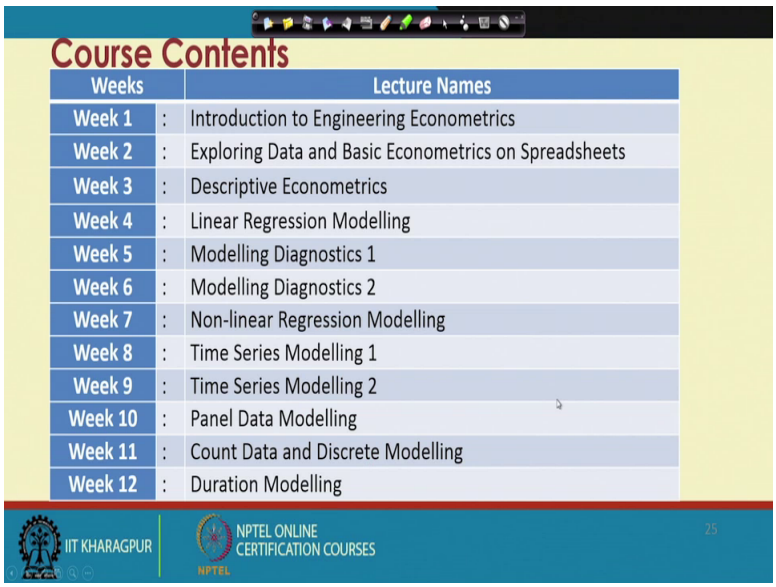
Engineering Econometrics
Prof. Rudra P. Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 03
Introduction to Engineering Econometrics (Contd.)

Hello everybody, and this is Rudra Pradhan here welcome to Engineering Econometrics and today we will start with lecture 3.

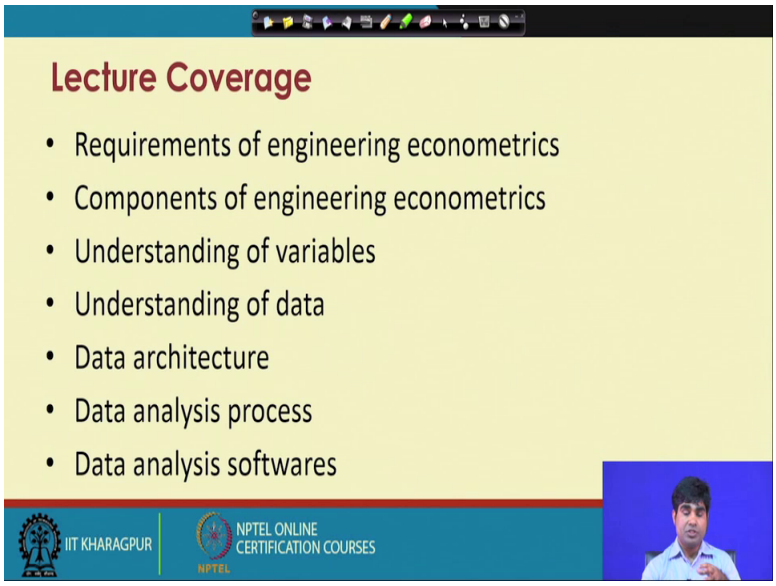
And again we will continue with the Introduction to Engineering Econometrics, let me highlight the a lecture contents.

(Refer Slide Time: 00:34)



And the coverage of this lecture will be like this.

(Refer Slide Time: 00:37)



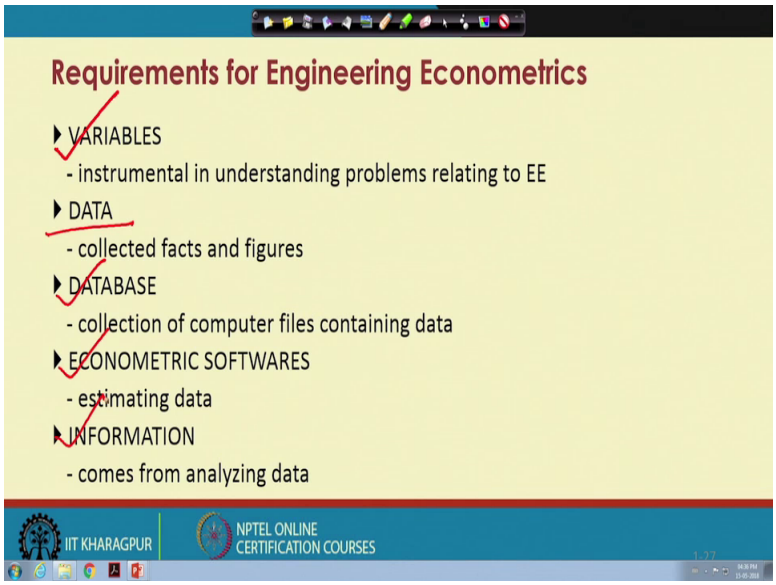
So, we will start with you know at requirements of engineering econometrics, components of engineering econometrics, understanding variables, understanding of data, then data architecture then data analysis process and finally, we will discuss something related to data analysis software's.

