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 predict whether the percentages for *all* male physicians would be lower for those taking aspirin than for those taking placebo.

1.2 Poverty and race:

a) The aspects referring to description are the percentages of the 60,000 households (8.0% of whites, 23.4% of blacks, and 22.7% of Hispanics) 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 at least 22% but no greater than 25% is an example of inference.

💻1.3 GSS and heaven:

Yes, definitely: 64.8%; Yes, probably: 20.9%; No, probably not: 8.6%; No, definitely not: 5.8%

💻1.4 GSS and heaven and hell:

a) Yes, definitely: 66.5%; Yes, probably: 19.3%; No, probably not: 7.8%; No, definitely not: 6.5%

b) Yes, definitely: 55.4%; Yes, probably: 19.0%; No, probably not: 13.3%; No, definitely not: 12.3%; The percentage of people who believed in hell in 1998 was lower than the percentage who believed in heaven in that year.

💻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 819 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%).

1.8 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 Aspirin inference:

a) The sample includes all the male physicians who participated in the Physicians' Health Study.
b) The population would be all male physicians.
c) The inference is that aspirin would be more effective than placebo for *all* male physicians.

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, 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 graph shows fewer men and women as age increases. The bars on these graphs indicate thousands of people of a given gender and in a given age range. The very short bars toward the top indicate that there are very few men and women in their 70's and 80's in 1750.
b) For every age range, the bars are much longer for both men and women in 2000 than in 1750.
c) The bars for women in their 70's and 80's in 2000 are longer than those for men of the same age in the same year.
d) The bars of people who were born right after World War II, now middle-aged, are the longest bars for both women and men.

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 take an exit poll of 1000 voters and find that 0% or 100% voted for Smith.

b) As the sample size increases, the amount by which sample percentages tend to vary decreases. The estimates from larger samples, therefore, tend to be more accurate than estimates from smaller samples. Let's assume that 50% of the population of voters voted for Smith. With samples of just ten voters, it's easy to see that we could get a sample with zero or only one who voted for Smith, or even a sample with a much higher number who voted for Smith. With carefully selected samples of 1000 voters, however, it's much more likely that the percentage in the sample who voted for Smith is a more accurate estimate of the percentage of the population who voted for Smith. One would not expect to get percentages far off from 50%; for example, it would be very unlikely to find a sample of 1000 in which only 5 or in which 800 voted for Smith.

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
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 Internet poll:

An Internet poll is not a random sample because every person in the population does not have the same chance of being in the sample. Some people do not have computers, others don't have Internet access, still others do not visit the site on which the poll is posted, and some choose not to participate. Those with computers and Internet access who frequently surf the web would have a much higher chance of being in this study than those who don't meet those criteria.

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
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 CNN and Gallup were using the 35% of the sample who approved of how Bush is handling the presidency to make a prediction about the population – how many Americans in general approved of how Bush was handling the Presidency.

1.27 Bush vs. Kerry in other countries:
a) The results summarize sample data because not everyone in each country was polled.
b) The percentages reported here are descriptive in that they describe the exact percentages of the samples polled who preferred Kerry or Bush.
c) The inferential aspect of this analysis is that the BBC report is implying that these percentages provide information about the general population of each of these countries. The margin of error for the sample percentage gives information about the likely range in which the percentages fall in each of these countries.

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 CD's, the population is all customers to whom catalogs could be sent, and the sample is the 500 customers to whom catalogs actually are sent.

b) Example 4 suggests that we might determine that the average sales per person equaled $4. This would be a descriptive statistic in that it describes the average sales per person in the sample of 500 customers. If one were to use this information to make a prediction about the population, this would be an inferential statistic.

1.30 Believe in reincarnation?:

b) inferential statistics.

1.31 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.