Basic Statistics and Data Manipulation

HP System 45 Desktop Computer

Hewlett-Packard Desktop Computer Division
3404 East Harmony Road, Fort Collins, Colorado 80525

HP Computer Museum
www.hpmuseum.net

For research and education purposes only.

Table of Contents

Commentary

The introduction of this package of statistical programs marked the beginning of an effort by our statistical laboratory and the Hewlett-Packard Company to develop a unified set of statistical programs, most of which use the same data input and manipulation routines. The data analyst can input data to a common data base and perform several different types of data manipulations prior to subjecting the data to one or more of several statistical analysis procedures. We are confident that you will soon discover that the data manipulation programs available in this package (such as Transform, Recode and Sort) provide you with a set of sophisticated operations. These operations on your data set will undoubtedly help you get more insight into the structure of your data set.

As indicated above, the Data Manipulation portion of this package will be used as a "front-end" program to a number of other statistical routines. At the time of this writing there are six programs packages available which use the Data Manipulation package. These packages are:

1. **Regression Analysis Methods** which includes three programs:

    a. Stepwise Regression and other selection procedures

    b. Multiple Linear Regression

    c. Polynomial Regression

2. **Forecasting and Graphics** which includes several routines to smooth data sets and several routines to forecast into the future.
3. **Statistical Graphics** which includes several graphical procedures which can be used on one or more variables.
4. **General Statistics** which is a collection of parametric and nonparametric tests of hypotheses as well as a collection of routines which allow you to determine significance levels of various statistics.
5. **Non-Linear Regression** which allows you to fit non-linear models to multivariate data and also includes routines for residual analysis.
6. **Analysis of Variance** which allows you to analyze results from the most commonly used statistical designs. It includes multiple comparison plots, interaction plots and an orthogonal polynomial routine.

The programs in this package were written by David Deane of the Statistical Laboratory at Colorado State University.

Thomas J. Boardman
Professor-in-charge
Statistical Laboratory
Colorado State University
Fort Collins, CO 80523

Basic Statistics and Data Manipulation

Introduction

Description

This set of programs allows you to enter a data matrix into memory and to then perform various operations on the data. The data entry may be made via keyboard, mass storage, card reader, or other devices. The operations on the data set include editing, transforming, recoding, sorting, creating subfiles, naming, storing and listing.

Features include a provision for missing data values, a provision for incorporating a subfile structure, the ability to store the data matrix and related information, the ability to join two data sets, error detection, the ability to correct many possible errors, and YES/NO keys to speed program use.

More specific objectives and features are listed with the instructions for the individual programs.

System Configuration

9845B or 9845C (Opt. 001) Desktop Computer

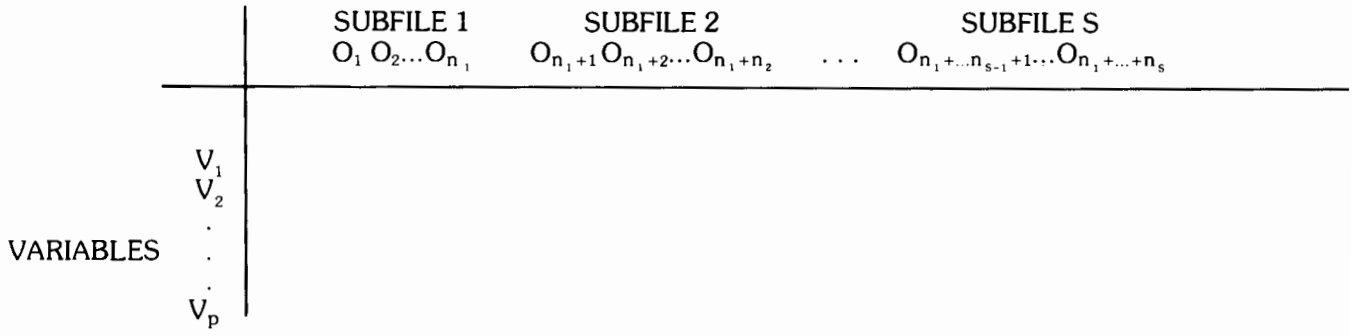
Special Considerations

Data Matrix Configuration

The data matrix incorporated in this program should be thought of as a p-by-n array whose columns correspond to observations and whose rows correspond to variables as shown below.

		OBSERVATIONS				
		O_1	O_2	O_3	...	O_n
VARIABLES	V_1				...	
	V_2				...	
	V_3				...	
	.	.	.	.		.
	.	.	.	.		.
	.	.	.	.		.
	V_p				...	

Subfiles may be created, in which case the structure becomes only slightly more complex as shown below.



Limitations

The programs in this package have been designed to run on any 9845B regardless of memory size. They will operate on a maximum of 50 variables and a maximum of 20 subfiles. A total of 1500 data values may be input. So, for example, if the data set has two variables, 750 observations may be input. To take advantage of the larger memory sizes which are available, some program modifications are necessary. These are explained in the Appendix.

Data File Configuration

The scratch file on the program medium, "DATA", and any files created to hold stored data and related information are configured as follows.

The data file is broken into logical records of 1000 bytes each. The first logical record is a "header file", which contains information pertinent to the data set stored in the remaining logical records. The header file contains the following information (variables):

- data set title (T$)
- number of observations (No)
- number of variables (Nv)
- variable names (Vn$(X))
- number of subfiles (Ns)
- subfile names (Sn$(*))
- subfile characterizations (Sc(*))

The remaining logical records contain D(*,*) – the data matrix.

Missing Values

If a data value is missing, it can be so designated by pressing the special function key labeled M.V., when entering the data from the keyboard. The number used to designate a missing value is −9999999.99999. The justification for this number is that (besides seeming unlikely to occur as a legitimate data point) it is easily picked out in a listing of the data. If the input is to be via cards, it may be more desirable to designate a missing value by a more easily typed number, for example, by 0 if zero is not a legitimate data point. The zeros could then be converted to the missing value recognized by the programs via a user-defined transformation.

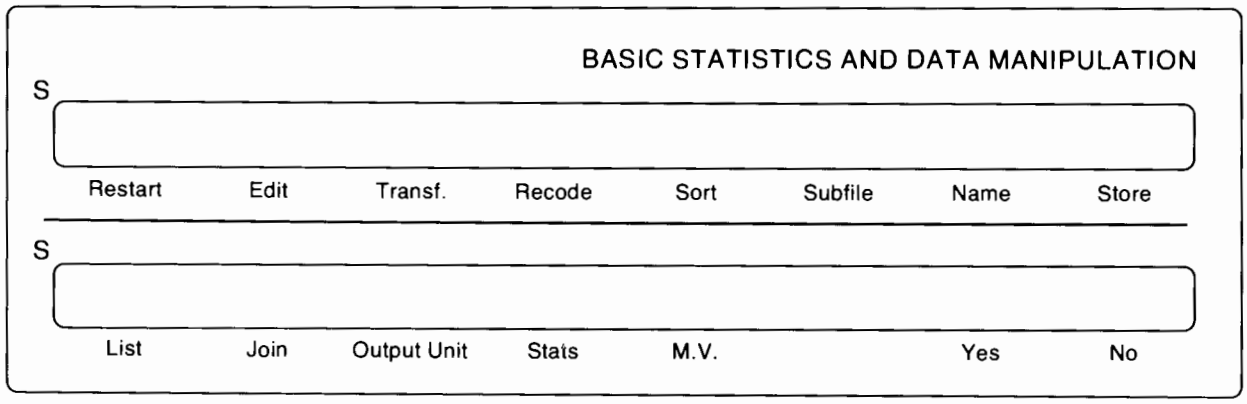
Incorrect Responses

If a response outside the range of plausible responses is input from the keyboard, a message so stating is displayed for about three seconds. Program execution is resumed by asking the question or a previous question again.

If a plausible response is given, but it is not correct, one of three possibilities exists. First, if an incorrect value has been entered for a data point, it may be corrected using the EDIT program. Second, in many cases, responses to several questions are printed on the CRT and then a question such as "Is the above information correct?" is asked. This allows any of the printed information to be changed. Third, if a YES/NO question is answered incorrectly or if the above options are not offered, the program can be re-started by pressing STOP, then RUN.

Special Function Keys

Most of the operations are selected by pressing the appropriate special function key, shown below. Two of the keys those labeled Yes and No, may be used as responses to any question requiring a yes or no answer.



Part Numbers

The component parts of the BASIC STAT AND DATA MANIPULATION package may be ordered separately:

User Instructions	09845-15101
Cartridges	09845-15104
Key Overlay:	7120-6711

Start

Object of Program

This program allows you to place a data matrix into memory. The data may be entered from the keyboard, from cards, or from some other input device. Conversely, the data may have been entered previously and stored in the program medium's scratch file ("DATA") or in a user created file on a tape cartridge, flexible disk or hard disk. In this case, the function of this program is to retrieve the previously stored data and place it into memory so that further operations can be performed. After the data is in memory, a listing option is provided.

Special Considerations

- The prompts concerning the data medium and program medium may cause confusion. The word "medium" is used since the set of programs making up the "BASIC STATISTICS AND DATA MANIPULATION" package may be tape, floppy disk or hard disk. Thus, the "program medium" refers to either the cartridge or the disk on which the programs making up this package are stored. Conversely, the "data medium" refers either to the cartridge or to the disk on which the file containing the data matrix resides. In some cases, the program medium and the data medium may be the same. However, this cannot be determined by the program and hence, the prompts are still displayed to make sure the correct medium is in the correct device.
- When entering data from the keyboard, an option to enter data one case at a time is offered. The following example will serve to explain this feature. Suppose an investigator has collected four observations on each of three variables. He has the following data matrix:

		Variable		
		1	2	3
Observation	1	10	2	5
	2	11	2	6
	3	9	3	7
	4	9	2	6

He elects to enter the data one case at a time. Then, when the prompt "Observation #, all variables (separated by commas)=?" is displayed, he enters 10, 2, 5 and presses CONT, etc. This allows for quick entry of the data.

- Because of the wide variety of formats that could be used when entering data via the card reader or "other" device, no attempt was made to program in the necessary statements. It will be necessary for you to provide the statements before using the program. The following procedure should be followed:

  For card reader input:

  1. Type: LOAD "START"
  2. Press: EXECUTE
  3. Type: EDITLINE Card
  4. Press: EXECUTE
  5. Type in and store the appropriate statements for card reader input.

  For "other" input:

  1. Type: LOAD "START"
  2. Press: EXECUTE
  3. Type: EDITLINE Other
  4. Press: EXECUTE
  5. Type in and store the appropriate statements for "other" input.

User Instructions

Getting Started

1. With the machine turned on, insert the "BASIC STATISTICS AND DATA MANIPULATION" cartridge into the tape drive.
2. Load the program into memory:
   a. Type: LOAD "AUTOST"
   b. Press: EXECUTE
3. Press: RUN
4. When "Is 'BASIC STATISTICS AND DATA MANIPULATION' overlay placed on keys?" is displayed:
   a. Press: YES when the overlay is in place.

Hard-Copy Options

5. When "HARD-COPY OPTIONS" is printed and "Option number=?" is displayed:

   a. Enter: 1, if no hard copy is desired.
   b. Press: CONT
   c. Go to step 9.

   or

   a. Enter: 2, if hard copy is desired from the machines's internal printer.
   b. Press: CONT
   c. Go to step 9.

   or

   a. Enter: 3, if hard copy is desired from an external printer.
   b. Press: CONT

6. When "Are you using an HP-IB Printer?" is displayed:

   a. Press: YES if an HP-IB Printer is being used.
   b. Go to step 7.

   or

   a. Press: NO if the printer is not HP-IB.
   b. Go to step 8.

7. When "Printer select code, bus address=?" [for example, 7,1] is displayed:

   a. Enter the select code, bus address for the HP-IB printer.
   b. Press: CONT
   c. Go to step 9.

8. When "Printer select code=?" is displayed:

   a. Enter the select code of the printer desired for hard copy.
   b. Press: CONT

Data Entry Modes

9. When "DATA ENTRY MODES" is printed and "Mode number=?" is displayed:

   a. Enter: 1, if data is to be entered from the keyboard.
   b. Press: CONT
   c. Go to step 10.

   or

   a. Enter: 2,if data is to be entered from a cartridge or a disk.
   b. Press: CONT
   c. Go to step 15.

   or

   a. Enter: 3, if data is to be entered via cards.
   b. Press: CONT
   c. Go to step 10.

   or

   a. Enter: 4, if data is to be entered from some other device.
   b. Press: CONT

Creating the Data Set

10. When "Project title for this data set (<= 80 characters)=?" is displayed:

    a. Enter up to 80 characters to be used as the name of the data set.
    b. Press: CONT

11. When "Number of variables=?" is displayed:

    a. Enter the number of variables you want in the data set.
    b. Press: CONT

12. When "Number of observations/variable=?" is displayed:

    a. Enter the number of observations for each variable in the data set.
    b. Press: CONT

13. When "VARIABLE NAMES" is printed and "Variable #I name (<= 10 characters)=?" is displayed:

    a. Enter the name of the Ith variable - it may be up to 10 characters long.
    b. Press: CONT
    c. Repeat steps a and b until all variables have been named, then go to step 14.

14. When "Is above information correct" is displayed:

    a. Press: YES if the variable names are as desired.

    b. Go to step 21.

    or

    a. Press: NO if changes to the variable names are desired.

    b. Go to step 13.

Using an Existing Data Set

15. When "Is data stored on the program medium's scratch file (DATA)?" is displayed:

    a. Press: YES if the data is stored in "DATA", the scratch file on the program tape.

    b. The data and related information are loaded into memory at this point.

    c. Go to step 21.

    or

    a. Press: NO if the data is stored on a user-created file.

16. When file name specifications are printed and "Data file name=?" is displayed:

    a. Enter the data file name (less than or equal to six characters) followed by a colon and the mass storage unit specifier.

    b. Press: CONT

17. When "Was data stored by this program?" is displayed:

    a. Press: YES if the data file was created by the "BASIC STATISTICS AND DATA MANIPULATION" routine or by any routine which stores a header file with the data (see Special Considerations).

    b. Go to step 18.

    or

    a. Press: NO if the file was created by a program which does not store header information along with the data.

    b. Go to step 10.

18. When "Is data medium placed in device?" is displayed:

    a. Press: YES when the data medium is in place.

19. The data and related information are loaded into memory at this point.

20. When "Is program medium replaced in device?" is displayed:

    a. Press: YES when the program medium is in place (or if it was never removed).

Summary of the Data Set

21. At this point a summary of the data set is printed.
22. If the input mode is via keyboard, card reader or "OTHER", go to step 25.
23. When "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the data is desired.

    b. Go to step 1 of the user instructions for the LIST program.

    or

    a. Press: NO if a hard-copy listing is not desired.

Selecting a Program

24. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to the user instructions for the selected key.

Data Input

25. At this point the subprogram "Input" is called. A subprogram is used here for the sole purpose of variable dimensioning.
26. When "DATA INPUT" is printed:

    a. If the data is being entered from the keyboard, go to step 32.

    or

    a. If the data is being entered from cartridge or disk, go to step 32.

    or

    a. If the data is being entered via cards, go to step 40.

    or

    a. If the data is being entered from some other device, go to step 41.

27. When the program notes are printed:

    a. If there is only one variable in the data set, go to step 28.

    or

    a. If there are two or more variables in the data set, go to step 29.

28. When the heading is printed and "Observation #I" is displayed:

    a. Enter the value of the Ith observation.

b. Press: CONT

c. Repeat steps a and b until all observations have been entered.

d. Go to step 42.

29. When "Enter data one case at a time, that is, by observation?" is displayed:

a. Press: YES if you desire to key in the values of all variables for a given observation at once.

b. Go to step 31.

or

a. Press: NO if you desire to key in the value of one variable at a time for a given observation.

30. When "Observation #I, Variable #J" is displayed:

a. Enter the value of the Ith observation, Jth variable.

b. Press: CONT

c. Repeat steps a and b until all variables for all observations are entered.

d. Go to step 42.

31. When "Observation #I, all variables (separated by commas)=?" is displayed:

a. Enter variable 1, comma, variable 2, comma, . . ., variable P for Observation #I.

b. Press: CONT

c. Repeat steps a and b until all variables for all observations are entered.

d. Go to step 42.

32. When "Are MISSING VALUES denoted by −9999999.99999?" is displayed:

a. YES if missing values were specified by pressing MISSING VALUE (M.V.) when the data was originally entered, or if there are no missing values.

b. Go to step 34.

or

a. Press: NO if missing values are specified by some other value.

33. When "Missing value=?" is displayed:

a. Enter the value that has been used to specify a missing data point.

b. Press: CONT

34. When "Is data in proper configuration, that is, variables = rows, observations = columns?" is displayed:

    a. Press: YES if the data matrix contains a variable in each row and an observation in each column.

    b. Go to step 36.

    or

    a. Press: NO if the data matrix is in a different configuration.

35. When "Data stored as contiguous array with observation = rows, variables = columns?" is displayed:

    a. Press: YES if the data matrix contains an observation in each row and a variable in each column.

    b. Go to step 36.

    or

    a. Press: NO if the data is configured in a different manner.

    b. Execution stops - the data set is unacceptable for use by this program and therefore must be keyed in.

36. When "Is data medium placed in device?" is displayed:

    a. Press: YES when data medium is in place.

37. At this point the data is loaded into memory.

38. When "Is program medium replaced in device?" is displayed:

    a. Press: YES when program medium is in place (or if it was never removed).

39. At this point the data is stored on the scratch file of the program medium (DATA). Go to step 42.

40. See Special Considerations.

41. See Special Considerations.

42. When "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the data is desired.

    b. Go to step 1 of the user instructions for LIST.

    or

    a. Press: NO if no listing is desired.

Selecting a Program

43. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys defined on the "BASIC STATISTICS AND DATA MANIPULATION" overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

The data listed below provides an example of most of the operations that can be performed by this package. There are five variables:

1. average monthly temperature (°C)
2. monthly production (kg.)
3. number of working days
4. number of people on the payroll
5. monthly water use (litres)

There are 17 observations, one per month.

```
*****************************************************************************
*                          DATA MANIPULATION                                *
*****************************************************************************
                                    DATA

Data file name:
Number of observations: 17
Number of variables: 4

Variables names:
   1. AGE
   2. HEIGHT
   3. WEIGHT
   4. PRESSURE


                                    DATA

     Variable # 1   Variable # 2   Variable # 3   Variable # 4
OBS#
   1
         24.00000       172.50000        66.10000        76.60000
   2
         41.00000       196.00000        55.60000        74.20000
   3
         28.00000       189.80000        69.80000        72.40000
   4
         46.00000       149.50000        71.70000        83.30000
   5
         46.00000       145.60000        66.50000        77.80000
   6
         52.00000       155.30000        56.80000        72.70000
   7
         45.00000       158.80000        67.60000        82.90000
   8
         43.00000       172.10000        68.70000        82.20000
   9
         40.00000       182.30000        77.20000        82.40000
  10
         34.00000       187.40000        70.90000        81.70000
  11
         27.00000       181.80000        74.60000        73.50000
  12
         28.00000       183.30000        74.00000        75.50000
  13
         30.00000       165.40000        62.30000        87.20000
  14
         40.00000       177.60000        59.80000        79.30000
  15
         36.00000       202.90000        78.90000        78.20000
  16
         44.00000       147.50000        64.50000        75.10000
  17
         37.00000       176.30000        80.50000        81.00000
```

Edit

Object of Program

This program is designed to allow you to perform a variety of editing procedures on the data matrix. The editing capabilities include: the changing of an incorrect data value, the deletion of a variable, the deletion of an observation, the addition of an observation, the insertion of an observation (if the data is ordered), and the addition of a variable. All of the above operations can be performed repeatedly; for example, three variables could be deleted in succession. After the data matrix has been edited, you are given the option of listing the edited data.

Special Considerations

Order of Corrections

As stated in the program note printed on the screen, the data is renumbered after deletions or insertions are performed. For this reason, if more than one deletion (insertion) is to be performed, it is recommended that the highest-numbered observation (or variable) be deleted, then the next highest-numbered, etc. For example, if observations three and eight are to be deleted, then it is recommended to delete observation eight first, then observation three. Notice that if observation three were deleted first, the subsequent renumbering would move observation eight to position seven. The recommendation is meant to alleviate confusion which may occur due to the renumbering.

Subfiles

Insertions or deletions of observations will affect the content of subfiles which are in existence at the time of editing; for example, if subfile one consists of the first 10 observations while subfile two consists of the last 20 and if observation five is deleted, then observation ten (formerly numbered 11) will have jumped from subfile two to subfile one. Thus, it may be necessary to change the subfile structure after editing. It is recommended that subfiles be created only after all editing has been performed.

Methods and Formulae

The data matrix is redimensioned into a row vector to facilitate the shuffling of elements necessitated by the editing operations. The vector contains all the observations of variable one, followed by the observations of variable two, etc. When an observation is inserted, for example, the elements of the data vector are shuffled one at a time to make room for the incoming observation. Similarly, when an observation is deleted, the remaining observations are "packed" together so that the resultant data vector has no "holes" between observations.

User Instructions

Press: The special function key labeled Edit.

Correcting a Data Value

1. When "Correct a data value?" is displayed:

   a. Press: NO if there are no data values present that need to be changed.

   b. Go to step 6.

   or

   a. Press: YES if there is a data point in memory that needs to be changed.

2. When "Observation number=?" is displayed:

   a. Enter the observation number of the data value that is to be changed.

   b. Press: CONT

   c. If there is only one variable in the data set, go to step 4.

3. When "Variable number=?" is displayed:

   a. Enter the variable number of the data value that is to be changed.

   b. Press: CONT

4. When "Old value='.' –correct value=?" is displayed:
   (The old value is displayed so one can be sure that the correction is being made as anticipated.)

   a. Enter the correct data value.

   b. Press: CONT

   c. A hard-copy note of the change is printed.

5. When "Correct another value?" is displayed:

   a. Press: YES if there is another data point in memory that needs to be changed.

   b. Go to step 2.

   or

   a. Press: NO if there are no more changes to be made to existing data points.

6. If there is only one variable in the data set, go to step 11.

Deleting a Variable

7. When "Delete a variable?" is displayed:

   a. Press: NO if no variables are to be removed from the data set.

   b. Go to step 11.

   or

   a. Press: YES if you desire to remove one or more variables from the data set.

8. When the program note is printed, and "Number of the variable to be deleted=?" is displayed:

   a. Enter the highest-numbered variable to be deleted. For example, if variables 2 and 7 are to deleted, enter 7 first (since the variables are renumbered after a deletion, and if 2 were deleted first, then variable 7 would be referred to as variable 6).

   b. Press: CONT

   c. A hard-copy note of the deletion is printed.

9. If there is only one variable remaining in the data set, go to step 11.

10. When "Delete another variable?" is displayed:

    a. Press: YES if you desire to remove another variable from the data set.

    b. Go to step 8.

    or

    a. Press: NO if no more variables are to be deleted from the data set.

Deleting an Observation

11. When "Delete an observation?" is displayed:

    a. Press: NO if no observations are to be removed from the data set.

    b. Go to step 14.

    or

    a. Press: YES if you desire to remove one or more observations from the data set.

12. When the program note is printed and "Number of the observation to be deleted=?" is displayed:

    a. Enter the highest-numbered observation to be deleted; for example, if observations 2 and 5 are to be deleted, enter 5 first (since the observations are renumbered after each deletion, if 2 were deleted first, then observation 5 would become observation 4).

    b. Press: CONT

    c. A hard-copy note of the deletion is printed.

13. When "Delete another observation?" is displayed:

    a. Press: YES if it is desired to remove another observation from the data set.

    b. Go to step 12.

    or

    a. Press: NO if no more observations are to be removed from the data set.

Adding an Observation

14. If the addition of one more observation will exceed available memory, go to step 23.
15. When "Add an observation?" is displayed:

    a. Press: NO if no observations are to be added to the end of the data set or inserted into the data set.

    b. Go to step 23.

    or

    a. Press: YES if you desire to add more observations to the end of the data set or to insert observations into the data set.

16. When "Are observations ordered, that is, should additions be inserted?" is displayed:

    a. Press: YES if the data follows some type of order and if you desire to insert an observation, for example, between observations 4 and 5.

    b. Go to step 19.

    or

    a. Press: NO if the observations are to be added to the end of the data set.

17. When "How many observations are to be added?" is displayed:

    a. Enter the number of observations to be added to the "end" of the data set.

    b. Press: CONT

18. When "Observation #'I', Variable #'J'=?" is displayed:

    a. Enter the data point corresponding to the Jth variable of the Ith observation.

    b. Press: CONT

    c. A hard-copy note of the addition is printed.

    d. Repeat steps a-c until data has been entered for all variables of each of the observations added.

    e. Go to step 23.

19. When the program note is printed and "Insertion to precede observation #?" is displayed:

    a. Enter the number of the observation which the insertion will precede. For example, if an observation is to be inserted between observations 8 and 9, enter 9. Since the observations are renumbered after each insertion, insertions should be made between the highest-numbered observations first. For example, if insertions are to be made between observations 4 and 5 as well as between observation 8 and 9, the latter should be performed first, that is, 9 should be entered first.

    b. Press: CONT

20. When "Observation #'I', Variable #'J'=?" is displayed:

    a. Enter the data point corresponding to variable J of the inserted observation.

    b. Press: CONT

    c. A hard-copy note of the insertion is printed.

    d. Repeat steps a-c until data has been entered for all variables.

21. If the addition of one more observation will exceed program limitations, go to step 23.

22. When "Insert another observation?" is displayed:

    a. Press: YES if you desire to insert another observation.

    b. Go to step 19.

    or

    a. Press: NO if no more observations are to be inserted.

Adding a Variable

23. If the addition of a variable will exceed program limitations, go to step 29.

24. When "Add a variable?" is displayed:

    a. Press: NO if no variables are to be added from the keyboard.

    b. Go to step 29.

    or

    a. Press: YES if you desire to add one or more variables from the keyboard.

25. When "Variable name (<= 10 characters)" is displayed:

    a. Enter the name of the variable to be added - it may be up to 10 characters long.

    b. Press: CONT

26. When "Variable #'I', Observation #'J'=?" is displayed:

    a. Enter the Jth observation of the variable being added.

    b. Press: CONT

    c. A hard-copy note of the addition is printed.

    d. Repeat steps a-c until all observations have been entered for the added variable.

27. If the addition of one more variable will exceed available memory, go to step 29.

28. When "Add another variable?" is displayed:

    a. Press: YES if you desire to add another variable from the keyboard.

    b. Go to step 25.

    or

    a. Press: NO if no more variables are to be added from the keyboard.

29. When "More corrections?" is displayed:

    a. Press: YES if more corrections to the data matrix are required.

    b. Go to step 1.

    or

    a. Press: NO if no more corrections to the data set are necessary.

30. At this point the corrected data matrix and related information are recorded in file "DATA" of the program medium.

31. "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the corrected data set is desired.

    b. Go to step 1 of the user instructions for LIST.

    or

    a. Press: NO if a hard-copy listing of the data is not required.

32. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

The program EDIT was run in the print all mode. The printout shows a correction, deletion of an observation, and addition of an observation. The edited data was then listed.

```
Correct a data value?
"YES"
Observation number = ?
11
Variable number = ?
2
Old value =  15022 -- Correct value =
?
15024
Observation # 11 Variable # 2 -- correct value =  15024
Correct another value?
"NO"

Delete a variable?
"NO"
Delete an observation?
"YES"
Number of the observation to be deleted = ?
10
Observation # 10 has been deleted, 16 observations remain.
Delete another observation?
"NO"

Add an observation?
"YES"
Are observations ordered, i.e., should additions be inserted?
"NO"
How many observations are to be added?
1
Observation # 17 Variable # 1  =
?
4.2
Observation # 17 Variable # 1  =  4.2
Observation # 17 Variable # 2  =
?
12707
Observation # 17 Variable # 2  =  12707
Observation # 17 Variable # 3  =
?
20
Observation # 17 Variable # 3  =  20
Observation # 17 Variable # 4  =
?
192
Observation # 17 Variable # 4  =  192
Observation # 17 Variable # 5  =
?
3614
Observation # 17 Variable # 5  =  3614

  Total number of observations now =  17

Add a variable?
"NO"
More corrections?
"NO"
List data?
"YES"
```

SAMPLE

OBS#	Variable # 1	Variable # 2	Variable # 3	Variable # 4	Variable # 5
1	14.90000	6396.00000	21.00000	134.00000	3373.00000
2	18.40000	5736.00000	22.00000	146.00000	3110.00000
3	21.60000	6116.00000	22.00000	158.00000	3180.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000
5	26.30000	13313.00000	25.00000	198.00000	3390.00000
6	27.20000	13108.00000	23.00000	194.00000	4287.00000
7	22.20000	10768.00000	20.00000	180.00000	3852.00000
8	17.70000	12173.00000	23.00000	191.00000	3366.00000
9	12.50000	11390.00000	20.00000	195.00000	3532.00000
10	6.40000	15024.00000	22.00000	200.00000	3896.00000
11	13.30000	13114.00000	19.00000	211.00000	3437.00000
12	18.20000	12257.00000	22.00000	203.00000	3324.00000
13	22.80000	13118.00000	22.00000	197.00000	3214.00000
14	26.10000	13100.00000	21.00000	196.00000	4345.00000
15	26.30000	16716.00000	21.00000	205.00000	4936.00000
16	4.20000	14056.00000	22.00000	205.00000	3624.00000
17	4.20000	12707.00000	20.00000	192.00000	3614.00000

Transform

Object of Program

This program allows you to transform one or two variables in the data matrix via 16 pre-specified functions or through a function which you specify. The transformed data may then be treated as a new variable, or it may replace the elements of an existing variable. Hence, transformations on more than two variables may be made iteratively or via a user-defined transformation.

The transformations available are:

1. a X^b+c
2. a log(bX)+c
3. a ln (bX)+c
4. a exp(bX)+c
5. a (b^{cX})
6. a cos(bX)+c
7. a sin(bX)+c
8. $a+\sqrt{\arcsin(bX)+c}$
9. aX+bY+c
10. aX^bY^c
11. a log(bX+cY)
12. a ln (bX+cY)
13. a cos(bX+cY)
14. a sin(bX+cY)
15. PROUND(X,a) [round to specified power of 10]
16. DROUND(X,a) [round to specified no. of digits]
17. User Defined

Special Considerations

Missing Values

None of the 16 pre-specified transformations are applied to missing values. Thus, missing values are unaffected by these transformations. However, this is not necessarily the case with the user-defined transformation. If you define a transformation and there are missing values, you must make provisions to ensure that the transformation is not applied to the missing values (unless, of course, this is desired). This may be accomplished as explained below.

User-Defined Transformation

Up to 10 lines of code may be used to define a transformation, namely lines 1330 through 1339. These lines may be typed in and stored successively prior to pressing 'CONT'. (See user instructions.) The following example shows the form of a typical user-defined transformation. Suppose the data set consists of four variables with 20 observations each. There are missing values, which are to be left unchanged. You desire to form variable five as the sum of the exponentials of variables one and three. The following sequence should be carried out at step three of the TRANSFORM User Instructions.

a. Type: 1330 D(Z,I)=EXP(D(1,I))+EXP(D(3,I))

b. Press: STORE

c. Type: 1331 IF D(1,I)=−9999999.99999 OR D(3,I)=−9999999.99999 THEN D(5,I)=1

d. Press: STORE

e. Press: CONT

The variable Z is used to identify variable in which the result of the transformation is to be stored. Notice that the elements of D(.,.) are the data - the first subscript refers to variable number while the second subscript refers to observation number (and should always be 'I' here). Line 1331 sets D(5,I)=1 if a missing value is present in either of variables one or three. Notice that D(Z,I) or D(5,I) may be used interchangeably since Z has been assigned the value of 5 previously in the program.

User Instructions

Press: The special function key labeled Transf

1. When the program notes and transformations are printed and "Transformation number=?" is displayed:

   a. Enter the number of the transformation which is desired (1-17, inclusive).

   b. Press: CONT

2. If the selected transformation was not 'User defined', that is, not transformation number 17, go to step 4.
3. When the program notes are printed and "Ready to continue?" is displayed:

   a. Type: "1330 D(Z,I)='defined transformation' "; the form of the defined transformation is discussed in Special Considerations.

   b. Press: STORE

   c. Press: YES when ready to continue.

   d. Go to step 11.

4. When the selected transformation is printed and "Number of the variable corresponding to X=?" is displayed:

a. Enter the variable number corresponding to X, where X refers to the variable in the selected transformation.

b. Press: CONT

5. If the selected transformation number is less than 9 or greater than 14, go to step 7.
6. When "Number of the variable corresponding to Y=?" is displayed:

a. Enter the variable number corresponding to Y, where Y refers to the variable in the selected transformation.

b. Press: CONT

7. When "Parameter a=?" is displayed:

a. Enter the value of parameter a, where a refers to the parameter in the selected transformation.

b. Press: CONT

8. If the selected transformation number is greater than 14, go to step 11.
9. When "Parameter b=?" is displayed:

a. Enter the value of parameter b, where b refers to the parameter in the selected transformation.

b. Press: CONT

10. When "Parameter c=?" is displayed:

a. Enter the value of parameter c, where c refers to the parameter in the selected transformation.

b. Press: CONT

11. When "Store transformed data in variable # (<='I')" is displayed:

a. Enter the number of the variable in which the transformed data is to be stored - it may be less than or equal to I.

b. Press: CONT

12. If the variable specified in step 11 existed previously, that is, if the transformed data is being stored in place of previously existing data, go to step 14.
13. When "Variable name (<= 10 characters)=?" is displayed:

a. Enter the name of the variable to contain the transformed data - it may be up to 10 characters long.

b. Press: CONT

14. If the transformation was 'User defined', go to step 16.

15. When "Is above information correct?" is displayed:

    a. Press: NO if a mistake has been made during entry of the information printed.

    b. Go to step 1.

    or

    a. Press: YES if no corrections are to be made.

16. At this point a hard-copy summary of the transformation is made (unless it was user defined), then the transformation is carried out.

