



hp calculators

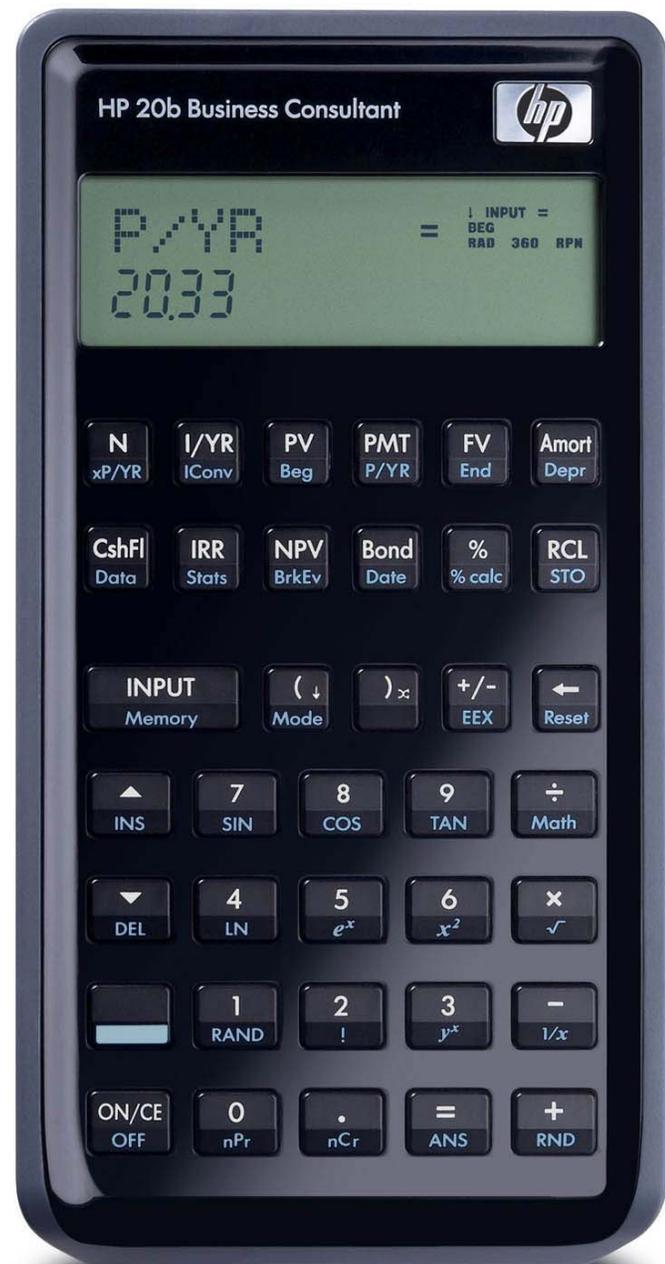
HP 20b Statistics – Regression

Regression and trend lines

Statistics on the HP 20b

The Best Fit feature

Practice solving problems involving regression



Regression and trend lines

A trend line is actually an equation of a line in the form $Y = aX + b$, where a 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.

Linear regression is also 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.

Other types of regression exist that fit curves to sets of data, rather than a straight line. These all work in a similar manner, by finding an equation for a curve that fits as closely as possible to the provided data points.

When choosing between differing models fit to a set of data points, the choice is often made for the model with the highest correlation, as this measure generally reflects how good a fit a model has to the data points – a model with a correlation of 0.7 is generally not as good a fit to data as a model with a correlation of 0.9. However, this approach should be weighted along with other factors – the model's usefulness and future expectations for example.

