List of Data Analytics MCQs

**1. Data Analytics uses \_\_\_ to get insights from data.**

1. Statistical figures
2. Numerical aspects
3. Statistical methods
4. None of the mentioned above

**Answer:** C) Statistical methods

To gain insights from data, Data Analytics use statistical approaches. Organizations can use data analytics to uncover trends and develop insights by analyzing all of their data (real-time, historical, unstructured, structured, and qualitative).

**2. Amongst which of the following is / are the branch of statistics which deals with the development of statistical methods is classified as \_\_\_.**

1. Industry statistics
2. Economic statistics
3. Applied statistics
4. None of the mentioned above

**Answer:** C) Applied statistics

The discipline of statistics that works with the development of statistical procedures is known as applied statistics. Planning for data collecting, maintaining data, analyzing, interpreting, and drawing conclusions from data, and finding issues, solutions, and opportunities utilizing analysis are all part of applied statistics. In data analysis and empirical research, these major fosters critical thinking and problem-solving skills.

**3. Linear Regression is the supervised machine learning model in which the model finds the best fit \_\_\_ between the independent and dependent variable.**

1. Linear line
2. Nonlinear line
3. Curved line
4. All of the mentioned above

**Answer:** A) Linear line

linear Regression is a supervised Machine Learning model that identifies the best fit linear line between the independent and dependent variables, i.e., the linear connection between the dependent and independent variables.

**4. Amongst which of the following is / are the types of Linear Regression,**

1. Simple Linear Regression
2. Multiple Linear Regression
3. Both A and B
4. None of the mentioned above

**Answer:** C) Both A and B

There are two forms of linear regression: simple and multiple. Simple Linear Regression is used when there is only one independent variable and the model must determine the linear connection between it and the dependent variable. Multiple Linear Regression is employed more than one independent variable in the model to determine the link.

**5. Amongst which of the following is / are the true about regression analysis?**

1. Describes associations within the data
2. Modeling relationships within the data
3. Answering yes/no questions about the data
4. All of the mentioned above

**Answer:** B) Modeling relationships within the data

Regression analysis is used to describe relationships within data, and so it is a collection of statistical methods for estimating relationships between a dependent variable and one or more independent variables. There are various types of regression analysis, including linear, multiple linear, and nonlinear. Simple linear and multiple linear models are the most frequent. Nonlinear regression analysis is typically employed for more difficult data sets with a nonlinear connection between the dependent and independent variables.

**6. Linear regression analysis is used to predict the value of a variable based on the value of another variable.**

1. True
2. False

**Answer:** A) True

Linear regression analysis predicts the value of one variable depending on the value of another. The variable we wish to forecast is referred to as the dependent variable. The variable we are utilizing to predict the value of the other variable is referred to as the independent variable

**7. A Linear Regression model's main aim is to find the best fit linear line and the \_\_\_ of intercept and coefficients such that the error is minimized.**

1. Optimal values
2. Linear line
3. Linear polynomial
4. None of the mentioned above

**Answer:** A) Optimal values

**Explanation:**

The basic goal of a Linear Regression model is to determine the best fit linear line and the ideal intercept and coefficient values such that the error is minimized. A linear regression model describes the relationship between one or more independent variables, X, and a dependent variable, y. A multiple linear regression model is a type of regression model that has numerous lines of regression. A multiple linear regression model is *yi*=*β*0+*β*1*Xi*1+*β*2*Xi*2+⋯+*βpXip*+*εi*, *i*=1,⋯,*n*

**8. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.**

1. True
2. False

**Answer:** A) True

In statistics, the actual value is the value derived from observation or measurement of the available data. It is also known as the observed value. The expected value is the predicted value of the variable based on the regression analysis. Linear regression is most commonly used to calculate model error using mean-square error (MSE). MSE is derived by measuring the distance between the observed and anticipated y-values at each value of x and then computing the mean of the squared distances.

**9. The process of quantifying data is referred to as \_\_\_.**

1. Decoding
2. Structure
3. Enumeration
4. Coding

**Answer:** C) Enumeration

Enumeration is the term for the process of quantifying data. Any quantifiable information that can be used for mathematical calculations or statistical analysis is referred to as quantitative data. This type of information aids in the development of real-world decisions based on mathematical derivations. To answer inquiries like how many, quantitative data is used. How often do you do it? How much is it? This information can be confirmed and validated.

**10. Text Analytics, also referred to as Text Mining?**

1. True
2. False

**Answer:** A) True

**Explanation:**

Text analytics uses a combination of machine learning, statistical, and linguistic tools to analyze vast amounts of unstructured material (text that does not have a preset format) in order to draw insights and trends. It enables corporations, governments, researchers, and the media to make critical decisions based on the vast amounts of data available to them

**11. \_\_\_ are used when we want to visually examine the relationship between two quantitative variables.**

1. Bar graph
2. Scatterplot
3. Line graph
4. Pie chart

**Answer:** A) Bar graph

**Explanation:**

Dots are used to indicate values for two different numeric variables in a scatter plot, also known as a scatter chart or a scatter graph. The values for each data point are indicated by the position of each dot on the horizontal and vertical axes. Scatter plots are used to see how variables relate to one another.

**12. A graph that uses vertical bars to represent data is called a \_\_\_\_.**

1. Bar graph
2. Line graph
3. Scatterplot
4. All of the mentioned above

**Answer:** A) Bar graph

**Explanation:**

A bar graph is a graph that employs vertical bars to represent data. Bar graphs are visual representations of data (usually grouped) in the shape of vertical or horizontal rectangular bars, with bar length proportional to data measure. Bar charts are another name for them. In statistics, bar graphs are one of the data management methods.

**13. Data Analysis is a process of,**

1. Inspecting data
2. Data Cleaning
3. Transforming of data
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

The process of reviewing, cleansing, and manipulating data with the objective of identifying usable information, informing conclusions, and assisting decision-making is known as data analysis. Data analysis is important in today's business environment since it helps businesses make more scientific decisions and run more efficiently.

**14. Least Square Method uses \_\_\_.**

1. Linear polynomial
2. Linear regression
3. Linear sequence
4. None of the mentioned above

**Answer:** B) Linear regression

**Explanation:**

Linear regression employs the Least Square Method. The least-squares approach is a type of mathematical regression analysis that determines the best fit line for a collection of data, displaying the relationship between the points visually. The relationship between a known independent variable and an unknown dependent variable is represented by each piece of data.

**15. What is a hypothesis?**

1. A statement that the researcher wants to test through the data collected in a study
2. A research question the results will answer
3. A theory that underpins the study
4. A statistical method for calculating the extent to which the results could have happened by chance

**Answer:** A) A statement that the researcher wants to test through the data collected in a studyp

**Explanation:**

A hypothesis is a proposition that a researcher wishes to evaluate using data from a study. A hypothesis is a conclusion reached after considering evidence. This is the first step in any investigation, where the research questions are translated into a prediction. Variables, population, and the relationship between the variables are all included. A research hypothesis is a hypothesis that is tested to see if two or more variables have a relationship.

**16. Linear-regression models are relatively simple and provide an easy-to-interpret mathematical formula that can generate \_\_\_.**

1. Predictions
2. Interpretation
3. Conclusion
4. None of the mentioned above

**Answer:** A) Predictions

**Explanation:**

Linear-regression models are straightforward and provide a basic mathematical method for generating predictions. Linear regression can be used in a variety of corporate and academic study.

**17. Amongst which of the following is / are the applications of Linear Regression,**

1. Biological
2. Behavioral
3. Social sciences
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Linear regression is utilized in a variety of fields, including biology, behavioral science, environmental research, and business. Linear regression models have proven to be a reliable and scientific means of forecasting the future. Because linear regression is a well-known statistical process, its properties are well understood and linear regression models may be trained quickly.

**18. With reference to data, dependent and independent variables should be quantitative.**

1. True
2. False

**Answer:** A) True

**Explanation:**

Dependent and independent variables should be quantitative when it comes to data. Both the dependent and independent variables should have a numerical value. Religious, major field of study and residential region categorical factors must be represented as binary variables or other sorts of contrast variables.

**19. For each value of the \_\_\_, the distribution of the dependent variable must be normal.**

1. Independent variable
2. Depended variable
3. Intermediate variable
4. None of the mentioned above

**Answer:** A) Independent variable

**Explanation:**

The dependent variable's distribution must be normal for each value of the independent variable. For all values of the independent variable, the variance of the dependent variable's distribution should be constant. The dependent variable should have a linear relationship with each independent variable, and all observations should be independent.