17. When "More transformations?" is displayed:

    a. Press: YES if more transformations are desired.

    b. Go to step 1.

    or

    a. Press: NO if no more transformations are desired.

18. At this point the altered data matrix and related information are stored on file "DATA" of the program medium.

19. When "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the data is desired.

    b. Go to step 1 of the user instructions for LIST.

    or

    a. Press: NO if a hard-copy listing of the data is not required.

20. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

Variable five, Water Use, was tranformed, converting litres to U.S. gallons. Transformation 1 was used, with a=0.2642, b=1,c=0.

```
Transformation number = ?
1
Number of the Variable number corresponding to X = ?
5
Parameter a = ?
.2642
Parameter b = ?
1
Parameter c = ?
0
Store transformed data in Variable # ( <= 6 )
?
5
Is above information correct?
"YES"
The following transformation was performed: a*(X^b)+c
  where a =  .2642
        b =  1
        c =  0
        X is Variable # 5
        Transformed data is stored in Variable # 5 (Water Use).


More transformations?
"NO"
List data?
"YES"
```

SAMPLE

OBS#	Variable # 1	Variable # 2	Variable # 3	Variable # 4	Variable # 5
1	14.90000	6396.00000	21.00000	134.00000	891.14660
2	18.40000	5736.00000	22.00000	146.00000	821.66200
3	21.60000	6116.00000	22.00000	158.00000	840.15600
4	25.20000	8287.00000	20.00000	171.00000	870.01060
5	26.30000	13313.00000	25.00000	198.00000	895.63800
6	27.20000	13108.00000	23.00000	194.00000	1132.62540
7	22.20000	10768.00000	20.00000	180.00000	1017.69840
8	17.70000	12173.00000	23.00000	191.00000	889.29720
9	12.50000	11390.00000	20.00000	195.00000	933.15440
10	6.40000	15024.00000	22.00000	200.00000	1029.32320
11	13.30000	13114.00000	19.00000	211.00000	908.05540
12	18.20000	12257.00000	22.00000	203.00000	878.20080
13	22.80000	13118.00000	22.00000	197.00000	849.13880
14	26.10000	13100.00000	21.00000	196.00000	1147.94900
15	26.30000	16716.00000	21.00000	205.00000	1304.09120
16	4.20000	14056.00000	22.00000	205.00000	957.46080
17	4.20000	12707.00000	20.00000	192.00000	954.81880

Recode

Object of Program

This program allows you to assign codes to various categories or classes of data. The categories are intervals along the real number line and 20 of these may be specified. The recoding is done on one variable at a time. The same coding scheme may be used iteratively on successive variables. A summary of the coding intervals, codes, and number of observations assigned to each code is printed as hard copy.

Special Considerations

Coding Schemes

Four coding schemes are available for the sole purpose of eliminating unnecessary entries from the keyboard. If the coding intervals are all of the same length and are contiguous, that is, together they form a connected interval, then the interval construction can be accomplished internally knowing only the interval length and lower limit for the first interval. Similarly, if the intervals are of equal length but noncontiguous, for example, [10,20), [25,35), [35,45), [50,60), then the lower limit of each interval needs to be specified but the upper limit may be computed internally. Hence, the coding schemes are meant only to minimize the amount of information which needs to be entered from the keyboard. Clearly, the coding intervals could all be constructed by requiring you to enter the lower and upper limits for each and every interval (which is necessary, and what is done if the intervals are unequal and non-contiguous).

Same Coding Scheme

The coding is carried out on one variable at a time. However, if you desire to code both variables one and two according to the same coding intervals, these intervals need to be contructed only once. A positive response to the option offered by "Use same coding scheme?" allows variable two to be coded according to the same scheme without constructing the intervals a second time. If, however, you desire to code variable two according to a different scheme, it is possible to construct a second set of coding intervals by giving a negative response to the above prompt.

Brackets

The brackets used to denote the coding intervals are meant to follow their usual mathematical interpretation, that is, the intervals are closed on the left and open on the right.

Observation not in an Interval

If an observation does not fall into any of the coding intervals, its value is not changed during the coding process.

User Instructions

Press: The special function key labeled Recode.

1. When the title is printed and "Store recoded data in variable # (<='I')?" is displayed:
   a. Enter the number of the variable in which the recoded data is to be stored - it may be less than or equal to I.
   b. Press: CONT
2. If the variable specified in step 1 existed previously, that is, if the recorded data is being stored in place of previously existing data, go to step 4.
3. When "Variable name (<= 10 characters)=?" is displayed:
   a. Enter the name of the variable which will contain the recoded data - it may be up to 10 characters long.
   b. Press: CONT
4. When "Number of the variable to be recoded=?" is displayed:
   a. Enter the number of the variable which is to be recoded.
   b. Press: CONT
5. When the data coding schemes are printed and "Option number=?" is displayed:
   a. Enter the option number associated with the desired recoding scheme.
   b. Press: CONT
6. When "Number of recoding intervals to be specified (<=20)=?" is displayed:
   a. Enter the number of categories into which the data will be divided, or equivalently, the number of codes which will be assigned.
   b. Press: CONT
7. If the specified coding scheme does not have equal intervals, go to step 9.
8. When "Length of each interval=?" is displayed:
   a. Enter the common increment to be used to generate the intervals.
   b. Press: CONT
9. If the data recoding scheme is non-contiguous, go to step 11.
10. When "Lower limit of first interval=?" is displayed:
    a. Enter the smallest number which will be assigned code #1.
    b. Press: CONT

11. Note: The following steps, 12 through 14, are repeated the number of times specified in step 6 until all information is input for each of the recoding intervals. Some of the steps are skipped for certain schemes, since the necessary information is computed internally. For scheme #1, only step 14 needs to be repeated. For scheme #2, only steps 13 and 14 need to be repeated. For scheme #3, only steps 12 and 14 need to be repeated. For scheme #4, steps 12-14 need to be repeated.
12. When "Lower limit of interval #'I'=?" is displayed:
    a. Enter the smallest number which will be assigned the code associated with interval #I.
    b. Press: CONT
13. When "Upper limit of interval #'I'=?" is displayed:
    a. Enter the upper bound of numbers which will be assigned the code associated with interval #I.
    b. Press: CONT
14. When "For data falling in interval #'I', code=?" is displayed:
    a. Enter the code which will be assigned to those data values falling in the Ith coding interval.
    b. Press: CONT
    c. Repeat as noted in instruction 11.
15. When "Is above information correct?" is displayed:
    a. Press: NO if a mistake has been made in entering the information.
    b. Go to step 5.

    or

    a. Press: YES if the recoding scheme is as desired.
16. At this point the recoding is carried out and a hard-copy note of the coding is printed.
17. When "Recode more data?" is displayed:
    a. Press: YES if you desire to recode more data.
    b. Go to step 18.

    or

    a. Press: NO if no more recoding is desired.
    b. Go to step 19.

18. When "Use same coding scheme?" is displayed:

    a. Press: YES if the same intervals and codes are desired, that is, if the identical recoding scheme is to be applied to another variable.

    b. Go to steps 1-4, then directly to 16.

    or

    a. Press: NO if a different recoding scheme is desired.

    b. Go to step 1.

19. At this point the altered data matrix and related information are stored on file "DATA" of the program medium.

20. When "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the data is desired.

    b. Go to step 1 of the user instructions for LIST.

    or

    a. Press: NO if no listing is desired.

21. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

Using the original data, variable 5, Water Use, was recoded, and the resulting data was stored as a new variable, 6. The purpose of the recoding operation was to simplify the data, replacing entire intervals by single numbers.

```
Store recoded data in Variable # (<= 6 )
?
6
Variable name (<= 10 characters) = ?
Coded Use
Number of the variable to be recoded = ?
5
Option number = ?
1
Number of recoding intervals to be specified (<=20) = ?
5
Length of each interval = ?
400
Lower limit of first interval = ?
3000
For data falling in interval 1 , code =
?
30
For data falling in interval 2 , code =
?
34
For data falling in interval 3 , code =
?
38
For data falling in interval 4 , code =
?
42
For data falling in interval 5 , code =
?
46
Is above information correct?
"YES"
```

Variable # 5 is recoded into 5 categories, and the recoded values are stored in Variable # 6 , where:

CATEGORY BOUNDS		# OBS	
LOWER	UPPER	CODED	CODE
3000.000	3400.000	8	30.000
3400.000	3800.000	4	34.000
3800.000	4200.000	2	38.000
4200.000	4600.000	2	42.000
4600.000	5000.000	1	46.000

Recode more data?
"YES"
Use same coding scheme?

Recode more data?
YES

Recode more data?
NO

SAMPLE

OBS#	Variable # 1	Variable # 2	Variable # 3	Variable # 4	Variable # 5	Variable # 6
1	14.90000	6396.00000	21.00000	134.00000	3373.00000	30.00000
2	18.40000	5736.00000	22.00000	146.00000	3110.00000	30.00000
3	21.60000	6116.00000	22.00000	158.00000	3180.00000	30.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000	30.00000
5	26.30000	13313.00000	25.00000	198.00000	3390.00000	30.00000
6	27.20000	13108.00000	29.00000	194.00000	4287.00000	42.00000
7	22.20000	10769.00000	20.00000	180.00000	3852.00000	38.00000

```
 8
        17.70000     12173.00000        23.00000       191.00000      3365.00000
           30.00000
 9
        12.50000     11390.00000        20.00000       195.00000      3532.00000
           34.00000
10
         6.90000     12707.00000        20.00000       192.00000      3614.00000
           34.00000
11
         6.40000     15022.00000        22.00000       200.00000      3896.00000
           38.00000
12
        13.30000     13114.00000        19.00000       211.00000      3437.00000
           34.00000
13
        18.20000     12257.00000        22.00000       203.00000      3324.00000
           30.00000
14
        22.80000     13118.00000        22.00000       197.00000      3214.00000
           30.00000
15
        26.10000     13100.00000        21.00000       196.00000      4345.00000
           42.00000
16
        26.30000     16716.00000        21.00000       205.00000      4936.00000
           46.00000
17
         4.20000     14056.00000        22.00000       205.00000      3624.00000
           34.00000
```

Sort

Object of Program

This program allows the data matrix, or subfiles thereof, to be sorted according to the values of one variable. For example, suppose an investigator has five observations of three variables, say height, weight and age and wanted to arrange the observations in ascending order according to age. This is accomplished by sorting the data matrix according to variable three.

Special Considerations

Subfile Structure Options

If subfiles are ignored, the entire data set will be sorted and, in the process, the composition of the subfiles is subject to change. The option of sorting certain subfiles may be used to sort a single subfile or a set of successive subfiles according to one variable. The option of sorting all subfiles may be used to sort every subfile. The options of sorting certain subfiles and sorting all subfiles treat each subfile as if it were a separate data set. Thus, the sort is done with respect to one subfile at a time.

What Happens

It is important to note that entire observations are moved when the sort is carried out. Thus, referring to the example given in Object of Program, a person's height and weight remain with the person's age as shown below.

Original Matrix

		Height	Weight	Age
	1	72	170	21
	2	70	165	25
OBSERVATION	3	69	150	20
	4	70	165	25
	5	73	160	19

Matrix Sorted By Age

VARIABLE

		Height	Weight	Age
	1	73	160	19
	2	69	150	20
OBSERVATION	3	72	170	21
	4	70	165	25
	5	70	165	25

User Instructions

Press: The special function key labeled Sort.