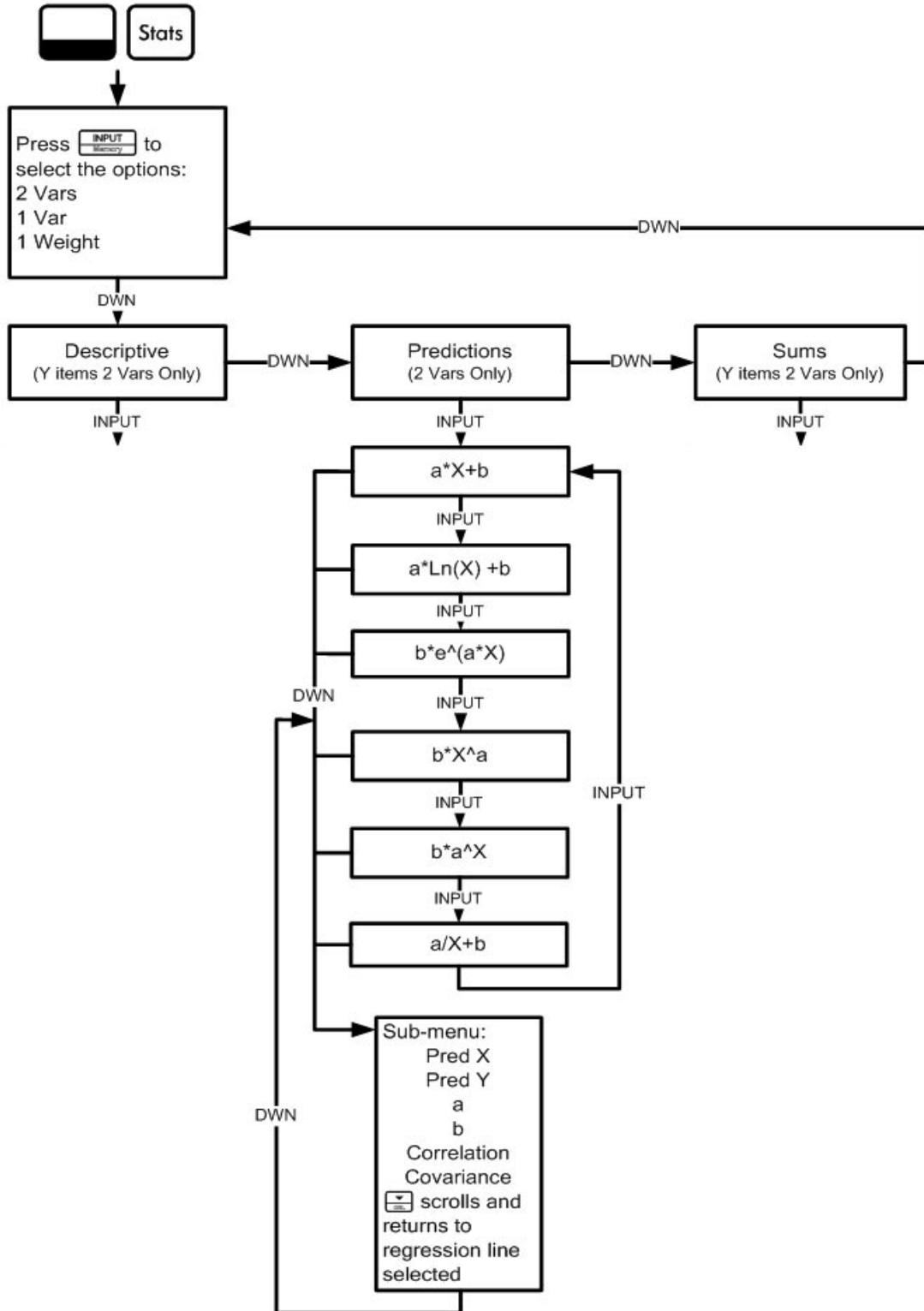
Statistics on the HP 20b

The HP 20b has many built-in statistics functions that apply to finding averages, standard deviations, standard error of the means as well as regression, correlation, and covariance. The HP 20b also accumulates many statistical sums for your use. Many of the HP 20b statistics functions are found in the  menu, shown in the menu map on the next page. The portion of the menu map for predictions is not shown in this map, since it is covered in a separate learning module.

The data the statistics functions use for computations must be entered first by pressing . If you enter the stats menu before you have entered any data, you will automatically be placed in the data menu. In this menu, enter a list of x values for one-variable statistics, a list of pairs, (x, y) for two-variable statistics, or a list of pairs, (x, y) for weighted, one-variable statistics, where the y values would be the weights. To enter data, key in a number and press . To enter a list of x values only, press   to bypass the entry of y values. To review the data items that are in the data menu, you can press  or  to scroll through the values. To clear the data menu while in the menu, press   , followed by pressing . To simply exit the menu, press .

When you press , your first choice needs to be the type of statistics you will be analyzing: 2 Variable, 1 Variable, or 1 Weight. Press  to scroll through the options. When the option you wish to use is displayed, press  to enter the second level of the menu, where you will need to choose descriptive, predictions, or sums by pressing  to move between the choices and by pressing  when the choice you want is displayed.

In this learning module, we will focus on the predictions sub-menu of the stats menu.



The Best Fit feature

The table below lists the different regression models built into the HP 20b. Pressing will cycle through the regression model displayed. Press when the model you wish to use is displayed to select that model.

NOTE: when a regression model is displayed, pressing will perform a “Best Fit” calculation and immediately change the displayed regression model choice to the regression model that best fits the data values, as determined by the model with the correlation whose absolute value is closest to 1. This feature can be very helpful if you are trying to predict values using a regression model.

HP 20b Statistics Menu – Predictions Regression models

Item	Description
$a \cdot X + b$	Linear regression.
$a \cdot \ln(X) + b$	Logarithmic regression.
$b \cdot e^{(aX)}$	Exponential regression using base e.
$b \cdot X^a$	Power regression using X as the base.
$b \cdot a^X$	Power regression using the coefficient a as the base.
$a/X + b$	Inverse regression.

The table below explains each of the entries in the Statistics – Predictions sub-menu in more detail.

HP 20b Statistics Menu – Predictions Sub-Menu items

Item	Description
Pred X	Predicted X value. Enter an X value to use for predicting a Y value and press <input type="button" value="INPUT"/> , or predict an X value based upon the value found in Pred Y by pressing <input type="button" value="="/> . Input/Output.
Pred Y	Predicted Y value. Enter an Y value to use for predicting an X value and press <input type="button" value="INPUT"/> , or predict an Y value based upon the value found in Pred X by pressing <input type="button" value="="/> . Input/Output.
a	The a coefficient for the selected regression model. Output only.
b	The b coefficient for the selected regression model. Output only.
Correlation	The correlation of the X and Y values. Values range from +1 to -1. Output only.
Covariance	The covariance of the X and Y values. Output only.

Practice solving 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. What are the a and b coefficients (for a linear model, these would be the slope and intercept), and what are the correlation and covariance?

Solution:

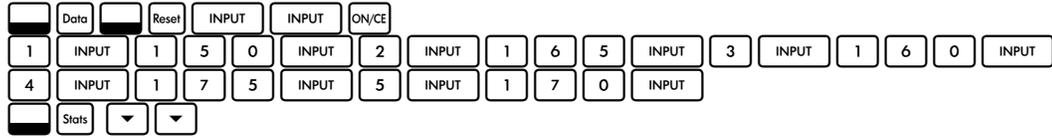


Figure 1



Figure 2



Figure 3



Figure 4

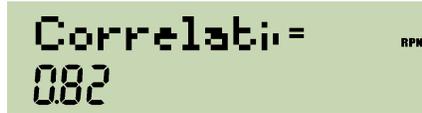


Figure 5



Figure 6



Figure 7



Figure 8

2 0 0 INPUT ▲ =

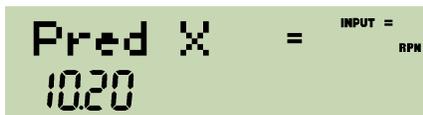


Figure 9

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). The correlation is 0.82 and the covariance is 12.50.

Example 2: Assuming you have just finished example 1 and that the data is still in the calculator, which model is the best fit, giving the highest correlation? What is that correlation?

Solution: Re-enter the stats menu and navigate to the sub-menu item showing the regression model choice.

ON/CE Stats ▼ ▼ INPUT



Figure 10

Now, press \square to have the HP 20b choose the Best Fit model – the model with the correlation whose value is closest to -1 or +1.

\square



Figure 11

▲ ▲



Figure 12

Answer: The best fitting model is the inverse regression, with a correlation of -0.87, compared to 0.82 for the linear model. This may be worth considering as an alternative.

Example 3: 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? 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? Is a linear regression line the best model? If not, what is?

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

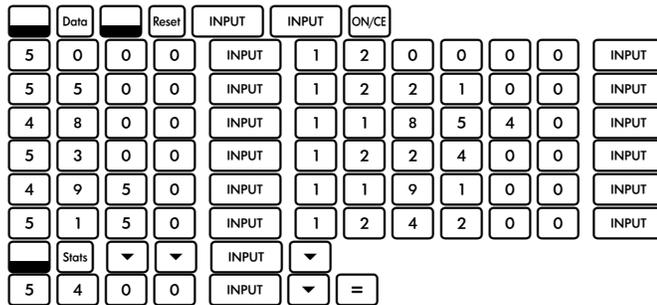


Figure 13



Figure 14



Figure 15



Figure 16



Figure 17



Figure 18

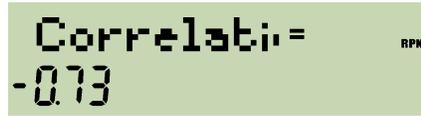


Figure 19

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. The correlation of the linear regression line is 0.72. The HP 20b computes the best fitting line as the inverse regression with a correlation of -0.73. Since this is barely “better” than the 0.72 of the linear model, there may be no compelling reason to switch from the linear model, particularly since the linear model provides a clearer meaning for fixed and variable costs.