Alan Donahue

DSC 530 Final Term Project

Neerja Sahu

Date: 20 Nov 21

Looking at the heart disease dataset, my statistical question is can heart disease be predicted? I decided to focus on the following five variables: age, sex, type of chest pain, cholesterol, and resting blood pressure. I decided on those five variables because of all the ads that focus on heart disease and what causes it.

Going through the data, I found that the age data had a normal distribution. I also found that males comprised the majority of the study as well as asymptomatic chest pain was the major type of chest pain. The outcome of my exploratory data analysis was that resting blood pressure alone was not good enough to predict heart disease.

I felt that I should have started to look at the different variables together instead of separate towards the end of the data analysis process. It might be case that it is multiple variables that affect heart disease instead of just a single variable. Additionally, potentially focusing on variables like RestingECG, Oldpeak, and ST\_Slope may have been better because those were part of a test on the heart. I would assume that those could potentially predict heart disease better.

Looking back at the project, I think the only major assumption I made was in choosing my variables as I continued down my data analysis process. I don’t think that is an issue because it is in the exploratory phase. So, if need be, I can go back and try out different variables instead.

I faced a few challenges during this project. First off, I didn’t want to utilize the code from the book. I wanted to find all the solutions through the Python modules that are widely used throughout the industry. This grew to be a challenge because I haven’t used the modules before. I think if I stuck to using them during my homework instead of the book’s code then it would have been smoother. Additionally, the process of testing the hypothesis proved to be challenging for me. While I know that the process is four steps: choose test statistic, define null hypothesis, compute p-value, and interpret. The coding portion of it and choosing the test statistic was confusing.

# References

Hermann, L. K., Weingart, S. D., Yoon, Y. M., Genes, N. G., Nelson, B. P., Shearer, P. L., . . . Henzlova, M. J. (2010). Comparison of frequency of inducible myocardial ischemia in patients presenting to emergency department with typical versus atypical or nonanginal chest pain. *National Library of Medicine*, Abstract.

Huizen, J., & Soliman, M. (2021, September 21). *What is serum cholesterol?* Retrieved from Medical News Today: https://www.medicalnewstoday.com/articles/321519

*Non-Cardiac Chest Pain*. (2015, June 25). Retrieved from Cleveland Clinic: https://my.clevelandclinic.org/health/diseases/15851-gerd-non-cardiac-chest-pain

*Silent Heart Attacks: What Do Asymptomatic Signs of a Heart Attack Mean?* (n.d.). Retrieved from Elite Care: https://elitecarehouston.com/silent-heart-attacks-what-do-asymptomatic-signs-of-a-heart-attack-mean/