

UNDERSTANDING DATA

STUDY NOTES

- People take decisions based on certain data or information.
- Data is crucial for decision making.
- When we process large amount of data with computer we see a hidden trait which is otherwise not visible to humans.
- Data can be:
 - ❖ Structured or organised
 - ❖ Unstructured having no pattern
- In order to process data, it should be first collected.
- After data collection, data is stored on a storage device from where it can be retrieved easily.
- Data processing processes data to see results and analysing those results to make conclusions.
- Measure of central tendency gives us some idea about the data.
 - ❖ **Mean:** Average of numeric values of an attribute
 - ❖ **Median:** After all values are sorted, median is the middle value.
 - ❖ **Mode:** the value that occurs maximum number of times.
- Degree of diversity in a group is a measure of variability or measure of dispersion.
 - ❖ Range is the difference between the maximum and minimum values of data.
 - ❖ Standard deviation is squared difference of each value from mean value of data. Smaller value means less spread and larger value means more spread.

QUESTION BANK

MULTIPLE CHOICE QUESTIONS

1. _____ is not a suitable choice for measuring central tendency if there are exceptionally large or small values in the data.

(a) Mean	(b) Median	(c) Mode	(d) Range
----------	------------	----------	-----------
2. Which of the following is not an input for data process cycle?

(a) Data collection	(b) Results	(c) Data preparation	(d) Data entry
---------------------	-------------	----------------------	----------------
3. What type of processing is done in data process cycle?

(a) Storing data	(b) Classifying data	(c) Retrieving data	(d) All of these
------------------	----------------------	---------------------	------------------
4. After sorting data in ascending or descending order the actual central value where the given data is actually divided into two parts is known as:

(a) Mean	(b) Median	(c) Mode	(d) Range
----------	------------	----------	-----------
5. If a list has even number of values then the _____ is calculated by taking the average of the two middle values.

(a) Mean	(b) Median	(c) Mode	(d) Range
----------	------------	----------	-----------
6. _____ data value occurs maximum number of times in a given data.

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

7. Data can be structured or unstructured.
(a) True (b) False
8. After processing data can give meaningful results.
(a) True (b) False
9. What is median for 2,9,3,7,5,4,3,2,10?
(a) 2 (b) 4 (c) 7 (d) 3
10. Difference between the maximum and minimum data is known as:
(a) Mean (b) Median (c) Mode (d) Range
11. While collecting data we store it in file or database for processing.
(a) True (b) False
12. Data is crucial for:
(a) Searching (b) Sorting (c) Decision making (d) Story telling
13. In order to make a good decision, data must not be:
(a) Gathered (b) Deleted (c) Processed (d) Analysed
14. Data processing with computers is easier than manual processing because:
(a) Computer stores data in electronic form (b) Computers are fast
(c) We are slow (d) All of these
15. The word data is plural form of word:
(a) Date (b) Daily (c) Datum (d) .dat
16. What is not true for data?
(a) We are dependent on it for decision making. (b) Data processing delays decision making
(c) Data shows no possibilities or hidden trait (d) Data must be maintained and updated
17. A store of information consisting of facts assumptions and rules which an AI system can use for decision making is known as:
(a) Database (b) Knowledge base (c) Web base (d) App base
18. Collection of characters, numbers and other symbols that represents values of some situations or variables is known as:
(a) String (b) Variables (c) Data (d) Integers
19. ICT stands for:
(a) Information and computer technology (b) Information and communication technology
(c) Informatics computer technology (d) Informatics communication technology
20. What is not an example of data?
(a) Name, age, gender (b) Transaction data generated through banking etc.
(c) Documents and web page (d) None of these
21. Scientists record data of experiments so that:
(a) It can help in sorting (b) Collected data may provide information
(c) So that labs are well utilized (d) Regular procedures are followed
22. What are the two types of data on the basis of format?
(a) Sorted and unsorted (b) Structured and unstructured
(c) Recorded and unrecorded (d) Ordered and unordered
23. Which of the following is not an example of storing and analysing data for decision making?
(a) Electronic voting machine (b) Post man
(c) Library (d) Search Engine
24. Data that is organised and recorded in a well-defined format is:
(a) Sorted (b) Ordered (c) Structured (d) Perfect
25. In structured data, column represents different data for a particular parameter called:
(a) Attribute (b) Characteristic (c) Variable (d) All of these

