End Course Summative Assignment

Applied statistics Interview Grind

Bhimanna Shegunashi

AlmaBetter

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.

Your solutions should be concise, well-structured, and effective in showcasing your problem-solving skills. In the video, use a dynamic approach to clarify the chosen questions, ensuring your explanations are easily comprehensible for a broad audience.

Note:

1. Make a copy of this document and write your answers.
2. Include the Video Link here in your document before submitting.

1. What is a vector in mathematics?

2. How is a vector different from a scalar?

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

4. How can vectors be multiplied by a scalar?

5. What is the magnitude of a vector?

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

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

8. What is a basis in linear algebra?

9. What is a linear transformation in linear algebra?

10. What is an eigenvector in linear algebra?

11. What is the gradient in machine learning?

12. What is backpropagation in machine learning?

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

14. How are partial derivatives used in machine learning?

15. What is probability theory?

16. What are the primary components of probability theory?

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

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

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

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

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

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

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

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

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

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

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

28. What is a joint probability distribution?

29. How do you calculate the joint probability distribution?

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

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

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

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

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

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

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

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

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

39. What is confidence interval estimation?

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

41. What is the difference between correlation and causation?

42. How is a confidence interval defined in statistics?

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

44. What is hypothesis testing in statistics?

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

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

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

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

49. How can sample size determination affect experiment design?

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

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

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

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

54. What are observational and experimental data in statistics?

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

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

57. What is Bessel’s correction?

58. What is kurtosis?

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

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

61. 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?

62. 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?

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

64. 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)?

65. What is the empirical rule in Statistics?

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

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

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

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

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

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

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

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

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

75. What is a Chi-Square test?

76. What is a t-test?

77. What is the ANOVA test?

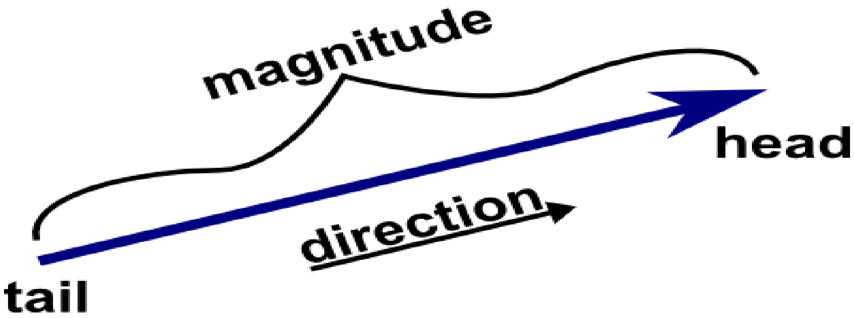
78. How is hypothesis testing utilised in A/B testing for marketing campaigns?

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

80. What is an inlier?

Solutions :-

1. **What is a vector in mathematics?**
   Ans. In mathematics, a vector is a quantity that has both magnitude and direction. It is often represented geometrically as an arrow, where the length of the arrow represents the magnitude of the vector, and the direction of the arrow represents the direction in which the quantity is pointing.



2. How is a vector different from a scalar?

Ans. A quantity that has magnitude but no particular direction is described as scalar. A quantity that has magnitude and acts in a particular direction is described as vector.

In summary, a scalar is a single numerical value representing magnitude, while a vector includes both magnitude and direction and is represented by a set of components. Scalars can be added, subtracted, and operated on using regular arithmetic, whereas vectors have specific rules for operations like addition, subtraction, and multiplication by scalars.

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

Ans. . There are many types of operations that can be performed on vectors.

- Addition of two vectors
- Subtraction of two vectors
- Multiplication of a vector with a scalar
- Product of two vectors
- Dot product
- Cross-product
- Unit vector
- Vector Projection
- Vector Norms

These operations are foundational for various data science tasks, including preprocessing, feature engineering, modelling, and analysis.

4. **How can vectors be multiplied by a scalar?**

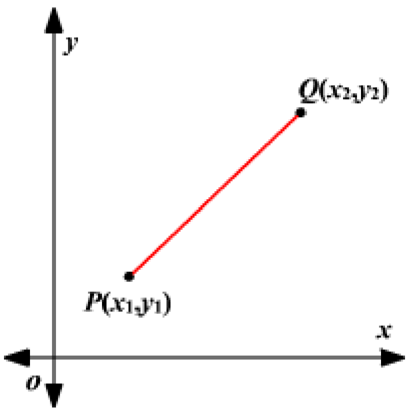
Ans. To multiply a vector by a scalar, we multiply the vector's magnitude by the scalar's magnitude. The result is a vector with the same direction as the original vector, but with a magnitude that is proportional to the scalar's magnitude.

If we multiply a vector by a positive scalar, the vector will go in the same positive direction. If we multiply a vector by a negative scalar, the vector will go in the same negative direction.

5. **What is the magnitude of a vector?**

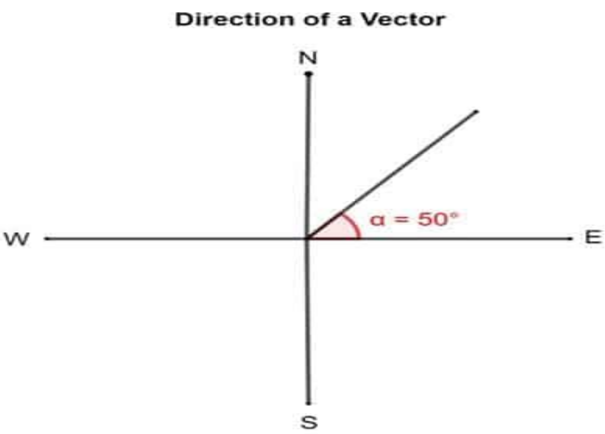
Ans. The magnitude of a vector is the distance between the initial point and the end point of the vector. The magnitude of a vector is denoted as ‖a‖.

The magnitude of a vector can be calculated using the Distance Formula. The formula is | →PQ |=√(x2−x1)**2+(y2−y1)**2



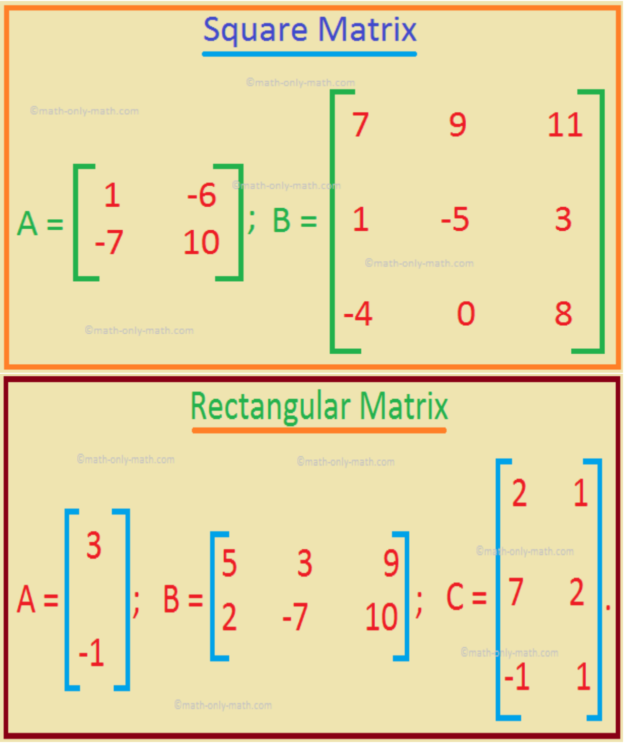
6. **How can the direction of a vector be determined?**

Ans. The direction of a vector in a two-dimensional plane is determined by the angle it makes with the positive x-axis. To find this angle, start from the positive x-axis (0 degrees) and move in an anticlockwise direction until reaching the vector. The angle can range from 0 to 360 degrees, representing all possible directions in the plane.



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

Ans. A square matrix is a matrix that contains the same number of rows and the same number of columns. If a matrix is not a square matrix, then it is known as a rectangular matrix. We can also say that the matrices which have different numbers of rows and columns are called rectangular matrices.



