

1.

The Central Limit Theorem (CLT) states that the distribution of a sample mean that approximates the normal distribution, as the sample size becomes larger, assuming that all the samples are similar, and no matter what the shape of the population distribution is.

It all has to do with the distribution of our population. This theorem allows you to simplify problems in statistics by allowing you to work with a distribution that is approximately normal.

2.

To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. This is called a sampling method. There are two primary types of sampling methods that you can use in your research:

- Probability sampling involves random selection, allowing you to make strong statistical inferences about the whole group.
- Non-probability sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect data.

3.

A type I error (false-positive) occurs if an investigator rejects a null hypothesis that is actually true in the population; a type II error (false-negative) occurs if the investigator fails to reject a null hypothesis that is actually false in the population

4.

A normal distribution is a type of continuous probability distribution in which most data points cluster toward the middle of the range, while the rest taper off symmetrically toward either extreme.

5.

Covariance is an indicator of the extent to which 2 random variables are dependent on each other. A higher number denotes higher dependency. Correlation is a statistical measure that indicates how strongly two variables are related.

6.

Uni variate statistics summarize only one variable at a time. Bivariate statistics compare two variables. Multivariate statistics compare more than two variables

7.

The technique used to determine how independent variable values will impact a particular dependent variable under a given set of assumptions is defined as sensitive analysis

The sensitivity is calculated by dividing the percentage change in output by the percentage change in input.

8.

The major purpose of hypothesis testing is to choose between two competing hypotheses about the value of a population parameter.

Null Hypothesis  $H_0$  : Statement being tested; Claim about  $\mu$  or historical value of  $\mu$  Given Null Hypothesis:  $\mu = k$   $k$  is a value of the mean given  $\mu$  is the population mean discussed throughout the worksheet

Alternative Hypothesis  $H_1$  : Statement you will adopt in the situation in which evidence(data) is strong so  $H_0$  is rejected

9.

Quantitative data are measures of values or counts and are expressed as numbers. Quantitative data are data about numeric variables (e.g. how many; how much; or how often). Qualitative data are measures of 'types' and may be represented by a name, symbol, or a number code.

10.

Range: the difference between the highest and lowest values.

The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

11.

A bell curve is a type of graph that is used to visualize the distribution of a set of chosen values across a specified group that tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side.

12.

IQR:

The interquartile range (IQR) tells you the range of the middle half of your dataset. You can use the IQR to create “fences” around your data and then define outliers as any values that fall outside those fences.

This method is helpful if you have a few values on the extreme ends of your dataset, but you aren't sure whether any of them might count as outliers.

Inter quartile range method

1. Sort your data from low to high
2. Identify the first quartile (Q1), the median, and the third quartile (Q3).
3. Calculate your IQR =  $Q3 - Q1$
4. Calculate your upper fence =  $Q3 + (1.5 * IQR)$
5. Calculate your lower fence =  $Q1 - (1.5 * IQR)$
6. Use your fences to highlight any outliers, all values that fall outside your fences.

13.

The p value is a number, calculated from a statistical test, that describes how likely you are to have found a particular set of observations if the null hypothesis were true. P values are used in hypothesis testing to help decide whether to reject the null hypothesis.

14.

Binomial probability refers to the probability of exactly x successes on n repeated trials in an experiment which has two possible outcomes (commonly called a binomial experiment). If the probability of success on an individual trial is p, then the binomial probability is  $nCx \cdot p^x \cdot (1-p)^{n-x}$ .

15.

Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups.

Application

A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables