UNIT- 1

Linear Algebra:

1. **Question:** What is the definition of a vector?
   A) A point in space
   B) A quantity with both magnitude and direction
   C) A matrix with one row
   D) A scalar quantity
   Answer: B) A quantity with both magnitude and direction
2. **Question:** In linear algebra, what is the determinant of a 2x2 matrix?
   A) ad - bc
   B) a + b
   C) a/b
   D) ac - bd
   Answer: A) ad - bc
3. **Question:** What is the result of multiplying a matrix by its inverse?
   A) Identity matrix
   B) Zero matrix
   C) Diagonal matrix
   D) Transpose matrix
   Answer: A) Identity matrix
4. **Question:** A system of linear equations with no solution is called:
   A) Inconsistent
   B) Homogeneous
   C) Consistent
   D) Singular
   Answer: A) Inconsistent
5. **Question:** The dot product of two vectors is also known as:
   A) Cross product
   B) Scalar product
   C) Matrix product
   D) Outer product
   Answer: B) Scalar product
6. **Question:** What is the rank of a matrix?
   A) The number of columns
   B) The number of rows
   C) The maximum number of linearly independent rows or columns
   D) The determinant of the matrix
   Answer: C) The maximum number of linearly independent rows or columns
7. **Question:** Eigenvectors of a matrix are associated with:

A) The identity matrix
B) The zero matrix
C) Eigenvalues
D) Transpose matrix

Answer: C) Eigenvalues

8. **Question:** What does the transpose of a matrix involve?

A) Changing the sign of each element
B) Interchanging rows and columns
C) Taking the square root of each element
D) Dividing each element by a constant

Answer: B) Interchanging rows and columns

9. **Question:** A linear transformation that stretches or compresses space along coordinate axes is called:

A) Rotation
B) Shear
C) Reflection
D) Scaling

Answer: D) Scaling

10. **Question:** The determinant of a square matrix is equal to zero if and only if:

A) The matrix is singular
B) The matrix is symmetric
C) The matrix is invertible
D) The matrix is orthogonal

Answer: A) The matrix is singular

Multivariate Calculus:

11. **Question:** What is the partial derivative of a function with respect to a variable?

A) The total derivative
B) The derivative at a specific point
C) The derivative with respect to only one variable, keeping others constant
D) The integral of the function

Answer: C) The derivative with respect to only one variable, keeping others constant

12. **Question:** The gradient of a scalar field is a:

A) Scalar
B) Vector
C) Matrix
D) Tensor

Answer: B) Vector

13. **Question:** What is the divergence of a vector field?

A) The curl of the vector field
B) The Laplacian of the vector field

C) The dot product of the gradient and the vector field
D) The cross product of the gradient and the vector field

Answer: C) The dot product of the gradient and the vector field

1. **Question:** What is a matrix?

A) A mathematical operation
B) A rectangular array of numbers
C) A set of linear equations
D) A complex number

Answer: B) A rectangular array of numbers

2. **Question:** The transpose of a matrix involves:

A) Inverting the matrix
B) Flipping its elements across the main diagonal
C) Multiplying it by its inverse
D) Swapping rows and columns

Answer: D) Swapping rows and columns

3. **Question:** What is the determinant of a 2x2 matrix [[a, b], [c, d]]?

A) ad - bc
B) a + b + c + d
C) (a + b)(c + d)
D) 2(ad - bc)

Answer: A) ad - bc

4. **Question:** The eigenvalues of a matrix are the solutions to the equation:

A) det(A) = 0
B) A^2 = I
C) A^T = A
D) Ax = b

Answer: A) det(A) = 0

5. **Question:** What is the dot product of two vectors a and b in R^n?

A) a * b
B) ||a|| * ||b||
C) Σ(ai * bi) for i=1 to n
D) a x b

Answer: C) Σ(ai * bi) for i=1 to n

Multivariate Calculus:

6. **Question:** The partial derivative of a function with respect to a specific variable measures:

A) The total change in the function
B) The rate of change with respect to all variables
C) The rate of change with respect to a single variable
D) The average rate of change

Answer: C) The rate of change with respect to a single variable

7. **Question:** What is the gradient of a scalar function?

   A) A vector of partial derivatives
   B) The Hessian matrix
   C) The Jacobian matrix
   D) The Laplacian operator

Answer: A) A vector of partial derivatives

8. **Question:** The line integral represents:

   A) The area under a curve
   B) The sum of function values along a curve
   C) The integral of a scalar field over a curve
   D) The integral of a vector field over a curve

Solving Simultaneous Equations:

1. **Question:** What is the solution to the system of equations 2x - y = 3 and x + y = 5?

