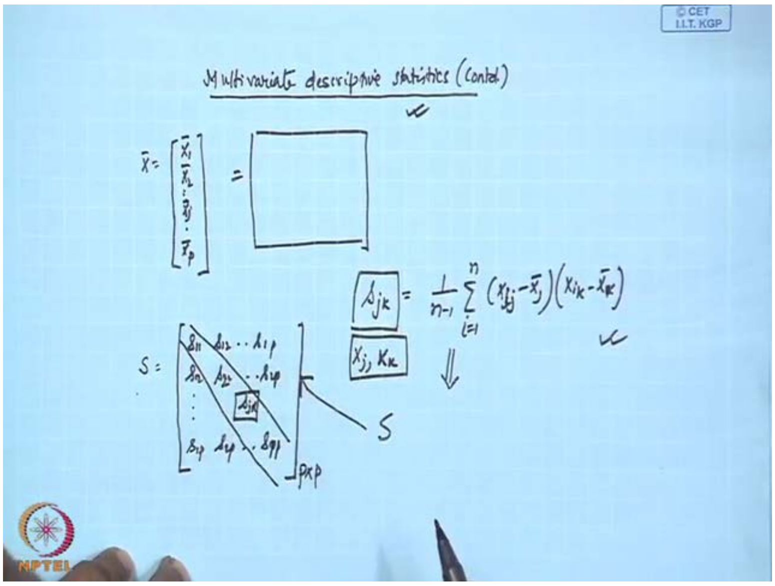
Applied Multivariate Statistical Modeling

Prof. J. Maiti
Department of Industrial Engineering and Management
Indian Institute of Technology, Kharagpur

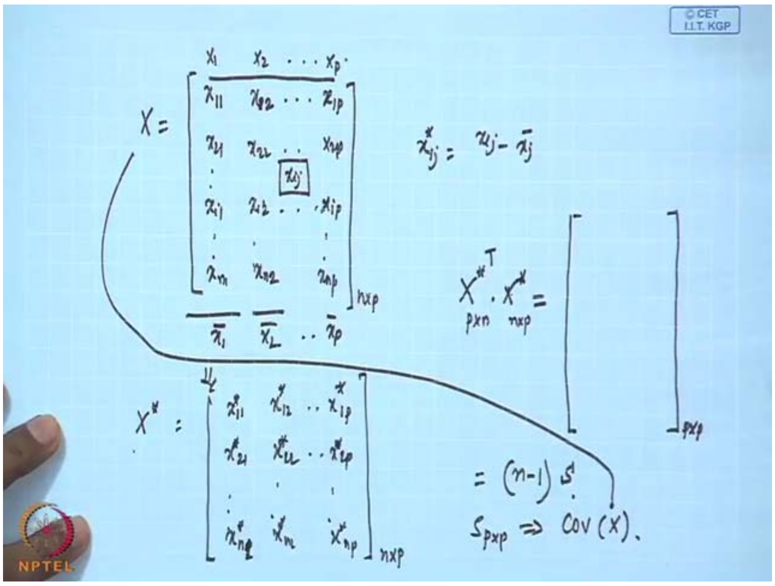
Lecture - 9
Multivariate Descriptive Statistics (Contd.)

(Refer Slide Time: 00:25)



Good afternoon, we will continue multivariate descriptive statistics today. Last class I have given you S j k the formula to compute the covariance between two variables X j and X k, and this is the sample covariance computation formula. Now, we want to use matrix here, primarily matrix multiplication to compute S, where S is this one, that P cross P matrix, diagonal elements will be the variance component and up diagonal elements will be the covariance component. We will compute these from the data matrix.

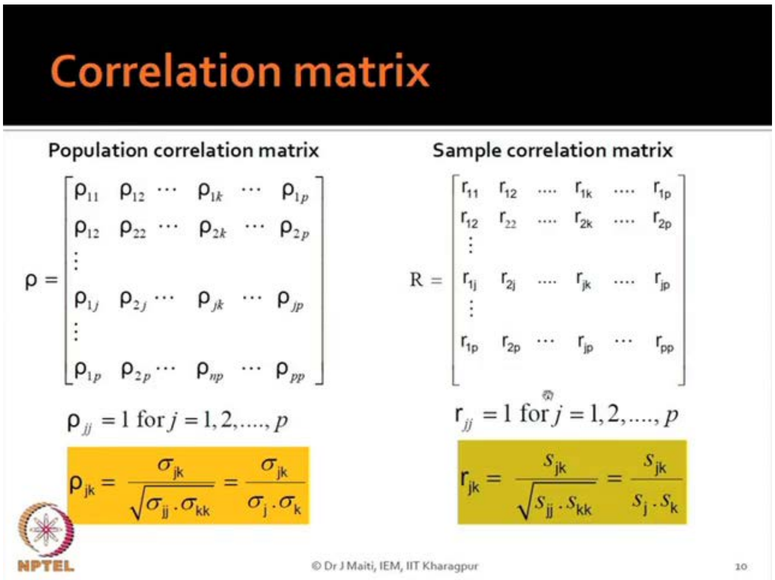
(Refer Slide Time: 01:20)



Last class you have seen the data matrix like this that x 11, x 2 1 like x i 1, x n 1, that is observation on variable x 1. Similarly, observation of variable x 2 is x 1 2, x 2 2 like x i 2 then x n 2, so you take all p variables then x 1 p, x 2 p, x i p and like this x n p. This is our n cross p matrix. Now, let us consider a general observation here, which is x i j, so if we create one more general observation x star i j, which is x i j minus x j bar, then what you do, if you subtract each of the elements by its respective mean. For example, for this one, if you subtract by x 1 bar, for these it is x 2 bar like these it is x p bar, then we will create another matrix, which we are denoting like this x star this is your x star 1 1, x star 2 1 so like this x star n 1 then x star 1 2, x star 2 2 like this x star n 2. So, x star 1 p, x star 2 p like this x star n p.

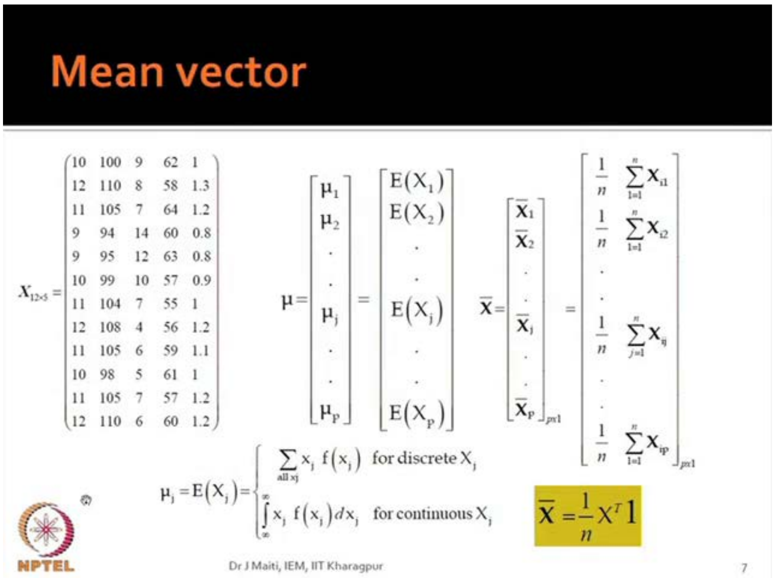
Now, if you create these x star transpose, what will be the order of this matrix p cross n and take a dot product with x star. This resultant matrix n cross p resultant matrix will be p by p cross matrix, which is nothing but n minus 1 into S, where S is the covariance matrix. So, S is p cross p that is the covariance of x, where you have taken this calculated this x. So, in matrix multi manipulation you are you are able to find the covariance matrix in just one go for all the variables. Now, you solve one small problem here.

