Quantitative Methods in Chemistry

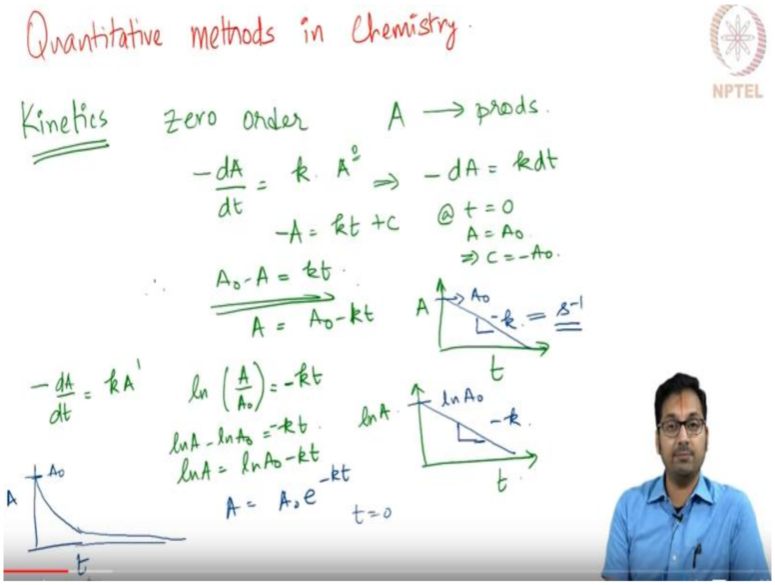
Prof. Dr. Aasheesh Srivastava, Dr. Bharathwaj Sathyamoorthy
Department of Chemistry
Indian Institute of Science Education and Research, Bhopal

Module No # 05

Lecture No # 25

Using spreadsheet and MATLAB towards data analysis with example of rate kinetics

(Refer Slide Time: 00:28)



Quantitative methods in Chemistry.

Kinetics zero order $A \longrightarrow prods.$

$-\frac{dA}{dt} = k \cdot A^0 \Rightarrow -dA = k dt$

$-A = kt + c$ @ $t = 0$, $A = A_0$ $\Rightarrow c = -A_0$.

$\therefore A_0 - A = kt$.

$A = A_0 - kt$

$-k = s^{-1}$

$-\frac{dA}{dt} = kA^1$

$\ln\left(\frac{A}{A_0}\right) = -kt$

$\ln A - \ln A_0 = -kt$.

$\ln A = \ln A_0 - kt$

$A = A_0 e^{-kt}$

$t = 0$

Welcome to the next lecture in the series of quantitative methods in chemistry where we are trying to look at how software could be used towards data input analysis and also plotting. In the last class we saw how when you are trying to determine the enthalpy change for a given process how when you make a measurement as a function of temperature you would be able to determine, what is the enthalpy change?

There you had a simple linear relationship between the independent variable temperature that you are changing to that of the variable that you wanted to measure which is enthalpy change not always you have a linear relationship and in today class we will be taking a look at few such examples. In order to do that, I am going to be invoking some of the basic concepts that we might have learnt in kinetics chemical kinetics.

So why do not we take a look at some of the equations that come across let us start with the 0 order experiment or 0 order rate kinetics and we are always talking about A going to products and this one is able to remember that the rate of change of reactants is given by the rate constant times A to the power 0 because it is a 0 order reaction. So this will be – dA = k dt this when integrated gives you the integrated rate law which is A= k t + c at time t = 0 we know that A = A naught of course there is minus sign here.

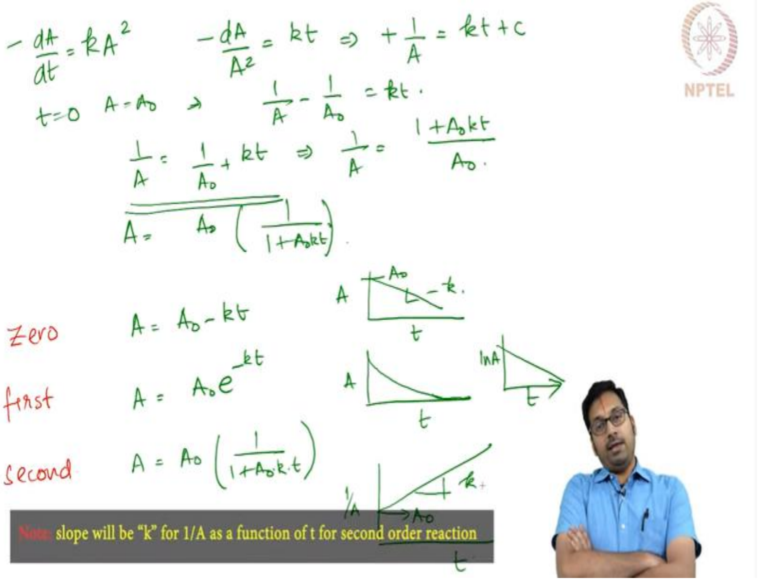
So therefore this implies C = - A naught therefore A naught – A = k t that is if you plot A as a function of t you are going to get something like this. If t is the variable and A is the concentration at any point of time you are going to get a constant line with the slope where the intercept is a concentration at initial time this slope is the rate constant. In this case the rate constant has the units of second inverse because at 0 order.

Now similarly for a first-order reaction you are going to have – dA / dt = k A power 1 this when you rearrange and do the reintegrated rate law similar to what we just did this is going to get reduced to lon of A over A naught = k t. So this can also be observed in the fact that if you are trying to plot as a function of time t. Let us say you are plotting lon of A so this can be written as lon of A – lon of A naught = k t it is – k t okay.

So what will end up happening lon of A is going to equal to lon of A naught – k t. So once again you are going to have a straight line whose intercept is lon of A naught at t = 0 the slope is – k right. So you would be able to determine same in the linear h plot on the other hand let us say you are plotting A as a function of t if you quickly rearrange you are going to get A = A naught power n – e power – k times t.

So once you have this what is going to end up happening at t = 0 you are going to have this as A naught and as a function of time this is going to go like this an exponential function. Of course this will not dye to 0 will dye to 0 only at infinity it will get very close to 0 at larger times. But it is going to not be equal to 0 until t reaches infinity.

