1. What is a vector in mathematics?

A vector is a physical quantity that has both magnitude and direction. It is typically represented as an arrow in space.
Eg :- force, velocity, acceleration,etc

2. How is a vector different from a scalar?

Let's compare the differences as a table.

Scalar	**Vector**
A quantity with only magnitude	A quantity with both magnitude and direction
It will be a single number (eg: 5)	It will be a tuple or array (eg: [4,3])
eg:- force, velocity, acceleration	eg:- temperature, speed, mass, length

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

(i) Vector addition :

We can add two vectors component-wise.

(3,2) + (6,1) = (9,3)

(i) Vector subtraction :

We can subtract two vectors similarly.

(3,2) - (6,1) = (-3,1)

(iii) Scalar multiplication :

We can multiply a vector with a scalar. It measures the similarity between two vectors.

5 (2,4) = (10,20)

(iv) Dot product :

Dot product of two vectors will be a scalar.

(5,4).(3,2) = 5x3 + 4x2 = 23

(v) Cross product :

Cross product of two vectors will be a vector. This vector will be perpendicular to two given vectors.

4. How can vectors be multiplied by a scalar?

Multiplying a vector by a scalar means scaling the vector's magnitude without changing its direction.

4 (2,7) = (8,28)

2 (3,2) = (6,4)

5. What is the magnitude of a vector?

Magnitude of a vector is the length of the vector. The magnitude of a vector v = (4,3) can be found as,

$$|v| = (4^2 + 3^2)^{0.5} = 5$$

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

Direction can be found by taking the inverse tan of y component divided by x component. The direction of a vector v = (4,3) can be found as,

$$\boldsymbol{\theta} = tan^{-1}(3/4)$$

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

In the square matrix, the number of rows and number of columns will be the same, whereas in the rectangular matrix, the number of rows and columns will not be the same. We can compute the determinant of a square matrix, but the determinant is not defined for a rectangular matrix.

8. What is a basis in linear algebra?

A basis of a vector space is a set of linearly independent vectors which allows every vector in the space to be uniquely represented as a combination of the basis vectors.

Consider 2D space :

Standard basis for 2D are,

$e_1 = (1, 0) \ \ and \ \ e_2 = (0, 1)$

Any vector v = (x,y) can be represented as $v = x.e_1 + y.e_2$

Consider 3D space :

Standard basis for 3D are,

$e_1 = (1, 0, 0)\ , \ \ e_2 = (0, 1, 0) \ \ and \ \ e_3 = (0, 0, 1)$

Any vector v = (x,y,z) can be represented as $v = x.e_1 + y.e_2 + z.e_3$

9. What is an eigenvector in linear algebra?

An eigenvector of a square matrix is a nonzero vector that, when multiplied by the matrix, results in a scalar multiple of itself.This means that the direction of the eigenvector remains unchanged, but its magnitude may be scaled by a factor called the eigenvalue.

$Av = \lambda v$

A is the matrix, v is the eigenvector and λ is the eigenvalue.

Suppose $A = [[4, 2], [1, 3]]$ and one of the eigenvector is $v = [2, 1]$ with eigenvalue 5.

$Av = [(4\times2 + 2\times1), (1\times2 + 3\times1)] = [10, 5] = 5[2, 1] = 5v$

10. What is the gradient in machine learning?

The gradient tells us how much a function changes if we change its inputs. The gradient tells you the direction where a function decreases the fastest.

In linear regression, we want to find the best line to fit the data.We calculate a gradient to see how much to change the slope and intercept. The model updates based on the gradient so that the error (loss) gets smaller.

11. What is backpropagation in machine learning?

Backpropagation is like a feedback loop that helps a neural network improve over time by learning from its errors. It is a method used to train neural networks by adjusting their weights to reduce errors.

The steps in neural networks :

- The network makes a prediction (Forward Pass).
- The error (difference between prediction and actual value) is calculated.
- The error is sent backward through the network to find which weights need to change (Backward Pass).
- The weights are updated to improve accuracy using Gradient Descent.

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

A derivative measures how a function changes as its input changes. It tells us the rate of change or the slope of a function at any point.

$m = dy / dx$

This is the equation for slope of a line, here it tells how the y value will vary when the x value varies.

13. How are partial derivatives used in machine learning?

A partial derivative measures how a function changes when only one variable is changed while keeping others constant.

For example if we have a function $F(x, y, z)$, the partial derivative of the function with respect to x tells how the function will change when the x changes holding y and z constants.Similarly, the partial derivative of the function with respect to y tells how the function will change when the y changes holding x and z constants.

In machine learning the partial derivatives are used in,

- Gradient Descent

- Backpropagation in Neural Networks
- Optimization of Functions, etc.

14. What is probability theory?

Probability theory is the branch of mathematics that deals with uncertainty. It helps us measure the likelihood of different outcomes in a random event. Range of probability is $[0, 1]$.

$P(A) = number\ of\ favorable\ outcomes\ /\ total\ number\ of\ outcomes$

15. What are the primary components of probability theory?

The components of probability are :

- Sample space (The set of all possible outcomes of an experiment)
- Events (A subset of the sample space, representing a specific outcome)
- Probability function ($P(A) = n(E) / n(S)$)
- Conditional probability (The probability of event A occurring given that event B has already occurred)
- Dependence and independence (In Independent Events, One event does not affect the other, whereas in Dependent Events One event does affect the other.
- Random variable (A function that assigns a numerical value to each outcome in a sample space)

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

Conditional probability is the probability of an event happening given that another event has already occurred.

If A and B are two events, the conditional probability of A occurring given that B has occurred can be calculated as,

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

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

Bayes' Theorem is a mathematical formula used to calculate the probability of an event occurring based on prior knowledge of related conditions. It updates probabilities when new information is available.

The formula is given by,

$P(A/B) = P(B/A).P(A) / P(B)$

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

A random variable is a variable that takes numerical values based on the outcome of a random event. A regular variable is a fixed number.

Regular variable	**Random variable**
Stores a fixed value	Represents uncertain outcomes
X = 5 (fixed number)	X = Number of heads in 3 coin flips
It does not change	Depends on the chance, it will change
One specific value at a time	Takes multiple values with probabilities

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

The Law of Large Numbers (LLN) states that as the number of trials in a random experiment increases, the average of the results gets closer to the expected value (true probability).

The key idea behind it is that The more we repeat an experiment, the closer the results get to the expected probability.

Example :

If you flip a fair coin 10 times, you might get 7 heads and 3 tails (70% heads instead of 50%).But if you flip the coin 100,000 times, the proportion of heads will be very close to 50%.

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

The Central Limit Theorem states that if we take many random samples from any population (regardless of its original distribution), the sampling distribution of the sample mean will approximate a normal distribution as the sample size increases.

It can be applied to any population distribution even if it's not normal.

Example :

A single roll of a die is uniformly distributed (each number 1-6 is equally likely).If we take many samples of 30 dice rolls and compute the mean for each sample, the distribution of these means will be approximately normal, even though the original die roll distribution is not.

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

A discrete probability distribution deals with countable outcomes like the number of heads in coin flips, while a continuous probability distribution represents uncountable(infinite possibilities) values like a person's height, where probabilities are assigned to ranges rather than specific values.

In discrete probability distributions, the probability is represented as a probability mass function. In continuous probability distributions, the probability is represented as a probability density function.

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

Mean : The sum of all values divided by the number of values.

Median : The middle value when data is sorted.

Mode : The most frequently occurring value.

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

Percentiles :

Divide data into 100 equal parts. The p-th percentile is the value below which p% of the data falls.Used to compare an individual value to the overall dataset.

Example:- A test score in the 90th percentile means it is higher than 90% of the other scores.

