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

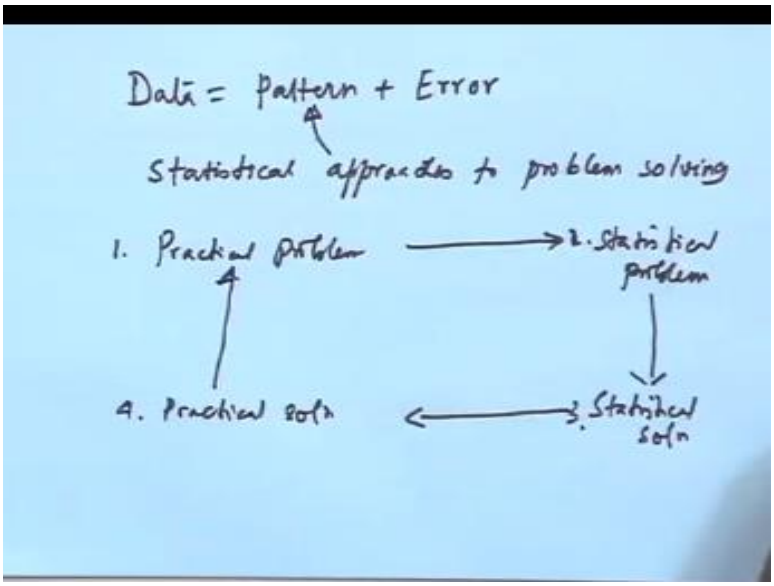
Lecture – 02

Topic

Introduction to Multivariate
Statistical Modeling
(Contd.)

So good morning last class what we have discussed that your data.

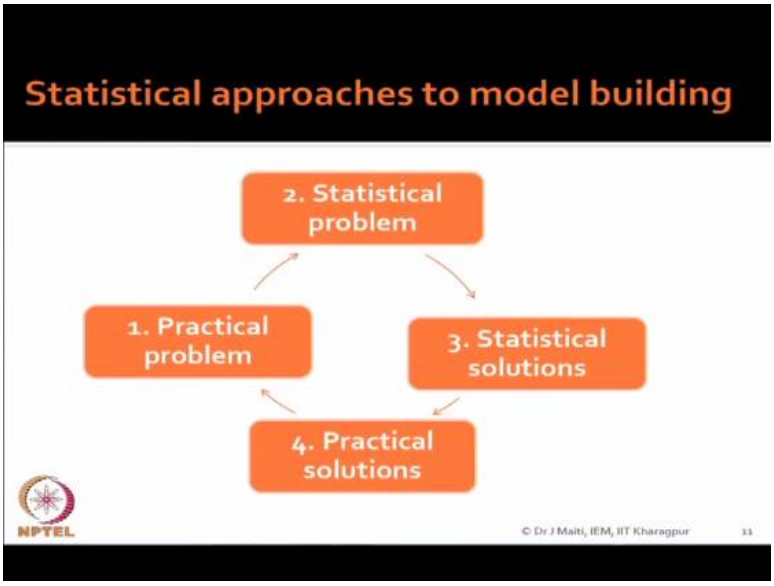
(Refer Slide Time: 00:27)



Equal to pattern + error so how do we extract pattern that was the issue so with respect to this we will discuss now what are the statistical approaches or approach statistical approaches to problem solving statistical approaches to problem solving if you see the slide you see that here we wills start with particle problem then this particle problem will be converted to statistical problem and statistical problem will us to generate statistical solution and this statistical solution so if I say that first one is particle problem second is statistical problem third one is statistical solution then this statistical solution must be practical enough .

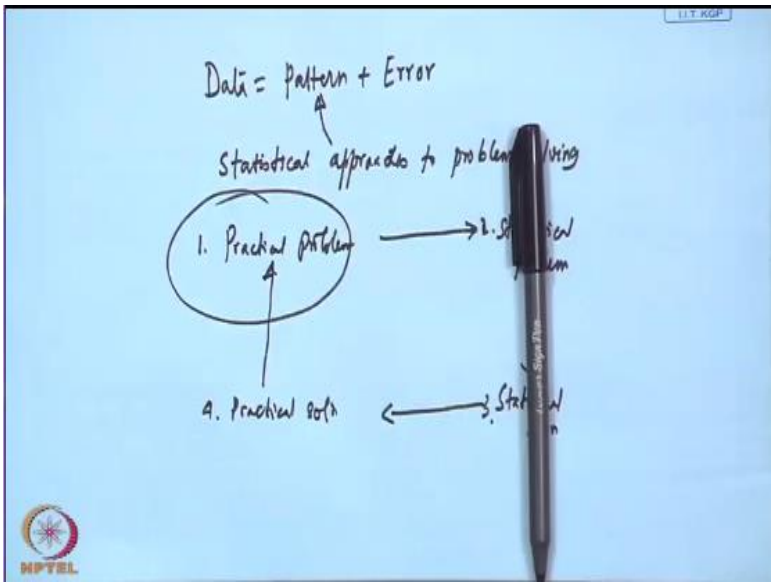
So that is practical solution now whatever solution you get statistical solution is converted to particle solution and it should be checked whether the particle problem is truly solved or not that is why this cycle and if you see the slide.

(Refer Slide Time: 02:39)



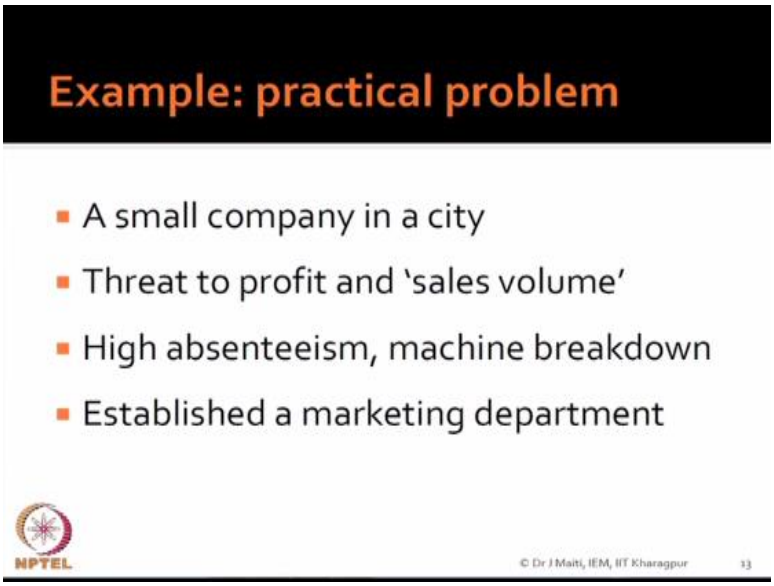
You see that it is a basically cyclic one your practical problem and that is the crux of the matter if you find out the particle problem nicely identify and defined the problem nicely because rest of the things are following it so the statistical problem then your statistical solution and finally the particle solution the many times we falter in.

(Refer Slide Time: 03:11)



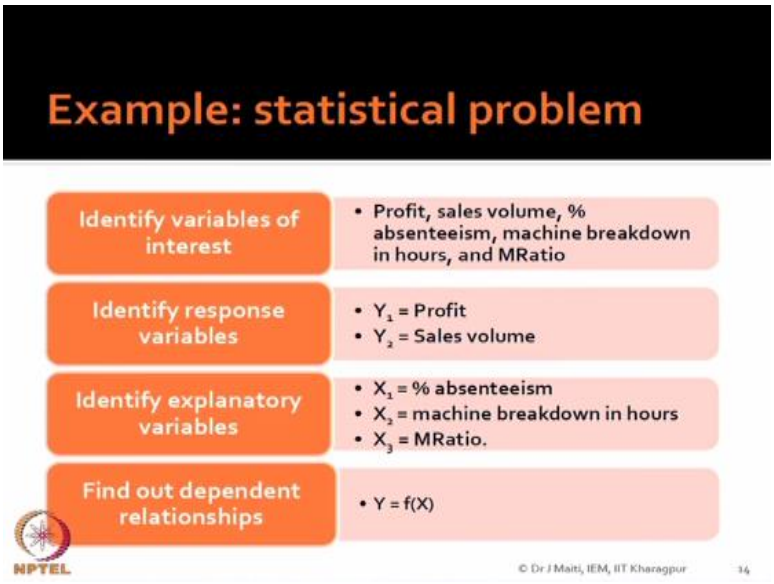
Effectively defining the practical problem getting me so what is practical problem getting me so what is practical problem.

(Refer Slide Time: 03:24)



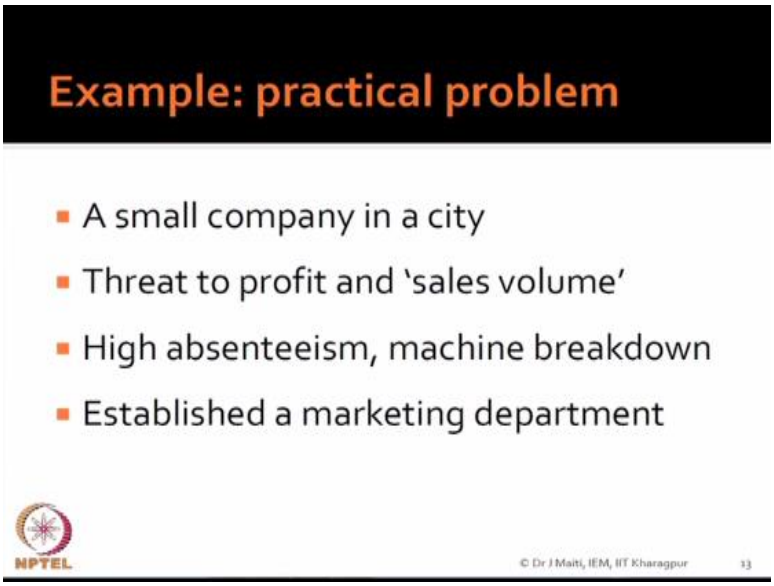
How do you understand a practical problem so I have started with a small example in last class that a small company necessity doing business primarily local at the local level and the company is monitoring it is profit and sale volume and it is observed that absenteeism is quite high and there are machine breakdowns also and in order to improve the that business performance marketing department is also established may be recently and then they want to check that how the marketing department is performing now here what is your.

(Refer Slide Time: 04:17)



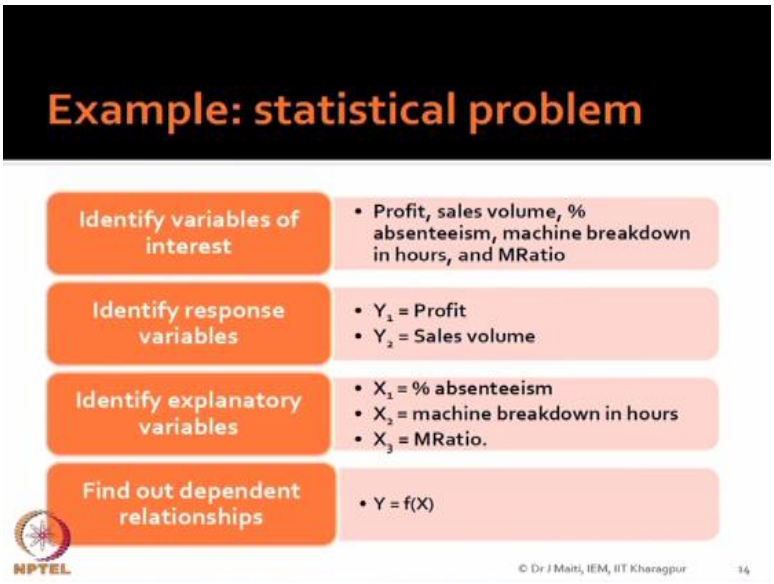
Your practical problem.

(Refer Slide Time: 04:23)



If I go back to this slide again what is the particle problem the particle problem is as I told you that threat to profit and sales volume high absenteeism and occasional to frequent break downs and may be the poor performance of the marketing department may not be but I am assuming that the performance or measuring the department also not good that means the thread to how to work on thread two profit and sales volume if high absenteeism is causing lower profit and what is the extend of that effect then it is better for the management to take actions. Now we will convert it into a statistical problem.

(Refer Slide Time: 05:27)



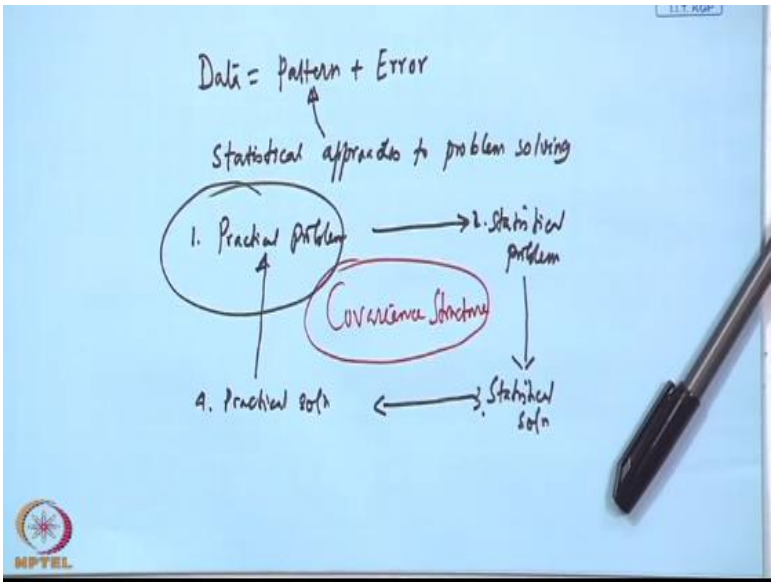
If you relay want to under build a statistical problem from particle problem these are the few very few items or steps which you must follow, first one is identify variables of interest in this example profit sales volume percentage absenteeism machine breakdown in hour and MRatio these are the variables what we have identified now identify response variables out of those variables all the variable what are the response variables now if here response variable by response variable what I mean to say by response variable we want to mean to say that these variables are effected by regions of other variable sin the system.

Getting me in other word what we can say depended variable so Y_1 and Y_2 are response variable they are depended variable why because the absenteeism machine break downs and MRatio can influence profit and sales volume so immediately the next objective is identity the explanatory variables and find out the dependence relationship getting me for these particular problem although it is little bit difficult at this movement that why suddenly we are going for this dependence modeling what are the different types of deference modeling techniques are there but you should not be what I mean to say you should not be governed by the techniques.

That is the one of the threads many times what happens you know may multivariate techniques and you think that your problem will fit to that techniques please do not do this so that is why here I have written in finding the relationship Y is function of x we have not said that it is a linear relationship or non linear relationship or it is a regression or it is something or other way of doing things it is nothing motioned.

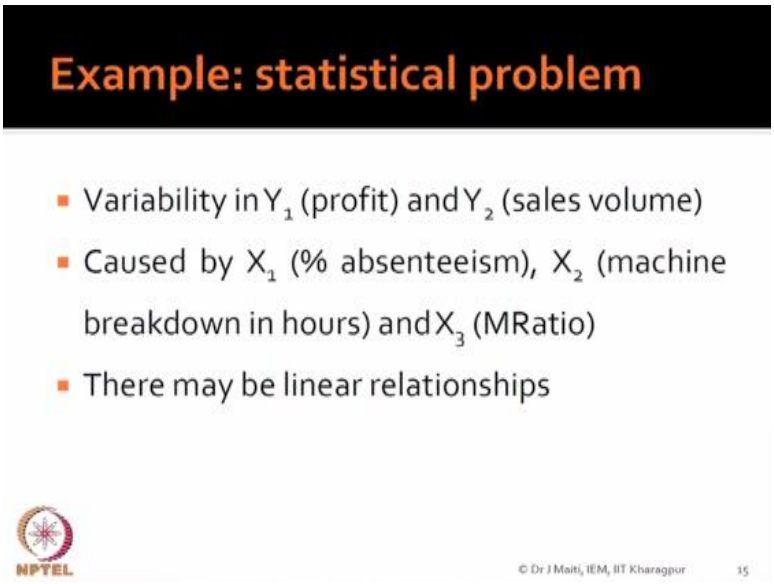
So you must be driven by the problem at hand so our problem at hand is thread to profit and thread to your sales volume and so we got the response variables where you want to concentrate and accordingly some other explanatory variables mean that variables which explain why there sis variability in profit and variability in sales volume and then function of Y effects in statistics modeling the variability is the crux of the matter please understand we you must understand the variability structure in case of one variable there is variants in case of several variable there is variants and covariance and in totality we say that is the covariance structure. So that covariant structure so that covariance structure.

(Refer Slide Time: 08:57)



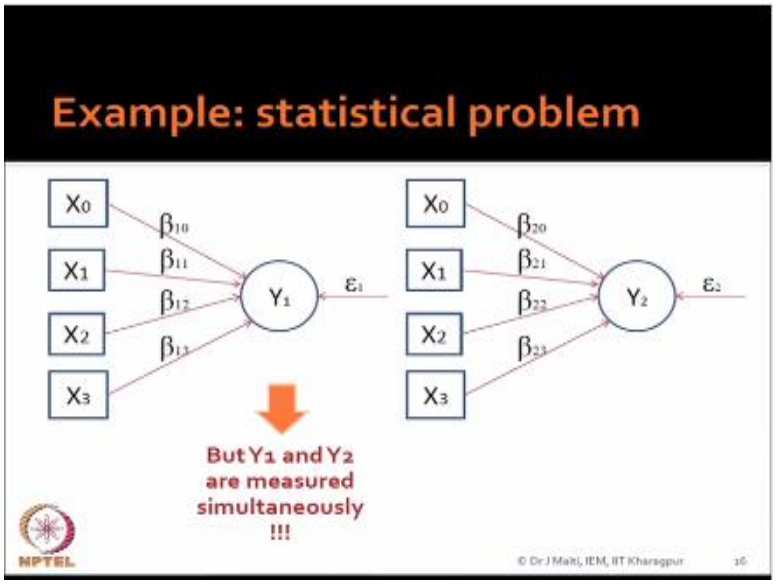
This will give you the pattern you want to extract this will be given by the covariance structure

(Refer Slide Time: 09:18)



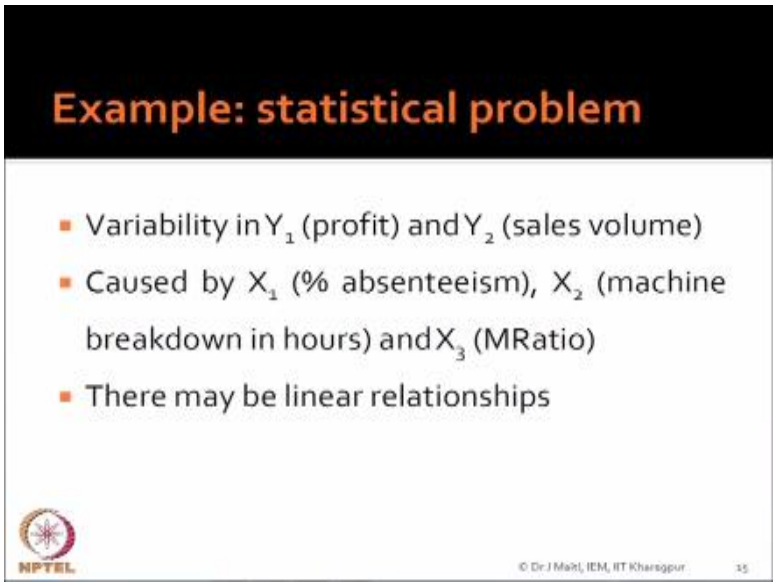
Okay so let us see the statistical problem here the statistical problem is that variability in profit and sales volume by variability we are saying that definitely there is high variability cost by x_1 which is percentage absenteeism x_2 machine break down in hours and x_3 M Ratio there may be linear relationships now you are coming to statistics domain so you have to assume something there may be linear relationships any problem up to this do you have any query any query from your side it is okay this side no problem yes then.

