

### **1. What is business analytics?**

Business analytics is the systematic process of examining and interpreting data to derive valuable insights that aid in making informed business decisions. It involves collecting and preparing data, performing statistical analysis, and employing various data modeling techniques. By identifying patterns, trends, and relationships within data, organizations can optimize operations, enhance strategic planning, and improve overall performance.

### **2. Can you name a trend in business analytics?**

One notable trend in business analytics is the increasing use of artificial intelligence (AI) and machine learning. Companies are adopting AI-powered analytics tools to analyze data more efficiently and make predictions with greater accuracy. These technologies help in automating repetitive tasks, such as data processing and recognition, and enable organizations to gain deeper insights from their data.

### **3. Define Big Data Analytics.**

Big Data Analytics refers to the process of analyzing vast and complex datasets that are too large and diverse to be handled by traditional data analysis tools. It involves using specialized techniques and technologies to extract valuable insights, patterns, and trends from these massive datasets.

### **4. Explain descriptive statistics briefly.**

Descriptive statistics is a branch of statistics that involves summarizing and presenting data in a meaningful way. It provides a snapshot of the main characteristics of a dataset, such as its central tendency (mean, median, mode), spread (range, variance, standard deviation), and distribution (normal, skewed).

**OR**

Descriptive statistics summarize data using measures like mean and standard deviation to provide insights into its characteristics.

### **5. What is the purpose of inferential statistics?**

The purpose of inferential statistics is to make predictions or draw conclusions about a larger population based on a smaller sample of data. It helps us generalize findings from sample data to the broader group or population it represents.

**6. Name a measure of central tendency.**

One common measure of central tendency is the "mean," which is also known as the average. To calculate the mean, you add up all the values in a dataset and then divide by the number of values. It represents the central or typical value in a set of data, helping to understand the center point around which the data tends to cluster.

**7. Can you briefly explain the Central Limit Theorem?**

The Central Limit Theorem (CLT) states that when you take many random samples from a population and calculate the means of those samples, the distribution of those sample means will be approximately normal, regardless of the population's original distribution.

**8. What is the Chi-square test used for?**

The Chi-square test is used to assess whether there is a significant association or relationship between two categorical variables in a dataset. It helps determine if the observed frequencies of different categories are significantly different from what we would expect by chance.

**9. Define Maximum Likelihood.**

The maximum likelihood estimation is a method that determines values for parameters of the model. It is the statistical method of estimating the parameters of the probability distribution by maximizing the likelihood function. The point in which the parameter value that maximizes the likelihood function is called the maximum likelihood estimate.

**10. What is Hypothesis Testing**

Hypothesis testing is a statistical method that is used in making a statistical decision using

experimental data Hypothesis testing is basically an assumption that we make about

population parameter. It evaluates two mutually exclusive statements about a population to

determine which statement is best supported by the sample data.

**11. Explain Bayes' Theorem briefly.**

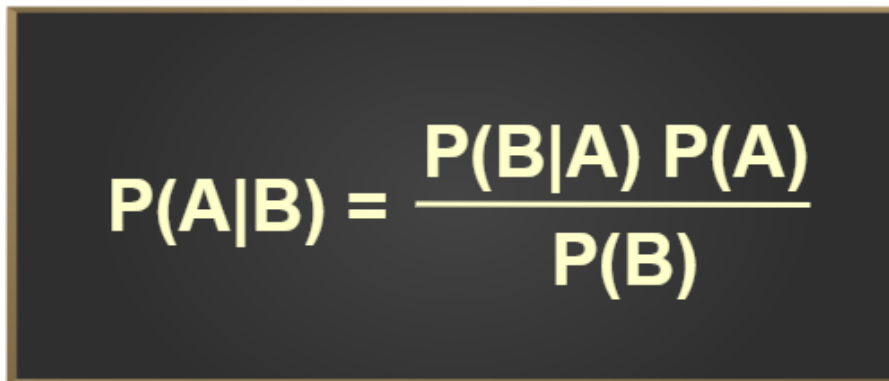
It is used to determine the conditional probability of event A when event B has already

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.

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


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

## 12. What is Conditional Probability?

The probability of an event A based on the occurrence of another event B is termed conditional Probability. It is denoted as  **$P(A|B)$**  and represents the probability of A when event B has already happened.

## 13. What is Joint Probability?

When the probability of two more events occurring together and at the same time is measured it is marked as Joint Probability. For two events A and B, it is denoted by joint probability is denoted as,  **$P(A \cap B)$** .

## 14. What is Uniform Distribution?

In statistics, uniform distribution is a term used to describe a form of probability distribution where every possible outcome has an equal likelihood of happening. The probability is constant since each variable has equal chances of being the outcome.

**15.What does "population" mean in statistics?**

In statistics, population refers to the entire group or set of individuals or items being studied or analyzed.

**16.What is Binomial Distribution?**

A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times. The binomial is a type of distribution that has two possible outcomes (the prefix “bi” means two, or twice).

For example, a coin toss has only two possible outcomes: heads or tails and taking a test could

have two possible outcomes: pass or fail.

**17.What is a binomial distribution used for?**

A binomial distribution models the number of successes in a fixed number of independent and identical trials, like coin flips.

**18.Provide an example of an exponential distribution.**

An example of an exponential distribution is the time between arrivals of customers at a retail store or the time between successive phone calls at a call center.

This distribution is often used to model events where occurrences are random, and there is a constant probability of an event happening within a certain time interval.

**19.Explain the concept of log-normal distribution.**

A log-normal distribution describes data that becomes approximately normally distributed

after taking the natural logarithm of the values. It's common in finance.

**20.What do sampling techniques involve?**

Sampling techniques select a subset of data from a larger population to make inferences and

draw conclusions about the entire population.

**21.How are probability theories used?**

Probability theories are used to model and analyze random events and uncertain phenomena

in various fields, including finance, engineering, insurance, and healthcare. They provide a

mathematical framework for calculating probabilities and making informed decisions based

on the available data.

## **22.How is the mean calculated?**

The mean is calculated by summing all values in a dataset and dividing by the total number

of values, providing a measure of central tendency.

## **23.In hypothesis testing, what is a p-value?**

In hypothesis testing, the p-value measures the strength of evidence against a null hypothesis.

A smaller p-value indicates stronger evidence against the null hypothesis.

## **24.What is sampling?**

Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate characteristics of the whole population.

## **25.What's the purpose of Maximum Likelihood estimation?**

Maximum Likelihood estimation aims to find the parameter values that maximize the

likelihood of observing the given data, helping estimate population parameters.

## **26.Can you provide a practical application of Bayes' Theorem?**

Bayes' Theorem is used in spam email detection to calculate the probability that an incoming

email is spam based on previous email characteristics.

## **27. When is the Central Limit Theorem relevant in real-world scenarios?**

The Central Limit Theorem is relevant when estimating characteristics of a large population

by analyzing random samples, ensuring that the sample means follow a normal distribution.

**28. How can a log-normal distribution be useful in finance?**

A log-normal distribution is useful in finance for modeling the distribution of asset prices, as it accommodates the right-skewed behavior often observed in financial markets.

**29. What is Standard deviation?**

Standard deviation is a statistical measure of the amount of variation or dispersion in a set of data. It provides valuable information about the spread of data points around the mean (average) of a dataset. In other words, it quantifies how individual data points deviate from the mean.

**30. How is the Central Limit Theorem useful in statistics?**

The CLT is essential because it allows us to make statistical inferences about population parameters using the normal distribution, even when the population is not normally distributed.

**31. What is the Chi-Square Test, and when is it used?**

The Chi-Square Test is a statistical test used to determine whether there is a significant association between categorical variables. It's commonly used in contingency table analysis and goodness-of-fit testing.

**32. What is probability, and how is it defined?**

Probability is a measure of the likelihood of an event occurring.

It is typically defined as the ratio of the number of favorable outcomes to the total number of possible outcomes.

The probability of an event ( $P(E)$ ) is defined as:

$$P(E) = \frac{\text{Number of favorable outcomes for event } E}{\text{Total number of possible outcomes in the sample space } S}$$

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

Discrete probability distributions are defined for countable or distinct outcomes, while continuous probability distributions are defined for continuous or uncountable outcomes.

### **34.What is Big Data?**

Big Data is a collection of data that is huge in volume yet growing exponentially with time. It is data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also data but with huge size.

### **35.What are the different types Of Big Data?**

There are three types of Big Data:

1. Structured
2. Unstructured
3. Semi-structured

### **36.What is Structured Big Data?**

Any data that can be stored, accessed and processed in the form of fixed format is termed as a 'structured' data.

### **37.What is Unstructured Big Data?**

Any data with unknown form or the structure is classified as unstructured data. In addition to the size being huge, un-structured data poses multiple challenges in terms of its processing for deriving value out of it.

### **38.What is Semi-structured Big Data?**

Semi-structured data can contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS.

### **39.What are the Characteristics Of Big Data?**

Big data can be described by the following characteristics:

- 1) Volume
- 2) Variety
- 3) Velocity
- 4) Variability

### **40.What is Variance?**

Variance is the expected value of the squared deviation of a random variable from its mean.

(Or)

It is the measurement of the distance of a set of random numbers from their collective average value.

#### **41. what is Central Tendency?**

Central tendency, in statistics, refers to the central or typical value in a set of data. It's a way to describe where most of the data points in a distribution cluster.

#### **42. What are the common measures of central tendency?**

There are three common measures of central tendency: 1. Mean 2. Median 3. Mode

**Mean (Average):** The mean is the most commonly used measure of central tendency. It is calculated by adding up all the values in a dataset and then dividing by the number of values. The formula for the mean is:

$$\text{Mean} = \frac{\text{Sum of All Data Points}}{\text{Number of Data Points}}$$

**Median:** The median is the middle value in a dataset when the values are arranged in ascending or descending order.

If there's an even number of data points, the median is the average of the two middle values.

**Mode:** The mode is the value that appears most frequently in a dataset. A dataset can have one mode (unimodal), more than one mode (multimodal), or no mode if all values occur with the same frequency. The mode is particularly useful for categorical or discrete data.

#### **43. What is Harmonic Mean?**

Harmonic mean is a type of average that is calculated by dividing the number of values in a data series by the sum of the reciprocals ( $1/x_i$ ) of each value in the data series.



$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_1} + \dots + \frac{1}{x_1}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

#### **44. What are the Types of Business Analytics Methods?**

Business analytics encompasses a range of methods and techniques for analyzing data to make informed business decisions.

1. Descriptive Analytics
2. Diagnostic Analytics
3. Predictive Analytics
4. Prescriptive Analytics

#### **45. Explain Descriptive analytics?**

Descriptive analytics is a type of business intelligence that involves analyzing and summarizing data to provide a descriptive overview of the data.

It helps organizations to understand their historical performance, identify trends, and detect patterns in their data. Descriptive analytics is useful for gaining insights into customer behavior, market trends, and business operations, and can help organizations make informed decisions.

#### **46. what is Diagnostic Analytics?**

Diagnostic analytics is a form of data analysis used to diagnose and identify issues within a system or process. It involves examining data to identify patterns, trends, and anomalies, and using that information to pinpoint the root cause of problems. This can help organizations improve their operations, reduce costs, and optimize performance.

#### **47. what is Predictive Analytics and explain?**

Predictive analytics is a branch of data analytics that uses statistical models and machine learning algorithms to forecast future events or behaviors. It involves analyzing historical data and identifying patterns to make predictions about future

outcomes, such as customer churn, job applicant success, or stock prices. Predictive analytics can help organizations make informed decisions by providing them with actionable insights and improving their overall business predictive accuracy.

#### **48. what is Prescriptive Analytics ?**

Prescriptive analytics is a type of advanced analytics that goes beyond descriptive and predictive analytics by providing recommendations or prescriptions for actions based on the analysis of historical and real-time data. It helps organizations make informed decisions by identifying the best course of action among various alternatives, considering constraints and limitations. In other words, prescriptive analytics tells you not just what will happen, but what you should do to achieve your goals.

#### **49. what is data wrangling?**

Data wrangling is the process of cleaning, transforming, and organizing data into a format suitable for analysis. It involves removing errors, handling missing values, and transforming data into a consistent format to ensure accuracy and efficiency in data analysis. Data wrangling is an essential part of the data science workflow, as it helps to prepare data for insights and predictions.

#### **50. what is EDA (Exploratory Data Analysis) ?**

EDA (Exploratory Data Analysis) is an essential step in data analysis that involves exploring and summarizing datasets to gain insights and identify patterns. It typically involves cleaning and organizing the data, visualizing distributions, identifying missing values, and performing statistical tests to understand the data's characteristics. EDA helps data analysts to evaluate the quality of the data, define the problem statement, and develop a hypothesis for further investigation.