Radio Shack

TRS-80

STATISTICAL ANALYSIS

Catalog Number 26-1703

User Instruction Manual

CUSTOM MANUFACTURED IN U.S.A. FOR RADIO SHACK TC A DIVISION OF TANDY CORPORATION

Statistical Analysis

A system of computer programs designed
for the routine analysis of data
in business, education, medicine,
government administration,
and other fields.

Written by

Dr. Stephen W. Hebbler

for use on the Radio Shack® TRS-80™
micro computer system.

© 1978 Radio Shack
℗ 1978 Radio Shack

Table of Contents

Introduction. 4

Description of the System. 4

Programs in the Statistical Analysis System 6

DATA FILE PREPARATION 7

DATA FILE LISTER. 11

RANDOM SAMPLE. 13

DESCRIPTIVE STATISTICS. 15

HISTOGRAM. 19

T-TEST 23

SIMPLE CORRELATION. 27

LINEAR REGRESSION. 31

Program Conversion for LEVEL II BASIC 36

Bibliography. 37

Appendix A: MAXIMUM DATA SET SIZES. 39

Appendix B: STATISTICAL ANALYSIS FILE STRUCTURE 41

Introduction

Statistical Analysis is a user oriented data analysis system designed for use on the Radio Shack TRS-80 micro computer. The system is ideally suited for applications in fields such as business, education, medicine, and government administration. The programs can be run by persons with little formal knowledge of data analysis techniques and no knowledge of computer programming. Each program in the system was written to "interact" with the user and to guide him/her in conducting routine analyses.

Description of the System

The Statistical Analysis system consists of eight computer programs stored on cassette tapes and a comprehensive manual which takes the user through each program step-by-step. The system includes five programs for describing data sets and conducting statistical data analyses; two "utility" programs for preparing, updating, and listing data files stored on tape; and a program to aid in selecting data samples. Programs supplied with the Statistical Analysis system are listed below.

DATA FILE PREPARATION
DATA FILE LISTER
RANDOM SAMPLE
DESCRIPTIVE STATISTICS
HISTOGRAM
T-TEST
SIMPLE CORRELATION
LINEAR REGRESSION

Statistical Analysis was designed to run under Radio Shack Level I BASIC. The amount of data which can be analyzed using the system depends upon how much memory (RAM) is installed in the TRS-80. Appendix A contains a chart comparing maximum data set sizes for different amounts of memory.

Data and Data Files

The user's data may be entered either from the TRS-80 keyboard or from a data file stored on cassette tape. The mode for data entry is selected by the user at the beginning of each program.

There are two "types" of data. The DESCRIPTIVE STATISTICS, HISTOGRAM, and T-TEST programs require that data values be input one at a time ("single" type data) whereas the SIMPLE CORRELATION and LINEAR REGRESSION programs require the user to enter data in pairs ("paired" type). Both types of data files can be prepared using the DATA FILE PREPARATION program. The structure of Statistical Analysis files is described in Appendix B.

When data are to be entered from the keyboard, the computer instructs the user to input either data elements or data pairs. When using a data file on cassette tape the computer displays the name of the data file on the screen while data are being read. Additionally, the wrong type of data cannot be read into a program – an error message is displayed and the program is terminated.

Some Words of Caution

Although many safeguards such as the ones above are built into the Statistical Analysis system, users are urged to become familiar with the programs, test them using sample sets of data, and follow the displayed instructions carefully. When in doubt, consult the user's manual.

Simple errors such as entering incorrect data, using the wrong data file (always check the data file name when it is displayed) or providing the computer with the wrong code when it asks for an instruction can result in output that is erroneous. Computer programmers refer to this phenomena as "Garbage in — Garbage out."

The statistical procedures used in the Statistical Analysis system require that all data values (whether entered as single values or as pairs) be actual "measurements." That is, the data values must NOT be codes referring to categories such as "1" for male, "2" for female; "1Ø" for New York, "6" for California; NOR rankings such as "1" for first or largest, etc. Examples of valid data include values such as temperature, age, test or attitude scores, time, cost, length, weight, miles-per-gallon, and numbers of people or objects, etc. (statisticians call these *interval scale measures*). For those interested, there are statistical procedures designed to analyze coded values and ranks — see the Bibliography.

Just like a calculator (or a person), a computer can never be perfectly accurate all the time. For short arithmetic expressions, the TRS-80 is accurate to the fifth or sixth decimal place. In longer, more complex expressions, a minute error in the sixth place can be magnified to where it becomes significant. All programmers have to cope with this kind of built-in error.

Finally, if you feel uncomfortable using one or more of the Statistical Analysis procedures, it would be wise to consult one of the books listed in the Bibliography. The publications were chosen with the understanding that users of the system would not be experts in statistics and, thus, most are of an elementary nature. The list includes books emphasizing several different fields.

NOTE: The Statistical Analysis program was originally written for a LEVEL I TRS-80 microcomputer. However, it is easily converted for use with a LEVEL II machine by using the program conversion tape that came with your microcomputer and a few simple program changes. (Refer to page 36).

Programs in the Statistical Analysis System

The following sections of the users' manual will describe the eight computer programs included in the Statistical Analysis system. Each section includes:

1. A description of the computer program and, where applicable, an introduction to the data analysis technique.
2. Step-by-step instructions for running the program.
3. A sample run of the program using a small set of data (the sample data may be run by the user for practice and to verify that the program is operating properly).
4. Messages and special considerations pertaining to the program.

Descriptions of the data analysis procedures are meant to be illustrative rather than comprehensive. For users who are unacquainted with a particular technique, a Bibliography is included in the manual.

NOTE: The value "99999" is used in the programs as the signal for "END OF DATA". Values between approximately 99998.9 and 99999.1 may, because of internal rounding, be seen by the computer as the "END OF DATA" signal. If values in that range occur in your data set, input 99999.1 instead. The effect on your results will be negligible.

DATA FILE PREPARATION

Description of the Program

Data may be stored on cassette tapes for use as input in any of the data analysis programs. The DATA FILE PREPARATION program allows the user to prepare, correct, update, and copy data files. The program automatically assigns a file type code to the file depending on the data analysis program(s) for which the user is preparing the data. Additionally, a name (entered by the user) is assigned to the data file and the data values and necessary signal values are written on tape.

The user can easily update a data file by entering appropriate instructions while the program is running. Data elements (or pairs) may be removed from an old data file each time the program is run and new data elements (or pairs) may be added. If the user instructs the computer to neither remove nor add data elements, the program will simply copy the file to another tape.

When updating or copying an old data file, the user can enter the original name of the old file or re-name the updated file (or copy). This feature is very useful for keeping tabs on file updates if the user includes the date as part of the file name (e.g. "SALES 11-11-77"). The file name may contain up to 16 characters.

How to Run DATA FILE PREPARATION

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will reply,

```
"PROGRAM FOR WHICH THE DATA ARE BEING PREPARED:
     1 = DESCRIPTIVE STATISTICS, HISTOGRAM, OR T-TEST
     2 = SIMPLE CORRELATION OR LINEAR REGRESSION  ?"
```

   Answer accordingly. If updating a tape, enter the program for which the old file was prepared.

3. The computer will then ask,

```
"THIS PROGRAM IS BEING RUN TO:
     1 = PREPARE A NEW DATA FILE
     2 = UPDATE AN OLD DATA FILE  ?"
```

   Enter the appropriate code (enter a "2" if you are only copying a tape). If you are preparing a new data file, skip to step #7.

