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

Lecture – 06
Exploring Data on Spreadsheets

Hello everybody, this is Rudra Pradhan here welcome to Engineering Econometrics and today we will start with unit 2 and that to exploring data on spreadsheets and as per the engineering econometrics requirement. In fact, in the first unit we have discussed details about the engineering econometrics the kind of you know requirement, the kind of you know scope, the kind of you know methodology, the kind of you know techniques availability, the kind of you know forecastings.

So, depending upon the particular you know engineering problems you know we have to choose a particular technique, we have to choose a particular you know methodology to do the kind of you know modeling, to do the kind of you know prediction and to do the kind of you know forecasting. But in reality whatever the kind of you know engineering problem, you can do better predictions, better modeling, better forecastings if you are you know data will be very supportive

So, one of the way to do or to analyze the engineering problem much better way, if your data is readily available as per the particular requirement in fact, we have already discussed the concept called as you know data structure. So, first of all any engineering problems when will you do modeling; so, you have to transfer into oh a form of model and a the model can be structured with the help of some of the decision variables.

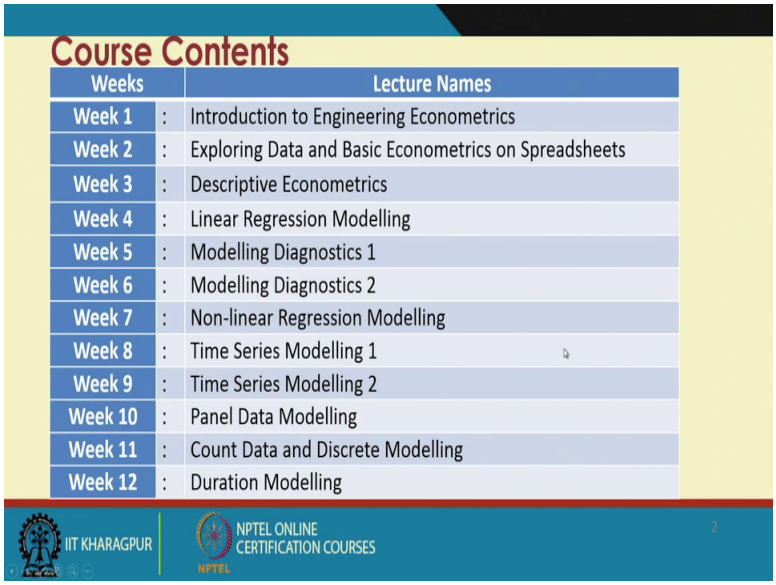
So, now, as per the kind of you know requirement, the kind of you know the kind of you know requirement in the sense the prediction requirement or forecasting requirement. So, you must have data you know behind these decision variables. So, now, the engineering econometrics requirement is that you know the problem can be transfer into a model with the help of you know decision variable. So that means, we should know, how the decision variables are you know integrated each other or functionally related to each other to analyze the engineering problem and address the engineering problem as per the particular requirement.

Now, empirically to validate this model we need actually data, we need actually engineering econometrics tools. So, that is why we have discussed in details, the kind of you know engineering econometrics techniques, the kind of you know data the kind of you know data use. So, here what is requirement means typically in this unit we like to know more about the data because it is the who will in the engineering econometrics, until unless you understand the data perfectly you are not in a position to pick up a good technique and a cannot do a better you know modeling and better prediction or better forecasting.

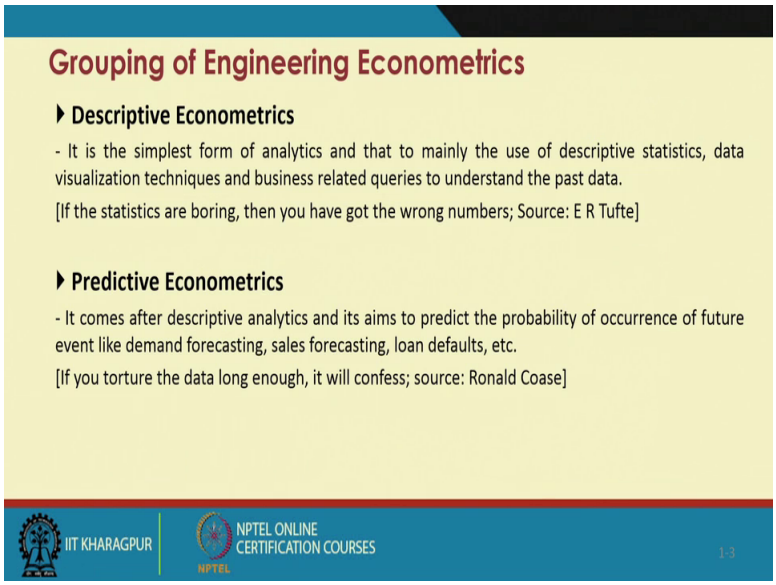
So, that is why understanding of data with respect to you know the kind of you know the data, kind of you know visualization process or something like that you are not in a position to do the better modeling or to do the better kind of you know forecasting you know something like that.

So, let us you know start with here the idea called as you know exploring data on spreadsheet because you know most important thing is here the kind of you know spread sheet uses.

(Refer Slide Time: 03:35)



(Refer Slide Time: 03:36)



So, when you are actually talking about data and that to decision variables. So, we need actually the use of you know spread sheet to report the data against the decision variables. So, now, when you report the data against the decision variables, so, the spreadsheet itself will give you a particular structure, how to proceed, what kind of you know technique you can use, how to do the kind of you know structuring destructuring as per the particular you know modeling requirement or the kind of you know technique requirement.

So, before you move to the use of spreadsheet for data or the kind of you know modeling, let me give you the you know some kind of you know hint about the engineering econometrics. So, of course, we have already discussed this in unit 2 unit 1.

So, here ones again you know we are highlighting this. So, in the first case means first two thing is it a called as you know descriptive econometrics, where you know we can say that you know it is the simplest form of analytics that to mainly the use of descriptive statistics, data visualization technique and business related queries to understand the first data.

If the statistic you know if you say that if the statistics are you know boring, then you have got the wrong numbers similarly a the kind of you know predictive econometrics you know it comes after descriptive analytics, and its objective is to predict the

probability of occurrence of future event like you know demand forecasting, sale forecastings, production forecasting and something like that.

So, here the idea is you know how best you can torture the data means in a long way to find out the reality. So, in one sense you have to read the data and then you have to actually explore or you know torture the data to get some kind of you know insides; that is the game between descriptive econometrics to predictive econometrics.

So, now with you know you are in a position to analyze both the econometrics and that to with the help of data only. So, without the availability of data, without the use of data you cannot actually start with descriptive econometrics or we cannot start with you know predictive econometrics.

So, that is how data is the big component for this kind of you know quantitive analysis and that to for this engineering econometrics. So, now, you know means technically what is the basic objectives behind this you know descriptive econometrics and a predictive econometrics.

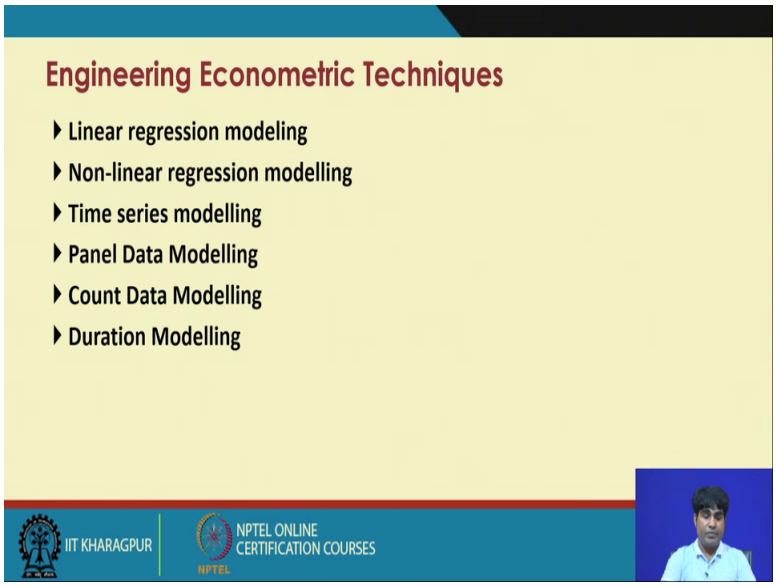
(Refer Slide Time: 06:31)

In next units we will discuss detail about this you know descriptive econometrics and a predictive econometrics, but in the mean time what I like to say in descriptive econometrics basically explains, what happened in the past and like you know many

organizations used DA descriptives econometrics as a part of you know business intelligence.

So, it is like you know you know descriptive analytics and similarly predictive econometrics, it is also something like called as you know predictive analytics. So, here the issues are you know what will happen in the future, and many organizations use predictive econometrics a to do the predictions and the kind of you know forecastings.

(Refer Slide Time: 07:22)



So, now before we going to details about the spreadsheet use. So, I like to point out here you know some of the techniques behind the engineering econometrics. So, sometimes what is happening, knowing the techniques. So, you can you know you know use the data accordingly or you can understand the data accordingly.

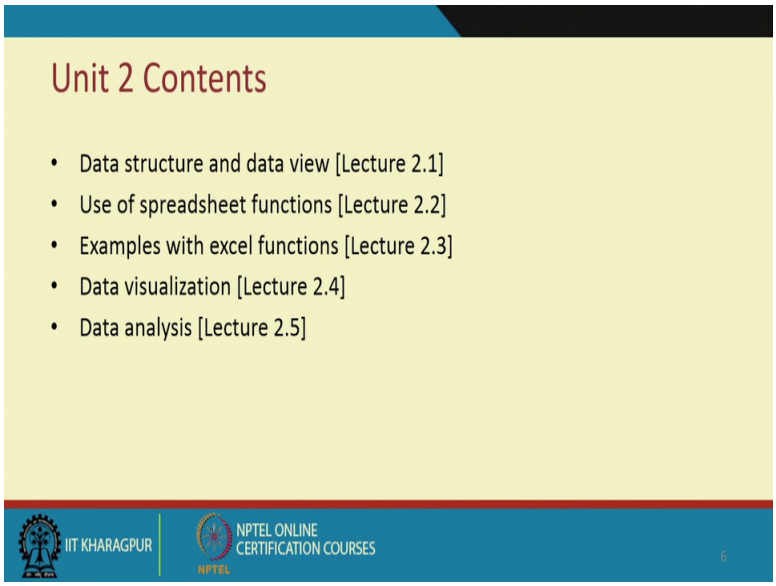
So, for as this subject is concerned, we are suppose to connect with linear regression modeling, non-linear regression modeling, time series modeling, panel data modeling then we have count data modeling and finally, we have a duration modeling. So, now, all these modeling depends upon the kind of you know data structure so; that means, these divisions are on the basis of you know pattern of data and the kind of you know problem.

Sometimes if the variable structure is the qualitative if the model structural will be different sometimes, if data is a kind of you know cross sectional type or time series type

then the modeling structure will be also different. So, that is why a you should actually know the problems, understanding the problems in both the aspects.

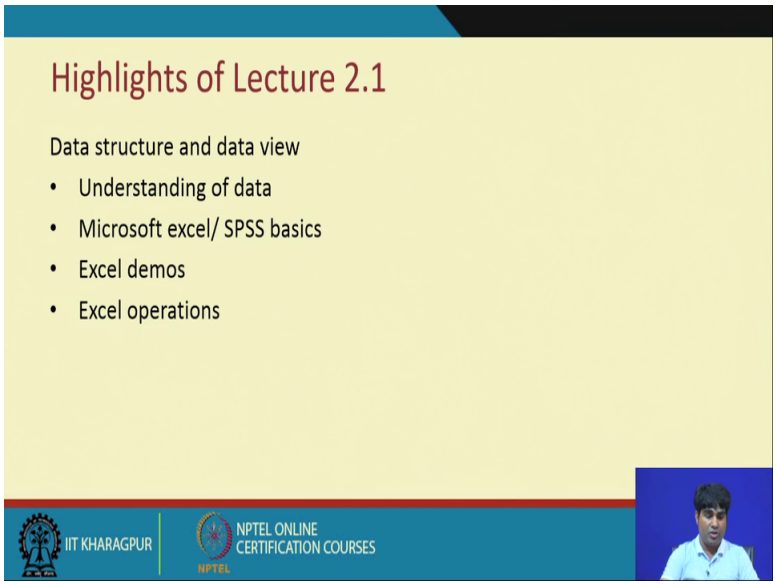
So, once you understand the problem both the aspects then you may be in a center position to decide, which particular mechanisms or which particular path we have to follow so, that you can do the better forecastings or better predictions as per the particular you know engineering requirement.

(Refer Slide Time: 08:53)



So, in this unit a you know we like to we like to know the data structure and data view, then we can know the use of spreadsheet functions and then we can connect with some examples that to with excel functions, then we will go for data visualizations and some kind of you know data analysis, that is at the basic you know on the basis of you know basic statistics.

(Refer Slide Time: 09:19)



So, let us start with you know the first one. So, that is called as you know data structure and data view. So, here the idea is actually on a the understanding of the data and the use of Microsoft excel. So, SPSS or how to enter the data, how to read the data, how to you know manipulate the data or you can say that you know process the data, then some kind of you know as spread sheet you know demos and some kind of you know excel operations.

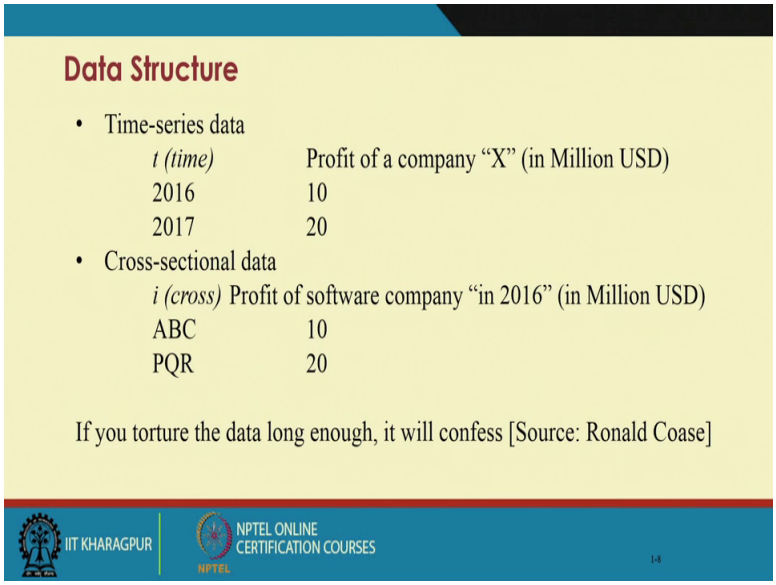
So; that means, a when you enter the data that is called as you know firsthand data sometimes we called as you know primary data. But you know the particular data may not be may not be directly used for the kind of you know modeling requirement. For the modeling requirement the first hand data can be transferred can be actually somehow to you know structured in a particular format as per the particular you know modeling requirement.

So, once you know all these details, then you may be in a position to do the kind of you know transpose transforms and in the kind of you know structuring restructuring as per the particular you know modeling requirement. So, when you have a when you have a data in a kind of you know spreadsheet let us say in excel, then you know doing any kind of you know transformations or you know transpose or the kind of you know operation. So, it is very easy to do that.

But you know manually it is not so easy to do because sometimes your data size maybe very high and your you know variable size is also very high, sometimes you need operations variable to variables and with a particular variable with a you know set up the data. So, all kinds of you know operations you can easily do in that in the help of you know excel sheet.

Sometimes some of the softwares if they need actually the final data which can be operative or you know for the functionality of a particular you know model. So, I mean see you assume that you know excel is a like you know kind of you know homework kind of you know things, where you have to process the data as per the particular you know modeling requirement.

(Refer Slide Time: 11:34)



And what I have mentioned earlier the kind of you know data structure the basic data structure is a time series data and cross sectional data, now when you club this two clubbed this two then it can be either called as you know pool and panel depending upon the kind of you know structuring and restructuring, but in the firsthand the reading will be the cross sectional type and time series types.

So, for the you know Layman's language, data is a data can be two types; that means, time series reporting and cross sectional reporting. So, then a you know after knowing the time series reporting and time series understanding cross sectional reporting and cross sectional understanding, then it can go for you know a pool and penal.

So, here the simple example is in nothing, but called as you know time series a reporting. The time series reporting here actually the let us say we have taken here actually t equal to 2016 and 2017, and a in the second column we have taken profit of a company X so; that means, it is the annual reporting and in million USD's.

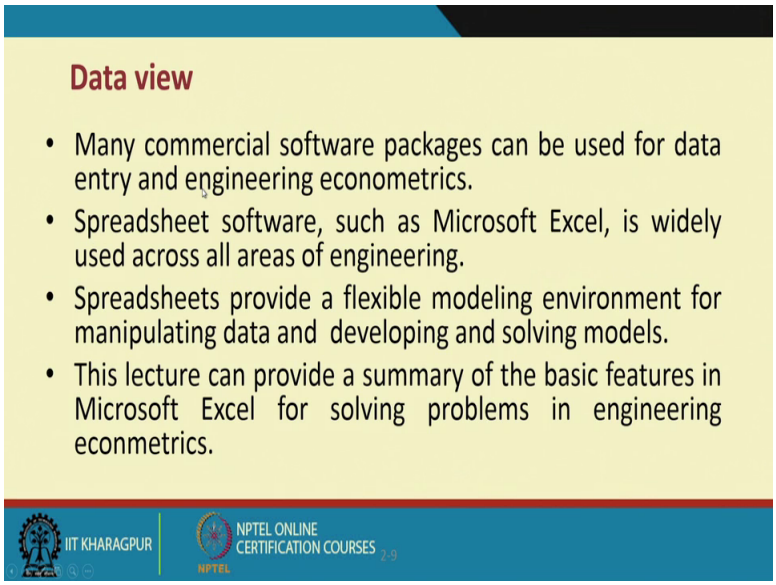
So, one thing very clear here; So, when you are reporting data with respect to time or with respect to any cross sectional unit. So, unit of measurement is very essential, but sometimes what happeneds with you know use of excels, we can actually a process the data in such a way the unit can you know removed ultimately for instants we can standardize the data

So, the way will standardize the data. So, by default the units can be remove in the process and regression is a kind of you know technique by default it will you know standardize the process and make the things in such a way, sometimes unit can be removed in the process of you know analysis a you know what we can do here in the first instance. So, the time series reporting means in the case of time series reporting.

So, your cross sectional unit will be fix, for instance with a particular organization what is the profit with different years and what is the profit with you know one year with you know different once like this. But in the case of you know cross sectional data it is the time unit fix, but cross sectional unit vary. For instance let us say in a year 2016, what is the profit of company x company y company z like this.

So; that means, a whether you are using time series data or cross sectional data, both cross sectional unit and time series unit will be there. In one instance cross sectional unit will be constant for a for a particular instance, then time will be vary, but in the case of you know cross sectional kind of you know situation where time at a particular point of time fix and cross sectional part will be a you know varying. So, this is how the a simple understanding about the time series data and cross sectional data.

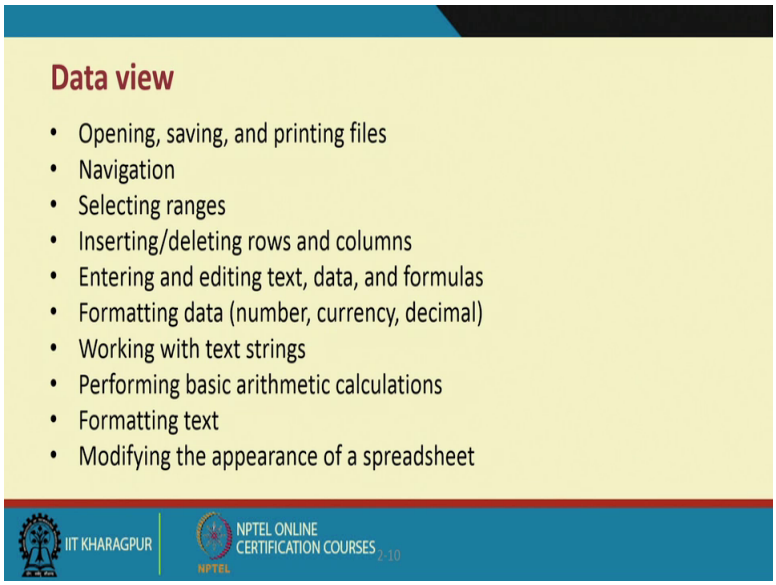
(Refer Slide Time: 14:39)



Now, so, for a data view is concern, many commercial software packages can be used for data entry and the kind of you know engineering econometrics.

So, what I have mentioned earlier, spreadsheet software such as you know Microsoft excel is a you know widely used across all areas of engineering and why because spreadsheets provides a flexible modeling environment for manipulating data and developing and solving various kind of you know engineering models. And this particular lecture can provide a summary of the basic features in Microsoft excel for solving problems as per the engineering econometrics requirement.

(Refer Slide Time: 15:25)



So, far as a data view is concerned so, you go to the excel and you should know all these kind of you know requirements. For instance you know in a kind of you know spreadsheet, the format will be like this it is a kind of you know matrix kind of you know things.

So, you will find a you know row wise a, a kind of you know box and column wise box. So, if you say that you know row wise a box will be decision variables say x 1, x 2, x 3, x 4, then column wise information will be either time specific or cross sectional specific and if you club these two again then it may be pool and you know panel specific.

So, now, let us assume that you know you have a decision variable let say x 1 and x 2, x 3 in the row wise a entry and column wise entry may be time series a you know you know reporting where you have early information let us say 2015, 2016, 2017 and so on.

Then you know while reporting all these information against all these decision variable then you can actually save the file, then you should know how to save the file, how to open the file then how to print the files as per your you know particular requirement.

Then you know navigation selecting ranges because you may have a data actually say 100 points or (Refer Time: 16:48) 1000 point sometimes statistics can be applied to a few data points only. For instance let us say you have 1000 data points from you know 2000 to 1 to 2017 in day wise informations for a particular variable or you know few variable.

You need actually for you know fast 100 entry or you know 200 entry; just you know you select these a particular you know items and then pick up a data analysis technique and you know give the options for you know analyzing or you know analysis, a then a you can choose a particular technique and then allow the software to you know run the models ultimately we will get the analysis.

So, you should know how to select the range row wise and column wise, then sometimes you may have option to insert and deleting rows and columns as per the particular requirement and why it is required here because sometimes there maybe chance that you know some missing observations are there..

Of course, in the last lectures we have already discuss that you know sometimes when there is a missing observations you can use interpolation and extrapolation to fill the gap as per the particular requirement or the particular you know techniques need.

But sometimes what happens the type of you know data when there is no such requirement. For instance if it is not time series data and it is a cross sectional repeating and that to there is a missing observations, then there is there is a high chance that you can actually a delete that particular you know row or you know delete that particular you know column.

So, you should know how to delete a particular row and particular column without disturbing the entire dataset. So, you can do that when you see that you know the particular row and particular column may have some problems, and deleting these rows can you know can give you better picture of the data to visualize and to do the analysis for the particular you know engineering requirement.

So, then again you should know how to entering the data, editing text data, formulas etcetera. For instance in a in a kind of you know excel scenario you may report the data which is actually qualitative in nature for a particular decision variable. For instance let us say the decision variable name is genders. So, the gender reporting by default will be male and female initially, when you get firsthand informations and in the excel sheet you can put a you know simply male against male you can put male against female you can put male and then you report all these information.

But when gender will be called as you know decision variable, then male females when you do kind of I mean the kind of you know modeling. So, the software may not read what is male and females and cannot do the kind of you know processing. So, in that case excel has the beauty that you know you just give the command that you know male can convert into 1 and female can be convert into 2, then you just you know you know do for you know all the columns or you know all the row as per the particular requirement then by default.

So, it will create a new row or new column and where the male entry you will convert into a particular number say 1 and female entry can be convert into a particular number say 2, the ultimately the qualitative informations will be quantified and then it will be readily available for you know immediate you know operation that to data analysis for the particular you know modeling requirement.

And it is very easy to do this for you know if you know the excel operations and that to the use of you know spreadsheet and manually you cannot just check and you know do continuously if your data set is very big.

So, in the in the use of you know a spreadsheet and excel, if you just you know do for one particular row or particular column, then you allow for every row and every column by default in within just you know 1 minute, you can you know do the transformations as per your you know particular requirement.

So, that is why you should not know how to enter the data, how to edit the data and how to transpose the data. And excel has also beauty to format the data so; that means, same data we start with you know numbers and it can transfer into currency format and it can be transfer into decimal format, it can be transfer into percentage format.

So, it has the option. So, that is why a means a the use of spreadsheet like excel is nothing, but a kind of you know softwares. In fact, in the excel we have a data analysis package, some of the some of the models can be analyze in the excel itself ok. So, you just installed the data analysis package and then you just give the command as per your you know particular requirement, then by default the data analysis package can a read the particular you know data and as per here it command.

So, it will a transpose these inputs to econometric output as per here you know engineering requirement.

So, that is why a knowing the excel use or knowing the spreadsheet use is very essential or you know kind of you know mandatory requirement for engineering econometrics. In fact, in the later stage when we use you know advanced kind of you know model like time series modeling panel data modeling, then you know any kind of you know advance modeling a nominal modeling a in that contest what will you do you know, you can do lots of you know transformation as per the particular you know requirement in that case what will you what will what will you do.

So, you should actually know the particular you know structure and then you just follow up as per your you know as per your you know need and unique kind of you know requirement. For instance you know you need actually a multiple entry or you know multiple kind you know setting.

So, excel can help you to you know do the you know necessary a kind of you know requirement. For instance if you have I if I have a kind of you know you know multivariate problems, where we need to you know club these two variables to work out the you know issue like you know interactive effect or you know dummy effect.

So, in that context, excel operation is very useful and what is happening in the advanced econometrics techniques you know, you need actually the basic operation first, then you will go for you know advance operations. So, in the advance technique it may not be very easy to do the operations, but excel can do very easily these kind of you know operations. So, that is why before using any kind of you know advance technique it is better to know how to do the kind of you know data entry and data operation etcetera.

But in fact, the beauty is that you know in the excel sheet, we have actually import option and export option and when we use advanced econometric software, you just open the econometric sofwares, then you give the import options or you know a export options. And if you do the import options by default, it will transfer the data from the excel sheet to here you know particular you know econometric softwares and after that you can do the kind of you know analysis. So, that is why a my suggestion is that you know in the firsthand a kind of you know requirement for any kind of you know engineering econometrics.

So, you report the data in a kind of you know excel sheet and spreadsheet, then as per the particular you know engineering problems requirement, you have to choose a particular you know engineering econometric tools to analyze the problem. What is happening means what is more important is that some of the problem is very technique specific right.

So, for instance it is a kind of you know problems where you use a panel data modeling, and panel data modeling you cannot do in each and every softwares. For instance a excelled excel you know spreadsheet we have a data analysis package, but you cannot do panel data modeling and SPSS we can do most of the modeling but you cannot do the panel data modeling.

So, for panel data modeling you have a different softwares all together. For instance you can use data you can use a views, but now you for instance you have data in a excels, but you can a you know what is actually panel data panel data is nothing, but actually clubbing both cross sectional and time series data.

So, that clubbing you can do in excel. Now, when you analyze the models as to study the impact of you know cross sectional unit and time series unit, while pulling the data; so, all these softwares cannot do that particular you know you know analysis. So, they particular you know standardize software will do the typical analysis and that will be the best for the particular you know requirement.

So, all the excel operation you can do, but ultimately a you go to that particular technique let us say strata or you know Shazam or Eviews, then a simply import the data from the excel spreadsheet and then allow the software or you know give the command to that particular software, do the panel data analysis then by default you will get the econometrics output as per the particular requirement.

So, likewise a you know you should know some of the basic arithmetic calculations there in the excel sheet. So, it can you know give you some kind of you know a you know average, for instance we have already discussed the concept called as you know moving average. And in the excel you have first hand informations and then what will you do? We need to normalize the data as I means to get the best I means best model fit.

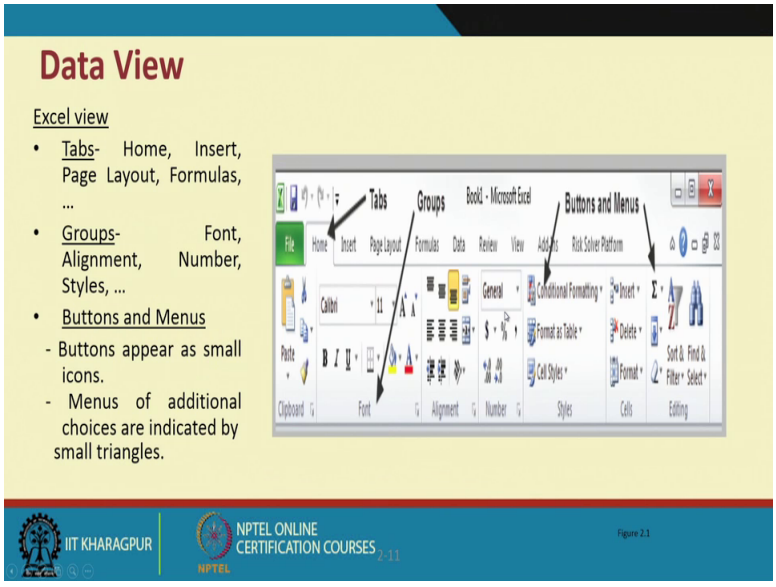
So, what will we do let us say you have 1000 data points and it is the annual data annual data or you know let us say it is a day wise data then what will you do the entire day wise data for you know one thousand data point. So, you can transfer into actually week wise data so; that means, technically every 7 a 7 data point you can do the average and then you can create a new entry where you know you have you know average data with respect to every 7 days.

So that means, technically instead of having 1000 data points, now it will be reduce to something you know average a you know something a less than that and all are in a average format that is the weekly average. And of course, this data size will be smaller one, but ultimately it is the average you know data and normalized data and this will this will have a you know (Refer Time: 27:58) and this will be give you some kind of you know high model accuracy.

And a excel is a kind of you know or spreadsheet is a kind of you know package, this kind of you know operation can be easily done and that will be very useful for you know your modeling requirement and the kind of you know software requirement.

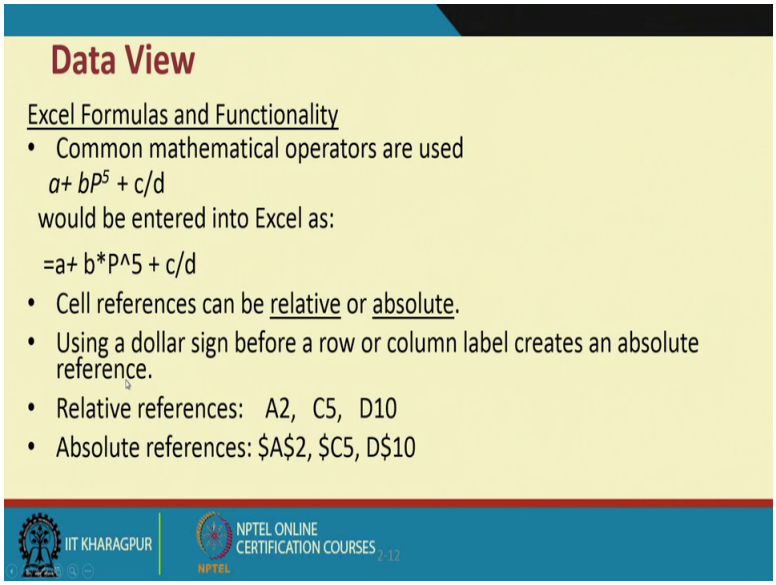
So, that is why you should know sum of the basic arithmetic, we have a plenty of you know arithmetic and basic statistics there you know you know statistical operation mathematical operation in the excel sheet you can do that and then as per the requirement you can you know you know all the details and perform accordingly.

(Refer Slide Time: 28:35)



So, usually excel look will be like this. So, we have a lots of you know options how to open the file how to close the file how to save the file, how to format the data, how to you know visualize the data every options are there, in the next lectures I will showing details you know opening a particular file and how to do the operations. So, sometimes you can do alignment to understand the data to visualize the data more perfectly.

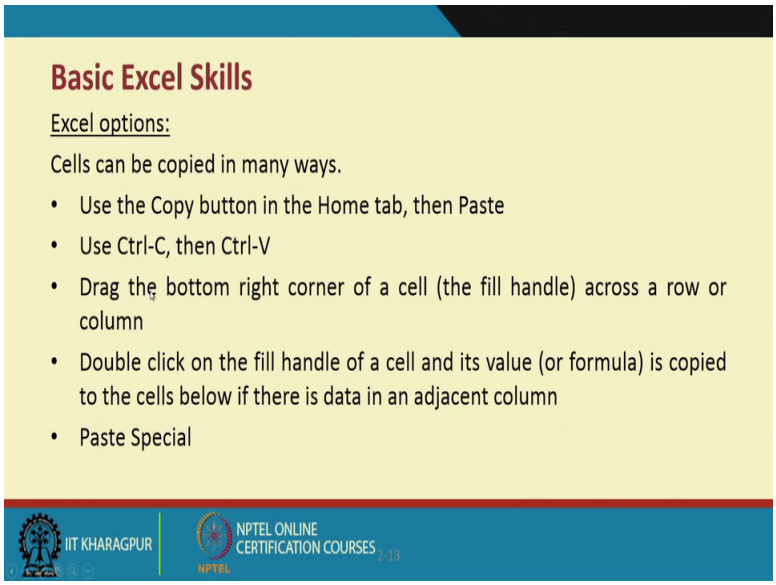
(Refer Slide Time: 29:04)



And then you can also do some kind of you know operations after you know entering the data, you can give the command or you know indicate the relative reference and absolute reference. The relative reference will be simply you know particular column or particular row, then absolute reference can indicate you a particular positions. So, these are all there in the excel sheet and you can do some kind of you know operations if you know the particular you know formula.

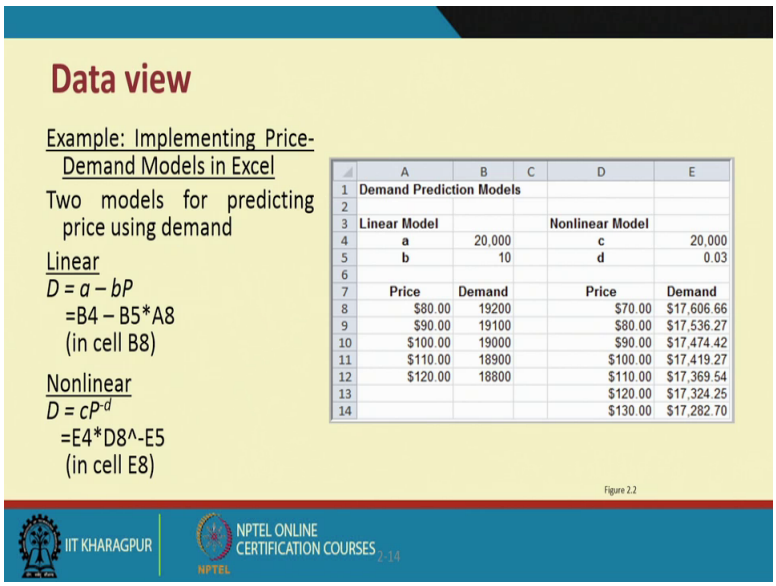
So, with the help of you know data with respect to un variable or few variables. So, you can you know integrate and with the help of a particular formula you can you know transport these two variable into un variable and then data can be operative for the you know use.

(Refer Slide Time: 29:50)



And then you must have a basic skill are to know how to copy, how to paste, how to transpose and a how to you know operate the kind of you know particular cell.

(Refer Slide Time: 30:05)



	A	B	C	D	E
1	Demand Prediction Models				
2					
3	Linear Model			Nonlinear Model	
4	a	20,000		c	20,000
5	b	10		d	0.03
6					
7	Price	Demand		Price	Demand
8	$80.00	19200		$70.00	$17,606.66
9	$90.00	19100		$80.00	$17,536.27
10	$100.00	19000		$90.00	$17,474.42
11	$110.00	18900		$100.00	$17,419.27
12	$120.00	18800		$110.00	$17,369.54
13				$120.00	$17,324.25
14				$130.00	$17,282.70

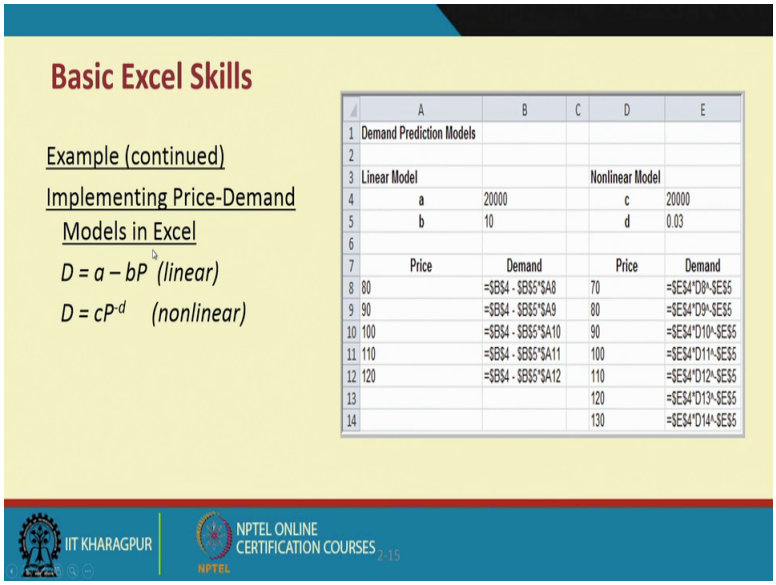
So, these are the particular you know requirement you will do, before you do the kind of you know use of you know spreadsheet and the kind of you know you know excel requirement. So, this is the typical you know box what I mention it is a kind of you know matrix a row wise and you know column wise box, you have to just enter and then operate as per your you know particular requirement.

The operation may operation may be very simple one, may be little bit you know complex, but if you know the details and the kind of you know use. So, everything is a very simple. If you do not know you know kind of you know structure restructures, then it will you know very complex for you.

But once you know the all these operationals kind of you know excel operations and something like that. So, then it is easy actually and for a data analysis and the kind of you know you know do the modeling a in the kind of you know engineering econometrics.

So, entering data, operating the data these are all you know very essential and very mandatory component. If you do not know how to enter the data, how to process the data and forget about the data analysis, in the use of techniques or engineering econometrics as per the particular you know engineering requirement.

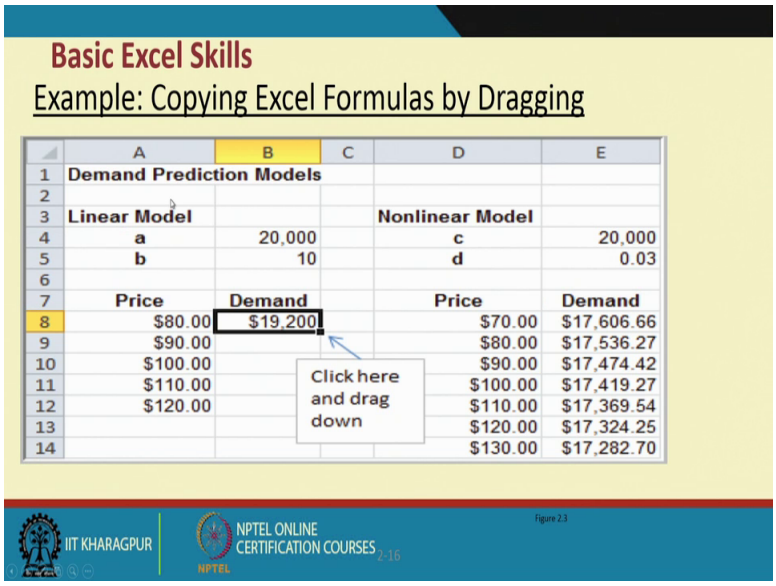
(Refer Slide Time: 31:21)



	A	B	C	D	E
1	Demand Prediction Models				
2					
3	Linear Model			Nonlinear Model	
4	a	20000		c	20000
5	b	10		d	0.03
6					
7	Price	Demand		Price	Demand
8	80	=$B$4 - $B$5*$A8		70	=$E$4*D8^-$E$5
9	90	=$B$4 - $B$5*$A9		80	=$E$4*D9^-$E$5
10	100	=$B$4 - $B$5*$A10		90	=$E$4*D10^-$E$5
11	110	=$B$4 - $B$5*$A11		100	=$E$4*D11^-$E$5
12	120	=$B$4 - $B$5*$A12		110	=$E$4*D12^-$E$5
13				120	=$E$4*D13^-$E$5
14				130	=$E$4*D14^-$E$5

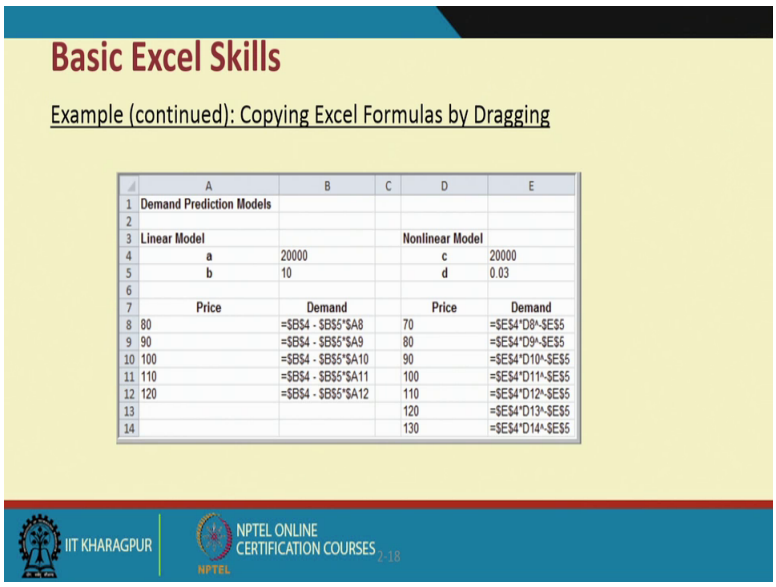
So, similarly a you must have some kind of you know skill to how to you know transport these two variable into one variable, with a different formula or you know different kind of you know structuring.

(Refer Slide Time: 31:34)



Like this which you have already done this here and similarly this is a kind of you know example.

(Refer Slide Time: 31:41)



So; that means, technically what I will like to say that you know, you must have some kind of you know basic knowledge or you know basic skill how to enter the data and how to use the data, how to read the data, how to operate the data, how to structure the data, how to normalize the data and these are all very essential for you know econometric modeling and the kind of you know engineering econometrics.

Because different models has a different kind of you know requirement and sometimes with the help of the kind of you know excel operations you may be in a position to do better modeling and you know better kind of you know you know use of engineering econometrics as per the particular you know engineering requirement.

So, in a in some what I like to say that you know for better modeling or you know better kind of you know analysis as per the particular you know engineering problems requirement. It is not essential that you have to know all kinds of you know engineering tools engineering econometrics tools, you are also want to know the data structure data understanding and the use of you know spreadsheet. So that means, you need two things the use of spreadsheet and the use of softwares.

The use of a you know econometrics software is a different entity all together and the reporting of the data the process of data is also different structure altogether. Of course, a particular software can do all these details, for instance x spaces is a standard package you can actually report you can do the processing you can do the analysis, but some of the advance component where you know x spaces is not handy.

So, that is why a for in that context you have to go for you know advance econometric tools and for that, you must have a basic data or you know actual a actual kind of you know spreadsheet or original spreadsheet through which you can do the processing as per the particular requirement and do the modeling to analyze the problems or you know do the kind of you know forecasting or something like that. So, with this we can conclude here and.

Thank you very much have a nice day.