Probability (1)

1. 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:
$\text{Median} = \dfrac{1}{2}(n+1)$ 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.
$\therefore \text{ Median} = 18 \qquad \{\because 18 \text{ is the middle data value}\}$

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

2. Analyzing Data

We have used stem-plots to organise 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:

$\text{Lowest score} = 3$

$\text{Highest score} = 62$

$\text{Range} = \text{Highest score} - \text{Lowest score} = 62 - 3 = 59$

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.

3. 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 **favourable 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:

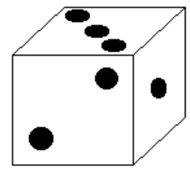
$$\Pr(E) = \frac{n(E)}{n(S)}$$

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 1: 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 2: 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 colour whereas diamonds and hearts are of red colour. 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.

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

c. A pack of 52 cards has 4 aces.

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

4. 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:

- $\Pr(\text{impossible event}) = 0$
- $\Pr(\text{certain event}) = 1$
- 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.

$\therefore \Pr(\text{a } 10) = 0$

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

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

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

5. 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 1: 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) $= \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{9}{10} = \frac{729}{1000}$

Example 2: A bag contains two red beads and two green beads. You reach into the bag and pull out a bead, replacing it with a red bead regardless of the color you pulled out. What is the probability that all beads in the bag are red after three such replacements?

Solution: The beads will all be red at the end of the third draw precisely when two green beads are chosen in the three draws. If the first bead drawn is green, then there will be one green and three red beads in the bag before the second draw.

So the probability that green beads are drawn in the first two draws is

$$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

The probability that a green bead is chosen, then a red bead, and then a green bead, is

$$\frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{32}$$

Finally, the probability that a red bead is chosen then two green beads is

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{16}$$

The sum of these probabilities is

$$\frac{1}{8} + \frac{3}{32} + \frac{1}{16} = \frac{9}{32}$$

6. Combinations and Permutations

In English we use the word "combination" loosely, without thinking if the order of things is important. In other words:

"My fruit salad is a combination of apples, grapes and bananas" We don't care what order the fruits are in, they could also be "bananas, grapes and apples" or "grapes, apples and bananas", it's the same fruit salad.

"The combination to the safe was 472". Now we do care about the order. "724" would not work, nor would "247". It has to be exactly 4-7-2.

If the order doesn't matter, it is a **Combination.**

If the order does matter it is a **Permutation.**

In other words: **A Permutation is an ordered Combination.**

1) Permutations

There are basically two types of permutation:

a) Repetition is Allowed: such as the lock above. It could be "333".
b) No Repetition: for example the first three people in a running race. You can't be first and second.

a) Permutations with Repetition

When you have n things to choose from ... you have n choices each time!
When choosing r of them, the permutations are: **n × n × ... (r times)**
(In other words, there are n possibilities for the first choice, THEN there are n possibilities for the second choice, and so on, multiplying each time.)
Which is easier to write down using an exponent of r: **n × n × ... (r times) =** $\mathbf{n^r}$

Example: in the lock above, there are 10 numbers to choose from (0, 1 … 9) and you choose 3 of them:
10 × 10 × ... (3 times) = 103 = 1,000 permutations

b) Permutations without Repetition

In this case, you have to reduce the number of available choices each time.
For example, what order could 16 pool balls be in?
After choosing, say, number "14" you can't choose it again.
So, your first choice would have 16 possibilities, and your next choice would then have 15 possibilities, then 14, 13, etc. And the total permutations would be:
16 × 15 × 14 × 13 × ... = 20,922,789,888,000
But maybe you don't want to choose them all, just 3 of them, so that would be only:
16 × 15 × 14 = 3,360

In other words, there are 3,360 different ways that 3 pool balls could be selected out of 16 balls.
But how do we write that mathematically? Answer: we use the "factorial function"

> **!** The factorial function (symbol: !) just means to multiply a series of descending natural numbers. Examples:
> 4! = 4 × 3 × 2 × 1 = 24
> 7! = 7 × 6 × 5 × 4 × 3 × 2 × 1 = 5,040
> 1! = 1
>
> Note: it is generally agreed that 0! = 1. It may seem funny that multiplying no numbers together gets you 1, but it helps simplify a lot of equations.

So, if you wanted to select all of the billiard balls the permutations would be:
16! = 20,922,789,888,000
But if you wanted to select just 3, then you have to stop the multiplying after 14. How do you do that?
There is a neat trick ... you divide by 13! ...

$$\frac{16 \times 15 \times 14 \times 13 \times 12 \ldots}{13 \times 12 \ldots} = 16 \times 15 \times 14 = 3{,}360$$

Do you see? 16! / 13! = 16 × 15 × 14

The formula is written:

$$\frac{n!}{(n-r)!}$$

where n is the number of things to choose from, and you choose r of them
(No repetition, order matters)

Examples: Our "order of 3 out of 16 pool balls example" would be:

$$\frac{16!}{(16-3)!} = \frac{16!}{13!} = \frac{20{,}922{,}789{,}888{,}000}{6{,}227{,}020{,}800} = 3{,}360$$

(which is just the same as: 16 × 15 × 14 = 3,360)
How many ways can first and second place be awarded to 10 people?

$$\frac{10!}{(10-2)!} = \frac{10!}{8!} = \frac{3{,}628{,}800}{40{,}320} = 90$$

(which is just the same as: 10 × 9 = 90)
Notation for Permutation

$$P(n,r) = {}^{n}P_{r} = {}_{n}P_{r} = \frac{n!}{(n-r)!}$$

Example: P(10,2) = 90

c) Permutations with Repeated Elements

Assume that we have an alphabet with k letters and we want to write all possible words containing n_1 times the first letter of the alphabet, n_2 times the second letter, ..., n_k times the kth letter. How many words can we write? We call this number $P(n; n_1, n_2, \ldots, n_k)$, where $n = n_1 + n_2 + \cdots + n_k$.

Example: With 3 a's and 2 b's we can write the following 5-letter words: *aaabb, aabab, abaab, baaab, aabba, ababa, baaba, abbaa, babaa, bbaaa.*

We may solve this problem in the following way, as illustrated with the example above. Let us distinguish the different copies of a letter with subscripts: $a_1 a_2 a_3 b_1 b_2$. Next, generate each permutation of this five elements by choosing 1) the position of each kind of letter, then 2) the subscripts to place on the 3 a's, then 3) these subscripts to place on the 2 b's. Task 1) can be performed in $P(5; 3, 2)$ ways, task 2) can be performed in 3! ways, task 3) can be performed in 2!. By the product rule we have $5! = P(5; 3, 2) \times 3! \times 2!$, hence $P(5; 3, 2) = 5!/(3!2!)$.

In general the formula is:

$$P(n; n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!\, n_2! \, \ldots \, n_k!} \, .$$

2) Combinations

There are also two types of combinations (remember the order does not matter now):

a) No Repetition: such as lottery numbers (2,14,15,27,30,33)
b) Repetition is Allowed: such as coins in your pocket (5,5,5,10,10)

a) Combinations without Repetition

This is how lotteries work. The numbers are drawn one at a time, and if you have the lucky numbers (no matter what order) you win!

The easiest way to explain it is to assume that the order does matter (ie permutations), then alter it so the order does not matter.

Going back to our pool ball example, let us say that you just want to know which 3 pool balls were chosen, not the order.

We already know that 3 out of 16 gave us 3,360 permutations.

But many of those will be the same to us now, because we don't care what order!

For example, let us say balls 1, 2 and 3 were chosen. These are the possibilities:

Order does matter	Order doesn't matter
1 2 3 1 3 2 2 1 3 2 3 1 3 1 2 3 2 1	1 2 3

So, the permutations will have 6 times as many possibilities.

In fact there is an easy way to work out how many ways "1 2 3" could be placed in order, and we have already talked about it. The answer is:

3! = 3 × 2 × 1 = 6

So, all we need to do is adjust our permutations formula to reduce it by how many ways the objects could be in order (because we aren't interested in the order any more):

$$\frac{n!}{(n-r)!} \times \frac{1}{r!} = \frac{n!}{r!(n-r)!}$$

That formula is so important it is often just written in big parentheses like this:

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

where n is the number of things to choose from, and you choose r of them
(No repetition, order doesn't matter)

It is often called "n choose r" (such as "16 choose 3")

Notation for Combination

$$C(n,r) = {}^nC_r = {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example: Our pool ball example (now without order) is:

$$\frac{16!}{3!(16-3)!} = \frac{16!}{3!\times 13!} = \frac{20{,}922{,}789{,}888{,}000}{6\times 6{,}227{,}020{,}800} = 560$$

Or you could do it this way:

$$\frac{16\times 15\times 14}{3\times 2\times 1} = \frac{3360}{6} = 560$$

It is interesting to also note how this formula is nice and symmetrical:

$$\frac{n!}{r!(n-r)!} = \binom{n}{r} = \binom{n}{n-r}$$

In other words choosing 3 balls out of 16, or choosing 13 balls out of 16 have the same number of combinations.

