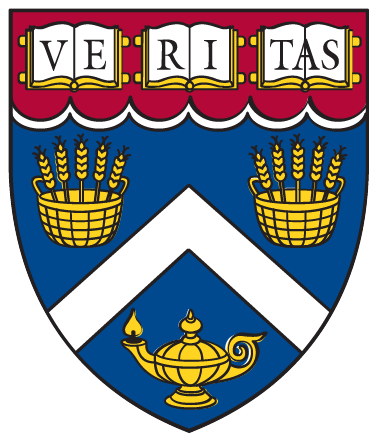
CSCI E-106: Data Modeling



Fall 2019

Dr.Hakan Gogtas

Assignment 8

Due: November 25, 2019 at 11:59pm 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 generated using knitr for the .Rmd file submitted in (1) where appropriate. Please, use RStudio Cloud for your solutions.

1. Refer to Brand preference data, build a model with all independent variables (45 pts)
2. Obtain the studentized deleted residuals and identify any outlying Y observations. Use the Bonferroni outlier test procedure with α = .10. State the decision rule and conclusion. (5pts)
3. Obtain the diagonal elements of the hat matrix, and provide an explanation for the pattern in these elements. (5pts)
4. Are any of the observations outlying with regard to their X values according? (5pts)
5. Management wishes to estimate the mean degree of brand liking for moisture content X1 = 10 and sweetness X2 = 3. Construct a scatter plot of X2 against X1 and determine visually whether this prediction involves an extrapolation beyond the range of the data. Also, use (10.29) to determine whether an extrapolation is involved. Do your conclusions from the two methods agree? (5pts)
6. The largest absolute studentized deleted residual is for case 14. Obtain the DFFlTS, DFBETAS, and Cook's distance values for this case to assess the influence of this case. What do you conclude? (5pts)
7. Calculate the average absolute percent difference in the fitted values with and without case 14. What does this measure indicate about the influence of case 14? (10pts)
8. Calculate Cook's distance D; for each case and prepare an index plot. Are any cases influential according to this measure? (5pts)
9. Find the two variance inflation factors. Why are they both equal to 1? (5pts)
10. Refer to the Lung pressure Data and Homework 7. The subset regression model containing first-order terms for X1 and X2 and the cross-product term X1X2 is to be evaluated in detail. (35 pts)
11. Obtain the residuals and plot them separately against Y and each of the three predictor variables. On the basis of these plots. should any further modification of the regression model be attempted? (5pts)
12. 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? (5pts)
13. Obtain the variance inflation factors. Are there any indications that serious multicollinearity problems are present? Explain. (5pts)
14. Obtain the studentized deleted residuals and identify outlying Y observations. Use the Bonferroni outlier test procedure with α= .05. State the decision rule and conclusion. (5pts)
15. Obtain the diagonal elell1ents of the hat matrix. Are there any outlying X observations? Discuss. (5pts)
16. Cases 3, 8, and 15 are moderately far outlying with respect to their X values, and case 7 is relatively far outlying with respect to its Y value. Obtain DFFITS, DFBETAS, and Cook's distance values for these cases to assess their inf1uence. What do you conclude? (10pts)
17. Refer to the Prostate Cancer data set in Appendix C.6 and Homework 7. For the best subset model developed in Homework 7, perform appropriate diagnostic checks to evaluate outliers and assess their influence. Do any serious multicollinearity problems exist here? (20pts)

