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
   ANS:- **True**
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
   ANS:-**CENTRAL LIMIT THEORM**

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

ANS**:- (a)**

3. 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
   ANS:- **(c)**

5. **Poisson**_____ random variables are used to model rates.

a) Empirical b) Binomial c) Poisson d) All of the mentioned

6. Usually replacing the standard error by its estimated value does change the CLT.

a) True b) False

ANS:- **False**

7. Which of the following testing is concerned with making decisions using data?

a) Probability b) Hypothesis c) Causal d) None of the mentioned

ANS:- **Hypothesis**

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0 b) 5 c) 1 d) 10

ANS:- **1**

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

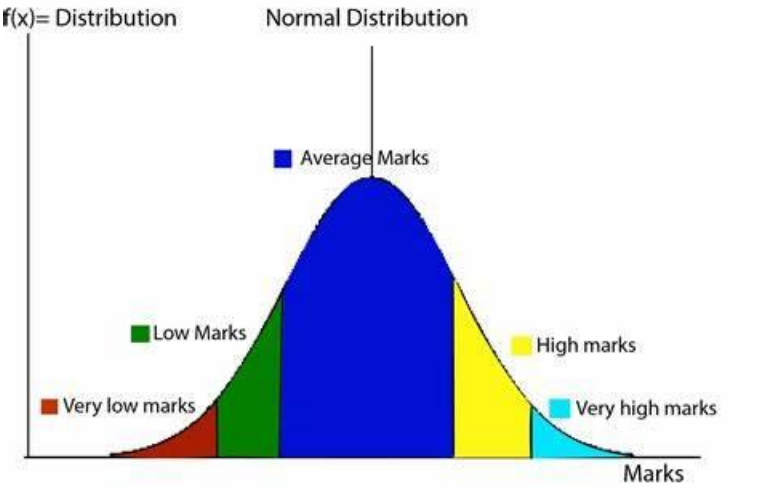
ANS:- **(b)**

WORKSHEET

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:- **Normal distribution is a continuous type of probability distribution where data is symmetrically distributed with no skew. When the given data is normally distributed in a form of a bell curve then it is referred to as normal distribution. The properties include: The mean , median, mode are same. The mean is where the curve is at its peak.**



11. How do you handle missing data? What imputation techniques do you recommend?

ANS:- **To handle missing data at first we need to analyse the data and if the missing values is affecting the whole data or not. If not, then deleting the missing data can be suggested . If the missing value does affect the whole data i.e. the original data while making a graphical analysis then the missing values can be filled with the mean of the variables.**

12. What is A/B testing?

ANS:-

13. Is mean imputation of missing data acceptable practice?

ANS:- **Advantages of imputing Mean data:**

-> It won't decrease the sample size.

-> This method is easy to understand and apply.

-> Sample mean is same for all the variables.

Disadvantages of imputing Mean data:

->Values imputed have correlation of zero with other variables, therefore relationship between variables does not become zero.

-> The confidence interval is narrow .

Therefore, it completely depends on the size of data, if there are more than 25% of missing values then it must be deleted otherwise imputation can be applied.

14. What is linear regression in statistics?

ANS:- **Linear regression is a model that is used to predict the relationship between two variables , one is target variable and the other is the feature variable. It is used to predict the effect of independent variable on dependent variable. For example what will be the rate of gold in 3 months using past figures.**

15. What are the various branches of statistics?

ANS:- **Descriptive Statistics: It deals with the presentation of data collection of data and Inferential Statistics: It deals with making future predictions and making right conclusions based on data derived from descriptive statistics.**