Minitab Newsletter

1. The Newsletter

This is the first newsletter about the Minitab Statistical Computing
System. We expect to produce these letters periodically and would welcome
hearing from users concerning novel uses in teaching or research, notes con-
cerning implementation on various computers; in short, anything of reader
interest for inclusion in future newsletters.

2. About Minitab II

Minitab II was developed by Thomas A. Ryan, Jr., Brian L. Joiner, and
Barbara F. Ryan. It has been available to users since late summer, 1974 and
incorporates all of our "second thoughts" since developing Minitab I. All
copies of Minitab sent out since September, 1974 have been Minitab II.

The new version continues the tradition of being designed for very naive
users, such as students in elementary, pre-calculus statistics courses, and
researchers who know little about either statistical analysis or computing.
Minitab II is even easier and more natural to use than the original Minitab.
It is much more flexible, and a number of new statistical capabilities have
been added. Considerable improvement has been made in interactive use. It
is still designed primarily for small to medium-sized data sets.

To illustrate how Minitab II differs from Minitab I (which was described
in the December, 1973 American Statistician), consider the following example:

```
READ DATA INTO COLUMNS C1-C3 (STREAM FLOW, DEPTH AND WIDTH)
0.636     0.34     1.6
0.319     0.29     1.9
  .         .        .
  .         .        .
  .         .        .
0.734     0.28     2.8
REGRESS C1 ON 2 PREDICTORS IN C2 AND C3, PUT RES IN C4, PRED IN C5
PLOT RESIDUALS IN C4 VS PREDICTED VALUES IN C5
NOTE NOW TRY RECIPROCAL TRANSFORMATION
DIVIDE 1 BY C1 PUT RECIPROCALS IN C6
REGRESS C6  2  C2, C3  C7, C8
PLOT C7 VS C8
STOP
```

Columns and constants are distinguished by the letter "C" instead of a decimal
point.

3. Updates

Updates for all copies of Minitab II distributed before June, 1975 have been mailed. If you are responsible for maintaining Minitab at your installation, and did not get a copy of the updates, please send your name and address to us on the form provided at the back of this newsletter.

4. The Minitab Handbook

The "Minitab Handbook" (alias "the red book" or the "Student Handbook for the Minitab Statistical Computing System") will be published by Duxbury Press, North Scituate, Massachusetts 02060, probably in Spring, 1976. Several new chapters have been added to those in the present, preliminary version of the Handbook, such as analysis of variance and non-parametric statistics. The Handbook explains how to explore data and carry out statistical analyses using Minitab, and is intended as a supplementary text for elementary (pre- or post-calculus) statistics courses. Our users report it is also widely used in consulting.

Until it is published, the preliminary edition of the Handbook will continue to be available from Penn State. The current price is $2.80. Purchase orders may be sent to: The University Park Bookstore, McAllister Building, University Park, Pennsylvania 16802. If you need a number of copies (more than 30 or so), please give as much advance notice as possible. If you want just one or two copies, you can write to the Penn State Statistics Department.

A reminder - the Handbook is copyrighted and cannot be reproduced without permission (unlike the Reference Manual, which may be reproduced at will).

5. The Minitab Reference Manual

A new edition of the Reference Manual is now being printed. This manual is designed for sophisticated users who want to know in precise detail how Minitab does various operations, how output can be controlled, etc. New features in the manual are an index, section names and numbers as page headers, an improved printing process, and a soft binding. The price is $2.00 and copies may be ordered from the Penn State Statistics Department.

Users may duplicate the reference manual themselves if they prefer; it is included on new tapes of the Minitab II System.

6. Projects Underway on the Minitab II System

Here is an outline of projects now under way at Penn State. We hope to have most of these features in the next updated version, which will be available in mid-October.

(1) Line Counter - We have been adding a line counter to new routines and inserting a counter in older routines for some time. We are now in the process of checking the line counter, and using it to conserve paper, especially for plots and contingency tables.

(2) <u>Histogram</u> - A new histogram routine has been written. This version provides better scaling and better handling of cases where more than 50 observations are in one class, optional user specification of scales, and fits on a 72-character teletype output. We are in the final stages of testing this routine; it will appear in the next updated version.

(3) <u>Random Number Generators</u> - New features being developed include a new uniform random number generator for the IBM 360/370 version, a machine-compatible random number generator to ease installation problems on other machines, and an improved method of generating normal random variables.

(4) <u>One-Way Analysis of Variance</u> - A new one-way analysis of variance command:

<u>ONEWAY</u> AOV, DATA IN <u>C</u>, CELL SUBSCRIPT IN <u>C</u>

is being developed. The output will be similar to the existing AOVONEWAY command, but the form of the data input will be much more convenient for many data sets encountered in consulting and research.

(5) <u>Two-Way Analysis of Variance</u> - The command:

<u>TWOWAY</u> AOV, DATA IN <u>C</u>, CELL SUBSCRIPTS IN <u>C</u> AND <u>C</u>

is being developed to handle a balanced two-factor analysis of variance.

