1 Errors (ϵ) in OLS

In ***ordinary least square*** **(OLS)**, we compute the **squared errors against the line** (SE_{line}) and let it take the minimum value. By the definition of SE_{line}:

$$\text{SE}_{\text{line}} = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2$$

Now we want to find the values β_0 and β_1, such that SE_{line} takes the minimum value. Therefore, we should have:

$$\frac{\partial \text{SE}_{\text{line}}}{\partial \beta_0} = 0\,, \text{and } \frac{\partial \text{SE}_{\text{line}}}{\partial \beta_1} = 0$$

Now, let's first re-write SE_{line} with respect to β_0, *i.e.* using β_0 as the variable:

$$\begin{aligned}
\text{SE}_{\text{line}} &= \sum_{i=1}^{n}[y_i^2 - 2y_i(\beta_0 + \beta_1 x_i) + (\beta_0 + \beta_1 x_i)^2] \\
&= \sum_{i=1}^{n}[y_i^2 - 2y_i\beta_0 - 2y_i\beta_1 x_i + \beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2] \\
&= \sum_{i=1}^{n}[\beta_0^2 + (2\beta_1 x_i - 2y_i)\beta_0 + (y_i^2 - 2y_i\beta_1 x_i + \beta_1^2 x_i^2)]
\end{aligned}$$

Now we let $\dfrac{\partial \text{SE}_{\text{line}}}{\partial \beta_0} = 0$, we have:

$$\frac{\partial \text{SE}_{\text{line}}}{\partial \beta_0} = \sum_{i=1}^{n}[2\beta_0 + (2\beta_1 x_i - 2y_i)] = 0$$

Divide by 2 at both sides, we have:

$$\sum_{i=1}^{n}(\beta_0 + \beta_1 x_i - y_i) = 0 \Rightarrow \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

Note that by definition, $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$. Therefore, we have:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = \sum_{i=1}^{n}\epsilon_i = 0 \tag{1}$$

Similarly, re-write $\mathrm{SE_{line}}$ with respect to β_1, *i.e.* using β_1 as the variable:

$$\mathrm{SE_{line}} = \sum_{i=1}^{n}[x_i^2\beta_1^2 + (2\beta_0 x_i - 2x_i y_i)\beta_1 + (y_i^2 - 2y_i\beta_0 + \beta_0^2)]$$

Now, we let $\dfrac{\partial \mathrm{SE_{line}}}{\partial \beta_1} = 0$, we have:

$$\frac{\partial \mathrm{SE_{line}}}{\partial \beta_1} = \sum_{i=1}^{n}(2x_i^2\beta_1 + 2\beta_0 x_i - 2x_i y_i) = 0$$

Divide by 2 at both sides, we have:

$$\sum_{i=1}^{n}(x_i^2\beta_1 + \beta_0 x_i - x_i y_i) = 0 \Rightarrow \sum_{i=1}^{n}(y_i - \beta_0 - x_i\beta_1)x_i = 0$$

Again, note that $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$ by definition. Therefore, we have:

$$\sum_{i=1}^{n}(y_i - \beta_0 - x_i\beta_1)x_i = \sum_{i=1}^{n}x_i\epsilon_i = 0 \tag{2}$$

Equations (1) and (2) are very important properties in OLS. They are the constraints that used up two degree of freedoms.

2 SST = SSR + SSE

During the lecture, we demonstrated that for each observation, the total deviation of y_i from its mean $\bar{y}$ consists of two parts: unexplained deviation due to error and deviation explained by the regression line. That is:

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

Once we collect the deviation for all observations and sum them up, we have:

$$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

We want to prove that SST = SSE + SSR.

Proof. We start with:

$$\begin{aligned}
\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 &= \sum_{i=1}^{n}[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\
&= \sum_{i=1}^{n}[(y_i - \hat{y}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})] \\
&= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= \text{SSE} + \text{SSR} + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})
\end{aligned}$$

Now we only need to prove that $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$. Expand the terms, we have:

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(\beta_0 + \beta_1 x_i - \bar{y}) \\
&= \sum_{i=1}^{n}[(y_i - \beta_0 - \beta_1 x_i)(\beta_0 - \bar{y}) + (y_i - \beta_0 - \beta_1 x_i)\beta_1 x_i] \\
&= \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(\beta_0 - \bar{y}) + \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)\beta_1 x_i \\
&= (\beta_0 - \bar{y})\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) + \beta_1\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i
\end{aligned}$$

Note that under the assumptions of OLS, both red terms are 0 according to equations (1) and (2). $\square$