Quartiles :

Divide data into 4 equal parts: Q1 (25th percentile), Q2 (50th percentile / Median), Q3 (75th percentile).

Helps identify spread and skewness in data. Used in box plots for detecting outliers.

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

Outliers are extreme values that differ significantly from other observations in a dataset. They can affect statistical analysis and machine learning models.

The method to detect outlier :

- IQR method to detect outliers :

$IQR = Q_3 - Q_1$

$Lower\ bound = Q_1 - 1.5(IQR)$

$Upper\ bound = Q_3 + 1.5(IQR)$

Any data point outside these bounds is an outlier.

- z- score method to detect outliers (when data follows normal distribution) :

$z = (x - \mu)/\sigma$

If $z > 3\ or\ z < -3,$ then the data point is an outlier.

visualisation methods (box plot, scatter plot) are also used to detect outliers.

The method to treat outlier :

- Remove Outliers (If they are due to errors or not meaningful).
- Transform Data (Apply log, square root, or normalization techniques).
- Impute Outliers (Replace with the median or nearest valid value).

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

The central limit theorem states that when we take a large number of random samples from any probability distribution (even discrete ones), the sampling distribution of the sample mean will approximate a normal distribution, regardless of the original distribution's shape.

The central limit theorem helps to approximate a discrete probability distribution.Consider an example of a fair six-sided die has the discrete probability distribution.

- Rolling a fair six-sided die, where the outcomes are {1, 2, 3, 4, 5, 6} with equal probability.
- Roll the die 30 times per sample and compute the sample mean for each set of 30 rolls.
- Repeat the Process Many Times (Collect thousands of sample means to create a distribution)
- Observe That the Sample Mean Distribution is Normal

Even though a single die roll is discrete, the distribution of sample means forms a bell-shaped normal curve. The central limit theorem allows us to approximate discrete distributions by making the sample mean distribution normal for large n.

26. What is a joint probability distribution?

A joint probability distribution describes the probability of two or more random variables taking specific values simultaneously. It is a fundamental concept in probability theory and statistics, particularly when dealing with multiple variables that may be related or dependent on each other.

The joint probability distribution gives the probability of two events happening together.It's calculated by dividing the frequency of occurrence by the total sample size.The sum of all joint probabilities must be 1.

27. How do you calculate the joint probability distribution?

Suppose a survey is conducted on 200 people from two cities (A & B) about their preferred drink: Tea or Coffee. The data is:

Preference	**City A**	**City B**	**Total**
Tea	60	50	110
Coffee	40	50	90
Total	100	100	200

Now, let's calculate the joint probabilities,

- **Probability of a randomly selected person being from City A and preferring Tea**:

  $P(A \cap T) = 60/200 = 0.3$

  - **Probability of a randomly selected person being from City A and preferring coffee**:

$P(A \cap T) = 40/200 = 0.2$

- **Probability of a randomly selected person being from City B and preferring Tea**:

$P(A \cap T) = 50/200 = 0.25$

- **Probability of a randomly selected person being from City B and preferring Coffee**:

$P(A \cap T) = 50/200 = 0.25$

The sum of all probabilities will be 1

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

Joint probability distribution describes the probability of two events occurring together. For example, the probability that a randomly chosen person prefers Tea and is from City A is $60/200 = 0.3$.

Marginal probability distribution describes the probability of a single event occurring, regardless of other factors. For example, the probability that a randomly chosen person prefers Tea (regardless of city) is $(60 + 50)/200 = 0.55$

In simple terms, joint probability considers two conditions together, while marginal probability sums over one variable, ignoring the other.

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

Covariance of a joint probability distribution measures the relationship between two random variables. It tells us how changes in one variable correspond to changes in another.

The formula is given by, $Cov(X, Y) = \Sigma\,(X - E[X])\,(Y - E[Y])$

- A positive covariance means both variables tend to increase together.
- A negative covariance means one variable increases while the other decreases.
- A zero covariance means no relationship between them.

