CHAPTER 1

Introduction to Statistics

1.1 AN OVERVIEW OF STATISTICS

1.1 TRY IT YOURSELF SOLUTIONS

1. The population consists of the responses of all ninth to twelfth graders in the United States. The sample consists of the responses of the 1501 ninth to twelfth graders in the survey. The sample data set consists of 1215 ninth to twelfth graders who said leaders today are more concerned with their own agenda than with achieving the overall goals of the organization they serve and 286 ninth to twelfth graders who did not say that.

2a. Population parameter, because the total spent on employees' salaries, $5,150,694, is based on the entire company.

b. Sample statistic, because 43% is based on a subset of the population.

3a. The population consists of the responses of all U.S. adults, and the sample consists of the responses of the 1000 U.S. adults in the study.

b. The part of this study that represents the descriptive branch of statistics involves the statement "three out of four adults will consult with their physician or pharmacist and only 8% visit a medication-specific website [when they have a question about their medication]."

c. A possible inference drawn from the study is that most adults consult with their physician or pharmacist when they have a question about their medication.

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.

Copyright © 2019 Pearson Education Ltd.

10. False. A sample statistic can differ from sample to sample.

11. Sample, because the collection of 95 shopkeepers is a subset within the population of the commercial complex's 550 shopkeepers.

12. Population, because it is a collection of the energy collected from all the solar panels on a photo voltaic power plant.

13. Population, because it is a collection of the height of each athlete participating in the Summer Olympics.

14. Sample, because the collection of every sixth person is a subset within the population of the departmental store's shoppers.

15. Sample, because the collection of the 10 patients is a subset of the population of 50 patients at the clinic.

16. Sample, because the collection of 25 households is a subset within the population of the neighborhood's 75 households.

17. Population, because it is a collection of all the gamers' scores in the tournament.

18. Population, because it is a collection of the ages of all presidents at the time of their election.

19. Sample, because the collection of top 10 taxpayers is a subset within the population of the country's total tax payers.

20. Sample, because the collection of the 20 air contamination levels is a subset of the population.

21. Population: Parties of registered voters
Sample: Parties of registered voters who respond to a survey

22. Population: Student donations at a food drive
Sample: Student donations of canned goods

23. Population: Ages of adults in the United States who own automobiles
Sample: Ages of adults in the United States who own Honda automobiles

24. Population: Incomes of home owners in Massachusetts
Sample: Incomes of home owners in Massachusetts with mortgages

25. Population: Collections of the responses of all U.S. adults
Sample: Collection of the responses of the 1020 U.S. adults surveyed
Sample data set: 42% of adults who said they trust their political leaders and 58% who said they did not

26. Population: Collection of fetal tobacco exposure of all infants
Sample: Collection of the fetal tobacco exposure of 203 infants
Sample data set: Infants with fetal tobacco exposure and their focused attention levels

27. Population: Collection of the influenza immunization status of all adults in the United States
Sample: Collection of the influenza immunization status of the 3301 U.S. adults surveyed
Sample data set: 39% of U.S. adults who received an influenza vaccine and 61% who did not

Copyright © 2019 Pearson Education Ltd.

28. Population: Collection of the responses of employees working in a foreign country
Sample: Collection of the responses of the 1500 employees who have settled in foreign countries
Sample data set: 62% of respondents said that they settle in the country where they work and 38% who said they did not

29. Population: Collection of the average hourly billing rates of all U.S. law firms
Sample: Collection of the average hourly billing rates for partners of the 159 U.S. law firms surveyed
Sample data set: The average hourly billing rate for partners of 159 U.S. law firms is $604.

30. Population: Collection of the responses of children
Sample: Collection of the responses of the 328 children living in a locality
Sample data set: 86% of respondents who said that they planned to visit their grandparents in vacations and 14% who said they did not

31. Population: Collection of blood donations collected globally
Sample: Collection of the 112.5 million blood donations collected globally
Sample data set: 50% of the donors who belong to high-income countries and 50% who do not

32. Population: Collection of the responses of laptop users
Sample: Collection of the responses of the 1468 laptop users
Sample data set: 81% of respondents who said that they preferred the use of mouse over touchpad and 19% who said they did not.

33. Population: Collection of the 1000 mutual funds listed on a recognized stock exchange
Sample: Collection of the 134 mutual funds of the 1000 mutual funds listed on a recognized stock exchange
Sample data set: Best mutual funds out of the 134 mutual funds listed on a recognized stock exchange

34. Population: Collection of parents of 13- to 17-year-olds
Sample: Collection of responses of 1060 parents of 13- to 17-year-olds surveyed
Sample data set: 636 parents who said they check their teen's social media profile and 424 parents who did not

35. Population Parameter. Forty out of 500 total students is a numerical description of the students who received a Grade C.

36. Sample statistic. The value 56.3% is a numerical description of a sample of college board members

37. Sample statistic. The value two million is a numerical description of a sample of civilian casualties during World War II.

38. Population parameter. The value 62% is a numerical description of the total number of governors.

39. Population Parameter. The entire population of employees working in the organization has been reviewed.

40. Sample statistic. The value 16% is a numerical description of a sample of teachers.

41. Sample statistic. The value 80% is a numerical description of a sample of U.S. adults.

42. Population parameter. The score 20.6 is a numerical description of the ACT scores for all graduates.

Copyright © 2019 Pearson Education Ltd.

43. The statement "50% are collected from high-income countries" is an example of descriptive statistics. Using inferential statistics, you may conclude that an association exists between income and the number of blood donations in a country.

44. The statement "81% preferred use of mouse over touchpad" is an example of descriptive statistics. Using inferential statistics, you may conclude that most laptop users prefer use of mouse over touchpad.

45. Answers will vary.

46. Answers will vary.

47. The inference may incorrectly imply that exercise increases a person's cognitive ability. The study shows a slower decline in cognitive ability, not an increase.

48. The inference may incorrectly imply that obesity trends will continue in future years. Even though the obesity rates have been increasing, that does not mean the rates will continue to increase for eternity.

49. (a) The sample is the results on the standardized test by the participants in the study.

(b) The population is the collection of all the results of the standardized test.

(c) The statement "the closer that participants were to an optimal sleep duration target, the better they performed on a standardized test" is an example of descriptive statistics.

(d) Individuals who obtain optimal sleep will be more likely to perform better on a standardized test then they would without optimal sleep.

1.2 DATA CLASSIFICATION

1.2 TRY IT YOURSELF SOLUTIONS

1. The city names are nonnumerical entries, so these are qualitative data. The city populations are numerical entries, so these are quantitative data.

2. (1) Ordinal, because the data can be put in order.

(2) Nominal, because no mathematical computations can be made.

3. (1) Interval, because the data can be ordered and meaningful differences can be calculated, but it does not make sense to write a ratio using the temperatures.

(2) Ratio, because the data can be ordered, meaningful differences can be calculated, the data can be written as a ratio, and the data set contains an inherent zero.

Copyright © 2019 Pearson Education Ltd.

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. Qualitative, because breeds of horses are attributes.

8. Qualitative, because ASCII codes are labels.

9. Quantitative, because blood pressure levels are numerical measurements.

10. Quantitative, because train speeds are numerical measurements.

11. Qualitative, because colors are attributes.

12. Quantitative, because width is a numerical measurement.

13. Quantitative, because weight is a numerical measurement.

14. Qualitative, because marital status is an attribute.

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

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

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

20. Ratio. A ratio of two data values can be formed, so one data value can be expressed as a multiple of another.

21. Horizontal: Nominal; Vertical: Ratio

22. Horizontal: Ordinal; Vertical: Ratio

Copyright © 2019 Pearson Education Ltd.

