

165
F-148

ABSTRACT

TEACHING STATISTICAL INFERENCE BY COMPUTER: PROBLEM SIMULATION FOR HYPOTHESIS TESTING AND EVALUATION

by Charles H. Frye

The purpose of this project was to develop and evaluate a computer-assisted instructional course in statistical inference (STAT). STAT is a laboratory course consisting of a sequence of twenty-four problems. Each problem presents an hypothesis-testing situation together with data that are sampled from simulated (randomly generated) populations. The student must translate these problems into appropriate statistical tests which he carries out, using the aids that are provided. The computer becomes a tool for efficient arithmetic manipulation of the data.

The student may choose his own method for satisfying the requirements of the problem, though he can request help from the computer if he needs it. During this time, the computer only responds to his various requests selected from a list of options--calculation assistance, procedure evaluation, help, etc.

Next, the computer tests the student over his work, giving feedback, diagnosing errors, providing help, or moving on, appropriately.

Charles H. Frye

Nine college student-volunteers took the course, doing as many problems as they could in fifteen hours of instruction. The data came from work records (teletype sheets), attitude scales, and performance test responses.

Motivation was generally high and attitudes were very positive. Comparison of performance levels indicated that the instruction was effective.

Much was also learned about various features of STAT that has implications for further developmental work.

TEACHING STATISTICAL INFERENCE
BY COMPUTER: PROBLEM SIMULATION FOR
HYPOTHESIS TESTING AND EVALUATION

By

Charles H. Frye

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

College of Education

1967

G 45683
8/22/67

ACKNOWLEDGMENTS

I am indebted to many people for the success of this project. First, I wish to acknowledge the co-sponsorship of the Learning Systems Institute at Michigan State University and System Development Corporation. Without their support in personnel, computer time, and line charges, the expenses would have been prohibitive. Joseph R. Rosenbaum and Samuel L. Feingold, both SDC employees, gave up several Saturdays to monitor subjects. This was over and above the working hours they spent helping to get the program ready.

I also wish to express my appreciation to the committee in charge, Dr. Norman Bell, Dr. David Krathwohl, Dr. Clesson Martin, Dr. Willard Warrington, and Dr. Ted Ward, for their guidance and contribution throughout the project. I especially want to thank Dr. Ward for assuming the responsibilities of thesis advisor when Dr. Krathwohl, the original thesis advisor, left to assume new responsibilities at Syracuse University. Dr. Ward provided valuable direction to the project and gave encouragement throughout.

Most of all, I would like to thank my wife, Pat, who in return for many hours of neglect, gave nearly as many hours of her time in typing all of the drafts, rough and finished.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF APPENDICES	vii
 Chapter	
I. INTRODUCTION	1
The Electronic Digital Computer	
CAI Defined	
Overview	
II. PURPOSE.	9
Increased Efficiency	
Motivated by Discovery	
Contributing Features of the Computer	
III. RELATED RESEARCH	16
Use of Computers for Instruction	
Relation of STAT to the Cited Projects	
IV. DESCRIPTION OF THE PROJECT MATERIALS.	27
Choice of a Language	
Choice of a Topic	
The Computer Program	
The Criterion Test	
The Pretest	
Description of the STAT Course	
The Computer Laboratory	
V. PROCEDURE FOR FIELD IMPLEMENTATION	74
The Survey Questionnaire	
The Attitude Questionnaire	
Sampling Population	
Schedule of the Experimental Sessions	
Telephone Interviews	

	Page
VI. EVALUATIVE DATA.	82
Guidelines For Data Collection	
Presentation of the Data	
Discussion of the Data	
VII. CONCLUSIONS AND IMPLICATIONS	128
Specifications For a CAI Author-Language	
Author-Languages Under Development	
BIBLIOGRAPHY.	156

LIST OF TABLES

Table	Page
1. Organization of the Twenty-Five Problems in STAT.	42
2. Summary Statistics for the Criterion Test Scores	95
3. Analysis of Criterion Test Scores.	95
4. Survey Questionnaire Results	95
5. Number of Problems and Problem-Types Com- pleted	97
6. Summary of the Guideline Data.	98
7. Interaction Among the Twenty-Eight Attitude Questionnaire Items.	102
8. Ranking of the Twenty-Eight Attitude Questionnaire Items.	102

LIST OF FIGURES

Figure	Page
1. Use of Option 10	54
2. Computer-Student Interaction at the Start. .	60
3. Sample of Lesson Termination and Student Records.	60
4. Flow Chart of the STAT Program	69
5. Sample Item From the Survey Questionnaire. .	75
6. Question I From the Attitude Questionnaire .	76
7. Means Employed to Satisfy or Avoid the STAT Evaluation Questions	108

LIST OF APPENDICES

Appendix	Page
A. Student's Guidebook to STAT.	160
B. Student's Reference Sheet for STAT Options .	170
C. Criterion Test.	173
D. Pretest.	177
E. Instructor's Guidebook to STAT	180
F. Illustrative Problem.	185
G. Survey Questionnaire	192
H. Attitude Questionnaire.	195
I. Table of Pretest Scores	199
J. Table of Criterion Test Scores	200
K. Sample Teletype Output from STAT	201
L. Table of Survey Questionnaire Results . . .	204
M. Data Used for Correlation Coefficients . .	205

I. Introduction

Many educators are making inquiry into the use of digital computers for instruction. Most large universities either have a computer complex located on their campus or have regular access to one. For example, the Western Data Processing Center located on the campus of the University of California at Los Angeles has invested more than eight million dollars in computing equipment. Similarly, Michigan State University has recently invested more than one and one-half million dollars in computer hardware. Suppes (1966) reports that by mid-1965 more than 800 computers were in service on American university campuses for which their budget reached \$175 million. In each case, the equipment represents the latest and most sophisticated equipment that is available on the commercial market.

Since World War II, education has felt the impact of a series of technological advances such as motion picture equipment, tape recorders (both audio and video), and television. These devices were used so successfully in military training that their potential for public education was obvious. The video tape recorder, the most recent innovation of those mentioned, is also attracting a great deal of attention. Each of these new devices have had associated with it a broad base of research conducted by

both commercial and academic investigators. That research activity continues today. Much of it is reported in the Audio-Visual Communication Review, published by the Department of Audio Visual Instruction of the National Education Association. Studies include the entire range from optimal equipment design to comparisons in teaching effectiveness.

The Electronic Digital Computer

The electronic digital computer ranks among the most recent of the technological advances. Many predict that it will bring about a greater revolution in education than any of the former devices. The history of the digital computer is brief. In 1946, the first all-electronic computer, the ENIAC, was constructed at the University of Pennsylvania and five years later the UNIVAC I appeared on the commercial market. After only two decades of existence, digital computers, large and small, are being used widely. Nearly every American citizen is in some way affected by one, whether through banking, buying an airplane ticket, or filing an income tax statement.

Though the computer originated at an educational institution, its development, like other innovations, was carried on largely by other interests.

A computer is a device which can receive, store, process, and transmit large quantities of information very quickly. Most computer operating times are measured in micro-seconds so, for computer-human interaction, the

time delay between an input message and the computer response seldom exceeds a few seconds. How the computer composes and displays the response to an input message is completely determined by a set of computer instructions called a program which is prestored in the electrostatic memory of the computer. The program specifies every action to be taken on an input message and determines every detail of the output response. Inputs can be obtained from human sources by means of such devices as typewriter keyboards, punched cards, tape, both paper and magnetic, and light pens. Once the inputs have been entered into the computer, they can be sorted and manipulated in almost any desired manner, limited only by the skill of the computer programmer and the storage capacity of the machine. The results of the processing can then be used to automatically initiate a transmission of information back to the human source. That information might be typed, printed, punched on cards, drawn graphically on a cathode ray tube, projected from film, displayed in lights, played from an audio tape or any combination of these. Computers have been programmed which allow persons to type questions in usual English grammar and if the information requested is part of the repertory of the program, the computer will find and type an intelligent reply back to the person (Green, 1963). Computers have also been programmed to ask questions and then evaluate the responses that are given and appropri-

ately modify subsequent interaction on the basis of that evaluation (Coulson, 1962).

Because computers have these capabilities which other forms of technology have not provided, many believe that these devices have more potential for educational use than any of the previous innovations.

Computer systems that are equipped to control instructional processes are known as computer-assisted instructional (CAI) systems. The complete system includes the computer together with all of the equipment required for computer-human interaction. It also includes a set of computer programs which make possible the preparation and subsequent execution of the lesson material.

Many informed people are predicting that CAI will be implemented in public education in the near future. They do not expect the computer to assume major responsibility for instruction for some time, if ever. They are not expecting computers to displace other innovations or replace teachers but rather to be used in those applications where their superiority is unquestioned.

It is reasonable to assume that computers will be introduced gradually. Perhaps schools will experiment with one or two typewriter terminals connected to a remote computer installation. Probably no school will have more than one classroom equipped for computer instruction within the next decade. For this reason,

computer-assisted instruction is likely to be limited to those subject matter areas where it demonstrates superiority over other methods.

CAI Defined

Instructional uses of the computer include a variety of applications. Dr. Glen Culler uses it on the Santa Barbara campus of the University of California to illustrate his lectures by displaying complex computer-generated graphical displays for the students. Six Queens High School students in New York used a computer from their home to help them with arithmetic homework. They used a special set of buttons which had been attached to their telephones. Computers are also being used for military and industrial training of key punch operators stenotype and other skill tasks. It is important, therefore, to define what is meant by computer-assisted instruction. Silvern and Silvern, (1966) define it in this way:

. . . the term CAI should be reserved for those particular learning situations in which a computer contains a stored instructional program designed to inform, guide, control and test the student until a prescribed level of proficiency is reached.

He suggests that other instructional uses be properly identified such as: CAT, for computer-assisted teacher, where the computer is being used as an aid in demonstrating problem solutions, and CAS, for computer-assisted student, when the computer is being used as a

tool to assist in problem solving. "To be CAI," he says, "the computer must actually instruct the student. . ."

Dorn (1967) suggests the term, computer-extended instruction (CEI) for those applications where the computer is being used as an instructional tool.

Overview

The following pages will describe the development and evaluation of a computer program for teaching statistical inference (STAT). There will be an attempt to show that the computer has a distinct advantage over other types of teaching devices for this kind of subject matter. The STAT program is a laboratory course in inferential statistics composed of twenty-five hypothesis-testing problems. Each problem is divided into three parts: (1) The problem statement and data generation, (2) data manipulation by the student and (3) evaluation. Problems are repeated with new data as necessary to bring the student's achievement up to a specified level.

The development of the STAT program represents an attempt to cause the computer to evaluate the student on the basis of his current work rather than judging him against prestored information. STAT "determines" which statistical procedures are appropriate and what responses are to be accepted as correct directly from the data that are randomly sampled from simulated "populations." Correct answers vary not only among problems but also for repeated problems.

The "discovery" method of learning influenced the development of the part of the STAT program where the student is left on his own to investigate the properties of the samples. He does this with the aid of a statistical procedure library and sophisticated calculator, both of which are built into the STAT program. The student may also request help from the computer to guide him in his investigation.

Nine graduate students took part in the evaluation of STAT. These were selected on the basis of their performance on a screening test that was devised for that purpose. They also recorded information on a survey questionnaire concerning their entering proficiency level over STAT-related subject matter. After going through the course, they responded to a criterion test and an attitude questionnaire. Twelve other comparable graduate students also responded to the same criterion test to provide a basis against which to assess the performance of the computer subjects. An important part of the data that were analyzed came from the teletype sheets which contained virtually all of the students' work. These data were related to the survey questionnaire responses according to fourteen guidelines that had previously been established. The results of this analysis provide a basis for assessing the success of the discovery approach as it was used here.

Finally, the report describes developmental work which is presently in progress that provides a more convenient communication language for a lesson author who desires to prepare a computerized lesson that incorporates methods like those which were used in the STAT course.

II. Purpose

The prime objective of this project was to explore the potential of a CAI system, especially those areas of CAI for which the computer is particularly well-suited and in which it holds a distinct advantage over all other instructional media.

Instead of asking what existing tasks the computer can take over, one can more profitably ask what new potential can be introduced by the computer. What things can the computer do better or faster than have been done before? What can be done now that was not practical before? What features of a computer are relevant to educational purposes? These are the questions that this project has undertaken to explore.

Increased Efficiency

One desirable outcome would be increased instructional efficiency. The word, efficiency, is taken in the context given it by Glaser (1964) as a description of learning that "leads to a high level of performance in the transfer situation." Efficiency is also meant to imply an economy in the amount of time expended on the learning task. To be an efficient method, the learner must acquire more useful knowledge per unit of time from a given set of materials using this computer-based course of instruction than he would otherwise.

Two features of the computer program suggests that it will be economical in terms of learning time: (1) the format of the program is basically multiple choice and (2) required calculations are performed faster and easier compared to using a desk calculator. Research in programmed instruction has shown that multiple choice response mode is not only adequate but requires less time (Silberman, 1962; Coulson and Silberman, 1960). It will be seen from the student's guidebook for STAT (Appendix A) that the procedures are activated in most cases by simply entering a one or two-digit code number selected from the student's reference sheet (Appendix B). Also, attention has been given, as Coulson (1964) has suggested, to the improvement of communication between the computer and the human. Relatively few of the common format and symbolic restrictions usually accompanying computers are imposed on the user.

Perhaps the greatest saving in time is due to the computational speed of the computer. The computer calculates the result for complex statistical procedures so fast that the student is seldom aware of a delay after his request has been inserted until the computer starts printing the reply.

Motivated by Discovery

Motivating Elements

Assuming that the computer user can accomplish more in less time than without the computer, what is there to

indicate that he will be motivated to use his additional time in a meaningful way? The rationale for this depends on the degree to which the author was successful in including elements of "discovery" in the computer course. The term "discovery" has had instructional significance in the past few years through the work of such men as Beberman (1958), Kersh (1958), Suchman (1961), Finlay (1960), and others. Bruner (1964a) defines it as, "a matter of rearranging or transforming evidence in such a way that one is enabled to go beyond the evidence so reassembled to additional new insights." Bruner suggests four benefits obtained by learning through discovery, the second of which is, "the shift from extrinsic to intrinsic rewards." Elaborating this point, he hypothesizes:

. . . to the degree that one is able to approach learning as a task of discovering something rather than "learning about" it, to that degree will there be a tendency for the child to carry out his learning activities with the autonomy of self-reward or, more properly by reward that is discovery itself.

.
I am suggesting that there are forms of activity that serve to enlist and develop the competence motive, that serve to make it the driving force behind behavior.

Kersh (1964) states, "the present results leave no doubt that there is a tendency for interest to accrue as a result of learning by discovery." He found that while a no-help discovery approach yields high motivation, a directed or rote learning approach produces the greater achievement gain in a given amount of time (Kersh, 1958;

Kersh, 1964). Getzels (1964) states the existence of an "optimum level of activation and stimulation." Above this level, the learner tends toward frustration and below it, he tends toward boredom. Kersh (1964) also suggests a "happy medium" between discovery and directed depending on the level of retention and transfer desired.

In introducing Bruner's chapter, DeCecco (1964) concludes that Bruner is, "attempting to gather evidence to show that each learner, in one sense, must be his own programmer."

In the same chapter, Bruner (1964a) states another hypothesis:

"It is my hunch that it is only through exercise of problem solving and the effort of discovery that one learns the working heuristic of discovery, and the more one has practice, the more likely is one to generalize what one has learned into a style of problem solving or inquiry that serves for any kind of task one may encounter--or almost any kind of task."

Discovery Elements in STAT

To clarify the relationship between the foregoing literature and the present project, notice Getzels' (1964) hierarchy of problem types. It is suggested that according to his eight classifications, which are arranged in an ascending order of complexity, the problems used in this course fall generally at level six:

6. The problem itself exists but remains to be identified or discovered (as in 4 and 5) and there is a standard for solving it, once the problem is discovered, known to the problem-solver and to the others (as in 1).

The practice exercises used in this computerized STAT course have several of the necessary elements to support a discovery interpretation. They coincide with Getzels' description of a discovery-type problem. The student must identify a testable hypothesis in each exercise. He knows that this is not a problem which has been solved before. No answer book exists. Many exercises allow the student to select the procedure to use. He may choose his own strategy which, if correctly used, will be positively reinforced. The "happy medium" between pure discovery and pure rote learning is provided for here by letting the student specify the kind and amount of help he wants when he wants it. It is presented to him only by his request. In this sense, the student is his own programmer. Enough help is available in the program to lead the student to an acceptable answer for each evaluation question. Thus, frustration should be avoided. "Knowledge of results" is given at selected points in the program, particularly where the student has made an error that needs correcting. The wording of the message he receives and the associated inconvenience should be negatively reinforcing as well as informative. For example, if the student requests a product-moment correlation on two groups of unequal size, a result will be calculated and printed but along with it the following message will be printed, "OPTION ILLEGALLY USED." Bruner (1964b) says that knowledge of results is useful or not

depending upon, "when and where the learner is able to put the corrective information to work." Hopefully, this information is immediately useful to him.

Thus, the combined force of these elements of the discovery method helped to sustain a very high motivation level over such a long duration that novelty alone was not a sufficient explanation of the subjects' perseverance.

Contributing Features of the Computer

An attempt was made to identify a set of capabilities that gave the computer a distinct advantage over other instructional systems. Four such capabilities were identified:

1. dynamic information storage and retrieval,
2. storage of procedures (set of manipulations) that can be utilized at will,
3. extensive branching, and
4. rapid computation.

Most instructional methods require information storage and retrieval. The information may be stored in a book, on film or in a computer.

For the sake of overall efficiency, it was decided that for this project, the computer would not be used to store large quantities of reading material. That could be done much more economically in a book. Rather, the goal was to use the computer's information storage and retrieval capabilities in a way that books and other media were unable to duplicate. The computer storage was

to be used for five main types of programmed material: (1) textual material for those cases where the line composition is changed dynamically as the course progresses, (2) textual material whose nature is such that it should be presented at a given point in the course, (3) programmed sets of instructions to do certain manipulative tasks when so requested by appropriate input messages, (4) programmed instructions that would control the order and presentation of the course and (5) records of student performance as they take the course. Very little instructional material in textual form went into the computer.

The choice of statistical inference as subject matter made the second, third, and fourth capabilities especially useful. The reasons for this will be discussed later.

In summary, the purpose of the project was to explore those particular characteristics of a CAI system which suggest a superior instructional capability and efficiency over existing methods and media.

III. Related Research

A nationwide effort to find a way to accelerate the instructional process was launched with Sputnik in 1957. Educators were suddenly made aware of the deficiencies in the curriculum, especially in the sciences and related fields. Teaching methods were considered to be outmoded and unable to keep pace with "space-age" demands.

Soon after Sputnik, B. F. Skinner's article, "Teaching Machines" (1958), appeared in a scientific magazine and the fire caught. Developing a machine that could automate parts of the instructional process looked like a panacea that would enable the schools to bridge the technological "gap." Within five years, hundreds of research projects were carried out. Scores of professional educators became "teaching machine" and "programmed learning" experts. Conferences were devoted to the subject. A flood of programmed learning articles appeared in many magazines and journals--both popular and professional.

Commercial organizations saw good prospects of a new and lucrative market. Several textbook publishers began marketing self-instructional programs, mostly in booklet form. Concurrently, many teaching-machine devices became available. Though many of the machines were very sophisticated, (and expensive), relatively few had more than a dozen hours of instructional material prepared for them.

Consequently, the booklet form became the most widely-used version of programmed instruction.

Research efforts have not verified the superiority of programmed instruction over other instructional approaches. After reviewing many research projects, Silberman (1962) reported several conflicting findings.

The enthusiasm over programmed instruction has since diminished. L. C. Silvern (1966) reported that more than 400 corporations which entered the programmed instruction field between 1960 and 1965 in the United States have since become bankrupt or have been absorbed by a holding company. Several of the basic concepts of programmed instruction have survived and are influencing many areas of education. Among these concepts were: (1) the instructional objectives must be specified in behavioral terms, (2) the concepts to be taught are broken down into small steps, usually called frames, each of which normally presents instructional material and elicits a response from the student that is relevant to the objectives, (3) sequences representing several levels of difficulty are included in the lesson so that the sequence of frames that is presented to the student will depend, at least in part, on his pattern of responses to previous questions, and (4) the entire lesson must undergo several evaluation-revision cycles to insure that the lesson does adapt to the individual needs of the learner and produce the desired level of performance. Mager (1962) applies these

principles not only to programmed instruction but to all teaching.

Use of Computers for Instruction

The potential of using a computer as an adaptive teaching machine soon became obvious. It could monitor student response and be more adaptive than any of the booklets or teaching machines in that changes in sequence could be made to depend on a history of the student's performance. Lesson changes could be made without the usual production problems. Many kinds of performance data could be automatically recorded. A 1961 survey (Kopstein and Shillestad, 1961) cited five projects that were using digital computers as teaching machines. Also in 1961 a conference on digital computer applications to education was held and the proceedings that were published include several more computerized instructional projects (Coulson, 1962). Among them is a description of a statistics course (Uttal, 1962) developed as a part of the IBM Research Computer Teaching Machine Project by Grubbs and Selfridge (1964). This project will be discussed in more detail later.

As computer utilization increased in business, schools also became interested. In addition to the university computing centers for research, computers were being installed to do other kinds of tasks such as keeping inventories, stock control, student scheduling, grade reporting, bus routing problems, bookkeeping, etc.

Goodlad and others (1966) present a survey of these kinds of computer uses. With computer costs being justified on other grounds, computer time could be made available for instructional purposes.

The literature pertaining to CAI is still relatively sparse. Computers were first used for instruction less than a decade ago. High equipment costs have since tended to restrain widespread use of CAI even on an experimental basis. The few CAI projects that have been carried out differ widely from one another. Because of this, resemblance between the computerized instruction used in this project (the STAT program) and most other known CAI projects is very slight.

Many instructional projects can now be identified in the general area of computerized instruction. Extensive descriptions of existing CAI systems and materials have been prepared by Dick (1965), Hansen (1966), Hickey and Newton (1966) and Zinn (1966). Several projects were selected from these as having certain features in common with the STAT course, including: (1) calculation assistance features, (2) course content area and (3) approach used for instruction.

Calculation Assistance Similarities

Kemeny and Kurtz (1966) developed a computer language called BASIC. Its purpose, like STAT, is to allow students to do statistical problems on the computer and provide them with a battery of statistical aids. However,

it is not, in any sense, an instructional course. It is rather a computer language (somewhat like FORTRAN) especially adapted to statistical problems and is less difficult to learn than conventional computer languages. BASIC and STAT both can be used to solve statistical problems in a CEI (Computer Extended Instruction) manner. The essential difference is that BASIC, though it has a more extensive computational capability, does not contain an instructional sequence.

Falkoff and Iverson (1966) have invented an instructional language called APL (for A Programming Language). APL, like BASIC, is a CEI system used by the student for solving algebraic problems. APL is similar to the calculation assistance portion of the STAT program which allows certain similar algebraic operations to be performed as required by the instructional sequence. APL also has a much more sophisticated set of algebraic capabilities than STAT because of its more general CEI application and like BASIC, contains no sequence of instructions.

Course Content Similarities

Grubbs and Selfridge (1964) prepared a CAI program to teach descriptive statistics. The course was prepared in the COURSEWRITER language (Maher, 1964), a computer language designed to facilitate CAI lesson preparation. The Grubbs and Selfridge course requires a specialized textbook. The computer guides the student through the text on the basis of his performance in problem-solving

sessions. The student uses a modest calculation capability called DESCAL, a part of the COURSEWRITER system which assists the student by performing indicated arithmetic operations in typed expressions. The computer also automatically records the student's performance as he proceeds through the lesson. Individualized lesson sequencing is made possible by associating "counters" with pertinent responses and altering the course sequence periodically on the basis of the accumulated contents of the counters. These techniques are fully described in the COURSEWRITER II manual (IBM, 1966). Apart from the performance records and use of counters, the course would operate much the same if the student instead worked on identical exercises in a booklet with the aid of a desk calculator, assuming he could be counted on to be honest in comparing his answers with the true ones and then to follow directions appropriately.

Method Similarities

Bushnell (1963) makes a distinction between "machine directivity" and "machine docility" to indicate whether the computer or the student is directing the activity. He notes that existing computer-based instruction wholly or at least partially directs the student. Allowing the student to request help and review is a form of machine docility. Bushnell proposed the adding of information retrieval capabilities to the instructional program to allow student-directed exploration of the subject matter.

Machine docility characterizes at least two projects and, because of this feature, they are relevant to the present effort. The first is the PLATO II system at the University of Illinois (Bitzer, Lyman and Easley, 1966; Bitzer, Braunfeld and Lichtenberger, 1962). PLATO II is a CAI system that generates informational displays on an "electronic blackboard" cathode ray tube. The displays can originate from the keyboard, computer or microfilm or a combination of all three. These displays are put together in an instructional sequence in such a way that the student responds frequently and his responses become a part of the total display. The computer scores his reply and moves him to the next place in the sequence. The computer is also responsive to "help" and "review" requests made by the student upon which appropriate remedial sequences would be initiated. When the particular deficiency has been corrected, the student presses the "aha" button and continues.

In an even greater measure is Wallace Feurzeig's (1964) Mentor system responsive to student behavior. This CAI system was designed to assist in the training of medical students for making diagnoses. It has been described as a sort of Twenty-Questions game toward pinpointing the exact ailment of a simulated patient. Using a list of acceptable vocabulary words, the medical student can ask for various types of physical characteristics of the patient as well as laboratory reports and, on

the basis of this information, he makes his diagnosis. The computer not only confirms his diagnosis but evaluates the body of information on which the diagnosis was based. Logical inconsistencies and information gaps are pointed out to the student and the process continues until a correct diagnosis has been made from an adequate supply of information.

Relation of STAT to the Cited Projects

The STAT course provides for a great deal of student-directed activity. The student has complete freedom in choosing his method and statistical aids while solving a problem. He initiates the computer action. Only during evaluation does the computer request responses from the student.

This student-directed activity primarily consists of a large battery of statistical procedures which he may use and a calculation assistance capability that gives him the ability to evaluate statistical formulas quite easily. It shares both of these features with BASIC and the latter with APL. However, both APL and BASIC allow much more complex kinds of mathematical operations because of their general-purpose nature. The calculation section of the STAT program essentially includes only those features needed for the lesson. Either BASIC or APL could be used to essentially duplicate any statistical operation or calculation feature in STAT but neither one is geared to any specific sequence of instruction. For example, a

person using APL or BASIC would be expected to supply the data to be manipulated by the program. In STAT, the current problem not only supplies the data, but also pre-stores them in such a way that they are appropriately referenced by all of the calculation features and, finally, calculates the necessary results from them to evaluate the student's work. This will be explained in more detail later.

Another instructional program in statistics was cited (Grubbs and Selfridge, 1964). On the surface, these projects may seem to have much in common. That this is not the case can be seen in at least four basic differences. First, the Grubbs project is essentially "machine directive." The student controls little more than the pace at which the lesson progresses. Second, the STAT course content is inferential statistics, a different topic than descriptive statistics. Whereas descriptive statistics implies the teaching of facts and formulas, inferential statistics implies the testing of hypotheses. Third, the problem exercises in the Grubbs course are completely specified including the precise numbers in the samples. Therefore, the answers are also known and written into the course. Different students working given exercises are working with the same lists of numbers. STAT constructs the samples for each problem from appropriately scaled random numbers. Students never work the same problem even if they repeat the exercise.

Fourth, unlike the Grubbs program, the rules for arriving at a solution for a hypothetical problem in the STAT course are not predetermined. Alternate approaches may be valid in which case either one is given credit.

The PLATO II project incorporates two features that are especially relevant to the STAT project: (1) it treats numbers as numeric values and performs arithmetic operations in the process of evaluating answers and (2) it provides a degree of "machine docility" in that the student can request a change of sequence by asking for a review. To understand the importance of the first property, the COURSEWRITER system used by Grubbs must consider these four numbers, 39, 39., 39.0, and 39.00, as for distinct correct replies to the problem, $13 + 26 = ?$. If additional trailing zeros are anticipated or if the acceptable answer falls within some interval, the answer set becomes prohibitively long. PLATO II and STAT are both equipped to interpret these answers numerically.

Feurzeig's project perhaps is most closely related to STAT in that the instructional strategies of the two programs are much alike. Both require the student to take the initiative and search out the relationships that exist and to make a decision based on that information. The course content, however, is completely different as can be seen from the description above.

One other project, not mentioned above, that shares most of the features of STAT and includes many more in

addition to them, is PLANIT (Feingold and Frye, 1966). Because PLANIT is a direct outgrowth of STAT, discussion is deferred until the last chapter.

IV. Description of the Project Materials

Several instruments were developed specifically for this project. Among them were two tests, two questionnaires, guidebooks, and a computer program. The latter embodies the instructional program. The other materials were developed for explanatory and evaluation purposes.

Choice of a Language

The STAT computer program consists of a set of coded, symbolic instructions that have been recorded on magnetic tape. When symbolic code is transferred or "loaded" into the core memory of a computer, the symbols are interpreted by the computer as step-by-step instructions to be performed in sequential order. Simplified coding schemes have been devised for various purposes. These are called programming "languages." Some of these languages are used widely among computer programmers. For example, COBOL is a business-oriented language. It might be used by a bank. On the other hand, FORTRAN and JOVIAL are scientific languages and are commonly used for research purposes. These are very powerful and flexible languages in terms of the kinds of complex computer operations that can be programmed rather easily. Another programming language has been more recently devised called COURSEWRITER (Maher, 1964; IBM, 1966). COURSE-

WRITER allows one to program certain kinds of instructional courses for the computer. However, COURSEWRITER was not suitable for use in this project. This will be discussed in more detail later. The JOVIAL language was used for this programming effort. It was both adequate for this task and available on the computer that was to be used. JOVIAL is a very powerful language. It is also fairly technical and requires some orientation before one can use it satisfactorily.

Choice of a Topic

Having obtained the use of a computer and having become familiar with the JOVIAL language, the next step was to decide on the topic for the instructional program. Psychological statistics was adopted as the topic.

The advantages that accrued to this choice included (1) computer dependence, (2) computational requirements, (3) logical presentation, (4) clarity of the instructional goals, (5) definable universe of acceptable behaviors and, (6) breadth of application. These will be discussed briefly in order.

Computer Dependence

Psychological statistics has already become computer-dependent to a large extent. Computers are being widely used for solving statistical problems. Nearly every statistical procedure commonly used today is available on some computer. Authors of textbooks are beginning to include computer program suggestions along with the

presentation of statistical procedures (Cooley and Lohnes, 1962; Harman, 1960). The largest segment of computer time in most university installations is devoted to handling statistical problems. It seemed highly desirable therefore, to use statistics as the subject matter for the instruction and obtain the added benefit of familiarizing the student with the operation of the computer.

Computational Requirements

It is well known that one of the things a computer does best is routine calculation. Though humans are entirely capable of performing these computations, when an excessive number are required the task is usually relegated to a computer. Most statistical procedures require extensive computation which largely explains why they are computer-dependent. Because of the computational power of computers, using statistics for the course content served to exploit the available potential.

Logical Presentation

Within most statistical problems it is possible to identify a set of subproblems. For example, the calculation of means and variances is often implied in the calculation of other procedures. Because of this property, it is possible to evaluate a student's progress toward the final solution. Through proper diagnosis and remediation, the student can be guided to the desired performance objective.

Clarity of the Instructional Goals

The objectives for a course in psychological statistics are readily definable in behavioral terms. The student is usually required to translate his particular research problem into an appropriate statistical test, carry the test through to a solution, and make a decision based on the solution, subject to the constraints placed on the problem. For example, one might be asked to test the mean difference between two independent groups with the constraint that the alpha error (probability of false rejection of an equal-means hypothesis) is less than five percent. He translates this problem into a t-test of uncorrelated means, and computes the statistic from his data which he uses to retrieve the proper table entry. Finally, he relates the table entry to the constraint and makes a decision regarding the significance of the difference that exists between the means of the two groups.

The fact that the desired student behavior can be specified so precisely, facilitates both the preparation of the instructional program and the evaluation of the students' performance. Both qualities are desirable for an investigation such as this.

Alternative Solutions are Acceptable

Several alternative methods are available for solving many of the statistical problems. Some are equally acceptable in certain situations. One of the advantages of using a computer for instruction is that it

can be made to evaluate the student's performance in a way that allows the student to obtain his solution by either acceptable method. For example, how severely should the sampling distribution deviate from normal before non-parametric tests are to be preferred over the parametric tests? Not even the "experts" always agree but a computer can rapidly obtain both solutions for the data and compare the discrepancy. In many cases, which method to use becomes a value-judgment for the student to make. Because the universe of extant overlapping statistical procedures is presently quite small, it is possible to allow the student a great deal of freedom in solving his problem and yet anticipate each of the possible approaches he might make. The evaluation would be considerably more difficult if the computer was programmed to ask, "What caused the Civil War?"

Breadth of Application

The topics that are usually presented in an introductory course to psychological statistics at the graduate level are quite uniform with few exceptions among the universities. The usual sequence includes probability distributions, sampling, estimators, distribution of means, tests based on the normal, t, chi-square, and F distributions, correlation, regression, non-parametric statistics and analysis of variance. Often, the sequence is a two-term course with the second term devoted primarily to analysis of variance. At Michigan

State University, an example of such a sequence is found in the education courses, "Quantitative Methods in Educational Research," denoted ED 969A and ED 969B.

The fact that essentially the same course sequence is taken by so many graduate students implies greater use for the finished program and an adequate supply of appropriate subjects for tryout purposes.

The Computer Program

The instructional program (STAT) was written in JOVIAL, a FORTRAN-like computer language for the IBM AN-FSQ-32 computer (Q-32) located at System Development Corporation (SDC) in Santa Monica, California. This computer is presently equipped to serve in excess of fifty simultaneous users. The users communicate with the computer via several teletypewriters, some of which are located on the premises of SDC and the remainder are connected to the computer through telephone-line channels.

The Q-32 at SDC is a fully time-shared computer, one of the few in this country. More will be said about this computer and the teletype laboratory later.

The programming of STAT (i.e. preparation of the STAT course for computer presentation) was begun during July, 1964, at SDC where the writer was employed as a participant in their Summer Student Associateship program. The programming effort was interrupted in September when he returned to school at Michigan State University but resumed again in January, 1965, over teletype lines from

the campus to SDC in Santa Monica and continued until May of that year. Mr. Samuel Feingold and Mr. Joseph Rosenbaum, both SDC employees, assisted with technical advice and the handling of those things that had to be done in Santa Monica. Much long distance toll time was avoided because they handled the magnetic computer tapes and printed program listings at the computer location and also spent several hours "trying out" the program so they could report segments that did not function properly. As a result of this experience, many insights were gained pertaining to the remote use of a computer.

In addition, Rosenbaum and Feingold monitored the group of subjects who took the course in the SDC laboratory. Their services were indispensable since the cost of having subjects operate remotely from Michigan would have been prohibitive.

When the programming was completed, STAT contained twenty-five statistical inference problems, a means of generating sample data for each problem, a battery of statistical procedures for investigating the properties of the data, a powerful computational aid and a set of evaluative questions associated with each problem. An executive program controlled the order of presentation of the problems, repeating some problems with new data under certain conditions. It controlled the computer-student interaction, executing requested procedures, accepting answers, providing feedback and keeping performance

records. The availability of certain statistical procedures to the student was also controlled by the executive program which limited access to groups of procedures dynamically based on one's performance on previous problems. The algorithm used by the executive program for controlling the problem sequence will be described later.

The work of preparing the STAT program was initiated by establishing a few specific instructional objectives for the course. The objectives were five-fold.

Objectives of the Instruction

Each student was expected to develop:

1. the ability to translate an experimental problem into appropriate statistical tests,
2. the knowledge necessary to carry through the tests,
3. the ability to draw conclusions which could be supported by the data,
4. the ability to perform the above three operations for univariate or bivariate hypothesis-testing situations and,
5. sufficient generality to test both normal and non-normal distributions.

The course was organized around the twenty-five statistical problems in the STAT program. Student performance on the problems were operationally related to the five objectives in the following way: For each type of problem situation contained in the program, the students would be expected to (1) choose an appropriate statistical test, the results of which would provide a sound

basis for accepting or rejecting the hypothesis, (2) correctly carry through the test, (3) accurately read the appropriate tables and, (4) make an appropriate decision based on a given level of confidence.

The problem set contains both univariate and bivariate problems. The sampling distributions used for the problems include normal, skewed normal, rectangular (uniform) and binomial distributions. The bivariate problems include both independent and matched samples.

The Criterion Test

A criterion test (Appendix C) was developed by the writer for evaluating the student's attainment of these objectives. The test items describe some novel hypothesis-testing situations and require the student to indicate which statistical procedures he would use for the test. It also contains questions pertaining to the correct application of selected tests and require the student to make certain decisions and relate those decisions to the outcomes which are provided. To do this last exercise correctly, it is necessary for the student to make satisfactory use of certain of the statistical tables.

For example, questions 10, 11 and 12 of the criterion test are related to the following stated situation:

You suspect that your class of 30 people has a very unusual spread (variance) of I.Q. scores. Each one happens to have taken the same I.Q. test that has a mean of 100 and standard deviation of 15. The computation

of an appropriate test statistic turns out:
 $(29)(116) / 225 = 14.9511$

10. What is the critical value from the table that tells you whether to reject at the .05 level of significance?
11. Would you reject for the above value, 14.9511?
12. If you had proposed a directional hypothesis, would you have suspected that the I.Q. spread in your class was more or less than normal? (answer more or less)

The problem statement presents an hypothesis-testing situation. The computation is included to (1) make the desired solution unambiguous, (2) provide a cue that should be recognized by the student who has successfully tested this kind of hypothesis during the instructional phase and (3) supply an outcome to use in another question. Question 10 requires the student to retrieve needed values from the appropriate table, question 11 asks for his decision which he obtains by relating the given outcome to the table values and question 12 tests the student's comprehension of the operation of the statistical procedure used to obtain the given outcome.

Similarly, the other questions relate to other types of hypothesis-testing situations.

The test takes about one-half hour to administer. The answers are all either multiple choice or short, constructed responses. No arithmetic is required in the test. The scoring procedures and results will be

described later.

The Pretest

The goals set forth for the course demanded a fairly high level of sophistication in statistical inference from the student. Recognizing the time limitations for giving the course, it was necessary to assume certain entry behaviors. A pretest (Appendix D), constructed by the writer, was designed to measure two important entry behaviors: (1) that the candidate could produce the correct statistical table entry when given necessary and sufficient information for using the table and (2) that he could transform reasonably complex statistical formulas into solvable equations when given the proper values to substitute. In the first case, satisfactory performance implied that the candidate could use either the Student's t table or the Chi-square table for either one-sided or two-sided cases and produce the correct table entry. Further, he was asked about the functional relationship between the significance level, degrees of freedom and table entry. The answers to all questions in this case could be found by looking at the table itself which the candidate was permitted to do. Question 4a of the pretest provides an example where the candidate is expected to retrieve a table entry:

- 4a. What would be the critical t value for rejecting the null hypothesis at the 5 percent level of significance on the basis of the t -test?

Question 7 pertains to the functional relationships found within the table:

7. What would be the effect on the critical t-value if the degrees of freedom are increased?

The effect of one- versus two-sidedness is the basis of question 8:

8. In general, t-values that are significant for a two-sided test (will, need not) be significant for a one-sided test.

Other similar questions provided additional information.

These behaviors were important prerequisites to taking the course because (1) they were not taught by STAT and (2) answer evaluation depended on the table values inserted by the student. The writer plans to incorporate several of the statistical tables in the STAT program as time and space permit.

On the second part of the pretest, the candidate had to express a mathematical formula in terms of the substituted values and then write the expression along a line using no superscripting, subscripting or underscoring. This constraint evolved from the limitations in character set on the teletype keyboard and the prospective subjects had to demonstrate that they could express themselves within that character set. Superscripts are denoted by inserting a double asterisk (**) between the base number and its exponent (e.g. 3^2 becomes 3**2). A slash (/) must replace division lines resulting in the need for more parentheses to group the divisors and

dividends. Radical (square root) signs are likewise restricted to a single character, $\sqrt{\quad}$, to be followed by a number or a parenthesized group.

For the performance measure in the pretest, several questions make reference to the following samples:

i	X _i	Y _i
1	22	21
2	26	9
3	19	13

Referring to the above X and Y columns of scores, the students were to write the equivalent numerical expressions for four formulas--for example, the formula:

$$\sum_{i=1}^3 Y_i^2$$

should cause the student to write:

$$21^{**}2 + 9^{**}2 + 13^{**}2$$

or something equivalent. Other similar exercises are included that differ only in level of difficulty.

Candidates could not miss more than one exercise from each of the two behaviors tested to be eligible as subjects. They had to satisfy at least five of the six exercises (questions 1, 2 and 3) related to formula substitution and at least seven of the eight questions (4 through 9) pertaining to the statistical tables.

One would only expect to find subjects with such a background among those who had taken a college course in

introductory statistics. A later chapter will describe the college students who were designated to be the population for the project. They knew the names of statistical procedures and had a limited concept regarding their use.

Neither the pretest nor the criterion test were designed to be comprehensive performance measures for a course in statistics. Rather, each was based on specific objectives that were to be measured. Certain entry behaviors are necessary if a student is to profit from the STAT course. These have been discussed. One can overcome the lack of these prerequisites to some extent by trial and error methods on the computer. However, for this project, it was desirable to select subjects who already exhibited these behaviors, who would not require extra help to obtain them. The pretest was used to select such subjects.

The criterion test, like the pretest, was based on a limited set of objectives. The test questions reveal many similarities to the activities required in the STAT course. It is not a comprehensive test over any area of statistics but rather a measure of the degree to which the stated objectives for STAT were attained. The author does not anticipate that either of these tests will become an integral part of the STAT course.

Description of the STAT Course

The STAT program might be viewed as a structured laboratory course in which statistical knowledge obtained in some traditional manner is to be applied to realistic problems. Certain statistical and computational aids are made available to the student as he attempts to solve each problem. When he has solved the problem to his own satisfaction, he then answers a series of evaluation questions over his work. After successfully answering the questions, he moves on to the next problem.

The procedural and calculation aids together with the high speed of the computer normally effect a significant saving in time over comparable work on a desk calculator allowing the student to give a greater portion of his time to the more relevant aspects of the problem.

The STAT program was organized around twenty-five statistical problems. The course includes both univariate and bivariate problems which require estimators to be calculated and hypotheses to be tested. Table 1 organizes the problems according to the kind of computational activity and properties of the data. Since two or more kinds of activity may be required in the same problem, certain problems are represented in more than one category. The estimator calculation category appears to lack entries. However, in nearly all of the problems, the calculation of one or more estimators is necessary before

TABLE 1. Organization of the Twenty-five Problems in STAT

Sampling Distribution	Problem Type				
	Hypothesis Testing		Estimator Calculation		Regression
	Univariate	Bivariate Independent	Bivariate Matched	Univariate	Bivariate Independent
Normal	2	3	4	1	0
Boarderline ^a	0	3	2	0	0
Non-normal ^b	0	2	5	0	0
Binomial	0	0	1	0	0

^aSomewhat deviated from true normal.

^bRectangular (uniform) and skewed normal.

Notes:

Each cell contains the number of problems belonging to that category. Certain problems fit more than one category, causing the total number of problems to exceed twenty-five.

one completes the problem. These problems were not included in the estimator calculation category even though the calculation of estimators is involved. The problems that are listed in the estimator calculation category are those for which the calculation of certain estimators is explicitly required and evaluated. Two-thirds of the problems are bivariate hypothesis-testing problems. Only one problem of the twenty-five has a binomial sampling distribution.

Each problem culminates with an evaluation section in which one or more questions are posed to the student. The question might require a numerical answer such as a correlation coefficient or ask the student to respond with a "YES" or "NO" decision based on the outcome of his work. The evaluation sections average nearly two questions for each problem-type. The maximum number of questions associated with any given problem is four. Appropriate feedback messages follow the answering of each question and all questions must be answered correctly in order to proceed in the problem sequence unless the student leaves the problem unsolved via the "STUCK" option.

After the STAT program had been developed to the point that it contained all that was planned for the project, it required just over sixteen thousand storage locations in the Q-32 computer. In addition to the program and the tests, guidebooks were prepared which

described the operation of the STAT program to those who were new to the system. Two versions of the guidebook were prepared, one for the student (Appendix A), and one for the instructor (Appendix E). The student's guidebook contained instructions for loading and operating the STAT program, a description of the available options, and instructions for communicating with the program. This guidebook also listed those sections in the text that related to the course.

The instructor's guidebook (Appendix E) contained additional information for altering the instructional program, for using STAT as a research tool and for interpreting the records which were being kept on the students.

An abbreviated form of the description of the options (Appendix B) was prepared separately as a convenience to the student for easy reference while working with the program. The numbers associated with each option provided a means of identifying to the computer what option the student desired to execute on any given occasion. These options included such things as "1. SAMPLE DATA" (for causing a randomly generated set of numbers to be printed on the typewriter), "21. MEAN" (for calculating the mean of the numbers that had been printed), replying with the result, "55. PRODUCT MOMENT CORRELATION COEFFICIENT" (for computing and printing the correlation coefficient of the two samples), "8. STUDENT RESPONSE AND SUMMARY" (for instructing the computer to

evaluate one's answers for the problem, etc.). Numbers, from 1 through 66, were assigned to various options in logical clusters of ten. Estimates were numbered in the twenties, non-parametric procedures in the forties, etc. All option numbers greater than twenty refer to statistical "LIBRARY" procedures which, when activated by typing the appropriate number, do the necessary calculation on the given data and reply with the result. The LIBRARY group of options is available to the student only when the executive portion of STAT activates it. All other options (twenty and below) are either always available or subject to other special controls. For example, Option 1 (SAMPLE) can only be exercised once for each new or repeated problem; Option 9 displays the answer for the most recently missed question provided that the student has attempted to answer first. Option 12 gives the student access to calculation assistance. It evaluates arithmetic statements, one line at a time. Certain mnemonics can also be used by the student to simplify his task (e.g. S1 in an expression will have the sum of the first data column substituted for it during execution. Similarly, the symbols, S2, SS1, SS2, SCP, SMD, and SSD will be replaced by the sum of column two, the sum of squares of column one, the sum of squares of column two, the sum of cross (pair) products, the sum of pair differences and the sum of squared pair differences, respectively). Factorial, combinatorial and square root

operations can similarly be referenced using mnemonics. Intermediate results (i.e. results of line evaluations) are assigned the names, L1, L2, ... Ln, by the program and can be referenced in succeeding expressions by name. Any of the above mnemonics can be a part of a longer expression. For example, suppose one wanted to calculate the biased variance of the first column of some data that contained ten numbers. The computer indicates that it is ready to accept his expression by typing:

ST = ?

Following the question mark (?), one might type the following statement for the desired variance (note the use of the double asterisk for exponentiation):

$(SS1 / 10) - (S1 / 10)**2$

The computer then evaluates the expression and the reply would resemble the following:

L14 = 216.5236

The symbol, L14, indicates the fourteenth line of computation for this problem. L14 could now be used as a legal constant name in subsequent expressions. The assigned values are maintained until the current problem has been completed and the next problem in the sequence has been presented.

Illustrative Problem

In order to clarify the operation of the program for the student, an illustrative problem was included in the problem sequence of STAT. This problem appears first and

is the only problem in the sequence for which the sample values are fixed so that they are replicated for anyone who begins the course. In addition to the illustrative problem in STAT, a copy of that problem as it would appear on the typewriter was prepared and made available to the subjects (Appendix F).

Normally, the STAT user is made aware that an input is expected of him because a question is typed and terminated by the symbols, = ? (which becomes a familiar cue to the student). All activity then ceases until the student types his reply and enters it into the computer. (The computer accepts the reply only after the student strikes the carriage return key). Since the reader is denied this running discourse which the student experiences, parts of the illustrative problem materials were underscored to identify that which the student typed as opposed to that which the computer typed. Similar underscoring techniques were also used in other appendices. Explanatory comments were written into the right margin to call attention to the various steps for the new student. As one observes from the illustrative problem, most of the communication between the student and the computer takes place in one of the following ways:

(1) the computer will type "ENTER OPTION NUMBER" and "OPTION = ?" to which the student responds with a number selected from the student's reference sheet, (2) the computer will indicate a state of readiness in "CALCULA-

TION ASSISTANCE" (Option 12) by typing "ST = ?" to which the student responds with an arithmetic expression that he wants evaluated, or (3) the computer will request a particular answer in the evaluation section by typing, "MEAN = ?," "VARIANCE = ?," etc., to which the student is expected to respond appropriately. This kind of dialogue continues throughout the twenty-five problems of the course. The student is free to consult his guidebook, reference sheet, illustrative problem, or the textbook at any time throughout the course.

Program Control Algorithm

The student exercises a great deal of control within each of the twenty-five problems. He can request options of his choosing. Most problems allow him to specify the size of the sample on which he will both work and be tested. Certain other features are not under his control. The problem sequence is controlled by the executive program as are the "LIBRARY" and evaluation sections. In general, the computer types the text of the next appropriate problem and then turns control over to the student by typing, "ENTER OPTION NUMBER. OPTION = ?." The student maintains control, typing option numbers at will.

Certain of the options are further subject to the control of the executive program, however, and in those cases the student receives an unexpected reply to his option request. Some of these cases follow: First, if a student types the number of an option which requires

certain characteristics that are not found in the data (i.e. two samples, equal sample sizes, etc.) a message will be typed back in place of the result. The message might be, "OPTION ILLEGALLY USED," or, "THIS OPTION REQUIRES TWO GROUPS." Second, if the student is repeating a given problem (with new data) and attempts to use a LIBRARY option (21-66), the computer replies, "THE LIBRARY IS NOT AVAILABLE FOR THIS PROBLEM." (This control can conveniently be changed by one who has the instructor's guidebook.) Third, if he enters a 9 after OPTION = ? without first having attempted to answer an evaluative question on a given problem, the computer will respond, "YOU ATTEMPT TO ANSWER FIRST." Fourth, if he enters any option number which computes its result as a function of the data and if he has not yet requested a sample for that problem, the computer will respond with, DRAW SAMPLE FIRST.

Many other messages are also possible within the student-controlled area, especially in "CALCULATION ASSISTANCE" (e.g. "PARENTHESES DO NOT MATCH," "ILLEGAL DIVISION BY ZERO," etc.).

Evaluation of Student Response

When the student decides he is ready to be evaluated, he so informs the computer by entering the number 8. The computer responds with, "ENTER THE FOLLOWING VALUES FROM YOUR RESULTS." It then proceeds to ask for various values that should have been calculated by the

student prior to evaluation. To check the student, the computer does all of the required calculations from the data (since the answers are not known prior to that time). The student is not required to follow the same procedures as the computer but his answers must be consistent with, or within the specified tolerances of, those derived by the computer. The program allows the student as much freedom as possible in choosing his methods for solving the problem as long as the resulting decisions are acceptable.

When the student enters "8" after "OPTION = ?," he is essentially turning control back to the executive program. The program requests the student's answers, one-by-one, and judges them against its own computed results. The judging of a given answer occurs immediately after that answer is typed in. If the answer is judged to be incorrect, the student is so informed and given a textbook reference related to the specific statistic on which the answer was to be based. The executive program then turns control back to the student with the messages, "ENTER OPTION NUMBER" and "OPTION = ?." The student resumes his work and, when ready, again types an "8" after "OPTION = ?." Now, the executive program finds the question which caused evaluation to terminate before and resumes at that point in the evaluation sequence.

There are two other ways, besides inserting an incorrect response, that a student can usurp control from the executive program while in the evaluation sequence. These are by (1) typing, "BACK" or (2) typing, "STUCK," as one's response to any test question. After typing "BACK" in place of the requested answer, the executive program records the use of that provision and returns control to the student by typing, "ENTER OPTION NUMBER" and "OPTION = ?." This provision was included in the program for those who prematurely request evaluation and afterward realize that more work must be done. The student could accomplish the same thing by responding incorrectly. However, in the early trials it was discovered that a poor psychological effect resulted from having to intentionally put in wrong answers. Though there was no penalty incurred by making errors, the students didn't like to make them. The "BACK" condition provided an acceptable solution.

The "STUCK" condition, also allowed in place of the answer to any test question, was provided as a convenience to the student who was unable to successfully complete the test questions. By typing "STUCK," the executive program would simply terminate the problem, store appropriate records and present the next problem in the sequence. Alternatively, the student could use Option 9 to obtain the correct answers. However, for those who had given up on the problem, that method resulted only in

a meaningless exercise. The "STUCK" provision was also useful as a recovery procedure on occasions when a program malfunction caused the computer to reject all answers for a given question. This occurred occasionally, especially during the checkout stages of the programming.

Often the executive program computes several answers to a given question. Sometimes certain wrong answers are discovered by first finding a mismatch between the answer and the correct result, then finding a match with an anticipated wrong answer that was computed. Two examples where this was done are: (1) when the unbiased variance is requested, the biased variance is also computed, and (2) the Pearson Product Moment Correlation Coefficient is also computed when the Spearman Rank correlation coefficient is requested. By computing alternative wrong answers, the computer can be more helpful in diagnosing common errors.

Occasionally, the executive program will compute alternate answers either one of which, if matched, will be accepted as correct. Instances of this occurred in problems where results from both t-tests and suitable non-parametric tests were acceptable answers.

When the requirements of the evaluation section have been fully satisfied, the computer types an appropriate stored comment such as "GOOD WORK. NO ERRORS." Following this, the parameters of the simulated population distribution are printed to provide information about the

inferences that were made. Having automatically recorded selected performance characteristics, the lesson moves on to the next problem in the sequence.

Help Options

Three options are designated "HELP" options. They are options 9, 10 and 11. You will recall that the use of Option 9 causes the computer to print the answer that the student most recently missed for any given problem. That answer might be a value, a "YES" or a "NO" giving him the necessary help when he is incapable of deriving it. Once he has the correct answer to insert, he can continue along the normal sequence. If he instead chooses to type "STUCK," he would miss the remaining evaluation for that problem.

