CSSE 386 – Data Mining with Programming
Rose-Hulman Institute of Technology

Quiz 02

Name (Print):_____   Section:_____

1. What is a univariate analysis?

    a.) Analysis of relationships between two variables

    b.) Analysis of a single variable

    c.) Analysis of missing data

    d.) Analysis involving more than two variables

2. Which of the following is not part of the 5-number summary used in box plots?

    a.) Minimum

    b.) Mean

    c.) Median

    d.) Q3

3. What does a boxplot represent visually? Select all that apply

    a.) Spread of data (interquartile range)

    b.) Median of the data

    c.) Outliers in the data

    d.) Relationships between two variables

4. Which term describes missing data that occurs completely at random, with no relationship to any observed or unobserved variable?

    a.) MAR

    b.) MNAR

    c.) MCAR

    d.) Data Imputation

5. Which of the following methods is best for visualizing the relationship between two continuous variables?

    a.) Boxplot

    b.) Histogram

    c.) Scatterplot

    d.) Bar chart

6. What are the common methods to handle missing data? Select all that apply

   a.) Deletion

   b.) Regression

   c.) Clustering

   d.) Imputation

7. What does the term MAR (Missing at Random) mean?

   a.) Missingness depends only on observed data

   b.) Missingness occurs completely randomly

   c.) Missingness cannot be explained by the data

8. Which of the following is a best practice for data visualization? Select all that apply

   a.) Always use 3D charts for better visualization

   b.) Avoid overloading your plot

   c.) Use clear and descriptive axis labels

   d.) Avoid using legends for categorical data

9. Which of the following methods are common for imputing missing data? Select all that apply

   a.) Mean or median substitution

   b.) Removing all columns with missing data

   c.) Predictive modeling (e.g., regression imputation)

   d.) Replacing missing data by NaN

10. What are some common reasons for missing data in datasets? Select all that apply

    a.) Data entry errors

    b.) Lack of data visualization tools

    c.) Non-response in surveys

11. Identify the scenario: A survey dataset is missing responses because some participants accidentally skipped questions due to a survey software glitch

    a.) MNAR

    b.) MCAR

    c.) MAR

12. A medical dataset is missing weight measurements, and the missingness is more common among older participants

    a.) MNAR

    b.) MCAR

    c.) MAR