**Statistics Worksheet**

1. A)
2. A)
3. B)
4. D)
5. C)
6. B)
7. B)
8. A)
9. C)
10. The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side. The area under the normal distribution curve represents probability and the total area under the curve sums to one.
11. Missing data handling can be very subjective. It depends upon the data we are dealing with and the type of column that we have.

The first step to deal with the missing data is to determine how much percentage of data is missing in a column. If the percentage is very high, say more than 70%, then the best option is to drop the entire column. Again, it is very subjective. If the column is of very high importance, it is not recommended to drop it. Other options of imputing the missing data are as follows:

1. Mean Imputation: This is done in case of continuous data. The missing values are replaced with mean of the column.
2. Substitution: Missing values are replaced with a new category, say ‘None’. This is used in case of categorical variables.
3. Mode Imputation: This is done in case of categorical data. The missing values are replaced with mode of the column.
4. Regression Imputation: The predicted value obtained by regressing the missing variable on other variables.
5. Multiple Imputation: Combination of more than 1 imputation techniques where a random variable is also introduced.
6. 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.
7. Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.

A linear regression line has an equation of the form *Y = a + bX*, where *X* is the explanatory variable and *Y* is the dependent variable. The slope of the line is *b*, and *a* is the intercept (the value of *y* when *x* = 0).

1. The two main branches of Statistics are:
2. Descriptive Statistics: Descriptive statistics deals with the collection of data, its presentation in various forms, such as tables, graphs and diagrams and finding averages and other measures which would describe the data.
3. Inferential Statistics: Inferential statistics deals with techniques used for the analysis of data, making estimates and drawing conclusions from limited information obtained through sampling and testing the reliability of the estimates.  
     
   But there are other forms as well which are used in statistical analysis like:
4. Predictive Analytics: If you want to make predictions about future events, predictive analysis is what you need. This analysis is based on current and historical facts.

Predictive analytics uses statistical algorithms and machine learning techniques to define the likelihood of future results, behavior, and trends based on both new and historical data.

1. Prescriptive Analytics: Prescriptive analytics aims to find the optimal recommendations for a decision making process.  It is all about providing advice.
2. Causal Analytics: When you would like to understand and identify the reasons why things are as they are, causal analysis comes to help.