Statistics and Probability

1. Statistics

A **population** is the whole set of items from which a data sample can be drawn; so a **sample** is only a portion of the population data. Features of a sample are described by **statistics**.

Statistical methods enable us to arrange, analyze and interpret the sample data obtained from a population. We gather sample data when it is impractical to analyze the population data as a smaller sample often allows us to gain a better understanding of the population without doing too much work or wasting precious time.

2. Measures of Central Tendency

We make inferences about a population from a sample set of observed values by finding the mean, median and mode. The mean, median and mode are collectively known as **measures of central tendency**.

1) Mean

The **mean** (or average) of a set of values is defined as the sum of all the values divided by the number of values. That is:

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

Symbolically, $\bar{x} = \dfrac{\sum x}{n}$

where $\bar{x}$ (read as 'x bar') is the mean of the set of x values,
$\sum x$ is the sum of all the x values, and
n is the number of x values.

Example: The marks of five candidates in a mathematics test with a maximum possible mark of 20 are given below:
14 12 18 17 13
Find the mean value.

Solution:

$$\bar{x} = \frac{\sum x}{n} = \frac{14+12+18+17+13}{5} = \frac{74}{5} = 14.8$$

So, the mean mark is 14.8

2) Median

The **median** is the middle value of the data set arranged in ascending order of magnitude.

E.g. the median of 3, 5 and 9 is 5 as it is the middle value.

In general:

Median $= \frac{1}{2}(n+1)$ th value, where n is the number of data values in the sample.

If the number of values in the data set is even, then the median is the average of the two middle values.

Example: Find the median of the following scores:

10 16 14 19 8 11

Solution:

Arrange the values in ascending order of magnitude:

8 10 11 14 16 19

There are 6 values in the data set.

$\therefore n = 6$

$$\text{Now, median} = \left(\frac{n+1}{2}\right)\text{th value where } n = 6$$

$$= \left(\frac{6+1}{2}\right) = \frac{7}{2} = 3.5\text{th value}$$

The third and fourth values, 11 and 14, are in the middle. That is, there is no one middle value.

$$\therefore \text{Median} = \frac{11+14}{2} = \frac{25}{2} = 12.5$$

Note: Half of the values in the data set lie below the median and half lie above the median.

3) Mode

The **mode** is the value (or values) that occurs most often.

E.g. the mode of the data set {3, 5, 6, 7, 7, 7, 8, 8, 9} is 7 as it occurs most often.

Example: The marks awarded to seven pupils for an assignment were as follows:

18 14 18 15 12 19 18

a. Find the median mark.

b. State the mode.

Solution:

a. Arrange the marks in ascending order of magnitude:

12 14 15 18 18 18 19

$$\text{Now, median} = \left(\frac{n+1}{2}\right)\text{th value where } n = 7$$
$$= \left(\frac{7+1}{2}\right)\text{th value} = \frac{8}{2}\text{th value} = 4\text{th value} = 18$$

Note:
The fourth mark, 18, is the middle data value in this arrangement.
$\therefore$ Median = 18 $\qquad$ {$\because$ 18 is the middle data value}

b. 18 is the mark that occurs most often.
$\therefore$ Mode = 18

3. Analyzing Data

We have used stem-plots to organize and display data sets. However, stem-plots do not provide enough information about the data. So, we use the range, mean, median, mode, quartiles and interquartile range to obtain additional information concerning the data. Then a box-plot is drawn to display the range, interquartile range, lower quartile, upper quartile and the median of the data set.

Range is defined as the difference between the highest and lowest scores. That is:
Range = Highest score – Lowest score

Example: Find the range of the following data set:
7 3 12 34 62

Solution:
Lowest score = 3
Highest score = 62
Range = Highest score − Lowest score = 62 − 3 = 59

Note: Range is a **measure of spread** and it tells us how much a data set is spread out or scattered. The mean, median and mode are called the **measures of central tendency** because these give us an indication of the middle value around which a set of values have a tendency to cluster.

4. Probability

An event (E) is a subset of the sample space. That is, an event is a subset of all possible outcomes. We refer to this subset of outcomes as **favorable outcomes**.

For example, the sample space for an experiment of tossing a fair coin is $S = \{H, T\}$, and the the two possible outcomes are the events $E_1 = \{H\}$ and $E_2 = \{T\}$.

Note that E_1 is the event that 'a head falls' and E_2 is the event that 'a tail falls'.

The probability of event E occurring is given by

$$\Pr(E) = \frac{\text{Number of outcomes in event } E}{\text{Number of outcomes in sample space } S}$$

This is often written as:

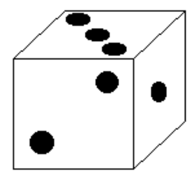
$$\Pr(E) = \frac{n(E)}{n(S)}$$

This result holds only if the outcomes of an experiment are equally likely.

