Probabilistic Systems Analysis for Civil Engineers

Instructor: **Yanlin Qi**

Institute of Transportation Studies
Department of Statistics
University of California, Davis

About the Course

Administrative

- To contact/meet
  - Office hours: To be posted on course webpage or by appointment
  - Email: ylqi@ucdavis.edu
    - Please keep email to a minimum.
    - The best options are to ask in person (e.g. after lecture, office hour)
- Credit hours: 4
- Lecture hours:
  - Tuesdays, Wednesdays, Thursdays 11:00 AM -1:15 PM
- Prerequisite: MAT 21C (Multivariable Calculus).

Assignments/Grading

- Assignments:
  - Homework: Typically focusing on a specific type of analysis and course material
  - Exams: Midterm and Final
    - Midterm Tuesday (Week 4), Jul. 16th in class
    - Final Thursday (Week 6), Aug. 1st in class
- Grading
  - Homework assignments (40%)
  - Midterm exam (25%)
  - Final exam (35%)

Course Description and Objectives

- Generalities
  - Probabilistic concepts and models
  - Engineering method and statistical thinking
  - Statistical analysis of engineering experimental/field data

- Course objectives:
  - To help you develop an understanding of principles of pro-bability and statistics, and
  - The application of those principles to life in general and engineering in particular.

Outline of the Course

1. Introduction	(Chapter 1, Sections 6.1)
2. Probability	(Chapter 2)
3. Discrete Random Variables	(Chapter 3)
4. Continuous Random Variables	(Chapter 4)
5. Joint Distributions	(Chapter 5)
[Midterm exam]	
6. Statistical Data Description	(Chapter 6)
7. Sampling Distributions	(Sections 7.1, 7.2)
8. Estimation	(Sections 7.3 – 7.4, 8.1 – 8.3)
9. Hypothesis Testing	(Chapters 9, 10)
10. Regression Analysis	(Chapter 11)
[Final exam]	

Homework

- 4 homework + 0-2 quizzes
- After due date and time – **no credit.**
- Lowest homework score will be dropped
- You can collaborate, though must turn in your own assignment with your own discussion
- 2 Types of assignments
    - Out of the classroom assignments (usually due 3 lectures later) – 4 assignments
    - In-class assignments (assigned and due in the same lecture) – 0-2 assignments, 20 minutes, announced via Canvas ahead of class

Bibliography

- Guide Textbook:
  - Douglas C. Montgomery and George C. Runger, *Applied Statistics and Probability for Engineers,* 7th ed. New York: John Wiley and Sons, 2018.

- Additional references:
  - Schaum's Outline of Probability and Statistics, 2nd or 3rd Ed. by Schiller, Alu Srinivasan and Spiegel.

  - Online Textbook for the Probability for Data Science class at UC Berkeley, By Ani Adhikari and Jim Pitman: http://prob140.org/textbook/content/README.html

Others

- 2-sided 8-1/2 notes (original handwriting) for final
- Turn in your crib sheet with exams
- Exams are closed-book, closed-notes.
- Neither the TA nor instructor will clarify any problems on the exam
- You need a calculator. No cellphones, no tables, no computers
- Check “Canvas” portal
- Exams are comprehensive

Academic integrity

- You are expected to adhere to the UC Davis code of academic conduct: *https://ossja.ucdavis.edu/code-academic-conduct*
- All assignments that are turned in for a grade must represent the student's own work.
- In cases where help was received, or teamwork was allowed, a notation on the assignment should indicate your collaboration.

Today!

Today

Introduction

- What are statistics?
- Importance of statistics
- Descriptive vs. Inferential Statistics

Background on some concepts

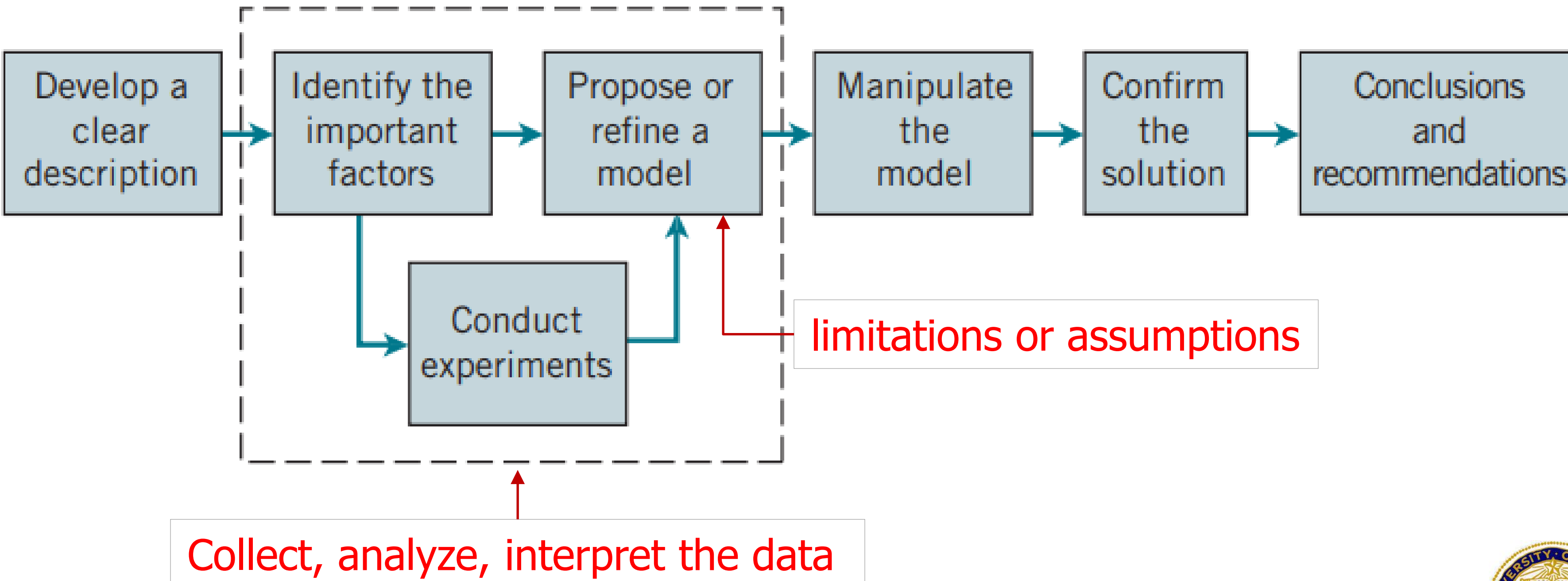
- Models
- Random variables
- Random experiments
- Sample spaces
- Events

I. Introduction

Engineering Method

- **An Engineer** formulates and solves problems by
    - Designing a new or refining an existing product/process
    - Making the efficient application of scientific principles

The engineering method:



What are Statistics?

Statistics: *Field of Study* that deals with collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes

Statistic: A ***single numerical value*** that represents a piece of data or a summary of data

Example: "The mean age of the participants in the study is a **statistic**."

Data → Statistics → Information

Statistics is the Science of Data!

Why bother??

Understanding Variability

- **Transportation:** Do you always get exactly the same mileage performance on every tank of fuel?

    - **Sources of variability:**
        - The type of driving (city versus highway)
        - The changes in the vehicle's condition over time
        - Valve wear
        - The brand/octane number of the gasoline used
        - Weather conditions

Why bother??

- **Variability:** The natural fluctuations or differences within a set of data or a process.
  - Slight variations could result serious system disruption

Example: Voice communication system

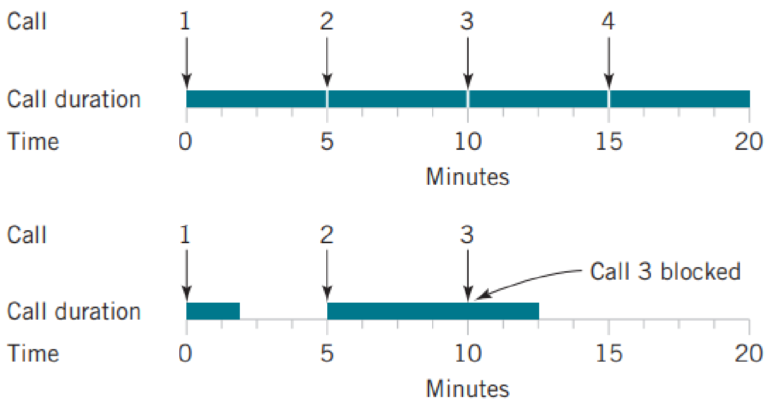


FIGURE 2.4

Variation causes disruptions in the system.

- **Sources of variability** include
  - Environmental factors
  - Individual differences
  - Operational conditions
  - Others

Role of Statistics

- Provides a framework for
  - Identifying, measuring, and quantifying this variability
  - Learning about
    - which potential sources of variability are the most important
    - which have the greatest impact on the outcome
  - Make informed decisions under uncertainty
  - Offering probabilistic models to estimate the likelihood of various outcomes

Statistical thinking - A useful way to incorporate variability into decision-making process!

Statistics...

