

Introduction to Statistics

1.1 AN OVERVIEW OF STATISTICS

1.1 Try It Yourself Solutions

- 1a. The population consists of the prices per gallon of regular gasoline at all gasoline stations in the United States. The sample consists of the prices per gallon of regular gasoline at the 900 surveyed stations.
- b. The data set consists of the 900 prices.
- 2a. Because the numerical measure of \$2,655,395,194 is based on the entire collection of player's salaries, it is from a population.
- b. Because the numerical measure is a characteristic of a population, it is a parameter.
- 3a. Descriptive statistics involve the statement "76% of women and 60% of men had a physical examination within the previous year."
- b. An inference drawn from the study is that a higher percentage of women had a physical examination within the previous year.

1.1 EXERCISE SOLUTIONS

1. A sample is a subset of a population.
2. It is usually impractical (too expensive and time consuming) to obtain all the population data.
3. A parameter is a numerical description of a population characteristic. A statistic is a numerical description of a sample characteristic.
4. Descriptive statistics and inferential statistics
5. False. A statistic is a numerical measure that describes a sample characteristic.
6. True
7. True
8. False. Inferential statistics involves using a sample to draw conclusions about a population.
9. False. A population is the collection of *all* outcomes, responses, measurements, or counts that are of interest.
10. False. A statistic can differ from sample to sample.

- 11.** The data set is a population because it is a collection of the heights of all the players on a school's basketball team.
- 12.** The data set is a population because it is a collection of the energy collected from all the wind turbines on the wind farm.
- 13.** The data set is a sample because the collection of the 500 spectators is a subset within the population of the stadium's 42,000 spectators.
- 14.** The data set is a population because it is a collection of the annual salaries of all pharmacists at a pharmacy.
- 15.** Sample, because the collection of the 20 patients is a subset within the population
- 16.** The data set is a population since it is a collection of the number of televisions in all U.S. households.
- 17.** Population, because it is a collection of all golfers' scores in the tournament
- 18.** Sample, because only the age of every third person entering the clothing store is recorded
- 19.** Population, because it is a collection of all the U.S. presidents' political parties
- 20.** Sample, because the collection of the 10 soil contamination levels is a subset in the population
- 21.** Population: Party of registered voters in Warren County
Sample: Party of Warren County voters responding to online survey
- 22.** Population: All students who donate at a blood drive
Sample: The students who donate and have type O⁺ blood
- 23.** Population: Ages of adults in the United States who own cellular phones
Sample: Ages of adults in the United States who own Samsung cellular phones
- 24.** Population: Income of all homeowners in Texas
Sample: Income of homeowners in Texas with mortgages
- 25.** Population: Collection of all adults in the United States
Sample: Collection of 1000 adults surveyed
- 26.** Population: Collection of all infants in Italy
Sample: Collection of 33,043 infants in the study

- 27.** Population: Collection of all adults in the U.S.
Sample: Collection of 1442 adults surveyed
- 28.** Population: Collection of all people
Sample: Collection of 1600 people surveyed
- 29.** Population: Collection of all registered voters
Sample: Collection of 800 registered voters surveyed
- 30.** Population: Collection of all students at a college
Sample: Collection of 496 students surveyed
- 31.** Population: Collection of all women in the U.S.
Sample: Collection of the 546 U.S. women surveyed
- 32.** Population: Collection of all U.S. vacationers
Sample: Collection of the 791 vacationers surveyed
- 33.** Population: Collection of all Fortune magazine's top 100 companies to work for
Sample: Collection of the 85 companies who responded to the questionnaire
- 34.** Population: Collection of all light bulbs from the day's production
Sample: Collection of the 20 light bulbs selected from the day's production
- 35.** Statistic. The value \$68,000 is a numerical description of a sample of annual salaries.
- 36.** Statistic. 43% is a numerical description of a sample of high school students.
- 37.** Parameter. The 62 surviving passengers out of 97 total passengers is a numerical description of all of the passengers of the Hindenburg that survived.
- 38.** Parameter. 52% is a numerical description of the total number of governors.
- 39.** Statistic. 8% is a numerical description of a sample of computer users.
- 40.** Parameter. 12% is a numerical description of all new magazines.
- 41.** Statistic. 44% is a numerical description of a sample of all people.
- 42.** Parameter. 21.0 is a numerical description of ACT scores for all graduates.

