**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: - 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 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 mentioned**

5. \_\_\_\_\_\_ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

**Answer: - c) Poisson**

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

a) True

b) False

**Answer: - b) 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

**Answer: - b) 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

**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: - Normal Distribution is also called as bell curve. It 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 a normal distribution the mean is zero and the standard deviation is 1. It has zero skew. The normal distribution is the most important probability distribution in statistics because it fits many natural phenomena. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution. The normal distribution model is motivated by the**[**Central Limit Theorem.**](https://www.investopedia.com/terms/c/central_limit_theorem.asp)

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

**Answer: -** **Data can be missing in the following ways:-**

* **Missing Completely At Random (MCAR):**When missing values are randomly distributed across all observations, then we consider the data to be missing completely at random.
* **Missing At Random (MAR):** The key difference between MCAR and MAR is that under MAR the data is not missing randomly across all observations, but is missing randomly only within sub-samples of data.
* **Not Missing At Random (NMAR):** When the missing data has a structure to it, we cannot treat it as missing at random.

**Imputation Techniques: -**

**1. Mean or Median Imputation**

## 2. Multivariate Imputation by Chained Equations (MICE)

## 3. Random Forest

12. What is A/B testing?

**Answer: - A/B testing also known as split testing. An AB test is an example of statistical hypothesis 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.**

13. Is mean imputation of missing data acceptable practice?

**Answer: - It is a non-standard, it uses Random Forest. It is use to predict the missing data. It also can be used for both i.e. continuous as well as categorical data and so it makes advantageous over other imputations.**

14. What is linear regression in statistics?

**Answer:-** **Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.**

**A linear regression line has an equation of the form *Y = mx + c*, where *X* is the explanatory variable and *Y* is the dependent variable. The slope of the line is *m*, and *c* is the intercept (the value of *y* when *x* = 0).**

15. What are the various branches of statistics?

**Answer: - Various branches of statistics are given below: -**

1. **Descriptive Methods**
2. **Analytical Methods**
3. **Inductive Methods**
4. **Inferential methods**
5. **Applied methods**