Deriving the Least-Squares Estimates for Simple Linear Regression

Dr. Maria Tackett

This document contains the mathematical details for deriving the least-squares estimates for slope (β_1) and intercept (β_0). We obtain the estimates, $\hat{\beta}_1$ and $\hat{\beta}_0$ by finding the values that minimize the sum of squared residuals (1).

$$SSR = \sum_{i=1}^{n}[y_i - \hat{y}_i]^2 = [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 = [y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i]^2 \tag{1}$$

Recall that we can find the values of $\hat{\beta}_1$ and $\hat{\beta}_0$ that minimize (1) by taking the partial derivatives of (1) and setting them to 0. Thus, the values of $\hat{\beta}_1$ and $\hat{\beta}_0$ that minimize the respective partial derivative also minimize the sum of squared residuals. The partial derivatives are

$$\begin{aligned}
\frac{\partial \text{SSR}}{\partial \hat{\beta}_1} &= -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
\frac{\partial \text{SSR}}{\partial \hat{\beta}_0} &= -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)
\end{aligned} \tag{2}$$

Let's begin by deriving $\hat{\beta}_0$.

$$\begin{aligned}
\frac{\partial \text{SSR}}{\partial \hat{\beta}_0} &= -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\
&\Rightarrow -\sum_{i=1}^{n} (y_i + \hat{\beta}_0 + \hat{\beta}_1 x_i) = 0 \\
&\Rightarrow -\sum_{i=1}^{n} y_i + n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \\
&\Rightarrow n\hat{\beta}_0 = \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i \\
&\Rightarrow \hat{\beta}_0 = \frac{1}{n}\Big(\sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i\Big) \\
&\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned} \tag{3}$$

Now, we can derive $\hat{\beta}_1$ using the $\hat{\beta}_0$ we just derived

$$\frac{\partial \text{SSR}}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Rightarrow -\sum_{i=1}^{n} x_i y_i + \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$(\text{Fill in } \hat{\beta}_0) \Rightarrow -\sum_{i=1}^{n} x_i y_i + (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\Rightarrow \bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{4}$$

$$\Rightarrow n\bar{y}\bar{x} - \hat{\beta}_1 n\bar{x}^2 + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\Rightarrow \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \hat{\beta}_1 n\bar{x}^2 = \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}$$

$$\Rightarrow \hat{\beta}_1 \Big(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \Big) = \sum_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}$$

$$\hat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}}{\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

To write $\hat{\beta}_1$ in a form that's more recognizable, we will use the following:

$$\sum x_i y_i - n\bar{y}\bar{x} = \sum (x - \bar{x})(y - \bar{y}) = (n-1)\text{Cov}(x, y) \tag{5}$$

$$\sum x_i^2 - n\bar{x}^2 - \sum (x - \bar{x})^2 = (n-1)s_x^2 \tag{6}$$

where $\text{Cov}(x, y)$ is the covariance of x and y, and s_x^2 is the sample variance of x (s_x is the sample standard deviation).

Thus, applying (5) and (6), we have

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum\limits_{i=1}^{n} x_i y_i - n\bar{y}\bar{x}}{\sum\limits_{i=1}^{n} x_i^2 - n\bar{x}^2} \\
&= \frac{\sum\limits_{i=1}^{n} (x-\bar{x})(y-\bar{y})}{\sum\limits_{i=1}^{n} (x-\bar{x})^2} \\
&= \frac{(n-1)\mathrm{Cov}(x,y)}{(n-1)s_x^2} \\
&= \frac{\mathrm{Cov}(x,y)}{s_x^2}
\end{aligned} \tag{7}
$$

The correlation between x and y is $r = \frac{\mathrm{Cov}(x,y)}{s_x s_y}$. Thus, $\mathrm{Cov}(x,y) = r s_x s_y$. Plugging this into (7), we have

$$
\hat{\beta}_1 = \frac{\mathrm{Cov}(x,y)}{s_x^2} = r\frac{s_y s_x}{s_x^2} = r\frac{s_y}{s_x} \tag{8}
$$