**Statistics Worksheet - 1**

In Q1 to Q9, only one option is correct, choose the correct option:

1. (A) True
2. (A) Central Limit Theorem
3. (B) Modeling bounded count data
4. (D) All of the mentioned
5. (C) Poisson
6. (B) False
7. (B) Hypothesis
8. (A) 0
9. (C) Outliners cannot conform to the regression relationship

In Q10 and Q15, are subjective answer type questions, Answer them briefly.

1. What do you understand by the term normal distribution?

**Answer:** A normal distribution is the most important probability distribution for independent and random variables. It is mostly known for its bell curve shaped. Normal distribution is mostly symmetrical around mean and the values are further away considered as equally in both directions. Extreme values are mostly unlikely in both tails of distribution. But not all symmetrical distribution is normal.

Normal distribution is also known as Gaussian distribution.

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

**Answer:** Missing data can occur in many ways or reasons. When mining the data, there will always be a chance to lose the data. It happens due to nonresponse, blanks, Incompatible version/libraries, missing data in the table etc.

There are 3 types of Imputation techniques such as,

* Numerical Variables – Mean, Median and Mode, End of tail etc.
* Categorical Variables – Frequency & Adding missing data.
* Random Sample Imputation

1. What is A/B testing?

**Answer:** A/B testing is a randomized process where two variables compare to find out which performs better. It is basically a hypothesis testing or inference where to make analytical method for making decisions based on sample variables/statistics.

1. Is mean imputation of missing data acceptable practice?

**Answer:** Mean imputation is generally considered as unbiased since its doesn’t show correlation.

1. What is linear regression in statistics?

**Answer:** Linear regression is a kind of statistical analysis that shows a relationship between two variables. There are two types of variables such as,

* Dependent Variable – Which you want to predict
* Independent Variable – Which you using to predict another variable.

1. What are the various branches of statistics?

**Answer:** There are 3 types of branches in statistics such as,

* Descriptive Statistics – Which describes the properties of sample and population data (Tools such as Mean, Variance, Skewness, Kurtosis)
* Inferential Statistics – Which is used to test hypothesis and draw conclusion (Tools such as Linear, ANOVA, Null hypothesis)
* Data Collection – It’s all about how data is collected. It is further classified/divided as frequency distribution, discrete and continuous data.