EXCELing with Mathematical Modeling
Prof. Sandip Banerjee
Department of Mathematics
Indian Institute of Technology Roorkee (IITR)
Week – 12
Lecture – 56 (Empirical Modeling)

Hello, welcome to the course EXCELing with Mathematical Modelling.

Today we will be talking about empirical models.

Now let us see, what do you mean by empirical model?

So, generally the trend is that when you construct a model you have some data and you use some mathematical tools to formulate the model and then with the help of the data you try to estimate the parameters, that is inside the model, that is the general trend.

So, we used the theory of one form of another, which is any mathematical tool to construct the model and then we use data, if it is available, to determine the value of the parameters that is inside the model.

And this is called model fitting and the resulting model will be called the analytical models.

As you will be noticing that the data which we try to fit it is not an exact fit but an approximate fit but we are ready to deal with it and most of the cases it gives some reasonable predictions.

Now, sometimes we do not use those mathematical tools but only rely on data that is we build the model solely guided by the data.

It depends what kind of situation you are in. So, sometimes you need a trend.

So, for example, say in COVID-19, so, daily there is a spike or daily there is some dynamics going on, whether it is rising or whether it is falling. So, I need daily trend. So, I need the curve what is happening like this. So, you have a lot of data.

So, with the help of this trend or this data, you try to build a model so that you can predict the next trend.

So, that kind of model we call it as empirical models, that is, you have only the data, you do not use any mathematical tools, you use some software to understand this trend and once you understand the trend, you can then predict what will be the rest of the values.

Yes, of course, while capturing the trend, you have to use certain mathematical formula, that is there.

But the actual thing is that this is totally based on data. So, that is what we call empirical models.

So, I will give you a start of how this is done. So, let us start with logarithmic model,

The equation will look in the form

$$y = a + b \ln x.$$

Then, we have power model which is of the form

$$y = ax^b.$$

Then, we have exponential model which is

$$y = ae^{bx}.$$

Here, all a's and b's are constants, they can be positive, they can be negative.

Also, there is a linear model which is

$$y = a + bx.$$

There can be a polynomial model which can be

$$y = a + bx + cx^2,$$

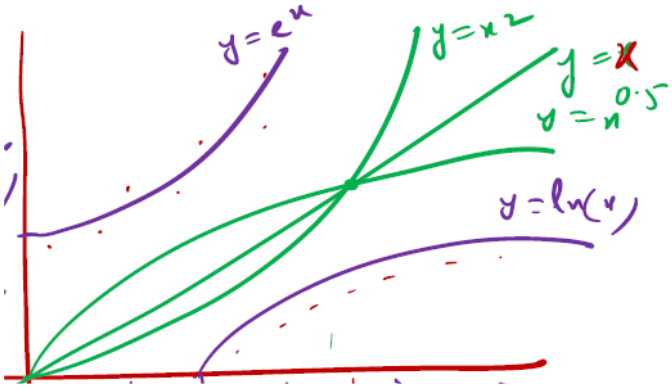
a specific one, this is a parabolic one.

So the question is, which of them to use?

Though I said that, okay, you build up the model with the help of the data, you find the trend, but that trend follows certain equations and those equations you have to choose.

So for example, if I plot the graphs, say, I have this equation, so this is $y = e^x$, this is, $y = \ln x$, then I can have linear one $y = x$, I can have a parabolic one, $y = x^2$, then I can have a power graph $y = x^{0.5}$.

So, all of them coincides at this point, should be (1, 1), but anyway.



So the idea is that the how to choose the equations like this.

So if you have the data and the data follows something like in this particular pattern, then you know okay I have to choose an exponential one.

If it follows a pattern like this, you know that I have to choose a logarithmic one and similarly for the linear or polynomial or the power model.

So how this will work?

So what you will have is that you will have the x values and the y values, say, this is x_1, x_2, x_3, y_1, y_2 and y_3.

So this is the set of data which you will get.