Note: The events are denoted by capital letters *A*, *B*, *C*, *D*, *E*, ...

Example: A die is rolled. Find:

a. the sample space for this experiment
b. the probability of obtaining an odd number
c. the probability of obtaining a number greater than 4.



Solution:

a. $S = \{1, 2, 3, 4, 5, 6\}$

b. Let A be the event that an odd number is obtained.

$\therefore A = \{1, 3, 5\}$

$$\Pr(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

c. Let B be the event that a number greater than 4 is obtained.

$\therefore B = \{5, 6\}$

$$\Pr(B) = \frac{n(B)}{n(S)} = \frac{2}{6} = \frac{1}{3}$$

Example: A pack of 52 playing cards consists of four suits, i.e. clubs, spades, diamonds and hearts. Each suit has 13 cards which are the 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, king and the ace card. Clubs and spades are of black color whereas diamonds and hearts are of red color. So, there are 26 red cards and 26 black cards.

Find the probability of drawing from a well-shuffled pack of cards:

a. a red card
b. the king of spades
c. an ace

Solution:

a. A pack of 52 cards has 26 red cards.

$$\Pr(\text{a red card}) = \frac{26}{52} = \frac{1}{2}$$

b. A pack of 52 cards has 1 king of spades.

$\therefore$ Pr(the king of spades) $= \dfrac{1}{52}$

c. A pack of 52 cards has 4 aces.

$\therefore$ Pr(an ace) $= \dfrac{4}{52} = \dfrac{1}{13}$

5. Range of Probability

If an event is **impossible**, its probability is 0. If an event is **certain** to occur, its probability is 1. The probability of any other event is between these two values. That is:

- Pr(impossible event) = 0
- Pr(certain event) = 1
- If A is any event, then $0 \le \Pr(A) \le 1$.

Event:	Impossible Event	Even Chance	Certain
Probability:	0	$\frac{1}{2}$	1

Example: A die is rolled. Find the probability of obtaining:

a. a 10

b. a number less than or equal to 6

Solution:

$S = \{1, 2, 3, 4, 5, 6\}$

a. It is impossible to obtain a 10.

$\therefore$ Pr(a 10) = 0

b. Let A be the event that a number less than or equal to 6 is obtained. Then:

$A = \{1, 2, 3, 4, 5, 6\}$

Now, $\Pr(A) = \dfrac{\text{Number of outcomes in } A}{\text{Number of outcomes in } S} = \dfrac{6}{6} = 1$

6. Independent Events

Definition: Two events, A and B, are **independent** if the fact that A occurs does not affect the probability of B occurring.

Some other examples of independent events are:

- Landing on heads after tossing a coin **AND** rolling a 5 on a single 6-sided die.
- Choosing a marble from a jar **AND** landing on heads after tossing a coin.
- Choosing a 3 from a deck of cards, replacing it, **AND** then choosing an ace as the second card.
- Rolling a 4 on a single 6-sided die, **AND** then rolling a 1 on a second roll of the die.

To find the probability of two independent events that occur in sequence, find the probability of each event occurring separately, and then multiply the probabilities. This multiplication rule is defined symbolically below. Note that multiplication is represented by AND.

Multiplication Rule:
When two events, A and B, are **independent**, the probability of both occurring is:
P(A and B) = P(A) · P(B)

Example
A school survey found that 9 out of 10 students like pizza. If three students are chosen at random with replacement, what is the probability that all three students like pizza?

P(student 1 likes pizza) $= \frac{9}{10}$

P(student 2 likes pizza) $= \frac{9}{10}$

P(student 3 likes pizza) $= \frac{9}{10}$

P(student 1 and student 2 and student 3 like pizza) $= \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{9}{10} = \frac{729}{1000}$

7. Conditional Probability - *How to handle* ***Dependent Events***

Dependent Event - they can be affected by previous events.

P(B|A) means "Event B **given** Event A"

In other words, event A has already happened, now what is the chance of event B?

P(B|A) is also called the "Conditional Probability" of B given A.

Formula:

"Probability Of" "Given"

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$

Event A Event B

"Probability of ***event A and event B*** *equals*
the probability of ***event A*** *times the probability of* ***event B given event A****"*

Rearrange, we have:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example 1: What is the probability of drawing 2 Kings from a deck of cards?

Let Event A be drawing a King first, and Event B be drawing a King second.
For the first card the chance of drawing a King is 4 out of 52
P(A) = 4/52

But after removing a King from the deck the probability of the 2nd card drawn is less likely to be a King (only 3 of the 51 cards left are Kings):
P(B|A) = 3/51

And so: P(A and B) = P(A) x P(B|A) = (4/52) x (3/51) = 12/2652 = 1/221
So the chance of getting 2 Kings is 1 in 221, or about 0.5%

Example 2: 70% of your friends like Chocolate, and 35% like Chocolate AND like Strawberry. What percent of those who like Chocolate also like Strawberry?

