SECTION 1.1

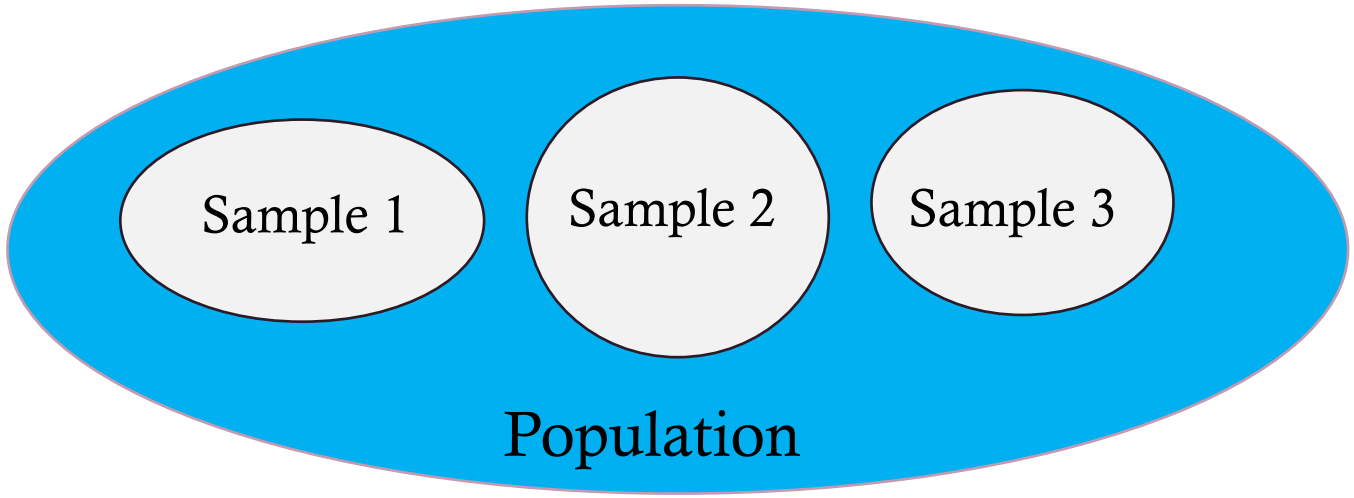
Populations, Samples and Processes

OBJECTIVES

- Learn the basic vocabulary of statistics.
- Distinguish between population and sample; parameters and statistics.

POPULATION & SAMPLE

- **Statistics** is the science of gathering, describing, and analyzing **data**.
- A **population** is a particular group of interest.
- A **sample** is a subset of the population from which data are collected.



POPULATION & SAMPLE

Example 1: The campus coffee shop of a college wants to know if students have tried the new flavor in their menu. They surveyed 120 randomly chosen students and found that 46 have tried it. Identify the population and the sample for this study.

Population: **All** students at the college

Sample: 120 randomly chosen students who were surveyed

Example 2: A personnel director is interested in determining how effective a new training course will be in improving the writing skill of her company's employees. The director randomly selects 30 employees and determines their writing skill both before and after taking the training course. Identify the population and the sample for this study.

Population: **All** employees in the personnel director's company

Sample: 30 employees in the company who were selected to take the new training course

VARIABLE

- A **variable** is a value or characteristic that changes among members of the population.
- **Data** are the counts, measurements, or observations gathered about a specific variable in a population in order to study it.
- A **census** is a study in which data are obtained from every member of the population.

VARIABLE

Examples: Determine the variable in the following studies.

1) A sociologist wishes to estimate the proportion of all adults in a certain region who have never married. In a random sample of 1,320 adults, 145 have never married, hence 145/1320 ≈ .11 or about 11% have never married.

   Variable: Marital status (Have never married / Have married)

2) After an airplane security scare on Christmas day, 2009, the Gallup organization interviewed 542 American air travelers about increased security measures at airports. The report stated that 78% of American air travelers are in favor of United States airports using full-body-scan imaging on airline passengers.

   Source: Jones, Jeffrey M. "In U.S., Air Travelers Take Body Scans in Stride." 11 Jan. 2010. http://www.gallup.com/poll/125018/Air-Travelers-Body-Scans-Stride.aspx (12 Dec. 2011).