Option 10 provides a limited reference to the textbook. After the student enters a "10" following "OPTION = ?," the computer will respond with "REQUEST HELP BY TYPING OPTION NUMBER" and "REQUEST = ?." The entire repertory of this option consists of page, section and formula references that are printed on the teletype corresponding to the option number that was put in. These references identify the formula on which the computer bases its computation for any given option number. The example in Figure 1 (also included in Appendix F) shows one use of Option 10. Any "LIBRARY" option number can be entered after "REQUEST = ?."

ENTER OPTION NUMBER

OPTION = ? 10

REQUEST HELP BY TYPING OPTION NUMBER.

REQUEST = ? 31

MEAN LIMITS: P311, S10.10, F10.10.1

Fig. 1. Use of Option 10. Help is being requested for the MEAN (Option 31). The reply consists of page, section, and formula reference numbers in the textbook.

The third help option, Option 11, contains the "STEPS TO THE SOLUTION." For each problem, a series of option numbers have been stored which correspond to the procedures that will be used by the computer for evaluating the student. When the student enters a new problem in the sequence, the executive program transfers that list of option numbers into the "STEPS" option so they may be called out by the student. Each time the student wants to query the computer to find out the next step to use, he makes use of Option 11. When the list has been exhausted, subsequent use of Option 11 will produce the message, "ALL STEPS COMPLETED," until the program has been advanced to another problem. Essentially, Option 11 (STEPS) produces Option 10 (HELP) information without having to specify the number of the option for which help is needed. This feature provides limited guidance to the student who cannot understand what is expected by the text of the problem alone. If the student executes each of the options indicated by STEPS, he should have enough information to satisfy all of the evaluation questions.

The use of STEPS (Option 11) is demonstrated in the illustrative problem (Appendix F).

It becomes evident, even as was mentioned in Chapter II, that the program relies heavily on the textbook for large quantities of textual material. When a student makes an error or requests help, he is directed into the textbook. Appropriate page references are also listed after each "LIBRARY" option in the student's reference sheet. The economy of this method is unquestionable. However, it does demand a well-written, self-sufficient textbook.

Some students adopted strategies for working through the problems which often made the available help more effective. This will be discussed in detail later.

Populations and Sample Data

For each problem, simulated populations of subjects have been programmed into the computer. The specified population distribution parameters make the randomly sampled data appear reasonable and appropriate for the problem.

Parameters that are prestored include means, standard deviations and correlation coefficients. Some parameters can be supplied randomly by the computer to give additional variety to the problem. An example of this would be fixing all parameters except a mean or the correlation coefficient which would be randomly supplied when the problem was presented.

The shape of the population distribution can also be specified. The distribution might be normal, uniform, binomial, or skewed in either direction.

The data that are printed for the student originate from a uniform random number generator, are transformed by the appropriate equation (depending on which distribution has been specified), and then are scaled according to the mean and standard deviation parameters. To generate two samples with a specified correlation, the first sample is generated as usual, then each number of the first sample is randomly altered to form its pair in the second sample. Samples with any specified correlation coefficient can be generated in this way.

Randomization is accomplished in such a way that, except for the first problem in the course which intentionally produces identical sets of data, the probability that any two sets of data will come out the same is very small.

Probability Distribution Options

Options 13, 14, 15, and 16 calculate probabilities for the binomial, hypergeometric, Mann-Whitney, and Wilcoxon paired observation probability distributions, respectively. Though the results are obtained by computation, identical results could be obtained from appropriate tables. Computation of exact probabilities for the Mann-Whitney and Wilcoxon distributions is based on a recursive formula which is a function of the sample

size and the shift parameter. Occasionally, extremely large numbers of calculations could be required for some samples. In cases where calculation time would be excessive, the normal approximation will be printed instead. Checks are included wherever necessary in the STAT program to protect the student against accidentally forcing the computer into an excessively long or endless sequence of operations which would adversely affect the response time.

Starting the STAT Course

After turning on the teletype, one begins by identifying himself to the computer by typing "LOGIN 1234 98625." The number 1234 represents the person's assigned man number and 98625 represents the account number under which he is working. (The student will have numbers assigned to use in place of these.) The computer replies, "\$OK LOG ON 24." The number 24 is the channel number of the teletype for that session. The next instruction to be typed is, "LOAD STAT." In response, the computer makes a copy of the STAT program that resides on magnetic disc, puts it in the proper working area of the computer for the given channel and then replies, "\$LOAD 24." If the program is not there, the tape reel number must be given to the computer operator and he places a copy on disc. In either event, after receiving the message, "\$LOAD OK," typing a "GO" command will put the STAT program into operation. After typing "GO" the computer re-

plies, "MSG IN" to acknowledge the input, then begins executing the STAT program. The next few messages are self-explanatory. Figure 2 illustrates the first few lines. Underlining has been added to identify that which the student has typed. Absence of underlining designates the computer's reply.

Notice in Figure 2 that "NO" has been answered to the question, "ARE YOU RESUMING A PREVIOUS LESSON:." If the student had wanted to pick up where he had previously left off, he would have answered YES and the computer would then request more information. That alternative will be explained later.

Termination Prior to Completion

Since several hours are required to complete the course of instruction, it was necessary to provide the means for a student to terminate and then, on the next session, to resume where he had left off. The computer time-sharing system does have the capability of saving one's program onto magnetic tape and reloading it when the program is to be resumed. However, doing this obviously requires the computer operator to maintain a separate tape reel for each person who uses STAT. Another method for interrupting the program was devised which fully satisfied this need without requiring computer storage facilities to maintain one's place. This was done by programming the computer to calculate and print a five-digit "continuation code" number for each

student who used Option 20, the termination option. Figure 3 illustrates the termination of the lesson.

The continuation code number contains (1) the re-entry point for resuming the lesson, (2) the repetition pattern which the executive program uses as an algorithm for advancing the student through the problem sequence, and (3) a check used to verify the other digits of the number. The first digit of the code number contains the maximum number of times that the student should have to repeat a given problem. The second digit contains the number of repetitions of a given problem for which the LIBRARY options should be made available. The last two digits indicate the position number in the problem sequence where the student will resume. The middle digit is assigned a value that will make it and the other digits sum to a multiple of ten. For example, the continuation code number 21908 would cause the program to resume on the eighth problem in the sequence, repeating each problem a maximum of two times with the LIBRARY options available only on the first presentation of each problem. The number is valid since twenty (i.e. $2 + 1 + 9 + 08$) is a multiple of ten. Constructing the number in this manner accomplishes three things; (1) it allows the student to resume where he left off without requiring separate computer storage for his lesson, (2) it retains the algorithm needed by the executive program from one session to the next, and (3) it provides a check whereby

LOGIN 1234 98625

\$OK LOG ON 24

LOAD STAT

\$LOAD 24

GO

\$MSG IN.

YOU ARE BEGINNING A COMPUTERIZED TRAINING
COURSE IN STATISTICAL APPLICATIONS.

TO BEGIN, TYPE IN YOUR STUDENT NUMBER.

NUM = ?

ARE YOU RESUMING A PREVIOUS LESSON: YES/NO

ANS = ? NO

PROBLEM 1.00

REQUEST A SAMPLE AND. . .

Fig. 2. Computer-student interaction at the
start of the STAT course. Underlining has been
added to identify that which the student has
typed.

OPTION = ? 20

YOUR CONTINUATION CODE IS:

21908

STUDENT NUMBER 1234

DATE 6 5 65, BEGIN 9 24, END 11 42

PROBLEM-TYPE-ERRORS-UNSUCCESSFUL-STEPS-TIME-BACK

1	6	0	0	3	41	0
2	8	0	0	2	20	0
3	11	0	0	2	10	0
4	12	1	1	2	65	2
5	12	0	0	0	12	1000

\$PGM CONCLUDED

Fig. 3. Sample of lesson termination and
student records. Underlining indicates that
which is typed by the student.

the computer can identify invalid numbers that result from typing errors and faulty memory. Originally, the number only contained a maximum of four digits; the fifth (check) digit was added to avoid the confusion which resulted when students put in the wrong number. If the submitted number was invalid, another number was requested. This technique has worked quite successfully thus far.

Students always re-enter the program beginning with the problem they last saw prior to termination. Because the data associated with the problem will not be the same as before, it is recommended that Option 20 should be used immediately after advancing to a new problem so that no effort will be wasted. In cases where Option 20 is used after work has been done on the problem, the computer asks the student to verify with a "YES" or "NO" whether he wishes to terminate anyway at the cost of losing his work on that problem.

Re-entering the Program

Recall that when the student enters the program, the computer asks:

ARE YOU RESUMING A PREVIOUS LESSON (YES/NO).

ANS = ?

In the illustrative problem, the student typed, "NO," and thus obtained the first problem in the sequence. If the student had already progressed part of the way through the course sequence, he would type, "YES," to

which the computer would respond:

```
ENTER YOUR CONTINUATION CODE.  
NUM = ?
```

After "NUM = ?," the student would type in the code number that was furnished by the computer at the termination of the previous session. The computer would either respond with, "ILLEGAL CODE NUMBER" and request another number or accept the number and resume at the specified position in the problem sequence. He then continues working the problems in the usual manner.

Record of Student Performance

As the student works through each problem, certain characteristics of his performance are automatically recorded. These records are made from internal states of the computer so that no indication of the action appears before the student. All of the records are printed out in a summary list at the end of every session. When the student requests termination (by using Option 20) or when he completes the last problem in the sequence, his performance records for that session are automatically printed out.

The records include: (1) identification of the type of problem, (2) number of errors made during evaluation, (3) number of times the answer give-away (Option 9) was used, (4) what helps were viewed from Option 11, (5) the time required to complete the problem, and (6) the number of times the student asked to leave the evaluation sequence

by typing "BACK" or "STUCK" in response to a test question. The summary also prints the student number, date, beginning time and ending time. Figure 3 shows a sample of the records which are printed when the session is terminated. The meaning of the continuation code has already been explained.

The records in Figure 3 cover a two-hour session on June 6, 1965 beginning at 9:24 a.m. until 11:42 a.m. for student number 1234. This student completed five problems during that session. The problems are identified under the TYPE column. The ERRORS column lists the number of questions that were answered incorrectly at first but later corrected. The UNSUCCESSFUL column shows how many times he was given the correct answer by using Option 9. STEPS indicates the number of times Option 11 was used in that problem. The numbers under TIME indicate minutes spent working on the problem. The BACK column tells how many times the student used the BACK option while being evaluated. If he used the STUCK option to leave a given problem, the number, 1000, appears in the BACK column for that option.

The records in Figure 3 indicate that the student had error-free performance in problem-types 6, 8, and 11 having spent more time on 6 than on 8 or 11. He had trouble on problem 12; the ERRORS column indicates that he was able to correct only one of his two mistakes. The UNSUCCESSFUL column lists the one he couldn't correct.

He used sixty-five minutes on that problem and used "BACK" twice to do more work. Because of his performance, he repeated problem 12 but after twelve minutes, he typed "STUCK" and then terminated when the next problem was presented.

These records present a concise summary of a student's work during the session. If more information is desired by the instructor, he can save the entire teletype tape which contains a permanent record of every detail of the student's work. Carbon copy rolls of teletype paper were used for this project so that the student could also retain a complete record of his work.

STAT Program Variations

The STAT program is designed to function in the described manner exactly as it is originally loaded into the computer. No instructor intervention is required. There are, however, several special options in the instructor's guidebook (Appendix E) which can be used to substantially alter the operation of the program. These options do not appear in the student's guidebook.

Modifications of problems

By using Option 555, one can change any of the following characteristics of a problem: the population parameters, the required significance level, sample size maximums, correlation coefficients, or its position in the lesson sequence.

The pattern of repetition may be changed by using Option 888. Instead of requiring the student to repeat the problem once without the use of the LIBRARY, the problems may each be repeated several times or not at all. The use of LIBRARY may be permitted or inhibited for as many of the repetitions as desired. The pattern of repetition can be made different for each problem.

Options 666 and 777 control the automatic sequencing from one problem to the next throughout the STAT program. Option 777 unlocks the sequence; Option 666 reinstates it. These options are especially useful for two purposes. First, an instructor may want to move the student to a new place in the sequence and have him continue the course from there. To do this, he waits until the student reaches a place in the program where the computer has printed, "OPTION = ?." The instructor then types in option number 777. The computer replies, "SELECT NEXT PROBLEM (0-24)" and "SELECT = ?." He enters the problem number and the computer immediately types out the text for the selected problem and again types, "OPTION = ?." He now uses Option 666 which locks the student in the sequence at the new position. The student continues working on the new problem in the usual manner.

Secondly, the instructor may wish to permit the student to select subsequent problems. In this case, the instructor uses Option 777 as in the first case but does not use Option 666. The student using a numbered list of

the problems, enters one of these numbers after "SELECT = ?" and begins working on that problem. When the student completes the problem, the computer will again request the next problem number that the student wishes to work, and the pattern continues until the student terminates.

Experimental use

Another special option, 999, allows the STAT program to be used for evaluation of real data rather than for instruction. After inserting option number 999, the computer requests the data and other pertinent information. Having input the data, the experimenter can then use any of the options to investigate properties of the data in the same way as if the data had been generated by STAT. The only difference in operation is found in some options which are no longer appropriate and are either deactivated or given other meanings. Option 1, for example, which usually produces a sample, now allows the experimenter to change significance level at any time. Option 8, evaluation, is made inactive. On the other hand, Calculation Assistance (Option 12) operates as usual. So do all of the options in the LIBRARY. Unlimited use of the LIBRARY is automatically allowed.

Technical Information

The JOVIAL version of the STAT instructional program requires approximately 16,000 words of computer memory per student. The Q-32 computer system makes it

convenient for each student to operate a separate copy of the program which he loads from disc storage. For smaller computer systems which are not time-shared, it would be possible to have each student typewriter controlled by the same program in multiprocessing fashion. In this way, a single programmed version of STAT could be modified to accommodate several students.

The entire program represents from ten to twenty hours of instruction, depending on the individual. Though the STAT program is written in JOVIAL, it requires only those programming features which are common in most compiler languages. Translation of the program into FORTRAN or some similar language should be quite routine and direct.

The actual programming task presented a number of challenging problems. Statistical procedures had to be general enough to accept any data that were generated and yet be sufficiently protected so that a student could not halt the program through improper actions. (Undefined operations, such as a division by zero, normally halt a computer program.) Since such interruptions would be very undesirable, numerous checks were included to avoid them, printing an error message to the student, instead. Evaluation had to take into account such things as rounding error discrepancies between the evaluated results and the student's answer.

The basic structure of the program is presented in flow chart form in Figure 4. The student enters the program at "START." The program moves through boxes one through four where the problem is structured, to Box 8 where it is presented, and then to Box 9 and waits for a reply. Typing a "1" moves him through Box 18 where his data is printed and then back to Box 9 again. Similarly, boxes 19, 21, 23 and 25 through 27 represent various option requests that take the student back to Box 9. Box 22 corresponds to Option 12, calculation assistance, where the student remains until he types, "END" and then returns to Box 9 again. Option 8 takes the student to Box 20 where he either advances to the next problem by route of boxes 10 through 17 or makes an error and goes through Box 13 to Box 9 again. If he gets to Box 17, he either moves on to Box 2 or repeats the problem at Box 8, depending on his performance. He terminates either by choosing Option 20 (Box 24) or by running out of problems (Box 4).

The Computer Laboratory

Though the STAT program could be implemented on one of several computers, the computer laboratory at SDC offered a number of important advantages.

The Q-32 computer rates among the most powerful of modern computers in terms of size, speed and reperatory of commands. It contains four 16,000 word core banks of which three, or approximately 47,000 words can be shared

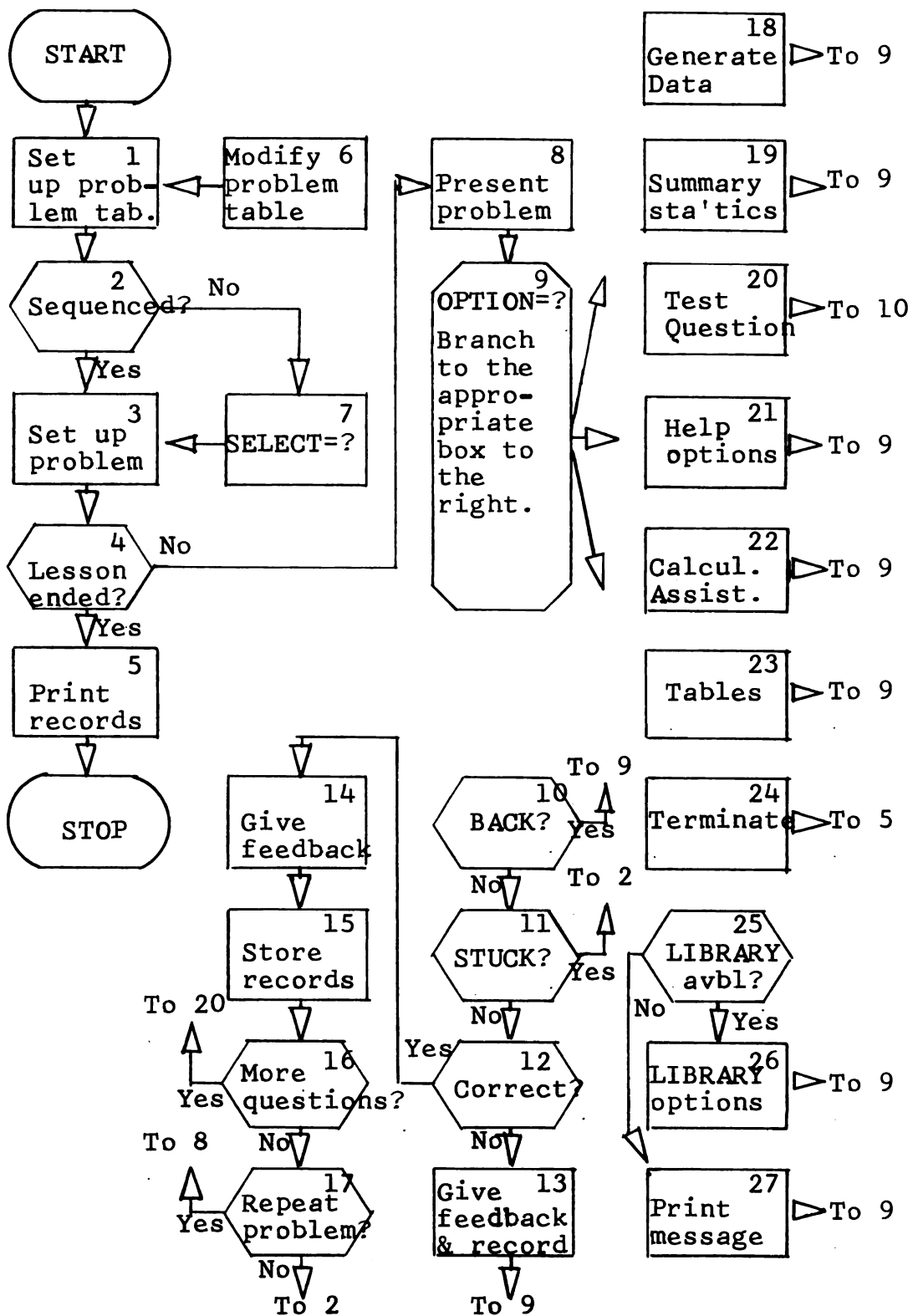


Fig. 4. Flow chart of the STAT program.

by all the users on the system. The remainder of the 64,000 words is occupied by the Time-Sharing System (TSS) executive program. The TSS program continuously monitors each of the fifty-two possible user channels, allocates drum and disc storage space, controls the input and output of all devices and cycles each user through the core memory of the computer where processing takes place. While all of this goes on, an individual user is unaffected by other users on the system. He works as if he has the computer to himself. Actually, he has the computer to himself for periods of six hundred milliseconds each time his turn in the cycle comes up. To accomplish this, the TSS executive program assigns space to each user on one of five high-speed drums and then "swaps" the programs, in turn, into and out of the four core banks of the computer. Inactive, but frequently used, programs reside on magnetic disc. The disc is also controlled by the TSS executive program so that programs may be moved from disc to drum to core and back again automatically as they are needed. The disc has a capacity of approximately 4.5 million words. In addition, a magnetic tape drive, coupled with the disc, is also controlled by the TSS executive program. The TSS program places the least active programs on tape when the disc becomes overloaded. All of these features are handled automatically by the TSS executive program. Yet, with all of these activities occurring, an individual user loses only about two seconds

in response time due to sharing the computer with several others. To accomplish this, the TSS executive program automatically senses those programs which require lengthy computation times with little or no human interaction and places these programs in a lower priority "production stack" to be executed during the many intervals when the computer would otherwise be idle.

In contrast to multiprocessing, where several users are serviced by one program, the SDC Time-Sharing System allows several users to operate independent programs simultaneously.

To use the STAT program, the student simply types, "LOAD STAT." The TSS responds to that request by allocating drum space to that student's channel and placing a copy of the STAT program from disc onto the allocated drum. The message, "\$LOAD 24" is typed on the student's teletype to signal the completion of the load. The loading process requires five to ten seconds. Then the student types, "GO." This is the signal to the TSS program to begin "swapping" that student's copy of the STAT program into and out of core along with the other programs that are already in the cycle. Unless an error occurs to stop the program, it continues to cycle through the core memory until the student has terminated and the channel is released by typing, "QUIT."

A wide range of input and output devices are available to the programmer including teletype, cathode ray

tube, light pen, Rand tablet, card and tape reading and punching equipment, random access disc and magnetic tape. The STAT program presently only uses the teletype for interaction with the student. A cathode ray tube could obviously be used to good advantage for displaying curves and graphs. However, the STAT program was designed to operate on a remote channel over telephone lines and the cathode ray tube display device cannot be used remotely. Therefore, its use was not included in STAT.

Several of the teletypes are located in a laboratory area on the floor above the computer. The teletypes have work tables conveniently situated next to them. It was there that the subjects in the project took the STAT course.

Six of the teletype channels are connected to TWX lines and four are on data phone lines. The writer did a large part of the programming over one of the TWX channels from East Lansing, Michigan to the computer in Santa Monica. There have been frequent users from the East Coast. It has been used from as far away as Denmark by Trans-Atlantic cable.

The computer complex and time-sharing system that SDC has developed is one of the very few such systems in existence. However, computer manufacturers are now building systems such as these. Within two years, the capabilities that have been described here will be widely available.

Summary

The appendices contain the materials which were developed for the project. These include a pretest and criterion test, two questionnaires, and printed instructions for operating the STAT computer program. Essential to operating the program is the student's reference sheet that contains a list of the options which are available. Since this list contains little explanation, the student's guidebook is also necessary for a more detailed explanation of the various options. The STAT program makes frequent references to the textbook written by Hays (Statistics For Psychologists, 1963). The illustrative problem is useful for orienting a novice to the computer program. It provides a guide for the student to follow as he works through the first exercise in the course.

V. Procedure For Field Implementation

The psychological instruments developed for evaluation purposes included two questionnaires and two performance tests. A survey-type questionnaire (Appendix G), was administered when the subjects began the STAT course and an attitude questionnaire (Appendix H) was given (together with the criterion test) when they finished.

The Survey Questionnaire

The purpose of the survey questionnaire was to find out the extent to which each of the computer subjects had been exposed to the statistical topics included in STAT. It posed the same three questions about each of thirty common statistical procedures. The questions were:

- Question 1. Have you become acquainted with this statistical procedure and its application (either in class or through independent study)?
- Question 2. Have you ever carried out the calculations for a statistical problem using this procedure?
- Question 3. Would you have answered either of the first two questions differently before you were contacted for this experiment?

The subject answered the questions for each procedure by circling a "Y" or "N" (Yes or No) corresponding to each question adjacent to the name of the procedure. For example, in Figure 5, the subject has indicated that he (1) is familiar with the arithmetic mean, (2) has computed

it, and (3) has not changed that status since he was contacted for the project.

Q.1	Q.2	Q.3	
<input checked="" type="radio"/> Y <input type="radio"/> N	<input checked="" type="radio"/> Y <input type="radio"/> N	<input type="radio"/> Y <input checked="" type="radio"/> N	1. Arithmetic Mean

Fig. 5. Sample item from the survey questionnaire (Appendix G).

The answers to this questionnaire revealed which procedures the STAT course would introduce to the student for the first time and also for which ones the subject had never attempted to perform the necessary calculations. This information provided a basis for part of the analysis of the data obtained from the experiment.

The Attitude Questionnaire

The purpose of the attitude questionnaire (Appendix H) was to obtain the following data: (1) an over all merit rating comparison between the computerized STAT course and the classroom presentation at the university, (2) attitude information about specific features of the STAT program and the way it operates, and (3) indication of attitude change, if any, regarding the use of a digital computer.

One item (Figure 6) was used to obtain an indication of the subject's appraisal of the STAT course relative to traditional classwork. In this item, the subject was instructed to divide one hundred points into two parts, thus weighting each method according to its over all merit.

- I. In your statistics class you probably had a lecture, text, and exercises to do. I want you to compare that method of learning how to use statistics with what you have been doing on the computer. Divide 100 points between the computer method and your former method according to which you think has the most over-all merit (50-50 division would mean undecided).

computer_____ former_____

Fig. 6. Question I from the Attitude Questionnaire. (See Appendix H.)

Attitudes about the program operation were sampled by having the subjects respond to each of twenty-eight statements on a four-point rating scale. The four categories were: (1) very true, (2) essentially true, (3) slightly true, and (4) not true. The subjects checked the category that best expressed their attitude about the truth of the statement. Fourteen statements reflected possible advantages of the computerized instruction and fourteen reflected possible weaknesses. The twenty-eight statements were mixed randomly to avoid response set. Statement 17, "I liked the idea of operating a computer" and statement 18, "I liked working on something new" were included to give an indication of the role played by the novelty of the situation.

The third part of the attitude questionnaire, the general reactions to the use of the computer, contained six questions. Had they used a computer prior to this experience? Did they come with apprehensions about using one? Had their attitudes about a computer changed during

the sessions? To what extent would they expect to use a computer in the future?

The subjects were encouraged to write any additional comments at the end of the questionnaire.

The questionnaire is three pages long but required only a few minutes to complete.

Sampling Population

Since this was a developmental project, rigorous experimental control of the sampling population was not a major concern. However, it was desirable to identify two comparable groups of students, one of which would take the STAT course and the other would not. With this arrangement, criterion test results of the two groups could be compared to provide some indication of achievement gain in the computer group.

Thus, students in a psychological statistics section at the University of California at Los Angeles (Psych 203B) and an equivalent section at Michigan State University (ED 969B) served as the sampling population. As a convenience logistically, all of the subjects who were to be given the computer course came from UCLA, while the ones who were to be used for comparison came from MSU.

The pretest was administered to eleven students at UCLA and sixty students at MSU without explanation. Then the UCLA students were told that volunteers would be paid to participate in a project at SDC if they satisfied the requirements on the pretest. Ten students at UCLA passed

the pretest and nine of them volunteered to participate at the rate of two and one-half dollars per hour of work at SDC. (No added travel allowances were given) These volunteers took the course in five three-hour Saturday sessions, an average of fifteen hours per student. They started the first session by completing the survey questionnaire (Appendix G) and then began working on the computer.

At the end of the fifth session, each subject completed the criterion test and attitude questionnaire. Only one subject completed the entire course. Two other subjects missed one session and therefore took the criterion test after four sessions.

The sixty MSU students in ED 969B also took the criterion test. It was administered to that entire section, again with no explanation, during the same week in which it was given to the nine subjects at SDC. However, only twelve of the test results were kept for the project, those of the twelve students who had passed the pretest. Neither questionnaire was given to the MSU students since they each pertained only to the computer program.

If the UCLA class had had a larger enrollment, it would have been preferable to have selected all of the subjects from that class. However, several similarities among the two classes suggested that meaningful information could be obtained by comparing selected students

from each. First, both the UCLA and the MSU sections were in the second quarter of two-quarter sequences. The course outline for both were nearly identical; both were working in analysis of variance techniques throughout the five-week tryout period, using the same textbook (Hays, 1963).

Secondly, both sequences were designed for preparing doctoral students to do their dissertation projects, probably accounting for the similarity that existed between the classes.

Thirdly, the same pretest was used for screening students in both classes. Thus, the two groups were similar enough for the comparison that was planned. However, the author does not wish to convey the idea that two treatments were being experimentally compared. Rather, the two groups were identified only to help make the criterion test results more meaningful.

Schedule of the Experimental Sessions

Both classes selected for the project were given the pretest during the last week of April, 1965. The computer group, made up of volunteers from those UCLA students who passed the pretest, began their first session on May 8, 1965 by filling out the survey questionnaire and then proceeded into the STAT course. All students continued their regular classwork at school.

Both classes had covered the material that was relevant to the STAT course during the previous quarter

and had moved on to analysis of variance (not included in STAT). Neither class discussed topics covered by STAT during the period of the computerized sessions.

The criterion test was administered during the first week of June, 1965. All of the students in the MSU class took it, but only the scores of those who had passed the pretest are reported.

From UCLA, only the computer subjects took the criterion test. They also completed the attitude questionnaire (Appendix H). For eight of the nine subjects, both instruments were administered at the end of the last Saturday session on June 5. (One subject had to terminate on the previous Saturday, so he was given both instruments then.)

Telephone Interviews

In addition to the other data that were collected, the author interviewed three subjects from the computer group by telephone at the end of the last session. Rosenbaum identified three subjects from the group in terms of the amount of difficulty they had encountered and how far they had progressed. The three subjects represented the two extremes and the middle. Each interview lasted approximately ten minutes.

The interviews were structured only to the extent that four prepared questions were posed at some time during the conversation to each of the three participating subjects. The four questions were: (1) Did you consider

the computer sessions to be an enjoyable experience?

(2) To what extent was the novelty of the situation

important? (3) Would you consider a laboratory such as

this to have sufficient merit to justify weekly assign-

ments on the computer in conjunction with classwork?

(4) What observations do you have in comparing the

merits of this method of learning statistics to that of

normal classroom procedures?

The subjects were encouraged to explain their attitudes. The conversations were not limited to the topics indicated in the above four questions, but other areas were probed as the subject brought them up. The information that was obtained during these interviews is discussed in Chapter VI.

VI. Evaluative Data

The completed project yielded an immense amount of data. There were pretest and criterion test scores for all subjects. In addition, each subject in the computer group responded to two questionnaires and produced several yards of teletype paper per student. In anticipation of these data, certain guidelines were previously established for organizing and tabulating the data.

Guidelines for Data Collection

Plans called for the collection of both objective and subjective data. The subjective data consisted of the response to the attitude questionnaire and the telephone interviews. The questions posed by both techniques have already been discussed. It was hoped that the telephone interviews would at a minimum manifest the biases of the three chosen subjects toward the STAT program and what difficulties they might have encountered in learning new material. If the subject was favorable toward the program, the question of novelty was probed to see if that was considered to be a major influence in the formation of his attitude.

The attitude questionnaire contained questions dealing with general attitudes toward the STAT course in relation to a more usual classroom situation, but it also contained several questions regarding isolated aspects of

the STAT program which were intended to provide diagnostic information which would be valuable for further development of computerized teaching techniques, especially in statistics. The construction of the attitude questionnaire was previously discussed.

Scoring Guidelines

The pretest and criterion test were used particularly for obtaining performance measures on the students. Two scores were taken from the pretest corresponding to the two topics tested. Each part of each question was worth one point, for a total of six and nine points on the respective topics. To be chosen as a subject, the candidates were required to earn at least five of the six possible points on the first topic, and eight of the possible nine on the second.

The criterion test yielded a single score ranging from zero to one hundred twenty-two points. These scores were derived from the seventeen items, most of which were multiple-choice. The items were weighted so that a two-choice item was worth four points, a three-choice item was worth six points, and all other items were each worth eight points. The method suggested by Coombs (1965) was used to score the multiple choice items to obtain an expanded scale for each item. Instead of choosing the correct alternative, the students were instructed to cross out the alternatives they knew to be wrong. For any multiple-choice question that has three

or more alternatives, this method provides more information about the students' achievement than simply choosing the correct alternative.

Constructed response items were worth eight points. Certain items instructed the student to name an appropriate procedure for designated conditions. If the student named a procedure that violated distribution assumptions, but was otherwise correct, he received half credit (four points). The scores obtained from the criterion test were used only for comparing the performance of the computer group with the no-treatment group.

Guidelines for Work Appraisal

The survey questionnaire and teletype sheets together produced the larger amount of objective data. For these data, the guidelines were particularly useful in that they provided a scheme for selecting, organizing, and tabulating pertinent observations. Responses on the survey questionnaire provided a basis for interpreting the information on the teletype sheets. Essentially, the survey questionnaire was designed to reveal the extent of the subject's previous familiarity with the statistical procedures in the STAT course.

The format of the survey questionnaire has already been discussed in Chapter IV. The first two questions were of primary interest. Question three was meant to identify anyone who might "bone up" after he was chosen for the project. It reads:

3. Would you have answered either of the first two questions differently before you were contacted for this experiment?

This third question received a negative answer for every procedure from all subjects. It received no further consideration. Responses to the first two questions placed the subject at one of three levels for each of the thirty statistical procedures named.

Level one: He has both encountered and practiced the statistical procedure.

Level two: He has encountered the statistical procedure but has not practiced it.

Level three: He has neither encountered nor practiced the statistical procedure.

These levels will be referred to in discussing the subject's work on the teletype sheet.

Recall that the four goals of the project (Chapter II) were: (1) to make effective use of information storage and retrieval (problem statement, help, diagnostic information, student records, etc.), (2) to make use of stored procedures (coinciding with Options 1-66), (3) to provide for extensive branching (the ability to use these procedures conveniently), and (4) to insure rapid computation (fast results from use of options with a minimum of time and effort required).

With the above goals in mind, the teletype sheets were expected to yield: (1) an estimate of the value of the information which the computer had stored by noting the occasions in which additional information was

requested from the computer, (2) which stored procedures were used and for what purpose, and (3) the speed and efficiency with which the subject arrived at an acceptable solution for each problem.

The following fourteen guidelines organized into three major groupings were chosen to provide a systematic method for examining the teletype sheets:

Performance Characteristic Guidelines:

1. Use made of the available help (STEPS, Option 11) before attempting to answer evaluation questions.
2. Errors in answering evaluation questions.
3. How much the "BACK" option was used.
4. How much the "STUCK" option was used.
5. Use of the option which provided the correct answer (Option 9) and reasons for its use if obvious.
6. The incidence of guessing.

LIBRARY Use Guidelines:

7. The extent to which the LIBRARY options were used.
8. The level three options that were used when level one or two options might have been used instead.
9. The frequency with which two or more appropriate solutions were applied to a given problem. E.g. using both a t-test and Mann-Whitney test for testing differences between means.
10. The use of LIBRARY options which obviously were used for motives other than answering the questions.
11. LIBRARY procedures correctly computed by the student. How many of these were

computed incorrectly at first and then subsequently corrected?

12. Of the correctly computed LIBRARY options, how many were in the level two or level three category?
13. Unsuccessful attempts to compute LIBRARY options.

Strategy Guidelines:

14. Evidence of problem-solving strategies.

The first five guidelines suggest inferences about the amount of direction the student may require on a given problem. Guideline one applies in cases where the student uses Option 11 to obtain additional cues in the form of suggested statistical procedures and their associated textbook page references that lead to an acceptable solution.

Errors (guideline two) also provide additional cues in that they result in feedback messages which suggest appropriate statistical procedures and provide references. At that point, the student may choose to implement the named procedure by using the appropriate LIBRARY option. This combination is usually adequate to help students through at least the first presentation of any problem. Since one cannot leave a problem until he can answer every question correctly, if he cannot compute the required answers, it may be necessary to obtain the correct answer through Option 9 or to type the word "STUCK" in order to move on to the next problem in the sequence. Evidence of either of these actions provides

reasonably good indication that the directions associated with that problem were insufficient at least for that student, hence they are noted in the third and fourth guidelines.

The significance of the use of both "BACK" and "STUCK" depends upon the conditions in which they were used. Repeated use of the BACK option would probably be due to inadequate cueing in the problem statement. If the student repeatedly requests evaluation (Option 8) before attempting to solve the problem, sees the question, types "BACK" and then proceeds to do that which is necessary to obtain the answer, then he is using that capability simply to obtain added cues.

Use of the "STUCK" option provides a way to circumvent difficulties encountered in solving a problem. "STUCK" should only be used as a last resort evidencing insufficient help provided for the student to adequately solve the problem. Frequent use of the "STUCK" option may indicate a degree of frustration with the course, especially if several attempts have been made.

The statement of the problem should be sufficiently clear so that the student will be ready for the kinds of questions that he will be asked when he requests evaluation. If the problem statement alone does not provide enough cues, then the HELP options (10 and 11) and the associated text that is referenced should provide the necessary additional cues. If neither of these is

sufficient to help the student to surmount the problem, then he may use Option 9 to ascertain what answer is expected at that point. The students were instructed that Option 9 should be used only as a last resort. The fifth guideline calls attention to Option 9 uses.

The teletype sheets contain all of the above information and reveal the actions that the students had to take to satisfy each of the problems.

The fifth and sixth guidelines might indicate a wearing off of the novelty effect if there is good evidence that the student begins guessing or requesting the correct answer simply to avoid work. The teletype records leave little doubt whether sufficient work was accomplished for a given problem to provide a basis for an intelligent answer. Absence of such work when the nature of the problem required it, is likely to indicate that the answer was a guess. In addition, certain topics have questions that require only a "yes" or "no" answer but then have subsequent questions which ask for values on which the previous answer was based. These also provide insight as to whether the first answer was a conclusion or a guess. The presence of frequent guessing probably indicates a "beat-the-machine" attitude rather than an earnest desire to learn since there was apparently nothing to be gained by guessing. There were no added incentives for special performance of any kind.

Guidelines seven through thirteen pertain to the use of stored procedures that produce the computed results by simply specifying the appropriate option number. It is of some interest to note which options were used by the various subjects and whether some were used much more frequently than others. This information would obviously depend largely on the type of problems that are included in the STAT course.

In addition to the enumerated list of options that were used, more information can be obtained by examining the options used in conjunction with the survey questionnaire (Appendix G) which reports the subject's experience with the various statistical procedures on which the options are based. If a student uses unfamiliar procedures (level two or level three) when a more familiar one (level one) could have been used, this would probably indicate that the ease of using stored procedures has contributed to the student's willingness to try the less familiar one. If two or more comparable procedures are used while solving a given problem, the student is probably comparing results to verify his conclusions. One of the desired outcomes of the ease of using procedures is that the student will use what capabilities are available to investigate various properties of the data. For example, he can easily perform a chi-square test or obtain a correlation coefficient to investigate dependence between groups or check the results of parametric against

non-parametric tests. According to the ninth guideline, instances of this sort of behavior will be identified.

Instances appropriate to the tenth guideline suggest that the student might be pursuing some interest of his own which the data has suggested to him. For example, he might want to know if data that contain only ones and zeros will yield the same correlation coefficient from the Pearson product method as from the Spearman rank method, though he knows that he will not be tested on this question. These kinds of activities would undoubtedly be less frequent in situations where the procedures had to be computed by hand.

Guidelines eleven and twelve pertain to obvious indications of learning taking place. If the student calculated a procedure correctly after earlier unsuccessful attempts or after indicating a lack of familiarity with the procedure, then learning has taken place. In many instances, one can estimate the amount of learning that has taken place by observing the kinds of errors made in earlier attempts.

Conversely, guideline thirteen pertains to those procedures that the student attempted to calculate but was unable to master. These occurrences indicate a lack of adequate cues, at least for that student.

The last guideline specifies the identification of strategies which the subjects may have developed for solving the problems. Two in particular were considered

in the design of the STAT program: (1) comparing computed results with procedure results to verify one's work before being tested, and (2) using procedures only (no computation) on the first presentation of the problem to find out what is expected and then doing the computation on the repetition of the problem. The STAT program is so designed that the problem will not be repeated if the student performs satisfactorily on the first presentation, calculating all of his results and using no procedures. However, it may be more efficient to go through the problem twice, taking advantage of the speed and ease of using procedures to determine what computation is necessary, than to do the problem once by computation alone. This is especially true if the student is not sure about what he should do.

Presentation of the Data

The data that are tabulated in this section will be discussed in the next section. More detailed tabulations of the data appear in the appendices to which references will be given.

Pretest Results

The pretest scores were only used for screening purposes to identify those to be chosen as subjects. The results were regarded as either pass or fail. The scores of those who passed appear in Appendix I.

Criterion Test Results

The criterion test, given at the close of the experiment, yielded one score per subject. The summary statistics of these results are shown in Table 2. The test scores are tabulated in Appendix J.

The author inserted these criterion test scores into the STAT program and used several of the options to analyze the data. A segment of the resulting teletype sheet has been reproduced in Appendix K. Table 3 presents some of the results from that analysis.

Survey Questionnaire Results

Recapitulating, the survey questionnaire (Appendix G) consisted of the names of thirty statistical procedures incorporated in the STAT program and a means for categorizing oneself into one of three levels for each named procedure. Level one signifies that the person is familiar with the procedure and has at some time attempted to use it. Level two signifies only a familiarity with the procedure while having never attempted to perform the necessary calculations. Level three signifies a lack of familiarity with the named procedure.

The list of procedure names included very common procedures (e.g. arithmetic mean, variance, standard deviations, and those that were not so well known--Mann-Whitney two-sample test, Fisher Exact test). Keeping in mind the fact that the subjects used for the project had all completed an introductory course in statistics, the level

one category was expected to receive the most tallies. The results are summarized in Table 4 and are presented in more detail in Appendix L. Most subjects placed such procedures as the arithmetic mean, variance, standard deviations, t-tests and Pearson product correlation in the level one category while the level three category was frequently used for the Mann-Whitney, Wilcoxon, and Fisher tests, regression procedures and tests based on the chi-square distribution. Such a breakdown is very suggestive of the relative amounts of attention usually given to these procedures in an introductory statistics course.

Teletype Sheet Results

A complete copy of all work done at the teletype by each of the nine subjects was retained by the author. The "results" alluded to in this topic heading consist of observations made from the teletype sheets in accordance with the guidelines that have been set forth. Obviously, the data from several hundred feet of teletype paper cannot be reflected in detail by a few tables. The tables will embody only that condensed form of the data from the teletype records which has been specified by the guidelines.

The project covered five three-hour sessions. Since the first problem was used to orient the subjects to the STAT program, and all reported performance data began with the second problem in the sequence. Data from the

TABLE 2. Summary Statistics for
the Criterion Test Scores

	UCLA Computer Group	MSU Comparison Group
Mean	78.02	48.25
Standard Deviation	19.63	10.66

TABLE 3. Analysis of Criterion
Test Scores

	Variance F-Ratio	Means T-Score	Mann-Whitney Probability
Test Statistic	3.4933	4.2339	0.0007
Degrees of Freedom	8, 11	8	
Significance	$p < .05$	$p < .01$	$p < .01$

TABLE 4. Survey Questionnaire
Results

Level One	Level Two	Level Three
60.0%	16.1%	23.9%

Notes:

Level one indicates experience with the named procedure; level two indicates familiarity but no computational experience; level three indicates a lack of familiarity. Percentages indicate combined response for each level.

(

(

(

i

(

(

twenty-third problem was also omitted because of a malfunction in that problem. The remaining twenty-three problems constituted the STAT course and subjects worked on as many problems as time allowed. The nine subjects worked an average of fifty-seven percent of the problems. The subject making the least progress completed only three problems in addition to the orientation problem. One subject completed the last problem with only a few minutes remaining. However, he had skipped two other problem types in the sequence.

Since certain performance characteristics would cause the problems to be repeated, most subjects repeated one or more of the problems. Table 5 shows the number of problems and problem-types that each subject completed. Individual differences are very apparent in this table.

Work on these problems, plus the orientation problem, occupied five three-hour sessions for all except subject number seven, who was able to attend only the first four sessions. An average of nearly three problems per subject were bypassed by using the "STUCK" option.

Table 6 presents the data relevant to the guidelines. The data for the first thirteen guidelines are reported in percentages in the three categories: average for the nine subjects, highest subject, and lowest subject, respectively. In general, the percentages are derived from the ratio of the number of occurrences to the number of opportunities. Specific ratios are included in Table 6.

TABLE 5. Number of Problems and Problem-Types Completed

Subject Number	Total Problems Completed	Problem-Types Completed	Problems Skipped
1	9	9	0
2	5	3	3
3	20	17	4
4	27	21	6
5	20	15	6
6	22	15	2
7	12	12	0
8	32	17	3
9	10	9	1
Average	17.4	13.2	2.8

Notes:

There were twenty-three problems in all (exclusive of the orientation problem) each of which could be repeated one or more times, depending on performance.

TABLE 6. Summary of the Guideline Data

Guideline	Average	Highest	Lowest
A. Performance Characteristics			
1. STEPS Option STEPS used/STEPS available	73%	100%	33%
2. Errors Questions missed/ questions asked	29	44	14
3. BACK Option "BACK" used/"BACK" available	6	12	0
4. STUCK Option "STUCK" used/"STUCK" available	13	33	0
5. Answer give-away (Option 9) Option 9 used/questions asked	4	12	0
6. Suspected Guessing Occurrences/questions asked	14	22	6
B. LIBRARY Uses			
7. Extent of LIBRARY options used Options used/options available	29	87	0
8. Level three options used in preference to level one or two options	(No unquestionable occurrences)		
9. Two solutions used for a given problem Occurrences/problems completed	2	8	0

TABLE 6. Continued

Guideline		Average	Highest	Lowest
10.	Use of unrelated LIBRARY options	(No unquestionable occurrences)		
11.	Computation of LIBRARY options			
	Number computed without error/number encountered	45%	82%	22%
	Number corrected/number encountered	26	46	9
12.	Computation of level two and level three LIBRARY options			
	Number computed without error/number encountered	32	75	0
	Number corrected/number encountered	20	43	0
13.	LIBRARY options not successfully computed			
	Unsuccessful attempts/number encountered	16	33	0
	Options avoided/number encountered	13	67	0
C. Problem-Solving Strategies				
14.	Strategy observed--number of subjects using observed strategy			
	a. Practice trial	(1)		
	b. Fact gathering	(2)		
	c. Prompt when necessary	(4)		

Note:

The three observations for each item reflect the average, the highest, and the lowest respective percentage scores among the nine UCLA subjects.

Attitude Questionnaire Data

Item number I of the attitude questionnaire (Appendix H) instructed the respondents to divide on hundred points between this computer method of instruction and classroom instruction, based on the over all merits of each. In doing so, an average of 45 points went to the computer method and 55 points to the classroom method. The distributions ranged from as high as 70 to 30 in favor of the computer, to as low as 10 to 90 in favor of the classroom.

Item number 11 of the questionnaire asked the subjects to rate each of twenty-eight statements on the four-point scale: very true, essentially true, slightly true, and not true. One half of the statements suggested some advantage of the STAT course while the other half suggested a disadvantage. The statements were randomized to minimize the effects of response set.

Two different scoring methods were used to tally the results of the data. Each method was designed to bring out a particular kind of information. The first method produced a contingency table to display the interaction effects between the kind of statement and the way it was marked (See Table 7). The assigned weights were as follows: tallies in the "very true" and "essentially true" categories were given weights, two and one, respectively, under "affirming the statement." Tallies in the "not true" and "slightly true" categories were given

the weights, two and one, respectively under "denying the statement." The entries in Table 7 depict relative amounts of agreement and disagreement with the positively-worded and negatively-worded statements.

A second scoring method was used to depict a simple ranking of all the statements from those most often rated "not true" to those most often rated "very true." To do this, the categories, "not true," "slightly true," "essentially true," and "very true," were assigned the weights, one, two, three, and four, respectively. The questionnaire responses, weighted in the described manner, were then accumulated to produce Table 8. The statements are identified by statement number. The positive and negative statements were separated into two columns for clarity. The accumulated results of the weighted responses appear in the last column.

The attitude questionnaire included four more items asking the respondents to reply to open-ended questions. Those results are not amenable to tabulation so discussion of them will be deferred until the next section.

Discussion of the Data

Discussion of the various sources of data will follow in the same order that they were presented in the last section.

Pretest

Implied in the discussion of the pretest data is the question of the comparability of the groups of subjects,

TABLE 7. Interaction Among the Twenty-Eight Attitude Questionnaire Items

	Fourteen Statements Implied Advantages	Fourteen Statements Implied Disadvantages
Affirming	141	47
Denying	35	158

Note:

Responses of "very true" and "essentially true" were counted two- and one-affirming, respectively, while "not true" and "slightly true" were counted two- and one-denying, respectively.

TABLE 8. Ranking of the Twenty-Eight Attitude Questionnaire Items

Rank	Statements Implied Advantages	Statements Implied Disadvantages	Accumulated Weight of the Response
1	17		31
2	21		30
3	11, 18, 23, 25	3	29
4	6	14	28
5	12, 26		26
6	1, 2, 4, 10		25
7	9		23
8		27	20
9		15, 16, 24	16
10		28	15
11		13	14
12		22	13
13		7, 8, 19, 20	12
14		5	11

Note:

Statements indicated by statement number (See Appendix H), were ranked on the basis of the sum of their weighted scores. Responses of "very true," "essentially true," "slightly true," and "not true" were counted as four, three, two, and one points, respectively.

the comparison group (all MSU students) and the computer group (all UCLA students). Certain similarities of the students in both groups have already been noted. The obvious difference was the percentage of the two classes that passed the pretest. Because of this difference, the pretest was useful for identifying an appropriate comparison group at MSU as well as for assuring that the proper entry behaviors existed in the computer group.

Criterion Test

The data from the criterion tests were used to compare the two groups, the UCLA computer group with the MSU comparison group. Since no special instruction was provided for the MSU group, the comparison is not between two methods. Rather, the higher mean criterion test score of the computer group only suggests that they did learn from the course. Inferences that are drawn from this comparison should be regarded as very tentative until further support can be obtained.

Had the experimental environment been a more rigorously controlled one, the difference of mean achievement between the two groups would have been dramatic. As Tables 2 and 3 show, the computer group not only demonstrated higher mean achievement levels but also significantly greater variance than the MSU group. An increase in the variance of achievement levels is a phenomenon that consistently accompanies individualization of instruction.

The differences in achievement are such that one could reject an "equal means" null hypothesis at very small levels of significance. Several factors could have contributed to this difference: novelty, sensitization toward the test, extrinsic motivation (from being paid volunteers), and quality of instruction. Observations during the course of the project verified that a great deal of intrinsic motivation also developed in the computer subjects that kept their attention fixed to the task throughout all of the sessions. Novelty, alone, would not have sustained their motivation that long. It is more likely that the discovery elements in STAT contribute to the developing of intrinsic rewards.

Some subjects volunteered to continue the course beyond the scheduled project sessions without reimbursement. For these subjects, at least, monetary rewards could not wholly account for their enthusiasm. Another subject, however, admitted that the salary offer was a primary consideration for his becoming a subject. In return, he achieved the lowest criterion test score of anyone in the computer group--lower than the mean score of the MSU group.

The foregoing remarks do not preclude the inference that a respectable part of the achievement difference favoring the computer group was due to the STAT course. This cannot be a conclusive statement at this juncture nor can a statistical probability be assigned to its

credibility. However, the test results certainly lend support to further investigation.

The Survey Questionnaire and Teletype Sheets

Most of the discussion about the survey questionnaire is also related to the teletype sheets. Table 4 shows the response to the survey questionnaire broken down into the three levels of familiarity--level one indicating previous experience in the calculation of the named procedures, level two indicating only an academic familiarity with the procedure, and level three meaning that the procedure was not a familiar one. Many procedures (60%) were given the level one (familiar) rating because the list of procedures included many that should have been in the repertoire of students in a second semester statistics course. Those procedures most often receiving the level three rating included tests based on the Chi-square distribution and other non-parametric tests (Mann-Whitney U-test, Wilcoxon Test of Paired Observations, and Fisher Exact Test). The subjects confirmed verbally that less class time had been devoted to these topics than to normal theory statistics.

The teletype sheets contained all of the actual work of the students on the STAT course during the five three-hour sessions. Also included in the sessions were the orientation problem (first session), the criterion test, and the attitude questionnaire (end of the last session).

Averages for the nine UCLA students show that they each completed about seventeen problems of which four were repeats. Some problems took much longer than others.

Subject number eight had the highest percentage of repeats which was largely due to the strategy that was employed. This will be discussed later. Subject number two completed the fewest problems. It was he who was chosen to be one of the three interviewed by telephone, having been identified earlier because he was experiencing difficulty with the course.

About thirteen percent of the problems fell into the "problems skipped" category. This number is not large enough to cause concern. However, it could, and in a few cases did, represent entire topics that were skipped. Guideline number four deals with that topic, so further discussion will be deferred until later.

The observations set forth in the guidelines were extracted from the teletype sheets together with appropriate survey questionnaire data (See Table 6). These data were reported as percentages since the significance of the number of times each of the various actions were taken depends on the opportunity for taking that action. For example, a fixed number of STEPS were available for each problem-type. It is of interest to know for what proportion of problems the subjects used the STEPS option and, since the number of completed problems varied among the subjects, a percentage score is the most meaningful.

Thus, one-hundred percent does not necessarily signify that all of the STEPS were seen by any student, only that some student employed the STEPS option for every new problem he worked. The principle is the same for all percentage scores in Table 6.

The STEPS option was used frequently--on about three problems out of every four. There were no constraints placed on the use of this option.

The average error rate was high. One subject missed nearly one out of every two questions asked. If one worked all twenty-three problems without making any errors or repeating any problems, a total of thirty-nine evaluation questions would have been asked. Repeated problems obviously increased the total number of questions per subject. Of those questions that were missed, attempts were made to recompute sixty-eight percent. Figure 7 shows the breakdown of how the errors were rectified. There were also several incidents, not reflected by the graph in Figure 7, where a subject would obtain the correct answer either by using Option 9 or the appropriate LIBRARY option and then attempt to derive that answer by recomputing the procedure, after which he would resubmit his answer to the missed question.

In those cases where the answer was recomputed unsuccessfully, the subjects usually used Option 9 rather than computing the answer for the third time. However, in several isolated instances, subjects recomputed the

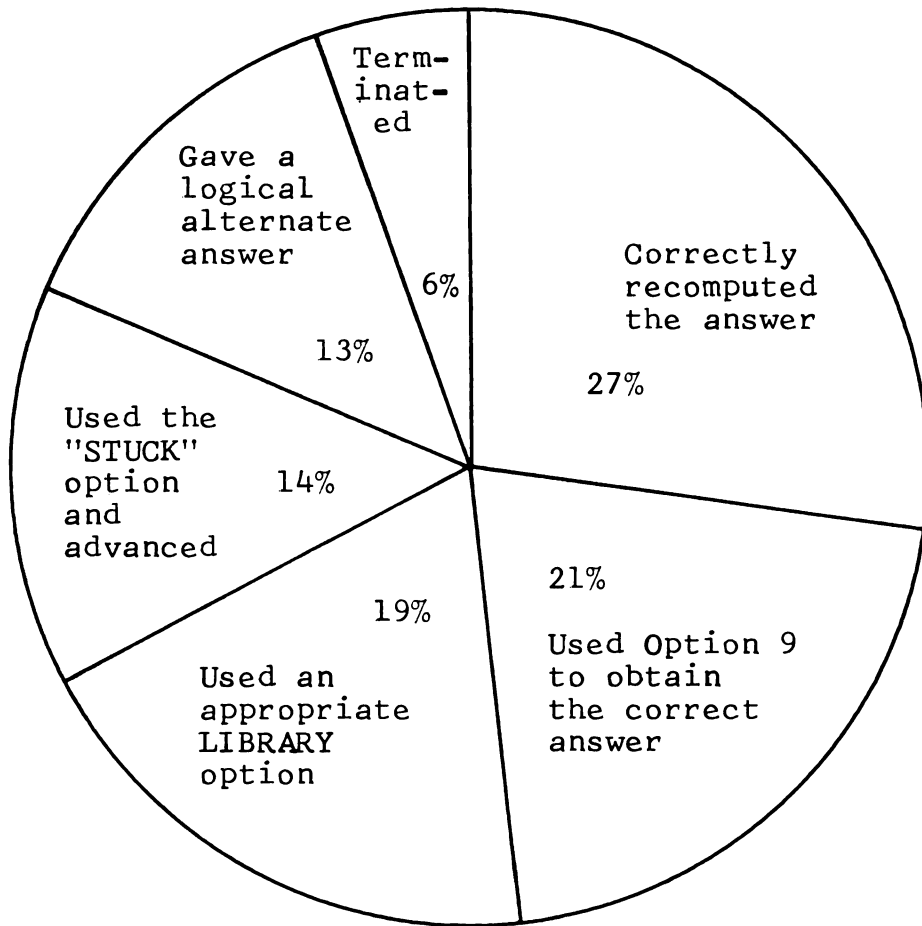


Fig. 9. Means Employed to Satisfy or Avoid the STAT Evaluation Questions. The graph reflects the subjects' actions subsequent to missing a question. Only those who employed the "STUCK" option were not presented with the same question again. Of the other means that were employed, all except the "termination" option provided another answer to submit when the question was seen again. Those who chose to terminate started the problem over again when they resumed.

The graph represents a composite of all questions which were missed by all nine subjects.

answer two and three times. Sometimes their dilligence paid off and sometimes not.

Use of the "BACK" option was negligible. No one made it a practice to view the evaluation question first before attempting to do the required work. Instead, the BACK option was used almost exclusively for the purpose for which it was designed--providing a means of shifting back into the calculation mode because a question was asked that had not been adequately anticipated. The subjects did not abuse this privilege.

The "STUCK" option was most frequently used on the repetition of a problem that had given the subject trouble on the initial presentation. Most subjects would at least make a token effort even though their misunderstanding still existed. In some cases they were successful, but often they would give up on the problem, use the "STUCK" option, and advance to the next problem in the sequence. In a very few cases, subjects used the "STUCK" option on the first presentation of a problem after having invested very little work in the problem. When this happened, one could find a previous problem that had certain similarities which had caused him trouble. In these few such instances, "STUCK" was used to by-pass the problem.

Requesting the answer through use of Option 9 was the second most frequently-used method for finally coming up with the right answer (See Figure 7). For one thing,

it was a sure method. Use of Option 9 would always yield the exact answer that was expected on the question just previously missed. Also, Option 9 could be used in any problem, whereas the LIBRARY options could only be used on the first presentation of a problem. In the opinion of the author, Option 9 was a little too convenient. The only constraints placed upon its use were: (1) that the student had to attempt to answer the question first, and (2) the exhortation on the student's reference sheet, "use as last resort." It would seem to be more desirable to require a more concerted effort from the student before Option 9 is made available. Use of Option 9 too often appeared to be a function of the amount of computation required. Option 9 was used more frequently for chi-square results than for any others, and chi-square procedures required the most computation. Non-parametric rank statistics came next, both in the amount of computation and use of Option 9. The subject usually did the required computation, taking from twenty to forty-five minutes, and then if the answer was wrong, he would use Option 9 instead of looking for his error. In a few cases, after having done similar computations for earlier problems, the subject would simply guess at the answer and then use Option 9, avoiding all further computation on that procedure. This was not true, though, for procedures with more moderate computational requirements such as t-tests and regression coefficients.

Another speculation concerning their lack of enthusiasm for the chi-square procedures is that the subjects entered the experiment with a negatively-biased attitude toward their validity. They reported verbally that their professor had expressed doubts about the usefulness of chi-square techniques, and had only superficially covered them in class. In the STAT course, use of Option 9 may have been employed simply to avoid investing time unnecessarily on this topic.

Some guessing did occur as the data for the sixth guideline (See Table 6) clearly showed, though not all of the guesses lacked a basis for forming a judgment. Many of the questions, typically those which asked for a decision on significance, were answered either "yes" or "no." Since all sample data originated from a random numbers generator, it was inevitable that differences between some samples would be more pronounced than others. Sometimes the samples exhibited characteristics which made the final decision obvious without any computation. In other cases, the means alone would provide an adequate basis for a decision if they were sufficiently different to make the test trivial. Between the extremes, was a range of situations where the need for computing a test is debatable. Therefore, many answers that were called guesses were really based on observations of extreme differences that made statistical testing unnecessary. Other guesses fell into the questionable region. Perhaps

the subject was quite sure of his answer. Some were outright guesses, revealed by a lack of previous computation coupled with an incorrect answer. The incidence of the proven guesses in the latter category was very small--an average of about one blind guess per subject.

In the main, there was little evidence of frustration. The characteristics exhibited in the first six guidelines would compare favorably with any typical assignment over comparable statistical subject matter. There will always be those who guess at and skip problems. The average time spent per problem is certainly no more, and quite likely less, than similar assignments would usually take. The orientation to the STAT system occupied from one to three hours per student, but even this has a counterpart in the familiarization of one's self with a desk calculator.

Subject number two found the STAT course to be too difficult and frustrating. It was he who used the "STUCK" option for the greatest percentage of problems. That he experienced frustration was later confirmed both in the telephone interview and on his attitude questionnaire. However, for the other eight subjects, the data from the first six guidelines gave no hint of general frustration with the course nor did they report being frustrated.

The next seven guidelines, seven through thirteen, pertain to the LIBRARY options. Use of the options is

the basis for four of these. It came as somewhat of a surprise that the LIBRARY options were used as sparingly as they were. Also, notice the wide discrepancy among subjects in the data for the seventh guideline (See Table 6). One subject used eighty-seven percent of the available LIBRARY options at some time during the course. That subject not only used more LIBRARY options, but used them more frequently--forty-four times compared to none. An average of fourteen requests were made from the LIBRARY among the nine subjects. Many of the subjects seemingly used the LIBRARY options as a last resort, much like Option 9, even though there was no such admonition made. They were told, however, that they would repeat the problem if LIBRARY options were used. The second time through the problem, the LIBRARY was not accessible. Apparently, this was enough to cause most of the subjects to regard the use of the LIBRARY as incurring a penalty. That this was not necessarily the case is indicated by the performance of subject number eight who used the LIBRARY more than three times the average. But most of the subjects seemingly didn't view the LIBRARY options as an aid to understanding and solving the problems, but rather as something to fall back upon when their computational efforts failed. Because of this apparent attitude, the data for guidelines eight, nine, and ten (or the lack of it) is not surprising. There were a few isolated occurrences that could have fit the eighth and tenth guideline

categories. However, in all such cases, it appeared equally likely that the option had been used mistakenly rather than by intent.

Guideline number nine had no occurrence of similar LIBRARY procedures being used to compare results for a single solution. There were, however, a few instances where a subject both used a LIBRARY procedure and computed its result in order to verify that he had done it correctly. The reported data for the ninth guideline pertained to these occurrences. There were also several instances where the subjects based their decisions on procedures which were different from the ones that were used to check their answer. Often, the procedures were sufficiently valid so that the answer submitted by the subject was acceptable. The most frequent example was the use either of a rank test or analysis of variance techniques in place of a t-test.

The data for guidelines eleven through thirteen in Table 6 reflect the subjects' performance in computing the various statistical procedures that they encountered in the course. An average of forty-five percent of the encountered procedures were computed correctly on the first try and ultimately, seventy-one percent were computed correctly. Correctness was inferred from having had an answer judged correct without having first used the corresponding LIBRARY option or Option 9 in that problem. In most cases, the students' work leading up to the

solution was also checked by the author. The individual differences are especially pronounced on these three guidelines as demonstrated by the second and third data columns in Table 6.

The percentages were smaller for the computation of the less familiar (only Level II and III) options. Fifty-two percent of these were computed correctly--thirty-two percent on the first try. Guideline number thirteen accounts for the twenty-nine percent of encountered options that were not correctly computed. Approximately half of these were attempted and the other half skipped. In all, the subjects at least attempted to calculate eighty-seven percent of all the statistical procedures they encountered in the course, each procedure being represented by a LIBRARY option number.

Correlation coefficients were computed between the criterion test scores and each of the first thirteen guidelines for which data are listed (See Table 6), and, in addition, the "problems completed" and "problem-types completed" categories from Table 5. In all, sixteen correlation coefficients were computed. Using the sixteen categories as predictor variables, a multiple correlation coefficient was also computed. The individual correlation coefficients were generally small. Only five of the predictors had more than ten percent of their variance in common with the criterion test data: (1) BACK option ($r = .556$), (2) STUCK option ($r = -.436$), (3) use

of Option 9 ($r = -.439$), (4) options computed wrongly and corrected ($r = .763$), and (5) option computation avoided ($r = -.447$). Each of these is consistent with the expectations of the author. One and four constitute actions that would tend to promote higher achievement in contrast to the other three, whose correlation coefficients are negative, corresponding to undesirable characteristics. The multiple correlation coefficient, based on all sixteen predictor variables, was .999. As the multiple correlation coefficient reflects, the resulting regression equation was highly accurate in predicting criterion test scores. The unusually high multiple "r" was reassuring in that it supported the relevance of the behavioral phenomena that were chosen to be observed--the basis for the guideline data.

Appendix M gives a breakdown of the data used for calculating the correlation coefficients.

The fourteenth guideline deals with problem-solving strategies which were evident in the project. Three were identified: (1) practice trial, (2) fact gathering, and (3) prompt when necessary (See Table 6). Of the three, the first was most easily identified as a planned strategy. It was only evident for one subject--subject number eight. The practice trial strategy amounted to using LIBRARY options almost exclusively on the first presentation of each problem type while apparently devoting attention to other than computational aspects of the

problem. Use of the LIBRARY options caused the problem to be repeated and the LIBRARY options to become inaccessible. Now, however, the subject has had experience with the problem and does the required computation. Table 5 showed subject number eight to have completed the most problems, largely due to the strategy used. However, only one subject completed more problem-types than she did, indicating that the additional problems seen may not have slowed her progression through the course. Since computation was the most time-consuming activity, it is likely that using the LIBRARY exclusively on the first presentation might have saved some time that others lost in misdirected computation.

Two subjects persisted in the strategy of immediately obtaining a certain set of information (fact gathering) when they began a new problem. That information consisted of the summary statistics (sums, sums of squares, etc.), Options 2 through 6, and STEPS to the solution, Option 11. Then, after obtaining this information, they proceeded to solve the problem.

The third strategy, prompt when necessary, used the above options, 2 through 6 and 11, only when the need for that information developed.

The remaining two subjects exhibited no uniform pattern of problem-solving behavior.

None of the strategies stood out as being superior with respect to any reasonable criterion. If problem-

types completed or criterion test scores were considered, there was more variance within strategies than between them. The first strategy involved the most extensive use of the features of the STAT course. However, no particular advantages of this strategy were clearly demonstrated.

Attitude Questionnaire Results

The attitudes toward the STAT course were extremely positive. However, conventional classroom teaching was rated slightly higher when a choice was made between the two (classroom, 55 points; STAT course, 45 points). This response was to be expected since the STAT course is not a self-contained course in statistics. Hence, if a student had to place sole reliance in one or the other, and was given a choice, the classroom is the better alternative.

Responses to the twenty-eight statements in the attitude questionnaire displayed strong, positive attitudes toward the STAT course. The ratings were weighted in order to summarize the respondents' assessments of the essential truth of the statements. Table 7 describes one of the two summaries and the weighting method that produced it. The entries in the table, the accumulated weighted scores, show such a strong dependence between the two variables that a test for independence would be trivial. The statements which suggested advantages drew more affirmative ("essentially true" and "very true")

responses while the statements which suggested inconveniences drew more denial ("slightly true" and "not true") responses. Only one respondent failed to exhibit this trend. Subject number two agreed with more statements, both positive and negative, than he denied. His weighted response totals were: affirming positive statements, 11; denying positive statements, 7; affirming negative statements, 14; denying negative statements, 7. His attitudes probably reflect the difficulties he had encountered in the course. He completed the fewest problems. His criterion test score was low. In interviewing him by telephone, some of these problems were discussed. He expressed doubt that he was ready for this kind of course. These interviews will be discussed more fully in the next section. In general, however, the positive statements were more readily affirmed.

Table 8 reflects an attempt to rank the twenty-eight statements according to the combined responses based on assigned weights of the rating categories. By assigning the weights, one to four, to the four rating categories, the accumulated responses for each statement were represented as the sum of the weighted responses. The statements were then ranked according to their scores, producing the information in Table 8.

The statement numbers in the second and third columns of Table 8 correspond to the statement numbers in the attitude questionnaire (Appendix H). Statement number

seventeen, "I liked the idea of operating a computer," received the highest endorsement among the nine subjects. Statement twenty-one, "I liked the ease of making computations" was next. Both of these were considered to be advantages of the STAT course. Statements eleven, eighteen, twenty-three, twenty-five, and three shared third place. Statement three ranked the highest among all of the statements of disadvantages. It read, "There were times when I didn't know what I was expected to do." Another prominent critical statement was the fourteenth, "I expected more explanation from the computer and less reliance on the book." Other than the third and fourteenth statements, none of the critical statements achieved the rank score of the lowest ranked positive statement, number nine. In addition, the two widest gaps in the distribution of accumulated weight scores occur just below the lowest-ranked positive statement. The statement that ranked lowest, number five, pertained to the option number format used throughout the course, "The use of numbers to request procedures was confusing." Clearly, it was not.

Certain characteristics seemed to be more prominent after the statements were ranked. Statements pertaining to novelty, ease of doing calculations, and enjoyment of the experience were ranked very highly. The STAT course obviously had the "novelty effect" working for it. If this is a transient effect, then it could hardly be

considered a lasting advantage. However, there was no discernable loss of enthusiasm through any of the five long and intensive sessions. That the "novelty effect" is alone an adequate explanation is debatable. Elements of "gaming" are also undeniably involved. The subjects appeared to be in a sort of competition with the decision structure of the course. This conclusion became evident as the subjects attempted to solve the problems while using as few of the built-in aids as necessary. Most subjects actively avoided taking actions that would cause them to repeat a problem as though this constituted a failure on their part.

Discovery elements described in Chapter II provide a rationale for the development of the kind of intrinsic motivation demonstrated by these subjects. The intrinsic motivation that develops from this kind of computer-human interaction is intense and, if novelty is the full explanation, no one seems to know how long it will last. According to the rank of statement seven, the respondents thought that it would last for some time. It is certainly advantageous while it exists.

"Ease of calculating" and "using stored procedures" were two of the primary objectives of the project. It was reassuring to note that the responses of the subjects were very positive for the statements which were related to those objectives--statements 5, 11, and 21. (The rank position of statement 5, "The use of numbers to request

procedures was confusing," suggests that the method was not at all confusing.)

Statements 15, 24, and 26 are all related to the "help" provisions in the STAT course. The position of the above statement numbers in the distribution of accumulated weights in Table 8 clearly indicate less enthusiasm for these options than for calculation capabilities.

Recall that all help consisted of referring the student to the appropriate pages and paragraphs in the textbook whether the help was requested by the student (Options 10 and 11) or provided involuntarily as a consequence of an incorrect answer to an evaluation question. The position of statement 3 in Table 8 indicates that the help was not specific enough to the need. Apparently, there were times when the subjects did not know how to proceed and there were no directions available that provided the necessary information. Statement 14, "I expected more explanation from the computer and less reliance on the book," was also frequently affirmed by the respondents. This bit of evidence supports a view which is in direct contrast to one of the goals of this project, that of avoiding the inclusion of static textual materials in the computer memory. Neither is it probable that the situation would have been improved much, if any, by having the computer type out, word-for-word, the explanations that the student was to read from the text. The

view taken by the author is that the help section, to be more effective, should include interactive remedial sequences to insure that the requested help is properly conveyed and a memory feature that considers former help requests and elaborates accordingly.

The first objective of the project was intended to preclude the use of expensive computer storage for static textual information. However, the above considerations suggest an expansion of that objective. Though the textual information may be static in that the computer effects no changes in the sentence structure, the need for a controlled presentation is sufficient to warrant the additional space required for its inclusion. If information is stored in the computer program, it should be organized in relative short segments, especially if it is to be displayed on a typewriter. Waiting for excessively long passages to be printed can be tedious. It is expected that future versions will contain additional information in the computer storage.

The remaining statements from the attitude questionnaire each pertain to some characteristic of the STAT course and the position of that statement in the list in Table 8 suggests something of the importance of that characteristic to the subjects. Statement 25, for example, ranks high, suggesting that the subjects liked the feature of choosing their own sample sizes and determining the kind of test they would make. Other

statements similarly reflect their preferences.

The attitude questionnaire concluded with six questions about the general reaction of the respondents to the use of the computer, each of which allowed space for a reply of two or three sentences in length. The responses were quite uniform and will be summarized briefly below:

1. What use have you made of the computer before?

Two subjects had used statistical computer programs to analyze data. A third subject had worked out some electrical engineering problems on a computer. The remaining subjects had no former experience.

2. Did you have misgivings about using a computer before you began here? (Please elaborate.)

One subject expected it to be more difficult, otherwise there were no misgivings.

3. In what ways might this experience have helped to change your attitude toward computational uses of the computer?

Most subjects expressed a positive attitude which they had held prior to the project. One subject now expects to benefit directly from computers as a result of this experience. Three expressed positive changes in attitude. One subject commented about the time it saved; another that it removed the distraction of doing computation.

4. Did this experience using the computer influence your thinking concerning future statistical work you might want to do?

Five subjects responded in the negative. The other four related the following kinds of applications that they had not previously considered: (1) use of the computer for all types of statistical work, (2) availability of so much statistical help, (3) instructional uses of the computer, and (4) speed and power of the computer.

5. Will you likely be using statistics in your career?

All subjects responded in the affirmative.

6. To what extent will your work be likely to involve a computer? Directly? Indirectly? Not at all? What uses:

Six subjects expected to submit data for analysis. One anticipated little need for computation aid. The remaining two planned to use computers in research, in test construction, and multi-variate analysis.

In brief, all subjects entered the project with positive feelings about what they could expect from a computer. The expectations of those who had no former experience were confirmed. The others had some awareness of what to expect. Only in the case of subject number two was there any doubt whether the experience was a positive reinforcement. He attributed his problem to a lack of readiness on his part rather than any failure of the instructional system. It should have been the function of the pretest to detect his deficiencies. The test could certainly be improved by current validation techniques. However, there will always be the "false posi-

tives" who will pass the most rigorous test. By putting more effort into the validation of the screening instrument, one could reduce the probability of putting unprepared students into the course.

Space was provided on the attitude questionnaire to cite suggestions regarding the course. Several subjects did. The suggestions given correspond to the conclusions that have been discussed throughout this chapter regarding ways of providing more adequate help and imposing fewer restrictions on the use of LIBRARY options. They will be a valuable aid in making improvements to the course.

The Telephone Interviews

The telephone interviews brought out nothing new. Each of the three subjects considered the course to be a valuable adjunct to classwork, providing insights through manipulating the data that they hadn't experienced in the classroom. Subject number two, one of the three interviewed, felt that he could get much more from the STAT course if he could come back to it at a later date. Several volunteered an interest in continuing work on the course. All of the subjects were interested in the research aspects of the STAT program as well. They wanted to insert their own data into the program and experiment with it. Unfortunately, no time was left over for this. When asked what they thought the long term effects would be, they thought that students would not lose interest,

especially if some improvements were made in the help area. Two of those interviewed volunteered that they thought they would remember what they had learned much longer than if the instruction had been more conventional. Several of the subjects, both in the telephone interviews and in footnotes on paper, expressed appreciation for a rewarding experience.

VII. Conclusions and Implications

Several different observations indicated that the nine experimental subjects did learn from the STAT course. Criterion test scores also verified this. Certain instructional goals had been laid down for the STAT course and these goals were achieved, to a greater or lesser degree, by each of the nine subjects. The author is confident that most subjects could have attained higher criterion test scores had they been given enough time to complete the course. However, the fact that the computer course did teach in some measure what it purported to teach was not considered to be a significant finding of this study. Most instructional systems, regardless of their quality, teach something.

In keeping with the stated purpose of the project, the STAT program did exemplify a computer-assisted instructional program utilizing that combination of capabilities which only a computer has offered to date: (1) dynamic information storage and retrieval, (2) easily accessed stored procedures, (3) extensive branching, and (4) rapid computation. The role of each of the four named capabilities is easily identifiable in STAT. Moreover, the data indicate that the STAT coursework was reasonably appropriate with reference to the stated instructional objectives.

Computation was relatively efficient. Fewer buttons were pushed on the teletype terminal to calculate results for statistical procedures than would have been pushed on a desk calculator to come up with the same answer. Since the data were automatically inserted into the calculation mode ready to be manipulated, this in itself effected a time advantage.

Every LIBRARY option that was relevant to the STAT course was used at least once. Some subjects used the LIBRARY options much more than other subjects did.

Perhaps the most relevant and practical conclusions of the project are its implications for future developments of this and other similar computer-assisted courses.

Specifications for a CAI Author-Language

Experiences with the STAT course have suggested many features that computer-assisted instruction should incorporate.

The STAT course was written in JOVIAL, a standard FORTRAN-like computer programming language. This necessitated the author to also be a programmer. People who are qualified to be course authors are usually not computer programmers. These subject matter specialists are public school teachers, college professors, curriculum personnel, and vocational educators who will use a computer system only if the advantages outweigh the inconveniences. Until now, one of the big obstacles has been the amount of training necessary to use one. These people

usually do not have the time required to learn computer programming techniques and, even if they did, the inconveniences are numerous. If this situation is to be remedied, some other intermediate computer language must be developed which has a syntax that is more congruent with the objectives of these instructional course authors than general purpose programming languages now provide.

The first feature that should characterize such an author-oriented language is: (1) an author should be able to prepare a computer-assisted course after only a brief orientation to the system. This situation is not currently true where lessons must be prepared in standard programming languages such as FORTRAN or JOVIAL. Just getting stored information out to the student requires page formatting and instructions that will transfer the information to the proper output device. Replies from the teletype must be dissected, converted, matched, and finally interpreted in some decision context. However, a course author is working in a much more restricted domain than the general computer programmer. Many of his desired activities could be anticipated by an appropriate intermediate program. For example, when he is typing text to be stored as a part of his lesson, he should be allowed to type it just as if he was using an ordinary electric typewriter and then expect it to reappear to the student in the same form as it was originally typed. If he wants to incorporate directions for requesting the student to

reply, he should only need to designate that desire, type a list of the numbers, words, or phrases that should be searched for in the reply, and associate some action with each of the anticipated possible replies, thus establishing a course sequence. A one-hour orientation session should be sufficient to familiarize the author with the above procedures. Included in this author-language must be convenient methods for correcting mistakes, preserving the lesson, and making it available to many students simultaneously.

The STAT course required approximately one-half of a man-year to prepare while providing only from twelve to twenty hours of instruction. Since such large preparation-to-instruction ratios could not be tolerated in practical situations, (2) an author should be able to prepare his lesson with maximum efficiency. The preparation-to-instruction ratio should be no larger in the worse case than it is for preparing comparable lessons independent of the computer. Where patterns can be identified in the lesson, the author-language should enable the author to use previously-prepared information to facilitate lesson preparation. Drill questions on arithmetic problems or spelling are particularly amenable to this. Perhaps certain sections of a lesson may be needed elsewhere. In these cases, a good author-language can lower the preparation-to-instruction time ratio by providing for a cross-reference to the other material rather than insert-

ing it twice or more. The STAT course allowed subjects to repeat problems, but with different samples. An author-language should allow this type of lesson to be built.

The author-language system should not be responsible for any more delays than absolutely necessary. Whenever an addition or a change was made to the STAT program, the entire program had to be processed or "compiled" before it was ready for use. If the computer is used heavily, it may take from hours to days to get this done. That time must also be taken into account in the total lesson-preparation time. A more desirable alternative would cause the course to be displayed "interpretively" from the lesson materials that were prepared by the author. With this method, the lesson would be ready to use the moment it was entered. The author could "try out" parts of his lesson while he was preparing it and make desired changes right then. If an error showed up while a student was working on the lesson, it could be corrected and the student could immediately resume.

One of the frequent criticisms of the STAT course is the rigidity of the sequence of instruction. Students could not review. Skipped problems were not presented later. Error feedback was cryptic and redundant. If one committed an error on a certain question twice in a row, he could get the same feedback message both times. The LIBRARY options were either all available or all disallow-

ed. No attempt was made to allow unconditional use of LIBRARY options for which the student had demonstrated competence. This would have effected a significant saving of instructional time, and would have enhanced the course. To have included that feature in the STAT course would have greatly increased the programming task. However, it was clear from observations that (3) the author must have complete control over the presentation of his lesson. This implies that the author must be able to identify entry points within his lesson and cause a "branch" of the instructional sequence to any one of these points to be contingent on some student performance characteristic. This also implies that lessons must be protected in such a way that only authorized persons will be able to display and change the "code" of the lesson that assesses performance and controls the sequencing. The student must only see that which he was intended to see. In the STAT course, use of Option 9 after a question was missed brought out the correct answer, indiscriminately. The author ought to have control over the availability of such options. In order to guarantee complete control over the lesson, (4) the author should be able to specify, in some uncomplicated way, exactly what part of the total lesson environment will be available to the student in any given segment of the lesson sequence. Instructional lessons are ordinarily composed of bits of learning material assembled into topics, each having a

prevailing contextual significance. An adequate author-language for a computer system would allow the author to establish an environmental context over designated portions of his lesson. Helps, prompts, feedback, messages, explanations, and computational aids would be relevant aspects of the contextual environment and appropriate controls could be effected.

To get the most for the money, (5) an adequate author-language should exploit the capabilities of the computer system. Many input and output devices are commercially available. In addition to the familiar typewriter, there are graphical displays, "light pens" for writing with light, electronic Rand tablets, mechanical plotters, audio message composers, micro-film displays, and many more. Certainly, the availability of a selection of input and output media enhances any educational system. The type of lesson to be taught will usually suggest the most desirable configuration of equipment. Monetary considerations also come in at this point. However, the computer itself inherently has a set of unique capabilities that can be exploited in any computer-assisted instructional system. Four were mentioned earlier--dynamic storage and retrieval, stored procedures, extensive branching and rapid computation. These capabilities are implicit in many of the specifications given in this chapter. If these characteristics did not exist, the computer would offer no more than a teaching machine does

at a fraction of the cost. Unfortunately, some developers have been satisfied if their computerized instructional system performs all the functions found in teaching machines and little else!

Computers are excellent monitors. The instructional system that displays the lesson to the student should automatically keep a comprehensive set of records tracing his progress through the lesson. The records should include such things as error information, response latencies, identification of anticipated replies, tracking at branch points, use of helps, prompts, and computational aids. By automatically storing such records independently for each student as he proceeds through the course, (6) the author should at anytime be able, by writing certain queries into his lesson, to alter the sequence of the lesson on the basis of the performance history of that student. It would be more desirable if the queries would utilize information which had been automatically stored so that the author would not have to anticipate future decisions by attaching "counters" to responses. But provisions should also be included to enable the author to base decisions on actions that he himself has specified which may not be included in the stored records.

The flexible decision capability was notably lacking in the STAT course. Decisions, there, were the same from problem to problem. Subjects soon became aware of the results of certain performance characteristics. They knew

that one use of a LIBRARY option would make them repeat the problem. Some subjects noted comments on their teletype sheets after seeing a certain error feedback message repeatedly. One subject replied, "no kidding!" Such rigidity is certain to have adverse effects on the learning task.

The computer is probably best known for its ability to manipulate numbers, yet relatively few CAI systems make much use of that capability. Hence, (7) the instructional system should include an extensive computational aid to be used both by the author and the student.

Authors should be allowed to specify numerical answers in the form of equations, formulas, or functions. If a statistical sample is part of the lesson for which the student is expected to compute a certain result, the author should be able to specify the correct answer by naming a function that will yield the result. There is no reason for him to compute the number; the computer can do that. Then, after establishing such a function, it might be used in other questions for other samples or in building other functions. Certain basic functions such as exponentiation, factorials, trigonometric functions and selected tables could be included in the system and in addition, provisions could be made for the author to write complex functions of his own. Having provided this capability for the author, he is then given the ability to allow all or part of the same capability to the

students. Prior to entering their answer, the students might use the computational facility to do their work in much the same way as the STAT subjects used the calculation assistance option. However, a further development of this capability should allow the author to write functions in addition to a basic set of "primitive" ones, and have dynamic program control over which are available to the student at any given place in the lesson. This would replace the LIBRARY options in the STAT course. This provision also would allow the student to write his own functions, much like the capability given to the course author, which are his to use, unconditionally. The fact that he could write such a function would clearly demonstrate his mastery of the procedure and then, having written the function, he would be relieved of the redundancy of performing similar computations for future problems. A reference to his function would do it for him.

The STAT course did not contain statistical tables. Rather, the student was asked to supply the table entry from a specified page, column, and line in the textbook. This was not the best solution. Several mistakes occurred from entering the wrong number. This affected not only the students' work, but evaluation as well. Such tables usually require large amounts of space within a computer but their inclusion is advisable, if possible, when they play such a crucial role.

Using the computer to prepare a lesson, calculate an arithmetic statement, define a function, edit a lesson, ask a question, or reply to a question presents no insurmountable challenge to a system designer if each kind of task is regarded as a "mode" of operation. Then, some set of symbols and/or command words are used to switch from one mode to the other. Thus, the lesson might ask a question and, before answering, the student might switch modes and perform some computations. Then he might type, "READY," to begin answering questions again. In the STAT course, typing the word, "END," caused a return from the calculation mode and Option 8 activated the evaluation mode.

The domain of possible replies were very limited in the STAT course, consisting mainly of numbers or the words, "YES" or "NO." Only in the calculation mode was there much freedom granted for composing a reply and, even there, strict adherence to the legal symbols was a must. To have a truly flexible author-language, (8) the author should be able to identify the form of a reply or elements within a reply without having to identify the entire reply. It would be extremely useful to be able to anticipate a certain word or phrase even if parts of it were misspelled. Often, the author would like to detect the presence of a "key word" or a group of such words. He should be able to discriminate a positively-worded reply from a negatively-worded one without imposing a

limited vocabulary upon the student. If he lists an algebraic formula as an answer, any algebraically equivalent formula should be recognized. In general, (9) if the author can specify exactly the characteristics of the reply that the lesson should recognize, it ought to do it. There will be limitations placed upon the course author by any instructional system which violates this ninth specification, but the degree to which this is implemented will be the measure of the system's responsiveness to the students.

Most of the subjects in the STAT project reacted favorably to the fact that their problems were unique. This characteristic requires the answers to be dynamically checked by the computer. They also expressed positive attitudes toward choosing their own sample sizes--also necessitating a dynamic-answer check since not even the course authors could anticipate the correct numerical answer. Dynamic checking of numerical answers has already been discussed under the topic of functions. However, (10) a flexible author-language will allow dynamic-answer checking of both numerical and non-numerical answers. To realize this feature, the author-language would need to allow the author to itemize a list of alternative answers, one of which will be correct. But, the author himself may not know which answer will be correct when he writes the lesson; the system will determine that at the moment the decision is required.

In order to accomplish this, the author-language must provide a means by which the author can specify the conditions under which one answer or another would be correct and also the action that is to be taken should the reply be right or wrong. For example, a statistical problem might ask if an hypothesized mean score is inside or outside a certain confidence interval constructed on a randomly-generated sample of scores. The sample generator parameters can be specified so that sometimes "inside" will be correct and sometimes "outside" will be. The author lists "inside" and "outside" as the two possible answers and also writes the statistical formula that will determine the correct answer. The actual determination of the correct answer will occur only when the student makes his reply to that question, using the same data he used as a basis of judgment.

Such a procedure could be used for a broad class of questions and answers. The effect would be that of making the system responsive to the interaction going on between the computer and the student. It is this kind of responsiveness that probably helped to motivate the subjects in the STAT course.

If an author-language was developed expressly for a restricted topic, then extensive course-building aids could be built into the languages. Suppose, for example, the topic was statistics; a comprehensive library of statistical procedures could form a part of the language.

However, practical considerations make a more general language necessary. The investment of time and money in such a language compels one to make it useful in a wide variety of applications. Hence, (11) an author-language should contain a repertoire of instructions which are sufficient to allow the most complex kinds of activity to be specified that the computer is capable of handling. This is generally referred to as the "power" of the system. It should be sufficiently powerful to allow the maximum capabilities to course authors who, through hours of use, become sufficiently acquainted with the system and want to make more sophisticated use of it. It might be used to conduct interviews or schedule classes.

The apparent dilemma between a system which is convenient enough for the naive as well as powerful enough for the sophisticated is not insoluble. The language can be constructed in "modular" form so that a naive user only concerns himself with a very small subset of the total capabilities. This subset, though small, can be completely adequate for the preparation of a wide variety of instructional lessons. The communication vocabulary should be mnemonic and easy to learn. A simplified user's manual will describe only the use of these capabilities. Then, as the author becomes more sophisticated, he graduates to a more detailed user's manual which describes more concise notations with new features

and applications. Though the author is guided by a different manual, he stays on the same author-language system, which is prepared to recognize the instructions from authors at any level.

The extremes of sophistication which the author-language should accept are not difficult to determine. At the naive extreme, acceptable language specifications can be determined empirically as soon as the system is put into use. At the other extreme, it is possible to allow the author to prepare program segments in the same computer programming language that was used to prepare the author-language itself. Whichever language it is (FORTRAN, JOVIAL, etc.) it will provide access to all of the features of the computer. Between the two extremes should be a sort of continuum whereby an author can grow in sophistication as he develops the need. Properly graduated user manuals will facilitate this.

Finally, (12) the author-language should incorporate an instructional management system. Reference has already been made to automatic record storing features. These records provide a history of the students' performance throughout the course. The records should be kept independently from the lesson. Any given lesson should remain under the control of the author at all times. That is to say, when a student begins the course, he does not retain his personal copy of the lesson as he might if the course was in booklet form. In any particular course,

all students will interact with a common lesson. Only his records are maintained independently. If, after school hours, the author modifies his lesson, all students would benefit from the change beginning at the time they resume, thus saving all reproduction costs.

In resuming, the student would identify himself and the course he is taking, causing the system to attach his records to the lesson again, continuing him exactly where he left off.

The student records would then comprise bases of information which could be used individually for assessing achievement and grading, or collectively for refining the course. Appropriate instructions would be included for obtaining such things as means, standard deviations, correlations, and item analyses. By effecting a transfer of student record data into the calculation mode, the course author could manipulate the data at will.

The STAT course made rudimentary attempts to provide some of these latter features, and the results were encouraging enough to show the advantages of such a system. Student records were printed at the end of each session, but not retained in the computer. Many of the features were available through the expenditure of more time and effort than is envisioned in a system incorporating instructional management. Provisions were included for making simple modifications to the problems in the STAT program. The author-language, on the other hand, would

go far beyond this.

Author-languages Under Development

Some work is currently being done in author-languages. The International Business Machine (IBM) Corporation has developed an author-language, called COURSE-WRITER II (IBM, 1966), that incorporates several of the features discussed in this chapter, and is easier to learn than conventional computer programming. It makes relatively few provisions for graduated levels of sophistication in users. It has a very limited computational capability. Dynamic answer checking in the form discussed here is nearly non-existent. The lessons produced by the author must be "compiled" before they are ready for student use. Each line of the lesson is prefixed by a mnemonic code that identifies the function of the line. It has a very adequate set of instructions for identifying a student's reply if it is spelled correctly. To identify incorrectly-spelled replies, a percentage match method is used. Too often this method incorrectly identifies irrelevant answers. With respect to instructional management, the record-keeping capabilities are good; the system keeps track of the student from session to session, placing him back into the proper point in the lesson. The summary statistics that are produced at the end of the course relieve human instructors of much non-essential clerical work. Unfortunately, the course author cannot utilize the records as a part of the decision

structure within his lesson. Rather, "counters" are provided that the author can assign to certain events and then query at a later point. Therefore, each decision must be properly anticipated. In general, COURSEWRITER represents a great improvement for an author who wishes to use a computer system but does not have the necessary programming skills.

Another author-language, PLANIT (Feingold and Frye, 1966) is under development at SDC. The PLANIT language is an outgrowth of the STAT project and incorporates most of the above features that were suggested. Eventually, all of these features will be included. To date, all of the twelve specifications have been implemented except that part of the instructional management system which analyzes the data resulting from student interaction. Since the STAT program provided a model for the development of PLANIT, the entire STAT course can now be written in the PLANIT language much more efficiently than was true in the original programming effort. But without the previous existence of the STAT program and the experience gained in the project, PLANIT would not exist in its present form.

By contrast, using COURSEWRITER would not facilitate the preparation of the STAT course. In fact, it is very doubtful if it could be done at all and retain its present characteristics. Primarily, the features suggested in the fourth, seventh, and tenth specification statements would

have to be added to COURSEWRITER to enable it to handle STAT.

Another CAI language that is being developed is called LYRIC (Silvern and Silvern, 1966). There are many similarities between LYRIC and COURSEWRITER II, both in their features and in the way lessons are prepared.

It is of some interest to note some comparisons between CAI languages, particularly between PLANIT and COURSEWRITER II because COURSEWRITER was available at the time STAT was being prepared, and found to be unsuitable for the kind of lesson preparation that was planned for STAT. Now, in view of the fact that PLANIT is an outgrowth of STAT, one might wonder what additional features had to be included in PLANIT to permit the preparation of STAT-like lessons.

At the outset, one notices that the COURSEWRITER II system and the associated IBM-1500 computer-based instructional system that uses it, uses a much richer student terminal than PLANIT. COURSEWRITER II can manipulate a printer, microfilm viewer, prestored audio message playback unit, and a cathode-ray tube (CRT). It can receive student responses from either a keyboard or light-projecting pen (touched to the face of the CRT). Suppes and Atkinson's CAI project (Bowen, 1967) in the Brentwood School at Palo Alto, California is using this CAI system with elementary school children. PLANIT, in its present stages of development restricts output messages to a

teletype printer and accepts replies only from a keyboard. It is very desirable to have a rich terminal and the PLANIT project includes plans to enrich the terminal capabilities before the system is made available for general use. However, this comes near the end of the project. The problems associated with assigning messages or displays to particular output devices are small compared to the problem of composing a suitable message or display. Similarly, the problem of accepting replies from a variety of input devices is also small compared to that of analyzing and evaluating that message after it has been received. For this reason, work on the PLANIT system has been concentrated on the internal processing problems, and for that the teletypewriter is a suitable interim device.

Internally, the PLANIT system can best be described as a merging of COURSEWRITER II and BASIC into one language. PLANIT has a CALC feature that is sufficiently sophisticated to do the kinds of mathematical work that APL or BASIC do, together with some of the prestored statistical functions of BASIC. More will be added as needed. Whereas BASIC follows language conventions that are similar to FORTRAN and APL uses Iverson notation, PLANIT's CALC follows the mathematical conventions normally found in textbooks, as closely as possible. As an example of the three methods, consider the operation denoted by the expression:

$$Y = \sum x_i, \quad (i = 1, 2, \dots, n)$$

In PLANIT, the corresponding line would look like this:

```
Y = SUM X(I)   FOR(I = 1, 2, N)
```

In BASIC, it could be accomplished with these lines:

```
      LET Y = 0
FOR I = 1 TO N
      LET Y = Y + X(I)
NEXT I
```

Finally, in APL, this short expression would do it:

```
Y ← +/X
```

APL is generally more concise in its notation than CALC, however, the Iverson notation is considerably different from the rules followed by most authors. As a simple example of this, consider the arithmetic expression:

```
(3)(4) + 5
```

The answer in APL is 27 (though a multiplication symbol must appear between the parenthesized terms), while in CALC, the expected answer of 17 is produced (and the multiplication symbol is optional).

However, the addition of BASIC or APL to COURSEWRITER II will not create another PLANIT. The PLANIT language is interrelated with CALC in such a way that anticipated answers for evaluation purposes can be written into the lesson in the form of CALC expressions. These produce the desired criteria for judging the student's answers at the appropriate time. Hence, two or three frames could be constructed that would iteratively produce an

infinite number of drill problems, from simple addition to testing hypotheses about mean differences. The student may be allowed to use as much of the CALC capability as the lesson author wishes to make available to him to assist in working on the problems. Again, as in STAT, the numbers are automatically prestored in CALC and are ready to be manipulated when the student needs them. COURSEWRITER II has no comparable feature. Its calculation capability, DESCAL, is available to the student, but is independent from any lesson.

PLANIT also includes a feature that allows a lesson author to punctuate blocks of frames with distinct sets of conditions providing a context for that part of the lesson. This includes the features in CALC that are to be made available, what base of information the student may query to obtain help, and what mathematical or statistical problem, if any, to associate with the next series of questions. CALC also provides a protected area into which a lesson author's evaluation formulas are put. While they are not available to the student, unless so specified, yet they are always operative for evaluating the students' answers. Again, COURSEWRITER II does not offer such features.

Both COURSEWRITER II and PLANIT accept and analyze constructed responses. In general, COURSEWRITER II does it by matching characters, singly or in specified groups. To attempt to recognize misspellings, COURSEWRITER II

looks for a specified percentage of the characters to be present. Any given characters can be purposely ignored (e.g. punctuation). Missing or incorrect characters can be individually brought to the attention of the student for correction.

PLANIT operates more on the concept of word and number units. Its individual character analysis capability is not nearly so well developed as COURSEWRITER. For individual character recognition, the lesson author must list the characters to be recognized in their expected order. However, this difference in concept has permitted the development of certain other features that COURSEWRITER II does not have. First, the author may cause his lesson to be responsive to misspellings through the use of a "phonetic encoder" that transforms words into a basic code according to certain rules of letter, and blend sounds so that a "match" occurs between the author's answer and the student's answer if the two "sound alike," as contrasted with a given percentage of identical characters.

PLANIT and COURSEWRITER II have similar "keyword" capabilities, based upon word units, that allows the lesson author to "find" certain key words within the student's reply. In addition, COURSEWRITER II allows one to "find" any selected sequence of characters within words and any group of key words that correspond to a specified number of a longer list of key words.

Both languages allow the restriction of the amount of time given for the student to respond.

For the remainder of the listed features, COURSE-WRITER II has no equivalent known to the author, nor does any other CAI language with which he is familiar. PLANIT, alone, has a "formulas" capability that evaluates a reply, consisting of an algebraic formula, according to the criterion that it must be "algebraically equivalent" to the formula listed by the lesson author. For example, the formula for the area of a triangle, $A = 1/2 (B) (H)$, would match $A = (B/2) (H)$ and $A = B (H/2)$, and all other equivalent rearrangements.

PLANIT allows answers to be interpreted as numbers and even recognizes numerical answers that fall within intervals if the lesson author so indicates. Suppose he asks for the value of pi, the mathematical symbol, correct to four places. He then writes his answer:

3.14159265 WITHIN .00001

So specified, he is prepared to recognize 3.1416 or anything closer to the exact answer. The answer, 3.14159, is as acceptable as the answer, 3.14160. The individual characters are no longer critically important.

PLANIT has a "multiple-choice" frame option which, if used, automatically restricts the acceptable answers to be the letter-tags of the listed answers. Any other response will be automatically rejected. Another frame option allows dichotomus answers (e.g. TRUE/FALSE) to be

listed together with an algebraic formula. The student will be required to respond with one of the two words and will be judged right or wrong according to the current evaluation of the formula. This frame is very valuable in situations where decisions must be based on numerical data.

PLANIT also allows the lesson sequence to be altered by student performance, but this is done without the use of counters. Instead, a "decision" frame allows the lesson author to "query" the students' records and "branch" accordingly. For example, he might include as part of his lesson,

IF ALL RIGHT 4-14 B: TOPIC2

where the student will go immediately to the frame bearing the label "TOPIC2" if the stated conditions for frames 4 to 14 were met. Otherwise, he will continue to the next line of the lesson. The possible variations that can compose a "decision statement" are very large. The lesson author can cause the branch decision to be based on any specified number of frames or groups of frames that were "right," "wrong" or "seen." He can similarly use for his branching criteria, time, specific answers, use of CALC functions, and particular response patterns. In each case, he can designate the frames for which his query should apply, and even whether or not the repetition of those frames should be considered. Though the possibilities are many, the language is relatively

simple. Suppose the lesson author wants to know if the student took too much time to answer frames 7 through 10 and 12 through 14, he might write his statement,

IF GR 10 MINUTES 7-10, 12-14

Other statements are analogous.

PLANIT allows lesson authors to include branches to other lessons, making it possible to "share" instructional sequences. The lessons are not "merged" together. Rather, each author maintains his own lesson. The lessons only "communicate" with each other in that case.

PLANIT also will compose feedback messages for the lesson author if so instructed. The messages will be short, reinforcing, and appropriate (e.g. YES, RIGHT, CORRECT, NO, WRONG, etc.). They will be selected randomly from either a "positive" or a "negative" list in accord with the evaluation of the student's reply.

Finally, PLANIT allows a lesson author to change its own set of primitive instructions. For example, in COURSEWRITER II, "qu" means question and "ca" means correct answer and are not subject to change. In PLANIT, KEYWORD is the name that controls the keyword matching process. The lesson author may type, "CHANGE KEYWORD TO KW" after which he uses his own term, KW, in place of the original. Similarly, "CHANGE" could be changed to "RENAME." This applies throughout PLANIT.

Both COURSEWRITER II and PLANIT and, in fact, most other CAI languages, include automatic features for

lesson building, editing, error monitoring, protection of course material (so that the student can only see what he is supposed to see), and methods of preserving lessons on various computer storage media. Lessons that are constructed under COURSEWRITER II must be "compiled" before they are ready for student use or author inspection. The length of time taken for this process depends on the schedule of activities on the computer. It can take a few minutes, but if other people are using the same system, it may take a day or more. PLANIT avoids this delay by operating "interpretively" on the material just as it comes from the author. Hence, frames can be "tried out" as soon as they are completed, and necessary changes can be made immediately. The cost of getting this added convenience is that of having a somewhat larger program. Ideally, both methods would be available--interpretive during lesson-building and a shorter compiled version for student use.

The plans for PLANIT include optional input and output devices, improved "keyword" capabilities, and a compiler for finished lessons. The author has not attempted to show how to write a lesson in either system. The references include users' guides that provide that information.

Significant developments are taking place in the field of computer-assisted instruction. These developments occur because a need exists for providing a means

whereby a larger percentage of the educational community can benefit from the advantages that a computer system has to offer. PLANIT is such a development and the STAT project was the occasion of that development.

Bibliography

- Beberman, M. Emerging Program of Secondary School Mathematics. Cambridge: Harvard Univer. Press, 1958.
- Bitzer, D. L., Braunfeld, P., and Lightenberger, W. W. "PLATO II: A Multiple-Student, Computer-Controlled Automatic Teaching Device." In J. E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: John Wiley and Sons, 1962. pp. 205-16
- Bitzer, D. L., Lyman, Elizabeth R., and Easley, J. A., Jr. "The Uses of PLATO, a Computer-Controlled Teaching System." Audiovisual Instruction, 1966, 11, 16-21.
- Bowen, E. "The Computer as a Tutor." Life, 1967, 62 (4), 68-81.
- Bruner, J. S. "The Act of Discovery." In J. P. DeCecco (Ed.), Human Learning in the School. New York: Holt, Rinehart and Winston, 1964a. pp. 256-70.
- Bruner, J. S. "Some Theorems on Instruction Illustrated with Reference to Mathematics." In E. R. Hilgard (Ed.), Sixty-third NSSE Yearbook. Chicago: University of Chicago Press, 1964b. pp. 306-35.
- Bushnell, D. R. "The Role of the Computer in Future Instructional Systems." AV Communication Review, 1963, 2, No. 2.
- Cooley, W., Lohnes, R. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley and Sons, 1962. 211 pp.
- Coulson, J. E. "A Computer-Based Laboratory for Research and Development in Education." In J. E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: John Wiley and Sons, 1962. pp. 191-204.
- Coulson, J. E. Present Status and Future Prospects of Computer-Based Instruction. SP Series 1629. Santa Monica, California: System Development Corporation, 1964. 15 pp. (Offset)

- Coulson, J. E. and Silberman, H. F. "Effects of Three Variables in a Teaching Machine." Journal of Educational Psychology. 1960, 51, 135-43.
- DeCecco, J. P. (Ed.) Human Learning in the School. New York: Holt, Rinehart and Winston, 1964. 636 pp.
- Dick, W. "The Development and Current Status of Computer-Based Instruction." American Educational Research Journal. 1965, 2, 41-54.
- Dorn, W. S. "Computers in the High School." Datamation. 1967, 10 (2), 34-38.
- Falkoff, A. D., and Iverson, K. E. The APL Terminal System: Instructions for Operation. Yorktown Heights, New York: International Business Machine Corporation, 1966. 37 pp. (mimeo)
- Feingold, S. L., and Frye, C. H. User's Guide to PLANIT: Programmed Learning for Interactive Teaching. Technical Memorandum, TM-3055/000/01. Santa Monica, California: System Development Corporation, 1966. 153 pp. (Offset)
- Feurzeig, W. "Conversational Teaching Machine." Datamation. 1964, 10, 38-42.
- Finlay, G. "Secondary School Physics: The Physical Science Study Committee." Am. J. Physics. 1960, 28, 286-93.
- Getzels, J. W. "Creative Thinking, Problem Solving, and Instruction." In Ernest R. Hilgard (Ed.), Sixty-third NSSE Yearbook. Chicago: University of Chicago Press, 1964. pp. 240-67.
- Glaser, R. "Implications of Training Research for Education." In Ernest R. Hilgard (Ed.), Sixty-third NSSE Yearbook. Chicago: University of Chicago Press, 1964. pp. 153-81.
- Green, B. F. Digital Computers in Research. New York: McGraw-Hill, 1963. 333 pp.
- Grubbs, R. E., and Selfridge, Lenore D. "Computer Tutoring in Statistics." Computers and Automation. 1964, 13, 20-26.

- Hansen, D. N. Applications of Computers to Research on Instruction. Stanford, Calif.: Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966. 17 pp. (Mimeo)
- Harman, H. H. Modern Factor Analysis. Chicago: University of Chicago Press, 1960. 469 pp.
- Hays, W. L. Statistics For Psychologists. New York: Holt, Rinehart and Winston, 1963. 717 pp.
- Hickey, A. E., and Newton, J. M. Computer-Assisted Instruction: A Survey of the Literature. Newburyport, Massachusetts: Entelek Inc., 1966. 31 pp.
- International Business Machines Corporation. 1500 Operating System Computer-Assisted Instruction Coursewriter II. White Plains, N. Y., 1966. 45 pp.
- Kemeny, J. G., and Kurtz, T. E. BASIC. Hanover, N. M.: Computation Center, Dartmouth College, 1966. 60 pp.
- Kersh, B. Y. "Directed Discovery Vs. Programmed Instruction." Title VII Project Number 907, Final Report. Monmouth, Oregon: Oregon State System of Higher Education, 1964. 77 pp. (Mimeo)
- Kersh, B. Y. "The Adequacy of 'Meaning' as an Explanation for the Superiority of Learning by Independent Discovery." J. educ. Psychol. 1958, 49 (5), 82-92.
- Kopstein, F., and Shillestad, Isabel. A Survey of Auto-Instructional Devices. Project No. 1710, Task No. 171007. Ohio: Aerospace Medical Laboratory, Wright-Patterson Air Force Base, 1961. 109 pp. (Offset)
- Mager, R. F. Preparing Objectives for Programmed Instruction. San Francisco: Fearon, 1962. 62 pp.
- Maher, Ann. Computer-Based Instruction (CBI): Introduction to the IBM Project. Research Report RC-1114. White Plains, N. Y.: International Business Machines Corporation, 1964. 13 pp.
- Silberman, H. F. Self-Teaching Devices and Programmed Materials. SP Series 663. Santa Monica, Calif.: System Development Corporation, 1962. 20 pp. (Mimeo)

Silvern, Gloria S., and Silvern, L. C. "Computer-Assisted Instruction: Specification of Attributes for CAI Programs and Programmers." Proceedings of 21st National ACM Conference, 1966. pp. 57-62.

