ST 518: Homework 1

Submit a .pdf file that includes the answers to both sets of questions, and also submit your .R script that you used to answer the questions in the R part.

Please feel free to discuss questions on the discussion board.

Conceptual Questions

1. (4 points) For each of the following, determine whether the response variable numerical or categorical. If the response variable is categorical, is it binary? If it is not binary, list possible categories for the response variable.
   (a) In a survey, college students were asked how many hours per week they spend on the internet.
   (b) In a survey, college students were asked, "What percentage of the time that you spend on the internet is not for course work?"
   (c) In a survey, college students were asked, "What is your primary mode of transportation when traveling between campus and home?"
   (d) In a survey, college students were asked whether or not they live on campus.
   (e) In a survey, college students were asked how many of their meals they prepare at home per week.
   (f) In a survey, college students were asked, "Which of the five main food groups constitutes the majority of your diet?"
   (g) In a survey, smart phone users were asked whether or not they have used a web-based taxi service like Uber or Lyft.
   (h) In a survey, smart phone users were asked how many times they used a web-based taxi service in the past three months.

2. (3 points) On a multiple-choice exam, each of the 20 questions has 2 possible answers and only one correct response. Suppose, for each question, one student selects his responses completely at random.
   (a) Let $X_i,\ i = 1, ..., 20$ represent whether the student answered the ith question correctly, and let Y represent the total number of answers the student gets correct. What is the distribution of X_i? What is the distribution of Y?
   (b) What is the probability that the student passes the test (i.e., scores at least 70%)? (You will learn more about this distribution in the next module, but you should be able to calculate this in R using the `pbinom` function.)

3. (2 points) Do you prefer taking courses online, on campus, or a hybrid of the two? Suppose these preferences occur with probabilities (p_1, p_2, p_3). For $N = 3$ independent subjects, let the observed frequencies be (n_1, n_2, n_3). That is, we observe n_1 subjects who prefer taking courses online, etc.
   (a) Explain how you can determine n_3 from knowing n_1 and n_2.
   (b) List all the possible observations (n_1, n_2, n_3), with $n = 3$.

R Question

4. (1 point) Consider the data in **ex2117** of the **Sleuth3** package.
   (a) In what format are these data (case, tabular, frequency, other)? Please explain.
   (b) Create a new file for the data, called **HW1data**. In that file, create a NumBoys column that represents the number of births that were boys (be sure to round to the nearest integer).
   (c) Use an appropriate R command or commands to change the format to something else (you can pick what).