   Variable: Opinion about increased security measures at airports (In favor / Against)

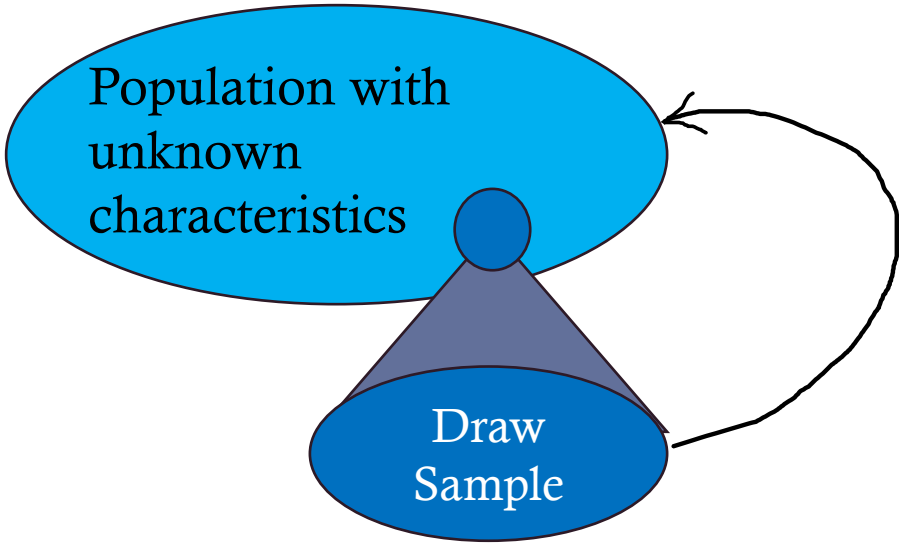
VARIABLE

- A **univariate** data set consists of observations on a single variable.
- **A bivariate** data when observations are made on each of two variables.
- **A multivariate** data arises when observations are made on more than one variable (so bivariate is a special case of multivariate).

Example:

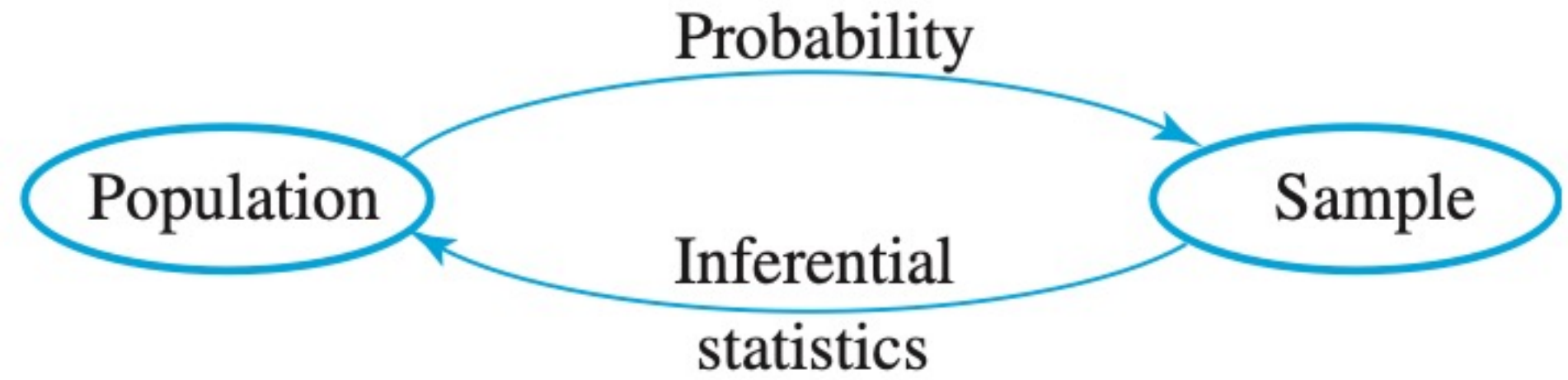
- Physical exam: (height, weight) → bivariate
- Automobile: (make, model, size, color) → multivariate

BRANCHES OF STATISTICS

Descriptive Statistics	Inferential Statistics
• Gathers, sorts, summarizes, and displays the data • Answers the questions such as: 1. Typical value? 2. How much variation? 3. Extreme values? 4. Shape or distribution of the data? 5. Relative position of a value in the data set? 6. Any relationships among variables? How strong is the relationship?	• Makes reasonable guesses about population characteristics using sample data 