1. At this point the subprogram "Sort" is called. A subprogram is used here for the sole purpose of variable dimensioning.
2. When "Number of the Variable on which to sort=?" is displayed:

   a. Enter the number of the variable on which the matrix will be sorted; that is, the variable whose observations will be arranged in ascending order.

   b. Press: CONT

3. If the data set contains no subfiles, go to step 7.
4. When the subfile structure options are printed and "Option number=?" is displayed:

   a. Enter 1 if you desire to ignore the subfile structure, that is, to lump all the observations together for the sort. Note that this may shift observations from subfile to subfile.

   b. Press: CONT

   c. Go to step 7.

   or

   a. Enter 2 if you desire to sort a set of one or more successive subfiles, for example, to sort subfiles 3 through 5. The sort will be done by subfile: subfile 3 will be sorted, then 4, and finally 5.

   b. Press: CONT

   c. Go to step 5.

   or

   a. Enter 3 if you desire to sort all of the subfiles. Each subfile will be sorted as if it were a separate data set.

   b. Press: CONT

   c. Go to step 7.

5. When "Number of first subfile=?" is displayed:

   a. Enter the number of the first subfile to be sorted.

   b. Press: CONT

6. When "Number of last subfile=?" is displayed:

   a. Enter the number of the last subfile to be sorted. All subfiles between and including the first and last specified will be sorted as though they were separate data sets.

   b. Press: CONT

7. At this point the desired sorting is performed, a hard-copy note of the sort is printed, and the sorted data is stored on file "DATA" of the program medium.
8. When "List data?" is displayed:

   a. Press: YES if a hard-copy listing of the sorted data is desired.

   b. Go to step 1 of the user instructions for LIST.

   or

   a. Press: NO if no listing is desired.

9. When "SELECT ANY KEY" is displayed:

   a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

   b. Go to step 1 of the user instructions for the selected key.

Example

The data was sorted by variable 2. That is, the observations were arranged so that monthly production was in ascending order. Notice that the other variables are carried by the sort operation, so that although the order of the observations has changed, each observation remains intact.

```
    Data set:
                                                SAMPLE
    has been arranged in ascending order according to Variable # 2
    (Subfiles ignored)

      Variable # 1    Variable # 2    Variable # 3    Variable # 4    Variable # 5
         Variable # 6
OBS#
    1
           10.40000      5736.00000        22.00000       146.00000      3110.00000
              30.00000
    2
           21.60000      6116.00000        22.00000       158.00000      3180.00000
              30.00000
    3
           14.90000      6396.00000        21.00000       134.00000      3373.00000
              30.00000
    4
           25.20000      8287.00000        20.00000       171.00000      3293.00000
              30.00000
    5
           22.20000     10768.00000        20.00000       180.00000      3852.00000
              30.00000
    6
           12.50000     11390.00000        20.00000       195.00000      3532.00000
              34.00000
```

Data set:

SAMPLE

has been arranged in ascending order according to Variable # 2
(Subfiles ignored)

OBS#	Variable # 1	Variable # 2	Variable # 3	Variable # 4	Variable # 5	Variable # 6
1	18.40000	5736.00000	22.00000	146.00000	3110.00000	30.00000
2	21.60000	6116.00000	22.00000	158.00000	3180.00000	30.00000
3	14.90000	6396.00000	21.00000	134.00000	3373.00000	30.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000	30.00000
5	22.20000	10768.00000	20.00000	180.00000	3852.00000	38.00000
6	12.50000	11390.00000	20.00000	195.00000	3532.00000	34.00000
7	17.70000	12173.00000	23.00000	191.00000	3366.00000	30.00000
8	18.20000	12257.00000	22.00000	203.00000	3324.00000	30.00000
9	6.90000	12707.00000	20.00000	192.00000	3614.00000	34.00000
10	26.10000	13100.00000	21.00000	196.00000	4345.00000	42.00000
11	27.20000	13108.00000	23.00000	194.00000	4287.00000	42.00000
12	13.30000	13114.00000	19.00000	211.00000	3437.00000	34.00000
13	22.80000	13116.00000	22.00000	197.00000	3214.00000	30.00000
14	26.30000	13313.00000	25.00000	198.00000	3390.00000	30.00000
15	4.20000	14056.00000	22.00000	205.00000	3624.00000	34.00000
16	6.40000	15022.00000	22.00000	200.00000	3896.00000	38.00000
17	26.30000	16716.00000	21.00000	205.00000	4936.00000	46.00000

Subfiles

Object of Program

This program allows you to specify subfiles or logical groupings of the observations. This may be accomplished by entering the number of observations in each subfile or by entering the observation number of the first observation in each subfile. Names for the subfiles are entered in both cases. A third option allows you to destroy the existing subfile structure.

Special Considerations

Use of Subfiles

Subfiles may be created in order to specify logical groupings of observations. A subfile structure allows you to consider each subfile as a separate data set or to lump all the subfiles together and analyze the overall data set. For example, suppose an investigator wished to measure several variables on 50 trout. He would like to analyze the data separately for each of the three varieties of the trout. He could form three separate data sets and do the individual analyses, then later join the three sets together for the overall analysis. However, since the same variables were measured on each variety of fish, this situation is well-handled by specifying a subfile for each variety. The subfile structure options make it possible to do the analysis by subfile as well as for the overall data set.

Editing and Sorting

Certain operations in the editing and sorting programs may cause observations to move from one subfile to another. To avoid undesired results such as this, it is recommended that subfiles be specified after any editing or sorting has been carried out.

User Instructions

Press: The special function key labeled Subfile.

1. When the subfile characterization options are printed and "Option number=?" is displayed:

   a. Enter 1 if you desire to specify subfiles by entering the number of observations in each subfile.

   b. Press: CONT

   c. Go to step 2.

   or

   a. Enter 2 if you desire to specify subfiles by entering the number of the first observation in each subfile.

   b. Press: CONT

   c. Go to step 2.

   or

a. Enter 3 if you desire to destroy the existing subfiles structure, that is, lump all the data together.

b. Press: CONT

c. Go to step 8.

2. When "Number of subfiles (<= 20)=?" is displayed:

a. Enter the number of subfiles which will be specified.

b. Press: CONT

3. When "Name of subfile #'I' (<= 10 characters)=?" is displayed:

a. Enter the name of the Ith subfile - if may be up to 10 characters long.

b. Press: CONT

4. If the option number you chose was 2, go to step 6.

5. When "Subfile #'I', number of observations=?" is displayed:

a. Enter the number of observations which will be in the Ith subfile.

b. Press: CONT

c. Repeat steps 3 and 5 until all subfiles have been specified.

d. Go to step 7.

6. When "Subfile 'I', number of first observation=?" is displayed:

a. Enter the number of the first observation in the Ith subfile.

b. Press: CONT

c. Repeat steps 3 and 6 until all subfiles have been specified.

7. When "Is above information correct?" is displayed:

a. Press: NO if a mistake has been made while entering the required information.

b. Go to step 2.

or

a. Press: YES if the subfile characterizations are as desired.

8. At this point a hard-copy record of the subfiles will be printed, and the information will be printed on file "DATA" of the program medium.

9. When "SELECT ANY KEY" is displayed:

a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

b. Go to step 1 of the user instructions for the selected key.

Example

The original data (before sorting and recoding) was grouped into two subfiles. The first 12 observations make up the first subfile (Fiscal year '76) and the remaining observations make up the second subfile (Fiscal year '77).

```
Option number = ?
2
Number of subfiles ( <=10 ) = ?
2
Name of Subfile # 1 ( <= 10 characters ) =
?
FY'76
Name of Subfile # 2 ( <= 10 characters ) =
?
FY'77
Subfile # 2 ; number of first observation =
?
13
Is above information correct?
"YES"
Subfile name:     beginning observation--number of observations
 1. FY'76                              1                        12
 2. FY'77                             13                         5

SELECT ANY KEY
```

Name

Object of Program

This program allows you to rename variables and/or to rename subfiles. These names are then stored, along with the data, on the program medium's scratch file ("DATA").

User Instructions

Press: The special function key labeled Name.

1. When "Rename data set?" is displayed:

   a. Press: NO to leave the data set name unchanged.

   b. Go to step 3.

   or

   a. Press: YES to change the data set name.

2. When "Name of data set (<= 80 characters)=?" is displayed:

   a. Enter the project title for the data set.

   b. Press: CONT

3. When the title is printed and "Name variables?" is displayed:

   a. Press: NO if no variables are to be renamed.

   b. Go to step 9.

   or

   a. Press: YES if you desire to rename some variables.

4. When "Name beginning with variable #?" is displayed:

   a. Enter the number of the first variable to be renamed.

   b. Press: CONT

5. When "Name ending with variable #?" is displayed:

   a. Enter the number of the last variable to be renamed.

   b. Press: CONT

6. When "Name of variable #'I' (<= 10 characters)=?" is displayed:

   a. Enter the name of the Ith variable - it may be up to 10 characters long.
   b. Press: CONT
   c. Repeat parts a and b until all specified variables have been renamed.

7. When "Is above information correct?" is displayed:

   a. Press: NO if a mistake has been made while entering variable names.
   b. Go to step 1.

   or

   a. Press: YES if the variable names are as desired.

8. At this point a hard-copy list of all the current variable names will be printed. If there are no subfiles in the data set, go to step 15.
9. When "Name subfiles?" is displayed:

   a. Press: NO if no subfiles are to be renamed.
   b. Go to step 15.

   or

   a. Press: YES if you desire to rename some subfiles.

10. When "Name beginning with subfile #?" is displayed:

    a. Enter the number of the first subfile to be renamed.
    b. Press: CONT

11. When "Name ending with subfile #?" is displayed:

    a. Enter the number of the last subfile to be renamed.
    b. Press: CONT

12. When "Name of subfile #'I' (<= 10 characters)=?" is displayed:

    a. Enter the name of the Ith subfile - it may be up to 10 characters long.
    b. Press: CONT
    c. Repeat parts a and b until all specified subfiles have been named.

13. When "Is above information correct?" is displayed:

    a. Press: NO if a mistake has been made while entering subfile names.
    b. Go to step 10.

14. At this point a hard-copy of the current subfile names is printed.
15. The revised information is stored on file "DATA" of the program medium.
16. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

The name of variable four was changed from 'Payroll' to 'Payroll (#)'.

```
Rename data set?
"NO"
Name variables?
"YES"
Name beginning with Variable # ?
4
Name ending with Variable # ?
4
Name of Variable # 4 ( <= 10 characters ) =
?
Payroll(#)
Is above information correct?
"YES"

SAMPLE

CURRENT VARIABLE NAMES
 1. Temp(C)
 2. Production
 3. Days
 4. Payroll(#)
 5. Water Use
 6. Coded Use

Name subfiles?
"NO"
SELECT ANY KEY
```

Store

Object of Program

This program allows you to store the data matrix and related information in a file so that it may be retrieved at a later date for further analysis. The program also allows you to specify the file name.

Special Considerations

Use of Program

The store feature will be useful in two different situations. First, if an investigator has a data set which he may want to analyze further at a later date, he may store it and retrieve it later via the BASIC STATISTICS AND DATA MANIPULATION routine. Secondly, if several people have access to the data input programs, it becomes mandatory that each be able to store his data set in a unique place. Note that if only one person uses the routine on one data set it is unnecessary to use the store feature since the data and related information are kept in "DATA" - the scratch file on the program medium.

Protecting Existing Data

The existence of a file is checked in the program in an attempt to avoid the accidental loss of existing data. Thus, when a file is specified to receive the data, an attempt is made to ensure that you are not accidentally storing the new data in a file which you did not know existed.

User Instructions

Press: The special function key labeled Store.

