

1.What is the main purpose of using a histogram in data analysis?

- To summarize statistical measures like mean and median
- To compare categorical data
- To show the relationship between two variables
- To display the frequency distribution of numerical data**

2. A frequency table is used to display data in the form of a matrix with rows and columns.

- True
- **False**

3. Box plots are primarily used to show the central tendency of data, such as mean and mode

- True
- False**

4. Which of the following are common steps in data preprocessing? (Check all that apply)

- Creating visualizations
- Data transformation**
- Data normalization**
- Handling missing values**

5.What is the primary purpose of calculating the standard deviation of a dataset?

- To display data visually
- To determine the central value
- To assess the average deviation from the mean**
- To find the most common value

6.In a box plot, the line in the center of the box represents the **median** of the data.

7. The process of **Data Wrangling** involves organizing raw data into a more structured format for analysis.

8. The function **head()** provides the first few rows of a Data Frame for an initial look at the data.

9.The process of transforming raw data into a more meaningful format is known as

- Data Storage
- Data Wrangling**
- Data Collection
- Data Visualization

10. Which visualizations are useful for displaying the distribution of numerical data?

- Box Plot
- Histogram
- Line Chart
- Bar Chart

11. When analyzing data, which measures can be used to understand the spread of the data?

- Median
- Standard Deviation
- Range
- Variance

12. The interquartile range (IQR) is used to measure the spread of data around the mean.

- True
- False

13. Unstructured data is challenging to store and analyze due to its lack of a fixed format.

- True
- False

14.To load a dataset in Python using Pandas, the command used is **pd.read_csv()** considering the variable is pd.

15.Which of the following is an example of structured data?

- Handwritten notes
- A video files
- A photo album
- A CSV file with customer details**

16.The command to plot a histogram of Math_Score using Matplotlib is plt.hist(df['Math_Score']) after setting up **matplotlib.pyplot** with import matplotlib.pyplot as plt.

matplotlib.pyplot

17.Variance and standard deviation both measure data variability, but standard deviation is the square root of variance.

- True**

- False

18.Z-scores indicate how many standard deviations a data point is from the mean.

- True**

- False

19.Which measure is most sensitive to extreme values in a dataset?

Mean

Median

Interquartile Range

Mode

20. The measure of how much values in a dataset vary from the mean is known as **Standard Deviation**.

21. The **mean()** function in Pandas can be used to compute the mean of a column in a DataFrame.

22.Categorical data can be displayed effectively using bar charts and pie charts.

True

False

23. Metadata is data that describes other data, providing context for better understanding.

True

False

24.Which of these measures is used to understand the central tendency of data?

- Variance
- Interquartile Range
- Standard Deviation
- Mean

25.Which data visualization method is best for identifying outliers in a dataset?

- Scatter Plot
- Box Plot
- Bar Chart
- Histogram

26.A **histogram** is a graphical representation that shows the frequency distribution of a dataset, often used for numerical data.

27. Which of these data types are considered categorical?

- Ordinal
- Interval
- Nominal
- Ratio

28. In data analysis, which method would you use to explore the relationship between two numerical variables?

- Pie Chart
- Frequency Table
- Box Plot

Scatter Plot

29. Nominal data has an inherent order or ranking associated with it.

- True
- False**

30. The method to check for missing values in each column of a DataFrame is **isnull().sum()**.

isnull()

31. Which of the following can help identify outliers in a dataset?

- Scatter Plot**
- Line Chart
 - Box Plot**
 - Z-scores**

32. Structured Data is the term used to describe data that is structured in a tabular format, with rows and columns.

33. Which data type is represented by values that can only be categorized and counted but have no specific order?

Interval

Ratio

Nominal

Ordinal

34. Structured data is organized and easy to analyze because it follows a fixed format.

True

False

35. What does the term "data cleaning" refer to in data analysis?

Visualizing data for better understanding

Formatting data to match the software requirements

Organizing data into tables and columns

Removing duplicate and inconsistent data