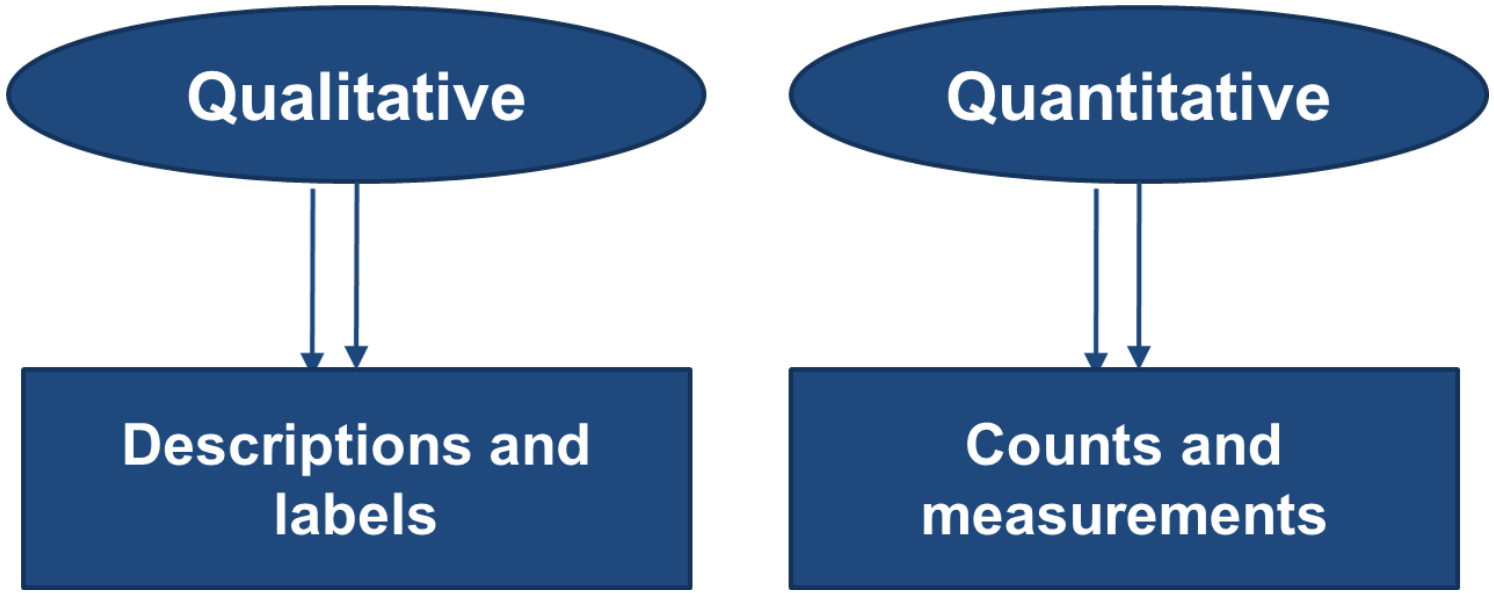
PROBABILITY & STATISTICS

Before we can understand what a particular sample can tell us about the population, we should first understand the uncertainty associated with taking a sample from a given population.



DATA CLASSIFICATION

- Qualitative data, also known as categorical data, consist of labels or descriptions of traits.
- Quantitative data consist of counts and measurements.



DATA CLASSIFICATION

Example: Classify the following data as either qualitative or quantitative.

a. Shades of red paint in a home improvement store

b. Rankings of the most popular paint colors for the season

c. Amount of red primary dye necessary to make one gallon of each shade of red paint

d. Numbers of paint choices available at several stores

e. The genders of the first 40 newborns in a hospital one year.

f. The fuel economy in miles per gallon of 20 new cars purchased last month.

THE PROCESS OF A STATISTICAL STUDY

Conducting a Statistical Study

1. Determine the design of the study.
   a) State the question to be studied.
   b) Determine the population and variables.
   c) Determine the type of the study, the sampling method.
2. Collect the data.
3. Organize the data.
4. Analyze the data to answer the question.

THE PROCESS OF A STATISTICAL STUDY

Example:

1. Does taking 80 mg of aspirin each morning reduce the risk of heart attacks? Population? Variables? Sampling method?
2. How to collect data? Observational study / Experiment; Population / Sample; Sampling Methods?
3. Organize data: graphical displays; numerical descriptions; compute statistics
4. Analyze the data to answer the question.

ENUMERATIVE VERSUS ANALYTIC STUDIES

Enumerative Study	Analytic Study
• The goal or purpose of the study is identifiable, i.e., not ambiguous. • The elements of the population are well defined and unchanging under this study. • The population could be an existing finite population about which one wants to draw conclusions. In this discussion, a *sampling frame*, which is a list of sample points to be collected is either available to an investigator or else can be constructed from definition of the study. • Examples: designed experiment, census	• Is not an Enumerative study • Draws conclusions about a process that does not even exit at the time of the study. Generally, here the result of the study is new because the objective is to improve things to be used in the future. • The study does not have a well-defined *sampling frame* and the impact of this study is highly localized and short term. • Example: in the production industry, where the new product is developed as an improvement over the existing one.

OBSERVATIONAL STUDY & EXPERIMENT

- An observational study observes data that already exist.
- An experiment generates data to be used in the study.

