

## 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 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 – A Normal distribution is referred to as standard state probability distribution that is symmetric bell-shaped curve where mean, median and mode is zero and Standard deviation is  $\pm 1$ .**

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

**Answer – There is NO good way to deal with missing data but you can either delete the data or you can use imputation techniques, recommended imputation method is replacing missing values with the mean or median values.**

12. What is A/B testing?

**Answer – A/B testing or split testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal, it is a method of comparing two versions of a webpage or app against each other to determine which one performs better.**

13. Is mean imputation of missing data acceptable practice?

**Answer – Yes, this has the advantage of being the simplest possible approach, and one that doesn't introduce any undue bias into the dataset.**

14. What is linear regression in statistics?

**Answer – Linear regression is a type of predictive analysis and These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.**

**Simplest form of linear regression =  $y = mx + c$**

- **y = Dependent variable**
- **x = Independent variable**

15. What are the various branches of statistics?

**Answer – There are mainly 2 branches of statistics (Descriptive statistics and Inferential statistics)**