

STATISTICS WORKSHEET

Q1.B

Q2.A

Q3. B

Q4. D

Q5. C

Q6. A

Q7.D

Q8.A

Q9.C

Q10.Normal Distribution : Normal distribution is also known as the 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,where it appears as bell curve in representation.

Q11. The deletion methods only work for certain datasets where participants have missing fields. There are several deleting methods – two common ones include Listwise Deletion and Pairwise Deletion. It means deleting any participants or data entries with missing values. This method is particularly advantageous to samples where there is a large volume of data because values can be deleted without significantly distorting readings. Alternatively, data scientists can fill out the missing values by contacting the participants in question. The problem with this method is that it may not be practical for large datasets. Furthermore, some corporations obtain their information from third-party sources, which only makes it unlikely that organisations can fill out the gaps manually. Pairwise deletion is the process of eliminating information when a particular data point, vital for testing, is missing. Pairwise deletion saves more data compared to likewise deletion because the former only deletes entries where variables were necessary for testing, while the latter deletes entire entries if any data is missing, regardless of its importance.

use two data imputation techniques to handle missing data: Average imputation and common-point imputation. Average imputation uses the average value of the responses from other data entries to fill out missing values. However, a word of caution when using this method – it can artificially reduce the variability of the dataset. Common-point imputation, on the other hand, is when the data scientists utilise the middle point or the most commonly chosen value. For example, on a five-point scale, the substitute value will be 3. Something to keep in mind when utilising this method is the three types of middle values: mean, median and mode, which is valid for numerical data (it should be noted that for non-numerical data only the median and mean are relevant

Missing data is a sad fact of life when it comes to data analytics. We cannot avoid situations like these entirely because there are several remedial steps data scientists

need to take to make sure it doesn't adversely affect the analytics process. While these methods are helpful, they are not fool proof because they are contentious, meaning, their effectiveness depends heavily on circumstances. The best option available is to work with powerful, processing tools that can make the data capturing and analysis process significantly easier. It is the best way to handle missing data.

(The simplest imputation method is replacing missing values with the mean or median values of the dataset at large, or some similar summary statistic. This has the advantage of being the simplest possible approach, and one that doesn't introduce any undue bias into the dataset.

Q12. A/B testing is basically statistical hypothesis testing, or, in other words, statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

Q13 The process of replacing null values in a data collection with the data's mean is known as mean. imputation. It leads to an underestimate of the standard deviation and it preserves the mean of the preserved data.

Which may not a good practice

Q14. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable which we want to predict is called the dependent variable. The variable which are using to predict the other variable's value is called the independent variable

Q15. Statistics:

Statistics is a study of presentation, analysis, collection, interpretation and organization of data

There are two main branches of statistics

- Inferential Statistic.
- Descriptive Statistic.

Inferential Statistics:

Inferential statistics used to make inference and describe about the population. These stats are more useful when its not easy or possible to examine each member of the population.

Descriptive Statistics:

Descriptive statistics are use to get a brief summary of data. You can have the summary of data in numerical or graphical form.