Assignment 3 for DS2311

Statistics Worksheet

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

Answer : A

Question 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: A

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

Answer : C

Question 4: Point out the correct statement.

Answer : A

Question 5: _____ random variables are used to model rates.

Answer: C

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

Answer: B

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

Answer: B

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

Answer: A

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

Answer: C

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

Answer : The Normal Distribution, also known as the Gaussian Distribution, is a continuous

probability distribution that is symmetric around its mean. The distribution is a bell-shape, and the shape is determined by two parameters: the mean and the standard deviation.

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

Answer: There are several strategies to handle missing data like KNN, Regression, Deletion and using Deep Learning Models.

Imputation Techniques that I recommend is KNN and Regression.

Question 12. What is A/B testing?

Answer: A/B testing is a popular statistical method used to determine which of two or more versions of a webpage, app, or email performs better.

Question 13. Is mean imputation of missing data acceptable practice?

Answer: Mean imputation is generally not considered best practice in most modern statistical and data analysis contexts.

Question 14. What is linear regression in statistics?

Answer: Linear regression is a statistical method used to model and analyze the relationship between a dependent variable (Y) and one or more independent variables (X). It's a fundamental tool in various fields like statistics, machine learning, economics, and social sciences.

Question 15. What are the various branches of statistics?

Answer:

- Descriptive Statistics: It involves organizing and summarizing data to give a picture of the structure of a population. Examples include mean, median, mode, range, etc.
- Inferential Statistics: It involves drawing conclusions about a population based on a sample. Examples include hypothesis testing, confidence intervals, etc.
- Probability Theory: It provides the mathematical foundation for the calculation of probabilities and the interpretation of the results of statistical tests. Examples include the law of large numbers, central limit theorem, etc.
- Bayesian Statistics: It involves updating our beliefs in the light of new evidence, using Bayes' theorem. It provides a principled way to handle uncertainty and model complexity.
- Multivariate Statistics: It involves dealing with data that has more than one dimension, such as multivariate analysis of variance, regression, etc.
- Time Series Analysis: It involves analyzing a sequence of data points indexed in time order, such as auto correlation, auto regressive models, etc.
- Non parametric Statistics: It involves statistical methods that do not assume a particular distribution of the data, such as permutation tests, non parametric regression, etc.
- Statistical Machine Learning: It involves the development of predictive models based on statistical techniques, such as linear regression, decision trees, neural networks, etc.

- Bayesian Machine Learning: It involves using Bayesian statistics to build probabilistic models of the data and their parameters, such as Gaussian processes, Bayesian neural networks, etc.