

## Statistics worksheet

### Answers

1. A) True
2. A) Central Limit Theorem
3. A) Modeling event/time data
4. D) All of above
5. C) Poison
6. B) False
7. C) Causal
8. A) 0
9. B) Outliers can be result of spurious or real processes
10. In Normal distribution the data is evenly distributed with mean of 0 and standard deviation on 1. In graphical representation it can be illustrated with a perfect bell-shaped curve where maximum data falls in centre and very less data falls at either ends of the curve. In normal distribution mean, median and mode are equal.
11. There are a few suggestion for handle missing data, depending of type of data set and importance of data in corresponding columns/features one can choose to below actions
  - a. Deleting rows.. can be considered if values in missing in important columns and rest columns are of less importance. This method is highly accurate and is mostly used to handle null values removing the data will be reduce the data size. This method is not recommended the percentage of missing values is high.
  - b. Replacing the missing values with mean/median/mode, it helps to avoid data reduction and is very useful when the data is linear. This method is useful in small data sets
  - c. Assigning an unique category. This will help when the data present have definite or known categories for Example if the data set is of cars it can be categorized into sedan, sports and SUV segments, so adding this feature will results into change of variance and prevent loss of data. However, adding an extra feature may result in poor model performance
  - d. Predicting the missing values: if the missing values are estimated to have less variance this method proves to be very helpful. One can check with algorithms which gives best accuracy.
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. A/B testing is one of the components of the overarching process of Conversion Rate Optimization (CRO), using which you can gather both qualitative and quantitative user insights. it eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.
13. Imputing the mean preserves the mean of the observed data. So, if the data are missing completely at random, the estimate of the mean remains unbiased. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution, as the resulting outcome can be either overestimated of under estimated.
14. Linear Regression is making estimation based on data available after analysing relationship between the coefficient and its variation on corresponding variable.

15. There are two main branches of statistics
  - a. Descriptive statistics
  - b. Inferential Statistics