

Statistical and Machine Learning (Spring 2018)

Mini Project 2

Instructions:

- Due date: Feb 15, 2018.
- Total points = 20.
- Submit a typed report.
- Submit only one report per group, and include a description of the contribution of each member.
- It is OK to discuss the project with other students in the class (even those who are not in your group), but each group must write its own code and answers. If the submitted report (including code and answer) is similar (either partially or fully) to someone else's, this will be considered evidence of academic dishonesty, and you will be referred to appropriate university authorities.
- Do a good job.
- You must use the following template for your report:

Mini Project #

Name

Names of group members

Contribution of each group member

Section 1. Answers to the specific questions asked

Section 2: R code. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.

-
1. Consider the crime data stored in `crime.csv`. We would like to understand how murder rate is related to the other variables in the dataset. Note that `state` is the “subject” here; it’s not a predictor, and `region` is a qualitative variable.
 - (a) Build a multiple linear regression model to predict murder rate based on the other variables. The model should have all the important variables and it should not have any unimportant variables. Be sure to explore the interactions as well. Perform model diagnostics to check the standard model assumptions and perform any transformations needed to obtain a model for which the assumptions reasonably hold.
 - (b) Use your final model to predict murder rate of a state whose predictor values are set at the average in the data for a quantitative predictor and the most frequent category for a qualitative predictor.