Skinner, B. F. "Teaching Machines." Science. 1958, 128 (969-77), 137-158.

Suchman, J. R. "Inquiry Training: Building Skills For Autonomous Discovery." Merrill-Palmer Quart. Behav. & Develpm. 1961, 7 (3), 147-169.

Suppes, P. "The Uses of Computers in Education." Information. San Francisco: Freeman, 1966. pp. 157-174.

Uttal, W. R. "On Conversational Interaction." In J. E. Coulson (Ed.), Programmed Learning and Computer-Based Instruction. New York: John Wiley and Sons, 1962. pp. 171-190.

Zinn, K. Computer Assistance for Instruction: A Review of Systems and Projects. CAIS Report 010. Ann Arbor: Center for Research on Learning and Teaching, University of Michigan, 1966. 63 pp. (Ditto)

Appendix A

Student's Guidebook to

STATISTICAL INFERENCE PROGRAM

STAT

Contents:

- Student's guide
- Loading instructions
- Description of options

STUDENT'S GUIDE

To begin your use of STAT a magnetic tape (the proper one) must be loaded into the computer. This is accomplished by sending instructions to the computer system via your teletypewriter (see LOADING INSTRUCTIONS, p. 2). The last instruction in the loading sequence initiates the execution of STAT. From this point on you will converse with STAT via the teletypewriter. You must be a polite conversationalist and wait for STAT to ask (also via the teletypewriter) for your response. STAT will "speak" first and will signify that it wants your response by ending its message with the equal sign, = . Always wait for this to happen before responding. Some typical queries by STAT appear below:

OPTION = Respond with any number 1-66 listed on the Student's Reference Sheet.

REQUEST = Respond with a number 21-66 on the Student's Reference Sheet.

COL = Respond with a '1' or '2' appropriately.

ANS = Respond with YES or NO.

Also, the computer will sometimes ask for tabled values. The following is printed by the computer and the value is to be supplied by the student. (Student replies will always be underlined.)

TYPE THIS T-TABLE ENTRY FROM HAYS, PAGE 674
FOR $2Q = .05$ AND $DF = 19$.
T = 2.093

As you see sometimes you will respond with a literal expression such as YES or NO or with an entry from the T-table, and sometimes (in the interests of efficiency) with a code number. The task you will be setting for STAT by using a particular code number as in the case of OPTION or REQUEST, above, is explained, beginning on page 3, DESCRIPTION of OPTIONS.

Basically, STAT will present problems for you to solve, will generate data for these problems, will help you to calculate your answers correctly, will evaluate your answers, and will guide you to sections of Hays' book when you need help--all in a dialogue between you and the computer.

This will become much clearer to you in a short time, as you work the illustrative example.

STATISTICAL INFERENCE PROGRAM-STAT

LOADING INSTRUCTIONS

To load the computer program, type everything below that is underlined. The computer will then type the non-underlined messages in turn.

LOGIN 1234 98625
\$OK LOG ON 18
LOAD STAT
\$LOAD OK
GO
\$MSG IN

Your instructor will give you your Student Number in place of 1234.

After a short time, the program will start and ask for your Student Number.

!QUIT
\$MSG IN

Use at the end of the instructional session, or to restart when the program malfunctions.

DESCRIPTION OF OPTIONS

When the computer prints OPTION = , the following procedures will be initiated and results printed in response to entering the corresponding number:

1. Sample data for problem; random data is generated in columns, for at most two columns.
2. Sum of each data column
3. Sum of squares for each data column
4. Sum of cross products
5. Sum of difference scores
6. Sum of squared difference scores
7. Sum of n^{th} power of data column
Data column number (1 or 2) will be requested
Value of n (exponent) will be requested; may be either an integer or decimal number.
8. Student response and summary--to be used when your answers are ready to be evaluated.
9. Answer to most recently-missed question. Use as a last resort. Each use will be recorded.
10. Help. Help may be obtained on any of the procedures listed below, from 21 on. The computer will respond with "REQUEST =." Enter the number (21 or up) corresponding to the procedure on which you want help. Help will be given in the form of a page, section, and formula reference in Hays. There is no penalty for the use of this option.
11. Steps required for solution.
This option prints the appropriate page, section, and formula reference in Hays for the procedure that the student should be currently using. Each use will yield the next step or procedure in the solution until the requirements of the solution have been satisfied.
12. Calculation Assistance Program

This option evaluates algebraic expressions which are input on a single line. The following arith-

metic operations are available in this option where m and n are integers, decimal numbers, or variables as listed below:

m+n	add
m-n	subtract
m*n or m n	multiply (space between m and n can replace asterisk)
m/n	divide
m**n	m to the n th power
Rn	square root of n--or R(m+...*n)
Fn	n factorial--or F(m+...*n)
mCn	$\frac{m!}{n! (m-n)!}$
() []	parentheses and brackets used in usual algebraic fashion to remove ambiguity.

These variables may be used in place of any m or n

Si=sum of data column i, where i=1,2
 SSi=sum of square of data column i, where i=1,2
 SCP=sum of cross products
 Lj=value previously computed for Lj (in LINE j of previous calculation)

An additional equal sign, =, may be typed by the student (in addition to that printed by the computer) to indicate where evaluation is to begin. This allows comments or labels to be inserted for reference only. Type END to exit from Calculation Assistance Program to another option.

Example: Calculation of product moment correlation coefficient.

OPTION = 12

CALCULATION ASSISTANCE PROGRAM

LINE 1.

ST = MEAN 1 = S1/20 (note extra equal sign
 L1 = 12.0000 after label, MEAN 1)

ST = MEAN 2 = S2/20
 L2 = 15.0000

ST = VAR 1 = SS1/20-L1**2 (note use of L1 for
 L3 = 6.0000 a value computed earlier, above)

ST = VAR 2 = SS2/20-L2**2
 L4 = 24.0000

ST = R = (SCP/20-L1*L2)/R(L3*L4)
 L5 = .8333

ST = END

OPTION = (Student then inserts the next
 option desired)

13. Binomial tables. Computer will request the following:

N = number of trials (or observations)
 P = probability of success on a trial
 ACCUMULATE PROBABILITIES FROM
 S = number of successes (smaller no.)
 TO
 S = number of successes (larger no.)

BINOMIAL ACC. PR. = value of $\sum_S P^S (1-P)^{N-S}$

14. Hypergeometric tables. Computer will request the following:

N = no. of elements in population
 R = total no. of specials in population
 S = sample size
 ACCUMULATE PROBABILITIES FROM
 D = number of specials in sample (smaller number)

TO
 D = number of specials in sample (larger number)

HYPERG. ACC. PR. = value of:

$$\sum_D \binom{R}{D} \binom{N-R}{S-D} / \binom{N}{S}$$

15. Mann-Whitney tables. Computer will request the following:

ENTER MIN. RANK SUM

S = minimum rank sum

ENTER THE NUMBER OF RANKS IN THE MIN. RANK SUM

N1 =

ENTER THE TOTAL NUMBER OF RANKS

N =

LEVEL OF SIGN = (Evaluates exact probability of
 attaining a value of S less than
 or equal to S value entered
 above)

NOTE: If the normal approximation is used in this
 calculation, e.g., for large samples,

APPROXIMATE is printed above LEVEL.

16. Wilcoxon paired observation table. The computer will request the information analogous to 15 above, and will compute the significance probability of the Wilcoxon test for two matched samples.
20. Termination of lesson. Printout of records. Computer will also print a code number which should be written down and saved to continue the lesson.

The following options are classed as Library procedures. These may be requested at any OPTION =

DCR will mean that the computer will request the data column (1 or 2) to which the procedure will be applied unless only one column exists. The specific request follows:

WHICH DATA COLUMN:
COL =

All confidence levels will either be the same as the one stated by the current problem or implied by the significance level stated in the current problem. The confidence level may be altered by simply using a different alpha level when reading values from tables in Hays. (confidence level = $1 - \alpha$ level)

21. Arithmetic mean. DCR
22. Biased variance and standard deviation (maximum likelihood estimate). DCR
23. Unbiased variance and standard deviation ("corrected" estimated variance). DCR
24. Standard Error of the mean. DCR
25. Pooled Standard Error for the difference of sample means.
26. Mean of difference scores.
27. Biased variance of difference scores.
28. Unbiased variance of difference scores.
31. Confidence limits for mean. DCR. Table entry may be requested.
32. Confidence limits for difference between means. Table entry may be requested.

33. Confidence limits for difference between means (correlated case). Table entry may be requested.
34. Criteria for testing that ρ (CORRELATION COEFFICIENT) equals zero. Table entry may be requested.
35. T-statistic for testing hypothesis about a mean. Computer will print "ENTER MEAN" and then request a value, "MEAN = " for the population mean under null hypothesis. DCR
36. T-statistic for test of difference between means (uncorrelated case). Hypothesized difference is requested.
37. T-statistic for test of difference between means (correlated case). Hypothesized difference is requested.
41. Significance probability for sign test on a single sample for testing hypothesis about a mean. Hypothesized value of population mean is requested. DCR. Non-parametric.
42. Significance probability for sample data, Mann-Whitney two sample test. Non-parametric. (See option 15)
43. Significance probability for sample data, Wilcoxon paired observations test. Non-parametric. (See option 16)
44. Spearman Rank Correlation Coefficient. Non-parametric.
45. Fisher Exact Test. Non-parametric
51. Chi-square test for goodness of fit, test statistic. DCR.
52. Confidence limits for Variance. DCR. A table entry may be requested.
53. Chi-square test statistic for a contingency table.
54. Contingency coefficient.
55. Pearson Product Moment Correlation Coefficient.
56. Criteria for testing that ρ (CORRELATION COEFFICIENT) equals zero. (Same as 34)

- 57. Test for correlation coefficient, T-statistic.
- 58. Phi coefficient (Point biserial correlation coefficient)
- 59. Spearman Rank Correlation coefficient. (same as 44).
- 60. T-statistic for rank correlation test (Approximation)

For the following, column one will be denoted X variables and column two, Y variables:

- 61. X predicted by Y: Slope, intercept and standard error of prediction
- 62. Y predicted by X: Slope, intercept and standard error of prediction
- 63. X predicted by Y: Standard error of beta and T-value for beta
- 64. Y predicted by X: Standard error of beta and T-value for beta
- 65. X predicted by Y: Prediction and confidence interval for the prediction. A predictor will be requested. A table entry may be requested.
- 66. Y predicted by X: Prediction and confidence interval for the prediction. A predictor will be requested. A table entry may be requested.

TEXTBOOK SECTIONS COVERED BY STAT

To obtain the best results from the program, the student should first have read the following sections in Hays, Statistics for Psychologists:

Chapter	Sections
5	5.9, 5.10, 5.11, 5.12, 5.13, 5.14, 5.16, 5.17, 5.18, 5.19
6	6.5, 6.6, 6.9, 6.13, 6.16, 6.17, 6.18, 6.21, 6.22, 6.23, 6.26
7	7.1, 7.2, 7.4, 7.5, 7.6, 7.10, 7.11, 7.12, 7.14, 7.15, 7.16, 7.19, 7.20, 7.21, 7.23,
8	8.1, 8.2, 8.3, 8.13
9	9.1, 9.2, 9.3, 9.4, 9.11, 9.18, 9.19, 9.20, 9.21, 9.22, 9.25, 9.30
10	All except 10.20, 10.21
11	11.3, 11.6, 11.7, 11.8
15	Intro., 15.1, 15.16, and 15.26
17	Intro., 17.1, 17.2, 17.3, 17.4, 17.5, 17.6, 17.7, 17.9, 17.10, 17.11, 17.12, 17.15
18	Intro., 18.4, 18.7, 18.8, 18.11, 18.12

You may want to correct a printing error in Hays in order to avoid confusion.

On page 320, the line above the equation for est. σ^2 now reads, "From 7.15.5 it follows that." The reference should be 7.16.5.

Appendix B

Student's Reference Sheet for STAT OPTIONS

OPTION =

CALCULATION

1. SAMPLE data
2. SUMS
3. Sums of SQUARES
4. Sum of CROSS PRODUCTS
5. Sum of DIFFERENCE SCORES
6. Sum of SQUARED DIFFERENCE SCORES
7. Sum of N^{th} POWERS

YOUR ANSWERS

8. Student RESPONSE and SUMMARY

HELP

9. ANSWER to most recently MISSED QUESTION (Use as a last resort)
10. HELP, OPTIONS 21 to 66
11. STEPS in SOLUTION

CALCULATION ASSISTANCE

12. CALCULATION ASSISTANCE program
Legal symbols: Sj SSj SCP SMD SSD Rn Fn
mCn Lj

PROBABILITY DISTRIBUTIONS

13. BINOMIAL
14. HYPERGEOMETRIC
15. MANN-WHITNEY

16. WILCOXON paired observation

LESSON TERMINATION

20. LESSON TERMINATION, records and restart code

ESTIMATES

21. MEAN.....	Page 161
22. BIASED VARIANCE and standard deviation...	177-8
23. UNBIASED VARIANCE, standard deviation....	207
24. STANDARD ERROR of the MEAN.....	202
25. POOLED STAND. ERROR, for differ. of sample means.....	320
26. MEAN of DIFFERENCE scores.....	335
27. BIASED VARIANCE of DIFFERENCE SCORES.....	335
28. UNBIASED VARIANCE of DIFFERENCE SCORES...	335

SAMPLE STATISTICS BASED ON THE T-DISTRIBUTION

31. CONFIDENCE LIMITS for MEAN.....	312
32. CONFIDENCE LIMITS for DIFFER. of MEANS...	321
33. CONF. LIM. for DIFF. of MEANS, CORRELATED case.....	334-5
34. Criteria for testing that RHO (CORRELATION COEFFICIENT) equals zero....	533
35. T-statistic for HYPOTH. TEST of MEAN.....	311
36. T-statistic for HYPOTH. TEST OF DIFFER. of MEANS.....	317
37. T-statistic, TEST of DIFF. of MEANS, CORRELATED case.....	334-5

NON-PARAMETRIC

42. MANN-WHITNEY TWO GROUP SIGNIF. probability.....	633
43. WILCOXON PAIRED OBSERVATION SIGNIF. probability.....	635

- 44. SPEARMAN RANK CORRELATION coefficient.... 643
- 45. FISHER EXACT TEST..... 598

STATISTICS BASED ON CHI-SQUARE DISTRIBUTION

- 51. GOODNESS-OF-FIT test statistic..... 586
- 52. CONFIDENCE LIMITS for VARIANCE..... 345
- 53. CHI-SQUARE test of INDEPENDENCE..... 589
- 54. CONTINGENCY COEFFICIENT..... 606

DEPENDENCE MEASURES

- 55. PRODUCT MOMENT CORRELATION coefficient... 505
- 56. Criteria for testing that ρ
(CORRELATION COEFFICIENT) equals zero.... 533
- 57. T-statistic for CORRELATION TEST..... 529
- 58. PHI-COEFFICIENT..... 604
- 59. SPEARMAN RANK CORRELATION..... 643
- 60. T-statistic for RANK CORRELATION test
(approx.)..... 646

REGRESSION

- 61. X on Y Slope, INTERCEPT and S.E. of
prediction..... 504-5
- 62. Y on X Slope, INTERCEPT and S.E. of
prediction..... 504-5
- 63. X on Y S.E. of BETA and T-value for BETA. 521
- 64. Y on X S.E. of BETA and T-value for BETA. 521
- 65. X on Y Prediction and CONFIDENCE INTERVAL 504-5,522
- 66. Y on X Prediction and CONFIDENCE INTERVAL 504-5,522

NOTE: Always wait for = ? before typing a message.

Strike CARRIAGE RETURN to enter typed messages into the computer.

To cancel a complete message, strike " (upper case 2) before striking CARRIAGE RETURN.

Appendix C

Name _____

Student Number _____

CRITERION TEST

Note: On all multiple choice, cross out the incorrect answers. Each successfully eliminated wrong answer will count points.

The first six questions are related to the following situation:

Mrs. X has developed two forms of an intelligence test and is proceeding to check them for equivalents. Both forms are administered to a random sample of 18 students and two I.Q. scores are obtained for each student. He chose a .05 significance level for his statistical tests.

A T value of 1.4 was computed to test whether the product-moment correlation coefficient was significantly greater than zero.

1. What is the critical value for concluding that the correlation coefficient is significantly different from zero?

- a) 1.684 b) 1.740 c) 1.746 d) 2.021
f) 2.120

2. Was it significant? a) Yes b) No

3. Approximately what value correlation coefficient must have been obtained?

- a) greater than .50 b) between .10 and .50
c) between -.10 and .10 d) between -.50 and
-.10 e) less than -.50

Chi-square goodness of fit tests dividing the data from each form into seven intervals produced computed values of 2.80 and 1.60 respectively.

4. What critical value did he find from the chi-square table?

- a) .710721 b) 7.26094 c) 9.48773
d) 24.9958

Finally, the response differences observed from the two groups were tested.

5. If the use of a T test was warranted, what critical T value would he find in the T table?

- a) 1.697 b) 1.740 c) 2.042 d) 2.110
e) 2.120

6. He found that the differences were not statistically significant. What conclusions can he make?

- a) The forms are enough alike to be used interchangeably.
b) The forms are producing radically different estimates of I.Q.
c) The two forms are not reliably measuring the same things.

Suppose you were interested in the relationship between military rank and the number of years spent in military service.

7. What statistical procedure would you use for this comparison?

You hypothesize that one can obtain a specific rank in less time in the Army than in the Navy. You contemplate setting up your design in one of two different ways. What statistical test would be appropriate for each of the following designs?

8. Randomly select one subject for each of the first ten ranks in each of the two branches of the service. Observe the length of time each one has spent by the two groups with what test?

9. Randomly select twenty subjects from the lowest non-commissioned officer rank in each branch (Army and Navy). Observe the length of time each one has spent in the service. Test the differences in time spent by the two groups with what test?

You suspect that your class of 30 people has a very unusual spread (variance) of I.Q. scores. Each one happens to have taken the same I.Q. test that has a mean of 100 and standard deviation of 15. The computation of an appropriate test statistic turns out:

$$(29) (116) / 225 = 14.9511$$

10. What is the critical value from the table that tells you whether to reject at the .05 level of significance?

11. Would you reject for the above value, 14.9511?

12. If you had proposed a directional hypothesis, would you have suspected that the I.Q. spread in your class was more or less than normal? (answer more or less)

You are given a column of X scores and a column of Y scores and are told that X is the independent and Y the dependent variable so using X to predict Y you find the slope (beta) and intercept (alpha) of the regression line that best fits the data. The computed values for the slope and intercept are:

$$\text{slope} = 5.0$$

$$\text{intercept} = 4.0$$

13. From an X score of 3 what do you calculate the predicted Y score estimate to be?

14. If a 95 per cent confidence interval for the true predicted Y score extends from 11.72 to 22.28, what would the predicted Y estimate be?

Each of three sections of a statistics course is taught by a different professor. You want to test whether the grades received by the students (A, B, C, or D are the grades) are independent of the professors who taught the sections.

15. Describe or diagram how the data would be arranged to be tested.

16. What significance test would you use?

17. What critical value would determine whether or not your hypothesis was supported at the .05 level of significance?

Appendix D

PRETEST

To enter a formula or mathematical expression into a computer via typewriter, the formula is typed in a line occupying a single space. Also certain conventions must be observed using the symbols on the typewriter, e.g.,

Division: $\frac{3}{5}$ becomes 3/5

Powers: 3^2 becomes 3**2 or $(3+5)^2$ becomes (3+5)**2

Square Root: $\sqrt{3+4}$ becomes $\sqrt{(3+4)}$

Usually additional parentheses (brackets, etc.) are required to exactly specify a multiplication or division, e.g.,

$\frac{3+4}{5-2}$ becomes (3+4)/(5-2)

$\frac{3}{5} \sqrt{6}$ becomes (3/5) $\sqrt{(6)}$ but

$\frac{3}{5 \sqrt{6}}$ becomes 3/(5 $\sqrt{(6)}$)

The first three problems are to be answered by writing the expressions that will give the correct answer without doing any calculations yourself, e.g.,

Add 3 to 4 and divide the answer by 5.

Answer: (3+4)/5. The answer 3+4/5 would be incorrect.

Use the following X and Y columns of scores where data are required.

i	X _i	Y _i
1	22	21
2	26	9
3	19	13

Obeying the conventions discussed above:

1. Write the numerical expression that computes the average of the X column of scores (add the scores and divide by the number of addends).
2. For the above X and Y columns of scores, write the equivalent numerical expression for:

a. $\sum_{i=1}^3 Y_i$

b. $\sum_{i=1}^3 X_i Y_i$

c. $\sum_{i=1}^3 Y_i^2$

d. $\frac{\sum_{i=1}^3 Y_i^2}{3} - \left[\frac{\sum_{i=1}^3 Y_i}{3} \right]^2$

3. Using the formula: $\bar{X} - \bar{Y} - (\mu_x - \mu_y)$

$$\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

where $\sigma = 4$, $\mu_x = 20$, $\mu_y = 15$, the numerical equation without parentheses is as follows:

$$22 + 26 + 19 / 3 - 21 + 9 + 13 / 3 - 20 - 15 / 4 \quad 1 / 3 + 1 / 3$$

Put in the required parentheses.

The following questions require the use of tables from Hays, page 674 ff or some other source.

4. Suppose you are going to do a two-sided t test and you have 20 degrees of freedom. a) What would be the critical t value for rejecting the null hypothesis at the 5 percent level of significance on the basis of the test?
b) If your obtained t value was 2.04, would you conclude that the difference was significant at the given level?
5. Suppose instead, the test was one-sided with 20 degrees of freedom. a) What would the critical t value be for the .05 significance level?
b) Would the t value 2.04 indicate significance for the one-sided test?
6. If the significance level were made smaller, would the computed t value have to be larger or smaller to show significance?
7. What would be the effect on the critical t value if the degrees of freedom are increased?
8. In general, t values that are significant for a two-sided test (will, need not) be significant for a one-sided test.
9. Referring to a chi-square table such as Hays, pages 675-76, suppose a chi-square test for independence produced a computed chi-square value of 16.54 with 10 degrees of freedom. Is it significant at the .05 level?

Appendix E

INSTRUCTOR'S GUIDEBOOK TO STAT

INSTRUCTOR'S GUIDE - Underlining will be the instructor's typed replies.

All actions may be taken at any OPTION = ?

- 1) To alter problems: (A sample problem has been altered.)

OPTION = ? 555

NUM = 5

Sequence position (0 to 29) of problem to be altered.

PGM(5,0) = 2

Type number of desired problem.

PGM(5,1) = 90

Range for column one. (6 S.D.).

PGM(5,2) = 100

Mean for column one.

PGM(5,3) = 20

Shift constant between means (random if zero).

PGM(5,4) = 120

Range for column two (equals range above if zero).

PGM(5,5) = 2

Repetitions of the problem.

PGM(5,6) = 0.01

Correlation (random if zero).

PGM(5,7) = 30

Maximum number of subjects per group.

PGM(5,8) = 1

Library inhibit after 1 of the 2 repetitions.

PGM(5,9) = .05

Significance level (alpha).

OPTION = ?

Resume the program again.

- 2) Allowing student selection of problems:

OPTION = ? 777

Student will be asked to supply his desired problem number with the message SELECT = .

OPTION = ? 666

Restores the fixed sequencing or problems.

- 3) Regulating problem presentations and LIBRARY access;

OPTION = ? 888

Allows the instructor to uniformly specify for all problems the number of trials per problem and access to the LIBRARY options (See PGM(5,5) and PGM(5,8) in 1) above). The computer will request the following:

TRIALS= ? 2 Each problem will be presented twice.
ACCESS= ? 1 LIBRARY options may be used on each
first presentation.

4) Use of program for real data:

OPTION= ? 999 External data fill. Information
and data will be requested.

5) Final record printout will list the following:

PROBLEM - problem count on this session

TYPE - problem identification number

ERRORS - corrected errors (if problem was finished)

UNSUCCESSFUL - uncorrected errors (number of uses of
Option 9)

STEPS - number of steps (Option 11) requested

TIME - time spent working problem and answering
questions

BACK - number of times BACK was used (An entry of
1000 indicates the use of STUCK.)

PROBLEM-TYPE LIST

Type No.

0. REQUEST A SAMPLE AND CALCULATE THE FOLLOWING ESTIMATES MEAN, VARIANCE, STANDARD DEVIATION AND CONFIDENCE INTERVAL FOR THE MEAN.
1. YOU HAVE I.Q. SCORES FROM A COMMON INSTRUMENT (MEAN, 100; S.D. 15) FOR A GROUP OF SUBJECTS AND YOU WANT TO TEST WHETHER THIS GROUP SHOULD BE REGARDED AS A RANDOM SAMPLE FROM THE POPULATION. MAKE SEPARATE TESTS FOR THE MEAN AND STANDARD DEVIATION.
2. AN EXPERIMENTAL GROUP OF SUBJECTS HAS RECEIVED SPECIAL INSTRUCTION WHICH WAS NOT GIVEN TO THE CONTROL GROUP. BOTH GROUPS ARE MADE UP OF SUBJECTS WHICH HAVE BEEN DRAWN AT RANDOM FROM THE SAME POPULATION. PREPARE A CONCLUSION ABOUT THE EFFECTIVENESS OF THE INSTRUCTION.
3. A PROBLEM-SOLVING EXPERIMENT USED ACHIEVEMENT AND LATENCY SCORES AS INDEPENDENT MEASURES. YOU QUESTION THE VALIDITY OF THIS AND COLLECT SIMILAR DATA. THE TWO ABOVE SCORES ARE OBTAINED FOR EACH SUBJECT.
4. YOU HAVE TRAINED A GROUP OF SUBJECTS USING A NEW TECHNIQUE TO IMPROVE READING SPEED. YOUR DATA CONSISTS OF READING RATES BEFORE AND AFTER THE TRAINING PERIOD. TEST THE EFFECTIVENESS OF YOUR METHOD.
5. YOU HAVE DEVELOPED TWO FORMS (A AND B) OF A SCALE WHICH MEASURES INTELLECTUAL CURIOSITY. BOTH FORMS WILL BE GIVEN TO A GROUP OF SUBJECTS. YOU NEED A MEASURE OF RELIABILITY BETWEEN A AND B AND ALSO YOU WANT TO KNOW WHETHER THE MEANS OBTAINED USING EACH FORM ARE SIGNIFICANTLY DIFFERENT.
6. TWO SECTIONS OF A GIVEN CLASS MEET DURING THE SAME HOURS AND TAKE THE SAME OBJECTIVE FINAL EXAM. OBTAIN THE SCORES FROM EACH CLASS AND TEST THE DIFFERENCES IN LEVEL OF ACHIEVEMENT OF THE TWO CLASSES.
7. IN A FIELD TEST OF A NEW SERUM A LARGE SAMPLE WAS DIVIDED INTO TWO TREATMENTS. TREATMENT 1 RECEIVED THE SERUM, TREATMENT 2 DID NOT. THE DATA RECORDS THE NUMBER OF CASES IN EACH CATEGORY FOR THE ENTIRE SAMPLE.
8. ON A 28-ITEM PRETEST, TEST THE HYPOTHESIS THAT EDUCATION GRADUATE STUDENTS GET MORE THAN ONE-HALF OF THE ITEMS CORRECT AS INDICATED BY THE SCORES.

