gu no.40

applications of the hp g series calculator in probability and statistics

gilberto e. urroz, ph. d., p.e.

Applications of the HP 48 G Series Calculator in Probability and Statistics

Copyright © 2000 Gilberto E. Urroz, Ph.D., P.E.
All Rights Reserved.

ISBN 1-58898-040-5

Applications of the HP 48 G Series Calculator in Probability and Statistics

Gilberto E. Urroz, Ph.D., P.E.
greatunpublished.com
No. 40

Applications of the HP 48 G Series Calculator in Probability and Statistics

Table of Contents

PART II- PROGRAMS FOR STATISTICAL ANALYSIS

PREFACE

The HP48G/GX calculator has a variety of statistical analysis tools that can be very helpful in the solution of problems in this course. This document guides you through several applications of those tools. The guide is divided into two parts: Part I - Use of Pre-programmed Features, and Part II - Programming Our Own Features. In Part I we use the HP48G/GX own features to solve problems in statistics. In Part II we will introduce a collection of programs developed using the HP48G/GX RPL programming language to solve statistical problems.
The presentations in Part I guide you step-by-step in the use of pre-programmed features and functions to solve statistical problems. Some of these features include:

- Calculation of descriptive measures for samples
- Determination of frequency distributions
- Barplots, Histograms and Scatterplots
- Factorials, permutations, and combinations
- Random number generator
- Discrete random variables
- Integrals for continuous random variables
- Upper-tail probability tables for Normal, t, Chi-squared, and F distributions
- Use of upper-tail probabilities for simple hypothesis testing
- Simple data fitting

Many of the procedures detailed in Part I can be programmed to facilitate calculations. We have programmed those procedures in a collection of sub-directories that can be transferred to your calculator via the infrared communication port or downloaded via a cable connection to a computer. Description of the programs' operation is provided in Part II. Sub-directories are included for the following procedures:

- Calculation of descriptive measures for samples, including frequency distributions and histograms
- Calculations with and plots of discrete and continuous probability distributions
- Generation of random number lists
- Generation of synthetic data corresponding to discrete and continuous probability distributions
- Inverse upper-tail probability tables for the Normal, t, Chi-squared, and F distributions
- Graphical check for normality of data
- Simulation of distribution of the mean
- Hypothesis testing on means and variances
- Generation of operating curves for hypothesis testing on the mean
- Multiple-linear and polynomial data fitting
- Chi-square applications for goodness of fit and RxC tables

It is not always feasible to keep all these procedures in the calculator due to the limited amount of memory available. Therefore, learn the procedures of Part I and keep this guide handy for future reference, since most of the procedures of Part II are also presented in Part I.

Note: The procedures presented in this guidebook were developed based on the textbook: Johnson, R.A. (1994) "Miller & Freund's Probability and Statistics for Engineers - Fifth *Edition,"* *Prentice* Hall, Upper Saddle River, NJ.

 © 2000 - Gilberto E. Urroz - All rights reserved

PART I – USE OF PRE-PROGRAMMED FEATURES

© 2000 – Gilberto E. Urroz – All rights reserved

1.0 General Features

In this section we will discuss some general features of the HP 48 G series calculator that can be used for statistical applications through pre-programmed features.

1.1 Notation

Keystrokes for the HP48 keys are shown between brackets. For example, [ENTER] indicates the use of the larger ENTER key. Keystrokes that require either [↰], [↱], or [α], are followed by the appropriate instruction, which is not the main key label. For example, the sequence [↰][PURG] indicates pressing the [↰] key, followed by the [EEX] key which has the instruction PURG as a secondary label.
Also, when using the white keys at the top of the keyboard in the calculator, we will indicate the appropriate instruction with italic characters between brackets. For example, to get to a specific subdirectory from the HOME directory, say STATS, we will indicate that the key [*STATS*] must be pressed. To find out which one of the white keys is the appropriate one, look at the labels associated with those keys that appear at the bottom of the display.
For entering numbers or variable names, we could, for example, indicate the keystroke sequence, as in the instruction: ['][α][α][S][T][A][T][S][α], the keystroke sequence to enter in the display the variable name 'STATS'. Or, we could simply indicate: press ['], and type STATS.

1.2 Creating a class directory

Create a directory in your HP48G/G+/GX that will contain problems for this class. Call the directory STATS. Use the following keystrokes:

[↱] [HOME]	Gets you into your HOME directory
[↱] [MEMORY] [NEW] [▼]	To create a new object
[α][α] [S][T][A][T][S] [α] [OK]	Enter directory name
[✓ CHK]	To define object as a directory
[OK]	The display now shows all variables in HOME directory

You can use the [▲] and [▼] keys to move through the display to see all variables in the HOME directory.
Press [ENTER] to get to the normal display.

1.3 Deleting variables

Sometimes you will need to delete variables in the directories and subdirectories to free up memory space. Here are some hints on how to perform such operations:
Press [VAR] to show all variables in the current directory. The variables will be shown at the bottom of the display, corresponding to the white buttons in the top row of the calculator.

To delete a single variable or subdirectory in the current directory: place the name of the variable in display's level 1 by pressing ['] and typing the variable name, or, simply, by pressing the white button corresponding to the variable after pressing [']. Then, press [↰] [PURG]. The variable name corresponding to the variable that you just purged will disappear from the list at the bottom of the display.

 © 2000 - Gilberto E. Urroz - All rights reserved

To delete several variables at once: Press [VAR] to show all variables. If there are more than six variables in any given directory, use the [NXT] or [↰][PREV] keys to view all of them. Create a list of the variables to be purged, by using [↰][{}]; then, press the white buttons corresponding to each of the variables to be purged. When the list is complete, press [ENTER]. To purge the list, press [↰][PURG]. The names of the purged variables will disappear from the bottom of the display.

To delete an entire sub-directory, enter the name of the sub-directory, between quotes, into level 1 of the display, then use the keystroke sequence: [↰][MEMORY][*DIR*][*PGDIR*].

1.4 Class subdirectories

To move to the HOME directory, press [↱][HOME]. The display will show all the directories and variables available in the HOME directory. One of them will be STATS. Press [*STATS*] to get into the class directory. It is a good idea to create a subdirectory for each problem to be solved to avoid mixing up the data matrices used by the STAT or PLOT features in the calculator. For example, create a subdirectory for the example (obtained from page 11 of the Miller & Freund's book). Call the subdirectory EXP11. To create this subdirectory follow these steps, from within the STATS directory:

[↱] [MEMORY] [*NEW*] [▼]	To create a new object
[α][α] [E][X][P][1][1] [α][α] [OK]	Enter directory name (EXP11)
[✓ *CHK*]	To define object as a directory
[OK][ENTER]	To get to the normal display.
[*EXP11*]	To enter subdirectory.

Note: Use [↰][UP] to move to an upper directory.

1.5 Transferring data with the HP48G/GX

Suppose you want to use the infrared port in the HP48G/G+/GX to transfer a variable called MYDAT to another calculator. The instructions to transfer files using the infrared port are given in page 8-2 (*Lesson 35: Transferring Objects Via Infrared*) in the *HP48 Series/Quick Start Guide*. We assume both the receiver and the sender calculators are in the appropriate directory. The steps to follow are:
Line up the infrared ports by lining up the marks (near the HP logo just above the display). The calculators should not be farther apart than 2 inches.
Receiver calculator:
Press [↱][I/O]
Select **Get from HP 48** from the menu and press [OK]
Sender calculator:
Press [↱][I/O]
Select **Send to HP 48** from the menu and press [OK]
Press [*CHOOS*] and select the name of the object to be transferred (MYDAT, in this case), then press [OK]
Press [SEND]

1.6 Undo and Arg

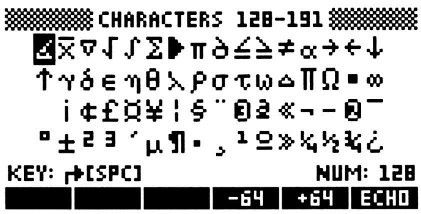
If you make a mistake in your calculations you may be able to recover your stack by pressing [↱][UNDO]. If you want to re-use the arguments from the latest calculation, try [↱][ARG].

 © 2000 - Gilberto E. Urroz - All rights reserved

1.7 How to type Greek letters and other characters

To type characters now shown in the keyboard, the general procedure is to use [↱][CHARS].

The calculator screen will look like this:



Then, use the arrow keys ([◄],[►],[▲], and [▼]) to move to any desired character. If the character is not available in the current screen, press the [*-64*] or [*+64*] buttons to move to the previous or next screen. Once the desired character is highlighted, press [*ECHO*] to copy it to the stack. Press [ENTER] to return to the stack.
The Greek letters more commonly used in Statistics are α, β, μ, ν, σ, ν, π, ρ, and $\bar{x}$. Another character used often in our course is $\bar{x}$ (x-bar). The shortcuts for typing these characters are as follows:

α: [α][↱] [A]	β: [α][↱] [B]	μ: [α][↱] [N]	ν: [α][↱] [G] (*)
σ: [α][↱] [S]	π: [↰][SPC]	ρ: [α][↱] [R]	$\bar{x}$: [α][↱] [X]

(*) The letter *nu* (ν) is not available in the HP48G/GX, thus, the letter gamma (γ) would have to do instead.

1.8 Changing the display format

Typically, the display format used is the standard format (STD), with a variable number of decimals. For example, if you enter the number 2.5 ([2][.][5]) when the standard format is active, the display will show just 2.5. Press now, [↱][LN], and you get .916290731874. The standard format uses up to 12 decimal places for non-integer results.

Use the following keystroke sequence to access the format change environment: [↰][MODES][FMT].
To fix the number of decimals places used, say to three decimal places, use the following keystroke sequence: [3][FIX]. The displayed result will now be 0.916.
To return to the standard display format you may press [STD]. The current display format will be marked by a dot in the corresponding white key at the top of the keyboard.

Note that in the format change environment there are also white keys labeled [SCI][ENG][FM,][ML]. The [SCI] and [ENG] keys refer to what calculator manufacturers call *Scientific* and *Engineering* notations, respectively, which provides results using powers of ten. The [FM,] button, when marked with a dot, changes the decimal point to a coma. Press [FM,] and the displayed number will read 0,916. Press [FM,] again to select the decimal point. Press [VAR] to return to your variable menu.

 © 2000 – Gilberto E. Urroz – All rights reserved

1.9 Changing the angle mode and coordinate system

To get into the angle/coordinate change environment, press [↰][MODES][*ANGL*].
In this environment, you will see the following white key labels:

[*DEG*][*RAD*][*GRAD*][*RECT*][*CYLIN*][*SPHER*]

referring to the angle mode as in (sexagesimal) DEGrees, RADians, or (decimal) GRADes; and to the RECTangular, CYLINdrical (polar), and SPHERical coordinate systems. For most applications we use the rectangular coordinate system, and the angles in degrees or radians.

As an example, press the keys labeled [RECT] and [DEG] to set rectangular coordinates and angle in degrees. Then, enter a vector of three components, say [2 3 5].

Use the following keystroke sequence:

[↰][[ ]][2][SPC][3][SPC][5][SPC][ENTER]

The calculator assumes that the three components of the vector correspond to the x,y,z, components of Cartesian or rectangular coordinates. If you now press [*CYLIN*], the three-dimensional vector gets transformed to

`[3.606 ∠56.310 5.00],`

where the symbol ∠ indicates an angle. (The character ∠ can be typed in by using the sequence [↰][SPC]). Notice that the characters R∠Z appear in the upper left corner of the display, indicating that the components of the vector are now the polar cylindrical coordinates

$r = 3.606$, $\theta = 56.310^{\circ}$, and, $z = 5.00$.

Pressing the key [*SPHER*] will produce the following vector:

`[6.165 ∠56.310 ∠35.796],`

which correspond to the spherical coordinates,

$\rho = 6.165$, $\theta = 56.310^{\circ}$, $\phi = 35.796^{\circ}$.

The upper left corner of the display will show the characters: R∠∠.

We could change our angle units to radians by pressing [*RAD*]. Notice that the vector in the display now reads:

`[6.164 ∠0.983 ∠0.625],`

and the upper left corner of the display shows the characters RAD in front of the cylindrical coordinate descriptor R∠∠. If you press the key [*GRAD*], the vector will be displayed as

`[6.164 ∠62.567 ∠39.773],`

while the upper left corner of the display shows the characters GRAD in front of the cylindrical coordinates descriptor R∠∠. The decimal GRADes are not very commonly used in practice.

 © 2000 - Gilberto E. Urroz - All rights reserved

A quick way to change angle units consists of pressing the keystroke combination: [↰][RAD]. That will toggle the RAD indicator at the top left corner of the display. When the RAD indicator disappears, the angle units are degrees. Try the following exercise:

[MTH][VECTR][NXT][RECT] [↰][[ ]][2][SPC][3][SPC][ENTER] [CYLIN]

If the RAD indicator is shown, you will get `[3.606 ∠0.983].`

Press [↰][RAD], to get `[3.606 ∠56.310].`

A quick way to change coordinate systems is to press [↱][POLAR], and watch for the coordinate announcer to change from blank (rectangular) to R∠Z (polar) or to R∠∠ (spherical). Enter

`[3.606 ∠56.310].`

into the display, and press [↱][POLAR] a couple of times to see the vector change coordinates.

As a reminder, recall that the basic transformation between angle units is as follows:

$$\theta^r/\theta^o = \pi/180,$$
$$\theta^r/\theta^d = \pi/200,$$
$$\theta^o/\theta^d = 90/100 = 9/10.$$

Quick conversions from degrees to radians, and viceversa, can be accomplished by using the sequence

[MTH][*REAL*][NXT][NXT][*D→R*], and [*R→D*], respectively.

For example, try:

[3][7] [MTH][*REAL*][NXT][NXT][*D→R*]

to convert 37° to 0.646 rad. Also, try

[3][.][1][4][MTH][*REAL*][NXT][NXT][*R→D*]

to convert 3.14 rad to 179.909°.

Coordinate transformations are given by the following:
Rectangular & cylindrical coordinates:

$$r = (x^2 + y^2)^{1/2}, \qquad \tan\theta = y/x,$$
$$x = r\cos\theta, \qquad y = r\sin\theta$$

Rectangular & spherical coordinates:

$$\rho = (x^2 + y^2 + z^2)^{1/2}, \qquad \tan\theta = y/x, \qquad \tan\phi = z/(x^2 + y^2)^{1/2},$$
$$x = \rho\cos\phi\cos\theta, \qquad y = \rho\cos\phi\sin\theta, \qquad z = \rho\sin\phi.$$

Make sure that you reset your coordinate system to rectangular.

 © 2000 - Gilberto E. Urroz - All rights reserved

1.10 HP48G/G+/GX constants

The following are the mathematical constants used by your calculator:

e: the base of natural logarithms.
i: the imaginary unit, $i^2 = -1$.
π: the ratio of the length of the circle to its diameter.
MINR: the minimum real number available to the calculator.
MAXR: the maximum real number available to the calculator.

To have access to these constants, use the combination:

[MTH][NXT][*CONS*].

The calculator display will show buttons corresponding to the following variables:

[*E*][*2.718*][*I*][*(0.00*][π][*3.141*].

Press [NXT] to get

[*MINR*][*1.000*][*MAXR*][*1.000*].

Press [*2.718*] to get the value of *e*, 2.718 in the display. If you press [*E*], you will get the variable name in the display, namely, `'e'`. To get the numerical value, press [↰][→NUM].

1.11 Physical constants available in the HP48G/G+/GX calculator

The keystroke sequence:

[↰][EQ LIB][*COLIB*][*CONLI*]

gives you access to a list of physical constants available for use in your HP48G/G+/GX calculator. Typically the buttons [*SI*] and [*UNITS*] will be selected, indicating that the constants are provided with units of the SI system. Use [▲] [▼] to see the entire list of constants.

The default list shows the most common constant symbol and its definition. To see the actual value of the constant, select the [*VALUE*] button. Pressing that button again will return you to the original list of constants. To get a constant value copied to the stack, press [*→STK*].

Although we probably will not be using any of these values in our statistical analysis, knowing where to find them can be useful in other applications.

1.12 Working with units in the HP48G/G+/GX calculator

Working with units entails searching through a number of menus in the calculator where the units are defined. Due to the lengthy process involved in assigning units when using the HP4G/G+/GX calculator, my personal preference is to work without units. When you do that, however, make sure that you are using a consistent system of units (i.e., SI, English).

To illustrate the use of units, let's calculate a force, F, given the mass, m = 3.5 kg, and the acceleration, a = 2.3 cm/s^2. From F = m a, we have, F = (3.5 kg)(2.3 cm/s^2). To perform this calculation use the following:

 © 2000 - Gilberto E. Urroz - All rights reserved

[3][.][5] Enter the numeric value of m
[↱][UNITS][MASS][KG] Select kg units in the "mass" menu.
[2][.][3] [↱][UNITS][[SPEED][CM/S] Select units of velocity since units of acceleration are not available.
[1][↱][UNITS][TIME][S] Select units of time (s), to calculate a = v/t.
[÷] Gives units of acceleration (cm/s^2)
[×] To calculate F = m·a (`8.050_Kg*cm/s^2`)
[ENTER][ENTER] Make two more copies of the result for future use.

Next, we demonstrate some operations that can be performed using the UNITS menu:

[↰][UNITS][UBASE] Converts to basic units of the SI system.
[⇦] Drop contents of level 1.
[↰][UNITS][UVAL] Eliminates units, retains numeric value.
[⇦] Drop contents of level 1.
[1][↱][UNITS][NXT][FORCE][LBF] Enters 1 lbf in level 1. We'll convert the value in level 2, which is in Kg cm/s2, into lbf (pound force), a unit in the English System.

[↰][UNITS][CONV] Converts value in level 2 to units in level 1.

The latest operation, i.e., *unit conversions*, is a very useful feature in the UNITS menu. As an example, if you want to know how many pints are there in a gallon, and how many cups in a pint, use the following keystroke sequences

[↱][UNITS][VOL][1][NXT][GAL][1][PT][↰][UNITS][CONV] (Result: 1 gallon = 8 pints); and,

[↱][UNITS][VOL] [1][NXT][PT] [1][NXT][CU][↰][UNITS][CONV] (Result: 1 pint = 2 cups).

1.13. Functions, lists and programs.

Defining a function consists in creating a variable that contains the definition of the function. For example, suppose we want to define the function

$$f(x) = x^3$$

There are two ways to enter the expression that defines the function:
Using the equation editor:

[↰][EQUATION] [α][↰][F] [↰][( )] [α][↰][X] [▶] [↰][=] [α][↰][X] [y^x][3] [ENTER]

Entering the expression directly into level 1 of the display:

['] [α][↰][F] [↰][( )] [α][↰][X] [▶] [↰][=] [α][↰][X] [y^x][3] [ENTER]

To define the function, use the keystroke sequence: [↰][DEF].

There will be a new white key named F, i.e., [*F*]. (Press [VAR] if needed.) To see how the function is stored into the variable f, press [↱][*F*]. Level 1 of the display now shows the following:

```
<< → x 'x^3' >>
```

 © 2000 - Gilberto E. Urroz - All rights reserved

This expression represents an RPL program that can be interpreted as follows*: take the value in level 1 and assign it to x (→ x), then, calculate the expression between quotes ('x^3').* The result is then placed in level 1 of the display.
Notice that the function definition procedure used above (namely, using the keystroke sequence [↰][DEF]), translates the function definition (in general, 'f(x) = expression containing x') into a program of the form

<< → x 'expression containing x'>>.

We will introduce other RPL programs in this guide. Please notice that RPL programs are always enclosed within the double quotes << >>.

Let's now evaluate our function f(x) for a given value of x, say, x = 5. First, you need to enter the value of x in level 1 of the display by pressing [5][ENTER]. Next, press the white button for F, i.e., [*F*]. The result is 125, or f(5) = 125.
We could evaluate the function using a list as the argument. (A list, in the HP48G/GX calculator, is a collection of objects contained between curly brackets {}. We will limit our discussion to lists of numbers.) For example, let's find the value of f({1 2 3}). First, enter the list in level 1 of the display:

[↰][{}] [1] [SPC] [2] [SPC] [3] [ENTER]

Then, press the button [F]. The result is {1 8 27}, i.e., f[{1 2 3}] = {1 8 27}. Notice that the program defining f(x) applied the function definition (namely, x^3) to each element of the argument list and produced a new list with the same number of elements.
Let's now try defining a different function,

$$g(x) = \frac{\sinh(x)}{1+x^2}.$$

Use the following keystroke sequence to create the variable g:

['] [α][↰][G] [↰][( )] [α][↰][X] [▶] [↰][=] [MTH][*HYP*][*SINH*] [α][↰][X]
[▶] [÷][↰][( )] [1] [+][α][↰][X] [y^x][2] [ENTER] [↰][DEF].

Press [VAR], if needed. Now, evaluate g(3.5), by entering the value of the argument in level 1 ([3][.][5][ENTER]) and then pressing [*G*]. The result is 1.2485..., i.e., g(3.5) = 1.2485... Try also obtaining g[{1 2 3}], by entering the list in level 1 of the display ([↰][{}] [1] [SPC] [2] [SPC] [3] [ENTER]), and pressing [*G*]. In this case we get a division error message. The calculator display looks as follows:

```
/ Error:
Invalid Dimension
4:
3:
2: { 1.17520119364 3. …
1:               {1 1 4 9}
```

Notice that the list in level 1 has four elements. Now, drop that list by pressing [⇦]. The list now in level 1 has only three elements. Apparently, the program defined by [*G*] tried to perform a division between two lists that have different number of elements, and that is what

 © 2000 - Gilberto E. Urroz - All rights reserved

causes the division error reported. (Most arithmetic operations between lists require lists with the same number of elements).

How did we end up with lists of different sizes if we started with a single list of size 3? The explanation lies in the way that the operator [+] applies to lists. The operator [+], when applied to lists, does not act as a simple addition, instead it is known as the *concatenation* operator and its function is to attach (or concatenate) the two lists. For our particular situation, the list x = {1 2 3} is squared in the denominator ($1+x^2$) of the expression that defines g(x) producing the list {1 4 9}. Then, this list is concatenated to the number 1 by way of the operator [+] in the denominator ($1+x^2$). This result in the list {1 1 4 9} = 1 + {1 4 9} shown in level 1 of the display (see above). The list in level 2 of the display represents SINH(x). In order to calculate the function g(x), the calculator tries to divide SINH(x), which is a list of three elements, by ($1+x^2$), which is a list of four elements, which results in the division error reported.

The moral of this exercise is the following: when trying to evaluate a function whose definition includes the plus (+) sign, if the function argument is a list the result most likely will produce an algebraic error. For example, when we evaluated f[{1 2 3}], we obtained a reasonable result (namely, {1 8 27}), because the definition of the function ($f(x) = x^3$) did not include any plus signs.

To get around using the [+] sign, the HP48G/GX calculator offers the function ADD that applies particularly to lists, but that can be used to represent addition of other numbers. Instead of using the definition currently stored for the variable [G], namely, `<< → x 'SINH(x)/(1+x^2) >>`, we'll replace it with the following program: `<< 'x' STO x SINH 1 x SQ ADD / 'x' PURGE >>`. To key in the program follow these instructions:

Keystroke sequence:	Produces:	Interpreted as:
[↰] [<< >>]	<<	Start an RPL program
['] [α] [↰] [X] [▶] [STO]	'x' STO	Store the contents of level 1 into variable x
[α] [↰] [X]	x	Place x in level 1
[MTH][*HYP*][*SINH*]	SINH	Calc. *sinh* of level 1
[1] [SPC] [α] [↰] [X] [↰] [x^2] [MTH] [*LIST*] [*ADD*]	1 x SQ ADD	Calculate ($1+x^2$)
[÷]	/	Divide
['] [α] [↰] [X] [▶] [↰] [PURGE]	'x' PURGE	Purge variable x
[ENTER]		Program shown in lev.1

To save the program use:

['] [α] [↰] [G] [STO]

Press [VAR], if needed. Now, evaluate g(3.5) by entering the value of the argument in level 1 ([3][.][5][ENTER]) and then pressing [G]. The result is 1.2485..., i.e., g(3.5) = 1.2485...Same as before. Try also obtaining g[{1 2 3}], by entering the list in level 1 of the display ([↰][{}] [1] [SPC] [2] [SPC] [3] [ENTER]), and pressing [G]. The result now is {0.5876... 0.7253... 1..001...}.

Note: This exercise shows how to create a program in the HP48G calculator using the calculator's RPL language. This is the language used in developing the programs presented in Part II of this guide. Programming the HP48G/GX is not an easy task, but can be learned with some effort. For starters, refer to section 29 and Appendices G and H in the ***HP 48G Series*** *User's Guide* (this manual is included with your HP48G/GX calculator), and to the ***HP 48G Series*** *Advanced User's Reference Manual* (this book is sold separately).

 © 2000 - Gilberto E. Urroz - All rights reserved

1.14. More operations with lists

In section 1.13 we presented a couple of examples of algebraic operations and functions applied to lists. Here we introduce a few more operations on lists. As you will see in the rest of this guide, lists are very useful for statistical analysis.
As indicated above, for the HP48G/GX calculator lists are collections of objects such as numbers, letters, character strings, variable names, and/or operators contained between curly brackets. Some examples of lists are: { t 1 }, {"BETA" h2 4}, {1 1.5 2.0}, {a a a a}, { {1 2 3} {3 2 1} {1 2 3}}, etc. For applications in statistics we will limit ourselves to numerical lists.

Simple operations with lists are accessible by using the keystroke sequence: [MTH][*LIST*].
The operations thus available are:

[*ΔLIST*][*ΣLIST*][*ΠLIST*][*SORT*][*REVLI*][*ADD*]

We already showed how to use [ADD]. The other operations in this set are defined as follows:

[*ΔLIST*]: produces a list of increments between consecutive elements in the original list. E.g., {1 2.1 3.5 4.2 3.8} [ENTER][*ΔLIST*] produces {1.1 1.4 0.7 -0.4}.

[*ΣLIST*]: calculates the sum of the elements in the list. E.g., {1 2.1 3.5 4.2 3.8} [ENTER] [*ΣLIST*] produces 14.6.

[*ΠLIST*]: calculates the product of the elements in the list. E.g., {1 2.1 3.5 4.2 3.8} [ENTER][*ΠLIST*]
produces 117.306.

[*SORT*]: sorts elements in increasing order. E.g., {1 2.1 3.5 4.2 3.8} [ENTER][*SORT*] produces {1 2.1 3.5 3.8 4.2}.

[*REVLI*]: reverses order of elements in list. E.g., {1 2.1 3.5 4.2 3.8} [ENTER][*REVLI*] produces {3.8 4.2 3.5 2.1 1}

Additional operations with lists are available by pressing [PRG][*LIST*]. In particular, we will describe the use of the operations [*OBJ→*] and [*→LIST*].

[*OBJ→*] is used to decompose a list into its elements. By using this operation the elements of the list are placed in different levels of the display, with level 1 displaying the number of elements, level 2 displaying the last element on the list, level 3 displaying the second-to-last element on the list, and so on. For example, {0.5 3.2 6.5} [PRG][*LIST*][*OBJ→*] produces the following display:

```
4:                    0.5
3:                    3.2
2:                    6.5
1:                      3   (← number of elements)
```

[*→LIST*] is used to create a list using the values available in the different display levels. The number of elements in the list must be placed in level 1. The value in level 2 will become the last element in the list, the value in level 3 will become the second-to-last element in the list, and so on. For example, the following display:

 © 2000 - Gilberto E. Urroz - All rights reserved

```
4:           0.5
3:           4.2
2:           8.5
1:             2  (← number of elements)
```

will produce this list: {4.2 8.5} when the keystroke sequence [PRG][*LIST*][→*LIST*] is used.

Notice that [*OBJ→*] and [→*LIST*] are inverse operations.

For more information on list operations refer to section 17 in the ***HP 48G Series****User's Manual* that is included with your calculator.

1.14.1. Programs for list operations

Following we present three programs that we developed for creating or manipulating lists. These programs are useful in handling numerical lists for some statistical applications.

The program listings are as follows:

LISC:

```
<<  SWAP 'n' STO 'x' STO 1 n FOR j x NEXT n →LIST { n x } PURGE >>
```

CRLST:

```
<<  'df' STO 'en' STO 'st' STO st en FOR  j j df STEP en st - df / FLOOR 1 +
{st en df } PURGE →LIST >>
```

CLIST:

```
<< REVLIST DUP DUP SIZE 'n' STO ΣLIST SWAP TAIL DUP SIZE 1 - 1 SWAP FOR j DUP
ΣLIST SWAP TAIL NEXT 1 GET n →LIST REVLIST 'n' PURGE >>
```

The following are the keystroke sequences to get some of the commands in the programs:

SWAP	[↰][SWAP]	**STO**	[STO]
FOR	[PRG][*BRCH*][*FOR*][*FOR*]	**NEXT**	[PRG][*BRCH*][*FOR*][*FOR*]
STEP	[PRG][*BRCH*][*FOR*][*FOR*]	**→LIST**	[PRG][*LIST*][→*LIST*]
PURGE	[↰][PURGE]	**REVLIST**	[MTH][*LIST*][*REVLI*]
DUP	[↰][EQUATION]	**SIZE**	[PRG][*LIST*][*ELEM*][*SIZE*]
ΣLIST	[MTH][*LIST*][*ΣLIST*]	**TAIL**	[PRG][*LIST*][*ELEM*][NXT][*TAIL*]
GET	[PRG][*LIST*][*ELEM*][*GET*]	**FLOOR**	[MTH][*REAL*][NXT][NXT][*FLOOR*]

The operation of these programs is as follows:

LISC: creates a list of n elements all equal to a constant c.

Operation: enter n, enter c, press [LISC]

Example: [5][ENTER][6][.][5][ENTER][LISC] creates the list: {6.5 6.5 6.5 6.5 6.5}

 © 2000 - Gilberto E. Urroz - All rights reserved

CRLST: creates a list of numbers from n_1 to n_2 with increment Δn, i.e., $\{n_1, n_1+\Delta n, n1+2\cdot\Delta n, \ldots n_1+N\cdot\Delta n\}$, where

$$N = floor(\frac{n_2 - n_1}{\Delta n}) + 1$$

Operation: enter n_1, enter n_2, enter Δn, press [CRLST]

Example: [.][5][ENTER][3][.][5][ENTER][.][5][ENTER][CRLST] creates the list: {0.5 1 1.5 2 2.5 3 3.5}

CLIST: creates a list with cumulative sums of the elements, i.e., if the original list is $\{x_1\ x_2\ x_3 \ldots x_N\}$, then CLIST creates the list:

$$\{x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots, \sum_{i=1}^{N} x_i\}$$

This type of procedure is useful when determining cumulative frequency or probability distributions.

Operation: place the original list in level 1, press [CLIST]

Example: {1 2 3 4 5}[ENTER][CLIST] produces {1 3 6 10 15}.

 © 2000 – Gilberto E. Urroz – All rights reserved

1.15. Graphics

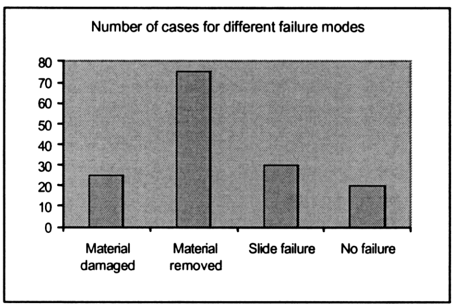
In this section we discuss a few of the many graphic features of the HP48G/GX calculator that can be useful for statistical applications. These include barplots, scatterplots, and function plots.

1.15.1. Barplots

Barplots are used for creating Pareto diagrams as well as histograms. Pareto diagrams are basically barplots comparing values of a certain variable corresponding to different (qualitative) categories. For example, suppose that an analysis of experiments on erosion control in hillslopes shows that, out of 150 tests, the following cases are observed:

Failure mode	No. of cases
Material damaged	25
Material removed	75
Slide failure	30
No failure	20

The values to be compared are the number of cases and the categories are the failure modes. A Pareto diagram for the data in the table above is shown below:



In the HP48G/GX, a rough Pareto diagram can be obtained by using the Barplot graphics feature. The numerical data must be stored in the statistics matrix ΣDAT. The categories, not being numerical values, can not be stored in the ΣDAT matrix, therefore, you only enter the numerical values and identify the categories in paper. The data from the table above can be entered in the statistics matrix as follows:

[↱][MATRIX] [2][5][ENTER][▼] [7][5][ENTER] [3][0][ENTER] [2][0][ENTER] [ENTER]
[↰][STAT] [*DATA*] [↰][*ΣDAT*]

Then, use the following keystroke to see the barplot:

[↰][STAT] [*PLOT*][*BARPL*]

Press [(X,Y)] to see coordinates of the cursor. Move the cursor using the arrow keys ([◀] [▶] [▲] [▼]) to determine the height of the bars. The values of x thus determined have no real meaning. Each bar represents the qualitative categories shown in the table above. Press [ON] to return to the normal display.

 © 2000 - Gilberto E. Urroz - All rights reserved

Barplots are also used to plot histograms. Frequency distributions and histograms are discussed in section 3.0 in Part I of this guide.

1.15.2. Scatterplots

Scatterplots are used to present numerical data in the form of (x,y) data points. The data must be stored in the statistical matrix ΣDAT. For example, if we want to plot the following data

x	y
0.20	3.16
0.50	2.73
1.00	2.12
1.50	1.65
2.00	1.29
4.00	0.47
5.00	0.29
10.00	0.01

We can enter them into the matrix ΣDAT using the following keystrokes:

[↱][MATRIX] 0.2 [SPC] 3.16 [ENTER][▼] 0.5 [SPC] 2.73[ENTER] 1 [SPC] 2.12 [ENTER]
1.5 [SPC] 1.65 [ENTER] 2 [SPC] 1.29 [ENTER] 4 [SPC] 0.47 [ENTER] 5 [SPC] 0.29 [ENTER]
10 [SPC] 0.01 [ENTER][ENTER]
[↰][STAT] [DATA] [↰][ΣDAT]

The ΣDAT matrix has the x values in the first column and the y values in the second column.

To plot the scatterplot use these keystrokes:

[↰][STAT] [PLOT][SCATR]

Try the following keystrokes to see the plot axes:

[EDIT][NXT][LABEL][MENU].

The plot shows a label X, and values of 0.2 to 10 for the x-axis. No information is provided for the y-axis, because it is located outside of the display window. To see the coordinates of the points, try the following keystrokes:

[NXT][NXT][PICT][(X,Y)]

Move the cursor towards any point to see the coordinates of that point. To return to normal display use:

[NXT][CANCL].

 © 2000 - Gilberto E. Urroz - All rights reserved

1.15.3. Function plots

In this section we present an example of an HP48G/GX plot of a function of the form y = f(x). In order to proceed with the plot, first, purge the variable x, if it is defined in the current directory (because x will be the independent variable in the calculator's PLOT feature, you don't want to have it pre-defined). Press [VAR] and check if one of the white keys is labeled [*X*]. If such variable exist, purge it before proceeding.

Let's plot the Standardized Normal curve, given by,

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})$$

Define the function using the following keystrokes:

['][↰][e^x][-][α][↰][Z][y^x][2][÷][2][▶][÷][√*x*][↰][( )][2][×] [↰][π][ENTER]

Store it into a variable called yN:

['][α][α][↰][Y][N][α][STO]

Enter the PLOT environment, [↱][PLOT], and select `Function` as the `TYPE`.

Highlight the `EQ` field, and press [*CHOOS*]. Use the [▲] [▼] keys to select yN, then press [OK].

Type z (lowercase) as the independent variable (`INDEP`) with range -4 to 4.

Reset the independent variable in the OPTS screen: [*OPTS*][▶][4][±][OK][4][OK][OK].

Place a check mark (✓) in the `AUTOSCALE` option.

Plot the graph: [*ERASE*][*DRAW*].

To see labels: [EDIT][NXT][*LABEL*][*MENU*]

To recover the menu: [NXT][NXT][PICT]

To trace the curve with the cursor: [TRACE][(X,Y)]

Use [▶] and [◀] to move along trajectory. Check that for z = 1.05 , y = 0.231. Also, check that for z = -1.48 , y = 0.134.

To recover the menu, and return to the PLOT environment, press [NXT][CANCL].

1.15.3.1. Some useful PLOT operations for function plots

In order to discuss these PLOT options, we'll modify the function to force it to have some real roots (Since the normal curve is totally contained above the x axis, it has no real roots.) First, highlight the field in front of `EQ:` in the PLOT environment. Then, press [*EDIT*]. Use the following keystrokes:

[▼][-][.][1][ENTER].

 © 2000 - Gilberto E. Urroz - All rights reserved

The function to be plotted is now,

$$f(z)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})-0.1$$

Before plotting, place a check mark (✓) in the AUTOSCALE option. To plot the graph, press

[*ERASE*][*DRAW*].

Press [*FCN*] to access the calculus menu. With this menu you can obtain additional information about the plot such as intersects with the x-axis, roots, slopes of the tangent line, area under the curve, etc.

For example, to find the root on the left side of the curve, move the cursor near that point, and press [*ROOT*]. You will get the result: ROOT: -1.6635.... Press [NXT] to recover the menu.

If you move the cursor towards the right-hand side of the curve and press [*ROOT*], the result now is ROOT: 1.6635... The calculator indicated, before showing the root, that the root was found through *SIGN REVERSAL*. Press [NXT] to recover the menu.

Pressing [*ISECT*] will give you the intersection of the curve with the x-axis, which is essentially the root. Press [*ISECT*]. You will get the same message as before, namely *SIGN REVERSAL*, before getting the result I-SECT: (1.6635..., 0.0000). The [*ISECT*] function is intended to determine the intersection of any two curves closest to the location of the cursor. In this case, where only one curve, namely, f(x) as defined above, is involved, the intersection sought is that of f(x) with the x-axis. Press [NXT] to recover the menu.

Place the cursor on the curve at any point and press [*SLOPE*]. For example, if you place the cursor at any point and press [*SLOPE*], you will get the value of the slope at that point. For example, at the negative root, SLOPE: 0.16670.... Press [NXT] to recover the menu.

To determine the highest point in the curve, place the cursor near the vertex and press [*EXTR*]. The result is EXTRM: (0,0.2989...). Press [NXT] to recover the menu.

Other buttons available in the first menu are [*AREA*] to calculate the area under the curve, and [*SHADE*] to shade an area under the curve. Press [NXT] to see more options. The second menu includes one button called [*VIEW*] that flashes for a few seconds the equation plotted. Press [*VIEW*]. Alternatively, you can press the button [*NXEQ*] to see the expression for the function f(x). Press [NXT] to recover the menu.

The button [*F(X)*] gives the value of f(x) corresponding to the cursor position. Place the cursor anywhere in the curve and press [*F(X)*]. The value will be shown in the lower left corner of the display. Press [NXT] to recover the menu.

Place the cursor in any given point of the trajectory and press [*TANL*] to obtain the equation of the tangent line to the curve at that point. The equation will be displayed on the lower left corner of the display. Press [NXT] to recover the menu.

If you press [*F'*] the calculator will plot the derivative function, f'(x) = df/dx, as well as the original function, f(x). Notice that the two curves intercept at two points. Move the cursor near the left intercept point and press [*FCN*][*ISECT*], to get I-SECT: (-0.6834...,0.21585). Press [NXT] to recover the menu.

To leave the FCN environment, press [*PICT*].

 © 2000 - Gilberto E. Urroz - All rights reserved

Press [*CANCL*] to return to the PLOT environment.

Please notice that the field in front of EQ: in the PLOT environment now contains a list. If you press [EDIT] you will notice that there are two equations in the list:

{ '-(2*z/2*EXP(-(z^2/2))/√(2*π))' 'EXP(-(z^2/2))/√(2*π)-.1'}

When a function is chosen in the EQ: field the calculator creates a variable called EQ (for Equation) containing that function. Originally, EQ contained only f(x). After we pressed the button [F'] in the [FCN] environment, the calculator automatically added f'(x) to the list of equations in EQ.

Press [ENTER] to return to the PLOT environment.

Press [ENTER] to leave the PLOT environment.

1.15.4. Saving a graph for future use

If you want to save your graph to a variable, get into the PICTURE environment by pressing

[↰][PICTURE].

Then, press

[EDIT][NXT][NXT][PICT→].

This captures the current picture into a graphics object. To return to the stack, press

[PICT][CANCL].

In level 1 of the stack you will see a graphics object described as `Graphic 131 × 64`. To store it into a variable, say FIG, type

['][α][α][F][I][G][α][STO].

Your figures now is stored in variable FIG.
To display your figure again, recall it to level 1 of the stack, by pressing [FIG]. Level 1 now reads `Graphic 131 × 64`. Enter the PICTURE environment, press

[↰][PICTURE].

Clear the current picture, [EDIT][NXT][ERASE].

Move the cursor to the upper left corner of the display, by using the [◄] and [▲] keys.

To display the figure currently in level 1 of the stack press [NXT][REPL].

To return to normal calculator function, press [PICT][CANCL].

 © 2000 - Gilberto E. Urroz - All rights reserved

1.15.5. Pie charts.

Pie charts are commonly used to describe data in a manner similar as that of Pareto charts. The HP48G/GX does not have a pre-programmed pie chart feature, however, an RPL program for pie chart generation is given in pages 2-49 to 2-53 of the ***HP 48G Series****Advanced User's Reference Manual*.

1.15.6. Other statistical plots.

Section 1 in Part II of these guides discusses other type of plots used commonly in statistical analysis including boxplots, dot plots, and ogives.

Note: for quick access to barplots and scatterplots, see section 15 in Part I of this guide

 © 2000 - Gilberto E. Urroz - All rights reserved

2.0 Entering Data: Vectors vs. Lists

To practice entering sample data in the calculator we'll use the data from problem 2.7. First, let's create a subdirectory within directory STATS that we will call P2.7. Follow these steps:

[↱][HOME].	Move to the HOME directory.
[STATS]	Go to the STATS subdirectory
[↱] [MEMORY] [*NEW*] [▼]	To create a new object
[α][α] [P][2][.][7] [α][α] [OK]	Enter directory name (P2.7)
[✓ *CHK*]	To define object as a directory
[OK][ENTER]	To get to the normal display.
[P2.7]	To enter subdirectory.

These data represents measurements of the same magnitude. Think of them as a single vector. Enter the data starting with the first number in the first column in page 19, and sweeping the table by columns until reaching the last number in the last column of the table. There are two ways to enter these data:

As a column vector(The STAT feature in the calculator requires a data matrix. For a single data set, as in this case, the matrix becomes a single-column vector).
Press [↱][MATRIX] to start the matrix editor.

Enter the first element, 32.5, and press [ENTER][▼]. The cursor is now in element (2,1), i.e., row 2, column 1, of the matrix. You can continue entering one element at a time, then, pressing ENTER. Use this procedure for entering the first column of data in page 19. Enter the following:

[ENTER] 27.3 [ENTER] 20.6 [ENTER] 25.4 [ENTER] 36.9 [ENTER].

You can also enter a list of values separated by spaces, before pressing the ENTER key. For example, to enter the second column of the list in page 19, enter

15.2 [SPC] 28.3 [SPC] 33.7 [SPC] 29.5 [SPC] 34.1 [SPC] 24.6 [SPC] [ENTER].

You'll notice that it takes the calculator some time to enter the data.

Once the second column of the data in page 19 has been entered, continue entering the remaining data until all values are in the vector. Press [ENTER] one last time to return to the main display.

Store the column vector in a variable called ARRAY as follows:

Press ['] and type ARRAY, then press [ENTER]. The display will show 'ARRAY' and the column vector will be in display level 2. Now press [STO].

At this point the data is stored in variable ARRAY. The name ARRAY should now be shown at the bottom of the display and on top of one of the white buttons in the top row of the calculator. To place the column vector in level 1 of the display, press the button ARRAY.

As a list, which is later transformed into a column vector
This approach takes less time than entering the data in the matrix editor as a column vector. To illustrate this approach, enter the data from problem 2.12 (P212 in sub-directory SDATA). First, we need to create a subdirectory for that problem that we'll call P212:

 © 2000 - Gilberto E. Urroz - All rights reserved

Press [↰][UP] to get to directory STATS.

Follow the procedure used to create subdirectory P2.7 to create now subdirectory P212.

Enter subdirectory P212.

We'll copy the data from P212 in the same order that we did for the data of P2.7, i.e., sweeping the data table by columns. To write the data as a list:

Press [↰][{}], and begin entering the data values separated by spaces, e.g., 65 [SPC] 43 [SPC] 88 [SPC] ... Continue until all data has been entered. Then press [ENTER]. The list will now be in level 1 of the display.

Store the list in a variable called DLIST. Press ['] DLIST [ENTER][STO].

The next step is to transform the list into a column vector. Follow these steps:

Press [VAR] [*DLIST*] to place the list in level 1 of the display.

Press [PRG] [*TYPE*] [*OBJ→*] to decompose the list into its components. These elements are listed in the order there were entered in the list and they occupy all but level 1 in the display. Level 1 contains the number of elements in the list (100, in this case).

To create a column vector we need to have all the elements of the vector occupying all but level 1 in the display. Level 1 is occupied by a list that, for the present case, looks like this {100 1}. This list indicates that we will be creating a matrix with 100 rows and 1 column. Since the number 100 is already available, to create the list {100 1} follow these steps:

Press [1] [ENTER]	This is the '1' in the list {100 1}
Press [2] [ENTER]	To indicate that we'll create a list with two elements.
[*→LIST*]	Creates the list {100 1}

To create the array, just press [→ARR]

Save the array in a variable called DVECT: Press ['] DVECT [ENTER][STO][VAR]. The original list is now a column vector that can be used for statistical analysis.

Keep subdirectories P2.7 and P212 with their corresponding data available for future use.

 © 2000 – Gilberto E. Urroz – All rights reserved

3.0 Frequency Distributions and Histograms

We'll use the information contained in a variable called DT1 to calculate frequency distributions. DT1 is simply a column vector containing the number of workers absent from a factory on 50 working days.

We will use these data to illustrate the use of frequency distributions, etc. We are required to use the following six classes 0-4, 5-9,10-14, 15-19, 20-24, and "25 or more" to construct (a) a frequency distribution, and (b) a percentage distribution.

First of all, we notice that the data is made of integer numbers. In the six classes defined above, *the class boundaries* are the numbers limiting each class, i.e., 0,4,5,9,10,14,15,19,20,24 and 25. The upper boundary of the sixth class is not defined (this is called an **open class**).

The class marks, i.e., midpoint of each class, can be calculated, at least for the first 5 classes by taking the average of the class boundaries. For example, for the first class, the class mark is: (0+4)/2 = 2; for the second one, it is (5+9)/2 = 7, and so on.

For the first five classes the class interval is the distance between consecutive class marks, i.e., 7-2 = 5.

Using the HP-48 STAT feature produces Table 1.

Please note that the calculator can only work with classes that have the same width, therefore, the last class listed above ("25 or more") can not be defined.
Also, notice that the calculator uses the terms BIN COUNT and BIN WIDTH for the number of classes and the width of each class. We will use six classes (BIN COUNT = 6) and widths of 5 for each class (BIN WIDTH = 5). Since we are using integer numbers using a BIN WIDTH of 5 includes all integers in any given class. For example, in the first class 0-4, there are 5 integer numbers included (0,1,2,3,4). Follow these steps:

Press [↱] [STAT][▼][OK] [*CHOOS*] to select "Frequencies..." from the STAT feature and to choose the data to be processed. Select MYDAT, and press [OK]. The data originally contained in DT1 is now stored in a variable called ΣDAT, which is the default name for the data matrix in STAT.

Press [▼] and enter the minimum value of data for frequency counting. For our case that will be zero. Press [OK].

Enter 6 for *BIN COUNT*, and press [OK].

Enter 5 for *BIN WIDTH*, and press [OK].

Press [OK], once more, to perform the frequency count.

At this point the display shows two vectors. The one in level 2 is the count for each class. You'll notice, for example, that class 1 has 4 elements, class 2 has 15, and the other counts are not visible. The vector in level 1 represents the number of **outliers** in the data, i.e., values that are either below the minimum value entered, or above the maximum, which is calculated by STAT. The maximum value in this frequency distribution is 29, since we are using 6 classes of 5 integer numbers each including 0. Display level 1 shows [0 3], for this case, indicating that there are no data values below 0, and 3 above 29. Checking the data in page 11, we notice that the values 32, 37 and 49 are the upper-end outliers of this data set.

 © 2000 - Gilberto E. Urroz - All rights reserved

To check the number of counts in each BIN, drop the vector in level 1: by pressing [⇦] and press [↰] [EDIT] to see all the elements in this vector. Using the up and down arrows ([▲] [▼])we can move through the vector. We note that the vector looks like this [[4] [15] [16] [8] [3] [1]]. We can summarize the information provided by the calculator in Table 1.
We can combine the counts of classes (bins) numbers 6 and 7 into a single class, identified as "25 or more," to comply with the requirements of the problem. Table 2 below shows the results when only 6 classes are used. We have also calculated the percentage frequency as:

Percentage frequency = frequency (or count)/total x 100

Table 2 also includes the class mark. In the text, the class mark of class j is referred to as x_j, while the frequency is referred to as f_j. The last column in the table shows a cumulative frequency for each class defined by the number of absences less than the upper class limit for each class. In other words, the cumulative frequency for class number 3 corresponds to the number of absences less than 14, and so on. Notice that the cumulative frequency of each class is computed by adding the individual frequencies of all the classes above it.

Table 1.

Bin No.	Number of absences	Count
1	0 - 4	4
2	5 - 9	15
3	10 - 14	16
4	15 - 19	8
5	20 - 24	3
6	25 - 29	1
7	> 29	3

Table 2.

Class Number	Number of absences	Class mark (x_i)	Frequency f_i	Percentage frequency	Cumulative frequency
1	0 - 4	2	4	8	4
2	5 - 9	7	15	30	19
3	10 - 14	12	16	32	35
4	15 - 19	17	8	16	43
5	20 - 24	22	3	6	46
6	25 or more	n/a	4	8	50
		Total	50		

Note that there is no value for the class mark in class number 6. That is because the upper class boundary is not defined. We can use as that upper boundary the maximum value in our data, which is 49, and modify Table 2 to read:

 © 2000 - Gilberto E. Urroz - All rights reserved

Table 3.

Class Number	Number of absences	Class mark (x_i)	Frequency f_i	Percentage frequency	Cumulative frequency
1	0 - 4	2	4	8	4
2	5 - 9	7	15	30	19
3	10 - 14	12	16	32	35
4	15 - 19	17	8	16	43
5	20 - 24	22	3	6	46
6	25 - 49	37	4	8	50
		Total	50		

3.1 Plot a histogram using frequency vector obtained through STAT and Frequencies

Let's plot a histogram for the data stored under variable DT2. First, perform the frequency distribution evaluation using a procedure similar to that used above.

The problem requires that we use the following classes: 15.0-19.9, 20.0-24.9, ..., 35.0-39.9, in evaluating the frequency distribution. The classes are of the same width (e.g., 24.9 - 20.0 = 4.9). However, the data in DT2 is accurate to the first decimal. Therefore, to avoid ambiguities in the tallying of the classes, we want to re-define the classes as follows: 15.00 - 19.99, 20.00 - 24.99, ..., 35.00 - 39.99. The actual bin width will be then 19.99 - 15.00 = 4.99. To determine the number of bins, we use the following information:

Minimum class boundary = 15.0;   Maximum class boundary = 39.99 ≈ 40.0;   Bin width = 4.99 ≈ 5.0

Therefore,   Number of bins = (Maximum c.b. - Minimum c.b.)/Bin width = (40.0 - 15.0) / 5.0 = 5.

To perform the frequency distribution, follow these steps:

[↱][STAT][▼][OK]	Select 'Frequencies ...'
[*CHOOS*] and select ARRAY [OK]	Select data array and place it into data matrix ΣDAT
[▼] 15.0 [OK]	Enter minimum class boundary
5 [OK]	Enter BIN COUNT
4.99 [OK]	Enter BIN WIDTH
[OK]	Perform frequency distribution calculations

Display level 2 shows the frequency counts, and display level 1 shows the outliers. There are no outliers in this case.

Steps to plot the histogram:

[⇦]	Drops the outliers vector.
[↰] [STAT] [*DATA*] [↰] [*ΣDAT*]	To store frequency vector in ΣDAT
[↰] [STAT] [*PLOT*] [*BARPL*]	To plot histogram
[CANCL]	To return to main display

 © 2000 - Gilberto E. Urroz - All rights reserved

3.2 Plot a histogram using PLOT and the original data vector

This is an alternative way to plot a histogram without having to calculate first the frequency distribution. We will use the data in variable DT2 again. To plot the histogram, follow these steps:

Press [↱][PLOT] to start the PLOT feature
Use [▲] and [▼] to highlight the space in front of TYPE:
Use [*CHOOS*], [▲] and [▼] to select **Histogram**, then press [OK]
Press [▼] to highlight the space in front of ΣDAT:
Use [*CHOOS*] and select DT2, then press [OK]

Since we have only one column in the ARRAY matrix, we don't care about changing the value of 1 in front of COL:

Use [▼] to highlight the space in front of WID: Enter 4.99 [OK].
Enter 15 [OK] 40 [OK] to define the limits of the horizontal view of the graph.

The underlined space before AUTOSCALE is now highlighted. At this point, you can place a check mark on it and let PLOT select the vertical scale by pressing [✓CHK].

Press [*ERASE*] to ensure that no previous graphs will be in the display.
Press [*DRAW*] to display the histogram.
Press [*CANCL*] to return to the PLOT display.

This display shows that the vertical scale (V-VIEW) ranges from -9 to 60. If you want to set your own scale, move the cursor to V-VIEW (do not place a check mark in front of AUTOSCALE) and change your parameters to, say, 0 and 20:

Press [▼] [▼] 0 [OK] 20 [OK]
Press [ERASE] [DRAW] to display histogram.

Press [CANCL] to go back to PLOT.
Press [ENTER] to return to the normal display.

Note: for quick access to histograms, see section 15 in Part I of this guide.

 © 2000 - Gilberto E. Urroz - All rights reserved

4.0 Calculation of descriptive measures

4.1 Calculation of means

This is a simple exercise to show the use of STAT to calculate the mean of a sample.
Enter the following data into a column vector: 5, 3, 1, 4, 2, 6, 9, 8.
Store the data into a variable called DT3.
Then, follow these steps:

[↱][STAT][OK]	Select "Single-var..."
[*CHOOS*] select DT3 [OK]	Select data in DT3 and place it in the data matrix ΣDAT
[▼][▼][◄][✓*CHK*][OK]	Select MEAN

The display shows: Mean: 4.75

4.2 Obtaining the median of a sample

The basic instructions to find the median of a sample of size *n* are as follows:

Sort the data in increasing order.
If *n* is odd, the median is observation number *(n+1)/2*.
If *n* is even, the median is the average of observations *n/2* and *(n+2)/2*.

To obtain the median of vector DT3, first we need to know the number of elements in the vector. This is obviously easy for vector DT3, but it may not be so if we have a larger vector. One way to obtain *n* is by using STAT as follows:

[↱][STAT] [▼][▼][▼] [OK]	Select "Summary stats..."
[*CHOOS*] select DT3 [OK]	Select data in DT3 and place it in the data matrix ΣDAT
Place a check mark in front of NΣ	Select the sample size
Press [OK]	

The display shows: NΣ : 8

To be able to obtain the median by the procedure shown above, we need to transform the column vector DT3 into a list. To do this, follow these steps:

[DT3]	The display shows the column vector DT3
[PRG][*TYPE*][*OBJ→*]	Displays vector elements in different levels of display, and a list {8 1}, indicating, for this case, that a matrix with 8 rows and 1 column was decomposed
[EVAL][⇦] [→*LIST*]	Decompose the list {8 1} into its components, 8 goes to level 2, 1 to level 1. Drop 1 from the display, and create a list with 8 elements.

The elements of the list just created occupied levels 2 through to 9 in the display.

[MTH][*LIST*][*SORT*]	Orders elements in list from smallest to largest.
[ENTER][ENTER]	Create two additional copies of the list for later use.
[PRG][*LIST*][*ELEM*][*SIZE*]	Determines the size of the list.

This last step determines that n = 8. While we don't really need to use the calculator to find out the size of a list of 8 elements, this step will be necessary for a larger list. Since n is even,

 © 2000 - Gilberto E. Urroz - All rights reserved

we need to find elements n/2 = 4, and (n+2)/2 = 5, and find their average to determine the median. To extract those elements from the ordered list, use these steps:

[⇦]	Drops the 8 from level 1 of the display.
[4][GET]	To extract element 4 from the list, we enter the list {4}.

The command GET extracts the element of the list in level 2 occupying the position indicated in the list in level 1.

[↰][SWAP]	Swaps the contents of levels 1 and 2 in the display. The ordered list is now in level 1, while the value of element number 4 in the ordered list is in level 2.
[5][GET]	Extracts element 5 from the list, and puts it in level 1.
[+][2][÷]	Add elements 4 and 5 (in levels 1 and 2) and divides by 2.

The result is the median of the list, 4.5.

If the size of the list had been odd, we needed only to extract element number (n+1)/2 to get the median.

This exercise shows you how to transform a column vector into a list, how to obtain the size of the list, and how to extract elements from a list. The procedure is, however, tedious. Luckily enough, the HP48G/GX, has a programmed procedure to extract the median of a column vector (or the medians of each column of a data matrix). This is explained in the following section.

4.3 Program for calculating the median of a sample

First, check if you have already installed the EXAMPLES directory in your calculator. Go to the HOME directory and check if there is a label EXAM on top of one of the white buttons. If that directory does not exists, of if you have created a directory called EXAM or EXAMPLES -- which is not the one provided by the HP48 -- you need to load the EXAMPLES directory first.
To load the EXAMPLES directory, type TEACH in level one of the display and press [ENTER]. The directory EXAMPLES will be placed in your HOME directory. The label on the white button corresponding to this directory reads only 'EXAM'.

To use the program for calculating the median of the vector DT1, follow these steps:

- Press [VAR][DT1] to list the contents of the column vector DT1 in level 1 of the display.
- Go back to the HOME directory, press [↱][HOME].
- Go into the directory that contains the program that calculates the median, press [EXAM][PRGS]

Now, store the vector in display level 1 into the variable ΣDAT, by pressing:

['][↱][CHARS][▶] [▶][▶][▶][▶][*ECHO*][ENTER][α][α][D][A][T][α][STO].

To get the median, press the white button [*MEDIA*]. The result shows a vector with a single element, [28.35]. The reason why the result comes as a vector is because the program MEDIA calculates the median of every column in the ΣDAT matrix. For our case, however, the ΣDAT matrix has only one column.

Note that we need to get to subdirectory HOME\EXAMPLES\PRGS before storing the vector in level into the data matrix ΣDAT. If we had saved the vector into ΣDAT while still in the

 © 2000 - Gilberto E. Urroz - All rights reserved

directory where DT1 is stored, or in any other directory, such variable would not be available in the PRGS subdirectory. This illustrates the point that you can have the same variable name in two different subdirectories storing completely different data. In other words, variables are not shared across subdirectories.

4.4 Obtaining descriptive measures and summary statistics of a sample

The STAT feature in the HP48 series calculator will let you find not only the mean and number of elements in a sample, but also other descriptive measures (standard deviation, variance, the total sum of the elements, the maximum and minimum values of the sample), as well as, summary statistics. Summary statistics are values such as Σx, Σx^2, for a sample. The elements of the sample are referred to as x_i.

To find the *descriptive measures* in the data in DT1, for example, use the following steps:

- Go to the appropriate subdirectory.
- Select DT1 as ΣDAT.

- Start the STAT feature and select the "Single-var..." option by pressing: [↱][STAT][OK].

- Using the arrow keys, i.e., [◀], [▶], [▲], [▼], and [✓CHK], place check marks in front of all descriptive measures of interest. For example, check MEAN, STD DEV, VARIANCE, TOTAL, MAXIMUM, and MINIMUM, for a complete set of descriptive measures.

- Press [OK] to calculate the descriptive measures. They will be shown in different levels of the main display clearly identified.

For the case of DT1, the values of the descriptive measures are listed below. The variables shown between parentheses after the values are commonly used to refer to these quantities.

Mean: 28.095 ($\bar{x}$) ; Std Dev: 5.2821228 (s); Variance: 27.90082 (s^2) ;

Total: 1685.7 (Σx); Maximum: 38.4 (x_{max}); Minimum: 15.2 (x_{min})

The 'Summary statistics' option of the STAT feature in the HP48 series calculator is designed to provide *summary statistics* for two data samples of the same size (NΣ). The samples are contained in columns of the ΣDAT matrix. One of the columns is referred to as x and the second one as y. By default, column 1 is x, and column 2 is y. However, you can select any other columns for x or y if ΣDAT contains more than two columns.

The 'Summary statistics' option of STAT allows you to calculate values such as Σx, Σy, Σx^2, Σy^2, Σxy and NΣ. If ΣDAT includes only one column, as in the case of problem 2.7, then only values such as Σx, Σx^2, and NΣ can be calculated.

For example, to calculate summary statistics for the data of DT1 follow these steps (Assuming you are in the proper subdirectory):

[↱][STAT][▼][▼][▼] [OK]	Selects "Summary stats..."
[CHOOS] select DT1 [OK]	Enters the data in DT1 into the matrix ΣDATA

Using the arrow keys, i.e., [◀], [▶], [▲], [▼], and [✓CHK], place check marks in front of all descriptive measures of interest, namely, as Σx, Σx2, and NΣ. Press [OK] to calculate statistics.

The display will show the following results: ΣX: 1685.7 (Σx); ΣX2: 49005.89 (Σx^2); NΣ: 60 (n)

 © 2000 - Gilberto E. Urroz - All rights reserved

4.5 Calculation of percentiles

The basic procedure to calculate the *100 p-th* Percentile ($0 < p < 1$) in a sample of size *n* is as follows:

- Order the n observations from smallest to largest.
- Determine the product *np*
- If *np* is not an integer, round it up to the next integer and find the corresponding ordered value.
- If *np* is an integer, say *k*, calculate the mean of the *k*th and *(k-1)*th ordered observations.

[Note: Integer rounding rule, for a non-integer x.yz…, if y ≥ 5, round up to x+1; if y < 5, round up to x.]

For example, in variable DT4 we have the data entered the same data as in array DT1, but this time as a list. If we want to calculate the 37^{th} percentile ($p = 0.37$) of that data set, we proceed as follows (assuming you are in the appropriate subdirectory):

[VAR][DT4] Places contents of DT4 in display level 1.
[MTH][LIST][SORT] Orders list from smallest to largest.
[ENTER] [ENTER] Creates two more copies of the list for later use.
[PRG][LIST][ELEM][SIZE]Gives *n* as 60 ($n = 60$)
[0][.][3][7][×] Enter *p* in level 1 of display and multiply *n* times *p*, to give 22.2.
Round that number up to 22.
[⇦] Drop 22.2 from level 1.
[2][2][GET] Indicates that element number 22 of the ordered list in display level 2 is
to be extracted. GET produces a value of 26.4 for the percentile.

We write the result as $P_{0.37} = 26.4$ for the data in DT4.

If we wanted to obtain the third quartile of the data in DT4, i.e., $Q_3 = P_{0.75}$, with n = 60 and p = 0.75, we find that $np = 45 = k$ is indeed an integer. Therefore, to determine Q_3 we extract from the ordered list elements number 44 (= *k-1*) and 45 (=*k*) and calculate their average. The procedure is as follows:

[VAR][DT4] Places contents of DT4 in display level 1.
[MTH][LIST][SORT] Orders list from smallest to largest.
[ENTER] [ENTER] Creates two more copies of the list for later use.
[PRG][LIST][ELEM] Displays programs that operate on elements of lists.
[4][4][GET] Gets element number 44 of the ordered list. Display level 1 shows a value of 31, i.e., $x_{44} = 31$, where x_i represents the *i*th element of the ordered list.
[↰][SWAP] Swaps objects in levels 1 and 2 of the display, placing the ordered list in level 1.
[4][5][GET] Gets element number 45 of the ordered list. Display level 1shows that $x_{45} =$ 31.3.
[+][2][÷] Calculates $Q_3 = (x_{44} + x_{45})/2$. Display level 1 shows that $Q_3 = 31.15$

Please notice that there is a variety of ways to calculate percentiles, and that the way presented above may not be the same utilized in your textbook.

 © 2000 - Gilberto E. Urroz - All rights reserved

4.6 A program for calculating percentiles

The HP48G provides a program for calculating (%TILE) under the directory HOME\EXAM. This program requires the user to place a list of data values in level 2 of the display, and the required percentile (a quantity from 0 to 100) in level 1 of the display. To access the program enter: [↱][HOME][EXAM]. The program %TILE is available in one of the white keys in the top of the keyboard.

To illustrate the use of the program we will use the data in DT4. Let's calculate the 37th percentile, using the following:

[VAR][DT4]	Places the list of data in level 1 of display.
[3][7][ENTER]	Places percentile value in level 1 and the list in level 2.
[↱][HOME][EXAM][%TILE]	Moves to directory EXAM and executes program.

The result in this case is $P_{0.37}$ = 26.6. Previously, we calculated $P_{0.37}$ to be 26.4 by using the procedure outlined earlier. The way that %TILE calculates percentiles is slightly from the one outlined above. However, the differences in the results should not be that significant.

Because the median is, by definition, the 50-th percentile, you need not keep both the MEDIAN and %TILE programs in your calculator. If you use lists for manipulating your data you need only to keep the %TILE program. The median can be calculated using: [5][0][%TILE].

See section 1 in Part II of this guide for instructions on programs for calculating descriptive measures of a sample. The programs in that section also allowed you to obtain frequency distributions and plot different type of graphs for data analysis.

<u>Note</u>: for quick access to calculating descriptive measures, see section 15 in Part I of this guide.

 © 2000 - Gilberto E. Urroz - All rights reserved

5.0 Calculation of descriptive measures in grouped data

Grouped data refers to data grouped into classes and presented as a group of class marks (x_i) and their corresponding frequency distribution (f_i). Assuming the we have a total of *n* data points grouped into *k* classes, the calculation of the mean and variance proceeds according to the formulas used below.

The HP48 series calculators can be used to calculate the summations in those formulas if the vectors *x* and *f*, containing class marks and frequencies, respectively, are entered as lists. To illustrate the calculations, we'll use the data in DT5. These data consist of a list of two lists: {{6.95 10.95 ... 30.95} {3 10 ...2}}. Separate the two lists and store them in variables *x* and *f*, respectively, by using:

[*DT5*][PRG][*TYPE*][*OBJ→*][⇦]	Decomposes list of two lists, drops number of elements (2).
['][α][↰][F][STO]	Store second list into variable *f*.
['][α][↰][X][STO]	Store first list into variable *x*.

To calculate Σ x_i f_i, for example, follow these steps:

[VAR][X][F][×]	Places the x and f lists in display levels 2 and 1, respectively, and multiply their corresponding elements creating a new list, which contains the elements of the third column in the table of page 32.
[MTH][LIST][ΣLIST]	Calculates the sum of all the elements in the list in display level 1. The display shows now that Σx_i f_i = 1508 (check Table in page 32).

To calculate the number of elements in the unordered sample, we use the following keystrokes:

[VAR][F][MTH][LIST][ΣLIST]	The display shows that n = 80.

Since Σ x_i f_i is in display level 2 and n is in display level 1, we can calculate $\bar{x} = \Sigma x_i f_i /n$, by pressing [÷], giving as a result $\bar{x}$ = 18.85.

Using the same data vectors x and f, we can calculate Σ x_i^2 f_i as follows:

[VAR][X][↰][x^2]	Places the x list in display levels 1, and takes the square of each element in list x.
[F][×]	Places the f list in display level 1, while the x list is moved to level 2. We, then, multiply the two lists element by element.
[MTH][LIST][ΣLIST]	Calculates the sum of all the elements in the list in display level 1. The display shows now that Σ x_i^2 f_i = 30857.

We can use the information obtained so far to calculate s^2 according to the formula shown below. We know that n = 80, Σ x_i f_i = 1508, and Σ x_i^2 f_i = 30857, then,

$$s^2 = [n\ \Sigma x_i^2\ f_i - (\Sigma x_i\ f_i)^2\ ]/[n\ (n-1)] = [80\times 30857\text{-}1508^2]/[80\times 79] = 30.77.$$

See section 2 in Part II of this guide for instructions on programs for calculating descriptive measures in grouped data.
Note: for quick access to summary statistics calculations, see section 15 in Part I of this guide.

 © 2000 - Gilberto E. Urroz - All rights reserved

6.0 Discrete random variables

Suppose that we want to calculate the mean, variance and standard deviation for a discrete random variable. First, we'll use the calculator to compute those parameters when the probability distribution is given as a table.

For example, using the data, x = {0 1 2 3 4 } and f = {0.05 0.20 0.45 0.20 0.10}, we begin by creating a subdirectory HOME\STATS\DDIST (for Discrete DISTributions), and moving into that subdirectory for the calculations. Next, enter the values of x and f as lists, namely, x = {0 1 2 3 4 }, and f = {0.05 0.20 0.45 0.20 0.10}. Recall that to store those any list into a variable you need to type in the list in level 1 of the display, then type the variable name between quotes, and press [STO]. Once both lists have been stored, you will have two white buttons at the top of the keyboard showing the names [X] and [F].

First, it is convenient to check if the probability distribution is valid by checking that every element in f is larger than or equal to zero, and by checking that they add up to 1.0. To check that they add up to one use:

[F][MTH][LIST][ΣLIST]	If display level 1 shows a 1.0, you can continue with the analysis. If not, the distribution is not valid.

To calculate the mean, μ, we need to multiply the lists x and f and sum the elements of the result, as follows:

[VAR][X][F][×][MTH][LIST][ΣLIST]	Display level 1 should show 2.1, i.e., $\mu = 2.1$.

To calculate the second moment with respect to the origin, μ'_2, we use:

[VAR][X][↰][x^2][F]	Calculates the list $x^2 \cdot f$.
[MTH][LIST][ΣLIST]	Calculates $\mu'_2 = 5.4$.

To calculate the variance, with μ in level 2 and μ'_2 in level 1 of the display, try the following:

[↰][SWAP][↰][x^2][-]	Calculates $\sigma^2 = \mu'_2 - \mu^2 = 0.99$

See section 3 in Part II of this guide for instructions on programs for calculating discrete distribution parameters. Part II also describes sub-directories that allow you to generate synthetic data obeying a given discrete probability distribution and to analyze such data.

 © 2000 - Gilberto E. Urroz - All rights reserved

7.0 Factorials, permutations, combinations, and discrete probability distributions

The factorial of an integer *n* is defined as: $n! = n\cdot(n-1)\cdot(n-2)\ldots3\cdot2\cdot1$. By definition, 0! = 1. Factorials are used in the calculation of the number of permutations and combinations of objects. For example, the number of permutations of *r* objects from a set of *n* distinct objects

$$_{n}P_{r} = n(n-1)(n-1)\ldots(n-r+1) = n!/(n-r)!$$

is

Also, the number of combinations of *n* objects taken *r* at a time is

$$\binom{n}{r} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

To simplify notation, I'll use P(n,r) for permutations, and C(n,r) for combinations.

The HP48 series calculators provide functions to calculate factorials, permutations and combinations by entering [MTH][NXT][*PROB*]. The white button's display now shows five functions:

[*COMB*][*PERM*][*!*][*RAND*][*RDZ*]

At this point, we are interested only in the first three of them. Their operations are described below:

[*COMB*]: Calculates the number of combinations of n (level 2) items taken r (level 1) at a time;
[*PERM*]: Calculates the number of permutations of n (level 2) items taken r (level 1) at a time;
[*!*]: Factorial of a positive integer (in level 1). For a non-integer, x [*!*], returns $\Gamma(x+1)$.

For example, to calculate 8!, enter

[8][MTH][NXT][*PROB*][*!*].

The display shows a value of 40320.

To calculate P(15,3), enter

[1][5][ENTER][3][ENTER][MTH][NXT][*PROB*][*PERM*],

or, if you are already in the MTH\PROB subdirectory, enter

[1][5][ENTER][3][ENTER][*PERM*].

The result is 2730.

To calculate C(15,3) enter

[1][5][ENTER][3][ENTER][MTH][NXT][*PROB*][*COMB*],

or, if you are already in the MTH\PROB subdirectory, enter

 © 2000 - Gilberto E. Urroz - All rights reserved

[1][5][ENTER][3][ENTER][*COMB*].

The result is 455.

Factorials and combinations are used in defining the binomial and other discrete probability distributions, as shown below.

7.1. Binomial probability distribution function

Create a directory HOME\STATS\DFUN (Discrete FUNctions). We will define some probability distributions and distributions functions in that directory.
To define the binomial distribution function, enter the following keystrokes:

['][α][α][↰][B][↰][P][↰][D] [↰] [( )] [↰] [X] [↰] [,] [↰] [N] [↰] [,] [↰] [P][α][▶][↰][=]
[MTH][NXT][*PROB*][*COMB*][α][α][↰][N][↰][,][↰][X] [α] [▶]
[×] [α][↰] [P][y^x] [α][↰] [X]
[×] [↰] [( )][1][-][α][↰] [P] [▶] [y^x] [↰] [( )] [α][↰] [N][-] [α][↰] [X]
[ENTER]

The calculator display will show:

```
1:     'bpd(x,n,p)=COMB(n,
       x)*p^x*(1-p)^(n-x)'
```

This statement defines a function *bpd* (binomial probability distribution) of a discrete random variable x = 0, 1, ..., n, with parameters n and p. We'll now define the function into a variable 'bpd'. *Enter*:

[↰][DEF].

Press [VAR], if needed. To see the contents of the variable 'bdp', enter:

[↱][*BPD*].

The display will show a program:

```
1:     << → x  n  p  'COMB(n,x
        )*p^(1-p)^(n-x)'
        >>
```

The sequence [↰][DEF] converts the function definition, *'bpd(x,n,p)=COMB(n,x)*p^x*(1-p)^(n-x)'*, into the program listed above.
To use the function we just defined, we need to enter three arguments, namely, x, n, and p, into display levels 3, 2, and 1, respectively. To do that, simply use the keystroke sequence:

x [ENTER] n [ENTER] p [ENTER].

To determine the value of bpd(x,n,p), press the white key corresponding to the label [*BPD*].
For example, to calculate bpd(2,5,0.5), enter:

[2][ENTER][5][ENTER][0][.][5][ENTER][BPD]   The display shows a value of 0.3125.

 © 2000 - Gilberto E. Urroz - All rights reserved

7.1.1 Binomial distribution function

To define the binomial distribution function, type in the following function definition:

```
'BDF(x,n,p)  = Σ(k=0,x,bdp(k,n,p))'
```

Then, press [←][DEF]. The variable BDF is now defined. If you press [→][*BDF*], you'll get the following program:

```
<<  → x  n  p ' Σ(k=0,x,bdp(k,n,p))' >>
```

{ **Note**: To get the symbol Σ, you can use the sequence [→][Σ]. (This corresponds to the key whose main function is [TAN])}

Calculate the value of BDF(3,5,0.5), by pressing:

[3][ENTER][5][ENTER][0][.][5][ENTER][BDF]   The display shows a value of 0.8125.

7.2 Other discrete probability distributions and distribution functions

The following function definitions will provide you with other probability distributions and distribution functions for discrete random variables:

Poisson Probability Distribution: `'PoPD(x,λ)  = λ^x*EXP(-λ)/x!`
Poisson Distribution Function: `'PoDF(x, λ)  = Σ(k=0,x,'PoPD(k,λ))'`
Hypergeometric Prob. Distr.: `'hpd(x,n,a,N)=COMB(a,x)*COMB(N-a,n-x)/COMB(N,n)'`
Geometric Distribution Function: `'gpd(x,p)  = p*(1-p)^(x-1)'`

Remember to press [←][DEF] after entering the function definition in level 1 of the display. To facilitate locating the functions, keep them all under the same directory DFUN. If you define all the functions suggested in this section, your white keys should have the following labels (they may be in different order):

[*BPD*][*BDF*][*POPD*][*PODF*][*HPD*][*GPD*]

Also, BDF uses *bpd*, and PoDF uses PoPD, therefore, you must define *bpd* and PoDF before using BDF or PoDF.

Note: If the functions in your directory are not in the order shown above and you want to place them in that order, try the following:

(1) Create a list that looks like this: { bpd BDF PoPD PoDF hpd gpd} and place it in level 1. To create the list use these keystrokes:

[←][{}] [BPD][BDF][POPD][PODF][HPD][GPD][ENTER]

(2) Use the following keystroke sequence to re-order the variables:

[←][MEMORY][DIR][ORDER].

 © 2000 - Gilberto E. Urroz - All rights reserved

7.2.1. Examples:

Calculate the following:	Keystrokes:	Result:
PoPD(x=3, λ=2.2)	[3][SPC][2][.][2][POPD]	0.19663867…
PoPF(x=5, λ = 1.5)	[5][SPC][1][.][5][POPF]	0.99554401…
hpd(x=2,n=10,a=20,N=100)	[2][SPC][1][0][SPC][2][0][SPC] [1][0][0][HPD]	0.31817062…
gpd(x=5,p=0.25)	[5][SPC][.][2][5][GPD]	0.07910156…

 © 2000 - Gilberto E. Urroz - All rights reserved

8.0 Generating random numbers

The HP48G series calculator has a random number generator that returns a real number between 0 and 1. The generator is able to produce sequences of random numbers. However, after a certain number of times (a very large number indeed), the sequence tends to repeat itself. For that reason, the HP48G generator is referred to as a *pseudo-random* number generator. To generate a random number with your calculator, press:

[MTH][NXT][*PROB*][*RAND*].

To generate a sequence of numbers just keep pressing the [*RAND*] white key.
If you want to generate a sequence of number and be able to repeat the same sequence later, you can change the "seed" of the generator by entering a given number in level 1 and pressing [*RDZ*] before generating the sequence. Random number generators operate by starting with a "seed" number that is transformed into the first random number of the series. The current number then serves as the "seed" for the next number and so on. By "re-seeding" the sequence with the same number you can reproduce the same sequence more than once. For example, try the following:

[.][2][5][*RDZ*]	Use 0.25 as the "seed."
[*RAND*]	First random number = 0.75285...
[*RAND*]	Second random number = 0.51109...
[*RAND*]	Third random number = 8.5429...E-2 = 0.085429....

Re-start the sequence:

[.][2][5][*RDZ*]	Use 0.25 once more as the "seed."
[*RAND*]	First random number = 0.75285...
[*RAND*]	Second random number = 0.51109...
[*RAND*]	Third random number = 8.5429...E-2 = 0.085429........

If you press [*RDZ*] with no value in the display, the generator will take a number based on the calculator's clock time and use it as the seed.
The pseudo-random number generator provided in your calculator produces random numbers with a *uniform distribution* in the interval [0,1]. To learn more about uniform distributions refer to a statistics textbook.
See section 5 in Part II of this guide for instructions on programs that generate lists of random numbers.

 © 2000 - Gilberto E. Urroz - All rights reserved

9.0 Continuous random variables

9.1. Calculating probabilities and distribution parameters

For probability distributions of continuous random variables, we are seldom interested in the distribution function (called *probability density function*, or *pdf*), i.e., we are seldom interested in finding f(x) = P(X=x). Quantities of interest will be the distribution functions (also known as *cumulative distribution functions* or *cdf*), i.e., F(x) = P(X ≤x), which are calculated using integrals. We can use the HP48G series calculator to evaluate such integrals either symbolically or numerically. Following we present some examples within a new subdirectory HOME\STATS\INTS :

1) Suppose that the pdf of a continuous random variable is given by $f(x) = K/(1+x^2)$, for $-\infty < x < \infty$. We are asked to find the value of K. By definition,

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

for this particular case, we can write

$$\int_{-\infty}^{+\infty} f(x)dx = K\int_{-\infty}^{+\infty}\frac{dx}{1+x^2} = K \cdot I = 1$$

We can calculate the value of the integral $I = {}_{-\infty}\int^{+\infty} dx/(1+x^2)$, by using the following keystroke sequence:

[↱][SYMBOLIC][OK]	Selects "Integrate..."
[1][÷][↰][( )][1][+][α][↰][X][y^x][2][OK]	Enters the integrand $1/(1+x^2)$
[α][↰][X][OK]	Enters integrating variable
[-][MTH][NXT][CONS][NXT][9.999][ENTER]	Enters -∞ ([1])as the lower limit
['][MTH][NXT][CONS][NXT][9.999][ENTER]	Enters +∞ ([2])as the upper limit
[*CHOOS*][OK]	Selects *symbolic* result ([3])
[OK]	Shows symbolic integral

([1]) The HP48G uses the value $9.99999999999\times 10^{499}$ as the largest possible value in its universe, therefore, that constant (also referred to as 'MAXR') is entered as infinity. Also, instead of using the actual value for the largest number, you can use the variable MAXR, i.e., press [MTH][NXT][*CONS*][*MAXR*].

([2]) To load the constant MAXR in this location, we need to enter ['] first. You can also start by pressing [+] before loading MAXR to avoid using ['].

([3]) **Warning** If you choose *Numeric,* instead of *Symbolic,* the calculator will try to use a numerical algorithm to calculate the integral. When your integration limits are -infinity to + infinity , this is not a good idea. Therefore, for improper integrals it is better to try a *Symbolic* integration. If there is no closed-form integral for your function, level 1 of the display will show an integral expression indicating that it can not calculate the integral. In that case, it is better to use the procedure for improper integrals suggested in page 20-2 in the ***HP 48G Series****User's Guide* and illustrated below.

 © 2000 - Gilberto E. Urroz - All rights reserved

At this point the display shows:

```
1:  '1*(ATAN(x)/∂x(x))|(x=9.99999999999E499) -
(1*(ATAN(x)/∂x(x))|(x=9.99999999999E499)'
```

{**Note**: To see the entire expression use [↰][EDIT]. Do not just press [▼] (which triggers EDIT automatically), because it will take you to the EQUATION editor, and it will consume a lot of time. The symbol `∂x()` in the HP48G simply means d/dx ( ), so, in this case, `∂x(x)` = 1.}

To calculate the value of the integral in level 1 of the display, simply press [↰][→ NUM]. Level 1 in the display shows a value of 3.14159265358 (i.e., π). Therefore, we find I = π, and, k = 1/I = 1/π = 0.318309.

2) The integral calculated above is an improper integral (i.e., one or both limits are ±∞). You can use the following procedure to transform the integral into a proper integral by using the variable *y= tan(x).* For any integral, the transformation is expressed by the following formula:

$$\int_a^b f(x)dx = \int_{\arctan(a)}^{\arctan(b)} f(\tan y)(1+\tan^2 y)dy$$

If *a* = -∞, then *arctan(a)* = -π/2. Also, if *b* = ∞, then *arctan(b)* = π/2. Also, *arctan (0) = 0.* For the integral, *I* = $_{-\infty}\int^{+\infty} dx/(1+x^2)$, the expression to use is then,

$$\int_{-\infty}^{\infty}\frac{dx}{1+x^2} = \int_{-\pi/2}^{\pi/2}\frac{1}{(1+\tan^2 y)}\cdot(1+\tan^2 y)dy = \int_{-\pi/2}^{\pi/2} dy = \pi$$

In this case, the integral simplifies so much that there is no need to use the integration feature of the calculator. The example below shows a different improper integral that uses the procedure outlined above.

3) Take the expression for the Standardized Normal distribution,

$$f(x) = \frac{1}{\sqrt{2\cdot\pi}}\exp(-\frac{x^2}{2}).$$

Prove that, for this distribution, $_{-\infty}\int^{+\infty} f(x)dx=1$. The variable transformation y = tan x, produces the following:

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\cdot\pi}}\exp(-\frac{x^2}{2})dx = \int_{-\pi/2}^{\pi/2}\frac{(1+\tan^2 y)}{\sqrt{2\cdot\pi}}\cdot\exp(-\frac{\tan^2 y}{2})dy$$

One way to enter the integral, without having to use the Symbolic environment, is to type it in the Equation writer using the following keystrokes:

 © 2000 – Gilberto E. Urroz – All rights reserved

[↰][EQUATION]	Starts equation editor
[↱][∫] [-][↰][π] [÷][2] [▶][▶]	Enters lower limit of integration
[↰][π] [÷][2][▶][▶]	Enters upper limit of integration
[↰][( )][1][+][TAN] [α][↰][Y] [▶][y^x][2] [▶][▶]	Enters (1+$\tan^2$y)
[÷][$\sqrt{x}$][2][×][↰][π] [▶][▶]	Enters $\sqrt{2}\cdot\pi$
[↰][e^x][-][TAN] [α][↰][Y] [▶][y^x][2][÷][2] [▶][÷][2]	Enters EXP(-$\tan^2$y/2
[▶][▶][▶][α][↰][Y]	Enters dy
[ENTER]	Enters expression in display

Level 1 of the display now shows the integral as:

```
∫(-π/2,π/2,(1+TAN(y)^2)/√2*π)*EXP(-TAN(y)^2/2),y)'
```

You may try to evaluate it symbolically by pressing [EVAL]. However, all what you get is the same expression. This indicates that the calculator can not evaluate it symbolically, as it did with the first example shown above. A numerical integration is in order, therefore, make sure that your angle mode is *radians* and press

[↰][→NUM]

Because this is the Standardized Normal distribution we know that the integral converges. However, it takes the calculator about 2 minutes to obtain the value of the integral by numerical methods (the value is 1).

4) Some integrals, for example,

$$\int_1^\infty \frac{dx}{x} = \int_{\pi/4}^{\pi/2} \frac{(1+\tan^2 y)}{\tan y} dy$$

do not converge to a value. And, in most cases, there is no way to tell from just looking at the integral that such is the case. (The case above is simple, since we know that $_1\int^x dx/x=\ln(x)$, therefore, $_1\int^\infty dx/x=\ln(\infty)=\infty$.) Let's try to evaluate the transformed integral shown above, while, at the same time, introducing another way to enter the integral in the display (i.e., by directly typing it into level 1), as follows:

['] [↱][∫] [↰][π] [÷][4] [↰][,]	Enters lower limit of integration
[↰][π] [÷][2][↰][,]	Enters upper limit of integration
[↰][( )][1][+][TAN] [α][↰][Y] [▶][y^x][2] [▶]	Enters (1+$\tan^2$y)
[÷][TAN] [α][↰][Y] [▶][↰][,]	Enters tan y, finish entering integrand
[α][↰][Y]	Enters the variable of integration (y)
[ENTER]	Enters expression in display

Level 1 of the display now shows the integral as:

```
'∫(π/4,π/2,(1+TAN(y)^2)/TAN(y),y)'
```

 © 2000 - Gilberto E. Urroz - All rights reserved

Note: Whether you use the Equation editor, the Symbolic environment, or you type the integral directly in level 1 of the display (as we just did), the calculator ends up with an expression of the form:

```
∫(lower limit, upper limit, integrand, variable of integration)'
```

Once you get the integral expression in level 1, pressing [EVAL] makes the calculator try to find a closed-form expression for the integral. If none is available, the calculator will show the integral expression again. At such point, a numerical integration may be attempted by pressing [↰][→NUM].

We could, at this point, force a numerical calculation by pressing [↰][→NUM]. Before trying that, however, press [ENTER] to have a second copy of the integral available for future use. Now, press [↰][→NUM].
The calculator will try a numerical integration at this point. However, you will notice after 2 or 3 minutes that there is no convergence. Press [ON] a couple of times to stop the process. Once the hourglass symbol disappears from the top of the display, you will get, in level 1, a value resulting from the most current operation before we stopped the numerical integration. Just ignore this value. Level 2 should show the second copy of the integral that we created earlier. As you can see, when trying a numerical integration, the calculator can not tell you whether the integral is converging or not. A waiting period of up to 5 minutes is the best judge here. For example, the integral shown above for the Standardized Normal distribution converges to a value in about 2 minutes.
Let's try to see what happens if we use [EVAL] instead of the numerical integration. Drop the value in level 1, by pressing [↰][DROP], and press [EVAL]. You get the following expression (Press [↰][EDIT] to see the entire expression in the display):

```
'1*(LN(SIN(y))/∂y(y))|(y=π/2)-
(1*(LN(SIN(y))/∂y(y))|(y=π/4))+∫(π/4,π/2,TAN(y)^2/TAN(y),y)'
```

Press [ENTER] and [EVAL] once more to get:

```
'-LN(SIN(π/4))+∫(π/4,π/2,TAN(y)^2/TAN(y),y)'
```

Pressing [EVAL] once more produces the same expression, i.e., the calculator can not simplify the expression anymore nor evaluate the integral.
Let's see if we can help the calculator by separating the last integral and trying to evaluate it. Press [ENTER], to keep an additional copy, just in case, then, press [↰][EDIT] to edit *-LN(SIN(π/4))+* out of the expression. Press [ENTER] when done to get

```
'∫(π/4,π/2,TAN(y)^2/TAN(y),y)'
```

Press [EVAL]. Still, we get the same expression, indicating that the calculator is unable to symbolically evaluate this integral. We could force a numerical integration, by pressing [↰][→NUM], but, as before, the calculator will take a long time and there will be no convergence of the solution.
Notice, however, that we can simplify the integrand even more since *TAN(y)^2/TAN(y)= TAN(y)*. We can let the calculator do the simplification by pressing [↰][SYMBOLIC][COLLECT]. The resulting integral is:

```
'∫(.25*π, .5*π,TAN(y),y)'
```

 © 2000 - Gilberto E. Urroz - All rights reserved

Now press [EVAL] to get:

'-LN(COS(y))/∂y(y) | (y=.5*π) - (-LN(COS(y))/∂y(y) | (y=.25*π))'

Press [EVAL] once more to get:

'-LN(COS(.5*π))+LN(COS(.25*π))'

Press [EVAL] even once more. We get the same expression, meaning that the calculator cannot simplify it anymore. If we press [↰][→NUM] in an attempt to get a numerical value out of the expression we'll get a complex value:

(25.6545,-3.1415)

In other words, '∫(.25*π,.5*π,TAN(y),y)= (25.6545,-3.1415)'

To calculate the value of the original integral, ∫(π/4,π/2,(1+TAN(y)^2)/TAN(y),y), we would have to add to the complex value shown above the value of the quantity we edited out earlier, namely, *-LN(SIN(π/4))*= 0.3465, to get (26.0011, -3.1416). However, the simplification performed above in which we replaced *TAN(y)^2/TAN(y)* with *TAN(y)* is flawed, since the original integrand, namely, *TAN(y)^2/TAN(y)*, is not defined for *y = π/2*. We must, by definition, try to evaluate the integral

'∫(π/4,π/2,TAN(y)^2/TAN(y),y)'

However, a numerical integration of this expression will not converge to a value.

The purpose of this exercise is to show you that you must have an understanding of the behavior of the integrand in order to determine whether the integral would converge to a value or not. As you saw in this example, a flawed simplification of the integrand can get you a result that may seem reasonable but is not.

5) On to a simpler example: If *f(x) = cos x*, for *0 < x < π/2*, and *f(x) = 0*, *elsewhere*, find *P(0<X< π /4)*. We need to calculate the following integral:

$$P(0 < X < \frac{\pi}{4}) = \int_0^{\pi/4} \cos(x)dx$$

Enter the following:

[↱][SYMBOLIC][OK]	Selects "Integrate..."
[COS][α][↰][X][ENTER]	Enters the integrand *cos(x)*
[α][↰][X][OK]	Enters integrating variable
[0][OK]	Enters 0 as the lower limit
[↰][π][÷][4][OK]	Enters π /4 as the upper limit
[*CHOOS*][OK]	Selects *symbolic* result (*)
[OK]	Shows symbolic integral

(*) If we had selected Numeric here, the calculator would have placed the numerical value (0.707106...) directly in level 1 of the display.

 © 2000 - Gilberto E. Urroz - All rights reserved

The display shows:

1: 'SIN(x)/δx(x) | (x=π/4) - (SIN(x)/δx(x) | (x=0))'

Now, press [↰][→ NUM] to get a numerical result, 0.707106….

The examples shown above illustrate how to use the integration features of the HP48G series calculator for distribution functions of continuous random variables. Integration is used to calculate the mean and variance of continuous pdf s.

6) To calculate the mean [$\mu = \int x \cdot f(x)dx$] of the pdf in case 5, use the following:

[↱][SYMBOLIC][OK]	Selects "Integrate..."
[α][↰][X][×][COS][α][↰][X][ENTER]	Enters the integrand $x \cdot cos(x)$
[α][↰][X][OK]	Enters integrating variable
[0][OK]	Enters 0 as the lower limit
[↰][π][÷][2][OK]	Enters π /2 as the upper limit
[CHOOS][▼][OK]	Selects *numeric* result
[OK]	Shows numeric value of integral

The display now shows: 0.57076 (i.e., μ = 0.57076)

Store it in a variable called μ:

['][↱][CHARS][▶][▶][▶][▶][▶][▼][▼][▼][ECHO][ENTER][STO]

7) To calculate the variance [$\sigma^2 = \int (x-\mu)^2 f(x)dx$] of the pdf defined in 2, whose mean was calculated in 3, use the following:

[↱][SYMBOLIC][OK]	Selects "Integrate..."
[↰][( )][α][↰][X][-][VAR][μ][▶]	
[y^x][2][×][COS][α][↰][X][ENTER]	Enters the integrand $x \cdot cos(x)$
[α][↰][X][OK]	Enters integrating variable
[0][OK]	Enters 0 as the lower limit
[↰][π][÷][2][OK]	Enters π /2 as the upper limit
[CHOOS][▼][OK]	Selects *numeric* result
[OK]	Shows numeric value of integral

The display now shows: 0.141592….(i.e., σ^2 = 0.141592…).

See section 7 in Part II of this guide for instructions on programs that calculate integrals for continuous probability distributions. Part II also has sub-directories that allow you to plot continuous pdf's and to generate synthetic data that follow a given continuous probability distribution. The following material describes some of those probability distributions.

 © 2000 – Gilberto E. Urroz – All rights reserved

9.2. Some continuous probability distributions

In this section we describe several continuous probability distributions including the gamma, exponential, beta, and Weibull distributions. These distributions are described in any statistics textbook. Some of these distributions make use of a mathematical function known as the Gamma function and defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

The gamma function has the property that ,

$$\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1), \text{ for } \alpha > 1,$$

therefore, it can be related to the factorial of a number, i.e.,

$$\Gamma(\alpha) = (\alpha-1)!,$$

when α is a positive integer.

9.2.1. The gamma distribution

The probability distribution function (pdf) for the gamma distribution is given by

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \cdot x^{\alpha-1} \cdot \exp(-\frac{x}{\beta}), for \quad x > 0, \alpha > 0, \beta > 0;$$

The corresponding (cumulative) distribution function (cdf) would be given by an integral that has no closed-form solution.

9.2.2.The exponential distribution

The exponential distribution is the gamma distribution with a = 1. Its pdf is given by

$$f(x) = \frac{1}{\beta} \cdot \exp(-\frac{x}{\beta}), for \quad x > 0, \beta > 0;$$

While its cdf is given by

$$F(x) = 1 - \exp(-x/\beta), \text{ for } x>0,\ \beta>0.$$

9.2.3. The beta distribution

The pdf for the gamma distribution is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}, for \quad 0 < x < 1, \alpha > 0, \beta > 0;$$

As in the case of the gamma distribution, the corresponding cdf for the beta distribution is also given by an integral with no closed-form solution.

 © 2000 – Gilberto E. Urroz – All rights reserved

9.2.4. The Weibull distribution

The pdf for the Weibull distribution is given by

$$f(x) = \alpha \cdot \beta \cdot x^{\beta-1} \cdot \exp(-\alpha \cdot x^{\beta}), \quad for\ x > 0, \alpha > 0, \beta > 0$$

While the corresponding cdf is given by

$$F(x) = 1 - \exp(-\alpha \cdot x^{\beta}), \quad for\ x > 0, \alpha > 0, \beta > 0$$

To define a collection of functions corresponding to the gamma, exponential, beta, and Weibull distributions, first create a sub-directory called CFUN (Continuous FUNctions) and define the following functions:

Gamma function:	`'fgamma(x) = (x-1)!'`
Gamma pdf:	`'gapd(x) = x^(α-1)*EXP(-x/β)/(β^α*fgamma(α))'`
Gamma cdf:	`'gadf(x) = ∫(0,x,gapd(t),t)'`
Beta pdf:	`'βpd(x) = 'fgamma(α+β)*x^(α-1)*(1-x)^(β-1)/(fgamma(α)*fgamma(β))'`
Beta cdf:	`'βdf(x) = ∫(0,x, βpd(t),t)'`
Exponential pdf:	`'expd(x) = EXP(-x/β)/β'`
Exponential cdf:	`'exdf(x) = 1 - EXP(-x/β)'`
Weibull pdf:	`'Wepd(x) = α*β*x^(β-1)*EXP(-α*x^β)'`
Weibull cdf:	`'Wedf(x) = 1 - EXP(-α*x^β)'`

Remember to press [↰][DEF] after entering each function definition in level 1. You also need to create a couple of variables, α and β, and load them with some values, say α = 1.0, β = 2.0. To do this use the following keystrokes:

[1]['][α][↱] [A][STO] [2]['][α][↱] [B][STO]

Finally, for the *cdf* for Gamma and Beta *cdfs*, you need to edit the program definitions to add →NUM to the programs produced by [DEF]. For example, for the Gamma cdf, use the following keystroke sequence:

[↱][GADF][▼][▼][▼][↰][→NUM][ENTER].

The program should look like this: `<< → x '∫ (0,x,gapd(t),t)' →NUM >>`

Store the new program into *gadf*, using:

[↰][GADF].

Repeat the procedure for *βdf*. If you want to order the variables in your directory, use the procedure shown in section 7.2, to get your white buttons to look like this:

[α][β][*FGAM*][*GAPD*][*GADF*][*βPD*][*βDF*][*EXPD*][*EXDF*][*WEPD*][*WEPF*]

 © 2000 - Gilberto E. Urroz - All rights reserved

9.2.5. Examples:
Unlike the discrete functions defined in section 7.2, the continuous functions defined in this section do not include their parameters (α and/or β) in their definitions, therefore, you don't need to enter them in the display to calculate the functions. However, those parameters must be previously defined by entering the corresponding values in the variables α and β. To enter these values use a procedure similar to that suggested above when the variables α and β were first defined.

Practice the following exercises:

1. For α = 3, β = 2, use the Gamma distribution to obtain f(12), and P(X>12).

First, store α and β:

[3][↰] [*α*] [2][↰][*β*].

Then, for f(12), enter

[1][2][*GAPD*],

thus f(12) = 0.02230. Also, P(X>12) = 1 - P(X<12) = 1 - F(12), enter

[1][ENTER][1][2][*GADF*][-],

thus P(X>12) = 0.05196. (The evaluation of the integral for GADF takes about twenty seconds)

2. For α = 2, β = 9, use the Beta distribution to obtain f(0.10) and P(X<0.10).

First, store α and β:

[2][↰] [*α*] [9][↰][*β*].

Then, for f(0.10), enter

[NXT][.][1][*βPD*],

thus f(0.10) = 3.87420. Also, P(X<0.10) = F(0.10), enter

[.][1][βDF],

thus P(X>12) = 0.2639010. (The evaluation of the integral for GADF takes from 5 to 10 seconds).

For α = 0.2, β = 0.333, use the Weibull distribution to obtain f(5) and P(X<5).
First, store α and β:

[VAR][.][2][↰] [α] [.][3][3][3][↰][β].

Then, for f(5), enter

[NXT][5][WEPD],

thus f(5) = 0.0161738. Also, P(X<5) = F(5), enter

[5][WEDF],

 © 2000 - Gilberto E. Urroz - All rights reserved

thus P(X<5) = 0.289518.

3. For β = 0.2, use the Exponential distribution to obtain f(2.5) and P(X>2.5).
First, store only β=0.2:

[VAR] [2][⬑][β].

Then, for f(2.5), enter

[NXT][2][.][5][EXPD],

thus f(2.5) = 1.8633×10^{-5}. Also, P(X>2.5) = 1 - P(X<2.5) = 1 - F(2.5), enter

[1][ENTER][2][.][5][EXDF][-],

thus P(X>12) = 0.00000372.

See section 8 of Part II for an expanded version of this sub-directory including instructions to plot the probability distribution functions.

9.3. Multiple integrals for multivariate probability distributions

Multiple integrals may result from calculations of probabilities for multivariate distributions of the form *f(x,y)* or *f(x,y,z)*. Multivariate distributions using polar coordinates are also possible.

1) For example, calculate the value of α for the multivariate probability distribution defined on the domain:

$$A: \{0 < r < a,\ 0 < \theta < 2\pi\},$$

and whose probability density function (pdf) is given by

$$f(r,\theta) = \alpha r^3.$$

In polar coordinates a differential of area is given by

$$dA = r\ dr\ d\theta,$$

The pdf must satisfy the condition that :

$$\iint_A f(r,\theta)dA = \int_0^{2\pi}\int_0^a \alpha\cdot r^4 dr\cdot d\theta = 1$$

We will use this problem to illustrate again the use of *the EQUATION editor* in the HP48G series calculator. Use the following keystrokes to build the double integral:

[⬑][EQUATION] [⮣][∫][0][▶][2][×][⬑][π][▶][⮣][∫][0] [▶] [α][⬑][A]
[▶] [α][⮣][A] [×] [α][⬑] [R] [y^x][4] [▶][▶] [α][⬑] [R] [▶][▶][α][⮣][F][ENTER]

The integral is shown in level 1 of the display as:

```
' ∫(0, 2*π, ∫(0,a,α*r^4,r),θ)'
```

 © 2000 - Gilberto E. Urroz - All rights reserved

To evaluate press:

[EVAL] [EVAL] [EVAL] [↰][SYMBOLIC][COLCT]

The result is `'0.4*a^5*α*π'`. Since the integral must be equal to 1, we write `'1=0.4*a^5*α*π'`, by pressing:

[↰][EDIT][▶][1][↰][=][ENTER]

Next, enter the following:

['] [α][↱][A] [ENTER]
[↰][SYMBOLIC] [ISOL]

to isolate α. The result is:

`'α = 1/π/(.4*a^5)'`

If you press [COLCT] the resulting expression is:

`'α = 2.5*a^-5/π'`

2) Try the following example: determine the constant K for a pdf defined in the Cartesian domain

A: {0 < x < 5, 0 < y < 1+x }

with

f(x,y) = K· x· (1-x·y)

In Cartesian coordinates a differential of area is given by

dA = dx dy.

Therefore, the multivariate pdf must satisfy:

$$\iint_A f(x,y)dA = K\cdot\int_0^5\int_0^{1+x} K\cdot x\cdot(1-x\cdot y)\,dx\cdot dy = 1$$

We will use this problem to illustrate the use of the EQUATION editor in the HP48G series calculator. Use the following keystrokes to build the double integral:

[↰][EQUATION] [↱][∫][0][▶][5][▶][↱][∫][0] [▶] [1][+] [α][↰][X]
[▶] [α][↰][X] [×] [↰][( )][1] [+][α][↰][X] [×] [α][↰][Y][▶][▶] [α][↰] [Y]
[▶][▶][α][↰] [X] [↰] [ENTER]

The integral is shown in level 1 of the display as:

`'∫(0, 5, ∫(0,1+x, x*(1+x*y),y),x)'`

 © 2000 - Gilberto E. Urroz - All rights reserved

Press [ENTER] to keep a second copy of the integral available in the stack.

Press [EVAL] three or four times. Because there is no closed-form expression for the integral, the calculator will stop the symbolic evaluation at a certain point. The only possibility now is a numerical integration, for which you need to use: [↰][→NUM]. The result, in this case, is 543.75. Press [1/x] to get the value of K = 0.00183. Press [⇦] to clear level 1 of the stack.

{**Note**: it will take your calculator about 2 minutes to produce a result.] To check the integration error, press [IERR]. For this case IERR = 5.43×10^{-9}, i.e., almost zero.}

As you see from this example, numerical integration of a double integral takes the calculator quite some time to finish. The more complicated the integrand or the integration limits, the longer it takes. For example, if we were to simplify the pdf used in this example to

$$f(x,y) = K\cdot x\cdot (1-y),$$

the integration will take much less time. Enter the integral in your calculator and time it:

```
'∫(0, 5, ∫(0,1+x, x*(1+y),y),x)'
```

I timed it at 15 seconds. The result is 180.2083.

 © 2000 - Gilberto E. Urroz - All rights reserved

10.0 The Normal distribution

The HP48G/GX calculator provides a number of functions to calculate probabilities using the Normal, Chi-squared (χ^2), Student-t, and F-distributions. These distributions are described in this and the following sections. In the HP48G/GX, tables of the Normal pdf, and what is known as the upper-tail distribution functions for the Normal, Chi-squared (χ^2), Student-t, and F-distributions are accessed through the keystroke sequence:

[MTH][NXT][PROB][NXT].

Upper-distribution functions provide the value of P(X>x) given x.

10.1 Normal distribution pdf:

The expression for the normal distribution pdf is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}),$$

where μ is the mean, and σ^2 the variance of the distribution.
To calculate the value of $f(\mu,\sigma^2,x)$ for the normal distribution, enter the following values: the mean, μ, in level 3; the variance, σ^2, in level 2; and, the value x in level 1, then enter

[MTH][NXT][PROB][NXT][NDIST].

For example, check that for a normal distribution, f(1.0,0.5,2.0) = 0.20755374. Use the following sequence:

[1][ENTER][.][5][ENTER][2][ENTER][MTH][NXT][*PROB*][NXT][*NDIST*]

10.2 Normal distribution cdf:

The HP48G series calculator has a function UTPN that calculates the upper-tail normal distribution, i.e.,

UTPN(x) = P(X>x) = 1 - P(X<x).

To obtain the value of the upper-tail normal distribution UTPN we need to enter the following values: the mean, μ, in level 3; the variance, σ^2, in level 2; and, the value x in level 1, then enter

[MTH][NXT][*PROB*][NXT][*UTPN*].

For example, check that for a normal distribution, with μ = 1.0, σ^2 = 0.5, UTPN(0.75) = 0.638163. Use the following sequence:

[1]ENTER][.][5][ENTER][.][7][5][ENTER][MTH][NXT][*PROB*][NXT][*UTPN*]

Different probability calculations for normal distributions [X is $N(\mu,\sigma^2)$] can be defined using *the function* UTPN, as follows:

 © 2000 - Gilberto E. Urroz - All rights reserved

$P(X<a) = 1 - UTPN(\mu, \sigma^2, a)$

$$P(a<X<b) = P(X<b) - P(X<a) = 1 - UTPN(\mu, \sigma^2, b) - (1 - UTPN(\mu, \sigma^2, a))$$
$$= UTPN(\mu, \sigma^2, a) - UTPN(\mu, \sigma^2, b)$$

$P(X>c) = UTPN(\mu, \sigma^2, c)$

Example: Using μ = 1.5, and σ^2 = 0.5, find (a) P(X<1.0); (b) P(X>2.0); (c) P(1.0<X<2.0).

P(X<1.0) = 1 - P(X>1.0) = 1 - UTPN(1.5, 0.5, 1.0). Enter:

[1][ENTER][1][.][5][SPC][.][5][SPC][1][MTH][NXT][*PROB*][NXT][*UTPN*][-].

The result is P(X<1.0) = 0.239750.

P(X>2.0) = UTPN(1.5, 0.5, 2.0). Enter:

[1][.][5][SPC][.][5][SPC][2] ([MTH][NXT][*PROB*][NXT]) [*UTPN*].

The result is P(X<2.0) = 0.239750.

P(1.0<X<2.0) = F(1.0) - F(2.0) = UTPN(1.5,0.5,1.0) - UTPN(1.5,0.5,2.0). Enter

[1][.][5][SPC][.][5][SPC][1][UTPN] [1][.][5][SPC][.][5][SPC][2][*UTPN*] [-]

The result is P(1.0<X<2.0) = 0.7602499 - 0.2397500 = 0.524998.

 © 2000 - Gilberto E. Urroz - All rights reserved

11.0 The Student-t, Chi-squared (χ^2), and F-distributions

11.1 The Student-t distribution

The Student-t, or simply, the t, distribution has one parameter ν, known as the degrees of freedom. The probability distribution function (pdf) is given by

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \cdot \sqrt{\pi\nu}} \cdot (1+\frac{t^2}{\nu})^{-\frac{\nu+1}{2}}, -\infty < t < \infty$$

where $\Gamma(\alpha) = (\alpha-1)!$ is the gamma function defined above.

The HP48G/GX provides for values of the upper-tail (cumulative) distribution function for the t-distribution using [UTPT] given the value of t and the parameter ν. The definition of this function is, therefore,

$$UTPT(\nu,t) = \int_t^\infty f(t)dt = 1 - \int_{-\infty}^t f(t)dt = 1 - P(T \le t)$$

To use this function, enter ν in level 2 and t in level 1, then press [UTPT]. Recall that to get to the probability functions you need to use the keystroke sequence:

[MTH][NXT][*PROB*][NXT].

For example, to calculate UTPT(5, 2.5), use the following:

[5][ENTER][2][.][5][ENTER][*UTPT*]

The result is: UTPT(5,2.5) = 2.7245...E-2
Alternatively, you can use:

[5][SPC][2][.][5][*UTPT*].

Different probability calculations for the t-distribution can be defined using the function UTPT, as follows:

$P(T<a) = 1 - UTPT(\nu,a)$

$P(a<T<b) = P(T<b) - P(T<a) = 1 - UTPT(\nu,b) - (1 - UTPT(\nu,a)) = UTPT(\nu,a) - UTPT(\nu,b)$

$P(T>c) = UTPT(\nu,c)$

 © 2000 - Gilberto E. Urroz - All rights reserved

11.2 The Chi-squared (χ2) distribution

The Chi-squared (χ2) distribution has one parameter ν, known as the degrees of freedom. The probability distribution function (pdf) is given by

$$f(x) = \frac{1}{2^{\frac{\nu}{2}} \cdot \Gamma(\frac{\nu}{2})} \cdot x^{\frac{\nu}{2}-1} \cdot e^{-\frac{x}{2}}, \nu > 0, x > 0$$

The HP48G/GX provides for values of the upper-tail (cumulative) distribution function for the χ^2-distribution using [UTPC] given the value of t and the parameter ν. The definition of this function is, therefore,

$$UTPC(\nu, x) = \int_t^{\infty} f(x)dx = 1 - \int_{-\infty}^{t} f(x)dx = 1 - P(X \leq x)$$

To use this function, enter ν in level 2 and x in level 1, then press [*UTPC*]. For example, to calculate UTPC(5, 2.5), use the following:

[5][ENTER][2][.][5][ENTER][*UTPC*]

The result is: UTPC(5,2.5) = 0.776495...
Alternatively, you can use:

[5][SPC][2][.][5][*UTPC*] .

Different probability calculations for the Chi-squared distribution can be defined using the function UTPC, as follows:

P(X<a) = 1 - UTPC(ν,a)

P(a<X<b) = P(X<b) - P(X<a) = 1 - UTPC(ν,b) - (1 - UTPC(ν,a)) = UTPC(ν,a) - UTPC(ν,b)

P(X>c) = UTPC(ν,c)

11.3 The F distribution

The F distribution has two parameters νN = numerator degrees of freedom, and νD = denominator degrees of freedom. The probability distribution function (pdf) is given by

$$f(x) = \frac{\Gamma(\frac{\nu N + \nu D}{2}) \cdot (\frac{\nu N}{\nu D})^{\frac{\nu N}{2}} \cdot F^{\frac{\nu N}{2}-1}}{\Gamma(\frac{\nu N}{2}) \cdot \Gamma(\frac{\nu D}{2}) \cdot (1 - \frac{\nu N \cdot F}{\nu D})^{(\frac{\nu N + \nu D}{2})}}$$

The HP48G/GX provides for values of the upper-tail (cumulative) distribution function for the F distribution using [UTPF] given the value of F and the parameters νN and νD.

 © 2000 - Gilberto E. Urroz - All rights reserved

The definition of this function is, therefore,

$$UTPF(\nu N,\nu D,F)=\int_t^{\infty} f(F)dF = 1-\int_{-\infty}^{t} f(F)dF = 1-P(\Im \le F)$$

To use this function, enter νN in level 3, νD in leve2, and F in level 1, then press [UTPF]. For example, to calculate UTPF(10,5, 2.5), use the following:

[1][0][ENTER][5][ENTER][2][.][5][ENTER][UTPF]

The result is: UTPF(5,2.5) = 0.776495…

Alternatively, you can use:

[1][0][SPC][5][SPC][2][.][5][UTPF].

Different probability calculations for the F distribution can be defined using the function UTPF, as follows:

P(F<a) = 1 - UTPF(νN, νD,a)

P(a<F<b) = P(F<b) - P(F<a) = 1 -UTPF(νN, νD,b)- (1 - UTPF(νN, νD,a))
= UTPF(νN, νD,a) - UTPF(νN, νD,b)

P(F>c) = UTPF(νN, νD,a)

See section 11 in Part II for a sub-directory with programmed features that allow you to calculate upper-tail probabilities for the Normal, t, Chi-squared, and F distributions. This sub-directory also allows you to find the inverse function, e.g., find t given ν and UTPF(ν,t). Plots of the functions can be obtained by using the programs described in section 10 of Part II of this guide. Below we present some applications of the functions UTPN, UTPT, UTPC, and UTPF in hypothesis testing.

© 2000 – Gilberto E. Urroz – All rights reserved

12.0 Hypothesis testing using P-values

Herein we discuss hypothesis testing on means and variances using a parameter called the *P-value*.

12.1. Inferences concerning one mean

12.1.1. Two-sided hypothesis

The problem consists in testing the null hypothesis H_o: $\mu = \mu_o$, against the alternative hypothesis, H_1: $\mu \neq \mu_o$ at a level of confidence $(1-\alpha)100\%$, or significance level α, using a sample of size n with a mean $\bar{x}$ and a standard deviation s. This tests is referred to as a *two-sided* or *two-tailed* test. The procedure for the test is as follows:

First, we calculate the appropriate statistic for the test (t_o or z_o) as follows:

- If n < 30 and the standard deviation of the population, σ, is known, use

$$z_o = \frac{\bar{x} - \mu_o}{\sigma / \sqrt{n}}$$

- If n > 30, and σ is known, use z_o as above. If σ is not known, replace s for σ in z_o, i.e., use

$$z_o = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}$$

- If n < 30, and s is unknown, use the t-statistic

$$t_o = \frac{\bar{x} - \mu_o}{s / \sqrt{n}}$$

with ν = n - 1 degrees of freedom.

Then, calculate the P-value associated with either z_o or t_o, and compare it to α to decide whether or not to reject the null hypothesis. The P-value for a two-sided test is defined as either

$$\text{P-value} = P(|z| > |z_o|),$$

or,

$$\text{P-value} = P(|t| > |t_o|).$$

The criteria to use for hypothesis testing is:

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

 © 2000 - Gilberto E. Urroz - All rights reserved

The P-value for a two-sided test can be calculated using the probability functions in the HP48G/GX as follows:

- If using z, P-value = 2·UTPN(0,1,$|z_o|$)
- If using t, P-value = 2·UTPT(ν,$|t_o|$)

For example, test the null hypothesis H_o: μ = 22.5 (= μ_o), against the alternative hypothesis, H_1: $\mu \neq 22.5$, at a level of confidence of 95% i.e., α = 0.05, using a sample of size n = 25 with a mean $\bar{x}$ = 22.0 and a standard deviation s = 3.5. We assume that we don't know the value of the population standard deviation, therefore, we calculate a t statistic as follows:

$$t_o = \frac{\bar{x} - \mu_o}{s/\sqrt{n}} = \frac{22.0 - 22.5}{3.5/\sqrt{25}} = -0.7142$$

The corresponding P-value, for n = 25 - 1 = 24 degrees of freedom is

P-value = 2·UTPT(24,-0.7142) = 2·0.7590 = 1.5169,

since 1.5169 > 0.05, i.e., P-value > α, we cannot reject the null hypothesis H_o: μ = 22.0.

12.1.2 One-sided hypothesis

The problem consists in testing the null hypothesis H_o: μ = μ_o, against the alternative hypothesis, H_1: $\mu > \mu_o$ or H_1: $\mu < \mu_o$ at a level of confidence (1-α)100%, or significance level α, using a sample of size n with a mean $\bar{x}$ and a standard deviation s. This tests is referred to as a *one-sided* or *one-tailed* test. The procedure for performing a one-side test starts as in the two-tailed test by calculating the appropriate statistic for the test (t_o or z_o) as indicated above. Next, we use the P-value associated with either z_o or t_o, and compare it to α to decide whether or not to reject the null hypothesis. The P-value for a two-sided test is defined as either

P-value = P(z > $|z_o|$),

or,

P-value = P(t > $|t_o|$).

The criteria to use for hypothesis testing is:

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

Notice that the criteria are exactly the same as in the two-sided test. The main difference is the way that the P-value is calculated. The P-value for a one-sided test can be calculated using the probability functions in the HP48G/GX as follows:

- If using z, P-value = UTPN(0,1,z_o)
- If using t, P-value = UTPT(ν,t_o)

For example, test the null hypothesis H_o: μ = 22.0 (= μ_o), against the alternative hypothesis, H_1: μ >22.5 at a level of confidence of 95% i.e., α = 0.05, using a sample of size n = 25 with a mean $\bar{x}$ = 22.0 and a standard deviation s = 3.5. Again, we assume that we don't know the

 © 2000 - Gilberto E. Urroz - All rights reserved

value of the population standard deviation, therefore, the value of the t statistic is the same as in the two-sided test case shown above, i.e., t_o = -0.7142, and P-value, for ν = 25 - 1 = 24 degrees of freedom is

P-value = UTPT(24, |-0.7142|) = UTPT(24,0.7124) = 0.2409,

since 0.2409 > 0.05, i.e., P-value > α, we cannot reject the null hypothesis H_o: μ = 22.0.

12.2 Inferences concerning two means

The null hypothesis to be tested is H_o: μ_1-μ_2 = δ, at a level of confidence (1-α)100%, or significance level α, using two samples of sizes, n_1 and n_2, mean values $\bar{x}_1$ and $\bar{x}_2$, and standard deviations s_1 and s_2. If the populations standard deviations corresponding to the samples, σ_1 and σ_2, are known, or if n_1 > 30 and n_2 > 30 (large samples), the test statistic to be

$$z_o = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

used is

If n_1 < 30 or n_2 < 30 (at least one small sample), use the following test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}\sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

12.2.1. Two-sided hypothesis

If the alternative hypothesis is a two-sided hypothesis, i.e., H_1: μ_1-μ_2 ≠ δ, The P-value for this test is calculated as

- If using z, P-value = 2·UTPN(0,1, $|z_o|$)
- If using t, P-value = 2·UTPT(ν, $|t_o|$)

with the degrees of freedom for the t-distribution given by ν = n_1 + n_2 - 2.
The test criteria are

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

12.2.2. One-sided hypothesis

If the alternative hypothesis is a two-sided hypothesis, i.e., H_1: μ_1-μ_2 < δ, or, H_1: μ_1-μ_2 < δ,, the P-value for this test is calculated as:

- If using z, P-value = UTPN(0,1, $|z_o|$)
- If using t, P-value = UTPT(ν, $|t_o|$)

 © 2000 - Gilberto E. Urroz - All rights reserved

The criteria to use for hypothesis testing is:

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

12.2.3. Paired sample tests

When we deal with two samples of size n with paired data points, instead of testing the null hypothesis, H_o: $\mu_1-\mu_2 = \delta$, using the mean values and standard deviations of the two samples, we need to treat the problem as a single sample of the differences of the paired values. In other words, generate a new random variable $X = X_1-X_2$, and test H_o: $\mu = \delta$, where μ represents the mean of the population for X. Therefore, you will need to obtain $\bar{x}$ and s for the sample of values of x. The test should then proceed as a one-sample test using the methods described in section 12.1.

12.3. Inferences concerning one variance

The null hypothesis to be tested is , H_o: $\sigma^2 = \sigma_o^2$, at a level of confidence $(1-\alpha)100\%$, or significance level α, using a sample of size n, and variance s^2. The test statistic to be used is a chi-squared test statistic defined as

$$\chi_o^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Depending on the alternative hypothesis chosen, the P-value is calculated as follows:

- H_1: $\sigma^2 < \sigma_o^2$, P-value = $P(\chi^2<\chi_o^2)$ = 1-UTPC(ν,χ_o^2)
- H_1: $\sigma^2 > \sigma_o^2$, P-value = $P(\chi^2>\chi_o^2)$ = UTPC(ν,χ_o^2)
- H_1: $\sigma^2 \neq \sigma_o^2$, P-value =2·min[$P(\chi^2<\chi_o^2)$, $P(\chi^2>\chi_o^2)$] = 2·min[1-UTPC(ν,χ_o^2), UTPC(ν,χ_o^2)]

where the function min[x,y] produces the minimum value of x or y (similarly, max[x,y] produces the maximum value of x or y). UTPC(ν,x) represents the HP48G/GX upper-tail probabilities for ν = n - 1 degrees of freedom.
The test criteria are the same as in hypothesis testing of means, namely,

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

Please notice that this procedure is valid only if the population from which the sample was taken is a Normal population. In order to check for normality of data, you can use the procedure programmed in sub-directory CHKN sub-directory described in section 12 of Part II of this guide.

As an example, consider the case in which σ_o^2 = 25, α=0.05, n = 25, and s^2 = 20, and the sample was drawn from a normal population. To test the hypothesis, H_o: $\sigma^2 = \sigma_o^2$, against H_1: $\sigma^2 < \sigma_o^2$, we first calculate

$$\chi_o^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(25-1)\cdot 20}{25} = 189.2$$

 © 2000 - Gilberto E. Urroz - All rights reserved

With $\nu = n - 1 = 25 - 1 = 24$ degrees of freedom, we calculate the P-value as,

$$\text{P-value} = P(\chi^2 < 19.2) = 1\text{-UTPC}(24,19.2) = 0.2587...$$

Since, 0.2587… > 0.05, i.e., P-value > α, we cannot reject the null hypothesis, H_o: $\sigma^2 = 25(= \sigma_o^2)$.

12.4. Inferences concerning two variances

The null hypothesis to be tested is , H_o: $\sigma_1^2 = \sigma_2^2$, at a level of confidence $(1-\alpha)100\%$, or significance level α, using two samples of sizes, n_1 and n_2, and variances s_1^2 and s_2^2. The test statistic to be used is an F test statistic defined as

$$F_o = \frac{s_N^2}{s_D^2}$$

where s_N^2 and s_D^2 represent the numerator and denominator of the F statistic, respectively. Selection of the numerator and denominator depends on the alternative hypothesis being tested, as shown below. The corresponding F distribution has degrees of freedom, $\nu_N = n_N-1$, and $\nu_D = n_D-1$, where n_N and n_D, are the sample sizes corresponding to the variances s_N^2 and s_D^2, respectively. The following table shows how to select the numerator and denominator for F_o depending on the alternative hypothesis chosen:

Alternative hypothesis	Test statistic	Degrees of freedom
H_1: $\sigma_1^2 < \sigma_2^2$ (one-sided)	$F_o = s_2^2/s_1^2$	$\nu_N = n_2-1$, $\nu_D = n_1-1$
H_1: $\sigma_1^2 > \sigma_2^2$ (one-sided)	$F_o = s_1^2/s_2^2$	$\nu_N = n_1-1$, $\nu_D = n_2-1$
H_1: $\sigma_1^2 \neq \sigma_2^2$ (two-sided)	$F_o = s_M^2/s_m^2$ $s_M^2=\max(s_1^2,s_2^2)$, $s_m^2=\min(s_1^2,s_2^2)$	$\nu_N = n_M-1, \nu_D = n_m-1$(*)

(*) n_M is the value of n corresponding to the s_M, and n_m is the value of n corresponding to s_m.

The P-value is calculated, in all cases, as: P-value = $P(F>F_o) = \text{UTPF}(\nu_N, \nu_D, F_o)$
The test criteria are:

- Reject H_o if P-value < α
- Do not reject H_o if P-value > α.

As an example, consider two samples drawn from Normal populations such that $n_1 = 21$, $n_2 = 31$, $s_1^2 = 0.36$, and $s_2^2 = 0.25$. We test the null hypothesis, H_o: $\sigma_1^2 = \sigma_2^2$, at a significance level $\alpha = 0.05$, against the alternative hypothesis, H_1: $\sigma_1^2 \neq \sigma_2^2$. For a two-sided hypothesis, we need to identify s_M and s_m, as follows:

$$s_M^2=\max(s_1^2,s_2^2) = \max(0.36,0.25) = 0.36 = s_1^2$$

$$s_m^2=\min(s_1^2,s_2^2) = \max(0.36,0.25) = 0.25 = s_2^2$$

 © 2000 - Gilberto E. Urroz - All rights reserved

Also,

$$n_M = n_1 = 21,$$

$$n_m = n_2 = 31,$$

$$\nu_N = n_M - 1 = 21-1=20,$$

$$\nu_D = n_m - 1 = 31-1 =30.$$

Therefore, the F test statistics is

$F_o = s_M^2/s_m^2=0.36/0.25=1.44$

The P-value is

P-value = $P(F>F_o) = P(F>1.44) = \text{UTPF}(\nu_N, \nu_D,F_o) = \text{UTPF}(20,30,1.44) = 0.1788\ldots$

Since $0.1788\ldots > 0.05$, i.e., P-value > α, therefore, we cannot reject the null hypothesis that H_o: $\sigma_1^2 = \sigma_2^2$.

<u>Note</u>: Section 14 in Part II of this guide presents instructions for programs that perform P-value-based hypotheses testing on the mean and variance.

 © 2000 – Gilberto E. Urroz – All rights reserved

13.0 Linear fitting using HP48G/GX 'Fit data' feature.

13.1. The Method of Least Squares

Let x = independent, non-random variable, and Y = dependent, random variable. The **regression curve** of Y on x is defined *as the relationship between x and the mean of the corresponding distribution of the Y's*.

Assume that the regression curve of Y on x is linear, i.e., mean distribution of Y's is given by

$$\alpha + \beta x.$$

Y differs from the mean ($\alpha + \beta x$) by a value ε, thus

$Y = \alpha + \beta x + \varepsilon$,

where ε is a random variable.

To visually check whether the data follows a linear trend, draw a **scattergram** or **scatter plot**

Suppose that we have *n* paired observations (x_i, y_i); we predict y by means of

$$\hat{y} = a + b x,$$

where a and b are constant.

Define the *prediction error* as,

$$e_i = y_i - \hat{y}_i = y_i - (a + b x_i).$$

The method of least squares requires us to choose a, b so as to minimize the *sum of squared errors* (SSE)

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

From the conditions

$$\frac{\partial}{\partial a}(SSE) = 0 \qquad \frac{\partial}{\partial b}(SSE) = 0$$

We get the, so-called, *normal equations*:

$$\sum_{i=1}^{n} y_i = a \cdot n + b \cdot \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i \cdot y_i = a \cdot \sum_{i=1}^{n} x_i + b \cdot \sum_{i=1}^{n} x_i^2$$

This is a system of linear equations with *a* and *b* as the unknowns.

 © 2000 - Gilberto E. Urroz - All rights reserved

13.1.1 HP48G application:

Create a directory called CFIT (Curve FITting) within directory STATS. Move into that directory and store the data from any example in your textbook. (Column 1 represents x, and column 2, y.) Then copy it into ΣDAT (Assume that the table containing the data is stored in variable P334):

[P334]['][↱][CHARS][▶][▶][▶][▶][▶] [ECHO][ENTER][α][α][D][A][T][ENTER][STO]

First, plot the data to see if it follows a linear trend:

[↰][STAT][PLOT][SCATR] (shows good linear trend) [CANCL].

Get summary statistics:

[↰][STAT][SUMS][ΣX][ΣY][ΣX^2][ΣY^2][ΣX*Y][NΣ].

Results in: $\Sigma x = 2000$; $\Sigma y = 8.35$; $\Sigma x^2 = 532000$; $\Sigma y^2 = 9.1097$; $\Sigma xy = 2175.4$, $n = 10$.

The normal equations are now:

$$\begin{aligned} 10\,a + 2000\,b &= 8.35 \\ 2000\,a + 532000\,b &= 2175.40 \end{aligned}$$

or, in matrix notation,

$$\begin{bmatrix} 10 & 2000 \\ 2000 & 532000 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 8.35 \\ 2175.4 \end{bmatrix}$$

Use the following commands to solve for the vector $[a \;\; b]^T$:

[↱][SOLVE][▲][▲][OK]
[↱][MATRIX]
[1][0][SPC][2][0][0][0][ENTER][▼]
[2][0][0][0][SPC][5][3][2][0][0][0][ENTER][ENTER]
[▼][↰][[ ]][8][.][3][5][SPC][2][1][7][5][.][4][OK]
[SOLVE]

Result is X: [6.92424242424E-2...

[EDIT]

Bottom display shows 6.92424E-2.

[▶]

Bottom display shows 3.8287E-3.

[ENTER][ON]

The results obtained are then a = 0.06924 and b = 0.00383.

13.1.2. Using the "Fit data" feature in the HP48G/GX.

The HP48G calculator has its own feature for determining the least-square linear fitting to a set of data points (x,y). It uses the ΣDAT matrix where x and y are stored in columns. To access this feature, use the following keystrokes:

[↱][STAT][▼][▼][OK]

 © 2000 - Gilberto E. Urroz - All rights reserved

The display shows the current ΣDAT, already loaded. Change your set up screen to the following parameters if needed:

```
X-COL: 1   Y-COL: 2
MODEL: Linear Fit
```

Then, press [OK], to get the following results:

```
1: '6.924242…E-2 + 3.828787…E-3*X'
2: Correlation: 0.9514813
3: Covariance: 56.1555
```

We can write the following equation: $\hat{y} = 0.06924 + 0.00383\ x$.

Notes:

- a,b are unbiased estimators of α, β.
- The Gauss-Markov theorem indicates that among all unbiased estimators for α and β, the least-square estimators (a,b) are the most efficient.

13.2 Covariance and Correlation

In general, the covariance of two random variables X,Y (i.e., populations) is defined as

$$Cov(X,Y) = E[(X-\mu_x)\ (Y-\mu_y)]$$

For a sample of data points (x,y), we define *covariance* as

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

The *sample correlation coefficient* for x,y is defined as

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y}$$

Where s_x, s_y are the standard deviations of x and y, respectively, i.e.

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

The values s_{xy} and r_{xy} are the "Covariance" and "Correlation," respectively, obtained by using the "Fit data" feature of the HP48G calculator.

© 2000 - Gilberto E. Urroz - All rights reserved

13.3. Additional equations and definitions

Let's define the following quantities:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)\cdot s_x^2 = \sum_{i=1}^{n} x_i^{\ 2} - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$S_y = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)\cdot s_y^2 = \sum_{i=1}^{n} y_i^{\ 2} - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^2 = (n-1)\cdot s_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

From which it follows that the *standard deviations* of x and y, and the *covariance* of x,y are given, respectively, by

$$s_x = \sqrt{\frac{S_{xx}}{n-1}} \qquad s_y = \sqrt{\frac{S_{yy}}{n-1}} \qquad s_{xy} = \frac{S_{yx}}{n-1}$$

Also, the *sample correlation coefficient* is

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}\cdot S_{yy}}}.$$

In terms of $\bar{x}$, $\bar{y}$, S_{xx}, S_{yy}, and S_{xy}, the solution to the normal equations is:

$$a = \bar{y} - b\bar{x} \qquad b = \frac{S_{xy}}{S_{xx}} = \frac{s_{xy}}{s_x^2}$$

The regression curve of Y on x is defined as $Y = \alpha + \beta x + \varepsilon$. If we have a set of n data points (x_i, y_i), then we can write

$$Y_i = \alpha + \beta x_i + \varepsilon_i. \qquad (i = 1,2,\ldots,n)$$

Where Y_i = independent, normally distributed random variables with mean $(\alpha + \beta x_i)$ and the common variance σ^2; ε_i = independent, normally distributed random variables with mean zero and the common variance σ^2.

Let y_i = actual data value, $\hat{y}_i = a + bx_i$ = least-square prediction of the data. Then, the *prediction error* is:

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i).$$

 © 2000 - Gilberto E. Urroz - All rights reserved

An estimate of σ^2 is the, so-called, **standard error of the estimate**

$$s_e^2 = \frac{1}{n-2}\sum[y_i - (a + bx_i)]^2 = \frac{S_{yy} - (S_{xy})^2 / S_{xx}}{n-2} = \frac{n-1}{n-2}\cdot s_y^2 \cdot (1 - r_{xy}^2)$$

13.4 Confidence intervals and hypothesis testing in linear fitting

Please refer to your statistics textbook for the following subjects:

- Statistics for inferences about α and β.
- Confidence limits for regression coefficients.
- Hypothesis testing on β.
- Confidence limits for $(\alpha + \beta x_o)$.
- Limits of prediction.

Procedure for calculating confidence intervals and testing hypothesis on β for linear fitting of data using the HP48 G calculator:

1) Enter (x,y) data into ΣDAT.

2) [↱][PLOT] with SCATTER, COLS: 1 2, and use appropriate H- and V-VIEWS to check linear trend. Press [*CANCL*][ENTER] to return to normal display.

3) [↱][STAT][▼][▼][OK], to fit straight line, and get *a, b,* s_{xy} (Covariance), and r_{xy} (Correlation).

4) [↰][STAT][1VAR][MEAN][SDEV] to obtain $\bar{x}$, $\bar{y}$, s_x, s_y.

5) Calculate

$$S_{xx} = (n-1)\cdot s_x^2 \qquad s_e^2 = \frac{n-1}{n-2}\cdot s_y^2 \cdot (1 - r_{xy}^2)$$

6) For either confidence intervals or two-tailed tests, obtain $t_{\alpha/2}$, with (1- α)100% confidence, from t-distribution with $\nu = n - 2$. (For example, use subdirectory NTCF/TDIST (section 11, Part II)).

7) For one- or two-tailed tests, use the appropriate value of t:

$$t = \frac{(a-\alpha)}{s_e}\sqrt{\frac{nS_{xx}}{S_{xx} + n(\bar{x})^2}}, \quad t = \frac{(b-\beta)}{s_e}\sqrt{S_{xx}}.$$

8) Reject the null hypothesis if `P-value` < α.

9) For confidence intervals use the appropriate formulas from your textbook.

 © 2000 - Gilberto E. Urroz - All rights reserved

14.0 Curvilinear regression using the HP48G/GX *'Fit data'* feature.

14.1. Linearized relationships.

Many curvilinear relationships "straighten out" to a linear form. For example, the different models for data fitting provided by the HP48G calculator can be linearized as described below:

Type of Fitting	Actual Model	Linearized Model	Independent variable ξ	Dependent Variable η	Covariance $s_{\xi\eta}$
Linear	y = a + bx	y = a + bx [same]	x	y	s_{xy}
Logarithmic	y = a + b ln(x)	y = a + b ln(x) [same]	ln(x)	y	$s_{\ln(x),y}$
Exponential	$y = a\,e^{bx}$	ln(y) = ln(a) + bx	x	ln(y)	$s_{x,\ln(y)}$
Power	$y = a\,x^{b}$	ln(y) = ln(a) + b ln(x)	ln(x)	ln(y)	$s_{\ln(x),\ln(y)}$

The sample *covariance* of ξ,η is given by

$$s_{\xi\eta} = \frac{1}{n-1}\sum(\xi_i - \overline{\xi})(\eta_i - \overline{\eta})$$

Also, we define the sample *variances* of ξ and η, respectively, as

$$s_\xi^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi})^2 \qquad s_\eta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\eta_i - \overline{\eta})^2$$

The sample *correlation coefficient* $r_{\xi\eta}$ is

$$r_{\xi\eta} = \frac{s_{\xi\eta}}{s_\xi \cdot s_\eta}$$

The general form of the *regression equation* is $\eta = A + B\xi$.

14.2. Other "linearized" relationships

Another equation that can be re-cast in the linear form ($\eta = A + B\xi$) is the *reciprocal (linear) function*

$$y = 1/(\alpha + \beta x),$$

which can be re-written as:

$$1/y = \alpha + \beta x.$$

In general, equations of the form

$$y = \alpha + \beta\, f(x), \text{ or } g(y) = \alpha + \beta x,$$

can be analyzed using a linear relationship, by taking either

$$\xi = f(x), \text{ or } \eta = g(y),$$

in the linearized form

 © 2000 - Gilberto E. Urroz - All rights reserved

$$\eta = A + B\xi.$$

For example, for centrifugal pumps, the relationship between the discharge, Q, and the energy head H, is usually a quadratic equation of the form

$$H = \alpha + \beta Q^2,$$

where α and β are constant. Therefore, if we take

$$\xi = Q^2 \text{ and } \eta = H,$$

the relationship is re-cast as

$$\eta = a + b\xi.$$

To illustrate this particular case we solve the following exercise. The data shown below comes from tests performed in a centrifugal pump:

Q(gpm)	H(ft)
0	100
110	90
180	80
250	60
300	40
340	20

In order to determine the relationship H-Q you need to perform a linear fitting between $\xi = Q^2$ and $\eta = H$. Here is a suggested procedure using the HP48G calculator:

1) Enter the values of Q as a list and place two copies of the list in the display:

[↰][{}][0][SPC][1][1][0][SPC][1][8][0][SPC][2][5][0][SPC][3][0][0][SPC][3][4][0][ENTER][ENTER]

2) Squared the list in level 1 (i.e., generate $\xi = Q^2$): [↰][x^2]

3) Enter the values of H as a list (i.e., enter $\eta = H$):

[{}][1][0][0][SPC][9][0][SPC][8][0][SPC][6][0][SPC][4][0][SPC][2][0][ENTER]

4) Create a matrix of three columns by pressing: [3][CRMAT] (*)

5) Save the matrix into a variable, say PUMP: ['][α][α][P][U][M][P][ENTER][STO]

6) Activate the data fit feature of the HP48G calculator: [↱][STAT][▼][▼][OK].

7) Load PUMP into ΣDAT, by selecting the field in front of `ΣDAT:` in the set-up screen, and using [CHOOS].

8) Also, change the following parameters: `X-COL: 2` `Y-COL: 3` `MODEL: Linear Fit`

9) Then, press [OK], to get the following results:

```
3: '100.357358269+-6.78644112646E-4*X'
2: Correlation: -.998337705738
1: Covariance: -1395200
```

Line 3 in the display gives the fitted equation, namely, $H = 100.36 - 6.79\text{x}10^{-6}\ Q^2$, with H(ft), Q(gpm).

 © 2000 - Gilberto E. Urroz - All rights reserved

The value of the correlation coefficient is very close to -1, indicating a good fitting of the data.

(*) The program [CRMT] allows you to put together a *p*×*n* matrix (i.e., *p* rows, *n* columns) out of *n* lists of *p* elements each. To use this program, enter the *n* lists in the order that you want them as columns of the matrix, enter the value of *n*, and press [CRMT]. To create the program enter the following keystrokes:

Keystroke sequence:	Produces:
[↰] [<< >>] [ENTER]	<<
[↰] [ENTER]	DUP
['] [α] [↰] [N] [▶]	'n'
[STO] [1]	STO 1
[↰] [SWAP]	SWAP
[PRG] [*BRCH*] [*FOR*] [*FOR*]	FOR
[α] [↰] [J]	j
[PRG] [*TYPE*] [*OBJ->*]	OBJ→
[*->ARR*]	→ARRY
[PRG] [*BRCH*] [*IF*] [*IF*]	IF
[α] [↰] [J] [SPC]	j
[α] [↰] [N]	n
[PRG] [*TEST*] [<]	<
[PRG] [*BRCH*] [*IF*] [*THEN*]	THEN
[α] [↰] [J] [SPC] [1] [+]	j 1 +
[↰] [STACK] [*ROLL*]	ROLL
[PRG] [*BRCH*] [*IF*] [*END*]	END
[PRG] [*BRCH*] [*IF*] [*NEXT*]	NEXT
[PRG] [*BRCH*] [*IF*] [*IF*]	IF
[α] [↰] [N] [SPC] [1]	n 1
[PRG] [*TEST*] [>]	>
[PRG] [*BRCH*] [*IF*] [*THEN*]	THEN
[1] [SPC]	1
[α] [↰] [N] [SPC] [1] [-]	n 1 -
[PRG] [*BRCH*] [*FOR*] [*FOR*]	FOR
[α] [↰] [J] [SPC]	j
[α] [↰] [J] [SPC] [1] [+]	j 1 +
[↰] [STACK] [*ROLL*]	ROLL
[PRG] [*BRCH*] [*IF*] [*NEXT*]	NEXT
[PRG] [*BRCH*] [*IF*] [*END*]	END
[α] [↰] [N] [SPC]	n
['] [α] [↰] [N] [▶]	'n'
[↰] [PURGE]	PURGE
[MTH] [*MATR*] [*COL*] [*COL→*]	COL→
[ENTER]	Program is displayed in level 1

To save the program:
['] [α] [α] [C] [R] [M] [T] [α] [STO]

 © 2000 - Gilberto E. Urroz - All rights reserved

14.3. Best data fitting.

The HP48G/GX can determine which one of its linear or linearized relationship offers the best fitting for a set of (x,y) data points. We will illustrate the use of this feature with an example. Suppose you want to find which one of the data fitting functions provides the best fit for the following data:

x	y
0.20	3.16
0.50	2.73
1.00	2.12
1.50	1.65
2.00	1.29
4.00	0.47
5.00	0.29
10.00	0.01

First, enter the data as a matrix, either by using the matrix editor and entering the data, or by entering two lists of data corresponding to x and y and using the program CRMT. To use the latter approach use the following keystrokes:

[↰][{}] [.][2][SPC] [.][5][SPC] [1][SPC] [1][.][5][SPC]
[2][SPC] [4][SPC] [5][SPC] [1][0] [ENTER]
[↰][{}] [3][.][1][6][SPC] [2][.][7][3][SPC] [2][.][1][2][SPC] [1][.][6][5][SPC]
[1][.][2][9][SPC] [.][4][7][SPC] [.][2][9][SPC] [.][0][1][ENTER]
[2][ENTER]
[CRMT]

Next, save this matrix into the statistical matrix ΣDAT, by using: [↰][STAT][DATA][↰][ΣDAT]

(If you press the white button labelled [ΣDAT], you will see the matrix displayed again in level 1 of the display).

Finally, the following instructions will allow you to find the best fit for your data:

[↱][STAT][▼][▼][OK]

The display shows the current ΣDAT, already loaded. Change your set up screen to the following parameters if needed:

```
X-COL: 1   Y-COL: 2
  MODEL: Best Fit
```

Press [OK], to get the following results:

```
1: '3.99504833324*EXP(-.579206831203*X)'
2: Correlation: -0.996624999526
3: Covariance: -6.23350666124
```

The best fit for the data is, therefore, $y = 3.995\, e^{-0.58 \cdot x}$.

Note (1): for quick access to barplots and scatterplots, see the next section (section 15) in Part I of this guide.

Note (2): See sections 15 and 16 in Part II of this guide for multiple-linear and polynomial data fitting.

 © 2000 – Gilberto E. Urroz – All rights reserved

15.0. Use of [↰][STAT] for data analysis, plots, and data fitting.

The keystroke combination [↰][STAT] provides direct access to several of the statistical functions in the calculator, namely:

[DATA][ΣPAR][1VAR][PLOT][FIT][SUMS]

Pressing the key corresponding to any of these menus provides access to different functions as described below.

[DATA]: Commands under this menu are used to manipulate the statistics matrix ΣDATA.

[Σ+]: add row in level 1 to bottom of ΣDATA matrix.
[Σ-]: removes last row in ΣDATA matrix and places it in level of 1 of the stack. The modified ΣDATA matrix remains in memory.
[CLΣ]: erases current ΣDATA matrix.
[ΣDAT]: places contents of current ΣDATA matrix in level 1 of the stack.
[↰][ΣDAT]: stores matrix in level 1 of stack into ΣDATA matrix.
[STAT]: returns to STAT menu.

[ΣPAR]: Commands under this menu are used to modify statistical parameters. The parameters shown in the display are:

Xcol: indicates column of ΣDATA representing x (Default: 1)
Ycol: indicates column of ΣDATA representing y (Default: 2)
Intercept: shows intercept of most recent data fitting ((Default: 0)
Slope: shows slope of most recent data fitting (Default: 0)
Model: shows current data fit model (Default: LINFIT)

n [XCOL]: changes Xcol to n.
n [YCOL]: changes Xcol to n.
[MODL]: lets you change model to LINFIT, LOGFIT, EXPFIT, PWRFIT or BESTFIT by pressing the appropriate button, or press [ΣPAR] to return to the ΣPAR menu.
[ΣPAR]: shows statistical parameters.
[RESET]: reset parameters to default values
[INFO]: shows statistical parameters
[NXT][STAT]: returns to [STAT] menu.

[1VAR] : Commands under this menu are used to calculate statistics of columns in ΣDATA matrix.

[TOT]: show sum of each column in ΣDATA matrix.
[MEAN]: shows average of each column in ΣDATA matrix.
[SDEV]: shows standard deviation of each column in ΣDATA matrix.
[MAXΣ]: shows maximum value of each column in ΣDATA matrix.
[MINΣ]: shows average of each column in ΣDATA matrix.
x_s, Δx, n [BINS]: provides frequency distribution for data in *Xcol* column in ΣDATA matrix with the frequency bins defined as $[x_s,x_s+\Delta x]$, $[x_s,x_s+2\Delta x]$,…, $[x_s,x_s+n\Delta x]$.

 © 2000 - Gilberto E. Urroz - All rights reserved

[NXT]: to access the second menu. Within this menu you will find the following commands:
[VAR]: shows variance of each column in ΣDATA matrix.
[PSDEV]: shows population standard deviation (based on *n* rather than on (*n-1*)) of each column
in ΣDATA matrix.
[PVAR]: shows population variance of each column in ΣDATA matrix.
[MINΣ]: shows average of each column in ΣDATA matrix.
[STAT]: returns to [STAT] menu.

[PLOT]: Commands under this menu are used to produce plots with the data in the ΣDATA matrix.
[BARPL]: produces a bar plot with data in *Xcol* column of the ΣDATA matrix.
[HISTP]: produces histogram of the data in *Xcol* column in the ΣDATA matrix, using the default
width corresponding to 13 bins unless the bin size is modified using
[↰][STAT][1BAR][BINS]. Press [CANCL] to return to normal display.
[SCATR]: produces a scatterplot of the data in Ycol column of the ΣDATA matrix vs. the data in
Xcol column of the ΣDATA matrix. Press [CANCL] to return to normal display.
Equation fitted will be stored in the variable EQ.
[STAT]: returns to [STAT] menu.

[FIT]: Commands under this menu are used to fit equations to the data in columns *Xcol* and *Ycol* of the ΣDATA matrix.
[ΣLINE]: provides the equation corresponding to the most recent fitting.
[LR]: provides intercept and slope of most recent fitting.
y [PREDX]: given y find x for the fitting y = f(x).
x [PREDY]: given x find y for the fitting y = f(x).
[CORR]: provides the correlation coefficient for the most recent fitting.
[COV]: provides sample co-variance for the most recent fitting
[NXT]: to access the second menu. Within this menu you will find the following commands:
[PCOV]: shows population co-variance for the most recent fitting.
[STAT]: returns to [STAT] menu.

[SUMS]: Commands under this menu are used to obtain summary statistics of the data in columns *Xcol* and *Ycol* of the ΣDATA matrix.
[ΣX]: provides the sum of values in *Xcol* column.
[ΣY]: provides the sum of values in *Ycol* column.
[ΣX^2]: provides the sum of squares of values in *Xcol* column.
[ΣY^2]: provides the sum of squares of values in *Ycol* column.
[ΣX*Y]: provides the sum of $x \cdot y$, i.e., the products of data in columns *Xcol* and *Ycol*.
[NΣ]: provides the number of columns in the ΣDATA matrix.

 © 2000 - Gilberto E. Urroz - All rights reserved

Example let ΣDATA be the matrix:

$$\begin{bmatrix} 1.1 & 3.7 & 7.8 \\ 3.7 & 8.9 & 101 \\ 2.2 & 5.9 & 25 \\ 5.5 & 12.5 & 612 \\ 6.8 & 15.1 & 2245 \\ 9.2 & 19.9 & 24743 \\ 10.0 & 21.5 & 55066 \end{bmatrix}$$

- Type the matrix in level 1 of the stack by using the MATRIX editor: [↱][MATRIX]. When done entering values, press [ENTER].

- To store the matrix into ΣDATA, use: [↰][STAT] [DATA] [↰][ΣDAT]

- Calculate statistics of each column: [STAT][1VAR]

[TOT] produces [38.5 87.5 82799.8]
[MEAN] produces [5.5. 12.5 11828.54...]
[SDEV] produces [3.39... 6.78... 21097.01...]
[MAXΣ] produces [10 21.5 55066]
[MINΣ] produces [1.1 3.7 7.8]
[NXT][VAR] produces [11.52 46.08 445084146.33]
[PSDEV] produces [3.142... 6.284... 19532.04...]
[PVAR] produces [9.87... 39.49... 381500696.85...]

- Generate a scatterplot of the data in columns 1 and 2 and fit a straight line to it:

[STAT][ΣPAR][RESET] resets statistical parameters
[NXT][STAT][PLOT][SCATR] produces scatterplot
[STATL] draws data fit as a straight line
[CANCL] returns to main display

- Determine the fitting equation and some of its statistics:

[STAT][FIT][ΣLINE] produces `'1.5+2*X'`
[LR] produces `Intercept: 1.5, Slope: 2`
3 [PREDX] produces 0.75
1 [PREDY] produces 3. 50
[CORR] produces 1.0
[COV] produces 23.04
[NXT][PCOV] produces 19.74

- Obtain summary statistics for data in columns 1 and 2: [STAT][SUMS]

[ΣX] produces 38.5
[ΣY] produces 87.5
[ΣX^2] produces 280.87
[ΣY^2] produces 1370.23
[ΣX*Y] produces 619.49
[NΣ] produces 7

 © 2000 - Gilberto E. Urroz - All rights reserved

- Fit data using columns 1 (x) and 3 (y) using a logarithmic fitting:

[NXT][STAT][ΣPAR][3][YCOL] select Ycol = 3, and
[MODL][LOGFI] select Model = Logfit
[NXT][STAT][PLOT][SCATR] produce scattergram of y vs. x
[STATL] show line for log fitting
Obviously, the log-fit is not a good choice.
[CANCL] returns to normal display.

- Select the best fitting by using:

[STAT][ΣPAR][MODL][BESTF] shows EXPFIT as the best fit for data
[NXT][STAT][FIT][ΣLINE] produces '2.6545*EXP(0.9927*X)'
[CORR] produces 0.99995... (good correlation)
2300[PREDX] produces 6.8139
5.2 [PREDY] produces 463.37

- To return to STAT menu use: [NXT][STATS]
- To get your variable menu back use: [VAR].

 © 2000 - Gilberto E. Urroz - All rights reserved

PART II – PROGRAMS FOR STATISTICAL ANALYSIS

 © 2000 –Gilberto Urroz – All rights reserved

In this section we will present a collection of sub-directories containing User RPL programs that can be used for a number of statistical applications. The programs will be made available by direct transfer through the infrared port of your calculator, or downloaded from the Internet at an address that will be provided to the students. The suggested storage order for the sub-directories is as follows:

Create a directory in your calculator to be named STATS
Within directory STATS, you will have the following sub-directories:

- ONEVAR - descriptive measures for a sample of ONE VARiable
- HISTO - frequency distribution and HISTOgram
- GRSTA - GRouped data STAtistics
- DDIST - Discrete DISTributions
- DFUNC - Discrete FUNCtions
- SIMUL - SIMULation: generation of random number lists
- SIMH - SIMulation Histogram - generation of synthetic random data
- INTS - INTegralS for continuous probability distributions
- CFUNC - Continuous FUNCtions (Gamma, Beta, Exponential, Weibull)
- PLTF - PLoT Functions
- SIMUC - SIMUlation for Continuous Synthetic Data (Gamma, Beta, Exp., Weibull, Normal)
- PLTF - PLoT special Functions (Normal, t, Chi-squared, F)
- NTCF - Normal, t, Chi-square, F distributions
- CHKN - CHecK data for Normality
- $\bar{x}$SIM - $\bar{x}$ distribution SIMulation
- HYPTST - HYPothesis TeSTing
- $\bar{x}$TST1 - hypothesis testing on one mean
- $\bar{x}$TST2 - hypothesis testing on two means
- σTST1 - hypothesis testing on one variance
- σTST1 - hypothesis testing on one variance
- OPCR - OPerating CuRves for hypothesis testing on mean values
- MLIN - Multiple LINear regression
- POLY - POLYnomial regression
- CFIT - Curve Fitting using HP48G/GX own features (described in sections 13 and 14, Part I)
- GDFIT - GooDness of FIT test
- RC - R x C tables
- BAYES - Simple Bayesian estimation
- SDATA - Data files

The directories are described in the following sections. Because each of the subdirectories may take quite a bit of memory in your calculator, you want to keep permanently only those that are more useful, namely:

For statistical analysis: ONEVAR, CHKN, HYPTST, MLIN, POLY, CFIT, GDFIT, RC
For generating synthetic data: SIMUL, SIMH, SIMUC
For probability calculations: DFUNC, CFUNC, NTCF

The other subdirectories can be purged once used for assignment solutions.

 © 2000 -Gilberto Urroz - All rights reserved

Note: most subdirectories have two variables called [*LVAR*] and [*ORDER*]. Pressing the [ORDER] key will organize the variables in a given subdirectory in the order indicated by LVAR, which is the preferred order described below for each subdirectory.

 © 2000 –Gilberto Urroz – All rights reserved

1.0 ONEVAR: Programs for single-variable data analysis

The sub-directory ONEVAR in directory STATS contains a series of programs to be used for statistical analysis of single-variable data.

The instructions for using this directory are the following:

- Enter list x: {x_1 x_2 … x_n} 'x' [STO] or [↰][*X*]
- Press [*DESC*] to obtain list of descriptive measures: mean absolute deviation (MAD), number of data points (n), mean, median, variance (s2), standard deviation (s), minimum value (min.x), maximum value (max.x), Range = max.x - min.x, first quartile (Q1), third quartile (Q3), Inter-quartile range (IQR =Q3 - Q1).
- Press [→*BXP*] to produce a box-and-whiskers plot of the data. Press [*CANCL*] to return to normal display.
- Press [*HISTO*] to move to sub-directory HISTO (see more instructions below or go to step 5).
- Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.
- Press [NXT][→ *XBA*] to get the mean value of list x.
- Press [→ *S2*] to get the variance of list x. To get the standard deviation, press [√*x*].
- Press [→*MED*] to get the median of the distribution.
- To calculate percentiles, enter the percentile number pth% (0 < pth% < 100) into level 1, then press [→*%TIL*]. For example,
- to get $Q_1 = P_{0.25}$, use [2][5][→*%TIL*].
- To get $Q_3 = P_{0.75}$, use [7][5][→*%TIL*].
- Press [→*MM*] to obtain the minimum and maximum value of list x. These values are useful to determine the histogram information.
- Press [NXT] to see the variables n, xbar, mdn, s2, MAD, xmin.
- Press [NXT] to see the variables xmax, range, Q1, Q2, AM (A-), AP (A+).
- Press [VAR] to return to main variable menu.
- To determine frequency distributions for a histogram enter subdirectory *Histo* by pressing [*HISTO*], then (*note*: make sure to press [*DESC*] before moving into [HISTO]:
- Enter a value for the number of bars to be displayed, *b*. Store it in variable [B]. For example, if you choose *b = 5*, enter [5][↰][*B*] or [5]['][*B*][STO].
- Enter a value for the minimum class boundary, y_{low}, store it in variable [*YLOW*]. For example, if you choose this minimum value to be 0, enter [0][↰][*YLOW*] or [0]['][*YLOW*][STO].
- Enter a value for the maximum class boundary, y_{high}, store it in variable [*YHIGH*]. For example, if you choose this minimum value to be 0, enter [0][↰][*YHIGH*] or [0]['][*YHIGH*][STO].
- Press [→ *FRQ*] to obtain frequency distribution. Level 3 represents the class marks. Level 2 represents the frequency counts in the NBIN classes, and level 1 shows the outliers in the x list, i.e., values of x < XBMI, and, x>XBMA.
- To plot the histogram, press [→*HIST*]. Press [*CANCL*] to return to normal display.
- Press [↑*ONE*] to return to sub-directory ONEVAR.
- To produce an ogive of "less than" cumulative frequencies, press [NXT][→*OGV*].
- To produce a dot plot of the data, press [NXT][→*DPL*]. Press [*CANCL*] to exit.

 © 2000 –Gilberto Urroz – All rights reserved

Example: within subdirectory ONEVAR try the following data analysis:
Store the following data in x: { 10 7 3 2 1 5 3 2 8 9 }.

Press: *Display shows:*

```
[DESC]
4:         median:  4
3:            mean: 5
2:               n:10
1:   MAD: 3.111111111
```

Press [⇦] as required to see more information, e.g., *s^2: 10.666666*; *s: 3.2652…; min.x: 1; max.x: 10; Range: 9; Q1: 2; Q3: 8.5; IntQ_Range: 6.5*.

[NXT][→*XBA*]	mean: 5 / n: 10
[→*S2*]	s^2: 3.26598….
[→*MED*]	median: 4
[2][5][→*%TIL*]	P 0.25: 2
[3][7][→*%TIL*]	P 0.37: 3
[→*MM*]	min.x:1 / max.x: 10
[↰][PREV][*HISTO*]	To move to HISTO sub-directory
[5][↰][*B*]	Store b = 5
[0][↰][*YLOW*]	Store BWIDTH = 2
[5][↰][*NBIN*]	Store NBIN = 5
[→*FRQ*]	c.marks: {1 3 5 7 … / freq.: { 1 4 1 1 3 … / outliers: [0 0]
[→*HIST*]	Plots histogram. Press [CANCL] to exit.
[→*DPL*]	Plots dot plot. Press [CANCL] to exit.
[NXT][→*OGV*]	Plots "less than" cumulative frequency ogive. Press [CANCL].
[↑*ONE*]	To return to ONEVAR upper directory.
[↰][UP]	To return to upper directory STATS.

 © 2000 -Gilberto Urroz - All rights reserved

© 2000 –Gilberto Urroz – All rights reserved

2.0 GRSTA: Programs for grouped-data statistics

Programs that can be used to calculate statistics for grouped data are contained in a directory called 'GRSTA' (GRouped-data STAtistics). The instructions for using the programs are the following:

- Enter list x and store it in variable 'x'.
- Enter list f and store it in variable 'f'.
- Press [→*XBA*] to get the total count and the mean value of list x.
- Press [→*S2*] to get the variance and standard deviation of list x.
- Press [→*CFR*] to get the cumulative frequency distribution.
- To calculate a percentile, enter the percentile (a number between 1 and 100), then press [→%TIL].
- Press [NXT] for the next menu. Press [→HIST] to plot a histogram of the data.
- Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.

To test the programs use the following data, x = {6.95 10.95 14.95 18.95 22.95 26.95 30.95}, and f = {3 10 14 25 17 9 2}. Press [→*XBA*], you should get a value of $\bar{x}$ = 18.85, and n = 80. Press [→*s2*] to obtain the sample variance s^2 = 30.77, and s = 5.547. Press [→*CFR*] to get the lists for x, f, and cf: { 3 13 27 52 69 78 80 }.

 © 2000 -Gilberto Urroz - All rights reserved

3.0 DDIST : Programs for distribution of discrete random variables.

Use this directory when you want to analyze the probability distribution of a discrete random variable given as a table or as a function. If the pdf is given as a table:

- Enter list x and store it in variable 'x'.
- Enter list f and store it in variable 'f'.
- Press [*CHKF*] to check that $\Sigma f = 1$. If that is not the case, then the pdf is valid.
- Press [→*CDF*] to get the corresponding cumulative distribution function.
- Press [μ] to calculate the mean of the distribution.
- Press [$\sigma 2$] to calculate the variance of the distribution.
- Press [NXT] for next menu. Press [*SKEW*] to calculate the skewness of the distribution.
- Press [*KURT*] to calculate the kurtosis of the distribution.
- Press [↑*DFU*] to go to subdirectory DFUNC (Discrete FUNCtions). To return to DDIST, find a white key called [↑*DDIS*].
- To generate a list of discrete, equally spaced values, enter the beginning value, the ending value, and the increment between values, and press [→*XLIS*].
- To generate a list of n values all equal to a constant c, enter n, then c, and press [LISC].
- Press [→*PLP*] to plot the pdf.
- Press [NXT] to get next menu. Press [→*PLD*] to plot the cdf.
- Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.

If the pdf is given as a function, say, f(x), first calculate the vector f by using the appropriate operations in the calculator. For example, if $f(x) = x^2/25$, first, place x in level 1 of the display, by pressing [X], then, press [←][x^2][2][5][ ]['][F][STO]. Then, proceed as in the case in which the pdf is given as a table.

As an example, enter the following values of x and f as lists, namely, x = {0 1 2 3 4 }, and f = {0.05 0.20 0.45 0.20 0.10}. Using the values of x and f already stored, we can perform the following operations by using the white keys:

Operation	Key	Result shown
Check if the distribution sums to 1.0:	[*CHKF*]	Σf: 1
Calculate the CDF	[→*CDF*]	cdf: {.05 .25 .7 .9 1}
Calculate the mean:	[μ]	μ: 2.1
Calculate the variance:	[$\sigma 2$]	σ^2: 0.99; σ: 0.9949
Calculate the skewness	[*SKEW*]	skew: 0.10354...
Calculate the kurtosis	[*KURT*]	kurt: 2.7545

If the probability distribution is given as a function, we can use a similar approach to check if it is a valid distribution and to calculate the parameters. For example, if f(x) = (x+1)/25, x = 1,2,3,4,5, we can enter the values of x into variable x, as we did before, and then calculate f(x) and store into F, as follows:

[X][1][ENTER][MTH][*LIST*][*ADD*] Calculates the list (x + 1).

