1972

Development of a student guided program in statistics using a small digital computer

Leland Richard Tack
Iowa State University

INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

A Xerox Education Company

72-20,004

TACK, Leland Richard, 1945-
DEVELOPMENT OF A STUDENT GUIDED PROGRAM IN STATISTICS USING A SMALL DIGITAL COMPUTER.

Iowa State University, Ph.D., 1972
Education, scientific

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Development of a student guided program in statistics using a small digital computer

by

Leland Richard Tack

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Education

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Area

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1972

TABLE OF CONTENTS

I. INTRODUCTION

Computer-assisted instruction has been said to (2) "be comparable to Gutenberg's invention of the printing press in terms of the potential effect it will have upon education." Stemming from a small idea in 1958, computer-assisted instruction has grown to a multimillion dollar investment. This growth has been by no means the result of the computer industry alone, for as we enter the decade of the seventies the demand for more and continuing education is increasing.

John Goodlad (21, p. 10) in his book on Computers and Information Systems in Education writes "American education has been guided by the following ideal: education is for all Americans." Indeed the United States has long professed equal educational opportunity for all its people. Yet much of the "free" public education has not been equal to the individuals partaking in the process if for no other than economic considerations the education must be mass education. And when the educational process loses the individual as its goal and concentrates on groups (classes, large lectures), the "fast" student may be slowed, the "slow" individual may be lost and only a few remain as being treated as individuals.

In the introduction to Computers in the Classroom, Margolin (32) writes: "There are many causes of the current crises in education. New demands and new resources are converging on this normally conservative area at a pace never before encountered. The pressures on education include:

1. The demand for education by an ever-widening population, reaching almost from the cradle to the grave.

2. Conflicting demands for greater specialization and greater general adaptability to the tremendously increased rate of change in the economy, coupled with the need for rapid retraining and more general education.

3. The need for more comprehensive education, making the individual more flexible and allowing him to participate more fully in our society.

4. The pressure to keep pace with the rapid social changes and the increases in mobility and fluidity of society.

5. The pressure for increased efficiency, centralization, and knowledge."

Pressures such as these are not and will not be easy to alleviate. As Stolurow (44) points out:

> Computer assisted instruction (CAI) is not the panacea for today's educational problems. . . . there is no single solution to problems such as these.

Twenty years ago the term computer was scarcely in use and our daily lives were by no means dependent upon its use. Yet today it is nearly impossible to proceed through a day where a computer has not had some role in the day's activities. Education also is capitalizing on the contribution a computer can make in its environment.

Although there are obvious business and record keeping functions which can be performed, the greatest potential for the computer in education is in the area of instruction. Presently proceeding under CAI (computer-assisted instruction or computer-aided instruction), CBI (computer-based instruction), CMI (computer-managed instruction), or a guise of other names, they all have the related objective of performing some role in the student's learning process.

Although we have already proceeded through three generations of computers at least in terms of physical attributes of the machine, computers in education are still in their infancy stage in terms of uses. In order to proceed toward the potential which the computer most likely has for education a crucial step must be taken, namely development. This study was entirely developmental in nature and explored the feasibility of using a computer for instruction by allowing the student to guide himself through the lessons. Two further considerations were made. One expressed by Rogers (37) is that the mode of instruction should be compatible with the subject matter, the subject matter in this study being statistics. The other consideration was that of using a relatively small computer but one which had a Cathode Ray Tube (CRT) and a Teletype to interact with the student.

Previous studies on using a computer in teaching statistics have been approached from using the computer as a high speed calculator-problem solver. Many hours of time consuming hand calculations could be eliminated by a use such as this. One example of this is the work done by Cooley (13) where students moved through a series of simulated computer experiments "designed to familiarize him . . . with: random number generation . . . sampling distribution of the correlation coefficient." Both batch processing of student problems and tutorial-type interaction at a terminal were used. An attempt was made to use the terminal to question the student on his problem but it was "concluded that this is too inefficient of computer time and terminal time."

In this study the computer is to be used for instruction and demonstration, not calculation. The calculating capabilities are secondary and used for instruction purposes.

A. Statement of the Problem

The problem of this study was to investigate the use of a small, general purpose computer, namely the PDP-12 (Programmed Data Processor-12), to present certain statistical concepts. Involved in the study was the consideration and feasibility of presenting material in an adjunct manner, in addition to the feasibility of using the system of the PDP-12's size for CAI. To be considered in the two modes of presentation, the CRT and the terminal, was the ability of the system to allow the student to guide himself through the statistical concepts selected.

Four subproblems were studied: (1) the feasibility of using a computer of this size for instructional purposes; (2) the development of curriculum material to be presented; (3) the development of the necessary software to implement the curriculum; (4) development of a way in which to best create lessons or units within a concept.

B. Purpose of the Study

The study had several major purposes. The first was the development of an approach to be used for presenting selected statistical concepts in a student controlled environment. By allowing the student to experiment in the use of a concept and by presenting it as graphically as possible it was thought an intuitive approach to statistics could be obtained. Since the approach was to be somewhat unstructured to the student, it needed to be very well structured in terms of monitoring the program.

A second purpose of this study was the development of the software to present these concepts. Several considerations were made, the most important one being the size of the computer on which this type of instruction was being planned but also a prime consideration in choosing the machine was the available storage area and the type of display device available.

The study was designed with the computer being a complement to the teacher, not a replacement. The computer would not be a stand-alone device which could administer a complete lesson. But it was felt that the instruction the computer was to perform was unique and could not be readily presented by another mode of teaching. The Cathode Ray Tube would permit rapid display of points along with the deletion and addition of various points and the calculation of the parameters involved.

Software to be developed was to include arithmetic routines, display routines and input-output routines. Also a method of building units from a building block approach was to be developed. These building blocks consisted of routines which would print, display points, accept points, and perform functions similar to these as needed. The purpose of these routines was the desire to construct teaching units as easily as possible.

The general purpose of this study was to examine the feasibility of using a computer of limited size to present selected statistical concepts through graphical means at an elementary intuitive level. At the inception of this study the IBM 360-65 did not have as a part of its time sharing facility a CRT which would adequately handle displaying of a variety of points. The PDP-12 was also chosen since to the author's knowledge such a computer had not been used for CAI on an individual basis.

This study was developmental in nature and an objective evaluation was not undertaken. The examination of the feasibility of the proposed system was the main concern of this research. The development of the system was felt to be a sufficient examination of the feasibility of the concepts.

II. REVIEW OF LITERATURE

In 1926, Sidney Pressey wrote (36, p. 374):

> The average teacher is woefully burdened by such routine of drill and information-fixing. It would seem highly desireable to lift from her shoulders as much as possible of this burden and make her free for those inspirational and thought-stimulating activities which are, presumably, the real function of the teacher.

Visualized by Pressey in his original work on a teaching machine and also recommended by him was what he called "adjunct autoinstructional programming." His use of the term adjunct was that this teaching machine was only part of the entire education process, but adjunct was not used with the denotation of being secondary. Until Pressey's work was picked up by Skinner, programmed instruction was essentially forgotten (40).

Skinner (40, p. 966) revised Pressey's work advancing the concept of programmed instruction. Stating that "the demand (for education) can not be met simply by building more schools and training more teachers" he went on to present his observations of why a teaching machine had merit. They were:

1. There is a constant interchange between program and student.
2. Like a good tutor, the machine insists that a point be thoroughly understood . . . ..
3. The machine presents just that material for which the student is ready.

4. The machine helps the student to come up with the right answers.

5. The machine reinforces the student for every correct response.

The manner in which he was conceiving programmed instruction was not to be a reality until the advent of the computer.

The first recorded instruction using a computer was the now historic project of Anderson and Rath in 1958 as reported by Atkinson and Wilson (2). At the IBM Wastone Research Center a computer was used to teach binary arithmetic and since their first simple experiment, the area of computer assisted instruction has grown many, many fold.

Computer assisted instruction is often broken down into three areas or levels. These levels as defined by Suppes (45) are:

1. Drill and practice
2. Tutored system
3. Dialogue system.

In addition Zinn (52) defines instructional uses of the computer in two more modes, those being simulation and scholarly aids.

This investigation will not attempt an exhaustive search of the literature in each of these areas. Several complete reviews have been done by Hickey (27) and McClain (31). Since the concern of this thesis lies mainly in the scholarly aids area, only the highlights of the other four areas will be mentioned. Then the remaining portion of this chapter will be divided up into reviews of scholarly aids, and a learning theory basis for scholarly aids.

A. Drill and Practice

Since this is a direct application of Skinner's theory of stimulus-reinforcement and one of the more developed areas of learning theory it also was one of the first areas to be developed for computer use. Suppes's (45) work at Stanford is probably one of the most widely known projects in the field. The original work began in 1964 and the longitudinal evaluations of his work are just beginning to be realized. The Stanford project is devoted mainly to mathematics and foreign languages, two subjects which have a fairly well set of structured rules behind them. The evaluations being conducted by Suppes are not completed but he has gathered evidence that this type of instruction is at least as effective as traditional instruction and further that additional benefits of increased independence of the student can be obtained.

In addition to Suppes' research with elementary students, research has been conducted by Uhr (50) to design a system in which drill and practice problems are generated from a set of rules. This was a very useful undertaking since the teacher interaction with the computer is high in terms of developmental time.

B. Tutorial Systems

Suppes (45) defines the tutorial systems in that the "aim is to take over from the classroom teacher the main responsibility for instruction." This is an argument many people would not agree with, feeling the computer should supplement the instructor, not supplant him.

Stolurow (43) sees the instruction in this mode as being designed by an instructional staff while the instructional "logic to be used for a particular student will be one that is generated by the system, based upon information it contains about the student and about what factors make a difference in instructional effectiveness."

Several CAI languages have been developed in this area to aid in the development of lessons. Systems Development Corporation's PLANIT (28) and IBM's COURSE WRITER are two of the more widely known languages in this area. While CAI languages exist for use in a large system, no language could be found for a system the size of the PDP-12.

The tutorial system should be designed with enough flexibility so that the slowest student does not fail yet the most intelligent student does not get bored. For the tutorial system to function as a tutor more work needs to be done in the area of identifying and prescribing instruction for individual differences.

C. Dialogue Systems

In a dialogue system a "natural language" approach to communicating with the learner is used. It is the highest level of interaction with the student and is still at the cutting edge of CAI development (47). Technical problems are the major hindrances at this time with recognition of the spoken or written word a major problem.

Projects which have been undertaken include the medical diagnosis by Feurfeig and the use of a dialogue system in the guidance and counseling area. The system used by Ellis on Information System for Vocational Decisions used sentence decomposition rules in order to understand the inquirers of the system. Although not all sentences could be decomposed correctly, communication between student and computer was possible.

Other efforts to develop procedures to handle the English language have been undertaken by Simmons et al. as referred to in Borko (6). Stolurow refers to the dialogue system as Inquiry (44), stressing the greatest asset of the system is that the student needs only ordinary communication skills to use the system. Again the ability of the computer to use English sentences as input is a problem.

D. Simulation

Simulation and gaming are yet another facet of CAI. The underlying basis of simulation is the processing of a given input through a model from which various algorithms have been programmed. Since the simulation is for instructional purposes most models are designed to increase the probability of the learner making desired decisions.

Various management games concerning the business world have been implemented to date. Included are models ranging from city management (27), to the buying and marketing of goods in various businesses. The decision making process which is to occur may be controlled by the algorithms internally, the instructor or other players in the simulation, or by both.

An area where simulation has been somewhat lacking is that of the social sciences. Although it is entirely the author's speculation, the lack of theory upon which to base models and their simulation may be one reason for this absence. A recent example of simulation in the social sciences is a dissertation by Thomas (48), using the computer to teach experimental design in the social sciences.

E. Scholarly Aids

The area of scholarly aids includes parts of all the other modes of CAI plus some unique contributions of its own. To avoid sounding as if scholarly aids is a catchall for CAI which doesn't fit any other classification some specific examples will be given. One of the ways in which a computer may be thought of as a scholarly aid is in the area of problem solving. With its considerable speed and calculating ability problems which were realistic but impossible to assign in the classroom can now be given and solved. At Iowa State University some introductory level statistics courses use a language called OMNITAB to aid in solving problems. This could be considered using the computer as a scholarly aid.

A further consideration for a scholarly aid is in the storage and retrieval of vast amounts of information. The information stored may be the primary purpose or it may be an auxiliary to the problem solving mode, where if a student needs more information to solve his problem he has it readily available. Marks (33) has developed a pilot system in which the student is in control of his learning environment. It consists of "a problem solving environment with an information file system to produce a powerful learning tool." Branstad (7) has carried this system further and developed a CAI system to teach formal computer languages.

F. Theory for Scholarly Aids

A criticism which has been directed toward the field of education and toward the preparation of "effective" teachers has been the lack of theory upon which instruction is based. As Stolurow has stated "CAI makes our meager knowledge of teaching patently obvious."

Teaching and learning are obviously not contagious terms. As Ausubel points out (3), B. O. Smith presents the idea that effective teaching may not necessarily be associated with the same model as effective learning. A problem with this type of approach is that teaching without learning is not effective teaching. A definition of teaching given by Ausubel (3, p. 11) seems quite acceptable for CAI:

> By teaching we mean primarily the deliberate guidance of learning processes along lines suggested by relevant classroom learning theory.

The procedure of CAI is the guiding of the learning process through the use of a mechanical device. Brunner (10) refers to three types of instructional devices, Type I being "devices for vicarious experience" such as films, TV, and books. Type II he defines as sequential programs which "have the function of helping the student grasp the underlying structure of a phenomenon." The essential point of Type II is the sequential aspect, assuming certain orders of presentation are better than others for a given subject area. The third type he has labelled as "automatizing devices" consisting of teaching machines to aid in learning.

The third type of device appears to be dependent upon I and II in that if it is to be essential it must also be a "vicarious experience" and the "orders of presentation of material and ideas . . . (should be) . . . more likely than others to lead the student to the main idea." How this automatizing device is to be used is an extension of the "arts of teaching" since the science of teaching is still somewhat lacking.

Brunner's (10) expressed theory then on the use of these devices can be summarized as follows:

> How these aids and devices should be used in concert as a system of aids is of course the interesting problem . . . In sum, then the teacher's task as communicator, model, and identification figure can be supported by a wise use of a variety of devices that expand experience, clarify it, and give it personal significance. There need be no conflict between the teacher and the aids to teaching.

Brunner's comments seem most appropriate for applying CAI in an educational setting. His concept of intuitive learning also seems to have merit with respect to a theory of instruction for the type of instruction proposed in this study. He defines intuition as follows:

> Intuition implies the act of grasping the meaning or significance or structure of a problem without explicit reliance on the analytic apparatus of one's craft. It is the intuitive mode that yields hypothesis quickly, that produces interesting combinations of ideas before their worth is known (9).

Ausubel has reacted to Brunner stating that possibly he has attempted to place too much faith in the intuitive approach. While Ausubel states that "by using an intuitive approach it is possible successfully to teach the elementary-school child many ideas in science and mathematics," he later expands on this. "Some abstractions are so inherently difficult or complex that they can not be made intuitively understandable to children below a certain cognitive maturity . . . Certain abstractions become relatively useless when restructured on an intuitive basis (3, p. 213)."

Perhaps the immediate use of CAI will not be as oriented as much toward an instructional device as it will be used as a research tool for testing models and theories of learning. Stolurow (43) puts forth this thought and goes on to delineate the process involved in putting a theory into practice on a computer. Although a theory may be complex and the strategies for its implementation surrounding it may be complex, if it can be well defined it can be programmed and processed.

The processes involved in formalizing a CAI program on a teaching theory as listed by Stolurow (43, p. 66) are:

1. The events must be defined.
2. The time sequence of the events must be mapped.
3. A description of the task (teaching) must be made in relationship to the subject matter.
4. A CAI system is chosen for the computer-student interaction.
5. The instruction materials are verified, in that they will run on a system.
6. To validate a model or theory, the system must be replicable.

Using Brunner's theory of instruction Grubb (24) undertook a study defined as "learner-controlled statistics." The purpose being to "direct the student through planned learning sequences and prompt him towards insightful solutions." In an attempt to recognize individual differences each student was allowed to "map" his own path through the course.

The system used was the IBM 1500 Instructional System utilizing a CRT keyboard and light pen. Observations up to this point indicate potential advantages and disadvantages. One of the advantages was each student's path through the materials is recorded, which allows the instructor some insight into what material is of most interest and to which students. It was featured that students now knowing which lessons to choose could be guided by what previous students had taken. A major disadvantage to a learner-controlled approach was in the size of computer-storage needed. The ratio of statements needed to present the same concept through a structured approach versus a learner-controlled approach was 1 to 5.

Cooley (13) found after using a computer to assist in the instruction of statistic courses in three different modes: (1) tutorial-dialogue, (2) exposition method, (3) computational exercises, that "CAI is most appropriately employed in the numerical demonstration of statistical concepts and for statistical laboratory exercise instruction."

In addition to Grubb's and Cooley's projects on using statistics, McClain (31) recently completed a study developing probability units for statistics. In his research also done at Iowa State University, IBM's Conversational Programming System (CPS) was utilized in the design of an author language and control of the CAI system. Although no structural evaluation of the project was planned some of the weaknesses and strengths of the system were evident. The use and interest of students trying the system was encouraging. One of the problems encountered in lesson preparation was the difficulty of correcting errors or changing program and/or text.

G. Summary

In relation to this study the work that has been done in learner-controlled statistics is limited. An area of concern to all those who have worked in CAI has been lack of a teaching theory on which to pattern instruction. An appropriate summary of what has been and what will be is presented by Max Rafferty in the foreword to Data Processing for Education by Grossman and Howe (23):

> No one can foresee the ultimate contribution of these amazing instruments, but most knowledgeable scientists and scholars are convinced it will be immense . . . education is not exempt.

The educator who sees the computer as an oversized calculator has missed the essence of the computer evolution.

III. DESCRIPTION OF THE CAI SYSTEM

This research project was undertaken to study the feasibility of using a computer of limited size to present selected statistical topics. Since to understand the CAI system a knowledge of the computer system is needed this chapter will describe the various components used. Three parts are discussed: (1) the hardware component, (2) the software components, and (3) the educational component.

A. The CAI Hardware Component

One of the major feasibility aspects of this study was to examine the use of a small computer for the purposes of CAI. Also a computer having both a Teletype and a Cathode Ray Tube (CRT) was needed. The computer which met these requirements was the PDP-12 (Programmed Data Processor-12).

The PDP-12 was developed by Digital Equipment Corporation and was one in a series of PDP computers. Although basically an in-laboratory computer it was defined as a general purpose computer. The PDP-12 was available in three configurations of varied capabilities. The system available at Iowa State University consists of a basic system with a core memory of 4096 (4 K) twelve bit words. It is possible to expand this system to 32,768 (32 K) bytes of core memory.

In addition to variation in core memory, peripheral devices and other options are available. These include a

CRT, a Teletype, tape transports, analog handling input capabilities, and six relays for controlling external equipment. Those devices of interest in this project were the CRT, the Teletype and tape storage available.

Core memory was divided into 4 segments of 1 K each of which only two segments are available at any one time. One segment was designated as the instruction field and the other was the data field. The instruction and the data field do not necessarily need to be contiguous nor do they need to be distinct from each other. Data fields may be changed at any point in a program and were in effect as soon as the instruction (LDF) was given. But the instruction field was changed only after a change of instruction field (LDF) and a jump (JMP) to a location with the new field was given.

The CRT, a Type VR12, has a 6.5 x 9 inch or 58.5 square inch screen on which individual points and characters can be displayed. The grid dimensions, in points available, were 512 points by 512 points for a total of 262,144 points available for display. The horizontal and vertical distances between points were not equal with horizontal distance being 0.0176 inch and the vertical distance being 0.0127 inch. Full-size and half-size alphameric characters can be displayed by using one of the two display instructions and the Special Function Register (ESP). The display instruction of most concern in this study was point display (DIS). The approximate time to display one point is 23 microseconds.

Branstad (8) states points need to be regenerated at a rate of 33 per second to remain flicker free. A conservative estimate was therefore made that approximately 1000 points could be generated if all points resulted from the routine.

The input/output device used was a Model 33 ASR Teletype. It was the primary device used for entering programs and building the CAI units. In addition to the keyboard and printer, the Teletype was equipped with a papertape reader and punch. The maximum character rate transfer for either input or output is ten characters per second while data may be transmitted as either input or output simultaneously.

Two magnetic tapes were utilized in addition to the basic 4 K of core storage. The LINC tape system consisted of two DEC tape transports TYPE TU55 on a completely buffered sub-processor. One read/write head was utilized for the 10 track tape. Three tracks were utilized for data and 2 tracks were used by the tape processor for control. The other five tracks were utilized for redundantly recording the same bits of information.

The tapes were organized in blocks, each tape containing 512 blocks and each block containing 256 twelve bit words. On each tape certain sections were reserved for the DIAL-2 system, and therefore not available for CAI use or program storage. The LAP-6, DIAL-2 system utilized 120 blocks leaving the remaining blocks available for program storage.

The PDP-12 had two programming modes available, one the

LINC mode and the other the PDP-8/I mode. Both modes have equal status and can be used interchangeably throughout a program. The LINC mode was selected and the LINC language was used, although input and output instructions needed to be in the 8-mode. The LINC mode consisted of 65 instructions under the eleven functions of add, multiply, load, store, shift, operate, logical operations, skip, I/O, memory, and LINC tape.

At Iowa State University the PDP-12 was an open shop operation where the user was in complete control of the computer. Since this meant consideration of how a student was to get on and off the machine several switches on the console needed to be set. The console itself was located within the entire computer setup and was an intrinsic part of the system. To initiate the routine the LEFT and RIGHT switches on the console needed to be set in addition to depressing the I/O PRESET, DO, and START-20 switches in a given order. To assure this would be done, directions concerning the correct procedure to be used were given and are present in Appendix A.

B. The CAI Software Component

The basic language which was used in this study was the LINC mode language. Several languages were available on the PDP-12 with the LINC, PDP-8/I, and FOCAL-4K the ones available at Iowa State University. Several factors were considered in the language used. Since this study was to examine the feasibility of using a computer of limited size for the purpose of CAI, a high level CAI language was not felt applicable, although concern was expressed over the ease at which future units or concepts might be built.

The familiarity with the language and ease with which it could be programmed were prime considerations for choosing the language. As previously mentioned 65 instructions under eleven functions made up the language. The PDP-8/I language was utilized for the input and output of characters since the LINC mode could not handle this input and output function.

It was the intent of this study to use as much of the existing software as possible. For this reason the LAP-6 monitoring system was employed. It was this system which allowed the lesson designer to enter units at the terminal, edit his programs, save programs in either source or binary mode, and a variety of other functions up to and including assembly and loading binary into core memory. The LAP-6 system also detected certain programming errors such as undefined symbols and tags, double use of tags, exceeding push down stacks and errors of this nature.

The LAP-6 was also the system used to build programs and lessons. Programs could be built using any or all of three systems. On-line programming was one of the methods used to build programs. By entering programs via the Teletype, lessons or routines could be built with the current lines being displayed on the Cathode Ray Tube. The second method utilized to build programs was paper-punch tape. Repetitive routines or portions of programs were prepunched on paper tape and then entered as needed in programs.

The third method of building programs used several features of the DIAL-2, LAP-6, system. Routines or portions of programs could be saved using the Save Program (SP) command. This saved program then could be inserted at any point in the program by having the line after which the program was added be the current line. Then the command Add Program (AP) was given along with the name under which the program had been saved.

Programs were saved by assigning one to six characters as the program name. By giving the display index (DX) command, the names of the programs currently on file were displayed on the CRT. In addition to the program name the total number of blocks utilized by that program was given. Total available file space for program storage was 384 blocks.

Although the text portion varied from unit to unit and the order of presentation varied, certain routines necessary to carry out these functions remained the same.

The building blocks were PRINT, ADPT, DELETE, DSPP, LINE, and ASK respectively. Access to these routines was accomplished by performing a JMP routine instruction. For the lesson designer one additional aspect to the building blocks was needed, that being setting the beta register. The first 15 registers of an instruction field were known as beta registers. They were used for direct and indirect addressing, and they could also be used as counters.

These registers were set by either a SET instruction or by placing values in their address through the accumulator. Using the SET instruction provides for a straight forward manner in which an address or value was stored by a lesson designer. Two examples of these routines were the print routines and the displaying of points and lines. By utilizing all three methods of lesson building a building block approach to lesson design was used.

It was intended that someone with a minimal amount of understanding of the PDP-12 could design lessons. Through the use of the building block approach and the LAP-6 system a simple yet powerful method of lesson building was developed. LAP-6-DIAL-2 is defined as "an editor, filing system and assembler for use with the PDP-12 computer (15)."

The editor has capabilities of adding and deleting individual lines and characters or deleting entire pages. It was especially useful in the correction of the text to be printed. By using the moveable cursor, spelling or spacing errors could be quickly corrected and additional text added.

As was mentioned previously, a building block approach to lesson designing was undertaken. The major routines written as capable of being added directly were print, add points, delete points, display points, display line and accepting student input routines.

The print routine, PRINT, translated chopped ASCII code into full ASCII code and printed the character. The print routine was initiated by setting register 11 to the location of the first register. The location does not necessarily need to be in the same segment as the routine, only the absolute address needs to be given. Printing continued until the symbol was encountered in ASCII code.

The function of the add point routine was to allow the user to enter data points through the Teletype. Points could be added to existing storage or an entirely new set of points could be added. The storage utilized was that part of storage not currently being used in the active segment from registers 1700 to 1777. Checks were built into the program to insure only acceptable points were entered and the maximum number of points was not exceeded.

The delete point routine functions approximately the same as the add point routine. When a data pair was entered to be deleted, the current storage area was searched in an attempt to find the point. If it was not found, the user was told the point could not be found. If the point was found, the data pair located in the last storage location was moved into the deleted position and the point counter was reduced by one.

The display routine displayed the points currently located in the locations 1700 to 1777. The number of points to be displayed was set through the length register. The coordinates of a point were both placed in a single 12 bit register. The leftmost six bits contained the X coordinate and the rightmost six contained the Y coordinate.

Although the choice was somewhat arbitrary the points needed to be integer values and range from 1-15. The distance between the points was also arbitrary to an extent with the distance display points being 40 CRT micro-points. The distance between these micro-points is 0.0176 inch horizontally and 0.0127 inch vertically, the distances reflecting the 9 inch by 6.5 inch screen. The approximate distances between display points was then .7 inch by .5 inch.

In addition to displaying points the ordinate and abscissa were also displayed with the approximate numeric hash marks. This routine was later abandoned when the routine was found to be too time consuming for display generation.

One additional display routine was utilized in this study, that being a line routine. The line was generated one point at a time by using the slope and the intercept. Points were generated from left to right by starting with the intercept and adding a cummulative slope to it. Both positive and negative slopes could be generated. The intercept had to be less than 31 and greater than -15, with the starting point on the axis being corrected accordingly. The slope and the intercept were entered through the two beta registers 15 and 16.

Accepting student input was accomplished through a very short input routine. When input was to be accepted a JMP

IN command was given. This caused the program to jump to the input routine. When a key was struck the input was placed in the accumulator and the routine returned to the instruction immediately following the jump instruction.

Several additional routines were utilized from one unit to another. One of the routines was a transfer routine, TRANS, used to move points from the data segment or an unused segment to the instruction segment. Three control routines were used to control the display of points and lines. ROUT1 controlled the display points only, ROUT2 controlled the display of points and lines simultaneously, and ROUT3 controlled the display of two lines and a set of points. All three routines' primary function was to control the length of the time a display was held on the screen. A display would generally remain until the user struck the key.

The third routine differs from the other two in performing an additional function of register exchange for line generation. Since at times it was desirable to display two lines without duplication of the line generation, routine ROUT 3 was used. The routine basically exchanged the contents of the two registers which the line display needed. As will be mentioned later another aspect of this routine was the differential rate at which the two lines were generated. The logic behind this was a way of pointing out the differences between the solid and the flickering line.

C. The Educational Software Component

The teaching strategies used and the type of curriculum material to be presented were the third consideration of the system used. It was felt the teaching strategy to be used and the material to be presented were interrelated. Furthermore the educational software was connected to the computer software and the computer hardware.

The area of statistics was chosen as the curriculum content area. The curriculum content of the two introductory level statistics courses in The Professional Studies area of the College of Education was selected. More specifically two topics, correlation and simple linear regression, were selected. These two topics are presented in Appendix E.

The general philosophy behind the approach taken was the concept of the computer performing in a manner a teacher or a book could not. That is, the computer should be able to make a unique contribution to the learning process in conjunction with the teacher, the book, and the other media to be used. Pressey referred to this type of instruction as adjunct instruction.

The teaching strategy developed was greatly dependent upon the size of the computer, the Cathode Ray Tube, and the terminal. A programmed learning approach was not attempted or felt desirable. Instead an approach which allowed the user to interact with the computer was used. Thus an intuitive approach to learning was taken.

The intuitive approach was such a student would be allowed to explore a concept for as long or as often as he wanted. By making inputs in the way of deleting or adding points to a scatter plot, he could note changes and determine immediately if some idea he has on a concept was correct.

In order to successfully use the intuitive approach certain assumptions had to be made at various points in a lesson with respect to previous or present knowledge. To insure this knowledge two approaches were used, supplemental material and lessons designed to present material directly.

Therefore some instructional units consisted of text and displays with very little interaction with the student. After initial viewing and trial it was felt that asking students questions as they proceeded through this material would increase the machine-student interaction. Also the response an individual gave would then give him a yardstick to measure his understanding.

The responses given were not checked by the computer for their correctness. Questions were asked only when an estimate was needed. For example, if a scatter plot was shown, a question was asked to estimate the correlation coefficient. The answer was checked to determine if there was a sign, a decimal point, and two digits. One of the reasons an exact check was not used was the accuracy involved in the calculations versus the value entered externally. The calculation routines will be discussed at a later point.

The level at which the introductory statistics courses in Education are offered is such that understanding and appreciation of general concepts is as important as actual manipulation of formulas. By providing a method of allowing the student to "see" what is involved in a concept, less apprehension of the formulas involved and a better understanding of a concept should result.

The teaching strategy was such that when a student began using the system the first output he received was an introduction to using the computer, from this he selected a concept. Appendix B contains an example. After the concept has been selected a unit is selected. It was proposed and computer programs were written to allow a student to interrupt the program and select a new unit at any point in a program. Figure 1 depicts the logic to accomplish this.

The interrupt feature was incorporated in only one unit of each concept, that being the first unit. Problems resulted in turning the interrupt on and off throughout a program since the interrupt to be used was connected with all peripheral devices. Included among these was a clock which malfunctioned in the early part of this research and was not replaced.

The clock was to be used when the computer was waiting for a response. If after two minutes a student did not respond a reminder would be sent to him. Then after a set number of reminders a warning would be sent for an immediate

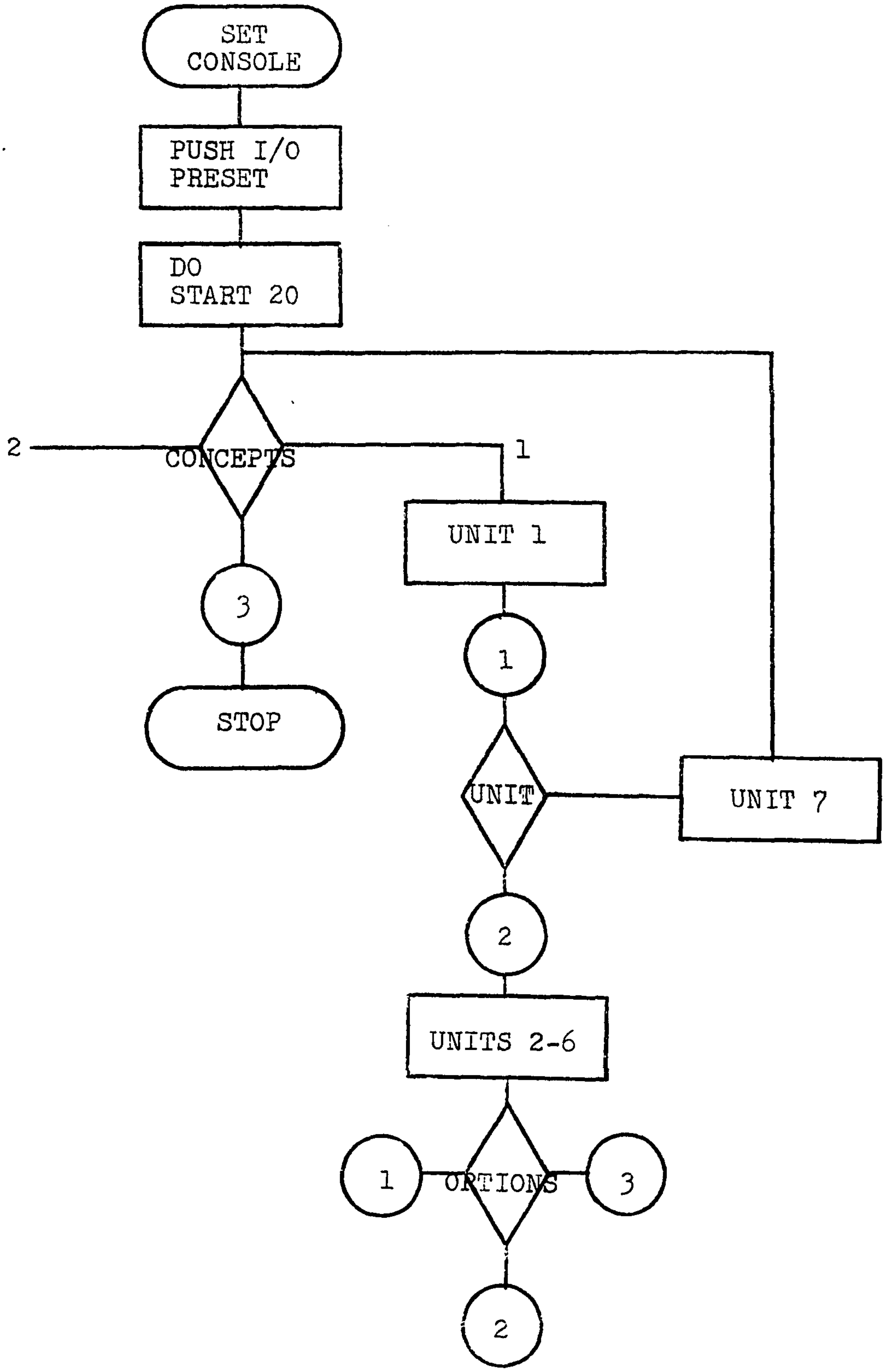

Figure 1. Flowchart of lesson logic

response, or the program would be terminated. If this could be incorporated the effect of a student leaving in the middle of a lesson or failing to stop the machine when finished would be minimal.

Although the lesson interrupt approach was not functionable it was still possible to stop within a unit and go to the unit selection position. This could be done by stopping the processor and reinitiating the program. This would amount to starting over but would allow for unit selection.

The teaching strategy employed was to an extent optional and to an extent fixed. The purpose of the project was to use the CRT, the Teletype and the computer as a unit and stress the Cathode Ray Tube for concept exploration. Therefore the teaching strategy would be a function of how the building blocks for a lesson were assembled.

The lessons were designed so that a student could guide himself through a concept as he felt necessary. Yet, there was enough direction given so that exploration would take place in a logical, meaningful manner. Formulas associated with a technique were not stressed and usually not given since it was felt another mode would better do this.

As mentioned previously, the two concepts chosen for instructional units were correlation and regression. From the investigator's experience in tutoring students in statistics these were two areas often misunderstood. The concepts included areas where a scatter plot had some meaning in describing the situation.

The decision was made to introduce regression prior to correlation which was to be explained in terms of regression. Several textbooks also use this approach, namely Snedecor and Cochran (42) and Huntsberger (30).

Each concept was divided into seven parts. The first unit in each concept was the same and is listed in Appendix C. The purpose of the unit was to present a listing of all the units available to a user. From this listing a selection as to which units he wanted to view could be made.

The remaining six units of each concept will be described in the remainder of this chapter. The regression concept will be described directly.

The second unit was on plotting a line. The purpose of this unit was to introduce how lines are plotted. By supplying the slope and the intercept an individual should be able to plot a line. The equation of the line was given and then the line was graphically displayed on CRT. An identification of the two numerical components was made and the user was asked to change first the slope, then the intercept and then the display.

At the end of the unit the option to continue adding more slopes and intercepts, go on to the next unit, or go on to some other unit was given. This unit prefaced the actual statistical units and served as a supplement to those who needed it.

The introduction to regression was made in unit three. The method used to introduce the unit was such that the least errors squared principle was stressed although not stated. A scatter plot was displayed first. To this scatter plot was added a possible location of a given X value. Next the line representing the mean was shown and finally the regression line of Y on X was shown. An explanation of what is meant by regressing Y on X is given and the last display shows both lines. In order to distinguish between the two lines, the regression line of X on Y was made to flicker.

Unit four explained the basis of linear regression, namely the least squares principle. This was done by taking a set of points and removing all but one point. Error was then defined to be the vertical distance between the point and the line. It was then noted if you have only two points the line passes directly through them without error. Next three points and the associated regression line were shown.

When a person reached unit five or if he selected unit five he could begin experimenting with some data points. Unit five was therefore designed to let the user enter data pairs and then scatter plot would be displayed. In addition the regression coefficients were calculated and the regression line was displayed. As part of the calculation routine the means of X and Y were printed, along with the regression model and the correlation coefficient.

Once a set of data had been entered, additional points could be added or deleted or a new set of points could be entered. It was hoped a student would explore using this unit. To insure exploration by a student on certain points unit six was added.

Unit six was to show the effects of points which were extremely deviant from the scatter plot. The effects of deleting these extreme points were next shown with an appropriate discussion. The student was then encouraged to add or delete extreme points to note their effects.

Unit seven was originally designed to display confidence regions for the predictions made. This was later abandoned and replaced with a unit which gave instructions to the user on how to sign off. It also provided for the reloading of the beginning of the CAI program where the concepts were selected.

The units in the correlation concept were established in approximately the same order as the units in the regression concept. Units two, five, six and seven were identical in purpose except adapted to correlation. The third unit within the correlation concept was connected very closely to unit two of the correlation concept.

The purpose of the second and third units was to establish an understanding for the meaning of correlation. Various scatter plots were displayed with either the correlation coefficient directly expressed or the user first asked to give an estimate of the correlation. The correlations displayed ranged from 1.00 to -1.00. When a correlation coefficient was entered by a user, a check was made on the sign, the decimal point, and two digits. Stressed then was the direction of the relationship and the idea the relationship was either perfect, 1.00, or less than perfect and denoted by a decimal number.

Unit four examined the relationship of correlation and regression not just from an algebraic standpoint but also from a geometric point of value. The angle formed by the two regression lines of X on Y and Y on X is directly related to the correlation between X and Y. The correlation is the cosine of this angle. Also the correlation may be expressed as the square root of the product of the slopes (beta coefficient) of the two regression lines.

Various correlations were shown to demonstrate this point. Values ranging from 1.00 to -1.00 were used. Since the lines would be perpendicular to each other when the correlation coefficient equals zero only one line was shown, that being the mean of Y on the regression of Y on X. The regression of X on Y would be the mean of X when the correlation between X and Y is zero.

IV. DESCRIPTION OF COMPUTER PROGRAMS

The programs used will be described in more or less the order they appear. Programs were written for special purpose needs and general building blocks. Special purpose programs will be described first.

The first program encountered was a bootstrap introduction routine. Since a user must initialize the program through the use of the console switches a simplified setting of the controls was needed. The starting procedure is to depress the following three switches in their respective order. They are I/O PRESET, DO, and START 20.

The I/O PRESET clears the accumulator and clears all input/output device flags and operations. The I/O PRESET also causes the Instruction Field register to be set to 2 and the Data Field register to be set to 3. The DO switch when depressed causes the processor to perform one instruction that being the one located in the LEFT SWITCH. If the instruction is a double word the LEFT and RIGHT switches are both used.

The START 20 causes the instruction in location 20 of the instruction register to be executed. The instruction is also incremented to the next instruction. In order to make this as simple as possible the LEFT switches were set to 0700 octal. This meant that left switches except 7-9 were set to zero. The RIGHT switches were all set to 0. The instruction of 0700,0000 is a double word instruction which says to read from tape, one block. The location of that block is 0000, the first useable block on the tape, and the tape unit is unit 0.

The block of tape memory is read into core, locations 0-377. The instruction read into location 20 is another read instruction which will fill up the remaining instruction and data fields. Instruction 20 is executed when the START 20 switch is depressed.

Once this has been done the instructions continue with the introduction printed and the computer waiting to go on to the concept selected.

When the concept is selected the program jumps to a program which fills up all four segments of core. The description of each segment will be given next.

Segment zero is the monitor segment of core. It controls the reading of additional or new instructions and data for their respective fields. Segment zero was chosen since if an interrupt occurs the program automatically jumps to location 40 of segment zero. If the interrupt was to change units, segment zero was the place to put the program to enter a new unit.

When an interrupt did occur it was processed to determine what the interrupt was. If the interrupt was caused by the clock, a PLEASE RESPOND command was given. If the interrupt was caused by a Teletype input, the input key had been struck. If the LINEFEED key had not been struck it was assumed the input was extraneous and the program returned to the point of interrupt and proceeded on. If any other type of interrupt occurred such as a tape motion, program control was returned to the point of interrupt and the program continued.

The programs residing in segment zero provided one other function, which was filling the instruction and data fields when a unit had been selected. This portion of the program could be entered either through an interrupt or through the end of a unit. When a unit is selected 7100 or 7700 is added to the unit number times 10. This created a read instruction which would read the unit from the appropriate blocks on tape. For example if unit two in the regression concept was selected the instruction would be 7120. This says read 8 blocks of tape beginning at location 120. The first four blocks of tape contain the instruction field and are placed in segment two. The last four blocks of tape contain the data field and are placed in segment three.

Since only one field can be an instruction field at one time, the instruction field to be, segment 2, is treated as a data field when segment zero is the current instruction. When the transfer of instruction and data have been completed, the data field is set to three and the instruction field is set to two. A jump to location 30 in the instruction field is initiated and the program proceeds from there.

The next program to be discussed is the one which resides in segment one. The routine is the calculation routine. This program given a set of data will calculate the mean of X and Y, the variance of each, the simple linear regression coefficients, and the correlation coefficient between X and Y.

The length of the program is such that it takes up all 1024 registers in the second K of storage. The arithmetic involved in processing the data was the major function and problem of this program. Built in hardware arithmetic routines consisted of add and multiply. To be built was a divide and a square root routine.

The decision was made that the right most four bits of the accumulator were to be used for the decimal portion of a number. Of the remaining eight bits seven were for the real portion of the number and the first bit for the sign. Single register arithmetic was performed throughout.

Both overflow and round off errors were problems. All arithmetic was in base-eight, which did not cause any problems. Problems were encountered in the area of the maximum number which could be handled. Using the available eight bits the maximum number was 275. Using the maximum eleven bits the maximum number was 2,047.

Obviously in statistical calculation the sums of numbers squared and the square of numbers summed can easily exceed the maximum if some restrictions are not placed on the numbers. The first restriction was that only integers would be accepted as input. The second restriction was that only the integers from zero to fifteen would be used. A further restriction was then placed on the value that the maximum number of points entered would not exceed thirty-two.

The expected value of the mean of points valued from 1 to 15 is 8. With 32 points entered the sum would be 256. Therefore with the average value used with the maximum number of points an overflow would not occur.

Using four bits for decimals, accuracy in the tenths place was usually maintained. Decimals were printed to the hundredths place to allow the user to round off. Division occurred by going through a divide routine twice, once to obtain the real portion of a number and once to obtain the decimal portion.

The statistical calculations to be performed were done in the following order. The sum and a count of the number entering was made first. In addition the numbers were squared and summed including the crossproducts. Next the means were calculated and the deviations squared. From these values the regression coefficients were calculated and then the correlation coefficient was calculated.

The routine also handled the printing routine to output the appropriate values. In addition it placed the calculated coefficients in their proper beta register in the instruction field to which it was to return.

Of all the programs written and needed the calculation routine proved to be the most difficult and unsatisfactory. This will be discussed more in the next chapter but it also deserves mentioning here. The intent of the instruction provided was not that the computer was to be a powerful calculating device but that the computer was to assist in demonstrating a particular concept. The calculating capabilities were to be secondary. The problem stemmed from the loss in accuracy and the resulting inaccurate display which an eleven or less bit number could produce. The problems are surmountable and powerful calculation routines do exist for the PDP-12. But these routines could not fit into core with the other units or current lesson.

The remainder of this chapter will be devoted to the contents of the other two segments of core, the instruction field and the data field. Rather than discussing each unit which was written individually, the routines that go into comprizing a unit will be discussed. These routines were referred to as building blocks using the idea of placing blocks of instructions together to build a unit.

Each unit is designed in approximately the same manner. Since in normal running or access to the instructional segment of core, location 20 is the first register used; this is where programming begins. Locations 20 through 27 are used for writing the program or unit on tape. The instructions in these first eight registers designated the tape and the place on the tape which the program is to be written.

After the routine is written on tape the processor is told to stop and will not proceed until told to do so.

Then beginning at location 30 the unit is written. This is true for all units so that control to and from units can be initiated at the same place. The unit is then built once an instructional strategy has been decided upon. The instructional strategy consists of the order and the content of text and point displays, and in addition the changes which are to be made in the point displays.

In order to accomplish using this instructional strategy, a process of setting what are referred to as beta registers were set which provided to a routine, information on location and length of content. For example if text is to be printed, register 11 was set to the location, an alphanumeric name, of the first character to be printed. After register eleven has been set the program jumps to PRINT and prints, beginning at the location specified in register 11.

Likewise, if next the program is to display points the location and number of these points must be given to the display routine. The point display routine was revised in the later stages of development so that instead of needing the location, all points were moved to one location prior to display and only the number of points was needed. This was done to facilitate easier addition and deletion of points.

After points were displayed the lesson designer might want to have the student add more points. This would require a JMP ADPT instruction to the add point routine. This routine then presents to the student the appropriate questions to allow him to enter more data. In order to explain this more fully a complete description of all the routines used will be given.

An example of how the instruction would look in an actual lesson is given in Appendix D. These instructions will print the text in Appendix C. The student may be asked to add points and then the points will be displayed. Then he may be asked to delete points and a new display can be generated. After the points are deleted a new set of regression coefficients are calculated and printed.

The first routine was the print routine. Two approaches were used in writing this routine. The initial approach was that the routine was to print a set number of points from some location. These characters were chopped ASCII characters and entered by the lesson designer in that form. Since this required the user to convert alphanumeric characters to numeric ASCII code a PL/1 program was written to perform this duty. Although this proved feasible it was not practical. Should a grammatical or spelling error occur the entire process had to be repeated and a new ordering of the packed characters had to be established. If the removal or addition of a single character was needed, a spacing problem arose since two characters were stored per location.

The second location used was one which used a LAP-6 pseudo-operation called TEXT. By typing the word TEXT F (string) F, all the characters between the two F's would be assembled and in turn could be printed. This made it possible for anyone to add and alter the text he wanted as he built his units. A unique name was associated with each TEXT pseudo-op and could be referred to for the address of the character string.

The PRINT routine itself used the address it was given and printed characters until the symbol, $, was encountered. Therefore when someone entered their text the dollar symbol was placed at the end of the text prior to the F. Upon completion of the printing the instruction counter was set to the instruction immediately following the jump (JMP PRINT). In all routines control was returned to the point where the jump occurred plus one address.

When points are to be displayed several routines are encountered. The first routine is the transfer routine, TRANS. This is used to transfer the points which are to be displayed from their storage location to location 1700 and following in the instruction field. Using a work area from which points were displayed was used since it was desirable to both add and delete points. Before you could jump to the TRANS routine two beta registers needed to be set. One register, 12, contained the number of points, and 13 was set to the location of the first data point.

The actual display of points occurred through the DSPP routine. The X and Y coordinates were packed such that both were placed in one word. The X coordinate took up the left most six bits and Y coordinate took up the right most six bits. In order to display a point the abscissa must be in the accumulator and the ordinate in beta register two. When this has been accomplished the point could be displayed. This process must be repeated for every point to be displayed.

The screen used to display points has 512 x 512 display points available. In order to display a point in proper perspective each point was multiplied by forty to obtain the coordinates of the point for display purposes.

Regeneration of the display was not handled by the display program. Three special routines were used to accomplish this. These routines referred to as ROUT1, ROUT2, ROUT3 performed the regeneration of three distinct displays. The amount of material to be displayed increased with each routine.

ROUT1 was used to regenerate a set of points only. In all three the display would stay on the screen until a key was struck. In ROUT2 a set of points and a line were regenerated. In ROUT3 a set of points and two lines were generated.

ROUT3 needs somewhat of a special explanation. The lines to be generated in both 2 and 3 were regression lines previously calculated or set. ROUT2 assumed the line is the regression of Y on X. The lines displayed in ROUT3 were the regression of Y on X and X on Y. In order to distinguish between the two or to allow the instruction to refer to one of the lines, the lines were not regenerated at the same rate. This caused one of the lines to flicker and could be referred to thusly.

In addition to displaying points then, a line or lines could also be displayed. ROUT3 used the same line display routine to display both lines. This was accomplished by interchanging the parameters which needed to be set before entering the routine.

The two parameters needed were the slope and the intercept. These were taken from beta registers fifteen and sixteen respectively. Given these two parameters the line was calculated one point at a time. Basically these were the steps involved. Given an intercept the slope was added to it. This then gave the proper coordinates for the point. A cumulative slope was also kept which was added to the intercept for the next point. This process was continued until the right or the top of the screen was exceeded.

Several problems arose in this approach due to accumulator overflow. The cumulative slope would at times appear to be negative due to its size. To avoid this a check was placed on the cumulative slope so that the intercept could be altered to absorb some of the cumulative slope.

In addition to this check, checks were needed to determine if the intercept was less than zero or greater than 15.

If the intercept was greater than 15 the line could not be displayed if the slope was positive. But if the slope was positive and the intercept was negative the line should be displayed. This involved setting the Y intercept to zero and the X intercept to some positive number. Negative slopes were handled in approximately the same manner. One notable exception in program performance with the negative slope and that was in the failure of the program to detect the completion of a line, which would cause two or more lines to form. Also extraneous points were generated given certain slope intercept combinations. This problem was not resolved.

In addition to the points and line display programs a third display program was written. This program displayed the X and Y axis for the first quadrant and in addition placed fletches on the axis for the fifteen numbers to be displayed. This program was not used in conjunction with the point display and line program because of the time it took to generate any two or all three of the displays. The flicker created was not at a tolerable level when the grid, a line, and points were displayed simultaneously.

The next two programs to be discussed are similar in concept and opposite in function. They are the add point routine and the delete point routine. The add point routine, ADPT, was written to allow a student to enter data. This could be new data or an addition to existing data pairs.

The student was first asked how many pairs of numbers he would like to add. He was next told how to enter the numbers. After this has been done, numbers are entered until a match is made on the number of points added, and the number of points to be added. If a number is added which exceeds 15 or if the input is not numeric the student is told the response is improper.

Points were added to the working area only. In order to keep track of the number of points the contents of beta register 12 were used as the current number of points in the working area. This number was added to 1700 to determine the next available location. Register 12 is updated to the current number of points.

The delete point routine, DELETE, began by asking the student how many points he wished to delete. The student was told how to enter points and then told to do so. When the coordinates of a point were entered a search was made of existing points to find a match. If a match was found the point's coordinates of the last point in the file was used in its place. The length of the file was then decreased by one.

If the coordinates input by the student could not be found in the file the user was informed of this. The program then waited for new input. Likewise, if a match was found, the program would wait for new input or continue on if appropriate.

Both the ADPT and DELETE programs were found to be not completely satisfactory even though they performed as expected. The lack of flexibility in changing the number of points to be added or deleted placed a restriction on the user. For instance once a user specified he was going to eliminate five points he had to delete five before the program would continue. Ideally this should have been designed to either eliminate one point and then ask if another point is to be removed or the option to change the number of points to be deleted throughout the program was needed.

A building block which was added late in the research's development was a routine called ASK. The purpose of this program was to allow the instructor to ask the student a question and then check the format of the input. The program was used to make estimates on correlation coefficients.

Almost every unit ended in approximately the same manner. The student was given a choice of either selecting a new unit, continuing on with the next one, or going through some aspect of the just finished unit again. Since very few changes would be needed from unit to unit this also could have served as a building block.

V. DISCUSSION

The research project was undertaken as a feasibility study on presenting a Computer Assisted Instruction unit in statistics on a small digital computer. The lessons presented were not necessarily unique presentations of statistical concepts but were presented to determine the feasibility of this approach.

Experimental research on evaluating this approach was not a part of the study undertaken. The opinions and reactions of students and interested staff was utilized to obtain a feeling for approaches and displays.

As in any developmental study the answer to the value of this is still speculative. The effects of the approach used are unknown. But one of the many steps in research is development, and hopefully a study such as this provides guidance and direction for future research.

The purpose of this chapter was to provide observations on the research undertaken. Also recommendations for future study on the PDP-12 and on CAI are made.

A. Observations

Without any hard data, the observations are for the most part subjective opinions of the investigator and others. The study was essentially devoted to two areas. The first was the man-machine interaction and the second was the software-hardware interaction needed to implement the desired man-machine interaction.

For CAI one of the obvious considerations is the means of communicating with the user and what is to be communicated. In this study one of the reasons the PDP-12 was chosen was its CRT since a quick graphical display device was needed. Also a Teletype was needed for communicating with the user and for the user to communicate with the program.

The CRT proved to be as useful as expected in displaying points and lines. If intuition can be effectively used and if observing the concept increases a person's knowledge of the concept then indeed, the CRT can be most useful. Since the image being displayed constantly needed to be regenerated, a restriction on the amount of material to be displayed was necessary. This restriction amounted to the total number of display coordinates it was possible to generate in a set length of time without the flickering of the screen being intolerable.

Ideally the CRT should be handled separately in the hardware configuration but this was one of the problems to work around in using the PDP-12. The approach taken to display points and lines seems reasonable, especially with a building block approach to lesson design. But a faster and more general display could possibly be generated if all coordinates of points to be activated were calculated and stored in a work area and then a display routine would take these coordinates and use them appropriately. This technique was not available in the core memory area. Dumping and loading tape would be required.

Since the Teletype output characters at the rate of 10 characters per second and the CRT was almost instantaneous, the CRT is a more preferable information giving device. But again core storage was at a premium and the Teletype could and was used to put forth the necessary information.

The relationship of the input and the output devices is an important one in CAI. In this study the input was only through the Teletype with the output through both the Teletype and the CRT. Some lessons were designed around the user entering or deleting data and observing results. The data points were displayed after the addition or deletion of points. Even though a listing of points could be obtained, a better "feel" of the scatter plot would have been brought forward if the user could have observed the display as he manipulated the points. The flicker problem was such in the display already written that in order to accomplish this double function a much more efficient program would need to be written.

The physical location of the CRT, and the Teletype was such that they were considered to be a part of the PDP-12's console. The CRT was actually located within the console directly above the console switches. The Teletype was moveable but obviously needed to be located in the proximity of of the CRT. Location proved to be more of a problem than anticipated.

The proximity of the CRT and console switches was taken into consideration for student use. The student was given complete control over the machine and the instructor merely had to give him the computer tape to execute a lesson. The Teletype was moved from approximately directly beneath the CRT to several feet to the right of the CRT and below the level of the screen. The student was required to look from the Teletype to the screen and back and forth. Instructions to look were not always explicitly given since it appeared obvious to look back and forth, or the screen would easily catch your eye when needed. It was also assumed the noise of the typewriter and the lack of image on the screen would divert the attention of the user to the Teletype when needed. Attention from the screen to the Teletype occurred correctly. But the opposite was not necessarily the case. The need for the placement of both devices within the student's peripheral vision was apparent.

The display of the scatter plot of points worked as planned even though the grid routine which was not implemented. As noted previously the routine was to provide the X and Y axis and the appropriate scale. Users of the system did not seem to have any trouble discerning the location of the points or the concept being brought out. Obviously the relations of the data points to each other was what was important and what the student was to learn.

One more point needs to be observed with respect to the Teletype. The speed of ten characters per second did not seem to be a problem for the user. But for the lesson designer, the amount of time it took to list a completed program was great. The complete listing of a program using all of core would take approximately 60-90 minutes.

A second area of observation on the hardware is on the overall storage of the computer. A major undertaking in this study was attempting to use a computer of limited size for CAI. Obviously CAI can take place on any computer which allows for remote or direct input/output, but the extent to which this is possible varies with the machine. The PDP-12 with 4 K of memory and two tapes was definitely limited in size. Yet it was unique in its display possibilities and user access.

It was felt that a structured-branching, programmed learning approach was not applicable or appropriate for a computer of this size or capabilities. The instruction to be given was keyed around the CRT and the student's intuitive explorations.

For many of the routines the size of the computer was more than adequate to the approach used. But in several a more in-depth look must be taken. Already referred to is the need for a continuous display and the extra memory needed to accomplish this. One area of serious concern and a problem was the arithmetic requirement needed.

The manipulation of the data involved in statistical routines is such that accuracy is a must. Using one K of memory and single register (12 bit) arithmetic did not provide the desired level of accuracy. However the rationale was used that the purpose of the problems were for demonstration and not for the user to calculate results of homework problems.

If a general answer was to be given as to the feasibility of using a computer of this size for CAI it would need to be positive. Indeed a larger system may be capable of more tasks but this system is capable of allowing a student to guide himself through various lessons.

The approach used to construct units proved to be one of the major assets of the system. Programs could be built and tried prior to entering on tape. This allowed for corrections or trial of changes without destroying the existing program. The ease with which units were entered on tape was considered important for lesson designers unfamiliar with the system. Along with the quickness of changing the program was the ease with which textual errors could be brought from storage, changed, saved, reassembled and resaved as a correct unit.

Since the units were independently saved on tape, a tape could be given to a student which he could use as instructed. Back-up systems containing source units and CAI units are easily maintained should a tape be destroyed.

Several observations should be made on the units. The design and the instructional philosophy was to allow a student to explore. Hopefully an attitude of "Let us see what would happen if I did this" would result. This in turn would lead to an attitude of "If I did this, this should result." The program was not to guide the student but let the student guide himself.

The units were structured in a hierarchy of events but were not and did not need to be followed. Asking questions and processing answers was avoided due to the involvement and accuracy needed to interpret correct and incorrect answers. The units were designed to represent the feasibility of this approach to teaching. Shortcomings in lessons might or might not be overcome with additional text and displays.

A definite limitation to further programming design was the time spent in developing the building blocks. Approximately twelve months elapsed in the total programming development. It is also interesting to note the hardware and software do not necessarily reflect on the sophistication of the material presented.

From a lesson designer's point the building block approach seems to be reasonable. The regression concept was used to guide the creation of the building blocks. The correlation concept was then built using the building blocks. This approach proved quite satisfactory with the correlation units taking considerably less time to construct.

B. Recommendations

Since this study has indicated it is feasible to use a "small" computer for CAI and it is also feasible to use the approach undertaken, recommendations for further research are made. The next obvious step in a project such as this is to experimentally test results in learning using this approach.

This should employ all the rigors associated with good research methodology to obtain objective information on the usefulness of this approach. The results of such research must be used in conjunction of what is known on other teaching strategies.

But prior to the implementation of such an evaluation, additional development is suggested. One of the major problems encountered was in the arithmetic routines. Additional study on implementing better arithmetic routines needs to be made. Since independent statistical packages exist for the PDP-12, methods of adapting them for the available storage are needed. These programs could give an additional feature in each lesson and be an aid in themselves.

Another method of approaching the task under investigation is through a language named FOCAL-4K with display capabilities. The FOCAL-4K language did exist at Iowa State University for the PDP-12 but without the display capabilities.

Since the FOCAL language is somewhat easy to use, this approach needs further consideration. Of special merit would be the calculating capabilities of this language.

Further research also needs to be done in the use of the Cathode Ray Tube in conjunction with and opposed to a Teletype. Even such elements as the physical placement of the two devices are important.

Additional study might be undertaken on alternate display routines, or new approaches to displaying points. One possibility might be to calculate all the coordinates of the points prior to actually displaying them. A consideration would need to be the amount of storage needed to accomplish this.

More units in statistics should be implemented such as graphically depicting analysis of variance or the shapes of probability distributions. Again the emphasis should be on student involvement and interaction with the programs.

An undertaking to be processed with the PDP-12 is the use of a higher level language to generate code for the PDP-12. For instance PL/1 might be used to interpret lesson code into PDP-12 code. Two benefits might result. One, a very easy and understandable method of lesson building should result. Secondly, programs could be listed from a high speed printer. Time would not need to be spent in obtaining program listings through the PDP-12 teletype.

Additional study might also be undertaken to expand the storage available by judicious use of the two magnetic tapes. In this study the tapes were used to initialize the programs and store programs. They might also be used to store core memory so that new instructions could be brought in and out and thus create longer programs. The tapes might also be used to contain a record keeping system.

A record of every student who signed on, what he selected, and how long he worked, would be kept. The student could note his progress and also the lessons most frequently chosen. A record keeping system is most important if the system is to be used in a large class setting.

The curriculum area chosen was done so because of the investigator's background and also because of a computer's ability to perform calculation in a rapid and precise manner. Other areas also might be considered such as displaying paradigms of statistical designs. There are also possibilities in areas where displaying objects may be more meaningful than verbalizing. Elementary education is possibly one area.

As a computer should function in the area it is best suited so should certain sizes and types of computers. A computer such as what was used in this study is small with respect to core storage. It can not perform all the duties and functions of a "large" computer. Yet its capabilities as to instructional powers are somewhat unexplored.

VI. SUMMARY

This study examined the feasibility of using a PDP-12 computer to teach selected statistical concepts. The study also examined a CAI unit building approach and the use of a Cathode Ray Tube as an integral part of a lesson. The emphasis of the study was on the development of the system. Problems centered around the software development in implementing the proposed CAI system.

The computer used was a PDP-12 produced by Digital Equipment Corporation. Core memory was 4 K words with additional storage available on two, 512 block tapes. The programming language used was an assembly language called LINC. All programs were entered through a DIAL-2, LAP-6 monitor program. This LAP-6 system allowed for on-line editing of programs and was used to assist in building the instructional units.

The software component involved the designing and building of "blocks" of programs to be used to develop units. The building blocks were designed so that they could be selected as needed when a unit was designed. The major routines built were DSPP,(display points), ADPT (add points), DELETE (delete points), LINE (line display), and PRNT (print text). A simple SET function was used prior to jumping to the routine.

Since the units were to be student guided a monitor routine to accomplish this was written to reside in core memory.

This routine controlled the filling and dumping of the instruction and data fields of core as new units were requested. A calculation program was also written to reside in core at all times. This program performed all statistical calculations needed and could only be entered by a change in instruction fields initiated by the current instruction field.

The third component developed was the educational component. Statistics was the general subject matter area selected with simple linear regression and correlation being the precise concepts chosen. An intuitive approach to lesson presentation was selected to make use of the Cathode Ray Tube and the Teletype.

Student interaction with the units was stressed and explanation was encouraged. Student inputs, data pairs, were displayed via the CRT and the resulting effects of these inputs to existing data were either displayed, output through the Teletype or both. The CRT and Teletype were to be used to perform functions unique to each and not readily performable through other instructional modes.

The regression units were developed along with the building blocks. The correlation units were built with the building blocks to determine the feasibility of using this approach. Except for minor changes in some programs the approach was an acceptable one.

Major problems occurred in anticipated areas due to the restricted core storage available and the length of programs needed. The calculation routine was not as accurate as desired nor was the CRT as versatile or as useful as desired. The display routines directly reflected the accuracy achieved in the calculation routines. If the inaccuracy was discernable it was obviously so, usually resulting in a regression line of opposite sign and inflated value.

Creating and maintaining displays proved to be a difficult task. The total number of points was restricted to a maximum of 32 as well as all points had to be integers. Nevertheless, displaying a set of points and two regression lines produced a flickering screen at a less than desirable level. Regardless of this flicker, it appeared the concept being brought forth was still evident. Prior to additional use of this system further investigation on other approaches to displays needs to be developed.

Although the student guided approach was a reality, it was not as functional as proposed. Interrupting lessons at any point was not always a feasible feature. Two types of interrupts were used: (1) software and (2) hardware. The hardware interrupt was accomplished by using the console and stopping the processor and then re-initializing the entire program. The software interrupt for lesson branching worked in selected units.

The basic question of feasibility has been answered in the affirmative. Now the next step of experimentally evaluating the approach is needed. It is recommended this be done and additional concepts be added. It is also recommended that prior to evaluation, further work be done on modifying and developing statistical packages to meet the needs of an approach, and a computer, such as was used in this study.

VII. BIBLIOGRAPHY

1. Atkinson, R. C. and Hanson, D. N. Computer-assisted instruction in initial reading: The Stanford project. Reading Research Quarterly 2: 5-25. 1966.

2. Atkinson, Richard C. and Wilson, H. A., eds. Computer-assisted instruction. New York, N.Y., Academic Press. 1969.

3. Ausubel, David P. Educational psychology, a cognitive view. New York, N.Y., Holt, Rinehart, and Winston. 1969.

4. Bitzer, Donald L., Lyman, Elisabeth R., and Easley, John A., Jr. The uses of PLATO: A computer controlled teaching system. Audiovisual Instruction 11, No. 1: 22-23. 1966.

5. Bloom, Benjamin S. Taxonomy of educational objectives, cognitive domain. New York, N.Y., David McKay Co., Inc. 1956.

6. Borko, Harold. Computer applications in the behavioral sciences. Englewood Cliffs, N.J., Prentice-Hall, Inc. 1962.

7. Branstad, Dennis K. A computer aided instructional system for teaching formal languages. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1970.

8. Branstad, Dennis K. A feasibility study of implementing a CAI system on a small digital computer. Unpublished M.S. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1968.

9. Brunner, Jerome S. On knowing, essays for the left hand. Cambridge, Mass., Belknap Press. 1962.

10. Brunner, Jerome S. Toward a theory of instruction. Cambridge, Mass., Belknap Press of Harvard University Press. 1966.

11. Bushnell, D. D. and Allen, D. W., eds. The computer in American education. New York, N.Y., John Wiley and Sons, Inc. 1967.

12. Butler, Cornelius. CAI in New York City. Educational Technology 9, No. 10: 84-87. October, 1969.

13. Cooley, William. Computer assisted instruction in statistics. Unpublished paper presented at Conference on Statistical Computation, University of Wisconsin Computing Center, April, 1969. Pittsburgh, Pa., University of Pittsburgh. 1969.

14. DeCecco, John P. The psychology of learning and instruction. Englewood Cliffs, N.J., Prentice-Hall, Inc. 1968.

15. Digital Equipment Corporation. System reference manual. Maynard, Massachusetts. Digital Equipment Corp. 1970.

16. Digital Equipment Corporation. Introduction to programming. Maynard, Massachusetts. Digital Equipment Corp. 1969.

17. Frye, Charles H. CAI languages: capabilities and applications. Datamation 14, No. 9: 34-37. Sept., 1968.

18. Glaser, R. and Ramage, W. W. The student-machine interface in instruction. Pittsburgh, Penn., University of Pittsburgh, Learning Research and Development Center. March, 1967.

19. Glauberman, M. H. Computers in education on RCA viewpoint. Educational Technology 1, No. 1: 63-67. Sept., 1969.

20. Goodlad, John I. Learning and teaching in the future. NEA Journal 57, No. 2: 49-51. Feb., 1968.

21. Goodlad, John I., O'Toole, John F., Jr., and Tyler, Louise L. Computers and information systems in education. New York, N.Y., Harcourt, Brace and World, Inc. 1966.

22. Gorber, Evan. Individualization of instruction for teacher corpsmen. Educational Technology 11, No. 8: 31-35. August, 1969.

23. Grossman, Alvin and Howe, Robert L. Data processing for education. Chicago, Ill., Educational Methods, Inc. 1965.

24. Grubb, Ralph E. Learner-controlled statistics. Programmed Learning 5, No. 1: 18-24. 1968.

25. Haga, Enoch, ed. Automated educational systems. Elmhurst, Illinois, The Business Press. 1967.

26. Hartman, Thomas F. Computer assisted instruction. Audiovisual Instruction 11, No. 1: 22-23. 1966.

27. Hickey, Albert E. Computer-assisted instructions: a survey of the literature. 3rd ed. Newburyport, Mass., Entelek, Inc. Oct., 1968.

28. Hickey, A. E. et al. Requirements for graphic teaching machines, final report. HEW Grant 7-31-0510-161. Boston, Mass., Northeastern University. 1962.

29. Holtzman, Wayne H., ed. Computer-assisted instruction, testing, and guidance. New York, N.Y., Hayes and Row, Publishers. 1970.

30. Huntsberger, David V. Elements of statistical inference. Boston, Mass., Allyn and Bacon, Inc. 1961.

31. McClain, Donald H. Development of a computer-assisted instruction unit in probability. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1970.

32. Margolin, Joseph B. and Misch, Marion R. Computers in the classroom. New York, N.Y., Sparten Books. 1970.

33. Marks, Martha. A student directed, computer assisted, programming environment. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1970.

34. Molner, Andrew and Sherman, Beverly. U.S. Office of Education support of computer activities. Educational Technology 11, No. 4: 5-9. April, 1969.

35. Ofiesh, Gabriel D. and Meierhenry, Wesley C., eds. Trends in programmed instruction. Washington, D.C., National Education Association. 1964.

36. Pressey, Sidney. A simple apparatus which gives tests and scores and teaches. School and Society 23, No. 586: 373-376. March 20, 1926.

37. Rogers, James L. Current problems in CAI. Datamation 14, No. 9: 28-33. Sept., 1968.

38. Rosenbaum, Joseph and Bennik, Fred O. CAI at Systems Development Corporation. Educational Technology 9, No. 2: 12-13. Feb., 1969.

39. Schwartz, H. A. and Long, H. S. Instruction by computer. Datamation 12, No. 9: 73-87. Sept., 1966.

40. Skinner, B. F. Teaching machines. Science 128: 969-977. 1958.

41. Snarnoff, David. No life untouched. Saturday Review 59, No. 3: 21. July 23, 1966.

42. Snedecor, George W. and Cochran, William G. Statistical methods. 6th ed. Ames, Iowa, The Iowa State University Press. 1967.

43. Stolurow, L. M. What is computer assisted instruction? Educational Technology 8, No. 15: 10-11. Aug., 1968.

44. Stolurow, Lawrence M. Computer assisted instruction. Detroit, Michigan, American Data Processing, Inc. 1968.

45. Suppes, Patrick. The uses of computers in education. Sci. American 215, No. 3: 207-220. 1966.

46. Suppes, Patrick and Jerman, M. Computer assisted instruction at Stanford. Educational Technology 11, No. 1: 22-24. Jan., 1969.

47. Suppes, Patrick and Morningstar, Mona. Computer-assisted instruction. Science 166, No. 17: 343-350. Oct., 1969.

48. Thomas, Rex A. Computerized simulation for teaching experimental design. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1970.

49. Treyise, Robert L. Report on a national conference on computer applications to learning. Ed. Technology 10, No. 12: 60-62. December, 1970.

50. Uhr, L. Teaching machine programs that generate problems as a function of interaction with students. National Conference Association for Computing Machinery Proceedings 24: 125-134. 1969.

51. Witt, Paul F., ed. Technology and the curriculum. New York, N.Y., Columbia University, Teachers College Press. 1968.

52. Zinn, Karl. Instructional programming languages. Ed. Technology 10, No. 3: 43-46. March, 1970.

VIII. ACKNOWLEDGMENTS

The author wishes to express his appreciation to those who have contributed to this research and my graduate program.

Appreciation is expressed to the Iowa State University faculty members who served on my thesis committee. A special thank you is given to Dr. Rex Thomas for his assistance and support throughout this project.

The author also is grateful for being a recipient of an Education Research Fellowship from the U.S. Department of Health, Education, and Welfare and extends his appreciation to Dr. Howe for helping to make this financial aid possible.

Deepest appreciation is expressed to my wife, Wendy, for her assistance, understanding, and encouragement given throughout this project.

IX. APPENDIX A. OPERATION PROCEDURES

OPERATION PROCEDURES

The PDP-12 is a general purpose computer of which you will have complete control. To start the computer the following events must be checked:

1. The tape marked 0 must be threaded on the tape transport marked 0.

2. The toggle switches should be set to write LOCK and REMOTE.

3. The BRIGHTNESS ON knob located by the Cathode Ray Tube should be rotated clockwise.

4. The Teletype switch should be turned to LINE.

5. On the Console twelve switches are labeled LEFT SWITCHES and twelve switches are labeled RIGHT SWITCHES.
   These switches should be set as follows.

   LEFT ↑↑↑ ↓↓↓ ↑↑↑ ↑↑↑

   RIGHT ↑↑↑ ↑↑↑ ↑↑↑ ↑↑↑

Read the remaining instructions BEFORE you do steps 6,7, and 8.

6. Depress I/O PRESTART

7. Depress DO

8. Depress START 20

If you wish to stop or if you or the machine are malfunctioning depress the STOP switch and then reset it.

To restart repeat instructions 6-8.

To continue from where you stopped depress the CONTINUE switch.

X. APPENDIX B. CONCEPT INTRODUCTION

```
****************************************************************************

WELCOME TO A FEASIBILITY STUDY USING THE PDP-12 COMPUTER TO
PRESENT SELECTED STATISTICAL TOPICS.  BEFORE I CONTINUE I WOULD
LIKE TO PRESENT THE METHOD BY WHICH YOU USE ME.
MY ORGANIZATION CONSISTS OF FILES CONTAINING UNITS WHICH YOU
MAY SELECT.  INPUTS FROM YOU ARE NORMALLY EXPECTED IN THE FORM
OF INPUT -RETURN- KEY.  YOU MAY ALSO INTERRUPT A PROGRAM AND
JUMP TO ANOTHER. TO DO THIS STRIKE THE -LINE FEED- KEY.
SHOULD YOU HAVE TROUBLE OR YOU WISH TO STOP:

     1. DEPRESS THE STOP SWITCH
     2. TURN THE TELETYPE OFF
     3. TURN THE CRT(BRIGHTNESS) TO OFF.


YOU HAVE TWO CONCEPTS TO CHOOSE BETWEEN:

     1. REGRESSION
     2. CORRELATION

SELECT AN OPTION....
```

XI. APPENDIX C. STUDENT TEXT

REGRESSION CONCEPT

UNIT 1

THIS UNIT CONTAINS A LISTING OF ALL OTHER UNITS IN THE CORRELATION PROGRAM. IF YOU DO NOT WISH TO SEE ALL THE UNITS AVAILABLE STRIKE THE RETURN KEY AND SELECT A UNIT.

UNIT 2. PLOTTING A LINE

UNIT 3. INTRODUCTION TO REGRESSION

UNIT 4. LEAST SQUARES PRINCIPLE

UNIT 5. USER ENTERS DATA

UNIT 6. EFFECTS OF OUTLIERS

UNIT 7. THE END

SELECT A UNIT NUMBER. >

UNIT 2 PLOTTING A LINE

THE LINE WHICH YOU WILL SEE DISPLAYED IS OF THE FORM

Y = AX + B.

WHERE A REPRESENTS THE SLOPE AND B REPRESENTS THE INTERCEPT.

THE EQUATION OF THE LINE IS Y= +02.00X+01.00

(DISPLAY)

CHANGE THE VALUE OF THE INTERCEPT.
GIVE ME A NUMBER OF THE FORM X.X > 1.1

THE EQUATION OF THE LINE IS Y= +02.00X+01.10

(DISPLAY)

CHANGE BOTH THE SLOPE AND THE INTERCEPT.

SLOPE= 2.5
INTERCEPT= 0.1

THE EQUATION OF THE LINE IS Y= +02.50X+00.10

(DISPLAY)

NEXT CHANGE THE VALUE OF THE SLOPE. THE INTERCEPT WILL BE RESET

TO 1.0.

SLOPE = 4.7

THE EQUATION OF THE LINE IS Y= +04.70X+01.00

(DISPLAY)

CHANGE BOTH THE SLOPE AND THE INTERCEPT.

SLOPE= 8.8
INTERCEPT= 8.8

THE EQUATION OF THE LINE IS Y= +08.80X+08.80

(DISPLAY)

CHANGE BOTH THE SLOPE AND THE INTERCEPT.

SLOPE= [illegible]
INTERCEPT= 3.0

THE EQUATION OF THE LINE IS Y= -0[illegible]X+03.00

(DISPLAY)

```
CHANGE BOTH THE SLOPE AND THE INTERCEPT.

SLOPE= 2.9
INTERCEPT= 0.0

THE EQUATION OF THE LINE IS  Y= +02.90X-00.00
```

(DISPLAY)

```
DO YOU WISH TO
                1. CHANGE MORE SLOPES AND INTERCEPTS
                2. CONTINUE WITH THE NEXT UNIT
                3. SELECT A NEW UNIT

SELECT AN OPTION....
```

UNIT 3. INTRODUCTION TO REGRESSION

SUPPOSE WE ARE GIVEN A SET OF POINTS HAVING X AND Y COORDINATES SUCH AS WHAT IS CURRENTLY BEING DISPLAYED.

(DISPLAY)

ANOTHER GUESS OF A POSSIBLE Y VALUE MIGHT BE THE MEAN. NOTE WHEN THE X VALUE IS SMALL THE PREDICTED Y VALUE IS LARGE AND VICE VERSA.

(DISPLAY)

WE WOULD LIKE TO PREDICT WHAT THE Y VALUE WOULD BE IF WE ARE GIVEN AN X. FOR EXAMPLE IF X = 7 WE MIGHT GUESS Y = 13. AS CAN BE SEEN ON THE DISPLAY THIS POINT DOES NOT FALL WITHIN THE CLUSTER OF POINTS CURRENTLY BEING DISPLAYED.

(DISPLAY)

OR WE COULD DECIDE TO DRAW SOME LINE(EQUATION) THROUGH THE POINTS WHICH COULD GIVE US THE BEST ESTIMATE OF Y GIVEN AN X. ONE WITH THE MINIMUM OF ERROR.
SUCH A LINE IS THE ONE ON THE SCREEN.

(DISPLAY)

THE LINE FORMED IS CALLED A REGRESSION LINE CALCULATED BY REGRESSING Y ON X. WHICH IS TO SAY WE ARE PREDICTING Y GIVEN AN X.

WE MIGHT ALSO REGRESS X ON Y (PREDICT X GIVEN A Y) WHICH IS THE FLICKERING REGRESSION LINE.

(DISPLAY)

@
OPTIONS,

1. SELECT A NEW UNIT
2. GO TO NEXT UNIT

SELECTION....

UNIT 4 LEAST SQUARES PRINCIPLE

THE LEAST SQUARES PRINCIPLE REFERS TO MINIMIZING THE SUM OF THE ERRORS SQUARED TO DETERMINE THE REGRESSION LINE. ERROR IS DEFINED AS THE DIFFERENCE BETWEEN OBSERVED VALUE AND PRE-DICTED VALUE(REGRESSION LINE).

FOR EXAMPLE THE MODEL ASSOCIATED WITH THE SET OF POINTS TO BE DISPLAYED IS

$$Y = 0.5X + 2.2$$

(DISPLAY)

REMOVING ALL BUT ONE POINT AND THE LINE THE DISTANCE FROM THE POINT TO THE LINE IS DEFINED AS ERROR.

(DISPLAY)

BY MINIMIZING THE SQUARE OF THIS DISTANCE WE HAVE OUR BEST RE-GRESSION LINE. THE FORMULA FOR DOING THIS ARE FOUND IN THE SYLLABUS.

IF YOU HAVE ONLY TWO POINTS THE LINE PASSES DIRECTLY THROUGH THE POINTS WITHOUT ERROR.

(DISPLAY)

NOW SUPPOSE WE HAVE THREE POINTS WHICH DO NOT ALL FALL ON A STRAIGHT LINE. THEN A LINE SUCH AS THIS WOULD PRODUCE THE LEAST AMOUNT OF ERROR.

(DISPLAY)

EVEN WHEN ALL THE POINTS ARE PRESENT THE IDEA IS TO CALCULATE A LINE WHICH HAS THE LEAST AMOUNT OF ERROR.

(DISPLAY)

SELECT AN OPTION   1. GO TO THE NEXT UNIT
                    2. SELECT A NEW UNIT

RESPOND....

```
            UNIT 5    USER MODE


THE FUNCTION OF THIS UNIT IS FOR YOU TO ENTER DATA AND I WILL
DISPLAY THE POINTS AND THE ASSOCIATED REGRESSION LINE.


HOW MANY PAIRS OF NUMBERS DO YOU HAVE? > 4

INPUT THE DATA PAIRS    X  -COMMA-   Y  -RETURN-


>1,1
>3,6
>7,9
>6,6
LISTING ? (YES OR NO)>NO

THE MEAN OF X =      +004.25

THE MEAN OF Y =      +005.50

REG. OF Y ON X:

      Y =   +001.06X +001.00

R =   +000.88



                    ( DISPLAY )



YOU HAVE FOUR OPTIONS:
         1. ENTER A NEW SET OF DATA
         2. ADD MORE POINTS TO CURRENT DISPLAY
         3. DELETE POINTS FROM CURRENT DISPLAY
         4. GO ON TO NEXT UNIT

SELECT AN OPTION.... 20

HOW MANY POINTS DO YOU WISH TO ADD? > 2
```

```
INPUT THE DATA PAIRS   X  -COMMA-  Y  -RETURN-


>10,11
>6,10
LISTING ? (YES OR NO)>YES
+0001,+0001
+0003,+0006
+0007,+0009
+0006,+0006
+0010,+0011
+0006,+0010

THE MEAN OF X =     +005.50

THE MEAN OF Y =     +007.19

REG. OF Y ON X:

      Y =  +001.00X +001.69

R =  +000.88
```

(DISPLAY)

```
SELECT AN OPTION.... 3

HOW MANY POINTS DO YOU WISH TO DELETE? >1
INPUT THE DATA PAIRS   X  -COMMA-  Y  -RETURN-


>6,10

THE MEAN OF X =     +005.38

THE MEAN OF Y =     +006.63

REG. OF Y ON X:

      Y =  +001.00X +001.25

R =  +000.88

SELECT AN OPTION....
```

****

UNIT 6 OUTLIERS

NOTE THE DISPLAY AND THE REGRESSION LINE.

THE MEAN OF X = +005.81

THE MEAN OF Y = +004.25

REG. OF Y ON X:

Y = +000.19X +003.19

R = +000.25

(DISPLAY)

I HOPE YOU NOTICED ONE SCORE OF POINT WAS NOT WITHIN THE CLUSTER OF OTHER POINTS. THIS POINT IS SUSPECT AND NEEDS FURTHER INVESTIGATION. FIRST WE WILL ELIMINATE THE POINT AND SEE WHAT HAPPENS.

THE MEAN OF X = +005.94

THE MEAN OF Y = +003.75

REG. OF Y ON X:

Y = +000.25X +002.31

R = +000.59

(DISPLAY)

NOTE THE BETTER FIT OF THE REGRESSION LINE THRU THE POINTS AND THE HIGHER CORRELATION.

INSTEAD OF ELIMINATING THE POINT WE MIGHT USE THE MEAN OF Y FOR THE Y VALUE OF THAT SCORE.

THE MEAN OF X = +005.56

THE MEAN OF Y = +003.63

REG. OF Y ON X:

Y = +000.31X +001.94

R = +000.50

(DISPLAY)

TO OBSERVE WHAT HAPPENS WITH EXTREME POINTS PRESENT ADD AND DELETE POINTS. NOTE TWO THINGS ARE AFFECTING THE REGRESSION LINE THE SMALL NUMBER OF POINTS AND THE OUTLIER ITSELF.

OPTIONS ARE
1. ADD POINTS
2. DELETE POINTS
3. CONTINUE ON

SELECT AN OPTION....

```
THANK YOU FOR YOUR ATTENTION.  IF YOU WISH TO CONTINUE SELECT

ONE OF THE OPTIONS WHICH WILL BE PRESENTED.  IF YOU ARE FINISHED

PLEASE DO THE FOLLOWING IN THE ORDER GIVEN.

1.   TURN THE TELETYPE TO OFF

2.   TURN THE BRIGHTNESS ON KNOB COUNTER-CLOCKWISE

3.   DEPRESS THE STOP SWITCH ON THE CONSOLE

THANK YOU AND I HOPE TO SEE YOU AGAIN.

YOU HAVE TWO CONCEPTS TO CHOOSE BETWEEN:

     1. REGRESSION
     2. CORRELATION

SELECT AN OPTION....
```

CORRELATION

UNIT 1

THIS UNIT CONTAINS A LISTING OF ALL OTHER UNITS IN THE CORRELATION PROGRAM. IF YOU DO NOT WISH TO SEE ALL THE UNITS AVAILABLE STRIKE THE RETURN KEY AND SELECT A UNIT.

UNIT 2. INTRODUCTION
UNIT 3. SCATTER PLOTS CORRELATED +1.0 TO -1.0
UNIT 4. RELATIONSHIP BETWEEN REGRESSION AND CORRELATION
UNIT 5. USER ENTERS DATA
UNIT 6. EFFECTS OF OUTLIERS

SELECT AN OPTION....

UNIT 2 INTRODUCTION

OFTEN IT IS DESIRABLE TO DETERMINE THE RELATIONSHIP BETWEEN TWO VARIABLES. CORRELATION IS ONE WAY OF DOING THIS SINCE IT EXPRESSES THE STRENGTH OF THE RELATIONSHIP AND THE DIRECTION OF THE RELATIONSHIP. THE RANGE OF CORRELATION COEFFICIENT IS FROM +1.0 TO -1.0. THE SIGN EXPRESSES THE DIRECTION OF THE RELATION AND THE NUMERAL EXPRESSES THE STRENGTH. THE CORRELATION COEFFICIENT MOST COMMONLY USED IS CALLED THE PEARSON PRODUCT MOMENT CORRELATION COEFFICIENT. FOR FORMULAS AND AN EXPLANATION OF HOW TO CALCULATE THE CORRELATION SEE THE HANDOUT.

THIS UNIT WILL PROCEED IMMEDIATELY INTO UNIT 3.

UNIT 3 EXAMPLES OF SCATTER PLOTS

LET US LOOK AT THE SCATTER PLOTS OF SOME SCORES AND THEIR CORRELATION. IN THIS FIRST EXAMPLE NOTE THE CLUSTERING OF POINTS AND THE UPWARD TREND OF THE POINTS. THE CORRELATION COEFFICIENT BETWEEN X AND Y IS +0.75

(DISPLAY)

IN THIS SECOND EXAMPLE I WILL DISPLAY THE POINTS AFTER WHICH YOU ARE TO GIVE ME AN ESTIMATE OF THE CORRELATION. PLEASE MAKE THE INPUT IN THE FORM -SIGN(+ OR -) DECIMAL POINT(.) TWO DIGITS AND THEN A RETURN

(DISPLAY)

CORRELATION =? > +.45
THE POINTS ARE MORE SCATTERED BUT THERE IS STILL A POSITIVE TREND.

THE CORRELATION IS +0.49

(DISPLAY)

NOTE THIS NEXT EXAMPLE AND THEN GIVE ME AN ESTIMATE FOR THE CORRELATIO

CORRELATION =? > +.05

THE CORRELATION COEFFICIENT COEFFICIENT IS APPROACHING ZERO.
THE SCATTER OF POINTS IS SOMEWHAT CIRCULAR.

THE CALCULATED CORRELATION IS +0.09

(DISPLAY)

IN THIS NEXT DISPLAY X AND Y ARE NEGATIVELY CORRELATED.

(DISPLAY)

CORRELATION =? > -.40

THE CORRELATION COEFICIENT IS -0.64. THE NEGATIVE SIGN

INDICATES THAT AS TH VALUE OF X INCREASES THE VALUE OF Y DECREASES.

(DISPLAY)

THE NEXT TWO DISPLAYS WILL SHOW A POSITIVE AND A NEGATIVE CORR.

OF 1.00. A POSITIVE CORRELATION WILL BE SHOWN FIRST.

(DISPLAY)

NEXT R = -1.00

(DISPLAY)

YOU HAVE THREE OPTIONS:

1. OBSERVE MORE SCATTER PLOTS AND THEIR CORRELATIONS
2. GO ON TO THE NEXT UNIT
3. SELECT A NEW UNIT

SELECT AN OPTION....

UNIT 4 CORRELATION AND REGRESSION

ALTHOUGH THE SCATTER PLOT OF THE POINTS GIVES US SOME FEEL FOR THE CORRELATION LOOKING AT THE REGRESSION LINES FORMED BY REGRESSING Y ON X AND X ON Y MAY BE HELPFUL IN DEPICTING THIS LINEAR RELATIONSHIP.

LOOK AT THIS EXAMPLE.

(DISPLAY)

FIRST WE WILL ADD THE REGRESSION LINE FOUND BY REGRESSING Y ON X.

(DISPLAY)

THE SECOND REGRESSION LINE IS THAT FOUND BY REGRES-SING X ON Y.

(DISPLAY)

NOTE SEVERAL POINTS:

1. THE SLOPES OF BOTH LINES ARE POSITIVE WHICH INDICATE THE CORRELATION BETWEEN X AND Y IS POSITIVE.

2. THE ANGLE BETWEEN THE TWO LINES IS VERY SMALL WHICH INDICATES THE HIGHNESS OF THE CORRELATION.

NOTE THESE TWO POINTS ON THE NEXT EXAMPLE.

R = +0.47

(DISPLAY)

IN THE NEXT EXAMPLE THE CORRELATION BETWEEN X AND Y EQUALS +1.00.

THE REGRESSION OF Y ON X AND X ON Y FORM THE SAME LINE.

(DISPLAY)

WHEN THE CORRELATION IS NEGATIVE BOTH REGRESSION LINES WILL HAVE NEGATIVE SLOPES SUCH AS THE NEXT EXAMPLE WHERE CORRELATION BETWEEN X AND Y EQUALS -0.64.

(DISPLAY)

CONSIDER ONE MORE EXAMPLE THAT BEING WHEN THE CORRELATION IS 0.00 OR CLOSE. ONLY ONE OF THE REGRESSION LINES IS DISPLAYED THAT BEING THE REGRESSION OF Y ON X. IF THE CORR. WAS EXACTLY 0.00 THE REGRESSION LINE ASSOCIATED WITH REGRESSION OF Y ON X WOULD BE THE MEAN OF Y. LIKEWISE THE REGRESSION LINE ASSOCIATED WITH THE REGRESSION OF X ON Y WOULD BE THE MEAN OF X. THE TWO LINES WOULD BE PERPENDICULAR TO EACH OTHER.

(DISPLAY)

OPTIONS 1. LOOK AT MORE EXAMPLES(NOT IN OPERATION)
2. GO TO UNIT 5
3. SELECT A UNIT

SELECT AN OPTION....

XII. APPENDIX D. EXAMPLE OF COMPUTER PROGRAM LISTING

```
                    *20
0020  0011  UNIT4,  CLF
0021  0705          WCG
0022  7640          7640
0023  0000          HLT
                    *30
0030  0011          CLF
0031  1020          LDA I
0032  0000          U41
0033  1120          ADA I
0034  2000          2000
0035  4011          STC 11
0036  6370          JMP PRNT
0037  0072          SET I 12
0040  0063          63
0041  0073          SET I 13
0042  1300          1300
0043  6246          JMP TRANS
0044  6277          JMP KOUT1
0045  1020          LDA I
0046  0237          U42
0047  1120          ADA I
0050  2000          2000
0051  4011          STC 11
0052  6370          JMP PRNT
0053  0075          SET I 15
0054  0014          0014
0055  0076          SET I 16
0056  0032          0032
0057  0065          SET I 5
0060  0025          0025
0061  0066          SET I 6
0062  7744          7744
0063  6311          JMP KOUT2
0064  1020          LDA I
0065  0307          U43
0066  1120          ADA I
0067  2000          2000
0070  4011          STC 11
0071  6370          JMP PRNT
0072  6326          JMP KOUT3
0073  1020          LDA I
0074  0356          U44
0075  1120          ADA I
0076  2000          2000
0077  4011          STC 11
0100  6370          JMP PRNT
0101  0072          SET I 12
0102  0050          50
0103  0073          SET I 13
0104  1370          1370
```

0063	0105	6246	JMP TRANS
0064	0106	0075	SET I 15
0065	0107	0007	07
0066	0110	0076	SET I 16
0067	0111	0102	0102
0070	0112	0065	SET I 5
0071	0113	0037	37
0072	0114	0066	SET I 6
0073	0115	7615	7615
0074	0116	6326	JMP ROUT3
0075	0117	1020	LDA I
0076	0120	0623	U45
0077	0121	1120	ADA I
0100	0122	2000	2000
0101	0123	4011	STC 11
0102	0124	6370	JMP PRNT
0103	0125	0072	SET I 12
0104	0126	0007	7
0105	0127	0073	SET I 13
0106	0130	1640	1640
0107	0131	6246	JMP TRANS
0110	0132	0075	SET I 15
0111	0133	0010	0010
0112	0134	.0076	SET I 16
0113	0135	0060	60
0114	0136	6311	JMP ROUT2
0115	0137	1020	LDA I
0116	0140	0726	U46
0117	0141	1120	ADA I
0120	0142	2000	2000
0121	0143	4011	STC 11
0122	0144	6370	JMP PRNT
0123	0145	0072	SET I 12
0124	0146	0064	64
0125	0147	0073	SET I 13
0126	0150	1540	1540
0127	0151	6246	JMP TRANS
0130	0152	0075	SET I 15
0131	0153	7762	7762
0132	0154	0076	SET I 16
0133	0155	0303	0303
0134	0156	0065	SET I 5
0135	0157	0025	25
0136	0160	0066	SET I 6
0137	0161	0445	445
0140	0162	6326	JMP ROUT3
0141	0163	1020	LDA I
0142	0164	1054	U47
0143	0165	1120	ADA I
0144	0166	2000	2000
0145	0167	4011	STC 11

```
0170  6370          JMP PRNT
0171  0072          SET I 12
0172  0053          53
0173  0073          SET I 13
0174  1460          1460
0175  6246          JMP TRANS
0176  0075          SET I 15
0177  0001          0001
0200  0076          SET I 16
0201  0165          165
0202  6311          JMP ROUT2          /SET 5 6 TOO A
          LARGE NUMBER
0203  1020          LDA I
0204  1420          1460
0205  1120          ADA I
0206  2000          2000
0207  4011          STC 11
0210  6370          JMP PRNT
0211  6234          JMP IN
0212  1460          SAE I
0213  0261          261
0214  6216          JMP .+2
0215  6227          JMP OP1
0216  1460          SAE I
0217  0262          262
0220  6222          JMP .+2
0221  6230          JMP OP2
0222  1460          SAE I
0223  0263          263
0224  6203          JMP .-21
0225  0600          LIF 0
0226  6102          JMP 102
0227  6203  OP1,    JMP 203
0230  1020  OP2,    LDA I
0231  0004          04
0232  0600          LIF 0
0233  6130          JMP 130
0234  0002  IN,     PDP
                    PMODE
4235  6036          KRB
4236  6031          KSF
4237  5236          JMP .-1
4240  6036          KRB
4241  6041          TSF
4242  5241          JMP .-1
4243  6046          TLS
4244  6141          LINC
                    LMODE
0245  6000          JMP 0
0246  1000  TRANS,  LDA
0247  0000          0
```

```
0250  4275          STC .+25
0251  1000          LDA
0252  0012          12
0253  0017          COM
0254  4017          STC 17
0255  1000          LDA
0256  0013          13
0257  1040          STA
0260  0276          TRAN1
0261  1120          ADA I
0262  7776          -1
0263  4013          STC 13
0264  0074          SET I 14
0265  1677          1677
0266  1033          LDA I 13
0267  1074          STA I 14
0270  0237          XSK I 17
0271  6266          JMP .-3
0272  1000          LDA
0273  0276          TRAN1
0274  4013          STC 13
0275  0016          NOP
0276  0016  TRAN1,  NOP
0277  1000  ROUT1,  LDA
0300  0000          0000
0301  4310          STC .+7
0302  0500          IOB
0303  6032          6032
0304  6441          JMP DSPP
0305  0500          IOB
0306  6031          6031
0307  6304          JMP .-3
0310  0016          NOP
0311  1000  ROUT2,  LDA
0312  0000          0000
0313  4325          STC END
0314  0500          IOB
0315  6032          6032
0316  6441          JMP DSPP
0317  6530          JMP LINE
0320  0500          IOB
0321  6031          6031
0322  6316          JMP .-4
0323  0500          IOB
0324  6032          6032
0325  0016  END,    NOP
0326  1000  ROUT3,  LDA
0327  0000          0000
0330  4325          STC END
0331  0500          IOB
0332  6032          6032
```

```
0333   6441            JMP DSPP
0334   6530            JMP LINE
0335   1000            LDA
0336   0015            15
0337   4353            STC .+14
0340   1000            LDA
0341   0016            16
0342   4356            STC.+14
0343   1000            LDA
0344   0005            5
0345   4015            STC 15
0346   1000            LDA
0347   0006            6
0350   4016            STC 16
0351   6530            JMP LINE
0352   1020            LDA I
0353   0016            NOP
0354   4015            STC 15
0355   1020            LDA I
0356   0016            NOP
0357   4016            STC 16
0360   6441            JMP DSPP
0361   6530            JMP LINE
0362   0500            IOB
0363   6031            6031
0364   6333            JMP .-31
0365   0500            IOB
0366   6032            6032
0367   6325            JMP END
0370   1000   PRNT,    LDA
0371   0000            0000
0372   1060            STA I
0373   0000            0000
0374   0011            CLR
0375   0500            IOB
0376   6044            6044
0377   1331   A1,      LDH I 11
0400   1460            SAE I
0401   0044            44
0402   6404            JMP .+2
0403   6373            JMP PRNT+3
0404   1460            SAE I
0405   0043            43
0406   6410            JMP .+2
0407   6433            JMP A4
0410   1460            SAE I
0411   0047            47
0412   6414            JMP .+2
0413   6436            JMP A5
0414   1120            ADA I
0415   7740            -37
```

```
0416  0451           APO
0417  6421           JMP .+2
0420  6423           JMP .+3
0421  1120           ADA I
0422  0100           0100
0423  1120           ADA I
0424  0237           0237
0425  0002   A3,     PDP
                     PMODE
4426  6041           TSF
4427  5226           JMP .-1
4430  6046           TLS
4431  6141           LINC
                     LMODE
0432  6377           JMP A1
0433  1020   A4,     LDA I
0434  0215           215
0435  6425           JMP A3
0436  1020   A5,     LDA I
0437  0212           212
0440  6425           JMP A3
0441  1000   DSPP,   LDA
0442  0000           0000
0443  1060           STA I
0444  0000           0000
0445  1000           LDA
0446  0013           13
0447  4501           STC SH1
0450  0073           SET I 13
0451  1677           1677
0452  1000           LDA
0453  0012           12
0454  0017           COM
0455  4007           STC 7
0456  1333           LDH I 13
0457  1333           LDH I 13
0460  1260           MUL I
0461  0040           40
0462  1120           ADA I
0463  0010           10
0464  4002           STC 2
0465  1333           LDH I 13
0466  1260           MUL I
0467  0040           40
0470  1120           ADA I
0471  7410           -367
0472  0142           DIS 2
0473  0227           XSK I 7
0474  6457           JMP .-15
0475  1000           LDA
0476  0501           SH1
```

```
0477  4013         STC 0013
0500  6444         JMP DSPP+3
0501  0000  SH1,   0000
0502  0011         CLR
0503  0067         SET I 7
0504  7400         -377
0505  0500         IOB
0506  6032         6032
0507  0000         HLT
0510  0516         RSW
0511  4015         STC 15
0512  0000         HLT
0513  0517         LSW
0514  4016         STC 16
0515  6530         JMP LINE
0516  0500         IOB
0517  6031         6031
0520  6515         JMP .-3
0521  0500         IOB
0522  6036         6036
0523  1460         SAE I
0524  0301         301
0525  6515         JMP .-10
0526  6507         JMP .-17
0527  0000         HLT
0530  1000  LINE,  LDA
0531  0000         0000
0532  1060         STA I
0533  0000         0000
0534  0062         SET I 2
0535  0007         7
0536  0063         SET I 3
0537  7000         -777
0540  1000         LDA
0541  0016         16
0542  0451         APO
0543  0017         COM
0544  0364         SCR I 4
0545  1260         MUL I
0546  0042         42
0547  5003         STC SAVE2
0550  1000         LDA
0551  0016         16
0552  0451         APO
0553  0017         COM
0554  1560         BCL I
0555  7760         7760
0556  1260         MUL I
0557  0042         42
0560  1120         ADA I
0561  0010         10
```

```
0562  0344              SCR 4
0563  3003              ADD SAVE2
0564  1120              ADA I
0565  7411              -366
0566  1040              STA
0567  1003              SAVE2
0570  1040              STA
0571  1004              SAVE3
0572  1000              LDA
0573  0015              15
0574  1040              STA
0575  1002              SAVE1
0576  0451              APO
0577  6723              JMP NEG
0600  1000              LDA
0601  0016              16
0602  0471              APO I
0603  6640              JMP RETURN
0604  1000              LDA
0605  1003              SAVE2
0606  1120              ADA I
0607  0366              366
0610  0243              ROL 3
0611  5003              STC SAVE2
0612  1000              LDA
0613  0015              15
0614  0341              SCR 1
0615  0017              COM
0616  5004              STC SAVE3
0617  1000              LDA
0620  1003              SAVE2
0621  3004              ADD SAVE3
0622  1040              STA
0623  1003              SAVE2
0624  0451              APO
0625  6633              JMP .+6
0626  1020              LDA I
0627  0001              1
0630  1140              ADM
0631  0002              2
0632  6617              JMP .-13
0633  1020              LDA I
0634  7410              -367
0635  1040              STA
0636  1003              SAVE2
0637  5004              STC SAVE3
0640  0011    RETURN,   CLR
0641  3004              ADD SAVE3
0642  0162              DIS I 2
0643  1000              LDA
0644  0015              15
```

```
0645  1140           ADM
0646  1002           SAVE1
0647  0471           APO I             /NEW
0650  6666           JMP .+16
0651  1000           LDA
0652  0015           15
0653  0017           COM
0654  3002           ADD SAVE1
0655  1120           ADA I
0656  0010           0010
0657  0304           ROR 4
0660  1560           BCL I
0661  7400           7400
0662  3003           ADD SAVE2
0663  5003           STC SAVE2
0664  1000           LDA
0665  0015           15
0666  1040           STA
0667  1002           SAVE1             /END OF NEW
0670  1120           ADA I
0671  0010           10
0672  0304           ROR 4
0673  1560           BCL I
0674  7400           7400
0675  3003           ADD SAVE2
0676  1040           STA
0677  1004           SAVE3
0700  0451           APO
0701  6711           JMP .+10
0702  1120           ADA I
0703  7401           -376
0704  0063           SET I 3
0705  1776           1776
0706  0471           APO I
0707  6533           JMP LINE+3
0710  6714           JMP .+4
0711  0223           XSK I 3
0712  6714           JMP .+2
0713  6533           JMP LINE+3
0714  1000           LDA
0715  0002           2
0716  1120           ADA I
0717  7402           -775
0720  0450           AZE
0721  6640           JMP RETURN
0722  6533           JMP LINE+3
0723  0011     NEG,  CLR
0724  3004           ADD SAVE3
0725  0162           DIS I 2
0726  1000           LDA
0727  0015           15
```

0775	1303	0403	0403
0776	1304	0404	0404
0777	1305	0405	0405
1000	1306	0406	0406
1001	1307	0503	0503
1002	1310	0504	0504
1003	1311	0505	0505
1004	1312	0506	0506
1005	1313	0507	0507
1006	1314	0510	0508
1007	1315	0604	0604
1010	1316	0605	0605
1011	1317	0606	0606
1012	1320	0607	0607
1013	1321	0610	0608
1014	1322	0611	0609
1015	1323	0704	0704
1016	1324	0705	0705
1017	1325	0706	0706
1020	1326	0707	0707
1021	1327	1005	0805
1022	1330	0710	0708
1023	1331	0711	0709
1024	1332	0712	0712
1025	1333	1006	1006
1026	1334	1007	1007
1027	1335	1010	1008
1030	1336	1011	1009
1031	1337	1012	1012
1032	1340	1106	1106
1033	1341	1107	1107
1034	1342	1110	1110
1035	1343	1111	1111
1036	1344	1112	1112
1037	1345	1113	1113
1040	1346	1206	1206
1041	1347	1207	1207
1042	1350	1210	1210
1043	1351	1211	1211
1044	1352	1212	1212
1045	1353	1213	1213
1046	1354	1310	1310
1047	1355	1311	1311
1050	1356	1312	1312
1051	1357	1313	1313
1052	1360	1412	1412
1053	1361	1413	1413
1054	1362	1414	1414
1055			*1370
1056	1370	0202	0202
1057	1371	0204	0204

```
0730  1140           ADM
0731  1002           SAVE1
0732  0451           APO
0733  6751           JMP .+16
0734  1000           LDA
0735  0015           15
0736  0017           COM
0737  3002           ADD SAVE1
0740  1120           ADA I
0741  0007           0007
0742  0364           SCR I 4
0743  3003           ADD SAVE2
0744  5003           STC SAVE2
0745  1000           LDA
0746  0015           15
0747  1040           STA
0750  1002           SAVE1
0751  1120           ADA I
0752  7767           -10
0753  0344           SCR 4
0754  3003           ADD SAVE2
0755  1040           STA
0756  1004           SAVE3
0757  0471           APO I
0760  6770           JMP .+10
0761  0063           SET I 3
0762  1777           1777
0763  1120           ADA I
0764  0376           376
0765  0451           APO
0766  6533           JMP LINE+3
0767  6773           JMP .+4
0770  0223           XSK I 3
0771  6773           JMP .+2
0772  6533           JMP LINE+3
0773  1000           LDA
0774  0002           2
0775  1120           ADA I
0776  7002           -775
0777  0450           AZE
1000  6723           JMP NEG
1001  6533           JMP LINE+3
1002  0000  SAVE1,   0000
1003  0000  SAVE2,   0000
1004  0000  SAVE3,   0000


                     *1300
1300  0302           0302
1301  0303           0303
1302  0304           0304
```

1060	1372	0205	0205
1061	1373	0210	0208
1062	1374	0303	0303
1063	1375	0403	0403
1064	1376	0406	0406
1065	1377	0410	0408
1066	1400	0412	0412
1067	1401	0503	0503
1070	1402	0505	0505
1071	1403	0507	0507
1072	1404	0604	0604
1073	1405	0611	0609
1074	1406	0614	0614
1075	1407	0705	0705
1076	1410	0707	0707
1077	1411	0712	0712
1100	1412	1005	0805
1101	1413	1010	0808
1102	1414	1013	0813
1103	1415	1015	0815
1104	1416	1104	1104
1105	1417	1106	1106
1106	1420	1107	1107
1107	1421	1111	1111
1110	1422	1114	9114
1111	1423	1206	1206
1112	1424	1210	1210
1113	1425	1211	1211
1114	1426	1212	1212
1115	1427	1213	1213
1116	1430	1215	1215
1117	1431	1307	1307
1120	1432	1405	1405
1121	1433	1407	1407
1122	1434	1411	1409
1123	1435	1413	1413
1124	1436	1415	1415
1125	1437	1513	1513
1126			*1460
1127	1460	0205	0205
1130	1461	0207	0207
1131	1462	0211	0211
1132	1463	0214	0214
1133	1464	0304	0304
1134	1465	0311	0311
1135	1466	0405	0405
1136	1467	0407	0407
1137	1470	0410	0410
1140	1471	0412	0412
1141	1472	0413	0413
1142	1473	0415	0415

