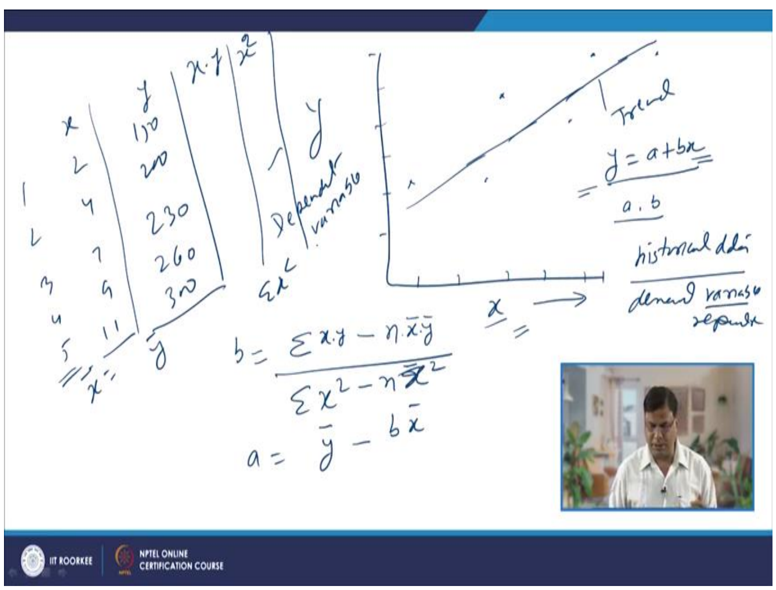
Principles of Industrial Engineering
Professor D. K. Dwivedi
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee
Lecture 52
Forecasting: Methods – V

Hello, I welcome you all in this presentation related with the subject Principles of Industrial Engineering. And in this presentation I will be talking about one method of the forecasting that is called casual method and in in this method basically, dependent variable is related with the independent variable, and there can be variety of approaches in casual methods like, simple regression analysis or multiple regression analysis.

Here, in simple regression analysis, we use one input variable and there will be one output response.

(Refer Slide Time: 1:20)



So here, basically, the output or the required variable y, is kept on the y-axis and independent variable, most of the time, period or time, is kept on the x-axis. So this may be like in months or the quarters or any, it may be any other independent variable also. And here, it may be demand, it may be sales or any other variable.

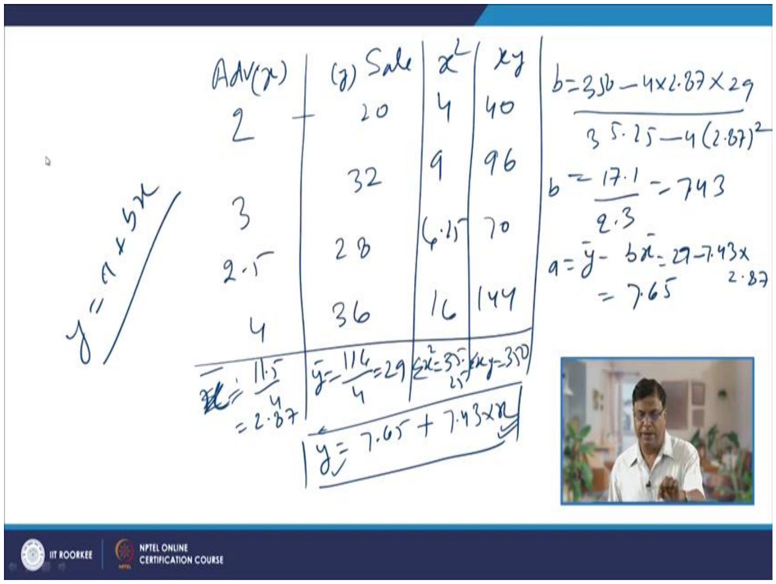
So basically, here the dependent variable is plotted like this, as a function of the x and we try to determine a trend line, the equation for the trend line corresponding to this variation, in y with respect to the x and that is expressed as y is equal to a plus b x. So here, say, there are different, there are in this one, basically x is the independent variable and y is the dependent variable, and a and b are the coefficients which are determined based on the historical data of demand or the variable and the response which is like the historical data of y and the x.

So relationship between the x and y is established through this kind of equation using the historical data of x and y. Say here, b is obtained using the equation, say here x and the y values and say few values are there like 2, 4's, 7, 9, 11 and corresponding values 150, 200, 230 then 260 and a 300. This is the kind of the, the y values are available for the corresponding x values.

So what we try to do is, using the, these values of the x and y, we try to calculate the, the coefficient values of a and b. So here, how many data points are there? Like 1, 2, 3, 4 and 5. So, the b coefficient value is calculated using the sum of all x and y, x into y product minus n, is the number of data point that is here, 5 into x bar, average x value and average y value, product of number of data points into average x and average of y, divided by sum of x squares or all x square values minus n into, n into x bar square.

This is how do we calculate the b, and a is calculated using y bar minus b into x bar. So for calculating a and b, what we need, the product of x into y, then we need, the x bar, average of x and average of y, then we also need x square values, then we need, sum of all these x square values.

(Refer Slide Time: 6:00)



So basically, these are the calculations that we have to make in. So say, we will take a simple example, like expenditure being made on the advertisement and its relationship with the sale. Data is identified say, some suitable unit we may assume like 2, 3, 2.5 and 4, and corresponding sale values are like 20, 32, 28, and 36.

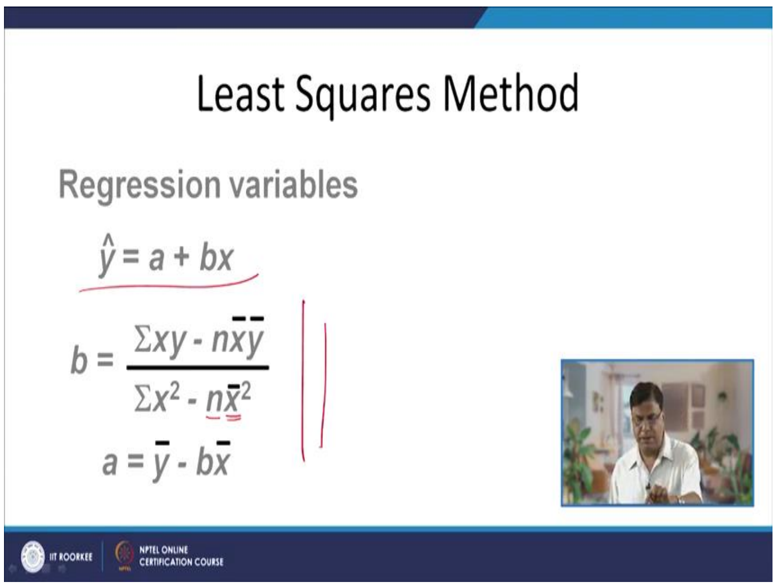
So this is x, that is independent variable and dependent variable sales is say y here. So x square will be here, like say, 4, 9, 6.25 and 16. Then x into y will be calculated, so it will be, 20 into 2, 40; 32 into 3, 96; then 2.8, 28 into 2.5, it is 70; and then 36 into 4, 144. Then what we will be doing, sum of, we will be determining the x bar, that is the average x value, so average x value will be like, here 7, 11.5 divide by 4. So, this so will be 2.87.

Similarly, y bar value is obtained from the sum of all these y values that will be 116 divide by 4 giving us 29. Then sum of all the x square values is obtained, that is coming 35.25, and then sum of all x into y values, that will be 350. Now what we have to do is, now we have to put all these values in the equation for b, so what we get in like 350 minus 4 into 2.87 into 29, divide by 35.25 minus 4 into 2.87 whole square.

So this will be giving us on solving, 17.1 divided by 2.3, and that in turn will be giving us the value of b equal to 7.43. So to calculate a, what we need, y bar minus b into x bar. So y bar here is equal to 29, y bar is equal to 29 here, so 29 minus b into x bar. So b is 7.43 and x bar is equal to 2.87. On solving this, we get the value of a equal 7.65. So our equation for y becomes equal to 7.65 plus the b is 7.43 into x.

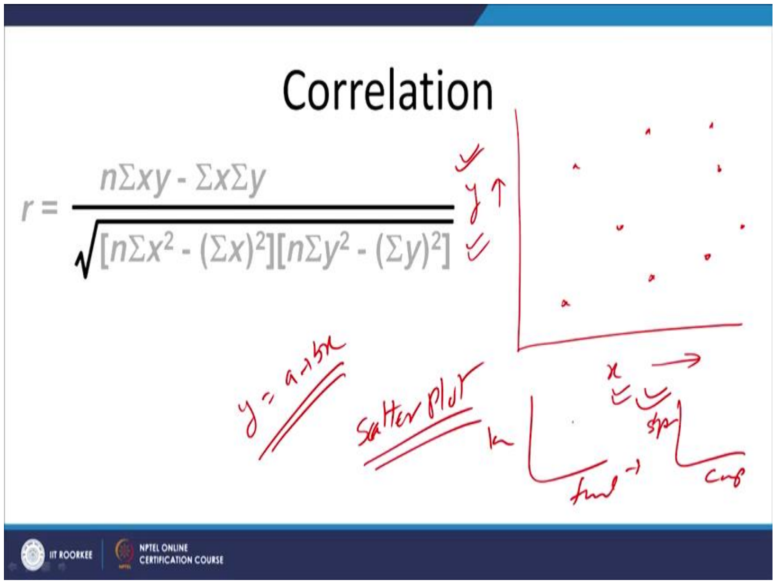
So this becomes the equation for simple regression equation showing the relationship between the sales and the expenditure on the advertisement. So this is how we can establish the casual relationship between, between the independent variable and the dependent variable.

(Refer Slide Time: 10:37)



So this is what has been explained in this regression method for calculating the y is equal to a plus b x and these are the equations which can be used to determine the relationship between the dependent variable and independent variable.

(Refer Slide Time: 10:57)

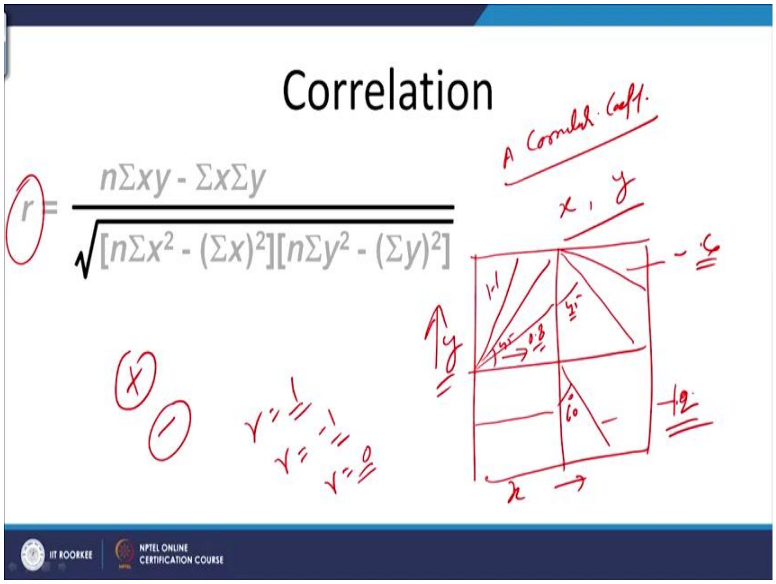


Many times what we see that the data is plotted in x-axis and y-axis, and we do not know really what is the kind of the relationship, there is no typical trend like what we have seen in where we can show the relationship y is equal to a plus b x.

But, if we do not have the kind of relationship, in linear form, then how to get the, how to have the understanding on the way by which independent variable x is affecting to the dependent variable y. Say, there are various values lying here and there. So how to have a kind of understanding the way by which x is affecting to the y.

So basically, we develop a scatter plot, and in a scatter plot, x showing the independent variable and y is showing the dependent variable. It can be like number of kilometers, the kilometers in y-axis and the fuel consumed in the x-axis. It can be the composition in x-axis and the strength in y-axis. So there can be number of things but there can be different, the way by which the y value is changing with respect to the x, that can vary significantly and to understand that relationship, basically we try to establish a correlation.

(Refer Slide Time: 12:50)



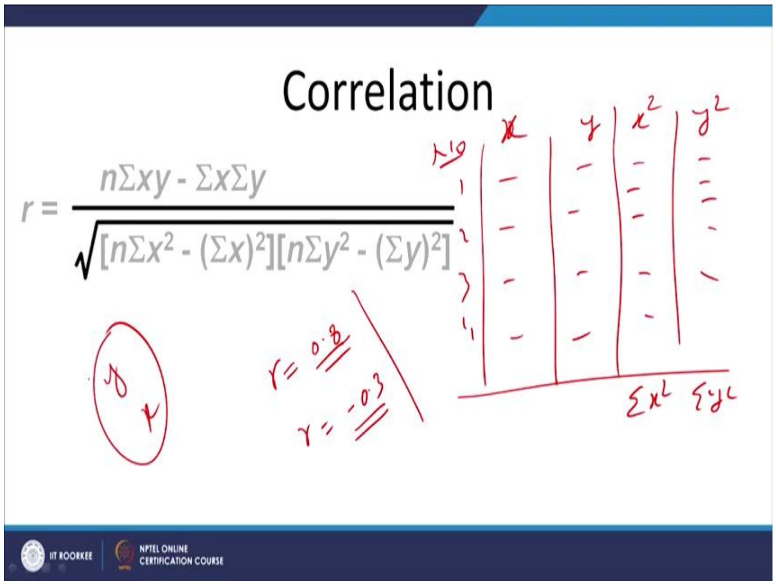
So a, a correlation coefficient is determined. Using the values of x and the corresponding values of y. There are different kinds of the trends which are observed as far as this relationship is concerned. It can be like this, so here x, this horizontal line will be showing x-axis and vertical line will be showing the, the y values, dependent variable.

So here, say, the for variation in x, if the change in x increases, the y like this where there is a direct relationship, so this is the case where you increase the x and the y will also be increased by the same, in the, the same way linearly with the, an angle of say 45 degree. So this kind of the, when this kind of the relationship exist, the coefficient, correlation coefficient value r, will be equal to 1. And when this value, value of y decreases linearly in negative way, again showing the 45 degree angle, in that case the relationship coefficient value r, becomes equal to the minus 1.

And when the value of the y does not change with the x, then the value becomes equal to 0. So which means, the value if higher than 45 degree, it may be like 1.1, if it is lesser than 45 degree, it may be like say, the 0.8 correlation, the coefficient may be 0.8 and if the, the slope is decreasing slowly it may be like minus 0.6.

So, if it is decreasing further sharply, angle greater than like say greater than 60, 45 degree and this angle say, is negative slope is of 60 degree, then in that case, like say, the correlation coefficient may be like minus point, minus 1.2. So, so whether the relationship is positive, whether there is a negative and how strong positive, how strong negative, it will be, that can be easily understood by calculating the correlation coefficient and correlation coefficient again can be calculated using the things in the same way.

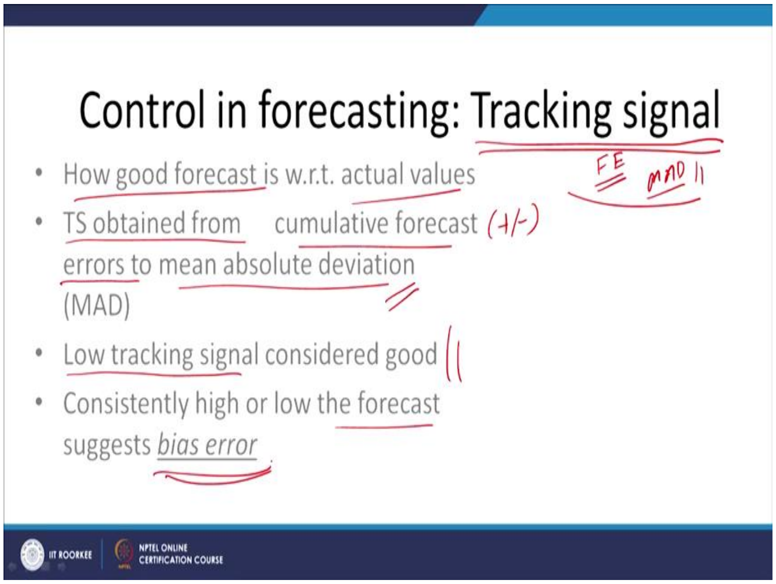
(Refer Slide Time: 15:36)



How? Like say, we have the values of x and corresponding values of y, and then we again calculate the things in the same way, like x, then we have to calculate the x square as well as the y square values and their sum also is to be calculated. So, and the number of data points which are there like say, so 1, 2, 3, 4 and here we may can have the different values, y values and their corresponding x square values and corresponding y square values their sum of x square and sum of y square and then these values can be put in to find out the correlation coefficient.

So if r is like say, the 0.8, the relationship is a positive and if it is, r is minus 0.3, so there is a weak negative correlation and a, a positive, strong positive correlation between the x and y values, or the relationship between the y and x is positive, means effect of x on the y can be positive or negative or it can be 0 that is what can be easily establish, established through this correlation coefficients.

(Refer Slide Time: 17:17)

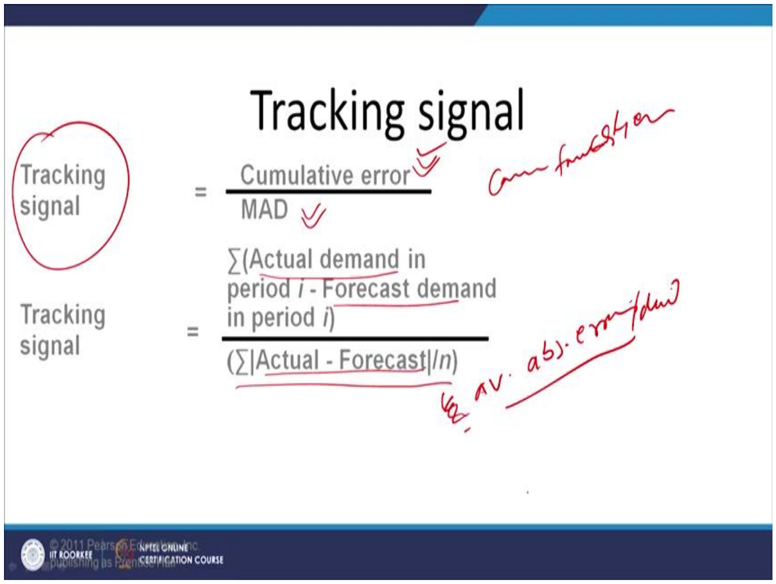


Now, it is important to see really, if we are doing the forecasting, then how good the forecasting is. So, one method was like, we calculate the forecasting error. But, like say, the forecasting error, like in form of a mean absolute deviation or the mean percentage, mean absolute percentage error or mean squared error.

So those things can be used, but whether we should continue with the current forecasting method or we should go with, we need to change the approach of the forecasting, we need to have a proper control over this through the tracking signal determination. So, how good forecasting is, with respect to the actual values, that is what can be done or can be determined through the tracking signals.

The tracking signals is obtained through the cumulative forecast errors to the mean absolute deviation. So the cumulative forecast errors, which considers both positive and negative signs, and mean absolute deviation does not consider the signs. Then, when we have the values of the tracking signals, the low tracking signal indicates that the forecasting approach is good and when we have the consistently high or the low values of the tracking signals, then it will show that the forecasting suggests the bias error in forecasting and we need to look into this one.

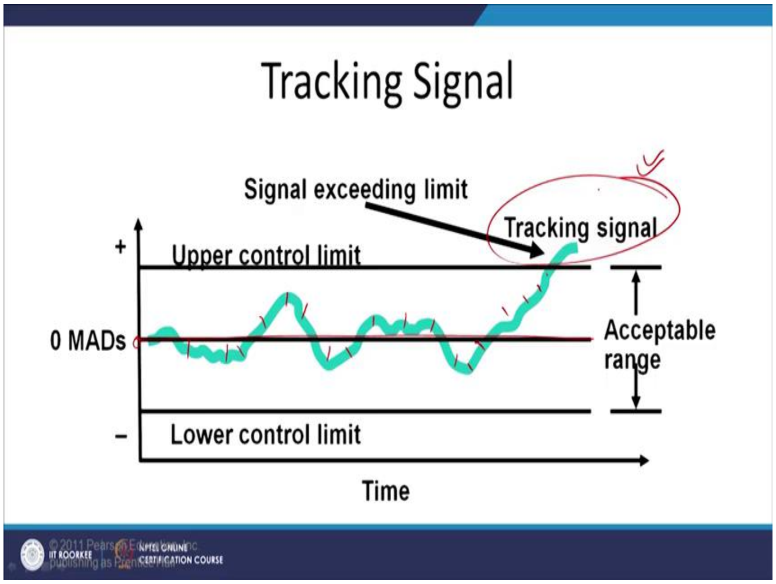
(Refer Slide Time: 19:23)



So, for tracking signals we need to, we need to calculate certain things like the tracking signal is calculated through the cumulative forecasting error and the mean absolute deviation. So sum of the, sum for given periods, sum of the actual demand minus the forecast demand divided by the sum of the absolute values of the difference in actual and forecast demand.

Basically this is mean absolute deviation, so the difference of the actual and the forecast demands and sum of all the difference of the actual forecast demands and there, divided by the number of data points which are being used. So basically this is average absolute error and here we use, sum of average, basically sum of the error or the deviation. Basically, this is the average absolute error and this is cumulative forecasting, casting error.

(Refer Slide Time: 21:00)

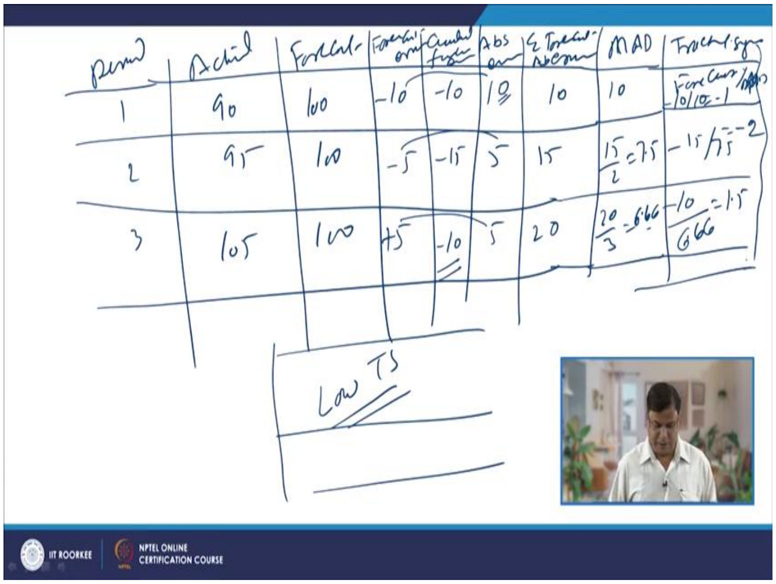


So the cumulative forecasting error and the mean deviation. So now, we will see one example for calculating the tracking signal related thing, so this is the mean absolute deviation line where it is 0, means, the forecasting and actual demands are exactly matching and the mean absolute deviation is 0.

But, it will hardly be happening, there will always be some kind of the difference and so the mean, the mean absolute deviation is plotted like this and when we find that the so basically, mean absolute deviation in the y-axis and when we plot that tracking signal and this will be indicating whether, whether the mean absolute deviation is within the limit or not.

So basically, we calculate first the tracking signal and then it is a plotted. So how to calculate the tracking signals?

(Refer Slide Time: 22:11)



So here, let us say, a very simple example, say the actual demand, forecast demand. And based on this and the period we can say, so like we have 3 data points. So period 1, 2 and 3. So the actual data is 90, 95 and 105 and say forecast is 100 in all these 3 cases.

So the forecast error is actual minus forecast, so here it is minus 10, minus 5 and 5, plus 5. So, now this is the, the forecast error. Then, we have the cumulative forecast error, cumulative forecast error, so here, the first is minus 10, second is minus 15 and then it will be again minus 10. Then, the mean absolute error, the absolute error, absolute error here is, absolute forecast error, the 10, 5, this is being taken directly from here, 5, 10, 5 and 5.

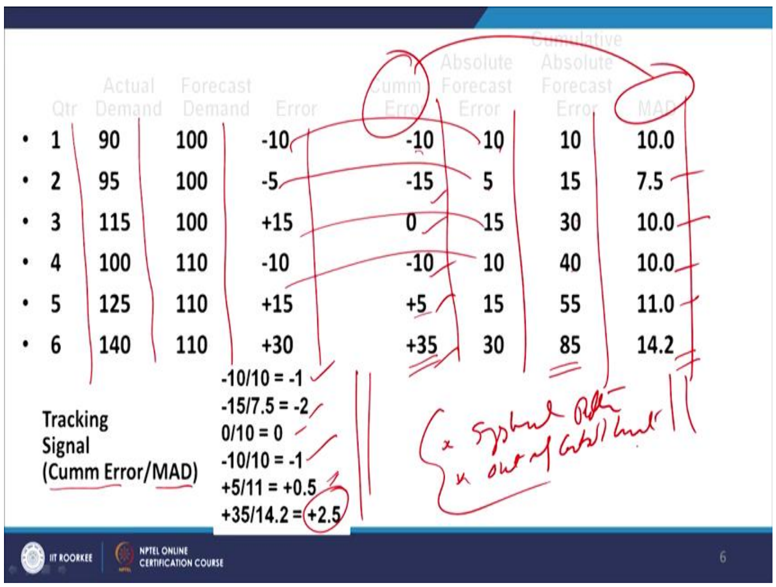
So, sum of the forecast error, sum of forecast error, here, sum of forecast error, absolute error basically, so this will be like 10, 15 and 20. Now, we have to calculate the MAD, mean absolute deviation, mean absolute deviation, so absolute error was 10, so here for 1 data it is just 10, for the 2 data point, it will be like 10 plus 5, 15; so 15 by 2, it will be 7.5; and for 3 data points, it will be 20 divide by 3, so we will have, the 6.66.

Now, to calculate the tracking signal here, what we have to do? We have to use the forecast error, forecast error divide by the MAD, mean absolute deviation. So the for first, it is 10 minus 10 by 10, so it will be minus 1. Then, for second, it is minus 15 by 7.5, so it will be 2, minus 2.

While for third it is, minus plus 5 forecast error is here it is cumulative one, is this so 15, 10, 15, 2.5 and it is minus 10 divide by 6.66; so here, it will be say 1.5.

So this is how we get the tracking signal and then tracking signal, as I have said if the, lower the value of the tracking signal is considered to be good because it will be closer, that demand, actual demand will be close to the forecast while if there is a consistent increase or decrease or it is showing particular pattern then it shows that tracking signal, this basically this tracking signal is plotted and we try to see that how much variation in the tracking signals.

(Refer Slide Time: 27:20)



So basically, the tracking signals will be indicating how good forecasting, there is one more example here related with the tracking signals, here that I can take up like say, for different quarters there is a different actual demand values and the corresponding forecast is 100 and the so based on the actual demand and forecast we have got the errors.

Like say, 90 minus 100, minus 10; 95 minus 100, minus 5; like that. And the cumulative errors are calculated here, minus 10, then minus 15, then minus 15 plus 15, 0; and then 0, then plus, minus 10, minus 10; and then plus 15 so plus 5 like this. The well, cumulative error is calculated and absolute forecast error is taken without sign directly from here like this. And then cumulative absolute forecast is calculated here 10, 15, 15, 15, 30, then 30 plus 10, 40; 40 plus 15, 55; and then 55 plus 30, 85.

And the MAD value is calculated and these MAD value and then the cumulative error. Cumulative error and MAD values are used to calculate the tracking signal. Cumulative error and MAD values are used to calculate the tracking signal. So here, minus 10 divide by 10 is minus 1, minus 15 divide by 7.5 is minus 2, then 0 divide by 10 is 0, then minus 10 divide by 10, again minus 1, then plus 5 divided by 11, so here point 5 and then 35 divided by 14.2 it will be giving us 2, plus 2.5.

So this is how the tracking signals can be calculated, these are plotted in control charts to see if it is within the limits or it is going beyond the limits or it is showing particular kind of pattern instead of having the random variation in the tracking signal value, so like, some systematic pattern or the, the values are going out of the control limits, beyond the control limits, then these will be indicating that there is a problem with the forecasting and we need to look into the, the way by which it is being done.

Now, I will summarize this presentation. Basically, in this presentation, I have talked about the simple regression analysis method, correlation coefficient and the track effectiveness or the utility of the tracking signal methodology for calculating the tracking signals. Thank you for your attention.