SAMPLING AND DATA: HOMEWORK

EXERCISE 1

For each item below:

1. Identify the type of data (quantitative - discrete, quantitative - continuous, or qualitative) that would be used to describe a response.
2. Give an example of the data.

- Number of tickets sold to a concert
- Amount of body fat
- Favorite baseball team
- Time in line to buy groceries
- Number of students enrolled at Evergreen Valley College
- Most–watched television show
- Brand of toothpaste
- Distance to the closest movie theatre
- Age of executives in Fortune 500 companies
- Number of competing computer spreadsheet software packages

EXERCISE 2

Fifty part-time students were asked how many courses they were taking this term. The (incomplete) results are shown below:

Part-time Student Course Loads

# of Courses	Frequency	Relative Frequency	Cumulative Relative Frequency
1	**30**	**0.6**	
2	**15**		
3			

1. Fill in the blanks.
2. What percent of students take exactly two courses? _____
3. What percent of students take one or two courses? _____

EXERCISE 3

Sixty adults with gum disease were asked the number of times per week they used to floss before their diagnoses. The (incomplete)results are shown below:

Flossing Frequency of Adults with Gum Disease

# of Flossings per Week	Frequency	Relative Frequency	Cumulative Relative Frequency
0	27	0.45	
1	18		
3			0.93
6	3	0.05	
7	1	0.02	

1. Fill in the blanks.
2. What percent of adults flossed six times per week?
3. What percent flossed at most three times per week?

EXERCISE 4

A fitness center is interested in the average amount of time a client exercises in the center each week. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic
- Variable
- Data

EXERCISE 5

Ski resorts are interested in the average age that children take their first ski and snowboard lessons. They need this information to optimally plan their ski classes. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample

- Parameter
- Statistic
- Variable
- Data

EXERCISE 6

A cardiologist is interested in the average recovery period for her patients who have had heart attacks. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic
- Variable
- Data

EXERCISE 7

Insurance companies are interested in the average health costs each year for their clients, so that they can determine the costs of health insurance. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic
- Variable
- Data

EXERCISE 8

A politician is interested in the proportion of voters in his district that think he is doing a good job. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter

- Statistic
- Variable
- Data

EXERCISE 9

A marriage counselor is interested in the proportion of clients she counsels that stay married. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic
- Variable
- Data

EXERCISE 10

Political pollsters may be interested in the proportion of people that will vote for a particular cause. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic
- Variable
- Data

EXERCISE 11

A marketing company is interested in the proportion of people that will buy a particular product. Define the following in terms of the study. Give examples where appropriate.

- Population
- Sample
- Parameter
- Statistic

- Variable
- Data

EXERCISE 12

Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. Suppose an airline conducts a survey. Over Thanksgiving weekend, it surveys 6 flights from Boston to Salt Lake City to determine the number of babies on the flights. It determines the amount of safety equipment needed by the result of that study.

1. Using complete sentences, list three things wrong with the way the survey was conducted.
2. Using complete sentences, list three ways that you would improve the survey if it were to be repeated.

EXERCISE 13

Suppose you want to determine the average number of students per statistics class in your state. Describe a possible sampling method in 3 – 5 complete sentences. Be detailed.

EXERCISE 14

Suppose you want to determine the average number of cans of soda drunk each month by persons in their twenties. Describe a possible sampling method in 3 - 5 complete sentences. Be detailed.

EXERCISE 15

726 distance learning student at Long Beach City College in the 2004-2005 academic year were surveyed and asked the reasons they took a distance learning class. (Source: Amit Schitai, Director of Instructional Technology and Distance Learning, LBCC). The results of the survey are listed in the table below.

Reasons for Taking LBCC Distance Learning Courses

Convenience	87.6%
Unable to come to campus	85.1%
Taking on-campus courses in addition to my DL course	71.7%
Instructor has a good reputation	69.1%

To fulfill requirements for transfer	60.8%
To fulfill requirements for Associate Degree	53.6%
Thought DE would be more varied and interesting	53.2%
I like computer technology	52.1%
Had success with previous DL course	52.0%
On-campus sections were full	42.1%
To fulfill requirements for vocational certification	27.1%
Because of disability	20.5%

Assume that the survey allowed students to choose from the responses listed in the table above.

1. Why can the percents add up to over 100%?
2. Does that necessarily imply a mistake in the report?
3. How do you think the question was worded to get responses that totaled over 100%?
4. How might the question be worded to get responses that totaled 100%?

EXERCISE 16

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

The following table was produced:

Frequency of Immigrant Survey Responses

Data	Frequency	Relative Frequency	Cumulative Relative Frequency
0	2	2/19	0.1053
2	3	3/19	0.2632
4	1	1/19	0.3158
5	3	3/19	0.1579

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	1.0000

1. Fix the errors on the table. Also, explain how someone might have arrived at the incorrect number(s).
2. Explain what is wrong with this statement: "47 percent of the people surveyed have lived in the U.S. for 5 years."
3. Fix the statement above to make it correct.
4. What fraction of the people surveyed have lived in the U.S. 5 or 7 years?
5. What fraction of the people surveyed have lived in the U.S. at most 12 years?
6. What fraction of the people surveyed have lived in the U.S. fewer than 12 years?
7. What fraction of the people surveyed have lived in the U.S. from 5 to 20 years, inclusive?

EXERCISE 17

A "random survey" was conducted of 3274 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 $2000 to spend, they would use it for computer equipment. Also, 66% of those surveyed considered themselves relatively savvy computer users. (Source: *San Jose Mercury News*)

1. Do you consider the sample size large enough for a study of this type? Why or why not?
2. 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?

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

1. With this additional information, do you feel that all demographic and ethnic groups were equally represented at the event? Why or why not?
2. With the additional information, comment on how accurately you think the sample statistics reflect the population parameters.

EXERCISE 18

1. List some practical difficulties involved in getting accurate results from a telephone survey.
2. List some practical difficulties involved in getting accurate results from a mailed survey.
3. With your classmates, brainstorm some ways to overcome these problems if you needed to conduct a phone or mail survey.

Exercises 19 – 22 refer to the following: A Lake Tahoe Community College instructor is interested in the average number of days Lake Tahoe Community College math students are absent from class during a quarter.

EXERCISE 19

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

EXERCISE 20

X = number of days a Lake Tahoe Community College math student is absent is an example of a

A. Variable
B. Population
C. Statistic
D. Data

EXERCISE 21

The instructor takes her sample by gathering data on 5 randomly selected students from each Lake Tahoe Community College math class. The type of sampling she used is

A. Cluster sampling
B. Stratified sampling
C. Simple random sampling
D. Convenience sampling

EXERCISE 22

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

A. Parameter
B. Data
C. Statistic
D. Variable

Questions 23 – 24 refer to the following relative frequency table on hurricanes that have made direct hits on the U.S between 1851 and 2004. Hurricanes are given a strength category rating based on the minimum wind speed generated by the storm. (http://www.nhc.noaa.gov/gifs/table5.gif)

Frequency of Hurricane Direct Hits

Category	Number of direct hits	Relative Frequency	Cumulative Frequency
1	109	0.3993	0.3993
2	72	0.2637	0.6630
3	71	0.2601	
4	18		0.9890
5	3	0.0110	1.0000

EXERCISE 23

What is the relative frequency of direct hits were category 4 hurricanes?

A. 0.0768
B. 0.0659
C. 0.2601
D. Not enough information to calculate

EXERCISE 24

What is the relative frequency of direct hits were AT MOST a category 3 storm?

A. 0.3480
B. 0.9231
C. 0.2601
D. 0.3370

Questions 25 thru 27 refer to the following: A study was done to determine the age, number of times per week and the duration (amount of time) of resident use of a local park in San Jose. The first house in the neighborhood around the park was selected randomly and then every 8^{th} house in the neighborhood around the park was interviewed.

EXERCISE 25

Number of times per week is what type of data?

A. qualitative
B. quantitative – discrete
C. quantitative - continuous

EXERCISE 26

The sampling method was

A. simple random
B. systematic
C. stratified
D. cluster

EXERCISE 27

Duration (amount of time) is what type of data?

A. qualitative
B. quantitative – discrete
C. quantitative - continuous