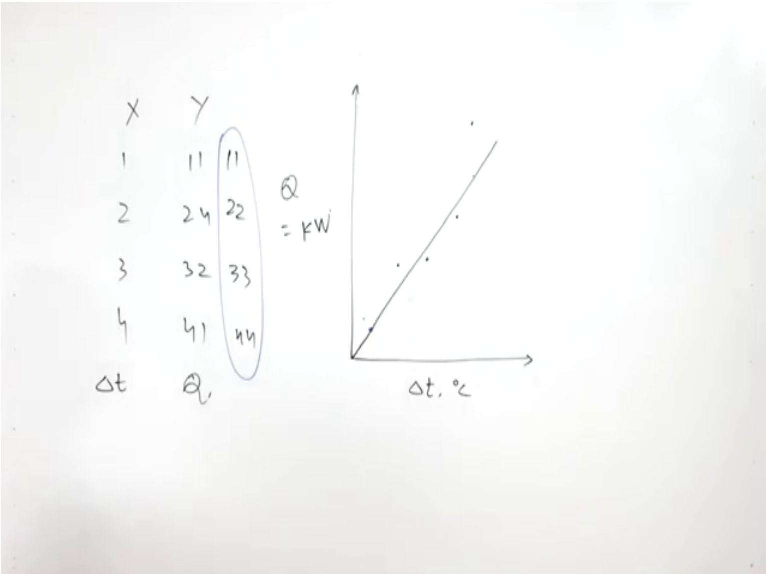
Mechanical Measurement Systems
Prof. Ravi Kumar
Department of Mechanical and Industrial Engineering
Indian Institute of Technology, Roorkee

Lecture - 10
Least Square Method

I welcome you all in this courses on course on Mechanical Measurement Systems, and today we will discuss the least square method of fitting the data. Now, suppose we conducted an experiment and we get certain data for example, expected suppose for the value of X and Y.

(Refer Slide Time: 00:38)



For the X value suppose the X value is 1 2 3 4 and Y value is 11 24 32 and 41.

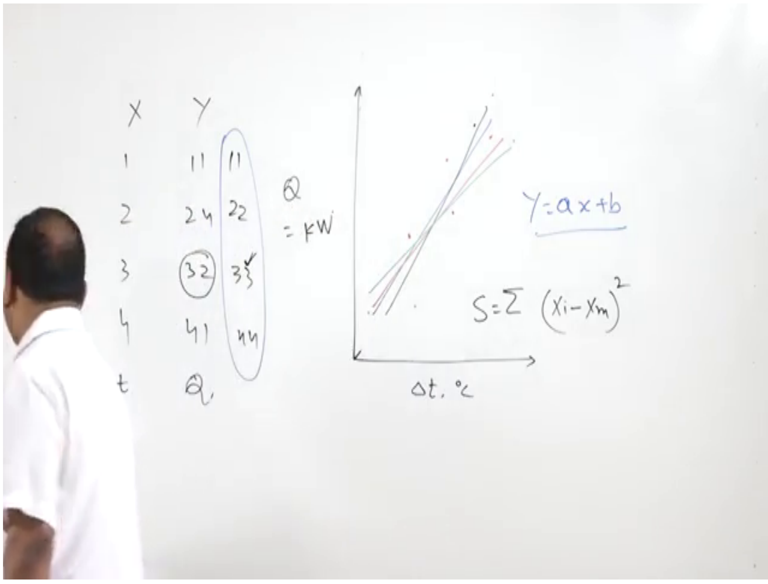
So, we will draw these data suppose this is temperature difference, and this is heat transfer coefficient, though temperature, so we cannot take the heat transfer coefficient because heat transfer coefficient may not increase with temperature difference. So, we can say that it is heat transfer right. So, temp because heat transfer is proportional to temperature difference and it is defined as the energy in transition by virtue of temperature difference.

So, energy transfer by virtue of temperature difference is heat transfer and we are now, plotting a graph between temperature and heat transfer. So, this is delta t temperature difference may be in degree centigrade and this is heat transfer in let us say in kilowatts or in watts, watts or kilowatts the unit does not matter here.

So, for different temperatures we will draw different value of heat transfer. Now, this is the experimental value ideal value you are expecting that it should be 11 22 33, 44, linear. So, this is the ideal variation which can be shown by this line right and the actual distribution because the data are scattered or 11 to 11 is ok. But for 24 it is 22, for 32 for 30 t, it is 30 sorry for 22 it is 24, for 22 it is 24, for 33 it is 32, for 44 it is 41 right. So, data are scattered this is the ideal variation, but data are scattered.

Now, we know this distribution. Suppose we do not know this distribution then we have only scatter data, then we have only scatter data. like this, and we want to know the trend of these data.

