DOCUMENT RESUME

ED 112 829 IR 002 410

AUTHOR Ryan, T. A., Jr.; And Others
TITLE Teaching Statistics with Minitab II.
PUB DATE Jun 75
NOTE 11p.; Paper presented at the Conference on Computers in the Undergraduate Curricula (6th, Fort Worth, Texas, June 16-18, 1975); Related document IR 002 399
AVAILABLE FROM Entire Proceedings; Ted Sjoerdsma, Treasurer, CCUC, 1248 Lindquist Center for Measurement, Iowa City, Iowa 52242 ($10.00, Checks payable to the University of Iowa)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS College Instruction; *Computer Oriented Programs; *Statistical Analysis; Statistics; Teaching Techniques
IDENTIFIERS Computer Augmented Statistics Instruction; *Minitab II

ABSTRACT

Minitab is a statistical computing system which uses simple language, produces clear output, and keeps track of bookkeeping automatically. Error checking with English diagnostics and inclusion of several default options help to facilitate use of the system by students. Minitab II is an improved and expanded version of the original Minitab which was based on the Omnitab System of the National Bureau of Standards. With the new system students can analyze more extensive and realistic data sets and, as a consequence, can devote more time to the concepts of statistics rather than to calculations. Minitab consists of a worksheet (normally 200 rows by 50 columns) and about 130 free format commands which provide a wide variety of capabilities. The system is in current use not only in academic settings but in industrial environments as well. Minitab is written in machine compatible FORTRAN IV and can be implemented on a variety of computers. Examples of a paired t-test, a data analysis, an illustration of the concept of a limit, a simulation, and a polynomial regression with data transformation are provided to show the ease and flexibility of Minitab II. Output exhibits for these examples are appended. (CH)

TEACHING STATISTICS WITH MINITAB II

T. A. Ryan, Jr.
Dept. of Statistics
The Penn. State Univ.
Univ. Park, Pa. 16802

B. L. Joiner
Dept. of Statistics
Univ. of Wisconsin
Madison, Wisc. 53706

B. F. Ryan
Dept. of Statistics
The Penn. State Univ.
Univ. Park, Pa. 16802

1. Introduction

Minitab is a general purpose statistical computing system which was originally developed for use by students in introductory, pre-calculus statistics courses. It has had a great impact on teaching these courses and many advanced courses as well. When the system was originally planned, the intent was to allow students to analyze more extensive and more realistic data sets, and thus get a better perspective on how statistics is really done. This objective has certainly been achieved. But what we didn't realize was that the basic emphasis of our teaching would change. Relieved of tedious calculations and worry about the technicalities of statistical procedures, we and the students are now free to concentrate on the important concepts of statistics — on exploring data and choosing good procedures to analyze the data.

The original Minitab system [6], based on the Omnitab System of the National Bureau of Standards [2, 3, 4, 9], was begun in 1971. It was used extensively at Penn State, both in courses and research, and was exported to over 65 institutions. Minitab II, which has just become available to users, is a significantly improved and expanded version.

Minitab can be readily installed, in either batch or interactive mode, on any medium to large computer.

2. How Minitab Works

Minitab consists of a worksheet (normally 200 rows by 50 columns) in which data are stored, plus a collection of about 130 pseudo English commands which operate on the data in the worksheet. The easiest way to see how Minitab works is to look at a simple example.

Problem: Suppose 3 exams were given in a course. A student's final grade is his average for the 3 exams. Find the final grade for each student, and find the class average of the final grades. There is one data card for each student containing his 3 exam grades.

Minitab program:

```
READ THE FOLLOWING GRADES INTO COLUMNS C1-C3
 85   82   90
 89   75   82
 95   85   72
100   98   95
 98   90   91
```

ADD C1-C3, STORE SUMS IN COLUMN C4
DIVIDE COLUMN C4 BY 3, PUT FINAL GRADES INTO C5
PRINT THE FINAL GRADES WHICH ARE IN C5
AVERAGE THE NUMBERS IN C5
STOP

The output is given in Exhibit 1.

Some basic points.

(1) Minitab is an interpreter. It reads (and prints) a command card, checks for errors, executes the command, and then goes on to the next command.

(2) The columns of the worksheet are denoted by C1, . . ., C50. In addition to the worksheet, Minitab can store up to 50 constants, denoted by K1, . . ., K50, and 10 matrices, denoted by M1, . . ., M10. (At present only a few of the planned matrix operations have been implemented.) We will use the word "argument" to refer either to a column, a stored constant, a matrix, or a number.

(3) On a command card, only the (first 4 letters of the) command name and the arguments (in the proper order) are used by the computer. All command and data cards are free format. For example, the third command card in the above program, contains the command name DIVIDE and arguments C4, 3, C5. We could just as well have written this command as: DIVIDE C4 BY 3, PUT IN C5, or even DIVI C4 3 C5. A list of consecutive column numbers may be indicated by using a dash, as in the READ command.

3. The Commands in Minitab

Minitab's 130 commands provides a wide variety of capabilities: including:

Arithmetic: ADD, MULTIPLY, SQRT, MAXIMUM, COUNT, PARSUMS, etc.

Functions: LOGTEN, LOGE, SIN, ABSOLUTE VALUE, etc.

Descriptive Statistics: AVERAGE, MEDIAN, HISTOGRAM, CORRELATION, etc.

Manipulations: SQRT, OMIT, CHOOSE, PICK, JOIN, etc.

Plots: PLOT, MPLOT (multiple), TPLOT (pseudo three dimensional), etc.

Basic Statistics: TINTERVAL, TTEST, CONTINGENCY TABLE, REGRESS (Simple, polynominial and multiple), AOVONEWAY, etc.

Data Generation: NRANDOM (normal), URANDOM (uniform), PRANDOM (Poisson), SAMPLE (without replacement), etc.

In addition, many other capabilities not expressly built in as commands can be readily done by linking together a few commands. For example SUBTRACT and TTEST will give a paired t-test, NSCORES and PLOT will do a normal probability plot, URANDOM and LOGE will give exponential random data, and so on.

Input and output is usually free format, but Fortran format statements can also be used when desired.

Other capabilities currently under development include (1) a very general analysis of variance capability, (2) matrix operations, (3) looping, (4) labeling variables, (5) robust statistical procedures, (6) non-linear

regression, (7) analysis of multidimensional contingency tables, and (8) Fortran-like algebraic statements: e.g., COMPUTE C5 = SIN(C1/2) + C2**5.

4. Some Examples

The following four examples show the ease and flexibility of using Minitab in undergraduate courses.

A. Paired t-test.

Exercise. A study was done to determine the effectiveness of cartoon vs. realistic pictures in visual aids. A short slide presentation containing 9 cartoon and 9 realistic slides was given to 34 University students. After the presentation, each student took a quiz which contained one question on each slide. Compare the effectiveness of cartoon vs. realistic slides. The data has been punched into cards. There is one card for each student containing his score on the 9 cartoon questions followed by his score on the 9 realistic questions.

Minitab does not have a command to do a paired t-test. However, since a paired t-test is just a test of $\mu = 0$ on the differences, we can use the following program:

```
READ CARTOON SCORES INTO C1, REALISTIC SCORES INTO C2
 (34 data cards go here)
SUBTRACT C2 FROM C1, PUT DIFFERENCES INTO C3
TTEST WITH MU = 0 ON DIFFERENCES IN C3
STOP
```

The output is in Exhibit 2. The results of the test show that the students did overwhelmingly better on the cartoon questions than on the realistic questions, suggesting that cartoon pictures make learning easier.

B. Data Analysis.

Exercise. The students in the visual aid study of Example A were also given the OTIS Quick Scoring Mental Ability Test, which yielded a rough estimate of their natural ability. This score was punched as the third number on each student's data card. What do the OTIS scores tell you about the experiment and the analysis of Example A?

