**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 id 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: b) Central Mean 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

**ANSWER:**

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**

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:** In a normal distribution, data is symmetrically distributed with no skew. When plotted on a graph, the data follows a bell shape, with most values clustering around a central region and tapering off as they go further away from the center. Normal distributions are also called Gaussian distributions or bell curves because of their shape.

* The mean, median and mode are exactly the same.
* The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.

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

**Answer:** Fill the missing data or Null values as follows:

1. Delete the unwanted data.
2. Fill the N/A values by mean, mode, by fillna method with the help of pandas.

12. What is A/B testing?

13. Is mean imputation of missing data acceptable practice?

**ANSWER:** Outliers data points will have a significant impact on the mean and hence, in such cases, it is not recommended to use the mean for replacing the missing values. Using mean values for replacing missing values may not create a great model and hence gets ruled out.

14. What is linear regression in statistics?

**ANSWER:** In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables also known as dependent and independent variables. The case of one explanatory variable is called simple linear regression for more than one the process is called multiple linear regression.

15. What are the various branches of statistics?

**ANSWER:** There are 2 main branches of statistics.

1). **Descriptive Statistics**: The branch of statistics that focuses on collecting, summarizing, and presenting a set of data. E.g. the variation in the weight of 100 boxes of cereal selected from a factory's production line.

2). **Inferential Statistics:** The branch of statistics that analyses sample data to draw conclusions about a population. E.g. In a class of 100 students of maths , 70 students got more than 60% of marks.