STATISTICS WORKSHEET

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.
d) All of the mentioned

5. _____ random variables are used to model rates.
c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
b) False

7. 1. Which of the following testing is concerned with making decisions using data?
b) Hypothesis - The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

8. 4. 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

WORKSHEET Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?
Answer: it is type of distribution where we can measure with the normal curve when we have continues data. It is also called as Belled shaped curve. In ideal condition mean=median=mode=0 and in standard deviation = +-1

11. How do you handle missing data? What imputation techniques do you recommend?
Answer: By decreasing the power of the analysis by decreasing the effective sample size. Techniques used are Knn imputation and Gaussian mixture imputation

12. What is A/B testing?
Answer: It is a hypothesis testing where A&B are variants in which A is controlled and B is variation.

13. Is mean imputation of missing data acceptable practice?
Answer: It is acceptable when the missing value is not large enough.

- 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?
Answer: It is a Statistical method for determining the relation between two or more variables. It involves using an independent variable which is x to predict dependent variable which is y

15. What are the various branches of statistics?
Answer:

1. Descriptive statistics – are procedures used to summarize, organize and make sense of a set of scores
2. Inferential statistics – are procedures used that allow researchers to generalize, observations made with samples to the large population from which they are selected