CSCI E-106:Assignment 5

Due Date: October 12, 2020 at 7:20 pm EST

Instructions

Students should submit their reports on Canvas. The report needs to clearly state what question is being solved, step-by-step walk-through solutions, and final answers clearly indicated. Please solve by hand where appropriate.

Please submit two files: (1) a R Markdown file (.Rmd extension) and (2) a PDF document, word, or html generated using knitr for the .Rmd file submitted in (1) where appropriate. Please, use RStudio Cloud for your solutions.

Problem 1

Refer to Plastic hardness data set. X is the elapsed time in hours? and Y is hardness in Brinell units. Build a model to predict Y. (30 points, 5 points each)

a-) Obtain the residuals ei and prepare a box plot of the residuals. What information is provided by your plot?

b-) Plot the residuals ei against the fitted values Y; to ascertain whether any departures from regression model (2.1) are evident. State your findings.

c-) Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

d-) Compare the frequencies of the residuals against the expected frequencies under normality, using the 25th, 50th, and 75th percentiles of the relevant t distribution. Is the information provided by these comparisons consistent with the findings from the normal probability plot in part (c)?

e-) Use the Brown-Forsythe test to determine whether or not the error variance varies with the level of X. Divide the data into the two groups, $X \leq 24$, $X > 24$, and use $\alpha = 0.01$. State the decision rule and conclusion. Does your conclusion support your preliminary findings in part (b)?

f-) conduct the Breusch-Pagan test to determine whether or not the error variance varies with the level of X. Use $\alpha = 0.01$. State the alternatives. decision rule, and conclusion. Is your conclusion consistent with your preliminary findings in part (b)?

Problem 2

Refer to Sales growth Data. (30 points, 10 points each)

a-) Divide the range of the predictor variable (coded years) into five bands of width 2.0, as follows: Band 1 ranges from X = -.5 to X = 1.5; band 2 ranges from X = 1.5 to X = 3.5; and so on. Determine the median value of X and the median value of Y in each band and develop the band smooth by connecting the five pairs

of medians by straight lines on a scatter plot of the data. Does the band smooth suggest that the regression relation is linear? Discuss.

b-) Create a series of seven overlapping neighborhoods of width 3.0 beginning at X = -.5. The first neighborhood will range from X = -.5 to X = 2.5; the second neighborhood will range from X = .5 to X = 3.5; and so on. For each of the seven overlapping neighborhoods, fit a linear regression function and obtain the fitted value $\hat{Y}_c$ at the center X_c of the neighborhood. Develop a simplified version of the lowess smooth by connecting the seven $(X_c, \hat{Y}_c)$ pairs by straight lines on a scatter plot of the data.

c-) Obtain the 95 percent confidence band for the true regression line and plot it on the plot prepared in part (b). Does the simplified lowess smooth fall entirely within the confidence band for the regression line? What does this tell you about the appropriateness of the linear regression function?

Problem 3

Refer to Plastic hardness Problem and data.(10 points, 5 points each)

a-) Obtain Bonferroni joint confidence intervals for β_0 and β_1, using a 90 percent family confidence coefficient. Interpret your confidence intervals.

b-) What is the meaning of the family confidence coefficient in part (a)?

Problem 4

Refer to Plastic hardness Problem and data. (25 points)

a-) Management wishes to obtain interval estimates ofthe mean hardness when the elapsed time is 20, 30, and 40 hours, respectively. Calculate the desired confidence intervals1. using the Bonferroni procedure and a 90 percent family confidence coefficient. What is the meaning of the family confidence coefficient here? (9 points)

b-) Is the Bonferroni procedure employed in part (a) the most efficient one that could be employed here? Explain. (8 points)

c-) The next two test items will be measured after 30 and 40 hours of elapsed time, respectively. Predict the hardness for each of these two items, using the most efficient procedure and a 90 percent family confidence coefficient. (8 points)