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) Modeling bounded count data

4. Point out the correct statement.
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. What do you understand by the term Normal Distribution?
Ans: A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans: We can delete the particular column if it is not contributing much to target variable.

Or we can use Imputation technique for the missing or nan values. Fill the missing values with mean values of that Column or use mode or frequent occurred values for categorical data.

I recommend to use simple Imputer. Where, replace missing values using a descriptive statistic (e.g. mean, median, or most frequent) along each column, or using a constant value.

12. What is A/B testing?
    Ans: From the Data Science perspective, A/B testing is a form of statistical hypothesis testing or a significance test.

13. Is mean imputation of missing data acceptable practice?
    Ans: It is acceptable but sometimes it's a bad idea.
    As it might not b suitable in some kind of datasets.

- Mean imputation reduces the variance of the imputed variables.
- Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.
- Mean imputation does not preserve relationships between variables such as correlations.

14. What is linear regression in statistics?
Ans: Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.

15. What are the various branches of statistics?
Ans: 1) Data Collection
2) Descriptive Statistics
3) Inferential Statistics