Chapter 1 Notes

Statistics Definitions

Global definition: Statistics involves collecting, organizing, summarizing, presenting, and analyzing data, as well as making inferences, conclusions, and decisions based on data.

Statistical definition: A statistic is a numerical value calculated from data (e.g. mean, proportion, standard deviation).

Probability vs Statistics:

- **Probability (deductive):** Population → Sample
- **Statistics (inductive):** Sample → Population

Basic Terminology

Individuals: Objects on which data are collected (people, animals, plots of land, etc.).
Variable: Any characteristic of an individual.
Population: The entire group of individuals of interest.
Sample: A subset of individuals taken from the population.
Statistical Inference: Drawing conclusions about a population based on a sample.

Sampling Methods

Simple Random Sample (SRS):

- Every possible group of size n has an equal chance of being selected.
- Helps avoid bias in sampling.
- Can be selected using random number tables or software.

Stratified Random Sampling:

- The population is divided into homogeneous groups *(individuals are similar with respect to the variable being studied)* called strata.
- A simple random sample is taken from each stratum. *(one subgroup of the population created)*
- Ensures that important subgroups are neither over nor under represented.

Types of Variables

Categorical Variable: Places individuals into categories (e.g. gender, major). These are qualitative.

Quantitative Variable: Takes numerical values for which arithmetic operations are meaningful.

- Discrete
- Continuous

Distributions

Distribution: Describes what values a variable takes and how often those values occur. When examining a distribution, look for:

- **Shape**
- **Center**
- **Spread**
- **Outliers**

Outlier: An individual value that falls outside the overall pattern of the data.

Describing Distributions with Numbers

Central Tendency: Describes where the data cluster or center.

- Mean: average value
- Median: middle value

The mean is more sensitive to extreme values than the median.

Changing a single data value will always change the mean, but may not change the median.

If a distribution is exactly symmetric, the mean and median are equal.

Trimmed Mean: The mean computed after removing extreme values. It is more resistant to outliers than the mean but uses more information than the median.

Measures of Spread

Range: Maximum minus minimum (very sensitive to extreme values).

Sample Variance:

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Standard Deviation:

$$\boxed{s = \sqrt{s^2}}$$

Degrees of Freedom: The number of independent pieces of information available to estimate variability.

Chapter 2, Jan 9th

Experiment: A process that generates an outcome.

Sample Space (S): The set of all possible outcomes of an experiment.

Example 1:

Select 3 items from a production line. Each item can be classified as either defective (D) or non-defective (N).

$$S = \{DDD, DDN, DND, NDD, DNN, NDN, NND, NNN\}$$

Since each item has 2 possible outcomes,

$$|S| = 2^3 = 8$$

Example 2:

$$S = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event (A): A subset of the sample space S.

Examples of events:

$$A = \{DDD, DDN, DND, NDD\}$$

$$B = \{NNN\}$$

$$C = \{(x, y) \mid x^2 + y^2 \leq 4\}$$

Event Operations:

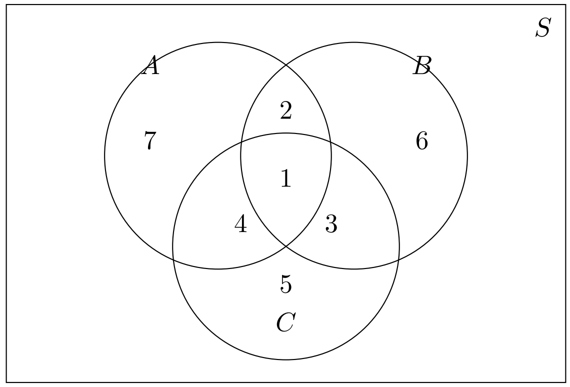
- Complement: A^c (or A')
- Intersection: $A \cap B$
- Union: $A \cup B$
- Null Event: $\varnothing$

If

$$A \cap B = \varnothing,$$

then A and B are mutually exclusive.

