MATH 2300: Statistical Methods

Introduction and Chapter 1

What is Statistics?

- What do you think of when you hear “statistics” ?

What is Statistics?

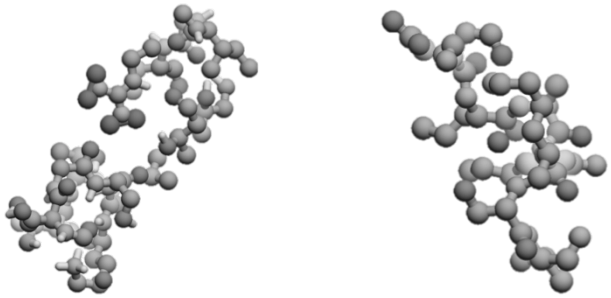
- Statistics is the *science* of collecting, organizing, summarizing, and interpreting **data**.
- It is also said to be the *science* of learning from **data**.
- Data: Information or facts that describe a group of interest
  - Not *"just"* numbers!
  - Context is key!

What I think of...

- Marketing and Politics
  - Conducting surveys to obtain opinions and feedback
- Public Health
  - Identifying sources of disease
  - Testing new drugs and treatments for diseases

What I think of...

- Bioinformatics and Molecular Biology
  - Developing new drugs and treatments
  - Studying the human genome
  - Using protein structures to predict protein functions

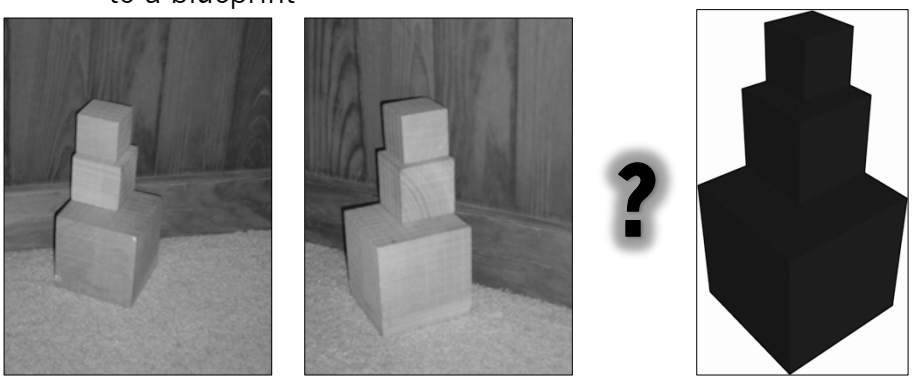


What I think of...

- Civil Engineering
  - Assessing effects of stress on structural elements
  - Studying impact of traffic flows on communities
- Materials Engineering
  - Studying properties of treatments to slow corrosion

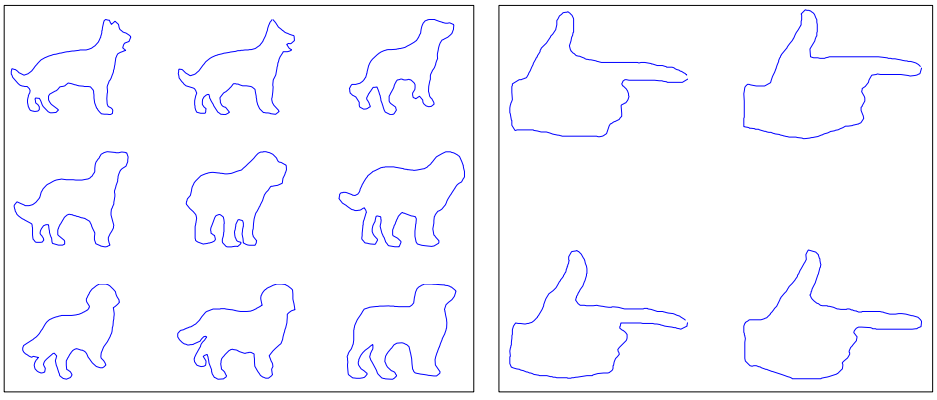
What I think of...

- Image and Shape Analysis
  - Facial recognition
  - Checking if a product was correctly made according to a blueprint



What I think of...

