Deriving OLS

EC 320 Handout

AUTHOR
Andrew Dickinson

Getting started

Our simple linear regression model is

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

We can decompose Y_i into estimated components:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i.$$

Thus, for any regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ we obtain estimates of u_i called residuals ($\hat{u}_i$). Residuals are the "errors" of the estimated regression line:

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i.$$

By squaring $\hat{u}_i$ and summing across i, we get obtain the sum of squared residuals, also known as the residual sum of squares (**RSS**):

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \text{RSS}.$$

Objective function

Our objective is to pick the $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the residual sum of squares:

$$\min_{\hat{\beta}_0,\, \hat{\beta}_1} \quad \text{RSS} = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

We will start by expanding the objective function and collecting like terms:

$$
\begin{aligned}
\text{RSS} &= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \\
&= \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) \\
&= \sum_{i=1}^n (Y_i^2 - Y_i \hat{\beta}_0 - Y_i \hat{\beta}_1 X_i - \hat{\beta}_0 Y_i + \hat{\beta}_0^2 + \hat{\beta}_0 \hat{\beta}_1 X_i - \hat{\beta}_1 X_i Y_i + \hat{\beta}_1 X_i \hat{\beta}_0 + \hat{\beta}_1^2 X_i^2) \\
&= \sum_{i=1}^n (Y_i^2 - 2Y_i \hat{\beta}_0 - 2Y_i \hat{\beta}_1 X_i + \hat{\beta}_0^2 + 2\hat{\beta}_0 \hat{\beta}_1 X_i + \hat{\beta}_1^2 X_i^2).
\end{aligned}
$$

First-order conditions

Now we will find the $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the residual sum of squares. First we need to take partial derivatives with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and then set these derivatives equal to zero. From this we obtain our first-order conditions:

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} (-2Y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 X_i) = 0$$

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} (-2Y_i X_i + 2\hat{\beta}_0 X_i + 2\hat{\beta}_1 X_i^2) = 0$$

The first-order conditions give us everything we need to find expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Solve for the intercept formula

We will start by solving for $\hat{\beta}_0$. From , we can isolate the term involving $\hat{\beta}_0$ on the left hand by subtracting the other two terms from both sides:

$$2 \sum_{i=1}^{n} \hat{\beta}_0 = 2 \sum_{i=1}^{n} Y_i - 2\hat{\beta}_1 \sum_{i=1}^{n} X_i.$$

Divide by 2 to obtain

$$\sum_{i=1}^{n} \hat{\beta}_0 = \sum_{i=1}^{n} Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i.$$

Notice that $\sum_{i=1}^{n} Y_i = n \frac{1}{n} \sum_{i=1}^{n} Y_i = n\bar{Y}$. Then becomes

$$n\hat{\beta}_0 = n\bar{Y} - \hat{\beta}_1 n\bar{X}.$$

Dividing by n, we obtain a simple formula for the intercept:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Solve for the slope formula

Next, we will solve for $\hat{\beta}_1$. From , we have

$$-\sum_{i=1}^{n} Y_i X_i + \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0.$$

Now plug in the expression for $\hat{\beta}_0$:

$$-\sum_{i=1}^{n} Y_i X_i + (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$-\sum_{i=1}^{n} Y_i X_i + \bar{Y} \sum_{i=1}^{n} X_i - \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0.$$

Isolate the $\hat{\beta}_1$ terms on the left-hand side:

$$\hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i.$$

Dividing both sides by $\left(\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i \right)$, we obtain a formula for the slope coefficient:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i}.$$

Rearranging the slope formula

We now have an expression for $\hat{\beta}_1$ in terms of data on X and Y, but we can rearrange terms to get a more familiar expression. To do this, we are going to subtract zero from both the numerator and the denominator of $\hat{\beta}_1$. Notice that

$$\sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X}$$

$$= \sum_{i=1}^{n} X_i - n\bar{X}$$

$$= \sum_{i=1}^{n} X_i - n\frac{1}{n} \sum_{i=1}^{n} X_i$$

$$= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i$$

$$= 0.$$

Using this trick, we will now subtract zero. Our choice of zero is strategic. If you have trouble following this step, you can try working backwards from the end.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i - \bar{X} \sum_{i=1}^{n} (Y_i - \bar{Y})}{\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i - \bar{X} \sum_{i=1}^{n} (X_i - \bar{X})}.$$

Distribute terms and pull constants into the sums to obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i - \sum_{i=1}^{n} \bar{Y} X_i - \sum_{i=1}^{n} \bar{X} Y_i + \sum_{i=1}^{n} \bar{Y} \bar{X}}{\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} \bar{X} X_i - \sum_{i=1}^{n} \bar{X} X_i + \sum_{i=1}^{n} \bar{X}^2}.$$

By factoring we obtain a beautiful expression for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}.$$

OLS formulas

We have derived OLS!

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}$$

Second-order conditions

We still need to check for convexity to make sure we minimized the objective function. We would feel silly if we maximized it. To make sure we minimized RSS, we will take second derivatives and check to see if they are positive.

$$\frac{\partial^2 \mathrm{RSS}}{\partial \hat{\beta}_0^2} = \sum_{i=1}^{n} 2 > 0$$

$$\frac{\partial^2 \mathrm{RSS}}{\partial \hat{\beta}_1^2} = \sum_{i=1}^{n} 2X_i^2 > 0$$

Crisis averted. Now we can say that we've minimized the residual sum of squares.