Deriving the least squares estimates using summation operators. This information is provided for those of you who wish to know more about the derivation of least squares estimates.

1 Finding the least squares estimates of α and β

Recall that we want to minimize the sum of squared deviations from our line. How do we write the sum of squared deviations mathematically?

$$\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2$$

The Y_i is just the actual data value for each Y. The $\hat{Y}_i$ is our predicted value for Y_i which is based on the line we drew. The letter i is the index for our summation notation, and $\sum$ tells us to some up all the squared deviations from 1 to N, where N is the number of observations (data points) we have. Our goal is to minimize this sum of squared deviations. Note that the lowest sum of squared deviations you can ever have is zero since we are adding together numbers that must all either be positive (because we are squaring numbers) or zero.

In order to minimize the sum of squared deviations $\left(\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2\right)$ we first need to substitute in for $\hat{Y}_i$.

What can we substitute in for $\hat{Y}_i$? Since $\hat{Y}_i$ is our predicted value of Y_i, we know that $\hat{Y}_i$ will be given to us by the equation of our line. So $\hat{Y}_i = \alpha + \beta X_i$, where X_i is the X value that corresponds to the Y value. So we substitute in $\alpha + \beta X_i$ for $\hat{Y}_i$. We now have our sum of squared deviations as:

$$\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i)^2$$

Recall that to minimize a function we need to take the derivative and set it equal to zero. But just what are we taking the derivative of? Well, we have two unknowns, α and β, that we are trying to estimate, so we need to take the derivative of our function with respect to α and also with respect to β. So we need two derivatives. Actually, we will take partial derivatives, which are denoted by ∂ rather than total derivatives, which would be the normal dy/dx derivatives most people are probably used to. Partial derivatives are easy to take – they just assume that other variables in the equation are constants. For example, if we take the partial derivative of our sum of squared deviations with respect to α, we just treat β as if it were a constant. So:

$$\frac{\partial}{\partial\alpha}\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i)^2 \quad = \quad -2\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i)$$

$$\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i)^2 \quad = \quad -2\sum_{i=1}^{N}X_i(Y_i - \alpha - \beta X_i)$$

Now, set both equations equal to zero, and solve for α and β:

$$-2\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i) \quad = \quad 0$$

$$-2\sum_{i=1}^{N}X_i(Y_i - \alpha - \beta X_i) \quad = \quad 0$$

I am not going through all of the math to solve for α and β so I will just skip to the answers:

$$\alpha = \frac{\sum_{i=1}^{N} Y_i}{N} - \beta \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\beta = \frac{N \sum_{i=1}^{N} X_i Y_i - \sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i}{N \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}$$

Notice that β is written only in terms of X_i and Y_i, so we can calculate β directly from the observed data. As for α, notice that it includes a β in its solution as well as X_i and Y_i. This is fine because we know that β only consists of X_i and Y_i. We could plug in the formula for β into the formula for α so that we would just have X_i and Y_i in the formula for α, but this would lead to a very messy formula for α. We can now obtain our least squares estimates for α and β, and find the estimated model.

You should now be able to compute the least squares estimates for a regression model with a constant and one independent variable (but I do not suggest doing this by hand – computers are very good at crunching numbers).