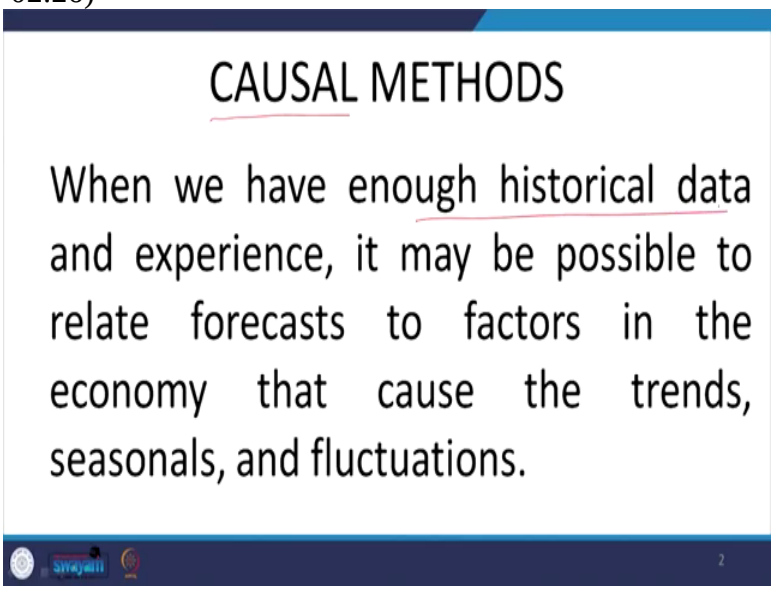
Production and Operation Management
Professor: Rajat Agarwal
Department of Management Studies
Indian Institute of Technology, Roorkee
Lecture 15
Causal or Explanatory Methods

Welcome friends, now this is the fifteenth session of our course on production operations management. In last few sections, we were discussing about various types of time series analysis which are also known as extrapolative methods. We also discussed some examples of different types of smoothing methods in our last session.

And we also discussed that when we are taking a particular time series method, determination of forecasting error is one very important aspect which determines the suitability of a particular method and for that purpose we have different measures of forecasting errors. We discussed average error, mean absolute deviation, mean squared error and mean absolute percentage error. And then by combining different types of forecasting errors we also discussed the concept of tracking signal.

And we discussed that tracking signal need to be random for a large sample and based on those different measures, we can determine the suitability of a model we can also determine suitability of process parameters for that particular model. But many a times, when we have large historical data with us, and with our own experience, we can also understand that there are few factors which are affecting our demand. Which are affecting the demand of products or services and now in such type of situations, we can go for a more accurate method of forecasting. Which is known as Causal or Explanatory Method and that is the topic of our today's discussion.

(Refer Slide Time: 02:26)

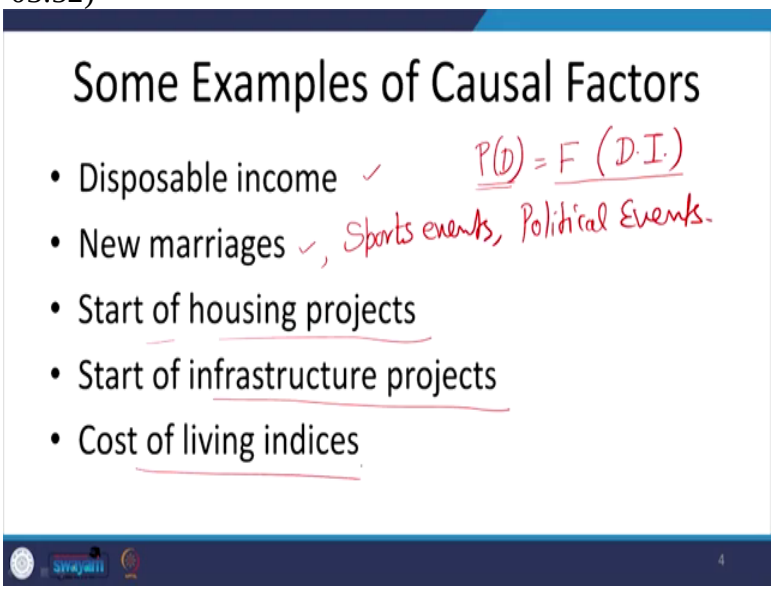


Now this causal or explanatory method, as the name indicates cause effect method. There is a cause and there is going to be some effect of that cause. And when we have the large amount of historical data and our own experience, because we are continuously dealing with that type of product, that type of service.

