End Course Summative Assignment

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

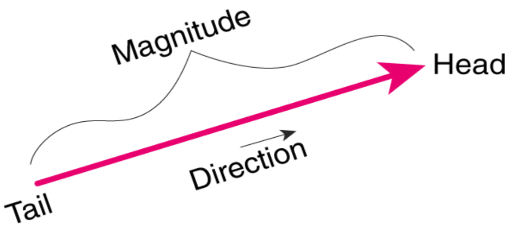
Imagine you are a dedicated student aspiring to excel in job interviews. Your task is to write the solutions for any 50 interview questions out of 80 total questions presented to you. Additionally, create an engaging video where you thoroughly explain the answers to any five of these questions.

Github Link: https://github.com/Dharmeshgadhiya161/Capstone-Project--Applied-statistics-interview-grind

Google Colab: *https://colab.research.google.com/drive/1ZPCPczMQFnP871wI70lQ6GHWQuI2FiWE?usp=sharing*

Capstone Project- Applied statistics interview grind

1) What is a vector in mathematics?



A vector is **a quantity that has both magnitude (length) and direction.**

The represented as a line with an arrow where the **tail** signifies the **starting point (initial point)** and the head represents the **ending point** (terminal point) indicating the **direction of the vector.**

Example: when you travel 16 kilometers south, your journey can be represented as a vector quantity.

2) How is a vector different from a scalar?

- Vector has **both magnitude and direction.**
- Scalar has **only magnitude** and no direction.
- Vector it is **multidimensional**.
- Scalar only **one dimensional**.
- Vector **cannot divide another vector.**
- Scalar quantity can **divide another scalar**.
- Scalars are simple numerical values that can describe quantities like mass, temperature, and speed.
- Vectors are represented by an **ordered set of values or coordinates**.
- Example:
- Scalar: Temperature (e.g., 30°C), Mass (e.g., 5 kg), Speed (e.g., 60 km/h)
- Vector: Velocity (e.g., 60 km/h east), Force (e.g., 10 N downward)

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

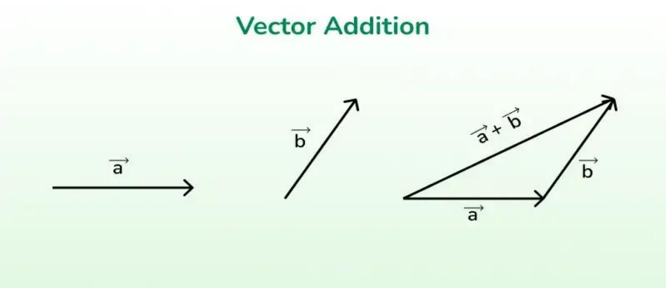
Key Vector Operations:

- Addition of Two Vectors
- Subtraction of Two Vectors
- Multiplication of Vector with Scala
- Product of Two Vector.
  1. Dot Product
  2. Cross-Product
- Vector projection.

These operations allow for various mathematical manipulations and calculations involving vectors.

Addition of Two Vectors.

Vector addition is a **sum of two or more vectors**. It combines the magnitudes and directions of the vectors to produce a single resultant vector.



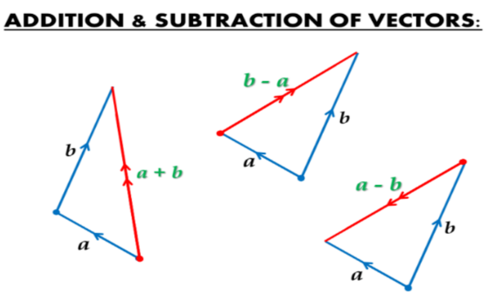
The **Triangle Law of Vector** Addition means that to add two vectors, place the second vector's starting point at the end of the first vector. Then, draw a new vector from the start of the first vector to the end of the second. the new vector represents the total effect of both vectors combined.

Example: u = (2,3), v = (3, −2)

The sum of these vectors is: u + v = (2+3,3+(−2)) = (5,1)

Subtraction of Two Vectors:

The same process as vector addition but with **the reversed vector**.



Vector subtraction the rule: **u−v = u+ (−v)**

Negative Vector (-v): The negative of a vector **v** is a vector of the same magnitude but in the opposite direction. **Addition of Vectors:** Instead of directly subtracting **v**, we add **-v** to **u**.

Example: u= (2,3), v= (3, −2)

u−v = (2,3) − (3, −2) = (2−3,3−(−2)) = (−1,5)

Multiplication of Vector with Scala:

A vector can be multiplied by a scalar (a real number), which scales its magnitude.

Example: 3 × (2, -1) = (6, -3)

Dot product of two vectors:

The dot product of two vectors is a scalar quantity that is obtained by multiplying the corresponding components of the vectors and then summing them up.

Mathematically, if we have two vectors A = (a1, a2, a3) and B = (b1, b2, b3), their

dot product **A · B is given by A · B = a1 * b1 + a2 * b2 + a3 * b3.**

Example: A= (3,4), B= (2, −1)

A·B = (3 × 2) + (4 ×−1) = 6−4 = 2

Cross product of two vectors:

The cross product of two vectors is a vector quantity that is orthogonal (perpendicular) to both of the original vectors.

It is calculated by taking the determinant of a special matrix formed by the components of the two vectors.

The cross product of two vectors A = (a1, a2, a3) and B = (b1, b2, b3)

is given by **A × B = (a2 * b3 - a3 * b2, a3 * b1 - a1 * b3, a1 * b2 - a2 * b1).**

Example: (1,2,3) × (4,5,6) = A×B = ((2×6−3×5), (3×4−1×6), (1×5−2×4))

= (12−15,12−6,5−8) = (−3 ,6, −3)

Vector projection

The **vector projection** of a vector **A** onto another vector **B** is a vector that represents the component of **A** in the direction of **B**. It is the **shadow** or **parallel component** of **A** along **B** when projected onto it.

$$\mathrm{Proj}_B A = \frac{A \cdot B}{|B|^2} B$$

- A·B is the **dot product** of A and B,
- |B|^2 is the **square of the magnitude** of B,
- The result is a vector in the direction of B.
- Example: A= (3,4) B= (1,2)

  A·B = (3×1) + (4×2) = 3 + 8 = 11
  |B|2 = (1+2)2 = 1+4 = 5

ProjBA= 5/11* B = 5/11*(1,2)
= (5/11,5/22)

4) How can vectors be multiplied by a scalar?

When a vector is multiplied by a scalar, each component of the vector is multiplied by the scalar value, **the vector's magnitude changes, but its direction remains the same.**

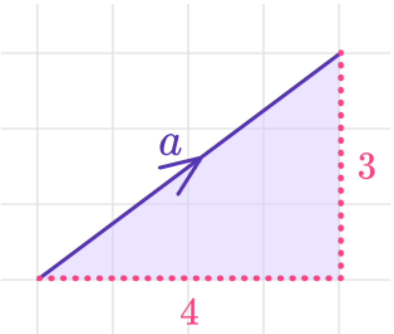
Example: if we have a vector **A = (a1, a2, a3)** and a **scalar c**

the scalar multiplication of the vector is **c * A = (c * a1, c * a2, c * a3).**

5) What is the magnitude of a vector?

The **magnitude of a vector represents its length or size in space**. It is **calculated using the Pythagorean theorem** (or the distance formula) in different dimensions

- **2D vector= |v| = $\sqrt{x^2 + y^2}$, 3D vector, and n-dimensional vector**



Example: v = (4,3) # Vector v has two components (x, y).
x=4 (horizontal component) y=4(vertical component)

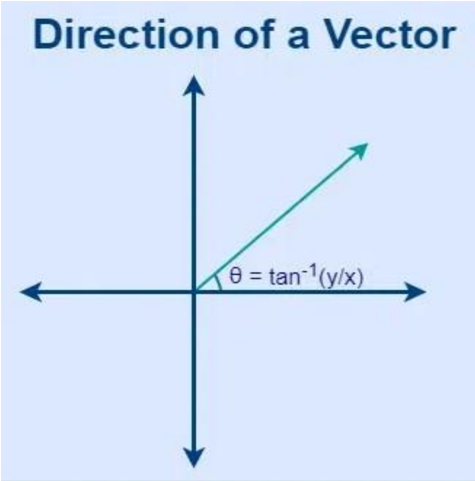
The magnitude of vector v a is: **$\sqrt{4^2 + 3^2} = \sqrt{16+9} = \sqrt{25} = 5$**
The length of the **vector is 5**.

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

The direction of a vector can be determined by finding the angle it makes with a reference axis or another vector.

This can be done using **trigonometric functions** such as sine, cosine, or tangent.

Alternatively, the direction can be represented using unit vectors or by specifying the angles of rotation needed to align the vector with the coordinate axes.



Formula:

calculate the **direction angle** (θ) is: **θ = tan^-1 (y/x).**

Example: - Vector: (3, 4)

Calculating direction: θ = tan^-1 (4 / 3)

θ ≈ 53.1 degrees

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

- A square matrix has an **equal number of rows and columns**.
- A rectangualer matrix where the **number of rows does not equal the number of columns.**
- Squer matrix **n × n (e.g., 2×2, 3×3, etc.)**
- Rectangular matrix **m × n where m ≠ n (e.g., 2×3, 3×4, etc.)**
- Square matrix are widely used in linear algebra, determinants, and eigenvalues, while rectangular matrix are common in data representation (e.g.,spreadsheets, transformation matrix).

8) What is a basis in linear algebra?

- A basis of a vector space is defined as a set of vectors that are both linearly independent and span the entire vector space, meaning every vector in the space can be expressed as a unique linear combination of the basis vectors.
- Key points about a basis:
- **Linear independence**: No vector in the basis can be written as a linear combination of the other vectors in the set.
- **Spanning**: The set of all possible linear combinations of the basis vectors includes every vector in the vector space.
- **Uniqueness**: Every vector in the space can be represented uniquely as a linear combination of the basis vectors.

9) What is a linear transformation in linear algebra?

A linear transformation (or linear map) in linear algebra is a function between **two vector spaces that preserves the operations of vector addition and scalar multiplication.**

- A linear transformation takes a vector and transforms it into another vector.
- It preserves the operations of vector addition and scalar multiplication.
- It respects the underlying structure of each vector space.

Additivity: $T(u + v) = T(u)+T(v)$ (The transformation of a sum is the sum of the transformations.)

Homogeneity (Scalar Multiplication Preservation):

$T(cv) = cT(v)$ (The transformation of a scaled vector is the scaled transformation.)

10) What is an eigenvector in linear algebra?

An eigenvector of a square matrix A is a nonzero vector v that remains in the same direction (up to a scalar multiple) when transformed by A.

Mathematically, it satisfies the equation: $\mathbf{Av=\lambda v}$

- A is an $n \times n$ matrix,
- v is the eigenvector (a nonzero vector),
- λ is the eigenvalue (a scalar associated with v).

Principal Component Analysis (PCA):

- PCA uses eigenvectors (principal components) of the covariance matrix to reduce dimensionality while preserving variance.
- The eigenvectors with the largest eigenvalues capture the most significant patterns in the data.

11) What is the gradient in machine learning?

The gradient is the vector of partial derivatives of a function with respect to its input variables. It is used in machine learning for optimizing models by adjusting their parameters in the direction of the steepest ascent.

Gradient Descent Formula: $\theta=\theta-\alpha\nabla J(\theta)$

- θ = model parameters (weights, biases)
- α = learning rate (step size)
- $\nabla J(\theta)$ = gradient of the loss function

12) What is backpropagation in machine learning?

Backpropagation is an algorithm used in machine learning to calculate the gradient of a loss function with respect to the parameters of a neural network. It is used to optimize the network by adjusting the weights and biases

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

In calculus, the derivative represents the rate of change of a function with respect to its input variable. Geometrically, it corresponds to the slope of the tangent line to the function's graph at a particular point. The derivative provides information about how the function is changing locally, whether it is increasing or decreasing, and the steepness of the curve.

14) How are partial derivatives used in machine learning?

Partial derivatives are used in machine learning, particularly in the field of deep learning, for optimizing neural networks. By calculating the partial derivatives of the loss function with respect to the network's weights, a process known as backpropagation, the network's parameters can be updated to minimize the loss. Partial derivatives provide information about how each weight affects the overall loss, enabling the network to learn and improve its predictions.

15) What is probability theory?

Probability theory is the branch of mathematics that deals with studying random events and the likelihood of their occurrence. It provides a framework for understanding uncertainty and making predictions based on statistical analysis.

