Math 1040 StatCrunch Exercise 2

Exercises from Elementary Statistics 12th ed. By Mario F. Triola

The purpose of this exercise is to introduce you to some more features of the StatCrunch software that is available with your textbook and MyStatLab program. The skills you learn in this exercise will be needed for your term project and for some of your homework. You will learn how to calculate the values that numerically summarize data, how to calculate lower and upper fences, how to find outliers of a data set, and how to create a boxplot and a modified boxplot

Part 1: calculating summary statistics

Using the StatCrunch software, load the data from Data Set 10: Cigarette Tar, Nicotine, and Carbon Monoxide. (If you cannot recall how to accomplish this or any of the previously learned tasks, please refer to StatCrunch Exercise 1) This data set contains several columns of variables. We will be working only with the MnTar data here, so you may wish to delete other columns. These values specifically represent the amount of tar in milligrams per cigarette found in a sample of Menthol cigarettes.

Begin by clicking on “Stat” on the toolbar at the top of the spreadsheet. Choose “Summary Stats” and click “Columns”. Choose your data by clicking on the title of the column, “MnTar”. To apply the default options, now click on “Compute!”. Click on “Options”, “Copy”, and follow the instructions to copy the table of default summary statistics into the document file that you will submit (Word is a good choice). Save this document. You will be adding to it!

Notice that the table as copied may not look right. In some word processing documents, your text and numbers do not fit inside the cells causing it to wrap. You should never turn in a college level report with such a messy table. For this exercise, as well as your term project, you must be sure that your tables are well presented and are easy to read. One way to make it look better is to highlight the text in your word processing document and change the font to a smaller size. Another way to keep numerical values from wrapping is to round the values to fewer decimal places. Try it!

You can make the text fit properly, but there are still some problems. Can you explain the meaning of standard error (Std. Err.)? If not, then you had better not include it in a report! Also note that we will seldom place as much importance on the variance as we do on the standard deviation, so you may not even want to include the variance. In StatCrunch, close the window with the summary statistics data and let’s try again.

Again click on “Stat”, “Summary Stats”, and select the column tar in menthol cigarettes. But now scroll down and look at the other options. There is a menu that allows you to select the statistics you want in your output table and remove the statistics that you do not want. You may also change the order in which they appear by entering them in the order you prefer. Try double clicking on “*n*” in the left column. Now the only statistic in the right column is the sample size, *n*. Hold down the “ctrl” key on your keyboard and click on “mean” in the left column and see that appears on the right side now. For now, let’s not include “Variance” and “Std. Err.” in the output. Do include the sample size, mean, standard deviation, median, range, minimum, maximum, first and third quartiles, and interquartile range. Compute again and copy the new output table into your document file. Place it below your first table and make sure it is formatted so that the text does not wrap and all values are easy to read. Save your file.

Notice that the computer does not give you the mode. The easiest way to find the mode of this data set is to sort it and look to see what, if any, value occurs most often. Sort the column of tar amounts in menthol cigarettes. Use the arrow key on your keyboard to scroll down the column and determine the mode. In the document file that you will be turning in, type the value of the mode under the two tables of statistics you have already copied.

Part 2: finding outliers

While your data is sorted, you may also look for outliers. Page 121 of your book explains one method for identifying outliers using the lower and upper fences:

 Lower fence = Q1 – 1.5(IQR)

 Upper fence = Q3 + 1.5(IQR)

Recall that Q1 is the first quartile, Q3 is the third quartile, and IQR is the inter-quartile range. The IQR is the difference between Q3 and Q1, and is the range of the middle 50% of the data. Any data value falling below the lower fence or above the upper fence would be considered an outlier. (See page 121 of your book for an example.)

Calculate the values of the lower and upper fences for the cigarette data (Notice that you have already found all the values you need to make your calculations.) In the document that you will be turning in, type these calculations and the values of the lower and upper fence. List any outliers for this data set.

Part 3: Creating a boxplot, creating a modified boxplot

You should now have a good idea how StatCrunch can help you find your numerical summary statistics. Another tool for summarizing data that you learned in Chapter 3 is the boxplot. To create a boxplot for the cigarette data, click on “Graph” on the toolbar at the top of the screen and choose “Boxplot”. Select the MnTar data column. Under “Other options”, check “Draw boxes horizontally”, but do not change any other default settings. Give the graph an **appropriate title**, and click “Compute!” to create the box plot. Copy this box plot into your document file. Below the boxplot, **type a** **complete sentence** **explaining the shape** of the distribution of the tar data. Save your file.

Notice that the standard boxplot you just created shows only the 5-number-summary, but does not indicate the outliers. A modified boxplot will use the lower and upper fence values and show the outliers in the graph. Create a boxplot again with the MnTar column and this time under “Other options” check both choices to draw it horizontally and to use fences to identify outliers. Give the graph an **appropriate title**, and click “Compute!” to create the box plot. Compare the modified boxplot to your earlier calculations to make sure your list of outliers is in agreement with the outliers displayed in the plot. Copy this boxplot into your document file. Below the modified boxplot, **type complete sentences explaining** how the outliers in this data set could be misleading.

Part 4: Practicing your new skills

Now you will practice all these skills that you just learned with a new data set! Load the data from Data Set 8: Alcohol and Tobacco Use in Animated Children’s Movies. Use the “Tobacco Use” column for the following exercise.

Use StatCrunch to output summary statistics for this data set that include the sample size, mean, standard deviation, 5-number summary values, and interquartile range. Do not include the standard error. Copy the output table into your document file, formatting it so that it is easy to read.

Find the mode, if any, of this data set. Report your findings in your document file.

Calculate the lower and upper fences and list any outliers in the data. Report these values in your document file.

Create a modified boxplot for this data, give your boxplot an appropriate title using the StatCrunch editing options, and then copy it into the document you will be turning in. Type complete sentences under this boxplot that explain the shape of the distribution of the data and also how the outliers (if any) could affect the interpretation of the data.

Part 5: Wrapping up

Read back through the assignment and double check that you have completed each part. Make sure your document file is neat, orderly, and easy to follow. Make sure your name is on your assignment.

Print your document and turn it in by the posted due date.

Last updated August 2014