) contingency correlation Phi (per Guilford)
	>1 Dependent Qualitative	--- not applicable ---

Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	multiple R-squared
	>1 Dependent Quantitative	canonical correlation
	1 Dependent Qualitative	R-squared logit
	>1 Dependent Qualitative	canonical correlation
Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	multiple R-squared
	>1 Dependent Quantitative	
	1 Dependent Qualitative	R-squared logit
	>1 Dependent Qualitative	

QUESTION : Measure the significance of group differences

No distinction made between independent (predictor) and dependent variables	Quantitative	--- not applicable ---
	Qualitative	--- not applicable ---
	Mixed types	--- not applicable ---
Single Variable	--- not applicable ---	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	ANOVA - break IV into groups t-test - u1=u2
	>1 Dependent Quantitative	MANOVA -break IV into groups Hotelling's T-square
	1 Dependent Qualitative	ANOVA - break IV into groups; quantify DV ANCOVA
	>1 Dependent Qualitative	MANOVA -DV quantified MANCOVA - DV quantified Hotelling T-squared

Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	factorial ANOVA - break IV into groups
	>1 Dependent Quantitative	factorial MANOVA - break IV into groups factorial MANCOVA
	1 Dependent Qualitative	factorial ANOVA - break IV into groups; DV quantified
	>1 Dependent Qualitative	factorial MANOVA - DV quantified
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	ANOVA
	>1 Dependent Quantitative	Hotelling T-squared if 2 groups of IV MANOVA
	1 Dependent Qualitative	ANOVA - DV quantified ANCOVA - DV quantified contingency tables
	>1 Dependent Qualitative	MANOVA - DV quantified MANCOVA - DV quantified
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	multifactor ANOVA
	>1 Dependent Qualitative	multifactor MANOVA, MANCOVA (interaction) (separate tests on each IV)
	1 Dependent Qualitative	multifactor ANOVA - DV quantified
	>1 Dependent Qualitative	multifactor MANOVA - DV quantified

Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	multifactor ANOVA with quantiified IV grouped
	>1 Dependent Quantitative	multifactor MANOVA with quantified IV grouped
	1 Dependent Qualitative	multifactor ANOVA with quant IV grouped; DV quantified
	>1 Dependent Qualitative	multifactor MANOVA with Quant IV grouped; DV quantified

QUESTION : Prediction of group membership

No distinction made between independent (predictor) and dependent variables	Quantitative	
	Qualitative	
	Mixed types	
Single Variable	--- not applicable ---	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	maximum likelihood logit 2 categories of DV
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	--- no groups --- --- not applicable
	>1 Dependent Quantitative	--- no groups --- --- not applicable
	1 Dependent Qualitative	max. likelihood logit 2 categories of DV discriminant analysis
	>1 Dependent Qualitative	
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	--- no groups --- --- not applicable
	>1 Dependent Quantitative	--- no groups --- --- not applicable
	1 Dependent Qualitative	minimum logit chi-square maximum likelihood logit - if DV has exactly 2 categories
	>1 Dependent Qualitative	
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	--- no groups --- --- not applicable
	>1 Dependent Quantitative	--- no groups --- --- not applicable
	1 Dependent Qualitative	minimum logit chi-square max. likelihood logit DV has exactly 2 if categories
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

QUESTION : Structure of variables

No distinction made between independent (predictor) and dependent variables	Quantitative	principal components factor analysis cluster analysis
	Qualitative	nonmetric multi-dimensional scaling with ordinal data
	Mixed types	
Single Variable	--- not applicable ---	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Independent Variable (Predictor)	Dependent	
Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

QUESTION : Inferences about population - based on a sample (parameters or distribution)

No distinction made between independent (predictor) and dependent variables	Quantitative	t-test (u1 = u2) equal variances
	Qualitative	
	Mixed types	
Single Variable	t-test (u1=u2); equal variances binomial test p1=p2 chi-square skewness tests	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Mixed Types	1 Dependent Quantitative	
Qualitative and Quantitative	>1 Dependent Quantitative	
	1 Dependent Qualitative	
	>1 Dependent Qualitative	

QUESTION : Prediction of individual values

No distinction made between independent (predictor) and dependent variables	Quantitative	--- no predictor --- --- not applicable ---
	Qualitative	--- no predictor --- --- not applicable ---
	Mixed types	--- no predictor --- --- not applicable ---
Single Variable	--- nothing to be predicted --- --- not applicable ---	
Independent Variable (Predictor) 1 Quantitative	1 Dependent Quantitative	simple linear regression polynomial regression
	>1 Dependent Quantitative	
	1 Dependent Qualitative	simple linear regression (logistic function)
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Quantitative	1 Dependent Quantitative	multiple linear regression polynomial regression
	>1 Dependent Quantitative	canonical correlation (Aaker)
	1 Dependent Qualitative	logistic regression function
	>1 Dependent Qualitative	
Independent Variable (Predictor) 1 Qualitative	1 Dependent Quantitative	
	>1 Dependent Quantitative	--- not applicable --
	1 Dependent Qualitative	
	>1 Dependent Qualitative	
Independent Variable (Predictor) >1 Qualitative	1 Dependent Quantitative	multiple regression
	>1 Dependent Quantitative	
	1 Dependent Qualitative	logistic regression
	>1 Dependent Qualitative	

Independent Variable (Predictor) >1 Mixed Types Qualitative and Quantitative	1 Dependent Quantitative	multiple regression
	>1 Dependent Quantitative	
	1 Dependent Qualitative	logistic regression
	>1 Dependent Qualitative	

APPENDIX B

FRAMEWORK INCORPORATING FIRST EXPERT'S KNOWLEDGE

QUESTION : Strength of relationship among variables

Single Variable	-- not applicable -- no variables to have a relationship between

---- 2 variables in data set ----

VARIABLE 2 \ VARIABLE 1	dichot- omous	dichotomous (underlying normal)	ordinal	interval	categorical (>2 cat.)
dichotomous	15	16	17	18	25
dichotomous (underlying normal)	16	19	20	22	25
ordinal	17	20	21	24	26
interval	18	22	24	23	27
categorical (>2 category)	25	25	26	27	28

