**VARIANCE AND CO-VARIANCE**

Today, we are going to learn about various important concepts on COVARIANCE AND VARIANCE.

We often, got confused:

1. What is the difference between covariance and variance?
2. When, we already studied variance then why we need covariance?
3. What is the significance of covariance and variance?
4. What is formula and how to calculate both variance and covariance?
5. Is there any relation between covariance and variance?

Today all your doubts will be cleared regarding all above mentioned dilemma.

Before directly jumping onto COVARIANCE, it is extremely important to understand about the VARIANCE concept.

**WHAT IS VARIANCE?**

Variance is a topic comes under Dispersion concept, meaning it is a measure of how far a set of numbers is spread out from their average value.

More specifically, variance measures how far each number in the set is from the mean (average), and thus from every other number in the set. Variance is often depicted by this symbol: σ2.

**USE CASE:**

It is used by both **analysts and traders** to determine volatility and market security.

The square root of the variance is the standard deviation (SD or σ).

Before, going to formula and its application, let us understand the concept of variance through this wonderful diagram.

**Understanding the concept through geometrical representation**.

**Here, in the below fig., we have divided concept for our understanding into 4 parts.**

First part (1), depicts the sample on x-axis and we have represented the balls as occurrence of these numbers (that is, Frequency)

We have sample or distribution as (2, 4,4,4,5,7,9) and created a frequency distribution.

Second part (2), we calculate the centroid of the distribution given its mean.

Mean = 1+2+3+4+5+6+7+8+9 / 9 = 5 (marked by the arrow under 5 in the fig.)

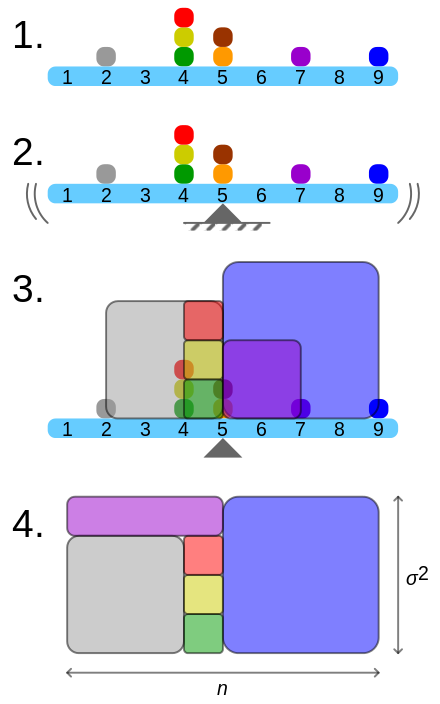
Third part (3), very important as here we are calculating the how far is each number from the mean value.

The dispersion of each number from mean is showed by different, for example,

* Distance of 9 from 5 is shown by violet colour
* Similarly, dispersion of 7 from 5 is shown by purple colour
* Dispersion of 4 from 5 is depicted by 3 colours (frequency) as orange, yellow and green
* Spread of 2 from 5 is shown by grey colour (complete square).

Fourth part (4), where we derived variance just by converting squares into rectangles and this will be represented as one of the sides as n (whole distribution is covered) and other as variance.

***Why variance, I know all are wondering, because summation of all deviations (3) for every n value is nothing but variance***. (will shortly, explain the formula too)



|  |
| --- |
| 2. - mean of distribution, arrow under 5 3. – distance from mean foe each number, by different colours  1. - frequency distribution 4. – converted square into rectangle, and one side is n and next is variance.  *GEOMETRICAL REPRESENTATION OF VARIANCE* |

**Before this, let me clear one thing, because you all have assumed that variance has to be calculated for all the data (population), because of the representation of (4) part.**

