STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False
   Bernoulli random variables are binary random variables that can only take two possible values: 1 (success) or 0 (failure).
2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?
   a) Central Limit Theorem
   b) Central Mean Theorem
   c) Centroid Limit Theorem
   d) All of the mentioned
   The Central Limit Theorem states that the distribution of averages (or sums) of independent and identically distributed (iid) variables, properly normalized, tends to follow a standard normal distribution as the sample size increases.
3. Which of the following is incorrect with respect to use of Poisson distribution?
   a) Modeling event/time data
   b) Modeling bounded count data
   c) Modeling contingency tables
   d) All of the mentioned
   The Poisson distribution is commonly used for modeling event/time data, where the focus is on the number of events occurring within a specific time period. It is particularly suitable for situations where events occur at a constant rate and independently of each other.

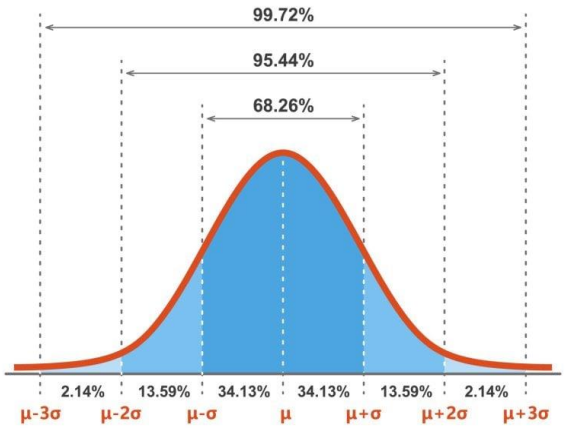
4. Point out the correct statement.
   a) The exponent of a normally distributed random variables follows what is called the log- normal distribution
   b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   c) The square of a standard normal random variable follows what is called chi-squared distribution
   d) All of the mentioned
5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False
7. 1. Which of the following testing is concerned with making decisions using data?
   a) Probability
   b) Hypothesis
   c) Causal
   d) None of the mentioned
8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
   a) 0
   b) 5
   c) 1
   d) 10

9. Which of the following statement is incorrect with respect to outliers?
    a) Outliers can have varying degrees of influence
    b) Outliers can be the result of spurious or real processes
    c) Outliers cannot conform to the regression relationship
    d) None of the mentioned

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Ans.10 The term "Normal Distribution" refers to a probability distribution that is symmetrically shaped and characterized by a bell-shaped curve. It is also known as the Gaussian distribution. The distribution is defined by two parameters: the mean, which represents the central tendency, and the standard deviation, which indicates the spread of the data. In a normal distribution, approximately 68% of the data falls within one standard deviation of the mean, about 95% falls within two standard deviations, and around 99.7% falls within three standard deviations.



11. How do you handle missing data? What imputation techniques do you recommend?
Handling missing data is an important step in data analysis. There are various techniques for handling missing data, and the choice of technique depends on the nature of the data and the specific analysis goals.
Mean/median imputation: Replace missing values with the mean or median of the available data for that variable.
Regression imputation: Predict missing values using regression models based on other variables that are correlated with the variable containing missing data.

12. What is A/B testing?
A/B testing, also known as split testing, is a statistical experiment used to compare two or more variants of a particular feature or design element to determine which one performs better. It is commonly used in various fields, including marketing, and website optimization.
The primary goal of A/B testing is to determine whether the changes made in the experimental group result in statistically significant improvements compared to the control group.

13. Is mean imputation of missing data acceptable practice?
Yes, Mean imputation involves replacing missing values with the mean of the available data for that variable. While it is an easy way to handle missing data.
Distortion of variability:- By replacing missing values with the mean, the imputed values are artificially assigned values that may not reflect the true variability of the variable.
Bias in relationships:- Mean imputation assumes that the missing data are missing completely at random , meaning the missingness is unrelated to the values of the variable or any other variables.

14. What is linear regression in statistics?

Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables, meaning that the

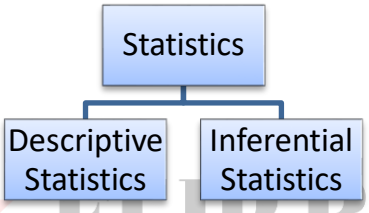
relationship can be represented by a straight line.
In linear regression, the goal is to find the best-fitting line that minimizes the difference between the predicted values and the actual observed values of the dependent variable.

$$Y = \beta 0 + \beta 1X + \varepsilon$$

where:

Y is the dependent variable or the variable to be predicted.
X is the independent variable or predictor variable.
β0 is the intercept, which represents the value of Y when X is 0.
β1 is the slope or regression coefficient, which represents the change in Y for a one-unit change in X.
ε is the error term or residual, representing the unexplained variation in Y.

15. What are the various branches of statistics?
Statistics is a field of study that involves the collection, analysis, interpretation, presentation, and organization of data.



Descriptive Statistics: Descriptive statistics involves summarizing and describing data using measures such as mean, median, mode, standard deviation, and graphs like histograms, bar charts, and scatter plots.

Inferential Statistics: Inferential statistics is concerned with drawing conclusions and making inferences about populations based on sample data. It involves estimation, hypothesis testing, and determining the level of confidence or significance in the results.