Statistics Worksheet-1

Ans 1: a

Ans 2: a

Ans 3: b

Ans 4: d

Ans 5: c

Ans 6: b

Ans 7: b

Ans 8: a

Ans 9: c

Ans 10: Normal Distribution: Also known as 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.

Ans 11: A common technique is to use the mean or median of the non-missing observations. This can be useful in cases where the number of missing observations are low.

However for large number of missing values, using mean or median can result in loss of variation in data and it is better to use imputations.

Following are some imputation techniques which can be used for handling missing data:

1- Mean or Median Imputation
2- Multivariate imputation by chained equations (MICE).
3- Random Forest

Ans 12: A/B testing- Also known as split testing, refers to a randomized experimentation process wherein two or more version of a variable (web page, page element, etc) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

It eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions.

Ans 13: Mean imputation (also called mean substitution) really oughtto be a last resort.

Following are some reasons:

1- Mean imputation does not preserve the relationships among variables.
2- It leads to an underestimate of standard errors.

Ans 14: Linear regression- It is a linear approach for modelling the relationship between a scaler response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression.

And for more than one, this process is called multiple linear regression.

Ans 15: There are two branches of statistics:

1- Descriptive
2- Infrential

Descriptive : procedures used to summarize, organize, and make sense of a set of scores or observations.

Inferntial: procedures used that allow researchers to infer or generalize, observations made with samples to the large population from which they were selected.