HP-71 STATISTICS

AMPI

Statistics Library

For The

Hewlett-Packard HP-71B

American Micro Products, Inc.

December 1984

Printed in U.S.A. 82489-90001

Disclaimer

Forward

Established in 1978, American Micro Products, Inc. (AMPI) is a software house specializing in writing programs for handheld computers. Unlike other existing computer software companies, AMPI started by developing programs and products for programmable calculators, forerunners to today's handheld computers. To date, AMPI has created over 250 programs for major handheld, booksize, and briefcase computer manufacturers. AMPI is very pleased that Hewlett-Packard Company chose it to write this statistics library to be used with the highly sophisticated HP-71B Computer.

William McNeil
President
American Micro Products, Inc.

Preface

American Micro Products, Inc. (AMPI) is pleased that you have selected the AMPI Statistics Library — a computer software library especially designed for the Hewlett-Packard HP-71B Computer. This AMPI Statistics Library contains 18 programs designed to be helpful to statisticians, mathematicians, engineers, scientists, and students for statistical evaluation of data.

You do not need previous programming experience to use these 18 programs. However, previous programming experience is useful when working with the toolbox portion of the library. A basic knowledge of statistics allows you to use these programs to their best advantage. This manual provides general information concerning statistics. Each program is discussed through a descriptive introduction, remarks, mathematical formulas, program operation, one or more sample problems, and step-by-step instructions.

The AMPI Statistics Library is supplied on a 32K ROM (Read-Only Memory) module. This AMPI Statistics Library ROM module and your HP-71B provide you with the basic equipment for loading, running, and saving statistical data. Optional equipment that may be used with this Library include the HP 82401A HP-IL Interface, the HP 82400A Magnetic Card Reader, a cassette recorder, and a printer.

Contents

Part I

Basic Operations

Congratulations! As the user of a Hewlett-Packard HP-71B Computer and the AMPI Statistics Library, you now have at your command a complete general statistics system. This part of the manual contains introductory material designed to acquaint you with the library and its use with the HP-71B Computer.

Section 1

How To Use This Manual

The information contained in this manual assumes that you are familiar with your *HP-71 Owner's Manual,* especially Sections 1, 6, 8, 9, 11, 12, 13, and 14. In particular, you should be familiar with the following HP-71B operations: operating the keyboard and files, writing and running simple subprograms, manipulating data, using flags, and correcting errors. If you wish to use a printer with this library, you should know how to install and use a printer using the HP82401A HP-IL Interface.

The 18 programs in the *AMPI Statistics Library For The Hewlett-Packard HP-71B* are useful in statistical testing and evaluation of data. This AMPI Statistics Library manual has been written as a tutorial and reference guide. As a tutorial, it presents an introduction to general statistics. This manual, however, is not designed to take the place of a statistics textbook. For each of the 18 programs, there is a program introduction, mathematical formulas, remarks, program operation, comprehensive example, and step-by-step instructions showing keystrokes and displays for the example. Once you have familiarized yourself with the basic operation of each program, this manual also will serve as a reference. The contents, input/output, error messages, program and subroutine labels and register assignments, HPAF format specifications, glossary, references, and index will help you find specific information about any given example, of its practical usage, and about its operation with the HP-71B.

For your convenience, the manual is divided into nine parts.

- Part I contains basic operation information. Section 1 presents basic introductory material and explains how to use this manual. Section 2 is a brief explanation of how to start using this library with the HP-71B Computer. Section 3 is a detailed explanation of how to use this library with the HP-71B computer.

- Part II details eight statistical tests. Section 4 presents three general descriptive statistical programs. The Means and Moments Program and the Histogram Program allow the user to use either raw ungrouped or grouped data to evaluate these data. Various descriptive statistics can be calculated, including: arithmetic mean, geometric mean, harmonic mean, and the second, third, and fourth moments. If desired, the Histogram Program also provides a goodness-of-fit test to the Normal, Weibull, Exponential, Binomial, Poisson, and Uniform distributions. The Multiple Linear Regression Program allows the user to determine the relationship between a dependent variable and independent variables. Regression coefficients, correlation coefficients, and various other statistics can be calculated. Sections 5, 6, and 7 present programs which allow you to use raw data from which specific parametric statistics may be generated. The five programs described are: Paired t-Test and Unpaired t-Test (t-Statistics); One-way Analysis of Variance and Two-way Analysis of Variance (F-Statistics); and Contingency Table (Chi-square Statistics). Sections 8 and 9 present two programs which permit you to use raw data from which specific non-parametric statistics may be generated. The two programs are the Mann-Whitney U Test Program and the Kruskal-Wallis Test Program.

- Part III presents programs which permit you to evaluate the statistical data generated by the seven programs found in Part II. This virtually eliminates the need for lengthy book tables. The eight programs found in Sections 9 and 10 are: Student's t-Distribution, F-Distribution, Chi-square Distribution, Normal Distribution, Weibull Distribution, Exponential Distribution, Binomial Distribution, and Poisson Distribution.

- Part IV contains the appendices and an index. For your convenience, Appendix A summarizes the input and output for each program, Appendix B gives the program and subroutine labels and register assignments, and Appendix C lists potential error and status messages associated with the Library. A list of the names of the files used in this library is presented in Appendix E. Appendix D defines the HPAF file specifications. Appendix F presents a glossary of the statistical terms found in this manual. An up-to-date list of recommended readings, many of which are referenced in this manual, is found in Appendix G. The index lists all pertinent subject entries.

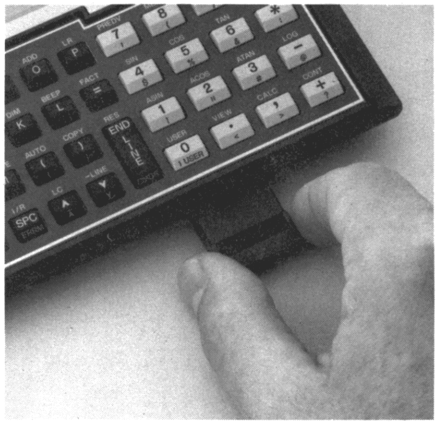
Section 2

Start-Up

The 32K ROM module containing the 18 programs in the AMPI Statistics Library can be plugged into any of the four ports on the front edge of the HP-71B computer. With the module label right-side up and the HP-71B's keyboard facing up, insert the module into one of the computer ports until it snaps in place.

PRECAUTIONS

- When a module is removed to make a port available for the AMPI Statistics Library ROM module, turn the HP-71B on by pressing **ON** and then off (by pressing **f** and then **OFF**) to reset the internal pointers **while the port is empty.**
- Before installing the module, make sure that you turn off the HP-71B



What the Statistics Pac Does

The AMPI Statistics Library Pac provides two main capabilities:

- Statistical tools that include descriptive statistics procedures and statistical inference procedures.
- These tools allow you to make statistical inferences, not statistical conclusions.

These applications are based on standard statistical procedures which are discussed in most statistical textbooks (see Appendix G for a list of recommended readings).

The AMPI Statistics Library Pac maintains current HP-71B machine status when the programs are running, except for the following changes:

- Option base zero is used.
- Option round near is used.
- Warning messages are suppressed.

These changes are reset upon termination of the programs. NOTE: The user can control various attributes of the HP-71B by using the **ATTN** key, entering a command, and using the **f** and **CONT** keys to resume the program. Commands to be entered may include: turning the printer on and off, changing from radians to degrees, or altering the delay function to control the speed and display time of the prompts.

All data and output involved with the AMPI Statistics Library Pac can be directed to the HP-71B liquid crystal display (LCD); thus, a printer or a video display is not required to use this pac. However, you most likely will find a printer or video display useful for viewing large amounts of input data, intermediate results, and final results. The programs contained in this pac will print (or display) the input and output in an easily understood format if you have an HP 82401A HP-IL Interface and a compatible printer or video display.

Using the AMPI Statistics Library

When you run the AMPI Statistics Library, there are 11 primary steps to go through to enter, edit, save, load, print, describe, and analyze your statistical data (using keyboard entry).

1. Plug the AMPI Statistics Library ROM into the HP-71B.
2. Type RUN AMPISTAT and then press **END LINE**.
3. Obtain automatically the main menu which allows you to select Data, Edit, Menu, and Quit options.
4. Select Data from the main menu and automatically obtain the data menu, which allows you to select Keyboard, Load, Save, and Print options.
5. Select Keyboard from the data menu and respond to optional special prompts.
6. Enter the data points and optionally save them in a file and/or print them on a printer (if one is attached).
7. Edit the data array, if necessary.
8. Specify the name of the AMPI Statistics Library program that you desire to use from the statistics menu.
9. Respond to optional control prompts.
10. Perform the statistical computations.
11. Exit the program.

Flow Chart of AMPI Statistics Library Menus

The above 11 primary steps are controlled by a series of menus. The flowchart shown in Figure 2-1 presents the primary steps involved in, and the menus associated with, the AMPI Statistics Library procedure.

Example: Student's t-Distribution

Without worrying too much about the meaning of all the keystrokes (they will be described in Sections 3 and 9), key in this first example and see how easily the AMPI Statistics Library handles statistical distributions. For this Student's t-Distribution Program example, you will do the following:

1. Plug the AMPI Statistics Library ROM into the HP-71B.
2. Type RUN AMPISTAT and then press **END LINE**.
3. Obtain automatically the main menu which allows you to select Data, Edit, Menu, and Quit options.
4. Select the Menu option from the main menu.
5. Select the Student's t-Distribution Program from the statistics menu.
6. Respond to control prompts.
7. Compute the results.
8. Exit the program.
9. Evaluate the results.

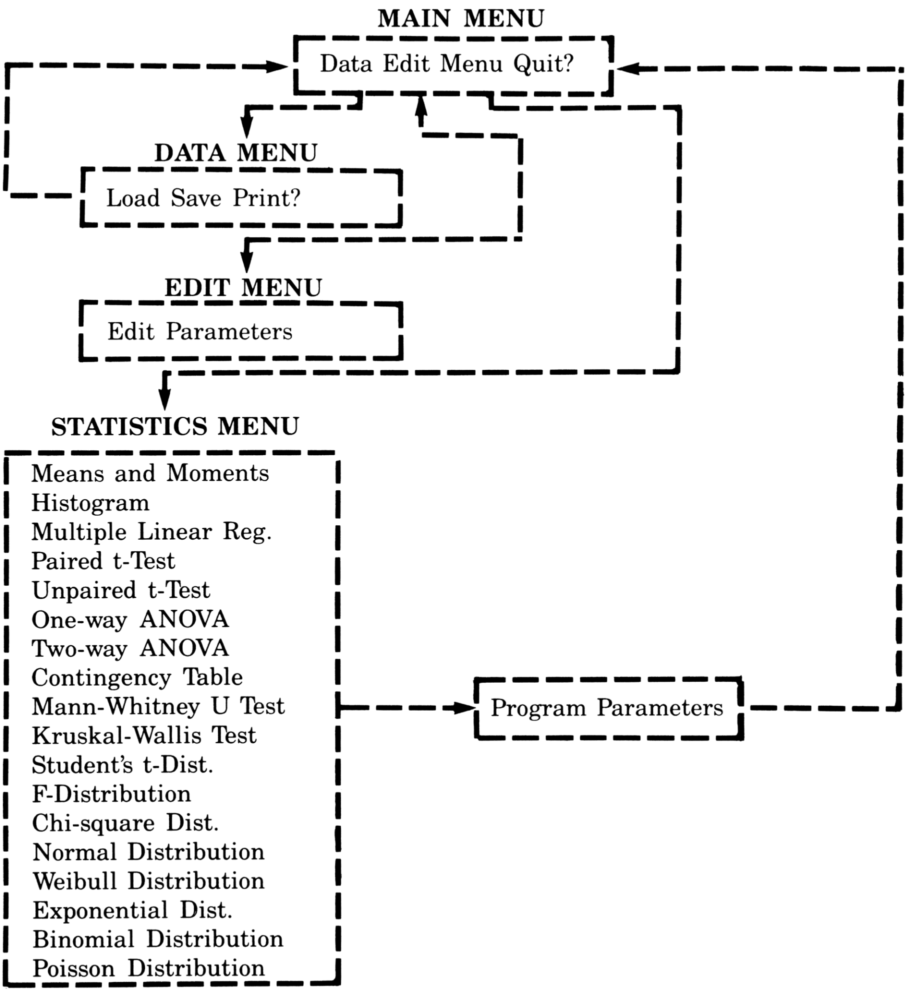


Figure 2-1. Flowchart of the AMPISTAT Statistics Pac procedure.

The example is as follows. Given a t-statistic value equal to 2 and degrees of freedom equal to 5, find the left tail area, P(t), and the right tail area of the distribution, Q(t), of these data. Input the data from the keyboard and view the results via the display.

Start with the AMPI Statistics Library ROM in one of the ports. The following input, responses, and results will be necessary to complete this example.

Keystrokes and Comments	Display and Comments
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
Type RUN and the name of the Library; then, press **END LINE**.	This is the main menu.
M	`Means and Moments`
Press **M** to start the statistics menu.	This is the first prompt of the scrolling statistics menu.
▼	`Student's t-Dist.`
Press ▼ several times to locate the Student's t-Distribution Program.	This is the Student's t-Distribution Program title.
END LINE	`Working...`
Press **END LINE** to select the Student's t-Distribution Program.	This is a status message which indicates that the HP-71B is calculating (no action is necessary).
	`t=0`
	This is a data input prompt which requires you to enter the value of the t-statistic; NOTE: the initial value for the t-statistic is set to zero.
2 **END LINE**	`df=0`
Enter the correct value for the t-statistic; then, press **END LINE**.	This is a data input prompt which requires you to enter the value of the degrees of freedom; NOTE: again, the initial value for the degrees of freedom is set to zero.
5 **END LINE**	`Edit Parameters?`
Enter the correct value for the degrees of freedom; then, press **END LINE**.	This is an edit prompt: to edit the input data, press **Y**; to continue the program without editing, press **N**.
N	`Working...`

Keystrokes and Comments	Display and Comments
Because we choose to continue without editing, press **N**.	This is a status message which indicates that the HP-71B is calculating the final results (no action is necessary).
	`P(t)=.949030260585`
	Assuming you do not have a printer or video display attached, the first of the result prompts is shown on the LCD; this prompt presents the left tail area. (If you do have a printer or video display attached, the results are output as shown in Part III, Figure 9-2.)
END LINE	`Q(t)=.050969739415`
Press **END LINE** or any key to continue	This result prompt gives the right tail area of the t-Distribution.
END LINE	`Repeat Results?`
Press **END LINE** to continue.	This control prompt allows you to repeat your results without going back through the entire program.
N	`Data Edit Menu Quit?`
Because we de not desire to repeat the results of this example, press **N**.	This again is the main menu.
Q	`Done`
Because we do not desire to continue the program, press **Q** to exit the program.	This is the final prompt and alerts you that the program is complete.
END LINE	`>_`
Press **END LINE** to exit the library.	This denotes that you are in BASIC
f OFF	
Press **f** and then **OFF** to turn the HP-71B off.	

In evaluating the results of this Student's t-Distribution example, we note that the left tail area, P(t), is equal to 0.949030260585 and the right tail area of the t-Distribution is equal to 0.050969739415.

Example: Means and Moments

Again, without worrying too much about the meaning of all the keystrokes (they will be fully described in Sections 3 and 4), key in this second example and see how easily AMPI Statistics Library handles descriptive statistics. For this Means and Moments Program example, you will do the following:

how easily AMPI Statistics Library handles descriptive statistics. For this Means and Moments Program example, you will do the following:

1. Plug the AMPI Statistics Library ROM into the HP-71B.
2. Type RUN AMPISTAT and then press **END LINE**.
3. Obtain automatically the main menu which allows you to select Data, Edit, Menu, and Quit options.
4. Select the Data option from the main menu.
5. Select the Keyboard option from the data menu and respond to optional special prompts.
6. Enter a set of data into a working array.
7. Save the data in a file.
8. Obtain automatically the main menu which allows you to select Data, Edit, Menu, and Quit options.
9. Select the Means and Moments Program from the statistics menu.
10. Respond to control prompts.
11. Compute the results.
12. Exit the program.
13. Evaluate the results.

The example is as follows. A packager wants a basic descriptive statistical analysis of the weights of 12 boxes of cereal taken at random from his production line over a 12 hour period. The weights, in grams, are as follows: 20.24, 19.36, 19.89, 20.05, 20.45, 20.83, 19.56, 20.15, 19.99, 20.38, 21.01, and 20.75. Find the means and moments of these data. Input the data via the keyboard, save the data, and view the results on the LCD.

Start with the AMPI Statistics Library ROM in one of the ports. The following input, responses, and results will be necessary to complete this example.

Keystrokes and Comments	Display and Comments
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
Type RUN and the name of the Library; then, press **END LINE**.	This is the main menu.
D	`Kbd Load Save Print?`

Keystrokes and Comments	Display and Comments
You need to enter data from the keyboard, so press **D** to access the data menu.	This is the data menu.
K	`Grouped Ungrouped?`
Press **K** to select entry via the keyboard.	This is the data format menu.
U	`Working...`
Because this example contains ungrouped data, press **U** to select ungrouped data entry.	This status message indicates that the HP-71B is calculating (no action is necessary).
	`No. of Var.=0`
	This prompt asks for you to enter the number of variables in your sample.
1 **END LINE**	`No. of Obs.=0`
This example has only one variable; thus, enter 1 as the value and then press **END LINE**.	This prompt asks for you to enter the number of observations contained in your sample.
12 **END LINE**	`Edit Parameters?`
This example has 12 observations; thus, enter 12 as the value and press **END LINE**.	This is an edit prompt: to edit the parameters, press **Y**; to continue the program without editing, press **N**.
N	`Edit Labels?`
Because we choose to continue without editing, press **N**.	This asks if you desire to add labels (i.e., names) to the variables. To add labels, press **Y**; to continue without adding labels, press **N**.
N	`Working...`
Because we do not want to add a label to the variable, press **N**.	This status message indicates that the HP-71B is calculating (no action is necessary).
	`X1(1)=0`
	The program branches to the editor and displays the first element of the array; all array values initially default to zero.
20.24 **END LINE**	`X1(2)=0`

Keystrokes and Comments	Display and Comments
Enter the correct value of the first array element; then, press **END LINE**.	The editor continues and displays the second element of the array; the value in this element also initially defaults to zero.
19.36 **END LINE** . . .	`X1(1)=20.24`
Enter the correct value of the second element; then, press **END LINE**. The program continues to step through each of the 12 elements of the entire array. Continue to enter the values from the example as the program calls for them; press **END LINE** after each entry.	The program goes through the entire array and wraps back to the first element.
Q	`Working...`
Press **Q** to exit the editor and return to the main menu.	This is a status message which indicates that the HP-71B is calculating (no action is necessary).
	`Data Edit Menu Quit?`
	This is the main menu.
D	`Kbd Load Save Print?`
You need to save the data array, so press **D** to select the data menu.	This is the data menu.
S	`SAVE:Filename?_`
Press **S** to start the process for saving data to a file.	The program prompts for the file name to save the data.
TEST **END LINE**	`Saving...`
Save the data into a file called TEST by typing TEST and then pressing **END LINE**.	The program displays this status message while the file is created and the data saved.
	`Data Edit Menu Quit?`
	The main menu automatically reappears.
M	`Means and Moments`
Press **M** to start the statistics menu.	This is the first prompt of the scrolling statistics menu.
END LINE	`Working...`

Keystrokes and Comments	Display and Comments
Because our example calls for the Means and Moments Program and the first display prompt of the scrolling statistics menu shows Means and Moments, press **END LINE** to select this program.	This is a status message which indicates that the HP-71B is calculating (no action is necessary).
	`Use n-1 for df?`
	This is a control prompt which asks if you want to use degrees of freedom equal to n-1 (i.e., if you are analyzing a sample drawn from a larger population) or degrees of freedom equal to n (i.e., if the data represent the entire population).
Y	`Edit Parameters?`
Because the number of observations represent a sample drawn from a larger population, use degrees of freedom equal to n-1. Thus, press **Y** for yes.	This is an edit prompt: to edit the parameter, press **Y**; to continue the program without editing,press **N**.
N	`Working...`
Because you desire to continue without editing, press **N**.	This is a status message which indicates that the HP-71B is calculating (no action is necessary).
	`Var.#=0`
	This is a special prompt which asks you to enter the number of the variable for which you desire to test.
1 **END LINE**	`Edit Parameters?`
Enter a value of 1; then, press **END LINE**.	This is an edit prompt: to edit the parameter, press **Y**; to continue the program without editing,press **N**.
N	`Working...`
Because you desire to continue without editing, press **N**.	This is a status message which indicates that the HP-71B is calculating the final results (no action is necessary).
	`#Data=12`
	Assuming you do not have a printer or video display attached, the first of the result prompts is shown on the LCD; this prompt gives the total number of observations in the sample. (If you do have a printer or video

Keystrokes and Comments	Display and Comments
	display attached, the results are output as shown in the example in Part II, Section 4, Figure 4-3.)
END LINE	`A Mean=20.2216666667`
Press **END LINE** or any key to continue.	This result prompt gives the arithmetic mean for the sample.
END LINE	`G Mean=20.2160383433`
Press **END LINE** or any key to continue.	This result prompt presents the geometric mean for the sample.
END LINE	`H Mean=20.2104006403`
Press **END LINE** or any key to continue.	This result prompt shows the harmonic mean for the sample.
END LINE	`Moment 2=.247887878783`
Press **END LINE** or any key to continue.	This result prompt gives the second moment for the sample.
END LINE	`Moment 3=-8.78149076342E-3`
Press **END LINE** or any key to continue.	This result prompt presents the third moment for the sample. NOTE: You have to use the ▶ and ◀ keys to scroll (i.e., from right to left and vice versa) the prompt.
END LINE	`Moment 4=.113603949422`
Press **END LINE** or any key to continue.	This result prompt presents the fourth moment for the sample.
END LINE	`Kurtosis=1.84876983925`
Press **END LINE** or any key to continue.	This result prompt shows the kurtosis for the sample.
END LINE	`Skewness=-7.11517051005E-2`
Press **END LINE** or any key to continue.	This result prompt gives the skewness for the sample. NOTE: Again, you have to use the ▶ and ◀ keys to scroll the prompt.
END LINE	`Repeat Results?`
Press **END LINE** to continue.	This control prompt allows you to repeat your results without going back through the entire program.

Keystrokes and Comments	Display and Comments
N	`Data Edit Menu Quit?`
Because we do not desire to repeat the results, press **N**.	This is the main menu.
Q	`Done`
Because we do not want to continue the program, press **Q** to exit the program. **END LINE**	This is the final prompt and alerts you that the program is complete.
	`>_`
Press **END LINE** to exit the library.	This denotes that you are in BASIC.
f OFF	
Press **f** and then **OFF** to turn the HP-71B off.	

In evaluating the results of this example, we note that the means and moments descriptive statistics give us the following information. There were 12 observations in the sample. The arithmetic mean or average, is 20.2216666667 grams; because our input data were measured in hundredths, this should be rounded to 20.22 grams. The geometric mean and harmonic mean, are 20.2160383433 and 20.2104006403 grams, respectively; rounded off, they are 20.22 and 20.21 grams. As can be seen, the geometric mean is less than (or equal to) the arithmetic mean, but is greater than the harmonic mean. The second moment, or variance of the sample, is 0.247887878788. The third and fourth moments are −8.78149076342E−3 and 0.113603949422, respectively. The skewness coefficient is −7.11517051005E−2, while the kurtosis coefficient is 1.84876983925. The negative third moment and negative skewness indicate that the frequency curve of the distribution is skewed to the left and, thus, has a longer "tail" to the left of the central maximum than to the right. Because kurtosis is less than three, the distribution is said to be platykurtic (i.e., a flat-topped distribution).

Section 3

Basics About the HP-71B and AMPI Statistics Library

The HP-71B contains a set number of kilobytes of Read-Only Memory (ROM). This internal ROM contains the operating system and all of the functions of the computer. You cannot write information to this memory; however, you can increase the capabilities of the HP-71B by adding ROM modules.

The HP-71B also contains a set number of kilobytes of Random-Access Memory (RAM), which (except for a small amount for operation of the HP-71B) is all available to you. To increase the amount of RAM, you can add up to four RAM modules to your HP-71B.

Inserting And Removing Application Devices

Application devices which may be attached to and used with the HP-71B in conjunction with the AMPI Statistics Library include ROM application modules, RAM memory modules, and the magnetic card reader. The HP-71B contains four external ports, an HP-IL port, and a Card Reader port. While the Card Reader port is designed especially to accept only the HP 82400A Magnetic Card Reader and the HP-IL port is specially designed for the HP 82401A HP-IL Interface, the application modules (ROM) and memory modules (RAM) may be plugged into any of the four external ports in any order. The placement of these ports on the HP-71B are illustrated in your *HP-71 Owner's Manual.*

Read-Only Memory (ROM) Modules

The 32K ROM application module provided with the AMPI Statistics Library enhances the HP-71B features by allowing you easy library insertion and direct keyboard access to the 18 programs without going through a separate loading sequence for each program. The Statistical ROM module can be plugged into any of the four ports on the front of the HP-71B.

PRECAUTIONS

- Make sure you turn off the HP-71B prior to installing any module by pressing **f** and then **OFF**.
- When a module is removed to make a port available for another module, press **ON**; then, press **f** followed by **OFF** to reset the internal pointers while the port is empty before installing the new module.
- For additional precautions, see your *HP-71 Owner's manual.*

To insert the AMPI Statistics Library ROM module, orient it so that the label is right-side up, hold the HP-71B with the keyboard facing up, and push in the module until it snaps into place. During this operation, be sure to observe the precautions noted above.

To remove the AMPI Statistics Library ROM module, use your fingernails to grasp the lip on the bottom of the front edge of the module and pull the module straight out of the port. Install a blank module in the port to protect the contacts inside the HP-71B port.

Random-Access Memory (RAM) Modules

The HP-71B Random-Access Memory (RAM) exists in two forms: main RAM and independent RAM. The initial amount of kilobytes in the RAM, and any RAM added to ports 1 through 4, is called main RAM; it is used by the HP-71B for operations, keeping files, and storing variables. Independent RAM is memory that is internally set aside from main RAM. You can use independent RAM like ROM; that is, you can use it for storage of data files or program files. The AMPI Statistics Library does not use independent RAM unless you specify such a device for storage. Main RAM within a port can be set aside as independent RAM; likewise, independent RAM within a port can be reclaimed back into main RAM.

Information concerning FREEing and CLAIMing RAM is found in Section 6 of your *HP-71 Owner's Manual*. Further instructions for installing, using, and removing RAM modules are included with each of these modules. Also, note the precautions given above for ROM modules.

Magnetic Card Reader

The optional HP 82400A Magnetic Card Reader is used for storing and retrieving data files to and from the memory of your HP-71B. For information on how to install and remove the card reader, to store a data file from memory (RAM or ROM) to a card in the card reader, and to retrieve a data file from a card in the card reader to the memory, please refer to Appendix C, "Using the HP 82400A Magnetic Card Reader" in your *HP-71 Owner's Manual*.

Running the AMPI Statistics Library

The AMPI Statistics Library start-up and initialization begins when you insert the AMPI Statistic Library ROM module into one of the four ports and then type RUN AMPISTAT and press the **END LINE** key. The display shows the main menu:

```
Data Edit Menu Quit?
```

You can access the entire AMPI Statistics Library through this main menu. The main menu is divided into four major options or parts — Data, Edit, Menu, and Quit.

The Data part of the AMPI Statistics Library contains commands for entering data points from the keyboard or from a file, saving entered data to a file, and printing the input data array. The Edit part contains an editor for examining or modifying the data points. The Menu part contains the statistical menu, which is your key to the 18 programs in the AMPI Statistics Library. Each of these three parts is called from, and returns to, the main menu. The last part of the main menu — Quit — allows you to quit or exit the AMPI Statistics Library; you will return to BASIC. The paragraphs that follow describe each of these four parts.

To assist the user, a flow chart of the steps involved in, and the menus associated with, the AMPI Statistics Library procedure is presented earlier (see Table 2-1).

Working With Data

From the main menu:

```
Data Edit Menu Quit?
```

select Data by pressing **D** (i.e, for Data). This displays the data menu:

```
Kbd Load Save Print?
```

All of the data menu options are selected from this menu. The data menu allows you to enter data points from either the keyboard or a pre-existing data file. The data menu allows you to save the data points in a file, and to print the data points.

> **Note:** If you desire not to use any of the data menu options, you can press **Q** (i.e., for Quit) to return to the main menu.

The Data Format

The data matrix used in this library basically consists of n variables and m_i observations for variable i.

Data may be entered in grouped or ungrouped format. For example, the following array represents an ungrouped data set with n variables of m_i observations each.

Variables

Observations				
	X1(1)	X2(1)	. . .	Xn(1)
	X1(2)	X2(2)	. . .	Xn(2)
	.	.	. . .	.
	.	.	. . .	.
	.	.	. . .	.
	X1(m_1)	X2(m_2)	. . .	Xn(m_n)

The following array represents a grouped data set with n variables of m_i observations each. Xi(Fj) is the associated frequency for variable i (where i = 1,...n) and observation j (where j = 1,...mi).

Variables

O	X1(1)	X2(1)	. . .	Xn(1)
b	X1(F1)	X2(F1)	. . .	Xn(F1)
s				
e	X1(2)	X2(2)	. . .	Xn(2)
r	X1(F2)	X2(F2)	. . .	Xn(F2)
v	.	.	. . .	.
a	.	.	. . .	.
t	.	.	. . .	.
i				
o	X1(m_1)	X2(m_2)	. . .	Xn(m_n)
s	X1(Fm_1)	X2(Fm_2)	. . .	Xn(Fm_n)

Setup using this basic format will be explained in each program description.

Entering Data From the Keyboard (Kbd)

Data entry for statistical programs presents a unique set of challenges because the rows (here denoted as observations) and columns (here represented as variables) of data for various samples are not always symmetrical. As shown above, basic sample data may be described as ungrouped (i.e., X1(1), X1(2), ... , X1(m_1) or grouped (i.e., X1(1) X1(F1), ... , X1(m_1) X1(Fm_1).

As shown in Table 3-1, the various programs included in this library have been classified as to the type of input data necessary for each program. Most of the programs require the user to set subscript limit(s) for the number of data input points to be entered; this is preset or fixed at the beginning of these programs by the user.

If you want to enter new data from the keyboard, follow the instructions presented in Table 3-2. The array has now been created with all elements initialized and set to zero (0), and (if for grouped data entry) with the frequency set to one (1). When you use the keyboard to enter the values of the elements, you simply edit their initial settings by going directly to the edit option of the main menu.

The procedure for editing an element is presented in the "Editing the Data" section.

Table 3-1. Keyboard input data guide to the AMPI Statistics Library programs for the HP-71B

Program Name	Sample Type	Requirements
Means and Moments	Ungrouped Grouped	One variable; minimum of 3 observations One variable; minimum of 3 observations
Histogram	Ungrouped Grouped	One variable One variable
Multiple Linear Regression	Ungrouped	Two or more variables. The last (nth) variable is taken to be the dependent variable; the first (n − 1) is taken to be the independent variable(s). All variables must have an equal number of observations, with a minimum of n
Paired t-Test	Ungrouped	One variable which equals the difference of the two variables under analysis
Unpaired t-Test	Ungrouped Grouped	Two variables Two variables
One-way Analysis of Variance	Ungrouped Grouped	Two or more variables Two or more variables
Two-way Analysis of Variance		
W/O Replication	Ungrouped	Four or more variables (i.e., rows x columns) with one observation (i.e., replicate)
With Replication	Ungrouped	Four or more variables (i.e., rows x columns) with equal number of observations (i.e., replicates) for each variable
Contingency Table	Ungrouped	Two or more variables with equal number of observations
Mann-Whitney Test	Ungrouped	Two variables
Kruskal-Wallis Test	Ungrouped	Two or more variables

Table 3-2. Steps for entering data via the keyboard

Step	Display	Instructions
1	Data Edit Menu Quit?	Press **D** to select Data menu
2	Kbd Load Save Print?	Press **K** to select the Keyboard option and to start the process of creating a data set from the keyboard.
3	Grouped Ungrouped?	This is the data format menu. If your data are grouped, press **G**; if your data are ungrouped, press **U**.
4	No. of Var.=0	Enter the number of variables; then, press **END LINE**.
5	No. of Obs.=0	Enter the number of observations in your sample or, if the variables have unequal numbers of observations, the maximum of these; then, press **END LINE**.
6	Edit Parameters?	To edit the variable and observation parameters, press **Y**; go to Step 4. To accept variable and observation parameters, press **N**; go to Step 7.
7	Edit Labels?	To edit variable labels, press **Y**; go to Step 8. To accept defaults (X1 to Xn, where n is the number of variables), press **N**; go to Step 10. (NOTE: Labels MUST contain from 1 to 10 characters; blanks are automatically removed from the beginning and ending of the label.)
8	X1=X1	To add labels to variables, type name of each variable (for X1 to Xn) and then press **END LINE** for each; to accept default, press **END LINE** for each.
9	Edit Labels?	To edit labels, press **Y**; go to Step 8. To accept labels, press **N**; go to Step 10.
10	X1(1)=0	The program branches to the editor (from the main menu) and displays the first element in the data array. (NOTE: If labels were added to variables, label names would replace X1 to Xn.) For ungrouped data, go to Step 12; for grouped data, go to Step 11.

Step	Display	Instructions
11	X1(F1)=0	This is the second element in the data array for the grouped variable. Enter the frequency for X1(1). Steps 10 and 11 are repeated until all elements are entered.
12	X1(2)=0	This is the second element in the data array for the ungrouped variable. Enter the value for this element; then, press **END LINE**. Step 12 is repeated until all elements are entered.

Loading Data From a File (Load)

If the data already exists in a file, you can load the data into an array by following the procedure outlined below.

Step	Display	Instructions
1	Data Edit Menu Quit?	Press **D** to select Data menu
2	Kbd Load Save Print?	Press **L** to select the Load option and to start the process of loading a data set.
3	Clear Data?	If data already exist in the array, the Load option can append new data items to the existing data set. To add new data to the array, press **N**. To clear out the existing array, press **Y**. This prompt will not appear if the array contains no data when the Load option is selected.
4	LOAD:Filename?_	Enter the name of the data file in which the data are stored; then, press **END LINE**. The file must be formatted as an HPAF file. Also, if a mass storage device is used, the filename must include the device specification (e.g., TESTDATA:TAPE). For detailed information on the HPAF format, refer to Appendix D; data files also are described in your owner's manual.
5	Loading... Working...	These messages appear while the AMPI Statistics Library is loading the file; you then return to the main menu.

Saving Data To a File (Save)

After data are entered, it is good practice to save these data to a file for future use. Instructions for saving data are as follows:

Step	Display	Instructions
1	`Data Edit Menu Quit?`	Press **D** to select Data menu.
2	`Kbd Load Save Print?`	Press **S** to select the Save option and to start the process of saving a data set in an HPAF file.
3	`Save as Ungrouped?`	This prompt appears only if the original data were input as grouped data. To save input data as grouped (e.g., to use the file with the AMPI Statistics Library), press **Y**; to save data as ungrouped (e.g., to use the file with the Curve Fit Pac), press **N**.
4	`SAVE:Filename?_`	Enter the name of the file on which to save the data; then, press **END LINE**. The file will be formatted to HPAF file specifications and may reside either in RAM or on a mass storage device. If the latter is used, the file name must include the device specifier (e.g., TESTDATA:TAPE).
5	`Overwrite File?`	If a file already exists with the file name you supplied, this prompt appears which asks if you desire to overwrite the data already on the file. To retain the original file and not overwrite the data, press **N** (the program will return to the main menu). To overwrite the previous data on the file, press **Y**.
6	`Saving...`	This message appears while the AMPI Statistics Library is saving the data; you then return to the main menu.

Printing the Data (Print)

Once data are entered, you can obtain a printed copy of the data array if you have a printer attached.

Step	Display	Instructions
1	`Data Edit Menu Quit?`	Press **D** to select data menu.
2	`Kbd Load Save Print?`	Press **P** to select the printer option and to start the process of printing the data array. AMPI Statistics Library will send the data array to the current PRINTER IS device.

Step	Display	Instructions
3	`Working...` `Printing...`	These messages appear while printing is in progress. AMPI Statistics Library then returns you to the main menu.

Editing the Data

The Edit option allows you to enter and modify items within the data array. From the main menu:

```
Data Edit Menu Quit?
```

the Edit option is selected by pressing **E**. The next display shows:

```
Edit Labels?
```

which requires you to respond by (1) pressing the **Y** key for editing the variable labels, or (2) pressing the **N** key to continue the program without editing the variable labels.

The process of editing labels was presented in Table 3-2 and will not be repeated here. The format for editing data is similar to that of entering data (which also is presented in Table 3-2); because of its importance to all the programs, further details are discussed here.

The next display contains the value of the first element in the data array. The data array format for the AMPI Statistics Library is:

Ungrouped: `X1(1)=############`

Grouped: `X1(1)=############`
`X1(F1)=###########`

which consists of three items:

- A letter giving the type of element displayed (i.e., X for an independent or dependent variable, and F for the frequency of a variable).
- A number following the X giving the variable (i.e., column) label.
- A number in parenthesis giving the observation (i.e., row) number.

The current value of the element is presented after the equal sign. (NOTE: If data are being entered from keyboard, each element in the array is initially set to zero (0) and must be edited to reflect the correct current value; also, each frequency is initially set to one (1) for grouped data. Inf and NaN are not acceptable entries; also, frequency values must be integers and greater than or equal to one (1).)

Editing an Element

All of the HP-71B line editing features (i.e., ▸, ◂, **I/R**, **BACK**, and **-CHAR**) are available in the data array editor. For additional information of the use of these features, please refer to Section 1, "Getting Started," in the *HP-71 Owner's Manual.*

When an element is visible on the display, it may be edited and the new value entered into the array. **The element is edited by placing the cursor over the current value and typing the new value over the current value; it is entered by pressing END LINE**. Also, the **ATTN** key may be used to clear the input value; then, the value may be re-entered.

When editing an element, you can use both numbers and numeric expressions for the value. For example, the numeric expression (1+SQR(25))/3 is just as acceptable as 2 for an entered value.

Moving Around the Data Array

When in the Edit option, a number of keys have been redefined to assist you in moving about within the data array and to help you to insert or delete rows and columns (and, where applicable, replicates). These keys are divided into the following groups:

- The direction keys for moving through the data array (i.e., **W**, **D**, **X**, and **A**).

- The maximum direction keys for moving anywhere in the data array (i.e., the gold prefix key **f** in combination with **W**, **D**, **X**, and **A**).

- The command keys for manipulating variables and observations (i.e., **U**, **O**, **M**, and **F**).

- The endline direction key (i.e., **S**) in combination with the direction keys (i.e., **D**, **X**, and **S**).

- The quit key (i.e., **Q**).

While in the edit option, the blue prefix key **g** in combination with a letter allows you to enter a function (or letter) which includes a redefined key (e.g., the letter **S** in SIN, the letter **F** in INF, and the letter **Q** in SQR).

The following keyboard illustration (Figure 3-3) shows the keys that are redefined when you are in the Edit option of the main memory.

The Direction Keys

The direction keys — **W**, **D**, **X**, and **A** — are found in the cross on the keyboard diagram (Figure 3-3). These direction keys are used in combination with the **f** key to move anywhere in the data array. When you get to the row and column (and replicate when applicable) desired, you can view the contents of the element through this "window" at that location and/or modify it.

