

## 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
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
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
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
5. \_\_\_\_\_ random variables are used to model rates.  
a) Empirical  
b) Binomial  
☒ c) Poisson  
d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.  
a) True  
☒ b) False
7. I. Which of the following testing is concerned with making decisions using data?  
a) Probability  
☒ b) Hypothesis  
c) Causal  
d) None of the mentioned
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
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

**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?
11. How do you handle missing data? What imputation techniques do you recommend?
12. What is A/B testing?
13. Is mean imputation of missing data acceptable practice?
14. What is linear regression in statistics?
15. What are the various branches of statistics?

### **Answers to Q10 - Q15**

10. A normal distribution is a probability distribution that signifies that data is symmetric about the mean, whereas outliers are more than three standard deviations away from the mean.

11. Missing data can be handled in a variety of ways. The technique used will depend on the kind and size of data that is under study. Some of these are: Most Frequent Value, K Nearest Neighbors, Mean Value, Median Value, Next or Previous Value, Linear Interpolation.

12. A/B testing is a method of comparing two iterations of a website or software to see which one performs better.

13. Yes. Mean imputation is one of the techniques used to handle missing data.

14. Linear regression is a statistical model used to predict the value of a variable (dependent) based on the value of another variable (independent).

15. Descriptive statistics and inferential statistics.

