1. **Explain what the difference between descriptive statistics and inferential statistics.**
   The difference between the two methods is that descriptive statistics is limited to only the data collected and the objective is to present and summarize it. Whereas inferential statistics applies to a larger population than those observed, arriving to a conclusion about the whole population based on the sample.

2. **The 2010 General Social Survey asked the question, "After an average work day, about how many hours do you have to relax or pursue activities that you enjoy?" to a random sample of 1,155 Americans. The average relaxing time was found to be 1.45 hours. Determine which of the following is a sampling unit, a variable, a statistic, or a population parameter.**
   a. An American in the sample. - sampling unit
   b. Number of hours spent relaxing after an average work day. - variable
   c. 1.65 – statistic
   d. Average number of hours all Americans spend relaxing after an average work day. - parameter

3. **Consider a list of all current graduate seniors at American colleges and universities. A random sample of 10,000 current graduating seniors is selected from the list. After analyzing the data, the conclusion is that current graduating seniors at American colleges and universities take on average 7 math courses during their four years in school. This conclusion is an example of:**
   a. Descriptive statistics
   b. Exploratory data analysis
   c. Inferential statistics
   d. Randomized experiment
   e. None of the above

4. **Consider a list of all current graduate seniors at American colleges and universities. A random sample of 10,000 current graduating seniors is selected from the list. Based on the sample seven is the average number of math courses that they take during their four years in school.**

   a. Descriptive statistics
   b. Exploratory data analysis
   c. Inferential statistics
   d. Randomized experiment
   e. None of the above

5. Explain the difference between a survey and a census.

A survey is based on results from a (representative) sample, or subset, of the population, while a census is based on results from the entire population.

6. In what type of study would you generally know the value of the parameter?

a. Experiment
b. Census
c. Sample Survey
d. All of the Above

7. What is the name of a study that the purpose is to determine whether a treatment causes a change in the variable of study?

a. Observational Study
b. Randomized Experiment
c. Census
d. Sample Survey
e. More than one of the above

8. Lauren wants to determine what the average house price is for single family homes in Los Angeles, California. She randomly selects 50 single family homes in Los Angeles and looks up the most recent value for each of these 50 homes. What is the sample in this study?

a. All residents of Los Angeles, California
b. All single-family homeowners in Los Angeles, California
c. All single-family homes in Los Angeles, California
d. The 50 selected homes in Los Angeles, California
e. The homeowners of the 50 selected homes in Los Angeles, California

9. Which of the following is an example of a self-selected (volunteered) sample? More than one may apply.

a. Researcher splits their population into a group of a males and a separate group of females; she takes a random sample of 25 males and a separate random sample of 25 females
b. Radio asks people listening to call in and say whether or not they want a new stoplight at a busy intersection
c. Student selects 15 classmates to answer a survey about university fees
d. Researcher at a major university randomly selects 50 students to participate in a survey. She sends the survey to the 50 selected students via email and 45 of them return the survey. – this error is called no-response
e. More than one of the above

10. Consider the list below of different sampling methods. Which of these sampling methods don't provide a representative sample?

a. 1 only
b. 3 and 4
c. 1, 3, and 4
d. 2 and 5
e. 2, 3, 4, and 5

1. Simple Random Sample
2. Convenience Sample
3. Cluster Sample
4. Stratified Sample
5. Voluntary Response Sample

11. A study finds that an increase in glove sales is associated with an increase in snowboarding accidents. Does that mean that gloves cause snowboarding accidents? Explain your answer. What is one possible lurking variable that could explain this correlation?

A lurking variable is a variable that is neither an explanatory variable nor a response variable but influences either the explanatory variable and/or the response variable or, the relationship between them. In this question, the season of the year for snowboarding and skiing influences the increase in the sales of gloves and accidents. As there are lurking variables that are not addressed in the study, we cannot establish a cause-effect relationship.

12. A university wants to determine what fraction of its undergraduate student body support a new $25 annual fee to improve the student union. For each proposed method below, indicate whether the method is reasonable or not. Explain.

a. Survey a simple random sample of 500 students.
b. Stratify students by their field of study, then sample 10% of students from each stratum.
c. Cluster students by their ages (e.g. 18 years old in one cluster, 19 years old in one cluster, etc.), then randomly sample three clusters and survey all students in those clusters.
d. Design a randomized experiment.
e. None of the above

(a) Simple random sampling is okay. In fact, it's rare for simple random sampling to not be a reasonable sampling method! (b) The student opinions may vary by field of study, so stratifying by this variable makes sense and would be reasonable. (c) Students of similar ages are probably going to have more similar opinions, and we want clusters to be diverse with respect to the outcome of interest, so this would not be a good approach. (Additional thought: the clusters in this case may also have very different numbers of people, which can also create unexpected sample sizes.)

13. **In a study where researchers collected data to examine the relationship between air pollutants and pre-term births in Southern California, they measured the air pollution levels by air quality monitoring stations. Length of gestation data were collected on 143,196 births between the years 1989 and 1993, and air pollution exposure during gestation was calculated for each birth. They found that higher air pollution levels are associated with pre-term births.**

    a. What type of study is this?

    Observational Study

    b. Identify the population of interest and the sample in this study.

    Population – Pre-term births in Southern California

    Sample – 143,196

    c. Identify the explanatory and response variables.

    Explanatory – air pollution

    Response – Occurrence of a pre-term birth ("1" if yes, "0" if no)

    d. Could you identify a lurking variable in this study?

    Mother's health, age, sedentary lifestyle (smoking habits, alcohol consumption etc), diet, genetic history, type of dwelling, household income, education, etc.

    e. Can we conclude that higher air pollution levels cause pre-term births? (Establish causal relationships)

No, we cannot establish a causal link in this study because the study design (observational study) doesn't address the lurking variables that might affect the relationship between pollution and pre-term births.

    f. Comment on whether the results of the study can be generalized to the population.

No, the results of the study cannot be generalized to the population at large since the sample is not representative of the population.

14. **Buteyko shallow breathing technique reduces asthma symptoms and improves quality of life. As part of this study 600 asthma patients aged between 18-69 years who relied on medication for asthma treatment were recruited and randomly assigned to two groups: one practiced the Buteyko method and the other did not. Those in the Buteyko group experienced, on average, a significant reduction in asthma symptoms and an improvement in quality of life.**

    a. What type of study is this?
    b. Identify the population of interest and the sample in this study.
    c. What are the cases?
    d. What is (are) the response variable(s) in this study?
    What is (are) the explanatory variable(s) in this study?

f. Can we establish a causal link between the explanatory and response variables?

g. Can the results of the study be generalized to the population at large?

a. What type of study is this?
   Randomized experiment
b. Identify the population of interest and the sample in this study.
   Population – Asthma patients aged between 18-69 years who relied on medication
   Sample – the 600 asthma patients selected
c. What are the cases?
   Each of those 600 asthma patients recruited will be a case.
d. What is (are) the response variable(s) in this study?
   Response variable – Whether a case exhibits the symptoms of Asthma ("1" if yes, "0" if no)
e. What is (are) the explanatory variable(s) in this study?
   Explanatory variable – Practice of the Buteyko shallow breathing technique (Yes/No)
f. Can we establish a causal link between the explanatory and response variables?
   We can conclude that practicing Buteyko shallow breathing technique improves the quality of life in the study because the study is based on the random allocation of patients to the treatments.

g. Can the results of the study be generalized to the population at large?
No, the results of the study cannot be generalized to the population at large since the sample is not representative of the population.