Final_project

MA 677

4/17/2022

Throughout the semester, I have emphasized that the goal of this course is to prepare you for a program of exploration and learning as a necessary component of successful statistical practice in data science. In the second half of the semester I have intentially chosen texts that you could begin reading in class and complete after the end of the semester. The final project is intended as a final push out of the nest.

You may start this final assignment from either *In All Likelihood* or *Introduction to Empirical Bayes*. Each starting place has it's own instructions.

In All Likelihood

Begin by reading Chapter 4, working through the examples as usual.

Using R, prepare answers to exercises 2.25, 4.39, and 4.27.

Now consider the data in the spreadsheet file Illinois_rain_1960-1964.xlxs which reports amounts of precipitation during storms in Illinois from 1960 to 1964. These data were gathered in a study of the natural variability of rainfall. The rainfall from summer storms was measured by a network of rain gauges in southern Illinois for the years 1960-1964 (Changnon and Huff, 1967). The average amount of rainfall (in inches) from each storm, by year, is contained in the spreadsheet.

Use the data to identify the distribution of rainfall produced by the storms in southern Illinois. Estimate the parameters of the distribution using MLE. Prepare a discussion of your estimation, including how confident you are about your identification of the distribution and the accuracy of your parameter estimates.

Using this distribution, identify wet years and dry years. Are the wet years wet because there were more storms, because individual storms produced more rain, or for both of these reasons?

To what extent do you believe the results of your analysis are generalizable? What do you think the next steps would be after the analysis? An article by Floyd Hufff of the authers of the 1967 report is included.

Introduction to Empirical Bayes

Continue reading *Introduction to Empirical Bayes*, chapters 5, 6, 11, 12, and 13.

Now, read *Computer Age Statistical Inference*, Chapter 6. This chapter takes you to the next step in using empirical bayes and contains four examples – insurance claims, estimating species discovery, estimating the size of Shakespeare's vocabulary, and estimaing the number of lymph nodes.. In each example, the result of the empirical Bayes analysis is given.

Using R, reproduce the analyses in Chapter 6 describing how you are proceeding and providing commentary. Note that at the end of the chapter, the authors noted that empirical Bayes a commonly used to support analysis of False-discovery rates which is discussed in Chapter 15.