This is a fairly openended exercise, where a student is free to explore the data using various techniques. Here suppose we just look at the relationship between the realistic scores and the OTIS scores. We usually expect a high correlation between ability and test scores. The Minitab program:

```
READ CARTOON SCORES INTO C1, REALISTIC SCORES INTO C2, OTIS SCORES INTO C3
(34 data cards)
CORRELATION BETWEEN C2 AND C3
STOP
```

finds a correlation of only 0.237. A plot often shows what's really going on in a data set; the command

```
PLOT REALISTIC IN C2 VS. OTIS IN C3
```

Produced the plot in Exhibit 3. Two things are immediately apparent — both of which tend to lower the correlation coefficient. First there is an outlier. A check of the data shows that this same student also had a very low cartoon score. Perhaps he should be omitted from analysis — at least temporarily. The plot also shows a truncation effect in the students' cartoon scores. Those with high OTIS scores tended to run into a ceiling on the test, indicating that the slide presentation and exam was too easy for these students. This may mean loss of information.

C. Calculus.

Although Minitab was designed as a statistical system, its flexible structure encourages its use in other fields. The following example shows how Minitab can illuminate the concept of a limit.

Exercise. Use Minitab to study the limit of (sin x)/ as x → 0. A student might use the following program:

```
GENERATE THE INTEGERS FROM 1 TO 100 INTO C1
DIVIDE 1 BY C1, PUT INTO C2 (X-VALUES)
SIN OF C2, PUT INTO C3
DIVIDE C3 BY C2, PUT INTO C4 (SINX/X)
PRINT X IN C2, SINX/X IN C4
STOP
```

He could then use the same technique to find the limit from below, and observe that in both cases (sin x)/x converges to 1, as x approaches 0.

D. Simulation.

Minitab can generate data from most of the distributions used in statistics. This allows students to do the usual simulation exercises — coin tossing, drawing balls from urns, observations of x, generating confidence intervals, generating data from a given regression model, etc.

The following example, taken from the Minitab Handbook [8], shows how simulation can be used to illustrate a fairly advanced concept — the Arc Sine Law [1].

Losing Streaks. Gamblers, sports fans and investors in the stock market all claim they have "winning streaks" and "losing streaks". In this example, we will use simulation to show that these streaks may be nothing more than chance fluctuations. Consider the following simple gambling game. You and your opponent each bet a dollar. Then you toss a fair coin. If it comes up tails, you win one dollar. If it comes up heads, you lose one dollar. Suppose you continue to play this game for a long time. Would you expect to be ahead for a short while, then behind for a short while, then ahead again, and so on? Or would you expect to be ahead for a very long time (in a winning streak) and then behind for a long time (in a losing streak)?

The following program simulates this game for 60 plays:

```
READ VALUES INTO C1, PROBABILITIES INTO C2
                 -1                          .5
                  1                          .5

DRANDOM 60 OBS., VALUES IN C1, PROB. IN C2, PUT IN C3
PARSUM C3, PUT PARTIAL SUMS IN C4
GENERATE 60 INTEGERS INTO C5
WIDTH 60 HEIGHT 20 FOR PLOT
PLOT C4 VS. C5
STOP
```

In this program we first generated a sequence of sixty random +1's and -1's to indicate for each toss whether you won or lost. They were put in C3. The command PARSUM calculated your total winnings or losses after each toss. The last three commands plot your successive "winnings" as the game progresses. This plot is shown in Exhibit 4. It's quite vivid from the plot that you were in a very long losing streak. In fact, this sort of pattern of long winning or losing streaks is quite typical. Usually you will spend most of the time in long winning streaks, or in long losing streaks. So it may not be "luck" after all that gives gamblers their winning and losing streaks. It may be just the laws of chance.

5. Philosophy and Use of Minitab

Basic to the development of Minitab is a strong belief that the system must be very easy to use, and at the same time be a genuine statistical tool. By this we mean that it should be useful for doing real statistics, not just simple textbook problems.

The best evidence that Minitab is easy to use comes from an introductory statistics course given at Penn State. Each year, over 1,000 students take this course. Although most of these students have never used a computer before and have poor mathematical backgrounds, they are able to write, punch, and run Minitab programs after one hour of instruction. Except for this first session, very little class time is spent on Minitab, per se. As new statistical methods are introduced, the corresponding commands are mentioned. Typically 10 problems involving the computer are assigned each term. Many of the students in this course continue to use Minitab in subsequent courses and in research.

