DOCUMENT RESUME

ED 190 111 IR 008 596

AUTHOR Stockburger, David W.
TITLE Teaching Statistical Concepts Using Microcomputer Simulations.
PUB DATE Apr 80
NOTE 32p.: Paper presented at the Southwestern Psychological Association Meeting (Oklahoma City, OK, April 1980). Best copy available.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction: Computer Oriented Programs: *Concept Teaching: Educational Technology: *Microcomputers: *Simulation: *Statistics; *Teaching Methods

ABSTRACT

This paper discusses the potential for the use of microcomputers as an aid in the teaching of statistics, and reviews five of the approaches that have been taken: (1) CAI question and answer dialogs: (2) statistics as a tool in the teaching of a computer language: (3) the computer as a computational tool: (4) computer generated tests and homework: and (5) the computer as simulator. The description of a microcomputer simulation program designed for an intermediate statistics course on research and design (ANOVA) includes an example of its application in a student exercise involving F-Ratios. Some guidelines on the purchase of a microcomputer system to be used in a statistics laboratory are provided. The paper concludes with a brief discussion of the obstacles to be overcome in the development of instructional systems, and the bibliography lists 24 references. (RAA)

TEACHING STATISTICAL CONCEPTS USING MICROCOMPUTER SIMULATIONS

David W. Stockburger

Southwest Missouri State University

A symposium paper presented at

Southwestern Psychological Association Meeting

April, 1980

Running Head: Microcomputer Simulations

TEACHING STATISTICAL CONCEPTS USING MICROCOMPUTER SIMULATIONS

David W. Stockburger

The microcomputer sill soon become standard equipment in a statistics laboratory because of its economy, power, and speed. Not only can statistical computations be performed on these machines, but computer simulations may illustrate statistical concepts. Strategies for developing programs which actively involve the student in stages of learning will be discussed. An example simulating experiments and the F-distribution will be presented.

With the advent of microcomputers the day of powerful, inexpensive computers predicted in the late 1960's (Cooley, 1969) are here. In this paper I will first review some of the various approaches that have been taken in applying computers to aid in the teaching of statistics, presenting an example of one approach in detail. I will then discuss the necessary equipment to implement a statistics laboratory. I will conclude with a discussion of the obstacles which must be overcome in the development of instructional systems.

During the late 1960's and early 1970's a number of individuals attempted to develop computer-assisted instructional systems to aid in the teaching of introductory statistics (Tubb, 1977). Before they began, most realized that it would be prohibitively expensive to implement the programs and systems that they developed on a large scale. Many of these applications were feasibility studies to judge the impact of this new technology on the teaching of statistics. It is worthwhile, therefore, to review some of the successes and failures of attempts to interface large and mini computers with statistics courses in order to better direct the development of microcomputers.

Attempts to integrate the computer into statistics courses may be placed into one of five categories:

I. CAI question-answer dialogs.

II. As a means of teaching a computer language.

III. As a computational tool.

IV. To generate statistics tests and homework.

V. As simulations.

These categories are not exclusive but are valuable as a reference point.

I. CAI QUESTION-ANSWER DIALOGS

Much of the early CAI work done in America was in the area of tutorials (Atkinson, 1969; Wassertheil, 1969; Rosenbaum, et. al., 1969). In an index to computer-assisted instruction (Lekan, 1970) under the twenty packages described with the heading of "Statistics" twelve were categorized as "tutorials".

In almost all cases where student evaluation of the systems was performed, the results were favorable to the computer method (Tubbs, 1977). By the middle and late 1970's, however, psychologists and statisticians had become disenchanted with this method of instruction (Elton, 1978). This disenchantment came from two sources:

1. It was too expensive.

The expense involved two aspects: the time it took to develop the materials and the student time on the machines. It takes a great deal of time and energy to create a dialog anticipating all possible thought patterns of introductory statistics students. This cost is relatively constant no matter what kind of hardware is used. These dialogs required extensive data bases and are at this time difficult to implement on a microcomputer. This may change in the fairly near future as witnessed by the availability of the PILOT language on the APPLE II tm microcomputer. If the cost of microcomputers continues to decrease, the monetary expense critism may also be unjustified.

2. The computer is not able to match its human counterpart for range and subtley of conversation.

Except for situations where drill-and-practice are appropriate, the computer is a relatively poor conversationalist, mindlessly repeating segments already covered and not quite realizing exactly what the student did not understand. A human teacher can do the job faster and better. As psychologists and computer scientists obtain a greater understanding of language, dialog, and learning processes, a more capable CAI may result. It could also be improved greatly by the addition of speech synthesis. Until that time other avenues of computer aided learning (CAL) will probably be more productive.

There are some CAI programs encompassing drill and practice which stand out as being successful. Anderson (1977) describes a system called CAPS (computer-assisted problem-solving system) which seems to fit this bill. One part of CAPS asks students to estimate parameters of distributions on the basis of a graphic display. For example, a scatterplot is illustrated on a CRT and the student must guess the correlation coeficient within some specified range. This drill continues until the student gets a certain number of estimates correct. This program was written for a PDP-10 minicomputer and would appear to be within the power range of a microcomputer.

II. STATISTICS AS A TOOL IN THE TEACHING OF A COMPUTER LANGUAGE.

Some authors have attempted to teach students a computer language such as FORTRAN (Lyczak, 1980) by having students program problems in statistics. While this approach may be valuable as a means to teach programming and perhaps review statistics at the same time, I doubt seriously whether it would work in an introductory statistics class. I speak from the personal experience of having a problem-solving class attempt to create simple algorithms in BASIC for a simulated rat to run a maze. It takes a great deal of class and instructer time to make students feel comfortable at a computer counsol with a low-level language.

Other authors have integrated a number of BASIC programs into an introductory statistics text (Price, 1979; Bulgren, 1971). These programs generally do statistical computations or simulations. Even with this approach I believe it would be difficult to teach enough about a low-level computer language and still have much time left to present statistical concepts.

III. THE COMPUTER AS A COMPUTATIONAL TOOL.

It is ironic that while most statisticians find that computers are a fact of life in the performance of their jobs, students are taught to use antiquated formulas designed for obsolete machines. One approact to integrate computers into the introductory statistics coursa is as a computational tool.

For the past six years I have taught my introductory statistics students how to use "canned" statistical packages to analyze data. The first three years I used MINITABS and the last three SPSS. I have found that while teaching SPSS takes slightly more class time (about three to five hours total), its general availability and ability to grow with the students is worth the additional teaching effort. I consider this to be an integral part of the course and generally get very favorable student reactions, as have others (Swanson, 1976).

The difficulty with this approach is that the university computer center has for the last few years become increasingly crowded. It would be ideal if the students could learn to use these canned statistics packages on a microcomputer. Unfortunately, a "canned" statistics package such a SPSS is not going to fit on a microcomputer. In order to most fully take advantage of the power of microcomputers in this area I can visualize three options:

1. The microcomputer is used as a terminal connected to a larger system which has the ability to compute statistics in an interactive mode. An example of this type of system would be conversational SPSS or CADA (Computer-assisted data analysis; Novick, Hamer, and Chen, 1979).

2. A small version of a larger statistical package resides in the microcomputer. When I teach SPSS to introductory students I present only a few file creation, file modification, and statistical commands such as BREAKDOWN, CROSSTABS, and PEARSON CORR. It would appear to be possible to put a subset of commands such as this on a small machine such that direct transfer of training would result when the student wished to run an analysis outside the power of the microcomputer.

The CADA system mentioned earlier is being modified to run on a microcomputer (personal communication, 1980). This interactive system is directed toward a Bayesian analysis with special attention given to the justification of the prior distribution. Miller (1979) and Anderson (1977) both describe a microcomputer system which computes statistics interactively with the student

3. The microcomputer writes programs for the students which are then run on a statistical package which resides in a larger machine. This hybred product would interact with the students in a dialog in order to create a program with the correct syntax. When completed the program would automatically be sent to the larger machine, run, and returned. This would free the student from JCL codes etc. and reduce the wasted time waiting for a program which has a small syntax error. I have not found a system such as this in my literature search.

IV. COMPUTER GENERATED TESTS AND HOMEWORK

A number of software systems described earlier have provisions to generate different homework and test problems for students Miller (1979; Anderson, 1977). A large project undertaken at the University of New Hampshire (Busbee, Merrill, and Warren, 1979) is collecting a data base of questions for introductory statistics tests and programs for editing and generating tests. The goal for this system is competency-based testing over a broad range of statistics courses taught in a variety of diciplines. The system is currently running on a PDP-10 and plans are underway to convert it to run on a microcomputer (Busbee, personal communication, 1980).

I have undertaken a project to write programs to generate both homework and the problem section of my tests over the last six months and have been pleased with the results. In these programs the instructer first selects the homework he or she wishes to print from a menu of possibilities which appear on the CRT screen (Figure 1). The BASIC programs then generate both problems and answers, an example of which appears in Figure 2. On the basis of this I can be sure that the student (or somebody!) has actually worked the problems instead of copying the answers from a neighbor or student from last semester. The answers allow the student to observe the problem worked correctly. Grading takes slightly longer, but I feel that it is worth the additional effort. It takes between one and two hours of computer time to generate and print 32 homeworks of four pages each, but this work is done when the computer would not otherwise be used.

V. THE COMPUTER AS SIMULATOR

This is the area where the computer has been most successfully utilized in aiding in the teaching of statistics. Elton and Waterford (1978) argue that computer simulations may teach concepts which are virtually impossible to teach in any other manner. Of the twenty program systems listed by LeKan (1970), five were classified as simulations.

Many of the fundemental comcepts in statistics are of the form "I wonder what would happen if ...?", especially if I sampled from a given population or performed the same experiment repeatedly. A number of programs exist to illustrate principles of the sampling distribution and hypothesis testing (Price, 1979; Lyczak, 1980; Diegart, 1974; Moore, 1973; Tanis, 1973; Rubner, Behr, and Baker, 1974). When a student review was performed on these programs they were generally favorable, with negative comments centering around the availability of a terminal (Anderson, 1977).

A system which deserves mention is one described by Beaujon (1970). Using an IBM 2250 terminal connected to a larger IBM 360 computer, he wrote a program which could display a number of different discrete and continuous probability functions. A random sample could be taken from these distributions and displayed as an overlay on the original distribution, illustrating limit functions. This program, written in Basic Assembler Language, occupied 30K with 8K reserved for the screen display (1024 x 1024). This is beyond the current limitations of microcomputers which typically have a screen of 48 x 128 and a much less powerful instruction set, but presents interesting possibilities.

Another approach to computer simulation is that of experiments, as characterized by the versions of EXPER SIM (MESS) and Cognitive Psychology (Brewley, 1978). An experiment is described to the students who then go to the computer and generate simulated results. This gives the student an opportunity to manipulate variables, collect data, and analyze it in much the same manner as a psychologist without the expense of actually running subjects. Snyder (1977) describes the successful application of a version of EXPER SIM in a course on research design.

I would now like to describe a simulation program I have written to assist in teaching an intermediate statistics course on research and design (ANOVA). This progam utilized aspects of both of the preceeding approaches to simulation. The program was written in BASIC on a POLYMORPHIC 88 which at that time had neither a disc nor a printer. This program is presented in Figure 3. The program could easily fit into 8K of memory plus BASIC and the operating system.

The program, called SMS (Score Model Simulation) was used in conjunction with Lee's (1976) text on experimental design and analysis. The score model for a single factor design

$$Xij = u + ai + eij$$

was first explained to the students. The symbols in the score model were given the following meaning:

Xij - score of subject j appearing in treatment i

u - population mean

ai - effect of treatment i

eij - error for subject ij

The students were then asked to do a number of simulations using the program.

After loading in both BASIC and SMS on the cassette the students were prompted to respond to a number of questions about the simulation being performed (Figure 4). The questions included the following:

1. The number of treatments (2-5)
2. The number of subjects in each treatment (2-12)
3. The relative size of the error term (1-10)
4. The distribution of the error term (uniform or approximate normal)
5. The size of each of the treatment effects
6. The number of simulations

Upon completion of this task the simulations began. Each simulation presented the students with raw scores, means and standard devitions for all groups, mean squares between and within, dfs, and the F-ratio (Figure 5). The student then plotted the distribution of F-ratios on a piece of graph paper as they appeared on the screen (Figure 6). Depending upon the type of error distribution this could take between 15 minutes for the uniform and 45 minutes for the normal to plot 100 F-ratios. At the fastest speed it required that the student work rapidly while the slowest speed ran the risk of boring the student.

Each student was given the assignment of exploring what effect changing one or more of the model parameters had on the distribution of F-ratios. The first step in this process was to generate a distribution where the null hypothesis was true, namely that all treatment effects were zero (ai=0). This distribution was then compared with values abtained from a traditional table of F-ratios. This exercize had the effect of demonstrating the meaning of the F-ratio tables.

By selecting values of the treatment effects other than zero the students could observe the creation of a non-central F-distribution. An estimate of the power and probability of a type II error could then be estimated from these distributions.

The difficulty with this approach was that it required a good deal of the students' time. Observing two or three distributions unfold before their eyes was a valuable exercize, any more and it became drudgery. The solution to this problem came about with the aquisition of a dot-matrix printer (Integral Data 440). Now, after completing three different distributions, the students may enter the number of simulations they desire, the values for each simulation, and then go home. The computer works all night and the student picks up the results the next morning (Figure 7). Even more economically, the student may pick up a power analysis of each simulation the next morning without the distributions themselves.

I found this exercize to be extremely valuable in portraying the concepts behind the ANOVA model. It has a number of features which I believe contributed to its success:

1. It required student interaction with a display which changed quickly enough to keep up their interest.
2. After a student mastered the concept it was no longer necessary to continue repetitious activity.
3. Each student had something to contribute to the class when he or she was finished.

4. It presented the concepts in a way that would not be possible using a conventional text and lecture format. The idea of the score model and the assumptions underlying it were made clear to the student. Effects of the violation of the assumptions were also demonstrated.

The system was not without faults, however:

1. Using a tape recorder required a number of operations to be performed before the program could be run. The tape reader was also prone to make ereors. This has been remedied with the acquisition of of a disc drive which loads the program automatically after hitting RESET.
2. It takes a good deal of time on the computer. More than once I could not do development work on the machine because students were using it. Timesharing is a possibility, but this is some distance in the future.

On the basis of this experience I believe that a number of computer programs could be written for a microcomputer which would greatly enhance the teaching of statistics. Not all individuals are as enthuiastic. Tubb (1977) writes, "Very, very rarely, the computer may be used to generate pseudo-random data for subsequent analysis. Such data are useful for illustrating the properties of a particular technique; they do not illustrate many of the problems of less idealized situations." This may be true, but I believe the computer can simulate situations where the assumptions underlying the analysis of data are violated.

THE DESIGN OF A MICROCOMPUTER SYSTEM FOR A STATISTICS LABORATORY

Rather than extol the virtues of one microcomputer manufacturer, processer, bus structure, etc. I would like to present some guidelines on the purchase of a microcomputer system to be used in a statistics lab.

I. A Minimum System

In order for the computer to be effectively used in the statistics laboratory I feel that the following equipment is necessary:

A. Central Processor, Keyboard, and CRT

The capacity for upper and lower case will be appreciated by the user as will a color display.

B. Memory

A minimum of 32k is necessary to implement many of the computer-assisted learning systems that were previously mentioned.

C. Disc Drive

A cassette tape drive only leads to frustration and wasted time, although it may prove useful as a back-up system.

D. Serial Port and Modem

These are necessary to connect to the outside world.

The total cost of such a system would probabily cost between $2000 and $3500 on today's market. One major consideration in deciding to purchase a given machine is the willingness of the vendor to support the application you have in mind.

II. An Upgraded System

This minimal system will allow one user at a time to access the machine. This system will permit developement and testing of software, but has serious limitations. I would suggest the pruchase of the following items to upgrade the minimal system:

A. Printer

Hard copy is an absolute must if any kind of serious work is to be done on the microcomputer. It is also necessary for the generation of tests and homework.

B. More Memory

It seems as if too much memory is impossible.

C. A Second, possibly Third Disc Drive

In this case the main drive may contain the operating system and language overhead while the other drives contain specific student information, overlays, etc.

D. Terminals

These may be minimal system microcomputers themselves. The main computer keeps records, stores programs, etc, while the students actually work at the terminals.

The cost for all this hardware may seem extravagent. One must remember, however, the cost of a single Freiden or Monroe calculator just 15 years ago. Miller (1979) argues that it would cost more than half a million dollars to implement 38 terminals on an IBM 360/158 using CMS, excluding the cost of the terminals themselves. This cost is far greater than the cost of 38 mirocomputer systems. This is not done without sacrifice, however. The sacrifices include speed, software, resale value, and high-speed printing. In most situations involving computer assisted learning, these disadvantages will not cause a great deal of concern.

OBSTACLES TO BE OVERCOME

I. Expense

This can be calculated in terms of the monetary cost of purchasing hardware and software and in the time it takes to get it running correctly. The monetary expense may be justified by the argument that this technique provides a learning experience that can be acheived in no other manner. The time expense is more difficult to justify. There are very few rewards, financial or otherwise, for developing CAI or CAL programs. Perhaps the time has come to reevaluate performance criteria for faculty members.

II. Transportability

This is perhaps the greatest obstacle to be overcome. For example, most BASICs share elements in common (Isaacs, 1976), but input/output, disc storage, etc. are often machine or operating system dependent. Greater standardization is clearly needed.

III. Politics

Statistics courses are taught in any number of diciplines: math, business, psychology, sociology, economics, etc. Each structures the course around its own dicipline and emphasises certain concepts. Basic elements, however, are shared. It appears that very little information is shared across dicipline boundaries, resulting in many different attempts to reinvent the wheel. The effort seen in Busbee, et. al. (1979) is a welcome exception.

IV. Communications

Even within a single dicipline many individuals are not aware of developments in this area. It is one thing to read a description of a system and another to actually have seen it work and use it. Many individuals will not purchase software until they have seen it work. Because of the transportability problem this is often difficult or impossible.

COMMUNICATION NETWORKS AS A MEANS OF OVERCOMING THE OBSTACLES

I did not adequately explain the need for a modem in the minimal system because I wanted to reserve it until now. The modem is used to connect computers using existing telephone lines. I believe that many of the obstacles may be reduced if not overcome by connecting computers through a communications network.

At some centralized location a facility for storage of programs and software systems is provided. An individual at another university could dial up the central facility, perhaps through an organization like EDUNET, tell it what kind of microcomputer or terminal he was using, and then actually use the system. If it was suitable for his own purposes he could buy the program through the service.

This would require a central facility which had accessability to and expertise in many different machines and which had an individual on their staff who understood the problems involved in the teaching of statistics. Perhaps an organization like CONDUIT may fit this bill.

CONCLUSION

The potential for the use of microcomputers as an aid in the teaching of statistics is just now being discovered. If the obstacles hindering this development may be overcome, it may someday have a profound effect on the teaching methods in all statistics courses.

REFERENCES

Anderson, T. et. al. Computer assisted problem solving in an introductory statistics course. Technical Report N. 56, Bolt, Beraned, and Newman, Inc., Center for the Study of Reading, Illinois University, Urbana, 1977.

Atkinson, R. C. Computerized instruction and the learning process. In COMPUTER-ASSISTED INSTRUCTION: A BOOK OF READINGS, ed. by R. C. Atkinson and H. A. Wilson, New York: Academic Press, 1969, 143-166.

