Design and Optimization of Energy Systems

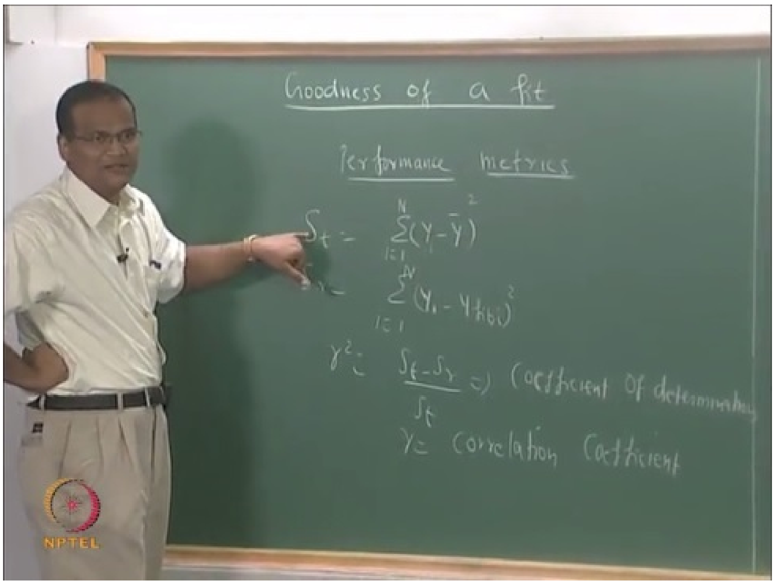
Prof. C. Balaji
Department of Mechanical Engineering
Indian Institute of Technology, Madras

Lecture No. # 18

Least Square Regression contd…

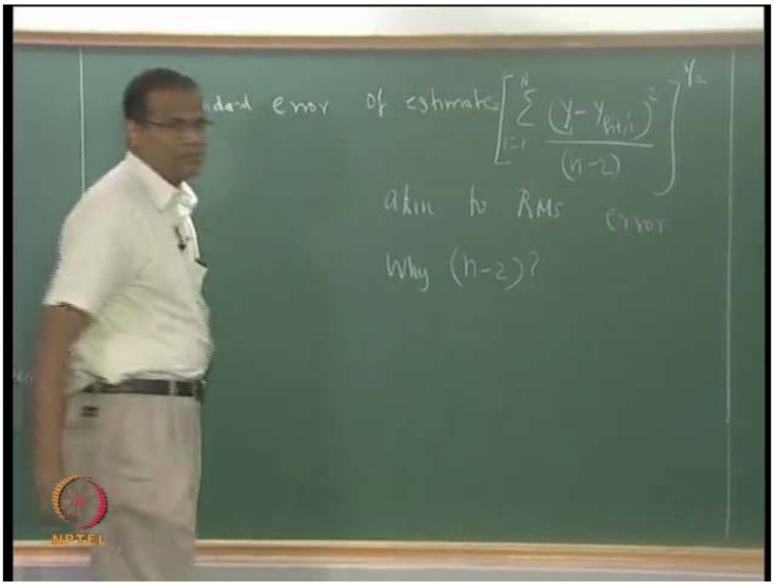
We are looking at least square regression. We worked out a problem in the last class. There are some more performance metrics, which we need to discuss. All these come under the category of basically characteristics, which determine the goodness of a fit; how good or how bad a fit is.

(Refer Slide Time: 00:36)



The performance metrics are; first, before the correlation, you had all the data and you had the Y bar, that is, Y mean. The sum of the deviation from the mean of all the data; this is S t; this you can call it as total, sum total or whatever. S r is a residue, which is Y i minus Y fits whole square sigma. Now, the coefficient of determination is basically, what is the percentage or how much reduction is brought about in S t by proposing a correlation wherein you fit Y as a function of X. And, root of this coefficient of determination is a correlation coefficient.

(Refer Slide Time: 01:23)



There is one more perform… There is one more quantity of interest, which is basically the standard error of the estimate. The standard error of the estimate is Y i minus y fit i whole square divided by n minus 2 whole to the power of r to whole to the power of half. It is something like an RMS error. RMS error is Root Mean Square error. So, there are n data points. So, you… It is basically that, numerator is actually (Refer Slide Time: 02:14) S… The numerator is actually S r. But, why are we dividing by n minus 2 or not n. What is this n minus 2 d? Suppose we ask this question… Any guesses?

Student: There are 2 points (( ))

2 points?

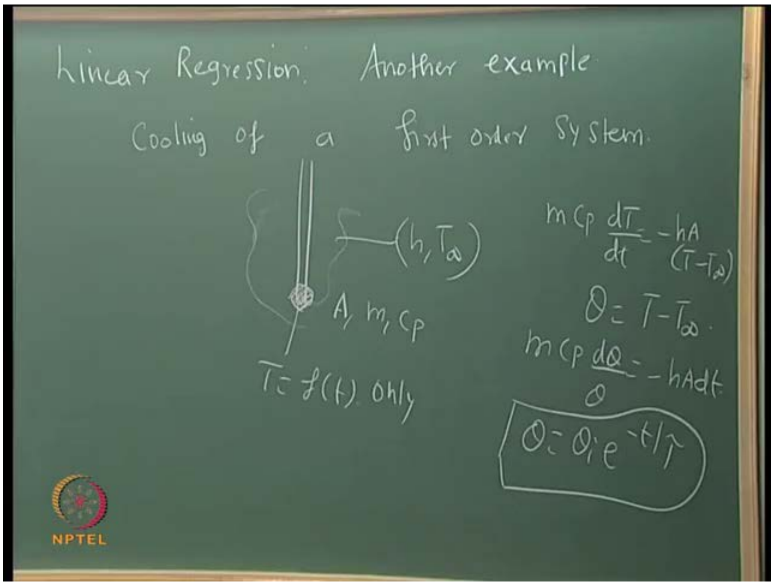
Student: Be an exact fit.

Correct, if it is a straight line; if Y is equal to AX plus B; if you have 2 points; there is no question of the error. It will be an exact fit. Therefore, this formula breaks down or this formula is meaningless if you try to do a linear fit for 2 points. So, when n is equal to 2, the denominator becomes 0, becomes infinity and all that; which does not mean that standard error is infinity; that means stupid, do not use this formula for n equal to 2; that is what it says. So, we want to be careful. So, n is much much greater than 2. Usually in experiments, we have 20 data points, 30 data points and 40 data points. Suppose I develop a correlation – Nusselt number is a function of Reynolds number, Prandtl number; I have 40 or 50 data points. Instead of dividing by 48, if I divide by 46; it will not make much difference. I will get the same standard error. So, this n equal to n and n minus 2 looks so different if n is very small. If n is large, it does not matter at all.

Now, any correlation, which has a high r square, is it necessarily a good correlation? Any correlation, which has a high r square, is a good correlation purely from a statistical prospective. But, we do not know whether there is a there is a causal relationship between X and Y. For example, the number of rainy days in a year; you put it as a fraction. Number of rainy days in India in a year; number of T 20 matches won by India in a particular year. And, look at the data for the last 5 years. It may have a very beautiful correlation. But, the correlation is not there; you are just trying to plot two irrelevant things. So, that is called as the spurious correlation.

The other frequent example is salaries of everybody… Salaries of all the people are going up. For example, if you take the prof salaries, it may not go at the rate at which we desire. But, let us say profs salaries are going up. You plot last 10 years; it will show an increasing trend. You take a completely arbitrary quantity; hopefully an arbitrary quantity; consumption of alcohol for example. Last 10 years, it is also growing up; it may beautifully correlate. We can ask for 0.958; it does not; you cannot start writing a theory and write a paper based on that. So, you have to know what are the variables you are correlating; whether that Y is really a function of X. That you will have to figure out. This can come when you do experiments or when you are looking at the non-dimensional form of the equations; they are arriving at the variables and so on. So, you have to be very strong in physics to figure this out. Statistically, you can generate many correlations and fits and so on. But, you have to be careful. The physics should be there. All right?

(Refer Slide Time: 05:48)

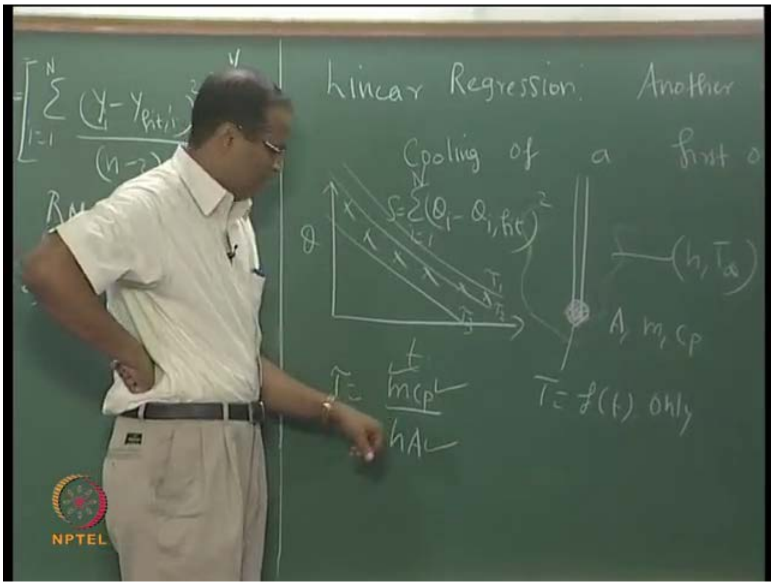


Now, let us look at… I briefly talked about this in the last class – cooling of a first order system, mercury in glass thermometer kept in a patient’s mouth and so on. If you have something, which has surface area A, mass m; and, outside you have h and T infinity. This is the thing under consideration. This you can take as the supporting rod or thermocouple leads or whatever. So, this is under consideration. The whole thing is T is a function of temperature only. The governing equation is… This is called a first order system. There are some inherent assumptions. The heat transfer coefficient is constant; it does not depend on the temperature of the body; which is a questionable assumption.

The mass is constant; specific heat is constant – all these are not questionable assumptions; these are all ok. But, the heat transfer coefficient is constant. Under natural convection, this is not so. The heat transfer coefficient is really variable. But, we keep it constant. Next – because the body is cooling and it is dissipating the heat to the surroundings, the temperature of the surroundings is not increasing. That is why it is the first order system. But, we can complicate this by having a second order system. How can I have a second order system? For example, there is a young child, small child, infant, is crying for milk. So, the mother is trying to prepare baby food. So, she is putting lactogen or whatever and hot water; and then, she is preparing; it is too hot. So, what they first do is they will pour this to see the temperature; they will check the temperature if you have seen carefully. First they will check the temperature; if it is too hot, what they will do is; you put it in a bucket or a tub full of water.

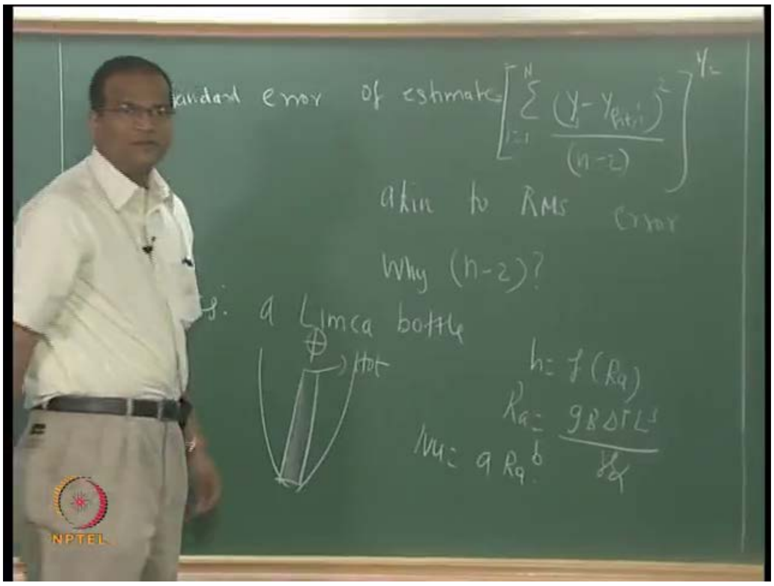
Now, it is m c p d T by d t is minus h A delta T. Unfortunately, the T infinity is the temperature of the surrounding water. That fellow will start going up, because this heat is going there. Now, we can model it as the second order system; we can work out the solution for that also. But, this is not a heat transfer class. We can do that in a heat transfer class; it is possible to do. That is the second order system, where T infinity keeps changing. But, in this case, it is relatively straightforward; we can keep T infinity constant. And, what is the goal of this?

(Refer Slide Time: 09:25)



It is possible for you to estimate the time constant, if you have a temperature versus time history. If you have a temperature versus time history; basically, the funda is very simple; you can generate different curves for different values of tau 1, tau 2, tau 3 and so on. There will be one particular value of tau at which sigma theta i minus theta i fit whole square; i is equal to 1 to n, is minimum. That is the best estimate of the time constant for the system. From the time constant tau, if you know m, if you know c p, if you know area; it gives you a very simple and inexpensive method to determine the heat transfer coefficient. So, this is the measurement of heat transfer coefficient by conducting unsteady heat transfer experiments. But, the assumption here is; the heat transfer coefficient is a constant. Unfortunately, the heat transfer coefficient is not a constant. The heat transfer coefficient is usually a variable, because the heat transfer coefficient depends on the temperature difference, that is, the Rayleigh number. You remember?

(Refer Slide Time: 10:42)



For example, let us take a Limca bottle. For argument sake, I will make it a cylinder like this. Now, it is hot; I want it to cool down. So, I keep it in the refrigerator. How will the boundary layer develop? Top view is like this. Natural convection boundary layer; how will you develop? What… It will develop like this. What will happen is; this boundary layer – the heat transfer… So, the boundary layer is actually an impediment to the… The boundary layer is an impediment to the heat transfer. The thickness of the boundary layer, rather the heat transfer coefficient, is inversely proportional to the thickness. So, if you are somehow able to get the average thickness of the boundary layer, the heat transfer coefficient is inversely proportional to the average thickness. So, we do boundary layer theory; we do solutions of (( )) equations; we do experiments basically to get this boundary layer thickness, so that from boundary layer thickness, we will get the heat transfer coefficient.