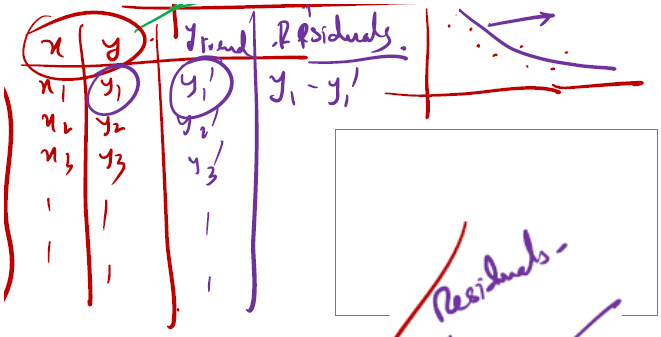
So once you get the data, you plot the data, like this.

And, then you choose any of the trend.

So once you choose the trend, then you will get a curve or a straight line like this.

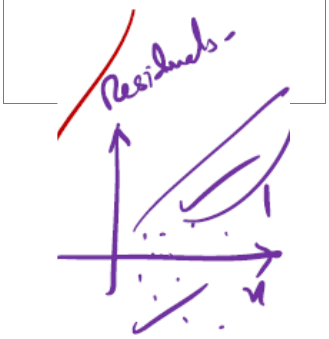
So, you get this trend, you find the equation of this trend and you calculate these values y_1, y_2, y_3 and so on and then there is something called residuals.

So, the residuals means the difference of the data which is given and the observed and the value which we have calculated.



You take the difference of $y_1 - y_1'$

Then you plot, so, this is the difference between the actual data and the one which has been predicted by certain software. In this case, it will be Microsoft Excel.



So, then you plot them.

Here is the x values and here will be the residuals.

So, along the axis y equal to 0, these residuals will be there.

Some of them will be positive, some of them will be negative.

Generally, if it is half of them in the positive direction and half of them in the negative direction and the rest is very close to y equal to 0, it can be up, it can be down, we say okay that this is a good fit.

The residuals should not show any trend. By trend, I mean say it can be say like this or it can be say like this.

So, if they follow certain trend then this is not a very good fit then you change the model and you move on.

So that is how it is going to work.

Let us now look into the numerical or the Excel version that how we will use Excel to empirically create a model.

You will be given a data something like this these are the x values and these are the y values.

So the x value says that in certain countries, these are all countries different countries and so number of persons per the number of physicians available, so 275 for one physician, like that and the life expectancy is 74 years.

So let us first see the trend so you highlight these data's, you go to insert, select the scatter diagram or the chart and plot it like this.

So, I will not use the grid lines and let me try to fit say logarithmic model.

So I put empirical model.

