Problem Set 1

Instructions: Answer the following questions in their entirety in a separate document. Submit your completed problem set as a PDF document via email to `sean.corcoran@ vanderbilt.edu`. Use your last name and problem set number as the filename (e.g., *Fauci Problem Set 1.pdf*). Working together is encouraged, but it is expected that all submitted work be that of the individual student.

1. (**23 points**) The pro-school choice organization EdChoice conducts a regular opinion survey on matters related to school enrollment. You can find highlights from the July 2021 edition here: `https://edchoice.morningconsultintelligence.com/assets/127517.pdf` and detailed results at the following website: `https://edchoice.morningconsultintelligence.com/assets/127480.pdf`. Refer to this report to answer the questions below.

   (a) The EdChoice report contains many interesting statistics, including items related to support for school choice, teacher salaries, vaccination, and in-person schooling during COVID-19. Based on a quick skim of this report, would you say these are intended as descriptive or inferential statistics? Briefly explain your answer. (**3 points**)

   (b) Consider the data the authors used for this report. What was the unit of observation? Was the data a population or a sample? How were the data obtained? How many observations were in the data? (**4 points**)

   (c) Consider the following list of measures used in this report. For each, state whether the measure is: (1) quantitative or categorical; (2) nominal, ordinal, or interval/ratio scale. Base your answer on how they describe the outcome, and provide a brief explanation if necessary. (**6 points**)

      i. Feelings about whether K-12 education is generally going in the right direction.

      ii. How much trust one has in principals to make good decisions about education.

      iii. If given the option, what type of school parents would select for their child.

      iv. School parents' comfort with their children returning to school in person.

      v. Parents' opinion on homeschooling since COVID-19 began.

      vi. How much parents would be willing to spend on a monthly basis to participate in a "learning pod".

   (d) What is an example of a dichotomous variable used in this report? (**2 points**)

(e) Take a look at Table EC1_1 in the detailed report. Which subgroups had the most positive view of K-12 education? Which subgroups had the most negative view? Do you see any interesting differences in the responses by gender, race, income, or political party? (**4 points**)

(f) What is the population of interest in this study, the sampling frame, and sampling method? Briefly summarize. Note: more details on the report's methodology are provided here: `https://edchoice.morningconsultintelligence.com/methodology/` (**4 points**)

2. (**6 points**) A sociologist wants to estimate the average age at marriage for women in New England in the early 18th century. She finds within her state archives of marriage records for a large Puritan village for the years 1700-1730. She then takes a sample of those records, noting the age of the bride for each. The average age in these records is 24.1 years. Using a statistical method covered later, the sociologist estimates the average age of bridges at marriage for the population to be between 23.5 and 24.7 years.

(a) What part of this example is *descriptive?*

(b) What part of this example is *inferential?*

(c) To what population does the inference refer?

3. (**3 points**) Identify the following quantitative variables as either discrete or continuous and identify their most likely level of measurement (nominal, ordinal, interval, or ratio).

(a) Number of students enrolled at Vanderbilt University in any particular term

(b) Distance an individual can run in five minutes

(c) Hair length

(d) Number of hot dogs sold at baseball games

(e) Self-concept as measured by the degree of agreement with the statement "I feel good about myself," on a five-point scale where 1 represents strongly disagree, 2 represents disagree, 3 represents neutral, 4 represents agree, and 5 represents strongly agree

(f) Lack of coordination as measured by the length of time it takes an individual to assemble a certain puzzle.

4. (**10 points**) Visit the **amazing** website `www.socialexplorer.com`. Create a user account with your Vanderbilt email address so that you can have Basic access to this site. When prompted, you can affiliate with Vanderbilt's institutional subscription to get full functionality. From the main page, select "Explore Maps" and "Explore" under United States. On the next screen, select "Change Data." Use the Social Explorer site to find answers to the following questions.

(a) In the list of variables for 2019, identify one categorical variable and one quantitative variable. (Consider how the variable was measured in the original data collection, not how it is summarized at a higher level). (**2 points**)

(b) In 2019, what is the median household income in Tennessee? In the Nashville-Davidson-Murfreesboro-Franklin metropolitan statistical area (MSA)? Are these statistics based on a population or sample? Explain how you know. (**4 points**)

(c) In 2020, what percentage of voters in Tennessee voted for the Democratic candidate (Joseph Biden) in the presidential election? What percentage of voters in Davidson County (Nashville) voted for the Democratic candidate? Are these statistics based on a population or sample? Explain how you know. (**4 points**)