STAT 614 - HW 1

Due: Tuesday, September 8, 2020 in Blackboard (go to the Homework folder under the Homework/Classwork content area) by 11:59pm.

Instructions: Please type your solutions in a separate document and upload the document in Blackboard. Submit *two* documents:

1. Your answers to questions below as a **pdf**.
2. An R script file with only the code needed to generate your output for Problem 1. I should be able to run this without errors!

Do not copy and paste solutions into Blackboard – use a separate document. Include supporting work (plots, etc.) when appropriate, but do not copy all computer output. Select only relevant output. Select only the minimum syntax code needed to reproduce your results.

Notes:

- This HW gives you practice on finding numerical and graphical summaries in R. These summaries are our primary tools for describing the distribution (i.e. behavior) of a variable.
- Graphs from R and other software can be copied and pasted into documents. RStudio has an excellent export feature for plots.
- You will also identify the key components of the design of a study.

Problem 1

The effects of exposure to lead on the psychological and neurological well-being of children were studied by Landrigan et al. (1975). Complete raw data for this study are in the data set lead.csv on Blackboard. The data in lead.csv describe a group of children who lived in one of three areas (`area`) near a lead smelter in El Paso, Texas. An exposed group of 46 children were identified who had blood-lead levels $\geq$ 40 μg/ml in 1972 (or for a few children in 1973). This group is defined by the variable `GROUP` = 1. A control group of 78 children were also identified who had blood-lead levels < 40 μg/ml in both 1972 and 1973. This group is defined by the variable `GROUP` = 2.

1. How many individuals are in the data set?
2. How many variables are in this data set?
3. Can you tell if any of the variables are categorical (i.e. qualitative)? Identify specific ones.

Two important variables that were studied were (1) MAXFT = the number of finger-wrist taps in the dominant hand (a measure of neurological function) and (2) IQF = the Wechsler full-scale IQ score. You will explore the relationship of lead exposure to one of these two outcome variables.

4. Is this an observational study or a randomized experiment? Explain why.
5. How many individuals have MAXFT scores measured? How many have IQF scores measured?
6. **Pick one of** `MAXFT` or `IQF` of interest to you. We are primarily interested in comparing the distribution of the outcome of interest (MAXFT or IQF) for the two different groups of children (GROUPS 1 and 2, those children with elevated blood-lead levels $\geq$ 40 μg/ml and those with lower levels, < 40 μg/ml, respectively.)

a. What are the mean and median of the outcome of interest (MAXFT or IQF) for each GROUP?

b. Describe the shape of the distribution (i.e. histogram) of the outcome for each GROUP.
c. What information can we get from the Boxplot of the outcome for each GROUP?
d. Based on these summaries, what is your assessment of the differences between the two groups of children on the outcome of interest? Discuss the role of randomization in this study.

References:

- Rosner B, (2005) *Fundamentals of Biostatistics*, Thomson.
- Landrigan PJ, Whitworth RH, Baloh RW, Staehling NW, Barthel WF, Rosenblum BF (1975) Neuropsychological dysfunction in children with chronic low-level lead absorption. *Lancet*, 1, 708-715.

Problem 2

Re-visit the study design for Case Study 1.1.2 Sex Discrimination in Employment from Chapter 1 of the textbook or our class notes. Briefly contrast the case study design with that of the study described in the *New York Times* article "Bias Persists for Women of Science, a Study Finds'' and in the manuscript *Science Faculty's Subtle Gender Biases Favor Male Students* (both are given in Blackboard under the *Homework* tab.) Briefly describe the overall goals of each study and give the strengths and weakness of each. Which study do you find has more compelling evidence for the hypotheses of interest?

Problem 3

Please take the brief and anonymous *Introductory Survey* located in Blackboard under the *Homework* tab.

Problem 4

Please send me an email and tell me at least one thing about yourself. This can be anything - you get to decide what you would like to share with me. It could be about your hopes and wishes for this class, any concerns you have about taking your courses online, a fun fact about yourself I wouldn't otherwise learn…whatever you want me to know about you. You can share whatever you want to share!

Also, feel free to ask me something in this email, whether related to the course or not. My hope is this will be the beginning of a dialogue and you become comfortable communicating with me in a variety of formats. I will respond to your email – and I try to respond to all email within 24 hours but because I have nearly 50 students this semester it may take me a little longer this time to get back to you!