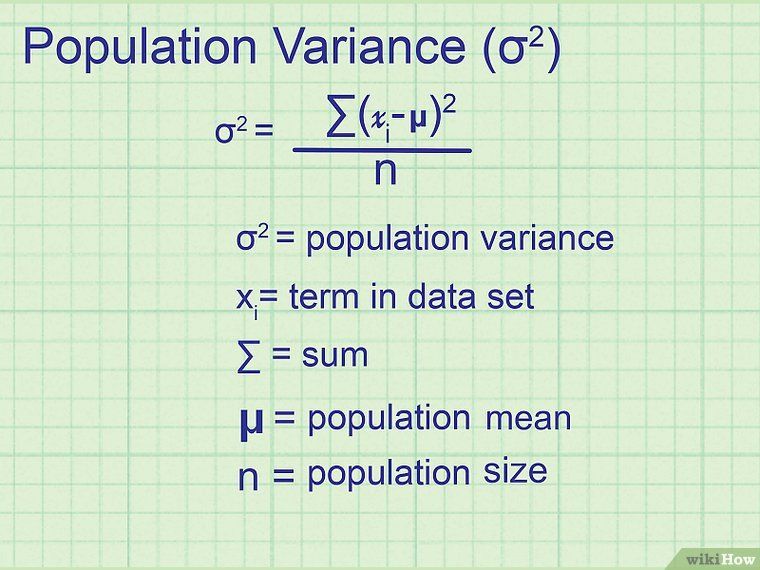
**Yes, the assumption is wrong, let us see how!**

**Now, variance can be calculated for population data and for sample data, but the formula sightly differs in each case.**

**Population data:**

When variance is calculated from observations, those observations are typically measured from a real-world system. If all possible observations of the system are present then the calculated variance is called the population variance.

**FORMULA FOR POPULATION VARIANCE:**



**AGAIN, A DILEMMA, IF WE HAVE VARIANCE POPULATION FORMULA, THEN WHY THIS FORMULA CAN NOT BE USED ALL THE TIME.**

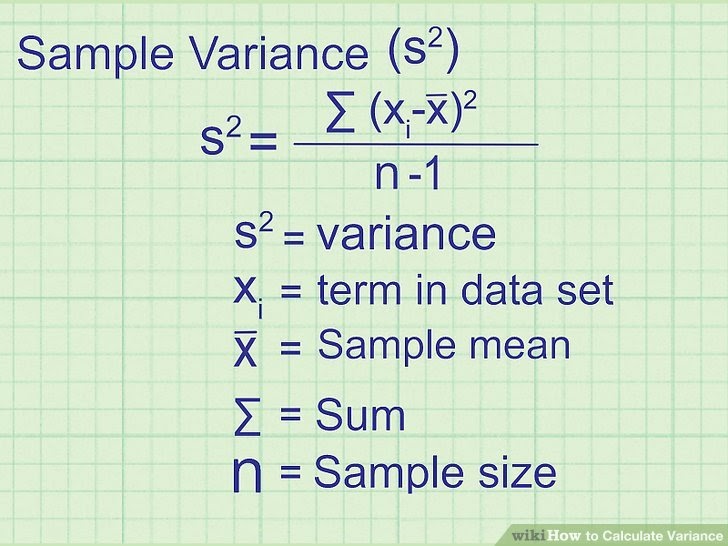
NO WORRIES, HERE IS THE LOGICAL POINT

WE DO NOT ALWAYS HAVE THE DATA OF THE POPULATION WITH US AND NEIGHTER ON ALL THE TOPICS, SO WE ALWAYS CONSIDER SAMPLE DATA OVER POPULATION DATA.

**Sample data:**

Normally, however, only a subset is available, and the variance calculated from this is called the sample variance. The variance calculated from a sample is considered an estimate of the full population variance.

**FORMULA FOR SAMPLE VARIANCE:**



**I know you all are thinking why n-1:**

I will answer this also, when calculating a sample variance to estimate a population variance, the denominator of the variance equation becomes n− 1 so **that the estimation is unbiased and does not underestimate the population variance.**

**How Do I Calculate Variance?**

**Follow these steps to compute variance**:

1. Calculate the mean of the data.
2. Find each data point's difference from the mean value.
3. Square each of these values.
4. Add up all of the squared values.
5. Divide this sum of squares by n – 1 (for a sample) or N (for the population).

## Why Is Standard Deviation Often Used More Than Variance?

Standard deviation is the square root of variance. It is sometimes more useful since taking the square root removes the units from the analysis.

**This allows for direct comparisons between different things that may have different units or different magnitudes**.

For instance, to say that increasing X by one unit increases Y by two standard deviations allows you to understand the relationship between X and Y regardless of what units they are expressed in.

## Advantages and Disadvantages of Variance

Statisticians use variance to see how individual numbers relate to each other within a data set**, rather than using broader mathematical techniques such as arranging numbers into quartiles**.

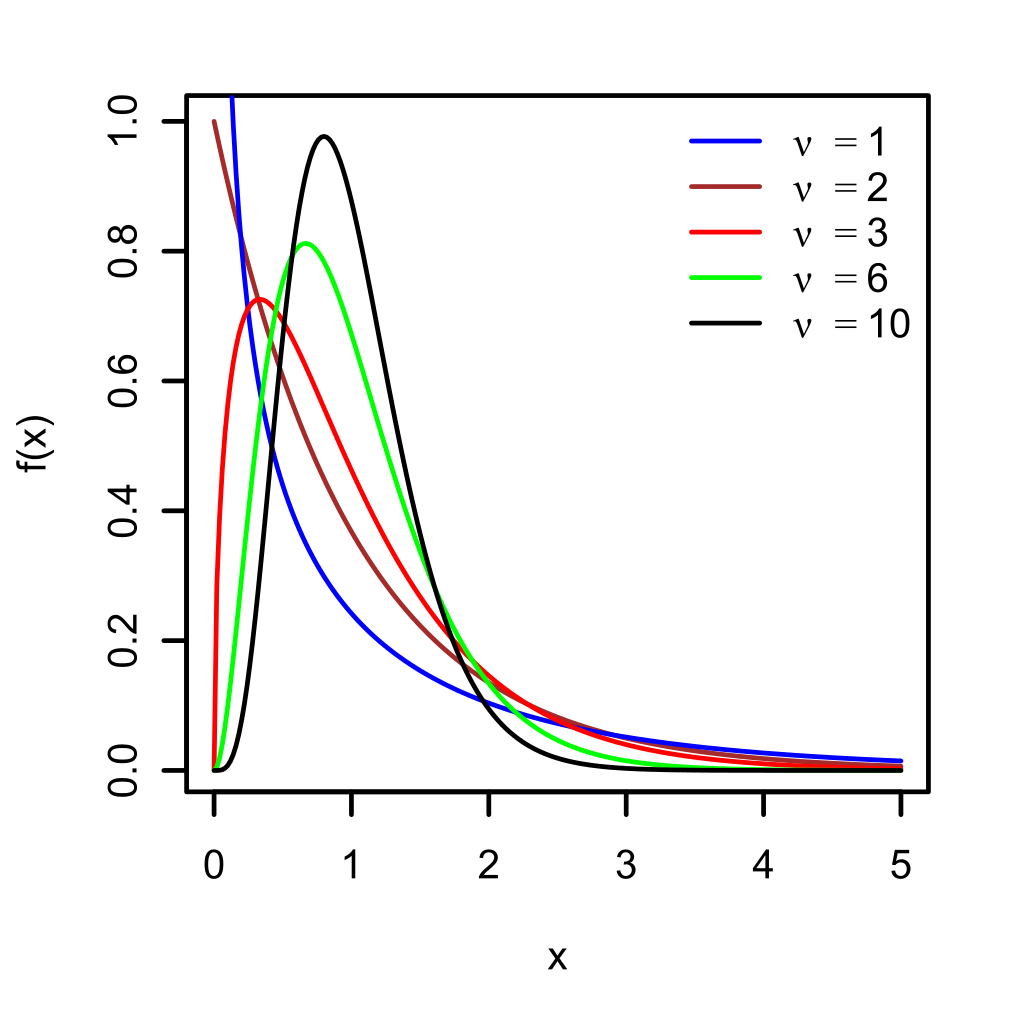
**The advantage of variance is that it treats all deviations from the mean as the same regardless of their direction**. The squared deviations cannot sum to zero and give the appearance of no variability at all in the data.