1. When the program notes are printed and "Name of data file=?" is displayed:

   a. Enter the name of the file in which the data matrix and related information are to be stored - refer to the program note for the proper form of the file name.

   b. Press: CONT

2. When "Is data medium placed in device 'I'?" is displayed:

   a. Press: YES when the medium on which the data and related information are to be stored is in place.

3. When "Does this file already exist?" is displayed:

   a. Press: YES if the file has been previously created (and is of the form specified in the Special Considerations) and you desire to "print over" the information that is currently in the file.

   b. Go to step 4.

   or

   a. Press: NO if the file has not been previously created.

NOTE

This step is an attempt to prevent you from accidentally writing over an existing file. If the response to the question in step 3 is negative and the program finds a file by this name, a note will be displayed and control will pass to step 1.

4. At this point, the file will be created if necessary and the data matrix along with the related information will be stored in the specified file. A hard-copy note of the storage will be printed.
5. When "Is program medium replaced in device?" is displayed:

   a. Press:YES when the program medium is in place (or if it was never removed).

6. When "SELECT ANY KEY" is displayed:

   a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

   b. Go to step 1 of the user instructions for the selected key.

List

Object of Program

This program allows you to obtain a listing of the data matrix. The listing will appear on the device that has been specified for hard-copy in the START routine or in the output unit routine.

User Instructions

Press: The special function key labeled List.

1. When the data listing options are printed and "Option number=?" is displayed:

   a. Enter 1 if you desire to obtain a hard-copy listing of all data - the listing will be by observation.

   b. Press: CONT

   c. Go to step 6.

   or

   a. Enter 2 if you desire to obtain a hard-copy listing of data by variable - it may be all the data or just a set of successive variables.

   b. Press: CONT

   c. Go to step 2.

   or

   a. Enter 3 if you desire to obtain a hard-copy listing of data by observation - the observations are to be successive but need not be the entire data set.

   b. Press: CONT

   c. Go to step 4.

2. When "List beginning with variable #?" is displayed:

   a. Enter the number of the first variable to be included in the list.

   b. Press: CONT

3. When "Ending with Variable #?" is displayed:

   a. Enter the number of the last variable to be included in the list (all variables between the first and last variables specified will be included in the list).

   b. Press: CONT

   c. Go to step 6.

4. When "List beginning with Observation #?" is displayed:

   a. Enter the number of the first observation to be included in the list.

   b. Press: CONT

5. When "Ending with Observation #?" is displayed:

   a. Enter the number of the last observation to be included in the list.

6. At this point the hard-copy listing of the data will be printed.
7. When "SELECT ANY KEY" is displayed:

   a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

   b. Go to step 1 of the user instructions for the selected key.

Join

Object of Program

This program allows you to join or combine two data sets into a single unit. One data set must be in memory and the other data set must have been previously stored by the BASIC STATISTICS AND DATA MANIPULATION program. Two options are available. First, observations may be added together (if both sets have the same number of variables). Second, variables may be added together (if both sets have the same number of observations).

Special Considerations

Adding Observations

Suppose data on six variables was gathered in each of the 52 weeks in 1975, analyzed, and stored on an auxiliary data tape. Suppose the same variables were measured in 1976, analyzed, and stored. If the investigator is interested in lumping the two sets of data together for an overall analysis, he may use the Add Observations option of the joining routine. One set of data must be retrieved via the START routine. Then, after pressing the JOIN key, the second set may be retrieved and the joining carried out. Notice that the variables must be in the same order in the two data sets.

Adding Variables

Suppose an investigator measured five variables on each of 50 mice in an experiment. These were analyzed and stored on disk. Later, he realized that three more variables were of interest. He measured these variables on the mice in the same order as before and analyzed them. All eight variables measured on each mouse could be combined into a single data set via the joining routine .

Subfiles

If variables are added, the subfile structure assigned to the resultant data set is the subfile structure of data set #1, that is, the data set that is in the machine prior to the joining operation. If observations are added, the following procedures are employed: 1) If no subfiles exist in either data set, the resultant set has no subfiles. 2) If data set #1 has no subfiles, but data set #2 does, then a subfile named "SET #1" is created which consists of data set #1 and the subfiles of data set #2 remain unchanged. 3) If data set #1 contains subfiles, but data set #2 does not, then a subfile named "SET #2" is created which consists of data set #2 and the subfiles of data set #1 remain unchanged. 4) If both data sets contain subfiles, all of the subfiles of data set #1 are retained and as many subfiles of data set #2 are retained as possible - the upper limit of total subfiles for the resultant set being 20.

User Instructions

Press: The special function key labeled Join.

1. At this point the subprogram "Join" is called. A subprogram is used here for the sole purpose of variable dimensioning.
2. When the joining options are printed and "Option number=?" is displayed:

   a. Enter 1 if you desire to add **variables** from an external data medium to those which are currently in memory.

   b. Press: CONT

   c. Go to step 3.

   or

   a. Enter 2 if you desire to add **observations** from an external data medium to those which are currently in memory.

   b. Press: CONT

3. When the requirements are printed and "OK to continue?" is displayed:

   a. Press: NO if the requirements have not been met.

   b. Go to step 16.

   or

   a. Press: YES if you feel that the requirements have been satisfied.

4. When "Project title for the combined data set (<= 80 characters)=?" is displayed:

   a. Enter up to 80 characters to be used as the name of the joined data set.

   b. Press: CONT

5. When the program note is printed and "File name of data set #2=?" is displayed:

   a. Enter the name of the data file in which the data to be joined to the set already in memory resides. Refer to the program note for the proper form of the file name.

   b. Press: CONT

6. When "Is data set #2 medium placed in device 'I'?" is displayed:

   a. Press: YES when the medium on which data set #2 resides in place.

7. At this point, the number of variables, number of observations, and subfile structures are printed for data sets 1 and 2 as well as for the resultant set. The requirements are checked internally and if they have not been met, control is passed to step 16.

8. When "OK to continue?" is displayed:

   a. Press: NO if, after reviewing the summary, you realize that the resultant data set will not turn out as expected (remember, the subfile structure may be changed at a later time).

   b. Go to step 12.

   or

   a. Press: YES if the resultant data set will be as expected.

9. At this point a summary of the variable names for data sets 1 and 2 as well as for the resulting joined data set are printed.

10. When "OK to continue?" is displayed:

    a. Press: NO if, after reviewing the summary, you desire to abort the joining operation (remember, the variable names may be changed at a later time).

    b. Go to step 12.

    or

    a. Press: YES if the joining operation is to be carried out.

11. At this point the joining operation is performed and a hard-copy summary of the resultant data set is printed.

12. When "Is program medium replaced in device?" is displayed:

    a. Press: YES when the program medium is in place (or if it was never removed).

13. If the joining operation was aborted, go to step 16.

14. At this point the resultant data set and related information are printed on file "DATA" of the program medium.

15. When "List data?" is displayed:

    a. Press: YES if a hard-copy listing of the resultant data set is desired.

    b. Go to step 1 of the user instructions for LIST.

    or

    a. Press: NO if no listing is desired.

16. When "SELECT ANY KEY" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to step 1 of the user instructions for the selected key.

Example

Three observations were entered by keyboard and stored in file "TDATA5". Then the JOIN program added these observations to the end of the original data set to form a new set.

```
Option number = ?
2
OK to continue?
"YES"
Project title for combined data set (<= 80 characters) = ?
SAMPLE
File name of data set #2 = ?
TDATA5:F8
Is data set #2 medium placed in device F8
?
"YES"
OK to continue?
"YES"
OK to continue?
"YES"
Is program medium replaced in device?
"YES"
List data?
"YES"
```

SAMPLE

OBS#	Variable # 1	Variable # 2	Variable # 3	Variable # 4	Variable # 5
1	14.90000	6396.00000	21.00000	134.00000	3373.00000
2	18.40000	5736.00000	22.00000	146.00000	3110.00000
3	21.60000	6116.00000	22.00000	158.00000	3180.00000
4	25.20000	8287.00000	20.00000	171.00000	3293.00000
5	26.30000	13313.00000	25.00000	198.00000	3390.00000
6	27.20000	13100.00000	23.00000	194.00000	4287.00000
7	22.20000	10768.00000	20.00000	180.00000	3852.00000
8	17.70000	12173.00000	23.00000	191.00000	3366.00000
9	12.50000	11390.00000	20.00000	195.00000	3532.00000
10	6.90000	12707.00000	20.00000	192.00000	3614.00000
11	6.40000	15022.00000	22.00000	200.00000	3896.00000
12	13.30000	13114.00000	19.00000	211.00000	3437.00000
13	18.20000	12257.00000	22.00000	203.00000	3324.00000
14	22.80000	13118.00000	22.00000	197.00000	3214.00000
15	26.10000	13100.00000	21.00000	196.00000	4345.00000
16	26.30000	16716.00000	21.00000	205.00000	4936.00000
17	4.20000	14056.00000	22.00000	205.00000	3624.00000
18	25.30000	9315.00000	20.00000	183.00000	3356.00000
19	12.40000	11298.00000	19.00000	203.00000	4205.00000
20	18.60000	14653.00000	21.00000	189.00000	4256.00000

Output Unit

Object of Program

This program allows you to change the device on which the hard-copy output will be printed, or conversely, to specify that no hard-copy is desired.
In lieu of this routine, the hard-copy device may be changed at any time by the following procedure:

Non HP-IB Printer

1. Type: Hc=
2. Key in the select code of the printer desired.
3. Press: EXECUTE
4. Type: Hcbus=999
5. Press: EXECUTE

HP-IB Printer

1. Type: Hc=
2. Key in select code of the printer desired.
3. Press: EXECUTE
4. Type: Hcbus=
5. Key in the bus address of the HP-IB device.
6. Press: EXECUTE

User Instructions

Press: The special function key labeled Output Unit.

1. When the hard-copy options are printed and "Option number=?" is displayed:

   a. Enter 1 if no hard-copy is desired.
   b. Press: CONT
   c. Go to step 5.

   or

   a. Enter 2 if a hard-copy is desired from the computer's internal printer.
   b. Press: CONT
   c. Go to step 5.

   or

   a. Enter 3 if a hard-copy is desired from an external printer.
   b. Press: CONT

2. When "Are you using an HP-IB Printer?" is displayed:
   (See Appendix for explanation of HP-IB.)

   a. Press: YES if an HP-IB printer is being used.
   b. Go to step 3.

   or

   a. Press: NO if the printer is not HP-IB.
   b. Go to step 4.

3. When "Printer select code, bus address=? [for example, 7,1]" is displayed:

   a. Enter the select code, bus address for the HP-IB printer.
   b. Press: CONT
   c. Go to step 5.

4. When "Printer select code=?" is displayed:

   a. Enter the select code of the printer desired for hard-copy.
   b. Press: CONT

5. When "SELECT ANY KEY" is displayed:

   a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.
   b. Go to step 1 of the user instructions for the selected key.

Basic Statistics

Object of Program

This program computes a variety of summary statistics for data which was entered via the BASIC STATISTICS AND DATA MANIPULATION program set. The statistics may be computed by subfile or for the entire data set (ignoring subfiles). Basic statistics which are computed include: number of observations, number of missing values, sum, mean, variance, standard deviation, coefficient of skewness, coefficient of kurtosis, coefficient of variation, standard error of the mean, and confidence interval of the mean. An option is available to obtain the correlation matrix. Order statistics computed include: the maximum, the minimum, range, and midrange. Additional order statistics which may be obtained include: the median, 25th percentile, 75th percentile, Tukey's middlemeans, and user-specified percentiles.