110

More than 2 variables --

all quantitative -- factor analysis
multiple r-squared

all categorical (2 or >2) categories --
r-sqrd logit analysis

mixed quantitative and categorical --
r-sqrd logit analysis;
multiple r-sqrd with indicator variables

15 - Phi
16 - none; can use Phi
17 - rank biserial correlation
18 - point biserial
19 - tetrachoric r
20 - none
21 - Kendall's tau or Spearman rank coefficient
22 - biserial
23 - simple regression
24 - none - convert var2 to ranks and use 21
25 - 2 x M contingency table
26 - M x N contingency table
27 - convert interval variable to categories and use 26
28 - M x N contingency table

REFERENCES

Glass, G.V. and Stanley, J.C. Statistical Methods in Education and Psychology

Guilford, J.P. Fundamental Statistics in Psychology and Education

Morris, C.N. and Rolph, J.E. Introduction to Data Analysis and Statistical Inference

QUESTION : Measure the significance of group differences (location)

Single Variable	--- not applicable ---

--- more than 1 variable ---

DEPENDENT

INDEP.	1 quan	>1 quan	1 qual	>1 qual	mixed
1 quan	1	2, 3	1	2, 3	2
>1 quan	5	6	5	6	6
1 qual	1,7,8, 9	2,3	1,9,10	2,3	2
>1 qual	5	2	1	6	6
>1 mixed	5	2	1	6	6

1 -- ANOVA *, **
2 -- MANOVA *, **
3 -- Hotelling's T-sqrd
5 -- factorial ANOVA *, **
6 -- factorial MANOVA *, **
7 -- Rank sum test
8 -- Wilcoxon signed ranks
9 -- t-test (2 groups)
10 -- contingency tables

* ANOVA and MANOVA are designed for categorical independent variables but may be used with continuous independent variables if analyst can break measurements into artificial categories.

** ANOVA and MANOVA can be used for qualitative dependent variables IF the qualitative dependent variables can be quantitized along at least an ordinal scale. For example, 3 categories of patient appearances (flushed, pale, and jaundiced) could be quantitized as flushed = 4, pale = 2 and jaundiced = 1 if a flushed appearance were 4 times as good as a jaundiced appearance.

REFERENCES

Neter, J. and Wasserman, W. Applied Linear Statistical Models

Tabachnick, B.G. and Fridell, L.S. Using Multivariate Statistics

QUESTION : Prediction of group membership

No distinction made between independent (predictor) and dependent variables	Quantitative	--- no predictor --- question not applicable
	Qualitative	--- no predictor --- question not applicable
	Mixed types	--- no predictor --- question not applicable
Single Variable	--- not applicable ---	

--- More than 1 variable ---

OUTCOME

PREDICT	1 quan	1 categ. dichot.	1 categ non-dich	>1 quant	>1 categ. non-dich	>1 dich	mixed
1 quan	113	114	114	114	114	114	114
1 categ dichot	113	115	114	114	114	114	114
1 categ non-dich	113	115	114	114	114	114	114
>1 quant	113	115,116	116	113	116	116	114
>1 categ non-dich	113	115,116, 118	116	113	116	116	114
>1 categ dichot	113	115.116, 118	116	113	116	116	114
>1 mixed	113	116,119	116,119	113	116	116	114

113 = redefine outcome variables as categorical
114 = no technique
115 = maximum likelihood logit *
116 = discriminant analysis **
117 = no techniques
118 = minimum logit chi-square *
119 = classification trees (non-parametric technique)

REFERENCES

* Morris, C.N. and Rolph, J.E. Introduction to Data Analysis and Statistical Inference

** Tabachnick, B.G. and Fridell, L.S. Using Multivariate Statistics

QUESTION : Structure of variables

No distinction made between independent (predictor) and dependent variables Any number of variables > 1	Quantitative (continuous)	principal components factor analysis cluster analysis
	Quantitative (Ordinal)	nonmetric multi-dimensional scaling (ordinal data)
	Categorical	no techniques ????
	Mixed types (ordinal and interval)	nonmetric multi-dimensional scaling
	Mixed types (categorical and quantitative)	no techniques ??

REFERENCES

Aaker, D. Multivariate Analysis in Marketing: Theory and Application

Anderberg, M.R. Cluster Analysis for Applications

QUESTION : Inferences about population - based on a sample (parameters or distribution)

--- Single population ---

Quantitative	t-test (ul=a); Chi-sqr test for population variance = a; Chi-sqr for population fitting a distribution; Kolmorgorov-Smirnoff for fitting a distribution; z-test for proportion of population possessing some attribute
Qualtitative	Chi-sqr test for fitting a distribution; z-test for proportion of population possessing some attribute

--- 2 Independent Populations ---

Quantitative	t-test for equal population means (samples have the same variance) (pooled variance) z-test and n>30 for equal population means (the samples have different variances) F-test for equal population variances z-test for equal proportions of the populations possessing some attribute z-test for difference between the population means (large sample) z-test for difference between the population means (small sample) (pooled deviation)
Qualitative	t-test for equal proportions of the populations possessing some attribute
2 paired samples (Quantitative)	z-test for equal means t-test for equal variances

--- more than 2 populations ---

quantitative	ANOVA for equal means Bartlett's test for equal variances (any sample size) Hartley's test for equal variances (equal sample sizes)

REFERENCES

Groebner, D.F. and Shannon, P.W. Business Statistics

Walpole, R.E. and Myers, R.H. Probability and Statistics for Engineers and Scientists

QUESTION : Prediction of individual values

No distinction made between independent (predictor) and dependent variables	Quantitative	--- no predictor - question not applicable
	Qualitative	--- no predictor - question not applicable
	Mixed types	--- no predictor - question not applicable
Single Variable	--- nothing to be predicted --- question not applicable	

OUTCOME VARIABLE

PREDICTOR	1 Interval/ Ordinal	1 Dich	1 Nondich	>1 any type
1 Interval/ Ordinal	85,87	90	90	no techniques
1 Dichot	88	??	??	no techniques
1 NonDich	88	??	??	no techniques
>1Interval/ Ordinal	86,87,97	91, 97	91,97	98
>1 Dichot	96,97	??	??	
>1 Nondich	96,97	??	??	
>1 Mixed	96,97	92,97	92,97	

85 = simple linear regression
86 = multiple linear regression
87 = investigate curvilinear regression model
88 = none
89 = simple linear regression
90 = simple linear regression with indicator variables for the outcome variable
91 = multiple linear regression with indicator variables for indicator variables for outcome variables
92 = multiple regression with indicator variables for the outcome and predictor variables
96 = multiple regression with indicator variables for predictor variables
97 = multiple regression with interaction
98 = canonical correlation

REFERENCES

Neter, J., Wasserman, W. and Kutner, M.H. Applied Linear Regression Models

Tabachnick, B.G. and Fridell, L.S. Using Multivariate Statistics

APPENDIX C

KNOWLEDGE BASE RULE SETS IN IF/THEN FORMAT

RULE SET 1

Strength of Relationships Between Variables

IF number of variables = 1
THEN question not applicable

IF number of variables = 2
and variable 1 = dichotomous
and variable 2 = dichotomous
THEN use Phi

IF number of variables = 2
and variable 1 = dichotomous
and variable 2 = dichotomous with an underlying
distribution
THEN no statistic; use Phi

IF number of variables = 2
and variable 1 = dichotomous
and variable 2 = ordinal
THEN use rank biserial correlation

IF number of variables = 2
and variable 1 = dichotomous
and variable 2 = interval
THEN point biserial correlation

IF number of variables = 2
and variable 1 = dichotomous
and variable 2 = categorical with more than 2
categories
THEN use 2 x n contingency table

IF number of variables = 2
and variable 1 = dichotomous with an underlying
    distribution
and variable 2 = dichotomous with an underlying
    distribution
THEN use tetrachoric correlation

IF number of variables = 2
and variable 1 = dichotomous with an underlying
    distribution
and variable 2 = ordinal
THEN no technique available

IF number of variables = 2
and variable 1 = dichotomous with an underlying
    distribution
and variable 2 = interval
THEN use biserial correlation

IF number of variables = 2
and variable 1 = dichotomous with an underlying
    distribution
and variable 2 = categorical with more than 2
    categories
THEN use 2 x m contingency tables

IF number of variables = 2
and variable 1 = ordinal
and variable 2 = ordinal
THEN use Kendall's Tau

IF number of variables = 2
and variable 1 = ordinal
and variable 2 = interval
THEN no technique; convert interval to ranks and use
    Kendall's Tau

IF number of variables = 2
and variable 1 = ordinal
and variable 2 = categorical with more than 2 categories
THEN use M x N contingency tables

IF number of variables = 2
and variable 1 = interval
and variable 2 = interval
THEN use Pearson's correlation

```
IF number of variables = 2
and variable 1 = interval
and variable 2 = categorical with more than 2 categories
THEN convert interval data to categorical and use M x N
    contingency table

IF number of variables = 2
and variable 1 = categorical with more than 2 categories
and variable 2 = categorical with more than 2 categories
THEN use M x N contingency table

IF number of variables > 2
AND all quantitative
THEN use factor analysis OR
     use multiple r-squared if independent/dependent

IF number of variables > 2
AND all categorical ( >= 2 categories)
THEN use r-sqrd logit

IF number of variables > 2
AND mixed quantitative and categorical
THEN use r-sqrd logit OR
     use multiple r-sqrd with indicator variables
```

RULE SET 2

Measure the Significance of Group Differences

IF you are not able to distinguish between independent and dependent variables
THEN the question is inappropriate for the data

IF 1 dependent quantitative variable
and 1 independent quantitative variable
THEN use ANOVA redefining independent variable as categorical

IF 1 dependent quantitative variable
and >1 independent quantitative variable
THEN use factorial ANOVA redefining independent variable as categorical

IF 1 dependent quantitative variable
and 1 independent qualitative variable
THEN use ANOVA OR
use t-test if only 2 groups

IF 1 dependent quantitative variable
and >1 independent qualitative variable
THEN use factorial ANOVA

IF 1 dependent quantitative variable
and mixed independent (quantitative and qualitative) variables
THEN use factorial ANOVA redefining the quantitative independent variables as categorical

IF >1 dependent quantitative variable
and 1 independent quantitative variable
THEN use MANOVA redefining independent variable as categorical
OR use Hotelling's T if only have two groups

IF >1 dependent quantitative variable
and >1 independent quantitative variable
THEN use factorial MANOVA redefining independent variable as categorical

IF >1 dependent quantitative variable
and 1 independent qualitative variable
THEN use MANOVA OR
use Hotelling's t-test if only 2 groups

IF >1 dependent quantitative variable
and >1 independent qualitative variable
THEN use factorial MANOVA

IF >1 dependent quantitative variable
and mixed independent (quantitative and qualitative) variables
THEN use factorial MANOVA redefining the quantitative independent variables as categorical

IF 1 dependent qualitative variable
and 1 independent quantitative variable
THEN use ANOVA redefining independent variable as categorical if you can quantify the dependent variable

IF 1 dependent qualitative variable
and >1 independent quantitative variable
THEN use factorial ANOVA redefining independent variable as categorical if you can quantify the dependent variable

IF 1 dependent qualitative variable
and 1 independent qualitative variable
THEN use ANOVA if you can quantify the dependent variable OR
use t-test if only 2 groups OR
use contingency tables

IF 1 dependent qualitative variable
and >1 independent qualitative variable
THEN use factorial ANOVA if you can quantify the dependent variable

IF 1 dependent qualitative variable
and mixed independent (quantitative and qualitative) variables
THEN use factorial ANOVA redefining the quantitative independent variables as categorical and if you can quantify the dependent variable

IF >1 dependent qualitative variable
and 1 independent quantitative variable
THEN use MANOVA redefining independent variable as
categorical if you can quantify the dependent
variable OR
use Hotelling's T-test if only two groups

IF >1 dependent qualitative variable
and >1 independent quantitative variable
THEN use factorial MANOVA redefining independent
variable as categorical if you can quantify
the dependent variable

IF >1 dependent qualitative variable
and 1 independent qualitative variable
THEN use MANOVA if you can quantify the dependent
variable OR
use Hotelling's t-test if only 2 groups

IF >1 dependent qualitative variable
and >1 independent qualitative variable
THEN use factorial MANOVA if you can quantify the
dependent variable

IF >1 dependent qualitative variable
AND mixed independent (quantitative and qualitative)
variables
THEN use factorial MANOVA redefining the quantitative
independent variables as categorical and if you
can quantify the dependent variable

IF mixed dependent (quantitative and qualitative)
variables
AND 1 independent quantitative variable
THEN use MANOVA redefining independent variable as
categorical if you can quantify the qualitative
dependent variables

IF mixed dependent (quantitative and qualitative)
variables
AND >1 independent quantitative variable
THEN use factorial MANOVA redefining independent variable
as categorical if you can quantify the
qualitative dependent variable

IF mixed dependent (quantitative and qualitative)
variables
AND 1 independent qualitative variable
THEN use MANOVA if you can quantify the qualitative
dependent variable

IF mixed dependent (quantitative and qualitative)
variables
AND >1 independent qualitative variable
THEN use factorial MANOVA if you can quantify the
qualitative dependent variable

IF mixed dependent (quantitative and qualitative)
variables
AND mixed independent (quantitative and qualitative)
variables
THEN use factorial MANOVA redefining the quantitative
independent variables as categorical and if
you can quantify the qualitative dependent
variable

RULE SET 3

Predicting Group Membership

IF you are not able to distinguish between independent
and dependent variables
THEN the question is inappropriate for the data

IF the number of variables = 1
THEN the group membership cannot be predicted

IF 1 quantitative outcome variables
AND 1 quantitative predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variables
AND 1 dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variable
AND 1 non-dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variables
AND >1 quantitative predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variables
AND >1 dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF 1 quantitative outcome variables
AND >1 mixed categorical and quantitative predictor
variables
THEN redefine the outcome variables as categorical

IF 1 categorical dichotomous outcome variables
AND 1 quantitative predictor variable
THEN no technique; need more than one outcome variable
to define a group

IF 1 categorical dichotomous outcome variables
AND 1 dichotomous categorical predictor variable
THEN use maximum likelihood logit

IF 1 categorical dichotomous outcome variables
AND 1 non-dichotomous categorical predictor variable
THEN use maximum likelihood logit

IF 1 categorical dichotomous outcome variables
AND >1 quantitative predictor variable
THEN use maximum likelihood logit OR
use discriminant analysis

IF 1 categorical dichotomous outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN use maximum likelihood logit OR
use discriminant analysis OR
use minimum logit chi-square

IF 1 categorical dichotomous outcome variables
AND >1 dichotomous categorical predictor variable
THEN use maximum likelihood logit OR
use discriminant analysis OR
use minimum logit chi-square

IF 1 categorical dichotomous outcome variables
AND >1 mixed categorical and quantitative predictor variables
THEN use discriminant analysis OR
use classification trees (non-parametric method)

IF 1 categorical non-dichotomous outcome variables
AND 1 quantitative predictor variable
THEN no technique; need more than one outcome variable to define a group

IF 1 categorical non-dichotomous outcome variables
AND 1 dichotomous categorical predictor variable
THEN no technique available

IF 1 categorical non-dichotomous outcome variables
AND 1 non-dichotomous categorical predictor variable
THEN no technique available

IF 1 categorical non-dichotomous outcome variables
AND >1 quantitative predictor variable
THEN use discriminant analysis

IF 1 categorical non-dichotomous outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN use discriminant analysis

IF 1 categorical non-dichotomous outcome variables
AND >1 dichotomous categorical predictor variable
THEN use discriminant analysis

IF 1 categorical non-dichotomous outcome variables
AND >1 mixed categorical and quantitative predictor variables
THEN use discriminant analysis OR
use classification trees (non-parametric method)

IF >1 quantitative outcome variables
AND 1 quantitative predictor variable
THEN redefine the outcome variables as categorical

IF >1 quantitative outcome variables
AND 1 dichotomous categorical predictor variable
THEN no technique available

IF >1 quantitative outcome variable
AND 1 non-dichotomous categorical predictor variable
THEN no technique available

IF >1 quantitative outcome variables
AND >1 quantitative predictor variable
THEN redefine the outcome variables as categorical

IF >1 quantitative outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF >1 quantitative outcome variables
AND >1 dichotomous categorical predictor variable
THEN redefine the outcome variables as categorical

IF >1 quantitative outcome variables
AND >1 mixed categorical and quantitative predictor variables
THEN redefine the outcome variables as categorical

IF >1 categorical dichotomous outcome variables
AND 1 quantitative predictor variable
THEN no technique available

IF >1 categorical dichotomous outcome variables
AND 1 dichotomous categorical predictor variable
THEN no technique available

IF >1 categorical dichotomous outcome variables
AND 1 non-dichotomous categorical predictor variable
THEN no technique available

IF >1 categorical dichotomous outcome variables
AND >1 quantitative predictor variable
THEN use discriminant analysis

IF >1 categorical dichotomous outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN use discriminant analysis

IF >1 categorical dichotomous outcome variables
AND >1 dichotomous categorical predictor variable
THEN use discriminant analysis

IF >1 categorical dichotomous outcome variables
AND >1 mixed categorical and quantitative predictor variables
THEN use discriminant analysis

IF >1 categorical non-dichotomous outcome variables
AND 1 quantitative predictor variable
THEN no technique; need more than one outcome variable to define a group

IF >1 categorical non-dichotomous outcome variables
AND 1 dichotomous categorical predictor variable
THEN no technique available

IF >1 categorical non-dichotomous outcome variables
AND 1 non-dichotomous categorical predictor variable
THEN no technique available