**One drawback to variance, though, is that it gives added weight to outliers. These are the numbers far from the mean. Squaring these numbers can skew the data.**

**Another pitfall of using variance is that it is not easily interpreted.** Users often employ it primarily to take the square root of its value, which indicates the standard deviation of the data.

**CONCLUSION: Variance can be positive and zero and also be affected by the outliers but variance can not be negative.**

|  |
| --- |
| I think it also important to mention, so that there will be no confusion on the topic variance.  Whether the n size should be greater or less while performing any calculation,  So, the MORE the **n-size** the **better** will be the shape of the DISTRIBUTION and the more the variance the better the shape of the distribution. |



The fig., above is for sample variance and for black distribution, the higher the variance, better the shape of the distribution as compared to other values of variance.

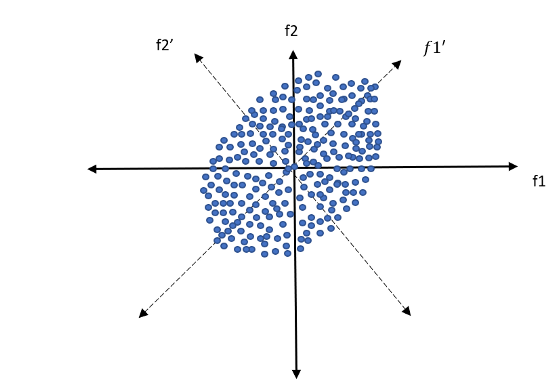
**COVARIANCE:**

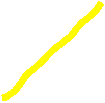
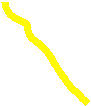
**covariance** is a measure of the joint variability of two random variables. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values (that is, the variables tend to show similar behaviour), the covariance is positive.

 In the opposite case, when the greater values of one variable mainly correspond to the lesser values of the other, (that is, the variables tend to show opposite behaviour), the covariance is negative.

The sign of the covariance therefore shows the tendency in the linear relationship between the variables and for non-linear curves spearman rank correlation is used.







**Geometric visualisation of covariance**

Here, a basic knowledge of quadrants is required for understanding the relationship between two data set or variables (in our example, random f1 and f2).

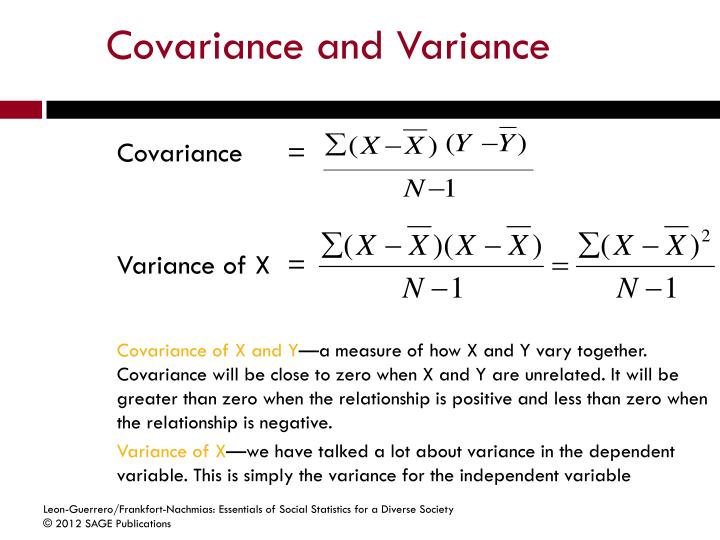
**In the first quadrant** the relation is positive as both f1 and f2 are positive in this quadrant, the **dotted line f1’ shows positive perfect relationship** between each other.

