

Cost estimation using Linear Regression

Linear regression calculates the equation for a line that "best fits" a set of ordered pairs 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. The equation of the line produced by linear regression is in the form $Y = mX + b$, where m is the slope of the line and b is the Y-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 . When the X value is a measure of time (months or years, for example), the equation is specifically referred to as a trend line.

Linear regression is often used to estimate the fixed and variable components from 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 away from the range of 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 1,500 a month, making cost predictions using forecasted production levels of 5,000 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.

On the HP 35s, values are entered into the statistical / summation registers by keying in the number (or pair of numbers) desired and pressing $\Sigma+$. This process is repeated for all numbers or pair of numbers. When entering a pair of numbers in RPN or algebraic mode, key the Y value, press ENTER , then key the X value and press $\Sigma+$.

To view the linear regression results, press L.R. . The HP 35s displays a menu of relevant values. Items on this menu are viewed by pressing the \leftarrow or \rightarrow cursor keys of the HP 35s.

This menu allows you to predict a value for X given a Y value, or predict a value for Y given an X value. It also displays the linear regression line's correlation, slope, and y-intercept. 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. To use a value displayed on the menu, press the ENTER button and the value will be copied for further use. This is illustrated in the problems below.

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:








<u># Chairs Made</u>	<u>Total Costs</u>
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. Be sure to clear the statistics / summation memories before starting the problem.

 CLEAR 

In RPN or algebraic mode, press:

To view the linear regression results, press  L.R.. Figure 2 displays the menu shown.

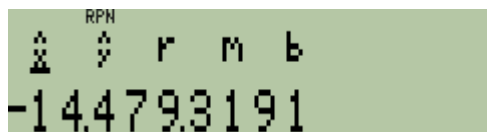


Figure 2

In either RPN or algebraic mode, press:    to view the slope.

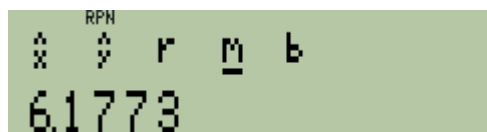



Figure 3

In either RPN or algebraic mode, press:  to view the y-intercept.

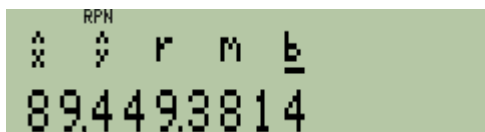


Figure 4

In either RPN or algebraic mode, press: $\leftarrow \leftarrow$ to view the correlation.

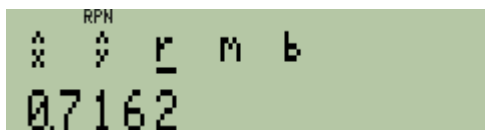


Figure 5

To estimate costs if 5,400 chairs are made, do the following:

In either RPN or algebraic mode, press: \leftarrow 5 4 0 0 \leftarrow L.R. \rightarrow

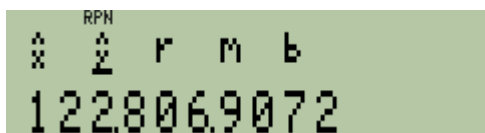


Figure 6

To estimate the number of chairs made if the costs were \$125,000, do the following:

In either RPN or algebraic mode, press: \leftarrow 1 2 5 0 0 0 \leftarrow L.R.

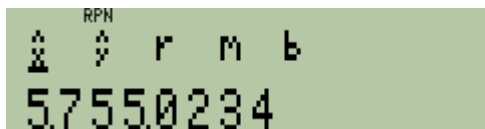


Figure 7

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.

<u># Machine Hours</u>	<u>Total Costs</u>
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?

HP 35s Cost Estimation using Linear Regression

Solution: The X values will be the number of machine hours used. The Y values will be the total costs. Be sure to clear the statistics / summation memories before starting the problem.

▶ **CLEAR** **4**

In RPN or algebraic mode, press:

5 **5** **0** **0** **0** **ENTER** **3** **5** **0** **Σ+**
5 **7** **3** **0** **0** **ENTER** **3** **7** **5** **Σ+**
5 **8** **1** **0** **0** **ENTER** **4** **0** **0** **Σ+**
5 **6** **2** **5** **0** **ENTER** **3** **6** **0** **Σ+**

To view the linear regression results, press **↵** **L.R.**. Figure 8 displays the menu shown.

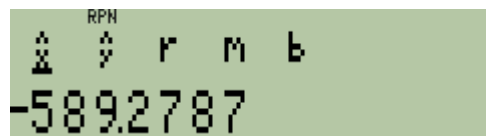


Figure 8

In either RPN or algebraic mode, press: **>** **>** **>** to view the slope.

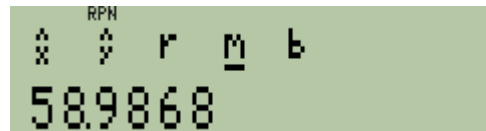


Figure 9

In either RPN or algebraic mode, press: **>** to view the y-intercept.

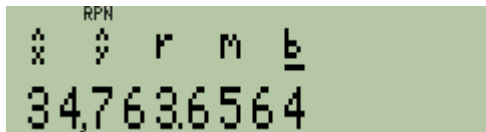


Figure 10

In either RPN or algebraic mode, press: **<** **<** to view the correlation.

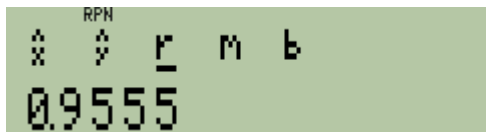


Figure 11

To estimate costs if 380 machine hours are used, do the following:

In either RPN or algebraic mode, press: **C** **3** **8** **0** **↵** **L.R.** **>**

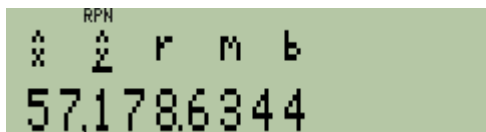


Figure 12

HP 35s Cost Estimation using Linear Regression

To estimate the number of number of machine hours used if costs were \$60,000, do the following:

In either RPN or algebraic mode, press: **C** **6** **0** **0** **0** **0** **↵** **L.R.**

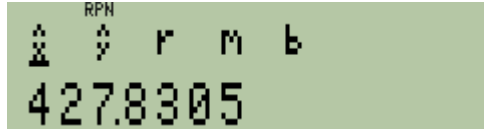


Figure 13

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.