FlipRobo

Name- Neha Somvanshi

Batch- DS2307

Statistics - STATISTICS WORKSHEET-1

Q1) Bernoulli random variables take (only) the values 1 and 0.

a) True

b) False

Q2) 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

Q3) 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

Q4) 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

Q5) _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the mentioned

Q6) Usually replacing the standard error by its estimated value does change the CLT.

a) True

b) False

Q7) Which of the following testing is concerned with making decisions using data?

a) Probability

b) Hypothesis

c) Causal

d) None of the mentioned

Q8) Normalized data are centered at_____ and have units equal to standard deviations of the original data.

a) 0

b) 5

c) 1

d) 10

Q9) 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

Q10) What do you understand by the term Normal Distribution?

A10) Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

Q11) How do you handle missing data? What imputation techniques do you recommend?

A11) There are several strategies you can employ to deal with missing data:

- Domain Knowledge
- Indicator Variables
- Impute Missing Data
- Predictive Modeling

Q12) What is A/B testing?

A12) A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

Q13) Is mean imputation of missing data acceptable practice?

A13) Mean imputation of missing data is a practice that is often discouraged in modern statistics and data analysis due to its potential drawbacks and limitations. While it is a simple method, it can lead to biased or inaccurate results in certain situations.

Q14) What is linear regression in statistics?

A14) Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Q15) What are the various branches of statistics?

A15) Their two main branches of statistics are:

- Descriptive Statistics.
- Inferential Statistics.