Probability is quantified as a number between **0** and **1 .**

Formula :- ***P(E) = Number of Favorable Outcomes /Total Number of Outcomes***

Example :- coin-tossing case.

- In tossing a coin, there are two outcomes: Head or Tail.
- The Probability of occurrence of Head on tossing a coin is P(H) = 1/2
- The Probability of the occurrence of a Tail on tossing a coin is P(T) = 1/2

16) What are the primary components of probability theory?

The primary components of probability theory include probability **axioms and rules**, **conditional probability** and **Bayes theorem**, **random variables**, and the **law of large numbers and central limit theorem.**

Probability axioms and rules:

- The probability of an event is greater than or equal to zero P(E) >= 0
- P(Ω) = 1 (The probability of the sample space is always 1).
- If events A and B are disjoint, P(A∪B) = P(A)+P(B).

Conditional probability:

The probability of event A happening given that B has already occurred.

Formula**: P (A | B) = P(A∩B)/P(B)**

- P(A|B) The probability of event A occurring given that B has already occurred (Conditional Probability).
- P(A∩B) The probability of both A and B occurring together (Joint Probability).
- P(B) The probability of event B occurring independently (Marginal Probability).

Bayes theorem: is a mathematical formula that calculates the probability of an event based on prior knowledge or observations.

Formula: **P(A|B)=P(B|A)P(A)/P(B)**

- P(A|B) - Probability of event A happening given that B is true (posterior probability).
- P(B|A) - Probability of event B happening given that A is true (likelihood).
- P(A) - Independent probability of event A (prior probability).
- P(B) - Independent probability of event B (normalizing constant).

Random variables:

A **random variable** is a function that assigns numerical values to the outcomes of a random experiment.

Types of Random Variables:

- Discrete Random Variable
- Continuous Random Variable

Law of large numbers.

The law of large numbers states that as the sample size of a random variable increases, the sample mean converges to the population mean. It is a fundamental concept in probability theory used to make predictions based on large datasets.

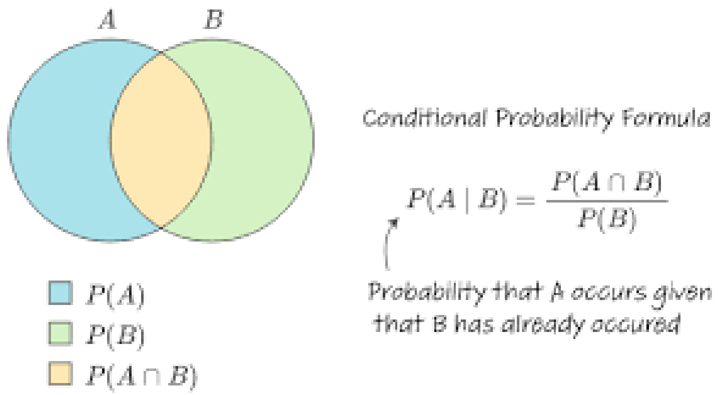
Central limit theorem.

The central limit theorem states that as the sample size of a random variable increases, the distribution of the sample means approaches a normal distribution. It is used to make predictions about the mean of a population based on a sample.

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

Conditional probability is a measure of the probability of an event occurring, given that another event has already occurred. If the event of interest is A and event B has already occurred, the conditional probability of A given B is usually written as P(A/B).

Formula: **P(A|B) = P (A ∩ B) / P(B)**



- P (A ∩ B) represents the probability of both events A and B occurring simultaneously.
- P(B) represents the probability of event B occurring.

Example: Rolling a Die 🎲

Problem: What is the probability of rolling a 4, given that the number rolled is even?

- P(A): Probability of **rolling a 4** = **1/6** (as there is only one way to roll a 4 on a die).
- P(B): Probability of rolling an **even number** -> Possible outcomes: {2, 4, 6} So P(B)= 3/6 = **1/2**.
- P (A ∩ B): Probability of rolling both a 4 and an even number = 1/6 (since the only way to achieve both is by rolling a 4).
- P(A|B) = P (A ∩ B) / P(B)
- P(A|B) = (1/6) / (1/2) = 2/6 = 1/3
- P(A|B) = 1/3
- The probability of rolling a 4, given that the number rolled is even, is 1/3.

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

Bayes theorem is a mathematical formula that calculates the probability of an event based on prior knowledge or observations. It updates our beliefs about an event based on new information.

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

- P(A|B) - Probability of event A happening given that B is true (posterior probability).
- P(B|A) - Probability of event B happening given that A is true (likelihood).
- P(A) - Independent probability of event A (prior probability).
- P(B) - Independent probability of event B (normalizing constant).

Example: Medical Diagnosis

- Suppose a doctor wants to determine the probability that a patient has a **disease (A)** given that they tested **positive (B)** for it.
- P(A) = 0.01 - 1% of people have the disease.
- P (B | A) = 0.95 - If a person has the disease, they test positive 95% of the time.
- P(B) = 0.05 - 5% of people test positive overall.

P(A|B) = P(B|A) · P(A)/P(B)

P(A|B) = 0.95 × 0.01 / 0.05 = 0.0095 / 0.05= 0.19

The probability that a person **actually has the disease** given they tested positive is **19%**.

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

A **random variable** is a variable in statistics that **represents a numerical outcome from a random experiment**.

Example: Rolling a die, X can be 1-6

Regular variable, which has a **fixed value**.

Example: x=5 (fixed)

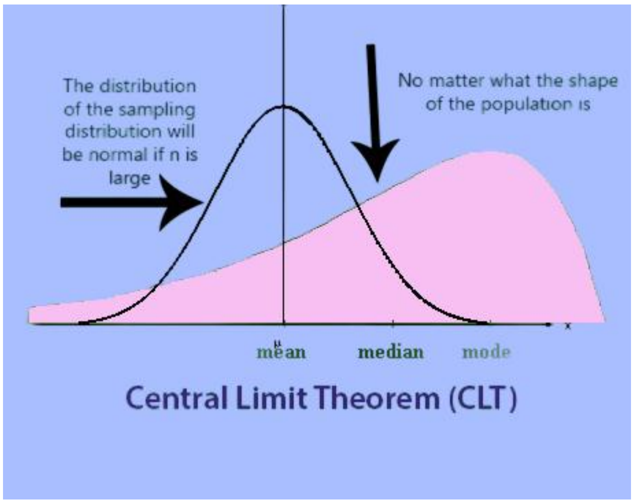
20) What is the law of large numbers, and how does it relate to probability theory?

The law of large numbers states that as the sample size of a random variable increases, the sample mean converges to the population mean. It is a fundamental concept in probability theory used to make predictions based on large datasets.

In simple terms, if you repeat an experiment **many times**, the results will stabilize around the expected probability.

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

The central limit theorem states that as the sample size of a random variable increases, the distribution of the sample means approaches a normal distribution. It is used to make predictions about the mean of a population based on a sample.



The Key for the CLT:

- A sufficiently large sample size can predict the characteristics of a population more accurately.
- Sample sizes equal to or greater than 30 are often considered sufficient for the CLT to hold.
- A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- CLT is useful in finance and investing when analyzing a large collection of securities to estimate portfolio distributions and traits for returns, risk, and correlation.

Some of the key problems that the CLT solves are:

- **Estimating Population Parameters:** The CLT helps us estimate population parameters, like the **mean** and **standard deviation**, by using a sample of the data. This is important because it allows us to make inferences about the entire population based on a smaller sample.
- **Hypothesis Testing:** The CLT is used to **test hypotheses** about population parameters, such as whether two means are statistically different or not.
- **Confidence Intervals:** The CLT helps us create confidence intervals, which are ranges of values that we are confident contain the true population parameter.
- **Model Fitting:** The CLT is also used in model fitting where it helps us estimate the parameters of a statistical model by using data.

The CLT has some limitations.

- **Sample Size:** The CLT assumes that the sample size is large enough for the normal distribution to emerging. The CLT may not be applicable if the sample size is too small.

- **Independence:** The CLT assumes that the samples are independent of each other. If the samples are not independent, the CLT may not be applicable.
- **Outliers:** The CLT is sensitive to outliers, which can skew the distribution and affect the accuracy of the results.
- **Non-Normal Distributions:** The CLT only applies to populations with a normal distribution. The CLT may not be applicable if the population is not normally distributed.

22) What is the difference between discrete and continuous probability distributions?

Discrete probability distributions have a **finite or countable number of possible values**

- Probability is assigned to specific values.
- The sum of all probabilities is 1.
- Usually represented by a probability mass function (PMF).

Examples

- Rolling a Die 🎲 : Outcomes: {1, 2, 3, 4, 5, 6}
- Flipping a Coin 🪙 : Outcomes: {Heads, Tails}
- Number of Sales per Day: {0, 1, 2, 3, …}

Continuous probability distributions have an **infinite number of possible values.**

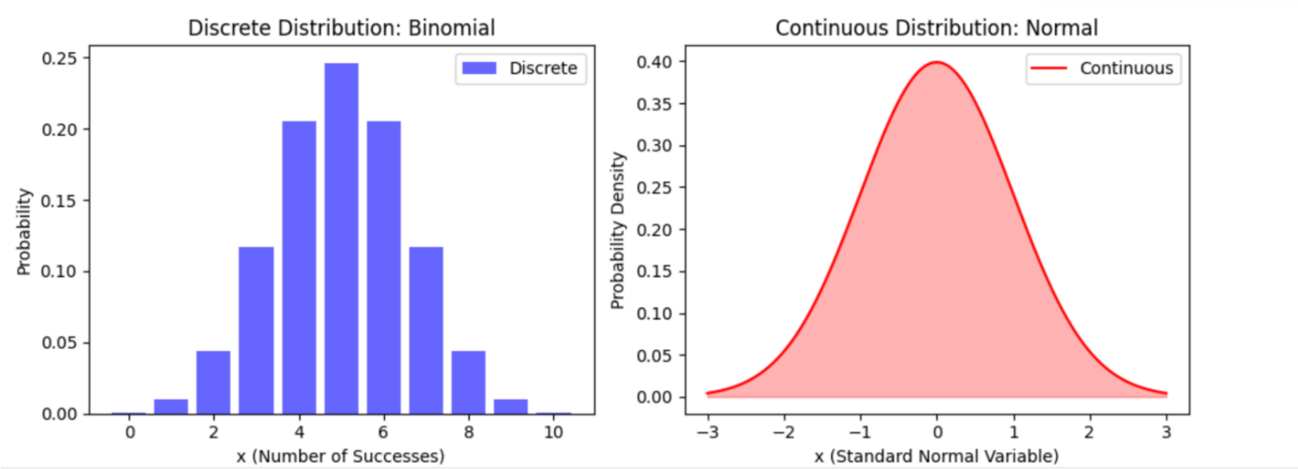
- Probability is assigned over an interval (not specific values).
- Probability is represented using a probability density function (PDF).
- The area under the curve represents probability.

Examples

- Height of People 📏 (e.g., 170.5 cm)
- Time Taken to Complete a Task ⏱ (e.g., 3.45 seconds)

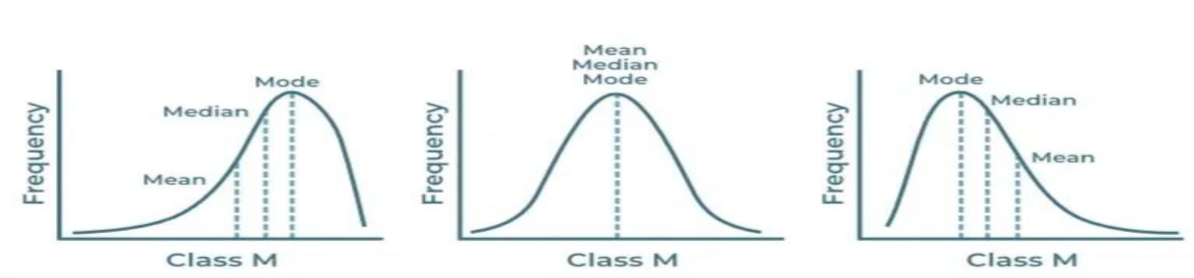
- Temperature 🌡 (e.g., 25.6°C)

Probability Distribution



23) What are some common measures of central tendency, and how are they calculated?

Common measures of central tendency include the **mean, median, and mode**.



Mean: The mean is calculated by **summing all values and dividing by the number of values**.

- **Formula:** Mean = (Sum of all values) / (Total number of values).
- **Example:** For the data set {1, 2, 3, 4, 5}, the mean is (1+2+3+4+5) / 5 = 3.

Median: The median is the **middle value when values are arranged in order**.

- **Odd number of values:** The median is the middle value when the data is sorted in ascending order.
- **Even number of values:** The median is the average of the two middle values when the data is sorted in ascending order.
- **Example:** For the sorted data set {1, 2, 3, 4, 5}, the median is 3. For the sorted data set {1, 2, 3, 4}, the median is (2+3)/2 = 2.5.

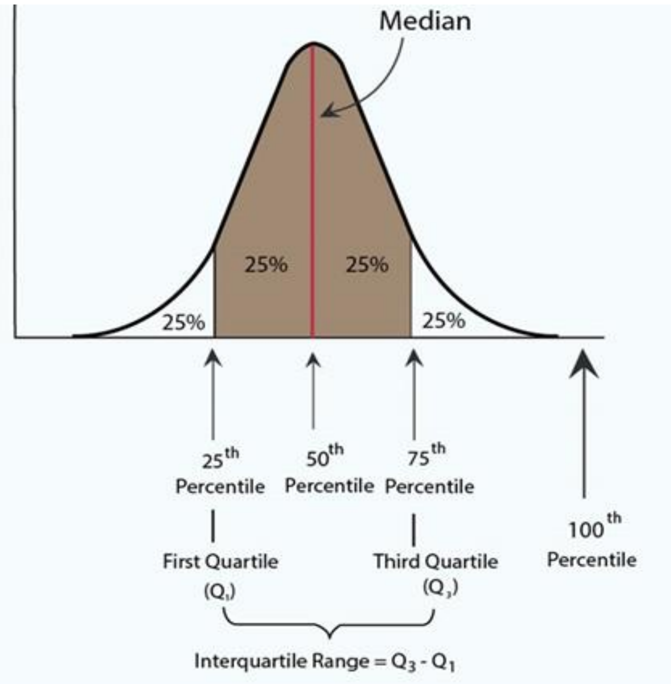
Mode: The **mode is the value that appears most frequently**.

- Identify the value that occurs most often in the data set.
- **Example:** For the data set {1, 2, 2, 3, 4, 4, 4, 5}, the mode is 4.

24) What is the purpose of using percentiles and quartiles in data summarization?

Percentiles and quartiles are statistical measures used to summarize and interpret data distribution. They help understand **distribution, central tendency**, **data spread, variability, and outliers, making them useful** in fields like machine learning, finance, and healthcare.

They provide a way to **summarize large datasets** and identify important trends.



Percentiles 📈 A **percentile** is a value below which a given percentage of data falls.

Purpose: Helps in ranking and comparison (e.g., test scores, income levels). Identifies **outliers** and extreme values. Used in **machine learning**, finance, and medical statistics.

Example: If a student scores in the **90th percentile** on an exam, it means they performed better than **90% of the students**.

Quartiles 📊 **:** Quartiles **divide data into four equal parts**:

- **Q1 (25th percentile)** → Lower quartile (25% of data falls below).
- **Q2 (50th percentile)** → Median (middle value of dataset).
- **Q3 (75th percentile)** → Upper quartile (75% of data falls below).

Purpose:

- Helps in understanding **data spread** and distribution.
- Used in **box plots** to visualize data variation.
- Helps detect **skewness** and outliers.

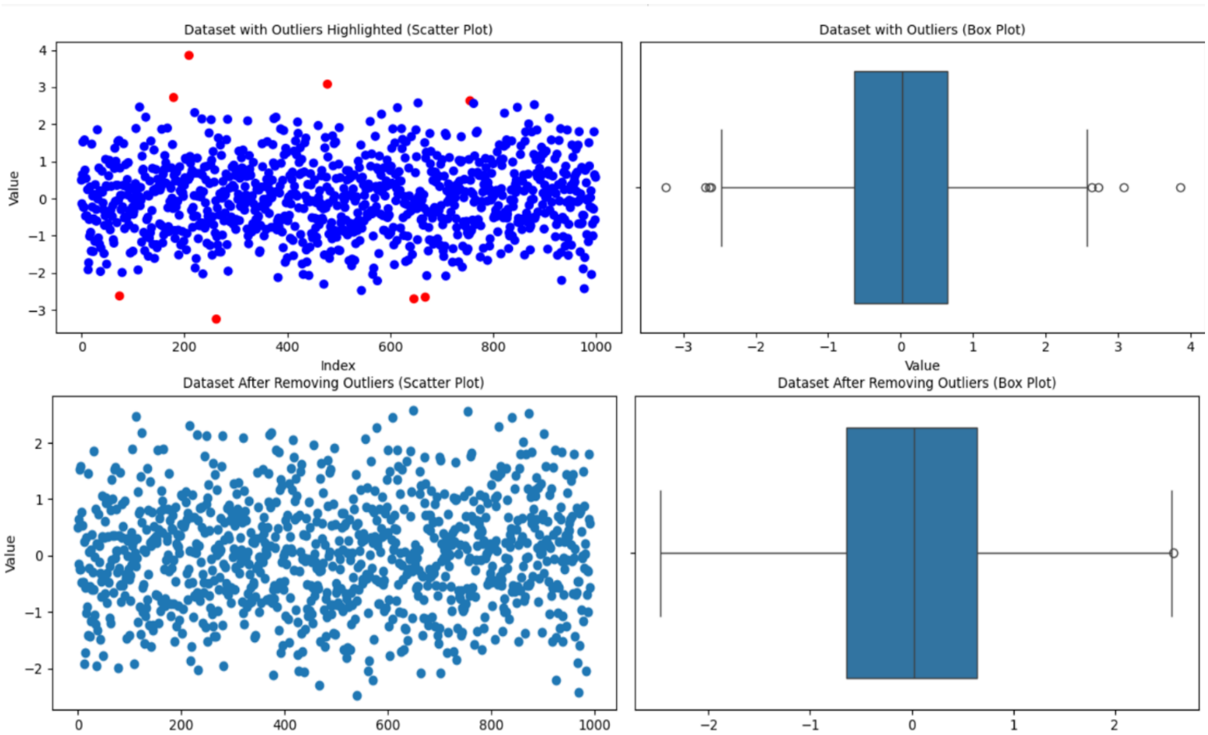
Example: If Q1 = 30, Q2 = 50, and Q3 = 70, then **50% of the data falls between 30 and 70**.

25) How do you detect and treat outliers in a dataset?

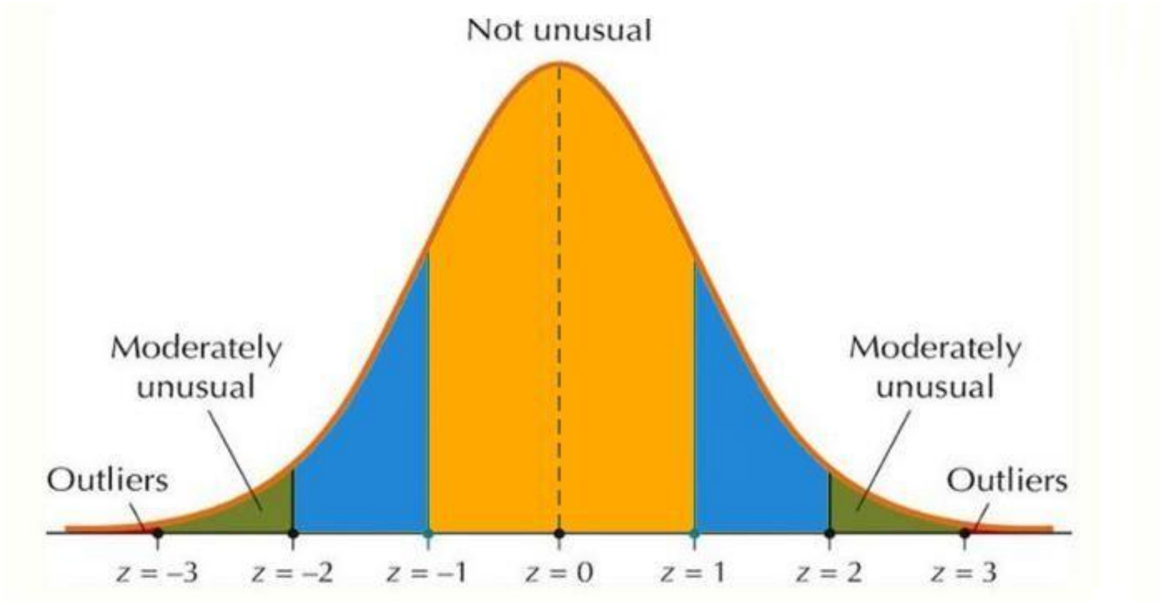
Outliers can be detected using various techniques, including **scatter plots, box plots, and z-scores.** Once identified, outliers can be treated using methods such as removal, transformation, or substitution.

Outlier Detection Techniques.

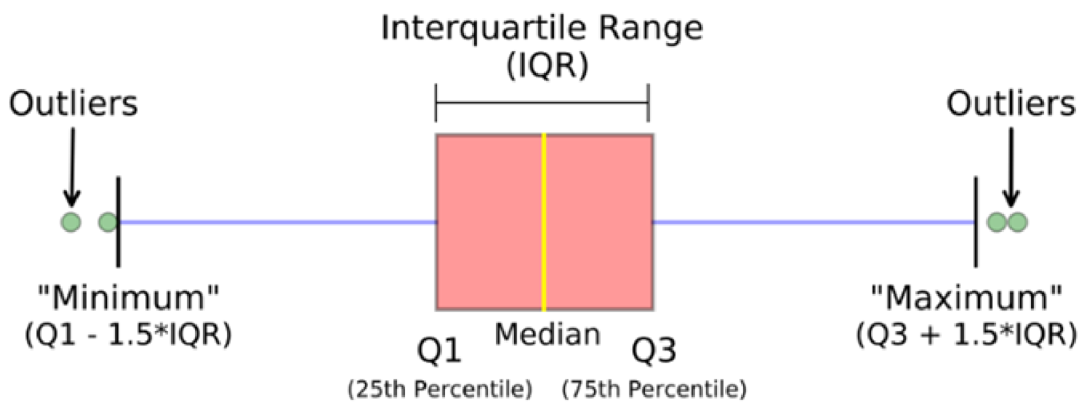
- **Visual:**
  - **Scatter Plots:** These plots help identify outliers by visually showing data points that deviate significantly from the general trend or cluster.
  - **Box Plots:** Box plots visually represent the distribution of data, with outliers shown as points beyond the whiskers (typically 1.5 times the interquartile range).



- **Statistical Methods:**
  - **Z-score:** The Z-score method for outlier detection uses a dataset's **standard deviation and its mean to identify data points** that are significantly different from the majority of the other data points.
  - Z-score notion. The Z-score for a value of **x in the dataset with a normal distribution with mean μ and standard deviation σ** is given by.
  - **$z = (x - \mu)/\sigma$**
  - Z-score takes the following values as shown below: The Z-score is equal to zero when x = .µ The Z-score is ± 1, ± 2, or ± 3, depending on whether x is ± 1, ± 2, or ± 3, respectively.



- **Interquartile Range (IQR) :** IQR outlier detection method involves calculating the first and third quartiles (Q1 and Q3) of a dataset and then identifying any data points that fall beyond the range of Q1 - 1.5 * IQR to Q3 + 1.5 * IQR, where IQR is the difference between Q3 and Q1. Data points that fall outside of this range are considered outliers.
- **IQR (Inter Quantile Range) = Q3– Q1**
- **Lower Bound Limit = Q1 – 1.5 x IQR**
- **Upper Bound Limit = Q3 + 1.5 x IQR**



26) How do you use the central limit theorem to approximate a discrete probability distribution?

