



DEPARTMENT OF STATISTICS AND DATA
SCIENCE

Jahangirnagar University
Take-home Exam-2024
Subject: Applied Regression

Course No. 307
Time: 15 days
Full Marks: 20

1. **Grocery Retailer:** A large, national grocery retailer tracks productivity and costs of its facilities closely. Data saved in **GroceryRetailer.csv** file were obtained from a single distribution center for a one-year period. Each data point for each variable represents one week of activity. The variables included are:

- The number of cases shipped (X_1)
 - The indirect costs of the total labor hours as a percentage (X_2)
 - A qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X_3)
 - The total labor hours (Y)
- (i). Obtain the scatter plot matrix and the correlation matrix. What information do these diagnostic aids provide here?
- (ii). Write a multiple regression model to the data for three predictor variables. State the estimate regression function.
- (iii). Obtain the residuals and prepare a box plot of the residuals. What information does this plot provide?
- (iv). Plot the residuals against Y , X_1 , X_2 , X_3 , and X_1X_2 on separate graphs. Also prepare a normal probability plot. Interpret the plots and summarize your findings.
- (v). Prepare a time plot of the residuals. Is there any indication that the error terms are correlated? Discuss.
- (vi). Conduct the Brown-Forsythe test for constancy of the error variance, using $\alpha = 0.01$. State the decision rule and conclusion.
- (vii). Test whether there is a regression relation, using a level of significance of 0.05. State the alternatives, decision rule, and conclusion. What does your test result imply about β_1 , β_2 , and β_3 ? What is the P-value of the test?
- (viii). Calculate the coefficient of multiple determination R^2 . How is this measure interpreted here?
- (ix). From separate shipments with the following characteristics must be processed next month:

X_1	X_2	X_3
230,000	7.50	0
250,000	7.30	0
280,000	7.10	0
340,000	6.90	0

Management desires predictions of the handling times for these shipments so that the actual handling times can be compared with the predicted times to determine whether any are out of line. Develop the needed predictions, using the most efficient approach and a family confidence coefficient of 95%.

- (x). Three new shipments are to be received, each with $X_1 = 282,000$, $X_2 = 7.10$, and $X_3 = 0$.
- (a). Obtain a 95% prediction interval for the mean handling time for these shipments.

- (b). Convert the interval obtained in part (a) into a 95% prediction interval for the total labor hours for the three shipments.
content...
-