Q Quit	**W** Up	**E**	**R**	**T**	**Y**	**U** Define	**I**	**O** Add	**P**
A Left	**S** Endline Direction	**D** Right	**F** Go To	**G**	**H**	**J**	**K**	**L**	**=**
Z	**X** Down	**C**	**V**	**B**	**N**	**M** Delete	**(**	**)**	**END LINE**
ON Clear	**f** Gold Prefix	**g** Blue Prefix	**RUN**	▶	◀	**SPC**	▲	▼	

Figure 3-3. Redefined HP-71B keyboard keys for editing data arrays.

Using the **W**, **D**, **X**, and **A** keys allow you to move, respectively, up, right, down, and left. As an example, use the following 5 by 5 data array.

X1(1)	X2(1)	X3(1)	X4(1)	X5(1)
X1(2)	X2(2)	X3(2)	X4(2)	X5(2)
X1(3)	X2(3)	X3(3)	X4(3)	X5(3)
X1(4)	X2(4)	X3(4)	X4(4)	X5(4)
X1(5)	X2(5)	X3(5)	X4(5)	X5(5)

As shown in the following table, the direction keys, used alone, move you through a data array one element at a time.

Starting Element	Key	Direction	Destination
X3(3)	**W**	Up	X3(2)
X3(3)	**D**	Right	X4(3)
X3(3)	**X**	Down	X3(4)
X3(3)	**A**	Left	X2(3)

You can move to one of the boundaries of the data array by pressing the **f** key in combination with a direction key; think of the **f** as denoting the word “far.”

- The **f D** moves to the far right boundary element of the data array. (NOTE: If an observation does not exist in the far right boundary element, **f D** moves to the first observation in the far right boundary variable of the data array.)

- The **f A** moves to the far left boundary element of the data array. (NOTE: If an observation does not exist in the far left boundary element, **f A** moves to the first observation in the far left boundary variable of the data array.)

- The **f W** moves to the far up boundary element of the data array.

- The **f X** moves to the far down boundary element of the data array.

Using the same 5 by 5 data array used above, this movement is shown in the following table.

Starting Element	Key	Direction	Destination
X3(3)	**f W**	Far Up	X3(1)
X3(3)	**f D**	Far Right	X5(3)
X3(3)	**f X**	Far Down	X3(5)
X3(3)	**f A**	Far Left	X1(3)

All the direction keys move across the boundaries of the data array. In other words, they allow you to "wrap around" to the opposite of the data array. For example in the above 5 by 5 data array, if you start at the X1(1) element and press **f A**, you will move to the X5(1) element (i.e., you move to the far-right boundary element of the data array).

> **Note:** If you are in the Edit option and desire to enter an expression using letters that have been re-defined by the AMPI Statistics Library, you must press down the **g** key first before typing the letters.

The Command Keys

The command keys — **O**, **U**, **M**, and **F** — can be associated (see Figure 3-3) with their gold, shifted functions on the HP-71B keyboard. The command keys and their functions are:

- The **O** key selects the ADD command.

- The **U** key selects the DEFine command.

- The **M** key selects the DELETE command.

- The **F** key selects the GOTO command.

The ADD command — the **O** key — is used to easily add an observation or variable to the data array. When adding a new variable, the ADD command must be used before using the DEFine command. To use this command:

1. Press the **O** key, which causes the display to show:

```
ADD:Var. or Obs.?
```

2. Do one of the following:

- To add a variable to the data array, press **V**; the display first will show:

```
Add Var. at?
```

To add a variable at the end of the array, enter one more than the total number of variables as the address for the new variable in the data array. This is automatically assigned and is displayed as the default value; to accept this default value, press **END LINE**.

To add a variable at any position of the array, enter the desired position and press **END LINE**. The remaining variables will increment by one, and retain their current label assignments. If you are unsure of the assignments, use the Edit Labels sequence (for viewing via the display) or the Print function (for viewing via the printer).

The next prompt will display:

```
Label=NEW
```

which requires you to enter the name of the new variable. Then, press **END LINE**. NOTE: If no name is entered, the default name is NEW.

- To add an observation for a single variable to the data array, press **O**; the display will first show:

```
Var. Number?
```

which requires that you enter the variable number to which you are adding the new observation; then, press **END LINE**. The next prompt shows:

```
Add Obs. at?
```

To add an observation at the end of the existing observations, enter one more than the total number of observations as the address for the new observation in the data array. This is automatically assigned and is displayed as the default value; to accept this default value, press **END LINE**.

To add an observation anywhere among the existing observations, enter the desired address and press **END LINE**. The remaining observation numbers will be increased by one; the variables will retain their current label assignments. If a new observation is added to grouped data, the frequency for the new value will default to one.

- To exit the ADD command, press the **Q** key; the program will return to the regular edit option.

3. In summary, should the address coincide with an existing variable or observation, the data array will expand (or open) to create a space for the new variable or observation. The new variable or observation will be filled with default values equal to zero upon initialization; frequency values will default to one.

If you add a new variable or observation at the address of an existing variable or observation, the existing variable or observation address (and all those variables or observations beyond it) will increase by one. As an example, if you enter a new variable at variable 4, old variable 4 becomes variable 5, old variable 5 becomes variable 6, and so on.

The DEFine command — **U** key — is used to assign values to a variable's observations (i.e., columns). If a variable does not exist, you must first create the variable by using the ADD command. To use the DEFine command:

1. Press the **U** key, at which time the display prompt first shows:

```
DEF:Sequence Compute?
```

2. Press **S** to automatically assign values to all observations or to sort a variable; press **C** to assign values to a variable using an equation or other variables.

3. If **S** is selected for sequence, the following prompt appears:

```
DEF:Sort Auto?
```

If **S** again is selected (this time, denoting sort), the display shows:

```
Var. Number?
```

which requires that you enter the number of the variable which you desire to sort; press **END LINE**. You will return to regular editing. Note that both ungrouped and grouped data may be sorted. The observations of the variable selected will be sorted in increasing order; other variables will not be affected.

If **A** is selected, the display shows:

```
Var.,Start,Step?
```

which requires entering the following generalized format:

<var number>,**<start value>**,**<step size value>**

where:

<var number> = the number of variable to sequence
<start value> = the starting value
<step size value> = the value of each step size

Then, press **END LINE**. (NOTE: The values for start and step must be entered as a number and not as Inf or NaN.)

This can be very useful when you desire to enter data that have a constant interval between points. As an example, sequence variable 1 beginning at 0 degrees Centigrade and at a constant of 10 degree steps. This would be entered as:

1,0,10 **END LINE**

If in this example all values are to equal a constant of 10 degrees, the entries would be:

1,10,0 **END LINE**

After the information has been entered, the program returns to regular editing.

4. If **C** is selected for compute, the following prompt appears:

`Var. Number?`

which requires you to enter the number of the variable to which you desire to assign values; press **END LINE**. The next prompt shows:

`Compute X##=X`

where the two pound symbols (##) denote the variable number that you entered for the previous prompt; the default assignment leaves the variable unchanged. To have the program compute new values, enter any function (e.g., *, /, −, and +) to represent the meaning of the variable, and press **END LINE**. (NOTE: Do not use Inf or NaN in equations; also, results should be checked for invalid computations resulting in Inf or NaN. Failure to correct such entries prior to exiting the data editor will result in loss of data.)

Examples of compute functions for variable X1 and their meanings are:

Display/Function	**Meaning**
`Compute X1=X1`	Default expression provided by program; leaves variable unchanged
`Compute X1=X1-100`	Scale by subtracting 100
`Compute X1=100`	Set X1 to a constant 100
`Compute X1=1/2*X1`	Scale by multiplication
`Compute X1=X1^2`	Scale by a power
`Compute X1=(X2+X3)*X4`	Define by combining other variables in an expression (Note that expressions must follow BASIC's syntax requirements.)

Display/Function	Meaning
`Compute X1=SIN(X2)`	Define by using keywords or functions NOTE: Any combination of functions also is allowed

Compute does not differentiate between grouped and ungrouped data. Thus, you should never use compute with grouped data, as the observations and their frequencies would be defined by the functions.

Like all keyboard commands, the DEFine command can be used at any time while you are in the edit option, regardless of where you are in the data array. (NOTE: If you accidently get into the DEFine command, press **Q** to return to regular editing.)

The DELETE command — the **M** key — is used to delete a variable or observation from the data array, or to clear or reduce the number of observations for a variable. To accomplish this:

1. Press the **M** key; the display shows:

```
DELETE:Var. or Obs.?
```

2. Do one of the following:

- To delete a variable from the data array or to reduce the number of observations in a variable, press **V**; the program will display:

  ```
  Delete Var. Number?
  ```

  with the current variable number being the default value. Enter the variable number to delete; press **END LINE**. The display next will show:

  ```
  Starting Obs.?
  ```

  Enter the starting observation number for the variable that you wish to delete; press **END LINE**. The observation and all observations after the number entered, will be deleted (e.g., if X1(3) is deleted and there is an X1(4), then X1(4) also will be deleted). Note that to completely delete a variable, enter a starting observation number of 1. Also note that all variables after the deleted variable will be reduced by one (e.g., if variable X3 is deleted, and there is an X4, then X4 will become X3 and all following variables will be reduced by one), while retaining their current labels.

- To delete an observation from the data array, press **O**; the display will show:

  `Var. Number?`

  requires you to enter the number of the variable in which the observation occurs and press **END LINE**. The display next shows:

  `Delete Obs. Number?`

  with the current observation number being the default value; press **END LINE**. The following observations are all reduced by 1 (i.e., Delete X3(2); then X3(3) becomes X3(2), X3(4) becomes X3(3), etc.).

- To exit the DELETE command, press the **Q** key; the program will return to regular editing.

3. After accepting the default or entering the address of the variable or observation to delete and pressing **END LINE**, the program will display either:

   `DeleteVar.##?`

   or

   `DeleteX#Obs.##?`

   where the pound (#) signs denote the variable and/or observation address selected above.

4. Press the **Y** key for yes to delete, or the **N** key for no to exit.

After the deletion, the program returns to regular editing at the first element.

The GOTO command — the **F** key — allows you to move directly to a specified element in the data array. In order to do this:

1. Press the **F** key; the display shows:

   `GOTO:Var.,Obs.?`

2. Enter the variable number, a comma, and the observation number of the element in the data array; then, press the **END LINE** key. The program displays the element for edit or review.

The Endline Direction Key

When the **END LINE** key is pressed to enter an updated value into a data array, the next data array element is automatically displayed for editing. The direction that the program moves to display the next element is called the endline direction. The endline direction is set with the **S** key, the key in the middle of the direction keys (see Figure 3-3).

The default direction for the next element is down; thus, when the **END LINE** key is pressed the next observation down is displayed. As an example, the second observation is displayed after you enter the value in the first observation. Using the default endline direction, you can easily input your data into a matrix by editing each observation and pressing the **END LINE** key.

There are three possible endline directions:

- To the right along the variables. This is set by pressing the **S** key (for set endline direction) followed by pressing the **D** key (the right direction key).
- Down the observations. This is set by pressing the **S** key (for set endline direction) followed by pressing the **X** key (the down direction key).
- No motion. This is set by pressing the **S** key (for set endline direction) followed by pressing **S** again. This will cause the same element to be displayed after pressing the **END LINE** key.

When you press the **S** key, the display will show:

```
Direction:S,X,D or Q?
```

Press **S**, **X**, or **D**, depending on the endline direction that you desire. As noted above, when you press one of these endline direction keys, the program sets the endline direction and returns you to the last element displayed. The following table summarizes the effects of setting the endline direction.

Starting Element	Key	Endline Direction	Destination
X3(3)	**S D**	Right	X4(3)
X3(3)	**S X**	Down	X3(4)
X3(3)	**S S**	No Motion	X3(3)

Also, by pressing the **Q** key, you can escape the endline direction command; this will return you to the Edit option.

Exiting the Edit Option

To exit or quit the data array editor, press the **Q** key. This allows you to return to the main menu. The **Q** key can be used regardless of where you are in the data array. When exiting the data array, intermediate calculations are computed and stored. Thus, it is advisable to complete all value editing prior to pressing the **Q** key.

The Statistical Menu

Once your data have been entered in a data array and, optionally, saved in a file and/or printed, you are ready to begin the AMPI Statistics Library programs. The statistical menu (see Figure 2-1) procedure involves:

1. Specifying the AMPI Statistics Library program.
2. Responding to optional special prompts.
3. Responding to optional control prompts.
4. Performing the statistical computations.
5. Exiting the program.

Specifying the AMPI Statistics Library Programs

Specifying individual AMPI Statistics Library programs is accomplished from the main menu:

```
Data Edit Menu Quit?
```

by pressing **M**. The resulting display shows the statistical menu; this menu is discussed in the following paragraphs.

To select and run the 18 programs found in this library, a special menu has been designed to assist you in quickly finding each program. This scrolling menu, referred to as the statistics menu, allows you to find (i.e., by pressing the ▲ and ▼ keys) and to select (i.e., by pressing the **END LINE** key) the statistical program you desire. (NOTE: By pressing **Q** while in the statistical menu, you will return to the main menu.) The statistical programs used in this menu are defined in Table 3-3.

For easy access, the 18 programs have been grouped according to their statistical function. These groupings correspond to the order of the sections found in Parts II and III of this manual.

The AMPI Statistics Library Programs

The AMPI Statistics Library programs described in Parts II and III all follow the same format. Each program description contains a general introduction, equations, remarks, user information form, and one or more examples with process instructions. For readability and ease of program operation, the format is different for the program user instruction form and the example process instructions.

The user information form takes you through the entire sequence of steps (except for error and status messages) necessary to complete each program. For each step, a sequential number is given, the display is shown, the response and comments are given, and the next "go to" step is listed. Thus, although sometimes complex, the various branches of each program can be easily followed. The sequence of prompts shown in each user information form assumes that: (1) you have loaded the AMPI Statistics Library ROM into the HP-71B; (2) you are running the AMPI Statistics Library; (3) you have already entered, edited, printed, saved, and/or

Table 3-3. AMPI Statistics Library programs

Function / Type / Program Name	Statistics Menu Title
General Statistical Tests	
Descriptive Statistics	
Means and Moments	Means and Moments
Histogram	Histogram
Multiple Linear Regression	Multiple Linear Reg.
Parametric Tests	
t-Statistics	
Paired t-Test	Paired t-Test
Unpaired t-Test	Unpaired t-Test
F-Statistics	
One-way Analysis of Variance	One-way ANOVA
Two-way Analysis of Variance	Two-way ANOVA
Chi-square Statistics	
Contingency Table	Contingency Table
Non-parametric Tests	
Rank Statistics	
Mann-Whitney U Test	Mann-Whitney U Test
Kruskal-Wallis Test	Kruskal-Wallis Test
Statistical Distributions	
Sampling Distributions	
Student's t-Distribution	Student's t-Dist.
F-Distribution	F-Distribution
Chi-square Distribution	Chi-square Dist.
Probability Distributions	
Normal Distribution	Normal Distribution
Weibull Distribution	Weibull Distribution
Exponential Distribution	Exponential Dist.
Binomial Distribution	Binomial Distribution
Poisson Distribution	Poisson Distribution

loaded your input data; (4) the input data in memory are valid for each particular program (see Table 3-1); and (4) a printer is not attached to the HP-71B.

The process instructions for the two examples encountered earlier in Part I, Section 2, are similar in format to those found in Parts II and III in that they all present the keystrokes and displays for each example; they differ in that no detailed comments are presented in the examples in the latter Sections. The process instructions for the text examples take you only through the sequence of steps (including loading the AMPI Statistics Library ROM, data entry, and output of printed results, but excluding error and status messages) necessary to complete each example. Thus, these simplified process instructions present the specific keystrokes and display prompts inherent to each example.

Part II

Statistical Tests

Part II presents 10 useful statistical tests. These are here divided into descriptive statistical tests, parametric tests, and non-parametric tests.

The procedures for describing groups of data are called descriptive statistics; the procedures for determining the probability that specific samples of observations are the result of chance variation are termed inferential statistics. Section 4 of this manual discusses general descriptive statistics.

A parametric statistical test is a test whose model specifies four conditions about the parameters of the population from which the sample was drawn. These conditions are:

- The observations must be independent;
- The observations must be taken from normally distributed populations;
- The populations must have the same variance (or, a known ratio of variances); and
- The means of the normal populations must be linear combinations of effects related to columns and/or rows.

These conditions are not ordinarily tested; instead, they are assumed to hold true. The meaningfulness of the results of a parametric test depends on the validity of these assumptions. Parametric tests require that the values under analysis result from measurement in the strength of at least an interval scale. Sections 5, 6, and 7 present five popular parametric tests.

A non-parametric statistical test is a test whose model does not specify conditions about the parameters of the population from which the sample was drawn. Non-parametric statistical tests have two assumptions:

- The observations are independent; and
- The variable under study has underlying continuity.

These assumptions are much weaker than those associated with parametric tests. In addition, non-parametric tests do not require measurements as strong as those associated with parametric tests. Most non-parametric tests apply to data in an ordinal scale; some also apply to data in a nominal scale.

The advantages of non-parametric statistical tests are:

- Probability statements obtained from most non-parametric tests are exact probabilities, regardless of the shape of the population distribution from which the random sample was taken;

- If sample sizes as small as N = 6 are used, there is no alternative to using a non-parametric statistical test unless the nature of the population is known exactly;

- There are non-parametric statistical tests for analyzing samples from several different populations;

- There are non-parametric statistical tests for treating (1) data which are inherently in ranks, and (2) data whose seemingly numerical values have the strength of ranks;

- There are non-parametric statistical tests for analyzing data which are measured on a nominal scale (i.e., simple classificatory); and

- Non-parametric statistical tests are usually easier to learn and use than are parametric tests.

The disadvantages of non-parametric tests are:

- Non-parametric statistical tests are wasteful of data if all the assumptions of the parametric statistical model are met and if the measurement is of the required strength;

- There are no non-parametric statistical tests available for testing interactions in the analysis of variance model, unless special assumptions are made concerning additivity; and

- Non-parametric statistical tests and their accompanying tables of significance are scattered in the literature and thus are inaccessible to all but the professional statistician.

Section 8 presents two widely used non-parametric tests.

Section 4

Descriptive Statistics

A basic and fundamental element in any science is the descriptive stage. One must describe the data (or facts) accurately as they are. If one does not do this, an analysis of these data and their causes is premature. The basic question "What?" precedes the question "How?"

Often a person finds himself (or herself) in a position of having so much information or data that he (or she) cannot absorb all of it adequately. Thus numerical summaries are necessary to describe concisely (and accurately) the properties of the observed distribution of the data. Quantities which provide such a numerical summary are known as descriptive statistics.

Two kinds of descriptive statistics are discussed in this section — statistics of location and statistics of dispersion. Statistics of location (or measures of central tendency) describe the position of a sample along a given measurement (or dimension) delineating a variable. A statistic of location must provide a representative value for the entire sample. Statistics of location do not describe the shape of distributions. Thus, statistics of dispersion provide quantitative measures of frequency distribution.

Means and Moments

The Means and Moments Program allows you to describe statistically a sample population and characterize its distribution. Specifically, the means and moments of a statistical sample of a population provide a measure of central tendency for the sample and describe the shape and symmetry of a distribution. In other words, the program provides both statistics of location and statistics of dispersion. This program allows you to find several values associated with means and moments of statistical data. These are defined and discussed in the following paragraphs.

A common statistic of location is the sample size, usually defined as the number of items in a sample. The most familiar statistic of location, however, is the arithmetic mean, commonly called the mean or average. Means calculated from transformed variables (i.e., original variables that have been transformed into their logarithms or reciprocals) and retransformed back into the original scale will not be the same as those arithmetic means computed on the original variables. The retransformed mean of logarithmically transformed variables is known as the geometric mean. The reciprocal of the arithmetic mean of reciprocals is known as the harmonic mean.

Although not determined by this program, two other statistics of location — the median and mode — are of interest because of their use in defining other terms and in discussing frequency distributions. The median is the value of the data variable in an ordered array that has an equal number of data variables on either side of it. The mode is the value represented by the greatest number of individuals in a frequency distribution.

Moment statistics are useful in measuring the nature and amount of departure by an observed frequency distribution from a normal distribution. Normal distributions are symmetrical. (Frequency distribution, including normal distributions, will be discussed in Section 10.) There are two kinds of such departure from normality — skewness and kurtosis.

Skewness is another name for asymmetry and denotes that one tail of the distribution curve is drawn out more than the other. Distribution curves are either symmetrical or skewed to the right or left. The relationship between skewness and the relative position of the mean and median is illustrated in Figure 4-1.

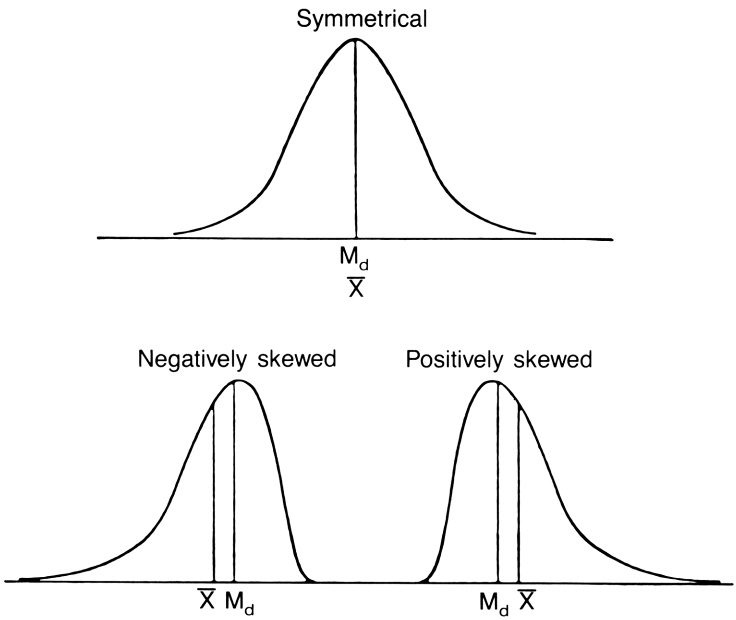


Figure 4-1. Relationship between skewness and the relative positions of the mean and median.

Kurtosis is another name for the peakedness of a distribution curve. A leptokurtic curve denotes that there are more data points near the mean and at the tails, with fewer data points in the intermediate regions of the curve. A platykurtic curve has fewer data points at the mean and at the tails, with more data points in the intermediate regions of the curve. A mesokurtic curve is not very peaked or very flat-topped and thus represents the normal distribution. A bimodal distribution has two peaks (equal or unequal in height) while a multimodal distribution has more than two peaks; both are examples of an extreme platykurtic distribution. Figure 4-2 presents several types of distribution curves.

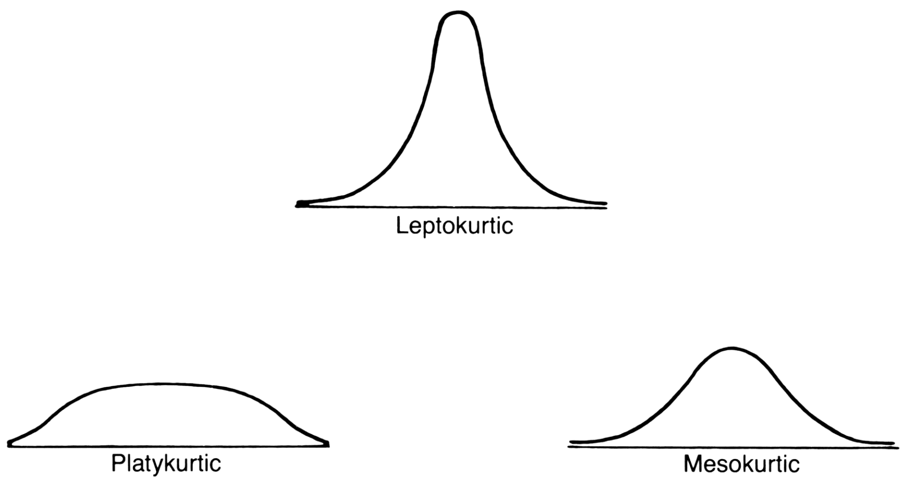


Figure 4-2. Examples of kurtosis.

Equations

Given a set of data points:

Ungrouped data	X(1), X(2), ... , X(n)
Grouped data	X(1), X(2), ... , X(n)
	F(1), F(2), ... , F(n)

where n is the number of observations and F(i) is the frequency of group i, the Means and Moments Program computes the following statistics.

Sample size, denoted in this program as # Data, is defined as:

Ungrouped data:

$$\# \text{ Data} = n$$

Grouped data:

$$\# \text{ Data} = \sum_{i=1}^{n} F(i)$$

where the capital Greek sigma, Σ, simply means the sum of the items indicated, i=1 means that the items should be summed starting with the first one and ending with the nth one as indicated by n on top of the Σ, and F(i) is the frequency measurements within a group.

The arithmetic mean (usually represented by $\overline{X}$ or $\bar{x}$, but denoted as A Mean in this program) is defined as follows.

Ungrouped Data:

$$\text{A Mean} = \frac{\sum_{i=1}^{n} X(i)}{\#\ \text{Data}}$$

Grouped Data:

$$\text{A Mean} = \frac{\sum_{i=1}^{n} F(i)\ X(i)}{\#\ \text{Data}}$$

The geometric mean, represented by G Mean in this program, is found by either of two equations.

Ungrouped Data:

$$\log\ \text{G Mean} = \frac{\sum_{i=1}^{n} \log\ X(i)}{\#\ \text{Data}}$$

Grouped Data:

$$\log\ \text{G Mean} = \frac{\sum_{i=1}^{n} F(i)\ \log\ X(i)}{\#\ \text{Data}}$$

The harmonic mean, denoted as H Mean in this program, is defined as follows.

Ungrouped Data:

$$\text{H Mean} = \frac{\#\ \text{Data}}{\sum_{i=1}^{n} \frac{1}{X(i)}}$$

Grouped Data:

$$\text{H Mean} = \frac{\#\ \text{Data}}{\sum_{i=1}^{n} \frac{F(i)}{X(i)}}$$

The relation between arithmetic mean, geometric mean, and harmonic mean is shown by the following comparison.

$$\text{A Mean} \geqslant \text{G Mean} \geqslant \text{H Mean}$$

The equality sign here holds true only if all samples are identical.

The rth, or generalized, moment about the mean is defined as follows.

Ungrouped Data:

$$\text{rth Moment} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})^r}{\#\ \text{Data}}$$

Grouped Data:

$$\text{rth Moment} = \frac{\sum_{i=1}^{n} F(i)\ (X(i) - \text{A Mean})^r}{\#\ \text{Data}}$$

Using these generalized equations, the first moment (here denoted as Moment 1) about the mean will always be zero for any distribution as shown below.

Ungrouped Data:

$$\text{Moment 1} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})}{\#\ \text{Data}}$$

Grouped Data:

$$\text{Moment 1} = \frac{\sum_{i=1}^{n} F(i)\ (X(i) - \text{A Mean})}{\#\ \text{Data}}$$

The second moment (represented by Moment 2 in this program) is defined as the variance, σ^2, by the following equations.

Ungrouped Data:

Population Variance:

$$\text{Moment 2} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})^2}{\#\ \text{Data}}$$

Sample Variance:

$$\text{Moment 2} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})^2}{\#\ \text{Data} - 1}$$

Grouped Data:

Population Variance:

$$\text{Moment 2} = \frac{\sum_{i=1}^{n} F(i)\ (X(i) - \text{A Mean})^2}{\#\ \text{Data}}$$

Sample Variance:

$$\text{Moment 2} = \frac{\sum_{i=1}^{n} F(i)\ (X(i) - \text{A Mean})^2}{\#\ \text{Data} - 1}$$

The variance is a measure of the variability (i.e., dispersion of a population). It will later be used to calculate skewness and kurtosis. Whereas "# Data" represents sample size or n, "# Data − 1" here denotes sample size minus one or n − 1 which is usually referred to as the degrees of freedom. The n form of degrees of freedom is used when the data represent the entire population; the n − 1 form is used when the data represent a sample drawn from a larger population.

The third moment of the population, denoted by Moment 3 in this program, is defined by the following formulas.

Ungrouped Data:

$$\text{Moment 3} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})^3}{\#\ \text{Data}}$$

Grouped Data:

$$\text{Moment 3} = \frac{\sum_{i=1}^{n} F(i)\ (X(i) - \text{A Mean})^3}{\#\ \text{Data}}$$

The third moment is used to determine the skewness or symmetry of the distribution. When the third moment equals zero, the distribution is symmetrical. When the third moment is a positive value, the distribution is skewed to the right; when it is a negative value, the distribution is skewed to the left. The distribution may be evaluated using a t-Test (see Section 5) or by using a skewness coefficient (here denoted by Skewness).

$$\text{Skewness} = \frac{\text{Moment 3}}{(\text{Moment 2})^{3/2}}$$

The fourth moment (denoted in this program by Moment 4) about the mean measures the kurtosis (i.e., shape) of a distribution. The fourth moment of the population is defined by the following equation.

Ungrouped Data:

$$\text{Moment 4} = \frac{\sum_{i=1}^{n} (X(i) - \text{A Mean})^4}{\#\ \text{Data}}$$

Grouped Data:

$$\text{Moment 4} = \frac{\sum_{i=1}^{n} F(i)\,(X(i) - \text{A Mean})^4}{\#\ \text{Data}}$$

The kurtosis coefficient (represented in this program by Kurtosis) is determined by the following formulas.

$$\text{Kurtosis} = \frac{\text{Moment 4}}{(\text{Moment 2})^2}$$

or

$$\text{Kurtosis} = \frac{\text{Moment 4}}{\sigma^4}$$

For a mesokurtic (i.e., normal) distribution, the kurtosis coefficient is equal to three. For a leptokurtic distribution, the kurtosis coefficient is greater than three. For a platykurtic distribution, the kurtosis coefficient is less than three.

Remarks

The Means and Moments Program requires one variable. To calculate the means, one observation is required. To calculate the standard deviation or the variance (i.e, Moment 2 in this program), two observations are necessary.

The geometric mean will not be calculated if you have any value equal to, or less than, zero. Similarly, the harmonic mean will not be calculated when any value is equal to zero. In this case, these statistics have the value "NaN" (i.e., Not-a-Number).

The Means and Moments Program does not compute the standard deviation, σ. This value, however, may be simply determined by taking the square root of the second moment (or variance, σ^2), which is represented by Moment 2 in this program.

One of the options you have in the Means and Moments Program is to use n or $n-1$ for degrees of freedom, depending on whether your data represent an entire population or a sample from it, respectively. The degrees of freedom affects Moment 2, Kurtosis, and Skewness.

User Information Form

This section presents a listing and brief tutorial explanation of the Means and Moments Program's display prompts and options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details see Part I To select Statistical Menu option, press **M**	2
		To select Quit option, press **Q**; for details see Part I	
2.	`Means and Moments`	First program title; press **END LINE**	3
3.	`Use n-1 for df?`	To use degrees of freedom equal to n − 1, press **Y**	4
		To use degrees of freedom equal to n, press **N**	4
4.	`Edit Parameters?`	To edit parameters, press **Y**	3
		To continue without editing parameters, press **N**	5
5.	`Var.#=0`	Enter the number of the variable that you wish to test; then, press **END LINE** (NOTE: $1 \leq$ Var. # $\leq$ # of variables; Var. # must be an integer.)	6
6.	`Edit Parameters?`	To edit parameters, press **Y**	5
		To continue without editing parameters, press **N**	7
7.	`#Data=###############`	Number of observations: if output is to display, press any key; if output is to printer, no action necessary	8

Step	Display	Response and Comments	Go To
8.	AMean=################	Arithmetic mean: if output is to display, press any key; if output is to printer, no action necessary	9
9.	GMean=################	Geometric mean: if output is to display, press any key; if output is to printer, no action necessary	10
10.	HMean=################	Harmonic mean: if output is to display, press any key; if output is to printer, no action necessary	11
11.	Moment2=##############	Second moment: if output is to display, press any key; if output is to printer, no action necessary	12
12.	Moment3=##############	Third moment: if output is to display, press any key; if output is to printer, no action necessary	13
13.	Moment4=##############	Fourth moment: if output is to display, press any key; if output is to printer, no action necessary	14
14.	Kurtosis=##############	Kurtosis: if output is to display, press any key; if output is to printer, no action necessary	15
15.	Skewness=##############	Skewness: if output is to display, press any key; if output is to printer, no action necessary	16
16.	Repeat Results?	To repeat output results, press **Y**	7
		To continue program, press **N**	1

Example

1. Ungrouped Data:

In 1984 at Sarajevo, Yugoslavia, the 43 women that completed the Women's Giant Slalom all finished their combined two ski runs within 42.47 seconds of each other. Given the following official results, determine the means and moments of their times.

Save the data on cassette tape and print the results.

Placement	Contestant	Country	Time (Sec.)
1	Debbie Armstrong	United States	140.98
2	Christin Cooper	United States	141.40
3	Perrine Pelen	France	141.83
4	Tamara McKinney	United States	141.98
5	Marina Kiehl	West Germany	142.03
6	Blanca Fernandez	Spain	142.14
7	Erika Hess	Switzerland	142.51
8	Olga Charvatova	Czechoslovakia	142.57
9	Liisa Savijarvi	Canada	142.73
10	Anne-Flore Rey	France	142.95
11	Carole Merle	France	143.27
12	Michela Figini	Switzerland	143.34
13	Maria Epple	West Germany	143.65
14	Anni Kronbichler	Austria	144.17
15	Monika Hess	Switzerland	144.58
16	Andrea Leskovsek	Yugoslavia	144.61
17	Diana Haight	Canada	144.66
18	Cindy Nelson	United States	144.88
19	Petra Wenzel	Liechtenstein	144.94
20	Veronika Sarec	Yugoslavia	145.01
21	Irene Epple	West Germany	145.52
22	Nusa Tome	Yugoslavia	146.21
23	Mateja Svet	Yugoslavia	146.22
24	Michaela Gerg	West Germany	146.28
25	Daniela Zini	Italy	146.33
26	Alexandra Marasova	Czechoslovakia	146.37
27	Roswitha Steiner	Austria	146.56
28	Ivana Valesova	Czechoslovakia	146.71
29	Nadeida Andreeva	Soviet Union	146.85
30	Dorota Tlalka	Poland	146.90
31	Ewa Grabowska	Poland	147.55
32	Paoletta Magoni	Italy	147.87
33	Laurie Graham	Canada	148.42
34	Sylvia Eder	Austria	149.03
35	Liliana Ichim	Romania	152.08
36	Teresa Bustamante	Argentina	154.19
37	Michele Dombard	Belgium	154.79
38	Nanna Leifsdotir	Iceland	154.84
39	Magdalena Birkner	Argentina	159.70
40	Gabriela Angaut	Argentina	163.16
41	Geraldina Bobbio	Argentina	167.51
42	Xuefei Jin	China	170.35
43	Lina Aristodimou	Cyprus	183.45

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Load Save Print?`

Keystrokes	Display
K	`Grouped Ungrouped?`
U	`No. of Var.=0`
1 **END LINE**	`No. of Obs.=0`
43 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
N	`X(1)=0`
140.98 **END LINE**	`X(2)=0`
141.40 **END LINE** • • •	NOTE: The program continues to step through each of the 43 elements of the entire array. Continue to enter the values from the example as the program calls for them; press **END LINE** after each entry. After all 43 elements have been entered, the editor wraps back to the first element.
183.45 **END LINE**	`X(1)=140.98`
Q	`Data Edit Menu Quit?`
D	`Kbd Load Save Print?`
S	`SAVE:Filename?_`
SLALOM **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
END LINE	`Use n-1 for df?`
N	`Edit Parameters?`
N	`Var.#=0`
1 **END LINE**	`Edit Parameters?`
N	(See Figure 4-3 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

Figure 4-3 presents the results which point out that the arithmetic mean, or average time, for the event was 148.398139535 seconds; because our input data were measured in hundredths, this should be rounded to 148.40 seconds. The rounded off geometric mean and harmonic mean are, respectively, 148.17 and 147.96 seconds. As expected, the geometric mean is less than (or equal to) the arithmetic mean, but is greater than the harmonic mean. The second moment, or variance of the event, is 73.57. The third and fourth moments are 1457.84 and 45663.22, respectively. The skewness coefficient is 2.31, while the kurtosis coefficient is 8.44. The positive third moment and positive skewness indicate that the distribution is skewed to the right and, thus, has a longer "tail" to the right of the central maximum than to the left. Because kurtosis is greater than three, the distribution is said to be leptokurtic (i.e., it has more data points near the mean and at the tails, with fewer data points in the intermediate regions of the curve).

Data Format Ungrouped

No. of Var. = 1
No. of Obs. = 43

Variable Labels

X1 = X1

X1

X(1) = 140.98
X(2) = 141.4
X(3) = 141.83
X(4) = 141.98
X(5) = 142.03
X(6) = 142.14
X(7) = 142.51
X(8) = 142.57
X(9) = 142.73
X(10) = 142.95
X(11) = 143.27
X(12) = 143.34
X(13) = 143.65
X(14) = 144.17
X(15) = 144.58
X(16) = 144.61
X(17) = 144.66
X(18) = 144.88
X(19) = 144.94
X(20) = 145.01
X(21) = 145.52
X(22) = 146.21
X(23) = 146.22
X(24) = 146.28
X(25) = 146.33
X(26) = 146.37
X(27) = 146.56
X(28) = 146.71
X(29) = 146.85
X(30) = 146.9
X(31) = 147.55
X(32) = 147.87
X(33) = 148.42
X(34) = 149.03
X(35) = 152.08
X(36) = 154.19
X(37) = 154.79
X(38) = 154.84
X(39) = 159.7
X(40) = 163.16
X(41) = 167.51
X(42) = 170.35
X(43) = 183.45

Means and Moments

Var. # = 1

df = n

Data = 43
A Mean = 148.398139535
G Mean = 148.169517007
H Mean = 147.958343236
Moment 2 = 73.5665918874
Moment 3 = 1457.84232268
Moment 4 = 45663.215297
Kurtosis = 8.43733347431
Skewness = 2.31041288866

Figure 4-3. Printed results for Means and Moments Program example problem.

Histogram

It is useful to describe large amounts of data in the form of frequency tables. Such tables assist statisticians in summarizing and analyzing data. Graphical presentation of frequency tables offer another tool to enhance further the ability of statisticians to analyze data. In the preceding paragraphs we have used frequency distribution curves in discussing data. Additional ways in which data may be depicted graphically include pie charts, bar charts, histograms, and frequency polygons.

A pie or circle chart partitions a group of data measurements using percentages similiar to that of slicing a pie. A bar chart, diagram, or graph presents meristic data in the form of bars, where the abscissa (the X coordinate or horizontal axis) represents the variable and the ordinate (the Y coordinate or vertical axis) represents the frequencies. In a bar chart, the height of each rectangular cell or bar is proportional to the number of data measurements or observations per cell, the width of all the bars are equal (and have no significance), and the bars usually do not touch each other; thus, the area of each bar is proportional to the frequency of data observations in each cell. A histogram (a type of bar chart) presents continuous variables in the form of bars, where the abscissa represents the intervals of measured values and the ordinate represents the frequencies of occurrence. In a histogram, the width of each bar represents a class interval of the frequency distribution and the bars touch each other; this shows that the actual limits of the classes are contiguous and the midpoint of each bar corresponds to the class mark. A frequency polygon is simply a histogram which has the midpoints of the tops of each rectangle connected with straight lines and the rectangles blotted out. Smoothing this line through these midpoints would create a frequency distribution curve.

Comparison of Figures 4-4, 4-5, 4-6, and 4-7 illustrates the differences between a bar chart, a histogram, a frequency polygon, and a frequency distribution curve.

Because a histogram, frequency polygon, and frequency distribution curve can be generated from the same data as a frequency distribution, the means and moments of a histogram may be computed and statistically compared to various theoretical histograms. The Histogram Program provides a goodness-of-fit test for the following theoretical histograms:

- Normal distribution
- Weibull distribution
- Exponential distribution
- Binomial distribution
- Poisson distribution
- Uniform distribution

This goodness-of-fit test calculates the chi-square statistic (χ^2) and its probability $Q(\chi^2)$. In addition to these two statistical values, the program generates the arithmetic mean, geometric mean, harmonic mean, second moment, third moment, fourth moment, kurtosis, and skewness, using the cell means and frequencies.

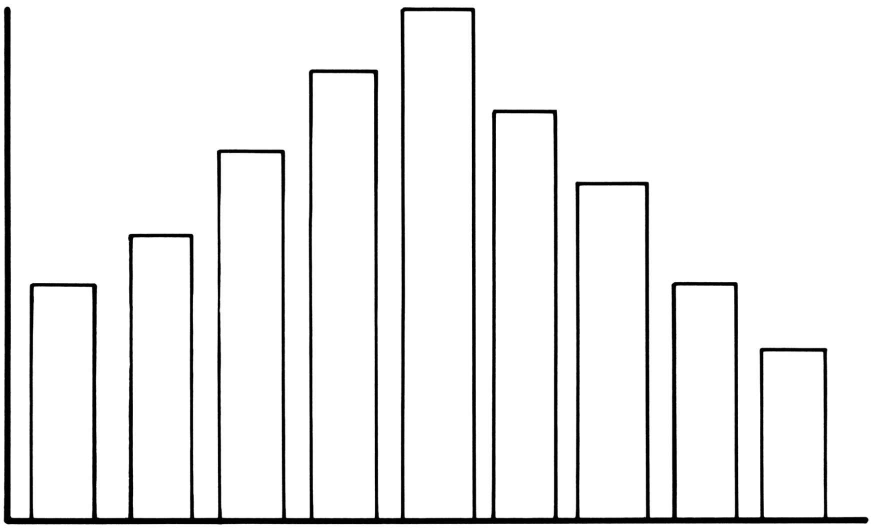


Figure 4-4. A typical bar chart.

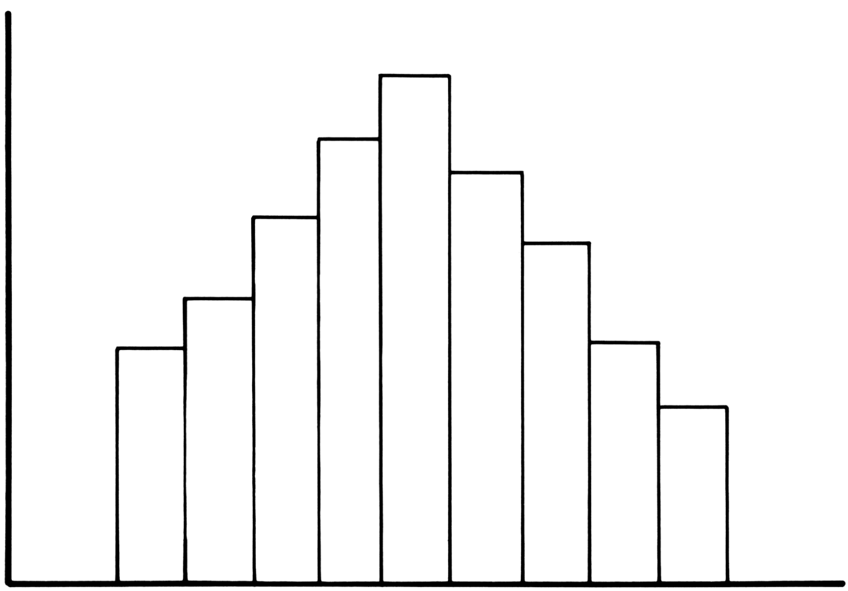


Figure 4-5. A typical histogram.

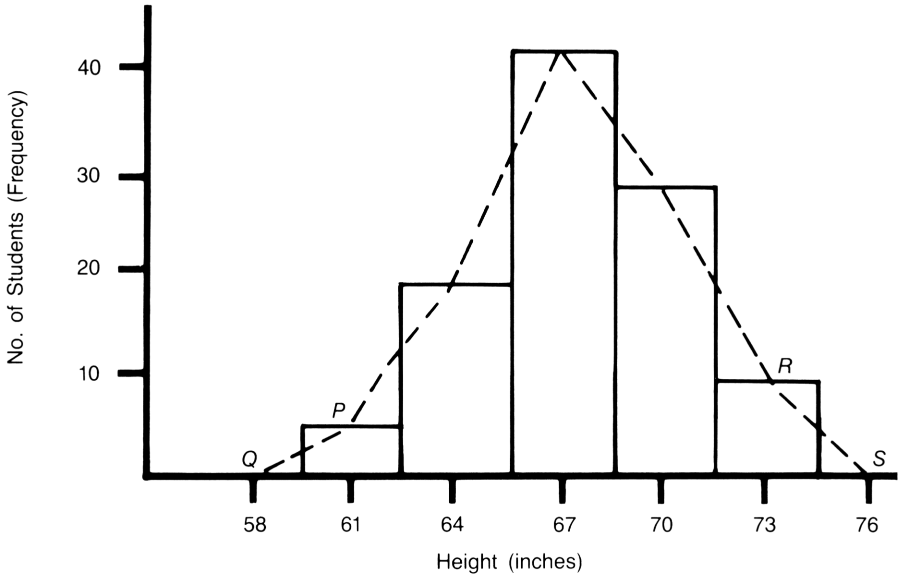


Figure 4-6. A typical frequency polygon.

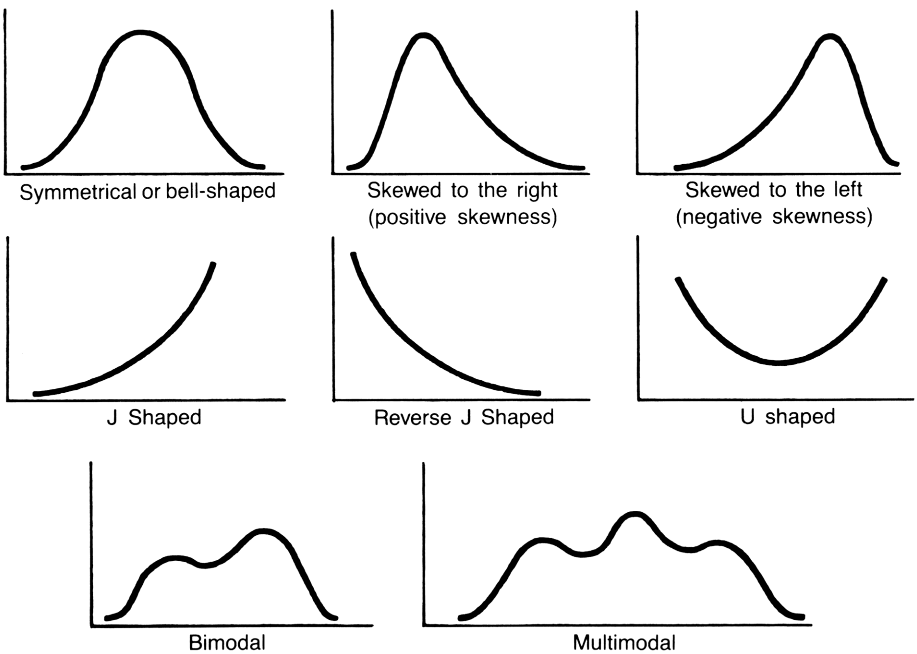


Figure 4-7. Typical frequency distribution curves.

The Histogram Program has applications in many general business and scientific fields, including product testing, probability analysis, and engineering evaluations.

Equations

The mathematical formulas and discussions for the means and moments may be found under the Means and Moments Program in the first portion of this section. Because the Histogram Program may exclude some of your values, the results may be different from those obtained using the Means and Moments Program (see further discussion under Remarks).

Mathematical formulas and discussions for five of the probability distributions — Normal, Weibull, Exponential, Binomial, and Poisson — can be found in Section 10 of this manual.

The mathematical formula for the uniform distribution is simple, as shown below:

$$\frac{\text{Total Number of Observations}}{\text{Number of Cells}}$$

The number of observations or occurrences in every cell must be zero or a positive number in order for it to be a true uniform distribution. Although realistically impossible, this theoretical distribution often (depending on the data) calculates a fraction of an occurrence in each cell.

Chi-square is used to test for the goodness-of-fit between the observed data and the expected or predicted data of all the theoretical histograms. The chi-square statistic is calculated by the following equation.

$$\chi^2 = \sum_{i=1}^{n} \frac{(\text{expected value} - \text{observed value})^2}{\text{expected value}}$$

where the number of cells in the histogram is n and where the number of degrees of freedom (df) is usually $n-1$. (NOTE: The number of degrees of freedom varies with the theoretical distribution; see Remarks.)

The probability, $Q(\chi^2)$, of rejecting a true null hypothesis, H_0, is computed using the chi-square distribution.

Remarks

The expected value used in calculating the chi-square statistic cannot equal zero as valid data would not produce an expected frequency of zero. However, in extreme instances (e.g., in the tails of the various distribution curves), a value that is very near to zero may be rounded to zero internally and therefore will cause an illogical calculation. In such cases, an error message will be displayed because the chi-square statistic cannot be calculated. (NOTE: If the HP-71B freezes and pressing **f** followed by **CONT** does not help, recovery may require that you press **ATTN**, followed by typing END ALL and pressing **END LINE**.)

For statistical purposes, one degree of freedom is lost when calculating chi-square in comparing original data with any one of the theoretical distributions. During the comparison for each parameter that is estimated from the original histogram, an additional degree of freedom is lost. In other words, if you let the program calculate parameters rather than provide (or estimate) them, one additional degree of freedom is lost for each calculated parameter. These extra degrees of freedom are not lost if the parameters are not used as calculated from the original data that you entered, but are entered as hypothesized values. As an example, the calculated parameters from the normal distribution are the mean and standard deviation; if calculated, the total degrees of freedom to be lost in this comparison would be three (i.e., remaining df = n − 3). With the Poisson distribution, the total degrees of freedom lost would be two (i.e., remaining df = n − 2) because the mean is the only calculated parameter. With the binomial distribution, the total degrees of freedom lost also will be two (i.e., remaining df = n − 2) when the probability p is calculated. With the Weibull, exponential, and uniform distributions, there are no estimates accepted; thus, the total degrees of freedom lost is one (i.e., remaining df = n − 1).

This information concerning degrees of freedom is very important in determining whether or not you have enough cells in your histogram to make a valid comparison of your data with one of the theoretical distributions. As an example, you cannot validly compare your data with a normal distribution curve when accepting the calculated parameters unless you have at least four cells for your data. For a normal curve, the degrees of freedom remaining after subtracting for estimators is n − 3; in our example, n = 4 and the degrees of freedom would then be 4 − 3, thus leaving the minimum of one as the number of remaining degrees of freedom for a valid comparison.

The number of cells in your histogram must be equal to or greater than three and must be a positive whole number. The value selected is used in the calculations throughout the program and establishes how many cells into which the histogram will be divided. For grouped data, the number of cells must be equally spaced and in increasing order.

The cell width value must be positive, but does not have to be a whole number. The cell width value is the number of measured units into which you want to categorize your values as one cell. As an example, if your data varies between 1 and 200 and you want 10 divisions or cells, your cell width will be 20. For grouped data, the cell width value is assumed to be the difference between the means.

The program allows you to set a minimum value for the X scale of the histogram. Usually this value is zero, because most data start from zero. But if all of your data for a particular set of observations were between 300 and 800, you should set the X minimum value slightly less than 300. The X minimum value and cell boundary values should never be set at an attainable level by the data because all data points that fall on the dividing line will be placed in the cell to the right of the boundary value. A value of 800 would be excluded from the highest cell if the calculated maximum X value was set at 800. A more appropriate calculated X value would be 800.5 in this case. If all of your values are integers, the use of a non-integer lower boundary prevents any entries from occurring on the boundaries. For grouped data, the X minimum value is equal to X_1 minus half the cell width.

Because the boundary limits that you set may exclude some of your data values, the results from this program may be different from those obtained by using the Means and Moments Program. Also, the Histogram Program calculates the means and moments on cell means and frequencies (grouped data) rather than raw ungrouped data.

The Histogram Program does not compute the standard deviation, σ. This value, however, may be simply calculated by taking the square root of the second moment (or variance, σ^2), which is represented by Moment 2 in this program.

User Information Form

This section presents a listing and brief tutorial explanation of the Histogram Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details see Part I To select Statistical Menu option, press **M**	2
		To select Quit option,press **Q**; for details see Part I	
2.	`Means and Moments`	First program title; press ▼ until "Histogram" appears on the LCD	3
3.	`Histogram`	Program title; press **END LINE**	4
4.	`Var.#=0`	Enter value for number of variables; then, press **END LINE** (NOTE: 1≤ Var. # ≤ # of variables; Var. # must be an integer.)	
5.	`# of Cells=0`	Enter value for number of cells (integer, 2 < # of Cells ≤ MAX-REAL); then, press **END LINE**	6
6.	`Width=0`	Enter width of cell (any number, 0 < Width ≤ MAXREAL); then, press **END LINE**	7
7.	`Minimum=0`	Enter value for minimal acceptable data value (integer, $-\infty <$ Minimum $\leq \infty$); then, press **END LINE**	8
8.	`Edit Parameters?`	To edit parameters, press **Y**	4
		To continue program without editing, press **N**	9

Step	Display	Response and Comments	Go To
9.	Output Groups?	To view groups, press **Y**	10
		To continue program without viewing groups, press **N**	14
10.	Cell#=####	The number of the cell (this begins with 1): if output is to display, press any key; if output is to printer, no action necessary	11
11.	Freq=####	The frequency value of the cell designated in Step 10: if output is to display, press any key; if output is to printer, no action necessary	12
12.	Mean=####	The mean of the cell designated in Step 10: if output is to display, press any key; if output is to printer, no action necessary. (NOTE: Steps 10, 11, and 12 are repeated for the number of cells entered in Step 5.)	13
13.	Repeat Results?	To repeat results, press **Y**	10
		To continue program, press **N**	14
14.	Use n-1 for df?	To use degrees of freedom equal to n − 1, press **Y**	15
		To use degrees of freedom equal to n, press **N**	15
15.	Edit Parameters?	To edit parameters, press **Y**	14
		To continue program without editing, press **N**	16
16.	Output Statistics?	To view means and moments, press **Y**	17
		To continue program without viewing means and moments, press **N**	27
17.	#Data=###############	Number of observations: if output is to display, press any key; if output is to printer, no action necessary	18
18.	AMean=###############	Arithmetic mean: if output is to display, press any key; if output is to printer, no action necessary	19

Step	Display	Response and Comments	Go To
19.	GMean=###############	Geometric mean: if output is to display, press any key; if output is to printer, no action necessary	20
20.	HMean=###############	Harmonic mean: if output is to display, press any key; if output is to printer, no action necessary	21
21.	Moment2=#############	Second moment: if output is to display, press any key; if output is to printer, no action necessary	22
22.	Moment3=#############	Third moment: if output is to display, press any key; if output is to printer, no action necessary	23
23.	Moment4=#############	Fourth moment: if output is to display, press any key; if output is to printer, no action necessary	24
24.	Kurtosis=#############	Kurtosis: if output is to display, press any key; if output is to printer, no action necessary	25
25.	Skewness=#############	Skewness: if output is to display, press any key; if output is to printer, no action necessary	26
26.	Repeat Results?	To repeat means and moments, press **Y**	17
		To continue program, press **N**	27
27.	Theo Histogram?	To execute theoretical histogram, press **Y**	28
		To continue program without executing theoretical histograms, press **N**	1
28.	Normal Distribution	This is usually the first distribution menu prompt. (When returning from Step 56, the distribution menu will be the last one selected.) To select a distribution, scroll menu by pressing ▲ or ▼ until desired distribution appears on display	
		For normal distribution	29
		For Weibull distribution	34
		For exponential distribution	38
		For binomial distribution	41
		For Poisson distribution	46
		For uniform distribution	50

Step	Display	Response and Comments	Go To
29.	Normal Distribution	To select normal distribution, press **END LINE**	
		If # of Cells is > 3	30
		If # of Cells is ≤ 3	31
		To select another distribution	28
30.	Use Calc Param?	To use computed mean and standard deviation, press **Y**	51
		To use hypothetical mean and standard deviation, press **N**	31
31.	Mean=0	Enter hypothetical mean (any number, $-\infty \leq$ Mean $\leq \infty$); then, press **END LINE**	32
32.	Std Dev=0	Enter hypothetical standard deviation (any number, $0 <$ Std Dev $\leq \infty$); then, press **END LINE**	33
33.	Edit Parameters?	To edit parameters, press **Y**	31
		To continue program without editing, press **N**	51
34.	Weibull Distribution	To select Weibull distribution, press **END LINE**	35
		To select another distribution	28
35.	Shape=0	Enter value for shape parameter (any number, $0 <$ Shape $\leq \infty$); press **END LINE**	36
36.	Scale=0	Enter value for scale parameter (any number, $0 <$ Scale $\leq \infty$); press **END LINE**	37
37.	Edit Parameters?	To edit parameters, press **Y**	35
		To continue program without editing, press **N**	51
38.	Exponential Dist.	To select exponential distribution, press **END LINE**	39
		To select another distribution	28
39.	Shape=0	Enter value for shape parameter (any number, $0 <$ Shape $\leq \infty$); press **END LINE**	40

Step	Display	Response and Comments	Go To
40.	Edit Parameters?	To edit parameters, press **Y**	39
		To continue program without editing, press **N**	51
41.	Binomial Distribution	To select binomial distribution, press **END LINE**	42
		To select another distribution	28
42.	Use Calc Param?	To use the calculated probability value, press **Y**	43
		To use hypothetical probability, press **N**	43
43.	n=0	Enter value for number of trials (integer, $0 \leq n \leq 200$); then, press **END LINE**	
		If **Y** at Step 42	45
		If **N** at Step 42 (or # of Cells = 2)	44
44.	p=0	Enter hypothethical probability value (any number, $0 \leq p \leq 1$); press **END LINE**	45
45.	Edit Parameters?	To edit parameters, press **Y**	43
		To continue program without editing, press **N**	51
46.	Poisson Distribution	To select Poisson distribution, press **END LINE**	
		If # of Cells > 2	47
		If # of Cells = 2	48
		To select another distribution	28
47.	Use Calc Param?	To use computed success parameter, press **Y**	51
		To use hypothetical success parameter, press **N**	48
48.	Mean=0	Enter expected hypothetical success parameter (any number, $0 \leq$ Mean $\leq \infty$); press **END LINE**	49
49.	Edit Parameters?	To edit parameters, press **Y**	48
		To continue program without editing, press **N**	51

Step	Display	Response and Comments	Go To
50.	`Uniform Distribution`	To select uniform distribution, press **END LINE**	51
		To select another distribution	28
51.	`Output Exp-Freq?`	To output the expected frequencies, press **Y**	52
		To continue program without seeing the expected frequencies, press **N**	53
52.	`Ex-F(i)=##############`	Expected frequency of each cell: if output is to display, press any key; if output is to printer, no action necessary	53
53.	`df=###################`	Degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	54
54.	`Chi^2=################`	Chi-square: if output is to display, press any key; if output is to printer, no action necessary	55
55.	`Q(Chi^2)=#############`	Cumulative distribution function: if output is to display, press any key; if output is to printer, no action necessary	56
56.	`Repeat Results?`	To repeat output results, press **Y**	52
		To continue program, press **N**	27

Example

Construct a histogram of three (3) cells, whose cell width is ten (10) and whose minimum value is zero (0). The values to be grouped into cells are −1, 1, 5, 9, 10, 12, 14, 21, 34, and 50, respectively. Find the means and moments of the histogram data. (Note that the restrictions given will exclude three of the values.) Test these data against the theoretical histogram distributions. Save the data and print the results. To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Load Save Print?`
K	`Grouped Ungrouped?`

Keystrokes	Display
U	`No. of Var.=0`
1 **END LINE**	`No. of Obs.=0`
10 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
N	`X1(1)=0`
−1 **END LINE**	`X1(2)=0`
1 **END LINE**	`X1(3)=0`
5 **END LINE**	`X1(4)=0`
9 **END LINE**	`X1(5)=0`
10 **END LINE**	`X1(6)=0`
12 **END LINE**	`X1(7)=0`
14 **END LINE**	`X1(8)=0`
21 **END LINE**	`X1(9)=0`
34 **END LINE**	`X1(10)=0`
50 **END LINE**	`X1(1)=-1`
Q	`Data Edit Menu Quit?`
D	`Kbd Load Save Print?`
P	`Data Edit Menu Quit?`
D	`Kbd Load Save Print?`
S	`SAVE:Filename?_`
HISTO **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Histogram`
END LINE	`Var.#=0`
1 **END LINE**	`#ofCells=0`
3 **END LINE**	`Width=0`
10 **END LINE**	`Minimum=0`

Keystrokes	Display
END LINE	`Edit Parameters?`
N	(See Figure 4-8 for output) `Output Groups?`
Y	(See Figure 4-8 for output) `Repeat Results?`
N	`Use n-1 for df?`
N	`Edit Parameters?`
N	`Output Statistics?`
Y	(See Figure 4-8 for output) `Repeat Results?`
N	`Theo Histogram?`
Y	`Normal Distribution`
END LINE	`Mean=0`
12 **END LINE**	`Std Dev=0`
7 **END LINE**	`Edit Parameters?`
N	(See Figure 4-8 for output) `Output Exp-Freq?`
Y	(See Figure 4-8 for output) `Repeat Results?`
N	`Theo Histogram?`
Y	`Normal Distribution`
▼ or ▲ to find	`Weibull Distribution`
END LINE	`Scale=0`
1 **END LINE**	`Shape=0`
0.23 **END LINE**	`Edit Parameters?`
N	(See Figure 4-8 for output) `Output Exp-Freq?`
Y	(See Figure 4-8 for output) `Repeat Results?`

Keystrokes	Display
N	Theo Histogram?
Y	Weibull Distribution
▼ or ▲ to find	Exponential Dist.
END LINE	Shape=0
1.2 **END LINE**	Edit Parameters?
N	(See Figure 4-8 for output) Output Exp-Freq?
Y	(See Figure 4-8 for output) Repeat Results?
N	Theo Histogram?
Y	Exponential Dist.
▼ or ▲ to find	Uniform Distribution
END LINE	(See Figure 4-8 for output) Output Exp-Freq?
Y	(See Figure 4-8 for output) Repeat Results?
N	Theo Histogram?
N	Data Edit Menu Quit?
Q	Done
END LINE f OFF	

Figure 4-8 presents the results of this example. Again, note that the restrictions eliminated three of the original data values. The final histogram had three cells, with a frequency of 3, 3, and 1, and a mean of 5, 15, and 25, respectively. The means and moments section showed the arithmetic mean to be 12.1. As expected, the geometric mean is less than the arithmetic mean and is greater than the harmonic mean. The second moment, or variance of the sample, is 49.0. The positive third moment and positive skewness indicate that the frequency curve is skewed to the right. Because the kurtosis coefficient is less than three, the distribution is said to be platykurtic. The histogram more closely follows the uniform distribution.

Data Format Ungrouped

No of Var. = 1
No of Obs. = 10
Variable Labels
X1 = X1
X1
X1(1) = −1
X1(2) = 1
X1(3) = 5
X1(4) = 9
X1(5) = 10
X1(6) = 12
X1(7) = 14
X1(8) = 21
X1(9) = 34
X1(10) = 50

Histogram

Var. #1 = 1
# of Cells = 3
Width = 10
Minimum = 0

Cell # = 1
Freq = 3
Mean = 5
Cell # = 2
Freq = 3
Mean = 15
Cell # = 3
Freq = 1
Mean = 25

df = n

Data = 7
A Mean = 12.1428571429
G Mean = 10.0763536106
H Mean = 8.33333333333
Moment 2 = 48.9795918369
Moment 3 = 157.434402327
Moment 4 = 5047.89670969
Kurtosis = 2.10416666665
Skewness = .459279326754

H:Normal Distribution

Mean = 12
Std Dev = 7
Ex − F(1) = 2.71283980353
Ex − F(2) = 3.40131715526
Ex − F(3) = .885843041202
df = 2
Chi^2 = 9.24586585627E − 2
Q(Chi^2) = .954822968216

H:Weibull Distribution

Scale = 1
Shape = .23

Ex − F(1) = 5.7189673678
Ex − F(2) = .32584679057
Ex − F(3) = .955185841631
df = 2
Chi^2 = 23.2409699946
Q(Chi^2) = .000008980231

H:Exponential Dist.
Shape = 1.2

Ex − F(1) = 6.99831741367
Ex − F(2) = .001682181893
Ex − F(3) = .000000404439
df = 2
Chi^2 = 2477905.23734
Q(Chi^2) = 0

H:Uniform Distribution

Ex − F(1) = 2.33333333333
Ex − F(2) = 2.33333333333
Ex − F(3) = 2.33333333333
df = 2
Chi^2 = 1.14285714286
Q(Chi^2) = .564718122007

Figure 4-8. Printed output for Histogram Program example.

Multiple Linear Regression

In statistics, regression is a term used to evaluate a known (or suspected) dependency relationship between two or more variables. A straight-line relationship between only two variables involves simple linear regression. Multiple linear regression, on the other hand, involves a dependent variable which is a function of more than one independent variable. Linear regression forces a straight line graph through a set of data regardless of the plotted appearance of the data. In recent years, regression analysis has been widely used as a basic modeling tool for predicting the behavior of one variable based on limited tests or observations of another variable; this has been called predictive modeling. Regression analysis also aids you in understanding how close a relationship exists between (or among) variables within a certain range of the dependent variable's values.

Equations

The mathematical formula for simple linear regression is:

$$Y = a + bX$$

where Y is the dependent variable, X is the independent variable, the constant a denotes the Y-intercept (or population parameter), and the constant b represents the slope.

The mathematical equation for multiple linear regression is:

$$Y = a + b_1X_1 + b_2X_2 + \ldots b_pX_p$$

where Y is the dependent variable, the X_j are the independent variables, the constant a denotes the Y-intercept (or population parameter), and p is the number of independent variables. The b_j terms are the partial regression coefficients of Y on the independent X variables. The Y-intercept, a, is the value of Y when all of the X_j's are equal to zero. The Y-intercept, a, is forced to zero for the standardized model of the multiple regression equation.

The standardized model is defined by the equation:

$$Y = \sum_{j=1}^{p} b_jX_j$$

where b_j is the standard regression coefficient.

The non-standard model is defined by the equation:

$$Y = B_0 + \sum_{j=1}^{p} B_jX_j$$

where B_0 is the intercept, and B_j is the non-standardized regression coefficient.

The mathematical equation for the error sum of squares, SSE, is:

$$SSE = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$$

$$= \sum_{i=1}^{n} (\hat{Y}_i)^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$$

where Y_i is the observed value for Y and $\hat{Y}_i$ is the predicted or estimated value.

The following abbreviated forms are used to simplify mathematical notation:

$$\sum_{i=1}^{n} x_{ji}^2 = \sum_{i=1}^{n} X_{ji}^2 - \frac{(\sum_{i=1}^{n} X_{ji})^2}{n}$$

$$\sum_{i=1}^{n} x_{ji} y_i = \sum_{i=1}^{n} X_{ji} Y_i - \frac{(\sum_{i=1}^{n} X_{ji})(\sum_{i=1}^{n} Y_i)}{n}$$

$$\sum_{i=1}^{n} x_{ji} x_{ki} = \sum_{i=1}^{n} X_{ji} X_{ki} - \frac{(\sum_{i=1}^{n} X_{ji})(\sum_{i=1}^{n} X_{ki})}{n}$$

The matrix of simple correlation coefficients, $r_{j,k}$, is determined by the equation:

$$r_{j,k} = \frac{\sum_{i=1}^{n} (x_{ji} x_{ki})}{\sqrt{(\sum_{i=1}^{n} x_{ji}^2)(\sum_{i=1}^{n} x_{ki}^2)}}$$

The mathematical equation for the standardized regression coefficient is:

$$b_i = \sum_{j=1}^{p} (r_{j,y})(r_{j,i})^{-1}$$

where $(r_{j,i})^{-1}$ is the inverted matrix value at row j and column i. Here, the intercept b_0 is always equal to zero.

The coefficient of determination, R^2, is found by using the mathematical equation:

$$R^2 = \sum_{i=1}^{p} b_i r_{i,y}$$

The standard error of the standardized regression coefficient, SE b(i), is found by the following mathematical equation:

$$\text{SE } b(i) = \sqrt{\frac{(1 - R^2)(r_{i,y}^{-1})}{(n - p - 1)}}$$

where n is the total number of Y values and p is the number of independent variables.

The non-standardized regression coefficient is defined by the mathematical equation:

$$B_j = (b_j)\sqrt{(\sum_{i=1}^{n} y_i^2)(\sum_{i=1}^{n} x_{ji}^2)}$$

The standard error of the non-standardized regression coefficient is determined by the equation:

$$\text{SE } B_j = \sqrt{\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_{ji}^2}}$$

The intercept is calculated by the equation:

$$B_0 = \bar{Y} - \sum_{j=1}^{p} (B_j)(\bar{X}_j)$$

where $\bar{X}$ and $\bar{Y}$ are means of the X and Y values.

The mathematical equations for the various sums of squares, with their respective degrees of freedom, are as follows.

Total sum of squares:

$$SST = \sum_{i=1}^{n} y_i^2$$

$$dfSST = n - 1$$

Regression sum of squares:

$$SSR = \sum_{j=1}^{p} (b_j r_{j,y})$$

$$dfSSR = p$$

Error sum of squares:

$$SSE = SST - SSR$$

$$dfSSE = n - p - 1$$

The standard error of the estimate (i.e., of the variance about the regression), SE, is found by the mathematical equation:

$$SE = \sqrt{\frac{SSE}{n - p - 1}}$$

The F-statistic is determined by the equation:

$$F = \frac{(SSR)(n - p - 1)}{(SSE)(p)}$$

Remarks

The Multiple Linear Regression Program uses the least squares method to define the straight line of best fit for a given set of data. The least squares method determines the vertical deviation or distance between each data point and the predicted line of best fit. It defines this line of best fit as that which results in the minimum value for the sum of the squares of these distances. This is best known as the deviations' sum of squares or the error sum of squares.

A simple correlation coefficient matrix is the first calculation performed. The correlation coefficient, r, is a set of values without units between -1 and $+1$ (i.e., $-1 \leq r \leq +1$). There is no correlation between variables when the correlation coefficient is zero. If the correlation coefficient is positive, one variable increases as the other variable increases. If the correlation coefficient is negative, one variable decreases as the other variable increases. If the correlation coefficient is significant, the value of the predictive equation for the regression line is important; this may imply a cause and effect relationship between the two variables.

An inversion of the correlation coefficient matrix is performed, and the standardized coefficients and the coefficient of determination are computed. The coefficient of determination measures the proportion of the total variability in the dependent variable, Y, which may be attributed solely to its dependence on the independent variable, X.

The program computes the standard error of the standardized regression, the non-standardized regression coefficient, the standard error of the non-standardized regression coefficient, the total sum of squares, the regression sum of squares, the error sum of squares, and the standard error of the estimate.

To test the null hypothesis that there is no dependence of variable Y on the independent variable X

$$H_0 : B_1 = B_2 = \ldots = P_p = 0$$

the F-statistic is computed.

You may decide to go beyond the scope of this program and test each of the regression coefficients in the multiple linear regression using the Student's t-Test if the Multiple Linear Regression Program determines a significant value for the F-statistic with a numerator degrees of freedom equal to the regression degrees of freedom and a denominator degrees of freedom equal to the error degrees of freedom. To test the null hypothesis

$$H_0 : B_j = 0$$

the Student's t-value may be found using the equation:

$$t = \frac{B_j}{\sqrt{(\Sigma y) C_{ij}}}$$

where C_{ij} is the value at row i and column j in an inverted sum of squares and cross products matrix.

User Information Form

This section presents a listing and tutorial explanation of the Multiple Linear Repression Program display prompts and options.

Step	Display	Response and Comments	Go To
1.	Data Edit Menu Quit?	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	Means and Moments	First program title; press ◆ until "Multiple Linear Reg." appears in LCD	3
3.	Multiple Linear Reg.	Program title; press **END LINE**	4
4.	Output Corr Matrix?	To output the correlation matrix, ri(j), press **Y**	5
		To continue program and not output the correlation matrix, press **N**	7
5.	ri(j)=###############	Elements of correlation matrix, where ri(j) denotes the correlation between Xi and Xj, and where the last variable is always the independent variable: if output is to display, press any key; if output is to printer, no action necessary	6
6.	Repeat Results?	To repeat correlation matrix, press **Y**	5
		To continue program, press **N**	7
7.	Std Regression Coeff?	To compute the standard regression coefficient, b(i), press **Y**	8
		To compute the non-standard regression coefficient, B(i), press **N**	8
8.	R^2=#################	Coefficient of determination: if output is to display, press any key; if output is to printer, no action necessary	9
9.	SST=##################	Total sum of squares; if output is to display, press any key; if output is to printer, no action necessary	10
10.	df/SST=###############	Degrees of freedom for total sum of squares: if output is to display, press any key; if output is to printer, no action necessary	11

Step	Display	Response and Comments	Go To
11.	SSRg=#################	Regression sum of squares: if output is to display, press any key; if output is to printer, no action necessary	12
12.	df/SSRg=##############	Degrees of freedom for regression sum of squares: if output is to display, press any key; if output is to printer, no action necessary	13
13.	SSE=#################	Error sum of squares: if output is to display, press any key; if output is to printer, no action necessary	14
14.	df/SSE=###############	Degrees of freedom for error sum of squares: if output is to display, press any key; if output is to printer, no action necessary	15
15.	SE=##################	Standard error: if output is to display, press any key; if output is to printer, no action necessary	16
16.	F=###################	F-statistic: if output is to display, press any key; if output is to printer, no action necessary	
		If **Y** was selected in Step 7	18
		If **N** was selected in Step 7	17
17.	B(0)=================	Constant for non-standardized regression coefficient: if output is to display, press any key; if output is to printer, no action necessary	21
18.	Repeat Results?	To repeat output results, press **Y**	8
		To continue program, press **N**	
		If **Y** was selected in Step 7	19
		If **N** was selected in Step 7	22
19.	b(i)=================	Standardized regression coefficient: if output is to display, press any key; if output is to printer, no action necessary	20
20.	SEb(i)=##############	Standardized error of the standard regression coefficient: if output is to display, press any key; if output is to printer, no action necessary (NOTE: Steps 19 and 20 are repeated for all of the dependent variables)	21

Step	Display	Response and Comments	Go To
21.	`Repeat Results?`	To repeat output results, press **Y**	8
		To continue program, press **N**	24
22.	`B(i)=#################`	Non-standardized regression coefficient: if output is to display, press any key; if output is to printer, no action necessary	23
23.	`SEB(i)=###############`	Standard error of the non-standard regression coefficient: if output is to display, press any key; if output is to printer, no action necessary (NOTE: Steps 22 and 23 are repeated for all of the dependent variables)	24
24.	`Repeat Results?`	To repeat output results, press **Y**	
		If **Y** was selected in Step 7	19
		If **N** was selected in Step 7	22
		To continue program, press **N**	1

Example

Dr. Cornelius Husker — a Nebraska agronomist better known to his friends and cohorts as "Cornhusker" — has been investigating the effects of plant-available inorganic and organic phosphorus on corn to find the source of the nutrient utilized by the plant. He obtained the following results from 10 samples.

Sample Number	Inorganic Phosphorus	Organic Phosphorus	Plant-available Phosphorus
1	0.4	18	62
2	1.8	23	68
3	3.1	25	72
4	4.7	32	83
5	9.4	39	92
6	11.6	41	95
7	18.1	47	99
8	23.1	50	102
9	26.8	53	106
10	29.9	56	113

Analyze these data using multiple linear regression; find the standard and non-standard regression coefficients. Save the data and print the results. To solve this example, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	Data Edit Menu Quit?
D	Kbd Load Save Print?
K	Grouped Ungrouped?
U	No. of Var.=0
3 **END LINE**	No. of Obs.=0
10 **END LINE**	Edit Parameters?
N	Edit Labels?
Y	X1=X1
InorgP **END LINE**	X2=X2
OrgP **END LINE**	X3=X3
PAP **END LINE**	Edit Labels?
N	InorgP(1)=0
0.4 **END LINE**	InorgP(2)=0
1.8 **END LINE**	InorgP(3)=0
3.1 **END LINE**	InorgP(4)=0
4.7 **END LINE**	InorgP(5)=0
9.4 **END LINE**	InorgP(6)=0
11.6 **END LINE**	InorgP(7)=0
18.1 **END LINE**	InorgP(8)=0
23.1 **END LINE**	InorgP(9)=0
26.8 **END LINE**	InorgP(10)=0
29.9 **END LINE**	OrgP(1)=0
18 **END LINE**	OrgP(2)=0
23 **END LINE**	OrgP(3)=0
25 **END LINE**	OrgP(4)=0
32 **END LINE**	OrgP(5)=0
39 **END LINE**	OrgP(6)=0

Keystrokes	Display
41 **END LINE**	OrgP(7)=0
47 **END LINE**	OrgP(8)=0
50 **END LINE**	OrgP(9)=0
53 **END LINE**	OrgP(10)=0
56 **END LINE**	PAP(1)=0
62 **END LINE**	PAP(2)=0
68 **END LINE**	PAP(3)=0
72 **END LINE**	PAP(4)=0
83 **END LINE**	PAP(5)=0
92 **END LINE**	PAP(6)=0
95 **END LINE**	PAP(7)=0
99 **END LINE**	PAP(8)=0
102 **END LINE**	PAP(9)=0
106 **END LINE**	PAP(10)=0
113 **END LINE**	InorgP(1)=.4
Q	Data Edit Menu Quit?
D	Kbd Save Load Print?
S	SAVE:Filename?_
CORN **END LINE**	Data Edit Menu Quit?
D	Kbd Save Load Print?
P	Data Edit Menu Quit?
M	Means and Moments
▼ or ▲ to find	Multiple Linear Reg.
END LINE	Output Corr Matrix?
Y	(See Figure 4-9 for output) Repeat Results?

Keystrokes	Display
N	Std Regression Coeff?
Y	(See Figure 4-9 for output) Repeat Results?
N	(See Figure 4-9 for output) Repeat Results?
N	Data Edit Menu Quit?
M	Multiple Linear Reg.
END LINE	Output Corr Matrix?
N	Std Regression Coeff?
N	(See Figure 4-9 for output) Repeat Results?
N	(See Figure 4-9 for output) Repeat Results?
N	Data Edit Menu Quit?
Q	Done
END LINE f OFF	

Figure 4-9 presents the results of this example. Each of the three correlation coefficients have a value of 0.9; thus, the variables are all highly correlated with each other. The coefficient of determination, R^2, for the standard regression coefficient and the non-standard regression coefficient is the same value, 0.993792704991. This indicates that 99.4% of the total variability in the dependent variable, Y, may be attributed to the two independent variables, X1 and X2.