The central limit theorem relies on **the concept of a sampling distribution**, which is the probability distribution of a statistic for a **large number of samples taken from a population.**

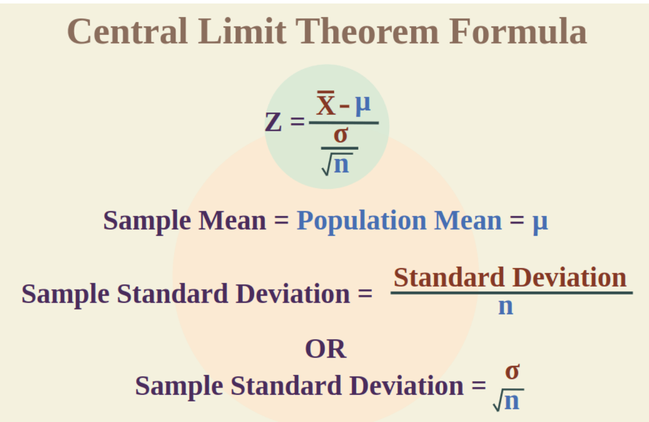
- **Identify the Discrete Distribution**: Determine the discrete probability distribution you want to approximate (e.g., binomial, Poisson).
- **Check for Large Sample Size**: Ensure that the sample size (n) is large enough for the CLT to apply effectively. A common rule of thumb **is n ≥ 30**, but this can vary depending on the skewness of the original distribution.
- **Calculate the Mean (μ) and Standard Deviation (σ) of the Distribution**: Find the mean (μ) and standard deviation (σ) of the original discrete distribution.
- **Approximate with a Normal Distribution**: Use the CLT to approximate the distribution of the sum (or average) of the random variables with a normal distribution.

Central limit theorem formula

Imagine you have a population where the data follows some random variable X and this population has:

- Mean μ the average of the population
- Standard deviation σ

let's say we take a sample of size n from this population and calculate its mean X¯ then the Z-Score is given below:



Conditions of the central limit theorem

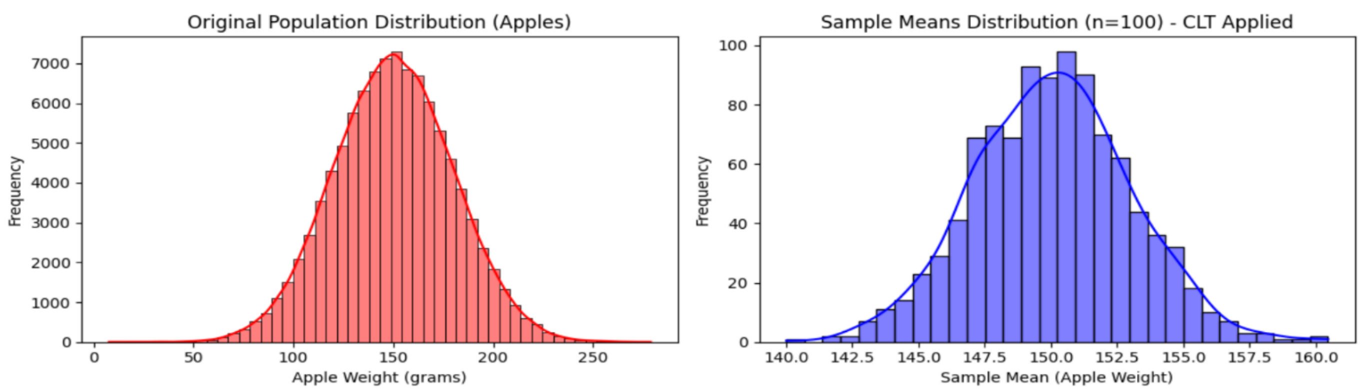
- The central limit theorem states that the sampling distribution of the mean will always follow a normal distribution under the following conditions:

