

## Answers to MCQs

1. a) True

2. a) Central Limit Theorem

3. b) Modeling bounded count data

4. d) All of the mentioned

5. c) Poisson random variables are used to model rates.

6. b) False - Usually replacing the standard error by its estimated value does not change the CLT.

7. b) Hypothesis

8. a) Normalized data are centered at 0

9. c) Outliers cannot conform to the regression relationship.

10. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

11. We can use either **isna()** or **isnull()** method to detect missing values in data. We can get the total of missing values in each column with **sum()** or take the average with **mean()**.

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 drives business metrics.

**13. Mean imputation** is typically considered terrible practice since it ignores feature correlation. Secondly, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

**14. Linear regression** is a type of statistical analysis used to predict the relationship between two variables. It assumes a linear relationship between the independent variable and the dependent variable, and aims to find the best-fitting line that describes the relationship. The line is determined by minimizing the sum of the squared differences between the predicted values and the actual values

**15.** There are two main branches of statistics: **descriptive statistics and inferential statistics**. **Descriptive statistics** deals with the presentation and collection of data. This is usually the first part of a statistical analysis. **Inferential statistics**, as the name suggests, involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics.