MICROSTAT
RELEASE 2.0

TABLE OF CONTENTS

Analysis Programs

***** Appendices *****

FOREWORD

Microstat consists of a library of statistical programs that perform most common statistical tests and procedures. It was developed for "real life" applications using fairly large data sets. Computational algorithms were selected for accuracy and speed. This manual describes the contents of Microstat, Release 2.0 and includes many new features (most of which, by the way, were suggested by previous users of Microstat...we do listen to you and value your feedback).

One of the most important and unique aspects of Microstat is its file orientation. All of the programs that require data read it from data files created by the Data Management Subsystem (DMS). This method is superior to direct (keyboard) data entry for several reasons. The most obvious advantage is that the data may be listed, verified and, if necessary, edited prior to use. With direct entry, there may be no error recovery procedure or, worse yet, an error may never be detected. With the data stored in a file, it can be analyzed by several different programs and new files can be created by partitioning and merging existing data files. Furthermore, the data can be transformed, ranked, sorted and cases added or deleted.

The system has been "human engineered" to be flexible and failsafe in operation, yet easy to use. Its algorithms have been selected with the utmost care so that users can have complete confidence in the analyses that result from the system. (In a review of Microstat in the March 16, 1981 issue of **Infoworld** the author examined the regression program and found that "...MICROSTAT outperforms several mainframe programs that were tested".) While Microstat was designed for rigorous use in a business and research environment, it is sufficiently easy to use that it can be used by students as well.

The purpose of this manual is to show you how to use the Microstat system to its maximum potential. The system was designed to be used by following instructions and prompts displayed on the monitor (i.e., CRT) as the programs are executed. Indeed, you could run all of the programs without reading this manual, **but** you will miss many of the features of Microstat if you do. The manual provides supplementary discussion along with examples, test data and sample output. In many cases, the sample output has a file lable that references common statistics texts or journal articles so that you can compare the output of Microstat to those produced on "mainframe" computers. We think you will be pleasantly surprised as the results.

One thing that the manual is not and that is a statistics textbook. While Microstat would be very useful in a statistics course, it is assumed that the user does have an elementary knowledge of the statistical procedures and terminology used.

Ecosoft, Incorporated, makes no representations or warranties, implied or otherwise, with respect to the contents hereof. Neither Ecosoft nor its authorized agents assume responsbility and shall have no liability, consequential or otherwise, of any kind arising from the use of this manual, program material or any part thereof. Further, Ecosoft reserves the right to make changes in its software, manuals and related material without the obligation of Ecosoft or its agents to notify any persons of such changes.

SYSTEM REQUIREMENTS

Although Microstat is now available for two different operating systems and three versions of Basic, this manual refers to Release 2.0 Microstat for use with Digital Research's CP/M operating system and Microsoft's Basic-80 interpreter (Relase 5.03 or later). Since the user's system must hold the operating system, the interpreter and the program in memory with some memory left for data, we recommend 48K (or more) of memory. **Note:** even though Microstat 2.0 has more statistical tests and features than before, we have reduced program size by over 20K compared with earlier versions. This means, of course, more memory is now available for data than with earlier versions of Microstat.

Data Size. One of the most-often asked questions we get is how large can a data set be in Microstat. The answer depends upon the specific procedure you are using. Where possible, we have redesigned those programs that previously needed the entire file in memory to function properly. The regression analysis, for example, now calculates how much memory is needed and, if the file won't fit into memory, it is "paged" into memory as needed. If the residuals are desired, it will make a second pass on the data. On the other hand, the sort program still requires the file to reside in memory.

The amount of data that can be handled is also a function of the precision of the numbers used in the data file. Basic-80 permits either single or double-precision numbers; the former requiring 4 bytes of memory per number and the latter 8 bytes. Everything else being equal, using single-precision numbers in a file will permit about twice as much data to be processed as double-precison numbers. Microstat allows you to determine whether a file will contain single or double-precison numbers (Release 2.0 allows this to be done on a file-by-file basis). To illustrate, if there is 20K of memory available after the program is loaded, a data file of about 5000 single-precison numbers or 2500 double-precision numbers could be processed. The actual data that can be used will vary with the specific program (i.e, does it need the entire file in memory, how big is the program itself, the precison used for the data file and the "overhead" it might need for intermediate results).

Disk Space. Microstat 2.0 consists of approximately 190K of programs (not including sample data files). One single-density 8" disk drive can, therefore, contain the entire Microstat package. Most 5 1/4" double density drives will have the programs divided between two disks. While Microstat can be run with just one 5 1/4" drive, it would involve considerable "disk switching" so we recommend a two-drive system.

Video Display. Microstat assumes at least a 16x64 display area for program output display. Some programs will automatically take advantage of larger display size (e.g., 24x80) if it is available. The program that is used to install Microstat (the PARM program discussed later) has been changed to make it possible to impliment the CLEAR SCREEN function for almost any video device using one or two control characters.

Printer. Although most of the Microstat programs provide for printer output, it is not required for program execution. If a printer is used, Microstat assumes it can printer 80 characters on a line.

SYSTEM POWER-UP ("COLD START")

All of the programs in Microstat are designed to use the CP/M operating system and Basic-80. Loading the CP/M operaing system will vary from system to system, with many of them automatically loading it once power is applied to the system. Do what you must to get CP/M into operation. Consult your computer system's manual if you are not sure.

You will know when the CP/M operating system has control by the presence of the following prompt on the video device (CRT):

A>

The "A" says that the currently "logged in" (i.e., active) disk drive is drive A and the greater-than sign means that CP/M is waiting for you to tell it to do something. This is a good time to make a working copy of the Microstat disk(s).

Making Copies. To make a copy of Microstat, place a disk that has been formatted for your system into drive B and the Microstat disk in drive A. If for some reason you haven't bought our (I)nterchange(TM) program, use PIP to transfer the programs to the blank disk. Consult your manual for instructions. Once a copy is made, place the original disk(s) in a safe place. Any updates or changes will only be made if the original disk(s) are returned.

Note: You are permitted to make copies of Micorstat for your personal use only. If you wish to license multiple (CPU) copies, we will provide additional disks and manuals for a nominal fee. We have great plans for future additions to Microstat that we think you'll appreciate. We simply ask that you help prevent unauthorized copies of Microstat from being made...to everyone's long-run benefit.

Since Microstat needs the Basic-80 interpreter to run its programs, the next task is to load Basic-80 into memory. We will assume that the disk in drive A contains both the CP/M operating system and Basic-80 and that the file name for Basic-80 is MBASIC.COM. To load Basic-80, enter the following:

A>**MBASIC /F:5**

and press the RETURN key (what you actually type in is in emphasized print). The F:5 part of the command tells Basic-80 to load the interpreter and reserve space in memory for up to 5 data files to be in use at one time. Failure to load Basic-80 in this manner will produce a bad-file-number error message.

Now that both the operating system and interpreter are in memory, remove the disk from drive A and place the working copy of Microstat into the drive. It's worth remembering at this point that computers aren't too clever without your help. Even though you've changed the disk in drive A, the computer doesn't know it. What must be done now is read into memory the file directory of the Microstat disk. To do this, enter:

RESET

and press the RETURN key. The disk drive will turn on, read in the new file directory and display "ok"; now you both know the disk was changed. Any time you change disks when Basic-80 is in control, you must use the RESET command to "log in" the file directory of the new disk. Failure to do so usually results in a BDOS error.

Parameter Initialization

The Microstat disk contains a short data file called PARMD that is used to store information that is unique to your particular system. It is also used to pass information between programs in Microstat. Its purpose is to increase speed of use by avoiding repetitive inputs.

The parameter file must be established prior to using Microstat for the first time. Generally, it should not be necessary to repeat this procedure unless you change the parameters at some time in the future (e.g., if you buy a new computer).

In the discussion that follows, we will assume that CP/M and Basic-80 (B-80 from now on) have already been loaded and that the Microstat disk is in drive A. (If not, follow the "cold start" instructions given earlier.) Now enter:

CHAIN "PARM"

and press the RETURN key.

CLEAR SCREEN-HOME CURSOR. Within a few seconds you will see a message stating that Microstat has been changed with respect to earlier versions to allow for more universal implementation of the CLEAR SCREEN function.

Most video display terminals (or CRT's) have a one or two character control codes that will clear the CRT screen (i.e, erase the contents on the CRT) and home the cursor (i.e., place the cursor in the upper left part of the screen). For example, the control code to do this on the SOROC terminal is an ESCAPE followed by an asterisk ("*"). Since these codes are sent to the CRT as ASCII characters, the number equivalent of ESCAPE in ASCII is 27 while the asterisk is 42.

The parameter program (PARM) will ask you to enter these numbers followed by the word END after the number(s) have been entered. The program displays the proper codes for the: SOL or VDM-1 video board, SOROC, ADM-3, ADDS-100, INTERTUBE, HAZELTINE, INFOTON, HEATH-ZENITH, TELEVIDEO and SWTP terminals. Simply enter the numbers for these CRT's as given in the program followed by the word END.

If you entered the proper codes, the program immediately uses to codes to clear the screen and home the cursor. If this does not happen, simultaneously press the CONTROL and C keys (i.e., a CONTROL-C) to abort the program and then type RUN to re-run it and try again using a different code(s). **Note:** if you have no idea what the codes are, try those listed for the various terminals. There's a good chance that one of them will work for your CRT. (Some terminals will even respond to two different code sequences; the TVI 920 responds to either a simple 26 or the 27 and 42 sequence of the SOROC.) Lastly, the program will only accept the first two numbers as valid control numbers.

If you've tried all of the codes and still can't find the proper sequence, place the phone near the computer with the CRT documentation, give us a call and we'll try to work it out over the phone. The CLEAR SCREEN function is simply to "beautify" the output of Microstat and has no effect on its computations.

Numeric Precision. The program then informs you that Microstat no longer requires you to specify all data files to be either single or double precision. The numeric precision of each data file is now declared when the file is created by the Data Management Subsystem (DMS). This is a much more flexible approach and should result in better use of both disk and memory space. The question of "system precision" for data files is no longer meaningful to Microstat.

Number of Video Lines. Since most people have trouble reading at 9600 baud (i.e., 200 words/second), most of the programs in Microstat will pause after the screen becomes filled with data so that you can review it before it "scrolls off" the screen. Usually, pressing any key will continue with the next page of information. To do this, the program needs to know how many lines can be presented on the screen at one time. Enter the number of lines (usually either 16 or 24).

Line Width of CRT. Several programs might need to display more information than the width of the screen permits. By knowing how many characters will fit on each line, these programs can "partition" the data to make the output more readable. The large correlation matrix in the sample printouts section of this manual is an example of this partitioning. The normal answer to this question will usually be either 64 or 80.

Line Width of Printer. This question is similar to the above, but pertains to the printer. Microstat assumes an 80 character line width for the printer in most programs. Some programs will partition the data if the width is less than 80 characters per line.

Maximum Number of Variables. Most of the programs in Microstat allocate data space dynamically and your answer to this question determines how much space must be reserved for data in your system. As a general rule of thumb, estimate what you think will be the maximum number of variables you will need, add one or two to it and enter the number. If you get "out of memory" messages, you can either reduce this number or buy more memory.

Default Data Drive. Earlier versions of Microstat used a dedicated data drive concept where all data files resided on one drive. Microstat 2.0 changes this somewhat by allowing you to "spread" the data files over as many drives as you wish. To do this, when files are created in DMS you will be asked what drive will hold the new data file. If no drive is given, the default data drive is asssumed (usually drive A). You should enter the drive prefix you wish to be the default data drive (i.e., A,B, C,D,E,F, or G with drive A being the normal default). If you have a single drive system, the answer must be A.

The Delay. The program will now clear the screen and display a message to inform you that there will be a delay while the program looks around in memory for some specific information contained in the B-80 interpreter. Since B-80 does not offer a convenient means of changing between the printer and the CRT, earlier versions of Microstat had to duplicate every PRINT statement that might go to either the printer or the CRT.

If the program finds what it needs, your new version of Microstat will be about 20K shorter than before (including the new programs and features that have been added). If it can't find the necessary information, you will be so informed. Microstat will still function even if the search fails. However, output cannot be sent to the printer until we send you a new disk with the duplicated PRINT statements. We have yet to find a version of B-80 for which the program failed to find what it needed. **Note:** the search may take a while (possibly two or three minutes), but eventually it will terminate the search.

The program will then activate the disk drive(s), depending upon your answer to the default-drive question and write the PARMD data file. It will then automatically chain you to the main program menu (called MAIN.BAS).

The main program menu presents the major Microstat program options available to the user. By entering the letter associated with the desired option, Microstat will then load and execute the program option selected. Some of the program options themselves have "sub-options" available as well (these are discussed later in the manual). The purpose of the main program menu, therefore, is to direct the user to the general field of analysis (e.g., ANOVA). After the program loads the ANOVA program, the user then is presented the sub-options that allow a specific test within the general field to be performed (e.g., one-way ANOVA, two-ANOVA or random blocks ANOVA).

The general program options as presented in the MAIN program menu are as follows:

```
---------------------------MICROSTAT------------------------------

A. DATA MANAGEMENT SUBSYST          H. REGRESSION ANALYSIS
B. DESCRIPTIVE STATISTICS           I. TIME SERIES ANALYSIS
C. FREQUENCY DISTRIBUTIONS          J. NONPARAMETRIC STATISTICS
D. HYPOTYHESIS TEST: MEAN           K. CROSSTAB / CHI-SQUARE TESTS
E. ANALYSIS OF VARIANCE             L. PERMUTATIONS / COMBINATIONS
F. SCATTERPLOT                      M. PROBABILITIY DISTRIBUTIONS
G. CORRELATION MATRIX               N. HYPOTHESIS TEST: PROPORTIONS
                                    O. [TERMINATE]

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

Each of the major program options presented above are discussed later in the manual. No doubt you could run each of the options above without reading the rest of this manual. If you don't read the rest of the manual, however, you will miss some of the features that give Microstat much of its power and make it so easy to use. We strongly urge you to read the manual, especially the Defaults and Conventions section that follows.

Convetions, Defaults and Procedures

In order to provide consistency and promote ease of use, the following conventions are used in Microstat.

1. **Options.** In many programs, the user is presented a list of options, the selection of which determines the nature of what is to be done next. By selecting the desired option, the user controls the operation of the program. (This is what is meant by the term "interactive"; it's a give-and-take process between you and the computer.) Wherever possible, program options are indicated by a single letter. To select an option, simply enter the letter for the option wanted. It is generally <u>not</u> necessary to press the RETURN key after entering the letter.

If you press the RETURN key <u>without</u> entering a letter, the **default option** is automatically selected. The default option is always the first option in the list unless otherwise indicated in the option list. (The sub-options list in the DMS shows an example of this exception.) If you enter a letter that is not in the list of options, you will be asked to re-enter your selection.

2. **YES-NO Defaults.** Whenever there is a question that requires a YES or NO response, the question will be followed by either (Y,N) or (N,Y). If you enter either a Y or N, the program proceeds accordingly without the need to press the RETURN key. However, if you press RETURN without entering a letter, the default option is automatically selected. The default option is always the first letter in the (Y,N) or (N,Y) field. For example, if the question is: PRINTER OUTPUT (N,Y), simply pressing RETURN is the same as entering an N because N is the first option in the field. The defaults were chosen to be the most common response or, in some cases, the least consequential if an error is made. If you type in any character other than those listed, the default is also selected.

3.**Other Data Entry.** While single letter program options may be selected without pressing the RETURN key, all other forms of data entry <u>will</u> require pressing the RETURN key. For example, when you enter numbers that will become part of a data file, you will have to press RETURN after each number is entered. This is standard procedure and will not be mentioned in subsequent discussion.

4. **End of Program Options (EOP).** In most of the programs, output is first sent to the CRT for display. After the output has been displayed, the user is given the option of having the output repeated on the printer or CRT. This can be a real timesaver, since you don't have to make notes to yourself about the data that gave you the desired results and then re-run the program to get a printout. Since most programs end with this option, we will refer to it as EOP in subsequent discussion.

5. **Panic Button.** A program can be stopped by simultaneously pressing the CONTROL and C key (i.e., a CONTROL-C). This will abort the execution of the program and return control to the B-80 interpreter. **NOTE:** we strongly advise against doing this, especially during disk activity. It should only be used if you must terminate the program (e.g., blue smoke coming out of the computer).

6.**Non-coma Output.** In many programs the computer will go into a loop that may take some time to complete (e.g., reading a long data file). Rather than have the computer look like it's in a coma, the program will display a number on the screen that shows the progress of the computer within the loop. This is done so you know the computer is not "locked up" (and to discourage you from pressing the "panic button").

7. **Error Indication.** The bell will sound on the terminal if there is an input error, provided the terminal supports this feature. Usually this will be caused by entering improper data.

8. **Instruction Conventions.** In subsequent discussion, capital letters will refer to messages, output or prompts displayed on the CRT as the programs are executed. Underlining indicates that the user is expected to enter a response for use in the program.

9. **Program Sequence and Procedures.** Wherever possible, program sequence and procedure has been duplicated for purposes of consistency. In order to avoid repetition in subsequent discussion, these procedures are discussed here in full and are only mentioned by heading later. These are:

a. **Input/Output File Selection.** At or near the beginning of many programs, a message such as the following is given:

```
OPEN FILE: B:TEST  (PRESS 'RETURN' TO USE OPEN FILE)
ENTER FILE NAME:______
```

The program is now asking you to enter the name of the file with which you wish to work. (Remember that underlining indicates a user response is being requested.) If you press RETURN with no entry, the program assumes you wish to continue working with the currently open file (which happens to be called TEST and is located on drive B).

If you wish to change files, you should enter the new file name. **Note:** Microstat 2.0 now checks for a drive prefix. That is, if you simply enter a file name, such as TUBE, the program will look for the file on the default data drive (as determined when you ran PARM and held in the PARMD data file). If you enter B:TUBE, the program overrides the default data drive prefix and looks on drive B for the TUBE data file. If the program cannot find the data file, you will be asked to re-enter the file name.

b. **Select Cases.** The option list is:

```
OPTIONS: A. ALL CASES
         B. SUBSET OF CASES
```

Simply pressing RETURN selects option A (this is a default, remember?) and will use all of the cases in the data file. If option B is entered, you will be asked for the beginning and ending case numbers (inclusively), ignoring all others. For example:

ENTER BEGINNING CASE NO.: 8 ENDING CASE NO.: 14

would use cases 8 through 14 for analysis in the program.

c. **Printer Output.** Usually, the following is displayed on the CRT:

PRINTER OUTPUT (N,Y):_____

You may respond with a Y or N, while a RETURN with no entry selects the default option which is NO in this case. When printer is selected, all output is sent to the print device (which is assumed to be turned on).

d. **Job Title.** This simply allows the user to place an identifying header on the printed output. This can be useful to "jog" one's memory on a printout made several months ago that might otherwise be unclear. It would appear on the screen as:

ENTER JOB TITLE:__________

The job title can be up to 50 characters, or press RETURN if no title is wanted.

e. **Variable Selection.** In some programs, you may wish to use the variables in a sequence other than the one in which they are stored in the data file. The variables can be entered by selecting the numbers that refer to each variable in the desired sequence. (Press RETURN after each entry.) **However**, if you will be using the variables in their present numeric sequence, simply pressing the RETURN key with no entry selects the current variable and increments the index by one. This allows contiguous blocks to be selected by just pressing RETURN. For example, suppose you wanted to select the following sequence of variables:

1,2,3,12,7,8,9,10

You could do this with th following keystrokes:

RETURN,RETURN,RETURN,12RETURN,7RETURN,RETURN,RETURN,RETURN

As you press RETURN each time, the program displays the index number of the variable selected, just as if you had entered it. The commas, of course, are not entered and are added just for clarity. While it may appear that there are a few too many returns, don't forget that RETURN must be pressed to enter the 12 and the 7.

After the variables have been entered, you will be asked:

SELECTION OK (Y,N):___

If an N is entered, the sequence must be re-entered. A Y or a RETURN assumes the selection was correct.

f. **Group Selection.** Several programs require that subsets of data be selected. The program will ask for the beginning and ending case for each group, inclusively. It functions in similar fashion to the SUBSET of CASES option in b. above. To illustrate, suppose you want two groupings of a variable; the first group cases 1-5 and the second 6-10 and there are 10 cases for the variable. This could be done by entering:

RETURN,5RETURN,RETURN,RETURN

In this case, the last return will subsititute N (the total number of cases, or 10 in the illustration) for the last RETURN.

g.**End of Program Options.** The EOP was mentioned earlier, but this fills in some of the details about how most programs end. After the program has been run, you will usually see something like the following (there are some deviations between programs):

```
OPTIONS: A. REPEAT OUTPUT ON PRINTER
         B. REPEAT OUTPUT ON SCREEN
         C. MORE COMPUTATIONS FROM CURRENT FILE
         D. CHANGE FILES
         E. [TERMINATE]
```

These should be fairly clear as to what happens. Options C and D may cause the file to be re-read in some programs (done for reasons of memory limitations). In programs that have multiple functions (e.g., FPCP does factorials, permutations and combinations), the E option returns you to the sub-menu in the program (i.e., do you want to do more factorials, permutations or combinations) rather than back to MAIN. Entering E at this point will then return you to MAIN.

While these defaults and conventions may seem like a lot to remember right now, they will become second nature to you in a very short while and significantly improve the ease and speed of use of Microstat.

DATA MANAGEMENT SUBSYSTEM

In order to handle moderately large data sets, Microstat was designed to be a file-oriented package. The heart of Microstat is the Data Management Subsystem (DMS) that creates all of the data files for subsequent use in the programs. For those of you that have a special statistical routine that is not presently part of Microstat, DMS is stored on the disk in source code as DATOP.BAS. This should be helpful in interfacing your routine into the general Microstat framework.

Each data file created by DMS actually produces two data files. The first is called the **header file** and contains information about the number of variables and cases, variable names, a file lable (if wanted), the precision of the data file and the name of the second data file. The second file is a standard random access file that contains the actual numeric data.

For example, suppose you want to create a data file called TEST and place it on drive B. The program will create the header file with the name TEST and also a second (random access) file called TESTR. The creation of TESTR is automatically done by DMS. The name of this file as stored in the header file will be B:TESTR, since we wish to place it on drive B. (If no drive prefix was specified, the default data drive is used as held in the PARMD data file.) All of the programs in Micorstat read the header file first and display its contents before computations begin.

When you select the DMS from the main program menu (i.e., MAIN), the following sub-system menu for the DMS is dispalyed:

```
-------------------DATA MANAGEMENT SUBSYSTEM--------------------

A. ENTER DATA                       G. DELETE CASES
B. LIST DATA [DEFAULT]              H. VERTICAL AUGMENT
C. EDIT DATA                        I. SORT
D. EDIT FILE HEADER                 J. RANK-ORDER
E. DESTROY FILES                    K. LAG TRANSFORMATIONS
F. MOVE / MERGE / TRANSFORM         L. FILE TRANSFER
                                    M. [TERMINATE]

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

The program will then be waiting for you to enter the letter corresponding to the desired option. Each of the options is discussed below.

A. ENTER DATA

At the beginning of this option, you will be given two sub-options; A. START NEW DATA FILES,and B. ADD DATA TO EXISTING FILE. Each of these is now discussed.

A. START NEW FILE. The program will request the following information to be entered:

1. WHICH DATA DRIVE. This is a new feature of Microstat and allows you to use data drives other than the default data drive. The question also will display the drive prefix for the default data drive. If you simply press RETURN, the default data drive as held in PARM (and displayed with the question) is selected. If you wish to use a different data drive, **enter the letter of the data drive only.** You should <u>not</u> enter the colon (i.e., enter a B but not B:).

2. NEW FILE NAME. The name of the data file can now be up to 7 characters in length, rather than 5 as was the case in the earlier versions. The program checks the default data drive (only) to make sure that the new file name is not the same as that of an existing file. Obviously, you should not use PARMD as a dat file, nor the program names. If the file already exists, you will be asked to re-enter the new file name.

3. FILE LABLE. A file lable can be up to 24 characters in length (as indicated by the "dots" following the question) and is used to identify the contents of the file. If you do not wish to use the file lable, simply press RETURN.

4. NUMBER OF VARIABLES. The maximum number of variables your system will support (as held in PARMD) is displayed as part of the prompt. You should enter the number of variables that will be in the new file, subject to the system maximum.

5. VARIABLE NAMES. The program will now request the name for each of the variables; each of which is limited to a maximum of 5 characters. Any names that have less than 5 characters are right-justified in the header file.

6. NUMBER OF CASES. If known, the exact number of cases (or observations per variable) should be entered. If you do not know the exact number, enter your best estimate. You can add cases

later or quit data entry (the "E" option discussed below) if the number entered is incorrect. The file header is automatically adjusted for the actual number entered.

7. PRECISION OF FILE. If you enter an S, single-precision numbers are used while a D denotes double-precision. **NOTE:** double precision numbers use twice as much file space (8 versus 4 bytes) and do slow down disk I/O speed somewhat. Unless you are using very large numbers, there is little to be gained by using double-precision numbers.

After you have entered this question, the disk drive will activate and write the header file to the disk and issue a message stating that the file has been written.

8. DATA ENTRY. Each entry will be prompted with the case number, the variable number and variable name. You should then enter the number associated with the displayed case-variable combination. If you detect an error before pressing the RETURN key, you can correct it with the DEL or RUB key (or CONTROL-H). If you detect an error after pressing the RETURN key, make a note of the case number and correct it later with the EDIT option of DMS (discussed below). There is no reason to "start over". If you enter an alpha character, you will be asked to re-enter the number.

If you press RETURN with no entry, the current case number is entered as the number. This can be very useful in some instances (e.g., time series) where you want the case number to represent time periods. Also, you may want to include the case number as a variable so the data can be restored to its orginal sequence after a sort has been performed.

The program will continue to prompt for data until all of the data have been entered. The data drive will activate after each case is entered (you may appreciate this when some kicks the plug out after 499 cases have been entered). Unless you are a real speed typist, disk I/O will not slow you down at all.

If you wish to end the data entry before all of the cases have been entered (or if you over-estimated the number of cases), you can invoke the **E option** by entering either E or END when the program prompts for the next case. For example, if you specified 100 cases, but only need 90, enter an E for the first variable number for the 91st case. The program will adjust the file header to reflect the smaller number.

The program will then display a message that data entry is complete and return the user to the DMS menu.

B. ADD DATA TO EXISTING FILE. This permits you to add data to an existing data file. The program will ask how many additional cases are to be added and will proceed with data entry in the same fashion as if it were a new data file. The file header will also be adjusted to reflect the increased number of cases. After the new data have been added to the file, the user is returned to the DMS menu.

B. LIST DATA

This is the second major option in the DMS menu and is the default option for the entire DMS menu. That is, a RETURN for the option selection for DMS activates this section of the program.

1. File Selection. The first part of LIST DATA asks for the name of the file to be listed. **NOTE:** Microstat 2.0 now recognizes a drive prefix to override the default data drive. For example, if the default data drive is drive A, but you wish to work with a file on drive B called TEST, you would enter the file name as B:TEST. If no drive prefix is given, the program looks on the default data drive for the file.

2. The program then presents the: printer output option, the all-subset of cases option and an all-subset of variables option.

a. All-Subset of Variables. This option has two possible actions. If you select All variables, the program will print all of the variables in the file using two decimal places for all of the numbers (i.e., 10F2 format). If the raw data will not fit in a field of ten, the program does an "overflow" printing of the number. This tends to jumble the normally neat formatted output.

NOTE: if there are more variables than will fit on the screen at one time (e.g., six variables for an 80 column CRT), the program will display the first six variables for all cases. It will then display the next six (or less) variables starting with the first case through the last case. This repeats until all cases and variables have been displayed. This avoids the "messed up" display that would normally result if Microstat simply dumped the data on the CRT.

b. Listing a Subset of Variables. If you select this option, the following is displayed on the CRT:

A. 10F0	XXXXXXXXX.	
B. 10F2 (Default)	XXXXXXX.XX	
C. 10F4	XXXXX.XXXX	
D. 10F6	XXX.XXXXXX	
E. 10E1	X.XE+DD	(DD = Exponent)
F. 10E3	X.XXXE+DD	(DD = Exponent)
G. INTEGER	XXXXXXXXXX	
H. FLOATING POINT	free format	

The program will first ask how many variables you wish to display (the prompt will state how many variables are in the open file). Enter the number you wish to display (it may be one or all of them). If you press RETURN with no entry, it signals that you don't want to list any variables and the LIST option ends and returns you to the DMS menu. If you do enter a number, the program will then request the format for that many variables.

Example: let's assume you selected three variables out of ten for display and these are variable numbers 4,5 and 6 to be displayed in the G (integer), B (10F2) and B formats respectively. To do this, enter a number 4 for the first variable to be displayed. The program will then ask for its format; enter a G. The program will then prompt for the second variable; just press RETURN since the program will automatically increment the variable number by 1 if RETURN is entered witout a number. Now enter a B for the format. When the program requests for the third variable, press RETURN again, which will cause variable 6 to be selected. The program is now asking for its format; press RETURN which causes the program to select the previsouly entered format (i.e., the B format).

We strongly urge you to experiment with this feature. You'd be surprised at the clarity that can result in a report that's not cluttered up with a bunch of unnecessary zeroes.

The variables may be listed in any sequence with any format. After you have entered your variable-format response(s), the program asks if the selection is OK. If not, you can start over by entering an N while a simple RETURN will cause the formatted listing to proceed. After the listing is complete, you will be returned to the DMS menu.

. EDIT

The EDIT option lets you to change the entries of an individual case. You are first asked the case number to be changed whereupon the program displays the current contents of that case. The program then asks for the values for each variable to be entered. If the value is wrong, enter the new value. If the value is correct, press RETURN with no value entered which retains the current value for the variable.

After all variables have been edited, you will be asked if there are additional cases to be edited. If so, the procedure starts over. If not, you ae returned to the DMS menu.

D. EDIT FILE HEADER

This option lets you change either the file lable or the variable names. It will first ask if you wish to change the file lable. If so, enter the new file lable. If not, it then asks if the variable names are to be changed. If so, each variable is prompted with the current variable name and asks for the new one to be entered. Each name is limited to 5 characters. If no change is wanted, press RETURN with no entry. This is repeated for all variables. The program will then write the new file header to the disk and return you to the DMS menu.

E. DESTROY FILES

This option allows you to remove data files that are no longer needed (thus freeing up file space). The program first displays all of the data files on the default data drive. It then asks how many files are to be destroyed; enter the appropriate number. **Note:** these will usually be multiples of 2, since each file has a header and random file associated with it. The program then asks if you're sure you want these files destroyed. Any response other than Y returns you to the DMS menu. A Y will destroy the requested files.

F. MOVE / MERGE /TRANSFORM

MOVE, MERGE and TRANSFORM (or MMT) gives you a great deal of power and flexibility in terms of manipulating data. As the name implies, this option has three different functions. MOVE simply transfers data from one file to another. However, since subsets of a file can be used and the sequence of the variables changed, it is more than a simple file copy. MERGE is similar to MOVE, but allows the variables to be selected from one or two data files. The real power in MMT, however, is found in the TRANSFORM options.

a. All three options begin by asking if there will be one or two input files and then request the name(s) of the file(s). If two input files are selected, they cannot have the same name. Also, both files must have the same number of cases.

b. The program then asks if transformations will be performed on the data. If the answer is N, the program skips directly to the output phase described below.

c. Transfromation Selection. If you indicated that transforms will be performed on the data, a list of the transformation codes is displayed on the CRT. A list of the codes is presented below.

Transform Codes

(X1 and X2 are variables, A and B represent constants)

A. 1/X	Reciprocal
B. LOG(X1)	Common log (base 10)
C. LN(X1)	Natural log (base e)
D. EXP(X1)	Natural antilog (e^X1)
E. X1+X2	Add variables X1 and X2
F. X1-X2	Subtract variables X1 and X2
G. X1*X2	Multiply variables X1 and X2
H. X1/X2	Divide variables X1 and X2
I. X1^A	Raise variable X1 to the Ath power
J. A+B*X1	Linear Transformation
K.SUMX1,X2	Summation of X1 through X2

These transforms can save you a considerable amount of pencil-pushing with a little thought. For example, if you wanted to duplicate a variable, transformation J with A=0 and B=1 will do it. A little thought about the transformation codes at the time a file is created should result in your being able to create a considerable variety of variables from one data set.

After the transformaion codes are displayed, the variable names and numbers are displayed. If two file were selected for input, both sets of names are displayed with an index number. For example, if the first file has three variables and the second has four variables, the variable names would have index numbers 1 through 7. The first variable created by a transformation would be number 8.

Under the listing of the variables, you will see:

```
NO.   CODE   OPERATION   X1   X2   --A-- --B--   -NAME
```

Under the NO. column you would see the index number of the first transformation (8 in the example). The cursor would now be positioned under the CODE column waiting for you to tell the program what transform will be performed. If you wanted to do a linear transform, you would enter a J at this point (no RETURN is needed). The cursor will then move to the X1 column and wait for a variable X1's index number to be entered. The cursor will then move to the A column and wait for you to enter the A constant. If you were duplicating a variable, you would enter a 0 and press RETURN. The cursor would then move to the B column and wait for the coefficient to be entered. After it was entered, the cursor would move to the NAME column and wait for you to enter the name for the new variable.

The cursor will then move back to the NO. column and display the next index number (e.g., 9) and wait in the CODE column. As each code is entered, the program moves to those columns needed for that transform, skipping those it does not need. For example, codes A-H would not stop in the A or B columns. This procedure can be repeated up to the maximum number of variables you specified when PARM was run. If you wish to end the transformations before the maximum number is reached, enter L for the next transformation code.

The program will then ask if the transforms are correct. If you respond with N, the transforms start over. If you respond with Y, the program asks how many variables will be written to the file and the output phase begins.

Error Recovery. If you note that you entered the wrong transformation code, simply press RETURN with no entry when the cursor is in the X1 column. This will cause the line to be repeated. If you detect the error after the X1 field is passed, you have two options: 1) enter L for the next transform and say the transforms are incorrect and start over, or 2) you can leave the "bad" transform in and ignore it during the output phase.

d. Output Phase. The program will now ask how many variables you wish to output. If there were 7 original variables and 4 transformations, up to 11 variables could be written to the disk. If you only want 6 variables in the new file, enter a 6. The program then asks if you wish to output all or a subset of cases; respond accordingly. The program will then ask you for the name of the output file. It cannot be the name of an existing file.

The program then asks you to enter the index numbers of the variables to be written to the file. If you said there were 6 and their indexes were 2,3,8,9,10 and 11, you can enter the following:

2RETURN,RETURN,8RETURN,RETURN,RETURN,RETURN

which will select and display the desired indexes for the variables. You are then asked if the selection is OK. If an N is entered, the program asks for the indexes to be re-entered. Any other response will cause the data to be written to the new file specified. Upon completion, you are returned to the DMS menu.

DELETE CASES

The steps involved are:

a. Input File Specification (i.e., enter the file name to be used).

b. Enter the number of cases to be deleted.

c. Select the cases to be deleted. The procedure is the same as described in the Variables Selection part of the Defaults section of the manual. The cases may be entered in any order.

The output file is the same as the input file. The remaining cases are "compacted" and the file header is adjusted accordingly. Upon completion, you are returned to the DMS menu.

H. VERTICAL AUGMENT

This program permits cases from one file to be appended to another file to create a new file. The first step is to select two input files. The two input files must have the same number of variables, of course. If desired, both input files can have the same name. You are then asked if a subset of cases will be used; respond accordingly.

The last step is to select the output file; it cannot be either of the two input file names. The variable names from either file can be used for the output file. A new file lable must also be entered. The program then writes the new file and returns you to the DMS menu.

I. SORT

This program re-arranges the cases in a file according to one or more variables selected as sort-keys. The sorting is done in ascending order. The first step is to enter the name of the file to be sorted. You must then enter the index of the variables to be used as sort keys. The program uses a modified Shell-Metzner sort.

Each variable may be thought of as a sort-key. The program asks for the variables that are to be used for the sort-keys. The sort is performed in the sequence in which the variables are entered.

Note 1. The sorted file replaces the original file. If you want to preserve the original file, use MMT or FILE TRANSFER to make a copy of the file before running the sort program. A scrath file is no longer used since the program pulls the entire file into memory before doing the sort.

J. RANK-ORDER

The program replaces the data in the selected variable with ranks. Tied ranks are averaged. The first step is to input the file name to be used. You are then asked to enter the number of variable(s) to be created into ranks and then their indexes. Their order of selection is not relevant.

The program will then rank the variables and write the new file to the disk. Note 1 of the sort program also applies to this program. You are then returned to the DMS menu.

L. LAG TRANSFORMATIONS

This program is useful for creating data files for auto-correlation and forecasting analysis. It is assumed that each case represents a sequential time period.

The operational procedure is the same as for MMT, except that there may be only one input file and you are asked to enter the number of lag periods. The number of lag periods is entered as k. The following operations are avaiable. (Note: X(t) = data for the t-th cases; X(t-k) = data from the k-th prior case)

A. LAG (The new variable contains X(t-k))
B. DIFFERENCE (X(t)-X(t-k))
C. % CHANGE (((X(t)-X(t-k))/X(t-k))*100)
D. RATIO CHANGE (X(t)/X(t-k))

The number of cases output is automatically adjusted for the number of lag periods in the header file. Upon completion, you are returned to the DMS menu.

L. FILE TRANSFER

The program amounts to a way of duplicating an existing data file and is useful when subsequent programs permanently alter the contents of the file (e.g., SORT) and you wish to preserve the contents of the original file. The first question asked is for the drive prefix for the input and output files and they may be on the same or different drives.

The program does allow disks to be switched in the drives if need be. Upon completion, you are asked to place the Microstat disk into drive A if it was removed as part of the transfer. The program then returns you to the DMS menu.

This completes the discussion of the DMS section. Most of the programs in DMS are in source code (i.e., DATOP.BAS can be listed directly) so you can see how the data files in Microstat are formed. This will enable you to write any specialty routines you may wish to use that are not part of Microstat.

We think you will find the DMS a powerful ally in your work and very easy to use after a little experience wth it. We are, of course, open to suggestions for additions to DMS (or any other part of Microstat). We intend to keep Microstat the best statistics package on the market and your feedback will help us do just that.

The rest of the manual discusses the various statistical analysis programs available in Microstat. The discussion is presented in outline form, drawing upon the information presented in the Defaults and Conventions section of the manual.

DESCRIPTIVE STATISTICS

This program calculates commonly-used measures of descriptive statistics. It utilizes a correction-vector input algorithm that minimizes the chances of overflow and inaccuracy due to large sums of squares. The program does not calculate the median. This can be done, however, by using the SORT option to find the middle case or the RANK-ORDER option to find the case with the middle rank.

Drawing upon the information contained in the Default and Conventions section of the manual for details, the operation of this (and subsequent) program is presented in outline form.

a. Input file selection

b. Select cases to be used

c. Printer output wanted?

d. Job title is requested if printer output was selected

e. Options:

```
ARITHMETIC MEAN (Y,N)?
SAMPLE STANDARD DEVIATION AND VARIANCE (Y,N)?
POPULATION STANDARD DEVIATION AND VARIANCE (Y,N)?
STANDARD ERROR (Y,N)?
MAXIMUM, MINIMUM (Y,N)?
SUM, SUM OF SQUARES, DEVIATION SS (Y,N)?
MOMENTS ABOUT MEAN, SKEWNESS, KURTOSIS (Y,N)?
```

The DEVIATION SS is the sum of the squared deviations around the mean (i.e., the numerator term of the variance). Don't forget that a simple RETURN selects the default option YES.

f. SELECTION OK (Y,N)?___ (If N entered, process starts over)

g.ENTER VARIABLE NO. TO BE OUTPUT (E=END,L=LIST VAR. NAMES):

If a valid variable index number is entered, the output of the selected statistics is presented. If L is entered, a list of the variable names for the file is displayed. Step g. is repeated until E is entered, which will cause the EOP options (i.e. End of Program) to be displayed. Note that the program calculates the statistics for all of the variables so the file is read only once. Once E is entered, you are returned to the main program menu (i.e., MAIN).

FREQUENCY DISTRIBUTIONS

This program will generate the frequency distributions and histograms for quantitative data set (grouped frequency distribution) or qualitative data coded as specific numeric values (count individual values). Note: since there was no way for us to anticipate the magnitude of values to be used in Microstat and that the histogram had to fit most common printers, a relative histogram is printed. If you think about it, a 1,000 item count on an 80-column printer is a bit difficult without relative printing.

The steps to be followed are:

a. Input file selection

b. OPTIONS: A. GROUPED FREQUENCY DISTRIBUTION
B. COUNT INDIVIDUAL VALUES
C. (TERMINATE)

c. Printer output?

d. ENTER VARIABLE NUMBER FOR FREQUENCY DISTRIBUTION:

e. Select cases (subsets permissible)

f. Enter the information for the distribution selected in option b. The input will vary according to the option selected. These are:

For Grouped Frequency Distribution:

ENTER INTERVAL WIDTH:____

ENTER LOWER LIMIT OF FIRST INTERVAL:____

For Count Individual Values:

HOW MANY INDIVIDUAL VALUES TO BE COUNTED:___ (50 max.)

ENTER VALUES TO BE COUNTED:

VALUE 1:___
VALUE 2:___
(etc.)

Prompting will continue until all values have been entered. Pressing RETURN with no entry enters the previous value incremented by one.

h. The data file will not be read and the output displayed. In addition to the actual count for each interval or value, the percent, cumulative frequency and cumulative percent will also be displayed. If the values are outside the specified limits, a message is displayed to inform you how many values were outside the specified limits.

i. EOP

HYPOTHESIS TESTS: MEANS

a. Input file selection

b. OPTIONS: A. MEAN VS. HYPOTHESIZED VALUE
B. DIFFERENCE BETWEEN MEANS: PAIRED OBSERVATIONS
C. DIFF. BETWEEN TWO GROUP MEANS: LARGE SAMPLE
D. DIFF. BETWEEN TWO GROUP MEANS: SMALL SAMPLE

c. According to the option selected in b, you will answer the following questions:

Mean vs. Hypothesized value: Paired Observations

1. Select cases

2. ENTERVARIABLENO.TO BETESTED(E=END,L=LISTVAR. NAMES):

When doing paired test, the variable selected must be the difference between the two variables tested. If the data are not in this form, use MMT to create the appropriate variable.

3. ENTER HYPOTHESIZED VALUE:___
or ENTER HYPOTHESIZED DIFFERENCE:___ (This is normally 0)

Difference Between Two Group Means:

1. Group selection

2. ENTER VARIABLE NO. TO BE TESTED (E=END,L=LIST VAR.NAMES):

d. Once you've reached this stage, the disk will activate, read the data and display the results.

e. EOP

ANALYSIS OF VARIANCE

a. Input file selection

b. OPTIONS: A. ONE-WAY ANOVA
B. RANDOMIZED BLOCKS ANOVA
C. TWO-WAY ANOVA

c. According to the option selected, you are asked:

One-Way ANOVA: Groups are defined as subsets of cases wth a variable.

1. ENTER NUMBER OF GROUPS:___

2. Group selection; they do not have to be of equal size.

3. ENTER VARIABLE NUMBER TO BE ANALYZED:___

4. Job Title

5. OPTIONS: A. OUTPUT TREATMENT MEANS ON SCREEN
B. OUTPUT TREATMENT MEANS ON PRINTER
C. SUPPRESS TREATMENT MEANS

Treatment means are only output as the data file is read (i.e., if you ask for the output to be repeated as EOP, the ANOVA table is repeated but not the treatment means).

Randomized Blocks ANOVA: Each variable selected will represent a treatment and each case will represent a block.

1. ENTER NUMBER OF BLOCKS:___

2. ENTER NUMBER OF TREATMENT GROUPS:___

3. Select variables to represent treatment groups as prompted. Pressing RETURN automatically increments the variable number by one.

4. Job title

5. Output treatment and block means (i.e., save as option 5 in the One-way ANOVA).

Two-Way ANOVA. Variables will represent the "column" treatments and subsets of cases will represent the "row" treatments. The number of replications in each cell must be equal.

1. ENTER NO. OF REPLICATIONS PER CELL:___

2. ENTER NO. OF ROWS:___

3. ENTER NO. OF COLUMNS:___

4. Select variables representing the column treatments.

5. Job title.

6. Output treatment and cell means? (same as option 5 above)

d. The data file will nowbe read and the ANOVA table requested will be displayed.

e. EOP

SCATTERPLOTS

a. Input file selection

b. Printer output. Note: if you select screen output, you cannot repeat the results as a normal EOP option. This is due to the possible differences between screen and printer size. The program can, of course, be run again.

c. ENTER VARIABLE NUMBER OF HORIZONTAL AXIS:___

d. ENTER VARIABLE NUMBER OF VERTICAL AXIS:___

e. Job title (printer only)

f. DO YOU WANT AUTOMATIC SCALING (Y,N):___

With automatic scaling, the minimum and maximum values of each variable is used as the endpoints for the axes. With manual scaling, the minimum and maximum values are displayed, but the user enters the endpoints to be used.

g. The file is read and the plot calculated. The algorithm has been changed in such a way that the program no longer needs to sort the data before the plot. The values are scaled to fit in a string that is 2 characters less than the screen width (i.e., 78 characters for most CRTs) by using an INT function. Because truncation does occur, the resulting plot is not a "true" X-Y plot but will be reasonably close. (If anyone has a better way of doing it, we're open to suggestions.)

The "*" indicates the presence of a single data point. If more than one data point occurs at the same coordinates in the plot, it will show the number of overlapping points up to 9. Beyond that value, a pound sign ("#") is displayed at that point in the plot.

h. EOP, subject to limitation mentioned above.

CORRELATION MATRIX

a. Input file selection.

b. Job title.

c. Printer output?

d. OPTIONS: A. OUTPUT CORRELATION MATRIX
B. OUTPUT SSCP AND VAR-COVAR
C. ALL OF THE ABOVE

Option B produces the raw sum of squares/cross products, adjusted sum of squares/cross products (i.e., deviation SSCP), variance/covariances as well as the correlation. This output will be in tabular rather than matrix form and cannot be repeated as a normal EOP option. The program can, of course, be run again.

e. The data file is now read and the requested output displayed. If the correlation matrix is too large to fit on the selected output device, it will be partitioned until it's been completely displayed (see sample in Appendix A). If 99.999 is displayed as a correlation, it means at least one of the variables was a constant.

f. EOP, subject to limitation mentioned above.