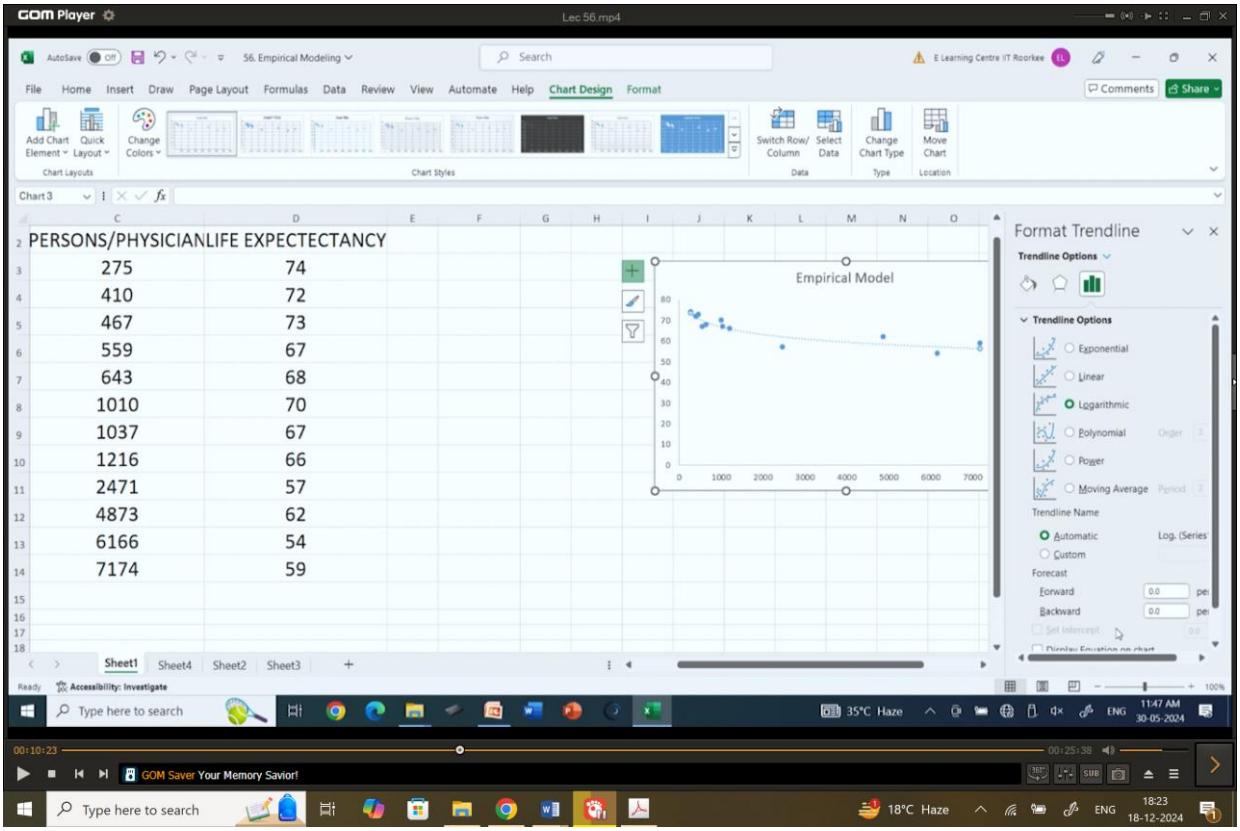
So next is I highlight this and I have to fit a trend which is already inbuilt in Excel.

You click this plus and you see this trend line.

So you click this arrow and so many opens.

Go to more option.

And you will see that there is exponential, linear, logarithmic, polynomial, power, moving average, everything is there.



So, say I choose the logarithmic one.

I click this and you can see that the trend line changes and if you go a little below, you will see display equation on the chart.

So, the line which has been chosen is logarithmic, which is of the form
y equal to a plus b ln x
but I do not know the value of a and b but this particular display equation will actually give you the equation.

So, if I click this I get the equation.

So, I have this take it here you can see the equation let us make it a bit bigger, say 16, yeah, now you can see.

So I have these data, I am trying to fit logarithmic model and I got the equation to be like this. So, let me first calculate these values. So, this is the data which is given and this value will I will get from here.

So, this is equal to minus 5.288 multiplied by ln x. So, x is this value bracket close plus 103.4 and I get this value.

I calculate this, then comes the residuals.

So, the residuals will be the difference between the data and the one which you got from the trend line.

So, this is the difference and you drag and calculate the value.

So, next what we will do is that we will plot the x values and the residuals.

Then you go to insert, go to this chart and get the points.

So, let us make this a bit smaller so that both the charts can fit.

So, I just name it as logarithmic model.

So, as I was telling that how will you know that this is a good fit?

So, we will know it from here that if you see the number of points in the top of y equal to 0.

