Assignment – 2

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.

Ans :d) All of the mentioned

5. _____ random variables are used to model rates.

Ans: a) Empirical

6. 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: d) None of the mentioned

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

Ans: normal distribution is a part of probability distribution normal distribution bell curve graph is give a idea a how data is and there is also 2 possibilities left curve and right curve

We have formula to delete this outliers but its depends upon data we delete this or not and we have to standardise data (-1 to 1)

if the sample size is more than 30 samples so this data is normaly distributed

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

Ans: we have multiple option to handle of missing data like mean ,median,mode and fillna or fill data is upword ya downword either if the missing data is small so we can delete this data

so specified data if we want to fill missing data but it's depend upon the data and variables

If the data is continuous so I personally choose mean mode cause its less risky to fill null values

And if data is discrite data so I choose mode method or median method .

12. What is A/B testing?

Ans: A/B testing is basically is the split to various differnet webside and test which webside is impact more so it was very helpful to focus on things and also we also find why this webside is imapact more so we can improve this

13. Is mean imputation of missing data acceptable practice?

Ans: short ans is yes, but its also depend upon data and variables we can use this or not

Cuz data is very sensible and if we fill wrong data it's impact on prediction and accuracy also

14. What is linear regression in statistics?

Ans: linear regression is one of the ml algorithm and it's help to prediction

Also help to which mode we invest more to get more revenue.

15. What are the various branches of statistics?

Ans: mean,median, mode (central tendancy)

Variance and standard deviation (measure of dispertion)