MCQ's

1. What is Data Science primarily concerned with?
   a. Designing new computer hardware
   b. The study of planets and galaxies
   c. Extracting knowledge and insights from data
   d. Creating video games
   Answer: c) Extracting knowledge and insights from data

2. The term "Big Data" refers to:
   a. Large datasets that cannot be processed by traditional computing techniques
   b. The amount of data stored in a big building
   c. Data about large objects
   d. Complex data structures in programming
   Answer: a) Large datasets that cannot be processed by traditional computing techniques

3. Getting past the hype of Data Science involves:
   a. Ignoring data analysis
   b. Focusing only on theoretical aspects
   c. Understanding the practical applications and limitations of Data Science
   d. Assuming all data is useful
   Answer: c) Understanding the practical applications and limitations of Data Science

4. "Datafication" is a process that involves:
   a. Turning physical data into digital data
   b. Transforming aspects of the physical world into quantified data
   c. Deleting unnecessary data files

d. Data encryption and security

Answer: b) Transforming aspects of the physical world into quantified data

5. Statistical inference is used to:

a. Make predictions about game scores

b. Design computer algorithms

c. Draw conclusions about populations based on samples

d. Calculate the speed of light

Answer: c) Draw conclusions about populations based on samples

6. A "population" in statistical terms refers to:

a. The total number of people in a country

b. The complete set of data points or individuals considered in a study

c. A group of similar species in biology

d. The number of people living in urban areas

Answer: b) The complete set of data points or individuals considered in a study

7. Statistical modeling is important because it:

a. Helps in creating video games

b. Allows for predictions and inferences based on data

c. Is only used in financial markets

d. Creates physical models of buildings

Answer: b) Allows for predictions and inferences based on data

8. Overfitting a model means:

a. The model is too simple

b. The model performs well on training data but poorly on unseen data

c. The model is universally applicable

d. The model uses too little data

Answer: b) The model performs well on training data but poorly on unseen data

9. R is a programming language used for:

a. Writing operating systems

b. Statistical computing and graphics

c. Developing mobile apps

d. 3D modeling

Answer: b) Statistical computing and graphics

10. Which of the following is NOT a basic data type in R?

a. Vectors

b. Lists

c. Tables

d. Matrices

Answer: c) Tables

11. Probability distributions in statistical modeling are used to:

a. Predict weather patterns exclusively

b. Describe the likelihood of outcomes

c. Store large datasets efficiently

d. Encrypt sensitive information

Answer: b) Describe the likelihood of outcomes

12. In the context of R, what does the R Environment Setup involve?

a. Installing R and configuring the software to work with other programming languages

b. Creating virtual reality environments

c. Setting up a gaming console

d. Configuring a router for internet access

Answer: a) Installing R and configuring the software to work with other programming languages

13. Which statement best describes the current landscape of perspectives in Data Science?

a. It is unanimously agreed upon with no differing viewpoints

b. It is focused solely on the use of big data

c. It encompasses a wide range of disciplines and methodologies

d. It is limited to statistical analysis

Answer: c) It encompasses a wide range of disciplines and methodologies

14. Fitting a model in statistics means:

a. Adjusting the model so it can physically fit into a designated space

b. Tailoring the model to match the training data as closely as possible

c. Choosing the smallest model for computational efficiency

d. Making the model visually appealing

Answer: b) Tailoring the model to match the training data as closely as possible

15. Programming with R involves:

a. Only the use of drag-and-drop interfaces

b. Writing code to perform statistical analysis and graphical representation

c. Developing mobile applications

d. 3D printing

Answer: b) Writing code to perform statistical analysis and graphical representation

16. A sample in statistical terms is:

a. A small music clip

b. A subset of a population used for analysis

c. An example of a database

d. A type of computer virus

Answer: b) A subset of a population used for analysis

17. The main purpose of statistical inference is to:

a. Create beautiful graphs

b. Provide exact measurements of population parameters

c. Estimate population parameters based on a sample

d. Record data in a laboratory notebook

Answer: c) Estimate population parameters based on a sample

18. Big Data and Data Science hype often leads to:

a. Underestimation of data's value

b. An accurate understanding of what data can do

c. Overexpectations and misconceptions about what data science can achieve

d. The belief that traditional data analysis methods are sufficient

Answer: c) Overexpectations and misconceptions about what data science can achieve

