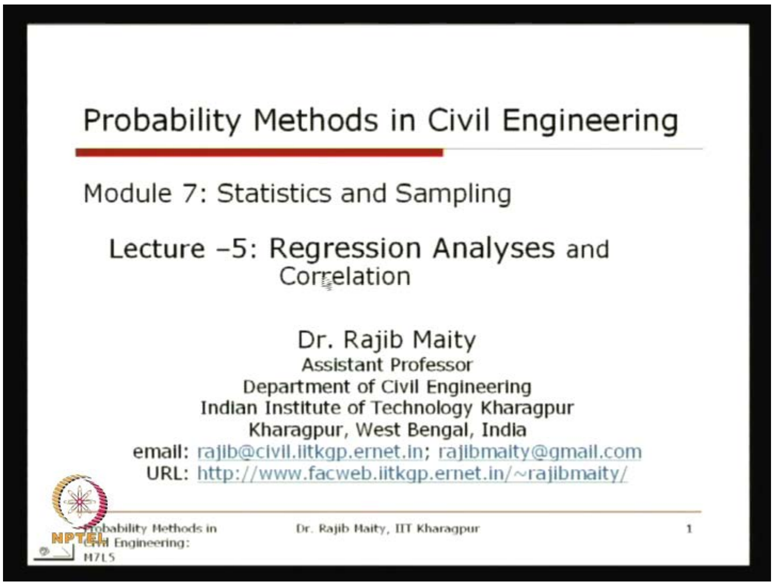
Probability Methods in Civil Engineering

Prof. Dr. RajibMaity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture No. # 39
Regression Analyses and Correlation

Hello and welcome to this lecture.Inthis lecture, may be in this or the next lecture, we will cover the topic on this regression analyses.

(Refer Slide Time: 01:55)

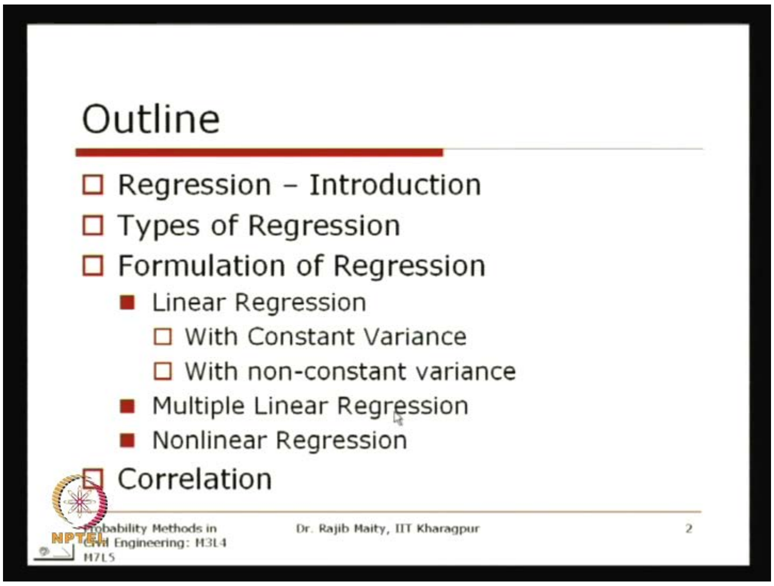


Therearedifferent types ofregression analyses say, first we will start with thatsimple linear regression and after that there are different types.Multipleregressions are there and then non-linearnon-linear regressions are there. So,we will see andbasicfundamental things; basic concept we will understand.Basically, whenin a, in many application fields, obviously, including civil engineering there aremany randomvariables are there which aresupposed to have somerelationshipto their and in this,through this analyses we tried tocapture that we try to model that relationship.

Now,if the relationship is linearthen we generally go for this linear regression and sometimes we have seen that may be the linearrelationship is not sufficient. So, there we have to go to the non-linearregressionanalyses.Sometimes,the target variableis dependent only on onevariable or sometimes thatresponse variable or the target variable can depend on more than oneand dependent variable. So, in that case we generally go for these multiple regressions.So,all these things we will learnin this lecture or this may continue to the next lecture also. So, this is our today's lecture title is regression analyses andcorrelation and this correlation means herethatwe have alreadydiscussed earlier that thiscorrelation when we discuss this random variable and all.

So, here also we will see that howthis regression analyses.In this regression analyses correlation is an important part. So, we will just see in the light of this regression analyses also towards the end of thislecture.

(Refer Slide Time: 02:24)

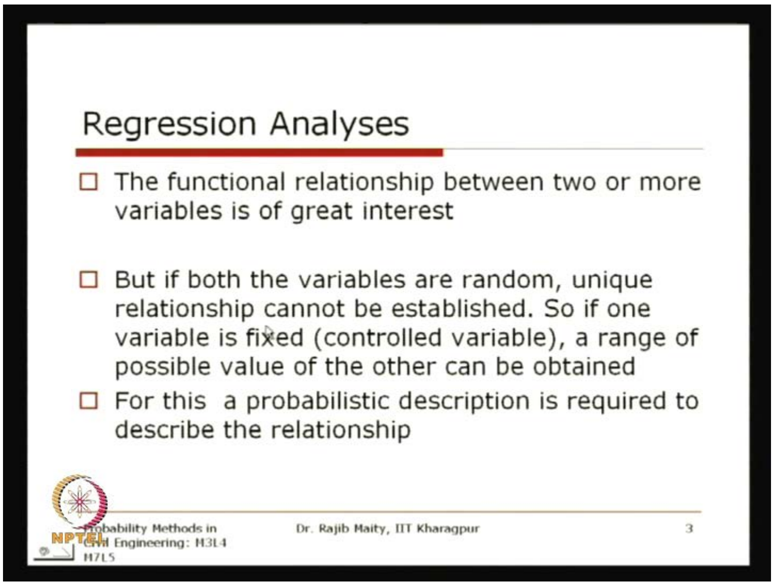


So, our outline of thistoday's lecture is that first we will go through some introduction and then we will discuss about somethat different types of regression then you formulation of this regression in this there arelinear regression as I mentioned and in this linear regression also it may have the constant variance orit may have the non constantvariance. So, this non constant may bethe variable variance may be the other word, but, just to avoid tosimilar word. So, it has used as this non constantvariance.

So,thisnon constant variance and this constant variance means in general for the linear regression when we refer to we refer to this constant variance we means over the entire range of the dependent variable the variance of theresponse variableremains same that is what is the I can say that by default case, but, sometimes or it can beobserved that these variance may also varyover thedifferentrange of that dependent variable. So, in that case, we have to go for this non constantvariance also.

Then, if there are more thanone dependent variable then we have to go for this multiple linear regression andif the relationship we see that may not be linear sometimes some other non-linear relationship may havebetter, can better extend thetarget variable then we can go for this non-linear regressionand then as I told that there is… So, we will see that correlation basically this will be a majorthat how strong the relationshiphas been captured through that model that we have developed through thisregression; so, that we will see.

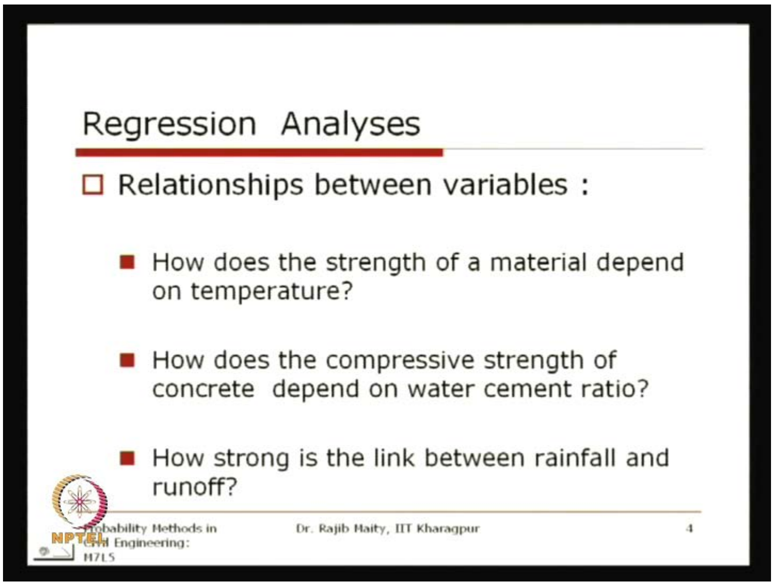
(Refer Slide Time: 04:09)



Well,in thisregression analyses the fundamental that, sorry, the functional relationship between two or more variables is of great interest as I mentioned that there may bemanyvariables which arewhich we can see thatthere could be acould be a relationship by the linear ornon-linear and this kind of relationship basically if we just take the observed data and plotted through somescatter diagram and thenitself by visual inspection itself we can see that there arewhether there are some types ofrelationship is there or not. So, if we can see than we canwe can think ofthis type ofregression analyses to capture thatthat particularrelationship.

So, hereif boththe variables are random, unique relationship cannot be established. So, you know thatunique relationship here what is meant that it may not be that one to one relationshipthere could be some eventhere could be some randomness is both the variable. So, if one variable is fixed and that is knownthat is termed as a control variable or that what I mentioned is thatis the dependentvariable; the range of possible values of the other can be obtained through thisanalyses.Forthis a probabilistic description is required to describe this relationship and basically this is what is you will get through this regression analyses.

(Refer Slide Time: 05:45)

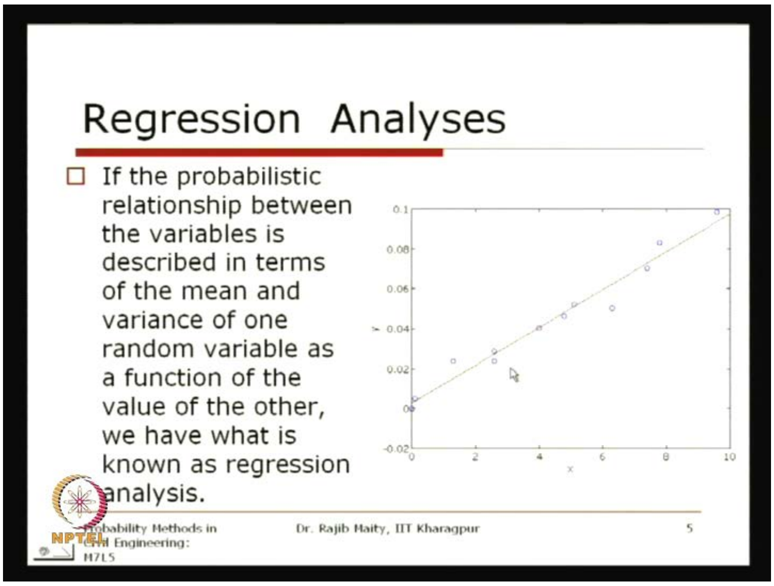


So,in this, the type of question particularly if I concentrate to this different afield of application in civil engineering then this type of analyses will give me the answers to a kind of this type of question say that how does the strength of material depend on the temperature. So, if the temperature I vary. So, how the strength of material whether it will increase or decrease or how the relationship is.