Example:

1) You want to determine the average age of college students across the nation.
2) Researchers wish to determine if flu shots actually help prevent severe cases of the flu.

OBSERVATIONAL STUDIES

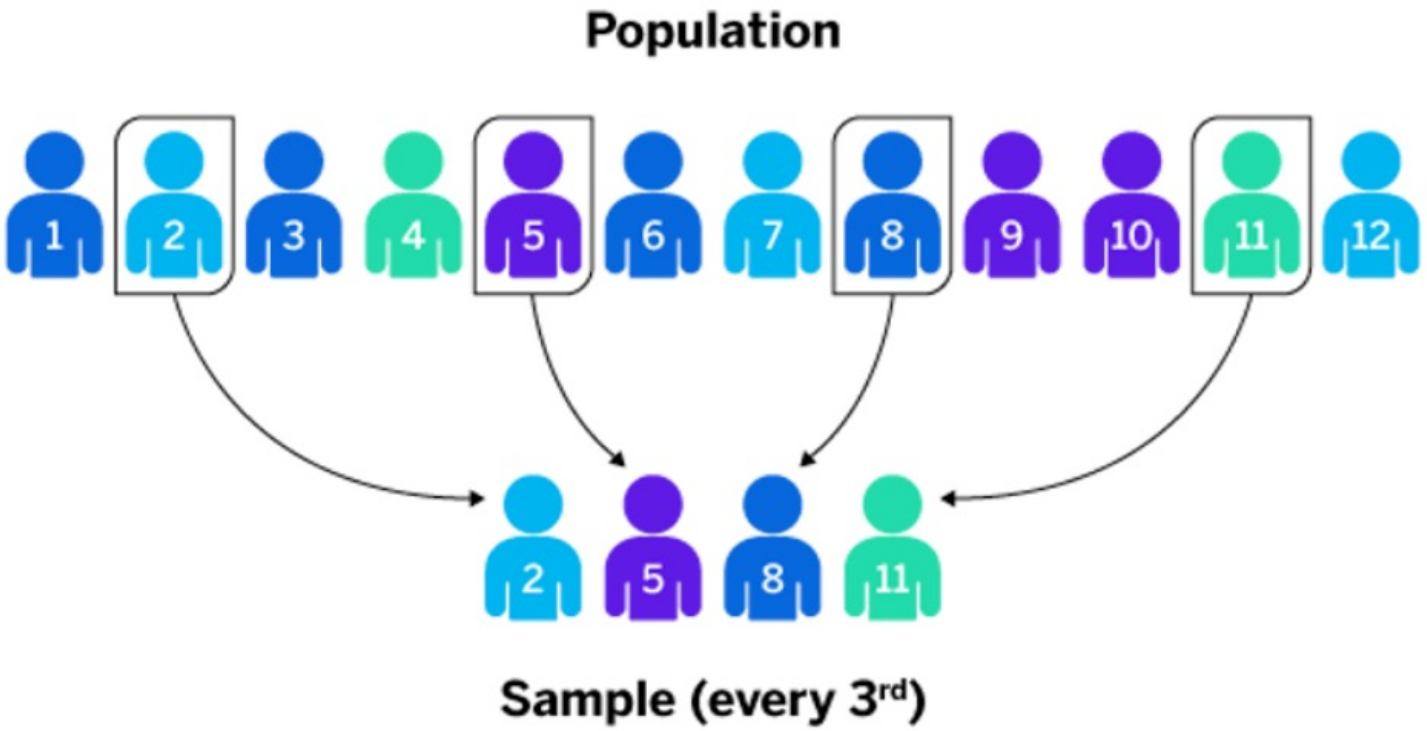
- A representative sample has the same relevant characteristics as the population and does not favor one group from the population over another.
- Sample Methods:

1) Random Sampling: Every member of the population has an equal chance of being selected.
2) Simple Random Sampling: Every sample of the population has an equal chance of being selected.
3) Stratified Sampling: A few members from each stratum (or group) are randomly selected.
4) Cluster Sampling: All members from a few randomly chosen clusters (group) are selected.
5) Systematic Sampling: Every n^{th} member of the population is chosen.
6) Convenience Sampling: The sample is chosen because it is convenitent for the researcher.

STRATIFIED SAMPLING

CLUSTER SAMPLING

SYSTEMATIC SAMPLING



EXPERIMENT

- A treatment is some condition that is applied to a group of subjects in an experiment.
- Subjects are people or things being studied in an experiment.
- Participants are people being studied in an experiment.
- The response variable is the variable in an experiment that responds to the treatment.
- The explanatory variable is the variable in an experiment that causes the change in the response variable.