Section 1.1

For each of the following eight exercises, identify: a. the population, b. the sample, c. the parameter, d. the statistic, e. the variable, and f. the data. Give examples where appropriate.

44. A cardiologist is interested in the mean recovery period of her patients who have had heart attacks.

a.) Population would be all her patients who have had a heart attack

b.) Could use stratified sampling to sub-sample between different age groups and different types of chronic diseases such as diabetes, chronic high blood pressure, cardiovascular diseases, etc. Or she could also use a sample of patients who only had acute causes of a heart attack such as panic attack/ stress induced, serious injury, drug overdose, etc.

c.) The parameter would be the estimated time it took for all of her heart attack patients to recover from the heart attack.

d.) The statistic is the mean time it took for the sample group to recover from a heart attack.

e.) The Variable is the recovery period

f.) The data is the recorded recovery times

Use the following information to answer the next three exercises: A Lake Tahoe Community College instructor is interested in the mean number of days Lake Tahoe Community College math students are absent from class during a quarter.

50. What is the population she is interested in?

a. all Lake Tahoe Community College students

b. all Lake Tahoe Community College English students

c. all Lake Tahoe Community College students in her classes

d. all Lake Tahoe Community College math students

52. The instructor's sample produces a mean number of days absent of 3.5 days. This value is an example of a:

a. parameter.

b. data.

c. statistic.

d. variable.

Section 1.2

For the following exercises, identify the type of data that would be used to describe a response (quantitative discrete, quantitative continuous, or qualitative), and give an example of the data.

53. number of tickets sold to a concert

- Quantitative Continuous

54. percent of body fat

- Quantitative Continuous

55. favorite baseball team

- Qualitative

74. A "random survey" was conducted of 3,274 people of the "microprocessor generation" (people born since 1971, the year the microprocessor was invented). It was reported that 48% of those individuals surveyed stated that if they had $2,000 to spend, they would use it for computer equipment. Also, 66% of those surveyed considered themselves relatively savvy computer users.

a. Do you consider the sample size large enough for a study of this type? Why or why not?

- I think it should be descent sample size, but it depends on how they sampled the population.

b. Based on your "gut feeling," do you believe the percents accurately reflect the U.S. population for those individuals born since 1971? If not, do you think the percents of the population are actually higher or lower than the sample statistics? Why?

- Based on my "gut feeling", I think the first statistic of 48% is probably accurate (depending on what is meant by 'computer', which I assume they mean laptop or desktop). However, the 66% of surveyed individuals considering themselves relatively savvy is a little tricky. One, because that is a subjective question, and two, the key word in that statement is 'relatively'. I think an important piece of information would be what year was this survey conducted. For instance, if this was 90's to early 2000's, then there could have been much less individuals of computer owners/ users within their peer group to compare their skills to. I believe this 66% is lower, but this is based on the number of families I know, that usually rely on one individual to set up and trouble shoot their electronics. But of course, this is just based on my "gut feeling" and rough analysis of available observations (available observations = off the top of my head).

Additional information: The survey, reported by Intel Corporation, was filled out by individuals who visited the Los Angeles Convention Center to see the Smithsonian Institute's road show called "America's Smithsonian."

c. With this additional information, do you feel that all demographic and ethnic groups were equally represented at the event? Why or why not?

- I do not believe all demographic groups were equally represented, based on the new information. This is because I think a good majority of visitors would be involved in academia, where computer use is more

common. Also, advertisement for something like this would most likely have been more saturated at academic institutions than it would have been for the general public.

d. With the additional information, comment on how accurately you think the sample statistics reflect the population parameters.

- I do not think it reflects population parameters very well. This would be accurate for a very specific population. Also, I think these parameters would fluctuate with the development of computers.

76. In advance of the 1936 Presidential Election, a magazine titled Literary Digest released the results of an opinion poll predicting that the republican candidate Alf Landon would win by a large margin. The magazine sent post cards to approximately 10,000,000 prospective voters. These prospective voters were selected from the subscription list of the magazine, from automobile registration lists, from phone lists, and from club membership lists. Approximately 2,300,000 people returned the postcards.

a. Think about the state of the United States in 1936. Explain why a sample chosen from magazine subscription lists, automobile registration lists, phone books, and club membership lists was not representative of the population of the United States at that time.

- At the time the Great Depression was still affecting Americans. So, this sample would be biased towards the wealthier voters who could afford an automobile, telephone, and club memberships. It is hard to say how much of a financial burden a magazine subscription would have been so I don't know how many economic groups would have been sampled from the magazine subscription. However, the content of the magazine could have been more biased towards one candidate over the other, which could cause more of their subscribers to lean more towards the bias; which might be different from subscribers of a different magazine.

b. What effect does the low response rate have on the reliability of the sample?

- The size of data may not be large enough to make a good representation of the population. Especially when compounded with a sampling method that may already include bias. The fewer results might also contain a higher proportion of participants with stronger opinions.

c. Are these problems examples of sampling error or nonsampling error?

- It is a little bit of both. The way they chose who to survey was a sampling error, while the non-response of 77% of those surveyed is a non-sampling error.

d. During the same year, George Gallup conducted his own poll of 30,000 prospective voters. These researchers used a method they called "quota sampling" to obtain survey answers from specific subsets of the population. Quota sampling is an example of which sampling method described in this module?

- It could be either stratified or cluster. If they sampled individuals within each subset then it is stratified, but if they sampled entire subsets, then it is clustering.

Section 1.3

81. Sixty adults with gum disease were asked the number of times per week they used to floss before their diagnosis. The (incomplete) results are shown in Table 1.34.

# Flossing per Week	Frequency	Relative Frequency	Cumulative Relative Freq.
0	27	0.45	
1	18		
3			0.933
6	3	0.05	
7	1	0.0167	

Table 1.34 Flossing Frequency for Adults with Gum Disease

a. Fill in the blanks in Table 1.34.

# Flossing per Week	Frequency	Relative Frequency	Cumulative Relative Freq.
0	27	0.45	0.45
1	18	0.3	0.75
3	11	0.1833	0.9333
6	3	0.05	0.9833
7	1	0.0167	1

b. What percent of adults flossed six times per week?

- 5%

c. What percent flossed at most three times per week?

- About 18%

82. Nineteen immigrants to the U.S were asked how many years, to the nearest year, they have lived in the U.S. The data are as follows: 2; 5; 7; 2; 2; 10; 20; 15; 0; 7; 0; 20; 5; 12; 15; 12; 4; 5; 10.

Table 1.35 was produced.

Data	Frequency	Relative Frequency	Cumulative Relative Freq.
0	2	2/19	0.1053
2	3	3/19	0.2632
4	1	1/19	0.3158
5	3	3/19	0.4737
7	2	2/19	0.5789
10	2	2/19	0.6842
12	2	2/19	0.7895
15	1	1/19	0.8421
20	1	1/19	1

Table 1.35 Frequency of Immigrant Survey Responses

a. **Fix the errors in Table 1.35. Also, explain how someone might have arrived at the incorrect number(s).**

Data	Frequency	Relative Frequency	Cumulative Relative Freq.
0	2	0.105263158	0.105263158
2	3	0.157894737	0.263157895
4	1	0.052631579	0.315789474
5	3	0.157894737	0.473684211
7	2	0.105263158	0.578947368
10	2	0.105263158	0.684210526
12	2	0.105263158	0.789473684
15	2	0.105263158	0.894736842
20	2	0.105263158	1

- They did not count the frequency of 15 and 20 years correctly. And then they just filled in a one on the last box of Cumulative relative frequency, instead of actually calculating it. Also, if they would have stuck to one format for the numbers, it would have been easily noticed.

b. Explain what is wrong with this statement: "47 percent of the people surveyed have lived in the U.S. for 5 years."

- It should 47% have lived 5 years or less, or 15.8% have lived in the U.S. for five years.

c. Fix the statement in b to make it correct.

-"15.7 percent of the people surveyed have lived in the U.S. for 5 years."

d. What fraction of the people surveyed have lived in the U.S. five or seven years?

- 5/19

e. What fraction of the people surveyed have lived in the U.S. at most 12 years?

-2/19

f. What fraction of the people surveyed have lived in the U.S. fewer than 12 years?

-13/19

g. What fraction of the people surveyed have lived in the U.S. from five to 20 years, inclusive?

-13/19