Modelling and Analytics for Supply Chain Management
Professor Anupam Ghosh
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur
Lecture 48
Forecasting: Trend: Regression and Holt's Method

(Refer Slide Time: 0:22)

Hello and welcome to Modelling and Analytics for Supply Chain Management. We are into week 9; lecture 48 and we are into one of the most important portions, as a beginning point of supply chain, which with we start that is forecasting in supply chain and today's topic is again, the one that has been most widely studied all over the world, that is regression analysis in forecasting. Most widely used, most widely studied.

Now see, what we have, all along we have learned is, all along in this topic over the last three weeks of studying forecasting, we have basically assumed that my demand is constant. Demand is constant that minor aberration is due to individual taste and preferences that are never, never stable and we give with an example that every day we have food, breakfast, lunch, dinner. Do we exactly eat the same quantity of food? Answer is 'No'.

But we have a basic level that this much I will consume. That minor aberrations will be there which we call as 'noise'. So, all along, all the methods of forecasting that we have learned; simple average, weighted average, exponentials smoothing, seasonality, all assume that your sales are more or less constant with the minor noise, that is minor aberrations.

Today we will we see, but what happens in the real world is your sales actually is never constant. And if your sales is constant, if your sales are not growing, then your organization will also not grow and prosper. So, sales have to increase and that is, taken that into consideration, forecasting has to consider and to take how much the sales is increasing over the last, let us say, 5 years, 6 years, 7 yeas, 2 years, 3 years. So, how much the sales is increasing?

So, what is the trend? That is the meaning of the word. What is the trend? Is the sales increasing? Is the sales decreasing? Is the sale is constant? Now, what this is happening, it is a good sign for business if it shows an increasing trend. Now, you understand what is trend. Trend in English word, in statistics, in mathematics, same meaning. Trend. What is the trend? English meaning and this is same. So, what is the trend?

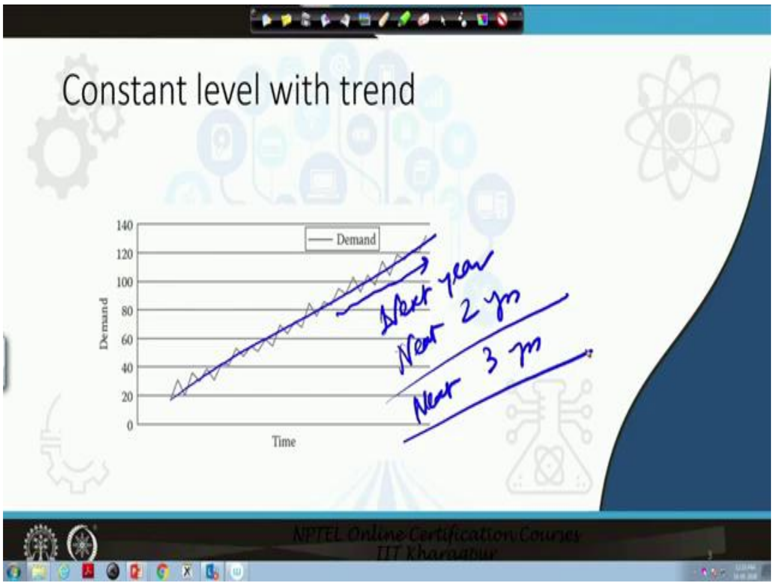
Now, how to find out what is the trend? How much is the increase? We are saying that it is an increasing trend, but how much is the increase? How much is the increase sale? Year on year, every year. How much is the increase? Year on year, that is previous year and this year. How much is the increase? That is found out by something called ‘regression analysis’.

So, regression analysis will tell us how much is the increase and what will be the increase based on this trend. So, regression, all of you have heard about this. All of you have studied about this and I am repeating regression has been the most widely used technique as far as forecasting is concerned, as far as prediction is concerned, most widely used forecasting technique in the world. Even today, though more and more advanced methods are coming in. So, let us talk about what is regression and how to (())(0:03:51) about it.

(Refer Slide Time: 3:57)

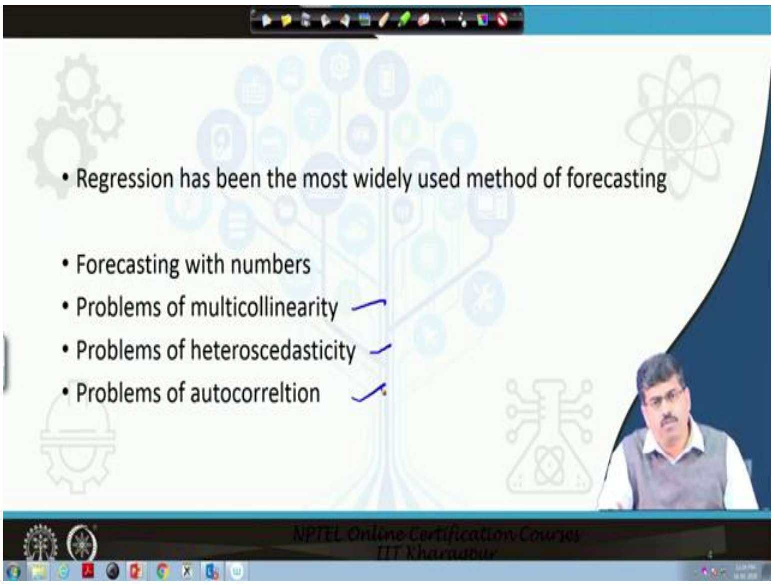
So, in today’s module, we will deal with regression and we will just touch up on Holt’s method, which we will take up in the next week.

(Refer Slide Time: 4:06)



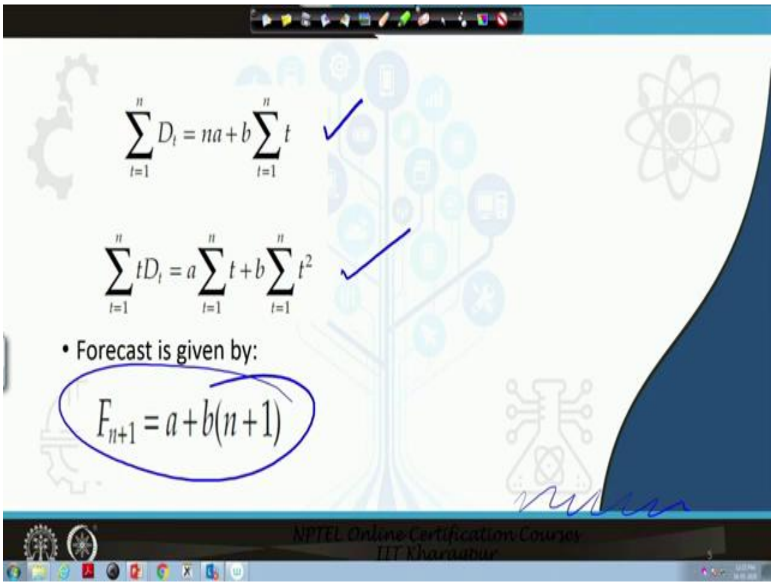
Now see, this is what I was talking about, that my sale is increasing, my sale is showing a trend. That minor zigzags, as I say, consumption of food, daily basis. This is what is, I am, so what am I doing? I am basically predicting how the sales will behave over the next year. How the sales will behave over the next 2 years? Given the same level of increase, how much the sale will behave over the next 3 years? Etc, etc, etc. So, this is what is our agenda today.

(Refer Slide Time: 4:59)



Yeah, regression, as we have are mentioning, regression has been the most widely used method of forecasting. Forecasting with numbers, but there are certain issues with regression. There are some problems that come up. Problems of multicollinearity oh sorry. Problems of multicollinearity, problems of heteroscedasticity and problem of autocorrelation. No rocket science, no high sounding words, nothing complicated, we will come to it and we will discuss this as we progress, okay, forecasting with numbers.