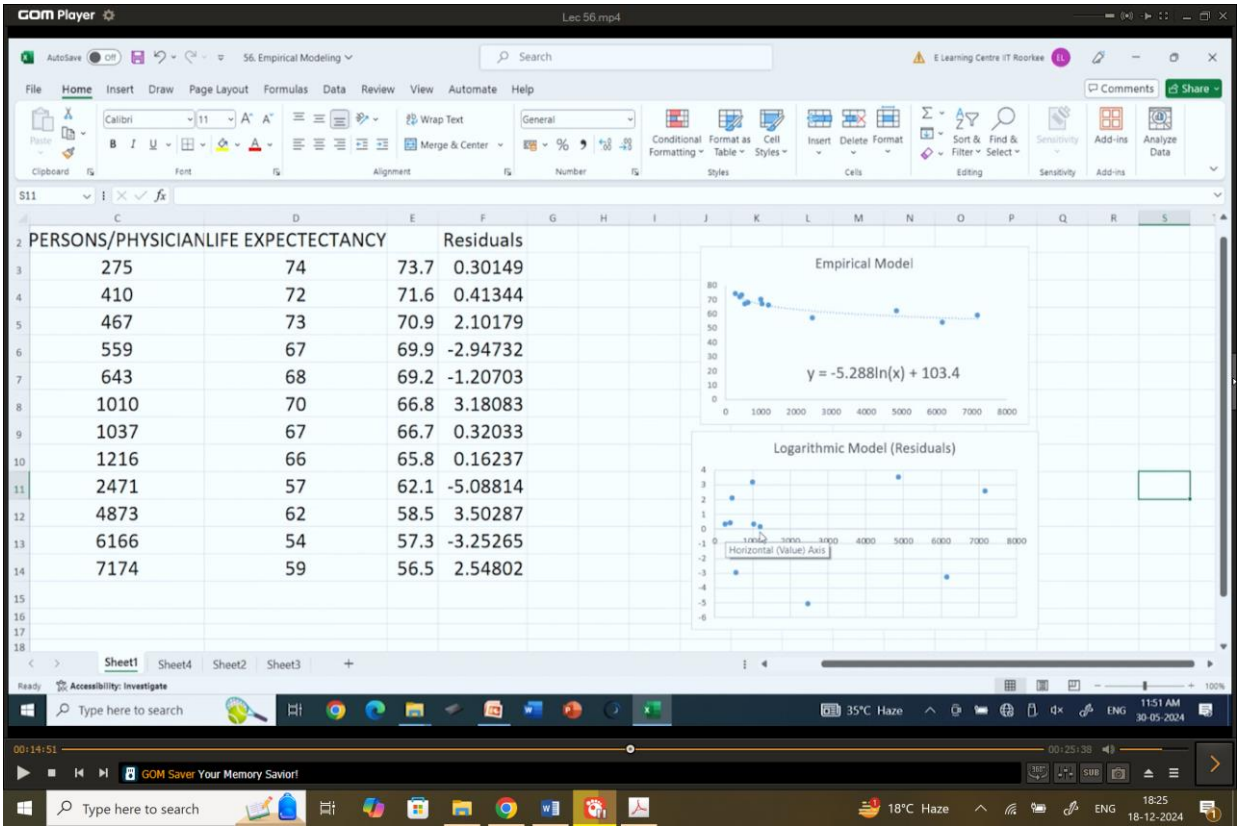
So, y equal to 0 is this straight line. This is the x axis whose equation is y equal to 0.

So, you can see that 1, 2, 3, 4, four points and here also 1, 2, 3, 4. So, four points up, four points down and the rest 4 though they are on the above y equal to 0, but they are very close to y equal to 0.

So, and in fact one of them almost touches y equal to 0.

So, that is why these two variations are very very less.

And hence we say that this empirical model, which is following this logarithmic model is a good fit or is a good model.



And also the residuals do not follow any trend. So, by trend, I mean that you will see in an example that it is follow a certain way of a curve. So, we say but here it looks like they are random.

So, if you get something like this where half of it on one side of y equal to 0 and the other half is on the other side and then the rest of the points are very close to y equal to 0, we know that, okay, this is a good empirical model.

Let us try to fit another model say power model to the same set of data. Let us see what we get.

So we have a power model. So as I told you these were the countries, these were the persons and this is the life expectancy.

So, as usual, first we plot this, we go to insert, we go to charts and scatter diagram.

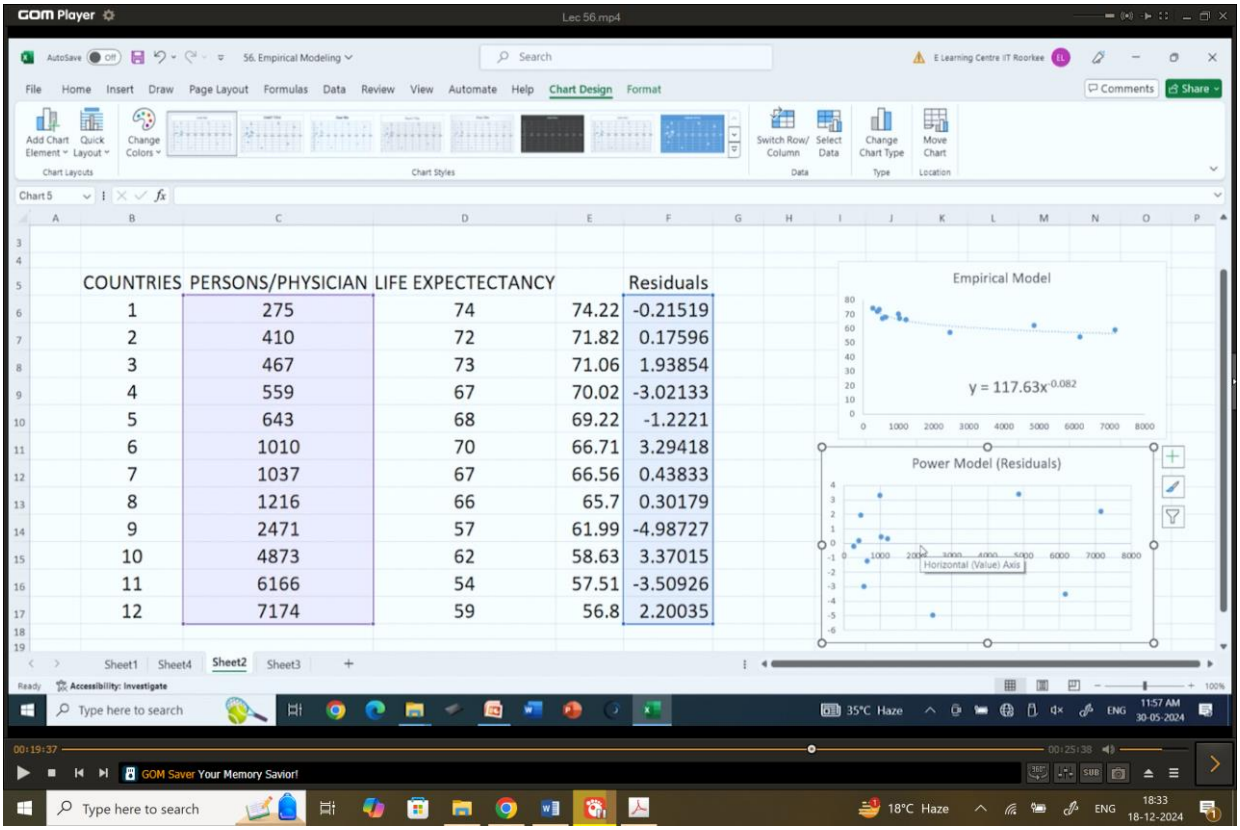
So, I name it empirical model.

Insert so change the chart title empirical model and then I try to create the trend right to the this, I remove the grid lines I go to this trend line and I will choose power. So, I go to more option and here is the power.

So, the moment I click this you can see that there is a change here and I want the equation. So, display this. So, we will take this equation somewhere here and increase the font size so that you can see.

So, this is the equation of the power model which is of the form $y = ax^b$. So, the value of a is 117.63 and b is minus 0.082. I reduce this a bit and then I calculate this is equal to 117.63 multiplied by x to the power minus 0.082. Okay and I get this values, I calculate them, then I calculate residuals.

So, the residuals are going to be the difference between the data and the trend equation get this value you drag it.



COUNTRIES	PERSONS/PHYSICIAN	LIFE EXPECTECTANCY		Residuals
1	275	74	74.22	-0.21519
2	410	72	71.82	0.17596
3	467	73	71.06	1.93854
4	559	67	70.02	-3.02133
5	643	68	69.22	-1.2221
6	1010	70	66.71	3.29418
7	1037	67	66.56	0.43833
8	1216	66	65.7	0.30179
9	2471	57	61.99	-4.98727
10	4873	62	58.63	3.37015
11	6166	54	57.51	-3.50926
12	7174	59	56.8	2.20035

So, please note this is the difference so this minus this or this minus this really does not matter there is only the change of signs.

