Unit 2) Data Representation and Sampling Technique

❖Graphical representation of data

As more and more data are available to us today, there are several varieties of charts and graphs than before. In reality, the amount of data that we produce, acquire, copy, and use now will be nearly doubled by 2025. Data visualisation is therefore crucial and serves as a powerful tool for organisations. One can benefit from graphs and charts in the following ways:

- Encouraging the group to act proactively.
- Showcasing progress toward the goal to the stakeholders.
- Displaying core values of a company or an organization to the audience.

Moreover, data visualisation can bring heterogeneous teams together around new objectives and foster the trust among the team members. Let us discuss about various. Here we want to discuss three graphs.

1. Histogram
2. Box plots
3. Scatter plots

1. Histogram

A histogram visualises the distribution of data across distinct groups with continuous classes. It is represented with set of rectangular bars with widths equal to the class intervals and areas proportional to frequencies in the respective classes. A histogram may hence be defined as a graphic of a frequency distribution that is grouped and has continuous classes. It provides an estimate of the distribution of values, their extremes, and the presence of any gaps or out-of-the-ordinary numbers. They are useful in providing a basic understanding of the probability distribution.

- **Constructing a Histogram:**

To construct a histogram, the data is grouped into specific class intervals, or "bins" and plotted along the x-axis. These represent the range of the data. Then, the rectangles are constructed with their bases along the intervals for each class. The height of these rectangles is measured along the y-axis representing the frequency for each class interval. It's important to remember that in these representations, every rectangle is next to another because the base spans the spaces between class boundaries.

- **Use Cases:**

When it is necessary to illustrate or compare the distribution of specific numerical data across several ranges of intervals, histograms can be employed. They can aid in visualising the key meanings and patterns associated with a lot of data. They may help a business or organization in decision-making process. Some of the use cases of histograms include-

- Distribution of salaries in an organisation
- Distribution of height in one batch of students of a class, student performance on an exam,
- Customers by company size, or the frequency of a product problem.

Best Practices

- **Analyse various data groups:** The best data groupings can be found by creating a variety of histograms.

- **Break down compartments using colour:** The same chart can display a second set of categories by colouring the bars that represent each category.

Example

The following daily wages and workers create histogram chart using data.

Daily wages	36-38	38-40	40-42	42-44	44-46	46-48	48-50	50-52	52-54	54-56
workers	2	4	10	15	19	19	15	10	4	2

2. Box plots

When displaying data distributions using the five essential summary statistics of minimum, first quartile, median, third quartile, and maximum, box-and-whisker plots, also known as boxplots, are widely employed. It is a visual depiction of data that aids in determining how widely distributed or how much the data values change. These boxplots make it simple to compare the distributions since it makes the centre, spread, and overall range understandable. They are utilised for data analysis wherein the graphical representations are used to determine the following:

1. Shape of Distribution
2. Central Value
3. Variability of Data

Constructing a Boxplot:

The two components of the graphic are described by their names: the box, which shows the median value of data along with the first and third quartiles (25 percentile and 75 percentile), and the whiskers, which shows the remaining data. The 3rd quartile's difference from the first quartile of data is called the interquartile range. The highest and minimum points in the data can also be displayed using the whiskers. The points beyond 1.5 ´ interquartile range can be identified as suspected outliers.

Use Cases:

A boxplot is frequently used to demonstrate whether a distribution is skewed and whether the data set contains any potential outliers, or odd observations. Boxplots are also very useful for comparing or involving big data sets. Examples of box plots include plotting the:

- Gas efficiency of vehicles
- Time spent reading across readers

Best Practices

- **Cover the points within the box:** This aids the viewer in concentrating on the outliers.
- **Box plot comparisons between categorical dimensions:** Box plots are excellent for quickly comparing dataset distributions.

3. Scatter plots

Scatter plot is the most commonly used chart when observing the relationship between two quantitative variables. It works particularly well for quickly identifying possible correlations between different data points. The relationship between multiple variables can be efficiently studied using scatter plots, which show whether one variable is a good predictor of another or whether they normally fluctuate independently. Multiple distinct data points are shown on a single graph in a scatter plot. Following that, the chart can be enhanced with analytics like trend lines or cluster analysis. It is especially useful for quickly identifying potential correlations between data points.

Constructing a Scatter Plot:

Scatter plots are mathematical diagrams or plots that rely on Cartesian coordinates. In this type of graph, the categories being compared are represented by the circles on the graph (shown by the colour of the circles) and the numerical volume of the data (indicated by the circle size). One colour on the graph allows you to represent two values for two variables related to a data set, but two colours can also be used to include a third variable.

Use Cases:

Scatter charts are great in scenarios where you want to display both distribution and the relationship between two variables.

- Display the relationship between time-on-platform (How Much Time Do People Spend on social media) and churn (the number of people who stopped being customers during a set period of time).

- Display the relationship between salary and years spent at company.

Best Practices

- **Analyse clusters to find segments:** Based on your chosen variables, cluster analysis divides up the data points into discrete parts.
- **Employ highlight actions:** You can rapidly identify which points in your scatter plots share characteristics by adding a highlight action, all the while keeping an eye on the rest of the dataset.
- **mark customization:** individual markings Add a simple visual hint to your graph that makes it easy to distinguish between various point groups.

Probability theory

Concept of probability

There are a number of events in day-to-day life about which one is not sure whether it will occur or not. But one is always curious to know what chance is there for happening or event to occur. For instance, one may be interested to estimate whether it will rain today or not, one would like to evaluate his chance of winning for head is a definite number of tosses, what is the chance that there are four aces in one hand in a game of card among four players, etc.

The numerical evaluation of chance factor of an event is known as a probability.

The probability is to numerically express the possibility of this uncertain event.

Define the random experiment

The exponent which can be independently repeated under the identical condition and all its possible outcomes are known but which of the outcomes will appear cannot be predicting with certainty before conducting the experiment is called a random experiment.

An experiment whose outcomes are known before the experiment is done but certain outcome cannot be predicted before the experiment is happens is known as a random experiment.

Define the random variable

Random variable is real valued functions consider each and every value of sample space. A variable whose values can be obtained from the results of a random experiment is called a random variable. A random variable is a function associated with a sample space of a random experiment, and it takes different values with different probabilities. A random variable can either be discrete or continuous.

Define sample space

The set of all possible outcome of a random experiment is called the sample space.

- **Finite Sample Space: -**If the total number of a possible outcomes in the sample space is finite then it is called a finite sample space.
- **Infinite Sample Space: -**If the total number of possible outcomes in the sample space is infinite then it is called an infinite sample space.

Define the Events

A subset of the sample space of random experiment is called Event.

The collection of possible outcomes which are favourable to a happening out of total outcomes is called an Event.

- **Impossible Event: -**An event which is certain not to occur is called an impossible event. It denoted by ϕ or { }.
- **Certain Event: -** A special subset of the sample space of random experiment is called certain event.
- **Complementary Event:** - The complement of any event A is the event [not A], i.e., the event that A does not occur. It is denoted by A'.
- **Interception of Event: -** Suppose A and B are two events of finite sample space where event A and Occur simultaneously is called the intersection event. It is denoted by $A \cap B$.
- **Union of Events: -** Suppose event A and B are any two events of the finite sample space where the event A occurs or the B occurs or both the event A and B occurs is called the union event. It is denoted by $A \cup B$.
- **Mutually Exclusive Event: -** Suppose A and B are any two events of finite sample space where event A and B do not occur together means $A \cap B = \emptyset$ or in other word event B does not occur when event A occur and event A does not

occur when event B occurs then the event A and B are called mutually exclusive event.

- **Exhaustive Event: -** If the group of the favourable outcomes of event of random experiment in the sample space, then the events are called the executive event. Suppose A and B are any two events of sample space where the event A and B are called exhaustive events if the union $A \cup B$ of the two event A and B is a sample space. That is $A \cup B = U$.
- **Mutually Exclusive and Executive Events: -** Suppose A and B are two event of a finite sample space these two events A and B are called the mutually exclusive and exhaustive events. That is $A \cap B = \emptyset$ and $A \cup B = U$.
- **Elementary Event: -** The event formed by all the subset of single element of the sample space of a random experiment are called elementary event. The elementary events are mutually exclusive and executive event.

Property of probability of an element A

a) Probability is a pure number. i.e., it has no unit.
b) It is a relative measure.
c) $0 \leq P(A) \leq 1$, i.e., probability can never be negative and cannot exceed unit(one).
d) If A1, A2, …., Ak are k mutually exclusive and exhaustive events in the sample than $\sum_{i=1}^{k} P(A_i) = 1 \rightarrow P(U) = 1$.
e) P(ϕ) = 0, i.e., Probability of an impossible event is 0.

Mathematical definition of probability:

Suppose there are the total an outcome in finite sample space of random experiment which are mutually exclusive in exhaustive and equiprobable. if m outcome among there is favourable for an event A then the probability of the event A is $\frac{\mathbf{m}}{\mathbf{n}}$**.** The probability of event A denoted by P(A).

P(A) = probability of event A

$$= \frac{\text{Favourable outcomes of event A}}{\text{Toatl number of mutually exclusive, exhaustive and eui}-\text{probable outcomes of sample specae}}$$

$$= \frac{m}{n}$$

Other formula

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

$$P(A' \cup B') = P(A \cap B)' = 1 - P(A \cup B)$$

$$P(A' \cap B') = P(A \cup B)' = 1 - P(A \cup B)$$
$$P(A - B) = P(A \cap B') = P(A) - P(A \cap B)$$
$$P(B - A) = P(A' \cap B) = P(B) - P(A \cap B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Example

1. **If 3 coins are tossed at a time find the probability of following event.**
   - **A. Getting at least one head**
   - **B. Getting at most one tail**
   - **C. Getting more than one tail**
   - **D. Getting head and tail alternately**
   - **E. Getting more number of tail than head**

Ans: - The sample space of random experiment of tossing coin 3 time

Total outcomes (n) = $(2)^3$

= 8

U = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

A. Getting at least one head

Favourable outcomes (m) =7

A = {HHH, HHT, HTH, HTT, THH, THT, TTH}

$$P(A) = \frac{m}{n}$$
$$= \frac{7}{8}$$

B. Getting at most one tail

Favourable outcomes (m) =4

B = {HHH, HHT, HTH, THH}

$$P(B) = \frac{m}{n}$$
$$= \frac{4}{8}$$
$$= \frac{1}{2}$$

C. Getting more than one tail

Favourable outcomes (m) =4

C = {HTT, THT, TTH, TTT}

$$P(C) = \frac{m}{n}$$
$$= \frac{4}{8}$$

$$= \frac{1}{2}$$

D. Getting head and tail alternately
   Favourable outcomes (m) =2
   D = {THT, HTH}

$$P(D) = \frac{m}{n}$$

$$= \frac{1}{2}$$

E. Getting more number of tail than head
   Favourable outcomes (m) =4
   E = {HTT, THT, TTH, TTT}

$$P(E) = \frac{m}{n}$$
$$= \frac{4}{8}$$
$$= \frac{1}{2}$$

2. **Two balance dice are thrown simultaneously. Find the probability of the following events:**
   - **A. The sum of number on the dice is 8**
   - **B. The sum of number on the dice is more than 9**
   - **C. The sum of numbers on the dice multiple of 3**
   - **D. The product of number on the dice is 12.**

Ans: - The sample space of random experiment of two balance dice are thrown simultaneously

Total outcomes (n) = $(6)^3$
= 36

U = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}

A. The sum of number on the dice is 8
   Favourable outcomes (m) = 5

A = {(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)}

$$P(A) = \frac{m}{n}$$
$$= \frac{5}{36}$$

B. The sum of number on the dice is more than 9

Favourable outcomes (m) = 6

B = {(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)}

$$P(B) = \frac{m}{n}$$
$$= \frac{6}{36}$$
$$= \frac{1}{6}$$

C. The sum of number on the dice is multiple of 3

Favourable outcomes (m) = 12

C = {(1, 2), (2, 1), (1, 5), (2, 4), (3, 3), (4, 2), (5, 1), (3, 6), (4, 5), (5, 4), (6, 3), (6, 6)}

$$P(C) = \frac{m}{n}$$
$$= \frac{12}{36}$$
$$= \frac{1}{3}$$

D. The product of the number o the dice is 12

Favourable outcomes (m) = 4

D = {(3, 4), (4, 3), (6, 2), (2, 6)}

$$P(D) = \frac{m}{n}$$
$$= \frac{4}{36}$$
$$= \frac{1}{9}$$

3. **The marks of the ten students in the subject of statistics are given below 50, 64, 65, 35, 55, 52, 63, and 30 if one is selected at randomly then what is the probability that his marks would be greater than the average marks.**

Ans: Total no. of Students (n) = 10

Probability that one is selected at randomly and his marks would be greater than the average marks.

$$\bar{x} = \frac{\sum x}{n} = \frac{495}{10} = 49.5$$

Favourable outcomes (m) = 6

A = {50, 60, 65, 55, 52, 63}

$$P(A) = \frac{m}{n}$$
$$= \frac{6}{10}$$
$$= 0.6$$

4. **Find the probability of getting vowels in the first, third and sixth place when all the letters of the word ORANGE are arranged in the all-possible ways.**

Ans: - The total number of ways of arranging these these six letters.

Total outcomes (n) = 6!

