Statistics Worksheet – 1

A1.a):- True

A2.a):- central Limit Theorem

A3.b):- Modeling Bounded Count Data

A4.d):- All Of the Mentioned

A5.c):- Poisson

A6.b):- False

A7.b):- Hypothesis

A8.a):- 0

A9.c):- outliers cannot conform to the regression relationship

A10:- 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.

A11:- The best way to handle missing data is by imputation or the removal of data.

❖ **Imputation methods:-**

1. <u>Mean, Median and Mode</u>:- This is one of the most common methods of imputing values when dealing with missing data. In cases where there are a small number of missing observations, We can calculate the mean or median of the existing observations.

2. <u>Time-Series Specific Methods</u>:- Another option is to use time-series specific methods when appropriate to impute data. There are four types of time-series data:-
   - No trend or seasonality.
   - Trend, but no seasonality.
   - Seasonality, but no trend.
   - Both trend and seasonality.

- ❖ **Removal of Data:-** There are two primary methods for deleting data when dealing with missing data.
  1. <u>List wise</u> :- In this method, all data for an observation that has one or more missing values are deleted. The analysis is run only on observations that have a complete set of data. If the data set is small, it may be the most efficient method to eliminate those cases from the analysis
  2. <u>Dropping variables</u>:- If data is missing for more than 60% of the observations, it may be wise to discard it if the variable is insignificant.

A12. A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B.

It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective.

A13. 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. That's a good thing.

also, by imputing the mean, you are able to keep your sample size up to the full sample size. That's good too.

A14. In statistics, **linear regression** is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression, for more than one the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

A15. Statistics plays a main role in the field of research. It helps us in the collection, analysis and presentation of data. In this blog post we will try to learn about the two main branches of statistics that is descriptive and inferential statistics.

1. **Descriptive Statistics:-** It describes the important characteristics/ properties of the data using the measures the central tendency like mean/ median/mode and the measures of dispersion like range, standard deviation, variance etc.

2. **Inferential Statistics:-** It is about using data from sample and then making inferences about the larger population from which the sample is drawn. The goal of the inferential statistics is to draw conclusions from a sample and generalize them to the population. It determines the probability of the characteristics of the sample using probability theory. The most common methodologies used are hypothesis tests, Analysis of variance etc.