INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Marketing Research

Lec -23
Correlation & Regression

Dr. Jogendra Kumar Nayak
Department of Management Students
Indian Institute of Technology Roorkee

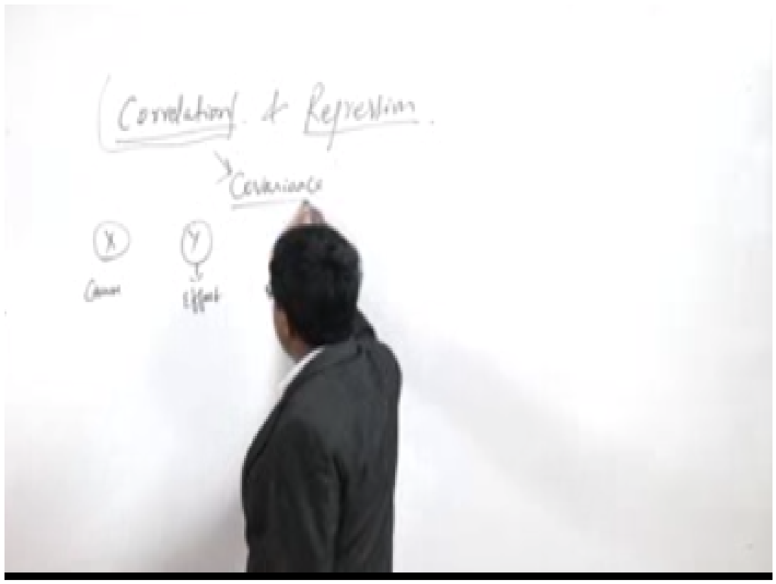
Welcome everyone to the class of marketing research and analysis till that we have discussed on several important factors several important studies which are associated with the subject of marketing research formula example we have described we have discussed about issues like that hypothesis development and then we started with how to test an hypothesis then we discussed about the test of means which includes the T test F test and the Z test basically T, Z and F basically.

So today we will discuss on something similar something important which is highly relevant right in any study and especially and also in the marketing research study let us say let us taken an example suppose we see that there is a cold wind blowing right automatically something comes to your mind we say we felt that may be today it would rain right so why do we say that right or sometimes we feel if more people are coming to a let us say shop there would be a higher sales but it is necessary.

People have might have come for some other reason right they might have visited the store to see a new product or something but whatever with a number of food falls we fell that more people would by products and more sales would happened suppose somebody gives a better ad a good ad very attractive ad let us say and then he feels that if people have liked the ad then would also buy the products right.

So all these things that we are doing markets are actually using this as a very important tool to understand and then understand the consumers psychology the behavior and then try to predict make a predictive analyzes okay so this study what we are discussing today is more importantly called as core relation.

(Refer Slide Time: 02:31)

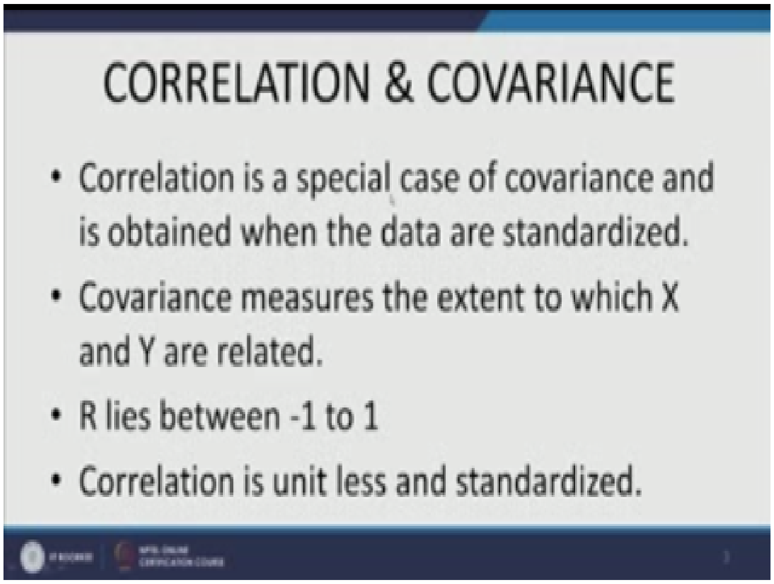


And then we will go into regression okay so the first part being the core relation so which we will discuss at the movement correlation and the we will go into regression so let us see what is a correlation a correlation is basically nothing but a statically technique that is use to measure and describe relationship between two variable suppose let us say there are two variables x and y right generally we say y is the dependent variable or the effect right.

The effect and x being the independent variable or the cause right so whenever we are trying to establish the relationship between two variables we are actually trying to say there any correlation is there any relationship between them now why do we need this relationships this relationship as I said are helpful to make some predictions right it will helps us to understand gain insides about certain features certain tents in the market etc okay.

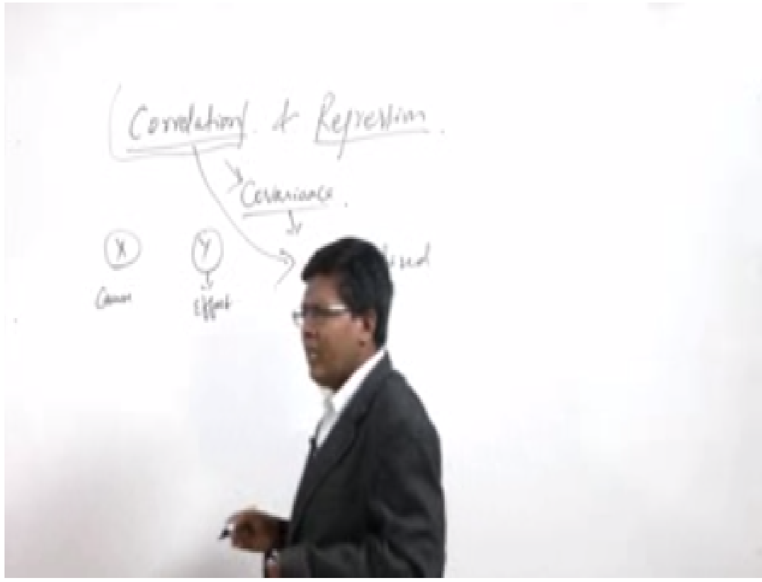
So let us see what is basically a correlation but remember Mainer times there is also another term which is largely used when you talk about correlation that is a word called covariance also students and researchers might have come through these words very often what is this covariance a co variants is as I if you can see this slides.

(Refer Slide Time: 04:10)



A co variance is nothing but a special case of the is a you know the correlation is special case of covariance that means what a covariance is also a correlation but a correlation is not necessarily covariance is that means what happens all covariance's are.

(Refer Slide Time: 04:28)



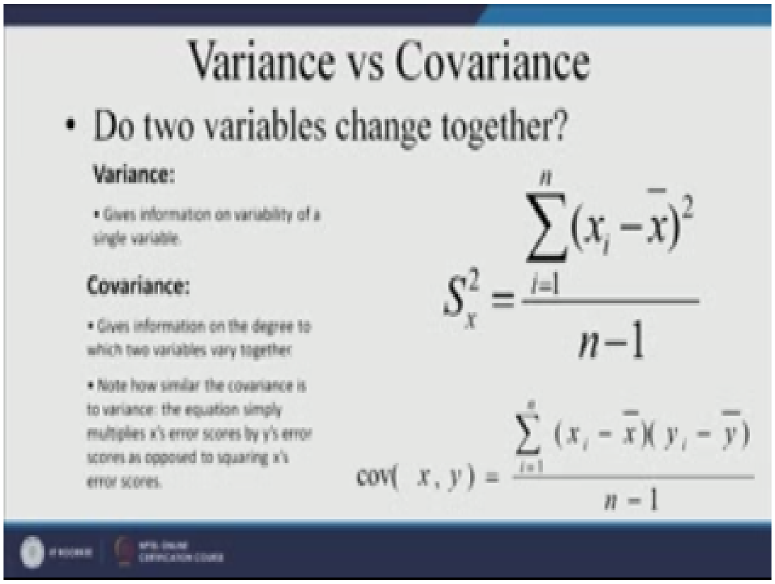
Correlations basically why because a covariance is has it is own units but a correlation on the other hand is something that is standardized in nature is standardized, so it is basically a correlation is a covariance which is standardized okay, so this is if you can see this right covariance also measures the same as I said so it is a correlation also right, and the value between of correlation as it is said is basically falls between – 1 to + 1 okay now what is the value of two things.

If the correlation value of two things is – 1 what do we understand we will just simply stay if there is if a x and y are having a – relationship that means their – 1 in the relationship it is inversely proportional it is inversely proportional that means what if x increases y decreases if x increases y decreases suppose it is a case of + 1 then we would say it is a highly positive relationship it is a this one was a negative relationship when you said – 1 and this is a positive relationship where you say.

Of there is a plus 1 right so the value of correlation falls in between – 1 to + 1 suppose something 1 if the you know the relationship the r which is basically correlation is denoted as r if r = 0 that means you can say two variables x and y are not correlated okay, so let us understand the first terms this terms variance covariance and etc. right so what does variance basically stay now if you remember variance, variance was nothing but we said the square of the standard deviation so what is standard deviation.

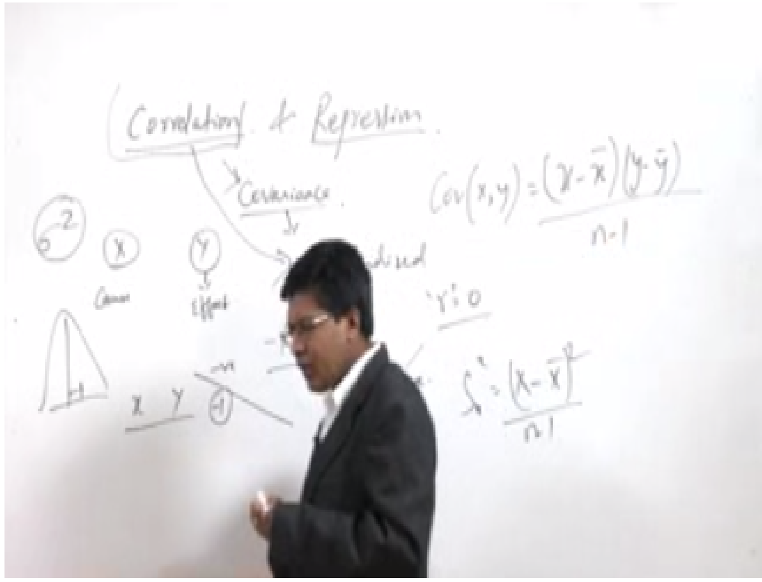
Now you said the standard deviation is something that the value the value of the observed item from the mean okay, now how far the observed item from the average or the mean is is termed as the standard deviation, so standard deviation square is called as r variance okay now if you see this now this is.

(Refer Slide Time: 07:05)



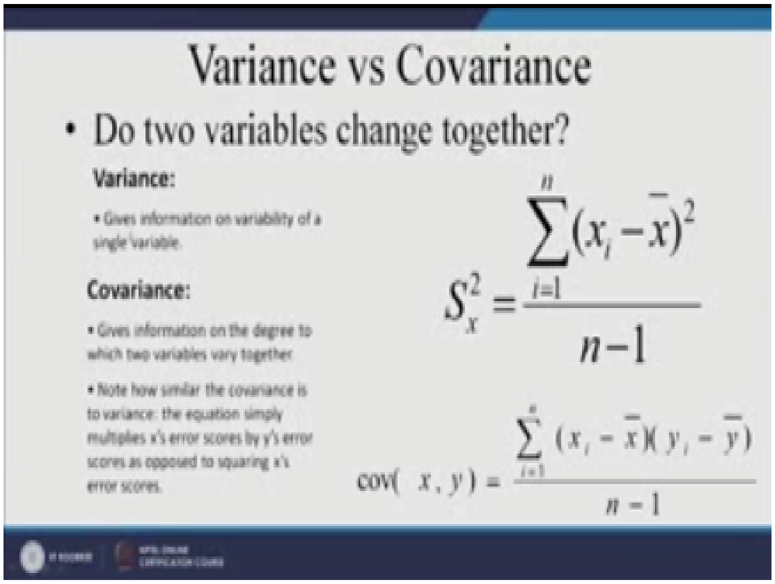
If you look at this S^2 = what it is saying is x – x bar x – x^2 = x – x bar square.

(Refer Slide Time: 07:21)



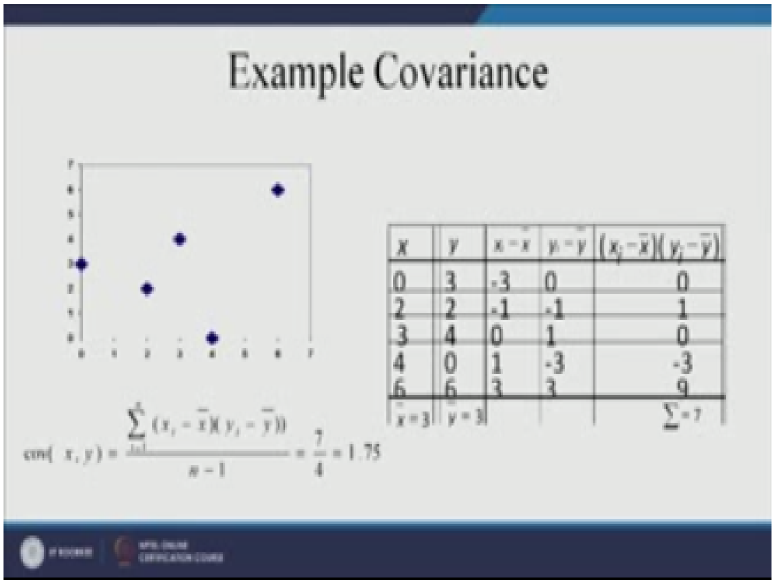
Okay divide by n – 1 now this is the variance that means the how much when there is only 1 variable how much is that variable deviating from it is mean from that we find out the variance okay then what is this covariance now covariance this is generally a doubt which students are others researchers have very large extent because for example we say the structural equation modeling is a covariance structure now what is it covariance mean to understand that you have to understand covariance is of b2 of x and y is equal to how it is mentioned is $(x-\bar{x})(y-\bar{y})$ right, divided by n-1. So if I say if I understand it in this way let us go to this.

(Refer Slide Time: 08:22)



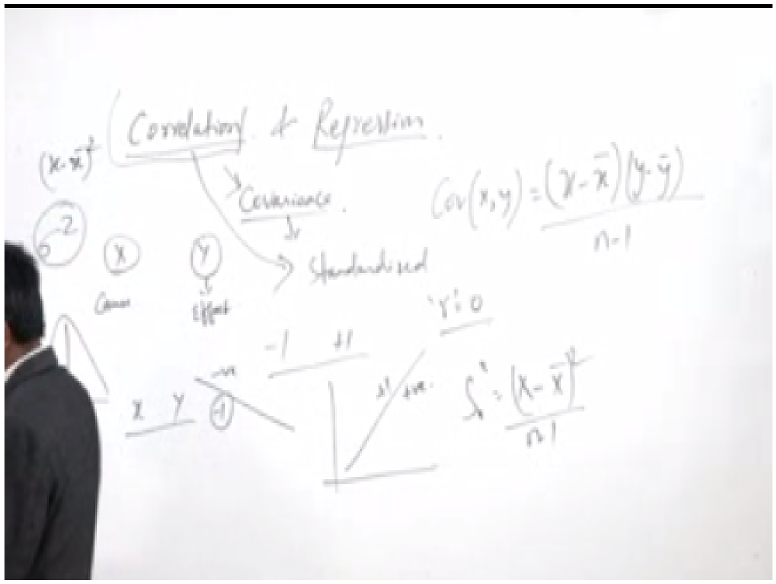
Variance gives information on variability of a single variable and covariance gives information when the degree to which two variables co vary or vary together that means it is like if you understand just imagine like a type of war right, if tow groups are pulling each other how, when one, how would they vary, how would the you know how would they move so basically when x and y are there with a change in x what is the change in y, so how would that vary together that is what covariance structure tells us, okay. So let us take a simple covariance case, now this is a problem if you can see.

(Refer Slide Time: 09:07)



The values of x are given to 0,2,3,4,6 right the mean of the independent variable is equal to 3 right, on the other hand the dependent variable is taken as 3,2,4,0,6 and mean of the dependent variable is 3. Now when I am measuring suppose I would need to measure the let us say individual decorations suppose thus variance then what I would have done, I would have done the $(x-\bar{x})^2$ or $(y-\bar{y})^2$.

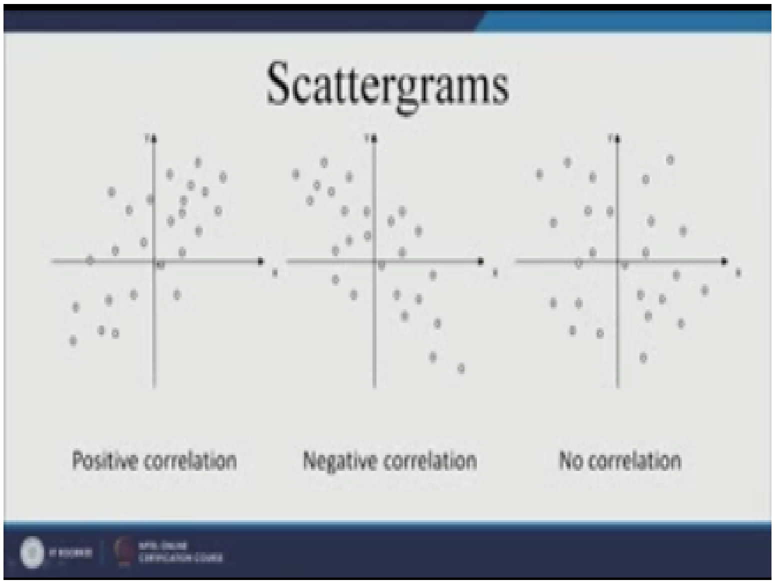
(Refer Slide Time: 09:45)



So $(x-\bar{x})^2$ I would have done, but in this case I am not measuring the individual variables variance rather I am measuring the covariance so how do they co vary the x and y, to do that what I have done is as I explained there $x-\bar{x}$ multiplied by $y-\bar{y}/n-1$ so how much it is coming, now $x-\bar{x}$ is let us say the total has been done, so let us say individually if you do 0-3 right, +2-3 so you have to do it and final it has been done here, so -3,-1,0,1,3 $x-\bar{x}$, $y-\bar{y}$ is 0,-1.1.-3.3 so thus the product, the product of the two is 1.75.

So basically you have to understand that because this is not covariance is not a standardized structure, so it will have its own unit, okay but when you do a correlation, a correlation will come to that, a correlation is standardized okay, as I was explaining this you can check.

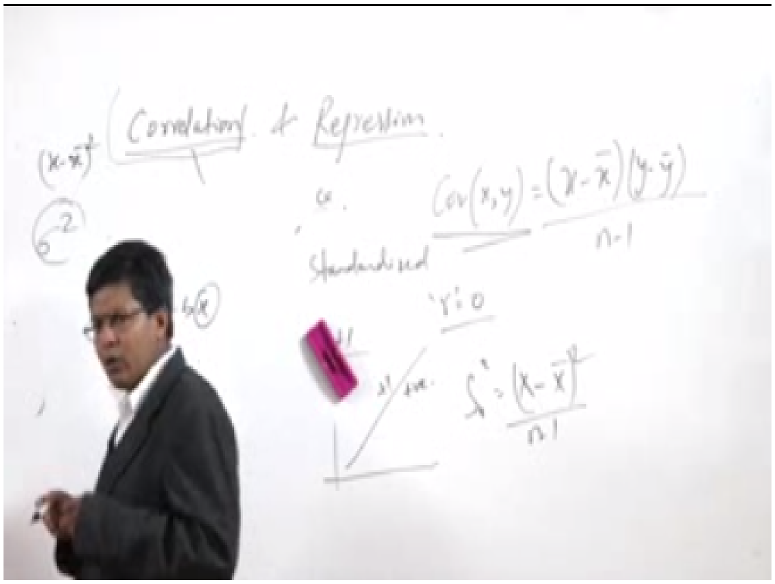
(Refer Slide Time: 10:56)



So when the data are positively correlated now if you see the data is moving up right, on a positive direction that means this is to the x the positive side of x so as the value of x is changing why it is growing right, so we say it is a positive correlation. Now this is a case of negative correlation as x is improving right, y is tending to decrease okay, so negative correlation, similarly this is something where we cannot understand anything because if you look at the scatter plot you will see the scatter plot, you will see that more the data is dispersed around right, so in such conditions we say that there is no relationship how do we find a correlation between the values x and y okay.

