Statistics assignment worksheet

1. Bernoulli random variables take (only) the values 1 and 0.
Answer-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?
Answer-b) Central Mean Theorem
3. Which of the following is incorrect with respect to use of Poisson distribution?
Answer- a) Modeling event/time data
4. Point out the correct statement.
Answer- d) All of the mentioned
5. _____ random variables are used to model rates
Answer-b) Binomial

6.Usually replacing the standard error by its estimated value does change the CLT.
Answer- a) True
7. 1. Which of the following testing is concerned with making decisions using data?
Answer-b) Hypothesis

8 Normalized data are centered at_____and have units equal to standard deviations of the original data.
Answer-a) 0
9. Which of the following statement is incorrect with respect to outliers?
Answer-c) Outliers cannot conform to the regression relationship
10. What do you understand by the term Normal Distribution?
Answer-The Normal Distribution is a bell shaped frequency distribution curve of a continuous random variable. normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of Normal distributions are symmetrical, but not all symmetrical distributions are normal
11. How do you handle missing data? What imputation techniques do you recommend?
Answer- Average imputation and common point imputation are the two techniques in which average imputation uses the average value of responses from the other data entries to fill out the missing values

12. What is A/B testing?
Answer- A/B testing is also known as split run testing. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or two-sample hypothesis testing as used in the field of statistics.
13. Is mean imputation of missing data acceptable practice?
Answer-Mean imputation isthepracticeof replacing null values in a data set with the mean of the data.Mean imputation is generally bad practicebecause it doesn't take into account feature correlation.
14. What is linear regression in statistics?
Answer- Linear regression is the statistical method that allows us to summarize the relationship between two continues variables i.e quantitative variables The equation of linear regression is similar to the slope formula what we have learned before in earlier classes such as linear equations in two variables. It is given by; "Y= a + bX Now, here we need to find the value of the slope of the line, b, plotted in scatter plot and the intercept, a

15 What are the various branches of statistics?

Answer- there are mainly two branches of statistics
1 descriptive statistics
2 inferential statistics

Descriptive Statistics

In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or standard deviation.

Descriptive statistics is a way to organise, represent and describe a collection of data using tables, graphs, and summary measures.

the collection of people in a city using the internet or using Television.

Descriptive statistics are also categorised into four different categories:

- Measure of frequency
- Measure of dispersion
- Measure of central tendency
- Measure of position

Inferential Statistics

This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.

Inferential Statistics is a method that allows us to use information collected from a sample to make decisions, predictions or inferences from a population. It grants us permission to give statements that goes beyond the available data or information.

deriving estimates from hypothetical research.