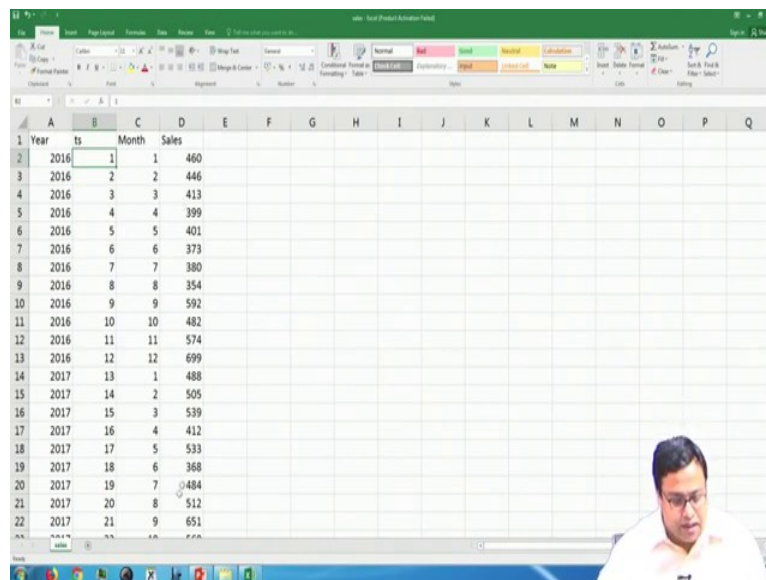


Marketing Analytics
Professor Swagato Chatterjee
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur
Lecture 20
Demand Forecasting and Pricing (Contd.)

Hello everybody. Welcome to Marketing Analytics course. This is Doctor Swagato Chatterjee from Vinod Gupta School of Management, IIT Kharagpur who is taking this course. We are in week four and we are discussing demand forecasting and pricing. So still now we have talked about a little bit of pricing and we have talked about demand functions. And now in this particular class I will discuss about how to use analytics to forecast a demand.

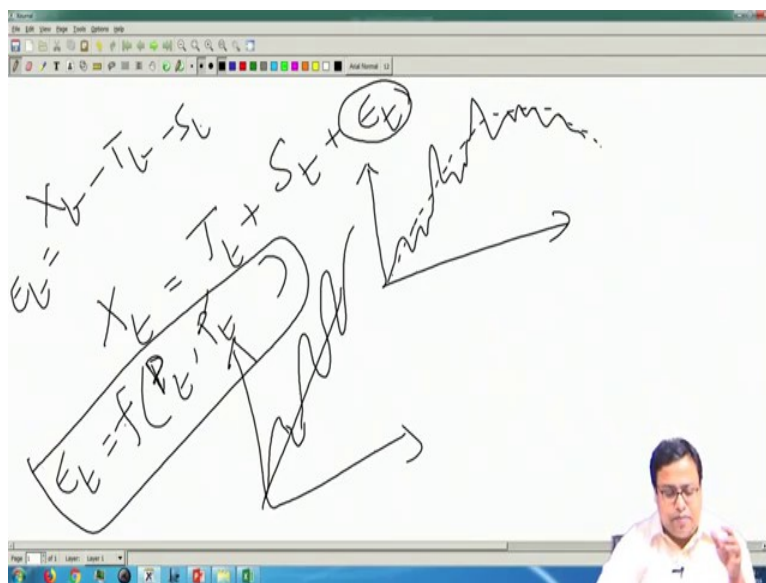
(Refer Slide Time: 0:58)



Year	ts	Month	Sales
2016	1	1	460
2016	2	2	446
2016	3	3	413
2016	4	4	399
2016	5	5	401
2016	6	6	373
2016	7	7	380
2016	8	8	354
2016	9	9	592
2016	10	10	482
2016	11	11	574
2016	12	12	699
2017	13	1	488
2017	14	2	505
2017	15	3	539
2017	16	4	412
2017	17	5	533
2017	18	6	368
2017	19	7	484
2017	20	8	512
2017	21	9	651

So in your data set, you will have a file called computer, I would say Sales. Excel file. So, Sales. Excel file will have a data set which looks like this, which has if you carefully see that it has year component and then the timestamp. Time stamp means whether, it is not timestamp actually, it is a time period component. And the month and the sales, so for 2016, from 2016 till 2019 I have 4 years of data. And I have monthly data of sales of something. And the timestamp is nothing but the month-year combination of that. So, as TS increases, you go, come closer to the current thing. And now our job is to try to estimate what this particular thing is, how I can forecast the next demand.

(Refer Slide Time: 01:42)



So what happens is, in real life situation, what happens is that if you have X_t , X_t is considered to have three components. $X_t = T_t + S_t + E_t$. So there are lots of ways of actually estimating a time series. But here we will be focusing on the simplest way, what we will break this particular thing into three parts, one is the trend component, one is the seasonality component and one is the error component.

So trend, what is trend component? So, let us say, if I have a time series which looks like this, if this is the time series, you can easily see that I can plot a middle value which goes like this, the dotted one. So this is the trend, the overall, the overall function I would say function of this particular thing whether it is going up, going down, going up and then coming down, or something like that. That is trend. And then, so one a little bit of less undulations are there. And then seasonality is more related to certain patterns which come back after certain period of time.

