**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.
   1. True
   2. False
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?
   1. Central Limit Theorem
   2. Central Mean Theorem
   3. Centroid Limit Theorem
   4. All of the mentioned
3. Which of the following is incorrect with respect to use of Poisson distribution?
   1. Modeling event/time data
   2. Modeling bounded count data
   3. Modeling contingency tables
   4. All of the mentioned
4. Point out the correct statement.
   1. The exponent of a normally distributed random variables follows what is called the log- normal distribution
   2. Sums of normally distributed random variables are again normally distributed even if the variables are dependent
   3. The square of a standard normal random variable follows what is called chi-squared distribution
   4. All of the mentioned
5. random variables are used to model rates.
   1. Empirical
   2. Binomial
   3. Poisson
   4. All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
   1. True
   2. False
7. 1. Which of the following testing is concerned with making decisions using data?
   1. Probability
   2. Hypothesis
   3. Causal
   4. None of the mentioned
8. 4. Normalized data are centered at and have units equal to standard deviations of the original data.
   1. 0
   2. 5
   3. 1
   4. 10
9. Which of the following statement is incorrect with respect to outliers?
   1. Outliers can have varying degrees of influence
   2. Outliers can be the result of spurious or real processes
   3. Outliers cannot conform to the regression relationship
   4. None of the mentioned

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

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

Normal Distribution is a symmetrical distribution and has equal mean , median and mode.

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

There are many ways to handle missing data.some of them are:

* Droping the column
* Deleting the row, if missing data are less.
* Mean substitution
* Median substitution
* Frequently observed data or Maximum likelihood
* The KNN : by finding the closest neighbors

1. What is A/B testing?

A/B testing is also known as bucket testing or split-run testing.A/B tests consist of a randomized experiment with two variants, A and B.It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

1. Is mean imputation of missing data acceptable practice?

Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations.

1. What is linear regression in statistics?

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.

1. What are the various branches of statistics?

Two branches, *descriptive statistics* and *inferential statistics*

#### Descriptive Statistics:The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

#### Inferential Statistics:The branch of statistics that analyzes sample data to draw conclusions about a population.