19. What does "getting past the hype" of Data Science require?

a. Ignoring all data analysis techniques

b. A realistic understanding of its capabilities and limitations

c. Focusing solely on Big Data technologies

d. Abandoning statistical methods

Answer: b) A realistic understanding of its capabilities and limitations

20. What is the significance of "Datafication" in today's world?

a. It signifies the end of traditional data analysis

b. It represents the transformation of social actions into online quantifiable data

c. It is a process of making data less important

d. It focuses only on personal data collection

Answer: b) It represents the transformation of social actions into online quantifiable data

21. Which of the following best represents the idea behind statistical modeling?

a. Ignoring data to make decisions based on intuition

b. Using mathematical formulas to make predictions about the weather only

c. Utilizing data, statistical methods, and probability theories to simulate real-world processes

d. Creating visual art from data

Answer: c) Utilizing data, statistical methods, and probability theories to simulate real-world processes

22. What role does probability distribution play in data science?

a. It determines the color scheme for data visualizations

b. It is used to understand the spread and central tendency of data

c. It predicts the exact outcomes of future events

d. It designs the user interface for data applications

Answer: b) It is used to understand the spread and central tendency of data

23. How does overfitting affect the performance of a statistical model?

a. It ensures the model performs optimally on all future data

b. It makes the model too generalized to make accurate predictions

c. It reduces the model's ability to generalize from the training data to new data

d. It decreases the time required to train the model

Answer: c) It reduces the model's ability to generalize from the training data to new data

24. In R programming, vectors are:

a. 3D models for graphic design

b. Tools for hardware repair

c. Data structures that contain elements of the same type

d. Functions for web development

Answer: c) Data structures that contain elements of the same type

25. When setting up the R environment, it is important to:

a. Ensure compatibility with video game consoles

b. Install and configure relevant packages and software for data analysis

c. Connect the computer to a satellite dish

d. Prepare physical lab equipment

Answer: b) Install and configure relevant packages and software for data analysis

26. The term 'population' in data science refers to:

a. The total number of animals in a forest

b. Every individual or item within the scope of an investigation

c. A group of people living in a city

d. The data collected from a sample

Answer: b) Every individual or item within the scope of an investigation

27. Samples are used in statistical analysis because:

a. It is impractical to collect data from the entire population

b. Samples are more accurate than populations

c. Populations are always too small to analyze

d. It is cheaper to study samples than to play video games

Answer: a) It is impractical to collect data from the entire population

28. A key aspect of statistical inference is:

a. Assuming that samples and populations are the same

b. Making predictions about a population based on the study of a sample

c. Ignoring variability within data

d. Focusing solely on qualitative data

Answer: b) Making predictions about a population based on the study of a sample

29. Fitting a statistical model involves:

a. Literally shrinking the model to fit into a smaller space

b. Making subjective judgments without data

c. Adjusting the model parameters to best capture the relationship between variables

d. Designing the graphical interface of the model

Answer: c) Adjusting the model parameters to best capture the relationship between variables

30. Overfitting can be avoided by:

a. Using simpler models with fewer parameters

b. Collecting more data from the internet

c. Ignoring all data except for the largest values

d. Focusing exclusively on the model's performance on training data

Answer: a) Using simpler models with fewer parameters

31. The main purpose of using R in data science is to:

a. Develop large-scale industrial machinery

b. Perform statistical analysis and graphical representation of data

c. Play video games

d. Build social media platforms

Answer: b) Perform statistical analysis and graphical representation of data

32. Lists in R are:

a. Data structures that can only contain numbers

b. Used exclusively for text processing

c. Data structures that can contain elements of different types

d. Not used in R programming

Answer: c) Data structures that can contain elements of different types

33. The process of 'Datafication' transforms:

a. Physical books into digital formats exclusively

b. Any aspect of the living world into quantifiable data

c. Data into physical objects

d. Digital photos into paintings

Answer: b) Any aspect of the living world into quantifiable data

34. In the context of Big Data, 'Volume' refers to:

a. The amount of data that can fit into a traditional database

b. The sound level of data centers

c. The size of the datasets which are too large to be handled by standard databases

d. The physical dimensions of data storage devices

Answer: c) The size of the datasets which are too large to be handled by standard databases

