STATISTICS WORKSHEET-1

1. Bernoulli random variables take (only) the values 1 and 0.

a) True b) False

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?

a) Central Limit Theorem

b) Central Mean Theorem

c) Centroid Limit Theorem

d) All of the mentioned

ANS = A) Central limit theoram

3. Which of the following is incorrect with respect to use of Poisson distribution?

a) Modeling event/time data b) Modeling bounded count data

c) Modeling contingency tables d) All of the mentioned

ANS = b) modelling bounded count data

4. Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution

b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

c) The square of a standard normal random variable follows what is called chi-squared distribution
d) All of the mentioned

ANS = D) All of the mentioned

5. _____ random variables are used to model rates.

a) Empirical b) Binomial c) Poisson d) All of the mentioned

ANS = C) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

a) True b) False

ANS = B) False

7. Which of the following testing is concerned with making decisions using data?

a) Probability b) Hypothesis c) Causal d) None of the mentioned

ANS = B) Hypothesis

8. Normalized data are centered at_____and have units equal to standard deviations of the original data.

a) 0 b) 5 c) 1 d) 10

ANS = A) 0

9. Which of the following statement is incorrect with respect to outliers?

a) Outliers can have varying degrees of influence

b) Outliers can be the result of spurious or real processes

c) Outliers cannot conform to the regression relationship d) None of the mentioned

ANS = C) outliers cannot conform to the regression relationship

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

ANS= normal distribution is one kind of the probability distribution. The normal distribution show the data near to the mean value using the bell curve. Normal distribution is the ideal distribution.in the normal distribution the mean is 0 and the standard deviation is +-3.all Normal distribution is symmatric distribution but not all the symmatric distribution are not the normal distribution. Data compose in the normal distribution is around 99.7% the distribution of the data in 3 standard deviation in the first deviation the data around 66% is in first SD. In 2^{nd} SD is around 95% of data and lastly the 3^{rd} is around 99.7% of all the data conjumed. The left 0.3% data is called the outliers.

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

In the real world in the datasets the some values or the some data are missing in that datasets so when we get output on that dataset we are not getting the exact value or answer so we firstly fill that missing data. In filling the data we use sevral of methods but we use normally 3 or 4 methods the first and very useful method to fill the data is mean of the that column but when the data are rising way or data in decrease way that situation the mean imputation tech are not getting right answer the second method is the mode of the values. And 3^{rd} method is the median of that column and the last method are used is the getting the column as a class and getting the blank values using the nearest class of 0 or 1 this is the imputation techniques me recommend.

12. What is A/B testing?

The A/B testing is also called the split test. In that testing we get the two variance or two version first is A and second is B in that A/B testing we changes the data or values but the main work of the A/B testing is when we change A value the B value will be constant the second value are not changes simultaneously B to A that time we create the situation where the only that thing will be changing and rest things are as same.

13. Is mean imputation of missing data acceptable practice?

It is acceptable when the missing value proportion is not large enough.But, when the missing values are large enough and you impute them with themean, the standard errors will be lesser than what they actually would have been.Small standard errors can lead to small p-values and this can create problems for us,because some variables will start appearing significant, which are ideally not significant.

14. What is linear regression in statistics?

Linear regression is an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events. The independent variable is also the predictor or explanatory variable that remains unchanged due to the change in other variables. However, the dependent variable changes with fluctuations in the independent variable. Thus, linear regression is a supervised learning algorithm that simulates a mathematical relationship between variables and makes predictions for continuous or numeric variables such as sales, salary, age, product price, etc.This analysis method is advantageous when at least two variables are available in the data.

15. What are the various branches of statistics?

Mainly in statistics two main branches are available. The first one is the descriptive statistics and second one is inferential statistics.

In descriptive statistics there are 3 type:-

1) Frequency distribution 2) central tendency 3) variability

In central tendency there also 3 type of data 1) mean 2) mode 3) median

In variability there also 3 type of data 1) Range 2) Varience 3) standard deviation