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. 1. 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

WORKSHEET

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

A) Normal distribution is also called as Gaussian distribution or bell curve. The data is symmetrically distributed with no skew. In this the mean, median and mode are equal.

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

A) Imputation means replacing the missing data with substituted value based on available information.

Imputation Techniques-

Single Imputation-

1. Hot Deck - Selected similar record
2. Cold Deck- Selects donors from another data set (Past surveys)
3. Mean Substitution- Replacing with the mean of variable.
4. Non-Negative matrix factorization- Ignore missing data.
5. Regression- Predict observed value of a variable based on other variables.

Multiple Imputation: 1. Missing completely at Random (MCAR)

2. Missing at Random (MAR)

3. Missing not at Random (MNAR)

4. Structured Missingness

Normal Imputation- By using mean, median, mode if data is numerical.

12. What is A/B testing?

A) It is also called as Bucket testing or split testing. It is a way to compare multiple version of a single variable to decide which performs better/ more effective and help in decision making. With this business can change their user interface for more reachability.

13. Is mean imputation of missing data acceptable practice?

A) Mean imputation is not acceptable because it is easy but it has 2 major problems.

1. Mean imputation does not preserve the relationships among variables

2. It leads to an underestimate of standard errors.

14. What is linear regression in statistics?

A) Linear regression is a statistical model which estimates relationship between a scalar response and one or more explanatory variables (dependent and independent variables).

15. What are the various branches of statistics?

A) There are 2 main branches in statistics.

1. Descriptive Statistics- Involves organising, summarizing and displaying data.

a. Measure of central tendency

a1. Mean

a2. Median

a3. Mode

b. Measure of Variability

b1. Range

b2. Variance

b3. Dispersion

2. Inferential Statistics- Uses sample to draw conclusion.