8. **What is a basis in linear algebra?**

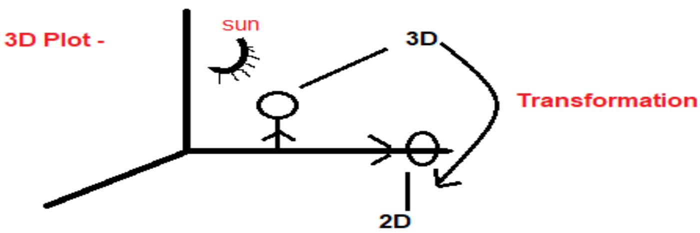
Ans. A basis is a set of linearly independent vectors that can be used to express any other vector in a given vector space.

9. **What is a linear transformation in linear algebra?**

Ans. A transformation is a technique in which a vector gets converted into another by keeping a unique element from each of the original vectors and assigning it into the resulting vector. This process maps one vector space into another.

Condition for linear transformation:-

1. T(u + v) = T(u) + T(v)
2. T(cu) = cT(u)
3. T(0) = 0



Note:- T= Transformation u,v = vector c = scalar value

10. What is an eigenvector in linear algebra?

Ans. An eigenvector is a vector that maintains its direction after undergoing a linear transformation, which means an eigenvector doesn't change its direction after transformation.

Formula for eigenvector:-

$$A.V = \lambda.V$$

Where, A = Matrix , V = eigenvector , λ = eigenvalue

11. What is the gradient in machine learning?

Ans. 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.

12. What is backpropagation in machine learning?

Ans. 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?

Ans. 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?

Ans. Partial derivatives are used in machine learning to update each weight independently. They are used to calculate the gradient of the error curve with respect to each weight in turn. Partial derivatives and gradient vectors are used in machine learning algorithms to find the minimum or maximum of a function. They are used in the training of neural networks, logistic regression, and many other classification and regression problems.

15. What is probability theory?

Ans. Probability theory is a field of mathematics and statistics that is concerned with finding the probabilities associated with random events. There are two main approaches available to study probability theory. These are theoretical probability and experimental probability.

Theoretical probability is determined on the basis of logical reasoning without conducting experiments. In contrast, experimental probability is determined on the basis of historic data by performing repeated experiments.

Probability formula:-

$$P(A) = \frac{\text{Number of favourable outcomes in A}}{\text{Total numbers of outcomes}}$$

16. What are the primary components of probability theory?

Ans. 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.

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

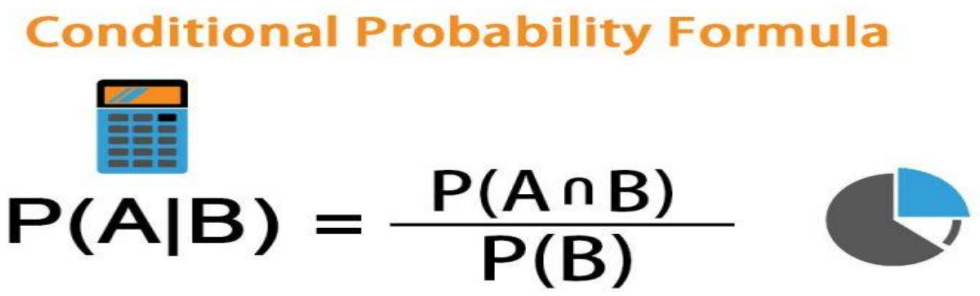
Ans. Conditional probability is the likelihood of an event occurring based on the occurrence of a previous event.

It is calculated by multiplying the probability of the preceding event by the updated probability of the succeeding event.

The formula for conditional probability is

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

In this formula, $P(A \cap B)$ represents the probability of both events A and B occurring simultaneously, and $P(B)$ represents the probability of event B occurring.



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

Ans. Bayes theorem is also known as the Bayes Rule or Bayes Law. It is used to determine the conditional probability of event A when event B has already happened. The general statement of Bayes' theorem is "The conditional probability of an event A, given the occurrence of another event B, is equal to the product of the event of B, given A and the probability of A divided by the probability of event B." i.e.

Formula for Bayes' Theorem:-

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

Where:

P(A) = The probability of A occurring.

P(B) = The probability of B occurring.

P(A|B) = The probability of A given B.

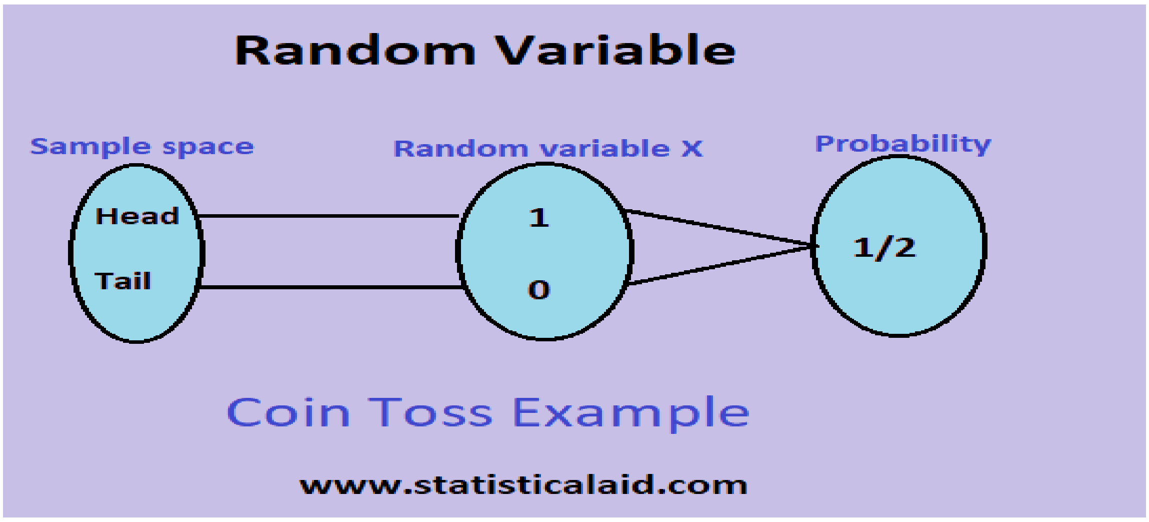
P(B|A) = The probability of B given A.

P(A∩B) = The probability of both A and B occurring.

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

Ans. Random variables are used to describe the possible outcomes of a random process. A random variable is a function that maps each outcome of a random process to a numerical value. For example, if we roll a die, the random variable can be defined as the number that appears on the top face.

The main difference between a random variable and a regular variable lies in their behaviour with respect to randomness and uncertainty. A random variable represents uncertain quantities that depend on random outcomes, while a regular variable represents known and fixed values.



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

Ans. The law of large numbers is a fundamental concept in probability theory that describes the relationship between the sample size of a random variable and its expected value. Here are some important points to understand about the law of large numbers:

1. The law of large numbers states that as the sample size of a random variable increases, the sample mean will approach the expected value of the variable. This means that the more data you have, the more accurate your estimate of the true underlying probability distribution will be.
2. This law applies to both discrete and continuous random variables, and it is a key concept in many areas of statistics and machine learning.
3. The law of large numbers is closely related to the central limit theorem, which states that the distribution of the sample means approaches a normal distribution as the sample size increases.
4. The law of large numbers has important implications for decision-making and risk management. It suggests that making decisions based on a large sample size is generally more reliable and accurate than making decisions based on a small sample size.
5. In practice, the law of large numbers is often used in simulations and statistical modelling to generate more accurate estimates of probabilities and other statistical measures. For example, if we want to estimate the probability of a rare event occurring, we can use the law of large numbers to simulate many trials and calculate the proportion of trials in which the event occurs.

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

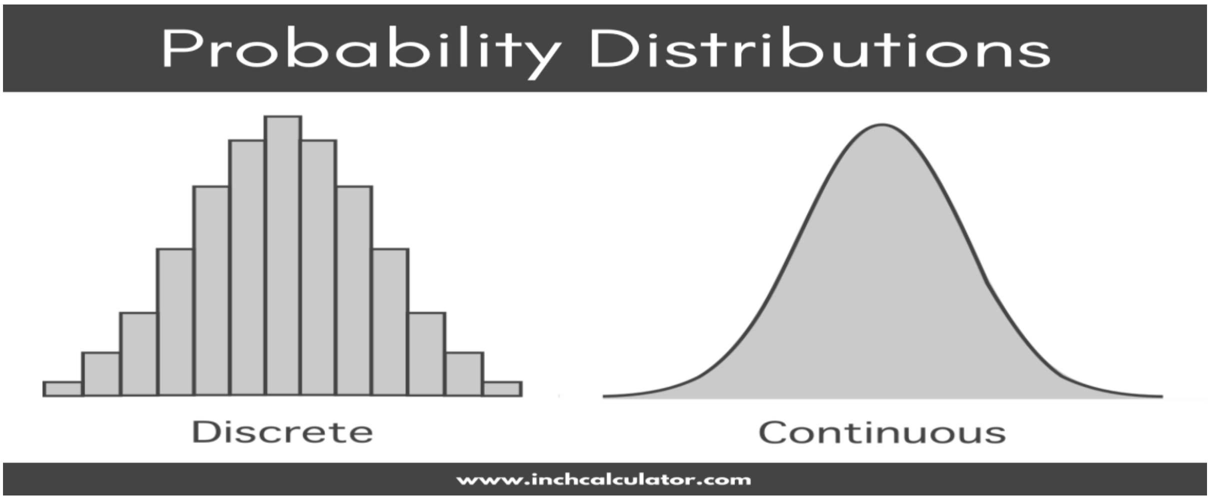
Ans. The central limit theorem is one of the most important theorems in probability theory and statistics. It states that, under certain conditions, the sum of a large number of independent and identically distributed random variables will be approximately normally distributed, regardless of the underlying distribution of the individual random variables. Here's an example of how the central limit theorem works in practice. Suppose we want to know the average height of all students in a school, but it's not practical to measure the height of every student. Instead, we can take a random sample of students and calculate the sample mean. By the central limit theorem, as the sample size increases, the distribution of the sample mean approaches a normal distribution, allowing us to make inferences about the population mean.

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

Ans. Random variables can be classified into two types: discrete and continuous. A discrete random variable can take on a countable number of values, while a continuous random variable can take on any value in a continuous range. Examples of discrete random variables include the number of heads in multiple coin flips, while an example of a continuous random variable is the height of a randomly selected person.

A **discrete probability distribution** is one that models systems with finite or countable outcomes. The data can only take on certain values, such as integers. A common example of a discrete distribution is the **binomial distribution**.

A **continuous probability distribution** is one that models systems with infinite possible values within a range. The data can take on any value within the specified range. A common example of a continuous distribution is the **normal distribution**.



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

Ans. There are three most common measures of central tendency mean, median, and mode.

Mean: It is the arithmetic average of a dataset and is calculated by summing all values in the dataset and dividing by the number of observations. Mean can be sensitive to outliers and extreme values in a dataset. The formula for calculating the mean is:

mean = (sum of all values) / (number of observations)

Median: It is the middle value in a dataset when the values are arranged in ascending or descending order. Median is less sensitive to outliers compared to mean. In case of an even number of observations, the median is calculated as the average of the two middle values.

Mode: It is the most frequently occurring value in a dataset. Mode can be used for both numerical and categorical data. A dataset can have one or more modes, or it may have no mode at all.

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

Ans. Percentiles and quartiles are important measures of relative standing or position of a value in a data set. Percentiles divide a dataset into 100 equal parts, each representing 1% of the data. For example, the 75th percentile is the value below which 75% of the data falls.

Quartiles divide a dataset into four equal parts, each representing 25% of the data. The first quartile (Q1) is the 25th percentile, the second quartile (Q2) is the 50th percentile (also known as the median), and the third quartile (Q3) is the 75th percentile.

Percentiles and quartiles can be calculated using various methods, such as interpolation, nearest rank, and the inverse of the cumulative distribution function. The most commonly used method for calculating percentiles and quartiles is the interpolation method. This method estimates the percentile value by interpolating between the two nearest values in the dataset.

In Python, we can use the NumPy library to calculate percentiles and quartiles. The percentile() function can be used to calculate any percentile value, and the quantile() function can be used to calculate quartiles.

Quartile Formula

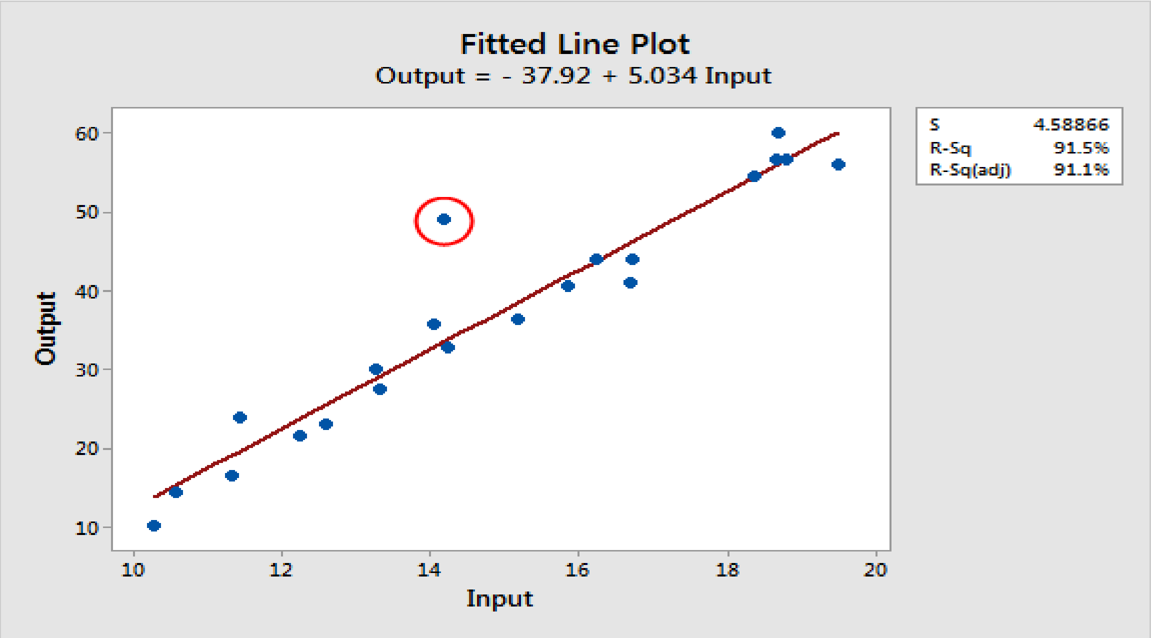
The Quartile Formula For Q1 $= \frac{1}{4}(n+1)^{th}$ term

The Quartile Formula For Q3 $= \frac{3}{4}(n+1)^{th}$ term

The Quartile Formula For Q2 $= Q3 - Q1$ (Equivalent to Median)

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

Ans. Outliers are data points that are significantly different from other data points in a dataset. Outliers can significantly affect the results of data analysis, so it is important to detect and treat them appropriately.



Here are some important points to consider for outlier detection and treatment:

1. Identify outliers: Use statistical methods such as box plots, scatter plots, or z-scores to identify potential outliers in your dataset.

2. Determine the cause: Investigate the cause of outliers, whether it is due to data entry errors or a legitimate deviation from the norm.

3. Decide on treatment: Decide whether to remove outliers or to treat them. Depending on the cause and impact of outliers, different treatments can be applied, such as replacing the outlier with a more appropriate value, removing the outlier entirely, or keeping the outlier in the analysis but using robust statistical methods.

4. Apply treatment: Apply the chosen treatment method to the outliers.

5. Re-evaluate the dataset: After treating the outliers, re-evaluate the dataset to ensure that the outliers no longer significantly impact the results of the analysis.

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

Ans. The central limit theorem states that as the sample size increases, the distribution of the sample mean approaches a normal distribution. This can be used to approximate the distribution of a discrete probability distribution with a normal distribution.

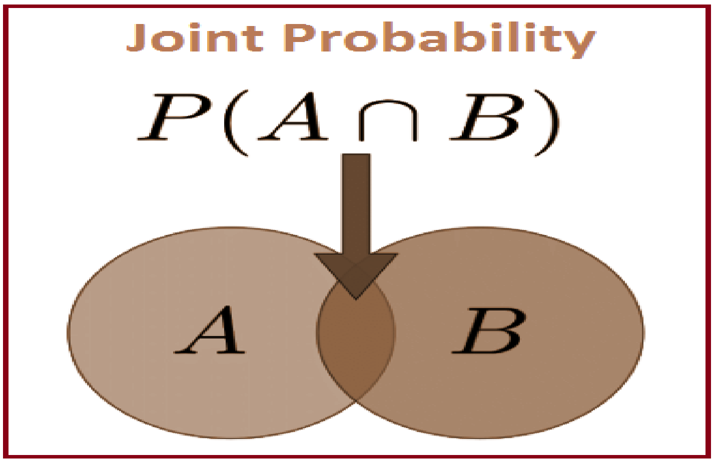
27. How do you test the goodness of fit of a discrete probability distribution?

Ans. Discrete Probability Distributions provide a way to model and analyse random phenomena that can take on only a finite or countable number of possible outcomes. Examples of discrete probability distributions include the Binomial distribution, Poisson distribution, Geometric distribution, and Hypergeometric distribution.

These distributions can be used to calculate probabilities of certain events occurring, such as the probability of a certain number of successes in a fixed number of trials (Binomial) or the probability of a certain number of arrivals in a given time interval (Poisson).It is important to choose the appropriate distribution for the situation, as using the wrong distribution can lead to incorrect results. In order to use these distributions in practice, we often need to calculate expected values, variances, and standard deviations. This can be done using formulas specific to each distribution.

28. What is a joint probability distribution?

Ans. A joint probability distribution represents a probability distribution for two or more random variables. Instead of events being labelled A and B, the condition is to use X and Y as given below. f(x,y) = P(X = x, Y = y) The main purpose of this is to look for a relationship between two variables.



29. How do you calculate the joint probability distribution?

Ans. Examples of how to calculate the joint probability distribution.

Number of possible outcomes when a die is rolled = 6

i.e. {1, 2, 3, 4, 5, 6}

Let A be the event of occurring 3 on the first die and B be the event of occurring 3 on the second die.

Both the dice have six possible outcomes, the probability of a three occurring on each die is 1/6.

P (A⋂B)

where,

A, B= Two events

P(A and B),P(AB)=The joint probability of A and B

P(A) =1/6

P(B)=1/6

P(A,B) = 1/6 x 1/6 = 1/36

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

Ans. A joint probability distribution is like a table that shows the chances of two events happening at the same time. For example, think of rolling two dice - it's the probability of getting a specific number on each die simultaneously.

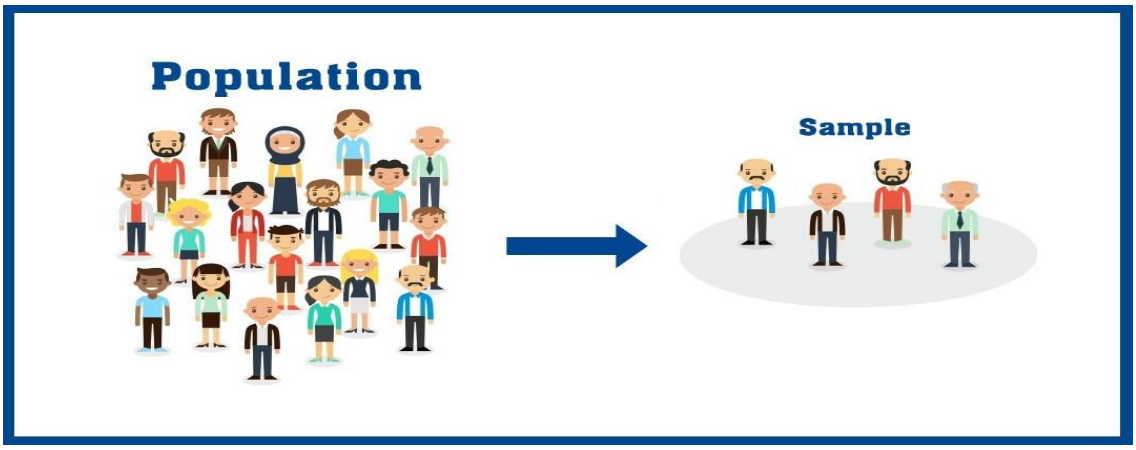
Marginal probability is just looking at one die at a time. It's the chance of getting a certain number on one die, no matter what the other die shows. For example, the chance of rolling a 3 on one die.

This table not only tells us about the chances of each combination but also helps us figure out the probability of each die separately. If we look at all the combinations where, say, the first die is 1, and add up those probabilities, we get the chance of rolling a 1 on the first die (that's the "marginal distribution" for the first die). Same goes for the second die.

In summary, a joint probability distribution gives us the combined probabilities, but it's also like a treasure map that helps us find the individual probabilities for each event (marginal probabilities). It's super useful in understanding the relationship between different events.

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

Ans. Sampling refers to the process of selecting a subset, known as a sample, from a larger group, known as a population. The sample is carefully chosen to be representative of the population's characteristics, allowing researchers to draw meaningful conclusions about the entire population based on the analysis of the sample data. Sampling involves techniques that aim to minimize bias and ensure the sample accurately reflects the diversity and variability present in the population.



Importance:

Time and Cost Efficiency: Collecting data from an entire population can be impractical, time-consuming, and expensive. Sampling allows researchers to obtain accurate information using a smaller budget and within a reasonable timeframe.

Inference: Statistical analysis of a well-selected sample can provide insights and draw conclusions about the entire population. This is the basis of inferential statistics, which generalize findings from the sample to the larger population.

Feasibility: In cases where the population is too large or inaccessible, sampling provides a feasible way to study and understand the population's characteristics.

Reduced Data Collection Effort: Instead of gathering data from all individuals, researchers can focus on collecting data from a smaller group, making data collection more manageable.

Risk Reduction: Sampling allows researchers to evaluate hypotheses and test new ideas on a smaller scale before implementing them on the entire population, reducing potential risks and errors.

Ethics: In cases where it is not feasible or ethical to collect data from every individual, sampling provides a way to gather relevant information without invading privacy or causing harm.

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

Ans. There are four types of sampling methods commonly used in statistical inference.

1. Random Sample: In a random sample, every member of the population has an equal chance of being selected. Our sampling frame should include the whole population. To conduct this type of sampling, we can use tools like random number generators or other techniques that are based entirely on chance.

2.Systematic sampling: Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

3.Stratified sampling: Stratified sampling involves dividing the population into subpopulations that may differ in important ways. It allows us to draw more precise conclusions by ensuring that every subgroup is properly represented in the sample.

To use this sampling method, we divide the population into subgroups based on the relevant characteristic.

Based on the overall proportions of the population, we calculate how many people should be sampled from each subgroup. Then we use random or Systematic sampling to select a sample from each subgroup.

4.Cluster sampling: Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, we randomly select entire subgroups.

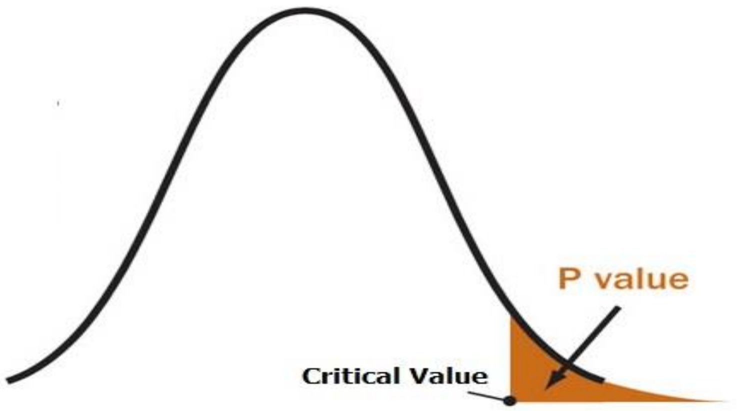
If it is practically possible, we might include every individual from each sampled cluster. If the clusters themselves are large, we can also sample individuals from within each cluster using one of the techniques above.

This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It’s difficult to guarantee that the sampled clusters are really representative of the whole population.

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

Ans. The p-value is the probability of obtaining a result equal to or more extreme than what was observed, assuming the null hypothesis is true. The p-value measures how likely it is that any observed difference between groups is due to chance. The lower the p-value, the greater the statistical significance of the observed difference.

The p-value is calculated by calculating the likelihood of the test statistic, which is the number calculated by a statistical test using your data. The p-value is the smallest level of significance that would lead to rejection of the null hypothesis.



34. What is confidence interval estimation?

Ans. In statistics, a confidence interval is a range of values that contains the true value of an estimated value. The confidence interval is based on the estimated value and the investigator's desired level of confidence.

The confidence interval is calculated from an estimate of how far away the sample mean is from the actual population mean. The confidence interval conveys how precise the measurement is.

The confidence interval is based on:

- The point estimate, e.g., the sample mean
- The investigator's desired level of confidence

The confidence level is often expressed as a percent. Common levels are: 90, 95, 98, 99.

For example, if an estimated value is 50 and the confidence interval of 80% is ±5%, then there is an 80% probability that the true value is between 45 and 55.

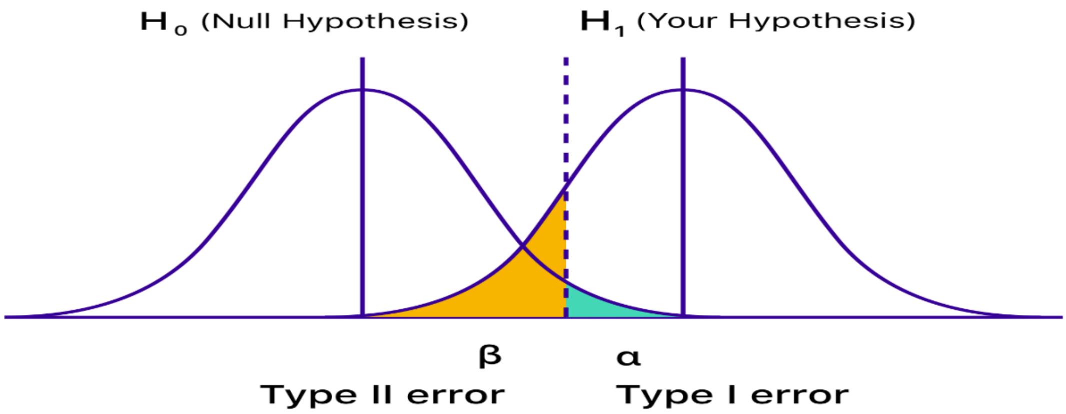
35. What are Type I and Type II errors in hypothesis testing?

Ans.

Type I Error: A Type I error is the mistake of rejecting the null hypothesis when it is true.

The symbol α (alpha) is used to represent the probability of a type I error.

Type II Error: A Type II error is the mistake of failing to reject the null hypothesis when it is false. The symbol β (beta) is used to represent the probability of a type II error.



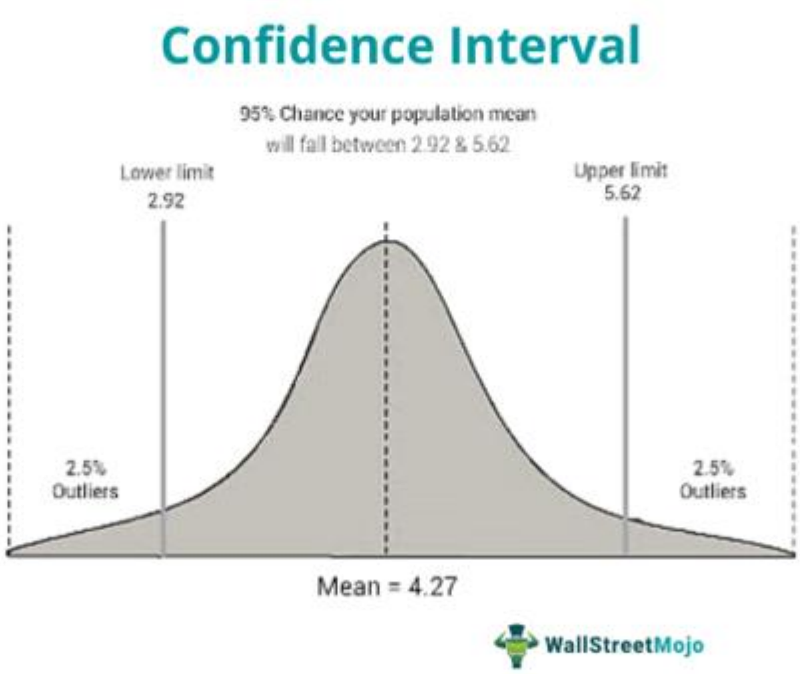
36. How is a confidence interval defined in statistics?

Ans. Confidence level is the probability that the true population parameter falls within the interval estimate of the sample statistic. It is important to choose an appropriate confidence level for an interval estimate because it determines the range of values that we can be reasonably confident contains the true population parameter.

The most common confidence levels used in statistics are 90%, 95%, and 99%. To choose an appropriate confidence level, we need to consider the desired level of accuracy and the potential consequences of making a Type I or Type II error

The formula for a confidence interval is:

CI = point estimate ± margin of error



37. What is hypothesis testing in statistics?

Ans. Hypothesis Testing is a type of statistical analysis in which you put your assumptions about a population parameter to the test. It is used to estimate the relationship between 2 statistical variables.

examples of statistical hypothesis :

- A teacher assumes that 60% of his college's students come from lower-middle-class families.
- A doctor believes that 3D (Diet, Dose, and Discipline) is 90% effective for diabetic patients.

Hypothesis Testing Formula

$Z = (\overline{x} – \mu0) / (\sigma /\sqrt{n})$

- Here, $\overline{x}$ is the sample mean,
- $\mu0$ is the population mean,
- σ is the standard deviation,
- n is the sample size.

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

Ans. The null hypothesis is a kind of hypothesis which explains the population parameter whose purpose is to test the validity of the given experimental data. This hypothesis is either rejected or not rejected based on the viability of the given population or sample.

The Null Hypothesis is the assumption that the event will not occur. A null hypothesis has no bearing on the study's outcome unless it is rejected.

H0 is the symbol for it, and it is pronounced H-naught.

The Alternate Hypothesis is the logical opposite of the null hypothesis. The acceptance of the alternative hypothesis follows the rejection of the null hypothesis. H1 is the symbol for it.

