ISyE 403 Regression and Forecasting
Homework 5
Due March 4, 2016

Solve the following problems by using Minitab or R, and submit the relevant parts of the outputs.

Refer to the *Shipment* data for the problems 1-3.

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).

1. Produce the matrix plot (or all scatter plots), and interpret the possible relations among all variables.

2. First-order model.

a. Solve the first-order model, i.e., $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. By setting $\alpha = 0.05$, is the model appropriate? Are the predictors significant? You can perform hypothesis tests by considering p values only. Submit the output.

b. Consider the shipment of a package which weighs 6 pounds, and the distance is 100 miles. Estimate the cost of shipment, $\hat{y}$, and corresponding 95% confidence interval and 95% prediction interval. Submit the output.

c. Check the random error assumptions, in particular, check whether $E(\varepsilon) = 0$ or not, the normality and identical distribution (variance) assumptions. Submit the output. Note that the independent distribution assumption is not a concern for this problem.

3. Now consider the following full second-order model for the same problem, i.,e.,
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \varepsilon.$$

a. Solve the model. By setting $\alpha = 0.05$, is the model appropriate? Are the predictors significant? You can perform hypothesis tests by considering p values. Submit the output.

b. Remove the nonsignificant variable(s) by maintaining hierarchy. Re-run the model. Submit the output.

c. Consider the shipment of a package which weighs 6 pounds, and the distance is 100 miles. Estimate the cost of shipment, $\hat{y}$, and corresponding 95% confidence interval and 95% prediction interval. Submit the output. Did you get the same results as in problem 2?

d. Is this model better than the model in question 2? Answer by conducting nested (partial) F test. Also compare the adjusted R^2 and *MSE* (or s).

4. 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. By using the attached StateCost data, solve the model to study whether the qualitative variable State with three categories is significant or not? What is your base level? Discuss whether the expected costs in each state are identical or not. Use $\alpha = 0.05$. Submit the output.