Section 1.1 Using Data to Answer Statistical Questions

1.1 Aspirin the wonder drug

a) Aspects of the study that have to do with design include the sample of 77,700 individuals, the randomization with assigning individuals in two groups (aspirin and placebo), and the plan to obtain percentages of each group who died of cancer.
b) Aspects having to do with description include the actual counts of individuals in the sample who died of cancer (i.e., 562 for those taking aspirin and 664 for those taking placebo).
c) Aspects that have to do with inference include the use of statistical methods to conclude that taking aspirin reduces the risk of dying of cancer.

1.2 Poverty and age

a) The aspects referring to description are the percentages of the 30,000 addresses eligible to complete the questionnaire (21.1% of children under the age of 18, 13.5% of people aged 18 to 64, and 10.0% of people aged 65 and older) who were in poverty.
b) The statistical method that predicted that the percentage of all Americans aged 18 to 64 in 2014 who were in poverty was between 13.2% and 13.8% is an example of inference.

1.3 GSS and heaven

Yes, definitely: 64.6%; Yes, probably: 20.8%; No, probably not: 8.7%; No, definitely not: 5.9%

1.4 GSS and heaven and hell

a) Yes, definitely: 64.3%; Yes, probably: 20.8%; No, probably not: 8.8%; No, definitely not: 6.0%
b) Yes, definitely: 52.6%; Yes, probably: 20.3%; No, probably not: 14.8%; No, definitely not: 12.3%; The percentage of "yes, definitely" responses was higher for belief in heaven in 2008.

1.5 GSS for subject you pick

The results for this item will be different depending on the topic that you chose.

Section 1.2 Sample Versus Population

1.6 Description and inference

a) With description, we are summarizing a group of numbers. We can use description with either samples or populations. With inferences, we use data from samples to make conclusions or predictions about populations. For example, if we ask a sample of adults how many pets they own, and take the mean number of pets, that number is a description. If we use that number to predict the mean number of pets owned by the whole population, the predicted mean (or the predicted range for the mean) would be an inference.
b) Descriptive statistics would be useful to summarize data from a population. With a census, it would be unwieldy to examine everyone's ages, for example, but it would be useful to know a mean age. Inferential statistics are not needed, however, because we already have information about the population; we don't need to predict it.

1.7 Ebook use

a) The sample is the sample of 170 students of Winthrop University during the Spring Semester 2014.
b) The population consists of all registered students in Winthrop University.
c) The statistic is the percentage of sampled students who had used ebooks for their academic work (i.e., 45%).

1.8 Concerned about global warming?

a) The sample is the set of polled Floridians. The population is the set of all adult Florida residents.
b) The percentages quoted are statistics since they are summaries of the sample.

1.9 Preferred holiday destination

a) Each faculty member of the college of business is a subject.
b) The sample is the number of faculty members interviewed about their preferred holidays' destination.
c) The population is all faculty members in the college of business.

1.10 Is globalization good?

a) The samples are those people selected from each country to participate in the survey. The populations are all adults in Africa and all adults in North America.
b) These are statistics because they represent a summary of the sample data.

1.11 Graduating seniors' salaries

a) These are descriptive statistics. They are summarizing data from a population – all graduating seniors at a given school.
b) These analyses summarize data on a population – all graduating seniors at a given school; thus, the numerical summaries are best characterized as parameters.

1.12 At what age did women marry?

a) The mean age of 24.1 years for this sample is descriptive.
b) The historian estimates the age for the whole population of brides in early 19th century New England, estimating the average age to fall between 23.5 and 24.7. This is inferential.
c) The inference refers to the population of all New England brides between the years of 1800 and 1820.
d) The average of 24.1 years is based on a sample and is therefore a statistic.

1.13 Age pyramids as descriptive statistics

a) The bar graph for 1750 shows shorter and shorter bars as age increases indicating that there were few Swedish people who were old in 1750.
b) For every age range, the bars are much longer for both men and women in 2010 than in 1750.
c) The bars for women in their 70's and 80's in 2010 are longer than those for men of the same age in the same year.
d) The first manned space flight took place in 1961 so that people born during this era would fall in the 45–49 year old category. This is the largest five-year group for both men and women.

💻1.14 Gallup polls

Responses to this exercise will differ depending on the studies that students choose.

a) The descriptive statistic will be a summary of data, without any prediction or population estimate. It might be a mean rating for a given attitude, for example.
b) The inferential statistical analysis will have some kind of prediction or estimation; for example, the inferential statistic might include the margin of error for a mean, indicating that the population mean likely falls somewhere in a given range.

1.15 Graduate studies

a) Yes, the populations are the same for the two studies.
b) Two samples are chosen randomly, they are independent from each other and almost unlikely to be the same.
c) The sample proportions are almost certainly different because they are obtained from two independent random samples that are almost certainly different.

1.16 Samples vary less with more data

a) It would be more surprising to flip a coin 500 times and observe all heads.
b) As the sample size increases, the amount by which sample proportions tend to vary decreases. The estimates from larger samples, therefore, tend to be more accurate than estimates from smaller samples. When the coin is flipped just 5 times, it's easy to see that we could get a sample with all heads. However, when the number of flips is increased to 500, it is much more likely that the sample proportion is near the population proportion of 0.5. It would be extremely unlikely to observe very few heads or almost all heads in 500 flips of a fair coin.

1.17 Comparing polls

a) $\left(1/\sqrt{n}\right) \times 100\% = \left(1/\sqrt{1000}\right) \times 100\% = 0.0316 \times 100\% = 3.16\%$ (rounds to 3.2%)
b) The first four polls are all within the margin of error; however, Rand favored Obama slightly, and Fox underestimated Obama's margin. Generally, the polls are fairly accurate.

1.18 Margin of error and *n*

a) $\left(1/\sqrt{n}\right)\times 100\% = \left(1/\sqrt{900}\right)\times 100\% = (1/30)\times 100\% = 0.033\times 100\% = 3.3\%$. This suggests that between 69.7% and 76.3% of Canadians agree that the Liberals should not make any changes to the country's voting system without a national referendum first.

b) $\left(1/\sqrt{n}\right)\times 100\% = \left(1/\sqrt{1600}\right)\times 100\% = (1/40)\times 100\% = 0.025\times 100\% = 2.5\%$. This suggests that between 70.5% and 75.5% of Canadians agree that the Liberals should not make any changes to the country's voting system without a national referendum first.

1.18 (continued)

c) $\left(1/\sqrt{n}\right)\times 100\% = \left(1/\sqrt{2500}\right)\times 100\% = (1/50)\times 100\% = 0.02\times 100\% = 2\%$, which suggests that between 71% and 75% of Canadians agree that the Liberals should not make any changes to the country's voting system without a national referendum first.

As *n* increases, the margin of error decreases, and the sample becomes more accurate reflection of the population.

1.19 Smoking cessation

a) iii

b) Yes. Because the employees were assigned to treatments randomly, the study provides us with convincing evidence that the difference was due to the effect of the financial incentive.

Section 1.3 Using Calculators and Computers

1.20 Data file for friends

The results for this exercise will be different for each person who does it. The data files, however, should all look like this:

Friend	Characteristic 1	Characteristic 2
1		
2		
3		
4		

For each friend, you'll have a number or label under characteristics 1 and 2. For example, if you asked each friend for gender and hours of exercise per week, the first friend might have m (for male) under Characteristic 1, and 6 (for hours exercised per week) under Characteristic 2.

1.21 Shopping sales data file

Customer	Clothes	Sporting goods	Books	Music CDs
1	$49	$0	$0	$16
2	$0	$0	$0	$0
3	$0	$0	$0	$0
4	$0	$0	$92	$0
5	$0	$0	$0	$0

1.22 Sample with caution

a) The population of all American internet users between the ages of 18 to 29 years is not a random sample of the general population of all American internet users because all American internet users who are of ages less than 18 years or more than 29 years are excluded from sampling.

b) The population of all American internet users between the ages of 18 to 29 years is not a random sample of the general population of all American internet users because all American internet users who are of ages less than 18 or more than 29 as well as because all American internet non-users of all ages are excluded from sampling.

1.23 Create a data file with software

Your MINITAB data (from Exercise 1.21) will be in the following format, although it will reside in the cells of the MINITAB worksheet.

Customer	Clothes	Sporting goods	Books	Music CDs
1	49	0	0	16
2	0	0	0	0
3	0	0	0	0
4	0	0	92	0
5	0	0	0	0

1.24 Use a data file with software

See solution for Exercise 1.21 for format of data in MINITAB.

1.25 Simulate with the Sampling Distribution for the Sample Proportion web app

a) These will be different each time this exercise is completed.
b) Regardless of the specific graphs constructed in (a), you will see that the amounts by which sample percentages tend to vary get smaller as the sample size n gets larger.
c) The practical implication of this is that larger sample sizes tend to provide more accurate estimates of the true population percentage value.

1.26 Margin of error

a) Answers will vary.
b) $\left(1/\sqrt{n}\right) \times 100\% = \left(1/\sqrt{1000}\right) \times 100\% = 0.0316 \times 100\% = 3.16\%$ (rounds to 3%)
c) Answers will vary.
d) Answers will vary.

1.27 Ebola outbreaks

The answer to this problem is based on a random process. This leads to potentially different answers each time it is performed. The binomial distribution (see Section 6.3) says that 14 or fewer people who died should occur in only about 1 in 100 simulations, so most students will likely not see any of these situations.

Chapter Problems: Practicing the Basics

1.28 UW Student survey

a) The population is the entire UW student body of 40,858. The sample is the 100 students who were asked to complete the questionnaire.
b) This value would not necessarily equal the value for the entire population of UW students. It is quite possible that the sample of 100 is not exactly representative of the whole student body. This percentage is only an estimate of the percentage of all students who would respond this way. It is unlikely that any single sample of 100 would have a percentage that was exactly the percentage of the entire population.
c) The numerical summary is a sample statistic because it only summarizes for a sample, not for a population.

1.29 Euthanasia

a) The population is all American adults.
b) The sample data are summarized by a proportion, 0.598.
c) The population proportion who would commit suicide.

1.30 Mobile data costs

This is an inferential statistic because the *Australian Communications and Media Authority* used a sample statistic: *the average cost per 1000 MB of free monthly mobile data allowance calculated from a sample data of 19 observations ($5.4)* to make a prediction about the population of all users: "*the average cost per 1000 MB of free monthly mobile data is most likely to be between $3.24 and $7.56*".

1.31 Breaking down Brown versus Whitman

a) The results summarize sample data because not every voter in the 2010 California gubernatorial election was polled.

b) The percentages reported here are descriptive in that they describe the exact percentages of the sample polled who were Democrat and voted for Brown, who were Republican and voted for Brown and who were Independent and voted for Brown.

c) The inferential aspect of this analysis is that the exit poll results were used to predict what percentage of each of the three parties (Democrat, Republican and Independent) voted for Brown in the 2010 California gubernatorial election. The margins of error give a likely range for the population percentages for each of the three parties.

1.32 Online learning

a) The sample is the 100 students surveyed. The population is all students in this school.

b) (i) Descriptive statistics would give us information about the preferences of the 100 students in the sample.

(ii) Inferential statistics allow us to draw a conclusion about the preferences of the student body.

1.33 Marketing study

For the study on the marketing of digital media, the population is all Facebook users, and the sample is the 1000 Facebook users to whom the ad was displayed. Example 5 suggests that we might determine that the average sales per person equaled $0.90. This would be a descriptive statistic in that it describes the average sales per person in the sample of 1000 potential customers. If one were to use this information to make a prediction about the population, this would be an inferential statistic.

1.34 Support of labor unions

a) $\left(1/\sqrt{n}\right) \times 100\% = \left(1/\sqrt{1540}\right) \times 100\% = 0.025 \times 100\% = 2.5\%$.

b) Between 50.5% and 55.5%

c) ii. Inferential statistics

1.35 Multiple choice: Use of inferential statistics?

The best answer is (c).

1.36 True or false?

False: We often want to describe the sample AND make inferences about the population.

Chapter Problems: Concepts and Investigations

1.37 Statistics in the news

If your article has numbers that summarize for a given group (sample or population), it's using descriptive statistics. If it uses numbers from a sample to predict something about a population, it's using inferential statistics.

1.38 What is statistics?

Answers will vary.

1.39 Surprising suicide data?

The likelihood of getting this result is extremely small.

1.40 Create a data file

See solution for Exercise 1.23 for format of data in MINITAB.

Chapter Problems: Student Activities

1.41 Getting to know the class

Answers will vary.