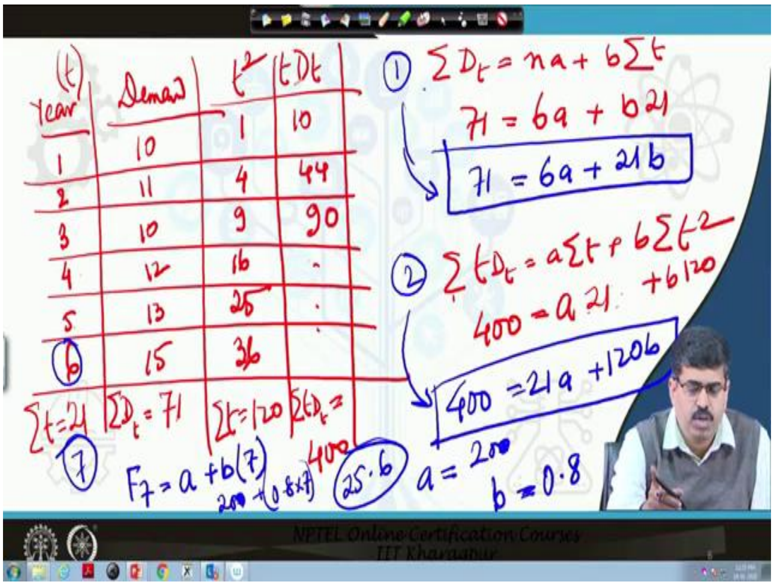
(Refer Slide Time: 5:41)



Now, let us just consider, in case you have an exercise book and pen, pencil and paper in front of you, let us just write down these three formulas. Summation D t is equal to n a plus b summation t. Okay. Let me read it once again, summation D t is n a plus b summation t. Summation t D t is equal to a summation t plus b summation t square. Very simple.

Summation D t is equal to n a plus b summation t. Summation t D t is equal to a summation t plus b summation t square. And forecasting is given by F n plus 1, that is period, next period, forecasting F n plus 1 is equal to a plus b n plus 1.So, let us see, let us do a problem and try to solve it. Let us see what happens and then we will discuss so many issues in forecasting.

(Refer Slide Time: 6:31)



Let us say, this month or year. I hope you are able to see year. 1, 2, 3, 4, 5, 6. Okay. What is the demand? 10, 11, 10, 12, 13, 15. So, you can clearly see an increasing trend. Clear, what is our formula? Let us say t square; 1, 4, 9, 16, 25, 36. Since, summation D t is equal to n a plus b summation t. Summation D t is equal to n a plus b summation t.

Let us put the summation D t. what is the summation? 60, 68, 70, 71. Okay. 50, sorry 60, 70, 71. So, what is my first formula? Summation D t equals n a plus b summation t. For this problem, we will have some of these issues because I am regularly changing between the mouse and the pen. What is summation t? Summation t is 6 plus 5, 11, 15, 18, 20, 21. Agreed? So, summation D t is equal to n a plus b summation t.

Now, what is your D t? Demand is 71, is equal to, n is 6, 6 time periods. 6 a plus b into summation t is 21. So, in other way around, it is 71 is equal to 6 a plus 21 b, this is equation 1. Let us go to the earlier one, summation t D t, which we have just calculated, is equal to a summation t plus b summation t square. t D t is equal to a summation t plus b summation, so, summation, we will have to introduce something here, summation t D t, so summation t D t 10, 44, 90, 12 into 16, and 13 into, whatever we get. In this way, you get something, summation t D t.

So, let us put an arbitrary number. Let us put an arbitrary number of, let us say, 400. t D t is equal to 400. Summation t is t plus b summation t square. This, b summation, so summation t D t is equal to a summation t plus b summation t square. So, summation t Dt is equal to, this one , we have assumed 400 and a summation t square, assume your summation t square is, let us say, 120, so a 120 plus b into, a summation t, sorry, a 21, this will be 21, plus b summation t square, is 120. So, your next equation now becomes 400 is equal to 21 a plus 120 b.

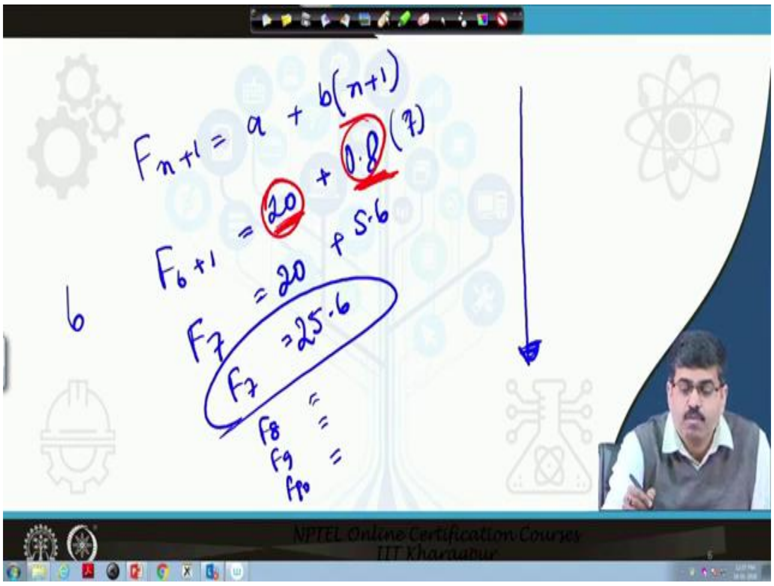
So now, what you do is this blue coloured equations, this blue coloured equations you solve and what you will get is, you will get an, sorry, you will get an a. These blue coloured equations, you solve. What you will get? You will get a value of a and you will get a value of b. Now, assume your a value is 200 and b value is let us say, 0.8. Now, these two other values that are required. So, you will forecast, so based on these two equations, the first equation, equation 1 and equation 2, we get these corresponding values, and we solve them and we get what is a and what is b?

Now, based on these, we will forecast for what? We will forecast for period 7. If this period 7, we will forecast based on what? Based on the formula that we are given, F n plus 1 is equal to a into b n plus 1. What is the (for) F n plus 1 is equal to a into b into n plus 1. So, I think we will do it here only because if you go to the next slide there will be problems, so, F 7 is equal to a plus b 7.

What was the formula? F n plus 1, that is F, n is 6, plus 1, 7 is equal to a plus b into n plus 1. a, we have just got, assumed, we have 200. You have got a, plus b was 0.8 into 7, that is 5.6. So, 5.6 plus 200 is 205.6 is your forecast for the next year. Now, we have got this because you have arbitrally assumed a to be 200. Ideally, a will be something like your 20 or something like that. So, you can easily assume, a is 20. So, it is that your, this will be 25.6, let us assume this as 25.6.

So, F7 is equals to a plus b7 a plus b into n plus 1. So, this is the formula. So, very simple one. Solve these two equations and just put F n plus 1. This has become a bit hazy I will say, but if you can rub it, so we will take a new slide.

(Refer Slide Time: 15:07)

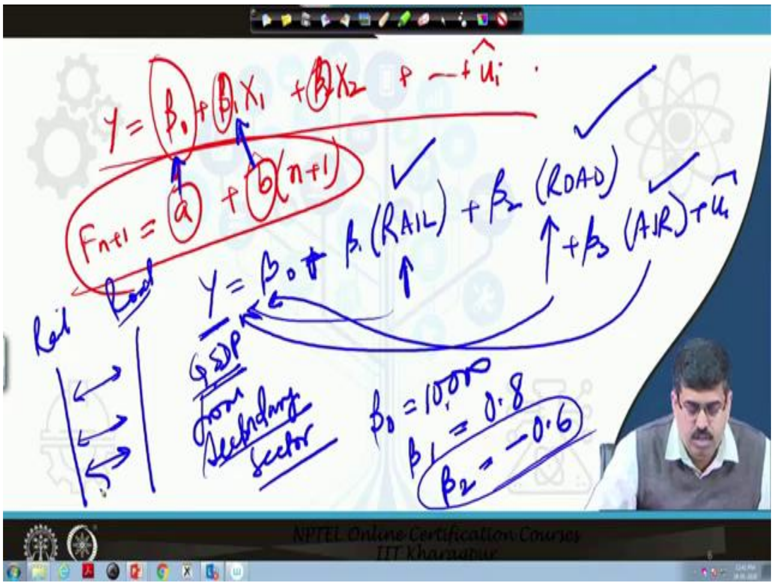


What we do is we say F n plus 1 is equal to a plus b into n plus 1. We have data up to six time periods, so, F, n was 6, 6 plus 1, that is you are forecasting for period 7, is equal to a, we have assumed 200 then we said let us assume 20 because the data was that manner only. b was 0.8; into n plus 1 is 7. So, 20 plus 5.6. So, F7 is 25.6. This is a simple one.

Now, what we normally do in regression is, if we are doing it in excel, we just keep on dragging the mouse, dragging the curser rather. We just keep on dragging the curser and what will happen? We will get F7, F8, F9, F10. In that way, we will keep on getting some values or the other for as many years as you want. Why? Why and how we can do that?

Why and how we can do that is because everywhere we are assuming this level is constant. This a is constant. Everywhere we are assuming that this 0.8, this trend is constant. So, level and trend, both are constant. We are assuming, okay, in this regression. Now, Holt's method takes care of adjusting for the level and the trend both. Here, we assume that the level, that is 20 fixed consumption and the trend of consumption, both are fixed over the time period. Holt's method we will say no, we can adjust them. So, this is it. This is the basic of regression. Now, I want to spend some time on something else.

(Refer Slide Time: 17:10)



See, regression has some issues with it. What are the issues? See sometimes you will get regression, what is, regression is a form of what? Y is equal to beta naught plus beta 1 X1 plus beta 2 X2 plus up to Ui predictive hat. This is your regression equation. Now, this beta naught was basically your a and all these are your that b. This is what we mean. This a is corresponding regression equation. This a is beta 1 and b is beta 1, this way. So, this is what we look at in the regression.

Now, logically, let us consider an equation wherein we are saying that Y is a function of beta naught plus beta 1 RAIL plus beta 2 ROAD plus beta 3 AIR plus Ui hat, this Ui hat is the error term. Just keep it at the side for some time being. Now, what we are saying is that the development, let us say Y is your GSDP; Gross State Domestic Product. The Gross State Domestic Product of a state depends on how much infrastructure you have in the state. Gross Domestic Product of a state depends on how much infrastructure you have in that state.

Now, what is happening? So, your Gross Domestic Product if a function of how much railway tracks you have? How many kilometres of good roads you have? And how many and how much is the air traffic in your sate? Logically, GSDP from, let us make it more specific, GSDP from secondary sector. Let us make it more specific. That is industry, GSDP from industry.

Now, logically, if you increase railway tracks, your GSDP should increase, State Domestic Product, output should increase. If you increase roadways, more number of good quality, international standard roads, your output from the state should increase. If we have more air travel, more number of air landing facilities, again your production facilities will increase.

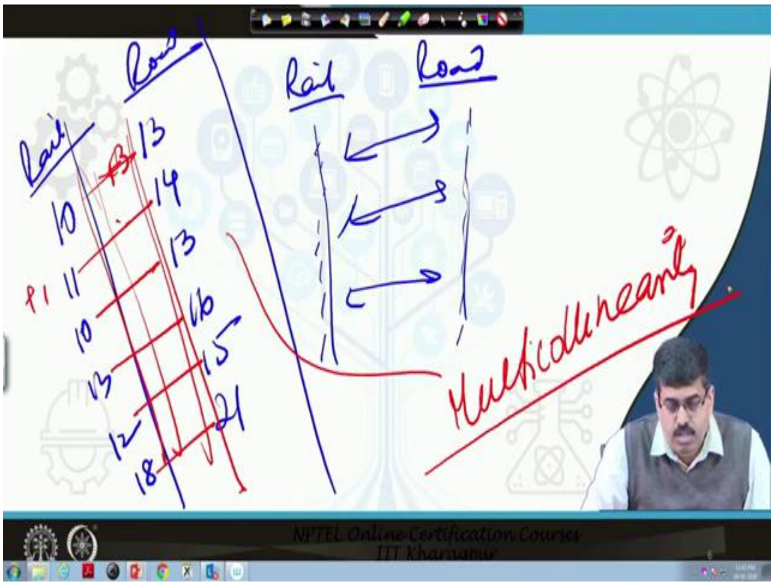
So, logically, for this equation, let us say beta 0 is 10,000 that means even if nothing is there, still 10,000 small-scale units will be there. Beta 1 let us say is 0.8 that means that if you increase railway track kilometres by one kilometre, your GSDP increases by 0.8 units of output.

But, beta 2 is coming out as minus 0.6, which essentially means that if you reduce roadways, your output will increase. Gross Domestic Product will increase. But, is that possible? Is that logical? You are saying that instead of increasing your infrastructure, you reduce infrastructure and show your output you will increase. This is what you get if you out data on infrastructure, rail, road, airways in the excel sheet and do a regression.

Some betas will come minus but logic says that the betas or the b should not come minus for such a equation. Some equation, it may come, but with this equation, it should not come. What is happening you know, this is happening because of something called 'multicollinearity', which we, just before sometime we had say in the title, just immediately after the title slide what we want to cover. This is what we call as multicollinearity.

Multicollinearity means the data behaves in such a manner that one data set, the rail data set, those numbers and the road data set those numbers, these are having a correlation, these numbers are having a correlation. That is what is called as multicollinearity.

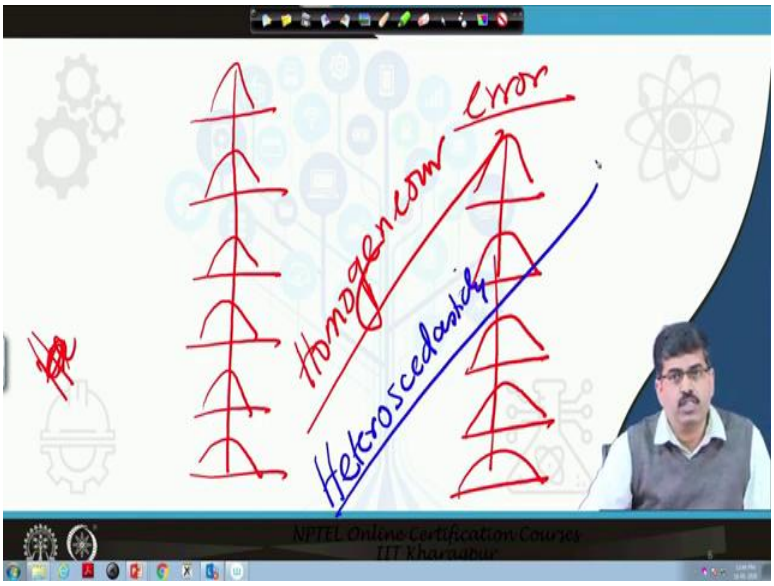
(Refer Slide Time: 21:41)



So, you have rail, you have road, and these are your data points. Now, these data points are correlative. See, mathematics do not understand relationship among variables. Mathematics only understand numbers. So, here is a number series; 10, 11, 10, 13, 12, 18. Here is another, that is rail. Here is another called 13, 14, 13, 16, 15, 21. What is happening? Let me take the red colour, this thing. 10 plus 3. So, all are plus 1 this way. So, 10-13, 11-14, 10-13, 13-16, so both the data series, both the data sets are behaving in the same manner.

So, your numbers are saying that they are correlated. Logic is saying that they are not correlated but your number are saying. So, this is the problem of multicollinearity. They are collinear. That is there is a high degree of correlation among them.

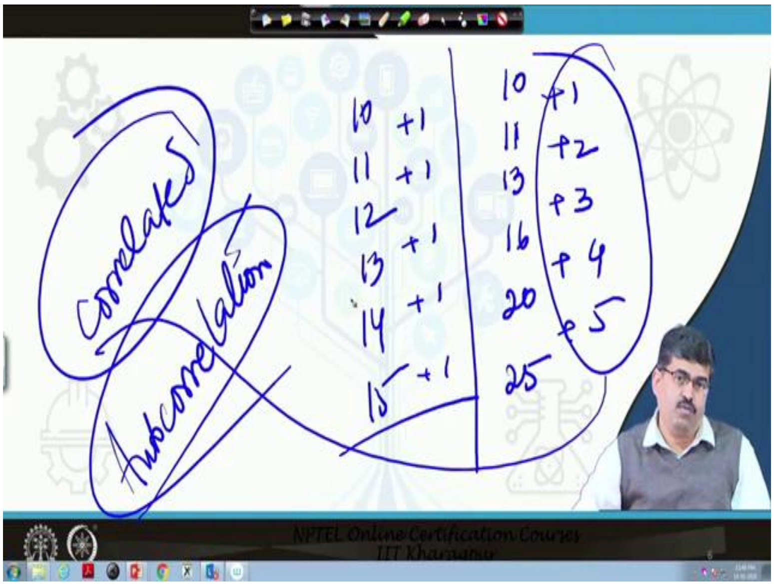
(Refer Slide Time: 23:20)



Now, the second problem is, after you solve the equation, there was something called the error terms. Now, that error terms should be like this that is they should have equal variance. The error term should have equal variance. Error terms should be like this that is they should have equal variance. Then the model is fine. If it does not, that is they should have equal variance means that is they are homogeneous.

When they do not have an equal variance, then we have something called heterogeneity. When it is not homogeneous, we have something called heterogeneity and that, if it is there in the data, is called as ‘heteroscedasticity’. So, the error variances are not uniform.

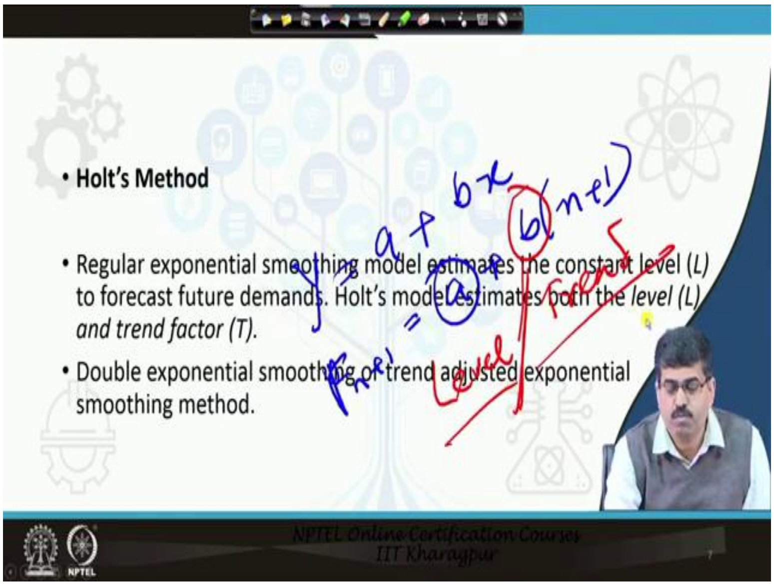
(Refer Slide Time: 24:44)

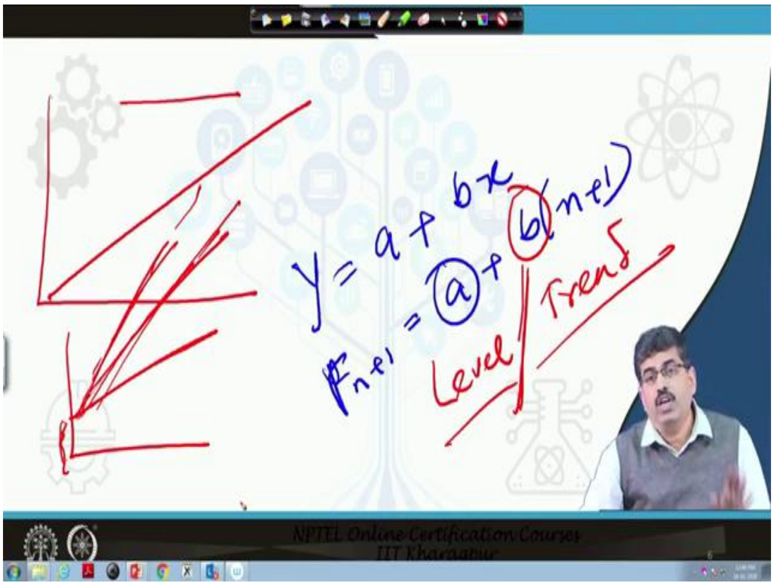


The third point that we have is, third point that we have is, for any data, let us say 10, 11, 12, 13, 14, 15, see, plus 1, plus 1, plus 1. Done. 10, 11, 13, 16. What is happening? Plus 1, plus 2, plus 3, plus 4, plus 5. Correlated? So, 1, 2, 3, 4, 5, some correlation is coming. This is called as autocorrelation.

So, what we want to see is do not take a regression result by its face value. Look at the numbers, look at the logic and see whether the numbers are logical, whether the numbers behave in this manner; multicollinear, heteroscedastic, autocorrelated, whether the number behaves in this manner? If the numbers behave in this manner, then there are ways by which you can correct multicollinearity, heteroscedasticity and autocorrelation. There are ways by which you can do that. But otherwise, if the numbers are behaving like this, you will have to be a bit careful.

(Refer Slide Time: 26:20)



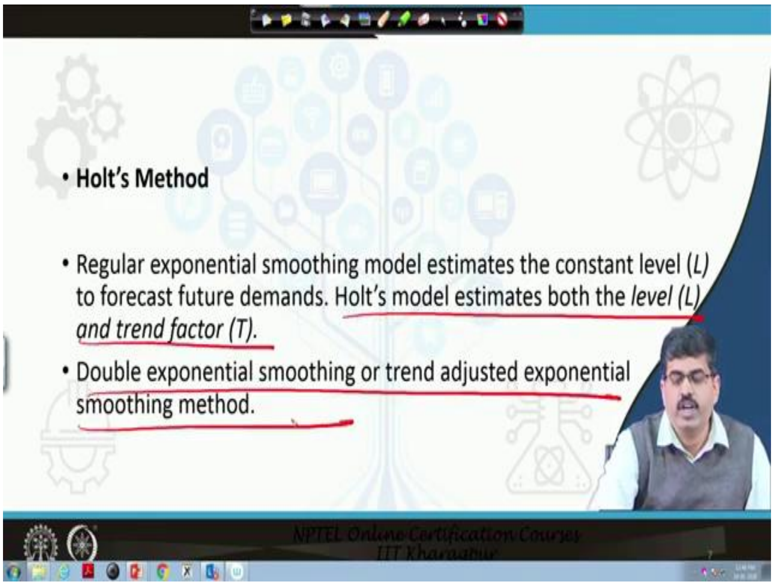


Now, this is that and as we say this Y is equal to a plus b x or F n plus 1 is equal to a plus b into n plus 1. Now, what we said it is that this is called as the 'level' and this is called as the 'trend'. Now, Holt's method, what we do is, it estimates I am purposefully not deleting the blue ink just to keep it there. It estimates the constant level to forecast future demands. It estimates the level and the trend. Now, let us go back. Let us make the slide clear.

Here, in regression, we assume that the level and trend remains same. So, based on the same level and trend, we were forecasting. But, will this level remain fixed? Automatic consumption may not be. Will this same trend remain or will it become this or will it become very high consumption?

So, Holt's model assumes that both the level and the trend will change and in such a situation we use a method called Holt's method.

(Refer Slide Time: 27:57)



That is why Holt's model estimates both the level and the trend. Double exponential smoothing or trend adjusted exponential smoothing. This is also it is called. Now, this, we will pick when we start off with the next week's session. Thank you.