Answers

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. **Normal Distribution**

Normal distribution is also called as Gaussian distribution which is a continuous distribution that is symmetrical around its mean and most of the observations cluster around the central peak. It has a bell-shaped curve when the observations are plotted. It’s mean, median and mode are equal.

1. **Handling Missing Data and imputation techniques.**

Missing data can be present in a dataset due to various reasons. It can be handled with any of the following methods.

1. Drop the missing data
2. Replace the missing data
3. Consider the missing data as it is.

There are several replace methods or imputation methods in python to replace the missing data. Few of them are simple imputation method, label encoding, get dummy method or the pandas replace method. Both simple imputation method and pandas inbuilt replace method is more useful and simpler to use. It can be used for string values, categorical data, integers and floats.

1. **A/B Testing.**

It is one of the most randomized controlled experiments which helps to identify the best of two versions of same variable in a controlled environment. It allows us to choose the best by looking at the analytics results obtained with two possible alternatives A and B.

1. **Is mean imputation of missing data acceptable practice?**

Mean imputation is one of the methods to handle missing data in dataset. It is the process of computing the mean of observed values and imputing missing values with the computed mean value. Even though it is one of popular method used among data users, there are drawbacks for this method.

1. Mean imputation reduces variance
2. Mean imputation shrinks standard errors and confidence intervals
3. Mean imputation distorts relationships between variables
4. **Linear regression in statistics?**

It is a way to explain the relationship between a dependent variable (target) and one or more explanatory variables(predictors) using a straight line. There are two types of linear regression - Simple and Multiple. Linear regression is only dealing with continuous variables.

1. **Various branches of statistics**

There are three branches of statistics: *Data Collection, Descriptive Statistics and Inferential Statistics.*

1. **Data Collection**

It consists of the collection of data for analysis. Data can be collected from primary source or secondary source. Data collected can be either including the whole population or with the sample distribution. Population distribution is made up of all the classes or values of variables.

Sample distribution is the finite subset of a population, selected from the population with the objective of investigating its properties. It is a representative part of the population.

1. **Descriptive Statistics**

It is the process of presenting the collected data. Data aspects can be presented either visually or numerically. Descriptive statistics is the study of distributions of variables. It can be univariate, bivariate or multivariate analysis.

1. **Inferential Statistics**

It is concerned with drawing inferences and conclusions from the findings of a study.

There are two areas of statistical inferences: Statistical estimation and testing of Hypothesis.