= 720

A = the probability of getting vowels in the first, third and sixth place when all the letters of the word ORANGE.

O			A			E

Favourable outcomes (m) = 3! $\times$ 3!

= 36

$$P(A) = \frac{m}{n}$$
$$= \frac{36}{720}$$
$$= \frac{1}{20}$$

5. **Find the probability of 5 Mondays in the months of February of a leap year**

Ans: - In a leap year total no. of days in the months of February = 29

- 7 days $\times$ 4 weeks in every month = 28

Extra day 1

In a week everyday come only once. Hence in 4 weeks every day comes 4 times the sample space for extra day is expressed as follows

U = {Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday}

Total outcomes (n) = 7

A = The probability of 5 Mondays in the months of February of a leap year

Favourable outcomes (m) = 1

$$P(A) = \frac{m}{n}$$
$$= \frac{1}{7}$$

6. **Find the probability of having 53 Thursday in day a leap year**

Ans: - Total no. of days in a leap year = 366
- 7 days × 52 weeks in a year = 364

Extra day = 2

The additional 2 days can be as follows which given the sample space for this experiment.

U = {Sunday-Monday, Monday-Tuesday, Tuesday-Wednesday, Wednesday-Thursday, Thursday-Friday, Friday-Saturday, Saturday-Sunday}

Total outcomes (n) = 7

A = The probability of 5 Mondays in the months of February of a leap year

Favourable outcomes (m) = 2
{Tuesday-Wednesday, Wednesday-Thursday}

$$P(A) = \frac{m}{n}$$
$$= \frac{2}{7}$$

7. **One card is randomly selected from the packet of 52 cards find a probability that the selected card is**
   i) Diamond or king card
   ii) neither a diamond nor a king card.

Ans: one card is randomly selected from the packet of 52 cards
Total outcomes (n) = 52

i) Selected card is diamond or king card

$$P(A \cup B) = \text{ Selected card is diamond or king card}$$

Event A: Selected card is diamond

Favourable outcomes (m) = 13

$$P(A) = \frac{m}{n}$$
$$= \frac{13}{52}$$

Event B: Selected card is king

Favourable outcomes (m) = 4

$$P(B) = \frac{m}{n}$$
$$= \frac{4}{52}$$

Event A∩B: Selected card is diamond and king

Favourable outcomes (m) = 1

$$P(A \cap B) = \frac{m}{n}$$
$$= \frac{1}{52}$$

$$P(A \cup B) = \text{ Selected card is diamond or king card}$$
$$= P(A) + P(B) - P(A \cap B)$$
$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52}$$
$$= \frac{4}{13}$$

ii) Selected card neither a diamond nor a king card

A’ = Event that the selected card is not diamond

B’ = Event that the selected card is not king

A’∩B = Event that the selected card neither a diamond nor a king card

$$P(A' \cap B') = P(A \cup B)'$$
$$= 1 - P(A \cup B)$$
$$= 1 - \frac{4}{13}$$
$$= \frac{9}{13}$$

8. **The probability that a Pearson from a group reads English newspaper is 0.55, the probability that he reads Hindi newspaper is 0.69 and the probability that he reads both the newspaper English and Hindi is 0.27. find the probability that a person**

selected at random from this group

i) read at least one of the English and Hindi newspaper.

ii) does not read any of newspaper English and Hindi.

Ans: Event A = Probability that a Pearson from a group reads English newspaper

Event B = Probability that a Pearson from a group reads Hindi newspaper

Event $A \cap B$ = The probability that he reads both the newspaper English and Hindi

$P(A) = 0.55$, $P(B) = 0.69$, $P(A \cap B) = 0.27$

i) Read at least one of the English and Hindi newspaper

$P(A \cup B)$ = The probability that a person selected at random from this group read at least one of the English and Hindi newspaper

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.55 + 0.69 - 0.27 \\ &= 0.97 \end{aligned}$$

ii) Does not read any of newspaper English and Hindi

Event A' = Probability that a Pearson from a group not reads English newspaper

Event B' = Probability that a Pearson from a group not reads Hindi newspaper

Event $A' \cap B'$ = The probability that he not reads both the newspaper English and Hindi

$$\begin{aligned} P(A' \cap B') &= P(A \cup B)' \\ &= 1 - P(A \cup B) \\ &= 1 - 0.97 \\ &= 0.03 \end{aligned}$$

9. **For the two event A and B is the sample space of a random experiment P(A') = 0.3, P(B)= 0.6 and $P(A \cup B) = 0.83$. Find $P(A \cap B')$ and $P(A' \cap B)$.**

Ans.: Here, $P(A') = 0.3$ $P(A) = 1-P(A') = 1- 0.3 = 0.7$

$P(B) = 0.6$ and $P(A \cup B) = 0.83$

First, we will find $P(A \cap B')$.

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

$0.83 = 0.7 + 0.6 - P(A \cap B)$

$P(A \cap B) = 0.7 + 0.6 - 0.83$

$P(A \cap B) = 0.47$

Now,

$P(A \cap B') = P(A) - P(A \cap B)$

= 0.7 – 0.47

= 0.23

Required probability = 0.23

$P(A' \cap B) = P(B) - P(A \cap B)$

= 0.6 – 0.47

= 0.13

Required probability = 0.13

10. **For two events A and B in the sample space of a random experiment P(A) = 2P(B) = 4P(A ∩ B) = 0.6. Find the probability of the following events: (1) $A' \cap B'$ (2) $A' \cup B'$ (3) A – B (4) B – A**

It is given that P(A) = 2P(B) = 4P(A ∩ B) = 0.6. Hence,

P(A) = 0.6 2P(B) = 0.6 4P(A ∩ B) = 0.6

∴ P(B) = 0.3 ∴ P(A ∩ B) = 0.15

(1) Probability of events A' ∩ B' = P(A' ∩ B')

= P(A ∪ B)'

= 1 – P(A ∪ B)

= 1 - [P(A) + P(B) - P(A ∩ B)]

= 1 - [0.6 +0.3 - 0.15]

= 1- 0.75

= 0.25

Required probability = 0.25

(2) Probability of events A' ∪ B' = P(A' ∪ B')

= P(A∩ B)'

= 1 - P(A ∩ B)

= 1 - 0.15

= 0.85

Required probability = 0.85

(3) Probability of events A – B = P(A-B)

= P(A) – P(A ∩ B)

= 0.6 – 0.15

= 0.45

Required probability = 0.45

(4) Probability of B – A = P(B-A)

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

$= 0.3 - 0.15$

$= 0.15$

Required probability = 0.15

11.**Two events A and B in the sample space of a random experiment mutually exclusive. If 3P(A) = 4P(B) = 1 then find P($A \cup B$).**

