



class: 11th

Subject: Computer Science

Unit 5: Data Analytics

❖ Important MCQs:

1. Which statement best describes data analytics?

(a) Storing data in databases

(b) Examining data to extract useful insights and patterns ✓

-
- (c) Deleting unnecessary data
 - (d) Converting data into images

2. Why is statistics important in data analysis?

- (a) It increases data size
- (b) It simplifies and summarizes large datasets
- (c) It removes all errors
- (d) It stores data permanently

3. Which measure of central tendency is MOST affected by outliers?

- (a) Median
- (b) Mode
- (c) Mean
- (d) Range



4. Which measure is BEST to use when data contains extreme values?

- (a) Mean
- (b) Median
- (c) Mode

(d) Variance

5. If all values in a dataset are the same, what will be the variance?

(a) High

(b) Low

(c) Zero

(d) Undefined

6. What happens to the mean if one value increases significantly?

(a) Mean remains same

(b) Mean increases

(c) Mean decreases

(d) No effect

7. In a dataset, if two values occur most frequently, the dataset is called:

(a) Unimodal

(b) Bimodal

(c) Trimodal

(d) Normal

8. Which measure gives the MOST accurate center when data is skewed?

- (a) Mean
- (b) Median
- (c) Mode
- (d) Variance

9. Which of the following best describes variance?

- (a) Middle value of dataset
- (b) Average of all values
- (c) Spread of data from mean
- (d) Most frequent value



StudyNotes360.com

10. If variance is high, what does it indicate?

- (a) Data points are close together
- (b) Data points are widely spread
- (c) Data is constant
- (d) Data has no mean

11. Standard deviation is the:

- (a) Square of variance

(b) Square root of variance ✓

(c) Sum of values

(d) Mean of dataset

12. Which dataset has greater consistency?

(a) High variance

(b) Low variance ✓

(c) High mean

(d) High mode

13. If mean = median = mode, the distribution is:

(a) Skewed

(b) Symmetrical (normal) ✓

(c) Random

(d) Bimodal

14. Which measure helps identify the most common value quickly?

(a) Mean

(b) Median

(c) Mode ✓

(d) Variance

15. Which measure requires data to be arranged in order?

- (a) Mean
- (b) Median
- (c) Mode
- (d) Variance

16. In which situation is mode NOT useful?

- (a) Finding most common value
- (b) When all values are unique
- (c) Identifying repeated data
- (d) Categorical data

17. Which measure of dispersion is expressed in same unit as data?

- (a) Variance
- (b) Standard Deviation
- (c) Mean
- (d) Mode

18. A dataset with values close to mean will have:

- (a) High variance
- (b) Low variance

(c) No mean

(d) High mode

19. Which of the following is TRUE about median?

(a) It is affected by outliers

(b) It is always the highest value

(c) It divides data into two equal halves

(d) It is same as mean

20. Which concept helps in understanding how far values deviate from mean?

(a) Central tendency

(b) Dispersion

(c) Mode

(d) Median

21. Standard deviation is defined as:

(a) Square of mean

(b) Square root of variance

(c) Sum of all values

(d) Difference between values

22. The main advantage of standard deviation over variance is:

- (a) It is always larger
- (b) It is in the same unit as data ✓
- (c) It removes errors
- (d) It ignores mean

23. A small standard deviation indicates:

- (a) Data is widely spread
- (b) Data is closely clustered around mean ✓
- (c) Data has no variation
- (d) Data is incorrect

24. A large standard deviation indicates:

- (a) Data is close to mean
- (b) Data is widely spread ✓
- (c) Data is constant
- (d) Data is sorted

25. Which class shows more consistency?

- (a) Class A (SD = 3.55) ✓
- (b) Class B (SD = 21.26)

-
- (c) Both are equal
 - (d) Cannot be determined

26. Probability is used to measure:

- (a) Data storage
- (b) Likelihood of an event occurring
- (c) Data size
- (d) Data type