$$\frac{16!}{3!(16-3)!} = \frac{16!}{13!(16-13)!} = \frac{16!}{3!\times 13!} = 560$$

b) Combinations with Repetition

Assume that we have a set A with n elements. Any selection of r objects from A, where each object can be selected more than once, is called a *combination of n objects taken r at a time with repetition*. For instance, the combinations of the letters *a, b,* c, *d* taken 3 at a time with repetition are: *aaa, aab, aac, aad,abb , abc, abd , acc, acd , add,bbb , bbc, bbd , bcc ,bcd , bdd , ccc,ccd, cdd, ddd.* Two combinations with repetition are considered identical if they have the same elements repeated the same number of times, regardless of their order.

Example: Assume that we have 3 different (empty) milk containers and 7 quarts of milk that we can measure with a one quart measuring cup. In how many ways can we distribute the milk among the three containers?

We solve the problem in the following way. Let x_1, x_2, x_3 be the quarts of milk to put in containers number 1, 2 and 3 respectively. The number of possible distributions of milk equals the number of non negative integer solutions for the equation $x_1 + x_2 + x_3 = 7$. Instead of using numbers for writing the solutions, we will use strokes, so for instance we represent the solution $x_1 = 2, x_2 = 1, x_3 = 4$, or 2 + 1 + 4, like this: II+I+IIII. Now, each possible solution is an arrangement of 7 strokes and 2 plus signs, so the number of arrangements is $P(9; 7, 2) = 9!/(7!2!)$

The general solution is:

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

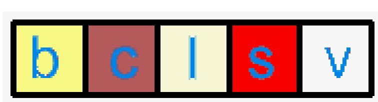
where ***n*** is the number of things to choose from, and you choose ***r*** of them
(Repetition allowed, order doesn't matter)

Example: Let us say there are five flavors of ice cream: **banana, chocolate, lemon, strawberry and vanilla**. You can have three scoops. How many variations will there be?

Let's use letters for the flavors: {b, c, l, s, v}. Example selections would be:

- {c, c, c} (3 scoops of chocolate)
- {b, l, v} (one each of banana, lemon and vanilla)
- {b, v, v} (one of banana, two of vanilla)

(And just to be clear: There are **n=5** things to choose from, and you choose **r=3** of them.
Order does not matter, and you **can** repeat!)



Think about the ice cream being in boxes, you could have 5-1 = 4 separators and between separators, you can put 1, 2, or 3 scoops. Therefore, you will have 4 separators + 3 scoops = 7 things to arrange.

This is like the permutation with repeated elements.

$$\frac{7!}{3!\times 4!} = \frac{5040}{6\times 24} = 35$$

By using the formula $\binom{n+r-1}{r} = \binom{n+r-1}{n-1} = \frac{(n+r-1)!}{r!(n-1)!}$ we have:

$$\frac{(5+3-1)!}{3!(5-1)!} = \frac{7!}{3!\times 4!} = \frac{5040}{6\times 24} = 35$$

►Questions in class

1. Four distinct circles are drawn in a plane. What is the maximum number of points where at least two of the circles intersect?

2. What is the probability that a randomly drawn positive factor of 60 is less than 7?

3. Coin *A* is flipped three times and coin *B* is flipped four times. What is the probability that the number of heads obtained from flipping the two fair coins is the same?

4. Three tiles are marked X and two other tiles are marked O. The five tiles are randomly arranged in a row. What is the probability that the arrangement reads XOXOX?

5. Integers a, b, c, and d, not necessarily distinct, are chosen independently and at random from 0 to 2007, inclusive. What is the probability that ad - bc is even?

6. Coming out of the grocery store, Ebree has eight coins, of which none is a half-dollar, that add up to $1.45. Unfortunately, on the way home she loses one of them. If the chances of losing a quarter, dime or nickel are equal, which coin is most probably lost?

7. Two different numbers are selected at random from {1, 2, 3, 4, 5} and multiplied together. What is the probability that the product is even?

8. Three distinct integers are selected at random between 1 and 2016, inclusive. Which of the following is a correct statement about the probability p that the product of the three integers is odd?

(A) $p < \frac{1}{8}$ (B) $p = \frac{1}{8}$ (C) $\frac{1}{8} < p < \frac{1}{3}$ (D) $p = \frac{1}{3}$ (E) $p > \frac{1}{3}$