Name- Sayan Paul

1.a. True

2.a Central limit theorem

3.b. Modeling bounded count data

4.d. All the mentioned

5.c. Poisson

6.b False

7.b. Hypothesis testing

8.a. 0

9.c. Outliers cannot conform to the regression relationship

10. Frequency of data is highest at the mean of the data. As we go away from the mean on either sides the frequency of data decreases gradually. When normally distributed data is plotted against frequency, a bell shaped curve is obtained.

11.If the number of rows containing NaN value is very less as compared to the total number of rows in the dataset, then we can delete the rows using dropna(axis=0). But if the number of rows containing NaN values is not negligible as compared to the total number of rows in the dataset, then we have to replace the NaN values with mean, median or mode of the data present in that column. This task can be executed using replace() method of pandas or SimpleImputer of scikitlearn library.

12. A/B testing is a user experience research methodology. A/B testing is a way to compare two variables of a single variable typically by testing a subject's response to variant A against variant B and determining which of the two variants is more effective.

Ex- Website or app of a particular company is working better?

13. Mean imputation is a good practice but not always the best practice. We have to observe the trends in the data before doing any imputation. If we observe that a particular data is appearing many times in a column, then it is better to do mode imputation. Also median imputation can be done in some cases.

14. Linear regression is a very common ML algorithm based on statistics. A linear equation if formed using the train_x and train_y dataset. The linear equation consists of all the relevant features in the table . The weightage of the features are given in the form of coefficients of the equation.
When the train_x and train_y data are plotted, it is noted that the best fit line is created in such a way that the square of distance of the points from the straight line should be minimum.

15. The branches of statistics are descriptive statistics and inferential statistics.