DOCUMENT RESUME

ED 079 993 EM 011 369

AUTHOR Edgar, S. Michael
TITLE Teaching Statistics, While Simultaneously Saving Time, Chalk, Etc...
INSTITUTION Texas Univ., Arlington. Dept. of Business Administration.
PUB DATE Jun 73
NOTE 9p.; Paper presented at the Conference on Computers in the Undergraduate Curricula (Claremont, California, June 18-20, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Calculation; *College Instruction; *Computer Assisted Instruction; Computer Programs; Higher Education; *Instructional Innovation; Man Machine Systems; Program Descriptions; *Statistics
IDENTIFIERS 360 APL Interactive Terminal

ABSTRACT

The interactive computer terminal provides an opportunity to teach statistics innovatively. A new computer-assisted instructional (CAI) program helps students by taking over the burden of computation. The instructor introduces a theoretical topic, describes its use, works examples, and has the students work a problem, doing the computation manually. Additional problems are then worked on the computer. The computer programs are written in conversational style to facilitate clear understanding and to reinforce for the students the steps followed. The program: 1) provides an introduction on methodology; 2) asks for data and shows how to enter them; 3) checks data accuracy and conducts calculations; and 4) gives results. Repetition is available and students control what they receive. All relevant portions of a topic are presented at one time, and the user can sample all aspects of the problem which interest him before arriving at the final solution. Students who use the CAI program save time, achieve better overall understanding of statistics, and learn to deal with statistical material from other areas. Their manual computational skills, however, are not well developed. (PB)

TEACHING STATISTICS, WHILE SIMULTANEOUSLY SAVING TIME, CHALK, ETC . . .

S. Michael Edgar
Department of Business Administration
University of Texas
Arlington, Texas 76010
(817) 273-3707

The availability of an interactive computer terminal in the classroom has provided a unique chance to experiment with a different method of teaching the traditional statistics curriculum. Typically, a topic is introduced both verbally and by illustrative example. Then the students solve prepared homework problems in hopes that this will reinforce the theory being presented. Finally, if time allows, variations on the originally proposed example are introduced. This further illustrates the solution methodology, and acquaints the student with the idea that once he has reached an answer, he has not necessarily solved the problem. Following up on the variations of the original example is usually too time consuming, so the results and implications are only alluded to and not explored.

This procedure presents several exasperating situations for the student and the instructor. For all but the simplest problems, many students encounter frustrating computational difficulties which effectively prevent them from learning from the example or from trying other prepared illustrations of the same topic. If they successfully solve two or three of the problems, many quit because the computational stress leads them to rationalize that they have learned "all they need to know about the topic." Thus, they frequently miss the nuances and subtlety of the theory which becomes evident by respecifying certain of the assumptions and/or modifying the data entry, etc. Most of the students end up lacking knowledge about "nonstandard" applications of the theory, and "learning" only what they can memorize out of the textbook.

Access to a 360/50 APL interactive terminal in the classroom has allowed me to continue with the traditional approach to teaching statistics, but freed my students and me from the drudgery of manually handling all of the numbers and mathematical manipulations. This is not to say, however, that we neglect the step-by-step process. On the contrary, it still appears necessary to force the student to "do one for himself" . . . or maybe two or more. But the time saved comes when he gains facility with the manual computations. Once this facility is achieved, there seems no reason to limit a student's investigative interest by continually forcing him to content with computational problems when the computer can do the same work more quickly and accurately. This is when the computer becomes valuable to the student.

From the instructor's point of view, the computer is useful almost from the time he introduces each new topic. Following a verbal introduction to a "new" theoretical topic, and some discussion of situations in which it may be used, the instructor frequently works through some simple examples of the applications of this theory. Then he usually has the class solve some problems on their own before allowing them to use the computer for their solutions. But this does not mean that the instructor is prevented from using the computer for illustrative purposes after initially demonstrating the computational methodology. In fact, its availability allows the instructor to change assumptions and/or data in the problem analysis, thereby illustrating the "sensitivity" of the methodology and the problem specification, and displaying the ability to improve upon the previous solution by making minor (or even major) changes in the previous problem set-up.

In the classroom, I frequently work through a complete example one time to illustrate the method of interpreting what is asked for or required, how to set up the data in an appropriate manner, and how to manipulate the computations to reach a conclusion. Then I go to the terminal and retrace the same steps, except that I enter the data as directed by the computer rather than in the "normal" sequence. The stored computer programs are written in conversational style for two reasons: (a) to explain to the student what is being asked for or what computations are currently being performed; and (b) to reinforce in his mind the sequence of steps to achieve a solution. Although this conversation becomes somewhat repetitious for some who are doing their third (or so) replication of a problem, it has proven helpful for others.

Typically a program begins with a short introduction of what that method will do or what the limitations are within the program. Then it asks for the data, showing how it is to be entered. After a data accuracy check, the calculation process begins. Frequently, some intermediate results are given; and occasionally, a plot of the results is available. A multinomial regression example follows.

```
        MULTREG

MULTIPLE REGRESSION SIMULATES A LINEAR RELATIONSHIP BETWEEN A
DEPENDENT VARIABLE (Y) AND TWO OR MORE INDEPENDENT VARIABLES
(X2, X3,...). THIS FUNCTION MAY BE ALTERED BY VARYING THE
NUMBER AND COMBINATION OF VARIABLES. THE STEPWISE OPTION
SELECTS THE INDEPENDENT VARIABLES IN ORDER OF THEIR IMPOR-
TANCE...AS DETERMINED BY THEIR RELATIVE CONTRIBUTION TO THE
UNEXPLAINED SUM OF SQUARES.

NUMBER OF VARIABLES?  MAXIMUM OF SIX (FIVE INDEPENDENT).
?:
      4
 ENTER THE NUMBER OF OBSERVATIONS...NO MAXIMUM NUMBER.
?:
      16
ENTER THE DATA AS FOLLOWS: ONE VALUE FOR Y AND ONE EACH FOR
X2, X3, ... SPACE BETWEEN EACH VALUE. E.G., IF Y=35, X2=17,
X3=45 AND X4=7; THE APPROPRIATE DATA ENTRY IS:  35 17 45 7
?:
      16 180 2800 90
?:
      14.9 195 3000 90
?:
      14.1 160 2700 80
?:
      18.5 235 3200 110
?:
      15.3 175 2700 100
?:
      21.9 285 3400 110
?:
      15.2 210 3000 100
?:
      18.2 230 3300 100
?:
      17 235 3500 100
?:
      16.8 295 2800 90
?:
      20 290 3500 110
?:
      14.8 150 2500 80
?:
      13 100 2100 80
?:
      20.3 290 3400 110
?:
      18.2 255 3400 100
?:
      17.1 220 2900 100
```

CHECK THE FORM AND ENTRIES OF THE DATA MATRIX.

16	180	2800	90
14.9	195	3000	90
14.1	160	2700	80
18.5	235	3200	110
15.3	175	2700	100
21.9	285	3400	110
15.2	210	3000	100
18.2	230	3300	100
17	235	3500	100
16.8	295	2800	90
20	290	3500	110
14.8	150	2500	80
13	100	2100	80
20.3	290	3400	110
18.2	255	3400	100
17.1	220	2900	100

IF THE DATA ARE NOT CORRECT, TYPE NO

VAR NUM	MEAN	STD DEV
1	16.95625	2.447302937
2	219.0625	56.95667213
3	3012.5	403.1128874
4	96.875	10.78192933

SIMPLE CORRELATION BETWEEN VARIABLES 1 AND 2 = 0.8682309786
SIMPLE CORRELATION BETWEEN VARIABLES 1 AND 3 = 0.8270492848
SIMPLE CORRELATION BETWEEN VARIABLES 1 AND 4 = 0.8585466126

DO YOU WANT THE OTHER SIMPLE CORRELATIONS?
YES
SIMPLE CORRELATION BETWEEN VARIABLES 2 AND 3 = 0.825168087
SIMPLE CORRELATION BETWEEN VARIABLES 2 AND 4 = 0.765683405
SIMPLE CORRELATION BETWEEN VARIABLES 3 AND 4 = 0.8225321512