Second is say thathow does the compressive strength of the concrete depend on the water cement ratio. So, if I increase the water cement ratio then what will happen to the compressive strength or if I decrease it what will happen. So, these are some two variables are considered.Similarly, what we can say that whether that target variable here is the compressive strength may have instead of this only that water cement ratio.Therecould have been other factors as well that canbe influencing to this. So, then what will happen?That one target variable and more than onedependent variable - sothat, multipleregressioncan come into the picture.

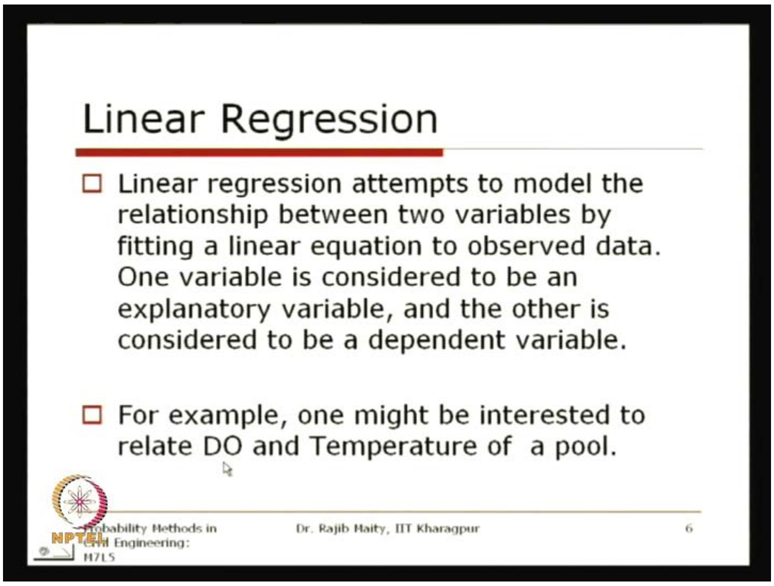
And, third: sothat, how strong the link between the rainfall runofffor a given catchmentor for givenarea. So, how. So, rainfall and runoff if the rainfall is more runoff can be more. So, how strong is that relationship? So, this type ofanswerwe can getthrough thisregression analyses.

(Refer Slide Time: 07:16)



Theprobabilisticrelationship between the variable is described in terms of the mean and variance of one random variable as a function of the value of the otherwe have what is known as theregression analyses. So, say for example,as I was telling just by if I just plot thatthrough a scatter plot the what is the observe data that we are having the paired observe datapaired in the sense here that we are talking about the twovariables first. So, this is one variable is x and other one is the y. Now, if I just plot it,theseblue circles,,you can see that this is thepaired data and. So, we can see that ifx increases y cany is also increasing and vice versa ifx is decreasing y is decreasing. So, whether nowcan we just estimate one relationship between this x and y. So, that estimate that estimate of this functional relationshipis that regression analyses that we will getthrough this analyses.

(Refer Slide Time: 08:24)

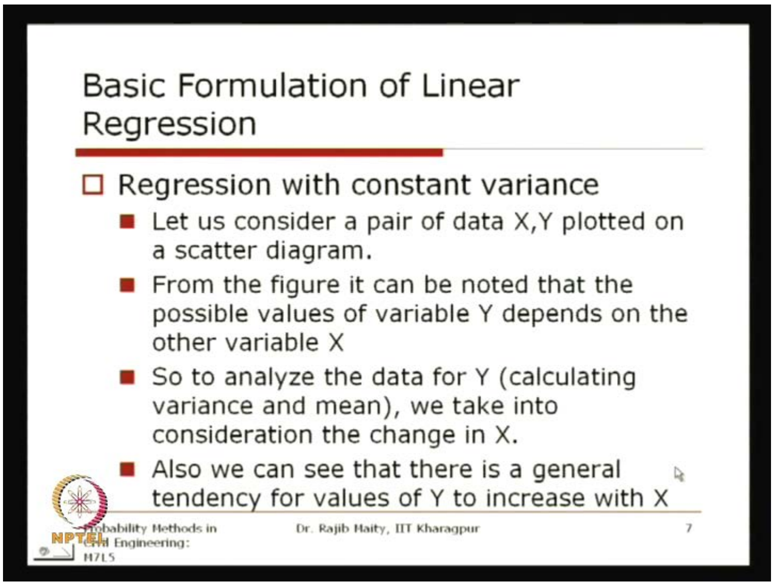


So, first we will take thatlinearregression for examples are thediagram that is shown here we can see and we can expect that there could be a linear relationship canhave it here. So, but, in many other cases where if just is lookingthis scatterplot we can see that initially it may be increasing and later on it may not increase in that rate. So, there could be we can expect that there might bea non-linear relationship can happen. So, the first what we are taking up is that linear regression;where the expectation that the relationshipis linearbetween thedependent variables and the target variable.

So, thelinear regression attempts to model the relationship between two variables by fitting a linear equation to the observed data one variable is considered to be an explanatory variable and other is considered to be adependent variable. So, thatis whatour target. So, in thisexample that we have seen what we can use is that this variable x we can use as to be that your dependent variable and this is a y is my target variable. So, I can use the information of x and I can model this y it can be it could not be opposite also if we can we if weestimate a x with respect to the variable y. So, then we generally say thatthat x is regressed on y andin other waythe y is regressed on x.

Forexample, that one might be interested to relate the dissolve oxygen and the temperature of a pool. So, whether the dissolve oxygen and temperature these two data is generally first collected and then we can see that whether thetheir relationship how the relations how they vary with respect to each other whether the must whether in the sense that I cansee it in the both sides whether that DOgiven the temperature orthe temperature given what is the DO, but,sometimes in case of thisthe practical considerationmay be we are interested to know that our what is our target what is thewhat should be thedependent variable and what should thattarget variable for example, the example that is given here the dissolve oxygen and the temperature generally what we see is that temperature we use as a dependent variable and this dissolve oxygen is the target variable. So, this depends on the in what area in what practical field that we arethat we are applyingthisanalyses.

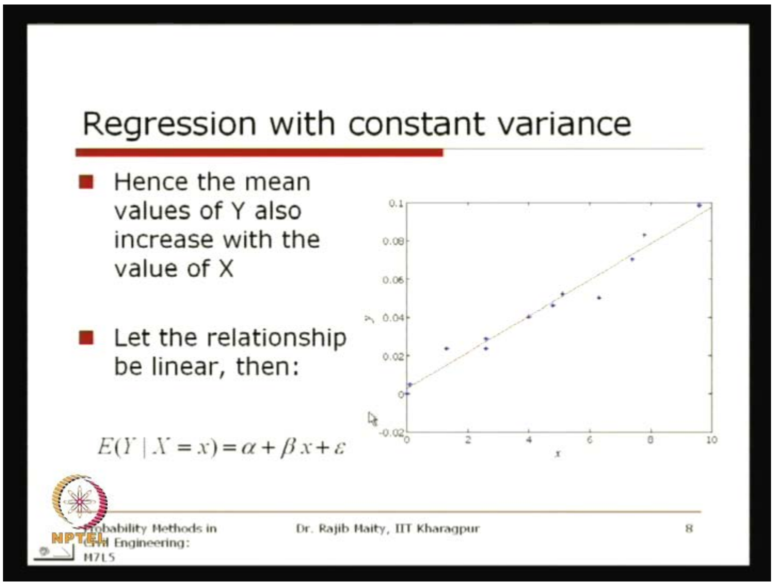
(Refer Slide Time: 11:10)



The basic formulation oflinear regressionwith the constant variance is first. So, here as I was statetelling as a starting that when we aretaking that the constant the variance of the dependent variableover the entire range of the dependant variable it remains constant. So, in that casewe generallysay that this regression with a constant variance andby default when you say the regression analyses we generally mention that it is with the constant variance. So, the non constant case is a special case that we will take that we will see after some time.

So, in case of thisregression with constant variancelet us consider a pair of data XYplotted on the scatter diagram asjustfew slidebefore you have seenfrom the figure it can be noted that the possible values of the variable Ydepends on the other variable X. So, to analyze the data for Ycalculating variance and mean we take into consideration the change in Xand also we can see that there is a general tendency of the values of Yto increase with the X. So, these are some of this example is given with respect to that plot the scatter plot that we have seenfew slides before.

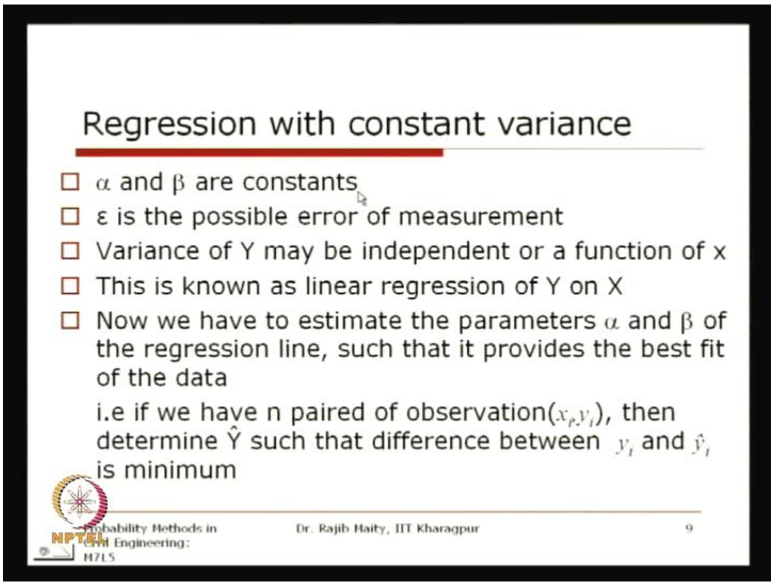
(Refer Slide Time: 12:35)



So,here again thesimilar plot has been shownhere. So, here that onevariable is Xand other one is the Y. So, herewe are taking the case that we will regress Yon X. So, So, Xis our dependent variable and Yis our target variable. So, here you can see that when Xincreases Yalso increases and vice versa. So, we have tofit a linear relationship between this Xand Y. So, hence the mean value of mean value mean values of Yalso increases with the value of the X. So, as Xincreases that mean value or the in the statistical sense the expected value of the Yalsoincreases. So, the relationship let the relationship be linear because we are discussing thislinear regression now. So, the expected value of the Ygiven Xa particular value X. So, you know. So, this is the conditional expectation. So, if I just take what is the expectation of the Yyou know the expectation of the y means without any other information. So, whatever the Ywe see that it canfrom this diagram we can see that it varies from 0to 0.1 say. So, whatever thevalues the range that we see we will just take its mean and that is the expected value of the Y.