As we have already discuss you know engineering econometrics is a kind of you know quantitative subject and it includes the kind of you know problem that is the theory, then it will be connected with a data mathematical application and statistical application and then we will empirically verify to a to this theory and looking for some kind of you know a new insides.

So, in total it is the integration of theory data, mathematical tools and statistical tools. And the most important component is the data through which we can connect theory with you know model, and then in the process of you know you know that is the empirical process. We will get some kind of you know new insides and that will address the particular you know engineering problems in a kind of you know much better way.

So, that is how we are here to know certain you know a requirements. So, what are the specific requirements before we go for you know the kind of you know engineering econometrics.

(Refer Slide Time: 02:16)



Technically engineering econometrics a is a kind of you know tool baskets. So, where we will be first pick up the particular you know engineering problem, and then we like to pick up certain techniques and then connect with data, and then empirically process is to get some kind of you know econometric estimated output and that output can be helpful to analyze the problem you know in a kind of you know better shape.

So, the particular requirements are like this. So, we have to understand very carefully the variables that is the decision variables. So, in the in the econometrics kind of you know a requirement that is the engineering econometrics requirement. So, one of the most important is the data, and data will be very meaningful when you are connect with a kind of you know concept called as you know variables that is the decision variables.

And you know we have already you know discuss that you know the entire subject is you know more or less connected with a regression modeling. So, when you talk about the regression modeling, a the simple structure is the establishment of relationship among the decision variables.

So, we will discuss details about the regression modeling after this lecture, but before we start the process of you know regression modeling. So, we must know certain you know requirements, until unless you clear about these requirements and the kind of you know concepts. So, you are not in a position to do some kind of you know you know empirical

testing's or you know the use of engineering econometrics to the particular you know engineering problem.

So, the first and requirement the need of you know understanding is the variables. So, since it is a kind of you know regression modeling application. So, the variables can be two types so, the dependent variables and independent variable, and in the regression modeling. So, at least two variables must be there in the system and out of which one must be dependent variable, another must be independent variables.

So, from you know engineering to engineering, the concept of you know dependent variable, independent variable varies from you know one point to another point which I will address in a much better way in the next slide. But in the mean times in the simple language, variables means it is a kind of you know instrumental in understanding problems relating to engineering econometrics so; that means, through variables.

So, the theory can be quantify and first we quantify in a kind of you know mathematical form and then we transfer into a statistical form and after that we can connect with data and looking for estimation process that is the empirically investigation process and then we look for the kind of you know new insights on the basis of you know empirically estimated derive output.

And the second requirement is the data so; that means, the first requirement is the variables and the second requirement is the data. So, now, data itself cannot be meaningful until unless you know point out the variable structure. So, any information behind this you know variable is called as you know data.

So, that can be represented in various ways. So, data can be for a organization we for that particular variables in that true with respect to different point of time. So, there are many ways you can actually understand the concept of you know data. In fact, we have you know separate slides to detail about the data structuring data architecture etcetera.

But in the simple language, you know data means collected facts and figures for the decision variables. So, that decision variables reporting maybe for a particular organization, for a particular department and for a particular department with different point of time for a particular organization with the different point of time, like that there are many different ways you can actually specify the kind of you know data as per the

engineering econometrics requirement. Then next concept is a the database so; that means, any information related to variable is nothing what called as you know data.

So, now for engineering econometrics and that too when we solve some kind of you know engineering problems, you know you need actually data for this decision variables and. So, it is not so, easy to have data and you know just you know you know collate and you know start working, it is not you know all altogether easy process. So, there are two ways you can have the data for the kind of you know engineering econometrics requirement. So, it is a primary source that is called as a primary data, and the secondary source it is called as you know secondary data.

Primary source means you go to the particular you know correspondent or a particular organization at a particular point of time and then have the information corresponding to that particular you know decision variable. If that is the case that is that particular collection process or the reporting process is called the primary data and sometimes when somebody collect a the data from the first hand, that is called as a primary data and when it will be recorded somewhere else, then it is called as you know secondary data.

That means if the same data will be used by second person then it will be treated as you know secondary data. So, that is how the database structure is there. So, database you will be find plenty of database, where you know a particular organization will keep the data in storage, because of their requirement or because of their business or something like that and a researcher or the kind of you know investigators his or her job is to have this data from the database as per the engineering econometrics requirement, and then go for the kind of you know empirical process or the kind of you know empirical testing.

So, understanding variables data and data bases are you know mandatory requirement for engineering econometrics. Because knowing the database it will be very cost effective otherwise a you know every time you have to go to the respondents or the particular organization and collect the data or you know report the data.

So, now, if some data is same data is available somewhere else, which you may have knowledge very sufficient knowledge then it is better to go there and you can have the data for your you know requirement and of course, some database is a free and some database is a bad kind of you know concept.

So, whatever maybe the issue so, whether it is a free service or you know paid service, but it is a kind of you know basket where data is readily available, you just go and have it as per your you know requirement and then go for the kind of you know empirical testing.