Data Format Ungrouped
No. of Var. = 3
No. of Obs. = 10
Variable Labels
X1 = InorgP
X2 = OrgP
X3 = PAP
InorgP
X1(1) = .4
X1(2) = 1.8
X1(3) = 3.1
X1(4) = 4.7
X1(5) = 9.4
X1(6) = 11.6
X1(7) = 18.1
X1(8) = 23.1
X1(9) = 26.8
X1(10) = 29.9
OrgP

X2(1) = 18
X2(2) = 23
X2(3) = 25
X2(4) = 32
X2(5) = 39
X2(6) = 41
X2(7) = 47
X2(8) = 50
X2(9) = 53
X2(10) = 56

PAP

X3(1) = 62
X3(2) = 68
X3(3) = 72
X3(4) = 83
X3(5) = 92
X3(6) = 95
X3(7) = 99
X3(8) = 102
X3(9) = 106
X3(10) = 113

Multiple Linear Reg.
r2(1) = .966474025183
r3(1) = .944227274585
r3(2) = .994666532177

R^2 = .993792704991
SST = 2673.6
df/SST = 9
SSRg = 2657.00417609
df/SSRg = 2
SSE = 16.5958239068
df/SSE = 7
SE = 1.53975062484
F = 560.352692854

b(1) = −.2592540833
SE b(1) = .115975795134
b(2) = 1.2452288696
SE b(2) = .115975795134

R^2 = .993792704991
SST = 2673.6
df/SST = 9
SSRg = 2657.00417609
df/SSRg = 2
SSE = 16.5958239068
df/SSE = 7
SE = 1.53975062484
F = 560.352692854
B(0) = 32.9092487178

B(1) = −.409815352625
SE B(1) = .183328496793
B(2) = 1.60347060358
SE B(2) = .149341043052

Figure 4-9. Printed output for Multiple Linear Regression Program example problem.

Section 5

t-Statistics

This section discusses t-Statistics, which allow you to enter raw input data so as to generate specific statistics. Within t-Statistics, there are two t-Test programs for the HP-71B—the paired t-Test and the unpaired t-Test.

Paired t-Test

The paired t-Test commonly is used to compare or evaluate paired observations in order to calculate a t-Statistic. Unlike the unpaired t-Test (see next program), the paired t-Test assumes that the difference of the two means (d_i) comes from a normally distributed population of differences. Furthermore, the two populations to be tested by the paired t-Test do not necessarily have normal distributions and statistically equivalent variances.

The t-Statistic is useful for comparing the means of different populations. Examples of uses for the t-Statistic include comparison of: two different diets, the effect of two drugs on a test animal, differences between tests, and the quality of two products.

Equations

The t-Statistic is calculated by the equation:

$$t = \frac{\bar{d}\sqrt{n}}{S_{\bar{d}}}$$

where $\bar{d}$ is the mean difference between pairs, S is the standard deviation of the differences, and n is the number of pairs sampled. This distribution has $n - 1$ degrees of freedom. The t-Statistic is used to test the hypothesis that the means of the two sampled populations, X and Y, of n paired observations are statistically equivalent. It tests whether the mean of sample differences between pairs of observations (i.e., X minus Y) is significantly different from a hypothetical mean, which the null hypothesis puts at zero (i.e., H_0: Difference $= 0$).

The mean of the differences between the paired values ($\bar{d}$) of X and Y is calculated by the equation:

$$\bar{d} = \frac{\sum_{i=1}^{n} (x_i - y_i)}{n}$$

The standard deviation of the differences between the points ($S_{\bar{d}}$) is calculated by the equation:

$$S_{\bar{d}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2 - \frac{1}{n}(\sum_{i=1}^{n} (x_i - y_i))^2}{n - 1}}$$

and has n − 1 degrees of freedom.

Remarks

The Paired t-Test Program uses only ungrouped data, and requires one variable which equals the difference of the two variables under analysis (denoted as X and Y in the preceding discussion). After entering the two variables, you must use the compute function of the data editor to compute a variable with the values Xi − Yi.

User Instruction Form

This section presents a listing and brief explanation of the Paired t-Test Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	`Means and Moments`	First program title; press ▼ until "Paired t-Test" appears on the LCD	3
3.	`Paired t-Test`	Program title; press **END LINE**	4
4.	`Diff.Var.#=0`	Enter the variable number which contains the differences (integer, 1 ≤ Diff. Var. # ≤ N, the # of Variables in the data base); then, press **END LINE**	5
5.	`Edit Parameters?`	To edit input parameters, press **Y** and then **END LINE**.	4
		To continue program without editing, press **N** and then **END LINE**	6
6.	`#Data=##############`	Number of observations for X and Y: if output is to display, press any key; if output is to printer, no action necessary	7
7.	`Mean=################`	Arithmetic mean for X minus Y: if output is to display, press any key; if output is to printer, no action necessary	8

Step	Display	Response and Comments	Go To
8.	`StdDev=##############`	Standard deviation for X minus Y: if output is to display, press any key; if output is to printer, no action necessary	9
9.	`df=######`	Degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	10
10.	`t=###################`	t-statistic: if output is to display, press any key; if output is to printer, no action necessary	11
11.	`SE=##################`	Standard error: If output is to display, press any key; if output is to printer, no action necessary	12
12.	`Repeat Results?`	To repeat output results, press **Y**	6
		To continue program, press **N**	1

Example

Dr. Oliver Virgil Waite, a research physician (known to his close associates as "O. Vir. Waite"), is conducting an experiment to compare two different diets for humans. Dr. Waite has found 10 sets of overweight twins, all the same age, for testing the two diets. He randomly selects one twin from each pair and places them on Diet A for a four month period. The remaining twins are placed on Diet B for the same four month period. At the conclusion of the experiment, Dr. Waite weighs each twin to see how much weight each has lost. The results are presented below:

	Twins									
Diet	1	2	3	4	5	6	7	8	9	10
A	5.6	3.7	4.8	5.4	3.5	3.4	3.6	4.2	4.9	5.7
B	9.2	7.1	7.9	9.5	8.0	7.6	7.3	8.1	8.3	9.8

Using the t-statistic, evaluate these paired observations to test the hypotheses:

$$H_0 = \mu_1 = \mu_2 \text{ and } H_A = \mu_1 \neq \mu_2$$

where μ_1 and μ_2 are the means of the twin's weight loss scores for the two diets. Enter the data as pairs via the keyboard, label the variables, save the data, and print the input data and the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
K	`Grouped Ungrouped?`
U	`No. of Var.=0`
2 **END LINE**	`No. of Obs.=0`
10 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
Y	`X1=X1`
DIET A **END LINE**	`X2=X2`
DIET B **END LINE**	`Edit Labels?`
N	`DIET A(1)=0`
S	`Direction:S,X,D or Q?`
D	`DIET A(1)=0`
5.6 **END LINE**	`DIET B(1)=0`
9.2 **END LINE**	`DIET A(2)=0`
3.7 **END LINE**	`DIET B(2)=0`
7.1 **END LINE** • • •	NOTE: The program continues to step through each of the 20 elements of the entire array. Continue to enter the values from the example as the program calls for them; press **END LINE** after each entry. After all 20 elements have been entered, the editor wraps back to the first element.
9.8 **END LINE**	`DIET A(1)=5.6`
O	`ADD:Var. or Obs.?`
V	`Add Var. at?3`
END LINE	`Label=NEW`
DIFF **END LINE**	`DIFF(1)=0`
U	`DEF:Sequence Compute?`
C	`Var. Number?3`

Keystrokes	Display
END LINE	`Compute X3=X3`
X1 − X2 **END LINE**	`DIFF(1)=-3.6`
END LINE	`DIET A(2)=3.7`
END LINE	`DIET B(2)=7.1`
END LINE	`DIFF(2)=-3.4`
END LINE • • •	NOTE: Because we selected the **D** direction for the editor, the values are shown across the rows (i.e., observations). The program continues to step through each of the elements of the entire array. Continue through the values from the example as the program calls for them; press **END LINE** after each appears. After all elements have been viewed, the editor wraps back to the first element.
END LINE	`DIET A(1)=5.6`
Q	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
S	`SAVE:Filename?_`
DIETS **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
P	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Paired t-Test`
END LINE	`Diff.Var.#=0`
3 **END LINE**	`Edit Parameters?`
N	(See Figure 5-1 for printed output.) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	Done
END LINE **f** **OFF**	

The results are shown in Figure 5-1. There is a difference in the means of the two diets; with df of 9 and t value of −17.1, which is significant at the 0.001 level.

Data Format Ungrouped

No. of Var. = 2
No. of Obs. = 10

Variable Labels

X1 = DIET A
X2 = DIET B
X3 = DIFF

DIET A

X1(1) = 5.6
X1(2) = 3.7
X1(3) = 4.8
X1(4) = 5.4
X1(5) = 3.5
X1(6) = 3.4
X1(7) = 3.6
X1(8) = 4.2
X1(9) = 4.9
X1(10) = 5.7

DIET B

X2(1) = 9.2
X2(2) = 7.1
X2(3) = 7.9
X2(4) = 9.5
X2(5) = 8
X2(6) = 7.6
X2(7) = 9.3
X2(8) = 8.1
X2(9) = 8.3
X2(10) = 9.8

DIFF

X3(1) = −3.6
X3(2) = −3.4
X3(3) = −3.1
X3(4) = −4.1
X3(5) = −4.5
X3(6) = −4.2
X3(7) = −5.7
X3(8) = −3.9
X3(9) = −3.4
X3(10) = −4.1

Paired t-Test

Diff. Var. # = 3

# Data = 10
Mean = −4
Std Dev = .737864787372
df = 9
t = −17.1428571429
SE = .233333333333

Figure 5-1. Printed results of Paired t-Test Program example.

Unpaired t-Test

The unpaired t-Test, the second of the t-Test programs, is used to determine significant differences between population distributions when there is no correlation between the two populations. The unpaired t-Test commonly is used to compare or evaluate the means of two different populations — especially if the two have normal distributions and statistically equivalent variances.

The t-Statistic is useful for comparing the means of different populations. Examples of uses for the t-Statistic include: comparison of two different weight control diets, comparison of the effect of two drugs on a test subject, comparison of educational tests, and comparison of products.

Equations

For the unpaired t-Test, the t-Statistic is calculated by the equation:

$$t = \frac{\bar{x} - \bar{y} - d}{S_{\bar{x}-\bar{y}}}$$

In this equation, $\bar{x} - \bar{y}$ is equal to the difference between the means, d is the hypothetical diffrence, and $S_{\bar{x}-\bar{y}}$ is the standard error of the means.

Standard error, using the pooled variance (S_p^2),is calculated by the equation:

$$S_{x-y} = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} * \sqrt{S_p^2}$$

where n_1 is equal to the number of observations of the X variable, and n_2 is equal to the number of observations of the Y variable.

The pooled variance (S_p^2) is calculated by the equation:

$$S_p^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n_1} + \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n_2}}{n_1 + n_2 - 2}$$

The t-Statistic is compared with the critical value of t, using $n_1 + n_2 - 2$ degrees of freedom.

Please note that the hypothesized difference is X minus Y.

Remarks

The Unpaired t-Test Program uses both ungrouped and grouped data. These data require two variables.

User Information Form

This section presents a listing and brief tutorial explanation of the Unpaired t-Test Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	Data Edit Menu Quit?	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	Means and Moments	First program title; press ▼ until "Unpaired t-Test" appears on LCD	3
3.	Unpaired t-Test	Program title; press **END LINE**	4
4.	Hyp. Diff=0	Enter hypothesized difference between the means of the two samples (any number, $-\infty \leqslant$ Hyp. Diff $\leqslant \infty$); then, press **END LINE**	5
5.	X Var. #=0	Enter number of variable selected as X (number selected must be equal to or less than number of variables entered under the data option); then, press **END LINE**	6
6.	Y Var. #=0	Enter number of variable selected as Y (number selected must be equal to or less than number of variables entered under the data option); then, press **END LINE**	7
7.	Edit Parameters?	To edit the values entered for the above special parameters, press **Y**	4
		To accept the values entered for the above special parameters, press **N**	8
8.	X#Data=##############	Number of observations for X: if output is to display, press any key; if output is to printer, no action necessary	9
9.	XMean=################	Arithmetic mean for X: if output is to display, press any key; if output is to printer, no action necessary	10
10.	XStdDev=#############	Standard deviation for X: if output is to display, press any key; if output is to printer, no action necessary	11
11.	Y#Data=##############	Number of observations for Y; if output is to display, press any key; if output is to printer, no action necessary	12

Step	Display	Response and Comments	Go To
12.	YMean=################	Arithmetic mean for Y: if output is to display, press any key; if output is to printer, no action necessary	13
13.	YStdDev=#############	Standard deviation for Y: if output is to display, press any key; if output is to printer, no action necessary	14
14.	df=######	Degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	15
15.	t=####################	t-statistic: if output is to display, press any key; if output is to printer, no action necessary	16
16.	PoolVar.=############	Pooled variance: if output is to display, press any key; if output is to printer, no action necessary	17
17.	SE=###################	Standard error: If output is to display, press any key; if output is to printer, no action necessary	18
18.	Repeat Results?	To repeat output results, press **Y**	8
		To continue program, press **N**	1

Example

Beatrice Honeycutt (known to her close friends as "Bea Honey") keeps honeybees on her father's farm and sells honey to help pay for her college tuition. In all, Bea has 30 beehives, all the same size, located 10 each in three different types of fields. This season, Bea's bees yielded the following pounds of honey per beehive for the three fields.

Red Clover	Alfalfa	Vetch
48.5	46.2	51.3
47.9	47.4	52.7
46.3	44.9	50.9
45.4	43.2	53.0
44.3	40.7	51.8
48.9	45.2	58.4
47.7	42.6	55.3
46.1	41.9	54.5
45.4	43.8	53.7
46.9	49.4	52.6

Is there a difference in the production between Red Clover and Alfalfa? Between Red Clover and Vetch? Between Alfalfa and Vetch?

The hypotheses for each test are:

$$H_0: \mu_1 = \mu_2 \text{ and } H_A: \mu_1 \neq \mu_2$$

Enter these data, label the variables by their respective crops, save the data onto a file named BEAHONEY, print the input data, and print the results. To solve this example problem, use the following keystrokes.

Keystrokes	**Display**
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
K	`Grouped Ungrouped?`
U	`No. of Var.=0`
3 **END LINE**	`No. of Obs.=0`
10 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
Y	`X1=X1`
RED CLOVER **END LINE**	`X2=X2`
ALFALFA **END LINE**	`X3=X3`
VETCH **END LINE**	`Edit Labels?`
N	`RED CLOVER(1)=0`
48.5 **END LINE**	`RED CLOVER(2)=0`
47.9 **END LINE** • • •	NOTE: The program continues to step through each of the 30 elements of the entire array. Continue to enter the values from the example as the program calls for them; press **END LINE** after each entry. After all 30 elements have been entered, the editor wraps back to the first element.
52.6 **END LINE**	`RED CLOVER(1)=48.5`
Q	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
S	`SAVE:Filename?_`
BEAHONEY **END LINE**	`Data Edit Menu Quit?`

Keystrokes	Display
D	`Kbd Save Load Print?`
P	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Unpaired t-Test`
END LINE	`Hyp. Diff=0`
END LINE	`X Var. #=0`
1 **END LINE**	`Y Var. #=0`
2 **END LINE**	`Edit Parameters?`
N	(See Figure 5-2 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
M	`Unpaired t-Test`
END LINE	`Hyp. Diff=0`
END LINE	`X Var. #=0`
1 **END LINE**	`Y Var. #=0`
3 **END LINE**	`Edit Parameters?`
N	(See Figure 5-2 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
M	`Unpaired t-Test`
END LINE	`Hyp. Diff=0`
END LINE	`X Var. #=0`
2 **END LINE**	`Y Var. #=0`
3 **END LINE**	`Edit Parameters?`
N	(See Figure 5-2 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are presented in Figure 5-2. There is a difference in honey production between Red Clover and Alfalfa (significant at the 0.05 level), between Red Clover and Vetch (significant at the 0.001 level), and between Alfalfa and Vetch (significant at the 0.001 level).

Data Format Ungrouped

No. of Var. = 3
No. of Obs. = 10

Variable Labels

X1 = RED CLOVER
X2 = ALFALFA
X3 = VETCH

RED CLOVER

X1(1) = 48.5
X1(2) = 47.9
X1(3) = 46.3
X1(4) = 45.4
X1(5) = 44.3
X1(6) = 48.9
X1(7) = 47.7
X1(8) = 46.1
X1(9) = 45.4
X1(10) = 46.9

ALFALFA

X2(1) = 46.2
X2(2) = 47.4
X2(3) = 44.9
X2(4) = 43.2
X2(5) = 40.7
X2(6) = 45.2
X2(7) = 42.6
X2(8) = 41.9
X2(9) = 43.8
X2(10) = 49.4

VETCH

X3(1) = 51.3
X3(2) = 52.7
X3(3) = 50.9
X3(4) = 53
X3(5) = 51.8
X3(6) = 58.4
X3(7) = 55.3
X3(8) = 54.5
X3(9) = 53.7
X3(10) = 52.6

Unpaired t-Test

Hyp. Diff = 0
X Var. # = 1
X Var. # = 2

X# Data = 10
XMean = 46.74
XStd Dev = 1.49829532766
Y# Data = 10
YMean = 44.53
YStd Dev = 2.64451213732
df = 18
t = 2.29929904841
Pool Var. = 4.61916666667
SE = .961162490599

Unpaired t-Test

Hyp. Diff = 0
X Var. # = 1
Y Var. # = 3

X# Data = 10
XMean = 46.74
XStd Dev = 1.49829532766
Y# Data = 10
YMean = 53.42
YStd Dev = 2.2215109973
df = 18
t = −7.88341229513
Pool Var. = 3.59
SE = .847348806573

Unpaired t-Test

Hyp. Diff = 0
X Var. # = 2
Y Vay. # = 3

X# Data = 10
XMean = 44.53
XStd Dev = 2.64451213732
Y# Data = 10
YMean = 53.42
YStd Dev = 2.2215109973
df = 18
t = −8.13968938303
Pool Var. = 5.96427777778
SE = 1.09217926897

Figure 5-2. Printed output from the Unpaired t-Test Program example.

Section 6

F-Statistics

The Student's t-Test, used in Section 5, is no longer applicable when trying to compare the means of more than one population. In its place, you would use the analysis of variance, whose acronym is ANOVA, as a standard statistical method for testing the hypothesis that a number of populations have equal means. In the analysis of variance, the mean of each sample is used to estimate the variance of the populations. This estimate of the total population variance is then compared to the estimate of the population variance derived from differences between individual elements in the sample. The F-distribution is used for the actual test (see Section 9).

There are two basic assumptions in using ANOVA where tests of significance are made. These are as follows.

- The sum of the treatment effects is zero. The difference between the mean of the sample of the population (or treatment) and the grand mean (the average of the sample mean) is a treatment effect.

- Experimental errors are random, are independently and normally distributed about a zero mean, and have a common variance.

One-way Analysis of Variance

The simplest type of ANOVA is the single-classification analysis of variance, often called the one-way ANOVA. This ANOVA assumes that the variance related to the treatment of the population is determined by a single factor.

There can be any number of sample (or treatment) groups when using the one-way ANOVA. Each sample group, j, is composed of any number of observations, X_j. Here, j equals 1, 2, . . . , n, n is the number of groups, i equals 1, 2, . . . , m_j, and m_j is the number of observations in group j. The sample groups may have an unequal number of observations. The sum of squares among groups is used by the program to estimate the variance from the sample means. The F-statistic is calculated from several measures of variance and degrees of freedom. These are the error sum of squares, error degrees of freedom, groups sum of squares, groups degrees of freedom, total sum of squares, total degrees of freedom, groups mean square deviation, and error mean square deviation. The null hypothesis

$$H_0: \overline{X}_1 = \overline{X}_2 = \ldots = \overline{X}_n$$

may be rejected if the calculated F value is equal to or greater than the critical value of F.

Equations

For the one-way ANOVA, the total number of observations, N, is calculated by following equation:

$$N = \sum_{j=1}^{n} m_j$$

The equation for the error sum of squares, SS_{error}, is as follows:

$$SS_{error} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} (X_{ji} - \bar{X}_j)^2 \right)$$

The equation for the error degrees of freedom ,df_{error}, is as follows:

$$df_{error} = N - n$$

The equation for the groups sum of squares, SS_{groups}, is as follows:

$$SS_{groups} = \sum_{j=1}^{n} m_j(\bar{X}_j - \bar{X})^2$$

where $\bar{X}$ is the grand mean.

The equation for the grand mean is:

$$\bar{X} = \frac{\sum_{j=1}^{n} \sum_{i=1}^{m_j} X_{ji}}{N}$$

The equation for the groups degrees of freedom, df_{groups}, is:

$$df_{groups} = n - 1$$

The equation for the total sum of squares, SS_{total}, is as follows:

$$SS_{total} = \sum_{j=1}^{n} \left(\sum_{i=1}^{m_j} (X_{ji} - \bar{X})^2 \right)$$

The equation for the total degrees of freedom, df_{total}, is:

$$df_{total} = N - 1$$

The equation for the groups mean square deviation, MS_{groups}, is:

$$MS_{groups} = \frac{SS_{groups}}{df_{groups}}$$

The equation for the error mean square deviation, MS_{error}, is:

$$MS_{error} = \frac{SS_{error}}{df_{error}}$$

The equation for the F-statistic, F, for the one-way ANOVA is:

$$F = \frac{MS_{groups}}{MS_{error}}$$

Remarks

This program uses ungrouped and grouped data. These data require two or more variables; the variables may have an unequal number of observations.

The data for the One-way Analysis of Variance Program should be set up so that each variable contains all the observations for one treatment. As an example, if the treatment is temperature, with four settings (e.g., 110, 120, 130, and 140 degrees), then variables X1, X2, X3, and X4 contain all observations for the 110, 120, 130, and 140 degree settings, respectively.

User Information Form

This section presents a listing and brief tutorial explanation of the One-way ANOVA Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	`Means and Moments`	First program title; press ▼ until "One-way ANOVA" appears on the LCD	3
3.	`One-way ANOVA`	Program title; press **END LINE**	4
4.	`# of Variables=0`	Enter the number of variables to be tested (integer, 2 ≤ # of Variables ≤∞; press **END LINE**	5

Step	Display	Response and Comments	Go To
5.	Edit Parameters?	To edit the above control prompt, press **Y**	4
		To continue program without editing, press **N**	6
6.	Var.#1=0	Enter the number of the variable which is to be used as variable number one in the test; then, press **END LINE**. (Note that the program repeats this prompt for all variables.)	7
7.	Edit Parameters?	To edit the above control prompts, press **Y**	6
		To continue program without editing, press **N**	8
8.	SST=##################	Total sum of squares: if output is to display, press any key; if output is to printer, no action necessary	9
9.	dfT=##################	Total degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	10
10.	SSW=##################	Within (row) groups sum of squares: if output is to display, press any key; if output is to printer, no action necessary	11
11.	dfW=##################	Within (row) groups degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	12
12.	MSW=#################	Within (row) groups mean square: if output is to display, press any key; if output is to printer, no action necessary	13
13.	SSE=##################	Error sum of squares: if output is to display, press any key; if output is to printer, no action necessary	14
14.	dfE=######	Error degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	15
15.	MSE=#################	Error mean square: if output is to display, press any key; if output is to printer, no action necessary	16

Step	Display	Response and Comments	Go To
16.	`F=####################`	F-statistic: if output is to display, press any key; if output is to printer, no action necessary	17
17.	`Repeat Results?`	To repeat output results, press **Y**	8
		To continue program, press **N**	1

Example

Using the One-way Analysis of Variance Program, analyze the example problem used in the Paired t-Test Program. The hypotheses are:

$$H_O : \mu_1 = \mu_2 \text{ and } H_A : \mu_1 \neq \mu_2$$

where μ_1 and μ_2 are the means of the twin's weight loss scores for the two diets. Enter the data by loading the file saved in the Paired t-Test Program; then, print the input data and the results.

To solve this example problem, use the following keystrokes. The file DIETS refers to the data saved in sample problem for the Paired t-Test Program presented earlier.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
L	`LOAD:Filename?_`
DIETS **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
P	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`One-way ANOVA`
END LINE	`# of Variables=0`
2 **END LINE**	`Edit Parameters?`
N	`Var.#1=0`
1 **END LINE**	`Var.#2=0`

Keystrokes	Display
2 **END LINE**	`Edit Parameters?`
N	(See Figure 6-1 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are presented in Figure 6-1. At df 1 and 18, the F value of 96.96 is significant at the 0.10 level. There is a difference in the means of the two diets.

DIET A

X1(1) = 5.6
X1(2) = 3.7
X1(3) = 4.8
X1(4) = 5.4
X1(5) = 3.5
X1(6) = 3.4
X1(7) = 3.6
X1(8) = 4.2
X1(9) = 4.9
X1(10) = 5.7

DIET B

X2(1) = 9.2
X2(2) = 7.1
X2(3) = 7.9
X2(4) = 9.5
X2(5) = 8
X2(6) = 7.6
X2(7) = 9.3
X2(8) = 8.1
X2(9) = 8.3
X2(10) = 9.8

DIFF

X3(1) = −3.6
X3(2) = −3.4
X3(3) = −3.1
X3(4) = −4.1
X3(5) = −4.5
X3(6) = −4.2
X3(7) = −5.7
X3(8) = −3.9
X3(9) = −3.4
X3(10) = −4.1

One-way ANOVA

# of Variables = 2
Var. #1 = 1
Var. #2 = 2
SST = 94.852
dfT = 19
SSW = 80
dfW = 1
MSW = 80
SSE = 14.852
dfE = 18
MSE = .825111111111
F = 96.9566388365

Figure 6-1. Printed results for the One-way ANOVA Program example problem.

Two-way Analysis of Variance

The two-way analysis of variance — often shortened to two-way ANOVA — is used when two separate treatments (or factors) may be causing the variance between the means. In this two-way ANOVA, replicate and non-replicate observations are possible for each group.

For the two-way ANOVA, one treatment, or factor, will be designated as A, the other treatment designated as B, the number of replicates as r, the number of treatment (or factor) A as a, and the number of treatment B as b. In the following formulas, each variable will contain a triple subscript (e.g., X_{ijk}). For the two-way ANOVA without replication, the replicate is treated as a value of one.

The two-way analysis of variance allows the user to test for three hypotheses. These are as follows:

H_{0_A}: Treatment A has no effect.

H_{0_B}: Treatment B has no effect.

$H_{0_{AB}}$: Interaction of Treatment A and B has no effect.

The two-way ANOVA with replication allows the user to test for all three of these hypotheses; the two-way ANOVA without replication usually allows the user to test for only the first two hypotheses. Although it is common to have a two-way ANOVA without replication, it is usually advantageous to have replicate data for each combination of treatments; these replicates constitute a cell containing only a single reading.

A common use of the two-way ANOVA without replication is the repeated testing of the same individuals. In other words, you may have a group of individuals to be tested repeatedly over a period of time. In this situation, one factor would be the individuals, while the other factor would be the time dimension (i.e., the fixed treatment effect). "Repeated testing" over time, as in this example, is more often a treatment effect, rather than a true statistical replication.

Equations

The equations for the various degrees of freedom needed for the Two-way Analysis of Variance Program are:

$$df_{total} = N - 1$$

$$df_{cells} = ab - 1$$

$$df_{Treatment\ A} = a - 1$$

$$df_{Treatment\ B} = b - 1$$

$$df_{AB} = (a - 1)(b - 1)$$

$$df_{error} = N - ab$$

The equation for the correction term, CT, is as follows:

$$CT = \frac{\left(\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}\right)^2}{N}$$

The equation for the total sample number, N, is as follows:

$$N = a\ b\ r$$

The equation for the total sum of squares, SS_{total}, is:

$$SS_{total} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}^2 - CT$$

The equation for calculating the cells sum of squares, SS_{cells}, is as follows:

$$SS_{cells} = \frac{1}{r} \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\sum_{k=1}^{r} X_{ijk}\right)^2 - CT$$

The equation for the error sum of squares, SS_{error}, is:

$$SS_{error} = SS_{total} - SS_{cells}$$

The equation for the treatment A sum of squares, $SS_{treatment\ A}$, is as follows:

$$SS_{treatment\ A} = \frac{1}{b\ r} \sum_{i=1}^{a} \left(\sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}\right)^2 - CT$$

The equation for the treatment B sum of squares, $SS_{treatment\ B}$, is as follows:

$$SS_{treatment\ B} = \frac{1}{a\ r} \sum_{j=1}^{b} \left(\sum_{i=1}^{a} \sum_{k=1}^{r} X_{ijk}\right)^2 - CT$$

The equation for the A $\times$ B interaction sums of squares, SS_{AB}, is:

$$SS_{AB} = SS_{cells} - SS_{treatment} - SS_{treatment\ B}$$

The equation for the treatment A mean square, $MS_{treatment\ A}$, is as follows:

$$MS_{treatment\ A} = \frac{SS_{treatment\ A}}{df_{treatment\ A}}$$

The equation for the treatment B mean square, $MS_{\text{treatment B}}$, is as follows:

$$MS_{\text{treatment B}} = \frac{SS_{\text{treatment B}}}{df_{\text{treatment B}}}$$

The equation for the A $\times$ B interaction mean square, MS_{AB}, is:

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$$

The equation for the error mean square, MS_{error}, is as follows:

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{df_{\text{error}}}$$

Finally, the equations for calculating the F-statistics for the three hypotheses are as follows:

$$H_{0_A}: \; F = \frac{MS_{\text{treatment A}}}{MS_{\text{error}}}$$

$$H_{0_B}: \; F = \frac{MS_{\text{treatment B}}}{MS_{\text{error}}}$$

$$H_{0_{AB}}: F = \frac{MS_{AB}}{MS_{\text{error}}}$$

Remarks

The Two-way Analysis of Variance Program uses the sum of squares among groups to estimate the variance from the sample means. The F-statistic is computed from several measures of variance and degrees of freedom. The initial hypothesis may be rejected if the calculated F-statistic is equal to or greater than the critical value of F.

The two-way ANOVA without replication uses computations that are similar to those for the two-way ANOVA with replication, except that the expressions to be evaluated are much simpler. Since replication equals one, much of the summation may be omitted. Here, the subgroup sum of squares is the same as the total sum of squares. The error sum of squares based upon variation within a treatment is absent. After subtracting the sum of squares for columns (factor or treatment A) and the sum of squares for rows (factor or treatment B) from the total sum of squares, you are left with only a single sum of squares. This is equivalent to the interaction sum of squares in the two-way ANOVA with replication, but which now is the only source for an error. This sum of squares also is called the remainder sum of squares or the discrepant. Since there is no interaction in the two-way ANOVA without replication, the row and column mean squares are tested over the error mean square.

For the two-way analysis of variance, the input data must be ungrouped. The input data should be set up differently, however, depending on whether they are replicated or non-replicated.

The Two-way Analysis of Variance Program requires four or more variables with an equal number of observations in each variable. In the case of non-replicated input data (i.e., data containing one replicate), each variable contains one observation. Replicated input data (i.e., data containing "r" replicates) have two or more observations for each variable. This is to say that X1 represents the replication class at "row 1, column 1," composed of "r" replicates.

There is a relationship in this program between the number of variables and the number of "rows" and "columns." Picture the grouping of raw data for the ANOVA table as a three-dimensional cube. The sides of the cube represent the number of different A (i.e., width) and B (i.e., height) factors, while the depth of the cube represents the number of replicates. In this visualization, there is a fixed number of A factors for each B factor and vice versa. The request for the number of rows is asking, in essence, for the number of A factors. Of course, what you call an A factor and a B factor is entirely arbitrary.

What all this means to the user is that for a given number of A factors, a, and B factors, b, with r replicates, the data file will be composed of a $\times$ b variables with r observations each. The program, using the "number of rows" data, determines exactly how many A factors and B factors there are in the raw data set. The program takes the first "a" variables as row 1, the second "a" variables as row 2, et cetera. Thus, although the labeling of factor A and B is simply dependent on the user's application, the number of "rows" is important for accurate and expected results.

To illustrate this discussion, the following table presents an example of a two-way ANOVA with replication.

Table 6-1. Anova Table; A factors = 3, B factors = 2, Replicates = 4

		A factors		
B factors	Row 1	1, 2, 3, 4	5, 6, 7, 8	9, 10, 11, 12
	Row 2	13, 14, 15, 16	17, 18, 19, 20	21, 22, 23, 24

The data file would look like the following:

	Variables					
	X1	X2	X3	X4	X5	X6
Observations	1	5	9	13	17	21
	2	6	10	14	18	22
	3	7	11	15	19	23
	4	8	12	16	20	24
		Row 1			Row 2	

User Information Form

This section presents a listing and brief tutorial explanation of the Two-way Analysis of Variance Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	Data Edit Menu Quit?	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	Means and Moments	First program title; press ▼ until "Two-way ANOVA" appears on the LCD	3
3.	Two-way ANOVA	Program title; press **END LINE**	4
4.	# of Rows	Enter number of rows (integer, 2 ≤ # of Rows < N, the # of Variables in the data base); then, press **END LINE**	5
5.	Edit Parameters?	To edit parameter, press **Y**	4
		To continue program without editing, press **N**	6
6.	SST=##################	Total sum of squares: if output is to display, press any key; if output is to printer, no action necessary	7
7.	dfT=##################	Total degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	8
8.	SSE=##################	Error sum of squares: if output is to display, press any key; if output is to printer, no action necessary	9

Step	Display	Response and Comments	Go To
9.	dfE=######	Error degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	10
10.	MSE=##################	Error mean square: if output is to display, press any key; if output is to printer, no action necessary	11
11.	SSW=##################	Factor A row sum of squares: if output is to display, press any key; if output is to printer, no action necessary	12
12.	dfW=##################	Factor A row degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	13
13.	MSW=#################	Factor A row mean square: if output is to display, press any key; if output is to printer, no action necessary	14
14.	F_W=###################	F-statistic for Factor A: if output is to display, press any key; if output is to printer, no action necessary	15
15.	SSC=##################	Factor B column sum of squares: if output is to display, press any key; if output is to printer, no action necessary	16
16.	dfC=##################	Factor B column degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	17
17.	MSC=#################	Factor B column mean square: if output is to display, press any key; if output is to printer, no action necessary	18
18.	F_C=###################	F-statistic for Factor B: if output is to display, press any key; if output is to printer, no action necessary	
		If non-replicated data	23
		If replicated data	19
19.	SSI=##################	A × B interaction sum of squares: if output is to display, press any key; if output is to printer, no action necessary	20

Step	Display	Response and Comments	Go To
20.	dfI=###################	A×B interaction degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	21
21.	MSI=###################	A×B interaction mean square: if output is to display, press any key; if output is to printer, no action necessary	22
22.	Fi=####################	F-statistic for interaction: if output is to display, press any key; if output is to printer, no action necessary	23
23.	Repeat Results?	To repeat output results, press **Y**	6
		To continue program, press **N**	1

Example

Phil Lum, a research chemist who works for a manufacturer of photographic film, customarily mixes the chemical ingredients for the film emulsion and allows them to set for a certain time and temperature for normal film processing. As a test, he wishes to know whether there is a significant interaction between time and temperature for a particular set of ingredients. Since this problem involves two factors (i.e., time and temperature) plus a possible interaction, he replicates the experiment to obtain a separate estimate of error. Phil Lum's film emulsion test results are presented in the following table as a percentage change from the normal processing time of the film. Note that there are really four variables with two observations each.

	Temperature (F)	
Time (Hours)	**60**	**70**
1	15	−15
	25	−30
2	5	55
	45	45

Using the two-way ANOVA, evaluate these ungrouped, replicated observations. Test the following three null hypotheses at the 95% confidence limit.

- There is no difference between the two time periods (i.e, the changes in time of the holding mixtures produce no effect).

$$H_A\colon \mu_1 = \mu_2$$

- There is no difference between 60 and 70 degrees F (i.e., the changes in temperature produce no effect).

$$H_B\colon \mu_1 = \mu_2$$

- There is no interaction between the times and the temperatures (i.e., there is no interaction of time and temperature).

$$H_{AB}\colon \mu_1 = \mu_2$$

where μ_1 and μ_2 are the means of the two sets of data. Enter the data via the keyboard, label the variables, save the data, and print the input data and the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
K	`Grouped Ungrouped?`
U	`No. of Var.=0`
4 **END LINE**	`No. of Obs.=0`
2 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
Y	`X1=X1`
1HR/60F **END LINE**	`X2=X2`
1HR/70F **END LINE**	`X3=X3`
2HR/60F **END LINE**	`X4=X4`
2HR/70F **END LINE**	`Edit Labels?`
N	`1HR/60F(1)=0`
15 **END LINE**	`1HR/60F(2)=0`
25 **END LINE**	`1HR/70F(1)=0`
−15 **END LINE**	`1HR/70F(2)=0`
−30 **END LINE**	`2HR/60F(1)=0`

Keystrokes	Display
5 **END LINE**	2HR/60F(2)=0
45 **END LINE**	2HR/70F(1)=0
55 **END LINE**	2HR/70F(2)=0
45 **END LINE**	1HR/60F(1)=15
Q	Data Edit Menu Quit?
D	Kbd Save Load Print?
S	SAVE:Filename?_
FILMTEST **END LINE**	Data Edit Menu Quit?
D	Kbd Save Load Print?
P	Data Edit Menu Quit?
M	Means and Moments
▼ or ▲ to find	Two-way ANOVA
END LINE	#of Rows
2 **END LINE**	Edit Parameters?
N	(See Figure 6-2 for output) Repeat Results?
N	Data Edit Menu Quit?
Q	Done
END LINE f OFF	

Results are presented in Figure 6-2. By using the F-distribution Program, it can be shown that F (with df at 1 and 4) equals 7.71 at the 95% confidence level. Thus, Phil Lum rejects the null hypothesis for time and interaction since their F values are greater than 7.71. The F value for the temperature hypothesis is 0.605; therefore, he does not reject this hypothesis.

From this, Phil Lum concludes that the film emulsion experiment: (1) does not show an effect associated with a change in temperature, and (2) is significantly affected by a change in time. However, because the interaction between time and temperature is significant, a change in temperature may affect the potency of the time factor.

Data Format Ungrouped

No. of Var. = 4
No. of Obs. = 2

Variable Labels

X1 = 1HR/60F
X2 = 1HR/70F
X3 = 2HR/60F
X4 = 2HR/70F

1HR/60F

X1(1) = 15
X1(2) = 25

1HR/70F

X2(1) = −15
X2(2) = −30

2HR/60F

X3(1) = 5
X3(2) = 45

2HR/70F

X4(1) = 55
X4(2) = 45

Two-way ANOVA

# of Rows = 2

SST = 6446.875
dfT = 7
SSE = 1012.5
dfE = 4
MSE = 253.125
SSW = 3003.125
dfW = 1
MSW = 3003.125
Fw = 11.8641975309
SSC = 153.125
dfC = 1
MSC = 153.125
Fc = .604938271605
SSI = 2278.125
dfI = 1
MSI = 2278.125
Fi = 9

Figure 6-2. Printed results of Two-way ANOVA Program example problem.

Section 7

Chi-square Statistics

Contingency Table

When observed frequencies occupy a single row, the resulting table is called a one-way classification table. When observed frequencies occupy several rows and several columns, the resulting table is called a two-way classification table, which is better known as a contingency table. An example of a theoretical contingency table is presented in Table 7-1.

Table 7-1. Theoretical contingency table

Row	Column 1	2	•	•	•	n	Total
1	f_{11}	f_{21}	•	•	•	f_{n1}	$\sum_{j=1}^{n} f_{j1}$
2	f_{12}	f_{22}	•	•	•	f_{n2}	$\sum_{j=1}^{n} f_{j2}$
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•
m	f_{1m}	f_{2m}	•	•	•	f_{nm}	$\sum_{j=1}^{n} f_{jm}$
Total	$\sum_{i=1}^{m} f_{1i}$	$\sum_{i=1}^{m} f_{2i}$	•	•	•	$\sum_{i=1}^{m} f_{ni}$	$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ji}$

Contingency tables are used to test the null hypothesis that the frequencies of one variable are independent of those of the second variable. Data should be entered one column at a time.

Equations

Where r is the number of rows and c is the number of columns, the expected frequencies (F_{ij}) of a contingency table containing r $\times$ c cells are calculated using:

$$F_{ji} = \frac{(R_i)(C_j)}{N}$$

where R_i = the sum of frequencies for a row;

C_j = the sum of frequencies for a column; and

$$N = \sum_{i=1}^{r} \sum_{j=1}^{c} f_{ji}.$$

The expected contingency table values are calculated and compared with the actual values. The null hypothesis — that one variable is independent of another — is tested by the chi-square statistic χ^2. This statistic is calculated by:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - F_{ij})^2}{F_{ij}}$$

where f is the observed frequency.

The degrees of freedom, df, are calculated by the formula:

$$df = (r - 1)(c - 1)$$

The null hypothesis may be rejected if the calculated chi-square value is greater than the critical value.

Remarks

The Contingency Table Program uses only ungrouped data. These data require two or more variables with an equal number of observations. All data entered in the contingency table must be greater than or equal to ($\geq$) zero. Also, you cannot enter a negative number.

User Instruction Form

This section presents a listing and brief tutorial explanation of the Contingency Table Program's display prompts and options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical Menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	

Step	Display	Response and Comments	Go To
2.	`Means and Moments`	First program title; press ▼ until "Contingency Table" appears on the LCD	3
3.	`Contingency Table`	Program title; press **END LINE**	4
4.	`df=####################`	Degrees of freedom: if output is to display, press any key; if output is to printer, no action necessary	5
5.	`Chi^2=################`	Chi-square statistic: if output is to display, press any key; if output is to printer, no action necessary	6
6.	`Repeat Results?`	To repeat output results, press **Y**	4
		To continue program, press **N**	1

Example

The following data were collected for testing the independence of hair color and sex in humans.

Sex	Hair Color				Total
	Black	Brown	Blond	Red	
Male	128	172	64	36	400
Female	110	130	128	32	400
Total	238	302	192	68	800

Test the following hypotheses:
H_O: Human hair color is independent of sex in the sample, and

H_A: Human hair color is not independent of sex in the sample

Using the Contingency Table Program, enter these data via the keyboard, save the file, and print the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`

Keystrokes	Display
K	`Grouped Ungrouped?`
U	`No. of Var.=0`
2 **END LINE**	`No. of Obs.=0`
4 **END LINE**	`Edit Parameters?`
N	`Edit Labels?`
Y	`X1=X1`
MALE **END LINE**	`X2=X2`
FEMALE **END LINE**	`Edit Labels?`
N	`MALE(1)=0`
128 **END LINE**	`MALE(2)=0`
172 **END LINE**	`MALE(3)=0`
64 **END LINE**	`MALE(4)=0`
36 **END LINE**	`FEMALE(1)=0`
110 **END LINE**	`FEMALE(2)=0`
130 **END LINE**	`FEMALE(3)=0`
128 **END LINE**	`FEMALE(4)=0`
32 **END LINE**	`MALE(1)=128`
Q	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
S	`SAVE:Filename?_`
HAIR **END LINE**	`Data Edit Menu Quit?`
D	`Kbd Save Load Print?`
P	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Contingency Table`
END LINE	(See Figure 7-1 for output) `Repeat Results?`

Keystrokes	Display
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are shown in Figure 7-1. This shows that human hair is dependent on sex for this sample; using df of 3 and a chi-square value of 28.77, which is significant at the 0.005 level.

Data Format Ungrouped
No. of Var. = 2
No. of Obs. = 4
Variable Labels
X1 = MALE
X2 = FEMALE

MALE

X1(1) = 128
X1(2) = 172
X1(3) = 64
X1(4) = 36

FEMALE
X2(1) = 110
X2(2) = 130
X2(3) = 128
X2(4) = 32

Contingency Table

df = 3
Chi^2 = 28.7710315914

Figure 7-1. Printed output for Contingency Table Program example problem.

Section 8

Rank Statistics

The next two programs in this library utilize non-parametric techniques or distribution-free methods. These methods are not dependent on a given distribution (such as the normal distribution in the case of the ANOVA), but usually will work for a wide range of different distributions. The two programs involve tests that are all based on the idea of ranking the variates in a sample after pooling all groups and considering them as a single sample for purposes of ranking.

Mann-Whitney U Test

The Mann-Whitney U Test Program, or simply Mann-Whitney Test Program, is a non-parametric test which evaluates the differences between sample means. This test is used most often in place of the two-sample t-test. Furthermore, it may be used when the basic assumptions make it impossible to use the t-test. The Mann-Whitney test is usable under almost all conditions; one exception is that the sampled populations must be continuous.

The Mann-Whitney U test compares the means of two populations having the same distribution. The program calculates the Mann-Whitney statistic, mean, variance, and Z statistic for each sample, using the standard normal form of the Mann-Whitney test. The program ranks the measurements from smallest to largest. The ranks of the measurements — not the actual measurements — are used by the program in the calculations. The null hypothesis, H_0, to be tested is that the means of the two samples are equivalent.

Equations

First, the measurements in each sample are ranked, with each measurement being assigned a specific rank, a_i and b_i. Next, the sum of the ranks for each sample is calculated using the following equations:

$$R_1 = \sum_{i=1}^{n_1} a_i$$

$$R_2 = \sum_{i=1}^{n_2} b_i$$

The equation for the Mann-Whitney statistic, U, for sample A is:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

where n_1 and n_2 are the number of observations in samples A and B respectively.

The equation for the Mann-Whitney statistic, U, for sample B is:

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

The equations for sample A and sample B yield different U's. The smaller value is U and the larger value is U′. You should check which value you have by applying the transformation:

$$U = n_1 n_2 - U'$$

If U or U′ is larger than the critical value of U, then the null hypothesis, H_0, must be rejected.

For those data where n_1 or n_2 are too large to determine the critical value of U, the significance of U may be determined by the Z statistic. The equation for the Z statistic is as follows:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sigma_U}$$

where the term

$$\frac{n_1 n_2}{2}$$

is usually called the ranked sum mean with a variance of σ_U^2.

The equation for the error of the distribution, σ_U, is:

$$\sigma_U = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

Finally, the critical value of the hypothesis is equal to the critical value for Z.

$$H_0\colon \mu_1 - \mu_2 = C$$

Remarks

The Mann-Whitney U Test Program uses ungrouped data. These data may contain several variables, but only two variables may be tested at a time.

User Information Form

This section presents a listing and brief tutorial explanation of the Mann-Whitney U Test Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	Data Edit Menu Quit?	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	Means and Moments	First statistics menu prompt; press ▼ until the "Mann-Whitney U Test" program title appears on the LCD	3
3.	Mann-Whitney U Test	Program title; press **END LINE**	4
4.	X Var. #=0	Enter number of variable to use as variable X (integer, 1 ≤ X Var. # ≤ N, the # of Variables in the data base); then, press **END LINE**	5
5.	Y Var. #=0	Enter number of variable to use as variable Y (integer, 1 ≤ Y Var. # ≤ N, the # of Variables in the data base); then, press **END LINE**	6
6.	Edit Parameters?	To edit parameters, press **Y**	4
		To continue program without editing, press **N**	7
7.	Output Ranks?	To output values and ranks, press **Y**	8
		To continue program, press **N**	13
8.	X(1)=#################	Sorted X(i) data: if output is to display, press any key; if output is to printer, no action necessary	9
9.	Rank=#################	Rank of sorted X(i) data: if output is to display, press any key; if output is to printer, no action necessary (NOTE: Steps 8 and 9 are repeated for all X(i) data)	10
10.	Y(1)=#################	Sorted Y(i) data: if output is to display, press any key; if output is to printer, no action necessary	11
11.	Rank=#################	Rank of sorted Y(i) data: if output is to display, press any key; if output is to printer, no action necessary (NOTE: Steps 10 and 11 are repeated for all Y(i) data)	12

Step	Display	Response and Comments	Go To
12.	`Repeat Results?`	To repeat output results, press **Y**	8
		To continue program, press **N**	13
13.	`RSumX=##############`	Sum of rankings in sample X: if output is to display, press any key; if output is to printer, no action necessary	14
14.	`W(X)=#################`	Mann-Whitney statistic for sample X: if output is to display, press any key; if output is to printer, no action necessary	15
15.	`Z(X)=#################`	Z statistic for sample X: if output is to display, press any key; if output is to printer, no action necessary	16
16.	`RSumY=##############`	Sum of rankings in sample Y: if output is to display, press any key; if output is to printer, no action necessary	17
17.	`W(Y)=#################`	Mann-Whitney statistic for sample Y: if output is to display, press any key; if output is to printer, no action necessary	18
18.	`Z(Y)=#################`	Z statistic for sample Y: if output is to display, press any key; if output is to printer, no action necessary	19
19.	`RMean=###############`	Ranked mean sum: if output is to display, press any key; if output is to printer, no action necessary	20
20.	`RankVar=#############`	Variance of the ranked mean: if output is to display, press any key; if output is to printer, no action necessary	21
21.	`Repeat Results?`	To repeat output results, press **Y**	13
		To continue program, press **N**	1

Example

Using the same data for the Unpaired t-Test Program example problem, determine if there is a difference in the production between Red Clover and Alfalfa?

Between Red Clover and Vetch? Between Alfalfa and Vetch? The hypotheses for each test are:

$$H_0: \mu_1 = \mu_2 \text{ and } H_A: \mu_1 \neq \mu_2$$

Load the data from the file BEAHONEY, print the ranks, and print the results. To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	Data Edit Menu Quit?
D	Kbd Save Load Print?
L	LOAD:Filename?_
BEAHONEY **END LINE**	Data Edit Menu Quit?
M	Means and Moments?
▼ or ▲ to find	Mann-Whitney U Test
END LINE	X Var, #=0
1 **END LINE**	Y Var, #=0
2 **END LINE**	Edit Parameters?
N	Output Ranks?
Y	(See Figure 8-1 for output) Repeat Results?
N	(See Figure 8-1 for output) Repeat Results?
N	Data Edit Menu Quit?
M	Mann-Whitney U Test
END LINE	X Var, #=0
1 **END LINE**	Y Var, #=0
3 **END LINE**	Edit Parameters?
N	Output Ranks?
Y	(See Figure 8-1 for output) Repeat Results?
N	(See Figure 8-1 for output) Repeat Results?

Keystrokes	Display
N	Data Edit Menu Quit?
M	Mann-Whitney U Test
END LINE	X Var. #=0
2 **END LINE**	Y Var. #=0
3 **END LINE**	Edit Parameters?
N	Output Ranks?
Y	(See Figure 8-1 for output) Repeat Results?
N	(See Figure 8-1 for output) Repeat Results?
N	Data Edit Menu Quit?
Q	Done
END LINE f OFF	

The results are shown in Figure 8-1. When comparing red clover and alfalfa, U is 22 and 78, respectively. The critical value for U (as found in statistical tables such as Rohlf and Sokal 1968), where the sample size for the X and Y variable is each 10 and using a two-tailed test, is 77 at the 0.05 level. Because one of the U's values is larger than the critical value of U, we conclude that the means of the red clover and alfalfa data are not the same and that there is a difference in production. The same reasoning applies when comparing the red clover and vetch data and the alfalfa and vetch data. Notice that in both cases one U is 100 and the other U is 0. The occurrence of zero in each instance results because the data are so different that the two sets of observations do not merge. Again, because one of the U's values is larger than the critical value of U, we conclude that there is a difference in the means of both the red clover and vetch data and the alfalfa and vetch data and that there is a difference in production in both cases.

Mann-Whitney U Test

X Var. # = 1
Y Var. # = 2

X(1) = 48.5
Rank = 18
X(2) = 47.9
Rank = 17
X(3) = 46.3
Rank = 13
X(4) = 45.4
Rank = 9.5
X(5) = 44.3
Rank = 6
X(6) = 48.9
Rank = 19
X(7) = 47.7
Rank = 16
X(8) = 46.1
Rank = 11
X(9) = 45.4
Rank = 9.5
X(10) = 46.9
Rank = 14
Y(1) = 46.2
Rank = 12
Y(2) = 47.4
Rank = 15
Y(3) = 44.9
Rank = 7
Y(4) = 43.2
Rank = 4
Y(5) = 40.7
Rank = 1
Y(6) = 45.2
Rank = 8
Y(7) = 42.6
Rank = 3
Y(8) = 41.9
Rank = 2
Y(9) = 43.8
Rank = 5
Y(10) = 49.4
Rank = 20

R Sum X = 133
W(X) = 22
Z(X) = −2.11660104886
R Sum Y = 77
W(Y) = 78
Z(Y) = 2.11660104886
R Mean = 50
Rank Var = 175

Figure 8-1. Printed output for Mann-Whitney U Test Program example problem (red clover versus alfalfa).

Mann-Whitney U Test

X Var. # = 1
Y Var. # = 3

X(1) = 48.5
Rank = 9
X(2) = 47.9
Rank = 8
X(3) = 46.3
Rank = 5
X(4) = 45.4
Rank = 2.5
X(5) = 44.3
Rank = 1
X(6) = 48.9
Rank = 10
X(7) = 47.7
Rank = 7
X(8) = 46.1
Rank = 4
X(9) = 45.4
Rank = 2.5
X(10) = 46.9
Rank = 6
Y(1) = 51.3
Rank = 12
Y(2) = 52.7
Rank = 15
Y(3) = 50.9
Rank = 11
Y(4) = 53
Rank = 16
Y(5) = 51.8
Rank = 13
Y(6) = 58.4
Rank = 20
Y(7) = 55.3
Rank = 19
Y(8) = 54.5
Rank = 18
Y(9) = 53.7
Rank = 17
Y(10) = 52.6
Rank = 14

R Sum X = 55
W(X) = 100
Z(X) = 3.7796447301
R Sum Y = 155
W(Y) = 0
Z(Y) = −3.7796447301
R Mean = 50
Rank Var = 175

Figure 8-1. Printed output for Mann-Whitney U Test Program example problem (red clover versus vetch).

Mann-Whitney U Test

X Var. # = 2
Y Var. # = 3

X(1) = 46.2
Rank = 8
X(2) = 47.4
Rank = 9
X(3) = 44.9
Rank = 6
X(4) = 43.2
Rank = 4
X(5) = 40.7
Rank = 1
X(6) = 45.2
Rank = 7
X(7) = 42.6
Rank = 3
X(8) = 41.9
Rank = 2
X(9) = 43.8
Rank = 5
X(10) = 49.4
Rank = 10
Y(1) = 51.3
Rank = 12
Y(2) = 52.7
Rank = 15
Y(3) = 50.9
Rank = 11
Y(4) = 53
Rank = 16
Y(5) = 51.8
Rank = 13
Y(6) = 58.4
Rank = 20
Y(7) = 55.3
Rank = 19
Y(8) = 54.5
Rank = 18
Y(9) = 53.7
Rank = 17
Y(10) = 52.6
Rank = 14

R Sum X = 55
W(X) = 100
Z(X) = 3.7796447301
R Sum Y = 155
W(Y) = 0
Z(Y) = −3.7796447301
R Mean = 50
Rank Var = 175

Figure 8-1. Printed output for Mann-Whitney U Test Program example problem (alfalfa versus vetch).

Kruskal-Wallis Test

The second non-parametric method to be presented is the Kruskal-Wallis Test. This test is analogous to the one-way, or single classification, analysis of variance (i.e., the general case with the number of samples designated as a and the number of variates per sample designated as n_i). The test assumes that the variable under study has an underlying continuous distribution and requires at least ordinal measurement of that variable.

The Kruskal-Wallis Test Program is general in nature and may be applied to a one-way ANOVA with equal size and to those with varying or unequal sample sizes. The program first ranks all of the variates from the smallest to the largest, without regard to division into groups. Ties may be a frequent problem during such a procedure. For these ties, the Kruskal-Wallis test computes the average of the ranks occupied by the tied values. Next, each original variate is replaced by its rank or average rank. The H-statistic is then computed. It alone is appropriate for data which contain no ties; when ties are present, the H-statistic must be divided by a correction factor, D.

The H-statistic is distributed approximately as $\chi^2_{[a-1]}$ for large samples when the null hypothesis is true. For tests at

$$\alpha = 0.10 \text{ or } \alpha = 0.05$$

the χ^2 approximation is very good even with n as small as 5. However, for

$$\alpha = 0.01$$

the test is conservative for small values of n; it rejects less than 1% of the tests if the hypothesis is true.

Equations

All observations are ranked first from smallest to largest and pooled together into a single group. In case of ties, average ranks are computed. Next, each observation in the original data table is replaced by its rank or average rank.

The sum of the ranks for each group is defined separately by

$$\text{Sum of Ranks for Each Group} = \left(\sum^{n_i} R\right)_i$$

where R is the rank (or average) rank and n_i is the number of variates per sample.

The mathematical equation for the H-statistic is:

$$H = \left[\frac{12}{\left(\sum^{a} n_i\right)\left(\sum^{a} n_i + 1\right)} \sum^{a} \frac{\left(\sum^{n_i} R\right)_i^2}{n_i} \right] - 3 \left(\sum^{a} n_i + 1 \right)$$

where $\sum^{a} n_i$ equals the sum of the sample sizes of the entire analysis; the numbers 12 and 3 are constants, and $(\sum^{n_i} R)_i$ is the sum of the ranks for the ith group.

When ties occur, the H-statistic must be divided by a correction term, D, whose equation is as follows:

$$D = 1 - \frac{\sum^{m} T_j}{(\sum^{a} n_i - 1)\sum^{a} n_i\,(\sum^{a} n_i + 1)}$$

where T_j is a function of the t_j, the number of variates tied in the jth group of ties. (Note that this t has no relation to Student's t.) The function is:

$$T_j = t_j^3 - t_j$$

computed easiest as

$$(t_j - 1)t_j\,(t_j + 1)$$

In most cases the tied group will range from $t = 2$ ties to $t = 10$ ties.

Remarks

The Kruskal-Wallis Test Program uses ungrouped data. These data may contain several variables; the program tests two or more variables at a time.

User Information Form

This section presents a listing and brief tutorial explanation of the Kruskal-Wallis Test Program's prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	`Means and Moments`	First statistics menu prompt; press ◆ until the "Kruskal-Wallis Test" program title appears on the LCD	3
3.	`Kruskal-Wallis Test`	Program title; press **END LINE**	4

Step	Display	Response and Comments	Go To
4.	#ofVariables=0	Enter number of variables (integer, 2 ≤ # of Variables ≤ MAXREAL); then, press **END LINE**	5
5.	Edit Parameters?	To edit parameter, press **Y**	4
		To continue program with data as entered, press **N**	6
6.	Var.#1=0	Enter number of variable to be used as first variable; then, press **END LINE** (NOTE: This prompt repeats for the number of variables, j, entered in Step 4; the actual number in the prompt increases from 1 to j)	7
7.	Edit Parameters?	To edit parameters, press **Y**	6
		To continue program with data as entered, press **N**	8
8.	Output Ranks?	To output ranks, press **Y**	9
		To continue program, press **N**	13
9.	RSum1=##############	Sum of rankings in group j: if output is to display, press any key; if output is to printer, no action necessary	10
10.	X1(1)=###############	Sorted data for Xj(i): if output is to display, press any key; if output is to printer, no action necessary	11
11.	Rank=################	Rank of sorted data: if output is to display, press any key; if output is to printer, no action necessary (NOTE: Step 9 is repeated for all Xj data; Steps 10 and 11 are repeated for all Xj(i) data)	12
12.	Repeat Results?	To repeat output results, press **Y**	9
		To continue program, press **N**	13
13.	H=####################	Kruskal-Wallis H-statistic: if output is to display, press any key; if output is to printer, no action necessary	14

Step	Display	Response and Comments	Go To
14.	D=####################	Correction term: if output is to display, press any key; if output is to printer, no action necessary	15
15.	Repeat Results?	To repeat output results, press **Y**	13
		To continue program, press **N**	1

Example

Using the Kruskal-Wallis Test Program, analyze the example problem used in the Paired t-Test Program. The hypotheses are:

$$H_0: \mu_1 = \mu_2 \quad \text{and} \quad H_A: \mu_1 \neq \mu_2$$

where μ_1 and μ_2 are the means of the twins weight loss scores for the two diets. Enter the data by loading the file saved in the Paired t-Test Program; then, print the ranks and the results. To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	Data Edit Menu Quit?
D	Kbd Save Load Print?
L	LOAD:Filename?_
DIETS **END LINE**	Data Edit Menu Quit?
M	Means and Moments?
▼ or ▲ to find	Kruskal-Wallis Test
END LINE	# of Variables=0
3 **END LINE**	Edit Parameters?
N	Var.#1=0
1 **END LINE**	Var.#2=0
2 **END LINE**	Var.#3=0
3 **END LINE**	Edit Parameters?
N	Output Ranks?

Keystrokes	Display
Y	(See Figure 8-2 for output) `Repeat Results?`
N	(See Figure 8-2 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are presented in Figure 8-2. Again, the test shows a difference in the two diets. The H value is 25.8 which is considerably higher than the chi-square value of 12.838 (df = 3), which is significant at the 0.005 level. Although the D value is given, the program automatically divides by this value in case of ties. Thus, you do not have to divide to obtain the H value.

Kruskal-Wallis Test

# of Variables = 3

Var. #1
Var. #2
Var. #3

R Sum 1 = 155	R Sum 2 = 255	R Sum 3 = 55
X1(1) = 5.6	X2(1) = 9.2	X3(1) = −3.6
Rank = 19	Rank = 27	Rank = 7
X1(2) = 3.7	X2(2) = 7.1	X3(2) = −3.4
Rank = 14	Rank = 21	Rank = 8.5
X1(3) = 4.8	X2(3) = 7.9	X3(3) = −3.1
Rank = 16	Rank = 23	Rank = 10
X1(4) = 5.4	X2(4) = 9.5	X3(4) = −4.1
Rank = 18	Rank = 29	Rank = 4.5
X1(5) = 3.5	X2(5) = 8	X3(5) = −4.5
Rank = 12	Rank = 24	Rank = 2
X1(6) = 3.4	X2(6) = 7.6	X3(6) = −4.2
Rank = 11	Rank = 22	Rank = 3
X1(7) = 3.6	X2(7) = 9.3	X3(7) = −5.7
Rank = 13	Rank = 28	Rank = 1
X1(8) = 4.2	X2(8) = 8.1	X3(8) = −3.9
Rank = 15	Rank = 25	Rank = 6
X1(9) = 4.9	X2(9) = 8.3	X3(9) = −3.4
Rank = 17	Rank = 26	Rank = 8.5
X1(10) = 5.7	X2(10) = 9.8	X3(10) = −4.1
Rank = 20	Rank = 30	Rank = 4.5

H = 25.8179390163
D = .999555061179

Figure 8-2. Printed output for Kruskal-Wallis Test Program example problem.

Part III

Statistical Distributions

In the previous parts of this manual we have referred to various statistical distributions. In Part III, two particular types of statistical distributions are discussed — sampling distributions and probability distributions.

Section 9

Sampling Distributions

For each sample drawn from a given population, various statistics can be computed (e.g., the mean and standard deviation) which vary from sample to sample. Thus, we can obtain a distribution of the statistic which is known as the sampling distribution. As an example, if the particular statistic used is the sample mean, the distribution is called the sampling distribution of means or the sampling distribution of the mean. For each sampling distribution, one also can calculate various statistics (e.g., the mean and standard deviation). Thus, you can obtain the mean and standard deviation of the sampling distribution of means, etc.

In summary, sampling distributions are probability distributions of means and variances, as well as those of random variables corresponding to other statistics. Discussed in this section are three sampling distributions useful in defining the rejection criteria when hypothesis testing:

- Student's t-Distribution;
- F-Distribution; and
- Chi-square Distribution.

By comparing the probability of a calculated value from one of these three tests with the probability in the rejection criteria, the decision to reject or not reject the null hypothesis can be made. Thus, these sampling distributions are useful in determining the rejection criteria when hypothesis testing.

Student's t-Distribution

The Student's t-Distribution, or just simply the t-Distribution, is the expected distribution of the ratio of the arithmetic mean minus the parametric mean to the sample standard deviation, or error. The Student's t-Distribution was discovered in 1908 by W.S. Gosset, who published under the pseudonym "Student"; thus the name Student's t-Distribution. It was perfected in 1924 by R.A. Fisher. This universally accepted and used statistical distribution revolutionized the statistics of small samples.

The Student's t-Distribution is very important in statistical analysis where the true mean, μ, is estimated from the sample mean, $\bar{x}$. The t-Distribution does not require that the true standard deviation, σ, be known nor estimated from the sample deviation, $S_{\bar{x}}$. The Student's t-Distribution does require that a sample be taken from a normally distributed population and that this sample have a mean, $\bar{x}$, and a standard deviation, $S_{\bar{x}}$.

The Student's t-Distribution Program computes the left tail area, P(t), given the degrees of freedom, df, and the t-statistic, t.

The curve of the Student's t-Distribution approaches the shape of the standard normal distribution ($\mu = 0$, $\sigma = 1$) more closely as the number of degrees of freedom increases. When the number of degrees of freedom equals infinity, the t-Distribution is the normal distribution. Figure 9-1 shows Student's t-Distributions for 1 and 2 degrees of freedom compared with the normal frequency distribution.

Equations

The left tail area, P(t), is determined from the following equation:

$$P(t) = \int_{-\infty}^{t} f(u)du$$

The Student's t-Distribution Program computes the left tail area, P(t), using the relationship between the F-distribution and the Student's t-Distribution. This relationship is expressed as:

$$P(t{:}\ df) = 1 - 1/2[Q(t^2{:}\ 1, df)]$$

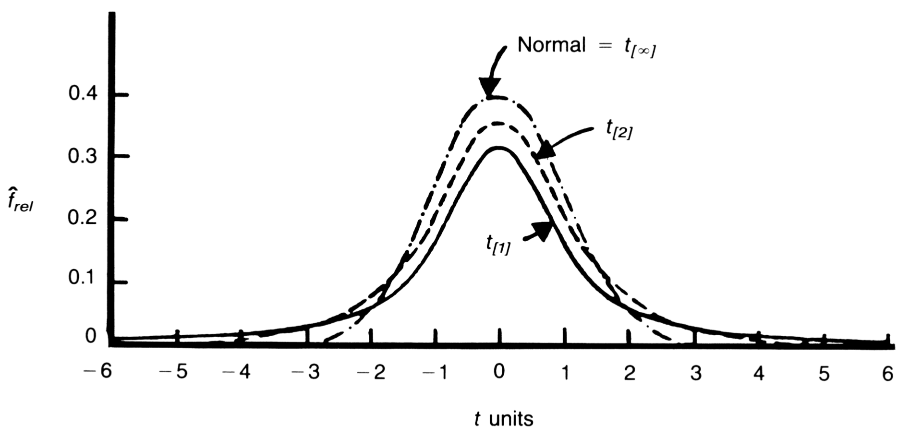


Figure 9-1. Frequency curves of Student's t-Distributions for 1 and 2 degrees of freedom compared with the normal distribution.

User Instruction Form

This section presents a listing and brief tutorial explanation of the Student's t-Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	`Means and Moments`	First statistics menu prompt; press ➡ until the "Student's t-Dist." program title appears on the LCD	3
3.	`Student's t-Dist.`	Program title; press **END LINE**	4
4.	`t=0`	Enter value for t-statistic (any number, $-\infty \leq t \leq \infty$); then, press **END LINE**	5
5.	`df=0`	Enter value for degrees of freedom (integer, $1 \leq df \leq 200$); then, press **END LINE**	6
6.	`Edit Parameters?`	To edit parameters, press **Y**	4
		To accept parameters without editing, press **N**	7
7.	`P(t)=##################`	Left tail area; if output is to display, press any key; if output is to printer, no action necessary	8
8.	`Q(t)=##################`	Right tail area; if output is to display, press any key; if output is to printer, no action necessary	9
9.	`Repeat Results?`	To repeat output results, press **Y**	7
		To continue program, press **N**	1

Example

Given a t-statistic of 4.5 with 10 degrees of freedom, calculate the left tail area, P(t), and the right tail area of the t-Distribution, Q(t). Print the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Student's t-Dist.`
END LINE	`t=0`
4.5 **END LINE**	`df=0`
10 **END LINE**	`Edit Parameters?`
N	(See Figure 9-2 for results) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

As shown in Figure 9-2, the left tail area, P(t), is 0.999428447455 and the right tail area of the t-Distribution, Q(t), is 0.000571552545.

```
Student's t-Dist.

t=4.5
df=10

P(t)=.999428447455
Q(t)=.000571552545
```

Figure 9-2. Printed output from Student's t-Distribution Program example.

F-Distribution

The F-Distribution, like the Student's t-Distribution, is a sample distribution; it is named in honor of R.A. Fisher. The F-Distribution and Student's t-Distribution are closely related. However, unlike the Student's t-Distribution, the F-Distribution is determined by two values for degrees of freedom, df_1 and df_2.

Because the F-Statistic is really the ratio of two independent chi-square distributions (each divided by its respective degrees of freedom), the F-Distribution Program utilizes this ratio to compare the variances of normal populations. Given the degrees of freedom found in the numerator and denominator of the ratio and the F-Statistic, the right tail area of the appropriate F curve, Q(F), can be calculated.

As shown in Figure 9-3, which presents three F-Distribution curves, very small degrees of freedom cause the distribution to be J-shaped. As the degrees of freedom increase, the shape of the distribution becomes humped and strongly skewed to the right. Most central F-Distributions have an average of 1.00, and approximately 5% of the frequency distribution falls beyond 3.18, as illustrated in Figure 9-4. If an F-ratio is selected at random from this distribution, the probability that an F-ratio with a value greater than 3.18 will be selected is equal to 0.05.

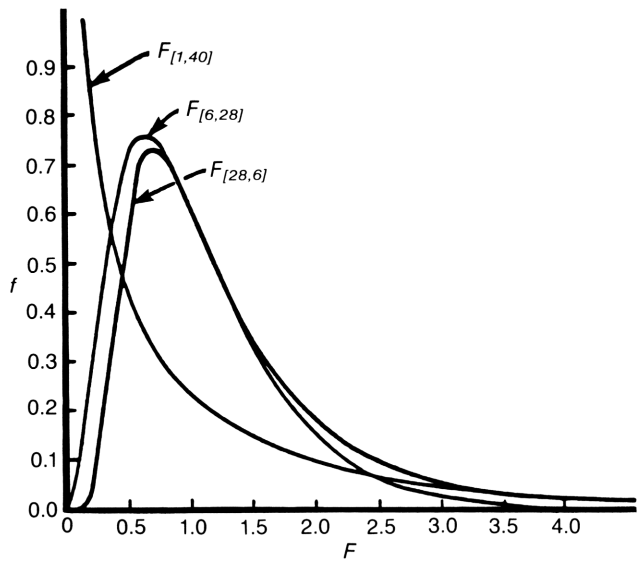


Figure 9-3. Three F-Distribution curves, with degrees of freedom equal to 1 and 40, 6 and 28, and 28 and 6.

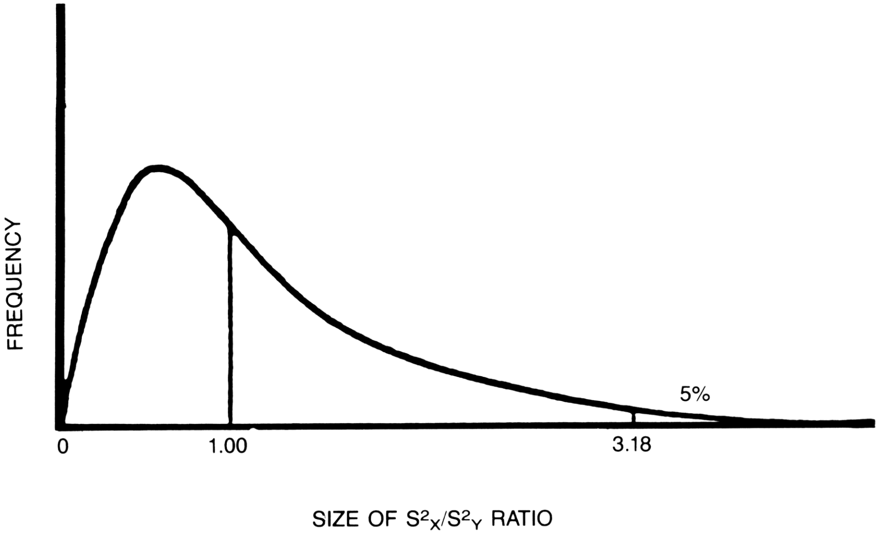


Figure 9-4. Frequency distribution $S^2_{\bar{x}}/S^2_{\bar{y}}$ with $df_x = 9$, $df_y = 9$, and $\sigma^2_x = \sigma^2_y$.

Equations

The equation for the F-Statistic, F, is as follows:

$$F = \frac{X^2/(N_1 - 1)}{Y^2/(N_2 - 1)} = \frac{X^2/df_1}{Y^2/df_2}$$

where X and Y are independent random variables with degrees of freedom, df_1 and df_2, respectively.

Three serial expansion equations are utilized by the F-Distribution Program to compute the right tail area of the F curve, Q(F). If the degrees of freedom in the numerator, df_1, is even, the following equation is used.

$$(F\colon df_1, df_2) = \chi^{\frac{df_2}{2}} \left[1 + \frac{df_2}{2}(1 - \chi) + \frac{df_2(df_2 + 2)}{2 \times 4}(1 - \chi)^2 + \ldots\right.$$

$$\left.\frac{df_2(df_2 + 2)\ldots(df_2 + df_1 - 4)}{2 \times 4 \ldots (df_1 - 2)}(1 + \chi)^{\frac{df_1 - 2}{2}}\right]$$

where $\chi = \dfrac{df_2}{df_2 + df_1 F}$

If the degrees of freedom in the denominator of the ratio, df_2, is even, the following equation is used.

$$Q\,(F\colon df_1,\, df_2) = 1 - (1-\chi)^{\frac{df_1}{2}} \left[1 + \frac{df_1}{2}\chi + \frac{df_1(df_1+2)}{2 \times 4}\chi^2 + \ldots\right.$$

$$\left.\frac{df_1\,(df_1+2)\ldots(df_2+df_1-4)}{2 \times 4 \ldots (df_2-2)}\,(1+\chi)^{\frac{df_1-2}{2}}\right]$$

When df_1 and df_2 are odd, the equation is very complex; for those interested, please refer to equation 26.6.8, page 946, in Abramowitz and Stegun (1972).

User Instruction Form

This section presents a listing and brief tutorial explanation of the F-Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	`Means and Moments`	First statistics menu prompt; press ⏷ until the "F-Distribution" program title appears on the LCD	3
3.	`F-Distribution`	Program title; press **END LINE**	4
4.	`df1=0`	Enter value for numerator degrees of freedom (integer, 1 ≤ df1 ≤ 200); then, press **END LINE**	5
5.	`df2=0`	Enter value for denominator degrees of freedom (integer, 1 ≤ df2 ≤ 200); then, press **END LINE**	6
6.	`F=0`	Enter value for F-Statistic (any number, 0 ≤ F ≤ ∞); then, press **END LINE**	7
7.	`Edit Parameters?`	To edit parameters, press **Y**	4
		To accept parameters without editing, press **N**	8

Step	Display	Response and Comments	Go To
8.	P(F)##################	Left tail area: if output is to display, press any key; if output is to printer, no action necessary	9
9.	Q(F)=#################	Right tail area: if output is to display, press any key; if output is to printer, no action necessary	10
10.	Repeat Results?	To repeat output results, press **Y**	8
		To continue program, press **N**	1

Example

Given an F-Statistic of 3.526 with numerator degrees of freedom, df1, equal to 1 and denominator degrees of freedom, df2, equal to 3, determine the left tail area, P(F), and the right tail area of the F curve, Q(F). Print the input and results. To solve this example, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	Data Edit Menu Quit?
M	Means and Moments
▼ or ▲ to find	F-Distribution
END LINE	df1=0
1 **END LINE**	df2=0
3 **END LINE**	F=0
3.526 **END LINE**	Edit Parameters?
N	(See Figure 9-5 for output) Repeat Results?
N	Data Edit Menu Quit?
Q	Done
END LINE f OFF	

As presented in Figure 9-5, the left tail area, P(F), is 0.842958108211 and the right tail area, Q(F), is 0.157041891789 for this example.

```
F-Distribution

df1=1
df2=3
F=3.526

P(F)=.842958108211
Q(F)=.157041891789
```

Figure 9-5. Printed output results for F-Distribution Program example.

Chi-Square Distribution

The chi-square distribution, χ^2, is a function whose values range from zero to infinity. The chi-square distribution is used as a measure of how a sample distribution deviates from a theoretical distribution. In the chi-square test, the actual frequencies are measured against the expected frequencies for corresponding variables. The larger the chi-square value, the larger the difference between the actual and expected frequency values. A chi-square of zero denotes a perfect fit, while a large chi-square value is a very bad fit. Thus, the chi-square distribution is referred to as a measure of goodness of fit.

The shape of the chi-square distribution, like the t-distribution, is related to changes in the number of degrees of freedom. Unlike the t-Distribution, the chi-square distribution usually is skewed positively. Figure 9-6 presents a chi-square distribution for three different degrees of freedom. The larger the degrees of freedom, the more dispersion in the distribution.

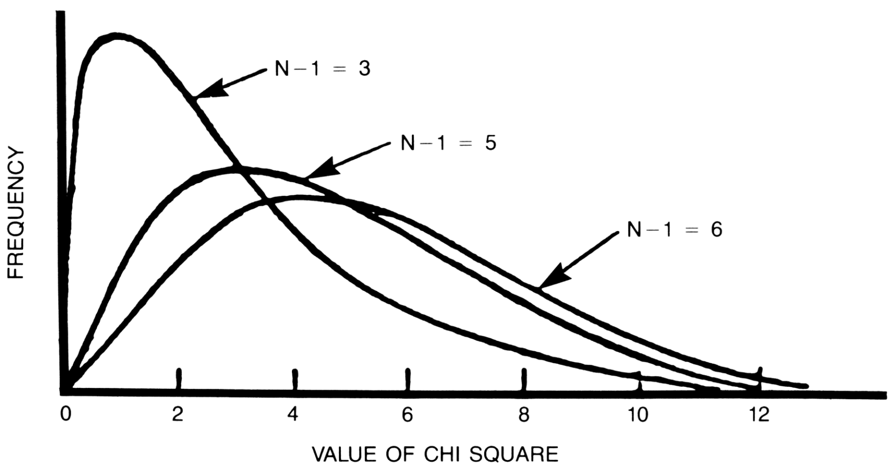


Figure 9-6. Chi-square distribution for df = 3, 5, and 6.

Equations

The right tail area, $Q(\chi^2)$, is calculated using an expansion series when the degrees of freedom are odd. The equation used is:

$$Q(\chi^2{:}df) = 2\,Q(\chi) + 2\,P(\chi) \sum_{r=1}^{\frac{df-1}{2}} \frac{X^{2r-1}}{1 * 3 * 5 * \ldots * (2r - 1)}$$

where $Q(\chi)$ and $P(\chi)$ are calculated using the normal distribution, and where:

$$\chi = \sqrt{\chi^2}$$

When the degrees of freedom are even, the formula is:

$$Q(\chi^2{:}df) = e^{-m} \sum_{j=0}^{C-1} \frac{m^j}{j!}$$

where:

$$C = \frac{df}{2}$$

$$m = \frac{\chi^2}{2}$$

Then, the following computation may be made:

$$P(\chi^2) = 1 - Q(\chi^2) = Pr(X \leq \chi\ )^2$$

Remarks

The Chi-Square Distribution Program calculates the right and left tail areas, $Q(\chi^2)$ and $P(\chi^2)$ respectively, given the chi-square statistic and the associated degrees of freedom. This program enables you to reject or not to reject an hypothesis where the observed differences between two or more sample populations, proportions, or percentages are the result of chance or are significantly different. If the null hypothesis is true, and the proportions of the "n" populations are indeed equal, then chances of obtaining a large chi-square value are small. We decide beforehand how small a chance we will accept; then, a chi-square value larger than this will cause us to reject the hypothesis.

User Information Form

This section presents a listing and brief tutorial explanation of the Chi-square Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	Data Edit Menu Quit?	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	Means and Moments	First statistics menu prompt; press ➡ until the "Chi-square Dist." program title appears on the LCD	3
3.	Chi-square Dist.	Program title; press **END LINE**	4

Step	Display	Response and Comments	Go To
4.	`df=0`	Enter value for degrees of freedom (integer, 1 ≤ df ≤ 200); then, press **END LINE**	5
5.	`Chi^2=0`	Enter value for chi-square (any number, 0 ≤ Chi^2 ≤ ∞); then, press **END LINE**	6
6.	`Edit Parameters?`	To edit parameters, press **Y**	4
		To accept parameters without editing, press **N**	7
7.	`P(Chi^2)=##############`	Left tail area: if output is to display, press any key; if output is to printer, no action necessary	8
8.	`Q(Chi^2)=##############`	Right tail area: if output is to display, press any key; if output is to printer, no action necessary	9
9.	`Repeat Results?`	To repeat output results, press **Y**	7
		To continue program, press **N**	1

Example

Given a chi-square value of 14 with 23 degrees of freedom, what is the left tail area, P(CHI^2), and the right tail area, Q(CHI^2)? Print the input and the results. To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Chi-square Dist.`
END LINE	`df=0`
23 **END LINE**	`Chi^2=0`
14 **END LINE**	`Edit Parameters?`
N	(See Figure 9-7 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`

Keystrokes	Display
Q	Done
END LINE f OFF	

The results are presented in Figure 9-7. For this example, the left tail area, P(Chi^2), is 0.073128701383 and the right tail area, Q(Chi^2), is 0.926871298617.

Chi-square Dist.

df=23
Chi^2=14

P(Chi^2)=.073128701383
Q(Chi^2)=.926871298617

Figure 9-7. Printed output results for the Chi-square Distribution Program example.

Section 10

Probability Distributions

The specialized form of a frequency distribution showing the values of a data set with their probabilities of occurrence is known as a probability distribution. The probability function — a mathematical expression relating a variable to its probability of occurrence — defines the values of the variable within the data set.

This section presents five probability distributions, each defined by its own specialized probability function. They are:

— Normal Distribution
— Weibull Distribution
— Exponential Distribution
— Binomial Distribution
— Poisson Distribution

Normal Distribution

The normal distribution has been called the most important and useful continuous distribution used in statistics because an abundance of situations may be described by it. For example, a few common measurements which are modeled adequately by the normal distribution are the heights and weights of men and women, the number of kernels on ears of corn, aptitude test scores, and life expectancies of automobile batteries. The mathematics of the normal distribution are easy to use and have resulted in a large amount of theory and practical applications.

A graphic presentation of the normal distribution is called the normal curve (sometimes it is referred to as the Gaussian curve or Laplacian curve). It is a bell-shaped curve as shown in Figure 10-1.

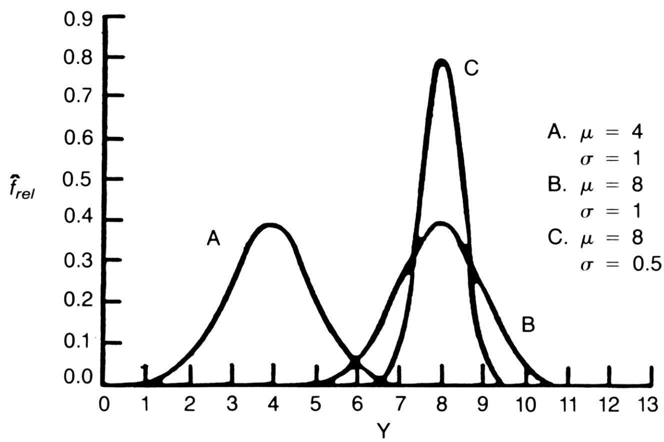


Figure 10-1. Normal curves.

Two parameters determine the normal probability density function: the parametric mean μ and the parametric standard deviation σ. These two statistics determine the location and shape, respectively, of the distribution. These concepts are illustrated by the three normal curves (each of which represents the same total frequencies) depicted in Figure 10-1. Curves A and B are at different locations on the X axis because they have different means. Curves B and C have identical means but different standard deviations. Because the standard deviation of curve C is only one-half that of curve B, it is much narrower to accommodate the same total frequency. In actuality, the density curves shown in Figure 10.1 never reach the X axis even though they appear that way on the diagram.

As noted earlier, a normal curve has a standard form with a mean of zero and a standard deviation of one; it also is symmetrical about the mean. Furthermore, the mean, median, and mode of the normal distribution are all at the same point. In a normal frequency distribution, certain percentages of expected frequencies lie within known limits as follows:

$\mu \pm 1\ \sigma$ contains 68.26% of the items;
$\mu \pm 2\ \sigma$ contains 95.46% of the items; and
$\mu \pm 3\ \sigma$ contains 99.72% of the items.

Furthermore, the following are true:

50% of items fall between $\mu - 0.674\ \sigma$ and $\mu + 0.674\ \sigma$,

95% of items fall between $\mu - 1.960\ \sigma$ and $\mu + 1.960\ \sigma$, and

99% of items fall between $\mu - 2.576\ \sigma$ and $\mu + 2.576\ \sigma$.

These percentages, which represent certain portions of the area under a normal curve, are calculated using integration of the normal probability function. See Figure 10-2 for a clarification of these concepts in graphic form.

Equations

The normal probability density function is derived from the equation (Abramowitz and Stegun 1972; Spiegel 1961):

$$f(x) = \frac{1}{\sigma\sqrt{2\ \pi}}\ e^{-[(x-\mu)^2/2\sigma^2]}$$

where f(x) indicates the height along the Y axis, or ordinate, of the curve, and represents the actual density of the items. The constants π and e are present in the equation; π is the ratio of the circumference of a circle to its diameter (π = 3.1415926536), and e is the base of the system of natural logarithms (e = 2.7182818285). As you will note, Figure 10-2 also illustrates the theoretical cumulative normal distribution function; this is the total frequency of the normal probability density function from negative infinity to any point along the X axis, or abscissa.

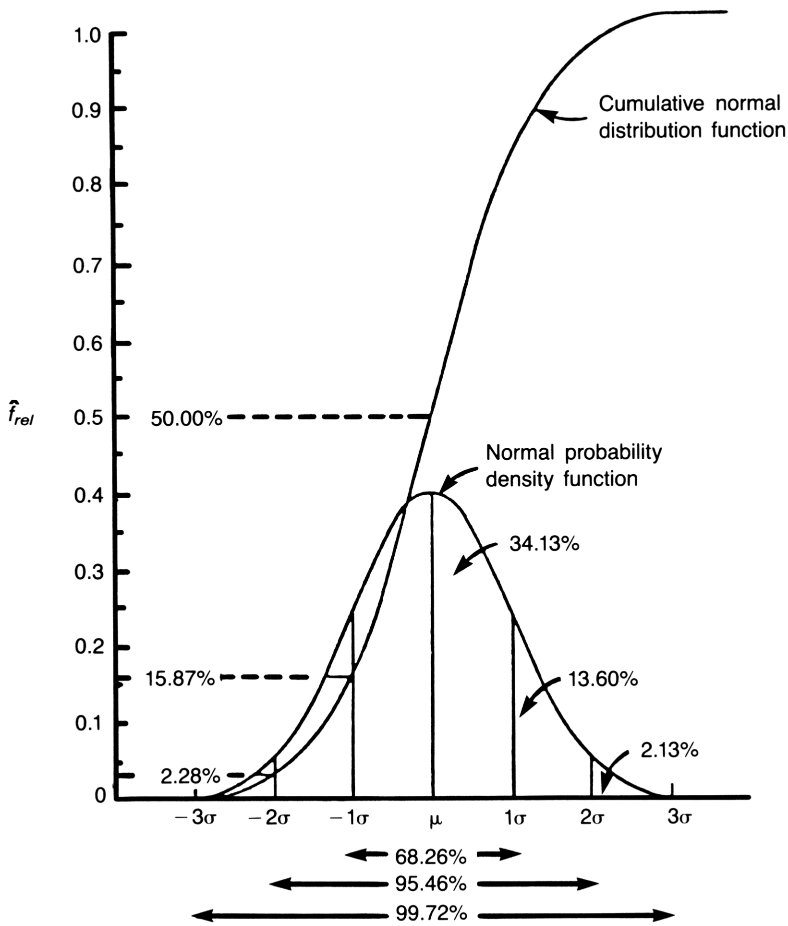


Figure 10-2. Graphic presentation of the normal probability density function and left tail area.

Because the normal distribution plays an important role in statistical theories, as well as in practical applications, it is beneficial to place it in a standard form. Using a linear transformation, the standard normal distribution is derived by the formula:

$$u = \frac{x - \mu}{\sigma}$$

to give the equation:

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)u^2}$$

The standard variate u obtained from the transformation is normally distributed with a mean of zero (0) and a standard deviation of one (1).

The right tail area, Q(u), which is the probability that an observed value is > u, is computed by the equation:

$$Q(u) = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{(-t^2/2)} dt$$

which can be approximated adequately with the formula:

$$Q(u) = f(u)\ (b_1 t^1 + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 + t(u)$$

where:

$$t = \frac{1}{(1 + pu)}$$

and the constants are:

$p =$	0.2316419	$b_4 =$	−1.821255978
$b_1 =$	0.319381530	$b_5 =$	1.330274429
$b_2 =$	−0.356563782	$t(u) <$	7.5E−08
$b_3 =$	1.781477937		

The left tailed area is computed by the equation:

$$P(u) = 1 - Q(u)$$

The right tail area, Q(u), and the left tail area, P(u), are shown in Figure 10-3. The area under the curve that is labeled P(u) represents the probability that the sample mean will fall below (or equal to) the value of u. Conversely, the area labeled Q(u) represents the probability that the sample mean will fall above the value of u.

For the equation to calculate the variate, u (where P(u) = 1 − Q(u), please refer to formula 26.2.23, page 933, in Abramowitz and Stegun (1972).

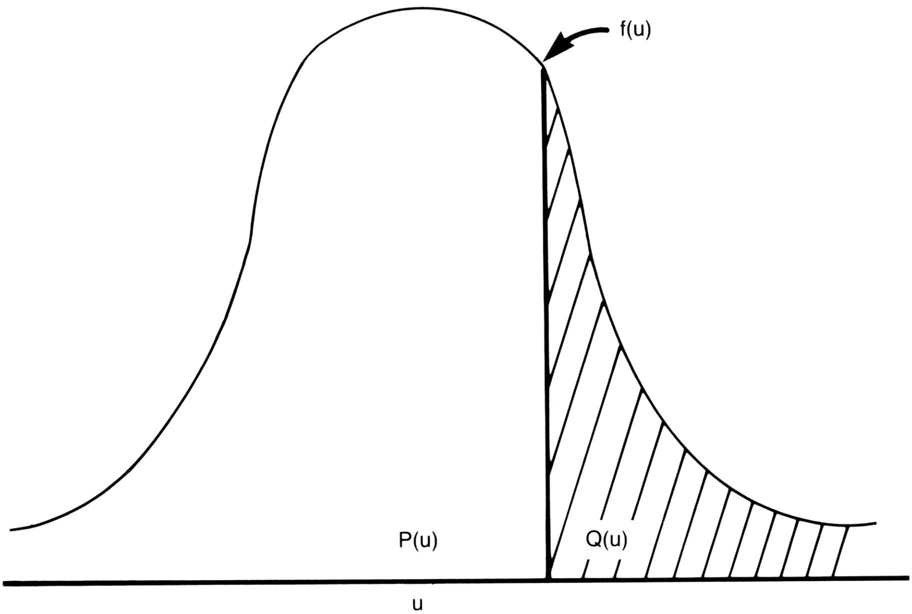


Figure 10-3. Right-tail cumulative density function Q(u) and inverse probability function P(u).

User Information Form

This section presents a listing and brief tutorial explanation of the Normal Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	
2.	`Means and Moments`	First statistics menu prompt; press ➡ until the "Normal Distribution" program title appears on the LCD	3
3.	`Normal Distribution`	Program title; press **END LINE**	4

Step	Display	Response and Comments	Go To
4.	Compute Q(u)?	To compute the right tail area, Q(u), press **Y**	5
		To compute the standard variate, u, press **N**	6
5.	u=0	Enter value for standard variate (any number, $-\infty \leq u \leq \infty$); then, press **END LINE**	7
6.	Q(u)=0	Enter value for right tail area (any number, $0 \leq Q(u) \leq 1$); then, press **END LINE**	7
7.	Edit Parameters?	To edit parameters, press **Y**	
		If Y selected in Step 4	5
		If N selected in Step 4	6
		To accept parameters without editing, press **N**	8
8.	f(u)=#################	Standard normal distribution function: if output is to display, press any key; if output is to printer, no action necessary	9
9.	P(u)=#################	Left tail area: if output is to display, press any key; if output is to printer, no action necessary	
		If Y selected in Step 4	10
		If N selected in Step 4	11
10.	Q(u)=#################	Right tail area: if output is to display, press any key; if output is to printer, no action necessary	12
11.	u=####################	Standard variate: if output is to display, press any key; if output is to printer, no action necessary	12
12.	Repeat Results?	To repeat output results, press **Y**	8
		To continue program, press **N**	1

Example

First, given a standard variate of 2.45, what is the right tail area, Q(u)? Second, given a right tail area, Q(u), of 0.24753, what is the standard variate, u? Print the input and the results. To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Normal Distribution`
END LINE	`Compute Q(u)?`
Y	`u=0`
2.45 **END LINE**	`Edit Parameters?`
N	(See Figure 10-4 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
M	`Normal Distribution`
END LINE	`Compute Q(u)?`
N	`Q(u)=0`
0.24753 **END LINE**	`Edit Parameters?`
N	(See Figure 10-4 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are shown in Figure 10-4. For the first part of the example, the standard normal distribution function, f(u), is 1.98373543918E − 2, the left tail area, P(u), is 0.99285718535, and the right tail area, Q(u), is 7.14281465047E − 3. For the second part of this example, the standard normal distribution function, f(u), is 0.316164511937, the left tail area, P(u), is 0.75247, and the standard variate, u, is 0.681988358063.

```
Normal Distribution

u=2.45

f(u)=1.98373543918E-2
P(u)=.99285718535
Q(u)=7.14281465047E-3

Normal Distribution

Q(u)=.24753

f(u)=.316164511937
P(u)=.75247
u=.681988358063
```

Figure 10-4. Printed output for Normal Distribution Program example.

Weibull Distribution

The Weibull distribution is quite similar to the exponential distribution, which will be covered in the next section of this manual. The Weibull distribution has its most widespread application in reliability testing as related to engineering and product manufacturing. For example, the Weibull distribution is quite useful in evaluating or predicting failure rates of machined metal parts in such items as automobiles, aircraft engines, or train cars.

Equations

The density function f(u) is found by the following equation, given the two Weibull parameters m and n and the standard variate u (Considine 1976):

$$f(u) = \frac{nu^{n-1}}{m} e^{-(u)^n/m}$$

The right tail area, Q(u), is calculated by the equation:

$$Q(u) = e^{-u^n/m}$$

where: $u^n/m \geq 0$, since $0 < Q(u) \leq 1$.

The left tail area also is calculated so that, given u, the standard variate, Q(u), may be computed.

The scale and shape parameters used in the Weibull distribution generally are presented graphically using Weibull probability paper (for an example, see Figure 10-5). An alternative method for finding these parameters is an iterative technique using the two equations:

$$\hat{b} = [(\frac{1}{n}) \sum_{i=1}^{n} X_i^{\hat{c}}]^{1/\hat{c}}$$

$$\hat{c}' = \frac{n}{[(1/\hat{b})^{\hat{c}} \sum_{i=1}^{n} X_i^{\hat{c}} \ln X_i - \sum_{i=1}^{n} \ln X_i}$$

where: X_i = observed data;

n = number of observations;

$\hat{b}$ = scale parameter;

$\hat{c}$ = shape parameter; and

$\hat{c}'$ = new shape parameter.

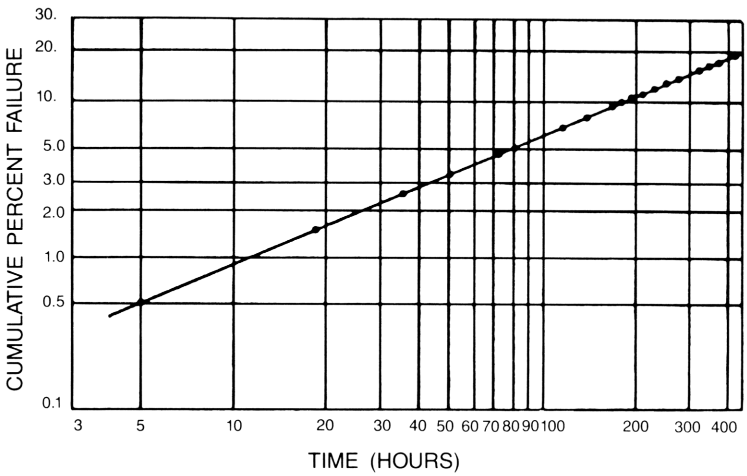


Figure 10-5. Representative Weibull distribution.

Substituting an initial guess for $\hat{c}$ into the previous equations, $\hat{c}'$ is calculated. Then, $\hat{c}'$ becomes the new $\hat{c}$ value and is used to calculate another $\hat{c}'$. If the initial guess for $\hat{c}'$ was a good one, this iteration technique will converge eventually so that the final calculated $\hat{c}'$ is equal to the new $\hat{c}'$ Finally, $\hat{b}$ and $\hat{c}$ may be substituted into the equations:

$$n = c' \quad \text{and} \quad m = bc'.$$

Remarks

The Weibull Distribution Program requires input of two characteristic parameters: scale and shape. Scale is represented by m and shape by n; both of these must be positive real numbers. An observed value also is needed as input. The program calculates the right tail area, Q(u), the left tail area, P(u), and the probability density function, f(u). The inverse Weibull distribution may be used to determine the value of u, given Q(u).

User Information Form

This section presents a listing and brief tutorial explanation of the Weibull Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu	
		For Data option and Edit option details, see Part I	
		To select Statistical menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	

Step	Display	Response and Comments	Go To
2.	Means and Moments	First statistics menu prompt; press ▼ until the "Weibull Distribution" program title appears on the LCD	3
3.	Weibull Distribution	Program title; press **END LINE**	4
4.	Compute Q(u)?	To compute the right tail area, Q(u), press **Y**	5
		To compute the Weibull variate, u, press **N**	6
5.	u=0	Enter value for Weibull variate (any number, $0 \leqslant u \leqslant \infty$); then, press **END LINE**	7
6.	Q(u)=0	Enter value for right tail area (any number, $0 \leqslant Q(u) \leqslant 1$); then, press **END LINE**	7
7.	Scale=0	Enter value for scale parameter (any number, $0 < \text{Scale} \leqslant \infty$); then, press **END LINE**	8
8.	Shape=0	Enter value for shape parameter (any number, $0 < \text{Scale} \leqslant \infty$); then, press **END LINE**	9
9.	Edit Parameters?	To edit parameters, press **Y**	
		If Y selected in Step 4	5
		If N selected in Step 4	6
		To accept parameters without editing, press **N**	10
10.	f(u)=#################	Weibull distribution density function: if output is to display, press any key; if output is to printer, no action necessary	11
11.	P(u)=#################	Left tail area: if output is to display, press any key; if output is to printer, no action necessary	
		If Y selected in Step 4	12
		If N selected in Step 4	13
12.	Q(u)=#################	Right tail area: if output is to display, press any key; if output is to printer, no action necessary	14

Step	Display	Response and Comments	Go To
13.	u=####################	Standard variate: if output is to display, press any key; if output is to printer, no action necessary	14
14.	Repeat Results?	To repeat output results, press **Y**	10
		To continue program, press **N**	1

Example

First, given a Weibull variate of 1.68, a scale parameter of 3, and a shape parameter of 2, what is right tail area, Q(u)? Second, given a right tail area, Q(u), of 0.36971, a scale parameter of 1, and a shape parameter of 1, what is the variate, u? Print the input and the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	Data Edit Menu Quit?
M	Means and Moments
▼ or ▲ to find	Weibull Distribution
END LINE	Compute Q(u)?
Y	u=0
1.68 **END LINE**	Scale=0
3 **END LINE**	Shape=0
2 **END LINE**	Edit Parameters?
N	(See Figure 10-6 for printed output.) Repeat Results?
N	Data Edit Menu Quit?
M	Weibull Distribution
END LINE	Compute Q(u)?
N	Q(u)=0
0.36971 **END LINE**	Scale=0
1 **END LINE**	Shape=0

Keystrokes	Display
1 **END LINE**	`Edit Parameters?`
N	(See Figure 10-6 for printed output.) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are presented in Figure 10-6. For the first part of this example, Weibull density function, f(u), is 0.437153313023, the left tail area, P(u), is 0.609684541942, and the right tail area, Q(u), is 0.390315458058. For the second part of the example, the Weibull density function, f(u) is 0.36971, the left tail area, P(u), is 0.63029, and the Weibull variate, u, is 0.995036364447.

Weibull Distribution

u = 1.68
Scale = 3
Shape = 2

f(u) = .437153313023
P(u) = .609684541942
Q(u) = .390315458058

Weibull Distribution

Q(u) = .36971
Scale = 1
Shape = 1

f(u) = .36971
P(u) = .63029
u = .995036364447

Figure 10-6. Printed output for Weibull Distribution Program example.

Exponential Distribution

The exponential distribution is a member of the family of Weibull distribution, with a shape parameter of one (1). As in the Weibull Distribution Program, the Exponential Distribution Program requires the entry of the scale parameter m, which must be real and positive.

Equations

The density function f(u) is calculated by the equation:

$$f(u) = \frac{1}{m} e^{-u/m}$$

The right tail area, Q(u), is defined by the equation:

$$Q(u) = e^{-u/m}$$

when u is greater than or equal to zero and m is greater than zero, or $Q(u) = 0$ for other values of the u and m parameters.

Applicable references for the exponential distribution equations are Abramowitz and Stegun (1972), Hays (1973), and Considine (1976).

Remarks

The density function, f(u), and the right tail area, Q(u), are calculated by this program. Figure 10-7 illustrates a typical exponential distribution pattern.

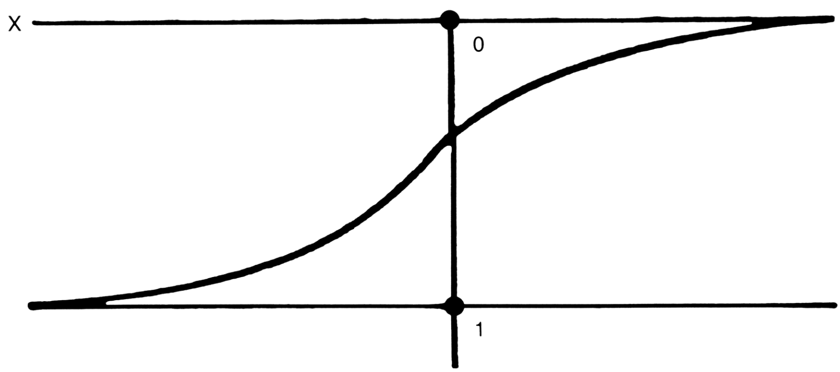


Figure 10-7. An exponential distribution.

User Information Form

This section presents a listing and brief tutorial explanation of the Exponential Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	`Means and Moments`	First statistics menu prompt; press ▼ until the "Exponential Dist." program title appears on the LCD	3
3.	`Exponential Dist.`	Program title; press **END LINE**	4
4.	`Compute Q(u)?`	To compute the right tail area, Q(u), press **Y**	5
		To compute the variate, u, press **N**	6
5.	`u=0`	Enter value for standard variate (any number, $0 \leqslant u \leqslant \infty$); then, press **END LINE**	7
6.	`Q(u)=0`	Enter value for right tail area (any number, $0 \leqslant Q(u) \leqslant 1$) then, press **END LINE**	7
7.	`Scale=0`	Enter value for scale parameter (any number, $0 < \text{Scale} \leqslant \infty$); then, press **END LINE**	8
8.	`Edit Parameters?`	To edit parameters, press **Y** If Y selected in Step 4 If N selected in Step 4	 5 6
		To accept parameters without editing, press **N**	9
9.	`f(u)=##################`	Exponential density function: if output is to display, press any key; if output is to printer, no action necessary	10
10.	`P(u)=##################`	Left tail area: if output is to display, press any key; if output is to printer, no action necessary If Y selected in Step 4 If N selected in Step 4	 11 12

Step	Display	Response and Comments	Go To
11.	`Q(u)=##################`	Right tail area: if output is to display, press any key; if output is to printer, no action necessary	12
12.	`u=#####################`	Exponential variate: if output is to display, press any key; if output is to printer, no action necessary	13
13.	`Repeat Results?`	To repeat output results, press **Y**	9
		To continue program, press **N**	1

Example

First, given an exponential variate of 3.4 and a scale parameter of 2, what is the right tail area, Q(u)? Second, given a right tail area, Q(u), of 0.125932 and a scale parameter of 1, what is the variate, u? Print the input and the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Exponential Dist.`
END LINE	`Compute Q(u)?`
Y	`u=0`
3.4 **END LINE**	`Scale=0`
2 **END LINE**	`Edit Parameters?`
N	(See Figure 10-8 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
M	`Compute Q(u)?`
N	`Q(u)=0`
0.125932 **END LINE**	`Scale=0`

Keystrokes	Display
1 **END LINE**	`Edit Parameters?`
N	(See Figure 10-8 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

Figure 10-8 presents the results. For the first part of this example, standard normal density function, f(u), is 9.13417620265E − 2, the left tail area, P(u), is 0.817316475947, and the right tail area, Q(u), is 0.182683524053. For the second part of the example, the exponential density function, f(u) is 0.125932, the left tail area, P(u), is 0.874068, and the exponential variate, u, is 2.07201320025.

```
Exponential Dist.

u=3.4
Scale=2

f(u)=9.13417620265E-2
P(u)=.817316475947
Q(u)=.182683524053

Exponential Dist.

Q(u)=.125932
Scale=1

f(u)=.125932
P(u)=.874068
u=2.07201320025
```

Figure 10-8. Printed output for the Exponential Distribution Program example.

Binomial Distribution

When the outcome of an event or selection has only two possibilities (success or failure), numerous such events can be evaluated using statistics dealing with binomial distribution (bi = two). The binomial distribution is applicable to many everyday situations, such as predicting the number of times that a coin will land on "tails" in a given number of coin tosses, or in finding the number of free throws that a basketball player will make in a known number of attempts. The binomial distribution is applicable when the variables have the following properties:

- Each trial results in only one of two possible outcomes, success or failure.
- The experiment or observation contains n identical trials.
- Each successful single trial has a probability equal to p, and the probability is the same for each trial. Conversely, the probability of a failure q also is constant, and is equal to 1 − p.
- Each trial is totally independent of other trials.
- During n trials, k number of successes are observed.

Equations

For the Binomial Distribution Program, the mean (μ) is calculated by the equation:

$$\mu = np$$

The standard deviation (σ) is found by the equation:

$$\sigma = \sqrt{(np\ (1 - p))}$$

The equation for determining the probability function for a binomial distribution F(p) is:

$$f(k;\ n,p) = \binom{n}{k}\ (p^k)(1-p)^{n-k}$$

when k = 0, 1, 2, 3, . . ., n. The equation becomes:

$$f(k;\ n,p) = 0$$

when k > n or k < 0. Also note that n is the number of trials, k is the number of successes, and f(k; n,p) is the probability of exactly k successes in n trials.

The left tail area, P(k), is calculated by using the equation:

$$P(k) = \sum_{j=0}^{k} f(j;\ n,p)$$

The probability of k or fewer successes is found by:

$$F(k;\, n,p) = P(k)$$

The probability of more than k successes is found by:

$$1 - F(k;\, n,p) = Q(k)$$

Applicable references for the binomial distribution include Spiegel (1961), Abramowitz and Stegun (1972), and Mendenhall (1965).

Remarks

The Binomial Distribution Program calculates the probability of obtaining exactly k number of successes f(k), the probability of obtaining k or fewer successes P(k), the probability of obtaining more than k successes, Q(k), the mean (μ), and the standard deviation (σ) for the number of trials being used. Figure 10-9 shows two typical binomial distributions.

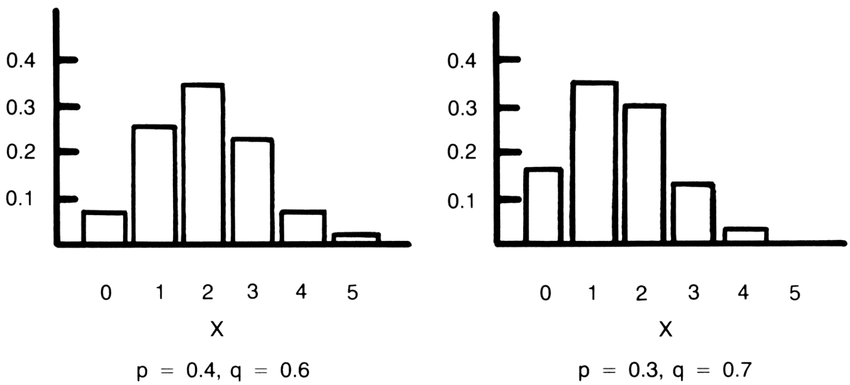


Figure 10-9. Two typical binomial distributions.

User Information Form

This section presents a listing and brief tutorial explanation of the Binomial Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M**	2
		To select Quit option, press **Q**; for details, see Part I	

Step	Display	Response and Comments	Go To
2.	Means and Moments	First statistics menu prompt; press ▼ until the "Binomial Distribution" program title appears on the LCD	3
3.	Binomial Distribution	Program title; press **END LINE**	4
4.	n=0	Enter value for number of trials (integer, $0 \leq n \leq 200$); then, press **END LINE**	5
5.	p=0	Enter value for probability (any number, $0 \leq p \leq 1$); then, press **END LINE**	6
6.	k=0	Enter value for number of successes (integer, $0 \leq k \leq n$, as entered in Step 4); then, press **END LINE**	7
7.	Edit Parameters?	To edit parameters, press **Y**	4
		To accept parameters without editing, press **N**	8
8.	Mean=#################	Mean: if output is to display, press any key; if output is to printer, no action necessary	9
9.	StdDev=###############	Standard deviation: if output is to display, press any key; if output is to printer, no action necessary	10
10.	f(k)=#################	Probability function: if output is to display, press any key; if output is to printer, no action necessary	11
11.	P(k)=#################	Cumulative probability: if output is to display, press any key; if output is to printer, no action necessary	12
12.	Q(k)=#################	Inverse cumulative probability: if output is to display, press any key; if output is to printer, no action necessary	13
13.	Repeat Results?	To repeat output results, press **Y**	8
		To continue program, press **N**	1

Example

If you flipped a two-sided coin 100 times, what is the probability of exactly 20 heads? Of 20 or less? Of more than 20? Print the input and the results.

To solve this example problem, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Binomial Distribution`
END LINE	`n=0`
100 **END LINE**	`p=0`
0.5 **END LINE**	`k=0`
20 **END LINE**	`Edit Parameters?`
N	(See Figure 10-10 for printed results.) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

The results are shown in Figure 10-10. The probability of exactly 20 heads, f(k), is 4.22816326764E − 10, the probability of 20 or less, P(k), is 5.57954452868E − 10, and the probability of more than 20, Q(k), is 0.999999999442.

```
Binomial Distribution

n=100
p=.5
k=20

Mean=50
Std Dev=5
f(k)=4.22816326764E-10
P(k)=5.57954452868E-10
Q(k)=.999999999442
```

Figure 10-10. Printed output for the Binomial Distribution Program example.

Poisson Distribution

The Poisson distribution, like the binomial distribution, is a discrete (discontinuous) random variable; the random variable is the number of observed successes or events. The Poisson distribution is quite common in quality control of manufacturing processes when numbers of samples are large and "success" is comparatively small (e.g., number of defects in factory made bolts).

The Poisson distribution also is a good approximation of the binomial distribution, in which the probability of success p is very small. Like the binomial, the Poisson distribution is based on the assumption that the probability of success is constant from trial to trial. However, in a Poisson distribution, the trials can be considered as a continuum of minute intervals of time, distance, size, or area rather than the well-defined trials of the binomial distribution. An example is the number of arrivals at a toll booth in a one-hour span — an arrival may occur at any one instant, and in this case the number of arrivals is the Poisson variate.

Equations

As an approximation to the binomial distribution, when n (number of trials) is large and p (probability of success) is small, the distribution probability function f(k) is found by using the following equation (Spiegel 1961; Freund and Williams 1972; Abramowitz and Stegun 1972; Larson 1975):

$$f(k) = (e^{-m}) \frac{m^k}{k!}$$

where: $k = 0, 1, 2, 3, \ldots, n$
$m = np$.

The inverse cumulative probability, Q(k), is calculated by the equation:

$$Q(k) = \sum_{k+1}^{\infty} f(j) = 1 - P(k)$$

The Poisson Distribution Program calculates the probability of exactly k successes, f(k), the probability of k or fewer successes, P(k), and the probability of more than k successes, Q(k).

Remarks

The Poisson Distribution Program calculates the individual probability function, f(k), and the cumulative probability, P(k), by using the expected number of successes m in a sample derived from the sample size and probability of occurrence. Figure 10-11 gives a typical Poisson distribution (in bar chart form) for which the mean μ is very small. Figure 10-12 illustrates the Poisson distribution for a variety of μ values.

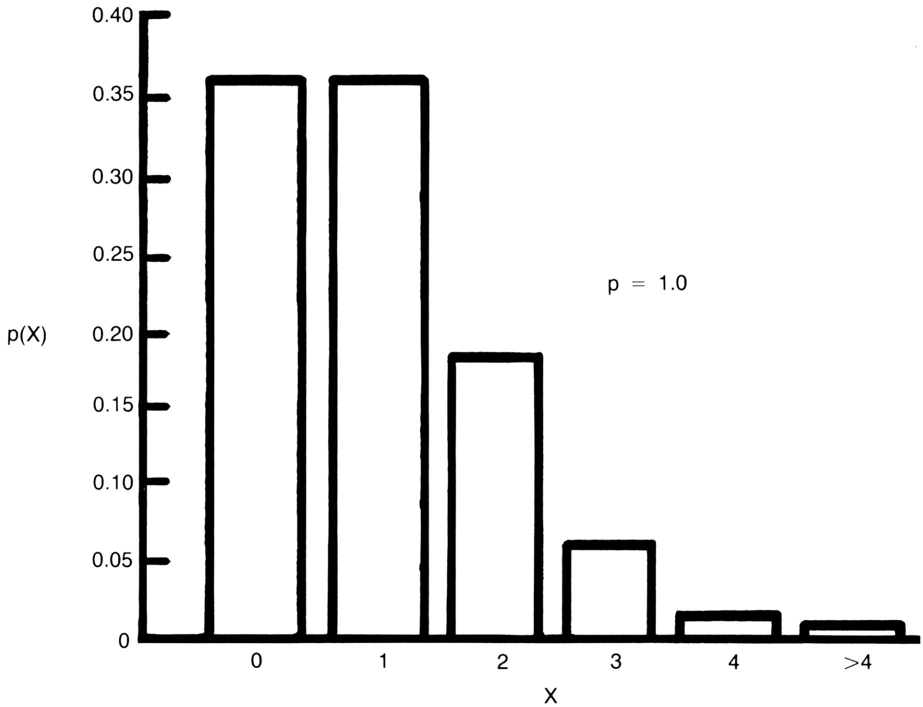


Figure 10-11. A typical Poisson distribution in bar chart form.

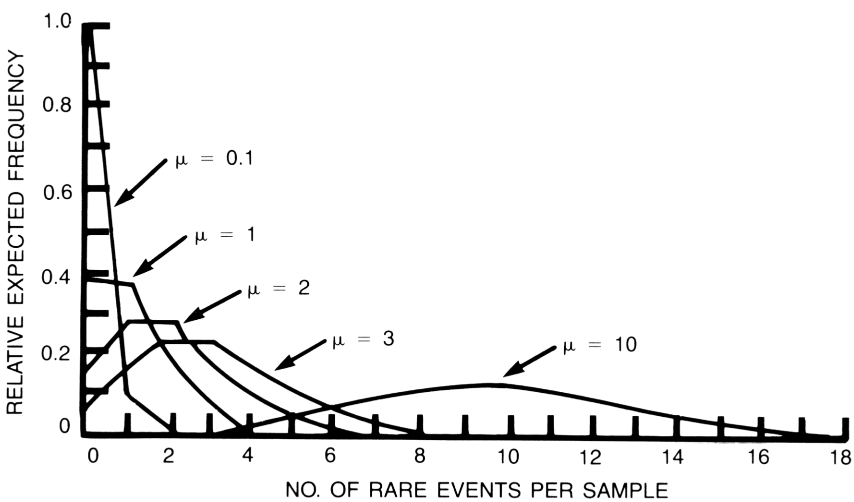


Figure 10-12. Poisson distribution for various mean values.

User Information Form

This section presents a listing and brief tutorial explanation of the Poisson Distribution Program's display prompts and their options.

Step	Display	Response and Comments	Go To
1.	`Data Edit Menu Quit?`	Main menu For Data option and Edit option details, see Part I To select Statistical menu option, press **M** To select Quit option, press **Q**; for details, see Part I	 2
2.	`Means and Moments`	First statistics menu prompt; press ▼ until the "Poisson Distribution" program title appears on the LCD	3
3.	`Poisson Distribution`	Program title; press **END LINE**	4
4.	`Mean=0`	Enter value for expected success parameter (any number, $0 \leq$ Mean $\leq \infty$); then, press **END LINE**	5
5.	`k=0`	Enter value for number of successes (integer, $0 \leq k \leq n$, the # of trials); then, press **END LINE**	6
6.	`Edit Parameters?`	To edit parameters, press **Y**	4
		To accept parameters as entered, press **N**	7
7.	`f(k)=#################`	Probability function: if output is to display, press any key; if output is to printer, no action necessary	8
8.	`P(k)=#################`	Cumulative probability: if output is to display, press any key; if output is to printer, no action necessary	9
9.	`Q(k)=#################`	Inverse cumulative probability: if output is to display, press any key; if output is to printer, no action necessary	10
10.	`Repeat Results?`	To repeat output results, press **Y**	7
		To continue program, press **N**	1

Example

Using the Poisson distribution to approximate the binomial distribution, find the probability of four (4) successes in 300 trials, assuming that p = 0.001. (Remember that the expected success parameter, m = np = 300 × 0.001 = 0.3.) Print the output. To solve this example, use the following keystrokes.

Keystrokes	Display
RUN AMPISTAT **END LINE**	`Data Edit Menu Quit?`
M	`Means and Moments`
▼ or ▲ to find	`Poisson Distribution`
END LINE	`Mean=0`
0.3 **END LINE**	`k=0`
4 **END LINE**	`Edit Parameters?`
N	(See Figure 10-13 for output) `Repeat Results?`
N	`Data Edit Menu Quit?`
Q	`Done`
END LINE f OFF	

Figure 10-13 presents the results. The probability function, f(k), is 2.5002614948E−4, the cumulative probability, P(k), is 0.99998421496, and the inverse cumulative probability, Q(k), is 0.00001578504.

Poisson Distribution

Mean=.3
k=4

f(k)=2.5002614948E−4
P(k)=.99998421496
Q(k)=.00001578504

Figure 10-13. Printed output for the Poisson Distribution Program example.

Part IV

Appendices and Index

Appendix A. Program Inputs/Outputs Summary

This appendix summarizes the input and output prompts (edit, status, and error prompts are omitted) for each of the 18 programs contained in this AMPI Statistics Library. Please note that the abbreviation for each input is followed by a colon and a flashing cursor (denoted in this manual as a "—"). Likewise, the abbreviation for each output is followed by an equal sign and a series of pound (#) signs denoting numbers. The heading "Response and Comments" assumes that a printer is not attached to the HP-71B computer. The programs are listed in the order that they appear in this manual.

Table A-1, which follows this listing, presents limitations and restrictions for data inputs.

Screen	Response and Comments
PROGRAM: MEANS AND MOMENTS	
`Means and Moments`	Program title; press **END LINE**
Inputs:	
Special prompts:	
`Use n-1 for df?`	Press **Y** (to use degrees of freedom equal to n − 1) or **N** (to use degrees of freedom equal to n)
`Var.#=0`	Enter number of the variable that you wish to test; press **END LINE**
Outputs:	
`#Data=###############`	Number of observations; press any key
`AMean=###############`	Arithmetic mean; press any key
`GMean=###############`	Geometric mean; press any key
`HMean=###############`	Harmonic mean; press any key
`Moment2=#############`	Second moment; press any key
`Moment3=#############`	Third moment; press any key
`Moment4=#############`	Fourth moment; press any key
`Kurtosis=############`	Kurtosis; press any key
`Skewness=############`	Skewness; press any key
PROGRAM: HISTOGRAM	
`Histogram`	Program title; press **END LINE**

Screen	Response and Comments
Input:	
Special prompts:	
`Var.#=0`	Enter value for number of the variable that you wish to test; press **END LINE**
`#ofCells=0`	Enter value for number of cells; press **END LINE**
`Width=0`	Enter value for width of cells; press **END LINE**
`Minimum=0`	Enter minimum acceptable data value; press **END LINE**
`Output Groups?`	Press **Y** (to output group or cell frequency and mean) or **N** (to continue without seeing groups)
`Use n-1 for df?`	Press **Y** (to use degrees of freedom equal to n − 1) or **N** (to use degrees of freedom equal to n)
`Output Statistics?`	Press **Y** (to output descriptive statistics) or **N** (to continue program without seeing descriptive statistics)
Outputs:	
If **Y** to output groups prompt:	
`Cell#=#####`	Number of cell; press any key
`Freq=#####`	Frequency of cell; press any key
`Mean=#####`	Mean of cell; press any key
If **N** to output statistics prompt:	
`#Data=###############`	Number of observations used; press **END LINE**
`AMean=###############`	Arithmetic mean; press any key
`GMean=###############`	Geometric mean; press any key
`HMean=###############`	Harmonic mean; press any key
`Moment2=############`	Second moment; press any key
`Moment3=############`	Third moment; press any key
`Moment4=############`	Fourth moment; press any key
`Kurtosis=############`	Kurtosis; press any key
`Skewness=############`	Skewness; press any key
Inputs (for theoretical histograms):	
Special prompts:	
`Theo Histogram?`	Press **Y** (to execute theoretical histogram) or **N** (not to execute theoretical histogram)
If **Y** to theoretical histogram question:	
Control prompts:	
`Normal Distribution`	Press ◀ or ▶ to find normal distribution, then **END LINE**

Screen	Response and Comments
If normal distribution selected:	
`Use Calc Param?`	Press **Y** (to use computed mean and standard deviation) or **N** (to use hypothetical mean and standard deviation)
If **N** to use calculated parameters question:	
Data prompts:	
`Mean=0`	Enter hypothetical mean value; press **END LINE**
`Std Dev=0`	Enter hypothetical standard deviation value; press **END LINE**
If **Y** to theoretical histogram question:	
Control prompts:	
`Weibull Distribution`	Press ▼ or ▲ to find Weibull distribution, then **END LINE**
If Weibull distribution selected:	
Data prompts:	
`Scale=0`	Enter scale parameter value; press **END LINE**
`Shape=0`	Enter shape parameter value; press **END LINE**
If **Y** to theoretical histogram question:	
Control prompts:	
`Exponential Dist.`	Press ▼ or ▲ to find exponential distribution, then **END LINE**
If exponential distribution selected:	
Data prompt:	
`Shape=0`	Enter shape parameter value; press **END LINE**
If **Y** to theoretical histogram question:	
Control prompts:	
`Binomial Distribution`	Press ▼ or ▲ to find binomial distribution), then **END LINE**
If binomial distribution selected:	
Control prompts:	
`Use Calc Param?`	Press **Y** (to use the calculated probability) or **N** (to use hypothetical probability)
Data prompts:	
`n=0`	Enter value for number of trials; press **END LINE**
If **N** to using calculated parameter question:	
Data prompts:	
`p=0`	Enter hypothethical probability value; press **END LINE**

Screen	Response and Comments
If **Y** to theoretical histogram question:	
Control prompts:	
`Poisson Distribution`	Press ▼ or ▲ to find Poisson distribution, then **END LINE**
If Poisson distribution selected:	
`Use Calc Param?`	Press **Y** (to use computed success parameter) or **N** (to use hypothetical probability)
If **N** to using calculated parameter question:	
Data prompt:	
`Mean=0`	Enter expected hypothetical success parameter; press **END LINE**
If **Y** to theoretical histogram question:	
Control prompts:	
`Uniform Dist.?`	Press ▼ or ▲ to find uniform distribution, then **END LINE**
Outputs (for theoretical histograms):	
`Output Exp-Freq?`	Press **Y** (to output theoretical histogram expected frequency and chi-square statistics) or **N** (to output just chi-square statistics)
`Ex-F(i)=####`	Expected frequency of each cell; press any key
`df=#####`	Degrees of freedom; press any key
`Chi^2=#####`	Chi-square; press any key
`Q(Chi^2)=#####`	Right tail area; press any key

PROGRAM: MULTIPLE LINEAR REGRESSION

Screen	Response and Comments
`Multiple Linear Reg.`	Program title; press **END LINE**
Inputs:	
Special prompts:	
`Output Corr. Matrix?`	Enter **Y** (to output a correlation matrix) or **N** (to continue program without seeing the correlation matrix)
`Std Regression Coeff?`	Enter **Y** (to compute the standard regression coefficient) or **N** (to compute the nonstandard regression coefficient)
Outputs:	
If **Y** for correlation matrix:	
`ri(j)=################`	Elements of correlation matrix, where ri(j) denotes the correlation between Xi and Xj, and where the last variable is always the independent variable; press any key after each appears

Screen	Response and Comments
If **Y** for standard regression coefficient question:	
`R^2=##################`	Coefficient of determination; press any key
`SST=##################`	Total sum of squares; press any key
`df/SST=##`	Degrees of freedom for SST; press any key
`SSRg=#################`	Regression sum of squares; press any key
`df/SSRg=##`	Degrees of freedom for SSR; press any key
`SSE=##################`	Error sum of squares; press any key
`df/SSE=##`	Degrees of freedom for SSE; press any key
`SE=###################`	Standard error; press any key
`F=####################`	F-statistic; press any key
`b(i)=#################`	Standard regression coefficient; press any key
`SEb(i)=###############`	Standard error of b(i); press any key
If **N** for standard regression coefficient question:	
`R^2=##################`	Coefficient of determination; press any key
`SST=##################`	Total sum of squares; press any key
`df/SST=##`	Degrees of freedom for SST; press any key
`SSRg=#################`	Regression sum of squares; press any key
`df/SSRg=##`	Degrees of freedom for SSR; press any key
`SSE=##################`	Error sum of squares; press any key
`df/SSE=##`	Degrees of freedom for SSE; press any key
`SE=###################`	Standard error; press any key
`F=####################`	F-statistic; press any key
`B(0)=#################`	Constant for non-standard regression coefficient; press any key
`B(i)=#################`	Non-standard regression coefficient; press any key
`SEB(i)=###############`	Standard error of B(i); press any key

PROGRAM: PAIRED t-TEST

`Paired t-Test`	Program title; press **END LINE**

Inputs:

Special prompts.

`Diff.Var.#=0`	Enter the variable number which contains the differences; then, press **END LINE**

Outputs:

`#Data=##############`	Number of observations; press any key
`Mean=###############`	Mean; press any key
`StdDev=#############`	Standard deviation; press any key
`df=##`	Degrees of freedom; press any key
`t=####################`	t-statistic; press any key
`SE=###################`	Standard error; press any key

Screen	Response and Comments
PROGRAM: UNPAIRED t-TEST	
`Unpaired t-Test`	Program title; press **END LINE**
Inputs:	
Special prompts:	
`Hyp. Diff=0`	Enter hypothesized difference between the means of the two samples; press **END LINE**
`X Var. #=0`	Enter the number of the variable selected as X; then, press **END LINE**
`Y Var. #=0`	Enter the number of the variable selected as Y; then, press **END LINE**
Outputs:	
`X#Data=##############`	Number of observations for X; press any key
`XMean=################`	Mean for X; press any key
`XStdDev=##############`	Standard deviation for X; press any key
`Y#Data=##############`	Number of observations for Y; press any key
`YMean=################`	Mean for Y; press any key
`YStdDev=##############`	Standard deviation for Y; press any key
`df=##`	Degrees of freedom; press any key
`t=####################`	t-statistic; press any key
`PoolVar.=############`	Pooled variance; press any key
`SE=###################`	Standard error; press any key
PROGRAM: ONE-WAY ANOVA	
`One-way ANOVA`	Program title; press **END LINE**
Inputs:	
Special prompts.	
`# of Variables=0`	Enter number of variables to be tested; then, press **END LINE**
`Var. #1=0`	Enter the number of the variable which is used as variable number 1 (prompt repeats for number of variables tested); then, press **END LINE**
Outputs:	
`SST=##################`	Total sum of squares; press any key
`dfT=#####`	Total degrees of freedom; press any key
`SSW=#####`	Between groups sum of squares; press any key
`dfW=#####`	Between groups degrees of freedom; press any key
`MSW=#####`	Between groups mean square; press any key
`SSE=#####`	Error sum of squares; press any key
`dfE=#####`	Error degrees of freedom; press any key
`MSE=#####`	Error mean square; press any key
`F=######`	F-statistic; press any key

Screen	**Response and Comments**
PROGRAM: TWO-WAY ANOVA	
`Two-way ANOVA`	Program title; press **END LINE**
Inputs:	
Control prompts:	
`#ofRows=0`	Enter number of rows; then, press **END LINE**
Outputs:	
If two (2) or more observations per variable:	
`SST=###################`	Total sum of squares; press any key
`dfT=#####`	Total degrees of freedom; press any key
`SSE=###################`	Error sum of squares; press any key
`dfE=#####`	Error degrees of freedom; press any key
`MSE=###################`	Error mean square; press any key
`SSW=###################`	Factor A row sum of squares; press any key
`dfW=#####`	Factor A row degrees of freedom; press any key
`MSW=###################`	Factor A mean square; press any key
`Fw=####################`	F-statistic for Factor A rows; press any key
`SSC=###################`	Factor B sum of squares; press any key
`dfC=#####`	Factor B degrees of freedom; press any key
`MSC=###################`	Factor B mean square; press any key
`Fc=####################`	F-statistic for Factor B rows; press any key
`SSI=###################`	$A \times B$ interaction sum of squares; press any key
`dfI=#####`	$A \times B$ interaction degrees of freedom; press any key
`MSI=###################`	$A \times B$ interaction mean square; press any key
`Fi=####################`	F-statistic for interaction; press any key
If number of observations per variable is (one) 1:	
`SST=###################`	Total sum of squares; press any key
`dfT=#####`	Total degrees of freedom; press any key
`SSE=###################`	Error sum of squares; press any key
`dfE=#####`	Error degrees of freedom; press any key
`MSE=###################`	Enter mean square; press any key
`SSW=##################`	Factor A row sum of squares; press any key
`dfW=#####`	Factor A row degrees of freedom; press any key
`MSW=###################`	Factor A mean square; press any key
`Fw=####################`	F-statistic for Factor A rows; press any key
`SSC=###################`	Factor B sum of squares; press any key
`dfC=#####`	Factor B degrees of freedom; press any key
`MSC=###################`	Factor B mean square; press any key
`Fc=####################`	F-statistic for Factor B rows; press any key

Screen	Response and Comments
PROGRAM: CONTINGENCY TABLE	
Contingency Table	Program title; press **END LINE**
Inputs:	
No special prompts.	
Outputs:	
df=#####	Degrees of freedom; press any key
Chi^2=#################	Chi-square statistic; press any key
PROGRAM: MANN-WHITNEY U TEST	
Mann-Whitney U Test	Program title; press **END LINE**
Inputs:	
Special prompts.	
X Var. #=0	Enter number of variable to be used as variable X; then, press **END LINE**
Y Var. #=0	Enter number of variable to be used as variable Y; then, press **END LINE**
Output Ranks?	Type **Y** (to output ranks) or **N** (to continue program)
Outputs:	
If **Y** to output ranks:	
X(i)=##################	Sorted X data; press any key
Rank=#####	Rank of sorted X; press any key
Y(i)=##################	Sorted Y data; press any key
Rank=#####	Rank of sorted Y; press any key
If **Y** or **N** to output ranks:	
RSumX=##############	Sum of rankings in sample X; press any key
W(X)=##################	Mann-Whitney statistic for sample X; press any key
Z(X)=##################	Z Statistic for sample X; press any key
RSumY=##############	Sum of rankings in sample Y; press any key
W(Y)=##################	Mann-Whitney statistic for sample Y; press any key
Z(Y)=##################	Z Statistic for sample Y; press any key
RMean=###############	Ranked mean sum; press any key
RankVar=#############	Variance of the ranked mean sum; press any key
PROGRAM: KRUSKAL-WALLIS TEST	
Kruskal-Wallis Test	Program title; press **END LINE**

Screen	Response and Comments
Inputs:	
Special prompts.	
`# of Variables=0`	Enter number of variables; then, press **END LINE**
`Var. # 1=0`	Enter number of variable to be used as first variable; then, press **END LINE** (prompt is repeated for number of variables)
`Output Ranks?`	Type **Y** (to output ranks) or type **N** (to continue program)
Outputs:	
If **Y** to output ranks:	
`RSum i=###############`	Sum of rankings in group i; press any key
`Xj(i)=################`	Sorted data; press any key
`Rank=#####`	Rank of sorted data; press any key
If **Y** or **N** to output ranks:	
`H=#####################`	Kruskal-Wallis H-statistic; press any key
`D=#####################`	Correction term; press any key

PROGRAM: STUDENT'S t-DISTRIBUTION

Screen	Response and Comments
`Student's t-Dist.`	Program title; press **END LINE**
Inputs:	
Data prompts:	
`t=0`	Enter t-statistic value; press **END LINE**
`df=0`	Enter degrees of freedom value; press **END LINE**
Outputs:	
`P(t)=#################`	Left tail area; press any key
`Q(t)=#################`	Right tail area; press any key

PROGRAM: F-DISTRIBUTION

Screen	Response and Comments
`F-Distribution`	Program title; press **END LINE**
Inputs:	
Data prompts:	
`df1=0`	Enter numerator degrees of freedom value; press **END LINE**
`df2=0`	Enter denominator degrees of freedom value; press **END LINE**
`F=0`	Enter F-statistic value; press **END LINE**

Screen	Response and Comments
Outputs:	
P(F)=##################	Left tail area; press any key
Q(F)=##################	Right tail area; press any key
PROGRAM: CHI-SQUARE DISTRIBUTION	
Chi-square Dist.	Program title; press **END LINE**
Inputs:	
Data prompts:	
df=0	Enter degrees of freedom value; press **END LINE**
Chi^2=0	Enter chi-square value; press **END LINE**
Outputs:	
P(Chi^2)=##############	Left tail area; press any key
Q(Chi^2)=##############	Right tail area; press any key
PROGRAM: NORMAL DISTRIBUTION	
Normal Distribution	Program title; press **END LINE**
Inputs:	
Special prompts:	
Compute Q(u)?	Press **Y** (to compute Q(u)) or **N** (to compute u)
Data prompts:	
If **Y** to compute Q(u) question:	
u=0	Enter standard variate; press **END LINE**
If **N** to compute Q(u) question:	
Q(u)=0	Enter right tail area; press **END LINE**
Outputs:	
If **Y** or **N** to compute Q(u) question:	
f(u)=#################	Standard normal density function; press any key
P(u)=#################	Left tail area; press any key
If **Y** to compute Q(u) question:	
Q(u)=#################	Right tail area; press any key
If **N** to compute Q(u) question:	
u=####################	Standard variate; press any key

Screen	Response and Comments
PROGRAM: WEIBULL DISTRIBUTION	
`Weibull Distribution`	Program title; press **END LINE**
Inputs:	
Special prompts:	
`Compute Q(u)?`	Press **Y** (to compute Q(u)) or **N** (to compute u)
Data prompts:	
If **Y** to compute Q(u) question:	
`u=0`	Enter Weibull variate value; press **END LINE**
If **N** to compute Q(u) question:	
`Q(u)=0`	Enter right tail area value; press **END LINE**
If **Y** or **N** to compute Q(u) question:	
`Scale=0`	Enter scale parameter value; press **END LINE**
`Shape=0`	Enter shape parameter value; press **END LINE**
Outputs:	
If **Y** or **N** to compute Q(u) question:	
`f(u)=##################`	Density function; press any key
`P(u)=##################`	Left tail area; press any key
If **Y** to compute Q(u) question:	
`Q(u)=##################`	Right tail area; press any key
If **N** to compute Q(u) question:	
`u=#####################`	Variate; press any key
PROGRAM: EXPONENTIAL DISTRIBUTION	
`Exponential Dist.`	Program title; press **END LINE**
Inputs:	
Special prompts:	
`Compute Q(u)?`	Press **Y** (to compute Q(u)) or **N** (to compute u)
Data prompts:	
If **Y** to compute Q(u) question:	
`u=0`	Enter variate; press **END LINE**

Screen	Response and Comments
If **N** to compute Q(u) question:	
`Q(u)=0`	Enter right tail area; press **END LINE**
If **Y** or **N** to compute Q(u) question:	
`Scale=0`	Enter scale parameter; press **END LINE**
Outputs:	
If **Y** or **N** to compute Q(u) question:	
`f(u)=##################`	Density function; press any key
`P(u)=##################`	Left tail area; press any key
If **Y** to compute Q(u) question:	
`Q(u)=##################`	Right tail area; press any key
If **N** to compute Q(u) question:	
`u=#####################`	Variate; press any key
PROGRAM: BINOMIAL DISTRIBUTION	
`Binomial Distribution`	Program title; press **END LINE**
Inputs:	
Data prompts:	
`n=0`	Enter number of trials value; press **END LINE**
`p=0`	Enter probability value; press **END LINE**
`k=0`	Enter number of successes value; press **END LINE**
Outputs:	
`Mean=##################`	Mean; press any key
`StdDev=###############`	Standard deviation; press any key
`f(k)=##################`	Probability function; press any key
`P(k)=##################`	Cumulative probability; press any key
`Q(k)=##################`	Inverse cumulative probability; press any key
PROGRAM: POISSON DISTRIBUTION	
`Poisson Distribution`	Program title; press **END LINE**
Inputs:	
Data prompts:	
`Mean=0`	Enter expected success parameter value; press **END LINE**
`k=0`	Enter number of successes value; press **END LINE**

Screen	Response and Comments
Outputs:	
f(K)=##################	Distribution probability function; press any key
P(K)=##################	Cumulative probability; press any key
Q(K)=##################	Inverse cumulative probability; press any key

Table A-1. Input restrictions for the AMPI Statistics Library programs

		Restrictions[a]		
Program	**Input**	**Number Type**	**Low Range**	**High Range**
MM	Var. #	Integer	≥ 1	$\leq N$
HIST	Var. #	Integer	≥ 1	$\leq N$
HIST	# of Cells	Integer	≥ 2	$\leq$ MAXREAL
HIST	Width	Any Number	> 0	$\leq$ MAXREAL
HIST	Minimum	Integer	$\geq -\infty$	$\leq \infty$
H:ND	Mean	Any Number	$\geq -\infty$	$\leq \infty$
H:ND	Std. Dev	Any Number	> 0	$\leq \infty$
TTP	Diff. Var. #	Integer	≥ 1	$\leq N$
TTU	Hyp. Diff.	Any number	$\geq -\infty$	$\leq \infty$
TTU	X Var. #	Integer	≥ 1	$\leq N$
TTU	Y Var. #	Integer	≥ 1	$\leq N$
AOV1	# of Variables	Integer	≥ 2	$\leq$ MAXREAL
AOV2	# of Rows	Integer	≥ 2	$< N$
AOV2	No. of Var.	Integer	≥ 4	$\leq N$
AOV2	No. of Obs.	Integer	≥ 1	$\leq \infty$
MW	X Var. #	Integer	≥ 1	$\leq N$
MW	Y Var. #	Integer	≥ 1	$\leq N$

		Restrictions@		
Program	**Input**	**Number Type**	**Low Range**	**High Range**
KW	# of Variables	Integer	≥ 2	$\leq$ MAXREAL
TD	t	Any number	$\geq -\infty$	$\leq \infty$
TD	df	Integer	≥ 1	≤ 200
FD	df1	Integer	≥ 1	≤ 200
FD	df2	Integer	≥ 1	≤ 200
FD	F	Any number	≥ 0	$\leq \infty$
CHI2	df	Integer	≥ 1	≤ 200
CHI2	Chi^2	Any number	≥ 0	$\leq \infty$
ND	u	Any number	$\geq -\infty$	$\leq \infty$
ND	Q(u)	Any number	≥ 0	≤ 1
WD	u	Any number	≥ 0	$\leq \infty$
WD	Q(u)	Any number	≥ 0	≤ 1
WD	Scale	Any number	> 0	$\leq \infty$
WD	Shape	Any number	> 0	$\leq \infty$
ED	u	Any number	≥ 0	$\leq \infty$
ED	Q(u)	Any number	≥ 0	≤ 1
ED	Scale	Any number	> 0	$\leq \infty$
BD	n	Integer	≥ 0	≤ 200
BD	p	Any number	≥ 0	≤ 1
BD	k	Integer	≥ 0	$\leq$ # of successes
PD	Mean	Any number	≥ 0	$\leq \infty$
PD	k	Integer	≥ 0	$\leq$ # of trials
DE**	No. of Var.	Integer	≥ 1	$\leq \infty$
DE**	No. of Obs.	Integer	≥ 1	$\leq \infty$

@ For values of $-\infty$, $+\infty$, and MAXREAL, see your *HP-71B Owner's Manual;* N = current # of Variables in data base; DE = Data Editor.

Appendix B. The Toolbox Approach

This appendix describes the "toolbox approach" to using the subprograms found in this AMPI Statistics Library. This toolbox approach allows you, the user, to personalize this statistical library to your own specifications. In other words, the toolbox approach is designed so that the program is user replaceable, user accessable, and user modifiable.

This appendix describes the subprograms and routines that are available in the AMPI Statistics Library, their syntax, and their calling relationships. The list of variables provided with these routines can be used separately — without using the entire AMPI Statistics Library — to alter the input and output. If so desired, you can easily substitute your own formulas or modify those already used, by saving the subprograms in RAM with the same name. The programs are divided into two sections and stored in two files — STP (i.e., STatistical Programs) and DIST (i.e., DISTributions). The file STP contains all the statistical analysis programs, while the file DIST contains all the statistical distribution programs.

Each of the subprograms describe the required syntax in a CALL statement to the user-accessible subprograms in this AMPI library. The actual variable names that you choose to use are immaterial. It is important to remember that the number, type, and location of the arguments agree with the specifications that follow. The input and output parameter matrices have been standardized to allow similar access to each subroutine. These matrices must be dimensioned by the user and appropriate values stored or initialized.

STP File

The STP file contains 14 subprograms:

MM2	Means and Moments
HIST21	Histogram
HIST22	Histogram
MLR21	Multiple Linear Regression
MLR22	Multiple Linear Regression
TTP2	Paired t-Test
TTU2	Unpaired t-Test
AV12	One-way Analysis of Variance
AV22	Two-way Analysis of Variance
CONT2	Contingency Table
MW2	Mann-Whitney U Test
KW2	Kruskal-Wallis Test
COMP	Rank Comparison Routine
SORT	Column Sorting Routine

Each of these 14 subprograms have four parameters, I, O, X, and S. Here, the parameter I is the input, O is the output, X is the raw data, and S is the summation data. In the following listings, frequency, F, equals one (1) for ungrouped data.

The summary statistical data stored in the parameter S are as follows.

DIM S(8,C) where C = Number of Variables

S(0,0) =	1 for ungrouped; 2 for grouped
S(1,0) =	1 for ungrouped and equal number of observations; 0 for grouped or unequal number of observations
S(2,0) =	1 if all values are positive or zero; 0 if any value is negative
S(3,0) = Σ X	Sum of X for all variables and observations; all frequencies assumed to be 1
S(4,0) = $\sum_{i=1}^{C} S(2,i)$	Sum of X times the frequency for all observations
S(5,0) = $\sum_{i=1}^{C} S(1,i)$	Sum of the frequencies for all observations
S(6,0) = 0	Application flag
S(7,0) = 0	Application error flag

For each of the following, i = 1 to C:

S(0,i) = #Obs	Number of observations
S(1,i) = Σ F	Sum of the frequencies
S(2,i) = Σ FX	Sum of the frequency times X
S(3,i) = Σ F(X-S(8,i))^2	Sum of the frequency times (X-S(8,i))^2
S(4,i) = Σ F(X-S(8,i))^3	Sum of the frequency times (X-S(8,i))^3
S(5,i) = Σ F(X-S(8,i))^4	Sum of the frequency times (X-S(8,i))^4
S(6,i) = Σ F/X	Sum of the frequency divided by X; if X = 0, then INF
S(7,i) = Σ F x LOG10(X)	Sum of the frequency times Log10(X); if X$\leq$0, then INF
S(8,i) = S(1,i)/S(2,i)	Sum of the frequencies for all observations divided by the sum of X times the frequency for all observations (variable i mean)

The MM2 Subprogram

Description: The Means and Moments Subprogram, MM2, calculates the arithmetic mean, harmonic mean, geometric mean, second moment, third moment, fourth moment, kurtosis, and skewness. The MM2 Subprogram appears in file STP. It uses ungrouped or grouped input data.

CALL MM2(X(,),S(,),I(),O())

Inputs:

I(1)	= Var. #	Variable Number
S(1,I(1))	= Σ F	Sum of the frequency
S(2,I(1))	= ΣFX	Sum of the frequency times X

S(3,I(1))	$= \Sigma\ F(X-\overline{\overline{X}})^2$	Sum of the frequency times $(X-\overline{\overline{X}})^2$
S(4,I(1))	$= \Sigma\ F(X-\overline{\overline{X}})^3$	Sum of the frequency times $(X-\overline{\overline{X}})^3$
S(5,I(1))	$= \Sigma\ F(X-\overline{\overline{X}})^4$	Sum of the frequency times $(X-\overline{\overline{X}})^4$
S(6,I(1))	$= \Sigma$ F/X	Sum of the frequency divided by X
S(7,I(1))	$= \Sigma$ F $\times$ LOG10(X)	Sum of the frequency times Log 10(X)
S(8,I(1))	$= \overline{\overline{X}}$	Mean

Set I(3) = 1 for n − 1 df or I(3) = 0 for n df

Outputs:

O(1)	= # Data	Number of observations
O(2)	= A Mean	Arithmetic mean
O(3)	= G Mean	Geometric mean
O(4)	= H Mean	Harmonic mean
O(5)	= Moment 2	Second moment
O(6)	= Moment 3	Third moment
O(7)	= Moment 4	Fourth moment
O(8)	= Kurtosis	Kurtosis
O(9)	= Skewness	Skewness

The HIST21 Subprogram

Description: The Histogram Subprogram, HIST21, performs grouping of data for the HIST22 Subprogram. The HIST21 Subprogram groups data into cells, given: number of cells, minimum acceptable data value, and width of a cell. The HIST21 Subprogram is found in File STP. This subprogram uses ungrouped data which contain one variable.

CALL HIST21(X(,),S(,),I(),O())

Inputs:

DIM O(# of Cells * 3 + 13) Column of X to group

I(1)	= Column #
I(2)	= # of Cells
I(3)	= Width
I(4)	= Minimum
X(0,I(1))	= # of data points in Column I(1)
X(#,I(1))	= Data points

Outputs:

O(10 + 1) = Mean Cell 1

O(10 + 2) = Frequency Cell 1

•

•

•

O(10 + 2(I(2))) = Frequency Cell I(2)

The HIST22 Subprogram

Description: The Histogram Subprogram, HIST22, calculates the theoretical histograms and the chi-square statistic; it uses O(6) to contain the distribution number to be fitted. The HIST22 Subprogram is found in File STP. This subprogram uses grouped data with one variable.

CALL HIST22(X(,),S(,),I(),O())

Inputs:

General:

O(7)	=	# Cells
O(0)	=	Second Moment
O(8)	=	Mean
O(9)	=	# Data
O(10)	=	Width/2
O(6)	=	Distribution to Fit
O(5)	=	Calculate Parameters (1, yes; 2, no)
O(11)	=	Mean Cell 1
O(12)	=	Frequency Cell 1
•		
•		
•		
O(10 + 2(O(7)))	=	Frequency Cell O(7)

For Normal Distribution Fit:

O(6)	=	14
I(1)	=	Mean
I(2)	=	St Dev
O(4)	=	17

For Weibull Distribution Fit:

O(6)	=	15
I(1)	=	Scale
I(2)	=	Shape
O(4)	=	17

For Exponential Distribution Fit:

O(6)	=	16
I(1)	=	Shape
O(4)	=	17

For Binomial Dist Fit:

To Calculate Parameters:

O(6)	=	17
O(5)	=	1
I(1)	=	n

To Not Calculate Parameters:

O(6)	=	17
O(5)	=	0
I(1)	=	n
I(2)	=	p

For Poisson Distribution Fit:

O(6)	=	18
I(1)	=	Mean

For Uniform Distribution Fit:

O(6)	=	19

Outputs:

S(7,0)	=	0, no error; 1, expected frequency equals zero; 2, negative cell mean
O(11+2(O(7)))	=	Expected X Frequency Cell 1
•		
•		
•		
O(10+3(O(7)))	=	Expected X Frequency O(7)
O(11+3(O(7)))	=	df
O(12+3(O(7)))	=	Chi^2
O(13+3(O(7)))	=	Q(Chi^2)

The MLR21 Subprogram

The Multiple Linear Regression Subprogram, MLR21, calculates the upper triangular correlation matrix. This subprogram is located in File STP.

CALL MLR21(X(,),S(,),I(),O())

Inputs:

DIM O(12 + (# Var)(# Obs))

I(1)	=	0	Error Flag
I(2)	=	# Var	# Variable
I(3)	=	# Obs	# Observation
S(8,i)	=		Column i Mean
X(i,j)	=		Data Points

Outputs:

I(1)	=	1, if an error; 0, if not an error

Upper Triangular Matrix Minus Diagonal as Vector:

O(1)	=	R2(1)
•		
•		
•		
O((I(2) − 1)*I(2)/2)	=	R I(2)(I(2) − 1)

The MLR22 Subprogram

Description: The Multiple Linear Regression Subprogram, MLR22. calculates the regression coefficient and the regression statistics. This subprogram also is in File STP.

CALL MLR22(X(,),S(,),I(),O())

Inputs:

I(0)	=	0	Error Flag
I(1)	=	1	Standard Regression Coefficient
		0	Non-Standard Regression Coefficient
I(2)	=	# Var	# Variable
I(3)	=	# Obs	# Observation
S(8,i)	=		Column i Mean
X(i,j)	=		Data points

Upper Triangular Matrix Minus Diagonal as Vector:

O(1) = R2(1)
•
•
•
O((I(2) − 1)∗I(2)/2) = R I(2)(I(2) − 1)

Outputs:

General:

O(1)	=	R^2
O(2)	=	SST
O(3)	=	df/SST
O(4)	=	SSRg
O(4)	=	df/SSRg
O(6)	=	SSE
O(7)	=	df/SSE
O(8)	=	SE
O(9)	=	F
I(0)	=	0, Not an Error 1, Error

Standard Regression Coefficient (if I(1) = 1):

I(1)	=	1
O(13)	=	b(1)
O(14)	=	SEb(1)
•		
•		
•		
O(14 + I(2) − 1)	=	SEb(I(2) − 1)

Non-Standard Regression Coefficient (if I(1) = 0):

I(1)	=	0
O(10)	=	B0
O(13)	=	B(1)
O(14)	=	SE B(1)
•		
•		
•		
O(14 + I(2) − 1)	=	SE B(I(2) − 1)

The TTP2 Subprogram

Description: The Paired t-Test Subprogram, TTP2, calculates the number of observations, mean, standard deviation, degrees of freedom, t-statistic, mean difference, and standard error. The TTP2 Subprogram appears in file STP. It uses only ungrouped input data.

CALL TTP2(X(,),S(,),I(),O())

Inputs:

I(1)	= Diff. Var. #	Variable number which contains the differences between X and Y
S(1,I(1))	= Σ F	Sum of the frequency
S(2,I(1))	= ΣFX	Sum of the frequency times X
S(3,I(1))	= Σ F$(X-\overline{X})^2$	Sum of the frequency times $(X-\overline{X})^2$

Outputs:

O(1)	= # Data	Number of observations
O(2)	= Mean	Arithmetic mean for X minus Y
O(3)	= Std Dev	Standard deviation for X minus Y
O(4)	= df	Degrees of freedom
O(5)	= t	t-Statistic
O(6)	= SE	Standard error

The TTU2 Subprogram

Description: The Unpaired t-Test Subprogram, TTU2, calculates the number of observations for X and Y, mean for X and Y, standard deviation for X and Y, t-statistic, pooled variance, and standard error. The TTU2 Subprogram appears in file STP. It uses ungrouped and grouped input data.

CALL TTU2(X(,),S(,),I(),O())

Inputs:

I(1)	= Hyp. Diff	Hypothesized difference between means of two samples
I(2)	= Var. X	Number of the variable selected as X
I(3)	= Var. Y	Number of the variable selected as Y
S(1,I(2))	= Σ F	Sum of the frequencies for X
S(1,I(3))	= Σ F	Sum of the frequencies for Y
S(2,I(2))	= ΣFX	Sum of the frequency times X
S(2,I(3))	= ΣFY	Sum of the frequency times Y
S(3,I(2))	= Σ F$(X-\overline{X})^2$	Sum of the frequency times $(X-\overline{X})^2$
S(3,I(3))	= Σ F$(Y-\overline{Y})^2$	Sum of the frequency times $(Y-\overline{Y})^2$

Outputs:

O(1)	= X# Data	Number of observations for X
O(2)	= XMean	Arithmetic mean for X
O(3)	= XStd Dev	Standard deviation for X
O(4)	= Y# Data	Number of observations for Y
O(5)	= YMean	Arithmetic mean for Y
O(6)	= YStd Dev	Standard deviation for Y
O(7)	= df	Degrees of freedom
O(8)	= t	t-statistic
O(9)	= Pool Var.	Pooled variance
O(10)	= SE	Standard error

The AV12 Subprogram

Description: The One-way Analysis of Variance Subprogram, AV12, calculates the total sum of squares, total degrees of freedom, within (i.e., between) groups sum of squares, within (i.e., between) groups degrees of freedom, within (i.e., between) groups mean square, error sum of squares, error degrees of freedom, error mean square, and F-statistic. The AV12 Subprogram appears in file STP. It uses ungrouped and grouped input data.

CALL AV12(X(,),S(,),I(),O())

Inputs:

I(1)	= # of Variables	Number of variables used, n
I(2)	= Var. #1	Number of variable to be used as variable number one
•		
•		
•		
I(n + 1)	= Var. #n	Number of the variable to be used as last variable number (i.e., equals n)
S(1,I(j))	= Σ F	Sum of the frequencies, where where $j = 2,\ldots,n + 1$
S(2,I(j))	= ΣFX	Sum of the frequency times X, where $j = 2,\ldots,n + 1$
S(3,I(j))	= $\Sigma F(X - \overline{X})^2$	Sum of the frequency times $(X - \overline{X})^2$, where $j = 2,\ldots,n + 1$
S(8,I(j))	= $\overline{X}$	Mean
X(0,0)	= 1 or 2	If 1, ungrouped; if 2, grouped
X(0,I(j))	= # of points	Number of points

Outputs:

O(1)	= SST	Total sum of squares
O(2)	= dfT	Total degrees of freedom
O(3)	= SSW	Within (i.e., between) row groups sum of squares
O(4)	= dfW	Within (i.e., between) row groups degrees of freedom
O(5)	= MSW	Within (i.e., between) row groups mean square

O(6)	= SSE	Error sum of squares
O(7)	= dfE	Error degrees of freedom
O(8)	= MSE	Error mean square
O(9)	= F	F-statistic

The AV22 Subprogram

Description: The Two-way Analysis of Variance Subprogram, AV22, calculates the total sum of squares, total degrees of freedom, error sum of squares, error degrees of freedom, error mean square, factor A sum of squares, factor A degrees of freedom, factor A mean square, F-statistic for factor A, factor B degrees of freedom, factor B sum of squares, factor B mean square, F-statistic for factor B, $A \times B$ interaction sum of squares, $A \times B$ interaction degrees of freedom, $A \times B$ interaction mean square, and F-statistic for $A \times B$ interaction. The AV22 Subprogram appears in file STP. It uses only ungrouped input data with and without replication.

CALL AV22(X(,),S(,),I(),O())

Inputs:

I(1)	= # of Rows	Number of rows
I(2)	= # of Var.	Number of variables
I(3)	= # of Obs.	Number of observations

Outputs:

With and Without Replication:

O(1)	= SST	Total sum of squares
O(2)	= dfT	Total degrees of freedom
O(3)	= SSE	Error sum of squares
O(4)	= dfE	Error degrees of freedom
O(5)	= MSE	Error mean square
O(6)	= SSW	Factor A sum of squares
O(7)	= dfW	Factor A degrees of freedom
O(8)	= MSW	Factor A mean square
O(9)	= Fw	F-statistic for Factor A
O(10)	= SSC	Factor B sum of squares
O(11)	= dfC	Factor B degrees of freedom
O(12)	= MSC	Factor B mean square
O(13)	= Fc	F-statistic for Factor B

With Replication:

O(14)	= SSI	$A \times B$ interaction sum of squares
O(15)	= dfI	$A \times B$ interaction degrees of freedom
O(16)	= MSI	$A \times B$ interaction mean square
O(17)	= Fi	F-statistic for interaction

The CONT2 Subprogram

Description: The Contingency Table Subprogram, CONT2, calculates the degrees of freedom and chi-square statistic. The CONT2 Subprogram appears in file STP. It uses only ungrouped input data.

CALL CONT2(X(,),S(,),I(),O())

Inputs:

I(1)	= # of Columns	Number of matrix columns
I(2)	= # of Rows	Number of rows in matrix
S(1,0)	= 1	
S(2,0)	= 1	
X(i,j)	= Raw data	
X(#,0)	= Σ Col	Sum of the rows
S(2,#)	= Σ X	Sum of the columns
S(3,0)	= Σ Σ X	Sum of the rows and columns

Outputs:

O(1)	= df	Degrees of freedom
O(2)	= Chi^2	Chi-square statistic

The MW2 Subroutine

Description: The Mann-Whitney Subroutine, MW2, determines the sorted X and Y data, the ranks of each of these sorted X and Y data, the sum of rankings for sample X and Y, the Mann-Whitney statistic for sample X and Y, the Z statistic for sample X and Y, the ranked mean sum, and the variance of the ranked mean sum. The MW2 Subroutine is located in File STP. This subroutine uses ungrouped data which contain two variables. The subroutine COMP does the ranking.

CALL MW2(X(,),S(,),I(),O())

Inputs:

I(0)	= 0
I(1)	= Variable 1, X
I(2)	= Variable 2, Y
X(0,I(1))	= # of Observations for Variable 1, X
X(0,I(2))	= # of Observations for Variable 2, Y

DIM O(9+X(0,I(1))+X(0,I(2))) Output for ranks

Outputs:

O(1)	= R Sum X	Sum of rankings for X
O(2)	= W(X)	Mann-Whitney statistic for X
O(3)	= Z(X)	Z-statistic for X
O(4)	= R Sum Y	Sum of rankings for Y
O(5)	= W(Y)	Mann-Whitney statistic for Y
O(6)	= Z(Y)	X-statistic for Y
O(7)	= R	Mean Ranked mean sum

O(8) = Rank Var   Variance of the rankedmean sum
O(9+1) =   Rank of X, Obs 1
•
•
•
O(9+X(0,I(1))) =   Rank of X, Obs X(0,I(2))
•
•
•
O(9+X(0,I(1))+X(0,I(2))) = Rank of Y, Obs X(0,I(2))

The KW2 Subroutine

Description: The Kruskal-Wallis Subroutine, KW2, determines the sorted data, the ranks of each of these sorted data, the sum of the rankings, the Kruskal-Wallis H-statistic, and the correction term. The KW2 Subroutine is located in File STP. This subroutine uses ungrouped data which contain two or more variables. The subroutine COMP does the ranking.

CALL KW2(X(,),S(,),I(),O())

Inputs:

I(0) = 1
I(1) = # of Variables, n
I(2) = Variable 1
I(3) = Variable 2
•
•
•
I(I(1)+1) = Variable n
X(0,I(2)) = # of observations for Variable 1
X(0,I(3)) = # of observations for Variable 2
•
•
•
X(0,I(I(1)+1)) = # of observations for Variable n

Outputs:

O(1) = H   Kruskal-Wallis H-statistic
O(2) = D   D-value or correction term
O(4) =   Ranked sum of Variable 1
O(4+1) =   Rank of Var 1, Obs 1
•
•
•
O(4+X(0,I(2))) =   Rank of Var 1, Obs X(0,I(2))
•
•
•

O(4 + X(0,I(2)) + 1) = Ranked sum of Var 2
O(4 + X(0,I(2)) + 2) = Rank of Var 2, Obs 1

•
•
•

O(4 + X(0,I(2)) + 1 + X(0,I(3))) = Rank of Var 2, Obs X(0,I(3))

•
•
•

$$O(4 + n - 1 + \sum_{j=1}^{n} X(0,I(I(1)+j))) = \text{Rank of Var n, Obs } X(0,I(1+n))$$

The COMP Subprogram

Description: The Ranking Subprogram, COMP, determines the ranking for the MW2 and KW2 Subprograms. The COMP Subprogram is located in File STP. This subprogram uses ungrouped data which contain two or more variables.

CALL COMP(X(,),S(,),I(),O())

Inputs:

I(0) = 0 Called by the MW2 Subroutine
I(0) = 1 Called by the KW2 Subroutine

For Mann-Whitney:

I(1) = Variable 1, X
I(2) = Variable 2, Y
X(0,I(1)) = # of Observations for Variable 1, X
X(0,I(2)) = # of Observations for Variable 2, Y

For Kruskal-Wallis:

I(1) = # of Variables, n
I(2) = Variable 1
I(3) = Variable 2

•
•
•

I(I(1) + 1) = Variable n
X(0,I(2)) = # of observations for Variable 1
X(0,I(3)) = # of observations for Variable 2

•
•
•

X(0,I(I(1) + 1)) = # of observations for Variable n

Outputs:

For Mann-Whitney:

$$\begin{array}{ll}
O(9+1) & = \text{Rank of X, Obs 1} \\
\quad\vdots & \\
O(9+X(0,I(1))) & = \text{Rank of X, Obs X(0,I(2))} \\
\quad\vdots & \\
O(9+X(0,I(1))+X(0,I(2))) & = \text{Rank of Y, Obs X(0,I(2))}
\end{array}$$

For Kruskal-Wallis:

$$\begin{array}{ll}
O(4+1) & = \text{Rank of Var 1, Obs 1} \\
\quad\vdots & \\
O(4+X(0,I(2))) & = \text{Rank of Var 1, Obs X(0,I(2))} \\
\quad\vdots & \\
O(4+X(0,I(2))+2) & = \text{Rank of Var 2, Obs 1} \\
\quad\vdots & \\
O(4+X(0,I(2))+1+X(0,I(3))) & = \text{Rank of Var 2, Obs X(0,I(3))} \\
\quad\vdots & \\
O(4+n-1+\sum_{j=1}^{n} X(0,I(I(1)+j))) & = \text{Rank of Var n, Obs X(0,I(1+n))}
\end{array}$$

The SORT Subprogram

Description: The Column Sorting Subprogram, SORT, does sorting for a column of X. The SORT Subroutine is located in File STP. This subprogram uses ungrouped or grouped data. The method used by SORT is the Shell-Metzner sorting algorithm.

CALL SORT(X(,),S(,),I(),O())

Inputs:

I(1)	= Column #	Column number to be sorted
X(0,I(1))	= 1	Ungrouped data
	2	Grouped data

Outputs:

The column specified by I(1) is sorted in increasing order and replaces column I(1) of X.

Toolbox Example

The following example gives an indication as to how the toolbox approach may be used to create a separate program. The example shows a paired t-test program using the toolbox.

```
10 DESTROY ALL
20 INPUT '# OF POINTS=';I5
30 DIM I(1),O(6),X(I5,1),S(8,1)
40 I(1)=1
50 FOR I4=1 TO I5
60 INPUT 'X,Y=';A1,A2
70 X(I4,1)=A1-A2
80 S(1,1)=S(1,1)+1
90 S(2,1)=S(2,1)+X(I4,1)
100 NEXT I4
110 FOR I4=1 TO I5
120 S(3,1)=S(3,1)+(X(I4,1)-S(2,1)/S(1,1))^2
130 NEXT I4
140 CALL TTP2(X(,),S(,),I(),O())
150 PRINT '# DATA=';O(1)
160 PRINT 'MEAN=';O(2)
170 PRINT 'STD DEV=';O(3)
180 PRINT 'df=';O(4)
190 PRINT 't=';O(5)
200 PRINT 'SE=';O(6)
210 END
```

DIST File

The DIST file contains seven subprograms (one subprogram has two distributions):

TD2	Student's t-Distribution
FD2	F-Distribution
CHI22	Chi-square Distribution
ND2	Normal Distribution
WD2	Weibull and Exponential Distributions
BD2	Binomial Distribution
PD2	Poisson Distribution

Each of these seven subprograms uses two arrays, I() and O(). Here, the array I() is the input; O() is the output.

The TD2 Subprogram

Description: The Student's t-Distribution Subprogram, TD2, calculates the right tail area and the left tail area. The TD2 Subprogram appears in file DIST.

CALL TD2(I(),O())

Inputs:

I(1)	= t	t-Statistic
I(2)	= df	Degrees of freedom

Outputs:

O(1)	= P(t)	Left tail area
O(2)	= Q(t)	Right tail area

The FD2 Subprogram

Description: The F-distribution Subprogram, FD2, calculates the right tail area and the left tail area. The FD2 Subprogram appears in file DIST.

CALL FD2(I(),O())

Inputs:

I(1)	= df1	Numerator degrees of freedom
I(2)	= df2	Denominator degrees of freedom
I(3)	= F	F-statistic

Outputs:

O(1)	= P(F)	Left tail area
O(2)	= Q(F)	Right tail area

The CHI22 Subprogram

Description: The Chi-square Distribution Subprogram, CHI22, calculates the right tail area and left tail area. The CHI22 Subprogram appears in file DIST.

CALL CHI22(I(),O())

Inputs:

I(1)	= df	Degrees of freedom
I(2)	= Chi^2	Chi-square statistic

Outputs:

O(1)	= P(Chi^2)	Left tail area
O(2)	= Q(Chi^2)	Right tail area

The ND2 Subprogram

Description: The Normal Distribution Subprogram, ND2, calculates the density function, the right tail area, and either the standard variate or the left tail area. The ND2 Subprogram appears in file DIST. If O(4) = 17, the subprogram computes Q(u); otherwise, it computes u.

CALL ND2(I(),O())

Inputs:

I(1)	= u	Standard variate (for O(4) = 17)
I(1)	= Q(u)	Right tail area (for O(4) $\neq$ 17)

Outputs:

For O(4) = 17:

O(1)	= f(u)	Standard normal distribution function
O(2)	= P(u)	Left tail area
O(3)	= Q(u)	Right tail area

For O(4) $\neq$ 17:

O(1)	= f(u)	Standard normal distribution function
O(2)	= P(u)	Left tail area
O(3)	= u	Standard variate

The WD2 Subprogram

Description: The Weibull Distribution and Exponential Distribution Subprogram, WD2, calculates the density function, right tail area, left tail area, and standard variate. The WD2 Subprogram appears in file DIST. If O(4) = 17, the subprogram computes Q(u); otherwise, it computes u.

CALL WD2(I(),O())

Inputs:

For O(4) = 17:

I(1)	= u	Standard variate
I(2)	= Scale	Scale parameter
I(3)	= Shape	Shape parameter (Note: For Exponential Distribution, I(3) = 1)

For O(4) $\neq$ 17:

I(1)	= Q(u)	Right tail area
I(2)	= Scale	Scale parameter
I(3)	= Shape	Shape parameter (Note: For Exponential Distribution, I(3) = 1)

Outputs:

For O(4) = 17:

O(1)	= f(u)	Density function
O(2)	= P(u)	Left tail area
O(3)	= Q(u)	Right tail area

For O(4) $\neq$ 17:

O(1)	= f(u)	Density function
O(2)	= P(u)	Left tail area
O(3)	= u	Standard variate

The BD2 Subprogram

Description: The Binomial Distribution, BD2, calculates the mean, standard deviation, probability function, right tail area, and left tail area. The BD2 Subprogram appears in file DIST.

CALL BD2(I(),O())

Inputs:

I(1)	= n	Number of trials
I(2)	= p	Probability
I(3)	= k	Number of successes

Outputs:

O(1)	= Mean	Arithmetic mean
O(2)	= Std Dev	Standard deviation
O(3)	= f(k)	Probability function
O(4)	= P(k)	Left tail area
O(5)	= Q(k)	Right tail area

The PD2 Subprogram

Description: The Poisson Distribution Subprogram, PD2, computes the distribution probability function, right tail area, and left tail area. The PD2 Subprogram appears in file DIST.

CALL PD2(I(),O())

Inputs:

I(1)	= Mean	Expected success parameter
I(2)	= k	Number of successes

Outputs:

O(1) =	f(k)	Distribution probability function
O(2) =	P(k)	Left tail area
O(3) =	Q(k)	Right tail area

Toolbox Examples

The following examples give an indication as to how the toolbox approach may be used to create separate programs. The first example shows a normal distribution program using the toolbox with flags to calculate Q(u) given u. Note the use of O(4) = 17 in line 40 to control whether the normal distribution program calculates u or Q(u).

```
20 DIM I(1),O(4)
30 INPUT 'u:';I(1)
40 O(4)=17
50 CALL ND2(I(),O())
60 PRINT 'f(u)=';O(1)
70 PRINT 'P(u)=';O(2)
80 PRINT 'Q(u)=';O(3)
90 END
```

The next example presents a Student's t-distribution program using the toolbox without flags.

```
10 DIM I(2),O(2)
20 INPUT 't';I(1)
30 INPUT 'df';I(2)
40 CALL TD2(I(),O())
50 PRINT 'P(t)=';O(1)
60 PRINT 'Q(t)=';O(2)
70 END
```

The following example gives a Student's t-distribution program using the toolbox as a subprogram.

```
>CALL ST(2,5,C,D)

10 SUB ST(A,B,C,D)
20 DIM I(2),O(2)
30 I(1)=A @ I(2)=B
40 CALL TD2(I(),O())
50 C=O(1) @ D=O(2)
60 END SUB
```

The final example shows a Student's t-distribution program using the toolbox as a function.

```
>FNT(2,5)

10 DEF FNT(A,B)
20 DIM I(2),O(2)
30 I(1)=A @ I(2)=B
40 CALL TD2(I(),O())
50 FNT=O(1)
60 END DEF
```

Appendix C. Program Error/Status Message Summary

The AMPI Statistics Library programs display certain messages under specific conditions. Some of these are only status messages. Other messages occur in response to an error. For example, if you are using the AMPI Statistics Library as a toolbox, an incorrectly typed or constructed command will produce an error. Also an error in a subprogram may produce an error message and halt execution of the program.

Errors in subprograms are usually related to bad data that the program was unable to detect before entering the routine. To prevent loss of data and allow recovery from the subprogram level, use the END statement when in the suspend mode and then **f CONT** to return to the AMPI Statistics Library. This produces garbage results, but in most cases will prevent loss of data. However, in the event of an error such as in the SUM routine, no recovery is possible. Thus, it is recommended that data be saved in RAM or an external device prior to running the statistics to prevent loss of data.

Several of the error messages are related to the amount of available memory in the HP-71B. The AMPI Statistics Library contains tests for low memory conditions. If you encounter an error or warning that refers to low memory conditions, you should interrupt the program by pressing **ATTN**, catalog the memory files, and purge unneeded files to make more memory available. Then, to proceed with the program, press **f** and then **CONT**.

This appendix summarizes the error and status messages found within the AMPI Statistics Library. The user is referred to pages 378 to 393 of the *HP-71 Reference Manual* for a complete listing of errors, warnings, and system messages related to the HP-71B computer. For a description of error and warning conditions, refer to Section 9, pages 162 to 177, of your *HP-71 Owner's Manual*. For messages related to devices that can be connected to the HP-71B, please refer to the user manuals for those devices.

Error/Status Message	Explanation/Correction Procedure
`Array Too Large`	The data array is too large and there is not enough memory in the HP-71B to hold it. Reduce the size of the data array or purge any unnecessary files or programs from the memory. The program goes back to the main menu.
`Done`	The program has ended and data have been cleared from the memory. The HP-71B is now ready for the next task. Pressing any key wil restore the BASIC prompt. To turn off the HP-71B, press **f** and then **OFF**. To restart the AMPI Statistics Library, press **RUN**.
`Expected Freq=0`	The expected frequency is equal to zero.

Error/Status Message	Explanation/Correction Procedure
File Not Found	The file specified in the load operation was not found. Enter valid filename or specify correct source for file.
Insolvable Matrix	The matrix is not solvable (i.e., singular or near singular).
Illegal Entry	The entry value is illegal (i.e., out of range). See Appendix A for valid input and their ranges.
Invalid Address	The given array element in a GOTO command does not exist. Program automatically returns to the previous position.
Invalid Compute	Compute statement is not a valid BASIC expression. Compute statement must follow BASIC syntax restrictions and variables must be in the form X# = , where the pound sign is the number of the desired variable. The Histogram Program cannot use a calculated parameter for the theoretical histogram because of an invalid value, such as p > 1.
Invalid Data Format	The format for the input data file being loaded (i.e., read) is incorrect (i.e., mismatch between grouped and ungrouped data).
Invalid Filespec	The filename specified is not valid. See your owner's manual for proper filename format.
Invalid Input	The input value is invalid. See Appendix A for input data restrictions.
Invalid Label	The label entered is invalid (i.e., too long). Enter a valid label; a valid label contains ≤ 10 characters or symbols. Null labels are not allowed.
Invalid Negative Data	Occurs in the Histogram Program when a cell mean is less than zero. Occurs in the Contingency Table Program if a negative X value exists.
Invalid # of Var.	The file being added to the data array has a different number of variables from that in the array. Correct the number of variables in the file so as to match the number of variables in the array by adding or deleting variables.

Error/Status Message	Explanation/Correction Procedure
Loading...	The HP-71B is loading your data. No action is necessary.
NaN	Result of invalid computation. Acronym for Not-a-Number.
No Data	The data matrix is empty; data are not available for the Library's procedures. The procedure selected requires data to be entered by the keyboard or load option before the procedure can be executed.
Nonexistent Obs.	The observation number entered in the ADD, DELETE, or DEFine command does not exist. Review observations and enter a valid observation number.
Nonexistent Var.	The variable number entered in the ADD, DELete, or DEFine command does not exist. Review variables and enter a valid variable number.
No Printer	Occurs on Print option if no printer is attached to HP-71B.
Not Enough Mem	The HP-71B does not have enough memory to edit the data or to run the program. Delete unwanted files or programs from memory. To do this, pause the program by pressing **ATTN**, free the memory, then press **f CONT** to resume.
Not HPAF File	The filename specified in the Load option is not in the HPAF format. Refer to Appendix D for correct HPAF format specifications.
Printing...	The HP-71B is printing. No action is necessary.
Saving...	The HP-71B is saving your data. No action is necessary.
Working...	The HP-71B is calculating. No action is necessary.
You Need 1 Obs.	The variable you specified for deletion of an observation contains only one observation. You are not allowed to delete the last observation of a variable using the Delete Obs. procedure. To delete the variable, use the Delete Var. procedure.

Error/Status Message	Explanation/Correction Procedure
You Need 1 Var.	The data set contains only one variable. You are not allowed to delete the last variable from a data set with the Delete Var. procedure. To clear all data, use the Data Keyboard option or the Clear option in the load procedure, or exit and restart the program.

Appendix D. Applications File Format (HPAF) Specifications

The AMPI Statistics Library stores data in files according to a prescribed format — the Hewlett-Packard Applications File format (HPAF). HPAF allows exchange of data between various Hewlett-Packard programs (e.g., *HP-71 Curve Fitting Pac*). The HPAF format provides room for information that describes the structure of the data so that various programs may make use of and exchange the data.

HPAF files are type DATA (see pages 246 through 250 of the *HP-71 Owner's Manual*). They may reside in either the HP-71B memory or a mass storage device.

The HPAF files are composed of three major sections — the header, the data records, and an optional descriptor block. An example of such a file as created by the AMPI Statistics Library is described in the following table.

Record	Contents	Data Type@	Comments
Direct Standard Header:			
0	HPAFNN...N	S	HPAF and N for # of Variables
1	(# Obs)(G)	N	G = 1 Ungrouped; 2 Grouped
2	(# Obs)(G)+3	N	G = 1 Ungrouped; 2 Grouped
Direct Data:			
3	X1(1), X2(1), ..., Xn(1)	N	Values for observation 1
•	X1(F1), X2(F1), ..., Xn(F1)	N	Frequencies for observation 1 if grouped data
•	•		
•	•		
•	•		
•	•		
•	X1(j), X2(Fj), ..., Xn(j)	N	Values for observation n
(# Obs)(G)+2	X1(Fj), X2(Fj), ..., Xn(Fj)	N	Frequencies for observation n if grouped data
Sequential Descriptors:			
(# Obs)(G)+3	"DATATYPE"	S	Data type field
	1	N	
	G	S	G = 1 Ungrouped; 2 Grouped
	"LABELS"	S	Label field

#VAR	N	Number of variables
"LABEL1"	S	Label 1 field
"LABEL2"	S	Label 2 field
•		
•		
•		
"LABEL#VAR"	S	Label # Var field
"#DATA"	S	Number of data field
#VAR	N	Number of variables
"#ENTRIES1"	S	Number of data points or groups for variable 1
•		
•		
•		
"#ENTRIES#VAR"	S	Number of data points or groups for variable n

@ N = Number; S = String.

Header Information

The header must contain the following items:

Record	Description
0	Record 0 contains a type string. The first four characters indicate the file is an HPAF file. The remaining characters describe the number of data items in each record and their type. For example, in "HPAFNNS" the characters "NNS" indicate that there are three items in each record (the first two are numbers and the third a string), with a maximum of 400 items.
1	Record 1 contains the number of data records that contain information.. This number can be less that the total available records (allowing room for additional records to be added later followed by the optional descriptor block).
2	Record 2 contains the record of the optional block. If no descriptor block is present, this number should be zero.

Data Records

The data records start at record 3 and must end before the descriptor block. Note that all items for each record must fit within each logical record so that any record can be accessed randomly. To compute the optimal logical record length for the file, remember that each number written in the record occupies 8 bytes, and each string occupies 3 bytes plus the number of bytes in the string. In addition, 1 byte is reserved in the record length for the carriage return character. For example, if each record is going to hold two numbers and a 10 character string, the record length must be at least 30 bytes (i.e., $2 \times 8 + 3 + 10 + 1$ bytes). For more information about creating data files, refer to "Data Files," in section 14 of the *HP-71 Owner's Manual*.

Descriptor Block

The descriptor block is optional. If present, the descriptor block must come after the data records, and record 2 must contain the record number of the first item in the block. Information in the descriptor block consists of: (1) tags or string, which identify the type of information that follows, (2) followed by the number of items associated with the tag, and (3) followed by the items themselves stored as strings.

TAG,number of items,string item one, string item two,...

The information in the descriptor block can be written serially, or, if the logical record size is sufficiently large, written one tag to a record. In either case, the descriptor block must be able to be read serially.

The AMPI Statistics Library Files

The HP-71B AMPI Statistics Library can read any HPAF file, even if it has been generated by another program. The Library assumes zero descriptor blocks and ungrouped data. String-type data items will be ignored automatically. For instance, if a file has a string NSSNSNN, columns 1, 4, 6, and 7 will be considered the first, second, third, and fourth columns in the data array.

Appendix E. File Names Used in This Library

This appendix contains a list of the file names used in this library. All of these files can be copied to main memory, assuming that you have enough memory available. Most of these files are in ROM and, since your HP-71B searches main RAM first when it looks for a file name, be sure not to have any files with the following names in main RAM when you are using the AMPI Statistics Library module. To do so may destroy any file(s) with the same name(s) already in main RAM or cause substitution of routines or program errors.

File Name	File Type	Description
AMPISTAT	BASIC	Contains the program for running the AMPI Statistics Library.
STP	BASIC	Contains the programs for the statistical tests.
DIST	BASIC	Contains the programs for the statistical distributions.
AMPILEX	LEX	Contains the keywords KEYWAIT$ and MSG$, the message table used by the AMPI Statistics Library.
STKEYS*	KEY	Saves STKEYZ while the AMPI Statistics Library is in the data editor.
STKEYZ	KEY	Contains key definitions for the AMPISTAT array editor.
USERKEYS*	KEY	Saves previously user-defined keys while the library is running.

* A reserve file created in RAM when AMPISTAT is running. If you already have a file with this name in RAM, it will be destroyed when you run AMPISTAT.

Appendix F. Glossary

The science of statistics has developed many specialized words, terms, and jargon. This glossary is designed to assist the user in reading this AMPI Statistics Library. Also, this glossary may be used as a reference manual while working with general statistics. The definitions contained here were derived from the textbooks and other references listed in Appendix G. Additional information concerning the items defined in this glossary and words not listed may be found in the text of this library and in the references found in Appendix G.

A

Abscissa: The horizontal line, or X axis, on a plane; it also is the line perpendicular to the Y axis which represents a zero value for Y.

Accuracy: The closeness of a measured or computed value to its true value.

ANOVA: An acronym denoting analysis of variance.

Arithmetic Mean: The mean (or average) of the sample data; it provides an unbiased estimate of the parametric mean; designated by $\overline{X}$.

Assumption: The supposition that something is true.

Attributes: Those variables that cannot be measured, but must be expressed qualitatively.

B

Bimodal: A frequency curve having two maxima or modes.

Binomial: A mathematical expression consisting of two terms connected by either a plus or minus sign (e.g., $q + p$). For this Library, it is a discrete probability distribution in which the probability of an event happening exactly k times in n trials is given by:

$$p(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

where $k = 0, 1, 2, 3, \ldots, n$ and $n! = n(n-1)(n-2)\ldots(1)$. Also, $p(k)$ for each $k = 0, 1, 2, 3, \ldots, n$ corresponds to successive terms in the binomial expansion formula:

$$(q + p)^n = q^n + nC_1q^{n-1}p + nC_2q^{n-2}p^2 + \ldots p^n$$

where 1, nC_1, nC_2, . . . are the binomial coefficients (Spiegel 1961).

Biometry: The application of statistical methods to the solution of biological problems; also called biostatistics.

C

Case: All the data associated with one specific parameter being measured.

Central Tendency: A measurement which indicates where the middle is located.

Chi-Square: The function whose value is the sum, over all data, of the squares of the weighted differences $(Y - F)/W$ of the dependent variables Y and the model F. If the Y's are normally distributed about the mean F and variance X^2, then chi-square is $\chi^2(\nu)$ distributed with ν (the number of data points minus the number of model parameters) degrees of freedom. Chi-square is denoted symbolically as χ^2.

Continuous Variables: Those variables which, at least theoretically, can assume an infinite number of values between any two fixed points.

Curve: A one-dimensional continuum of points.

D

Data: The facts you collect during observations; used as a plural noun.

Datum: A single fact collected during observations; the singular of data.

Degrees of Freedom: The quantity n or $n - 1$, where n is the sample size upon which a variance has been based; it ranges from zero to infinity; it is usually designated by the abbreviation df.

Dependent Variable: The measured or Y value that depends on the independent variables (X's) in the data set.

Derived Variables: Those variables which are based on two or more independently measured variables whose relations are expressed by a certain term (e.g., ratio, percentage).

Discontinuous Variables: Those variables that have only certain fixed numerical values, with no intermediate values possible in between; also known as meristic or discrete variables.

Distribution: The relative frequency in a finite or infinite population with which a variable quantity (or variate) assumes a particular value; the dispersion of measurement about one central point; also the arrangement of a set of numbers classified according to some property (e.g., classified by frequency, time, size, or location).

E

Enumeration Data: Attributes combined with frequencies into tables for statistical analysis.

Experimental Error: A measure of the variation which exists among observations on experimental units that are treated alike.

Exponential: Something related to an exponent, generally, related to 10 raised to some exponent x; also, involving a variable or unknown quantity.

F

F-Statistic: The ratio of two variances.

G

Geometric Mean: The retransformed mean of the logarithmically transferred variables.

Grand Mean: The average of the sample means.

Grouped Data: Data organized and summarized into classes or categories; data which tell you the number of cases with the same actual value.

H

Harmonic Mean: The reciprocal of the arithmetic mean of reciprocals.

HPAF Format: Hewlett-Packard Application File format. It refers to a standard DATA file format used by HP-71B application pacs.

I, J

Independent Variable: A variable controlled by the experimenter. It usually represents a variable whose value is selected rather than measured.

Individual Observations: Measurements taken on the smallest sampling unit.

K

Kurtosis: The degree of peakedness (shape) in a distribution.

L

Left Tail Area: The cumulative distribution function.

Leptokurtic: A distribution whose curve has a relatively high peak; there are more items near the mean and at the tails, with fewer items in the intermediate regions.

M

Mean: The center of a distribution, with the center being the sum of all scores in the distribution divided by the total number of scores; the arithmetic average of your measurements; a measure of central tendency.

Measurement Variables: Those characters which can be expressed in a numerically ordered fashion.

Median: That value of the variable in an ordered array that has an equal number of items on either side of it; that value which divides a frequency distribution into two halves.

Mode: The value that occurs most frequently in a given series.

M Parameter: For the Poisson Distribution Program, it is the number m of successes which should occur in a given sample size; that is, the sample size N multiplied by the probability of success p; or, it is the mean rate λ at which successes occur in a process times a given period of time t.

N

Non-parametric Tests: Statistical testing techniques that are not dependent on a given distribution (e.g., the normal distribution in the case of the analysis of variance).

Normal Distribution: A distribution in which the data points are expected to lie on the middle line and those points that are not on the middle line are expected to be equally distributed on either side of the middle line; those values on the right of the middle line are positive, while those on the left are negative; also, a continuous frequency distribution whose graphic representation is a bell-shaped curve that is symmetrical about the mean (the standard form has a mean of 0 and a variance of 1).

O

Ordinate: The vertical line, or Y axis, on a plane; also, the line perpendicular to the X axis which represents a zero value for X.

P, Q

Paired Measurements: Measurements involving two or more scores for each case.

Parameter: A variable which can be measured quantitatively; sometimes, an arbitrary constant. One of the unknown values that determine a model.

Parametric Mean: The true mean of a sample which locates the center of the sample population; designated by the lowercase Greek letter mu, μ.

Parametric Tests: Statistical tests that are based on normal distributions.

Platykurtic: A distribution having a curve which is flat topped; there are fewer items at the mean and at the tails than in a normal curve.

Poisson Distribution: A discrete probability distribution which is regarded as an approximation of the binomial distribution when the number of events N is large (≥ 50) and the probability of success p is small (≤ 0.1), in other words, when $Np \leq 5$.

Pooled Variance: An estimate of the population variance calculated by pooling the individual estimates from two samples of that population, when there is no one-to-one correlation between the samples.

Population: All possible values of a variable; the entire group that you want to examine.

Precision: The closeness of repeated measurements of the same quantity.

Probability: A measurement which denotes whether an event occurred simply by pure chance.

R

Range: The difference between the largest and the smallest items in a sample.

Ranked Variables: Those variables that cannot be measured, but at least can be ordered or ranked by their magnitude.

Real Number: That part of a complex number which is not imaginary.

Reliability: The dependability of a product in use.

Right Tail Area: The inverse cumulative distribution function.

S

Sample: Part of a population; that portion of the population that you measure.

Sample Mean: The mean resulting from the process of dividing the sum of the observations by the number of observations; usually denoted by $\overline{X}$.

Sample of Observations: A collection of individual measurements selected by a specific procedure.

Sample Size: The number of items in a sample; usually denoted by the letter n.

Sample Standard Error: The estimate of the standard error of the mean for a sample; designated by $S_{\bar{x}}$.

Simple Arithmetic Mean: The arithmetic mean for ungrouped data.

Skewness: Another word or name for asymmetry; one tail of the curve is drawn out more than the other tail.

Standard Deviation: Measures the variation, or spread, of the individual measurements; a measurement which indicates how far away from the middle your statistics are.

Standard Error: Measures the variation, or spread, of the means of the measurements.

Standard Variate: An observed value which has been normalized (i.e., converted) to fit a standard curve with a mean of 0 and a standard deviation of 1; for any observed point, it is the observed value minus the mean, divided by the standard deviation.

Statistic: The number that results from manipulating raw data according to a specified procedure.

Statistically Equivalent: Values that are alike and that represent the same parameters.

Statistics: The scientific study of numerical data based on natural phenomena.

Sum: The amount obtained by adding numbers or quantities; total; usually denoted by a capital Greek sigma, Σ.

Symmetry: A balanced or normal distribution where the mean and median are identical and the frequency to the right of the mean mirrors that to the left of the mean.

T

t-Statistic: A statistical test value defined as $t = X(X^2/df)1/2$ where X^2 is a random variable following an independent chi-square distribution.

Treatment: A procedure whose effect is to be measured and compared with the effect of other procedures.

Trial: An act or process of testing through use and experience.

U

Ungrouped Data: Raw unorganized data; a listing of all the actual data values.

V

Variable: The actual property measured by the individual observations; also called character.

Variance: The sum of the squares of the deviates divided by one less that the total number of deviates; a measurement which indicates how far away from the middle your statistics are.

Variate: A single measurement or observation of a given variable.

W, X, Y, Z

Weighted Arithmetic Mean: The arithmetic mean for grouped data.

Appendix G. Recommended Readings

Abramowitz, M., and I.A. Stegun, eds. 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* 10th printing with corrections. U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series No. 55. U.S. Government Printing Office, Washington, D.C. 1,046 pp.

Blalock, H.M., Jr. *Social Statistics.* 2nd ed. McGraw-Hill Book Company, New York, New York. 583 pp.

Callamaras, P. 1983. "Statistical Software: A Buyer's Guide to Seven Major Types of Statistics Packages." *Popular Computing* 2(12):206-220.

Cochran, W.G., and G.M. Cox. 1957. *Experimental Designs.* 2nd ed. John Wiley and Sons, Inc., New York, New York. 617 pp.

Considine, D.M. 1976. *Van Nostrands's Scientific Encyclopedia.* 5th ed. Van Nostrand Reinhold Company, New York, New York. 2,370 pp.

Croxton, F.E., D.J. Cowden, and S. Klein. 1967. *Applied General Statistics.* 3rd ed. Prentice-Hall, New York, New York. 754 pp.

Dixon, W.J., and F.J. Massey, Jr. 1969. *Introduction to Statistical Analysis.* 3rd ed. McGraw-Hill Book Company, New York, New York. 638 pp.

Draper, N.R., and H. Smith. 1981. *Applied Regression Analysis.* 2nd ed. John Wiley and Sons, Inc., New York, New York. 709 pp.

Fisher, R.A., and F. Yates. 1963. *Statistical Tables for Biological, Agricultural, and Medical Research.* 6th ed. Hafner Publishing Co., Inc., New York, New York. 146 pp.

Freund, J.E. 1962. *Mathematical Statistics.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 390 pp.

Freund, J.E., P.E. Livermoore, and I. Miller. 1960. *Manual of Experimental Statistics.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 132 pp.

Freund, J.E., and I. Miller. 1965. *Probability and Statistics for Engineers.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey. 432 pp.

Freund, J.E., and F.J. Williams. 1972. *Elementary Business Statistics: A Modern Approach.* 2nd ed. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Garson, G.D. 1976. *Political Science Methods.* Holbrook Press, Inc., Boston, Massachusetts. 540 pp

Hewlett-Packard Company. 1981. *HP 82161A Digital Cassette Drive Owner's Manual.* Hewlett-Packard Company, Corvallis, Oregon. 22 pp.

Hewlett-Packard Company. 1981. *HP-83/85 General Statistics Pac*. Hewlett-Packard Company, Corvallis, Oregon. 77 pp.

Hewlett-Packard Company. 1982. *HP-41C Stat Pac*. Hewlett-Packard Company, Corvallis, Oregon. 76 pp.

Hewlett-Packard Company. 1983. *HP-71 Owner's Manual*. Hewlett-Packard Company, Corvallis, Oregon. 306 pp.

Hewlett-Packard Company. 1983. *HP-71 Reference Manual*. Hewlett-Packard Company, Corvallis, Oregon. 404 pp.

Hewlett-Packard Company. 1983. *HP 82401A HP-IL Interface Owner's Manual*. Hewlett-Packard Company, Corvallis, Oregon. 254 pp.

Hewlett-Packard Company. 1983. *HP-IL/RS-232-C Interface Owner's Manual*. Hewlett-Packard Company, Corvallis, Oregon. 102 pp.

Hewlett-Packard Company. 1984. *HP 2225B Ink Jet Printer Owner's Manual*. Hewlett-Packard Company, Corvallis, Oregon. 69 pp.

Hewlett-Packard Company. 1984. *HP 82484A Curve Fitting Pac Owner's Manual for the HP-71*. Hewlett-Packard Company, Corvallis, Oregon. 125 pp.

Hu, Teh-nei. 1973. *Econometrics: An Introductory Analysis*. University Park Press, Baltimore, Maryland.

Klugh, H.E. 1974. *Statistics: The Essentials for Research*. 2nd ed. John Wiley and Sons, Inc., New York, New York.

Larson, H.J. 1975. *Statistics: An Introduction*. John Wiley and Sons, Inc., New York, New York.

Mendenhall, W. 1966. *Introduction to Statistics*. Wadsworth Publishing Company, Inc., Bellmont, California.

Mendenhall, W., L. Ott, and R.F. Larson. 1974. *Statistics: A Tool for the Social Sciences*. Duxbury Press, North Scituate, Massachusetts. 505 pp.

Natrella, M.G. 1966. *Experimental Statistics*. U.S. Department of Commerce, National Bureau of Standards. Handbook 91. U.S. Government Printing Office, Washington, D.C.

Pearson, E.S., and H.O. Hartley. 1966. *Biometrika Tables for Statisticians*. Vol. 1, 3rd ed. Cambridge University Press, Cambridge, England. 264 pp.

Phillips, L. 1973. *Statistical Thinking: A Structured Approach*. W.H. Freeman and Company, San Francisco, California.

Rohlf, F.J., and R.R. Sokal. 1969. *Statistical Tables*. W.H. Freeman and Company, San Francisco, California. 253 pp.

Scheffe, H. 1959. *The Analysis of Variance*. John Wiley and Sons, Inc., New York, New York. 477 pp.

Seal, H.L. 1968. *Multivariate Statistical Analysis for Biologists*. Methuen and Co. Ltd., London, England. 209 pp.

Selby, S.M., ed. 1971. *Standard Mathematical Tables*. 19th ed. The Chemical Rubber Co., Cleveland, Ohio. 710 pp.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company, Inc., New York, New York. 312 pp.

Snedecor, G.W., and W.G. Cochran. 1980. *Statistical Methods*. 7th ed. Iowa State University Press, Ames, Iowa. 507 pp.

Sokal, R.R., and F.J. Rohlf. 1981. *Biometry: The Principles and Practice of Statistics in Biological Research*. 2nd ed. Freeman and Co., San Francisco, California. 859 pp.

Spiegel, M.R. 1961. *Schaum's Outline of Theory and Problems of Statistics*. McGraw-Hill Book Company, New York, New York. 359 pp.

Spiegel, M.R. 1975. *Schaum's Outline of Theory and Problems of Probability and Statistics*. McGraw-Hill Book Company, New York, New York. 372 pp.

Steel, R.G.D., and J.H. Torrie. 1980. *Principles and Procedures of Statistics, with Special Reference to the Biological Sciences*. McGraw-Hill Book Company, New York, New York. 633 pp.

Young, R.K., and D.J. Veldman. 1965. *Introductory Statistics for the Behavioral Sciences*. Holt, Rinehart, and Winston, New York, New York.

Yule, G.U., and M.G. Kendall. 1950. *An Introduction to the Theory of Statistics*. 14th ed. Hafner Publishing Co., Inc., New York, New York. 701 pp.

Zar, J.H. 1984. *Biostatistical Analysis*. 2nd ed. Prentice- Hall, Inc., Englewood Cliffs, New Jersey. 718 pp.

Index

Page numbers in bold type indicate primary references; page numbers in regular type indicate secondary references.

A

B

C

D

E

F

G

H

I, J

K

L

M

N

T

U, V

W, X, Y, Z

HP-71 STATISTICS

82489-90001

Printed in USA