WORKSHEET SET 1 - STATISTICS 1

ANSWERS

Internship 28
Name: Vageesh

A1. A) True

A2. A) Central Limit Theorem

A3. B) Modeling bounded count data

A4. D) All of the mentioned

A5. E) Poisson

A6. B) False

A7. B) Hypothesis

A8. A) 0

A9. C) Outliers cannot conform to the regression relationship

A10. A Normal Distribution is the case where the data is symmetricaly distributed. If a line/curve is made to represent the distribution, then it will be symmetrical and makes a bell shape. There is a mid region where most of the data points are accumulated and they gradually decrease while moving either left or right direction. In a normal distribution mean, mode and median are equal and data is distributed symmetrical about the mean.

A11. Missing data in a dataset is handled by imputation technique (replacing missing data with some value) or deleting the entire record/row if most of the cells are empty or insignificant.
I choose mean imputation or imputing zero or other constant value in place of missing data as they are easy and quick ones to handle small data sets.

A12. A/B Testing is a metod to run simultaneous tests between models or products with two versions/variants to get data samples from a test or a target user base (in case of product). A measuring factor is finalised before starting the test which will help in the judgement of the better variant out of two. After the test is complete, the data generated for both variants is analysed and compared.

A13. Mean imputation for imputing missing data is common used technique but not a preferred practice in statistics for making estimate/predictions due to its many disadvantages like variance of the data is highly affected if more values are imputated by this method.

A14. In statistics, linear regression describes the relation between two variables in a data. An independent variable and other is dependent variable which changes with the increase in value of the independent variable. The relation can be defined by a linear equation ‘y = a +bx’. Here y is dependent variable and x is independent variable. ‘b’ is the slope of the line representing the x-y relation. It helps to analyse the data set for growth, downfall, variance etc.

A15. There are two branches of statistics. Descriptive and Inferential Statistics.
Descriptive Statistics:
On the basis of a sample data collected, finding variance in the data and values like mean, mode and median. Can be represented in the form of graph.
Inferential Statistics:
Techniques used to collect data and analyse it to make estimations / predictions and also testing the reliability of the of those estimates.