

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.

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. Identify the column(s) which contain the data of interest.
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).

Here is what you should do today:

1. Load the data set titled *Word Counts*. 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 males in one column and all of the counts for females into another column. This can be accomplished through the following:
 - (a) Click on "Data" in the menu bar.
 - (b) Select "Stack Columns" and select all the columns with (M) for male. Store the labels in var13 and the data in var14.
 - (c) Click "Stack Columns" and verify that your data is now combined on your spreadsheet. You should change the name of the var14 column label to Male.
 - (d) Repeat this by putting the female data in columns var15 and var16.
3. Using what you learned in Lab 2, create and compare a histogram of the Male and Female word data. Do they appear to have the same general shape? Would you consider these samples to have come from a population that was Normally distributed?
4. 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?

5. Create side-by-side boxplots for both variables. Do these boxplots reinforce the description you gave above? How many outliers are present? Did you notice those outliers earlier?
6. Now repeat the steps above separately for sample 3 (which was taken on students in Mexico) and sample 5. How would you describe and compare these two samples using relevant displays and summary statistics?