Now,this when you are fitting this relationship; that means, ; that means, it is a condition on the given value of this Xnow if I give some value of this Xat 6. So, a in this part what is the expected value of this Y. So, this is now becomes the condition and this conditional meanisexpressed through this linear relationship which is alpha plus beta x plus epsilon. So, this is you knowthis is the equation of that of the straight line plus some error term should be there to express that what is thatvalue of that the mean value of this y.

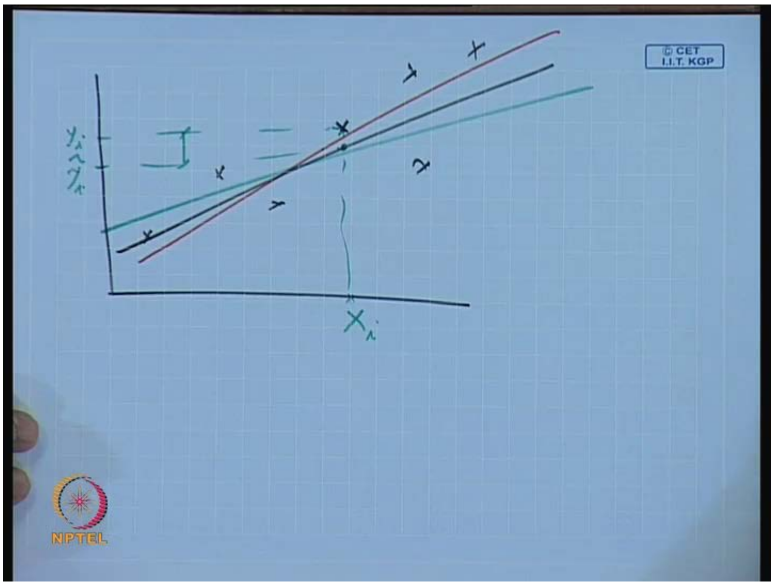
(Refer Slide Time: 14:42)



Now this alpha and beta are the constant and epsilon is the positive error of this measurement of the sorry possible error of measurement. So, if when we take that dataobserve that data. So, there could be some in that measurement. So, there could be some errors. So, that error is expressed throughthis epsilon variance of Ymay be independent or a function of Xthis is known as the linear regression of Yon Xthat is what I wastelling. So, it is Yon Xit can be expressed in otherway also that is Xon Y. So, the relationship will change that expectation of Xgiven Yis equals to some constant plusthe beta multiplied by that your Yplusepsilon. So, that is the observational error.

So,now we have to estimate the parameters of this alpha and beta of the regression line such that it provides the best fit of the data now this best fit of the data meansif I justsee this one this scattered diagram. So, there could be the various possiblelines that I cannot think through these points now which line should be the best fit line. So, what ismeant here is this.

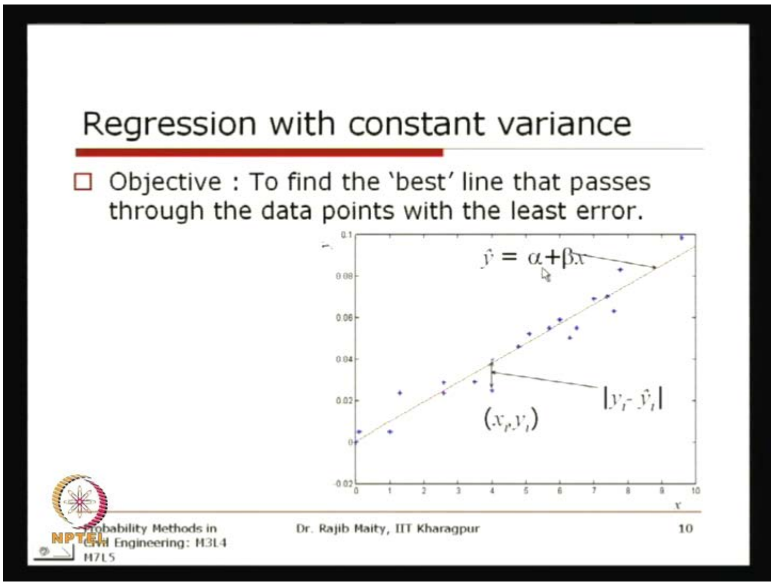
(Refer Slide Time: 16:09)



So,, if this is these are the data points then there could be the there could be some lines which canbedescribed throughdifferentstraight lines now out of these lines the possible lines which one should be the best fit now this to get that best fit.So, to get that best fit we have to follow somemethodology which is known as the method of least square to estimate thatto get that the line that is best fitting through this points and based on that we will get what is thethat estimate of those regression constants. So, this is what is mentioned here that issuch that its provide the best fit of the data that is if we have npaired of thisobservation xi y i. So, these are the paired observation and these are one pair is one point on that diagram thendetermine y cap in such that the difference between that yi and yi cap is minimum.

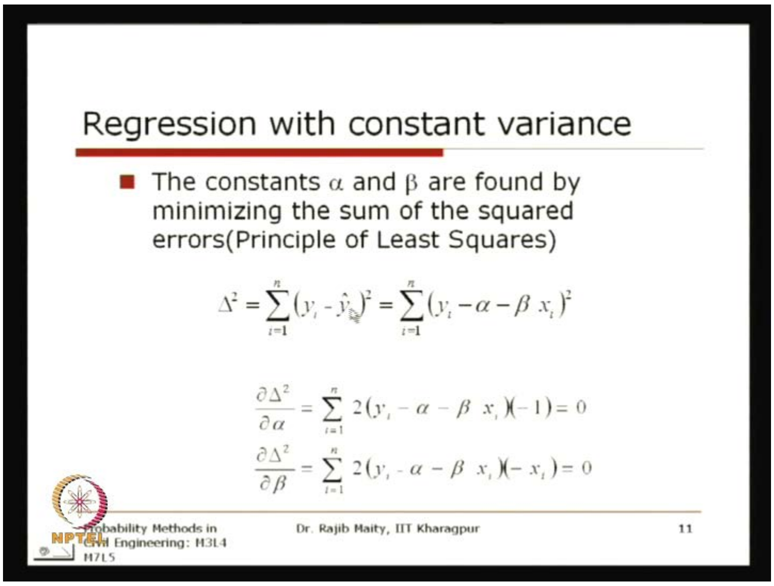
Now, what is this y i iand yi cap isif I referred to this diagram is this. So, this is your that point where you can see that this is your somethis is that yi and if the whatever suppose the this black line is your best fitline then this with respect to this x with respect to this x i. So, the estimate is this one. So, this is your yi cap. So, the difference between these two is the error which should be minimized. So, now, as close as thispoint to this observation and this is for all the points then that line should be the best fit line. So, the difference betweenthat is why the difference between this yi and yi capshould be minimum to declare that the line is a best fitting through the data.

(Refer Slide Time: 18:23)



So, it is nowexplained here. So, our objective here to find the best linethat passes through the data pointswith the least error. So, now, thisblue stars are the observeddata and this is the estimate of this regression line which is alpha plus beta x now this difference from what is the point that you can see and what is this correspondingpoint on this regression line there is a red line shown here is yourerror. So, this mode of this yi minus yi capis the absolute error for the point xi y i. So,yi is known is the is the observed one and that yi cap if I just put here this xi then alpha plus beta these two as a constantputting this xi what we will get that will be your yi cap.

(Refer Slide Time: 19:23)



So, the constants alpha and beta are found by the minimizing the sum of squared errors sum of squared errors and this is known as this principle of least square. So, what is done is that this is the error that is yi minus yi cap this is the error and that error is squared and summed up for all the observations. So, in this diagram if wesee,this is the error yi minus yi cap and this is obtained for all these data pointsand this errors for individual point is first square up andthat. So, that square error is summed of for all thenobservationsthat isavailable. So, this is giving is the sum of square errors.

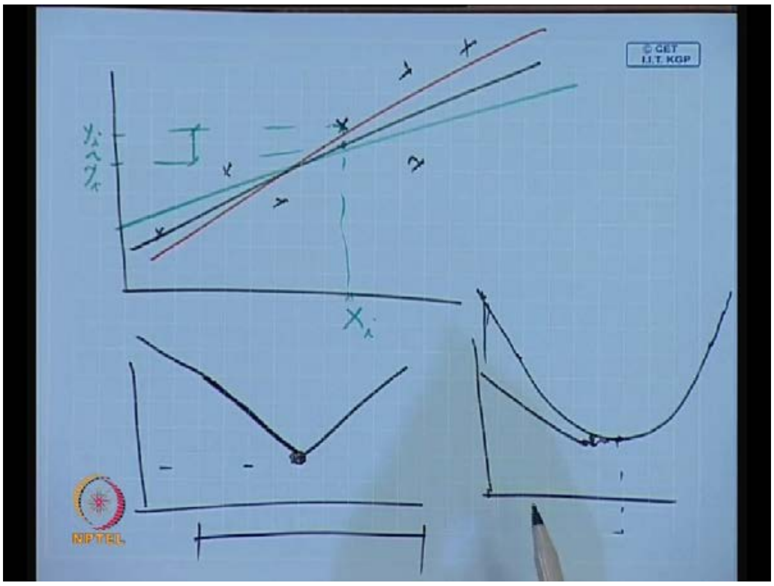
Ah now this sum of square error now if I just replaced this y cap from that regression line which is your alpha plusbeta xi and we are taking this minus. So, yi minus alpha minus beta xi whole square is give you that sum of square errors now toget thatestimate of this alpha and beta. So, this error should befor this alpha and beta the value of alpha and beta should be such that because these are the two constants which is basically which is determined everything about that straight line. So, error - this quantity should be minimum.

Now,to ensure thatwe have to take thispartial derivative of this sum of square error with respect to each thisparameters alpha and beta and this has to be equated to 0. So, we are having twounknowns and we are having twosimultaneouslinear equation that we can solved to get what is the estimate of this beta before I proceed I need to take some time to explain this one why we have taken thissquare andthisand we are using this as that total error.

So, because you can see the first thing the firstdirectthing that you canthat you can have from this diagram is that. So, for some points the error will be negative and some point the error will be positive -the error will be positive depending on whether the point is belowthe regression line or above the regression line.