- The sample size is sufficiently large. **This condition is usually met if the sample size is n ≥ 30.**
- The samples are independent and identically distributed (i.i.d.) random variables. This condition is usually met if the sampling is random.
- The population's distribution has finite variance. Central limit theorem doesn't apply to distributions with infinite variance, such as the Cauchy distribution. Most distributions have finite variance.

```python
# Step 1: Generate Normally Distributed Population (Apple Weights)
population_mean = 150  # Mean weight of apples
population_std = 30    # Standard deviation
population_size = 100000
population = np.random.normal(population_mean, population_std, population_size)
# Step 2: Take Multiple Random Samples & Compute Sample Means
sample_size = 100
num_samples = 1000
sample_means = [np.mean(np.random.choice(population, sample_size)) for _ in range(num_samples)]

# Step 3: Plot the Results
plt.figure(figsize=(12, 4))
# Plot the Population Distribution
plt.subplot(1, 2, 1)
sns.histplot(population, bins=50, kde=True, color='red')
plt.title("Original Population Distribution (Apples)")
plt.xlabel("Apple Weight (grams)")
plt.ylabel("Frequency")

# Plot the Sample Means Distribution
plt.subplot(1, 2, 2)
sns.histplot(sample_means, bins=30, kde=True, color='blue')
plt.title(f"Sample Means Distribution (n={sample_size}) - CLT Applied")
plt.xlabel("Sample Mean (Apple Weight)")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```



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

The chi-squared goodness-of-fit test is most common method for discrete distributions. It compares observed and expected frequencies to determine whether the data follows a hypothesized distribution.

28) What is a joint probability distribution?

A joint probability distribution describes the probabilities of two or more **random variables occurring simultaneously/ together**, using a joint probability mass function (PMF) for discrete variables (P (X= x, Y= y) = P (x, y)) and a joint probability density function (PDF) for continuous variables.

29) How do you calculate the joint probability distribution?

A **joint probability distribution** describes the probability of two random variables occurring **together**. The method of calculation depends on whether the variables are **discrete** or **continuous**.

Discrete: If the random variables are discrete (can only take on a finite or countable number of values), we use the **joint probability mass function (PMF)**

Discrete Random Variables (Using PMF):

- **f (x, y) = P (X = x, Y = y).**
- The joint PMF, denoted as f (x, y) for random variables X and Y, gives the probability of X taking the value x and Y taking the value y:

Example (Discrete): Given a table of joint probabilities for two discrete variables, X and Y:

X/Y	Y=0	Y=1	P(X) (Marginal)
0	0.1	0.2	0.3 (0.1 + 0.2)
1	0.3	0.4	0.7 (0.3 + 0.4)

To find **the marginal distribution of P(X)**, sum the probabilities across rows:

P(X=0) = P (X=0, Y=0) +P (X=0, Y=1) = 0.1 + 0.2= 0.3

P(X=1) =P (X=1, Y=0) +P (X=1, Y=1) = 0.3 + 0.4= 0.7

To find **marginal distribution of Y** P(Y), sum the probabilities across **columns**:

P(Y=0) = P (X=0, Y=0) + P (X=1, Y=0) = 0.1 + 0.3 = 0.4

P(Y=1) = P (X=0, Y=1) + P (X=1, Y=1) = 0.2 + 0.4 = 0.6

X/Y	Y=0	Y=1
0	0.1	0.2
1	0.3	0.4
P(Y)	0.4	0.6

the marginal distributions are:

- **P(X=0) = 0.3**, **P(X=1) = 0.7**
- **P(Y=0) = 0.4**, **P(Y=1) = 0.6**

Continuous: If the random variables are continuous (can take on any value within a range), the joint probability is represented by a joint **probability density function (PDF).**

- **Continuous Random Variables (Using PDF):**
    - The joint PDF, denoted as f (x, y), describes the probability density of X and Y at a particular point (x, y).
    - **P (a ≤ X ≤ b, c ≤ Y ≤ d) = ∫f (x, y) dx dy.**

Example 2 (Continuous):

If you have a joint PDF f (x, y) = 1/2 * x * y for 0 < x < 1 and 0 < y < 2, you can calculate the probability of the probability that X and Y are both between 0 and 1.

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

Joint probability distribution describes the probability of multiple events occurring together.

- **Formula**: P (X=x, Y=y) represents the probability of X taking value 'a' and Y taking value 'b' together.
- Discrete Random Variables (PMF) and Continuous Random Variables (PDF)
- **Example**: P (Rain, Cold) - Probability of both rain and cold weather.

Marginal probability distribution focuses on the probability of a single variable, ignoring the values of other variables

- **Formula**: P (X=x) = ∑yP(X=x ,Y=y) (for discrete variables) P(X=a) represents the probability of X taking value 'a', regardless of the value of Y.
- **Example**: P(Rain) - Probability of rain, regardless of temperature.

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

Covariance measures the relationship between two random variables X and Y. It indicates how changes in one variable are associated with changes in another. In the context of **joint probability distributions**, covariance helps in understanding whether two variables move together (positive covariance) or in opposite directions (negative covariance).

Formula for Covariance: Cov(X, Y) = E(XY) - E(X)E(Y).

- Cov(X , Y) represents the covariance between random variables X and Y.
- E[XY] is the expected value of the product of X and Y.
- E[X] is the expected value of X.
- E[Y] is the expected value of Y.

Interpretation:

- **Cov(X, Y) > 0** → Positive relationship (both increase together).
- **Cov(X, Y) < 0** → Negative relationship (one increases, the other decreases).
- **Cov(X, Y) = 0** → No linear relationship.

The strength of the relationship, ranging from **-1 to +1.**

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

Two random variables are independent if their joint probability distribution can be expressed as the product of their marginal probability distributions.

That is if P (X=x, Y=y) = P(X=x) * P(Y=y) for all possible values of x and y.

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

The correlation coefficient is a standardized version of the covariance, which allows for comparison of the degree of relationship between variables on different scales.

- The **correlation coefficient (ρ) is calculated as**: ρ = Cov (X, Y) / (σX * σY)
- Cov (X, Y) is the covariance between variables X and Y,
- σX and σY are their respective standard deviations of X and Y, respectively.

- **Interpretation:**
    - The correlation coefficient ranges from -1 to +1.
    - A value of +1 indicates a perfect positive linear relationship.
    - A value of -1 indicates a perfect negative linear relationship.
    - A value of 0 indicates no linear relationship.

34) What is sampling in statistics, and why is it important?

Sampling is the process of selecting a subset (sample) from a larger population to analyze and draw conclusions about the entire population. Instead of studying every individual or data point, we use a representative sample to make inferences.

Sampling Important :

Saves Time & Cost ⏳ 💰

- Collecting data from an entire population can be expensive and time-consuming.
- A well-chosen sample allows researchers to analyze trends faster.

Feasibility & Practicality 🏗

- In cases where testing the whole population is impossible (e.g., testing the durability of car tires), sampling makes research practical.

Accuracy & Efficiency 🎯

- A properly selected sample can provide accurate estimates of the population parameters.
- Statistical techniques reduce errors and biases in sampling.

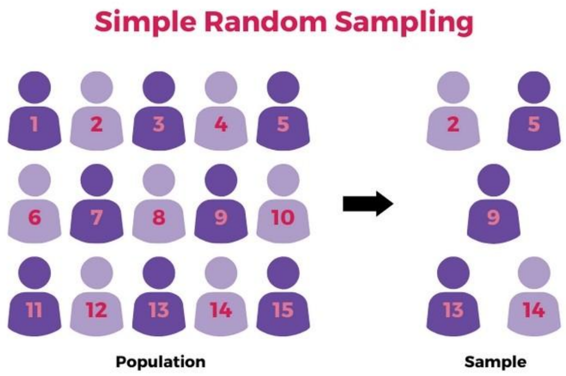
Basis for Statistical Inference 📊

- Sampling enables hypothesis testing, confidence intervals, and predictions

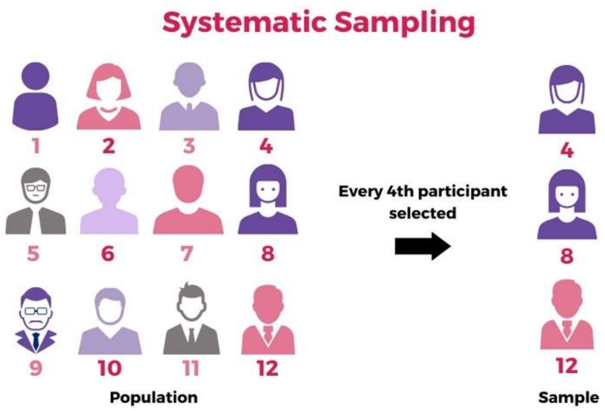
35) What are the different sampling methods commonly used in statistical inference?

There are several sampling methods, including **simple random sampling, stratified sampling, systematic sampling, cluster sampling, and convenience sampling**. Each method has its own advantages and disadvantages, and the choice of method depends on the research objectives and available resources.

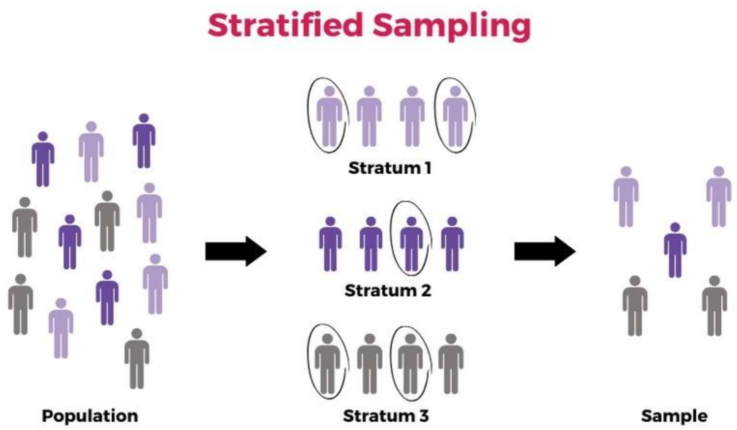
- **Simple Random Sampling:** Each member of the population has an equal chance of being selected.



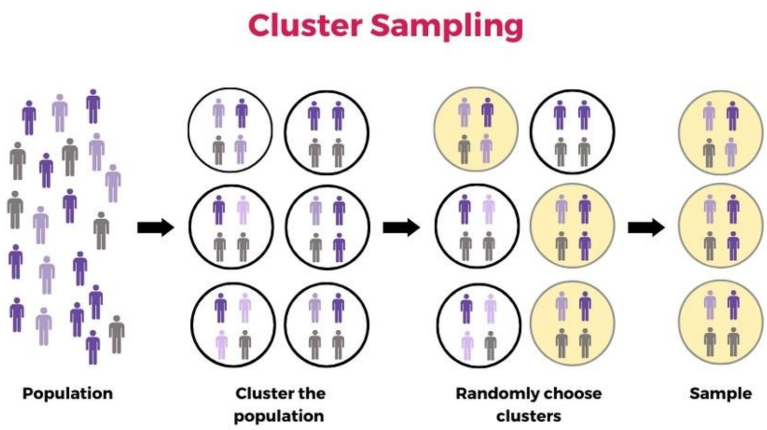
- **Systematic Sampling:** Selects individuals at regular intervals from a list or ordered population.



- **Stratified Sampling:** Divides the population into subgroups (strata) and then selects individuals randomly from each stratum.



- **Cluster Sampling:** Divides the population into clusters, and then randomly selects entire clusters to sample.



Non-Probability Sampling

- Judgment or purposive or deliberate sampling
- Convenience sampling
- Quota sampling
- Snow ball sampling

Judgment or purposive or deliberate sampling

- Judgment sampling, also known as purposive or deliberate sampling, involves selecting specific individuals or cases based on the researcher's judgment or specific criteria.
- This method is often used when the researcher believes that certain participants can provide valuable insights into the research topic.

Convenience Sampling

- Convenience sampling involves selecting individuals who are easily accessible or convenient for the researcher to reach.
- This method is simple and quick but may lead to biased results.

Quota Sampling

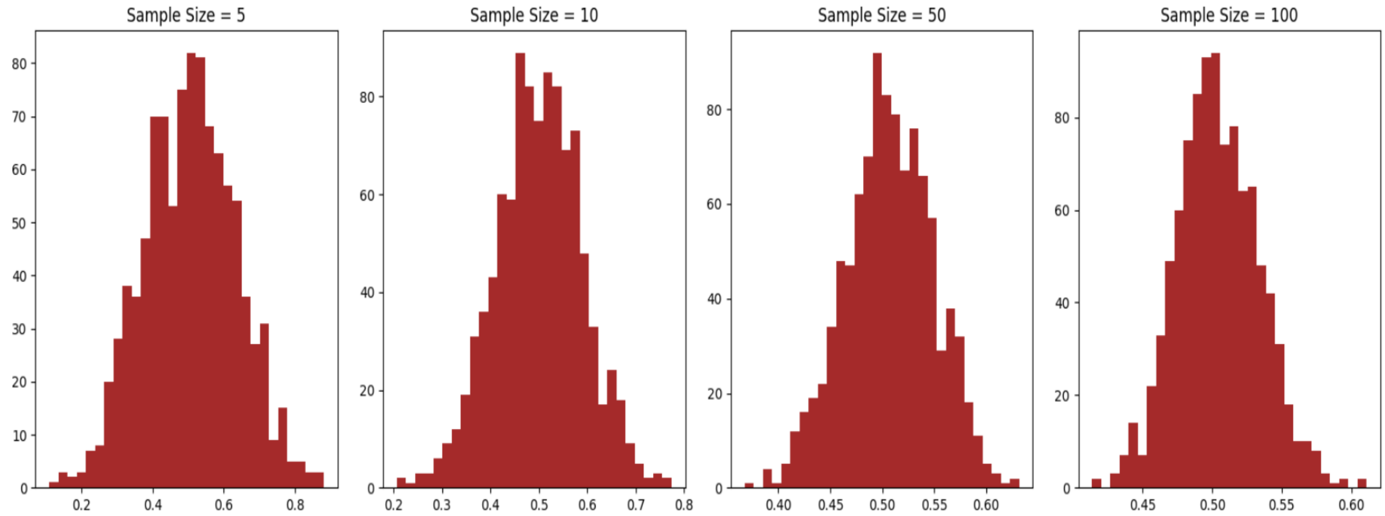
- Quota sampling involves dividing the population into subgroups (quotas) based on specific characteristics and then selecting participants from each subgroup.
- Quota sampling aims to ensure that the sample reflects the population's diversity in terms of the selected characteristics.

Snow ball sampling

- Snowball sampling is used when the population is hard to reach, and participants are recruited through referrals from existing participants.
- It's often used in studies involving hidden or marginalized populations.

36) What is the central limit theorem, and why is it important in statistical inference?

The central limit theorem states **that when independent random variables are added, their sum tends toward a normal distribution**, regardless of the shape of the original distribution. It is important because it allows us to make assumptions about **the sampling distribution of the sample mean**, even if the **population distribution is not known**.



Why is the Central Limit Theorem important?

1. **Sampling and Estimation:** The CLT is crucial for inferential statistics because it allows us to make inferences about population parameters based on samples. It provides a solid foundation for constructing confidence intervals and performing hypothesis tests, making statistical estimates more reliable and accurate.
2. **Simplification of Complex Distributions:** In many real-world scenarios, data may not follow a normal distribution, and their mathematical behavior can be quite complex. The CLT allows us to treat the sampling distribution of the mean as approximately normal, making statistical analysis much more manageable and feasible.
3. **Basis for Hypothesis Testing:** Hypothesis tests often rely on the assumption of normality, and the CLT allows us to apply these tests even when dealing with non-normally distributed populations, provided the sample size is large enough.
4. **Modeling and Simulation:** Many modeling and simulation techniques leverage the normal distribution due to its well-known properties. The CLT enables researchers to model and simulate complex phenomena by aggregating the effects of many random variables.