So, the next thing is you plot the x and the residuals as a scattered diagram so power mortal let's see duals. So, this is the line y equal to 0 here and you can see this is 1, 2, 3, 4 again 1, 2, 3, 4.

So, 4 on the one side of y equal to 0, 4 on the other side, here also this is 5, this is also 5 and these two are quite.

So, I can take this one as one 5 and this one as another 5 and this and this are very close to y equal to 0.

So, again this is a good empirical model.

So, with this set of data both the logarithmic trend as well as the power model trend, they both fits well and you can follow any one of them.

Let us go to another one.

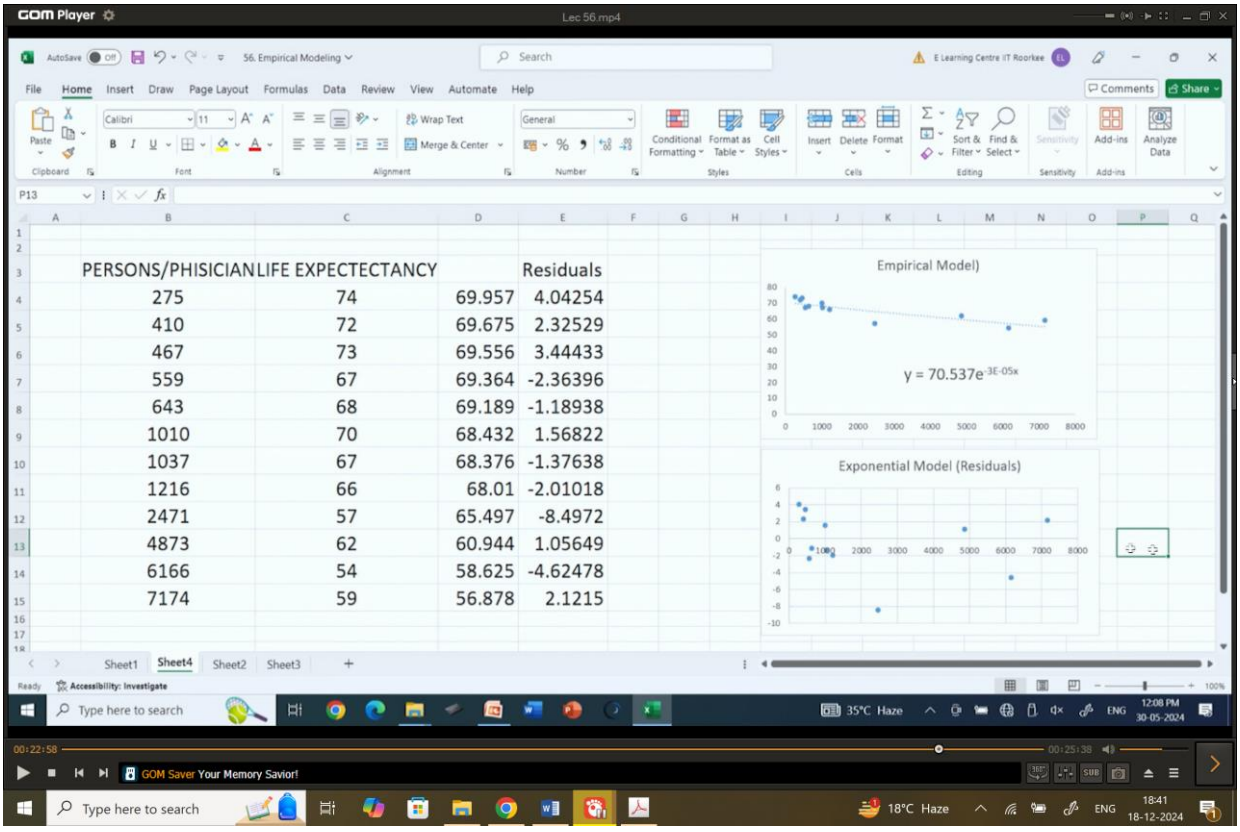
So, here, let us try another empirical model say exponential model.

So, I go to this, go to insert, I scatter and I get this model.

Empirical model, go to this I do not like grid lines go to trend line click this and exponential I can go to more options and I click exponential though it looks like a straight line the moment we click the equation display equation it will be clear that it is an exponential model and not straight line one.

So, and I change it to 16.

So, now you have a look that this is the model.



So, if I want to plot this, this is equal 70.537 multiplied by exp minus 0.12343 multiplied by x.

Then I calculate the residuals okay.

So now we calculate the residuals which is this data value minus the value which you got from the trend equation.

You calculate the residuals and then you plot the X and the residuals and you get the exponential model residual chart

Let me now go back to the slides and quickly explain the trends.

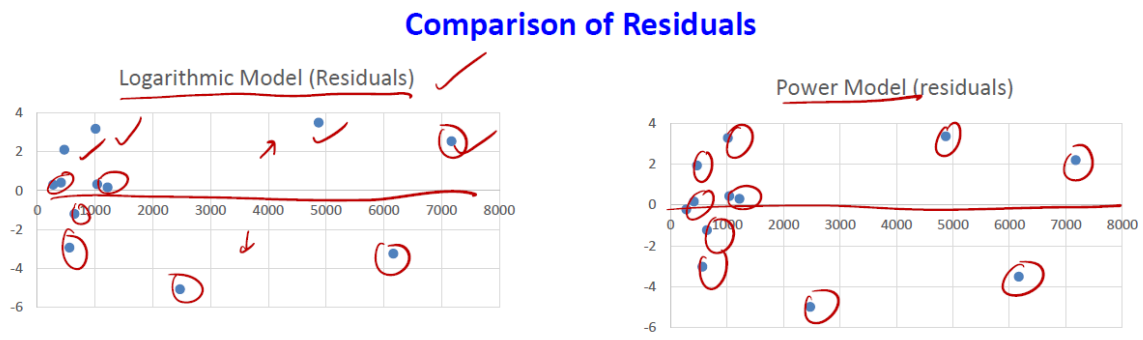
So, let us quickly comparison the residuals which we already calculated in our spreadsheet.

So, as explained that in the logarithmic model, you can see that 1, 2, 3, 4 and 1, 2, 3, 4. So, this is the line y equal to 0.

So, 4 of them lies below y equal to 0, 4 of them lies above and these two are very very close to the line y equal to 0.

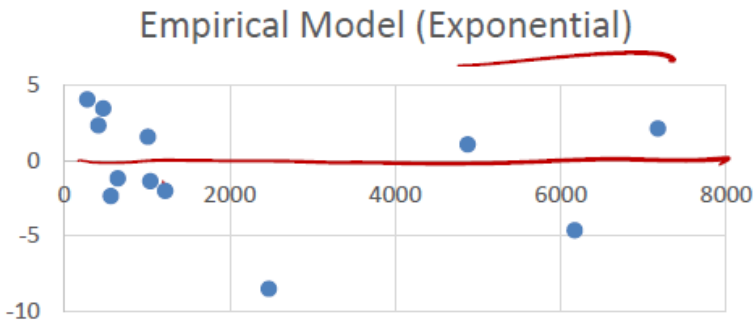
So, if we get such a thing, we say that the empirical model is good.

The power model also gives a similar kind of thing that you have 1, 2, 3, 4 and again 1, 2, 3, 4. This is the line y equal to 0 and we have these two and these two are quite close to y equal to 0 and we say that your power model also behaves well.



So, either you can use the logarithmic model or you can use the power model for this.

So, if you look at to the exponential model in this case you can see though 6 of them lies in the top and we have 6 of them lying at the bottom of y equal to 0 but their difference is quite high.



So the conclusion is that the logarithmic model and the power model, they are better empirical models in this case than the exponential model.

So with this we come to an end of this lecture where you get a rough idea that how to create an empirical model from the data.

In my next lecture, we will going through the estimation of parameters which is again one of the very important aspects in mathematical modelling.

Till then, bye-bye.