Distributions of OLS Coefficients

BIO210 Biostatistics

Extra Reading Material for Lecture 39

Xi Chen
School of Life Sciences
Southern University of Science and Technology

Spring 2024

During the lecture, we said that we use ordinary least squares (OLS) in simple linear regression not only because it is easy to understand and apply, but also because the resulting coefficients have nice distributions. Furthermore, the OLS estimators for the coefficients are unbiased, which means that on average, they accurately estimate the true values of the population parameters.

Under the assumptions of OLS, the slope, the intercept and the predicted value all follow normal distributions. Now let's just have a look at their means and variances.

Before we proceed, let's clarify our notations. In our population, we are interested in two joint random variable $(\boldsymbol{X}, \boldsymbol{Y})$. We think there is a linear relationship between them and would like to investigate the relationship. That is, the relationship between $\boldsymbol{Y}|\boldsymbol{X}$ and $\boldsymbol{X}$. The model is:

$$\boldsymbol{Y}|\boldsymbol{X} = \beta_0 + \beta_1 \boldsymbol{X} + \boldsymbol{\epsilon}$$

For each specific data point, we focused on the value of $\boldsymbol{Y}$ whenever $\boldsymbol{X}$ takes a value x. Due to this reason, we treat $\boldsymbol{X}$ as the known in simple linear regression, because we can only look at $\boldsymbol{Y}$ whenever $\boldsymbol{X}$ takes a specific value. Therefore, we tend to write x in the lower case.

$\boldsymbol{\epsilon}$ is the error, which is a random variable and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\epsilon}}^2)$ under the assumptions of OLS, where $\sigma_{\boldsymbol{\epsilon}}^2$ is called the **common variance of the error**. We would like to stick to our convention to use upper-case letters for random variables, but the upper-case of $\boldsymbol{\epsilon}$ is basically $\boldsymbol{E}$, which will cause confusing. Therefore, we will use the bold symbol $\boldsymbol{\epsilon}$ to represent the random variable and the regular symbol ϵ the value.

Similarly, our population parameters include the population slope β_1 and the population intercept β_0. We will use the bold symbols $\boldsymbol{\hat{\beta}_1}$ and $\boldsymbol{\hat{\beta}_0}$ to represent their estimators, respectively, and the regular symbols $\hat{\beta}_1$ and $\hat{\beta}_0$ their estimates, respectively.

When we have a sample with size n, we have (x_i, y_i) pairs. In OLS, we used the following formula to estimate the population parameters:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and } \beta_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x}$$

Each time we have a different samples, those values will change. Therefore, both the slope and intercept are random variables. We could write them in the estimator format:

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and } \boldsymbol{\beta}_{\mathbf{0}} = \bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}} \cdot \bar{x}$$

For each specific data points:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i = \hat{y} + \epsilon_i$$

The sample regression line is (note that we re-write $\hat{y}$ to $\hat{\boldsymbol{\mu}}_{y|x}$):

$$\hat{\boldsymbol{\mu}}_{y|x} = \hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}} x$$

Here, it is time to summarise the distributions of all those random variables. They have nice distributions. That is, they are all normally distributed:

Summary

$$\boldsymbol{\epsilon} \sim \mathcal{N}(0,\ \sigma_{\boldsymbol{\epsilon}}^2)$$

$$\boldsymbol{\beta}_{\mathbf{1}} \sim \mathcal{N}\left(\beta_1,\ \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\sum_{i=1}^{n}[x_i - \bar{x}]^2}\right)$$

$$\boldsymbol{\beta}_{\mathbf{0}} \sim \mathcal{N}\left(\beta_0,\ \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

$$\hat{\boldsymbol{\mu}}_{y|x} \sim \mathcal{N}\left(\hat{\mu}_{y|x},\ \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]\right)$$

The distribution of the error is our assumption. Let's see if we can make sense the rest three.

