

STATISTICS WORKSHEET-1

Only one correct answer. Choose the correct option to answer your question

1. a) True
2. a) Central Limit Theorem
3. b) Modelling bounded count data
4. d) All of the mentioned
5. c) Poisson
6. b) False
7. b) Hypothesis
8. a) 0
9. c) Outliers cannot conform to the regression relationship

subjective answer type questions, Answer them in briefly.

10.

- Normal distribution, also known as the Gaussian distribution, which 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.
- The normal distribution model is motivated by the Central Limit Theorem.
- This theory states that averages calculated from independent, identically distributed random variables have approximately normal distributions, regardless of the type of distribution from which the variables are sampled

11.

- Missing data is an inevitable part of the process. Many data researchers pour a lot of resources, time and energy into making sure the data set is as accurate as possible.
- Sometimes, data sets come up short, no matter how many times data scientists clean and prepare it.
- **The best way to handle such situations is to develop contingency plans to minimise the damage.**

Few imputation techniques which I would recommend are:

- **Hot deck imputation:** A randomly chosen value from an individual in the sample who has similar values on other variables.
- **Cold deck imputation:** A systematically chosen value from an individual who has similar values on other variables.
- **Regression imputation:** The predicted value obtained by regressing the missing variable on other variables.
- **Interpolation and extrapolation:** An estimated value from other observations from the same individual. It usually only works in longitudinal data.

12.

A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13. Is mean imputation of missing data acceptable practice?

Ans: **YES**

14.

linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables.

15.

There are three branches of Statistics such as data collection, descriptive statistics and inferential statistics.