So, we also understand how this system is behaving. So, with that experience and with the help of data, we may develop forecast with the respect of those factors which are going to affect the demand of the products or service in our consideration and here it is quite possible that you develop a type of cause effect relationship which we say is the mathematical equation. and that mathematical equation will give us more accurate forecast than the time series analysis.

Then you see this determination of relationship is the hallmark of this causal method. What type of relationship you develop, the more accurate relationship you have more accurate forecast you will have for your causal analysis.

(Refer Slide Time: 03:52)



Some of the examples which can affect the demand of your products can be the disposable income. How the disposable income of a particular society is changing, and accordingly if you have some kind of relationship of your products demand, then based on this function of disposable income, if you can develop a relationship, you will find that demand of your product is a function of disposable income and I need to determine this relationship. Then I will know that in year 2025, probably this will be the disposable income of India and if I substitute that disposable income accordingly I will know what will be the demand of my products in 2025.

So, from some other sources from some other independent sources, I know that how these factors are going to change incoming time and based on my that knowledge I will substitute those values in this relationship and that will give me the demand of my products or services. Similarly, there can be some more examples like new marriages.

So, how many new marriages are going to happen in the coming winter season and depending upon that estimate I can think of that this much demand of gold will be there in the market. Because marriages are affecting the demand of gold, marriages may affect the demand of gems and jewelry, marriages may affect the demand of designer crops.

So, there are large number of products for which demand is dependent on marriages. So, marriages can be one of the factors which can affect the demand of certain products. Even in

some of the cases marriages may affect the demand of automobiles also. So, or you can have let us say some sports events, you can think of some political events.

Elections were held in 2019 and because of general election demand of certain products increased. Now if you develop a relationship that the election and some demand of a product let us say the demand of helicopters, is directly proportional to the political events.

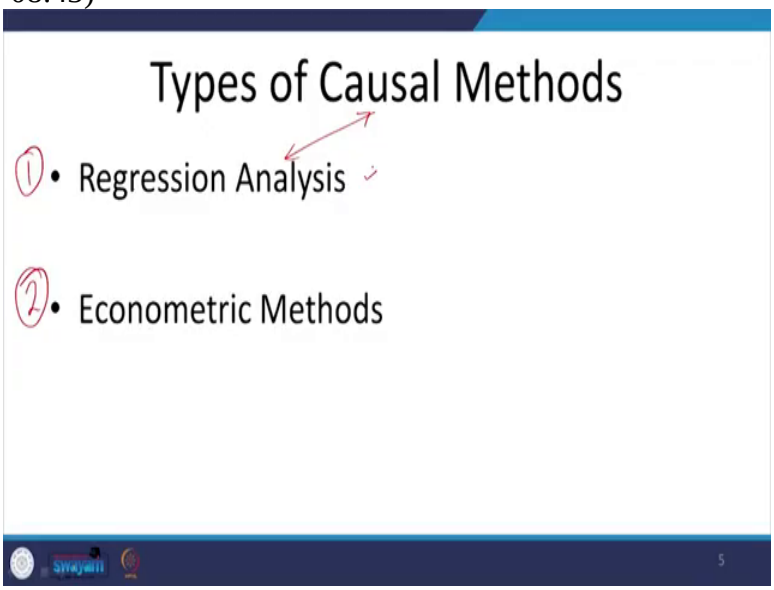
So, if it is a state assembly, then what will be the demand of helicopters, and if it is a general election what will be the demand of helicopters, so now if you have that type of relationship between the demand of helicopters and elections and now in 2024 when the elections are going to happen, using this relationship you can determine that how many helicopters will be required by different political parties for their political campaigns.

So, that is a kind of relationship we need to develop. Similarly, start of housing projects, start of infrastructure projects, cost of living indices. So, there can be many more such factors like how much you are going to spend on advertisements. So, if you are a FMCG company, so maybe the sales of your company is dependent on your advertisement expenses. So, if you develop a proper relationship between advertisement expenses and the sales of your company, then that relationship will help you that if you increase the advertisement expenses by 10 percent, how much your sales is going to increase. So, that type of relationships we develop in our causal analysis.

So, this is just a exemplary list for your understanding. But it is not the complete, there can be as per the situation, you can think of as many causal factors which affect the demand of your product. But the important point for us is that, we should be able to get the values of these causal factors from some independent source.

If the value from some reliable independent sources are not available, even if we are able to develop a mathematical relationship, we will not be able to forecast properly. So, it is very important that you take those causal factors for which the values are available to you from some reliable sources for the coming periods. So, that is the most important point in selecting this causal factors, that from where you are getting the data for those causal factors for future periods.

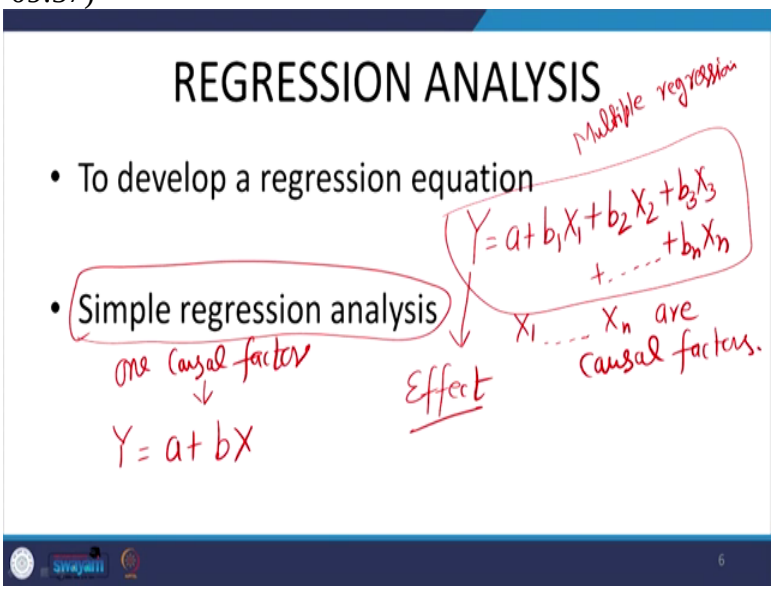
(Refer Slide Time: 08:43)



Now in the causal factors, in the causal methods, we have two important type of methods, one is regression analysis. Many a times in our production operation management we normally interchangeably use regression analysis with causal methods. Most of us feel that regression and causal methods are interchangeably, but it is important to understand that regression analysis is one type of causal method.

The other type of causal method is econometric methods. So these are the two most popular type of causal methods. One is regression analysis and another is econometric method. In this particular session, we will be focusing more on regression method and we will also have a bit of introduction about econometric. But our major focus will be on regression analysis.

(Refer Slide Time: 09:57)



Now what is regression analysis? Now in this regression analysis we develop a regression equation. We develop a regression equation and this regression equation is like this way. Y equals to a plus b1 X1 plus b2 X2 plus b3 X3 upto bn Xn. Where X1 to Xn are causal factors and Y is the effect. So, you have your effect in our case, because we are interested in knowing the forecast.

So, for us effect is in terms of demand of the product for a future time. So that is the effect, now this effect is depending on various causal factors and these causal factors are X1,X2 up to Xn. So, we develop this type of general regression equation. So, this is I am giving you the most general form of a regression equation.

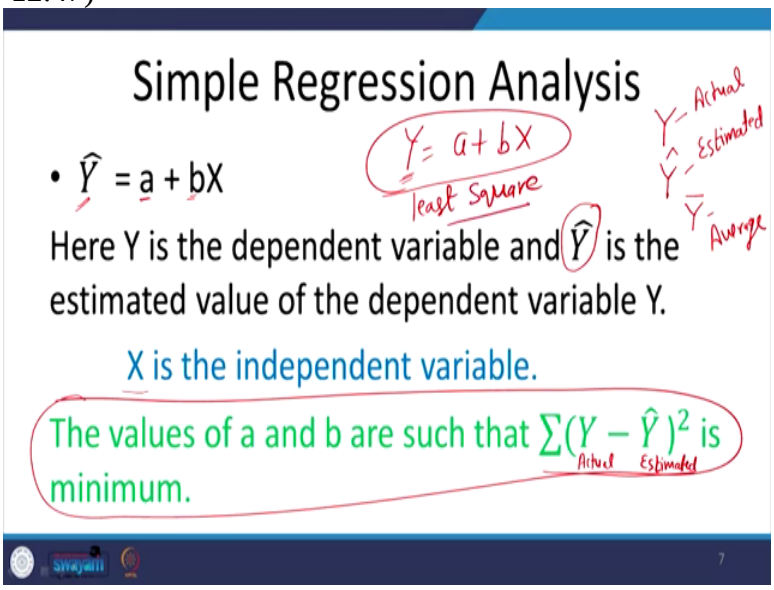
But in our course right now we are going to discuss only the simple regression analysis. In this simple regression analysis we take only one independent factor. We take only one causal factory. So, therefore the shape of our equation will be like Y equals to a plus bX.

