EXCELing with Mathematical Modeling

Prof. Sandip Banerjee
Department of Mathematics
Indian Institute of Technology Roorkee (IITR)

Week – 12

Lecture – 58 (Estimation of Parameters-II)

Hello, welcome to the course EXCELing with Mathematical Modelling.

Today we will be discussing a very important aspect of mathematical modeling, that is, estimation of parameters.

We have already seen how you can estimate the parameters in a function.

So, today we will be taking a differential equation, which will represent a model and the differential equation will contain some parameters.

So, we will see that how you can estimate those parameters using the data.

So basically you will be having a set of data, say, the time and the variable x. So this will be some t_1, t_2 and so on, this will be some x_1, x_2 and so on.

So by looking at the plot of this graph, so you will be plotting this too and you will be getting some values, some graph like this.

So by looking at this graph, you have to decide that what kind of model can fit this particular data.

So after you have decided what kind of model it is, say, this is the model, then you have to use the method of least squares to estimate the parameter values.

So, let us directly go to the excel sheet and see that how this is done.

So, as you can see that I have the data of E. coli strains. This is the growth data and the time is there and the growth is here.

So, I need to fit a suitable model and estimate those parameters with the help of this data.

Now, since it is a growth of this particular strain, so we will be using the growth model and the growth model which we have studied are either logistic growth or Gompertzian growth, which are a bit realistic than the exponential growth.

So, we will be taking the logistic one and also may be the Gompertzian one and we can compare which one is better.

So, if you try to fit the logistic growth, then your equation will be of the form

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{k}\right)$$

So, my parameters here is r and k, but before you state this equation whether it is a logistic one or whether it is a Gompertzian one, that is,

$$\frac{dx}{dt} = rln\left(\frac{k}{x}\right)$$

where r is the intrinsic growth rate and K is the carrying capacity.

So as I was telling before you write this equation you just see that how this particular data looks like once they are plotted.

So I highlight the time and the growth and let me plot a scattered diagram.

So I go to insert, I choose the scattered diagram and I plot the data.

So, this is, I write parameter estimation.

So, this is the first part that is done.

I assume that let us this particular curve is modelled by this equation the logistic growth, that is,

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{k}\right)$$

So, I have two parameters here, one is this r, another is this k. So, I increase the font size.

And since we will be using the Euler's method I have to take a h and the value of h, I have to put, say, h = 1.

The next thing is the predicted growth.

So this is the one which you already get from the data and then you make the prediction with the help of this particular equation.

So the predicted growth, so I just write here predicted growth.

Now in the predicted growth, I need an initial condition, which I coincide with this value, you could have put something else also.

But since you already have the data, so I coincide the first value that at point 5, the predicted growth also coincides with the actual growth data.

And then I have to attribute some values of r and k, you choose that arbitrarily, and then I calculate the rest of the values using the formula,

$$y_1 = y_0 + hf(x_0)\,.$$

In this case since I am following this equation let me take it as x_0, so x_0 plus h which is 1 and it is a constant. So, I make it a constant multiplied by r. So, this is r again, a constant, r times x_0, which is this value multiplied by $1 - \frac{x_0}{k}$.

k is again a constant that's it so quickly let me check it so this is y_0 plus k6 which is plus h, multiplied by r, okay and then I just drag it, okay.

So, now what you do is, you got the time, you have the data, now you have predicted the growth by using this model.

So, now let us see what is the difference between the data and the one which you have predicted.

So, I will add the curve. So, I will say select data and I will add data.

So, I name it as best fit or let me write predicted.

Now, x value, I choose the time, you highlight this and put the cursor down and press enter.

So, x values have been here.

The y values, I can press this and the y values is the one, which is the predicted one.

You highlight all of them and enter.

So all the x and the y values have been inserted and you click OK.

So now if you see the graph, you now see that this is the data and the orangish one is the predicted value.

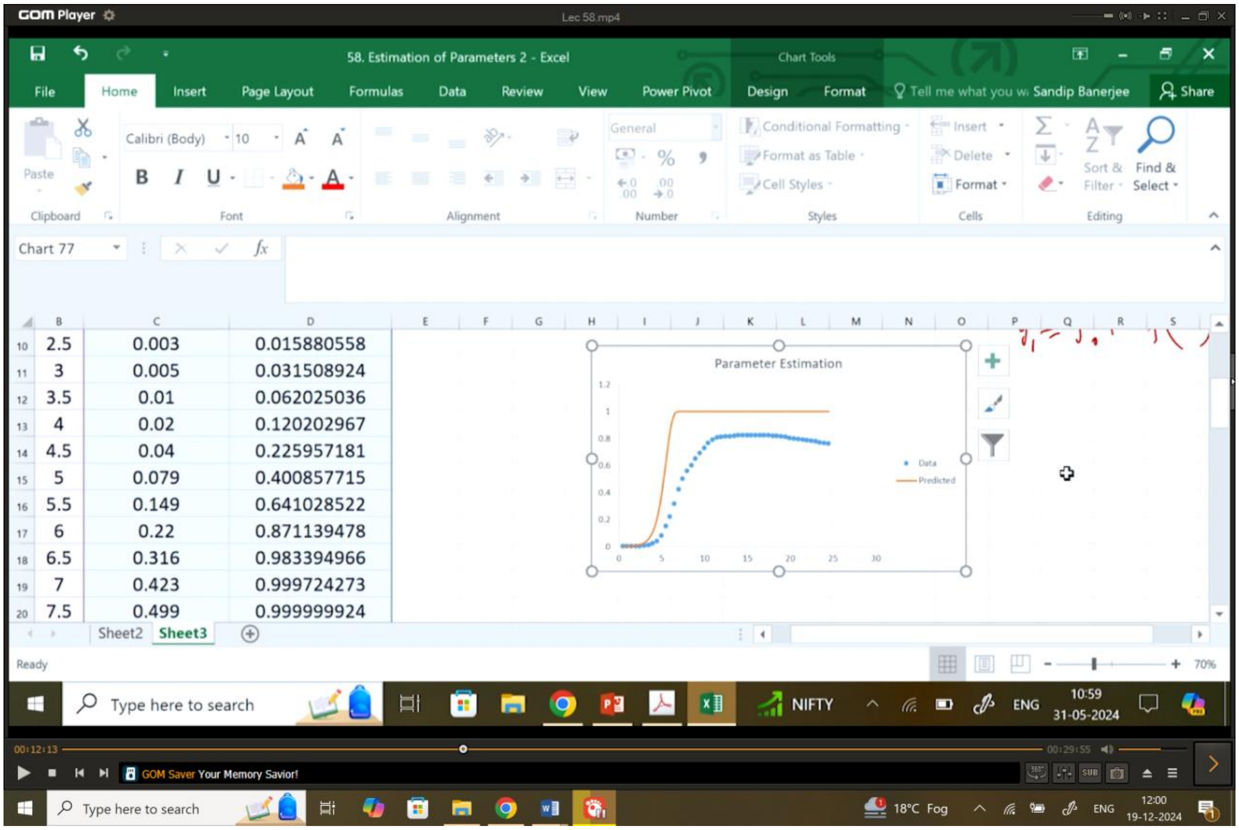
Now, quickly let me change this change chart type okay I want this one to be data, so put the legend, change the series one to data, so I go to design, I select data, so this we have already done so I am not explaining much, since series one, I edit and I name it as data okay.

So the blue one is the data and orange one is the predicted value.

Let me change the data as points.

So here is the data, I make it as points.

So now I have this particular figure where these dots in blue are the data and this is the one which is predicted.



So now how to estimate it using the method of least squares.

So the first thing you do is you try to bring a good match by manipulating these values.