Now,when we are taking this square obviously, those sign is going because we are interested to this what is the deviation from this regression line whether it is on the positive side or on the negative side that we are not interested when we are looking for the best fit line. So, whatever the error that we get if we take the square; obviously, that sign will go,but ,this can also be, think of that if we just take thatabsolute value of that error as it is shown it here then also that sign can go and if we just add them up then,what we will get is thatalso it will give an the absoluteerror summation of the absolute error.

(Refer Slide Time: 22:59)

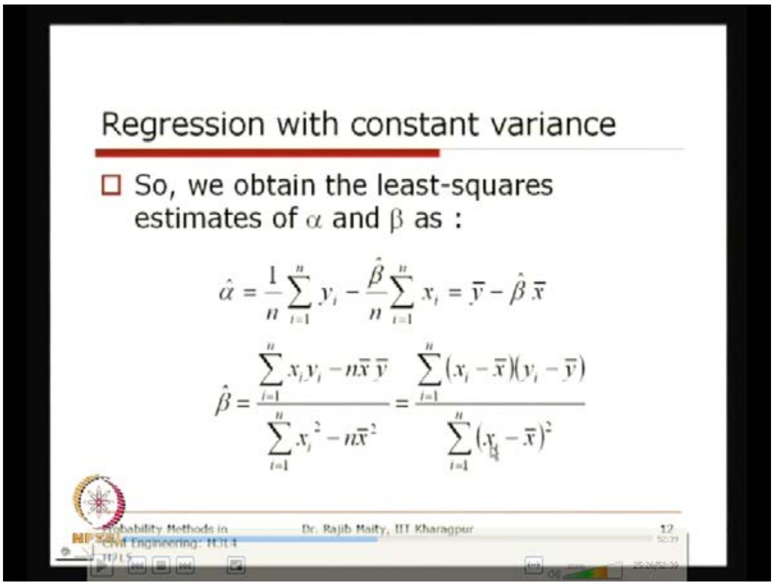


But, generally when we go for this least square technique we take it to be the square and thenwe do this partial derivative this is because you know when wetake the error now if we justseethat error and that error if we take it as a linearfunction basically what we are if we are minimizing it then this one basically our point issuppose this is our target point now this is the over that the possible range of the parameters now when we take this absolute error then the change with respect to thatparameter it will be the linear one and when it take it to thisto the square or the. So, this will become basically a quadraticfunction.

Now, what will happen if if our estimates are far away from what is the optimum value.Then you know from the optimization technique. So,if it is far away then the next step basically it will go very close to thatoptimumvalue and once it is comes to the optimumvalue then, the steps will be smaller steps.But,in this case generally thatthe step size are always same because this variation is linearhere, but, means this is basically when you go for this optimization optimizing theparameters that time it has been seen that this taking square is betterthan this taking thislinear function

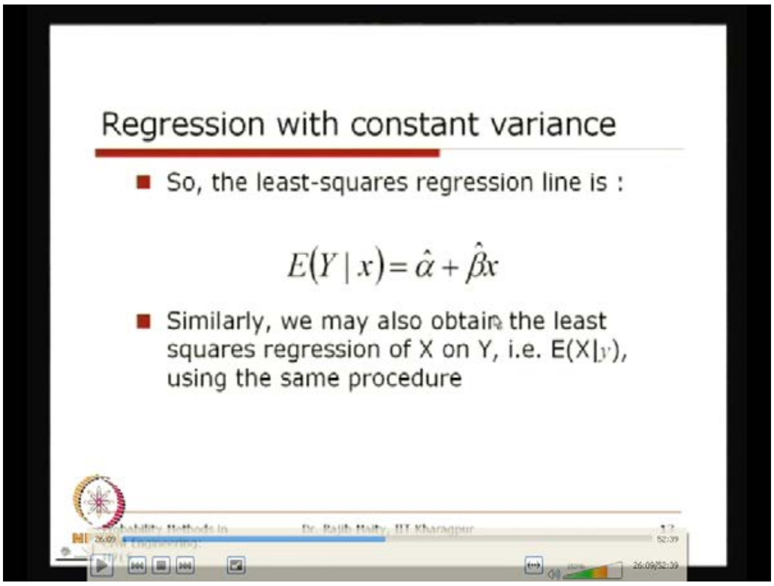
So, that is why sofar as that sum of it is we are all generally interested for this sum ofsquare error. So, what is done in thisprinciple ofleast square technique?

(Refer Slide Time: 24:50)



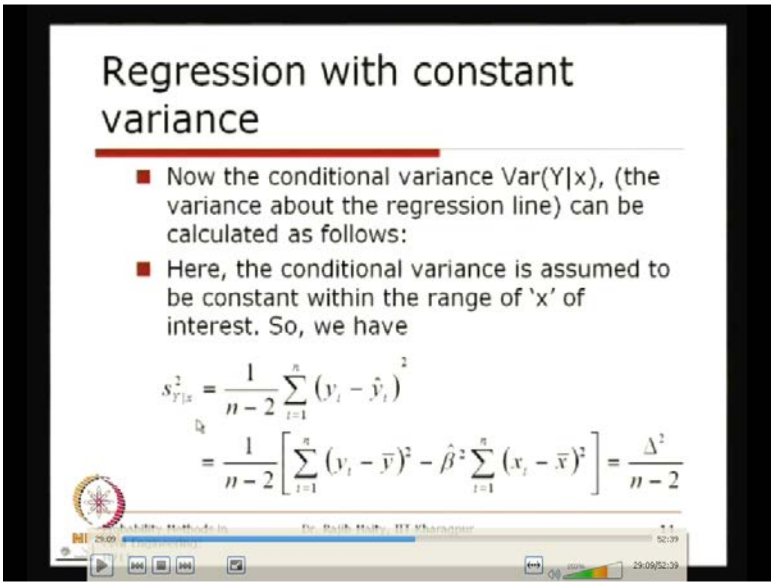
Well, we got thisafter, through this partial derivative we getthese two simultaneous linear equation whereby solving them we can get the estimates like this that alpha cap.Thisis now, this cap symbol is given; when we are referring to that it is theestimate. So, this alpha cap if we can solve it and we can. So,that it will be the y bar minus beta cap x bar this y bar and x bar are the mean of this observed data and this beta cap is the estimate of this beta can be shown that it is the summation of this xi minus x bar multiplied by yi minus y bar multiplication of them sum it over the all n observation divided by xi minus x bar squaresum it overthisall n observation.

(Refer Slide Time: 25:34)



So, these two are the estimate of this alpha and beta. So, the least square regression line is the expectedvalue of this y given that x is equals to alpha cap plus beta cap x. Similarly, we may also obtain the least square regression of this x on y as I was mentioning that is that expectedvalue of this x given y using the sameprocedure.But, here it will come as their dependent variable will be y.Obviously,that alphaand this beta the estimate of these regression parameters; obviously, will change through thatif we follow that procedure whatever we have done.

(Refer Slide Time: 26:12)



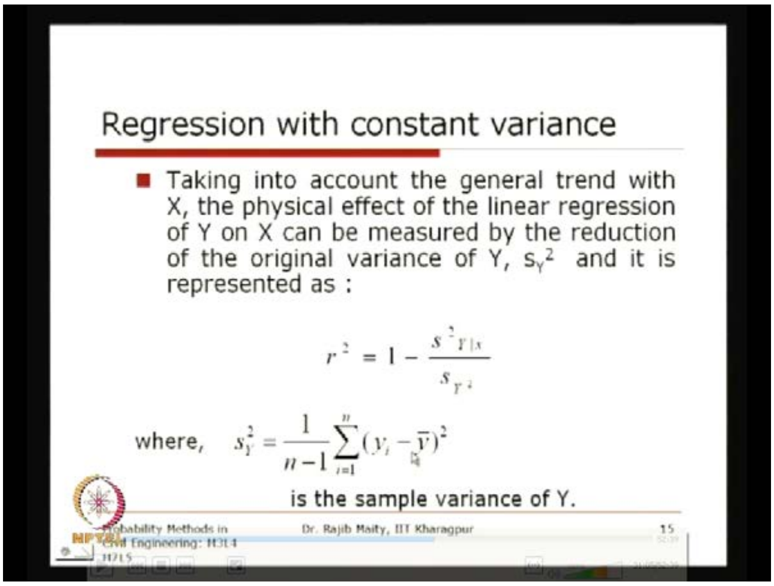
So, now the conditionalvariance - that is now the conditional variance of thisnow the variance of y given x. So, whatever we have got if that just now is that expected value of y given x. So, now, we are interested to know, what is the conditional variance of y given that x? So, thevariance about the regressionlinebasically,what we meant here is that if I just referred to this diagram that this is the, if this is the y this is y. So, we can see that it is varying from this 0to 1. So, whatever the yobserved data that we have got that we got and which we know how to obtained that it is a sampleestimate of this variance; if we do sothat will give you the variance of the y.

Now, after we get this regression line now, what is the variability of the y with respect to the regression line? So, basically we are looking through this access and we see that how it is varying across thisregression line. So, that is what isrefer toas means pictorially as this variance of y given x now if we want to estimate that one if we want to calculate that one this can be calculated as follows.Here, the conditional variance is assumed to be constant within the range of this x. So, we have this s square y given x is equals to 1by n minus 2 i equals to 1to n yi minus yi cap whole square.

So, this is the basically the estimate from this regression andthere arethis n minus 2is to make it that what it is calledthatunbiased and this you know that in the standard deviation we have seen that one degree of freedom is lost and that is why wemake it that n minus 1that we discussed in the earlier lectures and here one moredegrees of freedom is lost when we are estimating thatregression line.

Basically, there areif we see that there are two parameters that both alpha and beta has to be estimate through this regression line and that is why the two degrees of freedom is lost. So, make this estimate unbiased it is 1by n minus 2that we have to make. So, we can just do thiswe cansometimesfor this we can make that yi minusy bar that is a mean of y whole square minus beta cap squarei to 1to n xi minus x bar square. So, this is justfrom thisequation and you can see that this is basically that error and so, sum of square errors. So, which is that delta square by n minus 2that is thatconditional variance of y given x.