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

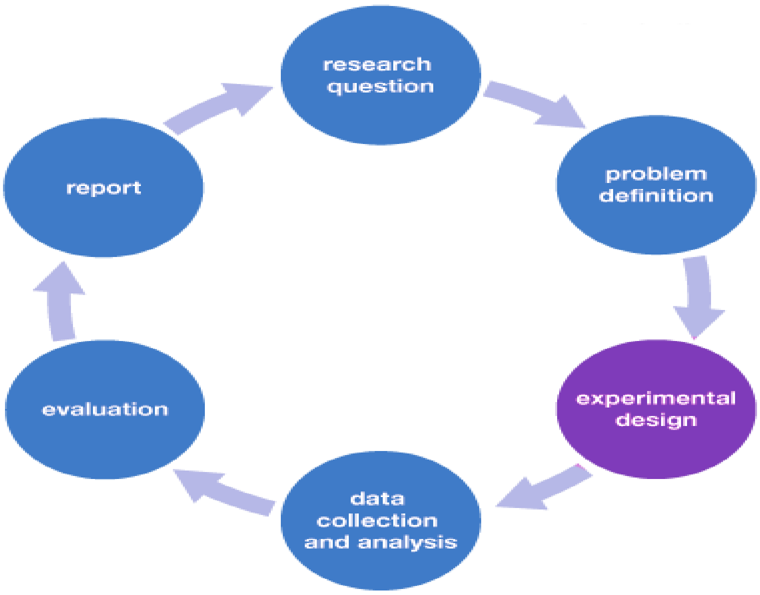
Ans. Difference between one-tailed and two-tailed tests.

One-Tailed	Two-Tailed
A test of any statistical hypothesis, where the alternative hypothesis is one-tailed either right-tailed or left-tailed.	A test of a statistical hypothesis, where the alternative hypothesis is two-tailed
For one-tailed, we use either > or < sign for the alternative hypothesis.	For two-tailed, we use ≠ sign for the alternative hypothesis.
When the alternative hypothesis specifies a direction then we use a one-tailed test.	If no direction is given then we will use a two-tailed test.
Critical region lies entirely on either the right side or left side of the sampling distribution.	Critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic
Here, the Entire level of significance (α) i.e. 5% has either in the left tail or right tail.	It splits the level of significance (α) into half.
Rejection region is either from the left side or right side of the sampling distribution.	Rejection region is from both sides i.e. left and right of the sampling distribution.
It checks the relation between the variables in a single direction.	It checks the relation between the variables in any direction.
It is used to check whether one mean is different from another mean or not.	It is used to check whether the two mean different from one another or not.

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

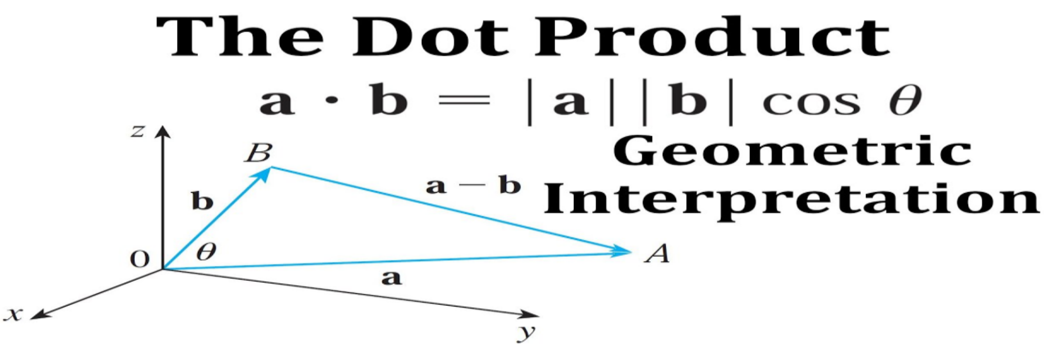
Ans. Experimental design in statistics refers to the process of planning and conducting experiments to answer research questions or test hypotheses.

Experimental design is important, especially when trying to avoid bias. We need to set up our study in a way that it will result in not only interpretable outcomes but also useful ones. The latter can then be taken out for data analysis.



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

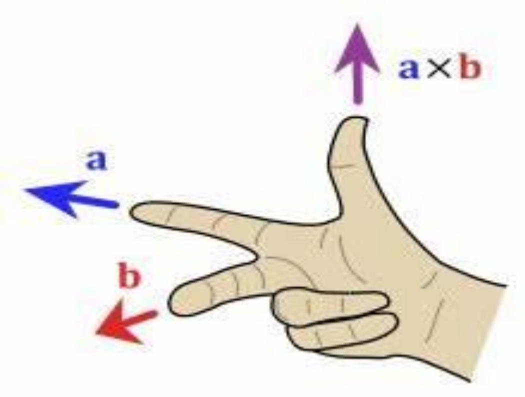
Ans. The dot product of two vectors is the product of their magnitudes and the cosecant of the angle between them. The dot product is written as a.b=|a||b|cosθ. The geometric interpretation of the dot product is the length of the projection of vector a onto the unit vector b.



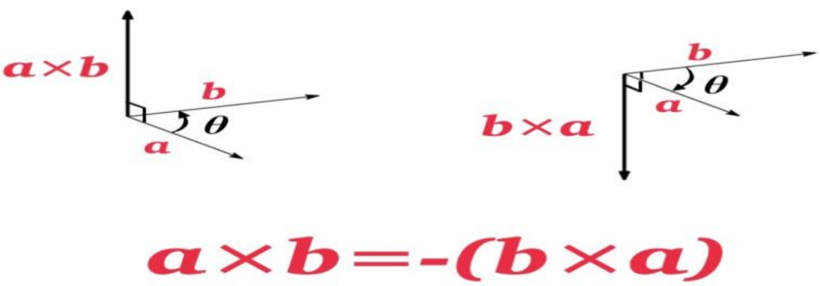
The dot product represents how close two vectors are to pointing in the same direction. The dot product is zero if the vectors are perpendicular. It is increasingly positive if the vectors are going in the same direction, and increasingly negative if they are pointing in opposite directions.

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

Ans. The cross product of two vectors is a vector that is perpendicular to both vectors. The magnitude of the cross product is the area of the parallelogram that the vectors span. The direction of the cross product is perpendicular to both vectors. The exact direction is determined by the right-hand rule.



The right-hand rule states that you should rotate your right hand's fingers from the side of the vector that comes first in the product to the side of the vector that comes second in the product. The direction of the extended thumb will be the direction of the product vector.



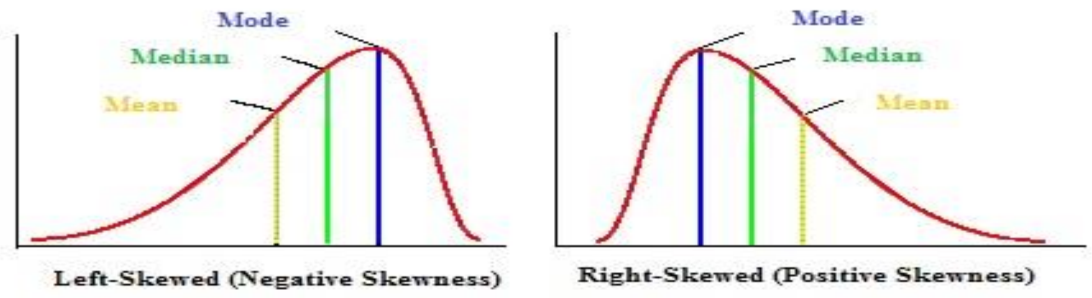
43. What is the left-skewed distribution and the right-skewed distribution?

Ans. left-skewed distribution: When the distribution has a long tail towards the left side, then it is known as a left-skewed or negative-skewed distribution. In the negative-skewed distribution, the concentration of data points towards the left tail is more than the right tail.

In the left-skewed distribution: Mode > Median > Mean

Right-skewed distribution: When the distribution has a long tail towards the right side, then it is known as a right-skewed or positive-skewed distribution. In the right-skewed distribution, the concentration of data points towards the right tail is more than the left tail.

In the right-skewed distribution: Mean > Median> Mode.



44. What is Bessel's correction?

Ans. In statistics, Bessel's correction is the use of n-1 instead of n in the formula for the sample variance and sample standard deviation. This method corrects the bias in the estimation of the population variance.

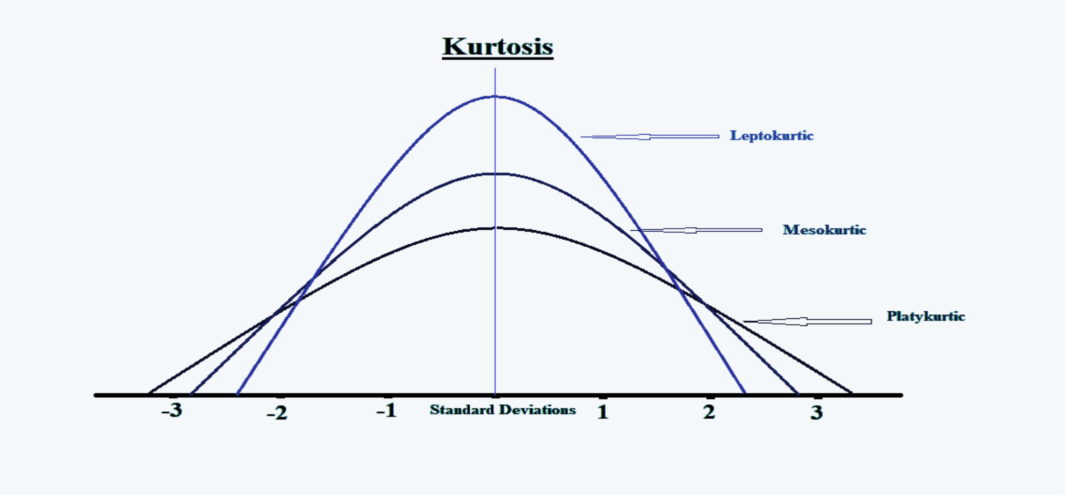
The 1/n variance formula is systematically biased. In nearly every case, it gives a lower estimate than we would make if we had the population mean available.

Bessel's correction is used because the mean of samples has an error or bias from the true mean. Dividing by (n-1) gives a better estimate.

As our sample size increases, the difference gets smaller. This is because our sample is more likely to capture the extreme values in the population and therefore our sample variance will approach the population variance.

45. What is kurtosis?

Ans. Kurtosis is a statistical measure used to describe a characteristic of a dataset. When normally distributed data is plotted on a graph, it generally takes the form of a bell. This is called the bell curve. The plotted data that are furthest from the mean of the data usually form the tails on each side of the curve. Kurtosis indicates how much data resides in the tails.



Distributions with a large kurtosis have more tail data than normally distributed data, which appears to bring the tails in toward the mean. Distributions with low kurtosis have fewer tail data, which appears to push the tails of the bell curve away from the mean.

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

Solution. Formula = Possible outcomes / Total outcomes

Calculation,

When two dice are rolled,

Total outcomes = 36

Possible outcome of getting 5 = (1,4), (4,1), (2,3) , (3,2)

⇒Number of possible outcome of getting 5 = 4

Possible outcome of getting 8 = (2,6), (6,2), (3,5), (5,3), (4,4)

⇒Number of possible outcome of getting 8 = 5

⇒Total number of possible outcomes = 4 + 5 = 9

Probability = Possible outcomes / Total outcomes

⇒probability = 9 /36 = 1 / 4

Ans. The probability of throwing two fair dice when the sum is 5 and 8 is 1/4.

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

Ans. The main difference between descriptive and inferential statistics is descriptive statistics summarize and describe data, while inferential statistics draw conclusions and make predictions about populations based on sample data.

Difference between Descriptive and Inferential statistics:

Sr. No.	Descriptive Statistics	Inferential Statistics
1	It gives information about raw data which describes the data in some manner.	It makes inferences about the population using data drawn from the population.
2	It helps in organizing, analysing, and to present data in a meaningful manner.	It allows us to compare data, and make hypotheses and predictions.
3	It is used to describe a situation.	It is used to explain the chance of occurrence of an event.
4	It explains already known data and is limited to a sample or population having a small size.	It attempts to reach the conclusion about the population.
5	It can be achieved with the help of charts, graphs, tables, etc.	It can be achieved by probability.

48. 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?

Solution.

$$\chi = \boldsymbol{\mu} + \boldsymbol{Z\sigma}$$

Here,

μ = Mean

Z = 1.20

σ = Standard deviation

X = Value to be calculated

Therefore, X = 160 + (1.20 * 15) = 173.8

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

Ans. The degree of freedom (DOF) is a term that statisticians use to describe the degree of independence in statistical data. A degree of freedom can be thought of as the number of variables that are free to vary, given one or more constraints. When you have one degree, there is one variable that can be freely changed without affecting the value for any other variable. As a data scientist, it is important to understand the concept of degree of freedom, as it can help you do accurate statistical analysis and validate the results.

The formula to determine degree of freedom :

D = N - 1

Where,

D = Degree of freedom

N = Sample of size

50. 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)?

Solution.

The probability of not seeing a supercar in 20 minutes is:

= 1 - P(Seeing one supercar) = 1 - 0.3

= 0.7

Probability of not seeing any supercar in 60 minutes is:

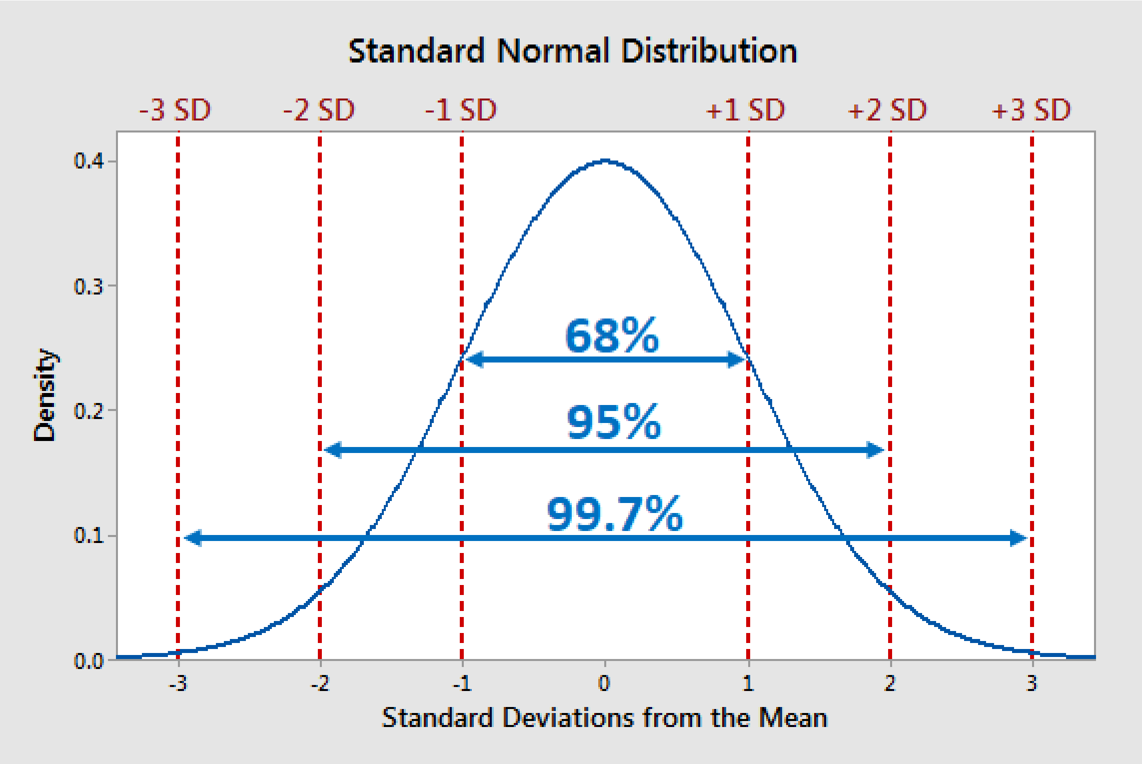
= (0.7)**3 = 0.343

Hence, the probability of seeing at least one supercar in 60 minutes:

= 1 - P(Not Seeing any supercar) = 1 - 0.343 = **0.657**

51. What is the empirical rule in Statistics?

Ans. The empirical rule, also sometimes called the three-sigma or 68-95-99.7 rule, is a statistical rule which states that for normally distributed data, almost all observed data will fall within three standard deviations (denoted by the Greek letter sigma, or σ) of the mean or average (represented by the Greek letter mu, or μ) of the data.



In particular, the empirical rule predicts that in normal distributions, 68% of observations fall within the first standard deviation ($\mu \pm \sigma$), 95% within the first two standard deviations ($\mu \pm 2\sigma$), and 99.7% within the first three standard deviations ($\mu \pm 3\sigma$) of the mean.

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

Ans. The width of a confidence interval (CI) in statistics is affected by several factors:

Sample Size (n): As the sample size increases, the width of the confidence interval decreases, holding all other factors constant. Larger sample sizes provide more precise estimates of the population parameter.

Level of Confidence: Higher levels of confidence (95% confidence) lead to wider confidence intervals because a higher level of confidence requires a larger range to capture the true population parameter with that level of certainty.

Standard Deviation (σ): A larger standard deviation in the population results in a wider confidence interval because the data points are more spread out from the mean, requiring a wider range to achieve a certain level of confidence.

Sampling Distribution: The shape of the sampling distribution and the distribution of the population affect the width of the confidence interval. For example, if the population is highly skewed or has heavy tails, the confidence interval may be wider.

Estimator's Variability: The variability of the estimator being used (sample mean or proportion) affects the width of the confidence interval. Less variability in the estimator leads to a narrower interval.

Choice of Estimator: The specific estimator used to estimate the population parameter can impact the width of the confidence interval. Different estimators may produce narrower or wider intervals based on their properties.

Population Size: In some cases, particularly in finite populations, the size of the population can impact the width of the confidence interval. For larger populations, the effect may be negligible.

Assumptions and Conditions: Assumptions made about the data and the statistical model being used can influence the width of the confidence interval. Violations of assumptions might widen the interval.

Understanding these factors helps in selecting an appropriate sample size, level of confidence, and statistical approach to construct a confidence interval that meets the desired precision and reliability for a given study or analysis.

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

Ans. A sampling error occurs when the sample used in the study does not represent the entire population. Although sampling errors occur frequently, researchers always include a margin of error in their conclusions as a matter of statistical practice.

The margin of error is the amount allowed for a miscalculation to represent the difference between the sample and the actual population.

Sampling is a type of analysis where a small sample of observations is chosen from a larger population. The selection bias process can produce both sampling errors and non-sampling errors.

Steps to reduce sampling errors:

- **Increase sample size**
  A larger sample size is more accurate because the study gets closer to the actual

  population size.
- **Divide the population into groups**

Test groups according to their size in the population instead of a random sample. For example, if people of a specific demographic make up 20% of the population, make sure that your study is made up of this variable to reduce sampling bias.

- **Know your population**
  Study your population and understand its demographic mix. Know what demographics use your product and service and ensure you only target the sample that matters.

54. What is a Chi-Square test?

Ans. A chi-squared test (symbolically represented as χ2) is basically a data analysis on the basis of observations of a random set of variables. Usually, it is a comparison of two statistical data sets. This test was introduced by Karl Pearson in 1900 for categorical data analysis and distribution. So it was mentioned as Pearson’s chi-squared test.

The chi-square test is used to estimate how likely the observations that are made would be, by considering the assumption of the null hypothesis as true.

A hypothesis is a consideration that a given condition or statement might be true, which we can test afterwards. Chi-squared tests are usually created from a sum of squared falsities or errors over the sample variance.

The formula for chi-square can be written as:

$$X^2 = \sum \frac{(\text{Observed value - Expected value})^2}{\text{Expected value}}$$

55. What is a t-test?

Ans. A t-test is a statistical test that compares the means of two groups. It's often used in hypothesis testing to determine whether a process or treatment has an effect on a population.

T-tests are used when data sets follow a normal distribution and have unknown variances. For example, we use a t-test to determine if the mean of the length of petals of a flower belonging to two different species is the same.

t-tests are used to:

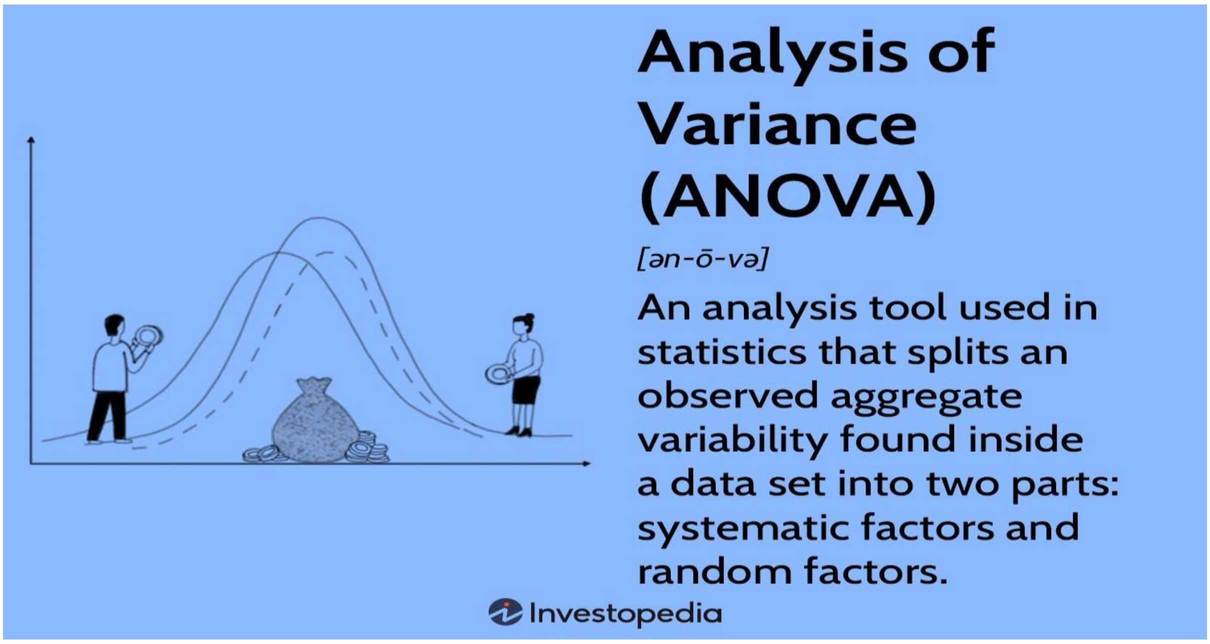
- Determine if there is a significant difference between the means of two groups
- Test the difference in a mean from a hypothesized value
- Test the difference in means between two normally distributed populations with equal variances

T-tests are one of the most widely used statistical hypothesis tests in pain studies.

56. What is the ANOVA test?

Ans. ANOVA, or Analysis of Variance, is a statistical technique used to analyze the differences among group means in a sample. It's a parametric hypothesis test that assesses whether there are statistically significant differences between the means of three or more independent groups. ANOVA compares the variability within each group (within-group variance) to the variability between the groups (between-group variance).

There are several types of ANOVA, including one-way ANOVA (comparing one factor), two-way ANOVA (comparing two factors), and more complex designs involving multiple factors and interactions.



57. What is an inlier?

Ans. An inlier, also known as an inlying value or an inlying point, is a data point or observation that fits well within a given dataset or model. Inliers are typically consistent with the majority of the data and are not considered outliers.

In the context of outlier detection and data analysis, an inlier is a point that is considered to be part of the "normal" or "typical" behaviour of the data. It aligns with the general pattern or trend exhibited by the majority of the dataset. Inliers contribute to a better understanding of the central tendencies and relationships within the data.

Github Link - https://github.com/BhimuShegunashi/Applied-statistics-Interview-Grind.git

Video Presentation Link-
https://drive.google.com/file/d/1E5xGf_5XyGrrbaKxpcT21iNMBCtsghFB/view?usp=sharing

PPT Presentation Link -
https://docs.google.com/presentation/d/197HJvtpRp87j4LhPd9xwyzNLrffR5PBf/edit?usp=sharing&ouid=102572852756707196661&rtpof=true&sd=true