The heat transfer coefficient is going to be a function of the Rayleigh number. So, the Rayleigh number is g beta delta T L cube by nu alpha; L could be the length or it could also be the diameter; that we will not worry. The appropriate length scale has to be taken. This delta T keeps decreasing for a problem like this (Refer Slide Time: 12:22). So, I you want to appropriately model this, you will say that, the Nusselt number goes as a into Rayleigh number to the power of b. If you model this as Nusselt is equal to a Rayleigh number to the power of b, this Rayleigh number has got the temperature delta T; delta T is T minus T infinity, which is nothing but theta. Therefore, in this time constant, you have to substitute for this heat transfer coefficient as Nusselt number; and finally, declare that, the unknowns are a and b. And then, using the least square regression, it is possible for you to get a and b. This Sparrow did. And, he wrote a paper 30 years back. It is in the ASME, Journal of Heat Transfer.

But now, what my students are doing is; if there is a convection and radiation, how to simultaneously estimate Nusselt number as well as emissivity; or using unsteady, can I simultaneously get c p, if I do not know what the material is; c p as well as h; or, if this is not a lumped capacitance system; temperature is a function of x and t; can I simultaneously get alpha and emissivity and so on. So, these are all inverse heat transfer problem; where, you actually run some experiments; you know that, there is a mathematical model and use some least square minimization or least square regression and fare it out or invert or retrieve the parameters.

Now, we will solve a small problem on this, which I composed before coming to the class. So, hopefully it works. Problem number 18; Akshay, what is it? Problem number 18 – consider the cooling of a first order system of mass m… Consider the cooling of a first order system… Consider the cooling of a first order system of mass m… Consider the cooling of a first order system of mass m, specific heat c p and surface area A… Consider the cooling of a first order system of mass m, specific heat c p and surface area A at an initial temperature of T i; at an initial temperature of T i degree celsius; T subscript i. Such a system is placed in cold quiescent air. What does it mean?

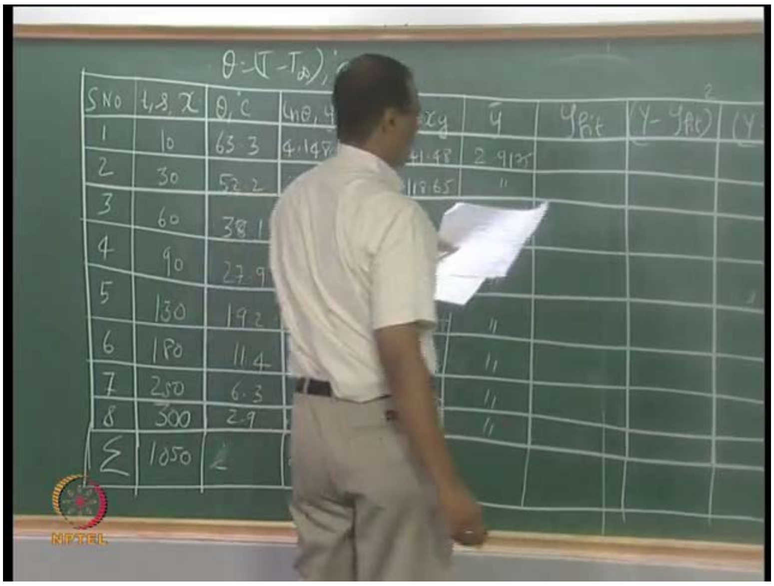
Student: Stationary.

Still air. Such a system is placed in cold quiescent air at T infinity equals 30 degree celsius… at t infinity 30 degrees celsius with the heat transfer coefficient of… with the heat transfer coefficient of h equal to 8 watts per meter square kelvin. T infinity is equal to 30 degree celsius, h is equal to 8 watts per meter square Kelvin. From the experimental data given below… From the experimental data given below… From the experimental data given below, determine the time constant tau. From the experimental data given below, determine the time constant tau and the initial temperature t i. Evaluate the standard error of the estimate. The standard error is in the estimate of the temperature. Evaluate the standard error of the estimate and the correlation coefficient.

Evaluate the standard error of the estimate of the temperature and the correlation coefficient.

