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

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

3. Which of the following is incorrect with respect to use of Poisson distribution?
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 (this one is the correct ans.)

5. _____ random variables are used to model rates.
c) Poisson

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

7. Which of the following testing is concerned with making decisions using data?
b) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

9. Which of the following statement is incorrect with respect to outliers?
c) Outliers cannot conform to the regression relationship

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

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

Ans: - Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve".

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

Ans:- Missing data can be dealt with in a variety of ways.

One of the simplest way is we delete the row who have missing values but that's not the perfect way to handle it we have to analyse our data set.

Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

- Mean imputation
  Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.
- Hot deck imputation
- Cold deck imputation
- Regression imputation
- Stochastic regression imputation

12. What is A/B testing?

Ans: - A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

13. Is mean imputation of missing data acceptable practice?

Ans:- Acceptable but dangerous .
Bad practice in general
If just estimating means: mean imputation preserves the mean of the observed data
Leads to an underestimate of the standard deviation
Distorts relationships between variables by "pulling" estimates of the correlation toward zero

14. What is linear regression in statistics?

Ans: - Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of an output variable (or response) based on the value of an input (or predictor) variable.

15. What are the various branches of statistics?

Ans: - Two branches, descriptive statistics and inferential statistics, comprise the field of statistics.
Descriptive statistics is considered as the first part of statistical analysis which deals with collection and presentation of data. Scientifically, descriptive statistics can be defined as brief explanatory coefficients that are used by statisticians to summarize a given data set.

Inferential statistics are techniques that enable statisticians to use the gathered information from a sample to make inferences, decisions or predictions about a given population. Inferential statistics often talks in probability terms by using descriptive statistics. These techniques are majorly used by statisticians to analyse data, make estimates and draw conclusions from the limited information which is obtained by sampling and testing how reliable the estimates are.