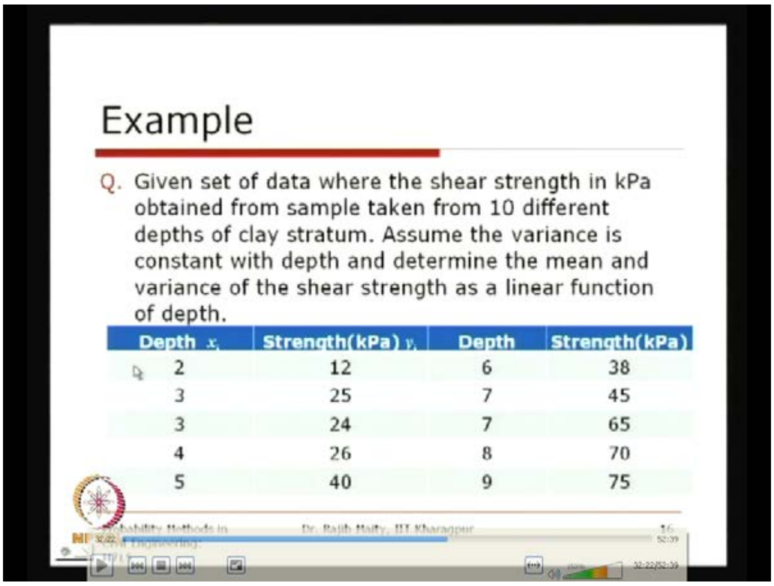
(Refer Slide Time: 29:12)



Now, taking into account the general trend with x the physical effect of this linear regression,what actually is happening through this linear regression is that, y onx that is the regression y on x can be measured by a reduction of the original variance of this y. So, the original variance of this y you know that which is that a s y square which I can get the data of this y and we can estimate what is thatwhat is thisvariance from the sample and that.But,through this regression when we do it that there is a regression there is a reduction in thatvariance. So, which can be expressed through this one minus thatvariance of this y given x divided by variance of thisvariance of y, sorry, thissquare will that Spower square. So, what you can see is that this is the original variance that was there in the data y and this is the variance after the regression that we that we got. So, this is basically how much is the reduction then that reduction is that what is the total minus what we got after this regression divided by what was their the total.

So, this is thatreduction in that variance and later on,we will justshow you that this can be approximated to basically the correlation coefficient.Obviously, the square root of this one is, can be approximatelyequal to the correlationcoefficient and that is basically the measure of how strong the relationship that we have measured for this Sy square; means this is the one by that you know it is the sample estimate of thisvariance of the data y. So, 1by n minus 1 summation of from 1to n yi minus yi,sorry, y bar square.

(Refer Slide Time: 31:07)

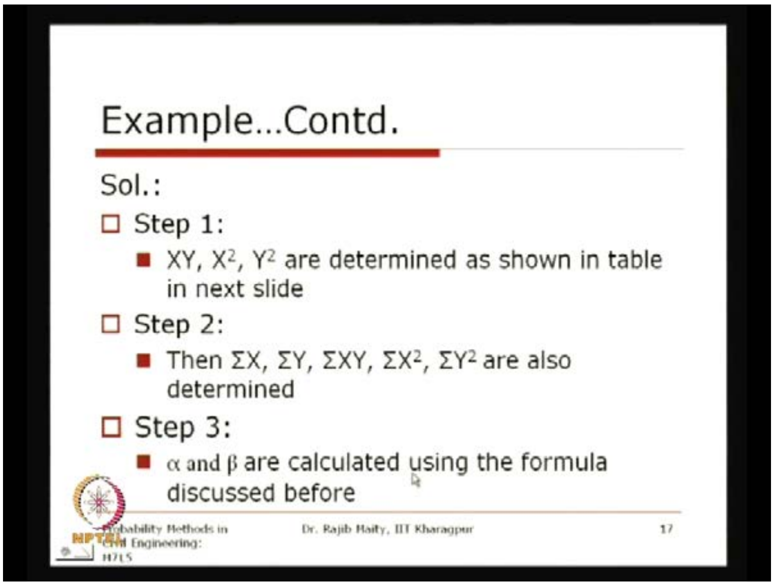


We will take one problem.Whatever we have discussed through the, for this linearregression, given the data where the shear strength in kilo pascal obtained from the sample taken from 10 different depths of this clay stratum, assume that the variance is constant with the depth and determine the mean and variance of the shear strength as a linear function of the depth.

So, here you can see that there aredepths are given the depth at 2,3again this 3. So, there are 10 suchdepths are taken.Thereare some depths are same; you can see here and we are getting thisdata. So, for the strength thatin kilo pascal,we are having these10differentdata set; this is thedepth and this is basically this depth is going to 3,3,4,5,6, 7,7,8,9and these are the correspondingstrengths. So, this 10 data set that we are having and we will follow whatever we have discussed just to find out the relationship between strength and thedepth.

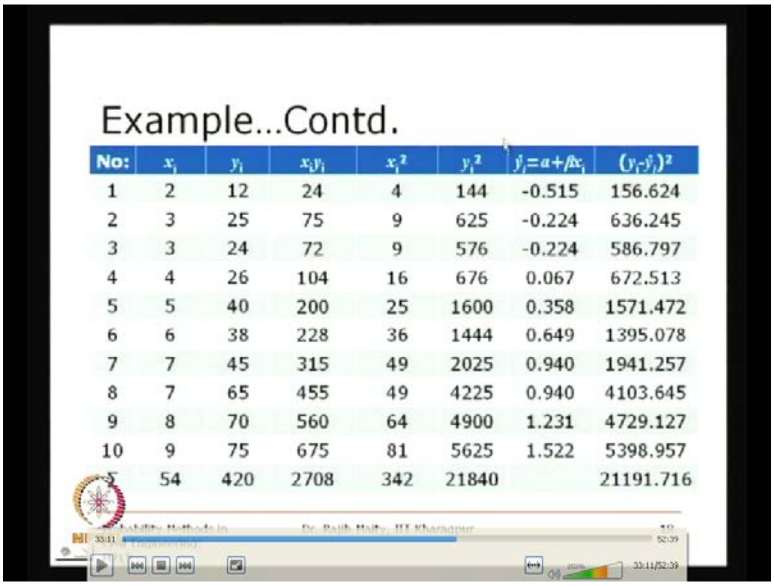
So, we willregress the strength on depth. So, our we can say that our variable y is here the strength and the x is heredepth.

(Refer Slide Time: 32:25)



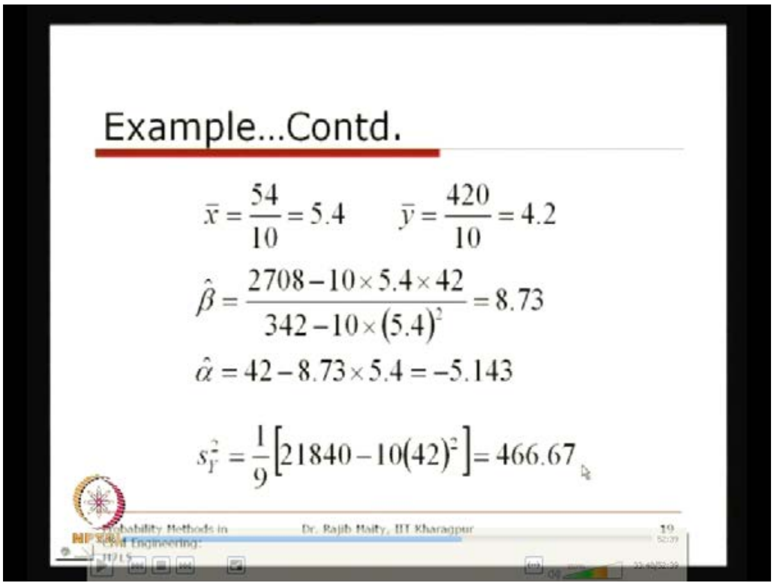
So, to get this estimate through this least square technique,we will first get theestimate of this parameter; that is XY,Xsquare,Ysquare are determined and this will be, we will show in this table and then the summation of this Xsummation of Ysummation,XYsummation of Xsquare and Ysquare and then, the alpha beta areobtained for using theformula that we obtained through thatfrom that least square estimate.

(Refer Slide Time: 32:52)



So, this is thedata for this different depth andfor this 10data sets are there xy,xi square y i square and then,these things we will just see. So, first we are having up to this and we are having their summation also.

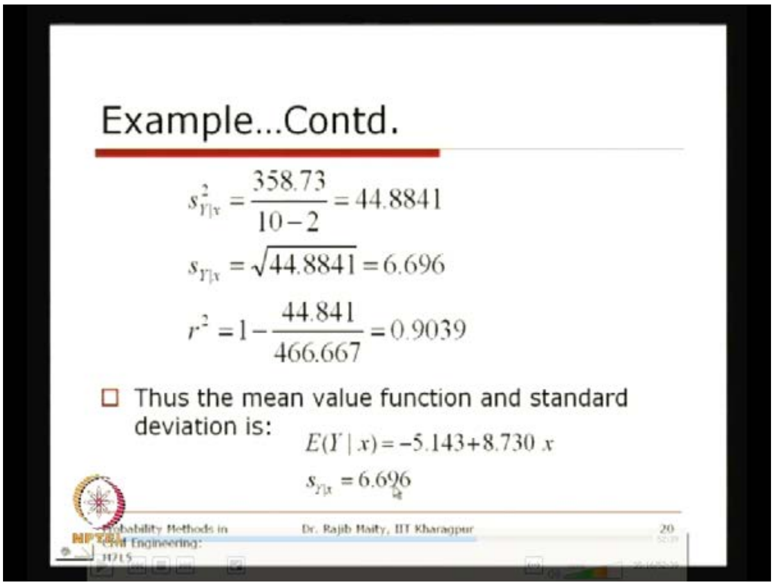
(Refer Slide Time: 33:14)



So, up to this of this table we know and using this information that is what is the power x bar is 5.4y bar is 4.2,sorry,it is 42this will be 424420by 10; so, it is 42.

So, this beta 1you know that this expression we will use. So, and we will get thatestimate of this 8.73 and alpha cap is the estimate of the minus 5.143 and this sy square there is a variance the or the total variance I can say now the total variance of this y is 466.67 now.

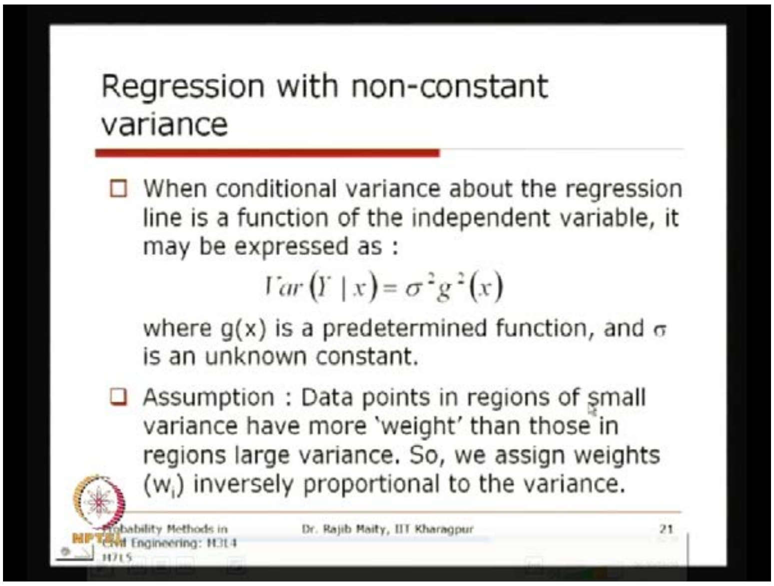
(Refer Slide Time: 33:49)



So, this now Sy given that x is your this 44.88 and the standard deviation is 6.696 andthisr square is equals to 1minusthese how much is the reduction is that 0.9039; and this alpha and beta that you can see it here. So, alpha is minus 0.50.143 and beta is this 1. So, this regression equation comes like this; that expected value of this y even x is that minus 0.50.143 plus 8.730 x.Now, using these things basicallythis relationship in the table we got thisexpression first. So, we are putting this x input and that alpha estimate and beta estimate and we get this one.

From here we are getting what is their errorof this that is y minus y caps. So, this minusthis and that square will give you that basically what is thisone that we get that error square now if we sum it up this is basically the sum of square error and we are using that information to estimate this what is that reduction in thisthat variabilityvariance in y. So, this one we have seen. So, this is finally,thatexpected mean is expressed through this expression and expectedvariance is 6.697 and obviously, this is constant over the entire range of x.

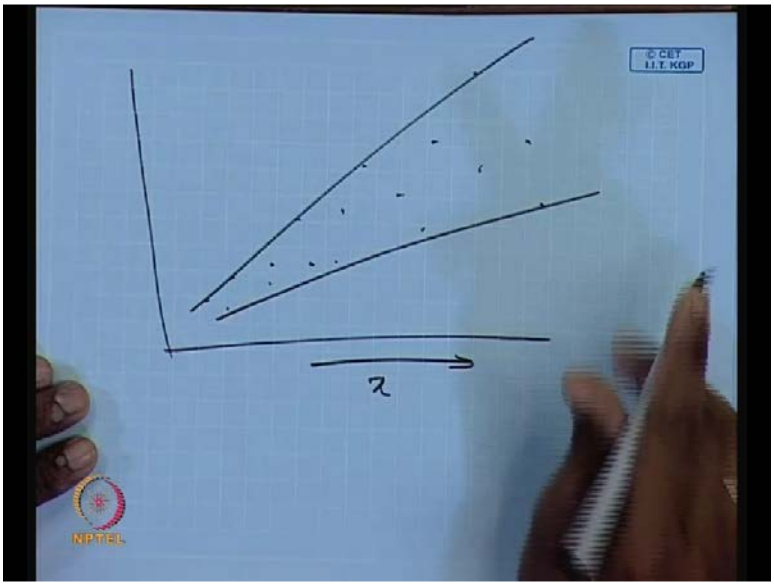
(Refer Slide Time: 35:22)



So, now we will deal with the regression with the non constant variance now when the conditional variance about the regression line is a function of independent variable, it may be expressed as variance of Ygiven x is equal to sigma square multiplied by g square x.Now, this g x basically is the predetermined functions.Somefunction is that, how it isvarying and this should be multiplied with the sigma square and when it is variance you know that any function or constant that is multiplied with thisvariance sothat, we make it square.

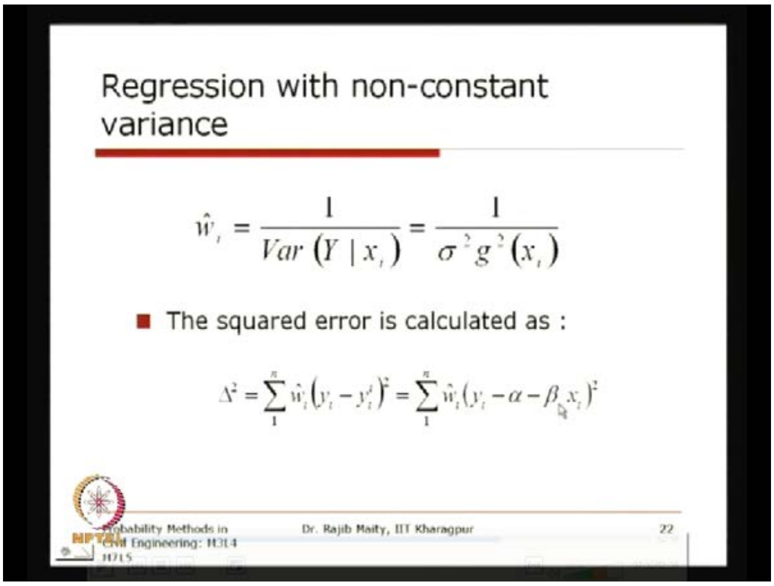
That is, we discussed in the earlier lectures. So, this is that sigma square g square x. So, now, thissigma is an unknown constant and here the assumption is that data points in the region of this small variance have more weight and than those in the region of this large variance. So, we assign the weightwi inversely proportional to the variance. So, some weight we have to put and our assumption is that when the data is having the small variance.

(Refer Slide Time: 36:33)



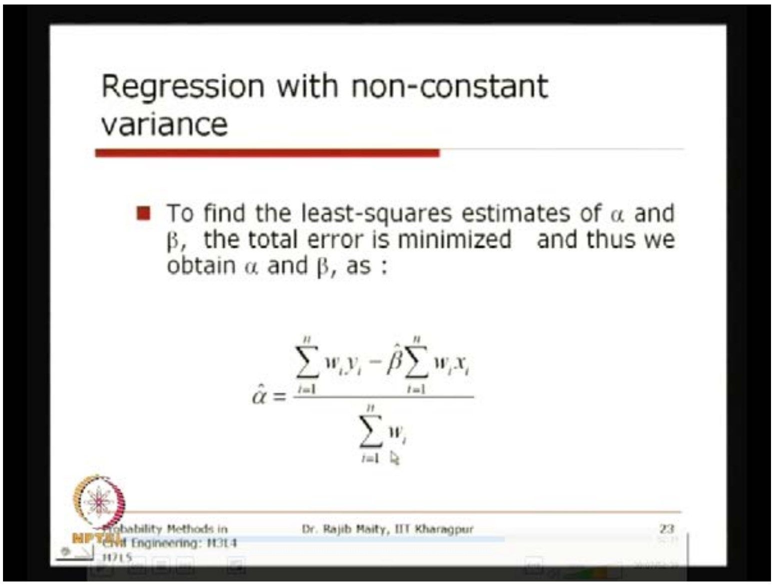
Now, if you see thisdiagram basically,if I say thatthis is varying means, suppose this what we can see in this literally. So, we can easily see that as it is goingand so,if this the x as this x is varying basically the range is changing. So,here the weight will be more in this zone where the variance is less and here the weight will be less where the variance is morebasically that is what. So,in this wayit isinversely proportional to the variance.

(Refer Slide Time: 37:03)



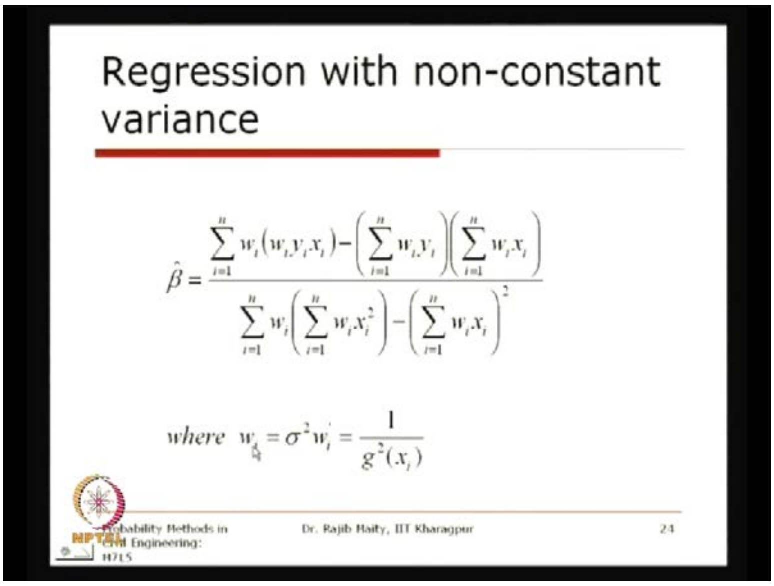
So, this how the weights are given. So, 1by variance ofYgiven xi which is the 1by sigma square g square xi the square error is calculated as this sigma square is equals to that this weight isweights we will put and then that your that difference square and sum it up. So,this is theiequals to1to n.

(Refer Slide Time: 37:33)



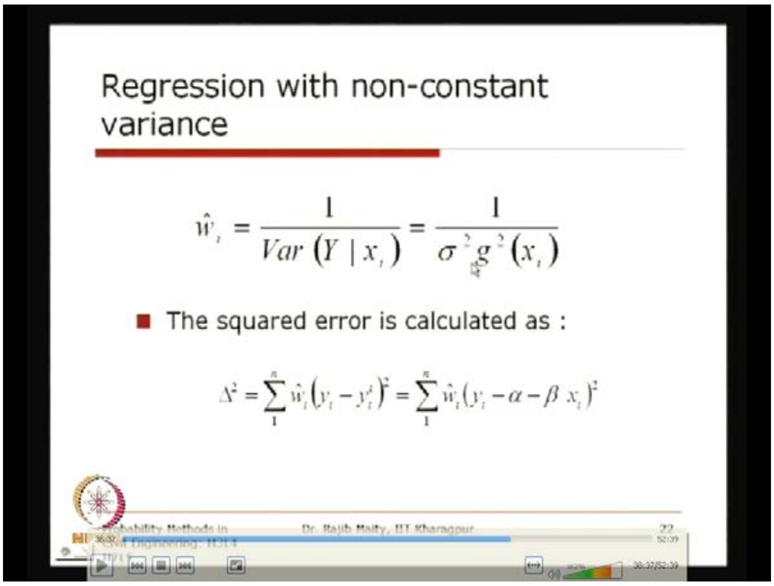
So,now this yi cap again that estimate will get from this alpha minus beta xi now to find thatleast square estimate of alpha and beta the total error isminimizeand thuswe obtain this alpha and beta and following the same principal that we have discussed for this constantvariance we will just get that error and error is partialderivativeis taken equated to 0and after solving those equation we will get the estimate of alpha is equals to through this expressionsay wi into yi minus beta cap of this wi xi divided by wi summation of all this w i.

