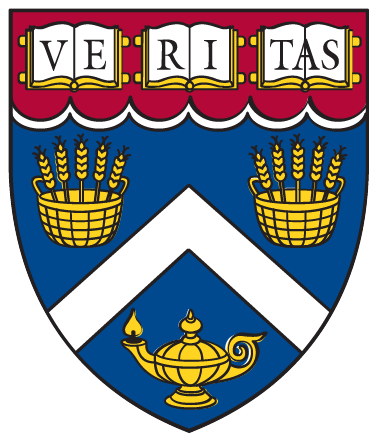
CSCI E-106: Data Modeling



Fall 2019

Dr.Hakan Gogtas

Assignment 3

Due: Monday, 10/07/19 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 the CDI data set. Using R2 as the criterion, which predictor variable accounts for the largest reduction in the variability in the number of active physicians?
2. Refer to the CDI data set in Appendix C.2 and Project l.44. Obtain a separate interval estimate of β1, for each region. Use a 90 percent confidence coefficient in each case. Do the regression lines for the different regions appear to have similar slopes?
3. Refer to GPA data:
4. Set up the ANOVA table.
5. What is estimated by MSR in your ANOVA table? by MSE? Under what condition do MSR and MSE estimate the same quantity?
6. Conduct an F test of whether or not β1 = 0. Control the α risk at .01. State the alternatives, decision rule, and conclusion.
7. What is the absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?
8. Obtain r and attach the appropriate sign.
9. Which measure, R2 or r, has the more clear-cut operational interpretation? Explain.
10. Refer to Crime rate data.
11. Compute the Pearson product-moment correlation coefficient r12.
12. Test whether crime rate and percentage of high school graduates are statistically independent in the population; use a α = .01. State the alternatives, decision rule, and conclusion.
13. Compute the Spearman rank correlation coefficient rs.
14. Test by means of the Spearman rank correlation coefficient whether an association exists between crime rate and percentage of high school graduates. State the alternatives, decision rule, and conclusion.

