Programs for Probability and Statistics with the HP 48 G Calculator

By

Gilberto E. Urroz, Ph.D., P.E.

Distributed by

infoClearinghouse.com

PART II – PROGRAMS FOR STATISTICAL ANALYSIS

In this document we will present a collection of sub-directories containing User RPL programs that can be used for a number of statistical applications. The programs are available at:

http://www.engineering.usu.edu/cee/faculty/gurro/MyBooks/StatPrograms.zip

You will need a cable to transfer the programs from the computer to your calculator. Information on HP's connection cable can be found at:

http://www.hp.com/calculators/accessories/connect.html

Please report any errors in this document to: gurro@cc.usu.edu

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.

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.

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.

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.

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 }.

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 Σf = 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).

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
[*σ2*]	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:	[*σ2*]	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).

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.

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][*→XLIS*][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.

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.

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}.

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* (*tab*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:

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.

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 [→*σ2*] 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, $\mu = 0.0$
[→*σ2*]	(*) Calculate the variance, $\sigma^2= 1.0$
[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.

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 α = 3 and β = 2 in the corresponding variables in directory CFUNC. Then, press the following keys:

[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.

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}

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].

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, $\alpha = P(Z>z)$.

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 $\sigma^2 = 0.5$. 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 $\sigma^2 = 0.16$. 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 $\sigma^2 = 0.25$, 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 $\sigma^2 = 0.9$. Use the following keystrokes:

[2][.][5][↩][*μ*] [.][9][↩][*σ2*] [.][0][5] [*α→X*]   The results are: α: 0.05, x: 4.06.

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] [$Z\rightarrow\alpha$] 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][$\alpha\rightarrow Z$] 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] [$T\rightarrow\alpha$] 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] [$\alpha\rightarrow T$] 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 $\alpha = P(\Xi^2>\chi^2)$, if χ^2 = 2.5. Use the following keystrokes:

[8][↰][ν] [2][.][5] [$X2\rightarrow\alpha$] The results are: X^2: 2.50, α: 0.96.

Determine χ^2 if $\alpha = P(\Xi^2>\chi^2)$ = 0.05 with ν = 5 degrees of freedom. Use the following keystrokes:

[5][↰][ν] [.][0][5] [$\alpha\rightarrow X2$] 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 $\alpha = P(\Im>F)$, if F = 0.5. Use the following keystrokes:

[5][↰][νN] [8][↰][νD] [.][5] [$F\rightarrow\alpha$] The results are: F: 0.50, α: 0.76.

Determine F if $\alpha = P(\Im>F)$ = 0.05 for with νN= 15, and νD= 5,. Use the following keystrokes:

[1][5][↰][νN] [5][↰][νD] [.][0][5] [$\alpha\rightarrow F$] The results are: α: 0.05, F: 4.61.

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.

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.

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.

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**,

$$f(x) = 1/10, \text{ for } x = 0, 1, 2, \ldots, 9;$$

(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'` ?

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.

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.

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: μ = 2.3, for a population having a standard deviation σ = 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: μ > 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.

The result is: *One-Tail, z - test, Reject H0: μ = 2.3*. Press [OK], and get the following additional results: *α = 0.01, P-value = 6.6E-4, z = 3.2115, σ = 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: μ ≠ 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 [↰][$\mu 0$]	Enter population mean to be tested ($\mu 0$)
-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:

[$\bar{x}$*SN*][*OK*]	Select the $\bar{x}$SN option.
0.05 [↰][α]	Enter level of significance.
4.0 [↰][$\mu 0$]	Enter population mean to be tested ($\mu 0$)
1.5 [↰][σ]	Enter known population standard deviation.
2.5 [↰][$\bar{x}$]	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 = 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 $\sigma = 1.5$ 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. $\bar{x}$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 ($\bar{x}_1$, s_1, n_1, $\bar{x}_2$, s_2, n_2) are known, use the option [$\bar{x}$*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 $\bar{x}$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 [↰][*σ1*] and [↰][*σ2*]. 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][+/−] [↰][*σ1*] (and/or [1][+/−] [↰][*σ2*]). 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][+/−] [↰][*σ1*], and [2][+/−] [↰][*σ2*].
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 (σ1, σ2); 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 [↰][*σ1*] and [↰][*σ2*]. 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][+/–] [↰][*σ1*] (and/or [1][+/–] [↰][*σ2*]). 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][+/–] [↰][*σ1*], and [2][+/–] [↰][*σ2*].
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 of freedom if a t test is used (ν); and, the populations standard deviations ($\sigma 1$, $\sigma 2$). 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: $\mu_1 - \mu_2 = \delta$).
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 ($\sigma 1$, $\sigma 2$). 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].

<u>NOTE</u>: 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_o: $\mu_1 - \mu_2 = 2.0$, 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: $\mu_1 - \mu_2 \neq 2$.

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,* $\bar{x}2 = 4.26$, *n2 = 5, s1 = 1.3779,* $\bar{x}1 = 4.266$, *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,* $\bar{x}2 = 4.26$, *n2 = 5, s1 = 1.3779,* $\bar{x}1 = 4.266$, *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*, $t = -4.1722$, $\nu = 9$, $\sigma 2 = -1$, $\sigma 1 = -1$, $s2 = 1.802$, $\bar{x}2 = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\bar{x}1 = 4.266$, $n1 = 6$.

14.2.4. Repeat example 14.1.1 for δ = 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*, $t = -4.1722$, $\nu = 4$, $\sigma 2 = -1$, $\sigma 1 = -1$, $s2 = 1.802$, $\bar{x}2 = 4.26$, $n2 = 5$, $s1 = 1.3779$, $\bar{x}1 = 4.266$, $n1 = 6$.

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:

$$\nu = \frac{(\frac{\sigma_1}{n_1})^2 + (\frac{\sigma_2}{n_2})^2}{\frac{(\frac{\sigma_1}{n_1})^2}{n_1 - 1} + \frac{(\frac{\sigma_2}{n_2})^2}{n_2 - 1}}$$

This test is referred to as the *Smith-Sathertwaite test*.
If the two samples come from populations having the same standard deviation then,

$$\nu = n_1 + n_2 - 2.$$

For example, in example 14.2.4, which uses a Smith-Sathertwaite test, we found $\nu = 4$, 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 $\nu = 9$. 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: $\mu_1 - \mu_2 > 4.0$.

<u>*Solution*</u>: 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* $= 5.5608E-52 \approx 0$, $z = -15.124$, $\sigma 2 = 1.5$, $\sigma 1 = 1.5$.

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-81 ≈ 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 $\sigma = 1.5$ 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-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 $\sigma = 1.5$, $n = 16$) to see the recommendation of rejecting H_o: $\mu = 1$, 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*].

- **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 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.
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 (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.
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: $\sigma_1^2\neq\sigma2_o^2$. 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 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 [*x̄TST1*] 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.

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

$$P(Z>z_\alpha) = \alpha = 1 - F(z_\alpha) = UTPN(0,1,z_\alpha).$$

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.

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 $\beta = L(\mu)$ 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.

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:

$$y^{\wedge} = b_o + b_1x_1 + b_2x_2 + \ldots + b_kx_k.$$

from a series of n points $\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i), i = 1,n\}$. 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 $\mathbf{b} = [b_o, b_1, b_2,\ldots, b_k]$. 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 $\{x_1\ x_2\ \ldots\ x_k\}$ and press [→Y] to get $y = b_o + b_1x_1 + b_2x_2 + \ldots + b_kx_k$.
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 $\mathbf{y}^{\wedge}$. This vector contains the regression values corresponding to the original values of the independent variables, i.e., $\mathbf{y}^{\wedge} = \mathbf{X}\,\mathbf{b}$. (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, $\mathbf{e} = \mathbf{y} - \mathbf{y}^{\wedge}$.
17. Press [*SSE*] to see the error sum of squares, $SSE = \Sigma(y_i - y^{\wedge}_i) = \mathbf{e}\cdot\mathbf{e}$.
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 $\bar{y}$.
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:

$$y^{\wedge} = 23721.32 + 960.92\, x_1 + 2975.65 x_2$$

or

$$\text{Income(\$)} = 23721.32 + 960.92\cdot\text{(Age)} + 2975.65\cdot\text{(Years of college)}$$

- Press [*YH*] to see the vector $\mathbf{y}^{\wedge}$ = [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, $\mathbf{e} = \mathbf{y} - \mathbf{y}^{\wedge}$ = [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 = $\Sigma(y_i - y^{\wedge}_i) = \mathbf{e}\cdot\mathbf{e}$ = 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 $\beta_o, \beta_1, \beta_2 \ldots \beta_p$, are the solution to a set of linear equations given, in matrix form, by **X**·**b** = **Y**, where the vectors **b** and **Y,** and the matrix **X**, are defined as follows:

$$\mathbf{b} = [\beta_o\ \beta_1\ \beta_2 \ldots \beta_p]^T ; \qquad \mathbf{Y} = [\Sigma y\ \ \Sigma xy\ \ \Sigma x^2 y \ldots \Sigma x^p y]^T$$

$$\mathbf{M} = \begin{bmatrix} 1 & x_1 & \ldots & x_1^p \\ 1 & x_2 & \ldots & x_2^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \ldots & x_n^p \end{bmatrix}$$

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 $y = b_o + b_1 x + b_2 x^2 + \ldots + b_p x^p$.
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, $\hat{y}_i = b_o + b_1 x_i + b_2 x_i^2 + \ldots + b_p x_i^p$, 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. <u>Example</u>
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:

$$y^{\wedge} = 11.93 \; -0.71 \, x -0.50 \, x^2 + 0.013 \, x^3 -0.01x^4.$$

- Press [*YBAR*] to see $\bar{y}$ = 8.944.
- Press [*YH*] to see the list $\mathbf{y}^{\wedge}$ = {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 = $\Sigma(y_i - y^{\wedge}_i)$ = 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 (ν), 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)