(6) <u>Non-Parametric Commands</u> - We are now designing one- and two-sample rank tests (Wilcoxon and Mann-Whitney tests) and the corresponding confidence intervals. A new efficient method of computing these confidence intervals now makes them feasible on larger data sets for the first time.

(7) <u>Variable Labels</u> - Labels can be defined for columns, constants, and matrices. These appear on the output and can be used as arguments in commands if enclosed in single quotes.

(8) <u>Stored Commands and Looping</u> - A primitive looping feature is now being tested and a more sophisticated one is being planned.

(9) <u>Smaller Version of Minitab</u> - Under development. Present version can be overlayed into about 12K words of core.

7. <u>Long-Term Plans for Minitab</u>

We welcome suggestions for improvements to Minitab and would like to hear your priorities; we also welcome support for projects.

Our highest priority at present is a more general analysis of variance capability. The projected solution centers around a model card, where the model will be given to the computer in virtually the same form as it is expressed in most textbooks. For example:

$$Y(I,J) = M + A(I) + B(I,J) + ERROR$$

expresses the usual two-stage nested model, translated into Latin letters for the computer.

Next on the list of priorities is the ability to handle stored files of commands or data, especially on interactive systems. We intend to extend the READ (and SET) commands to handle files containing data as follows:

READ FROM UNIT K INTO C,...,C
READ FROM FILE F INTO C,...,C

Reading from files will have to be machine-dependent.

To execute files of stored instructions, the command will probably be

EXECUTE FILE F

or

EXECUTE FILE F K TIMES.

(The latter is essentially a loop command.) To make stored commands more useful, we will probably allow "dummy" arguments D1, D2,..., to be replaced at execution time, as in the following example:

STORE COMMANDS IN FILE F3
AVERAGE D1
HISTOGRAM D1
END STORING COMMANDS
EXECUTE FILE F3 4 TIMES, WITH D1=C1-C4

We also plan to implement some new matrix commands.

We urge you to comment on what your priorities for Minitab are. What would you like to see us do? What do you want us not to do? Have you added features that might be of interest to others?

8. Notes on Implementation

Over 150 copies of the Minitab system have been distributed to a variety of institutions in the U.S. and foreign countries. Some institutions exper- ience no difficulty in implementation, while others seem to have more trouble. Considerable effort has gone into making Minitab II easy to implement on a wide variety of computers, and we are continually seeking ways to make implementation even easier. (For example, we are now working on ways to simplify overlaying.) Here is a brief outline of information presently available on implementation.

IBM 360/370. Most of the development of Minitab has been done on various models of IBM 360 and 370 at Penn State (currently a 370/168 running under OS). We test Minitab using Penn State's modified version of the WATFIV compiler and FORTRAN G and H (opt=2). The version on our system library is compiled in H (opt=2), and runs in batch mode in a 120K-byte (30K word) region. The average job runs in slightly under one second elapsed time and about 1/4 second CPU

time. Minitab runs on similar computers interactively under TSO and other systems. No difficulty should be experienced in putting Minitab on any of the larger IBM 360 or 370 models (although Cornell reports a difficulty in getting one or two routines to compile under H (opt=2)).

On smaller 360/370's, running under DOS, one difficulty has been pointed out to us: there is a limit of 50 subroutines, other than system subroutines, which can be linked with one program. Since Minitab has more than 50 subroutines, this requires putting some of Minitab's subroutines temporarily in the relocatable library, link-editing, and then removing the subroutines from the library. Also, overlaying is a bit more of a hassle under DOS. We hope the projected smaller version of Minitab may ease some of these problems.

DECsystem-10. Robert F. Kohm, at Alcoa Laboratories, has used Minitab extensively on their DEC. He reports that Minitab is very heavily used by engineers at the Alcoa Technical Center, and has given us numerous detailed comments which have improved Minitab greatly, especially in interactive use. He now runs it overlayed in 12 or 15K words, and has added features for command and data file storage, CALCOMP plotting, etc. We hope to add many more of his improvements to Minitab II in the future.

Xerox. Minitab was routinely installed on a Xerox Sigma 7 used by the Penn State Library system. Overlayed, it runs in about 12K words and is used interactively. It was also routinely installed at Texas Christian University on their Sigma 9 for use at the 1975 Conference on Computers in the Undergraduate Curriculum.

Univac. B. Joiner is now at the University of Wisconsin and runs Minitab in both batch and interactive modes on a Univac 1110. It runs in 47K words of memory (with a 5K word worksheet) without overlaying.

CDC. Minitab has been successfully installed on a number of CDC computers, both 3000 series and 6000 series. National Bureau of Standards at Boulder, Colorado reports no difficulty in putting the latest version on their CDC 6600.

Burroughs. Minitab has been put on several Burroughs models. One minor problem was discovered, which has been corrected in the current version.

PDP-11. Several people have obtained copies of Minitab for PDP-11's, but we have at present no details concerning their success or failure in implementing it.

PLEASE: We would appreciate any details anyone can give us concerning their implementation of Minitab. We want to eliminate as many difficulties as possible, both minor and major. A minor difficulty for one implementer may be a major hurdle for someone else.

9. An Example in Data Analysis

Recently we took another look at the "stack loss" data which has appeared in several statistics texts (Brownlee's Statistical Theory and Methodology;

Draper and Smith's <u>Applied Regression Analysis</u>;and most recently in Daniel and
Wood's <u>Fitting Equations to Data</u>). The problem involves developing an equation
to relate a chemical plant's indirect measure of the yield of nitric acid to
three variables: x_1 = air flow, x_2 = temperature, and x_3 = concentration of
nitric acid.

Daniel and Wood's creative analysis showed that some of the anomalies in
the data could be explained by the fact that they occurred on days following
large changes in x-values which may indicate a lag effect. We followed this
suggestion one step further by adding in the <u>changes</u> in the x's from the
previous day as additional variables. The following program was used:

```
READ X VARIABLES INTO C1-C3, Y INTO C4
    (21 data cards)
PICK ROWS 2 TO 21 OF C1, PUT INTO C11        (TODAY)
PICK ROWS 1 TO 20 OF C1, PUT INTO C21        (YESTERDAY)
SUBTRACT C21 FROM C11, PUT INTO C31          (CHANGE)
    .
    .      (similarly for the other 2 x-variables)
    .
PICK ROWS 2 TO 21 OF C4, PUT INTO C4     (TODAY-Y)
REGRESS Y IN C4 ON 6 PREDICTORS, X IN C11-C13, CHANGES IN C31-C33
STOP
```

When this program was run, it revealed several interesting facts. First,
the change variables were highly significant, indicating a strong dependence
on the amount of change in the x-variables. Second, x_3, which was not signi-
ficant in any of the analyses presented in the three textbooks, became signi-
ficant once the change variables were included. Third, the residuals were
reduced. In particular, the residual for the observation that had appeared to
be an outlier, was now brought right into line.

10. <u>Further Information About Minitab</u>

If you would like to obtain any further information about Minitab, please
fill in the attached sheet. Also, if you have any comments, we would appreciate
hearing from you.

11. <u>Ordering Minitab</u>

The price for the Minitab System includes one copy each of the Handbook,
Reference Manual, and implementation guide, as well as the program itself. If
the program is ordered on tape, the tape also includes a copy of the Reference
Manual, a short write-up, a test deck, and the results of running the test deck.

Information Request - Mailing List

_____ I would like to receive future newsletters.

_____ Please send me information on Minitab II.

_____ We currently use Minitab I on a _____ Computer.

_____ We currently use Minitab II on a _____ Computer.

Send this form (or the relevant information) to:

    Professor Thomas A. Ryan, Jr.
    Statistics Department
    215 Pond Laboratory
    The Pennsylvania State University
    University Park, Pennsylvania 16802

Your Name _____

Institution _____

Address _____

        _____

        _____

_____ We have Minitab II, but have not received the updates to June 1975.

The person responsible for maintaining and updating Minitab at our installation is:

    Name _____

    Institution _____

    Address _____

            _____

            _____

Order Form for Minitab II   Send to: Professor T. A. Ryan, Jr.
                                       Statistics Department
                                       215 Pond Laboratory
                                       The Pennsylvania State University
                                       University Park, Pa. 16802

The charges below are to cover the cost of handling, the cost of materials, and surface mail. These prices became effective August 1, 1975 and are subject to change. We include one copy each of the Minitab Handbook, the Minitab Reference Manual, and an implementation guide. The cost of the reel of tape is included in the price.

		Number	Amount
9 track, 1600 bpi, EBCDIC, OS labelled tape	$50/$100*	___	___
9 track, 800 bpi, EBCDIC, unlabelled tape	$50/$100*	___	___
7 track, 800 bpi, BCD, unlabelled, unblocked tape	$50/$100*	___	___

Other tape specifications, such as ASCII, DEC tapes, etc.
(These tapes may involve a delay of several weeks.)
Specify: _____ $100/$200*   ___   ___

_____

EBCDIC (029) cards	$125/$250*	___	___
BCD (026) cards	$125/$250*	___	___

(If interpreting is not needed, subtract $25.)

Extra Handbooks - $2.80 each.
(Please send orders for five or more to the University
Park Bookstore, at the address given in Section 4.)   ___   ___

Extra Reference Manuals - $2.00 each   ___   ___

                              Total   ___

Send above materials to: Name _____
                          Institution _____
                          Address _____
                          _____

Send updating information to: Name _____
(if different from above)   Address _____
                               _____

Send invoice to: Name _____
(if different from above) Address _____
                          _____

*The lower price applies to colleges, non-profit organizations, and Pennsylvania companies.