5. MULTIPLE REGRESSION ANALYSIS

5.2. Inferences Concerning the Overall Usefulness of the Multiple Regression Model

Example: Effect of age and miles driven on the price of Orion cars:

(a)
The purpose is to perfrom a hypothesis test for the usefulness of the overall regresson model where there is more than one predictor variable. This can only be done with multiple regression ANOVA.

$H_0 : \beta_1 = \beta_2 = 0$ (The overall model is not useful)
$H_a : \text{At least one of the } \beta \text{ is not equal } 0$ (The overall model is useful).

$n = 11, \; p = 2$
$SS_{Error} = SS_{Total} - SS_{REGR} = 9708.545 - 9088.314 = 630.232$

$F = \frac{SS_{REGR}/p}{SS_{Error}/(n-p-1)} = \frac{MS_{REGR}}{MS_{Error}} = \frac{9088.314/2}{630.232/(11-2-1)} = 58.612$

$df = (p, n-p-1) = (2, 8)$
P-value: $P < 0.001$, extremely strong evidence against H_0.
$P < \alpha$ (0.05), therefore reject H_0.

At the 5% significance level, the data provide sufficient evidence to conclude that the overall multiple regression model is useful for making predictions, that is, the variables age and/or miles driven, taken together, are useful for predicting price.

b. What percentage of the variation in Orion price is explained by the regression model? Determine the unadjusted percentage.

$R^2 = \frac{SS_{REGR}}{SS_{Total}} = \frac{9088.314}{9708.545} = 0.93611$

c. What percentage of the variation in Orion price is explained by the regression model? Determine the adjusted percentage and compare it with the unadjusted percentage calculated in part (b).

$MS_{Total} = \frac{SS_{Total}}{n-1} = \frac{9708.545}{11-1} = 870.8545$

$R^2_{Adj} = 1 - \frac{MS_{Error}}{MS_{Total}} = 1 - \frac{77.529}{870.8545} = 0.920$

Thus, 92% of the variation (adjusted) in price is explained by the regression model.

Comparison: Since the number of coefficients (3) is not very close to the sample size (11), there is not a very large difference between these two coefficients. Also, since R is very close to 1, the adjustment doesn't make much difference. However, R^2_{adj} is slightly more accurate. So, we conclude that 92% of the variation in the observed values of the response variable is accounted for by the regression analysis model of price against age and miles driven.

5.3 Inferences Concerning the Usefulness of Particular Predictor Variables: The Multiple Regression t-test and Confidence Interval for Particular

Example (Orion Prices):

a. At the 5% significance level, test whether the data provide sufficient evidence to conclude that the number of miles driven, in conjunction with age, is useful for predicting price.

Regression equation: $\hat{y} = 183.035 - 9.504x_1 - 0.821x_2$

$H_0 : \beta_2 = 0$ (Miles driven is not useful for making predictions)
$H_a : \beta_2 \neq 0$ (Miles driven is useful for making predictions)

$t = \frac{\beta_2}{SE(\hat{\beta}_2)} = \frac{-0.821}{0.255} = -3.219$

$df = n - p - 1 = 11 - 2 - 1 = 8$
P-value: $(0.005 < P < 0.01) \times 2 \Rightarrow 0.01 < P < 0.02$, there is strong evidence against H_0.
$P < \alpha$ (0.05), therefore reject H_0.

At the 5% significance level, the data provide sufficient evidence to conclude that, in conjunction with age, the number of miles driven is useful for predicting price.

b. Calculate a 95% confidence interval for the partial slope for miles driven.

At $df = n - p - 1 = 8$, for a 95% CI: $t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.306$

$\hat{\beta}_2 \pm t_{\alpha/2} \times SE(\hat{\beta}_2) = -0.821 \pm 2.306 \times 0.255 = -0.821 \pm 0.588 = (-1.409, -0.233)$

Conclusion: We can be 95% confident that the partial slope for miles driven is between -1.409 and -0.233 (reduction in price per miles driven).