For example, let us say, the sales of a cold beverage is going up, why? Because population is going up and if population is going up and their purchasing power is going up, then the sales of cold beverage will go up over years. So that is something that is a positive trend. But in the summer season, the sales goes up much higher than the winter season. So, you will see that within that going up, within the sales going up like this, there will be a jump in the summer season and little bit dip in the winter season and again a jump and dip, again a jump and dip and so on.

So, that part is called seasonality which, where the pattern actually repeats itself over certain period of time. So when this particular seasonality happens for a very large period of time, we call it cycle which is generally we do not use to forecast because forecasting is generally done for a shorter period of time and then the trend is like this. The rest of the part is error. Now, this error part is something which your time series cannot explain.

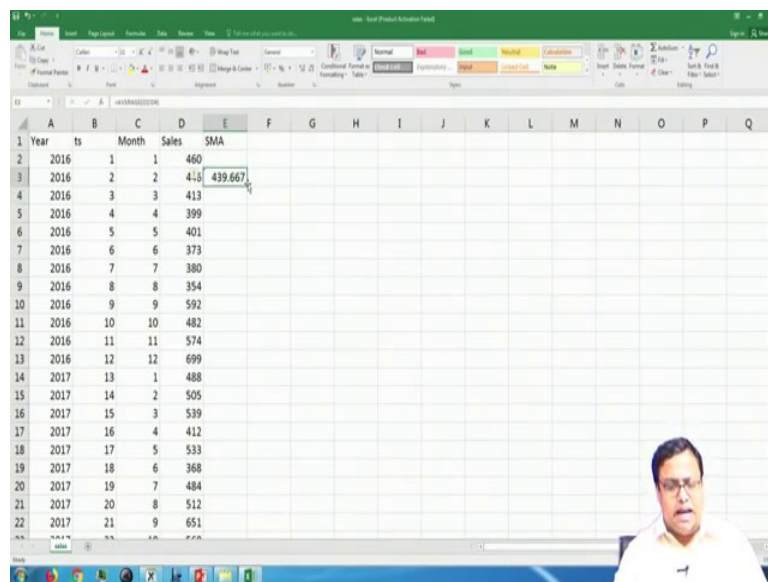
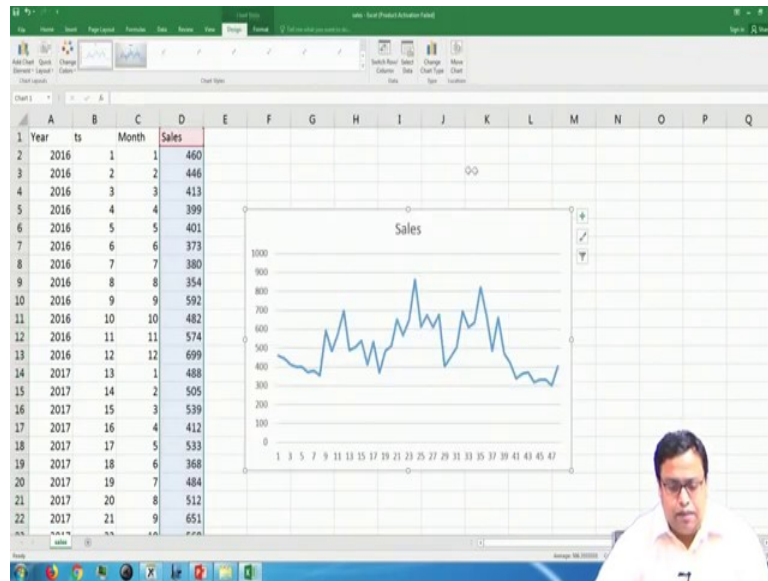
What is time series cannot explain, what does that mean? That means that for example, let us say, if I ask you that how much will you score in a math exams which will come in the next two months, the best bet of your math's exam score that will happen in last two months, in the next two months is whatever you scored in the last two months probably. So what I am trying to say is that your recent past performance of math's is the closest predictor of your recent future performance.

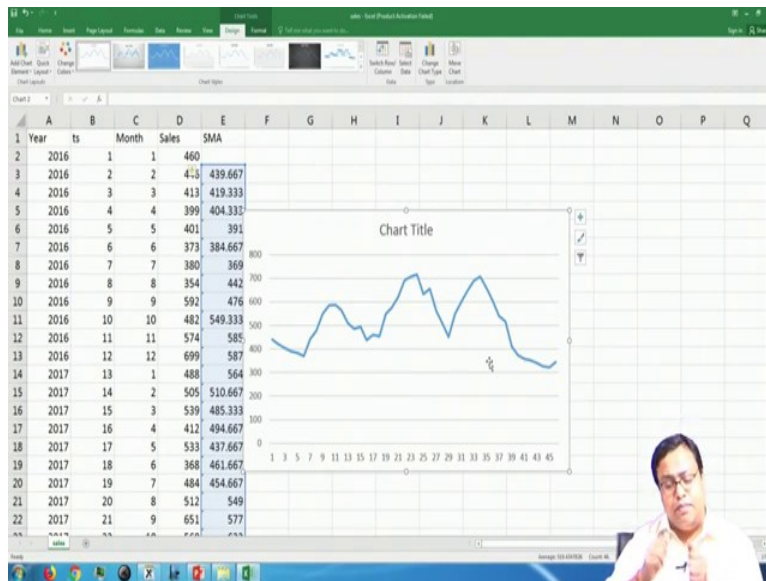
So similarly, for any time series there are some part of the time series is past which explains what will happen in the future. So that is the part where we do time series analysis. The trend is how it actually talks about, how the past of this particular observation is going to affect the future. And seasonality is also related to how the past, not exactly past, how the time will impact the changes in the x variable.

