# 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 Theorem

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. Modeling bounded count data

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

   Ans. D) All of the mentioned

5. _____random variables are used to model rates.
   a) Empirical
   b) Binomial
   c) Poisson
   d) All of the mentioned

   Ans. Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   a) True
   b) False

   Ans. 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

   Ans. Hypothesis

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

   Ans.0

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. 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. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a bell curve.

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

Ans. Here are the most common ways of handling missing data

**Zero Replacement**: Here, you replace the missing value with zero irrespective of everything.

**Min or Max Replacement**: Replace the missing value with the minimum or maximum value of a feature.

**Mean/ Median/ Mode Replacement**: Replace missing value with mean or median or most frequent feature value.

Also, one can replace the value of the missing cell with the previous cell's value. This kind of technique is popular while inputting time series data. For example, if the price of an instrument is missing on the i-th day, it makes sense to replace it with the (i-1)-th day's price.

12. What is A/B testing?

Ans. A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. `Is mean imputation of missing data acceptable practice?

Ans. Yes, the quick and easy workaround is to substitute a mean for numerical features and use a mode for categorical ones. Even better, someone might just insert 0's or discard the data and proceed to the training of the model.

14. What is linear regression in statistics?

Ans. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

15. What are the various branches of statistics?

Ans. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.

FLIP ROBO