27. The probability of getting heads in a fair coin toss is:

- (a) 0
- (b) 1
- (c) $1/2$
- (d) 2



28. Probability values always lie between:

- (a) -1 to 1
- (b) 0 to 1
- (c) 1 to 10
- (d) 0 to 100

29. The formula for probability is:

(a) Total outcomes / Favorable outcomes

(b) Favorable outcomes / Total outcomes

(c) Mean / Total values

(d) Variance / Mean

30. Which of the following is a real-life application of probability?

(a) Writing essays

(b) Predicting weather

(c) Drawing diagrams

(d) Typing documents

31. The MOST critical factor affecting analysis results is:

(a) Data size

(b) Data quality and relevance

(c) Data color

(d) Data format

32. Data collection is defined as:

(a) Deleting data

(b) Gathering relevant information for a purpose

(c) Analyzing data

(d) Visualizing data

33. Selecting a data collection method depends on:

(a) Internet speed

(b) Research objective and data type

(c) Data size only

(d) Software used

34. Which of the following is NOT a data collection method?

(a) Surveys

(b) Observations

(c) Experiments

(d) Compilation



35. Surveys are MOST suitable when:

(a) Behavior is observed

(b) Large structured data is needed

(c) Variables are controlled

(d) No interaction is required

36. A key characteristic of surveys is:

(a) Random observation

(b) Predefined set of questions

(c) No structure

(d) Only experiments

37. Which method involves collecting data without direct questioning?

(a) Survey

(b) Experiment

(c) Observation

(d) Interview

38. Observation is BEST when:

(a) Opinions are needed

(b) Natural behavior is studied without interference

(c) Data is numerical only

(d) Surveys are available

39. Experiments are different because they:

(a) Only observe data

(b) Manipulate variables to study effects

(c) Use no data

(d) Ignore results

40. In experiments, the purpose of changing variables is to:

(a) Store data

(b) Measure their effect on outcomes

(c) Reduce data

(d) Sort data

41. Data preparation is necessary to:

(a) Increase data size

(b) Ensure accurate and reliable analysis

(c) Delete all errors

(d) Avoid data collection

42. Which step is included in data preparation?

(a) Data cleaning

(b) Data deletion

(c) Data ignoring

(d) Data duplication

43. If data is not properly prepared, it may result in:

(a) Accurate results

(b) Misleading conclusions

(c) Faster processing

(d) Better storage

44. Handling missing data during preparation helps to:

(a) Reduce dataset size

(b) Improve accuracy and reliability

(c) Delete important values

(d) Increase errors

45. Estimating missing values using existing data is called:

(a) Observation

(b) Experiment

(c) Imputation

(d) Visualization

46. Data cleaning mainly focuses on:

(a) Creating graphs

(b) Fixing errors and inconsistencies in data

(c) Collecting data

(d) Storing data

47. Which of the following is NOT a data cleaning issue?

- (a) Missing values
- (b) Duplicate entries
- (c) Incorrect data
- (d) Data visualization

48. If duplicate entries exist in a dataset, they should be:

- (a) Ignored
- (b) Removed or corrected during cleaning
- (c) Increased
- (d) Converted

49. Raw data is often:

- (a) Always correct
- (b) Error-free
- (c) Incomplete or inconsistent
- (d) Fully analyzed

50. The main purpose of data cleaning is to ensure:

- (a) Faster storage
- (b) Accurate analysis results

-
- (c) Bigger dataset
 - (d) Random output

51. Data transformation is the process of:

- (a) Deleting data
- (b) Changing data format for analysis
- (c) Collecting data
- (d) Ignoring errors

52. Which of the following is an example of data transformation?

- (a) Correcting spelling mistakes
- (b) Converting data into categories or summaries
- (c) Removing duplicates
- (d) Collecting responses

53. Aggregating student scores into average marks is an example of:

- (a) Data cleaning
- (b) Data transformation
- (c) Data collection
- (d) Observation

54. Data transformation helps to:

- (a) Make data harder to analyze
- (b) Improve data usability for analysis or modeling ✓
- (c) Remove all data
- (d) Increase errors

55. Missing data means:

- (a) Extra data
- (b) Incomplete or unavailable values in dataset ✓
- (c) Clean data
- (d) Duplicate data

56. One simple method to handle missing data is:

- (a) Always delete dataset
- (b) Remove rows with missing values if few ✓
- (c) Increase missing values
- (d) Ignore analysis

57. Imputation means:

- (a) Deleting records
- (b) Estimating missing values using available data ✓

(c) Sorting data

(d) Collecting data

58. Flagging missing data means:

(a) Filling values

(b) Marking missing entries for awareness

(c) Removing dataset

(d) Ignoring dataset

59. Removing records with missing data is acceptable when:

(a) Missing data is very large

(b) Missing data is very small and does not affect analysis significantly

(c) Data is already complete

(d) Data is not important

60. A disadvantage of removing missing data is:

(a) Better accuracy

(b) Loss of useful information

(c) Faster processing

(d) Improved dataset quality

61. Statistical modeling is used to:

- (a) Store data permanently
- (b) Analyze data and make predictions about the future ✓
- (c) Delete errors from data
- (d) Create databases

62. The FIRST step in model development is:

- (a) Train model
- (b) Collect data
- (c) Define the problem ✓
- (d) Evaluate model



63. Defining the problem means:

- (a) Choosing algorithm
- (b) Understanding what needs to be predicted or analyzed ✓
- (c) Cleaning data
- (d) Running model

64. After defining the problem, the next step is:

- (a) Evaluate model
- (b) Collect data ✓

-
- (c) Choose output
 - (d) Ignore variables

65. In model development, algorithms are used to:

- (a) Delete data
- (b) Build models from data patterns ✓
- (c) Store results
- (d) Clean errors only

66. Training a model means:

- (a) Testing final output
- (b) Teaching model using data to make predictions ✓
- (c) Deleting dataset
- (d) Visualizing graphs

67. Model evaluation is done using:

- (a) Same data used for training only
- (b) New or unseen data ✓
- (c) Random guessing
- (d) No data

68. The purpose of evaluation is to check:

-
- (a) Data size
 - (b) Accuracy of predictions
 - (c) Storage capacity
 - (d) Data format

69. Linear regression is used to:

- (a) Classify data
- (b) Predict continuous values based on relationship between variables
- (c) Store data
- (d) Remove duplicates

70. In linear regression, the independent variable is:

- (a) Output
- (b) Input variable that influences result
- (c) Error term
- (d) Constant value

71. The dependent variable is:

- (a) Input
- (b) Output or predicted value

(c) Slope

(d) Intercept

72. The slope (β_1) represents:

(a) Starting value

(b) Change in Y for each unit change in X

(c) Error

(d) Total sum

73. The intercept (β_0) represents:

(a) Value of Y when X = 0

(b) Maximum value

(c) Slope

(d) Average value



74. In the regression equation, $Y = \beta_0 + \beta_1 X$, Y represents:

(a) Independent variable

(b) Dependent variable

(c) Slope

(d) Error

75. The error term (ϵ) in regression represents:

(a) Perfect prediction

(b) Difference between actual and predicted values

(c) Mean value

(d) Slope value

76. If slope = 40, it means:

(a) No change in output

(b) Each customer increases earnings by 40 Rs

(c) Earnings decrease

(d) Fixed earnings only

77. The intercept in the given example is:

(a) 40

(b) 100

(c) 500

(d) 22

78. Predicted earnings depend on:

(a) Only intercept

(b) Intercept and slope with input variable

(c) Only error

(d) Only mean

79. If prediction is not exactly equal to actual value, the difference is called:

(a) Mean

(b) Error

(c) Median

(d) Mode

80. Real-world data predictions are often:

(a) Always perfect

(b) Slightly different from actual values due to variation

(c) Always incorrect

(d) Not useful

81. Logistic regression is used to predict:

(a) Continuous values

(b) Categorical outcomes (Yes/No)

(c) Mean values

(d) Variance

82. Logistic regression outputs values between:

-
- (a) -1 to 1
 - (b) 0 to 100
 - (c) 0 to 1
 - (d) 1 to 10

83. Logistic regression is mainly used for:

- (a) Clustering data
- (b) Classification problems
- (c) Data cleaning
- (d) Data storage

84. Linear regression and logistic regression differ because logistic regression:

- (a) Predicts exact numbers
- (b) Predicts probabilities instead of exact values
- (c) Uses no variables
- (d) Only stores data

85. In logistic regression, output represents:

- (a) Exact score
- (b) Probability of an event occurring

(c) Data mean

(d) Data mode

86. Clustering is used to:

(a) Predict future values

(b) Group similar data together based on features

(c) Remove errors

(d) Store data

87. In clustering, grouping is based on:

(a) Random selection

(b) Similarity of characteristics

(c) Alphabet order

(d) Data size

88. K-means clustering requires:

(a) No groups

(b) Predefined number of clusters (K)