(Refer Slide Time: 04:42)



$-\frac{dA}{dt} = kA^2 \qquad -\frac{dA}{A^2} = kt \Rightarrow +\frac{1}{A} = kt + c$

$t=0 \quad A = A_0 \rightarrow \quad \frac{1}{A} - \frac{1}{A_0} = kt.$

$\frac{1}{A} = \frac{1}{A_0} + kt \Rightarrow \frac{1}{A} = \frac{1 + A_0kt}{A_0}.$

$A = A_0 \left(\frac{1}{1 + A_0kt}\right).$

Zero $\quad A = A_0 - kt$

first $\quad A = A_0e^{-kt}$

second $\quad A = A_0\left(\frac{1}{1 + A_0k \cdot t}\right)$

Note: slope will be "k" for 1/A as a function of t for second order reaction

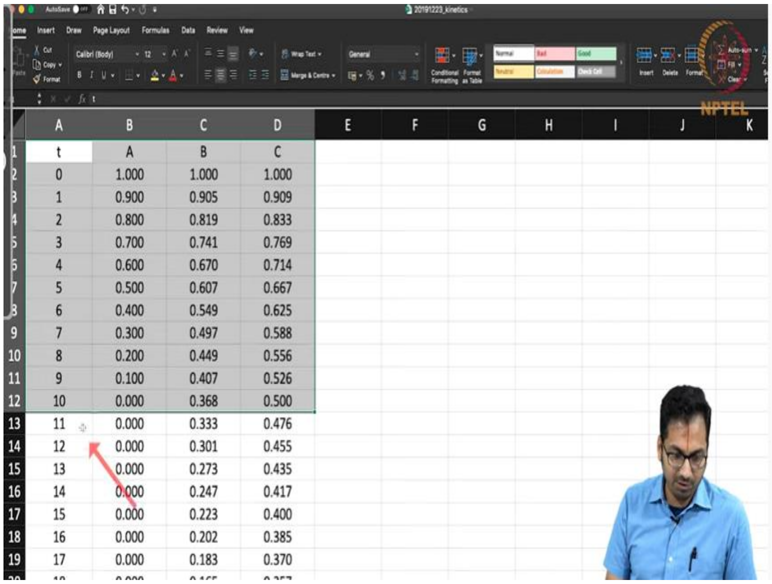
Similarly for second order it is going to be – d A / d t = k a square. So this is going to integrate into d A is k t implies + 1 over A = k t + c t = 0 A = A naught which implies 1 over A – 1 over A naught = k t. you can quickly rearrange the same way that we did earlier so this is going to be 1 over A = 1 over A naught + k t which implies 1 over A is 1 + A naught times k t divided by A naught therefore A as a function of t is going to be something like this.

So therefore 0 first and second order reactions let us try to summarize what we saw you are going to have A = A naught – k t in this one you are going to A equal to A naught times e power – k t in this case you are going to have A = A naught times 1 / 1 + A naught times k times t. If you are plotting A as a function of t you are expected to get a straight line with the intercept being A naught and the slope being – k and in this case if you plot A as a function of t.

They are going to get that dies are 0 at o for first order reaction if you are plotting A as a function of P then you are going to get an exponential dk as a function of time where A reaches 0 only at t infinity on the other hand if you plot lon of A as a function of t you get a straight line and in this case what you are able to realize if you plot 1 over A as a function of time then you are going to get a curve like that where at time 0 you get A naught and the slope is going to be equal to t.

In today's example we will be taking a look at simulated data for 0 first order and second order reactions and try to find out how a software could be used to distinguish apart from these different order of reactions. As always we will start with a spreadsheet program.

(Refer Slide Time: 07:16)



t	A	B	C
0	1.000	1.000	1.000
1	0.900	0.905	0.909
2	0.800	0.819	0.833
3	0.700	0.741	0.769
4	0.600	0.670	0.714
5	0.500	0.607	0.667
6	0.400	0.549	0.625
7	0.300	0.497	0.588
8	0.200	0.449	0.556
9	0.100	0.407	0.526
10	0.000	0.368	0.500
11	0.000	0.333	0.476
12	0.000	0.301	0.455
13	0.000	0.273	0.435
14	0.000	0.247	0.417
15	0.000	0.223	0.400
16	0.000	0.202	0.385
17	0.000	0.183	0.370

So if you realize here data has been provided where you have concentrations as a function of time. Since this is simulated data looks beautiful with no noise. But in reality when you are measuring it depending upon the variable that you used to measure for instance if you are getting the absorbance it is going to have some noise that comes from the spectrophotometer if you are going to be measuring conductance it is going to have it is own error that comes up.

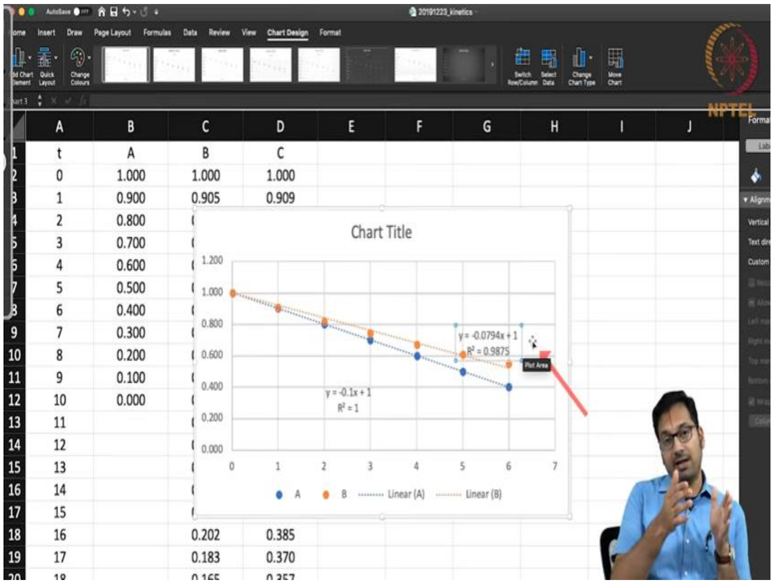
But in today's class for the sake of clarity we will not be introducing any uncertainties with any of these measurements just so that we understand how data can be once again analyzed using a software program. So the data that we are going to be using today is given as a in the format here were we have already seen in the previous class how to carefully enter data. So this is the time variable where time increases in a certain unit and let us says that it is given in the resolution of seconds meaning that the 0 seconds 1 second and data is acquired every second.

And this is one example where the concentration of the reactant changes as a function of this time A and this is concentration of V that changes as a function of time and this is concentration of c changing as a function of time let us assume these 3 are independent reactions meaning that they have their own order and we can assume that these are the only reactants that are required to form the products. Once again all that we are trying to do is how software could help us analyze data so therefore we are simplifying cases.

So that you can understand the simplest of case and extend it to any case that you might require in the research that you might end up pursuing. So let me let us first ask which is the easiest one to plot here if it is a 0 order reaction. What is going to end up happening is that when you plot the change of concentration of the reactants as a function time the 0 order is going to give you a linear curve.

So let us try to do it if any one of them is indeed 0 order reactions that are going to be ending up as a linear curve. So I am going to select a bunch of data and if you are able to realize after a point after the tenth second there is no more of a present. So I can actually go ahead and eliminate the data set so that plotting becomes easier. How would we plot this is you select column A and column B and you can go to insert.

(Refer Slide Time: 09:39)



t	A	B	C
0	1.000	1.000	1.000
1	0.900	0.905	0.909
2	0.800		
3	0.700		
4	0.600		
5	0.500		
6	0.400		
7	0.300		
8	0.200		
9	0.100		
10	0.000		
11			
12			
13			
14			
15			
16		0.202	0.385
17		0.183	0.370

And then you can go to go to the window that says insert a scatter plot and you are able to see a scatter plot here. And what you will realize that it does appear a little linear. So why do not we also include the other plots of A, B and C so the way to do it is to select all the 4 then say insert and insert a scatter plot and what you are able to realize here is that different things happen. So let us try to zoom in to the initial portion of this graph so the way you can do that is by formatting axis here.

So let us say that instead of plotting all the points we will only plot the first points for the first 20 seconds and what you are able to realize is that r that seems to be a blue line that seems like a straight line that it is an orange line that also appears straight for a little while but then there is a curvy part that comes after let us say 12 seconds the same seems to be the inference for the gray curve.

We will soon understand what different orders they belong to and 1 point that we are able to realize is that for any kinetic reaction one has to acquire good number of data points in the time domain so as to distinguish. If we are just taking the initial points we are able to realize that all this could seem like a straight line but they actually are not but of these three it appears that the first one could be the straight line.

So why do not we try fitting it with the trend line so we say right-click add a trend line when you add a trend line what you are able to realize here is that it already fits it to the linear curve by default it is fitting it to a linear curve and the best way to determine whether it is doing a good job is by showing the fit and what you are able to see here is that it appears that it is an ideal fit meaning that the r-square that comes up is equal to 1.

You are able to realize the R square is equal to 1 this indicates that the fit is excellent and what you are able to realize is that there is a certain slope that is associated with it and that is a there is a certain intercept that is associated of course the intercept is the first data point itself and if you remember and recollect just a moment back with the kind of equation we wrote for the 0 order is nothing but A = A naught – k times t and this is of the same form where A naught comes as the intercept and the k comes as the slope here.

So what you are able to realize here this is indeed a 0 order reaction that is the reaction that involves A is a 0 order reaction and the rate constant is 0.1 second inverse. We can try to do the same thing for the orange curve as well but one has to be very careful doing this let us try to do that carefully and then you can always say add a trend line when you say you add a trend line you are also able to get something here.

But interesting line of what ends up happening let us take a look at the equation and the R square you are able to realize the R square for such a fit is very bad what do we mean by bad an R square of 0 means there is no fit at all data are scattered there is no correlation between the X and Y axis on the on the other hand R square or 1 indicates that the independent and the dependent variable x and y respectively are indeed correlated.

On the other hand if you get something like this where you do get an equation and you do have some points that pass so for the sake of clarity let me remove the grey points and you will realize that the orange dotted line which is the fit does not have many points that end up passing through it okay. So in order to exemplify this let us try to show something more we can say minimum can be set to 0.

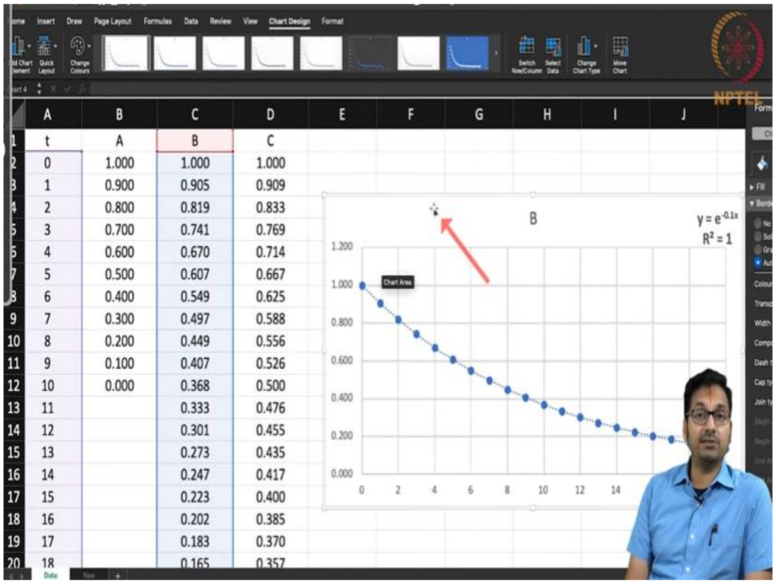
So what you are able to realize is that the fitted line that is the orange starts do not seem to be passing over the orange data points which indicates that the fit is very bad the same is indeed reflected by the r square that has come for such a fit the r squared of 0.5 indicates it is a very bad fit. Other hand if you take a look at the blue fitted line you are able to realize that it comes up as a beautiful fit with the R square of 1.

Of course this synthetic data so you get R Square of 1 what but generally whenever you get an r-squared of 0.99 and you try to believe this is ok and you move on. Now let us try to see what happens when you reduce the number of data sets meaning that let us say you only plot the first 10 data points what ends up happening. So we can try to fit the one that we found out as the 0 order reaction of course intersect we set to 1 not to 0.

Because we know for a fact that the concentration that you started with on the other hand let us try to fit the orange line then we can say you display the equations. Let us play R square and display the intercept a certain set to 1. So interestingly enough what you end up getting for the other orange data point is that the R square is still good this is confusing. What is happening here? Is that you are taking different sets of data points for the same data spanning a larger time you get a very poor fit for the linear line.

But the same set of points where shorter time you get a very reasonable fit if you are looking at R square is about 0.99 and can we believe this is first order or 0 order for that matter actually not if you are fitting something linear to the concentration of reactant linear as a function of time then it indicates that it is 0 order reaction. But this need not be the case this just indicates the fact that as you acquire more and more data points different models can be discerned models meaning you are acquire a given number of samples before you are sure that the data that getting can be reliably interpreted. So let us go a little forward let us not worry for now what is happening here we will come back to this question may be a little later.

(Refer Slide Time: 16:08)



t	A	B	C
0	1.000	1.000	1.000
1	0.900	0.905	0.909
2	0.800	0.819	0.833
3	0.700	0.741	0.769
4	0.600	0.670	0.714
5	0.500	0.607	0.667
6	0.400	0.549	0.625
7	0.300	0.497	0.588
8	0.200	0.449	0.556
9	0.100	0.407	0.526
10	0.000	0.368	0.500
11		0.333	0.476
12		0.301	0.455
13		0.273	0.435
14		0.247	0.417
15		0.223	0.400
16		0.202	0.385
17		0.183	0.370
18		0.165	0.357

So first thing that we are able to realize is that A is indeed a 0 order reaction now let us take look at what is B let us try to plot a the concentration change of B as a function of time. You get something like this let us take a look at the data points as well not just looking at the software what happens is that the initial data points there is a significant drop in the concentration. The first one drops about 9.5% then you are having about a 9% drop then you are having about a 6% drop not 6 unit drop and then you are having about a small sets of changes that starts coming in right.

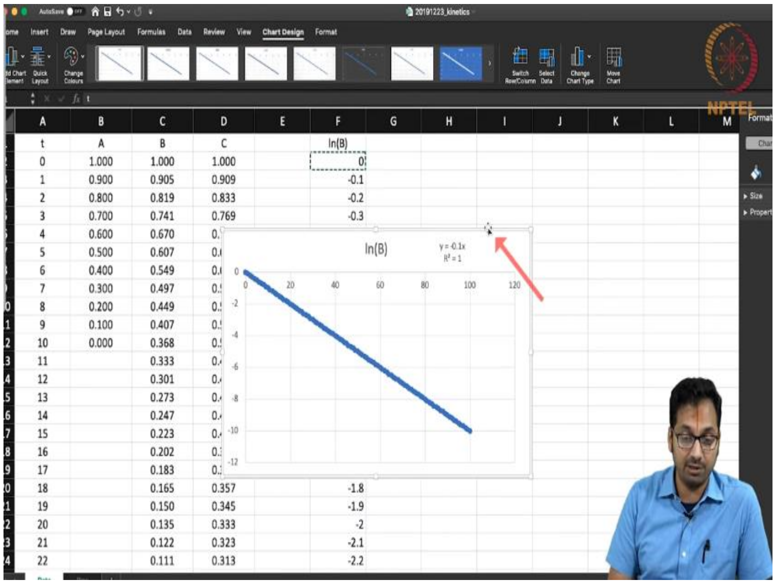
But as you start going further and further you realize the drop actually smaller and even if you go further if you are able to realize for the same 1 second resolution the fall is much more smaller right. So this could indicate probably it is an exponential drop that comes up why do not we try fitting it to a trend line? The first thing that program a spreadsheet program does is to fit it to a linear line but you are able to realize that the linear dotted line does not seem to pass most of the data points okay. If you would like to see we can take a look at the R square okay. This is what we saw a moment back we see that the R square is pretty bad meaning that you are getting something like 0.47 or rather 0.5 which is not a good fit at all.

