P8130 Fall 2020: Biostatistical Methods I

Homework 1

Due Friday, 9/25 @ 5:00pm

P8130 Guidelines for Submitting Homework

- Your homework must be **submitted through CourseWorks**. No email submissions!
- **Only one PDF file should be submitted**, including all derivations, graphs, output, and interpretations. When handwriting is allowed (this will be specified), scan the derivations and merge ALL PDF files (http://www.pdfmerge.com/).
- You are encouraged to use R for calculations, but you must show all mathematical formulas and derivations. **Please submit your commented code as a separate R/RMD file.**
- To best follow these guidelines, we suggest using Word (built in equation editor), R Markdown, Latex, or embedding a screenshot or scanned picture to compile your work.

DO NOT FORGET: You are encouraged to collaborate on homeworks, explain things to each other, and test each other's knowledge. *But Do NOT* hand out answers to someone who has not done any work. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. **Write your own solutions!**

Problem 1 (20p)

Antibody testing is a serological test used to detect the presence of antibodies to SARS-CoV-2, the virus that causes COVID-19. A positive antibody test is presumed to mean that a person has been infected with the virus at some point in the past. Current research is now focused on associations between antibodies and loss of smell/taste. The dataset "Antobodies.csv" contains a subset of data collected at a medical center (Spring 2020) and has been used to investigate the IG-M antibody and how these values vary across age, gender and self-reported smell loss (normal/altered) for patients diagnosed* with COVID-19.

*Note: COVID infection was first detected using PCR testing.

1) Using the entire sample, provide descriptive statistics for all variables of interest.

a) Total sample size for each variable (N); Mean/SD, Median/IQR, Min and Max for continuous variables; Frequency/Percentages for categorical variables; Number of missing values for each variable. (8p)

b) Generate a histogram for the Ig-M values and comment on its shape. (2p)

2) Provide descriptive statistics for all variables, stratified by smell category.

a) Provide descriptive statistics for all the other variables by the two smell categories (normal vs altered) and present them in a tabular form (see example below). Briefly comment on the differences observed between the two groups. (6p)

Hint: tableby() in R can easily generate this, but feel free to create your own table in Word.

	Normal (N=XXX)	Altered (N=XXX)
AgeCategory 18-30 31-50 51+		
Gender Male Female		
Antibody_IgM		

b) Use the Ig-M variable to generate side-by-side histograms and boxplots by smell categories (normal vs altered). Make sure you label your figures appropriately and briefly discuss the trends observed. (4p)

Problem 2 (15p)

Find a story in the media (newspaper, online) that discusses the results of scientific research. If the same story was reported in several media outlets, comment on the differences re: how the facts were presented. Every (decent) reported article should have references of the original research study, ideally published in a peer-reviewed scientific journal.

After reading the original reference(s), provide a summary* including the following:

- Main goal(s) of the study
- Aspects of study design, i.e., randomized vs observational study, sample selection and sample size, potential biases, generalizability, etc. If you feel confident, you can also comment on the analysis, but it is not required for this assignment.
- Decide how seriously you should take the results of the research based on the evidence (or lack of it) supporting the findings.

*Note: Your summary, including your recommendations should **not exceed one page**. Link to the media post and journal article(s) should be listed as references at the end of the summary.