STATISTICS WORKSHEET

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:
c) The square of a standard normal random variable follows what is called chi-squared distribution

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. 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. 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

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

Normal Distribution: Normal distribution, also known as Gaussian distribution, is a symmetric probability distribution that is characterized by a bell-shaped curve. In a normal distribution, the majority of the data falls around the mean, and the probabilities decrease as you move away from the mean in both directions. The standard normal distribution has a mean of 0 and a standard deviation of 1, and any normal distribution can be transformed into this standard form using the z-score.

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

Handling Missing Data: Handling missing data is crucial in data analysis. Imputation techniques involve replacing missing values with estimated values
Some common imputation techniques :
include mean imputation, median imputation, mode imputation, and more advanced methods like k-nearest neighbors (KNN) imputation or regression imputation.
The choice of technique depends on the nature of the data and the extent of missingness.

12.What is A/B testing?

A/B Testing: A/B testing, or split testing, is a statistical method used in marketing and product development to compare two versions (A and B) of a variable to determine which one performs better. It involves randomly assigning subjects to two groups and exposing them to different variations, then analyzing the differences in outcomes to assess the impact of changes.

13.Is mean imputation of missing data acceptable practice?

Mean Imputation of Missing Data: Mean imputation involves replacing missing values with the mean of the observed values. While it is a simple and quick method, it has limitations. It can introduce bias, underestimate variability, and distort relationships in the data. In some cases, mean imputation may not be appropriate, especially if data is not missing completely at random.

14.What is linear regression in statistics?

Linear Regression in Statistics: Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and aims to find the best-fitting line (regression line) that minimizes the sum of squared differences between observed and predicted values.

15.What are the various branches of statistics?

Branches of Statistics: Various branches of statistics include:

Descriptive Statistics: Involves summarizing and describing the main features of a dataset.
Inferential Statistics: Involves making inferences or predictions about a population based on a sample of data.
Biostatistics: Applies statistical methods to biological and medical data.
Econometrics: Applies statistical methods to economic data.
Social Statistics: Focuses on statistics related to social phenomena.
Mathematical Statistics: Involves theoretical aspects of statistical methods and their mathematical foundations.