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

Two random variables X and Y are independent if their joint probability distribution equals the product of their individual (marginal) probabilities for all possible values.

$P(X = x, Y = y) = P(X = x).P(Y = y)$

This means that knowing the value of X does not provide any information about Y, and vice versa.

Suppose we roll two fair dice. The probability of getting a 3 on the first die and a 5 on the second die is:

$P(X = 3, Y = 5) = P(X = 3).P(Y = 5) = 1/6 \times 1/6 = 1/36$

Since the joint probability is simply the product of individual probabilities, X and Y are independent.

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

The correlation coefficient and covariance both measure the relationship between two random variables in a joint probability distribution.

The relationship between the correlation coefficient (ϱ) and the covariance is given by,

$\rho(X, Y) = COV(X, Y) / \sigma_X.\sigma_Y$

The correlation coefficient is a normalized version of covariance, making it dimensionless and easy to interpret.

- Covariance can take any value, making it difficult to interpret across different datasets.
- The correlation coefficient is always in the range **[−1,1]**

Depends on the value of correlation coefficient, we can measure the correlation,

- $\rho = 1$ → Perfect positive correlation
- $\rho =- 1$ → Perfect negative correlation
- $\rho = 0$ → no correlation

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

Sampling in statistics is the process of selecting a subset of individuals or observations from a larger population to analyze and draw conclusions about the entire population. It allows researchers to study a manageable portion of data instead of the whole population.

It is important because,

- Studying an entire population is often impractical, sampling makes analysis faster and more cost-effective
- In many cases, collecting data from every individual is impossible
- A well-chosen sample can provide reliable estimates about the population without needing exhaustive data collection

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

Sampling methods in statistics are divided into two main types: Probability Sampling and Non-Probability Sampling.

(1) Probability Sampling (Random Selection):

Every member of the population has a known, non-zero chance of being selected. This method ensures unbiased and representative samples.

- **Simple Random Sampling** – Every individual has an equal chance of being selected
- **Stratified Sampling** – The population is divided into subgroups (strata) based on a characteristic, and a random sample is taken from each.
- **Cluster Sampling** – The population is divided into groups (clusters), and entire clusters are randomly selected.
- **Systematic Sampling** – Selecting every k-th individual from a list after a random start.

(2) Non-Probability Sampling (Non-Random Selection) **:**

Not every member has a known or equal chance of being selected. It is used in exploratory research, qualitative studies, and when random sampling is impractical.

- **Convenience Sampling** – Selecting participants based on ease of access.
- **Quota Sampling** – Selecting a sample based on specific characteristics, similar to stratified sampling but without random selection.
- **Judgmental (Purposive) Sampling** – The researcher chooses participants based on expertise or criteria.

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

The Central Limit Theorem (CLT) states that when we take many random samples of size n from a population (regardless of its original distribution), the sampling distribution of the sample mean will approximate a normal distribution, given that the sample size is sufficiently large.

It is important because ;

- It allows us to use normal distribution-based techniques (like confidence intervals and hypothesis testing) even when the original population is not normally distributed.
- Many statistical tests assume normality; the CLT justifies their use when sample sizes are large.
- Used in areas like finance, machine learning, quality control, and social sciences for estimating population parameters.

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

Parameter Estimation :

Parameter estimation is the process of determining an unknown population parameter (such as the mean, proportion, or variance) using sample data. It provides a numerical estimate of the parameter, which can be a single value (point estimation) or a range of values (interval estimation).

For example, if we want to estimate the average height of students in a university, we might take a sample and compute the sample mean as an estimate of the true population mean. Additionally, we can construct a confidence interval to express the range in which the true mean is likely to fall.

Hypothesis Testing :

Hypothesis testing is a method used to make decisions about a population parameter based on sample data. It involves formulating a hypothesis (a claim about the population), collecting data, and using statistical tests to determine whether the claim is likely to be true. Hypothesis testing starts with a null hypothesis (H_0), which represents the assumption of no effect or no difference, and an alternative hypothesis (H_a), which represents the claim we want to test.