9. TWO SIMILARLY NURTURED GROUPS OF RATS WERE TAUGHT TO RUN A MAZE TO REACH A FOOD BOX AND THE NUMBER OF TRIALS REQUIRED FOR A PERFECT RUN WERE RECORDED. ONE GROUP WAS GIVEN THE BENEFIT OF WHITE ARROWS INDICATING THE CORRECT ALLEY AT EACH CHOICE POINT. TEST THE EFFECTIVENESS OF THE ARROWS.

10. IN YOUR CLASS OF 30 PEOPLE, YOU BECOME INTERESTED IN THE NUMBER WHO WEAR GLASSES. A '1' INDICATES THAT THEY DO WEAR GLASSES, A '0' THAT THEY DO NOT. TEST WHETHER ONE SEX IS MORE LIKELY TO WEAR GLASSES THAN THE OTHER.

11. TWO GROUPS OF STUDENTS TOOK A 28-ITEM PRETEST. THE FIRST GROUP HAD THE LISTED PREREQUISITES, THE SECOND GROUP DID NOT. TEST THE IMPORTANCE OF THE PREREQUISITES AS SHOWN BY THE PRETEST.

12. YOU HAVE DEVELOPED AND ADMINISTERED A 20-ITEM QUESTIONNAIRE ON RACE RELATIONS TO 200 SUBJECTS, 100 EACH FROM TWO RACES. YOUR RESPONSES ARE CATEGORIZED EITHER POSITIVE OR NEGATIVE. YOUR DATA CONTAINS THE NUMBER OF POSITIVE ANSWERS FOR EACH GROUP ON EACH QUESTION. TEST WHETHER THE GROUP RESPONSES DIFFERED.

13. YOU HAVE CANVASSED 15 LIVING UNITS ON CAMPUS WITH A SINGLE QUESTION TO BE ANSWERED EITHER YES OR NO. YOUR DATA CONSISTS OF THE TOTAL NUMBER OF RESPONSES FOR EACH UNIT. TEST WHETHER THE UNITS AGREE.

14. YOUR EXPERIMENT REQUIRES TWO SCORES FOR EACH SUBJECT X (FIRST COLUMN) AND Y (COLUMN TWO). ONE SUBJECT MOVED BEFORE THE SECOND MEASURE COULD BE OBTAINED. FROM THE FIRST SCORE, 22, YOU MUST PREDICT THE MISSING SCORE. CALCULATE THE MAXIMUM ERROR YOU WOULD EXPECT IN YOUR PREDICTION.

15. FOR PREDICTING COLUMN TWO (Y) FROM COLUMN ONE (X) CALCULATE THE FOLLOWING ESTIMATES: SLOPE (BETA), INTERCEPT, CORRELATION COEFFICIENT AND STANDARD ERROR OF ESTIMATE.

16. CENTRAL JUNIOR AND HIGH STUDENTS COME FROM EIGHT ELEMENTARY SCHOOLS IN THE DISTRICT. SCHOOL NO. 6 IS IN A LOW SOCIO-ECONOMIC AREA. FOR THOSE WHO ENTERED JUNIOR HIGH EIGHT YEARS AGO, ALL HAVE EITHER GRADUATED OR DROPPED OUT. YOUR DATA CONTAINS DROP-OUT TOTALS FOR EACH ELEMENTARY SCHOOL ATTENDED. TEST WHETHER SOCIO-ECONOMIC LEVEL IS A MAJOR FACTOR IN HIGH SCHOOL DROP-OUT CASES.

17. YOU ARE EVALUATING A NEW MATH PROGRAM USED IN ONE FIFTH-GRADE CLASSROOM. IN CHOOSING YOUR CONTROL GROUP FROM ANOTHER FIFTH-GRADE SECTION, YOU MAY EITHER CHOOSE AT RANDOM OR HOMOGENEOUSLY MATCH.

18. IN A CERTAIN PSYCHOLOGICAL TEST, A GROUP OF SUBJECTS ARE ASKED TO STRAIGHTEN A PICTURE IN A TILTED ROOM. EACH SUBJECT DOES THIS TWICE: FIRST, WHEN ONLY THE PICTURE CAN BE SEEN AND THEN WHEN THE ENTIRE ROOM IS ALSO VISIBLE. TEST THE HYPOTHESIS THAT THE MEANS ARE SIGNIFICANTLY DIFFERENT.

19. DO COLLEGE GRADUATES EARN MORE ANNUALLY THAN NON-COLLEGE GRADUATES. TEST THE MEAN DIFFERENCE.

20. TEST THE RELATIONSHIP BETWEEN THE NUMBER OF YEARS SPENT IN HIGHER EDUCATION AND ANNUAL SALARY.

21. ON THE BASIS OF MID-TERM AND FINAL EXAM SCORES, TEST FOR SIGNIFICANT ACHIEVEMENT GAIN BETWEEN EXAMS.

22. TEST THE CORRELATION COEFFICIENT FOR SIGNIFICANCE WHEN NORMALITY ASSUMPTIONS ARE NOT MADE FOR THE FOLLOWING DATA. (THE TEST WILL BE AN APPROXIMATION.)

23. USE THE FIRST COLUMN SCORES TO PREDICT THE SECOND COLUMN. CHOOSE AN ADDITIONAL COLUMN ONE SCORE AND CALCULATE THE CORRESPONDING PREDICTED SCORE.

24. IT IS SAID THAT ENGINEERING STUDENTS HAVE HIGHER NON-VERBAL I.Q.'S THAN DO STUDENTS IN THE BEHAVIORAL SCIENCES. GROUP ONE IS DRAWN FROM ENGINEERS, GROUP TWO FROM THE BEHAVIORAL SCIENCES. TEST THE ABOVE STATEMENT.

Appendix F

ILLUSTRATIVE PROBLEM

!LOGIN 1234 98625

In place of 1234 type in your assigned student
I.D. No.
Type in only underlined information.

\$OK LOG ON 24
LOAD STAT

Requesting load of statistics program

\$LOAD 24

Indicates successful load of program

GO

This command starts the lesson.

\$MSG IN.

YOU ARE BEGINNING A COMPUTERIZED TRAINING COURSE IN
STATISTICAL APPLICATIONS.

TO BEGIN, TYPE IN YOUR STUDENT NUMBER.

NUM = ? 1234

Note: From this point on, all responses are
made only after an equal sign is typed out, as
in: NUM = , or ANS = .

ARE YOU RESUMING A PREVIOUS LESSON: YES/NO.

ANS = ? NO

PROBLEM 1.00

REQUEST A SAMPLE AND CALCULATE THE FOLLOWING ES-
TIMATES: MEAN, VARIANCE, STANDARD DEVIATION AND CONF-
IDENCE INTERVAL FOR THE MEAN.

USE THE .95 LEVEL OF CONFIDENCE.

ENTER OPTION NUMBER.

OPTION = ? 1

ENTER DESIRED SAMPLE SIZE--MAXIMUM
OF 35 PER GROUP.

NUM = ? 10

THE SAMPLE VALUES FOLLOW:

SUBJECT --SCORES

1-	30
2-	50
3-	40
4-	30
5-	50
6-	50
7-	59
8-	69
9-	40
10-	40

ENTER OPTION NUMBER.

OPTION = ? 2

See Student's Reference Sheet for Option number.

COL 1 SUM = 458.0000

ENTER OPTION NUMBER.

OPTION = ? 3

COL 1 SUM SQ. = 22342.0000

ENTER OPTION NUMBER.

OPTION = ? 12

CALCULATION ASSISTANCE PROGRAM.

Essentially, Option 12 allows one to enter a line of arithmetic to be evaluated. The result will be printed on the following line opposite an identification.

LINE 1 .

ST = ? MEAN=30+50+"

MEAN=(30+50+40+30+50+50+59+69+40+40)/10

See NOTE: at bottom of Students Reference Sheet on Cancellation of a message.

L1 = 45.8000

ST = ? MEAN=S1/10

L2 = 45.8000

ST = ? S1/10

Note advantage obtained by using operator S1 for

sum of column one.

L3 = 45.8000

ST = ? UNBIASSED EST. OF POPN. VAR.=(22342-10*45.8**2)/9

Notice anything bracketed by two equal signs will be regarded as a comment and not part of the expression.

L4 = 151.7333

ST = ? (SS1-10*L3**2)/9
PARENTHESES DO NOT MATCH.

ST = ? (SS1-10*L3**2)/9
L5 = 151.7333

ST = ? EST. OF STANDARD DEVIATION OF POPN.=R151.7333
L6 = 12.3180

ST = ? R(L5)
L7 = 12.3180

ST = ? RL5
L8 = 12.3180

Three examples of obtaining Standard Deviation:
Line 6 with comment
Line 7 without comment
Line 8 same as line 7 but without parentheses.

ST = ? END

The "END" message permits us to exit from this option and enter another.

Let us now proceed to use some of the Library Options. See the Reference Sheet on your table; Options numbered 21 and over belong to the Library and will only be available to you the first time you go through a problem. Let us begin with Option 21.

ENTER OPTION NUMBER.

OPTION = ? 21

MEAN = 45.8000

You notice this is the same value you computed with Option 12, Calculation Assistance Program above.

ENTER OPTION NUMBER.

OPTION = ? 23

VARIANCE = 151.7333
STD. DEV.= 12.3180

This is an easier way to calculate the variance, but we repeat, Options with numbers greater than 21 will not be available to you the second time you are asked to do a problem.

Suppose we have forgotten how to calculate confidence limits for the population mean. We can of course turn to the text or we can use Option 10 for help. We got the number 31 from our Student Reference Sheet which also has a page reference to the text by Hays--however, the computer will furnish more precise references to the text.

ENTER OPTION NUMBER.

OPTION = ? 10

REQUEST HELP BY TYPING OPTION NUMBER.

REQUEST = 31

MEAN LIMITS: P311, S10.10, F10.10.1

If we are completely at a loss as to how to begin to solve a problem, we can use Option 11. Each use of this option will provide a step in the problem solution and a reference to the text for this step.

ENTER OPTION NUMBER.

OPTION = ? 11

SOLVE FOR EACH QUANTITY SPECIFIED BY THE STEPS.

STEP 1

MEAN: P161, S6.5, F6.5.1

ENTER OPTION NUMBER.

OPTION = ? 11

STEP 2

UNBIASED VARIANCE, STD. DEV: P206, S7.14, F7.14.4

ENTER OPTION NUMBER.

OPTION = ? 11

STEP 3

MEAN LIMITS: P311, S10.10, F10.10.1

ENTER OPTION NUMBER.

OPTION = ? 11

STEP 4
ALL STEPS COMPLETED.

Let us proceed once again to Option 12 - Calculation Assistance Program to compute the confidence interval limits. Using the T table at the back of Hays we find 2.262 is the value we require for 9 degrees of freedom and $2Q = .05$. Notice the line no.'s continue to accumulate as they will until we finish a problem. We can still use previous evaluated L's in our calculations as well as the square root function R, the sum of column arrays, S1, etc.

ENTER OPTION NUMBER.
OPTION = ? 12

CALCULATION ASSISTANCE PROGRAM
LINE 9.

ST = ? LWR. LIM. = L2-2.262*(L6/R10)
L9 = 36.9888

ST = ? UPPR. LIM. = S1/10+2.262*R(L5/10)
L10 = 54.6112

ST = ? 10**52
L11 = 1.0E+52

Very large numbers are printed in scientific notation i.e. 1.0×10^{52}

ST = ? 10**L6
L12 = 2079730896000.0000

ST = ? END

Alternatively if you are unsure of your calculating ability use Option 31

ENTER OPTION NUMBER.
OPTION = ? 31

SPECIFY WHETHER THE LIMITS ARE FOR A ONE-SIDED OR TWO-SIDED TEST. ENTER 1 OR 2.

NUM = ? 2

TYPE THE T-TABLE ENTRY FROM HAYS, PAGE 674, FOR $2Q = 0.050$ AND $DF = 9$.

T = ? 2.262

UPPER BOUND = 54.6112
LOWER BOUND = 36.9888

ENTER OPTION NUMBER.

OPTION = ? 8

ENTER THE FOLLOWING VALUES FROM YOUR RESULTS:

MEAN = ? 45.8

VARIANCE = ? 153.000

INCORRECT VARIANCE.

CORRECT BEFORE PROCEEDING.

UNBIASED VARIANCE, STD. DEV: P206, S7.14, F7.14.4

We are now illustrating how you will enter your problem solutions. Notice this was not the value we calculated above for the unbiased estimate of variance, and the computer tells us it is incorrect.

As a last resort, if we cannot get the computer to accept our answer, we use Option 9 which will give us what the computer believes is a correct answer. Do not use this option unless you are really stumped.

ENTER OPTION NUMBER.

OPTION = ? 9

-ANSWER- **** 151.7333

We now resubmit our answers. (Enter all correct answers) The confidence interval we calculated above clearly shows that the population mean of 50 is within limits at the .95 level of confidence.

ENTER OPTION NUMBER.

OPTION = ? 8

ENTER THE FOLLOWING VALUES FROM YOUR RESULTS:

MEAN = ? 45.8

VARIANCE = ? 151.7333

THE POP. MEAN = 50.0000

IS THIS MEAN IN THE 95% CONFIDENCE INTERVAL: YES/NO

ANS = ? YES

NOW YOU ARE CORRECT.

THE PARAMETERS OF THE POPULATIONS FROM WHICH YOU SAMPLED ARE LISTED BELOW:

GROUP 1 MEAN	--	VARIANCE	--	STD. DEV.
50.0000		277.7778		16.6667

PROBLEM 2.00

REQUEST A SAMPLE AND CALCULATE THE FOLLOWING ESTIMATES: MEAN, VARIANCE, STANDARD DEVIATION AND CONFIDENCE INTERVAL FOR THE MEAN.

USE THE .95 LEVEL OF CONFIDENCE.

ENTER OPTION NUMBER.

OPTION = ?

You are now on your own. Feel free to experiment with this second attempt at Problem 00. Notice the Problem numbering system; 2.00 indicates the 2nd pass through problem 00. Problem 00 does not count, so have fun.

Appendix G

SURVEY QUESTIONNAIRE

Name _____

Student No. _____

Answer the following questions either yes or no. The same three questions are being asked about each of the 30 statistical procedures listed below. Circle Y for yes or N for no.

Question 1. Have you become acquainted with this statistical procedure and its application (either in class or through independent study)?

Question 2. Have you ever carried out the calculations for a statistical problem using this procedure?

Question 3. Would you have answered either of the first two questions differently before you were contacted for this experiment?

<u>Q.1</u>	<u>Q.2</u>	<u>Q.3</u>	
Y N	Y N	Y N	1. Arithmetic mean
Y N	Y N	Y N	2. Population variance or standard deviation
Y N	Y N	Y N	3. Unbiased variance or standard deviation estimate
Y N	Y N	Y N	4. Estimated standard error of the mean
Y N	Y N	Y N	5. Pooled estimate of the standard error of the mean
Y N	Y N	Y N	6. Mean of differences between paired scores
Y N	Y N	Y N	7. Unbiased estimate of variance of paired score differences
Y N	Y N	Y N	8. Confidence limits (critical values) based on a single group mean

Y N	Y N	Y N	9.	Confidence limits (critical values) based on the difference between two means.
Y N	Y N	Y N	10.	Confidence limits (critical values) based on the mean of the differences between paired observations.
Y N	Y N	Y N	11.	T-test for a single group mean against a specified population mean
Y N	Y N	Y N	12.	T-test for the difference between means
Y N	Y N	Y N	13.	T-test for the mean difference of paired observations
Y N	Y N	Y N	14.	Sign test
Y N	Y N	Y N	15.	Mann-Whitney two sample test
Y N	Y N	Y N	16.	Wilcoxon test on paired observations
Y N	Y N	Y N	17.	Spearman rank correlation coefficient
Y N	Y N	Y N	18.	Fisher Exact (Fisher-Irwin) test
Y N	Y N	Y N	19.	Chi-square test of goodness of fit
Y N	Y N	Y N	20.	Confidence limits for a single group variance (based on chi-square)
Y N	Y N	Y N	21.	Chi-square test of independence (association)
Y N	Y N	Y N	22.	Contingency coefficient
Y N	Y N	Y N	23.	Pearson product moment correlation coefficient
Y N	Y N	Y N	24.	Confidence limits for correlation coefficient
Y N	Y N	Y N	25.	T-test on correlation coefficient
Y N	Y N	Y N	26.	Phi coefficient
Y N	Y N	Y N	27.	Regression formula (slope and intercept)

- | | | | | |
|-----|-----|-----|-----|---|
| Y N | Y N | Y N | 28. | Standard error of prediction in regression |
| Y N | Y N | Y N | 29. | Standard error of the slope (beta) coefficient |
| Y N | Y N | Y N | 30. | Confidence interval for the true score based on the predicted score |

Appendix H

ATTITUDE QUESTIONNAIRE

Name _____

Student Number _____

I. In your statistics class you probably had a lecture, text, and exercises to do. I want you to compare that method of learning how to use statistics with what you have been doing on the computer. Divide 100 points between the computer method and your former method according to which you think has the most over-all merit (a 50-50 division would mean undecided).

computer _____ former _____

II. Indicate the extent to which each of the following statements describe your opinion of the program

Very Essentially Slightly Not
True True True True

--	--	--	--

1. I liked the summary information given at the end of each problem.

--	--	--	--

2. I liked doing wrong things on purpose to see how it would answer me.

--	--	--	--

3. There were times when I didn't know what I was expected to do.

--	--	--	--

4. I liked the idea that each problem was a little different from what others were working.

--	--	--	--

5. The use of numbers to request procedures was confusing.

--	--	--	--

6. I liked the ability to use alternate methods for solving problems.

--	--	--	--

7. This method soon becomes boring.

Very Essentially Slightly Not
True True True True

--	--	--	--

8. I learn something by computing statistical procedures by hand that is missing here.

--	--	--	--

9. I liked having the computer check my work rather than being checked against an answer book.

--	--	--	--

10. I liked being able to get necessary help without going to a person.

--	--	--	--

11. I liked giving more attention to the problem being tested and less to the computations.

--	--	--	--

12. I liked the problems that required the least calculations.

--	--	--	--

13. Problems of similar types should have been grouped together.

--	--	--	--

14. I expected more explanation from the computer and less reliance on the book.

--	--	--	--

15. The process for obtaining help was too confusing.

--	--	--	--

16. The exercises were not as meaningful as coordination with the text would have made them.

--	--	--	--

17. I liked the idea of operating a computer.

--	--	--	--

18. I liked working on something new.

--	--	--	--

19. I just followed formulas and had too little an idea what I was doing.

Very Essentially Slightly Not
True True True True

--	--	--	--

20. I didn't see any purpose in having to calculate any of the procedures.

--	--	--	--

21. I liked the ease of making computations.

--	--	--	--

22. I needed personal encouragement.

--	--	--	--

23. I liked the ability to try different things with the data.

--	--	--	--

24. The help was inadequate.

--	--	--	--

25. I liked the opportunity to exercise some choice in setting up the problems.

--	--	--	--

26. I liked the fact that help was given only by request.

--	--	--	--

27. The problems didn't go far enough in the understanding that they required.

--	--	--	--

28. I would rather have selected which problems I wanted to work.

III. GENERAL REACTIONS TO THE USE OF THE COMPUTER

1. What use have you made of the computer before?

2. Did you have misgivings about using a computer before you began here? (Please elaborate.)

3. In what ways might this experience have helped to change your attitude toward computational uses of the computer?

4. Did this experience using the computer influence your thinking concerning future statistical work you might want to do? (Please elaborate.)

5. Will you likely be using statistics in your career?

6. To what extent will your work be likely to involve a computer? Directly? Indirectly? Not at all?
What uses:

Please note here any other suggestions regarding the instruction you have just received.

Please accept these words of appreciation for your part in the evaluation of this instructional procedure. The information which you have supplied will be extremely valuable in determining its worth.

Charles H. Frye, Michigan State University
& System Development Corporation

Appendix I

Table of Pretest Scores

Student Number	UCLA (Computer Group)		MSU (Comparison Group)	
	Topic 1.	Topic 2.	Topic 1.	Topic 2.
1	1	1	1	1
2	1	1	1	0
3	0	0	1	1
4	0	0	1	0
5	0	0	1	0
6	0	0	0	0
7	1	0	0	1
8	1	0	0	0
9	1	0	1	1
10 ^a	0	1	0	1
11	-	-	1	0
12	-	-	1	0

^aThe UCLA student number 10 passed the pretest but was unable to be a subject.

Note:

Pretest results are for those admitted to the project. Topic 1 refers to the first section of the pretest (items 1 to 3). Topic 2 refers to the remaining items. The entries are the number of errors (maximum of one) for that topic.

Appendix J

Table of Criterion Test Scores

Student Number	UCLA (Computer Group)	MSU (Comparison Group)
1	104.7	53.4
2	51.4	53.4
3	67.4	41.4
4	47.7	45.4
5	105.2	66.4
6	83.7	52.7
7	70.7	65.4
8	79.7	50.4
9	91.7	39.2
10	-	42.4
11	-	42.4
12	-	26.9
Mean	78.0222	48.250
Std. Dev.	19.6273	10.6641

Note:

Criterion test scores are reported as the number of points earned out of one hundred twenty-two possible.

Appendix K

Sample Teletype Output from STAT

OPTION = ? 999

ENTER THE N VALUES:

N1 = ? 12

N2 = ? 9

FIRST GROUP:

DATA = ? 53.4

DATA = ? 53.4

DATA = ? 41.4

DATA = ? 45.0

DATA = ? 66.4

DATA = ? 52.7

DATA = ? 65.4

DATA = ? 50.4

DATA = ? 39.2

DATA = ? 42.4

DATA = ? 42.4

DATA = ? 26.9

SECOND GROUP:

DATA = ? 104.7

DATA = ? 51.4

DATA = ? 67.4

DATA = ? 47.7

DATA = ? 105.2

DATA = ? 83.7

DATA = ? 70.7

DATA = ? 79.7

DATA = ? 91.7

SPECIFY A SIGNIFICANCE LEVEL FOR CONSTRUCTION OF CONFIDENCE LIMITS, ETC.

ALPHA = ? .05

USE OPTION 1 TO CHANGE ALPHA.

OPTION = ? 21

WHICH DATA COLUMN:

COL = ? 1

MEAN = .

48.2500

OPTION = ? 21

WHICH DATA COLUMN:

COL = ? 2

MEAN = 78.0222

OPTION = ? 22

WHICH DATA COLUMN:

COL = ? 1

BIASED

VARIANCE = 113.7225

STD. DEV. = 10.6641

OPTION = ? 22

WHICH DATA COLUMN:

COL = ? 2

BIASED

VARIANCE = 385.2328

STD. DEV. = 19.6273

OPTION = ? 23

WHICH DATA COLUMN:

COL = ? 1

UNBIASED

VARIANCE = 124.0609

STD. DEV. = 11.1383

OPTION = ? 23

WHICH DATA COLUMN:

COL = ? 2

UNBIASED

VARIANCE = 433.3869

STD. DEV. = 20.8179

OPTION = ? 24

WHICH DATA COLUMN:

COL = ? 1

S.E. OF MEAN = 3.2153

OPTION = ? 24

WHICH DATA COLUMN:

COL = ? 2

S.E. OF MEAN = 6.9393

OPTION = ? 25

POOLED SE OF MD = 7.0319

OPTION = ? 36

ENTER HYPOTHEZIZED POPULATION MEAN DIFFERENCE.

NUM = ? 0

T-VALUE = -4.2339

OPTION = ? 42

MANN-WHITNEY TWO-SAMPLE PROBABILITIES

MIN. RANK SUM = 89.0000

LEVEL OF SIGN. = 0.0007

OPTION = ? 12

CALCULATION ASSISTANCE PROGRAM

LINE 1.

ST = ? F-RATIO = 433.3869/124.0609

L1 = 3.4933

ST = ? 3.2153**2

L2 = 10.3382

ST = ? 6.9393**2

L3 = 48.1539

ST = ? 7.0319**2

L4 = 49.4476

ST = ? CORRECTED D.F. = (L4**2/(L2**2/13 + L3**2/10))-2

L5 = 8.1835

ST = ? END

Note:

This is a copy of a portion of the interaction with the STAT program being used to analyze the criterion test scores. Underlining was added to indicate that which the author typed.

Appendix L

Table of Survey Questionnaire Results

Student Number	Level One	Level Two	Level Three
1	28	0	2
2	15	4	11
3	8	12	10
4	21	3	6
5 ^a	12	1	2
6	20	3	7
7	16	7	7
8	14	7	9
9	19	4	7
Percentages	60.0	16.1	23.9

^aCategorized only fifteen of the thirty named procedures.

Note:

Survey questionnaire results reporting the number of statistical procedures which each subject placed into the three level categories. Level one signifies experience with the named procedures, level two signifies only familiarity without computational experience, and level three signifies a lack of familiarity.

Appendix M

Data Used For Correlation Coefficients

Subject	Criterion Test	Guideline 1	Guideline 2	Guideline 3	Guideline 4	Guideline 5	Guideline 6	Guideline 7	Guideline 9	Guideline 11a	Guideline 11b	Guideline 12a	Guideline 12b	Guideline 13a	Guideline 13b	Total Problems Completed	Problem-Types Completed
1	104.7	88.9	21.7	11.8	0.0	5.9	5.9	6.7	0.0	44.4	33.3	50.0	50.0	11.1	11.1	9	9
2	51.4	66.7	35.0	0.6	33.3	11.1	11.1	13.3	0.0	22.2	11.1	0.0	0.0	0.0	66.7	5	3
3	67.4	70.6	34.5	4.9	16.7	2.4	22.0	53.3	8.3	46.2	15.4	50.0	25.0	23.1	15.4	20	17
4	47.7	59.1	32.0	6.7	18.2	11.7	20.0	40.0	0.0	40.0	20.0	66.7	33.3	33.3	6.7	27	21
5	105.2	66.7	44.3	6.1	23.1	6.1	16.3	26.7	0.0	35.7	42.9	0.0	0.0	21.4	0.0	20	15
6	83.7	100.0	33.9	10.6	8.3	2.1	12.8	20.0	4.2	27.3	45.5	0.0	25.0	27.3	0.0	22	15
7	70.0	91.7	14.3	0.0	0.0	0.0	9.1	0.0	0.0	81.8	9.1	75.0	0.0	9.1	0.0	12	12
8	79.7	82.4	22.5	12.0	8.6	0.0	10.9	86.7	2.9	58.3	25.0	28.6	42.9	16.7	0.0	32	17
9	91.7	33.3	22.6	8.7	9.1	0.0	21.7	13.3	0.0	45.5	36.4	33.3	33.3	0.0	18.2	10	9
R _{xyi}		.102	-.026	.556	-.436	-.439	-.216	-.169	-.120	.038	.763	-.267	.233	-.132	-.447	-.099	-.110

Notes:

All data came from the UCLA computer group and are explained in Chapter VI (Also see Tables 5 and 6 and Appendix J). The criterion test is the dependent variable.