INTRODUCTION

1. WHAT IS STATISTICS?

2. STATISTICS IN OUR EVERYDAY LIFE

3. TWO BASIC CONCEPTS: POPULATION AND SAMPLE

4. THE PURPOSEFUL COLLECTION OF DATA

5. USING STATISTICS WISELY

WHAT IS STATISTICS?

- The word **STATISTICS** originated from the Latin word "status", meaning "state".

- For a long time, it was identified solely with the displays of data and charts pertaining to the economic, demographic, and political situations prevailing in a country.

- **STATISTICS** as a subject provides a body of principles and methodology for designing the process of data collection, summarizing and interpreting the data, and drawing conclusions or generalities.

WHAT IS STATISTICS?

- Statistics is the science of describing or making inferences about the world from a sample of data.

STATISTICS

Descriptive   Inferential

STATISTICS IN OUR EVERYDAY LIFE

- Fact finding through the collection and interpretation of data is not confined to professional researchers.

- In our attempts to understand issues of environmental protection, the state of unemployment, or the performance of competing football teams, numerical facts and figures need to be reviewed and interpreted.

- In our day-to-day life, learning takes place through an often implicit analysis of factual information.

STATISTICS IN OUR EVERYDAY LIFE

- After the data are collected, statistical methods are available that summarize and describe the prominent features of data.

- These are commonly known as **DESCRIPTIVE STATISTICS.**

DESCRIPTIVE STATISTICS

- Descriptive statistics are methods for organizing and summarizing data.

STATISTICS IN AID OF SCIENTIFIC INQUIRY

- The phrase scientific inquiry refers to a systematic process of learning. A scientist sets the goal of an investigation, collects relevant factual information (or data), analyzes the data, draws conclusions, and decides further courses of action.

- Factual information is crucial to any investigation. The branch of statistics called **experimental design** can guide the investigator in planning the manner and extent of data collection.

STATISTICS IN AID OF SCIENTIFIC INQUIRY

- The goal of statistics is to help researchers organize and interpret the data.

"Data don't make any sense, we will have to resort to statistics."

POPULATION AND SAMPLE

- A **UNIT** is a single entity, usually a person or an object, whose characteristics are of interest.

- The **POPULATION OF UNITS** is the complete collection of units about which information is sought.

Population
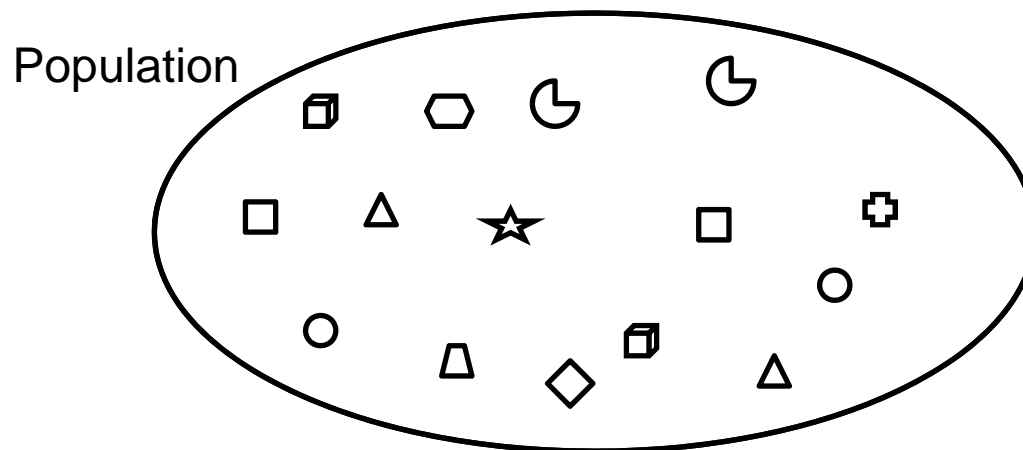
WHY SAMPLE?

- Measuring all units (trees, products, birds, etc.) is impractical, if not impossible.

- Sampling just a few units saves money.

- Sampling just a few units saves time.

- Some measurements are destructive: cutting down trees to inspect ring patterns or stem analysis capturing wildlife to examine their morphology, etc.

Sampling makes statistical methods attractive and powerful.

POPULATION AND SAMPLE

- A distinction between the data set that is actually acquired through the process of observation and the vast collection of all potential observations that can be conceived in a given context.

- The statistical name for the former is **sample;** for the latter, it is **population,** or **statistical population.**

