STATISTICS WORKSHEET-1

1.Bernoulli random variables take (only) the values 1 and 0.
Ans: 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?
Ans: a) Central Limit Theorem

3.Which of the following is incorrect with respect to use of Poisson distribution?
Ans: b) Modeling bounded count data

4. Point out the correct statement.
Ans: d) All of the mentioned

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

6. Usually replacing the standard error by its estimated value does change the CLT.
Ans: a) True

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

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

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

10. **Normal distribution**
In a normal distribution, data is symmetrically distributed with no skew. Most values cluster around a central region, with values tapering off as they go further away from the center. The measures of central tendency (mean, mode and median) are the same in a normal distribution.

11. In statistics, imputation is **the process of replacing missing data with substituted values**. When substituting for a data point, it is known as "unit imputation"; when substituting for a component of a data point, it is known as "item imputation".

Imputation techniques:

- Mean or Median imputation
- Multivariate Imputation by Chained Equations (MICE)
- Random Forest

12. **A/B Testing:**

A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better

13. For studies that compare different statistical methods, the number of imputations should be even larger than the percentage of missing observations, usually between 100 and 1000, to control the Monte Carlo error.

14. Simple linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative

15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics