

1. Bernoulli random variables take (only) the values 1 and 0. True
2. c) Centroid Limit Theorem
3. b) Modelling bounded count data
4. d) All of the mentioned
5. c) Poisson
6. Usually replacing the standard error by its estimated value does change the CLT. False
7. Which of the following testing is concerned with making decisions using data?
Hypothesis
8. Normalized data are centered at ____0 ____ and have units equal to standard deviations of the original data.
9. Which of the following statement is incorrect with respect to outliers?
c) Outliers cannot conform to the regression relationship
10. Normal Distribution: A normal distribution (bell curve), is a distribution that occurs naturally. This creates a distribution that resembles a bell. The bell curve is symmetrical. Half of the data will fall to the left of the mean and half will fall to the right. The empirical rule tells you what percentage of your data falls within a certain number of standard deviations from the mean:
 - 68% of the data falls within one standard deviation of the mean.
 - 95% of the data falls within two standard deviations of the mean.
 - 99.7% of the data falls within three standard deviations of the mean.
11. How do you handle missing data? What imputation techniques do you recommend?

Understanding the nature of missing data is critical in determining what treatments can be applied to overcome the lack of data.

Data can be missing in the following ways:

Missing Completely At Random (MCAR): When missing values are randomly distributed across all observations, then we consider the data to be missing completely at random. A quick check for this is to compare two parts of data – one with missing observations and the other without missing observations. On a t-test, if we do not find any difference in means between the two samples of data, we can assume the data to be MCAR.

Missing At Random (MAR): The key difference between MCAR and MAR is that under MAR the data is not missing randomly across all observations, but is missing randomly only within sub-samples of data. For example, if high school GPA data is missing randomly across all schools in a district, that data will be considered MCAR. However, if data is randomly missing for students in specific schools of the district, then the data is MAR.

Not Missing At Random (NMAR): When the missing data has a structure to it, we cannot treat it as missing at random. In the above example, if the data was missing for all students from specific schools, then the data cannot be treated as MAR

12. What is A/B testing?

It is 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.

13. Is mean imputation of missing data acceptable practice?

No, it is not acceptable practice. Mean imputation preserves the mean of the observed data. It also leads to an underestimate of the standard deviation. It also distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. What is linear regression in statistics?

It quantifies the relationship between one or more variable and one outcome variable. Linear regression is commonly used for predictive analysis and modelling. For example, it can be used to quantify the relative impacts of age, gender. Linear regression is also known as multiple regression, multivariate regression and regression.

15. What are the various branches of statistics?

- Descriptive Statistics
- Inferential Statistics

Descriptive Statistic: It deals with the presentation and collection of data. For example, a physicist studying turbulence in the laboratory needs the average quantities that vary over small intervals of time. The nature of this problem requires that physical quantities be averaged from a host of data collected through the experiment.

Inferential Statistics: It involves drawing the right conclusions from the statistical analysis that has been performed using descriptive statistics. In the end, it is the inferences that make studies important and this aspect is dealt with in inferential statistics.