**20. Residual plot helps in analyzing the model using the values of residues.**

1. True
2. False

**Answer:** A) True

**Explanation:**

The residue plot aids in the analysis of the model by displaying the values of the residues. It's shown as a line between the projected values and the residual. Their values are all the same. The point's distance from 0 indicates how inaccurate the prediction was for that number. If the value is positive, the probability of success is minimal. If the value is negative, the probability of success is high. A number of 0 implies that the forecast is perfect. The model can be improved by detecting residual patterns.

**21. Amongst which of the following is / are not a major data analysis approach?**

1. Predictive Intelligence
2. Business Intelligence
3. Text Analytics
4. Data Mining

**Answer:** A) Predictive Intelligence

**Explanation:**

The practice of collecting data about consumers' and potential consumers' behaviors/actions from a number of sources and perhaps integrating it with profile data about their qualities is known as predictive intelligence.

**22. By 2025, the volume of data will increase to,**

1. TB
2. YB
3. ZB
4. EB

**Answer:** C) ZB

**Explanation:**

It is projected that 2.5 quintillion bytes of data are created every day, with the volume of digital data expected to reach Zeta Byte by 2025.

**23. Alternative Hypothesis is also called as?**

1. Null Hypothesis
2. Research Hypothesis
3. Simple Hypothesis
4. None of the mentioned above

**Answer:** B) Research Hypothesis

**Explanation:**

The alternative hypothesis is the assertion that is being tested against the null hypothesis. Ha or H1 are common abbreviations for alternative hypotheses. The alternative hypothesis is the hypothesis that is inferred from a null hypothesis that has been rejected. It is best stated as an explanation for why the null hypothesis was rejected. It is also known as the research hypothesis. Unlike the null hypothesis, the researcher is usually most interested in the alternative hypothesis.

**24. If the null hypothesis is false then which of the following is accepted?**

1. Alternative Hypothesis.
2. Null Hypothesis
3. Both A and B
4. None of the mentioned above

**Answer:** C) Both A and B

**Explanation:**

The alternative hypothesis is accepted if the null hypothesis is untrue. An alternative theory is a proposition that a researcher is testing in hypothesis testing. From the researcher's perspective, this assertion is correct, and it finally proves to reject the null hypothesis and replace it with a different one. The difference between two or more variables is anticipated in this hypothesis.

**25. Amongst which of the following is / are not an example of social media?**

1. Twitter
2. Instagram
3. Both A and B
4. None of the mentioned above

**Answer:** C) Both A and B

**Explanation:**

Social media is a type of computer-based technology that allows people to share their ideas, thoughts, and information with others via virtual networks and communities. Social media is an internet-based platform that allows people to share content such as personal information, documents, films, and images quickly and electronically.

**26. Velocity is the speed at which the data is processed -**

1. True
2. False

**Answer:** A) True

**Explanation:**

The rate at which data is generated, distributed, and gathered is referred to as data velocity. High data velocity is created at such a rapid rate that it necessitates the use of specialized processing techniques. The faster data can be captured and processed, the more valuable the data collected will be and the longer it will hold its worth.

**27. \_\_\_ refers to the ability to turn your data useful for business.**

1. Value
2. Variety
3. Velocity
4. None of the mentioned above

**Answer:** A) Value

**Explanation:**

The ability to turn our data into business value is referred to as value. The usefulness of obtained data for our business is referred to as data value. Data, regardless of its magnitude, is rarely useful on its own; to be useful, it must be transformed into insights or knowledge, which is where data processing comes in.

**28. Correlation is the relationship between two variables -**

1. One
2. Two
3. Zero
4. All of the mentioned above

**Answer:** B) Two

**Explanation:**

Correlation is the strength of a relationship between two variables, and the Pearson's correlation coefficient measures how strong that relationship is. The correlation of two variables is the statistical link between them. A positive correlation means that both variables move in the same direction, while a negative correlation means that when one variable's value rises, the other variable's value falls.

**29. The Mean Squared Error is a measure of the average of the squares of the residuals.**

1. True
2. False

**Answer:** A) True

**Explanation:**

The degree of inaccuracy in statistical models is measured by the mean squared error (MSE). The average squared difference between observed and expected values is calculated. The MSE equals zero when a model has no errors. Its value rises as the model inaccuracy rises. The mean squared deviation is another name for the mean squared deviation (MSD). The average squared residual is represented by the mean squared error in regression.

