Handout 1

1. Using relevant examples, distinguish between the following:

   - a) Population and Sample
   - b) Descriptive and Inferential Statistics
   - c) Categorical and Numerical Data
   - d) Systematic and Multistage sampling

2. Are the following variables categorical (qualitative) or numerical (quantitative)? In addition, state the level of measurement (in this case, nominal, ordinal, discrete and continuous) for each variable.

   - a) The average miles driven per week commuting to and from school

   - b) Income tax brackets in dollars

   - c) Marital status

   - d) A 2020 Gallup Poll question asked "In politics, as of today, do you consider yourself a Republican, Democrat or and Independent?" The possible responses were Democrat, Republican, Independent, Other and No response. What kind of variable is the response?

   - e) A pharmaceutical company conducts an experiment in which a subject takes 100 mg of a substance orally. The researchers measure how many minutes it takes for half of the substance to exit the bloodstream. What kind of variable is the company studying?

3. For each study below, **specify the type of study (observational or experimental), the type of bias (if present), and primary outcome/response**. For selected studies describe the population of interest and the data sample.

   - A restaurant invited a random sample of 50 recent customers to complete an online survey and rank their experience. Restaurant owners suspect that only customers who had really strong feelings about their experience (positive or negative) completed the survey, while customers with a neutral experience did not.

   - An organization also wants to know if drinking caffeinated coffee causes hyperactivity in college students. They select a random sample of college students and randomly assign them to drink coffee with or without caffeine. After an hour, the researchers test the students' concentration behavior.

4. Using the *Focus* dataset posted on Canvas, answer the following questions **using R software**,

   - a) How many participants were included in the study?
   - b) What does each row represent?
   - c) What is the mean and standard deviation of the variable GPA?
   - d) Create a frequency table for the variable CLASS?
   - e) Create a contingency table using the variables CLASS and SEX?
   - f) Considering the variables GPA and Age, what is the explanatory and response variable?
   - g) Construct a scatter-plot of GPA versus Age. What can you say about the graph?