Data Science Class Test

1. What is the main objective of descriptive statistics?
   A. To summarize data meaningfully
   B. To test hypotheses
   C. To predict future values
   D. To estimate population parameters
   Answer: A
2. Which measure of central tendency is least affected by extreme values?
   A. Mean
   B. Mode
   C. Range
   D. Median
   Answer: D
3. The mean of a dataset is calculated as:
   A. The middle value
   B. The most frequent value
   C. Sum of all values divided by total values
   D. Difference between maximum and minimum
   Answer: C
4. Mode is most useful when the data is:
   A. Numerical and continuous
   B. Highly skewed
   C. Categorical
   D. Normally distributed
   Answer: C
5. If a dataset contains extreme outliers, which measure is most reliable?
   A. Mean
   B. Median
   C. Variance
   D. Range
   Answer: B
6. What does the range represent in statistics?
   A. Average distance from the mean
   B. Spread of middle values
   C. Difference between highest and lowest values
   D. Relationship between variables
   Answer: C
7. Variance measures:
   A. Direction of relationship
   B. Average squared deviation from the mean
   C. Absolute deviation from median
   D. Spread between quartiles
   Answer: B

8. Standard deviation is preferred over variance because it:
   A. Is easier to calculate
   B. Uses original data values
   C. Has squared units
   D. Is in the same unit as data
   Answer: D
9. Which statistic measures how two variables vary together?
   A. Correlation
   B. Variance
   C. Mean
   D. Covariance
   Answer: D
10. A negative covariance indicates that variables:
   A. Are unrelated
   B. Move in opposite directions
   C. Move together
   D. Are constant
   Answer: B
11. Correlation values always lie between:
   A. 0 and 1
   B. –10 and +10
   C. –1 and +1
   D. –100 and +100
   Answer: C
12. A correlation value close to +1 indicates:
   A. Weak relationship
   B. No relationship
   C. Strong positive relationship
   D. Strong negative relationship
   Answer: C
13. Inferential statistics is mainly used to:
   A. Clean data
   B. Visualize patterns
   C. Describe samples
   D. Draw conclusions about a population
   Answer: D
14. Which of the following is NOT an inferential technique?
   A. Hypothesis testing
   B. Confidence intervals
   C. Mean calculation
   D. Sampling
   Answer: C
15. Simple random sampling ensures:
   A. Only experts are chosen
   B. Equal chance for every element
   C. Groups are selected
   D. Samples are ordered
   Answer: B

16. Stratified sampling divides population based on:
    A. Time
    B. Random order
    C. Geographic location
    D. Subgroups or strata
    Answer: D
17. Data wrangling mainly involves:
    A. Model deployment
    B. Feature selection
    C. Data preparation
    D. Data visualization
    Answer: C
18. Missing data can be handled by all EXCEPT:
    A. Imputation
    B. Deletion
    C. Ignoring analysis completely
    D. Replacement with mean/median
    Answer: C
19. Outliers can significantly affect:
    A. Median
    B. Mode
    C. Mean
    D. Frequency
    Answer: C
20. Duplicate records should be removed because they:
    A. Increase variance
    B. Bias results
    C. Improve accuracy
    D. Normalize data
    Answer: B
21. Standardization transforms data to have:
    A. Range 0 to 1
    B. Mean = 0 and Std = 1
    C. Only positive values
    D. No outliers
    Answer: B
22. Normalization is especially useful when:
    A. Data has text values
    B. Data has missing values
    C. Features have different scales
    D. Dataset is small
    Answer: C
23. Feature engineering aims to:
    A. Reduce dataset size
    B. Remove noise only
    C. Create informative features
    D. Visualize data
    Answer: C

24. Encoding categorical variables is required because:
    A. Algorithms work only with numbers
    B. Categories are unordered
    C. Text increases accuracy
    D. It removes duplicates
    Answer: A
25. Label encoding assigns:
    A. Binary vectors
    B. Frequencies
    C. Numerical labels to categories
    D. Probabilities
    Answer: C
26. Feature transformation helps mainly in:
    A. Improving model performance
    B. Reducing labels
    C. Removing rows
    D. Visualizing trends
    Answer: A
27. Merging datasets is done to:
    A. Remove columns
    B. Combine information
    C. Normalize values
    D. Reduce features
    Answer: B
28. Splitting datasets is required to:
    A. Increase training data
    B. Remove bias
    C. Validate model performance
    D. Reduce noise
    Answer: C
29. Time-series plots are used to show:
    A. Category distribution
    B. Spatial patterns
    C. Changes over time
    D. Correlation strength
    Answer: C
30. Which Python library is commonly used for interactive maps?
    A. Seaborn
    B. Matplotlib
    C. Plotly
    D. Folium
    Answer: D
31. Folium visualizations are based on:
    A. Static images
    B. Excel charts
    C. Web-based maps
    D. Bar plots
    Answer: C

32. Plotly dashboards are mainly known for being:
    A. Static
    B. Text-based
    C. Interactive
    D. Offline only
    Answer: C
33. Dashboards help decision-making by:
    A. Showing raw data
    B. Highlighting key metrics
    C. Cleaning data
    D. Encoding features
    Answer: B
34. Pivot tables are used to:
    A. Clean missing data
    B. Summarize large datasets
    C. Normalize values
    D. Encode categories
    Answer: B
35. Conditional formatting is used to:
    A. Apply formulas
    B. Sort data
    C. Highlight patterns
    D. Merge cells
    Answer: C
36. Sparklines are:
    A. Large dashboards
    B. Miniature charts in cells
    C. Pivot tools
    D. Filters
    Answer: B
37. Data storytelling focuses on:
    A. Statistical depth
    B. Model tuning
    C. Clear communication of insights
    D. Code optimization
    Answer: C
38. Effective data storytelling requires:
    A. Complex formulas
    B. Clear visuals and narrative
    C. More tables
    D. Raw data only
    Answer: B
39. Structuring reports improves:
    A. Dataset size
    B. Coding speed
    C. Understanding of results
    D. Feature count
    Answer: C

40. A good data presentation should primarily:
    A. Use maximum text
    B. Avoid charts
    C. Focus on insights
    D. Show raw numbers
    Answer: C