Introduction to Statistics

1.1 Review and Preview
1.2 Statistical Thinking and Critical Thinking
1.3 Types of Data
1.4 Collection Sample Data
What is Statistics

- Statistics is a collection of methods for:
  - Planning studies or experiments
  - Gathering data
  - Organizing, summarizing, analyzing, interpreting data
  - Drawing conclusions based on data
Two parts of statistics we will study

DESCRIPTIVE

INFERENTIAL
DESCRIPTIVE

Involves collecting, organizing, summarizing, and presenting data with charts, graphs, and tables.
DESCRIPTIVE

Examples:
- Average amount spent on a haircut
- Pie chart of hair colors
INFERENTIAL

- Procedures that are used to draw conclusions, make decisions or predictions about a population based on a sample.
Examples

- There is a relationship between smoking and lung cancer.
- Based on last years figures, it is predicted that 52% will vote in this years election.
POPULATION

the complete collection of ALL measurements or data that are being considered.
SAMPLE

the sub-collection of members selected from the population
PARAMETER vs STATISTICS

If it is not stated (or clear from the context) whether you have a population or a sample, assume you have a SAMPLE.
PARAMETER

- a value that describes a population.
- uses greek letters ($\mu$, $\sigma$, $\alpha$, $\beta$)

Example
  - The average SAT score in CA in 1990 was 897.
STATISTICS

- a value that describes a sample.
- Uses Roman letters (x, s, a, b)

Example
- The average age of a sample of men in Orlando, FL was 43 years.
QUESTION

- PARAMETER or STATISTIC.

The actual average height of all adult human males in the US is 5' 9.4"
QUESTION

- PARAMETER or STATISTIC.

60% of the households sampled from the US own more than one car.
A mall survey found 36% of women prefer lipstick to lip balm.
QUESTION

- PARAMETER or STATISTIC.

Currently there are 10,713 house units in Delano, CA.
MISUSES OF STATISTICS

- Evil intent on the part of dishonest people.
- Unintentional errors on the part of people who do not know any better.

- Need to learn how to distinguish between conclusions that are valid and seriously flawed.
PERCENTAGES

- Misleading or unclear percentages are sometimes used.
  - For example, if you take 100% of a quantity, you take it all. 110% of an effort does not make sense.
BAD SAMPLES

- Voluntary response sample (or self-selected sample) one in which the respondents themselves decide whether to be included

- In this case, valid conclusions can be made only about the specific group of people who agree to participate
Other Pitfalls you should be aware of

- Misleading Conclusions
- Loaded Questions
- Missing Data
- Precise Numbers
- Order of Questions
- Correlation and Causality
- Self Interest Study Groups
- GRAPHS (Chapter 2)
VARIABLE

a characteristic that changes for different individuals or objects under study.

Variables can be classified as qualitative (categorical) or quantitative (numerical).
QUANTITATIVE VARIABLE

Can be further classified into discrete data or continuous data.
DISCRETE

finite number of values, data variables that can be counted.
Examples

- The number of students in the room
- The number of games won
CONTINUOUS

results from infinitely many possibilities, data obtained through measuring. Corresponds to a continuous scale with no gaps, interruptions or jumps.
CONTINUOUS

- Example
  - Temperature
  - Weight
  - Time
  - Distance
Levels of Measurements

- The levels of measurements or ways to classify the data are:
  - Nominal
  - Ordinal
  - Interval
  - Ratio
NOMINAL

- data that consists of names, labels or categories. Data can not be arranged in an order.

- Examples
  - colors: red, blue, green, yellow
  - Survey responses: yes, no, undecided
ORDINAL

- categorical data that can be ranked.

- Examples
  - course grades: A, B, C, D, F
  - ranks of colleges
INTERVAL

numerical data that can be ranked; the differences between units of measure do exist; however, there is no true zero.

Examples
- Body temperature
- Years
RATIO

numerical data that can be ranked; the differences and ratios between units of measure do exist, and there exists a true zero.

- Example
  - The cost of a textbook
  - Distance traveled
Example

- The cost of a textbook
  - Zero represents the textbook being free, or you not purchasing a textbook
- Distance traveled
  - Zero represents you have not traveled.
EXAMPLE

Classify each as nominal-level, ordinal-level, interval-level, or ratio level data.

a) Sizes of cars
b) Nationality of each student
c) IQ of each student
d) Weights of new born babies
Ways to Collect Data

- The ways to collect data are:
  - Random sample
  - Simple random sample
  - Systematic Sampling
  - Convenience Sampling
  - Stratified Sampling
  - Cluster Sampling
RANDOM SAMPLING

- each member of the population has an equal chance of being selected.
A simple random sample of $n$ subjects is selected in a way that every possible sample of the same size $n$ has the same chance of being chosen.
SYSTEMATIC SAMPLING

- select a starting point then select every kth member of the population.
CONVENIENCE SAMPLING

- Pick sample members in an easy (non-random) way.

Hey! Do you believe in the death penalty?
STRATIFIED SAMPLING

- Subdivide population into at least 2 different groups/strata so subjects in same stratum share some trait, then draw a sample from each stratum.
CLUSTER SAMPLING

- Divide the population into sections/clusters, next randomly select some clusters last choose ALL members in those clusters.

Interview all voters in shaded precincts.
EXAMPLE

- Consider the following situations listed in a-d and identify which sampling method is being used. Then write the corresponding number in the blank.

Convenience Sampling
Stratified Sampling
Cluster Sampling
Systematic Sampling
EXAMPLE

To estimate the proportion of Bakersfield College students who prefer hip-hop music to other forms of music, a researcher obtains a listing of all the students at Bakersfield and contacts every 20th student on the list.
A statistics instructor teaching 300 students wants to determine if the class understands different sampling methods. She selects a sample of 10 students, all from the 1st row.
EXAMPLE

To determine employee satisfaction, corporate management for Target divide all stores into 3 regions: West Coast, East Coast, and Midwest. Then 100 employees are randomly selected from each region to complete an employee survey.
EXAMPLE

Sharp Healthcare wants to administer random drug tests in its hospitals. An outside research company randomly selects 4 hospitals and then drug tests every employee at those hospitals.
TYPES OF STUDIES

Observational Study – The experimenter records the outcomes of an experiment without control.

Experimental Study – The experimenter intervenes by administering treatment to the subjects in order to study its effects on the subject.
GROUPS

Treatment Group – the group that is being treated.

Controlled Group – the group that is not being treated.
EXAMPLE

Lipitor is a drug that is supposed to lower the cholesterol level. To test the effectiveness of the drug, 100 patients were randomly selected and 50 were randomly chosen to use Lipitor. The other 50 were given a placebo that contained no drug at all.
EXAMPLE

a) What is the treatment?

b) Identify the treatment group and the control group.

c) Is this an observational or experimental study?

d) What factor could confound the result?