35. What does the 'Variety' aspect of Big Data signify?

a. The number of different languages spoken by data scientists

b. The types of snacks available in data centers

c. The range of data types and sources, including structured and unstructured data

d. The various sizes of databases

Answer: c) The range of data types and sources, including structured and unstructured data

36. In Data Science, 'Velocity' refers to:

a. The speed at which a hard drive spins

b. The rate at which data flows into an organization from various sources

c. The speed of the internet connection in a data center

d. The pace at which data scientists work

Answer: b) The rate at which data flows into an organization from various sources

37. A 'probability distribution' in statistics is

a. A method for distributing prizes in a lottery

b. A graphical representation of the frequencies of musical notes in a song

c. A mathematical function that provides the probabilities of occurrence of different possible outcomes

d. A distribution method for software updates

Answer: c) A mathematical function that provides the probabilities of occurrence of different possible outcomes

38. 'Matrices' in R are used to

a. Control the movements of characters in video games

b. Store two-dimensional data structures with elements of the same type

c. Create three-dimensional animations

d. Design user interfaces

Answer: b) Store two-dimensional data structures with elements of the same type

39. The current landscape of perspectives in Data Science is characterized by

a. Uniformity and agreement across all sectors

b. Diverse opinions, methodologies, and applications

c. A focus solely on theoretical research

d. An exclusive emphasis on technology companies

Answer: b) Diverse opinions, methodologies, and applications

40. The concept of 'Statistical Inference' allows data scientists to

a. Ignore data and make decisions based on intuition

b. Make predictions or decisions about a population based on sample data

c. Focus exclusively on data visualization

d. Study the behavior of animals

Answer: b) Make predictions or decisions about a population based on sample data

41. R's programming environment is favored for Data Science because it:

a. Is only suitable for game development

b. Offers a wide variety of packages for statistical analysis and graphical models

c. Cannot handle large datasets

d. Is the only programming language available

Answer: b) Offers a wide variety of packages for statistical analysis and graphical models

42. Overfitting a model can lead to

a. A perfect fit for any dataset

b. Improved model performance on future, unseen data

c. Poor generalization to new data

d. A decrease in computational time

Answer: c) Poor generalization to new data

43. The main advantage of using samples in statistics is to

a. Completely eliminate the need for data collection

b. Make data analysis more time-consuming and expensive

c. Provide insights into a population without studying every individual

d. Increase the complexity of statistical models

Answer: c) Provide insights into a population without studying every individual

44. A key feature of R that distinguishes it from other programming languages is its

a. Inability to perform arithmetic operations

b. Focus on statistical computing and graphics

c. Use exclusively in web development

d. Lack of support for data analysis

Answer: b) Focus on statistical computing and graphics

45. The process of transforming every aspect of a physical world into data is known as

a. Virtualization

b. Datafication

c. Digitization

d. Data encryption

Answer: b) Datafication

46. The concept of 'Big Data' is primarily about:

a. The physical size of data storage devices

b. The amount, speed, and variety of data being generated

c. The popularity of data science as a field

d. The budget allocated to IT departments

Answer: b) The amount, speed, and variety of data being generated

47. In statistical modeling, the term 'fitting a model' typically involves

a. Physical assembly of computer parts

b. Manual entry of data into spreadsheets

c. Adjusting the model to accurately reflect the observed data

d. Painting models for visual aesthetics

Answer: c) Adjusting the model to accurately reflect the observed data

48. What does 'Data Science' aim to achieve with 'Big Data'?

a. To store as much data as possible without analysis

b. To make data inaccessible for security reasons

c. To extract insights and knowledge from large volumes of data

d. To replace traditional science fields

Answer: c) To extract insights and knowledge from large volumes of data

49. The prevention of overfitting can be aided by

a. Ignoring validation and testing datasets

b. Using complex models with many parameters

c. Applying techniques like cross-validation and regularization

d. Collecting less data

Answer: c) Applying techniques like cross-validation and regularization

50. Why is programming with R particularly suitable for data analysis?

a. It is the only language that can process text files

b. It offers a comprehensive ecosystem of packages for data manipulation, calculation, and graphical display

c. It is faster than all other programming languages

d. It automatically prevents overfitting in statistical models

Answer: b) It offers a comprehensive ecosystem of packages for data manipulation, calculation, and graphical display

