Pre-lab 7

Stats 413: Applied Regression Analysis

due Nov 12, 2020 before the first lab session

Pre-labs are due on Canvas on the due date. For problems that require programming, please properly comment your code and submit it together with any output. You are encouraged to collaborate on pre-labs with classmates, but the final write-up (including any code) must be your own.

Data link: http://users.stat.umn.edu/~sandy/alr4ed/data/

9.16 Florida election 2000 (Data file: `florida`) In the 2000 election for U.S. president, the counting of votes in Florida was controversial. In Palm Beach County in south Florida, for example, voters used a so-called butterfly ballot. Some believe that the layout of the ballot caused some voters to cast votes for Buchanan when their intended choice was Gore.

The data from Smith (undated) has four variables, `County`, the county name, and `Gore`, `Bush`, and `Buchanan`, the number of votes for each of these three candidates. Draw the scatterplot of `Buchanan` versus `Bush`, and test the hypothesis that Palm Beach County is an outlier relative to the simple linear regression mean function for E(`Buchanan`|`Bush`). Identify another county with an unusual value of the Buchanan vote, given its Bush vote, and test that county to be an outlier. State your conclusions from the test, and its relevance, if any, to the issue of the butterfly ballot.

Next, repeat the analysis, but first consider transforming the variables in the plot to better satisfy the assumptions of the simple linear regression model. Again test to see if Palm Beach County is an outlier, and summarize.