**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

**Answer: 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) Modelling event/time data

b) Modelling bounded count data

c) Modelling contingency tables

d) All of the mentioned

**Answer: b (Modelling Bounded Count Data)**

4. Point out the correct statement.

a) The exponent of 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

**Answer: d (All of the Above)**

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: 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: 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)**

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

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

**Answer:** NORMAL DISTRIBUTION is also called as BELL CURVE, is a distribution that occurs naturally in many situations. For Examples in Exams, most of the students score Average Marks while less Students score Bad marks and less number of students score Good marks.  
In Normal Distribution the mean is 0 & the Standard Deviation is 1.

Normal Distribution are Symmetrical but not all the Symmetrical Distributions are Normal.

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

**Answer:** A better approach towards handling missing data is **Imputation**. Imputation means to replace or fill the missing data with some value.

Another Approach to handle the Missing data is to remove the Data means removes all data for an observation that has one or more missing values

**What imputation techniques do you recommend?**

The simplest imputation method is replacing missing values with the mean or median values of the dataset at large, or some similar summary statistic. This has the advantage of being the simplest possible approach.

12. What is A/B testing?

**Answer:** An AB testing, a process whereby a hypothesis is made about the relationship between two data sets and those data sets are then compared against each other to determine if there is a statistically significant relationship or not.

A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

**Answer:** Using Mean is not a Good Practice as it reduces the model’s accuracy and bias the results but Mean Imputation is easy to use and It can be applied to both continuous and categorical variables which makes it advantageous over other imputation algorithms

14. What is linear regression in statistics?

**Answer:** **Linear Regression in Statistics:** It is the next step up after correlation. It is used when we want to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable.

It is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regressions

**There is different type of Regression Techniques like:**

Logistic Regression

Ridge Regression

Lasso Regression

Linear Regression etc.

15. What are the various branches of statistics?

**Answer:** There are 2 Main Branches of Statistics:

A: **Descriptive Statistics**: It Deals with the presentation & collection of Data. This is used to describe and understand the feature and characteristics of a specific Data set by giving short summaries about the Data.

1. Mean
2. Median
3. Mode

These are the part of Descriptive Statistics.

B: **Inferential Statistics:** It helps us to make predictions from the Data and make inferences from the sample and generalize them into Population

It tries to make conclusion (prediction) about the population that is beyond the data available