But other than my past, other than my time factor, there can be so many other factors which might, which are cross-sectional, which is at that time period, which might impact my performance. For example, let us say, in case of your score in math's, probably your health whether you are okay on that particular day or not will actually impact your performance. So, that is the component, all those components are stored in E_t .

So, if there is no such component, if absolutely there is no such component, then this E_t will have a mean value of 0 and absolutely, but if by chance that if there is certain things which might impact then you have to find out $E_t = X_t - T_t - S_t$. And then you have to find out E_t as a function of various stuff. For example, let us say in this context price, so price at t^{th} time period, price of the product and price of the competing product, P^t . So, all of these things might impact and this is where the demand function actually comes in $[E_t = f(P_t, P)]$

(Refer Slide Time: 06:25)





So you, if you get a trend and seasonality, let us say from this data that we have is a sales data. Now, this sales has happened because of multiple reasons. Something is actually related to its past sales, so if I remove that past sale component, the whatever effect of past sales on present sales, remove that component and if I also remove the seasonality component that okay, in certain seasons the production goes up, that is why the sales goes up. I will actually remove those part and then I will keep the rest.

But the time factor is not anymore affecting. I will get some error. Now, by chance if I also have some price data here, for every time period price of the product and price of the competing product, so price let us say something like this. So, if I have this data, these two columns, then what I will do? I will try to find out how the error part whatever is remaining is related to this price. So, I will do a regression on the error with price and price dash as my independent variables, and I will try to create a demand function out of it.

So, before I do that, I have to do the, find out how to remove the trend and seasonality component of a data set. So that is something that is more important to do. So what I am going to do here is the first job is to try to plot in. And the plot looks like this. So, I can easily see at least in this particular plot that you can see that, so there is a pattern that, it is 8 let us say, and this one is around 18. So 8 to 18, 11 is something that there are coming back.

Now this is 18. I will explain if the, this is 29, so then here also you will see that the jump is 11. And then if it is 29, then this one is 40, 41. So I can say that 11 or 12 will be something where the pattern is repeating itself. Now, then you should not only see the picture, you should also focus on the common sense. So, it is monthly data. So probably, it is not 11, 12 is

the, so it is only because of these 3 years I got 11 but probably it is a 12 number of months actually which create the seasonality factor.

So what I will do is I will know that, so this guy has a seasonality of 12 months, something like that. So one job, one good job is to do that, so I will create first thing is I will try to find out the trend. So the easy way to find out the trend is to do a smoothening. Now, you can also see that 7, here it is 8 and here it is I think 10. And then here it is 13 and then this one is probably 16. And this one is 18, so 2, 3, 3, 2. And then this one is let us say 18 to 22, 4.

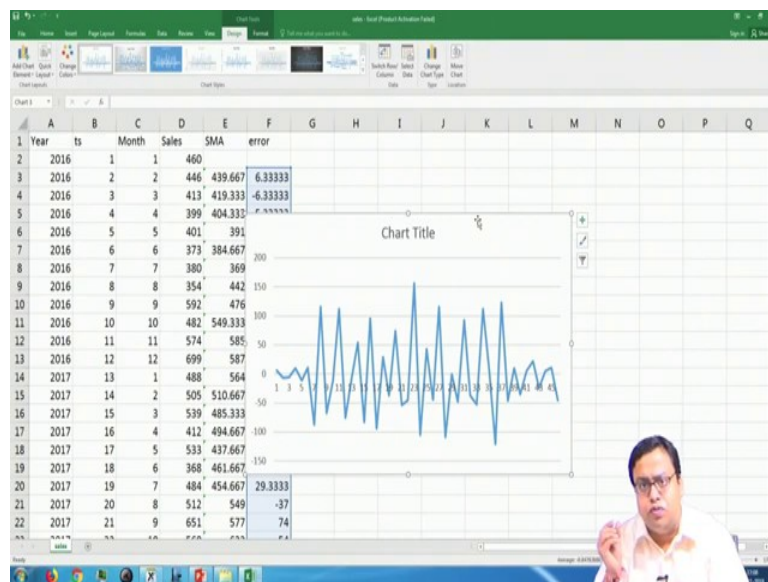
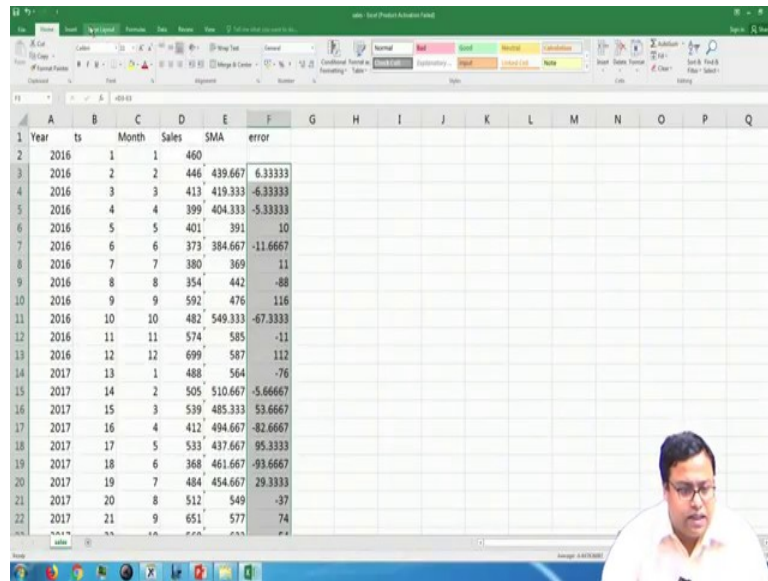
And then this one is probably another 22 to 325, so average 3, in average 3 there is a small jump and coming down, small jump and coming down, small jump and coming down. So, I can also think about that as my seasonality component or something like that. So keeping that in mind, I will be finding out the trend. The easy of finding out the trend is something called simple moving average method.

