Lecture 14 Simple Linear Regression

### Ordinary Least Squares (OLS)

Consider the following simple linear regression model

\[ Y_i = \alpha + \beta X_i + \varepsilon_i \]

where, for each unit \( i \),
- \( Y_i \) is the dependent variable (response).
- \( X_i \) is the independent variable (predictor).
- \( \varepsilon_i \) is the error between the observed \( Y_i \) and what the model predicts.

This model describes the relation between \( X_i \) and \( Y_i \) using an intercept and a slope parameter. The error \( \varepsilon_i = Y_i - \alpha - \beta X_i \) is also known as the residual because it measures the amount of variation in \( Y_i \) not explained by the model.

We saw last class that there exists \( \hat{\alpha} \) and \( \hat{\beta} \) that minimize the sum of \( \varepsilon_i^2 \). Specifically, we wish to find \( \hat{\alpha} \) and \( \hat{\beta} \) such that

\[
\sum_{i=1}^{n} \left( Y_i - (\hat{\alpha} + \hat{\beta} X_i) \right)^2
\]

is the minimum. Using calculus, it turns out that

\[
\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad \quad \hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}
\]

### OLS and Transformation

If we center the predictor, \( \tilde{X}_i = X_i - \bar{X} \), then \( \tilde{\bar{X}} \) has mean zero. Therefore,

\[
\hat{\alpha}^* = \bar{Y} \quad \quad \hat{\beta}^* = \frac{\sum \tilde{X}_i (Y_i - \bar{Y})}{\sum \tilde{X}_i^2}
\]

By horizontally shifting the value of \( X_i \), note that \( \beta^* = \beta \), but the intercept changed to the overall average of \( Y_i \)

Consider the linear transformation \( Z_i = a + bX_i \) with \( \bar{Z} = a + b\bar{X} \). Consider the linear model

\[ Y_i = \alpha^* + \beta^* Z_i + \varepsilon_i \]
Let $\hat{\alpha}$ and $\hat{\beta}$ be the estimates using $X$ as the predictor. Then the intercept and slope using $Z$ as the predictor are

$$
\hat{\beta}^* = \frac{\sum (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum (Z_i - \bar{Z})^2} = \frac{\sum (a + bX_i - a - b\bar{X})(Y_i - \bar{Y})}{\sum (a + bX - a - b\bar{X})^2} = \frac{\sum (bX_i - b\bar{X})(Y_i - \bar{Y})}{\sum (bX - b\bar{X})^2} = \frac{b\sum (X_i - \bar{X})(Y_i - \bar{Y})}{b^2 \sum (X - \bar{X})^2} = \frac{1}{b}\hat{\beta}
$$

and

$$
\hat{\alpha}^* = \bar{Y} - \hat{\beta}^* \bar{Z} = \bar{Y} - \hat{\beta}\bar{X} - a = \bar{Y} - \hat{\beta}\bar{X} - \frac{a\hat{\beta}}{b} = \hat{\alpha} - \frac{a}{b}\hat{\beta}
$$

We can get the same results by substituting $X_i = \frac{Z_i - a}{b}$ into the linear model.

$$
Y_i = \alpha + \beta X_i + \epsilon_i
$$

$$
Y_i = \alpha + \beta \left( \frac{Z_i - a}{b} \right) + \epsilon_i
$$

$$
Y_i = \alpha - \frac{a\beta}{b} + \frac{\beta}{b} Z_i + \epsilon_i
$$

$$
Y_i = \alpha^* + \beta^* Z_i + \epsilon_i
$$

**Properties of OLS**

Given the estimates $\hat{\alpha}$ and $\hat{\beta}$, we can define (1) the estimated predicted value $\hat{Y}_i$ and (2) the estimated residual $\hat{\varepsilon}_i$.

$$
\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i \\
\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i
$$

The least squared estimates have the following properties.

1. $\sum \hat{\varepsilon}_i = 0$

$$
\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i) = \sum_{i=1}^{n} Y_i - n\hat{\alpha} - \hat{\beta} \sum_{i=1}^{n} X_i = n\bar{Y} - n\hat{\alpha} - n\hat{\beta}\bar{X} = n(\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X}) = n(\bar{Y} - (\bar{Y} - \hat{\beta}\bar{X})\bar{X}) = 0
$$
2. $\bar{X}$ and $\bar{Y}$ is always on the fitted line.

$$\hat{\alpha} + \hat{\beta}\bar{X} = (\bar{Y} - \hat{\beta}\bar{X}) + \hat{\beta}\bar{X} = \bar{Y}$$

3. $\hat{\beta} = r_{XY} \times \frac{s_Y}{s_X}$, where $s_Y$ and $s_X$ are the sample standard deviation of $X$ and $Y$, and $r_{XY}$ is the correlation between $X$ and $Y$. Note that the sample correlation is given by:

$$r_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_X s_Y}.$$ 

Therefore,

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \times \frac{n-1}{n-1} \times \frac{s_Y}{s_Y} \times \frac{s_X}{s_X}$$

$$= \frac{\text{Cov}(X,Y)}{s_X^2} \times \frac{s_Y}{s_Y} \times \frac{s_X}{s_X}$$

$$= \frac{\text{Cov}(X,Y)}{s_X s_Y} \times \frac{s_Y}{s_Y} \times \frac{1}{s_X}$$

$$= r_{XY} \times \frac{s_Y}{s_X}$$

4. $\hat{\beta}$ is a weighted sum of $Y_i$.

$$\hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - X)^2}$$

$$= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - X)^2} - \frac{\sum (X_i - \bar{X})\bar{Y}}{\sum (X_i - X)^2}$$

$$= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - X)^2} - \bar{Y} \sum (X_i - \bar{X})$$

$$= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - X)^2} - 0$$

$$= \sum \left( \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})} \right) Y_i$$

$$= \sum w_i Y_i, \quad w_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$$
5. \( \hat{\alpha} \) is a weighted sum of \( Y_i \).
\[
\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} = \sum \frac{1}{n} Y_i - \bar{X} \sum w_i Y_i = \sum v_i Y_i , \quad v_i = \frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum (X_i - \bar{X})^2}
\]

6. \( \sum X_i \hat{\epsilon}_i = 0. \)
\[
\sum X_i \hat{\epsilon}_i = \sum (Y_i - \hat{Y}_i)X_i = \sum (Y_i - \hat{\alpha} - \hat{\beta} X_i)X_i \\
= \sum (Y_i - (\bar{Y} - \hat{\beta} \bar{X}) - \hat{\beta} X_i)X_i \\
= \sum ((Y_i - \bar{Y}) - \hat{\beta} (X_i - \bar{X}))X_i \\
= \sum X_i (Y_i - \bar{Y}) - \hat{\beta} \sum X_i (X_i - \bar{X}) \\
= \beta \sum (X_i - \bar{X})^2 - \beta \sum X_i (X_i - \bar{X}) \\
= \hat{\beta} \sum [(X_i - \bar{X})^2 - X_i (X_i - \bar{X})] \\
= \beta \sum [X_i^2 - 2x_i \bar{X} + \bar{X}^2 - X_i^2 + X_i \bar{X}] \\
= \beta \sum [\bar{X}^2 - X_i \bar{X}] \\
= \beta \bar{X} \sum (X_i - \bar{X}) = 0
\]

Note that \( \hat{\beta} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \)

7. Partition of total variation.
\[
(Y_i - \bar{Y}) = Y_i - \bar{Y} + \hat{Y}_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)
\]
- \( (Y_i - \bar{Y}) \): total deviation of the observed \( Y_i \) from a null model \( (\beta = 0) \).
- \( (\hat{Y} - \bar{Y}) \): deviation of the modeled value \( \hat{Y}_i (\beta \neq 0) \) from a null model \( (\beta = 0) \).
- \( (\hat{Y} - Y_i) \): deviation of the modeled value from the observed value.

Total Variations (Sum of Squares)
- \( \sum (Y_i - \bar{Y})^2 \): total variation: \((n - 1) \times V(Y)\).
- \( \sum (\hat{Y} - \bar{Y})^2 \): variation explained by the model
• $\sum (\hat{Y} - Y_i)^2$: variation not explained by the model; residual sum of squares (SSE), $\sum \hat{\varepsilon}_i^2$

Partition of Total Variation

$$\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$$

Proof:

$$\sum(Y_i - \bar{Y})^2 = \sum\left[(\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)\right]^2$$
$$= \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2 + 2\sum(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)$$
$$= \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$$

where,

$$\sum(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = \sum\left[(\hat{\alpha} + \hat{\beta}X_i) - (\hat{\alpha} + \hat{\beta}\bar{X})\right]\hat{\varepsilon}_i$$
$$= \sum\left[\hat{\beta}(X_i - \bar{X})\right]\hat{\varepsilon}_i$$
$$= \hat{\beta}\sum X_i\hat{\varepsilon}_i - \hat{\beta}\bar{X}\sum\hat{\varepsilon}_i = 0 + 0$$

8. Alternative formula for $\sum \hat{\varepsilon}_i^2$

$$\sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - \bar{Y})^2 - \sum(\hat{Y}_i - \bar{Y})^2$$
$$= \sum(Y_i - \bar{Y})^2 - \sum(\hat{\alpha} + \hat{\beta}X_i - \bar{Y})^2$$
$$= \sum(Y_i - \bar{Y})^2 - \sum(\bar{Y} + \hat{\beta}X\bar{X} + \hat{\beta}X_i - \bar{Y})^2$$
$$= \sum(Y_i - \bar{Y})^2 - \sum\left[\hat{\beta}(X_i - \bar{X})\right]^2$$
$$= \sum(Y_i - \bar{Y})^2 - \hat{\beta}^2\sum(X_i - \bar{X})^2$$

or

$$\sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - \bar{Y})^2 - \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})^2}{\sum(X_i - \bar{X})^2}$$

9. Variation Explained by the Model:
An easy way to calculate $\sum (\hat{Y}_i - \bar{Y})^2$ is

$$
\sum (\hat{Y}_i - \bar{Y})^2 = \sum (Y_i - \bar{Y})^2 - \sum (\hat{Y}_i - \bar{Y})^2
$$

$$
= \sum (Y_i - \bar{Y})^2 - \left[ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - X)^2} \right]^2
$$

$$
= \left[ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - X)^2} \right]^2
$$

10. Coefficient of Determination $r^2$:

$$
r^2 = \frac{\text{variation explained by the model}}{\text{total variation}} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}
$$

(a) $r^2$ is related the residual variance.

$$
\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \bar{Y})^2 - \sum (\hat{Y}_i - \bar{Y})^2
$$

$$
= \left[ \sum (Y_i - \bar{Y})^2 - \sum (\hat{Y}_i - \bar{Y})^2 \right] \times \frac{\sum (Y_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}
$$

$$
= (1 - r^2) \times \sum (Y_i - \bar{Y})^2
$$

Diving the right-hand side by $n - 2$ and the left-hand side by $n - 1$, for large $n$

$$
s^2 \approx (1 - r^2) \text{Var}(Y)
$$

(b) $r^2$ is the square of the sample correlation between $X$ and $Y$.

$$
r^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}
$$

$$
= \frac{\left[ \sum (X_i - \bar{X})(Y_i - \bar{Y}) \right]^2}{\sum (Y_i - \bar{Y})^2 \sum (X_i - X)^2}
$$

$$
= \left[ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_i - X)^2}} \right]^2 = r_{XY}^2
$$

Statistical Inference for OLS Estimates

Parameters $\hat{\alpha}$ and $\hat{\beta}$ can be estimated for any given sample of data. Therefore, we also need to consider their sampling distributions because each sample of $(X_i, Y_i)$ pairs will result in different estimates of $\alpha$ and $\beta$. 

6
Consider the following distribution assumption on the error,

\[ Y_i = \alpha + \beta X_i + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2) \]

The above is now a statistical model that describes the distribution of \( Y_i \) given \( X_i \). Specifically, we assume the observed \( Y_i \) is error-prone but centered around the linear model for each value of \( X_i \).

\[ Y_i \overset{iid}{\sim} N(\alpha + \beta X_i, \sigma^2) \]

The error distribution results in the following assumptions.

1. Homoscedasticity: variance of \( Y_i \) is the same for any \( X_i \).
   \[ V(Y_i) = V(\alpha + \beta X_i + \varepsilon_i) = V(\varepsilon_i) = \sigma^2 \]

2. Linearity: the mean of \( Y_i \) is linear with respect to \( X_i \)
   \[ E(Y_i) = E(\alpha + \beta X_i + \varepsilon_i) = \alpha + \beta X_i + E(\varepsilon_i) = \alpha + \beta X_i \]

3. Independence: \( Y_i \) are independent of each other.
   \[ Cov(Y_i, Y_j) = Cov(\alpha + \beta X_i + \varepsilon_i, \alpha + \beta X_j + \varepsilon_j) = Cov(\varepsilon_i, \varepsilon_j) = 0 \]

We can now investigate the bias and variance of OLS estimators \( \hat{\alpha} \) and \( \hat{\beta} \). We assume that \( X_i \) are known constants.

- \( \hat{\beta} \) is unbiased.

\[
E[\hat{\beta}] = E \left[ \sum w_i Y_i \right], \quad w_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \\
= \sum w_i E[Y_i] \\
= \sum w_i (\alpha + \beta X_i) \\
= \alpha \sum w_i + \beta \sum w_i X_i = \beta
\]
Because

\[
\sum w_i = \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) = 0
\]

\[
\sum w_i X_i = \frac{1}{\sum (X_i - \bar{X})^2} \sum X_i (X_i - \bar{X}) = 0
\]

\[
= \frac{1}{\sum X_i (X_i - \bar{X}) - \sum \bar{X} (X_i - \bar{X})} \sum X_i (X_i - \bar{X})
\]

\[
= \frac{1}{\sum X_i (X_i - \bar{X}) - \bar{X} \sum (X_i - \bar{X})} \sum X_i (X_i - \bar{X})
\]

\[
= \frac{1}{\sum X_i (X_i - \bar{X}) - 0} \sum X_i (X_i - \bar{X}) = 1
\]

- \( \hat{\alpha} \) is unbiased.

\[
E[\hat{\alpha}] = E[\bar{Y} - \hat{\beta}\bar{X}] = E[\bar{Y}] - \bar{X} E[\hat{\beta}]
\]

\[
= E \left[ \frac{1}{n} \sum Y_i \right] - \bar{X} = \frac{1}{n} \sum E [\alpha + \beta X_i] - \bar{X}
\]

\[
= \frac{1}{n} (n\alpha + n\beta X_i) - \bar{X}
\]

\[
= \alpha + \beta \bar{X} - \bar{X} = \alpha
\]

- \( \hat{\beta} \) has variance \( \frac{\sigma^2}{\sum_{i=1}^{n} (X_i - X_1)^2} = \frac{\sigma^2}{(n-1)s^2_x} \), where \( \sigma^2 \) is the residual variance and \( s^2_x \) is the sample variance of \( x_1, \ldots, x_n \).

\[
V[\hat{\beta}] = V \left[ \sum w_i Y_i \right], \quad w_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}
\]

\[
= \sum w_i^2 V [Y_i]
\]

\[
= \sigma^2 \sum w_i^2
\]

where

\[
\sum w_i^2 = \frac{1}{[\sum (X_i - X)^2]^2} \sum (X_i - \bar{X})^2
\]

\[
= \frac{1}{\sum (X_i - \bar{X})^2}
\]
Therefore, the sampling distributions for the regression parameters are:

- $\hat{\alpha}$ has variance $\sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$

$$V[\hat{\alpha}] = V\left[\sum v_i Y_i\right], \quad v_i = \frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum(X_i - X)^2}$$

$$= \sum v_i^2 V[Y_i]$$

$$= \sigma^2 \sum \left(\frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum(X_i - X)^2}\right)^2$$

$$= \sigma^2 \left[\sum \frac{1}{n^2} - \frac{\bar{X}^2}{\sum(X_i - X)^2} - 2 \frac{\bar{X}(X_i - \bar{X})}{n \sum(X_i - X)^2}\right]$$

$$= \sigma^2 \left[\frac{1}{n} - \frac{\bar{X}^2}{\sum(X_i - X)^2} \sum(X_i - \bar{X})^2 - \frac{2\bar{X}}{n \sum(X_i - X)^2} \sum(X_i - \bar{X})\right]$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - X_1)^2}\right)$$

- $\hat{\alpha}$ and $\hat{\beta}$ are negatively correlated.

$$\text{Cov}(\hat{\alpha}, \hat{\beta}) = \text{Cov}\left(\sum v_i Y_i, \sum w_i Y_i\right), \quad v_i = \frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum(X_i - X)^2}, \quad w_i = \frac{X_i - \bar{X}}{\sum(X_i - X)^2}$$

$$= \sum v_i w_i \text{Cov}(Y_i, Y_i)$$

$$= \sum v_i w_i \sigma^2$$

$$= \sigma^2 \sum \left(\frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{\sum(X_i - X)^2}\right) \left(\frac{X_i - \bar{X}}{\sum(X_i - X)^2}\right)$$

$$= \sigma^2 \left[\sum \left(\frac{X_i - \bar{X}}{n \sum(X_i - X)^2}\right) - \sum \frac{\bar{X}(X_i - \bar{X})}{\sum(X_i - X)^2} \frac{X_i - \bar{X}}{\sum(X_i - X)^2}\right]$$

$$= \sigma^2 \left[\frac{\sum(X_i - \bar{X})}{n \sum(X_i - X)^2} - \frac{\sum \bar{X}(X_i - \bar{X})^2}{\sum(X_i - X)^2}\right]$$

$$= \sigma^2 \left[0 - \frac{\bar{X} \sum(X_i - \bar{X})}{\sum(X_i - X)^2}\right]$$

$$= -\frac{\sigma^2 \bar{X}}{\sum(X_i - X)^2}$$

Therefore, the sampling distributions for the regression parameters are:

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - X_1)^2}\right]\right) \quad \hat{\beta} \sim N\left(\beta, \sigma^2 \left[\frac{1}{n^2} + \frac{2\bar{X}}{n \sum(X_i - X)^2}\right]\right)$$
However, in most cases, we don’t know what $\sigma^2$ and will need to estimate it from the data. One unbiased estimator for $\sigma^2$ is obtained from the residuals.

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

By replacing $\sigma^2$ with $s^2$, we have the following sampling distributions for $\hat{\alpha}$ and $\hat{\beta}$.

$$\hat{\alpha} \sim t_{n-2} \left( \alpha, s^2 \left[ \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - X_1)^2} \right] \right)$$

$$\hat{\beta} \sim t_{n-2} \left( \beta, \frac{s^2}{\sum_{i=1}^{n} (X_i - X_1)^2} \right)$$

Note that our degrees of freedom is $n-2$ because we estimated two parameters ($\alpha$ and $\beta$) in order to calculate the residuals. Similarly recall that $\sum (Y_i - Y) = 0$.

Knowing that $\hat{\alpha}$ and $\hat{\beta}$ follow a t-distribution with degrees of freedom $n-2$, we are able to construct confidence interval and carry out hypothesis test as before. For example, a 90% confidence interval for $\beta$ is given by

$$\hat{\beta} \pm t_{0.05, n-2} \times \sqrt{\frac{s^2}{\sum_{i=1}^{n} (X_i - X_1)^2}} .$$

Similarly, consider testing the hypotheses $H_0: \beta = \beta_0$ versus $H_A: \beta \neq \beta_0$. The corresponding t-test statistic is given by

$$\frac{\hat{\beta} - \beta_0}{s^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} (X_i - X_1)^2} \right]} \sim t_{n-2} .$$

**Prediction**

Based on the sampling distribution of $\hat{\alpha}$ and $\hat{\beta}$ we can also investigate the uncertainty associated with $\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0$ on the fitted line for any $X_0$. Specifically,

$$\hat{Y}_0 \sim N \left( \alpha + X_0 \beta, \sigma^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} (X_i - X)^2} \right] \right).$$

Note that $\alpha + X_0 \beta$ is the mean of $Y$ given $X_0$. The textbook uses the notation $\mu_0$. Replacing the unknown with our estimates, we have

$$\hat{Y}_0 \sim t_{n-2} \left( \hat{\alpha} + X_0 \hat{\beta}, \frac{s^2}{n} \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} (X_i - X)^2} \right] \right).$$

- The fitted value $\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0$ has mean $\alpha + X_0 \beta$ and variance $\sigma^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} (X_i - X)^2} \right]$. 

$$E(\hat{Y}_0) = E(\hat{\alpha} + \hat{\beta}X_0) = E(\hat{\alpha}) + X_0 E(\hat{\beta}) = \alpha + X_0 \beta$$
\[ V(Y_0) = V(\hat{\alpha} + \hat{\beta}X_0) = V(\hat{\alpha}) + X_0^2V(\hat{\beta}) + X_0\text{Cov}(\hat{\alpha}, \hat{\beta}) \]
\[ = \sigma^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right) + \sigma^2 \frac{X_0^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} - 2\sigma^2 \frac{X_0\bar{X}}{\sum(X_i - \bar{X})^2} \]
\[ = \sigma^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right] \]

We can also consider the uncertainty of a new observation \( Y_0^* \) at \( X_0 \). In other words, what values of \( Y_0^* \) do I expect to see at \( X_0^* \)? Note that this is different from \( \hat{Y}_0 \) which is the fitted value that represents the mean of \( Y \) at \( X_0 \). Since

\[ Y_0^* = \alpha + \beta X_0 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2), \]

by replacing the unknown parameters with our estimates, we obtain

\[ Y_0^* \sim t_{n-2} \left( \hat{\alpha} + \hat{\beta}X_0, \; s^2 + s^2 \left[ \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2} \right] \right). \]

Again, the above distribution describes the actual observable data, instead of model parameters; interval estimates are also known as *prediction intervals*, instead of confidence intervals.