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) Outliers cannot conform to the regression relationship

10. The term Normal Distribution is a continuous probability distribution that is symmetrical around its mean, most of the observations cluster around the central peak, and the probabilities for values further away from the mean taper off equally in both directions.

11. When there is data missing in the data set the simple method would be to drop the rows of the missing data but if the other values in the row are essential for analysis the other method would be to usually replace the values using the mean in each column if the data is numeric and not skewed. If the data is numeric and skewed, we would replace the missing values using the median. If the data is string, then the mode is used to replace the missing string values.
Imputation Techniques that are recommended:

1. Complete Case Analysis (CCA)
   In this we directly remove the rows that have missing data i.e we consider only those rows where we have complete data.
2. Arbitrary Value Imputation
   In this we either put the missing data as “Not defined” or “Nan” or we group the missing values in a column and assign them to a new value that is far away from the range of that column.
3. Frequent Category Imputation
   In this we replace the missing value with the variable with the highest frequency or in simple words replacing the values with the Mode of that column.

12. A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics. It's a common method used in marketing, web design, product development, and user experience design to improve campaigns and goal conversion rates.

13. Mean imputation of missing data is not an acceptable practice for the following reasons:

- It reduces the variance of the imputed variables.
- It shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.
- It does not preserve relationships between variables such as correlations.
- It can lead into severely biased estimates even if data are Missing completely at random (MCAR).

14. Linear regression is a kind of statistical analysis that attempts to show a relationship between two variables, it looks at various data points and plots a trend line. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. It is commonly used for predictive analysis and modeling.

15. The branches of statistics are:

- Data collection
- Descriptive statistics
- Inferential statistics