BeauJon, H. An interactive, graphical display system for illustrating elementary properties of statistical distributions. Master's Thesis, University of North Carolina, 1970.

Busbee, A. D., Merrill, H. D., and Warren, J. A. Statistical test item collection system. Paper presented at the American Statistical Association Meeting, 1979.

Bulgren, William. A COMPUTER-ASSISTED APPROACH TO ELEMENTARY STATISTICS EXAMPLES AND PROBLEMS. Belmont, CA: Wadsworth, 1971.

Corley, W. W. Computer-assisted instruction in statistics. Technical Report, Office of Naval Research, 1969.

Grubb, Ralph E. Learner-controlled statistics. In COMPUTER-ASSISTED INSTRUCTION: A BOOK OF READINGS. ed by R. C. Atkinson and H. A. Wilson, New York: Academic Press, 1969, 167-173.

Diegert, C. Some experience with interactive computing in teaching introductory statistics. Paper presented at the meeting of The Shared Educational Computing, 1974

Elton, L. R. B. and Waterworth, G. M. Trends in CAL. in INTERACTIVE COMPUTER GRAPHICS IN SCIENCE TEACHING. ed by J. McKenzie, L. Elton, and R. Lewis, New York: Wiley, 1978.

Isaacs, G. L. AN UPDATE TO INTERDIALECT TRANSLATABILITY OF THE BASIC LANGUAGE, Iowa City, Iowa: CONDUIT, 1976.

LeKan, Helen A. INDEX TO COMPUTER-ASSISTED INSTRUCTION (2nd edition), Boston: Sterling Institute, 1970.

Lyczak, Richard. STATISTICS BY COMPUTER: ELEMENTARY FORTRAN AND BASIC PROGRAMMING FOR STATISTICAL ANALYSIS. New York: Duxbury, 1980.

Miller, William G. Microprocessors as an adjunct to statistics instruction. Paper presented at meeting of American Educational Research Association, 1978.

Novick, M. R., Homer, R. M., and Chen, J. J. The computer-assisted data analysis (CADA) monitor (1978). THE AMERICAN STATISTICIAN, 1979,4, 219-220.

Price, Janet. INTRODUCTORY PSYCHOLOGICAL STATISTICS: STUDENT MANUAL. Wentworth, NH: COMPress, Inc., 1979.

Rosenbaum, J. Computer-based instruction in statistical inferrence; final report. Washington D. C.: National Science Foundation, 1967.

Rubner, V., Behern, G., and Baker, F. B. Stat-concept: An interactive computer package supporting a first course in educational statistics. Paper presented at the meeting of the American Educational Research Association, 1974.

Snyder, Rita. Computer simulations in teaching psychology. Paper presented at the meeting of the American Educational Research Association, 1977.

Swanson, J., Riederer, S., and Weekly, H. Using OMNITAB to teach applied statistics. Washington D.C.: National Science Foundation, 1973.

Tanis, E. A. A computer laboratory for mathematical probability and statistics. Paper presented at The Conference on Computers in the Undergraduate Curricula, 1973.

Thomas, Warren. The development of a statistical experiment simulator. Washington D.C.: Office of Education, 1972.

Tubb, Gary. Current use of computers in the teaching of statistics. Paper presented at The Computer Science and Statistics meeting, 1977.

Wasserthiel, S. Computer assistance in statistics. IMPROVING COLLEGE AND UNIVERSITY TEACHING, 1969, 17, 264-266.

Figure 2

Example Homework - Student Version

Page 2

PATTY BREID
Psychology 200 Stockburger HOMEWORK 6
Spring, 1980

2. Two individuals, Harold and Maude, start a program of physical fitness. Each night for twenty nights they attempt to do as many sit-ups as they possibly can. The following is a record o

DAY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
HAROLD	9	11	14	13	12	12	15	14	16	17	15	16	19	18	17	20	18	20	18	21
MAUDE	11	10	11	9	9	12	14	11	16	17	18	17	18	14	22	17	20	22	25	24

(a.) Draw scatterplots of these data using some means to differentiate the two individuals.
(b.) Compute the best-fitting regression lines for both individuals. Draw them on the scatterplots.
for Harold Y' = _____ + _____ X
for Maude Y' = _____ + _____ X
(c.) Which model best fits the data? Why?
(d.) Discuss the MEANING of the differences between these parameters.

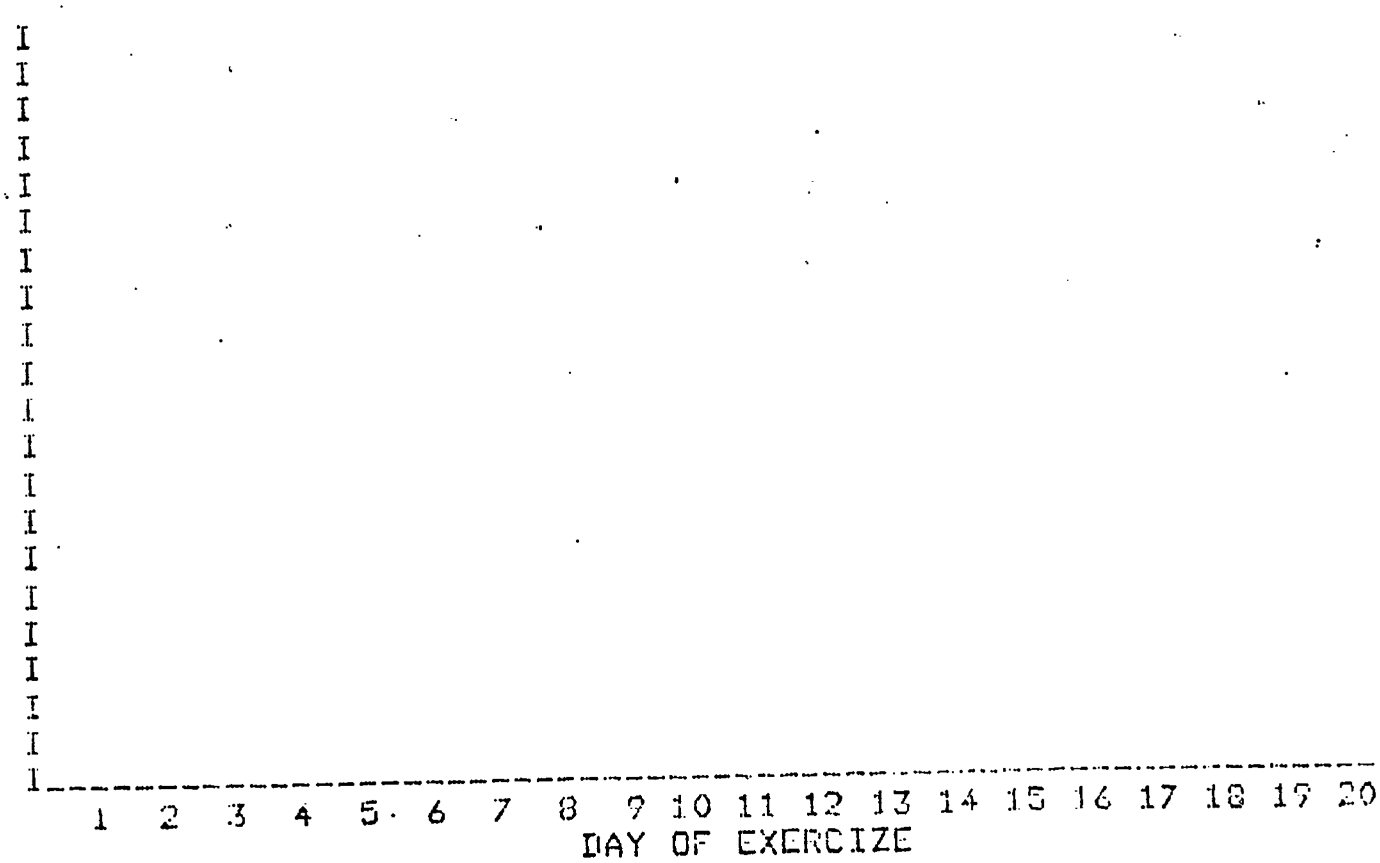


Figure 2 (cont)

Example Homework - Corrected Version

Page 2

PATTY BREID
Psychology 200 Stockburger HOMEWORK 6
Spring, 1980

2. Two individuals, Harold and Maude, start a program of physical fitness. Each night for twenty nights they attempt to do as many sit-ups as they possibly can. The following is a record o

DAY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
HAROLD	11	10	12	14	15	15	12	17	13	15	17	17	18	17	20	21	20	21	20	22
MAUDE	5	12	8	11	12	15	9	11	12	12	17	16	18	18	16	18	22	19	22	25

(a.) Draw scatterplots of these data using some means to differentiate the two individuals.

(b.) Compute the best-fitting regression lines for both individuals. Draw them on the scatterplots.

for Harold Y' = 10.373684 + .56917293 * X

for Maude Y' = 6.626316 + .78796992 * X

(c.) Which model best fits the data? Why?

Standard Error of Estimate

for Harold Sy.x = 3.8576532

for Maude Sy.x = 3.8215807

(d.) Discuss the MEANING of the differences between these parameters.

