

BIOL 322 (BIOSTATISCS) – REPORT 1

Topics covered: plots, descriptive statistics and frequency distributions

Due - Thursday October 4, 2018; **Handling** – Reports should be printed (hand written reports won't be accepted) and handed to Dr. Peres-Neto on the day of the deadline either during his office hours (10:30am to 11:30am (SP-437.13 LOY) or in classroom (before 11:45am but not after class; SP S110 LOY).

Important notes:

A) In this assignment, you will generate your own data. Think of each series of data that you will generate as a sample from a population of interest. Please send the data to your TA by email (by the same due date and time) in csv format. This can be done in excel and if you don't know how to save an excel spreadsheet in csv format, follow the instructions in this video: https://www.youtube.com/watch?v=3jW_ZJpVy_0

You will generate 6 files csv files and one RStudio file by the time you finish your report.

That said, you may want to first use the function `c` in R so that you can play with your data to achieve the desired effect requested by each task. For example, we can create a series of 11 numbers using `c` as follows (as seen in the tutorials):

```
X <- c(13,40,62,66,111,124,135,140,157,163,193)
```

Ooops...that did not work to generate the type of data I needed, so go there and change the numbers, e.g.,

```
X <- c(1,40,42,66,111,124,135,140,157,163,345)
```

Until you find the desired effect requested. Once you generate the series with the desired effect, then enter it in excel and save it in a csv format. This will help you practice saving data in csv and loading data in R. That's the way we analyze real data!

B) Place each data that you generated in a separate csv file (i.e., one data per csv file) using the name given in each task below.

C) You will use R to work on this assignment; send to your TA the R file (produced in RStudio) with the commands you used to generate the results. *The code should all go into one single file*, but in the RStudio file it should appear as follows:

```
# complete name and student number
#task 1
data1 <- read.csv...
enter lines of code to solve task 1
#task 2
data2 <- read.csv...
enter lines of code to solve task 2
#task 3
data3 <- read.csv...
enter lines of code to solve task 3
```

And so on for each task.

Really important: before you send the R file to your TA, make sure that all the code runs fine and without errors. And remember, you have to send the data in a csv file. Data using the c command as shown in A above won't be accepted.

D) There is no strict format for the report and you can use as many pages as necessary. But try to make the report as organized as possible. Use Arial or Times New Roman as fonts (12 points preferably for text; don't worry about the font size for graphs though).

Do not write code directly into the report; codes should be in separate files as explained in each task (see below).

E) Copying (from RStudio) and pasting graphs in your document. Once a graph is produced, click "export" in the graph window (lower right corner) and then click "copy to keyboard". Once the new window appears, click "Copy plot". Then use the paste command in your text editor to paste the graph in the report.

F) QUESTIONS? If you have questions, send it to me via Moodle. Note, however, that I will copy your question and answer it to all students as they may serve useful to others (your name won't appear in the message). Please be as precise as possible. Only questions regarding requests that were not clear are allowed; specific questions on how to

run analysis, etc, won't be answered; if you attended tutorials you should be more than able to work on the tasks using R.

G) DO NOT FORGET to write your complete name and student number on the printed report and the email with files sent to your TA.

Tasks:

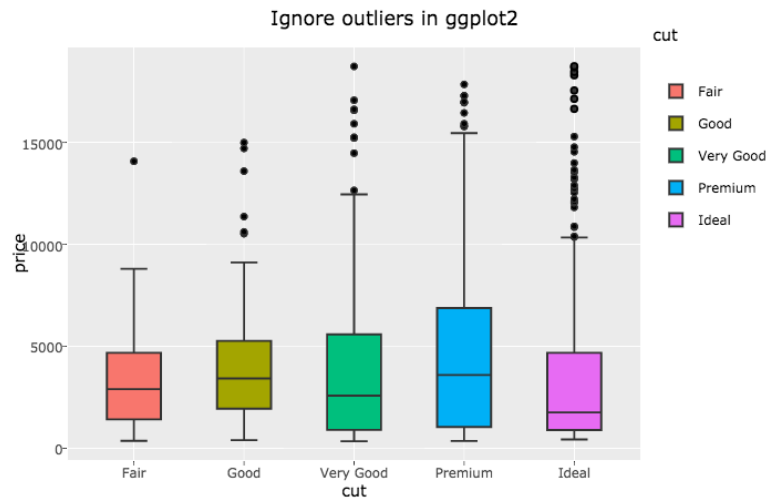
1) Generate a series of numbers containing a minimum of 10 and a maximum of 30 numbers (discrete or continuous) that is clearly left-skewed distributed. For the series, calculate the mean, median, variance, standard deviation, coefficient of variation and the interquartile range. Produce a histogram and a boxplot (graphs and not values to produce the boxplot) for the series. Presentation of results: values for the descriptive statistics should be presented in a table and graphs (histogram and boxplot) should be copied and pasted in your report; file name for the data should be: series1.csv

2) Generate a series of numbers containing a minimum of 10 and a maximum of 30 numbers (discrete or continuous) that is clearly right-skewed distributed. For the series, calculate the mean, median, variance, standard deviation, coefficient of variation and the interquartile range. Produce a histogram and a boxplot (graphs and not values to produce the boxplot) for the series. Presentation of results: values for the descriptive statistics should be presented in a table and graphs (histogram and boxplot) should be copied and pasted in your report; file name for the data should be: series2.csv

3) Generate a series of numbers containing a minimum of 10 and a maximum of 30 numbers (discrete or continuous) that is clearly symmetrically distributed (unimodal and not uniform, i.e., with one clear peak in the distribution). For the series, calculate the mean, median, variance, standard deviation, coefficient of variation and the interquartile range. Produce a histogram and a boxplot (graphs and not values to produce the boxplot) for the series. Presentation of results: values for the descriptive statistics should be presented in a table and graphs (histogram and boxplot) should be copied and pasted in your report; file name for the data should be: series3.csv

4) Find a published paper reporting results in the form of a boxplot. The boxplot should contain at least 2 series of data (distributions; see example below for 5 distributions). In your report, provide a small description of the study (3 to 5 sentences) and write the complete reference of the paper in the report (also send a pdf copy of the paper along with your report to your TA). Copy and paste the Boxplot you selected from the paper in your report and describe whether you considered the distribution of each series (each box)

right-skewed, left-skewed or symmetrical. In general, each series has a title (e.g., in the boxplot below the first series or distribution is “Fair” and the last is “Ideal”); refer to the title of the series when reporting your interpretation about the type of distribution (i.e., right-skewed, left-skewed or symmetrical).



source: <https://stackoverflow.com/questions/43499229/removing-outliers-from-boxplot-and-plotly>

One way to find papers with boxplot (though you can always browse journals) is to search in Google a topic of your preference adding the word boxplot. For example, when I searched: “boxplot behaviour ecology” (without the quotation marks) and went to “Images” in Google, I selected a boxplot that looked interesting which led me to the following paper: <http://rspb.royalsocietypublishing.org/content/282/1810/20150061> Really interesting paper, by the way.

5) Generate two series of numbers (say X and Y) that have a negative trend (values in the X series go up while values in the Y series go down in relation to one another).

Produce a scatter plot. File name should be: series4.csv

6) Generate two series of numbers (say X and Y) that have a positive trend (values in the X and Y series go up in relation to one another). Produce a scatter plot describing the relationship of the two series. File name should be: series5.csv

7) Generate two series of numbers (say X and Y) having 10 observations each. Both series X and Y should have a mean of 5.0, but series X should have a much larger standard deviation in contrast to series Y. Each series should be placed in a separate column in file series6.csv