**STATISTICS WORKSHEET-1**

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: Modelling bound 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:** 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. Usually replacing the standard error by its estimated value does change the CLT. a) True b) False

**Ans: 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

**Ans: 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

**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: Outliers can have varying degrees of influence**

**Ans: Outliers can have varying degrees of influence**

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

**Ans:** Normal distribution is an arrangement of data/values in the middle of the range. It always produces a bell shaped curve with symmetry towards the end. As it is centrally aligned mean ,median and mode are all the same. It is also known as Gaussian distribution.

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

**Ans:** We have to see the feature of missing value first then depending on the dependency of it on the outcome and on the other features it should be handled.

Different types of imputation techniques are:

1. Impute missing values with mean or median.
2. Impute with frequent data i.e mode.
3. Impute with arbitrary values but rarely recommended.
4. Using KNN method i.e nearest neighbor values.

12. What is A/B testing?

**Ans:** A/B testing is a kind of split testing where the dataset is divided into two part and subjected to hypothesis testing to find the variation among both the group.

The groups are selected by random selection process to avoid the biasedness.

13. Is mean imputation of missing data acceptable practice?

**Ans:** Not always , but it totally depends on the type of dataset and its dependency on the null values that how it will be effecting the outcomes.

14. What is linear regression in statistics?

**Ans:** Linear Regression predicts the value of a variable based on the value of the other variable. It draws a relationship between an explanatory data (independent value, x) and dependent value, y i.e represented by a straight line y=a+bx

15. What are the various branches of statistics?

**Ans:** Descriptive Statistics and Inferential statistics.

Descriptive statistics describes that is going on in a dataset or population whereas inferential statistics refers to the findings done on the sample data and the impact of it on the population as a whole. It actually describes data.

Inferential statistic is a way of making inferences about the population based on samples. It helps in predicting data.