26. A row in a table represents:
 (a) An observation of different attributes (b) One computer record
 (c) Neither (a) nor (b) (d) Both (a) and (b)
27. What is not an example of structured data?
 (a) Credit cards (b) Stock information (c) Newspaper (d) Geolocation
28. Data stored in computer in a tabular form is an example of:
 (a) Data science (b) Structured data (c) Data visualization (d) All of these
29. What is not true for structured data?
 (a) Standardized format (b) Well-defined structure
 (c) Follows persistent order (d) Cannot be stored in database
30. Which of the following is an example of unstructured data?
 (a) Email (b) Stock exchange (c) Geo location (d) Credit cards
31. Which of the following is not an example of unstructured data?
 (a) Email (b) Newspaper (c) Web pages (d) Geo location
32. Text documents, business reports, social media messages are examples of:
 (a) Irrelevant data (b) Structured data (c) Unstructured data (d) Corporate data
33. Which of the following is not a metadata for email?
 (a) Subject (b) Recipient (c) Attachment (d) Reply button
34. In an email, the subject, recipient, main body attachment etc. are all example of:
 (a) Structured data (b) Unstructured data (c) Meta data (d) No data
35. Metadata is:
 (a) Data about data (b) Data without data (c) Nested data (d) Structured data
36. Our interaction with digital media is creating a lot of:
 (a) Confusion (b) Data (c) Entertainment (d) Advancement
37. Structured data is not stored in:
 (a) CSV (b) Excel (c) MySQL (d) None of these
38. The next step after data collection is:
 (a) Storage (b) Processing (c) Mapping (d) Deleting irrelevant data.
39. Data can be stored in:
 (a) HDD (b) Pendrive (c) Memory card (d) All of these
40. In data collection, we should not use computer to:
 (a) Add data (b) Modify data (c) Steal data (d) Delete data
41. Data needs to be _____ in order to make conclusions.
 (a) Processed (b) Stored (c) Collected (d) Created
42. Arrange steps involved in Data process cycle:
 (i) Storage of data (ii) Retrieval, classification and updating of data.
 (iii) Data Collection and preparation (iv) Reporting results and processing systems
 (a) (iii), (i), (ii), (iv) (b) (i), (ii), (iii), (iv) (c) (i), (iv), (iii), (ii) (d) (iv), (i), (ii), (iii)
43. Measure of _____ is a single value that gives an idea about data:
 (a) Central tendency (b) Midpoint of data (c) Value of data (d) Cumulation of data
44. Which is not a measure of central tendency?
 (a) Mean (b) Median (c) Sum (d) Mode
45. What would be the mean of [23, 45, 87, 40, 50]?
 (a) 89 (b) 48 (c) 49 (d) 36

46. What is the median for [2,6,6,8,4,2,7,9]?
 (a) 6 (b) 12 (c) 8 (d) 1
47. Mode occurs minimum number of times in a data.
 (a) True (b) False
48. Data processing cycle involves:
 (a) Input and storage of data (b) Processing of data
 (c) Generating output (d) All of these
49. _____ is revealed when we summarise data using statistical techniques:
 (a) Data types (b) Data count (c) Data characteristics (d) None of these
50. What is not used for data summarisation?
 (a) Mean (b) Median (c) Mode (d) Product

INPUT TEXT BASED MCQs

Read the following passage and answer the following questions (51 to 54).