And a next item which we need for you know engineering econometrics the econometrics software's. So, now, you know we are in a kind of you know concept called as you know analytics environment or analytics world, where we have a plenty of you know data that is what we called as you know big data so; that means, there is a huge pool of you know data.

Of course it depends upon you know problem to problem or you know variables to variable. So, assuming that you know for a particular problem for a particular variables you know we have you know big you know data points; that means, that is what technically called as a sample size.

So, usually in a in a kind of you know econometrics process larger the sample size; that means, larger the you know date availability, better is the model accuracy or better is the estimation process. That is not 100 percent correct, but a somehow it is a significantly or considerably true that you know increasing the sample size or increasing the data points your model accuracy or model fitness will be much better.

And when you have a plenty of you know data so; that means, a large sample size and there are a large number of variables involved in the process. In fact, I will tell you the a structure of you know regression modeling, where we can have a simple regression modeling, multiple regression modeling, multivariate regression modeling. So, when we start from simple to multiple, multiple to multivariate then the degree of complexity will start increasing.

So, when the degree of complexity will start increasing that is one way, and again the degree of complexity will be increase with respect to increase of you know sample size. So, in whatever you know case when the complexities very high. So, manually to solve the problem; that means, testing in the model empirical is not so easy.

So, in that contest we look for some software's, which can you know easily process the kind of you know you know the engineering requirement and then come with a kind of

you know estimated output, which can give you better results as per your you know requirements. So that means, the same thing which you are doing a earlier manually. So, now, you will do through mechanical process that through you know software's.

In fact, software software's are you know it is a kind of you know business now, you will defined plenty of you know econometrics software's which you can actually use to have the empirical you know output and this empirical output will be very useful for your you know engineering problems requirement. So, we have a typical slides to tell you the number of you know a software's you know, which are available in the market you can use anyone to you know use in your you know engineering econometrics.

And sometimes when your problem is very big and very complex so, that times. So, the use of econometric software is a kind of you know mandatory and typically when the problem is actually research oriented, then in that context. So, it is not so, easy to handle the you know issue manually that is the first issue.

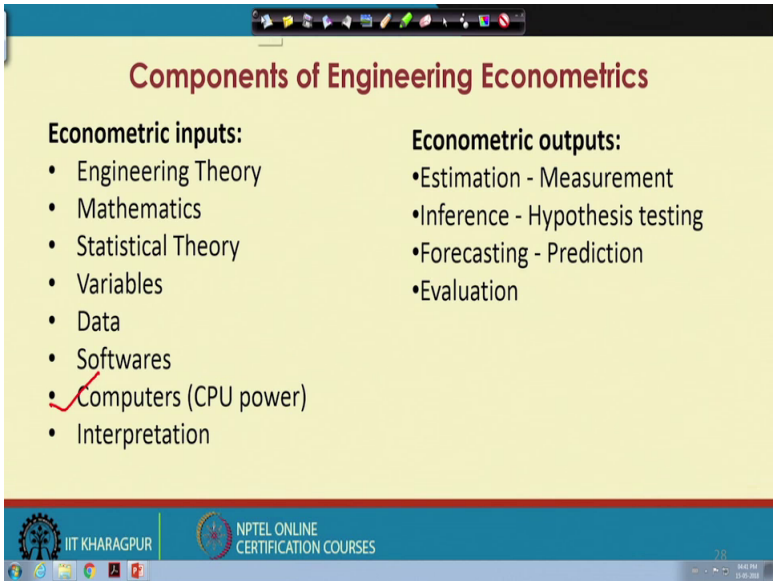
And second issue it can give you a quick service or you know quick output. So, it means a you know the thing which you can do manually 2 hours. So, within 2 minutes you can you know get the results; so, that is how a software's is a very easy to you know handle the process or you know manage the situation. So, what is the a suggestion here is that you know you must be acquainted with econometric software's if you are you know working on engineering econometrics. And last, but not the least is the structure called as you know information's that is comes from you know analyzing the data.

So; that means, we start with some you know information that is called as you know data first hand information and data is a kind of you know input in the econometric process, then in the econometric process you will be also get some kind of you know quantitative information that is called as you know econometric output.

So, that is a usually we like to say that you know analyzing the data. So, now, in the first hand first hand you know input that is the that is the structure called as you know data, again the second hand you know information that is again called as a econometric output. So, with see what is the you know information behind this you know empirical results and how best we can use as per your you know engineering requirement.

So, I like this. So, what we can say that you know, there are you know huge number of requirements in the case of you know engineering econometrics.

(Refer Slide Time: 14:37)



So, that is why we like to understand all this things. Let us you know start the you know little bit understanding about the econometrics process what are the actually components under you know a engineering econometrics. Of course, I have a highlighted the details in the previous slide, you know what is happening heres. So, we have a two different you know umbrella.

So, in one umbrella we have a econometric inputs and another umbrella we have a econometric outputs. Of course, the econometric output is derived from the econometrics input umbrella. So, in the in input baskets you know you have engineering theory then you know mathematical models and statistical theory or statistical model and variables details variables are there.

