STATISTICS WORKSHEET-1

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

   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?**

   a) Central Limit Theorem

3. **Which of the following is incorrect with respect to use of Poisson distribution?**

   b) Modeling bounded count data

4. **Point out the correct statement.**

   b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

5. **_____random variables are used to model ratgs.**

   c) Poisson

6. **10. Usually replacing the standard error by its estimated value does change the CLT.**

   b) False

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

   b) Hypothesis

8. **Normalized data are centered at_____and have units equal to standard deviations of the original data.**

   a) 0

9. **Which of the following statement is incorrect with respect to outliers?**

   c) Outliers cannot conform to the regression relationship

10. **What do you understand by the term Normal Distribution?**

The normal distribution is a function that describes how the values of a variable are distributed. It is a distribution where most of the value cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions.

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

Average imputation and common-point imputation. Average imputation uses the average value of the responses from other data entries to fill out missing values. Common-point imputation, on the other hand, is when we utilise the middle point or the most commonly chosen value. For example, on a five-point scale, the substitute value will be 3. When utilising this method is the three types of middle values: mean, median and mode, which is valid for numerical data.

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

A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics.

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

- Bad practice in general
- If just estimating means: mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by “pulling” estimates of the correlation toward zero

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

In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).

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

There are TWO branches of statistics:

- **Descriptive statistics:** It deals with the presentation and collection of data. This is usually the first part of a statistical analysis.
- **Inferenatial statistics:** It involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.