STATISTICS-WORKSHEET1

1) **A) true**
2) **a) Central Limit Theorem**
3) **b) Modeling bounded count data**
4) **d) All of the mentioned**
5) **c) Poisson**
6) **b) False**
7) **b) Hypothesis**
8) **a) 0**
9) **c) Outliers cannot conform to the regression relationship**

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

In order to be considered a normal distribution, a data set (when graphed) must follow a bell-shaped symmetrical curve centered around the mean.

Normal distributions follow the empirical rule, also called the 68-95-99.7

rule.

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

Missing data can be handled by either imputation or removing data.

Imputation technique:

1) impute using mean/median values
2) impute using most frequent or zero/constant values
3) impute using k-nn
4) multivariate imputation by chained equation

12. What is A/B testing?

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.

13. Is mean imputation of missing data acceptable practice?

No, 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?

Linear regression is a basic and commonly used type of predictive analysis. 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

1) data collection

2) descriptive statistics

3)inferential statistics