HEWLETT-PACKARD

Step-by-Step Examples for Your HP-28C

Probability and Statistics

Probability and Statistics

Step-by-Step Examples for Your HP-28C

HEWLETT PACKARD

Edition 1 August 1987
Reorder Number 00028-90043

Notice

The information contained in this document is subject to change without notice.

Hewlett-Packard makes no warranty of any kind with regard to this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Hewlett-Packard shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Hewlett-Packard assumes no responsiblity for the use or reliability of its software on equipment that is not furnished by Hewlett-Packard.

Portable Computer Division
1000 N.E. Circle Blvd.
Corvallis, OR 97330, U.S.A.

Printing History

Edition 1 August 1987 Mfg No. 00028-90058

Welcome...

... to the HP-28C Step-by-Step Booklets. These books are designed to help you get the most from your HP-28C calculator.

This booklet, *Probability and Statistics*, provides examples and techniques for solving problems on your HP-28C. A variety of statistics matrix manipulations and statistical function computations are designed to familiarize you with, and build upon, the statistics capabilities built into your HP-28C.

Before you try the examples in this book, you should be familiar with certain concepts from the owner's documentation:

- The basics of your calculator – how to move from menu to menu, how to exit graphics and edit modes, and how to use the menu to assign values to, and solve for, user variables.
- Entering numbers and algebraic expressions into the calculator.

Please review the section "How to Use This Booklet." It contains important information on the examples in this booklet.

For more information about the topics in the *Probability and Statistics* booklet, refer to a basic textbook on the subject. Many references are available in university libraries and in technical and college bookstores. The examples in the booklet demonstrate approaches to solving certain problems, but they do not cover the many ways to approach solutions to mathematical problems.

Contents

How To Use This Booklet

Please take a moment to familiarize yourself with the formats used in this booklet.

Keys and Menu Selection: A box represents a key on the calculator keyboard.

[ENTER]
[1/x]
[STO]

[ARRAY]
[PLOT]
[ALGEBRA]

In many cases, a box represents a shifted key on the HP-28C. In the example problems, the shift key is NOT explicitly shown (for example, [ARRAY] requires the press of the shift key, followed by the ARRAY key, found above the "A" on the left keyboard).

The "inverse" highlight represents a menu label:

≡ DRAW ≡ (found in the [PLOT] menu)
≡ ISOL ≡ (found in the [ALGEBRA] menu)
≡ ABCD ≡ (a user-created name, found in the [USER] menu)

Menus typically include more menu labels than can be displayed above the six redefinable menu keys. Press [NEXT] and [PREV] to roll through the menu options. For simplicity, [NEXT] and [PREV] are NOT shown in the examples.

Solving for a user variable within SOLVR is initiated by the shift key, followed by the appropriate user-defined menu key:

[ ] ABCD.

The keys above indicate the shift key, followed by the user-defined key labeled "ABCD". Pressing these keys initiates the Solver function to seek a solution for "ABCD" in a specified equation.

The symbol [<>] indicates the cursor-menu key.

Interactive Plots and the Graphics Cursor: Coordinate values you obtain from plots using the [INS] and [DEL] digitizing keys may differ from those shown, due to small differences in the positions of the graphics cursor. The values you obtain should be satisfactory for the Solver root-finding that follows.

Display Formats and Numeric Input: Negative numbers, displayed as

```
-5
-12345.678
[[-1,-2,-3 [ -4,-5,-6 [ ...
```

are created using the [CHS] key.

```
5 [CHS]
12345.678 [CHS]
[[1 [CHS] ,2 [CHS] , ...
```

The examples in this book typically specify a display format for the number of decimal places. If your display is set such that numeric displays do not match exactly, you can modify your display format with the [MODE] menu and the FIX key within that menu (e.g. [MODE] 2 FIX).

Combinations and Permutations

The HP-28C programs that follow provide simple building blocks for combinatorial analysis. Complex problems are readily evaluated by combining the results left on the stack, or by using the programs below as subroutines.

Permutations: Given X distinct objects, the number of ways to select and arrange Y of these objects in different order is computed by the formula below.

$$_XP_Y = \frac{X!}{(X-Y)!}$$

Clear the stack and key in the permutations program.

```
[CLEAR] [MODE] STD [USER]
« → x y 'FACT(x)÷
FACT(x-y [ENTER]
```

```
2:
1: « → x y 'FACT(x)/
   FACT(x-y)' »
ORDER CLUSR MEM
```

Store the program in the variable *PERM*.

```
'PERM [STO]
```

```
3:
2:
1:
PERM
```

Combinations: Combinations ignore the order in the Y objects chosen, and are computed by the formula below.

$$_XC_Y = \frac{X!}{Y!(X-Y)!}$$

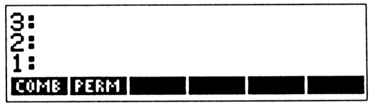
Key in the combinations program.

```
« → x y 'PERM(x,y)÷
FACT(y [ENTER]
```

```
2:
1: « → x y 'PERM(x,y)/
   FACT(y)' »
PERM
```

Store the program in the variable *COMB*.

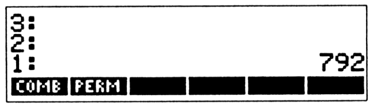
' COMB [STO]



Example: Compute how many five-person basketball squads can be formed from 12 players. The computation to be made is ${}_{12}C_5$.

With the program *COMB* keyed in as above, key in the parameters and evaluate the formula.

12 [ENTER]
5 [COMB]

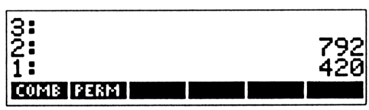


792 squads can be formed. Any combination of five players is acceptable, since the combination program was used to compose the number of teams.

Example: For the problem above, what if one of the two tallest players *must* be on the squad, and these two players never play at the same time?

There are now ten players from which to select the four remaining positions, and two ways to select the fifth. Thus, compute ${}_{10}C_4 \cdot 2$.

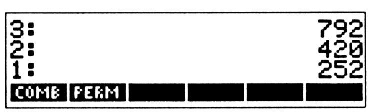
10 [ENTER]
4 [COMB]
2 [×]



Example: Compute the number of options lost if both tall players from the previous example foul out.

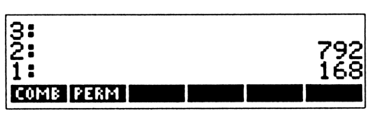
Form the five-person squad from the remaining players.

10 [ENTER]
5 [COMB]



The options lost are computed by subtracting.

[−]



168 squad combinations were lost as a result.

Example: Compute the number of permutations of the twelve original players that are possible.

12 [ENTER]
5 [PERM]

```
3:                      792
2:                      168
1:                    95040
 COMB  PERM
```

For large values of X and Y, it may be desirable to use a program that computes the value of the combination or permutation formula by explicitly multiplying the appropriate terms of the factorials. This can improve the accuracy of the result.

For example, rather than evaluating $_XP_Y$ by $FACT(X) \div (FACT(X-Y)$, compute it as the product of the appropriate terms:
$X \cdot (X-1) \cdot (X-2) \cdots (X-Y-1)$.

Key in the following program and compute $_{12}P_5$ from the previous example.

```
« → x y « x y - 'y'
STO x WHILE x 1 - y >
REPEAT x 1 - DUP 'x'
STO × END
```
[ENTER]

```
1: « → x y « x y - 'y'
   STO x WHILE x 1 - y
   > REPEAT x 1 - DUP
 COMB  PERM
```

'PER2 [STO]

```
3:                      792
2:                      168
1:                    95040
 PER2  COMB  PERM
```

12 [ENTER]
5 [PER2]

```
3:                      168
2:                    95040
1:                    95040
 PER2  COMB  PERM
```

An example of the accuracy difference between the two approaches can be seen by computing $_{20}P_{10}$.

Purge the variables created in this section.

{ 'COMB' 'PERM' 'PER2 [PURGE]

Statistics Matrix Setup

This section describes the structure of the statistics matrices used in the remainder of this booklet, and provides a number of techniques for manipulating the data within the statistics matrix. An approach to managing grouped data is also described.

Initialization and Data Entry

The statistics calculations throughout this booklet generally operate on single or paired columns of data collected in the variable ΣDAT. This statistics matrix can, however, hold additional data vectors, providing for multiple pairing and analysis.

Ungrouped Data Matrix: The statistics matrix ΣDAT for ungrouped data has the form shown below.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

The matrix shown above has n sets (vectors) of statistics data, each containing m data points.