   A) x = 2, y = 3
   B) x = 1, y = 4
   C) x = 3, y = 2
   D) x = 4, y = 1
   Answer: B) x = 1, y = 4

2. **Question:** In a system of three equations with three unknowns, how many equations are needed to guarantee a unique solution?

   A) 1
   B) 2
   C) 3
   D) 4
   Answer: C) 3

Matrix Transformations:

3. **Question:** Which type of matrix transformation preserves the shape and size of objects but may involve rotation or reflection?

A) Translation
B) Scaling
C) Rotation
D) Shearing
Answer: C) Rotation

4. **Question:** In a linear transformation, if T(v) = 0 only when v = 0, what type of transformation is T?

A) Shearing
B) Reflection
C) Invertible
D) Scaling
Answer: C) Invertible

Determinants and Inverses:

5. **Question:** If the determinant of a matrix A is zero, what does it imply about the invertibility of A?

A) A is invertible
B) A is not invertible
C) No conclusion can be drawn
D) A is a square matrix
Answer: B) A is not invertible

6. **Question:** What is the determinant of the identity matrix?

A) 0
B) 1
C) -1
D) Depends on the size of the matrix
Answer: B) 1

Matrices Changing Basis:

7. **Question:** In the context of changing basis, what does the transition matrix represent?

A) Eigenvalues
B) Eigenvectors
C) Change of basis
D) Orthogonal matrix
Answer: C) Change of basis

8. **Question:** If P is the transition matrix from basis B to basis C, what is the inverse of P used for?

A) Changing basis from C to B
B) Transforming matrices in basis C
C) Finding eigenvalues
D) Determining orthogonality
Answer: A) Changing basis from C to B

Orthogonal Matrices:

9. **Question:** What property does an orthogonal matrix satisfy?
A) Determinant is zero
B) Inverse is its transpose
C) Trace is zero
D) Eigenvalues are complex
Answer: B) Inverse is its transpose

10. **Question:** If Q is an orthogonal matrix, what is Q^T?
A) Q
B) -Q
C) Q^(-1)
D) I (Identity matrix)
Answer: A) Q

Eigenvalues and Eigenvectors:

11. **Question:** In the equation $Av = \lambda v$, what does λ represent?
A) Determinant of A
B) Eigenvalue
C) Inverse of A
D) Trace of A
Answer: B) Eigenvalue

12. **Question:** If a matrix has n distinct eigenvalues, how many linearly independent eigenvectors does it have?
A) n
B) 1
C) Depends on the size of the matrix
D) Always 2
Answer: A) n

14. **Question:** In the context of multivariate Taylor series, what is the role of the Hessian matrix?
A) Determines the coefficients of the series
B) Represents the partial derivatives of the function
C) Indicates the convergence of the series
D) Provides information about curvature at the expansion point

Answer: A) Determines the coefficients of the series

13. Question: In multivariable calculus, the Jacobian matrix represents the derivatives of what kind of functions?

A) Scalar functions
B) Vector functions
C) Matrix functions
D) Tensor functions
Answer: B) Vector functions

13. **Question:** The Hessian matrix is associated with which mathematical concept?

1. A) Integration
2. B) Differentiation
3. C) Determinants
4. D) Series expansion
5. **Answer:** C) Determinants

UNIT-2

Pandas for Data Handling:

1. **Question:** Which Pandas function is used to read a CSV file into a DataFrame?

A) **read_excel**
B) **read_csv**
C) **load_data**
D) **import_data**
Answer: B) **read_csv**

2. **Question:** What Pandas method is used to drop missing values from a DataFrame?

A) **remove_na**
B) **drop_missing**
C) **dropna**
D) **clean_data**
Answer: C) **dropna**

NumPy for Numerical Operations:

3. **Question:** In NumPy, what function is used to create an identity matrix?

A) **eye**
B) **ones**
C) **identity**

D) **zeros**

Answer: A) **eye**

4. **Question:** How can you concatenate two NumPy arrays vertically?

A) **vstack**

B) **concatenate_vertical**

C) **merge_vertical**

D) **stack_vertical**

Answer: A) **vstack**

Matplotlib for Visualization:

5. **Question:** What Matplotlib function is used to create a scatter plot?

A) **plot**

B) **scatter**

C) **draw_scatter**

D) **create_plot**

Answer: B) **scatter**

6. **Question:** Which Matplotlib command is used to add a title to a plot?

A) **title**

B) **add_title**

C) **set_title**

D) **plot_title**

Answer: C) **set_title**

Seaborn for Statistical Visualization:

7. **Question:** In Seaborn, what function is used to create a box plot?

A) **plot_box**

B) **create_boxplot**

C) **boxplot**

D) **draw_box**

Answer: C) **boxplot**

8. **Question:** How can you create a violin plot in Seaborn?

A) **plot_violin**

B) **draw_violin**

C) **violinplot**

D) **create_violin**

Answer: C) **violinplot**

Plotly for Interactive Visualization:

9. **Question:** Which Plotly module allows you to create interactive plots in Python?

A) **plotly.graph_objects**
B) **plotly.figure_factory**
C) **plotly.subplots**
D) **plotly.express**
Answer: A) **plotly.graph_objects**

10. **Question:** How can you add interactivity to a Plotly scatter plot?
A) **add_interactivity**
B) **scatter.interactive()**
C) **plotly.interactive**
D) **plotly.offline.plot**
Answer: D) **plotly.offline.plot**

Scikit-learn for Machine Learning:

11. **Question:** What is the primary purpose of Scikit-learn?
A) Web development
B) Machine learning
C) Data visualization
D) Natural language processing
Answer: B) Machine learning

12. **Question:** Which Scikit-learn module contains the **train_test_split** function?
A) **sklearn.preprocess**
B) **sklearn.model_selection**
C) **sklearn.split**
D) **sklearn.training**
Answer: B) **sklearn.model_selection**

TensorFlow for Deep Learning:

13. **Question:** What is TensorFlow primarily used for?
A) Data visualization
B) Machine learning
C) Web development
D) Deep learning
Answer: D) Deep learning

14. **Question:** How can you install TensorFlow in Python?
A) **pip install tf**
B) **pip install tensorflow**
C) **conda install tensorflow**
D) **install tf**
Answer: B) **pip install tensorflow**

SQLite for Database Handling:

15. **Question:** In Python, which library is commonly used for SQLite database operations?
    A) **sqlalchemy**
    B) **sqlite3**
    C) **pandas_sql**
    D) **sqlitealchemy**
    Answer: B) **sqlite3**
16. **Question:** What method is used to execute an SQL query in SQLite using the **sqlite3** library?
    A) **execute_query**
    B) **query**
    C) **execute**
    D) **run_query**
    Answer: C) **execute**

Altair for Declarative Visualization:

17. **Question:** In Altair, what is used to represent data on the x-axis or y-axis?
    A) Channels
    B) Marks
    C) Encodings
    D) Layers
    Answer: C) Encodings
18. **Question:** How can you create a scatter plot in Altair?
    A) **alt.Chart.scatter()**
    B) **alt.scatter()**
    C) **alt.create_scatter()**
    D) **alt.plot.scatter()**
    Answer: B) **alt.scatter()**

NetworkX for Network Analysis:

19. **Question:** What does NetworkX provide support for in Python?
    A) Statistical analysis
    B) Network analysis
    C) Image processing
    D) Natural language processing
    Answer: B) Network analysis
20. **Question:** How can you add nodes to a graph in NetworkX?
    A) **add_nodes()**
    B) **graph.add_nodes()**
    C) **add_nodes_from()**

D) **graph.add_nodes_from()**
Answer: C) **add_nodes_from()**

Statsmodels for Statistical Modeling:

21. **Question:** In Statsmodels, which module is commonly used for linear regression?
A) **statsmodels.regression**
B) **statsmodels.linear**
C) **statsmodels.model**
D) **statsmodels.api**
Answer: D) **statsmodels.api**
22. **Question:** How can you fit a linear model using Statsmodels?
A) **fit_model()**
B) **linear_model.fit()**
C) **sm.fit()**
D) **sm.OLS().fit()**
Answer: D) **sm.OLS().fit()**

PyTorch for Deep Learning:

23. **Question:** What is PyTorch known for in the context of deep learning?
A) High-level abstraction
B) Dynamic computational graph
C) Limited community support
D) Strict static typing
Answer: B) Dynamic computational graph
24. **Question:** How can you install PyTorch in Python?
A) **pip install pytorch**
B) **pip install torch**
C) **conda install pytorch**
D) **install torch**
Answer: B) **pip install torch**

Bokeh for Interactive Visualization:

25. **Question:** What does Bokeh specialize in?
A) Static visualizations
B) Interactive visualizations
C) Statistical modeling
D) Network analysis
Answer: B) Interactive visualizations

1) **What is the Hessian matrix used for in optimization?**

A) Gradient descent B) Second-order optimization C) Feature scaling D) Principal component analysis

2) **Answer: B) Second-order optimization**

3) **In multivariate calculus, what does the chain rule allow us to calculate?**

A) Derivatives of composite functions B) Integrals of composite functions C) Partial derivatives of multivariate functions D) Taylor series expansion

Answer: A) Derivatives of composite functions

4) **What does the power series expansion aim to represent a function as?**

A) A linear equation B) A quadratic equation C) An infinite sum of terms D) A constant value

Answer: C) An infinite sum of terms

5) **In linearization, what is the primary purpose of the tangent line to a curve at a specific point?**

