Statistical Methods in Linguistics

Homework 1

Due: February 7, 2018 at 4pm

1. Participants in the pilot experiment from class gave the following ratings for Item 6:

   1, 2, 1, 3, 1, 2, 5, 1, 2, 2, 3, 1, 7, 2, 3, 1

   For this problem, you are expected to do the problems by hand and show your work.

   (a) Give a measure of the central tendency of the data (mean or median).

   (b) Give a measure of the spread of the data (SD or IQR).

   (c) Why did you choose those statistics in (a) and (b) rather than the alternatives?

   (d) Are there any outliers in this data? If so, what are they and what method did you use?

   (e) Provide a graphical representation of the ratings (histogram or boxplot).

2. You perform an experiment that involves a magnitude estimation task. The ratings given to Item 6 are the following:

   125, 145, 150, 163, 133, 140, 145, 160, 125, 150, 200, 95, 135, 140, 155, 145

   Are there any outliers in the data set? What are they, and which method did you use to determine which ones they are? Why that method?

3. Load the 'english' data from the languageR package in R.

   (a) Just to make sure you know what you're looking at: Please give a general description of the experiment that this data comes from (what was being tested).

   (b) Plot a histogram of the response time for the lexical decision task – the variable name is RTlexdec. Describe what this tells you about the data. (Please print the histogram and submit it with your homework.)

   (c) Are there outliers in this data? Select a method that eliminates outliers, then plot a histogram of the new response time data. (Please submit the new histogram as well.)

   (d) Provide a brief report of the exploratory statistics you performed on this data, and what questions you might ask going forward.

4. Construct a data set that will have two outliers if you use the non-conservative IQR-based method to determine outliers, but only one if you use the conservative method.

5. Apply the mean-based 2SD method for finding outliers on the problem you constructed in the last problem. How many outliers does this method say there are?

6. Describe a data set for which the median would be a less representative measure than the mean. Use either a graphic representation or sample numbers to illustrate.