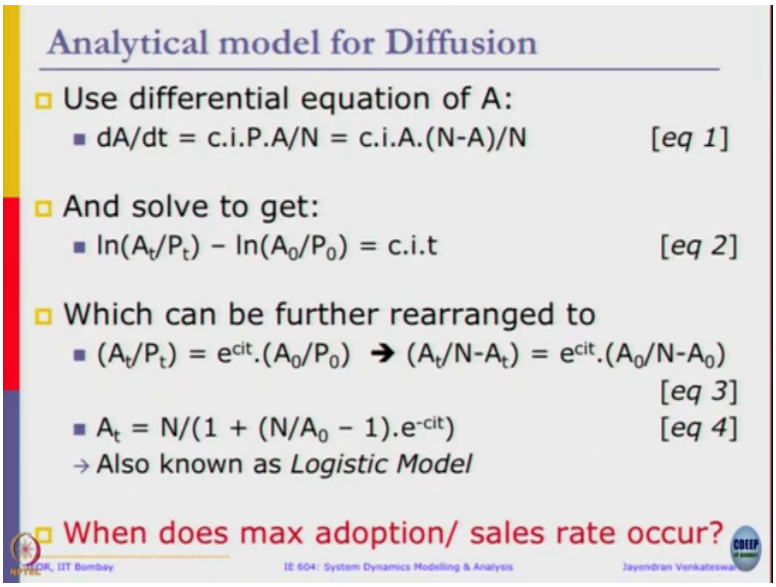
Introduction to System Dynamics Modeling
Prof. Jayendran Venkateswaran
Department of Industrial Engineering and Operations Research
Indian Institute of Technology, Bombay

Diffusion Model: Fitting Data Estimating Parameters
Lecture – 12.2
Diffusion Model and parameters estimation - II

So, we have seen these, use a differential, start a differential equation of A.

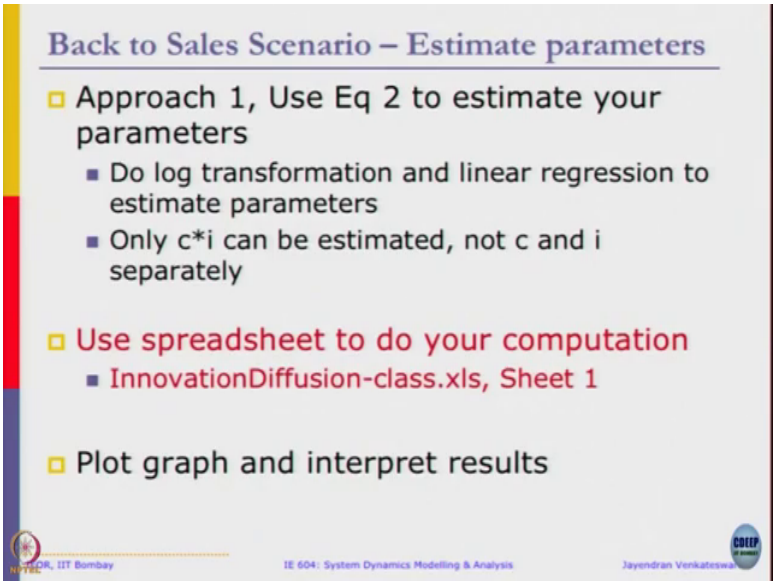
(Refer Slide Time: 00:19)



That is adoption rate A R is nothing, but d A by dt which is A adopters and we solved to get the equation in logarithm of A t by P t minus logarithm of A naught by P naught is c into i into t. Some we further rearranged together equation 3 as well as equation 4. This equation 4 is also called as logistic model a special case of which will be sigmoid curve. Since we have