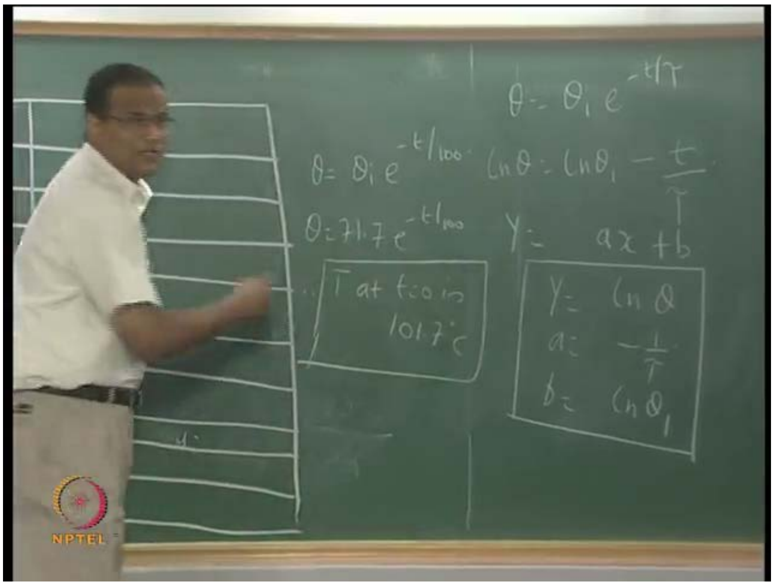
The standard error is not in the tau, because this correlation is basically you are… You will get a correlation to get the temperature. Whenever you use that correlation, you get a temperature; it is associated with the standard error. So, first, you have to do some tricks to convert it into a linear regression. It is not a power log form; it is an exponential form. I will give you the data. So, you have to open out a 70 mm column; 1 2 3 4 5 6 7 8 9 10; I have 10 columns – Y, Y bar, Y fit, Y minus Y fit, Y minus Y bar; so many columns are there.

(Refer Slide Time: 18:05)



So, 10 seconds; 10, 30, 60, 90, 130, 180, 250, 300; theta is T minus T infinity. So, you can start with this.

(Refer Slide Time: 20:11)



Correct? First, reduce it to the standard equation of a straight line. I can straight away call this is as x (Refer Slide Time: 20:59). Now, I have to open a column – ln theta; this will become y. Correct? Then x square, xy, y bar, y fit, Y minus y fit the whole square, Y minus y bar the whole square. So, what is the last row?

Student: Total; summation total.

Sigma of all.

Even in the quiz and the exam, you will have to show all the steps. Even if you have a calculator, which does all these; I want all the columns, I want all the entries to be filled up. Is it right? Please go ahead; I will just take attendance. Next? 1050; good; sigma y?

Student: 23.36

Here you have these basic things. Now, you have to get x square and xy. How many of you got the x square?

Student: (( ))

Some 2,11,260 something? So, x square is 100. Correct? How much is it?

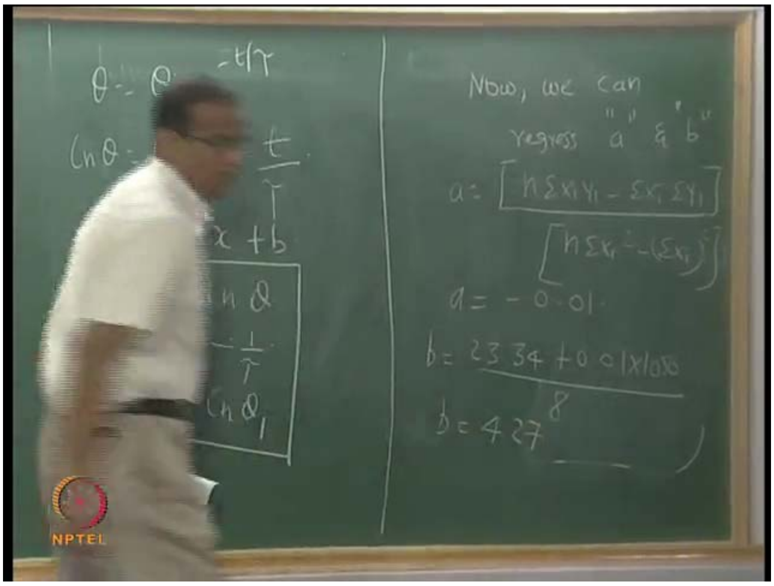
Student: 214500

214500. xy? 2275.3… 22?

Student: 2280

2280. When you take this log and all that, it depends on where you stop now; second decimal place or… There will be some small variations; that is all right. Time constant – instead of 200 seconds, you may get 198 or 203; it is ok. You should not get 2000 instead of 200. All right? Now, with this, you can work out a and b. Correct?

(Refer Slide Time: 25:58)



Now, we can regress a and b; or, regress for a and b; whatever. How much is it, Deepak?

Student: Minus 0.01.

Minus 0.01. How much is it?

Student: 4.

