Lecture 21
Software Implementation of Multiple Linear

Welcome to the lecture so in this lecture we are going to start the new topic and that is about forecasting, so we will study the forecasting in linear regression model.

(Refer Slide Time: 00:24)
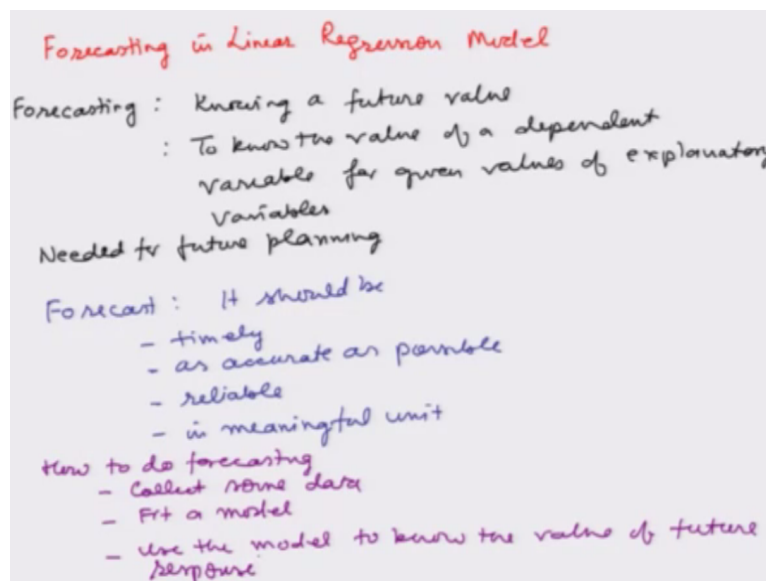
That was in important objective of this course, so the first question arises what is a forecasting? Forecasting is simply knowing a future value. Now there can be various ways to do the forecasting, but in this course we are interested and we are aiming to do the forecasting using the setup of multiple linear regression model and in that contest the forecasting is simply to know the value of dependent variable for some known or given values of explanatory variables or independent variables.

Now this forecasting is very important and this is a very important aspect of any statistical outcome any statistical analysis and an important role of this forecasting is that it is needed for the future planning. Forecasting is very an important aspect of any experimental study and this forecasting is needed for a better planning, the planning can be at the laboratory level, at an institute level, at a city level or even at a country level.

So whenever we are talking of the forecast the forecast should have some important properties, right. For example it should be timely that it should not take very long time to forecast the value. Because at the time it may be possible that the time has already over, and it should be as accurate as possible. Thirdly this should be as far as possible reliable and it should be in some meaningful units.

So that it can have some simple interpretation and that can be understood by each and everyone. So the next issue this how to do forecasting, so the first option is that first collect some data based on that fitter model. For example we have learnt how to fit a model using the linear regression analysis and use this model that you have obtain to know the value of future response, right. So rather I would say that good modeling is the backbone of success of good forecasting.

(**Refer Slide Time: 04:16**)
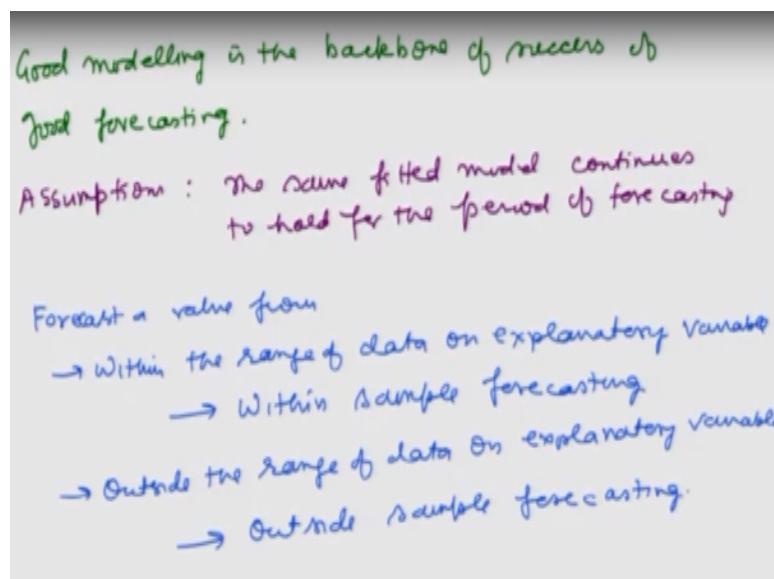
And this is based on certain types of a assumptions, the first assumption are a very important assumption is that the same fitted model continues to hold for the period of forecasting. Now when we are trying to discuss about the forecast then we have two types of forecasting, one is within sample forecasting and another is outside sample forecasting, so when we are trying to forecast a value from within the range of data on explanatory variable.

Then this called as within sample forecasting, and when we are trying to forecast a value from outside the range of data on explanatory variable then this is called as outside sample forecasting right, so let us try to understand this concept of within sample and outside sample

using the linear regression model in a little bit more detail, right. So we are trying to consider here a model like y=x beta + epsilon where is a n cross one vector.

(Refer Slide Time: 06:52)
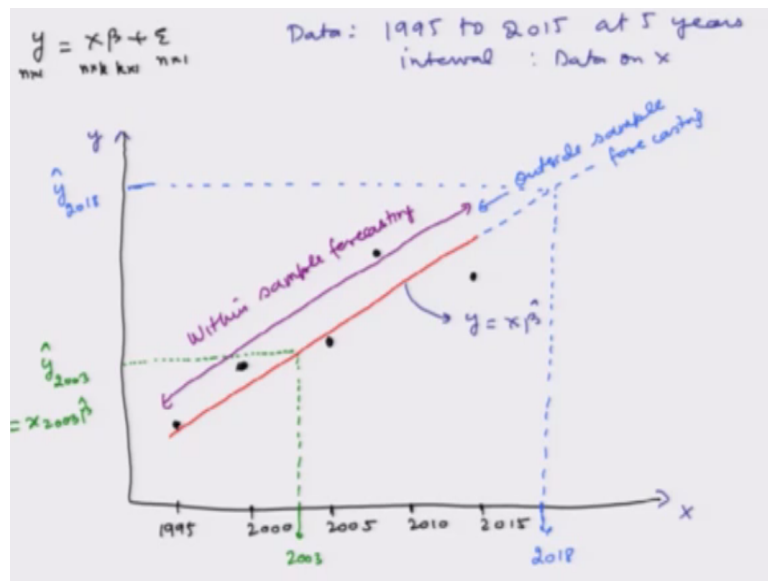
X is n cross k matrix beta is a k cross1 vector and epsilon is a n cross 1 vector, right. So now the first thing what we try to do is the following; that we try to plot this data, now suppose we have got a some data at an interval of five years, say 1995 say here 2005,2010, and 2015, right.

And suppose this beta is given like this, here, here, and say here and say somewhere here. So one can see her that this is showing a sort of linear trend and a line like this one can be fitted to the given set of data. So if you try to understand this line is nothing, but y=x beta hat, okay and then you have to keep in mind that data is available from 1995 to 2015 at five years of interval.

And this is a actually data on here on x, okay, yeah, here this is here x, and this is here y, okay. So now suppose our objective is that we want to know what is the value or what is the forecast in 2003, so now 2003 is lying somewhere here, and then what to we are trying to aim is to know the value of y using this fitted line and we obtain here say here as a y hat at 2003, and this can be obtain as say x 2003 beta hat that we will see it later on.

Okay, so, in this case what are we doing that we are trying to forecast a value of y for the given range of x, the given range of x is from 1995 to 2015, so this is called a within sample

forecasting, so in case if someone is trying to forecast any value from here to here that is between 1995 to 2015 that will be within sample forecasting.

Okay, now we try to consider another aspect another aspect is that suppose we are interested in making a forecast for say 2018. Now in order to know the value of y in 2018, first of all we have to assume that this fitted line y=x beta hat this continues further and then based on that we will try to find out the forecast at x=2018 and somewhere here, this will be your y2018, right.

So whatever we are going to forecast in the range, which is lying outside the range of x variable this is nothing, but outside sample forecasting. So one can also observe here once phenomenon that we have fitted here a model which is only between 1995 and 2015, and this is given by this line in red color, so when we are trying to forecast a value beyond 2015 we do know the future.

And it is possible that beyond 2015 this line may go into some other direction also and in that case we have to be careful and then we have to be watchful and then we have to do the forecasting accordingly, but in our set up we will assume that the same model continue further, okay.

(Refer Slide Time: 12:26)
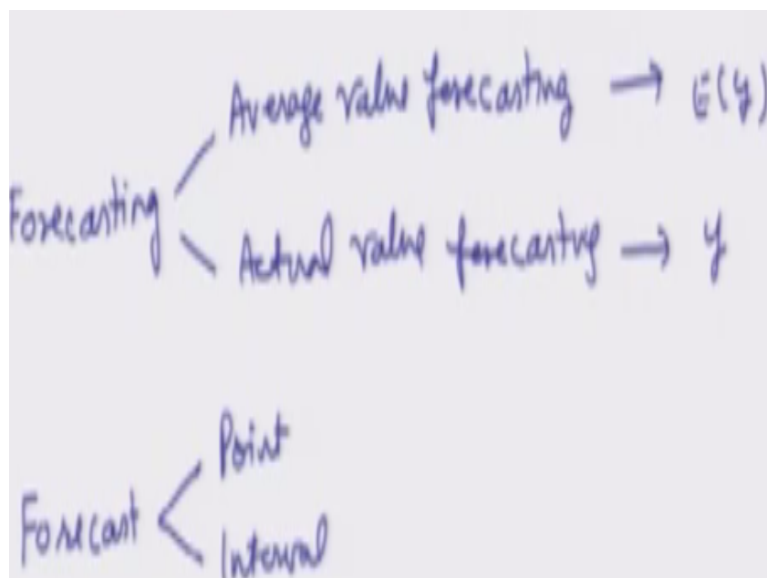
Now we come to another aspect and we try to discuss the average value forecasting and actual value forecasting. So when we are discussing forecasting there can be two types of forecasting in which we are interested. Average value forecasting and another is actual value

forecasting, so under the model y=x beta + epsilon this I can say that when we are interested in average value forecasting that means I need to forecast average value of y that is expected value of y.

And when we are interested in actual value forecasting then we are essentially interested in the value of y. Let me give you a simple example to understand what is the difference between average value forecasting and actual value forecasting, right, suppose there is some manufacture of medicine and the manufactures claims that there company has developed a new medicine which can control the body temperature for say for certain numbers of hours.

Now they want to know that how much is the time in which the new medicine can control the body temperature. Now they come to us and as an statistician we have to understand what really they want, so in order to conduct the experiment we will try to choose say some number of people having body temperature which is higher than the normal value we will give them the medicine and we will try to record the time by wish the medicine is effective in controlling the body temperature.

And then suppose if I have taken this fifty patients then we will have fifty values, now the company is not interested in knowing all those fifty values, they want to know a single value and this will be the value which they are going to use for the marketing of their medicine. For example if we come out to with a sort of forecasting that the medicine can control the body temperature for say ten hours.

And suppose this is higher than the period of body temperature controlled by the earlier medicine then company would like to use this information to market their medicine, so in this case as a statistician we have to understand that the manufacture of the medicine is interested in forecasting an average value. Now there is another aspect this medicine is going to be use by individual patients.

And when this information is passed onto the patient that this medicine can control the body temperature say up to ten hours then this the average value, average value means in some case this may be eight hours, in some case this may be twelve hours, in some case it may be

nine hours, in some case it may be thirteen hours and so on, but the average value of all the time period is close to ten.

Now the patients says well I am not going to use this medicine again and again for say so many times I am going to use it only once, so please let me know what is the actual time in which the body temperature will remain in control by using this medicine. So now as an statistician we have to understand that the patient is not interested in the average value, but he wants to know the actual value.

So in this case we will try to find out the actual value forecasting, okay, so we will try to consider here both the cases, now hen we are trying to make a forecast then forecast can be made as point forecast or as an interval forecast, so this is similar to the estimation problem. For example we have estimated the parameter beta a point estimate as well as in the form of confidence interval.

Similarly on the same lines the forecast can also be made as a point forecast and this can also be made in the form of an interval that is called as prediction interval. So, our objective here is that we are going to consider both the cases within sample prediction, outside sample prediction and under those two cases we will try to consider the average value forecasting and actual value forecasting.

And for each of the case we will try to obtain the point forecast as well as interval forecast and after that we will try to expose these results to some dataset. So we conclude this lecture here and in this lecture I have I try to give you some idea about the basic of a forecasting and now we need to construct a predictor and we need to find out its statistical properties so that we will try to do in the next lecture, till then goodbye.