Assignment #4
Crystal Mosley

Model 1—

How many observations are in the sample data?

To find the total observations in the data sample, we look at the ANOVA Corrected Total and also understand that one Degree of Freedom from the Parameter Estimates is devoted towards estimating the intercept. This means there are 72 observations.

Write out the null and alternate hypotheses for the t-test for Beta1.

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

Testing the significance of regression—the hypothesis for β_1 is meant to be an inference of whether or not there is a linear relationship between the dependent and independent variable.

Compute the t- statistic for Beta1.

Equation for t-stat: $t_0 = \frac{\beta\ carat\ 1}{se(\beta\ carat\ 1)}$

Note: t_i = [DF = n – dim(Model)]

$$t_0 = \frac{2.18604}{0.41043} = 5.3262$$

Compute the R-Squared value for Model 1.

Equation for r-squared value: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

- SST (Total Sum of Squares) – total variation in the sample
- SSR (Regression Sum of Squares) – variation in the sample that has been explained by the regression model
- SSE (Error Sum of Squares) – variation in the sample that cannot be explained

$$R^2 = 1 - \frac{630.3595}{2756.3686} = 0.7713$$

Compute the Adjusted R-Squared value for Model 1.

Equation for adjusted r-squared value: $R^2_{ADJ} = 1 - \frac{\frac{SSE}{(n-k-1)}}{\frac{SST}{(n-1)}} = 1 - \frac{\frac{SSE}{(n-p)}}{\frac{SST}{(n-1)}}$

- k = number of predictor variables included in the regression model
- p = total number of parameters in the model

- $p = k + 1$ if the model includes an intercept term
- $p = k$ if model does *not* include an intercept term

$$R^2_{ADJ} = 1 - \left[\frac{\frac{630.3595}{(72-5)}}{\frac{2756.3686}{(72-1)}}\right] = 0.7577$$

Write out the null and alternate hypotheses for the Overall F-test.

$$H_0 : \beta_1 = \ldots = \beta_k = 0 \text{ vs. } H_1 : \beta_i \neq 0$$

Testing the significance of regression—this equation determines if there is a linear relationship between the dependent variable and any of the regressing variables.

Compute the F-statistic for the Overall F-test.

Equation for f-stat of overall f-test: $F_0 = \frac{\frac{SSR}{k}}{\frac{SSE}{(n-p)}}$

$$F_0 = \frac{\frac{2126.0090}{4}}{\frac{630.3595}{(72-5)}} = 56.4926$$

Model 2—

Model 1 and Model 2 are a pair of models. Does Model 1 nest Model 2 or does Model 2 nest Model 1? Explain.

$$Y_{\text{Model 1}} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4$$
$$Y_{\text{Model 2}} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6$$

Since the predictor variables in Model 1 are a subset of Model 2, then Model 1 is nested by Model 2.

Write out the null and alternate hypotheses for a nested F-test using Model 1 and Model 2.

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_6 = 0 \text{ vs. } H_1 : \beta_i \neq 0$$

Compute the F-statistic for a nested F-test using Model 1 and Model 2.

Equation for f-stat of nested f-test: $F_0 = \frac{\frac{[SSE\,(RM) - SSE\,(FM)]}{(\dim(FM) - \dim(RM))}}{\frac{(SSE\,(FM))}{[n - \dim(FM)]}}$

- FM = full model
- RM = reduced model

$$F_0 = \frac{\frac{(630.3595-572.6091)}{(7-5)}}{\frac{(572.6091)}{(72-7)}} = 3.2777$$

Additional Questions—

Compute the AIC values for both Model 1 and Model 2.

Equation for Akaike Information Criterion: $AIC = n \text{ x } ln(\frac{SSE}{n}) + 2p$

Model 1 → $AIC = 72 \text{ x } ln(\frac{630.3595}{72}) + 2(5) = 166.2129$

Model 2 → $AIC = 72 \text{ x } ln(\frac{572.6091}{72}) + 2(7) = 163.2946$

Compute the BIC values for both Model 1 and Model 2.

Equation for Bayesian Analogues: $BIC = n \text{ x } ln(\frac{SSE}{n}) + p \text{ x } ln(n)$

Model 1 → $BIC = 72 \text{ x } ln(\frac{630.3595}{72}) + 5 \text{ x } ln(72) = 177.5962$

Model 2 → $BIC = 72 \text{ x } ln(\frac{572.6091}{72}) + 7 \text{ x } ln(72) = 179.2313$

Compute the Mallow's Cp values for both Model 1 and Model 2.

Equation for Mallow's C_p value: $C_p = \frac{SSRp}{MSE} - n + 2 \text{ x } p$

Model 1 → $C_p = \frac{630.3595}{9.4084} - 72 + 2 \text{ x } 5 = 5.0000$

Model 2 → $C_p = \frac{572.6091}{8.8094} - 72 + 2 \text{ x } 7 = 7.0000$

Verify the t-statistics for the remaining coefficients in Model 1.

Equation for t-stat: $t_0 = \frac{\beta\ carat\ 1}{se(\beta\ carat\ 1)}$

Parameter: Intercept

$$t_{intercept} = \frac{11.3303}{1.9941} = 5.6819$$

Parameter: X2

$$t_{x2} = \frac{8.2743}{2.3391} = 3.5374$$

Parameter: X3

$$t_{x3} = \frac{0.4918}{0.2647} = 1.858$$

Parameter: X4

$$t_{x4} = \frac{-0.4936}{2.2943} = -0.2151$$

Verify the Mean Square values for Model 1 and Model 2.

Equation for mean square: $Mean_x = \frac{SSR}{DF}$

Model 1 → $Y_{\text{Model 1}} = \frac{2126.0090}{4} = 531.5023$

Model 2 → $Y_{\text{Model 2}} = \frac{2183.7595}{6} = 363.9599$