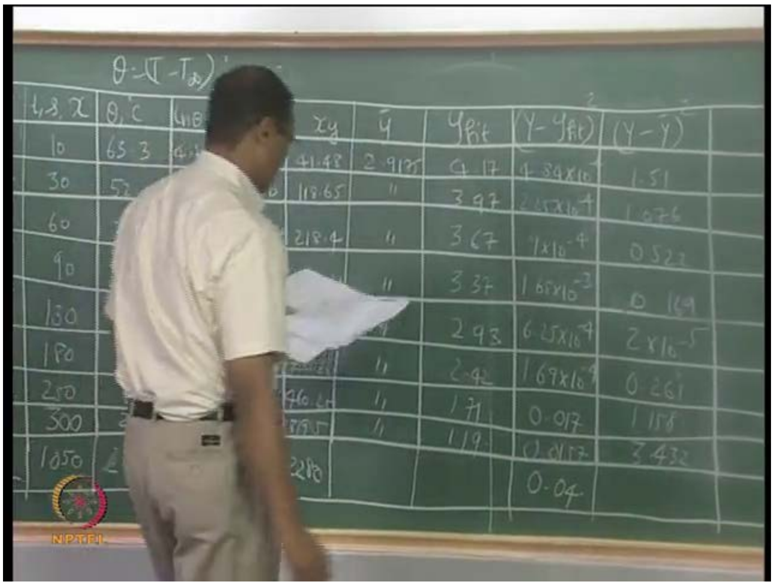
4.?

Student: 4.27.

Is equal to 4.27. I want each one of you to get this; pretty straightforward. So, in the days before regression, before Gauss, probably people plotted the thing on a log sheet or… And then, they took as scale and they drew an approximate curve and got all these intercept and all these. But, now, we have advanced regression and all these. There is no need to do all. There is no need to do that desi kind of curve fitting and what you do in school; trying to get the intercept, slope and… You can get accurate values. Of course, you can mathematically complicate the model further and further. And, beyond a certain point, it is not possible to do hand calculations; you can write computer programs. There are plenty of tools, which are available. MS-excel can do little bit of regression. You have got data fit for windows; you have got axiom. Matlab can do it; you can have sigma plot. There are so many software packages, which are available.

There is something called data fit for windows, which has a 30 day free trial. Any fit between x and y you want; you name it; you put fancy fit – y equal to a to the power of b to the power of t square plus a b t square; something; whatever you want you put – a, b, c, d; it will just give out the values of a, b, c, d and all that. So, non-linear regression is possible. Tools are available to do non-linear regression. But, basically, the physics must come from the user; you must know that y goes with x or t in this particular form. Once you have this form, then if you have done good experiments or you have got good simulation data, you can use the simulation data or experimental data with a powerful equation and go ahead and do curve fitting or regression and get the best estimation.

(Refer Slide Time: 29:38)



Now, you can get the y bar. What is y bar? 2.9175. y fit – you can invert it. So, finally, theta is equal to theta i into e to the power of minus t by 100. What is theta i?

Student: 71.7

70?

Student: 71.7

Point?

Student: 7

Therefore, this is where the question is posed. What is the initial temperature? Please look at the question. What is the initial temperature? See initially, it is cooling down like mad. Exactly at the zero-th second, we do not know what the initial temperature is. That is why it is always subjected to error. So, you can use regression and go back and find out this theta i. If you have already measured this, then you can verify whether the measurement you have made is accurate or not, because whether it collaborates with your findings from the regression. Are you getting my point?

It is very difficult to measure the initial temperature to be exactly… When you are trying to measure, already one second has elapsed. How good is the time constant? When you look at the stop watch and try to look at the thermometer, that 0 is a critical point. So, we will leave. Normally, because we are having so many data points, it is possible for us not only to retrieve tau; it is also possible for us to retrieve theta i. Then, you have an additional measurement. Suppose it says 100 degrees or 100.5 or something and this says 101; it is all right. So, they are of same order. So, y fit… Can you tell me the y fit values please? 4.?

Student: 17

4.17

Student: 3.

97

Student: 3.67, 3.37, 2.97, 2.42, 1.71, 1.19

1?

Student: 19

y minus y fit? Is it ok? 4.17, 3.97, 3.6, 3.3, 2.9; 2.90 you are getting; is it?

Student: 2.93

Please do these 2 columns; you will get s t and s r and then you can get the correlation coefficient. Can you tell me these values? Please do not feel lazy; complete it. You still do not know whether your estimates are good or not; unless you fill… unless you work out the last 2 columns, you will not know whether the correlation coefficient is good. Somebody can tell me. Vikram?

Student: 4.84 into 10 power minus 4

4?

Student: 84 into 10 power minus 4

4?

Student: 2.25 into 10 power minus 4; 9 into 10 power minus 4; 1.68 into 10 power minus 3.

Varun?

Student: 6.25 into 10 power minus 4.

Minus?