**30. Logistic regression is used to find the probability of event = Success and event = \_\_\_\_.**

1. Failure
2. Success
3. Both A and B
4. None of the mentioned above

**Answer:** A) Failure

**Explanation:**

The likelihood of event=Success and event=Failure is calculated using logistic regression. When the dependent variable is in nature, we should utilize logistic regression. For classification difficulties, logistic regression is commonly employed. There is no requirement for a linear relationship between the dependent and independent variables in logistic regression. Because it uses a non-linear log transformation on the anticipated odds ratio, it can handle a wide range of relationships.

**31. A good data analytics solution includes a viable self-service \_\_\_.**

1. Data mining
2. Data wrangling
3. Data warehouse
4. None of the mentioned above

**Answer:** B) Data wrangling

**Explanation:**

A smart data analytics solution incorporates self-service data wrangling and data preparation features so that data may be simply and quickly gathered from a range of incomplete, difficult, or messy data sources and cleansed for mashup and analysis.

**32. To glean insights from the data, many analysts and data scientists rely on \_\_\_.**

1. Data mining
2. Data visualization
3. Data warehouse
4. All of the mentioned above

**Answer:** B) Data visualization

**Explanation:**

Many analysts and data scientists use data visualization, or the graphical depiction of data, to assist individuals visually explores and finds patterns and outliers in the data in order to get insights. Data visualization features are included in a good data analytics system, making data exploration easier and faster.

**33. Predictive analytics involves taking historical data -**

1. True
2. False

**Answer:** A) True

**Explanation:**

The approach or practice of utilizing data to generate projections about the possibility of certain future events in your organization is known as predictive analytics, which is a form of advanced analytics. Predictive analytics models unknown future occurrences by combining historical and current data with advanced statistics and machine learning approaches. It is commonly characterized as utilizing data science and machine learning to learn from an organization's previous collective experience in order to make better decisions in the future.

**34. With reference to Predictive analytics, it allows organizations to predict customer behavior -**

1. True
2. False

**Answer:** A) True

**Explanation:**

Predictive analytics enables businesses to forecast consumer behavior and business results by combining historical and real-time data. Furthermore, predictive modeling is a subset of this activity that entails constructing and maintaining models, testing and iterating with existing data, and embedding models into applications.

**35. Customer analytics refers -**

1. Customer Relationship Management: churn analysis and prevention
2. Marketing: cross-sell, up-sell
3. Pricing: leakage monitoring, promotional effects tracking, competitive price responses
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Customer analytics includes churn analysis and prevention, marketing: cross-sell and up-sell, and pricing: leakage monitoring, promotional effects tracking, and competitive price reactions.

**36. \_\_\_ is the cyclical process of collecting and analyzing data during a research study.**

1. Extremis Analysis
2. Constant analysis
3. Interim Analysis
4. All of the mentioned above

**Answer:** C) Interim Analysis

**Explanation:**

The cyclical process of gathering and assessing data throughout a research Endeavour is known as interim analysis.

**37. An advantage of using computer programs for qualitative data is that they \_\_\_.**

1. Can reduce time required to analyze data
2. Help in storing and organizing data
3. Make many procedures available that are rarely done by hand due to time constraints
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Qualitative data is that they can reduce time required to analyze data, help in storing and organizing data and make many procedures available that are rarely done by hand due to time constraints.

**38. Data Modeling is the process of analyzing the data objects -**

1. True
2. False

**Answer:** A) True

**Explanation:**

The practice of evaluating data items and their relationships with other things is known as data modeling. It's utilized to look into the data requirements for various business activities. The data models are constructed in order to store the information in a database.

**39. \_\_\_ are the basic building blocks of qualitative data.**

1. Categories
2. Data chunk
3. Numeric figures
4. None of the mentioned above

**Answer:** A) Categories

**Explanation:**

The fundamental building elements of qualitative data are categories. The descriptive and conceptual results gathered through surveys, interviews, or observation is referred to as qualitative data. We can explore concepts and further explain quantitative outcomes by analyzing qualitative data.

**40. Metadata and data modeling tools support the creation and documentation of models -**

1. True
2. False

**Answer:** A) True

**Explanation:**

Models representing the structures, flows, mappings and transformations, connections, and quality of data may be created and documented using metadata and data modeling tools.