SAMPLE

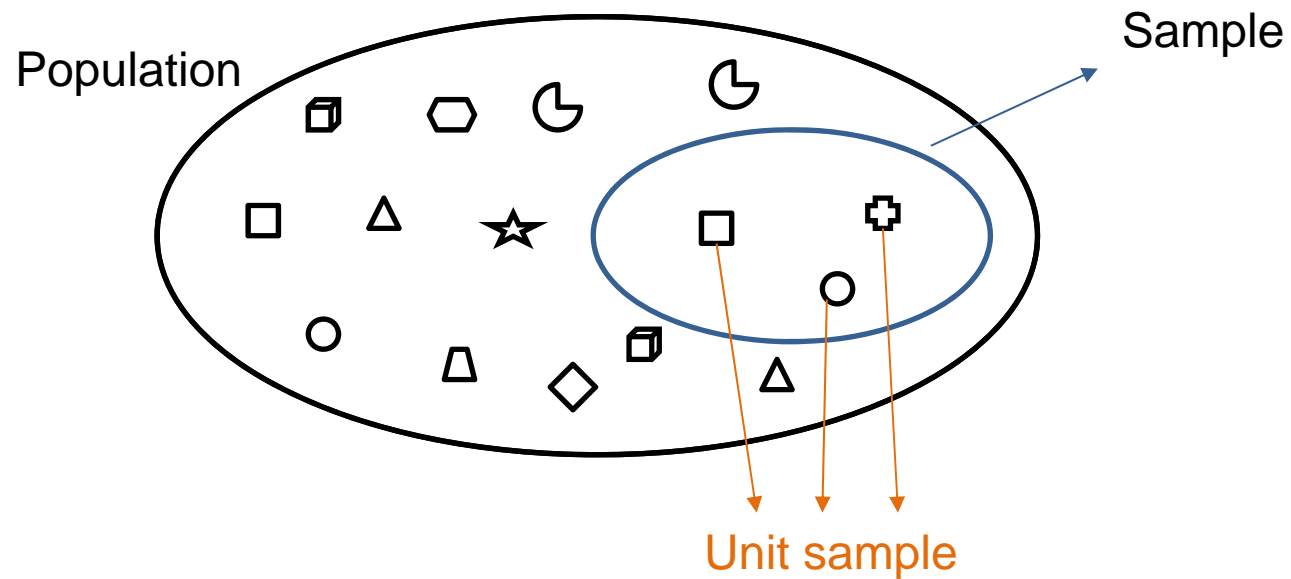- A sample is a group of units selected from a population.

TABLE 1 Populations, Units, and Variables

Population	Unit	Variables/Characteristics
Registered voters in your state	Voter	Political party Voted or not in last election Age Sex Conservative/liberal
All rental apartments near campus	Apartment	Rent Size in square feet Number of bedrooms Number of bathrooms TV and Internet connections
All campus fast food restaurants	Restaurant	Number of employees Seating capacity Hiring/not hiring
All computers owned by students at your school	Computer	Speed of processor Size of hard disk Speed of Internet connection Screen size

POPULATION AND SAMPLE

- A **STATISTICAL POPULATION** is the set of measurements (or record of some qualitative trait) corresponding to the entire collection of units about which information is sought.

- A **SAMPLE** from a statistical population is the subset of measurements that are actually collected in the course of an investigation.

SAMPLE PROBLEM 1

- Questions concerning the effect on health of two or fewer cups of coffee a day are still largely unresolved. Current studies seek to find physiological changes that could prove harmful.

- An article carried the headline "**Caffeine Decreases Cerebral Blood Flow**" describes a study which establishes a physiological side effect—a substantial decrease in cerebral blood flow for persons drinking two to three cups of coffee daily.

SAMPLE PROBLEM 1

- The cerebral blood flow was measured twice on each of 20 subjects. It was measured once after taking an oral dose of caffeine equivalent to two to three cups of coffee and then, on another day, after taking a look-alike dose but without caffeine. The order of the two tests was random and subjects were not told which dose they received. The measured decrease in cerebral blood flow was significant.

- Identify the population and sample.

THE PURPOSEFUL COLLECTION OF DATA

- Just making the decision to collect data to answer a question, to provide the basis for taking action, or to improve a process is a key step.

- Once that decision has been made, an important next step is to develop a **STATEMENT OF PURPOSE** that is both specific and unambiguous.

THE PURPOSEFUL COLLECTION OF DATA

- If the subject of the study is public transportation being behind schedule, you must carefully specify what is meant by late. Is it 1 minute, 5 minutes, or more than 10 minutes behind scheduled times that should result in calling a bus or commuter train late?

- Words like soft or uncomfortable in a statement are even harder to quantify. One common approach, for a quality like comfort, is to ask passengers to rate the ride on public transportation on the five-point scale.

THE PURPOSEFUL COLLECTION OF DATA

In any context, to pursue potential improvements of a process, one needs to focus more closely on particulars. Three questions

WHEN **WHERE** **WHO**

should always be asked before gathering further data. More specifically, data should be sought that will answer the following questions.

THE PURPOSEFUL COLLECTION OF DATA

- **WHEN** do the difficulties arise? Is it during certain hours, certain days of the week or month, or in coincidence with some other activities?

- **WHERE** do the difficulties arise? Try to identify the locations of bottlenecks and unnecessary delays.

- **WHO** was performing the activity and who was supervising? The idea is not to pin blame, but to understand the roles of participants with the goal of making improvements.

OBJECTIVES OF STATISTICS

The major objectives of statistics are:

- To make **inferences** about a population from an analysis of information contained in sample data. This includes assessments of the extent of uncertainty involved in these inferences.

- To **design the process and the extent of sampling** so that the observations form a basis for drawing valid inferences.

INFERENTIAL STATISTICS

Two main methods:

- Estimation

- Hypothesis testing

USING STATISTICS WISELY

1. Compose a clear statement of purpose and use it to help decide upon which variables to observe.

2. Carefully define the population of interest.

3. Whenever possible, select samples using a random device or random number table.

USING STATISTICS WISELY

4. Do not unquestionably accept conclusions based on self-selected samples.

5. Remember that conclusions reached in TV, magazine, or newspaper reports might not be as obvious as reported. You must be aware that the advocate, often a politician or advertiser, may only be presenting statistics that emphasize positive features.