Statistics worksheet

Q.no 1

Ans (a)

Q.no 2

Ans (a)

Q.no 3

Ans (b)

Q.no 4

Ans (d)

Q.no 5

Ans (c)

Q.no 6

Ans (b)

Q.no 7

Ans (b)

Q.no 8

Ans (a)

Q.no 9

Ans (c)

Q.no 10

Ans: The normal distribution is a continuous probability distribution function also known as Gaussian distribution which is symmetric about its mean and has a bell-shaped curve.

Q.no 11

Ans: Missing data can occur when no information is provided for one or more items or for a whole unit. Missing data is a very big problem in real life scenarios. Missing data also refer to as NA(Not available) values in pandas.

There are various imputation techniques used to handle missing data for example

1. Mean imputation
2. Simple or multiple imputation
3. Regression imputation
4. Interpolation and extrapolation
5. Hot deck imputation
6. Cold deck imputation
7. Stochastic imputaion

Q.no 12

Ans: A/B testing is a method of experimentation where two variants of a page are shown to users at random, and statistical analysis is used to determine which version perfoms better for a given conversion goal. Version A is the control, while version b is the variation. It is also known as bucket testing or split-testing.

Q.no 13

Ans: It is acceptable when the missing value proportion is not large enough. But when the missing values are large enough and you impute them with the mean, the standard errors will be lesser than what they actually would have been.

Q.no 14

Ans: Linear regression is a supervised learning algorithm that uses the relationship between data points to form a best fit line called a regression line.

If there is strong relationship in the data then the regression line is used to calculate the value of a variable, the target variable, based on one or more features.

Q.no 15

Ans: There are two kinds of Statistics, which are descriptive Statistics and inferential Statistics. In descriptive Statistics, the Data or Collection Data are described in a summarized way, whereas in inferential Statistics, we make use of it in order to

explain the descriptive kind. Both of them are used on a large scale. Also, there is another kind of Statistics where descriptive transitions into inferential Statistics.