(Refer Slide Time: 5:14)



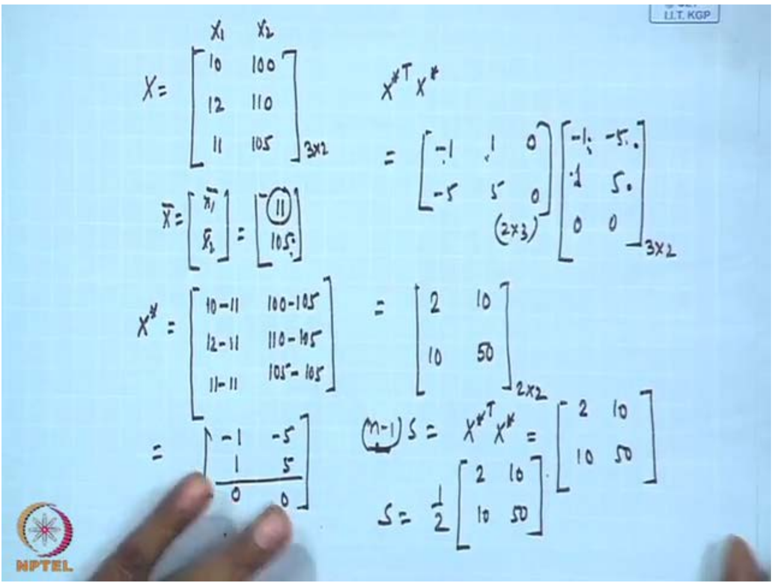
For example, suppose this one you see that you take this one that first 10 12 11 100 110 and 105, this data matrix.

(Refer Slide Time: 5:16)



X equal to 10 12 11 100 110 and 105.

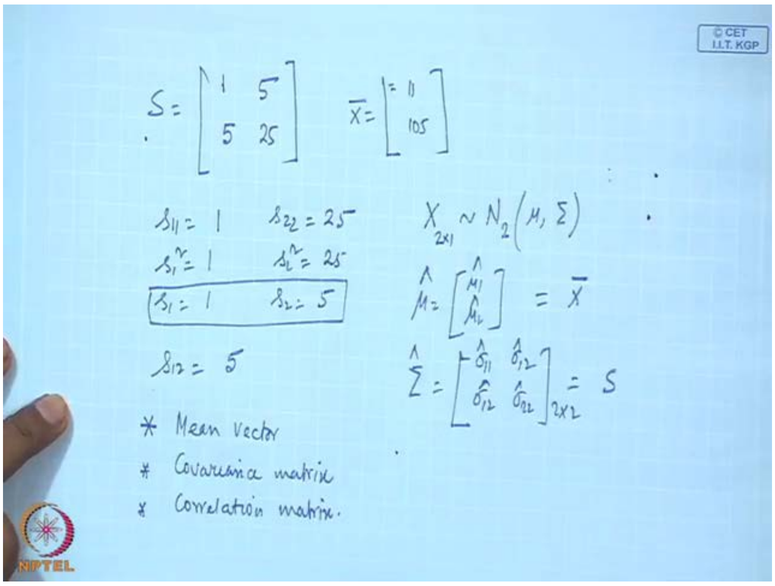
(Refer Slide Time: 05:21)



So, what will be your x 1 bar or if I say x bar equal to x 1 bar and x 2 bar, what will be its value it will be 10 plus 12, that is 33 by 3, 11 and this one will be 100 110 105 105. Now, what you are creating, you are creating X star. What is this X star? You want that each element on x 1 will be subtracted by its mean 11. Similarly, each element of x 2 will subtracted by its mean 105 so then this one will be 10 minus 11, 12 minus 11, 11 minus 11, second one will be 100 minus 105, 110 minus 105, 105 minus 105. Then what is happening here, you are getting this is 10 minus 11 minus 1 plus 1 zero, then 100 minus that is minus 5 plus 5 zero. So, sincerely that zero is element is coming here.

Let us see that what will happen if we do like this x transpose x, what will happen here minus 1 1 0 minus 5 5 0, multiplied by a x star minus 1 1 0 minus 5 5 0. If you multiply what will happen, this is basically 2 cross 3 matrix, this one is 3 cross 2, so you want to get matrix called 2 cross 2. So, this time this minus 1 into minus 1, that is plus 1, plus 1 plus zero for every this 2. Now, second one minus 1 into minus 5 that is plus 5, 1 into 5 plus 5 that is 10, so this 1 minus 5 into 1 5 5 into 1 5 5 plus 5 10 and this one 25 plus 25 that is 50. So, we say n minus 1 into S equal to X star transpose X star, so which is here 2 10 10 50. So, what is n value here 3, so minus 1 is 2, so s will be 1 by 2 2 10 10 50 then what is this value now.

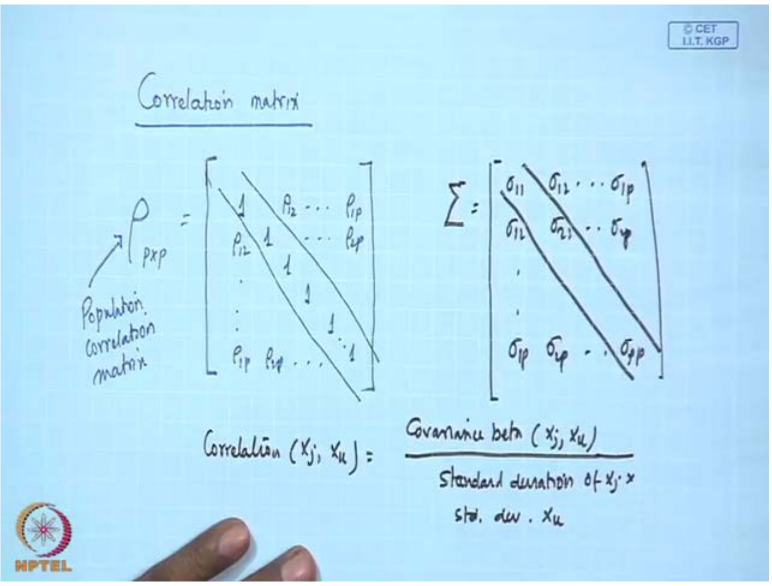
(Refer Slide Time: 08:59)



Our S is that means 1 5 5 25, so you have already computed x bar, which is your 11 105 and your S is this. So, that mean S 1 1 is 1, S 2 2 is 25, what does it mean S 1 square equal to 1 and S 2 square equal to 25, S 1 equal to 1, s 2 equal to 5, that is a standard deviation both the cases and S 1 2, which is 5 the covariance between x 1 and x 2. So, if I say that my population is multivariate normal that is 2 cross 1, this is the variable, so it is basically N 2 mu and sigma, then my mu is mu 1 and mu 2 and sigma will be 2 cross 2 sigma 1 1, sigma 1 2, sigma 1 2, sigma 2 2, this is 2 cross 2.

So, we can now say, this is our x bar the estimate in this manner we will precede. So, that please remember multivariate descriptive statistics has 3 components, one is mean vector, second one is. Covariance matrix, third one is correlation matrix. Now, we will discuss about correlation matrix. Now, population correlation matrix is denoted by rho, this is population correlation matrix.

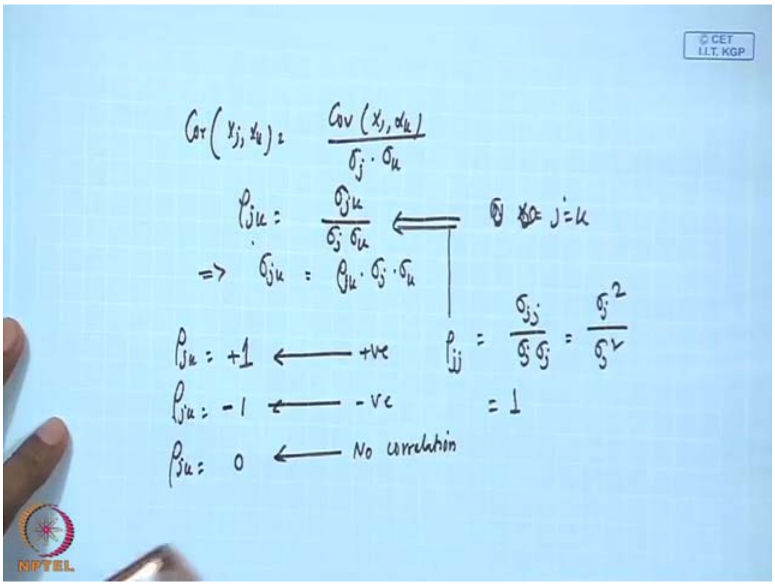
(Refer Slide Time: 11:23)



Basically, what I mean to say that the population characterized by p variable, then will you get p cross p matrix for the population correlation matrix, like p cross p for some population covariance matrix. Duty here is that your diagonal element will be 1, this is the correlation of the same variable with it. And up diagonal variable element will be writing like this rho 1 2, like rho 1 p here also rho 1 2, rho 2 p. So, like this rho 1 p, rho 2 p, it will continue like this. Now, if this is the case we find out a relationship between rho and sigma.

What is sigma, population covariance matrix that we have seen sigma 1 1, sigma 1 2, sigma 1 p, sigma 1 2, sigma 2 2, sigma 2 p like this 1 p, sigma 2 p, sigma p p. So, the crux of the matter is the diagonal elements are variance that is means, the same variable varying with it that is single here, diagonal correlation. So, if you see what is the correlation between x j and x k, then you can write this one as covariance between x j x k by standard deviation of x j times standard deviation of x k. So, mathematical what we will write basically mathematically that correlation of x j x k is cov of x j x k divided by sigma j sigma k.

(Refer Slide Time: 14:35)

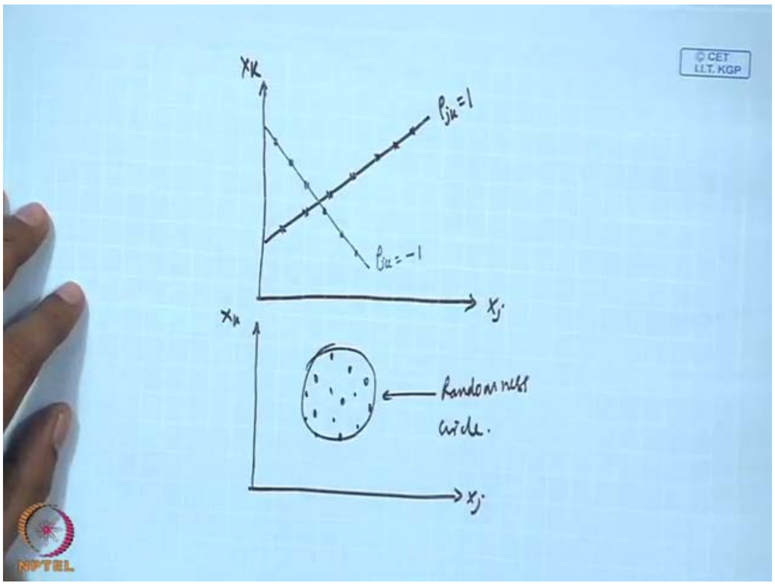


Now, you have seen from this figure, we are saying that correlation between j and k is somewhere like this rho j k, here also we are talking about that covariance between this is this like this. So, then if we use this same notation I can write rho j k equal to sigma j k divided by sigma j sigma k, that is the relationship. What is the relationship then covariance between 2 variables is the correlation between the 2 variables times its standard deviations.

Now, what will happen to this correlation when sigma j, what I mean to say j equal to k. You see what will happen to this that means it will be j j, j equal to k j j, then sigma j j by sigma j and we have discussed earlier that sigma j j is nothing but sigma j square. So, sigma j square which is one and as a result you are getting all 1. So, conceptually that the same variable you are trying to find out the covariance between the same variable, then that will be give the variance.

And here what you are doing, you are basically standardizing it by dividing the standard deviation to the co variance component. So, as a result this standardization effect is bringing you that all the diagonal elements will be 1. The same thing will happen for sample data also. What do mean by suppose correlation between j and k is plus 1, correlation between j k is minus 1, correlation between j k is 0. This one is saying that perfectly positively correlated and one stands for the perfect correlation, this minus stands for negative, which is why that was positively correlated, this is negatively correlated, and this one is having no correlation positive negative and no correlation.

(Refer Slide Time: 17:43)

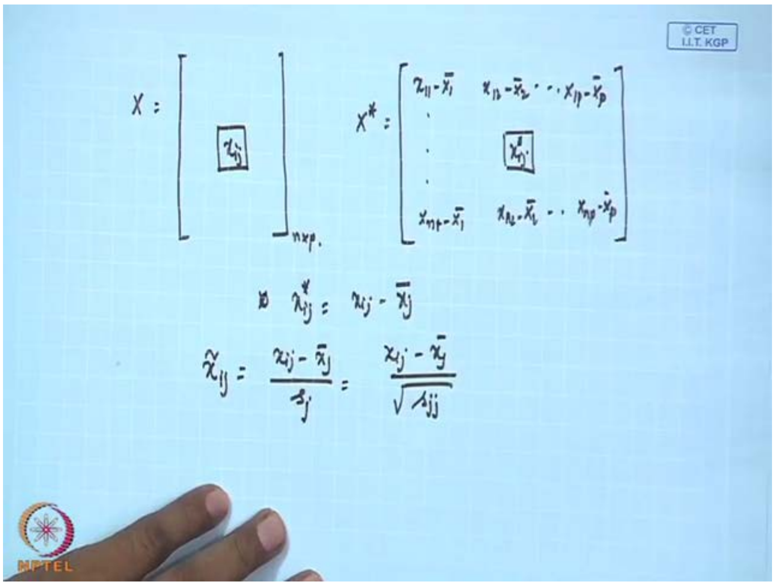


And if you draw this, suppose you have two variables this side suppose x j and this side x k and if you draw scatter plot for positive correlation, you may get like this that is rho j k equal to one. And negatively correlation means x j will increase x k will decrease or vice versa. So, you can think like this, here what is happening in this case this rho j k equal to minus 1. Assuming that all the values of x j and x k falling on this line, that is why negative that mean here when x j increasing, x k is decreasing.

And in the in this case when x j is increasing, x k also increasing, all are falling under the positive 1 possible when it will increase both the variable co vary in the same direction, negative means in the opposite direction and 1 is possible, when you will find a perfect straight line if you do curve fitting. And when you will get zero suppose, your points are like this, this is your x j, this y axis is x k, points are like this.

