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

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

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

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

d) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

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

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

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

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?

Ans) A normal distribution, sometimes called the bell curve, is a distribution that occurs naturally in many situations.

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

Ans) Techniques to handle missing data-

1)Use deletion methods to eliminate missing data. The deletion methods only work for certain datasets where participants have missing fields.

2)Use regression analysis to systematically eliminate data.

Mean imputation. Perhaps the easiest way to impute is to replace each missing value with the mean of the observed values for that variable.

12. What is A/B testing?

Ans)A/B testing 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?

Ans)Yes, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased

14. What is linear regression in statistics?

Ans)linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line.

15. What are the various branches of statistics?

Ans)There are two branches of statistics:

1)Descriptive statistics :: Deals with the presentation and collection of data. This is usually the first part of a statistical analysis.

2) inferential statistics :: involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.