

# ISyE 6414 Regression Analysis

## Homework 6

Fall 2022

*For all of the problems in this assignment, please submit the relevant outputs and your R codes (if you used R). In addition, unless otherwise indicated, assume that  $\alpha = 0.05$ .*

Consider the regional express delivery company problem that you studied in HW5. To remind you, we had defined: 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). We used the data set `HW5ShipmentData.csv`. Answer Questions 1–4.

1. Solve the complete model:  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1 x_2 + \epsilon$ . Which predictors are significant? You can perform the hypothesis tests by considering  $p$ -values only.
2. What is the effect of a one-mile increase in distance on the cost of shipment when the weight is held constant at 5 pounds?
3. Check the random error assumptions for the complete model, in particular, check whether  $E(\epsilon) = 0$  or not, the normality, and the identical distribution (variance) assumptions. What is your conclusion?
4. You checked random error assumptions for the reduced model,  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ , in HW5. What would you conclude when you compare the error assumptions of the reduced model and the error assumptions of the complete model? You can refer to the HW5 solutions that we posted.

\*\*\*\*\*

5. After running a linear regression model,  $y \sim x_1 + x_2 + x_3$  with a sample of 24 observations, the model adequacy was investigated, and the Durbin–Watson test statistic was found to be 0.829. Check the independence assumption of the errors by applying the Durbin–Watson test (test for both positive correlation and negative correlation). Assume level  $\alpha = 0.10$ .

Consider the hospital stay problem that you studied in HW4 by using `homework04Hospital.csv` data. To remind you, we had defined:  $y$  = monthly labor hours required;  $x_1$  = monthly X-ray exposures;  $x_2$  = monthly occupied bed days; and  $x_3$  = average length of patients' stays (in days). Answer Questions 6–8.

6. Do you identify potentially unusual observations? Answer by producing and listing standardized residual and Cook's distance measure for each observation.

7. Consider the solution you obtained for this problem (you can refer to the HW4 solutions that we posted). Can you identify counterintuitive results in the solutions by considering the estimated coefficients and their standard errors? Why is that?
8. Produce the variance inflation factor (VIF) for each predicting variable and comment on the results.