



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 being summarized, whereas inferential statistics applies to a larger population than those observed

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 1.65 hours. Determine which of the following is an observation, a variable, a sample statistic, or a population parameter.

- An American in the sample.- **observation**
- Number of hours spent relaxing after an average work day. - **variable**
- 1.65 – **sample statistic**
- Average number of hours all Americans spend relaxing after an average work day. - **parameter**

3. 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.

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

- Experiment
- Census**
- Sample Survey
- All of the Above

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

- Observational Study
- Randomized Experiment**
- Census
- Sample Survey
- More than one of the above

6. 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?

- All residents of Los Angeles, California
- All single-family homeowners in Los Angeles, California
- 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

7. Which of the following is an example of a self-selected sample?

- 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 asks 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

8. “What is the average age of people in Salt Lake City?”

Population: All people living in Salt Lake City.

Sample: 10 people picked randomly from all people living in Salt Lake City.

Is the sample representative?

Yes

9. Consider the list below of different sampling methods. Which of these sampling methods are not biased sampling methods?

- | | |
|-------------------|------------------------------|
| a. 1 only | 1. Simple Random Sample |
| b. 3 and 4 | 2. Convenience Sample |
| c. 1, 3, and 4 | 3. Cluster Sample |
| d. 2 and 5 | 4. Stratified Sample |
| e. 2, 3, 4, and 5 | 5. Voluntary Response Sample |

10. 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.

- 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.



(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 the 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.)

11. In a study where researchers collected data to examine the relationship between air pollutants and preterm 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 preterm births.

a. What type of study is this?

Observational Study

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

Preterm births in Southern California

c. Identify the explanatory and response variables.

Explanatory – air pollution

Response – preterm birth

d. Can we conclude that higher air pollution levels cause preterm births? (Establish causal relationships)

No, we cannot establish a causal link between the explanatory and response variables since the study is observational.

e. Comment on whether or not 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 random.

12. Buteyko shallow breathing technique reduces asthma symptoms and improve quality of life. As part of this study 600 asthma patients aged 18-69 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?

Randomized experiment



- b. Identify the population of interest and the sample in this study.
Population – Asthma patients aged 18-69 who relied on medication
Sample – the 600 asthma patients selected
- c. What are the cases?
The data collected from the 600 asthma patients recruited
- d. What is (are) the response variable(s) in this study?
Response variable – Asthma symptoms
- e. What is (are) the explanatory variable(s) in this study?
Explanatory variable – Practice of the buteyko shallow breathing technique
- 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 assignation 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 random.

13. Jim is conducting a health survey of residents of the Brazos Valley. He asks participants their age, height, weight, type of insurance, marital status, income, number of people living in their household, and number of days they were sick in the last month. How many of his variables were qualitative and how many of his variables were quantitative?

- a. 2 qualitative and 6 quantitative**
- b. 3 qualitative and 5 quantitative
- c. 4 qualitative and 4 quantitative
- d. 5 qualitative and 3 quantitative
- e. 6 qualitative and 2 quantitative

14. A survey was conducted to study the smoking habits of UK residents. Below is a data matrix displaying a portion of the data collected in this survey. Note that “£” stands for British Pounds Sterling, “cig” stands for cigarettes, and “N/A” refers to a missing component of the data.

No	Sex	Age	Marital	Gross Income	Smoke	# Weekends	# Weekdays
1	Female	42	Single	£2,600	Yes	12 cig/day	12 cig/day
2	Male	44	Single	£12,400	No	N/A	N/A
3	Male	53	Married	£36,400	Yes	6 cig/day	6 cig/day



....							
....							
1691	Male	40	Single	£4,500	Yes	8 cig/day	8 cig/day

- What does each row of the data matrix represent?
Each observation in the sample
- How many participants were included in the survey?
n=1691
- Indicate whether each variable in the study is numerical or categorical. If numerical, identify as continuous or discrete. If categorical, indicate if the variable is ordinal.

15. A survey of a random sample of 100 nurses working at a large hospital asked how many years they had been working in the profession. Their answers are summarized in the following (incomplete) table. Fill in the blanks in the table and round your answers to two decimal places:

# of years	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency
<5	25	25	0.25	0.25
5-10	30	55	0.30	0.55
>10	100-55=45	100	.45	1

16. What proportion of nurses have five or more years of experience?

0.75

17. What proportion of nurses have ten or fewer years of experience?

0.55