Q-1 How to the Statistics significant of an insight assessed?

Hypothesis testing is guided by statistical analysis. Statistical significance is calculated using a p-value, which tells you the probability of your result being observed, given that a certain statement (the null hypothesis) is true.[1] If this p-value is less than the significance level set (usually 0.05), the experimenter can assume that the null hypothesis is false and accept the alternative hypothesis. Using a simple t-test, you can calculate a p-value and determine significance between two different groups of a dataset.

1. Define your hypotheses.
2. Set the significance level to determine how unusual your data must be before it can be considered significant.
3. Decide to use a one-tailed or two-tailed test.

Q-2 What is mean?

**Mean** is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

**Mean** = (Sum of all the observations/Total number of observations)

**Example:**

What is the mean of 2, 4, 6, 8 and 10?

**Solution:**

First, add all the numbers.

2 + 4 + 6 + 8 + 10 = 30

Now divide by 5 (total number of observations).

Mean = 30/5 = 6

Mean Symbol (X Bar)

The symbol of mean is usually given by the symbol ‘x̄’. The bar above the letter x, represents the mean of x number of values.

X̄ = (Sum of values ÷ Number of values)

X̄ = (x1 + x2 + x3 +….+xn)/n

Mean Formula:- x̄=∑ x/n

Q-3 What is the meaning of Standard Deviation?

**Standard deviation** is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. The standard deviation is calculated as the square root of variance by determining each data point's deviation relative to the mean.

If the data points are further from the mean, there is a higher deviation within the data set; thus, the more spread out the data, the higher the standard deviation.

**KEY TAKEAWAYS:**

1.Standard deviation measures the dispersion of a dataset relative to its mean.

2.It is calculated as the square root of the variance.

3.Standard deviation, in finance, is often used as a measure of a relative riskiness of an asset.

4.A volatile stock has a high standard deviation, while the deviation of a stable blue-chip stock is usually rather low.

Q-4 What is correlation?

**Correlation** refers to the statistical relationship between two entities. In other words, it's how two variables move in relation to one another. Correlation can be used for various data sets, as well. In some cases, you might have predicted how things will correlate, while in others, the relationship will be a surprise to you. It's important to understand that correlation does not mean the relationship is causal.

**Types of correlation coefficients**

While correlation studies how two entities relate to one another, a correlation coefficient measures the strength of the relationship between the two variables. In statistics, there are three types of correlation coefficients. They are as follows:

**1.Pearson correlation**: The Pearson correlation is the most commonly used measurement for a linear relationship between two variables. The stronger the correlation between these two datasets, the closer it'll be to +1 or -1.

**2.Spearman correlation**: This type of correlation is used to determine the monotonic relationship or association between two datasets. Unlike the Pearson correlation coefficient, it's based on the ranked values for each dataset and uses skewed or ordinal variables rather than normally distributed ones.

**3**.**Kendall correlation**: This type of correlation measures the strength of dependence between two datasets.

Q-4What is the meaning of covariance?

**Covariance** is a measure of the relationship between two random variables and to what extent, they change together. Or we can say, in other words, it defines the changes between the two variables, such that change in one variable is equal to change in another variable. This is the property of a function of maintaining its form when the variables are linearly transformed. Covariance is measured in units, which are calculated by multiplying the units of the two variables.

**Types of Covariance**

Covariance can have both positive and negative values. Based on this, it has two types:

1.Positive Covariance

2.Negative Covariance

**Positive Covariance** -

If the covariance for any two variables is positive, that means, both the variables move in the same direction. Here, the variables show similar behaviour. That means, if the values (greater or lesser) of one variable corresponds to the values of another variable, then they are said to be in positive covariance.

**Negative Covariance -**

If the covariance for any two variables is negative, that means, both the variables move in the opposite direction. It is the opposite case of positive covariance, where greater values of one variable correspond to lesser values of another variable and vice-versa.

Q-5 Where is Inferential statistics used?

**Inferential statistics** are generally used in two ways:

* to set parameters about a group and then create hypotheses about how data will perform when scaled.
* Inferential statistics are among the most useful tools for making educated predictions about how a set of data will scale when applied to a larger population of subjects. These statistics help set a benchmark for hypothesis testing, as well as a general idea of where specific parameters will land when scaled to a larger data set, such as the larger set’s mean.
* This process can determine a population’s z-score (where a subject will land on a bell curve) and set data up for further testing.

Q-6 What is One sample t-test?

A **one sample t-test** is used to test whether or not the mean of a [population](https://www.statology.org/population-vs-sample/) is equal to some value.

**This tutorial explains the following**:

* The motivation for performing a one sample t-test.
* The formula to perform a one sample t-test.
* The assumptions that should be met to perform a one sample t-test.
* An example of how to perform a one sample t-test.

**A one-sample t-test always uses the following null hypothesis:**

* **H0:**μ = μ0 (population mean is equal to some hypothesized value μ0)

The alternative hypothesis can be either two-tailed, left-tailed, or right-tailed:

* **H1 (two-tailed):**μ ≠ μ0 (population mean is not equal to some hypothesized value μ0)
* **H1 (left-tailed):**μ < μ0 (population mean is less than some hypothesized value μ0)
* **H1 (right-tailed):**μ > μ0 (population mean is greater than some hypothesized value μ0).

Q-7 What is the relationship between Standard deviation and standard variance?

Variance is equal to the average squared deviations from the mean, while standard deviation is the number’s square root. Also, the standard deviation is a square root of variance. Both measures exhibit variability in distribution, but their units vary: Standard deviation is expressed in the same units as the original values, whereas the variance is expressed in squared units.

Variance and Standard Deviation are the two important measurements in statistics.

Variance is a measure of how data points vary from the mean, whereas standard deviation is the measure of the distribution of statistical data.

The basic difference between both is standard deviation is represented in the same units as the mean of data, while the variance is represented in squared units.

Q-8 What is One way ANOVA test?

A **one-way ANOVA** (“analysis of variance”) compares the means of three or more independent groups to determine if there is a statistically significant difference between the corresponding population means.

This tutorial explains the following:

* The motivation for performing a one-way ANOVA.
* The assumptions that should be met to perform a one-way ANOVA.
* The process to perform a one-way ANOVA.

One-Way ANOVA: The Process-

A one-way ANOVA uses the following null and alternative hypotheses:

1. H0 (null hypothesis): μ1 = μ2 = μ3 = … = μk (all the population means are equal)
2. H1 (alternative hypothesis): at least one population mean is different from the rest.