Grouped Data Matrix: The approach used to manage grouped data in this booklet is to collect data using the same functions as for ungrouped data, including data entry and removal and data pair selection. However, once the grouped data has been entered, the data matrix is stored in another matrix variable, and the data is expanded into ΣDAT as if it were ungrouped data. This approach has a disadvantage in terms of memory consumption (since the data is effectively retained twice in the machine); however, it greatly simplifies the steps and programs to compute basic and advanced statistics on the data, since many powerful functions for ungrouped data are built into the HP-28C.

Thus, the grouped data matrix shown below is transformed to the ungrouped form shown earlier prior to calculating statistics for the data.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} & g_1 \\ x_{21} & x_{22} & \cdots & x_{2m} & g_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} & g_n \end{bmatrix}$$

Data vector one in the matrix above occurs g_1 times, data vector two occurs g_2 times, and so on.

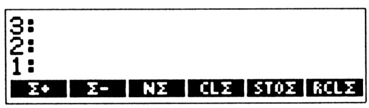
Initialization and Data Entry Examples

At the beginning of each new problem, the current statistics matrix is cleared by pressing CLΣ. A statistics matrix may also be saved for later use by recalling and assigning it to another matrix variable.

Example: Clear the stack and the current statistics matrix, and then enter the following ungrouped statistical data.

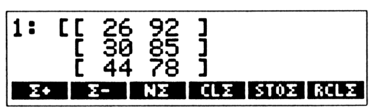
$$\begin{bmatrix} x_i & y_i \\ \text{--} & \text{--} \\ 26 & 92 \\ 30 & 85 \\ 44 & 78 \\ 50 & 81 \\ 62 & 54 \\ 68 & 51 \\ 74 & 40 \end{bmatrix}$$

CLEAR STAT CLΣ
[26,92 Σ+
[30,85 Σ+
[44,78 Σ+
[50,81 Σ+
[62,54 Σ+
[68,51 Σ+
[74,40 Σ+

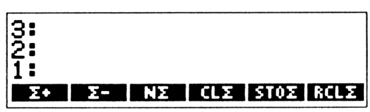


Now recall the statistics matrix and copy it to another variable named *UNGR*.

RCLΣ



'UNGR STO



Example: Clear the current statistics matrix and enter the following set of grouped data.

$$\begin{bmatrix} x_i & y_i & g_i \\ \text{--} & \text{--} & \text{--} \\ 4.8 & 15.1 & 1 \\ 5.2 & 11.5 & 3 \\ 3.8 & 14.3 & 1 \\ 4.4 & 13.6 & 6 \\ 4.1 & 12.8 & 2 \end{bmatrix}$$

```
CLΣ
[4.8,15.1,1   Σ+
[5.2,11.5,3   Σ+
[3.8,14.3,1   Σ+
[4.4,13.6,6   Σ+
[4.1,12.8,2   Σ+
```

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Recall this matrix of grouped data, copy to the variable *GRUP*, and re-display the matrix.

```
RCLΣ  'GRUP STO
RCLΣ
```

```
1: [[ 4.8 15.1 1 ]
    [ 5.2 11.5 3 ]
    [ 3.8 14.3 1 ]
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Data Removal

The last data vector in the statistics matrix is easily removed by using Σ−. Removal of a specified vector other than the last entry is accomplished with the program below.

Last Data Vector Removal: Remove the last data vector from the ungrouped data matrix and display and view the matrix contents.

First, recall the ungrouped data from the variable *UNGR*.

[USER] UNGR

```
1:  [[ 26  92 ]
     [ 30  85 ]
     [ 44  78 ]
GRUP ΣDAT UNGR
```

Make this matrix the current statistics matrix.

[STAT] STOΣ

```
3:
2:
1:
Σ+   Σ-   NΣ   CLΣ  STOΣ RCLΣ
```

Now remove the last data vector.

Σ−

```
3:
2:
1:                [ 74  40 ]
Σ+   Σ-   NΣ   CLΣ  STOΣ RCLΣ
```

Recall the matrix to the stack and examine the contents.

RCLΣ
[VIEW↓]
[VIEW↓]
[VIEW↓]
[ATTN]

```
1:  [[ 26  92 ]
     [ 30  85 ]
     [ 44  78 ]
Σ+   Σ-   NΣ   CLΣ  STOΣ RCLΣ
```

Note that there are now six pairs of data in the statistics matrix. *UNGR* still contains, of course, the original data set.

Arbitrary Data Vector Removal: Key in the program below for removing a specified data vector from the current statistics matrix.

Program	**Comments**
`« NΣ SWAP - → n`	Compute number of rows below row to be discarded.
`« 1 n 1 + START Σ- NEXT`	Stack rows from the bottom through the discard row.
`DROP`	Drop the discard row.
`IF n 0 ≠ THEN 1 n START Σ+ NEXT END » »`	Put the rows below the discard row back into the statistics matrix.

ENTER 'DELI STO

Example: Remove the third data vector from the current statistics matrix.

CLEAR
3 USER DELI

```
3:
2:
1:
DELI GRUP ΣDAT UNGR
```

Display the matrix.

STAT RCLΣ <>

```
1: [[ 26 92 ]
    [ 30 85 ]
    [ 50 81 ]
    [ 62 54 ]
```

Data Column Extraction

The program *GET1* retrieves a column of data from the current statistics matrix. The desired column number is passed to the program in level 1 of the stack.

Program	**Comments**
`«RCLΣ DUP TRN STOΣ`	Save the original data and transpose to drop unwanted columns in row form.
`SWAP 1 + NΣ DUP2 IF ≤ THEN START Σ- DROP NEXT ELSE DROP2 END`	Drop all rows (columns) beyond specified one.
`Σ- SWAP STOΣ »`	Get specified column, restore data.

[ENTER] `'GET1` [STO]

Example: Get a vector containing the elements of the first column of the current statistics matrix.

`1` [USER] [GET1]

```
3:
2: [[ 26 92 ]  [ 30 85…
1:    [ 26 30 50 62 68 ]
ΣDAT GET1 DELI GRUP UNGR
```

Data Sets With More Than Two Variables

The user variable ΣPAR contains a list of four real numbers. The first two numbers determine the columns of the statistics matrix operated on by the statistics functions of the HP-28C.

The command COLΣ takes two column numbers from the stack and stores them as the first two objects in the list in the variable ΣPAR.

Example: For the multiple-pair data set below, specify columns one and three as the pair for analysis. (This capability will be discussed further in later examples in this booklet.)

$$\begin{bmatrix} t_i & x_i & y_i \\ \text{--} & \text{--} & \text{--} \\ 1 & 26 & 92 \\ 2 & 30 & 85 \\ 3 & 44 & 78 \\ 4 & 50 & 81 \\ 5 & 62 & 54 \\ 6 & 68 & 51 \\ 7 & 74 & 40 \end{bmatrix}$$

Clear the stack and the current statistics matrix, and enter the data above.

CLEAR STAT CLΣ
[1,26,92 Σ+
[2,30,85 Σ+
[3,44,78 Σ+
[4,50,81 Σ+
[5,62,54 Σ+
[6,68,51 Σ+
[7,74,40 Σ+

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Now specify columns one and three as the pair of data vectors for analysis.

1,3 COLΣ

```
3:
2:
1:
 COLΣ | CORR | COV | LR | PREDV |
```

Recall the statistics matrix parameters to examine the columns specified.

USER ΣPAR

```
3:
2:
1:                 { 1 3 0 0 }
 ΣPAR | ΣDAT | GET1 | DELI | GRUP | UNGR
```

Columns one and three are the currently specified data vectors. This process is useful for multiple data vector manipulations or regressions on multiple sets of data with the same base- or time-line.

Grouped Data Matrix Transformation

The program below transforms grouped data in the current statistics matrix to an ungrouped form in the current statistics matrix.

Program	Comments
`«RCLΣ 'GD' STO RCLΣ`	Recall the grouped data and save it in *GD*.
`CLΣ ARRY→ LIST→ DROP → n m`	Transform grouped data to element form and save the dimensions.
`« 1 n START`	Loop *n* times.
`m ROLLD m 1 - 1 →LIST →ARRY → ar`	Save g_i in stack, place data in temporary vector for expansion.
`« 1 SWAP START ar Σ+ NEXT »`	Setup and accumulate *ar* g_i times.
`NEXT » »`	Repeat outer loop.

[ENTER] `'XFRM` [STO]

Example: Use the program *XFRM* to transform the grouped data matrix *GRUP*.

With the program entered and stored, recall the grouped data matrix and make it the current statistics matrix.

[CLEAR]
[USER] GRUP
[STAT] STOΣ

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Now run the program *XFRM* on the grouped data.

[USER] XFRM

```
3:
2:
1:
 ΣDAT | GD | ΣPAR | GET1 | DEL1 | GRUP
```

Recall the current statistics matrix and review the data.

[STAT] RCLΣ

```
1: [[ 4.1 12.8 ]
    [ 4.1 12.8 ]
    [ 4.4 13.6 ]
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Use [VIEW↓] to scan the matrix. Note how the program builds the transformed, ungrouped matrix from the bottom to the top of the grouped data.

Purge the grouped data matrix *GD*. The original data exists in *GRUP*.

'GD [PURGE]

Basic Statistics for Multiple Variables

This section provides keystrokes and programs to calculate a variety of basic statistics on the current statistics matrix. These statistics include means, standard deviations, variances, and covariances on both samples and populations, correlation coefficient, coefficients of variation, sums of data and sums of products of data, normalized data, moments, and delta percents on paired statistics.

The current statistics matrix is assumed to be ungrouped data in the calculations that follow. For grouped data, the statistics matrix should be transformed by the program *XFRM* described in the previous section.

Sums and Means

Sums and means for each column of statistics in the current statistics matrix are easily calculated on the HP-28C. The mean and sum are computed from the formulas

$$\text{mean} = \sum_{i=1}^{n} \frac{x_i}{n}$$

and

$$\text{sum} = \sum_{i=1}^{n} x_i$$

where x_i is the ith coordinate value in a column, and n is the number of data vectors.

Example: Compute the sums and means for the ungrouped data stored in the variable *UNGR*.

First, clear the stack and recall the matrix.

CLEAR USER UNGR

```
1: [[ 26 92 ]
    [ 30 85 ]
    [ 44 78 ]]
UNGR
```

Specify this matrix as the current statistics matrix.

STAT STOΣ

```
3:
2:
1:
Σ+ Σ- NΣ CLΣ STOΣ RCLΣ
```

Compute the column totals.

TOT

```
3:
2:
1:              [ 354 481 ]
TOT MEAN SDEV VAR MAXΣ MINΣ
```

Compute the means. Change the display setting to two digits following the decimal point.

MEAN MODE 2 FIX

```
3:
2:      [ 354.00 481.00 ]
1:         [ 50.57 68.71 ]
STD [FIX] SCI ENG DEG [RAD]
```

Standard Deviation, Variance, and Covariance

Both sample and population statistics are readily computed using the built-in functions of the HP-28C. For the population statistics, a short program as described in the "STAT" section of the Reference Manual makes calculation of the population statistics easy.

Standard Deviation

The standard deviation of the sample and population are given by the following formulas.

Sample Standard Deviation:

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

Population Standard Deviation:

$$\sigma_x = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^{\frac{1}{2}}$$

Variance

The variance of the sample and population are given by the following formulas.

Sample Variance:

$$s_x^{\,2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Population Variance:

$$\sigma_x^{\,2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

For the formulas above, x_i is the ith coordinate value in a column, $\overline{x}$ is the mean of the data in this column, and n is the number of data vectors.

Covariance

The covariance of the sample and population are given by the following formulas.

Sample covariance:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_{im_1} - \overline{x}_{im_1})(x_{im_2} - \overline{x}_{n_2})\right]$$

Population Covariance:

$$\sigma_{xy} = \frac{1}{n}\left[\sum_{i=1}^{n}(x_{im_1} - \overline{x}_{im_1})(x_{im_2} - \overline{x}_{n_2})\right]$$

In the above formulas, n is the number of data vectors, x_{im_1} is the ith coordinate value in column m_1, and $\overline{x}_{m_1}$ is the mean of the data in column m_1.

Computing Sample and Population Statistics

A program of the general form `«MEAN Σ+ fn Σ- DROP»`, where `"fn"` is replaced by the appropriate HP-28C function `(SDEV, VAR,` or `COV)`, will compute the population statistics for the specified function.

Example: Compute the sample and population standard deviation, variance, and covariance for the current statistics matrix.

First, key in the population statistics programs. (A fast way to key in the second and third programs is to duplicate and EDIT the previous program with the function change.)

```
[CLEAR]
«MEAN Σ+ SDEV Σ- DROP
[ENTER]
«MEAN Σ+ VAR Σ- DROP
[ENTER]
«MEAN Σ+ COV Σ- DROP
[ENTER]
```

```
2: « MEAN Σ+ VAR Σ- DR...
1: « MEAN Σ+ COV Σ-
   DROP »
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Store the population statistics programs.

```
'COVP [STO]
'VARP [STO]
'SDVP [STO] [USER]
```

```
3:
2:
1:
 SDVP | VARP | COVP | ΣDAT | XFRM | ΣPAR
```

Now compute the sample and population standard deviations.

[CLEAR]
[STAT] SDEV
[USER] SDVP

```
3:
2:          [ 18.50 20.00 ]
1:          [ 17.13 18.51 ]
SDVP VARP COVP ΣDAT XFRM ΣPAR
```

Compute the sample and population variances.

[CLEAR]
[STAT] VAR
[USER] VARP

```
3:
2:      [ 342.29 399.90 ]
1:      [ 293.39 342.78 ]
SDVP VARP COVP ΣDAT XFRM ΣPAR
```

Compute the sample and population covariances. Specify columns one and two in *ΣPAR* if they have been set otherwise.

[CLEAR]
[STAT] `1,2` COLΣ
COV
[USER] COVP

```
3:
2:               -354.14
1:               -303.55
SDVP VARP COVP ΣDAT XFRM ΣPAR
```

Correlation Coefficient and Coefficient of Variation

The correlation coefficient and coefficient of variation are computed by the following formulas.

Correlation Coefficient

$$\frac{\sum_{i=1}^{n}(x_{im_1}-\overline{x}_{m_1})(x_{im_2}-\overline{x}_{m_2})}{\left[\sum_{i=1}^{n}(x_{im_1}-\overline{x}_{m_1})^2\sum_{i=1}^{n}(x_{im_2}-\overline{x}_{m_2})^2\right]^{\frac{1}{2}}}$$

Coefficient of Variation

$$V_x=\frac{s_x}{\overline{x}}\cdot 100$$

The terms are defined in the previous problem section.

Compute the statistics above for the *grouped* statistics data *GRUP*.

First, recall and transform the grouped data into the current statistics matrix.

[CLEAR]
[USER] [GRUP]

```
1: [[ 4.80 15.10 1.00 ]
    [ 5.20 11.50 3.00 ]
    [ 3.80 14.30 1.00 ]
GET1 DELI GRUP UNGR
```

[STAT] [STOΣ]

```
3:
2:
1:
Σ+ Σ- NΣ CLΣ STOΣ RCLΣ
```

Now transform the grouped data into its ungrouped form.

[USER] [XFRM]

```
3:
2:
1:
ΣDAT GD SDVP VARP COVP XFRM
```

Purge *GD*, since the original data remains in *GRUP*.

Compute the statistics.

'GD PURGE
STAT CORR

```
3:
2:
1:                   -0.62
COLΣ CORR COV LR PREDV
```

The correlation coefficient is calculated with a built-in function.

The coefficients of variation are calculated by entering the following program.

```
«SDEV ARRY→ DROP
MEAN ARRY→ LIST→
DROP → n « n 1 FOR x
x n + ROLL n 1 +
ROLL ÷ 100 × -1 STEP
n 1 →LIST →ARRY
```

ENTER <>

```
1: « SDEV ARRY→ DROP
   MEAN ARRY→ LIST→
   DROP → n « n 1.00
   FOR x x n + ROLL n
```

Store the program in the variable *VCO*.

'VCO STO
CLEAR

```
4:
3:
2:
1:
```

Compute the coefficients of variation.

USER VCO

```
3:
2:
1:         [ 9.93 8.42 ]
VCO ΣDAT SDVP VARP COVP XFRM
```

The statistics matrices *GRUP* and *UNGR* are not used in further examples. Purge them and the variables *VCO*, *SDVP*, *VARP*, and *COVP*.

{ 'GRUP' 'UNGR' 'VCO' 'SDVP' 'VARP' 'COVP PURGE

Sums of Products

The HP-28C matrix functions provide an easy method of computing sums of products of the statistical data. The product of two columns of statistics with its matrix transpose will produce the result shown below.

Let the current statistics matrix be

$$M = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

Then $M^T \cdot M$ is an $m \times m$ matrix of the form

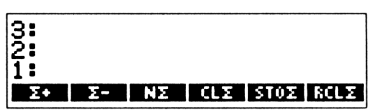
$$\begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \cdots & \sum x_{i1}x_{im} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 & \cdots & \sum x_{i2}x_{im} \\ \cdots & \cdots & \cdots & \cdots \\ \sum x_{i1}x_{im} & \sum x_{i2}x_{im} & \cdots & \sum x_{im}^2 \end{bmatrix}$$

Example: Compute the sums of products on all pairings of the following data.

$$\begin{bmatrix} 10 & 4 & 7 \\ 20 & 5 & 3 \\ 3 & 9 & 2 \\ 5 & 2 & 1 \\ 7 & 4 & 5 \end{bmatrix}$$

Key in the statistics data.

CLEAR
STAT CLΣ
[10,4,7 Σ+
[20,5,3 Σ+
[3,9,2 Σ+
[5,2,1 Σ+
[7,4,5 Σ+



Compute the sums of products.

RCLΣ RCLΣ
ARRAY TRN SWAP ×