Special Considerations

Hard-copy Output

If a hard-copy of the statistics is not being made, the program halts occasionally so that you may study the results. In this case, it is necessary only to press 'CONT' to continue program execution.

Additional Order Statistics

If the option to obtain additional order statistics (Tukey's middlemeans and percentiles) is exercised, the data matrix is sorted and the observations of each variable are arranged in ascending order. At the end of the program the original data matrix is re-loaded. Thus, if the program is aborted, that is, if another key is pressed before the re-loading can occur, the data matrix will be in the sorted state. Hence, if the portion of the program used to calculate additional order statistics is accessed, abortion of the program should be discouraged.

Methods and Formulae

Ith Variable

Let N(J) be the number of observations of the Jth variable in the data set or subfile, whichever is applicable. Let D(I,J) be the Jth observations of the Ith variable. The following formulas are computed for the Ith variable.

- Sum: $S(I)= \sum_{J=1}^{N(I)} D(I,J)$

- Mean: $M(I) = \frac{S(I)}{N(I)}$

- Variance: $V(I) = \frac{\sum_{J=1}^{N(I)} (D(I,J))^2 - N(I)(M(I))^2}{N(I)-1}$

- Standard deviation: $Sd(I) = (V(I))^{1/2}$

- Second moment: $M_0(I) = \frac{(N(I)-1)V(I)}{N(I)}$

- Skewness: $Sk(I) = \frac{\sum_{J=1}^{N(I)} (D(I,J))^3 - 3M(I) \sum_{J=1}^{N(I)} (D(I,J))^2 + (2M(I))^3}{(M_0(I))^{3/2}\, N(I)}$

- Kurtosis:

$$K(I) = \frac{\sum_{J=1}^{N(I)} (D(I,J))^4 - 4M(I) \sum_{J=1}^{N(I)} (D(I,J))^3 + 6(M(I))^2 \sum_{J=1}^{N(I)} (D(I,J))^2 - 3(M(I))^4 N(I)}{(M_0(I))^2 N(I)} - 3$$

t-value

Let C be the confidence coefficient for a confidence interval on the mean. The following operations are used to obtain the desired t-value.

$$P = \frac{1-\frac{c}{100}}{2}$$

$$V = \left(\ln\left(\frac{1}{p^2}\right)\right)^{1/2}$$

$$X = 2.5155174+.802853V+.010328V^2$$

$$Y = 1 + 1.432788V + .189269V^2 + .001308V^3$$

$$Z = V-\frac{X}{Y}$$

$$M = N(J)-1$$

- Then the desired t-value is:

$$T = Z + \frac{Z^3+Z}{4M} + \frac{5Z^5+16Z^3+3Z}{96M^2} + \frac{3Z^7+19Z^5+17Z^3-15Z}{384M^3}$$

$$+ \frac{79Z^9+776Z^7+1482Z^5-1920Z^3-945Z}{92160M^4}$$

- Standard error: $Se(I) = \frac{(V(I))^{1/2}}{(N(I))^{1/2}}$

- Confidence interval on mean: $M(I) \pm T(Se(I))$

- Coefficient of variation: $Cv(I)=\left|\frac{(V(I))^{1/2}}{M(I)}\right|(100)$

- Correlations: Suppose we have the following data matrix:

		OBSERVATION				
		1	2	3	4	5
VARIABLE	1	5	M	3	4	5
	2	6	7	M	6	4
	3	1	3	2	1	1

An M denotes a missing value. When computing the correlation between variables 1 and 2, we discard observations 2 and 3 since variable 1 is missing a data value for observation 2 and variable 2 is missing the data value for observation 3. However, when computing the correlation between variables 1 and 3, we need only discard observation 2. Similarly, the correlation between 2 and 3 is computed by discarding only observation 3. Hence, the correlations may be based on different numbers of observations. An observation is thrown out if and only if a data value from that observation is missing from one of the two variables for which the correlation is being computed. With this in mind, let N(I,J) be the number of observations used to compute the correlation between variables I and J. Then, the correlation is:

$$C(I,J)=\frac{\sum_{K=1}^{N(I,J)} D(I,K)D(J,K)-\frac{\sum_{K=1}^{N(I,J)} D(I,K)\sum_{K=1}^{N(I,J)} D(J,K)}{N(I,J)}}{\left[\sum_{K=1}^{N(I,J)}(D(I,K))^2-\frac{\left(\sum_{K=1}^{N(I,J)} D(I,K)\right)^2}{N(I,J)}\right]^{\frac{1}{2}}\left[\sum_{K=1}^{N(I,J)}(D(J,K))^2-\frac{\left(\sum_{K=1}^{N(I,J)} D(I,K)\right)^2}{N(I,J)}\right]^{\frac{1}{2}}}$$

Ranges and Percentiles

Let M(I) be the largest data value of the Ith variable, m(I) be the smallest data value of the Ith variable.

1. Range: $R(I) = M(I) - m(I)$

2. Midrange: $Mr(I) = \frac{M(I)+m(I)}{2}$

3. The percentiles are computed as follows: Let P be the percentile in question. If P*N(I) is an integer, the P(I)=(D(I,P*N(I))+D(1,Q))/2, where Q is the next integer value between P*N(I) and the observation index of the median. If P*N(I) is not an integer, the P(I)=D(I,N(I)*P+Q) where

$$Q = \begin{cases} 1 \text{ if } P \leq 50 \\ -1 \text{ if } P > 50 \end{cases}$$

The median refers to the 50th percentile.

Tukey's Middlemeans

- Midmean: Mm (I) =1/N Σ all observations between (and including, if applicable) 25th and 75th percentile.
- Trimean: Tm(I) =¼ (25th percentile + 2(median) + 75th percentile).
- Midspread: Ms(I) = 75th percentile - 25th percentile.

User Instructions

Press: The special function key labeled Stats.

1. If there is only one variable in the data set, go to step 5.
2. When "Summary statistics on all variables?" is displayed:

   a. Press: YES if summary statistics are desired for all of the variables in the data set.

   b. Go to step 5.

   or

   a. Press: NO if summary statistics are desired for only one variable or a set of successive variables in the data set.

3. When "Number of first variable=?" is displayed:

   a. Enter the number of the first variable for which summary statistics are desired.

   b. Press: CONT

4. When "Number of last variable=?" is displayed:

   a. Enter the number of the last variable for which summary statistics are desired. Summary statistics will be calculated for all variables between and including the first and last specified.

   b. Press: CONT

5. If the data set contains no subfiles, go to step 9.

6. When the subfile structure options are printed and "Option number=?" is displayed:

   a. Enter 1 if you desire to ignore the subfile structure, that is, to lump all the observations together and calculate summary statistics on the data set as a whole.

   b. Press: CONT

   c. Go to step 9.

   or

   a. Enter 2 if you desire to calculate summary statistics for only one subfile or for each of a set of adjacent subfiles; for example, for subfiles 3 through 5. The statistics will be calculated by subfile, that is, calculated as though each subfile were a separate data set.

   b. Press: CONT

   c. Go to step 7.

   or

   a. Enter 3 if you desire to calculate summary statistics for all of the subfiles. The statistics will be calculated as though each subfile were a separate data set.

   b. Press: CONT

   c. Go to step 9.

7. When "Number of first subfile=?" is displayed:

   a. Enter the number of the first subfile for which summary statistics will be calculated.

   b. Press: CONT

8. When "Number of last subfile=?" is displayed:

   a. Enter the number of the last subfile for which summary statistics will be calculated. Statistics will be calculated for all subfiles between and including the first and last subfiles specified as though they were separate data sets.

   b. Press: CONT

9. At this point a heading and basic statistics will be output. If no hard-copy printer has been specified, the program will pause occasionally to allow examination of the output and "Press 'CONT' when ready." will be displayed. To resume execution, press: CONT.

10. When "Confidence coefficient for confidence interval on the mean . . ." is displayed:

    a. Enter the confidence coefficient or confidence level to be used in constructing a confidence interval on the mean; for example, enter 95 for a 95% confidence on the mean.

    b. Press: CONT

11. At this point several more statistics will be output. If the output is to the screen, it will be necessary to press 'CONT' to resume program execution after examining the output.

12. If summary statistics are being computed for a single variable, go to step 15.

13. When "Correlation matrix?" is displayed:

    a. Press: NO if correlations are not desired.

    b. Go to step 15.

    or

    a. Press: YES to obtain a matrix of correlations among the variables.

14. At this point the correlation matrix will be computed and output. If output is to the screen, it will be necessary to press 'CONT' to resume program execution after examining the correlations.

15. At this point several order statistics are computed and output.

16. When "More order statistics (Tukey's middlemeans & percentiles)?" is displayed:

    a. Press: NO if no further order statistics are desired.

    b. Go to step 21.

    or

    a. Press: YES if Tukey's middlemeans (see Methods and Formulae), the 25th, 50th, and 75th percentiles, are desired.

17. At this point the data matrix is sorted by variable. Then, several order statistics are computed and output.

18. When "Other percentiles?" is displayed:

    a. Press: NO if no further percentiles are desired.

    b. Go to step 21.

    or

    a. Press: YES if additional percentiles are desired.

19. When "Press 'NO' when finished - Percentile=?" is displayed:

    a. Enter the percentile which is desired; for example, enter 90 to obtain the 90th percentile. The number entered must be between 1 and 99, inclusive.

    b. Press: CONT

    or

    a. Press: NO if no more percentiles are desired.

    b. Go to step 21.

20. At this point, the desired percentile is computed for each variable under consideration and the results are output. Go to step 19.

21. If summary statistics are being computed for more than one subfile, steps 9-20 are repeated for each subfile.

22. If the additional order statistics were computed for the data set (or any subfile), the original data matrix is re-loaded at this point.
23. When "SELECT ANY KEY?" is displayed:

    a. Press: Any of the keys identified on the BASIC STATISTICS AND DATA MANIPULATION overlay.

    b. Go to the start of the user instructions.

