In this section, we will develop methods for finding confidence intervals and for performing hypothesis tests about the slope and the intercept of the regression line.

Assumptions for the Simple Linear Regression Model

Throughout this section we shall make the following assumptions:

1. Y is related to X by the simple linear regression model
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1, \ldots, n,$$
i.e.,
$$E(Y \mid X = x) = \beta_0 + \beta_1 x.$$
2. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent of each other.
3. The errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ have a common variance σ^2.
4. The errors are normally distributed with mean 0 and variance σ^2, that is,
$$\varepsilon_i \mid X \sim N(0, \sigma^2).$$

Methods for checking these four assumptions will be considered in a later chapter. In addition, since the regression model is conditional on X, we assume that the values of the predictor variable $x_1, x_2, \ldots, x_n$ are known fixed constants (as they should be once the data are collected).

Inferences About the Slope of the Regression Line

Recall from Section 2.1 that the least squares estimate of β_1 is given by
$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Under the regression assumptions listed above,
$$E(\hat{\beta}_1 \mid X) = \beta_1, \qquad \mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma^2}{S_{XX}},$$
and therefore the distribution of $\hat{\beta}_1$ is:
$$\hat{\beta}_1 \mid X \sim N\left(\beta_1,\ \frac{\sigma^2}{S_{XX}}\right).$$

Standardization and Test Statistic

Since σ^2 is unknown, we replace it with the estimator $\hat{\sigma}^2$. The estimated standard error of $\hat{\beta}_1$ is
$$\mathrm{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{XX}}},$$
and the corresponding test statistic is
$$T = \frac{\hat{\beta}_1 - \beta_1}{\mathrm{se}(\hat{\beta}_1)} \sim t_{n-2}.$$

Hypothesis Test for the Slope

To test the null hypothesis

$$H_0 : \beta_1 = \beta_1^0,$$

the test statistic is

$$T = \frac{\hat{\beta}_1 - \beta_1^0}{\mathrm{se}(\hat{\beta}_1)},$$

which under H_0, follows a t-distribution with $n-2$ degrees of freedom.

– Note: When the test of whether or not $\beta_1 = 0$ leads to the conclusion that $\beta_1 \neq 0$, the association between Y and X is sometimes described to be a linear association.

Confidence Interval for the Slope

A $100(1-\alpha)\%$ confidence interval for β_1, the slope of the regression line, is given by

$$\hat{\beta}_1 \ \pm \ t_{\alpha/2,\, n-2} \cdot \mathrm{se}(\hat{\beta}_1).$$

Example 1

Continued with the Production data.

a. Find a 95% confidence interval for β_1.

With 95% confidence, the average production time is estimated to increase by between __________ and __________ minutes for each additional unit in the order.

b. Find the p-value for testing $H_0 : \beta_1 = 0$ v.s. $H_a : \beta_1 \neq 0$. Draw a conclusion based on the p-value with $\alpha = 0.05$.

c. Test whether the average production time increases by more than 15 seconds for each additional unit in the order with $\alpha = 0.05$. Report the p-value and draw a conclusion.

Example 2

Continued with the FreshmanGPA data.

a. Find a 95% confidence interval for β_1. Interpret the interval in context of the data.

b. Find the p-value for testing $H_0 : \beta_1 = 0$ v.s. $H_a : \beta_1 \neq 0$. Draw a conclusion based on the p-value with $\alpha = 0.01$.

c. Test whether the average GPA increases by more than 0.1 for each additional 5-point increase in the ACT score. Report the p-value and draw a conclusion with $\alpha = 0.05$.

Inferences About the Intercept

Recall from Section 2.1 that the least squares estimate of β_0 is

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Under the regression assumptions, we know that

$$E(\hat{\beta}_0 \mid X) = \beta_0, \qquad \mathrm{Var}(\hat{\beta}_0 \mid X) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}} \right),$$

and therefore the distribution of $\hat{beta}_0$ is:

$$\hat{\beta}_0 \mid X \sim N\Bigg(\beta_0,\ \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}} \right)\Bigg).$$

Standardization and Test Statistic

Standardizing of $\hat{\beta}_0$ gives

$$Z = \frac{\hat{\beta}_0 - \beta_0}{\sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}}} \ \sim\ N(0,1).$$

Since σ is unknown, we replace it with the estimator $\hat{\sigma}$. The estimated standard error of $\hat{\beta}_0$ is

$$\mathrm{se}(\hat{\beta}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}}.$$

The corresponding test statistic is

$$T = \frac{\hat{\beta}_0 - \beta_0}{\mathrm{se}(\hat{\beta}_0)} \sim t_{n-2}.$$

The procedure for conducting a hypothesis test and constructing a confidence interval for the intercept follows the same reasoning as that used for the slope.

– Note: the interpretation of the intercept is only appropriate when the scope of the model is valid to be extended to $X = 0$.

Example 3

Continued with the Production data.

a. Find a 95% confidence interval for β_0.

With 95% confidence, the average production setup time is estimated to be between __________ and __________ minutes.

b. Find the p-value for testing $H_0 : \beta_0 = 0$ v.s. $H_a : \beta_0 \neq 0$. Draw a conclusion based on the p-value.

Example 4

Continued with the FreshmanGPA data. Find and interpret a 95% confidence interval for β_0.