Chapter 1

INTRODUCTION

1.1 The *population* consists of the entire set of responses from all teenagers, 13 to 17 years old, in the United States while the *sample* consists of the responses of the particular 1055 teens contacted in the telephone survey.

1.2 (a) A student at your college is the *unit*.

(b) Total monthly entertainment expenses is the *variable of interest*.

(c) The *statistical population* is the set of data consisting of the total monthly entertainment expenses for each student at your school.

1.3 (a) A person living in Chicago is the *unit*.

(b) The *variable of interest* is the characterization of a Chicago resident as an illegal alien or not.

(c) The *statistical population* consists of the collection of illegal alien characterizations of each resident of the city of Chicago.

1.4 (a) The *statistical population* consists of the height measurements of male students on campus. The *sample* consists of the height measurements of the members of the basketball team.

(b) The sample is likely to be non-representative of the population, as basketball players tend to be much taller than the typical student.

(c) There are many ways to choose a sample. One method is to use a table of random digits to select names from a student directory.

1.5 (a) An individual golfer is the *unit*.
(b) The *statistical population* consists of the collection of all possible decisions regarding "choice of hole" made by all golfers. The *sample* consists of the set of responses for this particular group of 46 golfers.

1.6 An individual adult is the *unit*. The *statistical population* consists of the entire set of data concerning the number of leisure hours per week each adult has in the entire surveying area, while the *sample* consists of the collection of such data for the group of 1,010 adults chosen to answer the question.

1.7 No, because a self-selection bias is likely to exist since only people who are interested in this particular exam are likely to answer, and such people perceive such a problem with the values.

1.8 The *population* consists of the entire collection of yes/no responses to program purchase from every reader while the *sample* consists of the yes/no responses from those readers who actually sent in the completed form. This is apt to be a very biased sample because persons who have not purchased the program are not as likely to take the time to fill out the form and send it in as readers who have purchased the program.

1.9 The newspaper is suggesting that the *population* is the collection of preferences for each adult in the city while the *sample* is the collection of preferences of the particular persons who sent in their votes. This sample is apt to be non-representative because those persons in the sample are self-selected. Only the few who feel very strongly positive will likely send in a vote.

1.10 (a) The sample is the characteristic of lying regularly or not reported by the 200 students.
(b) This is anecdotal. Data are not given to support this statement.
(c) The sample is the purchase/not purchase data for the 50 persons interviewed.

1.11 (a) This is anecdotal. No data given.
(b) The yes/no answer regarding multiple credit cards, for each of the 22 students, is the sample on which the statement is based.
(c) The yes/no answer regarding destination outside the continental United States, for each of the 55 people at the airport, is the sample on which the statement is based.

1.12 The term "comfortable" is not well defined. You could ask each person in a given group to write with the pen for ten minutes and then rate comfort on a five point scale from very uncomfortable to very comfortable. *Purpose*: Determine the average comfort rating on this scale.

1.13 The term "too long" is not well defined. By asking a number of people, we may determine that 5 minutes is too long. Further, the time will not be the same for all people. One improved statement of the purpose is:

Purpose: Determine if over half the persons take over 5 minutes to get cash during the lunch hour.

1.14 *Purpose*: For a fixed check-in date and fixed check-out date, take a sample of people and find the average time it takes each to book a hotel in San Francisco for the given dates using the Internet.

1.15 First number the classrooms 1 to 35. In Table 1, we started in row 20 using columns 29 and 30. Reading downward, and ignoring 00 and numbers above 35, we selected rooms 8, 7, 1, and 9. Answers will vary.

1.16 It is simpler to select 6 persons who will not go on the bus. Number the students from 1 to 50. In Table 1, we started in row 51 using columns 9 and 10. Reading downward, and ignoring 00 and numbers above 50, we selected students 23, 1, 44, 37, 19, and 26. Answers will vary.

1.17 We started in row 10 and read down column 9 and then down column 6 from the top. We ignored the second digit in a pair, and kept reading, where the two digits were the same. That type of assignment of students is not allowed. Answers will vary.

20 pairs of random digits

4,0	1,2	4,2	5,2	2,1
5,1	3,2	2,0	7,6	5,4
0,1	2,5	2,7	3,6	5,4
2,4	5,7	3,5	6,7	7,2

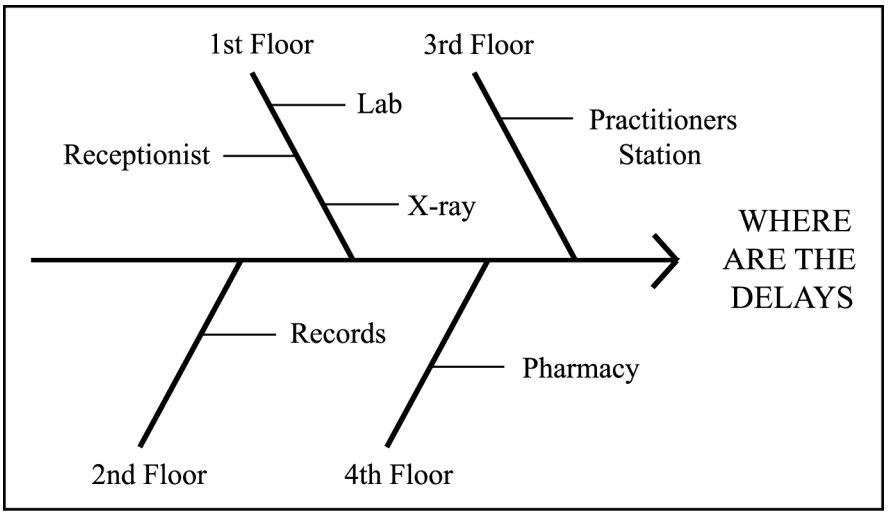
(a) $4/20 = 0.20$

(b) $9/20 = 0.45$

(c) $7/20 = 0.35$

1.18 At the lab, receptionist and x-ray.

1.19



1.20 (a) The statement must refer to an average amount per person. Clearly, some persons create much more solid waste and others less.

(b) Most likely from a sample of solid waste, the average was 4.6 pounds of solid waste per person per day. Certainly, that average for the whole population of the United States is unknown.

(c) You would prefer a nation wide sample. If restricted to households, you could conceivably use random numbers to select from the census listings.

1.21 (a) The miniature poodles could never be observed even if they greatly outnumber the Great Danes. Only the big dogs can volunteer to show they were inside the fence.

(b) Persons who call-in their opinions are self-selected because they have strong opinions. This is analogous to the big dogs who are the only volunteers to show they were inside the fence.