Derivation of OLS Estimator

In class we set up the minimization problem that is the starting point for deriving the formulas for the OLS intercept and slope coefficient. That problem was,

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \tag{1}$$

As we learned in calculus, a univariate optimization involves taking the derivative and setting equal to 0. Similarly, this minimization problem above is solved by setting the partial derivatives equal to 0. That is, take the derivative of (1) with respect to $\hat{\beta}_0$ and set it equal to 0. We then do the same thing for $\hat{\beta}_1$. This gives us,

$$\frac{\partial W}{\partial \hat{\beta}_0} = \sum_{i=1}^{N} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{2}$$

and,

$$\frac{\partial W}{\partial \hat{\beta}_1} = \sum_{i=1}^{N} -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{3}$$

Note that I have used W to denote $\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. Now our task is to solve (2) and (3) using some algebra tricks and some properties of summations. Lets start with the first order condition for $\hat{\beta}_0$ (this is Equation (2)). We can immediately get rid of the -2 and write $\sum_{i=1}^{N} y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = 0$. Now lets rearrange this expression and make use of the algebraic fact that $\sum_{i=1}^{N} y_i = N\bar{y}$. This leaves us with,

$$N\hat{\beta}_0 = N\bar{y} - N\hat{\beta}_1\bar{x}. \tag{4}$$

We simply divide everything by N and amazing, we have the formula that Professor Sadoulet gave in lecture! That is,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}. \tag{5}$$

Now lets consider solving for $\hat{\beta}_1$. This one is a bit more tricky. We can first get rid of the -2 and rearrange Equation (3) to get $\sum_{i=1}^{N} x_i y_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2 = 0$. Now lets substitute our result for $\hat{\beta}_0$ into this expression and this gives us,

$$\sum_{i=1}^{N} x_i y_i - (\bar{y} - \hat{\beta}_1\bar{x})x_i - \hat{\beta}_1 x_i^2 = 0 \tag{6}$$

Note that the summation is applying to everything in the above equation. We can distribute the sum to each term to get,

$$\sum_{i=1}^{N} x_i y_i - \bar{y}\sum_{i=1}^{N} x_i + \hat{\beta}_1\bar{x}\sum_{i=1}^{N} x_i - \hat{\beta}_1\sum_{i=1}^{N} x_i^2 = 0. \tag{7}$$

We have of course used the property that you can always pull a constant term out in front of a summation. Lets again use the property that $\sum_{i=1}^{N} y_i = N\bar{y}$ (and of course this also means that $\sum_{i=1}^{N} x_i = N\bar{x}$). We apply these facts to Equation (7) and solve for $\hat{\beta}_1$. This gives,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}. \tag{8}$$

Doesn't quite look like the formula from class, right? Well, let us just use a couple more tricks. You can either look up or derive for yourself that $\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}$. You can also easily derive that $\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - N\bar{x}^2$. These two can be derived very easily using algebra. Now we substitute these two properties into (8) and we have something that looks very, very familiar:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}. \tag{9}$$

All done!