Chapter 1
Statistics: The Art and Science of Learning from Data

SECTION 1.1: PRACTICING THE BASICS

1.1. Aspirin and heart attacks

a) Aspects of the study that have to do with design include the sample, the randomization of the halves of the sample to the two groups (aspirin and placebo), and the plan to obtain percentages of each group that have heart attacks.
b) Aspects having to do with description include the actual percentages of the people in the sample who have heart attacks (i.e., 0.9% for those taking aspirin and 1.7% 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 having a heart attack.

1.2 Poverty and race

a) The aspects referring to description are the percentages of the 50,000 households (14.7% of whites, 30.4% of blacks, and 11.1% of Asians) who had incomes below the poverty level.
b) The statistical method that predicted that the percentage of *all* black households in the United States that had income below the poverty level was between 28.6% and 32.2% 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: PRACTICING THE BASICS

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 Number of good friends

a) The sample is the 840 respondents to the General Social Survey question, "About how many good friends do you have?"
b) The population is the American adult public.
c) The statistic reported is the percentage of respondents having only 1 good friend (i.e., 6.1%).

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 EPA

a) The subjects in this study are cars – specifically, new Honda Accords.
b) The sample is the few new Honda Accords that are chosen for the study on pollution emission and gasoline mileage performance.
c) The population is all new Honda Accords.

1.10 Babies and social preference

a) The sample includes the 16 six-month-old infants who were shown the sequence of videos.
b) The population would be all six-month-old infants.
c) The inference is that six-month-old infants have both the ability to recognize and the preference to align themselves with the helpful figure from the video.

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 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 National service

a) The populations are the same for the two studies. Two separate samples are taken from the same population.
b) The sample proportions are not necessarily the same because the two random samples may differ by chance.

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.

SECTION 1.3: PRACTICING THE BASICS

1.17 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.

1.18 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

1.19 Sample with caution

A sample of individuals with children who read the Ann Landers column is not a random sample of individuals with children because every member of the population does not have the same chance of being in the sample. Many individuals with children may not read Ann Landers while others who do read the column may choose not to participate in the survey.

💻1.20 Create a data file with software

Your MINITAB data (from exercise 1.18) will be in the following format, although it will reside in the cells of the MINITAB worksheet.

Customers	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

💻1.21 Use a data file with software

See solution for 1.20 for format of data in MINITAB.

1.22 Simulate with the *sample for a population* applet

a) These will be different each time this exercise is completed.
b) Regardless of the specific graphs constructed in part 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.

1.23 Is a sample unusual?

It would be surprising to get a percentage that's more than 20 points from the true population percentage with a sample of 50 people. If you use the applet to conduct a simulation, you'll see that most of the time, the samples fall within 14 points of the true population percentage – from about 56 to 84.

CHAPTER PROBLEMS: PRACTICING THE BASICS

1.24 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.

1.25 ESP

a) The population of interest is all American adults (the population from which this sample was taken).
b) The sample data are summarized by giving a proportion of all subjects (0.638) who said that they had at least one such experience, rather than giving the individual data points for all 3887 sampled subjects.
c) We might want to make an inference about the population with respect to the proportion who had had at least one ESP experience. We would use the sample proportion to estimate the population proportion.

1.26 Presidential popularity

This is an inferential statistic because Gallup was using the 45% of the sample who approved of how Obama is handling the presidency, along with the margin of error, to make a prediction about the population – how many Americans in general approved of how Obama was handling the Presidency.

1.27 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.

1.28 Reducing stress

a) The sample is the 100 students who were asked if they preferred to have a several-day period between the end of classes and the start of final exams. The population is all students in this school.
b) In this study, (i) descriptive statistics would give us information about the preferences of the 100 students in the sample, whereas (ii) inferential statistics would allow us to draw a conclusion about the preferences of the student body in general.

1.29 Marketing study

a) 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.

b) Example 4 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.

1.30 Multiple choice: Believe in reincarnation?

b) inferential statistics.

1.31 Multiple choice: Use of inferential statistics?

c) to make predictions about populations using sample data

1.32 True or false?

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

CHAPTER PROBLEMS: CONCEPTS AND INVESTIGATIONS

1.33 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.

1.34 What is statistics?

This answer will be different depending on the question chosen by the student.

1.35 Surprising ESP data?

This result would be very surprising with such a large sample. You'll notice that when you use the applet to simulate this study, you will get a sample proportion as large as 0.638, when the true proportion is 0.20, only VERY rarely. With such a large sample, if randomly selected, you'd expect a sample proportion very close to the population proportion.

1.36 Create a data file

See solution for Exercise 1.20 for format of data in MINITAB.

CHAPTER PROBLEMS: STUDENT ACTIVITIES

1.37 Activity 3

Answers will vary.