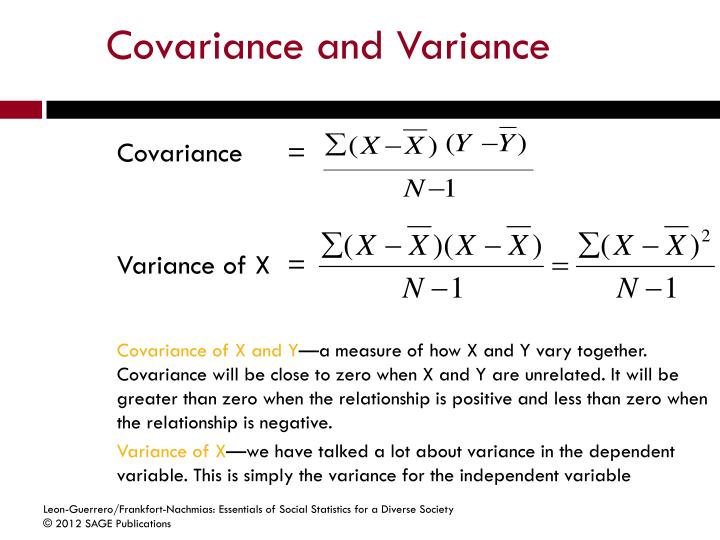
**In the second quadrant** the relation is negativeas the data set f1 is negative and f2 is positive so thecovariance is negative.

**Similarly, for third and fourth quadrant, I have marked the relation in the diagram itself, kindly, review.**

**Now, for the origin, horizontal and vertical lines, covariance is zero as there is no specific relation between f1 and f2.**

**FORMULA FOR COVARIANCE**





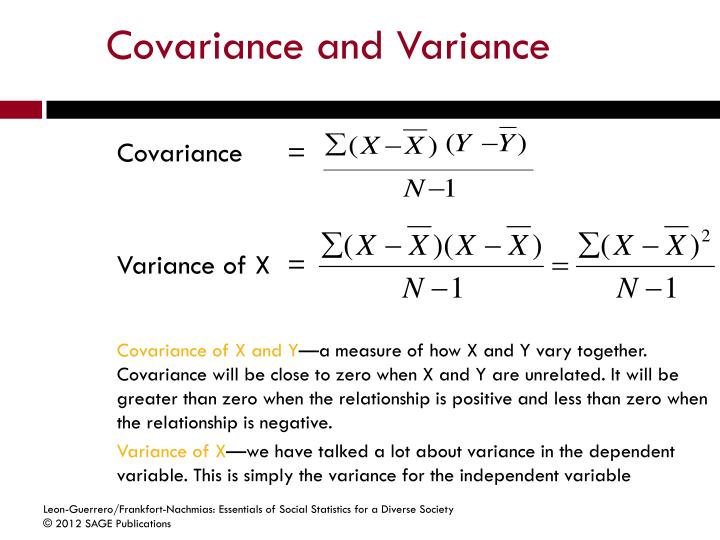
**DIFFERENCE BETWEEN COVARIANCE AND VARIANCE**

**Variance only talks about the dispersion, spread or deviation from mean but in what direction does they follow, can not be concluded only by variance.**

**To know the relationship, we need covariance.**

**Now, why covariance?**

**Because, Co means two and variance means dispersion relation which means, tells about the relationship of two variables or dataset simultaneously as mentioned above section of covariance.**