4. If you entered a "2", the computer will ask,

```
"HOW MANY DATA ELEMENTS (OR PAIRS) ARE TO BE REMOVED ?"
```

   Type in the number of data elements (or pairs) to be removed.

   If you do not wish to remove incorrect or obsolete data elements or pairs, enter a "Ø".

If you did not enter a "Ø", the reply, "BEGIN ENTERING THE ELEMENT (OR PAIR) #'S" will appear. BE CAREFUL! If you have recently updated the old file, the current data values may not be in their original positions since removal of an element (or pair) from the data requires some of the remaining data to be moved to new positions in the file.

NOTE: If you are not sure exactly what the data file "looks like", run DATA FILE LISTER. Each data element (or pair) has a number which is printed by this program.

Begin entering the element (or pair) numbers for the data values that you want removed from the old file. Enter one element (or pair) number after each question mark appears.

5. If you choose not to remove data elements from the old file (or after you enter the numbers of the element(s) to be removed) the computer will display, "LOAD FILE TO BE UPDATED. SET TO 'PLAY' (1=READY)?" Be sure the old data file is rewound and set the recorder controls to "play." Then type "1" and hit ENTER.

6. The computer will begin reading the data and display the name of the file (check to be sure you have loaded the correct tape).

   When the data have all been read, the computer will print, "2Ø DATA ELEMENTS (OR PAIRS) WERE READ" ("2Ø" is, of course, only an example).

   If data elements or pairs were removed from the old file, the computer will reply "2 DATA ELEMENTS (OR PAIRS) WERE REMOVED". Both of the messages above should be checked to verify that all the old data were read and that the correct number of data values were removed.

7. Depending on which program you choose at instruction #2, the computer will reply with either,

   "BEGIN ENTERING YOUR DATA ELEMENTS.
   SIGNAL END OF DATA WITH 99999. ?"

   or

   "BEGIN ENTERING YOUR DATA PAIRS (X,Y).
   SIGNAL END OF DATA WITH 99999,99999. ?"

   If you will be entering single data elements, type your first value after the question mark and hit ENTER. Another question mark will appear. Continue to enter the remaining data values then enter "99999".

   If you will be entering data pairs, enter each pair after a question mark, separating the X and Y (first and second) values with a comma. After all your data pairs have been entered, type and enter "99999,99999".

If you are copying a tape, or if no new data are to be added, enter only the end of data signal, "99999" or "99999,99999", after the question mark.

8. The computer will respond, "NEW DATA COUNT = 21 DATA ELEMENTS (OR PAIRS)." For updated files, this count takes into consideration the element(s) removed and those added.

   The question, "NAME FOR THE NEW DATA FILE?" will also be printed. You may select any name up to 16 characters in length. It is usually best to give the file a name which will remind the user of what the data values represent (e.g., SALES FOR JUNE, HT VS. WT, TAXUPDATE12Ø376, etc.). When updating or copying a file, you may give the new file the old name or a new one.

9. The computer will print, "LOAD A BLANK TAPE AND SET CONTROLS FOR RECORDING (1=READY) ?" You may either load a different tape into the cassette recorder or rewind and cue the tape containing the old data file. It is NOT ADVISABLE, however, to "write over" the old file if you do not have a backup tape (copy). This is especially true if the old data file is important and/or lengthy. When you have prepared the tape for recording (don't forget to "cue" tapes which have plastic leaders!) type "1" and hit ENTER. The new tape will be written and the program will stop.

Sample Run:

```
        D A T A   F I L E   P R E P A R A T I O N

PROGRAM FOR WHICH THE DATA ARE BEING PREPARED:
     1=DESCRIPTIVE STATISTICS, HISTOGRAM, OR T-TEST
     2=SIMPLE CORRELATION OR LINEAR REGRESSION   ?1


THIS PROGRAM IS BEING RUN TO:
     1=PREPARE A NEW DATA FILE
     2=UPDATE AN OLD DATA FILE   ?2
```

```
HOW MANY DATA ELEMENTS (OR PAIRS) ARE TO BE REMOVED   ?2
BEGIN ENTERING THE ELEMENT (OR PAIR) #'S
?14
?18

LOAD FILE TO BE UPDATED. SET TO 'PLAY' (1=READY)   ?1
```

```
DATA FILE BEING READ = FICA WITHHELD
20 DATA ELEMENTS (OR PAIRS) WERE READ
2 DATA ELEMENTS (OR PAIRS) WERE REMOVED
BEGIN ENTERING YOUR DATA ELEMENTS.
SIGNAL END OF DATA WITH 99999.

?452
?323
?510
?99999
```

```
NEW DATA COUNT = 21 DATA ELEMENTS (OR PAIRS).
NAME FOR THE NEW DATA FILE ?FICAUP6/2/77
LOAD A BLANK TAPE AND SET CONTROLS FOR RECORDING (1=READY)?1
```

Messages and Special Considerations

"WRONG DATA FILE TYPE" means that you instructed the computer to read a single type data file (entered a 1 when asked "PROGRAM FOR WHICH DATA IS BEING PREPARED") then loaded a tape containing paired type data — or vice versa.

"DUPLICATE:
BEGIN ENTERING THE ELEMENT (OR PAIR) #'S"
means that you entered a duplicate element number when listing the elements to be removed. You will need to enter the element numbers again.

"CAN'T REMOVE #38" means that you asked the program to remove a data element (or pair) which did not exist on the data file (the computer reached the end of the data before the displayed element was reached).

REMEMBER: Whenever a data element (or pair) is removed from a file, the element (or pair) numbers change for some of the remaining data. Use DATA FILE LISTER to check on the positions of data values. Data values which are added during an update are positioned at the end of the old file.

The number of data elements or pairs which can be removed from an old file is directly related to the number of data elements or pairs already on the file. For example, with 4K RAM Memory and 400 elements (or 200 pairs) on the old file, only 10 elements (5 pairs) can be removed using one run of the program. If you accidentally ask for too many values to be removed, you will receive a message of "SORRY" before the computer finishes reading the file to be updated. No harm will be done, however, and you may re-run the program specifying fewer records to be removed. For every element less than 400, another value may be removed.

DATA FILE LISTER

Description of the Program

Data files may be listed on the TRS-80 screen using DATA FILE LISTER. The program offers fast, moderate, and slow listing speeds and lists both single and paired type data files. The file type and file name are displayed along with each data element (or pair) and the associated element (or pair) number.

How to Run DATA FILE LISTER

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will reply, "LISTING SPEED (1=SLOW, 2=MODERATE, 3=FAST) ?" Select a speed and enter the appropriate code. The slow speed is useful if you wish to check the values on the screen against a written or typed list while the program is running.

3. The computer will ask, "DATA FILE TAPE LOADED AND SET TO PLAY (1=YES) ?" Be certain your data tape is rewound and ready to play, then enter a "1". The computer will begin reading the data and display the name of the file and the data file type (SINGLE or PAIRED). After all the data have been read the data element (or pair) numbers, beginning with "1", will be displayed along with the data (or pair) values.