- Image and Shape Analysis
  - Distinguishing between objects



Chapter 1

The Nature of Statistics

Some Questions to Ask Before Producing Data

- What is the group of interest?
- What information about the group are we interested in?
- How do we collect this information?

The Population

- The entire group of individuals about which we want to get information.
- If we can quickly and/or cheaply do so, we perform a **census**:
  - The collection of data from *every* member of a population.
- Often, it is too time-consuming or expensive to obtain data for the entire population.
- How do we get around these constraints?

The Sample

- The subset of the population from which we actually collect data.
- Selected according to some prescribed manner.
- For this sample data, we:
  - Organize
  - Describe
  - Analyze

Example

- Each week, the Gallup Poll questions a group of 1500 adult U.S. residents to determine national opinion on a wide variety of issues.

- Population =
- Sample =

Example

- The 2000 Census tried to gather basic information from every household in the United States. A "long form" requesting much more information was sent to about 17% of households.

- Population =
- Sample =

Example

- There are reports that variation in output voltage from regulator devices is affecting the performance of the complete product. To assess this, 5 regulators are sent to the lab for study.

- Population =
- Sample =

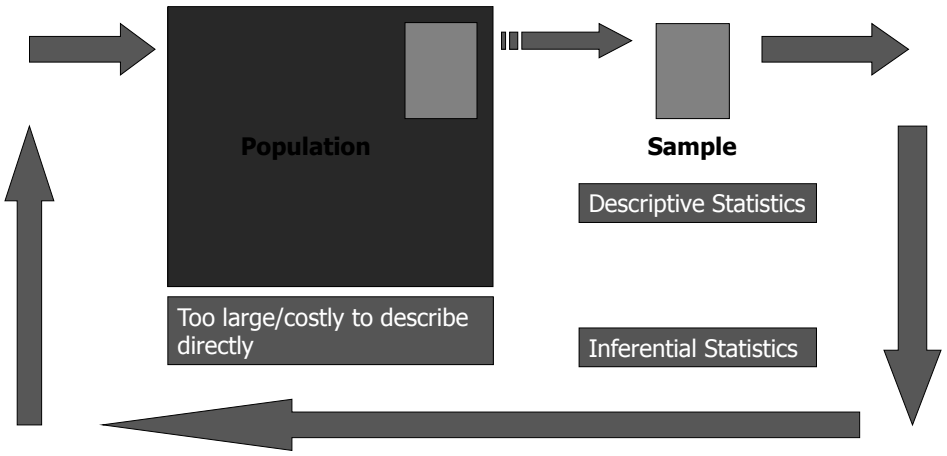
How do we describe objects in a data set?

- A **variable** is any characteristic whose value may change from one object to another in a population.
- Univariate data
  - Observations on a single variable
- Bivariate data
  - Observations on each of two variables
  - We will see this in Chapter 4
- Multivariate data
  - Observations made on more than one variable

Two Branches of Statistics

- Descriptive Statistics:
  - Summarize and describe key features of the data
  - Includes graphical methods and calculation of numerical summary measures
  - Discussed in Ch. 2 – 4
- Inferential Statistics:
  - Generalize and draw conclusions about the population from the sample data
  - Discussed in Ch. 8 – 14

The Process of Statistical Analysis



Conclusions are Uncertain!

- Variation is everywhere.
  - Is a 1% increase in profit due to changes or just a result of natural fluctuations?
- The data we have might not tell the full story.
  - Is there crucial information we don't have that may explain the results?
- How do we choose our samples to get the "best" samples?
  - *Random* sampling!
  - **Probability** helps us understand randomness
  - Ch. 5 – 7

Simple Random Samples (SRS)

- Every possible sample of a specified size *has an equal chance* of being selected.
- Every individual also *has an equal chance* of being selected for the sample.