REGRESSION ANALYSIS

The regression program in Microstat 2.0 now supports simple, multiple and stepwise multiple regression. The latter feature is new and causes no degradation in speed or accuracy of the earlier version. Also, the data file no longer has to fit in memory to run. The program calculates the amount of memory needed and, if the file won't fit and the residuals are selected, a second pass is made on the data. Although the need for a second pass on large data files does slow things down somewhat, it is still relatively fast.

a. Input file selection

b. ENTER VARIABLE NUMBER OF DEPENDENT VARIABLE:___

c. Job title

d. The disk will now activate, read the data file and perform some preliminary calculations.

e. ENTER NUMBER OF PREDICTOR VARIABLES:___
(predictor = independent)

f. ENTER INDEXES OF PREDICTOR VARIABLES:___

The variables will be listed on the CRT with their means and standard deviations. The chosen dependent variable will be the last variable in the list and the remaining variables re-numbered 1 through N-1. The new numbers are the indexes referred to in the question. As usual, pressing RETURN will automatically enter the previous index incremented by one.

g. SELECTION OK (Y,N)?__ (If N, f. is repeated)

h. The program now asks if you wish to do a stepwise or full model regression. If the full model regression is selected, the program skips to section i below. Otherwise, you are asked:

F TO ENTER:___ F TO REMOVE:___

TOLLERANCE (DEFAULT = 0.001):___

MAXIMUM NUMBER OF STEPS (DEFAULT = X):___

If you press RETURN for the F values, the default value of 3 is used for each F so selected. The default for the maximum number of steps is equal to the number of predictor variables plus 2 and is printed in the prompt in place of the "X" above.

i. DO YOU WANT RESIDUALS CALCULATED (N,Y):___ If you respond with Y, you are also asked:

DO YOU WANT DURBIN-WATSON TEST (N,Y):___

j. Printer output?

k. The output that results is:

Means and standard deviations.

Regression output: regression coefficient, standard error, F value, partial r^2, constant term, standard error of estimate, R squared (or r squared, as needed), multiple R (or r), analysis of variance table, the residuals and Durbin-Watson (if selected). If the stepwise regression was selected, the variable entered-removed is displayed, plus the tolereance and F to enter. Due to the new information needed for stepwise, the t values are no longer printed, but may be determined by the ratio of the coefficient to its standard error.

1. OPTIONS: A. ANOTHER SET OF PREDICTOR VARIABLES
   B. CHANGE DEPENDENT VARIABLE
   C. REPEAT OUTPUT FROM PREVIOUS ANALYSIS
   D. TERMINATE

If option C is selected, you ae given the choice of sending output to the screen or printer. If stepwise was selected, only the final regression results can be repeated. If D is selected, you are returned to the main program menu.

We'd like to point out that the regression program is not a simple algorithm published elsewhere. A good test of test of the regression can be found in the article referenced in the file lable of the sample regression in Appendix A. (JASA is the Journal of the American Statistical Association). If you run that "pathological" data set in any other regression program, we'd like to know the results. You will find that our regression outperforms some regression algorithms used on mainframe machines.

TIME SERIES ANALYSIS

a. Input file selection

b. OPTIONS: A. MOVING AVERAGE
   B. CENTER MOVING AVERAGE AND DE-SEASONALIZATION
   C. EXPONENTIAL SMOOTHING

Moving Average:

1.ENTERVARIABLE NO.TO BE AVERAGED(E=END,L=LISTVAR.NAMES):

2.Select cases

3.ENTER NUMBER OF PERIODS IN MOVING AVERAGE:___

4.Job title

5.Printer output?

6.Data file read and output displayed.

7. EOP

Centered Moving Average and De-Deasonalization:

1. ENTER VAR. NO. TO BE AVERAGED (E=END,L=LIST VAR. NAMES):

2. Select cases

3. OPTIONS: A. QUARTERLY DATA
            B. MONTHLY DATA

If a file contains data for a partial year, a message to that fact is issued and the program will not analyze the data for the partial year.

4. Job title

5. Printer output?

6. The file is read and the seasonal indexes calculated by averaging the ratios of the data to the centered moving average.

7. OPTIONS: A. USE INDEXES SHOWN
            B. OUTPUT INDEXES ON PRINTER
            C. INPUT OTHER INDEXES

If option C is selected, the program prompts for the new indexes. Options B and C return to step 7 when completed.

8. Output results.

9. EOP

Exponential Smoothing:

1. ENTER VAR. NO. TO BE SMOOTHED (E=END,L=LIST VAR. NAMES):

2. Select cases

3. Job title

4. ENTER SMOOTHING FACTOR:___

5. OPTIONS: A. USE FIRST DATA POINT AS INITIAL VALUE
            B. INPUT INITIAL VALUE

6. File is read and results displayed.

7. EOP

NONPARAMETRIC TESTS

At the beginning of the program, the following options are displayed: (It should be noted that all of the nonparametric tests will not fit into a single program with much memory left for data. For this reason, some of the test result in another program being loaded into meory.)

OPTIONS: A. WALD-WOLFOWITZ RUNS TEST
B. WILCOXON RANK-SUM TEST FOR TWO GROUPS
C. KRUSKAL-WALLIS ONE-WAY ANOVA BY RANKS
D. KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST
E. KOLMOGOROC-SMIRNOV TWO GROUP TEST
F. WILCOXON SIGNED-RANK TEST
G. ABSOLUTE NORMAL SCORES TEST
H. FRIEDMAN TEST
I. KENDALL COEFFICIENT OF CONCORDANCE
J. SIGN TEST
K. (TERMINATE)

Note: several of the options (described below) assume that certain variables have been converted to ranks prior to their execution. As the data are read, the ranks are summed and then compared to the formula for the sum of N integers. If the sums do not agree, the program displays a CHECCKSUM ERROR message. The analysis can be continued, but must be interpreted with caution.

Wald-Wolfowitz Runs Test

This test calculates the number of runs above and below a specified "split" value.

1. Input file selection

2. ENTER VARIABLE TO BE ANALYZED:___

3. ENTER SPLIT VALUE:___

If the data are already coded as two dichotomous values, specify any value between the coded values as the split vale.

4. Job title

5. The file is read and the output displayed. It will show the number of cases above and below the split value, the number of runs above and below the split value and the z approximation.

6. EOP

Wilcoxon Rank-Sum Test for Two Groups
Kruskal-Wallis One-Way ANOVA by Ranks

Since both of these tests are operationally the same except that the latter inputs more than two data groups, they are discussed together.

1. Input file selection

2. ENTER VARIABLE NO. TO BE ANALYZED:___

The variable to be tested must have been converted to ranks with the RANK-ORDER option in DMS prior to analysis.

3. Enter group selection

4. Job title

5. The data is read and output displayed

6. EOP

Kolmogorov-Smirnov Two Group Test

This test compares two frequency distributions from two independent groups. Data entry is from the keyboard since most frequency distributions have relatively small numbers of cases.

1. ENTER NUMBER OF CLASSES:___

2. ENTER OBSERVED FREQUENCIES FROM TWO DISTRIBUTIONS

CLASS	GROUP 1	GROUP 2
1	_______	_______

Prompting will continue until both distributions have been entered. Note that the input consists of actual frequencies, not cumulative frequencies.

3. Job title

4. Output includes the observed frequencies, cumulative frequencies for each group, D MAX (the class associated with D MAX will be identified with an "*") and the critical value of D MAX at the .05 and .01 significance levels.

5. EOP

Wilcoxon Signed Rank Test
Absolute Normal Scores Test

The operational procedures are the same for both tests, so they are discussed together. The input data file must contain one variable that is the actual difference between two variables being compared and another variable that contains the ranks of the absolute difference. Code F of MMT may be used to do the first, while the second may be created by squaring the difference variable (G or I of MMT) and then using the RANK-ORDER option of DMS to rank the squared differences. (Squaring serves the same purpose as absolute value.)

1. Input file selection

2. ENTER VARIABLE NUMBER CONTAINING DIFFERENCES:___
   ENTER VARIABLE NUMBER CONTAINING RANKS:___

3. Job title

4. The file is read and the results displayed. The Absolute Normal Scores Test involves calculating z values via the inverse normal distribution function. These z values are flashed on the CRT as the data are read.

5. EOP

Friedman Test
Kendall Coefficient of Concordance

Because of their similarity, they are discussed together.

1. Input file selection

2. Job title

3. OPTIONS: A. VARIABLES = ITEMS, CASES = JUDGES
   B. VARIABLES = JUDGES, CASES = ITEMS

The data must be in the form of rankings of items by judges. Each variable can represent an item and the cases for that variable ae the rankings (option A), or each case can contain the rankings by judges represented by the variables (option B).

4. File is read and output displayed.

Output for Friedman test: rank sums for each item, chi-square value, degrees of freedom (d.f) and a Multiple Comparison Value are given. The Multiple Comparison Value indicates the difference between rank sums that would be necessary for the items to be different at the .05 significance level.

Output for Kendall Test: The coefficient of concordance (W), average rank-order correlation, chi-square and d.f. are all displayed.

5. EOP

Sign Test

The best way to do a sign test, given the inherent power of Microstat, is to generate the entire binomial distribution with the probability distribution program and find the exact number of occurences corresponding to the required level of significance (i.e., let n = the number of cases and p = .50)

With a large sample, you could also use the PROPORTIONS VS> HYPOTHESIZED VALUE test of the HYPOTHESIS TESTS FOR PROPORTIONS program (i.e., use .5 as the hypothesized value) as found in the main program menu.

This ends the section on nonparametric statistics.

CROSSTAB / CHI-SQUARE STATISTICS

The following options are given at the beginning of the program.

```
OPTIONS: A. CROSSTAB
         B. CONTINGENCY TABLE (DATA FILE INPUT)
         C. CONTINGENCY TABLE (KEYBOARD INPUT)
         D. GOODNESS OF FIT TEST
```

CROSSTAB

This test generates a two-way contingency table by counting selected values from two variables.

1. Input file selection

2. ENTER VARIABLE TO REPRESENT ROWS:___
   ENTER VARIABLE TO REPRESENT COLS:___

3. ENTER NUMBER OF ROWS IN TABLE:___
   ENTER NUMBER OF COLS IN TABLE:___

4. The program will now prompt for the specific values to be counted for rows and columns. Pressing RETURN will increment the previous value by one.

5. Select cases

6. Job title

7. The data is read and the two variables crosstabulated to generate a contingency table. The chi-square will be calculated as described below for options B and C.

Contingency Table (data file input)
Contingency Table (keyboard input)

Both options calculate the chi-square statistic for a previously-generated contingency table. Option B assumes that the table has been stored in a data file, while C permits direct keyboard entry. The maximum size for a table is 20 rows by 5 columns.

The output is the same for both options and includes the chi-square statistic and degrees of freedom. If d.f. = 1 (i.e., a 2x2 table), the chi-square with continuity correction factor is also displayed.

Prior to output there is an option to display observed frequencies, expected frequencies, observed percentages or expected percentages. All four combinations can be displayed as part of the EOP options.

Data File Input:

1. Input file selection (each variable will represent a column and each case a row).

2. Job title

3. Data read and output displayed

4. EOP

Keyboard Input:

1. ENTER NUMBER OF ROWS:___
   ENTER NUMBER OF COLS:___

2. Program prompts for input of cell frequencies.

3. Results displayed

4. EOP

GOODNESS OF FIT TEST

This program calculates chi-square given the input of observed and expected values. The input may be either frequencies or proportions and is from the keyboard.

1. OPTIONS: A. INPUT OBSERVED FREQUENCIES
   B. INPUT OBSERVED PROPORTIONS
   C. INPUT EXPECTED FREQUENCIES
   D. INPUT EXPECTED PROPORTIONS

2. ENTER NUMBER OF CLASSES:___

3. ENTER NUMBER OF PARAMETERS ESTIMATED FROM SAMPLE DATA:___

4. ENTER SAMPLE SIZE:___

5. Job title

6. The program now prompts for the observed and expected values. The input will be frequencies and/or proportions as selected in step 1.

7. ARE TOTALS CORRECT (Y,N):___

The question is asked at the end of data input and after the totals are displayed. For frequency input, the total equals sample size; for proportions it should be 1.00 (with rounding error). If the answer to the question is N, step 6 is repeated.

8. Output includes the observed and expected frequencies and proportions, chi-square and d.f.

9. WANT KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST (Y,N):

If you respond with Y, D MAX will be displayed and the class corresponding to D MAX indicated.

10. EOP

This concludes the section on crosstabs and chi-square.

FACTORIALS / PERMUTATIONS / COMBINATIONS

For each of these options, enter N and R as prompted and the results will be dsplayed. The EOP will allow the results to be printed on the printer.

The factorials involved in these programs (and in the PROBABILITY DISTRIBUTIONS program) are calculated three ways. Values up to 33 factorial are calculated directly (i.e., "brute force"); this is as far as you can go in some packages. Microstat continues up to one million factorial (1,000,000!). Factorials from 34 to 300 are calculated by an accumulation of logarithms ("Lincoln Logs" algorithm), while values in excess of 300 use Stirling's approximation with an extended precision multiplication routine.

The Permutations and Combinations options calculate factorials only if other, more direct, methods will not work.

PROBABILITY DISTRIBUTIONS

The following distributions are available:

```
OPTIONS: A. BINOMIAL
         B. HYPERGEOMETRIC
         C. POISSON
         D. EXPONENTIAL
         E. NORMAL
         F. F DISTRIBUTION
         G. STUDENT'S t
         H. CHI-SQUARE
         I. (TERMINATE)
```

Distributions A through C are discrete while D through H are continuous probability distributions.

BINOMIAL, HYPERGEOMETRIC, POISSON

1. Enter input parameters:

a. Binomial: N=number of trials, P=probability of occurrence
b. Hypergeometric: Population size, sample size, number of possible occurences.
c. Poisson: Mean rate of occurence

2. Enter X1 ad X2, the minimum and maximum values to be calculated. Pressing RETURN with no entry causes X1 to be set to 0 and X2 to be set to the sample size (X2 = 1000 for Poisson). In other words, if you want the entire distribution, press RETURN twice for X1 and X2. If a single probability is wanted, X1=X2.

3. Printer output?

4. The program then displays the probability and cumulative probability for each value of X1 and X2. Note that the cumulative probability is cumulative starting at X1.

The output of the probabilities is formatted to 5 places. The output will not be displayed until the cumulative probability is greater than .000005 and the output stops when it exceeds displayed.

5. EOP. Note that repeating does not require re-calculation of the initial probability.

EXPONENTIAL, NORMAL, F, STUDENT'S t, CHI-SQUARE

These distributions calculate the probability below (p) and above (1-p) specified input values.

