# STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.
   a) True
   b) False

**Answer:** a) 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

**Answer:** a) 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

**Answer:** b) 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

FLIP ROBO

**Answer:** d) All of the mentioned

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

**Answer:** c) Poisson

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

**Answer:** 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

**Answer:** b) 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

**Answer:** a) 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

**Answer:** c) Outliers cannot conform to the regression relationship

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

**Answer:** The probability distribution known as the normal distribution, sometimes it is referred to as the Gaussian distribution or the standard normal distribution, which predicts all of its values symmetrically, with the majority of the outcomes clustered around the probability's mean. The values that are being plotted above or below the mean is equal.

FLIP ROBO

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

**Answer:** There are two conditions to handle the missing data either remove the missing (NaN values) from data through using EDA method or replace the missing values with most frequent data or average of data that are available in data set.

Imputation Techniques Recommended –

1. Supervised Machine Learning
   1.1. Regression
   1.2. Classification
2. EDA method

12. What is A/B testing?

**Answer:** A/B testing is used to compare the effectiveness of two models by analyzing the data through statistical analysis like mean, median, mode, percentile & standard deviation & compare the same which will give the result into binomial metrics.

13. Is mean imputation of missing data acceptable practice?

**Answer:** Mean imputation is referred to replacing the missing values of data to either mean value (in case of numerical data) or use a mode (frequently used category).
This is a NOT acceptable practice as it leads to get to ML for any model to either biased or one sided as it

sometimes shows the correlation between the replaced data with the other categories totally unacceptable.

14. What is linear regression in statistics?

**Answer:** Linear regression is one of the most used method for predictive analysis for any kind of data. It works over the linear equation i.e. y = a + bx + e where y is the dependent variable & x is the independent variable. It also provides the continuous label (output) where as Logistic regression provides only binary output.

15. What are the various branches of statistics?

**Answer:** There are two main branches for statistics that are used for predictions: -

1. Descriptive Statistics – This is a branch of statistics which implements the method of collecting & presenting the statistical data.
2. Predictive Statistics – This is the branch of statistics which deals with the techniques used for analysing the data & drawing conclusions.

FLIP ROBO