Example

When the basic statistics key (Stats) is pressed, basic statistics and order statistics are printed, first for the entire data set, and then for individual subfiles.

```
*****************************************************************************
                          SUMMARY STATISTICS                                 *
*                            ON DATA SET:                                     *
*                               SAMPLE                                        *
*****************************************************************************


-----------------------------------------------------------------------------
(SUBFILES IGNORED)
-----------------------------------------------------------------------------

                              BASIC STATISTICS

VARIABLE      # OBSERVATIONS    # MISS. VALUES            SUM             MEAN
Temp(C)                  20                0       366.50000         18.32500
Production               20                0    232643.00000      11632.15000
Days                     20                0       425.00000         21.25000
Payroll                  20                0      3751.00000        187.55000
Water Use                20                0     73590.00000       3679.50000
-----------------------------------------------------------------------------

                                                      COEF OF           COEF OF
VARIABLE            VARIANCE     STANDARD DEV.       SKEWNESS          KURTOSIS
Temp(C)             51.93987           7.20693         -.51175           -.87415
Production     9253846.44947        3042.01355         -.64356           -.42042
Days                2.19737           1.48235          .55329            .23443
Payroll           421.62895          20.53361        -1.37170            .95821
Water Use      242672.26316         492.61777          .99654            .11191
-----------------------------------------------------------------------------

                              STANDARD ERROR    95 % CONFIDENCE INTERVAL ON MEAN
VARIABLE      COEF VARIATION      OF THE MEAN       LOWER LIMIT       UPPER LIMIT
Temp(C)              39.32841          1.61152          14.95123          21.69877
Production           26.15177        680.21491       10208.09757       13056.20243
Days                  6.97578           .33146          20.55607          21.94393
Payroll              10.94834          4.59145         177.93764         197.16236
Water Use            13.38817        110.15268        3448.89172        3910.10828
-----------------------------------------------------------------------------

                            CORRELATION MATRIX

                Production        Days     Payroll    Water Use
Temp(C)          -.1165685    .2535951   -.1443165    .1428234
Production                    .1245049    .8729479    .6347921
Days                                     -.0272339   -.1524027
Payroll                                              .4291423


-----------------------------------------------------------------------------
```

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
Temp(C)	27.20000	4.20000	23.00000	15.70000
Production	16716.00000	5736.00000	10980.00000	11226.00000
Days	25.00000	19.00000	6.00000	22.00000
Payroll	211.00000	134.00000	77.00000	172.50000
Water Use	4936.00000	3110.00000	1826.00000	4023.00000

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
Temp(C)	18.50000	12.90000	25.25000
Production	12482.00000	10041.50000	13215.50000
Days	21.00000	20.00000	22.00000
Payroll	194.50000	181.50000	201.50000
Water Use	3484.50000	3340.00000	4050.50000

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
Temp(C)	19.29000	18.78750	12.35000
Production	12303.30000	12055.25000	3174.00000
Days	21.20000	21.00000	2.00000
Payroll	193.50000	193.00000	20.00000
Water Use	3544.00000	3589.87500	710.50000

VARIABLE	10-th percentile
Temp(C)	6.65000
Production	6256.00000
Days	19.50000
Payroll	152.00000
Water Use	3197.00000
VARIABLE	90-th percentile
Temp(C)	26.30000
Production	14837.50000
Days	23.00000
Payroll	205.00000
Water Use	4316.00000

```
*************************************************************************************
                                SUMMARY STATISTICS                                   *
*                                  ON DATA SET:                                      *
*                                    SAMPLE                                          *
*************************************************************************************
```

Subfile: FY'76

BASIC STATISTICS

VARIABLE	# OBSERVATIONS	# MISS. VALUES	SUM	MEAN
Temp(C)	12	0	212.60000	17.71667
Production	12	0	128130.00000	10677.50000
Days	12	0	257.00000	21.41667
Payroll	12	0	2170.00000	180.83333
Water Use	12	0	42330.00000	3527.50000

VARIABLE	VARIANCE	STANDARD DEV.	COEF OF SKEWNESS	COEF OF KURTOSIS
Temp(C)	50.45242	7.10299	-.24164	-1.10813
Production				
10346097.9091 M7D.5D				
3216.53508	-.47514	-1.24452		
Days	2.99242	1.72986	.53169	-.47442
Payroll	563.60606	23.74039	-.78573	-.59264
Water Use	113867.00000	337.44185	.94917	.15129

VARIABLE	COEF VARIATION	STANDARD ERROR OF THE MEAN	95 % CONFIDENCE INTERVAL ON MEAN LOWER LIMIT	UPPER LIMIT
Temp(C)	40.09212	2.05046	13.20247	22.23086
Production	30.12442	928.53370	8633.28011	12721.71989
Days	8.07718	.49937	20.31728	22.51605
Payroll	13.12833	6.85326	165.74549	195.92118
Water Use	9.56603	97.41107	3313.04397	3741.95603

CORRELATION MATRIX

	Production	Days	Payroll	Water Use
Temp(C)	-.2644635	.4396054	-.1782662	-.0077412
Production		.1294077	.9312213	.6450021
Days			.0040583	.0388568
Payroll				.4892023

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
Temp(C)	27.20000	6.40000	20.80000	16.80000
Production	15022.00000	5736.00000	9286.00000	10379.00000
Days	25.00000	19.00000	6.00000	22.00000
Payroll	211.00000	134.00000	77.00000	172.50000
Water Use	4287.00000	3110.00000	1177.00000	3698.50000

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
Temp(C)	18.05000	12.90000	23.70000
Production	11781.50000	7341.50000	13111.00000
Days	21.50000	20.00000	22.50000
Payroll	191.50000	164.50000	196.50000
Water Use	3413.50000	3329.50000	3733.00000

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
Temp(C)	18.01667	18.17500	10.80000
Production	11405.50000	11003.87500	5769.50000
Days	21.16667	21.37500	2.50000
Payroll	187.16667	186.00000	32.00000
Water Use	3452.00000	3472.37500	403.50000

VARIABLE	10-th percentile
Temp(C)	6.90000
Production	6116.00000
Days	20.00000
Payroll	146.00000
Water Use	3180.00000
VARIABLE	90-th percentile
Temp(C)	25.20000
Production	13114.00000
Days	23.00000
Payroll	198.00000
Water Use	3852.00000

Subfile: FY'77

BASIC STATISTICS

VARIABLE	# OBSERVATIONS	# MISS. VALUES	SUM	MEAN
Temp(C)	8	0	153.90000	19.23750
Production	8	0	104513.00000	13064.12500
Days	8	0	168.00000	21.00000
Payroll	8	0	1581.00000	197.62500
Water Use	8	0	31260.00000	3907.50000

VARIABLE	VARIANCE	STANDARD DEV.	COEF OF SKEWNESS	COEF OF KURTOSIS
Temp(C)	60.11125	7.75314	-.91030	-.30986
Production	4953615.26714	2225.67187	-.06985	-.38572
Days	1.14286	1.06904	-.75000	-.50000
Payroll	65.41071	8.08769	-.76833	-.75334
Water Use	380730.85714	617.03392	.33651	-1.17044

VARIABLE	COEF VARIATION	STANDARD ERROR OF THE MEAN	95 % CONFIDENCE INTERVAL ON MEAN LOWER LIMIT	UPPER LIMIT
Temp(C)	40.30225	2.74115	12.75397	25.72103
Production	17.03652	786.89384	11202.91630	14925.33370
Days	5.09069	.37796	20.10602	21.89398
Payroll	4.09244	2.85943	190.86170	204.38830
Water Use	15.79102	218.15443	3391.50799	4423.49201

CORRELATION MATRIX

	Production	Days	Payroll	Water Use
Temp(C)	.0040190	-.0844548	-.4503797	.2109508
Production		.4515051	.4876946	.6262209
Days			.2643635	-.3473772
Payroll				.2501814

ORDER STATISTICS

VARIABLE	MAXIMUM	MINIMUM	RANGE	MIDRANGE
Temp(C)	26.30000	4.20000	22.10000	15.25000
Production	16716.00000	9315.00000	7401.00000	13015.50000
Days	22.00000	19.00000	3.00000	20.50000
Payroll	205.00000	183.00000	22.00000	194.00000
Water Use	4936.00000	3214.00000	1722.00000	4075.00000

TUKEY'S HINGES

VARIABLE	MEDIAN	25-th %-ile	75-th %-ile
Temp(C)	20.70000	15.30000	25.70000
Production	13109.00000	11777.50000	14354.50000
Days	21.00000	20.50000	22.00000
Payroll	200.00000	192.50000	204.00000
Water Use	3914.50000	3340.00000	4300.50000

TUKEY'S MIDDLEMEANS

VARIABLE	MIDMEAN	TRIMEAN	MIDSPREAD
Temp(C)	21.22500	20.60000	10.40000
Production	13132.75000	13087.50000	2577.00000
Days	21.25000	21.12500	1.50000
Payroll	199.75000	199.12500	11.50000
Water Use	3860.25000	3867.37500	960.50000

VARIABLE	10-th percentile
Temp(C)	4.20000
Production	9315.00000
Days	19.00000
Payroll	183.00000
Water Use	3214.00000
VARIABLE	90-th percentile
Temp(C)	26.10000
Production	14653.00000
Days	22.00000
Payroll	205.00000
Water Use	4345.00000

Appendix

HP-IB

"HP-IB" stands for Hewlett-Packard Interface Bus and is our version of IEEE Standard 488,1975 which is a universal standard interface. In this program package it is used to refer to the interface between the computer and the 9872A Plotter. The select code may be found on the interface card. The bus address is found on the back of the plotter itself in binary form. For further information consult the Programming or Operating and Service Manual for the plotter.

Mass Storage Changes

To change the mass storage to a different tape drive, or to a disk, follow these instructions.

1. Type: MASS STORAGE IS ":unit specifier"
2. Press: EXECUTE

Some examples are ":T15" for the right hand side tape drive and ":F8" for the flexible disk master.

Changes Necessary for Larger Data Sets

CAUTION

INCREASING THE DATA HANDLING CAPACITY MAY CAUSE SEVERAL PROBLEMS. FIRST, IF THE NUMBER OF VARIABLES IS INCREASED, DATA WHICH WAS STORED WITH THE ORIGINAL ROUTINES WILL NOT BE DIRECTLY ACCESSIBLE BY THE ALTERED ROUTINES. SECONDLY, IF THE NUMBER OF DATA POINTS IS INCREASED, THERE MAY NOT BE ENOUGH ROOM ON THE PROGRAM TAPE TO STORE THE ENLARGED DATA SET. TO FIND OUT, PROCEED AS FOLLOWS:

1. Make sure nothing of value is in the scratch file "DATA". If there is, use the STORE routine to save it.
2. Type: PURGE "DATA"
   Press: EXECUTE
3. Compute the size of data file you need. Let # = 464+20*n where n is the maximum number of variables you wish the programs to handle.
4. TYPE: CREATE "DATA",2+Nv*No*8DIV #, # where # is the number calculated in step 3.
5. Press: EXECUTE
6. If you obtain an error, you must transfer all programs to a larger storage medium in order to enlarge the data set.

To take advantage of the larger memory sizes available on the 9845B, certain lines in the Basic Statistics and Data Manipulation routines must be changed.

For example, to change the maximum number of data points from 1500 to 3000, and the maximum number of variables from 50 to 100 in file START, use the following procedure:

1. Type: LOAD "START"
2. Press: EXECUTE
3. Type: EDITLINE Com
4. Press: EXECUTE
5. By editing, make the line read:

   Com: COM D(1,3000), T$[80], Vn$(100) [10], Sn$(20)[10], Nv, Mnv, No, Mno, Sc(20), Hc, Hcbus, Ns, Mns
   Press: STORE

6. Do the same for the line labeled Scom
7. Type: EDITLINE Obs
8. By editing, make the line read:

   Obs: Mno = 3000

9. Press: STORE
10. Type: RE-STORE "START"
11. Press: EXECUTE

The following table shows all lines which need to be changed to handle more variables and/or observations.

The lines that are labeled Com or Scom should be changed as follows:

COM D(1,m), T$([80], Vn$(n) [10], Sn$(20) [10], Nv, Mnv, Mno, Sc(20), Hc, Hcbus, Ns, Mns

where m refers to the maximum number of data points and n refers to the maximum number of variables.

The other lines, e.g., L1, L2, etc. should be made to read exactly as shown in the table. The symbol, #, seen in the table next to file STORE should be replaced by the number 464 + 20 * n where n is the maximum number of variables you specified.

File	Lines
START	Com
	Scom
	Obs: Mno=m
LIST	Com
EDIT	Com
BASIC	Com
	L1: INTEGER L(n)
	L2: DIM M(n), V(n), Tf(n)
TRANSF	Com
SUBFIL	Com
NAME	Com
HCOPY	Com
STORE	Com
	L1: IF X=1 THEN CREATE F$, 2+Nv*No*8 DIV #, #
JOIN	Com
	Scom
SORT	Com
	Scom
RECODE	Com

Part No. 09845-15101
Rev. C

Printed in U.S.A.
May 6, 1980