Example (Venn Diagram):



$$A = \{DDD, DDN, DND, NDD\}, \quad B = \{NNN\}$$

$$A \cup B = \{DDD, DDN, DND, NDD, NNN\}$$

$$A \cap B = \varnothing$$

Jan 12th

Chapter 2: January 12

Review

1. *Experiment: A process that generates an outcome.*
2. *Sample Space* (S)*: The set of all possible outcomes of an experiment.*
3. *Event Operations:*
   - *Complement:* A' (A^c)
   - *Intersection:* $A \cap B$
   - *Union:* $A \cup B$
   - *Null Event:* $\varnothing$
4. *If* $A \cap B = \varnothing$*, then* A *and* B *are called mutually exclusive.*

$$(A \cap B)' = A' \cup B'$$

$$(A \cup B)' = A' \cap B'$$

$$A \cap \varnothing = \varnothing$$

$$A \cup \varnothing = A$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Probability

$P(A)$ = probability of event A: the proportion of times the event occurs in infinitely many repetitions of the experiment.

Theorem2.1:

$$0 \leq P(A) \leq 1$$

$$P(A) + P(A') = 1$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\begin{aligned} P(A \cup B \cup C) = P(A) &+ P(B) + P(C) \\ &- P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &+ P(A \cap B \cap C) \end{aligned}$$

Mutually Exclusive Events

Definition: If $A_1, A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

If

$$A_1 \cup A_2 \cup \cdots \cup A_n = S,$$

then $\{A_1, A_2, \ldots, A_n\}$ is a partition of S.

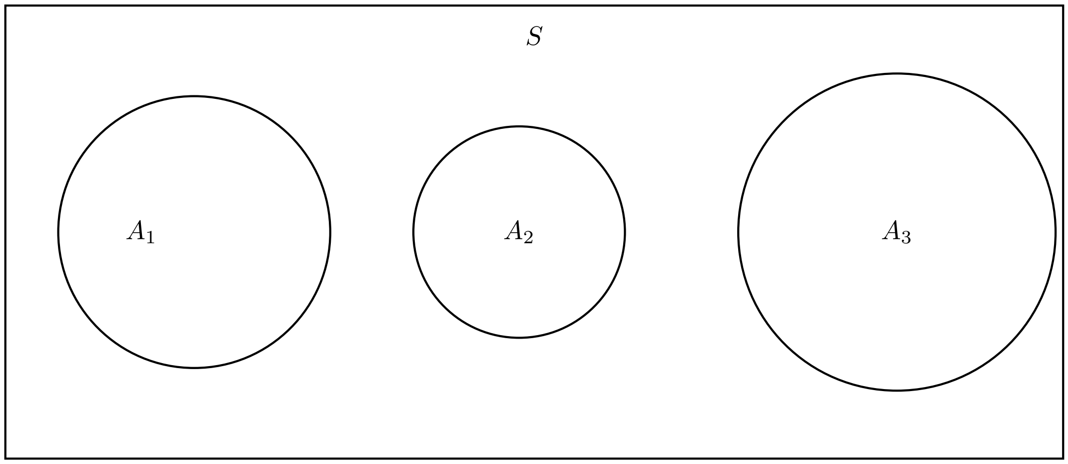


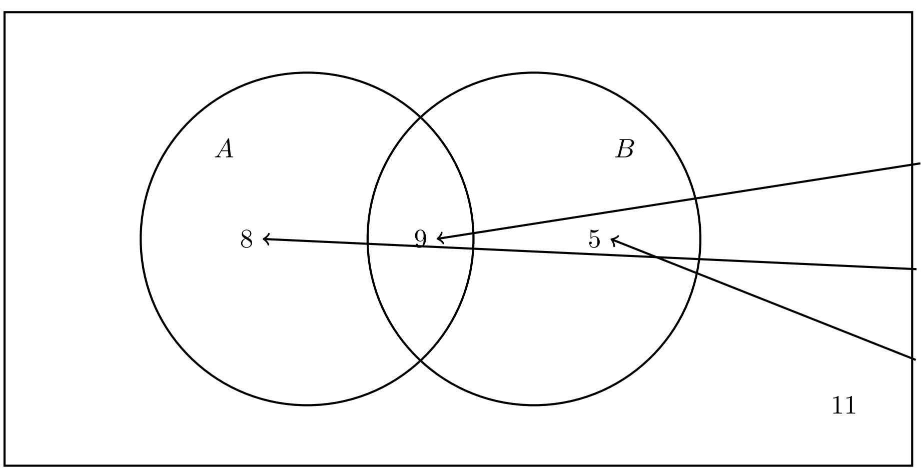
Figure 1: Partition of the sample space S into A_1, A_2, A_3