1 $\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \beta_1$

The OLS estimator for the slope is:

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Proof. To simplify notation, we denote

$$\mathrm{S}_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

which is just a number. Therefore, we have:

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})}{\mathrm{S}_{xx}}\right] = \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\sum_{i=1}^{n}\left[(x_i - \bar{x})\boldsymbol{Y}_i - (x_i - \bar{x})\bar{\boldsymbol{Y}}\right]\right] \\
&= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\sum_{i=1}^{n}(x_i - \bar{x})\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\sum_{i=1}^{n}(x_i - \bar{x})\right]
\end{aligned}$$

Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$, we have:

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] &= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\sum_{i=1}^{n}(x_i - \bar{x})\boldsymbol{Y}_i\right] = \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + \boldsymbol{\epsilon}_i)\right] \qquad (1) \\
&= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\sum_{i=1}^{n}\left(\beta_0 x_i + \beta_1 x_i^2 + x_i\boldsymbol{\epsilon}_i - \beta_0\bar{x} - \beta_1 x_i\bar{x} - \bar{x}\boldsymbol{\epsilon}_i\right)\right] \\
&= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\beta_0\sum_{i=1}^{n}x_i + \beta_1\sum_{i=1}^{n}x_i^2 + \sum_{i=1}^{n}x_i\boldsymbol{\epsilon}_i - \beta_0\sum_{i=1}^{n}\bar{x} - \beta_1\bar{x}\sum_{i=1}^{n}x_i - \bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\right]
\end{aligned}$$

Note that $\sum_{i=1}^{n}x_i = n\bar{x}$. In addition, $\boldsymbol{\epsilon}_i$ are the random variables and everything else are just constants, so the above equation becomes:

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] &= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\beta_0 n\bar{x} + \beta_1\sum_{i=1}^{n}x_i^2 + \sum_{i=1}^{n}x_i\boldsymbol{\epsilon}_i - \beta_0 n\bar{x} - \beta_1 n\bar{x}^2 - \bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\right] \\
&= \frac{1}{\mathrm{S}_{xx}}\mathbb{E}\left[\beta_1\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}\right) + \sum_{i=1}^{n}x_i\boldsymbol{\epsilon}_i - \bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\right] \\
&= \frac{1}{\mathrm{S}_{xx}}\left[\beta_1\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}\right) + \mathbb{E}\left[\sum_{i=1}^{n}x_i\boldsymbol{\epsilon}_i\right] - \mathbb{E}\left[\bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i\right]\right]
\end{aligned}$$

Under the assumptions of OLS, the magenta terms are both 0, so we have:

$$\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \beta_1 \cdot \frac{\sum_{i=1}^{n}x_i^2 - n\bar{x}}{\mathrm{S}_{xx}}$$

Since the numerator and the denominator in the fraction is the same[1], we have:

$$\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \beta_1$$

$\square$

2 $\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) = \dfrac{\sigma_{\boldsymbol{\epsilon}}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Just like how we derived the expectation of $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, you just need to be patient. There are more than one way of achieving this goal.

Proof. **Method 1: derivation by definition**. This is a hard way. Start with the second equality of equation (1), we have:

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) &= \mathbb{E}\left[\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right]\right)^2\right] = \mathbb{E}\left[\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}} - \beta_1\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i - \beta_1 \cdot \mathrm{S}_{xx}}{\mathrm{S}_{xx}}\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i - \sum_{i=1}^n \beta_1 (x_i - \bar{x})(x_i - \bar{x})}{\mathrm{S}_{xx}}\right)^2\right] \qquad (2)
\end{aligned}$$

Now let's do some algebraic manipulations of the numerator of the above equation:

$$\begin{aligned}
&\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i - \sum_{i=1}^n \beta_1 (x_i - \bar{x})(x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})[\boldsymbol{Y}_i - \beta_1(x_i - \bar{x})] \\
&= \sum_{i=1}^n (x_i - \bar{x})[\beta_0 + \beta_1 x_i + \boldsymbol{\epsilon}_i - \beta_1(x_i - \bar{x})] \\
&= \sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 \bar{x} + \boldsymbol{\epsilon}_i) = \sum_{i=1}^n (\beta_0 x_i + \beta_1 \bar{x} x_i + x_i \boldsymbol{\epsilon}_i - \beta_o \bar{x} - \beta_1 \bar{x}^2 - \bar{x}\boldsymbol{\epsilon}_i) \\
&= \beta_0 \sum_{i=1}^n x_i + \beta_1 \bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n x_i \boldsymbol{\epsilon}_i - \sum_{i=1}^n \beta_o \bar{x} - \sum_{i=1}^n \beta_1 \bar{x}^2 - \bar{x} \sum_{i=1}^n \boldsymbol{\epsilon}_i \qquad (3)
\end{aligned}$$

