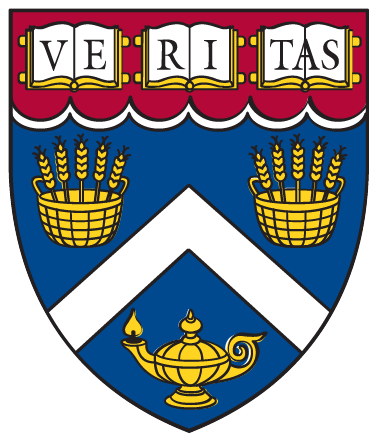
CSCI E-106: Data Modeling



Fall 2019

Dr.Hakan Gogtas

Assignment 5

Due: November 4, 2019 at 7pm 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 Plastic hardness data (20pts)
2. Using matrix methods, obtain the following: (1) (X’X)-1, (2) b, (3) Yhat, (4) H, (5) SSE, (6) s2(b), (7) s2(pred) when Xh= 30. (10 pts)
3. From part (a6), obtain the following: (1) s2{bo}; (2) s{bo, bl }; (3) s{bl }. (5pts)
4. Obtain the matrix of the quadratic form for SSE. (5pts)
5. Refer to the Brand preference data. In a small-scale experimental study of the relation between degree of brand liking (Y) and moisture content (X1) and sweetness (X2) of the product. (25 pts)
6. Obtain the scatter plot matrix and the correlation matrix. What information do these diagnostic aids provide here? (5pts)
7. Fit regression model to data. State the estimated regression function. Interpreted regression coefficients? (5pts)
8. Obtain the residuals, and prepare box plot of the residuals. What information does this plot provide? (5pts)
9. Plot the residuals against Yhat, X1, X2, and X1X2 on separate graphs. Also prepare a normal probability plot. Interpret the plots and summarize your findings. (5pts)
10. Conduct the Breusch-Pagan test for constancy of the error variance. State the alternatives, decision rule, and conclusion. (5pts)
11. Refer to Problem 2 (Brand preference data) (20 pts)
12. Test whether there is a regression relation, using α =0.01. State the alternatives, decision rule, and conclusion. What does your test imply about β1 and β2? (5pts)
13. Estimate β1 and β2 jointly by the Bonferroni procedure, using a 99 percent family confidence coefficient. Interpret your results. (5pts)
14. Obtain an interval estimate of E{Yh} when Xh1 = 5 and Xh2 = 4. Use a 99 percent confidence coefficient. Interpret your interval estimate. (5pts)
15. Obtain a prediction interval for a new observation Yh(new) when Xh1 = 5 and

Xh2 = 4. Use a 99 percent confidence coefficient. (5pts)

1. Refer to Commercial properties data. The age (X1), operating expenses and taxes (X2), vacancy rates (X3), total square footage (X4), and rental rates (Y). (25pts)
2. Obtain the scatter plot matrix and the correlation matrix. Interpret these and state your principal findings. (5pts)
3. Fit regression model for four predictor variables to the data. State the estimated

regression function. (5pts)

1. Obtain the residuals and prepare a QQ plot of the residuals. Does the distribution appear to be fairly symmetrical? (5pts)
2. Plot the residuals against Y, each predictor variable, and each two-factor interaction terms on separate graphs. Also prepare a normal probability plot. Analyze yours plots and summarize your findings. (5pts)
3. Divide the 81 cases into two groups. placing the 40 cases with the smallest fitted values into group 1 and the remaining cases into group 2. Conduct the Brown-Forsythe test for constancy of the error variance, using α= .05. State the decision rule and conclusion. (5pts)
4. Refer to Problem 4 (Commercial properties data) (10 pts). Three properties with

the following characteristics did not have any rental information available.



1. Based on the data above in the table. Develop separate prediction intervals for the rental rates of these properties, using a 95 percent statement confidence coefficient in each case. Can the rental rates of these three properties be predicted fairly precisely? What is the family confidence level for the set of three predictions? (10pts)