- Rely upon the calculation of numbers
- Rely upon **how the numbers are chosen** and how the statistics are **interpreted**
- Sometimes have *problematic* interpretations

Statistics Example #1: Ice Cream

A new advertisement for Ben and Jerry's ice cream introduced in late May of last year resulted in a 30% increase in ice cream sales for the following three months. **Thus, the advertisement was effective.**

...Problems with this logic?

Statistics Example #2: Polling Stations and Crime

The more polling stations in a city, the more crime there is. **Thus, polling stations lead to crime.**

...Problems with this logic?

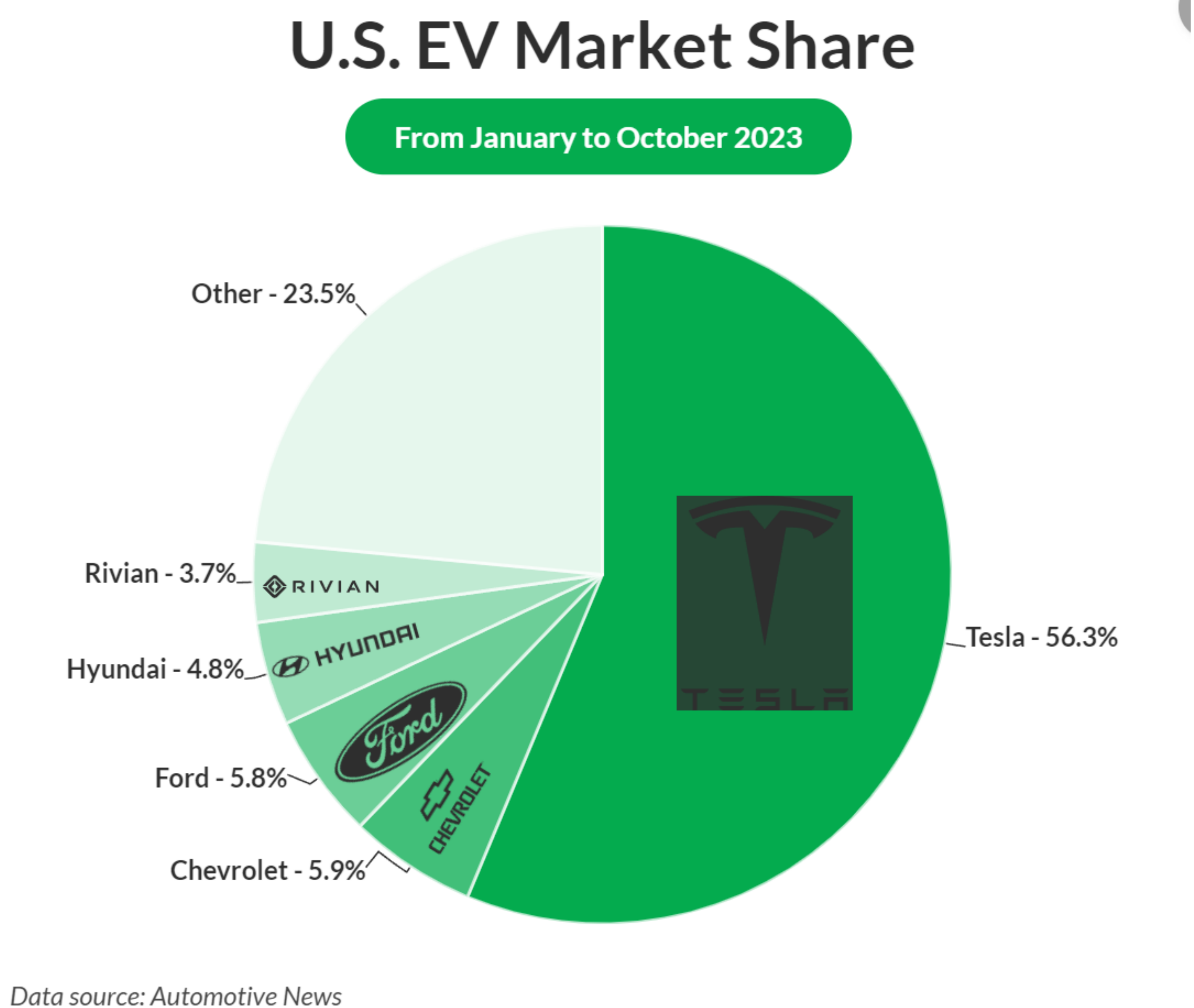
Concepts

Descriptive vs Inferential Statistics

Descriptive Statistics

- Used for organizing, summarizing, and describing data
  - Examples: percentages, frequencies, averages, graphs, tables

U.S. EV Market Share

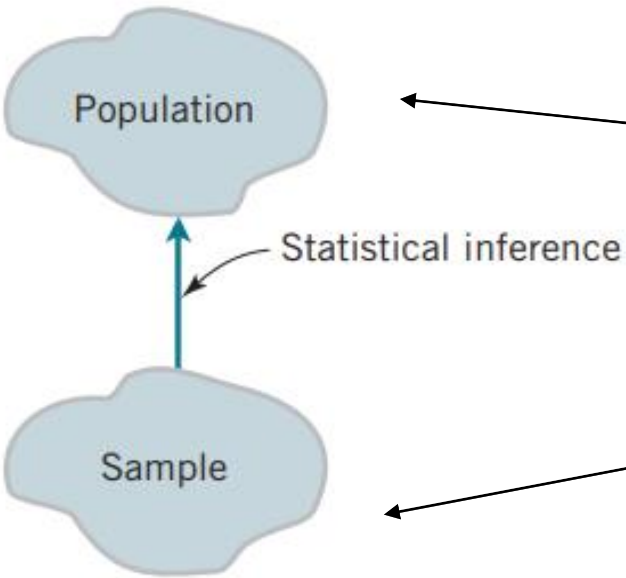


Inferential Statistics

- Inferential statistics can be used for two purposes:
  - To aid scientific understanding: **Estimating the probability** of a statement is True or False
  - To aid in making sound decisions: **Estimating which alternative** among a range of possibilities is most desirable.

Populations and Samples

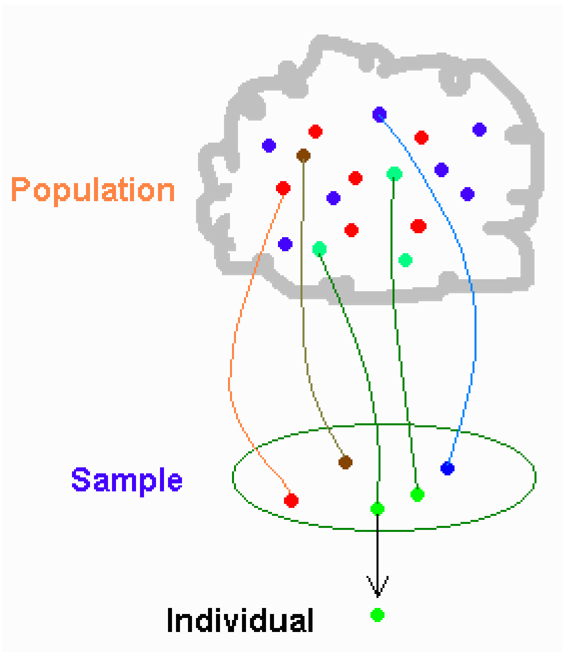
- Inferential Statistics (generalizing from a <u>sample</u> to the <u>population</u>)

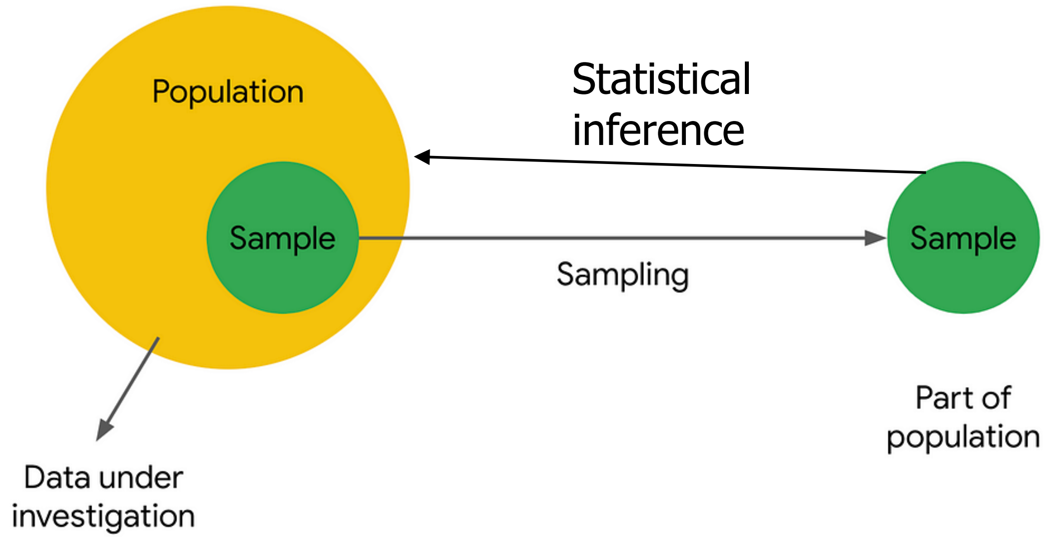


Population: the **entire** set of individuals or items that **share at least one property** of interest.

Sample: a **subset** of the population selected for measurement or observation

Population vs Sample





Sampling Example 1

- Fifty bottles of water were randomly selected from a large collection of bottles in a company's warehouse.
- These fifty bottles are referred to as the
  A. parameter.
  B. population.
  C. sample.
- The large collection of bottles is referred to as the
  A. parameter.
  B. population.
  C. sample.

- Which of the following statements is true regarding a population?
  A. It is a collection individuals, objects, or measurements
  B. It must be a large number of values
  C. It must refer to people

Sampling Example

- A substitute teacher wants to know how students in the class did on their last test
- She asks only the **10 students sitting in the front row** to report how they did on their last test and she concludes from them that **the class did extremely well**

- What is the sample?
- What is the population?
- Any problems?

Simple Random Sampling

- Every member of the population has an **equal chance** of being selected into the sample
- The selection of one member is **independent** from the selection of another member
- Thus, it is selection by **pure chance**

Decrease the bias in the sample!

Models

Models

- A **model** is a **simplified representation** of reality used to understand, analyze, predict, or control a complex system or phenomenon.

Real-world Bridge
(Physical Phenomenon)

- **Physical Model**

A scale model of the bridge in a wind tunnel

- **Mathematical Model**

Equations representing the forces and stresses on the bridge components.

- Example Equations:
  - $\text{Force} = \text{Mass} \times \text{Acceleration}$
  - $\text{Stress} = \frac{\text{Force}}{\text{Area}}$

- **Statistical Model**

Data collected on load patterns and environmental conditions to predict failure points

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_2$$

Models

Why Use Models?

- Understanding complex systems
- Prediction about future behavior and outcomes
- Identifying the best solutions and optimizing performance
- Clear and concise way to communication

Model structure: Variables and parameters

- What are Variables?
  - Symbols or names that represent values which can change within a given context or model
  - **Dynamic:** Variables can change and take on different values.
- **Types:**
  - **Continuous** variables (can take any value within a range),
  - **Discrete** variables (can take specific values).

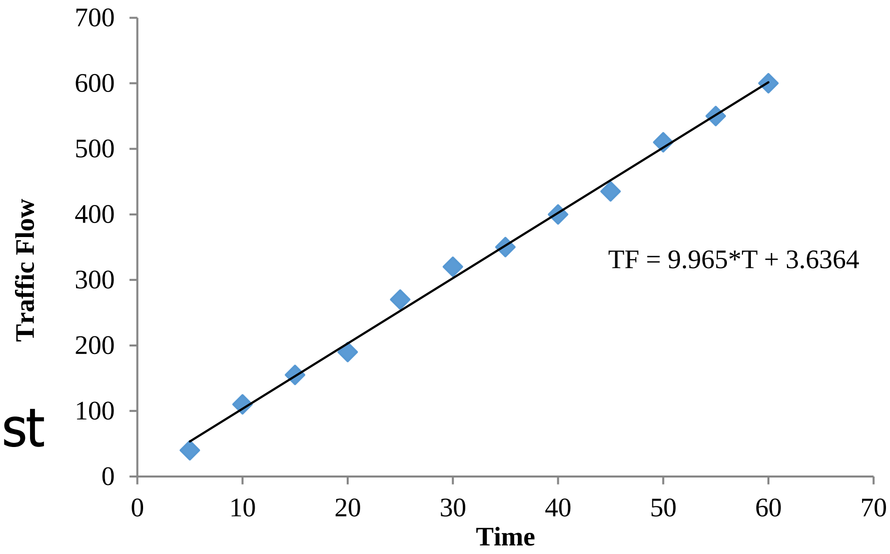
Dependent or Independent Variables:

- **Independent:** Not affected by other variables
- **Dependent:** Affected by other variables

Ex: In the equation:

$$y = mx + b$$

- x and y are **variables**
- x: **Model input** - Independent
- y: **Model output** - Dependent
- m, b: **Model parameters**, they adjust the general structure of the model



Mean, Median and Mode

Mean, Median and Mode

- **Mean** of the sample (N)

  Mean: $\bar{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$

- **Median** of the sample: **the middle number**
  - Arrange in ascending and descending order

  For odd N: $Y_1, Y_2, Y_3, Y_4, \underline{\mathbf{Y_5}}, Y_6, Y_7, Y_8, Y_9$

  For even N: $Y_1, Y_2, Y_3, Y_4, [\mathbf{(Y_5+Y_6)/2}]\ Y_7, Y_8, Y_9, Y_{10}$

- **Mode** of the sample

  The most frequent measurement

  [1,2,**3**,2,5,**3**,4,6,**3**]

Random Variables, Experiments, Spaces

Random Variables (RVs)

- A convenient way to think of a RV is by using the model:

$$X = \mu + \varepsilon$$

  - where μ is a constant and ε is a random disturbance

- μ: Identifies the central tendency of measurements.
- ε: Explains the variability around this central value.