Ans: Since 3P(A) = 4P(B) = 1

$3P(A) = 1 \qquad 4P(B) = 1$

$\therefore P(A) = \frac{1}{3} \qquad \therefore P(B) = \frac{1}{4}$

As the events A and B are mutually exclusive ($A \cap B = \emptyset$),

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

$$= \frac{1}{3} + \frac{1}{4}$$

$$= \frac{7}{12}$$

Required probability = $\frac{7}{12}$

12. **For three mutually exclusive and exhaustive events A, B and C in the sample space of a random experiment 2P(A) = 3P(B) = 4P(C). Find P($A \cup B$) and P($B \cup C$).**

Ans: Taking 2P(A) = 3P(B) = 4P(C) = x,

$2P(A) = x \qquad 3P(B) = x \qquad 4P(C) = x$

$\therefore P(A) = \frac{x}{2} \qquad \therefore P(B) = \frac{x}{3} \qquad \therefore P(C) = \frac{x}{4}$

Since A , B and C are mutually exclusive and exhaustive events,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$\therefore \frac{x}{2} + \frac{x}{3} + \frac{x}{4} = 1$$

$$\therefore \frac{6x+4x+3x}{12} = 1$$

$$\therefore 13x = 12$$

$$\therefore x = \frac{12}{13}$$

Thus,

P(A) = $\frac{x}{2} = \frac{\frac{12}{13}}{2} = \frac{6}{13}$

P(B) = $\frac{x}{3} = \frac{\frac{12}{13}}{3} = \frac{4}{13}$

P(C) = $\frac{x}{4} = \frac{\frac{12}{13}}{4} = \frac{3}{13}$

Now, the probability of required events,

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

$$= \frac{6}{13} + \frac{4}{13}$$

$$= \frac{10}{13}$$

Required probability = $\frac{10}{13}$

$$P(B \cup C) = P(B) + P(C)$$

$$= \frac{4}{13} + \frac{3}{13}$$

$$= \frac{7}{13}$$

Required probability = $\frac{7}{13}$

QUESTIONS

1. If 3 coins are tossed at a time find the probability of following event.
   A. Getting at least one tails
   B. Not getting a single head
   C. Getting at most two head
   D. Getting more than one head
   E. Getting more number of head than tail
   F. Getting at most three tails

[Ans: 1/8, 1/8, 7/8, 1/2, 1/2, 1]

2. If two balanced coins are tossed, then find the probability of
   A. Getting at most one tail
   B. Getting one head and one tail

[Ans: 3/4, 1/2]

3. Two balanced dice marked with number 1 to 6 are thrown simultaneously. Find the probability that
   A. The sum of number on the dice is 5
   B. The sum of number on the dice is more than 10
   C. The sum of numbers on the dice multiple of 4
   D. Both the dice show same numbers
   E. Sum of numbers on the dice is 1
   F. Sum of the number on the dice is 12 or more
   G. Sum of the number on the dice is 12 or less

[Ans: 1/8, 1/12, 1/4, 1/6, 0/36, 0/36, 1]

4. One family randomly selected from the families having two children find the probability that
   A. One child is boy and one child is girl
   B. At least one child is boy among the two children of the selected family

[Ans: 1/2, 3/4]

5. The sample space for the random experiment of selecting numbers is U {1, 2, 3,……,150} and all the outcomes in the sample space are equiprobable. Find the probability that the number selected is:
   A. a multiple of 2
   B. not a multiple of 2
   C. a multiple of 5
   D. not a multiple of 5
   E. a multiple of both 2 and 5

[Ans: 1/2, 1/2, 1/5, 4/5, 1/10]

6. Find the probability of getting U in the first place and E in the last place when all of the word **UNIQUE** are arranged in the possible ways.

[Ans: 1/30]

7. The runs for by 10 players are follows 25, 32, 40, 20, 36, 45, 50, 30, 25, 37 one player is selected at randomly the find the probability that
   A. Whose run more than the average runs
   B. Whose runs are less than the average runs

[Ans: 0.5, 0.5]

8. The run of 11 cricketers in a day match argue 90, 50, 30, 100, 5, 40, 0, 25, 70, 10, 20 if one cricketer is selected at randomly the what is the probability that he's run could be less than the average runs.

[Ans: 6/11]

9. Find the Probability having 53 Fridays in a year which is not a leap year. **[Ans: 1/7]**
10. Find the probability of having 5 Tuesdays in the month of august of any year. **[Ans: 3/7]**

11. Two cards are drowned from a pack of 52 cards what is the probability that
   A. both cards are king
   B. both cards are king or both queens
   C. both cards are king and queens.

[Ans: 1/221, 2/221, 8/663]

12. Two cards are drowned from a pack of 52 cards what is the probability that
   A. Both cards are same suit
   B. Both cards are same colour

[Ans: 4/17, 25/51]

13. One card is randomly selected from the packet of 52 cards find a probability that the selected card is
   i) Heart or jack card
   ii) neither a Heart nor a jack card.

[Ans: 4/13, 9/13]

13. Two aircrafts dropped bomb to destroy a bridge the probability that a bomb dropped from the first aircraft hits the target is 0.9 and the probability that a bomb from the second aircraft hits the target is 0.7. the probability of one one bomb drops from both the aircraft hitting the target is 0.63 the bridge is destroyed even if one bomb drops on it. Find the probability that the bridge is destroyed.

[Ans: 0.97]

14. For three mutually exclusive and exhaustive events A, B and C in the sample space of a random experiment 4P(A) = 5P(B) = 3P(C). Find $P(A \cup C)$ and $P(B \cup C)$.

[Ans: 35/47, 32/47]

15. For two events A and B in the sample space of a random experiment P(A') = 2P(B') = 3P(A ∩ B) = 0.6. Find the probability of the following events:
   i. A – B
   ii. B – A

[Ans: 0.2, 0.5]

16. **F**or two events A and B in the sample space of a random experiment P(A) = 0.6 P(B) = 0.5 and P(A ∩ B) = 0.15. Find the probability of the following events: (1)
   i. P(A')
   ii. P(B – A)
   iii. $P(A \cap B')$
   iv. $P(A' \cap B')$
   v. $P(A' \cup B')$

[Ans: 0.4, 0.35, 0.45, 0.05, 0.85]

❖ Probability distribution

A probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.

Type of Probability Distribution

There are two types of Probability Distribution

1. Discrete probability distribution
2. Continuous probability distribution

1. Discrete probability distribution

If a random variable can take finite values or countable infinite values it is known as a discrete variable. The number of heads obtained in tossing of two coins is a discrete variable. Similarly, the number of children in a family, number of accidents are examples of discrete variables. If the number of defective screws manufactured by a company is denoted by x then the values of x can be 0, 1, 2, 3, 4, 5, 6...... etc. The values of discrete variable in this case are countable infinite values.