(Refer Slide Time: 03:04)



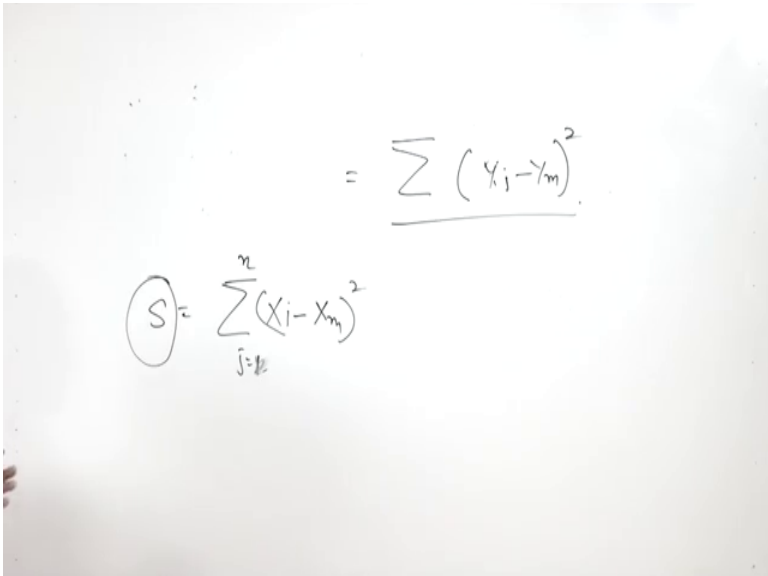
Now, in order to find the trend of this data we can have line like this. We first of all let us assume that the variation of data is linear. So, Y is equal to ax plus b right.

So, we can have this the best fit line, we can have this also a distribution of line trend of the data or we can have this is the best fitting line, we can have a number of lines we can have let us say this also a best fitting line.

Now, which one we should choose? Which one we should take? Now, for this the least square method comes into the picture. The least square method says measured value X i for this is equal to some kilowatts measured value minus predicted value from this line. That is a difference suppose the measured value here is 32 right this is the measured value this is the this is the data, but this is the actual we want 33. So, it is difference between 33 and 32.

Now, let us take square of this that is the difference or we can say residue of the data actual value and predicted value predicted value by the trend. So, difference of these two squared it is a squared and if you take sum of these squares. I will write on, I will rub on rub this off and then. So, here this is the value we have with us and this is the ideal value predicted by the trend this is the best fit line this is by the best fit line and this is the value with us.

(Refer Slide Time: 05:25)



So, the residue will be X i or here it is shown as y. So, you can take Y i right, and then Y m this one whole square right the residues whole square of the residue and if we take sum of these residues then it becomes sum of the square of the residues right. So, we can always write in terms of suppose X i instead of this X and Y is let us remove this otherwise this will create confusion.

So, there is a data X i right and the predicted for the best is X m then we can take i is equal to n, sorry i is equal to 1 to n whole square this is the sum of the residues. So, sum of the residues is taken, and this has to be minimum in least square method sum of these square has to be minimum right

Now, in order to have for example, we take n number of observations, n number of observations, and all the observations do not agree amongst themselves right all the observations I am taking another case where we are measuring certain parameter and all those observations are not agreeing among themself. So, there is a variation right. Then there is a variation in that case what we can do we can have mean value of observations this is another example of this least square we can have mean value of observations.

So, some of the residues in case will be i is equal to 1s to 2 n X i minus X m whole square suppose I am measuring data sorry the diameter of the shaft diameter of the shaft is 10 centimeters. But I am getting observations 9.99 or 10.01, ten point this normally happens those who conduct experiment they very well understand this.

(Refer Slide Time: 07:19)

$$0 = -2\left(\sum_{i=1}^{n} X_i - nX_m\right.$$

$$S = \sum_{i=1}^{n} (X_i - X_m)^2$$

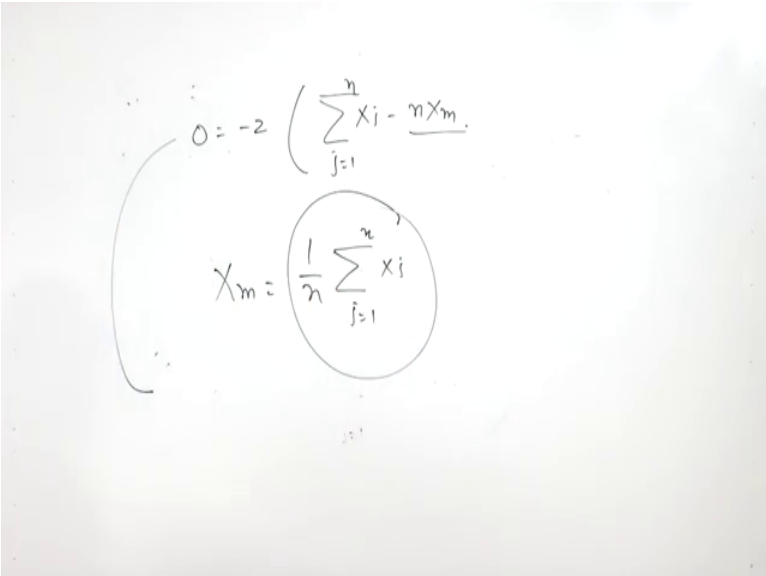
$$\frac{\partial S}{\partial X_m} = 0 = \sum_{i=1}^{n} -2(X_i - X_m)$$

So, if we take sum of all. So, if we take average of all observations and find the deviation of the reading from these observations right and this square of sum of the square of this will be S. Now, I want to minimize this I want to minimize this. So, I will take delta S by delta X m and when we partially differentiate this has to be 0 or it is equal to i is equal to 1 to n minus 2 X i minus X m, minus 2 we have taken because this is minus.

So, minus 2 is here right, and this or we can further write this as this as 0 is equal to this is a square right minus 2 sum of i is equal to 1 to n X i minus n, X m because X m is constant we are taking the average of the deviation right. If you multiply it by n you will get n X m right.

Now, here if we want to take this equation as a final equation then X m is equal to 1 by n sigma X i, i is equal to 1 to n this is nothing but average, average of the n number of readings ok.

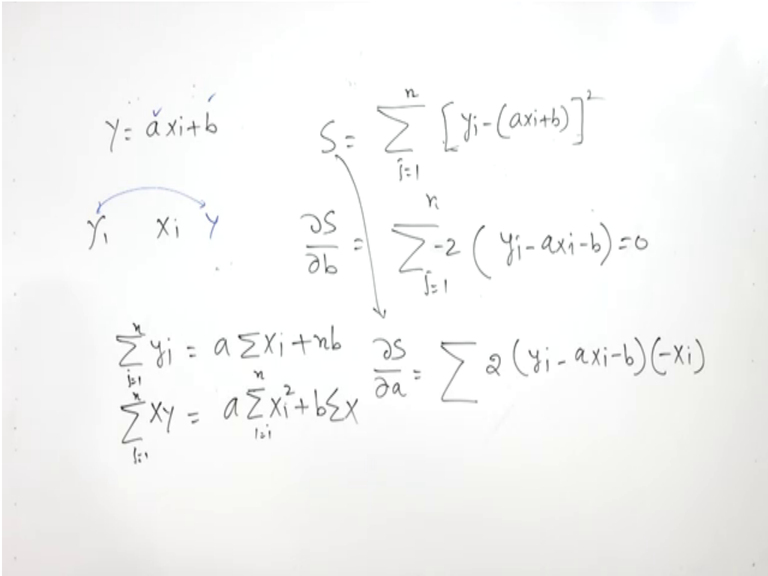
(Refer Slide Time: 09:33)



So, for the analysis of least square method we have to take arithmetic; this is not average this is arithmetic average this expression is arithmetic average. We can have different type of there is a variety of many ways you can take the average, but for the least square method for this particular application we have to go for arithmetical average of the data

Now, let us go back to the least square method.

(Refer Slide Time: 10:28)



Now, S is 0 Y is equal to ax i plus b we develop this equation linear equation and we have set of data Y i sorry yes Y i and X i. Using these data we calculate the parameter a and b and for different value of X i we get predicted value of Y. Now, this predicted value of Y will not be same as Y i there is going to be some difference because data are scattered ok.

So, in that case we will have to take sum of the squares of residues and sum of the squares of the as a residues can be expressed as yi minus y. So, Y is equal to axi plus b above the square. So, this is the sum of the residues and now, we will have to differentiate it with respect to a with respect to b because X i is constant the value of Y will depend upon a and b, the value of a and b xi is calls X i is given. So, value of Y will depend upon a and b. So, we will optimize for a b.

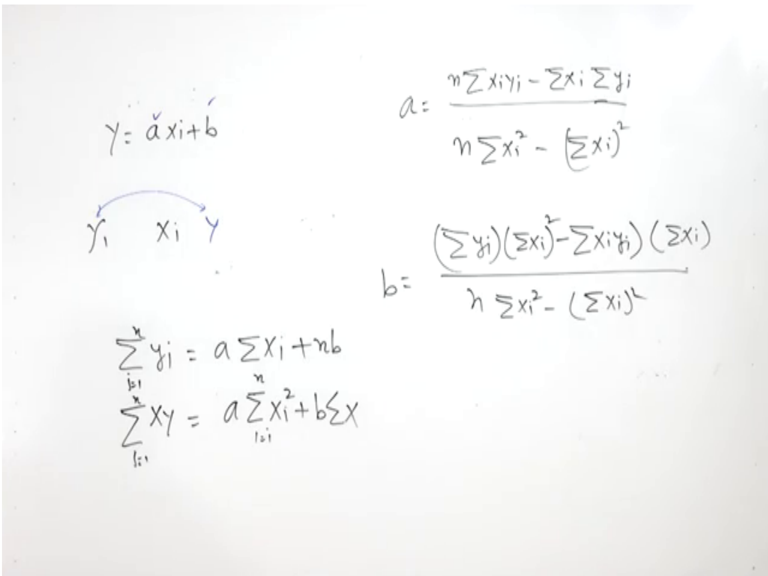
So, if now, we differentiate this partially differentiate this with respect to b. Now, we when we differentiate this with respect to b then we get i is equal to 1 to n 2, yi minus axi minus b sorry minus b will not be there because we are differentiating with respect to b. So, now, we will do partial differentiation of S with respect to b and we will get i is equal to 1 to n yi minus axi minus b multiplied by minus 2. Now, this is going to be 0 in order to maximize this.

So now, we can get sigma Y i or we can write here one equation sigma yi is equal to a sigma X i sigma b, b is a constant. So, in n numbers it will be appearing n times. So, here we can write nb. Now, again this equation is differentiated with respect to a. Now, we are differentiating this equation against a I mean with respect to a then we will get 2 yi minus axi minus b multiplied by minus X i.

If you further simplify this we will be getting another equation sigma xy is equal to a sigma X i square plus b sigma X and this is all from i is equal to 1 to n, i is equal to 1 to n, all i is equal to 1 to n 2 n and so on, I mean of both the equations

So, we have two simultaneous equations consisting of a and b is variables because a and b are variable that is why we have differentiated the equation with respect to a and b. Now, if you solve these two equations the final expression you are going to get for a is n sigma X i Y i minus sigma X i multiplied by sigma Y i divided by m sigma X i square minus sigma X i whole square. This is the value for a.

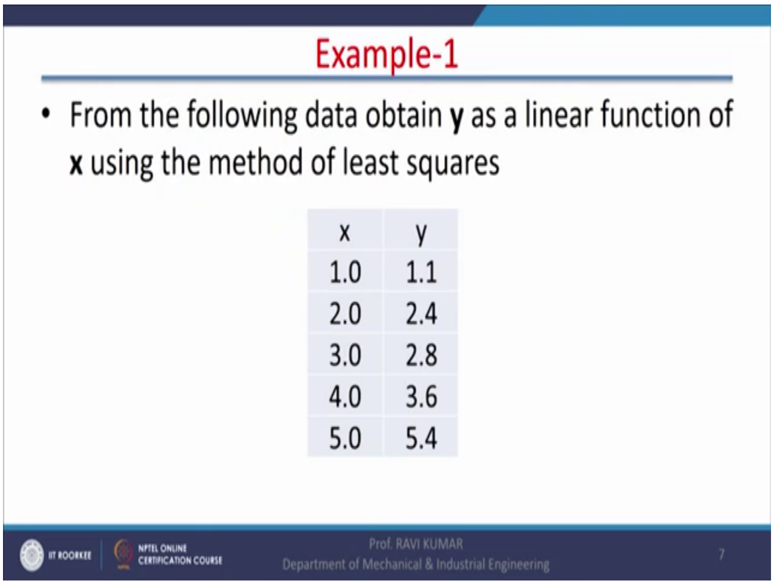
(Refer Slide Time: 14:51)



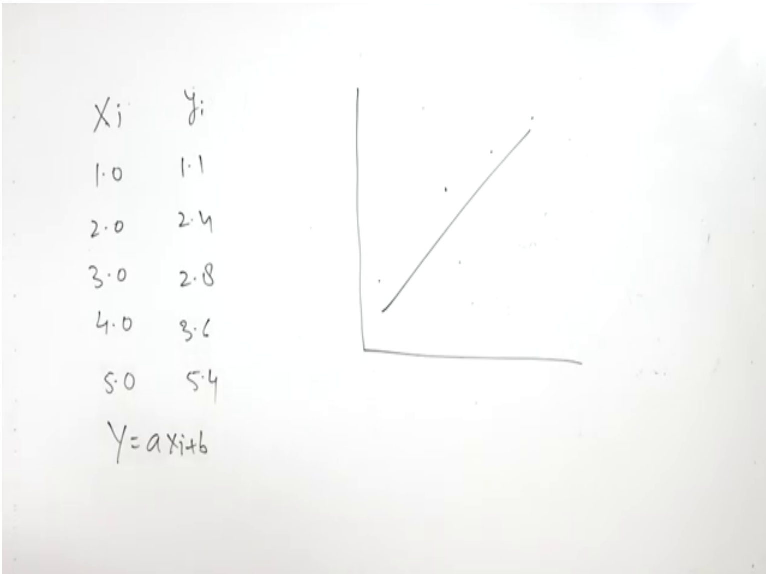
And value for b is going to be equal to sigma y i sigma X i whole square minus sigma X i yi sigma X i divided by same n sigma X i squared minus sigma X i whole square. So, this is value for b.