(Refer Slide Time: 10:16)



I want to show then we say that fine that is a your relationships statistical way you can examine that relationship and we also assume that their the relationship is linear now let us see then pictorially if you see this you see this particular figure here what we have said that Y1 is effected by X1 X2 and X3 what is Y1, Y1.

(Refer Slide Time: 10:54)



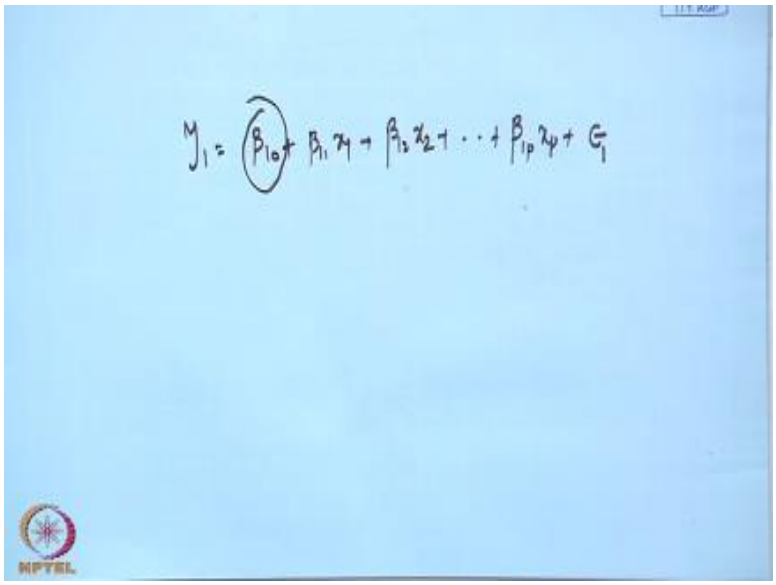
We have said Y1 is profit correct so Y1 is profit X1 is percentage of contingent X2 is break down and X3 is machine hours okay so Y1 is profit X1 X2 X3 have send assume and your break down hours and M ratio now this is the linear relationship between Y1 and X1 X2 and X3 you may be wondering that what is X0 then okay that X0 what is constant value which will be later on in multiple relation you will be knowing that X0 will be giving may value of 1 are all observation that one value will be given.

And β10 will be the inter sect means irrespective of the effect of all those experimental variable considered here but there will be steal some value for Y and that will be determine by β10 then slowly you weight later on we will see but the sole purpose of this particular figure is that what I mean to say that Y1 if you want a linear relationship you can pictorially represent like this similarly for sales volume also you can represent like this in the pictorial representation please keep in mind two things.

First fall the arrow head, arrow head you see that it is suppose X1 versus Y1 if you consider that the arrow a arrow basically starts from X1 and ends with Y1 and arrow head is there Y1 level so that it simply indicates that Y1 is a depended or effective variable and X1 is the costal variable or

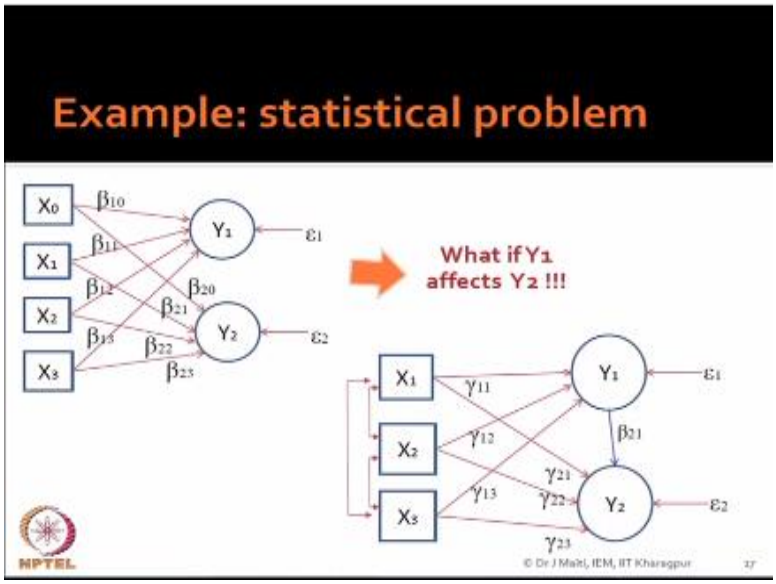
exponential variable and another is we this E1 that is basically some error component and I told you that data had two bars one is pattern and error that apart from this error these rest of the things are basically pattern β10 β11 β12 β13 that will be replying the pattern now similarly for Y2 but for this particular company your Y1 and Y2 they are sales and volume and your profit and for the same month that is simultaneously occurring so if you go by two different models linear models we will suffice that is what I have written that Y1 Y2 and major simultaneously they are cohering also yes, yes.

(Refer Slide Time: 14:33)



Like β1p X_p + E1 so what is this now what I we are what I am saying now further that, that Y1 Y2 your measuring simultaneously do you want to keep the structure in take while estimating while modeling what is the structure here both things you are measuring simultaneously so you see the next slide here.

(Refer Slide Time: 15:15)



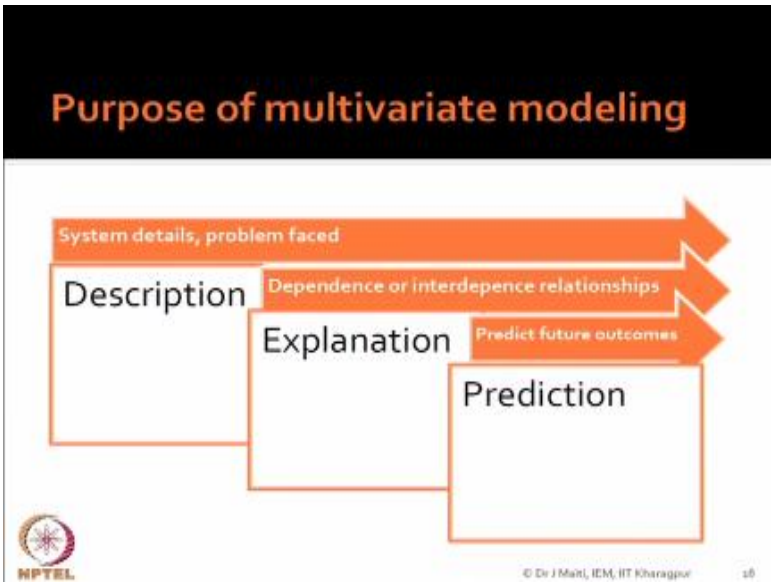
The next slide left one then what you were doing here your basically signore both this occurring simultaneously and they are multimedia observations only so what I want I want to keep them in one model I will not go for two linear models only in one linear model I will do this okay but there is another problem what is other problem, other problem is Y1 is basically and Y2 there maybe relationship between the two it may sure happen that Y1 may affect Y2 so that means that Y1 is not only depended variable or X1 variable it also becomes a casual variable or exploiter variable.

It all depends on the situations suppose in your situation this is not the case no problem but many times what will happen you will find out that this type of structure is there. So when such structure is present in the practical problem how can we ignore it, you cannot ignore it, that means you have to keep the practical that behavior real behavior intact and then find out the model.

Not the other way around find out a model, hit the data and accordingly you say that is the behavior of the system studied, it is not like this okay. So actually what happen there are few

models by saying this few models are, I can say discussed pictorially one is multiple regression, then multivariate regression then path model, later us see what are all those things.