Say I put it 2 and let us see how the behaviour of the graphs, okay so it gives something like this so 2 will not work, so let me put 0.,1 okay again it goes down 0.5 okay goes up 0.3, okay somewhat like this 0.4 okay, 0.7 and so on.

So, like this you bring the curve close to the data. So, it is coming close.

So, you can play with this a bit so that a little good match is found and then what you do is we use the method of these squares.

So that will be the value which you got from the data minus the predicted value, which gives you the residuals.

You square them which will give the residual square.

You sum them and then minimize the residuals, which is just the error.

So what I will do is, this is equal to the data value and the predicted value.

This will be 0 obviously, because both are same and if I drag it, I get some value.

Now, this is residuals, but let me directly calculate the residual square, because we will be needing them actually, so I make them square.

So, this is the residuals square, I just increase the font size, residuals square.

And then what you have to do is you have to take the sum of this square of the residuals.

So, this is equal to sum, you highlight this, press shift, go drag all the values this much close the bracket and enter and you get the sum.

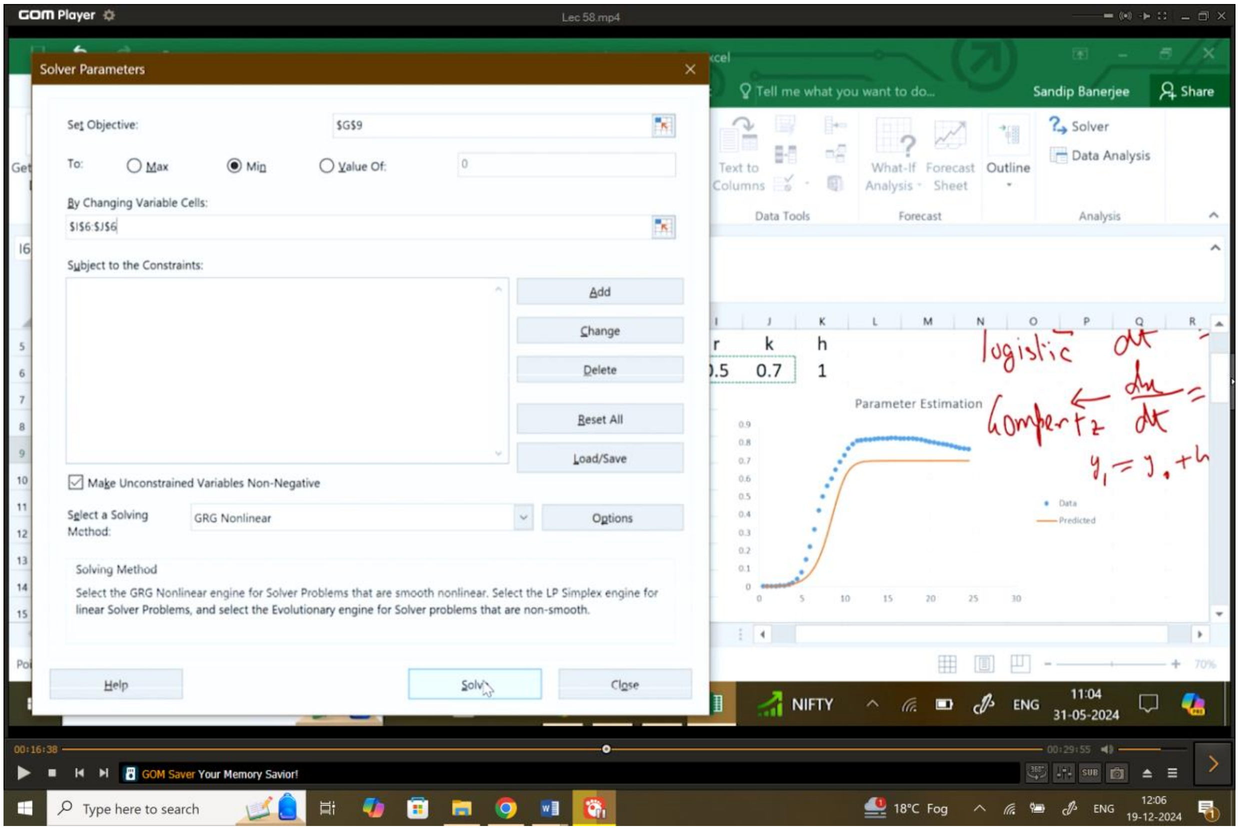
So once you get the sum of the square of the residuals, now I have to minimize this value and for that I will be using the solver.