(Refer Slide Time: 38:08)



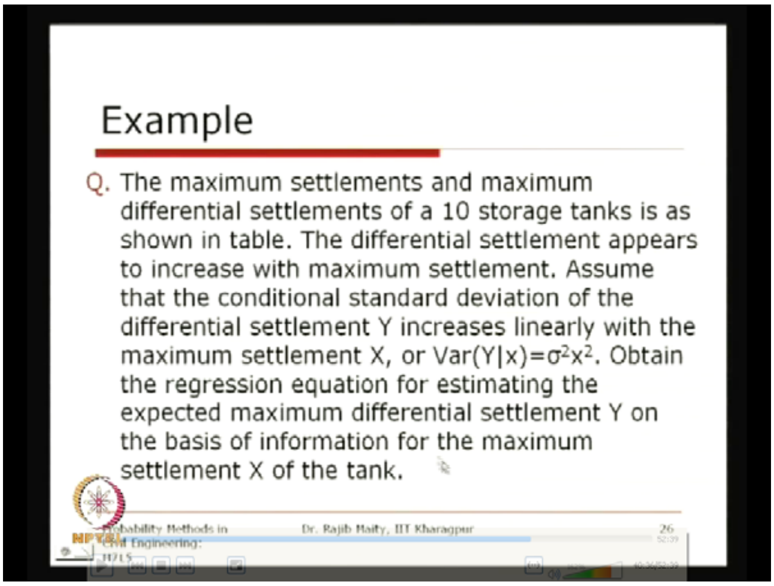
And, beta cap will be obtain through this expression even though this expression looksthroughlike a little bit cumbersome, but,thing is that this is we get following the same principal that we havedone for this constantvariance only thing here the one that weight function is coming and which the weight you can see that this weight is equal to sigma square wi prime and this sigma square.Ifwe just multiply whatever the equation that we have used here.

(Refer Slide Time: 38:36)



So,this sigma square will be cancelled. So, it will be 1by gx square.So,now this g x square generally same functionwe will used and thatfunction of this xshould be there to when we are determining thiswi to get the estimate of this alpha and beta and the conditional variance iscalculated as S i square is the s square g x square and Sxi –this is the standard deviation, the square root of this positive square root of this. So, s multiplied by the g x;you can see here that this conditionalstandard deviation is a function of that x where this Shere is that summation ofi equals to 1- 1to n wi into yi minus alpha cap minus beta cap xi whole square divided by n minus 2.

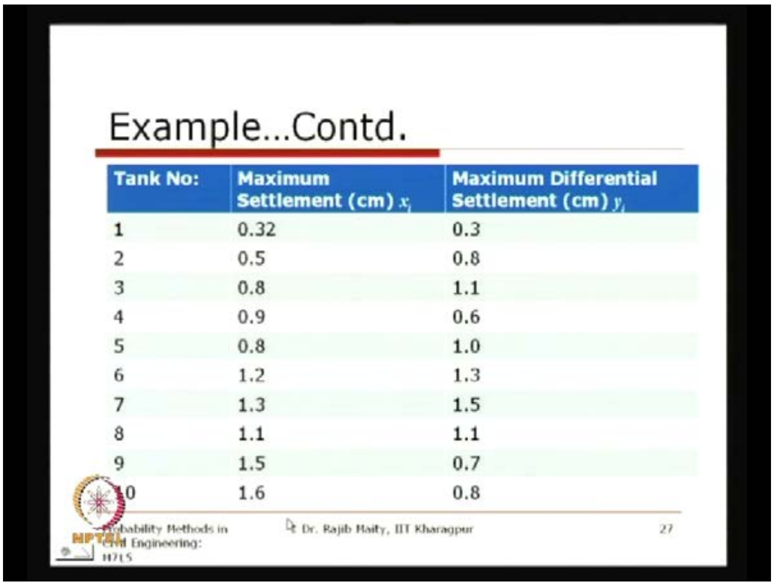
(Refer Slide Time: 39:30)



Now we will take one example on this one.Itwill be more clear in thatway where thevariance is dependent on the value of X.Themaximum settlement and the maximum differentialsettlement of 10storage tanks, this is wrong; of 10storage tanks is as shown in the table.Thedifferential settlement appears to increase with the maximum settlement assumed that the conditional standard deviationof the differential settlement Yincreases linearly with the maximum settlement Xor this is what is told.Thatis, linearly it increases; that means, that g x that the function that we have told this is g x is equals to X.

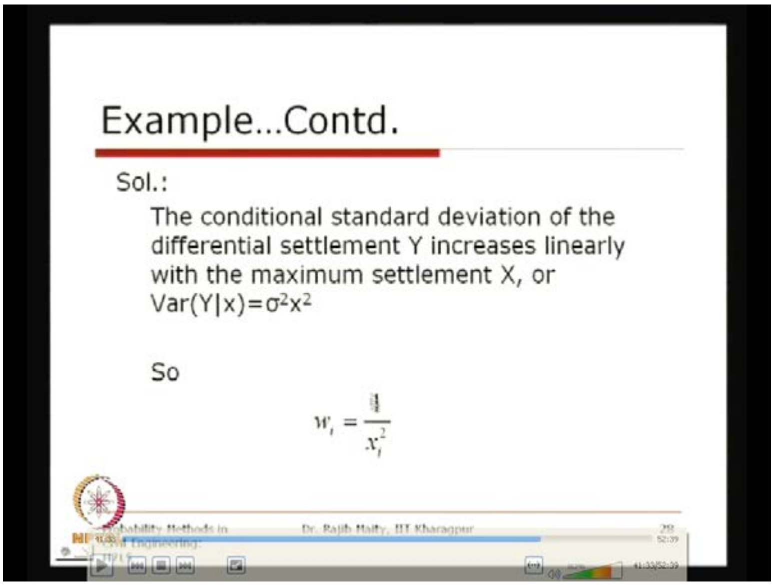
So, the variance of Y given Xis equals to sigma square x’s square;that function square obtained, the regression equation for estimating the expected maximum differential settlement y on the basis of the information for the maximum settlement of X of that tankto do this one; this is the data that 10 differentdata set is given here.

(Refer Slide Time: 40:39)



So, this is the maximum settlement there is a maximum differential settlement. So, here we have to regressed thatmaximum differential settlement on this maximum settlement. So, our variable here in thisfollowing the notation that we have used is the this is of oury and this is our x.

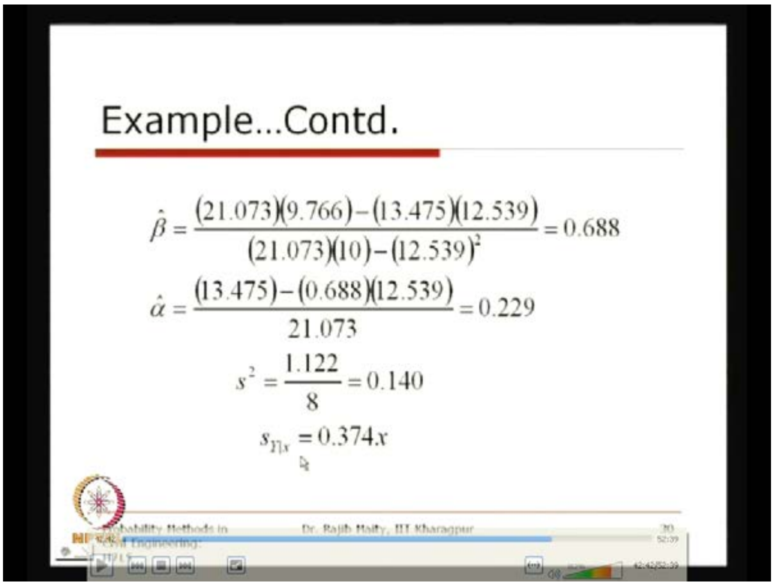
(Refer Slide Time: 41:05)



So,the conditional standard deviation of the differentialsettlement Y increases linearly with the maximum settlement Xorvariance Yon condition x is equals to sigma square x's square.

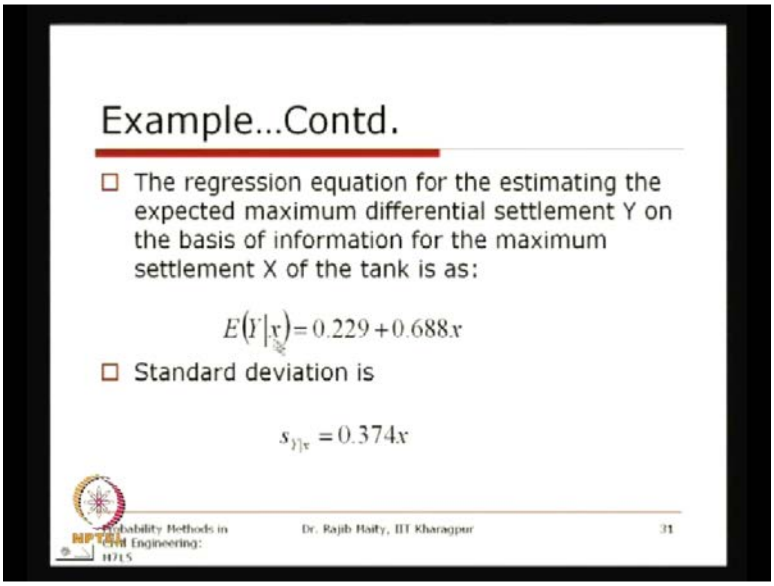
So, this is the relationship that is given. So, this x this function actually is predetermined as we haveseen in that theory. So, here that wi is the inverse of that functions. So, 1by xi square. So, this is the weight age that is the wi and with that w i,for all these we will get this weight andbasically, this is input.This is also we know the observed data this is the weight which is a inverse to this xi and. So, these things we cancalculate wi xi wi yi wi xi yi and wi xi square.

(Refer Slide Time: 42:09)



So, if we use this one andthenup to this of this table we can calculate and based on this we will estimate that alpha and betaand here the beta is estimate of this beta is 0.688 and estimate of this alpha is0.229 and the Ssquare is your0.140 and thisstandard deviation of y given x equals to 0.374 x.

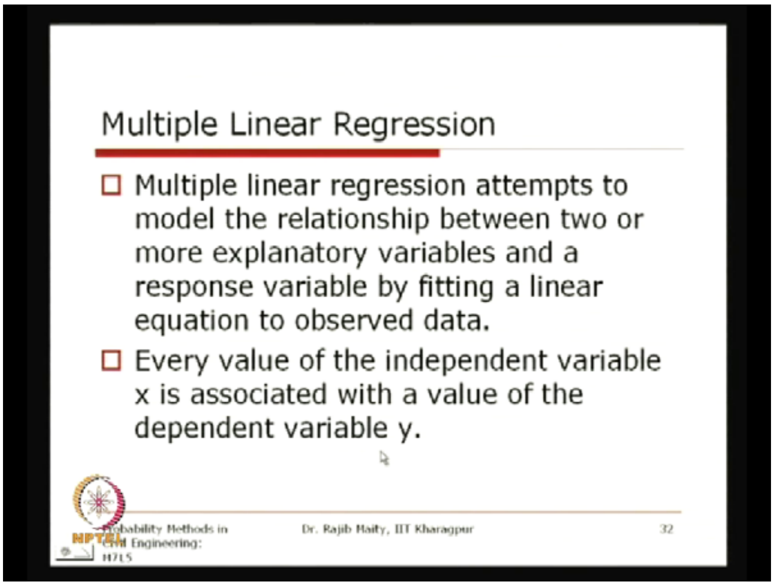
(Refer Slide Time: 42:45)



So, you can see that as x increases thisstandard deviation also increases which is the function of this xof thevariable x.Now, the expected value of this yis theregression equation for the estimating the expected maximum differential settlement Yon the basis of information for the maximum settlement Xof the tank is as expected value of this y given x is equals to 0.229 plus 0.688 x.

