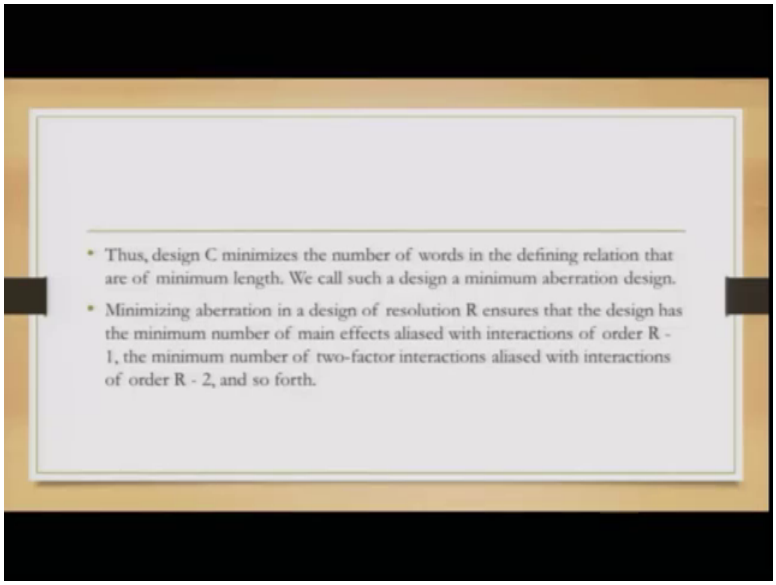
Total Quality Management - II
Prof. Raghunandan Sengupta
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 27
Two level Fractional Factorial Design – V

Welcome back my dear students. A very good morning good afternoon good evening to all of you, all the participants and all the students for the TQM-II lecture series and this is the 27th lecture under a the NPTEL MOOC series which means that the total number of lectures as you know is total 40 in number. So, we are I will just bit a little bit more than the half. So, if you remember that you are discussing the factor designs and how 2 to the power k, k is being the different combinations, you can take depending on the effects of the factors can be taken to consider different type of examples for that. So, we continue on those same lines.

(Refer Slide Time: 01:07)



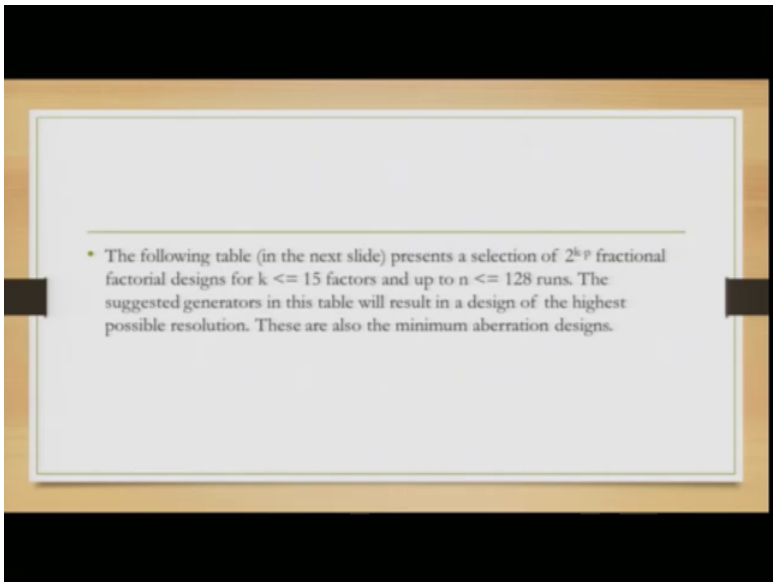
So, when thus when you are designing. So, thus design C minimizes the number of words in defining the relations or relationship between the variables and if you remember we would try to basically take the collective effect of the weight the factors in order to basically have the maximum effect and; obviously, the error would be minimized; what is the assumptions for the errors whatever distribution you will take? So, all these things we have been repeating in the last 27 lectures; obviously, or 26 lectures, I am sorry for

that and if you remember, we did discuss about the distributions in the first 7 lectures so; obviously, add on to that discussion for the 7 first 7 lectures and in the later part.

So, you will understand the repetitions are taking care of the concepts which are important. So, to continue; so, thus C minimizes the number of words in defining the relation that are of minimum length minimum length means, you are trying to basically minimize some distance function. So, we call such a design a minimum aberration design.

So, minimizing aberrations in a design of resolution are on depending on the level of accuracy you want, robustness you want, sensitivity is also a factor, it will ensure that the design has the minimum number of main effects aliased with interaction on the older R minus one effects. So, which means that minimum number of 2 factor interaction aliases with the interaction of order R minus 2, then R minus 3, R minus 4 would happen which will give you the overall picture; what is the effect the following table which will consider in the next slide presents the selection of the k 2 to the power k minus p.

