Statistics Worksheet 1

Name: Avichal Vishwakarma

Batch: 1844

Ans01- A) True

Ans02- A) Central Limit Theorem

Ans03- B) Modeling bounded count data

Ans04- D) All of the mentioned

Ans05- C) Poisson

Ans06)- B) False

Ans07)- B) Hypothesis

Ans08)- A) 0

Ans09)- C)

Ans10)- Normal Distribution is also called “Bell Curve”, it represents the frequency of occurrence. The frequency is more in the center and gradually reduces as the standard deviation is increased. Usually, it is symmetric about the mean frequency of the dataset.

Ans11)- There are various method to deal with missing data using python.

The basic method includes:

- Replacing the missing data with mean or mode (df.fillna())
- Deleting rows/columns containing missing data(df.dropna())

The advanced method includes the use of imputation techniques. Imputation techniques are selected by data scientists based on various factors. The recommended imputation techniques are:

- KNN Imputer
- Iterative Imputer

Ans12)- A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs. variant B and concluding which of the two variants is more effective.

Ans13)- Mean Imputation is not considered an ideal practice because of the following three problems:

- Mean imputation reduces the variance of the imputed variables.
- Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.
- Mean imputation does not preserve relationships between variables such as correlations.

Ans14)- Linear regression is a statistical technique which can be used to predict the value of an unknown variable based on the value of a known variable.

Ans15)- There are two broad categories of statistics:

1. Descriptive Statistics
2. Inferential Statistics

Descriptive statistics: Descriptive statistics is used to present the data in an understandable way, so that a meaningful description can be made.

Inferential Statistics: This is a branch of statistics which deals with techniques used for analysis of data, making estimates that lead to predictions and drawing conclusions or inferences.