Why is Minitab so easy to use? First its language is simple, natural, and consistent. To illustrate what we mean by natural, let's look at one small example. Consider the two commands ADD C1 TO C2, PUT IN C3 and ADD 1 TO C2, PUT IN C3. It's clear that the first command adds 2 columns together while the second adds the number 1 to (all entries in) a column. In Omnitab (and the original Minitab) these 2 commands are written ADD 1 TO 2, PUT IN 3 and ADD 1.0 TO 2, PUT IN 3. Omnitab uses the presence or absence of a decimal point to distinguish numbers from column numbers. One or two artificialities such as this are tolerable. But too many and students get annoyed and confused by the special rules and conventions they must remember.

Second, Minitab's output has been designed (and redesigned) to be as clear, as concise, and as pedologically sound as possible. Consider the regression output in Exhibit 6. To a beginning student, the most important part of regression is the regression equation, so this is what appears first. Having the equation in this form also helps a student orient himself to the tables which follow.

Third, Minitab keeps track of all bookkeeping automatically. For example, you don;t have to specify how many data cards there are to be read. Minitab counts them for you. You can also freely mix different length columns.

Fourth, Minitab has default options, so that easy things are easy to do. If you want to plot column 1 vs. column 2, you can use the command PLOT C1 VS. C2, and never worry about specifying scales. If however, you want to control the scales, or the size of your plot, you can use the more advanced plotting capabilities.

And finally, Minitab has extensive error checking with good, clear, English diagnostics. All error messages come from the Minitab system and not from the computer's operating system.

The best evidence that Minitab is a genuine statistical tool comes from its research users. For example, Minitab is heavily used by the Penn State Statistical Consulting Center — especially for regression analysis, and for analyzing small data sets. It is also used in industrial environments, e.g., at The Alcoa Technical Center where it's used primarily by experimenters who are not professional statisticians. Of course, Minitab is a limited tool — it does not have any multivariate analysis, (no factor analysis, etc.), nor any time series, and it cannot handle extremely large data sets. But, within its scope, it is extremely useful.

Because students know they are using a real tool, and not just a teaching device, they are very enthusiastic about Minitab and about the computer. It also appears that once students have "conquered" the computer with an easy-to-use system such as Minitab, they can then go on to more sophisticated packages like SPSS, BMD, or Omnitab with very little difficulty.

6. How Minitab Can Change the Way you Teach — One Example

Beginning students find it easier to appreciate statistical concepts if the concepts are extensively illustrated by examples, and particulary if the examples involve good graphical techniques. Here we outline how

polynomial regression and data transformations can be explained to students who have studied only simple regression and testing.

The Example: The classical pendulum experiment from physics is designed to show the relationship between the length of a pendulum and its cycle time (or period). We can use the following data [5] to explore this relationship:

length (L)	175.2	151.5	126.4	101.7	77.0
time for one cycle (T)	2.650	2.468	2.256	2.024	1.764

A plot of T vs L, given in Exhibit 5, shows a very strong linear relationship, indicating that the equation $T = B_0 + B_1 L$ should provide a reasonable fit. The output from this fit is given in Exhibit 6. (How much of this output should be explained to students depends on how much statistics they know). The results from this first fit look fairly good (high R^2, significant coefficients) but we shouldn't quit yet.

At this point it is convenient to introduce residual plots as a powerful visual technique for examining lack of fit. One such plot is shown in Exhibit 7. This plot indicates that there was some curvative in the original data. An obvious thing to try next is the quadratic, $T = B_0 + B_1 L + B_2 L^2$. As far as students are concerned, fitting this equation is just an obvious extension of simple regression, and easy to do using the Minitab program:

```
READ LENGTH INTO C1, CYCLE TIME INTO C2
  (data cards)
MULTIPLY C1 BY C1, PUT INTO C3
REGRESS TIME IN C2, ON 2 PREDICTORS IN C1, C3, STORE RESIDUALS IN C4
PLOT RESIDUALS IN C4 VS LENGTH IN C1
STOP
```

This quadratic equation fits the data very well, with no lack of fit indicated by the residual plot. Of course, we know from physics that the "true" relationship between L and T is, $T = B_1\sqrt{L}$. We can also try fitting this model. Again, the concept of data transformations is very natural, and easily explained to students. This fit is also very easy to do using the Minitab program:

```
READ L INTO C, T INTO C2
  (data cards)
SQRT OF C1, PUT INTO C3
NOCONSTANT (IN FOLLOWING REGRESSION COMMAND, FIT EQUATION
  WITHOUT B0 TERM)
REGRESS  T IN C2 ON 1 PRED. IN C3, STORE RES. IN C4
PLOT RES. IN C4 VS LENGTH IN C1
STOP
```

7. Availability and Implementation of the Minitab System

Minitab is written in machine compatible FORTRAN IV and has been inplemented on a wide variety of computers, including IBM 360 models 30 and up, IBM 370, small and large CDC, UNIVAC, Burroughs, Xerox, DECsystem-10. It requires about 35K words of memory, and can be overlayed into less than half that space on many computers. It runs in either batch or interactive mode. A copy of the system is available, either on tape or cards, at a nominal charge from the Penn State Statistics Dept. Documentation includes (1) a 90 page Reference Manual and (2) a Student Handbook (about 200 pages), which includes an introduction to the Minitab system, material on how to use Minitab for statistical analysis, examples and exercises.

REFERENCES

[1] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, 3rd ed., John Wiley & Sons, 82.

[2] Hilsenrath, J., Ziegler, G. G., Messina, C. G., Walsh, P. J., and Herbold, R. J. (1966). *Omnitab: A Computer Program for Statistical and Numerical Analysis*, National Bureau of Standards Handbook 101, U.S.G.P.O., Washington, D.C.

[3] Hogben, David, Peavy, Sally T., and Varner, Ruth N. (1971). *Omnitab II User's Reference Manual*, National Bureau of Standards, Technical Note 552, 264 pages, U.S.G.P.O. Washington, D. C.

[4] Jowett, D., Chamberlain, R. L., Mexas, A. G. (1972). "Omnitab—A Simple Language for Statistical Computations," *Journal of Statistal Computation and Simulation, 1*, 129-147.

[5] Mandel, John (1973). "The Acceleration of Gravity" in Mosteller, Frederick, et. al., eds., *Statistics by Example, Detecting Patterns*, Addison-Wesley, 109-118.

[6] Ryan, T. A. and Joiner, B. L. (1973). "Minitab: A Statistical Computing System for Students and Researchers," The American Statistician, 27, 222-225.

[7] Ryan, T. A., Joiner, B. L. and Ryan, B. F. (1974). *Minitab II Reference Manual* (preliminary edition). Statistics Dept., The Pennsylvania State University.

[8] Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1974) *Student Handbook for the Minitab Computing System* (preliminary edition). Statistics Dept., The Pennsylvania State University.

[9] Swanson, J. M., Riederer, S. and Weekly, H. (1973). "Using Omnitab to Teach Applied Statistics", "Proceedings of a Fourth Conference on Computers in the Undergraduate Curricula." The Claremont Colleges, Claremont Calif.

Exhibit 1

```
-- READ THE FOLLOWING GRADES INTO COLUMNS C1-C3
COLUMN    C1          C2          C3
COUNT      5           5           5
          85.         82.         90.
          89.         75.         82.
          95.         85.         72.
         100.         98.         95.
          98.         90.         91.

-- ADD C1-C3, STORE SUMS IN COLUMN C4
-- DIVIDE COLUMN C4 BY 3, PUT FINAL GRADES INTO C5
-- PRINT THE FINAL GRADES WHICH ARE IN C5
COLUMN    C5
COUNT      5
       85.6667     82.0000     84.0000     97.6667     93.0000

-- AVERAGE THE NUMBERS IN C5
   AVERAGE =       88.467

-- STOP
```

Exhibit 2

```
-- READ CARTOON SCORES INTO C1  REALISTIC SCORES INTO C2

   SUBTRACT C2 FROM C1,PUT DIFFERENCES INTO C3
-- TTEST WITH MU=0 ON DIFFERENCES IN C3
   C3      N =  34      MEAN =        1.0000      ST.DEV. =        1.28

   TEST OF MU =     0.0    VS. MU N.E.     0.0
   T =   4.558
   THE TEST IS STATISTICALLY SIGNIFICANT AT ALPHA = 0.0000

-- STOP
```

Exhibit 3

```
      -- PLOT REALISTIC IN C2 VS. OTIS IN C3
           2
       9.0+          *  * *   *   *   *
          -
          -  *   *   **  *           * * *
          -             ** **     *
          -
       6.0+        **        *    **    *
          -
          -       *  **      **
          -        *
          -
       3.0+       *     *
          -
          -     *
          -
          -
       0.0+                         *   outlier
          +---------+---------+---------+ 3
         95.                125.
                  110.                140.
```

Exhibit 4

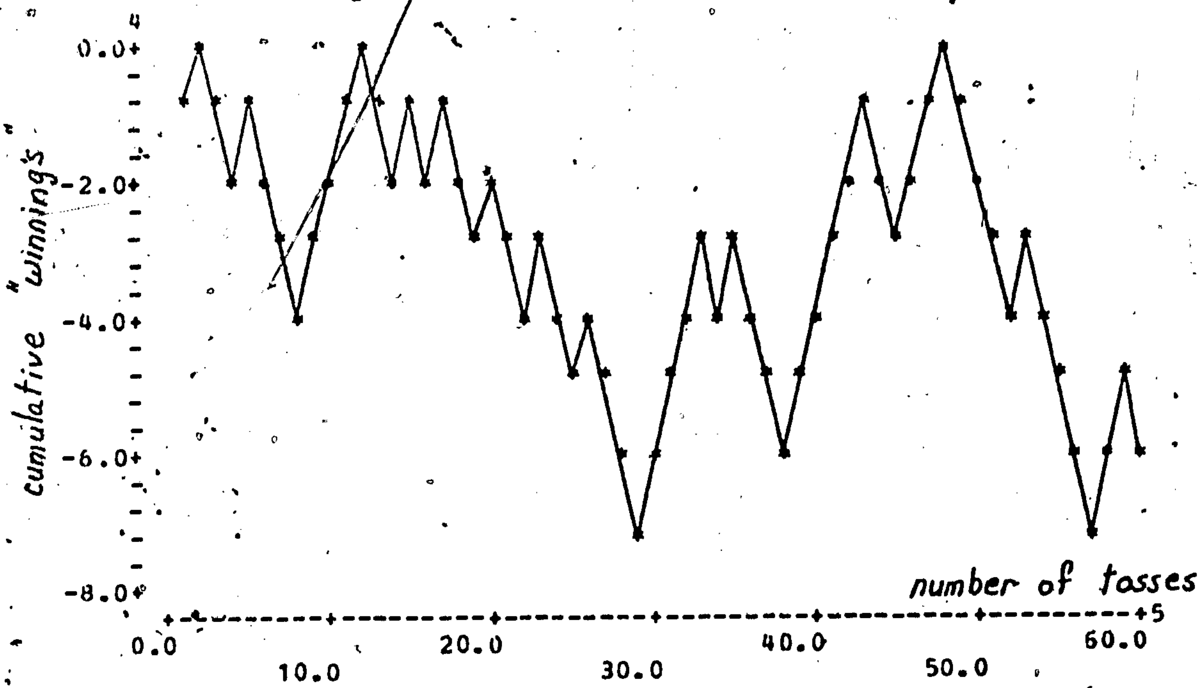

Exhibit 5

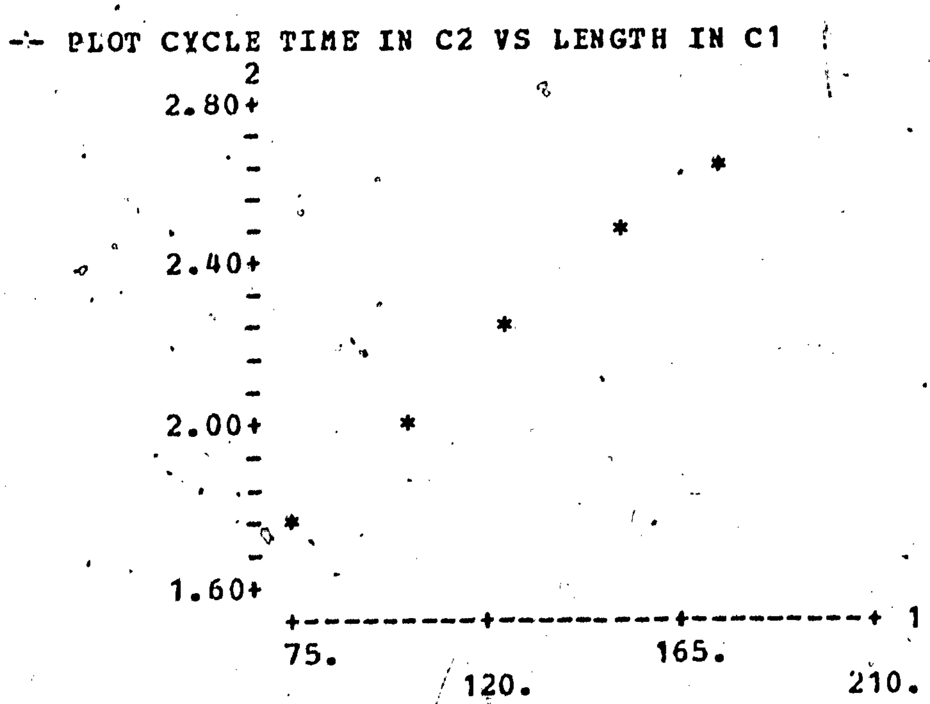

Exhibit 6

```
-- REGRESS CYCLE TIME IN C2 ON 1 PREDICTOR IN C1
THE REGRESSION EQUATION IS
Y= 1.09      + (0.900D-02) X1
             X IN        Y IN      PRED Y     STD DEV                 STD.
ROW        COL. 1      COL. 2     VALUES     PRED Y       RES.      RES.
 1        175.200      2.650      2.672      0.019      -0.022     -1.388