 © 2000 -Gilberto Urroz - All rights reserved

Note that here we could not use [X][1][+] because it will only add the number 1 at the end of the list. Addition is, therefore, the only arithmetic operation for lists where, when adding a number to each element of the list, you need to use the [ADD] command in [MTH][LIST] instead of [+].

[2][5][÷]	Calculates the list (x + 1)/25.
[VAR]['][*F*][STO]	Stores the values for the probability distribution f.
[*CHKF*]	Checks that Σf = 1. Display shows: .8.

The calculations must stop here because the condition Σf = 1 is not satisfied. Just to make sure that your programs are working properly, you should obtain the following results:

[μ]	Display shows: μ: 2.8
[$\sigma 2$]	Display shows: σ^2: 3.36; σ: 1.833...

If we want to generate a list of values {0.0 0.5 1.0 1.5 2.0 ... 16.5}, enter the following: 0 [ENTER] 16.5 [ENTER] 0.5 [ENTER][→*XLIS*]. The result will be the desired list, which we can store into the variable x. [VAR][↰][*X*]. Suppose that we define a uniform distribution for this list. We need to know the number of elements in x: [*X*][PRG][*LIST*][*ELEM*][*SIZE*]

This indicates that there are 34 elements in the list, therefore, the uniform probability distribution is such that, for any value of x, P(X=x) = f(x) = 1/34 = 2.941176E-2. To generate the probability distribution we use the program LISC. Enter: 34 [ENTER] 34 [ENTER][1/x][ENTER][LISC]. Store this list into f : [↰][*F*]. Now, perform the following calculations:

Operation	Key	Result shown
Check if the distribution sums to 1.0:	[*CHKF*]	Σf: 1
Calculate the mean:	[μ]	μ: 8.25
Calculate the variance:	[$\sigma 2$]	σ^2: 24.062; σ: 4.905
Calculate the skewness	[*SKEW*]	skew: -5.083E-12 ≈ 0
Calculate the kurtosis	[*KURT*]	kurt: 1.7979
Plot the pdf	[→*PLP*]	uniform plot
Plot the cdf	[→*PLD*]	linear plot

(See the description of DFUNC for additional use of this DDIST sub-directory).

 © 2000 -Gilberto Urroz - All rights reserved

4.0 DFUN: Probability distributions and distribution functions of discrete random variables

This directory contains definitions for the following functions:

* Binomial probability distribution: `'bpd(x,n,p)=COMB(n,x)*p^x*(1-p)^(n-x)'`
* Binomial distribution function: `'BDF(x,n,p) = Σ(k=0,x,bdp(k,n,p))'`
* Poisson Probability Distribution: `'PoPD(x,λ) = λ^x*EXP(-λ)/x!`
* Poisson Distribution Function: `'PoDF(x, λ) =Σ(k=0,x,'PoPD(k,λ))'`
* Hypergeometric Prob. Distr.: `'hpd(x,n,a,N)=COMB(a,x)*COMB(N-a,n-x)/COMB(N,n)'`
* Geometric Distribution Function: `'gpd(x,p) = p*(1-p)^(x-1)'`

Operation of functions:

*	Binomial probability distribution:	x [ENTER] n [ENTER] p [ENTER] [*BPD*]
*	Binomial distribution function:	x [ENTER] n [ENTER] p [ENTER] [*BDF*]
*	Poisson Probability Distribution:	x [ENTER] λ [ENTER] [*POPD*]
*	Poisson Distribution Function:	x [ENTER] λ [ENTER] [*PODF*]
*	Hypergeometric Prob. Distr.:	x [ENTER] n [ENTER] a [ENTER] N [ENTER] [*HPD*]
*	Geometric Distribution Function:	x [ENTER] p [ENTER] [*GPD*]

Examples for calculation of single values of these functions were presented in Sections 7.1 and 7.2 of Part I of this guide. In this section we'll present operations with list of data. Press [NXT] to access the following features:

For the pdf's used in this sub-directory, the values of the random variable are positive integers. Therefore, to generate a list of integers, enter the lowest value, then the largest value, then press [→*XLIS*].

When you have a list of values of x and of the pdf f(x) in levels 2 and 1, respectively, press [→*ΣDA*] to create the statistic matrix ΣDAT.

To calculate the cumulative distribution function, place the pdf in level 1 and press [→*DFN*].

To plot f(x)-vs-x, create the statistic matrix ΣDAT as indicated in 2, above, then press [→*PLT*].

To plot F(x)-vs-x, place x in level 1 and f(x) in level 2. Calculate the cdf as indicated in step 3, above. Then create the statistic matrix ΣDAT, as indicated in 2, above. Finally, press [→*PLT*]. To move to sub-directory DDIST, press [↑*DDIST*].

For example, suppose that we want to get the entire binomial distribution for n=10, p = 0.35. Here is the way to proceed:

Generate list {0 1 … 10}and keep two extra copies handy: [0][SPC][1][0][→*XLIS*][ENTER][ENTER]
Generate pdf, enter n and p, then press [*BPD*]: [1][0][ENTER][.][3][5][ENTER][*BPD*]
Check the pdf using subdirectory DDIST: [↑*DDIS*]. Store the pdf (level 1) into f. The list x moves to level 1. Store it in x.

 © 2000 -Gilberto Urroz - All rights reserved

The following steps are performed within DDIST:

Operation	Key	Result shown
Check if the distribution sums to 1.0:	[*CHKF*]	Σf: 1
Calculate the mean:	[*μ*]	μ: 3.5
Calculate the variance:	[*σ2*]	σ^2: 2.275; σ:
Calculate the skewness	[NXT][*SKEW*]	skew: 0.19889806
Calculate the kurtosis	[*KURT*]	kurt: 2.83956043
Plot the pdf	[→*PLP*]	(see display)
Plot the cdf	[NXT][→*PLD*]	(see display)

Plots can also be obtained from within sub-directory DFUNC, as follows: First, get x and f(x) back into levels 2 and 1, respectively, of the display: [VAR][*X*][*F*]. Then, move back to sub-directory DFUNC, by [NXT] [↑*DFU*].
Create a ΣDAT matrix by pressing [→*ΣDA*]. A message indicating that the matrix is "READY" will be shown in level 1. Now, press [→*PLT*]. The display shows a barplot. Press [*CANCL*] to get back to normal display.
Go back to sub-directory DDIST, and get another copy of x and f(x), as done before. Then, return to sub-directory DFUNC.
Press [→*DFN*] to get the cdf. Then, create a new ΣDAT matrix by pressing [→*ΣDA*]. When "READY", press [→*PLT*] to plot the cdf. . Press [*CANCL*] to get back to normal display.

Similar procedures can be performed for the Poisson distribution with λ = 1.2, for x = 0, 1, 2, … 25.
Generate list {0 1 … 25}and keep two extra copies handy: [0][SPC][2][5][→*XL/S*][ENTER][ENTER]
Generate pdf, enter λ, then press [*BPD*]: [1][.][2][ENTER][VAR][*POPD*]
Check the pdf using subdirectory CDIST: [NXT][↑*DDIS*]. Store the pdf (level 1) into f. The list x moves to level 1. Store it in x.
The following steps are performed within DDIST:

Operation	Key	Result shown
Check if the distribution sums to 1.0:	[*CHKF*]	Σf: 0.9999 (*)
Calculate the mean:	[*μ*]	μ: 1.1999
Calculate the variance:	[*σ2*]	σ^2 :1.2 / σ: 1.095
Calculate the skewness	[*SKEW*]	skew: 0.9128
Calculate the kurtosis	[*KURT*]	kurt: 3.83333
Plot the pdf	[→*PLP*]	(see display)
Plot the cdf	[→*PLD*]	(see display)

(*) Because the Poisson distribution is defined for x = 0, 1, 2, … . You will never get exactly 1.0 here for such distribution.

Now let's check the plots within sub-directory DFUNC, as follows: First, get x and f(x) back into levels 2 and 1, respectively, of the display: [VAR][*X*][*F*]. Then, move back to sub-directory DFUNC, by pressing [NXT][↑*DFU*].

Create a ΣDAT matrix by pressing [NXT][→*ΣDA*]. A message indicating that the matrix is "READY" will be shown in level 1. Now, press [→*PLT*]. The display shows a barplot. Press [*CANCL*] to get back to normal display.

Go back to sub-directory DDIST, and get another copy of x and f(x), as done before. Then, return to sub-directory DFUNC.

 © 2000 -Gilberto Urroz - All rights reserved

Press [→*DFN*] to get the cdf. Then, create a new ΣDAT matrix by pressing [→*ΣDA*].
When "READY", press [→*PLT*] to plot the cdf. Press [*CANCL*] to get back to normal display.

 © 2000 -Gilberto Urroz - All rights reserved

5.0 SIMUL: Programs for random number generation.

Contains programs for generating random numbers and random number lists.

- To generate a random number between 0 and 1, press [→RAN]
- To generate an integer random number of three digits, i.e., between 0 and 1000, press [RAN3]
- To generate a list of integer random numbers between 0 and 1000, enter the number of element sin the list and press [RNL3].
- To generate a list of n integer random numbers between 0 and 10m, use these keystrokes:
- n [ENTER] m [ENTER] [RNLS]
- To re-start the HP48 G random number generator enter a 'seed' number and press [→RDZ]
- To get to directory SIMH (for analysis of random number lists) press [↑SIM]

Examples:

To generate one random number between 0 and 1, press [→RAN]. One possible outcome is: 0.73136244.
To generate one random number between 0 and 1000, press [RAN3]. One possible outcome is 772.
To generate 10 random numbers between 0 and 1000, use: [1][0][RNL3]. One possible outcome is {990 249 867 34 588 25 807 724 103 56}.
To generate a list of 10 integer random numbers between 0 and 100, use: [1][0][ENTER][2][ENTER][RNLS]. One possible outcome is: { 64 23 54 81 70 67 66 23 73 34}.

 © 2000 -Gilberto Urroz - All rights reserved

6.0 SIMH: Programs for generating discrete synthetic data and generating histograms of the same.

The programs in this subdirectory are used to analyze lists of random numbers, and to generate values of a random variable x based on the list of random numbers. Also, a histogram for the list of random numbers can be obtained.

First, a list of random numbers must be generated in subdirectory SIMUL. This list must then be stored in variable r in subdirectory SIMH. We'll assume that the random numbers are integers between 0 and 1000:

- To go to subdirectory SIMUL, press [↑*SIM*]. Generate a list of random numbers as indicated above and keep it in level 1. Within subdirectory SIMUL, press [↑*SIM*] to get to SIMH.
- Store list of random numbers by using: [↰][*R*].
- Enter a list of possible values of x into variable *tbx* (*ta*b*le* of *x*), e.g.,
  {0 1 2 3} ['][*TBX*][STO].
- Enter a list of class boundaries of numbers between 0 and 1000, e.g.,
  { 0 287 757 958 1000} ['][*BOUN*][STO]

Note: when assigning values of x in program →XLIST, x will take the value tbx_i, if

$$bound_i \leq x < bound_{i+1}$$

- Press [→*XLIST*] to generate a list of variables taking values from *tbx*. The list is stored in x.
- Press [→*HIST*] to get a frequency count of the random number list *r*.
- Plot the histogram by pressing: [↰][STAT][*PLOT*][*BARPL*]. Press [*CANCL*] [VAR] to return to the main display.
- To see the frequency counts as a list, press [↱][*SUM*]
- With the frequency count list in level 1, press [*N*] [÷], to get the relative frequency count.
- Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.

To illustrate the way the program operates, we have generated a list of 5 random numbers using the program RNL3 in subdirectory SIMUL, i.e., within the subdirectory SIMUL we used [5][*RNL3*]. The list generated was placed in level 1 of the display, in this case, it was { 89 743 800 118 791}. We pressed [↑*SIM*] within subdirectory SIMUL to get to subdirectory SIMH, and stored the list into variable r, by pressing:

[↰][*R*][STO].

We also entered the values of Bound, i.e., for this case we would have entered {0 287 757 958 1000} and stored it in *Bound*. We also entered the values of tbx (table of x), i.e., {0 1 2 3} and stored them in the corresponding variable. To generate the list of variables, x, we pressed [→XLIS]. To see the list generated press [↱][*X*]. We get { 0 1 2 0 2}. What we have done was to generate a list of 5 values of the variable x from the possible values listed in tbx by using the list of random numbers stored in r according to the following table:

 © 2000 -Gilberto Urroz - All rights reserved

x	r From	To	probability
0	0	286	0.286
1	287	756	0.470
2	757	957	0.201
3	958	1000	0.043
		sum	1.000

Next, we can get frequency distribution and plot a histogram of the random numbers by pressing [→*HIST*]. Press [*CANCL*] [VAR] to return to the main display. The frequency count determines the number of random numbers falling within each interval. The column vector in the display is the frequency count shown as a column vector, which corresponds to the contents of variable ΣDAT. The result is also stored in *sum* (as a list). To see the frequency counts, press [↱][SUM]. You'll get {2 1 2 0} for this case.

This example is very poor in the sense that the number of random numbers generated is very small. Try using a list of 100 random numbers to generate a histogram. Use the following:

[NXT][↑*SIM*][1][0][0][*RNL3*][↑*SIM*][↰][*R*][→*HIST*][↱][*SUM*]

The frequency count is now {30 42 21 7}. Press [NXT][*N*] [÷], to get the probability distribution given by the histogram. You'll get { 0.3 0.42 0.21 0.07}. These values are close to the values of 0.286, 0.47, 0.201 and 0.043 given in the table above.

(**Note**: when you perform this exercise in your calculator, you'll get slightly different results because you will probably get a different list of random numbers than mine.)

To generate a list of values of x, press [→*XLIS*]. To see the list, press [↱][*X*]. It will take your calculator about 1 minute to finish generating the list.

 © 2000 -Gilberto Urroz - All rights reserved

7.0. INTS: programs for probabilities and population parameters for continuous distributions

Enter an expression for f(x) between quotes (e.g., ' COS(x)', lower case x is required).
Enter lower limit of integral into [*A*].
Enter upper limit of integral into [*B*].
Press [→*PRO*] to calculate P(a<X<b).
Press [→μ] to calculate the mean of the distribution. (Use appropriate limits).
Press [→$\sigma 2$] to calculate the variance of the distribution. (Use appropriate limits).
Press [NXT] for next menu. Press [*SKEW*] for skewness of distribution. (Use appropriate limits).
Press [*KURT*] to calculate the kurtosis of the distribution. (Use appropriate limits).

If the integration limits includes infinity , use the MAXR constant from the calculator instead of infinity:

[MTH][NXT][*CONS*][NXT][*MAXR*]

Example: Using the Standardized Normal distribution, $f(x) = \exp(-x^2/2)/(2\pi)^{1/2}$, $-\infty<x<\infty$. Find: P(1<X<2); mean, variance, skewness, and kurtosis of the distribution.

['][↰][e^x][-][α][↰][X][y^x][2][÷][2] [▶]	
[÷][√*x*][↰][( )][2][×][↰][π] [ENTER]	Define f(x)
[↰][*F*]	Enter f(x)
[1][↰][*A*] [2][↰][*B*]	Enter a = 1, b = 2
[→*PRO*]	Calculate P(1<X<2) = 0.1359.

For the distribution parameters change the integration limits as follows:

[0][↰][*A*]	Enter a = 0
[MTH][NXT][*CONS*][NXT][*MAXR*]	Get MAXR in level 1
[ENTER][+/-]	Get -MAXR in level 1
[VAR][↰][*A*]	Enter a = -MAXR (i.e., -calculator infinity)
[↰][*B*]	Enter b = MAXR (i.e., calculator infinity)
[→μ]	(*) Calculate the mean, μ = 0.0
[→$\sigma 2$]	(*) Calculate the variance, σ^2= 1.0
[NXT][SKEW]	(*) Calculate the skewness, skew = 0.0
[KURT]	(*) Calculate the kurtosis, kurt = 3.0

(*) The calculator takes some time to calculate these values. Give each calculation anywhere from 2 to 5 minutes.

Note: Because the calculator uses a numerical approach to calculate the integrals, if the limits of integration include ±∞, it will take a while for the integral to converge. Therefore, try first to find a closed-form integral by hand before attempting the numerical calculation. If the numerical calculation is the only possibility, be patient with your calculator.

 © 2000 -Gilberto Urroz - All rights reserved

8.0 CFUNC: Programs for continuous probability distributions

These programs are used for calculating pdf and cdf for the Gamma, Beta, exponential, and Weibull distributions. This sub-directory includes another sub-directory, called PLTF, which can be used for plotting pdfs. See also section 9.2 for the function definitions and simple examples.

- Press [READ] to see a message.
- Enter values of α and β, corresponding to the same parameters in the Gamma, Beta, exponential, and Weibull distributions. For the exponential distribution only the parameter β is needed.
- Enter a value of x and press the appropriate white button to get any of the following:

[FGAM]	The gamma function (needs only x)
[GRPD]	The gamma pdf (needs α and β) ($x > 0$)
[GRDF]	The gamma cdf (needs α and β) ($x > 0$)
[NXT][βPD]	The beta pdf (needs α and β) ($0 < x < 1$)
[βDF]	The beta cdf (needs α and β) ($0 < x < 1$)
[EXPD]	The exponential pdf (needs only β) ($x > 0$)
[EXDF]	The exponential cdf (needs only β) ($x > 0$)
[WEPD]	The Weibull pdf (needs α and β) ($x > 0$)
[WEDF]	The Weibull cdf (needs α and β) ($x > 0$)

- Press [VAR][NXT][NXT][PLTF] to move to a sub-directory that plots the functions.
- Press [COPY] to copy the values of α and β from the upper directory.
- Press [↱][PLOT] to get into the HP48G/GX PLOT environment. Please refer to section 1.15 in Part I of this guide for more information on HP48G/GX plots.
- Select the function that you want to plot by pressing [CHOOS]. The functions correspond to variables with the names listed above, i.e., FGAM, GRPD, GRDF, etc.
- Change the INDEP variable name to x (lower-case x), if needed.
- Select an appropriate range for x (H-VIEW).
- Use AUTOSCALE for the y range.

You may need to redefine the INDEPendent variable values by pressing [OPTIONS] and changing the LO and HI values of INDEP. If you enter this screen, press [OK] to return to the PLOT environment.

- Press [ERASE][DRAW] to plot the function.
- Press [CANCL] to return to PLOT environment.
- Press [ENTER] or [ON] to get to normal display.
- Press [↰][UP] to return to CFUNC.
- Press [↰][UP] to return to the upper directory STATS.

See section 9.2 for examples of evaluating the functions. The following is an example for a plot. First, store the values $\alpha = 3$ and $\beta = 2$ in the corresponding variables in directory CFUNC. Then, press the following keys:

 © 2000 -Gilberto Urroz - All rights reserved

[VAR][NXT][NXT][PLTF]	Move to sub-directory PLTF
[COPY]	Copy the values of α and β from CFUNC.
[↱][PLOT]	Get into the HP48G/GX PLOT environment.
[CHOOS]	To select the function to plot.
For this case, select βpd, and press [OK].	Place a check mark (✓) on AUTOSCALE.
[ERASE][DRAW]	To plot the function. (Change plots limits if you want to see more detail.)
[CANCL][ON]	To return to normal display.

© 2000 -Gilberto Urroz - All rights reserved

9.0 SIMUC: programs to generate synthetic data based on continuous probability distributions

These programs generate lists of data that follow certain continuous probability distributions including the uniform, exponential, Weibull, normal and lognormal. The variables α and β used in this sub-directory correspond to the parameters of the same name for the uniform [$f(x) = 1/(\beta-\alpha)$, $\alpha < x < \beta$], exponential, and Weibull distributions. For the normal distribution, $\alpha = \mu$, and $\beta = \sigma$. Finally, for the log-normal distribution, α and β, are defined in its pdf: $f(x) = x^{-1}\cdot\exp(-(\ln x-\alpha)/2\beta^2)/(2\pi\beta)^{1/2}$, $x>0$, $\beta>0$.
To use this sub-directory, first enter α and β in the corresponding variables, then, enter the number of data points you want to generate. Finally, press the appropriate button for the distribution selected, i.e.,

[UNIF]	Uniform distribution
[EXPF]	Exponential distribution
[WEIF]	Weibull distribution
[NORF]	Normal distribution
[NXT][LOGN]	Log-normal distribution
Press [↑CHK]	To go to the CHKN (check for Normality) sub-directory.

Example: Generate 10 data points for a log-normal distribution with $\alpha = 12.5$, $\beta = 2.5$.

[1][2][.][5][↰][α]	Enter α
[2][.][5][↰][β]	Enter β
[1][0][NORF]	Generates synthetic data: {8.73 16.28 17.80 ... 11.01}

 © 2000 -Gilberto Urroz - All rights reserved

10.0 PLTF: plot continuous functions (Normal, t, Chi-square, F)

This sub-directory is used for plotting pdf for the Normal, t, Chi-square (χ^2), and F distributions. For their definitions refer to sections 10 and 11 of Part I of this guide. To plot any of these pdf press the corresponding key:

[NORM]	Normal distribution ($-\infty < x < \infty$)
[TDIST]	Student t distribution ($-\infty < t < \infty$)
[CHISQ]	Chi-square (χ^2) distribution ($\chi^2 > 0$)
[FDIST]	F distribution (F > 0)

Pressing the appropriate key will provide information on the next step. For example, if you want to plot the t distribution, press [TDIST]. The calculator will display a message indicating:

```
Enter ν, Press [→][PLOT].
```

Press [OK] to return to normal display. A white key labeled [ν] will be available for you to store the value of n. Say, ν = degrees of freedom = 10, enter:

[1][0][↵] [ν]

Next, press

[↱][PLOT]

This will get you into the HP48G/GX PLOT environment. Select the appropriate range for the INDEP variable, which is already defined as t. For example, select

```
H-VIEW: -5   5
```

Place a check mark in the AUTOSCALE, then press

[ERASE][DRAW].

Press [CANCL] to return to the PLOT environment, and [ENTER] to return to normal display. To get back to the first menu in the sub-directory, press [VAR].
Experiment with the other plots by pressing the appropriate key: [NORM][CHISQ] or [FDIST].

© 2000 -Gilberto Urroz - All rights reserved

11.0 NTCF - Normal, t, Chi-square, and F distributions

This sub-directory includes programs that allow you to calculatethe cumulative distribution function (cdf) or its inverse for the normal distribution with mean μ and variance σ^2 [*PNOR*].

Also, the sub-directory includes programs for calculating the upper-tail cdf or its inverse for the following distributions:

- the normal distribution with mean μ and variance σ^2 [*NORM*],
- the standardized normal distribution ($\mu = 0$, $\sigma^2 = 1$) [*SNOR*],
- the t distribution with ν degrees of freedom [*TDIST*],
- the Chi-square (χ^2) distribution with ν degrees of freedom [*CHISQ*], and,
- F distribution with νN degrees of freedom in the numerator and νD degrees of freedom in the denominator [*FDIST*].

In this sub-directory, the upper-tail probability distributions are referred to by the variable α. For example, for the standardized normal distribution, α = P(Z>z).

The main display in the subdirectory will show the buttons referred to above:

[*PNOR*][*NORM*][*SNOR*][*TDIST*][*CHISQ*][*FDIST*]

The following are applications of the programs in this sub-directory:

Probability calculations for the normal distribution. Press [VAR][*PNOR*]:
Determine P(X < 2.5) for a normal distribution with mean μ = 1.5, and variance σ^2 = 0.5. Use the following keystrokes:

[1][.][5][↰][*μ*] [.][5][↰][*σ2*] [2][.][5] [*X→P*]   The results are: x: 2.5, P: 0.92.

Determine x if P(X < x) = 0.35 for a normal distribution with mean μ = 2.5, and variance σ^2 = 0.16. Use the following keystrokes:

[2][.][5][↰][*μ*] [.][1][6][↰][*σ2*] [.][3][5] [*P→X*]   The results are: P: 0.35, x: 2.34.

Upper-tail probability calculations for the normal distribution. Press [VAR][NORM]:
For a normal distribution with mean μ = 3.5, and variance σ^2 = 0.25, determine the value of α = P(X>x), if x = 2.5. Use the following keystrokes:

[3][.][5][↰][*μ*] [.][2][5][↰][*σ2*] [2][.][5] [*X→α*]   The results are: x: 2.50, α: 0.98.

Determine x if α = P(X > x) = 0.05 for a normal distribution with mean μ = 2.5, and variance σ^2 = 0.9. Use the following keystrokes:

[2][.][5][↰][*μ*] [.][9][↰][*σ2*] [.][0][5] [*α→X*]   The results are: α: 0.05, x: 4.06.

 © 2000 -Gilberto Urroz - All rights reserved

Upper-tail probability calculations for the standardized normal distribution. Press [VAR][*SNOR*]: For the standardized normal distribution, determine the value of α = P(Z>z), if z = 1.1. Use the following keystrokes:

[1][.][1] [*Z→α*] The results are: z: 2.50, α: 0.14.

Determine z if α = P(Z > z) = 0.10 for the standardized normal distribution. Use the following keystrokes:

[.][1][*α→Z*] The results are: α: 0.05, z: 1.28.

Probability calculations for the standardized normal distribution. Press [VAR][*SNOR*]:

Determine the probability P(Z<z) for z = 0.25. Use the following keystrokes:

[.][2][5][z→P] The results are: z: .25, P: .598706...

Determine z if P(Z<z) = 0.75 for the standardized normal distribution. Use the following keystrokes:

[.][7][5] [P→z] The results are: P: .75, z: .773372...

Upper-tail probability calculations for the t distribution. Press [VAR][*TDIST*]: For a t-distribution with ν = 12 degrees of freedom, determine the value of α = P(T>t), if t = 0.5. Use the following keystrokes:

[1][2][↵][*ν*] [.][5] [*T→α*] The results are: t: 2.50, α: 0.31.

Determine t if α = P(T > t) = 0.05 for with ν = 5 degrees of freedom. Use the following keystrokes:

[5][↵][*ν*] [.][0][5] [*α→T*] The results are: α: 0.05, t: 2.01.

Upper-tail probability calculations for the Chi-squared (χ^2) distribution. Press [VAR][*CHISQ*]: For a χ^2-distribution with ν = 8 degrees of freedom, determine the value of α = P(Ξ^2>χ^2), if χ^2 = 2.5. Use the following keystrokes:

[8][↵][*ν*] [2][.][5] [*X2→α*] The results are: X^2: 2.50, α: 0.96.

Determine χ^2 if α = P(Ξ^2>χ^2) = 0.05 with ν = 5 degrees of freedom. Use the following keystrokes:

[5][↵][*ν*] [.][0][5] [*α→X2*] The results are: α: 0.05, X^2: 11.07.

Upper-tail probability calculations for the F distribution. Press [VAR][*FDIST*] For an F-distribution with νN= 5, and νD= 8, determine the value of α = P(ℑ>F), if F = 0.5. Use the following keystrokes:

[5][↵][*νN*] [8][↵][*νD*] [.][5] [*F→α*] The results are: F: 0.50, α: 0.76.

Determine F if α = P(ℑ>F) = 0.05 for with νN= 15, and νD= 5,. Use the following keystrokes:

[1][5][↵][*νN*] [5][↵][*νD*] [.][0][5] [*α→F*] The results are: α: 0.05, F: 4.61.

 © 2000 -Gilberto Urroz - All rights reserved

12.0 CHKN : programs to check for normality of data

The suggested procedure for using this sub-directory is as follows:

- Enter value of list x: $\{x_1\ x_2 \ldots x_n\}$ 'x' [STO]
- Press [→ ΣDAT] to create a two column array called 'ΣDAT ' that contains in column one the standardized values (z) of the sorted data x, i.e., 'DATA', and, in column 2, the corresponding normal scores (m), i.e., 'scores'.
- Press [→PLT] to plot scores vs. data.
- Press [NXT][↑SIM] to get to subdirectory SIMUC (see section 9, Part II).
- Press [↑SDA] to get to subdirectory SDATA (see section 22, Part II).

As an example, store the list {3 5 6 1 2 } into x. Press [→ΣDA] to build the variable ΣDAT. Then press [→PLT]. The five point seem to fall along a straight line, to check if that is the case, press [STATL]. This will trace the best-fit line through the data. Press [CANCL] to return to normal display. Press [EQ] to get the equation of the best-fit line. For this case: `'4.43337E-13 + .73895*X'`, i.e.,

$$m = 4.43337E\text{-}13 + .73895*z \approx 0.73895*z.$$

For normal data, we expect m = z. To find out what the correlation coefficient is for this fitting, use the "Fit data..." feature in the HP48G/GX as explained in sections 13 and 14 of Part I of this guide. Use the following keystroke sequence:

[↱][STAT][▼][▼][OK]

The display shows the current ΣDAT, already loaded. Change your set up screen to the following parameters if needed:

```
X-COL: 1   Y-COL: 2
MODEL: Linear Fit
```

Then, press [OK], to get the following results:

```
1: '4.43337E-13 + .73895*X'
2: Correlation: 0.986835465419
3: Covariance: 0.738951493934
```

A correlation of 0.986 indicates that the linear fit is excellent. However, as indicated above, for a Normal distribution we would expect m ≈ 0.99*z, rather than m ≈ 0.73895*z. In spite of that, considering that we are testing only five points, a linear fitting as obtained above, can be considered excellent for normally distributed data.

Let's try another exercise. Move to directory SIMUL, and enter:

[5][0][ENTER][2][ENTER][RNLS]	Generates a list of 50 random numbers (0<r<100)
[↰][UP][CHKN]['][X][STO]	Store into variable x in CHKN
[→ΣDAT]	It will take a little while to get the process completed.
[→PLT]	The graph will show a scatter plot.
[STATL]	Draws the best straight line through the points.

You will notice that the straight line fits most of the data points, but the scatter plot diverges from the straight line at both the lower and upper values of x. This is a typical behavior of a uniform (rather than a normal) distribution. Recall that the random number generator we are using produces numbers that are uniformly distributed.

 © 2000 -Gilberto Urroz - All rights reserved

To check the equation of the straight line, first, press [CANCL][VAR] to return to normal display, then, press [VAR] and press [NXT] until you find a white button labeled [EQ], press that button. For my case, I get the value: '-7.103..E-13…+ .929166…*X', which is approximately, Y = 0.929*X. For full normality we expect to get Y = X. The data I generated, however, is very close to normal.

For additional exercises, generate data using the programs of subdirectory CSIMUL, then bring the data lists to the CHCKN sub-directory.

 © 2000 –Gilberto Urroz – All rights reserved

13.0 $\bar{x}$SIM: programs to SIMulate distributions of the mean ($\bar{x}$)

The purpose of the programs in this directory is to generate a list of mean values corresponding to a given distribution and verify that the mean values follow approximately a normal distribution. (This property is known as the *Central Limit Theorem*). These programs are used for this specific demonstration and will not be used very frequently. However, they illustrate the point that you can use your calculator to demonstrate certain properties of the distribution of $\bar{x}$ by generating synthetic data and analyzing them with the calculator itself. The only limitation is the speed of the calculator.

There are two programs in this directory used to generate lists of the mean values ($\bar{x}$) of sets of 10 data points that follow either a discrete distribution (GEND) or a normal distribution (GENN). To operate the programs enter the number (*n*) of $\bar{x}$ values that you want, then press the appropriate white key (either GEND or GENN). These programs require that you have the sub-directories SIMUL, SIMH and SIMUC available under directory STATS. The programs GENN and GEND work by generating *n* sets of 10 synthetic data points that follow either the discrete distribution set up in SIMH or the normal distribution obtained from SIMUC. The programs then calculate the mean values of each set of 10 points and display the list of $\bar{x}$ values in level 1 of the calculator's display.

The list of $\bar{x}$ values generated by either GEND or GENN can be analyzed using sub-directory ONEVAR\HISTO. You will find that the histogram of the $\bar{x}$ distribution resembles that of a normal distribution. You can also analyze the list of $\bar{x}$ values with sub-directory CHKN to verify that the mean values are normally distributed.