(Refer Slide Time: 17:24)

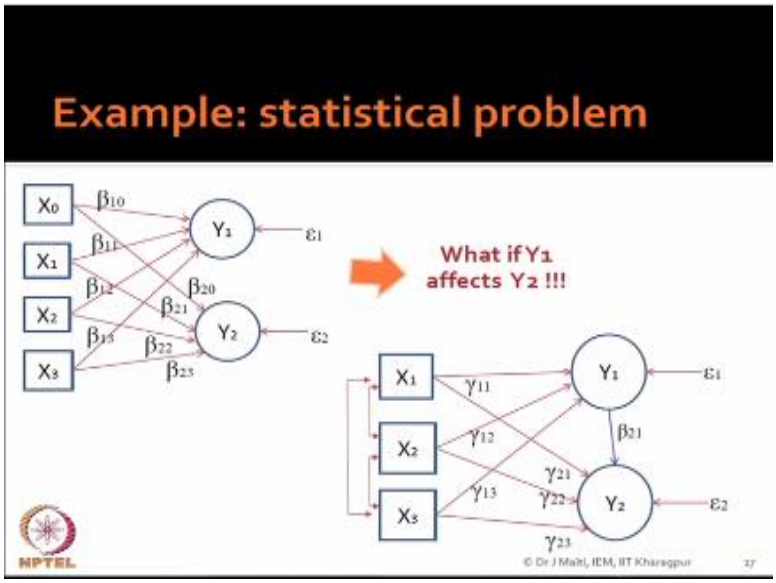


Now what is the, what purpose effectively basically from statistical sense, I said that what is the purposes of multivariate modeling in the beginning, but from statistical sense what are the purpose. You see what is the first purpose is description, any model you build it definitely describe what is the problem attend and what are the purpose of that study and what are the different variables involved in this particular problem and what are the, how this variables are measured, how this variables I can say are stored or kept, what from what data source it is found out, whether it is a primary data, I mean just we have gone and collected or you have taken from some other source.

So all those things are coming under description. Then these description part should be done ritual in fact everything should be done ritually, descriptive part you should not falter, because this is the problem definition part. Then explanation, what is explanation? The relationships between the variables amongst the variables what is the relationship that is coming under explanation, then prediction.

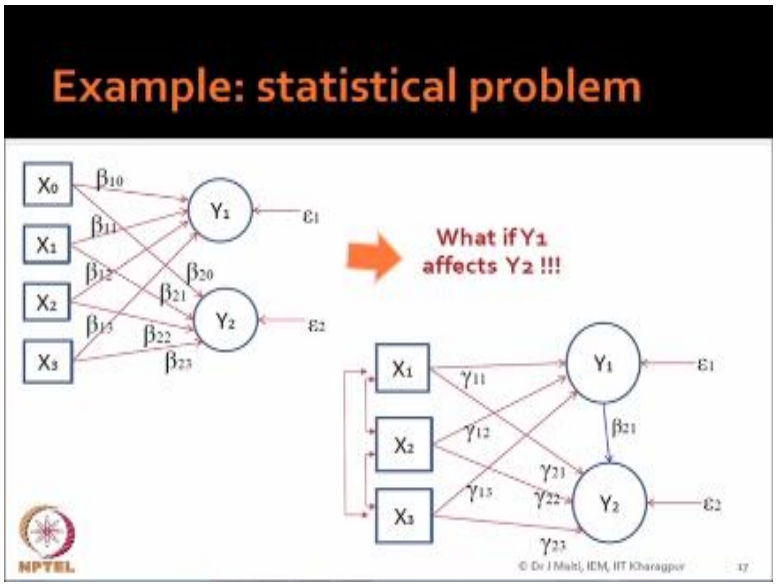
Okay, so whenever we talk about any model we talk about description on the problem, we talk about explanation of the relationships of the variables building which is basically used to build the model and prediction, now many models are not able to predict. So that means in a model if in any model if the first two portion is missing then that is not a model. You will just develop a structure like this schematic diagram like this and describe something and you will say it is my model that is a description part only. That is not the model you have to explain the relationship that means in here.

(Refer Slide Time: 19:55)



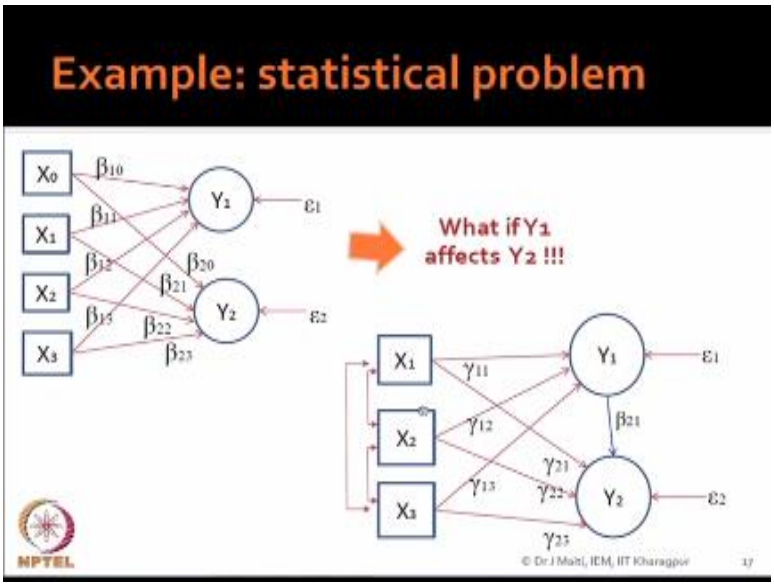
In this particular diagram this is the description plus relation, where is the relation, relation is γ1, γ2, γ13, γ21, γ22, these are relations and without this γ and β values this is only a description. If you use certain model which will give you all those βγ estimates and then explanation is completed.

(Refer Slide Time: 20:31)



Now using this, this side the left hand side plus the estimates of βγ in errors if you predicts some values for Y_1 or for Y_2 that is what is prediction, any model. I understand the value of this is the first class the second basically second lecture here. Now this what I am trying to give a pictorial view only here, the how this β_1 γ_1 or $\beta\gamma_{11}$ β_1 all those things I mean what is the issue and how it will be estimated all those things will come slowly in the subsequent lectures you will be knowing the estimation position.

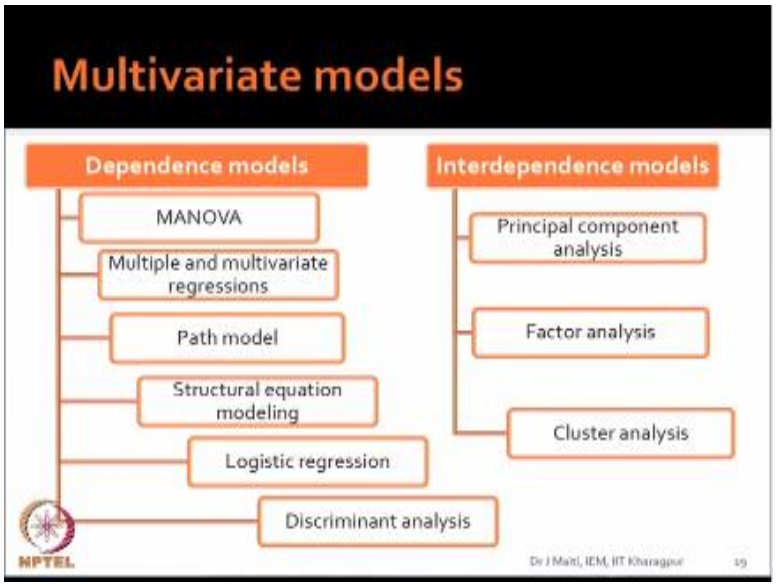
(Refer Slide Time: 21:45)



Where this is the only diagram, this is the only diagram. Okay, fine I understand now, your question is this one, this side. Actually in this case this is a representation of multivariate regression here we come and consider all this variables are to be purely independent, and in this modular case this is a pictorial representation of path model. Here we consider that the independent variable can vary, can coherent getting me but in reality you will not find out that all variables are independent truly independent.

So this is a better representation okay, surely I just if multiply is a problem then this is the case selected on after in few subsequent lectures it is there, okay.

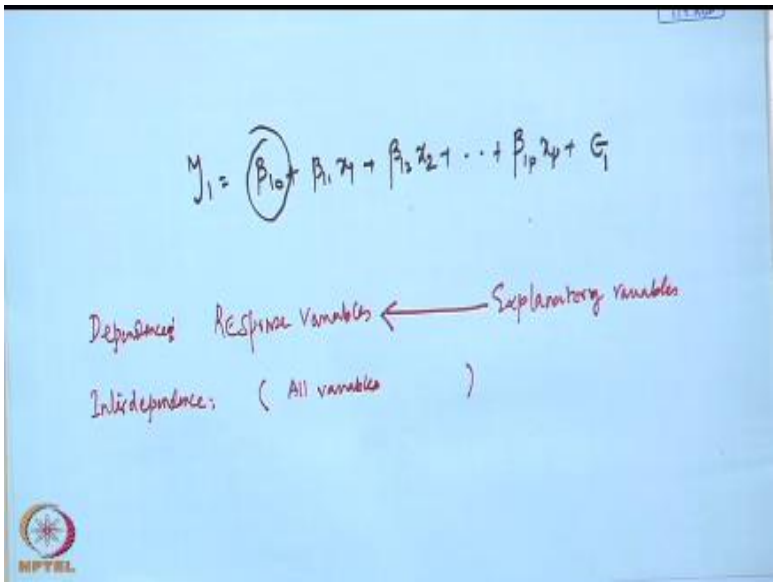
(Refer Slide Time: 22:56)



Now what are the different multivariate models that will be discussed in this subject, okay so I have given you so idea of that what is multivariate analysis, what is multivariate statistical modeling, what do you mean by multivariate and also different data types and how practical problem can be converted to statistical problem and in statistical domain what are the problems and then there are we will now, we will see that there are different types of multivariate models and primarily at these models we will discussed in this subject, okay.

So multivariate models are categorized into two broad groups, one is dependence model another one is interdependence model. By dependence model we will say that there are two sides of variables.

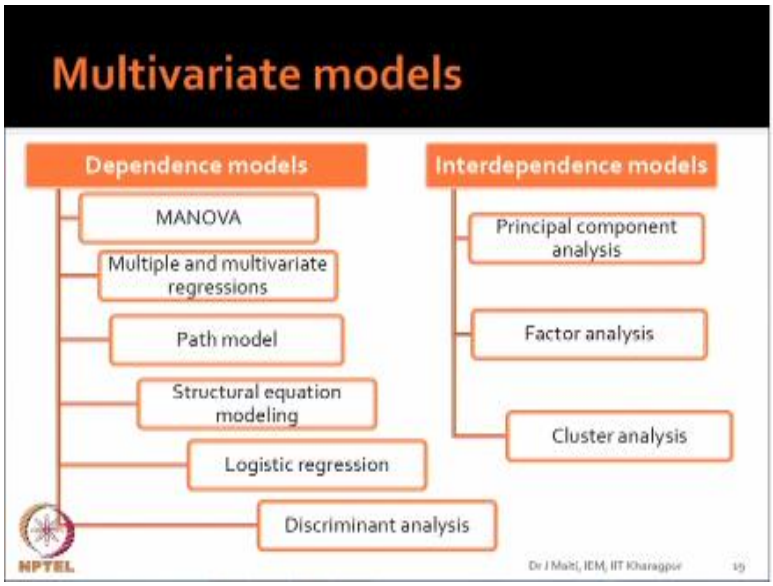
(Refer Slide Time: 23:59)



One is respond side or response variables another is explanatory variables, getting me there are two sides explanatory variables are used to explain the response variables so that is what is dependence, that is the dependence structure, okay and interdependence if you see thus interdependence model that is no response side all variables are coming under one bracket, all variables under one bracket.

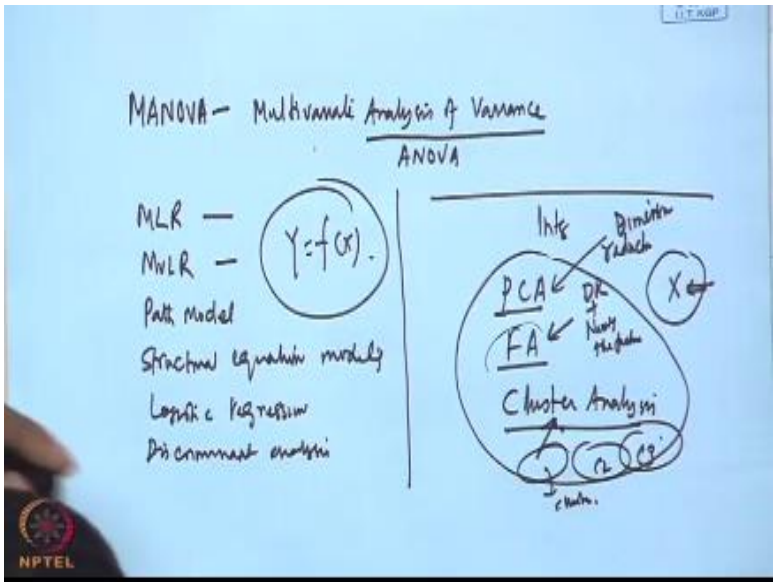
What do you mean to say there is relationship definitely, but within these are same set of variables there is no categorization of that once it is dependent and variable or response variable otherwise it is explanatory variables so based on this concept, so there are dependence model.

(Refer Slide Time: 25:17)



There are interdependence model, so in dependence model MANOVA is there which is multivariate analysis of variance.

(Refer Slide Time: 25:29)



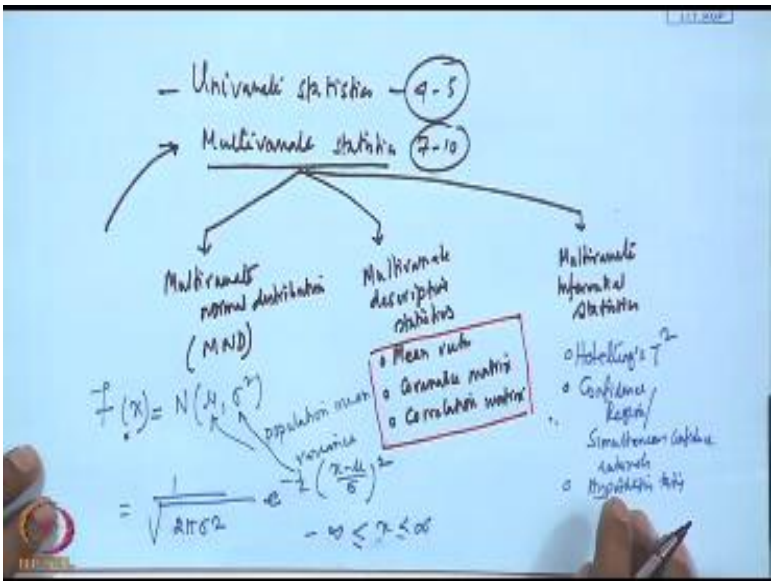
Multivariate analysis of variance, I think all of you know that what is analysis of variance that is ANOVA and it is presume that the ANOVA is known the only MANOVA can be understood. Then multiple linear regression, then multivariate linear regression then your path model structural equation modeling is another component. In addition logistic regression discriminate analysis. There are many more dependence model, but this sets of models what is the basic picture is that we have two sides of variables one can be affected by other side of variables and we can build model like Y=f(x) this type of relationship model we can get.

Other sets, other group that is what I say the interdependence model, so under interdependence model case there are principle component analysis, factor analysis, cluster analysis. Here we will not segregate the variables into dependent or independent side with everything as x one set of variable and then we want to see the covariance structure of this x variables there are many x variables and based on this structure our primary objective in principle component analysis is reduction of dimensions, so dimension reduction.

Factor analysis not only dimension reduction plus naming the factors. So dimension reduction plus naming the factor. In cluster analysis we want to group the individuals not the variables,

factor analysis and PCA we basically group the variables so then we reduce the dimension and in cluster analysis we will not group the variable, we will group the individuals or items or objects on which that I has collected and then we make several clusters, okay cluster 1, cluster 2, cluster 3 so like this several clusters will be formed. Before that, because we will try to cover so many models it may not be possible to cover all but substantial number will be or models will be occurred here. But before that what the basics of multivariate statistical is very, very important that will be covered.

(Refer Slide Time: 29:05)



So as per as the sequence of lecture is concern so after this in next lecture we will covering your univariate statistics, a few lectures on univariate statistics followed by multivariate statistics, okay. So there may be few hours let it be four hours or five be hours 4 to 5 hours of this univariate case then multivariate statistics may be around 7 to 10 hours will be on multivariate statistics and understanding of multivariate statistics is very, very important for all of us.

Otherwise, you will not be able to explain the multivariate models and know what are those models, why this hour from this modular coming, what is the utility of model we will say many things and you will not be able to explain, so the backbone for all those multi variate models

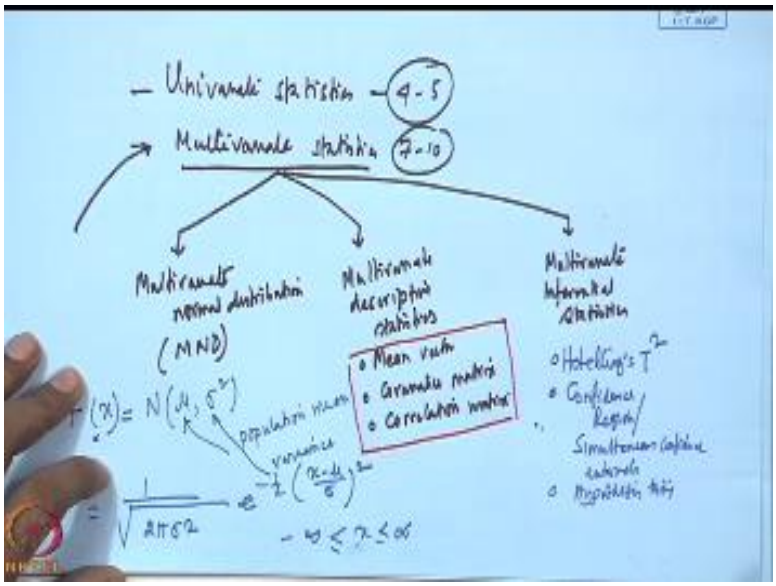
what I shown you just few minutes back is multi vitiate statistics under multi variate statistics there will be multi variant normal distribution which will be denoting like MND and followed by that will be multi variant decrepit statistics which covers primarily mean vector co variants matrix co relation matrix so under multi variate decrepit statistics will be concentrating on this.

Then under multi variate statistic the next issue is multi variate inferential statistics okay so under multi variate inferential statistic what are the thing will be covering we will be covering first is Hotelling T^2 getting me then you are in that is confidence region and simultaneous confidence interval then your hypnosis testing so whatever you have learn in variate statistics that in variate inference statistic that counterpart in multi variate doming that will be discussed up to hypnosis testing okay.

And it is that in case of one population in case of two populations all those things will be discussed there and you will see later on that as I told you that multi variate your variate normal distribution is the crack of the meter so in variate normal distribution all of you know and that we will denoted like this N μ and σ^2 where μ is the population mean that population is characterize by one variable that is x and σ^2 is the variants component variants of the population with respect to x.

And this one is $1/ \sqrt{2\pi}$ of σ^2 $e^{-x- \mu/ \sigma 2}$ where your x value lays – infinitive to + infinitive okay. It is possible so now what will be the multi variable counterpart of this so normal distribution if you draw.