Here we study Binomial Distribution which discrete probability distribution.

Binomial Distribution

In random experiment which done in single trial and outcome are of the trial either success or failure then such trial is known as Bernoulli trials.

Considering n independent Bernoulli trials each trial has two outcome is considered as success with the the probability p and failure with probability q that is p – q = 1, among this trial here the probability of x success follows binomial distribution. Since x denoted the number of successes in n independent trials

Therefore x is random variable having value X = 0, 1, 2, …. , n let probability of x success and n - x failure having the order of probability of success and failure respectively then it is denoted with $p^x q^{n-x}$

The possible ways of x success among these trails are C_x^n

Therefore, the x^{th} in n independent trial

$$P(x) = C_x^n . p^x . q^{n-x} \qquad ; x = 0, 1, 2, \ldots. \ , n$$

- **Examples of Bernoulli's Trails are**
  - Tossing a coin (head or tail)
  - Throwing a die (even or odd number)
  - Student's performance an exam (Pass or fail)

Conditions for binomial distribution

The probability of success is denoted by ‘p’ and that of failure by ‘q’, such that p + q = 1. The distribution can be obtained under the following experimental conditions:

- The number of trials ‘n’ is finite.
- The trials are independent of each other.
- Success probability ‘p’ is constant for each trial.
- Each trial has only one of the two possible results either success or failure.

The best examples of binomial distribution are, tossing of coins or throwing of dice or drawing cards from a pack of cards.

Characteristics of Binomial Distribution

1) This is a distribution of a discrete variable.
2) n, p are the parameters of Binomial distribution.
3) The mean of Binomial distribution is np which shows the average number of successes.
4) Variance of binomial distribution is npq
5) Standard deviation of binomial distribution is $\sqrt{npq}$
6) When p and q are equal i.e., p = q = 1/2 Binomial distribution is a symmetrical distribution.
7) When p< 1/2, its skewness is positive and when p> 1/2 its skewness is negative.
8) The variance of Binomial distribution is always less than its mean.
9) When number of trials n is very large, and p and q are not very small, Binomial distribution tends to Normal distribution.
10) When number of trials n is very large and p or q is very small, Binomial distribution tends to Poisson distribution.

➢ **Example**

1. **The probability of winning a one-day cricket match of India against Australia is 1/3. India and Australia are going to play three one day cricket matches. Find the probability that**
   A. **India will lose all the three one day matches**
   B. **India will win at least one one-day matches.**

Ans: Probability of winning a one-day cricket match of India against Australia $= p = \frac{1}{3}$

India and Australia are going to play three one day cricket matches = n = 3

here $q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

A. India will lose all the three one day matches

x = 0

$$P(x) = C_x^n . p^x . q^{n-x}$$
$$P(0) = C_0^3 . \left(\frac{1}{3}\right)^0 . \left(\frac{2}{3}\right)^{3-0}$$
$$= 1 \times 1 \times \frac{8}{27}$$
$$= \frac{8}{27}$$

B. India will win at least one one-day matches

x = 1, 2, 3

P(x ≥ 3) = P(1) + P(2) + P(3)

$$= C_1^3 . \left(\frac{1}{3}\right)^1 . \left(\frac{2}{3}\right)^{3-1} + C_2^3 . \left(\frac{1}{3}\right)^2 . \left(\frac{2}{3}\right)^{3-2} + C_3^3 . \left(\frac{1}{3}\right)^3 . \left(\frac{2}{3}\right)^{3-3}$$
$$= \left(3 \times \frac{1}{3} \times \frac{4}{9}\right) + \left(3 \times \frac{1}{9} \times \frac{2}{9}\right) + \left(1 \times \frac{1}{27} \times 1\right)$$
$$= \frac{12}{27} + \frac{6}{27} + \frac{1}{27}$$
$$= \frac{19}{27}$$

OR short cut method

P(x ≥ 3) = 1 + P(0)

$$= 1 - \left(C_0^3 . \left(\frac{1}{3}\right)^0 . \left(\frac{2}{3}\right)^{3-0}\right)$$
$$= 1 - \left(1 \times 1 \times \frac{8}{27}\right)$$

$$= 1 - \left(\frac{8}{27}\right)$$
$$= \frac{19}{27}$$

2. **A chance of winning the match of India against Pakistan is 5:2 if a series of three matches is to be played than what is the probability of the India to be win at most one match.**

Ans: A chance of winning the match of India against Pakistan is 5:2

$p = \frac{5}{7}, \qquad q = 1 - p = 1 - \frac{5}{7} = \frac{2}{7}, \qquad n = 3$

Probability of the India to be win at most one match

x = 0, 1

P(x ≤ 1) = P(0) + P(1)

$$= C_0^3 . \left(\frac{5}{7}\right)^0 . \left(\frac{2}{7}\right)^{3-0} + C_1^3 . \left(\frac{5}{7}\right)^1 . \left(\frac{2}{7}\right)^{3-1}$$
$$= \left(1 \times 1 \times \frac{8}{343}\right) + \left(3 \times \frac{5}{7} \times \frac{4}{49}\right)$$
$$= \frac{8}{343} + \frac{60}{343}$$
$$= \frac{68}{343}$$

3. **30% of attacking aircraft are expected to be shoot down before reaching the targets what is the probability that one out of five attacking aircraft will reach the target.**

Ans: Attacking aircraft are expected to be shoot down before reaching the targets = 30%

q = 0.3, p = 1 – q = 1 – 0.3 = 0.7, n = 5

Probability that one out of five attacking aircraft will reach the target

x = 1

$$P(x) = C_x^n . p^x . q^{n-x}$$
$$P(1) = C_1^5 . (0.7)^1 . (0.3)^{5-1}$$
$$= 5 \times (0.7) \times (0.0081)$$
$$= 0.02835$$

4. **The probability that a bomb dropped from the plane will hit target is 2/5. Two bombs are enough to destroy a bridge. If 4 bombs are dropped on the bridge, find the probability that**

A. **The bridge will be destroyed**
B. **The bridge will be partially destroyed**
C. **The bridge will be saved.**

Ans: The probability that a bomb dropped from the plane will hit target = 2/5

$$p = \frac{2}{5}, \qquad q = 1 - p = 1 - \frac{2}{5} = \frac{3}{5}, \qquad n = 4$$

