STATISTICS WORKSHEET-1

1. (b) False

2. (a) Central Limit Theorem

3. (b) Modeling bounded count data

4. (d) All of the mentioned

5. (c) Poisson

6. (b) False

7. (b) Hypothesis

8. (a) 0

9. (c) Outliers cannot conform to the regression relationship

10. Normal Distribution: It is also called as the "Gaussian distribution". The Normal distribution is mostly seen as continuous distribution in nature like height, age marks, salary etc. This distribution is most widely used distribution of all distributions. Every event in this distribution is independent from one another. Here mean, median and mode are equal as all are lined up such that the center of the distribution is the mean. The Shape of the normal Distribution curve is like a bell so due to which its curve is also called as Bell curve.

11. Missing data in statistics affects the validity of the final result. There are two primary methods to solve the error which are: Imputation and removal of data

    Imputation method: In this process missing data are replaced by the estimated values. This approach preserves all cases by replacing the missing data with a probable value estimated by other available information.

    Removal of data: In case of random missing data then related data can be deleted to reduce biasness.

    Regression Imputation is the best way in handling missing data as in this process missing data are replaced with a probably estimated value by other available information. This process has an advantage that it retains a great deal of data over the other methods and avoids significantly altering the standard deviation or the shape of the distribution. Standard error is reduced while sample size is increased in this process.

12. A/B Testing: It is basically called Statistical Inferences and it is an analytical method for making decisions that estimates population parameters based on sample statistics. A/B testing consist

of randomized experiment with two variants named A and B. A/B tests are widely considered the simplest form of controlled experiment. Many professions use the data from A/B tests. This includes data engineers, marketers, designers, software engineers, and entrepreneurs.

13. Mean imputation replaces missing values of certain variable by the mean of non missing cases of that variable.

It is not acceptable in terms of missing data due to its major disadvantages which are:

a) Mean imputation reduces the variance of the imputed variables.

b) Mea n imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence level.

c) Mean imputation doesn't preserves relationships between variables such as correlations.

14. Linear regression in statistics is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables which are known as dependent and independent variables. In linear regression the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data.

The simplest form of the regression equation with one dependent and one independent variables is defined by the formula: y= c + b*x, where

y = estimated dependent variable score

c = constant

b = regression coefficient and x = score on the independent variable

There are two types of linear regression:

a) Simple linear regression

b) Multiple linear regression

15. Statistics is a branch of mathematics used to summarize, analyse and interpret a group of number or observation.

There are two branches of Statistics:

a) Descriptive Statistics: It deals with the presentation and collection of data. This is actually the first part of a statistical analysis. It further contains Measures of center and measures of dispersion.

b) Inferential Statistics: Procedure that allows researchers to infer or generalize observations made with the samples to the large population from which they were selected.