HP-80 APPLICATION NOTES

PUBLISHED AS A SERVICE FOR USERS OF THE HP-80 FINANCIAL POCKET CALCULATOR

MARCH 28, 1974 NO. 80–010

LINEAR REGRESSION CALCULATIONS

GENERAL

Least squares linear regression is a statistical method for finding a straight line that best fits a set of data points, thus providing a relationship between two variables. The HP-80 Trend Line (TL) calculation performs linear regression calculations, but requires that input data be evenly spaced and in chronological order. If data is not evenly spaced, the calculations described below can be used to develop a regression line.

Given the observations of two variables, the HP-80 user can solve for the ***slope*** (m), and ***y-intercept*** (b) of the standard regression line equation, $y = mx + b$. From this equation, the dependent variable (y) can be predicted for any given independent variable (x). In addition, the following procedures calculate the ***correlation coefficient*** (r), which measures the linear relationship between the two variables ($-1 \leqslant r \leqslant 1$), the ***coefficient of determination*** (r^2), which indicates the goodness of fit of the line to the data points ($0 \leqslant r \leqslant 1$), and the ***standard error*** (s) of the estimate of y on x, which is a measure of the scatter about the regression line of y on x.

The following symbolic values will be used to demonstrate the keystroke sequences below.

Input data: $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$

where n = number of observations
x = independent variable
y = dependent variable

Notation:

$\Sigma x_i = x_1 + x_2 + \ldots + x_n$	sum of x values
$\Sigma y_i = y_1 + y_2 + \ldots + y_n$	sum of y values
$\Sigma x_i^2 = x_1^2 + x_2^2 + \ldots + x_n^2$	sum of squares of x values
$\Sigma y_i^2 = y_1^2 + y_2^2 + \ldots + y_n^2$	sum of squares of y values
$\Sigma x_i y_i = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$	sum of x y products

σ_x = standard deviation of x values
σ_y = standard deviation of y values
m = slope
b = y-intercept
r = correlation coefficient
r^2 = coefficient of determination
s = standard error of the estimate of y on x

HEWLETT hp PACKARD 19310 Pruneridge Avenue, Cupertino, California 95014, Telephone 408-996-0100

GENERATING TREND LINE AND ESTIMATING VALUES

Keystrokes:

1. Solve for Σx_i and store it; solve for Σx_i^2 and σ_x and write these down.

   (gold key) [CLEAR CLX] x_1 [Σ+] x_2 [Σ+] . . . x_n [Σ+] [STO] ⟶ Σx_i

   [R↓] [R↓] ⟶ Σx_i^2 write this down*

   [R↓] [R↓] [$\bar{x}$] [x⇄y] ⟶ σ_x write this down

2. Solve for Σy_i, Σy_i^2, and σ_y and write these down.

   (gold key) [CLEAR CLX] y_1 [Σ+] y_2 [Σ+] . . . y_n [Σ+] ⟶ Σy_i write this down*

   [R↓] [R↓] ⟶ Σy_i^2 write this down

   [R↓] [R↓] [$\bar{x}$] [x⇄y] ⟶ σ_y write this down

3. Solve for $\Sigma x_i y_i$ and write it down.

   x_1 [SAVE↑] y_1 [×] x_2 [SAVE↑] y_2 [×] [+] . . . x_n [SAVE↑] y_n [×] [+] ⟶ $\Sigma x_i y_i$ write this down

4. Leaving $\Sigma x_i y_i$ on the display, continue as follows to calculate the slope (m). Write this answer down.

   [RCL] Σy_i [×] n [÷] [−] Σx_i^2 [RCL] [SAVE↑] [×] n [÷] [−] [÷] ⟶ m write this down*

5. Calculate, store and write down the y-intercept (b).

   [RCL] [×] Σy_i [x⇄y] [−] n [÷] ⟶ b write this down*

6. Now the slope (m) and y-intercept (b) can be properly located in the operational stack and storage register so the [TL] key can be used to find the corresponding y value for any given x value (x_k).

   m [SAVE↑] [SAVE↑] [SAVE↑] b [STO]

7. x_k [n] [TL] ⟶ y_k

   (This step may be repeated for any x value)

*These intermediate results must be written down in order to complete this set of keystroke procedures; i.e., in order to determine the equation for the regression line. All other results are optional, in that they are only required for computing the correlation coefficient (r), coefficient of determination (r^2), and standard error (s) in the following section.

(NOTE: If a high degree of accuracy is desired, press (gold key) 6 before writing down intermediate results.)

