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 800 surveyed stations.

b. The data set consists of the 800 prices.

2a. Because the numerical measure of $5,150,694 is based on the entire collection of employee’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 “31% support their kids financially until they graduate college and 6% provide financial support until they start college.”

b. An inference drawn from the survey is that a higher percentage of parents support their kids financially until they graduate college.

1.1 EXERCISE SOLUTIONS

1. A sample is a subset of a population.

2. It is usually impractical (too expensive and/or 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. The two main branches of statistics are 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 sample statistic can differ from sample to sample.
11. The data set is a population because it is a collection of the revenue of each of the 30 companies in the Dow Jones Industrial Average.

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. The data set is a sample because the collection of the 20 patients is a subset of the population of 100 patients at the hospital.

16. The data set is a population because it is a collection of the number of televisions in all U.S. households.

17. The data set is a population because it is a collection of all the golfers’ scores in the tournament.

18. The data set is a sample because only the age of every third person entering the clothing store is recorded.

19. The data set is a population because it is a collection of all the U.S. presidents’ political parties.

20. The data set is a sample because the collection of the 10 soil contamination levels is a subset of the population.


       Sample: Parties 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


24. Population: Incomes of all homeowners in Texas Sample:

       Incomes of homeowners in Texas with mortgages

25. Population: Collection of the responses of all adults in the United States Sample:

       Collection of the responses of the 1015 U.S. adults surveyed

26. Population: Collection of the heart rhythms of all infants in Italy

       Sample: Collection of the heart rhythms of the 33,043 infants in Italy in the study
27. Population: Collection of the immunization status of all adults in the U.S.
   Sample: Collection of the immunization status of the 12,082 U.S. adults surveyed

28. Population: Collection of the factors for choosing a hotel of all adults in the United States
   Sample: Collection of the factors for choosing a hotel of the 1012 U.S. adults surveyed

29. Population: Collection of the average billing rates of all U.S. law firms
   Sample: Collection of the average billing rates of the 55 U.S. law firms surveyed

30. Population: Collection of the travel plans of all students at a college
    Sample: Collection of the travel plans of the 496 students surveyed at a college

31. Population: Collection of the effect of sleepiness on all pilots
    Sample: Collection of the effect of sleepiness on the 202 pilots surveyed

32. Population: Collection of the responses of all major-appliance shoppers
    Sample: Collection of the responses of the 961 major-appliance shoppers surveyed

33. Population: Collection of the starting salaries at all 500 companies listed in the Standard & Poor’s 500
    Sample: Collection of the starting salaries at the 65 companies listed in the Standard & Poor’s 500 that were contacted by the researcher

34. Population: Collection of the times spent per day to using entertainment media by all third- to twelfth-grade students
    Sample: Collection of the times spent per day to using entertainment media by the 2002 third- to twelfth-grade students surveyed

35. Statistic. The value $68,000 is a numerical description of a sample of annual salaries.

36. Statistic. The value 38% is a numerical description of a sample of college board members.

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. The value 60% is a numerical description of the total number of governors.

39. Statistic. The value 8% is a numerical description of a sample of computer users.

40. Parameter. The value 78% is a numerical description of all voters in a county.

41. Statistic. The value 52% is a numerical description of a sample of U.S. adults.
42. Parameter. The score 21.1 is a numerical description of ACT scores for all graduates.

43. The statement “20% admit that they have made a serious error due to sleepiness” is an example of descriptive statistics.

   An inference drawn from the sample is that an association exists between sleepiness and pilot error.

44. The statement “23% bought extended warranties” is an example of descriptive statistics.

   An inference drawn from the sample is that most major-appliance shoppers do not buy extended warranties.

45. Answers will vary.

46. (a) The sample is the responses of the volunteers in the study.

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

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

   (d) An inference drawn from the sample is that individuals who are not sleep deprived will be 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 memories 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) The inference may incorrectly imply that the 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 names: Nonnumerical City
      Populations: Numerical

   c. City names: Qualitative
      City Populations: Quantitative
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, the data can be written as a ratio, meaningful differences can be calculated, 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 levels.

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. Quantitative, because heights of hot air balloons are numerical measurements.

8. Quantitative, because carrying capacities of pickups are numerical measurements.

9. Qualitative, because the colors are attributes.

10. Qualitative, because student ID numbers are labels.

11. Quantitative, because weights of infants are numerical measurements.

12. Qualitative, because species of trees are labels.

13. Qualitative, because the poll responses are attributes.

14. Quantitative, because wait times at a grocery store are numerical measurements.
15. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say one year is a multiple of another.

16. Ordinal. Data can be arranged in order, but differences between data entries are not meaningful.

17. Nominal. No mathematical computations can be made and data are categorized using numbers.

18. Ratio. Data can be ordered and meaningful differences can be calculated. A length of 0 means it lasts for 0 minutes. A ratio of two data entries can be formed so that one data entry can be meaningfully expressed as a multiple of another.

19. Ordinal. Data can be arranged in order, or ranked, but differences between data entries are not meaningful.

20. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say one time is a multiple of another.

21. Horizontal: Ordinal; Vertical: Ratio

22. Horizontal: Ordinal; Vertical: Ratio

23. Horizontal: Nominal; Vertical: Ratio

24. Horizontal: Interval; Vertical: Ratio

25. (a) Interval  (b) Nominal  (c) Ratio  (d) Ordinal

26. (a) Interval  (b) Nominal  (c) Interval  (d) Ratio

27. Qualitative. Ordinal. Data can be arranged in order, but differences between data entries are not meaningful.

28. Qualitative. Nominal. No mathematical computations can be made, and data are categorized by political party.

29. Qualitative. Nominal. No mathematical computations can be made and data are categorized using names.

30. Quantitative. Interval. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say one score is a multiple of another.

31. Qualitative. Ordinal. Data can be arranged in order, but differences between data entries are not meaningful.

32. Quantitative. Ratio. Data can be ordered and meaningful differences can be calculated. A ratio of two data values can be formed so that one data entry can be meaningfully expressed as a multiple of another.

33. An inherent zero is a zero that implies “none.” Answers will vary.

34. Answers will vary.
1.3 DATA COLLECTION AND EXPERIMENTAL DESIGN

1.3 Try It Yourself Solutions

1a. The study does not apply a treatment to the elk.

b. This is an observational study.

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. The gum and the DVD could be confounding variables.

b. Two experiments could be done; one using the gum and the other using the DVD. Or just conduct one experiment using either the gum or the DVD.

3. Sample answers:

a. 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 using the students in a randomly chosen class. This sampling technique is 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. This sampling technique is 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 a part of a population but does not change existing conditions.

2. A census includes the entire population; a sampling 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 under the same or similar conditions. Replication is important because it enhances the validity of the results.
5. False. A placebo is a fake treatment.
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. Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.
12. Experiment. The study applies a treatment (2000 milligrams per day of acetyl-L-carnitine) to the subjects.
13. Experiment. The study applies a treatment (different genres of music) to the subjects.
14. Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.
15. (a) The experimental units are the 250 females ages 30-35 in the study. The treatment is the new allergy drug.
   (b) A problem with the design is that there may be some bias on the part of the researcher if the researcher 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.
16. (a) The experimental units are the 80 people with early signs of arthritis. The treatment is the experimental sneaker.
   (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 and 20 females in each treatment group.
17. Answers will vary. Sample answer: Starting at the left-most number in row 6:
    28/70/35/17/09/94/45/64/83/96/73/78/
    The numbers would be 28, 70, 35, 17, 9, 94, 45, 64, 83, 96, 73, 78.
18. Answers will vary. **Sample answer:** Starting with the left-most number in row 10:

421/030/278/173/920/562/977/267/812/249/252/

The numbers would be 421, 30, 278, 173, 920, 562, 267, 812, 249, 252.

19. Answers will vary.

20. Answers will vary.

21. Answers will vary. **Sample answer:** Number the volunteers from 1 to 18. Using the random number table in Appendix B, starting with the left-most number in row 16:

29/55/31/84/32/13/63/00/55/29/02/79/18/10/17/49/02/77/90/31/50/91/20/93/99

Treatment group: Maria, Adam, Bridget, Carlos, Susan, Rick, Dan, Mary, and Connie.

Control group: Jake, Mike, Lucy, Ron, Steve, Vanessa, Kate, Pete, and Judy.

22. Answers will vary. **Sample answer:** Using a random number generator:

Treatment group: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 16, 17, 19, 21, 25, 29, 39, 40, 43, 45, 46, 47, 48, 49, 51, 52, 53, 55, 56, 57, 58, 59, 60, 61, 62, 64, 65, 66, 67, 69, 71, 72, 75, 76, 77, 79, 83, 84, 90

Control group: 8, 10, 11, 13, 14, 16, 17, 19, 21, 24, 25, 29, 39, 40, 43, 45, 46, 47, 48, 49, 51, 52, 53, 55, 56, 57, 58, 59, 60, 61, 62, 64, 65, 66, 67, 69, 71, 72, 75, 76, 77, 79, 83, 84, 90

23. Simple random sampling is used because each telephone number has an equal chance of being dialed, and all samples of 1400 phone numbers have an equal chance of being selected. The sample may be biased because telephone sampling only samples those individuals who have telephones, who are available, and who are willing to respond.

24. Stratified sampling is used because the persons are divided into strata (rural and urban), and a random sample is selected from each stratum.

25. Convenience sampling is used because the students are chosen due to their convenience of location. 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.

26. Cluster sampling is used because the disaster area is divided into grids, and 30 grids are then entirely selected. A possible source of bias is that certain grids may have been much more severely damaged than others.

27. 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.

28. 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.

29. Stratified sampling is used because a sample is taken from each one-acre subplot (stratum).

30. Simple random sampling is used because each telephone number has an equal chance of being dialed, and all samples of 1012 phone numbers have an equal chance of being selected. The sample may be biased because telephone sampling only samples those individuals who have telephones, who are available, and who are willing to respond.
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 because it would be easy to randomly select subscribers and then record their favorite movie types.

33. The question is biased because it already suggests that eating whole-grain foods improves your health. The question might be rewritten as “How does eating whole-grain foods affect your health?”

34. The 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 affect the risk of a crash?”

35. The survey question is unbiased because it does not imply how much exercise is good or bad.

36. The question is biased because it already suggests that the media have a negative effect on teen girls’ dieting habits. The question might be rewritten as “Do you think the media have 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. Stratified sampling ensures that each segment of the population is represented.

38. Sample answer: Observational studies may be referred to as natural experiments because they involve observing naturally occurring events that are not influenced by the study.

39. Open Question
   Advantage: Allows respondent to express some depth and shades of meaning in the answer. Allows for new solutions to be introduced.
   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.

40. (a) Advantage: Usually results in a savings in the survey cost.
   Disadvantage: There tends to be a lower response rate and this can introduce a bias into the sample. Only people with strong feelings might respond.
   (b) Sampling technique: Convenience sampling.

41. Answers will vary.