

Statistics worksheet 1 Solutions.

1. A) TRUE
2. A) Central limit theorem
3. B)Modelling 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. A normal distribution is the term for a probability bell curve. Showing that data near the mean are more frequent in occurrence than data far from the mean.
11. Imputation is the process of substituting and estimate for missing values and analysing the entire data set if the imputed values were the true observed values. We can use either `isna()` an a 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 is a way to compare two versions of a single variable, typically by testing a subjects response to variant a against b, and determining which of the two variant is more effective.
13. Mean imputation is typically considered terrible practice since it ignores feature correlation.
14. Linear regression is the most basic and commonly used predictive analysis. regression estimates are used to describe data and to explain the relationship.
15. There is two main branch of statistics:
 - 1.descriptive statistics,
 - Measure of Central tendency(mean, mode, median)
 - Measure of dispersion(variance and standard deviation)
 - 2.Inferential statistics (Sampling data and infer the result to describe entire population.)