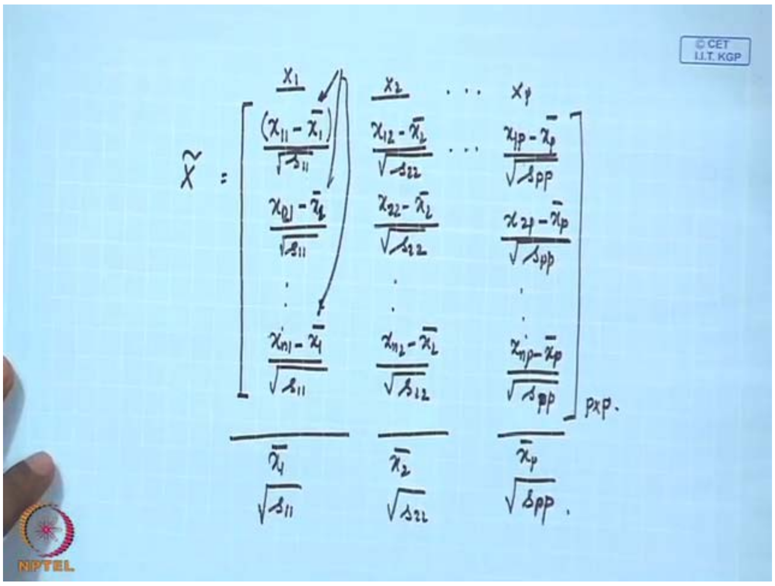
You cannot find circle, this points are resembles circle it’s totally random, so when you find this type of randomness, that mean it is resembling a circle. There is no relation because you see you take any direction; you will not get any pattern here, which is the meaning of correlation coefficient like rho j k. You convert it into the same how do you calculate?

(Refer Slide Time: 20:10)



You have data set like x, where same data set n cross p and I have already given you that we have converted this x star, where each of the observation was subtracted by its mean like this. This n 1 minus x 1 bar, n 2 minus x 2 bar, n p minus x p bar, this was we created earlier. And here we created one general x i j, here also we have created that x i j star, where x star i j is x i j minus x j bar. Now, let us create another variable, let us write like this x tilde i j, this one is x i j minus x j bar by s j, what are you doing? You are first finding out the mean subtracted value, and then dividing it by the corresponding standard deviation. I can write it like this, x i j minus x j bar square root of s j j.

(Refer Slide Time: 22:22)

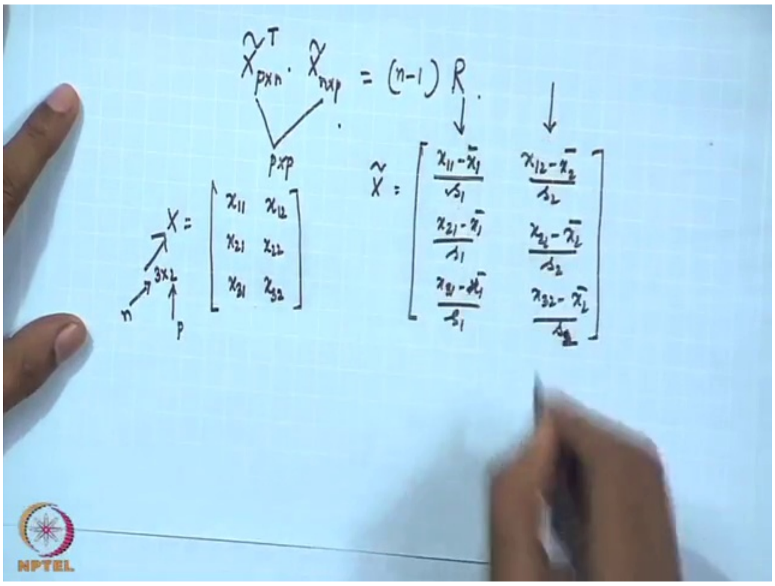


If you follow these then you will create a matrix which is x tilde, this one will be look like this x 1 one minus x 1 bar, these divided by S 1 1 square root then x 2 1 minus x 1 bar divided by square root of S 1 1, same manner all the observation in the x variable is x n 1 minus x 1 bar divided by square root of S 1 1. For x 2 what you will do x 1 2 minus x 2 bar divided by square root of S 2 2, this is 2 2 minus x 2 bar divided by S 2 2 like this, x n 2 minus x 2 bar divided by s 2 2.

So, if you go in this manner for p th variable then you will write x 1 p minus x p bar divided by square root of x p p, then x 2 p minus x p bar divided by square root of S p p, same manner x n p minus x p bar divided by square root of S p p. This is a transform at data matrix p cross p, when it is x 1, all will be 1 1, x 2 all will be 2, when it is p all will be x p p. See the similarity is all these are x 1, when each observation is subtracted by the same mean vector of the variable 1, and each that sub resultant quantity is divided by square root of s 1 1.

Similarly, here x 2 bar square root of S 2 2, similarly here x p bar square root of S p p. So, you see that earlier ultimately, if you see the covariance and correlation relationship, you have found out that see j k, when you say j k, we will basically the co variance component is divided by the corresponding standard deviation. So, in order to achieve this, what we are doing here, we are now dividing each of the observation by the corresponding standard deviation.

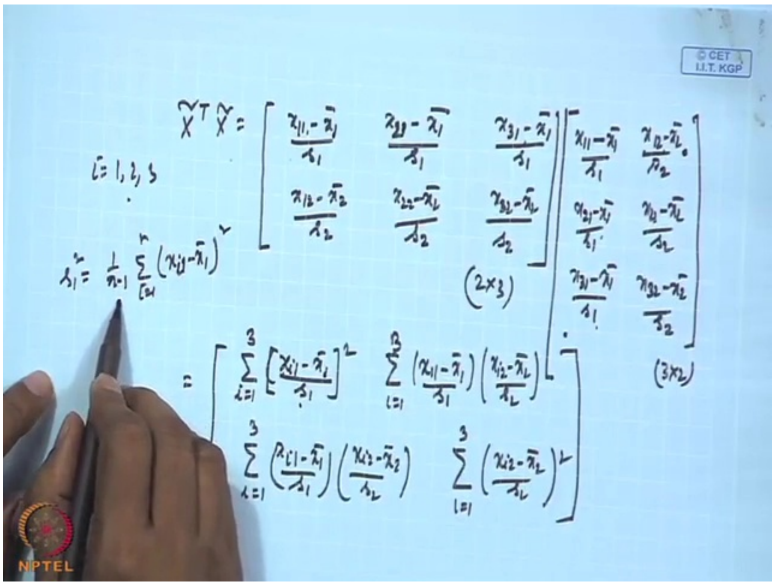
(Refer Slide Time: 25:52)



Now, if you find out this one with this so x transpose then it will be p cross n, if you give the transpose here x tilde transpose dot product x tilde, that will be your n cross p. So, the resultant quantity will be your p cross p then what will happen? This will not be identity matrix; you see you will get this one as n minus 1 into r. If you want to check it, you check it very simply, suppose my data matrix is like this, I will take only three values x 1 1, x 2 1, x 3 1 then here the second variable you take x 1 2, x 2 2 and x 3 2.