EXPONENTIAL

1. ENTER MEAN RATE OF OCCURENCE:___

2. OPTIONS: A. CALCULATE PROBABILITY GIVEN X
   B. CALCULATE X GIVEN PROBABILITY

3. Enter either P or X are determined in step 2.

4. Results are displayed.

5. EOP

NORMAL

1. OPTIONS: A. CALCULATE PROBABILITY GIVEN Z
   B. CALCUALTE Z GIVEN PROBABILITY

2. Enter P or Z depending upon option from step 1.

3. Results displayed.

4. EOP

F DISTRIBUTION

1. ENTER D.F.   NUMERATOR:___
   ENTER D.F   DENOMINATOR:___
2. ENTER F:___
3. Results displayed.
4. EOP

STUDENT'S t

1. ENTER D.F.:___
2. ENTER   t:___
3. Results displayed.
4. EOP

CHI-SQUARE

1. ENTER D.F.:___
2. ENTER CHI-SQUARE:___
3. Results displayed.
4. EOP

HYPOTHESIS TESTS: PROPORTIONS

The program has the following options available:

OPTIONS: A. TWO PROPORTIONS FROM INDEPENDENT GROUPS
B. PROPORTIONS VS. HYPOTHESIZED VALUE

TWO PROPORTIONS FROM ONE GROUP:

C. (MUTUALLY EXCLUSIVE CATAGORIES)
D. (OVERLAPPING CATAGORIES)

After the option is selected, the following is also given:

OPTIONS: A. INPUT PROPORTIONS
B. INPUT NUMBER OF OCCURENCES

For example, if a proportion was 17/39, option B would allow an input of 17 whereas option A would require an input of .435897. The sample size (39 in the example) would be requested later and the program calculate the proportion. Whenever the program requires a proportion, it will prompt for P or X depending upon which option was selected.

The operational procedure for each is identical:

1. Enter input proportions and sample size as prompted.

2. Results displayed.

3. EOP

CONCLUDING COMMENTS

The remainder of the manual is devoted to sample printouts of program results from Mcirostat. Most of these contain file lables that reference common textbooks and journal articles that permit you to compare the results from Microstat to those produced on much larger machines. We think you'll be surprised at the results possible with Microstat.

The fact that you bought this manual and have read it this far suggests that you need a statistics package. While we hope that you choose Microstat, we urge you to compare Microstat with any other package for tests included, ease of use and, most importantly, numeric accuracy. We think you'll find that we're tough to beat...at any price.

Lastly, if there is some test or procedure that you think we should include, please let us know about it. After all, it was your comments that produced this most current release. We have plans for additional "library" routines in the future plus several other "surprises" in the offing. In short, we're here to serve your needs as best we can and genuinely value want to know what those needs are. Keep in touch!

APPENDIX A: SAMPLE OUTPUT

DATA MANAGEMENT SUBSYSTEM

What follows in this section are examples of: ENTER DATA, LIST DATA, EDIT DATA, DELETE CASES and VERTICAL AUGMENT. You may wish to replicate the sequence below. For the sake of brevity, the printing of the file header has been left out.

The operations represent:

1. Input of DAT1 and DAT2 using option A of ENTER DATA.
2. Use of EDIT DATA option to change variable Y, case 1 of DAT2 from 10500 to 10783.
3. Both files appended using VERTICAL AUGMENT.
4. Cases 1 and 7 removed using DELETE CASES.
5. Another case added using option B of ENTER DATA.

DAT1

	--X--	--Y--
1	14	7666
2	21	8385
3	17	8866
4	18	9219
5	24	9978

DAT2

	--X--	--Y--
1	19	10500
2	24	11000
3	19	11148

	--X--	--Y--
1	19	10783
2	24	11000
3	19	11148

	--X--	--Y--
1	14	7666
2	21	8385
3	17	8866
4	18	9219
5	24	9978
6	19	10783
7	24	11000
8	19	11148

	--X--	--Y--
1	21	8385
2	17	8866
3	18	9219
4	24	9978
5	19	10783
6	19	11148

	--X--	--Y--
1	21	8385
2	17	8866
3	18	9219
4	24	9978
5	19	10783
6	19	11148
7	25	305

HEADER DATA FOR: B:2WAY LABEL: SCHAUM: P. 232
NUMBER OF CASES: 12 NUMBER OF VARIABLES: 3

	A1	A2	A3
1	70	83	81
2	79	89	86
3	72	78	79
4	77	77	74
5	81	87	69
6	79	88	77
7	82	94	72
8	78	83	79
9	80	79	75
10	85	84	68
11	90	90	71
12	87	88	69

HEADER DATA FOR: B:PDIFF LABEL: B & P : P.441
NUMBER OF CASES: 20 NUMBER OF VARIABLES: 4

	1977	1976	DIFF	RANK
1	5.58	4.14	1.44	18.00
2	6.62	6.38	0.24	7.00
3	1.86	1.69	0.17	5.50
4	6.05	5.73	0.32	8.00
5	6.44	6.33	0.11	3.00
6	8.22	7.66	0.56	15.00
7	14.16	8.36	5.80	20.00
8	7.29	3.67	3.62	19.00
9	3.86	4.19	-0.33	9.00
10	4.87	4.50	0.37	10.00
11	1.36	2.44	-1.08	17.00
12	4.27	3.24	1.03	16.00
13	3.03	2.52	0.51	13.00
14	2.64	2.67	-0.03	1.00
15	2.38	2.90	-0.52	14.00
16	1.93	1.80	0.13	4.00
17	1.53	1.46	0.07	2.00
18	4.61	4.16	0.45	11.00
19	2.31	2.48	-0.17	5.50
20	1.25	0.76	0.49	12.00

HEADER DATA FOR: B:TTEST LABEL: LAPIN P.607
NUMBER OF CASES: 10 NUMBER OF VARIABLES: 2

	DATA	RANK
1	35.60	1.00
2	38.70	2.00
3	42.30	3.00
4	42.80	4.00
5	47.20	7.00
6	45.30	5.00
7	46.40	6.00
8	50.10	8.00
9	53.10	9.00
10	61.40	10.00

HEADER DATA FOR: B:KENDL LABEL: SIEGEL: P.230
NUMBER OF CASES: 6 NUMBER OF VARIABLES: 3

	--X--	--Y--	--Z--
1	1	1	6
2	6	5	3
3	3	6	2
4	2	4	5
5	5	2	4
6	4	3	1

HEADER DATA FOR: B:1WAY LABEL: SCHAUM: P.225
NUMBER OF CASES: 12 NUMBER OF VARIABLES: 2

	WPM	RANK
1	79	8.50
2	83	11.00
3	62	3.00
4	51	1.00
5	77	7.00
6	74	6.00
7	85	12.00
8	72	5.00
9	81	10.00
10	65	4.00
11	79	8.50
12	55	2.00

The following output illustrates various transformations performed on file HOMEWORK (See Appendix C for listing of test data files.) The variable labeled e^LN represent the inverse of the previous natural log transformation and thus returns the variable to its original value. The variable 7++11 is the sum of variables 7 through 11. The other transformations are self-explanatory from the variable names.

The data were listed by specifying subsets of variables during the LIST DATA option and thus different formats were selected for each variable.

HEADER DATA FOR: TRATEST.2 LABEL: TRANSFORMATIONS TEST
NUMBER OF CASES: 22 NUMBER OF VARIABLES: 13 SIZE: 7 BLOCKS

	--X--	--Y--	1/Y	LOG Y	LN Y	e^LN
1	2	19	.0526	1.2788	2.9444	19.00
2	4	50	.0200	1.6990	3.9120	50.00
3	4	42	.0238	1.6232	3.7377	42.00
4	6	79	.0127	1.8976	4.3694	79.00
5	7	81	.0123	1.9085	4.3944	81.00
6	7	99	.0101	1.9956	4.5951	99.00
7	8	130	.0077	2.1139	4.8675	130.00
8	10	149	.0067	2.1732	5.0039	149.00
9	11	170	.0059	2.2304	5.1358	170.00
10	11	132	.0076	2.1206	4.8828	132.00
11	13	160	.0063	2.2041	5.0752	160.00
12	14	160	.0063	2.2041	5.0752	160.00
13	16	149	.0067	2.1732	5.0039	149.00
14	18	140	.0071	2.1461	4.9416	140.00
15	19	140	.0071	2.1461	4.9416	140.00
16	20	110	.0091	2.0414	4.7005	110.00
17	20	120	.0083	2.0792	4.7875	120.00
18	22	135	.0074	2.1303	4.9053	135.00
19	23	91	.0110	1.9590	4.5109	91.00
20	24	101	.0099	2.0043	4.6151	101.00
21	25	102	.0098	2.0086	4.6250	102.00
22	25	80	.0125	1.9031	4.3820	80.00

HEADER DATA FOR: TRATEST.2 LABEL: TRANSFORMATIONS TEST
NUMBER OF CASES: 22 NUMBER OF VARIABLES: 13 SIZE: 7 BLOCKS

	X + Y	X - Y	X * Y	X / Y	Y^.5	7++11	8+2*Y
1	21	-17	38	.1053	4.3589	46.4642	46
2	54	-46	200	.0800	7.0711	215.1511	108
3	46	-38	168	.0952	6.4807	182.5760	92
4	85	-73	474	.0759	8.8882	494.9642	166
5	88	-74	567	.0864	9.0000	590.0864	170
6	106	-92	693	.0707	9.9499	717.0206	206
7	138	-122	1040	.0615	11.4018	1067.4633	268
8	159	-139	1490	.0671	12.2066	1522.2737	306
9	181	-159	1870	.0647	13.0384	1905.1031	348
10	143	-121	1452	.0833	11.4891	1485.5724	272
11	173	-147	2080	.0813	12.6491	2118.7304	328
12	174	-146	2240	.0875	12.6491	2280.7366	328
13	165	-133	2384	.1074	12.2066	2428.3140	306
14	158	-122	2520	.1286	11.8322	2567.9608	288
15	159	-121	2660	.1357	11.8322	2709.9679	288
16	130	-90	2200	.1818	10.4881	2250.6699	228
17	140	-100	2400	.1667	10.9545	2451.1212	248
18	157	-113	2970	.1630	11.6190	3025.7820	278
19	114	-68	2093	.2527	9.5394	2148.7921	190
20	125	-77	2424	.2376	10.0499	2482.2875	210
21	127	-77	2550	.2451	10.0995	2610.3446	212
22	105	-55	2000	.3125	8.9443	2059.2568	168

SORT

File TEST1 was sorted first on variable 2 (GPA) and then on variable 7 (ACCT). Thus the GPA measure is sorted within the ACCT classification.

HEADER DATA FOR: SORTTEST,2   LABEL: ACCT=MAJOR, GPA=MINOR
NUMBER OF CASES: 52   NUMBER OF VARIABLES: 7   SIZE: 8 BLOCKS

	MOTIV	GPA	SAT-V	SAT-M	SEX	ATTND	ACCT.
1	7	2.20	450	520	1	2	0
2	7	2.32	500	600	1	4	0
3	5	2.35	400	620	0	3	0
4	2	2.40	590	540	0	1	0
5	8	2.45	392	558	0	4	0
6	5	2.45	460	520	0	2	0
7	7	2.47	420	550	0	3	0
8	8	2.50	450	500	0	2	0
9	5	2.50	480	500	0	4	0
10	6	2.53	480	520	1	2	0
11	7	2.60	540	450	0	2	0
12	2	2.66	450	540	0	4	0
13	3	2.70	532	433	1	2	0
14	7	2.80	400	400	0	1	0
15	5	2.80	643	732	0	3	0
16	6	2.88	336	413	1	3	0
17	7	2.90	650	550	0	3	0
18	7	3.00	380	680	0	3	0
19	2	3.00	590	610	1	3	0
20	7	3.18	500	540	1	3	0
21	10	3.20	520	540	0	2	0
22	5	3.29	510	420	1	2	0
23	3	3.29	520	470	1	2	0
24	8	3.40	600	620	1	1	0
25	6	3.46	690	628	1	3	0
26	5	3.52	610	600	0	2	0
27	7	3.67	600	750	0	2	0
28	7	3.75	558	705	0	2	0
29	9	3.75	600	690	0	2	0
30	6	3.80	590	710	0	2	0
31	8	3.94	530	760	0	3	0
32	9	3.95	670	760	1	0	0
33	4	3.96	740	760	1	2	0
34	4	2.40	520	540	1	2	1
35	8	2.60	530	620	0	2	1
36	2	2.70	550	630	0	3	1
37	8	2.75	500	510	1	1	1
38	4	2.88	540	560	1	2	1
39	10	3.15	480	530	0	2	1
40	8	3.23	720	700	0	3	1
41	2	3.25	480	660	1	2	1
42	3	3.25	490	620	0	2	1
43	5	3.36	480	570	1	2	1
44	1	3.40	450	650	0	1	1
45	1	3.50	550	600	0	2	1
46	4	3.50	630	570	0	3	1
47	5	3.57	650	720	0	3	1
48	6	3.65	580	630	1	1	1
49	5	3.66	710	610	1	3	1
50	8	3.70	780	740	0	1	1
51	4	3.85	580	580	1	1	1
52	8	3.90	580	600	1	1	1

RANK ORDER

The MOVE/MERGE/TRANSFORM option was used to output the variable GPA twice, once with the name GPA and once with the name RANK. The RANK ORDER program was then used to convert the second variable to ranks.

HEADER DATA FOR: RANKTEST,2 LABEL: TEST OF RANK-ORDER PROG.
NUMBER OF CASES: 52 NUMBER OF VARIABLES: 2 SIZE: 3 BLOCKS

	GPA	RANK
1	2.88	20.50
2	3.00	23.50
3	2.45	6.50
4	2.80	18.50
5	2.35	3.00
6	2.47	8.00
7	2.20	1.00
8	3.40	34.50
9	2.66	14.00
10	2.50	9.50
11	2.45	6.50
12	3.25	29.50
13	3.15	25.00
14	2.53	11.00
15	2.50	9.50
16	3.36	33.00
17	3.25	29.50
18	2.32	2.00
19	2.75	17.00
20	3.18	26.00
21	3.29	31.50
22	3.20	27.00
23	2.40	4.50
24	3.29	31.50
25	3.94	50.00
26	2.60	12.50
27	2.70	15.50
28	2.60	12.50
29	2.88	20.50
30	3.50	37.50
31	2.70	15.50
32	3.75	45.50
33	3.65	41.00
34	3.90	49.00
35	3.85	48.00
36	2.40	4.50
37	3.80	47.00
38	3.00	23.50
39	3.67	43.00
40	3.75	45.50
41	3.40	34.50
42	3.52	39.00
43	3.50	37.50
44	2.80	18.50
45	3.57	40.00
46	2.90	22.00
47	3.95	51.00
48	3.46	36.00
49	3.66	42.00
50	3.23	28.00
51	3.96	52.00
52	3.70	44.00

LAG TRANSFORMATIONS

Variable --Y-- of the HOMEWORK file was transformed with each of the options of the LAG TRANSFORMATIONS option. Number of lag periods=1. Note that the file size is reduced by 1 because of the lag period.