What is simple moving average? Simple moving average is saying that, average of these three terms. So average of one point earlier, now and one point later, something like that. Now, here I have chosen the timer period for the simple moving average is 3. You can choose as 5 or 7. By chance if you do choose even number, then you have to do double smoothing. Two times you have to do the smoothing. So, once with the even number and the second time with 2 as your time period.

But if you do odd smoothing, then you are more or less safe. Then that it is centralized. So I will go ahead with doing with odd number. So 3 is the thing that we chose by seeing that curve, so 3. So because there is no past data for this particular sale, this is remains blank and because there is no past data for this particular sale, this also remains blank. So these are the two values that remains blank.

Now, if I just try to plot this table, whatever I created. The simple moving average table if I try to plot, you will see that it is a little bit of smoothened version, the undulations has come down. But this looks very-very beautiful. That it goes up and comes down, goes up and comes down, goes up and comes down and in the SMA itself, in the, I would say in the moving average itself, there is a seasonality component. That is there, that is number one.

(Refer Slide Time: 11:26)



Now if I find out the error, so how much did I do mistake, how much I could not explain? That is nothing but this - this, something like this and then I, if I just drag it up, it is like this. Now if I just try to plot this part which is the part which I could not explain, it looks like this. So this is very, very I would say peculiar. So, if you see that this particular guy has a, it is more or less the average value is 0, more or less it has come down to the average value, the trend value, the trend component, the increasing or decreasing component it is not there.

But there is a seasonality component that is there. And I think the seasonality size is almost 3, So, every three time periods there is a jump and coming down. So you will see that this peak, this lower is 0.7 and this lower is 0.9 and this lower is 12. And then this lower is 15 and this

one is 17, this one is 19. So, almost 2 or 3, something like that, this one is 21. And then this one is probably 24, so 2 or 3 is something where things are repeating itself. This is zig zig, zig zig zig. So, I can take also 2 or 3 depending on the situation.

So, let us go ahead because there are lots of, so for example, 24, this is 26. This is 28 and this is 30. This one is 33. So I can take also 2 probably as my in the picking up. So how much will be the every jump? So let us do that. Now there are two jumps. One is to find out the seasonality component and another thing is to find out the SMA, how I can estimate this SMA.

(Refer Slide Time: 13:25)

Year	ts	Month	Sales	SMA	error
2016	1	1	460		
2016	2	2	446	439.667	6.33333
2016	3	3	413	419.333	-6.33333
2016	4	4	399	404.333	-5.33333
2016	5	5	401	391	10
2016	6	6	373	384.667	-11.6667
2016	7	7	380	369	11
2016	8	8	354	442	-88
2016	9	9	592	476	116
2016	10	10	482	549.333	-67.3333
2016	11	11	574	585	-11
2016	12	12	699	587	112
2017	13	1	488	564	-76
2017	14	2	505	510.667	-5.66667
2017	15	3	539	485.333	53.6667
2017	16	4	412	494.667	-82.6667
2017	17	5	533	437.667	95.3333
2017	18	6	368	461.667	-93.6667
2017	19	7	484	454.667	29.3333
2017	20	8	512	549	-37
2017	21	9	651	577	74

Year	ts	Month	SMA	predicted error	rmse
2016	1	1			628.557
2016	2	2	439.6667	530.4997	8250.649
2016	3	3	419.3333	519.3331	9999.947
2016	4	4	404.3333	468.2498	4085.31
2016	5	5	391	428.4164	1399.99
2016	6	6	384.6667	414.0832	865.3348
2016	7	7	369	430.9998	3843.977
2016	8	8	442	483.4998	1722.229
2016	9	9	476	506.9164	955.8241
2016	10	10	549.3333	545.583	14.06469
2016	11	11	585	582.9164	4.341235
2016	12	12	587	650.7793	4067.8
2017	13	1	564	628.557	4167.61
2017	14	2	510.6667	530.4997	393.3511
2017	15	3	485.3333	519.3331	1155.982
2017	16	4	494.6667	468.2498	697.8528
2017	17	5	437.6667	428.4164	85.56678
2017	18	6	461.6667	414.0832	2264.183
2017	19	7	454.6667	430.9998	560.12
2017	20	8	549	483.4998	4290.283

Year	ts	Month	SMA	predicted error
2016	1	1		
2016	2	2	439.6667	530.4997
2016	3	3	419.3333	519.3331
2016	4	4	404.3333	468.2498
2016	5	5	391	428.4164
2016	6	6	384.6667	414.0832
2016	7	7	369	430.9998
2016	8	8	442	483.4998
2016	9	9	476	506.9164
2016	10	10	549.3333	545.583
2016	11	11	585	582.9164
2016	12	12	587	650.7793
2017	13	1	564	628.557
2017	14	2	510.6667	530.4997
2017	15	3	485.3333	519.3331
2017	16	4	494.6667	468.2498
2017	17	5	437.6667	428.4164
2017	18	6	461.6667	414.0832
2017	19	7	454.6667	430.9998
2017	20	8	549	483.4998