4. After all the data have been listed, the computer will respond, "(1=LIST DATA AGAIN, 2=STOP) WHICH ?"

Sample Run:

```
        D A T A   F I L E   L I S T E R

LISTING SPEED (1=SLOW, 2=MODERATE, 3=FAST) ?2

DATA FILE TAPE LOADED AND SET TO PLAY (1=YES) ?1
```

```
FILE NAME = HT VS. WT                    FILE TYPE = PAIRED
-------------------------------------------------------------

  DATA ELEMENT                        VALUE(S) READ
  (OR PAIR) #                          FROM TAPE
  ------------                        -------------

       1                               68        16Ø
       2                               69        18Ø
       3                               7Ø        17Ø
       4                               7Ø        2ØØ
       5                               71        21Ø
       6                               71        18Ø
       7                               72        19Ø
       8                               72        2ØØ
       9                               73        21Ø
      1Ø                               74        2Ø5

  (1=LIST DATA AGAIN, 2=STOP) WHICH ?
```

Messages and Special Consideration

NOTE: If you want to stop the listing in order to examine a data value, hit the button marked "BREAK". The computer will stop listing and the words "BREAK AT ØØØ" and "READY >" will appear. In order to restart the program, you must type "CONT" then hit ENTER. The listing may be stopped momentarily by simply holding down any key on the keyboard.

"NOT A 'GENERAL DATA ANALYSIS' DATA FILE." means that the file which was loaded contained neither single nor paired type data and was probably not prepared for use in Statistical Analysis programs.

RANDOM SAMPLE

Description of the Program

Procedures such as the t test and linear regression require a randomly selected sample of measurements as input data. RANDOM SAMPLE aids the user in selecting such a sample, Sampling procedures available include sampling with and sampling without replacement. The sample data element numbers are printed out in numerical order for convenience.

How to Run RANDOM SAMPLE

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will ask "WHAT IS THE TOTAL POPULATION SIZE?" Enter the number of persons, objects, packages, etc. in the total group.

3. "WHAT SIZE SAMPLE DO YOU DESIRE?" will be displayed. Enter the number of persons, objects, etc. that you want as your sample.

4. The computer will reply,

```
"SAMPLING PROCEDURES AVAILABLE:
      1=SAMPLING WITH REPLACEMENT
      2=SAMPLING WITHOUT REPLACEMENT  WHICH ?"
```

   Select a procedure and enter the appropriate value. If you enter a "1", each member of the total group can be selected as a member of the sample more than once.

5. The computer will print "COMPUTER AT WORK - PLEASE BE PATIENT". Depending on the size of the sample you are selecting it may take quite a while before the selection is completed.

6. The element (or pair) numbers which make up your sample will be displayed on the screen. Copy these down for use in preparing your sample data. If your sample is larger than 48, the computer will stop listing numbers as it fills the screen. The message, "ENTER A NUMBER TO CONTINUE LIST" will appear. When you have copied the element numbers from the screen, type a number and hit ENTER.

7. When all sample element numbers have been listed the computer will print, "(1=SELECT ANOTHER SAMPLE, 2=STOP) WHICH ?"

Sample Run:

```
                R A N D O M   S A M P L E

WHAT IS YOUR TOTAL POPULATION SIZE  ?1ØØ

WHAT SIZE SAMPLE DO YOU DESIRE ?2Ø

SAMPLING PROCEDURES AVAILABLE:
     1=SAMPLING WITH REPLACEMENT
     2=SAMPLING WITHOUT REPLACEMENT  WHICH ?2

COMPUTER AT WORK - PLEASE BE PATIENT.
```

```
YOUR SAMPLE WILL CONSIST OF MEASUREMENTS ON THE 2Ø DATA
ELEMENTS (OR PAIRS) NUMBERED:

  6          7          17         18
 21         23          3Ø         34
 35         39          48         59
 6Ø         63          65         67
 7Ø         76          83         86

(1=SELECT ANOTHER SAMPLE, 2=STOP)  WHICH ?
```

Messages and Special Considerations

The largest total population size from which a sample can be drawn is 32767.

If you run this program for practice using the population and sample sizes above and obtain "results" that are different from the sample run, don't panic — that is the way the program is supposed to work! Remember, it is drawing a *random* sample.

DESCRIPTIVE STATISTICS

Description of the Program

DESCRIPTIVE STATISTICS provides the user with an overall picture of his/her data. Output from the program includes measures of central tendency (mean, median, and mode); measures of dispersion (variance, standard deviation, and range); minimum and maximum data values; and sample size (N).

The mean is the arithmetic average of all the data values and the median is the value above and below which half the data elements fall (in the case of an even number of data elements, the median is the arithmetic mean of the two "middle" data values). The mode is the value which occurs most often in the data set. If no value occurs more than once or if there is a "tie" for the greatest frequency of occurrence, an appropriate message is displayed in place of the numeric value for the mode.

Variance is the total of the squared deviations of each data value from the mean divided by the sample size minus 1. Standard deviation is the square root of the variance and can be used to directly compare the dispersion (spread) of values in different data sets. Range is simply the maximum data value minus the minimum data value.

How to Run DESCRIPTIVE STATISTICS

1. Load the program into the TRS-80 and enter "RUN."

2. The computer will ask, "HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?" Answer accordingly.

   If you enter a "1" the computer will respond with, "BEGIN ENTERING YOUR DATA. SIGNAL END OF DATA WITH 99999. ?" Type your first data value after the question mark and hit ENTER. Another question mark will appear. Continue to enter the remaining data values then enter "99999".

   If you enter a "2" the computer will ask, "DATA FILE TAPE LOADED (1=YES) ?" Be sure that your data file tape is loaded and rewound and that the recorder controls are set to "play". Then enter a "1". The computer will begin reading your data and the name of the data file will appear on the screen. Check the name of the file to be certain that the correct data are being read.

3. The message "DATA BEING SORTED--PLEASE BE PATIENT" will appear followed by the program results.

   NOTE: Depending on the size of your data set, it may take quite a while for your data to be sorted. For other information concerning sorting time for your data see Messages and Special Considerations.

4. After the results of the program have been displayed, the computer will ask, "WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?"

Sample Run:

```
D E S C R I P T I V E   S T A T I S T I C S

HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?
```

```
BEGIN ENTERING YOUR DATA.
SIGNAL END OF DATA WITH 99999.
?64
?48
?55
?68
?72
?59
?57
?61
?63
?60
?60
?43
?67
?70
?65
?55
?56
?64
?61
?60
?99999
```

OR

```
DATA FILE TAPE LOADED (1=YES) ?1

DATA FILE BEING READ=DISHWASHERS
```

```
DATA BEING SORTED
PLEASE BE PATIENT
```

```
D E S C R I P T I V E   S T A T I S T I C S

MEASURES OF CENTRAL TENDENCY:

   MEAN              = 6Ø.4
   MEDIAN            = 6Ø.5
   MODE              = 6Ø

MEASURES OF DISPERSION:

   VARIANCE          = 48.9897
   STD. DEV.         =  6.99927
   RANGE             = 29
            (MINIMUM= 43, MAXIMUM = 72)

SAMPLE SIZE (N)      = 2Ø

WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?
```

Messages and Special Considerations

"WRONG DATA FILE TYPE (SEE MANUAL)" means that the data file which was loaded contained data pairs (one X measurement, one Y measurement) rather than single data values on one variable.

" THE DATA DOES NOT HAVE A MODE" is printed when each data value occurs exactly once.

"MORE THAN ONE MODE EXISTS IN THE DATA" is printed when there is a "tie" for the greatest frequency of occurrence or when each data value occurs more than once and the same number of times (rectangular distribution).

The variance and standard deviation are calculated assuming that the data were collected on a sample rather than the total population. This is a necessary condition when working with samples in data analysis and results in negligible differences when the total population is used.

DATA SORTING: Calculation of the median and mode requires that the data be in "order". After all the data have been entered, DESCRIPTIVE STATISTICS sorts the data and performs the necessary calculations. Sorting is one of the most demanding operations which can be performed on a computer — as the number of elements is increased, the sorting time required increases at an even greater rate. Sorting time is also affected by the order in which the data are entered. If the data are input in ascending order (smallest values first — largest values last) DESCRIPTIVE STATISTICS will automatically bypass the sorting operation. Examples of sort times are given in the table on the next page.

No. of Elements	Order of Elements	Approx. Time Required
1 to 400	Perfect Ascending	No more than 6 seconds
20	Mixed	10 seconds
100	1 element out of order	2 minutes
200	Descending (Backwards)	15 minutes
400	Descending (Backwards)	Over 60 minutes

The considerations above apply only to DESCRIPTIVE STATISTICS. Ordering of data has no affect on the other programs in the Statistical Analysis system.

HISTOGRAM

Description of the Program

HISTOGRAM allows the user to obtain a graphic description of his/her data set. The histogram is drawn with from one to eight equal intervals as selected by the user. Both frequencies and percentages are labeled on the histogram and each interval is plotted with considerable accuracy. The number of intervals on the histogram can be changed at will without the need for re-entering the data.

How to Run HISTOGRAM

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will ask, "HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?" Answer accordingly.

   If you enter a "1" the computer will respond with, "BEGIN ENTERING YOUR DATA. SIGNAL END OF DATA WITH 99999. ?" Type your first data value after the question mark and hit ENTER. Another question mark will appear. Continue to enter the remaining data values then enter "99999".

   If you enter a "2" the computer will ask, "DATA FILE TAPE LOADED (1=YES) ?" Be sure that your data file tape is loaded and rewound and that the recorder controls are set to "play". Then enter a "1". The computer will begin reading your data and the name of the data file will appear on the screen. Check the name of the file to be certain that the correct data are being read.

3. The number of data elements and the minimum and maximum data values will be printed. The computer will then ask, "HOW MANY INTERVALS FOR HISTOGRAM (1 THROUGH 8) ?" Enter the number of intervals you want the histogram to have.

   NOTE: At least one of the intervals on the histogram must contain six or more values in order for the plotting function to operate properly. If, after calculating the frequency for each interval, none of the intervals contain at least six values, an error message will be displayed (see Messages and Special Considerations).

4. The computer will plot one horizontal and two vertical lines on the screen and will label the horizontal line with the interval boundaries. After a short time (depending on the size of the data set) the frequencies and percentages will be placed along the vertical lines and the histogram intervals will be plotted.

5. "ENTER A NUMBER TO CONTINUE PROGRAM ?" will also appear on the screen. When you have finished viewing the histogram, type and enter any number.

6. The program will respond, "(1=NEW INTERVALS, 2=RUN AGAIN, 3=STOP) WHICH ?" If you enter a "2", the program will start over and ask you for a new data set.

7. If you enter a "1", the program will again print the number of data elements and the minimum and maximum data values then ask, "HOW MANY INTERVALS FOR HISTOGRAM (1 THROUGH 8) ?" You may redraw the historgram changing the number of intervals as many times as you like.

Sample Run:

```
                    H I S T O G R A M

HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?
```

```
BEGIN ENTERING YOUR DATA.
SIGNAL END OF DATA WITH 99999.
?64
?48
?55
?68
?72
?59
?57
?61
?63
?60
?60
?43
?67
?70
?65
?55
?56
?64
?61
?60
?99999
```

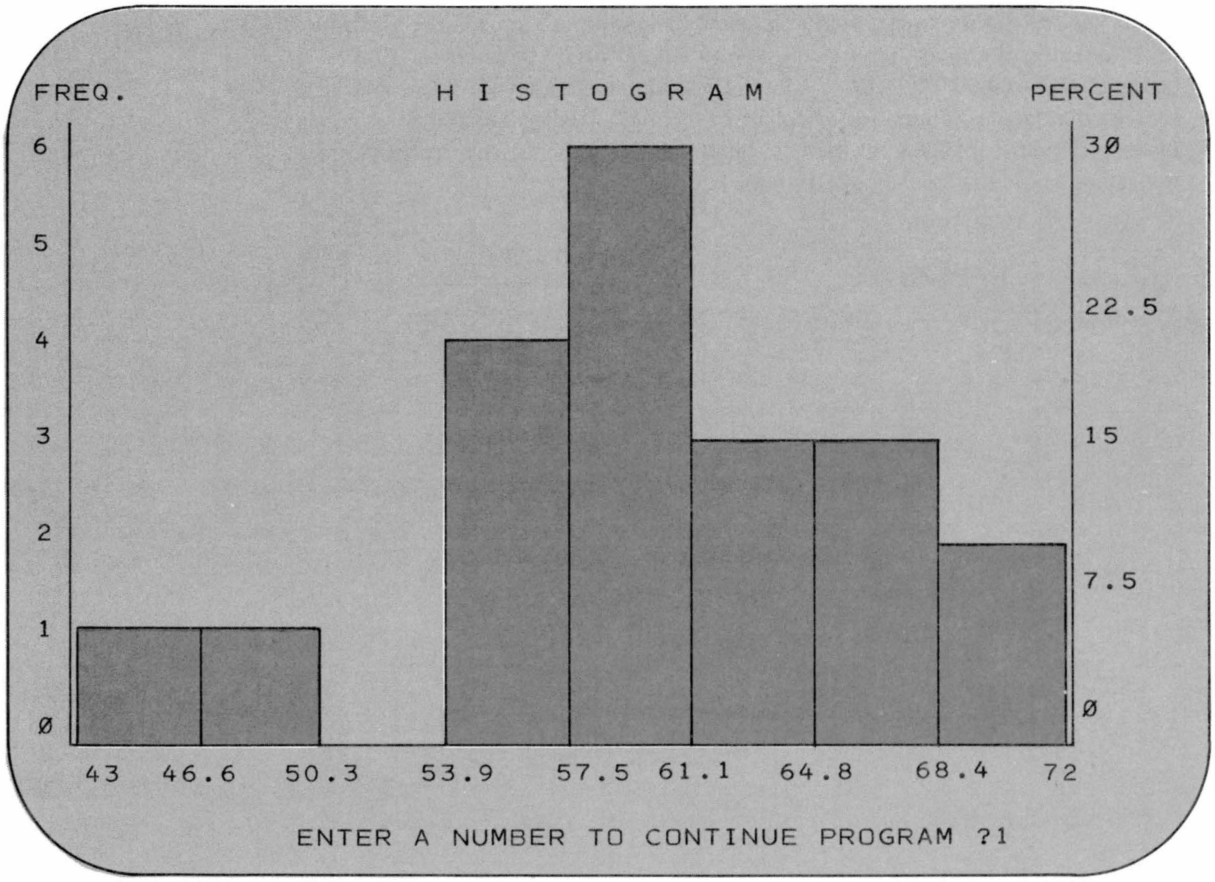
OR

```
DATA FILE TAPE LOADED (1=YES) ?1

DATA FILE BEING READ=FICA WITHHELD
```

```
NUMBER OF DATA ELEMENTS (N) = 2Ø
MINIMUM DATA VALUE          = 43
MAXIMUM DATA VALUE          = 72

HOW MANY INTERVALS FOR HISTOGRAM (1 THROUGH 8) ?8
```




```
(1=NEW INTERVALS, 2=RUN AGAIN, 3=STOP) WHICH ?
```

Messages and Special Considerations

"WRONG DATA FILE TYPE (SEE MANUAL)" means that the data file which was loaded contained data pairs (one X measurement; one Y measurement) rather than single data values on one variable.

