End Course Summative Assignment

Video Link: **interview_ques_statistics_video**

Problem Statement: Write the Solutions to the Top 50 Interview Questions and Explain any 5 Questions in a Video

1. What is a vector in mathematics?

In mathematics, a vector represents a quantity that has both the size and the direction. We can think of it like an arrow showing both how much of something there is and which way it's going. For instance, imagine a wind blowing in a certain direction with a certain strength - that's a vector.

Picture yourself behind the wheel of a car. Your speedometer informs you that you're travelling at 60 miles per hour. However, merely stating your speed isn't sufficient. To provide a complete description of your motion, you must also indicate the direction in which you're moving, such as "north" or "east."

In this case:

- The magnitude of your velocity (speed) is 60 miles per hour.
- The direction of your velocity is the direction in which you're driving.

So, velocity can be represented as a vector: $\vec{v}$ = 60 mph-east.

2. How is a vector different from a scalar?

The differences can be stated as below:

Vectors	Scalars
Vector quantities have both magnitude and direction.	Scalar quantities have magnitude or size only.
Vector quantities can exist in one, two, or three-dimension.	Scalar exists in one dimension only.
Represented by arrows or bold letters.	Represented by regular letters or numbers.
Changes in a vector quantity can involve adjustments to its magnitude, direction, or both at the same time.	Whenever there is a change in a scalar quantity, it can correspond to a change in its magnitude.
When performing mathematical operations	Performing any mathematical operation with

with two or more vectors, the outcome can either be a scalar or another vector. For instance, when you calculate the dot product of two vectors, you get a scalar value, whereas the cross product, addition, or subtraction of two vectors yields another vector.	more than two scalar quantities will result in another scalar quantity.
E.g-> 1. The ***velocity*** vector tells us not only how fast you're moving but also in which direction you're heading. 2. The ***displacement*** of the person is a vector because it indicates both the distance and the direction travelled. For instance, if the person walks 10 metres north, the displacement vector would be represented as 10 metres north.	E.g-> 1. The ***speed*** scalar gives us information about how fast you're going but doesn't specify the direction of travel. 2. ***Distance*** is considered a scalar quantity because it only has magnitude and no direction associated with it.For example, if you're measuring the distance between two cities on a map.

3. What are the different operations that can be performed on vectors?

In mathematics, there are various operations applicable to vectors. Among the essential ones are:

- **Vector Addition:** Combining two vectors to produce a new vector that represents their combined effect. For example,

  If you have $\vec{a} = (a_x, a_y)$ and $\vec{b} = (b_x, b_y)$, then their difference $\vec{a} + \vec{b}$,

  $$\vec{c} = \vec{a} + \vec{b}$$

  $$\vec{c} = (a_x + b_x,\ a_y + b_y)$$

- **Vector Subtraction:** The difference between two vectors can be calculated by subtracting the corresponding components of one vector from the other. For example,

  If you have $\vec{a} = (a_x, a_y)$ and $\vec{b} = (b_x, b_y)$, then their difference $\vec{a} - \vec{b}$,

  $$\vec{c} = \vec{a} - \vec{b}$$

  $$\vec{c} = (a_x - b_x, a_y - b_y)$$

- **Dot Product (Scalar Product):** It involves multiplying corresponding components of the vectors and then adding up the outcomes. For example,

  If you have $\vec{a} = (a_x, a_y)$ and $\vec{b} = (b_x, b_y)$, then dot product can be represented as,

$$\vec{c} = \vec{a} . \vec{b}$$

- **Cross Product (Vector Product):** A mathematical operation that accepts two vectors as input and produces another vector that is perpendicular to both of the initial vectors.

  If you have $\vec{a} = (a_x, a_y)$ and $\vec{b} = (b_x, b_y)$, then cross product can be represented as,

$$\vec{c} = \vec{a} \times \vec{b}$$

- **Scalar Multiplication:** This operation involves multiplying each component of the vector by the scalar value.

  If you have $\vec{a} = (a_x, a_y)$ and a scalar *k,* then scalar multiplication will be,

$$k\vec{a} = k(a_x + a_y)$$

4. How can vectors be multiplied by a scalar?

Scalar multiplication involves scaling the vector by a single numerical value (scalar), which affects both its magnitude and direction. Here's how scalar multiplication works:

If you have two dimensional vector $\vec{a} = (a_x, a_y)$ and a scalar *k,* then scalar multiplication will be,

$$k\vec{a} = (ka_x + ka_y)$$

If you have three dimensional vector $\vec{a} = (a_x, a_y, a_z)$ and a scalar *k,* then scalar multiplication will be,

$$k\vec{a} = (ka_x + ka_y + ka_z)$$

Here's what happens when you multiply a vector by a scalar:

- The magnitude (length) of the vector is multiplied by the absolute value of the scalar. If the scalar is negative, the direction of the resulting vector is reversed, but its magnitude remains the same.
- If the scalar is greater than 1, the resulting vector is longer than the original vector in the same direction.

- If the scalar is between 0 and 1, the resulting vector is shorter than the original vector in the same direction.
- If the scalar is 0, the resulting vector is the zero vector (a vector with all components equal to zero), regardless of the original vector.

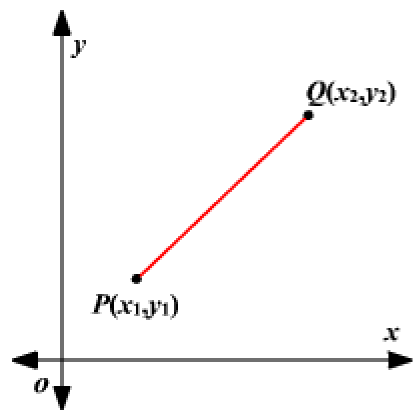
5. What is the magnitude of a vector?

The magnitude of a vector refers to its size or length, representing the extent or strength of the quantity it denotes, regardless of its direction. For instance, if the magnitude of the velocity vector is 50 kilometres per hour, it signifies that the car is moving at a speed of 50 kilometres per hour, without specifying its direction.

If you have two dimensional vector $\vec{a} = (a_x, a_y)$, then its magnitude will be

$$|\vec{a}| = \sqrt{a_x^2 + a_y^2}$$

6. How can the direction of a vector be determined?



The direction of a vector is established by the angle it creates with a horizontal line. Typically, this angle is measured counterclockwise from the positive x-axis in two-dimensional space.

The direction of a vector whose endpoints are given by the position vectors $P(x_1, y_1)$ and $Q(x_2, y_2)$,

$$(x, y) = (x_2 - x_1, y_2 - y_1)$$
$$\alpha = tan^{-1}|y/x|$$

The direction of the vector θ is calculated by using the following rules depending on which quadrant (x, y) lies in:

Quadrant in which (x, y) lies	θ (in degrees)
1	α
2	180-α
3	180+α
4	360-α

7. What is the difference between a square matrix and a rectangular matrix?

The differences can be stated as below:

Square Matrix	Rectangular Matrix
Matrix with equal number of rows and columns (nxn)	Matrix with unequal number of rows and columns (mxn)
Example: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}_{(3x3)}$	Example: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}_{(4x3)}$
Symmetrical in shape.	Asymmetrical in shape.
Determinant and eigenvalues are well-defined.	Determinant or eigenvalues are not well-defined.
Inverse may exist if nonsingular.	Inverse generally does not exist.

8. What is a basis in linear algebra?

In the realm of linear algebra, a basis comprises vectors that fulfil two conditions: linear independence and spanning the entire vector space.

Linear independence: Linear independence denotes that within a set of vectors, none can be represented as a combination of the others using scalar multiplication. Put simply, no vector in the set can be expressed as a sum of the other vectors multiplied by scalars.

Spanning: Spanning implies that a set of vectors covers the entire vector space if any vector within that space can be expressed as a combination of the vectors in the set.

Essentially, any vector in the space can be reached by scaling and adding the vectors from the set.

Let's denote a vector space **V** and a basis **B** for **V** by:

$$V = span(\vec{v_1}, \vec{v_2}, \vec{v_3}, \ldots\ldots, \vec{v_n})$$

where $\vec{v_1}, \vec{v_2}, \vec{v_3}, \ldots\ldots, \vec{v_n}$ are the set of basis vectors ***B***.

Then, the formula for expressing a vector $\vec{v}$ in terms of the basis vectors $\vec{v_1}, \vec{v_2}, \vec{v_3}, \ldots\ldots, \vec{v_n}$ wil be,

$$\vec{v} = c_1\vec{v_1} + c_2\vec{v_2} + \ldots\ldots c_n\vec{v_n}$$

where $c_1, c_2, \ldots, c_n$ are scalars.

9. What is a linear transformation in linear algebra?

In linear algebra, a linear transformation is a function that preserves vector addition and scalar multiplication. Simply put, it is a mapping that transfers vectors from one vector space to another while preserving the key properties of vector operations.

Suppose U(F) and V(F) are two vector spaces over the field F.
Then, a mapping f: UV is considered a linear transformation if the following conditions are satisfied:

1. **Additivity:** $f(\vec{\alpha} + \vec{\beta}) = f(\vec{\alpha}) + f(\vec{\beta})$
2. **Homogeneity:** $f(c\vec{\alpha}) = cf(\vec{\alpha})$

Here, $\vec{\alpha}, \vec{\beta} \in U$ and $\forall c \in F$

10. What is an eigenvector in linear algebra?

A square matrix's eigenvector is a non-zero vector that when multiplied by a given matrix yields a scalar multiple of the vector. Assume A is a square matrix and v is a non-zero vector. The product of A and v is defined as the product of a scalar quantity λ and the supplied vector, such that:

$$A\vec{v} = \lambda\vec{v}$$

Where:

- $\vec{v}$ = Eigenvector
- λ be the scalar quantity that is termed as eigenvalue associated with given matrix A.

11. What is the gradient in machine learning?

Finding the optimum value (Minimum/Maximum) of an objective function is the goal of the iterative optimization process known as gradient descent. When updating a model's parameters to minimise a cost function, it is one of the most popular optimization approaches used in machine learning projects.
The primary aim of gradient descent is to identify the ideal model parameters that provide the best accuracy on training and testing datasets. A gradient is a vector in gradient descent that indicates the direction of the function's steepest increase at a given moment. By moving against the gradient, the algorithm can progressively drop towards the function's lower values until it reaches the function's minimum.
The gradient descent algorithm for linear regression can be expressed as:

$$\theta_j = \theta_j - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial \theta_j}$$

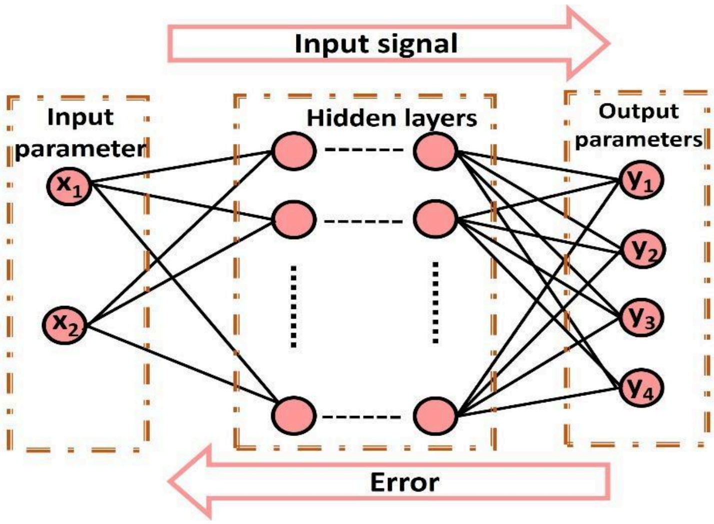
Where:

- J → objective function
- α → learning rate
- θ_j → weight of the hypothesis.

12. What is backpropagation in machine learning?

In machine learning, backpropagation is a go-to technique for training neural networks, especially those that are feed-forward. This method is all about fine-tuning the network's weights and biases to lower the error or loss function. Every round of training, known as an epoch, involves adjusting these parameters to move closer to the optimal values that minimise the error. It's like finding your way down a hill by following the slope. And to do this, we often rely on popular optimization methods like gradient descent or stochastic gradient descent.

Now, computing the gradient in backpropagation is crucial because it tells us how to tweak those weights and biases to reduce the error. We employ a fundamental mathematical concept called the chain rule from calculus to navigate through the intricate layers of the neural network. It's like untangling a tangled mess, step by step, to understand how each parameter affects the final outcome. This process helps the model learn and improve its performance over time.



The Backpropagation algorithm works by two different passes, they are:

- **Forward pass:**
  In the forward pass of a neural network, the input data is first sent to the input layer. These raw inputs are then processed in the hidden layers, where computations are performed. If there are multiple hidden layers, like in the illustration, the output of one hidden layer can become the input for the next. Before applying the activation function, bias is added to the weighted sum of inputs in each neuron of the hidden layer.

  Activation functions such as ReLU are frequently applied in the hidden layers of neural networks. ReLU operates by returning the input value if it's positive; otherwise, it returns zero. This inclusion of non-linearity in the model aids in capturing intricate patterns within the data, enhancing the network's ability to learn and generalise effectively. The outputs from the last hidden layer are then fed into the output layer to compute the final prediction. The output layer may use the softmax function, which converts the weighted outputs into probabilities for each class. This helps in making predictions by assigning probabilities to each class.
- **Backward pass:**
  During the backward pass, the network learns from its mistakes by transmitting the error back through the layers, helping it improve. To assess the error from the forward pass, a widely utilised approach is mean squared error (MSE), which quantifies the disparity between predicted and actual outputs.

  Once the error is computed at the output layer, it's propagated backward through the network, layer by layer. A key part of this backward pass is calculating the gradients for each weight and bias. These gradients guide us in adjusting the weights and biases to minimise errors in the next forward pass. The chain rule is employed to efficiently compute these gradients.

Moreover, the activation function also plays a vital role in backpropagation. It calculates gradients using the derivative of the activation function, aiding in error propagation and adjustment of weights and biases.

13. What is the concept of a derivative in calculus?

The derivative of a function shows how fast it changes with respect to its input. It's like finding the slope of a curve at a particular spot. For example, if you have a function that tells you where an object is over time, its derivative tells you how fast the object is moving at any moment.

One can understand derivatives in the following ways:
(a) A derivative measures how a function changes with respect to the change in its input. It indicates the rate at which the function's output changes with respect to changes in its input.
(b) Geometrically, the derivative at a point on a curve represents the slope of the tangent line to the curve at that point. When the derivative is positive, it means the function is increasing at that point; if it's negative, it's decreasing. Zero indicates a local maxima or minima.
(c) Derivatives provide the instantaneous rate of change of a function at a specific point, rather than an average rate of change over an interval.

Mathematically the derivative of a function $f(x)$ with respect to x is denoted by, $f'(x)$ or $\frac{df}{dx}$.
In order to train machine learning models, optimization strategies like gradient descent depend on derivatives. They guide the process of adjusting model parameters to minimise a loss function and improve model performance.

14. How are partial derivatives used in machine learning?

In the field of machine learning, partial derivatives are essential, especially for optimization techniques like gradient descent. These derivatives aid in our understanding of how a function alters in light of its input variables. In machine learning, where complex models and high-dimensional data are frequently encountered, an understanding of partial derivatives becomes essential to efficiently optimise model parameters.

Partial derivatives are important in machine learning for optimising models, such as neural networks, and understanding feature impact. They're crucial in algorithms like gradient descent, guiding parameter updates to minimise the cost function. In neural networks, backpropagation uses partial derivatives to compute gradients for updating weights. Additionally, they inform feature engineering by revealing how feature changes affect the cost function, aiding in feature selection and creation.

If a function of many variables is only related to one of the variables while keeping the others constant, it is said to have a partial derivative in multivariable calculus. For example $f'\left(x_1,\ x_2, x_3, \ldots . x_n\right)$ is the partial derivative of a function $f\left(x_1,\ x_2, x_3, \ldots . x_n\right)$ with respect to x_i.

15. What is probability theory?

Probability theory is a branch of mathematics and statistics focused on determining the likelihood of random events occurring. It employs two approaches: theoretical probability and experimental probability. Theoretical probability relies on logical deduction without conducting actual experiments. In contrast, experimental probability is derived from historical data obtained through repeated experiments.

$$P(A) = \frac{Number\ of\ possible\ outcomes\ to\ A}{Total\ number\ of\ possible\ outcomes}$$

Example: Suppose we aim to determine the likelihood of rolling a 4 on a fair six-sided die. Since there's only one way to roll a 4, out of a total of six possible outcomes {1, 2, 3, 4, 5, 6}, we have 1 favourable outcome. Utilising probability theory, we calculate the probability of rolling a 4 by dividing the number of favourable outcomes (1) by the total possible outcomes (6). Therefore, the probability of rolling a 4 is 1/6 or approximately 0.167.

16. What are the primary components of probability theory?

The primary components of probability theory include:

1. **Sample Space:** The sample space represents all potential outcomes of a random experiment. For instance, if we were to simultaneously flip a fair coin and roll a fair six-sided die, the sample space could be listed as:
2. (H,1), (H,2), (H,3), (H,4), (H,5), (H,6), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6).
3. **Events:** Subsets of the sample space representing specific outcomes or combinations of outcomes.
   Now in the above example if we want to find only the outcomes which have even numbers; then the set of all such possibilities can be given as:
   (H,2), (H,4), (H,6), (T,2), (T,4), (T,6)
4. **Probability Measure:** We give an event A a probability measure, P(A). The probability of the event is indicated by this number, which ranges from 0 to 1. It is

highly unlikely that the event A will occur if P(A) is near to 0. Conversely, if P(A) is near 1, then A has a high probability of happening.

5. **Probability Axioms:** The basic rules for probabilities include three main ideas. First, probabilities can't be negative. Second, when you add up the chances of all possible outcomes, it equals 1. And third, if you have events that can't happen at the same time, the probability of them both happening is the sum of their individual probabilities.
   Axioms of Probability:
   Axiom 1: For any event A,
   $$P(A) \geq 0$$
   Axiom 2: Probability of the sample space S is,
   $$P(S) = 1$$
   Axiom 3: If A1,A2,A3,⋯ are disjoint events, then
   $$P(A1 \cup A2 \cup A3.....) = P(A1) + P(A2) + P(A3).....$$
6. **Random Variables:** Random variables are quantities that represent the outcomes of a random process. They can take on either discrete or continuous values.
   - **Discrete random variables** have a finite or countably infinite number of distinct outcomes. For example, consider an experiment where a coin is flipped three times. If we let X represent the number of heads obtained, it's a discrete random variable with values of 0, 1, 2, or 3.
   - On the other hand, **continuous random variables** can take on any value within a specific range or interval, with an infinite number of possible outcomes. For instance, an experiment measuring the amount of rainfall in a city over a year or determining the average height of a randomly selected group of 25 people involves continuous random variables.
7. **Probability Distributions:** A probability distribution is a statistical tool that illustrates all possible values a random variable can take within a defined range, along with their corresponding likelihoods. While the range is determined by the minimum and maximum possible values, the exact positioning of each potential value on the distribution is influenced by various factors.
8. **Expected Value:** The expected value of a random variable is the weighted average of its possible outcomes, taking into account their respective probabilities. It helps anticipate what outcomes are likely to occur over a large number of trials.
9. **Variance and Standard Deviation:** Measures of the spread or dispersion of a probability distribution.
10. **Joint, Marginal, and Conditional Probabilities:** Joint probability demonstrates the possibility of many events occurring together, while marginal probability

focuses on the probability of individual events. Conditional probability determines the likelihood of an event provided that another event has occurred..

17. What is conditional probability, and how is it calculated?

Conditional probability measures the chance of an event happening once another event has already taken place. In simpler terms, it calculates the probability of one event occurring under the condition that another event has already happened. It's denoted as P(A | B), indicating the probability of A occurring given that B has already occurred.

$$P(B|A) = \frac{P(A \cap B)}{P(B)}$$

Where:

- The notation P(A ∩ B) denotes the probability of both events A and B happening at the same time.
- Meanwhile, P(B) represents the probability of event B occurring.

For example, let us consider the case of rolling two dice, the sample space will be,

{(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)}

Here, Sample space is $S = 36$.

Let us consider two events,

- Event A represents rolling a 5 on the first die.
- Event B represents obtaining a sum of 7.

Exactly, we can calculate these probabilities using conditional probability.

Then, $P(B|A)$ is the probability of getting a sum of 7 when we know that the first die shows a 5.

and, $P(A|B)$ is the probability of getting a 5 on the first die when we know that the sum of the two dice is 7.

We can calculate both as below,

1. **P(B|A):**
   When event A has occurred (rolling a 5 on the first die), there are six possible outcomes: (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), and (5, 6). Among these outcomes, only one of them, (5, 2), results in a sum of 7.

   Thus, $P(B|A) = \frac{1}{36}$.
2. **P(A|B):**

When event B has occurred (obtaining a sum of 7), the possible outcomes are (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1). Among these outcomes, only one of them, (5, 2), has a 5 on the first die.

Thus, $P(A|B) = \frac{1}{36}$.

18. What is Bayes theorem, and how is it used?

Bayes' Theorem is a mathematical principle that is used in probability and statistics to compute conditional probability. Essentially, it helps determine the likelihood of an event occurring given its relationship to another event.

Mathematically, Bayes' Theorem is expressed as:

$$P(A|B) = \frac{P(B|A)\ x\ P(A)}{P(B)}$$

Where:

- P(A|B) → represents the probability of event A occurring given that event B has occurred.
- P(B|A) → represents the probability of event B occurring given that event A has occurred.
- P(A) and P(B) are the probabilities of events A and B occurring, respectively.

For example: An urn holds 5 red and 5 black balls. One ball is randomly drawn, its colour noted, and returned to the urn. Then, two more balls of the same colour are added. Subsequently, a second ball is drawn randomly. What's the likelihood that this second ball is red?

Let's use Bayes' theorem to solve this problem:

- A: The event where the second ball drawn is red.
- B: The event where the first ball drawn is red.

We want to find, P(A|B), the probability that the second ball drawn is red given that the first ball drawn is red.

By Bayes' theorem:

$$P(A|B) = \frac{P(B|A)\ x\ P(A)}{P(B)}$$

Now let us interpret values,

- **P(B|A)**: The probability of drawing the first ball red given that the second ball drawn is red is equivalent to the probability of drawing a red ball initially, which is $\frac{1}{2}$.
- **P(A)**: Probability of drawing the second ball red. This is the same as the probability of drawing a red ball initially, which is also $\frac{1}{2}$.

- **P(B)**: Probability of drawing the first ball red. Since each color has an equal chance initially, this is $\frac{1}{2}$.

Now, we plug these values into Bayes' theorem:

$$P(A|B) = \frac{\frac{1}{2} x \frac{1}{2}}{\frac{1}{2}}$$

$$P(A|B) = \frac{1}{2}$$

Therefore, the likelihood of the second ball drawn from the bag being red is. $\frac{1}{2}$.

19. What is a random variable, and how is it different from a regular variable?

A **Regular Variable** is represented by an alphabetical character which stands for an unknown value.
For Example:
Given the equation: $2x + 5 = 15$
Here, x is the variable whose value is unknown, solving it,

$$2x = 15 - 5 = 10$$

$$x = \frac{10}{2}$$

Hence, upon evaluation, **x=5**
However, a **Random Variable** in statistics behaves a bit differently. Instead of just one value, it can actually represent a whole set of possible values, and it's chosen randomly from among those. Unlike the variables we're familiar with in algebra, which typically represent just one value at a time, a Random Variable has this flexibility to take on different values randomly from its set.
For Example:
If there is a random variable X such that,

$$X = \{0, 1, 2, 3, 4, 5\}$$

In the case of a random variable like X, which has possible values of 0, 1, 2, 3, 4 or 5 , each of these values could occur randomly. However, the probability of each value occurring might be different. Random variables can be either **discrete** or **continuous**.
To prevent mixing it up with regular variables, we use capital letters for random variables.

20. What is the law of large numbers, and how does it relate to probability theory?

The Law of Large Numbers, a fundamental concept in statistics and probability theory, states that as the quantity of independent and identically distributed random variables rises, their average will converge towards the expected value or mean of the underlying distribution.

It suggests that as the number of trials or experiments increases, the average outcome will approach the expected value of the random variable. Put simply, the more times you repeat an experiment, the closer the average result will be to what you expect.

In probability theory, the Law of Large Numbers establishes a mathematical basis for comprehending and forecasting the behaviour of random events. It guarantees that despite significant variations in individual outcomes, the average behaviour across numerous trials will exhibit consistent patterns. This principle holds critical importance in diverse domains like statistics, finance, and science, where uncertainty and randomness are prominent factors.

21. What is the difference between discrete and continuous probability distributions?

The differences can be stated as below:

Discrete Probability Distribution	Continuous Probability Distribution
Describes probabilities associated with discrete variables, which can only take on specific, separate values (usually integers).	Describes the probability for continuous variables, which have the ability to assume any value within a specified interval.
The probability mass function (PMF) is used to represent the probabilities of each possible value of the discrete variable.	The probability density function (PDF) is used to represent the probabilities of different values within the continuous range of the variable.
Probability is assigned to individual points or intervals on a discrete scale.	Probability is assigned to intervals or ranges of values on a continuous scale, rather than individual points.
Examples include the probability of rolling a particular number on a six-sided die or the number of heads obtained in multiple coin flips	Examples include the distribution of heights or weights in a population or the time taken for a process to complete.

22. What are some common measures of central tendency, and how are they calculated?

Measures of central tendency assist in determining the typical value within a dataset. The most frequently used measures include:

1. **Mode:** The mode represents the value that occurs with the highest frequency in a dataset. A dataset can have one mode (unimodal), multiple modes (multimodal), or no mode if all values occur with equal frequency.
2. **Median:** The median is the central value of a dataset when the values are sorted in ascending or descending order. If the dataset has an odd number of values, the median is the middle value. If it has an even number of values, the median is the average of the two middle values.
3. **Mean:** The mean is obtained by adding up all the values in a dataset and then dividing by the total number of values.

Let us see an example:

Suppose we have the following exam scores: 85, 90, 78, 92, 85 and 88.

1. **Mode:** In this case, the mode is 85 because it appears more frequently than any other score in the dataset.

Marks	Frequency
85	2
90	1
78	1
92	1
88	1

2. **Median:** To find the median, we start by arranging the scores in ascending order: 78, 85, 85, 88, 90, 92. As there are 6 scores (an even number), we'll first identify the two middle values.

$$m1 = \frac{n}{2}\ and\ m2 = \frac{n}{2} + 1$$

After solving, $m1 = 3\ and\ m2 = 4$. The 3rd and 4th terms of the distribution are 85 and 88 respectively. Lastly, we will add both and divide by 2 to calculate the median value.

$$Median = \frac{85+88}{2} = 86.5$$

Hence, the median is **86.5.**

3. **Mean:** To calculate the mean, we add up all the scores and then divide by the total number of scores.

$$Mean = \frac{85+90+78+92+85+88}{6} = 86.6$$

The mean of the scores is **86.6**.

23. What is the purpose of using percentiles and quartiles in data summarization?

Quartiles and percentiles are metrics that measure the variability or spread of data. They both fall under the category of quantiles and are essential statistical measures used for data summarization and analysis. Here's how they serve in data summarization:

1. **Understanding Distribution:** Percentiles and quartiles help understand how data is spread out across different values. By dividing the dataset into intervals based on percentiles or quartiles, you can see how the data is distributed from the lowest to the highest values.
2. **Identifying Central Tendency:** Quartiles, specifically the second quartile, serve as a measure of central tendency, equivalent to the median. They split the dataset into four equal segments, where the median marks the value below which 50% of the data falls and above which the remaining 50% falls. Percentiles offer a more detailed breakdown by dividing the data into hundredths.
3. **Handling Outliers:** Percentiles and quartiles offer robust measures of central tendency, less susceptible to outlier influence than the mean. They provide clearer insights into typical dataset values, even in the presence of extreme data points.
4. **Comparing Datasets:** Percentiles and quartiles enable easy comparison between different datasets. By comparing the distribution of percentiles or quartiles, you can quickly assess similarities or differences between datasets in terms of their central tendency and spread.
5. **Identifying Skewness and Spread:** Percentiles and quartiles help identify skewness in the data distribution. For instance, if the difference between the third quartile (Q3) and the first quartile (Q1) is large, it indicates a wide spread or variability in the dataset.

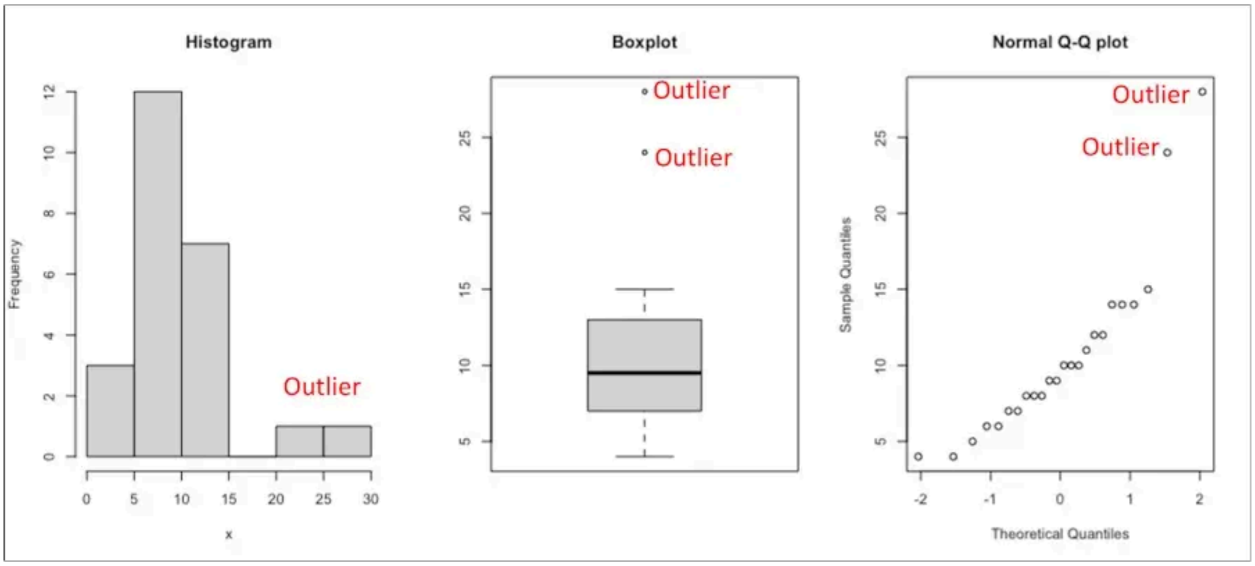
In summary, percentiles and quartiles are vital for summarising data, offering valuable insights into its distribution, central tendency, and variability. These insights are crucial for making well-informed decisions across various fields such as finance, healthcare, and social sciences.

24. How do you detect and treat outliers in a dataset?

Outliers are data points that differ dramatically from the rest of the dataset. They're typically unusual or abnormal observations that can throw off the overall pattern of the data. Outliers often arise as a result of inconsistencies in data entry or errors in observations.

Detecting Outliers:

- **Visual Inspection:** Plotting the data using histograms, box plots, scatter plots, or Q-Q plots allows for visual identification of outliers. Unusual points that stand far from the bulk of the data may indicate outliers.



- **Statistical Methods:** Statistical techniques like z-score, interquartile range (IQR), or modified z-score quantify how much a data point deviates from the mean or median. For instance, the **Z-score** measures how many standard deviations a data point is from the mean. Data points with a Z-score above a certain threshold (typically 2 or 3) are considered outliers and can be flagged or treated accordingly. Also, the **IQR** is the range between the first quartile (25th percentile) and the third quartile (75th percentile) of the data. Outliers are often identified as data points that fall below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR.
- **Machine Learning Techniques:** Advanced methods such as clustering or anomaly detection algorithms can automatically identify outliers. These algorithms learn patterns in the data and flag observations that deviate significantly from those patterns as outliers.

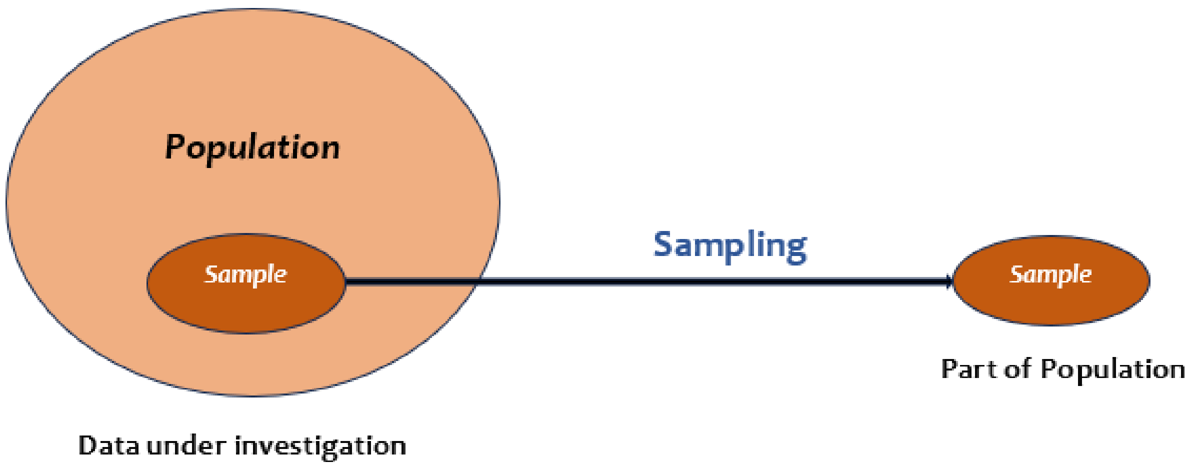
Handling Outliers:

- **Understanding the Data**: Comprehensive data exploration is essential to discern the nature and origin of outliers, distinguishing between genuine anomalies and data errors.

- **Data Transformation:** Applying transformations such as logarithmic or square root functions can normalise the data distribution, reducing the influence of outliers.
- **Winsorization**: This method involves capping extreme values by replacing them with values at a specified percentile, ensuring the model's robustness to outliers.
- **Trimming:** Trimming involves discarding a certain percentage of extreme data points, tailoring the dataset to focus on the core distribution and mitigate outlier impact.
- **Imputation:** For erroneous or missing values suspected to be outliers, imputation techniques like mean, median, or regression-based estimations can provide more reliable data for model training.
- **Robust Statistical Methods:** Leveraging robust statistical metrics such as median and median absolute deviation (MAD) instead of mean and standard deviation can enhance model resilience to outliers.
- **Domain Knowledge Integration:** Incorporating domain expertise helps discern whether outliers are meaningful outliers or anomalies, guiding the decision to retain or treat them.
- **Model Adaptation:** Adapting machine learning models to handle outliers explicitly, or using models inherently robust to outliers like decision trees, can improve predictive performance.
- **Segregation and Separate Modelling:** In certain cases, creating separate models to analyse outliers separately from the main dataset can provide deeper insights and prevent outlier influence on the primary model.

25. What is sampling in statistics, and why is it important?

Sampling is a technique that enables us to gather information about a population by studying or examining a smaller group (sample), without needing to examine every single individual of the population.



Sampling is essential for several reasons:

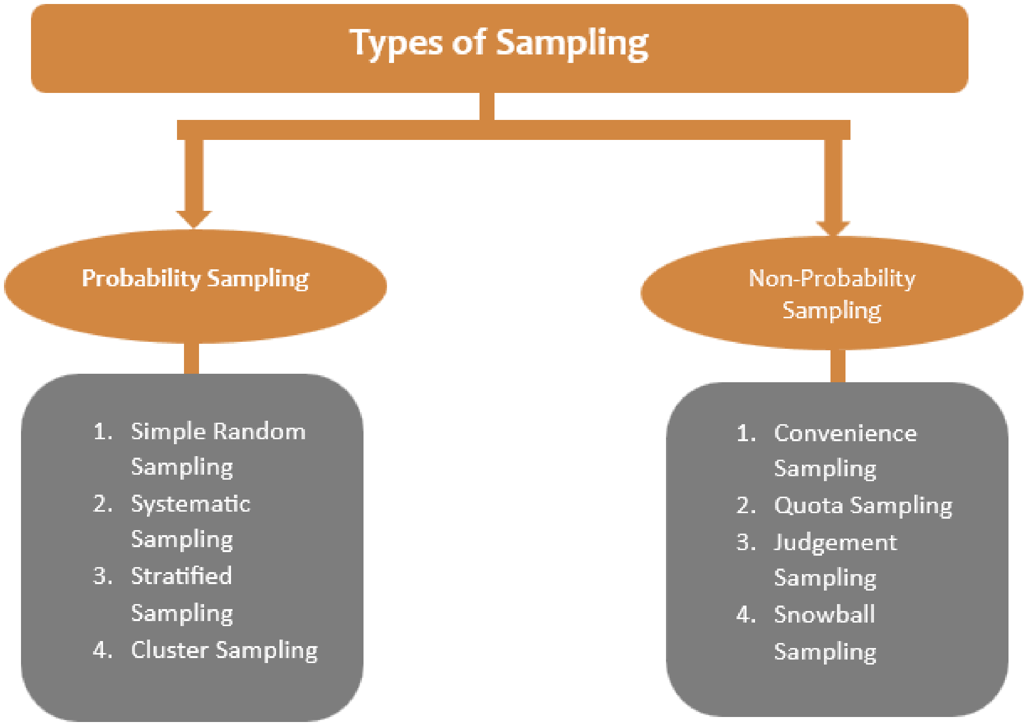
1. **Efficiency:** Selecting a sample takes less time and resources compared to studying every individual in a population.

2. **Cost-effectiveness:** Sampling is a more economical approach than attempting to study the entire population, especially when populations are large or dispersed.

3. **Practicality:** Analysing a sample is more manageable and feasible than analysing an entire population, making it easier to draw conclusions and make inferences about the population's characteristics.

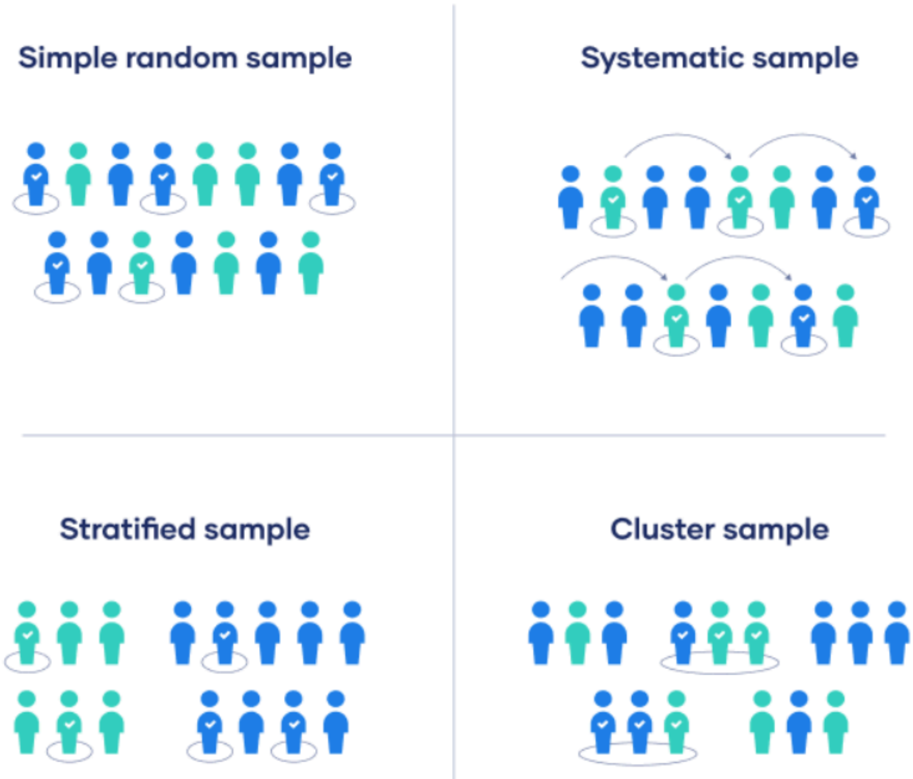
26. What are the different sampling methods commonly used in statistical inference?

Below are the types of sampling methods commonly used:



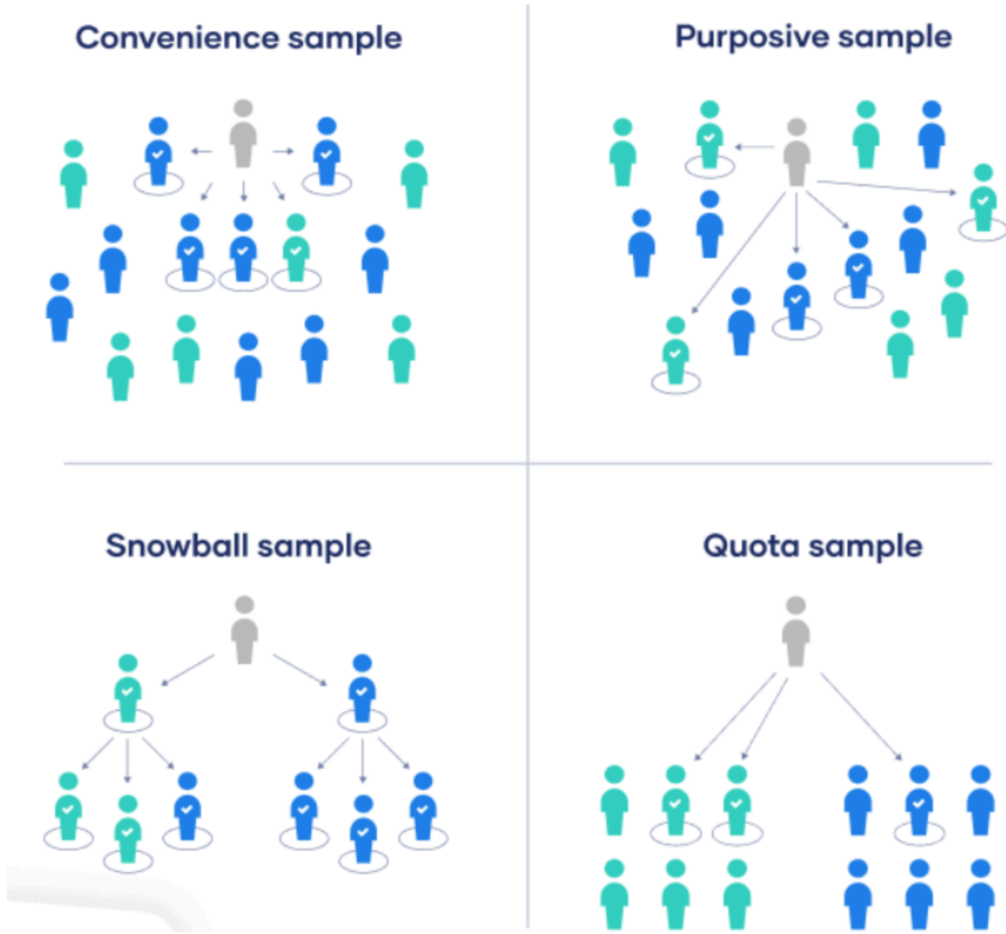
Probability Sampling: In probability sampling, each member of the population has an equal opportunity to be chosen. There are four types of probability sampling:

1. **Random Sampling:** This technique involves selecting individuals purely by chance, ensuring that each member of the population has an equal probability of being chosen. While it is the most straightforward sampling method, it may not always capture enough individuals with the desired characteristics.
2. **Systematic Sampling:** In this approach, the first individual is chosen randomly, and subsequent selections are made at regular intervals determined by a fixed "sampling interval." For example, if the population size is x and the sample size is n, then every x/nth individual from the initial selection is chosen.
3. **Stratified Sampling:** This method involves dividing the population into subgroups (strata) based on specific traits such as gender or category. Samples are then selected independently from each subgroup, allowing for a more targeted representation of different population characteristics.
4. **Cluster Sampling:** In cluster sampling, the population is divided into clusters, and entire clusters are randomly selected to be included in the study. This approach is useful when it is impractical or inefficient to sample individuals directly, making it more feasible to sample groups or clusters instead.



Non-Probability Sampling: In non-probability sampling, not all elements have an equal chance of being selected. There are four types of non-probability sampling,

1. **Convenience Sampling:** In non-probability sampling, not all elements have an equal chance of being selected. There are four types of non-probability sampling:
2. **Quota:** In quota sampling, items are chosen based on predetermined characteristics of the population. For example, individuals with numbers in multiples of four may be selected for the sample.
3. **Judgement/ Purposive :** This type of sampling relies on the researcher's expertise to select a sample that aligns closely with the research objectives.
4. **Snowball:** When the population is difficult to access, snowball sampling can be used. This method involves recruiting participants through referrals from existing participants, gradually increasing the sample size like a rolling snowball.



27. What is the central limit theorem, and why is it important in statistical inference?

In probability theory, the central limit theorem (CLT) tells us that as the sample size increases, the distribution of a sample variable tends to resemble a normal distribution, irrespective of the actual shape of the population distribution.

- Sample sizes of 30 or more are considered adequate for the CLT to apply.
- Additionally, an important feature of the CLT is that the average of sample means and standard deviations will match the population mean and standard deviation.

28. What is the difference between parameter estimation and hypothesis testing?

The differences can be stated as below:

Parameter Estimation	Hypothesis Testing
Involves estimating the unknown parameters of a population based on sample data.	It concentrates on drawing conclusions or making decisions about population characteristics using information obtained

	from samples.
Aimed at determining the best estimate or values for population parameters, such as mean or standard deviation.	It's focused on evaluating whether a statement or hypothesis about population parameters aligns with the data gathered from samples..
Common methods include point estimation (e.g., finding the sample mean) and interval estimation (e.g., constructing confidence intervals).	Common methods include comparing sample statistics to hypothesised values or conducting tests using p-values.
An outcome usually refers to a particular numerical figure or a set of values that depict the estimated parameter.	An outcome typically involves deciding whether to accept or reject a null hypothesis based on the evidence provided by the sample data.
Examples include estimating the population mean or variance based on sample data.	Examples include testing whether a new drug has an effect, comparing means of two groups, or testing a hypothesis about a population proportion.

29. What is the p-value in hypothesis testing?

The p-value, a probability ranging from 0 to 1 under the assumption that the null hypothesis holds true, is also referred to as the Probability Value. It represents the lowest significance level at which the null hypothesis is rejected.

The interpretation of the p-value is as follows:

- A low p-value, usually below 0.05, indicates robust evidence against the null hypothesis, implying its rejection.
- Conversely, a high p-value suggests weak evidence against the null hypothesis, implying insufficient grounds for its rejection.

Here is how it works,

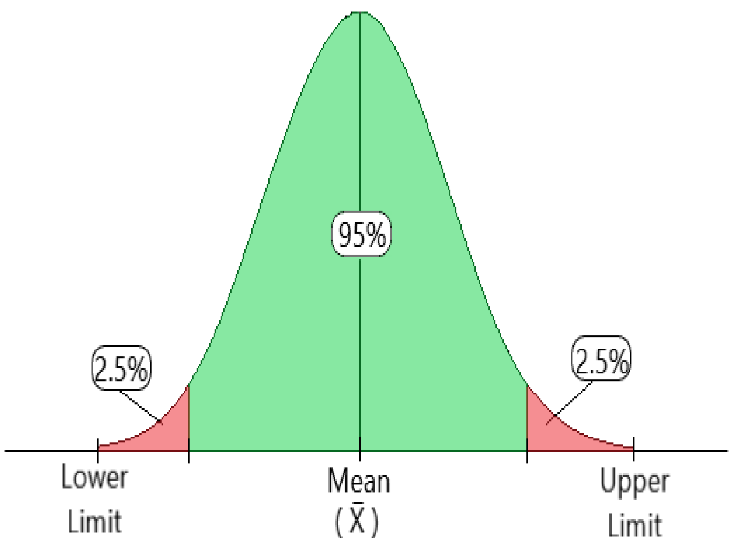
1. **Formulate Hypothesis:** In hypothesis testing, you start with a null hypothesis (H_o) which typically represents a default or no-effect assumption, and an alternative hypothesis(H_1), which represents what you're trying to show.
2. **Collect data:** You collect data relevant to your hypothesis.
3. **Calculate a test statistic:** The choice of test statistic depends on the specific hypothesis test you're conducting. For example, in a t-test, the test statistic is the t-value.
4. Determine the p-value
5. **Make a decision:** You compare the p-value with a predetermined significance level, often represented as α. Commonly used significance levels include 0.05,

0.01, and 0.001. If $p <= \alpha$, reject the null hypothesis and if $p > \alpha$, fail to reject the null hypothesis where, $\alpha \rightarrow level\ of\ significance$.

30. What is confidence interval estimation?

A confidence interval signifies the span of values where you anticipate your estimate to lie if you were to conduct the test again, with a specified level of confidence. It's computed by determining the average of your estimate and then adding and subtracting the variability in that estimate.

Imagine you're trying to guess the average height of people in *abc* city. You take a sample of people and calculate an average height. Now, a confidence interval is like a guess about where you think the real average height is for everyone in the city. But, since you didn't measure everyone, you're not 100% sure. So, you give a range of heights you're pretty confident the real average falls within. For example, if you say you're 95% confident, it means you think your guess will be right 95 times out of 100.



We construct a confidence interval to help estimate what the actual value of unknown population mean is,

$$\overline{x} \pm z_{\alpha/2}(\sigma/\sqrt{n})$$

Where:

- $\overline{x} \rightarrow Sample\ Mean$
- $z_{\alpha/2} \rightarrow Z\ score\ with\ \alpha\ significance\ value$
- $\sigma \rightarrow Standard\ Deviation\ of\ population$
- $n \rightarrow Sample\ size$

31. What are Type I and Type II errors in hypothesis testing?

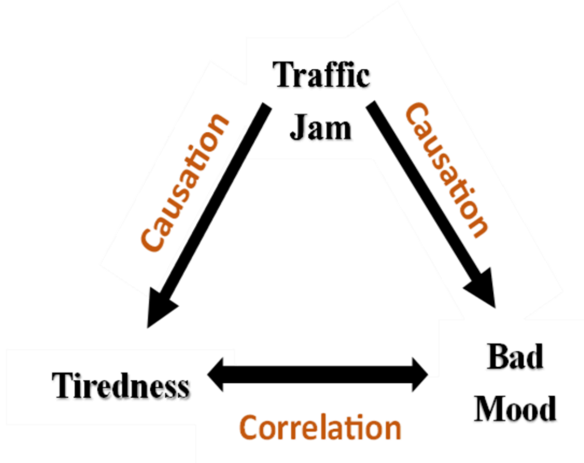
Type I and Type II errors are two types of errors that can arise in hypothesis testing:

1. **Type Ⅰ (False Positive) :** This occurs when a true null hypothesis is rejected. In other words, it occurs when you mistakenly conclude that there is a significant effect or difference when there isn't one in reality.
   Example: You incorrectly conclude that a new drug is effective in treating a disease when, in fact, it has no effect.
   Example: You conclude that a new drug is effective in treating a disease when, in fact, it has no effect.
2. **Type Ⅱ (False Negative) :** This occurs when a false null hypothesis is not rejected. In other words, it happens when you fail to detect a real effect or difference that actually exists.
   Example: You fail to conclude that a new drug is effective in treating a disease when, in reality, it does have an effect.

Null Hypothesis (H_o)	**True**	**False**
Reject	**Type Ⅰ Error** False Positive	Correct Decision True Positive
Fail to reject	Correct Decision True Negative	**Type Ⅱ Error** False Negative

32. What is the difference between correlation and causation?

The differences can be stated as below:



Correlation	Causation
Correlation indicates a relationship between the values of two variables.	Causation refers to one event causing another event to happen.
Correlation does not imply causation	Causation implies correlation, but correlation does not always imply causation.
Both independent and dependent variables are necessary for correlation analysis.	Both independent and dependent variables are required to establish causation.
Example: Tiredness and bad mood are correlated, meaning they often occur together, but one does not necessarily cause the other.	Example: Tiredness and bad mood are caused by traffic jams, indicating a direct influence of one event on the other.

33. What is experiment design, and why is it important?

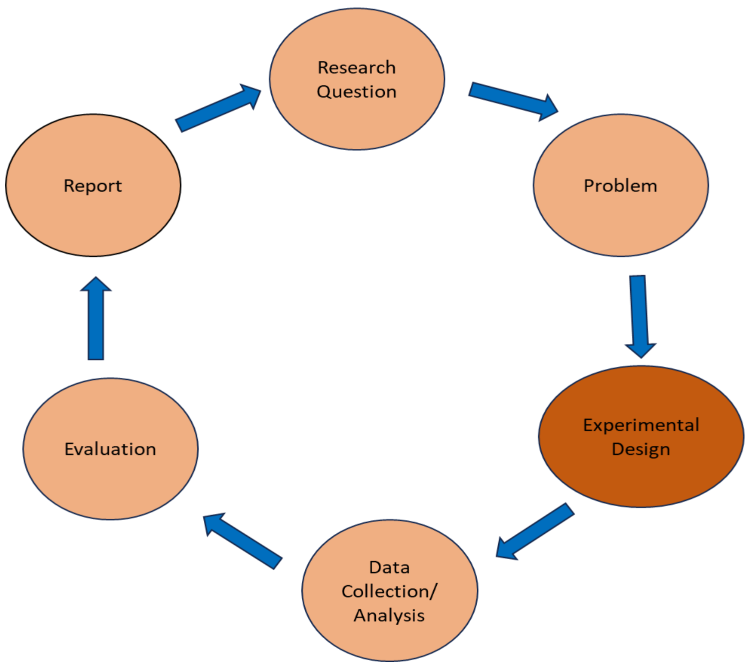
Experimental design entails conducting research in a methodical and regulated fashion to improve accuracy and derive particular conclusions regarding a hypothesis statement. Its main objective is to assess the influence of an independent variable on a dependent variable.

Overview of various experimental designs:

1. **Between-Subjects Design**: Different groups of participants receive different treatments.
2. **Within-Subjects Design:** Participants undergo all treatments, serving as their own controls.

3. **Factorial Design:** Manipulates multiple factors simultaneously to examine their combined effects.
4. **Cross-Over Design:** Participants receive multiple treatments in a specific sequence, serving as their own controls.
5. **Matched-Subjects Design:** Participants are matched based on characteristics before random assignment to treatment groups.
6. **Quasi-Experimental Design:** Lacks random assignment and may rely on existing characteristics or natural variations.

Each design offers unique advantages and is chosen based on the research question, feasibility, and ethical considerations.



Experimental design is of paramount importance in research for several reasons:

1. **Validity of Results:** A good experiment makes sure that any effects seen are because of the things we changed on purpose, not because of other factors we didn't consider. This makes the results more trustworthy.
2. **Reliability and Reproducibility:** When experiments are well-designed, it's easier for other scientists to repeat them and get similar results. This helps confirm findings and build trust in scientific discoveries.
3. **Efficiency:** Proper experimental planning ensures we use our resources wisely by figuring out the best sample size, conditions, and methods. This cuts down on waste and makes the research process more efficient.
4. **Generalizability:** By carefully selecting participants and designing study conditions representative of real-world scenarios, experimental design enhances

the external validity of research findings, facilitating their applicability to broader populations or contexts.

5. **Control of Bias:** Good experimental design helps us reduce bias and factors that could skew results, making sure our conclusions are accurate and reliable.

34. What are the key elements to consider when designing an experiment?

When designing an experiment, several key elements need to be carefully considered to ensure its success:

1. **Research Question or Hypothesis:** Clearly define the question or hypothesis that the experiment aims to address. This provides a focused direction for the study.
2. **Variables:** Identify the independent variable and dependent variable. Additionally, consider any potential confounding variables that could influence the results and how they will be controlled.
3. **Experimental Design:** Choose a suitable experimental design (such as between-subjects, within-subjects, or factorial design) depending on the research inquiry and practical constraints.
4. **Sampling:** Identify the target population and choose a sample that is representative. Take into account variables like sample size, sampling technique (such as random sampling or stratified sampling), and sampling frame.
5. **Randomization:** Implement randomization procedures to assign participants to different experimental conditions. Random assignment helps minimise bias and ensures that the groups are comparable.
6. **Control Group:** Include a control group that does not receive the experimental treatment or receives a placebo. This allows for comparison and helps isolate the effects of the independent variable.
7. **Manipulation:** Clearly state how the independent variable will be modified or applied in the experiment. Maintain uniformity and standardisation in the manipulation method.
8. **Measurement:** Choose the relevant measures and instruments to examine the dependent variable(s). Consider the reliability, validity, and sensitivity of the measures.
9. **Experimental Procedure:** Develop a detailed protocol outlining the steps involved in conducting the experiment. Standardise procedures to minimise variability and ensure consistency across participants and conditions.
10. **Data Analysis:** Determine the statistical techniques and analyses that will be used to analyse the data. Consider the assumptions underlying the chosen analyses and how they align with the research question.

11. **Ethical Considerations:** Ensure that the experiment follows ethical norms and principles, such as gaining informed consent from participants, maintaining their confidentiality, and minimising any risks or harm.
12. **Practical Constraints:** Consider practical constraints such as time, budget, equipment availability, and logistical considerations when designing the experiment.

By carefully considering these key elements, researchers can design experiments that are methodologically sound, ethically conducted, and capable of producing meaningful and valid results.

35. What are observational and experimental data in statistics?

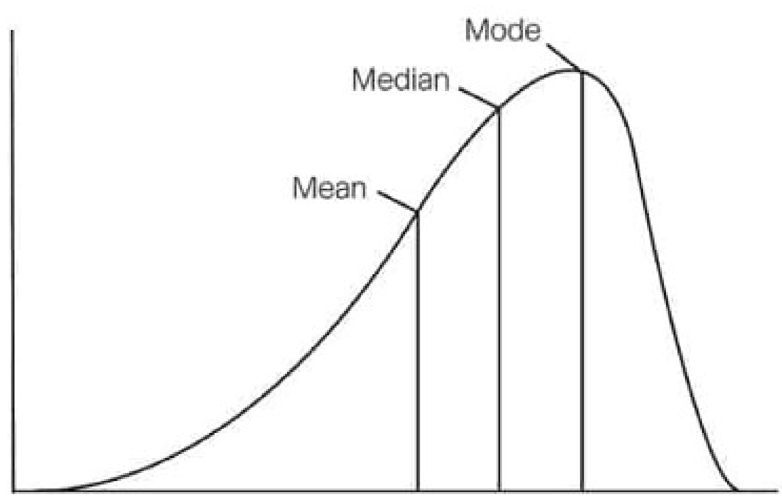
The differences can be stated as below:

Observational Data	Experimental Data
Observational data is collected through observation. This means that anything that can be heard or seen is collected.	It’s collected by doing experiments through the scientific method with a prescribed methodology.
Data collected by observing and recording events or behaviours.	Data collected by actively manipulating variables and observing the effects on other variables.
No intervention or manipulation of variables by the researcher.	Researchers deliberately introduce interventions or treatments to manipulate variables.
Limited ability to manage or influence other factors that might affect the observed outcomes.	Researchers can exert greater influence over potential variables that might distort results through the implementation of control groups and randomization methods.
It's harder to be sure about what causes what (Causality) because we can't control all the variables.	It's easier to figure out what causes what (Causality) because we can manipulate variables and keep other factors in check.
Example: Observing the relationship between smoking and lung cancer by comparing groups of smokers and non-smokers.	**Example:** Evaluating how well a new drug works by randomly assigning people to either take the drug or a placebo and then seeing what happens.

36. What is the left-skewed distribution and the right-skewed distribution?

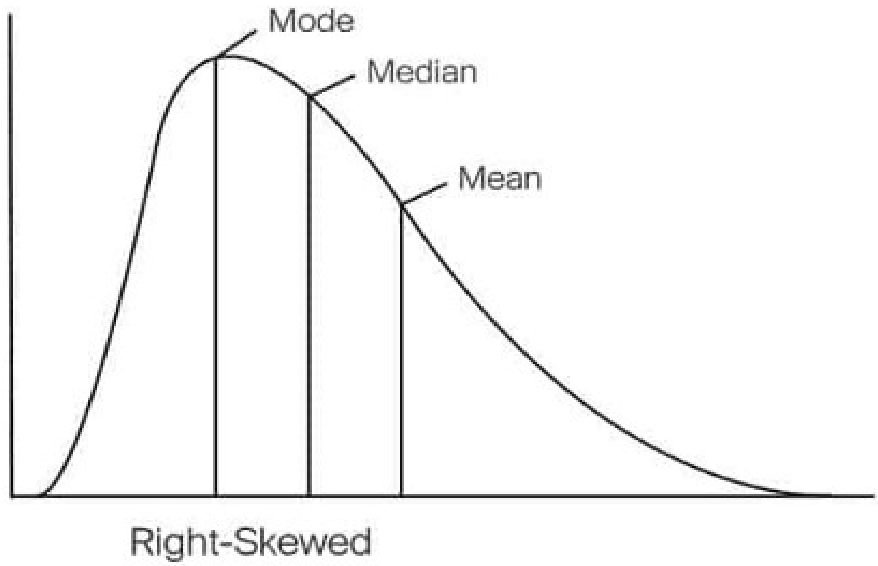
Left-skewed and right-skewed distributions, also known as negatively skewed and positively skewed distributions respectively, refer to the shape of the distribution of data points.

Left-skewed (Negatively Skewed) Distribution: In a left-skewed distribution, the left tail of the distribution is longer or stretches out farther than the right tail. Most data points are concentrated on the right side of the distribution, with fewer data points on the left side.



Left-Skewed (Negative Skewness)

Right-skewed (Positively Skewed) Distribution: In a right-skewed distribution, the right tail of the distribution is longer or stretches out farther than the left tail. The bulk of the data points are clustered on the left side of the distribution, with fewer data points on the right side.



37. What is Bessel's correction?

Bessel's correction involves adjusting formulas like those for sample variance and sample standard deviation by using n - 1 instead of n. This tweak compensates for the tendency of these sample statistics to underestimate the actual parameters in the population. So, it's like a fix to make sure our calculations better reflect reality, especially when working with smaller samples.

Bessel's Correction in Variance and Standard Deviation Calculations-

We usually use the following formula to find the sample variance:

$$S^2 = \frac{\sum(X-\overline{X})^2}{N-1}$$

The formula for the sample standard deviation is:

$$s = \sqrt{\frac{(x-\overline{x})^2}{n-1}}$$

Both of the formulas have n-1 in the denominator, here n is the sample size.

When we deal with the entire population, the calculations for parameters like population variance or standard deviation are correct because we have the entirety of the data. But when we take just a sample of the dataset, the answers won't be as accurate because we are only working with a portion of the whole set.

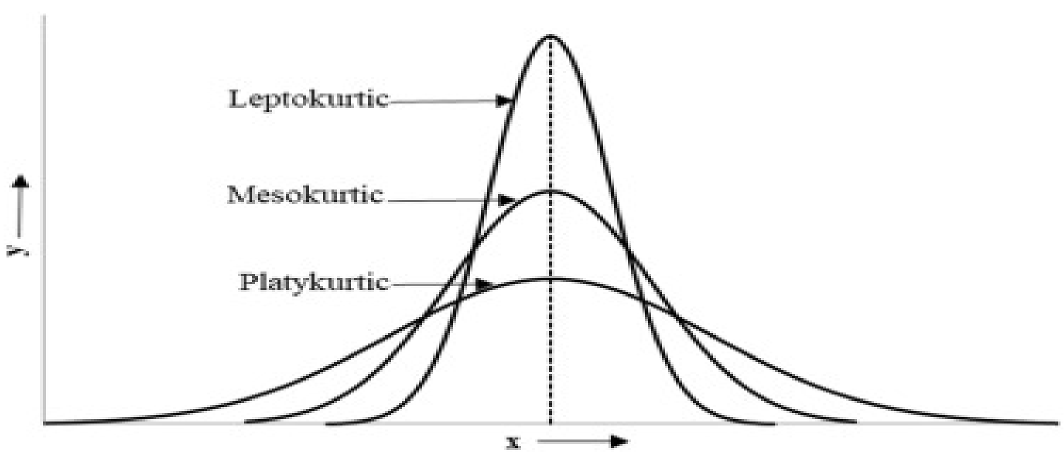
Sample variance and standard deviation are based on the sample mean (x̄), not the population mean (μ). This affects how we calculate sums of squares in the formulas, as any value in the sample will be closer to x̄ than to μ.

Generally, when we perform a calculation with 'n' in the denominator, the results tend to be higher compared to when we use 'n-1'. By subtracting 1 from the sample size, we are making a significant adjustment for the smaller sum of squares, especially when the sample size is not large.

When to use ?
Bessel's correction has a significant impact on small sample sizes. If your sample is small, it's unlikely to provide an accurate estimate of the population mean. For large sample sizes, Bessel's correction is not necessary because sample statistics may closely match population parameters. We can use Bessel's correction only with large samples when we need to approximate the population mean. Otherwise, we can omit the correction if only the sample mean is needed.

38. What is kurtosis?

Kurtosis measures the extent of the tails in a distribution, indicating whether extreme values are present. It assesses the frequency of outliers in the distribution.



Types of Kurtosis:
The categories of kurtosis are established based on the excess kurtosis of a specific distribution. Excess kurtosis compares the kurtosis of a distribution to that of a normal distribution, where the kurtosis equals 3. The excess kurtosis is calculated using the provided formula.

$$Excess\ Kurtosis = Kurtosis - 3$$

1. **Mesokurtic**: Data exhibiting a mesokurtic distribution displays an excess kurtosis close to zero or zero. This suggests that if the data conforms to a normal distribution, it aligns with a mesokurtic distribution.
2. **Leptokurtic:** The term "lepto" denotes narrowness or skinniness. Leptokurtic denotes a positive excess kurtosis (k > 0). A leptokurtic distribution demonstrates heavy tails on both sides, indicating the presence of significant outliers.
3. **Platykurtic:** The prefix "platy" originates from the Greek word "platus" meaning flat. A platykurtic distribution exhibits a negative excess kurtosis (k < 0). The kurtosis denotes a distribution with flattened tails, signifying the existence of minor outliers.

39. What is the probability of throwing two fair dice when the sum is 5 and 8?

Sum of 5: There are four possible combinations to get a sum of 5: (1, 4), (2, 3), (3, 2), and (4, 1).
Sum of 8: There are five possible combinations to get a sum of 8: (2, 6), (3, 5), (4, 4), (5, 3), and (6, 2).
The total number of possible outcomes when throwing two fair dice,

$$\text{Total outcomes} = \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$$

Therefore, the probability of throwing two fair dice and getting a sum of either 5 or 8 is the sum of these probabilities:

$$P(\text{sum}= 5 \text{ or } 8) = P(\text{sum}=5)+P(\text{sum}=8) = \frac{4}{36} + \frac{5}{36}$$

$$P(\text{sum}=5 \text{ or } 8) = \frac{9}{36} = \frac{1}{4}$$

40. What is the difference between Descriptive and Inferential Statistics?

The differences can be stated as below:

Descriptive Statistics	Inferential Statistics
This involves summarising and describing the attributes and qualities of a dataset.	This involves drawing conclusions, predictions, or generalisations about a population using data obtained from a sample.
Involves organising, presenting, and summarising data using measures such as mean, median, mode, standard deviation, and graphs.	This entails utilising data from a sample to make inferences or predictions about a broader population.
Examples comprise computing central	Examples include hypothesis testing,

tendency measures (mean, median, mode) and variability measures (range, standard deviation) for a dataset.	confidence interval estimation, and regression analysis.
Usually employed to depict and investigate the attributes of a dataset without extending conclusions beyond the available data.	Usually applied when researchers aim to extend findings from a sample to a broader population or when making predictions using data..

41. Imagine that Jeremy took part in an examination. The test has a mean score of 160, and it has a standard deviation of 15. If Jeremy's z-score is 1.20, what would be his score on the test?

The formula for z-score is,

$$z = \frac{x - \mu}{\sigma}$$

Where:

- **z** → z-score.
- **x** → individual data point.
- **μ** → mean of the dataset.
- **σ** → standard deviation of the dataset.

So,

$$x = z * \sigma + \mu$$

$$x = 1.20 * 15 + 160 = 178$$

Therefore, Jeremy's test score is **178**.

42. In an observation, there is a high correlation between the time a person sleeps and the amount of productive work he does. What can be inferred from this?

The significant correlation observed between the duration of sleep and the level of productivity implies a robust association between these two variables. Here are some potential inferences that can be made:

1. **Sleep Quality and Productivity:** Sleep quality is positively correlated to productivity. This shows that people who get sufficient sleep show higher levels of productivity. This is relevant because quality sleep may lead to better concentration and overall well-being, which can enhance productivity.
2. **Importance of Rest:** The correlation highlights the importance of good rest and sleep in maintaining productivity levels. Lack of sleep or poor sleep quality could potentially lead to decreased productivity due to fatigue, reduced focus, and lower energy levels.

3. **Workplace Policies:** Another conclusion drawn from sleep and productivity correlation can be implementation of policies or initiatives by workplaces to promote healthy sleep habits among employees. This may include flexible work schedules, promoting work-life balance, and providing resources for managing stress and improving sleep quality.
4. **Individual Well-being:** It suggests that individuals should prioritise getting enough sleep to optimise their productivity levels and overall performance in various aspects of life, including work.

43. What is the meaning of degrees of freedom (DF) in statistics?

Degrees of freedom represent the count of independent variables that can be estimated within a statistical analysis. They indicate the quantity of items that can be randomly chosen before constraints need to be applied to the data.

Example: Imagine you have a dataset of five positive integers, and the average of all five numbers is six. Say we already know four of the numbers are {3, 8, 5, and 4}, then the fifth number has to be 10 to hit that average. Since we could have picked any four numbers, the degrees of freedom here are four.

$$Degrees\ of\ freedom(df) = n - 1$$

Where, $n \rightarrow\ size\ of\ sample$

44. If there is a 30 percent probability that you will see a supercar in any 20-minute time interval, what is the probability that you see at least one supercar in the period of an hour (60 minutes)?

To determine the likelihood of encountering at least one supercar within a 60-minute timeframe, we can apply the complement rule. This rule asserts that the probability of an event happening is equivalent to 1 minus the probability of the event not happening.

Given that the probability of spotting a supercar within a 20-minute interval is 0.30, the probability of not spotting a supercar within 20 minutes is:

$P(not\ seeing\ a\ supercar) = 1 - P(seeing\ a\ supercar)$
$P(not\ seeing\ a\ supercar) = 1 - 0.3 = 0.7$

The probability of not seeing any supercar within a 60-minute period can be calculated as:

$P(not\ seeing\ a\ supercar\ in\ 60\ min) = 0.70^3 = 0.343$

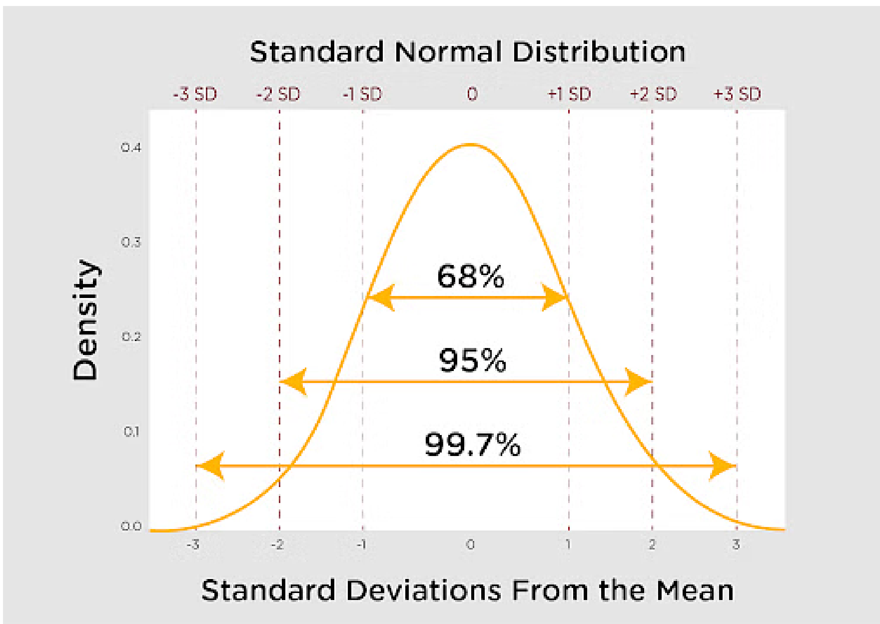
Therefore, the probability of encountering at least one supercar within a 60-minute period can be determined as:

$$P(seeing\ at\ least\ one\ supercar\ in\ 60\ min) = 1 - P(not\ seeing\ a\ supercar\ in\ 60\ min)$$
$$= 1 - 0.343 = 0.657$$

Therefore, the likelihood of observing one supercar within a 60-minute timeframe is **0.657**.

45. What is the empirical rule in Statistics?

The empirical rule, sometimes called the three-sigma rule or the 68-95-99.7 rule, suggests that almost all data points within a normal distribution will fall within three standard deviations (σ) from the mean (μ). This can be visualised in the following diagram:



- About 68% of the data lies within one standard deviation (σ) of the mean (μ).
- Approximately 95% of the data falls within two standard deviations (σ) of the mean (μ).
- Nearly 99.7% of the data is within three standard deviations (σ) of the mean (μ).

46. What is a Chi-Square test?

The chi-square test is a statistical method utilised to determine the significance of correlation between two categorical variables. As a non-parametric test, it doesn't rely on assumptions about the distribution of the data. This test relies on comparing observed and expected frequencies within a contingency table. It's commonly used for feature selection purposes, aiming to discern if the relationship observed between two sampled categorical variables is representative of the true relationship in the population.

The Chi-Squared distribution falls under the category of continuous probability distributions. It is characterised as the summation of the squares of k independent standard random variables:

$$\chi_c^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where,

- c represents degree of freedom,
- O_{ij} indicates observed frequency, and
- E_{ij} is the expected frequency in the ij cell.

Following are the steps to perform Chi-square test:

1. Define hypothesis: A hypothesis is a statement that determines a relationship or difference between variables. In the context of the chi-square test, the null hypothesis suggests that there is no significant correlation between the two categorical variables, while the alternative hypothesis proposes that a significant correlation exists between them.
2. Create a contingency table and find the expected values using the expression:

   $$E_{ij} = \frac{R_i \times C_j}{N}$$

   Where:

   - R_j represents the total of rows
   - C_j represents the total of columns.
3. Compute the Chi-square statistic and determine the degrees of freedom using the following formula:

   $$dof = (m - 1) \times (n - 1)$$

Here, m and n represent the number of categories in each of the categorical variables.

4. Accept or reject the null hypothesis: To decide whether to accept or reject the null hypothesis, compare the computed chi-square statistic to the critical value from the chi-square distribution table at the chosen significance level (e.g., 0.05).
   - Reject the null hypothesis if the chi-square value exceeds the critical value, demonstrating a significant correlation between the variables.
   - Do not reject the null hypothesis if the chi-square value is less than or equal to the critical value, suggesting no meaningful relationship between the variables.

47. What is a t-test?

The t-test is a statistical method employed to determine if there are significant differences between the means of two groups and their associations. It analyzes sample data taken from these groups to draw conclusions about the entire populations they represent.

Here are the steps to conduct a t-test:

1. **Formulate hypotheses:** Establish the null hypothesis stating that there is no significant difference between the means of the two groups, and the alternative hypothesis suggesting that a significant difference exists.

2. **Compute the t-value:** Calculate the t-value, which measures the difference between the means of the two groups relative to the variance within the groups.

3. **Determine degrees of freedom:** Determine the degrees of freedom based on the sample sizes of the two groups, which are used in the calculation of the t-value.

4. **Calculate the p-value**: Evaluate the probability, or p-value, which represents the likelihood of obtaining a t-value as extreme as or more extreme than the one observed, assuming the null hypothesis is true.

5. **Interpret the results:** If the p-value is less than the chosen significance level (e.g., 0.05), reject the null hypothesis, concluding that there is a significant difference between the means of the two groups. If the p-value is greater than 0.05, fail to reject the null hypothesis, indicating insufficient evidence to conclude a significant difference.

48. What is the ANOVA test?

ANOVA, or analysis of variance, is a statistical method that examines whether the means of three or more groups are significantly different from each other. It compares the group means to assess the influence of different factors. The null hypothesis, which assumes that all group means are equal, is tested against the alternative hypothesis, which proposes that at least one group mean differs from the others.

Here are the prerequisites for using ANOVA:

1. The dependent variable should have an approximately normal distribution within each group, especially crucial for smaller sample sizes.

2. Samples must be randomly selected and independent of each other.

3. All groups should have equal standard deviations.

4. Each data point should be exclusively associated with one group, without any overlap or data sharing between groups.

There are two primary types of ANOVA:

- **One-way ANOVA:** This form is employed when there's only one independent variable with more than two levels or groups. It determines if there are significant differences among the means of these groups.
- **Two-way ANOVA:** Building upon the one-way ANOVA, this type involves two independent variables. It allows for the investigation of each variable's main effects and their interaction. The interaction effect examines whether the influence of one variable on the dependent variable varies based on the level of the other variable.

49. How is hypothesis testing utilised in A/B testing for marketing campaigns?

A/B testing is an application of hypothesis testing that involves comparing two versions of a product or feature to determine which one performs better.

Let's imagine an online retailer who wants to boost his sales:
Objective: Increase sales on the platform.
Metric to Measure: Revenue increase. This could be other relevant metrics like number of new customers, user engagement, etc.

#Step 1: Identify a Change

Suppose an online store aims to enhance sales by improving its website's "Add to Cart" button. The current design features a green button labelled "Add to Cart".

#Step 2: Create Two Versions

- *Version A (Null): Retains the current website design with the existing "Add to Cart" button.*
- *Version B (Alternative): Introduces a redesigned website with a more prominent "Add to Cart" button, now displayed in red instead of green.*

#Step 3: Split Your Audience

The company randomly divides its user base into two equal groups:

- 50% of users view Version A.
- 50% of users view Version B.

#Step 4: Run the Test

The test is conducted over a specified duration, such as 30 days. During this period, data on user engagement and revenue metrics are collected for both groups.

#Step 5: Analyse the Results

Following the testing period, the company evaluates the collected data:
Did Version B exhibit higher revenue generation compared to Version A? Was there an increase in the number of customers making purchases? Did the average spending per customer rise?

#Step 6: Make a Decision

Based on the data analysis, the company determines its course of action:

- If Version B demonstrates superior performance in desired metrics, the new "Add to Cart" button design is implemented for all users.
- If there is no significant disparity or if Version A proves more effective, the company opts to retain the original design and considers alternative strategies.

50. What is the difference between one-tailed and two tailed t-tests?

The differences can be stated as below:

One-tailed t-test	Two-tailed t-test
It indicates a specific direction of change, either an increase or decrease, in the variable being assessed.	It is non directional, does not specify whether the difference will be positive or negative.

One critical region (either on the right or left side of the distribution, depending on the direction specified in the hypothesis).	Two critical regions (both tails of the distribution).
Used when there's a specific expectation or prediction about the outcome being tested.	Used when there's no specific expectation or when the researcher wants to detect any kind of difference.
Hypothesis Statement Example: "The sample mean is significantly greater than the population mean." Or "The sample mean is significantly less than the population mean."	Hypothesis Statement Example: "The implementation of a new teaching method leads to a statistically significant improvement in student test scores compared to the traditional teaching approach."
Example Scenario: Testing whether a new drug increases performance on a task	Example Scenario: Testing whether a teaching method has any effect on student test scores

51. What is an inlier?

An inlier is a data point that falls within the expected range of a statistical distribution but is erroneous. Due to their similarity to accurate data points, identifying and correcting inliers can cause big challenges.

REST QUESTIONS-

21. What is the central limit theorem, and how is it used?

25. How do you use the central limit theorem to approximate a discrete probability distribution?

27. How do you test the goodness of fit of a discrete probability distribution?

28. What is a joint probability distribution?

29. How do you calculate the joint probability distribution?

30. What is the difference between a joint probability distribution and a marginal probability distribution?

31. What is the covariance of a joint probability distribution?

32. How do you determine if two random variables are independent based on their joint probability distribution?

33. What is the relationship between the correlation coefficient and the covariance of a joint probability distribution?

42. How is a confidence interval defined in statistics?

43. What does the confidence level represent in a confidence interval?

44. What is hypothesis testing in statistics?

45. What is the purpose of a null hypothesis in hypothesis testing?

46. What is the difference between a one-tailed and a two-tailed test?

49. How can sample size determination affect experiment design?

50. What are some strategies to mitigate potential sources of bias in experiment design?

51. What is the geometric interpretation of the dot product?

52. What is the geometric interpretation of the cross-product?

53. How are optimization algorithms with calculus used in training deep learning models?

55. How are confidence tests and hypothesis tests similar? How are they different?

66. What is the relationship between sample size and power in hypothesis testing?

67. Can you perform hypothesis testing with non-parametric methods?

68. What factors affect the width of a confidence interval?

69. How does increasing the confidence level affect the width of a confidence interval?

70. Can a confidence interval be used to make a definitive statement about a specific individual in the population?

71. How does sample size influence the width of a confidence interval?

72. What is the relationship between the margin of error and confidence interval?

73. Can two confidence intervals with different widths have the same confidence level?

74. What is a Sampling Error and how can it be reduced?