

## Proof of why the Linear Regression Equation is the “Best Fit” Equation

In project #3, we saw that, for a given data set with linear correlation, the “best fit” regression equation is

$$\hat{y} = b_0 + b_1x$$

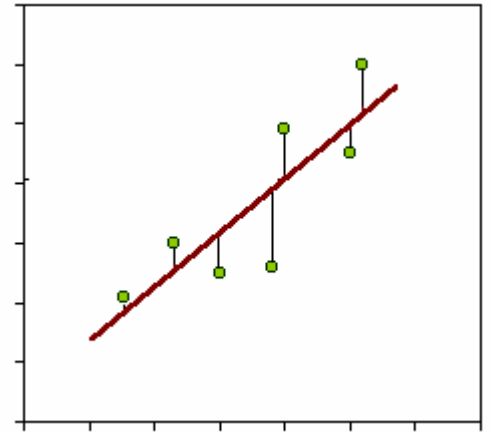
where

$$b_1 = \frac{n\sum(xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \text{ and } b_0 = \bar{y} - b_1\bar{x}.$$

This paper will prove why this is indeed the best fit line.

As you recall from regression, the regression line will not pass through each and every data point unless there is a perfect correlation. Since the  $y$  – values are predicted, and the data we use are observed, there will usually be some kind of difference between the predicted and observed  $y$  – values. These differences are called **residuals**.

Some residuals are positive (if the observed values lie above the best fit line), some are negative (if they lie below), and the rare ones are zero (if they lie exactly on the line. In the diagram at right, we see 4 positive residuals and 3 negative ones (best fit line is red).



The general form of a residual is as follows:

$$\text{residual} = \text{observed } y - \text{expected } y = y - \hat{y}$$

This looks a lot like the first part of the definition of standard deviation (and later, goodness of fit tests). A tempting idea would be to sum up the residuals and minimize that value. However, one small problem exists:

$$\sum (y - \hat{y}) = 0$$

Always. The same thing happened when we looked at standard deviation. So how did we alleviate that? We squared the results before summing. And that’s what we’ll do here: find the equation  $y = b_0 + b_1x$  such that

$$\sum (y - \hat{y})^2, \text{ or } \sum (y - b_0 - b_1x)^2 \text{ is as small as possible.}$$

Now...here’s where the explanation escapes the scope of our course. In order to minimize a function (and our best fit equation  $y = b_0 + b_1x$  is really a multi-variable function), we need to use something from calculus called *partial differentiation*, which is differentiation with respect to a certain variable. Let  $S = \sum (y - b_0 - b_1x)^2$ , the quantity which we want to minimize.

$$\frac{dS}{db_1} = -2\sum x(y - b_0 - b_1x)$$

$$\frac{dS}{db_0} = -2\sum (y - b_0 - b_1x)$$

Now...in order to find relative extrema (in this case, a minimum), we need to set each of these derivatives to zero<sup>1</sup>:

$$\begin{aligned}
 0 &= -2 \sum x(y - b_0 - b_1 x) \\
 0 &= \sum (xy - xb_0 - b_1 x^2) \\
 0 &= \sum xy - \sum xb_0 - \sum b_1 x^2 \\
 0 &= \sum xy - b_0 \sum x - b_1 \sum x^2 \quad (*)
 \end{aligned}$$

$$\begin{aligned}
 0 &= -2 \sum (y - b_0 - b_1 x) \\
 0 &= \sum y - \sum b_0 - \sum b_1 x \\
 0 &= \sum y - nb_0 - b_1 \sum x \\
 b_0 &= \frac{\sum y - b_1 \sum x}{n} \quad (**)
 \end{aligned}$$

Substituting (\*\*) into (\*), we have

$$0 = \sum xy - b_0 \sum x - b_1 \sum x^2 = \sum xy - \left( \frac{\sum y - b_1 \sum x}{n} \right) \sum x - b_1 \sum x^2$$

and then, after some rad algebra, we arrive at the desired result:

$$\begin{aligned}
 0 &= \sum xy - \left( \frac{\sum y - b_1 \sum x}{n} \right) \sum x - b_1 \sum x^2 \\
 0 &= \sum xy - \left( \frac{\sum x \sum y - b_1 \sum x \sum x}{n} \right) - b_1 \sum x^2 \\
 0 &= n \sum xy - \sum x \sum y + b_1 (\sum x)^2 - nb_1 \sum x^2 \\
 nb_1 \sum x^2 - b_1 (\sum x)^2 &= n \sum xy - \sum x \sum y \\
 b_1 (n \sum x^2 - (\sum x)^2) &= n \sum xy - \sum x \sum y \\
 b_1 &= \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}
 \end{aligned}$$

The derivation for  $b_0$  is far easier; simply look at (\*\*) in a different way :

$$\begin{aligned}
 b_0 &= \frac{\sum y - b_1 \sum x}{n} \\
 b_0 &= \frac{\sum y}{n} - \frac{b_1 \sum x}{n} \\
 b_0 &= \bar{y} - b_1 \frac{\sum x}{n} \\
 b_0 &= \bar{y} - b_1 \bar{x}
 \end{aligned}$$

---

<sup>1</sup> The derivative is the slope of the tangent line to the curve; if the slope is zero, the tangent is horizontal, which implies a high (or low) point. In this case, it must be a minimum, since the function  $S = \sum (y - b_0 - b_1 x)^2$  is quadratic in the two variables of interest.