Now, suppose we have the value of X i and y i we can comfortably calculate the value of a and b and can find the best fit equation. Let us take one example in this regard one example in this regard.

(Refer Slide Time: 16:13)



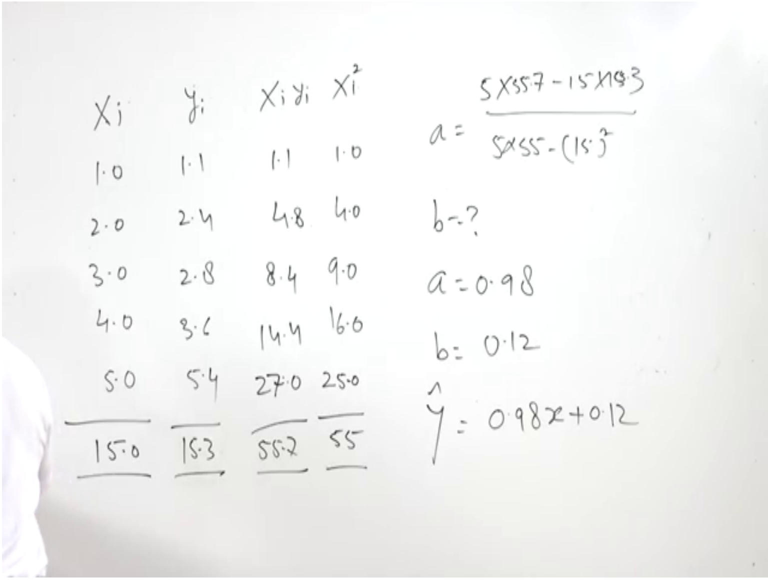
(Refer Slide Time: 16:29)



The value of X i sorry and Y, so X i and y i. Let us take the X i value as 1.0, 2.0, 3.0, 4.0 and 5.0; y i has value 1.1, 2.4, 2.8, 3.6, and 5.4.

Now, if I draw this graph, this data on the graph. So, when I draw this data on the graph we get certain scatter of the data. it is not linear it is obvious you can look at the data it is not linear. So, the data are scared scattered around certain best fit line and now, we have to find this best fit line and the best fit line is X is equal to sorry Y is equal to axi plus b, fine.

So, in order to find this first of all let us calculate the value of X and Y.



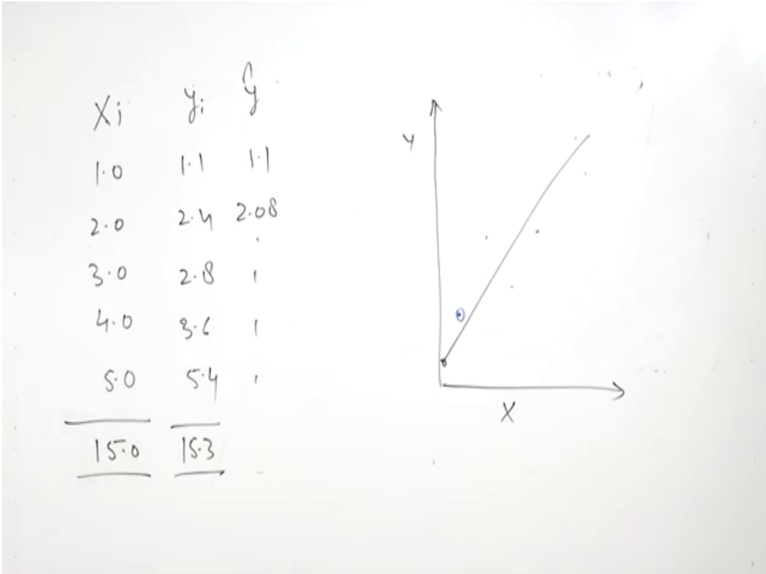
Because this term will be required in calculating the value of a and b. So, here the X and Y value is 1.1, I have already calculated it is 4.8, 8.4, 14.4 sorry 4.8, 14.4 and 27.0 0 and one more term is required if you remember that is X square X is square xi square. So, again we will take X i square 1.0, 4.0, 9.0, 16.0 and 25.0 sorry 20.0.

Now, we will calculate the value of a and b. So, we will use the scene equation for a we need the sigma of all this. So, sigma of all these is for this it is 15.0, sigma for this is 15.3, this is the sigma values and for this it is sigma X i is 55.7, is 55.7 right and for X square the sigma is 55.

Now, for calculating a we will have to make use of that formula n n is 5, sigma X i Y i it is 55.7 right, minus sigma X i that is 15. And then sigma Y i that is 15.3 divided by again n sigma X square that is n n is 5, 5 into 55 sigma X square minus sigma X whole square minus 15 whole square right and likewise we can calculate the value of b also. And I have calculated both the values and it is a is coming as a is 0.98 and b is coming as 0.12.

It means the y predicted value of y is 0.98 x plus 0.12. If you put X is equal to 1 then exactly we are getting 1.1. Now, if you put X is equal to 2 X is equal to 3. So, we will be getting altogether different readings for different values of X and different values of X. So, predicted values are for example, X is equal to 1 it is it is 1.1 right, for X is equal to 2 It is 2.08 and so on, I mean we can calculate all the readings.

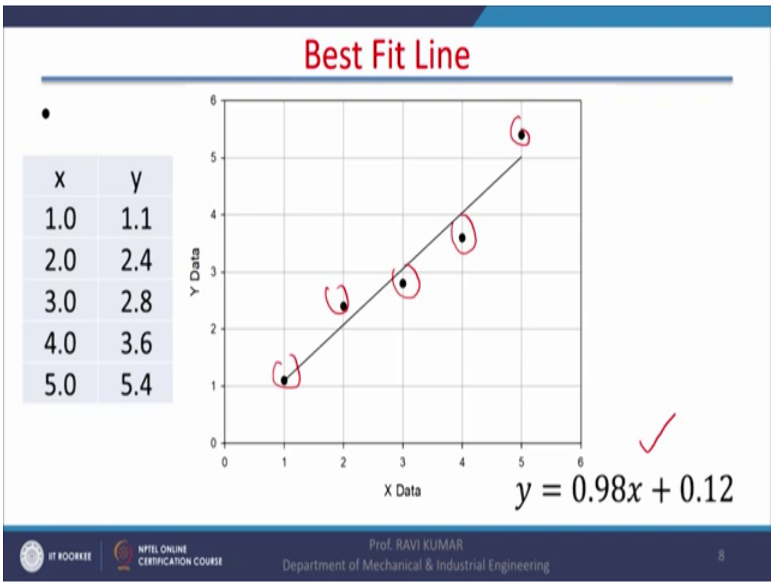
(Refer Slide Time: 21:37)



Now, if we draw a graph for this in straight line Y is equal to ax plus b. So, Y is equal to ax plus b will be like this something like this is the best fit line this is X and this is Y. And for one it is 1.1 it is ok, for 2 right it is actually it is 2.08 and, but actually it is 2.4 predicted is 2.08. So, we may be getting somewhere here point this point somewhere here. Likewise we will be getting points scattered around this.

Now, these graphs I have actually done this and this graphs I am showing on this screen.

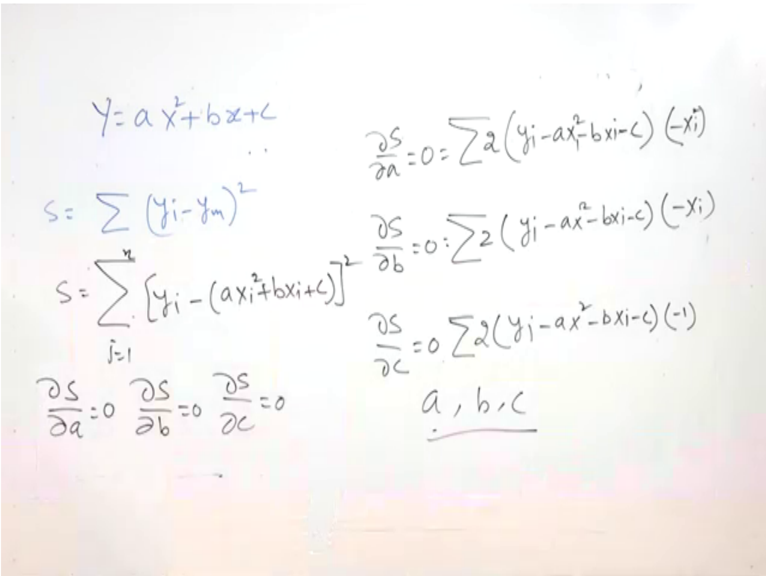
(Refer Slide Time: 22:49)



And we will be getting exactly this type of graph where the value of X and value of Y are given and when we find the value of Y using this equation. When we calculate the value of Y using this equation we get different points surrounding the best fit line ok. So, this is the best fit line best possible trend, this is the best possible trend which is depicted with this line

Now, we have taken only one case. Suppose instead of this we have equation like this Y is equal to ax square plus bx plus c because we have to find best fit line for all the all the possible trends. So, if this is not a linear this is a non-linear equation and it has 3 parameters.

