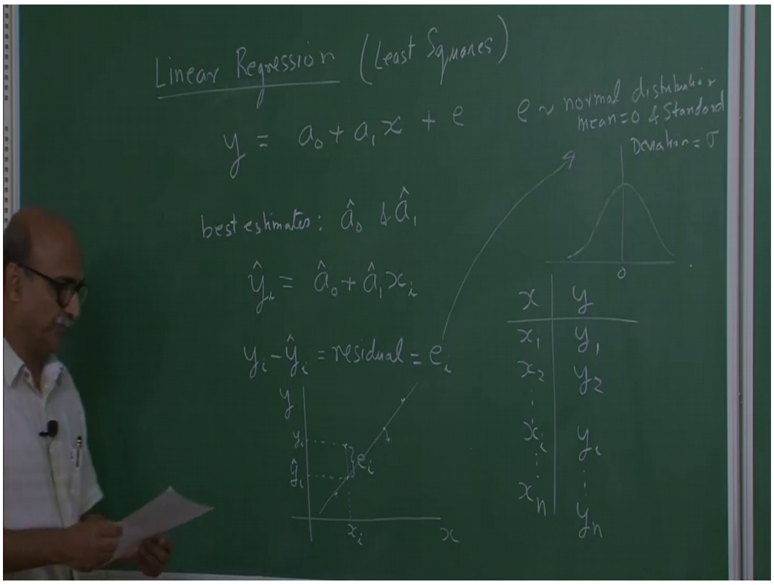
Heat Treatment and Surface Hardening - II
Prof. Kallol Mondal
Prof. Sandeep Sangal
Department of Material Science & Engineering
Indian Institute of Technology, Kanpur

Lecture – 32
Linear regression (least squares) method to find the value of n and k in Avrami equation

So, in the last lecture, we had we were discussing how to analyze experimental data with respect to the Avrami equation, where we have a set of experimental data points of the phase transformed fraction of the phase transformed as a function of time and based on this data one wants to determine the parameters n and k. For this purpose, we had linearized this equation the Avrami equation and then reduced it to a simple linear equation of the form y equals a 0 plus a 1 x, where x was simply log of the time and y was a double log log of minus log of 1 minus the fraction transform.

(Refer Slide Time: 00:47)



So, we transformed the experimental data as well to these quantities that we did in the last lecture then x versus y if you plot it should come as a straight line. Now the objective here is that we want to determine the best values of a naught and a 1 which best represents a straight line through these set of points x and y.

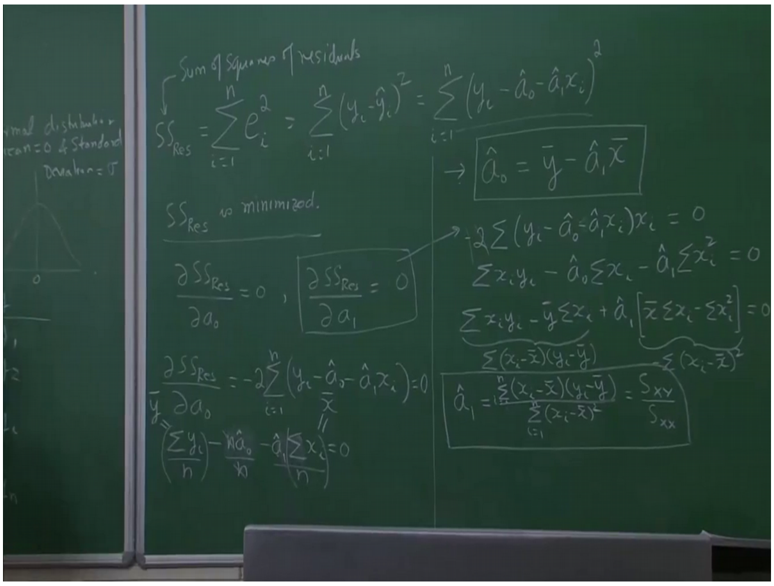
In order to do this we will use the method of linear regression or it is also called as simply least squares technique and one shall see how that is done. Now in any equation of this form where experimental data is also involved, there is always a statistical variation of values. So, such a equation is also we add an error term to it, this error is simply a statistical fluctuation of the response y when the independent variable x takes up a certain value this we assume that this error follows a normal distribution the normal distribution is a bell shaped curve with mean of 0 and standard deviation of sigma. So, since it is a an error term mean is 0. So, the average of all the errors put together should reduced on to 0.

Now, we wish to estimate or the best estimates the best estimates for a 0 and a 1, I will denote them as a naught cap and a 1 cap once these best estimates have been made then the predicted value of y from these estimates we will call that as y cap which is equal to a naught cap plus a 1 cap x. Now if this is the i th value of the independent variable like x i then this is the i th value response. So, I have a set of data here x and y. So, I have x 1 responses y 1 when x takes of the value of x 2 the responses y 2 and so on if it at the response of the i th value of x is x i the response is y y i and so on up till n if you are if I have n data points that I have collected. Now this is the predicted value at the i th response and here this is the measured value at the i th value of the independent variable x that I have experimentally setup in this case it is log time. So, I have this which means I can write down the difference between the measured y i and the predicted y i cap that I obtain.

If I have the estimates for a 0 and a 1 this difference is called the residual and I will denote it as e i which is actually related to this e that I was discussing and this e i as well would follow a distribution with mean equal to 0 and standard deviation equal to sigma graphically one can understand what is happening here if I just plot my data points and through this data I draw the best fit line and I look at different value; suppose this is my i th this is x i, then this is y i while this point on the line is y i cap or the predicted value and this is e i the difference between y i and y i cap e i will be positive if y i is above the line the measured value is above the line; e i will be negative which in this case in this case this is the measured value or the observed value and this is the predicted value in this case e i would be the residual would be negative.

So, some of the residuals are going to be positive while the other residuals are going to be negative. So, what we do is we do not work with a residual e i, but we square the residual. So, instead of because some of them are positive and some of them are negative if we square them then everything will become positive.

(Refer Slide Time: 07:19)



So, e i square and then I sum up all of the individual residuals from i equal to 1 to n, this sum I will call as SS sum of the squares of the residual. So, let me just make a point here sum of the squares of residuals.

Now, or n let me just further expand this instead of writing e i, let me write it as y i minus y i cap squared, we want to find now a naught cap values for or the estimate for a naught and a 1 is a naught cap and a 1 cap such that S sum of the squares of the residual SS residual is minimized. So, we need to minimize the sum of the squares of the residuals. So, how do we do this how does well it is SS residual is a function of all of this. So, we need to minimize it by taking certain derivatives and setting them to 0. So, what we do let me expand this further y i minus now instead of writing y i cap i expand y i cap as a 0 cap plus a 1 cap x i and hence this will become a 0 cap minus a 1 cap x i. Now in order to minimize this sum of the squares, I need to differentiate SS residual with respect to parameters a 0 cap and a 1 cap. So, getting the derivative dell SS residual dell a naught set it to 0. Similarly, I take the derivative of sum of the squares of the residual with respect to a 1 and set as well to 0.

So, let us take the derivatives, now if I first let me take the derivative of with respect to a naught. So, if I do that and this should be square here I look at this function and take the derivative with respect to a naught only keeping a 1 cap as constant this will taking a partial derivative this will become minus 2 that comes from the power and a 2 comes from the power the negative sign come from taking the derivative inside well derivative of a with respect to a naught cap would be minus 1 and hence I get minus 2 summation and the stuff inside remember we are taking the derivative with respect to a naught and a 1 and not with respect x this should be set to 0.

If I expand this further individually this 2 and negative sign will go and what I will have is summation y i minus summation a naught cap minus summation a 1 cap x i equal to 0 well a naught cap is a constant and hence it is in a sense that in the this comes out of the summation and hence this whole thing basically you are adding a naught cap n times. So, this is simply become n times a naught cap where n is the number of data points this a 1 cap comes out of the summation and can be written as a 1 cap out here, then I can divide this entire equation everywhere by n.

So, I divide this by n, divide here by n, divide here by n and the right hand side remains 0. So, this n and this n; this n and will cancel at what i am left with is this term what is this term this is the average value of y simply add up all the y S divide by n giving you an average y similarly this term here summation x i upon n is an average value of y. So, let me call this value as y bar and let me call this value as x bar from this it is easy to write that a naught cap equals y bar minus a 1 cap x bar. So, this is one important equation, I have which relates a naught to the average y which that is average of the all the measured values average of all the independent values that I got from the experiment while a 1 cap I still need to determine.