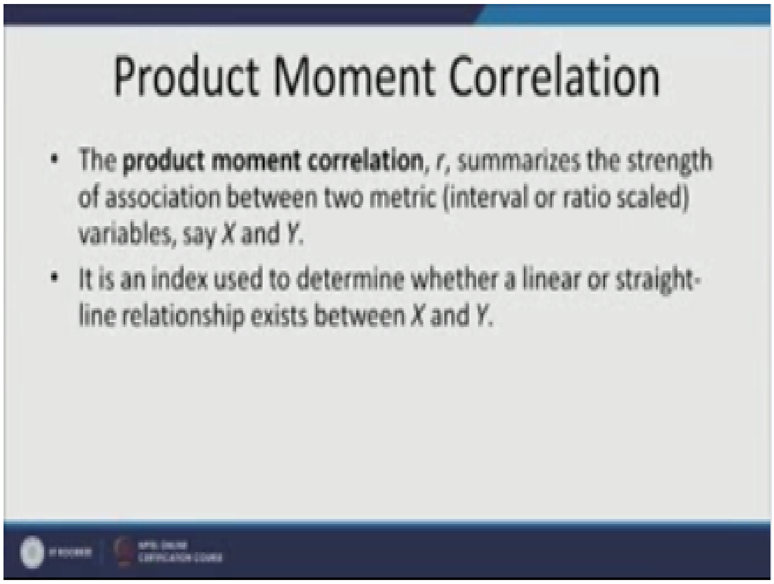
Now product moment correlation are which I was just explaining it summarizes the strength of association between two metric variables right, so what did I mean by that okay, so when I am saying, when I am doing a correlation I am like in the last session also I think I told okay, regression is basically uses two metric variables.

(Refer Slide Time: 12:07)



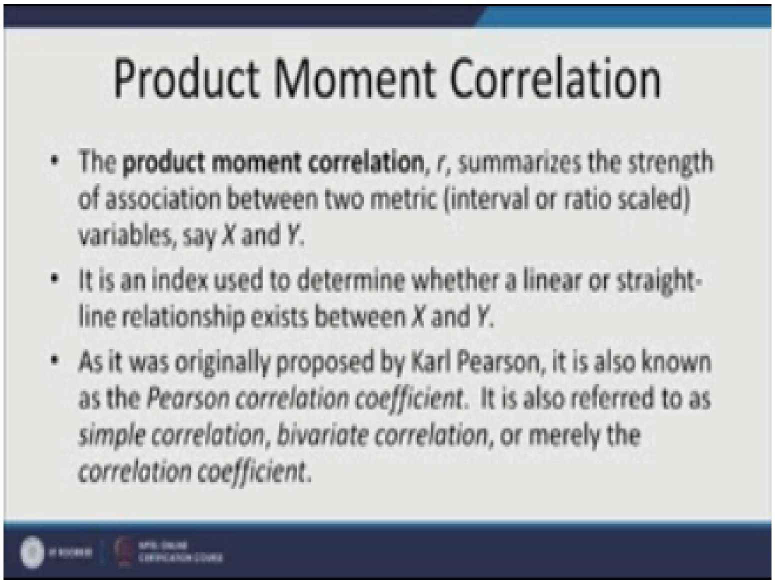
So y=a A+b (x) where x and y I have both measured in a interval or a ratio scale that is a continuous say scale basically right, and similarly the correlation is also takes into account that these two variables must measured in a printable or a ratio scale okay.

(Refer Slide Time: 12:31)



It is an index used to determine whether a linear straight line relationship exists between x and y so the assumption of regression is that it is a linear relationship right, that means the relationship is the change in value is quite similar is same it is moving in a particular proportion let us say right. So that is what it finds out.

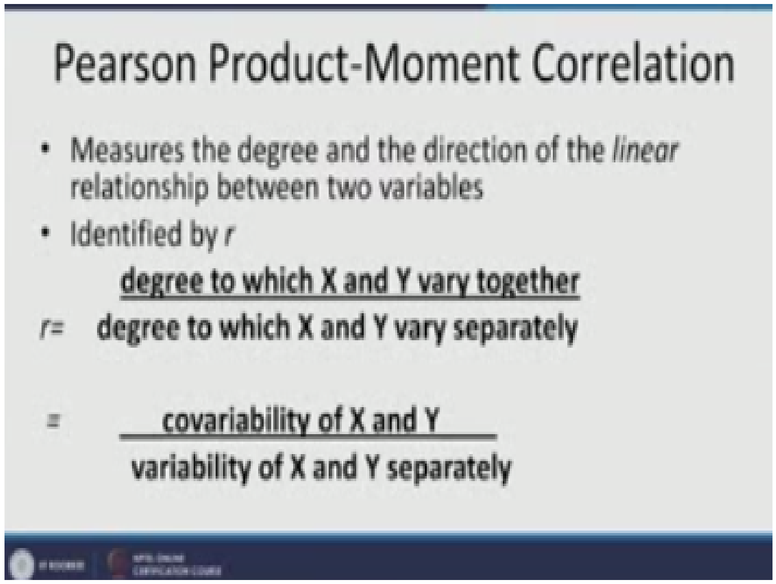
(Refer Slide Time: 12:56)



It was originally proposed as it if I have written if you see it is also sometimes call as a Karl Pearson's correlation coefficient or it is also defined a simple co relation co efficient and or you can say you know the by variant co relations also so do not ever get confusion if you see different names right so by variant correlations simple co relation Karl Pearson co relations right.

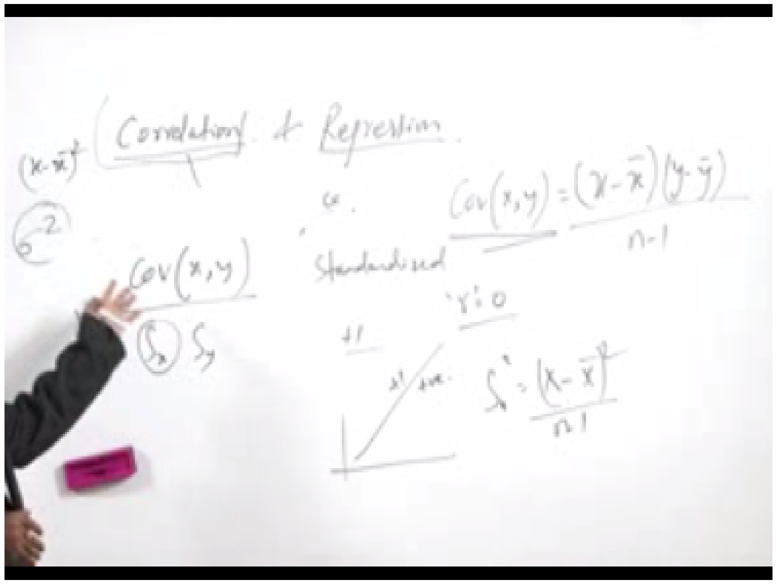
So all these are the same thing it is more or less it is same thing right now how do you explain the R now as I said R if you see if you remember this so it says if you go to it now what is it saying R = the co variants right between x y / what that are degree to which x and y varying separately right so that means what x suppose this is my standard deviation of x this is my standard deviation of y right.

(Refer Slide Time: 13:58)



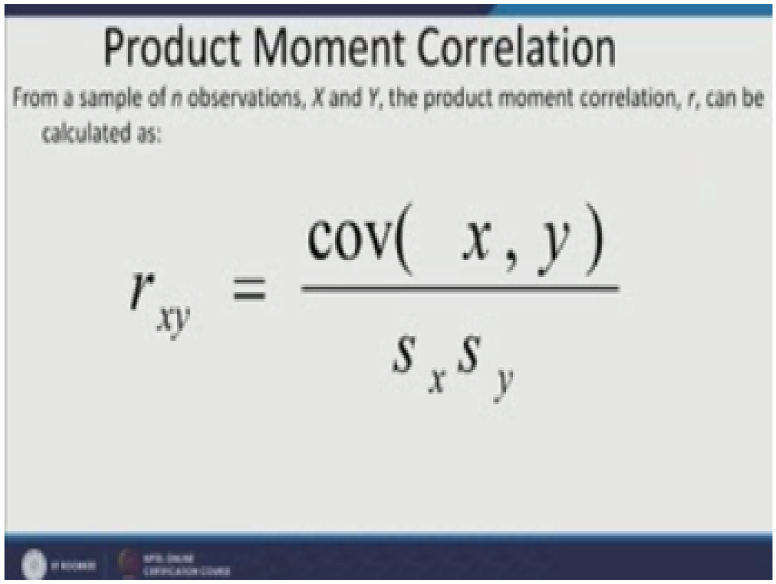
So it says co variability of x and y together upon the standard deviation of x / the standard deviation of y will tell me the R value.

(Refer Slide Time: 14:09)



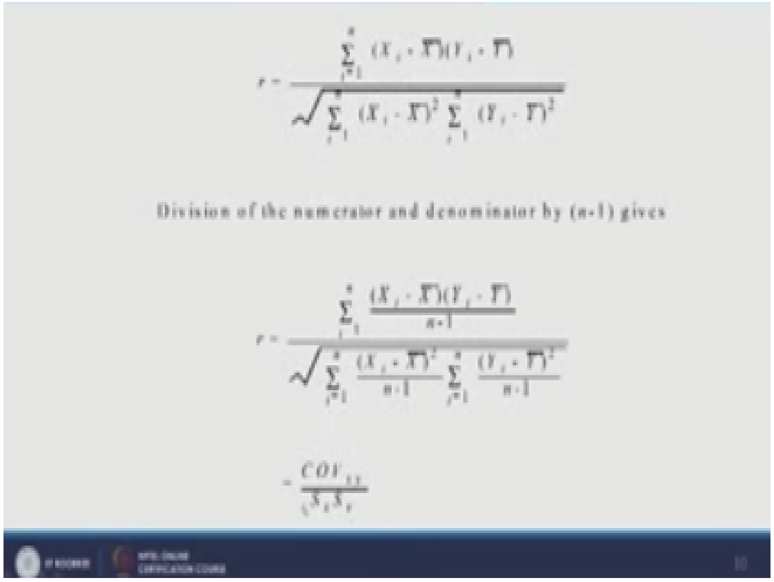
Now if you can see this automatically you can understand that it is a unit less thing right there is no unit because a units will be cut each other right at the numerator in the denominator okay. So this is how explains for any sample of n observations x and y.

(Refer Slide Time: 14:31)



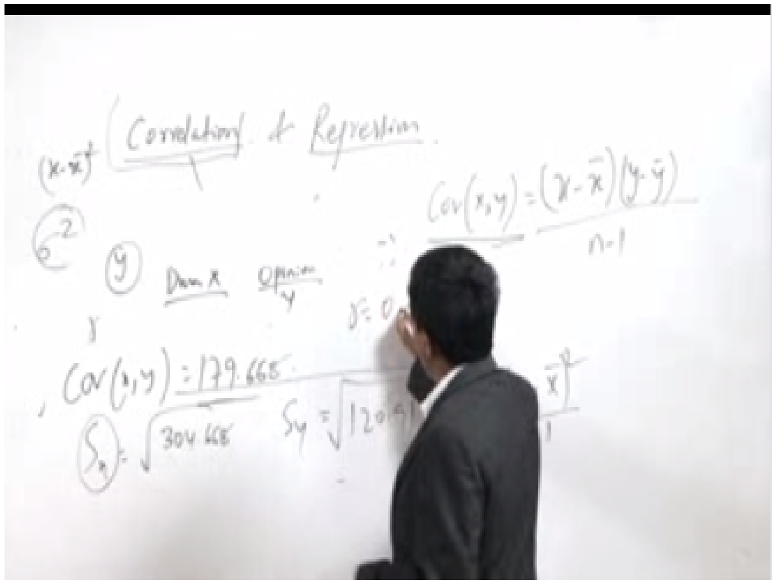
The product movement co relation is calculated as I have mention there co variants of x and y / the sampler standard deviations of x multiply by the standard deviation of y okay.So if you look at this is very interesting.

(Refer Slide Time: 14:50)



So I have said this right from here if you remember only this much which I just explain you do not have to remember anything else right you can make your formula. Now how that is let us see now suppose I am saying co variants of x and y / sx / s of y let us say so how can I do it what I can do from here now co variants we knew is what is co variants we are written is x – x bar right / y – y bar okay so x – x bar x y – y bar / now what is sx and sy.

(Refer Slide Time: 15:29)

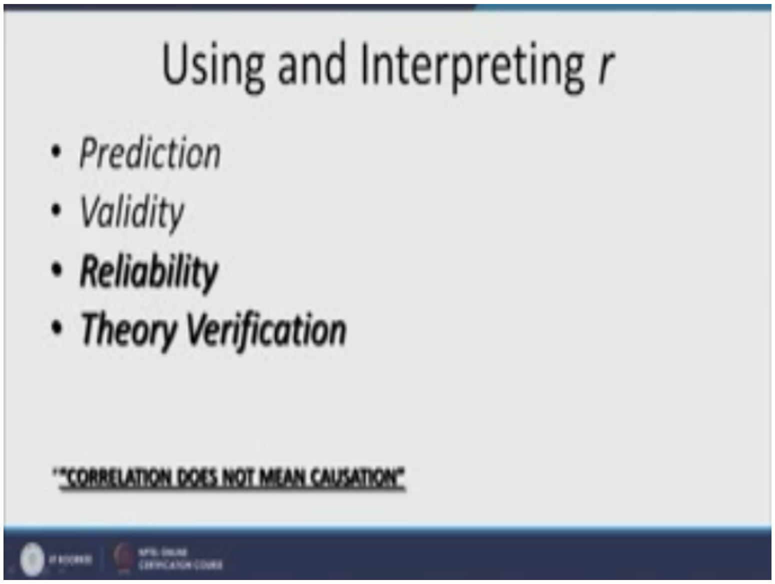


Now √ of because why it is a √ now because we are talking about the variants at this and the standard deviation is the √ n of the variants right. So it will be x – x bar square / this is also divided by n – 1 degree of freedom n -1 x y – y bar square / n – 1, so you do not have to remember anything now you just see if you look at the R, r is equal to only thing we know r is nothing but the ratio of the covariant the co variants means how two variables are changing according together divide by the individual deviations right.

Now when I am doing this if this is co variants is x – x bar right in to y - y bar / n – 1similarly divided by whole divide by √ of x – x bar this is the individual x variants right multiplied by the y variants and the route over. So if we do this so we are getting the same thing again if you and the n- 1 n- 1 will obviously get you know crosses' so what remains is x – x bar this is the formula so x – x bar x y – y bar / √ x – x bar square x y – y bar square.

So this is the basic you know formula right, so do not forget to write the summation right. So this is the basic formula right so how do you use r what is the use of r how do where do we use it r has a very large use as I said it the higher value of r as I said if the more the value of or the close the value of r is to one you will say that there is a any strong relationship let us say if a company is announcing that it is going to come with the new product in the market right only an announcement it has not actually done it is just made an announcement suppose it has made an announcement that is going to come with the new product in to the market would people buy the shares of that company right is there any relationship between the announcement made by company and the share holders interest in investing in the particular companies share now to do that to get an idea this is where the correlation comes into play.

(Refer Slide Time: 18:29)

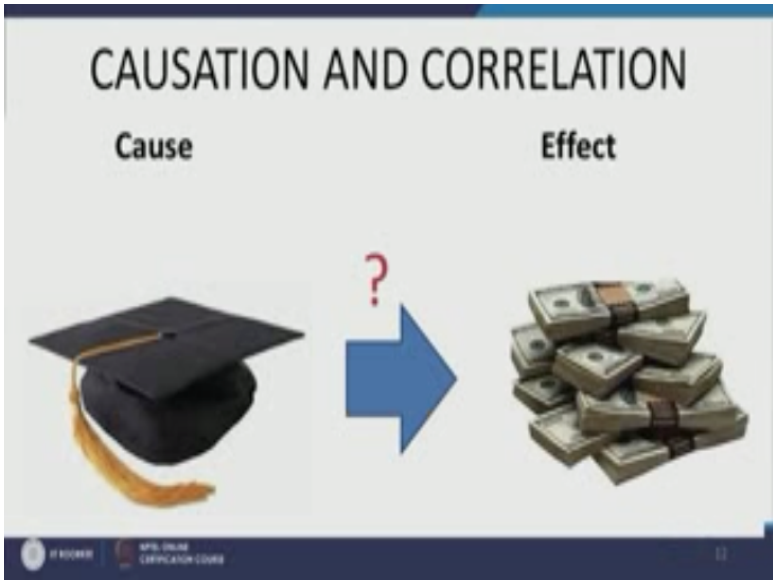


And thus R is useful in not only several things for example predicting right so it helps you to predict right it helps you to validate and also check the reliability of the relationships right between x and y right and many times the theory is verified suppose you have got a theory you have got a new theory and you say may be you know there is a relationship between glacier melting let say glacier melting.

And let say flying of plains over that particular area let say I do not know no if I am proposing a new theory or I am proposing a new kind of knowledge now how to check that the value of R or the correlation value comes into handy it helps us right so as I have written one more thing you to understand many at times students and researchers get confused.

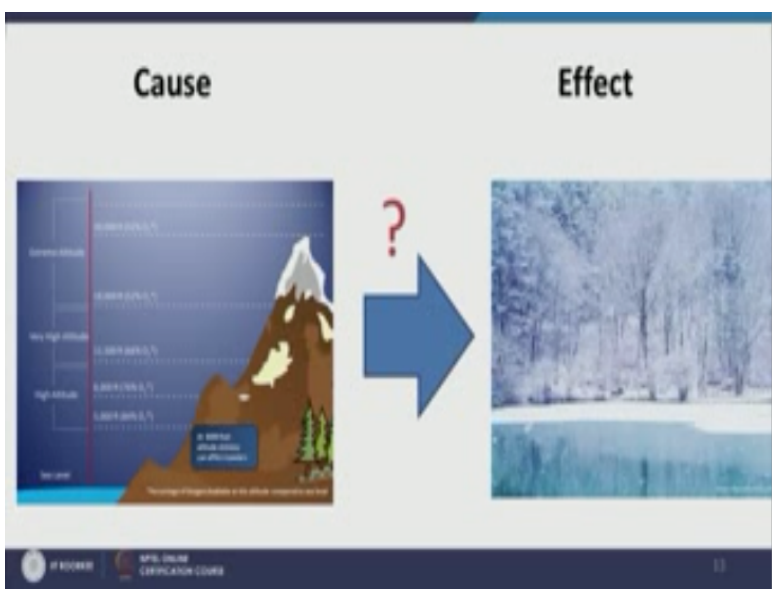
A correlation and causation or the same thing now we have understood variance co variance correlation I am coming to four thing called causation if you it is written here you can see correlation and causation are generally understood to be the same thing but it is not correct it is absolutely not correct right.

(Refer Slide Time: 19:56)



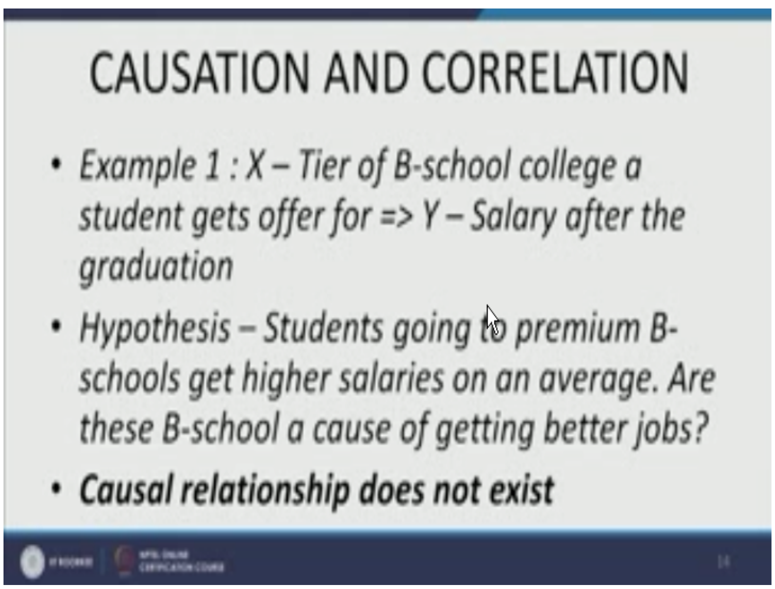
There is a line of difference between causation and correlation right let us see that I have brought two photographs check this photograph now if you look at this is the relationship it says what happens this is something like somebody is getting a degree right and his earning chances are improving we are saying so is it a causation or correlation let us think about it right.