This will be the shape of our regression equation because in simple regression equation simple regression analysis there is only single causal factor. So, this is Y equals to a plus bX. When you have more than one causal factor, it becomes a case of multiple regression equation. So, this is this which I wrote earlier this is the case of multiple regression.

The meaning of simple and multiple is coming because of how many factors are there in your independent side. How many causal factors are there, that number of causal factors are same

whether it is simple regression analysis or multiple regression analysis. So, we are going to discuss just simple regression analysis. Here we have the equation of Y equals to a plus bX.

(Refer Slide Time: 12:47)



Now in the simple regression analysis as we just discussed, the equation of regression line is Y cap equal to a plus bX. Now the sign Y cap is very important, because the equation is of the type Y equals to a plus bX. But for different values of X, we will estimate the value of Y. So, Y cap when I write, it is the estimated value of Y for a particular value of X.

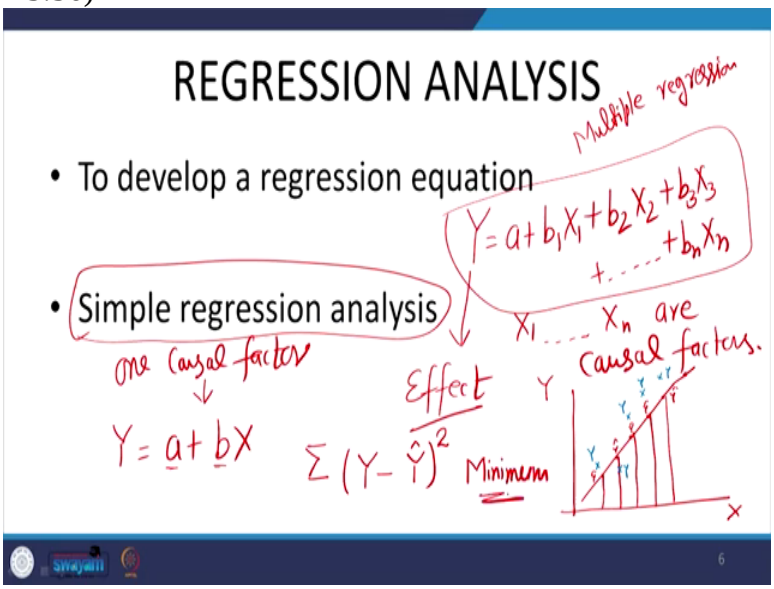
There is we will see in the coming slides, that what is the meaning of estimated value and the actual value of Y. So, this is the actual value of Y equals to a plus bX. But with the help of future periods, when we put the value of X what value of Y is coming by substituting the value of X that is the estimated value of Y.

So, there will be three types of symbols we will be using. One is Y, another is Y cap, another is Y bar. So, Y is the actual value of Y, this is the estimated value and this is the average value. So, different types of symbols you will see, in our discussions. So Y, Y cap and Y bar. These are the three types of symbols we will be using for giving our analysis with respect to our effect side and similarly you have this X as the independent variable and a and b are the coefficients of this equation, a and b are the coefficients of this equation and the way we try to find out the values of a and b, so that this particular term, sigma of Y minus Y estimated whole square is minimum.

Sigma of Y minus Y estimated whole squared is minimum, that is how we want to determine the values of a and b and because of this particular statement, because of this particular statement that the deviation of actual values of Y and the estimated values of Y should be minimum. The values of Y minus Y bar r are the difference of actual and estimated.

So, this difference of actual and estimated value should be minimum for that purpose, this line is also known as line of least square. This is also known as line of least square. This is also known as line of least square.

(Refer Slide Time: 15:56)



So, meaning is that, when we are developing this equation, it is like this way. This is the equation of a line, we understand Y equals to a plus bX. So, this is Y and this is X. So, this with respect to any value of X if I put that value of X in this equation, you will get a value of Y on these any of these points.

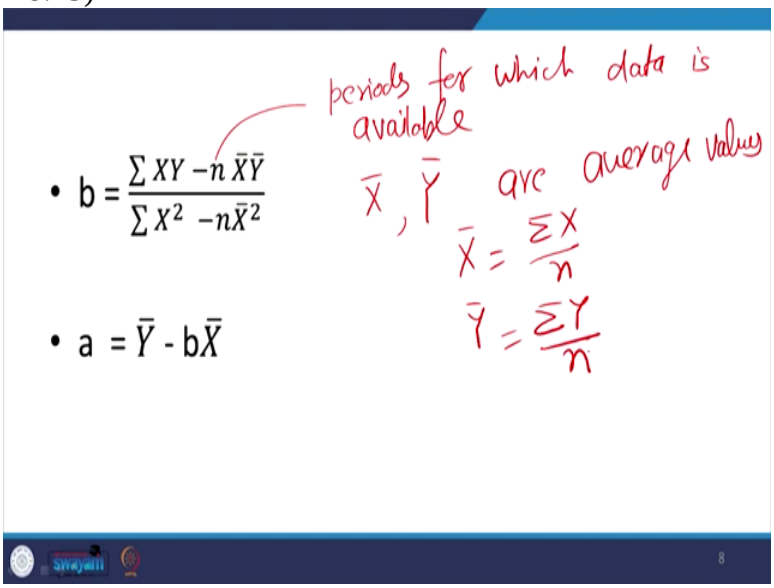
So, with corresponding to any value of X, these values of Y are there. But the actual values of Y maybe the actual values of Y for this particular value of X maybe here. For this particular value of X it may be here, for here it may be here. Then it may be here. Then it may be here. So, these are the actual values of Y. So, these are Y, Y, Y, Y, Y and values which are here these are Y cap the values which are on this straight line these are Y cap.

So, the idea of determining this a and b is that the difference between Y and Y cap should be as minimum as possible. So, that your this line should be determining as close to possible to the

actual values as much we can do. So, therefore the difference of Y minus Y cap and then because some of the differences are in positive and some of the differences are in negative.

So, we take their square, so that all terms becomes positive and then we take the sigma. So, we want that this particular value should be minimum. It means our line is giving maximum closeness to the actual value and this is how we determine the values of a and b. So, therefore this is also known as line of least square.

(Refer Slide Time: 18:25)



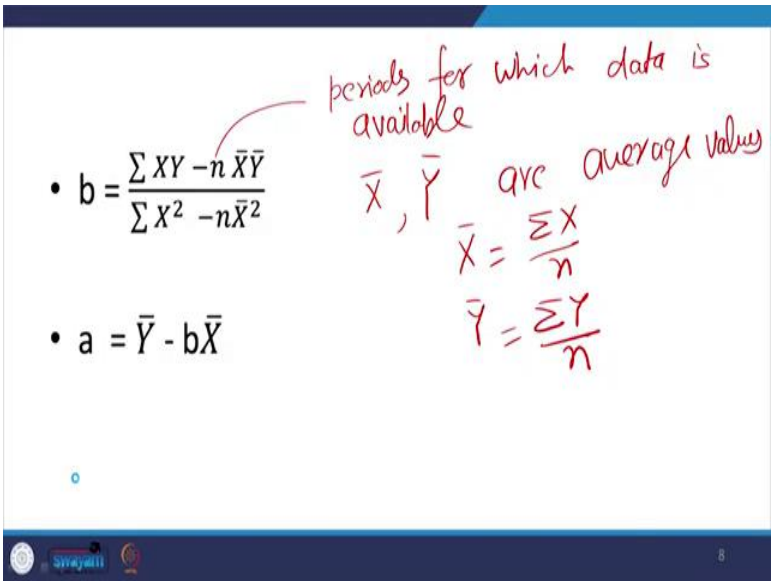
And for this purpose, the formula to calculate the values of b and a are sigma XY minus n, n is the number of observations or the periods for which data is available. As I told you, these bar values are average values. So, these bar values are determined as sigma X upon n, Y bar is determined as sigma Y upon n. So you have these values, and on the basis of this, we will calculate the values of b and a.

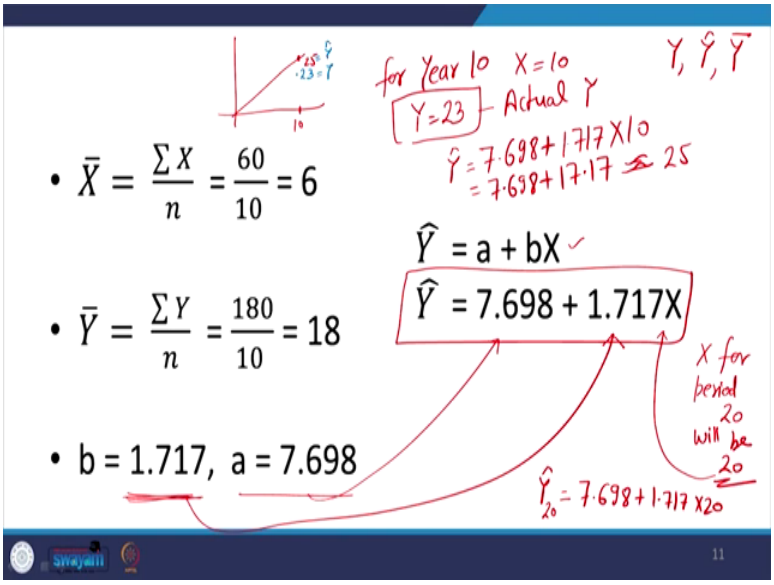
(Refer Slide Time: 19:18)

Year	Store Sales (Y) (in 10 thousand dollars)	Industry Sales X (Million dollars)
1	9	2
2	14	4
3	16	4
4	18	6
5	19	6
6	20	6
7	18	7
8	21	7
9	22	8
10	23	10

Year	Store Sales (Y) (in 10 thousand dollars)	Industry Sales X (Million dollars)	XY	X^2	Y^2
1	9	2	18	4	81
2	14	4	56	16	196
3	16	4	64	16	256
4	18	6	108	36	324
5	19	6	114	36	361
6	20	6	120	36	400
7	18	7	126	49	324
8	21	7	147	49	441
9	22	8	176	64	484
10	23	10	230	100	529
SUM	180 $= \Sigma Y$	60 $= \Sigma X$	1159 $= \Sigma XY$	406 $= \Sigma X^2$	3396 $= \Sigma Y^2$

calculated data





Now with the help of this tabular data, we will try to see how do we use this formula. Now we have some data for past 10 years. Now here we have two types of data, one is there is a store, and the sales data of that store is given in 10000 dollars and the industry data is available in million dollars. So, industry data is our independent variable that is X, you see and the store data is our dependent variable that is why.

So, now in previous 10 years, this industry sale is 2 million dollar. So, my store sale was 90000 dollars and so on for different periods, for last 10 years with respect to industry sales, how the store sales have varied that data is available. Now to apply our regression analysis on this data

what we are doing, that we are going to create a table. Where we have 3 additional columns, so these are given data, this much is the given data to us and this is calculated data.

So, I have added three more columns, in this table one is for XY, another is for X square and another is for Y square. So simple multiplication of column 2 and column 3, that is 9 into 2 18, 14 into 4 56 and so on have given me the values of this fourth column. Then in the fifth column the X squared values.

So, these are the square of third column values. 2 square 4, 4 square 16, 4 square 16, 6 square 36 and so on and in the last column we have the values which are the square of Y column that is the second column. 9 square 81,14 square 196 and so on and the last row of this table gives me the sigma value. So, this is sigma Y, this is sigma X this is sigma XY, this is sigma X square, this is sigma Y square. So, we have this table ready for our calculation. Now using these values of sigma X and sigma Y first we can determine the values of X bar and Y bar.

So, X bar is coming sigma X upon n, that is 60 by 10, n is for 10 periods we have the data. So, here this gives me the value of n equals to 10. So, now you have X bar equals to 60 by 10, that is 6 and Y bar sigma Y upon n, 180 upon 10 that is 18 and when I put these values of sigma X, X bar, Y bar, sigma XY, n, X square, Y square in the formula of calculation of of b and a, I will get the values of b is coming as 1.717. This is coming 1.717 and when I put the values of a X bar and Y bar in the formula of a, I get the value of a as 7.698.

The form of regression equation is Y bar, Y cap equals to a plus bX. So, now the desired regression equation in this case is Y cap equals to, the value of a goes here, value of b goes here. So, this becomes Y cap equals to 7.698 plus 1.717X. Now if I know the value of X, if I know the value of X for any future periods, I have the value of X for period 10 as 10.

Value of X as 10, for period 10. Now if I say that the value of X for period X for period 20 will be 20, let us say. So, now I can put this value of 20 here, and correspondingly for year twentieth, I can estimate the value of Y cap 698 plus 1.717 into 20 and this calculation will give me the forecast for twentieth year.

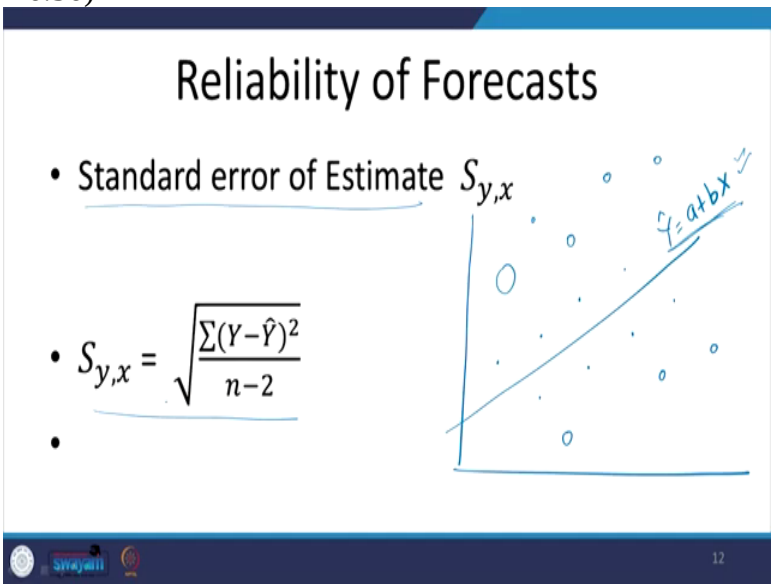
So, accordingly I can get the values of any particular year by knowing the value of X for that particular year. Now another important thing which I will like to tell you, that in the beginning I

told you that there are Y, Y cap and Y bar. Now let us say for year 10, the value of X was 10, for year 10, X is 10 and the value of Y was 23.

So, this is value of Y for year 10. But this is the actual value, actual Y. Now this equation is also with us. When I put this value of X equals to 10 in this equation. So, let us see how much is Y cap coming that is 7.698 plus 1.717 into 10. So, that makes 7.698 plus 17.17 and that is making me 697, this is going to be 25, somewhere around 25.

So, you see actual value of Y is 23. But the estimated value of Y is coming 25. Using this regression equation. So, the estimated values are those values which are which are estimated value, let us say for this regression line, the estimated value of X equals to n is 25. But the actual value of, is 23 only. So this is Y and this is Y cap. So, for each of those points, you can determine that what is Y cap and what is Y. So this is also very important for understanding the quality of our forecast. The relationship of Y cap and Y, this also helps us in determining the quality of our forecast.

(Refer Slide Time: 26:56)



And the quality of our forecast is measured by this term which is known as reliability of the forecast. Now what is the reliability of the forecast, for that purpose you see that we have a regression equation which is for Y cap equals to a plus bX.