51. Which type of data refers to qualitative characteristics of objects?

a) Numeric

b) Ordinal

c) Attributes

d) Continuous

Answer: c) Attributes

52. A binary attribute can have:

a) Multiple values

b) Only one value

c) Exactly two values

d) An infinite number of values

Answer: c) Exactly two values

53. Which of the following attributes is ordered?

a) Nominal

b) Binary

c) Ordinal

d) Asymmetric

Answer: c) Ordinal

54. What is a distinguishing feature of nominal attributes?

a) They have a natural order.

b) They can only have two values.

c) They categorize data without order.

d) They are numerical.

Answer: c) They categorize data without order.

55. Which of the following is a measure of central tendency?

a) Range

b) Variance

c) Median

d) Quartile

Answer: c) Median

56. What does variance measure?

a) Central tendency

b) Spread of data points

c) Order of data points

d) Peaks of a data set

Answer: b) Spread of data points

57. Continuous attributes are characterized by:

a) Integer values only

b) Values that can take any real number

c) Ordered categorical values

d) Non-numeric values

Answer: b) Values that can take any real number

58. The interquartile range measures:

a) Variability within the middle 50% of a dataset

b) The total range of the dataset

c) The variance of the dataset

d) The mean value of the dataset

Answer: a) Variability within the middle 50% of a dataset

59. Which attribute type is characterized by the absence of a mathematical notion of difference between values?

a) Ordinal

b) Nominal

c) Discrete

d) Continuous

Answer: b) Nominal

60. In statistical data description, the mode refers to:

a) The value that appears most frequently

b) The average value

c) The middle value when data is sorted

d) The difference between the highest and lowest values

Answer: a) The value that appears most frequently

61. Asymmetric attributes are those:

a) That can be easily mirrored

b) Where the mode equals the mean

c) That do not have a symmetric distribution of values

d) With exactly two values

Answer: c) That do not have a symmetric distribution of values

62. Discrete attributes are best described as:

a) Having values that are countable

b) Having values that can take on any real number

c) Being unordered categorical values

d) Being ordered but not numerically expressive

Answer: a) Having values that are countable

63. Which of the following best describes numeric attributes?

a) Qualitative

b) Quantitative

c) Unordered

d) Binary

Answer: b) Quantitative

64. The range of a dataset is defined as:

a) The average of all values

b) The difference between the largest and smallest values

c) The middle value of the dataset

d) The most frequent value in the dataset

Answer: b) The difference between the largest and smallest values

65. Standard deviation is used to:

a) Determine the central value of a dataset

b) Measure the spread or variability of a dataset

c) Identify the range of the dataset

d) Classify the data into quartiles

Answer: b) Measure the spread or variability of a dataset

66. Which of the following best defines an ordinal attribute?

a) An attribute with binary outcomes

b) An attribute that categorizes data without implying order

c) An attribute that can be ranked or ordered

d) An attribute with continuous numerical values

Answer: c) An attribute that can be ranked or ordered

67. How is mean calculated in a dataset?

a) By adding all the values and dividing by the number of values

b) By finding the value that appears most frequently

c) By determining the middle value when the data is ordered

d) By subtracting the smallest value from the largest value

Answer: a) By adding all the values and dividing by the number of values

68. A graphic display useful for showing the dispersion of data is:

a) Pie chart

b) Line graph

c) Box plot

d) Histogram

Answer: c) Box plot

Answer: c) Box plot

69. Which statistical measure is not affected by extreme values?

a) Mean

b) Median

c) Range

d) Standard deviation

Answer: b) Median

70. Which of the following is an example of a discrete attribute?

a) Height of students in a class

b) Weight of apples in a basket

c) Number of books on a shelf

d) Temperature in a room

Answer: c) Number of books on a shelf

71. The variance of a dataset provides information about:

a) The dataset's central tendency

b) The consistency of dataset values

c) The order of dataset values

d) The dataset's distribution shape

Answer: b) The consistency of dataset values

72. Which attribute type does not assume any order or ranking among its categories?

a) Binary

b) Nominal

c) Ordinal

d) Numeric

Answer: b) Nominal

73. In which type of data are the differences between values meaningful and quantifiable?

a) Nominal

b) Ordinal

c) Discrete

d) Continuous

Answer: d) Continuous

74. The first quartile of a dataset is:

a) The median of the dataset

b) The highest value of the dataset

c) The value below which 25% of the data falls

d) The average value of the dataset

Answer: c) The value below which 25% of the data falls

75. What characteristic defines a binary attribute?

a) It describes a quality without inherent numeric value.

b) It can take only two possible values.

c) It categorizes data in a naturally ordered way.

d) It measures data in a continuous manner.

Answer: b) It can take only two possible values.

76. What does the mode measure in a dataset?

a) Spread

b) Central tendency

c) Order

d) Variability

Answer: b) Central tendency

77. Which of the following is a characteristic of ordinal attributes?

a) Countability

b) Natural ordering

c) Binary values

d) Continuity

Answer: b) Natural ordering

78. How is the interquartile range (IQR) calculated?

a) Q3 - Q1

b) Maximum value - Minimum value

c) Mean value of the dataset

d) Sum of all values / Number of values

Answer: a) Q3 - Q1

79. Which term refers to the total span of values in a dataset, from the smallest to the largest?

a) Variance

b) Range

c) Median

d) Mode

Answer: b) Range

80. A dataset's standard deviation is a measure of:

a) How spread out the values are from the mean

b) The average value of the data

c) The middle value of a sorted dataset

d) The dataset's total span of values

Answer: a) How spread out the values are from the mean

81. Which of the following data types is considered quantitative?

a) Nominal

b) Ordinal

c) Binary

d) Numeric

Answer: d) Numeric

82. The mean is often referred to as:

a) The average

b) The midpoint

c) The most frequent value

d) The data range

Answer: a) The average

83. Which graph is best suited for displaying the frequency of nominal data?

a) Histogram

b) Pie chart

c) Scatter plot

d) Box plot

Answer: b) Pie chart

84. Which measure of central tendency can be used for nominal data?

a) Mean

b) Median

c) Mode

d) Variance

Answer: c) Mode

85. Continuous data is characterized by which of the following?

a) It can only take integer values.

b) It includes values that cannot be counted.

c) It cannot be measured, only categorized.

d) It represents categories with a natural order.

Answer: b) It includes values that cannot be counted.

86. How are ordinal attributes different from nominal attributes?

a) Ordinal attributes can be measured, while nominal cannot.

b) Nominal attributes have a natural ordering, while ordinal do not.

c) Ordinal attributes have a natural ordering, while nominal do not.

d) Nominal attributes can take on numerical values, while ordinal cannot.

Answer: c) Ordinal attributes have a natural ordering, while nominal do not.

87. The term "quartile" refers to:

a) A measure of variability based on dividing data into four equal parts

b) The average value of a dataset

c) The difference between the highest and lowest values in a dataset

d) A graphical representation of statistical data

Answer: a) A measure of variability based on dividing data into four equal parts

88. Which of the following is not a characteristic of asymmetric attributes?

a) Unequal distribution of data values

b) Data values can be ranked

c) Absence of a central peak in the data distribution

d) Presence of a predefined order among values

Answer: d) Presence of a predefined order among values

89. In data analysis, what is the significance of the median?

a) It is the most frequently occurring value.

b) It divides the dataset into two equal halves.

c) It is the sum of all data values divided by the number of values.

d) It represents the total range of the dataset.

Answer: b) It divides the dataset into two equal halves.