And more importantly rather than looking in the R square number you are able to realize the fitted line does not encompass most of the data points that you acquired meaning that it is a poor fit therefore the model that you are choosing is wrong. So instead of doing this why do not we try to look at what other models are possible let us try to fit it to an exponential. And let us see what happens you fit it to an exponential suddenly what you are able to see is that Y = e power – 0.1 times X + R square goes to one.

So this could indicate ha it is not a linear fit it is actually a exponential fit. If you recollect a moment back we just said for the first order reaction A = A naught times e power – k times t. So therefore this nicely fits into the equation since we are starting with 1 unit of concentration we are able to realize this is indeed a first-order reaction. And of course since it simulated data you are also able to see the fact that the R square is actually equal to 1 and if you see now the dotted line you are not even able to see the dotted line because all of them are indeed passing nicely through all the data points.

So let us expand it and take a look or we can try to reduce this again format axis. So that we can try to see how much we would like to take a look into what you are able to see the dotted line nicely encompasses all the data points. Once again if you recollect the older fit for the linear line you realize that it went something in a bizarre fashion let us see whether we can fit again the same data okay. So what you are able to realize is that this seems to nicely fit an exponential equation indicating it is a first-order reaction. Let us try to see whether it is indeed true how we do this we can take the lon of the concentration.

(Refer Slide Time: 19:20)



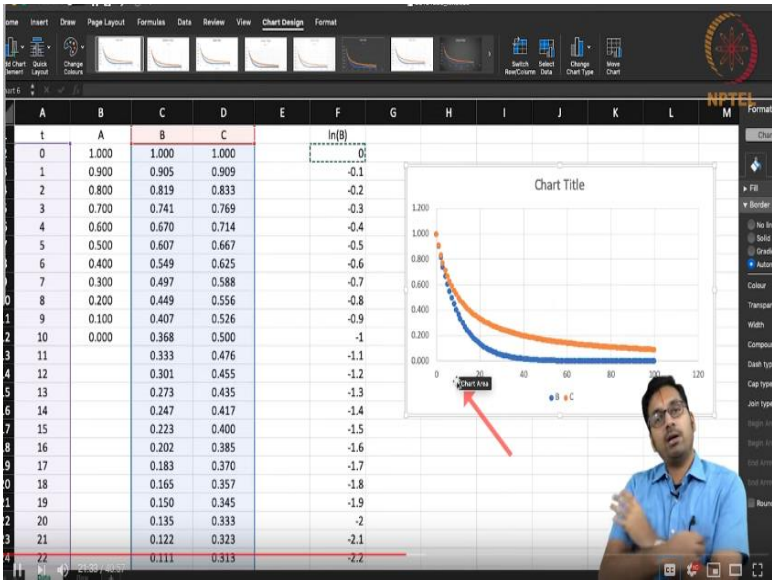
	A	B	C	D	E	F
1	t	A	B	C		ln(B)
2	0	1.000	1.000	1.000		0
3	1	0.900	0.905	0.909		-0.1
4	2	0.800	0.819	0.833		-0.2
5	3	0.700	0.741	0.769		-0.3
6	4	0.600	0.670	0.		
7	5	0.500	0.607	0.		
8	6	0.400	0.549	0.		
9	7	0.300	0.497	0.		
10	8	0.200	0.449	0.		
11	9	0.100	0.407	0.		
12	10	0.000	0.368	0.		
13	11		0.333	0.		
14	12		0.301	0.		
15	13		0.273	0.		
16	14		0.247	0.		
17	15		0.223	0.		
18	16		0.202	0.		
19	17		0.183	0.		
20	18		0.165	0.357		-1.8
21	19		0.150	0.345		-1.9
22	20		0.135	0.333		-2
23	21		0.122	0.323		-2.1
24	22		0.111	0.313		-2.2

So we have learnt in this class before that if you take Ln of something then it becomes lon and in copying that cell in so this is going to be a lon of B. So I can go ahead and copy this and paste it for all the lines going until a certain amount of time. So I am pasting it and then we get a set of numbers so now let us plot lon of B as a function of time if it indeed was a first order reaction you are going to get a linear line you end up getting that.

So this can also be fit to trend line and you can actually take a look at the intercept that is going to come up if you remember you got something like y is equal to E power – 0.1 times X here also you get a slope of -0.1 times X indicates that the k is 0.1 concentration inverse second inverse right. So what you are able to realize the same inference that we get through exponential fit you can also get it through a linearized fit.

However this I would not encourage you to linearize all the time because there is certain amount of problems that is associated in converting the error that you obtain let us say in concentration of B and we need to take log of it. We have learnt in this class how error propagates and that causes an under evaluation of the error which is not a good idea. But this is just a textbook knowledge that I wanted to exploit and see whether it works out ok.

(Refer Slide Time: 20:56)

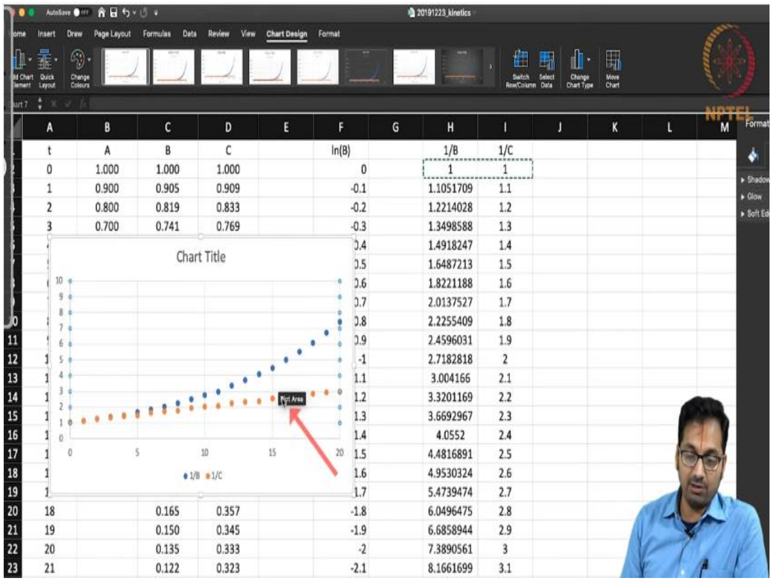


t	A	B	C		ln(B)
0	1.000	1.000	1.000		0
1	0.900	0.905	0.909		-0.1
2	0.800	0.819	0.833		-0.2
3	0.700	0.741	0.769		-0.3
4	0.600	0.670	0.714		-0.4
5	0.500	0.607	0.667		-0.5
6	0.400	0.549	0.625		-0.6
7	0.300	0.497	0.588		-0.7
8	0.200	0.449	0.556		-0.8
9	0.100	0.407	0.526		-0.9
10	0.000	0.368	0.500		-1
11		0.333	0.476		-1.1
12		0.301	0.455		-1.2
13		0.273	0.435		-1.3
14		0.247	0.417		-1.4
15		0.223	0.400		-1.5
16		0.202	0.385		-1.6
17		0.183	0.370		-1.7
18		0.165	0.357		-1.8
19		0.150	0.345		-1.9
20		0.135	0.333		-2
21		0.122	0.323		-2.1
22		0.111	0.313		-2.2

Now that we have seen reaction A is 0 order reaction B is first order natural progression is that reaction C is going to be second-order. Let us first the take a look at plotting reaction B and C and see how different the curves look. If you are able to see as a small change that comes up the blue is the fit for the first-order equation that we just did and the red the orange dots are the one for the reaction.

