hp calculators

HP 17bII+ Statistics – Linear Regression

Statistics on the HP 17bII+

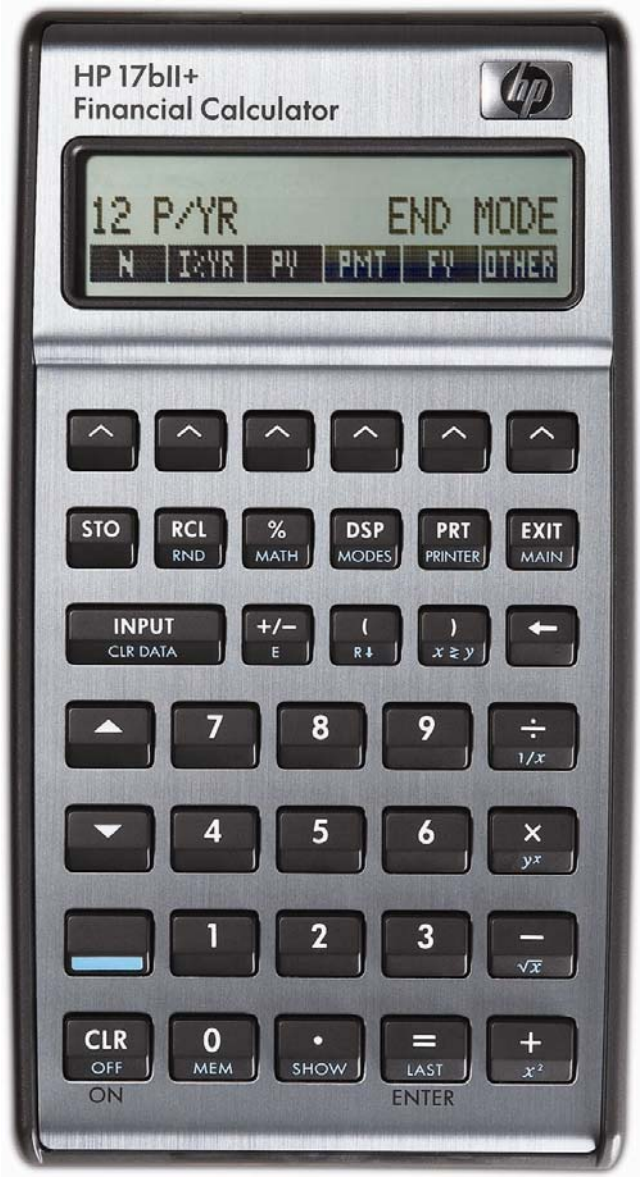
Linear regression and trend lines

Sum lists

Entering data values

Linear regression on the HP 17bII+

Practice solving problems involving linear regression



Statistics on the HP 17bII+

The HP 17bII+ has many built-in statistics functions that apply to finding averages and standard deviations as well as linear regression and correlation.

Linear regression and trend lines

A trend line is actually an equation of a line in the form Y = mX + b, where m is the slope of the line and b is the Y-intercept. 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. When the X value is a measure of time (months or years, for example), the equation is specifically referred to as a trend line. These are often used to predict future sales growth given past sales data. Be aware, however, 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. Also be aware that these predictions are linear in nature and make no adjustment for any seasonality that may exist.

Linear regression is also 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 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.

Sum lists

On the HP 17bII+, statistics problems are solved in the Sum environment. This is entered from the main menu by pressing SUM.

Figure 1

Figure 2

On this screen, the menu keys on the bottom of the display perform these functions:

CALC – Ends the entry of statistical data into the sum list and enters the statistics calculation environment.

TOTAL – Displays the total of the data values in the current sum list.

The other menu functions, INSR, DELET, NAME, and GET allow you to edit and save a sum list. They are covered in another learning module.

The HP 17bII+ uses a different approach to handling statistics problems than other HP calculators. Rather than storing the data values in registers, the HP 17bII+ stores data values in lists that can be named and saved for future use. This also provides the flexibility to have multiple lists of statistical data stored within the calculator at the same time, limited only by the available calculator memory.

Entering data values

The initial data value is keyed and entered using the INPUT CLR DATA key. The remaining data values are entered in the same manner. Data values can be keyed or computed. They are added to the list when you press INPUT CLR DATA = LAST. Note to RPN users: This is an instance where there is a difference between the INPUT CLR DATA and = LAST keys. If you wish to compute a value and add it to a sum list, use = LAST rather than INPUT CLR DATA to serve as ENTER. Pressing INPUT CLR DATA in RPN mode while entering data into a sum list will add the number displayed to the list.

When you have entered all data values, press EXIT MAIN followed by STAT to enter the calculation environment, which appears as shown below in figure 3. The menu choices of TOTAL, MEAN, MEDN, STDEV and RANG are covered in other learning modules. Pressing MORE shows the second page of the statistics calculation environment as shown in figure 4 below.

Figure 3

Figure 4

MIN , MAX , and SORT are covered in other learning modules. FRCST displays a series of menus for calculations involving two variables for curve fitting or statistical analysis requiring two lists.

Linear regression on the HP 17bII+

On the HP 17bII+, linear regression problems involve entering two sum lists – one for the dependent variable (X) and one for the independent variable (Y). These must be entered and saved with a name in order to associate them together for a regression problem. Once these have been entered, pressing FRCST will allow you to choose a list for the X values and a list for the Y values. Once chosen, the screen displays the forecasting menu as shown below. The first two menu positions represent the independent and dependent variables chosen as you entered the FRCST environment.