**41. The Process of describing the data that is huge and complex to store and process is known as \_\_\_.**

1. Analytics mining
2. Data cleaning
3. Big data
4. None of the mentioned above

**Answer:** C) Big data

**Explanation:**

Big data is a term used to describe the process of describing data that is large and difficult to store and interpret. Big data analytics is the use of advanced analytic techniques to very large, heterogeneous big data sets, which can contain structured, semi-structured, and unstructured data, as well as data from many sources and sizes ranging from terabytes to zettabytes.

**42. In descriptive statistics, data from the entire population or a sample is summarized with \_\_\_.**

1. Numerical descriptor
2. Decimal descriptor
3. Integer descriptor
4. All of the mentioned above

**Answer:** A) Numerical descriptor

**Explanation:**

Data from the full population or a sample is summarized using numerical descriptors in descriptive statistics.

**43. Customer behavior analytics is about understanding how your customers act -**

1. True
2. False

**Answer:** A) True

**Explanation:**

Understanding how your customers behave across each channel and interaction point is the goal of customer behavior analytics. Understanding consumer behavior may aid in customer acquisition, engagement, and retention for your company.

**44. Data Analysis is defined by the statistician?**

1. John Tukey
2. Hans Peter Luhn
3. Gregory Lon
4. None of the mentioned above

**Answer:** A) John Tukey

**Explanation:**

John Tukey, a statistician, defined data analysis. Tukey began his career in statistics, and he was fascinated with data analysis challenges and methodologies. Some people remember him for pioneering exploratory data analysis, but he also made significant contributions to analysis of variance, regression, and a wide range of applications. This study examines some of the most notable contributions in these fields.

**45. Amongst which of the following is / are the challenges overcome by the data strategy to make a business in a strong position -**

1. Data privacy, data integrity, and data quality issues that undercut your ability to analyze data
2. Inefficient movement of data between different parts of the business
3. Lack of deep understanding of critical parts of the business
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Data strategy aids in the development of a strong firm. It also puts a company in a good position to overcome obstacles. Issues with data privacy, integrity, and quality that limit your capacity to evaluate data Lack of understanding of important business components and the processes that keep them run Inefficient data transportation between different portions of the organization, or data duplication by several business units, as well as a lack of clarity about current business needs and goals.

**46. Tableau is a \_\_\_ tool.**

1. Visualization
2. Analytical
3. Data Exploration
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Tableau is a visualization software program. Tableau gives data scientists a versatile front-end for data exploration with the analytical depth they need. Data scientists may execute complicated quantitative studies in Tableau and communicate visual findings to encourage improved understanding and collaboration with data by utilizing advanced computations, R and Python integration, quick cohort analysis, and predictive capabilities.

**47. Big data analytics refers to collecting, processing, cleaning, and analyzing large datasets -**

1. True
2. False

**Answer:** A) True

**Explanation:**

Big data analytics is the process of gathering, processing, cleaning, and analyzing enormous datasets in order to assist businesses operationalize their data.

**48. Amongst which of the following is / are the features of Tableau for data analytics -**

1. Data Blending
2. Real time analysis
3. Collaboration of data
4. All of the mentioned above

**Answer:** D) All of the mentioned above

**Explanation:**

Tableau software's finest features are data blending, real-time analysis, and data collaboration. The beautiful thing about Tableau software is that it can be used without any technical or programming knowledge. The tool has piqued the curiosity of people from many walks of life, including business, researchers, and other industries.

**49. \_\_\_ is a category, also called supervised machine learning methods in which the data is split on two parts.**

1. Classification
2. Clustering
3. Data mining
4. None of the mentioned above

**Answer:** A) Classification

**Explanation:**

Classification is a type of supervised machine learning approach in which the data is divided into two parts: a training set and a validation set. A model is trained from the training set by extracting the most discriminative characteristics that are previously connected with known outputs. This model is then tested on a test set, in which we evaluate the learnt model's efficiency by creating appropriate outputs for a particular set of input values.

**50. Clustering belongs to \_\_\_ data analysis.**

1. Supervised
2. Unsupervised
3. Both A and B
4. None of the mentioned above

**Answer:** B) Unsupervised

**Explanation:**

Unsupervised data analysis includes clustering. Without any prior knowledge, the data's hidden structure is discovered and emphasized. Popular clustering techniques include K-means, K-nearest neighbors, and hierarchical clustering.