1)A)True

2)B)Central limit theorom

3) B)modeling bounded count data

4) D)All of the above

5) C)Poison distribution

6) A)True

7) B)Hypothesis testing

8) A)0

9) C)outliers cannot confirm to the regression relationship

10) In statistics normal distribution is also called as gaussian distribution.

It is a probability distribution that shows the symmetry around mean value that the data is near the mean are more frequent in occurrence than data far from mean.

Normal distribution model is key to central limit theorem (CLT) this theorem states that average 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 values or data that is not stored for some variables in given dataset.

NaN is a representation of missing values in pandas.

Errors regarding missing data is solved either by imputation or the removal of data.

Imputation is effective when percentage of missing data is low.

Most common methods of imputing values:

1] Mean, median, mode

2]Time series specific methods

3]Last observation carried forward (LOCF) and next observation carried backward (NOCB)

4]Linear interpolation

5]Seasonal adjustment with linear interpolation

6]K nearest neighbors

12)A/B testing also called as split testing or split-run testing it is a user experience research

methodology.

It includes application of statistical hypothesis testing or two-sample hypothesis testing.

It is used to compare multiple versions of a single variable and determines which of the variants is more effective.

13)No mean imputation is a terrible practice as it ignores feature correlation.

14)Linear regression is used to predict the value of variable based on the value of another variable.

The variable you want to predict is dependant variable and the variable which is used to predict the other variable's value is called independent variable.

The best fit line is a line that fits the given scatter plot in best way. Mathematically the best fit line is obtained by minimizing the residual sum of squares(RSS).

15)Statistics is defined as the discipline that concerns with the collection, organization, analysis, summarization, interpretation and presentation of data.

There are two main branches of statistics :

1]Descriptive statistics: It is a summary statistic that quantitatively describes features of collected information.

2]Inferential statistics: It is a process of data analysis to deduce properties of an underlying probability distribution.