And then you must have a data and then you must have a software's how to apply the data as per your requirement and then a you must have computers to handle the software's. It is not so, easy actually without something you know called as you know you know without the help of computer you can handle the software.

So; that means, we need computers for you know you know installing software's to work out the particular you know engineering problems as per the a you know as per the use of

you know econometrics and then finally, you go for the kind of you know interpretation. And then in the case of you know econometric output, we have estimations that is called as you know measurement and then inference and hypothesis testing's, and the last part of this particular process called as you know prediction and forecasting, and then we will go for the kind of you know evaluation.

So that means, it is a very lengthy process altogether it is not so easy actually. Identifying the inputs in the econometric process is a somehow a you know very simple, but you know getting the econometric output is not so simples.

Because in between there are lots of test, you know and there are lots of you know checks or you know it is a kind of you know ultimately continuous process ok. So, you have to continuously a testing estimating then finally, you have to you have to fix a particular you know output level, which is good for here you know engineering problems requirements.

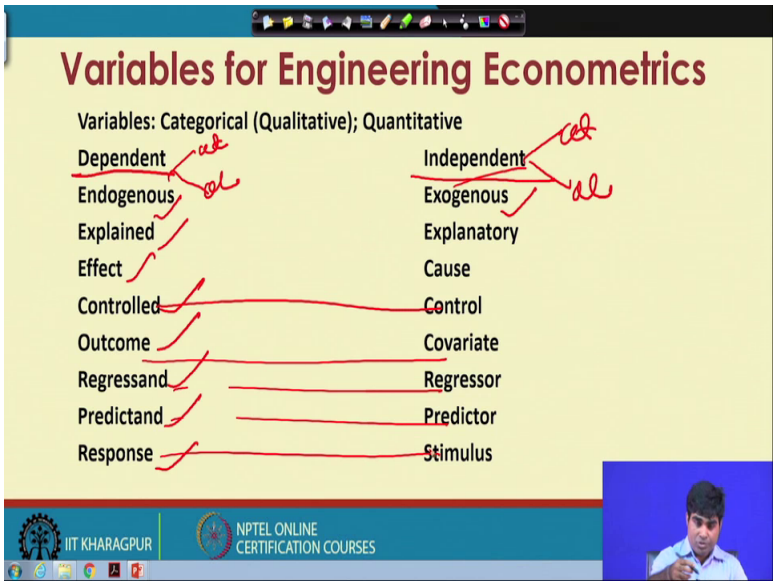
So that means, technically it is not to one way process altogether yes it looks like you know step by process, but it is look like you know one way process, but it is not like that. Yes, we have to follow the particular you know steps which you can which will mention in the later stage.

So, but ultimately you have to start with you know step by step by process and at a particular point of time, if you will not get the output desired output as per your you know requirement or something going wrong, then again he revisit the particular process. Again you start with you know particular step and agains you come to that point a.

So, again you have to check all these details, then if you satisfy and this is this is good for your you know analysis then you have to stop there. Otherwise it will be continuously you know you know iterative process you can say that you know iterative process, till you get the best output as per the particular engineering econometrics requirement right.

So, now in the in the next a process so, what we can do. So, this is how ok.

(Refer Slide Time: 18:18)



So, what I have mentioned earlier that you know one of the biggest requirement in the process of you know engineering econometrics is to understand the variables. So, what I have mentioned actually, variables can be two types a dependent variable and independent variable and again dependent variable can be a qualitative and can be quantitative that is a numeric.

And independent variables can be again you know quantitative and a qualitative which we have already highlighted in our previous lecture, but still I am once again you know I am highlighting this, because it is very important for any kind of you know regression modeling. Whatever, you apply any kind of you know engineering problem this clarity you must have, otherwise you know it is very difficult to you know you know go for the kind of you know processing.

Because if you are you know classification of dependent variable and classification independent variable is wrong, then you will not get you know best estimated output and you may not get the desired output as per your you know particular requirement. There is a kind of you know theorem called as a blue best linear unbiased estimator.

So, we will be discuss later in details about this particular you know theorem, but the theorem will give you the simple message that you know the estimated a model is a free from all kinds of in error and it is the best for any you know for analyzing that particular you know engineering problem.

So that means, technically the theorem cannot give the optimum result until unless everything will be in a systematics you know reporting or this a systematic way you have to you know a design. So, one of the requirement of this particular process is you know that variables understanding. So, let us see you know how many are dependent variable, how many are independent variable and how many are you know qualitative in nature, how many are quantitative in nature.

So, we need actually details kind of you know you know reporting. So, then you have to see how these variables are you know related to each other is it a bivariate case, is it a multiple case or it is a kind of you know multivariate case.

So, in the case of multivariate it is a very a complex kind of you know things, because every step you will find you know different kind of you know functional form of different kind of you know equation. Even if you are objective is a very you know specific still a there is a need of you know connection to reach that particular you know objective.

