**INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR**

**NPTEL
National Programme
on
Technology Enhanced Learning**

Applied Multivariate Statistical Modeling

Prof. J. Maiti

**Department of Industrial Engineering and Management
IIT Kharagpur**

Lecture – 08

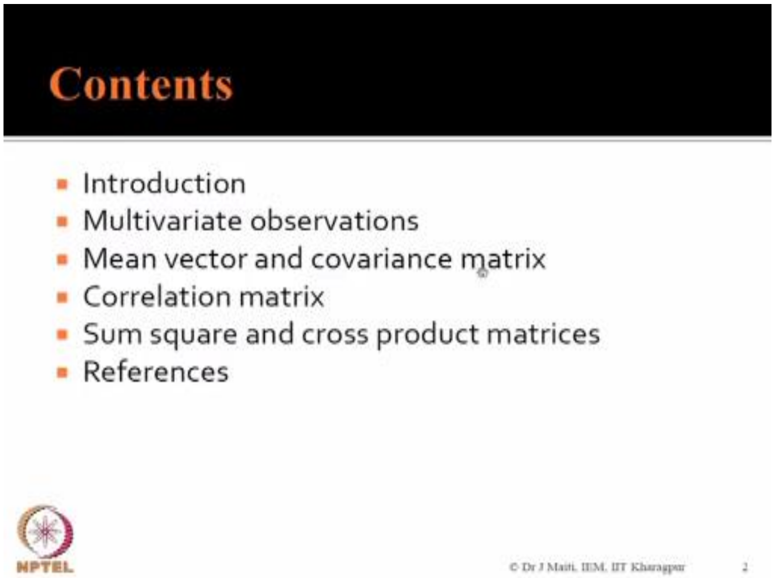
Topic

Multivariate Descriptive Statistics

Good afternoon, today we will formally enter into the multivariate domain. Now, topic is multivariate descriptive statistics.

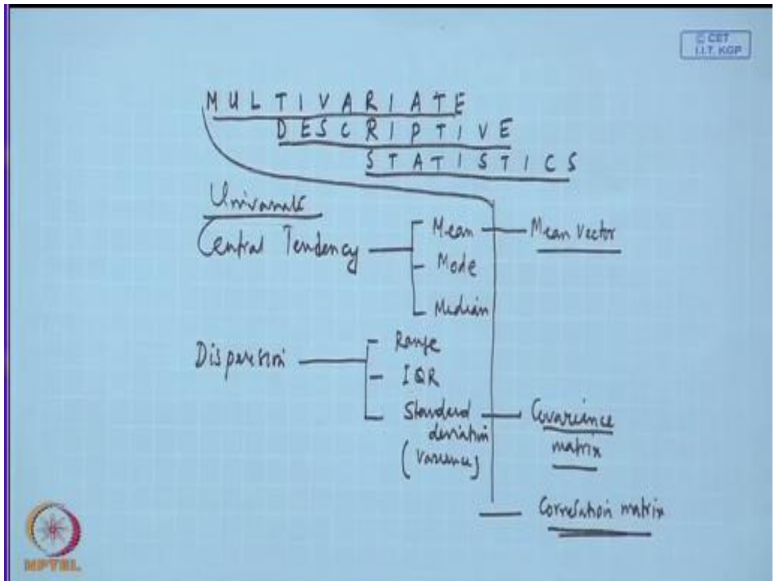
(Refer Slide Time: 00:32)

(Refer Slide Time: 00:35)



And the content of today's presentation includes multivariate observations, mean vector and covariance matrix, correlation matrix, sum square and cross product matrices. So, it will be purely data, we will be dealing with data with symbols like x so, in univariate statistics.

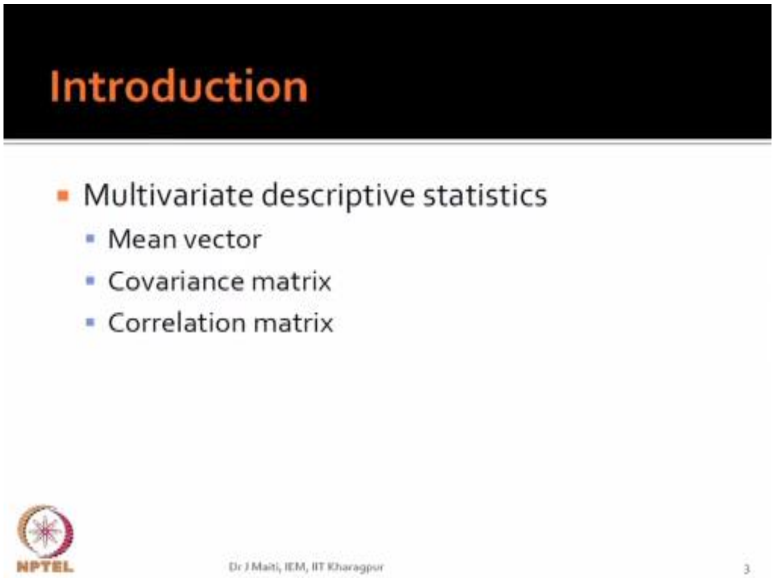
(Refer Slide Time: 01:05)



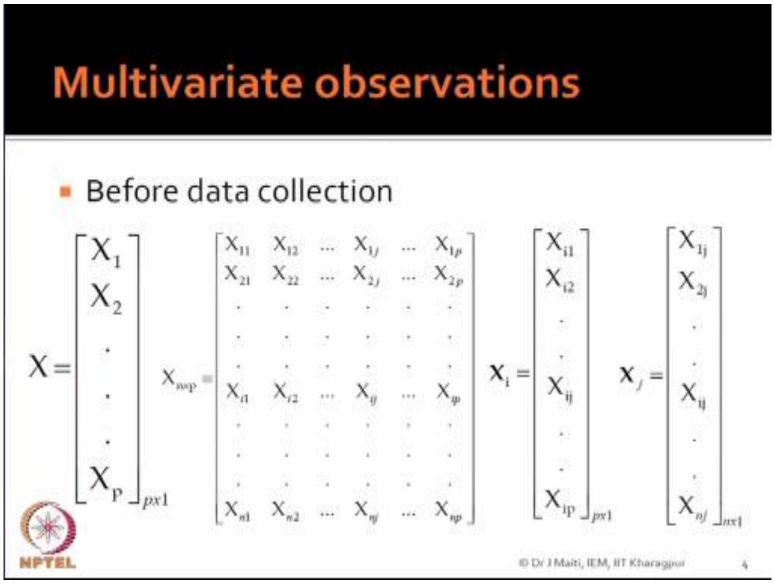
Univariate descriptive statistics you have seen central tendency measures as well as dispersion measures, correct univariate case you have seen earlier in this lecture classes. Now, central tendency we have measured in univariate case using mean, then your mode and median, these are the measures we have adopted. And under dispersion, we have used range then I think IQR Inter Quartel Range as well as standard deviation okay now, we will see some of the counter parts in the multivariate domain.

For example, mean will be mean vector when you go for this multivariate statistics site, you will find out that mean will be mean vector. And this standard deviation another component, which is the measure of dispersion that will be not standard deviation it is square that is the variance, variance will be a covariance matrix. In addition as there will be more than 1 variable so there will be correlation between variables, so we will be knowing correlation matrix also. Today’s discussion will be concentrated on mean vector, covariance matrix and correlation matrix.

(Refer Slide Time: 03:27)

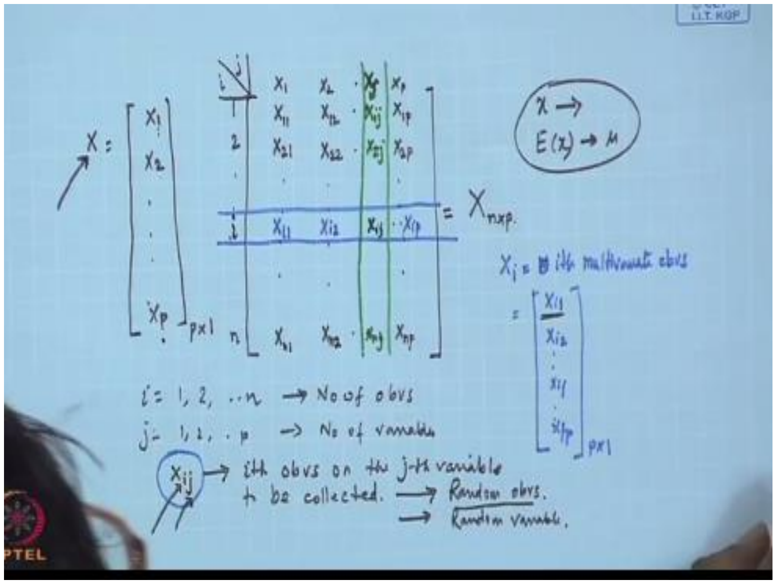


(Refer Slide Time: 03:30)



Okay now, think a little bit on abstraction level now that there is a multivariate population and that population is characterized by a variable vector.

(Refer Slide Time: 03:46)



Which is known as x there is p number of variables, which is characterizing the population of interest, so it is a p x 1 vector. What does it mean? We are saying we have created a vector X, which is p x 1 vector where, p stands for number of variables so there are variable 1, which is X_1, variable 2 is X_2 like variable p is Xp okay now, you think that you require to collect data on this p variables.

So I am writing here my first variable is X_1, second variable isX_2 like my last variable is X_p. And you are basically collecting data on this p variables, so you will be collecting observation 1, observation 2, like this n observations. So, essentially you are not in univariate domain, you are in multivariate domain and you are not only in 1x with several that n values that means 1 to n values.

Here you are with i=1 to n and j = 1 to p, so what does it mean i = 1 to n, these are the number of observations and j = 1 to p, these are the number of variables okay now, you think of a situation that you have your population you know that variables p variables are there, you have identified the variables now, you are planning to collect data you have not collected data you are you are planning to collect data then our nomenclature the way we will be writing here we will be writing like this. The general observation will be x_{ij} what does it mean? x_{ij} means that is the i^{th} observation on the j^{th} variable.

i^{th} observation on the j^{th} variable that is why i stands for 1 to n, which is number of observations, j is 1 to p number of variables x_{ij} which is the i^{th} observation on the j^{th} variables to be collected. If this is the case that means we are first writing the variable then, we are writing what is the observation number, and then we are writing what is the X_{21} so like this you are writing observation n variable1.

My observation on the second observation my observation on the second variables will be X observation 1 variable 2 X observation2 variable 2, so like this X observation n variable 2.So, in this manner you will go then X observation number 1 on variable p, observation number 2 on variable p like this you will be getting observation number n on variable p, this is the data matrix. And we will be denoting this as X boll X and the order will be n x p is visible okay this one I am saying this is a data matrix, which you are planning to collect.

If this is the case and you all know that X is a random vector here X_1 to X_p, X is random vector because all the variables are random variable here. So, X_{ij} is the i^{th} observation on j^{th} variable, this is also a random variable. Okay this one please keep in mind this is also a random variable we are writing it as random observation. So, as you have not collected data you have just planning to collect data any value of X_{ij} can be found depending on the spread of that variable you will be getting any value of this, but you do not know what that value. So, that is why we are saying it is a random it will be random observation and that X_{ij} is a random variable will be random variable.

Okay so, you cannot expect value of X_{ij} getting me you will get you will get you have some expectation, so that is why you will see in univariate case that if X is random variable then, expected value of X is μ. So this is the case then for every variable here, it has also some expected value, so that is what we will basically talk about the mean vector.

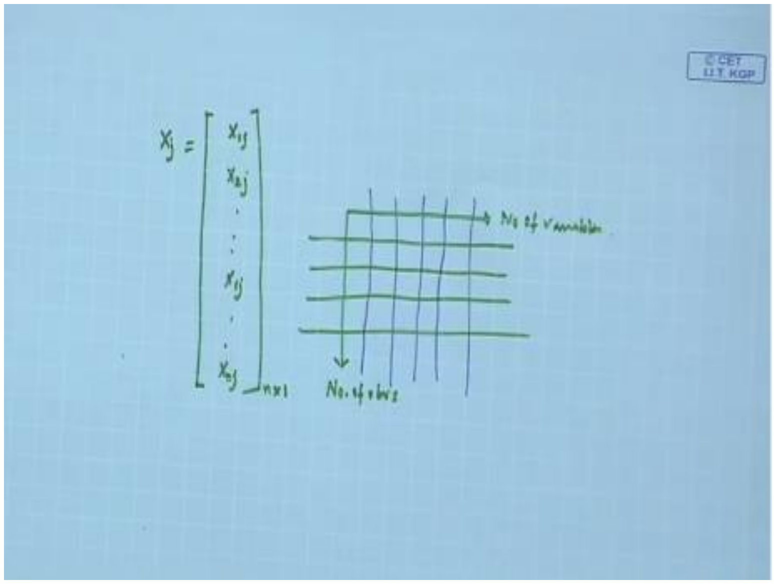
Okay so, before coming to mean vector another two important issues is there that to be discussed here. Suppose the observation is i^{th} observation and there is the variable is x_j, this also what will be there I told you that the general what I have given you here that this is the general observation x_{ij}. So, similarly, there will be a general observation, multivariate observation, general observation similarly, there will be a general variable.

Now, if suppose this one is i, this is the i^{th} multivariate observation So, your value will be X observation number i variable 1, observation i variable 2 like this if I go observation i on variable j then, observation i on variable p. So, if I write that X_i is the multivariate i^{th} multivariate observation can I not write like this that X_{i1}, X_{i2}, X_{ij}, X_{ip}, this is a p x 1 vector correct now, if I go by the variable wise so that means when I talk about the i^{th} multivariate observation, please keep in mind that in a particular observation on all the variables.

Why it is it that it is not that what is the in the univariate case what will happen you will get the one value only for that observation. As it is multivariate, you are getting all values on p variables, but the other important thing is that that means all the variables are occurring simultaneously, simultaneous occurrence that is why multivariate in nature. Now, if I go by our general variable which is x_j then, this observation will be X observation1 on variable j, observation 2 on variable j and observation i on variable j similarly, observation n on variable j.

So, you will be getting a general variable that means observations on a particular variable that we will write. So, I if we write like this.

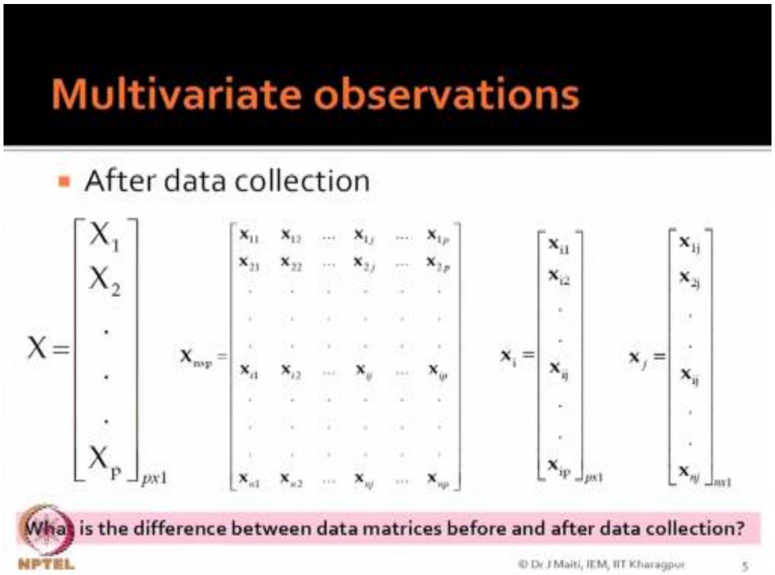
(Refer Slide Time: 13:35)



That x_j this is what we is the that is then observation on n observations on j variable then this is x_{1j}, x_{2j} like this x_{ij} then, x_{nj} it is a n x 1 matrix, So, that means essentially what is happening? You have one hand this side that the observations number of observations axis and the other hand you have number of variable axis. So, when you go row wise, you are basically talking about different multivariate observations and when you go column wise that means you are talking about n number of observations on different variables, that observation on n observation on variable 1 to variable p.

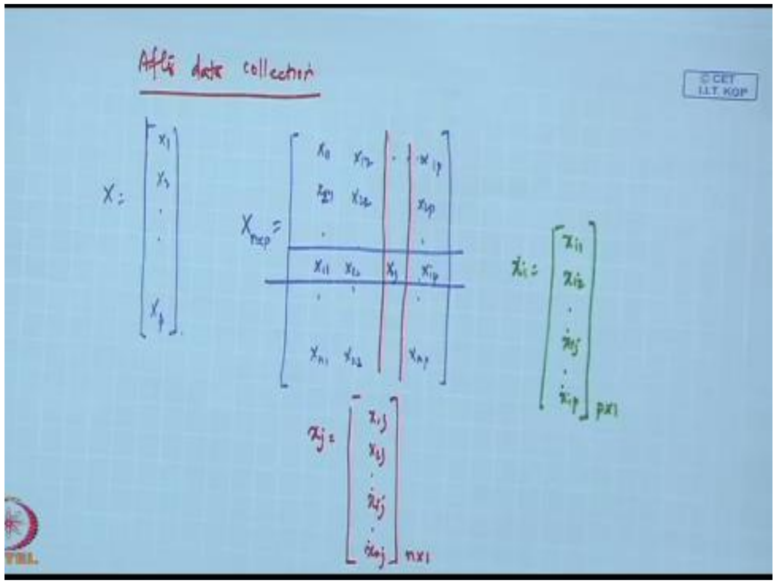
Okay so but what we have assumed here we have assumed that we have not collected the data, we are planning to collect data and as a result all the entries in this particular matrix are random in nature. What will happen if you collect the data?

(Refer Slide Time: 15:21)



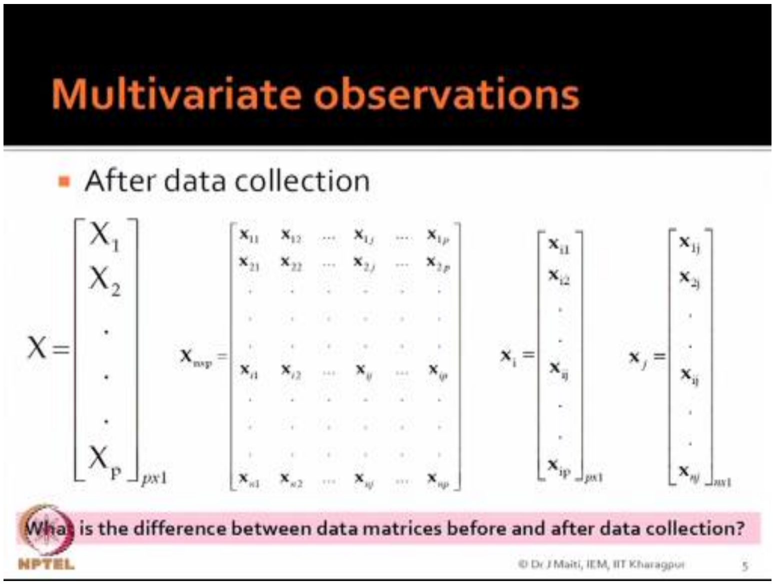
When you will collect data something you will collect data on.

(Refer Slide Time: 15:29)



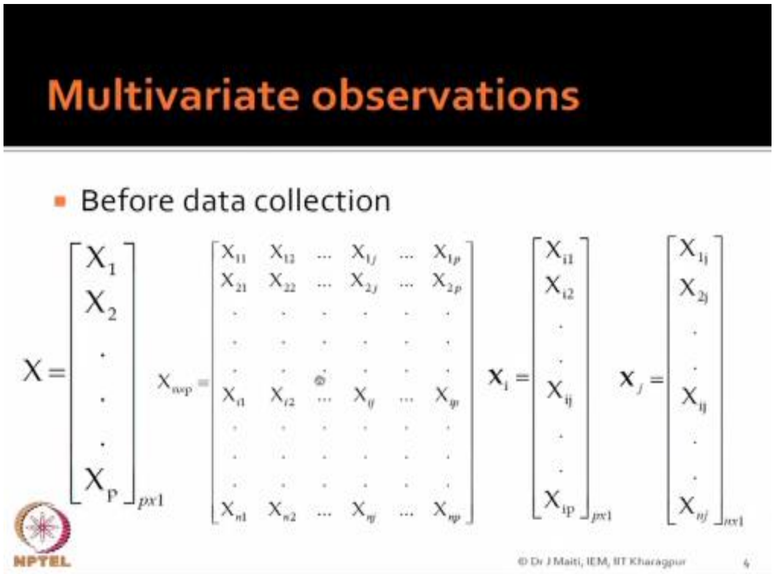
Same number of variables X_1 X_2 like X_p and your data matrix also will be n x p here also you will be getting i = 1 to n like X_{11}, X_{21}, X_{n1}, X_{12},X_{22},X_{n2} then X_{1p}, X_2p, X_{np}, this you will get. Here also you are X_{i1} then X_{i2} somewhere X_{ij} then X_{ip} that general observations will be there, which we will be able to write like this that X_i, which is X_{i1}, X_{i2}, X_{ij} then X_{ip} that is p x 1. And also you will be getting one general variable that observation we write that X_j. If we write, you will be getting X_{1j}, X_{2j} like this X_{ij} then, X_{nj} all remain same, it is after data collection. Then my question is, what is the difference between the first matrix and second matrix this is our second matrix.

(Refer Slide Time: 17:23)



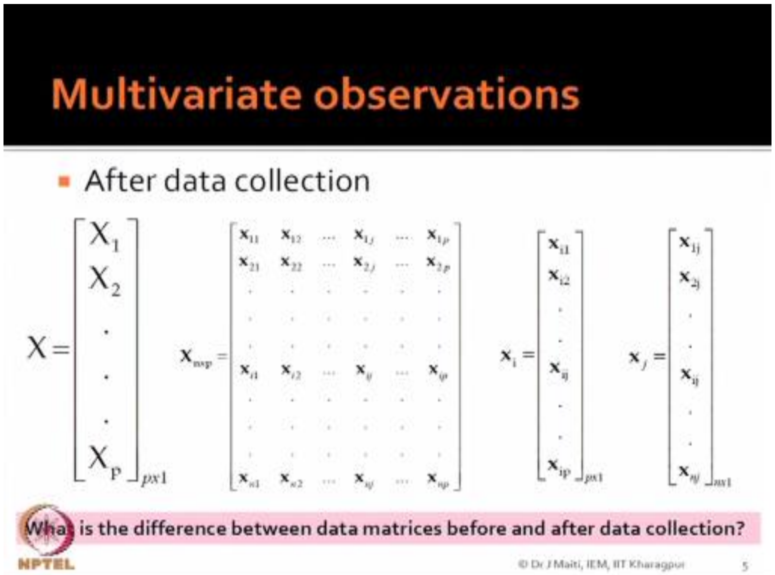
Where we said that after data collection.

(Refer Slide Time: 17:27)



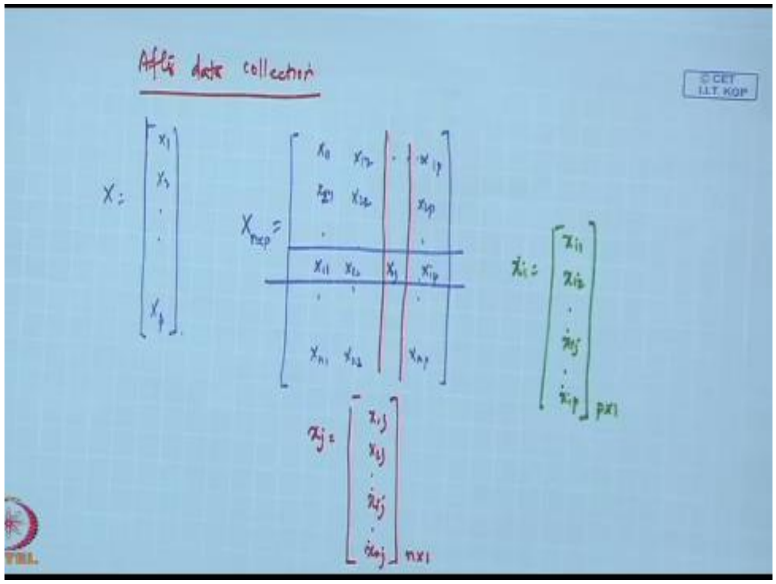
And this is our first matrix we say this is before data collection.

(Refer Slide Time: 17:30)



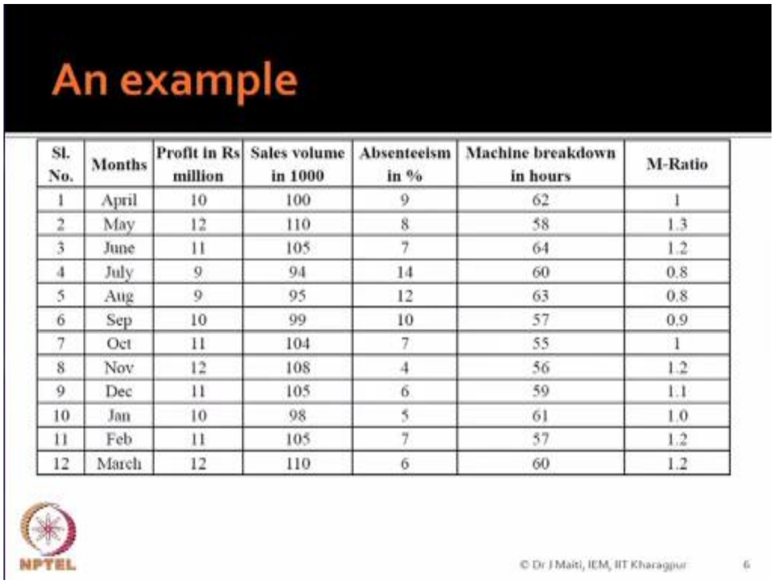
What is the difference a second matrix values are known.

(Refer Slide Time: 17:37)



So it is either fixed values now from the data collection point of view, they are fixed values. First matrix you do not know what value you will get, that is why in the first matrices when you do not know anything, you will expect some value for each of the variable that mean, you also expect some deviation of different values from the mean that will be your variance and you also expect the 2 variables will be co vary, that will be your covariance, that is from the population point of view.

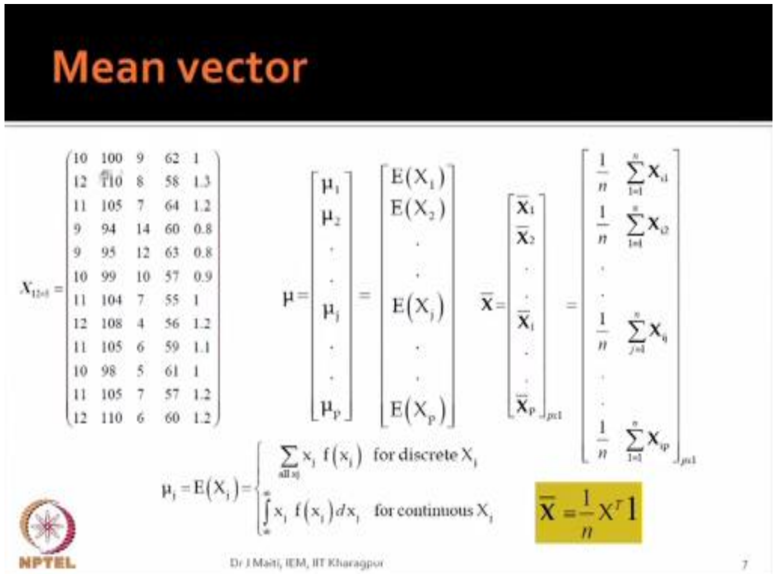
(Refer Slide Time: 18:20)