IF >1 categorical non-dichotomous outcome variables
AND >1 quantitative predictor variable
THEN use discriminant analysis

IF >1 categorical non-dichotomous outcome variables
AND >1 non-dichotomous categorical predictor variable
THEN use discriminant analysis

IF >1 categorical non-dichotomous outcome variables
AND >1 dichotomous categorical predictor variable
THEN use discriminant analysis

IF >1 categorical non-dichotomous outcome variables
AND >1 mixed categorical and quantitative predictor variables
THEN use discriminant analysis OR
use classification trees (non-parametric method)

IF mixed (quantitative and qualitative) outcome variables
THEN no technique available

RULE SET 3

Predicting Individual Values

IF you cannot distinguish between outcome and predictor variables
THEN the question is inappropriate for the data in the problem

IF the number of variables = 1
THEN there is no prediction possible; inappropriate question

IF the number of outcome variables > 1
AND the number of predictor variables = 1
THEN no techniques are available;
reconsider the question being asked about the data

IF the number of outcome variables > 1
AND the number of predictor variables > 1
THEN use canonical correlation

IF 1 quantitative outcome variable
AND 1 quantitative predictor variable
THEN use simple linear regression AND
investigate a curvilinear relationship

IF 1 quantitative outcome variable
AND 1 dichotomous predictor variable
THEN no technique available

IF 1 quantitative outcome variable
AND 1 non-dichotomous predictor variable
THEN no technique available

IF 1 quantitative outcome variable
AND >1 quantitative predictor variable
THEN use multiple linear regression OR
use multiple regression with interaction AND
investigate a curvilinear relationship

IF 1 quantitative outcome variable
AND >1 dichotomous predictor variable
THEN use multiple linear regression with indicator variables OR
use multiple regression with interaction

IF 1 quantitative outcome variable
AND >1 non-dichotomous predictor variable
THEN use multiple linear regression with interaction AND
use multiple regression with interaction

IF 1 quantitative outcome variable
AND >1 mixed (both qualitative and quantitative)
predictor variables
THEN use multiple linear regression with indicator
variables
use multiple regression with interaction

IF 1 dichotomous outcome variable
AND 1 quantitative predictor variable
THEN use simple linear regression using indicator
variables for the outcome variables

IF 1 dichotomous outcome variable
AND 1 dichotomous predictor variable
THEN no technique available

IF 1 dichotomous outcome variable
AND 1 non-dichotomous predictor variable
THEN no technique available

IF 1 dichotomous outcome variable
AND >1 quantitative predictor variable
THEN use multiple linear regression with indicator
variables for the outcome variables AND
use multiple regression with interaction

IF 1 dichotomous outcome variable
AND >1 dichotomous predictor variable
THEN no technique available

IF 1 dichotomous outcome variable
AND >1 non-dichotomous predictor variable
THEN no technique available

IF 1 dichotomous outcome variable
AND >1 mixed (both qualitative and quantitative)
predictor variables
THEN use multiple linear regression with indicator
variables
use multiple regression with interaction

IF 1 non-dichotomous outcome variable
AND 1 quantitative predictor variable
THEN use simple linear regression using indicator
variables for the outcome variables

IF 1 non-dichotomous outcome variable
AND 1 dichotomous predictor variable
THEN no technique available

IF 1 non-dichotomous outcome variable
AND 1 non-dichotomous predictor variable
THEN no technique available

IF 1 non-dichotomous outcome variable
AND >1 quantitative predictor variable
THEN use multiple linear regression with indicator
variables for the outcome variables AND
use multiple regression with interaction

IF 1 non-dichotomous outcome variable
AND >1 dichotomous predictor variable
THEN no technique available

IF 1 non-dichotomous outcome variable
AND >1 non-dichotomous predictor variable
THEN no technique available

IF 1 non-dichotomous outcome variable
AND >1 mixed (both qualitative and quantitative)
predictor variable
THEN use multiple linear regression with indicator
variables
use multiple regression with interaction

RULE SET 4

Inferences About a Populations

IF you have one sample from a single population
AND quantitative data
AND want to test mean equal to a specific value
THEN use a t-test

IF you have one sample from a single population
AND quantitative data
AND want to test population variance equal to a specific value
THEN use a chi-square

IF you have one sample from a single population
AND quantitative data
AND want to test proportion of a population equal to a specific value
THEN use a z-test

IF you have one sample from a single population
AND quantitative data
AND want to test whether the population fits a certain distribution
THEN use a Chi-square test OR
use Kolmogorov-Smirnov test

IF you have one sample from a single population
AND qualitative data
AND want to test proportion of a population equal to a specific value
THEN use a z-test

IF you have one sample from a single population
AND qualitative data
AND want to test whether the population fits a certain distribution
THEN use a Chi-square test OR
use Kolmogorov-Smirnov test

IF you have two samples from two independent populations
AND quantitative data
AND want to test for equal population means
AND have equal sample means
THEN use t-test with pooled deviations

IF you have two samples from two independent populations
AND quantitative data
AND want to test for equal population means
AND have unequal sample means
THEN use a large sample size and a z-test

IF you have two samples from two independent populations
AND quantitative data
AND want to test for equal population variances
THEN use F-test

IF you have two samples from two independent populations
AND quantitative data
AND want to test for equal population proportions
THEN use z-test

IF you have two samples from two independent populations
AND quantitative data
AND want to test for the difference between the
population means as equal to some value
THEN use t-test with pooled deviations (small samples) OR
use t-test with unpooled deviations

IF you have two samples from two independent populations
AND qualitative data
AND want to test for equal population proportions
THEN use t-test

IF you have two samples which are paired
AND quantitative data
AND want to test for equal population means
THEN use t-test

IF you have two samples which are paired
AND quantitative data
AND want to test for equal population variances
THEN use z-test

IF you have samples from more than 2 populations
AND quantitative data
AND want to test for equal means
THEN use ANOVA

IF you have samples from more than 2 populations
AND quantitative data
AND have equal sample sizes
AND want to test for equal variances
THEN use Hartley's test

IF you have samples from more than 2 populations
AND quantitative data
AND have unequal sample sizes
AND want to test for equal variances
THEN use Bartlett's test

RULE SET 5

Structure of the Data

IF variables = 1 in number
THEN cannot answer this question

IF variables > 1
AND data is quantitative and continuous
THEN use principal components OR
    use factor analysis OR
    use cluster analysis

IF variables > 1
AND data is quantitative and ordinal
THEN use non-metric multidimensional scaling

IF variables > 1
AND data is categorical
THEN no techniques are available

IF variables > 1
AND data is quantitative and both continuous and
    ordinal
THEN use non-metric multidimensional scaling

IF variables > 1
AND data is both quantitative and categorical
THEN no techniques are available

APPENDIX D

INSTRUCTIONS TO USERS

This appendix contains the written instructions to the users of the consultant system as well as the oral instructions given to each group before they participated in the experiment.

Instructions to participants - computer

Please read the System Instructions on the desks in front of you.

You may refer to these instructions as you use the system.

Each of you will be given 5 problems, one at a time. Attached to each of these problems is a form with questions which you are to answer about the problem. Please put the number from the sign up sheet that you signed as you entered the room on each problem. Feel free to write on the problem sheets or make notes on the problems as you go along. Do not be concerned if other people take more time or less time to solve the problems than you do.

Your task is to use the statistical mentor system to help you choose the correct statistic to analyze the problem. You should aanswer the questions presented to you by the system. Different sequences of questions will be asked of you depending upon what statistical question you are trying to answer. You will not always get the same number of questions or the same order of questions. Do not try to out-guess the computer. You know you are finished with a problem when you have chosen a specific statistic (for example the F-test or a t-test) and written it down on the answer sheet. Don't worry if the computer chooses a statistic of which you have never heard. That's one of the reasons this system can be useful.

The mentor system should be on the screen in front of you. You should see a screen which says 'Welcome to SAM'. If you do not have the welcome screen in front of you, please raise your hand. The system is controlled by function keys at the left of the keyboard, F1, F2, etc. Please take a few seconds to find these keys. If you cannot find these keys, please raise your hand. When you are instructed by SAM (the system) to press any key, you should press any key on the keyboard to tell the system that you are finished reading that screen and are ready to go to the next screen.

When you are finished with the each problem and have answered all of the questions about it, please raise your hand. I am timing to see how long each problem takes. Remember, the one person in the computer group who has the most problems correct in the shortest amount of time will receive $10. This means that accuracy is just as important as speed.

Please wait until I have passed out problem 1 to everyone before you begin. When you have finished problem 5, I will have a short questionnaire for you to fill out.

Are there any questions before we begin? If you have a question as you use the system, please raise your hand and I will come over to your desk.

Instructions to participant - textbook

Thanks for helping me test my statistical mentor system. You are the group who does not have to use the computer.

Each of you will be given 5 problems, one at a time. Attached to each of these problems is a form with questions which you are to answer about the problem. Please put the number from the sign up sheet that you signed as you entered the room on each problem. Feel free to write on the problem sheets or make notes on the problems as you go along. Do not be concerned if other people take more time or less time than you to solve these problems.

Your task is to choose the correct statistic by name to analyze the problem. For example, you may need a t-test or regression. You may use your textbook as much as you want to help you decide on the statistical technique to use. You do not have to do any calculations. In fact, there are no numbers in the problems on which to do calculations. You know you are finished with a problem when you have chosen a statistic and written in down on the answer sheet. You should also answer the other questions on the question sheet to the best of your ability.

When you are finished with the first problem and have answered all of the questions about it, please raise your hand. I am timing to see how long each problem takes. Remember, the one person in the textbook group who has the most problems correct in the shortest amount of time will receive $10. This means that accuracy is just as important as speed.

Please wait until I have passed out problem 1 to everyone before you begin. When you have finished problem 5, I will have a short questionnaire for you to fill out.

Are there any questions before we begin? If you have a question as you work, please raise your hand and I will come over to your desk.

Instructions for the Use of the Statistical Mentor

Using SAM (Statistical Analysis Mentor) is very easy. Most of the process is controlled by function keys, F1, F2, F3, etc. found on the left side of the keyboard. The definitions of these function keys can be found on the bottom of the computer screen.

When you are asked a question by the system, use the number keys above the keyboard or the letter keys to answer the question. Push the RETURN key when you are finished answering the question.

If at any time you see a message "USER ABORTED" on the screen followed by a 'B>' prompt, don't worry!! You simply hit a sequence of keys which took you out of the system. To get back into the system, type CHOOSE and push the RETURN key. You should see the WELCOME screen again.

To make the best use of the system in determining a statistical technique to use to solve your problem, use the detailed QUESTION screens at the beginning of the program. You can get to them from a REVIEW QUESTIONS function key or from the MAIN QUESTION MENU. Compare the problem you are trying to solve with the example questions on each of the QUESTION screens.

If you do not understand what all of the words mean, use the REVIEW DATA TYPES, especially for words like quantitative, qualitative, dichotomous, etc.

Make notes on the problem sheets. Think about what the different variables represent and how many of them are in each problem.

APPENDIX E

PROBLEMS USED IN THE EXPERIMENT

1. Property taxes are based on the assessed valuation of real estate. The higher the valuation the greater the tax on that property. In Boulder, Colorado, a controversy is taking place between some citizens and the city assessor. The citizen's groups claims that during a recent appraisal the residential property was increased in value by a larger average percentage than the commercial property. If the citizens group's claim is true, they will end up paying a greater relative share of property taxes than the owners of commercial property.

You are hired to study the situation. You selected a sample of 400 residential and 300 commercial properties and calculated the average percent increase for each class of property. What statistical technique are you going to use to tell the city assessor what to do about the citizen's group?

2. ABC Food Stores operates 2 stores in Tucson. Each store uses a different marketing approach. Store A caters to the high-income shopper by carrying specialty items and gourmet foods. The store's physical layout is spacious and much money and effort is spent keeping the floors and shelves clean. Store B is directed at lower-income shoppers and provides a good selection of basic food products.

ABC Management has recently performed a study of shoppers in the 2 stores. One manager has claimed that the dollar volume of purchases at store B is more consistent (has smaller variation) than the dollar volume of purchases in store A. The manager has the dollar volume for each of 45 days at both stores and the mean and standard deviation of this data. What technique should he use to determine if his claim is true or not?

3. A radio station in St. Louis recently decided to develop a theme song for the station. Three different songs were written and recorded by local musicians. The station manager wanted to be sure that the listening audience would like the theme song so she asked you, the marketing manager, to select some of the station's listeners at random, have them listen to the songs, and rate the songs on a 100 point scale.

You feel that age of the listener is something that should be controlled. You devised 3 age categories to match your listeners: yuppie (20 - 35), parents (36 - 59), and seniors (above 59). You selected 12 people, 4 in each category, and had then listen to the songs. Each person rated each sons. What technique are you going to use to decide if each group likes the songs an equal amount?

4. Automobile insurance companies have used age, sex, and marital status to determine rates for insurance. For example, single males under 25 years old are the worst risk and pay the highest rates. Females over 30 are considered the best risk and pay the lowest rates.

Recently, the Car Crash Insurance Company studied a sample of 1200 policy holders. The purpose of the study was to determine whether the number of accidents is really dependent upon age and sex. They collected data about the number of claims filed in the following categories: female vs. male; under 25, 25-49, 40-55, and over 55. Using this data, what technique should the insurance company use to decide whether age and sex are actually related to the number of claims filed?

5. The users of a desk top computer are either schools that use the computer for student training or businesses that use the computer in their day-to-day operations. An analyst for the We-Fix-It computer shop never has enough technicians available because some of the service calls seem to take longer than she expects when she sends out the technician. She wants to determine if there is any way she can predict ahead of time how long a service call will take. The only data she has to work with are the service call type (student or business), the number of minutes each of the different service calls took, and the age of the machine that was serviced (in months). What technique should she use to build her model of service call times?

These questions were used for each problem. Problems 1 and 2 asked only the first two questions while problems 3, 4, and 5 asked all four questions.

1. What generic question are you trying to answer in the problem? (Circle one answer.)

   a. Is there any relationship between the variables you are studying?

   b. Is there a difference between group responses for the groups you are studying?

   c. What individual or group prediction models can be built with the data you are studying?

   d. What inferences can be made about a population from your sample? Choose the sub-question below that you answered.

      d1. question about means

      d2. question about variances

      d3. question about proportions

      d4. question about distributions

   e. What is the structure of the data?

2. How many variables are in the problem? _____

3. List the variables in the problem.

4. What statistical technique did you choose? __________________

LIST OF REFERENCES

Anderberg, Micheal R. (1973) Cluster Analysis for Applications. New York: Academic Press.

Andrews, Franklin; Klein, Laura; Davidson, Terrence N.; O'Malley, Patrick M.; and Rodgers, Willard L. (1974) "A Guide for Selecting Statistical Techniques for Analyzing Social Science Data". Ann Arbor, MI: Institute for Social Research.

Beaujon, H. (1970) "An Interactive Graphical Display System for Illustrating Elementary Properties of Statistical Distributions". Master's Thesis, University of North Carolina.

Blum, Robert L. (1982) "Discovery and Representation of Causal Relationships From a Large Time-Oriented Clinical Data Base: The RX Project". From Lecture Notes in Medical Informatics. New York: Springer-Verlag.

Burton, R.R. (1982) "Diagnosing Bugs in A Simple Procedural Skill". Intelligent Tutoring Systems, edited by Sleeman and Brown. New York: Academic Press. pp. 157-184.

Caldwell, Steven B. (1983) "Combining Real and Generated Data in Lab Exercises Designed to Demonstrate Problems of Inference". Proceedings of the Section on Statistical Education, 1983. Washington, D.C.: American Statistics Association. pp. 53-56.

Chambers, John M. (1981) "Some Thoughts on Expert Software". Proceedings of the Computer Science and Statistics: 13th Annual Symposium on the Interface, edited by William F. Eddy. pp. 36-45.

Chase, W.G. and Simon, H.A. (1973) "Perception in Chess". Cognitive Psychology. January 1973, pp. 55-81.

Chervany, N.L.; Collier, R.O.; Fienberg, S.R.; Johnson, P.E.; and Neter, J. (1979) "A Framework for the Development of Measurement Instruments for Evaluating the Introductory Statistics Course". The American Statistician. February 1977, pp. 17-23.

Chi, Michelene, Glaser, Robert, and Rees, E. (1982) "Expertise in Problem Solving". Advances in Psychology of Human Intelligence, vol. 1, edited by Robert J. Sternberg. Hillsdale, New Jersey: Erlbaum Associates.

Cramer, S.G. and Cady, F.B. (1969) "Computerized Data Generation for Teaching Statistics". The American Statistician. December 1969, pp. 33-35.

Dawes, William S. (1983) "An Interactive Computer Program for Teaching Mid-Level Statistics". Proceeding of the Section on Statistical Education, 1983. Washington, D.C.: American Statistics Association. PP.57-59.

Egan, Dennis E. and Schwartz, Barry J. (1979) "Chunking in Recall of Symbolic Drawings". Memory and Cognition. March 1979, pp. 149-158.

Emond, W.J. (1982) "Some Benefits of Micro-Computers in Teaching Statistics". Computers and Education, vol. 6. pp. 51-54.

Feltovich, P.J. (1981) "Knowledge-based Components of Expertise in Medical Diagnosis". Technical Report #PDS-2, University of Pittsburgh Learning Research and Development Center. Pittsburgh, PA.

Fertuck, L. (1981) "The TREE System as a Teaching Aid in Statistics, Modeling and Business Courses". Computers and Education, vol. 5. pp.31-36.

Gagne, Ellen D. (1985) The Cognitive Psychology of School Learning. Boston: Little, Brown and Co.