DO YOU WANT THE PARTIAL CORRELATION COEFFICIENTS?
YES
ENTER THE SUBSCRIPTS FOR THE DESIRED PARTIAL CORRELATION
COEFFICIENT AS FOLLOWS...FOR R[12.3], ENTER 1 2 3
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
⎕:
1 2 3
THE PARTIAL CORRELATION COEFFICIENT IS 0.5850489633
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
⎕:
1 2 4
THE PARTIAL CORRELATION COEFFICIENT IS 0.6393442918
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
⎕:
1 3 2
THE PARTIAL CORRELATION COEFFICIENT IS 0.3946586358
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
⎕:

```
      1 3 4
THE PARTIAL CORRELATION COEFFICIENT IS 0.4144930271
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      1 4 2
THE PARTIAL CORRELATION COEFFICIENT IS 0.6071227825
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      1 4 3
THE PARTIAL CORRELATION COEFFICIENT IS 0.5576341297
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      1 2 3 4
THE PARTIAL CORRELATION COEFFICIENT IS 0.5432525143
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      1 3 2 4
THE PARTIAL CORRELATION COEFFICIENT IS 0.1123516056
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      1 4 2 3
THE PARTIAL CORRELATION COEFFICIENT IS 0.511420267
WHEN YOU WANT TO STOP, JUST TYPE A ZERO (0)
□:
      0
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT TO CONTINUE THE ANALYSIS?
YES
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ENTER THE SUBSCRIPTS FOR THE INDEPENDENT VARIABLES TO BE USED
IN THE ANALYSIS...SPACE BETWEEN EACH ENTRY.  E.G.,  2 3 4
□:
      2 3 4
ENTER THE SUBSCRIPT FOR THE DEPENDENT VARIABLE
□:
      1
 DO YOU WANT TO USE THE STEPWISE REGRESSION OPTION?
YES

              ******** STEPWISE REGRESSION ********

INTERCEPT IS 1.519743175

VAR
NUM COEFFICIENT     STD DEV        T VALUE
2  0.02002924437  0.008935704051  2.241484752
4  0.09666059443  0.04688579114   2.061618074
3  0.0005592880978  0.001427930927  0.3916772773

THE COEFFICIENT OF MULTIPLE CORRELATION IS 0.9200688524
THE STANDARD ERROR OF THE ESTIMATE IS 1.071912075
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT TO PERFORM THE F TEST?
YES
```

	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SUM OF SQUARES
EXPLAINED	76.05142904	3	25.35047635
UNEXPLAINED	89.839375	12	1.148995497

ENTER THE CRITICAL F VALUE FROM THE TABLE.
[]:
3.49
THE CALCULATED F VALUE IS 22.06316424
THE RELATIONSHIP IS STATISTICALLY SIGNIFICANT.

DO YOU WANT THE RESIDUALS AND THE DURBIN-WATSON?
YES.

OBSERVED Y	ESTIMATED Y	RESIDUAL
16	15.39046733	0.6095326658
14.9	15.80276362	-0.9027636194
14.1	13.96734769	0.1326523073
18.5	18.6490029	-0.1490029024
15.3	16.20099825	-0.9009982468
21.9	19.76232274	2.137677259
15.2	17.06980823	-1.869808229
18.2	17.63817955	0.561820454
17	17.85018339	-0.8501833875
16.8	17.69383044	-0.8938304371
20	19.91839777	0.08160222771
14.8	13.65519763	1.144802371
13	12.43002017	0.5699798284
20.3	19.86246896	0.4375310375
18.2	18.19483947	0.005160534834
17.1	17.21417186	-0.1141718632

THE DURBIN-WATSON D STATISTIC IS 1.858553438

DO YOU WANT ANY MORE REGRESSIONS WITH THIS DATA?
YES

ENTER THE SUBSCRIPTS FOR THE INDEPENDENT VARIABLES TO BE USED IN THE ANALYSIS...SPACE BETWEEN EACH ENTRY. E.G., 2 3 4
[]:
2 4
ENTER THE SUBSCRIPT FOR THE DEPENDENT VARIABLE
[]:
1
DO YOU WANT TO USE THE STEPWISE REGRESSION OPTION?
YES