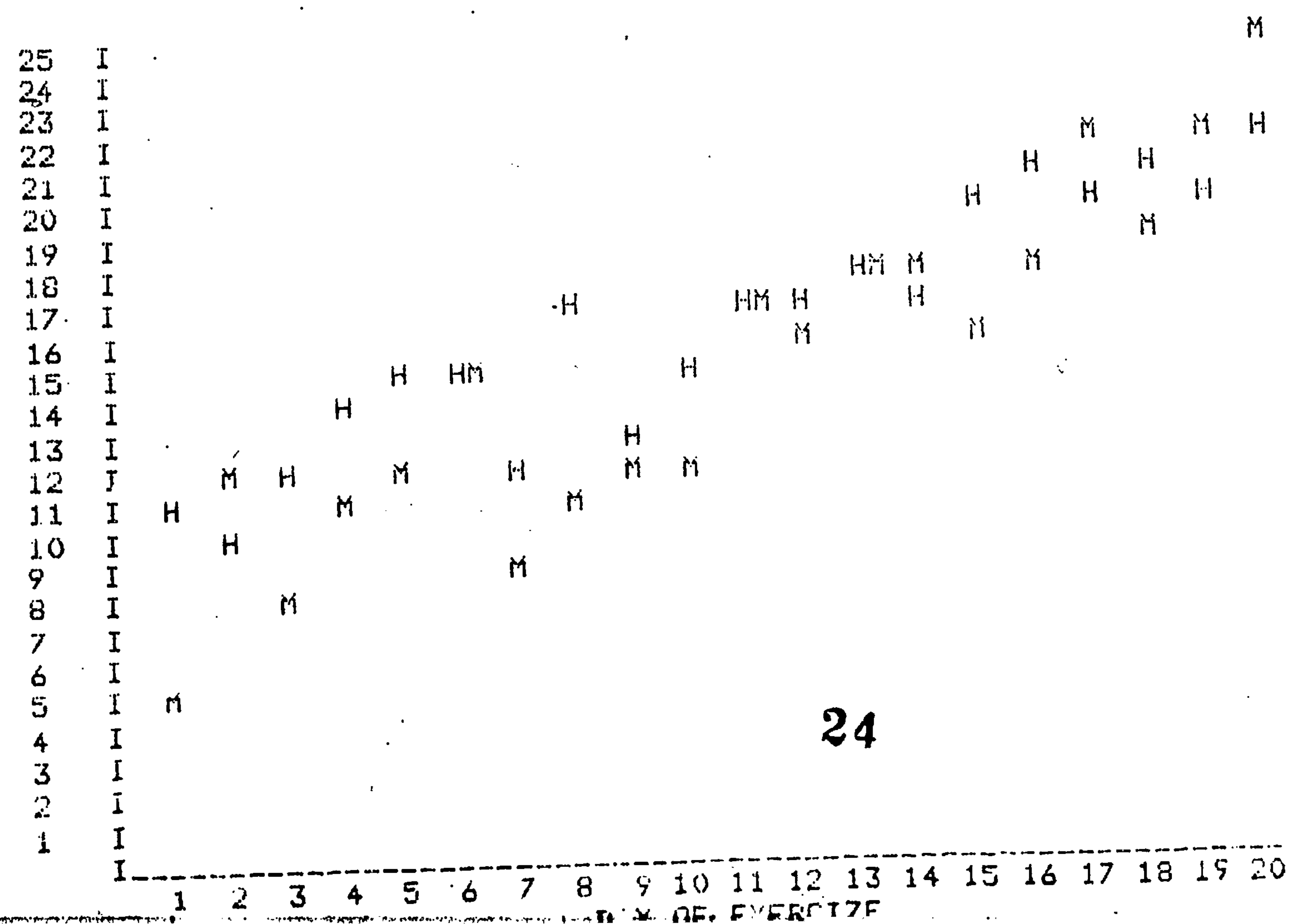


Figure 3

Program SMS - Score Model Simulation

```
10 REM BASIC PROGRAM TO SIMULATE ONE-WAY ANOVA EXPERIMENT
20 REM COPYRIGHT 1979  -  DAVID W. STOCKBURGER
30 REM                    PSYCHOLOGY DEPARTMENT
40 REM                    SOUTHWEST MISSOURI STATE UNIVERSITY
50 REM                    SPRINGFIELD, MISSOURI 65802
60 REM
70 REM T(I) IS THE SIZE OF THE TREATMENT EFFECT
80 REM X(I,J) IS THE RAW DATA MATRIX
90 REM X2(I) ARE THE MEANS
100 REM S2(I) ARE THE VARIANCES
110 FILE:2,LIST
120 REM
130 DIM T1(5,20)                     \REM TREATMENT EFFECTS
140 DIM P(6,20)                      \REM PARAMETERS
150 REM                                1=# OF GROUPS
160 REM                                2=ERROR VARIANCE
170 REM                                3=RANDOM NUMBERS
180 REM                                4=# OF SUBJECTS
190 REM                                5=DISTRIBUTION
200 REM                                6=# OF SIMULATIONS
210 DIM D$(1:10)                      \REM TEMPORARY STORAGE
220 DIM T(5),X(5,12),X2(5),S2(5)
230 DIM F1(500)                        \REM NEEDED FOR DISTRIBUTION
240 REM
250 REM INPUT RUN PARAMETERS
260 PRINT CHR$(12),\PLOT 0,47,0
270 INPUT "DO YOU WANT MULTIPLE SIMULATIONS (Y OR N)?  ",D$
280 IF D$="Y" THEN 1150
290 INPUT "NUMBER OF GROUPS (2 TO 5)   ",N1
300 IF N1>5 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 290
310 IF N1<2 THEN PRINT "** ERROR ** NUMBER TOO SMALL"\GOTO 290
320 INPUT "SIZE OF ERROR VARIANCE (1 TO 10)   ",N5
330 INPUT "START OF RANDOM NUMBERS (0 TO 1)   ",N4
340 IF N4>1 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 330
350 N4=RND(N4)
360 INPUT "NUMBER OF SUBJECTS IN EACH GROUP (2 TO 12)   ",N2
370 IF N2>12 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 360
380 INPUT "1 IF UNIFORM DISTRIBUTION, 2 IF APPROXIMATE NORMAL   ",N3
390 INPUT "NUMBER OF SIMULATIONS THIS RUN   ",N6
400 PRINT\PRINT "SIZE OF TREATMENT EFFECT IN EACH GROUP"
410 FOR I=1 TO N1
420 PRINT "SIZE OF TREATMENT EFFECT FOR GROUP ";I;
430 INPUT "    ",T(I)
440 NEXT
```

```
450 F[illegible]I=1 TO N[illegible]
460 P[illegible] CHR[illegible]\PLOT [illegible]47,0
470 PR[illegible] "CO[illegible] RAW SC[illegible]!!!"
480 I[illegible] T[illegible] 550
490 REM GENERATES DATA FROM A UNIFORM DISTRIBUTION
500 FOR I=1 TO N1\FOR J=1 TO N2
510 X(I,J)=RND(0)*N5*3.1
520 X(I,J)=X(I,J)+100+T(I)
530 NEXT\NEXT
540 GOTO 610
550 REM GENERATES DATA FROM AN APPROXIMATE NORMAL DISTRIBUTION
560 FOR I=1 TO N1\FOR J=1 TO N2\X1=0\FOR K=1 TO 10
570 X1=X1+RND(0)\NEXT\X1=X1/10
580 X(I,J)=X1*N5*10
590 X(I,J)=X(I,J)+100+T(I)
600 NEXT\NEXT
610 REM DISPLAY RAW DATA
620 PRINT CHR$(12),\PLOT 0,47,0
630 PRINT "RAW DATA                          GROUP"
640 PRINT "SUBJECT",
650 FOR I=1 TO N1
660 PRINT "  - ",%6I,I,
670 NEXT\PRINT
680 FOR J=1 TO N2\PRINT %6I,J,"   ",
690 FOR I=1 TO N1
700 PRINT "      ",%5F1,X(I,J),
710 NEXT\PRINT
720 NEXT\PRINT
730 REM NOW FIND MEANS AND VARIANCES
740 FOR I=1 TO N1\X2(I)=0
750 FOR J=1 TO N2\X2(I)=X2(I)+X(I,J)\NEXT\X2(I)=X2(I)/N2
760 NEXT
770 PRINT "MEAN     ",
780 FOR I=1 TO N1
790 PRINT "    ",%6F2,X2(I),
800 NEXT\PRINT
810 FOR I=1 TO N1\S2(I)=0\FOR J=1 TO N2
820 X1=X(I,J)-X2(I)
830 S2(I)=S2(I)+(X1*X1)\NEXT
840 S2(I)=S2(I)/(N2-1)
850 NEXT
860 PRINT "VARIANCE",\FOR I=1 TO N1
870 PRINT "   ",%7F1,S2(I),
880 NEXT\PRINT
890 REM NOW FIND MEAN SQUARES
900 REM FIND MEAN SQUARE BETWEEN
910 M1=0\FOR I=1 TO N1\M1=M1+X2(I)\NEXT\M1=M1/N1
920 PRINT "GRAND MEAN ",%6F2,M1,"    ",
930 M2=0\FOR I=1 TO N1\M2=M2+((X2(I)-M1)↑2)\NEXT
940 M2=N2*(M2/(N1-1))
950 PRINT "MS(BET) = ",M2,
960 REM FIND MEAN SQUARE WITHIN
970 M3=0\FOR I=1 TO N1\M3=M3+S2(I)\NEXT\M3=M3/N1
980 PRINT "   ",
990 PRINT "MS(WITH) = ",M3
1000 F=M2/M3
1010 PRINT "F(OBS) = ",F,"  DF(BET) = ",N1-1,"  DF(WITH) = ",N1*(N2-1)
1020 REM DELAY
1030 FOR I=1 TO 50\FOR J=1 TO 50\NEXT\NEXT
1040 NEXT
1050 REM DECISIONS ABOUT RERUNS
1060 PRINT "DO YOU WANT TO RUN ANOTHER SIMULATION?"
1070 PRINT "                         0 = NO"
1080 PRINT "                         1 = YES, NO CHANGES"
```

```
[illegible] PRINT "          2 - [illegible], NEW PARAMETERS"
1100 [illegible]
[illegible]
[illegible] =1 THEN 420
1130 IF [illegible]=2 THEN 29[illegible]
1140 GOTO 1060
1150 REM BEGINNING OF MULTIPLE SIMULATIONS
1160 INPUT "NUMBER OF DISTRIBUTIONS TO BE GENERATED (1-20)?  ",G1
1170 FOR G=1 TO G1
1180 FOR I=1 TO 500\F1(I)=0\NEXT
1190 PRINT CHR$(12),\PLOT 0,47,0\PRINT "*** SIMULATION # ",G," ***"
1200 INPUT "NUMBER OF GROUPS (2 TO 5)   ",P(1,G)
1210 IF P(1,G)>5 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 1200
1220 IF P(1,G)<2 THEN PRINT "** ERROR ** NUMBER TOO SMALL"\GOTO 1200
1230 INPUT "SIZE OF ERROR VARIANCE (1 TO 10)   ",P(2,G)
1240 INPUT "START OF RANDOM NUMBERS (0 TO 1)   ",P(3,G)
1250 IF P(3,G)>1 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 1240
1260 P(3,G)=RND(P(3,G))
1270 INPUT "NUMBER OF SUBJECTS IN EACH GROUP (2 TO 12)   ",P(4,G)
1280 IF P(4,G)>12 THEN PRINT "** ERROR ** NUMBER TOO LARGE"\GOTO 1270
1290 INPUT "1 IF UNIFORM DISTRIBUTION, 2 IF APPROXIMATE NORMAL   ",P(5,G)
1300 INPUT "NUMBER OF SIMULATIONS THIS RUN   ",P(6,G)
1310 PRINT\PRINT "SIZE OF TREATMENT EFFECT IN EACH GROUP"
1320 FOR I=1 TO P(1,G)
1330 PRINT "SIZE OF TREATMENT EFFECT FOR GROUP ",I
1340 INPUT "    ",T1(I,G)
1350 NEXT\NEXT
1360 FOR G=1 TO G1
1370 PRINT CHR$(12),\PLOT 0,47,0\PRINT "WORKING ON DISTRIBUTION  ",G
1380 FOR E=1 TO P(6,G)
1390 PLOT 0,44,0\PRINT "SIMULATION - ",E
1400 IF P(5,G)=2 THEN 1470
1410 REM GENERATES DATA FROM A UNIFORM DISTRIBUTION
1420 FOR I=1 TO P(1,G)\FOR J=1 TO P(4,G)
1430 X(I,J)=RND(0)*P(2,G)*3.1
1440 X(I,J)=X(I,J)+100+T1(I,G)
1450 NEXT\NEXT
1460 GOTO 1530
1470 REM GENERATES DATA FROM AN APPROXIMATE NORMAL DISTRIBUTION
1480 FOR I=1 TO P(1,G)\FOR J=1 TO P(4,G)\X1=0\FOR K=1 TO 10
1490 X1=X1+RND(0)\NEXT\X1=X1/10
1500 X(I,J)=X1*P(2,G)*10
1510 X(I,J)=X(I,J)+100+T1(I,G)
1520 NEXT\NEXT
1530 REM NOW FIND MEANS AND VARIANCES
1540 FOR I=1 TO P(1,G)\X2(I)=0
1550 FOR J=1 TO P(4,G)\X2(I)=X2(I)+X(I,J)\NEXT\X2(I)=X2(I)/P(4,G)
1560 NEXT
1570 FOR I=1 TO P(1,G)\S2(I)=0\FOR J=1 TO P(4,G)
1580 X1=X(I,J)-X2(I)
1590 S2(I)=S2(I)+(X1*X1)\NEXT
1600 S2(I)=S2(I)/(P(4,G)-1)
1610 NEXT
1620 REM NOW FIND MEAN SQUARES
1630 REM FIND MEAN SQUARE BETWEEN
1640 M1=0\FOR I=1 TO P(1,G)\M1=M1+X2(I)\NEXT\M1=M1/P(1,G)
1650 M2=0\FOR I=1 TO P(1,G)\M2=M2+((X2(I)-M1)↑2)\NEXT
1660 M2=P(4,G)*(M2/(P(1,G)-1))
1670 REM FIND MEAN SQUARE WITHIN
1680 M3=0\FOR I=1 TO P(1,G)\M3=M3+S2(I)\NEXT\M3=M3/P(1,G)
1690 F=M2/M3
1700 F=INT(F*5)\IF F>500 THEN F=500
1710 PLOT 0,41,0\PRINT F,
1720 IF F<1 THEN F=1
```

```
1730 F1(F)=F1(F) [illegible]
1740 PRINT F,F1(F)
1750 NEXT
1760 REM ROUTINE TO DRAW DISTRIBUTION
1770 PRINT:2,TAB(20),"F-DISTRIBUTION SIMULATION"
1780 PRINT:2,TAB(10),"Parameters for this simulation"
1790 PRINT:2,TAB(15),"Number of Groups = ",P(1,G),
1800 PRINT:2,TAB(46),"Number of Subjects/Group = ",P(4,G)
1810 PRINT:2,TAB(15),"Size of Error = ",P(2,G),
1820 PRINT:2,TAB(46),"Start of Random Numbers = ",P(3,G)
1830 PRINT:2,TAB(15),"Number of Simulations = ",P(6,G),
1840 IF P(5,G)=1 THEN PRINT:2,TAB(46),"UNIFORM ERROR DISTRIBUTION"
1850 IF P(5,G)=2 THEN PRINT:2,TAB(46),"NORMAL ERROR DISTRIBUTION"
1860 PRINT:2,TAB(15),"Size of Treatment Effects = ",
1870 FOR I=1 TO P(1,G)
1880 PRINT:2,T1(I,G),"    ",\NEXT\PRINT:2\PRINT:2
1890 PRINT:2,CHR$(31),TAB(20),
1900 FOR I=1 TO 50\PRINT:2,"+",\NEXT\PRINT:2
1910 R=0
1920 FOR J=1 TO 100
1930 R=R+.2
1940 PRINT:2,TAB(10),%5F2,R,TAB(20),"+",
1950 FOR I=1 TO F1(J)\PRINT:2,"*",\NEXT\PRINT:2
1960 NEXT
1970 PRINT:2,CHR$(29),CHR$(12),
1980 NEXT
1990 GOTO 290
```