Please be aware that the larger the value of n the longer the time it takes for the calculator to produce the requested list. For example, to generate 50 values of $\bar{x}$ it may take the calculator up to 5 minutes.

13.1 $\bar{x}$ data from a discrete probability distribution

When generating data from the discrete distribution (by using GEND), you need to set up the appropriate parameters in sub-directory SIMH, i.e., fill in the variables *tbx* and *bound* in that sub-directory, before pressing GEND in sub-directory $\bar{x}$SIM. For example, move to sub-directory SIMH and enter the following values: *tbx* = {0 1 2 3}, and *bound* = {0 287 757 958 1000}. (These are the same values described in section 6, part II, of this guide.) Then, return to sub-directory $\bar{x}$SIM. To generate a list of 10 $\bar{x}$ values, enter:

[1][0][GEND]

It takes the calculator about 1 minute to generate that list. If you try this in your calculator you will get a list different than mine, since the synthetic data is generated through random numbers. This is the results I got:

```
{0.8 0.7 1.4 1.3 0.6 0.7 1.3 0.9 1.1 0.7}
```

They represent mean values of 10 different sets of values that follow the discrete probability distribution set up in sub-directory SIMH.

 © 2000 -Gilberto Urroz - All rights reserved

13.2 $\bar{x}$ data from a Normal distribution

When generating data from the Normal distribution (by using GENN), you need to set up the appropriate parameters in sub-directory SIMUC, i.e., fill in the variables α and β in that sub-directory, before pressing GENN in sub-directory $\bar{x}$SIM. For example, move to sub-directory SIMUC and enter the following values: α = 12.5 and β = 2.5. (These values represent the mean, $\alpha = \mu$, and the standard deviation, $\beta = \sigma$, of the Normal distribution, as described in section 9, part II, of this guide.) Then, return to sub-directory $\bar{x}$SIM. To generate a list of 10 $\bar{x}$ values, enter:

[1][0][GENN]

It takes the calculator about 40 seconds to generate that list. Again, the list you get from your calculator will differ from mine. This is the results I got:

```
{13.53 13.77 10.69 12.42 13.01 12.60 12.64 13.90 13.41 13.22}
```

They represent mean values of 10 different sets of values that follow the Normal distribution set up in sub-directory SIMUC.

13.3 Proposed exercises

Exercise 1. This exercise consists on:
(1) generating 50 values of $\bar{x}$ corresponding to data that follow the **discrete uniform distribution**

f(x) = 1/10, for x = 0, 1, 2, …, 9;

(2) generating a histogram of the data using the classes: 2.0-2.9, 3.0-3.9, …, 6.0-6.9.

Here are the steps to use in your calculator (total time for completing the exercise may be about 10 minutes):

1. Move to sub-directory SIMH, and enter the values:
   tbx = {0 1 2 3 4 5 6 7 8 9}, *bound* = {0 99 199 299 399 499 599 699 799 899 1000}.
2. Move to sub-directory $\bar{x}$SIM, and press: [5][0][GEND]. Then, wait about 5 to 6 minutes.
3. Keep the list in level 1 and move to sub-directory ONEVAR.
4. Enter the list in level 1 into variable x: [↰][X]
5. Press [DESC]
6. Press [HISTO]
7. Enter: [2][↰][XBMI] [1][↰][BWIDT] [5][↰][NBIN] [→FRQ].
8. Press [→HIST] to plot the histogram. Is the histogram symmetrically shaped? Does it resemble the histogram of a normal distribution? Press [CANCL] to return to normal display.
9. Press [NXT][↑ONE] [X] to get a copy of the data values in level 1 of the display.
10. Move to sub-directory CHKN, and enter the list into variable x: [↰][X]
11. Press [→ΣDA] and wait about 1 minute until the message "Ready" is shown.
12. Press [→PLT]. Does the scatterplot show a good linear trend?
13. Press [STATL]. Dose the line plotted fits most of the data points in the graph?
14. Press [CANCL] [EQ]. Is the equation shown close to `'0.0 + 1.0*X'` ?

© 2000 -Gilberto Urroz - All rights reserved

Exercise 2. This exercise consists on: (1) generating 50 values of $\bar{x}$ corresponding to data that follow a Normal distribution with $\mu = 50$, $\sigma = 20$; (2) generating a histogram of the data using the classes: 2.0-2.9, 3.0-3.9, ..., 6.0-6.9. Here are the steps to use in your calculator (total time for completing the exercise may be about 6 minutes):

1. Move to sub-directory SIMUC, and enter the values: $\alpha = 50$, $\beta = 20$.
2. Move to sub-directory $\bar{x}$SIM, and press: [5][0][GENN]. Then, wait about 3 to 4 minutes.
3. Repeat steps 3 through 14 of Exercise 1. Except that you need to use suitable values for XBMIN, BWIDTH, and NBIN in step 7. Choose those values depending on the minimum and maximum value of your data. I suggest using XBMIN = 20, BWIDTH = 5, NBIN = 10.

The results of Exercises 1 and 2 should serve as evidence that the distribution of mean values follow the Normal distribution regardless of the original distribution of the data. You can try these exercises with different parameters (100 points instead of 50, different discrete distributions, different Normal parameters) to further convince you of that fact.

 © 2000 -Gilberto Urroz - All rights reserved

14.0 HYPTS - HYPothesis TeSting

This sub-directory includes four sub-directories that allow you to perform hypothesis testing on one mean, two means, one variance, and two variances. HYPTS also includes one sub-directory with data for use with the four hypothesis testing sub-directories. The operation of each of the five sub-directories is described below.

14.1 $\bar{x}$TST1 - Hypothesis testing on one mean

The basic ideas for hypothesis testing on one mean were presented in section 12.1 in part I of this guide. The procedures described in that section are programmed in this sub-directory. The following are the instructions on how to use the sub-directory:

- Press [*INFO→*] to get brief instructions in the use of this sub-directory. Press [*OK*] to return to normal display.

- There are two possibilities for hypothesis testing on one mean allowed in this subdirectory:

(1) If a sample of data points is known use the option [*XLIST*].
(2) If the sample statistics ($\bar{x}$, s, n) are known, use the option [*$\bar{x}$SN*].

- **Option** [*XLIST*]:

1. Press [*XLIST*] to get instructions for the hypothesis testing programs when you have a list of values, x, representing a sample. . Press [*OK*] to return to normal display.
2. Enter the sample list $\{x_1\ x_2\ ...\ x_n\}$ into variable [*x*] by using [↰][*X*].
3. Enter the level of significance for the test, α, by using [↰][α].
4. Enter the value of the population mean, μ_o, to be tested (the null hypothesis is H_o: $\mu = \mu_o$) by using [↰][*$\mu 0$*].
5. Enter the value of the population standard deviation, σ, if known, by using [↰][σ]. If the population standard deviation is not known enter a negative value into σ, for example: [1][+/−] [↰][σ].
6. For a one-sided or one-tailed test press [*ONET*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \mu_o$.
7. Press [OK] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); the population standard deviation (σ); the sample standard deviation (s); the number of data points in the sample (n); and, the mean value of the sample ($\bar{x}$). Use the [⇦] key to drop values from the display.
8. For a two-sided or two-tailed test press [*TWOT*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \mu_o$.
9. Press [OK] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); the population standard deviation (σ); the sample standard deviation (s); the number of data points in the sample (n); and, the mean value of the sample ($\bar{x}$). Use the [⇦] key to drop values from the display.
10. Press [VAR] to return to original menu.

 © 2000 -Gilberto Urroz - All rights reserved

- **Option [$\bar{x}$SN]:**

1. Press [$\bar{x}$SN] to get instructions for the hypothesis testing programs when you know the statistics of the sample. Press [*OK*] to return to normal display.
2. Enter the level of significance for the test, α, by using [↰][*α*].
3. Enter the value of the population mean, μ_o, to be tested (the null hypothesis is H_o: $\mu = \mu_o$) by using [↰][*μ0*].
4. Enter the value of the population standard deviation, σ, if known, by using [↰][*σ*]. If the population standard deviation is not known enter a negative value into σ, for example: [1][+/–] [↰][*σ*].
5. Enter the value of the sample mean, $\bar{x}$, by using: [↰][$\bar{x}$].
6. Enter the value of the sample standard deviation, s, by using: [↰][*s*].
7. Enter the value of the sample size, n, by using: [↰][*n*].
8. Press [NXT] to access the next variable menu needed for the hypothesis testing procedure.
9. For a one-sided or one-tailed test press [*ONET*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \mu_o$.
10. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); and, the population standard deviation (σ).Use the [⇦] key to drop values from the display.
11. For a two-sided or two-tailed test press [*TWOT*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \mu_o$.
12. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); and, the population standard deviation (σ). Use the [⇦] key to drop values from the display.
13. Press [VAR] to return to original menu.

- Press [↑*UP*] to move to the upper sub-directory HYPTS.

- Press [↑*HDA*] to move to the HDATA sub-directory containing data for hypothesis testing applications.

Examples:

14.1.1 At a level of significance of 0.01, test the null hypothesis: H_o: $\mu = 2.3$, for a population having a standard deviation $\sigma = 1.50$. In order to test the hypothesis use the following sample, x = { 3.2 4.5 5.3 2.3 6.1 4.2}. Use as the alternative hypothesis, H_1: $\mu > 2.3$.

<u>*Solution*</u>: Within sub-directory $\bar{x}$TST1 use the following keystrokes:

[*XLIST*] [*OK*]	Select the XLIST option.
{3.2 4.5 5.3 2.3 6.1 4.2} [ENTER][↰][*X*]	Enter x list.
0.01 [↰][*α*]	Enter level of significance.
2.3 [↰][*μ0*]	Enter population mean to be tested (μ0)
1.5 [↰][*σ*]	Enter known population standard deviation.
[*ONET*]	For one-sided test.

 © 2000 -Gilberto Urroz - All rights reserved

The result is: *One-Tail, z - test, Reject H0:* $\mu = 2.3$. Press [OK], and get the following additional results: $\alpha = 0.01$, *P-value = 6.6E-4*, $z = 3.2115$, $\sigma = 1.5$, $s = 1.377$, $n = 6$, $\bar{x} = 4.266$.

14.1.2 Repeat example 14.1.1 using the alternative hypothesis, H_1: $\mu \neq 2.3$

Solution: The input data is exactly the same as in example 14.1.1, however, for a two-tailed test we use the [TWOT] program. The result is the same as above, *Reject H0:* $\mu = 2.3$. Also, $\alpha = 0.01$, *P-value = 1.32E-3*, $z = 3.2115$, $\sigma = 1.5$, $s = 1.377$, $n = 6$, $\bar{x} = 4.266$.

14.1.3 Repeat example 14.1.1 for $\mu_o = 4.0$, assuming that the population standard deviation is unknown.

Solution: Change the values of the parameters μ_o and σ, as follows:

4 [↰][*μ0*]	Enter population mean to be tested (μ0)
-1 [↰][*σ*]	For unknown population standard deviation enter -1.

Press [*ONET*], for one-sided test. The result is: *One-Tail, t - test, Do not reject H0:* $\mu = 4$. Also, $\alpha = 0.01$, *P-value = 0.3277*, $t = 0.47405$, $\nu = 5$, $\sigma = -1$, $s = 1.377$, $n = 6$, $\bar{x} = 4.266$.

14.1.4 At a level of significance of 0.05, test the null hypothesis: H_o: $\mu = 4.0$, for a population having a standard deviation $\sigma = 1.50$. In order to test the hypothesis a sample of 50 elements is used having a mean value $\bar{x} = 2.5$, and a standard deviation s = 0.5. Use as the alternative hypothesis, H_1: $\mu > 2.3$.

Solution: Within sub-directory $\bar{x}$TST1 use the following keystrokes:

[*x̄SN*][*OK*]	Select the $\bar{x}$SN option.
0.05 [↰][*α*]	Enter level of significance.
4.0 [↰][*μ0*]	Enter population mean to be tested (μ0)
1.5 [↰][*σ*]	Enter known population standard deviation.
2.5 [↰][*x̄*]	Enter sample mean.
0.5 [↰][*s*]	Enter sample standard deviation.
50 [↰][*n*]	Enter sample size.
[NXT][*ONET*]	For one-sided test.

The result is: *One-Tail, z - test,Reject H0:* $\mu = 4$. Press [OK], and get the following additional results: $\alpha = 0.05$, *P-value = 7.687E-13*, $z = -7.07$, $\sigma = 1.5$.

14.1.5 Repeat example 14.1.4. if the population standard deviation is unknown.

Solution: Replace the current value of s with a negative number, say, -1,0, i.e.,

[↰][PREV]	Move to previous menu.
[1][+/-][↰][*σ*]	Enter s = -1.
[NXT][*ONET*]	For one-sided test.

The result is: *One-Tail, z - test, Reject H0:* $\mu = 4$. Press [OK], and get the following additional results: $\alpha = 0.05$, *P-value = 3.606E-100* ≈ 0, $z = -21.21$, $\sigma = -1$. Notice that the calculator still performed a z-test because the sample size is relatively large (n =

 © 2000 -Gilberto Urroz - All rights reserved

50 > 30), however, the z value changed because the value of s = 0.5 is used to define the z parameter rather than the value of σ = 1.5 used in example 14.1.4.

14.1.6 Repeat example 14.1.5 if the sample size is 16 elements only.

Solution: Replace the current value of s with a negative number, say, -1,0, i.e.,

[↰][PREV]	Move to previous menu.
16 [↰][*n*]	Enter n = 16.
[NXT][*ONET*]	For one-sided test.

The result is: *One-Tail, t - test, Reject H0:* $\mu = 4$. Also, $\alpha = 0.05$, *P-value = 2.16136E-9*, $t = -12$, $\nu = 15$, $\sigma = -1$. For this case, since n = 16 < 30 and σ is unknown, a t-test is performed.

14.1.7 If we were to use a value of σ = 1.5, as in example, 14.1.4, but with n = 16, the calculator would perform a z-test, with *P-value = 3.1671E-5, and z = -4.0.* The calculator would still recommend rejecting H0.

14.1.8 Try changing the value of μ_o to 1.0 and try a one-tailed test (with σ = 1.5, n = 16) to see the recommendation of rejecting H_o: μ = 1, with *P-value = 3.167E-5*, and *z = 4*.

14.2. $\overline{x}$TST2 - Hypothesis testing on two means

The basic ideas for hypothesis testing two means were presented in section 12.2 in part I of this guide. The procedures described in that section are programmed in this sub-directory. The following are the instructions on how to use the sub-directory:

- Press [*INFO→*] to get brief instructions in the use of this sub-directory. Press [*OK*] to return to normal display.

- There are three possibilities for hypothesis testing on one mean allowed in this subdirectory:

1) If there are two samples consisting of two lists of data points use the option [*XLIST*].
2) If statistics of the two samples ($\overline{x}_1$, s_1, n_1, $\overline{x}_2$, s_2, n_2) are known, use the option [*$\overline{x}$SN*].
3) If there are two paired samples use the option [*XPAIR*]. This option calculates the difference in corresponding values of the two samples creating a list of values that is analyzed using sub-directory $\overline{x}$TST1. The data currently stored in that sub-directory will be replaced by that calculated in XPAIR.

- **Option** [*XLIST*]:

1. Press [*XLIST*] to get instructions for the hypothesis testing programs when you have two lists of values, x1 and x2, representing two samples. Press [*OK*] to return to normal display.
2. Enter the sample list $\mathbf{x}_1$ into variable [*x1*] by using [↰][*x1*].
3. Enter the sample list $\mathbf{x}_2$ into variable [*x2*] by using [↰][*x2*].
4. Enter the level of significance for the test, α, by using [↰][α].
5. Enter the difference of the population means, δ, to be tested (the null hypothesis is H_o: $\mu_1 - \mu_2 = \delta$) by using [↰][δ].
6. Enter the value of the population standard deviations, σ_1 and σ_2, if known, by using [↰][$\sigma 1$] and [↰][$\sigma 2$]. If the two samples come from the same population use $\sigma_1 = \sigma_2$. If one (or both) of the population standard deviations is not known enter a negative value

 © 2000 -Gilberto Urroz - All rights reserved

into the unknown variable, for example: [1][+/–] [↰][$\sigma 1$] (and/or [1][+/–] [↰][$\sigma 2$]). If you suspect that the samples come from different populations but you don't know the values of σ_1 or σ_2, enter two different negative values for the two population standard deviations, for example:

[1][+/–] [↰][$\sigma 1$], and [2][+/–] [↰][$\sigma 2$].

7. Press [NXT] to access the next menu required for hypothesis testing on two means.
8. For a one-sided or one-tailed test press [*ONET*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu_1 - \mu_2 = \delta$).
9. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); the populations standard deviations ($\sigma 1, \sigma 2$); the samples standard deviations (s1 and s2); the number of data points in each sample (n1 and n2); and, the mean values of the samples ($\bar{x}1$ and $\bar{x}2$). Use the [⇦] key to drop values from the display.
10. For a two-sided or two-tailed test press [*TWOT*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \mu_1 - \mu_2 = \delta$).
11. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); the populations standard deviations ($\sigma 1, \sigma 2$); the samples standard deviations (s1 and s2); the number of data points in each sample (n1 and n2); and, the mean values of the samples ($\bar{x}1$ and $\bar{x}2$). Use the [⇦] key to drop values from the display.
12. Press [VAR] to return to original menu.

Option [*$\bar{x}$SN*]:

1. Press [*$\bar{x}$SN*] to get instructions for the hypothesis testing programs when you know the statistics of the samples. Press [*OK*] to return to normal display.
2. Enter the level of significance for the test, α, by using [↰][α].
3. Enter the difference of the population means, δ, to be tested (the null hypothesis is H_o: $\mu_1 - \mu_2 = \delta$) by using [↰][δ].
4. Enter the value of the population standard deviations, σ_1 and σ_2, if known, by using [↰][$\sigma 1$] and [↰][$\sigma 2$]. If the two samples come from the same population use $\sigma_1 = \sigma_2$. If one (or both) of the population standard deviations is not known enter a negative value into the unknown variable, for example: [1][+/–] [↰][$\sigma 1$] (and/or [1][+/–] [↰][$\sigma 2$]). If you suspect that the samples come from different populations but you don't know the values of σ_1 or σ_2, enter two different negative values for the two population standard deviations, for example:

   [1][+/–] [↰][$\sigma 1$], and [2][+/–] [↰][$\sigma 2$].
5. Enter the values of the samples means, $\bar{x}_1$ and $\bar{x}_2$, by using: [↰][$\bar{x}1$] and [↰][$\bar{x}2$].
6. Enter the values of the samples standard deviations, s_1 and s_2, by using: [↰][*s1*] and [↰][*s2*].
7. Enter the values of the sample sizes, n_1 and n_2, by using: [↰][*n1*] and [↰][*n2*].
8. For a one-sided or one-tailed test press [*ONET*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu_1 - \mu_2 = \delta$).
9. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees

© 2000 -Gilberto Urroz - All rights reserved

of freedom if a t test is used (ν); and, the populations standard deviations (σ1, σ2). Use the [⇦] key to drop values from the display.
10. For a two-sided or two-tailed test press [*TWOT*]. A message box identifies the type of test (one or two tails, z or t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu_1 - \mu_2 = \delta$).
11. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding z or t parameter; the degrees of freedom if a t test is used (ν); and, the populations standard deviations (σ1, σ2). Use the [⇦] key to drop values from the display.
12. Press [VAR] to return to original menu.

Option [*XPAIR*]:

1. Press [*XLIST*] to get instructions for the hypothesis testing programs when you have two lists of values, $\mathbf{x}_1$ and $\mathbf{x}_2$, representing two paired samples. It is assumed that the two samples come from the same population, and that the population standard deviation is unknown. Press [*OK*] to return to normal display.
2. Enter the sample list $\mathbf{x}_1$ into variable [*x1*] by using [↰][*X1*].
3. Enter the sample list $\mathbf{x}_2$ into variable [*x2*] by using [↰][*X2*].
4. Enter the level of significance for the test, α, by using [↰][*α*].

NOTE: In this test, we create a list of differences $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$, and test the null hypothesis, H_o: $\mu_{\Delta x} = \delta$.

5. Enter the population mean of the differences, $\mu_{\Delta x} = \delta$, to be tested by using [↰][*δ*].
6. For a one-sided or one-tailed test press [*ONET*]. A message box identifies the type of test (one or two tails, t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \delta$).
7. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding t parameter; the degrees of freedom for t (ν); *σ: -1* (required for using $\bar{x}$TST1, just ignore this output); the standard deviation of the differences ($s = s_{\Delta x}$); the number of data points in each sample (n); and, the mean values of the differences ($\bar{x} = \bar{x}_{\Delta x}$). Use the [⇦] key to drop values from the display.
8. For a two-sided or two-tailed test press [*TWOT*]. A message box identifies the type of test (one or two tails, t test) and gives the recommendation on whether or not to reject the null hypothesis, H_o: $\mu = \delta$).
9. Press [*OK*] to return to normal display. Shown in the display will be: the level of significance (α); the P-value for the test; the corresponding t parameter; the degrees of freedom for t (ν); *σ: -1* (required for using $\bar{x}$TST1, just ignore this output); the standard deviation of the differences ($s = s_{\Delta x}$); the number of data points in each sample (n); and, the mean values of the differences ($\bar{x} = \bar{x}_{\Delta x}$). Use the [⇦] key to drop values from the display.
10. Press [VAR] to return to original menu.

Press [↑*UP*] to move to the upper sub-directory HYPTS.

Press [↑*HDA*] to move to the HDATA sub-directory containing data for hypothesis testing applications.

Examples:

14.2.1. At a level of significance of 0.01, test the null hypothesis: H_0: $\mu_1 - \mu_2 = 2.0$, for two populations 1 and 2 having the same standard deviation $\sigma_1 = \sigma_2 = 1.50$. In order to test the hypothesis use the following samples taken from the corresponding populations, x_1 = { 3.2 4.5 5.3 2.3 6.1 4.2}, and x_2 = {1.5 6.2 4.8 3.6 5.2}. Use as the alternative hypothesis, H_1: $\mu_1 - \mu_2 \neq 2$.

 © 2000 -Gilberto Urroz - All rights reserved

Solution: Within sub-directory $\bar{x}$TST2 use the following keystrokes:

[*XLIST*] [*OK*]	Select the XLIST option.
{3.2 4.5 5.3 2.3 6.1 4.2} [ENTER][↰][*X1*]	Enter x_1 list.
{1.5 6.2 4.8 3.6 5.2 } [ENTER][↰][*X2*]	Enter x_2 list.
0.01 [↰][α]	Enter level of significance.
2 [↰][δ]	Enter population mean difference.
1.5 [↰][$\sigma 1$]	Enter known σ for population 1.
1.5 [↰][$\sigma 2$]	Enter known σ for population 2.
[NXT][*TWOT*]	For a two-sided test.

The result is: *Two-Tail, z - test, Do not reject H0:* $\mu_1 - \mu_2 = 2$. Press [OK], and get the following additional results: $\alpha = 0.01$, *P-value* = *2.819E-2*, $z = -2.1945$, $\sigma 2 = 1.5$, $\sigma 1 = 1.5$, $s2 = 1.802$, $\overline{x2} = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\overline{x1} = 4.266$, $n1 = 6$.

14.2.2. Repeat example 14.2.1 using the alternative hypothesis, $H_1: \mu_1 - \mu_2 > 2$.

Solution: The input data is exactly the same as in example 14.1.1, however, for a one-tailed test we use the [*ONET*] program. The result is the same as above, *Do not reject H0*: $\mu_1 - \mu_2 = 2$. Also, $\alpha = 0.01$, *P-value* = *1.409E-2*, $z = -2.19458$, $\sigma 2 = 1.5$, $\sigma 1 = 1.5$, $s2 = 1.802$, $\overline{x2} = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\overline{x1} = 4.266$, $n1 = 6$.

14.2.3. Repeat example 14.1.1 for $\delta = 4.0$, assuming that the populations standard deviations are unknown, but the same.

Solution: Change the values of the parameters δ and σ's, as follows:

[↰][PREV]	Move to previous menu.
4 [↰][δ]	Enter population mean difference to be tested (δ).
-1 [↰][$\sigma 1$]	For unknown population standard deviation enter -1.
-1 [↰][$\sigma 2$]	For unknown population standard deviation enter -1.

Press [NXT][*TWOT*], for one-sided test. The result is: *Two-Tail, t - test, Reject H0:* $\mu_1 - \mu_2 = 4$. Also, $\alpha = 0.01$, *P-value* = *2.403E-2*, $t = -4.1722$, $v = 9$, $\sigma 2 = -1$, $\sigma 1 = -1$, $s2 = 1.802$, $\overline{x2} = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\overline{x1} = 4.266$, $n1 = 6$.

14.2.4. Repeat example 14.1.1 for $\delta = 4.0$, assuming that the populations standard deviations are unknown, but we suspect they're different.

Solution: Change the values of the parameters μ_o and σ, as follows:

[↰][PREV]	Move to previous menu.
-2 [↰][$\sigma 2$]	Use $\sigma_2 = -2$, to make it different than $\sigma_1 = -1$.

Press [NXT][*TWOT*], for one-sided test. The result is: *Two-Tail, t - test, Do not reject H0:* $\mu_1 - \mu_2 = 4$. Also, $\alpha = 0.01$, *P-value* =*1.400E-2*, $t = -4.1722$, $v = 4$, $\sigma 2 = -1$, $\sigma 1 = -1$, $s2 = 1.802$, $\overline{x2} = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\overline{x1} = 4.266$, $n1 = 6$.

 © 2000 -Gilberto Urroz - All rights reserved

Note: When we suspect that the populations from where the samples were taken have different standard deviations, the degrees of freedom for the t-test are calculated using the following formula:

$$v = \frac{(\frac{\sigma_1}{n_1})^2 + (\frac{\sigma_2}{n_2})^2}{\frac{(\frac{\sigma_1}{n_1})^2}{n_1 - 1} + \frac{(\frac{\sigma_2}{n_2})^2}{n_2 - 1}}$$

This test is referred to as the *Smith-Sathertwaite test.*
If the two samples come from populations having the same standard deviation then,

$$v = n_1 + n_2 - 2.$$

For example, in example 14.2.4, which uses a Smith-Sathertwaite test, we found ν = 4, while in example 14.2.3, which has basically the same data as 14.2.4, but with the same unknown standard deviation for the populations we found ν = 9. For the same t parameter, t = -4.1722, the two examples produce opposite results ("Do not reject Ho" in 14.2.4 as opposite to "reject Ho" in 14.2.3).

14.2.5. At a level of significance of 0.05, test the null hypothesis: H_o: $\mu_1 - \mu_2 = 4.0$, for two populations (1 and 2) having standard deviations $\sigma_1 = \sigma_2 = 1.50$. In order to test the hypothesis a sample of 50 elements, having a mean value $\bar{x}_1 = 2.5$, and a standard deviation $s_1 = 0.5$, is taken from population 1. Meanwhile, a sample of 35 elements, having a mean $\bar{x}_1 = 3.5$, and a standard deviation $s_2 = 1.5$ is taken from population 2. Use as the alternative hypothesis, H_1: $\mu_1 - \mu_2 > 4.0$.

Solution: Within sub-directory $\bar{x}$TST1 use the following keystrokes:

[$\bar{x}$*SN*][*OK*]	Select the $\bar{x}$SN option.
0.05 [↰][α]	Enter level of significance.
4.0 [↰][δ]	Enter population mean to be tested (μ0)
1.5 [↰][*σ1*]	Enter standard deviation for population 1.
1.5 [↰][*σ2*]	Enter standard deviation for population 2.
2.5 [↰][$\bar{x}$*1*]	Enter mean for sample 1.
3.5 [↰][$\bar{x}$*2*]	Enter mean for sample 2.
[NXT]	To access the next menu.
0.5 [↰][*s1*]	Enter standard deviation for sample 1.
1.5 [↰][*s2*]	Enter standard deviation for sample 2.
50 [↰][*n1*]	Enter size for sample 1.
35 [↰][*n1*]	Enter size for sample 2.
[*ONET*]	For one-sided test.

The result is: *One-Tail, z - test, Reject H0:* $\mu_1 - \mu_2 = 4$. Press [OK], and get the following additional results: $\alpha = 0.05$, *P-value* = 5.5608E-52 ≈ 0, z = -15.124, σ2 = 1.5, σ1 = 1.5.

 © 2000 -Gilberto Urroz - All rights reserved

14.2.6. Repeat example 14.2.5. if the population's standard deviations are unknown, but suspected to be the same.

Solution: Replace the current value of σ_1 and σ_2 with the same negative number, say, -1,0, i.e.,

[↰][PREV]	Move to previous menu.
[1][+/-][ENTER][ENTER]	Place two copies of -1 in the display.
[↰][$\sigma 1$]	Enter σ_1 = -1.
[↰][$\sigma 2$]	Enter σ_2 = -1.
[*NXT*][*ONET*]	For one-sided test.

The result is: *One-Tail, z - test, Reject H0:* $\mu_1 - \mu_2 = 4$. Press [OK], and get the following additional results: $\alpha = 0.05$, *P-value* $= 9.3104E\text{-}81 \approx 0$, $z = -18.995$, $\sigma 2 = -1$, $\sigma 1 = -1$. Notice that the calculator still performed a z-test because the sample sizes are relatively large (n_1 = 50 > 30, and n_2 = 35 > 30). However, the z value changed because the values of s_1 = 0.5 and s_2 = 1.5 are used to define the z parameter rather than the value of σ = 1.5 used in example 14.2.5.

14.2.7. Repeat example 14.2.6 if the sample sizes are n_1 = 16 and n_2 = 35 elements.

Solution: Replace the current value of n_1 with 16, keeping the existing value of n_2 = 35.

16 [↰][*n1*]	Enter n = 16.
[*NXT*][*ONET*]	For one-sided test.

The result is: *One-Tail, t - test, Reject H0:* $\mu_1 - \mu_2 = 4$. Also, $\alpha = 0.05$, *P-value* $= 9.825E\text{-}18$, $t = -12.94$, $\nu = 49$, $\sigma 1 = \sigma 2 = -1$. For this case, since n1 = 16 < 30 and the σ's are unknown, a t-test is performed.

14.2.8. If we were to use a value of $\sigma_1 = \sigma_2$ = 1.5, as in example, 14.2.5, but with n_1 = 16, the calculator would perform a z-test, with *P-value = 1.1515E-28, and z = -11.0455*. The calculator would still recommend not rejecting H0.

Try changing the value of μ_o to 1.0 and try a one-tailed test (with σ = 1.5, n = 16) to see the recommendation of rejecting H_o: μ = 1, with *P-value = 3.167E-5*, and *z = 4*.

14.3 σTS1 - Hypothesis testing on one variance

The basic ideas for hypothesis testing on one mean were presented in section 12.3 in part I of this guide. The procedures described in that section are programmed in this sub-directory. The following are the instructions on how to use the sub-directory:

- Press [*INFO→*] to get brief instructions in the use of this sub-directory. Press [*OK*] to return to normal display.

- There are two possibilities for hypothesis testing on one mean allowed in this subdirectory:

(1) If a sample of data points from a normal population is known use the option [*XLIST*].
(2) If the sample statistics (s, n) are known, use the option [*SDATA*].

 © 2000 -Gilberto Urroz - All rights reserved

Option [*XLIST*]:

1. Press [*XLIST*] to get instructions for the hypothesis testing programs when you have a list of values, x, representing a sample. . Press [*OK*] to return to normal display.
2. Enter the sample list $\{x_1\ x_2\ \ldots\ x_n\}$ into variable [*x*] by using [↰][*X*].
3. Enter the level of significance for the test, α, by using [↰][*α*].
4. Enter the value of the population standard deviation, σ_o, to be tested (the null hypothesis is H_o: $\sigma^2=\sigma_o^2$) by using [↰][*σ0*].
5. Press [*TσLT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2<\sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
6. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); the degrees of freedom for χ^2 (ν); the sample standard deviation (s); and, the number of data points in the sample (n). Use the [⇦] key to drop values from the display.
7. Press [*TσGT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2>\sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
8. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); the degrees of freedom for χ^2 (ν); the sample standard deviation (s); and, the number of data points in the sample (n). Use the [⇦] key to drop values from the display.
9. Press [*TσNE*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2\neq\sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
10. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); the degrees of freedom for χ^2 (ν); the sample standard deviation (s); and, the number of data points in the sample (n). Use the [⇦] key to drop values from the display.
11. Press [VAR] to return to original menu.