For instance, if a researcher wants to test whether the average height of students is greater than 170 cm, they would conduct a hypothesis test using methods like the t-test or z-test. The result of the test helps decide whether to reject or fail to reject the null hypothesis based on a significance level.

In summary, parameter estimation provides an approximate value of a population characteristic, while hypothesis testing evaluates whether a claim about a population is statistically supported by the data. Both methods are essential in statistical analysis and are widely used in fields such as business, medicine, and engineering.

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

The p-value in hypothesis testing is the probability of obtaining results as extreme as (or more extreme than) the observed data, assuming the null hypothesis (H_0) is true.

It helps determine whether we should reject H_0:

- Small p-value (≤ 0.05) → Strong evidence against H_0, so we reject it.
- Large p-value (> 0.05) → Weak evidence against H_0, so we fail to reject it.

37. What is confidence interval estimation?

Confidence interval estimation is a statistical technique used to estimate an unknown population parameter based on sample data. Instead of providing a single estimated value (point estimate), a confidence interval gives a range of values that is likely to contain the true population parameter with a specified confidence level (eg: 95%).

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

Type I Error (False Positive) :

- Occurs when we reject the null hypothesis (H_0) even though it is actually true.
- It is denoted by α , which represents the significance level (eg: 5% or 0.05 probability of making this error).
- Example: A medical test incorrectly detects a disease in a healthy person.

Type II Error (False Negative) :

- Occurs when we fail to reject the null hypothesis (H_0) even though it is actually false.
- It is denoted by β , and $1-\beta$ represents the statistical power (the ability to detect a true effect).
- Example: A medical test fails to detect a disease in a person who actually has it.

For more clarity consider a scenario,

H_0: The person does not have COVID-19

H_a: The person has COVID-19

Type I Error: The test wrongly indicates the person has COVID-19 (false positive).

Type II Error: The test wrongly indicates the person does not have COVID-19 (false negative).

39. What is the difference between correlation and causation?

Correlation and causation are both about relationships between variables, but they mean different things.

- Correlation means two variables move together, but one does not necessarily cause the other.

  Example: People who carry lighters tend to get lung cancer more often. But lighters don't cause cancer—smoking does.

- Causation means one variable directly influences another.

  Example: Studies show that smoking causes lung cancer because of the harmful chemicals in cigarettes.

40. What is hypothesis testing in statistics?

Hypothesis testing is a statistical method used to make decisions or inferences about a population based on sample data. It helps determine whether a claim or assumption (hypothesis) about a population parameter is supported by evidence.

Key Steps in Hypothesis Testing:

1. Define the Hypotheses:
   - Null Hypothesis (H_0): Assumes no effect or no difference (status quo).
   - Alternative Hypothesis (H_a): Represents what we want to test (a significant effect or difference).
2. Select a Significance Level (α):
   - Typically 0.05 (5%) or 0.01 (1%), meaning there's a small chance of wrongly rejecting H_0.
3. Choose a Test Statistic and Compute It:
   - Common tests: t-test, z-test, chi-square test, ANOVA, etc.

4. Compare with the Critical Value or Compute the p-value:
    - p-value ≤ α → Reject H_0 (strong evidence against H_0).
    - p-value > α → Fail to reject H_0 (not enough evidence).

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

Total number of pairs in the sample space, $n(S) = 6\times6 = 36$

Probability of rolling a sum of 5 :

$E = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$

$P = n(E)/n(S) = 4/36 = 1/9 = 0.111$

Probability of rolling a sum of 8 :

$E = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$

$P = n(E)/n(S) = 5/36 = 0.139$

42. 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 given by,

$$Z = (X - \mu)/\sigma$$

$\mu = 160$, $\sigma = 15$, $Z = 1.2$

$X = Z\sigma + \mu = (1.2\times15) + 160 = 178$

The score in the test, $X = 178$

43. 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?

A high correlation between sleep time and productivity suggests that the amount of sleep a person gets is related to their level of productivity. If the correlation is positive, it means that more sleep is associated with higher productivity, possibly because sleep improves focus, energy, and cognitive function. However, correlation does not imply causation, so this does not necessarily mean that more sleep directly causes increased productivity. Other factors, such as a healthy lifestyle or stress levels, may influence both sleep and productivity

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)?

The probability of seeing at least one car means, total probability minus the probability of not seeing any car.

$P(seeing\ car) = 30\% = 0.3$

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

$P(not\ seeing\ a\ car\ in\ 60\ minutes) = 0.7^3 = 0.343$ (since 60 minutes contains 3 times 20 minutes)

$P(seeing\ at\ least\ a\ car\ in\ 60\ minutes) = 1 - 0.343 = 0.657$

The probability that you see at least one supercar in the period of an hour = **0.657**

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

A sampling error is the difference between a sample statistic (eg: sample mean) and the true population parameter (eg: population mean) due to the fact that only a subset of the population is analyzed rather than the entire population. It occurs because samples are not perfect representations of the population.

Methods to reduce sampling error :

- Increase Sample Size

  Larger samples tend to better represent the population and reduce variability.

- Use Random Sampling

  A well-randomized sample reduces bias and ensures all population members have an equal chance of selection.

- Stratified Sampling

  Dividing the population into meaningful subgroups (strata) and taking random samples from each ensures better representation.

- Reduce Selection Bias

  Ensure the sample is chosen properly to avoid over-representation of specific groups.

- Use Appropriate Sampling Techniques

  Methods like systematic, cluster, and stratified sampling can be more effective than simple random sampling, depending on the population structure.

46. What is an inlier?

An inlier is a data point that fits well within the expected range of a dataset and follows the general trend of the data. Unlike outliers, which are extreme values that deviate significantly from the rest of the data, inliers do not stand out and are considered normal observations.

They follow the general pattern or trend of the dataset.They do not significantly affect statistical calculations like mean and standard deviation.

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

The sample size has a direct impact on the width of a confidence interval. As the sample size increases, the confidence interval becomes narrower, meaning the estimate is more precise.

- Larger sample size → Narrower confidence interval (higher precision).
- Smaller sample size → Wider confidence interval (less precision).
- A narrower confidence interval means the estimate is more reliable, while a wider confidence interval indicates more uncertainty.

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

Yes, two confidence intervals can have different widths while maintaining the same confidence level. The width of a confidence interval depends on factors like sample size and variability, but not solely on the confidence level.

For example let’s consider two cases with small sample size and large sample size, In Case 1, with a small sample size (n=30), the 95% confidence interval is [65, 75], which is wider due to the smaller sample. In Case 2, with a larger sample size (n=300), the 95% confidence interval is [68, 72], which is narrower due to the increased sample size. Although both intervals have the same confidence level (95%), their widths differ because of the difference in sample sizes.

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

The margin of error (ME) is a key component of a confidence interval(CI). The confidence interval is built around the sample estimate (eg: sample mean) and extends by the margin of error on both sides.

A confidence interval is given by:

$CI = \bar{X} \pm MI$

- The margin of error determines the width of the confidence interval.
- Larger margin of error → Wider confidence interval (less precision).
- Smaller margin of error → Narrower confidence interval (more precision).

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

In statistics, degrees of freedom (DF) refer to the number of independent values that can vary in a statistical calculation without violating a given constraint. It plays a crucial role in various statistical tests, such as t-tests, chi-square tests, and regression analysis.

$DF = Total\, number\, of\, observations - Number\, of\, constraints$

Example :

If we have a sample of 5 numbers and we know their mean, then only 4 of the numbers are free to vary. The 5th number is determined by the constraint that their average must remain fixed (5th number should be a value such that average of 5 numbers will be the mean)