Statistics worksheet-1

1. a
2. a
3. b
4. d
5. c
6. b
7. b
8. a
9. c

Subjective answer type Questions:

10. Answer:

The Normal Distribution is a type of continuous data distribution. The data will be normally distributed when mean=median=mode=0 and standard deviation is +1 or -1. In ideal condition it will be zero but in real case it does not happen. The Normal Distribution is easy to work with mathematically. In many practical cases, the methods developed using Normal theory work well even when the distribution is not normal.

11. Answer:

The missing data can be handled by using Imputation or removal of data. Removing the data from the data set every time is not feasible so Imputation techniques are recommended for handling missing data. The Imputation techniques which I recommend are;

- Mean/Median/Mode Imputation
- Regression Imputation
- Multiple Imputation

12. Answer:

A/B testing is a basic randomized control experiment. It helps to compare two versions of a variable to find out which performs better in a controlled environment. It allows to decide the best by looking at the analysis results obtained with two possible alternatives A and B. It includes application of statistical hypothesis testing.
Benefits of A/B testing:

- Ease of analysis
- Quick results
- Reduced risks

13. Answer:

Mean imputation is a popular solution and simple process to handle with missing data despite its drawbacks. Even though it is a simple and quick process, the mean imputation is not an acceptable practice to handle with missing data because;

- It decreases the variance of the data while increasing bias. Due to reduced variance, it will be less accurate
- It does not preserve the relationships among the variables.

14. Answer:

Linear Regression is the simplest and most extensively used statistical technique for predictive modelling analysis. It is used for continuous values. It is a way to explain the relationship between a dependent variable and one or more explanatory variables using a straight line. There are two types of linear regression they are simple and multiple. Linear regression has many practical uses.

15. Answer:

Statistics deals with collection, analysis and interpretation of numerical data, the various branches of statistics are;

- Descriptive Statistics:

  It deals with the presentation and collection of data. It helps in summarizing and organizing any data set characteristics. It also helps in the representation of data in both classification and diagrammatic way. The descriptive statistics is further categorized into central tendency and dispersion of data.

- Inferential Statistics:

  It helps in finding the conclusion regarding the population after analysis on the sample drawn from it. It includes t-test, chi-square test, correlational test and ANOVA test.