SURV 616/686
Project #1
75 points

The dataset for this assignment is available on Canvas. The dataset is CHD.CSV (comma-delimited format). The dataset contains the following variables:

Variable Name	Variable Description
sbp	Systolic blood pressure
tobacco	Cigarettes per day
ldl	LDL cholesterol
adiposity	Measure of body fat
famhist	Family history of heart disease (CHD)
typea	Score on a test of Type A personality
obesity	Body Mass Index
alcohol	Ounces per day
age	Age
chd	Coronary Heart Disease; 1=present, 0=absent

The goal is to predict coronary heart disease (CHD) using the other available variables. For this homework, you will write 2-4 pages. **The focus of the paper will be on a description of your model selection strategy and an explanation of the results.** The paper should not be crammed with tables and figures and include little description. Therefore, <u>this paper will include not more than 4 tables or figures</u> (2 tables + 2 figures also = 4). Please do not include raw output from the R console in the paper. Format tables and figures for a quality presentation. You should include as an appendix the code that you used and any additional output that you feel is necessary. In the paper, please do the following:

1) Describe how you selected your final model. What preliminary analysis did you do? How did you determine which predictors to include? Did you try any transformations or interactions? How well does the final model fit the data? How have you assessed predictive accuracy?

2) Explain your final model. What is the meaning of the coefficients? You need not explain each coefficient, but explain any that are particularly important.

See the rubric for additional description of what should be included.