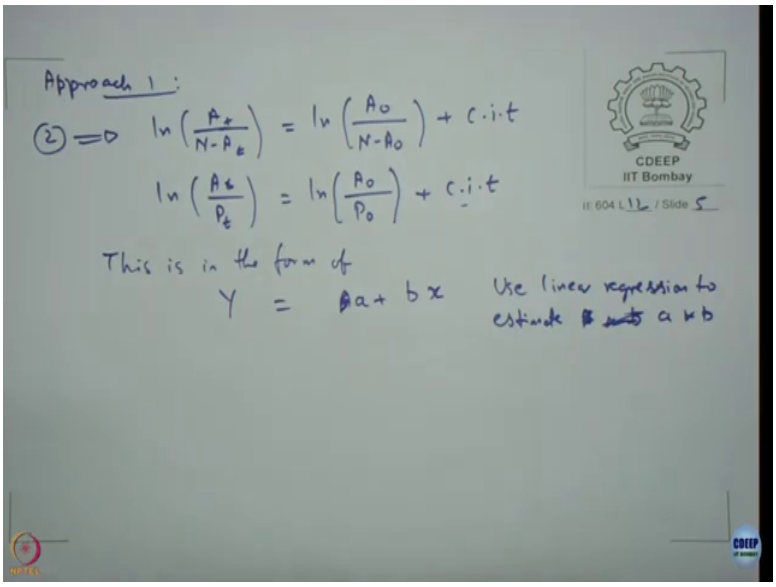
come across that the similar kind of model that we have seen when this max adoption or sales rate occur its figure that also.

(Refer Slide Time: 01:01)



Now, let us get back to example and you want estimate the parameters. The first approach is to use equation 2 to estimate the parameters. We are going to do log transformation and linear regression, estimate the parameters since only c into i can be estimated, c and i separately cannot be estimated. We do not have enough data for the that. So, we will just take the product and estimate it. So, this for we go to the spread sheet let us just quickly comeback to notes to see what is it that we are going to estimate.

(Refer Slide Time: 01:38)

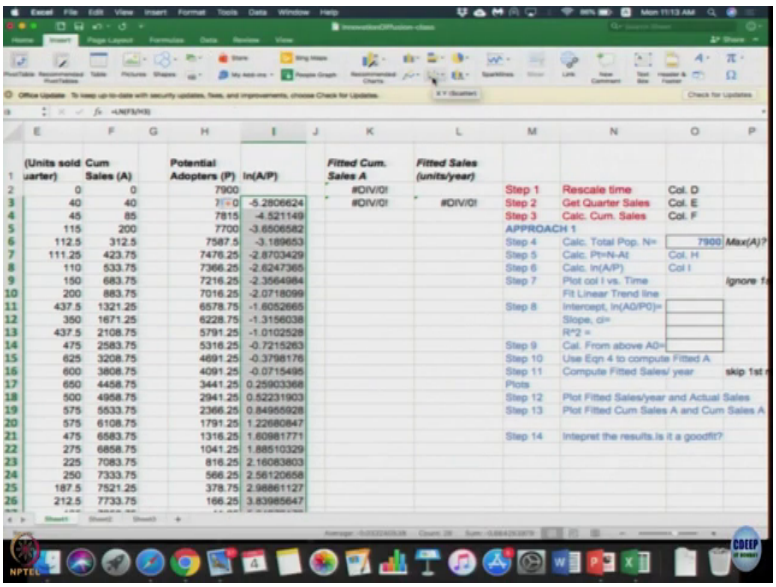


In approach 1, there is going to be estimate parameters. From equation number 2, we know that algorithm of A by N minus A t is equal to logarithm of A naught by N minus A naught plus c into i into t or logarithm of A t by P t is logarithm of A naught by P naught plus c into i into t. Again note that P t is N minus A t, right.

So, that is we have this. This is in linear regression form the form of Y equal to A plus b x. A is constant which is nothing, but this b is nothing, but c into t x is nothing, but your c into i rather b in to c into i x is your t, right. So, that we have. So, this is your Y variable, this is your A c into i is your b and t is your x and we can get the product. So if we have these two data values, we can plot a simple linear regression of it to find intercept and the slope right. That is precisely what we are going to do. This is also form then I am going to use linear regression the estimate a and b, ok.

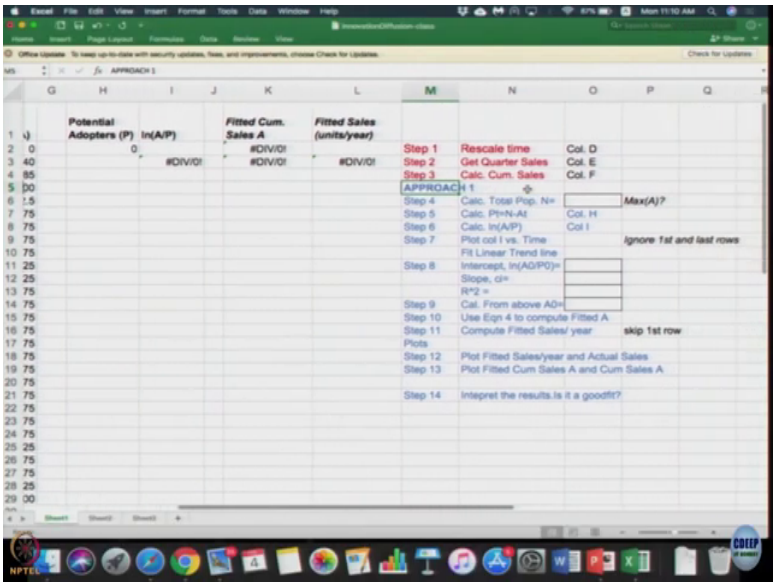
Sorry for confusion. Let us just call it small a. You already have capital A nus. So, just avoid confusion. So, then we will use regression do that lets precisely. What we are going to do, let us go back to a spread sheet, look after the spread sheet.

(Refer Slide Time: 04:10)



Before taking logarithm we need to figure out what is the potential adopters. So, for the first approach one we can see on the right side.

(Refer Slide Time: 04:20)

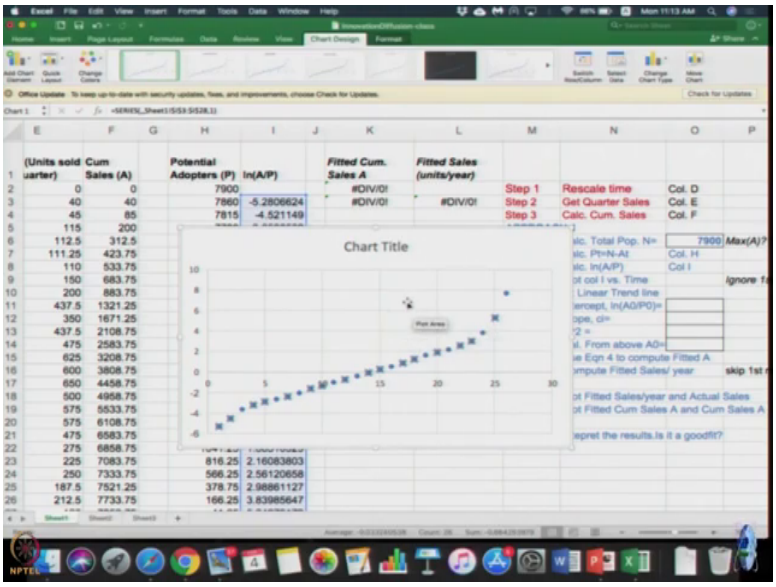


The your spread sheet that is from column M onwards first is step 4. Calculate total population M. So, we need to guess some value of M. So, for that you can just check what has been the total cumulative sales so far its say 7900. So that means there is total number of adopters that has been so any number greater than that is should be for sufficient. So, for basic purpose let us just assume that it is 7900 eqbal. So, we can just write the number as 7900 as a total population n.

Now, if you approach here to column H, we need to compute the P t. So, N minus A t the formula needs little update. Its dollar o dollar 6 minus F2 because total N does not change. So, update the column, update the formula minus the actual sales A t using that we can compute a value of P. So, do that drag it downwards till row 29. So, you should see that the value of P goes in the opposite direction as A which was expected. It goes on 790020.