P(Strawberry|Chocolate) = P(Chocolate and Strawberry) / P(Chocolate)
0.35 / 0.7 = 50%

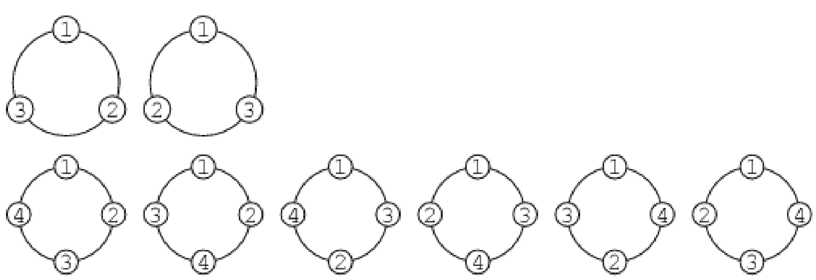
50% of your friends who like Chocolate also like Strawberry

8. Circular Permutation

The number of ways to arrange n distinct objects along a fixed (i.e., cannot be picked up out of the plane and turned over) circle is

$P_n = (n-1)!$.

The number is $(n-1)!$ instead of the usual factorial $n!$ since all cyclic permutations of objects are equivalent because the circle can be rotated.



For example, of the $3! = 6$ permutations of three objects, the $(3-1)! = 2$ distinct circular permutations are $\{1, 2, 3\}$ and $\{1, 3, 2\}$. Similarly, of the $4! = 24$ permutations of four objects, the $(4-1)! = 6$ distinct circular permutations are$\{1, 2, 3, 4\}$, $\{1, 2, 4, 3\}$, $\{1, 3, 2, 4\}$, $\{1, 3, 4, 2\}$, $\{1, 4, 2, 3\}$, and $\{1, 4, 3, 2\}$. Of these, there are only three free permutations (i.e., inequivalent when flipping the circle is allowed): $\{1, 2, 3, 4\}$, $\{1, 2, 4, 3\}$, and $\{1, 3, 2, 4\}$. The number of free circular permutations of order n is $P'_n = 1$ for $n = 1$, 2, and

$P'_n = \frac{1}{2}(n-1)!$
for $n \geq 3$, giving the sequence 1, 1, 1, 3, 12, 60, 360, 2520, ...

Questions in class

1. Three balls numbered 1, 2, and 3 are placed in a bag. A ball is drawn from the bag and the number is recorded. The ball is returned to the bag. After this has been done three times, what is the probability that the sum of the three recorded numbers is less than 8?

2. Twenty tickets are numbered from one to twenty. One ticket is drawn at random with each ticket having an equal chance of selection. What is the probability that the ticket shows a number that is a multiple of 3 or 5?

3. A driver approaching a toll booth has exactly two quarters, two dimes and two nickels in his pocket. He reaches into his pocket and randomly selects two of these coins. What is the probability that the coins that he selects will be at least enough to pay the 30-cent toll?

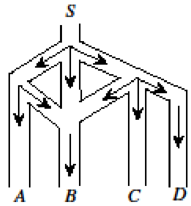
4. An unusual die has the numbers 2, 2, 3, 3, 5, and 8 on its six faces. Two of these dice are rolled. The two numbers on the top faces are added. How many different sums are possible?

5. A bag contains eight yellow marbles, seven red marbles, and five black marbles. Without looking in the bag, Igor removes N marbles all at once. If he is to be sure that, no matter which choice of N marbles he removes, there are at least four marbles of one color and at least three marbles of another color left in the bag, what is the maximum possible value of N?

6. Alice rolls a standard 6-sided die. Bob rolls a second standard 6-sided die. Alice wins if the values shown differ by 1. What is the probability that Alice wins?

7. In a dice game, a player rolls two dice. His score is the larger of the two numbers on the dice. For example, if he rolls 3 and 5, his score is 5, and if he rolls 4 and 4, his score is 4. What is the probability that his score is 3 or less?

8. Harry the Hamster is put in a maze, and he starts at point S. The paths are such that Harry can move forward only in the direction of the arrows. At any junction, he is equally likely to choose any of the forward paths. What is the probability that Harry ends up at B?



9. A die is a cube with its faces numbered 1 through 6. One red die and one blue die are rolled. The sum of the numbers on the top face of each die is determined. What is the probability that this sum is a perfect square?

10. The digits 2, 2, 3, and 5 are randomly arranged to form a four digit number. What is the probability that the sum of the first and last digits is even?