

STATISTICS WORKSHEET-1

- 1) A
- 2) A
- 3) B
- 4) D
- 5) C
- 6) B
- 7) B
- 8) A
- 9) C
- 10) Probability distribution which is symmetric about mean i.e. showing data near mean is more frequent in occurrence than the data farther away.
- 11) Missing data can be handled using either `fillna()` function where the missing data is filled with an alternate value like zero or text if the column is categorical or you can also use `fillna()` methods such as `ffill(forward)`, `bfill (backward)`. Another way is to use `replace()` function to replace the null value with a new alternative value. You can also use `dropna()` function that allows you to drop rows/column with null values.

Mean/Median imputation can be used for numeric variable.
Mode can be used for categorical variable.

12) A/B testing is an experiment to compare and find out which version of the variable works best in the environment. In this process product will be first split into two parts i.e. A & B . Part A of the product will remain unchanged while B will be changed and then based on customer's feedback, the tester will decide which part worked better.

13) Mean imputation is not considered good practice because it ignores feature correlation. For eg: if there are two columns of age and fitness and if we use mean, then the fitness for someone with age 90 will appear more than it should actually be.

14) Linear regression is an analysis in which we find the relationship between one independent and one dependent variable using a straight line.

15) There are 2 main branches of statistics :

Descriptive statistics : It uses summarized data from collection of data to perform analysis. It can be further divided into two subcategory, which are

- i) central tendency (mean, median, mode) and
- ii) dispersion of data i.e. how wide data is spread (range, skew)

Inferential statistics : It uses observations on sample from population to analyze whether the population are actually different.

Eg: Hypothesis testing