WORKSHEET

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

Ans: 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

Ans: 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

Ans: 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

Ans: D – All the mentioned

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans: C - Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False

Ans: 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

Ans: 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
Ans: 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
Ans: 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?

Ans: The Normal Distribution is a symmetrical probability distribution where most results are located in the middle and few are spread on both sides. It has the shape of a bell and can entirely be described by its mean and standard deviation.

Most of the datasets are follow the normal distribution.

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

Ans: We can handle missing data by:

- Mean/Median Value
- Delete rows with missing value
- Impute missing values for categorical variable
- Prediction of missing value
- KNN imputer
- Iterative imputer

Recommended techniques for handle missing data is **Mean value and KNN imputer.**

12. What is A/B testing?

Ans:

13. Is mean imputation of missing data acceptable practice?

Ans: Yes

14. What is linear regression in statistics?

Ans: Regression is used for relationship between dependent variable and independent variable.

A Linear Regression model predicts the dependent variable using a regression line based on the independent variables. The equation of the Linear Regression is:

y=a+b*x+e

Where, a is the intercept, b is the slope of the line, and e is the error term. The equation above is used to predict the value of the target variable based on the given predictor variable(s).

15. What are the various branches of statistics?

Ans: Descriptive Statistics and Inferential Statistics