So, in this case the data matrix is 3 cross 2, where n is 3 and p is 2. Then what you are creating here, you are creating x tilde. What is this? This is nothing but x 1 1 minus x p bar by I am writing S 1 only, instead of square root of S 1 1, that is S 1 I am writing. Then this will be what x 2 1 minus x 1 bar by S 1 and x 3 1 minus x 1 bar by s 1 and this will be x 1 2 minus x 2 bar by S 2 then x 2 2 minus x 2 bar by S 2, then x 3 2 minus x 2 bar by S 3. What will happen if you now do like another, keep in mind the variable and accordingly write down this one.

(Refer Slide Time: 28:05)



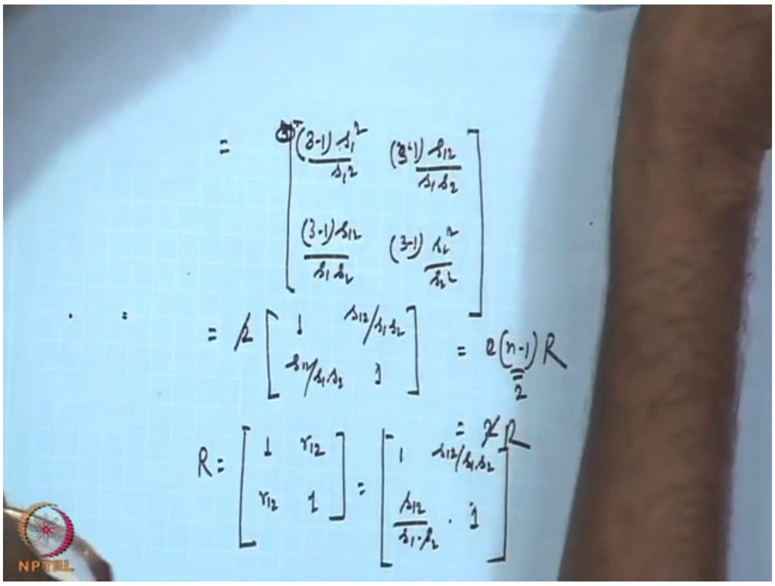
So, when if I do like this x tilde transpose x tilde, what will happen your x 1 1 minus x 1 bar by S 1 x 2 1 minus x 1 bar by S 1 x 3 1 minus x 1 bar by S 1. And then this one will be x 1 2 minus x 2 bar by s 2 x 2 2 minus x 2 bar by S 2 x 3 2 minus x 2 bar by s 2, this times you will be writing the same thing x 1 1 minus x 1 bar by S 1 x 2 1 minus x 1 bar by S 1 x 3 1 minus x 1 bar by s 1.

And then here x 1 2 minus x 2 bar by S 2 x 2 2 minus x 2 bar by S 2 x 3 2 minus x 2 bar by S 2, this one is 1 2 3. So, 2 cross 3, this is 3 cross 2 you see this into this plus this into this plus this into this, what is happening here x 1 1 minus x 1 bar by S 1 x 1 1 minus x 1 bar by S 1, you are getting a square. So, what you are doing then. You are basically getting some total if I write i equal to 1 to 3 because our observation is 1 2 3. Now, second one stands for the variable so x i 1 minus x 1 bar divided by divided by S 1 that square you are getting.

Then what will be this one, this cross this, now this cross this, what will we get x 1 1 minus x 1 bar x 1 2 minus x 2 bar. See you what you are getting here, you are also getting i equal to 1 2 3 x i 1 minus x 1 bar divided by S 1 into x 2 minus x 2 bar divided by S 2. Same quantity will be getting here i equal to 1 2 3 x i 1 minus x 1 bar by S 1 and x i 2 minus x 2 bar by S 2. Here you will be getting a square term i equal to 1 2 3 i 2 minus x 2 bar.

Getting any similarity here, any clue are you getting you seeing. We say what a sum total is of if I ask you what S 1 square is. Suppose n equal to 3, then what is s 1 1 square, 1 by n minus 1 sum total of i equal to 1 to n x 1 or x i 1 you write minus x 1 bar square S 1. So, this quantity is what n minus 1 into S 1 square.

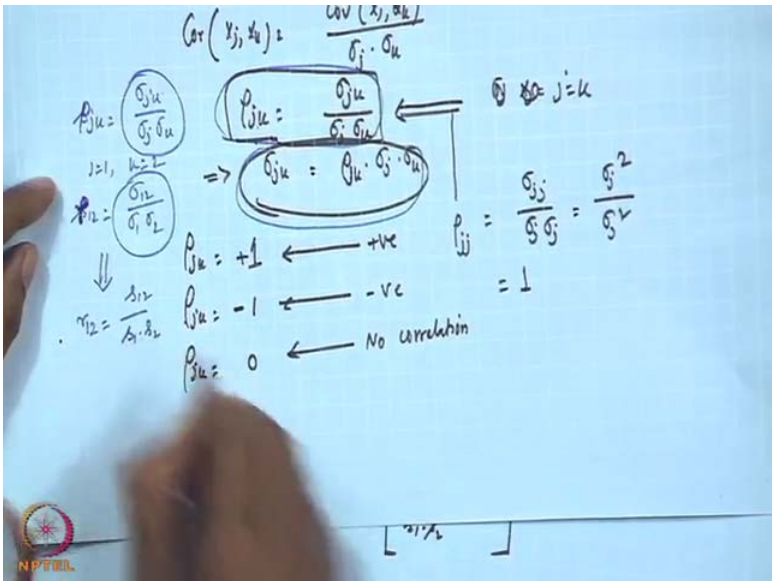
(Refer Slide Time: 32:04)



So, ultimately what will happen ultimately, you will get this quantity like this that we can write n minus 1 means 3 minus 1 3 minus 1 into s 1 square by s 1 square. Then second one what will happen, this is the co variance n minus 1, you will get S 1 2 by S 1 S 2 that is n is 3 here. So, 3 minus 1 S 1 2 by S 1 S 2 and then this one also you get 3 minus 1 S 2 square by S 2 square. Then if I just take out 3 minus 1 which is basically 2 then what I will get, I will get 1 s 1 2 by s 1 i s 2 s 1 2 by s 1 into s 2 and 1.

As a result what we writing this is n minus 1 r n minus 1 is 2 that mean this is 2 r. Now, 2 is 2 cancelled out so r is now 1 r 1 2 r 1 2 1, which is now 1 S 1 2 by S 1 into S 2 then S 1 2 by S 1 into S 2 into 1. I think you were seen earlier that the correlation and covariance, the relationship you were seen earlier the relationship what in population domain we say this one I have said to you, this as well as this. Now, instead of rho in the sample domain, what you can write?

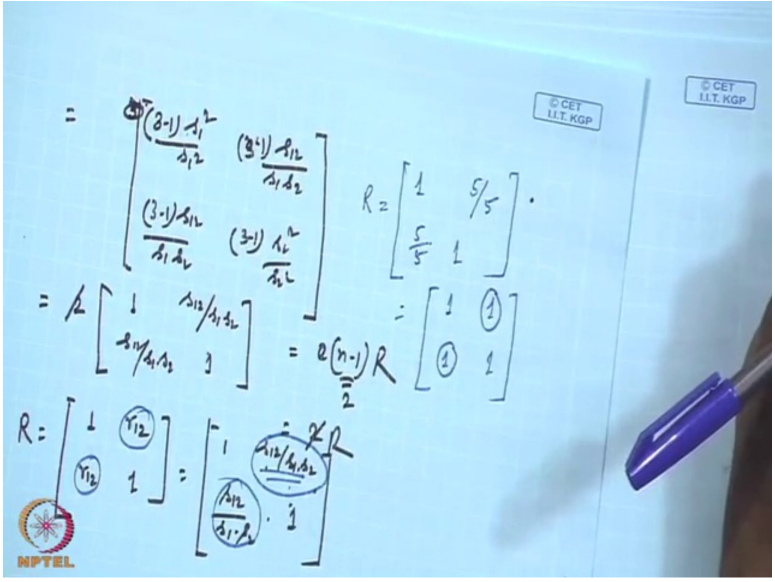
(Refer Slide Time: 33:58)



You can write S j k equal to sigma j k by sigma j sigma k. Now, if your j equal to 1 and k equal to 2 and this S 1 2 equal to sigma 1 2 by sigma 1 sigma 2, what is exactly happened here. This is not S j k, this is basically rho j k, this is rho j k then this will be your r j k, this is basically replaced by sigma and then it will be like this. So, if I can write rho j k here, rho 1 2 like this one.

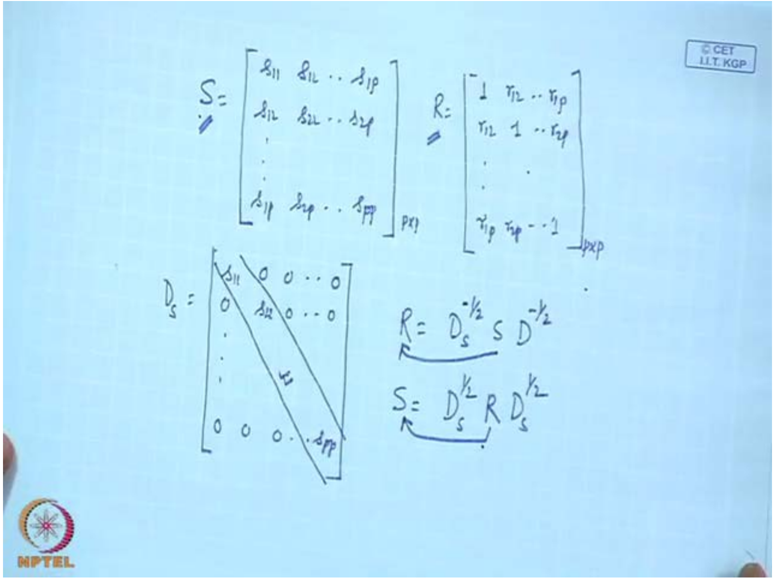
Now, in the sample domain when we go that will be your r 1 2 will be your S 1 2 by S 1 into S 2, that is why these r 1 2 is S 1 2 co variance by their corresponding standard deviation. So, for the same data set now can you compute this r value? I think we have computed one place r value, you have computed, we have computed here we have seen S equal to this. So, that means your standard deviation S 1 is given and S 2 is given for the same data.

(Refer Slide Time: 36:08)



If I want to compute my r, then what will be your r value? Your r value will be first that blindly you can write like this, diagonally it will be 1 r 1 2 will be S 1 2, so S 1 2 is 5. So, you can write 5 divided by their standard deviation S 1 is 1, S 2 is 5 so that mean 5 by 5 5 by 5, 1 1 1 1 1. You are getting perfect correlation, 1 means perfect correlation. So, this is what your multivariate descriptive statistic, we will talk about that is why the mean vector correlation matrix and covariance matrix. Now, you can very easily convert this one.

(Refer Slide Time: 37:22)

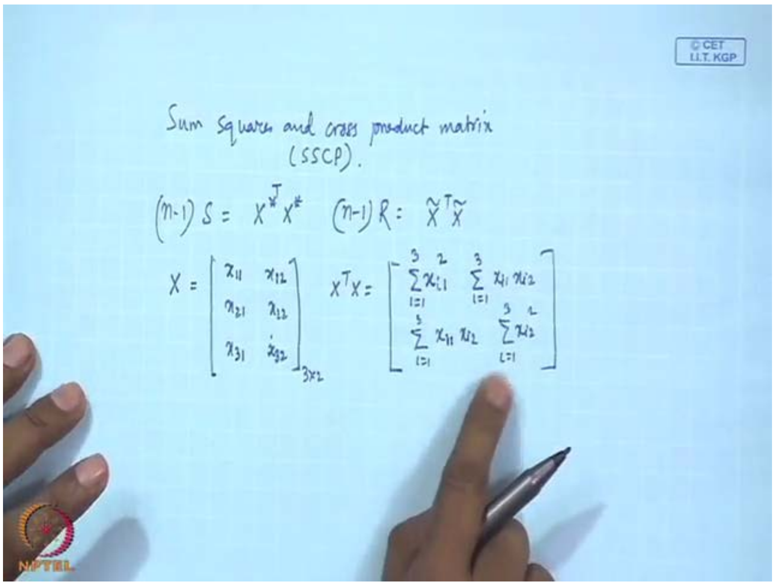


Suppose my covariance matrix is this, S 1 1 S 1 2 S 1 p 1 2 2 2 2 p like S 1 p 2 p p p. My correlation matrix is this 1 r 1 2 r 1 p r 1 2 again 1 r 2 p so like this r 1 p r 2 p 1. So, you create another diagonal matrix D s. This one is same p cross p matrix, this is p cross p this also p cross p, same p cross p matrix only the diagonal elements will be the variance component of diagonal will be 0. So, this is S 1 1 0 0 0 0 S 2 2 0 0 like 0 0 0 0 S p p. So, the diagonal elements of D s are the diagonal element of the covariance matrix, up diagonal elements are 0.

It will create like this and suppose you know S, you can just you with one trick you can find out that R is Ds to the power half S D s to the power minus half, both case minus half D s to the power minus half. If you use mat lab if use this mat lab now straight way, we will calculate all those things from the data, but suppose you want do the conversion, you mean in excel you can do this. What you are doing? Most of the time you may be knowing this one that variance component, once you know S you know the variance, co variance also, you want to calculate R.

Suppose, you know this variance component and correlation is known, correlation matrix is known, you want to go to co variance matrix from co relation matrix. What you have to do, you have to write like this plus R. So, this is basically from co variance to correlation and here correlation to covariance. Only thing you want to require in the second case, the variance component of all the variables considered.

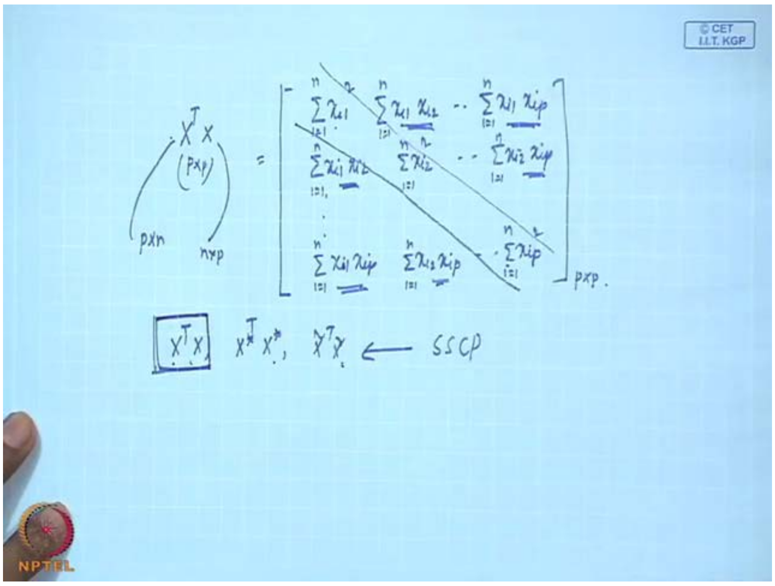
(Refer Slide Time: 40:31)



Another important concept in multivariate data analysis is sum square and cross product matrix, which is known as S S C P. So, if you see, when you calculate the correlation matrix, we are using the formula that one for n minus 1 cross S equal to X star transpose x star, we have used. We have used n minus 1 R equal to X tilde transpose, x tilde we have used where, both x star and x tilde are basically transformed matrix from the original data which is x.

So, suppose x I am writing like this, x 1 1 x 2 1 x 3 1 x 1 2 x 2 2 x 3 2, you do like this. Now, you calculate x transpose x, what will happen here, you will be getting like this x i 1 square i equal to 1 to 3, here the i equal to 1 to 3 x i 1 x i 2, here i equal to 1 to 3 x i 1 x i 2, here x i 2 square i equal to 1 to 3. I have taken 3 three cross 2, if you multiply we will be getting because earlier, we have seen subtracted by mean and for divided by standard deviation case, we have seen we got similar formula.

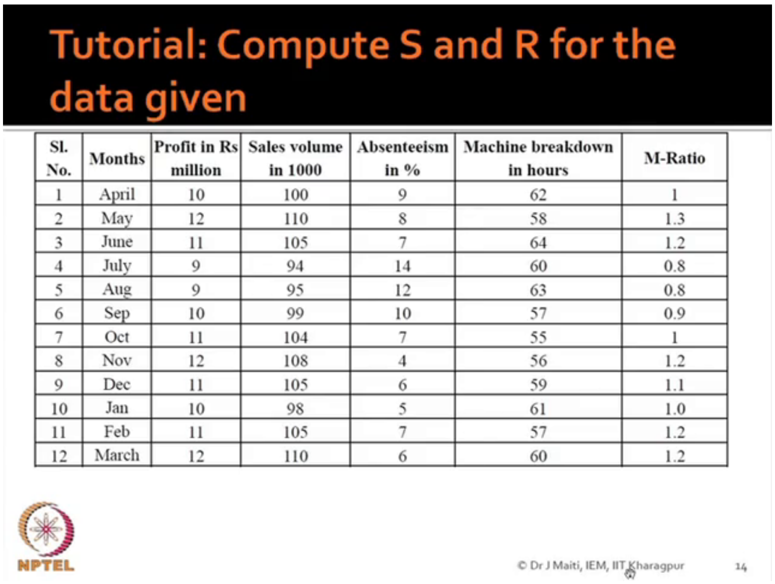
(Refer Slide Time: 42:45)



Now, the same thing if you think from the p cross p variable point of view then X transpose x will be a p cross p because x transpose this one is p cross n and this one is n cross p, you will be getting like this. So, your matrix will be like, this sum total x i 1 square then x 1 x i 2, then your x i 1 x i p, here x i 1 x i 2 here x i 2 square, then x i 2 x i p. In similarity, I am writing all those things x i 1 x i p x i 2 x i p then x i p square. So, it is a p cross p matrix then i equal i definitely equal to 1 to n all cases 1 to n.

This 3 matrices like x transpose x, x star transpose x star, x tilde transpose x tilde, these are all sum square and cross product matrix, all S S C P why? You see now, these are the some square, all diagonal elements are sum square and up diagonal you see cross product all up diagonal are cross product. So, sum squares for the variance cross product from the co variance that mean from this matrix also. Once you know these we can use these or these matrixes were ultimately, we can calculate the descriptive statistics like co variance and correlation matrix. This one is very-very important matrix, later on particularly in regression; you will be using this matrix. Now, let us see that how to calculate, suppose this is the problem given.

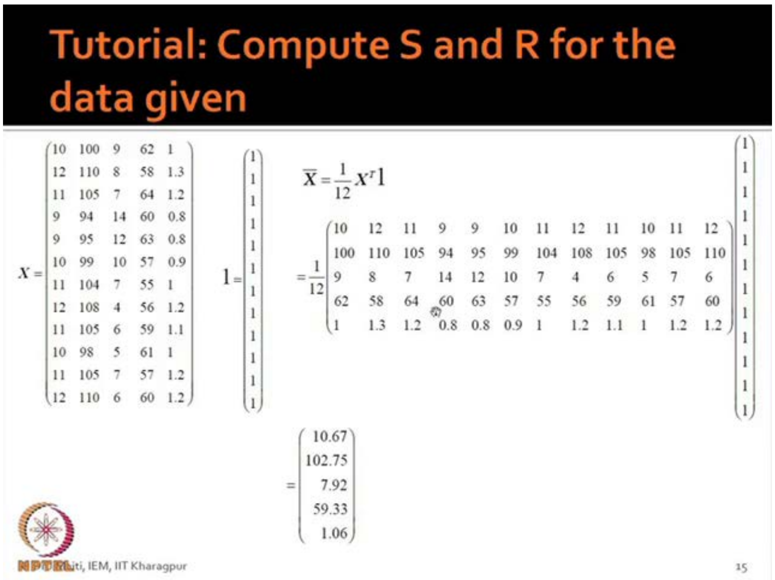
(Refer Slide Time: 45:24)



Sl. No.	Months	Profit in Rs million	Sales volume in 1000	Absenteeism in %	Machine breakdown in hours	M-Ratio
1	April	10	100	9	62	1
2	May	12	110	8	58	1.3
3	June	11	105	7	64	1.2
4	July	9	94	14	60	0.8
5	Aug	9	95	12	63	0.8
6	Sep	10	99	10	57	0.9
7	Oct	11	104	7	55	1
8	Nov	12	108	4	56	1.2
9	Dec	11	105	6	59	1.1
10	Jan	10	98	5	61	1.0
11	Feb	11	105	7	57	1.2
12	March	12	110	6	60	1.2

That compute S and R for the data given, this data said you have seen earlier so I have used excel only.

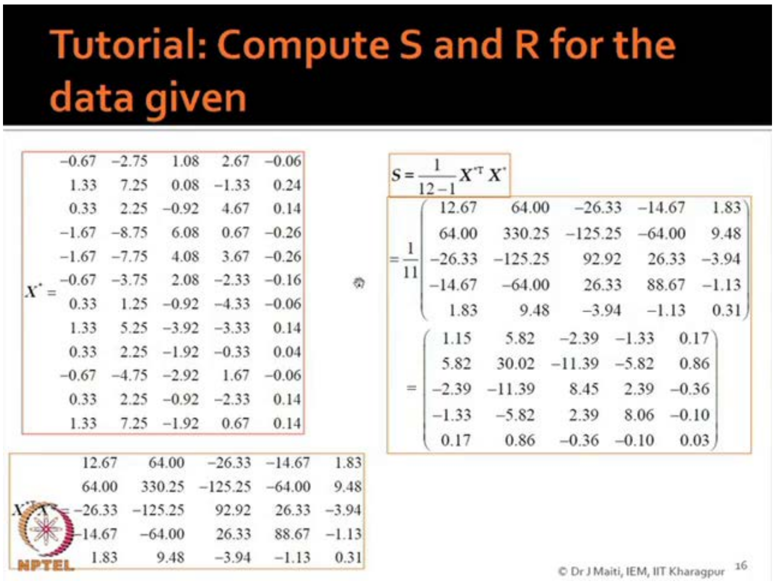
(Refer Slide Time: 45:41)



