

Statistics Assignment Answer Sheet

Q-1

1. A) True
 2. A) Central Limit Theorem
 3. B) Modeling 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 an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.
11. When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data.
- The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.
12. A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.
- For instance, let's say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

13. Bad practice in general

- If just estimating means: mean imputation preserves the mean of the observed data
- Leads to an underestimate of the standard deviation
- Distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

15. There are three real branches of statistics:

- data collection
- descriptive statistics
- inferential statistics