HEADER DATA FOR: LAGTEST,2 LABEL: VAR. Y LAGGED 1 PERIOD
NUMBER OF CASES: 21 NUMBER OF VARIABLES: 6 SIZE: 3 BLOCKS

	--X--	--Y--	LAG	DIFF.	%CHNG	RATIO
1	4	50	19	31	163.1579	2.6316
2	4	42	50	-8	-16.0000	.8400
3	6	79	42	37	88.0952	1.8810
4	7	81	79	2	2.5317	1.0253
5	7	99	81	18	22.2222	1.2222
6	8	130	99	31	31.3131	1.3131
7	10	149	130	19	14.6154	1.1462
8	11	170	149	21	14.0940	1.1409
9	11	132	170	-38	-22.3529	.7765
10	13	160	132	28	21.2121	1.2121
11	14	160	160	0	.0000	1.0000
12	16	149	160	-11	-6.8750	.9313
13	18	140	149	-9	-6.0403	.9396
14	19	140	140	0	.0000	1.0000
15	20	110	140	-30	-21.4286	.7857
16	20	120	110	10	9.0909	1.0909
17	22	135	120	15	12.5000	1.1250
18	23	91	135	-44	-32.5926	.6741
19	24	101	91	10	10.9890	1.1099
20	25	102	101	1	.9901	1.0099
21	25	80	102	-22	-21.5686	.7843

```
------------------ FREQUENCY DISTRIBUTIONS --------------------

HEADER DATA FOR: TEST1,2   LABEL: STATISTICS CLASS DATA
NUMBER OF CASES: 52   NUMBER OF VARIABLES: 7   SIZE: 8 BLOCKS

     VARIABLE: 3. SAT-V

    TEST OF GROUPED FREQUENCY DISTRIBUTION  (SAT VERBAL)

                                                ....CUMULATIVE...
 =====CLASS LIMITS====  FREQUENCY    PERCENT    FREQUENCY PERCENT
    300.00 <    400.00          3       5.77          3      5.77
    400.00 <    500.00         14      26.92         17     32.69
    500.00 <    600.00         21      40.38         38     73.08
    600.00 <    700.00         10      19.23         48     92.31
    700.00 <    800.00          4       7.69         52    100.00
                      TOTAL    52     100.00

 =====CLASS LIMITS====  FREQUENCY ............................
    300.00 <    400.00          3 =====
    400.00 <    500.00         14 ====================
    500.00 <    600.00         21 ==============================
    600.00 <    700.00         10 ==============
    700.00 <    800.00          4 ======


------------------ FREQUENCY DISTRIBUTIONS --------------------

HEADER DATA FOR: TEST1,2   LABEL: STATISTICS CLASS DATA
NUMBER OF CASES: 52   NUMBER OF VARIABLES: 7   SIZE: 8 BLOCKS

     VARIABLE: 6. ATTND

TEST OF NOMINAL FREQUENCY DISTRIBUTION (ATTENDANCE SELF-REPORT)

                                                ....CUMULATIVE...
  ===== VALUE =====     FREQUENCY    PERCENT    FREQUENCY PERCENT
          1.00                  9      17.65          9     17.65
          2.00                 23      45.10         32     62.75
          3.00                 15      29.41         47     92.16
          4.00                  4       7.84         51    100.00
                      TOTAL    51     100.00

 1 CASES WERE OUTSIDE SPECIFIED CLASS LIMITS

  ===== VALUE =====     FREQUENCY ............................
          1.00                  9 ============
          2.00                 23 ==============================
          3.00                 15 ===================
          4.00                  4 ======
```

```
------------------ HYPOTHESIS TESTS FOR MEANS ------------------

HEADER DATA FOR: PAIRDIFF,2    LABEL: B & L, P. 441
NUMBER OF CASES: 20    NUMBER OF VARIABLES: 4    SIZE: 2 BLOCKS
          DIFFERENCE BETWEEN MEANS: PAIRED OBSERVATIONS

             TEST OF PAIRED OBSERVATIONS T-TEST

        HYPOTHESIZED DIFF. =       .0000
                      MEAN =       .6590
                 STD. DEV. =      1.5243
                STD. ERROR =       .3408
                         N =      20    (CASES = 1 TO 20)

        T =    1.9335      (D.F. = 19)     VARIABLE:   DIFF
```

```
------------------ HYPOTHESIS TESTS FOR MEANS ------------------

HEADER DATA FOR: TTEST,2    LABEL: LAPIN, P. 607
NUMBER OF CASES: 10    NUMBER OF VARIABLES: 2    SIZE: 1 BLOCK
      DIFFERENCE BETWEEN TWO GROUP MEANS: SMALL SAMPLE

 SMALL SAMPLE T-TEST  (POOLED EXT. OF STD. ERROR OF DIFF.)

                     GROUP 1        GROUP 2
           MEAN =    41.3200        51.2600
      STD. DEV. =     4.3962         6.4555
              N =     5              5
          CASES =    1 TO 5         6 TO 10

                     DIFFERENCE =   -9.9400
      STD. ERROR OF DIFFERENCE =    3.4928

      T =   -2.8458      (D.F. = 8)     VARIABLE:   DATA
```

----------------- ANALYSIS OF VARIANCE ---------------------

ONE-WAY ANOVA

GROUP	MEAN	N
1	70.400	5
2	77.000	3
3	70.000	4
GRAND MEAN	71.917	12

----------------- ANALYSIS OF VARIANCE ---------------------

HEADER DATA FOR: ONE-WAY,2   LABEL: SCHAUM, P. 225
NUMBER OF CASES: 12   NUMBER OF VARIABLES: 2   SIZE: 1 BLOCK

ONE-WAY ANOVA

TEST OF ONE-WAY ANOVA WITH UNEQUAL GROUP SIZES

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RATIO
BETWEEN	103.717	2	51.858	.367
WITHIN	1273.200	9	141.467	.
TOTAL	1376.917	11		

----------------- ANALYSIS OF VARIANCE ---------------------

RANDOMIZED BLOCKS ANOVA

TREATMENT	MEAN	N
1	80.000	5
2	85.000	5
3	75.000	5

BLOCK	MEAN	N
1	86.000	3
2	84.667	3
3	80.667	3
4	74.333	3
5	74.333	3
GRAND MEAN	80.000	15

----------------- ANALYSIS OF VARIANCE ---------------------

HEADER DATA FOR: RBLOCK.2   LABEL: SCHAUM, P. 228
NUMBER OF CASES: 5   NUMBER OF VARIABLES: 3   SIZE: 1 BLOCK

RANDOMIZED BLOCKS ANOVA

TEST OF TWO-WAY ANOVA W/O INTERACTION (RANDOMIZED BLOCKS DESIGN)

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RATIO
TREATMENT	250.000	2	125.000	12.397
BLOCK	250.000	4	91.833	9.107
ERROR	80.667	8	10.083	
TOTAL	698.000	14		

---------------- ANALYSIS OF VARIANCE ---------------------

TWO-WAY ANOVA

COL	MEAN	N
1	80.000	12
2	85.000	12
3	75.000	12

ROW	MEAN	N
1	79.667	9
2	78.778	9
3	80.222	9
4	81.333	9

CELL MEANS

ROW	COL	MEAN	N
1	1	73.667	3
2	1	79.000	3
3	1	80.000	3
4	1	87.333	3
1	2	83.333	3
2	2	84.000	3
3	2	85.333	3
4	2	87.333	3
1	3	82.000	3
2	3	73.333	3
3	3	75.333	3
4	3	69.333	3

GRAND MEAN 80.000 36

----------------- ANALYSIS OF VARIANCE ---------------------

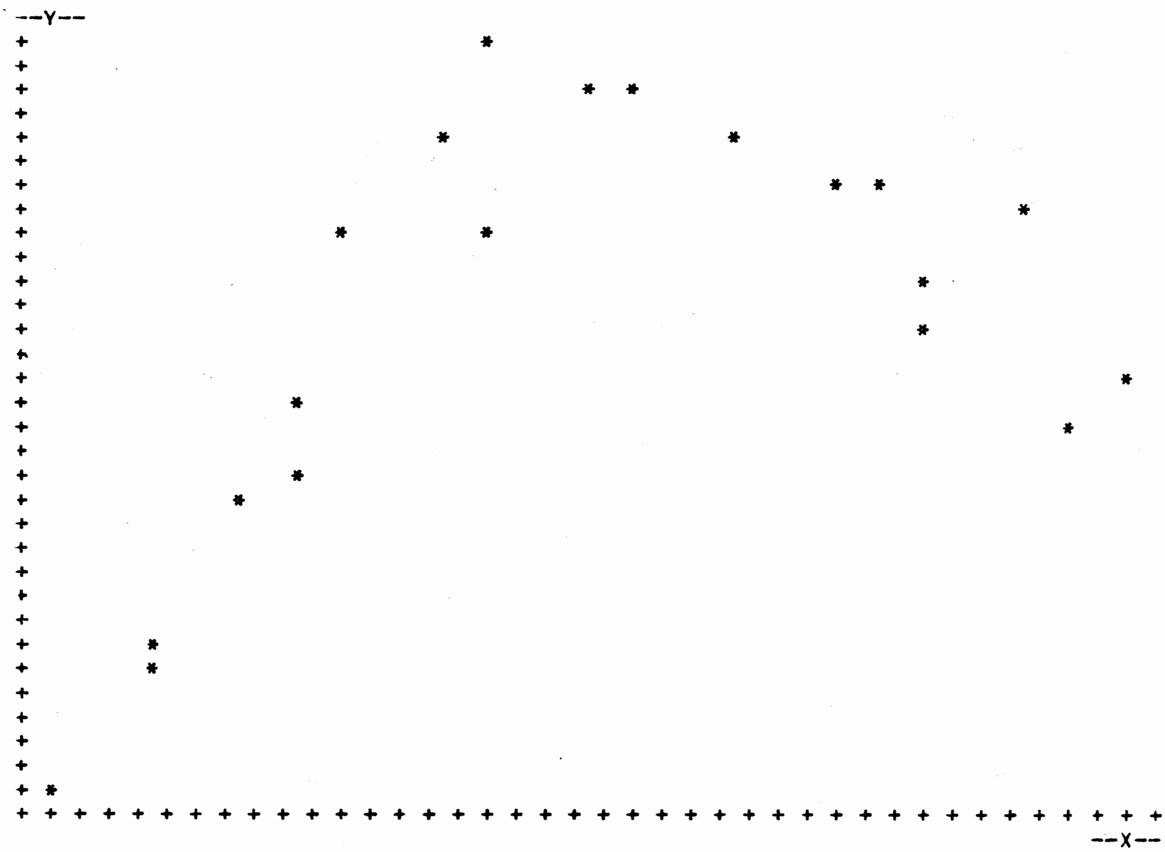
HEADER DATA FOR: TWO-WAY.2 LABEL: SCHAUM, P.232
NUMBER OF CASES: 12 NUMBER OF VARIABLES: 3 SIZE: 1 BLOCK

TWO-WAY ANOVA

TEST OF TWO-WAY ANOVA WITH INTERACTION

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RATIO
COLS	600.000	2	300.000	16.539
ROWS	30.889	3	10.296	.568
INTERACTION	533.778	6	88.963	4.905
ERROR	435.333	24	18.139	
TOTAL	1600.000	35		

SCATTERPLOT



TEST DATA FOR QUADRATIC REGRESSION

HORIZONTAL AXIS: --X--
LEFT ENDPOINT: 2 RIGHT ENDPOINT: 25

VERTICAL AXIS: --Y--
LOWER ENDPOINT: 19 UPPER ENDPOINT: 170

HEADER DATA FOR: HOMEWORK.2 LABEL: QUADRATIC REGRESSION
NUMBER OF CASES: 22 NUMBER OF VARIABLES: 3 SIZE: 2 BLOCKS

```
------------------- CORRELATION MATRIX -------------------------
HEADER DATA FOR: LONGLEY,2    LABEL: JASA, V.62, P.819-841
NUMBER OF CASES: 16    NUMBER OF VARIABLES: 7    SIZE: 3 BLOCKS

----------------------------------------------------------------

LONGLEY DATA (UNPERTURBED)
```

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--Y--	--Y--	6.844598E+10	1.850070E+08	1.233380E+07	1.0000
--X1-	--Y--	1.068162E+08	5.519500E+05	3.679667E+04	.9709
--X2-	--Y--	4.103227E+11	5.149960E+09	3.433307E+08	.9836
--X3-	--Y--	3.361978E+09	2.473680E+07	1.649120E+06	.5025
--X4-	--Y--	2.740942E+09	1.676540E+07	1.117693E+06	.4573
--X5-	--Y--	1.230685E+11	3.519200E+08	2.346133E+07	.9604
--X6-	--Y--	2.042837E+09	2.436000E+05	1.624000E+04	.9713

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X1-	--X1-	1.671721E+05	1.746860E+03	1.164573E+02	1.0000
--X2-	--X1-	6.467007E+08	1.595410E+07	1.063607E+06	.9916
--X3-	--X1-	5.289080E+06	9.388000E+04	6.258667E+03	.6206
--X4-	--X1-	4.293174E+06	5.235380E+04	3.490253E+03	.4647
--X5-	--X1-	1.921397E+08	1.102550E+06	7.350333E+04	.9792
--X6-	--X1-	3.180540E+06	7.638000E+02	5.092000E+01	.9911

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X2-	--X2-	2.553152E+12	1.481905E+11	9.879367E+09	1.0000
--X3-	--X2-	2.065054E+10	8.418660E+08	5.612440E+07	.6043
--X4-	--X2-	1.663295E+10	4.632070E+08	3.088047E+07	.4464
--X5-	--X2-	7.386802E+11	1.027860E+10	6.852400E+08	.9911
--X6-	--X2-	1.213117E+10	7.064000E+06	4.709333E+05	.9952

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X3-	--X3-	1.762543E+08	1.309836E+07	8.732240E+05	1.0000
--X4-	--X3-	1.314528E+08	-1.730690E+06	-1.153793E+05	-.1774
--X5-	--X3-	6.066486E+09	6.694130E+07	4.462753E+06	.6865
--X6-	--X3-	9.990586E+07	4.459500E+04	2.973000E+03	.6682

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X4-	--X4-	1.159817E+08	7.264560E+06	4.843040E+05	1.0000
--X5-	--X4-	4.923864E+09	2.646150E+07	1.764100E+06	.3644
--X6-	--X4-	8.153707E+07	2.073700E+04	1.382467E+03	.4173

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X5-	--X5-	2.213402E+11	7.258200E+08	4.838800E+07	1.0000
--X6-	--X5-	3.672577E+09	4.938000E+05	3.292000E+04	.9940

ROW	COL.	RAW SSCP	ADJUSTED SSCP	VAR-COVAR.	CORR
--X6-	--X6-	6.112146E+07	3.400000E+02	2.266667E+01	1.0000

```
------------------- CORRELATION MATRIX -------------------------
HEADER DATA FOR: LONGLEY,2    LABEL: JASA, V.62, P.819-841
NUMBER OF CASES: 16    NUMBER OF VARIABLES: 7    SIZE: 3 BLOCKS

----------------------------------------------------------------

LONGLEY DATA (UNPERTURBED)
```

	--Y--	--X1-	--X2-	--X3-	--X4-	--X5-	--X6-
--Y--	1.000						
--X1-	.971	1.000					
--X2-	.984	.992	1.000				
--X3-	.503	.621	.604	1.000			
--X4-	.457	.465	.446	-.177	1.000		
--X5-	.960	.979	.991	.687	.364	1.000	
--X6-	.971	.991	.995	.668	.417	.994	1.000

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

----- REGRESSION ANALYSIS -----

HEADER DATA FOR: B:LONG   LABEL: JASA V.62 PP.819-841
NUMBER OF CASES: 16   NUMBER OF VARIABLES: 7

TEST OF MULTIPLE REGRESSION B-80

INDEX	NAME	MEAN	STD.DEV.
1	--X1-	101.681	10.792
2	--X2-	387,698.438	99,395.000
3	--X3-	3,193.312	934.464
4	--X4-	2,606.688	695.920
5	--X5-	117,423.375	6,956.570
6	--X6-	1,954.500	4.761
DEP. VAR.:	--Y--	65,317.000	3,511.970

DEPENDENT VARIABLE: --Y--

VAR.	REGRESSION COEFFICIENT	STD. ERROR	T(DF= 9)	BETA
--X1-	14.9218	84.7603	0.1760	0.046
--X2-	-0.0357	0.0335	-1.0678	-1.011
--X3-	-2.0190	0.4882	-4.1360	-0.537
--X4-	-1.0331	0.2142	-4.8221	-0.205
--X5-	-0.0518	0.2258	-0.2295	-0.103
--X6-	1828.5194	455.4865	4.0144	2.479

CONSTANT: -3,480,963.153

STD. ERROR OF EST. = 304.828
R SQUARED = .995
MULTIPLE R = .998

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F RATIO
REGRESSION	184172546.743	6	30695424.4571	330.341
RESIDUAL	836283.450	9	92920.3834	
TOTAL	185008830.193	15		

```
                                                 STANDARDIZED RESIDUALS
      OBSERVED   CALCULATED    RESIDUAL -2.0                  0                  2.0
  1   60323.000   60055.652    267.348   |                  |        *          |
  2   61122.000   61215.767    -93.767   |               *  |                   |
  3   60171.000   60124.594     46.406   |                  |*                  |
  4   61187.000   61597.292   -410.292   |     *            |                   |
  5   63221.000   62911.645    309.355   |                  |         *         |
  6   63639.000   63888.362   -249.362   |          *       |                   |
  7   64989.000   65152.570   -163.570   |             *    |                   |
  8   63761.000   63773.958    -12.958   |                  *                   |
  9   66019.000   66005.265     13.735   |                  *                   |
 10   67857.000   67401.630    455.370   |                  |              *    |
 11   68169.000   68186.106    -17.106   |                 *|                   |
 12   66513.000   66552.086    -39.086   |                 *|                   |
 13   68655.000   68810.980   -155.980   |             *    |                   |
 14   69564.000   69649.697    -85.697   |               *  |                   |
 15   69331.000   68988.904    342.096   |                  |          *        |
 16   70551.000   70757.492   -206.492   |           *      |                   |
```

DURBIN-WATSON TEST = 2.5599

-------------------- TIME SERIES ANALYSIS --------------------

HEADER DATA FOR: LONGLEY,2   LABEL: JASA, V.62, P.819-841
NUMBER OF CASES: 16   NUMBER OF VARIABLES: 7   SIZE: 3 BLOCKS

LONGLEY DATA: UNEMPLOYMENT (X3), 3-TERM MOVING AVERAGE

	--X3-	3 TERM MOVING AVG.
1	2356.00	
2	2325.00	
3	3682.00	2787.67
4	3351.00	3119.33
5	2099.00	3044.00
6	1932.00	2460.67
7	1870.00	1967.00
8	3578.00	2460.00
9	2904.00	2784.00
10	2822.00	3101.33
11	2936.00	2887.33
12	4681.00	3479.67
13	3813.00	3810.00
14	3931.00	4141.67
15	4806.00	4183.33
16	4007.00	4248.00

-------------------- TIME SERIES ANALYSIS --------------------

HEADER DATA FOR: LONGLEY,2   LABEL: JASA, V.62, P.819-841
NUMBER OF CASES: 16   NUMBER OF VARIABLES: 7   SIZE: 3 BLOCKS

EXPONENTIAL SMOOTHING

LONGLEY DATA: UNEMPLOYMENT (X3), W=.5

SMOOTHING FACTOR= .5

	--X3-	SMOOTHED VALUE
1	2356.00	2356.00
2	2325.00	2340.50
3	3682.00	3011.25
4	3351.00	3181.13
5	2099.00	2640.06
6	1932.00	2286.03
7	1870.00	2078.02
8	3578.00	2828.01
9	2904.00	2866.00
10	2822.00	2844.00
11	2936.00	2890.00
12	4681.00	3785.50
13	3813.00	3799.25
14	3931.00	3865.13
15	4806.00	4335.56
16	4007.00	4171.28

```
-------------------- NON-PARAMETRIC TESTS --------------------

HEADER DATA FOR: WWTEST,2    LABEL: B. & L., P. 416
NUMBER OF CASES: 50    NUMBER OF VARIABLES: 1    SIZE: 2 BLOCKS

                 WALD-WOLFOWITZ RUNS TEST

  (DATA ARE CODED: 1=NON-DEFECTIVE ITEM, 2=DEFECTIVE ITEM)

         CASES BELOW = 44        CASES ABOVE = 6

          RUNS BELOW = 6          RUNS ABOVE = 6

                 TOTAL RUNS = 11

                 Z =   -.390
```

```
-------------------- NON-PARAMETRIC TESTS --------------------

HEADER DATA FOR: TTEST,2    LABEL: LAPIN, P. 607
NUMBER OF CASES: 10    NUMBER OF VARIABLES: 2    SIZE: 1 BLOCK

            WILCOXON RANK-SUM TEST FOR TWO GROUPS

   TEST OF WILCOXON RANK-SUM TEST FOR INDEPENDENT GROUPS

         SUM OF RANKS, GROUP 1 = 17       N1 = 5

         SUM OF RANKS, GROUP 2 = 38       N2 = 5

         Z1 = -2.193

         Z2 =  2.193
```

```
-------------------- NON-PARAMETRIC TESTS --------------------

HEADER DATA FOR: ONE-WAY,2   LABEL: SCHAUM, P. 225
NUMBER OF CASES: 12    NUMBER OF VARIABLES: 2    SIZE: 1 BLOCK

                   KRUSKAL-WALLIS TEST

       TEST OF KRUSKAL-WALLIS ONE-WAY ANOVA BY RANKS

              H =    .419           D.F. = 2
```

-------------------- NON-PARAMETRIC TESTS --------------------

HEADER DATA FOR: FRIEDMAN.2   LABEL: FRIEDMAN TEST DATA
NUMBER OF CASES: 11   NUMBER OF VARIABLES: 4   SIZE: 2 BLOCKS

FRIEDMAN TEST

DATA REPRESENTS 11 JUDGES' RANKINGS OF 4 PRODUCTS

ITEM	RANKSUM
1	30.0
2	24.0
3	38.0
4	18.0
TOTAL	110.0

CHI-SQUARE = 11.945   D.F. = 3

MULTIPLE COMPARISON VALUE (.05 LEVEL) = 15.56

-------------------- NON-PARAMETRIC TESTS --------------------

HEADER DATA FOR: KENDALL.2   LABEL: SIEGEL, P. 230
NUMBER OF CASES: 6   NUMBER OF VARIABLES: 3   SIZE: 1 BLOCK

KENDALL COEFFICIENT OF CONCORDANCE

TEST FOR CORRELATION OF 3 EXECUTIVES' RANKINGS OF 6 APPLICANTS

W = .162

AVERAGE RANK-ORDER CORRELATION = -.257

CHI-SQUARE = 2.429   D.F. = 5

-------------- CROSSTAB / CHI-SQUARE TESTS -----------------

DATA FROM SCHAUM, P. 211; KEYBOARD INPUT

OBSERVED FREQUENCIES

	1	2	3	TOTAL
1	120	20	20	160
2	50	30	60	140
3	50	10	40	100
TOTAL	220	60	120	400

CHI-SQUARE = 55.130   D.F.= 4

-------------- CROSSTAB / CHI-SQUARE TESTS -----------------

DATA FROM SCHAUM, P. 211; KEYBOARD INPUT

EXPECTED FREQUENCIES

	1	2	3	TOTAL
1	88.00	24.00	48.00	160.00
2	77.00	21.00	42.00	140.00
3	55.00	15.00	30.00	100.00
TOTAL	220.00	60.00	120.00	400.00

CHI-SQUARE = 55.130   D.F.= 4

-------------- CROSSTAB / CHI-SQUARE TESTS -----------------

DATA FROM SCHAUM, P. 211; KEYBOARD INPUT

OBSERVED PERCENTAGES

	1	2	3	TOTAL
1	30.00	5.00	5.00	40.00
2	12.50	7.50	15.00	35.00
3	12.50	2.50	10.00	25.00
TOTAL	55.00	15.00	30.00	100.00

CHI-SQUARE = 55.130   D.F.= 4

-------------- CROSSTAB / CHI-SQUARE TESTS -----------------

DATA FROM SCHAUM, P. 211; KEYBOARD INPUT

EXPECTED PERCENTAGES

	1	2	3	TOTAL
1	22.00	6.00	12.00	40.00
2	19.25	5.25	10.50	35.00
3	13.75	3.75	7.50	25.00
TOTAL	55.00	15.00	30.00	100.00

CHI-SQUARE = 55.130   D.F.= 4

-------------- CROSSTAB / CHI-SQUARE TESTS -----------------

GOODNESS OF FIT TEST

GOODNESS OF FIT TEST FOR POISSON DISTRIBUTION (B & L, P. 416)

	FREQUENCIES		PROPORTIONS	
CLASS	OBSERVED	EXPECTED	OBSERVED	EXPECTED
1	45	34.58	.4787	.3679
2	22	34.58	.2340	.3679
3	16	17.29	.1702	.1839
4	11	7.55	.1170	.0803
TOTALS	94	94.00	1.0000	1.0000

CHI-SQUARE= 9.390 D.F.= 2

KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST

CLASS CORRESPONDING TO LARGEST DIFFERENCE: 1

D MAX = .1108

FACTORIAL / PERMUTATIONS / COMBINATIONS

```
42! = 1.4050061E+51

289! = 2.0798695E+587

3600! = 2.5469452E+11241

THE NUMBER OF PERMUTATIONS OF 500 OBJECTS
    TAKEN 3 AT A TIME = 1.2425097E+8

THE NUMBER OF PERMUTATIONS OF 5423 OBJECTS
    TAKEN 51 AT A TIME = 2.2069601E+190

THE NUMBER OF COMBINATIONS OF 68 OBJECTS
    TAKEN 31 AT A TIME = 2.1912851E+19

THE NUMBER OF COMBINATIONS OF 4000 OBJECTS
    TAKEN 400 AT A TIME = 1.1208125E+563
```

BINOMIAL DISTRIBUTION
N = 94 P = .734

X	P(X)	CUMULATIVE PROBABILITY
49	.00001	.00001
50	.00001	.00002
51	.00003	.00005
52	.00007	.00012
53	.00015	.00027
54	.00032	.00059
55	.00064	.00123
56	.00123	.00247
57	.00227	.00473
58	.00399	.00872
59	.00671	.01543
60	.01081	.02624
61	.01662	.04287
62	.02442	.06728
63	.03422	.10150
64	.04574	.14724
65	.05825	.20550
66	.07063	.27613
67	.08145	.35758
68	.08924	.44681
69	.09279	.53960
70	.09144	.63104
71	.08529	.71634
72	.07518	.79152
73	.06252	.85404
74	.04896	.90300
75	.03603	.93903
76	.02485	.96388
77	.01603	.97992
78	.00964	.98956
79	.00539	.99495
80	.00279	.99773
81	.00133	.99906
82	.00058	.99965
83	.00023	.99988
84	.00008	.99996
85	.00003	.99999
86	.00001	1.00000
87	.00000	1.00000
88	.00000	1.00000

E(X) = 68.99600
STD. DEV. = 4.28403
VARIANCE = 18.35294

BINOMIAL DISTRIBUTION
N = 6 P = .5

X	P(X)	CUMULATIVE PROBABILITY
0	.01563	.01563
1	.09375	.10937
2	.23437	.34375
3	.31250	.65625
4	.23437	.89062
5	.09375	.98437
6	.01563	1.00000

E(X) = 3.00000
STD. DEV. = 1.22474
VARIANCE = 1.50000

HYPERGEOMETRIC DISTRIBUTION

THE POPULATION OF SIZE, 45600 OBJECTS
CONTAINS 2280 POSSIBLE OCCURENCES.

THE SAMPLE SIZE IS 60

X	P(X)	CUMULATIVE PROBABILITY
0	.04598	.04598
1	.14538	.19136
2	.22593	.41729
3	.23000	.64729
4	.17249	.81978
5	.10163	.92142
6	.04899	.97040
7	.01986	.99026
8	.00691	.99718
9	.00210	.99927
10	.00056	.99983
11	.00013	.99997
12	.00003	1.00000
13	.00001	1.00000
14	.00000	1.00000

E(X) = 3.00000
STD. DEV. = 1.68710
VARIANCE = 2.84631

POISSON DISTRIBUTION

MEAN RATE OF OCCURENCE = .6

X	P(X)	CUMULATIVE PROBABILITY
0	.54881	.54881
1	.32929	.87810
2	.09879	.97688
3	.01976	.99664
4	.00296	.99961
5	.00036	.99996
6	.00004	1.00000
7	.00000	1.00000

E(X) = .60000
STD. DEV. = .77460
VARIANCE = .60000

EXPONENTIAL DISTRIBUTION

MEAN RATE OF OCCURENCE = 1

X = 1.2
P = .6988, 1-P = .3012

EXPONENTIAL DISTRIBUTION

MEAN RATE OF OCCURENCE = 1

X = .69314717
P = .5000, 1-P = .5000

NORMAL DISTRIBUTION

Z = 1.96
P = .9750, 1-P = .0250

NORMAL DISTRIBUTION

Z = 1.6448535
P = .9500, 1-P = .0500

F DISTRIBUTION

D.F. NUMERATOR = 9
D.F. DENOMINATOR = 23

F = 3.3
P = .9900, 1-P = .0100

STUDENT'S T DISTRIBUTION

D.F. = 20

T = 1.7247
P = .9500, 1-P = .0500

CHI-SQUARE DISTRIBUTION

D.F. = 15

CHI-SQUARE = 24.996
P = .9500, 1-P = .0500

CHI-SQUARE DISTRIBUTION

D.F. = 500

CHI-SQUARE = 553.127
P = .9500, 1-P = .0500

HYPOTHESIS TESTS FOR PROPORTIONS

HYPOTHESIS TEST FOR TWO PROPORTIONS FROM INDEPENDENT GROUPS

P1 = .4500,   N1 = 200
P2 = .4000,   N2 = 250

Z = 1.067

HYPOTHESIS TEST FOR SAMPLE PROPORTION VS. HYPOTHESIZED VALUE

OBSERVED PROPORTION = .3750,   N = 160

HYPOTHESIZED PROPORTION = .3333

Z = 1.119

HYPOTHESIS TEST FOR TWO PROPORTIONS FROM ONE GROUP
(MUTUALLY EXCLUSIVE CATEGORIES)

P1 = .4000   P2 = .3000   SAMPLE SIZE = 400

Z = 2.408

HYPOTHESIS TEST FOR TWO PROPORTIONS FROM ONE GROUP
(OVERLAPPING CATEGORIES)

P1 = .2500   P2 = .2000   SAMPLE SIZE = 2000

OVERLAP PROPORTION = .0800

Z = 4.170

APPENDIX B: TEST DATA FILES

Microstat includes several test data files that can be used to reproduce the sample printouts in the manual. The data files include (only the header file is listed below):

LONG	HOME	TEST1	PDIFF	TTEST	WTEST
KENDL	1WAY	RBLCK	2WAY	FMAN	IEC32

We suggest that you use these sample data files for practice; you should easily reproduce the sample output for the files listed. Note that all of the sample data files are stored in single precision form. When you have finished with them, you can use the DESTROY FILES option in DMS to delete the files (also a form of practice!).

Some of the test data files (as referenced in the file lable) come from the following sources:

Beaton, Albert E., Rubin, Donald B. and Barone, John L.,"The Acceptability of Regression Solutions: Another Look at Computational Accuracy", _Journal of the American Statistical Association_, 71, (March, 1976), pp. 158-168. (This source contains the Longley data referenced below.)

Berenson, Mark L. and Levine, David M.,_Basic Business Statistics: Concepts and Applications_, Englewood Cliffs, NJ, Prentice-Hall,1979.

Kazmier, Leonard J.,_Theory and Problems of Business Statistics (Schaum's Outline Series)_, New York, McGraw-Hill, 1976.

Lapin, Lawrence, _Statistics for Modern Business Decisions_, 2nd Ed., New York, Harcourt-Brace-Janovick, 1978.

Longley, James W., "An Appraisal of Least Squares Programs for theElectronicComputer from the Point of ViewoftheUser", _Journalofthe American StatisticalAssociation_, 62, (Sept., 1967), pp. 819-841.

Siegel, Sidney,_Nonparametric Statistics for the Behavioral Sciences_, New York, McGraw-Hill, 1956.