A. The bridge will be destroyed
   x = 2, 3, 4

P(2 ≤ x ≤ 4) = P(2) + P(3) + P(4)

$$= C_2^4 . \left(\frac{2}{5}\right)^2 . \left(\frac{3}{5}\right)^{4-2} + C_3^4 . \left(\frac{2}{5}\right)^3 . \left(\frac{3}{5}\right)^{4-3} + C_4^4 . \left(\frac{2}{5}\right)^4 . \left(\frac{3}{5}\right)^{4-4}$$

$$= \left(6 \times \frac{4}{25} \times \frac{9}{25}\right) + \left(4 \times \frac{8}{125} \times \frac{3}{5}\right) + \left(1 \times \frac{16}{625} \times 1\right)$$

$$= \frac{216}{625} + \frac{96}{625} + \frac{16}{625}$$

$$= \frac{328}{625}$$

B. The bridge will be partially destroyed
   x = 1

$$P(x) = C_x^n . p^x . q^{n-x}$$

$$P(1) = C_1^4 . \left(\frac{2}{5}\right)^1 . \left(\frac{3}{5}\right)^{4-1}$$

$$= 4 \times \frac{2}{5} \times \frac{27}{125}$$

$$= \frac{216}{625}$$

C. The bridge will be saved
   x = 0

$$P(x) = C_x^n . p^x . q^{n-x}$$

$$P(0) = C_0^4 . \left(\frac{2}{5}\right)^0 . \left(\frac{3}{5}\right)^{4-0}$$

$$= 1 \times 1 \times \frac{81}{625}$$

$$= \frac{81}{625}$$

5. **There are 50 fishes in a pool, out of which 10 are golden, 5 fishes are taken at random from the pool. What is the probability that 2 fishes are golden out of 5?**

Ans: there are 50 fished in a pool N = 50
out of which 10 are golden X = 10
5 fishes are taken random from the pool n = 5
Probability that 2 fishes are golden x = 2

$p = \frac{2}{5}, \qquad q = 1 - p = 1 - \frac{2}{5} = \frac{3}{5}, \qquad n = 5$

$$P(x) = C_x^n . p^x . q^{n-x}$$
$$P(2) = C_2^5 . \left(\frac{2}{5}\right)^2 . \left(\frac{3}{5}\right)^{5-2}$$
$$= 10 \times \frac{4}{25} \times \frac{9}{25}$$
$$= \frac{360}{625}$$
$$= \frac{72}{25}$$

6. **In a competitive test there are 6 objective questions and three alternatives are given for each question. Only one of them is correct. A candidate does not know the correct answers of any of the questions and hence in each question he ticks any one of the alternatives randomly. Find the probabilities of getting**
   A. **All correct answers**
   B. **At least 4 correct answers**

Ans: There three alternatives are given for each question only one of them is correct = 1/3

$p = \frac{1}{3}, \qquad q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}, \qquad n = 6$

A. All correct answers
   x = 6
$$P(x) = C_x^n . p^x . q^{n-x}$$
$$P(2) = C_6^6 . \left(\frac{1}{3}\right)^6 . \left(\frac{2}{3}\right)^{6-6}$$
$$= 1 \times \frac{1}{729} \times 1$$
$$= \frac{1}{729}$$

B. At least 4 correct answers

   x = 4, 5, 6

   P(x ≥ 4) = P(4) + P(5) + P(6)

$$= C_4^6 . \left(\frac{1}{3}\right)^4 . \left(\frac{2}{3}\right)^{6-4} + C_5^6 . \left(\frac{1}{3}\right)^5 . \left(\frac{2}{3}\right)^{6-5} + C_6^6 . \left(\frac{1}{3}\right)^6 . \left(\frac{2}{3}\right)^{6-6}$$

$$= \left(15 \times \frac{1}{81} \times \frac{4}{9}\right) + \left(6 \times \frac{1}{243} \times \frac{2}{3}\right) + \left(1 \times \frac{1}{729} \times 1\right)$$

$$= \frac{60}{729} + \frac{12}{729} + \frac{1}{729}$$

$$= \frac{73}{729}$$

7. An accountant is to audit 24 accounts of a firm. 16 of these are high valued customers. If the accountant selects six accounts at random, what is probability of selection of:

A. At least three high valued customers?

B. At the most three high valued customers?

Ans: An accountant is to audit 24 accounts of a firm N = 24

16 of these are high valued customers X = 16

If the accountant selects six accounts at random n = 6

$$p = \frac{16}{24} = \frac{2}{3}, \qquad q = 1 - p = 1 - \frac{2}{3} = \frac{1}{3}, \qquad n = 6$$

A. At least three high valued customers

x = 3, 4, 5, 6 **OR** x = 1 – [0, 1, 2]

$$P(x \geq 3) = 1 - [P(0) + P(1) + P(2)]$$

$$= 1 - \left[C_0^6 . \left(\frac{2}{3}\right)^0 . \left(\frac{1}{3}\right)^{6-0} + C_1^6 . \left(\frac{2}{3}\right)^1 . \left(\frac{1}{3}\right)^{6-1} + C_2^6 . \left(\frac{2}{3}\right)^2 . \left(\frac{1}{3}\right)^{6-2}\right]$$

$$= 1 - \left[\left(1 \times 1 \times \frac{1}{729}\right) + \left(6 \times \frac{2}{3} \times \frac{1}{243}\right) + \left(15 \times \frac{4}{9} \times \frac{1}{81}\right)\right]$$

$$= 1 - \left[\frac{1}{729} + \frac{12}{729} + \frac{60}{729}\right]$$

$$= 1 - \left[\frac{73}{729}\right]$$

$$= \frac{656}{729}$$

B. At the most three high valued customers

x = 0, 1, 2, 3

$$P(x \geq 3) = P(0) + P(1) + P(2) + P(3)$$

$$= C_0^6.\left(\frac{2}{3}\right)^0.\left(\frac{1}{3}\right)^{6-0} + C_1^6.\left(\frac{2}{3}\right)^1.\left(\frac{1}{3}\right)^{6-1} + C_2^6.\left(\frac{2}{3}\right)^2.\left(\frac{1}{3}\right)^{6-2} + C_3^6.\left(\frac{2}{3}\right)^3.\left(\frac{1}{3}\right)^{6-3}$$

$$= \left(1 \times 1 \times \frac{1}{729}\right) + \left(6 \times \frac{2}{3} \times \frac{1}{243}\right) + \left(15 \times \frac{4}{9} \times \frac{1}{81}\right) + \left(20 \times \frac{8}{27} \times \frac{1}{27}\right)$$

$$= \frac{1}{729} + \frac{12}{729} + \frac{60}{729} + \frac{160}{729}$$

$$= \frac{233}{729}$$

8. Five coins are tossed simultaneously for 3200 times. Find the frequencies of different number of heads and tails and tabulate the results.

Ans: Here probability of getting head $p = \frac{1}{2}$

$\therefore q = 1 - p = \frac{1}{2}$ N = 3200, n = 5

Now for Binomial distribution

$$P(x) = C_x^n.p^x.q^{n-x}$$
$$= C_x^5.\left(\frac{1}{2}\right)^x.\left(\frac{1}{2}\right)^{n-x}$$

Putting x = 0, 1, 2, 3, 4, 5 we find the probability of different number of heads.