(Refer Slide Time: 20:17)



Similarly second is this is the ace where we are saying if you go up the you know altitude then the temperature will tend to go down right there will be less temperature okay now if you look at the two cases these two cases what is the causation and what is a correlation it will be very clear to you right so let me show you.

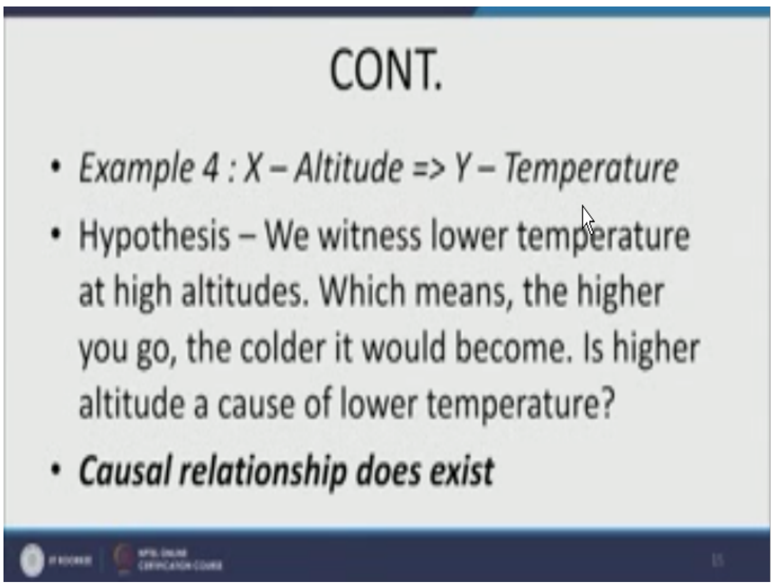
(Refer Slide Time: 20:46)



Now the first example what it is saying the tier 1 suppose in B school college a student gets offer right you know once the student graduates from the institute from a good school or nay were if good engineering college or anywhere right so salary after the graduation is better right so the hypothesis that is coming up is students going to premium these schools get higher salary or an average so are this B school is case or getting better jobs.

Is the question right but here we cannot say right we cannot say a student might have got better job because of some other talent of that he has got may be it is Percival right so in such condition a causal relationship does not exist we cannot say it is cause and effect is there only cause right there could be several other causes for the student has got good salary right.

(Refer Slide Time: 21:47)

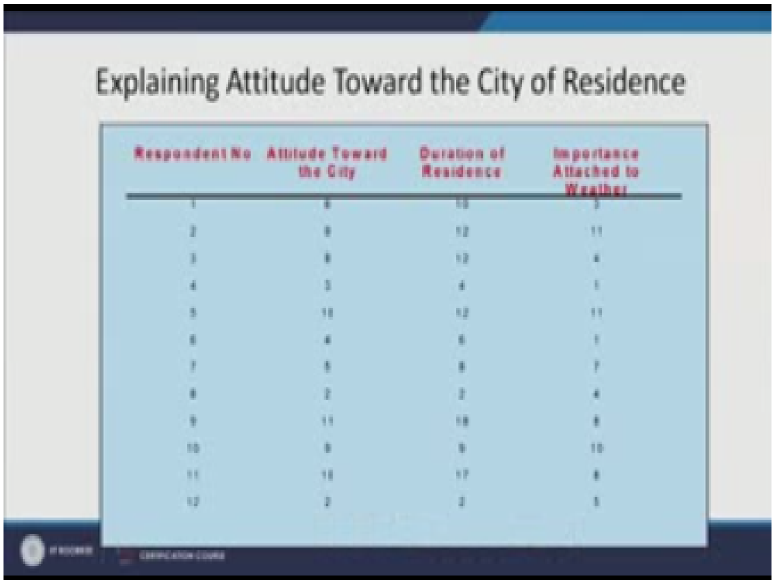


In the second case we said hypothesis was we witness the lower temperature higher altitudes it which means the higher you go the colder it would become so is higher altitude cause of lower temperature now we are asking this in this relationship if you see here I am saying a casual relation exist now why I am saying this because in the previous case we could not say that only degree was the cause there could be other relationships also right.

But in those case you cannot you know you cannot say that it is not effecting it is you cannot go to the viewers right so that a means what we are saying okay there is no other explanation so in the explanation why the temperature is going down is because higher altitude so that means what higher altitude will lead to temperature this is a causal relationship.

So that means what is the first was the case of correlation yes there is a correlation there is no doubt about it but students if they go to a better B school they would on better salaries there is a correlation but that was not case of causation right but in this case it is the aces' of causation right now let us take up a problem so we are this is the data if you see the data is given to you.

(Refer Slide Time: 23:09)

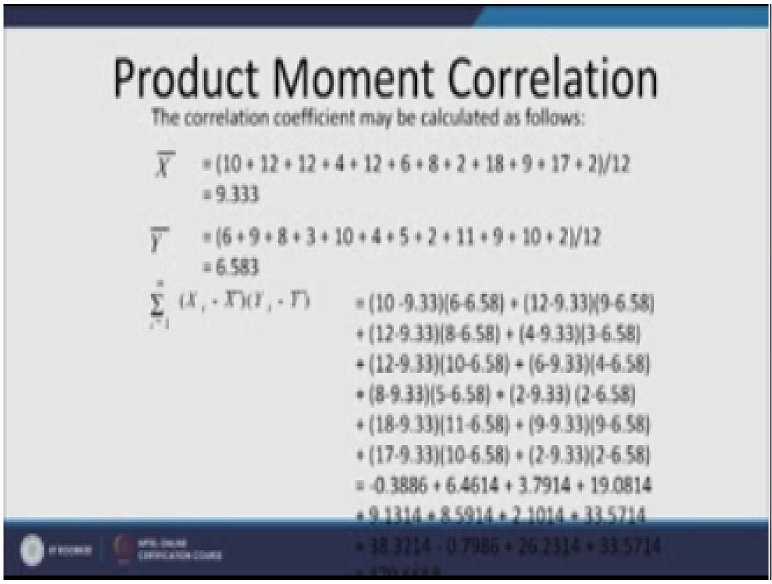


Now what we are going to do is, we are going to see how this has been measure right, now what we are saying the responded are given to you, the attitude towards city is given, now the attitude towards city is basically are Y, so the dependent variable. The independents variables are duration of residence, how much time the persons had stayed in that particular city right and the importance attached to the weather conditions of the city.

For example now we are saying what are the two things, the attitude towards the city is nothing but the amount of the day that person is stayed in that city and the importance to it. Whether it is not the 2 variables that can be there, there can be many such variables but in this case we have taken only two, we have taken only 1 in this study. So let say what it has done basically, so what has been done in the 1st case we have only taken the 2.

What are the 2? Now we have taken the duration of the residence, x and the opinion or the feeling about the city is y okay. Now let us look at this now if you go up again this is the x and this is my attitude in the y okay.