So, that is how your clear cut understanding must be a very essential. So, now, in the case of you know dependent variable and independent variable, which I have highlighted here. So, this is what the dependent variable and this is what the independent variable, this can be a this can be quantitative this can be qualitative and this can be also quantitative and this can be qualitative.

And this understanding or this reporting is very important, because on the basis of this you know the involvement of quantity variables or qualitative variables, your choice of the model will be also different. So that means, there are certain econometrics tools are there it is a variable specific. If you have if you are using qualitative variables a dependent in the dependent variable case, then your model structure will be different if it is in the case of independent variable, then the model structure will be also different.

So, that is why you must see how many variables are you know there in the system and how many are qualitatively qualitative in nature and how many are you know quantitative in nature.

And we are dealing with you know various engineering problem starting with you know aerospace engineering, biotechnology to mining engineering and you will be find you

know different turns or you know frequently used, there is a excellent journal called as you know techno matrix.

So, you will find you know different kind of you know engineering problems are there you know and discuss empirically. And you will be find the same dependent variable can be used differently by using the name of you know endogenous variables, explain variables, effect controlled, outcome, regressand, predictand response.

So, these are all you know called as a dependent variables and it is frequently used in different kind of you know engineering, but ultimately this is the basket of you know dependent variable. Where using dependent variable is fine otherwise instead of dependent variable you can say that you know it is a endogenous variables or you can say that it is a effect, you can say that it is a kind of you know controlled.

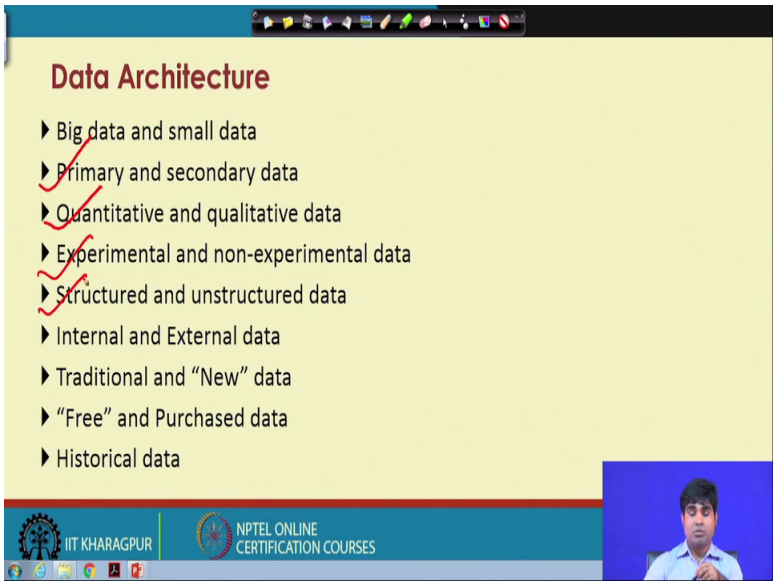
And on the contrary so, the independent variables can be reported like you know it is it can be called as you know exogenous variable, it can be also called as a explanatory variable; cause means if you say that independent variable is cause then the counterpart will be called as you know effect that is the dependent variable. So, where you will called as you know independent variable otherwise called as you know control, then the counterpart is called as a control.

So; that means, you know it is altogether you know have a different kind of you know naming procedure. So, you must have actually the particular structure to understand the kind of you know concept, and the detailed classification and you know similarity are there is a book called as you know Basic Econometrics by DN Gujarati.

If you go through that book in the first chapter itself you will be find this particular you know declarations, the a difference or the similarity between dependent variable and independent variable. So, accordingly so, what is the requirement that you know, you must be very clear about the dependent variable and independent variable.

For instance instead of you know independent if I will use covariate if you are not equant means you are not acquainted, then you know you may assume that the covariate term may be the dependent variables. So, then the entire problem will be different shape altogether. So, you must be very clear about this particular pictures right. So, ultimately.

(Refer Slide Time: 24:16)



So, we have to move again further to know about the data structures, what I have already mentioned there are you know couple of ways you can actually put the data understanding. So, it is question of you know big data versus small data so; that means, in the basket if you have actually big pull of you know data, that is called as you know big data and if it is a small structure of the data it is called as you know small data.

In fact, we have a concept called as you know large sample and small sample. So, the large sample concept means econometrically is you know it depends upon you know sample size a let us say 30 32. So, if it is crossing a above 32, then it is called as you know large samples otherwise it is called as you know small you know small samples. So that means, large sample and small sample depends upon the number of you know sample points.

Usually the bar is actually 30 32 and then if it is less than it is called as you know small samples otherwise it is called as you know large sample. So, in other words if you say big data it is means huge full of you know data and then if you say that you know small data that not necessarily mean that in a small samples, but it is a large sample, but it is the for instance 50 or 100 in comparison with you 5000 or 10,000.

If you say 10,000 it is a big data, than 100 may be a small data this the kind of you know example or you know understanding you must have. Then the data can be again can be a

primary kind of you know structure and secondary kind of you know structure, then you called as you know primary data and secondary data.

Similarly, it can be a quantitative and qualitative structure then quantitative data and qualitative data, which I have already highlighted. So that means, this is actually we have already highlighted and quantitative qualitative I have already also you know mention.

And experimental and non experimental data and some data are you know used in the particular you know investigation process that is called as you know experimental data and where the accuracy is must higher. In non experimental data the accuracy not so, high and you have to you know do lots of hardworking to get the best output in the process of you know empirical testing.

And then structured and unstructured data, so that means, once you have a data first hand data that may not be in a structured form. So, because your requirement maybe something different and usually first hand data, we called as you know unstructured and then we put in a kind of you know structured format depending upon your you know kind of you know requirement.

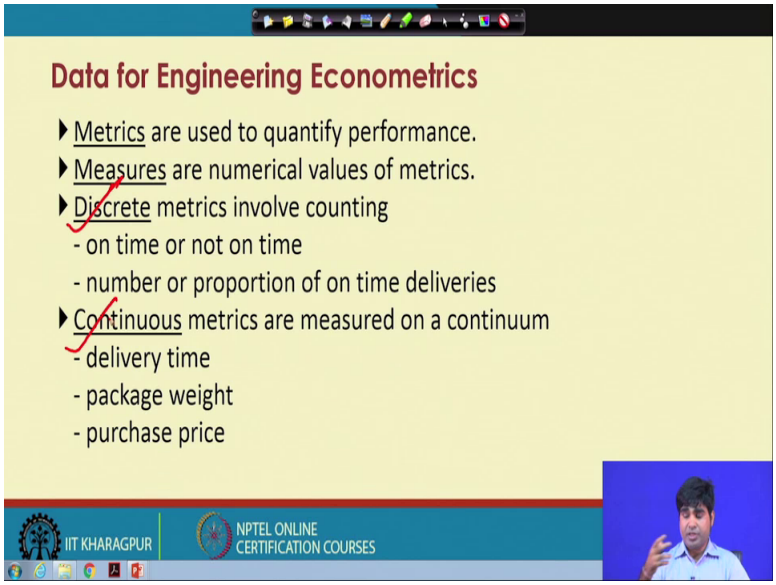
Sometimes you can do data transformation do data transfers something like that, and then put in a kind of you know structured way. So, that you know your analysis can be much better and you know here you know empirical investigation process will be you know very effective and a within a quick time you can get the empirical output.

Then internal data and external data; In fact, a you know external you will get lots of you know data and again this same data can be collected internally, and as a result you can called as you know internal data and external data.

Similarly data can be traditional and new data. Traditional data means it is already available then new data means a newly generated you know data maybe experimental process may be non experimental process and free data and purchased data, that is what I have already highlighted in the contest of you know database. Some data a base are you know freely available and some database are you know data will be available with you know kind of you know purchased only and then there is a concept called as you know historical data

We have a data structure called as you know time series data generally historical data is nothing, but called as you know time series data. So, this is altogether called as a data architecture through which you can analyze the particular you know process or you know most have a clear cut understanding about the data use in a kind of you know engineering econometrics, you know requirement.

(Refer Slide Time: 28:29)



Agains in the engineering econometrics so, far as a data is concerned, we use the term called as a metrics and that are used to quantify some kind of you know variables. Sometimes measures are you know numerical values for values of you know metrics and it maybe a discrete, it may be a continuous.

Discrete means it is kind of you know on time or not on time, this technically it can be a discrete and a it can be a continuous, but continuous it is a you know very good for the kind of you know empirical testing, that is like you know on delivery time, package weight, purchase price.

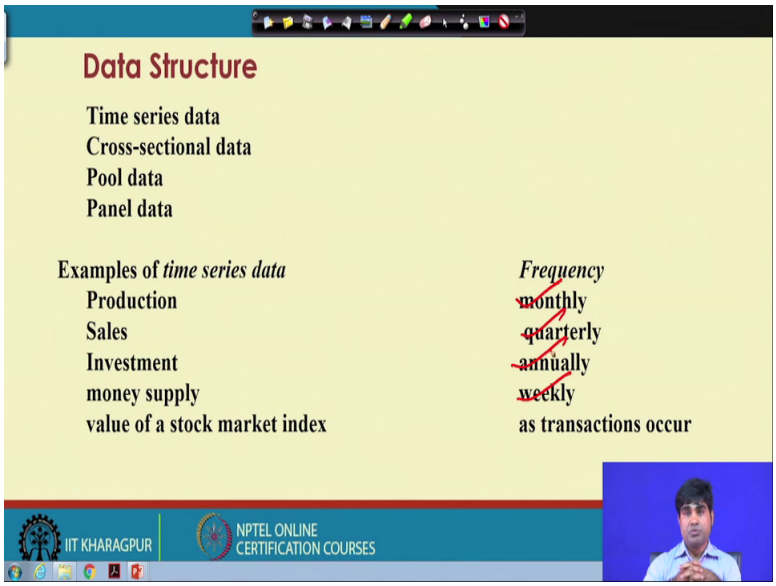
So, you must be very careful whether the available data is in a discrete format and the continuous format. Ultimately we need to analyze the particular problem if data will be availables corresponding to a particular you know variables. If data will not available, it is very difficult to analyze the process or you know doing the kind of you know empirical investigation. In fact, some information's or some reporting in the first hand is completely qualitative, like you know let us say gender impact on you know incomes.

So, in that case gender cannot be by default called as you know in a numeric format that is how structured and unstructured concept will be coming into the pictures.

So, if gender is a kind of you know variables, then the reporting will be either male and female. If you write male and female in the data in investigation process or data analysis process to work out so, what will you do? So, you transferred to male and female transfer male and female to some coding let say 1 2 or 0 1, then you analyze it is very easy for doing the kind of you know jobs.

So, that is how we need to know the kind of you know measures and that to the metrics measures are to reporting the data numerically as per the particular requirement, and the you know making the process easy to get the empirical output as per the engineering requirement.

(Refer Slide Time: 30:46)

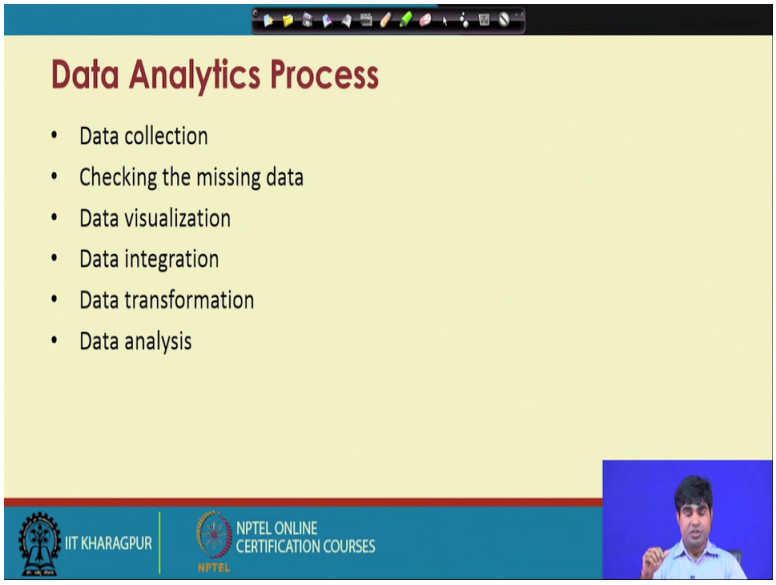


And by the way so, we are in the data investigation process and the type of data structure will be like this, and I will let you know in the later stage.

But what I will be like to tell you that you know when you go for empirical investigation process. So, data is the one of the vital component and the we can called as you know wheel in the process of you know in the process of you know engineering and. So, when we use the data for any kind of you know investigation process. So, the reporting of data can be time series, that is you know reported with respect to time for a particular

organization or a for a particular you know sector or something like that. So, the time may be annually like here we have already reported. So, it maybe monthly, it can be quarterly, even if sometimes it may be it may be also day wise reporting.

(Refer Slide Time: 31:38)



For instance you know monthly, quarterly, annually, weekly these are all you know time series reporting and a cross sectional data means, same you know one particular point of times and the data is reported for a particular organization, for a particular sector, particular unit or particular individual. So, then it is called as you know cross sectional reporting.

So, now, when your pulling time and cross, then that becomes called as you know pull data and then when you are you know targeting the impact of you know cross sectional impact and time series impact on the pull data, then it becomes called as you know panel data.

So, we have a panel data modeling during the times, I will highlight in details about the panel data structure start with you know a simple cross sectional time series pull and you know penal. This is how the hierarchy you have to maintained to understand the their data and use the data for the particular you know empirical investigation process; that means, in the same problems if you have a data with you know time series and you know cross sectional, then you can you can go for you know cross sectional modeling, you can go for time series modeling, you can go for you know panel data modeling.

So, in the later stage picking we pick up a particular you know engineering problem and we can use a data set, where we have a cross sectional data and time series data together then by default we can have a many different models depending upon the use of data for the particular you know investigation process.

The cross sectional data, cross sectional modeling time series data, time series modeling, then panel data panel data modeling. So, likewise it is a very interesting process. Once you understand the kind of you know data structure, then you can just connect the connect with a particular model and you will be find you know plenty of you know beautiful results to analyze the particular you know engineering problems.

So, this is all about the you know data structure and then we have we have the kind of you know concept called as you know data analysis means a data analytics process. So, ultimately when will be when we are you know dealing with you know data in the kind of you know engineering econometrics.

So, one of the big component is the data collections that is data gathering, then the checking the missing data. Sometimes you know when you are reporting a data for a particular variable with respect to time for a particular organization some point of time the data is missing or one point particular point of time with different organization some organization maybe missing, that is called as you know sign of you know missing data.

But ultimately when you are doing regression analysis or regression modeling or some kind of you know econometric modeling. So, in that contest you know you must have a data and; that means, uniform data you must have for all variables. If any kind of you know missing either you can drop that particular data point or you can delete the data point before you start the you know analysis.