A measure of central tendency is a single value that gives us some idea about the data. Three most common measures of central tendency are the mean, median, and mode. Instead of looking at each individual data values, we can calculate the mean, median and mode of the data to get an idea about average, middle value and frequency of occurrence of a particular value, respectively. Selection of a measure of central tendency depends on certain characteristics of data.

51. The mean is calculated by:
 (a) Average of numeric values of an attribute. (b) Value at middle position
 (c) Value at middle position of sorted data (d) Element that is occurring maximum number of times
52. What would be median for [85,95,90,101,102,110,110,111, 116]?
 (a) 100 (b) 102 (c) 90 (d) 115
53. What would be the mode for [85,95,90,101,102,110,110,111, 116]?
 (a) 110 (b) 102 (c) 111 (d) 116
54. Find the range for [85,95,90,101,102,110,110,111, 116]
 (a) 31 (b) 32 (c) 33 (d) 34

ANSWERS

Multiple Choice Questions

- | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1. (a) | 2. (b) | 3. (d) | 4. (b) | 5. (b) | 6. (c) | 7. (a) | 8. (a) | 9. (b) | 10. (d) |
| 11. (a) | 12. (c) | 13. (b) | 14. (d) | 15. (c) | 16. (c) | 17. (b) | 18. (b) | 19. (b) | 20. (d) |
| 21. (b) | 22. (b) | 23. (b) | 24. (c) | 25. (d) | 26. (d) | 27. (c) | 28. (b) | 29. (d) | 30. (a) |
| 31. (d) | 32. (c) | 33. (d) | 34. (c) | 35. (a) | 36. (b) | 37. (d) | 38. (a) | 39. (d) | 40. (c) |
| 41. (a) | 42. (a) | 43. (a) | 44. (c) | 45. (c) | 46. (a) | 47. (b) | 48. (d) | 49. (c) | 50. (d) |

Input Text Based MCQs

51. (a) 52. (b) 53. (a) 54. (a)

HINTS/EXPLANATION

1. Mean is not a suitable choice for measuring central tendency if there are exceptionally large or small values in the data.
2. Results are part of output stage of data process cycle.
3. Processing data in data process cycle involves storing data, retrieving data, classifying and updating data.

5. If a list has even number of values then the median is calculated by taking the average of the two middle values.
6. Mode data value occurs maximum number of times in a given data.
8. After processing data gives meaningful results.
9. [2,9,3,7,5,4,3,2,10] when sorted looks like [2,2,3,3,4,5,7,9,10]
Median is 4 as it is the mid point value.
11. While collecting data, we store it in file or database for processing.
12. Data is crucial for decision making.
14. Data processing with computers is easier than manual processing because computers store data in electronic form.
16. Data processing does not delay decision making.
18. Collection of characters, numbers and other symbols that represents values of some situations or variables is known as data.
20. Options (a) to (c) are examples of data.
21. Scientists record data of experiments so that collected data may provide information.
22. The two types of data on the basis of format are structured and unstructured.
24. Data organised and recorded in a well-defined format is said to be structured.
25. In structured data, column represents attribute/characteristic/variable.
28. Data stored in computer in tabular form is an example of structured data.
29. Structured data can be stored in the database hence (d) is not true.
31. Geolocation is not an example of unstructured data.
34. In an email, the subject, recipient, main body attachment etc. are all example of metadata.
35. Our interaction with digital media is creating a lot of data.
36. Our interaction with digital media is creating a lot of data.
38. Next step after data collection is data storage.
39. Data can be stored in HDD, pen drive and memory card.
42. (i) Data Collection and preparation (ii) Storage of data
(iii) Retrieval, classification and updating of data. (iv) Reporting results and processing systems
45. In the given list [23, 45, 87, 40, 50], $23 + 45 + 87 + 40 + 50 = 245$. Total number of elements = 5. Therefore, $245/5 = 49$.
47. mode data value occurs maximum number of times in a given data.
48. Data processing cycle involves:
(a) Input and storage of data (b) Processing of data
(c) Generating output
49. Data characteristic is revealed when we summarise data using statistical techniques.