Random Experiments

Definition:

- A **random experiment** is a process or procedure that generates a set of outcomes that are uncertain or unpredictable.

Characteristics:

- **Outcome Uncertainty:** The exact outcome cannot be predicted before the experiment.
- **Reproducibility:** The experiment can be repeated under the **same** conditions.
- **Sample Space:** The set of **all possible outcomes**

Examples:

- Rolling a die.
- Flipping a coin.

Sample Spaces

- **Sample Space:**
  - The set of **all possible outcomes** of a **random experiment**, denoted by S

- **Discrete Sample Space:** A sample space consists of a **finite** or **countable infinite** set of outcomes
  - e.g., $S = \{yes, no\}$, $S = \{yy, yn, ny, nn\}$

- **Continuous Sample Space:** A sample space contains an **interval** (either finite or infinite) of real numbers
  - e.g., $S = R^{+} = \{x | x > 0\}$, $S = \{x | 10 < x < 11\}$

Understanding sample space

- Consider an experiment that selects a cell phone camera and records the time of a camera flash:
- 1. Simply positive real time
- 2. Between 1.5 and 5 seconds
- 3. Objective: consider only whether the recycle time is low, medium, or high
- 4. Objective: only to evaluate **whether or not** a particular camera conforms to a minimum recycle-time specification

Tree Diagrams

- S can be described graphically with tree diagrams

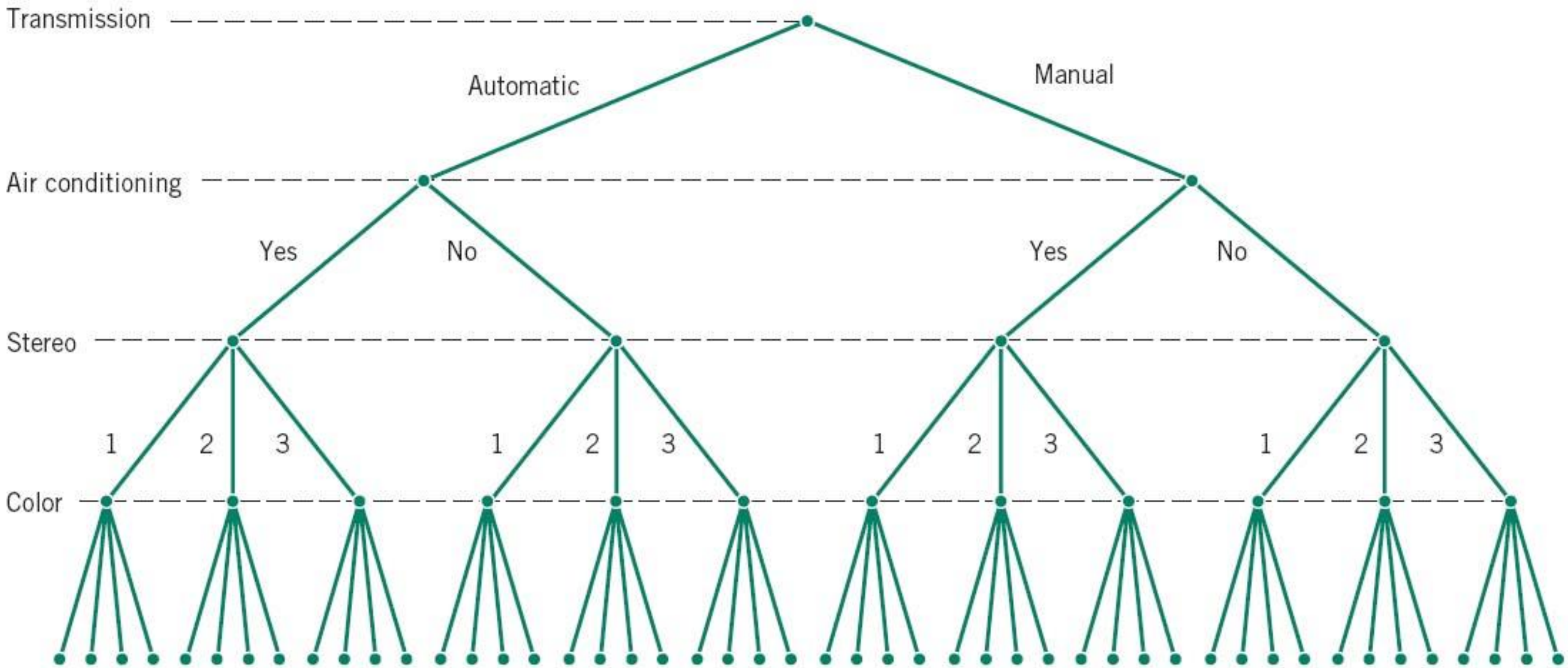
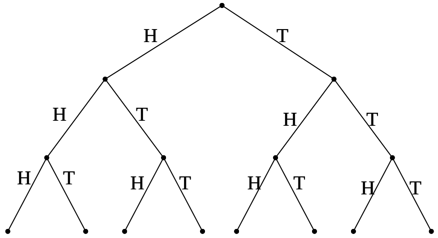


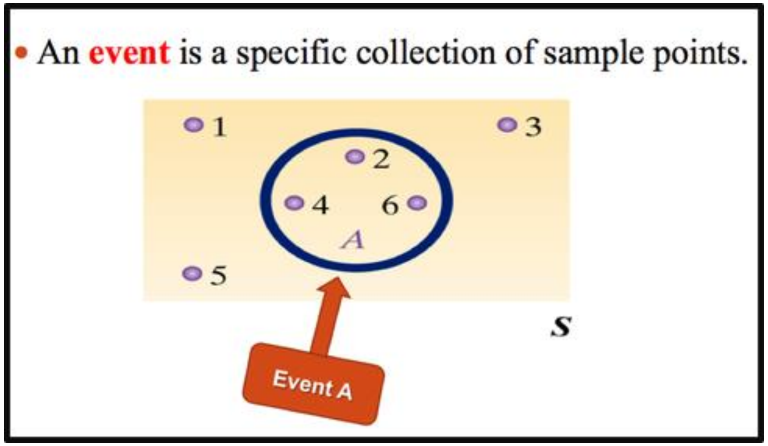
Figure 2-6 Tree diagram for different types of vehicles with 48 outcomes in the sample space.



Events

Events

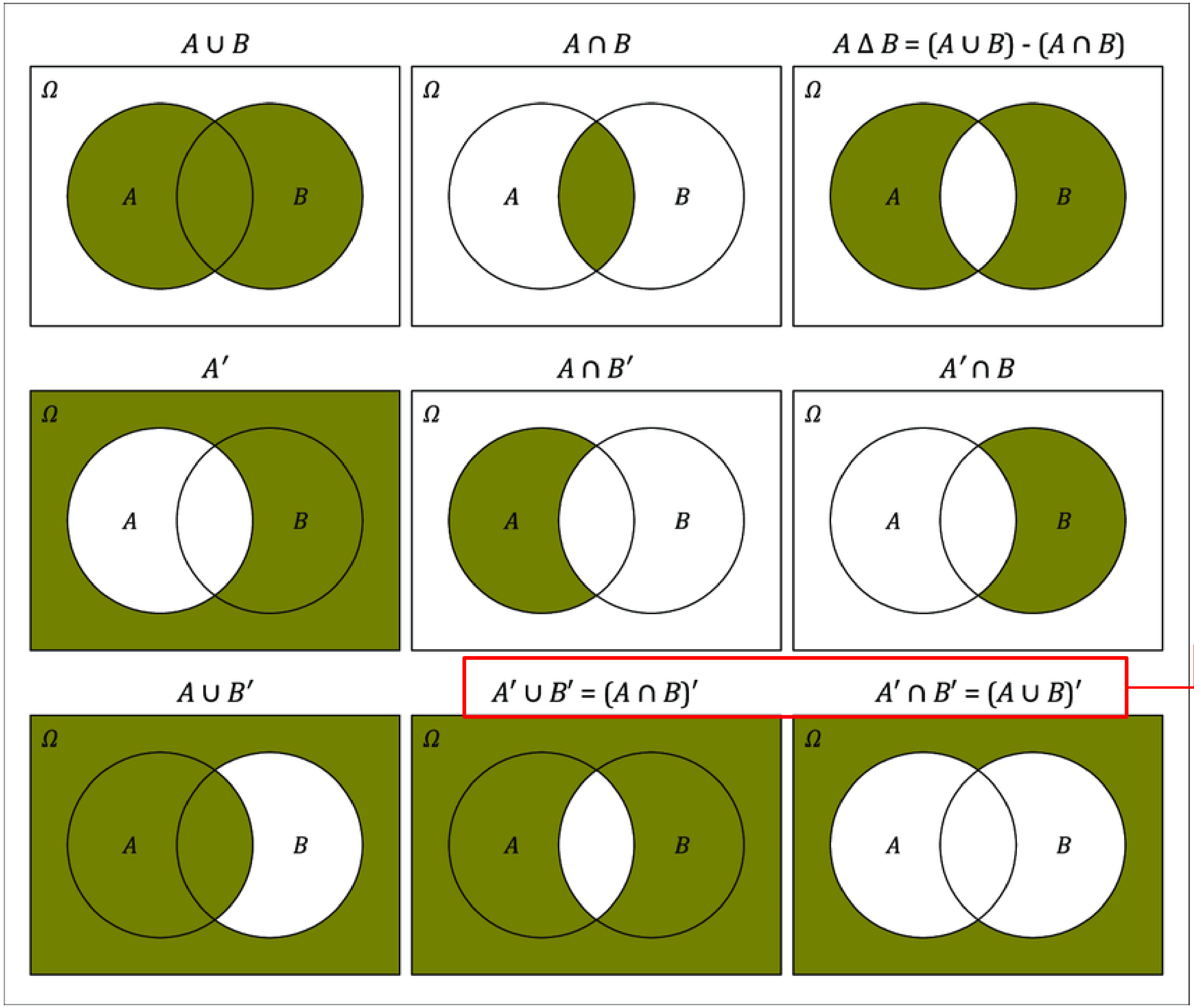
- An event is a **subset** of the sample space of a random experiment