90. What does a box plot display?

a) Frequency distribution of nominal data

b) Correlation between two numeric variables

c) Central tendency and dispersion of data

d) Time series data

Answer: c) Central tendency and dispersion of data

91. Nominal attributes are best analyzed using which statistical measure?

a) Mean

b) Median

c) Mode

d) Standard deviation

Answer: c) Mode

92. Discrete data can be described as:

a) Values that can take any number within a range

b) Values that represent counts of items

c) Values that are measured, not counted

d) Values that have a natural order

Answer: b) Values that represent counts of items

93. What does standard deviation help to understand in a dataset?

a) The most common value

b) The average value

c) The dispersion or spread of values around the mean

d) The sequence of data points

Answer: c) The dispersion or spread of values around the mean

94. The mode is particularly useful in which kind of data?

a) Continuous

b) Ordinal

c) Nominal

d) Binary

Answer: c) Nominal

95. Variance in a dataset measures:

a) The average distance of each data point from the mean

b) The difference between the highest and lowest data points

c) The data point in the middle of the dataset

d) The frequency of the most occurring data point

Answer: a) The average distance of each data point from the mean

96. When is the median more informative than the mean?

a) When the dataset is symmetric

b) When the dataset contains outliers

c) When the dataset is large

d) When the dataset values are all the same

Answer: b) When the dataset contains outliers

97. Continuous attributes can be divided into which two subcategories?

a) Nominal and ordinal

b) Binary and nominal

c) Interval and ratio

d) Discrete and categorical

Answer: c) Interval and ratio

98. Which attribute type implies a measure of quantity that can vary infinitely within a given range?

a) Ordinal

b) Nominal

c) Continuous

d) Discrete

Answer: c) Continuous

99. The process of categorizing nominal attributes into distinct groups without implying any order is called:

a) Ranking

b) Sorting

c) Classification

d) Sequencing

Answer: c) Classification

100. A characteristic feature of ordinal data compared to nominal data is:

a) The ability to perform arithmetic operations on ordinal data

b) The presence of a meaningful order or ranking

c) The limitation to binary values

d) The categorization without any order

Answer: b) The presence of a meaningful order or ranking

101. How do you create a numeric vector in R?

a) c(1, 2, 3, 4)

b) vector(1, 2, 3, 4)

c) list(1, 2, 3, 4)

d) matrix(1, 2, 3, 4)

Answer: a) c(1, 2, 3, 4)

102. Which operation can be used to add two vectors of equal length in R?

a) +

b) %

c) &

d) |

Answer: a) +

103. How can you subset the third element from a vector `v` in R?

a) v[3]

b) v(3)

c) v{3}

d) subset(v, 3)

Answer: a) v[3]

104. What function is used to create a matrix in R?

a) matrix()

b) c()

c) list()

d) array()

Answer: a) matrix()

105. How can you name the rows of a matrix `m` in R?

a) rownames(m) <- c("row1", "row2")

b) m.rows <- c("row1", "row2")

c) names(m)[1] <- c("row1", "row2")

d) m[rownames] <- c("row1", "row2")

Answer: a) rownames(m) <- c("row1", "row2")

106. What does the R function `factor()` do?

a) Creates a list

b) Creates a factor for categorical data

c) Transforms a vector into a matrix

d) Converts numeric data to character data

Answer: b) Creates a factor for categorical data

107. Which of the following is true about ordered factors in R?

a) They cannot have levels

b) They are used for numeric data only

c) They represent categorical data with an inherent order

d) They are identical to vectors

Answer: c) They represent categorical data with an inherent order

108. To subset all elements of a vector `v` except the fourth, which expression is correct?

a) v[-4]

b) v[4]

c) v[!4]

d) v[4, drop = FALSE]

Answer: a) v[-4]

109. How do you create a 3x3 matrix with elements from 1 to 9 in R?

a) matrix(1:9, nrow=3)

b) matrix(c(1,2,3,4,5,6,7,8,9), 3, 3)

c) array(1:9, dim=c(3,3))

d) Both a and c are correct

Answer: d) Both a and c are correct

110. In R, how can you convert a vector to a factor with two levels, "High" and "Low"?

a) factor(vector, levels = c("High", "Low"))

b) as.factor(vector, c("High", "Low"))

c) factor(vector, labels = c("High", "Low"))

d) vector(factor, levels = c("High", "Low"))

