Math 445 Midterm – Part II

Prof. Adam Loy

Due Wednesday, May 7, 2014 at 11:10 am

1. For many years until 1995, the US national speed limit was 55 miles per hour (mph). A federal law in late 1995 abolished the national limit and allowed states to set their own limits. Many, but not all, states promptly increased the freeway speed limit. The following data on change in freeway traffic fatalities were collected as part of a study on the effects of changing the speed limit.

   Thirty-two states increased the speed limit either at the beginning of, or during, 1996. Nineteen states left the speed limit unchanged. The data include the District of Columbia, which is why there are 51 observations. The data file, `speedlimit.txt`, contains the state, whether or not it increased the speed limit, and the percent change in traffic fatalities from 1995 to 1996. The percent change is used as a response because states with large populations will have more fatalities, both in 1995 and in 1996.

   Imagine you work in the Department of Transportation in one of the states that did not change its speed limit. Your boss has asked to analyze the data and write (this should be typed) a short report summarizing your findings. Your boss is specifically interested in:

   - Was the change in traffic fatalities the same in the two groups of states?
   - How large is the difference and how precisely is that estimated?

   Your report should include one sentence (or more if necessary) describing your analysis and an answer to each of your boss's questions, using appropriate statistical summaries. Your boss is a busy person so she only wants to see results from a single analysis. If you analyze the data many ways, present only the one that you believe is the most appropriate.

   In addition to the report for your boss, write a short overview of the methods that you used. In this overview, be sure to justify any assumptions that you need to make in order to perform your analysis. You should submit any R commands and output used in your final analysis as part of this overview.

2. We wish to estimate the variance, σ^2, in a population, and we have two potential estimators, which we have seen before in class:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

   You may use the fact that

$$\left(\frac{1}{\sigma^2}\right)\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sim \chi^2_{n-1}$$

   without proof.

   (a) Calculate the bias of both estimators.

   (b) Calculate the variance of both estimators.

   (c) Calculate the MSE of both estimators.

(d) What is the relative efficiency of $\widehat{\sigma}^2$ to s^2?

(e) Make an argument for which estimator is better to use in practice.