Homework 2

Introduction to Data Science (QR2)

due: **9/24/25**

Instructions: Please submit solutions on Populi. Only a knitted pdf of an `Rmarkdown` file will be accepted.

Problem 1: Refer to *The Art of Statistics* for the following questions.

a. In chapter 8 (pg. 222), the question is proposed: If we observe everything, where does probability come in? Provide your own answer and give a concrete example other than the book's Poisson example.

b. In chapter 8 (pg. 216), what is probability anyway? Pick two suggested answers and contrast/elaborate in your own words.

Problem 2: The data `predimed.csv` is from a randomized control trial (RCT) of diet variation among participants at risk for a cardiovascular event. The type of diet was randomly assigned, and a key outcome variable is whether or not the person had a cardiac event, denoted `event` in the data. There are two variants of the Mediterranean diet, one supplemented with nuts and the other supplemented with extra virgin olive oil (EVOO).[1]. We will use this data to estimate a variety of probabilities and assess the effectiveness of a Mediterranean diet.

a. Estimate the following probabilities: $P(\texttt{event})$, $P(\text{any } \texttt{MedDiet})$, $P(\texttt{event}, \text{any } \texttt{MedDiet})$, and $P(\texttt{event} \mid \text{any } \texttt{MedDiet})$.

b. Estimate $P(\texttt{event} \mid \texttt{Control})$. Using this result and the answer from the previous question, assess whether the Mediterranean diet has an effect on the chance of a cardiac event?

c. What are the effects on cardiac event likelihood of the Mediterranean diet on `Female` and `Male` subpopulations relative to the control diet?

Problem 3: Read about the crash and search for Air France Flight 447. Why did the initial search fail for so long, and how did probability accelerate the discovery of the wreckage (be as specific as possible)? Any formulas and figures are welcome!

[1] The original study is Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. *The New England Journal of Medicine*, 2013, 368:1279-1290.