The results can be represented in a table as follows:

Number of heads x	Probability $P(x)$	Expected frequency $= N \times P(x)$
0	$C_0^5.\left(\frac{1}{2}\right)^0.\left(\frac{1}{2}\right)^5 = \frac{1}{32}$	100
1	$C_1^5.\left(\frac{1}{2}\right)^1.\left(\frac{1}{2}\right)^4 = \frac{5}{32}$	500
2	$C_2^5.\left(\frac{1}{2}\right)^2.\left(\frac{1}{2}\right)^3 = \frac{10}{32}$	1000
3	$C_3^5.\left(\frac{1}{2}\right)^3.\left(\frac{1}{2}\right)^2 = \frac{10}{32}$	1000
4	$C_4^5.\left(\frac{1}{2}\right)^4.\left(\frac{1}{2}\right)^1 = \frac{5}{32}$	500

5	$C_5^5.\left(\frac{1}{2}\right)^5.\left(\frac{1}{2}\right)^0 = \frac{1}{32}$	100
Total	1	3200

9. A person tosses three coins simultaneously. He gets Rs. 5 if three heads appear, Rs. 3 if two heads appear, Rs. 2 if one head appears. He has to pay a penalty of Rs. 10 if no head appears. Find his expected amount.

Ans: suppose the penalty is Rs a if no head appears.

Head	Amount x	Probability $P(x)$	$x.P(x)$
0	$-a$	$C_0^3.p^0.q^3 = \frac{1}{8}$	$-\frac{a}{8}$
1	2	$C_1^3.p^1.q^2 = \frac{3}{8}$	$\frac{6}{8}$
2	4	$C_2^3.p^2.q^1 = \frac{3}{8}$	$\frac{12}{8}$
3	8	$C_3^3.p^3.q^0 = \frac{1}{8}$	$\frac{8}{8}$

For a fair game $Ex = 0$

$$\therefore Ex = 0$$

$$\therefore -\frac{a}{8} + \frac{6}{8} + \frac{12}{8} + \frac{8}{8} = 0$$

$$\therefore -a + 26 = 0$$

$$\therefore a = 26$$

$$\therefore He\ will\ loss\ Rs. 26\ if\ no\ head\ appears.$$

10. The mean and variance of binomial distribution are 15 and 6 respectively. Find the value of n and p.

Ans: $Mean = np = 15$

$$Variance = npq$$

$$= 6$$

$$q = \frac{variance}{mean} = \frac{npq}{np} = \frac{6}{15} = \frac{2}{5} = 0.6$$

$$q = \frac{6}{15} \qquad \therefore q = 1 - p = 1 - 0.6 = 0.4$$

$$np = 15$$

$$n(0.6) = 15$$

$$n = \frac{15}{0.6}$$

$$n = 25$$

11.For a binomial variate n = 10 and $P(x = 5) = 2. P(x = 4)$ find the values of p.

Ans: for binomial variate

$$P(x) = C_x^n . p^x . q^{n-x}$$

$$P(x = 5) = 2 \times P(x = 4)$$

$$\therefore \quad C_5^{10} . p^5 . q^5 = 2 \times (C_4^{10} . p^4 . q^6)$$

$$\therefore \quad 252. p^5 . q^5 = 2 \times (210. p^4 . q^6)$$

$$\therefore \quad \frac{p^5 q^5}{p^4 q^6} = \frac{420}{252}$$

$$\therefore \quad \frac{p}{q} = \frac{5}{3}$$

$$\therefore \quad p = \frac{5}{3} \mathrm{q}$$

$$Now \qquad p + q = 1$$

$$\therefore \quad p + \frac{5}{3} p = 1$$

$$\therefore \quad \frac{8}{3} p = 1$$

$$\therefore \quad p = \frac{5}{8}$$

➢ Question

1. A chance of winning the match of India against Pakistan is 2:1 if a series of three matches is to be played than what is the probability of the India to be win
   A. At least one match
   B. two matches

[Ans: 26/27, 12/27]

2. A chance of winning the match of India against Pakistan is 5:2 if a series of three matches is to be played than what is the probability of the India to be win
    A. At least one match
    B. Three matches

[Ans: 0.9768, 0.3645]

3. A chance of winning the match of India against Pakistan is 4:1 if a series of three matches is to be played than what is the probability of the India to be win
    A. At least one match
    B. two matches

[Ans: 0.992, 0.384]

4. If Wining Probability of India against South Africa is 3/5 and both teams will play 5 matches. So, find the probability that India wins 2 matches at least.

[Ans: 0.9129]

5. If the probability that a man aged 60 will live to be 70 is 0.65, what is the probability that out of 10 men aged 60, at least 7 will live up to 70?

[Ans: 0.5138]

6. A brokerage survey reports that 30 per cent of individual investors have used a discount broker, i.e., one which does not charge the full commission. In a random sample of 9 individuals. What is the probability that
    A. Exactly two of the sampled individuals have used a discount broker
    B. Not more than three have used a discount broker
    C. At least three of them have used a discount broker.

[Ans: 0.2656, 0.7297, 0.5372]

7. Assuming that half the population is rice eater and assuming that 100 investigators can take a sample of 10 individuals to see whether they are rice eaters. How many investigators would you expect to report that:
    A. Three people or less were rice eater.
    B. At least 7 people were rice eater.

[Ans: 0.1719, 0.1719]

8. There are 200 farms in a taluka among the bore wells made in these two 100 farms of the taluka, salted water is found in 20 farms. Find the probability of the event of not getting salty water in three out of five randomly selected farms from the taluka.

[Ans: 0.0729]

9. Seven coins are tossed simultaneously for 512 times. Find the frequencies of different number of heads and tails and tabulate the results.

[Ans:]

No of head x	0	1	2	3	4	5	6	7

Probability $P(x)$	$\frac{1}{128}$	$\frac{7}{128}$	$\frac{21}{128}$	$\frac{35}{128}$	$\frac{35}{128}$	$\frac{21}{128}$	$\frac{7}{128}$	$\frac{1}{128}$
Expected frequency $= N \times P(x)$	4	28	84	140	140	84	28	4

10. The mean and variance of binomial distribution are 18 and 4.5 respectively. Find the value of n and p.

[Ans: n = 24, p = 3/4]

11. The mean and variance of binomial distribution are 3.9 and 2.73 respectively. Find the value of n and p.

[Ans: n = 13, p = 0.3]

12. The parameter of a binomial distribution is 10 and 2/5. Calculated mean and variance.

[Ans: np = 4, npq = 12/5]

13. If the Variance and probability of failure are 3/4 and 3/4 respectively in binomial distribution then find mean.

[Ans: np = 1]

14. Mean and S.D. of binomial variable are 20 and 2 respectively then find its parameters

[Ans: n = 25, p = 4/5, q = 1/5]

15. For a binomial variate n = 8 and $2.P(x = 4) = P(x = 3)$ find the values probability of failure also find mean and variance.