Answer: c) factor(vector, labels = c("High", "Low"))

111. How do you retrieve the class of an object in R?

a) get.class(object)

b) class(object)

c) typeof(object)

d) object.class()

Answer: b) class(object)

112. Which function is used to summarize the levels of a factor in R?

a) summary()

b) levels()

c) factor_sum()

d) describe()

Answer: a) summary()

113. How can you check if two ordered factors are identical in R?

a) identical(factor1, factor2)

b) factor1 == factor2

c) all.equal(factor1, factor2)

d) compare(factor1, factor2)

Answer: c) all.equal(factor1, factor2)

114. What does the expression `v <- 1:5; v * 2` in R produce?

a) A vector with elements 2, 4, 6, 8, 10

b) A vector with elements 1, 2, 3, 4, 5

c) An error

d) None of the above

Answer: a) A vector with elements 2, 4, 6, 8, 10

115. When creating a data frame in R, which function is used?

a) data.frame()

b) dataframe()

c) create.data.frame()

d) new.data.frame()

Answer: a) data.frame()

116. How can you subset rows 2 and 3 from a matrix `m` in R?

a) m[2:3,]

b) m[, 2:3]

c) subset(m, 2:3)

d) m.rows[2:3]

Answer: a) m[2:3,]

117. Which arithmetic operation can be performed on two vectors of unequal length in R?

a) None

b) Addition

c) Multiplication

d) Both b and c via recycling

Answer: d) Both b and c via recycling

118. To create an array in R with dimensions 2x3x4, which function call is correct?

a) array(1:24, c(2, 3, 4))

b) matrix(1:24, nrow=2, ncol=3)

c) c(1:24, dim = c(2, 3, 4))

d) list(1:24, dim = c(2, 3, 4))

Answer: a) array(1:24, c(2, 3, 4))

119. What is the primary purpose of using factors in R?

a) To handle numerical calculations

b) To create matrices

c) To manage categorical data

d) To plot data

Answer: c) To manage categorical data

120. How can you convert a factor to its underlying numeric codes in R?

a) as.numeric(factor)

b) as.integer(factor)

c) unclass(factor)

d) Both a and b are correct

Answer: d) Both a and b are correct

121. Which statement is true about comparing ordered factors in R?

a) Ordered factors can only be compared using == operator.

b) Ordered factors cannot be compared.

c) You can use >, <, >=, <= to compare ordered factors.

d) Comparison operators convert factors to characters before comparing.

Answer: c) You can use >, <, >=, <= to compare ordered factors.

122. In R, what will `vec <- c("high", "medium", "low"); factor(vec, ordered=TRUE)` produce?

a) An ordered factor with levels "high" > "medium" > "low"

b) A numeric vector

c) A character vector

d) An error, because the levels are not specified

Answer: a) An ordered factor with levels "high" > "medium" > "low"

123. How do you assign names to the elements of a vector `v` in R?

a) names(v) <- c("first", "second", "third")

b) v.names <- c("first", "second", "third")

c) assign(v, c("first", "second", "third"))

d) v <- setNames(v, c("first", "second", "third"))

Answer: a) names(v) <- c("first", "second", "third")

124. What is the result of subsetting a data frame `df` with `df[1:3, ]`?

a) The first three columns of `df`

b) The first three rows of `df`

c) An error

d) A summary of the first three elements of `df`

Answer: b) The first three rows of `df`

125. How can you create a factor in R with levels explicitly ordered as "low", "medium", and "high"?

a) factor(c("low", "medium", "high"), levels = c("low", "medium", "high"), ordered = TRUE)

b) c("low", "medium", "high", ordered = TRUE)

c) as.ordered(c("low", "medium", "high"))

d) ordered(c("low", "medium", "high"))

Answer: a) factor(c("low", "medium", "high"), levels = c("low", "medium", "high"), ordered = TRUE)