

1. Data science is the process of diverse set of data through ?
 - A. Organizing data
 - B. Processing data
 - C. Analyzing data
 - D. All of the above
2. 2Which of the following language is used in Data science?
 - A.C
 - B. C++
 - C. R
 - D. Ruby
3. A _____ is a structured representation of data.
 - A. XML
 - B. Email
 - C. Text file
 - D. Database table
4. Which of the following is one of the key data science skills?
 - A. Statistics
 - B. Machine Learning
 - C. Data Visualization
 - D. All of the above
5. Which of the following step is performed by data scientist after acquiring the data?
 - A.Data Cleaning
 - B. Data Integration
 - C. Data Replication
 - D. All of the above
6. Which of the following is correct skills for a Data Scientist?
 - A. Probability & Statistics
 - B. Machine Learning / Deep Learning
 - C. Data Wrangling
 - D. All of the above.
7. Unstructured data is not organized.
 - A. TRUE
 - B. FALSE
 - C. Can be true or false
 - D. Can not say

8. Which of the following is not a application for data science?

- A. Recommendation Systems
- B. Image & Speech Recognition
- C. Online Price Comparison
- D. Privacy Checker

9. Normal Distribution is also known as_____.

- A. Class Distribution
- B. Gaussian Distribution
- C. Permanent Distribution
- D. None of the above

10. Some of the important libraries of Python that are used in Data Science are _____.

- A. Pandas
- B. Numpy
- C. Both A and B
- D. None of the above

11. an _____ is an observation that lies an abnormal distance from other in a random sample from a population.

- A. Range
- B. Quartiles
- C. Outliers
- D. None

12. Which of the following is the common goal of statistical modelling?

- A. Inference
- B. Summarizing
- C. Subsetting
- D. None of the above

13. Chi-square test is a_____ method.

- A. decision tree
- B. clustering
- C. hypothesis testing
- D. correlation testing

14. The mean, median and mode of the t-distribution are equal to

- A. Zero
- B. One
- C. Null
- D. Minus one

15. In hypothesis testing, the hypothesis which is tentatively assumed to be true is called the_____.

- A. correct hypothesis
 - B. null hypothesis
 - C. alternative hypothesis
 - D. level of significance
16. . Baye's estimator is always a function of.
- A. Minimal sufficient statistic
 - B. Sufficient statistic
 - C. Both A and B
 - D. None of these
17. . Which of the following testing is concerned with making decisions using data?
- A. Probability
 - B. Hypothesis
 - C. Causal
 - D. None of the mentioned
18. Point out the wrong statement.
- A. The normal distribution is asymmetric and peaked about its mode
 - B. A constant times a normally distributed random variable is also normally distributed
 - C. Sample means of normally distributed random variables are again normally distributed
 - D. None of the mentioned
19. The mean, median and mode of the t-distribution are equal to
- A. Zero
 - B. One
 - C. Null
 - D. Minus one
20. In hypothesis testing, the hypothesis which is tentatively assumed to be true is called the_____.
- A. correct hypothesis
 - B. null hypothesis
 - C. alternative hypothesis
 - D. level of significance
21. The standard deviation of a data set is the positive square root of the _____.
- A. Sample

- B. Variance
 - C. Population
 - D. Outliers
22. The square root of the variance is called the _____ deviation.
- A. empirical
 - B. mean
 - C. continuous
 - D. standard
23. . Concept hierarchies are a form of data smoothing. that can also be used for data
- A. binarization
 - B. discretization
 - C. missing
 - D. all of these
24. Which of the following is NOT data preprocessing techniques ?
- A. Data cleaning
 - B. Data integration
 - C. Data Handling
 - D. Data transformation
- 25.._____ Normalization performs a linear transformation on the original data.
- A. Z-score
 - B. Min-max
 - C. Smoothing
 - D. Aggregation