STATISTICS WORKSHEET-1

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

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

They were first called “normal” because the pattern occurred in many different types of common measurements. There are many normal curves. Even though all normal curves have the same bell shape, they vary in their center and spread. ... The mean of a normal distribution locates its center.

- A normal distribution is the proper term for a probability bell curve.

- In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
- Normal distributions are symmetrical, but not all symmetrical distributions are normal.
- In reality, most pricing distributions are not perfectly normal.

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

1. Deletions

    - Pairwise Deletion
    - Listwise Deletion/ Dropping rows
    - Dropping complete columns

2. Basic Imputation Techniques

    - Imputation with a constant value
    - Imputation using the statistics (mean, median, mode)

3. K-Nearest Neighbor Imputation

Imputation techniques:-

- Imputation with constant value:
- Imputation using Statistics:
- K_Nearest Neighbor Imputation:

12. What is A/B testing?

A/B testing is a shorthand for a simple randomized controlled experiment, in which two samples (A and B) of a single vector-variable are compared. These values are similar except for one variation which might affect a user's behavior. A/B tests are widely considered the simplest form of controlled experiment.

13. Is mean imputation of missing data acceptable practice?

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?

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables)

15. What are the various branches of statistics?

There are **two main branches** of statistics
- Inferential Statistic.
- Descriptive Statistic.

Inferential Statistics:
Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

Descriptive Statistics:
Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.