ISyE 6414 Regression Analysis

Homework 5

Fall 2022

A regional express delivery company conducted a study to investigate the relationship between the cost of shipment, y (in dollars), and the variables that control the shipping charge: package weight, x_1 (in pounds), and distance shipped, x_2 (in miles). By using the attached comma-deliminated data file `HW5ShipmentData.csv` and $\alpha = 0.05$, answer Questions 1–7. Submit all relevant outputs and your R codes (if you used R).

1. Produce the matrix plot, and interpret the possible relations among all variables.

2. Solve the first-order model, i.e., $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. Also, produce the ANOVA table. Is the model useful (significant) as a whole, i.e., apply an F-test. Are the predictors significant? You can perform the hypothesis tests by considering p-values only.

3. Calculate R^2 by using appropriate ANOVA table quantities. Justify your result by referring to the R^2 quantity found in your output.

4. Check the random error assumptions for the first-order model, in particular, check whether $\mathsf{E}(\epsilon) = 0$ or not, the normality, and the identical distribution (variance) assumptions. What are your conclusions?

5. Find a 95% *confidence* interval for the mean response when the predictors Weight = 6 and Distance = 150.

6. Find a 95% *prediction* interval for a single future observation when the predictors Weight = 6 and Distance = 150.

7. Now consider the following full second-order model for the same problem, i.e., $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$. Is this full model better than the reduced model in Question 2? Answer by performing the partial F-test.

A consulting firm markets a computerized system for monitoring road construction bids to various state departments of transportation. The firm wants to compare the mean annual maintenance costs accrued by the system users in three different states: Kansas, Kentucky, and Texas. Answer Questions 8–10 by using the attached `HW5StateCostData.csv` and $\alpha = 0.05$.

8. Solve the linear regression model to study whether the qualitative variable "State" with three categories is significant or not? What is your base level?

9. Write down the model and the dummy variables in the model. What does β_0 correspond to in terms of expected costs?

10. By using the hypothesis in Question 8, discuss whether the expected costs in each state are identical or not.