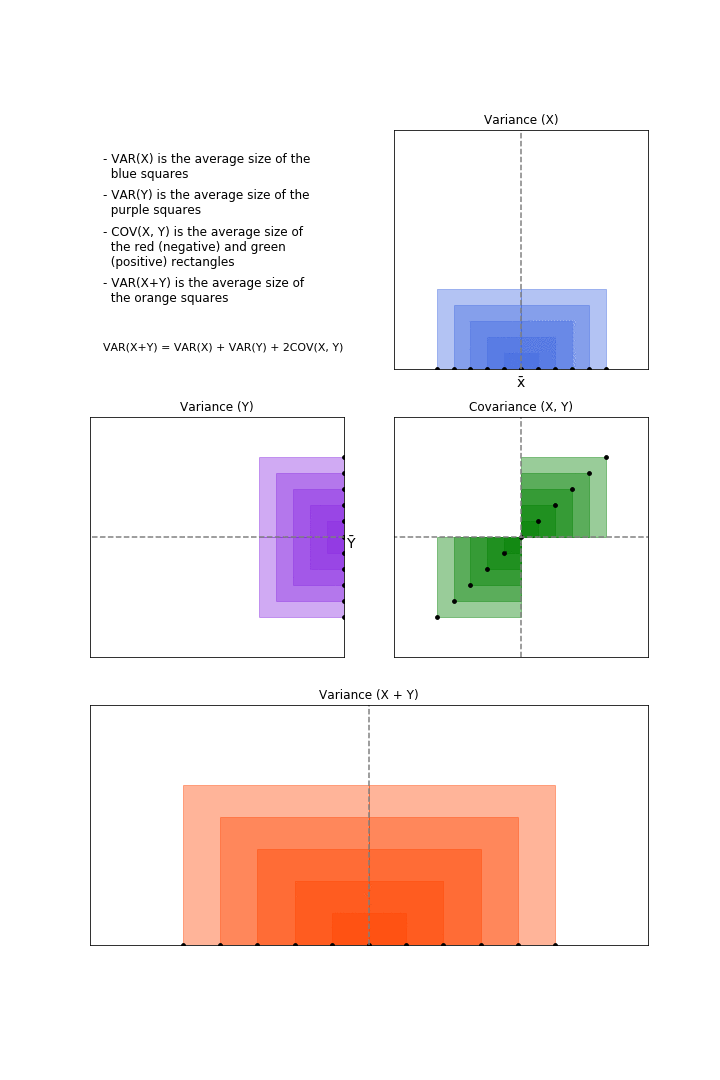
**If we find out the relation between x and x only then covariance and variance will be equal.**

**COV (X, X) = VAR(X); use the formula to see for the proof of the relation.**

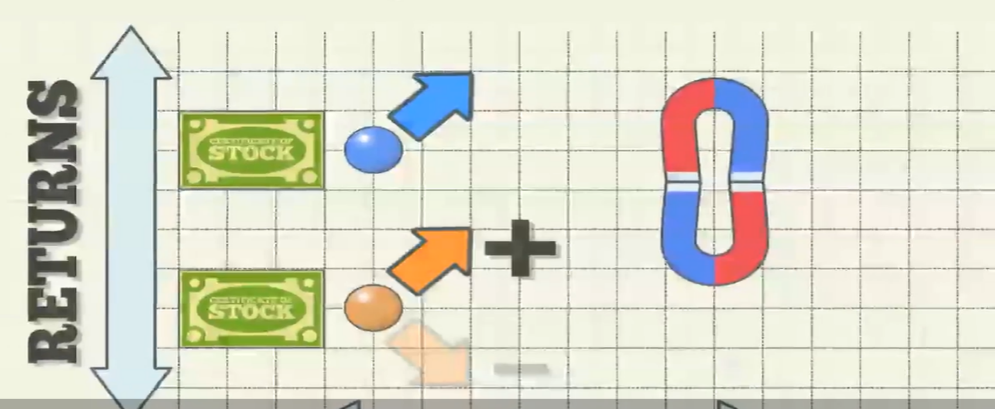
**In the fig., below, VAR(X) AND VAR(Y) only depicting dispersion from mean (x bar and y bar). Only COV (X, Y) depicts relationship between x any y**

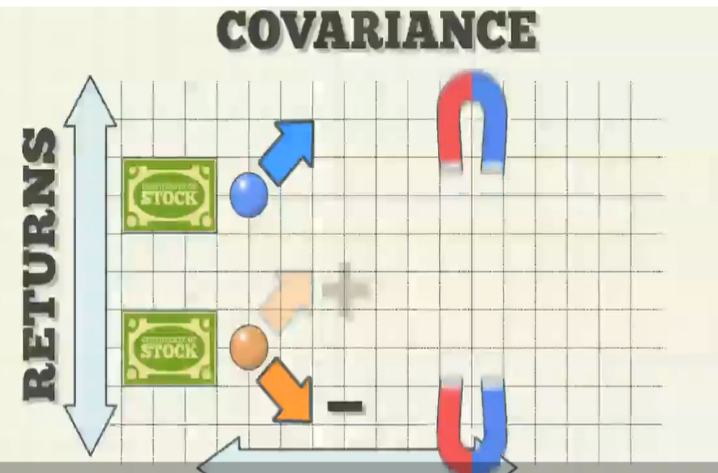
**Green shows positive relation in quadrant 1 and 3.**

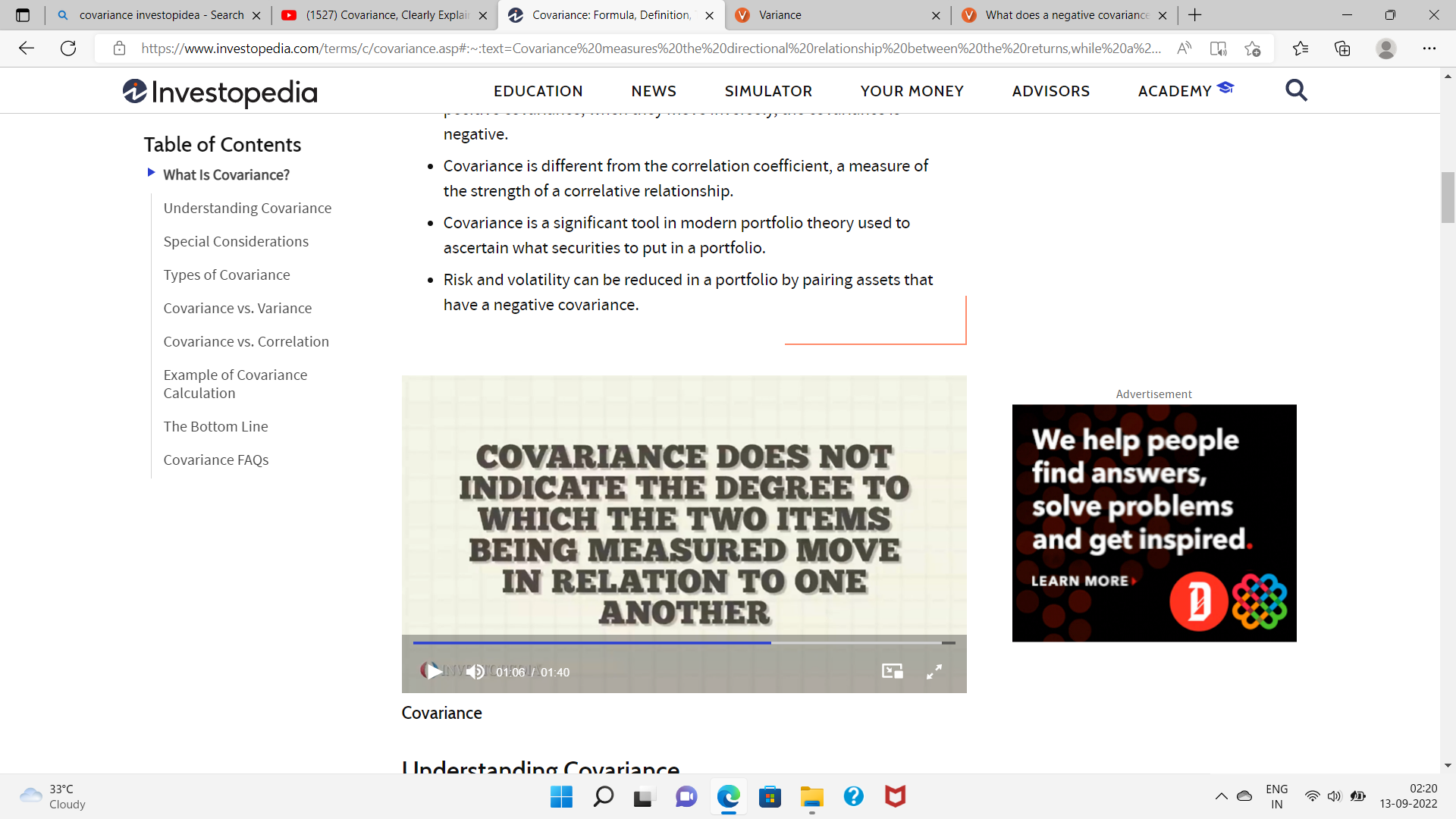
**Red shows negative relation in quadrant 2 and 4.**

