

Historical Foundations of Data Science

Data Science History · Answer Key · 18 Questions

1. In 1962, which statistician published 'The Future of Data Analysis,' arguably the first paper to propose data analysis as a distinct field separate from classical statistics?

- A) John Tukey**
- B) Ronald Fisher
- C) C.R. Rao
- D) Jerzy Neyman

2. Who coined the term 'Data Science' in his 1974 book 'Concise Survey of Computer Methods' while advocating for a change in statistics terminology?

- A) Peter Naur**
- B) William S. Cleveland
- C) C.F. Jeff Wu
- D) Andrew Ng

3. The term 'KDD' (Knowledge Discovery in Databases) was formally introduced in a workshop in 1989. Where was this inaugural international workshop held?

- A) IJCAI in Detroit**
- B) AAAI in Seattle
- C) SIGMOD in Portland
- D) ICML in Pittsburgh

4. Which famous 19th-century statistician was the first to use the term 'correlation' in his 1888 paper, 'Co-relations and their Measurement, chiefly from Anthropometric Data'?

- A) Francis Galton**
- B) Karl Pearson
- C) Charles Spearman
- D) Thomas Bayes

5. In 1996, at the first Knowledge Discovery and Data Mining conference (KDD-96), which researcher presented the seminal paper on the 'Random Forest' ensemble method?

- A) Leo Breiman**
- B) Yann LeCun
- C) Geoffrey Hinton
- D) Judea Pearl

6. What was the primary focus of the 1956 Dartmouth Summer Research Project on Artificial Intelligence, which laid the groundwork for future data-driven machine learning?

A) Symbolic logic and heuristic search

- B) Neural networks and backpropagation
- C) Bayesian statistics and inference
- D) Database normalization theory

7. Which algorithm, central to modern data science for identifying clusters, was proposed by Lloyd in 1957 but not published in a journal until 1982?

A) K-Means

- B) DBSCAN
- C) Mean Shift
- D) Expectation-Maximization

8. The 'Perceptron' was invented in 1958 by Frank Rosenblatt. On what specific machine was the first hardware implementation of this algorithm executed?

A) IBM 704

B) Mark I Perceptron

- C) UNIVAC I
- D) ENIAC

9. In 1972, the Nelder and Wedderburn paper introduced which statistical framework that generalized linear regression, ANOVA, and logistic regression?

A) Generalized Linear Models

- B) Support Vector Machines
- C) Hidden Markov Models
- D) Factor Analysis

10. Which mathematician and computer scientist authored 'The Art of Computer Programming' in 1968, establishing the formal analysis of algorithms crucial for data science?

A) Donald Knuth

- B) Edsger W. Dijkstra
- C) John von Neumann
- D) Claude Shannon

11. Before modern Big Data tools, which 1944 development by Claude Shannon provided the mathematical foundation for quantifying information and data transmission?

A) Information Theory

- B) Set Theory
- C) Game Theory
- D) Queueing Theory

12. The term 'Logistic Regression' was first introduced in 1958 by D.R. Cox. What was the original context of his research?

A) Bioassay and dose-response

- B) Stock market volatility
- C) Astrophysical measurements
- D) Linguistic frequency analysis

13. In 1995, what major breakthrough in statistical learning theory was introduced by Corinna Cortes and Vladimir Vapnik?

A) Support Vector Networks

- B) Recurrent Neural Networks
- C) Gradient Boosting Machines
- D) Deep Belief Networks

14. Which landmark 1936 paper by R.A. Fisher introduced Linear Discriminant Analysis for classifying biological data?

A) The use of multiple measurements in taxonomic problems

- B) Statistical Methods for Research Workers
- C) The Design of Experiments
- D) The Genetical Theory of Natural Selection

15. The term 'Big Data' was used in a 1997 paper by Michael Lesk. In what specific context did he use it to describe the difficulty of managing information?

A) Scaling storage systems

- B) Predicting weather patterns
- C) Genomic sequencing
- D) High-frequency trading

16. In the history of time-series analysis, who introduced the 'Autoregressive Integrated Moving Average' (ARIMA) models in their 1970 text?

A) Box and Jenkins

- B) Engle and Granger
- C) Hamilton and Sims
- D) Markov and Wiener

17. The first implementation of a programming language specifically for statistics, 'S', was developed at which research facility in 1976?

A) Bell Labs

B) Xerox PARC

C) MIT AI Lab

D) Stanford Research Institute

18. Which statistician formulated the 'Method of Least Squares' in 1805, which remains the cornerstone of linear regression in data science?

A) Adrien-Marie Legendre

B) Carl Friedrich Gauss

C) Pierre-Simon Laplace

D) Thomas Bayes