A) Minimizing the function B) Maximizing the function C) Approximating the function near that point D) Calculating the integral of the function

Answer: C) Approximating the function near that point

6) **Which term represents the second partial derivatives in the Hessian matrix?**

A) Jacobian B) Gradient C) Hessian D) Laplacian

Answer: C) Hessian

7) **What is the multivariate chain rule used for in calculus?**

A) Computing partial derivatives of composite functions B) Computing definite integrals C) Finding global extrema D) Solving differential equations

Answer: A) Computing partial derivatives of composite functions

8) **When building an approximate function using linearization, what is the process of finding the best-fitting line to the curve at a specific point called?**

A) Extrapolation B) Interpolation C) Regression D) Interpolation

Answer: B) Interpolation

9) **In the context of power series, what is the purpose of including more terms in the expansion?**

A) Reducing accuracy B) Improving accuracy C) Making the series divergent D) Converging to zero

Answer: B) Improving accuracy

10) **What does the multivariate Taylor series expansion provide for a function?**

A) First-order approximation B) Second-order approximation C) Infinite-order approximation D) Exact solution

Answer: C) Infinite-order approximation

11) **Which matrix is involved in the multivariate chain rule when dealing with vector-valued functions?**

A) Jacobian matrix B) Hessian matrix C) Covariance matrix D) Identity matrix

Answer: A) Jacobian matrix

12) **What is the primary purpose of the Hessian matrix in optimization algorithms?**

A) Finding the minimum of a function B) Evaluating the gradient C) Estimating the variance D) Computing the Jacobian

Answer: A) Finding the minimum of a function

13) **In linearization, what is the term used to describe the difference between the true function value and the value predicted by the linear approximation?**

A) Error term B) Residual C) Deviation D) Anomaly

Answer: A) Error term

14) **When using the power series expansion, what happens if the series converges at a specific value of ?** ***x*?**

A) The function is undefined B) The series diverges C) The function is continuous at that point D) The expansion is inaccurate

Answer: C) The function is continuous at that point

15) **In linearization, what is the slope of the tangent line to a curve at a specific point equal to?**

A) The first derivative of the function B) The second derivative of the function C) The value of the function at that point D) The average rate of change over an interval

Answer: A) The first derivative of the function

16) **What is the purpose of the multivariate Taylor series in approximating a function?**

A) Estimating the global minimum B) Evaluating the Hessian matrix C) Providing a polynomial approximation D) Solving differential equations

Answer: C) Providing a polynomial approximation

17) **When building an approximate function using linearization, what does the term "local linear approximation" refer to?**

A) An approximation that holds globally B) An approximation that holds near a specific point C) An exact representation of the function D) An approximation with a linear trend

Answer: B) An approximation that holds near a specific point

18) **Which of the following is a benefit of using the power series expansion to approximate functions?**

A) Simplicity in computation B) Accuracy over the entire domain C) Guaranteed convergence D) Independence from initial conditions

Answer: B) Accuracy over the entire domain

19) **What does the term "linearization" mean in the context of mathematical modeling?**

A) Making a function more complex B) Approximating a function with a linear one C) Transforming a function into a linear equation D) Calculating the slope of a tangent line

Answer: B) Approximating a function with a linear one

20) **In the Hessian matrix, what does each entry represent?**

A) Second partial derivative of the function B) Gradient of the function C) First partial derivative of the function D) Jacobian matrix

Answer: A) Second partial derivative of the function

21) **Which mathematical operation is central to the multivariate chain rule?**

A) Addition B) Subtraction C) Multiplication D) Division

Answer: C) Multiplication

22) **When using power series, what is the term for the mathematical function represented by the sum of the series?**

A) Polynomial B) Exponential C) Taylor series D) Generating function

Answer: A) Polynomial

23) **In linearization, what is the purpose of the residual or error term?**

A) It quantifies the accuracy of the linear approximation B) It adds complexity to the linear model C) It represents the slope of the tangent line D) It is the derivative of the function

Answer: A) It quantifies the accuracy of the linear approximation

24) **Which term is associated with the rate at which a power series converges to the function it approximates?**

A) Remainder term B) Residual term C) Error term D) Deviation term

Answer: A) Remainder term

25) **What is the primary advantage of using the multivariate Taylor series over a simple linear approximation?**

A) It provides a better fit over the entire domain B) It is computationally simpler C) It requires fewer terms in the series D) It is only applicable to univariate functions

Answer: A) It provides a better fit over the entire domain

26) **In power series expansions, what is the role of the remainder term?**

A) It represents the function itself B) It quantifies the error in the approximation C) It is the first term in the series D) It determines the convergence of the series

Answer: B) It quantifies the error in the approximation

27) **When using the Hessian matrix, what does a positive definite matrix imply about the function?**