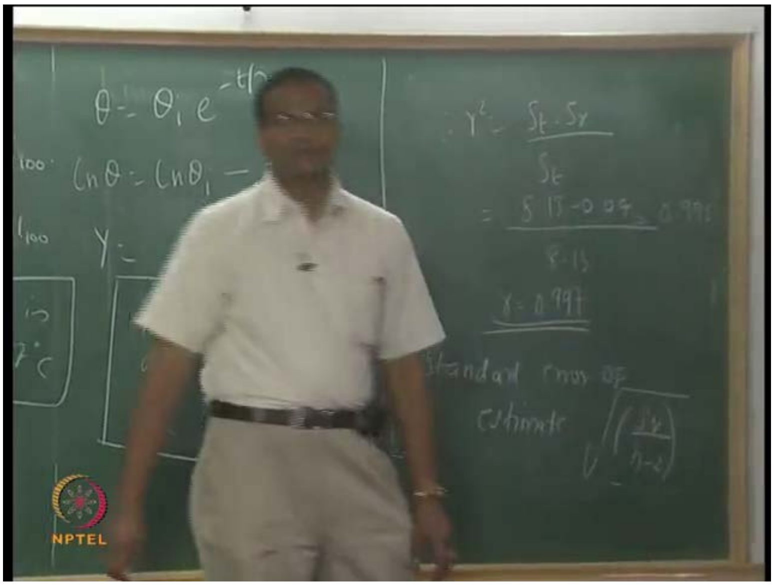
Student: 4; 1.69 into 10 power minus 4; 0.017

Last one?

Student: 0.0157

Get the sigma. 0.05, 04; how much is it? 0.04; good. Now, y minus y bar is easy. 1.51, 1.076. Is it correct? 0.522, 0.169, 2 into 10 power minus 5, 0.261, 1.158, 3.432. 8.1; is it? 8.13.

(Refer Slide Time: 35:53)



How much is it? 0.997?

Student: 0.997

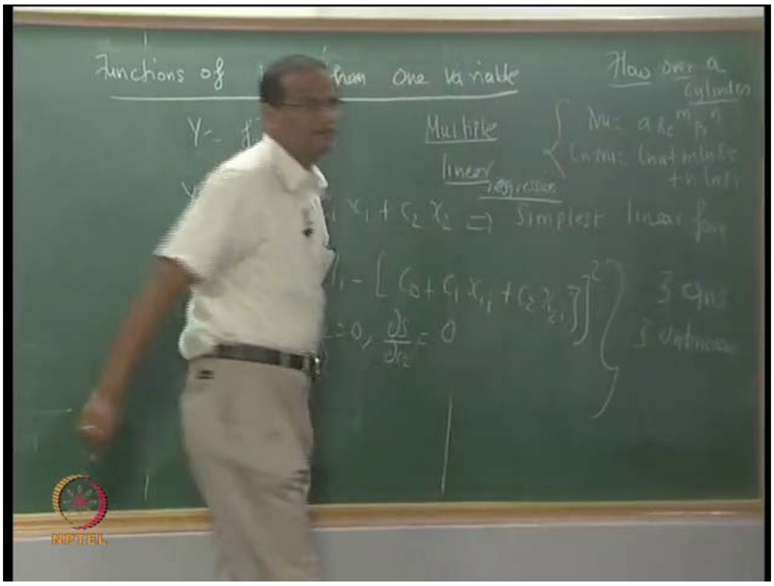
r? 0.997. It is a very terrific correlation. The correlation coefficient is very high, very good. So, it is a very accurate representation of the… It is a very accurate representation of the data; that means if you are able to regress the data; if you propose the data can be represented as e to the power of minus t by tau; then, 99.7 percent of the variance in the data can be explained by proposing this curve. Standard error of estimate – square root of s r by h minus 2. How much is it? 0.?

Student: 8

The standard error of the estimate is 0.08; about 0.1; which is ok; that means at a particular time t, if you are using the correlation and obtaining the theta, there is a 99 percent chance that this theta will lie between this mean and plus or minus 3 times 0.1. Correct? There is a 99 percent… You can have 99 percent confidence that, the theta will lie in that estimated value plus or minus 3 times 0.08. So, that gives you the confidence level with which you can predict. So, this is also an additional measure. It is an independent measure. This is an independent measure of what is the confidence you have in the estimates. Since it is 0.08 when theta is high, it is not a very serious… Theta is high if 0.08 is small. So, it is ok. But, when the theta is going down, then this will become comparable. Therefore, if you want to do the experiments intelligently, you must try to use the (Refer Slide Time: 38:47) maximum data when the system is hot. When the system is approaching the temperature of the surroundings; when the theta itself is very small… Are you getting the point? When sufficient time has elapsed; since the driving force for the heat transfer itself is small, it is more error prone; other effects may come into play.

In fact, you can subdivide the data into 3 sets: early part, middle part and final part. For each of these parts, you may get a different value of tau. These are all additional things when you do experiments, design of experiments and this things, you will learn all these. If you actually try to measure for some company or you are trying to do some research and all these, these are the additional factors, which you will have to consider. So, this will be a typical question in the quiz or this thing. So, it will take about half an hour; you patiently do. I will give you about 8 or 10 data points. You will have to do this patiently and… Maybe it would not be the linear form; I can give some other… I can give you some other forms. So, it may be possible that, I will give you some sentences in English; from that you convert it into mathematical form and get the equation – theta is equal to theta i and then use this data. It would not be plug and play; I will not say. Then, it will be a just mathematics class. x is known to vary with y like this; x and y are given. Find out the estimate. Where is that design and optimization of thermal systems or energy systems there. That modeling is also part of the course.

