

## **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

**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 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- 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. 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**

**Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.**

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

**Answer-**The Normal Distribution is a probability bell curve. It is symmetric about mean and indicates values near mean occurs more frequently than the values that are farther away from the mean.

In a Normal Distribution the mean is zero and standard deviation is one.

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

**Answer-**Imputer is nothing but filling null values. Fillna is the also techniques to fill null values . But there are advanced techniques available to fill the nulls , so technically we called as Imputations. To fill the null values we use imputation techniques. There are various imputation techniques we called as Imputers.

Imputer is basically classified as :

- Single Imputer
- Knn Imputer
- Iterative Imputer

### **12. What is A/B testing?**

**Answer-**A/B Testing is essentially an experiment where two or more versions of variables are shown to users at random , and statistical analysis is used to determine which variation perform better for a given conversion goal. Analysis of A/B Testing is :

- If the  $p\text{-value} < 0.05$  , we can say that the alternate hypothesis( $H_1$ ) is False which means we can reject the null( $H_0$ ) Hypothesis.
- If the  $p\text{-value} > 0.05$  ,we can say that the alternate hypothesis( $H_1$ ) is True which means we failed to reject the null( $H_0$ ) Hypothesis.

### **13. Is mean imputation of missing data acceptable practice?**

Answer-Yes, mean imputation of missing data is acceptable practice. Simple Imputer takes average of all data and fills mean in missing data.

### **14. What is linear regression in statistics?**

Answer-Linear Regression is one of the fundamental and widely known Machine Learning Algorithms. The building blocks of a Linear Regression model are :-

- Discrete/Continuous independent variable.
- A best-fit regression line.
- Continuous dependent variable i.e., A Linear Regression model predicts the dependent variable using a regression line.

The general equation of Linear Regression is :  $y = mx + c$

### **15. What are the various branches of statistics?**

Answer-There are two real branches of statistics:- Descriptive Statistics and Inferential statistics.