

Foundations of Data Science

Data Science · Answer Key · 20 Questions

1. Which of the following is a primary goal of data science?

- A) To create fictional narratives
- B) To extract knowledge and insights from data**
- C) To develop new forms of art
- D) To predict the weather with 100% accuracy

2. What type of data is represented by categories or labels, such as 'red', 'blue', or 'green'?

- A) Numerical data
- B) Ordinal data
- C) Categorical data**
- D) Time-series data

3. Which statistical measure represents the average of a dataset?

- A) Median
- B) Mode
- C) Standard Deviation
- D) Mean**

4. In data science, what does 'feature engineering' involve?

- A) Discarding irrelevant data
- B) Creating new features from existing ones**
- C) Visualizing raw data
- D) Training machine learning models

5. What is the term for a variable that is manipulated or changed in an experiment?

- A) Dependent variable
- B) Independent variable**
- C) Confounding variable
- D) Extraneous variable

6. Which programming language is widely used in data science for its extensive libraries?

- A) Java
- B) C++
- C) Python**
- D) Ruby

7. What is the process of identifying and removing errors or inconsistencies in data called?

- A) Data augmentation
- B) Data transformation
- C) Data cleaning**
- D) Data aggregation

8. What does 'overfitting' in machine learning refer to?

- A) A model that performs poorly on training data
- B) A model that performs poorly on new, unseen data**
- C) A model that is too simple to capture patterns
- D) A model that requires very little data

9. Which of the following is a common type of data visualization?

- A) Algorithm
- B) Database
- C) Scatter plot**
- D) Code

10. What is the primary purpose of a 'dataset' in data science?

- A) To describe a single event
- B) To store a collection of related data points**
- C) To define a mathematical formula
- D) To represent a user's opinion

11. What does 'algorithm' mean in the context of data science?

- A) A type of database
- B) A set of rules or instructions to solve a problem**
- C) A method for storing data
- D) A visual representation of data

12. What is the 'median' of a dataset?

- A) The most frequent value
- B) The average of the values
- C) The middle value when the data is ordered**
- D) The range of the data

13. Which of these is a supervised learning technique?

- A) Clustering
- B) Dimensionality Reduction
- C) Classification**
- D) Association Rule Mining

14. What is the process of combining multiple data sources into a single, unified view called?

- A) Data mining
- B) Data integration**
- C) Data warehousing
- D) Data analysis

15. What is a 'hypothesis' in scientific research?

- A) A confirmed fact
- B) A testable explanation or prediction**
- C) A final conclusion
- D) A data visualization

16. Which type of variable can only take specific, separate values, such as the number of students in a class?

- A) Continuous variable
- B) Discrete variable**
- C) Nominal variable
- D) Interval variable

17. What is the goal of 'exploratory data analysis' (EDA)?

- A) To finalize a model
- B) To create production-ready code
- C) To understand the main characteristics of a dataset**
- D) To document data sources

18. Which statistical concept measures the spread or dispersion of data around the mean?

- A) Mean
- B) Median
- C) Standard Deviation**
- D) Mode

19. What is 'machine learning'?

- A) A type of computer hardware
- B) A field of study that gives computers the ability to learn without being explicitly programmed**
- C) A method for physical computing
- D) A technique for data storage

20. What is a 'database' in data science?

- A) A single numerical value
- B) A tool for drawing charts
- C) An organized collection of data**
- D) A programming language