STATISTICS WORKSHEET 1

1. Bernoulli random variable takes only the values 0 and 1. The statement is **True**
2. **Central Limit Theorem** states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases
3. Modelling bounded count data is incorrect with respect to use of Poisson distribution. It is used for modelling **unbounded** count data
4. All of the mentioned statements are correct
5. **Poisson** distribution random variables are used to model rates
6. **False**. Usually replacing the standard error by its estimated value doesn't change the CLT.
7. **Hypothesis** testing is concerned with making decisions using data.
8. Normalized data are cantered at **0** and have units equal to standard deviations of the original data.
9. c) Outliers cannot conform to the regression relationship
10. **NORMAL DISTRIBUTION:**

    It represents a symmetric distribution where most of the observations cluster around the central peak called as mean of the distribution. It has a **bell curve** and is also called as **Gaussian** distribution. It has two parameters mean and standard deviation. It is very important because it helps to normalise the data. The mean, median and mode of normal distribution are **equal**.
11. The common ways of handling missing values is by deletions and imputations. Two types of deletions are listwise deletion and pairwise deletion. Imputation is the method of substituting missing data with substituted values. The best data imputation technique is to calculate a

statistical value for each column (such as a mean) and replace all missing values for that column with the statistic. It is a popular approach because the statistic is easy to calculate using the training dataset and because it often results in good performance.

12. A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment. It is an analytical method for making decisions that estimates the population parameters based on sample statistics.
13. Mean imputation of missing data is not an acceptable practise because it leads to an underestimate of standard errors and it does not preserve the relationship among variables.
14. Linear regression analysis is a commonly used and basic type of predictive analysis. It is used to predict the value of the variable based on the value of another variable.
15. Various branches of statistics are
    - Mathematical or theoretical statistics
    - Statistical methods or functions
    - Descriptive statistics
    - Inferential statistics