That you are trying to figure out what you are able to realize is that both seem to have a beautiful progression and this kind of seems similar one can always argue that this could be a first-order reaction with a slightly different rate constant may be it has a rate constant that is smaller than what we saw. However one good way of trying to see this we saw a moment back if you are able to plot 1 over A as a function of t. If you get a linear line then it is a second-order reaction.

(Refer Slide Time: 21:56)



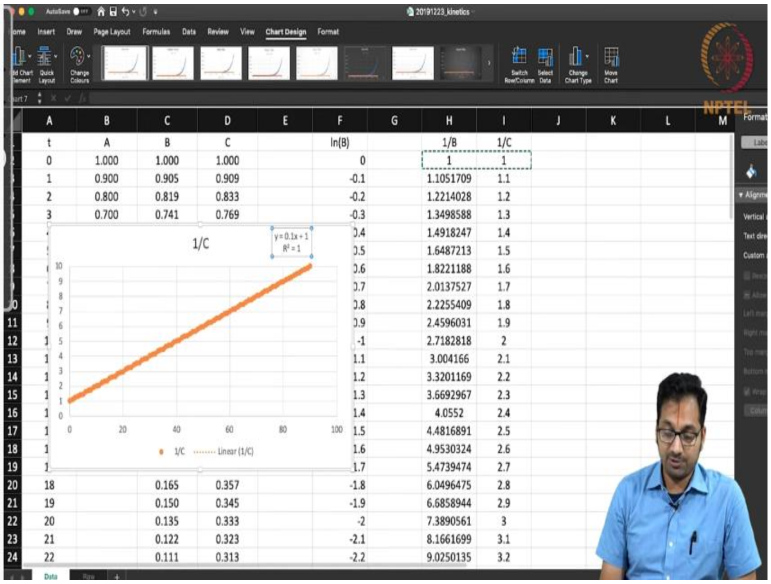
	A	B	C	D	E	F	G	H	I
	t	A	B	C		ln(B)		1/B	1/C
	0	1.000	1.000	1.000		0		1	1
	1	0.900	0.905	0.909		-0.1		1.1051709	1.1
	2	0.800	0.819	0.833		-0.2		1.2214028	1.2
	3	0.700	0.741	0.769		-0.3		1.3498588	1.3
						0.4		1.4918247	1.4
						0.5		1.6487213	1.5
						0.6		1.8221188	1.6
						0.7		2.0137527	1.7
						0.8		2.2255409	1.8
11						0.9		2.4596031	1.9
12						-1		2.7182818	2
13						1.1		3.004166	2.1
14						1.2		3.3201169	2.2
15						1.3		3.6692967	2.3
16						1.4		4.0552	2.4
17						1.5		4.4816891	2.5
18						1.6		4.9530324	2.6
19						1.7		5.4739474	2.7
20	18		0.165	0.357		-1.8		6.0496475	2.8
21	19		0.150	0.345		-1.9		6.6858944	2.9
22	20		0.135	0.333		-2		7.3890561	3
23	21		0.122	0.323		-2.1		8.1661699	3.1

So let us try to do 1 over B and 1 over C we know for a fact that B is a first-order reaction. So if you are plotting 1 over B it is not going to result in a straight line. But we will do this just for the sake so that we understand what we are trying to do so similarly you do 1 over C we are not using a dollar sign anywhere here just to make sure that all the cells indeed do get changed okay. So now we have gotten this why do not we plot this as a function of time?

So I am plotting 1 over B and 1 over C and whichever gives you a straight line as the second-order reaction which ever does not give you a straight line indicates it is not a second-order reaction okay. So there seems to be a problem here because the fit for blue or the number of points for blue increases significantly. So once again as always we would like to take a look at the initial set of data points.

So this is also having trouble because the y-axis is way too much so we can try to reduce this to something like 10 okay. So what you are able to realize here this is what is causing the trouble for visualizing the blue and the yellow points where the y-axis was overdone there. So now when I zoom in properly what you are able realize is that the orange line goes in a certain way well the blue line starts to deviate very fast which is known because we know for a fact it is a first order reaction for B.

(Refer Slide Time: 24:56)



t	A	B	C		ln(B)		1/B	1/C
0	1.000	1.000	1.000		0		1	1
1	0.900	0.905	0.909		-0.1		1.1051709	1.1
2	0.800	0.819	0.833		-0.2		1.2214028	1.2
3	0.700	0.741	0.769		-0.3		1.3498588	1.3
					0.4		1.4918247	1.4
					0.5		1.6487213	1.5
					0.6		1.8221188	1.6
					0.7		2.0137527	1.7
					0.8		2.2255409	1.8
					0.9		2.4596031	1.9
1					-1		2.7182818	2
1					1.1		3.004166	2.1
1					1.2		3.3201169	2.2
1					1.3		3.6692967	2.3
1					1.4		4.0552	2.4
1					1.5		4.4816891	2.5
1					1.6		4.9530324	2.6
1					1.7		5.4739474	2.7
18		0.165	0.357		-1.8		6.0496475	2.8
19		0.150	0.345		-1.9		6.6858944	2.9
20		0.135	0.333		-2		7.3890561	3
21		0.122	0.323		-2.1		8.1661699	3.1
22		0.111	0.313		-2.2		9.0250135	3.2

So therefore it is not going to be a linear line and you are exploding as the t starts to increase the 1 over B is increasing tremendously okay. So now let us try to fit these 2 data points I will actually remove for the sake of argument I will remove the blue data point and then we can try to format and see in it is full glory. And you are able to realize that orange line nicely increases as a function of time you can add a trend line as we have always done to get the linear curve.

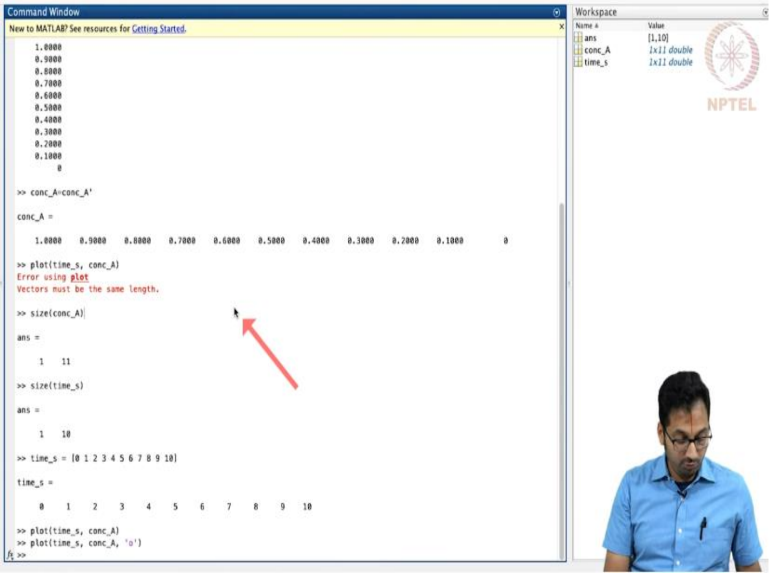
So what you are able to realize because you have got to set the intercept to 1 because your initial data point for time 0 it is concentration of 1 unit and what you are able to realize here the 1 over A = 1 over A naught + k t. And what you are able to realize here is that the y which is A and the intercept 1 is actually equal to A naught which is where you started with and the k comes as 0.1. In this case it will be 0.1 concentration power – 2 times second inverse.

So for all the 3 simulations what I did was to assume the same rate constant of 0.1 units of course the unit is change 0.1 second inverse 0.1 concentration in a second inverse and 0.1 concentration power -2 second inverse and what we are able to see is that for a 0 order data if you just plot concentration as a function of time if you get a linear line then it is 0 order if you plot lon of the concentration as a function of time.

If you get a linear line it is first order if you plot 1 over A 1 over concentration as a function of time you get a linear right it is second-order. And we also tried to do the opposite what happens if you plot 1 over A for first order you realize that it starts to explode and same could have been done for the first order. When you are doing lon of A as a function of time what will end up happening is that you are realized for a 0 order that also does not fit we are able to exemplify the fact here.

That as you start using computation programs or software more it helps us analyze this data are very much faster. Similar to a spreadsheet I am also going to be introducing you another way of the data and this way is by using software called MATLAB.

(Refer Slide Time: 26:01)



```
    1.0000
    0.9000
    0.8000
    0.7000
    0.6000
    0.5000
    0.4000
    0.3000
    0.2000
    0.1000
         0

>> conc_A=conc_A'

conc_A =

    1.0000    0.9000    0.8000    0.7000    0.6000    0.5000    0.4000    0.3000    0.2000    0.1000         0

>> plot(time_s, conc_A)
Error using plot
Vectors must be the same length.

>> size(conc_A)

ans =

     1    11

>> size(time_s)

ans =

     1    10

>> time_s = [0 1 2 3 4 5 6 7 8 9 10]

time_s =

     0     1     2     3     4     5     6     7     8     9    10

>> plot(time_s, conc_A)
>> plot(time_s, conc_A, 'o')
fx >>
```

Name	Value
ans	[1,10]
conc_A	1x11 double
time_s	1x11 double

So MATLAB is this program where we had already simulated some of the curves that we saw for the error propagation also the way we saw the Gaussian fix that came for different numbers that we had obtained experimentally. This is software which also works the same way like a spreadsheet does generally you can give variables the advantage of this software. However is that you can also read the excel sheets that you might have prepared for analysis and write program such that it can be semi-automated towards your data analysis.

I am just going to be giving you a brief introduction of MATLAB I am sure there are other tutorials available for you to understand how this can be used. So the way you saw in the excel sheet where a certain column has a certain running variable that we saved it as here also a similar thing can be done by giving different variables. So let us say we want to define time as something like time and seconds as time underscore S this is just a variable you can name it any possible way.

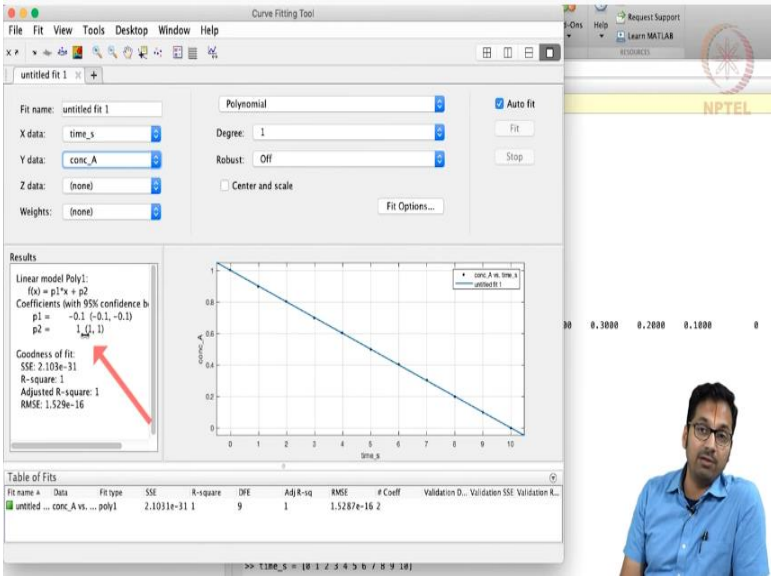
So let us just say it is time underscore S so I want to give it 10 such points what you are able to realize now the time is nothing but row vector which has 10 units in it basically 1 going until 10 similarly to the concentration of A. So we can say Conc of A you can do the same thing and the advantage is you can even copy paste from excel. What you are able to realize here the time is a row vector while concentration of A is a column vector.

So one can always make convert them into the other vector by using a transpose which I am sure you guys have learnt with basic matrices if you take a transpose a row matrix it becomes a column matrix. If you take a transpose of a column matrix it becomes a row matrix. We can try to plot now time underscore S in the X axis and concentration of A okay. So it is kind of giving you an error that says the vectors must be of same length.

Let us take a look at what is the length of the vectors you see that concentration of A is a 1 / 11 vector while concentration of time underscore S is 1 / 10. So this is something not okay so we can always introduce one more into this actually if you remember the time that we ended up defining should just have 0 seconds that go with it. So now if you say plot you would be able to see the plot.

You would be able to see the plot where this is time time axis this is the concentration and you are able to see a straight line remember we do not have points between 0 and 1 second but this software ends up showing this you. So rather than plotting it this way you are better off plotting it with data point in this aspect. So what you are able to realize is that there you that is the data point that you are trying to look at and as a function of time this indeed does change.

(Refer Slide Time: 29:43)



Fit name: untitled fit 1
X data: time_s
Y data: conc_A
Z data: (none)
Weights: (none)
Polynomial
Degree: 1
Robust: Off
☐ Center and scale
☑ Auto fit

```
Linear model Poly1:
    f(x) = p1*x + p2
Coefficients (with 95% confidence b
    p1 =     -0.1  (-0.1, -0.1)
    p2 =        1  (1, 1)

Goodness of fit:
  SSE: 2.103e-31
  R-square: 1
  Adjusted R-square: 1
  RMSE: 1.529e-16
```

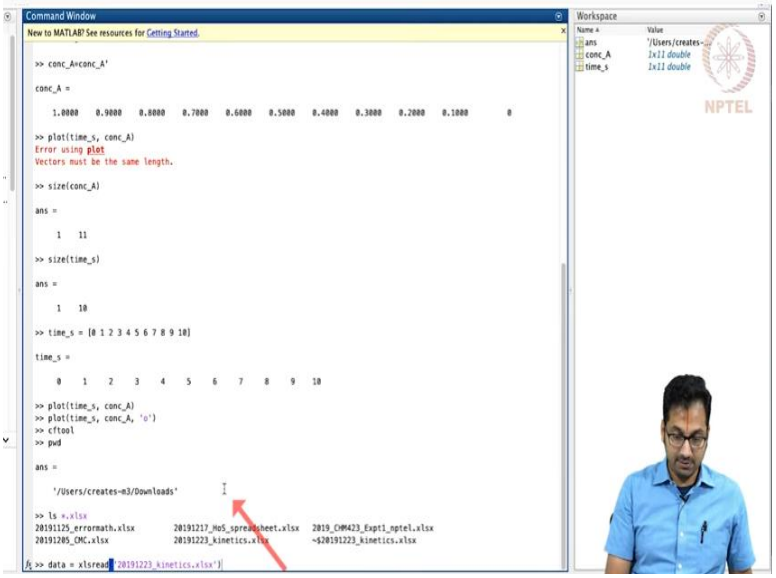
Fit name	Data	Fit type	SSE	R-square	DFE	Adj R-sq	RMSE	# Coeff	Validation D...	Validation SSE	Validation R...
untitled ...	conc_A vs. ...	poly1	2.1031e-31	1	9	1	1.5287e-16	2			

```
>> time_s = [0 1 2 3 4 5 6 7 8 9 10]
```

So the same way we ended up fitting in excel you can use a tool called curve fitting tool if you call this module of curve fitting tool then it is going to come up with a dialog box that looks like this where you and say okay my x axis is time by my y-axis is concentration of A. Then what you are able to realize is that even before we say anything it fitted it to a polynomial and is saying the polynomial as P 1 times X + P 2 where P1 here is the slope that comes as – 0 .1.

And the intercept comes as 1 and it is saying with a 91 95% confidence interval the range is – 0.1 to -0.1 meaning that there is literally no uncertainty. So is the case for the intercept which goes between 1 and 1 which means that the only value it can take as 1 and if you are able to realize the R square comes as 1 which once again is what we saw even in excel spreadsheet there are also other ways of judging the goodness of fit. So therefore you can understand how good you are model agrees with the data that you have generated.

(Refer Slide Time: 30:40)



```
>> conc_A=conc_A'

conc_A =

    1.0000    0.9000    0.8000    0.7000    0.6000    0.5000    0.4000    0.3000    0.2000    0.1000         0

>> plot(time_s, conc_A)
Error using plot
Vectors must be the same length.

>> size(conc_A)

ans =

     1    11

>> size(time_s)

ans =

     1    10

>> time_s = [0 1 2 3 4 5 6 7 8 9 10]

time_s =

     0     1     2     3     4     5     6     7     8     9    10

>> plot(time_s, conc_A)
>> plot(time_s, conc_A, 'o')
>> cftool
>> pwd

ans =

    '/Users/creates-m3/Downloads'

>> ls *.xlsx
20191125_errormath.xlsx         20191217_MoS_spreadsheet.xlsx   2019_CHM423_Expt1_nptel.xlsx
20191205_CMC.xlsx               20191223_kinetics.xlsx          ~$20191223_kinetics.xlsx

>> data = xlsread('20191223_kinetics.xlsx')
```

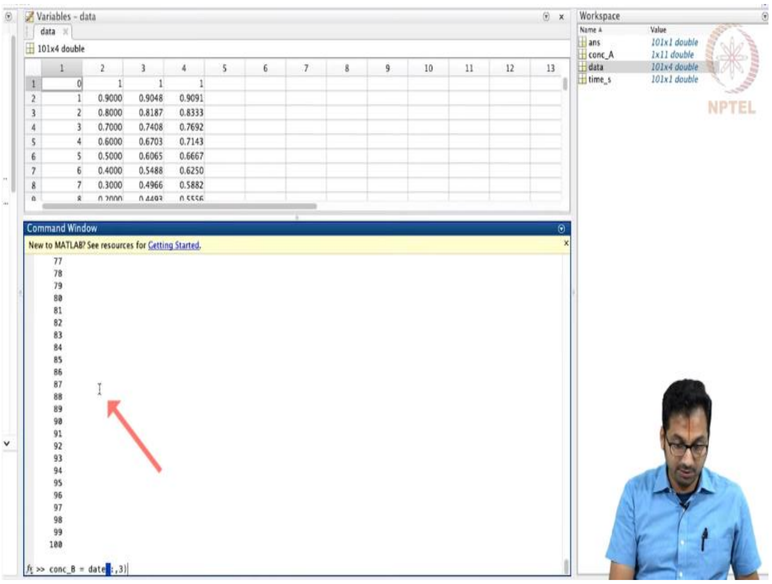
Name	Value
ans	'/Users/creates-...
conc_A	1x11 double
time_s	1x11 double

So similarly ok now that we have done this can we do the same thing for concentration of B. However if you remember we just saw the fact that if you are just taking the initial set of data points you could fit to something which fits reasonably to the linear plot however you will be misinterpreting a first order reaction as a 0 order reaction. So that is not something good so one way of doing it is to read the entire set of data so of data so how am I going to do it is by using this function called XLS are read.

So this is a predefined program within the software called MATLAB which can read excel files but before doing the same the first step that you need to do is to make sure you are the file set in the correct place. So what I am doing here is that I am using the command PWD which prints the directory that I am working in so this is like the way you might end up in a windows computer.

You might see it is a c colon slash windows slash downloads users slash downloads but in this case this is a user by name creates m3 and the files with in downloads is where you are sitting and if you say list the set of files that have the extension excel XS. Actually using this excel sheet to do our analysis.

(Refer Slide Time: 32:17)



101x4 double

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	1	1	1									
2	1	0.9000	0.9048	0.9091									
3	2	0.8000	0.8187	0.8333									
4	3	0.7000	0.7408	0.7692									
5	4	0.6000	0.6703	0.7143									
6	5	0.5000	0.6065	0.6667									
7	6	0.4000	0.5488	0.6250									
8	7	0.3000	0.4966	0.5882									
9	8	0.2000	0.4493	0.5556									

Name	Value
ans	101x1 double
conc_A	1x11 double
data	101x4 double
time_s	101x1 double

```
    77
    78
    79
    80
    81
    82
    83
    84
    85
    86
    87
    88
    89
    90
    91
    92
    93
    94
    95
    96
    97
    98
    99
   100

fx >> conc_B = date(:,3)
```

So now what we can do is that we can say data is equal to XLS read and then you can say 2019 the file name kinetics and then you hit enter. Then what you are able to realize is that MATLAB is already able to smartly read the different columns as well with the precision that you have may set the data set into write. If you had 4 decimals is going to read into 4 decimals. But what you are able to realize is that it has now converted the excel sheet into a matrix where the matrix has different lengths.

So basically you go until 100 seconds for all of this. So basically you have 101 data points including the 0 second. Now what we can try to do is to see how this data can be accessed so all of that is present within this variable called data. So if you say hit data and hit enter you are going to get this on the other hand you can also go to this workspace where all the different variables that you have are indeed set.

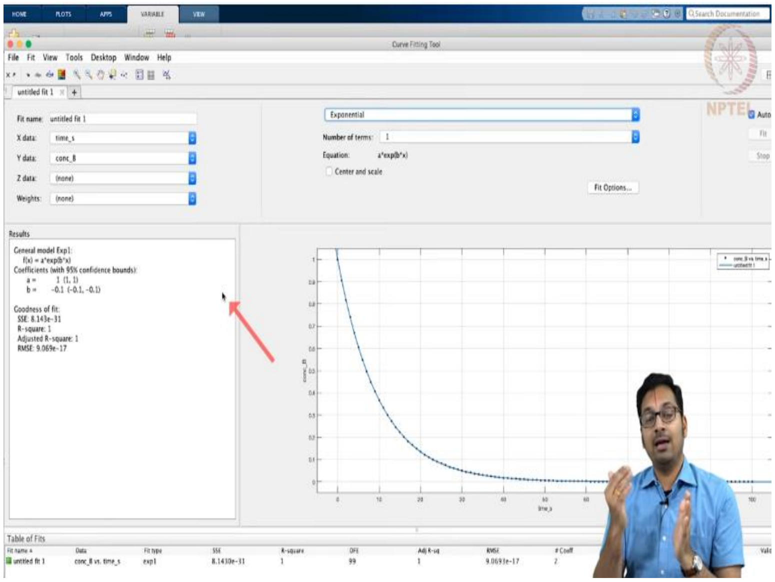
So if you open selection of data you are going to be realizing that the same excel sheet is what is displayed here. So this is kind of nice so that you can toggle between software that you are comfortable with and also use software which can do few of these automatedly. Now the exercise comes into the fact how I will define time from this. So basically if you are having 4 different columns I should be able to pick the certain column that I am interested in and the way we do it is that we say data bracket open colon comma 1.

This indicates just go pick all the elements from the first column if I hit enter what you are able to realize is that it is only picking the time axis here. So this is the nice way of handling the data now that your if your excel sheet is comprehensive different variables that have been carefully acquired this will be a very easy way for you to analyze the data. Let us take the example that we took earlier in the class where we were trying to understand what is the critical nicely concentration of a detergent and where you had spectrum that was acquired as a function of the surfactant concentration.

You are going to get many different fluorescence spectra that can be analyzed the same way because you started the excitation at the same wavelength and you had the emission recorded between 400 and 725 nanometers. So you are going to have all of that displayed nicely so you can call in different such files and read and do the data analysis by carefully calling only that given column. So now that we understand. So we can now redefined our time underscore s as that first column within that file the semicolon here what it ends up achieving is to not to display this data. If you wanted to display this data you can always say time underscore s and you are going to see that now the time underscore s has this variable.

Now let us go directly into concentration of B I can say concentration of B is if you remember the first column is time second column is concentration of A third column is concentration of B. So therefore what we have to do colon comma 3 so it is actually data not date. So now what ends up happening is that now concentration of B is saved into from the data a colon comma 3 it is now say save it into the variable concentration underscore B that is once again reopen the curve fit tool.

(Refer Slide Time: 35:29)



Fit name: untitled fit 1
X data: time_s
Y data: conc_B
Z data: (none)
Weights: (none)

Exponential
Number of terms: 1
Equation: a*exp(b*x)

```
General model Exp1:
    f(x) = a*exp(b*x)
Coefficients (with 95% confidence bounds):
      a =          1  (1, 1)
      b =       -0.1  (-0.1, -0.1)

Goodness of fit:
  SSE: 8.143e-31
  R-square: 1
  Adjusted R-square: 1
  RMSE: 9.069e-17
```

Fit name ▲	Data	Fit type	SSE	R-square	DFE	Adj R-sq	RMSE	# Coeff
untitled fit 1	conc_B vs. time_s	exp1	8.1430e-31	1	99	1	9.0693e-17	2

So now what we can do is to set it to time axis and then concentration of B so what you are able to see is the same set of plot that we saw a moment before. So instead of using a polynomial fit one can always change it to an exponential fit. If it exists and then try to see what works out and you are able to nicely see the blue line very beautifully passes through all the data points and you are able to get the same results that you obtained before that is the pre factor of the exponent is 1 and this the B which is the rate constant that you are getting here.

But the -k is k is equal to 0.1 and if it is beautiful and able to you are able to visualize what is going on so what we have done. So far is that if you nicely save your data into a software it can be recalled into different programs in this case I am just showing you an example of a spreadsheet and MATLAB which is a tool that you can use in order to plot your data on the other hand what you are able to realize is that exponential fits are nicely done and you can have different types of fit that one can do.

You can do a simple linear fit and see that the fitting is bad actually this is not even a linear fit let us go to polynomial and then let us say degree of polynomial is one you realize that the fit is extremely poor. Let us try to see how a second-order polynomial works okay. The r-square improves from 0.47. Let us take a look at it again the this is exactly the same r-squared we saw in the spreadsheet program.

So it goes from 0.48 to something like 0.82 so although the r-squared is getting better you are still able to realize that the points do not fall within the fitted blue line. So this is still not good let us try to take 3. Now we are starting to get closer but you are clearly able to see differences that come some of the data points are above the curve and some are below the curve but systematically they are off actually off on either side of the curve throughout the experiment the house r-squared has now become 0.9541.

Let us keep on increasing and see what ends up happening when you go to the power 4 the fit starts to definitely get better. But you are still able to see anomalies that come here where the blue line does not pass through the black data set the initial points are also not good and the final point for sure is going further away. Let us try to see what happen with 5 so as you are able to increase the polynomial nicely able to see what is going on.

But interestingly enough after the data point that they are acquired you realize that the number is going to negative and you know for a fact concentration cannot go negative. So one can keep on increasing to higher orders to see what ends up happening in fact here you are able to get an r-squared of 0.9999. So you might start to wonder right now. How is that an exponential fit that I wanted to get is also satisfied by a polynomial fit this is not surprising if you do a simple Taylor series expansion of e power X you get something like 1 + X square by 2 factorial + X cube by 3 factorial.

So X Square by 2 factorial cube by 3 factorial X power by 4 factorial and so on until infinity. So this is not a surprise as you keep increasing this set of numbers your R square gets become gets better and better and you are able to get something nice. But unfortunately your intuitive understanding of what is the fit is completely lost. So this is a very important point for one to understand although your model fits your data very well.

If it does not have a physical backing of why that model is chosen you cannot interpret your results as a quantitative chemist it is not just your duty to get your data in a reliable fashion with minimal errors. But also to fit it through and do the analysis with the proper background of which model has to be chosen. If you realize MATLAB is also trying to tell you a equation is badly conditioned remove repeated data points or try center and scaling.

So it is also trying to suggest what is going on we can try to see whether when you increase it to 8 you are starting to get things better. But still things do not improve until you go back to it exponential fit here it does not give you any such error right. So now what we are able to realize is that this type of data analysis can be easily done by using different software and how if you carefully and in an organized fashion save your data in such a spread sheet or any other format like a CSV comma separated value file you would be able to analyze this data.

I hope this week of lecture has helped you understand how to save the data spreadsheet how to analyze very simple linear fits within the spread sheet program. How to analyze some nonlinear fits in terms of the kinetics of reaction using the spread sheet program for 0 first and second order and also using MATLAB. As we go to the next week of the lecture we will be using these simple features of the programs in order to analyze our data to simulate some data.

So that so as to understand how we can sample better on the other hand also how to analyze our data with something that is a little more complicated than the simple lon with the linear relationship. Thank you very much.