```
1: [[ 583 205 176 ]
    [ 205 142 83 ]
    [ 176 83 88 ]]
[STD] FIX SCI ENG DEG [RAD]
```

MODE STD

Normalized Data

A column of data in the current statistics matrix can be normalized by transforming each element as shown below.

$$x_j' = \frac{x_j}{\sum_{i=1}^{n} x_i}$$

x_j' is created by dividing the original x_j by the sum of the column data.

Example: Compute a normalized vector from the second column of the data of the previous section.

First, use program *GET1* to extract the second column.

[CLEAR]
2 [USER] ≣ GET1 ≣

```
3:
2:
1:           [ 4 5 9 2 4 ]
 ΣDAT  XFRM  ΣPAR  GET1  DELI
```

Key in the program below. The program divides each element of the vector by the sum of the absolute values of each of the elements.

« DUP CNRM INV ×
[ENTER] [<>]

```
4:
3:
2:
1:     « DUP CNRM INV * »
```

Now execute the program.

[EVAL]

```
4:
3:
2:
1: [ .166666666667 .20...
```

≣ ARRY→ ≣ will break the vector into component form for examination. The program can be stored for repeated use if desired.

Delta Percent on Paired Data

The "delta percent" of a pair of columnar data can be computed by the program below.

$$\Delta\% = \frac{new - old}{old}.$$

Old represents the first column of data specified in ΣPAR, and *new* represents the second column of data specified in ΣPAR.

Delta Percent Program

For this program to work properly, you must first create the variable *ΣPAR*. If *ΣPAR* does not appear in the USER menu, you can create it by the keystrokes `1,2 COLΣ`. If you are working with only two columns, you can simplify the program below by removing the flexibility to specify the columns to be used in the computation (i.e., `ΣPAR 2 GET` can be replaced by the column number desired, and similarly for the first column.)

The program below assumes the program *GET1* is already resident in the HP-28C.

Program	**Comments**
`«ΣPAR 2 GET GET1 ARRY→ DROP ΣPAR 1 GET GET1 ARRY→ DROP`	Get the data from the two columns on the stack.
`NΣ 1 FOR x x NΣ + ROLL → new`	Roll down and store *new*.
`« NΣ ROLL → old`	Roll down and store *old*.
`«'(new-old)÷old' EVAL » »`	Compute the delta percent.
`-1 STEP NΣ 1 2 →LIST →ARRY »`	Count down until complete, then structure the data into a column.

[ENTER] `'DLTA` [STO]

Example: Compute the delta percent between columns one and three of the data in the 'Sums of Products' section.

Select columns one and three.

`1,3` [STAT] COLΣ

```
3:
2:
1:
COLΣ CORR COV LR PREDV
```

Now compute the delta percent between the pair of columns.

[USER] DLTA

```
1: [[ -.3 ]
    [ -.85 ]
    [ -.333333333333 ]
ΣDAT DLTA XFRM ΣPAR GET1 DELI
```

Purge variables created in this section.

`{'DLTA''XFRM''DELI''ΣPAR` [PURGE]

Moments, Skewness, and Kurtosis

For grouped or ungrouped data, moments are used to describe sets of data, skewness is used to measure the lack of symmetry in a distribution, and kurtosis is the relative peakness or flatness of a distribution.

For a given set of data

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}$$

the moments and moment coefficients are calculated by the following expressions.

1st Moment

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2nd Moment

$$m_2 = \frac{1}{n}\sum x_i^{\,2} - \overline{x}^{\,2}$$

3rd Moment

$$m_3 = \frac{1}{n}\sum x_i^{\,3} - \frac{3}{n}\overline{x}\sum x_i^{\,2} + 2\overline{x}^{\,3}$$

4th Moment

$$m_4 = \frac{1}{n}\sum x_i^{\,4} - \frac{4}{n}\overline{x}\sum x_i^{\,3} + \frac{6}{n}\overline{x}^{\,2}\sum x_i^{\,2} - 3\overline{x}^{\,4}$$

Moment Coefficient of Skewness

$$\gamma_1 = \frac{m_3}{m_2^{3/2}}$$

Moment Coefficient of Kurtosis

$$\gamma_2 = \frac{m_4}{m_2^{2}}$$

Calculating Moments, Skewness and Kurtosis

Two programs for computing moments, skewness, and kurtosis are described below. The first program requires specification of a column of data for the current statistics matrix. It calls the second program repeatedly to compute various sums of powers of the columnar data.

The program below assumes the program *GET1* is already resident in the HP-28C.

Program	**Comments**
`«GET1 NΣ 1 2 →LIST RDM`	Retrieve the specified column and transform it into a column vector.
`RCLΣ SWAP STOΣ`	Save the current statistics matrix on the stack, store the column vector.
`2 SUMO NΣ ÷ MEAN SQ - 'M2' STO`	Compute the second moment.
`3 SUMO 2 SUMO MEAN × 3 × - NΣ ÷ MEAN 3 ^ 2 × + 'M3' STO`	Compute the third moment.
`4 SUMO 3 SUMO 4 × 2 SUMO 6 × MEAN × - MEAN × - NΣ ÷ MEAN 4 ^ 3 × - 'M4' STO`	Compute the fourth moment.
`'M3÷M2^1.5' EVAL 'GM1' STO`	Compute the coefficient of skewness.
`'M4÷M2^2' EVAL 'GM2' STO`	Compute the coefficient of kurtosis.
`STOΣ »`	Restore the complete statistics array.

[ENTER] `'MOMS` [STO]

Program	**Comments**
`«→ p « RCLΣ ARRY→ DROP`	Store the power for $\sum x^p$ and place the data separately on the stack.
`0 1 NΣ START`	Zero for sum accumulation, set up loop count.
`SWAP p ^ +`	Compute x_i^p and sum.
`NEXT » »`	Complete loop, end program.

[ENTER] `'SUMO` [STO]

Example: Compute the first through fourth moments and the coefficients of skewness and kurtosis for the data below.

x_i	y_i
---	---
2.1	1.1
3.5	3.8
4.2	4.4
6.5	9.7
4.1	3.2
3.6	2.2
5.3	1.6
3.7	5.0
4.9	3.7

Clear the current statistics matrix and enter the data.

CLEAR STAT CLΣ
```
[2.1,1.1 Σ+
[3.5,3.8 Σ+
[4.2,4.4 Σ+
[6.5,9.7 Σ+
[4.1,3.2 Σ+
[3.6,2.2 Σ+
[5.3,1.6 Σ+
[3.7,5 Σ+
[4.9,1.7 Σ+
```

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Compute the first moments.

MEAN MODE 2 FIX

```
3:
2:
1:              [ 4.21 3.63 ]
 STD [ FIX ] SCI | ENG | DEG [ RAD ]
```

Specify the first column and compute the other x_i moments and coefficients.

1 USER MOMS

```
3:
2:
1:              [ 4.21 3.63 ]
 GM2 | GM1 | M4 | M3 | M2 | ΣDAT
```

Display the second, third, and fourth moments.

M2 M3 M4

```
3:                1.39
2:                0.39
1:                5.49
GM2 GM1 M4 M3 M2 ΣDAT
```

Display the coefficients of skewness and kurtosis.

GM1 GM2

```
3:                5.49
2:                0.24
1:                2.84
GM2 GM1 M4 M3 M2 ΣDAT
```

Repeat the process for the second column of data y_i.

CLEAR 2 MOMS
M2 M3 M4

```
3:                6.20
2:               21.26
1:              159.43
GM2 GM1 M4 M3 M2 ΣDAT
```

GM1 GM2

```
3:              159.43
2:                1.38
1:                4.14
GM2 GM1 M4 M3 M2 ΣDAT
```

Purge variables created in this section.

`{ 'MOMS' 'SUMO' 'M2' 'M3' 'M4' 'GM1' 'GM2` PURGE

Regression

A variety of regression techniques are performed easily with the HP-28C. Its built-in matrix manipulation and system solution capabilities, coupled with data and curve plotting make the HP-28C a very capable tool for regression analysis.

Curve Fitting

This problem section describes programs to compute linear, exponential, logarithmic, and power curve fits to a set of data points in the current statistics matrix. Any or all of the curve types may be selected to find a 'best' fit. The data and regression equation may be plotted, and estimates from the regression equation are easily computed with the Solver.

The programs and instructions that follow are designed for flexibility in trying different types of regressions on the same data. If your analysis requirements are for linear regression only, you should use the built-in commands for linear regression, described in the owner's documentation.

For a set of data points (x_i, y_i), the regression equations for four types of curves are shown below.

Straight Line (Linear Regression)

$$y = a + bx.$$

Exponential Curve

$$y = ae^{bx} \quad \text{where } a > 0.$$

Logarithmic Curve

$$y = a + b\ln(x)$$

Power Curve

$$y = ax^b \quad \text{where } a > 0.$$

The regression coefficients a and b are found by solving the following system of linear equations.

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} A \\ b \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix}$$

where the variables are defined below.

Regression	A	X_i	Y_i
Linear	a	x_i	y_i
Exponential	$\ln a$	x_i	$\ln y_i$
Logarithmic	a	$\ln x_i$	y_i
Power	$\ln a$	$\ln x_i$	$\ln y_i$

The coefficient of determination is

$$R^2 = \frac{A\sum Y_i + b\sum X_i Y_i - \frac{1}{n}(\sum Y_i)^2}{\sum(Y_i^2) - \frac{1}{n}(\sum Y_i)^2}$$

The programs below apply the least squares method, either to the original data or the transformed data as described above. For all regression types, the original data is restored after the computation of the regression equation. This allows for multiple regression types to be tried on the same data set.

Key in the four programs below. These programs define the data transformations and the equation-generating transforms for the general curve fitting program.

```
« « » « × + »
  FIT [ENTER]
« « LN » « × EXP SWAP
  EXP × » FIT [ENTER]
« « SWAP LN SWAP »
  « LN × + » FIT [ENTER]
« « LN SWAP LN SWAP »
  « SWAP ^ SWAP EXP × »
  FIT [ENTER]
```

```
1: «  « LN SWAP LN SWAP
   »  « SWAP ^ SWAP EXP
   * » FIT »
 LOG  ALOG  LN  EXP  LNP1  EXPM
```

Store the programs above. Note that the *LIN* program defines a 'null' transform to the data.

```
[USER]
'PWR [STO]
'LOGF [STO]
'EXPF [STO]
'LIN [STO]
```

```
3:
2:
1:
 LIN  EXPF  LOGF  PWR  GET1
```

Key in the general curve fitting program below. The program uses the transforms defined above in the calling programs, and returns the regression equation and coefficient of determination for a measure of the goodness of fit.

Program	**Comments**
`«→ xf1 xf2`	Store data transform and equation generator.
`«RCLΣ 'TMP' STO NΣ → n`	Save original data, get data count for looping.
`« 1 n START Σ- ARRY→ DROP xf1 EVAL NEXT`	Transform original data onto stack.
`n 2 2 →LIST →ARRY STOΣ »`	Put transformed data into current statistics matrix.
`LR DUP2 RCLΣ TRN RCLΣ ×`	Compute A and b, duplicate for regression equation and R^2.
`{2 1} GET × TOT {2} GET SQ NΣ ÷ - SWAP TOT {2} GET × +`	Compute numerator of R^2.
`NΣ ÷ MEAN Σ+ VAR {2} GET ÷`	Compute denominator of R^2 ($n \cdot VAR_p(Y_i)$.)
`TMP 'TMP' PURGE STOΣ`	Restore original data, purge temporary variable.
`ROT ROT 'X' xf2 EVAL STEQ RCEQ » »`	Generate regression equation and store.

[ENTER] 'FIT [STO]

Key in the plotting program below. The program scales the plotting region by the statistics data and overlays the data and curve plots.

«CLLCD SCLΣ DRWΣ DRAW [ENTER]

```
2:
1: « CLLCD SCLΣ DRWΣ
   DRAW »
STEQ RCEQ PMIN PMAX INDEP DRAW
```

Store the program in the variable *PLOT*.

'PLOT [STO]

```
3:
2:
1:
STEQ RCEQ PMIN PMAX INDEP DRAW
```

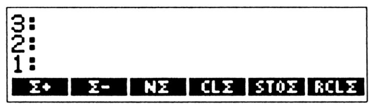
You may choose to enhance the program above by modifying the axes position according to the data set (e.g., such as the midpoints between the minimum and maximum points).

Example: Fit the following set of data into a straight line.

x_i	y_i
----	----
40.5	104.5
38.6	102
37.9	100
36.2	97.5
35.1	95.5
34.6	94

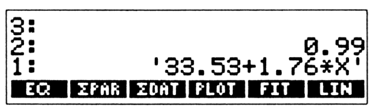
Clear the current statistics matrix and enter the data.

STAT CLΣ
[40.5,104.5 Σ+
[38.6,102 Σ+
[37.9,100 Σ+
[36.2,97.5 Σ+
[35.1,95.5 Σ+
[34.6,94 Σ+



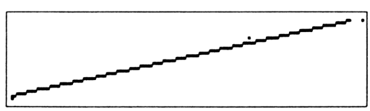
Compute the regression equation and coefficient of determination.

USER LIN



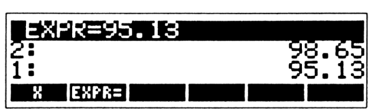
Plot the equation.

PLOT



Find estimates for $\hat{y}$ at $x = 37$ and $x = 35$.

ATTN SOLV SOLVR
37 X EXPR=
35 X EXPR=

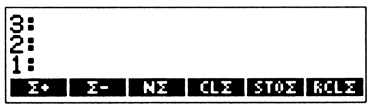


Example: Fit the following set of data into an exponential curve.

x_i	y_i
----	----
.72	2.16
1.31	1.61
1.95	1.16
2.58	.85
3.14	0.5

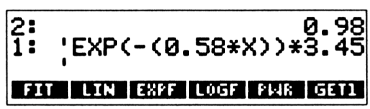
Clear the current statistics matrix and enter the data.

CLEAR
STAT CLΣ
[.72,2.16 Σ+
[1.31,1.61 Σ+
[1.95,1.16 Σ+
[2.58,.85 Σ+
[3.14,.5 Σ+



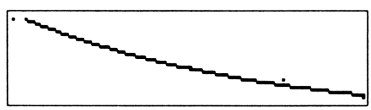
Compute the regression equation and coefficient of determination.

USER EXPF



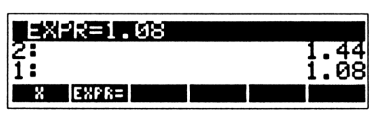
Plot the equation.

PLOT



Find estimates for $\hat{y}$ at $x = 1.5$ and $x = 2$.

ATTN SOLV SOLVR
1.5 X EXPR=
2 X EXPR=



Example: Fit the following set of data into a logarithmic curve.

x_i	y_i
3	1.5
4	9.3
6	23.4
10	45.8
12	60.1

Clear the current statistics matrix and enter the data.

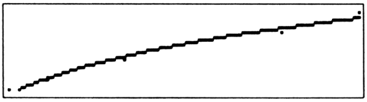
CLEAR
STAT CLΣ
[3,1.5 Σ+
[4,9.3 Σ+
[6,23.4 Σ+
[10,45.8 Σ+
[12,60.1 Σ+

Compute the regression equation and coefficient of determination.

USER LOGF

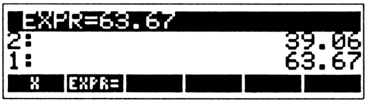
Plot the equation.

PLOT



Find estimates for $\hat{y}$ at $x = 8$ and $x = 14.5$.

ATTN SOLV SOLVR
8 X EXPR=
14.5 X EXPR=



Example: Fit the following set of data into a power curve.

x_i	y_i
----	----
10	.95
12	1.05
15	1.25
17	1.41
20	1.73
22	2.00
25	2.53
27	2.98
30	3.85
32	4.59
35	6.02

Clear the current statistics matrix and enter the data.

CLEAR
STAT CLΣ
[10,.95 Σ+
[12,1.05 Σ+
[15,1.25 Σ+
[17,1.41 Σ+
[20,1.73 Σ+
[22,2 Σ+
[25,2.53 Σ+
[27,2.98 Σ+
[30,3.85 Σ+
[32,4.59 Σ+
[35,6.02 Σ+

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

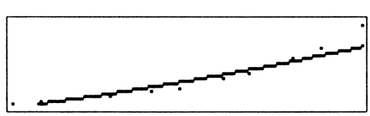
Compute the regression equation and coefficient of determination.

USER PWR

