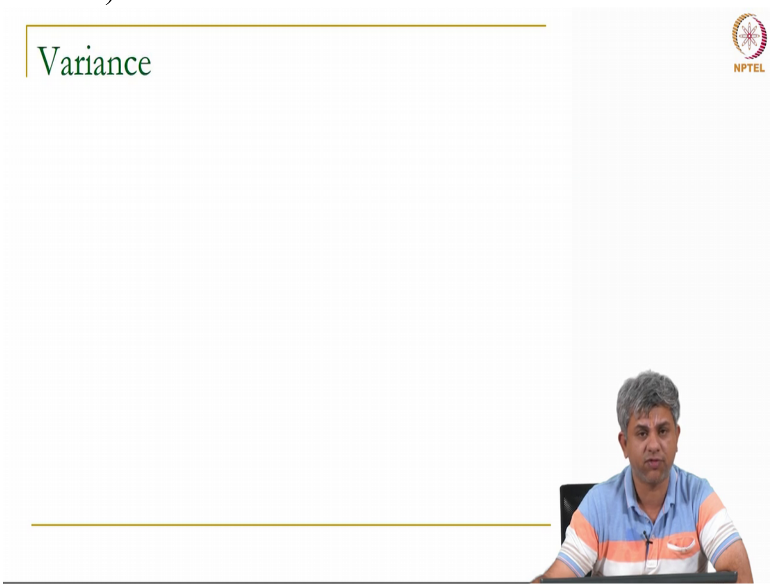
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology Madras
Variance Covariance

(Refer Slide Time: 00:13)

In the previous video we had looked at the idea of Expectation. In this video we will be looking at some additional ideas which are called variance and covariance. Again you will be familiar with some of these

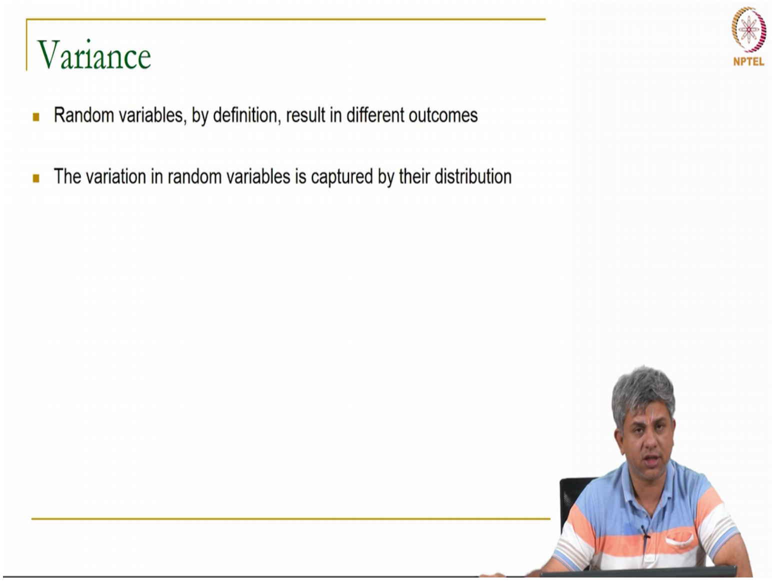
(Refer Slide Time: 00:28)



from school and some of these might be unfamiliar.

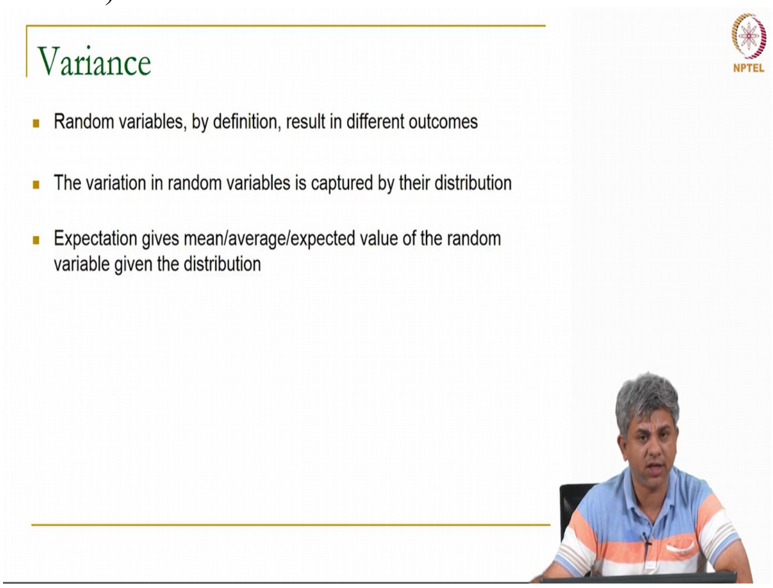
So let us continue

(Refer Slide Time: 00:33)



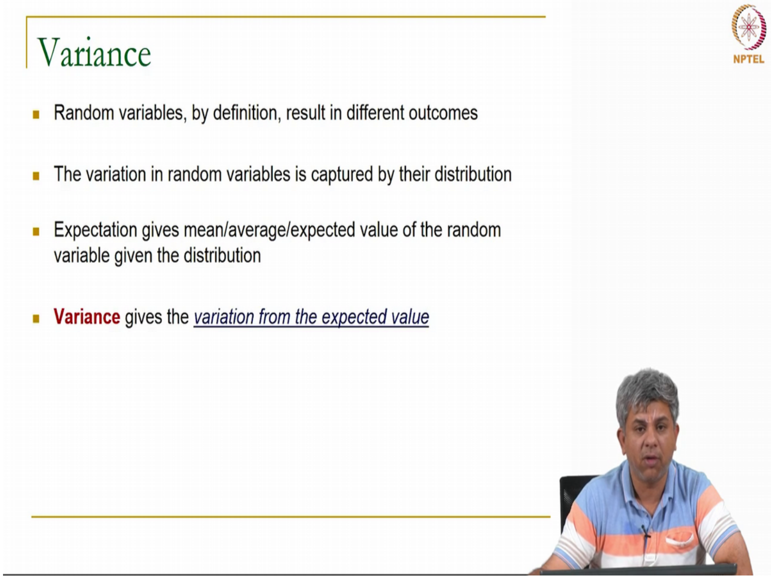
our discussion from where we had our expectation discussions. Once again the same

(Refer Slide Time: 00:37)



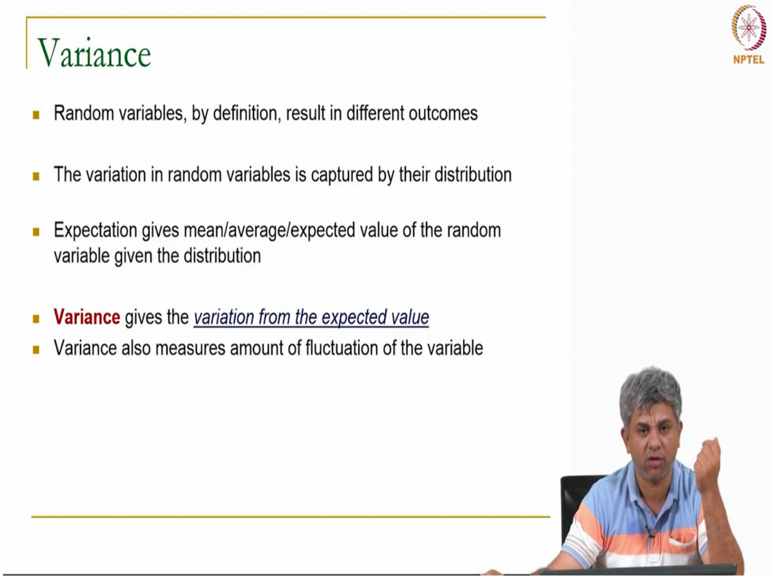
idea continues from before. If you have a random variable it is going to vary, Ok. And its variation is actually captured by its distribution. What expectation does is, on an average what can you expect? Ok this is what it will tell you.

(Refer Slide Time: 00:53)



But what we want to know is variance. Variance is, Ok, it is given that this is the expected value. For example expectation for dice throw is 3 point 5 but how much more can it go, how much less will it go? What is the variation from the expected value? What is the variation from the mean, from the expectation? This is what variance talks about,

(Refer Slide Time: 01:14)

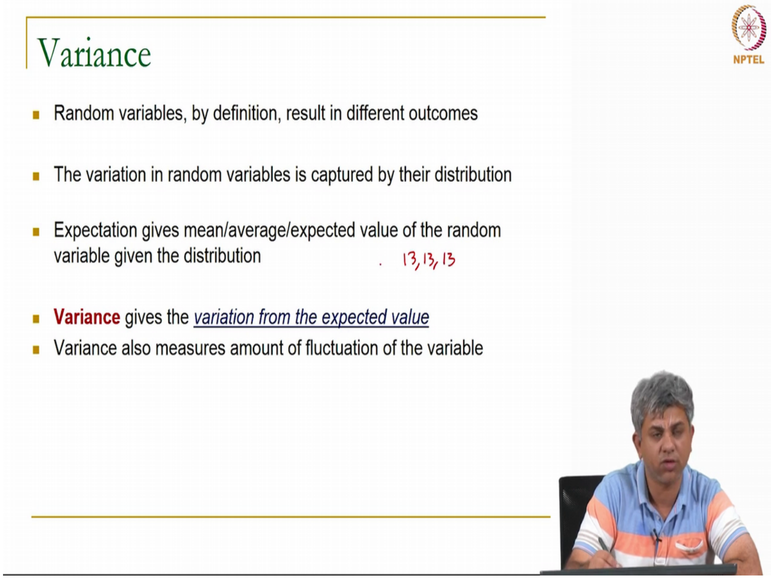


Ok.

Variance also measures; in many cases you can also measure how much does the quantity fluctuate? Ok. So it is possible, entirely possible for two distributions to have the entirely same expectation but different variances.

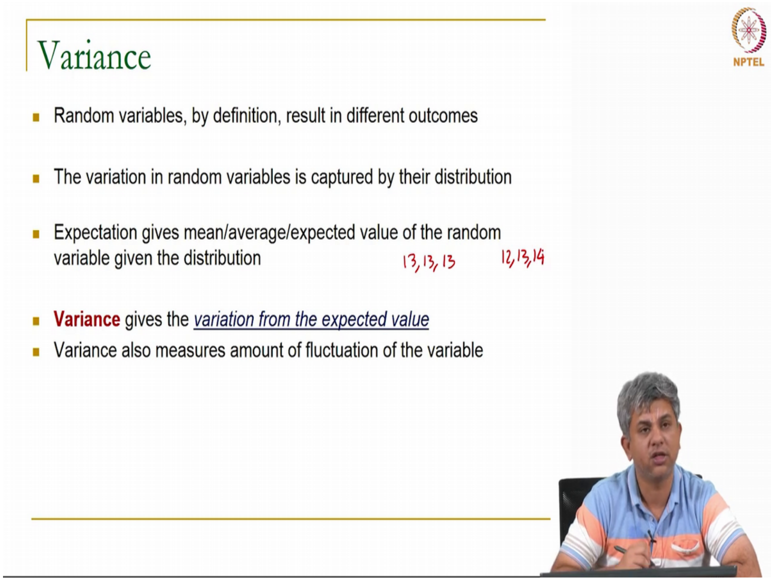
For example if I take the value 13, 13, 13 or, Ok

(Refer Slide Time: 01:33)



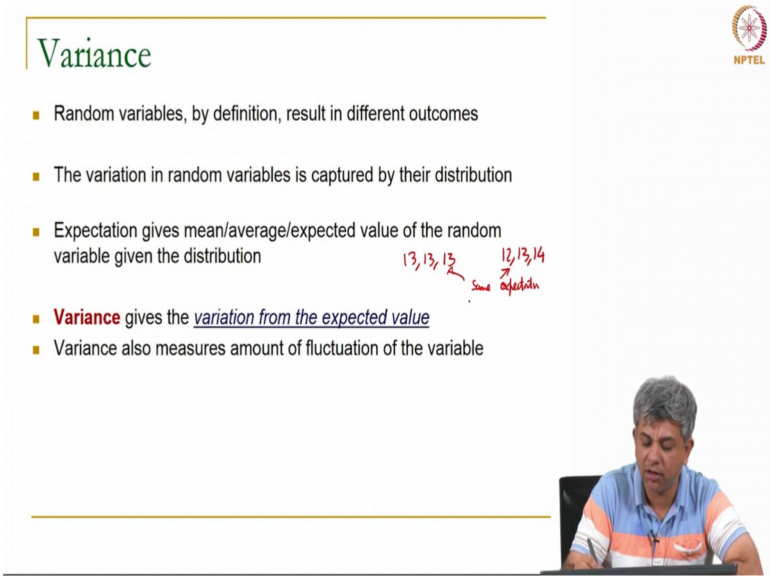
so let us say there is some random value which takes the value 13, 13, 13. In this case the expectation is 13. Another one

(Refer Slide Time: 01:42)



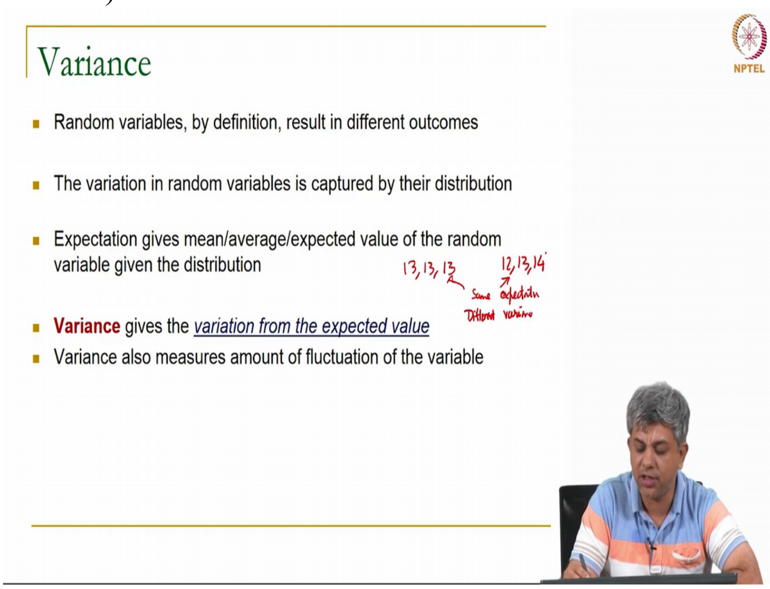
is 12, 13, 14. In these cases is the same expectation but

(Refer Slide Time: 01:53)



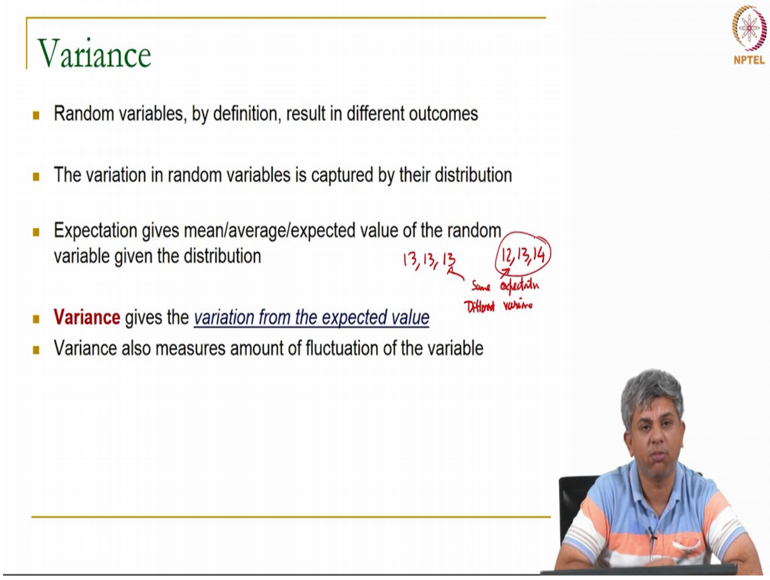
different variance, Ok.

(Refer Slide Time: 01:58)



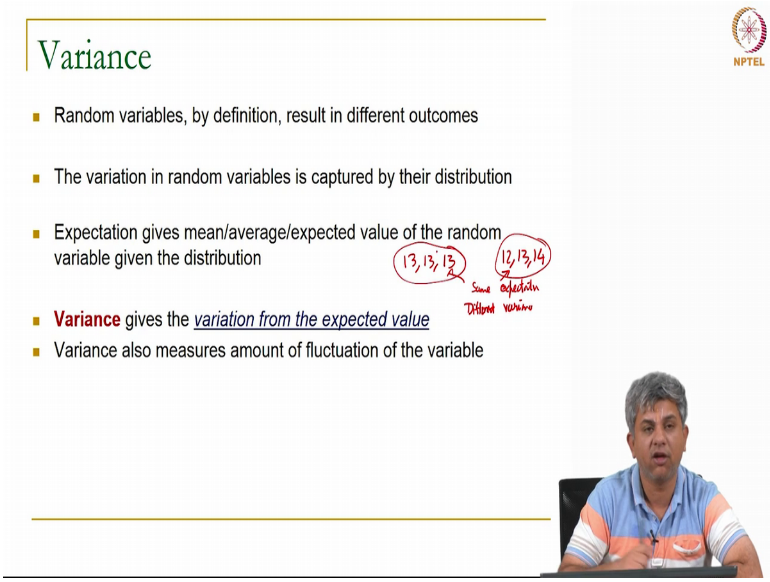
We see this one varies

(Refer Slide Time: 01:59)



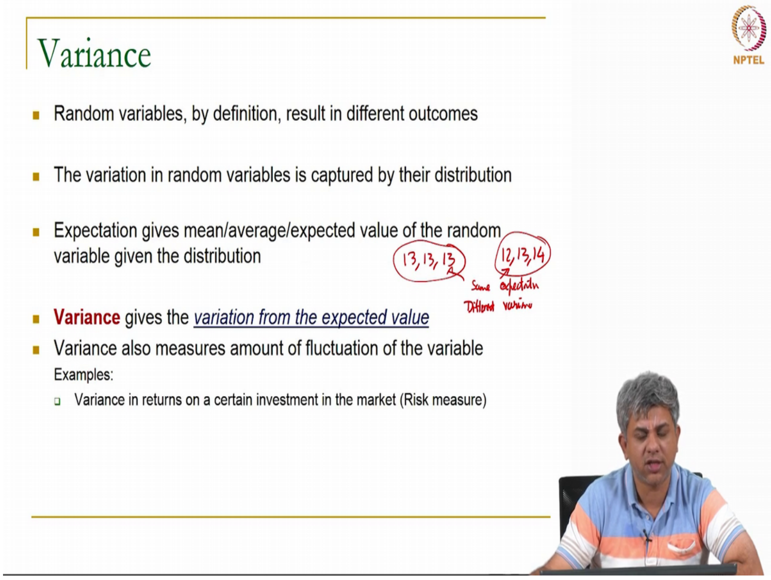
more from the mean and this one in fact does not vary

(Refer Slide Time: 02:03)



from the mean at all and we actually get variance is equal to 0. So

(Refer Slide Time: 02:08)

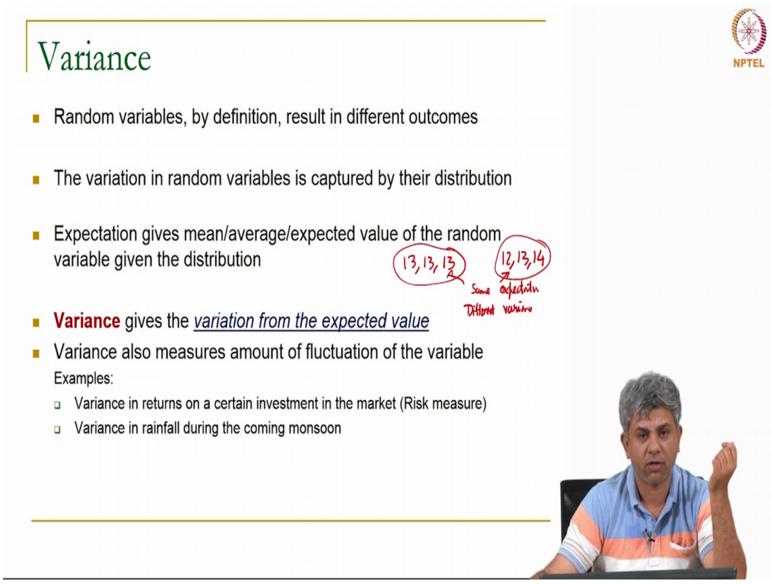


once again some practical example, if we go back to the investment example, expected returns is one thing but you would also like to know how much can change, Ok

So your expected returns might be x y z amount of Rupees, but your actual variation, you could actually have a profit or you could actually have a loss, even though your expected value might be a profit, Ok.

So some times and in fact lot of times in modern, what is called, modern portfolio theory usually the variance is what is used to find out what is the risk in investment.

(Refer Slide Time: 02:45)



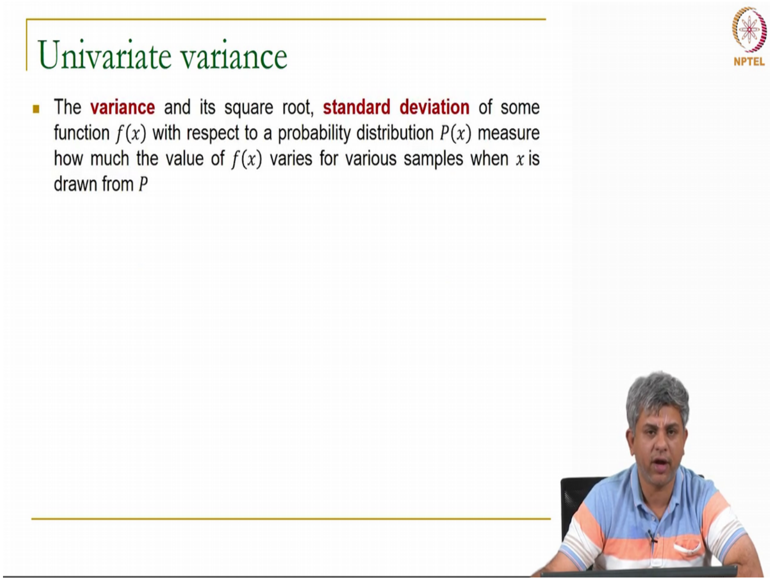
Though this is a little bit controversial but anyway, people, many people do use variance as a measure of risk.

So similarly I gave the example of the expected value of rainfall during the coming monsoon. But you could also look at the variance in rainfall during coming monsoon.

Not only do you want to know what the rainfall is but you also want to know what is the maximum it can go, you know in case you have to plan for, in a flood, Ok. Or what is the least that it can go in case you have to plan for drought, Ok. So variance measures these ideas.

