Chapter 1 Statistics, data and statistical thinking

1.1 The science of Statistics

Statistics is the science of_____. It involves <u>collecting, classifying, summarizing, organizing, analyzing, and interpreting</u> numerical information.

Example: Math Test scores for 50 students from a class with 200 students:
Data: 90, 85, 88, 76, ….., 63. (50 scores)

*We can use the following table to summarize the data.

score	frequence
>=90	
80-89	
70-79	
60-69	
<60	

*We also can do some calculation, such as

Average score = $\dfrac{sum\ of\ 50\ scores}{50}$

Range = max. score – min. score

*We may use the average of 50 students to estimate the average of the 200 students.

Main procedures in statistics:

- How to _____data---(process) sampling
- How to _____ data─(methods) descriptive statistics
- How to _____for a large data set---(methods) inferential statistics

1.2 Types of statistical applications

_____**statistics** utilizes numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present that information in a convenient form.
- Numerical: calculate _____, _____, _____ etc;
- Graphical: construct _____, _____, _____ etc.

_____**statistics** utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.
- Two important methods: _____ and _____

How to classify descriptive or inferential statistical studies:
Descriptive statistics: explores information for the own intrinsic interest of the data.
Inferential statistics: draws conclusions about a population based on sample information. (some key words: _____, _____)

Examples: classify each of the studies as either descriptive or inferential statistics.

1. Professional athlete salaries: In the Statistical Abstract of the United States, average professional athletes' salaries in baseball, basketball, and football were compiled and compared for the years 1990 and 2000.

Average Salary ($1000):

| Sport	1990	2000
Baseball(MLB)	598	1720
Basketball (NBA)	750	3522
Football	395	1071

(_____)

2. Thirty of the 198 students enrolled in Statistics were asked if they wanted Exam II to be a take-home or in-class assessment. Twenty, or about 67% of the students polled indicates a preference for an in-class exam. The professor concluded that the majority of the students would prefer an in-class exam for the second assessment. Did the professor perform a descriptive statistical study or an inferential statistical study?

(_____ statistics)

3. "U.S. Market Share for Credit and Debit Cards"
CardWeb.com, Inc. tracked all the credit or debit card purchases in the United States during 2005. The pie chart describes the results. (The graphic is an example of descriptive statistics.)
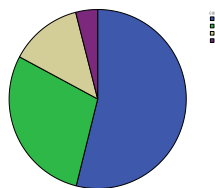
(_____ statistics)

4. There is a survey result. 48 out of 100 university students like football best. Based on this result, we conclude that almost half of university students like football best.

(_____ statistics)

1.3 Fundamental Elements of Statistics

- **An** _____: an object upon which we collect data.
  (e.g., one person, thing, transaction or event)
- **A** _____**: all** units of interest. (all people, objects, transactions or events)
- **A** _____: a subset of the units of a population.
- **A** _____: a characteristic or property of an individual experimental unit.
  (e.g., income, age, height, test scores, profit, cost, time; name, gender, race)
- _____: the value of a variable.
- _____: measurements for all E.U.s of a population.
- **A** _____ is an estimate or prediction or some other generalization about a population based on information contained in a sample.
- **A** _____: a statement about the degree of uncertainty associated with a statistical inference.

Example1. According to the state of the news media, 2006, the average age of viewers of "ABC World News Tonight" is 59 years. Suppose a rival network executive hypothesizes that the average age of ABC news viewers is less than 59. To test her hypothesis, she samples 500 ABC nightly news viewers and determines the age of each.

a. Describe the population.

The population is the set of units of interest, which is the set of _____

b. Describe the variable of interest.

The _____ of each viewer is the variable of interest.

c. Describe the sample.

The sample must be a subset of population, so the sample in this case is

_____ .

d. Describe the inference.

The inference of interest involves the generalization of the information contained in the sample of 500 viewers to the population of all ABC nightly news viewers. In particular, the executive wants to _____

She might accomplish this by calculating the average age of the sample and using the sample average to estimate the population average.

Example2. In order to get an estimate of average math test score of a class with 200 students. We randomly choose 50 students and calculate the average math test score for them.

a. Describe the population.

_____.

b. Describe the variable of interest.

_____.

c. Describe the sample.

_____ .

d. Describe the inference.

_____.

Example3. "Cola wars" is the popular term for the intense competition between Coca-Cola and Pepsi displayed in their marketing campaigns, which have featured movie and television stars, rock videos, athletic endorsements, and claims of consumer preference based on taste tests. Suppose, as part of a Pepsi marketing campaign, 1000 cola consumers are given a blind taste test. Each consumer is asked to state a preference for brand A and brand B.

a. Describe the population.

_____.

b. Describe the variable of interest.

_____.

c. Describe the sample.

_____ .

d. Describe the inference.

_____.

Four Elements of Descriptive Statistical Problems:

1. The _____ of interest
2. One or more _____ that are to be investigated
3. _____ summary tools
4. Identification of _____ in the data

Five Elements of Inferential Statistical Problems:

1. The _____ of interest
2. One or more _____ that are to be investigated
3. The _____ of population units
4. The inference about the _____ based on information contained in the _____
5. A measure of _____ for the inference

1.4 Types of Data

_____ **data** are measurements that are recorded on a naturally occurring numerical scale. (numerical observations)
For example: the number of students; the income; the height of a mountain, the stock index; the temperature (how many or how much)

_____ **data** are measurements that cannot be measured on a natural numerical scale; they can only be classified into one of a group of categories. (non-numerical or categorical observations)
For example: the brand of a car (Honda, Toyota, Ford); the hair color (black, brown, white); the name of the stocks, the types of blood, etc.

Example1. DDT: Chemical and manufacturing plants sometimes discharge toxic-waste materials such as DDT into nearby rivers and streams. These toxins can adversely affect the plants and animals inhabiting the river and the riverbank. The U.S. Army Corps of engineers conducted a study of fish in Tennessee River and its three creeks: Flint Creek, Limestone Creek, and Spring Creek. A total of 144 fish were captured, and the following variables were measured for each:
1. River/Creek where each fish was captured
2. Species(catfish, bass, or buffalo fish)
3. Length (centimeters)
4. weight (grams)
5. DDT concentration (parts per million)

Classify each of the five variables measured as quantitative or qualitative.
Quantitative variables: _____, _____, and _____
Qualitative variables: _____ and _____

Example2. The following table shows some information of employee in a company.

Name	Age	Gender	Salary
McKinley	52	F	$101,000
Logan	33	F	$82,000
……	……	……	……
St. Elias	46	M	$58,700
Angie	21	F	$43,900

What kinds of data are given in each column?

Quantitative data:_____

Qualitative data : _____

1.5 Collecting Data

Three ways to obtain data:
- from a _____(publications)
- from a _____(control)
- from an _____(e.g., a survey)

A _____ **sample** exhibits characteristics typical of those possessed by the
target population.
(Think: How to draw a **representative sample** for a political poll?)

How to get a representative sample: (randomly choose, no restriction)
A _____ **sample of n experimental units**: a sample selected from the population in
such a way that every different sample of size n has an equal chance of selection.

Example: Math test score: In order to get an estimate of average math test score of a class
with 200 students, we want to use 50 students as a sample. Which of the following sampling
leads to a representative sample?
1. only choose 50 top students.
2. only choose 50 bottom students.
3. randomly choose 50 students. (200 paper slips with the # assigned to each student,
randomly choose 50 from it)

Avoid using nonrandom samples:
- _____ (some E.U.s is excluded)
- _____ (unable to obtain data on all E.U.s in a sample)
- _____ (inaccurate data recorded)

1.6 The role of Statistics in critical thinking

Statistics has been widely used in many fields: Business, finance, industry, agriculture, social study, clinical study, government activities …

_____ involves applying rational thought and the science of statistics to critically assess data and inferences. Fundamental to the thought process is that variation exists in populations and process data.

Successful managers rely heavily on statistical thinking to help them make decisions.

Learning Objective of Chapter 1:

1. Classify descriptive and inferential statistics
2. Identify experimental unit, population, sample, variable
3. Identify quantitative (numerical) and qualitative (non-numerical, categorical)data