APPENDIX C. COMPUTATIONAL EQUATIONS

DESCRIPTIVE STATISTICS

arithmetic mean: $\bar{X} = \frac{\Sigma X}{n}$

sample standard deviation and variance: $s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}}$ $\qquad s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$

population standard deviation and variance: $\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$ $\qquad \sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$

standard error of the mean: $s_{\bar{x}} = \frac{s}{\sqrt{n}}$

sum: ΣX

sum of squares: ΣX^2

deviation sum of squares: $\Sigma(X - \bar{X})^2$

moments about the mean:

$$m_2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2$$

$$m_3 = \frac{1}{n} \Sigma x_i^3 - \frac{3}{n} \bar{x} \Sigma x_i^2 + 2\bar{x}^3$$

$$m_4 = \frac{1}{n} \Sigma x_i^4 - \frac{4}{n} \bar{x} \Sigma x_i^3 + \frac{6}{n} \bar{x}^2 \Sigma x_i^2 - 3\bar{x}^4$$

moment coefficient of skewness: $\gamma_1 = \frac{m_3}{m_2^{3/2}}$

moment coefficient of kurtosis: $\gamma_2 = \frac{m_4}{m_2^2}$

HYPOTHESIS TESTS: MEAN

A. MEAN vs. HYPOTHESIZED VALUE

$$t_{n-1} = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

B. DIFFERENCE BETWEEN MEANS: PAIRED OBSERVATIONS

$$t_{n-1} = \frac{\overline{D} - 0}{\frac{S_D}{\sqrt{n}}}$$

C. DIFFERENCE BETWEEN TWO GROUP MEANS: LARGE SAMPLE

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

D. DIFFERENCE BETWEEN TWO GROUP MEANS: SMALL SAMPLE (pooled variances)

$$t_{n_1+n_2-2} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)\ S_2^2}{n_1 + n_2 - 2}$$

ANALYSIS OF VARIANCE

A. ONE-WAY ANOVA

Source of variation	Sum of squares, SS	Degrees of freedom, df	Mean square, MS	F ratio
Between treatment groups, A	$SSA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N}$	$K - 1$	$MSA = \frac{SSA}{K - 1}$	$F = \frac{MSA}{MSE}$
Sampling error, E	$SSE = SST - SSA$	$N - K$	$MSE = \frac{SSE}{N - K}$	
Total, T	$SST = \sum_{i=1}^{n} \sum_{k=1}^{K} X^2 - \frac{T^2}{N}$	$N - 1$		

B. RANDOMIZED BLOCKS ANOVA

Source of variation	Sum of squares, SS	Degrees of freedom, df	Mean square, MS	F ratio
Between treatment groups, A	$SSA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N}$	$K - 1$	$MSA = \frac{SSA}{K - 1}$	$F = \frac{MSA}{MSE}$
Between treatment groups, or blocks, B	$SSB = \frac{1}{K} \sum_{j=1}^{J} T_j^2 - \frac{T^2}{N}$	$J - 1$	$MSB = \frac{SSB}{J - 1}$	$F = \frac{MSB}{MSE}$
Sampling error, E	$SSE = SST - SSA - SSB$	$(J - 1)(K - 1)$	$MSE = \frac{SSE}{(J - 1)(K - 1)}$	
Total, T	$SST = \sum_{j=1}^{J} \sum_{k=1}^{K} X^2 - \frac{T^2}{N}$	$N - 1$		

C. TWO-WAY ANOVA

Source of variation	Sum of Squares, SS	Degrees of freedom, df	Mean square, MS	F ratio
Between treatment groups, A	$SSA = \sum_{k=1}^{K} \frac{T_k^2}{nJ} - \frac{T^2}{N}$	$K - 1$	$MSA = \frac{SSA}{K - 1}$	$F = \frac{MSA}{MSE}$
Between treatment groups, B	$SSB = \sum_{j=1}^{J} \frac{T_j^2}{nK} - \frac{T^2}{N}$	$J - 1$	$MSB = \frac{SSB}{J - 1}$	$F = \frac{MSB}{MSE}$
Interaction (between factors A and B), I	$SSI = \frac{1}{n} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\sum_{i=1}^{n} X \right)^2 - SSA - SSB - \frac{T^2}{N}$	$(J - 1)(K - 1)$	$MSI = \frac{SSI}{(J - 1)(K - 1)}$	$F = \frac{MSI}{MSE}$
Sampling error, E	$SSE = SST - SSA - SSB - SSI$	$JK(n - 1)$	$MSE = \frac{SSE}{JK(n - 1)}$	
Total, T	$SST = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K} X^2 - \frac{T^2}{N}$	$N - 1$		

TIME SERIES ANALYSIS

A. MOVING AVERAGE

The moving average is the mean of a 'moving window' of k cases of a specified variable. Larger values of k yield more smoothing.

B. CENTERED MOVING AVERAGE AND DE-SEASONALIZATION

First a 4 term (quarterly) or 12 term (monthly) moving average is performed, then a 2 term moving average centers the data. Seasonal indexes are calculated by averaging ratios of the data to the centered moving average. Deseasonalization is performed by dividing the original data by the seasonal index.

C. EXPONENTIAL SMOOTHING

$$\varepsilon_i = WY_i + (1 - W)\varepsilon_{i-1}$$

where ε_i = value of the exponentially smoothed series being computed in time period i

ε_{i-1} = value of the exponentially smoothed series already computed in time period $i - 1$

Y_i = observed value of the time series in period i

W = subjectively assigned weight or smoothing coefficient (where $0 < W < 1$)

CORRELATION MATRIX

raw SSCP = $\sum \sum (X_i X_j) \quad i,j = 1, \ldots, m$

adjusted SSCP = $\sum \sum \left[(X_i - \bar{X}_i)(X_j - \bar{X}_j) \right]$

variance-covariance = $S_{x_i x_j} = \frac{\text{adjusted SSCP}}{m-1}$

correlation = $\frac{S_{x_i x_j}}{\sqrt{S_{x_i x_i}} \sqrt{S_{x_j x_j}}}$

REGRESSION ANALYSIS

Y = BO + B1*X1 + B2*X2 . . . . Bk*Xk

Where, Y = estimated value for dependent variable
BO = intercept; constant
B1....Bk = regression coefficients
X1....Xk = values for independent variables

Refer to Neter, John; Wassserman, Wm., Applied Linear Statistical Models , 1974, Irwin., or other texts for details of interpretation.

CROSSTAB/CHI-SQUARE TESTS

CONTINGENCY TABLES

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\chi^2 = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$ (Yates correction for continuity when d.f. = 1)

where

f_o = observed frequencies

$f_e = \frac{\Sigma r \, \Sigma k}{n}$ = expected frequencies

d.f. = (rows -1)(col - 1)

GOODNESS OF FIT TEST

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

d.f. = number of categories - number of parameters estimated - 1

NONPARAMETRIC TESTS

A. WALD-WOLFOWITZ RUNS TEST

$$Z \cong \frac{U - \left(\frac{2n_1n_2}{n} + 1\right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}}}$$

where U = total number of runs

n_1 = number of successes in sample

n_2 = number of failures in sample

n = sample size; $n = n_1 + n_2$

B. WILCOXON RANK-SUM TEST FOR TWO GROUPS

$$Z \cong \frac{T_{n_1} - \frac{n_1(n+1)}{2}}{\sqrt{\frac{n_1n_2(n+1)}{12}}}$$

where T_{n_1} = summation of the ranks assigned to the n_1 observations in the first sample

C. KRUSKAL-WALLIS ONE-WAY ANOVA BY RANKS

$$H = \left[\frac{12}{n(n+1)} \sum_{i=1}^{c} \frac{T_{n_i}^2}{n_i}\right] - 3(n+1)$$

where n = total number of observations over the combined samples; $n = n_1 + n_2 + \cdots + n_c$

n_i = number of observations in the ith sample; $i = 1, 2, \ldots, c$

$T_{n_i}^2$ = square of the summation of the ranks assigned to the ith sample

D. KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST

$$D = \text{maximum } |F_0(X) - S_N(X)|$$

$F_0(X)$ = theoretical cumulative distribution

$S_N(X)$ = observed cumulative relative frequencies

E. KOLMOGOROV-SMIRNOV TWO GROUP TEST

$$D = \text{maximum } [S_{n_1}(X) - S_{n_2}(X)]$$

$S_{n_1}(X)$ = cumulative frequency distribution from sample 1

$S_{n_2}(X)$ = cumulative frequency distribution from sample 2

F. WILCOXON SIGNED-RANKS TEST

$$Z \cong \frac{W - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$$

where W = sum of the positive ranks; $W = \sum_{i=1}^{n} R_i^{(+)}$

μ_W = mean value of W; $\mu_W = \dfrac{n(n+1)}{4}$

σ_W = standard deviation of W; $\sigma_W = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$

n = number of nonzero absolute difference scores in sample

G. ABSOLUTE NORMAL SCORES TEST

$$Z \cong \frac{K - \dfrac{\sum_{i=1}^{n} z_i}{2}}{\dfrac{\sqrt{\sum_{i=1}^{n} z_i^2}}{2}}$$

where K = sum of the positive normal scores; $K = \sum_{i=1}^{n} z_i^{(+)}$

H. FRIEDMAN TEST

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1)$$

where N = number of rows
k = number of columns
R_j = sum of ranks in jth column
$\sum_{j=1}^{k}$ directs one to sum the squares of the sums of ranks over all k conditions

I. KENDALL COEFFICIENT OF CONCORDANCE

$$W = \frac{s}{\frac{1}{12}k^2(N^3 - N)}$$

where s = sum of squares of the observed deviations from the mean of R_j, that is, $s = \sum\left(R_j - \frac{\Sigma R_j}{N}\right)^2$

k = number of sets of rankings, e.g., the number of judges

N = number of entities (objects or individuals) ranked

$\frac{1}{12}k^2(N^3 - N)$ = maximum possible sum of the squared deviations, i.e., the sum s which would occur with perfect agreement among k rankings

PROBABILITY DISTRIBUTIONS

A. BINOMIAL

$$P(X=x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

where n = number of trials
p = probability of success on any trial
q = 1 - p

B. HYPERGEOMETRIC DISTRIBUTION

$$P(X=x) = \frac{\binom{b}{x}\binom{r}{n-x}}{\binom{b+r}{n}}$$

where b = number of successes in the population
r = number of failures in the population
n = sample size

C. POISSON DISTRIBUTION

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots$$

where λ = mean rate of occurrence

D. EXPONENTIAL DISTRIBUTION

$$P(T < t) = 1 - e^{-\lambda t}$$

where λ = mean rate of occurrence

E. NORMAL DISTRIBUTION

$$F(x) = P(X \leqq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv$$

F. F DISTRIBUTION

$$f(u) = \begin{cases} \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \nu_1^{\nu_1/2}\nu_2^{\nu_2/2} u^{(\nu_1/2)-1}(\nu_2+\nu_1 u)^{-(\nu_1+\nu_2)/2} & u > 0 \\ 0 & u \leqq 0 \end{cases}$$

G. STUDENT'S t DISTRIBUTION

$$f(t) = \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\dfrac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \qquad -\infty < t < \infty$$

H. CHI-SQUARE DISTRIBUTION

$$P(\chi^2 \leqq x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^{x} u^{(\nu/2)-1} e^{-u/2}\, du$$

HYPOTHESIS TESTS: PROPORTIONS

A. TWO PROPORTIONS FROM INDEPENDENT GROUPS

$$Z \cong \frac{(p_{s_1} - p_{s_2}) - 0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

B. PROPORTION VS. HYPOTHESIZED VALUE

$$Z \cong \frac{p_s - p}{\sqrt{\frac{p(1-p)}{n}}}$$

C. TWO PROPORTIONS FROM ONE GROUP, MUTUALLY EXCLUSIVE CATEGORIES

$$Z = \frac{(p_1 - p_2) - 0}{\sqrt{\frac{1}{n}\left[p_1(1-p_1) + p_2(1-p_2) + 2\,p_1 p_2\right]}}$$

D. TWO PROPORTIONS FROM ONE GROUP, OVERLAPPING CATEGORIES

$$Z = \frac{(p_1 - p_2) - 0}{\sqrt{\frac{1}{n}\left[p_1(1-p_1) + p_2(1-p_2) + 2(p_1 p_2 - p_{12})\right]}}$$

APPENDIX D

Microstat File Structure

Even though the major part of DMS is in source code, a number of users have requested a more detailed description of how the files are used in Microstat. That's the purpose of this appendix. Both the header and the random access file are discussed.

Header File

The function of the header file is to stored the data necessary about the file for use in subsequent programs. It is a simple sequential file and is read as follows:

```
10 DIM A$(M),X(M):SP$=SPACE$(5)
.
.
100 OPEN "I",#1,N$:INPUT #1,Q5,N,M,D$
110 FOR J=1 TO M
120  INPUT #1,A$(J):RSET SP$=A$(J):A$(J)=SP$
130 NEXT J
140 INPUT #1,Z$
```

where:

Q5 = 4 if the numbers in the random file are single precision and will equal 8 if they are double precision numbers.

N = the number of cases in the random file

M = the number of variables in the random file

A$(J) = the variable name of the Jth variable

Z$ = the name of the header file plus "R"; hence it becomes the name of the random file (i.e., if the header file is TEST, Z$ will be TESTR)

Since the header file is always read first, the information needed to read the random access file that contains the actual numbers is now known. Note that Q5 tells how many bytes are needed for the random access FIELD statement, while N and M determine the number of cases and variables respectively. The normal method of reading the random file is as follows:

```
200 OPEN "R",#2,Z$,Q5:FIELD #2,Q5 AS T$
210 FOR J=1 TO N
220  FOR K=1 TO M
230   GET #2,K+(M*(J-1))
240   IF Q5=4 THEN X(K)=CVS(T$) ELSE X(K)=CVD(T$)
.
.      (usuallyMicrostatdoescalculations heresothe
.       entire file does not have to be read into memory)
.
400  NEXT K
410 NEXT J
420 CLOSE #2
```

where all of the variables have been defined at the time the header file was read. The only exception is X(K) which is simply the number after is has been converted by the CVS or CVD statement. Note that the K loop is equivalent of reading one case at a time. If the entire file is to be read, the J loop would be read from J=1 to N*M and the K loop removed.

The random file can be though of as a matrix that is M variables across and N cases deep. The method of access in line 230 above allows any individual element in the matrix to be accessed.

In both files, the read-write procedures are usually the same.

NOTE: Users of earlier versions of Microstat will have data files that cannot be read by the new version because of changes in the header file. It is **not**, however, necessary to re-enter all of the data. All that needs to be done is re-write the header file. The following steps will recreate the new file header.

1. Use your old version of Microstat to list the contents of the old file header. The LIST option of DMS will do this. You need to do this to find out the contents of the old header file so it can be re-entered in the new header file.

2. Use the DESTROY FILES option of DMS to destroy the old header file. Do **not** destroy its companion "R" (random) file.

3. Use the new Microstat CREATE FILE to create the new header file to replace the old header file just erased. Note: the drive prefix in the new version should be the drive that contains the old data file. The numeric precision must be the same as that

used to create the old data file.

4. When you see the message that "File XXXX has been created with YYYY cases...", CONTROL-C the program. This will be just prior to the point where the program asks you to start entering the data for the file. The new Microstat has now created the new header file and is asking you to enter its associated data. Since the data are already contained in the old random file, and there were no changes in it for the new version, there's no need to re-enter the actual numbers; hence the CONTROL-C. It's not very elegant, but it works and will save re-entering the numbers.

5. Repeat the steps for all of the old header files.