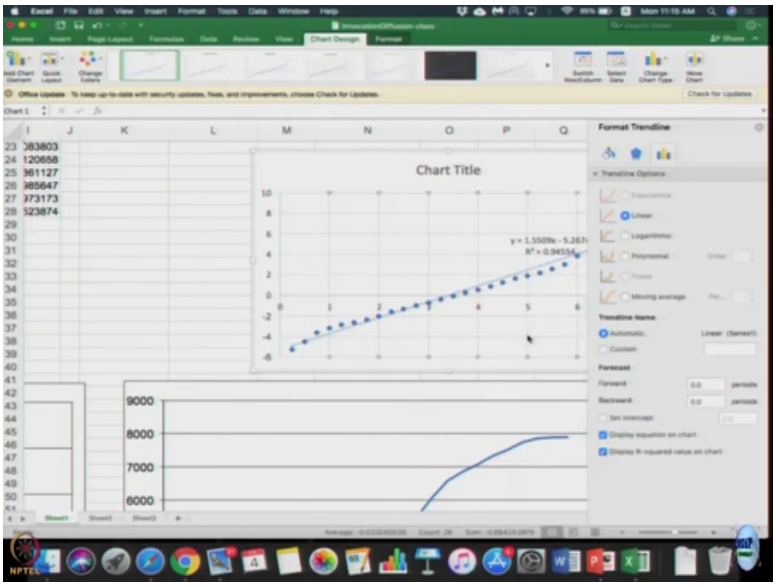
So, once you do that we need our Y column that is nothing, but logarithm of A by P. Sample has already been calculated for as logarithm of A divided by P. So, I am going to simply drag that column all the way to row number 28, have the step 6 in our approach step 6. Now step 7 is to plot column I verses time. So, to do that I hope all of you know how to plot it in excel. Select the column I click insert click scatter plot.

(Refer Slide Time: 07:08)



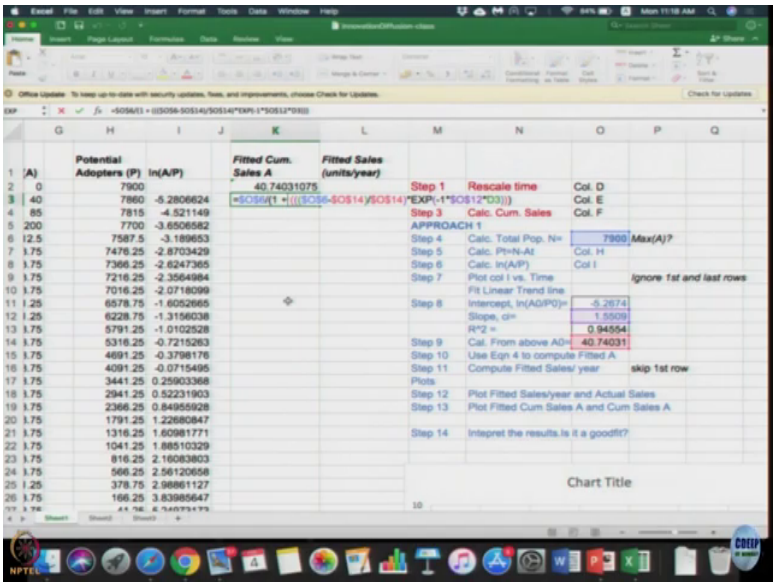
As soon as clicks scatter plot, we will get this right click, select data because you want to set the x axis, data x axis should be the time that series that we have got right by default. You just take 1 2 3 4 we just select the we have to series 1 select x values in the x values are 0.25 0.65 in column B. We will get a curve like this or a plot like this. We can move the plot down, sorry he does not hide any of the figure or text in the text we have and fit a regression.

(Refer Slide Time: 08:13)



It is quite simple and in excel you can just right click the data series or any of the dots right click it, click add trend line by default. It will be linear. Select linear display equation on chat display r square value on the chart and you are done. So, this graph you can see here we just selected both the columns R value as time is selected time as x axis and we selected logarithm of A by P as the y axis values we plotted them and we fit a regression line linear regression and you got A equation y equal to 1.5509 x minus 5.2674 with the r squared of 0.94. So, let us write these values.

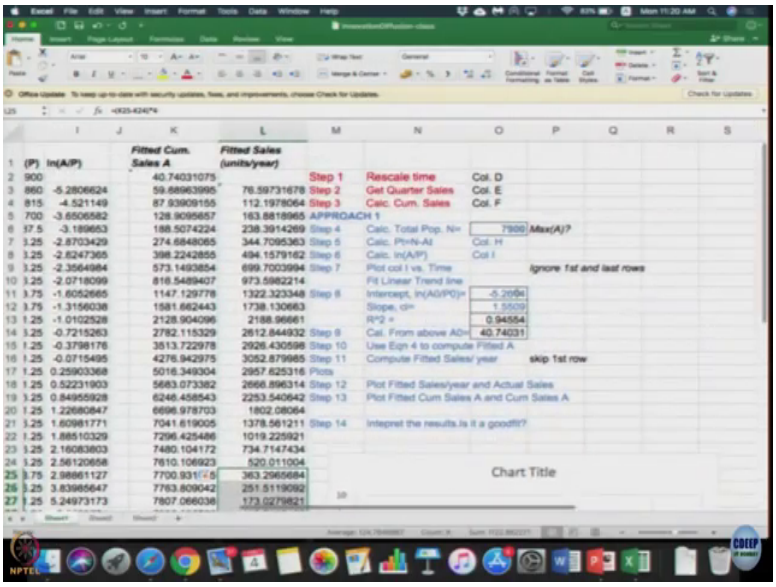
(Refer Slide Time: 09:22)



So, the intercept is minus 5.2674. This slope c i is 1.5509, R square is 0.94554. Even if you did not get it, you can copy the values that I am just typing here into the boxes that has been provided for you in column O step 8. So, I got the intercept slope as well as r squared values.

So, from this the intercept is nothing, but logarithm of A naught by P naught, right. So, we need to calculate A naught. How we will do that? Logarithm of A naught by P naught is equal to minus 5.267 and P naught value. What is initial value of P change? Go up and see initial value of P is 7900 that is also given. So, value of a naught will be e power intercept value into P naught which is I think, but total population n. So, I just wrote the equal to exp of O11 multiplied by O 6. So, A naught comes to 40.7 so, that is initial number of adopters say this is 40.7. As soon as to enter the values, you will see that in column K its first two rows are filled.

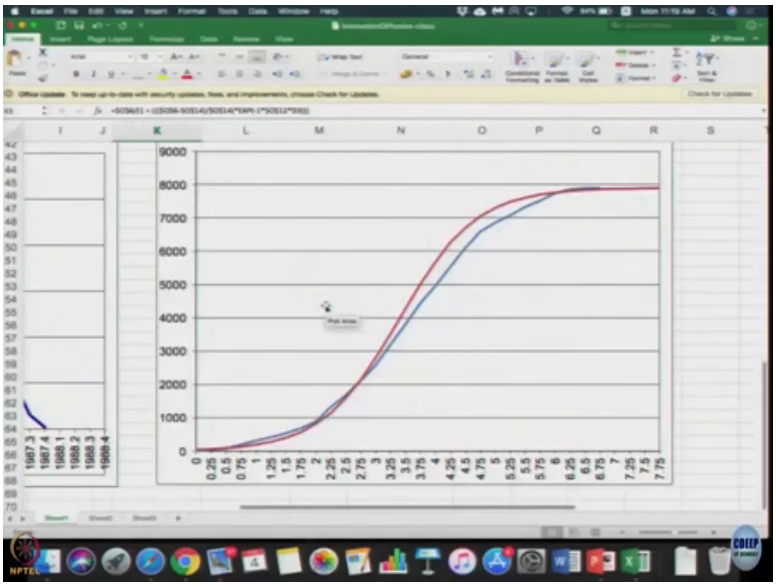
(Refer Slide Time: 11:20)



Now, we are moving to step 10. Step 10 says use the equation 4 to complete fitted A. Now I need to compute the sequence for A and for A we derived the equation 4 which gives a expiration for A which is nothing, but shown here in this formula N divided by 1 plus what is it N minus A naught by A naught into e power minus c into i into t which is what is coded for you.