Example:

A commercial land appraiser has examined 4 vacant lots in the downtown section of a local community, all of which have the same depths but different frontages and values as shown below. Based on this data, what is the relationship between frontage and lot value? What predicted value would a lot have with a 65 foot frontage? With a 50 foot frontage?

Lot Frontage (feet)	Lot Value
70.8	$10,100.00
60.0	$ 8,219.00
85.0	$15,000.00
75.2	$11,120.00

Procedure:	*See Displayed:*
1. (gold key) [CLEAR CLX] 70.8 [Σ+] 60 [Σ+] 85 [Σ+] 75.2 [Σ+] [STO] ⟶	291.00 (Σx_i, write this down)
[R↓] [R↓] ⟶	21492.68 (Σx_i^2, write this down)
[R↓] [R↓] [$\bar{x}$] [x⇄y] ⟶	10.37 (σ_x, write this down)

2. (gold key) CLEAR CLx 10100 Σ+ 8219 Σ+ 15000 Σ+ 11120 Σ+ ⟶ 44439.00 (Σy_i, write this down)

R↓ R↓ ⟶ 518216361.0 (Σy_i^2, write this down)

R↓ R↓ $\bar{x}$ x⇄y ⟶ 2858.33 (σ_y, write this down)

3. 70.8 SAVE↑ 10100 × 60 SAVE↑ 8219 × + 85 SAVE↑ 15000 × + 75.2 SAVE↑ 11120 × + ⟶ 3319444.00 ($\Sigma x_i y_i$, write this down)

4. RCL 44439 × 4 ÷ − 21492.68 RCL SAVE↑ × 4 ÷ − ÷ ⟶ 268.30 (slope, m, write this down)

5. RCL × 44439 x⇄y − 4 ÷ ⟶ -8408.80 (y-intercept, b, write this down)

(The equation of the regression line is: y = \$268.30 x −\$8408.80)

6. 268.30 SAVE↑ SAVE↑ SAVE↑ 8408.80 CHS STO ⟶ -8408.80

7. 65 n TL ⟶ 9030.70 (\$9,030.70, projected value of lot with 65 foot frontage)

50 n TL ⟶ 5006.20 (\$5,006.20, projected value of lot with 50 foot frontage)

CORRELATION COEFFICIENT AND COEFFICIENT OF DETERMINATION

Keystrokes:

1. Calculate the correlation coefficient (r).

   m SAVE↑ σ_x × σ_y ÷ ⟶ r

2. Leaving the correlation coefficient on the display, find the coefficient of determination (r^2).

   2 y^x ⟶ r^2

Example:

Using the data from the previous example, what is the correlation coefficient (r) and coefficient of determination (r^2)? The required values written down from that example are:

m = 268.30
σ_x = 10.37
σ_y = 2858.33

Procedure:	*See Displayed:*
1. 268.30 SAVE↑ 10.37 × 2858.33 ÷ ⟶	.97 (correlation coefficient, r)
2. 2 y^x ⟶	.95 (coefficient of determination, r^2)

STANDARD ERROR OF THE ESTIMATE OF y ON x

Keystrokes:

1. Σy_i^2 SAVE↑ b SAVE↑ Σy_i × − m SAVE↑
2. $\Sigma x_i y_i$ × − n SAVE↑ 2 − ÷ (gold key) $\sqrt{x}$ (y^x) → s

Example:

Using the data from the first example, what is the standard error(s) of the estimate of y on x?

Given values or values written down from intermediate calculations are:

$\Sigma y_i^2 = 518216361.0$
$b = -8408.80$
$m = 268.30$
$\Sigma y_i = 44439.00$
$\Sigma x_i y_i = 3319444.00$
$n = 4$

Procedure: ***See Displayed:***

1. 518216361 SAVE↑ 8408.8 CHS SAVE↑ 44439 × − 268.3 SAVE↑ → 268.30
2. 3319444 × − 4 SAVE↑ 2 − ÷ (gold key) $\sqrt{x}$ (y^x) → 802.56 (standard error, s)

Equations:

$$m = \frac{\Sigma x_i y_i - \dfrac{\Sigma x_i \, \Sigma y_i}{n}}{\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}}$$

$$b = \frac{1}{n}(\Sigma y_i - m \cdot \Sigma x_i)$$

$$r = \frac{m\,\sigma_x}{\sigma_y}$$

$$s = \left[\frac{\Sigma y_i^2 - b\,\Sigma y_i - m\,\Sigma x_i y_i}{n-2}\right]^{1/2}$$

$$\sigma_x = \left[\frac{\Sigma(x_i - \bar{x})^2}{n-1}\right]^{1/2}$$