Data Analysis and Decision Making - I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture - 27
Forecasting

A warm welcome to all my dear friends and students, a very good morning, good afternoon, good evening to all of you and this is the DADM which is Data Analysis and Decision Making I; course under VLFM MOOC series. And this is a 12 week course for 30 hours total number of lectures being 60 and each week we have 5 lectures; each lectures for half an hour. And I am Raghu Nandan Sengupta from IME Department IIT, Kanpur and we are going to start today the 27th lecture; so, we are in the sixth week.

Now if you remember we were discussing about different types of forecasting methods and there we considered that we can either use the exponential smoothing method or the weighted method. Then we had the exponential smoothing would be for 1 time period 2 time period 3 time period by the word time period I mean that standing today you are trying to utilize the past data.

So, 1 time period would basically mean you are putting weightages of alpha 1 and alpha 2 to the predicted and the forecasted values for t minus 1 such that the weights basically add up to 1. If it is 2 time period in the past; that means, you are trying to utilize the forecasted and the actual values for t minus 1 and t minus 2 and try to give weightages of alpha 1, alpha 2, alpha 3, alpha 4 such that you are able to predict the best what is the what best I keep repeating it; I will again repeat it such that the weights add up to 1.

And the best prediction basically you want to basically minimize some error. So, this in this case we are considering the error and the mean square errors such that you sum the square of the errors, differentiate with respect to the parameters put them to 0 and solve your problem.

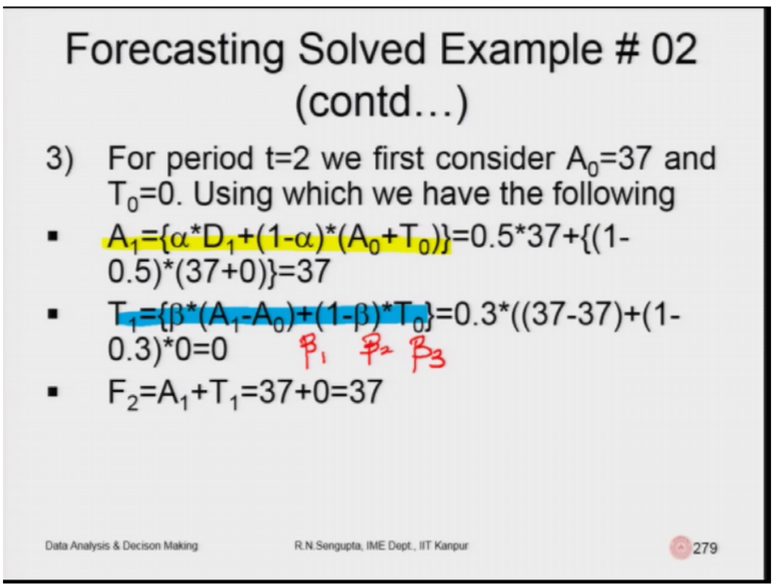
Now, I also did point out and; obviously, the question would be very relevant from all of your side is that if the parameters on the variables based on which the weights, based on which you are trying to calculate or themselves changing; so, that is possible. So, you basically you will have the trend and there would be some mind plus minus some deviation from the trend such that we are able to predict to the maximum possible extent.

Now when we consider the linear regression we had the concepts of trying to predict are the same in the sense you have the present data for all the past values for the y's and you regress on them trying to basically predict for the future. In the case of multiple linear regression again you have the present data, but the present data only is not y; there are basically facts of xs which are dependent variables some assumptions are also there.

If you remember, implicitly we do not assume any assumptions for the forecasting method trend analysis; only that we consider the errors to be normally distributed. And this normality of x's, y's covariances not existing between the x's covariance not existing between the errors and x those are whole set of assumptions we have assumed in the regression context.

Now, in the problem formulations very simply we consider the data being given; then we took the 3 month moving average, the 5 month moving average then we can consider the exponential smoothing. And in the exponential smoothing we consider the weights such that we will try to basically minimize the error that will come later. And in the later part we will consider today is basically the trend analysis and how the trends with respect to fluctuations would be considered in order to give us the best forecast for the future.

(Refer Slide Time: 04:15)



So, we are using the trend analysis and for the example number 2; if you refer to the last set of slides in the last class. So, for period 2 based on period 1; we will consider first consider the A 0. So, if you remember we will consider A 0 or the T 0; the trends and the fluctuations basically are 0 because you have to start at one point using which we will basically try to predict.

So, the values of so, A 1 let me highlight it. So, the value for A 1 is I am putting a weights of D 1 has been given away, D 1 is basically the demand or the y value D 1 has been given a weight of alpha. And obviously, if you if you remember we always half on the fact that 1 minus alpha is the other weight such that the sum of the weights is 1. So, 1 minus alpha would be the weight for the initial A naught and T naught value.

So considering 50 percent, 50 percent being the weights for alpha and 1 minus alpha; so, the value calculate for A 1 would be we are considering t time t period T 2 because we are starting at T 1 for which we are considering 0 value suffix 0. So, 0.5 into 37 plus 1 minus 0.5 which is again 0.5 into 37 plus 0; so, because we are considering A naught as 37 and that trend as 0. So, based on that we find not A 1 as 37; so, A naught was 37, A 1 was 37 now we need to find out T 1. So, T 1 would basically be the other parameter based on which we are trying to basically find out the actual predicted value.

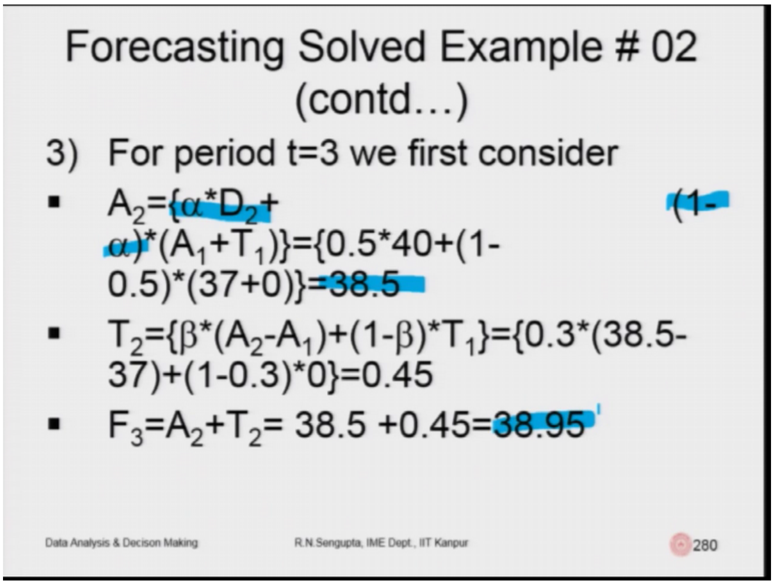
So, one are the weights for the past data and one is basically the weights on the trend based on which you are going to try find out in which direction the trend is moving. So, T 1 would be beta; we will consider the parameters as beta. So, beta into the trend difference A 1 minus A naught and plus 1 minus beta would be see again the weights on t naught.

So, we could have been several example we could have taken T naught, T minus 1, T minus 2. So, all these weights would have come here let me use a different. So, this in place of T naught we could have also considered the values of beta 1, beta 2, beta 3; so, corresponding to T minus 1, T minus 2.

So, in this case it would have been not only beta n 1 minus beta 1 minus beta. So, you would basically have a beta 1 here, a beta 2 here and a beta 3 here considering beta 1 is the weight for difference between in 1 minus A 0 beta 2 is basically the difference the weight we are giving to T naught and beta 3 is the weight we are trying to give to T minus 1.

So, here we consider only 2 betas. So, the sum is one hence the equation becomes 0.3 into the difference between A naught and A 1 which is 0 because 37 minus 37 is 0. And 1 minus 0.3 is basically 70 percent or 0.7 into T naught which you already considered as 0; so, the value becomes 0. So, there initially T naught is also 0 T 1 is also 0; based on that we find out the forecasted value for time period 2, time period 2 would be basically the trend plus the fluctuation; it comes out to be 37 plus 0 is equal to 37, I will come to the detailed tables within few minutes.

(Refer Slide Time: 08:02)



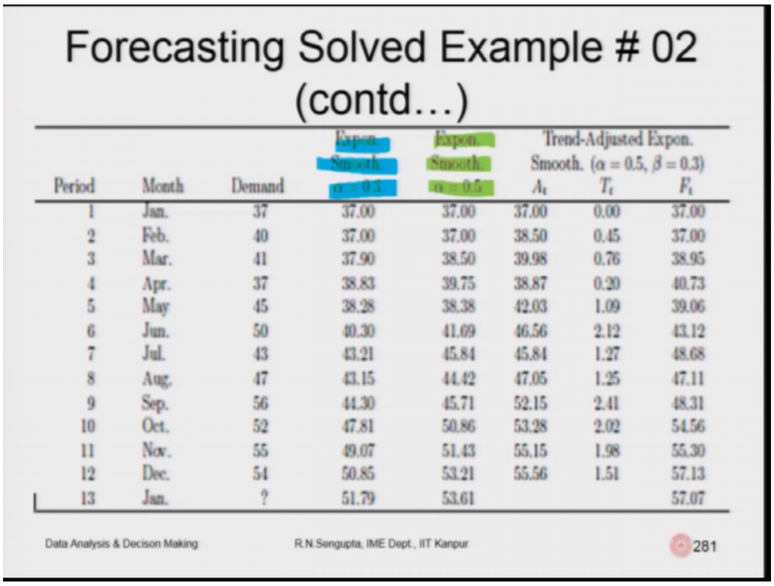
Now, for a time period 3 we need to first consider the formula then alpha plus. So, alpha plus D 2 would be the alpha would be they wait for demand 2 actual value for the last period. And 1 minus alpha, see it has shifted because due to the indentation there is no such problem as such. So, there is no missing values or formulas in side.

So, 1 minus alpha would be the weights given to A 1 plus T 1; trends and the fluctuations for time period 1. So, it is again alpha we have considered as 0.5 why we are considering 0.5? I am going to come to that later 50 percent into 40 plus 1 minus 50 percent into 37 plus 0; so, the value becomes 38.5.

Similarly, again T 2 would basically be given weights for to calculate T 2; we will give weights of beta and 1 minus beta 2 the values of the difference initially the different between A naught minus A 1 and in the second case 1 minus beta would be the weight given to T 1. So, based on that once you calculate it is 30 percent or 0.3 into 38.5 minus 37 plus 17 to 0; it comes out to 0.25.

So, now, the trend basically percentage has started; similarly, so because in the first case T 1 and T naught was 0. So, now, F 3 basically is the value given by a 2 plus T 2 and if you add them up the values corresponding the fact that. For the first time A 2, T 2 and all these values are changing with base back to the base value which we have started and taking. So, the F 3 value becomes 38.5 plus the fluctuation which is zero point percentage sort of thing or the variance sort of thing which is 0.45 and it becomes 38.9.

(Refer Slide Time: 09:59)



Forecasting Solved Example # 02 (contd…)

Period	Month	Demand	Expon. Smooth. $\alpha = 0.3$	Expon. Smooth. $\alpha = 0.5$	Trend-Adjusted Expon. Smooth. ($\alpha = 0.5, \beta = 0.3$) A_t	T_t	F_t
1	Jan.	37	37.00	37.00	37.00	0.00	37.00
2	Feb.	40	37.00	37.00	38.50	0.45	37.00
3	Mar.	41	37.90	38.50	39.98	0.76	38.95
4	Apr.	37	38.83	39.75	38.87	0.20	40.73
5	May	45	38.28	38.38	42.03	1.09	39.06
6	Jun.	50	40.30	41.69	46.56	2.12	43.12
7	Jul.	43	43.21	45.84	45.84	1.27	48.68
8	Aug.	47	43.15	44.42	47.05	1.25	47.11
9	Sep.	56	44.30	45.71	52.15	2.41	48.31
10	Oct.	52	47.81	50.86	53.28	2.02	54.56
11	Nov.	55	49.07	51.43	55.15	1.98	55.30
12	Dec.	54	50.85	53.21	55.56	1.51	57.13
13	Jan.	?	51.79	53.61			57.07

Now, the table which is shown here is basically goes like this. So, this is the extended calculation which we are doing for only 2 things which I have shown.

So, on the first column on the left hand side we have the periods. So, periods are basically for this case given for; 1 for January; January to December for 12 months which is 1 to 12. Demands are given in the third column where I am pointing my finger starting from 37 to 54 and we do not have the demand for the thirteen time period which is the January for the next year.

Based on the demand once we find it out we will compare our results that will come later on. Now we will consider the expansion smoothing same problems here considered for 30 percent and 50 percent. So, we write the values in the columns as I am working 37 and the other one is basically 50 percent. So, this is for the 50 percent now based on the fact we use the same formula; that means, if it is 30 percent we use 30 percent weightages and 70 percent weightages for the predicted and the actual value for one time period difference one step backwards.

In the 50 percent alpha is equal to 0.5; we use 50 percent and 50 percent. So, what is 0.5 and 1 minus 0.5, for the one step past data which is actual value and the predicted value for t minus 1. So, based on that we continue; so, given F 1, D 1 or F 1 sorry forecasted and the demand (Refer Time: 11:42); so I am using D 1 in place of y 1.

So, F 1 D 1 given we find out F 2 then we find then actually get the data which is D 2 and compare D 2 and F 2. Then given D 2 and f F 2 we already found out we go into stage 3 or time period 3, we find out F 3 so; that means, given alpha as 0.30 percent.

Then once the time period comes we find we are given the value of D 3 compared F 3 and D 3. Then we go to a time period 4 calculate F 4 then once the actual value D 4 is given you calculate D 4 and F 4. So, in this way this way we go step by step and the difference of the F i and D i which is D 1 minus F 1; get when you positive or negative whatever it is D 2 minus F 2, D 3 minus F 3, D 4 minus F 4 all this will be errors for epsilon one epsilon 2 epsilon 3 and so on and so forth; till the twelfth month.

And we squared this error sum them up differentiate with respect to alphas, put to 0 find it out this is what we are doing. Similarly, we do the same thing for alpha is equal to 0.5 in place of 30 percent, we have to use 50 percent and calculate the values. Now in the train adjusted exponential smoothing process we need to find out the trends and the fluctuations.

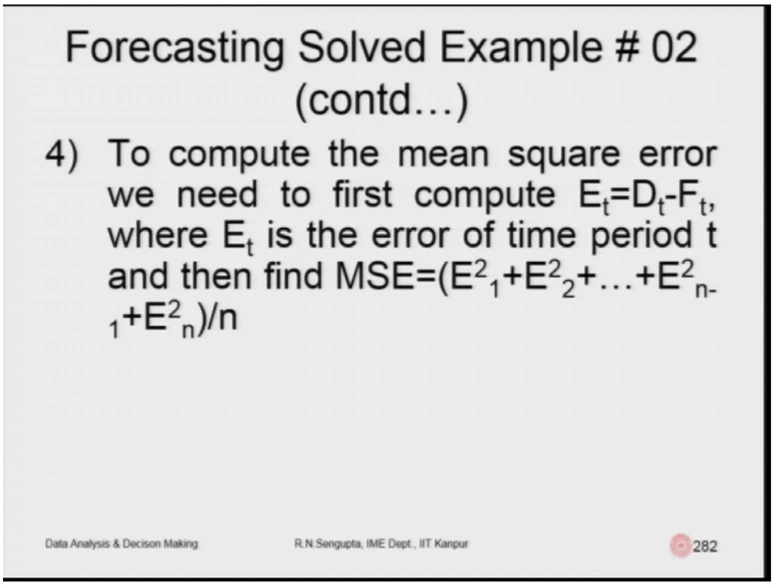
So, the trends of the fluctuations are technically the values based on which we find out A t and T. So, A naught T naught are considering as 0 based on that we proceed. So, the A naught sorry by mistake T naught would be given as 0 and A naught would be given the actual value which is predicted; the actual value on time period 1, which is 37 and F 1 would also be 37.

So, based on because the calculation will show that because T naught is 0; based on that you go step by step first find out the values of A and T; what at the particular time utilize that when that values of A and T to predict basically F for the next weighted time. So, A

1 T 1 calculated they give F 2; then we get D 2 then again we try to calculate A 2; A 2, T 2; based on that we try to basically calculate F 3, then D 3 is given, then we find A 3, T 3 given that we find out F 4 and a given F 4 we generally find out from the once the value comes we find out D 4 and continue in this case.

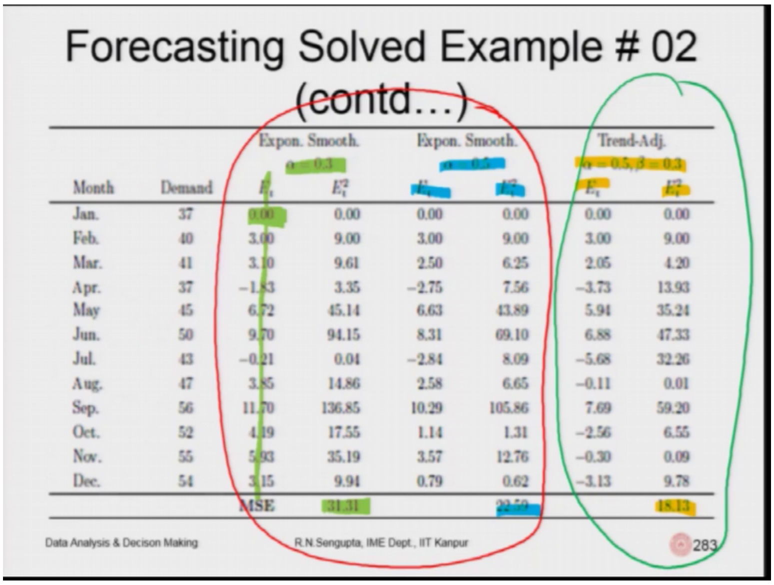
Now, again the question would come what we do how do we minimize? So, basically utilizing that alpha and beta you find out the reference of F i and D i; find out the errors which are epsilon is, square them up, sum them up and differentiate with respect to betas and alphas. You can do step wise; that means, to find out the trend as well as the actual values. And once given the alpha hats and beta hats you can proceed to find out for the later time period accordingly. Now we are also given to compute the mean square error; we need to basically compute the errors.

(Refer Slide Time: 14:52)



Errors I am considering naught epsilon as E; suffix t. So, that would be the difference between D t and F t the actual value and the forecasted value, where E t is the error of time period t and then we need to find out the mean square errors would E one square plus E 2 square dot dot till the values which you have which is E n square n time period, sum them up divided by the number of time periods which is n. So, we need to basically find out this calculations.

(Refer Slide Time: 15:19)



		Expon. Smooth. $\alpha = 0.3$		Expon. Smooth. $\alpha = 0.5$		Trend-Adj. $\alpha = 0.5, \beta = 0.3$	
Month	Demand	E_t	E_t^2	E_t	E_t^2	E_t	E_t^2
Jan.	37	0.00	0.00	0.00	0.00	0.00	0.00
Feb.	40	3.00	9.00	3.00	9.00	3.00	9.00
Mar.	41	3.10	9.61	2.50	6.25	2.05	4.20
Apr.	37	−1.83	3.35	−2.75	7.56	−3.73	13.93
May	45	6.72	45.14	6.63	43.89	5.94	35.24
Jun.	50	9.70	94.15	8.31	69.10	6.88	47.33
Jul.	43	−0.21	0.04	−2.84	8.09	−5.68	32.26
Aug.	47	3.85	14.86	2.58	6.65	−0.11	0.01
Sep.	56	11.70	136.85	10.29	105.86	7.69	59.20
Oct.	52	4.19	17.55	1.14	1.31	−2.56	6.55
Nov.	55	5.93	35.19	3.57	12.76	−0.30	0.09
Dec.	54	3.15	9.94	0.79	0.62	−3.13	9.78
		MSE	31.31		22.59		18.13

So, here are the charts. So, charts are basically the continuation example 2; we now trade basically try to find out that the error is based on the fact what are the alphas, what are the trends and all these things.

So, again the chart has the same information in the first column the serial number is not there. So, I have we have omitted it months are given January to November and because we are predicting for December; so, that would also be used. The demands given which is the second column now I have basically at the errors. So, the errors would be the demand minus the actual value like forecasted value. So, once we have for the first case we have considered, remember we have consider the actual value is equal to demand hence the error would be 0 here.

So, the second case it will be basically F 2 D 2 differences F 3 D 3 differences; whatever you follow the same nomenclature. If it is D i minus F i calculate it accordingly it can be positive negative whatever it is. So, once we do that we have the third column which is basically the errors. So, I just mark it out; now you square them considering the fact that this is this is I did not mention because I have been repeating time and again.

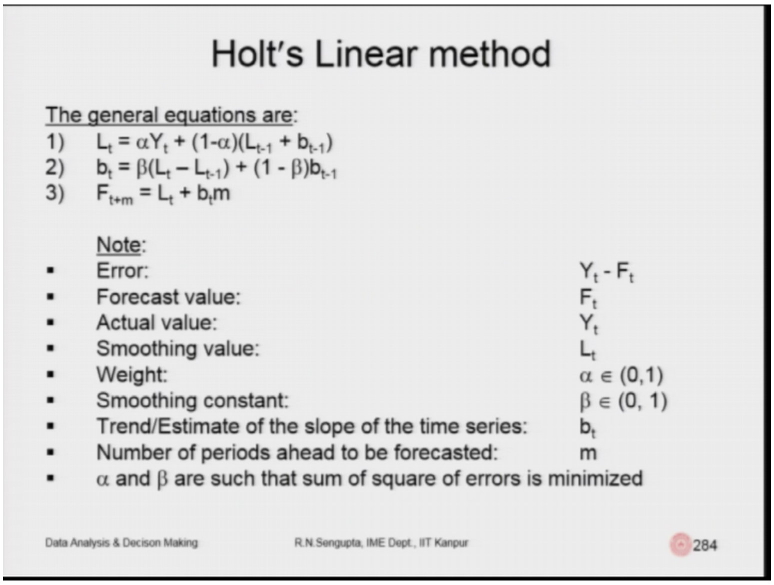
So, this is the errors based on the fact that I have used alpha as 30 percent or 0.3. Now in the fourth column once the errors are calculated, we basically find out the square of the errors and then sum them up. So, if this becomes 31.31; now again we do the same calculation for alpha is equal to 0.5; find out the errors then square the errors and then again sum them up. So, in the second case for the alpha is equal to 0.5 you have 22.59 as the error and 31.31 as the error for the case of alpha is equal to 0.3. So, we can definitely say using a weightages of 51 and 50 percent in the exponential smoothing method; we are getting a much less and less error square of the errors.

So; obviously, it will mean that on the prediction based on which we are doing for alphas, betas; if in case they are betas they would be the best; obviously, you can find out some values; obviously. So, iteration methods some values which will give us the best alpha. Now in the case of the trend, we do the same calculations; we find out for the alpha is equal 0.3 and beta is equal to 0.3. We find out the errors which is the second last column; I should use a different colours this also I should use a different colours sorry for that, but just give me a minute.

So, this is the error for 0.5; now I go to for trend analysis this is the error square and the values of the error square is coming out to be 18.13 which means that using the trend analysis there as a slowly decreasing. So obviously, there can be in case if we want to find out it may be possible in the exponential smoothing method which we are using; we can find out some alpha best which will give us the minimum error.

Then again we can do the same thing for the trend analysis and consider there is only one value of alpha and beta we are taking. We can consider different values on alpha and beta and can predict for which the errors squares some of those square of the error with the list and we can proceed accordingly.

(Refer Slide Time: 19:06)

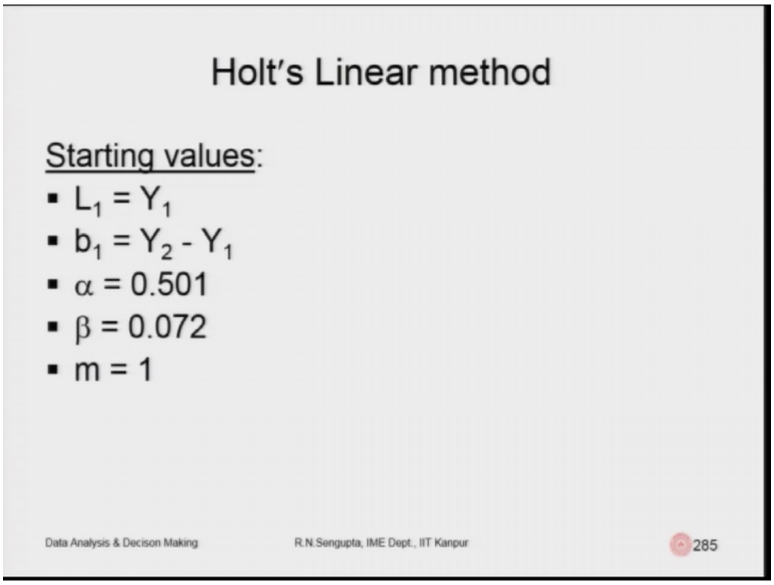


Now, in the Holt linear method the general equations are given by I will just first before coming to the equation; I should basically consider what are the nomenclature. So, the nomenclature errors are E capital E which is given by difference of Y t and F t; F t is the forecasted, Y t is basically the actual which is basically demand which you are considering for the examples.

The forecast is value values are given by F t, actual values are given by Y t or D t, the smoothing value is given by L suffix t. So, how it will smooth out the fluctuations and we also give which we have also consider we also give the this is the same repetition which we have done first I had discussed the problem then I am coming. So, we considered the trend the weights and the smoothing constant. So, alpha is the weights we are giving and the beta is the smoothing constant.

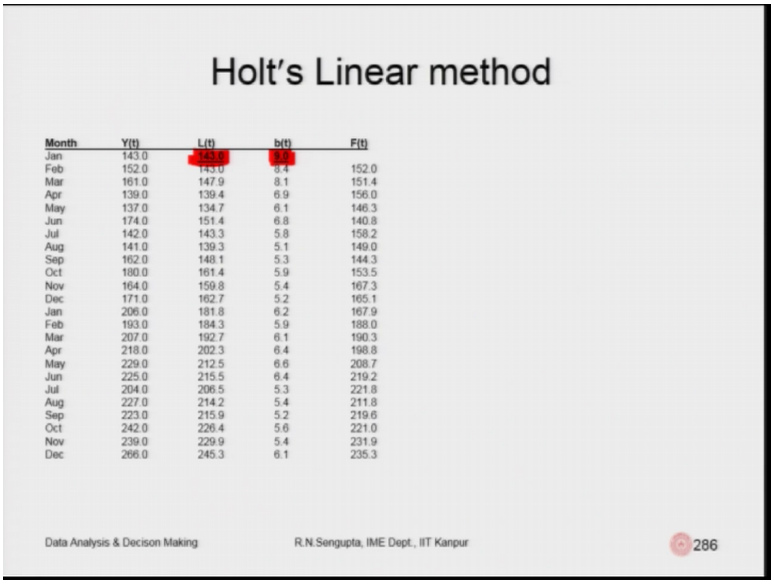
The trend of the estimate of the slope of the line is given by basically beta and the number of years ahead based on which we will try to basically forecast is m. So, it can be 1 period time period, 2 time period 3 time period so on and so forth. So; obviously, the answer remains that how do we find out alpha and beta the logical thing being that we will try to basically find out the sum those square of the errors, differentiate with respect to alpha and beta at stages, try to basically put to 0 which is minimizing the sum of the square of the errors, find the alpha and beta and proceed accordingly.

(Refer Slide Time: 20:44)



So again the Holt linear method with L 1 and Y 1 given as fixed which is the 37 value which you took; we consider that the difference of beta b 1 as in the case because they would be a trend and the fluctuations. So, we consider the difference about D 2 and D 1 which is Y 2 and Y 1 and alpha and beta are given arbitrarily I am considering as 0.501 and 0.072 and n; m you are considering the time period as 1; it can time period 2, 3, 4 whatever it is based on Z that we can do the calculation.

(Refer Slide Time: 21:20)



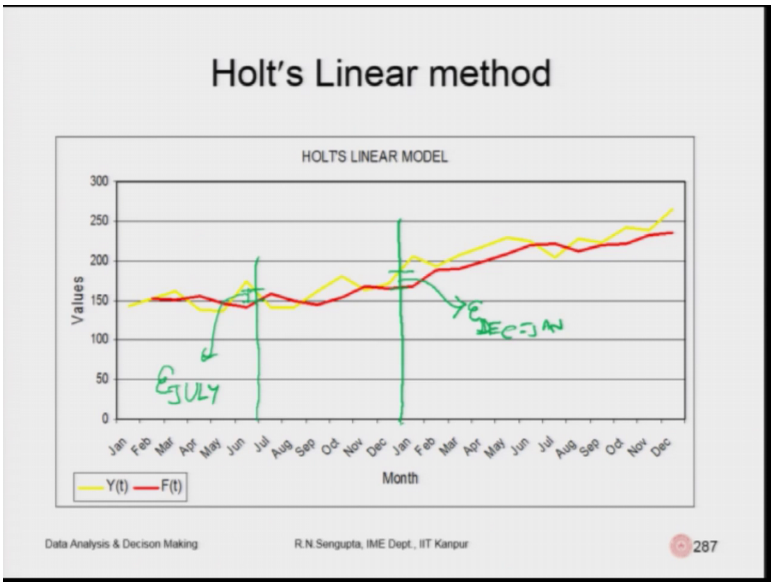
Month	Y(t)	L(t)	b(t)	F(t)
Jan	143.0	143.0	9.0	
Feb	152.0	143.0	8.4	152.0
Mar	161.0	147.9	8.1	151.4
Apr	139.0	139.4	6.9	156.0
May	137.0	134.7	6.1	146.3
Jun	174.0	151.4	6.8	140.8
Jul	142.0	143.3	5.8	158.2
Aug	141.0	139.3	5.1	149.0
Sep	162.0	148.1	5.3	144.3
Oct	180.0	161.4	5.9	153.5
Nov	164.0	159.8	5.4	167.3
Dec	171.0	162.7	5.2	165.1
Jan	206.0	181.8	6.2	167.9
Feb	193.0	184.3	5.9	188.0
Mar	207.0	192.7	6.1	190.3
Apr	218.0	202.3	6.4	198.8
May	229.0	212.5	6.6	208.7
Jun	225.0	215.5	6.4	219.2
Jul	204.0	206.5	5.3	221.8
Aug	227.0	214.2	5.4	211.8
Sep	223.0	215.9	5.2	219.6
Oct	242.0	226.4	5.6	221.0
Nov	239.0	229.9	5.4	231.9
Dec	266.0	245.3	6.1	235.3

So, the Holt linear method we consider January to December the Y t values are given at the second column, the L t values we will calculate based on this calculation, but the 143 which you are considering which I will highlight now using a red this 143 and 9 are the values which you are assuming for the starting; they can change they can be changed based on the fact that I am trying basically trying to find out the minimum of the sum of the squares.

So, as we proceed the values are for the L t values are 153 till 245 and the b value basically change because they are dependent on time they become 8.4, 8.1, 8.9. The detailed calculation which I have shown in another table for other example they would exactly be replicated here also, there is no change.

So, the concepts what I repeated what I said are exactly the same only I am trying to basically give different set of data to basically prove or highlight the important points. And then the F t values with a forecasted value are given in the last column; so, difference of these errors which is Y t and F t would give me the error; sum square them up error square and as the error square is found out find the sum of the errors, minimize with respect to alpha and beta and your problem is solved.

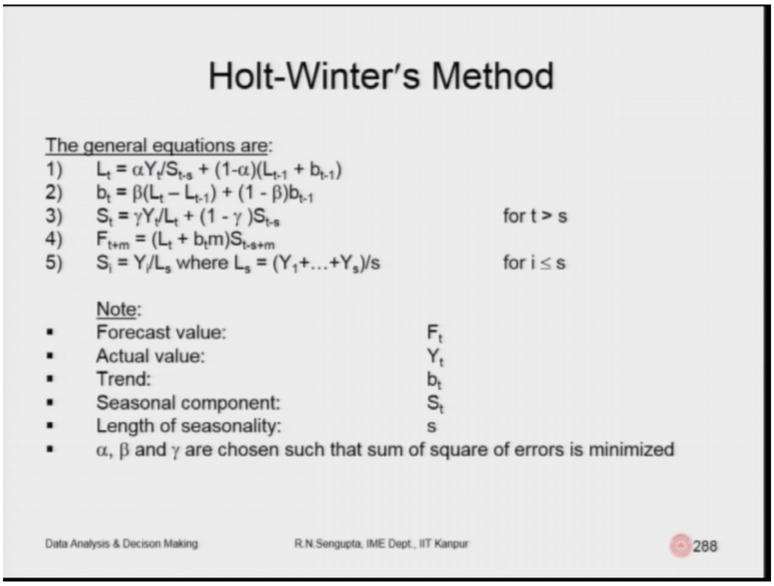
(Refer Slide Time: 22:54)



So, the Holt linear method I just plot for the values of alpha and beta which I have taken the yellow one and the red one and lines for the actual value and the and the predicted value; actual value for this set of problems is basically denoted by Y and the predicted value is basically denoted by F, suffix t b basically b being for the time period.

So, if you see the differences is; so if I consider January. So, consider this straight line; so this difference is basically the error for middle of January December. If I consider let we consider this if possible I think you should be possible; so, this is the error. So, this is the errors we are trying to basically find out and square them up and sum them up.

(Refer Slide Time: 24:06)



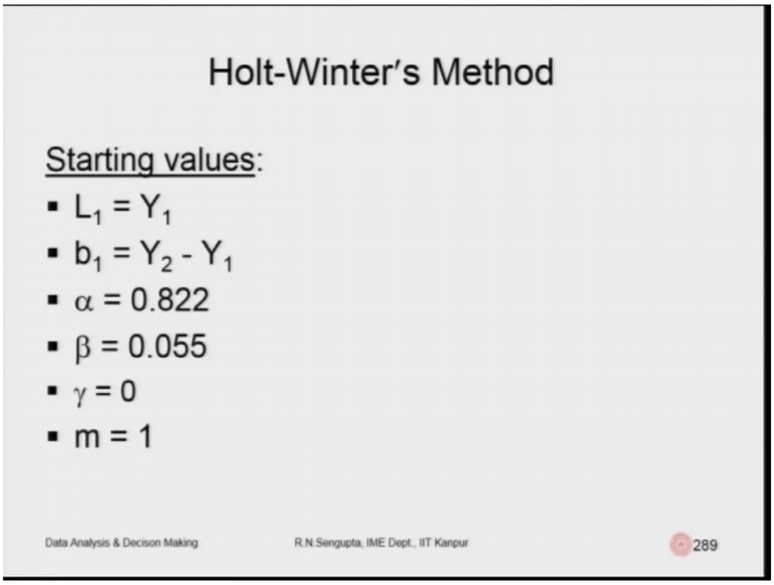
Now, in the Holt Winter's method in this case let me first discuss the nomenclature. So, the forecast value is there the actual value is there which has basically F t and Y t the trend is given by b t. So, the trend of the overall concept of the in which direction is moving positive or negative; the seasonal component would also come.

So, there are now 2 factor seasonality and trend. So, the length of the seasonality it will be considered at s and to what level you are going to do the prediction for the last time period was given by m if you remember. So, now, there would be 3 factors one would be alpha for the weights for the actual values point 1, point number 2 will be the beta would be the weights for the trend and point number 3 gamma would be the weights for the seasonality.

So, what you can do technically is that find out the trend minimize with respect to the gamma and find out gamma which is best. Once that is found out basically then you go to the values if it is step by step; then you basically go to the trend find out the errors, minimize with respect sound the squares with respect to the betas and find out the beta hats and finally, you find out the difference between the errors being actual value and the predicted value, square them up sum them up differentiate with respect to alphas.

Because you already done it for the differentiation of gamma and beta; then once you differentiate this with the sum of the squares with respect to alpha you put it to 0 and solve the problems accordingly.

(Refer Slide Time: 25:48)



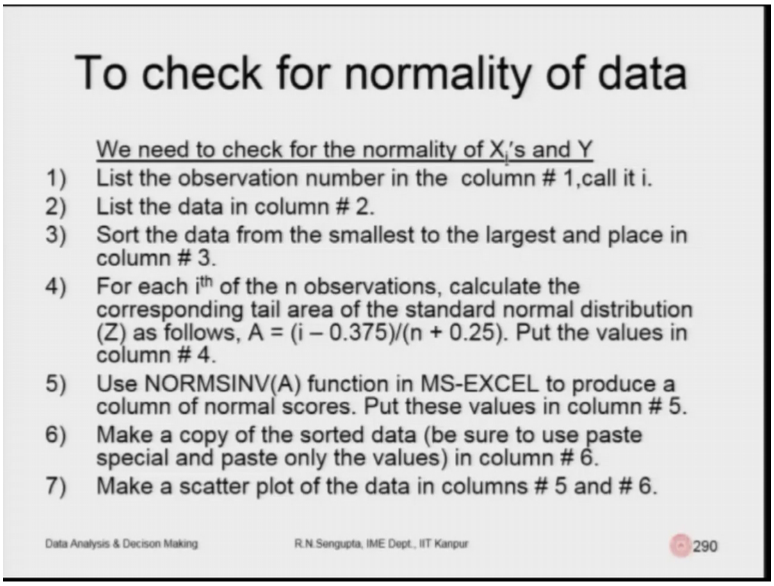
For the Holt Winter's method we consider L 1 and Y 1 as equal because that is the starting that half we have to consider. Then we consider the there are 2 things; so, the trend we will consider the Y 2 and Y 1, the actual values difference between time period 1 and time period 2 are equal such that b 1 is 0.

We considered a value of alpha as 0.822, beta as 0.055 and gamma value of 0 so; that means, we are considering weights being of 82 percent. The trend component being for about 5.5 percent and the seasonality component being of 0 percent; these are the initial values I am comes considering m as 1; that means, based on the fact that I am going to predict on the trend and front s is also 1; intrinsically which means the seasonality I am also considering as 1.

Here we are considering seasonality as 1 time period, but in say for example, for different type of products like steel prices, coal prices they would be seasonality the seasonality which will be happening. Consider that for different type of woolen garments; you have a seasonality of 12 months. So, in that case s would be 12 in case you have to actual index based on the index means the what is the nomenclature based on which the time period have being calculated. If they are 1 month in that case s is 12; in case it is basically 1 year in that case s is 1 because that unit would basically dictate what is the value of s and m.

Now, we need to basically check for normality of data. So, this is very important and we will basically consider the concept of the q q plots. So, q q plots would be considered in the next class, but initially let me start the discussion. So, we need to check the normality of the curves.

(Refer Slide Time: 27:48)



So, what you will do in excel sheet basically you list the observation numbers in column 1 column call it as i. So, I can be 1 2 3 4 till whatever values you have and list the data in column 2.

So, in column one you have thus index number serial number and column 2 you have basically the data. Now we will sort out the data which is basically in column 2 and put it separately and column 3 from the smallest to the highest; that means, you have rank them. So, as you rank them you find out the probability and basically your main essence is trying to plot it for each ith of this n observations calculate the corresponding tail area which is basically the overall area of coverage for the probability perspective from minus infinity to that point.

So, you are trying to basically integrate, but in this case as integration is not possible and then sum them up because they are a discrete points they are not discrete points; I should not use the word discrete points is basically the number of points are limited you are basically forming a sample. So, corresponding to the fact that A value is equal to i minus 0.375; why it is I will come to that divided by n plus 0.25, you put the calculate the values put in column 4.

Use norm inverse function from excel and basically put these values in column number 5 based on a column 4. Make a copy of the sorted data and basically use and try to put in column 6 and this basically try to plot the column 5 and 6. What is to be done? I will come into details when I basically go through the concept of normality plots and the q q plots in the next class. With this I will end this class and have a nice day.

Thank you very much for your attention.