43. The statement “56% are the primary investors in their household” is an application of descriptive statistics.

An inference drawn from the sample is that an association exists between U.S. women and being the primary investor in their household.

44. The statement “spending at least \$2000 for their next vacation” is an example of descriptive statistics.

An inference drawn from the sample is that United States vacationers are associated with spending more than \$2000 for their next vacation.

45. Answers will vary.

46. (a) The volunteers in the study represent the sample.

(b) The population is the collection of all individuals who completed the math test.

(c) The statement “three times more likely to answer correctly” is an application of descriptive statistics.

(d) An inference drawn from the sample is that individuals who are not sleep deprived will be three times more likely to answer math questions correctly than individuals who are sleep deprived.

47. (a) An inference drawn from the sample is that senior citizens who live in Florida have better memory than senior citizens who do not live in Florida.

(b) It implies that if you live in Florida, you will have better memory.

48. (a) An inference drawn from the sample is that the obesity rate among boys ages 2 to 19 is increasing.

(b) It implies the same trend will continue in future years.

49. Answers will vary.

1.2 DATA CLASSIFICATION

1.2 Try It Yourself Solutions

- 1a. One data set contains names of cities and the other contains city populations.

b. City: Nonnumerical
Population: Numerical

c. City: Qualitative
Population: Quantitative

- 2a.** (1) The final standings represent a ranking of basketball teams.
- (2) The collection of phone numbers represents labels. No mathematical computations can be made.
- b.** (1) Ordinal, because the data can be put in order
- (2) Nominal, because you cannot make calculations on the data
- 3a.** (1) The data set is the collection of body temperatures.
- (2) The data set is the collection of heart rates.
- b.** (1) Interval, because the data can be ordered and meaningful differences can be calculated, but it does not make sense writing a ratio using the temperatures
- (2) Ratio, because the data can be ordered, can be written as a ratio, you can calculate meaningful differences, and the data set contains an inherent zero

1.2 EXERCISE SOLUTIONS

1. Nominal and ordinal
2. Ordinal, interval, and ratio
3. False. Data at the ordinal level can be qualitative or quantitative.
4. False. For data at the interval level, you can calculate meaningful differences between data entries. You cannot calculate meaningful differences at the nominal or ordinal level.
5. False. More types of calculations can be performed with data at the interval level than with data at the nominal level.
6. False. Data at the ratio level can be placed in a meaningful order.
7. Qualitative, because telephone numbers are merely labels
8. Quantitative, because the heights of hot air balloons are a numerical measure
9. Quantitative, because the body temperatures of patients is a numerical measure.
10. Qualitative, because the eye colors are merely labels
11. Quantitative, because the lengths of songs on an MP3 player are numerical measures
12. Quantitative, because the carrying capacities of pickups are numerical measures
13. Qualitative, because the player numbers are merely labels
14. Qualitative, because student ID numbers are merely labels

15. Quantitative, because weights of infants are a numerical measure
16. Qualitative, because species of trees are merely labels
17. Qualitative, because the poll results are merely responses
18. Quantitative, because wait times at a grocery store are a numerical measure
19. Qualitative. Ordinal. Data can be arranged in order, but differences between data entries make no sense.
20. Qualitative. Nominal. No mathematical computations can be made and data are categorized using names.
21. Qualitative. Nominal. No mathematical computations can be made and data are categorized using names.
22. Quantitative. Ratio. A ratio of two data values can be formed so one data value can be expressed as a multiple of another.
23. Qualitative. Ordinal. The data can be arranged in order, but differences between data entries are not meaningful.
24. Quantitative. Ratio. The ratio of two data values can be formed so one data value can be expressed as a multiple of another.
25. Ordinal
26. Ratio
27. Nominal
28. Ratio
29. (a) Interval (b) Nominal (c) Ratio (d) Ordinal
30. (a) Interval (b) Nominal (c) Interval (d) Ratio
31. An inherent zero is a zero that implies “none.” Answers will vary.
32. Answers will vary.