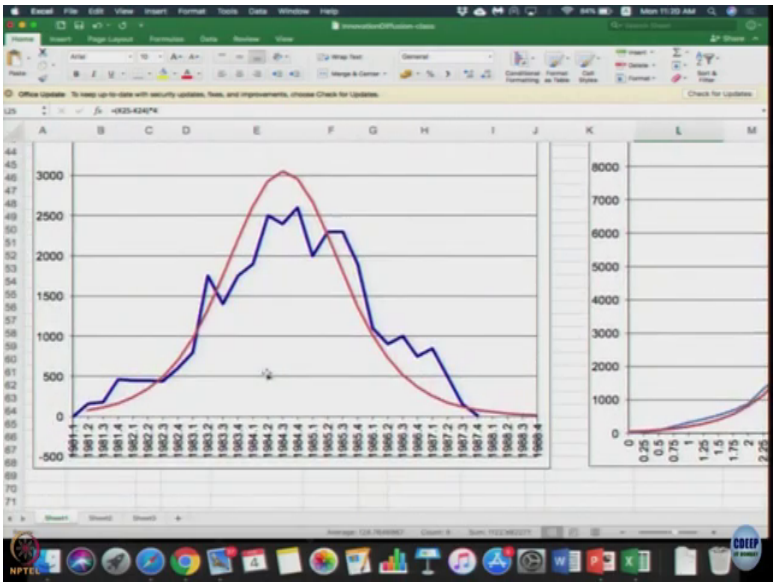
So, if you drag this since now this A t is just dependent on time t, you can drag it even beyond 6.5 because you are just now using an estimate. The values are actually doing a little bit of forecasting. So, you can drag the second row downwards ok. So, this gives our fitted cumulative sales A after completing column K you scroll down.

(Refer Slide Time: 12:42)



You will see two graphs; the blue line corresponds to the actual data and the red line corresponds to the fitted data. It is an approximate fit. So, let us see how the fitted data corresponds to the actual sales data that is step number 11, computed fitted sales per year.

(Refer Slide Time: 13:26)



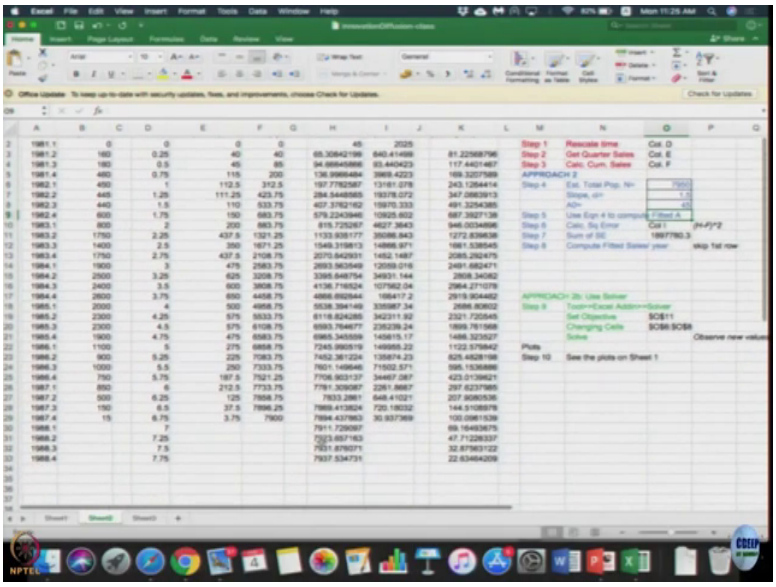
So, if you drag the formula for that all the way down someway beyond, so you see this graph that appears x axis. Here I am just all the formulas are prefilled. So, if you fill the basic values and drag it, you should get this curve. So, red curve gives the fitted curve which is kind of an approximation for this given data.