So go to data, go to the solver and this kind of window is opened.

So I have to minimize this sum of the square of the residuals.

So I click this and immediately, the cell number is highlighted here.

So, I have to minimize this value and by adjusting the value of r and k. So, here I will choose this r and k and immediately it has come here and then keep all them as it is and click solve.



So as you can see that it has automatically adjusted the values using this solver to give a fit.

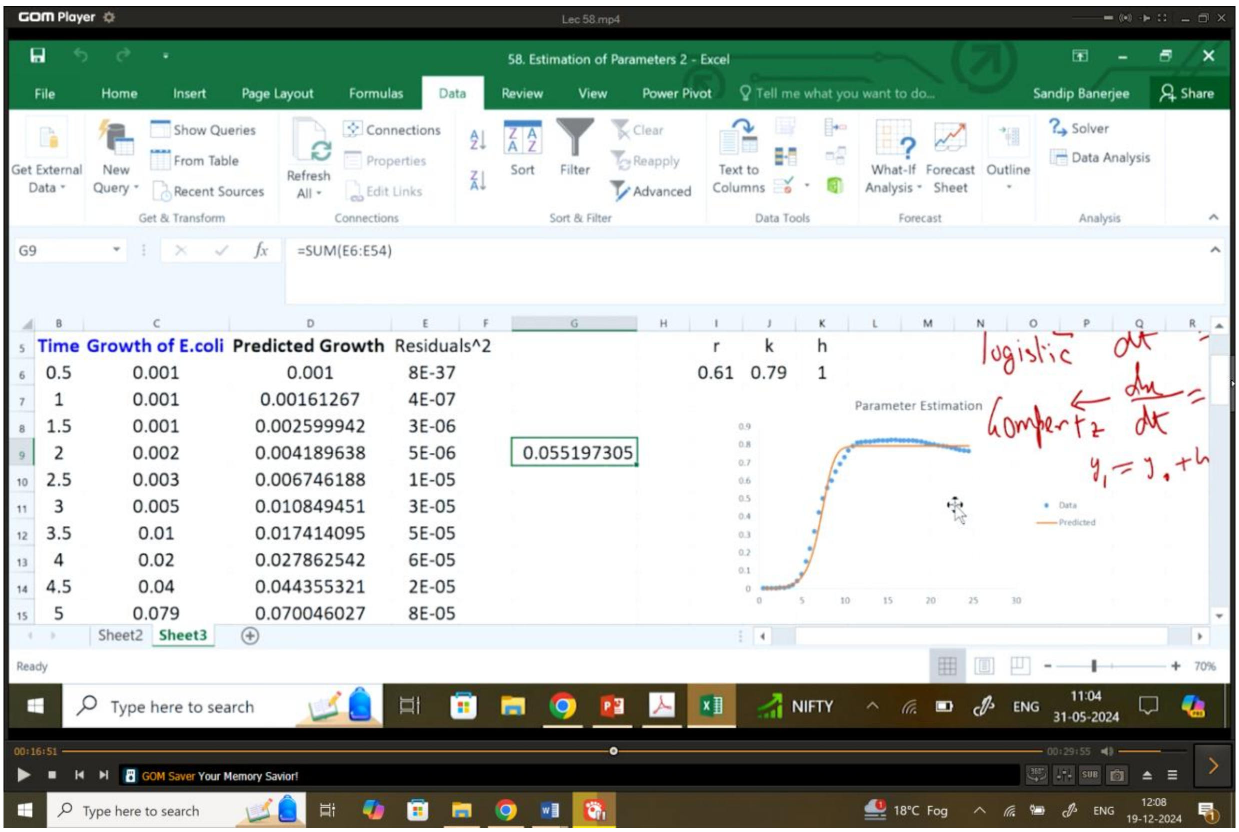
So this is the best fit that you can get for this particular data.

And please note this, this is a real life data so it is not that you can expect that

There will be an exact match of this curve with all the datas that is the best that you can get and the estimated parameter values are

$$r = 0.61 \text{ and } k = 0.79.$$

So, this is how it is done.



Let us move to another example where we take two equations.

Let me copy the data. So, this is the time and this is the measured value, that is, the one which you get from the data.

So, this is measured B. Now, my variables here are A and B and I put it in the form, say,

$$\frac{dA}{dt} = -k_1 A + k_2 B^2, \qquad \frac{dB}{dt} = k_1 A - 2k_2 B^2$$

So I have this system of equation, system of differential equation, ODE is where A and B is my variable and k_1, k_2 are my parameters.

So I have to estimate the values of k_1 and k_2 and the parameters which are given, I mean the data which is given is only of B and not of A. So the first thing is, let me see how these behave.

So, I just go to the scatter diagram and this.

So, this is how the data is behaving and let us now fit this equation in the data and see what the values of A and B we are getting.

So they are the predicted values, now.

So this is the measured and this is predicted A, predicted B. For that I need initial value of k_1 and k_2. So, as usual I put them first to be 1,1 and then we can always adjust them.

So, initial value this is, say A(0) = 1 and B(0) = 0.

So, I coincide this value with the data this I chose arbitrarily.

Now first let us calculate the predicted value of A and B. So this is equal to initially A_0 plus h. So I need a h, which again I take to be 1.

Okay, so we attribute some initial value of this k_1 and k_2 and say, I put 0.1, 0.3 and let us calculate the predicted value of A and B and we use Euler's method.

So, this is equal to the initial value of A, which is A_0 plus h times multiplied by this part, that is, $-k_1A + k_2B^2$.

Similarly, this is equal to B_0 plus h multiplied by $k_1A - 2k_2B^2$ and calculate.

So, highlight these two and drag them to all these values and I change the font to 20.

So let's now plot these values. So I will add data to it. So, select, add data, predicted A