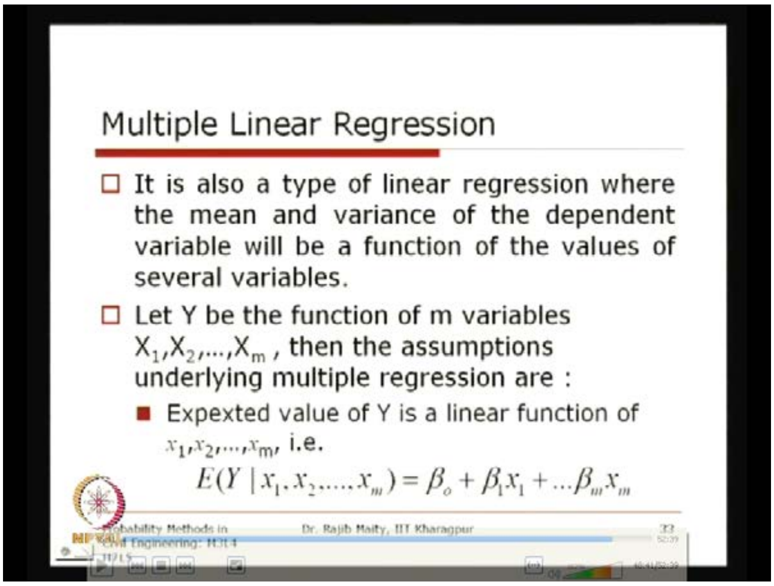
So, this is thatexpected value of y given x and this is expected value ofthis is thestandard deviation of y given x and now, using this relationship basically when how we are getting this 0.345 here that we have seen that Ssquare is this one.Basically, we are using this alpha and beta estimateto calculate this one first and this total we are gettingthis is the sum of square error weightage sum of square error and from there we are getting this s square and from there it wegetting thatgiven thatvariance, sorry, standard deviation of y given x 0.374 x.

(Refer Slide Time: 43:44)



So, next we will take that multiple linear regression and here you know that. So,far whatever we have discussed it is thatregression and onedependent variable, oneresponse, variable one targetvariable was there.Now, in case when we are having that more thanone random variable then we have to go for thismore than onedependent variable then we have to go to this multiple linear regression. So, this multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to the observed data.Every value of independent variable is associated with a value of the dependent variable y.

(Refer Slide Time: 44:42)



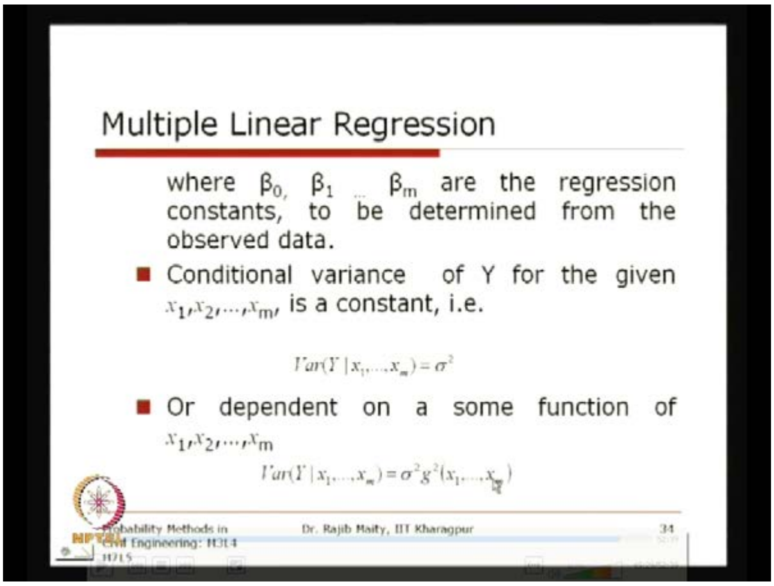
It is also a type oflinear regression where the mean and variance of the dependent variable will be afunction ofvalues of this several variables. So, here instead of that using that oneindependent variable that is the X; that Xand ourdependent variable was aYearliercase.

So,here what you can see is that Ybe the function of m variables instead of only one. So, far what we have discussed here is,Yis the function of m variables which is x 1x 2up to x m then the assumptions underlying the multiple regression are the expected value of Yis a linear function ofx 1,x 2up to x m that is the expected value of y given the information of thisindependent variable x 1,x 2up to x m equals to that beta naught beta 1x 1plus beta 2x 2plus up to in this way beta m x m.Onething I willjust, one correction I will just do before I proceedfurther in thislinear regression - withonebetween x and y. So, where only one inputwasthere I might have sometime mentioned that this x is in this expressionwhen we are regressing y on x,Imight have sometime mentioned that this x is yourdependent variable and y is thetarget variable.

So,the correction will be that x is your independent variable and y is your dependentvariable.Sometimes, for the y we canmention that this is the target variable, responsevariable, dependentvariable and all and basically, when we are referring to this o x this is the independent variable. Earlier in this case,when we were discussing the simple regression that time, only one dependent variable was there.Now, what we are discussing here is that, we are having more than onedependent,more than one independent variable to model thatdependent variable y and this is through a linearfunction which is thatbeta naught plus beta 1x 1plus up to this; up to the beta m x m. now, basically how the concept istaken through is that.

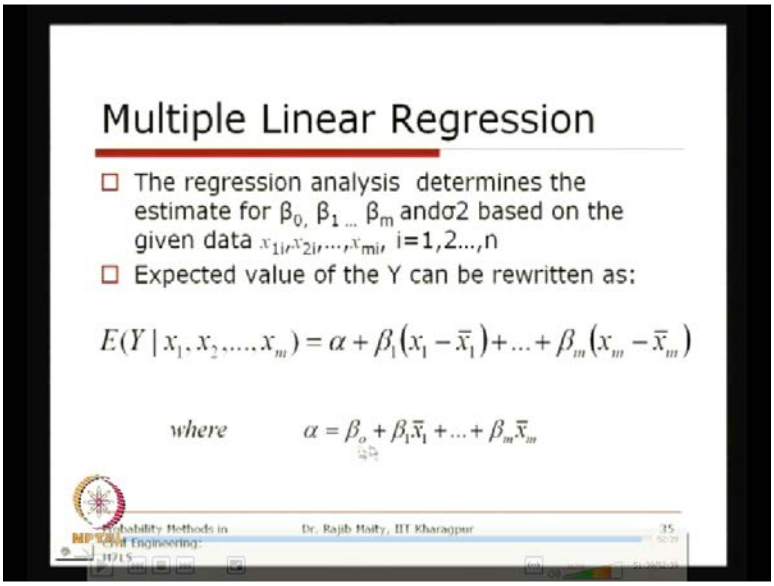
Now,for that when you see that there is only oneindependent variable and one dependent variable between x and y, we are basically fitting a straight line through the observed data point.Now, when we are having more than oneinput say for example, if I just say there are 2 inputs x 1and x 2and our target - our dependent variable is y then basically, you can visualize, you can conceptualize in this way that this is a 3 dimensional space over which the twoaxis is one for the x 1,other for the x 2and we are basically fitting onesurface; one straight, one linear surfacethrough the data point in the 3 dimensional space. So, this is in case of when there are two independent variables and one dependent variabley. now, similarly you can extend it to the higher dimension andthis; sothat, for the mindependent variable, the relationship is generally that b naught plus b 1x 1plus bup to the b m x m.Now,we will follow the similar procedure to estimatetheseparameters as well.Thatis, we have to first find out what is the error and that error should be squared up; sum them. So,there is the sum of square error then it is minimized with respect to the parameters to get thoseexpressions.

(Refer Slide Time: 48:44)



So,where this beta naught, beta 1,beta m are the regression constant to be determined from the observed data andthe conditional variance of Yfor the given x 1,x 2up to x m is a constant that isvariance of this Ygiven this input is equal to sigma square.Or, this is in case of when it is constant or it may be dependent on some function of this x 1,x 2,x m. So,when it is dependent, when it is varying when it is non constant as we have used in the simple regression case. So,this variance of y given x 1x 2x m is equals to sigma square multiplied by the square of some function of this x 1x 2up to x m.

(Refer Slide Time: 49:31)



Now thatexpression that is the regression analysesdetermines the estimates for this beta naught, beta 1,beta m and the sigma square, sorry, this will be square and the sigma square based on the given data x 1 I,x 2 I,x 3 I,up to x mi andi is varying from the 12n. So,we are having then n set of I can say that n set of data that isy 1,x 1,x 2,x 3,x m.Similarly,I will have another set ofthisdata. So, there are,m is the number for thenumber of the dependent variable and n is the number of what?Howmany sets of the observed data that is available to us.

So,based on thiswe can,whatever the expression that we have seenin this expression can beslightly modified as this one.Thatis, the alpha plus beta 1x 1minus x 1bar plus up to this that beta m x m minus x m bar. So,how we get this one is that, this x 1bar is the mean of whatever we have seen in this x, in thevariable of x 1and the x m bar is the mean of thatobserved data of the, for theindependent variable x m. So,basically what we are replacing is that thisconstant beta naught is basically a adjusted is basically replaced. So,this alpha you can see that this alpha is equals to that beta naught plusbeta 1x 1bar plus beta 2x 2bar plus up to this beta m x m bar.

So,herewhat we canget is that, from this expression we can estimate that is alpha, beta 1,beta 2,beta m and from that estimate of this alpha and obviously,beta 1,beta 2of the same.Ifwe put it here,we will get what is the estimate for this beta naught andthis one we will see.Wewill continue from this point onwards in our next lecture andwhat is in this linear regression part, what we have seen in today's lecture is the linear regression and linear regression with respect to the constant variance and someor the non constant variance.Wehave seen one example for each case and this. So, in the next class what we will see is thatmultiplelinear regression and then we will also see the non-linear regression and we will see theco relation as ameasure of how strong the relationshipis captured; that we will see in the next class; thank you.