[1] $\mathrm{S}_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2$, replace $\sum_{i=1}^n x_i$ with $n\bar{x}$, we have $\mathrm{S}_{xx} = \sum_{i=1}^n x_i^2 - n\bar{x}^2$

Note again $\sum_{i=1}^{n} x_i = n\bar{x}$. Therefore, equation (3) becomes:

$$\begin{aligned}
&\sum_{i=1}^{n}(x_i - \bar{x})\boldsymbol{Y}_i - \sum_{i=1}^{n}\beta_1(x-\bar{x})(x-\bar{x}) \\
&= \beta_0 n\bar{x} + \beta_1 n\bar{x}^2 + \sum_{i=1}^{n} x_i\boldsymbol{\epsilon}_i - n\beta_0\bar{x} - n\beta_1\bar{x}^2 - \bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i \\
&= \sum_{i=1}^{n} x_i\boldsymbol{\epsilon}_i - \bar{x}\sum_{i=1}^{n}\boldsymbol{\epsilon}_i = \sum_{i=1}^{n}\boldsymbol{\epsilon}_i(x_i-\bar{x}) \qquad (4)
\end{aligned}$$

Now put equation (4) back to equation (2), we have:

$$\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\boldsymbol{1}}\right) = \mathbb{E}\left[\frac{\left[\sum_{i=1}^{n}\boldsymbol{\epsilon}_i(x_i-\bar{x})\right]^2}{\mathrm{S}_{xx}^2}\right] = \frac{1}{\mathrm{S}_{xx}^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\boldsymbol{\epsilon}_i(x_i-\bar{x})\right)^2\right] \qquad (5)$$

Look at the expectation term, and expand the polynomial using the multinomial theorem, we have:

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{n}\boldsymbol{\epsilon}_i(x_i-\bar{x})\right)^2\right] &= \mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^2(x_i-\bar{x})^2 + 2\sum_{i\neq j}\boldsymbol{\epsilon}_i(x_i-\bar{x})\boldsymbol{\epsilon}_j(x_j-\bar{x})\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^2(x_i-\bar{x})^2\right] + 2\mathbb{E}\left[\sum_{i\neq j}\boldsymbol{\epsilon}_i(x_i-\bar{x})\boldsymbol{\epsilon}_j(x_j-\bar{x})\right] \qquad (6)
\end{aligned}$$

Under OLS assumptions, the errors are independent, so $\mathbb{E}\left[\boldsymbol{\epsilon}_i \cdot \boldsymbol{\epsilon}_j\right] = \mathbb{E}\left[\boldsymbol{\epsilon}_i\right] \cdot \mathbb{E}\left[\boldsymbol{\epsilon}_j\right]$ $= 0$, the second term of equation (6) becomes 0 like this:

$$\begin{aligned}
&2\mathbb{E}\left[\sum_{i\neq j}\boldsymbol{\epsilon}_i(x_i-\bar{x})\boldsymbol{\epsilon}_j(x_j-\bar{x})\right] = 2\sum_{i\neq j}\mathbb{E}\left[\boldsymbol{\epsilon}_i(x_i-\bar{x})\boldsymbol{\epsilon}_j(x_j-\bar{x})\right] \\
&= 2\sum_{i\neq j}(x_i-\bar{x})(x_j-\bar{x})\mathbb{E}\left[\boldsymbol{\epsilon}_i\cdot\boldsymbol{\epsilon}_j\right] = 0 \qquad (7)
\end{aligned}$$

Similarly, the first term of equation (6) becomes:

$$\mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{\epsilon}_i^2(x_i-\bar{x})^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\boldsymbol{\epsilon}_i^2(x_i-\bar{x})^2\right] = \sum_{i=1}^{n}(x_i-\bar{x})^2\mathbb{E}\left[\boldsymbol{\epsilon}_i^2\right]$$

Note the errors have a common variance: $\mathbb{V}\mathrm{ar}\left(\boldsymbol{\epsilon}_1\right) = \mathbb{V}\mathrm{ar}\left(\boldsymbol{\epsilon}_2\right) = \cdots = \mathbb{V}\mathrm{ar}\left(\boldsymbol{\epsilon}_n\right) =$

$\sigma_{\boldsymbol{\epsilon}}^2$, we have:

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \boldsymbol{\epsilon}_i^2 (x_i - \bar{x})^2\right] &= \sum_{i=1}^{n} (x_i - \bar{x})^2 \mathbb{E}\left[\boldsymbol{\epsilon}_i^2\right] = \sum_{i=1}^{n} (x_i - \bar{x})^2 \sigma_{\boldsymbol{\epsilon}}^2 \\
&= \sigma_{\boldsymbol{\epsilon}}^2 \cdot \mathrm{S}_{xx}
\end{aligned} \tag{8}
$$

Put equations (7) and (8) back to equation (6), we have:

$$
\mathbb{E}\left[\left(\sum_{i=1}^{n} \boldsymbol{\epsilon}_i (x_i - \bar{x})\right)^2\right] = \sigma_{\boldsymbol{\epsilon}}^2 \cdot \mathrm{S}_{xx}
$$

Put the above term back to equation (5), we have:

$$
\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) = \frac{1}{\mathrm{S}_{xx}^2} \cdot \sigma_{\boldsymbol{\epsilon}}^2 \cdot \mathrm{S}_{xx} = \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}} = \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
$$

□

That really involves in a lot of algebraic manipulation. Actually, there is a much simpler way.

Proof. **Method 2: using the fact that $\boldsymbol{Y}_i$ are independent.** We still should start with similar manipulations from equation (1):

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathbf{1}} &= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})}{\mathrm{S}_{xx}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\boldsymbol{Y}_i - \bar{\boldsymbol{Y}} \sum_{i=1}^{n} (x_i - \bar{x})}{\mathrm{S}_{xx}} \\
&= \frac{\sum_{i=1}^{n} (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}} \\
\Rightarrow \mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) &= \mathbb{V}\mathrm{ar}\left(\frac{\sum_{i=1}^{n} (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}}\right) \\
&= \mathbb{V}\mathrm{ar}\left(\frac{x_1 - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_1 + \frac{x_2 - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_2 + \cdots + \frac{x_n - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_n\right)
\end{aligned}
$$

Since $\boldsymbol{Y}_i$ are independent of each other, the variance of the sum is the sum of the variances. Then we have:

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) = \mathbb{V}\mathrm{ar}\left(\frac{x_1 - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_1\right) + \mathbb{V}\mathrm{ar}\left(\frac{x_2 - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_2\right) + \cdots + \mathbb{V}\mathrm{ar}\left(\frac{x_n - \bar{x}}{\mathrm{S}_{xx}} \cdot \boldsymbol{Y}_n\right) \\
&= \frac{(x_1 - \bar{x})^2}{\mathrm{S}_{xx}^2} \cdot \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_1\right) + \frac{(x_2 - \bar{x})^2}{\mathrm{S}_{xx}^2} \cdot \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_2\right) + \cdots + \frac{(x_n - \bar{x})^2}{\mathrm{S}_{xx}^2} \cdot \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_n\right)
\end{aligned} \tag{9}
$$

Note that each $\boldsymbol{Y}_i$ has the same variance:

$$\mathbb{V}\text{ar}\left(\boldsymbol{Y}_i\right) = \mathbb{V}\text{ar}\left(\beta_0 + \beta_1 x_i + \boldsymbol{\epsilon}_i\right) = \mathbb{V}\text{ar}\left(\boldsymbol{\epsilon}_i\right) = \sigma^2_{\boldsymbol{\epsilon}} \tag{10}$$

Putting equation (10) back to equation (9), we have:

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right) &= \frac{(x_1-\bar{x})^2}{\mathrm{S}^2_{xx}}\cdot\sigma^2_{\boldsymbol{\epsilon}} + \frac{(x_2-\bar{x})^2}{\mathrm{S}^2_{xx}}\cdot\sigma^2_{\boldsymbol{\epsilon}} + \cdots + \frac{(x_n-\bar{x})^2}{\mathrm{S}^2_{xx}}\cdot\sigma^2_{\boldsymbol{\epsilon}} \\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\mathrm{S}^2_{xx}}\cdot\sigma^2_{\boldsymbol{\epsilon}} = \frac{\sigma^2_{\boldsymbol{\epsilon}}}{\mathrm{S}_{xx}} \\
&= \frac{\sigma^2_{\boldsymbol{\epsilon}}}{\sum_{i=1}^n (x_i-\bar{x})^2}
\end{aligned}$$

□

3 $\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right] = \beta_0$

Since we already figured out $\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \beta_1$. the expectation of the intercept $\boldsymbol{\beta}_{\mathbf{0}}$ can be easily proved by using the fact that the sample regression line always passes through the point $(\bar{x}, \bar{y})$ under OLS, that is, $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$.

Proof. We start with the way how we get $\boldsymbol{\beta}_{\mathbf{0}}$:

$$\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right] &= \mathbb{E}\left[\bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x}\right] = \mathbb{E}\left[\bar{\boldsymbol{Y}}\right] - \mathbb{E}\left[\bar{x}\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{Y}_i\right] - \bar{x}\cdot\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^n \boldsymbol{Y}_i\right] - \bar{x}\cdot\beta_1 \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\boldsymbol{Y}_i\right] - \bar{x}\cdot\beta_1
\end{aligned} \tag{11}$$

Note that:

$$\mathbb{E}\left[\boldsymbol{Y}_i\right] = \mathbb{E}\left[\beta_0 + \beta_1 x_i + \boldsymbol{\epsilon}_i\right] = \beta_0 + \beta_1 x_i + \mathbb{E}\left[\boldsymbol{\epsilon}_i\right] = \beta_0 + \beta_1 x_i \tag{12}$$

Put equation (12) back to equation (11), we have:

$$
\begin{aligned}
\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right] &= \frac{1}{n}\sum_{i=1}^{n}(\beta_0+\beta_1 x_i) - \bar{x}\cdot\beta_1 \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\beta_0 + \beta_1\sum_{i=1}^{n}x_i\right) - \bar{x}\cdot\beta_1 \\
&= \frac{1}{n}(n\beta_0 + \beta_1\cdot n\bar{x}) - \bar{x}\cdot\beta_1 \\
&= \beta_0
\end{aligned}
$$

□

4 $\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right) = \sigma_{\boldsymbol{\epsilon}}^2\cdot\left[\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$

Here, we could use a similar strategy, starting with the fact that the point $(\bar{x},\bar{y})$ must be in the sample regression line, so $\bar{\boldsymbol{Y}} = \hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x}$

Proof.

$$\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right) = \mathbb{V}\mathrm{ar}\left(\bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x}\right) = \mathbb{V}\mathrm{ar}\left(\bar{\boldsymbol{Y}} + [-\bar{x}]\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right)$$

You see here, we already computed $\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right)$ previously, and $\mathbb{V}\mathrm{ar}\left(\bar{\boldsymbol{Y}}\right)$ is straightforward to calculate. If we know that $\bar{\boldsymbol{Y}}$ and $(-\bar{x})\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ are independent, then we are done. However, proving they are independent is slightly more complicated, which we will come back in a later time.

For now, we re-write $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ similar to the previous section:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right) &= \mathbb{V}\mathrm{ar}\left(\bar{\boldsymbol{Y}} - \bar{x}\cdot\frac{\sum_{i=1}^{n}(x_i-\bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}}\right) = \mathbb{V}\mathrm{ar}\left(\frac{\mathrm{S}_{xx}\cdot\bar{\boldsymbol{Y}} - \bar{x}\sum_{i=1}^{n}(x_i-\bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}}\right) \\
&= \frac{1}{\mathrm{S}_{xx}^2}\mathbb{V}\mathrm{ar}\left(\mathrm{S}_{xx}\cdot\bar{\boldsymbol{Y}} - \bar{x}\sum_{i=1}^{n}(x_i-\bar{x})\boldsymbol{Y}_i\right) \qquad (13)
\end{aligned}
$$

Taking the term inside the parentheses of equation (13) out to simply:

$$
\begin{aligned}
&\mathrm{S}_{xx} \cdot \bar{\boldsymbol{Y}} - \bar{x}\sum_{i=1}^{n}(x_i - \bar{x})\boldsymbol{Y}_i = \mathrm{S}_{xx} \cdot \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Y}_i - \sum_{i=1}^{n}\bar{x}(x_i - \bar{x})\boldsymbol{Y}_i \\
&= \left[\frac{\mathrm{S}_{xx}}{n} \cdot \boldsymbol{Y}_1 - \bar{x}(x_1 - \bar{x})\boldsymbol{Y}_1\right] + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} \cdot \boldsymbol{Y}_n - \bar{x}(x_1 - \bar{x})\boldsymbol{Y}_n\right] \\
&= \left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right] \cdot \boldsymbol{Y}_1 + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right] \cdot \boldsymbol{Y}_n
\end{aligned}
$$

