Midterm Two - Statistics 151a, Fall 2015

Due on 24 November, 2015

17 November, 2015

1 Resources

You must not consult any person or resource (online or otherwise) for your analysis. If you have any clarification questions about the data etc., you may contact Prof. Guntuboyina or Sujayam Saha via email. You may consult the materials from class that are posted on bcourses, the textbooks listed in the syllabus and documentation for R functions and packages. **You may not access any other resources**.

2 Submission

You are required to submit a short report describing your analysis. The printed reports will be due at the beginning of class on 24 November, 2015.

The report should contain an introduction, brief description of the data, a description of the analyses and results and general discussion/conclusions. Your R code should be included as an appendix at the end of the report (code should not appear at any other place in the report). Figures should be carefully chosen, labeled, and referred to in the text. The maximum number of pages allowed in the report is eight. Also note the following points:

- The report should be written in paragraphs and sentences.
- The text should form a logical narrative, and not be just a series of plots or statistics. The text should be written in the best order to make a coherent report, i.e., it should not simply follow the order in which you did the analysis.
- Include plots to supplement the narrative, and again they should be a logical part of the narrative of your results and appropriate to the analysis.
- Spell check and grammar check your report.
- Comment the R code appropriately so that it is easy to understand the code.

3 The Data

You are required to analyze the Boston Housing data which can be accessed from R via:

```
library(mlbench}
data(BostonHousing)
help(BostonHousing)
```

There is also a related dataset BostonHousing2 (see the help output above) which has more variables. You may work with either of the two datasets.

Data is collected from 506 census tracts in Boston (from 1970) about housing prices and other variables. Your task is to model the relationship between housing price (given by the variable *medv*) and the explanatory variables. On the basis of your analysis, can you come up with answers to interesting questions? For example, two such interesting questions are (a) Which variables mainly determine the median housing price in a tract?, and (b) What is the willingness to pay for clean air? etc. Can you come up with more such questions and reasonable answers to them on the basis of your analysis?