(c) Only one cluster

(d) No data

89. In K-means, K represents:

-
- (a) Number of variables
 - (b) Number of clusters
 - (c) Error rate
 - (d) Mean value

90. The main idea of K-means clustering is to:

- (a) Increase errors
- (b) Group similar data points together
- (c) Delete data
- (d) Predict continuous values

91. Students with similar performance are grouped based on:

- (a) Age only
- (b) Scores in subjects like math and English
- (c) Name length
- (d) Attendance only

92. Performance metrics are used to:

- (a) Collect data
- (b) Measure model performance
- (c) Delete models

(d) Train hardware

93. Error metric measures:

(a) Correct predictions

(b) Difference between predicted and actual values

(c) Data size

(d) Model type

94. Accuracy measures:

(a) Total data collected

(b) Percentage of correct predictions made by the model

(c) Error rate only

(d) Data cleaning quality

95. Ethical consideration in data modeling includes:

(a) Increasing data size

(b) Fairness, bias, and privacy concerns

(c) Faster computation

(d) Removing all errors

96. Data visualization is used to:

(a) Store data

(b) Represent data in visual form for better understanding ✓

(c) Delete data

(d) Encrypt data

97. The main benefit of data visualization is:

(a) Making data complex

(b) Identifying patterns and trends easily ✓

(c) Increasing data size

(d) Slowing analysis

98. Which visualization is BEST for comparing categories?

(a) Line graph

(b) Histogram

(c) Bar chart ✓

(d) Scatterplot

99. In a bar chart, comparison is shown using:

(a) Colors only

(b) Height or length of bars ✓

(c) Random points

(d) Time intervals

100. Line graphs are mainly used to show:

- (a) Categories
- (b) Trends over time
- (c) Data distribution
- (d) Data grouping

101. A line graph connects:

- (a) Bars
- (b) Data points with lines
- (c) Categories
- (d) Tables



102. Histograms are used to show:

- (a) Relationship between variables
- (b) Data distribution in intervals (bins)
- (c) Trends over time
- (d) Exact categories

103. In histograms, data is grouped into:

- (a) Lines
- (b) Bars only

(c) Bins or intervals

(d) Points

104. Scatterplots are used to show:

(a) Single variable distribution

(b) Relationship between two variables

(c) Time trends

(d) Categories comparison

105. Boxplots are used to show:

(a) Only mean values

(b) Data distribution, median, and outliers

(c) Only trends

(d) Only categories

106. Data visualization tools are used to:

(a) Store raw data

(b) Convert data into charts and graphs for better understanding

(c) Delete datasets

(d) Encrypt information

107. Which of the following is an easy-to-use data visualization tool?

- (a) Microsoft Excel
- (b) Google Sheets
- (c) Both a and b
- (d) Python only

108. Excel and Google Sheets are mainly used for:

- (a) Programming games
- (b) Creating charts and graphs from data
- (c) Writing documents
- (d) Designing hardware

109. In Excel, visualizations are created using the:

- (a) View tab
- (b) Insert tab
- (c) File tab
- (d) Home tab

110. The first step in creating a visualization is:

- (a) Choosing chart type

(b) Entering data into spreadsheet

(c) Adding labels

(d) Printing chart

111. Before creating a chart, data must be:

(a) Deleted

(b) Selected/highlighted

(c) Hidden

(d) Sorted randomly

112. Which of the following chart types is NOT commonly used in Excel?

(a) Bar chart

(b) Line graph

(c) Scatterplot

(d) Database table

113. A bar chart is selected when we want to:

(a) Show trends over time

(b) Compare categories visually

(c) Show distribution

(d) Show correlation

114. Line graphs are BEST used for:

(a) Comparing categories

(b) Showing changes over time

(c) Showing single values

(d) Showing groups

115. After selecting data, the next step is to:

(a) Delete it

(b) Choose chart type from Insert menu

(c) Close file

(d) Print data

116. Customizing a chart includes:

(a) Deleting axes

(b) Adding labels and titles to improve clarity

(c) Removing data

(d) Changing dataset

117. X-axis in a chart usually represents:

(a) Values

(b) Categories like time or groups ✓

(c) Errors

(d) Random numbers

118. Y-axis in a chart usually represents:

(a) Categories

(b) Numerical values or measurements ✓

(c) Names

(d) Labels only

119. The purpose of labeling a chart is to:

(a) Make it colorful

(b) Make it easier to understand and interpret ✓

(c) Increase data size

(d) Hide information

120. Data visualization helps in:

(a) Confusing data interpretation

(b) Understanding data patterns quickly ✓

(c) Removing useful data

(d) Slowing analysis process

❖ Important Short Questions:

1. Define data analytics.

Answer:

Data analytics is the process of examining raw data to discover useful information, patterns, and trends. It helps in making better and informed decisions based on data rather than guesswork.

Example:

A supermarket analyzes customer purchase data to find which products are most popular so they can increase stock of those items.

2. What is statistics?

Answer:

Statistics is a branch of mathematics that deals with collecting, organizing, analyzing, interpreting, and presenting data. It helps in simplifying large datasets into meaningful information.

Example:

Calculating the average marks of a class helps teachers understand overall student performance.

3. Define mean with formula.

Answer:

Mean is the average value of a dataset. It is calculated by adding all values and dividing by the total number of values.

Formula:

Mean = (Sum of all values) ÷ (Number of values)

Example:

Data: 10, 20, 30

Mean = $(10 + 20 + 30) \div 3 = 60 \div 3 = 20$

4. What is median?

Answer:

Median is the middle value of a dataset when it is arranged in ascending or descending order. If the number of values is odd, the middle value is the median. If even, the average of two middle values is taken.

Example:

Odd: 5, 10, 15 → Median = 10

Even: 5, 10, 15, 20 → Median = $(10 + 15) \div 2 = 12.5$

5. Define mode with example.

Answer:

Mode is the value that appears most frequently in a dataset. A dataset can have one mode, more than one mode, or no mode at all.

Example:

Data: 2, 3, 3, 5, 6

Mode = 3 (because it appears most often)

6. What are measures of central tendency?

Answer:

Measures of central tendency are statistical methods used to find the central or typical value of a dataset. The three main measures are mean, median, and mode.

Example:

In student marks, the average score (mean) represents overall class performance.

7. Define variance.

Answer:

Variance measures how far data values are spread from the mean. It shows the degree of variation in a dataset. A high variance means data is widely spread, while a low variance means data is close to the mean.

Example:

If two classes have different score spreads, the class with more scattered marks has higher variance.

8. What is standard deviation?

Answer:

Standard deviation is the square root of variance. It shows how much the data values deviate from the mean in the same unit as the original data, making it easier to interpret than variance.

Example:

If Class A has a small standard deviation, it means students' marks are close to the average.

9. Write formula of probability.

Answer:

Probability measures how likely an event is to occur.

Formula:

Probability = (Number of favorable outcomes) ÷ (Total number of outcomes)

Example:

In a coin toss:

Probability of heads = $1/2 = 0.5$

10. What is probability?

Answer:

Probability is the mathematical measure of the likelihood of an event happening. It always lies between 0 and 1, where 0 means impossible and 1 means certain.

Example:

Getting a head in a fair coin toss has a probability of 0.5 (50%).

11. Define data collection.

Answer:

Data collection is the process of gathering relevant and useful information from different sources for the purpose of analysis or research. It is the first step in any data analysis process and ensures that meaningful insights can be extracted later.

Example:

A school collects students' attendance records and exam marks to analyze performance.

12. Name three data collection methods.

Answer:

The three main data collection methods are:

1. Surveys

-
2. Observations
 3. Experiments

Example:

A company uses surveys to get customer feedback, observation to study customer behavior, and experiments to test new products.

13. What is a survey?

Answer:

A survey is a data collection method in which a set of predefined questions is asked to a group of people to collect information in a structured way. Surveys can be conducted online, through interviews, or questionnaires.

Example:

A supermarket asks customers about their favorite products using a questionnaire.

14. What is observation method?

Answer:

Observation is a data collection method in which data is collected by watching or monitoring subjects in their natural environment without interfering. It helps to study real behavior.

Example:

A restaurant manager observes which tables are most frequently chosen by customers.

15. What is an experiment in data collection?

Answer:

An experiment is a method in which one or more variables are controlled or changed to observe their effect on another variable. It is used to find cause-and-effect relationships.

