

## STATISTICS WORKSHEET-1

- 1 a) True
- 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 confirm to the regression relationship

10 NORMAL DISTRIBUTION- A normal distribution is a bell shaped frequency distribution curve of a continuous random variable. In normal distribution mean=median=mode=0 and standard deviation is  $\pm 1$

11 Missing data can be dealt in a number of ways the most common is to ignore it. Another common way is imputations. Imputation is the process of substituting an estimate for missing values.

### IMPUTATION TECHNIQUES

1 MEAN IMPUTATION

2 SUBSTITUTION

3 REGRESSION IMPUTATION

4 STOCHASTIC REGRESSION IMPUTATION

12 A/B testing also known as split testing is a methodology for comparing two versions of a web page or app against each other to determine which one performs better. It is essentially an experiment where two or more variants of a page are shown to users at random and statistical analysis is used to determine which variation performs better.

13 No, mean imputation of missing data is not acceptable practice because it ignores feature correlation and also mean imputation decreases the variance of our data while increasing bias, as a result of reduced variance the model is less accurate.

14 linear regression is defined as a statistical practice that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events.

15 There are two branches of statistics

- a) Descriptive statistics
- b) Inferential statistics

