4.6 Confidence Interval for the Slope of the Population Regression Line

Example of Finding a Confidence Interval For the Slope: Calculate a 95% confidence interval for the slope of the regression line for the relationship between Effort Index and Chemistry Performance.

At $df = n - 2 = 6 - 2 = 4,\ t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.776$

$SE(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{S_{xx}}} = \frac{5.8990}{\sqrt{14.8333}} = 1.5316$

$\hat{\beta}_1 \pm t_{\alpha/2} \times SE(\hat{\beta}_1) = 9.4719 \pm 2.776 \times 1.5316 = (5.22,\ 13.72)$

Interpretation: We can be 95% confident that the mean performance in Chemistry increases by between 5.22 and 13.72 per unit increase in Effort index.

Note: The confidence interval for the slope does not contain 0, which agrees with the hypothesis test perfromed above which concluded that the slope is not 0.

4.7 Confidence Intervals for Estimation of Mean Response and Predicted Response.

Example of Finding Mean Repsonse and Predicted Response:

> Find a 95% confidence interval for the mean height of all 3-year-old trees.
> Also, find a 95% prediction interval for the height of a 3-year-old tree.

At $df = n - 2 = 5,\ t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.571$

Calculate the point estimate: $\hat{y} = \hat{\beta}_1 + \hat{\beta}_1 x^* = -0.08087 + 0.66809 \times 3 = 1.923\ m$
So the mean response of a 3-year-old tree would be 1.923 m (on average).

Confidence interval for the mean response:
$\hat{y} \pm t_{\alpha/2} \times \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = 1.923 \pm 2.571 \times 0.39367\sqrt{\frac{1}{7} + \frac{(3-3.28571)^2}{13.42857}} = 1.923 \pm 0.3906 = (1.532, 2.314)\ m$

Interpretation: We can be 95% confident that the mean height of 3-year-old trees is between 1.532 m and 2.314 m.

Confidence interval for the predicted response of a single tree that is 3 years old (prediction interval):
$\hat{y} \pm t_{\alpha/2} \times \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = 1.923 \pm 2.571 \times 0.39367\sqrt{1 + \frac{1}{7} + \frac{(3-3.28571)^2}{13.42857}} = 1.923 \pm 1.0849 = (0.838, 3.008)\ m$

Interpretation: We can be 95% confident that the height of a 3-year-old tree is between 0.838 and 3.008 m.

Note: The prediction interval is always wider than the confidence interval because the estimate of a mean is always close to the population mean, whereas variation in all observed values is more disperse.

> Calculate $S_{xx} = (n-1)s_x^2 = (7-1)(1.496026)^2 = 13.42857$

4.8 Hypothesis Test for Linear Correlation

> *P-value for F-test = P-value for two-tailed t-test = P-value for two-tailed test for correlation (r)*

Example of Linear Correlation Hypothesis Test: At the 5% significance level, test whether there was a correlation between performance in Chemistry and Effort index (based on a random sample of 6 students).

$H_0 : \rho = 0$ (there is no correlation between performance in Chemistry and Effort index)
$H_a : \rho \neq 0$ (there is a correlation between performance and effort)

$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{140.5}{\sqrt{14.8333 \times 1470}} = 0.9515$

$df = n - 2 = 6 - 2 = 4$
P-value: $0.002 < P < 0.005$ (Exact P-value: 0.003474)
$P < \alpha$, reject H_0 with very strong evidence.

At the 5% significance level, there is sufficient evidence to conclude that there is a correlation between Effort Index and Chemistry performance.

Example on Biotechnology: Using the water fern Azolla to produce Hydrogen Fuel [Example Combining All Concepts]

(a) According to the regression model what would you predict to be the rate of the hydrogen production at 25% atmospheric nitrogen? Would this be a reliable estimate?

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 750.306 + (-9946) \times 25 = 501.656$

The predicted rate of H_2 production at 25% atmospheric N_2 is 501.656 nmol. This would be a reliable estimate since 25% is within the observed range of x (interpolation)

(b) What was the residual (error) of this regression model at an atmospheric N2 level of 60%?

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 750.306 + (-9946) \times 60 = 153.546$

$Residual = error = y_i - \hat{y} = observed - predicted = 66 - 153.546 = -87.546$