```
3:
2:                        0.94
1:            'X^1.46*0.03'
 EXPF | LOGF | PWR | GET1 |  |
```

Plot the equation.

PLOT



Find estimates for $\hat{y}$ at $x = 18$ and $x = 23$.

[ATTN] [SOLV] [SOLVR]
18 [X] [EXPR=]
23 [X] [EXPR=]

```
EXPR=2.52
2:                    1.76
1:                    2.52
 X  EXPR=
```

Multiple Regressions on the Same Data

Because the original, un-transformed data is restored to the current statistics matrix, repeated and different regressions can be tried on the same data set. The equation plots can also be overlayed with a simple program like the one that follows.

Example: Plotting Multiple Regressions

For the data entered for the power curve fit in the preceding example, plot the curves for both power and exponential regressions, and compare their relative coefficients of determination.

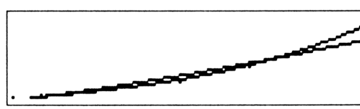
The program below performs the power curve fit, plots the data and curve, performs the exponential fit and draws it.

[CLEAR]
« PWR PLOT EXPF DRAW
[ENTER]

```
2:
1: « PWR PLOT EXPF DRAW
   »
STEQ RCEQ PMIN PMAX INDEP DRAW
```

Execute the program. Note: You may find it necessary to purge unused variables to provide sufficient space in the HP-28C for both the curve fitting program and graphics display memory.

[EVAL]



Now compare the equations and the coefficients of determination.

[ATTN] [<>]

```
4:                    0.94
3:         'X^1.46*0.03'
2:                    0.99
1:     'EXP(0.07*X)*0.41'
```

The exponential curve is a better fit.

Program *PLOT* is used in the Polynomial Regression section. Purge the other variables created in the section.

Multiple Linear Regression

This problem section provides a program for computing regression coefficients to a linear equation in two or three independent variables by the least squares method. The coefficient of determination is also computed, and point estimates based on the regression line can be computed.

Two Independent Variables

For a set of data points (x_i, y_i, t_i), the linear equation has the form

$$t = a + bx + cy \ .$$

Regression coefficients a, b, and c are calculated by solving the following system of equations.

$$\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum y_i x_i & \sum y_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum t_i \\ \sum x_i t_i \\ \sum y_i t_i \end{bmatrix}$$

The coefficient of determination is defined below.

$$R^2 = \frac{a\sum t_i + b\sum x_i t_i + c\sum y_i t_i - \frac{1}{n}(\sum t_i)^2}{\sum t_i^2 - \frac{1}{n}(\sum t_i)^2}$$

Three Independent Variables

For a set of data points (x_i, y_i, z_i, t_i), the linear equation has the form

$$t = a + bx + cy + dz \ .$$

Regression coefficients a, b, c and d are calculated by solving the following system of equations.

$$\begin{bmatrix} n & \sum x_i & \sum y_i & \sum z_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum y_i & \sum y_i x_i & \sum y_i^2 & \sum y_i z_i \\ \sum z_i & \sum z_i x_i & \sum z_i y_i & \sum z_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \sum t_i \\ \sum x_i t_i \\ \sum y_i t_i \\ \sum z_i t_i \end{bmatrix}$$

The coefficient of determination is defined below.

$$R^2 = \frac{a\sum t_i + b\sum x_i t_i + c\sum y_i t_i + d\sum z_i t_i - \frac{1}{n}(\sum t_i)^2}{\sum t_i{}^2 - \frac{1}{n}(\sum t_i)^2}$$

The following minimum condition for the number of data points n must be satisfied:

- $n \geq 3$ for the case of two independent variables
- $n \geq 4$ for the case of three independent variables.

Multiple Linear Regression Program

The program below finds the regression line for both two and three independent variables. It also calculates R^2.

Program	Comments
`« RCLΣ SIZE LIST→ DROP → m`	Begin to build the left-most matrix in the system of equations. m is the number of elements in each data vector.
`« 1 →LIST 1 CON ARRY→ DROP`	Generate NΣ 1's on the stack.
`RCLΣ TRN ARRY→ LIST→ DROP SWAP 1 + SWAP 2 →LIST →ARRY`	Combine the 1's into the array.
`DUP TRN ×`	Generate the first row and column and covariance data.
`m m 1 + 2 →LIST RDM`	Drop the last row by redimensioning.
`TRN ARRY→ DROP m 1 2 →LIST →ARRY → rhs`	Pull out the right-hand-side of the system solution and save.
`« m m 2 →LIST →ARRY INV rhs ×`	Form the left-hand-side matrix, invert, and compute the regression coefficients.
`DUP TRN rhs × ARRY→ DROP`	Compute the first m terms of the numerator of R^2.
`TOT m 1 →LIST GET SQ NΣ ÷ -`	Complete the numerator.
`NΣ ÷ MEAN Σ+ VAR m 1 →LIST GET ÷`	The denominator of R^2 is $n \cdot VAR_{population}$.
`Σ- DROP » » »`	Undo the change made to compute the population variance.

[ENTER] 'MLR [STO]

Example: Find the regression coefficients and coefficient of determination for the following set of data.

x_i	y_i	z_i	t_i
7	25	6	60
1	29	15	52
11	56	8	20
11	31	8	47
7	52	6	33

Clear the current statistics matrix and enter the data.

```
CLEAR
STAT  CLΣ
[7,25,6,60  Σ+
[1,29,15,52  Σ+
[11,56,8,20  Σ+
[11,31,8,47  Σ+
[7,52,6,33  Σ+
```

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Compute the regression coefficients and coefficient of determination.

```
USER  MLR
MODE  4  FIX
```

```
3:
2: [[ 103.4473 ]  [ -1…
1:                0.9989
 STD [ FIX ] SCI  ENG  DEG [ RAD ]
```

The coefficient of determination is 0.9989 .

Drop it and display the values for a , b , c , and d .

```
DROP  <>
```

```
1: [[ 103.4473 ]
    [ -1.2841  ]
    [ -1.0369  ]
    [ -1.3395  ]]
```

The regression line is $t = 103.4473 - 1.2841x - 1.0369y - 1.3395z$. You can use the HP-28C's algebraic features to generate this equation from the matrix displayed. This would allow solving for any variable of the equation, given values for the remaining variables.

You can also compute estimates for $\hat{t}$ by multiplying the regression coefficients matrix by a matrix of values for the independent variables.

Example: Find t for $x=7$, $y=25$, and $z=6$, and $x=1$, $y=29$, and $z=15$ for the problem above.

First make a copy of the coefficient matrix for the two computations of $\hat{t}$. Enter the first set of values for the independent variables. Note that a 1 is entered for the multiplication with the coefficient a.

ENTER
[[1,7,25,6 SWAP ×

```
4:
3:
2: [[ 103.4473 ]  [ -1…
1:          [[ 60.4985 ]]
```

The estimate $\hat{t}$ is 60.4985 .

Compute $\hat{t}$ for the second set of values.

DROP
[[1,1,29,15 SWAP ×

```
4:
3:
2:
1:          [[ 52.0000 ]]
```

Example: Find the regression line and the coefficient of determination for the following data.

x_i	y_i	t_i
1.5	0.7	2.1
0.45	2.3	4.0
1.8	1.6	4.1
2.8	4.5	9.4

Clear the stack and the current statistics matrix; enter the data.

CLEAR STAT CLΣ
[1.5,.7,2.1 Σ+
[.45,2.3,4 Σ+
[1.8,1.6,4.1 Σ+
[2.8,4.5,9.4 Σ+

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Find the regression line and coefficient of determination.

USER MLR

```
3:
2: [[ -0.0971 ]  [ 0.7…
1:                0.9984
 ΣDAT  MLR  PLOT  GET1
```

DROP

```
1: [[ -0.0971 ]
    [ 0.7914 ]
    [ 1.6269 ]]
ΣDAT | MLR | PLOT | GET1 |  | 
```

The regression line is $t = -.0971 + .7914x + 1.6269y$. The same techniques described in the previous example may be used for computing $\hat{t}$.

Save programs *MLR* and *PLOT* for the Polynomial Regression section. Purge the other variables created in this section.

Polynomial Regression

This problem section provides a general program for computing regression coefficients to a parabolic and cubic equation for a set of paired data points by the least squares method. The coefficient of determination is also computed, and point estimates based on the regression equation can be computed.

Parabolic Regression

For a set of data points (x_i, y_i), the parabolic equation has the form

$$y = a + bx + cx^2 .$$

Regression coefficients a, b, and c are calculated by solving the following system of equations.

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

The coefficient of determination is defined below.

$$R^2 = \frac{a\sum y_i + b\sum x_i y_i + c\sum x_i^2 y_i - \frac{1}{n}(\sum y_i)^2}{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2}$$

Cubic Regression

For a set of data points (x_i, y_i), the cubic equation has the form

$$y = a + bx + cx^2 + dx^3 .$$

Regression coefficients a, b, c and d are calculated by solving the following system of equations.

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \sum x_i^3 y_i \end{bmatrix}$$

The coefficient of determination is defined below.

$$R^2 = \frac{a\sum y_i + b\sum x_i y_i + c\sum x_i^2 y_i + d\sum x_i^3 y_i - \frac{1}{n}(\sum y_i)^2}{\sum(y_i^2) - \frac{1}{n}(\sum y_i)^2}$$

The following minimum condition for the number of data points n must be satisfied:

- $n \geq 3$ for Parabolic Regression
- $n \geq 4$ for Cubic Regression.

Polynomial Regression Programs

The programs below transform the data to a form directly useable in the Multiple Linear Regression program from the previous section. By modifying the statistics data inputs to the form $[x_i \; x_i^2 \; y_i]$, the Multiple Linear Regression Program, *MLR*, computes the regression coefficients and coefficient of determination for parabolic regression. Similarly, by including an x_i^3 term, the *MLR* program computes the coefficients for a cubic regression.

Program	**Comments**
`« SWAP DUP SQ 3 ROLL { 3 } →ARRY Σ+ »`	Form $[x_i \; x_i^2 \; y_i]$ and accumulate in the current statistics matrix.

[ENTER] ' PARA [STO]

Program	**Comments**
`« SWAP DUP DUP SQ SWAP 3 ^ 4 ROLL {4} →ARRY Σ+ »`	Form [x_i x_i^2 x_i^3 y_i] and accumulate in the current statistics matrix.

[ENTER] 'CUB [STO]

Example: Find the regression coefficients and coefficient of determination for the following set of data.

$$\begin{bmatrix} x_i & y_i \\ \text{---} & \text{---} \\ .8 & 24 \\ 1 & 20 \\ 1.2 & 10 \\ 1.4 & 13 \\ 1.6 & 12 \end{bmatrix}$$

Clear the current statistics matrix and enter the data.

[CLEAR]
[STAT] CLΣ
[USER]
.8,24 CUB
1,20 CUB
1.2,10 CUB
1.4,13 CUB
1.6,12 CUB

```
3:
2:
1:
ΣDAT CUB PARA MLR PLOT GET1
```

Compute the cubic regression coefficients and coefficient of determination.

MLR

```
3:
2: [[ 47.9429 ]  [ -9....
1:               0.8685
ΣDAT CUB PARA MLR PLOT GET1
```

The coefficient of determination is 0.8685.

Drop it and display the values for a, b, c, and d.

[DROP] [<>]

```
1: [[ 47.9429 ]
    [ -9.7619 ]
    [ -41.0714 ]
    [ 20.8333 ]]
```

The regression equation is $y = 47.9429 - 9.7619x - 41.0714x^2 + 20.8333x^3$. You can use the HP-28C's algebraic features to generate this equation from the matrix displayed. This would allow solving for any variable of the equation, given values for the remaining variables.

You can also compute estimates for $\hat{y}$ by multiplying the regression coefficients matrix by a matrix of values for the independent variables.

Example: Find $\hat{y}$ for $x = 1$ and $x = 1.4$ for the problem above.

Enter the regression equation and use Solver to compute $\hat{y}$.

'47.9429-9.7619×X-
41.0714×X^2+20.8333×X^3
ENTER

```
2: [[ 47.9429 ]  [ -9.…
1: '47.9429-9.7619*X-
   41.0714*X^2+20.8333*
   X^3'
```

Store the equation and compute $\hat{y}$ for $x = 1$.

SOLV STEQ SOLVR
1 X
EXPR=

```
EXPR=17.9429
2: [[ 47.9429 ]  [ -9.…
1:              17.9429
 X  EXPR=
```

Repeat for $x = 1.4$.

1.4 X
EXPR=

```
EXPR=10.9429
2:              17.9429
1:              10.9429
 X  EXPR=
```

With the regression equation entered above, compare the orginal statistics data to a plot of the equation.

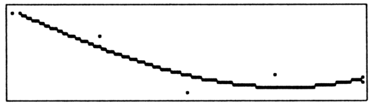
First, clear the current statistics matrix and enter the original data. (For larger matrices, the use of *GET1* and the ARRAY menu functions can be used to extract the first and last columns of data and construct the statistics matrix without re-entry of the data).

CLEAR
STAT CLΣ
[.8,24 Σ+
[1,20 Σ+
[1.2,10 Σ+
[1.4,13 Σ+
[1.6,12 Σ+

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Use the *PLOT* program from the "Curve Fitting" section to plot the data and the cubic equation.

[USER] ≣PLOT≣



Example: Find the parabolic equation and the coefficient of determination for the following data.

x_i	y_i
---	---
1	5
2	12
3	34
4	50
5	75
6	84
7	128

Clear the stack and the current statistics matrix; enter the data.

[CLEAR] [STAT] ≣CLΣ≣
[USER]
`1,5` ≣PARA≣
`2,12` ≣PARA≣
`3,34` ≣PARA≣
`4,50` ≣PARA≣
`5,75` ≣PARA≣
`6,84` ≣PARA≣
`7,128` ≣PARA≣

```
3:
2:
1:
PARA MLR PLOT GET1
```

Find the parabolic regression equation and coefficient of determination.

≣MLR≣
[DROP] [<>]

```
2:
1: [[ -4.0000 ]
    [ 6.6429 ]
    [ 1.6429 ]]
```

The regression line is $y = -4.0000 + 6.6429x + 1.6429x^2$. The same techniques described in the previous example may be used for computing $\hat{y}$.

Test Statistics and Confidence Intervals

Decisions based on sample data can be directed with the use of test statistics. A variety of test statistics for different hypotheses and assumptions can be calculated with the HP-28C. This section presents three such test statistics – paired t statistic, t statistic for two means, and chi-square statistic. Additional test statistics for different hypotheses are readily computed with similar, simple procedures. The test statistics are used in conjunction with the upper-tail probability commands of the HP-28C to determine confidence intervals.

Paired t Statistic

Given a set of paired observations from two normal populations with unknown means μ_1, μ_2

$$\begin{bmatrix} x_i & y_i \\ \text{---} & \text{---} \\ x_1 & y_1 \\ x_2 & y_2 \\ \cdots & \cdots \\ x_n & y_n \end{bmatrix}$$

the test statistic

$$t = \frac{\overline{D}}{s_D} \cdot \sqrt{n}$$

with $n - 1$ degress of freedom can be used to test the null hypothesis

$$H_0\colon \mu_1 = \mu_2 .$$

The variable definitions are

$$D_i = x_i - y_i$$

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$$

and

$$s_D = \left(\frac{\sum D_i^{\,2} - \frac{1}{n}(\sum D_i)^2}{n - 1} \right)^{\frac{1}{2}} .$$

Example: Test the null hypothesis that $\mu_1 = \mu_2$ for the following data pairs.

$$\begin{bmatrix} x_i & y_i \\ \text{---} & \text{---} \\ 14 & 17 \\ 17.5 & 20.7 \\ 17 & 21.6 \\ 17.5 & 20.9 \\ 15.4 & 17.2 \end{bmatrix}$$

Clear the current statistics matrix and key in the data.

[CLEAR] [MODE] 2 FIX

STAT CLΣ

```
[14,17 Σ+
[17.5,20.7 Σ+
[17,21.6 Σ+
[17.5,20.9 Σ+
[15.4,17.2 Σ+
```

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Use *GET1* to recover the two columns of data.

1 [USER] GET1

2 GET1

```
3:
2: [ 14.00 17.50 17.00...
1: [ 17.00 20.70 21.60...
 ΣDAT  GET1
```

Compute the difference of the data pairs.

[-]

```
3:
2:
1: [ -3.00 -3.20 -4.60...
 ΣDAT  GET1
```

Redimension and store the difference matrix as the current statistical matrix. (Save the original data if so desired.)

[STAT] NΣ 1 2 [LIST] →LIST

[ARRAY] RDM [STAT] STOΣ

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Compute the t statistic.

MEAN SDEV
STACK DUP2
÷ STAT NΣ √ ×

```
3:                -3.20
2:                 1.00
1:                -7.16
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

The mean $\overline{D}$ is -3.20 .
s_D is 1.00 .
t is -7.16 .
The degress of freedom are 4.00 .

Example: Determine if the hypothesis H_0 of the previous problem should be rejected at a 0.05 level of significance.

First create a program for a general solution to the Student's t distribution, as described in the "STAT" section of the Reference Manual.

CLEAR
« P N X UTPT - »
ENTER

```
3:
2:
1:       « P N X UTPT - »
 UTPC  UTPF  UTPN  UTPT
```

Store it for use in the Solver.

SOLV STEQ SOLVR

```
3:
2:
1:
 P  N  X  EXPR=
```

Enter the degrees of freedom and the level of significance. Note that for a two-tailed test at a 0.05 level of significance, you compute the value for the confidence interval $-t_{.975}$ to $t_{.975}$.

4 N .025 P
□ X

```
X: 2.78
Sign Reversal
1:                 2.78
 P  N  X  EXPR=
```

Thus the hypothesis is rejected since t falls outside the range $(-2.78,2.78)$.

Example: Compute the level of significance for which the hypothesis H_0 will be accepted.

Enter the t statistic computed in the first example. Note that the absolute value is input, corresponding to the upper-tail portion of the probability function.

`7.16` [X]
[ ] [P]
`2` [×]

```
3:
2:                2.78
1:            2.01E-3
  P    N    X   EXPR=
```

The probability is multiplied by 2 for the upper- and lower-tails of the probability function outside the range (−7.16,7.16).

Rather than using the program from the previous example, you can also compute the level of significance directly with the UTPT command. The keystrokes

`4,7.16` [STAT] [UTPT] `2` [×]

generate the same result as above.

t Statistic for Two Means

Suppose $\{x_1, x_2, \cdots, x_{n1}\}$ and $\{y_1, y_2, \cdots, y_{n2}\}$ are two independent random samples from two normal populations, with unknown means μ_1 and μ_2 and the same unknown variance σ^2.

The null hypothesis

$$H_0: \mu_1 - \mu_2 = d$$

can be tested with the t statistic

$$t = \frac{\overline{x} - \overline{y} - d}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} \left(\frac{\sum x_i^2 - n_1 \overline{x}^2 + \sum y_1^2 - n_2 \overline{y}^2}{n_1 + n_2 - 2}\right)^{\frac{1}{2}}}.$$

This t statistic has the t distribution with $n_1 + n_2 - 2$ degrees of freedom for testing the null hypothesis H_0.

Example: Test the null hypothesis that $H_0: \mu_1 = \mu_2$ (i.e. $d=0$) for the data below.

x_i	y_i
---	---
79	91
84	103
108	90
114	113
120	108
103	87
122	100
120	80
	99
	54

Clear the current statistics matrix and accumulate the x data.

CLEAR
STAT CLΣ
79 Σ+
84 Σ+
108 Σ+
114 Σ+
120 Σ+
103 Σ+
122 Σ+
120 Σ+

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Compute the mean, variance and number of data points and store.

MEAN 'MX STO
VAR 'VX STO
NΣ 'NX STO

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Clear the current statistics matrix and accumulate the y data.

CLΣ
91 Σ+
103 Σ+
90 Σ+
113 Σ+
108 Σ+
87 Σ+
100 Σ+
80 Σ+
99 Σ+
54 Σ+

```
3:
2:
1:
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Compute the t statistic. First compute the numerator. Recall that $d = 0$.

USER MX STAT MEAN −

```
3:
2:
1:                         13.75
 TOT | MEAN | SDEV | VAR | MAXΣ | MINΣ
```

Compute the first part of the denominator and divide.

USER NX 1/x
STAT NΣ 1/x
+ √ ÷

```
3:
2:
1:                         28.99
 Σ+ | Σ- | NΣ | CLΣ | STOΣ | RCLΣ
```

Compute the second part of the denominator and divide.

NΣ 1 [-]
VAR [×]
[USER] NX 1 [-]
VX [×] [+]
NX [STAT] NΣ [+] 2 [-]
[÷] [√] [÷]

```
3:
2:
1:                 1.73
 Σ+  Σ-  NΣ  CLΣ STOΣ RCLΣ
```

The t statistic is 1.73 with 16 degrees of freedom.

Example: Compute the level of significance for the two-tailed test on the range $(-1.73, 1.73)$ from the preceding example.

16,1.73 UTPT
2 [×]

```
3:
2:                 1.73
1:                 0.10
UTPC UTPF UTPN UTPT
```

The hypothesis cannot be rejected at, or below, this level of significance.

You may also choose to test the assumption made in this problem section that the unknown variances are equal. For this purpose, compute the F statistic

$$\frac{s_{max}^2}{s_{min}^2}$$

with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.
s_{max}^2 is the maximum of the sample variances s_1^2 and s_2^2.
s_{min}^2 is the minimum of the two sample variances.

You can then compute the level of significance with the UTPF command, using the same approach as the example above.

Purge the variables created in this section.

{ 'NX' 'VX' 'MX [PURGE]

Chi-Square Statistic

This section provides a simple program for computing the χ^2 statistic for the goodness of fit test.

The equation computed is

$$\chi_1{}^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

with $n - 1$ degrees of freedom.

O_i is the observed frequency.
E_i is the expected frequency.
n is the number of classes.

Example: Find the value of the chi-square statistic for the goodness of fit for the following data set.

$$\begin{bmatrix} O_i & E_i \\ \text{---} & \text{---} \\ 8 & 9.6 \\ 50 & 46.75 \\ 47 & 51.85 \\ 56 & 54.4 \\ 5 & 8.25 \\ 14 & 9.15 \end{bmatrix}$$

Clear the current statistics matrix and enter the data.

CLEAR
STAT CLΣ
`[8,9.6` Σ+
`[50,46.75` Σ+
`[47,51.85` Σ+
`[56,54.4` Σ+
`[5,8.25` Σ+
`[14,9.15` Σ+

```
3:
2:
1:
 Σ+  Σ-  NΣ  CLΣ  STOΣ  RCLΣ
```

Enter the program below to compute the chi-square statistic. The program has the same form as the Delta Percent Program from an earlier section. You may wish to refer to it for comments on the approach being followed. Note that this program differs by explicitly computing with columns 1 and 2 of the statistics matrix. It could be generalized for any pair of columns by retrieving the contents of *ΣPAR* for the column specification.

```
« 1 GET1 ARRY→ DROP 2
GET1 ARRY→ DROP NΣ 1
FOR x x NΣ + ROLL → ofq
«NΣ ROLL → efq
« '(ofq-efq)^2÷efq'
EVAL » » -1 STEP NΣ
1 2 →LIST →ARRY CNRM»
```

[ENTER]

```
1: « 1.00 GET1 ARRY→
   DROP 2.00 GET1 ARRY→
   DROP NΣ 1.00 FOR x x
→ARRY ARRY→ PUT GET PUTI GETI
```

Save the program in the variable *CHI* for repeated use.

'CHI [STO]

```
3:
2:
1:
→ARRY ARRY→ PUT GET PUTI GETI
```

Execute the program.

[USER] [CHI]

```
3:
2:
1:                  4.84
ΣDAT CHI GET1
```

The chi-square statistic χ^2 is 4.84 with 5 degrees of freedom.

Example: Compute the level of significance for the example above.

Enter the degrees of freedom and the χ^2 statistic, and compute the upper-tail probability for the χ^2 distribution.

5,4.84 [STAT] [UTPC]

```
3:
2:                  4.84
1:                  0.44
UTPC UTPF UTPN UTPT
```

Step-by-Step Examples for Your HP-28C

Probability and Statistics contains a variety of examples and solutions to show how you can solve your technical problems more easily.

- Combinations and Permutations
- Statistics Matrix Setup
  - Initialization and Data Entry, Grouped and Ungrouped Data
  - Data Removal and Extraction
- Basic Statistics for Multiple Variables
  - Sums, Means, Standard Deviation, Variance, Covariance
  - Sample and Population Statistics
  - Correlation Coefficient and Coefficient of Variation
  - Sums of Products, Normalized Data, Delta Percent on Paired Data
  - Moments, Skewness, and Kurtosis
- Regression
  - Curve Fitting — Linear Regression, Exponential, Logarithmic, and Power Curves
  - Multiple Linear Regression
  - Polynomial Regression
- Test Statistics and Confidence Intervals
  - Paired t Statistic, t Statistic for Two Means, Chi-Square Statistic

HEWLETT PACKARD

Reorder Number
00028-90043

0 88698 00003 8

00028-90058
Printed in U.S.A. 8/87