******** STEPWISE REGRESSION ********

INTERCEPT IS 1.861343345

VAR NUM	COEFFICIENT	STD DEV	T VALUE
2	0.02189845023	0.007304464855	2.997954082
4	0.1062996377	0.03858660145	2.754832862

```
THE COEFFICIENT OF MULTIPLE CORRELATION IS 0.9190019864
THE STANDARD ERROR OF THE ESTIMATE IS 1.0364219
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT TO PERFORM THE F TEST?
YES

                SUM OF      DEGREES OF    MEAN SUM
              SQUARES       FREEDOM     OF SQUARES
EXPLAINED   75.87516039        2        37.9375802
UNEXPLAINED 89.839375         13      1.074170355

ENTER THE CRITICAL F VALUE FROM THE TABLE.
[]:
      3.81
THE CALCULATED F VALUE IS 35.31802943
THE RELATIONSHIP IS STATISTICALLY SIGNIFICANT.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT THE RESIDUALS AND THE DURBIN-WATSON?
YES
OBSERVED Y       ESTIMATED Y        RESIDUAL
  16               15.37003178       0.6299682212
  14.9             15.69850853      -0.7985085322
  14.1             13.8690664        0.2309336028
  18.5             18.7004393       -0.2004392953
  15.3             16.3235359       -1.023535905
  21.9             19.79536181       2.104638193
  15.2             17.08998166      -1.023535905
  18.2             17.52795067       0.6720493328
  17               17.63744292      -0.6374429183
  16.8             17.88835355      -1.088353555
  20               19.90485406       0.09514594219
  14.8             13.65008189       1.149918105
  13               12.55515938       0.4448406163
  20.3             19.90485406       0.3951459422
  18.2             18.07541192       0.1245880772
  17.1             17.30896616      -0.2089661649

THE DURBIN-WATSON D STATISTIC IS 1.787153007
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT ANY MORE REGRESSIONS WITH THIS DATA?
YES
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
ENTER THE SUBSCRIPTS FOR THE INDEPENDENT VARIABLES TO BE USED
IN THE ANALYSIS...SPACE BETWEEN EACH ENTRY.  E.G.,  2 3 4
[]:
      2 3
ENTER THE SUBSCRIPT FOR THE DEPENDENT VARIABLE
[]:
      1
DO YOU WANT TO USE THE STEPWISE REGRESSION OPTION?
YES
```

******** STEPWISE REGRESSION ********

INTERCEPT IS 5.136566509

VAR NUM	COEFFICIENT	STD DEV	T VALUE
2	0.02501553892	0.009617578538	2.601022578
3	0.002104470372	0.001358888005	1.548670946

THE COEFFICIENT OF MULTIPLE CORRELATION IS 0.8900382702
THE STANDARD ERROR OF THE ESTIMATE IS 1.198444055

DO YOU WANT TO PERFORM THE F TEST?
NO

DO YOU WANT THE RESIDUALS AND THE DURBIN-WATSON?
NO

DO YOU WANT ANY MORE REGRESSIONS WITH THIS DATA?
YES

ENTER THE SUBSCRIPTS FOR THE INDEPENDENT VARIABLES TO BE USED IN THE ANALYSIS...SPACE BETWEEN EACH ENTRY. E.G., 2 3 4
?:
3 4
ENTER THE SUBSCRIPT FOR THE DEPENDENT VARIABLE
?:
1
DO YOU WANT TO USE THE STEPWISE REGRESSION OPTION?
YES

******** STEPWISE REGRESSION ********

INTERCEPT IS -1.997845063

VAR NUM	COEFFICIENT	STD DEV	T VALUE
4	0.125106383	0.05165138589	2.422130226
3	0.002268685215	0.001381502824	1.642186448

THE COEFFICIENT OF MULTIPLE CORRELATION IS 0.8844598851
THE STANDARD ERROR OF THE ESTIMATE IS 1.226652451

DO YOU WANT TO PERFORM THE F TEST?
NO

DO YOU WANT THE RESIDUALS AND THE DURBIN-WATSON?
NO

DO YOU WANT ANY MORE REGRESSIONS WITH THIS DATA?
NO

DO YOU WANT TO FORECAST WITH ONE OF THESE RELATIONSHIPS?
YES

```
ENTER THE SUBSCRIPTS FOR THE INDEPENDENT VARIABLES TO BE USED
IN THE ANALYSIS...SPACE BETWEEN EACH ENTRY.  E.G., 2 3 4
⎕:
      2 4
ENTER THE SUBSCRIPT FOR THE DEPENDENT VARIABLE
⎕:
      1
 AHHH, YES....THE REGRESSION EQUATION LOOKS LIKE:

      X1C = 1.861343345 + 0.02189845023 X2 + 0.1062996377 X4
ENTER THE PROJECTED INDEPENDENT VALUES (SPACE BETWEEN EACH)
⎕:
      245 110
THE FORECASTED DEPENDENT VARIABLE VALUE IS 18.9194238
IF YOU ARE THROUGH TYPE ZERO (SPACE) ZERO... E.G.,    0 0

ENTER THE PROJECTED INDEPENDENT VALUES (SPACE BETWEEN EACH)
⎕:
      295 110
THE FORECASTED DEPENDENT VARIABLE VALUE IS 20.01434631
IF YOU ARE THROUGH TYPE ZERO (SPACE) ZERO... E.G.,    0 0

ENTER THE PROJECTED INDEPENDENT VALUES (SPACE BETWEEN EACH)
⎕:
      0 0
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
DO YOU WANT TO FORECAST WITH ONE OF THESE RELATIONSHIPS?
NO
```

This is one of a complete set of computer-assisted-instructional programs which I have developed for a second semester introductory statistics course offered at the undergraduate level. Several characteristics of these programs become immediately obvious. Repetition is primarily in the verbal sections of the program because some students began losing interest when they were no longer able to visualize exactly what was going on at all times. Other students require little of this after several replications of the sample problems, so they can choose to eliminate a large part of the descriptive output. The data check has, more than once, prevented a student from reaching nonsensical or wholly erroneous results. Frequently, intermediate-step figures are given to assist the student in checking his own computations. Except for the data check and the initial simple correlations, the user has effective control over what he receives from the computer.

Another characteristic of this set of programs is that all of the relevant portions of a particular topic are presented at one time. This is also true in the lectures. I have found that this integrated-whole approach is better received than a parcel-by-parcel presentation. It is like putting together a puzzle . . . it is much easier if you can see a copy of the pattern before trying to piece together the parts. For example, the importance of the F-test and the Durbin-Watson for each repetition of the problem cannot be appreciated by most students without actually observing how they are used and what weight they may carry in determining which of the possible alternatives is "best". This leads to a third characteristic of these programs.

The user is able to sample all of the variations of a particular problem which interest him before arriving at a final solution. In most cases he would look at only one specification because other constraints limit the amount of time he can allocate to the problem. But, with a computer he has the advantage of being able to handle all of the alternative specifications in less time than he would spend manually reaching one solution. The sample problem illustrates that while the independent variable combinations X2 and X3, and X3 and X4 can be readily eliminated, the choice between X2, X3, and X4, and X2 and X4 is not nearly as clear cut. If the student was only looking for a "good" regression relationship, he might be tempted to stop with the first alternative. But, in fact, it is difficult to statistically select between these two alternatives. Frankly, this is one of the key points I try to bring to the student's attention. On what basis do you select between these two alternatives?

To date I have had just one class that was exposed to a complete set of these programs. The students individually elected to prepare their homework assignments either manually or with computer assistance. Not too surprisingly, most of them chose the former. But, those who were "daring" seemed to perform better throughout most of the semester.

One of the advantages which these "innovaters" realized was the ability not only to complete their assignments very quickly, but to also find time to work with more "real world" problems that they were interested in. In fact, they began to bring in statistics-related problems from other courses--sales forecasting, pro forma statement construction, motivation analysis. This obviously led to crossing the well-established lines of course content, and hopefully to fixing in their minds the usefulness of statistical tools rather than the futility they initially felt because they had to take this last statistics course to graduate.

Of course, there were negative aspects as well. Once away from the terminal, this group tended to have more trouble handling the manual operations. But they did appear to have the better grasp of the overall problem posed. Perhaps it all depends upon what you seek to achieve as a statistics instructor, whether you would consider this experiment a success or a failure. I had hoped to instill in the students a method of analysis, an ability to see the whole problem, and the importance of looking at all the possible alternatives before choosing a "best" solution. I feel that I was more successful using this approach than I had been previously, at least with those students who were willing to try the computer option.