 2        151.500      2.468      2.459      0.014       0.009      0.443
 3        126.400      2.256      2.233      0.011       0.023      1.039
 4        101.700      2.024      2.010      0.014       0.014      0.651
 5         77.000      1.764      1.788      0.019      -0.024     -1.520

                        ESTIMATES FROM LEAST SQUARES FIT
X        COLUMN       COEFFICIENTS          S.D.OF COEFF.        RATIO
0          **            1.0947               0.421D-01          26.00
1          1             0.90036D-02          0.321D-03          28.03
THE ESTIMATED ST. DEV. OF Y ABOUT THE REGRESSION LINE IS         0.0250
 WITH (  5- 2) =    3 DEGREES OF FREEDOM
R-SQUARED = 99.6 PERCENT
R-SQUARED = 99.5 PERCENT ( ADJUSTED FOR D.F.)

                             ANALYSIS OF VARIANCE
                            DF             SS                    MS=SS/DF

 TOTAL                       5          25.41
 DUE TO MEAN                 1          24.92
 TOTAL DUE TO REGRESSION     1          0.4914                   0.4914
 RESIDUAL                    3          0.1876D-02               0.6255D-03
X TRANSPOSE X INVERSE IN LOWER TRIANGLE
CORRELATIONS OF ESTIMATES OF COEFFICIENTS IN UPPER TRIANGLE
                    ***            1
***                2.83397        -0.96
 1             -0.208450D-01   0.164965D-03
```

Exhibit 7

-- PLOT RESIDUALS IN C3 VS LENGTH IN C1