1.3 DATA COLLECTION AND EXPERIMENTAL DESIGN

1.3 Try It Yourself Solutions

- 1a. (1) Focus: Effect of exercise on relieving depression
- (2) Focus: Success of graduates

- b.** (1) Population: Collection of all people with depression
 - (2) Population: Collection of all university graduates at this large university
- c.** (1) Experiment
 - (2) Survey
- 2a.** There is no way to tell why people quit smoking. They could have quit smoking either from the gum or from watching the DVD.
 - b.** Two experiments could be done; one using the gum and the other using the DVD.
- 3a.** Example: start with the first digits 92630782 ...
 - b.** 92 | 63 | 07 | 82 | 40 | 19 | 26
 - c.** 63, 7, 40, 19, 26
- 4a.** (1) The sample was selected by only using the students in a randomly chosen class. Cluster sampling
 - (2) The sample was selected by numbering each student in the school, randomly choosing a starting number, and selecting students at regular intervals from the starting number. Systematic sampling
- b.** (1) The sample may be biased because some classes may be more familiar with stem cell research than other classes and have stronger opinions.
 - (2) The sample may be biased if there is any regularly occurring pattern in the data.

1.3 EXERCISE SOLUTIONS

1. In an experiment, a treatment is applied to part of a population and responses are observed. In an observational study, a researcher measures characteristics of interest of part of a population but does not change existing conditions.
2. A census includes the entire population; a sample includes only a portion of the population.
3. In a random sample, every member of the population has an equal chance of being selected. In a simple random sample, every possible sample of the same size has an equal chance of being selected.
4. Replication is the repetition of an experiment using a large group of subjects. It is important because it gives validity to the results.
5. True
6. False. A double-blind experiment is used to decrease the placebo effect.

7. False. Using stratified sampling guarantees that members of each group within a population will be sampled.
8. False. A census is a count of an entire population.
9. False. To select a systematic sample, a population is ordered in some way and then members of the population are selected at regular intervals.
10. True
11. Use a census because all the patients are accessible and the number of patients is not too large.
12. Perform an observational study because you want to observe and record motorcycle helmet usage.
13. In this study, you want to measure the effect of a treatment (using a fat substitute) on the human digestive system. So, you would want to perform an experiment.
14. It would be nearly impossible to ask every customer whether he or she would still buy a product with a warning label. So, you should use a survey to collect these data.
15. Because it is impractical to create this situation, you would want to use a simulation.
16. Perform an observational study because you want to observe and record how often people wash their hands in public restrooms.
17. (a) The experimental units are the 30–35 year old females being given the treatment. One treatment is used.

(b) A problem with the design is that there may be some bias on the part of the researchers if he or she knows which patients were given the real drug. A way to eliminate this problem would be to make the study into a double-blind experiment.

(c) The study would be a double-blind study if the researcher did not know which patients received the real drug or the placebo.
18. (a) The experimental units are the 80 people with early signs of arthritis. One treatment is used.

(b) A problem with the design is that the sample size is small. The experiment could be replicated to increase validity.

(c) In a placebo-controlled double-blind experiment, neither the subject nor the experimenter knows whether the subject is receiving a treatment or a placebo. The experimenter is informed after all the data have been collected.

(d) The group could be randomly split into 20 males or 20 females in each treatment group.
19. Each U.S. telephone number has an equal chance of being dialed and all samples of 1400 phone numbers have an equal chance of being selected, so this is a simple random sample. Telephone sampling only samples those individuals who have telephones, are available, and are willing to respond, so this is a possible source of bias.