(Refer Slide Time: 33:55)

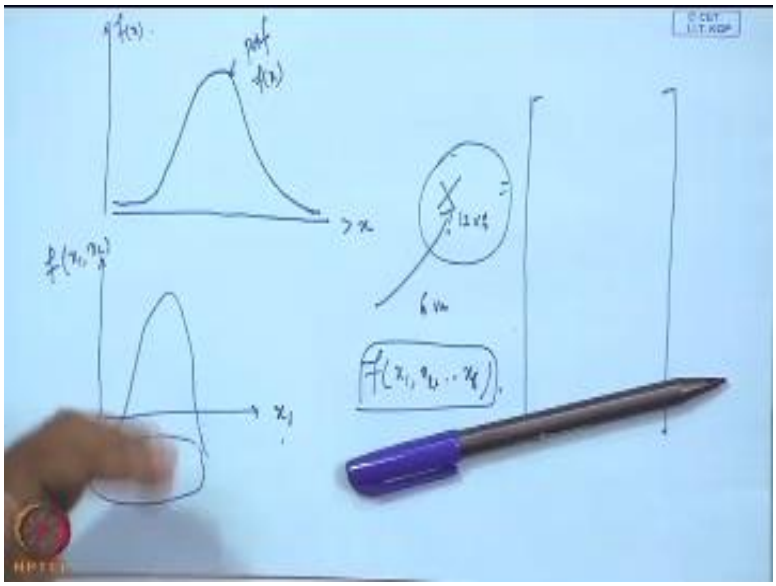


You will be finding like this, this is your uni variate normal distribution which known as probability density function or other way we say f_x this diamond direction x axis will be x variable and your y axis will be your f_x. So you have to understand it is multi variate counterpart that will be very difficult one it is not that easy suppose there are two variables x1 and x_2 and then there will be join distribution x1 and x_2.

So what is the distribution so something like this you will be getting a two dimensional picture when we talk about two variables when we talk about more than two variables that is not possible to draw but you have to think that there is some shape that have structure level of thinking is required for understanding multi variate normal distribution, and as I told you multi variate normal distribution is very, very important because all though models almost all the models in multi variate statistical modeling subject follows multi variate normal distribution.

Because if the data follows multi variate normal distribution then what will happen most of the test can be possible you can measure the goodness of it with appropriate statistical distribution of the statistics what will be used in those models, okay so you will know all those things that mean vector go variants matrix and co relation matrix like this.

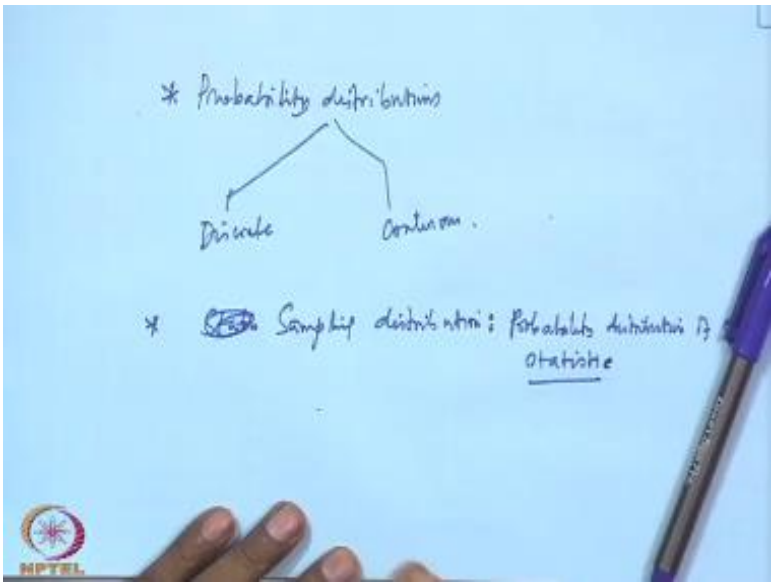
(Refer Slide Time: 36:00)



Now with respect to this small company data I think I have given you that they the data is 12 x how many 6 12 x 6 data was giving so you have seen that the matrix also you have seen the data matrix now if I ask you what will be the distribution of this s there are 6 variables getting me so it will be a 6 dimensional issue and your join distribution will be f x_1 x_2….. x_6 so how to get this distribution.

So you cannot work now in scalar doming like the in variate case we have to go to the matrix doming and as a result you will see that I have written under multi variate distribute that is few thing one is you mean vector, vector is coming into co ratio second one is co variants matrix third one is co- relation matrix okay so we will be discussing this after in variate statistics once this will come multi variate normal distribution and descript statistic is cover will go for multi variate inferential statistics which will basically discuss on the messy shop Hotelling T^2 will be using extensively the Hotelling distribution so apart from this.

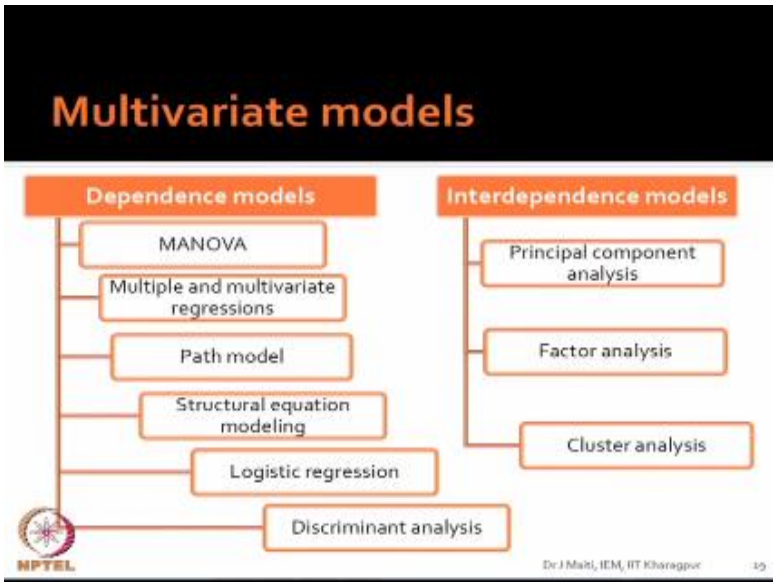
(Refer Slide Time: 37:44)



I request all of you to do one more thing that you please through probability distributions little bit of this because this knowledge is very important all discrete and continuous as well as continuous take up already distribution will not taught here okay I will definitely bring some all distributions but I cannot explain each of the distribution okay so you have to go through this another issue in this subject is that statistical distribution no not statistical this is sampling distribution.

It is also a probability distribution but it is related to that is probability distribution of statistic, or there will be statistics also but actually statistic and this statistic we are basically talking about random variable. I will explain in a variant multi period that based statistic case that how this, what is the statistic how it is random variable all those will be discuss to you, okay.

(Refer Slide Time: 39:31)



Now let us come back to this again.

(Refer Slide Time: 39:37)

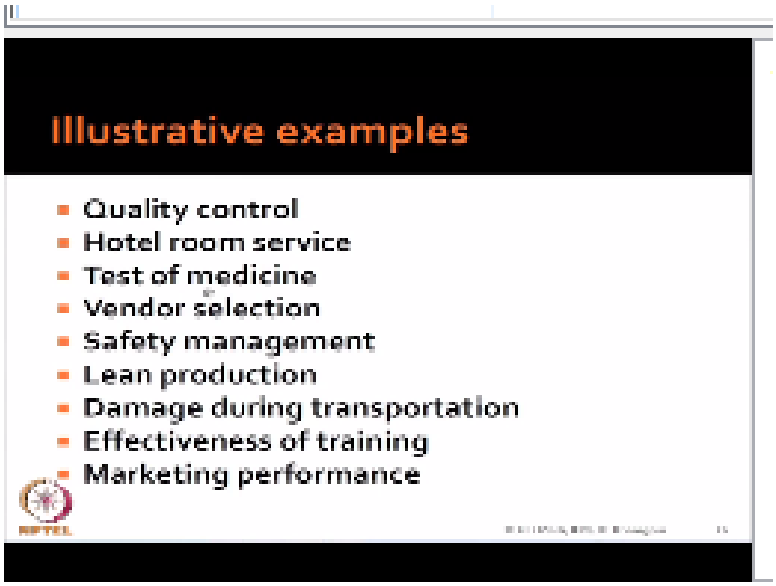
The introduction part that I told you that multi variate analysis is purely for the here for this particular subject here it is for the people who will be using it for serving the real life problems now by realize problems what do you mean here some illustrative examples quality control any idea about quality control. For example you think that suppose in Kharagpur area itself there is tata bearings the tata metallic's now tata bearings produce the ball bearings roller bearings this component they produced. Now this components has certain quality features for example the dimensions ,for example they strength now customers will require the bearings the product with certain quality features ,this quality features basically converted in to specifications .

If you produced beyond specifications means not within the specifications provided by the customer what will happen that product will be rejected by the customer but if you want to see that why you are producing rejects then you will find out that there is a problem with the process manufacturing process now manufacturing process variables will affect the product quality so it is product quality variables will be the dependent side and process side variables will be in the independent side you can model that product quality versus business variables another mass examples of quality control issues are there published in the literature okay, using statistical techniques .

(Refer Slide Time: 41:42)

Second example basically it is a service one hotel room service case service example ,there was you see that multi ah that hotel B hotels basically I am talking about this hotel they are basically survey providing service to many customers so now they also the equal to some sort of quality management otherwise impossible ,they will not lead to the profit they are interested in so they are also lot of data is generated and you have to use those mm data and develop model and solve the purpose for which the model is developed or the basically the hotel rooms are be much better.

(Refer Slide Time: 42:36)



Test of medicine another issue ,test of medicine every time every month you will see that there is a different medicine coming per different diseases for even a single diseases there may be four ,five medicines there ,now my question is how do i know that medicine works, medicine A works ,for some group B may works for some other group may be no the same diseases or some medicine performing better than the other medicine for the same disease .

So that type of testing is also possible using if you collect upper data that is possible now vendor selection is a very tricky problem very essential ,critical problem in all most all small medium large enterprises because you are vendor the supplier quality very important if you get poor items ,poor raw materials ,supplied by your vendor ultimately your product quality will suffer .

So how to go about it? What modeling can be possible they are so that you will select the base vendor they are different way of selection may be someone may go for not statically roots some other root but statics also multivariate statics also helps you in selecting vendor. Safety management so as you told your jobs few minutes back there are many variable effecting the people safety i want to know that what are the variables effecting more and what are the variable

effecting less so that I can take appropriate actions which will improve the overall safety standards of the plant of the flow of the work everywhere .

(Refer Slide Time: 44:36)

Now the lean production you see this lean production means what is basically there suppose inventory control is a big issue you are producing ten items but from raw materials you keeping huge items a huge raw materials in inventory it is a lost to the company so what you want you want to minimize the inventory but you cannot minimize the inventory unless you have some steps taken some buffer or something else is there through which you can satisfy the costumer at the same time based on your product requirement ,production requirement.

Now lean production says that you minimize everything in such a manner that you will satisfy the customer as well as from inventory point of view there is also a minimum level raw material inventory, working process inventory everything at the minimum level. There are very good models in work operation research area but statics also can be useful there.

(Refer Slide Time: 45:44)

Now damage during transportation another one you see the ,damage during the transportation was the logistic problem I will show you some i think one tutorial I will give later on that how to model the damage part with respect to that what mode of transport you are using what is the distance you are transporting what type of packaging system you are using so all those things ultimately lead to damage and its huge cost to the company because you are basically supplying made that product to the customers effectiveness of training is there.

Now question PPT based training here teaching is best or class board chalk and talk there is question like that in the training in industry level that weather it is basically class room training or on shop flow training what will be their teacher combination or faculty combination so many issues are there so if you collect data effectively you will be able to use multivariate statistical models to find out that which mode of training which type of training is better.

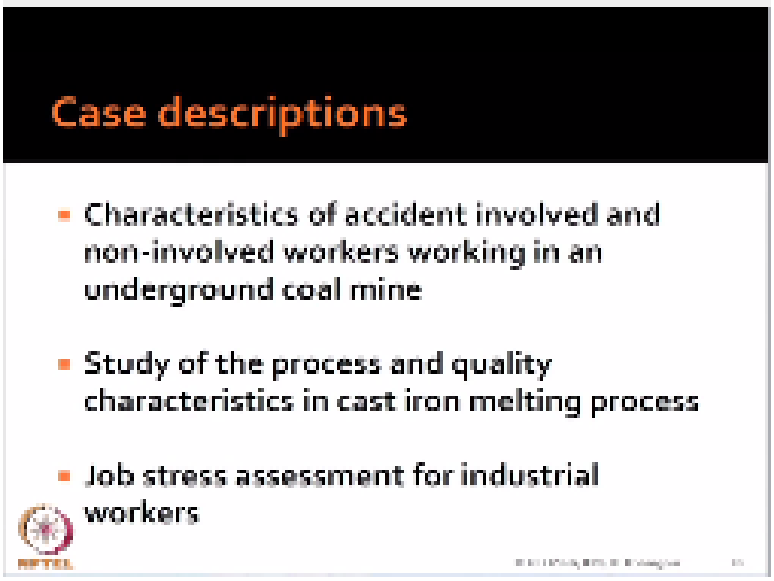
Marketing is a area where as safety sorry your statics is used very much for example if you go through any marketing journal you will be finding out that full of statics many models come from that side also marketing research that is the area where you will be used statics by

marketing purpose basically we have conducted one study to understand the positive intensions of customers and we used structural modeling ware.

So some illustrative example some other relative example can be framed but I you are the person who knows your system and please find out some case some example from your workspace your domain of expertise so that whatever I will teach here you will be able to translate to your own system.

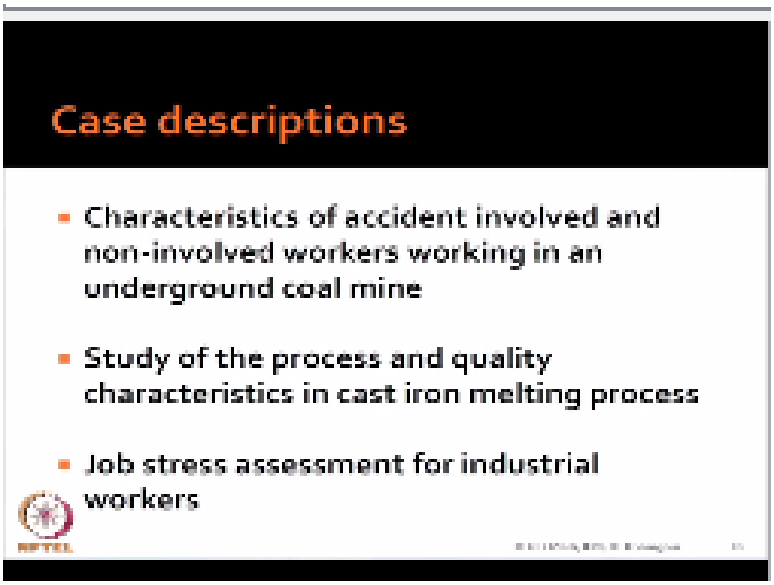
In admin science, management science social science every where this multivariate statics is used and you have to find out all those things and I will be giving one after another and you just based on one this my explanation the things you will be learning here you try to find out analogy to your system and develop accordingly and then learning will be complete otherwise learning will become incomplete ,you have to be very, very careful for this purpose

(Refer Slide Time: 48:53)



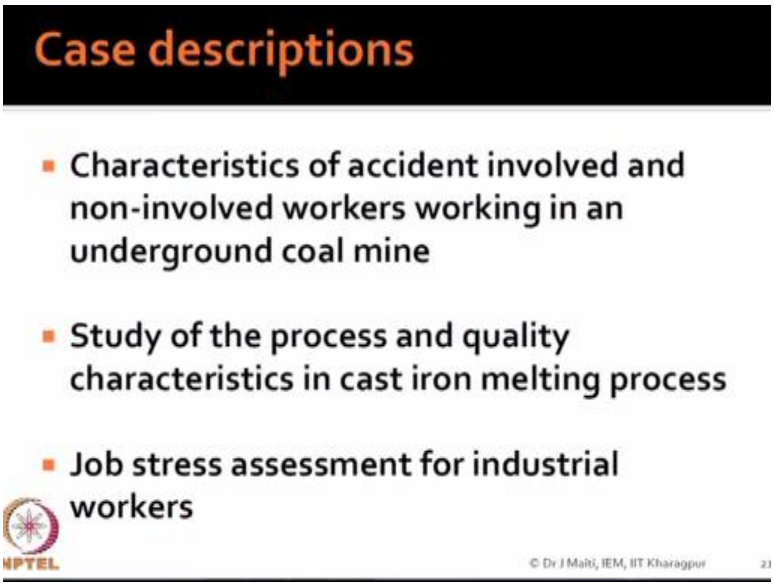
I will show you later on some cases so that you understand that totality as I told you starting from practical problem to practical solution that mean practical problem statistical problems ,statistical solution practical solution this is total cycles will be completed using this our cases.

(Refer Slide Time: 49:22)

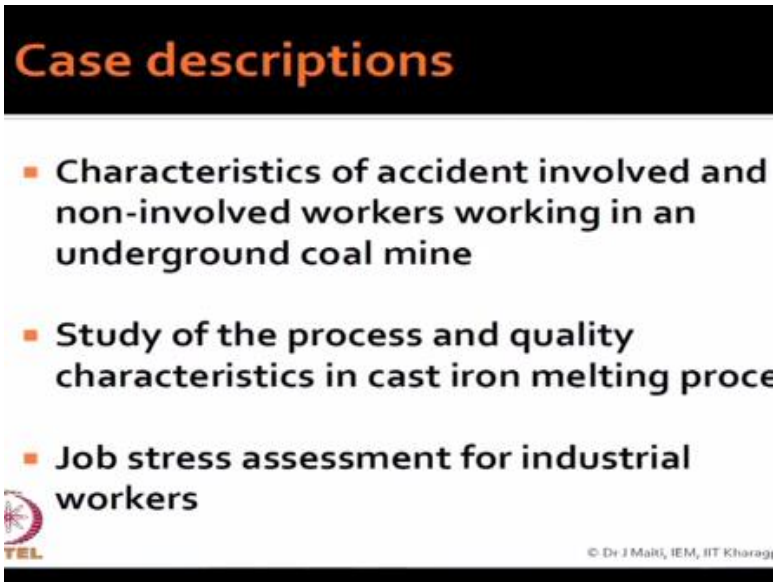


This is our first case is characteristics of accident involved and non involved workers working in an underground coal mine told you that in safety management just i told you that one that in safety management one of the issue is one of the theories was there which is not now that much popular the theory still working in that sense that some people are inherently accidental prom, in respective of that the situation they meet an accident. But some people to avoid accident, so that is the issue and they wish started thinking that people can be categorized, as accident prone or not accident prone, so there are 22 variables we have collected and I will show you this case.