Putting back to equation (13) and noting that all $\boldsymbol{Y}_i$ are independent and have a common variance $\sigma_{\boldsymbol{\epsilon}}^2$, we have:

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right) = \frac{1}{\mathrm{S}_{xx}^2}\mathbb{V}\mathrm{ar}\left(\left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right] \cdot \boldsymbol{Y}_1 + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right] \cdot \boldsymbol{Y}_n\right) \\
&= \frac{1}{\mathrm{S}_{xx}^2} \cdot \left[\mathbb{V}\mathrm{ar}\left(\left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right] \cdot \boldsymbol{Y}_1\right) + \cdots + \mathbb{V}\mathrm{ar}\left(\left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_n - \bar{x})\right] \cdot \boldsymbol{Y}_n\right)\right] \\
&= \frac{1}{\mathrm{S}_{xx}^2} \cdot \left(\left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right]^2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_1\right) + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_n - \bar{x})\right]^2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_n\right)\right) \\
&= \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}^2} \cdot \left(\left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_1 - \bar{x})\right]^2 + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} - \bar{x}(x_n - \bar{x})\right]^2\right) \\
&= \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}^2} \cdot \left[n \cdot \frac{\mathrm{S}_{xx}^2}{n^2} - 2 \cdot \frac{\mathrm{S}_{xx}}{n} \cdot \bar{x}\sum_{i=1}^{n}(x_i - \bar{x}) + \bar{x}^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right] \\
&= \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}^2} \cdot \left[\frac{\mathrm{S}_{xx}^2}{n} + \bar{x}^2\mathrm{S}_{xx}\right] \\
&= \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{\mathrm{S}_{xx}}\right] \\
&= \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]
\end{aligned}
$$

□

5 $\mathbb{E}\left[\hat{\boldsymbol{\mu}}_{y|x}\right] = \mu_{y|x}$

Now that we have derived the expectation of the slope $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ and the intercept $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$, it is very easy for us to figure out the expectation of the predicted value $\hat{\boldsymbol{\mu}}_{y|x}$.

Proof. We start with the definition: $\hat{\boldsymbol{\mu}}_{y|x} = \hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}x$, so we have:

$$\mathbb{E}\left[\hat{\boldsymbol{\mu}}_{y|x}\right] = \mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}x\right] = \mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right] + x\mathbb{E}\left[\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right] = \beta_0 + \beta_1 x = \mu_{y|x}$$

□

6 $\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\mu}}_{y|x}\right) = \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]$

We are in a similar situation, if we know that $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ are independent, then we would get:

$$\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\mu}}_{y|x}\right) = \mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}x\right) = \mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right) + x^2\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right)$$

Since we already computed $\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}}\right)$ and $\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\right)$ previously, we are done now. We might come back to the proof that $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ are independent, but for now let's start from scratch. The procedures look very similar to previous ones.

Proof. First, we start with the following expressions what we encountered repeated in the previous sections:

$$\hat{\boldsymbol{\beta}}_{\mathbf{0}} = \bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x} \text{ and } \hat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}}$$

Now, we start with the definition of $\hat{\boldsymbol{\mu}}_{y|x}$:

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\hat{\boldsymbol{\mu}}_{y|x}\right) &= \mathbb{V}\text{ar}\left(\hat{\boldsymbol{\beta}}_{\mathbf{0}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}x\right) = \mathbb{V}\text{ar}\left(\bar{\boldsymbol{Y}} - \hat{\boldsymbol{\beta}}_{\mathbf{1}}\bar{x} + \hat{\boldsymbol{\beta}}_{\mathbf{1}}x\right) \\
&= \mathbb{V}\text{ar}\left(\bar{\boldsymbol{Y}} + \hat{\boldsymbol{\beta}}_{\mathbf{1}} \cdot [x - \bar{x}]\right) \\
&= \mathbb{V}\text{ar}\left(\bar{\boldsymbol{Y}} + \frac{\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}} \cdot (x - \bar{x})\right) \\
&= \mathbb{V}\text{ar}\left(\frac{\mathrm{S}_{xx} \cdot \bar{\boldsymbol{Y}} + (x - \bar{x})\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i}{\mathrm{S}_{xx}}\right) \\
&= \frac{1}{\mathrm{S}_{xx}^2}\mathbb{V}\text{ar}\left(\mathrm{S}_{xx} \cdot \bar{\boldsymbol{Y}} + (x - \bar{x})\sum_{i=1}^n (x_i - \bar{x})\boldsymbol{Y}_i\right) \qquad (14)
\end{aligned}$$

Expand the terms inside the variance parentheses using the same tricks as before, and let $d = x - \bar{x}$:

$$\begin{aligned}
\mathrm{S}_{xx} \cdot \bar{\boldsymbol{Y}} + (x - \bar{x}) \sum_{i=1}^{n} (x_i - \bar{x}) \boldsymbol{Y}_i &= \mathrm{S}_{xx} \cdot \bar{\boldsymbol{Y}} + \sum_{i=1}^{n} d(x_i - \bar{x}) \boldsymbol{Y}_i \\
&= \frac{\mathrm{S}_{xx}}{n} \sum_{i=1}^{n} \boldsymbol{Y}_i + \sum_{i=1}^{n} d(x_i - \bar{x}) \boldsymbol{Y}_i \\
&= \left[\frac{\mathrm{S}_{xx}}{n} \boldsymbol{Y}_1 + d(x_1 - \bar{x}) \boldsymbol{Y}_1 \right] + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} \boldsymbol{Y}_n + d(x_n - \bar{x}) \boldsymbol{Y}_n \right] \\
&= \left[\frac{\mathrm{S}_{xx}}{n} + d(x_1 - \bar{x}) \right] \cdot \boldsymbol{Y}_1 + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} + d(x_n - \bar{x}) \right] \cdot \boldsymbol{Y}_n
\end{aligned}$$

Put the above expression to equation (14). Note that all $\boldsymbol{Y}_i$ are independent and have a common variance $\sigma_{\boldsymbol{\epsilon}}^2$, we have:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\mu}}_{y|x}\right) &= \frac{1}{\mathrm{S}_{xx}^2} \mathbb{V}\mathrm{ar}\left(\left[\frac{\mathrm{S}_{xx}}{n} + d(x_1 - \bar{x}) \right] \cdot \boldsymbol{Y}_1 + \cdots + \left[\frac{\mathrm{S}_{xx}}{n} + d(x_n - \bar{x}) \right] \cdot \boldsymbol{Y}_n \right) \\
&= \frac{1}{\mathrm{S}_{xx}^2} \cdot \left[\left(\frac{\mathrm{S}_{xx}}{n} + d(x_1 - \bar{x}) \right)^2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_1\right) + \cdots + \left(\frac{\mathrm{S}_{xx}}{n} + d(x_n - \bar{x}) \right)^2 \mathbb{V}\mathrm{ar}\left(\boldsymbol{Y}_n\right) \right] \\
&= \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}^2} \cdot \left[\left(\frac{\mathrm{S}_{xx}}{n} + d(x_1 - \bar{x}) \right)^2 + \cdots + \left(\frac{\mathrm{S}_{xx}}{n} + d(x_n - \bar{x}) \right)^2 \right] \qquad (15)
\end{aligned}$$

Taking the sum inside the square brackets out to simplify:

$$\begin{aligned}
&\left(\frac{\mathrm{S}_{xx}}{n} + d(x_1 - \bar{x}) \right)^2 + \cdots + \left(\frac{\mathrm{S}_{xx}}{n} + d(x_n - \bar{x}) \right)^2 \\
&= n \cdot \frac{\mathrm{S}_{xx}^2}{n^2} + 2 \cdot \frac{\mathrm{S}_{xx}}{n} \cdot d \sum_{i=1}^{n} (x_i - \bar{x}) + d^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \\
&= \frac{\mathrm{S}_{xx}^2}{n} + d^2 \cdot \mathrm{S}_{xx} \qquad (16)
\end{aligned}$$

Put equation (16) back to equation (15), we have:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\hat{\boldsymbol{\mu}}_{y|x}\right) = \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}^2} \cdot \left(\frac{\mathrm{S}_{xx}^2}{n} + d^2 \cdot \mathrm{S}_{xx} \right) = \frac{\sigma_{\boldsymbol{\epsilon}}^2}{n} + \frac{d^2 \cdot \sigma_{\boldsymbol{\epsilon}}^2}{\mathrm{S}_{xx}} \\
= \sigma_{\boldsymbol{\epsilon}}^2 \cdot \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]
\end{aligned}$$

$\square$