2.3 Confidence Intervals for the Population Regression Line

A common objective in regression analysis is to estimate the mean for one or more probability distributions of Y. Consider, for example, a study of the relation between level of work pay (X) and worker productivity (Y). The **average** productivity at high and medium levels of piecework pay may be of particular interest for purposes of analyzing the benefits obtained from an increase in the pay.

Prediction at $X = x^*$

First, recall that the population regression line at $X = x^*$ is given by

$$E(Y \mid X = x^*) = \beta_0 + \beta_1 x^*.$$

An estimator of this unknown quantity is the value of the estimated regression equation at $X = x^*$, namely,

$$\hat{y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*.$$

Under the assumptions stated in the previous section, it can be shown that

$$E(\hat{y}^*) = E(\hat{y} \mid X = x^*) = \beta_0 + \beta_1 x^*.$$

The variance of $\hat{y}^*$ is

$$\mathrm{Var}(\hat{y}^*) = \mathrm{Var}(\hat{y} \mid X = x^*) = \sigma^2 \left(\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{XX}} \right),$$

where $S_{XX} = \sum (x_i - \bar{x})^2$.

Therefore,

$$\hat{y}^* \mid X = x^* \;\sim\; N\Big(\beta_0 + \beta_1 x^*,\; \sigma^2 \left(\tfrac{1}{n} + \tfrac{(x^* - \bar{x})^2}{S_{XX}} \right)\Big).$$

Standardization and Test Statistic

Replacing σ by $\hat{\sigma}$ results in the t distribution:

$$T = \frac{\hat{y}^* - (\beta_0 + \beta_1 x^*)}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{XX}}}} \;\sim\; t_{n-2}.$$

Confidence Interval for the Regression Line

Thus, a $100(1-\alpha)\%$ confidence interval for

$$E(Y \mid X = x^*) = \beta_0 + \beta_1 x^*$$

is given by

$$\hat{y}^* \;\pm\; t_{\alpha/2,\, n-2}\, \hat{\sigma}\sqrt{\tfrac{1}{n} + \tfrac{(x^* - \bar{x})^2}{S_{XX}}}.$$

Example 1

Continued with the Production data. Find a 95% confidence interval for the average run time for orders that have 200 items produced.

Example 2

Continued with the FreshmanGPA data. Find a 95% confidence interval for the average GPA at the end of the freshman year for students who scored 30 on the ACT.