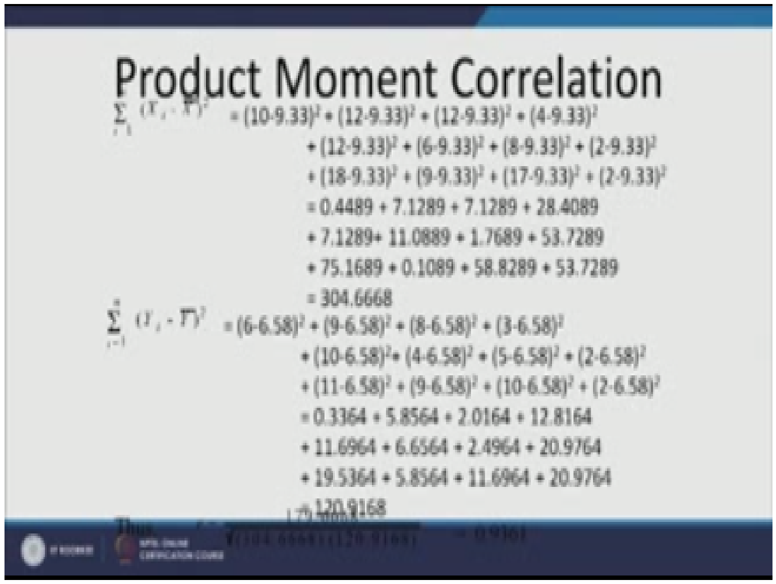
(Refer Slide Time: 24:47)



Now x bar = let see 10 + 12+ 12 + 4+ 12, 6, 8 + 2 + 18+ 9+ 17+ 2/12 which I s= 9.333 okay, on the other hand if you look at the dependent variable is 6+ 9+8+3+10+4 goes on till 2 right, so which is 6.583. Now we are interested to measure the R , now let see 1^{st} we need the co variance, the co variance we said x – x bar x y – y bar. So if we measure the x – x bar is known to you 9.333 and x is the individual value, so 10-9.333, 12-9.333, again12-9.333 till 2- 9.333 okay.

So when we get this we found this the value of this is something around 170 it is not clear at the bottom, somewhere it is coming to 170 okay.

(Refer Slide Time: 26:03)



So 179.66 so the co variance value, the co variance of x and y = 179.668, similarly what we need for the denominator for now we need standard deviation of x and y right, so do that we have calculated the x – x bar and we will take the √ of the variances right. so what we have been calculated 304.668 x – x bar the variance of the x is 304.668, similarly the variance of the y = 120.9168, so 304.66 and 120.91, so x = √304.668 I think it is clear because the variance, so when I require the x I will take the √ of it.

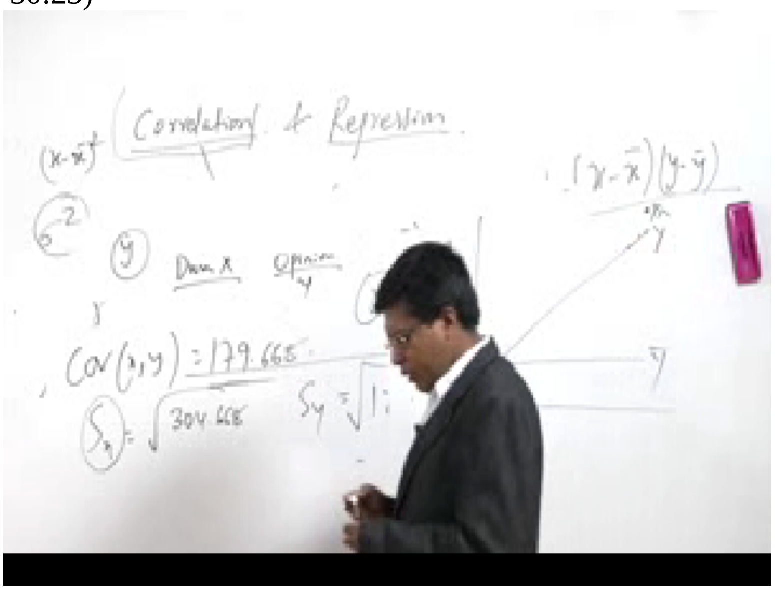
And similarly s(y) = √120.91, so if we calculate now we will find the value to be now 0.9361 so when we divided this value this 179.68/ this so we found that R = 0.9361. So now this tells us, this gives us the very clear picture that there is the strong relationship between the attitudes for the persons has for the particular city against the independent variable of how much he has state so if you have stayed long that means if he has stayed long in the long time in that city his attitude for the city is good right and if he had stayed for a lesser time his attitude for the city is not so good right as clear as that as simple as that it is a positive correlation.

It is a very high correlation right but many times although I am not telling you right now at this moment there are situation where extremely high relationships in certain situations are also not desirable but at the moment we will not talk about it so that is the special case of multiple linearity which we will go later on.

So the total variation so once you have calculated the R now this R that we have calculated we this R is if we take the r^2 is something a term called as coefficient of determination okay so what is basically this it is saying now in terms of recursion although I have not come to regulation so just imagine for the moment I will explain once I have to recursion so imagine that the higher the value of r is the higher is the explanation power of the independent variable for the dependent variable that means this if the r^2 is nothing but it is the ratio of explained variance divided by the total variance.

Now let me show you otherwise how it would look like in a graphical way if you want to see that suppose let us say this is something like you know the value y and this is something like the y right now and let us say this is my observed value of y right.

(Refer Slide Time: 30:23)



So my total variance is explained as in this case my total variances would be how much of variation is happening now this yi-y so this entire is my total variation right what is my explained now this is my $y=a+b_x$ right so y had the predicted value is what we are talking about that y had is y had let us say ssy or sum of square total $y=y-y^2$ okay on the other hand if we talk about there are two other variations variances that is coming up in this case one is called the explained variance.

And the unexplained variance now this part $y-y^2$ right y had predicable predicted and the mean now if you take this is called the explained variances right this is called the unexplained variances or it is also called as residual or error right however you may understand right and this is the regression basically the we talk about the ss regression we say is the explained variances so for any researcher the entire research the total variance is this total this is why the observed value but if I write this it will look like $y\text{-}y^2\text{=}y\text{-} \chi\text{+}y\text{-}y^2$ this is how it look like okay let me wind up at the moment we will continue in the next session we will come with little more of correlation.

And then we will get into regression right this is the very important very interesting and very useful subject because it is highly utilized in any kind of studies were you find a caution effect or you find correlation right so you need to study understand whether there are significant relationship between variables are not to do that this correlation and regression are very important right so we will continue in the next session with the same right thank you for this session.

For Further Details Contact

Coordinator.Educational Technolgy Cell
Indian Institute of Technology Roorkee
Roorkee 247 667
E-Mail Etcellitrke@gmail.com etcell@itr.ernet.in
Website: www.itr.ac.in/centers/ETC. www.nptel.ac.in

Production Team
Sarath Koovery
Mohan Raj. S
Jithin. K
Pankaj saini
Graphics
Binoy. V.P

Camera
Arun. S

Online Editing
Arun.S
Video Editing
Arun.S

NPTEL Cooridinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production