20. Because the persons are divided into strata (rural and urban), and a sample is selected from each stratum, this is a stratified sample.
21. Because the students were chosen due to their convenience of location (leaving the library), this is a convenience sample. Bias may enter into the sample because the students sampled may not be representative of the population of students. For example, there may be an association between time spent at the library and drinking habits.
22. Because the disaster area was divided into grids and thirty grids were then entirely selected, this is a cluster sample. Certain grids may have been much more severely damaged than others, so this is a possible source of bias.
23. Simple random sampling is used because each customer has an equal chance of being contacted, and all samples of 580 customers have an equal chance of being selected.
24. Systematic sampling is used because every tenth person entering the shopping mall is sampled. It is possible for bias to enter the sample if, for some reason, there is a regular pattern to people entering the shopping mall.
25. Because a sample is taken from each one-acre subplot (stratum), this is a stratified sample.
26. Each telephone has an equal chance of being dialed and all samples of 1012 phone numbers have an equal chance of being selected, so this is a simple random sample. Telephone sampling only samples those individuals who have telephones and are willing to respond, so this is a possible source of bias.
27. Answers will vary.
28. Answers will vary.
29. Answers will vary.
30. Answers will vary.
31. Census, because it is relatively easy to obtain the ages of the 115 residents
32. Sampling, because the population of subscribers is too large to easily record their favorite movie type. Random sampling would be advised since it would be too easy to randomly select subscribers then record their favorite movie type.
33. Question is biased because it already suggests that eating whole-grain foods is good for you. The question might be rewritten as “How does eating whole-grain foods affect your health?”
34. Question is biased because it already suggests that text messaging while driving increases the risk of a crash. The question might be rewritten as “Does text messaging while driving increase the risk of a crash?”
35. Question is unbiased because it does not imply how much exercise is good or bad.

- 36.** Question is biased because it already suggests that the media has a negative effect on teen girls' dieting habits. The question might be rewritten as "Do you think the media has an effect on teen girls' dieting habits?"
- 37.** The households sampled represent various locations, ethnic groups, and income brackets. Each of these variables is considered a stratum.
- 38.** Stratified sampling ensures that each segment of the population is represented.
- (a) Answers will vary.
 - (b) Answers will vary.
 - (c) Answers will vary.
 - (d) Answers will vary.
- 39.** Observational studies may be referred to as natural experiments because they involve observing naturally occurring events that are not influenced by the study.
- 40.** Open Question
Advantage: Allows respondent to express some depth and shades of meaning in the answer
Disadvantage: Not easily quantified and difficult to compare surveys
- Closed Question
Advantage: Easy to analyze results
Disadvantage: May not provide appropriate alternatives and may influence the opinion of the respondent
- 41.** (a) Advantage: Usually results in a savings in the survey cost
- (b) Disadvantage: There tends to be a lower response rate and this can introduce a bias into the sample.
Sampling Technique: Convenience sampling
- 42.** Answers will vary.
- 43.** If blinding is not used, then the placebo effect is more likely to occur.
- 44.** The Hawthorne effect occurs when a subject changes behavior because he or she is in an experiment. However, the placebo effect occurs when a subject reacts favorably to a placebo he or she has been given.
- 45.** Both a randomized block design and a stratified sample split their members into groups based on similar characteristics.

CHAPTER 1 REVIEW EXERCISE SOLUTIONS

1. Population: Collection of all U.S. adults

Sample: Collection of the 1000 U.S. adults that were sampled

2. Population: Collection of all nurses in San Francisco area

Sample: Collection of 38 nurses in San Francisco area that were sampled

3. Population: Collection of all credit cards

Sample: Collection of 39 credit cards that were sampled

4. Population: Collection of all physicians in the U.S.

Sample: Collection of 1205 physicians that were sampled

5. The team payroll is a parameter since it is a numerical description of a population (entire baseball team) characteristic.

6. Since 42% is describing a characteristic of the sample, this is a statistic.

7. Since “10 students” is describing a characteristic of a population of math majors, it is a parameter.

8. Since 50% is describing a characteristic of a sample of U.S. adults who say they oppose drilling for oil and gas in the Arctic National Wildlife Refuge, this is a statistic.

9. The average APR of 12.83% charged by credit cards is representative of the descriptive branch of statistics. An inference drawn from the sample is that all credit cards charge an APR of 12.83%.

10. 60% of all physicians surveyed consider leaving the practice of medicine because they are discouraged over the state of U.S. healthcare is representative of the descriptive branch of statistics. An inference drawn from the sample is that 60% of all physicians surveyed consider leaving the practice of medicine because they are discouraged over the state of U.S. healthcare.

11. Quantitative, because monthly salaries are numerical measurements

12. Qualitative, because Social Security numbers are merely labels for employees

13. Quantitative, because ages are numerical measurements

14. Qualitative, because zip codes are merely labels for the customers

15. Quantitative, because revenues are numerical measures

16. Qualitative, because marital statuses are attributes

17. Interval. It makes no sense saying that 100 degrees is twice as hot as 50 degrees.

- 18.** Ordinal. The data are qualitative but could be arranged in order of severity.
- 19.** Nominal. The data are qualitative and cannot be arranged in a meaningful order.
- 20.** Ratio. The data are numerical, and it makes sense saying that one amount is twice as large as another amount.
- 21.** Because CEOs keep accurate records of charitable donations, you could take a census.
- 22.** Because it is impractical to create this situation, you would want to perform a simulation.
- 23.** Perform an experiment because you want to measure the effect of training from animal shelters on inmates.
- 24.** Because it would be nearly impossible to ask every college professor about teaching classes online, you should take a survey to collect the data.
- 25.** The subjects could be split into male and female and then be randomly assigned to each of the five treatment groups.
- 26.** Number the volunteers and then use a random number generator to assign subjects to one of the treatment groups or the control group.
- 27.** Answers will vary.
- 28.** Sample. Take a survey, because asking all students at the university about their favorite spring break destinations would be nearly impossible.
- 29.** Because random telephone numbers were generated and called, this is a simple random sample.
- 30.** Because the student sampled a convenient group of friends, this is a convenience sample.
- 31.** Because each community is considered a cluster and every pregnant woman in a selected community is surveyed, this is a cluster sample.
- 32.** Because every third car is stopped, this is a systematic sample.
- 33.** Because grade levels are considered strata and 25 students are sampled from each stratum, this is a stratified sample.
- 34.** Because of the convenience of surveying people waiting for their baggage, this is a convenience sample.
- 35.** Telephone sampling only samples individuals who have telephones, are available, and are willing to respond.
- 36.** Due to the convenience sample taken, the study may be biased toward the opinions of the student's friends.
- 37.** The selected communities may not be representative of the entire area.

38. It may be difficult for the law enforcement official to stop every third car.

CHAPTER 1 QUIZ SOLUTIONS

1. Population: Collection of all men

Sample: Collection of 20,000 men in study

2. (a) Statistic. 19% is a characteristic of a sample of Internet users.

(b) Parameter. 90% is a characteristic of the entire Board of Trustees (population).

(c) Statistic. 55% is a characteristic of a sample of chief financial officers and senior comptrollers.

3. (a) Qualitative, since debit card pin numbers are merely labels.

(b) Quantitative, since a final score is a numerical measure.

4. (a) Ordinal. Badge numbers may be ordered numerically according to seniority of service, but no mathematical computations can be made.

(b) Ratio. It makes sense to say that the horsepower of one car was twice as many as another car.

(c) Ordinal, because data can be arranged in order but the differences between data entries make no sense

(d) Interval, because meaningful differences between entries can be calculated, but a zero entry is not an inherent zero.

5. (a) In this study, you want to measure the effect of a treatment (low dietary intake of vitamin C and iron) on lead levels in adults. You want to perform an experiment.

(b) Because it would be difficult to survey every individual within 500 miles of your home, sampling should be used.

6. Randomized Block Design

7. (a) Because people were chosen due to their convenience of location (on the campground), this is a convenience sample.

(b) Because every tenth part is selected from an assembly line, this is a systematic sample.

(c) Stratified sample because the population is first stratified and then a sample is collected from each stratum

8. Convenience

