

**101. How to calculate p-value using a manual method?****Ans.**

i. If sample size  $n < 30$ , sample mean ( $\bar{x}$ ), sample standard deviation, number of samples is given we calculate test statistics

$$\text{test statistics (z)} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

ii. Then we check value from z table.

iii. If it is a two tailed test we multiply result with 2. Other for one tailed test result will be same.

**102. What do we mean by – making a decision based on comparing p-value with significance level?  
What is the goal of A/B testing?****Ans.**

To determine the result of hypothesis testing the p-value is compared with the significance level. If p-value  $\leq$  significance level, we fail to reject null hypothesis. If p-value  $>$  significance level then we fail to reject the null hypothesis.

A/B testing is like hypothesis testing where we make decisions based on sample data for population parameter.

**103. What is the difference between a box plot and a histogram.****Ans.**

Histogram is a special kind of bar graph which displays a range of values divides in bins and frequency.

Boxplot displays 5 Number summary which includes min, max, 1st quartile, median (2nd quartile), 3rd quartile. It is helpful to identify outliers.

## Comparing Shapes of Histograms to the Corresponding Boxplot

104. A jar has 1000 coins, of which 999 are fair and 1 is double headed. Pick a coin at random, and toss it 10 times. Given that you see 10 heads, what is the probability that the next toss of that coin is also a head?

**Ans.**

The probability of tossing the jar of coin =  $1/1000 * 1 + 999/1000 * 0.5$

$$= 1001/2000 = 0.5$$

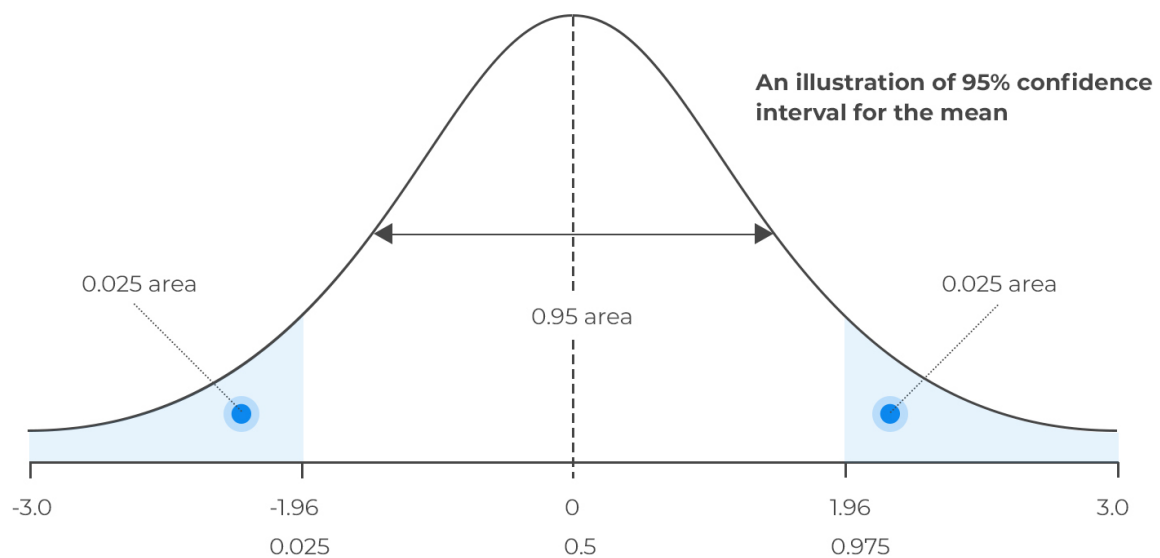
105. What is a confidence interval and how do you interpret it?

**Ans.**

Confidence interval is a range of values where population parameter lies. This means how certain or uncertain the estimated difference between variables is.



### 95% Interval



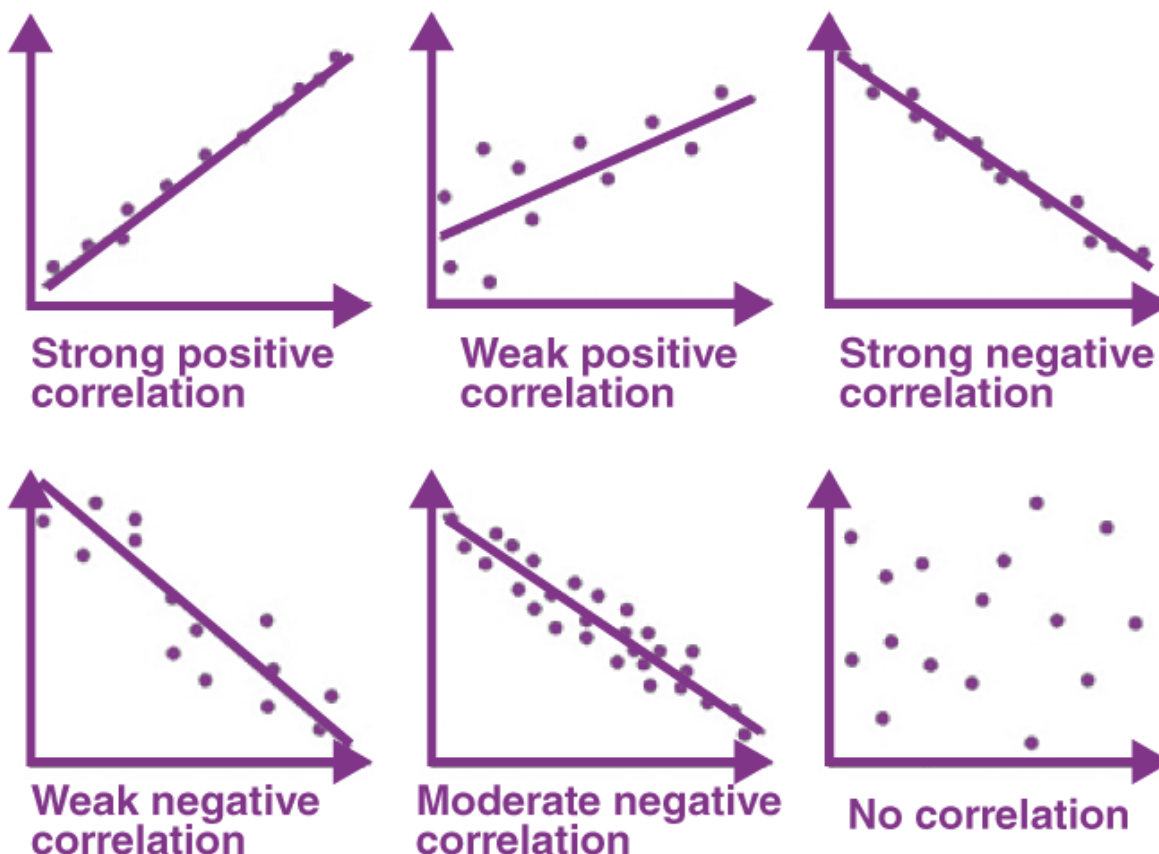
106. How do you stay up-to-date with the new and upcoming concepts in statistics?

**Ans.**

**107. What is correlation?****Ans.**

The correlation describes the size and direction of relationship between two variables. It defines the linear relationship between two variables.

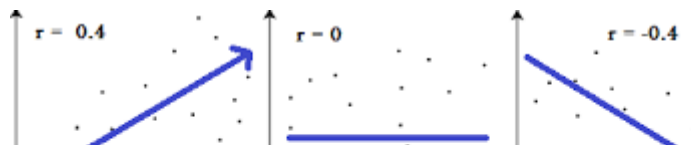
Drawback of correlation is that it has no limit to define the relation between variables.