Option [*SDATA*]:

1. Press [*SDATA*] to get instructions for the hypothesis testing programs when you know the statistics of the sample. Press [*OK*] to return to normal display.
2. Enter the level of significance for the test, α, by using [↰][*α*].
3. Enter the value of the population standard deviation, σ_o, to be tested (the null hypothesis is H_o: $\sigma^2=\sigma_o^2$) by using [↰][*σ0*].
4. Enter the value of the sample size, n, by using [↰][*n*].
5. Enter the value of the sample standard deviation, s, by using: [↰][*s*].

12. Press [*TσLT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2<\sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
13. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); and, the degrees of freedom for χ^2 (ν). Use the [⇦] key to drop values from the display.
14. Press [*TσGT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2>\sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
15. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); and, the degrees of freedom for χ^2 (ν). Use the [⇦] key to drop values from the display.

 © 2000 -Gilberto Urroz - All rights reserved

16. Press [*TσNE*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma^2 \neq \sigma_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
17. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding chi-square parameter (X^2); and, the degrees of freedom for χ^2 (ν). Use the [⇦] key to drop values from the display.
18. Press [VAR] to return to original menu.

- Press [↑*UP*] to move to the upper sub-directory HYPTS.

- Press [↑*HDA*] to move to the HDATA sub-directory containing data for hypothesis testing applications.

14.4 σTS2 - Hypothesis testing on two variances

The basic ideas for hypothesis testing on one mean were presented in section 12.4 in part I of this guide. The procedures described in that section are programmed in this sub-directory. The following are the instructions on how to use the sub-directory:

- Press [*INFO*→] to get brief instructions in the use of this sub-directory. Press [*OK*] to return to normal display.

- There are two possibilities for hypothesis testing on one mean allowed in this subdirectory:

(1) If a sample of data points is known use the option [*XLIST*].
(2) If the sample statistics ($\overline{s}$, n) are known, use the option [*SNDA*].

- **Option** [*XLIST*]:

1. Press [*XLIST*] to get instructions for the hypothesis testing programs when you have two lists of values, $\mathbf{x}_1$ and $\mathbf{x}_2$, representing two samples from the same normal population. Press [*OK*] to return to normal display.
2. Enter the sample list $\mathbf{x}_1$ into variable [*x1*] by using [↰][*x1*].
3. Enter the sample list $\mathbf{x}_2$ into variable [*x2*] by using [↰][*x2*].
4. Enter the level of significance for the test, α, by using [↰][α].
5. Press [*TσLT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 < \sigma_2^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
6. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding F parameter; the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom); the samples standard deviations (s1 and s2); and, the number of data points in each sample (n1 and n2). Use the [⇦] key to drop values from the display.
7. Press [*TσGT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 > \sigma_2^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
8. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding F parameter; the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom); the samples standard deviations (s1 and s2); and, the number of data points in each sample (n1 and n2). Use the [⇦] key to drop values from the display.

 © 2000 -Gilberto Urroz - All rights reserved

9. Press [*TσNE*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 \neq \sigma 2_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
10. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding F parameter; the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom); the samples standard deviations (s1 and s2); and, the number of data points in each sample (n1 and n2). Use the [⇦] key to drop values from the display.
11. Press [VAR] to return to original menu.

   - **Option** [*SNDA*]:

1. Press [*SNDA*] to get instructions for the hypothesis testing programs when you know the statistics of the sample. Press [*OK*] to return to normal display.
2. Enter the level of significance for the test, α, by using [↰][*α*].
3. Enter the values of the samples standard deviations, s_1 and s_2, by using: [↰][*s1*] and [↰][*s2*].
4. Enter the values of the sample sizes, n_1 and n_2, by using: [↰][*n1*] and [↰][*n2*].
12. Press [*TσLT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 < \sigma_2^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
13. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding F parameter; and, the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom). Use the [⇦] key to drop values from the display.
14. Press [*TσGT*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 > \sigma_2^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
15. Press [OK] to return to normal display. Shown in the display will be: the P-value for the test; the corresponding F parameter; and, the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom). Use the [⇦] key to drop values from the display.
16. Press [*TσNE*] to test the null hypothesis against the alternative hypothesis, H_1: $\sigma_1^2 \neq \sigma 2_o^2$. A message box identifies the alternative hypothesis and gives the recommendation on whether or not to reject the null hypothesis.
17. Press [OK] to return to normal display Shown in the display will be: the P-value for the test; the corresponding F parameter; and, the degrees of freedom for F (νD = denominator degrees of freedom, and νN = numerator degrees of freedom). Use the [⇦] key to drop values from the display.
18. Press [VAR] to return to original menu.

   - Press [↑*UP*] to move to the upper sub-directory HYPTS.

   - Press [↑*HDA*] to move to the HDATA sub-directory containing data for hypothesis testing applications.

Examples:

14.5. HDAT - Data for hypothesis testing programs

This sub-directory contains lists of data to be used with one or more of the four sub-directories contained in sub-directory HYPTS. A P followed by two numbers separated by a dot identifies the data. The first number indicates the chapter in the Textbook by Miller & Freund (all the

 © 2000 -Gilberto Urroz - All rights reserved

data is from Chapter 7), and the second number indicates the problem number. To use the data press the appropriate key, say *[P7.46]* for problem 7.46. This will show a list of tagged values. Move to the upper sub-directory by pressing [↑*UP*], and press [*$\bar{x}$TST1*] for one-mean hypothesis testing. Next, you need to decompose the data list into its individual components by using:

[PRG][*TYPE*][*OBJ→*]

The number in level 1 represents the number of element in the original list (4). Use the [⇦] key to drop that value from the display. Then, store the values of σ, μ0, α, and x in the corresponding variables by using the following keystrokes (the variables tag are automatically removed by the calculator):

[VAR][*XLIST*][*OK*] To get into the XLIST option.
[↰][*σ*] Store population standard deviation, σ. A value σ= -1, means σ is unknown.
[↰][*μ0*] Store value of population mean to be tested.
[↰][*α*] Store level of significance.
[↰][*X*] Store list of values x.
Now, press *[ONET]* or *[TWOT]* for one-sided or two-sided hypothesis testing. For example, [ONET] produces the following result:

```
One Tail. t - test
Do not reject
H0: μ = 29
```

Press [OK] and use the [⇦] key to check the following values in the main display:

α = 0.01, P-value = 0.9666, t = -2.3354, ν = 5, σ= -1 (i.e., unknown), s = 4.1952, n = 6, $\bar{x}$ = 25.

 © 2000 -Gilberto Urroz - All rights reserved

15.0 OPCR: programs for generating OPerating CuRves for hypothesis testing on the mean

In hypothesis testing we recognize a *Type I error*: rejecting the null hypothesis when it is true, and a *Type II error*: accepting the null hypothesis even though it is not true. The significance level α in a hypothesis testing process is also the probability of committing a Type I error. Associated with the Type II error is a probability referred to as β. The expression

$$\beta = L(\mu) = \text{probability of accepting the null hypothesis when } \mu \text{ prevails}$$

is known as the *operating curve*. In this section we present programs that produce one-sided and two-sided operating curves for hypothesis testing on the mean.

The problem under consideration is similar to that presented in section 12. 1 of Part I of this guide: suppose that we test the null hypothesis H_o: $\mu = \mu_o$, against the alternative hypothesis, H_1: $\mu \neq \mu_o$ (two-sided) or, H_1: $\mu > \mu_o$ or H_1: $\mu < \mu_o$ (one-sided) at a level of confidence $(1-\alpha)100\%$, or significance level α. Typically, the test is performed by using a sample of size n with a mean $\bar{x}$ and a standard deviation s. The sample is extracted from a population with a standard deviation σ. We want to find the probability of Type II error, β, as a function of μ. The procedure for obtaining β is presented in some detail in statistical textbooks. The general procedure for obtaining β given μ can be summarized in the following steps:
Given α, μ_o, σ, n, μ, find β. Let F(z) be the cumulative distribution function for the Standardized Normal distribution, i.e., Z is N(0,1). Also, let z_α be defined by

$$P(Z>z_\alpha) = \alpha = 1 - F(z_\alpha) = \text{UTPN}(0,1,z_\alpha).$$

First, find

$$d = |\mu - \mu_o|/\sigma.$$

Then, depending on the alternative hypothesis, find

$$\beta = F(|z_\alpha| - n^{1/2}\cdot d),$$

for a one-sided hypothesis, or

$$\beta = F(|z_\alpha| - n^{1/2}\cdot d) - F(-|z_\alpha| - n^{1/2}\cdot d),$$

for a two-sided hypothesis.

To develop an operating curve let $\mu = \mu_{str}, \mu_{str}+\Delta\mu, \ldots, \mu_{end}$, and calculate $\beta = L(\mu)$ following the procedure outlined above. Sub-directory OPCR has two programs that allow you to generate the data for one-sided or two-sided operating curves. The procedure to generate such curves is as follows:

1. Enter values of α, μ_o, σ, n, μ_{str}, μ_{end}, and $\Delta\mu$.
2. Press [→ΣD1] or [→ΣD2] to generate data for a one-sided or two-sided operating curve. In both cases, the calculator will display a message box indicating data are ready. Press [OK] to continue.
3. Press [ΣDAT] to see the data matrix. First column = μ, and second column = β.
4. Press [→PLT] to plot the operating curve as a scattergram.
   Press [CANCL] to return to normal display.
5. Press [→DRL] to draw a line through the points in the curve.
   Press [CANCL] to return to normal display.

 © 2000 -Gilberto Urroz - All rights reserved

15.1 Example

Given: α = 0.05, μ_o = 155.0, σ = 5.0, n = 10, μ_{str} = 145, μ_{end} = 165, and $\Delta\mu$ = 0.5, plot the one-sided and two-sided operating curve. The following is the suggested keystroke sequence to generate the operating curves:

0.05 [↰][α] 155 [↰][μo]	Enter values of α and μ_o.
5 [↰][σ] 10 [↰][N]	Enter values of σ and n.
145 [↰][μstr] 165 [↰][μend]	Enter values of μ_{str} and μ_{end}.
[NXT] 0.5 [↰][Δμ]	Enter value of $\Delta\mu$.
[→ΣD1]	Generate one-sided operating curve data (*).
[ΣDAT]	Displays ΣDAT matrix.

Use the matrix editor to see all elements. Start by pressing [▼], then use the arrow keys to move through the matrix. Press [ENTER] to return to normal display.

[→PLT]	Plot operating curve scatterplot.
[(x,y)]	Displays (x,y) coordinates of cursor.
[NXT]	To recover menu.

Move cursor using the arrow keys to specific points on the curve to read its coordinates.

[CANCL]	Returns to normal display.
[→DRL]	To draw a line through the points (*).
[CANCL]	Returns to normal display.

Let's try the two-sided operating curve now:

[VAR][NXT][→ΣD2]	Generate two-sided operating curve data (*).
[ΣDAT]	Displays ΣDAT matrix.
[→PLT]	Plot operating curve scatterplot.
[CANCL]	Returns to normal display.
[→DRL]	To draw a line through the points (*).
[CANCL]	Returns to normal display.

(*) Depending on the number of data points involved, these operations may take some time to complete. Be patient.

15.2. Obtaining single values of β

If you are interested only in a point value of β = L(μ) rather than on a long list as in the example above, you can use still the programs in this directory, by using values of μ_{str}, μ_{end}, and $\Delta\mu$ that include the value of μ of interest. For example, suppose that you want the one-sided value of β for μ = 155, using the values of α, μ_o, σ, and n from the example above. Choose μ_{str} = 155, μ_{end} = 155.5, and $\Delta\mu$ = 0.5,and press [→ΣD1]. Once the data are ready, press [ΣDAT] to see the value of interest. For this case, β = L(155) = 0.95.

 © 2000 -Gilberto Urroz - All rights reserved

16.0 MLIN: program for Multiple LINear regression

The programs in this sub-directory allow you to determine the coefficients, correlation coefficient, and other parameters of a multiple linear regression equation of the form:

$$\hat{y} = b_o + b_1x_1 + b_2x_2 + ... + b_kx_k.$$

from a series of n points $\{(x_{1i}, x_{2i}, ..., x_{ki}, y_i), i = 1,n\}$. There are k independent variables (x_i, i = 1,2,.., k), one dependent variable, y, and (k+1) coefficients, (b_i, i = 0,1, ..., k). The coefficients b_i are known as the *least-square estimators*.

To perform multiple linear regression using the programs in this sub-directory use this procedure:

1. To get started, clear the display, by pressing [DEL]. Next, we need to enter the values of each independent variable $\mathbf{x}_i$ as a list and place them in the display in the appropriate order, i.e., enter $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$. Next, we place the dependent variable **y** in the display, and press [→DAT] to set up the data matrix necessary to solve for the vector of coefficients **b** = $[b_o, b_1, b_2, ..., b_k]$. The display will show the values of k and n.
2. Press [→*SOL*] to solve for **b**. The display will show the vector b in level 2 and the correlation coefficient, r, in level 1.
3. Press [→*PLT*] to plot residual errors, e, vs. y.
4. Enter a list $\{x_1\ x_2\ ...\ x_k\}$ and press [→Y] to get $y = b_o + b_1x_1 + b_2x_2 + ... + b_kx_k$.
5. Press [*X*] to see a matrix **X** containing the values of the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$ in columns 2 through k+1. The first column is filled with 1's.
6. Press [*Y*] to see the vector **y**. (Both the matrix **X** and the vector **y** are created with [→*DAT*] in step 1, above).
7. Press [NXT] for the next menu.
8. Press [*B*] to see the vector **b**.
9. Press [*YH*] to see the vector $\hat{\mathbf{y}}$. This vector contains the regression values corresponding to the original values of the independent variables, i.e., $\hat{\mathbf{y}} = \mathbf{X}\ \mathbf{b}$. (A matrix product).
10. Press [*SE*] to see the standard error of estimate, s_e. This value is an estimator of the standard deviation of the distribution of each independent variable.
11. Press [*R*] to see the correlation coefficient for the multiple linear regression.
12. Press [*VARB*] to see the variance vector, i.e., a vector containing estimates of the least square estimators. The elements of this vector are defined by $VARB_{i,i} = var(b_i)$.
13. Press [*K*] to get the number of independent variables, k.
14. Press [NXT] to get to the next menu.
15. Press [*N*] to see the number of data points, n.
16. Press [*ER*] to see the error vector, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.
17. Press [*SSE*] to see the error sum of squares, $SSE = \Sigma(y_i - \hat{y}_i) = \mathbf{e}\cdot\mathbf{e}$.
18. Press [*SST*] to see the total sum of squares, $SST = \Sigma(y_i - \bar{y})$, where $\bar{y}$ is the mean value of y.
19. Press [*YBAR*] to see $\bar{y}$.

 © 2000 -Gilberto Urroz - All rights reserved

20. Press [*COV*] to see the covariance matrix, i.e., a matrix containing estimates of the variance and covariance of the least square estimators. The elements of this matrix are defined by $COV_{i,j} = cov(b_i,b_j)$. Also, $COV_{i,i} = cov(b_i,b_i) = VARB_i = var(b_i)$.
21. Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.

16.1. Example

The following data shows ages, years of college and yearly income of five engineers working in the same company.

age (years) x_1	years of college x_2	Income ($) y
37	4	71200
45	0	66800
38	5	75000
42	2	70300
31	4	65400

Fit an equation of the form to this data and use it to estimate how much a 40-year-old engineer with 4 years of college would make on the average if working for this company.

This is the procedure to follow using the programs of sub-directory MLIN:

{37 45 38 42 31} [ENTER]	Enter $\mathbf{x_1}$
{4 0 5 2 4} [ENTER]	Enter $\mathbf{x_2}$
{712 668 750 703 654} [ENTER]	Enter **y**/100
100 [×] [ENTER]	Now, enter **y**
[→*DAT*]	Prepare data for multiple linear regression. Display shows: k: 2 / n: 5
[→*SOL*]	Performs regression. Display shows **b** = [23721.32 96... / r: .9990232. (A value of r close to 1.0 shows excellent correlation)
{40 4} [ENTER]	Enter $x_1 = 40$, $x_2 = 4$
[→*Y*]	Estimated income for 40 years, 4 years of college.

Display shows: y = 74060.95

Additional information:

- Press [→*PLT*] to see residual errors vs. y. Press [*STATL*] to see the zero axis for errors. Press [*CANCL*] to return to normal display.

- Press [*X*] to see the matrix **X** = [[1 37 4] [1 45 0] [1 38 5] [1 42 2] [1 31 4]]. Use the matrix editor to see it in its entirety. To enter the matrix editor, just press [▼]. Use the arrow keys to move through the matrix. Press [ENTER] to return to normal display if using matrix editor.

- Press [*Y*] to see the vector **y** = [71200 66800 75000 70300 65400]. Use the matrix editor to see it in its entirety. Press [ENTER] to return to normal display if using matrix editor.

- Press [NXT][*B*] to see the vector **b**= [23721.32 960.92 2975.65]. This means that our regression equation can be written as:

© 2000 -Gilberto Urroz - All rights reserved

$$\hat{y} = 23721.32 + 960.92\, x_1 + 2975.65 x_2$$

or

$$\text{Income(\$)} = 23721.32 + 960.92\cdot(\text{Age}) + 2975.65\cdot(\text{Years of college})$$

- Press [*YH*] to see the vector $\hat{\mathbf{y}}$ = [71178.17 66962.94 75114.76 70031.48 65412.62]. Use the matrix editor to see it in its entirety. Press [ENTER] to return to normal display if using matrix editor.
- Press [*SE*] to see the standard error of estimate, s_e = 237.12.
- Press [*R*] to see the correlation coefficient for the multiple linear regression, r = 0.9990232.
- Press [*VARB*] to see the variance vector, VARB = [1594.26 35.29 93.87].
- Press [*K*] to get the number of independent variables, k = 2.
- Press [NXT] [*N*] to see the number of data points, n = 5.
- Press [*ER*] to see the error vector, $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ = [21.82 -162.94 -114.76]. Use the matrix editor to see it in its entirety. Press [ENTER] to return to normal display if using matrix editor.
- Press [*SSE*] to see the error sum of squares, SSE = $\Sigma(y_i - \hat{y}_i) = \mathbf{e}\cdot\mathbf{e}$ = 112456.38.
- Press [*SST*] to see the total sum of squares, SST = $\Sigma(y_i - \bar{y})$ = 57592000.00.
- Press [*YBAR*] to see $\bar{y}$ = 69740.
- Press [*COV*] to see the covariance matrix, COV = [[2541666.86 -55792.19 -125614.18] [-55792.19 1245.70 2569.27] [-125614.18 2569.27 8813.39]]. Use the matrix editor to see it in its entirety. Press [ENTER] to return to normal display if using matrix editor.

17.0 POLY: POLYnomial regression programs

Suppose that n data points (x_i,y_i) can be fitted to a polynomial relationship of the form

$$y = \beta_o + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p,$$

Where *p* is the order of the polynomial. The parameters β_o, β_1, β_2 … β_p, are the solution to a set of linear equations given, in matrix form, by $\mathbf{X}\cdot\mathbf{b} = \mathbf{Y}$, where the vectors **b** and **Y,** and the matrix **X**, are defined as follows:

$$\mathbf{b} = [\beta_o\ \ \beta_1\ \ \beta_2 \ldots \beta_p]^T; \qquad \mathbf{Y} = [\Sigma y\ \ \Sigma xy\ \ \Sigma x^2 y \ldots \Sigma x^p y]^T$$

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^p \\ 1 & x_2 & \ldots & x_2^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \ldots & x_n^p \end{bmatrix}$$

 © 2000 -Gilberto Urroz - All rights reserved

Subdirectory POLY in your HP48G calculator contains a program [→*SOL*] that calculates the parameters
in the vector **b** when given the values of x and y.

To perform a polynomial regression using the programs in this sub-directory use this procedure:

1. Store the values of x and y lists, and the value of p. [→*DAT*] to set up the data matrix necessary to solve for the vector of coefficients **b** = [b_o, b_1, b_2,..., b_k].
2. Press [→*SOL*] to solve for **b**. The display will show the values of the vector of coefficients **b** = [b_o, b_1, b_2,..., b_k], the value of p, and the correlation coefficient r. (Note: the time required for the calculator to perform the regression increases with the value of p, i.e., a regression with p = 5 will take much longer than a regression with p = 2).
3. Press [→*PLT*] to plot residual errors, e vs. y.
4. Enter a value of x and press [→*Y*] to get $y = b_o + b_1x + b_2x^2 + \ldots +b_px^p$.
5. Press [NXT] for the next menu.
6. Press [*N*] to see the number of data points, n.
7. Press [*B*] to see the vector **b**.
8. Press [*R*] to see the correlation coefficient for the multiple linear regression, r.
9. Press [*YBAR*] to see the mean value of y, $\bar{y}$.
10. Press [*YH*] to see the vector $\hat{\mathbf{y}}$. This vector contains the regression values corresponding to the original values of the independent variables, i.e, $\hat{y}_i = = b_o + b_1x_i + b_2x_i^2 + \ldots +b_px_i^p$, for i = 1, 2, ..., n.
11. Press [*SSE*] to see the error sum of squares, $SSE = \Sigma(y_i - \hat{y}_i)$.
12. Press [*SST*] to see the total sum of squares, $SST = \Sigma(y_i - \bar{y})$.
13. Press [*SE*] to see the standard error of estimate, s_e. This value is an estimator of the standard deviation of the distribution of the independent variable x.
14. Press [*X*] to see the matrix **X**.
15. Press [*Y*] to see the vector **Y**.
16. Press [↑*SDA*] to access sub-directory SDATA containing data files described in section 22.

17.1. Example

Fit a polynomial to the following data:

x	y
0	12.0
1	10.5
2	10.0
3	8.0
4	7.0
5	8.0
6	7.5
7	8.5
8	9.0

Use values of p = 2, 3, 4 and 5. And determine the values of r for each one of them. Then, select the value of p with the best value of r, and determine the coefficients of the polynomial regression equation. Finally, determine the value of y expected if x = 3.5.

Follow this procedure:
1) Store the values of x and y as lists into the corresponding variables. For example, for x, use the following keystrokes: [↰][{}][0][SPC][1] [SPC][2][SPC][3][SPC][4][SPC][5]...[8][ENTER][↰][*X*]
2) Store the value of p, p = 2. Use: [2][↰][*P*]
3) Press [→*SOL*]. The correlation coefficient is, r = 0.96056.
4) Store a new value of p, p = 3. Use: [3][↰][*P*]. The result is r = 0.96107.
5) Repeat step 4) for p = 4 and 5. Verify, that r(p=4) = 0.97051, and r(p=5) = 0.97051.

The results indicate that r reaches a maximum value of 0.9751 for p = 4. Using p = 5 does not improve in the value of r. To verify what is happening, we show below the values of the **b** vectors and the correlation coefficients, r, for each value of p:

p = 2,	**b** = [12.18 -1.85 0.18]	r = 0.96056
p = 3,	**b** = [12.12 -1.71 0.14 0.004]	r = 0.96107
p = 4,	**b** = [11.93 -0.71 -0.50 0.13 -0.01]	r = 0.97051
p = 5,	**b** = [11.93 -0.71 -0.50 0.13 -0.01 -1.32×10-12≈0]	r = 0.97051

The three first coefficients of the case p = 3 are very similar to the coefficients of the case p = 2, with the last coefficient of the case p = 3 being very close to zero. This explains the fact that the correlation coefficients for those two cases are very similar. The correlation coefficient improves a little when we use p = 4 or 5 (r is the same for this two cases), over either p = 2 or 3. Notice that the first four coefficients of the case p = 5 are the same as those of the case p = 4, with the last coefficient of the case p = 5 being almost zero. In other words, there is no improvement in the regression by using p = 5 over p = 4. There was a slight improvement on the regression going from p = 2 to p = 4. Let's select the case p =4 as the case with the best correlation. Repeat step 4 for p = 4.

6) Store the value of p, p = 4. Use: [4][↰][*P*]
7) Press [→*SOL*].
8) Press [3][.][5][ENTER][→*Y*], the result is y: 7.7695.

Additional information:

- Press [→*PLT*] to see residual errors vs. y. Press [*STATL*] to see the zero axis for errors. Press [*CANCL*] to return to normal display.

- Press [NXT][*N*] to see the number of points, n = 9.

- Press [*B*] vector **b** = [11.93 -0.71 -0.50 0.13 -0.01]. This means that our regression equation can be written as:

$$\hat{y} = 11.93 \ -0.71\ x \ -0.50\ x^2 + 0.013\ x^3 \ -0.01x^4.$$

- Press [*YBAR*] to see $\bar{y}$ = 8.944.

- Press [*YH*] to see the list $\hat{\mathbf{y}}$ = {11.93 10.84 9.43 8.21 7.48 7.36 7.78 8.48 8.99 }.

- Press [*SSE*] to see the error sum of squares, SSE = $\Sigma(y_i - \hat{y}_i)$ = 1.20.

- Press [NXT][*SST*] to see the total sum of squares, SST = $\Sigma(y_i - \bar{y})$ = 20.72
- Press [*SE*] to see the standard error of estimate, s_e = 0.41.
- Press [*X*] to see the matrix **X**.
- Press [*Y*] to see the vector **Y**.

18.0 CFIT - Curve Fitting using HP48G/GX own features

This sub-directory basically contains the program CRMT that allows you to create a matrix out of a number of lists. The lists become the columns of the matrix. The matrix thus created can be stored into the statistical matrix ΣDAT, and then you can use the HP48G/GX "Fit data..." feature to fit linear, exponential, logarithmic, or power relationships to the data. Examples for using CRMT and the "Fit data..." feature are shown in section 13 of Part I of this guide.

The sub-directory CFIT also contains a sub-directory called 'DATAFIT', which includes a number of variables with data matrices corresponding to problems in the textbook by Miller & Freund. The description of the problems is shown in the table below. To use these data, press the button [*DATAF*] within sub-directory CFIT. Within the sub-directory DATAFIT press the variable that you want to place in the display. (Use [NXT] to move through the sub-directory.) To get back to sub-directory CFIT, press the button [↑*CFIT*]. Next, store the matrix in the display into ΣDAT by using [↰][*ΣDAT*]. At this point you can use the HP48G/GX "Fit data..." feature to fit a relationship to the data in ΣDAT.

Variable name	Data from	Data format
XP331	Example from page 331-M&F	[[x col] [y col]]
XP334	Example from page 334-M&F	[[x col] [y col]]
XP350	Example from page 350-M&F	[[x col] [y col]]
DP350	Example from page 350 (modified)	[[x col] [y col] [log(y) col]
P11.1	Problem 11.1-M&F	[[x col] [y col]]
P11.3	Problem 11.3-M&F	[[x col] [y col]]
P11.6	Problem 11.6-M&F	[[x col] [y col]]
P11.9	Problem 11.9-M&F	[[x col] [y col]]
P1127	Problem 11.27-M&F	[[x col] [y col]]
P1130	Problem 11.30-M&F	[[x col] [y col]]
P1133	Problem 11.33-M&F	[[x col] [y col] [log(3-y) col]
P1174	Problem 11.74-M&F	[[x col] [y col]]

M&F = Miller and Freund's Probability and Statistics for Engineers.

19. 0 GDFIT - - GooDness of FIT test

Observed data; n: number of observations; expected - uniform in this case.
[→EXP]: expected frequency count [→X2]: calculates $\chi 2$

20.0 RC: R x C tables

Given observed data table (OTAB); [→X2] obtains degrees of freedom (v), expected data table (ETAB), and χ^2.

21.0 BAYES - Simple Bayesian estimation

(See section 7.3 in M&F). Use the SOLVE feature in the calculator to solve the different equations contained in variable [*BAYES*].

22.0 SDATA: data files

This sub-directory contains a number of variables with data from problems in the Textbook by Miller and Freund, and from other sources. The description of the problems is shown in the table below. To use these data, press the button corresponding to the variable you want to place in the display. Then, using the [NXT] button, move through the sub-directory and find any of the following keys:

[↑ONE]	To move to sub-directory ONEVAR
[↑GRS]	To move to sub-directory GRSTA
[↑DDIS]	To move to sub-directory DDIST
[↑SIM]	To move to sub-directory SIMH
[↑CHK]	To move to sub-directory CHKN
[↑MLI]	To move to sub-directory MLIN
[↑POL]	To move to sub-directory POLY
[↑GDF]	To move to sub-directory GDFIT
[↑RC]	To move to sub-directory RC

Press the key that will take you to the sub-directory of your choice. Store the data in the corresponding variable in the destination sub-directory.

Variable	*Data from*	*Can be used with*	*Data format*
XP11	Example from page 11-M&F	ONEVAR, CHKN	List of values
XP32	Example from page 32-M&F	GRSTA	{ x-list f-list}(*)
P2.7	Problem 2.7-M&F	ONEVAR, CHKN	List of values
P2.10	Problem 2.10-M&F	ONEVAR, CHKN	List of values
P2.12	Problem 2.12-M&F	ONEVAR, CHKN	List of values
P2.29	Problem 2.29-M&F	ONEVAR, CHKN	List of values
P2.32	Problem 2.32-M&F	ONEVAR, CHKN	List of values
P2.36	Problem 2.36-M&F	ONEVAR, CHKN	List of values
P2.38	Problem 2.38-M&F	GRSTA	{ x-list f-list}(*)
XP132	Example from page 132-M&F	SIMH	{ tbx-list bound-list}(*)
XP133	Example from page 133-M&F	SIMH	{ tbx-list bound-list}(*)
DUNI	Data from uniform distribution	ONEVAR, CHKN	List of values
DEXP	Data from exponential distribution	ONEVAR, CHKN	List of values
DWEI	Data from Weibull distribution	ONEVAR, CHKN	List of values
DNOR	Data from normal distribution	ONEVAR, CHKN	List of values
DLOG	Data from log-normal distribution	ONEVAR, CHKN	List of values
$\bar{x}$DAT	Data from $\bar{x}$SIM / GEND.	ONEVAR, CHKN	List of values
XP199	Example from page 199-M&F	ONEVAR, CHKN	List of values
P1138	Problem 11.38-M&F	MLIN	{x1-list x2-list y-list}(*)
P1180	Problem 11.80-M&F	MLIN	{x1-list x2-list y-list}(*)
P1134	Problem 11.34-M&F	POLY	{x-list y-list} (*)
POL1	Data for polynomial fitting	POLY	{x-list y-list} (*)
P9.49	Problem 9.49-M&F	GRSTA, GDFIT	{ x-list f-list}(*)
P9.43	Problem 9.43-M&F	RC	RxC table (matrix)
P9.44	Problem 9.44-M&F	RC	RxC table (matrix)
P9.71	Problem 9.71-M&F	RC	RxC table (matrix)

ABOUT THE AUTHOR

Gilberto E. Urroz is an Associate Professor of Civil and Environmental Engineering and a researcher at the Utah Water Research Laboratory, both at Utah State University, in Logan, Utah. He has been a teacher of engineering disciplines for more than 15 years both in his native Nicaragua and in the United States.

His teaching experience includes courses on introductory physics, engineering mechanics, probability and statistics for engineers, computer programming, fluid mechanics, hydraulics, and numerical methods. His research interests include mathematical and numerical modeling of fluid systems, hydraulic structures, and erosion control applications.

Dr. Urroz is an expert on the HP 48 G and HP 49 G series calculator and has written several books on applications of these computing devices to disciplines such as engineering mechanics, hydraulics, and science and engineering mathematics. His personal interests include reading, music, opera, theater, and taijiquan.

ABOUT GREATUNPUBLISHED.COM

greatunpublished.com is a website that exists to serve writers and readers, and remove some of the commercial barriers between them. When you purchase a greatunpublished.com title, whether you receive it in electronic form or in a paperback volume or as a signed copy of the author's manuscript, you can be assured that the author is receiving a majority of the post-production revenue. Writers who join greatunpublished.com support the site and its marketing efforts with a per-title fee, and a portion of the site's share of profits are channeled into literacy programs.

So by purchasing this title from greatunpublished.com, you are helping to revolutionize the publishing industry for the benefit of writers and readers. And for this we thank you.

This book is the result of more than four years of teaching probability and statistics for engineers at Utah State University.

The course involves heavy use of the HP 48 G series calculator for probability and statistical applications. The calculator's pre-programmed features and programs developed by the author are presented and demonstrated in the book with applied examples.

Subjects include

- Data manipulation
- Sample statistics
- Frequency analysis
- Random number generation
- Statistical graphical analysis
- Confidence intervals
- Hypothesis testing
- Operating curves
- Linear regression
- Multiple-linear regression
- Polynomial regression
- RxC tables

ISBN 1-58898-040-5
90000>
9 781588 980403

www.greatunpublished.com