Now in this particular regression equation, you will get all the values of Y cap on this regression line. But there will be actual values also, which will be like here. So, these are the actual values

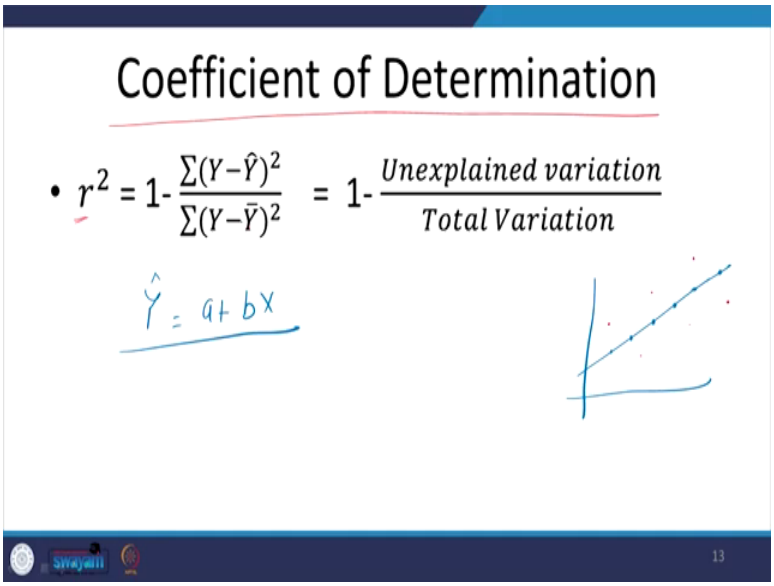
of Y. But there can be one more set of data for which the same regression equation may be applicable and that set of data maybe…

So, here we have two types of data, one data which is represented by a small dots around the regression equation and another set of data which is represented by small circles around the regression equation. You can easily see that points which are represented in the form of small dots are narrowly scattered around the regression equation.

But the circles, the data side which is represented by circles that is widely scattered around the regression equation. So, when the data set is widely scattered, around the line of regression in the equation our quality of forecast is not very good. The same equation same equation is representing both these data sets. But in one case, when the data is narrowly around the line of regression that is for that particular data this line is giving more accurate forecast or reliability of the forecast is good.

When the data set is widely scattered, the reliability of the forecast is not so good. So the measure of reliability of the forecast is given with the standard error of estimate and standard error of estimate is measured using this formula, Y minus Y cap whole square sigma, divided by n minus 2 and then the under root of this whole expression. So, if standard error of estimate is high, it means that data is widely scattered around your line of regression equation and therefore that is not a good measure of your forecast. The similarly one more issue is also there. That is the coefficient of determination.

(Refer Slide Time: 29:55)



Like when you are forecasting and you have regression equation of Y cap equals to a plus bX. Now there are going to be some variations, because as we just discussed that your actual points and the estimated points estimated points are coming on this line. But the actual points are here. So, your regression equation is not able to explain why the points are somewhere else, actual points which are not coinciding with your regression equation.

Because you are regression equation is not able to explain all the variations of your dependent variable. It can I explain to some extent variation of your dependent variable. But it cannot explain hundred percent variation of your dependent variable. So, for that purpose we use this measure which is known as coefficient of determination which is known as r square in very generic language. Many of us know it only by name r square.

So, this r square is 1 minus unexplained variation divided by total variation. And here you see, Y minus Y cap. That is the unexplained variation and Y minus Y bar that is the total variation. So, we calculate this coefficient of determination and if coefficient of determination is more than 65 percent 67 percent, we say that it is a good forecast.

But if coefficient of determination is less than that, then we say that it is not a good forecast. So, you have to you see that your coefficient of determination is also sufficiently good. So, that you can have a good forecast. So, two important measures are there to determine the quality of your causal method. One is the reliability of the forecast, which is known as standard error of

estimation and the second is the coefficient of determination, which is also known as r square. So, by combining these two things, you can see whether you are having a good forecast or not a good forecast.

In this particular case, we have discussed all the things with respect to our simple regression method. Where only one particular variation is there. But as I told you that in practice you may have more independent variables also and when you have more independent variables, then you will have a much better forecast. Because then you will be able to explain more variations of your dependent variable.

There will always be some kind of limitation with only single factor. So, if you are considering more than one factor, in determining the forecast, you will be able to explain most of the variations. So, when we are going for multiple regression method, the idea is to improve our r square. So, that we will be explain more and more variation. So, when we go in the courses of statistics and we discuss multiple regression normal we use 3 to 4 variables, which will affect the demand of the particular product.

So, the idea is to know how I can explain most of the variations. So, this simple regression method is just to give you a foundation of our regression analysis. But in practice we use most of the time multiple regression method, because we want to know maximum variations we want to know the control, over the maximum variation which can affect the demand of our product. So with this we come to end of this session. Thank you very much.