

DSC 241 – Homework 4

Problem 1. In this problem, you practice working with predictor variables that are discrete. Consider the **Boston** dataset in the package **MASS**. Take as response the median property value.

- a. Look at side-by-side boxplots for **medv** where the groups are defined by **chas**. Comment on what you observe. In particular, compare the different groups visually. Then fit a model explaining **medv** as a function of **chas**. Output an ANOVA table. What is the F-test testing? Is the result consistent with the boxplots?
- b. Repeat with **rad** in place of **chas**.
- c. Produce a nice boxplot display of **medv** where the groups are defined by **chas** and **rad** jointly. Comment on what you observe. Then look at an interaction plot. Then fit a model explaining **medv** as a function of **chas** and **rad** with interactions. Output an ANOVA table. What are the different F-tests testing? Compare with the previous F-test as appropriate. Are the results of these tests consistent with the plots you just looked at?
- d. It makes sense that median property value decreases with the percentage of lower status population **lstat**, and this is indeed what is observed here. Does the rate of decrease depend on whether the area borders the Charles River? Produce a plot that helps answer that question. Then formulate that into a hypothesis testing problem and perform an appropriate test.

Problem 2. Consider the same dataset and turn to the problem of fitting a polynomial model explaining **medv** as a function of **lstat**.

- a. Fit a polynomial model of degree 3 by least squares.
- b. Repeat with each robust method covered in the lecture notes/slides.
- c. Produce a scatterplot and overlay all these fits with different colors and a legend.

[HINT: use the function **predict**.]