So, that would come from this relation that we take the derivative of the sum of the squares of the residual with respect to a 1. Now when I do this, I get y again differentiate, but this time I differentiate with respect to a 1 cap and I will get the following result minus 2 summation y i minus a naught cap and this time this minus is coming from here when I differentiate a 1 cap, but i will be left with minus x i. So, minus i come out of the summation, but x i will remain in the summation as you will see. So, I am let me first write down the whole thing a naught cap minus a 1 cap x i times this x i that is coming when I differentiate with respect to a 1. So, x i will be left and it will be remain inside the summation this would be set to 0.

I can forget about this 2 and this minus sign, I am just expand this, when I do this, I will get summation x i y i minus a naught cap summation x i minus a 1 cap summation x i square. Now in order to this is equal to 0, now in order to solve this, I can substitute a naught cap from this; this result can be substituted into this equation and when I do this, I get summation x i y i minus y bar.

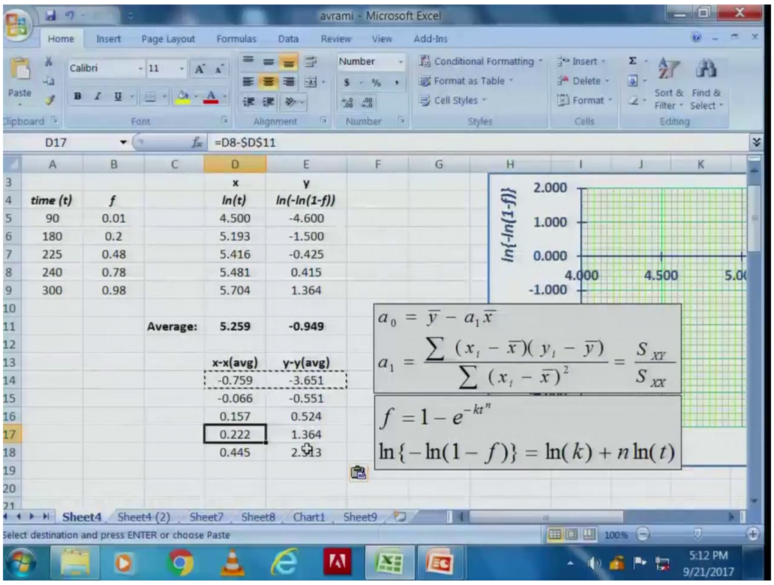
I am put coming here. So, I will get y bar summation x i and then a 1 cap will also come here as a result, I will get plus a 1 cap and I can take this brackets here and I can write x bar summation x i minus summation x i square and this is equal to 0. So, now, I have an expression in which only the parameter a 1 cup is there which means I should now be able to solve for a 1 cap from here and then that value can be substituted here to get a 0 cap.

Now, I am not going to prove this, but it is somewhat longest, but we can show that the term here in square brackets can be written as minus summation x i minus x bar square while the term this term can be written as summation x i minus x bar multiplied by y i minus y bar the way you can prove it is you expand this and eventually show that this is equal to this same thing do here you expand this and show that this is eventually equal to this term from this we can find right a 1 cap as summation x i minus x bar times y i minus y bar divided by summation x i minus x bar square where the summation is running from i equal to one to n i equal to 1 to n.

Now, just we can abbreviate this and write this expression as the numerator summation x i minus x bar times y i minus y bar a, I will write it with the symbol S x y sub S subscript x y and the denominator, I will write as S subscript xx. So, this forms an important equation which from the experimental data of x and y i will get a 1 cap. Similarly, once I get the value of a 1 cap, I will put it plug it into this expression and I know what is y bar and x bar and hence I will get the value of a naught cap. So, these are the 2 best fit values of the a model or the linear equation a naught plus a 1 x. So, once I have this, then I will have the straight line and I can draw the line and see how that line fits the data.

So, what I do now is actually look take our data of the eutectoid steel when fraction transformed as a function of time and see how we can fit the straight line through that data.

(Refer Slide Time: 21:18)

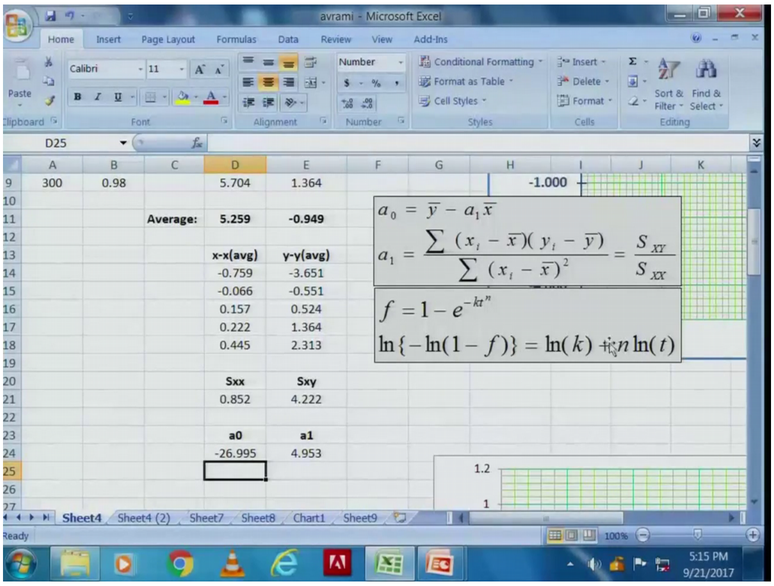


So, now let me start to look at this data. So, I have the same excel sheet that I had earlier in which I had already transformed log t and logoff minus log one minus f. So, let me this I will call it as x this is my x and this is my y and my equation is a naught plus a 1 x at these are the equation that I just wrote down on the board a 0 is equal to y bar minus a 1 x bar and a 1 x is equal to a summation x i minus x bar times y i minus y bar divided by summation x i minus x bar square which I have every abbreviated as S x y upon S xx. So, the one of the first things that I think; I need to do in order to calculate all this I need to give the average y and average x. So, let me write down first the average values. So, my x bar average is simply I can use the average function of this spreadsheet and this 5.259 that is shown is the average of all of these values. Similarly I will just copy this formula here and this gives me the average of all of these y values.

Now I need to calculate S; oh, before I calculate S xx I need to calculate x i a set of values x i minus x bar and a set of values y i minus y bar. So, let me call this as x minus x average and y minus y average now this set of values would be de e f the first x value is in cell D 5. So, D 5 minus the average x and I will anchorite. So, that it does not change when I copy the formula this gives me the first x minus x average value the second or rather the first y minus y; y average the corresponding y minus y average value would be the first y value which is e 5 which is 5 minus the average of all the y S anchorite and thus gives me the first pair. Now I just have to copy this at populated.

So, I have now just I had these 5 data points on I have correspondingly 5 sets of x minus x average and y minus y average.

(Refer Slide Time: 24:50)



Now, let me calculate S xx and S x y. Now S xx is the sum of the squares of the deviations x i minus x bar square. So, this is simply equal to the function called sums squared what it will do is the a set of values that are given in the arguments each of them are squared and then summed and that should do the purpose because I need to square these x minus x average values first and then add them up. So, this gives me the denominator of this equation for a 1 and then S x y, I will get by another function in excel which is sum of the products of the corresponding sum of the product of the of the corresponding a values. So, the first set of values x minus x average comma I give the second set and what this spreadsheet will do it will correspondingly first multiply this and this add to the product of the second pair and the third pair and. So, on this gives me S xx and S x y.

So, now I should be in a position to determine the parameters a 0 and a one. So, let me write a 0 here and a 1 here and below these, I will calculate the values of a 0 and a 1 well first I need to calculate a 1 then only I can get a 0. So, a 1 is equal to S x y upon S xx. So, S x y is this e 21 divided by the value in the cell D 21 which is S xx and this gives me the value of a 1 of the slope 4.953 a very close to 5. Now I will estimate a naught which is the average x average y, y bar; this is the equation minus value of the upper estimate a 1 multiplied by the average value of x and this gives me the value of minus 26.995. Now this is in a nutshell a sigma, the method of linear regression; we will look at this method a little bit in more detail as to how good the fate as and so on and in the next lecture. So, I will stop here; in the next lecture, I will discuss the values of n and k that I get out of this linear regression and look at how well the data fits to the experimental data points.