37) What is the difference between parameter estimation and hypothesis testing?

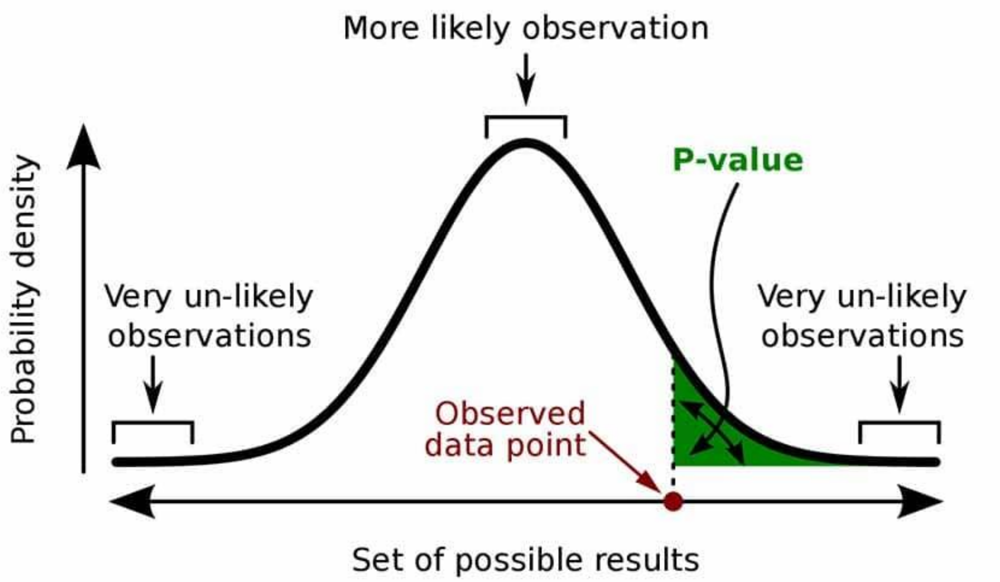
- Parameter estimation involves estimating unknown parameters, such as the population mean or variance, based on sample data.

- Hypothesis testing, on the other hand, involves making decisions about the population based on sample data, such as testing whether a specific hypothesis is true or not.

- Parameter estimation focuses on determining unknown population values.
- Hypothesis testing focuses on making decisions about population parameters based on statistical evidence.
- **Example:** Estimating the average height of students in a school based on a sample of students' heights.
- **Example:** Hypothesis Testing whether the average height of students in a school is different from a specific value (e.g., 5 feet) based on a sample of students' heights.

38) What is the p-value in hypothesis testing?

The p-value is the probability of obtaining a test statistic as extreme as, or more extreme than, **the observed value, assuming that the null hypothesis(H_0) is true.** It is used to determine the statistical significance of the results and helps in deciding whether to reject or fail to reject the null hypothesis.



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

- **Small p-value ($P < 0.05$)** → Strong evidence against H_0 → **Reject H_0**.
- **Large p-value ($p > 0.05$)** → Weak evidence against H_0 → **Fail to reject H_0**

A **p-value of less than 0.05** (or $P < 0.05$) is a commonly used threshold for statistical significance, meaning there's a **less than 5% chance of observing the results** (or more extreme results) if the null hypothesis is true.

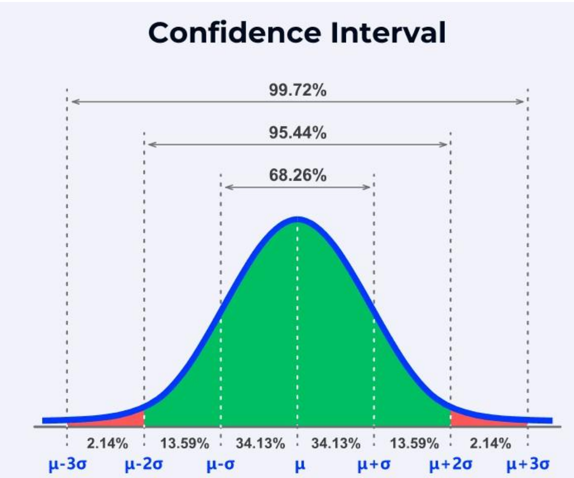
Decision Rule in Hypothesis Testing Using P-Value To decide whether to reject the **null hypothesis (H_0)**, compare the **p-value** to the **significance level (α)**:

- **If $p \le \alpha$: Reject H_0** (There is enough evidence to support H_1). The result is **statistically significant**.
- **If $p > \alpha$: Fail to reject H_0** (Not enough evidence to support H_1).
  The result is **not statistically significant**.

39) What is confidence interval estimation?

Confidence interval estimation is a method used to **estimate the range of values within which a population parameter** (like the mean or proportion) is likely to fall based on a sample. It provides a **range of plausible values rather than a single-point estimate**, and the confidence level represents the probability that the interval contains the true population parameter.

Example: A 95% confidence interval for a population mean might be 50 to 60, meaning that we can be 95% confident that the true population mean lies between 50 and 60.



Components:

Confidence level: The probability that the calculated confidence interval will contain the true population parameter. Commonly set at (e.g.95%, 99%), but can be adjusted depending on the desired level of certainty.

Lower and upper bounds: The two endpoints of the confidence interval, representing the range of plausible values for the population parameter.

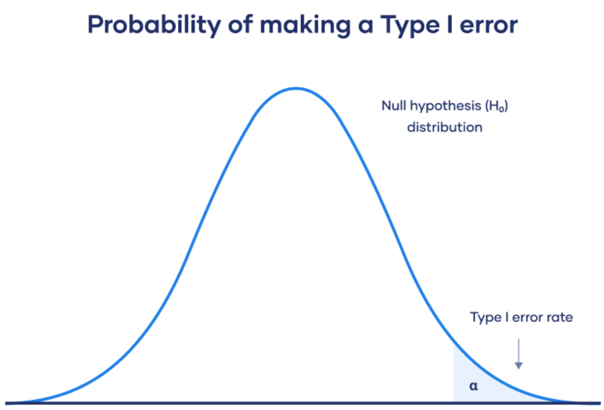
40) What are Type I and Type II errors in hypothesis testing?

In hypothesis testing, there are two types of errors that can occur when making a decision about the null hypothesis: **Type I error and Type II error**.

		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	We decide to reject the null hypothesis	Type I error (rejecting a null hypothesis when it is true)	Correct Decision
	We fail to reject the null hypothesis	Correct Decision	Type II error (failing to reject a false null hypothesis)

Type-I (False Positive) error occurs when the null hypothesis is rejected, even though it is true. It represents a false positive result.

- Type I error is denoted by α (alpha), which is also known as the significance level.
- By choosing a significance level, researchers can control the risk of making a Type I error.
- Null hypothesis and type 1 error
- If the resultant effect of this error is worse than a Type I error, one should consider alpha with a value higher than 0.10

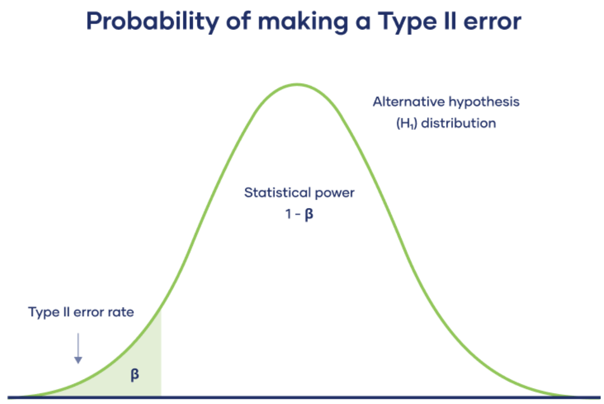


At the tail end, the shaded area represents alpha. It’s also called a critical region in statistics.

If your results fall in the critical region of this curve, they are considered statistically significant and the null hypothesis is rejected. However, this is a false positive conclusion, because the null hypothesis is actually true in this case!

Type-II (**False Negative**) error occurs when the null hypothesis is not rejected, even though it is false. It represents a false negative result.

- This means that the researcher fails to detect a significant effect or relationship when one actually exists.
- The probability of committing a Type II error is denoted by β (beta).
- Alternative hypothesis and type 2 error
- If the resultant of a Type I error is worse, one should set alpha with a value lower than 0.01.



The Type II error rate is beta (β), represented by the shaded area on the left side. The remaining area under the curve represents statistical power, which is 1 – β.

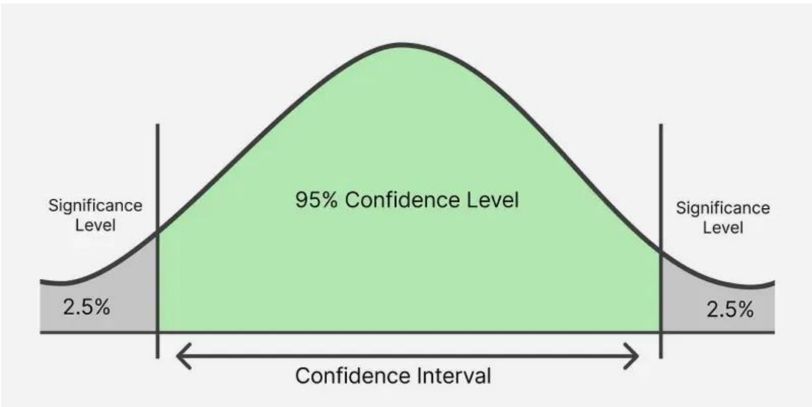
Increasing the statistical power of your test directly decreases the risk of making a Type II error.

41) What is the difference between correlation and causation?

Correlation refers to a statistical relationship between two variables, indicating how they move together. Causation, on the other hand, implies that one variable directly influences or causes a change in another variable. While correlation can suggest a potential relationship, it does not prove causation, as there may be other underlying factors or confounding variables at play.

42) How is a confidence interval defined in statistics?

A confidence interval is a **range of values that is constructed around an estimate** and is used to quantify the uncertainty associated with the estimate. It provides a level of confidence that the true population parameter lies within the interval.



How to Calculate **The interval is calculated using the following steps**:

1. Gather the sample data:
   - Collect a random sample from the population.
   - Determine the **sample size** (n).
2. Calculate the sample mean $\bar{x}$: sample statistic (e.g., sample mean, sample proportion).

3. Determine whether a population's standard deviation is known or unknown.

    - **If σ is known** → Use **Z-score** (Standard Normal Distribution).
    - **If σ is unknown** → Use **t-statistic** (Student's t-distribution).

4. Choose the Correct Distribution Based on the Critical Value /Confidence Level:
    - If σ (population standard deviation) is known, use Z-scores from the normal distribution (90 % CI → Z = 1.645, 95 % CI → Z = 1.96, 99 % CI → Z = 2.576)
    - If σ is unknown, use a T-score from the Student's t-table based on degrees of freedom (df = n−1).

5. Compute the Standard Error (SE).:
    - If **σ is known**: SE=σ/√n
    - If **σ is unknown** (using sample standard deviation s): SE=s/√n

6. Calculate the **Confidence Interva**l:
    - **Formula for confidence interval**: **CI= x̄ ± (Critical Value × SE)**
        - x̄ = Sample Mean
        - **Critical Value** = Z-score (if population standard deviation σ is known) or t-score (if σ is unknown)
        - **SE (Standard Error):** σ is known SE=σ/√n OR σ is unknown (using sample standard deviation s) SE=s/√n
    - **Lower Bound:** x̄ − (Critical Value × SE)
    - **Upper Bound:** x̄ + (Critical Value × SE)

Example: Estimating the Average Height of Students

Suppose we randomly sample **100 students** and find their **average height** to be **168 cm** with a **standard deviation of 6 cm**. We want to construct a **95% confidence interval** for the true average height of all students.

Step 1: Identify Given Values

- Sample mean (x̄) = 168 cm
- Standard deviation (s) = 6 cm
- Sample size (n) = 100
- Confidence level = 95%
- Critical Value for 95% CI → Z = 1.96 (from Z-table)

Step 2: Standard Error (SE)

SE= s / √n = 6/ √100 = 6/10= 0.6

~~**Step 3: Margin of Error (ME)**~~

~~ME = Z * SE = 1.96 * 0.6 = 1.176~~

Step 4: Confidence Interval

CL= $\bar{x}$ ± (Critical Value × SE) = 168 ± 1.96 * 0.6 = **168 ± 1.176**

Lower Bound= 168 − 1.176 = 166.82

Upper Bound= 168 + 1.176 = 169.18

Confidence Interval = (166.82 ,169.18)

Interpretation: We are **95% confident** that the **true average height** of all students in the school is between **166.82 cm and 169.18 cm**.

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

The confidence level is the percentage of times that the true population parameter will be within the calculated interval. It's a measure of how sure you can be that your estimate is accurate.

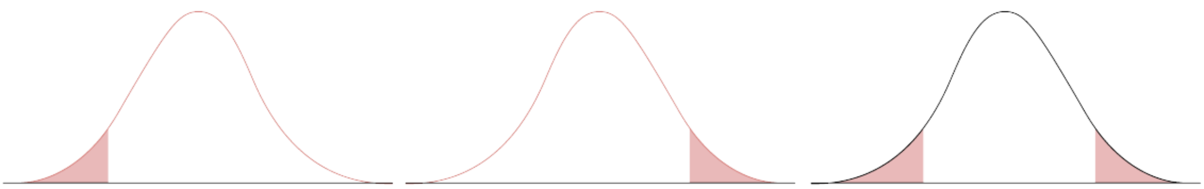
- A 95% confidence level means that if we were to take 100 different random samples and compute confidence intervals for each, about 95 of those intervals would contain the true population parameter.
- The confidence level is usually expressed as a percentage (e.g., 90%, 95%, or 99%).
- A **higher confidence level** (e.g., 99%) results in a **wider confidence interval**, while a **lower confidence level** (e.g., 90%) gives a **narrower confidence interval**.

44) What is hypothesis testing in statistics?

Hypothesis testing is a statistical method used to make inferences about **population parameters based on sample data**. It involves formulating a null hypothesis and an alternative hypothesis, collecting sample data, and evaluating the **evidence to determine whether there is enough evidence to reject the null hypothesis in favor of the alternative hypothesis**.

$H_0: \mu = claim$ $\quad$ $H_0: p = claim$ $\quad$ $H_0: \beta = claim$

$H_1: \mu < claim$ $\quad$ $H_1: p > claim$ $\quad$ $H_1: \beta \neq claim$



Null Hypothesis (H_0): The default assumption that there is **no effect or no difference**. It says there is no relationship between groups.

- We test the null hypothesis directly.

- Either reject H_0 or fail to reject H_0.
- Example: A company claims its website gets **50 visits per day**.:
  Null Hypothesis: H_0: The mean number of daily visits (μ) = 50.

Alternative Hypothesis (H_1 or H_a): The alternative hypothesis (denoted by H_1 or H_a) is the statement that the statistic has a value that somehow differs from the null hypothesis.

- The symbolic form of the **alternative** hypothesis must use one of these **symbols: ≠, <, >.**

The alternative hypothesis is typically what researchers are hoping to find evidence for, as it represents a new theory or idea that could advance our understanding of the world.

Example: **H_1**: The average daily visits **≠ 50** (two-tailed test).

- If we find strong evidence that actual visits are different, we **reject H_0 in favor of H_1**

There are **three types** of **alternative hypotheses**:

1 One-tailed alternative hypothesis.

2. Two-tailed alternative hypothesis.

3. Non-inferiority or equivalence alternative hypothesis:

Example: Testing a New Pain Medication 💊

A pharmaceutical company wants to test whether a new pain medication is **more effective** than an existing one.

- **Null Hypothesis (H_0)** (Default Assumption): The new medication is **no more effective** than the existing one.
  - **H_0**: The mean effectiveness score of the new medication **≤** the mean effectiveness score of the existing medication

- **One-Tailed Alternative Hypothesis:** Tests if the new medication is more effective than the existing one.
  - **H_1**: The new medication **>** the existing medication (right-tailed test).
  - Used when we expect a specific improvement.

- **Two-Tailed Alternative Hypothesis:** Tests if the new medication has a different effect (could be better or worse).
  - H_1: The new medication has a different effectiveness than the existing one (two-tailed test).
  - Used when we are unsure of the direction of the effect.

- **Non-Inferiority or Equivalence Alternative Hypothesis**:
  - Tests if the new medication is not significantly worse than the existing one.

- Used in clinical trials when a new treatment is expected to be at least as good.

Two Types:

- Non-Inferiority Test:
  - H_0: The new medication is worse than the existing one.
  - H_1: The new medication is not worse by more than a small margin (δ).
- Equivalence Test:
  - H_0: The two medications are not equivalent.
  - H1: The two medications are statistically equivalent within a margin ($\pm\delta$).

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

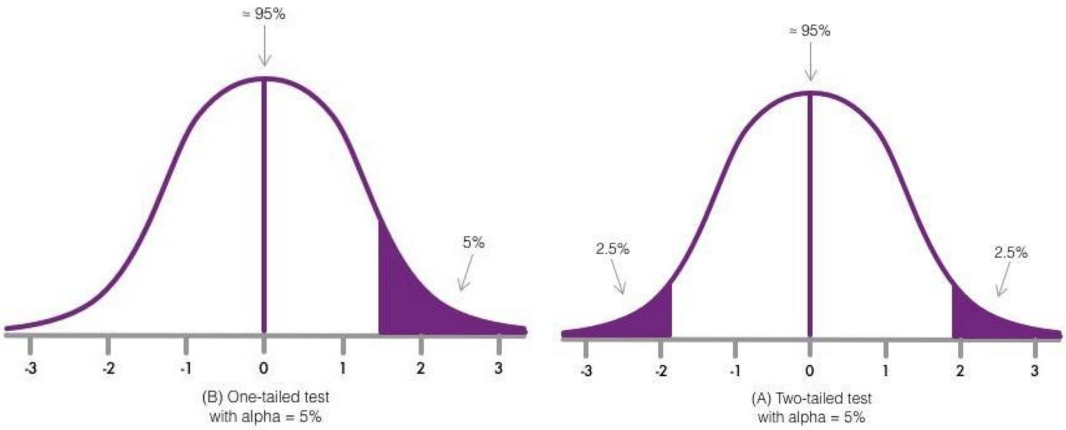
The default assumption that there is **no effect or no difference**. It says there is no relationship between groups.

- We test the null hypothesis directly.
- Either reject H_o or fail to reject H_o.
- Example: A company claims its website gets **50 visits per day**.:
  Null Hypothesis: H_0: The mean number of daily visits (μ) = 50.

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

In hypothesis testing, the choice between a one-tailed and a two-tailed test depends on the nature of the alternative hypothesis (HA) and how we assess significance.

- One-tailed test a of any statistical hypothesis, **where the alternative hypothesis is one-tailed either right-tailed or left-tailed**.
- One-tailed Test, **we use either > or < sign** for the alternative hypothesis.
- One-tailed test an Entire level of **significance** (α) i.e. 5% has either in the left tail or right tail.
- One-tailed test reject H_0 if **p-value < α** in one direction
- Example: Testing if a **new drug is more effective** than an old one (**H_1: $\mu > \mu_0$**)



- Two-tailed test of a statistical hypothesis, **where the alternative hypothesis** is two-tailed.
- Two-tailed a test we **use ≠ sign** for the alternative hypothesis.
- Two-tailed test it **splits the level of significance (α)** into half.
- Two-tailed reject H_0 if **p-value < α/2** in either direction
- Example: Testing if a **new drug has a different effect** than an old one (**H_1: $\mu \neq \mu_0$**)

47) What is experiment design, and why is it important?

Experiment design refers to the process of planning and organizing an experiment to gather data and test specific hypotheses.

It is **important** because a well-designed experiment allows researchers to control variables, minimize bias, and draw reliable conclusions about cause-and-effect relationships.

48) What are the key elements to consider when designing an experiment?

When designing an experiment, key elements to consider include defining the research question, determining the variables, selecting an appropriate sample or population, deciding on the experimental design (e.g., randomized controlled trial, factorial design), and establishing a suitable data collection and analysis plan.

49) How can sample size determination affect experiment design?

Sample size determination is a crucial aspect of experiment design. A larger sample size generally increases the statistical power of the study, allowing for more accurate detection of treatment effects. It also helps ensure that the study findings are representative of the target population.

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

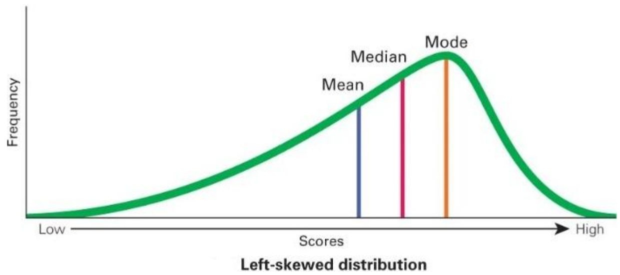
Strategies to mitigate bias in experiment design include randomization, blinding, careful selection and allocation of participants, proper measurement techniques, and careful control of confounding variables. These strategies help minimize systematic errors and increase the reliability of the study results.

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

- Confidence interval (CI) test estimates a range for a population parameter (e.g., mean, proportion)
- Confidence in output Provides an interval (e.g., $\mu \in [45, 55]$).
- Example Question: What is the **average height** of students with 95% confidence?
- Hypothesis tests a claim about a population parameter.
- Hypothesis in output Provides a decision (reject or fail to reject H0).
- Example Question: Is the **average height** of students **greater than 170 cm**?

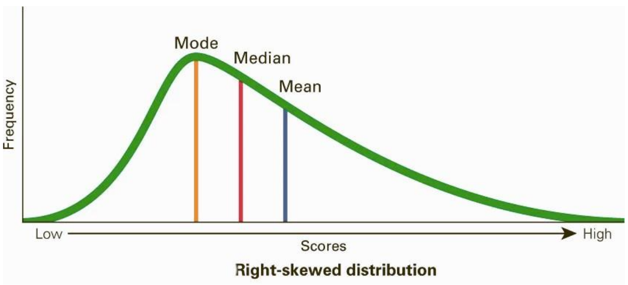
56) What is the left-skewed distribution and the right-skewed distribution?

A **left-skewed** distribution has a **longer tail on the left side of the data**, meaning most data points cluster towards the higher values with a few outliers on the lower end.



- Left skew is also referred to as negative skew.
- The mean of a left-skewed distribution is almost always **less than its median. (mean < median)**

A **right-skewed** distribution has a **longer tail on the right side**, where most data points cluster towards the lower values with a few outliers on the higher end.



Right skew is also referred to as positive skew.

The mean of a right-skewed distribution is almost always greater than its median. That's because extreme values (the values in the tail) affect the mean more than the median**. (mean > median)**

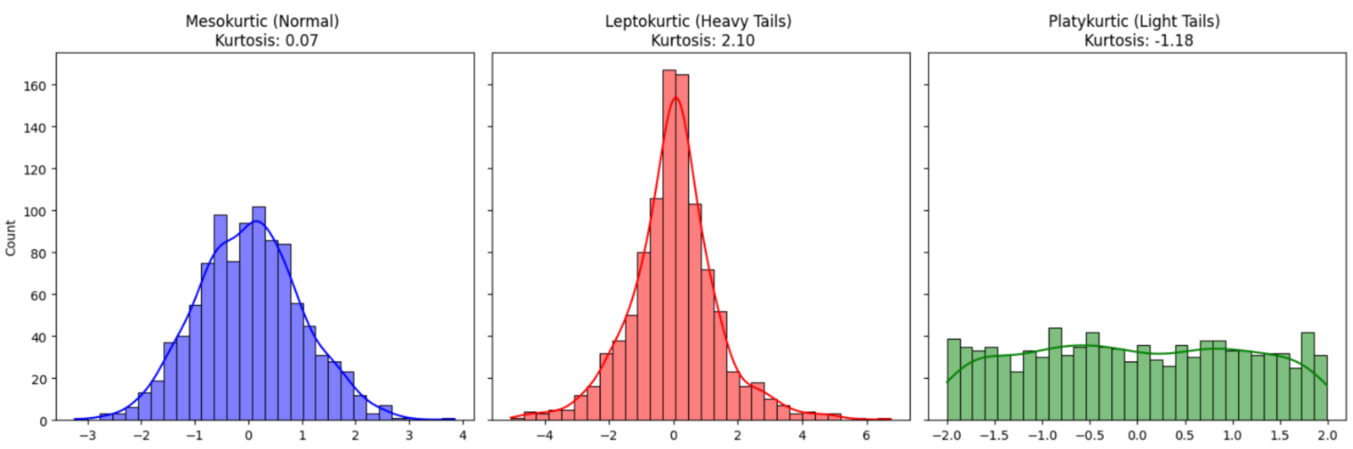
57) What is Bessel's correction?

Bessel's correction is the use of n-1 instead of n in the formula for sample variance and standard deviation. This correction is used to account for the bias in estimating population variance and standard deviation.

58) What is kurtosis?

Kurtosis is a statistical measure that describes the “tailedness” of a probability distribution. It indicates whether the data has heavy or light tails compared to a normal distribution. The kurtosis of a normal distribution equals 3.

- The three types of kurtosis:
  - Mesokurtic (Normal Distribution).
  - Leptokurtic (Heavy Tails).
  - Platykurtic (Light Tails).



1. **Mesokurtic (Normal Distribution)**: **Kurtosis ≈ 3**
   - Similar to a normal distribution, with moderate tails.
   - Example: Standard Normal Distribution (bell-shaped).
2. **Leptokurtic (Heavy-Tailed): Kurtosis > 3**
   - More peaked than a normal distribution, with **fatter tails** (more extreme values).
   - Example: Financial returns (higher risk of extreme losses/gains).
3. **Platykurtic (Light-Tailed): Kurtosis < 3**
   - Flatter than a normal distribution, with **thinner tails** (fewer extreme values).
   - Example: Uniform distribution.

59) What is the probability of throwing two fair dice when the sum is 5 and 8?

 Two fair dice: 6 * 6 = 36 total outcomes.

 When the sum is 5 and 8?:

Probability of Rolling a Sum of 5:

Possible pairs that sum to 5: (1,4), (2,3), (3,2), (4,1)

Total ways = 4

Probability of Rolling a Sum of 8:

Possible pairs that sum to 8: (2,6), (3,5), (4,4), (5,3), (6,2)

Total ways = 5

- Probabilities:

P (Sum = 5) = 4/36 = 1/9

p (sum= 8) = 5/36

Combined Probability (Sum = 5 or 8)

Since getting a sum of 5 and sum of 8 are mutually exclusive events, we add their probabilities:

P (Sum=5 or 8) = P (Sum=5) + P (Sum=8) = 4/36 + 5/36 = 9/36 = 1/4

Final Answer: The probability of rolling a sum of 5 or 8 is

1/4 or 25%

60) What is the difference between Descriptive and Inferential Statistics?

- Descriptive statistics are used to **summarize and describe the important characteristics** of a dataset.
- These statistics help to provide a quick overview of the data, identify patterns or trends, and determine if the data is skewed or has outliers.
- Example: Measures of Central Tendency: Mean, Median, Mode
- Example: Measures of Dispersion: Range, Variance, Standard Deviation
- Graphical Representations: Histograms, Pie Charts, Boxplots
- Inferential statistics allows us to make **predictions or draw conclusions about a larger population based on a sample of data**.

- Example: Hypothesis Testing (t-tests, chi-square tests, ANOVA), Confidence Intervals (Estimating population parameters), Regression Analysis (Predicting relationships), Probability Distributions (Normal, Binomial, Poisson)

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

Degrees of Freedom (DF) refer to the number of independent values in a dataset that can vary while estimating a statistical parameter. It represents the amount of information available to make an inference.

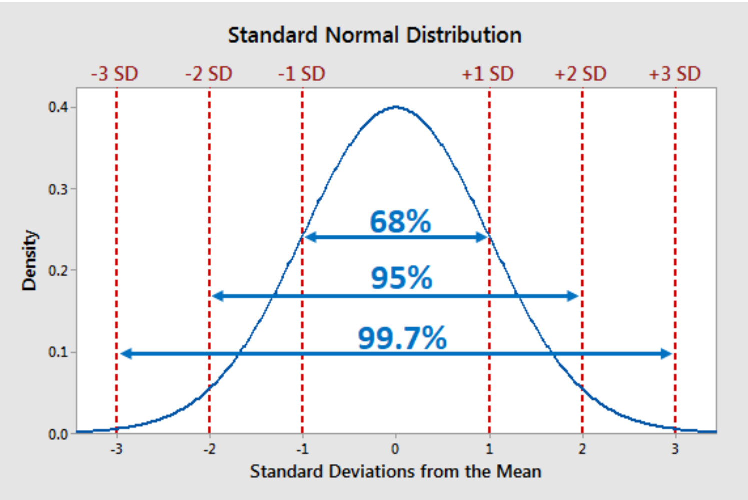
- They are used in hypothesis testing, including t-tests, chi-square tests, and ANOVA.
- They influence the shape of probability distributions, such as the t-distribution.

Degrees of Freedom Formula

- The formula to determine degrees of freedom is:
- Df = N− 1 where: Df= degrees of freedom N=sample size

65) What is the empirical rule in Statistics?

The Empirical Rule, also known as the 68-95-99.7 Rule, **states that for a normal (bell-shaped) distribution**, most of the data falls within **three standard deviations of the mean**.



📌 Empirical Rule:

- 68% of the data falls within 1 standard deviation of the mean: **$\mu - \sigma \leq X \leq \mu + \sigma$**
- 95% of the data falls within 2 standard deviations of the mean: **$\mu - 2\sigma \leq X \leq \mu + 2\sigma$**
- 99.7% of the data falls within 3 standard deviations of the mean: **$\mu - 3\sigma \leq X \leq \mu + 3\sigma$**

Example: Suppose exam scores in a class follow a normal distribution with **Mean (μ) = 70, Standard deviation (σ) = 10**

- Using the Empirical Rule:
  - **68% of students scored between 60 and 80 (70 ± 10)**
  - **95% of students scored between 50 and 90 (70 ± 20)**
  - **99.7% of students scored between 40 and 100 (70 ± 30)**

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

Power of a test refers to the probability of correctly rejecting the null hypothesis (H_0) when the alternative hypothesis (H_A) is true.

- Power=1−β

β is the probability of making a Type II error (failing to reject H_0 when H_A is true).

Relation Between:

- Larger Sample Size → Higher Power:
  - A larger sample size reduces the variability (standard error), making it easier to detect a true effect.
  - It decreases the chance of a **Type II error (β), increasing the power**.
- Smaller Sample Size → Lower Power:
  - A smaller sample size increases variability, making it harder to detect a true effect.
  - It increases the probability of failing to **reject H_0 when H_A is true.**

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

Yes! Non-parametric hypothesis tests are used when the data does not meet the assumptions of parametric tests (such as normality or equal variances). These tests do not rely on specific distributional assumptions, making them useful for small sample sizes, ordinal data, and skewed distributions.

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

The width of a confidence interval is influenced by several factors, including the sample size, the variability of the data, and the chosen confidence level. Generally, larger sample sizes and lower variability result in narrower confidence intervals.

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

Increasing the confidence level increases the width of the confidence interval. This is because a higher confidence level requires a wider interval to provide a higher level of certainty that the true parameter is captured within it.

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

No, it only provides a range of values that are likely to include the true population parameter with a certain level of confidence, based on a sample taken from that population.

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