[Ans: q = 1/3, np = 16/3, npq = 16/9]

16. For a binomial variate n = 6 and $2.P(x = 4) = P(x = 2)$ find the values probability of success also find mean and variance.

[Ans: p = 1/4, np = 3/2, npq = 9/4]

17. For a binomial variate n = 5 and $4.P(x = 2) = 6.P(x = 3)$ find the values probability of success also find mean.

[Ans: p = 0.4, np = 2]

18. For a binomial variate n = 5 and $P(x = 0): P(x = 1) = 1: 36$ find the values variance.

[Ans: p =3/4, npq = 9/4]

19. A person tosses three coins simultaneously. He gets Rs. 8 if three heads appear, Rs. 4 if two heads appear and Rs. 2 if one head appears. What penalty should be charged if no head appears in order that game is fair?

[Ans: 1.25]

2. Continuous probability distribution

Continuous data is the data that can be of any value. Over time, some continuous data can change. It may take any numeric value, within a potential value range of finite or infinite. If the variable which take all possible values within a given range and its value not countable than it is called continuous data. The continuous data can be broken down into fractions and decimals, i.e. according to measurement accuracy, it can be significantly subdivided into smaller sections.

Examples: Measurement of height and weight of a student, Daily temperature measurement of a place, Wind speed measured daily, etc.

Normal Distribution

Normal distribution is also known as normal probability distribution which is very useful for continuous random variables. Many statistical data concerned with business and economic problems are displayed in the form of normal distribution. Normal distribution is the cornerstone of the modern biostatistics. It is important for the reason that it plays a vital role in the theoretical and applied statistics. In many natural processes, random variation matches to a particular probability distribution which is known as the normal distribution. In 1733, English mathematicians deMoivre and Laplace first discovered normal distribution. Later 11 in 1812, German mathematician Gauss rediscovered it to analyse astronomical data, and it consequently became to be known as the Gaussian distribution.

Here we study Normal Distribution which Continuous probability distribution.

- **Definition**

A continuous random variable x is said to have random distribution with parameters of mean (μ) and standard deviation (σ) if its density function is,

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \times e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Here,

e and π are mathematical constants
μ is mean
σ is standard deviation
If μ=0 and σ=1 then the variate is called as standard normal variate

- **Properties of normal distribution:**

1. This is a distribution of a continuous variable.
2. μ and σ are the parameters of this distribution.
3. The curve of the normal distribution is symmetrical about mean and it is bell shaped.
4. Mean, median and mode are equal in this distribution.
5. Quartiles are equidistant from median.
6. Its skewness is zero.
7. The total area under the normal curve is 1. The following are some of the important areas of the normal curve:

   Area under $\mu \pm \sigma$ is 68.27%

   Area under $\mu \pm 1.96\sigma$ is 95%

   Area under $\mu \pm 2\sigma$ is 95.45%

   Area under $\mu \pm 2.58\sigma$ is 95%

   Area under $\mu \pm 3\sigma$ is 99.73%
8. The tails of the normal curve do not meet x axis. i.e., The curve is asymptotic to x axis.
9. Mean deviation about mean= $\frac{4}{5} \times$ Standard deviation (approximately)
10. Quartile deviation = $\frac{2}{3} \times$ Standard deviation (approximately)
11. Median deviation = standard deviation
12. The sum of two independent normal variates is also a normal variate.
13. When n is very large, and p and q are not very small, Binomial distribution tends to Normal distribution.
14. In fact, most of the distributions in statistics follow normal distribution when n is large.
15. The spread of a normal distribution is controlled by the standard deviation, σ.
16. The smaller the standard deviation, the more concentrated the data.

- **Questions**

1. Find the value of the average deviation if the standard deviation in a normal distribution is 5.
2. If the average deviation in the normal distribution is 12, find its Quartile deviation.
3. Find $Q1$, if the normal distribution has $Q3$ = 40 and mode = 2.
4. The Quartiles of a normal distribution are 8.64 and 14.32 respectively. So, find its mean and standard deviation.
5. If $Q3$ = 60 and the standard deviation is 3 in normal distribution, state the value of $Q1$ as well as the Mode.
6. If X is normal variable whose mean is 10 and S.D. is 2, p($x \leq 10$).
7. If X:N(25, 25), 1) p($20 \leq x \leq 35$) 2) p(x < 35).

8. The probability function of normal distribution is
$P(x) = \frac{1}{\sqrt{2\pi 5}} \times e^{-\frac{1}{50}(x-50)^2}$, Find p(x ≥ 60).
9. The probability function of normal variable x is $P(x) = \frac{1}{\sqrt{2\pi 5}} \times e^{-\frac{1}{50}(x-50)^2}$, find
   A. p(x ≤ 45)
   B. p(x ≥ 55).
10. The probability function of normal distribution is $P(x) = \frac{1}{\sqrt{2\pi 8}} \times e^{-\frac{1}{128}(x-40)^2}$ then find
   A. p(32 ≤ x ≤ 40)
   B. p(x ≤ 48).
11. The mean and standard deviation of the wages of 6000 workers engaged in a factory are Rs. 4800 and Rs. 400 respectively. Assuming the distribution to be normally distributed estimate
   A. Percentage of workers getting wages more than Rs. 5400.
   B. Number of workers getting wages between Rs. 4200 and Rs. 4600.
12. An aptitude test for selecting officers in a bank was conducted on 1000 candidates. The average score is 42 and the standard deviation of scores is 24. Assuming normal distribution for the scores, find
   A. The number of candidates whose scores exceeds 58.
   B. The number of candidates whose scores lies between 30 and 60.
13. The average the weights of 4000 girls of a university are normally distributed. The mean and S.D. of weights are 95 lbs and 7.5 lbs. How many girls weigh between 100 and 110 lbs?
14. The income distribution of officers follows a normal distribution. The average income of an officer was Rs. 1,50,000 and the S.D. was Rs. 5000. If there were 242 officers drawing salary then find above 1.62,000 Rs. how many officers were there in the company?
15. In an intelligence test administered to 1000 children the average score is 42 and its s.d. is 24. Assuming that the scores are normally distributed find.
   A. The number of children exceeding score 50
   B. The number of children getting score between 30 and 54
   C. The minimum score of the most intelligent 100 students.
16. A normal distribution has mean 77, find the variance if 20% of the area under normal curve lies to the right of 90.
17. In a normal distribution 7% of the observations are less than 35 and 89% of the observations are less than 63. Find mean and standard deviation of the distribution.

❖ Sampling Techniques