# STATISTICS WORKSHEET - 1

1. **Bernoulli random variables take (only) the values 1 and 0.**

**Ans:** a) 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) Modelling 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:** b) False

7. **Which of the following testing is concerned with making decisions using data?**

**Ans:** b) Hypothesis

8. **Normalized data are centred 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. **What do you understand by the term Normal Distribution?**

**Ans:** The Normal Distribution, commonly known as the Gaussian Distribution in probability theory and statistics, is the most significant continuous probability distribution. It is also known as a bell curve. In every physical discipline and economics, the normal distribution either nearly or exactly represents a huge number of random variables. Furthermore, it can be used to approximate other probability distributions, hence validating the use of the term 'normal' as in the most commonly used.

11. **How do you handle missing data? What imputation techniques do you recommend?**

**Ans**: Numerous strategies can be employed to address missing data. I think ignoring it is the most usual response. On the other hand, if you choose to remain undecided, your statistical programme will decide for you. Most of the time, your programme will remove items in a listwise order. Listwise deletion might or might not be a good idea, depending on why and how much data is lost. Imputation is another tactic that people who pay attention frequently employ. The process of imputed values being analysed as if they were the actual observed values involves replacing missing values with an approximation. The technics can be used are Mean imputation, Substitution, Hot deck imputation, Cold deck imputation, Regression imputation.

**12. What is A/B testing?**

**Ans:** A/B testing is a type of controlled experiment that is frequently used in business to evaluate and compare two different options or strategies, like two different versions of a website design or advertisement, to determine which is more effective at bringing in and keeping customers, or at achieving a particular goal, like higher sales or user engagement. It entails splitting a user base into two groups, exposing each group to one of two options, and then evaluating the outcomes to ascertain which option is superior.

**13. Is mean imputation of missing data acceptable practice?**

**Ans:** Since mean imputation ignores feature correlation, it is generally seen as a bad approach. The variance of the imputed variables is decreased using mean imputation. The majority of hypothesis tests and the confidence interval computation are rendered incorrect when mean imputation reduces standard errors. Correlations and other relationships between variables are not preserved by mean imputation.

**14. What is linear regression in statistics?**

**Ans:** Applying a linear equation to observed data is the goal of linear regression, which aims to illustrate the relationship between two variables. It is intended for one variable to function as an independent variable and the other as a dependent variable. For instance, a person's weight and height have a linear relationship. As a result, this indicates a linear relationship between the subject's height and weight. A person's weight increases in proportion to their height.

**15. What are the various branches of statistics?**

**Ans:** Descriptive Statistics: Ideas the area of statistics that is concerned with gathering, compiling, and displaying data.

Inferential Statistics: The concept of inferential statistics the area of statistics that uses sample data analysis to make population-level inferences.