23. Horizontal: Nominal; Vertical: Ratio

24. Horizontal: Interval; Vertical: Ratio

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

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

27. Quantitative. Ratio. A ratio of two data entries can be formed, so one data entry can be expressed as a multiple of another.

28. Quantitative. Ratio. A ratio of two data entries can be formed, so one data entry can be expressed as a multiple of another.

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

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

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

32. Quantitative. Ratio. A ratio of two data entries can be formed, so one data entry can be 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

1. This is an observational study.

2. There is no way to tell why the people quit smoking. They could have quit smoking as a result of either chewing the gum or watching the DVD. The gum and the DVD could be confounding variables. To improve the study, 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 answer: Assign numbers 1 to 79 to the employees of the company. Use the table of random numbers and obtain 63, 7, 40, 19, and 26. The employees assigned these numbers will make up the sample.

4. (1) The sample was selected by using the students in a randomly chosen class. This is cluster sampling.

Copyright © 2019 Pearson Education Ltd.

(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 is systematic sampling.

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 convenience sample is not representative of a 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 apply a treatment to the adults.

12. Experiment. The study applies a treatment (intensive program to lower systolic blood pressure) to the subjects.

13. Experiment. The study applies a treatment (different photographs) to the subjects.

14. Observational study. The study does not apply a treatment to the motorists.

15. 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.

16. 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.

Copyright © 2019 Pearson Education Ltd.

17. Answers will vary.

18. Answers will vary.

19. (a) The experimental units are the 500 females ages 25 to 45 years old who suffer from migraine headaches. The treatment is the new drug used to treat migraine headaches.

(b) A problem with the design is that the sample is not representative of the entire population because only females ages 25 to 45 were used. To increase validity, use a stratified sample.

(c) For the experiment to be double-blind, neither the subjects nor the company would know whether the subjects are receiving the drug or the placebo.

20. (a) The experimental units are the 31 patients with type 2 diabetes. The treatment is the dietary supplement designed to control metabolism in patients with type 2 diabetes.

(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) Divide the subjects into age categories and then, within each age group, randomly assign subjects to either the treatment group or the control group.

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
23/50/12/26/42/63/08/10/81/91/89/42/06/78/00/55/13/75/47/07/
Treatment group: Dennis, Alice, Arthur, Kate, Harry, Bertha, Bill, Mercer and Edgar. Control group: Lewis, Raj, William, Edwin, Zoya, Lara, Jennifer, Ahmed and Ronald.

22. Answers will vary. *Sample answer*: Using a random number table:
Treatment group: 1, 3, 4, 6, 9, 11, 15, 16, 17, 18, 20, 23, 26, 27, 29, 31, 32, 33, 37, 39, 42, 45, 46, 49, 50, 52, 55, 56, 58, 59, 61, 62, 65, 69, 70, 73, 74, 78, 79, 80.
Control group: 2, 5, 7, 8, 10, 12, 13, 14, 19, 21, 22, 24, 25, 28, 30, 34, 35, 36, 38, 40, 41, 43, 44, 47, 48, 51, 53, 54, 57, 60, 63, 64, 66, 67, 68, 71, 72, 75, 76, 77.

23. Cluster sampling is used because the constituency is divided into areas, and 12 areas are then entirely selected. A possible source of bias is that problems of the residents of one area might be different from that of the other area.

24. 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.

25. 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.

Copyright © 2019 Pearson Education Ltd.

26. Stratified sampling is used because the organization is divided into departments and a sample executive is taken from each department.

27. Simple random sampling is used because each house number has an equal chance of being selected, and all samples of 1638 house numbers have an equal chance of being selected. The sample is unbiased.

28. Systematic sampling is used because every sixth customer entering the parlor is sampled. It is possible for bias to enter into the sample, if, for some reason, there is a regular pattern to customers entering the shop.

29. Sampling, because the population of mobile phone purchasers is too large for their most popular model of mobile phone to be easily recorded. Random sampling would be advised because it would be easy to select mobile phone purchasers randomly and then record their most popular model of mobile phone.

30. Census, because it is relatively easy to obtain the heights of the 264 students.

31. 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?"

32. The survey question is unbiased because it does not suggest that one should drink a lot of water to stay hydrated.

33. The question is biased because it already suggests that listening to music while studying increases the chances of retention. The question could be rewritten as "Does listening to music while studying have an effect on retention?"

34. The survey question is unbiased.

35. 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.

36. *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.

37. Answers will vary.

38. Answers will vary.

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.

Copyright © 2019 Pearson Education Ltd.

CHAPTER 1 REVIEW EXERCISE SOLUTIONS

1. Population: Collection of the responses of all U.S. adults
Sample: Collection of the responses of the 4787 U.S. adults who were sampled
Sample data set: 15% of adults who use ride-hailing applications and 85% who do not

2. Population: Collection of the opinions on the current educational policy of all professors in the Pennsylvania state.
Sample: Collection of the opinions on educational policy of the 42 professors in the Pennsylvania state that were sampled.

3. Population: Collection of the responses of all U.S. adults
Sample: Collection of the responses of the 2223 U.S. adults who were sampled
Sample data set: 62% of adults who would encourage a child to pursue a career as a video game developer or designer and 38% who would not

4. Population: Collection of the responses of all U.S. children and adults ages 16 years and older
Sample: Collection of the responses of the 1601 U.S. children and adults ages 16 and older who were sampled
Sample data set: 48% of children and adults who have visited a public library or a bookmobile over a recent span of 12 months and 52% who did not

5. Population parameter. The value $22.7 million is a numerical description of the total infrastructure-strengthening investments.

6. Sample statistic. The value 29% is a numerical description of a sample of U.S. voters.

7. Parameter. The 12 students minoring in math is a numerical description of all physics majors at a university.

8. Sample statistic. The value 30% is a numerical description of a sample of U.S. workers.

9. The statement “62% would encourage a child to pursue a career as a video game developer or designer” is an example of descriptive statistics. An inference drawn from the sample is that a majority of people encourage children to pursue a career as a video game developer or designer.

10. The statement “48% have visited a public library or a bookmobile over a recent span of 12 months” is an example of descriptive statistics. An inference drawn from the sample is that about half of U.S. children and adults ages 16 years and older have visited a public library or a bookmobile over a recent span of 12 months.

11. Quantitative, because ages are numerical measurements.

12. Quantitative, because IQ levels are numerical measurements.

13. Quantitative, because revenues are numerical measures.

14. Qualitative, because genders are attributes.

Copyright © 2019 Pearson Education Ltd.

15. Interval. The data can be ordered and meaningful differences can be calculated, but it does not make sense to say that 84 degrees is 1.05 times as hot as 80 degrees.

16. Ordinal. The data are qualitative and could be arranged in order of income level.

17. Nominal. The data are qualitative and cannot be arranged in a meaningful order.

18. Ratio. The data are quantitative, and it makes sense to say that $53.2 million is 1.12 times as much as $47.5 million.

19. Experiment. The study applies a treatment (drug to treat hypertension in patients with obstructive sleep apnea) to the subjects.

20. Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.

21. *Sample answer:* The subjects could be split into male and female and then be randomly assigned to each of the five treatment groups.

22. *Sample answer:* Number the volunteers and then use a random number generator to assign subjects randomly to one of the treatment groups or the control group.

23. Simple random sampling is used because random telephone numbers were generated and called. A potential source of bias is that telephone sampling only samples individuals who have telephones, who are available, and who are willing to respond.

24. Convenience sampling is used because the professor sampled a convenient group of his students. The study may be biased toward the opinions of the professor's students.

25. Cluster sampling is used because each district is considered a cluster and every pregnant woman in a selected district is surveyed. A potential source of bias is that the selected districts may not be representative of the entire area.

26. Systematic sampling is used because every tenth house is surveyed. A potential source of bias is that the locality the researcher is using may be posh.

27. Stratified sampling is used because the population is divided by religious groups and then 50 voters are randomly selected from each religious group.

28. Convenience sampling is used because of the convenience of surveying students in just one school. A potential source of bias is that the school is located in the downtown area where a lot of junk food may be available.

29. Answers will vary. *Sample answer*: Using the random number table in Appendix B, starting with the left-most number in row 7: 681/088/926/694/730/957/617/502/348/464/655/449/658/318/
The random numbers are 88, 502, 348, 464, 449, 318.

Copyright © 2019 Pearson Education Ltd.

CHAPTER 1 QUIZ SOLUTIONS

1. Population: Collection of the school performance of all Korean adolescents
Sample: Collection of the school performance of the 359,264 Korean adolescents in the study

2. (a) Sample statistic. The value 52% is a numerical description of a sample of U.S. adults.

(b) Population Parameter. The 90% of members that approved the contract of the new president is a numerical description of all Board of Trustees members.

(c) Sample statistic. The value 25% is a numerical description of a sample of small business owners.

3. (a) Qualitative, because debit card personal identification numbers are labels and it does not make sense to find differences between numbers.

(b) Quantitative, because final scores are numerical measurements.

4. (a) Ordinal, because badge numbers can be ordered and often indicate seniority of service, but no meaningful mathematical computation can be performed.

(b) Ratio, because horsepower of one car can be expressed as a multiple of another.

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

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

5. (a) Observational study. The study does not attempt to influence the responses of the subjects and there is no treatment.

(b) Experiment. The study applies a treatment (multivitamin) to the subjects.

6. Randomized block design

7. (a) Convenience sampling is used because all the people sampled are in one convenient location.

(b) Systematic sampling is used because every tenth machine part is sampled.

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

8. Convenience sampling. People at campgrounds may be strongly against air pollution because they are at an outdoor location.

Copyright © 2019 Pearson Education Ltd.

CHAPTER 1 TEST SOLUTIONS

1. (a) Sampling, because the population of New Jersey is too large for the most popular type of investment to be easily recorded. Random sampling would be advised because it would be easy to select people from New Jersey randomly and then record their most popular type of investment.

(b) Census, because the population is small and it is relatively easy to obtain the ages of the 30 employees.

2. (a) Sample statistic. The value of 72% is a numerical description of a sample of U.S. adults ages 18 years and older.

(b) Population parameter. The average evidence based reading and writing score of 543 is a numerical description of all test takers in a recent year.

3. (a) Stratified sampling is used because the high school students are divided into strata (male and female), and a sample is selected from each stratum.

(b) Simple random sampling is used because each customer has an equal chance of being contacted, and all samples of 625 customers have an equal chance of being selected.

(c) Convenience sampling is used because a sample is taken from members of a population that are readily available. The sample may be biased because the teachers at that school may not be representative of the population of teachers.

4. (a) Quantitative. Ratio. The number of employees are numerical measurements. A ratio of two data values can be formed, so it makes sense to say that 40 employees are twice as many as 20 employees.

(b) Quantitative. Interval. The grade point averages are numerical measurements. Data can be ordered and meaningful differences can be calculated, but it does not make sense to say that a person with a 3.8 GPA is twice as smart as a person with a 1.9 GPA.

5. (a) The survey question is unbiased.

(b) The question is biased because it already suggests that the town's ban on skateboarding in parks is unfair. The question could be written as "What are your thoughts on the town's ban on skateboarding in parks?"

6. (a) Population: Collection of the responses of all U.S. physicians
Sample: Collection of the 19,183 U.S. physicians who were sampled.

(b) Both. Location, employment status, benefits received, and speciality are qualitative because they are attributes. Income and time spent seeing patients per week are quantitative because they are numerical measurements.

(c) Nominal: location, employment status, benefits received, specialty
Ratio: income, time spent seeing patients per week

(d) Observational study. The study does not attempt to influence the responses of the physicians and there is no treatment.

Copyright © 2019 Pearson Education Ltd.