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

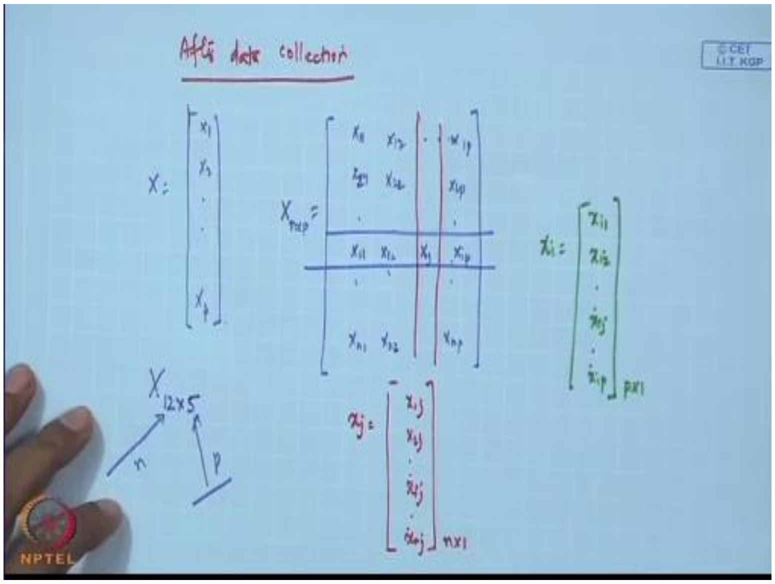
Now, let us see this data set okay now if I ask you what the x matrix, data matrix is. So, what is the n value, n is 12. Now, months we are excluding for the time being, although month can be in two data's variable. For the time being let the month is excluded then1, 2, 3, 4, 5 so your data matrix is 12 x 5, all the variables are measured for 12 observations. So, if you see this.

(Refer Slide Time: 19:07)



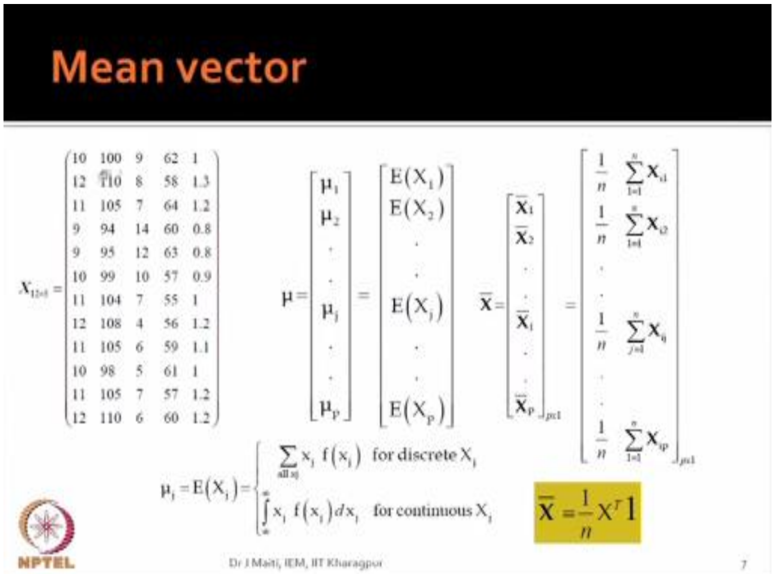
Then see this is the data matrix, this one 12 x 5. Oh this will not come here 12 x 5. So, the left hand side matrix, the data matrix is x.

(Refer Slide Time: 19:24)



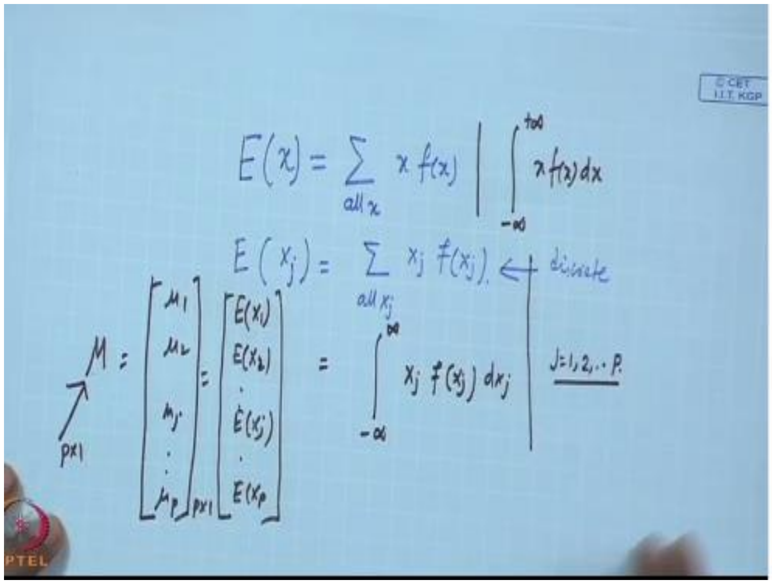
12 x 5, 12 stands for n, this stands for p okay now we will just find out our mh first we will see the expectation of the variable values, for each variable what is the expected value, then we will see that when we collect data, what that value is okay now, you see this slide.

(Refer Slide Time: 20:04)



That as there are p variables, so there are p means these are population parameters this μ is the vectors, which represent p means for the p variables, and it is mean vector and which is a parameter vector for the population and I have written here that it is expected value of X_1, expected value of X_2, expected value of X_j, expected value of X_p and you have already seen that.

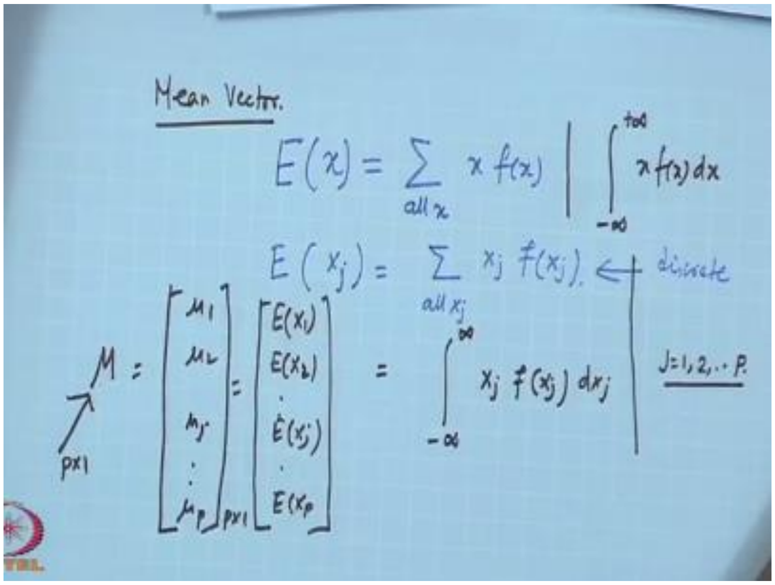
(Refer Slide Time: 20:43)



What is the expected value of X, what we have written earlier if your variable is discrete, you give a summation and then you say all x, xf(x). Here what happened we have basically so many variables at n we are writing like this expected value of X_j stand for the variable for a particular variable j^{th} variable, then we can write all $X_{j,}$ X_j $f(x_j)$, when it is a discrete variable. When it is continuous what will happen, what you write in continuous case here you have seen continuous case, you write integration - 2 + ∫xf(x)dx.

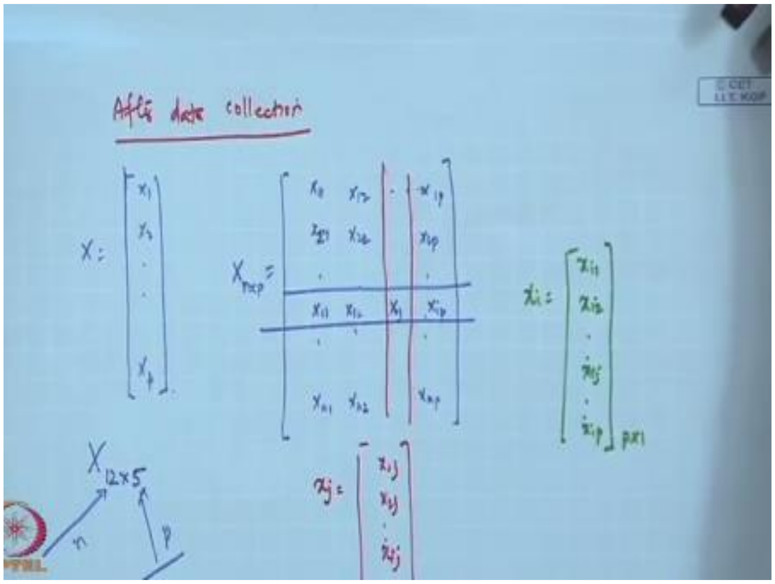
So, I am writing here for continuous case ∫ - ∞+∞ X j f X_j $f(x_j)$ dxj. So, for both the cases j = 1 to p, this is your expected value. So, when you write down for μ what is μ is a vector p x 1 vector, which is μ 1, μ 2 like μ j then your μp, which is p x 1 vector, this is the this is nothing but as we discussed now expected value of X_1, expected value of X_2 then expected value of X_j, then expected value of Xp and expectation we will calculate like this correct. So, where you collect data we have sets of matrices, 1 matrix we say that we have not collected data.

(Refer Slide Time: 23:07)



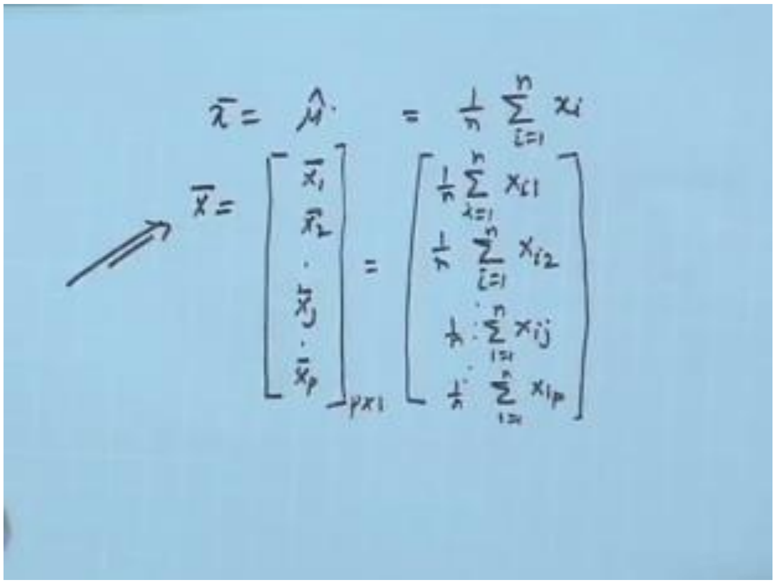
That is before data collection with respect to this we are developing this one, this we are saying that our topic now is mean vector. When you collect data like the 1 I have already given you that 12 x 5.

(Refer Slide Time: 23:38)



Then our data matrix is this after data collection. And we want to compute the average value only because we have seen in univariate case, univariate case what you have seen we have also seen that $\bar{x}$ is the estimate of μ.

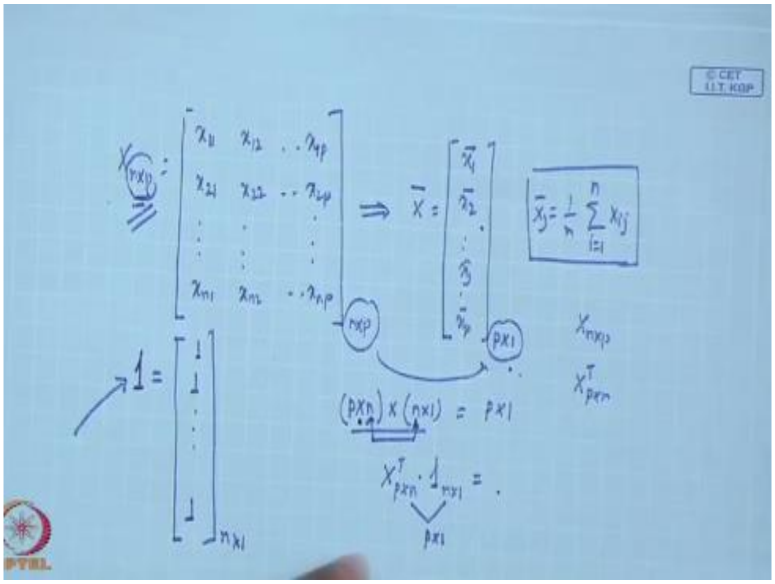
(Refer Slide Time: 23:55)



So, in our case we are writing $\bar{x}$ which is nothing but $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_j$, $\bar{x}_p$ that is p variable average vector .Now, what is your $\bar{x}$ in case of mean variant you have written 1 / n, sum total i = 1 to nX_i, So, that mean I can write this one now in the same manner that this one is i = 1 to n 1 / n i sum total i = 1 to n X_{i1} I can write, 1 stand for the variable i stands for the observation. Similarly, second one you can write sum total of i = 1 n X_{i2}. So, in that same manner for the j^{th} case, you can write i = 1 n X_{ij} and the last one you can write i = 1 to n $X_{ip,}$ this is my mean average vector for the sample collected on p variables.

Okay so, we will not go for individual mean calculation the average calculation instead of doing this, we want to do matrix calculation, vector matrix that matrix algebra calculation what we will do here see I have in order to calculate this $\bar{x}$, this $\bar{x}$ we will take this data matrix fast what is our earlier data matrix was there this data matrix was.

(Refer Slide Time: 26:12)

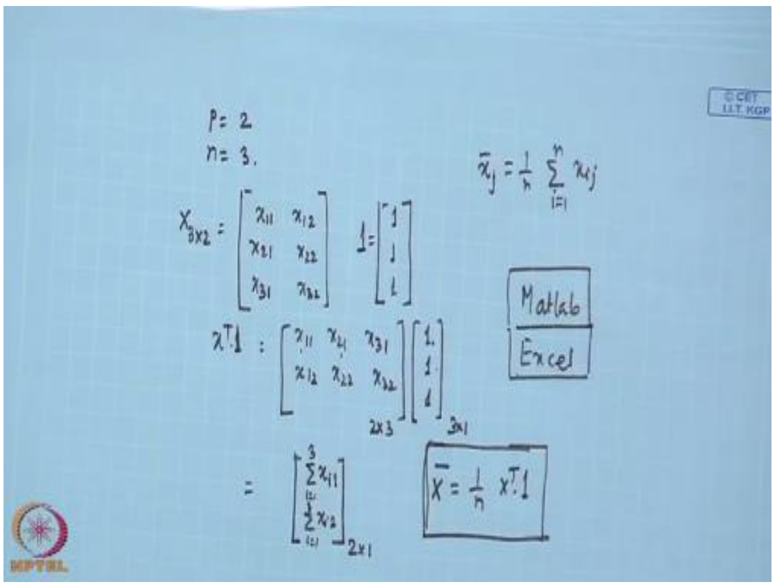


$X_{n \times p}$, which we say suppose X_{11}, $X2_1$ like this Xn_1, $X_{12,}$ X_{22} then your X_{n2}, So, like this X_{1p}, X_{2p} your X_{np} that was my sample data and it is n x p you want to compute $\bar{x}$ from this sample data and $\bar{x}$ is a p x 1 vector, that is $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_j$, $\bar{x}_p$ so and you know the formula also you know the general formula suppose, $\bar{x}_j$ is 1 / n sum total i = 1 to nX_{ij} that also you know. Now, in matrix in1 go you want to calculate all this things, what you require to do that means when you calculate in matrix domain, please remember I have n x p matrix I want a p x 1 matrix. I am going from suppose this is n x p.

So, you have nx p you are going to p x n that means what if I create one matrix, which is let it is 1 big symbol one which is all ones 1, 1 like this, there will be n ones n x 1. I am creating 1unit vector where all there are n elements in the vector and each element is one only. Getting me now, this I want to use this one, this unit vector with this data matrix in such a manner that I will be able to apply this computational formula and I will be getting the p x 1 vector. So, that means if I create like this suppose, p x n into n x 1 it is p x 1 from matrix multiplication point of view row column.

So, here number of column is number of row equality is there if I want to do this that means I have to transpose this matrix. I have a matrix called X with n x p, if I do transpose x^T this will be p x n row and column will be interchanged. So, I am doing this x^T this is p x n, I will take a dot product that is n x1. So, your resultant matrix will be definitely this will be cancelled out and it will be p x 1. Okay so, what will happen with this, you see now.

(Refer Slide Time: 29:55)

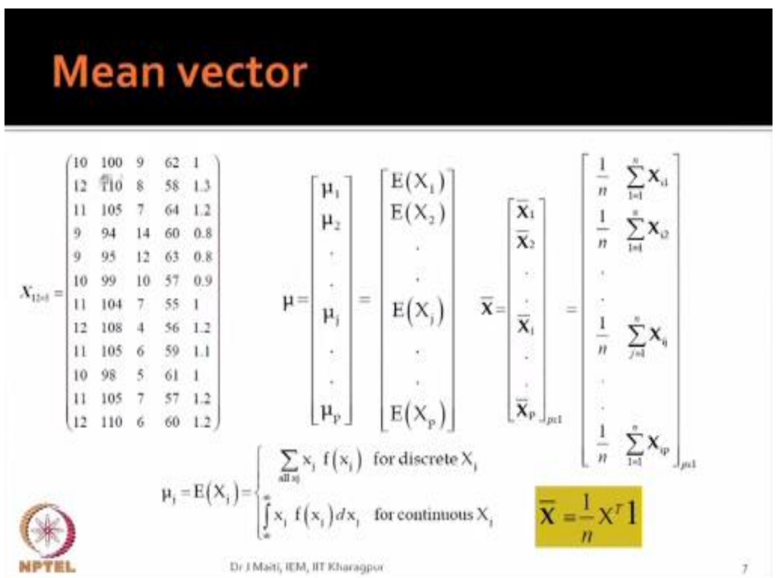


So, your data is like this so I am taking now let p = 2, n = 3 just to reduce the complexity of repeating the same calculation. Then my x matrix will be 3 x 2, which will be like this x_{11}, x_{21}, x_{31} then x_{12}, x_{22}, x_{32} as there are 3 observations. So, let you create 1 unit vector, which is 1 1 and 1 this 3 is required because my n is 3. So, I want to multiple x^T1,So, what I will do then x_{11}, x_{21}, x_{31} that is the first row, second row x_{12}, x_{22} and x_{32}.This is my second row because I have made the x matrix transpose, so then you multiply this by 1 1 and 1. And we all know this one is a x^T is 2 x 3 and 1 is 3 x 1.

