

1. A
2. A
3. B
4. D
5. C
6. B
7. B
8. A
9. C
10. **A normal distribution** sometimes called the bell curve, is a distribution that occurs naturally in many situations. It is the most common distribution function for independent, randomly generated variables. Its familiar bell-shaped curve is ubiquitous in statistical reports, from survey analysis and quality control to resource allocation.

**11. Following are the seven ways to handle missing data:**

**Listwise Deletion:** Delete all data from any participant with missing values. If your sample is large enough, then you likely can drop data without substantial loss of statistical power. Be sure that the values are missing at random and that you are not inadvertently removing a class of participants.

**Recover the Values:** You can sometimes contact the participants and ask them to fill out the missing values. For in-person studies, we've found having an additional check for missing values before the participant leaves helps.

## **Imputation**

Imputation is replacing missing values with substitute values. The following methods use some form of imputation.

**Educated Guessing:** It sounds arbitrary and isn't your preferred course of action, but you can often infer a missing value.

**Average Imputation:** Use the average value of the responses from the other participants to fill in the missing value. If the average of the 30 responses on the question is a 4.1, use a 4.1 as the imputed value. This choice is not always recommended because it can artificially reduce the variability of your data but in some cases makes sense.

**Common-Point Imputation:** For a rating scale, using the middle point or most commonly chosen value. For example, on a five-point scale,

substitute a 3, the midpoint, or a 4, the most common value (in many cases). This is a bit more structured than guessing, but it's still among the more risky options. Use caution unless you have good reason and data to support using the substitute value.

**Regression Substitution:** You can use multiple-regression analysis to estimate a missing value. We use this technique to deal with missing SUS scores. Regression substitution predicts the missing value from the other values. In the case of missing SUS data, we had enough data to create stable regression equations and predict the missing values automatically in the calculator.

**Multiple Imputation:** The most sophisticated and, currently, most popular approach is to take the regression idea further and take advantage of correlations between responses. In multiple imputation [pdf], software creates plausible values based on the correlations for the missing data and then averages the simulated datasets by incorporating random errors in your predictions.

12. **A/B testing is also called split testing or bucket testing.** It compares the performance of two versions of content to see which one appeals more to visitors/viewers. It tests a control (A) version against a variant (B) version to measure which one is most successful based on your key metrics. Regardless of the focus, A/B testing helps you determine how to provide the best customer experience (CX).
13. **The process of replacing null values in a data collection with the data's mean is known as mean imputation.**

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Also mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. **Linear regression in statistics: The process of finding a straight line (as by least squares) that best approximates a set of points on a graph.** It is a regression model that estimates the relationship between

one independent variable and one dependent variable using a straight line. Both variables should be quantitative. It is basically used to predict the value of dependent variable by considering the value of independent variable.

15. **Statistics is broadly divided into 2 types:** Descriptive statistics and Inferential statistics.

- i. **Descriptive Statistics:** Descriptive statistics is considered as the first part of statistical analysis which deals with collection and presentation of data. Scientifically, descriptive statistics can be defined as brief explanatory coefficients that are used by statisticians to summarize a given data set. Generally, a data set can either represent a sample of a population or the entire populations. Descriptive statistics can be categorized into

**Measures of central tendency:** Measures of central tendency specifically help the statisticians to estimate the center of values distribution. These measures of tendency are:

- **Mean:** This is the conventional method used in describing central tendency. Usually, to compute an average of values, you add up all the values and then divide them with the number of values available.
- **Median:** This is the score found at the middle of a set of values. A simple way to calculate a median is to arrange the scores in numerical orders and then locate the score which is at the center of the arranged sample.
- **Mode:** This is the frequently occurring value in a given set of scores.
- **Measures of variability:** The measure of variability help statisticians to analyze the distribution spread out of a given set of data. Some of the examples of measures of variability include quartiles, range, variance and standard deviation.

## ii. **Inferential Statistics**

Inferential statistics are techniques that enable statisticians to use the gathered information from a sample to make inferences, decisions or predictions about a given population. Inferential statistics often talks in probability terms by using descriptive statistics. These techniques are majorly used by statisticians to analyze data, make estimates and draw conclusions from the limited information which is obtained by sampling and testing how reliable the estimates are.

The different types of calculation of inferential statistics include:

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