

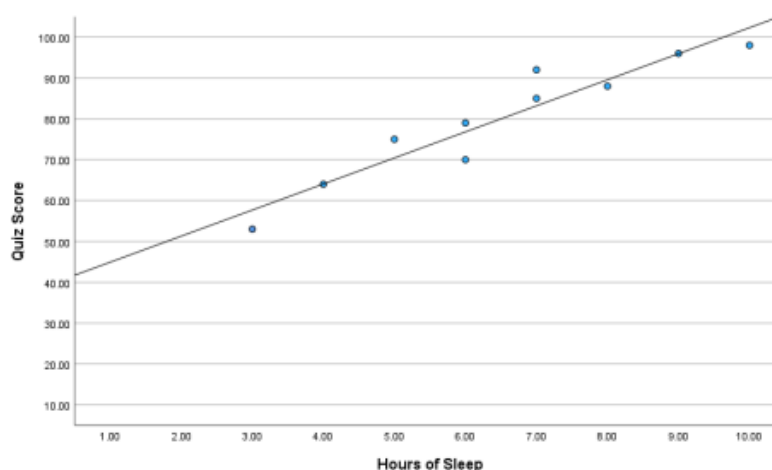
## 13.6: Visualizing Linear Regression

Regression builds for the same graph and foundation as correlation. Thus, bivariate regression can be graphed using a scatterplot and a regression line (See Chapter 12 for a review of scatterplots and fit lines). The **regression line** is used to summarize the change in  $Y$  that is associated with a change in  $X$ . The closer the dots tend to be to the regression line, the stronger the relationship and, thus, the more accurate the predictions will be based on the regression line. When there is a relationship between the variables, a regression line will either slope up or it will slope down when read from left to right. The angle of the line (i.e. the slope) indicates the direction of the correlation (either positive or negative).

Let's take a look at how regression works using Data Set 12.1 from Chapter 12. In Chapter 12, we supposed a researcher collected data from 10 college students to test the hypothesis that hours of sleep would positively relate to quiz scores. For regression we would hypothesize that hours of sleep would be useful in predicting quiz scores. Data Set 12.1 and the corresponding scatterplot for those data are included below for reference.

Data Set 12.1. Hours of Sleep and Quiz Scores ( $n = 10$ )

Participant Number	Sleep Hours	Quiz Score
1	7	92
2	8	88
3	9	96
4	6	70
5	6	79
6	4	64
7	5	75
8	10	98
9	3	53
10	7	85



Graph 12.2 Hours of Sleep and Quiz Scores with a Regression Line

### Predictions using Regression Lines

The stronger the correlation, the more accurate the predictions. This is because the regression line is being used to make predictions. Recall that a regression line is balanced to approximate the location of the dots with as little error as possible. Keep in mind that each dot is a bivariate data point. When we look at those data points in Graph 12.1, we can see that they are fairly close

to the positively sloping regression line. Visually, we can see that the line does a good job of estimating the location of the data and, thus, that it will likely be useful in estimating (predicting)  $Y$ -values using  $X$ -values.

### The Regression Equation

When regression is used, the regression line is being used to estimate scores of the  $Y$ -variable. Thus, the equation of the line is the formula used to predict (or estimate)  $Y$ -values. Linear graphs are summarized with the following equation:

$$\hat{Y} = b_0 + b_1 x$$

In this version of the linear equation  $\hat{Y}$  stands for a predicted  $Y$  value,  $b_0$  stands for the  $y$ -intercept of the line, and  $b_1$  stands for the slope of the line. A  $y$ -intercept is where the line crosses the  $y$ -axis; it represents what  $Y$  equals when  $X$  is zero. Notice that this is very similar to the structure some of us have seen before in math classes when we learned the linear equation; the version used in math classes is as follows:  $Y = mx + b$ . The symbols for the slope and  $y$ -intercept are different in statistics and math but they represent the same things. Specifically, in math the slope is called “ $m$ ” and the  $y$ -intercept is called “ $a$ ” while in statistics the slope is called “ $b_1$ ” and the  $y$ -intercept is called “ $b_0$ ”. In statistics, the  $y$ -intercept,  $b_0$ , is also often referred to as a constant because it refers to a single point on the graph. In addition, in statistics we often state the  $y$ -intercept first followed by slope times  $X$  whereas in math we may see these reversed (where the slope times  $X$  is stated first and the  $y$ -intercept is then added). However, these are just two different ways of writing the same formula. Thus, the two formulas are computationally the same, even though the symbols look different.

The regression equation is what is used to both summarize the regression line and to make predictions. When the slope and  $y$ -intercept are known, an  $X$ -value can be plugged into the linear equation to predict a  $Y$ -value. Slope and  $y$ -intercept are rarely calculated by hand in regression and, instead, are generally calculated using software such as SPSS. Therefore, the process for hand-calculating the slope and  $y$ -intercept are not included in this chapter and, instead, we will focus on how to read and use these values using SPSS as we progress through this chapter.

---

This page titled [13.6: Visualizing Linear Regression](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .