**Statistics Worksheet-4 Solution**

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. The **Normal distribution** is a probability function that describes how the values of a variable are distributed. It is a symmetric distribution where most of the observations cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions.

It 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.

1. There are many ways one could handle missing data - and there is no single answer to this problem. Some specific ways of handling this with imputation are:
2. Imputation with mean: Missing data is replaced by mean of the column
3. Imputation with median: Missing data is replaced by mean of the column
4. Imputation with Mode: Missing data is replaced with mode of the column
5. Imputation using the forward fill and backward fill to the null positions.
6. A/B testing (also known as split testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. AB testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.
7. The mean imputation is a bad practice actually as it preserves the mean of the observed data which collectively leads to an underestimate of the standard deviation. Distorts relationships between variables by “pulling” estimates of the correlation toward zero.
8. Linear regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.  The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = m\*x + c

where y = estimated dependent variable score,

c = constant,

b = regression coefficient,

and x = score on the independent variable.

1. Statistics are a form of mathematical analysis that uses quantitative models to give a set of experimental data or studies of real life. Statistics examine the methodology for collecting, reviewing, analysing, and making data conclusions. Some statistical measures include the following:

* Mean
* Regression analysis
* Skewness
* Kurtosis
* Variance
* Analysis of variance
* Descriptive Statistics

**Branches of Statistics**

**Descriptive Statistics**

Descriptive statistics is the first part of statistics that deals with the collection of data. Descriptive statistics are used in use to do various kinds of analysis on different studies.

**Mean**

Mean is a conventional method used to describe the central tendency. Typically, to calculate the average of values, count all values, and then divide them with the number of available values.

**Median**

It is the result that is in the middle of a set of values. An easy way to calculate the median is to edit the results in numerical journals and locate the result that is in the center of the distributed sample.

**Mode**

The mode is the frequently occurring value in the given data set.

**Inferential Statistics**

The inference statistics are techniques that enable statisticians to use the information collected from the sample to conclude, bring decisions, or predict a defined population.

Inference statistics often speak in terms of probability by using descriptive statistics.

Most predictions of the future and generalization on a population study of a smaller specimen are in the scope of the inference statistics. Besides, most of the social sciences experiments deal with the study of a small sample population that helps determine the behaviour of the community.

**Different types of inferential statistics include:**

* Regression analysis
* Analysis of variance (ANOVA)
* Analysis of covariance (ANCOVA)
* Statistical significance (t-test)
* Correlation analysis