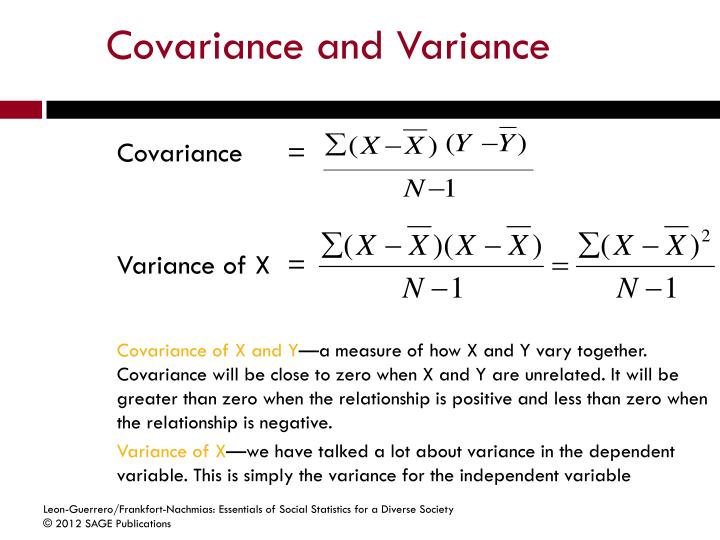


SUMMARY:

positive relation, attracts each other, depicted by magnets, as blue and orange dataset have positive returns from stock.

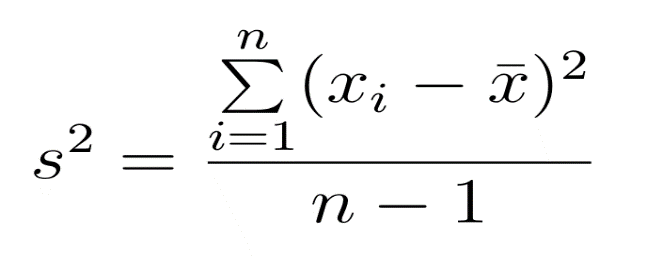
negative relation, oppose each other, depicted by each other, as blue have positive and orange have negative dataset, depicting negative returns from stock.

****



**NUMERICAL**:

TWO DATA SET: X = (2, 4, 6, 8, 10) Y = (1, 3, 8, 11, 12)

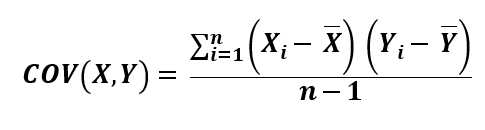
S (SQAURE) = VARIANCE 

VARIANCE IS A MEASURE OF HOW SPREAD OUT THE MEASURE OF DATA SET ARE

AVERAGE OF X = 2+4+6+8+10/5 = 6 VAR(X) = 40/4 = 10

AVERAGE OF Y = 35/5 = 7 VAR(Y) = 94/4 = 23.5

COVARIANCE OR COV(X, Y) HOW THE TRENDS OF TWO DATA SET RELATED

X-MEAN Y-MEAN 



COV (X, Y) = (-4)(-6)+ (-2)(-4)+ (0)(1)+ (2)(4)+ (4)(5)/ 4 = + 15

DEPICTS POSITIVE TREND (MOVING IN SAME DIRECTION) BUT CANNOT TELL HOW MUCH X AND Y ARE POSITIVELY RELATED.