But deleting the data omitting the data is not the solution, sometimes you know if you if it is a time series data, if you just omit the data then it may affect the time series features. So, you must have a some kind of you know mechanism how you have to address this particular situation.

Will it discuss in details in the later stage. So, there is a technique called as you know interpolation and extrapolation through which you can you know somehow manage the missing data to handle the situation, but sometimes when you use interpolation and

extrapolation to use the data, some other errors maybe a coming into the picture, but still we can manage how to handle the particular you know scenario. Then we have a data visualization and data visualization means regression basically represents the functional relationship between variables, we will discuss details in the later stage. So, since it is a functional relationship. So, we initially may not have any idea, whether is a linear kind of you know structure or not linear kind of you know structure.

So, in that case data visualization is a mandatory. So, once you visualize then you can pick up a appropriate functional form if you pick up appropriate functional form then your model will be very good and your estimation process will be very equally very good. So, you must be do the data visualization before you go for any kind of you know modeling.

And then data integration sometimes you know you have to club the data properly as per the particular you know requirement and data transformation, sometimes you know you have a data which I have lots of variety you know in a for a variable x and you have actually 10 data points. One data point 3 digit, one data point is a single digit, one data point is a double digit like this then you will define lots of volatile is there. So, when you have a large number of volatile, then model accuracy will be down in the process of you know empirical testing.

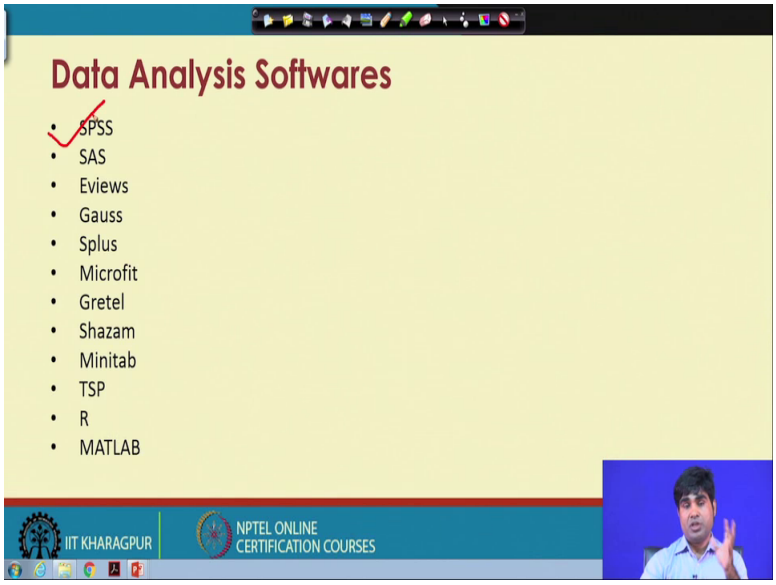
So, in order to avoid such kind of you know problems, we can go for you know data transformation. So, the there are many different ways you can go for you know data transformation like you know log transformation, fast different transformations and some kind of you know normalizations. So, we will discuss in details about the data transformation in the later stage and mostly do in the case of you know time series modeling and then finally, the concept called as you know data analysis.

And it is one of the biggest component in the engineering econometrics, if you know what kind of you know modeling to choose and how to go for you know estimation and what kind of you know analysis we will do what kind of you know output will be obtain and whether you can again continuously re estimate if your output is not coming as per your expectation.

So, the process you know data analysis is a very you know very complex and you know very complicated and for that you must be you must have a patience and you must know

all these obstacles and you know kind of you know issues and you know kind of you know problems, then you can address the issues and problems very effectively as per your you know problem requirement.

(Refer Slide Time: 37:56)



And finally the kind of you know called as you know data analysis software's. So, we have actually plenty of you know software's, which you can use for this you know data analysis, which you I have already highlighted and these are all the listed you know software's, which you can actually a used in the in the kind of you know engineering econometrics process.

Starting with a SPSS, SAS, Eviews, Gauss, Splus, Microfit, Gretel, Shazam, Minitab TSP, R, MATLAB; So, just we listed you know few there are plenty of you know econometric software's it is a kind of you know big business now. So, but most of the software's are you know more or less sames either you connect with the programming or recently if I you know window version just you know operate a in a kind of you know mode you will get the kind of you know output. So, you should know the tricks everythings before you go to the software.

So, the my suggestion is that you know before you enter to the software's to handle the you know to get the output or to you know make the process. So, you must be clear about the theory, clear about the models, most have a data, most have clear

understanding of the you know variables or the functionality, then you go for go to this you know go through the software process and then you can easily get the output.

Once everything is clear then definitely you will get the clear cut results and then you can equally actually analyze the problem more effectively. So that means, technically it is very complicated process, but you can simplify if you are understanding is a clear and some of the infrastructural support is a with you to deal with this kind of you know problem. So, with this we will stop here.

Thank you very much have a nice day.