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

Answer-a

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

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

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-c

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

Answer-c

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

Answer-b

7. 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

8. 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

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

## 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?**

Answer-

The Normal Distribution is defined by the probability density function for a continuous random variable in a dataset. It given by gaussian distribution.

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

Answer-

There are some general approaches to handle missing data by the data deletion and imputation techniques.

Several imputation techniques are such as.

1.Mean/median/ mode imputation

2.KNN imputation.

3.Linear regression imputation.

4.Predictive modelling.

**12. What is A/B testing?**

A/B testing, also known as split testing, is a method used to compare two versions of a product or service to determine which performs better. It's commonly used in marketing, user experience design, and various other fields to optimize outcomes.

**13. Is mean imputation of missing data acceptable practice?**

Answer-

Mean imputation, where missing values are replaced by the mean of the available data in that column, is a simple and straightforward method for handling missing data.

14. **What is linear regression in statistics?**

Linear regression is a fundamental and widely used statistical technique for modelling the relationship between a dependent variable and one or more independent variables. It assumes that this relationship is linear, meaning it can be represented by a straight line or plane.

**15. What are the various branches of statistics?**

There are several branches of statistics as per many aspect such that:

1.Descriptive Statistics

2.Inferential Statistics

3.Probability and random variables

4.Time series analysis.

5.Bayesian Statistics

6.Machine learning

==========
==========