Gale, William A. and Pregibon, Daryl. (1982) "An Expert System for Regression Analysis". Proceedings of the 14th Annual Symposium on the Interface of Computer Science and Statistics. New York: Springer-Verlag, pp. 110-117.

__________. (1985) "Artificial Intelligence Research in Statistics". The AI Magazine. Winter 1985. pp.72-75.

Gentleman, Jane F. (1977) "It's All a Plot". The American Statistician. November 1977, pp. 166-175.

Goldstein, I.P. (1982) "The Genetic Graph: A Representation for the Evolution of Procedural Knowledge". Intelligent Tutoring Systems, edited by Sleeman and Brown. New York: Academic Press. pp. 51-78.

Hahn, Gerald J. (1985) "More Intelligent Statistical Software and Statistical Expert Systems: Future Directions". The American Statistician. February 1985, pp. 1-8.

HaKong, L. and Hickman F.R. (1986) "Expert Systems Techniques: An Application in Statistics". Proceedings of the Fifth Technical Conference of the British Computer Society Specialist Group on Expert Systems, edited by Martin Merry. Cambridge, Great Britain: Cambridge University Press.

Harris, Richard J. (1975) A Primer for Multivariate Analysis. New York: Academic Press.

Jeffries, R., Polson, P.G., Razran, L. and Atwood, M.E. (1077) "A Process Model for Missionaries, Cannibals and Other River-Crossing Problems". Cognitive Psychology, vol. 9. pp. 412-440.

Johnson, Paul E. and Thompson, William B. (1981) "Strolling Down the Garden Path: Detection and Recovery from Error in Expert Problem Solving". Proceeding of the Seventh International Joint Conference on Artificial Intelligence, August 1981. pp. 214-217.

Johnson, Paul E. (1983) "What Kind of Expert Should a System Be?". The Journal of Medicine and Philosophy, vol. 8. pp. 77-97.

__________. (1984) "The Expert Mind: A New Challenge for the Information Scientist". In Beyond Productivity: Information System Development for Organization Effectiveness, edited by T.M.A. Bemelmans. The Netherlands: North Holland Publishing Co.

Larkin, Jill H. (1983) "The Role of Problem Representation in Physics". Mental Models, edited by Dedre Gentner and Albert L. Stevens. Hillsdale, New Jersey: Erlbaum Associates. pp. 75-99.

Larkin, J.H., McDermott, J., Simon, D.P., and Simon, H.A. (1980) "Expert and Novice Performance in Solving Physical Problems". Science. pp. 1335-1342.

Malhotra, A., Thomas, J.C., Carroll, J.M. amd Miller, L.A. (1980) "Cognitive Processes in Design". International Journal of Man-Machine Studies, vol. 12. pp. 119-140.

McCulloch, Charles E. (1983) "Problem Solving in Statistics". Proceedings of the Section on Statistical Education, 1983. Washington, D.C.: American Statistics Association.

Meeker, W.Q., Hahn G.J., and Feder, P.I. (1975) "A Computer Program for Evaluating and Comparing Experimental Designs and some Applications". The American Statistician, February 1975, pp. 60-64.

Morris, Carl N. and Rolph, John E. (1981) Introduction to Data Analysis and Statistical Inference. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Neter, J., Wasserman, W., and Kutner, M.H. (1983) Applied Linear Regression Models. Homewood, IL: Richard Irwin, Inc.

Norman, Donald A. (1982) Learning and Memory. New York: W.H. Freeman and Co.

Novick, Melvin R., Hamer, Robert M. and Chen, James. (1979) "The Computer-Assisted Data Analysis (CADA) Monitor". "The American Statistician. November 1979, pp. 219-220.
Novick, Melvin R. (1983) "Human Factors in Computer-Assisted Data Analysis". Proceedings of the Computer Science and Statistics: 13th Annual Symposium on the Interface. pp.33-36.

Rumelhart, David E. and Norman, Donald A. (1981) "Analogical Processes in Learning". In Cognitive Skills and Their Acquisition, edited by John Anderson. Hillsdale, NJ: L. Erlbaum Associates. pp.336-358.

SAS Institute, Inc. (1983) SAS Introductory Guide. Cary, N.C.: SAS Institute, Inc.

Seigel, Sidney. (1956) Non-parametric Statistics for the Behavioral Sciences. New York: McGraw-Hill.

Simon, D.P. and Simon, H.A. (1979) "A Tale of Two Protocols". In Cognitive Process Instruction, edited by Jack Lockhead and John Clements. Philadelphia: The Franklin Institute Press.

Sleeman, D. and Brown, J.S. (1982) Intelligent Tutoring Systems. New York: Academic Press.

Sleeman, D. (1982) "Assessing Aspects of Competence in Basic Algebra, edited by D. Sleeman and J.S. Brown. New York: Academic Press. pp. 185-200.

Southward, G.M., Urquhart, N.S., and Ortiz, M. (1983) "Computer Enriched Instruction of Intermediate Level Statistical Methods". Proceedings of the Section on Statistical Education, 1983. Washington, D.C.: American Statistical Association. pp. 6-9.

Stevens, Albert, Collins, Allan, and Goldin, Sarah E. (1982) "Misconceptions in Students' Understanding". Intelligent Tutoring Systems, edited by Sleeman and Brown. New York: Academic Press. pp. 13-24.

Tabachnick, Barbara and Fidell, Linda S. (1983) Using Multivariate Statistics. New York: Harper and Row.

Weber, E.S. (1984) User and Analyst Mental Models in the Implementation of Management Information Systems. Unpublished doctoral dissertation, University of Texas at Austin, Austin.

Wiser, Marianne and Carey, Susan. (1983) "when Heat and Temperature Were One". Mental Models, edited by Dedre Gentner and Albert L. Stevens. Hillsdale, N.J.: Erlbaum Associates. pp. 75-99.

Young, Kenneth C. (1982) "Micro-Computer Based Demonstration in Statistics". Proceedings of the Section on Statistical Education. Washington, D.C.: American Statistics Association.