- We can use basic set operations
  - Union of two events: $E_1 \cup E_2$
  - Intersection of two events: $E_1 \cap E_2$
  - Complement of an event E'

Operations on Events

- **Venn Diagrams** are used to portray relationships between events



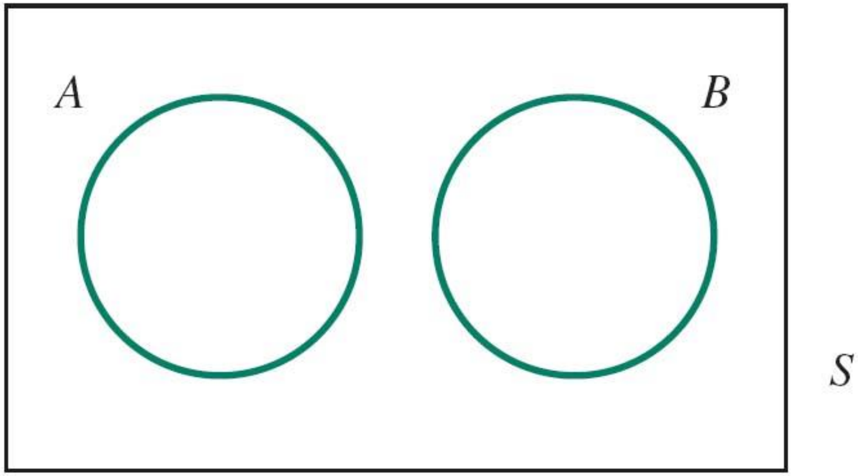
De Morgan's Law

Mutually Exclusive Events

- Two events such that:

$$E_1 \cap E_2 = \varnothing$$

are said to be mutually exclusive



Understanding Events in Sample Space

- Explore various events derived from
- Sample space S={yy,yn,ny,nn}

Defining Events

- Event E_1: At least one camera conforms.
  - $E_1 = \{yy, yn, ny\}$
- Event E_2: Both cameras do not conform.
  - $E_2 = \{nn\}$
- Event E_3: Null set (no outcomes).
  - $E_3 = \emptyset$
- Event E_4: The entire sample space.
  - $E_4 = S$
- Event E_5: Specific outcomes.
  - $E_5 = \{yn, ny, nn\}$

Operations on Events

- Union $E_1 \cup E_5$:
  - $E_1 \cup E_5 = S$
- Intersection $E_1 \cap E_5$:
  - $E_1 \cap E_5 = \{yn, ny\}$
- Complement E_1':
  - $E_1' = \{nn\}$

Summary

- The importance of Statistics
- The engineering or scientific method
- Examples of statistics
- Descriptive vs inferential statistics
- Review of concepts
  - Models
  - Mean, median and mode
  - Random variables, experiments and spaces

Next Class

- Counting Techniques
- Interpretation of Probability
- Addition
- Read Sections: 2-1, 2-2, 2-3