Year	ts	SMA	predicted error
2019	41	374.3333	428.4164
2019	42	358.6667	414.0832
2019	43	351.3333	430.9998
2019	44	341	483.4998
2019	45	327.6667	506.9164
2019	46	322	545.583
2019	47	345.3333	582.9164
2019	48	650.7793	
	1		628.557
	2		530.4997
	3		519.3331
	4		468.2498
	5		428.4164
	6		414.0832
	7		430.9998
	8		483.4998
	9		506.9164
	10		545.583
	11		582.9164
	12		650.7793

Now carefully see, if I do the first job first that if I try to create a function which will define what this SMA is, I will pick up this time series, two time series components. So timestamp, or let us say I will pick up the whole thing here and I am pasting it as special. And if I just zoom it up and I remove this, because this is the part that I do not focus on, this is the thing that I am interested in.

So I can do a regression on this one. I have to find out a linear equation, so I can find out, do a regression or I can easily just simply do a time series component. So, let us do a regression. And I say that okay, TS is 1 to 48, and timestamp is a numeric variable. And one job of this thing is to do a let us say, if there are 12 time periods, so 1, 2 and then 10, 12. So I will put a value of 1 here, all of these values are 1 let us say.

And I will see that okay, so if the predicted value here, predicted value here is if this value is 1, so if this, I will actually Vlookup this month value, the small table and I will put F4 because I will drag later and then 2 comma false. So you take from the second column and I will drag it up. So all 1's because they are, now what is the error part? The error part, how much is error remaining after this thing? Error is nothing but basically, or error square is this - this square. That is the error square that I got. And what is RMSE?

RMSE is equal to basically square root of mean of or average of, square root of average of this particular column. So I got the RMS. Now I have to find out such values here such that this minimizes. So what I will do, I will go to the data solver, I will minimize i15 by changing what? By changing these values and these values now can be negative as well. So not positive and solve.

And the solution is giving me something and carefully see what is solution giving. So the solution is saying that, that this is the thing that will repeat itself. This is the thing that will repeat itself. And how the repetition looks like? So this particular table, so 1 to 24, this will go on repeating itself. And you will find out this thing. So if you remember, the table was, the actual SMA values were these, the actual SMA value. So, if I just draw two lines, this is the, actual series 1 is the red line which is the actual line.

And the red line is the one that I predicted and they are more or less same. They are pretty good in terms of the predictions. So, I can say that, so this particular thing can actually talk about the trend part. So this is something that is done that I can find out. I can also run if you want, you can run a regression equation with timestamp(TS)-12 as your variable. So, this - 12 or this is square as your, as your x variable and try to predict.

Now, this is how I am trying to predict the trend part, how the trend will look like. So if I, if you ask me that okay, what will be the trend component here? I will find out, okay, so this value is 12. Whatever value is corresponding to 12 put that. So the value is 650. So I will just drag it up for the next probably 12 and then again 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. So, next one year's data I know that this one will be the trend component.

So this is something which I know. Now, next component, now the problem with this model is that I am saying that the trend is same, fixed. There is no going up or going down. So sometimes we prefer, that is why I am saying that you can also run a regression equation,

regression equation will give you a movement but it will be slightly going up. So you can actually instead of doing this, you can actually do this one as well.

(Refer Slide Time: 18:09)

Year	ts	Month	SMA	predicted	error
2016	1	1			
2016	2	2	439.6667	530.4997	8250.649
2016	3	3	419.3333	519.3331	9999.947
2016	4	4	404.3333	468.2498	4085.31
2016	5	5	391	428.4164	1399.99
2016	6	6	384.6667	414.0832	865.3348
2016	7	7	369	430.9998	3843.977
2016	8	8	442	483.4998	1722.229
2016	9	9	476	506.9164	955.8241
2016	10	10	549.3333	545.583	14.06469
2016	11	11	585	582.9164	4.341235
2016	12	12	587	650.7793	4067.8
2017	13	1	564	628.557	4167.61
2017	14	2	510.6667	530.4997	393.3511
2017	15	3	485.3333	519.3331	1155.982
2017	16	4	494.6667	468.2498	697.8528
2017	17	5	437.6667	428.4164	85.56678
2017	18	6	461.6667	414.0832	2264.183
2017	19	7	454.6667	430.9998	560.12
2017	20	8	549	483.4998	4290.283

Year	ts	Month	SMA
2016	1	1	
2016	2	2	439.6667
2016	3	3	419.3333
2016	4	4	404.3333
2016	5	5	391
2016	6	6	384.6667
2016	7	7	369
2016	8	8	442
2016	9	9	476
2016	10	10	549.3333
2016	11	11	585
2016	12	12	587
2017	13	1	564
2017	14	2	510.6667
2017	15	3	485.3333
2017	16	4	494.6667
2017	17	5	437.6667
2017	18	6	461.6667
2017	19	7	454.6667
2017	20	8	549

So, you can probably delete and try to find out SMA as a function of this time series component. So if you remember it went up and then came down and then again went up and then came down and so on. So, I can use this month 1 and then timestamp both as my x variable, these two is my x variable and this is my y variable and try to predict that. And that will also give me some kind of solution, the problem is that here you have to treat these values, the month values as categorical variable.

So, you have to open it in R file, change this one to a factor variable and then run the regression. This is your y variable, this timestamp is your x variable which is continuous and month should be a categorical variable because month 1, month 2, month 3 are actually January, February, March. This has to be treated as factor variable. You can find out equation for your trend component.

(Refer Slide Time: 19:11)

Year	ts	Month	Sales	SMA	error
2016	1	1	460		
2016	2	2	446	439.667	6.33333
2016	3	3	413	419.333	-6.33333
2016	4	4	399	404.333	-5.33333
2016	5	5	401	391	10
2016	6	6	373	384.667	-11.6667
2016	7	7	380	369	11
2016	8	8	354	442	-88
2016	9	9	592	476	116
2016	10	10	482	549.333	-67.3333
2016	11	11	574	585	-11
2016	12	12	699	587	112
2017	13	1	488	564	-76
2017	14	2	505	510.667	-5.66667
2017	15	3	539	485.333	53.6667
2017	16	4	412	494.667	-82.6667
2017	17	5	533	437.667	95.3333
2017	18	6	368	461.667	-93.6667
2017	19	7	484	454.667	29.3333
2017	20	8	512	549	-37
2017	21	9	651	577	74

$$X_t = T_t + S_t + E_t$$

$$X_t = T_t + S_t + E_t$$

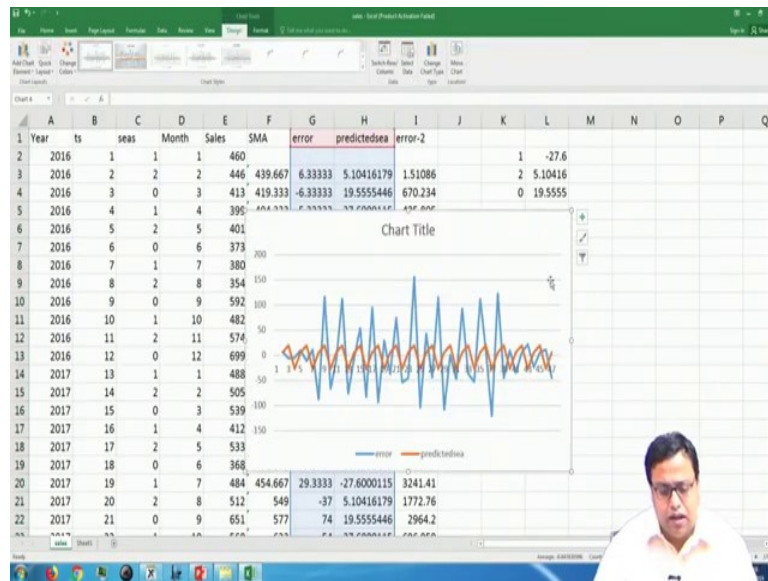
$$E_t = f(T_t, S_t)$$

Now what is to do about this one, the error part which is, now here what is this error part? The previous slide, the previous slide error part is we have actually already considered the....

(Refer Slide Time: 19:25)

Year	ts	seas	Month	Sales	SMA	error	predictedsea	error-2
2016	1	1	1	460				
2016	2	2	2	446	439.667	6.33333	5.10416179	1.51086
2016	3	0	3	413	419.333	-6.33333	19.5555446	670.234
2016	4	1	4	399	404.333	-5.33333	-27.6000115	495.805
2016	5	2	5	401	391	10	5.10416179	23.9692
2016	6	0	6	373	384.667	-11.6667	19.5555446	974.826
2016	7	1	7	380	369	11	-27.6000115	1489.96
2016	8	2	8	354	442	-88	5.10416179	8668.38
2016	9	0	9	592	476	116	19.5555446	9301.53
2016	10	1	10	482	549.333	-67.3333	-27.6000115	1578.74
2016	11	2	11	574	585	-11	5.10416179	259.344
2016	12	0	12	699	587	112	19.5555446	8545.98
2017	13	1	1	488	564	-76	-27.6000115	2342.56
2017	14	2	2	505	510.667	-5.66667	5.10416179	116.011
2017	15	0	3	539	485.333	53.6667	19.5555446	1163.57
2017	16	1	4	412	494.667	-82.6667	-27.6000115	3032.34
2017	17	2	5	533	437.667	95.3333	5.10416179	8141.3
2017	18	0	6	368	461.667	-93.6667	19.5555446	12819.3
2017	19	1	7	484	454.667	29.3333	-27.6000115	3241.41
2017	20	2	8	512	549	-37	5.10416179	1772.76
2017	21	0	9	651	577	74	19.5555446	2964.2

$$X_t = T_t + S_t + I_t$$
$$X_t - T_t = S_t + I_t$$



So we have considered as I told, X_t is equal to what? $X_t = T_t + S_t + E_t$. So what then? What is that? $X_t - T_t$ is absolutely seasonality and error part $X_t - T_t = S_t + E_t$. So this is the part that I will focus on. The seasonality part and the error part and I saw that the seasonality is either 2 months or 3 months or something like that. So I can do the same thing that I did earlier. I can say that okay, this is of 3 months. And the values are 1, 1, and 1. You can try out with two months also. I do not know which one will give better result.

The methodology is same. So I say that okay, Vlookup, Vlookup, basically I have to repeat this particular value every three times. So if I have to repeat this particular value every three months, I have to divide it by 3, this month I have to divide it by 3 and try to find out, so the timestamp, for example, this timestamp I will actually divide it by 3 and I will find out what is the remainder. So what was that?

So, remainder is probably, so if I am not wrong which is, I have to find out the remainder, mod. You can see, remainder after a number is divided. Mod of this by 3. So this, 3 basically. This comma 3. So that gives you the remainder. So see, 1, 2, 0; 1, 2, 0; 1, 2, 0 gets repeat. So this is the season component and I am saying 1, 2, 0, these are my three seasons. And I am saying that Vlookup, Vlookup of what? Vlookup of this particular value, the season value. For this one, it is this one is value and then what?

In the small table, and I have to put 4 because I will actually drag it and then 2, false. So that gives me the values and if I just drag everyone comes one. The last value will be not there. So this is my predicted seasonality, predicted season. And what means the error now? The error

now after seasonality, so error 2, error 2 is nothing but basically this, this is what? Is this, square it up, no, I have not squared it up.

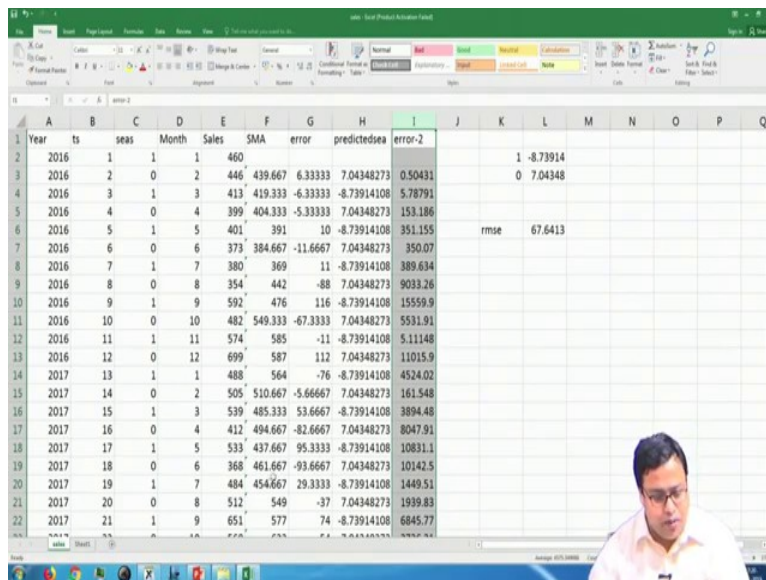
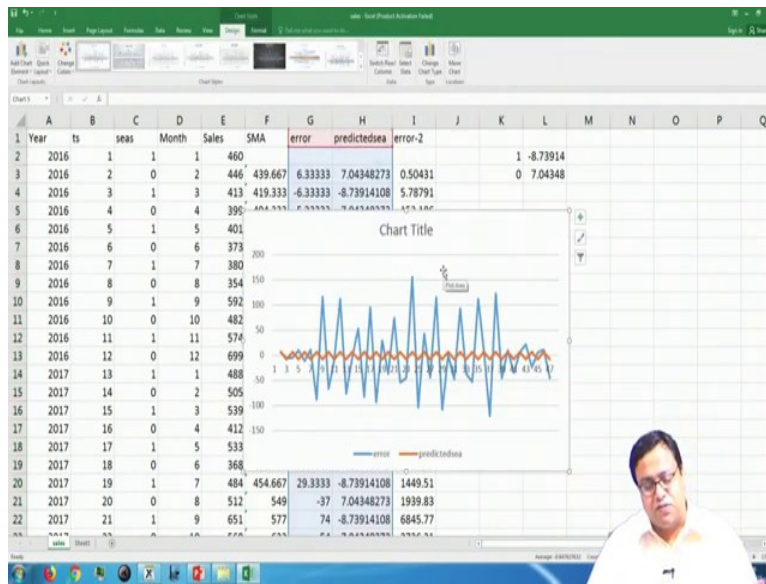
So basically, is equal to this error - this predicted square, square of that and then how much is the RMSE? The RMSE is nothing but the square root of average of this particular column. Not the 0, 1; not the last 0, 1. So, if I just do that and press an enter, I get the, so what is then? I have to again optimize this RMSE. How to optimize? I will go to data, click on solver, and the solver saying that this is my reference cell.

And which one I will change? I will change these three cells. And I unconstrained, so solve. And it gave me solution and it told me that this month should be the seasonality. 519-27, and this is the predicted values and this is the actual values. And now RMSE is coming as 65, you can also try out with two time periods and see that whether RMSE is coming up further lower.

And if I just try to predict, plot these two and see that how good it is, I will see that, not so good probably. I should have a better option than this one. So, I have to find out whether some other variable is predicting better or not. So, probably two time period or three time period. Two time period is something probably might work better. You have to check that.

(Refer Slide Time: 23:43)

Year	ts	seas	Month	Sales	SMA	error	predictedsea	error-2
2016	1	1	1	460				
2016	2	0	2	446	439.667	6.33333	7.04348273	0.50431
2016	3	1	3	413	419.333	-6.33333	-8.73914108	5.78791
2016	4	0	4	399	404.333	-5.33333	7.04348273	153.186
2016	5	1	5	401	391	10	-8.73914108	351.155
2016	6	0	6	373	384.667	-11.6667	7.04348273	350.07
2016	7	1	7	380	369	11	-8.73914108	389.634
2016	8	0	8	354	442	-88	7.04348273	9033.26
2016	9	1	9	592	476	116	-8.73914108	15559.9
2016	10	0	10	482	549.333	-67.3333	7.04348273	5531.91
2016	11	1	11	574	585	-11	-8.73914108	5.11148
2016	12	0	12	699	587	112	7.04348273	11015.9
2017	13	1	1	488	564	-76	-8.73914108	4524.02
2017	14	0	2	505	510.667	-5.66667	7.04348273	161.548
2017	15	1	3	539	485.333	53.6667	-8.73914108	3894.48
2017	16	0	4	412	494.667	-82.6667	7.04348273	8047.91
2017	17	1	5	533	437.667	95.3333	-8.73914108	10831.1
2017	18	0	6	368	461.667	-93.6667	7.04348273	10142.5
2017	19	1	7	484	454.667	29.3333	-8.73914108	1449.51
2017	20	0	8	512	549	-37	7.04348273	1939.83
2017	21	1	9	651	577	74	-8.73914108	6845.77

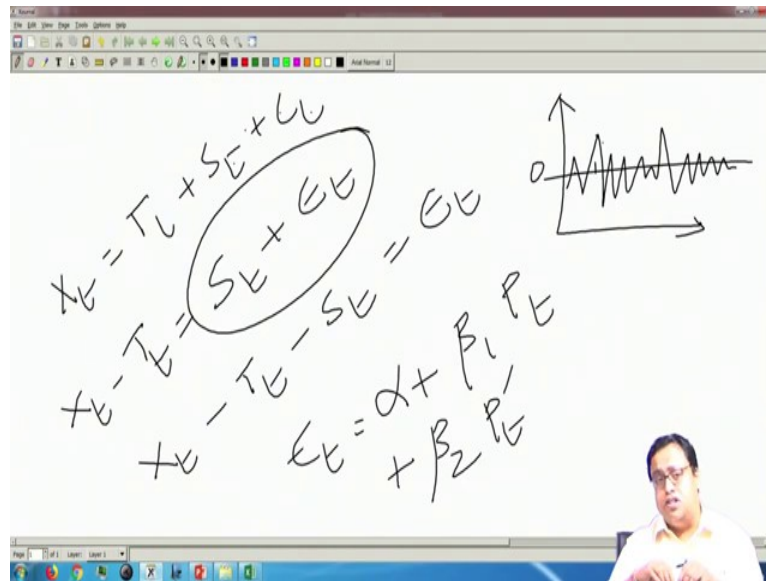


For example, how to do that two time period? The simple thing is if I just change it 2, then this one will be 1 and 0, so I remove this and then I put this as here, 1 and 0. And then what is the VlookUP of C3? In K2 to L3, 2 falls, so this part remains same. And then the error also remains same, RMSE remains same and then I will run the solver once more. 65, if you should remember that previous one was 65 or something like that.

Solver, solve it. So L2 to L3, solve. And now it went up actually. It did not work well. So I will go with the three time period only. And if I just see the picture here, the picture is coming up to be something like this, which is worse. So, previous one was better so something like this you can do to find out the undulations. Now, there can be other things. These undulations, they are one that we could not predict which is still remaining here.

The square of that is still remaining here, can be actually predicted by many other factors. Not everything is actually coming from your past. So you do not have to worry if there is something which is still left after removing out the seasonality component and trend component. That remaining part is something that you try to regress on price or let us say something else.

(Refer Slide Time: 25:18)



So what we do is in that case that we try to find out this that okay, so now I have found the seasonality component. So, I say that $X_t - T_t - S_t = E_t$. This should be, this should be something like, it should look like this. But it is, no trend component is there. And some, so absolutely this is 0 thing and it is actually distributed over 0.

Now what are these particular spikes? Why these spikes are coming? I will, what I will do is I will write $E_t = \alpha + \beta_1 P_t + \beta_2 P_t' + \dots$. And then if I have let us say advertisement expenditure, if I have price, if I have probably some other kind of promotions going on, all those variables will now come in. After the time series component has been taken care of, all those components will now come in to do the predictions.

There can be also some things where you can say that okay, so there were cases where we see that the sales actually is not only related to current pricing but also related to past pricing. For example, let us say during the festival season before the festival season or during the festival season, during the festival season prices goes up. Before the festival season they give huge discounts, 50 percent discount in pantaloons and let us say Big Bazaar and here and there. Lot

of huge huge discounts and people go and buy. Big billion day, after big billion day you will see for a very long period of time, probably 1 month or 2 month there will be a dip.

So the sales is very high in big billion day or etcetera. Will it actually impact the sales later because people has accumulated and stored lots of stuff and they have created their inventory and they will not buy? So you can say that the pricing at T-2, T-3 of time period is actually affecting your current sales, also the sales at that time actually affecting the current sale. And so those kind of factors are there that those are ergonomic issues that whether the price of two time periods back is dependent variable independent variable or the sales component is an independent variable.

But in general, if I actually remove the sales part, if I remove the seasonality part then I generally do not have an issue. There are other methods, there are other methods but here if you see I am only considering the past effect of three time periods at max. But there can be a scarce situation where the sales of the current time period is dependent on the sales of three time periods or two time periods back.

So we can also use concepts of ARIMA, ARIMAX is actually classic example. ARIMA part is the part where you take care of the component which is the time series component, you remove all of those stuff and then you also include the cross-sectional component in the x variables. So that is also something can be used in forecasting. We are not actually focusing on that part because that is something which will be, which itself is a course. The time series forecasting and etcetera itself is a, is probably 20 hours, 30 hours course and it will actually overshoot our this particular course's purpose.

So that is why we are not going to that but this is something that is very basic, anybody can use it in any context and can get fairly good result. So by saying that, I will be meeting you in the next video and we will talk about a little bit of a different kind of forecasting and different situation. We call it mass diffusion model where we will see in the next slide about it. Thank you very much for being with me.