Statistics Solutions

1. True (a)
2. Central Limit Theorem (a)
3. Modelling bounded count data (b)
4. All of the mentioned (d)
5. Poisson (c)
6. False(b)
7. Hypothesis (b)
8. 0 (a)
9. Outliers cannot conform to the regression relationship (c)

10. What do you understand by the term Normal Distribution?
    - The Data is usually said to be normally distributed when the mean, median, mode is equal. It is usually denoted by bell curve. Normal distribution is symmetrical, but not all symmetrical distributions are normal.
11. How do you handle missing data? What imputation techniques do you recommend?
    - Firstly, one needs to identify if the data is missing and if yes, we have to check if there is any patten is being followed in the missed data or if it is happening at random. And then we need to use appropriate solution depending on the kind of problem.
    - There are seven methods with the help of which one can make up for the missing data. They are:
      1.Mean Imputation 2. Substitution 3. Hot Deck Imputation 4. Cold Deck imputation
      5. Regression Imputation 6. Stochastic Regression Imputation
      7. Interpolation & Extrapolation

12. What is A/B Testing?
    - A/B testing is a method to compare two versions of a variable to measure which performs better in a controlled environment. It is a basic randomized control experiment.
13. What are the various branches of Statistics?
    - There are two main branches: Descriptive & Inferential
    - Descriptive Statistics: If the data can be explained without any statistical tools, then it is called as descriptive statistics (Usually for small groups of data).
    - Inferential Statistics: If the data is too large to compute, then we use Inferential statistics.
14. Is mean Imputation of missing data acceptable practice?
    - It is basically used to predict the missing data. It can be used for both continuous as well as categorical data. But it has some limitations too:
      1. Mean Imputation does not preserve the relationship among variables. It preserved the mean of observed data.
      2. Mean Imputation leads to underestimate the standard errors.

15. What is Linear Regression in Statistics?
    - It is a statistical model which helps us understand the relationship between two variables i.e., dependent variable and independent variable. The case of one explanatory variable is called simple linear regression and for more than one it's called multiple linear regression.