Stat 240. Handout 1.

Chapter1. What is statistics?

- The word **statistics** has two meanings:
- In the more common usage, *statistics* refers to numerical facts. The numbers that represent the income of a family, the age of a students……
- The second meaning of *statistics* refers to the field or discipline of study.

  Statistics is a group of methods used to collect, analyze, present, and interpret data and to make decisions.

- Every day we make decisions that may be personal, business related, or of some other kind. Usually these decisions are made under conditions of uncertainty.
- Decisions made by using statistical methods are called *educated guesses.*
- Decisions made without using statistical (or scientific) methods are *pure guesses* and, hence, may prove to be unreliable. For example, opening a large store in an area with or without assessing the need for it may affect its success.

The three main aspects of statistics:

1)Design:

Designing the process of data collection (Identify population, what kind and how much data needed, how to collect a sample)

2) Description:

The methods of summarizing/describing data.

3) Inference:

Infer "general rules" about a population from a sample.

Types of Statistics

- A data set in its original form is usually very large. Consequently, such a data set is not very helpful in drawing conclusions or making decisions. It is easier to draw conclusions from summary tables and diagrams than from the original version of a data set. So, we reduce data to a manageable size by constructing tables, drawing graphs, or calculating summary measures such as averages. The portion of statistics that helps us do this type of statistical analysis is called **descriptive statistics**.

- **Descriptive Statistics** *Descriptive statistics* consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.
- In statistics, the collection of all elements of interest is called a **population**. The selection of a few elements from this population is called a **sample.**

- A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples. For example, we may make some decisions about the political views of all college and university students based on the political views of 1000 students selected from a few colleges and universities. The area of statistics that deals with such decision-making procedures is referred to as **inferential statistics**. This branch of statistics is also called *inductive reasoning* or *inductive statistics.*
- **Inferential Statistics** *Inferential statistics* consists of methods that use sample results to help make decisions or predictions about a population.

- *Probability,* which gives a measurement of the likelihood that a certain outcome will occur, acts as a link between descriptive and inferential statistics. Probability is used to make statements about the occurrence or nonoccurrence of an event under uncertain conditions

Opinion poll

More Americans Think The Democratic Party Better Represents Their Values.

In the following poll, 49% of Americans say that the Democratic Party better represents their values, compared to 43% who say that of the Republican Party.

POLL RELEASES,
Source: Gallup Poll Releases. http://www.gallup.com

Question: were all Americans asked their preference?

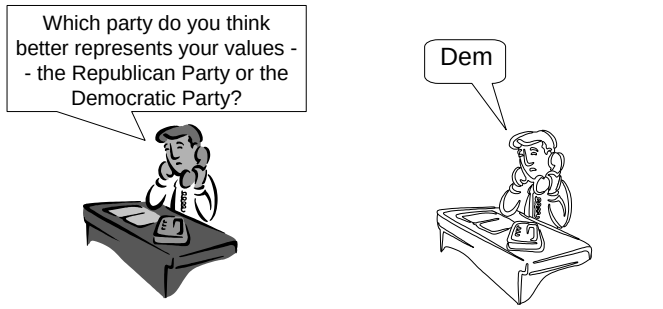
Sampling

- Interviewing all Americans would be impractical.
- Select a random sample.

Sampling: 1,000 adults, 18 years and older, were randomly selected and interviewed on the telephone.

- **Random selection** means that this group of individuals represents the population of American registered voters: A much smaller subset of the population of registered voters having the same characteristics:
- E.g. age distribution, gender, social classes etc.

Data Collection



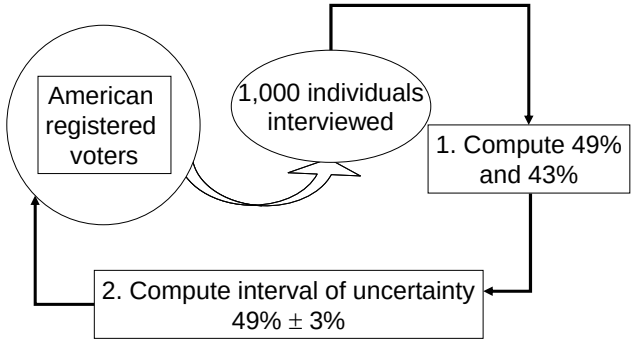
Sample data: Dem, Rep, Rep, Dem, Dem, Rep, Rep, ….

Data Description

1. Summary: 490 said Dem, 430 Rep
   It is better to compare **relative frequencies** .
   49% = 490/1,000 x 100 and 43% = 430/1,000 x 100
   These figures describe the sample preferences.
2. How can we deduce that "49% of Americans say that the Democratic Party better represents their values, compared to 43% who say that of the Republican Party."

3. Use statistical inference.

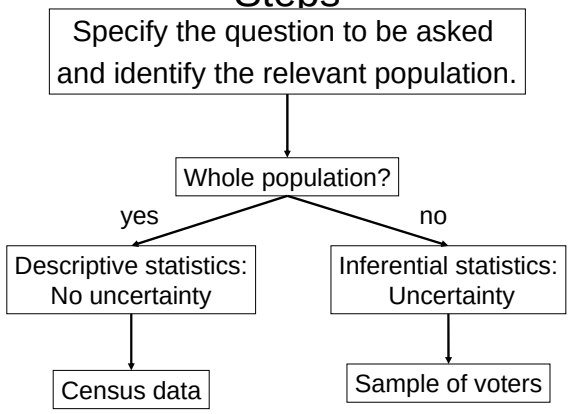
Summary



Steps

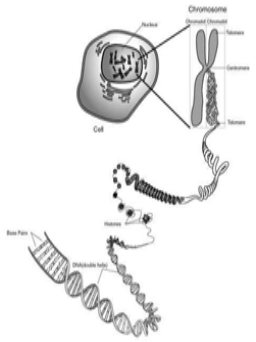
- Sampling:
  - Decide how to select the sample.
  - Collect the sample data.
- Analysis:
  - Describe the sample data, using graphs and numerical summaries.
  - Compute measures of uncertainty to project the sample summaries to the whole population.
- Report the results.
- In some cases, the whole relevant population is available.
  - Census data can be used to describe features of the population. Example- customers of a particular bank.

Steps



Forensic statistics

- DNA test results, first allowed as evidence in 1990, are changing the criminal justice system.
- DNA -- or deoxyribonucleic acid -- is the inherited, coded genetic material that determines much of one's appearance and physiology.
- **The power of DNA evidence lies in statistics**. Certain chromosome sites contain unique sequences of base pairs. Forensics technicians typically analyze three to five of these sites, called polymorphic markers, yielding a probability that only one person in, say, 100,000 or 1,000,000 could have the same profile.



How does it work?

- Forensic scientists extract DNA from samples of blood, semen, saliva, skin, or hair follicles found at the scene of the crime and develop a profile by analyzing several locations in the DNA where individual variations are common.
- This profile is compared with DNA taken directly from the suspect.
  - If the profiles do not match, the suspect is cleared -- as happens in about one-third of all rape cases, for example.
  - If the samples match, the weight of the evidence depends greatly on an understanding of the likelihood of such a match. **How likely it is that two matching samples came from different people?**