(Refer Slide Time: 50:28)



(Refer Slide Time: 50:42)



Wherever it is required based on the need of the model that will be described. The study of process quality, a process variable as well as quality characteristic, in a cast iron melting process, that is also real case study, like earlier one and that where we have, the two states of variables, we consider and we found out the relationship between the two states and based on this relationship how effectively the quality of the melting process can be improved.

Third one is job stress assessment for industrial workers. Ok? There from officers to that some where working like mechanic or the someone doing the welding, so different job profiles are there, the responsibilities and the rules vary. And accordingly what happen, they may suffer from different level of job stress. Ok, I will show you that how the job profile as well as the demographic value fit job stresses.

And you will see that what way it will be done and the similar situation you will face in your work also, and then straight away translate this to your work, it is possible. Getting me? Any questions? Because this the purely qualitative one lecture but next class it will be mathematical and if you, there are some, I think all of you know this, but there are famous quotes for statistics.

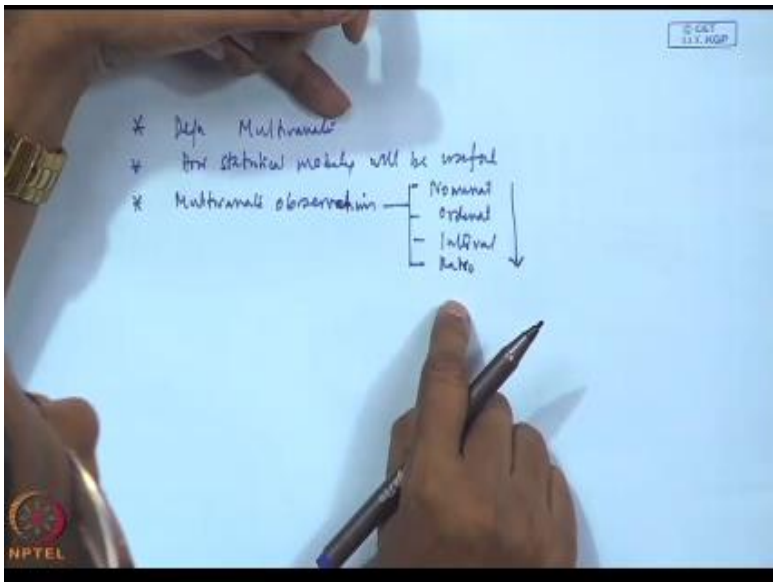
(Refer Slide Time: 52:53)

I will end today's, now this lecture with these quotes, you know this Bernard Shaw, what he said you see 'It is the mark of a truly intelligent person to be moved by statistics'. Okay now you know Mark Twain also, but what he says is that there are lies, damned lies and statistics, the two different philosophy, Mark Twain is saying that don't believe in statistics, because statistics is the maximum level.

But we believe in Edward Deming, because ultimately Edward Deming is the 'quality guru' that sense. He used statistics in the quality domain, quality management domain and he showed that statistics can do wonderful in improving quality in terms of totality of quality management, what he said "God we trust, all others must bring data". If there is no data, I will not believe okay.

So what we can summarize, what we have learned today, So today ultimately we started with the definition of the word multi variant, then we said that how statistical model will be useful and we also said that multi varied observations needs to be carefully measures, and you must be thorough about the data types, like nominal data, ordinal data, your interval data, ratio data.

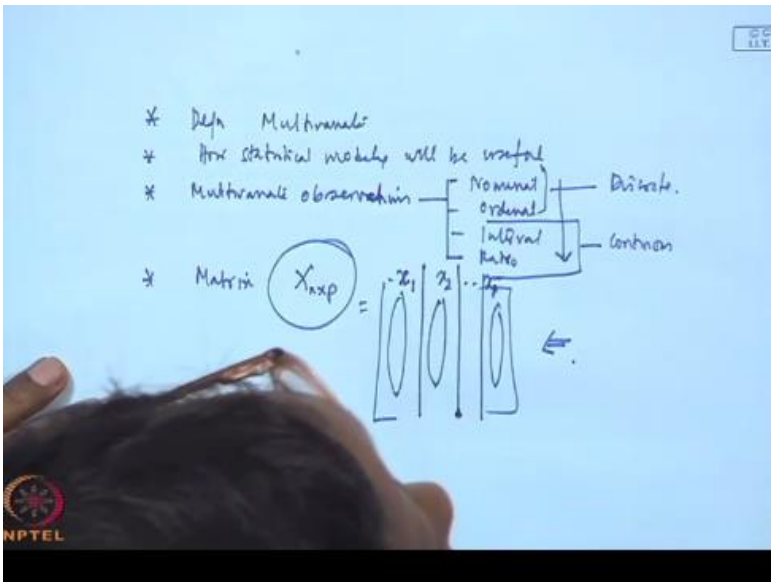
(Refer Slide Time: 55:25)



If possible, you go by this order means, if possible calculate ratio data. And we have seen also that ratio data and interval data are basically, comes under this continuous and first two come under discrete, but by discrete data you also mean have understand was a account data, that all please keep in mind. Ok? Then I told you basically that multi varied analysis means you have to work in the nominal matrix, because X_{n*p} that is what our multi varied observation.

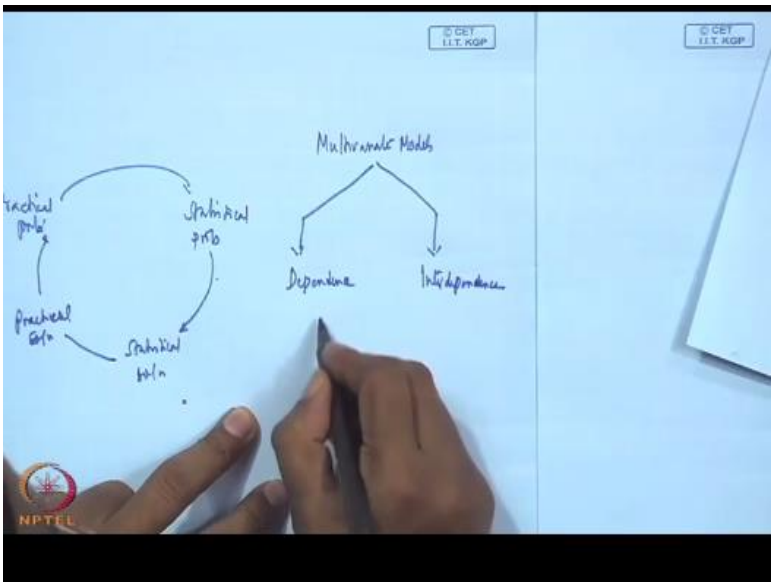
And this one can be written like this, there may be several, your victors like x1,x2…..xp, where x1 denotes the first column all observations, x2 denotes the second column all observations, this is the vector variable, vector 1 variable, vector 2 variable, vector 3 that is very, very important for all of us. Ok? Then another thing what we say is don't rely on technique.

(Refer Slide Time: 56:50)



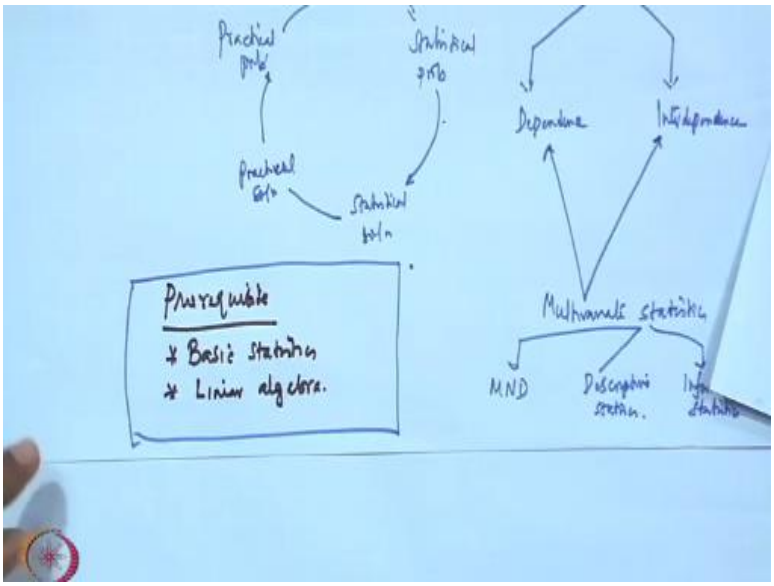
Do not be biased with techniques, you go by problem, go by the need of the, need means, what I mean to say, go by the problem and you carefully define this, and that definition that means you are basically going from the practical problem to statistical problem, statistical problem to statistical solution, statistical solution to practical solution, then practical solution to this and what you more you have understand here show that , you have understood that there are multi varied models, there are two states of modeling techniques, one is dependence another one is inter dependence.

(Refer Slide Time: 58:13)



But please keep in mind that these two will not be fully understood, if you do not understand multi variant statistics, mainly the descriptive statistic part. Three things you have to understand here, multi variant normal distribution, descriptive statistics and inferential statistics. Okay and last but that is also very, very important one is that, that important is that prerequisite, your prerequisite is basic statistics and I think linear algebra that will be better. Okay so thank you very much, we will meet tomorrow again.

(Refer Slide Time: 59:19)



NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in