**108. What types of variables are used for Pearson's correlation coefficient?****Ans.**

Pearson correlation coefficient was introduced to overcome the drawbacks of correlation.

Pearson correlation coefficient ranges between -1 to 1, which measures the strength and direction of relationship between two variables. The more the value towards +1 the more positively correlated the variables are, and the more the value towards -1 the more negatively correlated the variables are.

We use two or more continuous variables.



**109.** In an observation, there is a high correlation between the time a person sleeps and the amount of productive work he does. What can be inferred from this?

**Ans.**

Correlation doesn't imply here. It is only used to measure the linear relationship between sleep and amount of productive work.

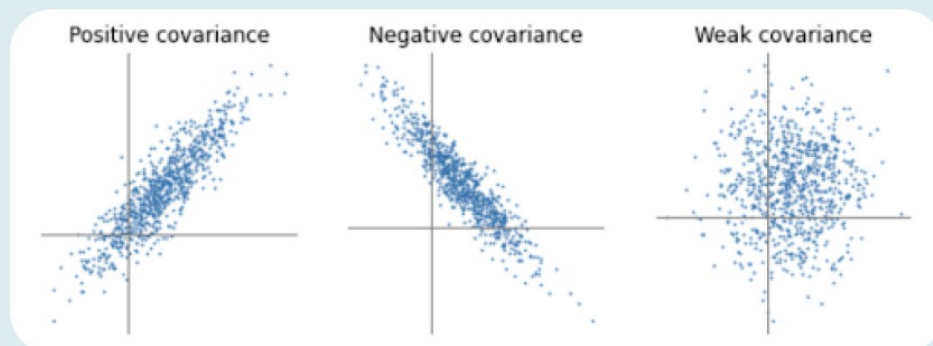
**110.** What is the meaning of covariance?

**Ans.**

Covariance is a measure of relationship between two random variables and to what extent they vary together.

If increase in one variable causes increase in another variable, is called positive covariance.

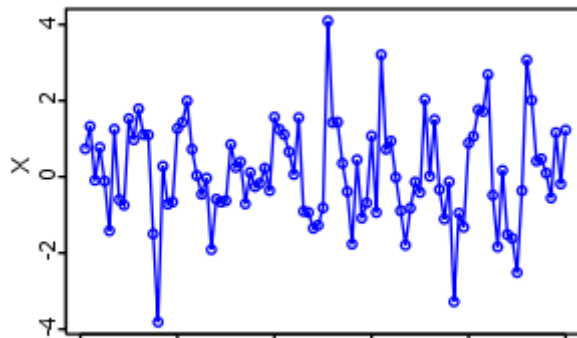
If decrease in one variable causes decrease in another variable, is called negative covariance.



**111.** What does autocorrelation mean?

**Ans.**

Autocorrelation is the relation between same variable in two successive time intervals. It measures how the lagged version of the variable is related to original version of it in time series.



**112. How will you determine the test for the continuous data?**

**Ans.**

To test continuous data we commonly use t-test statistical analysis. It is an appropriate test to compare between two continuous data which are both normally distributed. The commonly used forms of t-test are the test of hypothesis, the single sample sample test, paired test and two sample test and unpaired test.

**113. What can be the reason for non normality of the data?**

**Ans**

Some measurements naturally follow non normal distribution where too many extreme values in a dataset will result skewed distribution. We also get non-normality because of measurement errors, data-entry errors, and outliers, we can remove them from data for valid reasons.

**114. Why is there no such thing like 3 samples t- test?? why t-test failed with 3 samples?**

**Ans.**

T-test should be used to measure differences between more than two groups, because the error structure for t-test will under estimate the actual error when many groups are compared. Everytime you conduct a t-test there is a chance of type-1 error which is usually 5%. For two sample t-test in some data you will increase the chance of making mistake 10%.

In [ ]:

1

**Thanks**

**Github:** <https://github.com/saisubhasish> (<https://github.com/saisubhasish>)

In [ ]:

1

