Project 3 (Due on 11.30.0 pm Friday, October 20)

Notes:

- Write comments on each step. Submit a report.
- **You are supposed to work on this project entirely on your own. So, do not consult with anyone within or outside the class.**
- You are welcome to ask me questions. However, first, try to find the answer on your own. Don't be afraid to google! Google is the best friend of a graduate student!!

1. Uric Acid and Cardiovascular Risk Factors: The cardio dataset contains data on 998 individuals on the following variables

Table 1: Overview of the datasets.

#	Variable	Description
1	uric	Uric acid level
2	dia	Diastolic blood pressure
3	hdl	High-density lipoprotein cholesterol
4	choles	Total cholesterol
5	trig	Triglycerides level in body fat
6	alco	Alcohol intake (ml per day)

   (a) Make scatterplots of Uric acid level with other variables and calculate the corresponding correlation coefficients. Based on these, choose one numerical variable that you think may be used effectively to predict Uric acid. Highlight any potential outliers on the scatterplot of this variable with Uric acid. [2]

   Fit a simple linear regression model and carry out regression diagnostics (including Brown Forsythe Test). The analysis should include an assessment of the degree to which the key regression assumptions are satisfied. If an assumption is not met, attempt to remedy the situation. Comment on the fit of the final model using appropriate tests and statistics. [3]

   Use the final model to provide (1) interval estimates, and (2) prediction interval of the mean uric acid level for patients whose predictor variable values are at first, second, and third quartiles. [2]

   (b) Remove five outliers (justify your choices) and redo the above analysis. Compare the two sets of results. [3]