A) The function has no critical points B) The function has a local minimum C) The function is linear D) The function has a global minimum

Answer: B) The function has a local minimum

28) **In linearization, what is the term used to describe the interval over which the linear approximation is valid?**

A) Domain B) Range C) Neighborhood D) Derivative

Answer: C) Neighborhood

29) **Which of the following statements is true about the multivariate chain rule?**

A) It only applies to univariate functions B) It is a special case of the power series expansion C) It allows for the computation of partial derivatives of composite functions D) It is equivalent to the Hessian matrix

Answer: C) It allows for the computation of partial derivatives of composite functions

30) **In power series expansions, what is the significance of the convergence radius?**

A) It determines the number of terms needed for convergence B) It indicates the interval of x-values for which the series converges C) It measures the accuracy of the approximation D) It represents the rate of convergence

Answer: B) It indicates the interval of x-values for which the series converges

Question 1: What is the primary purpose of converting data into numerical format?

A. Enhancing data security B. Improving data visualization C. Reducing data storage requirements D. All of the above

Answer: B. Improving data visualization

Question 2: Which of the following is not a common method for converting categorical data into numerical format?

A. One-Hot Encoding B. Label Encoding C. Mean Encoding D. String Conversion

Answer: D. String Conversion

Question 3: In machine learning, why is it essential to convert textual data into numerical format?

A. Text data cannot be processed by machine learning algorithms. B. Numerical data allows for easier mathematical operations and analysis. C. Textual data occupies too much storage space. D. Text data is inherently inaccurate.

Answer: B. Numerical data allows for easier mathematical operations and analysis.

Question 4: What is the purpose of normalization when converting numerical data?

A. To scale data within a specific range B. To convert data into integers C. To remove missing values D. To convert data into binary format

Answer: A. To scale data within a specific range

Question 5: Which of the following is an example of a continuous variable?

A. Gender (Male/Female) B. Age C. Country D. Marital Status (Single/Married)

Answer: B. Age

Question 6: When dealing with time-series data, what is a common approach to convert timestamps into numerical features?

A. One-Hot Encoding B. Label Encoding C. Time Encoding D. Fourier Transform

Answer: C. Time Encoding

Question 7: What is the purpose of feature scaling in data preprocessing?

A. To convert features into categorical variables B. To standardize the range of numerical features C. To increase the dimensionality of the dataset D. To remove outliers from the data

Answer: B. To standardize the range of numerical features

Question 8: Which technique is suitable for handling missing numerical data in a dataset?

A. Mean imputation B. Mode imputation C. Regression imputation D. One-Hot Encoding

Answer: A. Mean imputation

Question 9: What is the role of encoding in converting data for natural language processing (NLP) tasks?

A. To compress the data B. To encrypt the data C. To convert text into numerical representations D. To remove outliers from the text

Answer: C. To convert text into numerical representations

Question 10: What is the significance of using dummy variables in regression analysis?

A. To create new variables B. To handle multicollinearity in categorical data C. To replace missing values D. To perform feature selection

Answer: B. To handle multicollinearity in categorical data

1. **Which of the following are types of correlation?**
    a. Positive and Negative
    b. Simple, Partial and Multiple

c. Linear and Nonlinear
d. All of the above

Answer: d

2. **Which of the following is true for the coefficient of correlation?**
    a. The coefficient of correlation is not dependent on the change of scale
    b. The coefficient of correlation is not dependent on the change of origin
    c. The coefficient of correlation is not dependent on both the change of scale and change of origin
    d. None of the above

    Answer: c

3. **Which of the following statements is true for correlation analysis?**
    a. It is a bivariate analysis
    b. It is a multivariate analysis
    c. It is a univariate analysis
    d. Both a and c

    Answer: c

4. **If the values of two variables move in the same direction, _____**
    a. The correlation is said to be non-linear
    b. The correlation is said to be linear
    c. The correlation is said to be negative
    d. The correlation is said to be positive

    Answer: d

5. **If the values of two variables move in the opposite direction, _____**
    a. The correlation is said to be linear
    b. The correlation is said to be non-linear
    c. The correlation is said to be positive
    d. The correlation is said to be negative

    Answer: d

6. **Which of the following techniques is an analysis of the relationship between two variables to help provide the prediction mechanism?**
    a. Standard error
    b. Correlation

c. Regression
d. None of the above

Answer: c

7. **Which of the following statements is true about the arithmetic mean of two regression coefficients?**
    a. It is less than the correlation coefficient
    b. It is equal to the correlation coefficient
    c. It is greater than or equal to the correlation coefficient
    d. It is greater than the correlation coefficient

    Answer: d

8. **What is the meaning of the testing of the hypothesis?**
    a. It is a significant estimation of the problem
    b. It is a rule for acceptance or rejection of the hypothesis of the research problem
    c. It is a method of making a significant statement
    d. None of the above

    Answer: b

9. **Which of the following statements is true about the null hypothesis?**
    a. Any wrong decision related to the null hypothesis results in two types of errors
    b. Any wrong decision related to the null hypothesis results in one type of an error
    c. Any wrong decision related to the null hypothesis results in four types of errors
    d. Any wrong decision related to the null hypothesis results in three types of errors

    Answer: a

10. **Which of the following statements is true about the type two error?**
    a. Type two error means to accept an incorrect hypothesis
    b. Type two error means to reject an incorrect hypothesis
    c. Type two error means to accept a correct hypothesis
    d. Type two error means to reject a correct hypothesis

    Answer: a

11. **Which of the following statements is true about the level of significance?**
    a. In testing a hypothesis, we take the level of significance as 2% if it is not mentioned earlier
    b. In testing a hypothesis, we take the level of significance as 1% if it is not mentioned earlier
    c. In testing a hypothesis, we take the level of significance as 10% if it is not mentioned earlier
    d. In testing a hypothesis, we take the level of significance as 5% if it is not mentioned earlier

    Answer: a

12. **The independent variable is used to explain the dependent variable in _____.**
    a. Linear regression analysis
    b. Multiple regression analysis
    c. Non-linear regression analysis
    d. None of the above

    Answer: a

13. **Which of the following statements is true about the regression line?**
    a. A regression line is also known as the line of the average relationship
    b. A regression line is also known as the estimating equation
    c. A regression line is also known as the prediction equation
    d. All of the above

    Answer: d

14. **Which of the following statements is true about the correlational analysis between two sets of data?**
    a. The correlational analysis between two sets of data is known as a simple correlation
    b. The correlational analysis between two sets of data is known as multiple correlation
    c. The correlational analysis between two sets of data is known as partial correlation
    d. None of the above

    Answer: a

15. **The original hypothesis is known as _____.**

a. Alternate hypothesis
b. Null hypothesis
c. Both a and b are incorrect
d. Both a and b are correct

Answer: b

1. What is data visualization?

a) The process of collecting data

b) The process of representing data in graphical form

c) The process of cleaning data

d) The process of storing data in databases

Answer: b) The process of representing data in graphical form

2. What is the primary goal of data visualization?

a) To collect more data

b) To represent data accurately

c) To make data more accessible and understandable

d) To hide data from the audience

Answer: c) To make data more accessible and understandable

3. Which of the following is NOT a common visual element used in data visualization?

a) Bar chart

b) Line chart

c) Data table

d) Pie chart

Answer: c) Data table

4. Why is data visualization important in data analysis?

a) It helps hide complex data patterns

b) It makes data analysis slower and less efficient

c) It enables quick understanding and insights from data

d) It is not important in data analysis

Answer: c) It enables quick understanding and insights from data

5. Which type of data visualization is best suited for showing the distribution of a single numerical variable?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: a) Bar chart

6. What does interactivity in data visualization allow users to do?

a) Change the color scheme of the visualizations

b) Add more data to the visualizations

c) Explore and analyze the data further by interacting with the visualizations

d) Hide certain data points in the visualizations

Answer: c) Explore and analyze the data further by interacting with the visualizations

7. How does data visualization facilitate better communication of insights?

a) By making the data more confusing and difficult to interpret

b) By simplifying complex data and presenting it visually

c) By excluding data points that are not relevant to the audience

d) By presenting the data in long paragraphs of text

Answer: b) By simplifying complex data and presenting it visually

8. Which data visualization is commonly used to show the relationship between two numerical variables?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: c) Scatter plot

9. What is the benefit of using line charts in data visualization?

a) They show the distribution of a single numerical variable

b) They are useful for comparing multiple categories at once

c) They are effective for showing trends and changes over time

d) They display the parts of a whole in a circular format

Answer: c) They are effective for showing trends and changes over time

10. What is the purpose of providing data context in data visualization?

a) To make the data more confusing

b) To present the data in multiple formats

c) To allow users to interact with the data visualizations

d) To help users understand the significance and implications of the data

Answer: d) To help users understand the significance and implications of the data

11. Which type of data visualization is suitable for comparing the proportion of different categories in a whole?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: d) Pie chart

12. How does data visualization support decision-making?

a) By making the data more complex and difficult to understand

b) By providing irrelevant data points for analysis

c) By enabling users to identify trends and patterns in the data

d) By hiding data from the users

Answer: c) By enabling users to identify trends and patterns in the data

13. Which data visualization is commonly used to compare values across different categories?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: a) Bar chart

14. What is the benefit of using color in data visualization?

a) To make the visualizations more monotonous

b) To hide data points

c) To highlight important data points and patterns

d) To decrease the visibility of the visualizations

Answer: c) To highlight important data points and patterns

15. Which data visualization is often used to show the correlation between two numerical variables?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: c) Scatter plot