So to get this curve you need to fill these values in column O. You can take a quick look at it, you can fill these values and drag the formula. It should work for you pausing for a minute. So, you can see to enter it, intercept is minus 5.26 slope c i is 1.55, A naught is 40.7. Even if you do not know formula you can just type the number. Well let us interpret this graphs idea is not just fit some curve, it should be an accurate fit. You can observe the initially the model seems to be underestimating here its seems to a reasonable fit.

Peak it seems to overestimate that seems to overestimate the peak and again starts underestimating the tail or rather the second half of the curve it seems to underestimate heavily. Here it seems to overestimate it, here the result is this cumulative graph with seems to kind of underestimate in the starting period and then over estimate here and towards the end kind of reach it much earlier than this can do. This is figure we have. Instead of this, this is one approach. See this approcach is heavily depended on your initial value of population.

Instead of this initial value of population, we can use any other values. We can put any of the values of c and I you see what happened right or can we use it existing equation number 4 to directly solve this which is what we are going to do it approach 2. Then approach 2 you are directly going to use equation number 4 and you are going to choose parameters N A A naught c into i. So, in the least squared error is minimized you can do it in trial and error method. We do not need to convert it into a regression model trying to do a logarithmic regression and it is not that great we can try to see how we can do that. So, in a same excel spread sheet go to sheet number 2, ok.

(Refer Slide Time: 16:10)



Again the first up to column D is exactly the same as the old one we had. The sales data that we can convert it into B by B 2 by 4, drag the sales data the some values in column O. You will see it is already pre filled for you that is it is 7900 slope is 1.5, A naught is 40. We do not know how we got it. Assume we got it we can always fit this curve A t right which is what let us complete the cumulative sales on the fitted equation. We know the equation.

So, I can we simply drag the equation till end. I can compute this square error for it I am going to drag the fitted sales also along with that. Now we can see that if we are going to change this initial population, I am just going to focus on column O and if I change that initial population suppose in the 7900, I put 8000. If you will observe the some of squared error will change. You can just take any number 8 1 or 7 9 5 0 is square error changes and if A naught changes from 40 to 45, again squared error changes.

You just drag all the formulas that is written except for column W, where that should be the sales value divided by 4 not time divided by 4. See if I change those values I am going to get arbitrary fits what is it fit. I request you to go back to sheet 1 and if you scroll I just put everything on one graph. The green line so, this green line will keeps shifting for whatever values I am going to enter in the sheet number 2 but I wanted the best fit.

So, to do that we can actually use the optimization solver to fit it where I minimize the mean square error. So, to do that you need to go to tools excel, add on include the solver. So, if you have it then you will go to tools. So, you do not have it go to the excel, add on add the solver, go to solver, set the objective as 11. I have given what is to be entered by changing all these three values N c A and A naught. Again what is to be entered I have written it here ah. Right here I have entered it here.

What is to be entered and that is it you do not need to worry about any constraints. It solve it keep solver solution oops and maximize it. Sorry to solver minimize it yes minimize the means square error by change the cells, solve it ok. I need to set it as initial values slope is data solver ok. Once you solve it, in this case we got a estimated total population of 8075, the slope of 1.24 and A naught as 89 and automatic you have seen that all your H and I would have changed and K is.

So we can simply go to sheet 1 and absorb that I am going to another curve here which is green curve which seems to this slightly better fit than the previous case. Still it seems to be over estimating in this chart. Peak seems to be much better on some under estimation is happening in the second half of the curve, but not as much as the previous one. So, this seems to be a better fit and here when you put the cumulative sales, you find that it seems to be a much better fit with your blue line and the green line.

So, now the drawbacks of this diffusion model is that I there has to be some initial number of adopters. So, to you get over that assumption, you have to update our model something called as a base diffusion model named of the person base which we look at it in next class, but for given diffusion model I hope you found out how to fit base on the existing data and in the

column side given what are the step to be. If you follow that, you should you would be able to get these answers.

Thank you.