Computer Lab Project No. 3

Describing, Exploring, and Comparing Data - Numerically

In today's lab we will explore obtaining and analyzing numerical summary statistics such as mean and standard deviation in StatCrunch. We will also study boxplots.

Info

Obtaining Summary Statistics:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Stat" in the menu bar.
4. Click on "Summary Stats".
5. Select the column(s) which contain the data of interest. By control-clicking, you can select more than one.
6. Click "Calculate" to obtain the descriptive statistics.

Boxplots:

1. Start StatCrunch.
2. Load relevant data into StatCrunch.
3. Click on "Graphics" in the menu bar.
4. Click on "Boxplot".
5. The optional "Group by" column can be used to compare boxplots across groups on a single graph.
6. Click "Next" to indicate whether or not to use fences for the boxplots. Note: The five-number summary is used by default.
7. Click "Create Graph" to construct the boxplot(s).

Do now

1. Load the word count data set from the textbook; in edition 14 of the textbook, this is data set 14. It is a count of the number of words spoken in a day for both men and women, taken in six different samples.
2. First for a bit of preprocessing, let's combine all of the counts for men in one column and all of the counts for women into another column. This can be accomplished through as follows:
   (a) Click on "Data" in the menu bar.
   (b) Select "Arrange ▶ Stack" and select all the columns beginning with (M) for male.
   (c) Click "Compute!" and verify that your data is now combined on your spreadsheet. Change the name of the new column, containing the word counts for men, to *Men*. Another column is created, containing the column header each word count came from. We won't need this column in what follows, so you may delete it.
   (d) Use the same process to stack the word counts of women in one column.

3. Using what you learned in Lab 2, create and compare a histogram of the men's and women's word data. By choosing both variables at the same time, you can create both graphs at once. In order to better be able to compare them, under "For multiple graphs," check "Use same X-axis" and "Use same Y-axis". To get the graphs to be displayed side-by-side, choose 1 row and 2 columns per page.

4. Ask yourself whether the graphs appear to have the same general shape? Would you consider these samples to have come from a population that was normally distributed? Why/why not?

5. Calculate the summary statistics and use them to comment on the central tendency and variation of the data. Does it appear that women are more talkative than men?

6. Create side-by-side boxplots for both variables. Do these boxplots reinforce the description you gave above? How many outliers are present?

Finish

- Download the histograms and box plots you created to your hard drive, as a png file. To do this:
    1. Click "Options" in the window displaying the graph, and select "Download".
    2. Type a name for your file, for example "Lab3-Histogram".
    3. Click "OK". The file will be saved in your Downloads folder.
- Create a document using your favorite word processor, for example google docs, in which you include the histograms, the box plots and the summary statistics (you can just copy and paste them), and answer the questions you were asked in the "Do now" section.
- Save this document as a pdf file. Name it "Lab3-Firstname-Lastname.pdf".
- Submit this file on Blackboard as your Lab 3 assignment.