Example:

A teacher tests whether students perform better with printed notes or digital notes.

16. What is data preparation?

Answer:

Data preparation is the process of cleaning, organizing, and formatting raw data so that it becomes suitable for analysis. It ensures data is accurate and consistent before use.

Example:

Removing duplicate entries and correcting missing values in student records.

17. Why is data preparation important?

Answer:

Data preparation is important because raw data often contains errors, missing values, or inconsistencies. Proper preparation ensures accurate and reliable analysis results.

Example:

If incorrect data is used in analysis, it may lead to wrong business decisions.

18. Define data cleaning.

Answer:

Data cleaning is the process of identifying and correcting or removing errors, missing values, and duplicate entries in a dataset to improve data quality.

Example:

Fixing spelling mistakes in student names or removing repeated records.

19. What is data transformation?

Answer:

Data transformation is the process of converting data into a suitable format for analysis. It may include changing structure, creating new variables, or summarizing data.

Example:

Converting raw student marks into grade categories like A, B, C.

20. What is missing data?

Answer:

Missing data refers to incomplete information in a dataset where some values are not recorded or unavailable. It can affect the accuracy of analysis if not handled properly.

Example:

A student's exam score is not recorded in the dataset.

21. Define imputation in missing data handling.

Answer:

Imputation is a technique used to handle missing data by replacing missing values with estimated or calculated values based on available data. It helps maintain dataset completeness without removing useful records.

Example:

If a student's marks are missing, we can replace them with the class average score.

22. What is data cleaning used for?

Answer:

Data cleaning is used to improve the quality of data by correcting or removing errors, duplicates, and missing or inconsistent values. It ensures that data is accurate and reliable for analysis.

Example:

Correcting spelling mistakes in names or removing duplicate student records.

23. What is duplicate data?

Answer:

Duplicate data refers to repeated entries of the same record in a dataset. It can distort analysis results and must be removed during data cleaning.

Example:

A student's record appears twice in an attendance sheet.

24. What is data transformation used for?

Answer:

Data transformation is used to convert raw data into a more useful and structured format for analysis. It helps in organizing data, creating new variables, or summarizing information.

Example:

Converting numerical marks into grades like A, B, C.

25. Why is handling missing data important?

Answer:

Handling missing data is important because incomplete data can lead to incorrect or biased results. Proper handling ensures accuracy, consistency, and reliability in analysis.

Example:

If many student marks are missing, ignoring them may give an unfair result of class performance.

26. What is statistical modeling?

Answer:

Statistical modeling is the process of using mathematical and statistical techniques to analyze data and make predictions about real-world situations. It helps in understanding relationships between variables.

Example:

Predicting monthly grocery expenses based on past spending data.

27. What are the steps of model development?

Answer:

The steps of model development are:

- Define the problem
- Collect data

-
- Choose an algorithm
 - Train the model
 - Evaluate the model

Example:

To predict exam performance, we first define the problem, collect study data, build a model, train it, and test its accuracy.

28. Define linear regression.

Answer:

Linear regression is a statistical method used to find the relationship between two variables, where one variable predicts the value of another using a straight-line equation.

Example:

Predicting shop earnings based on number of customers.

29. What is logistic regression?

Answer:

Logistic regression is a statistical method used to predict categorical outcomes such as yes/no or pass/fail. It gives results in probabilities between 0 and 1.

Example:

Predicting whether a student will pass or fail based on study hours.

30. What is clustering (K-means clustering)?

Answer:

Clustering is a machine learning technique used to group similar data points together based on their characteristics. K-means clustering divides data into K number of groups based on similarity.

Example:

Grouping students into clusters based on their math and English scores (high performers and average performers).

Note:

This chapter is designed to provide a solid foundation of knowledge, with the goal of deepening understanding and encouraging further exploration of the subject. The content has been carefully selected to support effective learning and inspire students to engage with the topic more deeply.

Author: Muhammad Asghar

Purpose: To contribute to education by offering insightful, valuable content that enhances learning and understanding.

Copyright & Usage Policy

© 2026 **Studynotes360.com**. All Rights Reserved.

No part of these notes may be reproduced, redistributed, or used for commercial purposes without explicit written permission from the author. These notes are intended solely for personal study and educational use.