(Refer Slide Time: 23:38)



So, again we will adopt the same method that sigma y i minus y m whole square S and here in this case S is equal to sigma y i minus axi square plus bxi plus c whole square sorry this is whole square and we will take sigma of this i is equal to 1 is a, i is equal to 1.

Now, we have 3 variables. So, we will take delta S upon delta a is equal to 0, del S upon del b is equal to 0 and del S upon del c is equal to 0. And when we get del S upon del a. So, del S upon del a and that is going to be equal to 0 is equal to again [noise 2 times yi minus axi square minus bxi plus c right multiplied by X i square.

Student: Minus a.

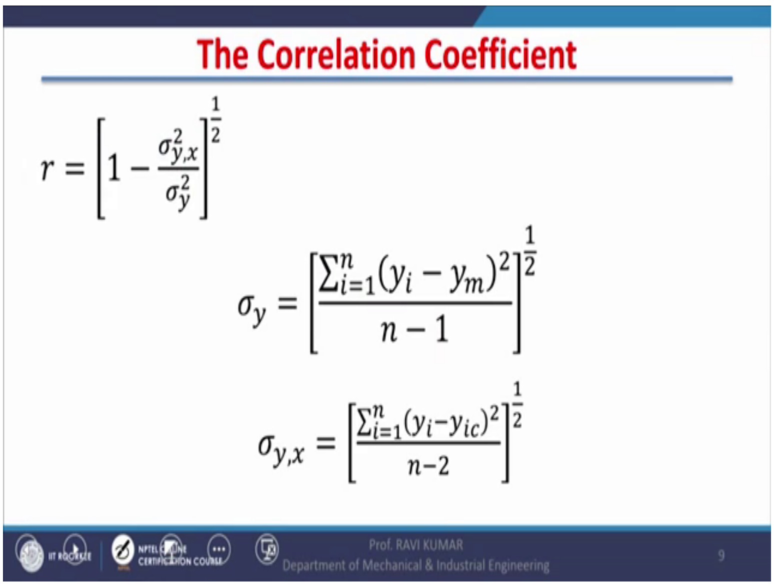
Sorry, sorry it is minus c so, minus c multiplied by X because it is multiple of a minus X i square. Now, similarly we will take delta S by delta b is equal to 0, is equal to 2 times again sigma sigma 2 times yi minus ax square minus bxi minus c multiplied by minus X i. And you will take delta S upon del c is equal to del upon del S upon del c is equal to again 2 y i minus ax square minus bxi minus c and this is minus 1, right.

And again we have three simultaneous equations and for these equations we can find the value of a b and c. And then we can solve these simultaneous equations and once we have the value of a b and c we can have predicted value then the experimental value and predicted value, predicted value by the best fit line and again we can draw the best fit line for the equation.

Now, these best fit lines are very useful I mean they show the trend of the data and once we have the best fit line just simply putting the value of any input parameter we can get the output parameter. And because they are best fit lines we can also understand we can also take into account the error the range of error or the range of error in the measurement because that is also imp explicit from the from the exploration of the best fit line itself because the error in the prediction for the best fit line will be minimum.

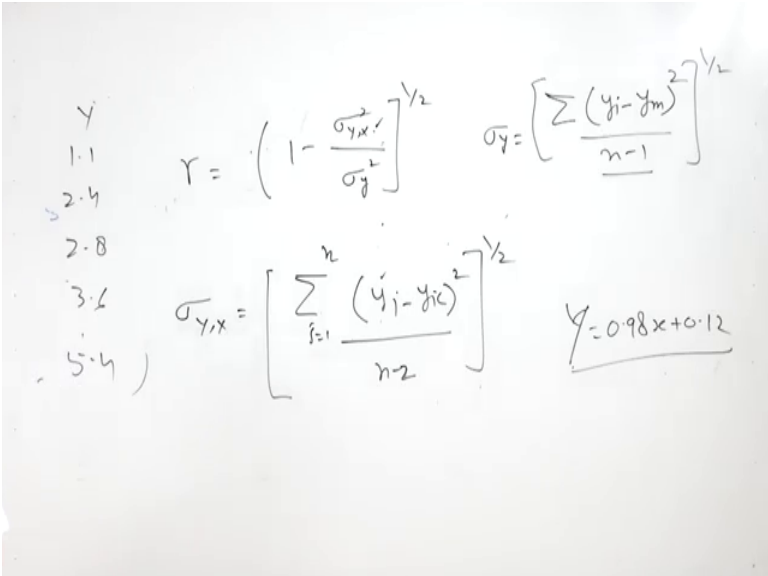
Now, once we have drawn the best fit line. Now, we have to check the goodness of it. Now, here also there is a check for goodness of it how best the line is fitted into the data. So, for the goodness of it we have to find the correlation coefficient.

(Refer Slide Time: 28:13)



Now, the correlation coefficient it is shown by r, r is equal to 1 minus sigma yx divided by sigma y square raised to power 1 by 2.

(Refer Slide Time: 28:24)



Now, what is sigma yx and why it is sigma y? This is the standard deviation of the data using the best sit best fit equation. So, yx, y is the a standard deviation of the data using the best fit equation. So, yx, y sigma yx and y this is the standard deviation of the data using best fit equation. So, it is going to be equal to sum of sum of I; not a n; i is equal to n y i minus y ic whole square divided by n minus 2 n raised to the power 1 by 2. y i is the given value of y and y ic is the value predicted by the best fit line right, we take difference of these two right in order to find this standard deviation for y and x the y which is data available with us minus data predicted by fellow of I predicted by the best fit line the difference of these two is square divided by n minus 2.

Here we are taking n minus 2 because we have here we have two restriction of a and b, otherwise it would have been n. So, here because this is a small sample and we are taking that is why we are taking n minus 2 and restriction is due to a and b.

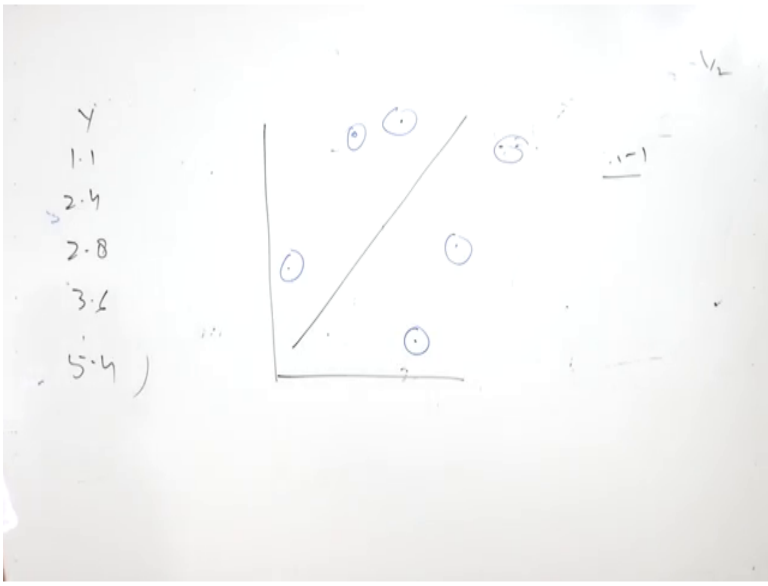
Now, another one is sigma y sigma y, sigma y is the standard deviation is sigma y i minus y m whole square divided by n minus 1 raised to power 1 by 2. Now, this y m is average of y. Suppose we have number of we had number of y if you remember the value of y was 1.1, 2.4, 2.8, 3.6 and 5.4. So, average of this is y m.

And the deviation of each from this average is y i minus y m whole square divided by a here we have taken n minus 1. Here we have taken n minus 1 because there is a restriction of data there only 5 data. So, there is one restriction it is not an infinite population. And here we have taken n minus 2, because once we have the value of Y is equal to 0.98 x plus 0.12 best fit line we can generate infinite number of data for given value of x. So, that is not in restriction, restriction is only for a and b that is why we have taken here n minus 2 and here we are taking n minus 1.

Now, putting these two values here we get the value of r. Now, if this is equal to this then r is equal to 0 and when this is equal to 0 this is equal to 0 then r is equal to 1, when r is equal to 1 it means perfect matching of the data or sometimes it is taken as r square this is removed or we can do it another way the value of r square. So, when this is equal to 0 then it is going to be equal to 1 right. So, the value has to be close to 0.9, for a good fit this value has to be very close to 0.9 in that is we say the data are the best fit data best fit lined it is very is a very good fit for the given set of data.

Sometimes what happens we have the best fit line, but the scattering is too much. So, I will give you an example for that also.

(Refer Slide Time: 33:38)



Suppose we have a scattering of data like this, and if you draw best fit line it will come something like this, but if you can see here the scattering is too much, when the scattering is too much or maybe some here some here when the scattering is too much then r square value is going to be very low. And then we can say that this best fit straight line is not to present representative of the trend of the data.

I think that is all for today we will take a few numericals on statistical methods in the next class. Next class we will be start with the uncertainty analysis, and after uncertainty lies analysis in the I mean in the next to next class we will take some numericals on these statistical methods to give you more insight of the phenomena.

Thank you.