Statistics Worksheet -1 Name of the Intern: Ravindra Hanchate.

1.Bernnoulli random variables take (only) the values 1 and O.
Ans: a). True.

2. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of standard normal as the sample size inreases?
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. 10.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 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: In normal distribution, data is symmetrically distributed with no skew. Most values cluster around the center region, with values tapering off as they go further away from the center. The measures of central tendency (mean,mode,median) are exactly the same in a normal distribution. A Normal distribution is the proper form for probability bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans: When dealing with missing data, we can use two primary methods to solve the missing data, imputation technique or removal of data.
When data is missing , it make sense to delete data as mentioned above. However, that may not be the most effective option. Instead of deletion of data, we have multiple imputation solutions for missing data. Depending why the data are missing, imputation methods can deliver reasonably reliable results.

Different types of Imputation techniques:
1. **Mean, Mode, Median**
2. **Time-series specific method.**
3. **Carried forward fill or backrward fill method.**
4. **Linear Interpolation.**
5. **Simple Imputer.**

12. What is A/B testing?
Ans: A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

13. Is mean imputation of missing data acceptable practice?
Ans: Mean imputation is simple yet not advisable, on the other hand there are many alternatives to mean imputation that provide much more accurate estimates and standard errors. The quick and easy workaround is to substitute a mean for numerical features and use a mode for categorical ones. Mean imputation ignores feature correlation.

14. What is linear regression in statistics ?
Ans: Linear regression analysis is used to predict value of a variable based on the value of another variable. The variable you want to predict is called dependent variable and the variable you are using to predict the other variables value is called independent variable.
In statistics , linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression and for more than one, the process is called multiple regression.

15. What are the various branches of statistics?
Ans: There are two main branches of statistics.
1. **Descriptive Statistics: It deals with the presentation and collection of data. This is usually the first part of Statistical analysis.**
2. **Inferential statistics: It involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. It is the inferences that make studies important and this aspect is dealt with in inferential statistics. Most predictions of the future and generalization of the population by studying smaller sample come under the purview of inferential statistics.**