Example

In a class of 33 students:

- 17 earned an A on the midterm
- 14 earned an A on the final
- 11 earned no A on either exam

Find the probability that a randomly selected student earned A's on <u>both</u> exams.



$$|A \cup B| = 33 - 11 = 22$$

$$\begin{aligned}|A \cap B| &= |A| + |B| - |A \cup B| \\ &= 17 + 14 - 22 = 9\end{aligned}$$

$$|A \setminus B| = |A| - |A \cap B| = 17 - 9 = 8$$

$$|B \setminus A| = |B| - |A \cap B| = 14 - 9 = 5$$

$$P(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{9}{33} \approx 0.273$$

Figure 2: Events A: A on midterm, B: A on final, with region counts and calculations

Theorem 2.2 (Equally Likely Outcomes):

If the sample space S has a finite number of outcomes and all outcomes are equally likely, then for any event A,

$$P(A) = \frac{|A|}{|S|}$$

where
A: the event of interest (a subset of the sample space S),
S: the sample space, i.e. the set of all possible outcomes.

Example 1: Poker Hands Basics

A standard deck has:

$$4 \text{ suits} \times 13 \text{ denominations (A,2,3,\ldots,Q,K)} = 52 \text{ cards.}$$

A poker hand consists of 5 cards chosen from 52:

$$|S| = \binom{52}{5} = 2{,}598{,}960.$$

Combinations Reminder

If there are 3 objects $\{A, B, C\}$ and we choose 2:

$$\binom{3}{2} = \frac{3!}{(3-2)!2!}.$$

Order does <u>not</u> matter.

Example 2: Probability of 2 Aces and 1 Jack

A 5-card hand contains:

- exactly 2 aces,
- exactly 1 jack,
- 2 cards that are neither aces nor jacks.

$$P(\text{2 aces and 1 jack}) = \frac{\binom{4}{2}\binom{4}{1}\binom{44}{2}}{\binom{52}{5}}.$$

Example 3: Probability of a Full House

A full house consists of:

- 3 cards of one denomination
- 2 cards of a different denomination

Number of full house hands:

$$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}.$$

Thus,

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}.$$

Example 4: Probability of Four of a Kind

A four of a kind consists of:

- 4 cards of the same denomination
- 1 remaining card of a different denomination

Number of such hands:

$$\binom{13}{1}\binom{4}{4}\binom{48}{1}.$$

Thus,

$$P(\text{four of a kind}) = \frac{\binom{13}{1}\binom{4}{4}\binom{48}{1}}{\binom{52}{5}}.$$

Example 5: Probability of Exactly One Pair

An **excatly** one-pair hand consists of:

- 1 pair
- 3 cards of different denominations, none matching the pair

Number of such hands:

$$\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3.$$

Thus,

$$P(\text{exactly one pair}) = \frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3}{\binom{52}{5}}.$$

Note*: Counting Patterns

$$\binom{a}{b}$$

Meaning: Choose b different items from a **at once**, order does not matter.

Key features:

- No repeats
- Grouped choice
- Used when items must be distinct

$$\binom{a}{1}^b$$

Meaning: Make b independent choices, each time choosing 1 item from c.

Key features:

- Repeats allowed
- Choices are independent
- Used when selections do not restrict each other

Rule to Remember:

$$\text{Different items, no repeats } \Rightarrow \binom{a}{b}$$

$$\text{Independent choices } \Rightarrow \binom{c}{1}^b$$