Figure 4

Example SMS Setup - Appears on CRT Screen

```
DO YOU WANT MULTIPLE SIMULATIONS (Y OR N)?  N
NUMBER OF GROUPS (2 TO 5)   5
SIZE OF ERROR VARIANCE (1 TO 10)   6
START OF RANDOM NUMBERS (0 TO 1)   .8629
NUMBER OF SUBJECTS IN EACH GROUP (2 TO 12)   8
1 IF UNIFORM DISTRIBUTION, 2 IF APPROXIMATE NORMAL   1
NUMBER OF SIMULATIONS THIS RUN   100

SIZE OF TREATMENT EFFECT IN EACH GROUP
SIZE OF TREATMENT EFFECT FOR GROUP  1    0
SIZE OF TREATMENT EFFECT FOR GROUP  2    0
SIZE OF TREATMENT EFFECT FOR GROUP  3    0
SIZE OF TREATMENT EFFECT FOR GROUP  4    0
SIZE OF TREATMENT EFFECT FOR GROUP  5    0
```

Figure 5

Example SMS Output - Appears on CRT Screen

RAW DATA

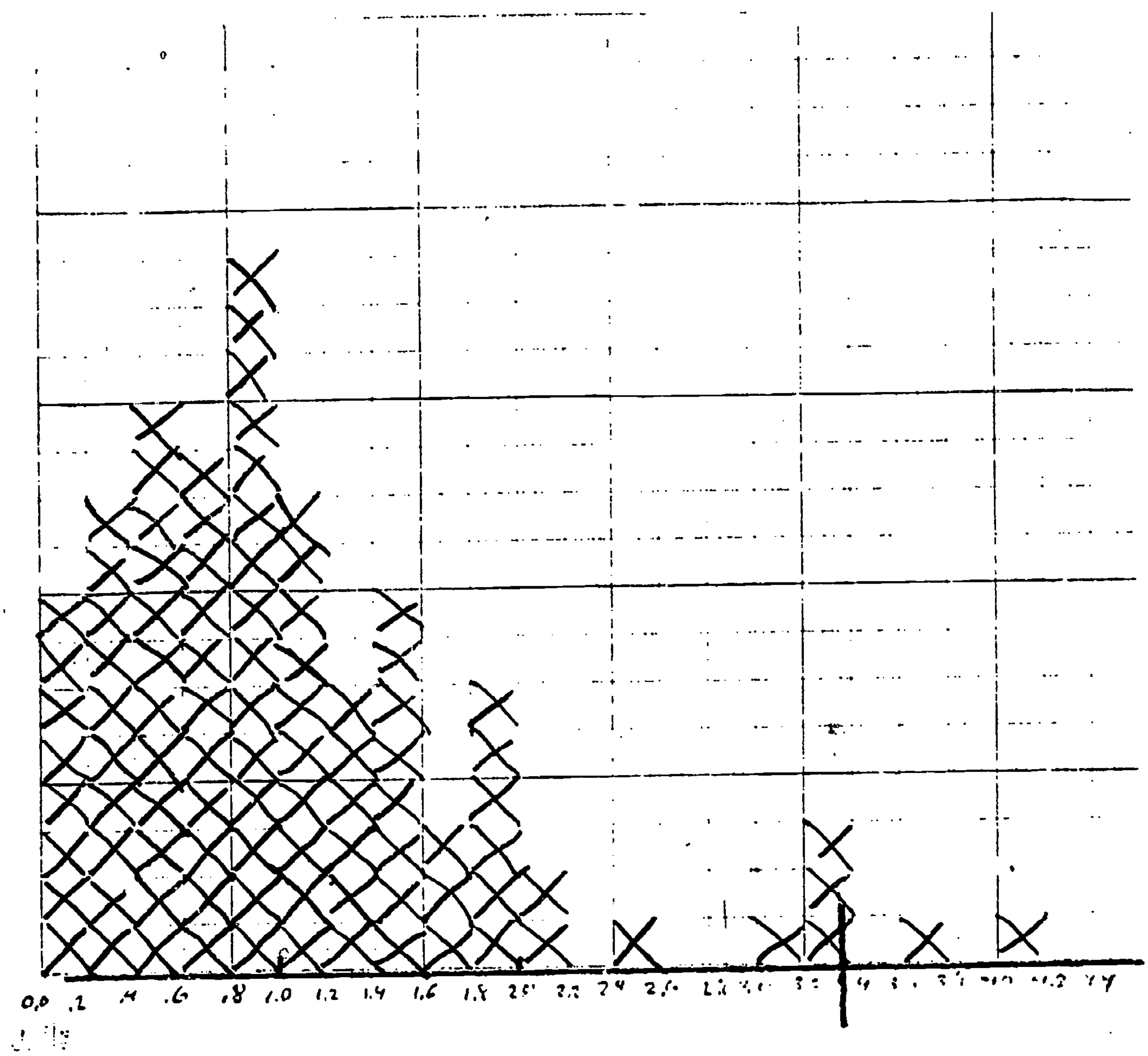
SUBJECT	GROUP 1	2	3	4	5
1	103.0	105.9	111.6	116.4	118.0
2	111.9	116.4	102.3	107.8	101.3
3	108.5	103.1	103.0	112.8	117.3
4	106.5	110.9	116.9	115.7	108.6
5	100.9	116.6	118.0	105.7	106.8
6	103.0	100.9	118.2	112.3	104.8
7	108.2	116.1	100.7	115.6	118.3
8	104.6	111.7	108.0	111.1	100.1
MEAN	105.83	110.19	109.84	112.16	109.38
VARIANCE	13.2	38.8	54.1	15.1	56.4

GRAND MEAN 109.48   MS(BET) = 42.331408   MS(WITH) = 35.52488

F(OBS) = 1.1915989   DF(BET) = 4   DF(WITH) = 35

Figure 6

F-distribution Drawn by Student



```
            F-DISTRIBUTION SIMULATION
   Parameters for this simulation
        Number of Groups =  5              Number of Subjects/Group =  12
        Size of Error =  1                 Start of Random Numbers =  .9978332

        Number of Simulations =  200   UNIFORM ERROR DISTRIBUTION
        Size of Treatment Effects =  0    0    0    0    0

       ++++++++++++++++++++++++++++++++++++++++++++++++++++
  .20  +****************************************
  .40  +*************************
  .60  +************************************
  .80  +**********************
 1.00  +**********************
 1.20  +**********
 1.40  +**************
 1.60  +**********
 1.80  +**********
 2.00  +****
 2.20  +***
 2.40  +***
 2.60  +*
 2.80  +*
 3.00  +
 3.20  +*
 3.40  +
 3.60  +***
 3.80  +
 4.00  +
 4.20  +**
 4.40  +
 4.60  +
 4.80  +
 5.00  +*
 5.20  +
 5.40  +
 5.60  +
 5.80  +
 6.00  +
```

Figure 7

F-distribution Drawn by SMS

