



## hp calculators

### HP 48GII Cost Estimation using Linear Regression

The STAT menu

Cost Estimation using Linear Regression

Practice estimating costs using linear regression



## The STAT menu

The Statistics menu is accessed from the RED shifted function of the  $\boxed{5}$  key by pressing  $\boxed{\rightarrow}$   $\boxed{STAT}$ . When pressed, a CHOOSE box is displayed with a number of problem areas within statistics where the HP 48GII can be applied.



Figure 1

The first choice deals with calculating many values related to the statistics of a single variable, such as the average, standard deviation and others. The second choice deals with placing data within bins or classes and is useful when dealing with frequency distributions or when there is a need to summarize data within a histogram. The third choice, Fit Data... is what we will use to compute trend lines. The fourth choice allows for the computation of summary statistics from a data set, such as the sum of all X values, the sum of all the Y values squared, and others. The fifth choice allows for the evaluation of many different hypothesis tests and the sixth choice allows for the construction of confidence intervals around a sample mean.

To develop a linear regression cost estimation equation, press the  $\boxed{\nabla}$  key until the third choice Fit Data.. is highlighted and press the  $\boxed{F6}$  key to register the choice of  $\boxed{\text{FIT DATA}}$  shown above the  $\boxed{F6}$  key at the right side of the display. The  $\boxed{ENTER}$  key can also be used to make selections from a CHOOSE box and will be used below. Alternatively, since the Fit Data.. choice is the third item in the list, the  $\boxed{3}$  key can be pressed to immediately choose this function.

The screen displayed allows for the entry of the existing data for the estimation of linear regression equations. The  $\Sigma$ DAT field at the top of the screen is where a matrix containing the statistics data is entered.



Figure 2

When the cursor is on this field, press the  $\boxed{FI}$  key at the top left of the keyboard to register the choice of  $\boxed{\text{FIT DATA}}$ . When pressed, the MatrixWriter is started to allow for the entry of the statistical data as shown in the screen below.

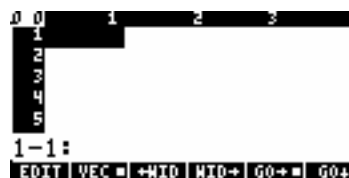


Figure 3

In many ways, this screen works like any spreadsheet. Enter numbers and they will go in the highlighted cell. The menu labels at the lower left corner of the screen,  $\boxed{\rightarrow}$  and  $\boxed{\nabla}$ , determine the direction the cursor moves after a data point has been entered, either right to the next column or down to the next row. In this example, the selection is to move right after each data point has been entered. This is indicated by the square present next to the  $\boxed{\rightarrow}$  menu label. To change the way the cursor will move, press the menu label for the direction desired and the square in the menu label will change accordingly. If a column is too small to show the data entered, the  $\boxed{\text{WID}}$  and  $\boxed{\text{NAR}}$  menu keys may be used to expand or shrink the area displayed for each column.

Enter the first X data value by keying in the numbers and pressing the  $\boxed{ENTER}$  key. The cursor will move to the right into the second column where the first Y data value should be keyed with the  $\boxed{ENTER}$  key pressed to accept this value. At this point, the cursor will be in column 3. Use the  $\boxed{\nabla}$  and  $\boxed{\leftarrow}$  keys to move back to the first column and enter the second X data value. Continue entering the pairs of numbers in columns 1 and 2 until all have been entered. Since the variable for  $\Sigma$ DAT will hold its contents from one

## HP 48GII Cost Estimation using Linear Regression

calculation to the next, be alert to extra rows of data that may exist below the last row entered for the current calculation. Any rows not needed can be deleted if the cursor is on the row to be deleted by pressing the **(NEXT)** key and then pressing the **DEL** menu label above the **(F2)** key. Continue entering the pairs of numbers in columns 1 and 2 until all have been entered. Press the **(UP)** key until the cursor is in row 1 and the screen would look something like this:

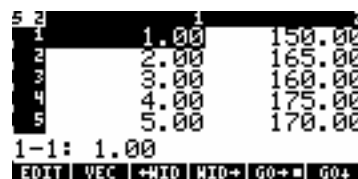


Figure 4

To accept the data as input, press the **(ENTER)** key to return to the previous menu. If an error is noticed, use the **(LEFT)**, **(UP)**, **(DOWN)**, **(RIGHT)** cursor keys to move to the cell with the incorrect value. Key in the correct value and press the **(ENTER)** key. Do this until all data values have been corrected and press the **(ENTER)** key again to return to the previous menu. The screen will now look something like this:



Figure 5

To make predictions using a linear regression model, leave the model choice set to the value of Linear Fit. Other choices include Logarithmic, Exponential, Power and Best fits. To access the prediction screen, press the **(F4)** key to register the choice of **PRED**. The screen displayed will look like this:

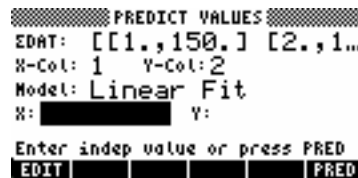


Figure 6

Move the cursor key down to the X: or Y: data entry locations. You may then enter an X or Y value for which you wish to predict the corresponding X or Y value. To actually predict the corresponding value, press the **PRED** key after entering the X or Y value. You may do this as often as you like by moving the cursor back and forth.

To exit from this data entry screen, press the **(ON)** key to return to the previous data entry screen. From this location, you can either press the **(ON)** key again to quit the STAT screen or press **QUIT** shown above the **(F6)** to exit the STAT screen but with the extra feature of copying to the screen information about the linear regression line that was used. Three items will be shown on the screen. The top level indicates the equation of the linear regression-determined line. The second level indicates the correlation of the line. The bottom level indicates the covariance of the line. The correlation will always be between -1 and +1, where values closer to -1 and +1 indicating a good "fit" of the line to the data. Values nearer to zero indicate little to no "fit." Little reliance should be placed upon predictions made where the correlation is not near -1 or +1. Exactly how far away from these values the correlation can be and the equation still be considered a good predictor is a matter of debate. Once these data items are viewed, pressing **(P) STAT** will return you to the STAT menu where you can reenter the Fit Data.. area if desired.

## Cost estimation using Linear Regression

Linear regression is a process which fits the equation of a line in the form  $Y = bX + a$ , where  $b$  is the slope of the line and  $a$  is the Y-intercept, to a given set of data. Linear regression calculates the equation for this line by minimizing the sum of the squared residuals between the actual data points and the predicted data points using the estimated line's slope and intercept. Once the slope and intercept have been calculated, it is fairly easy to substitute other values for X and predict a corresponding value for Y, or to substitute a value for Y and predict a value for X. Linear regression is often used to estimate the fixed and variable components from

## HP 48GII Cost Estimation using Linear Regression

a company's or department's total costs. In these circumstances, the values for X are usually the cost driver for the organization or department. Examples might include units produced, hours worked, hours of machine time, and others. The values for Y are the total cost for that level of X input. The computed slope of the linear regression line will indicate the variable cost per unit of X, while the computed Y-intercept will indicate the fixed cost.

In many or most circumstances, this type of cost analysis will generate slopes and Y-intercepts that make sense in the real world. It is sometimes possible, though, that the fixed cost component in particular may not make any sense. The generated Y-intercept (fixed cost) might be negative, for example, to make the linear regression line fit the observed cost data as closely as possible. Be aware, as well, that it is rarely a good idea to use such an equation to predict too far into the future from the actual data used, since circumstances can change rather quickly. In other words, if you fit a line using cost data for units produced from 500 to 1500 a month, making cost predictions using forecasted production levels of 5000 units a month may generate unreliable results. Also, since time is not a variable in these calculations, the order in which the costs are input as data points does not matter – you may enter the data points in any order desired.

## Practice estimating costs using linear regression

**Example 1:** Johnson's Chair Company has experienced the following costs for the first 6 months of the year:

# Chairs Made	Total Costs
5,000	\$120,000
5,500	\$122,100
4,800	\$118,540
5,300	\$122,400
4,950	\$119,100
5,150	\$124,200

What estimate would a linear regression equation produce for Johnson's fixed and variable cost? How good is the fit of the linear regression line generated (What is the correlation)? What are the total costs predicted if 5,400 chairs were to be made? If the total costs were \$125,000, how many chairs would you estimate had been produced?

**Solution:** The X values will be the number of chairs produced. The Y values will be the total costs.

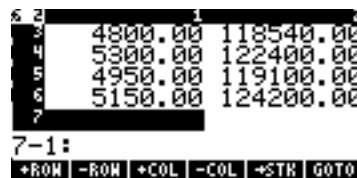
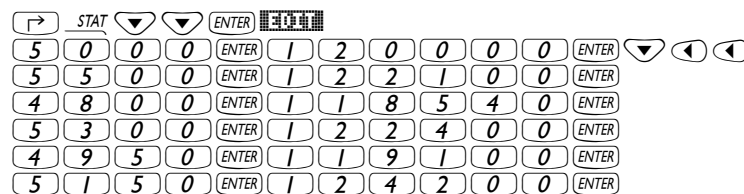


Figure 7



Figure 8

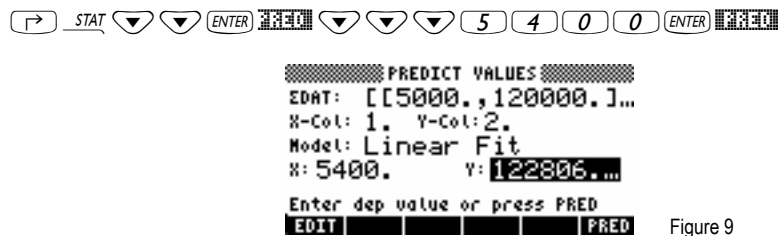


Figure 9

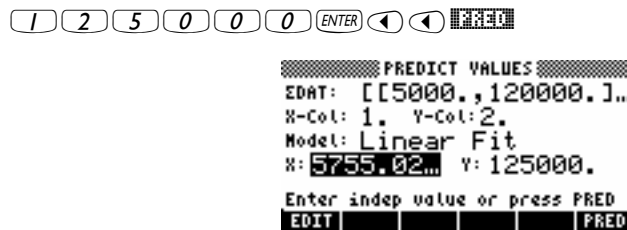


Figure 10



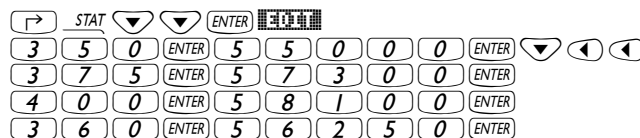
**Answer:** The linear regression equation generated is of the form:  $Y = 6.18X + 89449.38$ . The slope of 6.18 is the estimate for the variable cost and the Y-intercept of 89,449.38 is the estimate for the fixed cost. The correlation value of 0.72 is not as close to +1 as might be hoped, but still indicates a moderate fit. The total cost estimate if 5,400 chairs were made is \$122,806. The estimated number of chairs made if the total costs were \$125,000 is 5,755 chairs.

**Example 2:** The stamping department cost analyst is reviewing the total cost compared with the number of machine hours used for the last 4 weeks.

# Machine Hours	Total Costs
350	\$55,000
375	\$57,300
400	\$58,100
360	\$56,250

What estimate would a linear regression equation produce for the stamping department's fixed and variable cost? How good is the fit of the linear regression line generated (What is the correlation)? What are the total costs predicted if 380 machine hours are used next week? If the total costs were \$60,000, how many machine hours would you estimate had been used?

**Solution:** The X values will be the number of machine hours used. The Y values will be the total costs.



The next keystrokes are only needed if there is data in this input form beyond row 4 (as there would be if this problem is worked immediately after the previous problem).



If this problem is being worked without having just completed the previous example, then there may need to be more than 3 presses of the key or less presses or none.

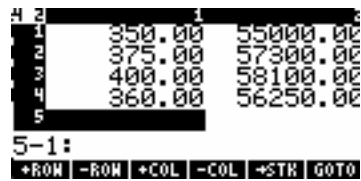


Figure 11

ENTER OK

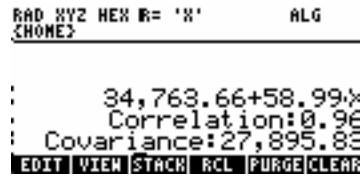


Figure 12

STAT 3 8 0 ENTER

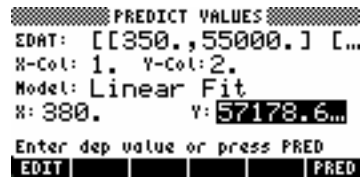


Figure 13

6 0 0 0 0 ENTER PRED

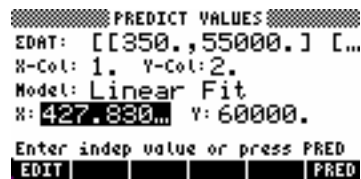


Figure 14

ON OK

**Answer:**

The linear regression equation generated is of the form:  $Y = 58.99X + 34,763.66$ . The slope of 58.99 is the estimate for the variable cost per machine hour and the Y-intercept of 34,763.66 is the estimate for the fixed cost. The correlation value of 0.96 is quite close to +1 and indicates a very good fit. The total cost estimate if 380 machine hours are used is \$57,178. The estimated number of machine hours used if the total costs were \$60,000 is 427.8 machine hours.