

15.2.1: Using Linear Equations

Before we start practicing calculating all of the variables in a regression line equation, let's work a little with just the equation on its own.

Regression Line Equations

As we just learned, linear regression for two variables is based on a linear equation:

$$\hat{Y} = a + (b * X)$$

where a and b are constant numbers. What this means is that for every sample, the intercept (a) and the slope (b) will be the same for every score. The X score will change, and that affects Y (or predicted Y , or \hat{Y}). Some consider the predictor variable (X) as an IV and the outcome variable (Y) as the DV, but be careful that you aren't confusing prediction with causation!

We also just learned that the graph of a linear equation of the form $\hat{Y} = a + (b * X)$ is a straight line.

? Exercise 15.2.1.1

Is the following an example of a linear equation? Why or why not?

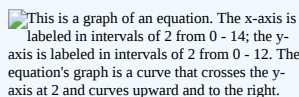
 This is a graph of an equation. The x-axis is labeled in intervals of 2 from 0 - 14; the y-axis is labeled in intervals of 2 from 0 - 12. The equation's graph is a curve that crosses the y-axis at 2 and curves upward and to the right.

Figure 15.2.1.1. Sample Plotted Line (CC-BY by [Barbara Illowsky & Susan Dean \(De Anza College\)](#) from [OpenStax](#))

Answer

No, the graph is not a *straight* line; therefore, it is not a linear equation.

The minimum criterion for using a linear regression formula is that there be a linear relationship between the predictor and the criterion (outcome) variables.

? Exercise 15.2.1.2

What statistic shows us whether two variables are linearly related?

Answer

Pearson's r (correlation).

If two variables aren't linearly related, then you can't use linear regression to predict one from the other! The stronger the linear relationship (larger the Pearson's correlation), the more accurate will be the predictions based on linear regression.

Slope and Y-Intercept of a Linear Equation

As we learned previously, b = slope and a = y -intercept. From algebra recall that the slope is a number that describes the steepness of a line, and the y -intercept is the y coordinate of the point $(0, a)$ where the line crosses the y -axis. Figure 15.2.1.2 shows three possible graphs of the regression equation ($y = a + bx$). Panel (a) shows what the regression line looks like if the slope is positive ($b > 0$), the line slopes upward to the right. Panel (b) shows what the regression line looks like if there's no slope ($b = 0$); the line is horizontal. Finally, Panel (c) shows what the regression line looks like if the slope is negative ($b < 0$), the line slopes downward to the right.

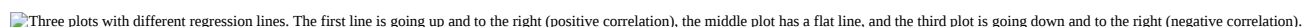
 Three plots with different regression lines. The first line is going up and to the right (positive correlation), the middle plot has a flat line, and the third plot is going down and to the right (negative correlation).

Figure 15.2.1.2: Three possible graphs of $y = a + bx$. (CC-BY by [Barbara Illowsky & Susan Dean \(De Anza College\)](#) from [OpenStax](#))

I get it, everything has been pretty theoretical so far. So let's get practical. Let's try constructing the regression line equation even when you don't have the scores for either of the variables. First, we'll start by identifying the variables in the examples.

✓ Example 15.2.1.1

Svetlana tutors to make extra money for college. For each tutoring session, she charges a one-time fee of \$25 plus \$15 per hour of tutoring. A linear equation that expresses the total amount of money Svetlana earns for each session she tutors is $y = 25 + 15x$.

What are the predictor and criterion (outcome) variables? What is the y -intercept and what is the slope? Answer using complete sentences.

Answer

The predictor variable, x , is the number of hours Svetlana tutors each session. The criterion (outcome) variable, y , is the amount, in dollars, Svetlana earns for each session.

The y -intercept is the constant, the one time fee of \$25 ($a = 25$). The slope is 15 ($b = 15$) because Svetlana earns \$15 for each hour she tutors.

Although it doesn't make sense in these examples, the y -intercept (a) is determined when $x = 0$. I guess with Svetlana, you could say that she gets \$25 for any sessions that you miss or don't cancel ahead of time. But geometrically and mathematically, the y -intercept is based on when the predictor variable (x) has a value of zero.

? Exercise 15.2.1.3

Jamal repairs household appliances like dishwashers and refrigerators. For each visit, he charges \$25 plus \$20 per hour of work. A linear equation that expresses the total amount of money Jamal earns per visit is $y = 25 + 20x$.

What are the predictor and criterion (outcome) variables? What is the y -intercept and what is the slope? Answer using complete sentences.

Answer

The predictor variable, x , is the number of hours Jamal works each visit. The criterion (outcome) variable, y , is the amount, in dollars, Jamal earns for each visit.

The y -intercept is 25 ($a = 25$). At the start of a visit, Jamal charges a one-time fee of \$25 (this is when $x = 0$). The slope is 20 ($b = 20$). For each visit, Jamal earns \$20 for each hour he works.

Now, we can start constructing the regression line equations.

✓ Example 15.2.1.2

Alejandra's Word Processing Service (AWPS) does word processing. The rate for services is \$32 per hour plus a \$31.50 one-time charge. The total cost to a customer depends on the number of hours it takes to complete the job.

Find the *equation* that expresses the total cost in terms of the number of hours required to complete the job. For this example,

- x = the number of hours it takes to get the job done.
- y = the total cost to the customer.

Answer

The \$31.50 is a fixed cost. This is the number that you add after calculating the rest, so it must be the intercept (a).

If it takes x hours to complete the job, then $(32)(x)$ is the cost of the word processing only.

Thus, the total cost is: $y = 31.50 + 32x$

Let's try another example of constructing the regression line equation.

? Exercise 15.2.1.4

Elektra's Extreme Sports hires hang-gliding instructors and pays them a fee of \$50 per class as well as \$20 per student in the class. The total cost Elektra pays depends on the number of students in a class. Find the equation that expresses the total cost in terms of the number of students in a class.

Answer

For this example,

- x = number of students in class
- y = the total cost

The constant is \$50 per class, so that must be the intercept (a).

So \$20 per student is the slope (b).

The resulting regression equation is: $y = 50 + 20x$

You can also use the regression equation to graph the line if you input scores from your X variable and your Y variable into the equation. Let's see what that might look like in Figure 15.2.1.3 for the equation: $y = -1 + 2x$

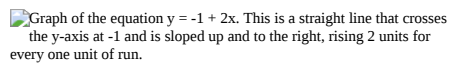
Graph of the equation $y = -1 + 2x$. This is a straight line that crosses the y-axis at -1 and is sloped up and to the right, rising 2 units for every one unit of run.

Figure 15.2.1.3: Regression Line for $y = -1 + 2x$. (CC-BY by Barbara Illowsky & Susan Dean (De Anza College) from OpenStax)

In the example in Figure 15.2.1.3 the intercept (a) is replaced by -1 and the slope (b) is replaced by 2 to get the regression equation ($y = -1 + 2x$). Right now, you are being provided these constants. Soon, you'll be calculating them yourself!

Summary

The most basic type of association is a linear association. This type of relationship can be defined algebraically by the equations used, numerically with actual or predicted data values, or graphically from a plotted. Algebraically, a linear equation typically takes the form $y = mx + b$, where m and b are constants, x is the independent variable, y is the dependent variable. In a statistical context, a linear equation is written in the form $y = a + bx$, where a and b are the constants. This form is used to help readers distinguish the statistical context from the algebraic context. In the equation $y = a + bx$, the constant b that multiplies the x variable (b is called a coefficient) is called the slope. The constant a is called the y -intercept.

The slope of a line is a value that describes the rate of change between the two quantitative variables. The slope tells us how the criterion variable (y) changes for every one unit increase in the predictor (x) variable, on average. The y -intercept is used to describe the criterion variable when the predictor variable equals zero.

This page titled 15.2.1: Using Linear Equations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michelle Oja.

- 12.2: Linear Equations by OpenStax is licensed CC BY 4.0. Original source: <https://openstax.org/details/books/introductory-statistics>.