So, we will be getting a resultant matrix, which is 2 x 1. So, then matrix multiplication point of view x_{11}x 1+ x_{21} x 1 so that is what i = 1, 2, 3 X_{i1}, second one will be i = 1, 2, 3 X_{i2} now, we have seen earlier that what is $\bar{x}_j$ that is 1 / n, sum total i = 1 to n x_{ij} So, if I divide this resultant thing by n, I will get the mean vector okay so, that means I can write $\bar{x}$ is 1 / n x^T this one or you will be able to do it very easily because this formulation is much better because you can multivariate domain hands on that calculation type you forget. You have to use Mat lab for understanding the computational part.

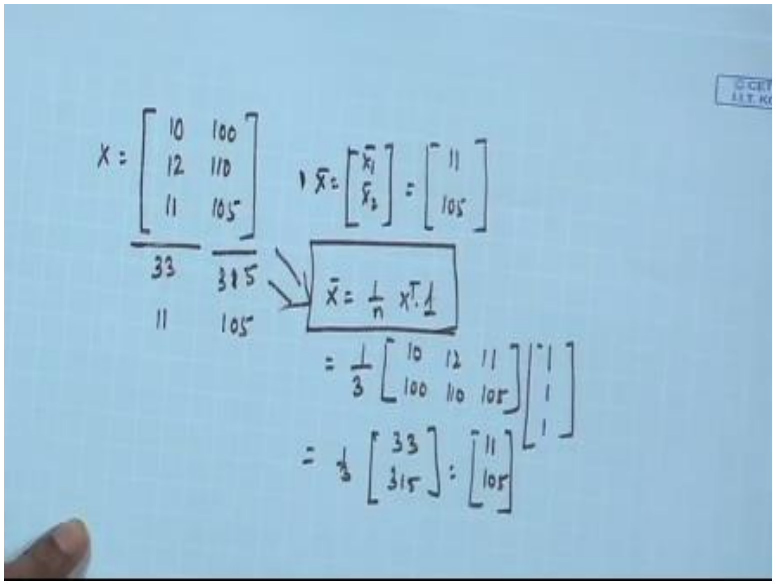
You can now a day excel is very powerful also so excel also you can use. Using excel you can use this formulation very easily I think I will be giving you tutorial and you will have to do this and then if you find problem talk to me again in my chamber also that is no problem okay so, will we be able to compute the mean vector log given data. Now, see this slide.

(Refer Slide Time: 33:24)



Suppose you take these first 3 values for the first variable and also first 3 values for second variable and use this that matrix multiplication.

(Refer Slide Time: 33:41)

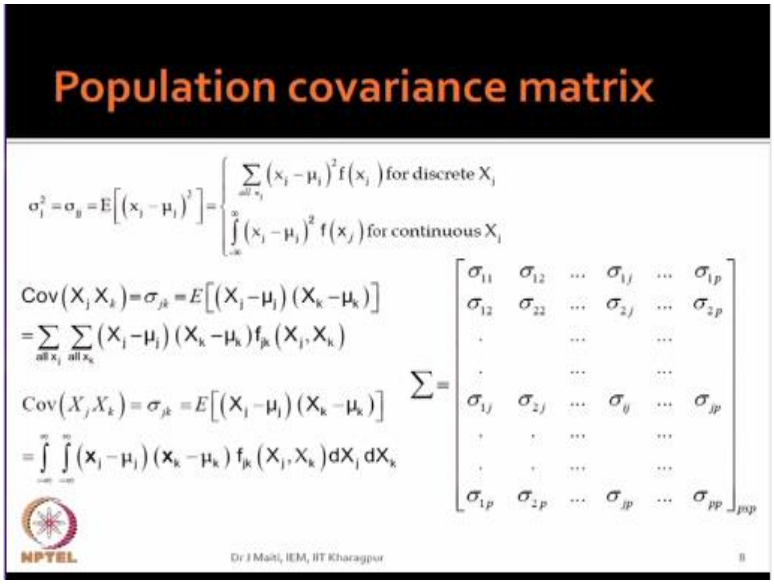


Can we not do what I said that we are taking only 2 variable case with 3 observations 10, 12, 11 then, 100,110 and 105. So, I want to get the $\bar{x}$ what we are saying $\bar{x}_1$ and $\bar{x}_2$. You can very easily you can go like this okay 10 that is 33 that total is 33, here the total will be 5 that will be 1, 0 1 1 1 and 3. 3 1 5. If you divide it by 3 it will be 11 that first one variable 11 and the second one will be 105.So, that means your $\bar{x}$ is 11 and 105 that you can very easily do here also, but I am saying that you use this formulation $\bar{x}$ = 1 b/ n x^T1.If you do like this 1 / 3 x^T will be 10, 12, 11, 100, 110, 105 and then 1 1 1 this is nothing but 1 / 3 into 33, 315 which is your same thing 11, 105.

It seems for that means here and here there is not much of difference in calculation. The reason is because of number of observation is also less number of variable is also less, but you have to have p means large value of number of variables with large number of observations. So, that individual calculation is not required straight away go for matrix multiplication. We will be using simple matrix multiplication, matrix inverse and other things in throughout I can say the lectures.

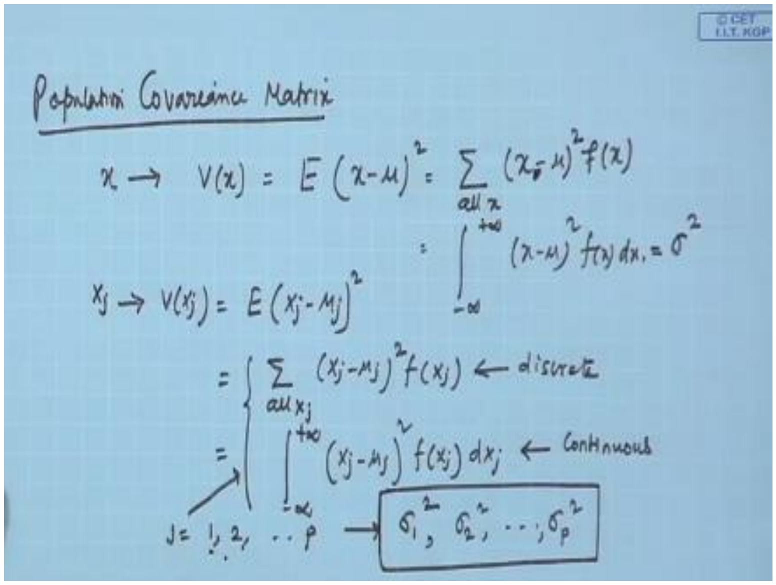
So, this is what is mean vector and from the population point of view and from sample point of view from sample point of view sample average is the estimate of population mean vector okay

(Refer Slide Time: 36:23)



Now, come to covariance matrix see you have to understand it now. Although things are very simple, but physical meaning of each of the items must be understood then only later on we will be talking about covariance matrix. We will not come back to the physical meaning of covariance matrix further, we will simply say covariance matrix then you will be able to catch what is covariance matrix immediately then only you will be able to relate the discussion that time okay. So you we are interested to know covariance matrix.

(Refer Slide Time: 37:06)

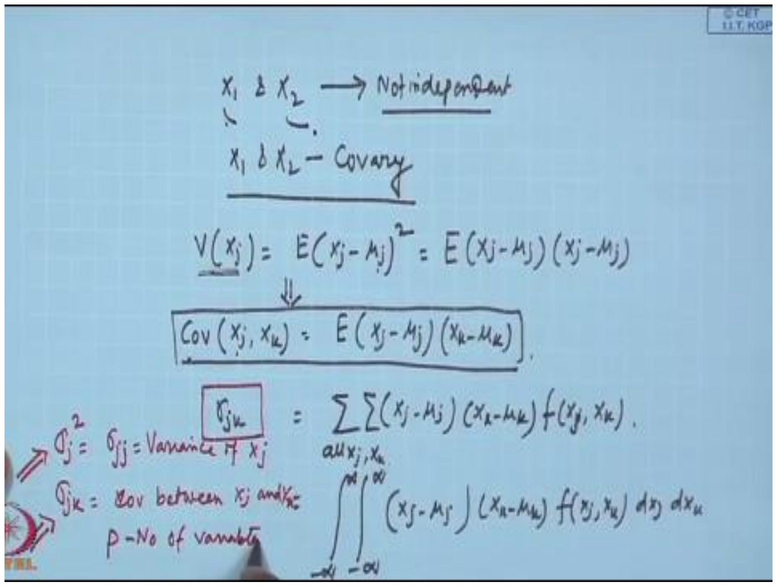


So, let us discuss from population point of view first, that is population covariance matrix correct if I say my x is a random variable univariate case then, if I ask you what is the variance of x then you will say that it is expected value of $x - \mu^2$ yes or no and you have also seen for your discrete case that all $x - \mu$ or $x_i - \mu^2$ then, f(x)i we are not using now let it be general case like this so that is why I have written all x.

And when you go for integration you write like this plus this $x - \mu^2$ f(x)dx, this is what this is the variance component, which is σ^2 . So now, I will do simple alteration instead of x will write x_j that means I want to know the variance of x_j I then you will write it is nothing but you will write $x_j - \mu_j$, only j will be the added there everywhere, then what you will write you will write sum total of all $(x_j - \mu_{j)}^2 f(x_j)$ this is for discrete case. And you write like this integration -∫ to +∫ $(x_j - \mu_{j)}^2 f(x_j)\, dx_j$, this is your continuous case.

Alright fine now, if this is the case and if I say what that means for j equal to if you put j =1, 2 like p in this formulation whether it is discrete and continuous, when you put j = 1 what you get, you will get σ_1^2 . When you put j = 2 you get σ_2^2. So, like this you will get σ_p^2 that means the variance component of all the variables coming from this equation, but we have seen that we have p number of variables. And what we are also assuming that these p numbers of variables are not independent to each other.

(Refer Slide Time: 40:47)



If x_1 is dependent on x_2 or x_1 and x_2 are not independent, what will happen for example, height versus weight of people? Nonzero, so that means they have correlation or otherwise I can say there is very much if someone height is more than other one it is naturally that weight also more naturally. But there are other parameters also, which govern that weight but naturally this is the case. So, when they are not independent a not independent they are dependent then what will happen.

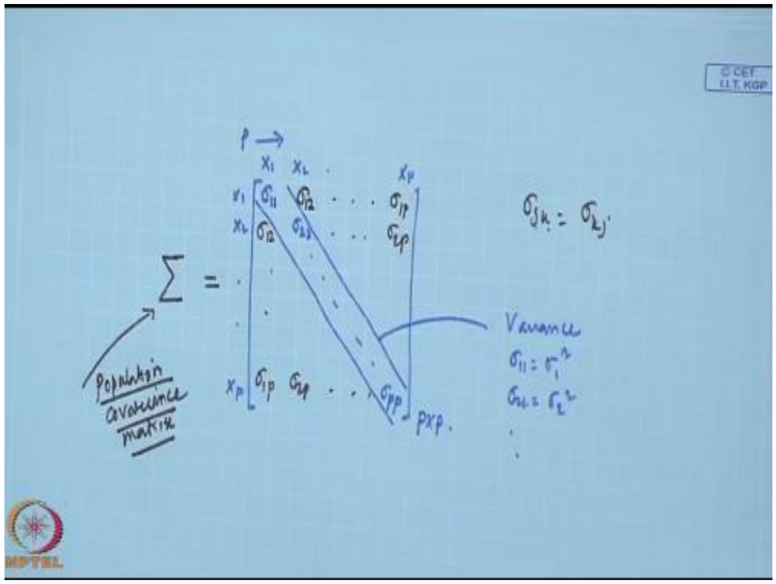
That means when x vary x_1 vary, x_2 also vary. So, there are simultaneous variability is known as what we want to say x_1 and x_2 co vary, they simultaneously vary if there is covariance means association between their realization of values of x_1 and x_2 fine so, now I will write again variance of xj we have written expected value of $(x_j - \mu_j)^2$ it is basically what this one if I just do some manipulation and if I write like this that there is a formulation called covariance between x_j and x_k, if I write like this $x_j - \mu_j$ x $x_k - x_j - \mu_k$ you see the similarity between this 2, when I am talking about variance of j, I am saying x_j - μ_j that this one I am further writing x_j - μ_j and x_j - μ_j, that is why this square is coming.

That means same variable I am saying that suppose it is repeating to creating two variables same on that it x_1 on x_1 that sense if you do like this. So, the covariance is this, as I am saying if x_j vary x_k also vary, there is a chance that is why I am expecting that what is the association between the two getting me So, in that case we can again write down suppose the same formula that all x_j, x_k, x_j - μ_j , x_j – μ_k what we will write here for probability density function.

Can we write that if x_k and x_j separately or we will write x_j joint probability getting me you have to write the joint probability here and continuous case you have to write like this. It is double integration or here I have written all x_j, x_k simultaneously one symbol I am giving or otherwise you have to write all $x_{j.}$ all x_k. What is the notation for this, we will use the notation for this is σ_{jk}. We have used σ_j for standard deviation, σ_{jj} for variance and σ_{jk} for covariance, so what I am writing here that this is σ_{jk}.

Now, be careful about the notation now that σ_j^2equal to σ_{jj}. Later on we will be using σ_{jj} σ_{11} that is the variance component, which is basically if I write σ_{11} it is σ_1^2.Then we will be using σ_{jk}, which is this is your variance of x_j and this one is covariance between x_j and x_k okay so, you have $\sigma_1$2 you have σ_{jk} and you have p number of variables, can we not find out the population covariance matrix now. We are now in a position to write down the population covariance matrix.

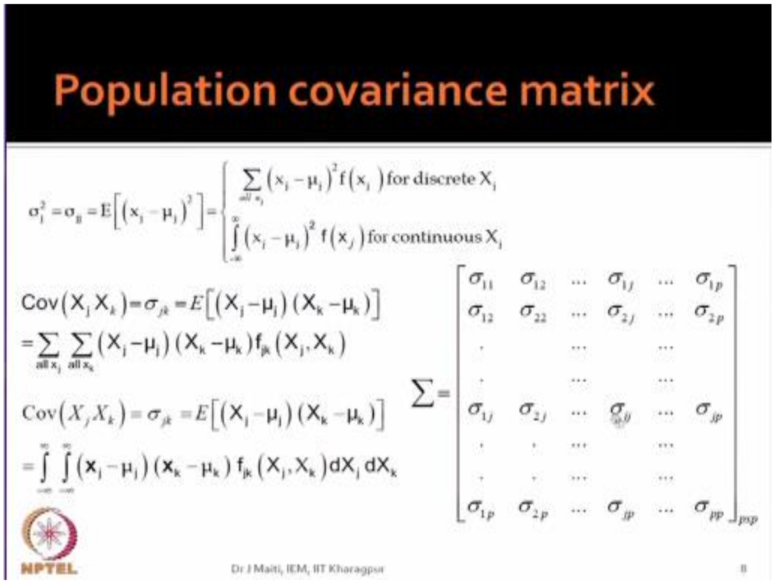
(Refer Slide Time: 46:44)



As there are p variables so covariance stands for every two variables. So, how many elements will be there. So let us write down like this, we will create a matrix p x p, p stands for number of variables so that when this is x_1, x_2 like x_p again x_1, x_2 like x_p .So, x_1, x_2 the variability then when x_1 is varying with x_1 the same variable variability is variance. So, this 1 will be σ_{11} for the second case x_2, x_2 this will be $\sigma_{22.}$ So, like this for p variable case x_p, x_p σ_{pp} this diagonal lines all the elements in the diagonal lines are variance component that I am saying this is basically variance component.

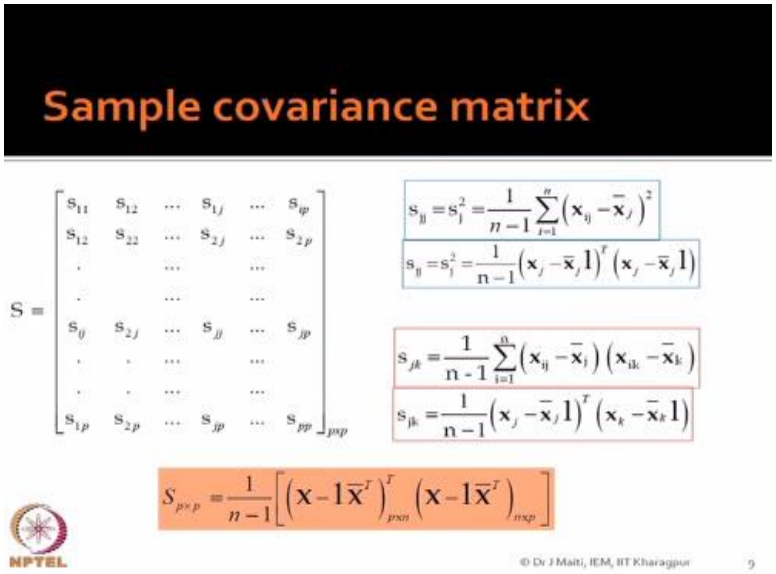
Variance part of the variable as I told you σ_{11} is equal to σ_{22} = equal to σ^2 like this variance. Then the off diagonal elements will be covariance, so what will be this σ_{12}, σ_{1p} again I am writing 12 instead of 2 1 that what is the assumption we are doing $\sigma_{jk} = \sigma_{kj}$ because j^{th} variable k^{th} variable two variables only, but in order we are changing. Then σ_{2p} like this you will get σ_{1p} σ_{2p} this. So, off diagonal elements are covariance part and diagonal elements are variance part, these resultant matrix in our class we will be denoting it like capital σ. So, keep in mind capital σ whenever we will be using, this is your population covariance matrix.

(Refer Slide Time: 49:26)



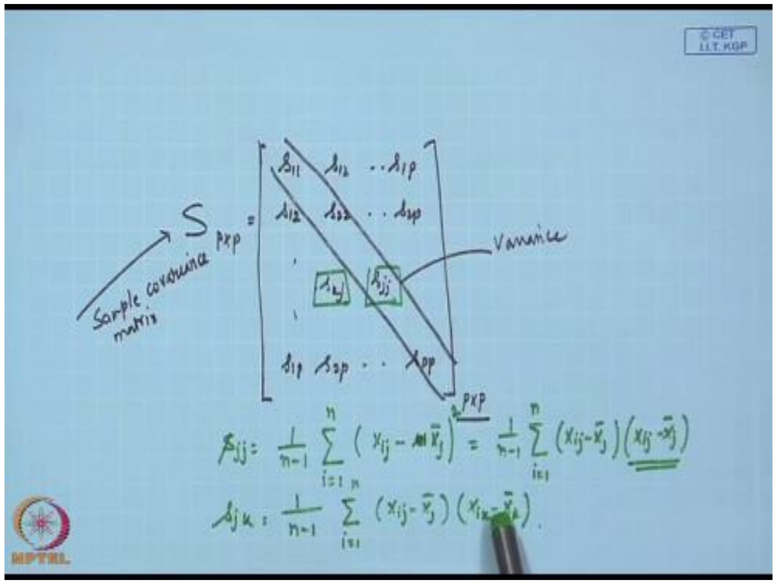
So, population covariance matrix looks like this the way same thing as see σ_{11}, σ_{12}, σ_{1j}, σ_{1p}, σ_{12}, σ_{22}, σ_{2j}, σ_{2p} like this. If there are p variables, there will be p x p elements that this side that will be p x p, the size of the matrix okay.

(Refer Slide Time: 50:08)



Now, we require to, know sample covariance matrix getting me now this matrix will this population covariance matrix, sample covariance matrix very-very vital component or multivariate statistics, very-very vital covariance matrix for the population for the sample. Now come to the sample part.

(Refer Slide Time: 50:41)

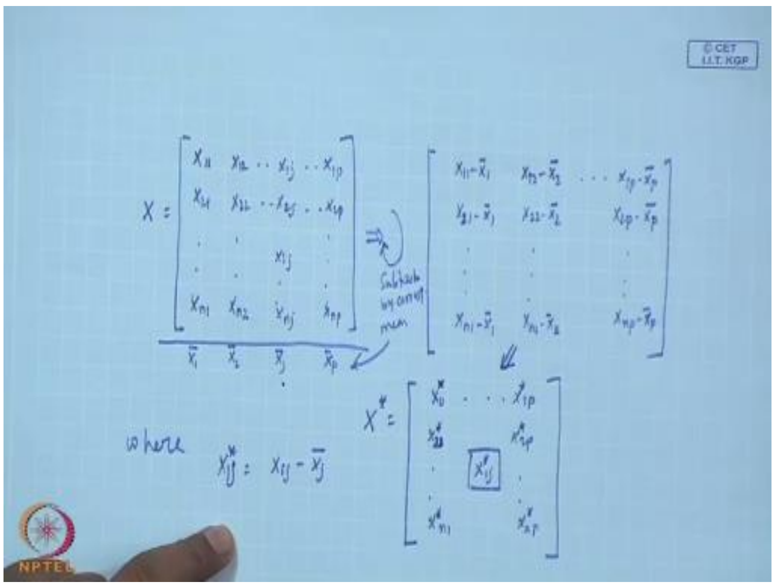


So sample case we will say sample covariance matrix is S. We will be denoting sample covariance matrix as S, this will also be p x p matrix. So, my matrix elements I can write like this S_{11}, S_{12} like S_{1p} again S_{12}, S_{22},S_{2p} like this S_{1p}, S_{2p}, …S_{pp}. So, these diagonal elements these are the variance part and off diagonal element will be the covariance part variance and covariance part. How do calculate $S1_1$, S_{12} like this all the elements of this matrix. So, the general one is here, it will be S_{jj} and somewhere here maybe your S_{kj} will be there or S_{jk} correct so, you can go by the same manner the way you developed in the univariate case, the variance computation.

So, S_{jj} what you will do we have seen that 1 / n - 1 sum total of i = 1 to n, you have written that x, you have written i then, you have written - μ that sense, but we will use here it is basically $\overline{x}$ we have use now j is coming into consideration we will write like this. We can use μ, but here it is μ is not available and we will not when you go for in the sample case, we will always write subtract by the sample average that is why n-1 is subtracted. If I use μ here it will be 1 / n so this square yes or no then, if I write I can write this one like this 1 / n- 1 sum total of i = 1 to n x_{ij} - $\overline{x}_j$ again I can write like this x_{ij} - $\overline{x}_j$ same thing.

So, using this I want to write s_{jk} , s_{jk} is 1 / n - 1 sum total of i = 1 to n, first I will keep the j variable as it is then second case you introduce k. What is happening here actually if you see in the covariance case or the variance case, the original data matrix is transformed that will capture, that concept we will take here you see x_{ij} - $\bar{x}_j$ that means for the j^{th} variable every element is subtracted by its average, for the k^{th} variable also every element is subtracted by its average. Now if this is the case, can I not write down the data matrix in this format like this? That my original data matrix is X.

(Refer Slide Time: 55:41)

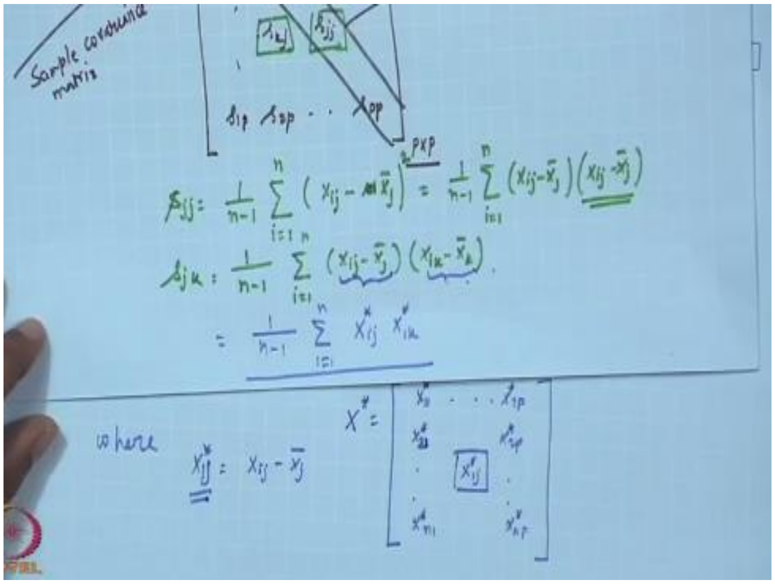


Which is x_{11}, x_{21}, x_{n1}, x_{12}, x_{22}, x_{n2} then x_{1j}, x_{2j}, x_{ij} then $x_{nj,}$ x_{1p}, x_{2p} then x_{np}. So, you have computed here this is $\bar{x}_1$, $\bar{x}_{2,}$ $\bar{x}_j$, $\bar{x}_p$ then, you are writing something you are converting this that some conversion is taking place here, that is subtraction of mean then what are you getting here? If I subtract by mean, I will be getting every observation is subtracted by its corresponding mean value.

Okay so, if I just after this basically it is a subtraction by corresponding mean which is this. So, instead of writing this minus this, if I write like this suppose I will write x* is like this x_{*11}, x_{*21} so like this x^{*}n1. Same manner I am writing x_{*1p}, x_{*2p} like this x_{*np}, somewhere there will be x_{*ij}.

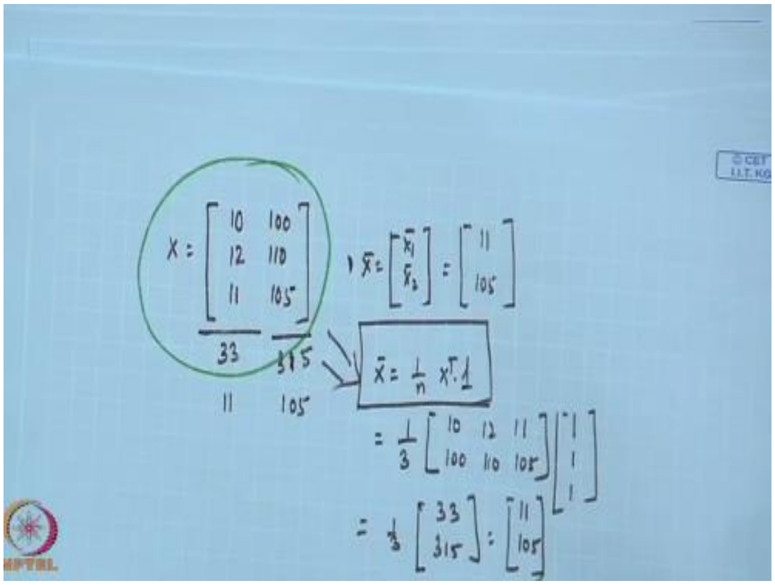
In general one where x_{*ij} is x_{ij} - $\overline{x}_j$ that means this matrix, this matrix same. Okay now, if I use this formula what will happen then?

(Refer Slide Time: 58:51)



In this case that means x_{ij} - $\overline{x}_j$ x^*_{ij} and this one will be x^*_{ik}.So, the resultant matrix will be then 1 / n -1 sum total i = 1 to n x^*_{ij} x^*_{ik} okay so, this type of conversion will take place.

(Refer Slide Time: 59:25)



And ultimately little more mathematics that we will see that. I think up to that to you calculate this using this formulation can you calculate. You take the first data point, you take same data point that first 3 variable values, I think I have given you. Suppose, this is my data points you already calculated mean value.

You have to now calculate the variance and covariance part because there are 2 variable only 1 covariance will be there. Then next class we will go for the matrix, how using matrix multiplication formula we will be able to calculate the covariance matrix totally then, the correlation matrix all those things thank you.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in