5.4. Confidence Interval and Prediction Interval for the Response Variable

Example (Price of Orions against age and miles driven):

> Find a 95% confidence for the mean price of all Orions that are 5 years old and have been driven 52,000 miles

At $df = n - (p+1) = 11 - (2+1) = 8,\ t_{\alpha/2} = t_{0.05/2} = t_{0.025} - 2.306$
$Price = 183.035 - 9.504 \times 5 - 0.821 \times 52 = 92.80 \text{ (hundreds of dollars)}$

$\hat{y_p} = t_{\alpha/2} \times SE(Fit) = 92.80 \pm 2.306 \times 2.74 = 92.80 \pm 6.32 = (86.48, 99.12)$

We can be 95% confident that the mean price of all Orions that are 5 years old and have been driven 52000 miles is between $8648 and $9912.

> Calculate a 95% prediction interval for the price of an Orion (any single observation) that is 5 years old and has been driven 52,000 miles

$\hat{y_p} \pm t_{\alpha/2} \times \sqrt{\hat{\sigma}^2 + [SE(Fit)]^2} = 92.80 \pm 2.306 \times \sqrt{8.805^2 + 2.74^2} = (71.54, 114.06)$

We can be 95% confident that the price of an Orion (any single observation) that is 5 years old and has been driven for 52,000 miles is between $7,154 and $11,406.

5.6. Interaction Models in Multiple Regression

Effect of BMI and Salt Intake (and their Interaction) on Systolic Blood Pressure:

(a) At the 5% significance level, perform a hypothesis test to determine whether the overall multiple regression model is significant or useful for making predictions about systolic blood pressure (SBP). Perform ALL steps of the hypothesis test.

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
H_a : At least one β_i is not 0

$F = \frac{SS_{REGR}/p}{SS_{Error}/(n-(p+1))} = \frac{MS_{REGR}}{MS_{Error}} = \frac{1330.00/3}{3.428/(14-(3+1))} = 1293.271$

df = (3, 10), P < 0.001, There is extremely strong evidence against null hypothesis.
$P < \alpha$ (0.05), reject H_0.

(b) At the 5% significance level, perform the most appropriate test to determine whether there is a positive relationship between salt intake and systolic blood pressure.

$H_0 : \beta_2 = 0$
$H_a : \beta_2 > 0$

$t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \frac{0.892}{0.350} = 2.5485$

At df = n - (p + 1) = 10, 0.01 < P < 0.02, very strong evidence against H_0.
$P < \alpha$ (0.05), reject H_0
Note: the exact P = 0.029/2 = 0.0145

At the 5% significance level, the data provide sufficient evidence to conclude that there is a significant positive relationship between salt intake and systolic blood pressure.

(c) Calculate a 95% confidence interval for the slope of the interaction term (representing interaction between BMI and sodium intake). Using this confidence interval, what conclusion can you make about the possible interaction between body mass index and sodium intake in their effect on systolic blood pressure? Explain your answer.

At $df = 10,\ t_{\alpha/2} = t_{0.025} = 2.228$

$\hat{\beta}_3 \pm t_{\alpha/2} \times SE(\hat{\beta}_3) = 0.015 \pm 2.228 \times 0.006 = (0.00163, 0.02837)$

Since 0 is not inside this interval we can be 95% confident that the slope of the interaction term is significant, that is, there is significant interaction between body mass index and salth intake in their effect on systolic blodd pressure.

(e) Find the standard error of the model (standard error of the estimate of the model)?

$MS_{Error} = \frac{SS_{Error}}{n-p-1} = \frac{3.428}{14-3-1} = 0.3428$

$SE(model) = \hat{\sigma} = \sqrt{MSE} = \sqrt{0.3428} = 0.585491$

(f) What percentage of the variation in systolic blood pressure is explained by (or accounted for by) the regression model? (Note: Determine the adjusted percentage.)

$R^2_{adj} = 1 - \frac{MS_{Error}}{MS_{Total}} = 1 - \frac{(n-1)SS_{Error}}{(n-1)SS_{Total}} = 1 - 0.0333 = 0.9666 = 96.66\%$

(g) Suppose that a person with a body mass index of 40 kg/m2 and daily sodium intake of 42 (in 100s of mg/day) had an observed systolic blood pressure reading of 163 mm Hg. What was the residual or error of this observation?

$\hat{y} = 108.726 - 0.218 \times 40 + 0.892 \times 163 + 0.15 \times 40 \times 42 = 162.67$
$Residual = Error = Observed - Predicted = 163 - 162.67 = 0.33$