

Statistics Worksheet -1

1. a)

2. a)

3. b)

4. d)

5. c)

6. b)

7. b)

8. a)

9. c)

10. Normal distribution: Normal distribution is the most widely known and used, of all distribution. It follows the following rules:

- 68% probability of the variable lying within 1 standard deviation of the mean.
- 95% probability of the variable lying within 2 standard deviations of the mean.
- 99% probability of the variable lying within 3 standard deviations of the mean.

11. To handle missing data, imputation technique that can be used is, if the field is numeric than we can fill it with the mean value but if there are outliers than we can fill the missing values with the median. In case of categorical field, we can fill the missing values with the mode(most occurring text value).

12. A/B testing is statistical hypothesis testing also known as statistical inference. It is used for making decisions that estimates population parameters based on sample statistics.

13. Yes, mean imputation of missing data is acceptable practice unless there are outliers.

14. Linear Regression in statistics is used to show a relationship between dependent and independent variable. There is one dependent variable and one or more independent variables. If the relationship between two variables is linear it can be summarized by a straight line. A straight line can be described by an equation:

$y=a+bx$ (a is called the intercept and b the slope of equation)

The slope is the amount by which y increases when x increases with 1 unit.

15. Branches of statistics:

Statistics is broadly categorized into two fields-

1. Descriptive Statistics
2. Inferential Statistics

a) Descriptive statistics- data is described using the mean, standard deviation, charts or probability distributions.

b) Inferential statistics- we estimate population parameters or the hypothesis testing is done to assess the assumptions made about the population parameters. We interpret the meaning of descriptive statistics by inferring them to the population.