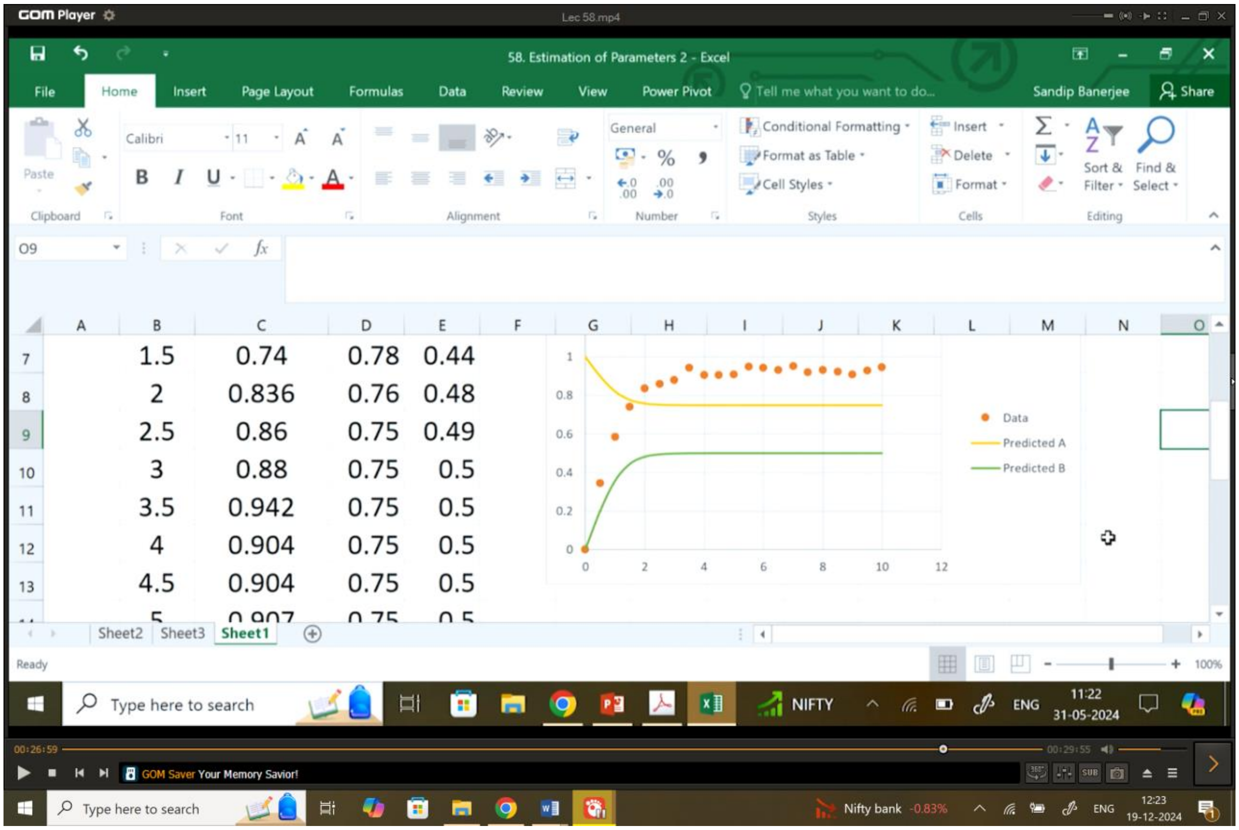
The X values will start from 0 till 10 and the Y values will start from , so this is the predicted A and enter and okay, and add another value, this is predicted B, the X value will start from here, all the times and the Y value is the predicted B value. And okay and okay.

So, let us see how the figures look like.

So, the chart looks like this and I need the legends.

Go to design, select data, the series edit and name it to data, and then the data's maybe in dots but this change to lines. So, change chart type.

So, let the data be scatter predicted A be a line and predicted B be also a line okay. And let us see if I change the colour, say I want a green, yeah. So this is now the data, the predicted A and the predicted B.



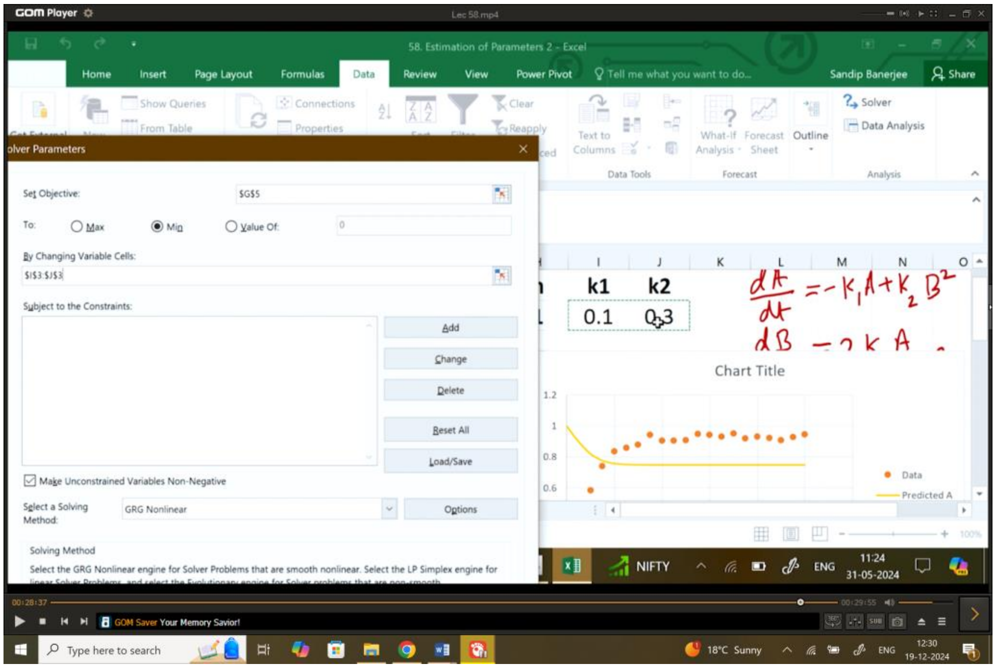
So the next step is you have to calculate the residuals.

So, now you are dealing with the data of B and the predicted value of B. So, everything will be calculated for B. So, this is equal to the residuals, which is the measured values from the data minus the predicted value and I have to calculate the square of the residuals. So, I square them and enter. So this gives me the square values of the residuals.

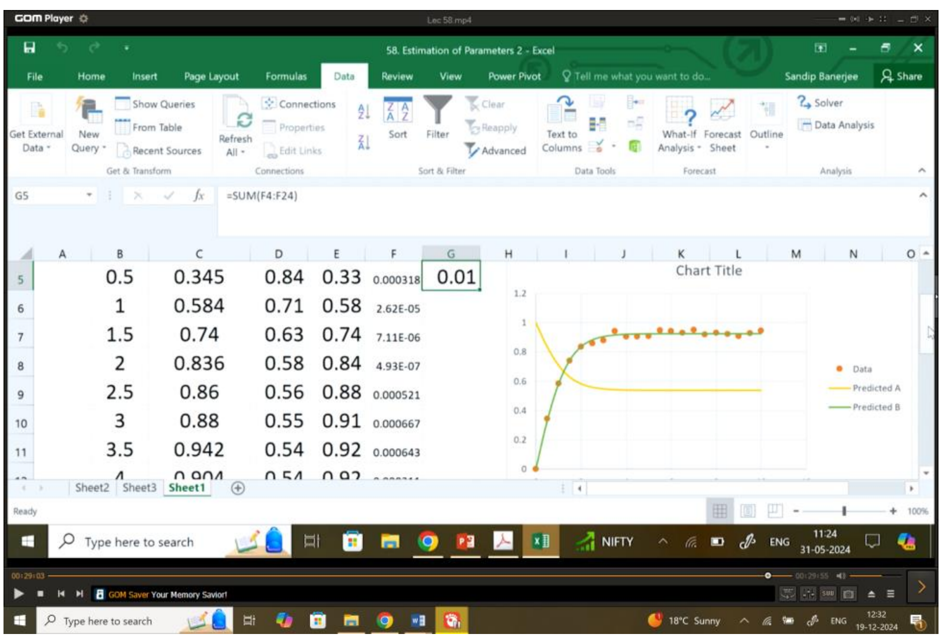
And the next thing is, I have to calculate the sum of the square of the residuals. Enter and I get the sum.

So next is I have to minimize this sum of the square of the residuals, go to the data, I use the solver.

So first I choose which I have to minimize that is this cell, so I click this and immediately it comes here, I have to minimize the value, so I click it here and I have to adjust the variable cell which is my k_1 and k_2 .



It came here and then keep them as it is and just use solve and okay. So now you see that this is your data and it is a very good match. Your predicted A is as it is whereas the predicted B is this green line.



and your estimated value of k_1 and k_2 came to be 0.16 and 0.1.

So, this is how you estimate parameter values using data in the case of differential equations.

In my next lecture, we will again continue with estimation of parameter, but with difference equations. Till then, bye-bye.