1143	1474	Ø5Ø3	Ø5Ø3
1144	1475	Ø5Ø4	Ø5Ø4
1145	1476	Ø5Ø6	Ø5Ø6
1146	1477	Ø6Ø2	Ø6Ø2
1147	15ØØ	Ø612	Ø612
115Ø	15Ø1	Ø615	Ø615
1151	15Ø2	Ø7Ø4	Ø7Ø4
1152	15Ø3	Ø7Ø6	Ø7Ø6
1153	15Ø4	Ø71Ø	Ø71Ø
1154	15Ø5	Ø712	Ø712
1155	15Ø6	Ø714	Ø714
1156	15Ø7	1ØØ5	1ØØ5
1157	151Ø	1ØØ6	1ØØ6
116Ø	1511	1Ø11	1Ø11
1161	1512	11Ø4	11Ø4
1162	1513	11Ø6	11Ø6
1163	1514	11Ø7	11Ø7
1164	1515	1112	1112
1165	1516	1115	1115
1166	1517	12Ø5	12Ø5
1167	152Ø	1213	1213
117Ø	1521	13Ø7	13Ø7
1171	1522	131Ø	131Ø
1172	1523	1311	1311
1173	1524	1312	1312
1174	1525	1314	1314
1175	1526	14Ø4	14Ø4
1176	1527	14Ø6	14Ø6
1177	153Ø	141Ø	141Ø
12ØØ	1531	1413	1413
12Ø1	1532	1415	1415
12Ø2			*154Ø
12Ø3	154Ø	Ø213	Ø213
12Ø4	1541	Ø214	Ø214
12Ø5	1542	Ø216	Ø216
12Ø6	1543	Ø311	Ø311
12Ø7	1544	Ø314	Ø314
121Ø	1545	Ø315	Ø315
1211	1546	Ø41Ø	Ø41Ø
1212	1547	Ø413	Ø413
1213	155Ø	Ø414	Ø414
1214	1551	Ø5Ø5	Ø5Ø5
1215	1552	Ø5Ø7	Ø5Ø7
1216	1553	Ø51Ø	Ø51Ø
1217	1554	Ø512	Ø512
122Ø	1555	Ø6Ø7	Ø6Ø7
1221	1556	Ø611	Ø611
1222	1557	Ø614	Ø614
1223	156Ø	Ø7Ø5	Ø7Ø5
1224	1561	Ø7Ø6	Ø7Ø6
1225	1562	Ø71Ø	Ø71Ø

```
1226        1563  0711               0711
1227        1564  0713               0713
1230        1565  1004               1004
1231        1566  1006               1006
1232        1567  1007               1007
1233        1570  1010               1010
1234        1571  1011               1011
1235        1572  1012               1012
1236        1573  1105               1105
1237        1574  1106               1106
1240        1575  1110               1110
1241        1576  1113               1113
1242        1577  1204               1204
1243        1600  1207               1207
1244        1601  1211               1211
1245        1602  1212               1212
1246        1603  1303               1303
1247        1604  1305               1305
1250        1605  1306               1306
1251        1606  1307               1307
1252        1607  1310               1310
1253        1610  1311               1311
1254        1611  1313               1313
1255        1612  1404               1404
1256        1613  1406               1406
1257        1614  1407               1407
1260        1615  1411               1411
1261        1616  1503               1503
1262        1617  1504               1504
1263        1620  1505               1505
1264        1621  1507               1507
1265        1622  1603               1603
1266        1623  1605               1605
1267                                 *1640
1270        1640  0204               0204
1271        1641  0405               0405
1272        1642  0606               0606
1273        1643  1007               1007
1274        1644  1210               1210
1275        1645  1411               1411
1276        1646  1612               1612
1277
1300
1301                                 SEGMNT 3
1302                                 *2000
1303                         U41,    TEXT Z
1304        0000  4043
1304
1305        0001  4740
1305        0002  4347
1305
```

```
1306        0003  4043
1306
1307        0004  4740
1307        0005  4347
1307
1310        0006  4043
1310
1311        0007  4740
1311        0010  4347
1311
1312        0011  4043
1312        0012  4740
1312        0013  4040
1312        0014  4040
1312        0015  4040
1312        0016  4040
1312        0017  4025
1312        0020  1611
1312        0021  2440
1312        0022  6440
1312        0023  4040
1312        0024  0317
1312        0025  2222
1312        0026  0514
1312        0027  0124
1312        0030  1117
1312        0031  1640
1312        0032  0116
1312        0033  0440
1312        0034  2205
1312        0035  0722
1312        0036  0523
1312        0037  2311
1312                                      UNIT 4   CORRELATION
                  AND REGRESSION
1313        0040  1716
1313        0041  4347
1313
1314        0042  4043
1314
1315        0043  4740
1315        0044  4347
1315        0045  0114
1315        0046  2410
1315        0047  1725
1315        0050  0710
1315        0051  4024
1315        0052  1005
1315        0053  4023
1315        0054  0301
1315        0055  2424
```

1315	0056	0522	
1315	0057	4020	
1315	0060	1417	
1315	0061	2440	
1315	0062	1706	
1315	0063	4024	
1315	0064	1005	
1315	0065	4020	
1315	0066	1711	
1315	0067	1624	
1315	0070	2340	
1315	0071	0711	
1315	0072	2605	
1315	0073	2340	
1315	0074	2523	
1315	0075	4023	
1315	0076	1715	
1315	0077	0540	
1315	0100	0605	
1315	0101	0514	
1315	0102	4006	
1315			ALTHOUGH THE SCATTER PLOT OF THE POINTS GIVES US SOME FEEL FOR
1316	0103	1722	
1316	0104	4347	
1316			
1317	0105	4043	
1317	0106	4724	
1317	0107	1005	
1317	0110	4003	
1317	0111	1722	
1317	0112	2205	
1317	0113	1401	
1317	0114	2411	
1317	0115	1716	
1317	0116	4014	
1317	0117	1717	
1317	0120	1311	
1317	0121	1607	
1317	0122	4001	
1317	0123	2440	
1317	0124	2410	
1317	0125	0540	
1317	0126	2205	
1317	0127	0722	
1317	0130	0523	
1317	0131	2311	
1317	0132	1716	
1317	0133	4014	
1317	0134	1116	
1317	0135	0523	

```
1317      0136  4040
1317      0137  0617
1317      0140  2215
1317      0141  0504
1317      0142  4002
1317                          THE CORRELATION LOOKING AT THE
              REGRESSION LINES  FORMED BY
1320      0143  3140
1320
1321      0144  4340
1321      0145  4322
1321      0146  0507
1321      0147  2205
1321      0150  2323
1321      0151  1116
1321      0152  0740
1321      0153  3140
1321      0154  1716
1321      0155  4030
1321      0156  4001
1321      0157  1604
1321      0160  4030
1321      0161  4017
1321      0162  1640
1321      0163  3140
1321      0164  1501
1321      0165  3140
1321      0166  0205
1321      0167  4010
1321      0170  0514
1321      0171  2006
1321      0172  2514
1321      0173  4011
1321      0174  1640
1321      0175  0405
1321      0176  2011
1321      0177  0324
1321      0200  1116
1321      0201  0740
1321      0202  2410
1321      0203  1123
1321      0204  4014
1321      0205  1116
1321      0206  0501
1321      0207  2240
1321                     REGRESSING Y ON X AND X ON Y MAY BE HELPFUL
                IN DEPICTING THIS LINEAR
1322      0210  4043
1322
1323      0211  4043
1323      0212  2205
```

```
1323        0213  1401
1323        0214  2411
1323        0215  1716
1323        0216  2310
1323        0217  1120
1323                     RELATIONSHIP.
1324        0220  5643
1324
1325        0221  4740
1325        0222  4347
1325        0223  1417
1325        0224  1713
1325        0225  4001
1325        0226  2440
1325        0227  2410
1325        0230  1123
1325        0231  4005
1325        0232  3001
1325        0233  1520
1325        0234  1405
1325                               LOOK AT THIS EXAMPLE.
1326        0235  5643
1326        0236  4744
1326                               $Z
1327                     U42,      TEXT Z
1330        0237  4043
1330
1331        0240  4740
1331        0241  4347
1331        0242  0611
1331        0243  2223
1331        0244  2440
1331        0245  2705
1331        0246  4027
1331        0247  1114
1331        0250  1440
1331        0251  0104
1331        0252  0440
1331        0253  2410
1331        0254  0540
1331        0255  2205
1331        0256  0722
1331        0257  0523
1331        0260  2311
1331        0261  1716
1331        0262  4014
1331        0263  1116
1331        0264  0540
1331        0265  0617
1331        0266  2516
1331        0267  0440
```

```
1331        0270  0231
1331        0271  4022
1331        0272  0507
1331        0273  2205
1331        0274  2323
1331        0275  1116
1331                               FIRST WE WILL ADD THE REGRESSION
                  LINE FOUND BY REGRESSING
1332        0276  0740                                -
1332        0277  4347
1332
1333        0300  4043
1333        0301  4731
1333        0302  4017
1333        0303  1640
1333                               Y ON X.
1334        0304  3056
1334        0305  4347
1334        0306  4400
1334                               $Z
```

XIII. APPENDIX E. REGRESSION AND CORRELATION

A. Simple Linear Regression and Correlation

One important concern of statistical methodology is the study of relationships among a set of variables. The simplest case of this sort would be the relationship between two variables. One way of expressing such a relation is to represent one variable as a linear function of the other. Usages of this approach would include prediction or causality. For purposes of discussion, denote the dependent or criterion variable as Y and the independent or prediction variable as X. Consider n values of (X, Y) pairs, denoted by (X_i, Y_i), $i = 1,2,\ldots,n$. The technique of simple linear regression, or the regression of Y on X, is designed to construct a line of the form

$$\hat{Y}_i = a+bX_i,$$

where

X_i = the ith value of X

b = the slope of the line

$\hat{Y}_i$ = the point on the line at $X = X_i$

a = the Y intercept, or the value of Y_i at $X_i = 0$,

such that this line best fits the n pairs of observations. The method of least squares is one technique for obtaining such a line. Define the deviation of a Y_i value from the Y_i value at $X = X_i$ as

$$d_i = Y_i - \hat{Y}_i.$$

The least squares method defines the values of a and b such that the sum of squares of the d_i values is a minimum. Mathematically, we have

$$Q = \sum_{i=1}^{n} d_i^{\,2} = \sum_{i=1}^{n} (Y_i - a-bX_i)^2$$

$$\frac{\partial Q}{\partial a} = -2\sum_{i=1}^{n} (Y_i - a-bX_i)$$

$$\frac{\partial Q}{\partial b} = -2\sum_{i=1}^{n} X_i(Y_i - a-bX_i).$$

Setting these partial derivatives equal to zero and solving for a and b, we obtain as the least squares values for a and b the following:

$$b = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})}$$

$$a = \bar{Y} - b\bar{X}.$$

One may illustrate these concepts by use of an example. Consider the six (X_i, Y_i) pairs below:

Pair	X	Y
1	1	2
2	2	4
3	3	4
4	4	5
5	5	6
6	5	3

Computation of the least squares values for a and b yield the results

$$\hat{Y}_i = 2.24 + .53X_i.$$

Figure 2 shows the (X_i, Y_i) pairs and the resulting regression line. Also noted is the deviation, d_i, for the sixth pair,

$$d_6 = Y_6 - Y_6$$
$$= 3 - 4.89$$
$$= - 1.89.$$

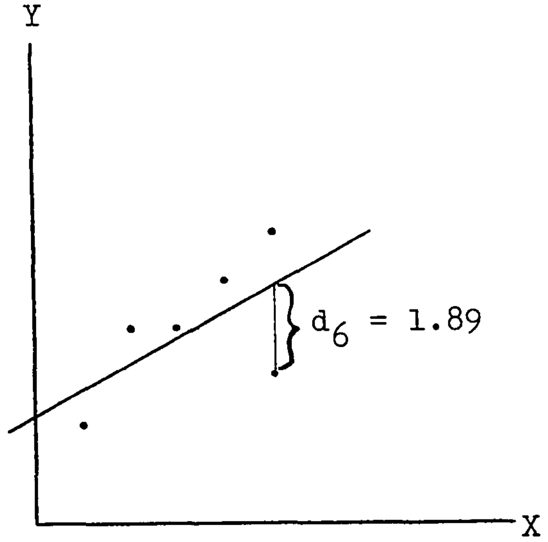

Figure 2. Regression and correlation example

Although one usually designates X as the independent variable and Y as the dependent variable, this need not be the case. The regression of X on Y can be defined and computed. For example, the application of least squares gives the values,

$$\hat{X}_i = 0.53 + 0.7Y_i.$$

The two regression lines are plotted in Figure 3.

One notes that the two regression lines do not coincide unless all the (X_i, Y_i) pairs form a perfect linear function, that is, all d_i's equal zero.

If the number of data points is small and one of the points is somewhat deviant, a considerable change in the regression line may result if that point is eliminated. Figure 4 shows the regression of Y on X with the sixth (X_i, Y_i) pair eliminated.

A method of assessing the joint relationship between two variables is by use of the product-moment correlation coefficient. In this situation both variables may be thought of as being neither independent nor dependent, and we are interested in measuring their covariation.

The product-moment correlation coefficient, r, is defined as

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}.$$

Although not proven here, r may range between the values - 1 and + 1 inclusively. A positive correlation of 1.00 would indicate a perfect linear relationship between the two variables in which both increase or decrease together. Similarly, a correlation of - 1.00 indicates a perfect linear relationship in which one variable increases as the other decreases. A correlation coefficient of zero indicates there is no

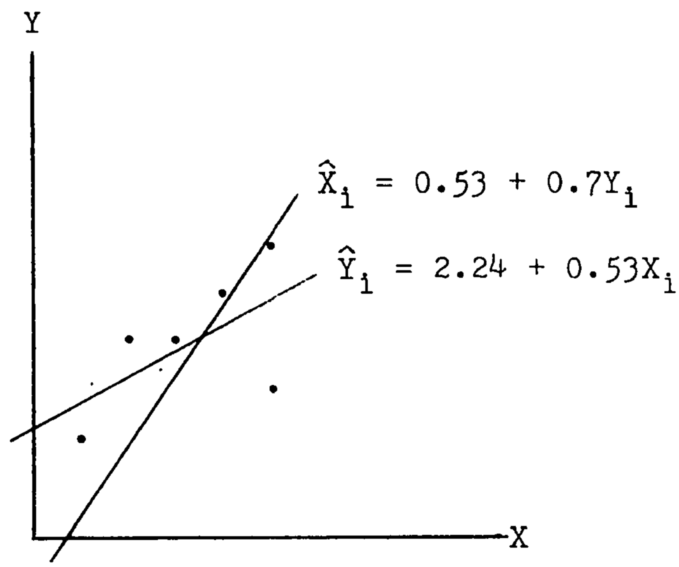

Figure 3. Regression of X on Y and Y on X

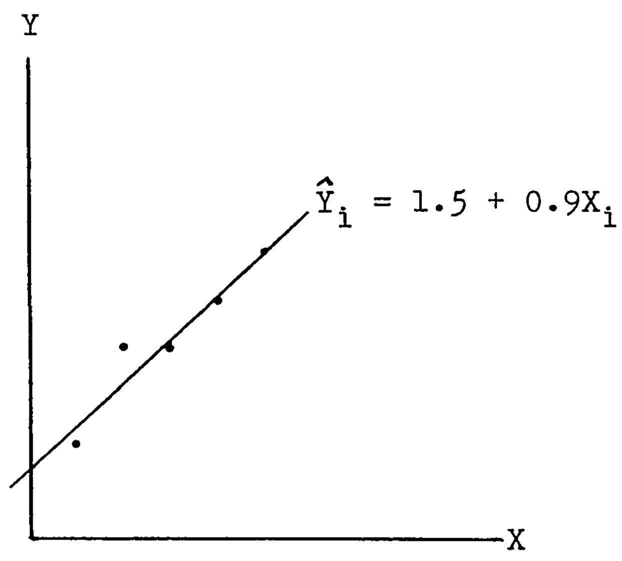

Figure 4. Example with deviant point eliminated

relationship between the two variables. Computing the value of r on the six (X_i, Y_i) pairs of the example data, r = 0.59.

The correlation coefficient can also be expressed as the geometric mean of the two regression coefficients, b_{yx} and b_{xy},

$$r = \sqrt{b_{yx} b_{xy}}.$$

As in regression analysis, if the number of points is small, the elimination of a somewhat deviant point will result in a significant change in the correlation coefficient. In Figure 4, where the sixth (X_i, Y_i) pair has been eliminated, the correlation coefficient is 0.96.