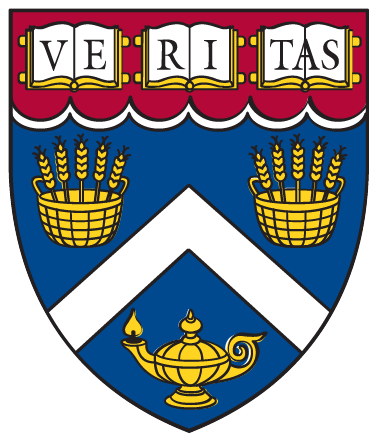
CSCI E-106: Data Modeling



Fall 2019

Dr.Hakan Gogtas

Assignment 6

Due: November 11, 2019 at 12am 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. An analyst wanted to fit the regression model Yi = β0 + β1 Xi1 + β2 Xi2 + β3 Xi3 + εi,

i = 1, ... , n, by the method of least squares when it is known that β2 = 4. How can the analyst obtain the desired fit by using a multiple regression computer program? (20pts)

1. Refer to the Commercial Properties data and problem in Assignment 5. (25 pts)
2. Obtain the analysis of variance table that decomposes the regression sum of squares into extra sums of squares associated with X4; with X1 given X4; with X2 , given X1 and X4; and with X3 , given X1, X2 and X4. (10pts)
3. Test whether X3 can be dropped from the regression model given that X1, X2 and X4 are retained. Use the F test statistic and level of significance .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test? (5pts)
4. Test whether both X2 and X3 can be dropped from the regression model given that X1 and X4 are retained; use α=.01. State the alternatives, decision rule, and conclusion. What is the P-value of the test? (5pts)
5. Test whether, β1 = -.1 and, β2 =.4; Use α=.01. State the alternatives, full and reduced models, decision rule, and conclusion. (5pts)
6. Refer to Brand preference data and problem in Assignment 5 (30 pts)
7. Transform the variables by means of the correlation transformation and fit the standardized regression model (10pts).
8. Interpret the standardized regression coefficient (5pts).
9. Transform the estimated standardized regression coefficients back to the ones for the fitted regression model in the original variables (5pts).
10. Calculate R2Y1, R2Y2, R212, R2Y1|2, R2Y2|1 and R2. Explain what each coefficient measures and interpret your results. (10pts)
11. Refer to the CDI data set. For predicting the number of active physicians (Y) in a county, it has been decided to include total population (X1) and total personal income (X2) as predictor variables. The question now is whether an additional predictor variable would be helpful in the model and, if so, which variable would be most helpful. Assume that a first-order multiple regression model is appropriate. (25 pts)
12. For each of the following variables, calculate the coefficient of partial determination given that X1 and X2 are included in the model: land area (X3), percent of population 65 or older (X4), number of hospital beds (X5), and total serious crimes (X6). (15pts)
13. On the basis of the results in part (a), which of the four additional predictor variables is best? Is the extra sum of squares associated with this variable larger than those for the other three variables? (5pts)
14. Using the F\* test statistic, test whether or not the variable determined to be best in part (b) is helpful in the regression model when X1 and X2 are included in the model; use α=.01. State the alternatives, decision rule, and conclusion. Would the F\* test statistics for the other three potential predictor variables be as large as the one here? (5pts)