Using excel I have created so this is my data matrix, I want to compute the mean then as there are n data points so I created one unit vector with n data points. So, my aim is 12 here. So, it is 12 cross 1 vector then X bar is 1 by 12 X transpose 1 when I multiply all those things, I got this values. So, profit mean is 10.67, then sense volume 1002.75, 7.92

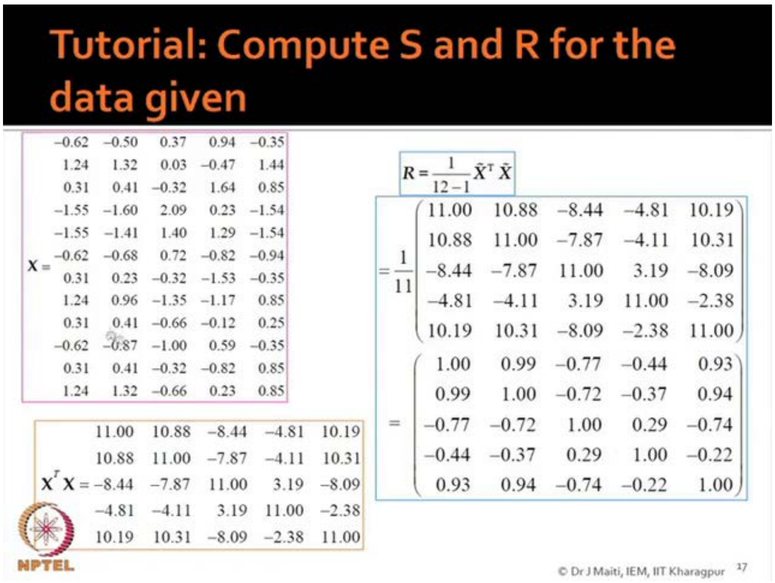
is your absent sing, then break down 59.33 and 1.06 is the m ratio case. So, your first step will be this, find out x bar.

(Refer Slide Time: 46:40)



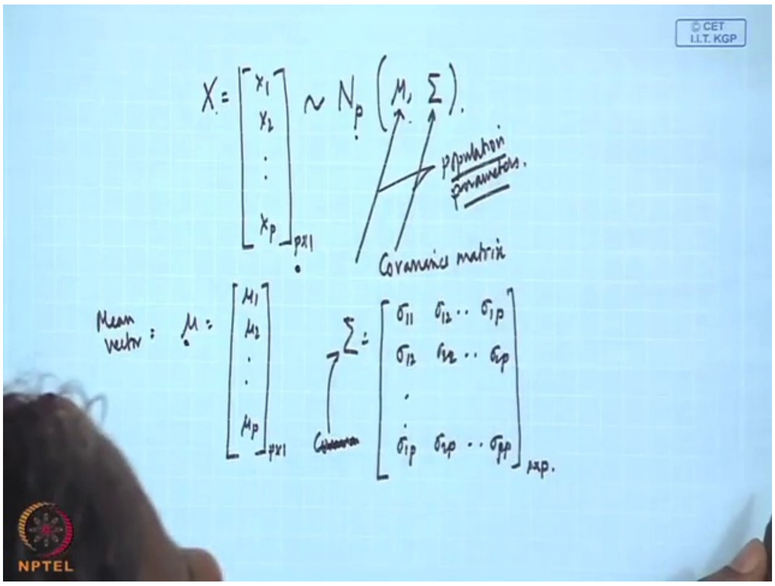
And you use this type of formulation, then what you require to calculate, you require calculating s. You require converting this x value to x star that means each of the suppose 10 minus 10.67 that is why minus 0.67 coming here X star. Once you get this, this is the formula S is 1 by n minus 1 X star transpose X star, this will give you this value.

(Refer Slide Time: 47:29)



On the left hand side the bottom portion, this is nothing but S S C P matrix X star transpose X star. You can do the same thing now; we are interested to know that tilde. Here this is basically X tilde, this one X tilde transpose x tilde, this transpose part is missing x tilde and R is 1 by 11 into X tilde transpose X tilde, you are getting like this. You see once you go by this way calculating, we will get all the diagonal elements 1, if you do not get that, there is a problem. You have any questions so far now. Although the next class I will be explaining in detail, when we will start that multivariate normal distribution. What we have assumed here? We have assumed here is x that is a variable vector x 1 x 2 dot dot x p, that is P cross 1.

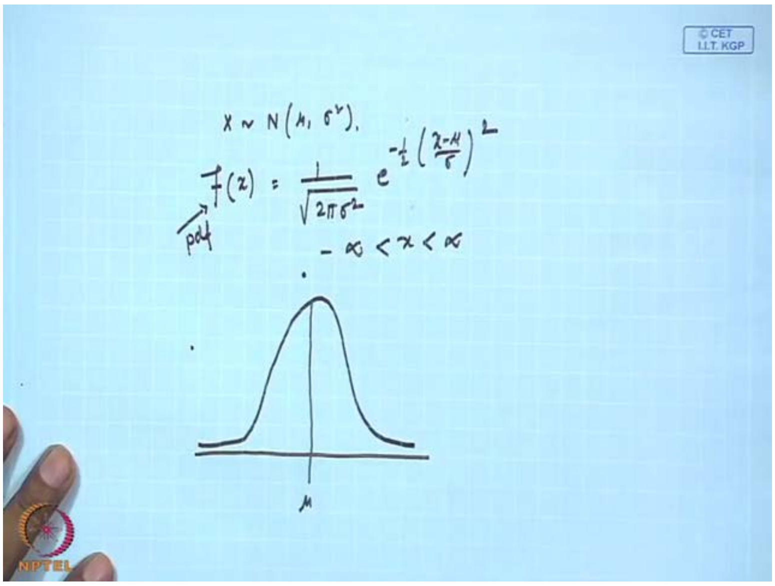
(Refer Slide Time: 48:34)



And we assume that this follows normal distribution that is multivariate normal, which will be denoted by like this n p and mu and sigma. Now, you are well accustomed with the nomenclature. Nomenclature in the sense you know that mean is mean vector, if there is p cross 1 variable, then my mean is again p cross 1 that is mu 1 mu 2 mu p. And you also know this one, this is nothing but covariance matrix so this covariance matrix is our p cross p matrix sigma 1 1 sigma 1 2 sigma 1 p sigma 1 2 sigma 2 2 sigma 2 p sigma 1 p sigma 2 p sigma p p.

Multivariate normal distribution is characterized by p variable with parameter that is mu that is a mean vector and covariance matrix. So, please remember these are population parameters, these two are population parameters. So, far we have not discussed about that whether the data is coming from multivariate normal or not but ultimately, we will be going to multivariate normal distribution because most of the models that will be relied on these assumption multivariate normality. Now, when p equals to 1 that is univariate normal, and now what is the probability density function of univariate normal.

(Refer Slide Time: 50:53)



Suppose x is a random variable, which is univariate normal with mu and with sigma square. Then if I want to know, what is your probability density function of x, this is your f x p d f, you will write 1 by root over 2 pie sigma square e to the power minus half x minus mu by sigma square, where minus infinite less than x less than plus infinite. So, this is what you have seen earlier that this is what our normal distribution is. So, what will be the equivalent distribution when number of variable is more than 1 that will be our starting point in the next class.

Thank you.