So let us look

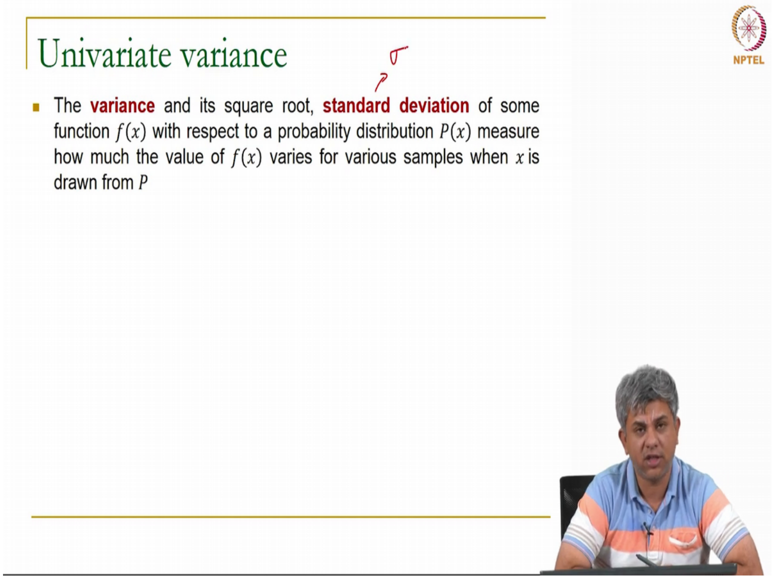
(Refer Slide Time: 03:21)



at mathematically what variance is. Once again we will start with a univariate case. Remember univariate simply means scalar. So if I take a single value x or a single random variable x which is a scalar, you look at two sorts of quantities, variance and standard deviation.

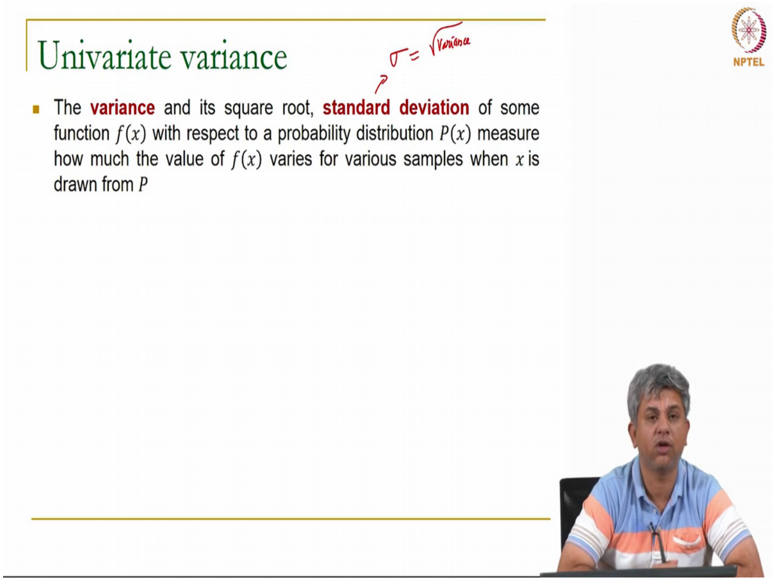
Standard deviation is usually denoted by sigma

(Refer Slide Time: 03:43)



and it is usually equal to, not usually, it is defined as being equal to

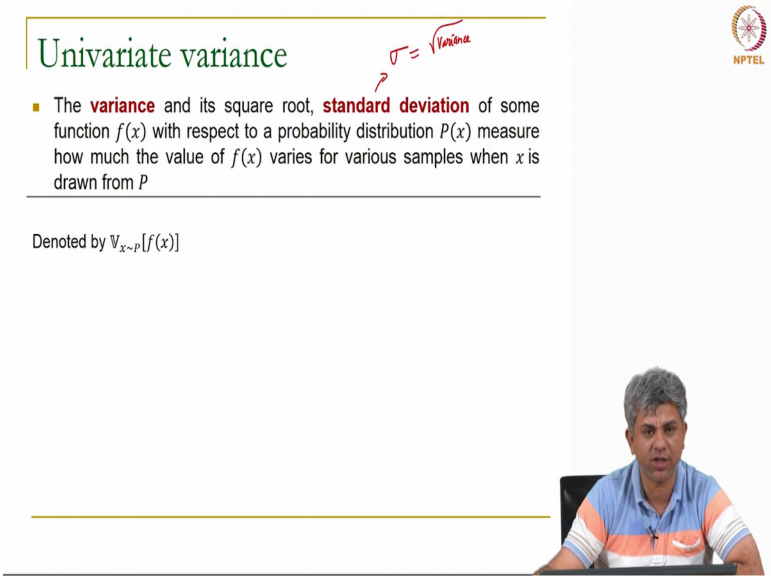
(Refer Slide Time: 03:49)



variance, Ok square root of variance.

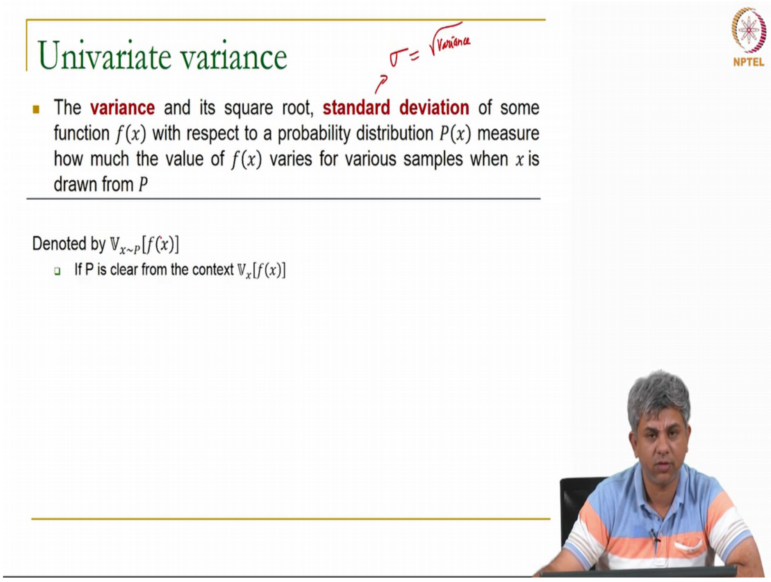
So what variance measures is how much does the value of f vary, Ok, vary from its expected value, Ok when x is drawn from P? Once again the notation,

(Refer Slide Time: 04:05)



if you want to be very precise you will say V drawn for x drawn from P of f of x,

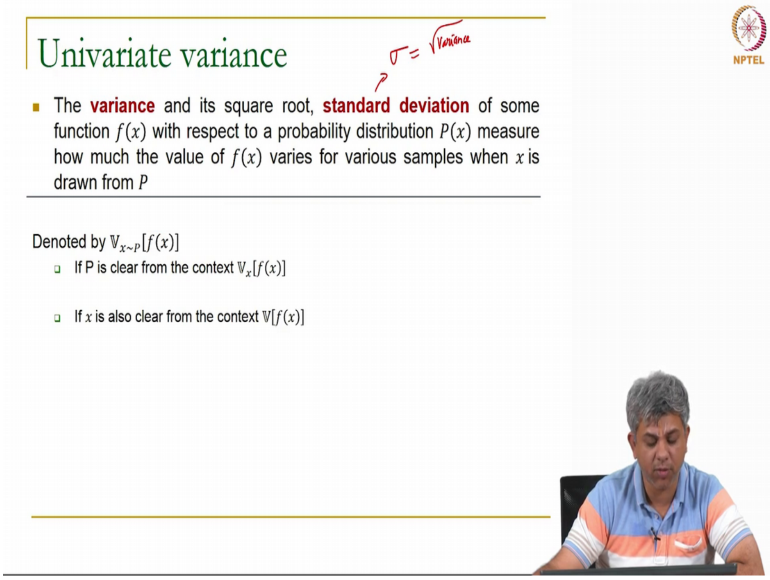
(Refer Slide Time: 04:14)



Ok.

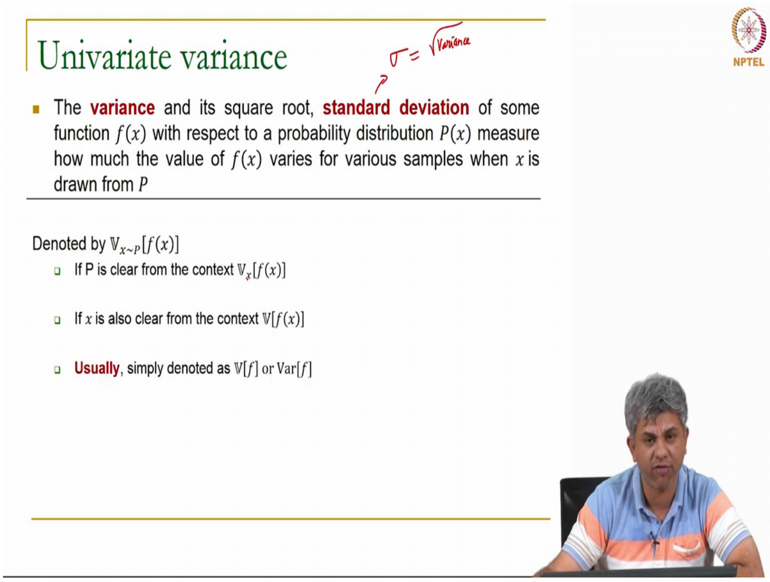
Once again if P is clear

(Refer Slide Time: 04:16)



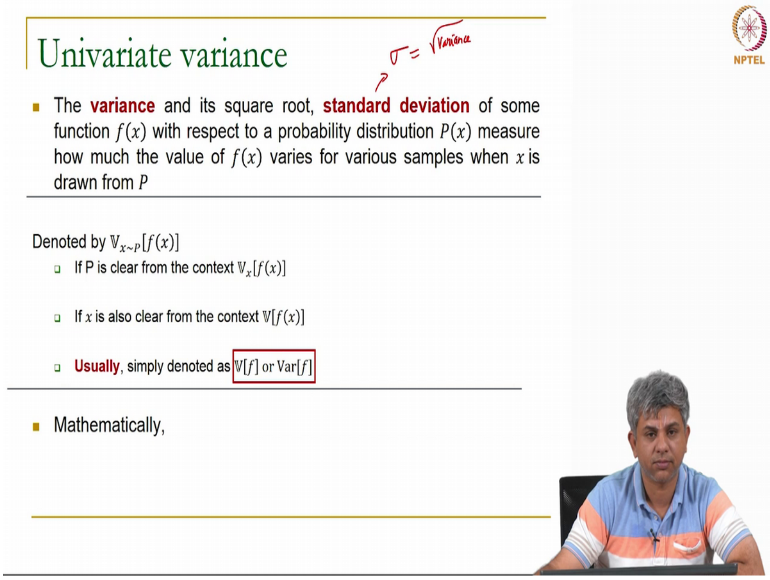
from the context we drop P and simply say V x. If x is also clear we say V of f

(Refer Slide Time: 04:22)



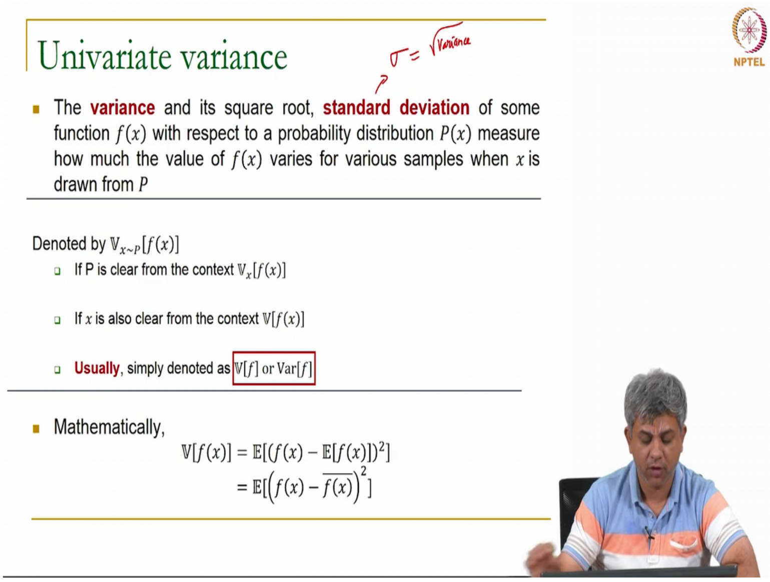
and usually, and this is unlike the expectation, this is the usual notation which is used, which is either V of f or Var of f, Var stands for variance.

(Refer Slide Time: 04:36)



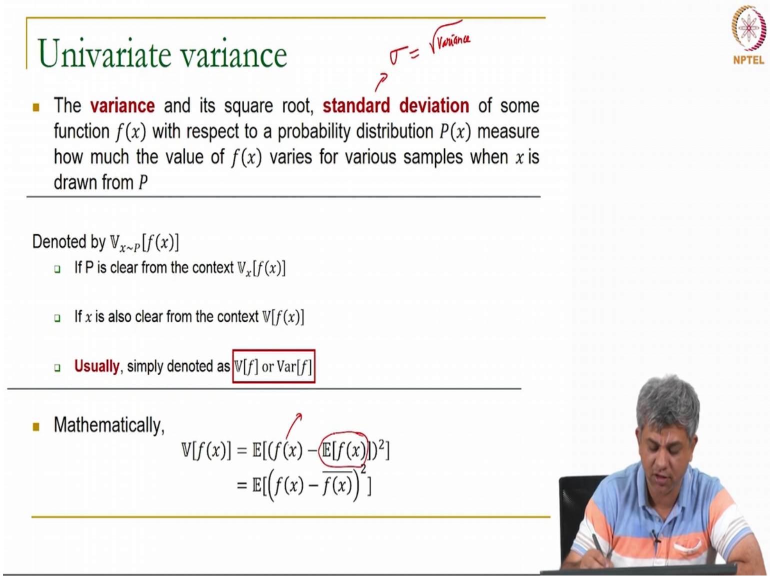
Mathematically

(Refer Slide Time: 04:39)



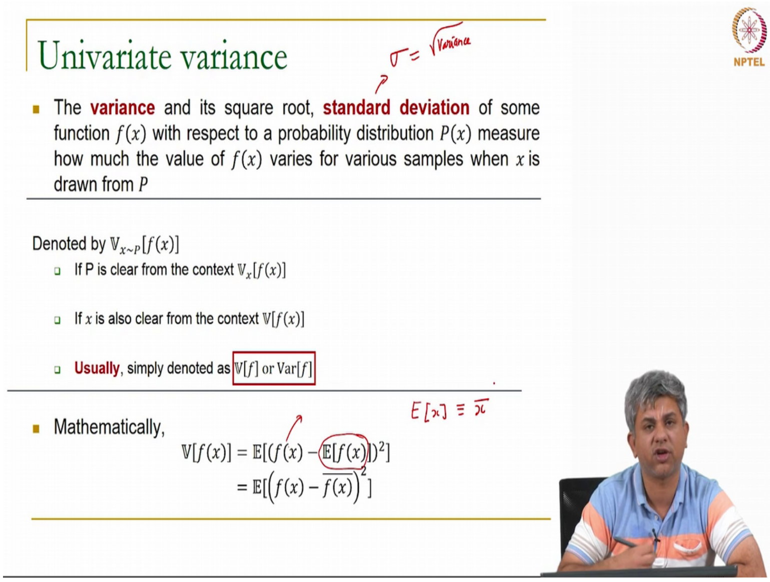
variance of f of x is the variation or the difference between f of x and the expected value of f of x,

(Refer Slide Time: 04:49)



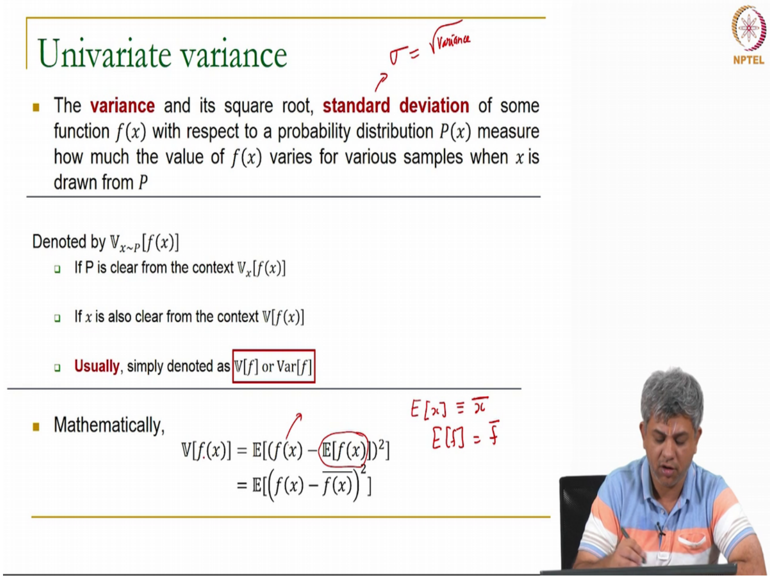
Ok. So this, I will, we use the notation

(Refer Slide Time: 04:56)



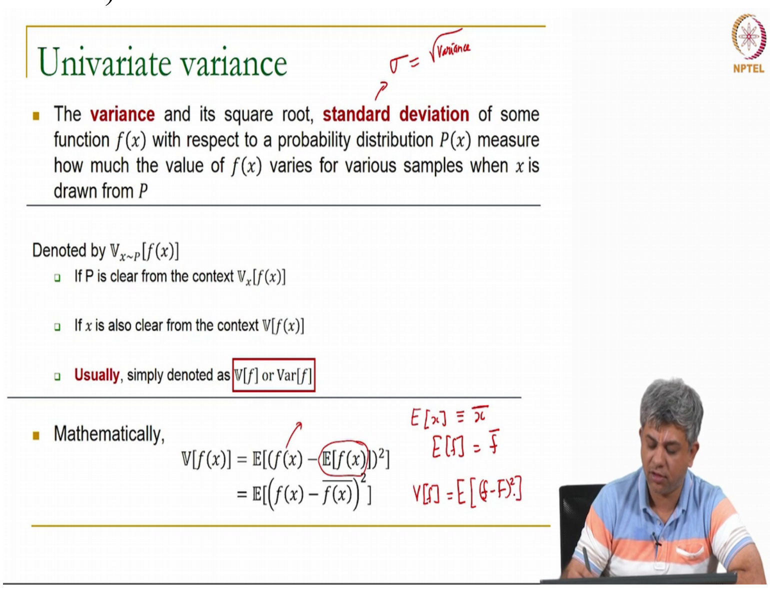
just for simplification x bar to denote expectation of x. So similarly expectation of f can be written as f bar.

(Refer Slide Time: 05:06)



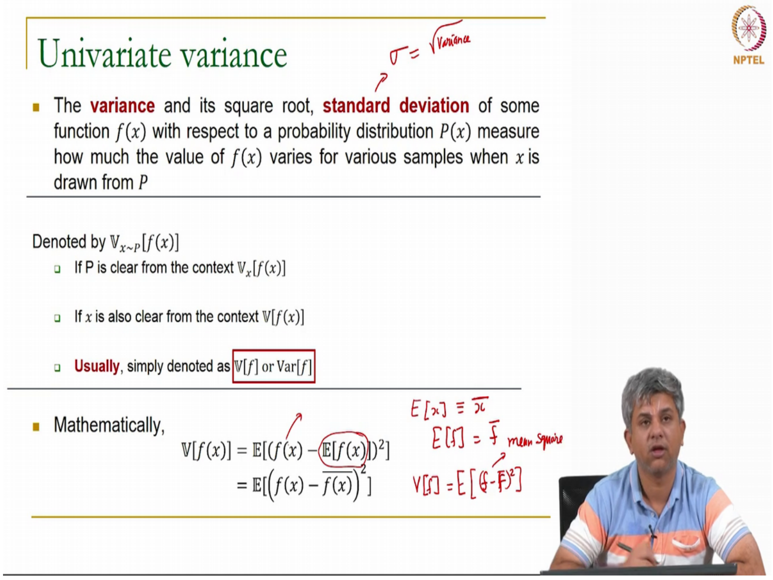
So what is variance of f, it is the expectation of f minus f bar square,

(Refer Slide Time: 05:17)



that is the variance from the mean square. So this is

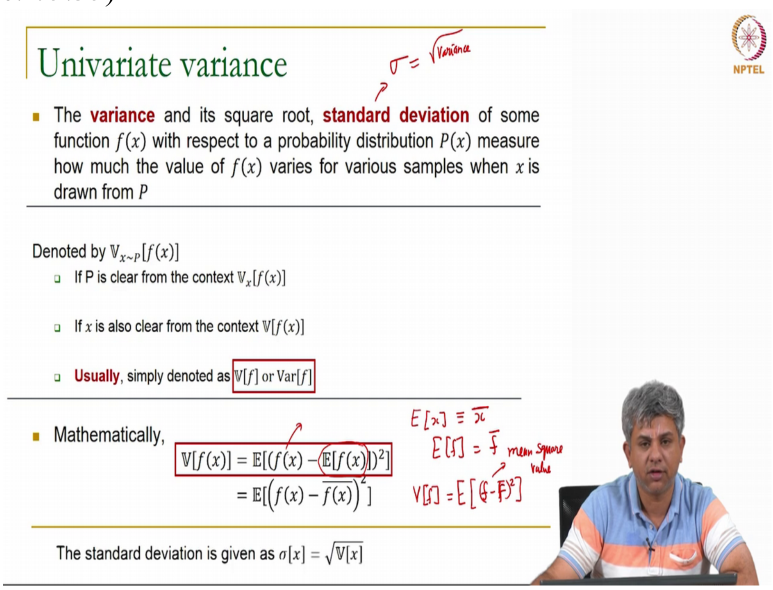
(Refer Slide Time: 05:26)



mean square value. So variance is essentially mean square.

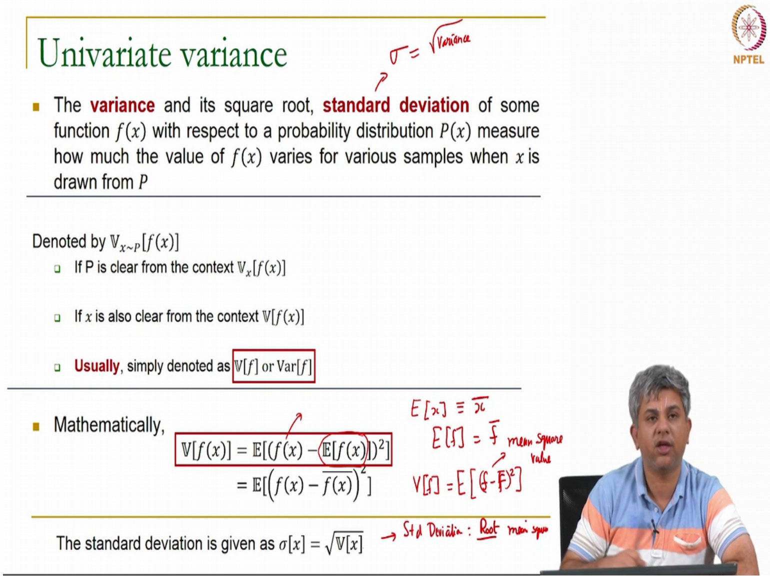
We can also define the

(Refer Slide Time: 05:35)



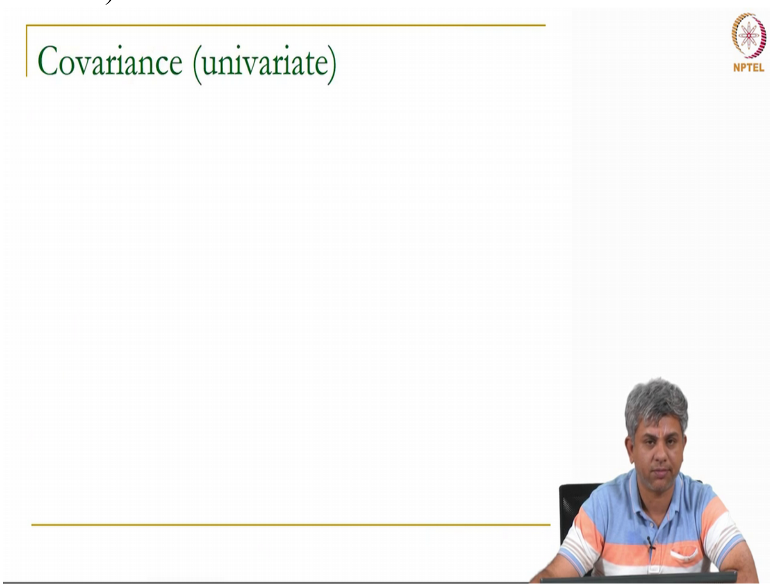
standard deviation. Standard deviation is square root of variance, Ok. So standard deviation is root mean square, Ok. Variance is mean square;

(Refer Slide Time: 05:55)



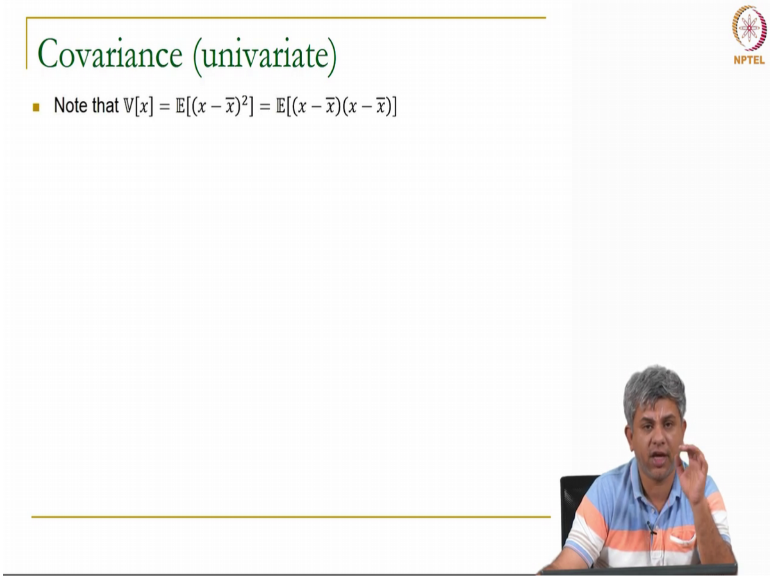
standard deviation is root mean square.

(Refer Slide Time: 05:58)



Now there is another idea, some of you might actually be unfamiliar with this idea. This is the idea of covariance. Now covariance is something we are interested in, if instead of having one variable x, you now have two variables, Ok. When I say

(Refer Slide Time: 06:17)

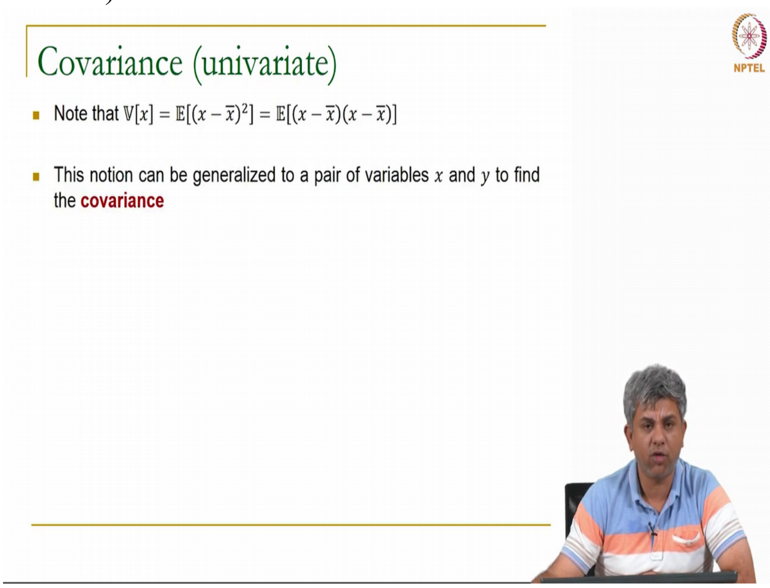


univariate what it means is each of x and y are individually univariate we look at a multivariate case a little bit later, Ok.

So notice that when we defined variance of x this was expectation of, I have not taken f so variance of f, x is simply x minus x bar square, mean square deviation. Now you can write it as expectation of x minus x bar multiplied by x minus x bar.

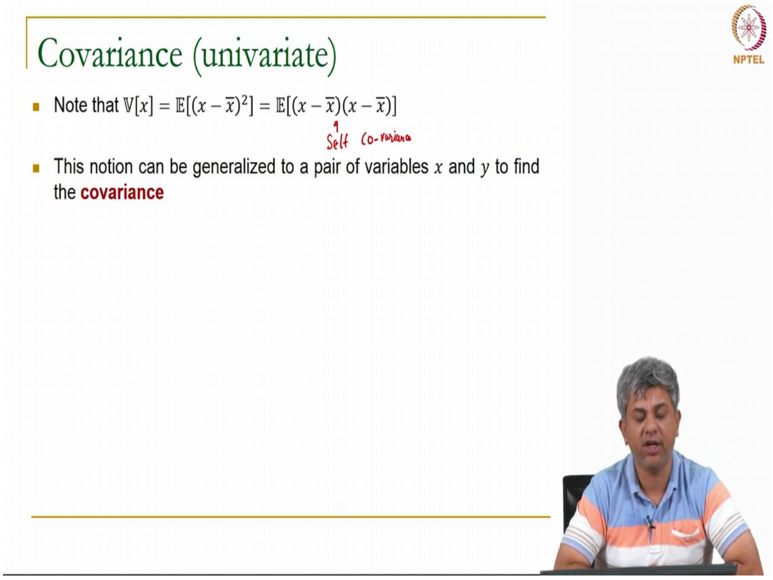
So this idea can now be

(Refer Slide Time: 06:48)



generalized to a pair of variables, Ok. This I can call self covariance. The meaning will become

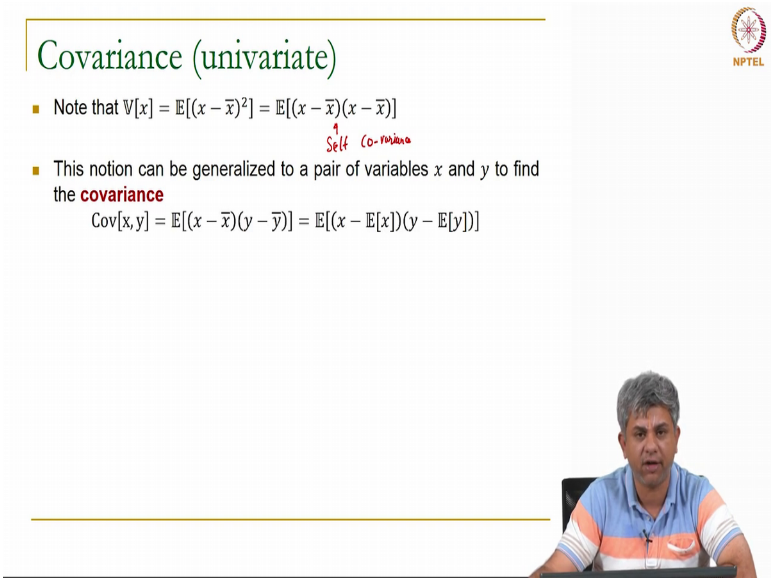
(Refer Slide Time: 07:00)



shortly clear as we go through the next couple of slides, Ok.

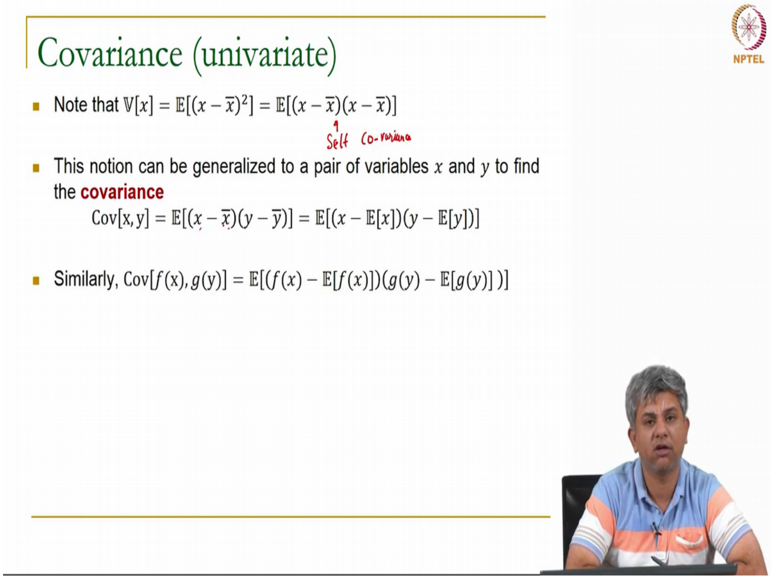
So suppose instead of one variable, you have two variables, we simply generalize this idea

(Refer Slide Time: 07:08)



and we define covariance of two variables x and y as expectation of x minus x bar

(Refer Slide Time: 07:17)

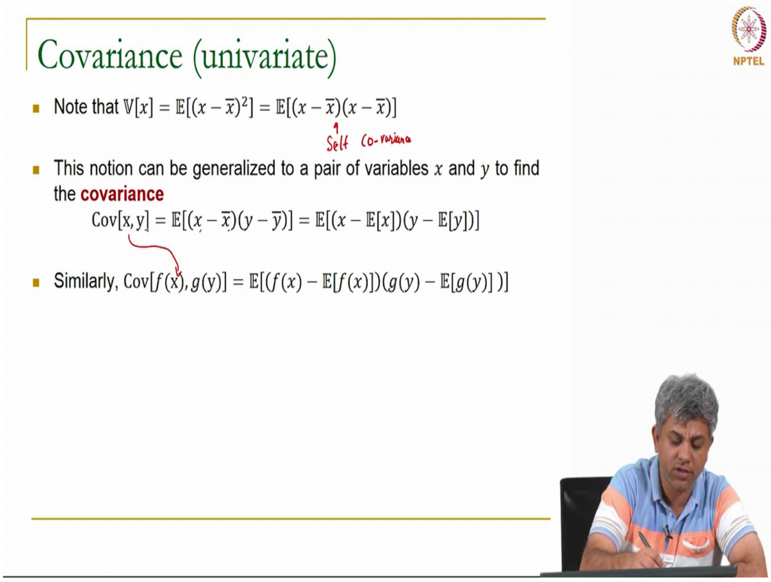


multiplied by y minus y bar, Ok.

What does that mean? This says that how much does x vary from its mean when y varies from its own mean, Ok. So this is what is called covariance. That is how much do x and y vary together, Ok.

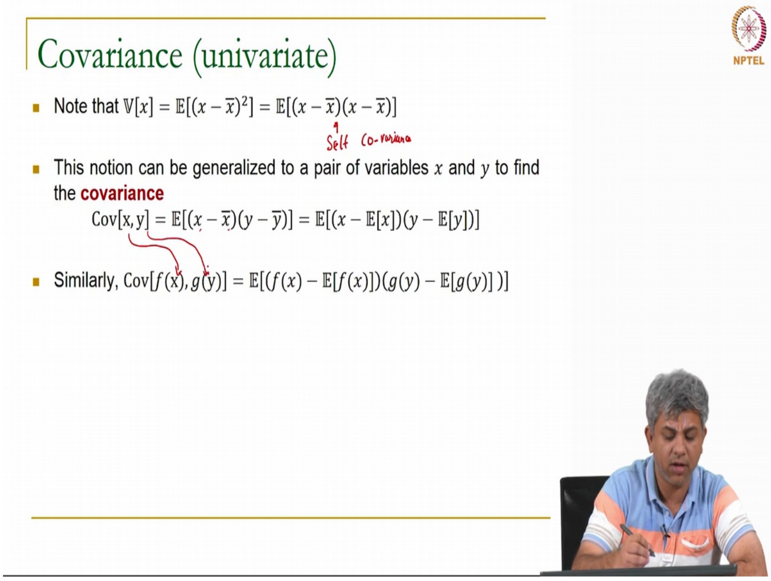
Now just like we defined it, again I use x bar just for compact notation, just like we can define covariance of x and y, you can also define covariance of functions of x and functions of y, all you need to do

(Refer Slide Time: 07:51)



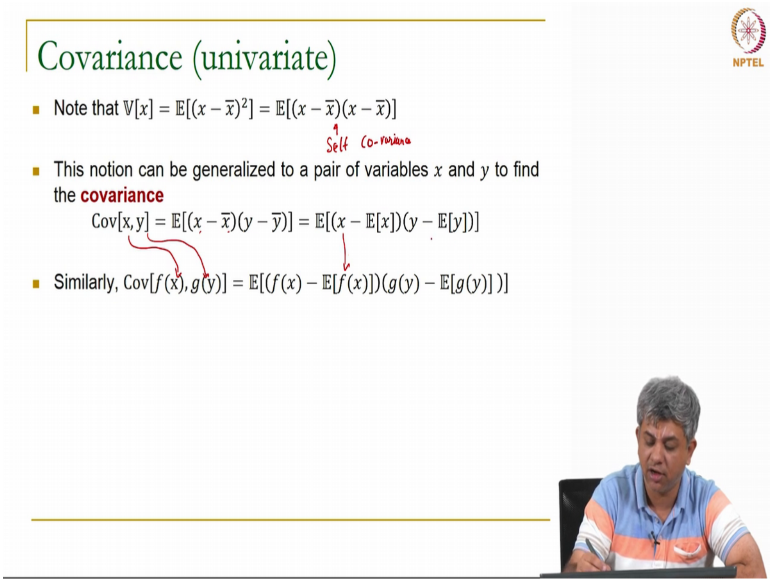
is replace

(Refer Slide Time: 07:53)



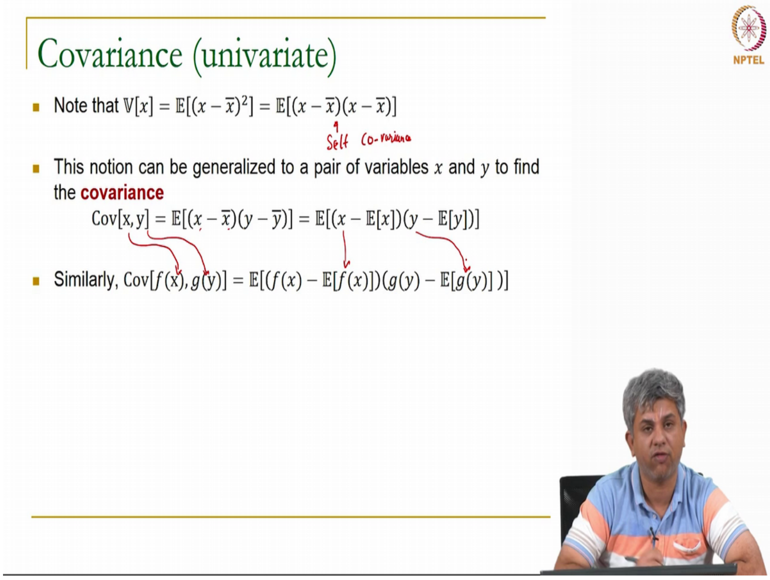
in these expressions x by f of x

(Refer Slide Time: 07:56)



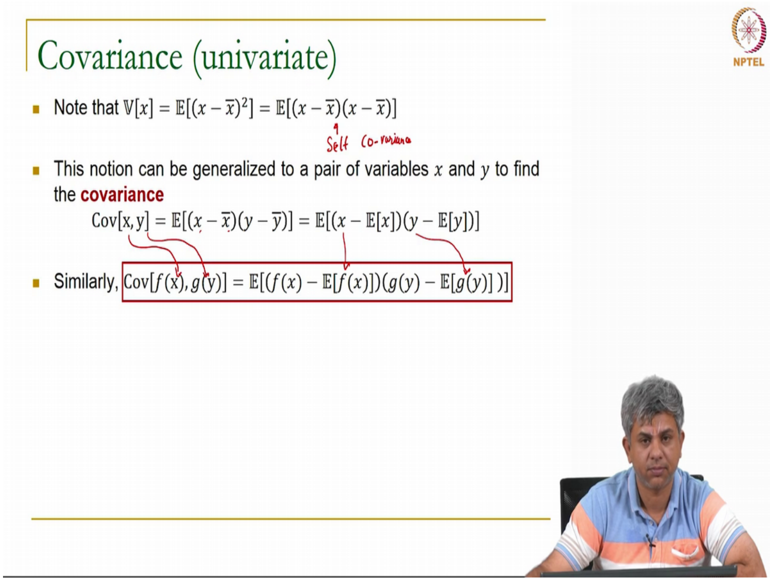
and y by g of y

(Refer Slide Time: 07:58)

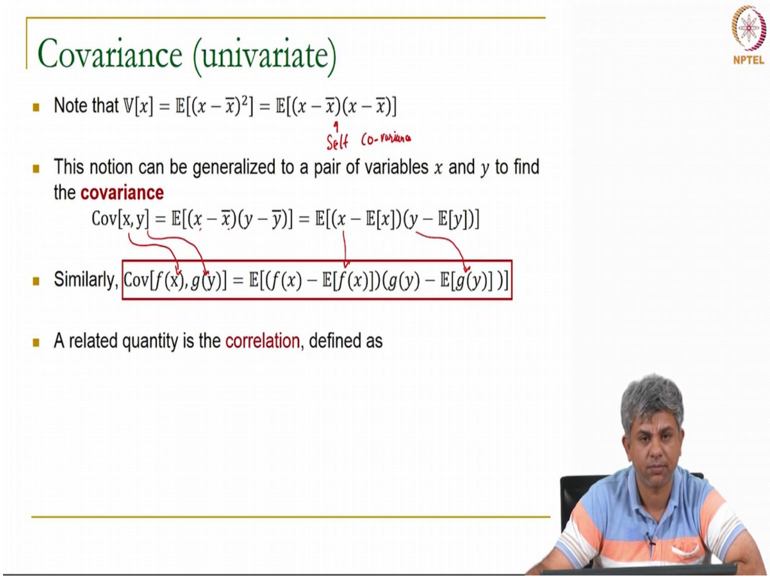


and you will get the corresponding covariance functions for f of x and g of y.

(Refer Slide Time: 08:05)

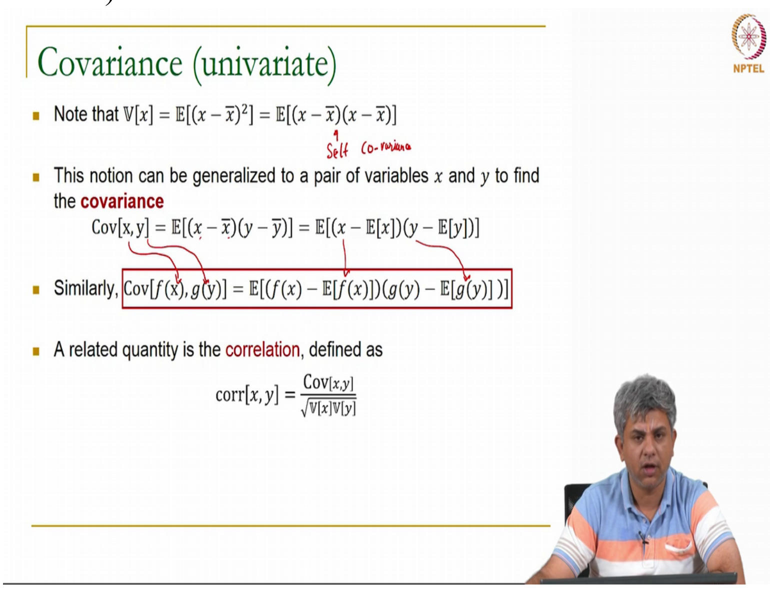


(Refer Slide Time: 08:06)



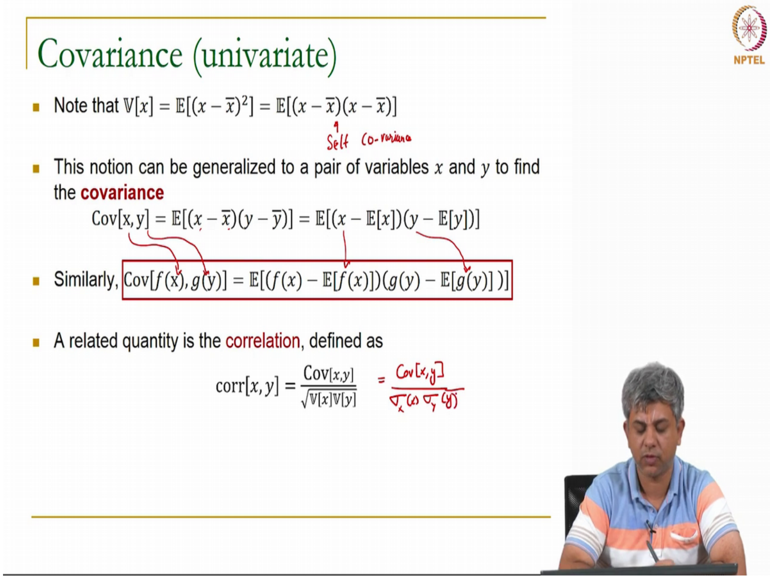
Now closely related quantity to covariance is something called the correlation.

(Refer Slide Time: 08:14)



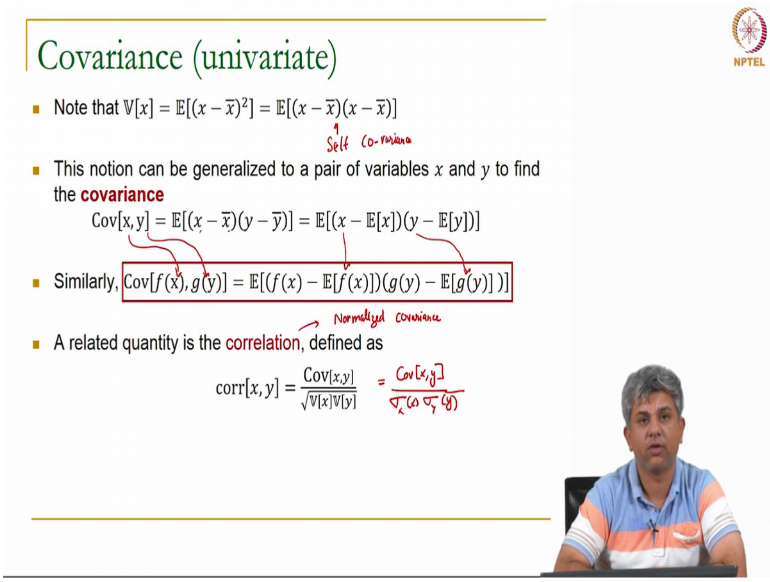
You will see that it is simply the covariance divided by the square root of variance of x and variance of y. Another way to say it is remember square root of variance of x is standard deviation of x.

(Refer Slide Time: 08:35)



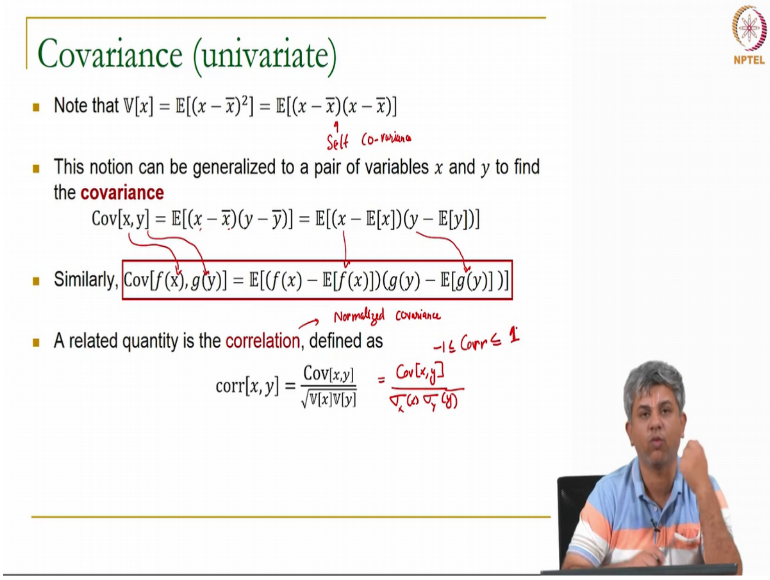
So essentially correlation is normalized, Ok.

(Refer Slide Time: 08:47)



What does normalized mean? Normalized means, you know just like we normalize the vector and make it into unit normal, similarly you are normalizing covariance so that its size stays between certain limits. So as it turns out, covariance will always, sorry correlation will always lie between minus 1 and

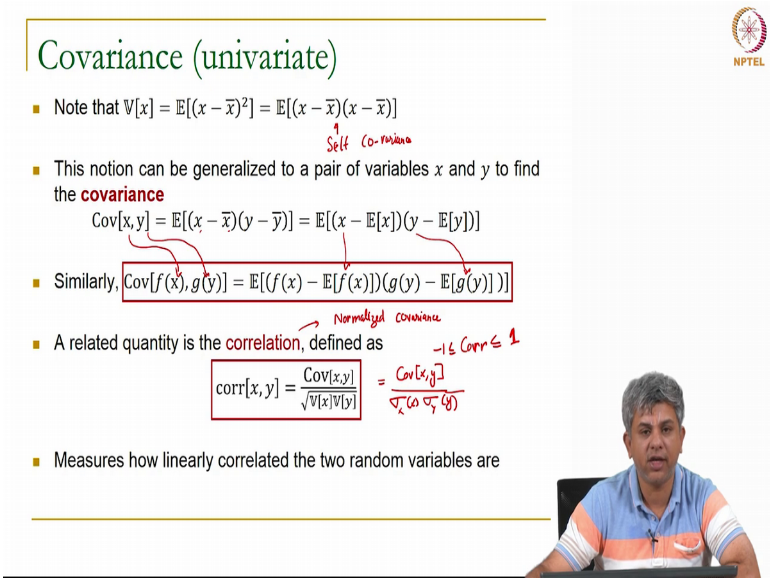
(Refer Slide Time: 09:10)



1, Ok.

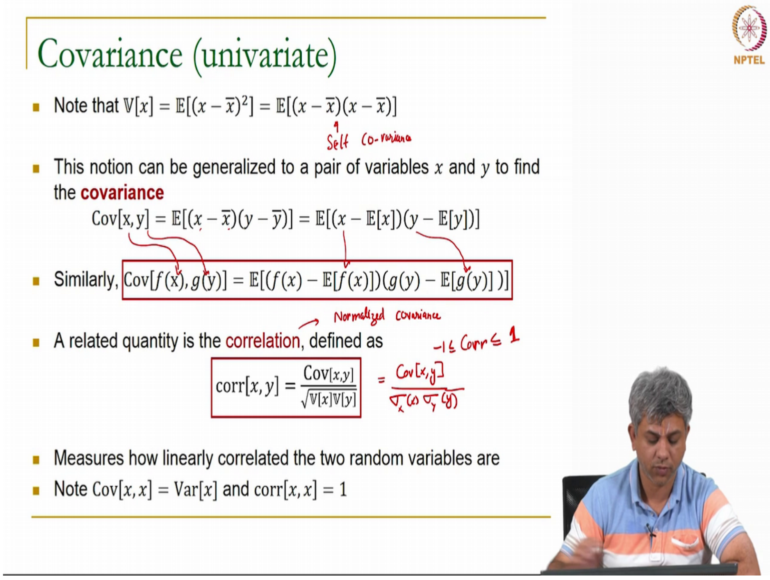
Now we had looked at the

(Refer Slide Time: 09:15)



ideas of variance and covariance.

(Refer Slide Time: 09:20)

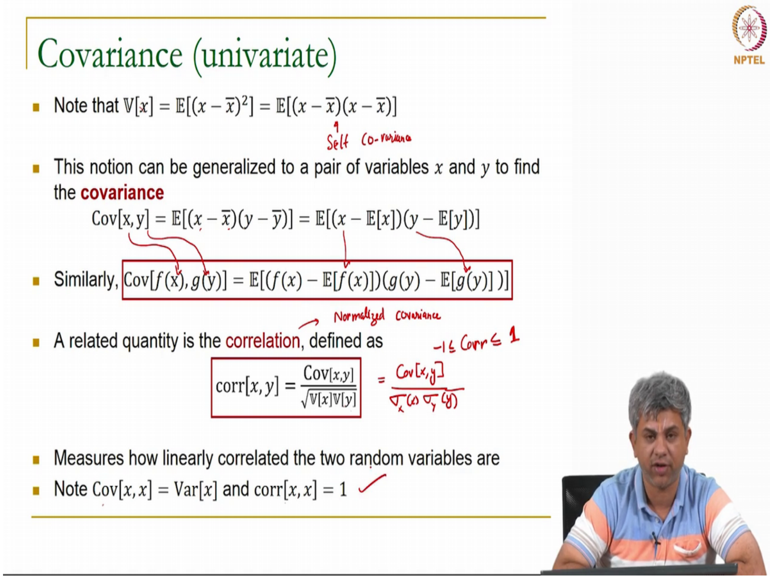


What does correlation indicate? Correlation indicates how linearly correlated, the word linearly will become clear in the next slide, how linearly correlated the two variables, the two random variables are.

For example let us say x is height and y is weight, Ok. What you expect is, as height increases weight also increases. Of course there is randomness here. There are tall people who can be very, very, very slim and there can be short people who could actually have higher weight than the taller person but nonetheless you can see that overall the trend will be that as x increases y also increases, Ok.

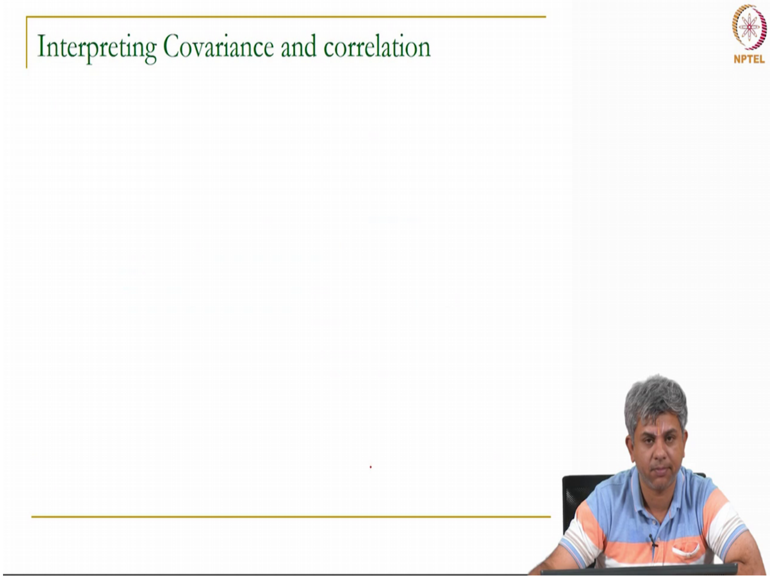
Notice that by definition, covariance of x comma x, Ok which is simply expectation of x minus x bar, x minus x bar it will simply give you variance of x, that was the way it was designed. And if I find the correlation of x and x, it makes complete sense that this is 1, Ok.
So you can see that since the numerator itself is variance of x, square root of variance of x multiplied by variance of x you will get correlation of

(Refer Slide Time: 10:34)



x comma x as 1,

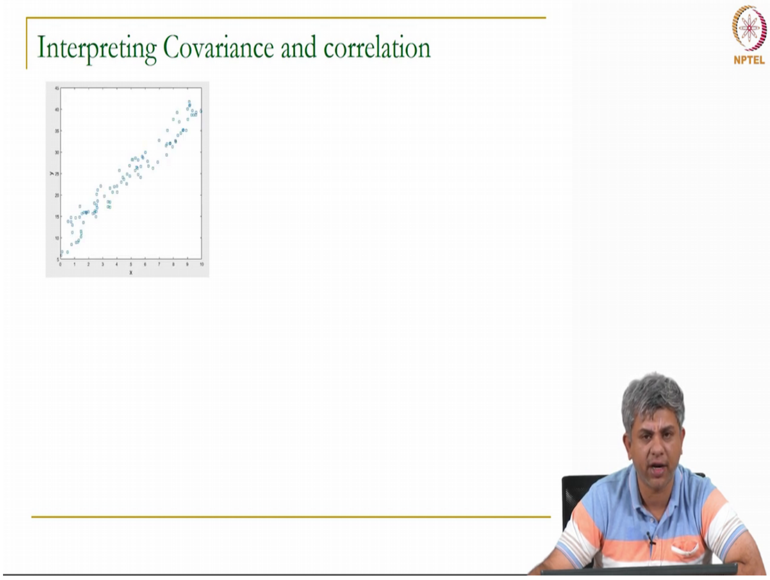
(Refer Slide Time: 10:36)



Ok

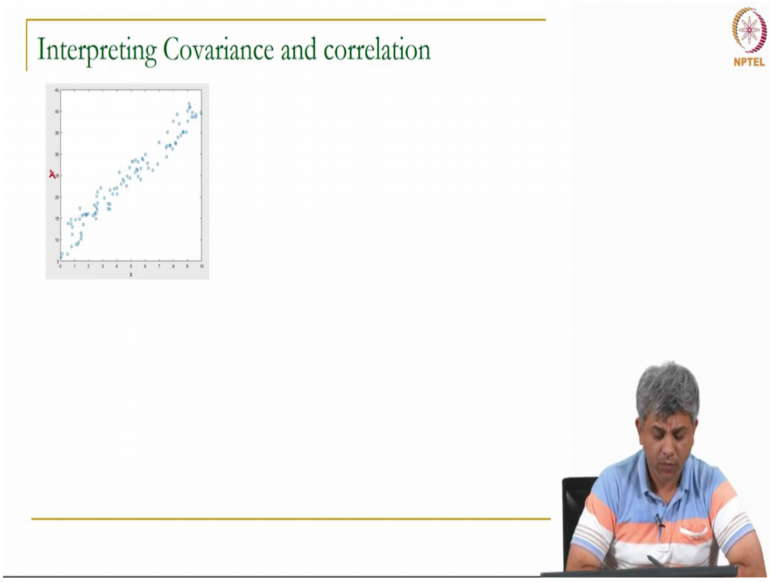
So let us now in this slide look at an interpretation of what covariance and correlation mean. I have already given you a slight

(Refer Slide Time: 10:47)



indication. So let us say this direction here is x and here it is y.

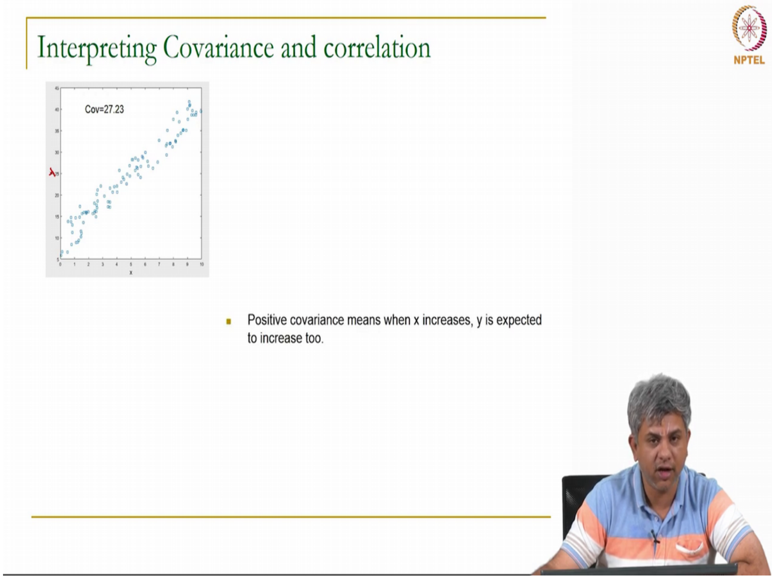
(Refer Slide Time: 10:55)



And let us say x and y are random variables. And you see this kind of variation amongst them, Ok.

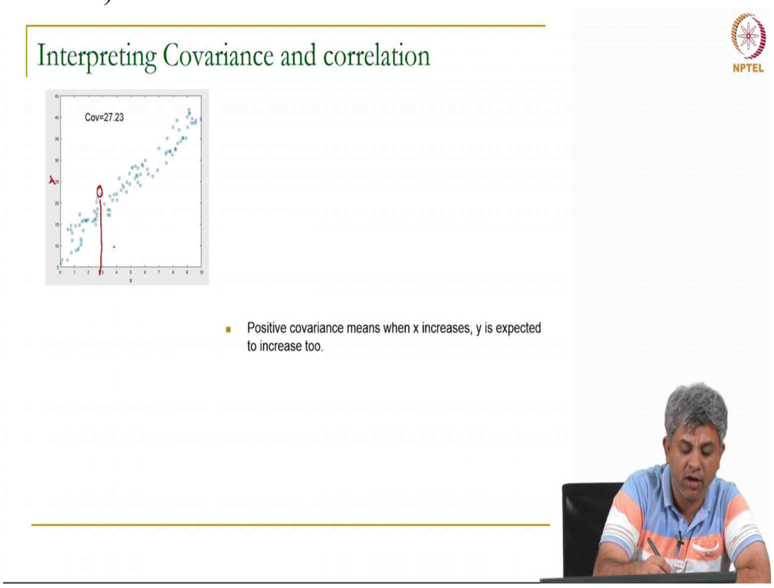
You can see that even though both are random, as x increases, y also increases. That is at least the overall trend, Ok. Now if I calculate the covariance in this case it comes to some 27 point something.

(Refer Slide Time: 11:19)



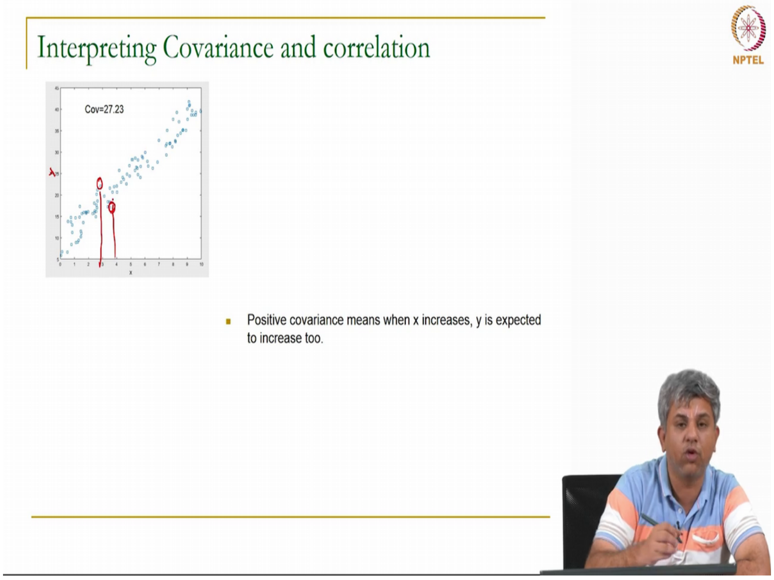
Now the general rule of, not rule of thumb, the general rule is that if you have positive covariance then it means that as x increases y is also expected to increase, even though in a certain few cases it can happen that, you know at this given x, you have a y

(Refer Slide Time: 11:36)



and at a slightly higher x, you have lower

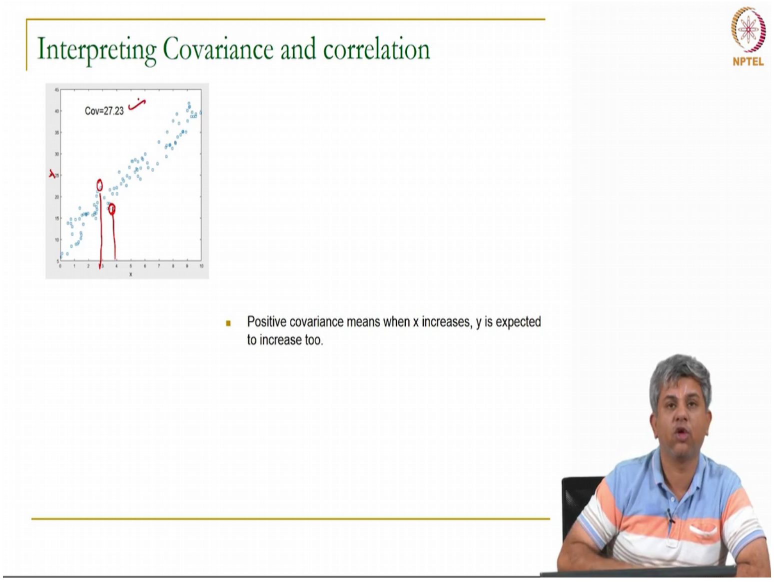
(Refer Slide Time: 11:39)



y.

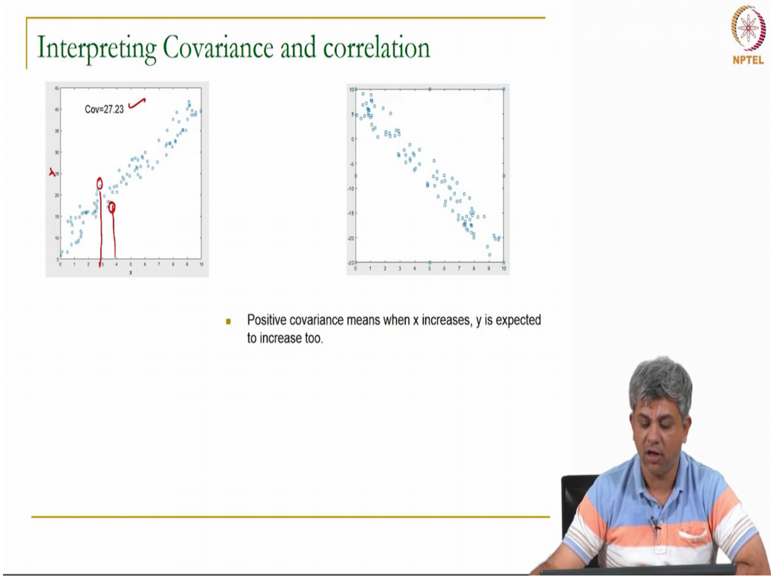
Even though that expectation could be violated, overall the trend is that as x increases, y increases. And that is what is indicated by

(Refer Slide Time: 11:50)



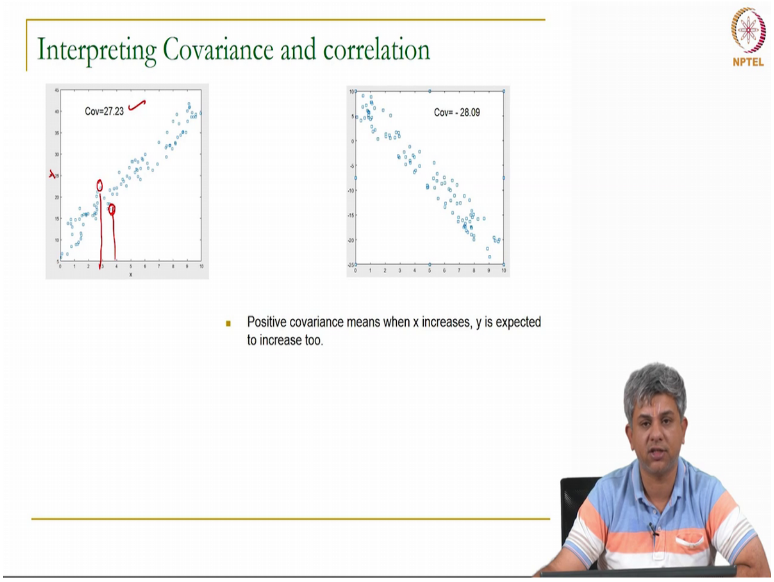
covariance, the value of covariance, Ok or the sign of covariance.

(Refer Slide Time: 11:54)



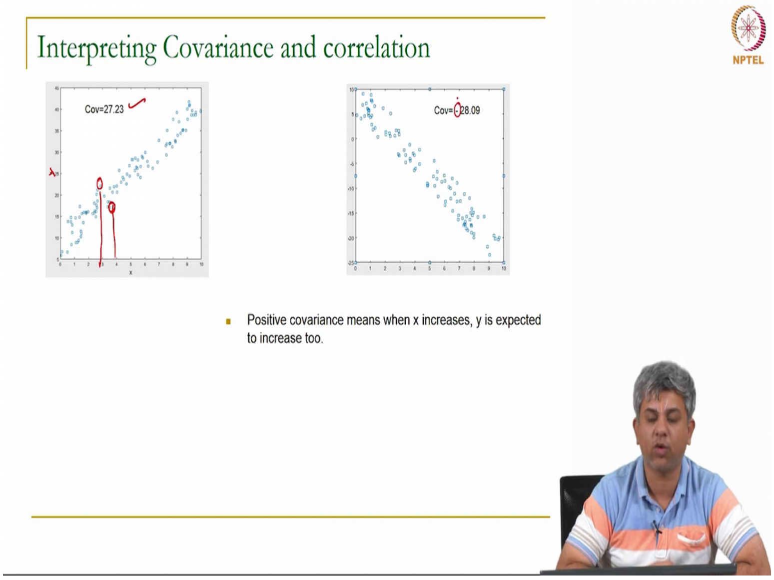
Now you can see a counter case. In this case what you notice is as x increases, y decreases.

(Refer Slide Time: 12:01)



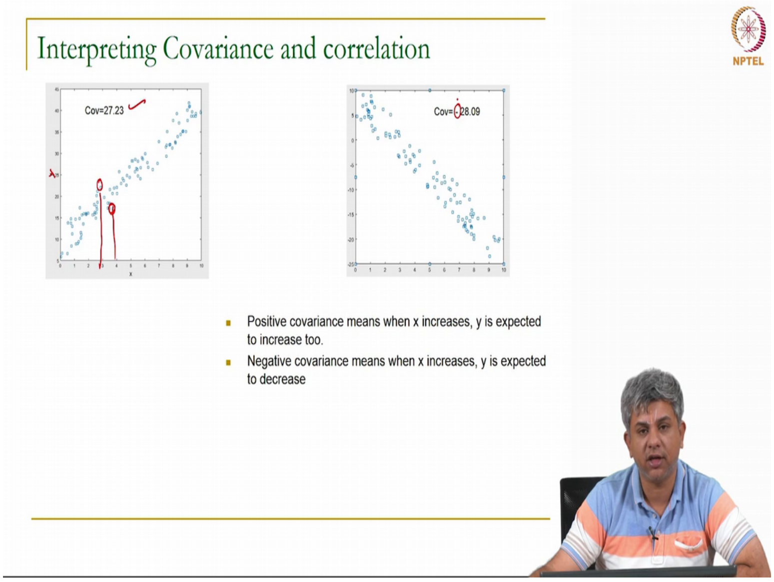
And sure enough the covariance in this case is negative, Ok. It is negative

(Refer Slide Time: 12:07)



28 point 0 9. And what negative covariance

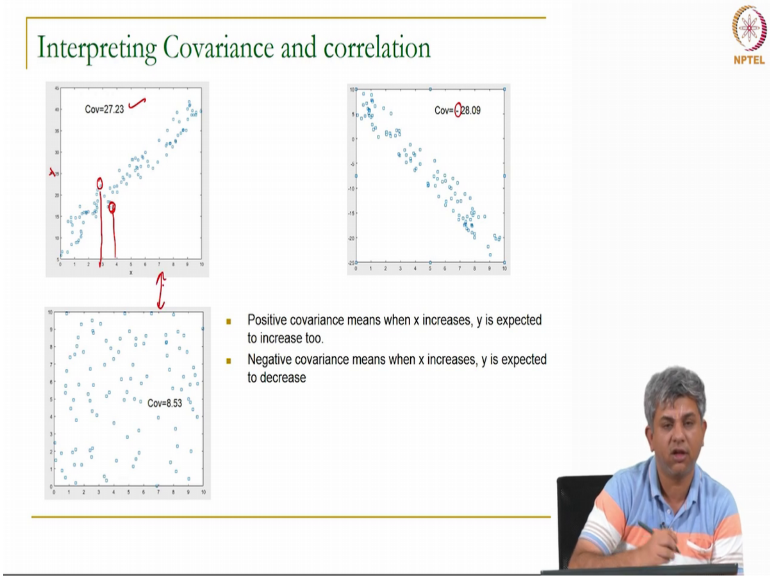
(Refer Slide Time: 12:11)



indicates is that as x increases, y actually decreases.

Let us look at the third case. You will see that you really cannot say anything about any trend. It seems like x and y have no relation whatsoever. If you find out the covariance in this case, it still gives the positive value. It is 8 point 5, so 8 point 5 3. But we would like to qualitatively distinguish between these two cases,

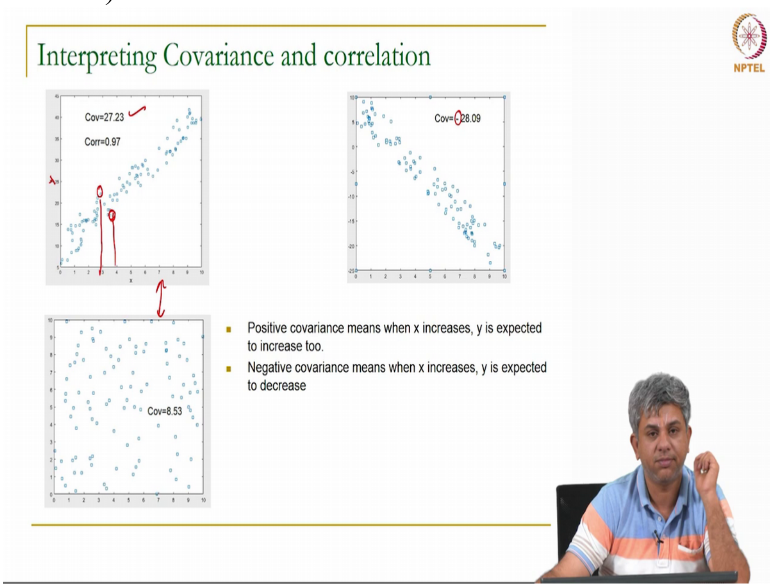
(Refer Slide Time: 12:41)



Ok.

Here covariance is positive but you see a clear trend. Here covariance is positive even though it is a smaller value, we can notice that it is a smaller value but it is all over the place, Ok. So can we relate these two? It turns out we can by looking at

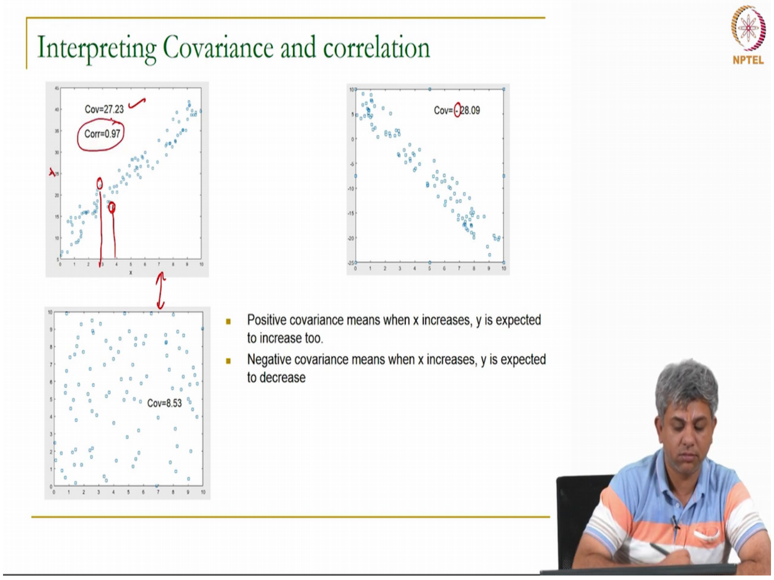
(Refer Slide Time: 12:59)



the correlation.

So notice that the correlation in this case

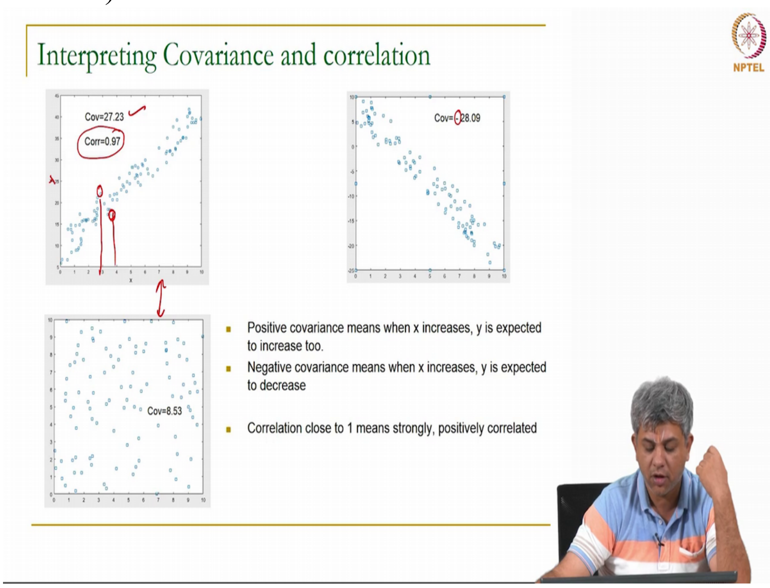
(Refer Slide Time: 13:04)



is point 9 7. This is what normalization means. Correlation is simply the numerator. It tells you what the trend is. But how positive is positive? We will not know unless we compare it with something and the comparison metric here is the variance of these two quantities.

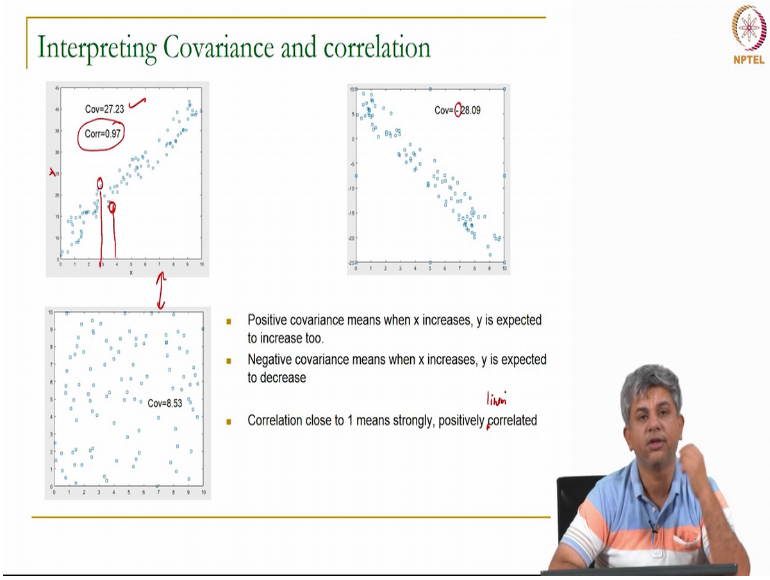
So if you find out the correlation, that is normalize the covariance divided by sigma x sigma y, then you get the correlation of point 9 7 which is very, very close to 1. Remember that if I had simply taken a linear relationship x versus x I would have got 1.

(Refer Slide Time: 13:39)



So correlation which is close to 1 means that there is a strong positively

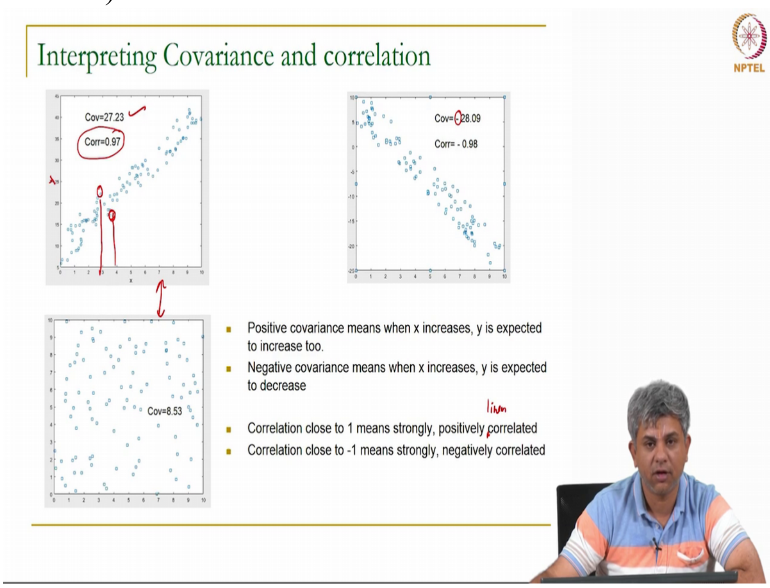
(Refer Slide Time: 13:52)



correlated, linearly positively correlated, Ok strong positively linear correlation between the two variables.

Similarly if I find out the correlation in this case, I find out that it is minus point 9 8.

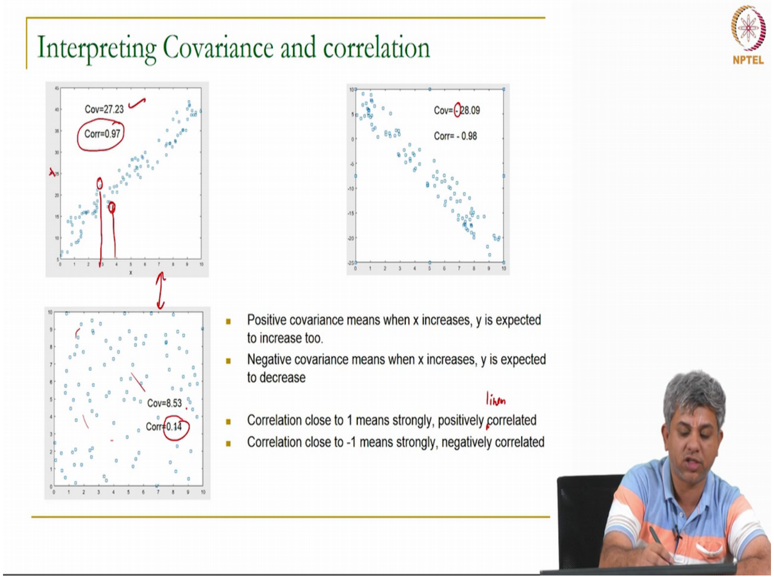
(Refer Slide Time: 14:02)



And it is strongly negatively correlated linearly. Now what we would expect is, in a case of this sort our correlation should be very low, Ok even though covariance is positive.

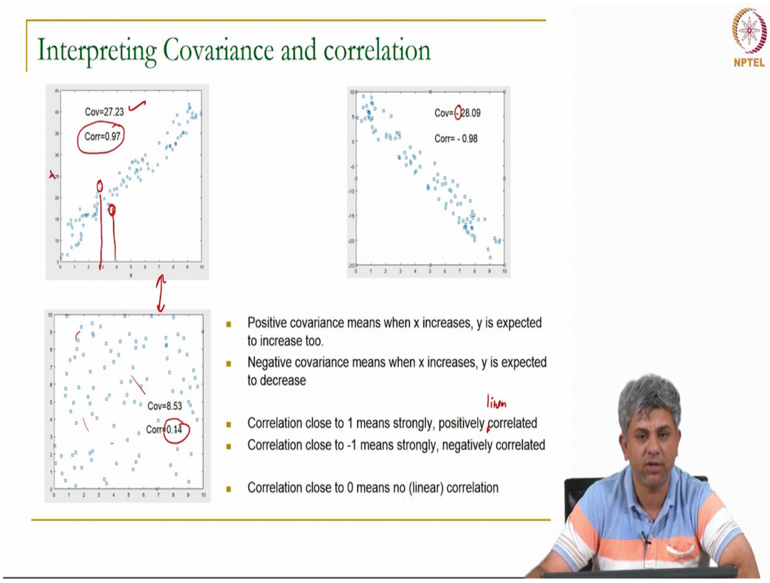
And indeed this is the case. So correlation in this case is point 1 4 which is

(Refer Slide Time: 14:20)



much smaller in comparison to point 9 7. In fact typical rule of thumb, if you take correlation below point 3 or even point 5, you might as well assume that the variables are not correlated. Of course the lower the correlation gets, the lesser actually the relationship between the two variables,

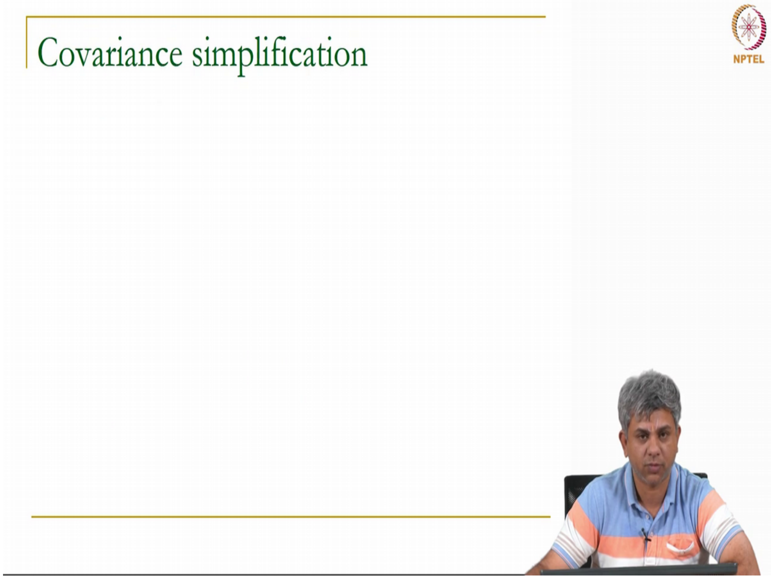
(Refer Slide Time: 14:37)



Ok.

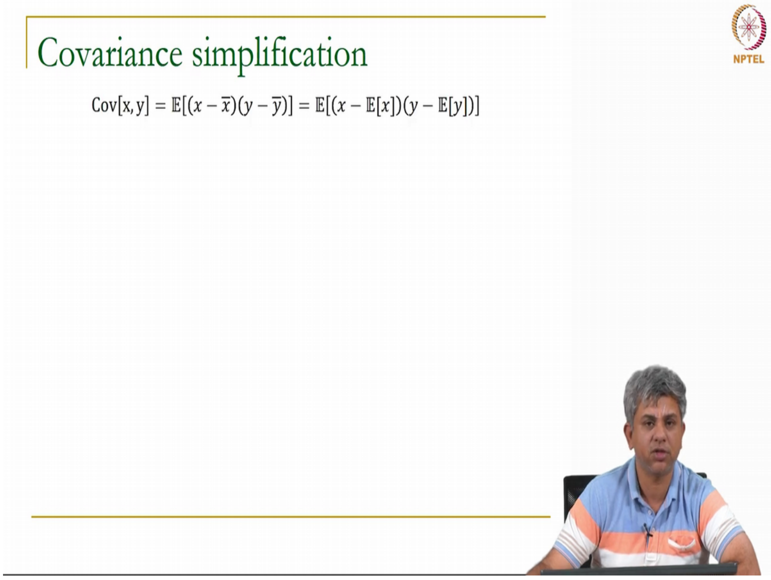
So correlation which is close to 0 means there is no real correlation between the two variables.

(Refer Slide Time: 14:45)



So let us look at some mathematical simplification of covariance which we will be using

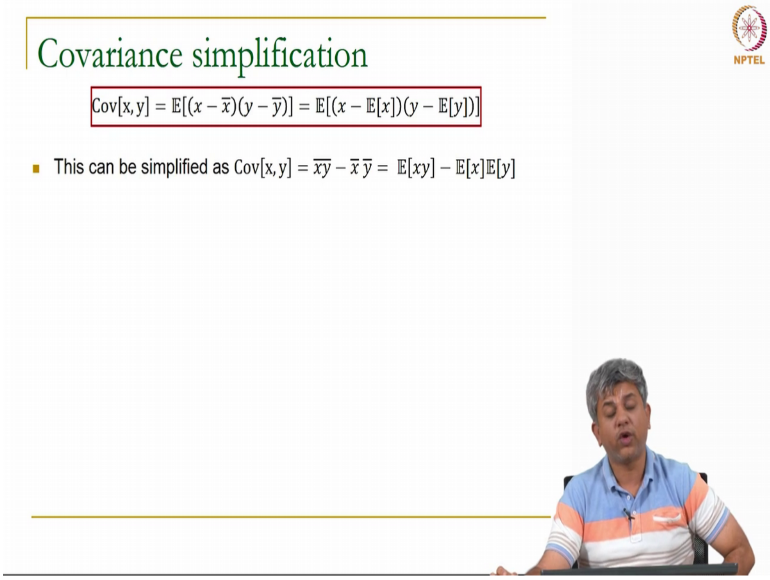
(Refer Slide Time: 14:51)



multiple times through this course.

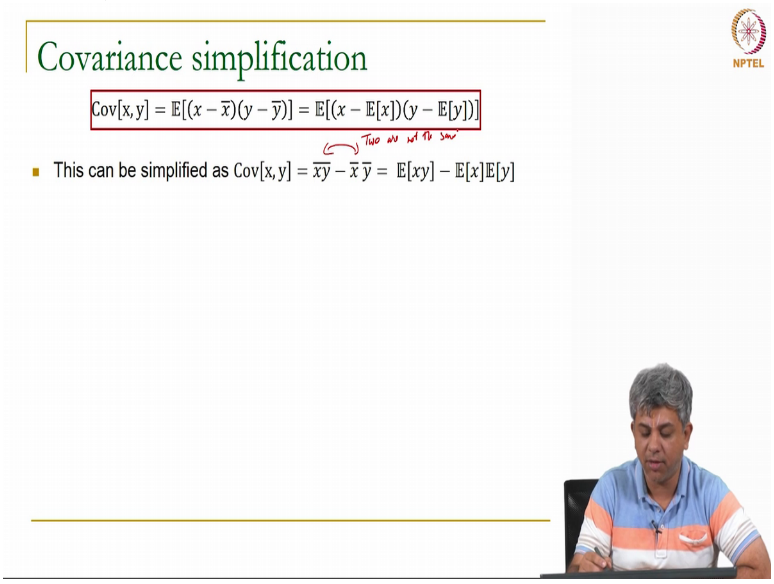
So recall that covariance of x comma y is expectation of x minus x bar multiplied by y minus y bar. You can

(Refer Slide Time: 15:03)



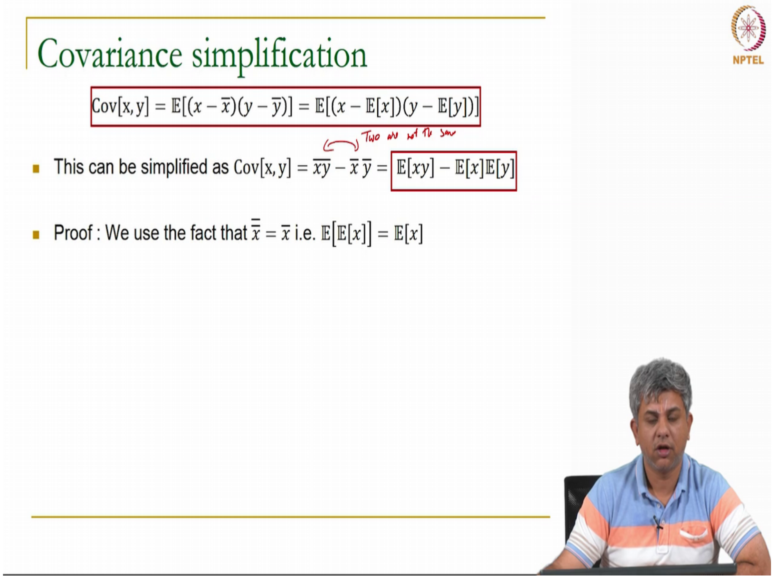
simplify this. This is usually useful for calculations, that is covariance of x y is x y bar minus x bar y bar. Notice that the two are not the same.

(Refer Slide Time: 15:21)



In terms of expectation notation, expectation of x y minus expectation of x expectation of y, Ok.

(Refer Slide Time: 15:30)

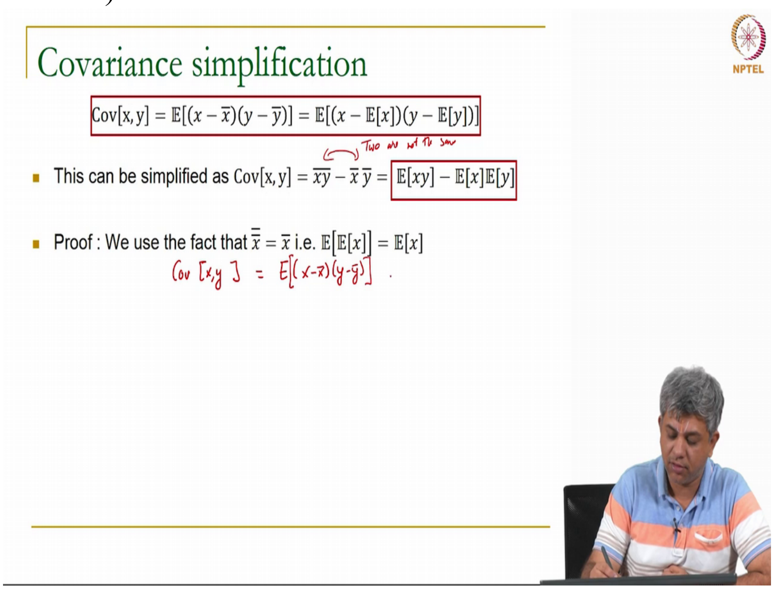


So let me quickly prove this. We will use the fact that the expectation of expectation or the average of an average is simply the same as the average because the average is one single value, Ok. At least it is useful to think of it that way.

So if I have already taken average of 10 values I have got one simple number. If I take average above that, it does not make any difference whatsoever. Ok so let us use this.

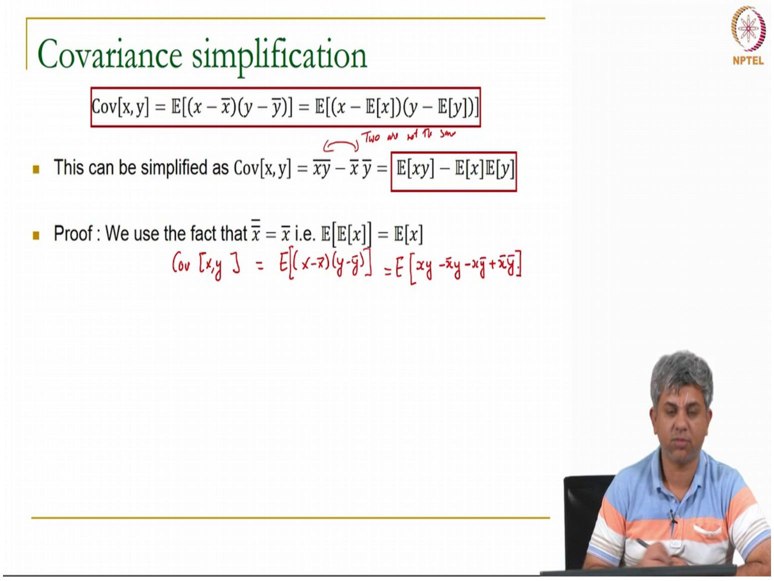
So remember that covariance of x y is expectation of x minus x bar y minus y bar;

(Refer Slide Time: 16:12)



this is expectation of, let us open this bracket up, x y minus x bar y minus x y bar plus x bar y bar,

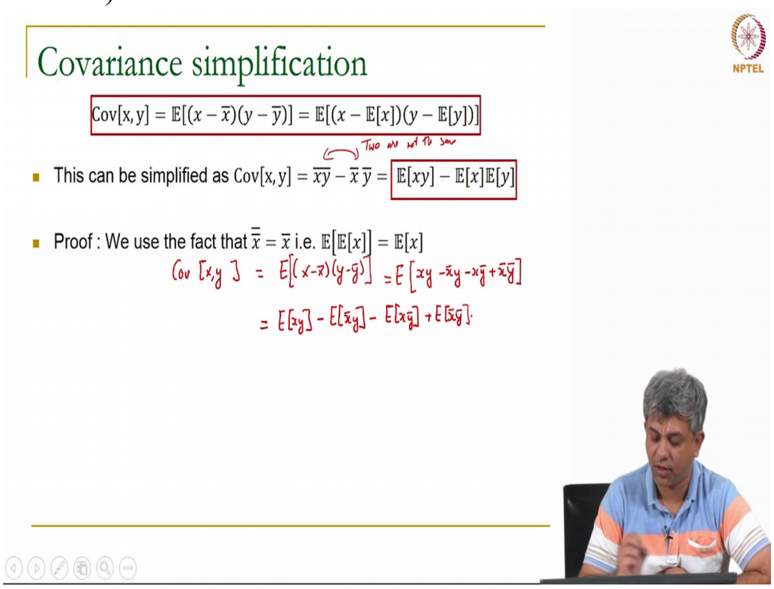
(Refer Slide Time: 16:27)



we can read each of these quantities separately.

Expectation of x y minus expectation of x bar y minus expectation of x y bar plus expectation of x bar y bar,

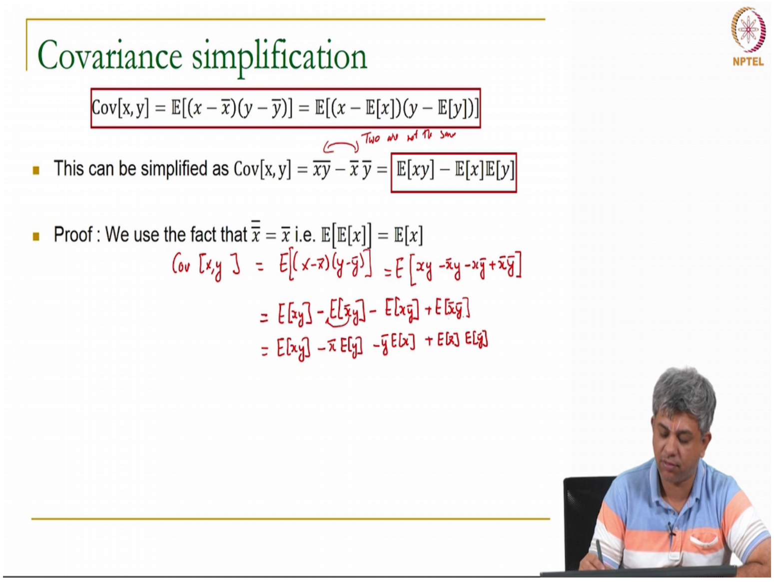
(Refer Slide Time: 16:52)



Ok. So the first quantity is simply expectation of x y. We cannot do anything with it. Now in this quantity x bar is a constant, it can be brought out, so you can write this as x bar times expectation of y.

Here you can bring out y bar, y bar times expectation of x Ok and this is, these are already average quantities so you can write it as expectation of x bar,

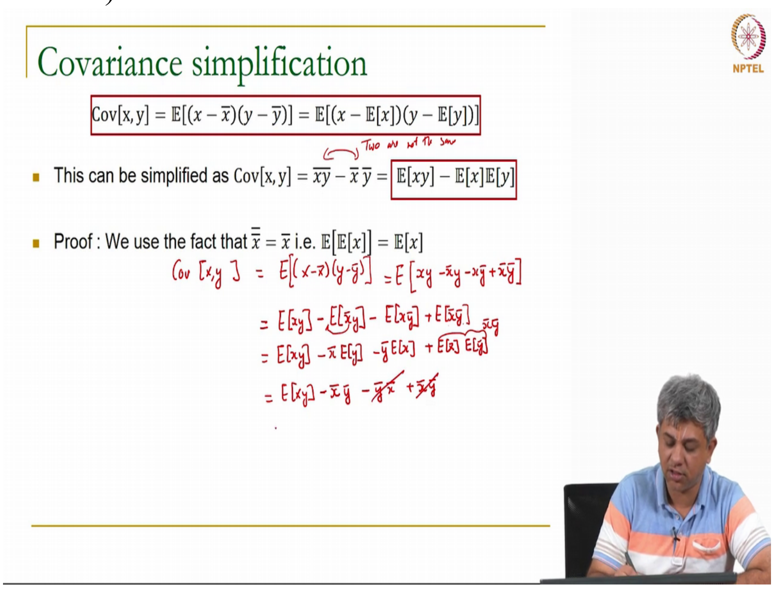
(Refer Slide Time: 17:24)



simply x bar y bar, Ok. Expectation above and beyond this does not make sense.

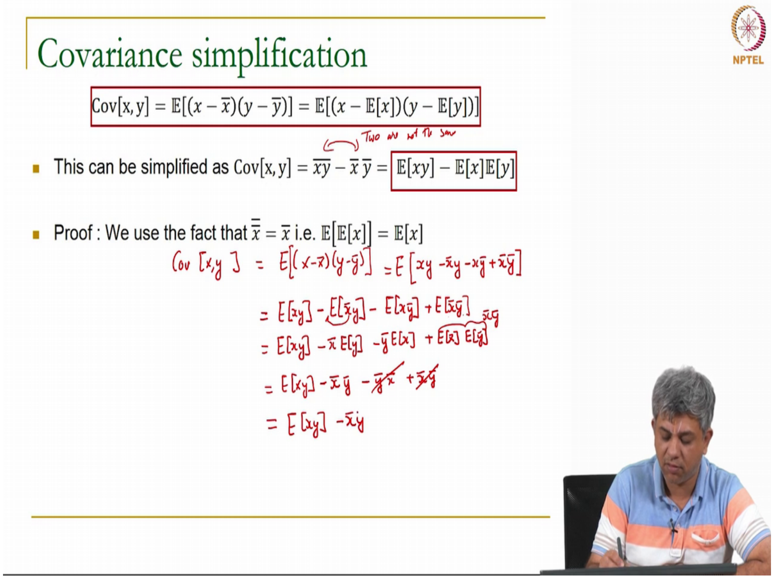
So I will write this as expectation of x y minus x bar, expectation of y is y bar, minus y bar, expectation of x is x bar plus x bar y bar. These cancel

(Refer Slide Time: 17:48)

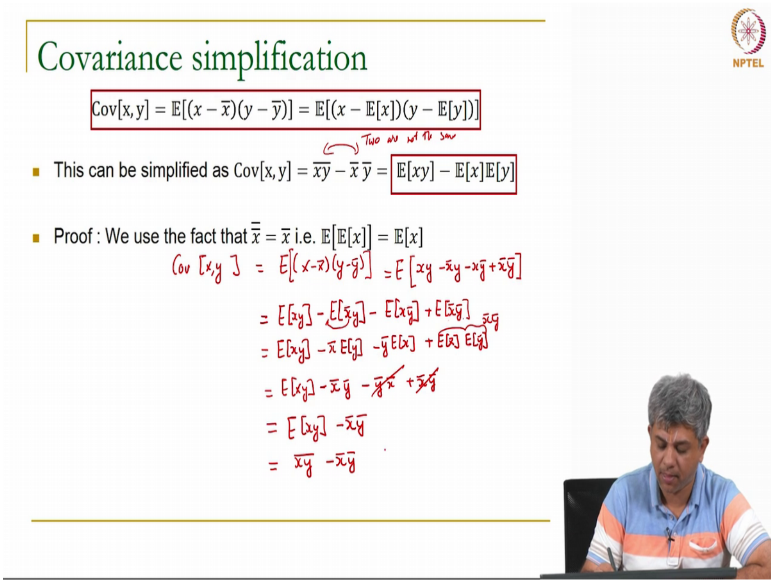


so you can write this as expectation of x y,

(Refer Slide Time: 17:55)

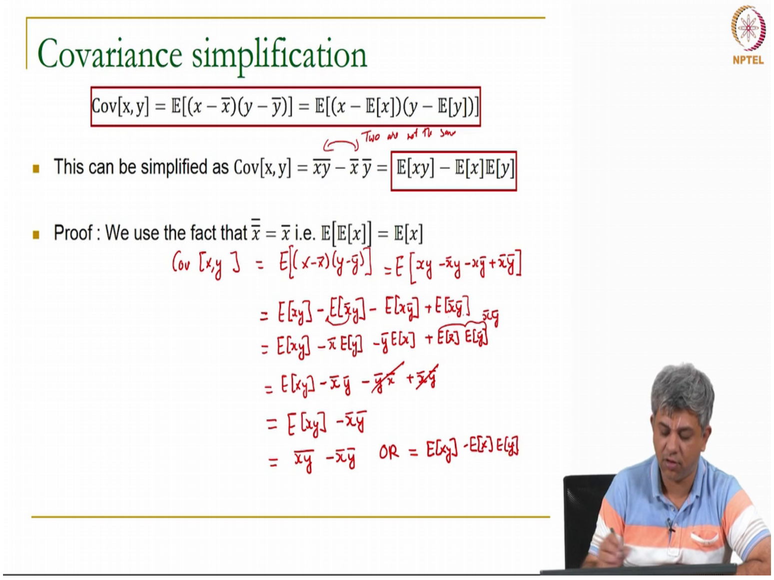


(Refer Slide Time: 18:00)



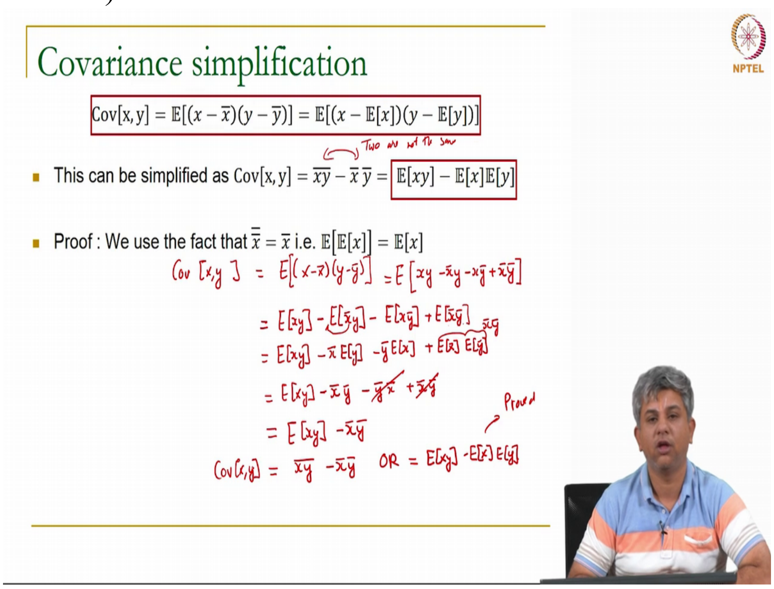
expectation of x y minus expectation of x times expectation of y.

(Refer Slide Time: 18:12)



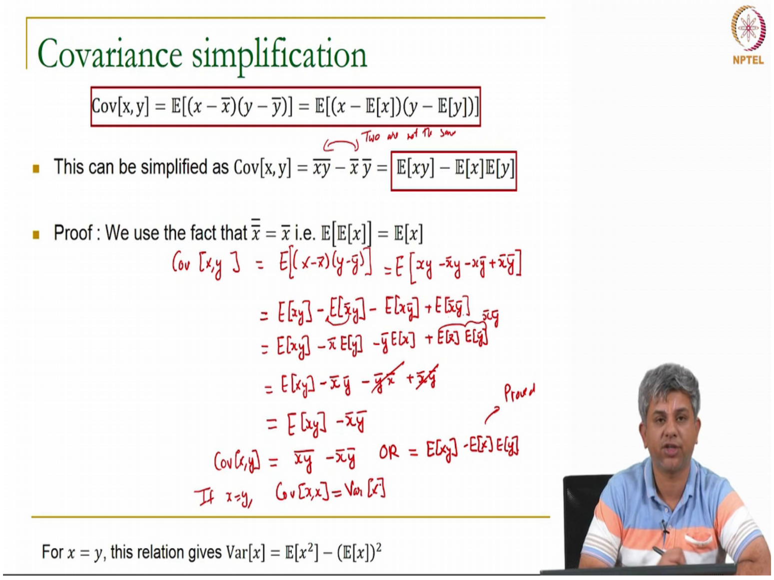
So we have proved the result, Ok.

(Refer Slide Time: 18:20)



There is a special case of this which we tend to use, which is if you said, if x equal to y then we have covariance of x comma x, remember covariance of x comma x is simply, by

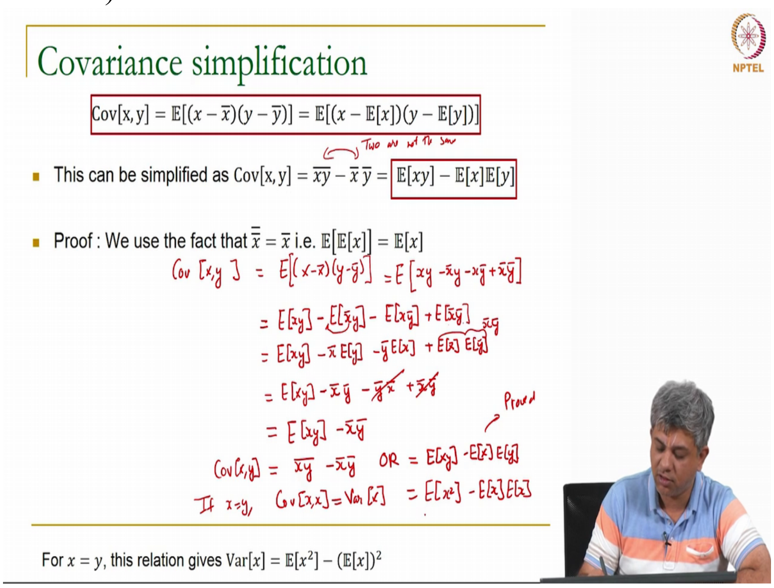
(Refer Slide Time: 18:38)



definition variance of x.

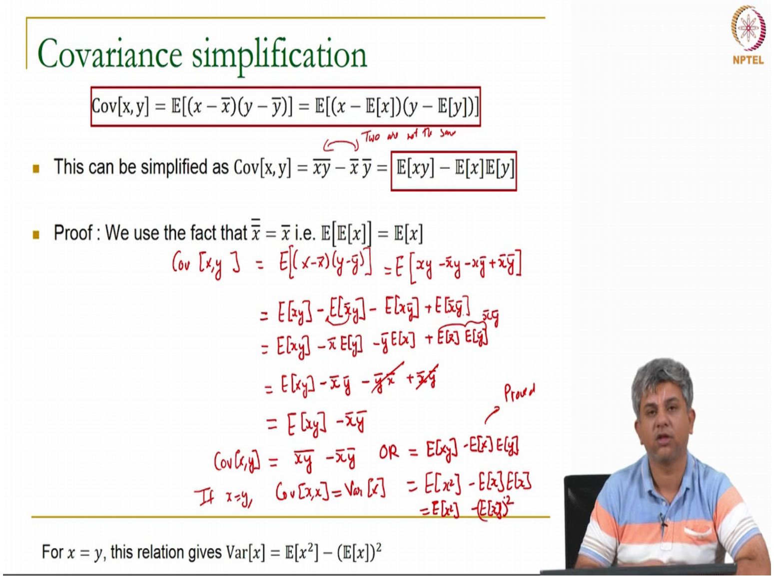
This now becomes expectation of x square minus expectation of x multiplied by expectation of x

(Refer Slide Time: 18:48)



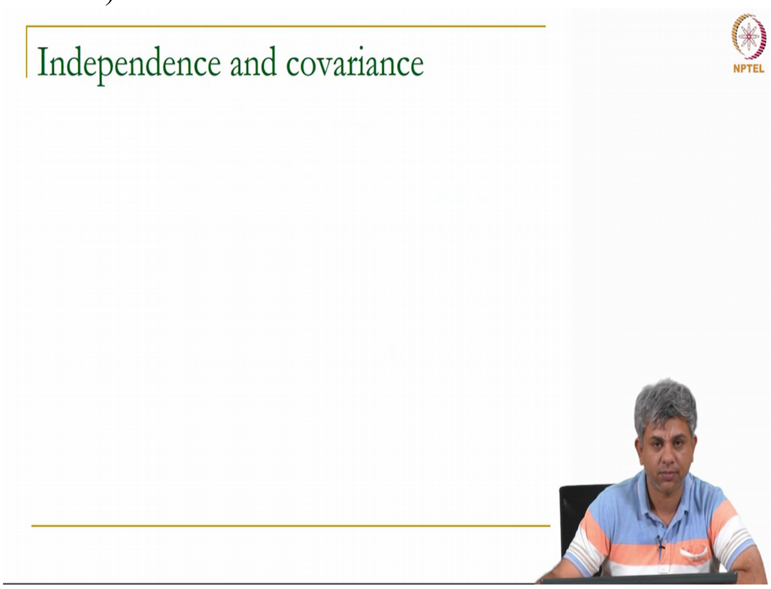
which is expectation of x square minus expectation of x

(Refer Slide Time: 18:57)



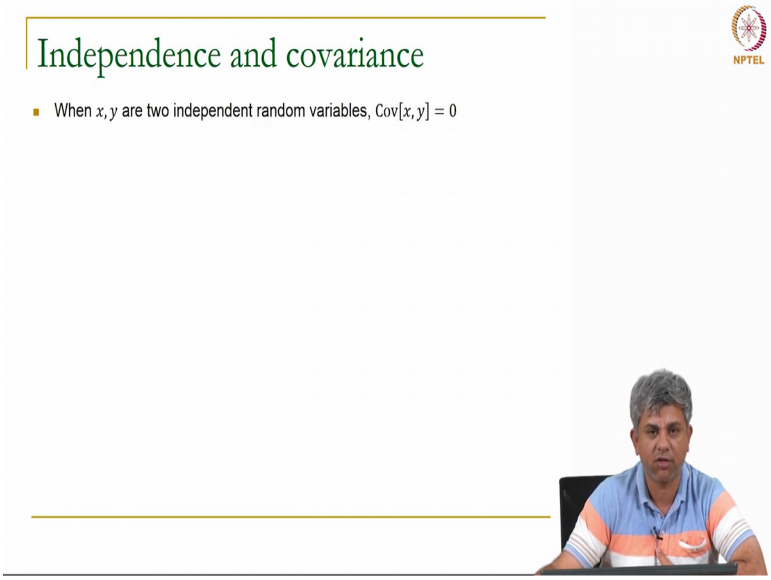
the whole square, Ok. So this is also a definition that we are, or equivalence that we will often use.

(Refer Slide Time: 19:06)



Let us come now; there is a certain point of covariance and independence that I wish to discuss.

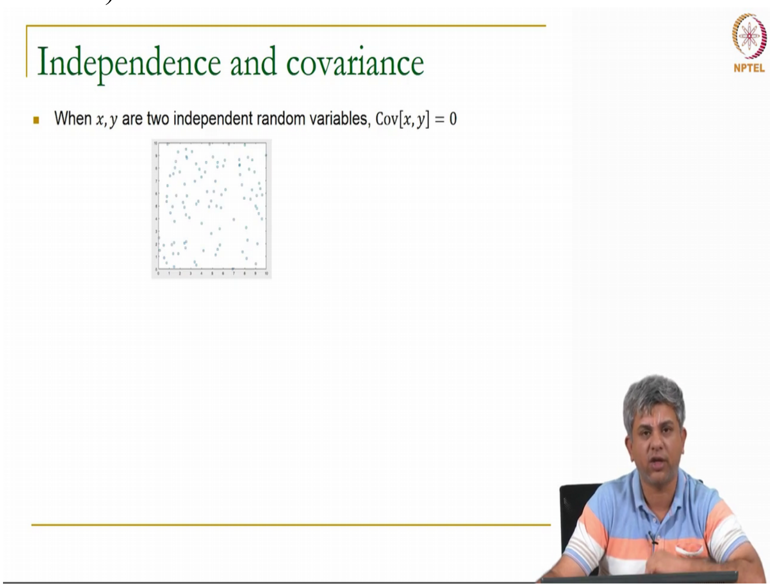
(Refer Slide Time: 19:13)



So we already saw that when x and y are two independent random variables then covariance of x and y is actually 0. So if x varies in one way and y varies in a completely independent way you will actually get covariance of x comma y is equal to 0.

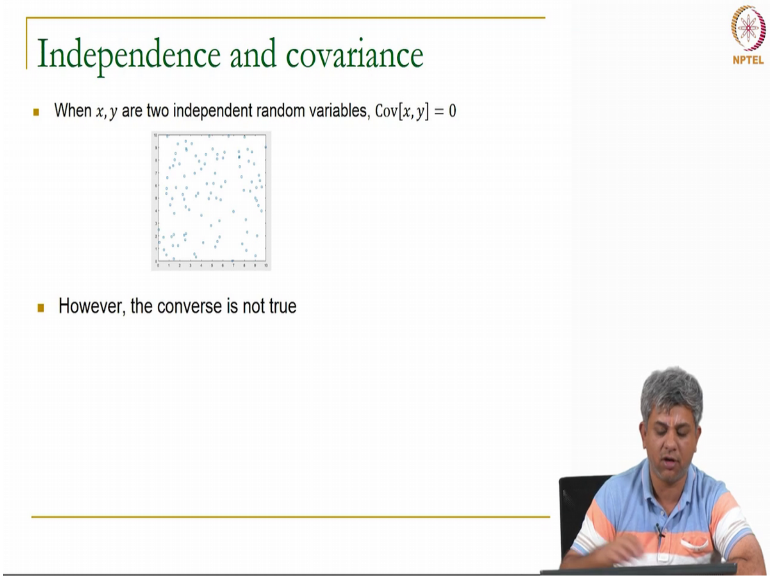
You can actually prove it from the previous slides results also

(Refer Slide Time: 19:33)



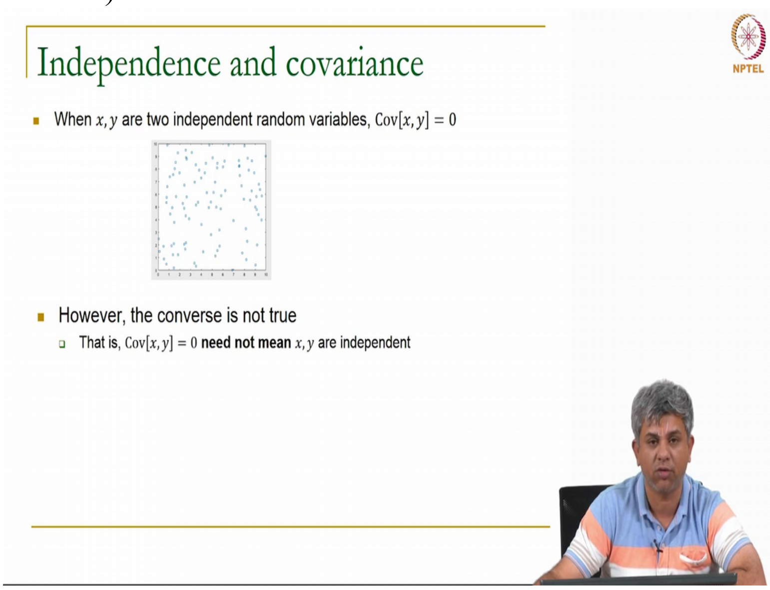
but I will not do that. I will just appeal to your intuition, Ok. So if you have,

(Refer Slide Time: 19:37)



you know sort of random spread of points of this sort, then covariance of x comma y is 0. However if you want to go the other way, Ok which is you want to say that if covariance is 0, x and y are independent

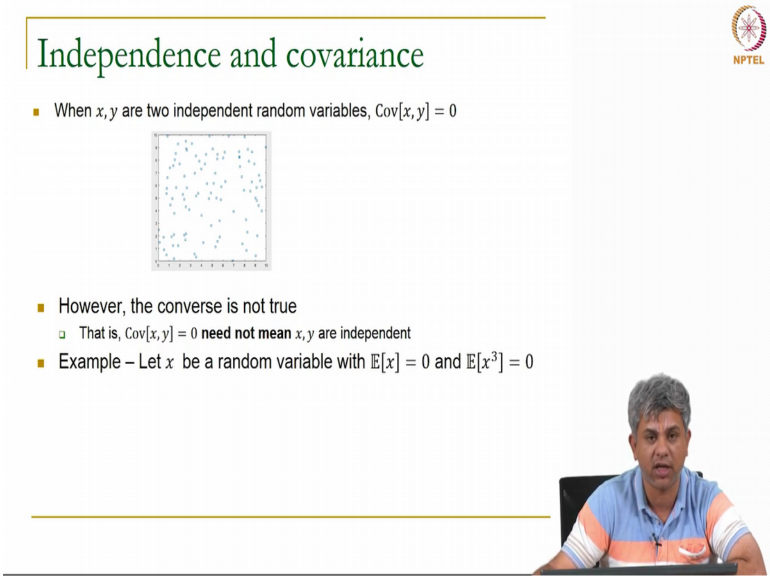
(Refer Slide Time: 19:52)



this is not true, Ok. So covariance of x being, x comma y being 0 need not necessarily mean that x and y are independent of each other.

Let us see

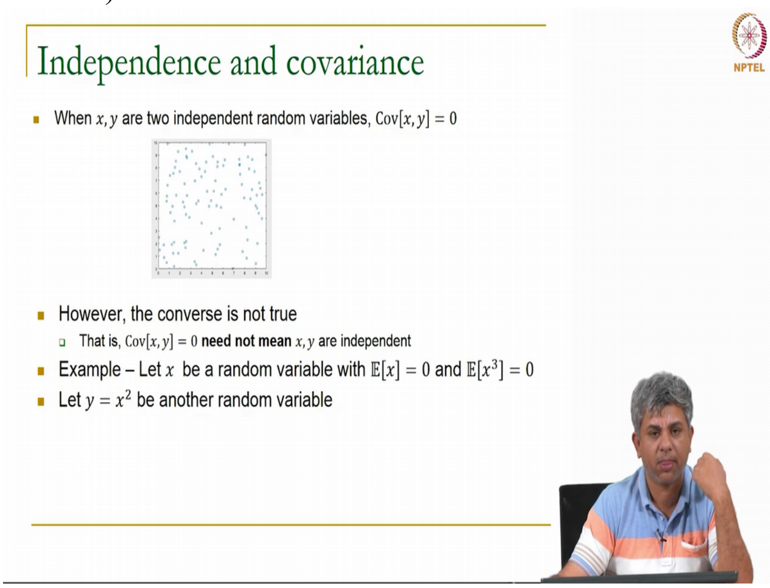
(Refer Slide Time: 20:02)



an example, ok. So let us say x is a random variable with expectation of x is 0 and also expectation of x cube being 0, Ok. This need not always necessarily be the case, but you can easily generate a set of random numbers for which this is true, Ok.

For example if I take x is a random variable, a uniform random variable with values going from minus 10 to 10 these are uniformly spaced and all values are equally possible then expectation of x will be 0, expectation of x cube will also be 0 as you can quite easily see,

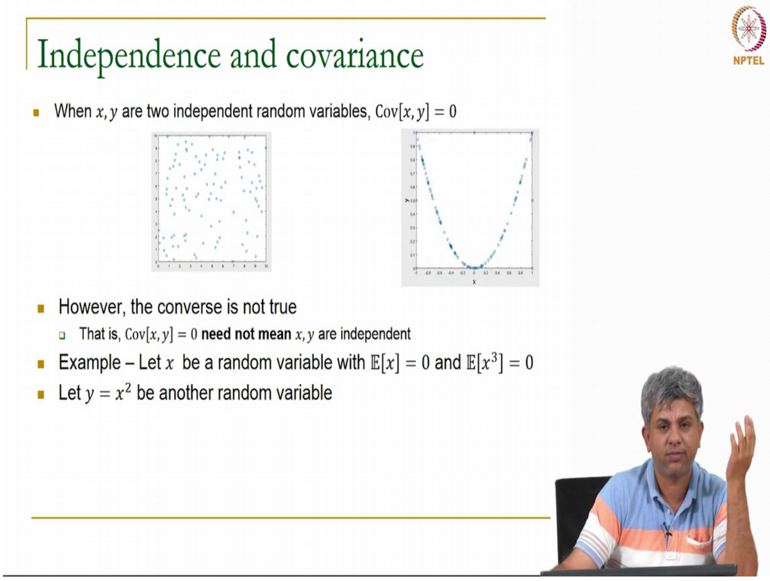
(Refer Slide Time: 20:38)



Ok.

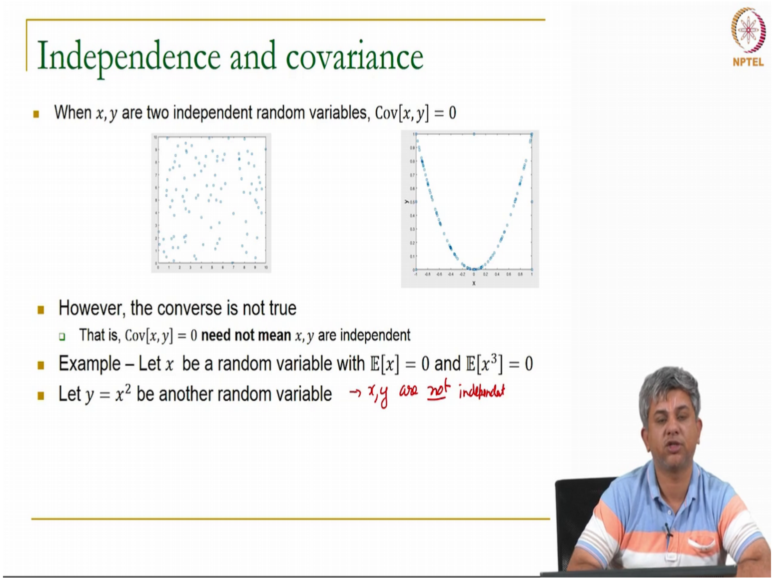
Now let us take y as x square, Ok. Now y is another random variable.

(Refer Slide Time: 20:45)



So if x is a random variable and y is simply x square you know you will find the distribution of x and y like this, Ok. Now clearly x and y are not independent, Ok. Obviously the value of y

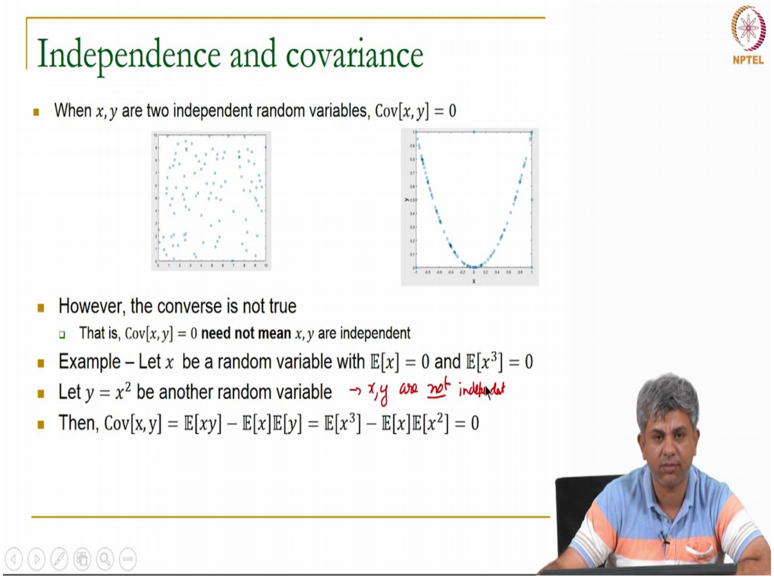
(Refer Slide Time: 21:04)



depends on the value of x.

However if we try and find out

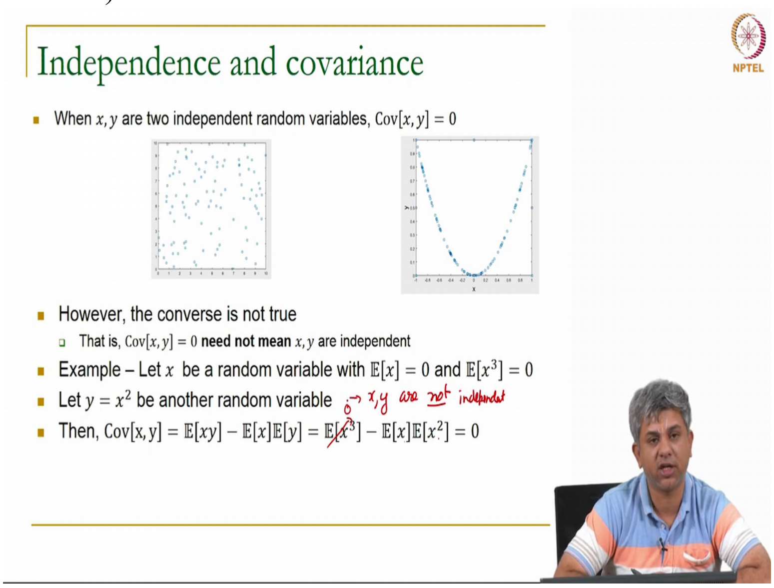
(Refer Slide Time: 21:09)



covariance of x comma y, the expression that we had in the previous slide was the expectation of x y minus expectation of x multiplied by expectation of y.

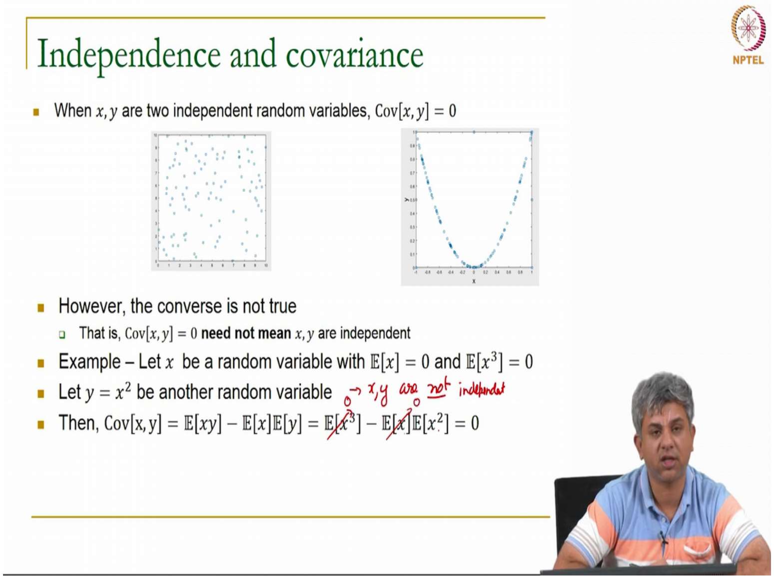
Now this is expectation of x cube minus expectation of x, expectation of x square. By construction

(Refer Slide Time: 21:27)



expectation of x cube was 0,

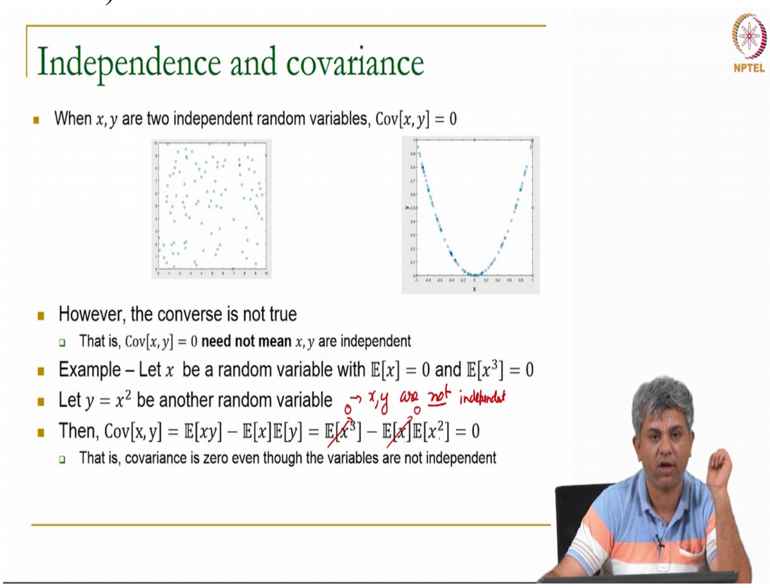
(Refer Slide Time: 21:30)



again by construction expectation of x was 0. So the covariance actually is 0.

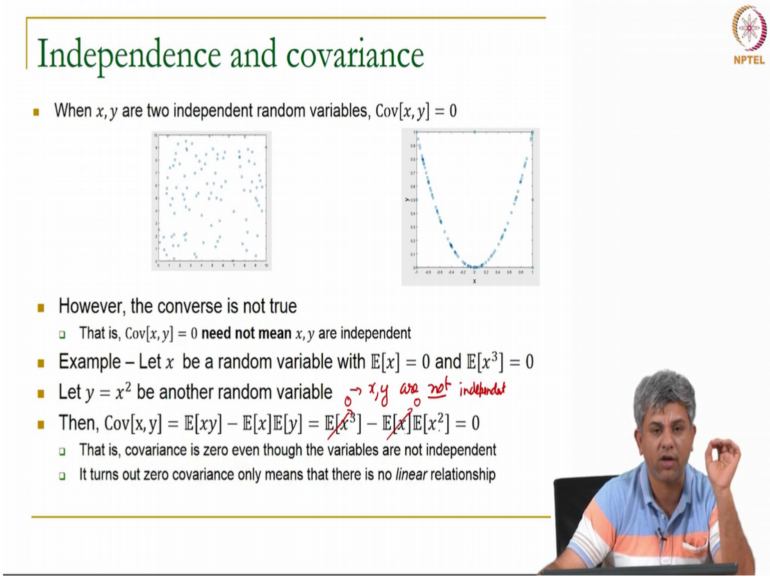
So for this case even though you can see a nice relationship between the two variables, covariance is actually 0,

(Refer Slide Time: 21:42)



Ok. So that is, the covariance is 0 even though the variables are not independent of each other,

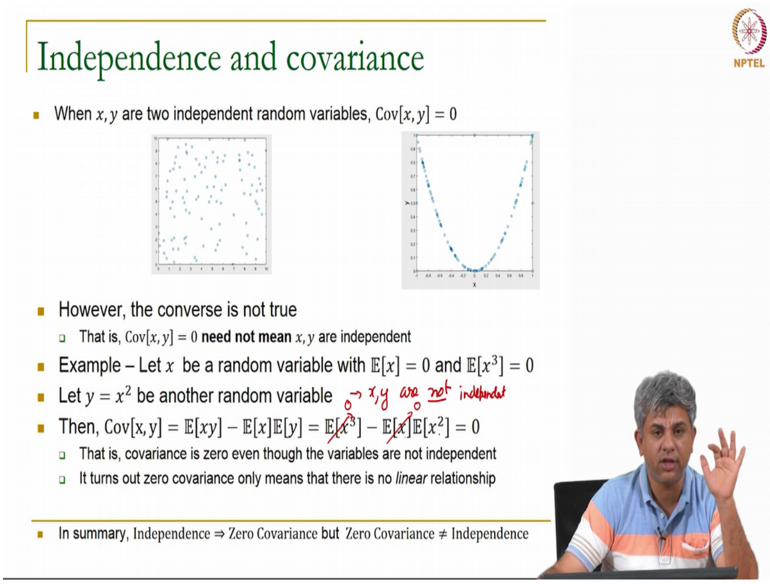
(Refer Slide Time: 21:49)



Ok.