Part 2: Download data visualization questions & answers for free

16. How does interactive data visualization benefit data analysis?

a) It allows users to hide data points they don’ t like

b) It slows down data analysis

c) It enables users to explore and analyze data in real-time

d) It decreases the visibility of data visualizations

Answer: c) It enables users to explore and analyze data in real-time

17. What is the primary purpose of storytelling in data visualization?

a) To make the data more complicated and confusing

b) To hide the data from the audience

c) To provide context and a narrative around the data

d) To reduce the amount of data presented

Answer: c) To provide context and a narrative around the data

18. What does a bubble chart represent in data visualization?

a) The distribution of a single numerical variable

b) The correlation between two numerical variables

c) The comparison of values across different categories

d) The relationship between three numerical variables

Answer: d) The relationship between three numerical variables

19. How can data visualization support data-driven decision-making?

a) By presenting data in a confusing and disorganized manner

b) By hiding relevant data points from users

c) By enabling quick understanding and insights from the data

d) By excluding certain data points in the visualizations

Answer: c) By enabling quick understanding and insights from the data

20. Which type of data visualization is suitable for showing the change in values over time?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: b) Line chart

21. Why is data visualization essential in presenting complex data sets?

a) To

make the data more difficult to understand

b) To make the data appear larger than it actually is

c) To simplify the data and make it easier to grasp

d) To hide data patterns from the audience

Answer: c) To simplify the data and make it easier to grasp

22. Which data visualization is commonly used to compare values between different

categories over time?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: b) Line chart

23. How does data visualization facilitate better communication of data insights?

a) By using complex jargon and technical terms

b) By presenting data in tables and long paragraphs of text

c) By transforming data into visual representations that are easy to understand

d) By excluding data points from the visualizations

Answer: c) By transforming data into visual representations that are easy to

understand

24. Which data visualization is commonly used to compare values across multiple categories and show trends over time?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: a) Bar chart

25. What is the significance of providing context in data visualization?

a) It makes the data more confusing

b) It allows users to interact with the visualizations

c) It helps users understand the significance and implications of the data

d) It hides data patterns from the audience

Answer: c) It helps users understand the significance and implications of the data

26. Which data visualization is suitable for displaying the distribution of a single numerical variable with outliers?

a) Bar chart

b) Line chart

c) Scatter plot

d) Box plot

Answer: d) Box plot

27. How can data visualization support effective communication of data insights?

a) By using small fonts and unclear labels

b) By presenting data in a disorganized manner

c) By using relevant visual elements to convey the message clearly

d) By excluding important data points from the visualizations

Answer: c) By using relevant visual elements to convey the message clearly

28. Which data visualization is best suited for showing the correlation between two numerical variables with a line indicating the trend?

a) Bar chart

b) Line chart

c) Scatter plot

d) Pie chart

Answer: c) Scatter plot

29. What is the purpose of using color gradients in data visualization?

a) To highlight specific data points

b) To hide data patterns

c) To make the data more difficult to understand

d) To make the visualizations less appealing

Answer: a) To highlight specific data points

30. How does data visualization benefit decision-makers in understanding complex data?

a) By providing raw data without any visual representations

b) By making the data more confusing and difficult to analyze

c) By transforming complex data into visual representations for quick understanding and insights

d) By excluding relevant data points from the visualizations

Answer: c) By transforming complex data into visual representations for quick understanding and insights

1. **Question:** In data analysis, what is a common method for handling missing numerical values?

   A) Deleting rows with missing values
   B) Replacing missing values with the mean
   C) Ignoring missing values
   D) Filling missing values with random numbers

   Answer: B) Replacing missing values with the mean

2. **Question:** What is the purpose of imputation in data analysis?

   A) Deleting missing values
   B) Replacing missing values with new data
   C) Filling in missing values with estimated or predicted values
   D) Ignoring missing values

   Answer: C) Filling in missing values with estimated or predicted values

3. **Question:** Which mathematical operation is commonly used to measure the central tendency of a dataset?

   A) Median
   B) Mode
   C) Range
   D) Standard Deviation

   Answer: A) Median

4. **Question:** What is the purpose of using interpolation for handling missing values?

   A) Estimating missing values based on existing data points
   B) Deleting missing values
   C) Replacing missing values with the mode
   D) Ignoring missing values

   Answer: A) Estimating missing values based on existing data points

5. **Question:** When is it appropriate to use forward fill to handle missing values in a time series dataset?

   A) When values are missing at the beginning of the time series
   B) When values are missing randomly
   C) When values are missing at the end of the time series
   D) When values are missing in the middle of the time series

Answer: A) When values are missing at the beginning of the time series

6. **Question:** What does the term "outlier" refer to in data analysis?
   A) Missing data points
   B) Unusually high or low values in a dataset
   C) Average values in a dataset
   D) Standardized values

Answer: B) Unusually high or low values in a dataset

7. **Question:** Which statistical measure is sensitive to outliers?
   A) Median
   B) Mean
   C) Mode
   D) Range

Answer: B) Mean

8. **Question:** What is the purpose of using the median instead of the mean for imputing missing values?
   A) Median is less affected by outliers
   B) Median is always equal to the mean
   C) Median is easier to calculate
   D) Median provides a larger value

Answer: A) Median is less affected by outliers

9. **Question:** Which method involves replacing missing values based on the values of its neighbors?
   A) Imputation
   B) Interpolation
   C) Extrapolation
   D) Regression

Answer: B) Interpolation

10. **Question:** What does the acronym "MICE" stand for in the context of imputing missing values?
   A) Multiple Imputation by Chained Equations
   B) Maximum Imputation for Complete Estimation
   C) Mean Imputation with Clustered Elements
   D) Missing Information Correction and Estimation

Answer: A) Multiple Imputation by Chained Equations

11. **Question:** Which method involves replacing missing values with the most frequently occurring value in a dataset?
   A) Mean imputation
   B) Mode imputation
   C) Median imputation
   D) Regression imputation

Answer: B) Mode imputation

12. **Question:** What is the primary purpose of exploratory data analysis (EDA)?

A) Cleaning the dataset
B) Summarizing the main characteristics of the dataset
C) Imputing missing values
D) Conducting hypothesis testing

13. If a dataset has missing values, which of the following is a common method to handle them?

A) Remove the entire row B) Replace with the mean/median C) Replace with zero D) Leave them as is

Answer: B) Replace with the mean/median

14. Which mathematical operation is used to find the average of a set of numbers?

A) Addition B) Subtraction C) Multiplication D) Division

Answer: D) Division

15. If a dataset has outliers, which measure of central tendency is less affected?

A) Mean B) Median C) Mode D) Range

Answer: B) Median

16. How is the range of a set of numbers calculated?

A) Subtract the smallest from the largest B) Add all numbers and divide by the count C) Multiply the smallest by the largest D) Take the square root of the product

Answer: A) Subtract the smallest from the largest

17. In data analysis, which method is commonly used to impute missing values based on the relationships with other variables?

A) Random imputation B) Mean imputation C) Regression imputation D) Mode imputation

Answer: C) Regression imputation

18. If a dataset has a skewness value of 0, what does it indicate about the distribution?

A) Positively skewed B) Normally distributed C) Negatively skewed D) No conclusion can be drawn

Answer: B) Normally distributed

19. Which measure of dispersion is influenced by extreme values?

A) Range B) Variance C) Standard deviation D) Interquartile range

Answer: B) Variance

20. If a dataset has missing values, which of the following is NOT a valid method to handle them?

A) Removing the entire column B) Interpolation C) Imputation using the mode D) Ignoring them in analysis

Answer: D) Ignoring them in analysis

21. In data analysis, which measure of central tendency is most affected by outliers?

A) Mean B) Median C) Mode D) Range

Answer: A) Mean

22. If a dataset has missing values, which statistical technique can be used to estimate and replace those missing values based on the observed values?

A) Clustering B) Principal Component Analysis (PCA) C) Multiple Imputation D) Cross-validation

Answer: C) Multiple Imputation

23. What is the absolute value of -9?

A) 9 B) -9 C) 0 D) 81

Answer: A) 9

24. In data analysis, which method can be used to identify and handle outliers?

A) Box plot B) Histogram C) Bar chart D) Pie chart

Answer: A) Box plot

25. If a dataset has missing values, what is a drawback of using mean imputation?

A) It is computationally expensive B) It may introduce bias C) It is only suitable for small datasets D) It cannot handle continuous variables

Answer: B) It may introduce bias

26. In data analysis, what does the term "skewness" measure?

A) Spread of the data B) Symmetry of the distribution C) Presence of outliers D) Central tendency

Answer: B) Symmetry of the distribution

27. If a dataset has missing values, which method involves using the values of neighboring data points to estimate and replace the missing values?

A) Mean imputation B) Interpolation C) Mode imputation D) Regression imputation

Answer: B) Interpolation

28. In data analysis, which method is used to detect and handle multicollinearity among variables?

A) Correlation analysis B) Factor analysis C) Variance inflation factor (VIF) D) Chi-square test

Answer: C) Variance inflation factor (VIF)

29. If a dataset has missing values, which imputation method is suitable for categorical variables?

A) Mean imputation B) Regression imputation C) Mode imputation D) Interpolation

Answer: C) Mode imputation