**Statistics Worksheet-1 Answers**

Q.1 (a)

Q.2 (a)

Q.3 (b)

Q.4 (d)

Q.5 (c)

Q.6 (b)

Q.7 (b)

Q.8 (a)

Q.9 (c)

Q.10 Normal distribution, 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. In graph form, normal distribution will appear as a bell curve.

Q.11 **Common Methods**

1. Mean or Median **Imputation**. When **data** is **missing** at random, **we** can use list-wise or pair-wise deletion of the **missing** observations. However, there can be multiple reasons why this may not be the most feasible option:

* There may not be enough observations with non-missing data to produce a reliable analysis
* In predictive analytics, missing data can prevent the predictions for those observations which have missing data
* External factors may require specific observations to be part of the analysis

In such cases, we impute values for missing data. A common technique is to use the mean or median of the non-missing observations.

1. Multivariate **Imputation** by Chained Equations (MICE) MICE assumes that the **missing data** are **Missing** at Random (MAR). It imputes data on a variable-by-variable basis by specifying an imputation model per variable.
2. Random Forest.

Q.12 A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

For instance, let’s say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

Q.13 Although imputing missing values by using the mean is a popular imputation technique, there are serious problems with mean imputation.

There are three problems with using mean-imputed variables in statistical analyses:

* Mean imputation reduces the variance of the imputed variables.
* Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval.
* Mean imputation does not preserve relationships between variables such as correlations.

Q.14 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. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

Q.15 The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important.