

With the regression equation obtained, one may answer the question: If applicant Sara Greenley's SAT score is 1110, what would one predict her freshman grade average to be? The question is answered by substituting $X = 1110$ into the regression equation and computing \hat{Y} :

$$\begin{aligned}\hat{Y} &= .012(1110) - 11.638 \\ \hat{Y} &= 1.682\end{aligned}$$

On the graph, this is equivalent to drawing a vertical line through $X = 1110$ parallel to the Y -axis and determining the Y value at the point of intersection of that vertical line and the regression line. This point of intersection is \hat{Y} for $X = 1110$.

The several computations described above may be performed by many small calculators designed for scientific purposes and by statistical programs for personal or larger computers. Typically, you are required to enter X_1 followed by Y_1 , then X_2 followed by Y_2 , and so on for all pairs of scores. When all data have been entered, a and b can be obtained with the push of one or two buttons or some other simple command. Some programs will even calculate the predicted value for a specific X score. Read the instruction manual for your particular calculator or program, and note that sometimes m rather than b is used to symbolize the slope and the word "constant" may be used to indicate a , the y -intercept. Procedures to use the StataQuest, MINITAB, and SPSS programs on a personal computer are given in the *Study Guide* that accompanies this text. However, it is recommended that students calculate a few regression examples or exercises "by hand" to become familiar with how the formulas work before moving to programmed calculators or computers.

Plotting the Regression Line

It is sometimes necessary to plot a regression line once you have obtained its algebraic equation. Two points determine a straight line, and the regression constants, a and b , will provide those two points. First, mark off a set of axes. Remember, the *variable doing the predicting* (variable X) goes on the horizontal axis (X -axis), and the *variable being predicted* (variable Y) goes on the vertical axis (Y -axis). Suppose the regression equation is $\hat{Y} = .75X + 2.00$. The regression constants then are $a = 2.00$ and $b = .75$. Since the value of a is the y -intercept, simply place a point on the Y -axis at the point corresponding to the value calculated for a . In this case $a = 2.00$, which would be a point on the ordinate 2.00 units above 0. If $a = -2.00$, the point would be on the ordinate but 2.00 units below 0.

Now the task is to determine any second point on the line. Recall that the slope represents the amount of vertical distance relative to horizontal distance between any two points. Thus, start with the point a that you just placed on the Y -axis and move up vertically b units. From there, move horizontally 1 X unit to the right. That is, the slope is .75, so move up .75 unit on the Y -axis for every 1 unit you move right on the X -axis. If the slope is $-.75$, move down .75 unit on the Y -axis for every 1 unit you move to the right on the X -axis.

The second point on the regression line may be determined more simply if the means of the X and Y distributions are known, because (\bar{X}, \bar{Y}) is always a point on the line. Thus, (\bar{X}, \bar{Y}) and a are two points on the line. Figure 6-7 illustrates how to plot a regression line from its equation.

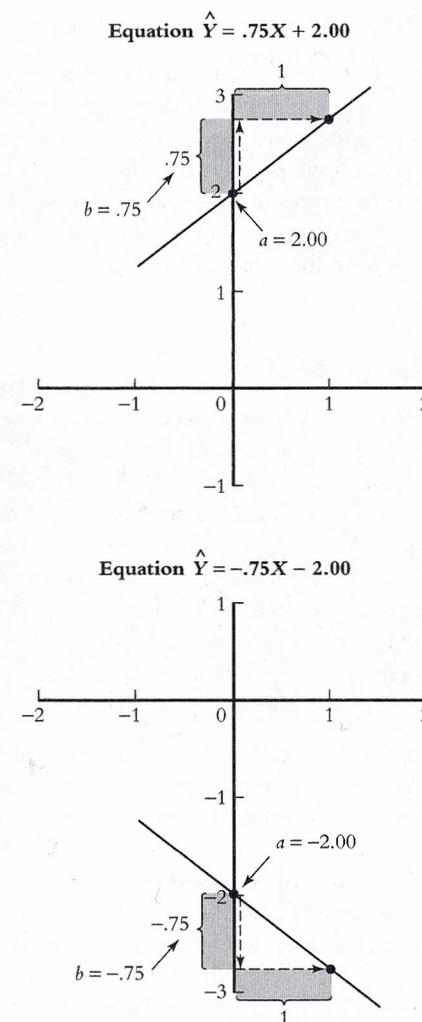


Figure 6-7 Plotting a regression line from the regression equation. At the top, the regression equation is $\hat{Y} = .75X + 2.00$, in which $a = 2.00$ and $b = .75$. Locate one point of the regression line on the vertical axis at $a = 2.00$. Determine the second point by moving up .75 unit ($b = .75$) and then to the right 1 unit. Then draw the regression line by connecting the two points. At the bottom, the regression equation is $\hat{Y} = -.75X - 2.00$, in which $a = -2.00$ and $b = -.75$. Again, locate one point on the vertical axis, but this time at $a = -2.00$. Determine the second point by moving *down* .75 unit (because b is negative; $b = -.75$) and then to the right 1 unit. Connect the two points. Alternatively, if (\bar{X}, \bar{Y}) is known, it and a constitute two points that specify the regression line.