Sum of Residuals

Eric Zheng

January 30, 2019

In this document, I'd like to present a short proof that the sum of the residuals in a least squares regression is equal to 0. The proof is actually well-known (and can easily be found via an online search), but it came to me as I was thinking about a question on our statistics midterm, and I could not resist the temptation to typeset it. (And since I'm a second-semester senior now, I can do whatever I want.)

1 Setting up the problem

In this section, I'll give a short introduction to the problem of the (linear) least squares regression model. If you're already familiar with least squares regression, you can probably skip this section.

In general, we are given a data set described by a bunch of points (x_i, y_i). Our goal is to find some linear model that "best" fits these data. In general, a linear model has the form:

$$\hat{y} = \beta_0 + \beta_1 x \tag{1}$$

with the two parameters β_0 and β_1 representing the y-intercept and slope respectively. Since our model is fully described by these two parameters, when we do a least squares regression, we're really trying to find the optimal values of β_0 and β_1.

But what do we mean by optimal? Usually, our model won't pass through all of the data points; i.e. there is an "error" between each actual data point (x_i, y_i) and its predicted location $(x_i, \hat{y}_i)$ based on the model. We can quantify this error and call it the *residual*; more precisely, we say that at each x_i, the residual is given by:

$$y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_i \tag{2}$$

Now, these residuals may be positive or negative (if the point lies above or below our line respectively), but we don't really care about the sign; all we want is the magnitude. So, we really care about the *squares* of the residuals. This is what we mean when we say a "least squares" regression analysis: we are trying to find a model, specified by the parameters β_0 and β_1, that minimizes the total squared residual. More precisely, if we quantify the total squared error in the model as a function:

$$F(\beta_0, \beta_1) = \sum (y_i - \beta_0 - \beta_1 x_i)^2 \tag{3}$$

we are trying to find β_0 and β_1 that minimizes F over the data points (x_i, y_i).

2 The proof

The hard part was really setting up the problem; with the tools of multivariate calculus, the proof is quite straightforward from here. We know that the minimum will occur when $\nabla F = 0$, which implies that:

$$\frac{\partial F}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \tag{4}$$

But wait, that term inside the sum is really the same as our definition of a residual from before! Thus, the sum of the residuals must be zero if we have found the true least squares regression line. (And, of course, the mean of the residuals is therefore also zero.) If you like, I will leave it up to you to apply the second partials test to show that this extremum is really a minimum.