

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 ( $-1 \leq r \leq 1$ ), the *coefficient of determination* ( $r^2$ ), which indicates the goodness of fit of the line to the data points ( $0 \leq r \leq 1$ ), 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:**  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

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

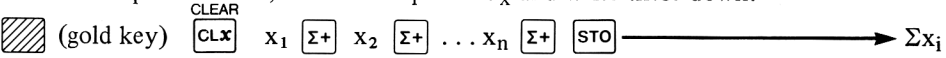
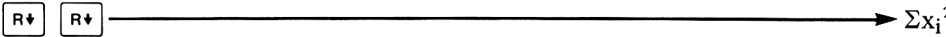
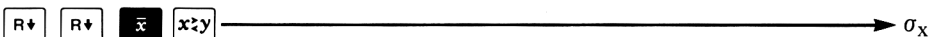
### Notation:

$\sum x_i = x_1 + x_2 + \dots + x_n$	sum of $x$ values
$\sum y_i = y_1 + y_2 + \dots + y_n$	sum of $y$ values
$\sum x_i^2 = x_1^2 + x_2^2 + \dots + x_n^2$	sum of squares of $x$ values
$\sum y_i^2 = y_1^2 + y_2^2 + \dots + y_n^2$	sum of squares of $y$ values
$\sum x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$	sum of $x y$ products
$\sigma_x$ = standard deviation of $x$ values	
$\sigma_y$ = 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 $x$	

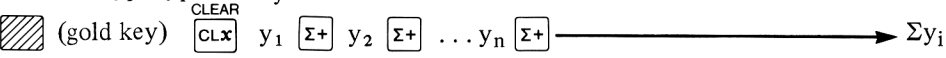
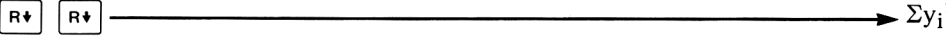
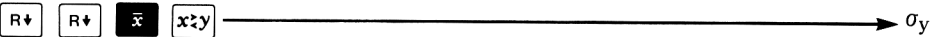
# GENERATING TREND LINE AND ESTIMATING VALUES

## Keystrokes:

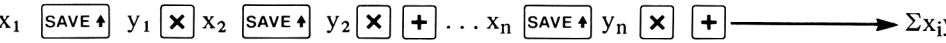
- Solve for  $\Sigma x_i$  and store it; solve for  $\Sigma x_i^2$  and  $\sigma_x$  and write these down.
 


write this down\*  
write this down
- Solve for  $\Sigma y_i$ ,  $\Sigma y_i^2$ , and  $\sigma_y$  and write these down.
 

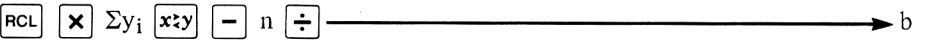
write this down\*  
write this down\*  
write this down
- Solve for  $\Sigma x_i y_i$  and write it down.
 




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



write this down\*
- Calculate, store and write down the y-intercept (b).
 



write this down\*
- 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 ( $x_k$ ).
 


- $x_k$  **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  6 before writing down intermediate results.)

## Example:

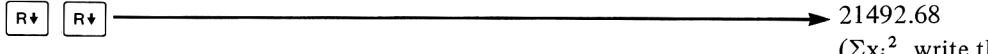
A commercial land appraiser has examined 4 vacant lots in the downtown section of a local community, all of which have the same depths but different frontages and values as shown below. Based on this data, what is the relationship between frontage and lot value? What predicted value would a lot have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

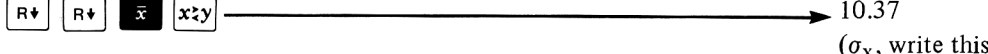
## Procedure:

- 

291.00  
( $\Sigma x_i$ , write this down)



21492.68  
( $\Sigma x_i^2$ , write this down)



10.37  
( $\sigma_x$ , write this down)

## See Displayed:

2.  $\square$  (gold key)  $\square$  CLEAR  $\square$  CLX 10100  $\square$   $\Sigma+$  8219  $\square$   $\Sigma+$  15000  $\square$   $\Sigma+$  11120  $\square$   $\Sigma+$   $\longrightarrow$  44439.00  
 ( $\Sigma y_i$ , write this down)
- $\square$  R $\blacktriangleright$   $\square$  R $\blacktriangleright$   $\longrightarrow$  518216361.0  
 ( $\Sigma y_i^2$ , write this down)
- $\square$  R $\blacktriangleright$   $\square$  R $\blacktriangleright$   $\square$   $\bar{x}$   $\square$   $\bar{y}$   $\longrightarrow$  2858.33  
 ( $\sigma_y$ , write this down)
3. 70.8  $\square$  SAVE  $\blacktriangleright$  10100  $\square$   $\times$  60  $\square$  SAVE  $\blacktriangleright$  8219  $\square$   $\times$   $\square$  + 85  $\square$  SAVE  $\blacktriangleright$  15000  
 $\square$   $\times$   $\square$  + 75.2  $\square$  SAVE  $\blacktriangleright$  11120  $\square$   $\times$   $\square$  +  $\longrightarrow$  3319444.00  
 ( $\Sigma x_i y_i$ , write this down)
4.  $\square$  RCL 44439  $\square$   $\times$  4  $\square$   $\div$   $\square$  - 21492.68  $\square$  RCL  $\square$  SAVE  $\blacktriangleright$   $\square$   $\times$  4  $\square$   $\div$   $\square$  -  $\square$   $\div$   $\longrightarrow$  268.30  
 (slope, m, write this down)
5.  $\square$  RCL  $\square$   $\times$  44439  $\square$   $\bar{y}$   $\square$  - 4  $\square$   $\div$   $\longrightarrow$  -8408.80  
 (y-intercept, b, write this down)
- (The equation of the regression line is:  $y = \$268.30 x - \$8408.80$ )
6. 268.30  $\square$  SAVE  $\blacktriangleright$   $\square$  SAVE  $\blacktriangleright$   $\square$  SAVE  $\blacktriangleright$  8408.80  $\square$  CHS  $\square$  STO  $\longrightarrow$  -8408.80
7. 65  $\square$  n  $\square$  TL  $\longrightarrow$  9030.70  
 (\$9,030.70, projected value of lot with 65 foot frontage)
- 50  $\square$  n  $\square$  TL  $\longrightarrow$  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 ( $r$ ).

$$m \square \text{SAVE } \blacktriangleright \square \sigma_x \square \times \square \sigma_y \square \div \longrightarrow r$$

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

$$2 \square y^x \longrightarrow r^2$$

### 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:

$$m = 268.30$$

$$\sigma_x = 10.37$$

$$\sigma_y = 2858.33$$

### Procedure:

1. 268.30  $\square$  SAVE  $\blacktriangleright$  10.37  $\square$   $\times$  2858.33  $\square$   $\div$   $\longrightarrow$  .97  
 (correlation coefficient,  $r$ )
2. 2  $\square$   $y^x$   $\longrightarrow$  .95  
 (coefficient of determination,  $r^2$ )

### See Displayed:

## STANDARD ERROR OF THE ESTIMATE OF y ON x

### Keystrokes:

- $\Sigma y_i^2$  [SAVE ↗] b [SAVE ↗]  $\Sigma y_i$  [X] [-] m [SAVE ↗]
- $\Sigma x_i y_i$  [X] [-] n [SAVE ↗] 2 [-] [÷] [gold key]  $\sqrt{x}$  [y<sup>x</sup>] → 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:

$$\begin{aligned} \Sigma y_i^2 &= 518216361.0 \\ b &= -8408.80 \\ m &= 268.30 \\ \Sigma y_i &= 44439.00 \\ \Sigma x_i y_i &= 3319444.00 \\ n &= 4 \end{aligned}$$

### Procedure:

- 518216361 [SAVE ↗] 8408.8 [CHS] [SAVE ↗] 44439 [X] [-] 268.3 [SAVE ↗] → 268.30
- 3319444 [X] [-] 4 [SAVE ↗] 2 [-] [÷] [gold key]  $\sqrt{x}$  [y<sup>x</sup>] → 802.56  
(standard error, s)

*See Displayed:*

### Equations:

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

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

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

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

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