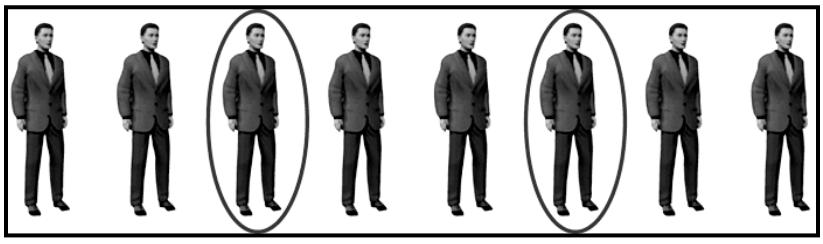
Example

Researchers wish to test a promising weight loss medication on a group of overweight persons. The list below needs to be randomly divided into two groups: a treatment and control group.

01	Birnbaum	35	11	Moses	25
02	Brown	34	12	Nevesky	39
03	Brunk	30	13	Obrach	30
04	Cruz	34	14	Rodriguez	30
05	Deng	24	15	Santiago	27
06	Hernandez	25	16	Smith	29
07	Jackson	33	17	Stall	33
08	Kendall	28	18	Tran	35
09	Loren	32	19	Wilansky	42
10	Mann	28	20	Williams	22

Systematic Samples

- Select some starting point and then select every k^{th} individual in the population.



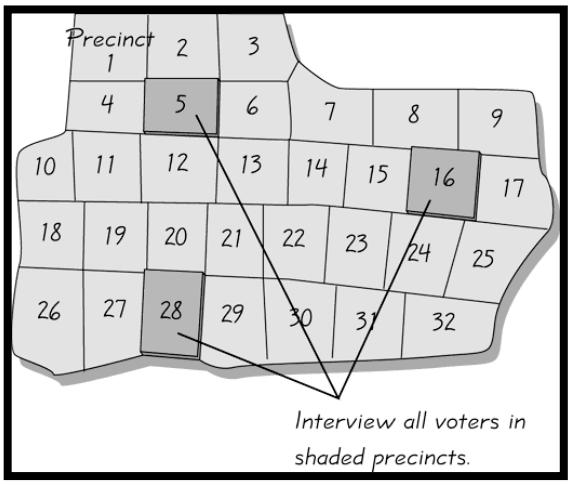
Stratified Samples

- First, divide the population into at least two subgroups.
  - All of the individuals in each subgroup should share some common characteristic.
  - The subgroups are called **strata**.
- Then draw a SRS from each subgroup.

Cluster Samples

- First, divide population into clusters
  - Cluster members do **NOT** share a common characteristic
  - Each cluster should be representative of the entire population
- Then randomly choose clusters to include
- Lastly, combine the chosen clusters

Cluster Samples



Representative Samples

- These are samples in which the relevant characteristics of the sample members are generally the same as the characteristics of the population.
- All of the previous methods *can* produce these.
- These samples will contain little or no **bias**.
  - A study is said to be **biased** if it *systematically* favors certain outcomes.

Poor Sampling Methods

- Voluntary Response Samples
  - Respondents choose whether to be included in the survey or not
    - Example: A poll on a website
- Convenience Samples
  - Researcher samples those who are readily available/ willing
    - Example: Polling only the students who sit near you in class

Basic Types of Studies

- Observational Studies
- Designed Experiments
- Meta-Analysis

Observational Studies

- **Observe** individuals and measure key variables or characteristics.
- No intentional influences on the responses.
- **Purpose**: To describe some group or situation

Designed Experiments

- Researchers apply some **treatment** to individuals in an attempt to impact the response.
- **Purpose**: To observe whether the treatment is associated with a change in the response

Meta-Analysis

- Study a topic that has been the subject of many previous studies to link them.
- **Purpose**: To find trends that were not apparent in individual studies.

Example

- Ask men and women to choose which features they consider essential in a health plan to see whether the required features differ across genders.

Designed Experiment, Meta-Analysis,
or Observational Study?

Example

- Test a group of college freshmen on economics. Divide them in half. Show one half a video presentation about economics and have the other half study the book. Test them all again to test for increased understanding.

Designed Experiment, Meta-Analysis, or Observational Study?

Example

- Every 3 years for the past 30 years, researchers looked at hiring trends for TTU graduates. You look at the results of all 10 studies to look for any trends over time.

Designed Experiment, Meta-Analysis, or Observational Study?