Larger sample sizes result in narrower confidence intervals. As the sample size increases, the estimate becomes more precise, reducing the uncertainty associated with the estimate and leading to a narrower interval.

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

The confidence interval (**CI**) and margin of error (**MOE**) are closely related concepts in statistics. The margin of error helps determine the width of the confidence interval, which provides a range where the true population parameter is expected to lie.

Key Relationships:

1. **Wider Confidence Interval = Larger Margin of Error**
    - Increasing the **confidence level** (e.g., from 95% to 99%) increases the **margin of error**, making the confidence interval wider.
2. **Narrower Confidence Interval = Smaller Margin of Error**
    - Increasing the **sample size (n)** decreases the margin of error, resulting in a **more precise** estimate.
3. **Standard Deviation (σ or s) Affects ME**
    - Greater variability (higher **σ**) increases the margin of error, making the CI wider.

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

Yes, it is possible for two confidence intervals with different widths to have the same confidence level. The width of the interval is influenced by factors like sample size and variability, while the confidence level is determined by the desired level of certainty, which may be the same for both intervals.

Example:

CI for a Large Sample (n = 500, 95% confidence level): Suppose we estimate the average income with a narrow CI of ($50,000 ± $1,000) → ($49,000, $51,000).

CI for a Small Sample (n = 30, same 95% confidence level): The confidence interval might be wider, e.g., ($50,000 ± $5,000) → ($45,000, $55,000).

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

Sampling error is the difference between a sample statistic (sample mean) and the true population parameter (population mean). It occurs because we are analyzing only a subset (sample) of the population rather than the entire population.

How Can Sampling Error Be Reduced?

Although sampling error **cannot be completely eliminated**, it can be **minimized** using the following strategies:

1. **Increase Sample Size (n)** 🏆
    - Larger samples provide more accurate estimates of population parameters and reduce variability.
    - Example: A survey of **10,000 people** is more reliable than one with **500 people**.
2. **Use Random Sampling** 🎲
    - Ensures that every individual has an **equal chance of being selected**, reducing selection bias.
    - Avoid **convenience sampling**, which can lead to non-representative data.
3. **Stratified Sampling** 📊
    - Divide the population into **homogeneous subgroups (strata)** and sample from each proportionally.
    - Example: Instead of randomly selecting students from a school, divide them into **grade levels** and sample proportionally from each.
4. **Reduce Measurement Errors** ✍️
    - Ensure **accurate data collection methods**, well-trained surveyors, and reliable instruments.

5. **Avoid Sampling Bias** 🚫
    - Make sure the sample **accurately reflects** the entire population.
    - Example: A survey on smartphone usage should include **both young and old users**, not just tech enthusiasts.
6. **Use a Larger Population Variability Estimate** 📈
    - When designing a study, accounting for higher **standard deviation (σ)** can lead to better sample size calculations.

75) What is a Chi-Square test?

A Chi-Square (χ^2) test is a non-parametric statistical test used to determine whether there is a significant association between categorical variables in a dataset.

The chi-square test is used to **analyze categorical data, which means data that can be sorted** into categories (e.g., yes/no, male/female, colors).

Chi-Square Test Formula

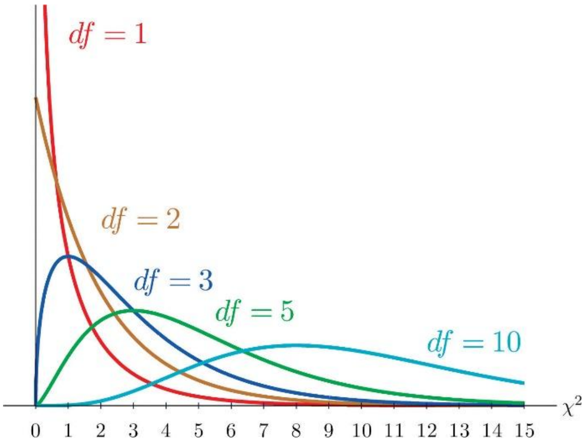
$$x_c^2 = \frac{\Sigma\,(O_i - E_i)^2}{E_i}$$

c = Degrees of freedom

O = Observed Value

E = Expected Value

The degrees of freedom in a statistical calculation represent the number of variables that can vary. The degrees of freedom can be calculated to ensure that Chi-Square tests are statistically valid. These tests frequently compare observed data with data expected to be obtained if a particular hypothesis were true.



Chi-square goodness of fit test

You can use a **chi-square goodness of fit test** when you have **one** categorical variable. It allows you to test whether the frequency distribution of the categorical variable is significantly different from your expectations. Often, but not always, the expectation is that the categories will have equal proportions.

Example: Hypotheses for chi-square goodness of fit test **Expectation of equal proportions**

- **Null hypothesis (*H_0*):** The bird species visit the bird feeder in **equal** proportions.
- **Alternative hypothesis (*H_A*):** The bird species visit the bird feeder in **different** proportions.

Expectation of different proportions

- **Null hypothesis (*H_0*):** The bird species visit the bird feeder in the **same** proportions as the average over the past five years.
- **Alternative hypothesis (*H_A*):** The bird species visit the bird feeder in **different** proportions from the average over the past five years.

Chi-square test of independence

You can use a **chi-square test of independence** when you have **two** categorical variables. It allows you to test whether the two variables are related to each other. If two variables are independent (unrelated), the probability of belonging to a certain group of one variable isn't affected by the other variable.

Example: Chi-square test of independence

- **Null hypothesis (*H_0*):** The proportion of people who are left-handed is **the same** for Americans and Canadians.
- **Alternative hypothesis (*H_A*):** The proportion of people who are left-handed **differs** between nationalities.

76) What is a t-test?

A t-test is a statistical test used to **compare the means of one or two groups to determine** if the differences between them are statistically significant.

It is commonly used when the sample size is small ($n < 30$) and the **population standard deviation is unknown**. which is similar to the normal distribution but has heavier tails.

The t-test is part of **hypothesis testing** where you start with an assumption the null hypothesis that the two-group means are the same. Then the test helps you decide if there's enough evidence to reject that assumption and conclude that the groups are different.

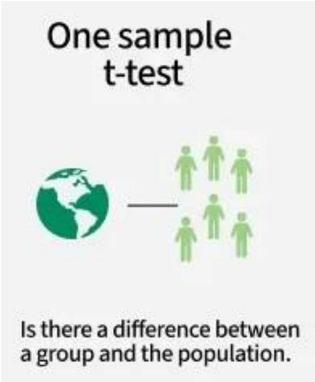
Assumptions of a t-test:

- Data is normally distributed (for small samples, normality is assumed).
- Samples are independent (except for paired t-tests).
- Equal variance (for independent t-tests, if variances are unequal, use Welch's t-test).

There are three types of t-tests:

1) One-sample t-test:

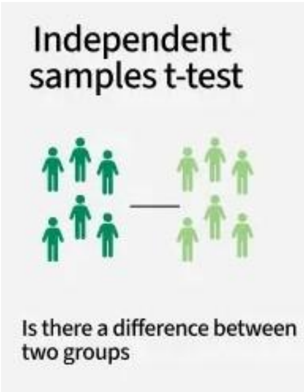
- The one-sample t-test is used to **compare the mean of a single sample** to a known population mean.



- Used to determine if the sample comes from a population with a specific mean.
- The null hypothesis states that there is no significant difference between the sample mean and the population mean, while the alternative hypothesis states that there is a significant difference.

2) Two-Sample T-Test (Independent T-Test):

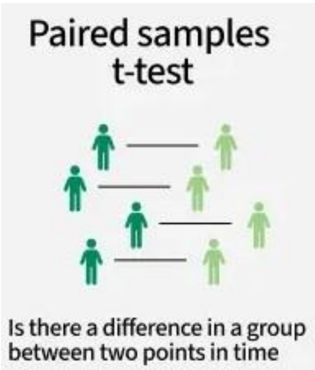
- The independent two-sample t-test is used to **compare the means of two independent samples**.



- Used to determine if there is a significant difference between the means of two groups.
- The null hypothesis states that there is no significant difference between the means of the two samples, while the alternative hypothesis states that there is a significant difference.

3) Paired t-test (dependent two-sample t-test):

- The paired t-test is used to **compare the means of two samples that are dependent or paired**, such as pre-test and post-test scores for the same group of subjects or measurements taken on the same subjects under two different conditions.



- Used to determine if there is a significant difference between paired observations.
- The null hypothesis states that there is no significant difference between the means of the paired differences, while the alternative hypothesis states that there is a significant difference.

77) What is the ANOVA test?

The ANOVA test (Analysis of Variance) is a statistical method used to determine whether there are statistically significant differences between the means of three or more independent groups.

It compares the variance between group means to the variance within groups to assess if the observed differences are likely due to chance or a true effect.

When comparing **more than two** group means, performing multiple t-tests increases the risk of **Type I error** (false positives).

Types of ANOVA:

1) **One**-Way **ANOVA**: Comparing **one categorical independent variable** with three or more groups. (e.g., Sales performance for three marketing strategies: A, B, C)

2)**Two**-Way **ANOVA**: Comparing **two categorical independent variables**. (e.g., Impact of both marketing strategy and region on sales.)

3) **Repeated** Measures **ANOVA**: Measuring the **same group multiple times under different conditions.** (e.g., Testing drug effectiveness at 1-month, 3-month, and 6-month intervals.)

78) How is hypothesis testing utilised in A/B testing for marketing campaigns

A/B testing (or split testing) is a controlled experiment where two versions of a marketing element (e.g., an ad, email, or webpage) are tested to determine which performs better. Hypothesis testing is the statistical backbone of A/B testing.

A/B testing is an experimental method in which two versions of anything are contrasted to see which is “better” or more effective.

Apply Hypothesis Testing in A/B Testing:

Define Hypotheses

- **Null Hypothesis (H_0):** There is **no significant difference** in the conversion rates between Version A and Version B.
- **Alternative Hypothesis (H_1 or H_a):** There is a **significant difference** in conversion rates between Version A and Version B.

Example:

- H_0: "The conversion rate of Ad A is equal to the conversion rate of Ad B."
- H_1: "The conversion rate of Ad A is higher than that of Ad B."

Collect Data

- Select a **random sample** of users and split them into **two equal groups**:
  - **Group A** → Sees the original version (Control)
  - **Group B** → Sees the new version (Variation)
- Track key performance metrics (e.g., **click-through rate (CTR), conversions, revenue**).

Choose a Statistical Test

- **Z-test or t-test:** Used for comparing two proportions (e.g., conversion rates).
- **Chi-Square Test:** Used if the data is **categorical** (e.g., button clicks vs. no clicks).

- **ANOVA (for multiple variations):** If testing **more than two versions**.

Set Confidence Level & Significance

- Typically, a **95% confidence level** ($\alpha = 0.05$) is used.
- If **p-value < 0.05**, reject H_0 → A significant difference exists.

Analyze Results

- Calculate:
  - **Conversion Rate (CR) = (Conversions / Visitors) × 100**
  - **P-value**: Probability of obtaining results as extreme as observed, assuming H_0 is true.
  - **Confidence Interval (CI)**: The range within which the true conversion rate lies.
- **Decision:**
  - If p-value < 0.05 → **Reject H_0** (Version B performs significantly better or worse).
  - If p-value ≥ 0.05 → **Fail to reject H_0** (No significant difference).

80) What is an inlier?

An inlier is a data point that lies within the expected range of a dataset and follows the general trend or distribution of the data. Unlike outliers, inliers do not significantly deviate from the other data points.