So what turns out is covariance is 0 only means that there is no linear relationship, Ok. So

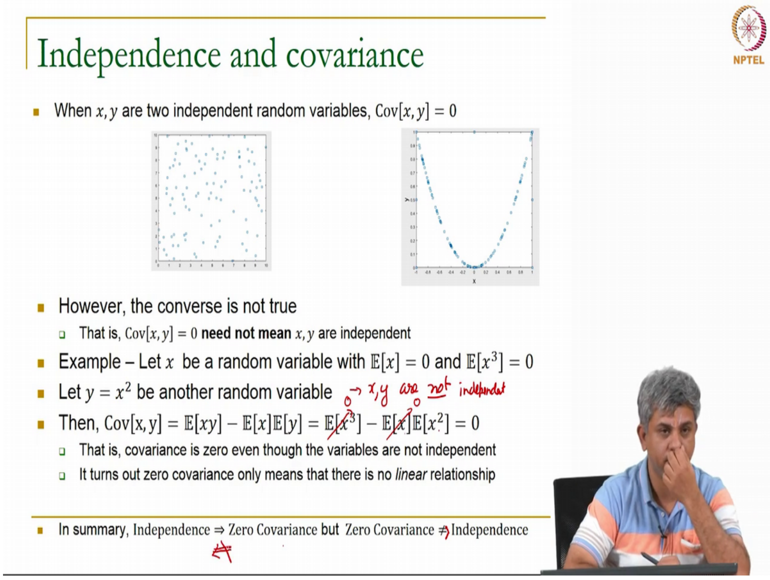
(Refer Slide Time: 22:01)



this can get a little bit confusing. So let me summarize this.

In case two variables are independent, Ok in this direction, so if independence is there, then there is zero covariance. But it does not necessarily mean that if there is zero covariance

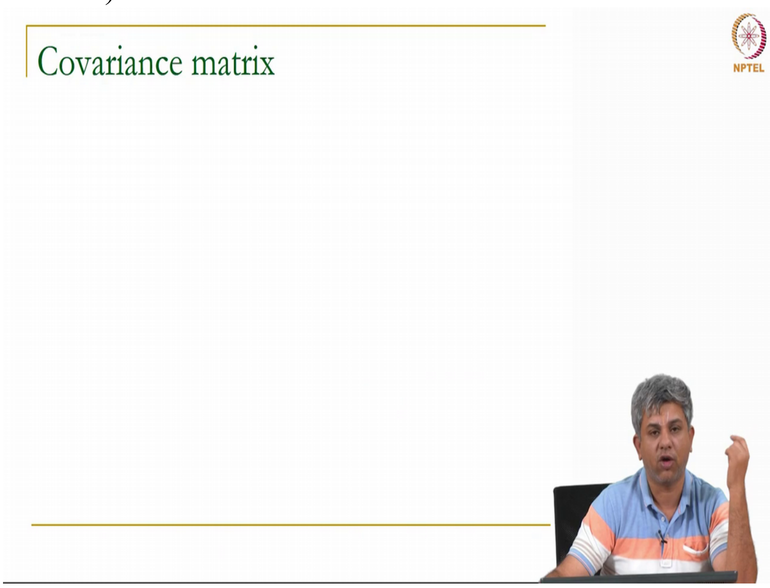
(Refer Slide Time: 22:22)



that there is independence.

If there is zero covariance, all you can say is that they are not linearly dependent on each other, Ok but you cannot talk about general independence, Ok. So covariance basically measures linear dependence of one

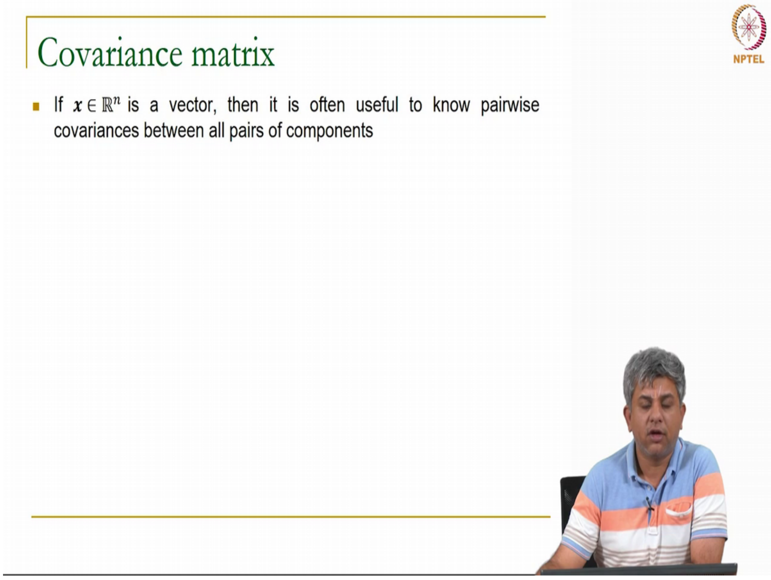
(Refer Slide Time: 22:38)



quantity on the other.

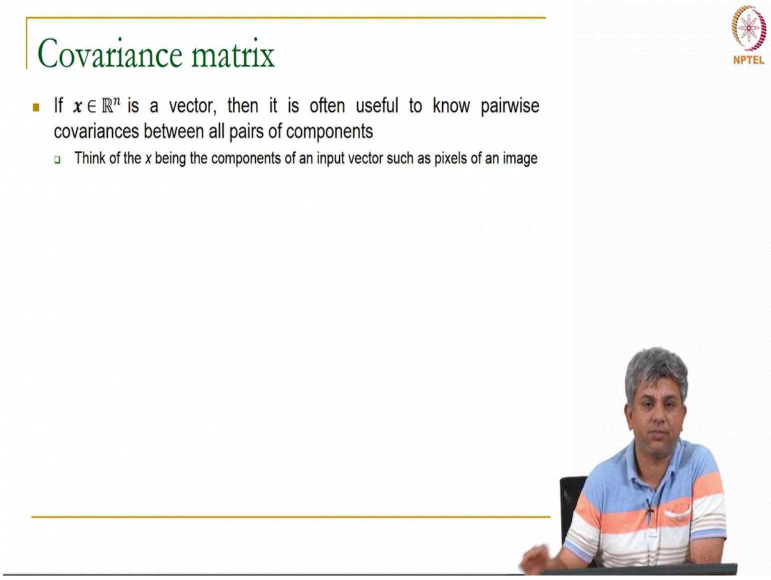
The final idea that I would like to discuss in this video is that of a covariance matrix. So, so far we have been looking at univariate x.

(Refer Slide Time: 22:51)



Now let us consider case such as an image where x is a vector,

(Refer Slide Time: 22:56)

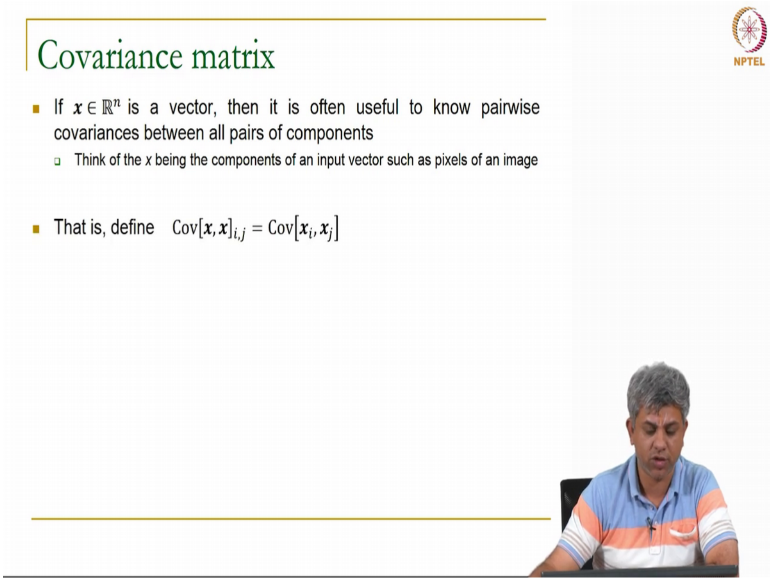


Ok. It could be not just an image, of course I will keep on using that example because that is the most intuitive example and we will be using it very often in this course. But it could be anything, Ok.

So not only would you like to know, you know how each, what is the probability of each pixel is, but you would also like to know what is the joint probability of, let us say one pixel being white and the other pixel being black which is usually how you can characterize images.

Or if you look at an input vector such as temperature, pressure, humidity, you know what is the probability that the temperature is so much and the pressure is something else, the joint probability, stuff like that. And how much does temperature vary from the mean given that humidity has varied from its mean. So that is usually how the covariance matrix is used,

(Refer Slide Time: 23:50)

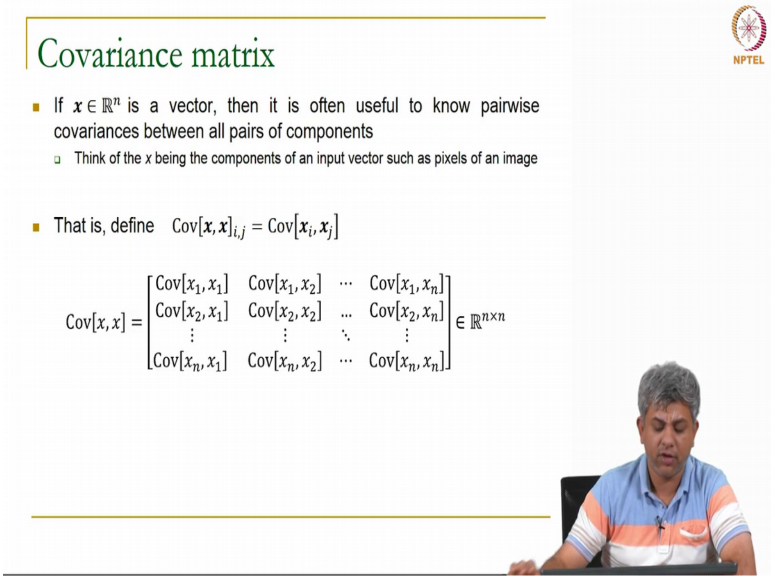


Ok.

So if x is a vector, Ok let us say belonging to R n. Let us say it has n components. So we define a pair wise correlation or pair wise covariance of each of the variables, temperature with pressure, temperature with humidity, pressure with humidity etc and find out and also find temperature with temperature which is simply a variance etc.

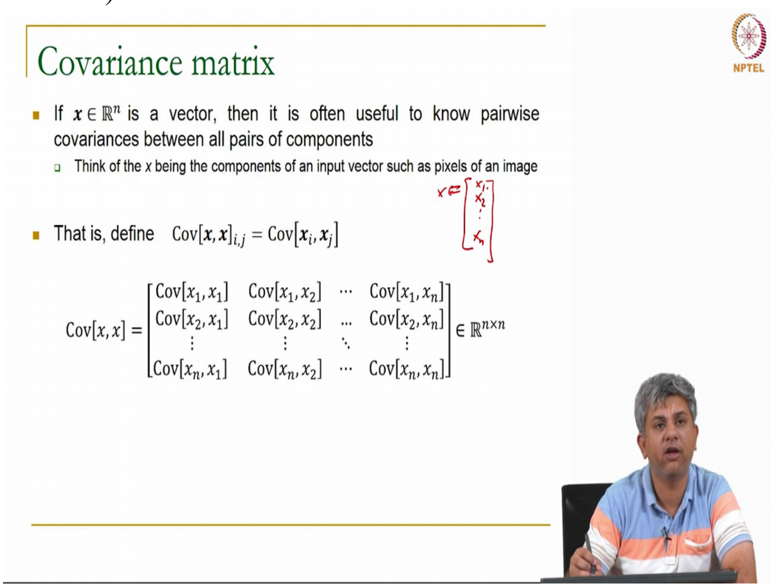
And that is how you define a covariance matrix.

(Refer Slide Time: 24:17)



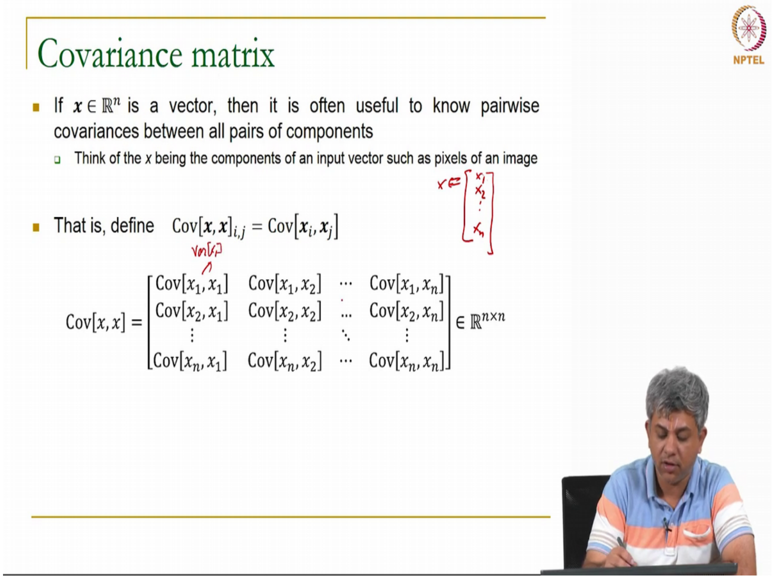
The covariance matrix is, even though it looks big, it is actually very simple to define. So if you have x which is belonging to R n, so let us say, x is x 1 through x n.

(Refer Slide Time: 24:34)



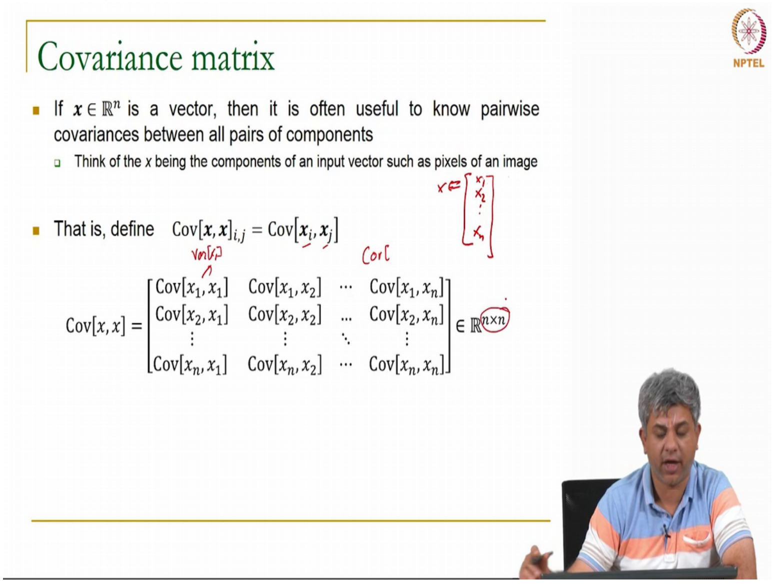
You find pair wise covariances. So the first entry of this matrix is covariance of x 1 with x 1 which is obviously, this is same as variance of x 1.

(Refer Slide Time: 24:45)



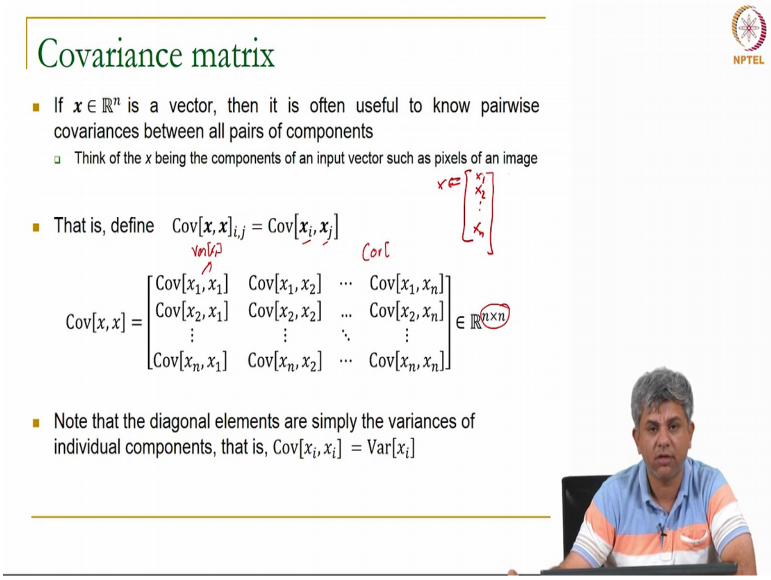
The second entry here is covariance of x 1 with x 2, so on and so forth. So covariance in general of x i with x j, Ok. This matrix will have size n cross n

(Refer Slide Time: 25:00)



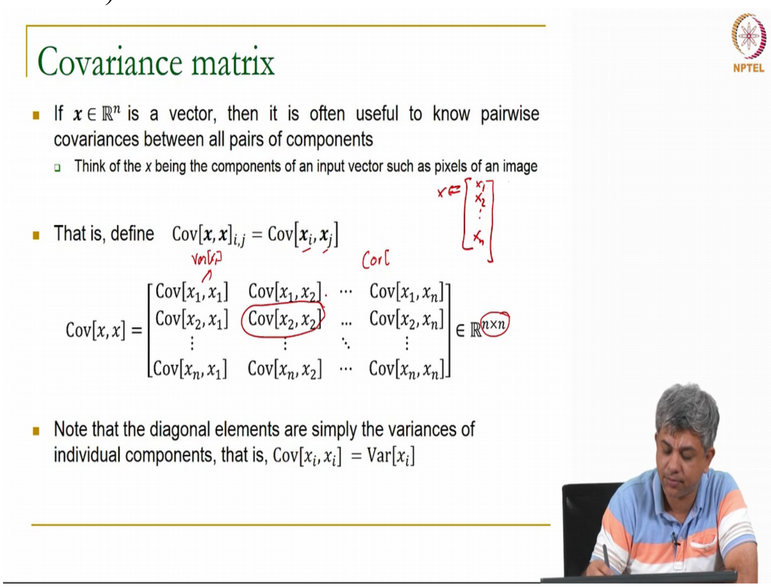
and

(Refer Slide Time: 25:02)



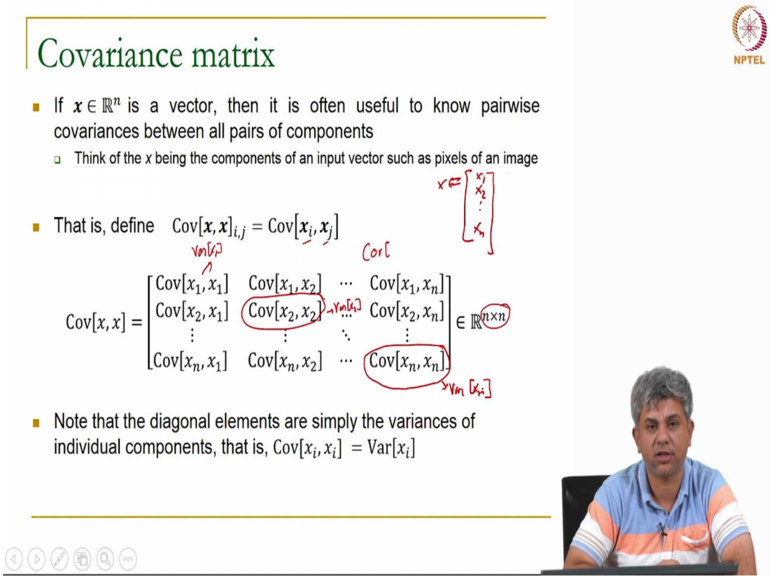
the diagonal entries of the matrix are simply the variances of the individual components. So,

(Refer Slide Time: 25:10)



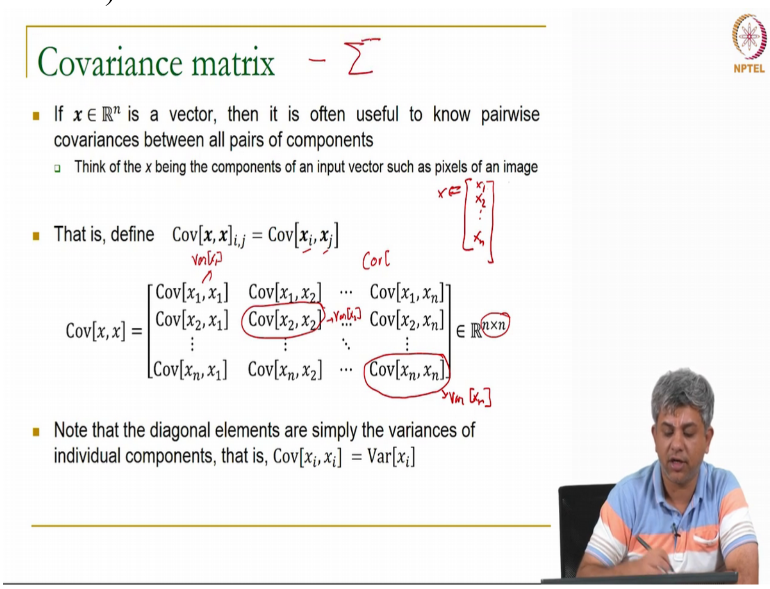
Ok so this one will be variance of x 2, this is variance of

(Refer Slide Time: 25:22)



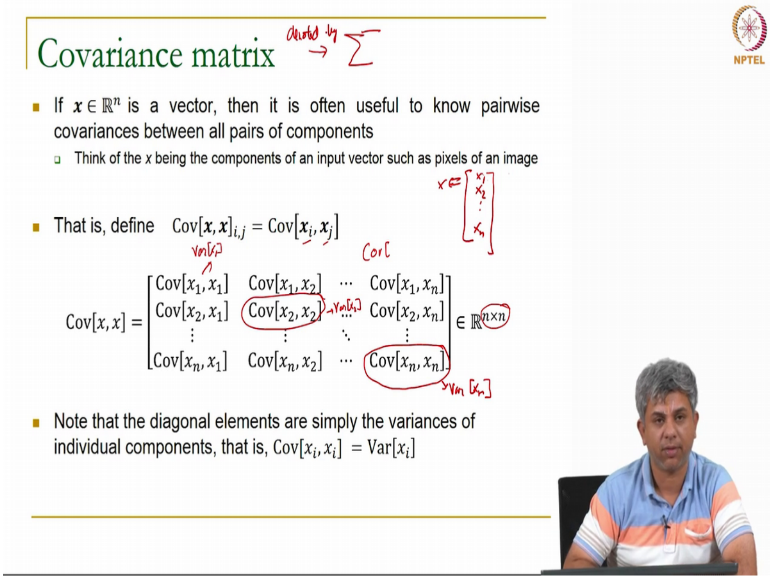
x n. So the covariance matrix is something that we will be using quite often. Sometimes

(Refer Slide Time: 25:31)



it is denoted by capital sigma.

(Refer Slide Time: 25:36)



So we will be seeing this in greater detail in future videos. Thank you.