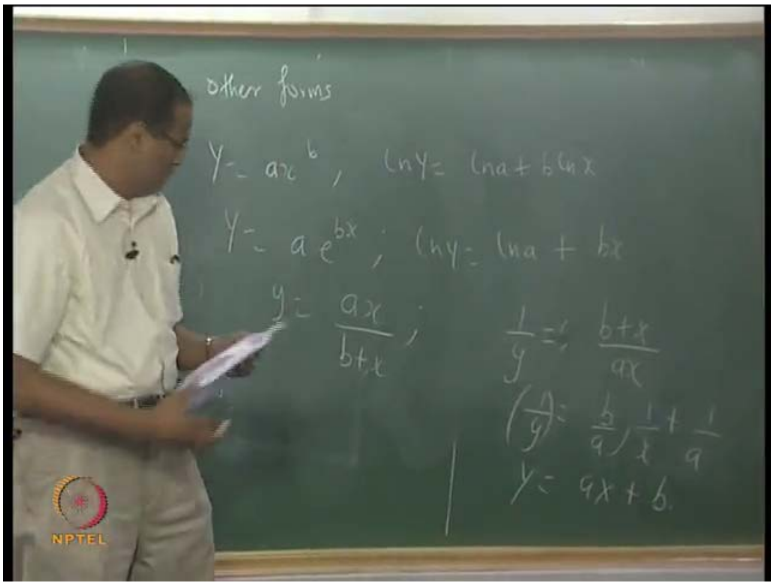
(Refer Slide Time: 40:46)



Now, there are some other forms, which are amenable to regression – linear regression. We will see those forms. Let us look at functions of more than one variable. Y is the function of f of x 1, x 2. The efficiency of an IC engine or the efficiency of fan and duct system; the function of two three parameters: flow rate, this thing, head and so on. Now, suppose I propose that Y is equal to C 0 plus C 1 x 1 plus C 2 x 2. This is the simplest linear form. Correct? Now, propose S. Are there too many brackets? So, that is y i minus y fit. This has to be minimized. So, I can do… 3 equations; 3 unknowns.

What is the immediate provocation for trying to do understand a form like that? For example, flow over a cylinder. If you have flow over a cylinder, the Nusselt number goes as a Reynolds number to the power of m Prandtl number to the power of n. This is also turbulent flow over a pipe; Dittus-Boelter equation, Sieder-Tate equation. If you take ln Nu… Is it a linear form? Y is equal to C 0 plus C 1 x 1 plus C 2 x 2; where, x 1 is Reynolds number; x 2 is a Prandtl number. Like that, there are several systems in heat transfer in thermal engineering; where, this power-law form can be reduced to the linear form. This power law… So, this is called multiple linear regression. So, it is not only that you have to stop with one variable; you can do several variables. It is not that you will always have only straight line. If it is a power-law form also; but, if you have Nu equal to a Re m Prandtl n plus C something, then you are stuck. And, it becomes a non-linear regression. So, for certain kinds of problems, you can do all these [FL]; you can do all these tricks. These will all work. Beyond a certain point, you have to do hands up and then go for non-linear regression. It is not that it cannot be done. You have to do non-linear regression, which is little more difficult. All right?

(Refer Slide Time: 45:02)



Now, what are the other forms, which can be reduced? Y is equal to a x to the power of b. What is our proposal? How do you convert to straight line?

Student: ln

ln Y is equal to ln a plus b ln x. Y is equal to a e to the power of b x. Same; ln Y is equal to ln a plus?

Student: bx

bx. I will give you… I will give you some more forms. Can this be done? Can this be done? How?

Student: 1 divided by y.

Good. So, take 1 by y. 1 by y is b plus x by a x. So, here 1 by y… So, this will become y is equal to ax plus b. If I put a, b, you will get confused. I will put Y is equal to m X plus C. So, a e to the power of x, a e to the power of minus b t, a x to the power of b, a x 1 to the power of b into x to the power of c; a x by b plus x, a by b plus x. a by b plus x can also be done. 1 by y is b plus x by a and all that. So, there are some forms, which are amenable to manipulation. If you are able to manipulate, then you can reduce it to form, which will eventually lead you to a straight line; which will lead you to a straight line; and, which… And, you can use Y is equal to m X plus C and so on. However, all these tricks will stop at some stage. There are some problems; regardless of you can do all sorts of manipulation, you will get stuck. For these problems, you will have to essentially go for non-linear regression. So, in the next class, we will discuss non-linear regression.