Figure 5

- [redacted] – This location represents the independent variable. Enter a value for the dependent variable and press this key to estimate the corresponding value of the independent variable. The name shown at this menu position will depend upon the list chosen. In the rest of the statistics options, this is referred to as the X-variable list.
- [redacted] – This location represents the dependent variable. Enter a value for the independent variable and press this key to estimate the corresponding value of the dependent variable. The name shown at this menu position will depend upon the list chosen. In the rest of the statistics options, this is referred to as the Y-variable list.
- CORR – Computes and displays correlation coefficient of the independent and dependent variables. This value will be between -1 and +1. The better the "fit" of the model to the X and Y variable lists, the closer this value will be to -1 or +1. Two lists that have no relationship to each other at all would have a correlation of 0.
- [redacted] – Computes and displays M. This is the slope for the linear regression model.
- [redacted] – Computes and displays B. This is the y-intercept for the linear regression model.
- MORE – Displays the second page of forecasting menu options as shown below in figure 6.

Figure 6

MODL – Displays a screen of choices for regression models to be used. The HP 17bII+ provides a choice of four models. Linear LIN, Logarithmic LOG, Exponential EXP and Power PWR fits.

W.MN – Computes and displays the weighted mean of the X variable list using the weights in the Y variable list.

G.SD – Computes and displays the standard deviation of the X variable list grouped by the frequencies specified in the Y variable list.

SIZE – Displays the number of items in either list.

MORE – Displays the third page of forecasting menu options as shown below in figure 7.

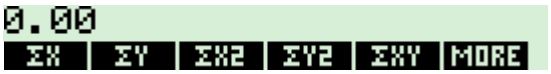


Figure 7

ΣX – Displays the sum of the items in the X variable list.

ΣX2 – Displays the sum of the square of the items in the X variable list.

ΣY – Displays the sum of the items in the Y variable list.

ΣY2 – Displays the sum of the square of the items in the Y variable list.

ΣXY – Displays the sum of the values in the X variable list multiplied by the values in the Y variable list.

MORE – Displays the first page of forecasting menu options again.

Practice solving linear regression problems

Example 1: John's store has had sales for the last 5 months of $150, $165, $160, $175, and $170. Use a trend line to predict sales for months 6 and 7 and also predict when estimated sales would reach $200.

Solution: [ ] EXIT MAIN SUM

If you already have at least one sum list, press: GET ÷NEW

1 INPUT CLR DATA
2 INPUT CLR DATA
3 INPUT CLR DATA
4 INPUT CLR DATA
5 INPUT CLR DATA EXIT MAIN

NAME NXYZ X INPUT CLR DATA

GET ÷NEW

1 5 0 MEM INPUT CLR DATA
1 6 5 INPUT CLR DATA
1 6 0 MEM INPUT CLR DATA
1 7 5 INPUT CLR DATA
1 7 0 MEM INPUT CLR DATA EXIT MAIN

NAME RSTUV S ABCDE A JKLM L ABCDE E RSTUV S INPUT CLR DATA

CALC MORE FRCST

Figure 8

X

Figure 9

SALES

Figure 10

X

Figure 11

B

Figure 12

6 X SALES

Figure 13

7 X SALES

Figure 14

2 0 MEM 0 MEM SALES X

Figure 15

Answer: The slope is 5 and the y-intercept is 149. The linear regression equation is therefore Y = 5X + 149. Sales in month 6 are predicted to be $179 and in month 7 are expected to be $184. Sales are predicted to reach $200 between months 10 and 11 (actual answer is 10.2).

Example 2: 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? 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.

[ ] [EXIT MAIN] [SUM]

If you already have at least one sum list, press: [GET] [→NEW]

[5] [0 MEM] [0 MEM] [0 MEM] [INPUT CLR DATA]
[5] [5] [0 MEM] [0 MEM] [INPUT CLR DATA]
[4] [8] [0 MEM] [0 MEM] [INPUT CLR DATA]
[5] [3] [0 MEM] [0 MEM] [INPUT CLR DATA]
[4] [9] [5] [0 MEM] [INPUT CLR DATA]
[5] [1] [5] [0 MEM] [INPUT CLR DATA] [EXIT MAIN]

[NAME] [ABCDE] [C] [FGHI] [H] [ABCDE] [A] [FGHI] [I] [RSTUV] [R] [INPUT CLR DATA]

[GET] [→NEW]

[1] [2] [0 MEM] [0 MEM] [0 MEM] [0 MEM] [INPUT CLR DATA]
[1] [2] [2] [1] [0 MEM] [0 MEM] [INPUT CLR DATA]
[1] [1] [8] [5] [4] [0 MEM] [INPUT CLR DATA]
[1] [2] [2] [4] [0 MEM] [0 MEM] [INPUT CLR DATA]
[1] [1] [9] [1] [0 MEM] [0 MEM] [INPUT CLR DATA]
[1] [2] [4] [2] [0 MEM] [0 MEM] [INPUT CLR DATA] [EXIT MAIN]

[NAME] [ABCDE] [C] [MOPQ] [O] [RSTUV] [S] [RSTUV] [T] [RSTUV] [S] [INPUT CLR DATA]

[CALC] [MORE] [FRCST]

Figure 16

[CHAIR]

Figure 17

[COSTS]

[M]

Figure 18

[B]

Figure 19

[CORR]

Figure 20

[5] [4] [0 MEM] [0 MEM] [CHAIR] [COSTS]

Figure 21

Figure 22

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 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.