

STATISTICS WORKSHEET- 1

Sol1] A (true)

Sol2] A (central limit theorem)

Sol3] B (modeling bounded count data)

Sol4] D (all of the mentioned)

Sol5] C (poisson)

Sol6] A (true)

Sol7] B (hypothesis)

Sol8] A (0)

Sol9] C (outliers cannot conform to the regression relationship)

Sol10] 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. in graph form, normal distribution will appear as a bell curve.

Sol11] Best techniques to handle missing data

- 1 use deletion methods to eliminate missing data. the deletion methods only work for certain datasets where participants have missing fields.
- 2 use regression analysis to systematically eliminate data.
- 3 Data scientist can use data imputation techniques.

I will recommend data imputation techniques i.e complete case analysis (CCA): this is a quite straight forward method of handling the missing data, which directly removes the rows that have missing data i.e we consider only those rows where we have complete data i.e data is not missing.

Sol12] A/B testing- (also known as bucket testing or split-run testing) is a user experience research methodology. A/B test consist of a randomized experiment with two variants, A and B. it includes application of statistical hypothesis testing or “two sample hypothesis testing” as used in the field of statistics.

A/B testing is a way to compare two versions of a single variable, typically by testing a subjects response to variant A against variant B, and determining which of the two variants is more effective.

Sol13] True, 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.

Sol14] In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables(also known as dependent and independent variables) linear regression has many practical uses.

Sol15] there are three real branches of statistics: data collection, descriptive statistics and inferential statistics.