"ONE INTERVAL MUST CONTAIN AT LEAST 6 VALUES TRY FEWER INTERVALS (YOU TRIED 5 LAST TIME)." is self-explanatory. Any data set containing a total of 6 or more values can be plotted (try asking for 1 interval) but you may have to settle for fewer intervals than you wanted.

Data values input in the HISTOGRAM program may range from −3276.7 to +3276.7 only. To insure that the program will run properly, and to provide for readable values on the X axis of the histogram figure, it is recommended that all data values be "coded" down to consist of a maximum of three (3) digits before the decimal point. Any valid coding scheme, such as division by a constant, subtraction of a constant, etc. may be used. The number of digits following the decimal point has no effect on the operation of the program.

For example, a researcher studying income might input each subject's income in "thousands of dollars" (an income of $15,000.00 would be input as 15, an income of $11,450.00 as 11.45, etc.). The resulting histogram would, therefore, present the data as "thousands of dollars of income". Coding can also be used to make histograms involving extremely small values more readable. In this case, the values would be "coded" up, or made larger.

T-TEST

Description of the Program

T-TEST allows the user to test whether the mean of a randomly selected sample of measurements is significantly different from the mean (or mu) of the total group (see the section on program: RANDOM SAMPLE). The *t* ratio, the ratio of (1) the observed difference between the sample and population means and (2) a special measure of dispersion, or spread, for the sample, is calculated when performing a *t* test. As the value of the *t* ratio increases, the probability that the observed difference occurred purely by "chance" decreases. When used in a quality control setting, the "accepted" value for the item being manufactured replaces the total group mean.

Output provided by the program includes the control value, sample mean, standard deviation, sample size, *t* ratio, difference (mean-mu), and degrees of freedom. In addition, the *t* ratio is tested for significance at the 5% level. Since the TRS-80 displays only upper-case characters, the terms *t* test and *t* ratio will be printed "T-TEST" and "T-RATIO". Examples of applications for the *t* test follow.

1. An automatic bagging machine fills 2000 bags with flour during a four hour run. Each bag should contain five pounds of flour. Although minor deviations from this value are acceptable, significant deviations from five pounds will indicate that the machine is in need of adjustment. During each four hour run of the machine, 20 bags of flour are randomly selected from the total number produced and carefully weighed. The weights of the sample bags are used as input for T-TEST. The quality control value of 5.Ø is also input into the program. If the difference observed between the sample mean and the control value (+ or −) could be expected to occur by chance less than five times in a hundred, production is halted.

2. Every year a particular small college performs a test to determine whether the academic aptitude of the incoming freshman class is significantly different (higher or lower) than the average for freshman classes during previous years. An academic aptitude examination is administered to all new freshman students. The mean score for random sample of current freshmen is compared to the overall mean for all previous freshmen using T-TEST.

3. As a final example, a company wants to test a new gasoline additive to see whether it produces a significantly *greater* mileage (MPG). Each car in a random sample of automobiles is provided with a carefully measured quantity of fuel and driven until the fuel is exhausted. The average mileage obtained by the sample is compared to the mileage one would expect due to experience with similar automobiles run without an additive.

   NOTE: In the last example the company would be interested only in additives which produce significantly *greater* mileage. In cases such as this, a directional or one-tailed *t* test is called for (see Messages and Special Considerations).

How to Run T-TEST

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will ask, "HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?" Answer accordingly.

3. "ENTER THE POPULATION MEAN (MU) OR YOUR QUALITY CONTROL VALUE ?" will appear on the screen. Type in the appropriate value then hit ENTER.

4. If you entered a "1" at step #2 the computer will respond with, "BEGIN ENTERING THE DATA FOR YOUR SAMPLE. SIGNAL END OF DATA WITH 99999. ?" Type your first data value after the question mark and hit ENTER. Another question mark will appear. Continue to enter the remaining data values then enter "99999".

   If you entered a "2" the computer will ask, "DATA FILE TAPE LOADED (1=YES) ?" Be sure that your data file tape is loaded and rewound and that the recorder controls are set to "play". Then enter a "1". The computer will begin reading your data and the name of the data file will appear on the screen. Check the name of the file to be certain that the correct data are being read.

5. The computer will display the results of the *t* test on the screen and ask, "WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?"

Sample Run:

```
                    T-TEST

HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?

ENTER THE POPULATION MEAN (MU) OR YOUR QUALITY CONTROL
VALUE ?58.Ø
```

```
BEGIN ENTERING THE DATA FOR YOUR SAMPLE.
SIGNAL END OF DATA WITH 99999.
?64
?48
?55
?68
?72
?59
?57
?61
?63
?6Ø
?6Ø
?43
?67
?7Ø
?65
?55
?56
?64
?61
?6Ø
?99999
```

OR

```
DATA FILE TAPE LOADED (1=YES) ?1

DATA FILE BEING READ = MATH1977
```

```
                T - T E S T   R E S U L T S


                MU OR CONTROL VALUE    = 58

                SAMPLE MEAN            = 6Ø.4
                STANDARD DEVIATION     = 6.99927
                SAMPLE SIZE (N)        = 2Ø

                T-RATIO                = 1.53346
                DIFFERENCE (MEAN-MU)   = 2.4
                DEGREES OF FREEDOM     = 19

WITH A SAMPLE SIZE OF 2Ø A T-RATIO OF (+ OR -) 1.53346
COULD HAVE OCCURRED BY CHANCE MORE THAN FIVE TIMES IN 1ØØ.

   WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?
```

Messages and Special Considerations

"WRONG DATA FILE TYPE (SEE MANUAL)" means that the data file which was loaded contained data pairs (one X measurement; one Y measurement) rather than single data values on one variable.

Users who are familiar with the *t* test procedure and wish to test the *t* ratio at a significance level other than 5%, or need to run a directional (one-tailed) test, can consult a statistics text which contains a table of critical values of *t* (see Bibliography). The program prints the DEGREES OF FREEDOM for this purpose.

NOTE: Source for values in statements 280 through 310 in the T-Test program were extracted from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research (6th Edition),* published by Longman, Essex; by permission of the authors and publisher.

SIMPLE CORRELATION

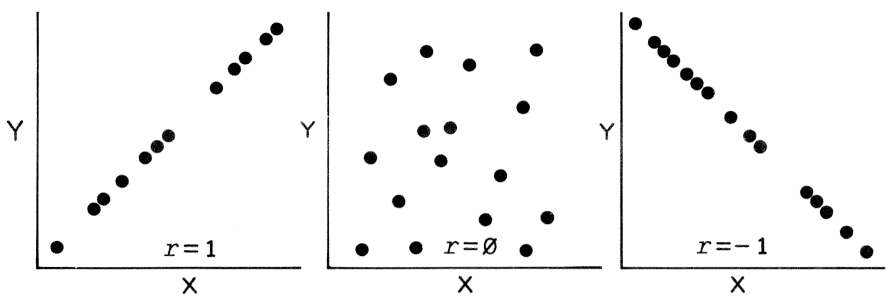
Description of the Program

The objective of SIMPLE CORRELATION is to describe the relationship (correlation) between two variables, or sets of measurements. Questions which the correlation technique might help answer include:

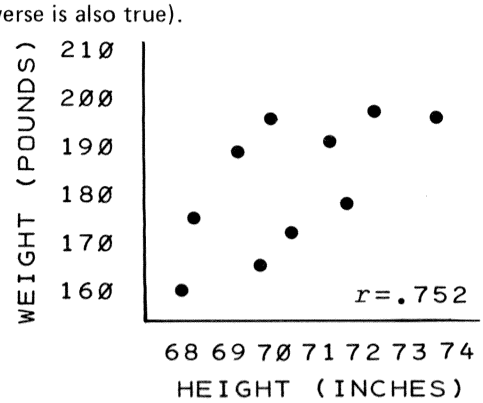
1. Is there a relationship between the number of consumers who buy a new car and the number of prime-time advertising minutes bought by an automobile dealership?

2. Is there a relationship between the number of reported influenza cases and the pollution count (index) in a particular community? Is age a related variable?

SIMPLE CORRELATION uses the *Pearson product-moment correlation coefficient (r)* to describe the degree of relationship within the data. We will let the letter R stand for the coefficient. The value of R depends on the relationship between the two variables and may range from −1 through Ø to +1. An R of Ø indicates that *no* relationship exists; the values −1 and +1 signify a perfect relationship. A positive coefficient indicates that "large" values on one of the variables are paired with "large" values on the other. A negative R occurs when "large" values on one of the variables are paired with "small" values on the other. Examples of different degrees of relationship between two variables are given in the scattergrams below.



If, for example, one were to measure the height and weight for a group of 10 people, there would be two measurement (X and Y) for each individual. These data are illustrated below using a scattergram. It can be seen that some degree of relationship exists between the two variables, height and weight, for the group in our example. That is, people who are tall tend to weigh more than those who are short (the reverse is also true).



A high degree of relationship between two variables must *never* be construed to mean that one of the variables "caused" the other. Correlational techniques cannot test causality. Consider the following example:

> The headlines in a local New York newspaper reads, "A High Degree of Relationship (r=–.98) Found Between the Consumption of Ice Cream Cones in New York City and the Number of Malaria Cases in a Small Undeveloped Village in Southern South America." The finding above might lead New Yorkers to believe that they can help to reduce the death rate in that village by eating more ice cream. Not so! A more reasonable (and logical) explanation for the observed correlation might be that; as the weather in New York becomes colder and the consumption of ice cream cones decreases, the weather in southern South America (southern hemisphere) becomes warmer, resulting in more cases of mosquito-carried diseases. Hence, the high negative correlation between the variables under study.

The program calculates means and standard deviations (S.D.) for the X and Y variables and counts the number of pairs input. The correlation coefficient and degrees of freedom are also calculated. Finally, the program informs the user of the probability that a correlation of the degree found in the data could have occurred by chance (in a random sample selected from a larger population) if the X and Y variables were actually not related (r=Ø) in the total population.

How to Run SIMPLE CORRELATION

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will ask, "HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?" Answer accordingly.

   If you enter a "1" the computer will respond with "BEGIN ENTERING YOUR DATA PAIRS (X,Y). SIGNAL END OF DATA WITH 99999,99999.?" Type your first data pair after the question mark (separate the X and Y values with a comma) and hit ENTER. Another question mark will appear. Continuing entering data pairs then type and enter "99999,99999" after the last pair.

   If you enter a "2" the computer will respond with "DATA FILE TAPE LOADED (1=YES) ?" Be sure that your data file tape is loaded and rewound and that the recorder controls are set to "play". Then enter a "1". The computer will begin reading your data and the name of the data file will appear on the screen. Check the name of the file to be certain that the correct data are being read.

3. The computer will display the results of the product-moment correlation on the screen and ask,

   "WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?"

Sample Run:

```
S I M P L E   C O R R E L A T I O N

HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?
```

```
BEGIN ENTERING YOUR DATA PAIRS (X,Y).
SIGNAL END OF DATA WITH 99999,99999.
?68,16Ø
?69,18Ø
?7Ø,17Ø
?7Ø,2ØØ
?71,21Ø
?71,18Ø
?72,19Ø
?72,2ØØ
?73,21Ø
?74,2Ø5
?99999,99999
```

OR

```
DATA FILE TAPE LOADED (1=YES) ?1

DATA FILE BEING READ = HT VS. WT
```

```
P R O D U C T - M O M E N T   C O R R E L A T I O N

          MEAN OF X = 71
          S.D. OF X = 1.732Ø5
          MEAN OF Y = 19Ø.5
          S.D. OF Y = 16.5
          # OF PAIRS= 1Ø

      CORRELATION COEFFICIENT (R) = .752
      DEGREES OF FREEDOM          = 9

WITH 1Ø DATA PAIRS, A CORRELATION COEFFICIENT OF .752
COULD HAVE OCCURRED BY CHANCE LESS THAN 5 TIMES IN 1ØØ.

WANT TO RUN ANOTHER SET OF DATA (1=YES, 2=NO) ?
```

Messages and Special Considerations

"WRONG DATA FILE TYPE" means that the data file which was loaded contained data values on one variable rather than data pairs (one X measurement; one Y measurement).

"MUST HAVE AT LEAST 3 PAIRS." is self-explanatory.

Users who are familiar with Pearson product-moment correlation and wish to test the correlation at a significance level other than 5% can consult a statistics text which contains a table of critical values of *r* (see the Bibliography). The program prints the DEGREES OF FREEDOM for this purpose.

NOTE: Source for values in statements 400 through 430 in the Simple Correlation program were extracted from Table VI of Fisher & Yates; *Statistical Tables for Biological, Agricultural and Medical Research (6th Edition),* published by Longman, Essex; by permission of the authors and publisher.

LINEAR REGRESSION

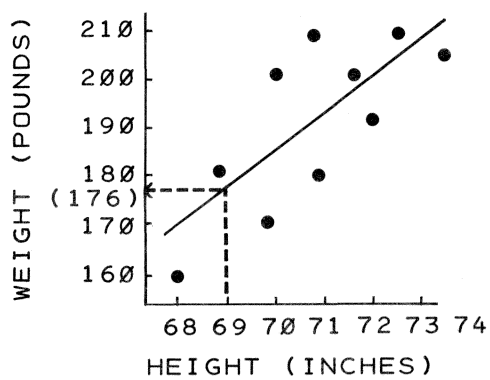
Description of the Program

LINEAR REGRESSION serves to locate the straight line which best "fits" through the data points for two variables or sets of measurements. We call this line the *regression* or *prediction* line and use it in predicting values of the Y variable for different values of X. Questions which LINEAR REGRESSION might help answer include:

1. Can the value of a person's home be predicted from his/her reported annual income?

2. Can a student's success in an advanced science course be predicted from the average of that student's grades in previous courses in mathematics and general science?

Consider again the example used in discussing SIMPLE CORRELATION – height and weight measurements on a sample of 10 people. We can predict a person's weight (Y) given his/her height. Using the regression line in the figure below, we would predict a weight of 176 pounds for a person who is 69 inches tall. If the ten people who made up our sample were randomly selected from a larger group, or population (see the section on program: RANDOM SAMPLE), we could use the same regression line for predicting the weight of any member in the population.



The general formula for the regression line is:

$Y = mX + b$ (Since the TRS-80 uses only uppercase, we will use):

Y = MX + B where

M is called the *slope* of the line, and B is the *Y intercept* and is the value of Y where the regression line crosses the vertical axis (line).

After the values for M and B have been found, predicted values of Y for any given value of X can be calculated by simply substituting X in the equation above. In our example, M = 7.16667 and B = –318.32; the prediction equation becomes:

Y′ = 7.16667(X)+(–318.32)

If we substitute 69 for X, we arrive at a predicted Y value (Y') of 176.18. This value agrees with that derived using the scattergram. For large sets of data, the arithmetic method of prediction is easier to use (LINEAR REGRESSION will calculate predicted values for you).

The degree to which the regression line "fits" the data depends upon how close to the line the data points (X,Y) are. If variables X and Y were perfectly related, all data points would fall directly *on* the regression line. The degree of relationship between the variables can be determined by running SIMPLE CORRELATION on the same data.

LINEAR REGRESSION is a multi-step program. The program first calculates and displays the slope and Y intercept for the data. When instructed by the user, the program plots a scattergram of the data on the screen and draws the regression line in the proper location (the regression line may be left out if you prefer). When instructed to "predict", the program asks the user for a value of X then calculates and displays the corresponding predicted value of Y. If the user enters a value of X which is not between the minimum and maximum X values in the original sample, a warning message is displayed along with the predicted Y value. After each of the steps described above is completed, the user is allowed to re-run any of the steps by entering an appropriate instruction code.

How to Run LINEAR REGRESSION

1. Load the program into the TRS-80 and enter "RUN".

2. The computer will ask, "HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?" Answer accordingly.

   If you enter a "1" the computer will respond with "BEGIN ENTERING YOUR DATA PAIRS (X,Y). SIGNAL END OF DATA WITH 99999,99999. ?" Type your first data pair after the question mark (separating the X and Y values with a comma) and hit ENTER. Another question mark will appear. Continue entering data pairs then type and enter "99999,99999" after the last pair.

   If you enter a "1" the computer will respond with "DATA LOADED (1=YES) ?" Be sure that your data file tape is loaded and rewound and that the recorder controls are set to "play." Then enter "1". The computer will begin reading your data and the name of the data file will appear on the screen. Check the name of the file to be certain that the correct data are being read.

3. The computer will display the slope and Y intercept for the regression line and ask, "(1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN) WHICH ?" You may select any of the options by entering the appropriate code.

4. If you enter a "1" the computer will ask "WANT REGRESSION LINE (1=YES, 2=NO) ?" Enter a "1" or a "2". The computer will draw and label a scattergram (X values on the horizontal axis, Y values on the vertical axis) and plot the data points.

If you wanted the regression line, the line will be drawn on the scattergram at the proper location. The computer will then display, "ENTER A 1 TO CONTINUE". When you have finished viewing the scattergram, type and enter a "1".

5. The computer will again reply, "(1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN) WHICH ?" If you enter a "2", the computer will display, "ENTER 99999 TO STOP" and will set up a table. A question mark will appear and the computer will wait for you to input a value for X. Enter any numeric value within the range of X values in your data. The predicted Y value will be displayed along with another question mark. When you want to stop predicting, enter "99999" in place of an X value.

   NOTE: Values of Y predicted from X values which lie outside the range of X for your data will probably be inaccurate. When an X value outside this range is entered, the predicted value will be accompanied by the message, "(X VALUE OUTSIDE ORIGINAL RANGE)".

6. When you have stopped predicting, the computer will reply, "(1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN) WHICH ?". If you enter a "3" the slope and Y intercept values will be printed and the message above will reappear.

Sample Run:

```
L I N E A R   R E G R E S S I O N

HOW WILL DATA BE ENTERED (1=KEYBOARD, 2=DATA FILE) ?
```

```
BEGIN ENTERING YOUR DATA PAIRS (X,Y).
SIGNAL END OF DATA WITH 99999,99999.
?68,160
?69,180
?70,170
?70,200
?71,210
?71,180
?72,190
?72,200
?73,210
?74,205
?99999,99999
```

OR

OR

```
DATA LOADED (1=YES) ?1
DATA FILE BEING READ = HT VS. WT
```

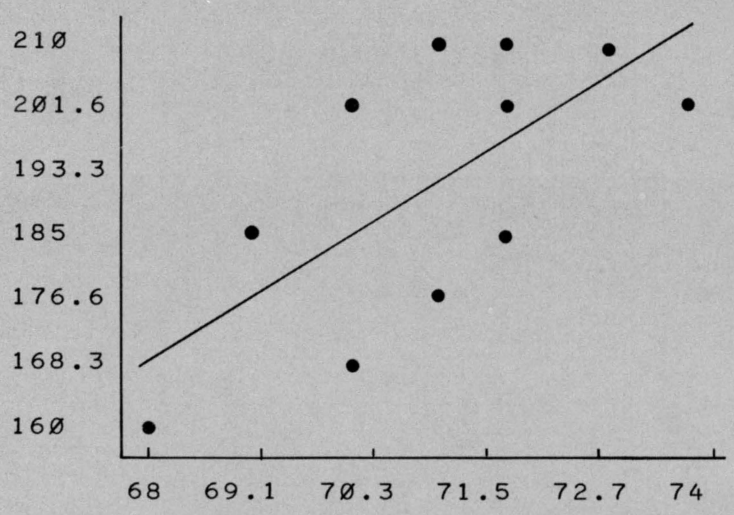
```
        SLOPE (M) OF THE REGRESSION LINE = 7.16667

        THE Y INTERCEPT (B) FOR THE LINE = -318.32

(1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN)  WHICH ?1
```

```
WANT REGRESSION LINE (1=YES, 2=NO) ?1
```

```
ENTER A 1 TO CONTINUE ?1
```




```
1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN) WHICH ?2
```

```
ENTER 99999 TO STOP

 X          PREDICTED Y
-----------------------

?69
              176.18
?70
              183.347
?75
              219.18 (X VALUE OUTSIDE ORIGINAL RANGE)
?65
              147.513 (X VALUE OUTSIDE ORIGINAL RANGE)
?99999

(1=PLOT, 2=PREDICT, 3=RESULTS, 4=STOP, 5=NEW RUN)  WHICH ?
```

Messages and Special Considerations

"WRONG DATA FILE TYPE" means that the data file which was loaded contained data values on one variable rather than data pairs (one X variable; one Y variable).

IMPORTANT: LINEAR REGRESSION calculates the statistics necessary for predicting values of Y from values of X. That is, the X variable is the *predictor.* It is important, therefore, when entering data or preparing a data file to place the predictor variable in the X or first position in the pair. In our example, we predicted people's weight from their height. We could predict height from weight by entering weight as the X variable and height as the Y variable and rerunning the program.

Since the X and Y variables are not labeled on the scattergram, it is important to remember that the X variable is plotted on the horizontal axis; the Y variable on the vertical axis.

Program Conversion for Level II BASIC

The Statistical Analysis program was originally written for the LEVEL I TRS-80 microcomputer. However, it is easily converted for use with a LEVEL II machine through the use of the program conversion tape that came with your microcomputer, and a few simple program changes. The conversion tape is loaded first like any other tape, then the program tape you wish to run is loaded. When the computer completes the conversion process and prints READY, enter the following:

```
5    DIMA (1ØØ)
```

Do this for all Statistical Analysis programs.

The Descriptive Statistics, Histogram, and Linear Regression programs require further changes.

Descriptive Statistics:

```
215    CLS : PRINT @ 47Ø , "DATA BEING SORTED" : PRINT @ 534 ,
       "PLEASE BE PATIENT"
```

Histogram:

```
31Ø    L = 9ØØ : PRINT @ L , INT(U * 1Ø)/1Ø ;
32Ø    FORJ = Ø to B – 1 : L = L + 6 : PRINT @ L , A(N + J);
37Ø    I = Ø : H = Ø : FORJ = 832 TO 64 STEP –128 : PRINT @ J , H ;
385    T = K/N : I = 1 : H = Ø : FORJ = 888 TO 12Ø STEP –192 : PRINT @
       J , H;
39Ø    PRINT @ Ø , "FREQ." ; TAB(22) ; "HISTOGRAM" ; TAB(57) ;
       "PERCENT" ;
51Ø    PRINT @ 97Ø, "ENTER A NUMBER TO CONTINUE PROGRAM" ;
```

Linear Regression:

```
48Ø    PRINT @ J , INT(1Ø * (U + I * (V – U)/6))/1Ø ;
5ØØ    PRINT @ J , INT(1Ø * (K + I * (L – K)/5))/1Ø ;
59Ø    PRINT @ Ø , "ENTER A 1 TO CONTINUE" ;
```

These changes can easily be made by using the EDIT function of the computer. See Chapter 9 of the LEVEL II BASIC Reference Manual for a complete description of this function.

It is a good idea to record these revised programs so it will not be necessary to make these changes again. To record the program, put a tape in the cassette player and set it to record. Enter CSAVE"P" and the tape will record. "P" is a file name and can be replaced by any single letter.

Bibliography

Arkin, H. and Colton, R. *Statistical methods (5th ed.).* New York: Barnes and Noble, 1970.

Bruning, J. and Kintz, B. *Computational handbook of statistics.* Glenview, Ill.: Scott, Foresman, 1968.

Downie, N. and Heath, R. *Basic statistical methods (2nd ed.).* New York: Harper and Row, 1970.

Edwards, A. *Statistical analysis (3rd ed.).* New York: Holt, Rinehart and Winston, 1973.

Franzblau, A. *A primer of statistics for non-statisticians.* New York: Harcourt, Brace and Jovanovich, 1958.

Freund, J. and Williams, F. *Elementary business statistics.* Englewood Cliffs, N.J.: Prentice Hall, 1964.

Friedman, H. *Introduction to statistics.* New York: Random House, 1972.

Hoel, P. *Elementary statistics (3rd ed.).* New York: John Wiley and Sons, 1971.

Hoel, P. and Jessen, R. *Basic statistics for business and economics.* New York: John Wiley and Sons, 1977.

Langley, R. *Practical statistics for non-mathematical people.* New York: Drake Publishers, 1971.

Naiman, A., Rosenfeld, R. and Zirkel, G. *Understanding statistics.* New York: McGraw-Hill, 1972.

Neter, J. and Wasserman, W. *Fundamental statistics for business and economics (4th ed.).* Boston: Allyn and Bacon, 1972.

Peters, W. and Summers, G. *Statistical analysis for business decisions.* Englewood Cliffs, N.J.: Prentice Hall, 1968.

Runyon, R. and Haber, A. *Fundamentals of behavioral statistics (2nd ed.).* Reading, Mass.: Addison-Wesley, 1971.

Shurter, R., Williamson, J. and Broehl, W. *Business research and report writing.* New York: McGraw-Hill, 1965.

Spence, J., Underwood, B., Duncan C. and Cotton, J. *Elementary statistics (2nd ed.)* New York: Appleton-Century-Crofts, 1968.

Spurr, W. and Bonini, C. *Statistical analysis for business decisions (3rd ed.).* Homewood, Ill.: Irwin, 1967

Stockton, J. *Business and economic statistics (3rd ed.).* Franklin, Tenn.: Southwestern, 1966.

Yamane, T. *Statistics, an introductory analysis.* New York: Harper and Row, 1964.

Appendix A

MAXIMUM DATA SET SIZES

Data Analysis Program	Size of Memory (RAM) in TRS-80	
	4K	16K
DESCRIPTIVE STATISTICS[a] HISTOGRAM T-TEST	4ØØ elements	Over 34ØØ elements
SIMPLE CORRELATION LINEAR REGRESSION	2ØØ pairs	Over 17ØØ pairs

[a]DESCRIPTIVE STATISTICS sorts all the data values entered. The time required to sort large sets of data may be excessive. For further information see Messages and Special Considerations in the DESCRIPTIVE STATISTICS section of the manual.

Appendix B

STATISTICAL ANALYSIS FILE STRUCTURE

Data files prepared for Statistical Analysis programs (using DATA FILE PREPARATION) are structured as follows:

RECORD #1: File Type Indicator

Single type data files contain a "1", paired type files contain a "2". This code allows Statistical Analysis programs to distinguish between the two types of data.

RECORD #2: File Name

The second record of each data file contains the name of the file in the form of an alphanumeric string.

RECORDS #3 TO LAST: Data Records

Each data record contains eight numeric values separated by commas. In the last data record, one of the data positions is occupied by the end of data signal, "99999" and the remaining positions are filled with zeros.

The number of records in a data file can be calculated as follows:

For single type data

#RECORDS = 2(file type & name) + #DATA ELEMENTS ÷ 8* + 1

For paired type data

#RECORDS = 2(file type & name) + (2 X #PAIRS) ÷ 8* + 1

*whole number less remainder (24÷8=3, 26÷8=3, 23÷3=7, etc.)

All Radio Shack computer programs are distributed on an "AS IS" basis without warranty.

Radio Shack shall have no liability or responsibility to customer or any other person or entity with respect to any liability, loss or damage caused or alleged to be caused directly or indirectly by computer equipment or programs sold by Radio Shack, including but not limited to any interruption of service, loss of business or anticipatory profits or consequential damages resulting from the use or operation of such computer or computer programs.

NOTE: Good data processing procedure dictates that the user test the program, run and test sample sets of data, and run the system in parallel with the system previously in use for a period of time adequate to insure that results of operation of the computer or program are satisfactory.

Refer to User's Manual for warranties. Failure to adhere to procedures set forth in User's Manual may result in the loss of warranties.

RADIO SHACK A DIVISION OF TANDY CORPORATION

U.S.A.: FORT WORTH, TEXAS 76102
CANADA: BARRIE, ONTARIO L4M 4W5

TANDY CORPORATION

AUSTRALIA	BELGIUM	U. K.
280-316 VICTORIA ROAD	PARC INDUSTRIEL DE NANINNE	BILSTON ROAD WEDNESBURY
RYDALMERE, N.S.W. 2116	5140 NANINNE	WEST MIDLANDS WS10 7JN

878–2980078 **PRINTED IN U.S.A.**