(Refer Slide Time: 02:49)

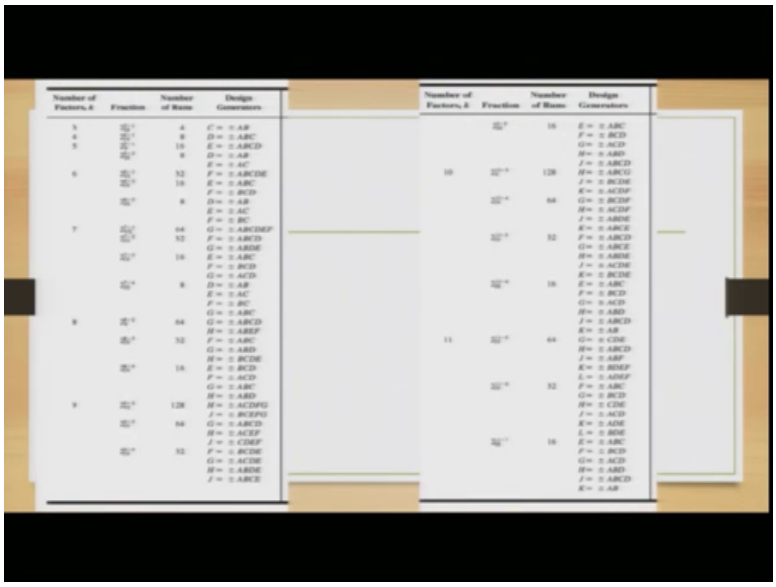


Now these are the factorial design. So, the concept of k and p have already been discussed in the in the factorial and design models though briefly I do understand that we did not go into the depth. So, we will again repeat that accordingly.

So, for a for presence a selection of 2 to the power k minus p fractional factorial designs for k less than is equal to 15 factors up to about 128 runs depending on n. So, if you remember each factors had samples. So, we consider initially the number of observations or I would not use the samples means the number of observations makes a sample. So, number of observations in each factors was considered equal.

So, if he had say for example, k factors and each of them had n number of observations, total set of observations was k into n and it could be different also. So, he could take the averages so on and so forth, we G which we did discuss in the 24th, 25th, 26th lecture. So, the suggester generators in this table will result in the design of the highest possible resolution depending on the accuracy we want. So, there are also the minimum aberration design which can be considered.

(Refer Slide Time: 04:10)



So, this sorry for the slide is a little bit too intense in the sense there are lot of information here and considering that I am only trying to give that in one slide it may not be immediately possible for you to and to read it plus please bear with me because everything has been taken from Montgomery's book so that, if you can either check it in the library or find it in your college and no need to buy which I keep saying that if you can get those informations, it would be very helpful to refer what I am saying with respect to the slides which are there. So, the number of factors which is k are given on the first leftmost column. So, we start with t 3 4 5 6 7 8 9, then you go to the next table

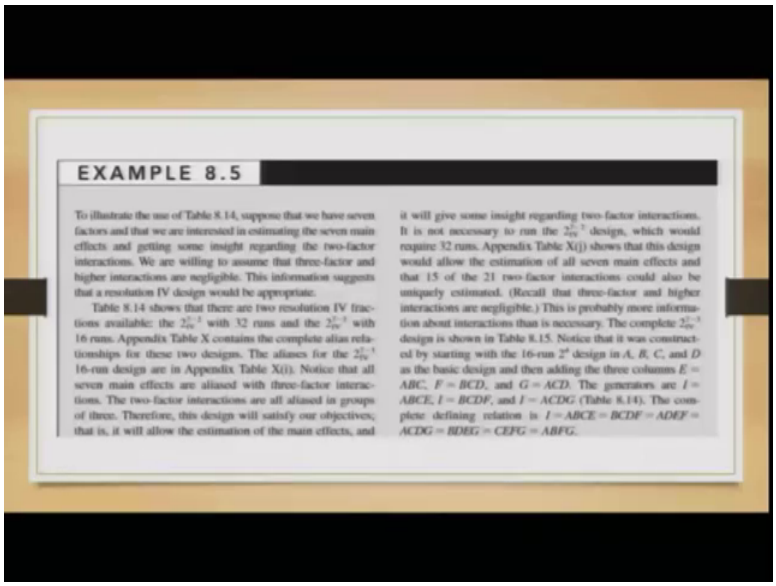
which is 10 11. So, there combinations increases. So, the fractions would be there would be if you remember it was 2 to the power k minus p.

So; obviously, we would have 2 to the power 3 minus 1, then you will have basic here 2 to the power 4 minus 1 2 to the power 5 minus 1 2 to the power 5 minus 2 2 to the power 6 minus 1 2 to the power 6 minus 2, it will continue and for the 11 when the number of factors are k is equal to 11. So, some of the combinations are 2 to the power 11; 11 is the factors minus p, it can be considered as 5 as 6 and 7 so on and so forth. So, the number of runs which you do depending on the combinations the design generators can be a combination of A B can be combination A B C can be combination A B C D depending on what do you think is the highest influence.

And the number of runs would be dictated according to the experiment and the number of ob observations which you have for each factors based on that; we have this table which gives you, again, I will repeat the first column number of factors then the fractions which is technically 2 to the power k minus p, then you have the number of runs and design generators which will basically, give you the best possible effect of the factors from which you can find out what are the minimum number of the optimal number of factors based on which you can predict the maximum amount of dependence that is what that is what you what you are trying to do and; obviously, there would be errors.

But as I mentioned that the white noise of the errors would have certain properties certain distributions certain mean value certain variances which we did discuss I will discuss that diamond again. So, consider this example; obviously, we have considered the aluminium, example, the smelting example different type of examples, then the battery one where do you want to design the battery at different temperature using different materials. So, we will continue on another consider another example also to illustrate the, this use of the table.

(Refer Slide Time: 07:07)



Which you have just consider suppose that we have 7 factors where the k is given and that they are interested in estimating the 7 main effects their individual effects and; obviously, will try to replace them in the combined one and getting some in insight regarding the 2 factor interaction.

So, we are willing to assume that the 3 factor and higher interactions are negligible so; obviously, three factors, four factor design would have been basically like two to the power k 3 to the power k 4 to the power k, but we are taking limiting into the power two factor model either on to basically two things minimize a number of calculations minimize the overall complications in the calculation, but still get the best effect based on which we can predict the effects of the factors.

So, this way and; obviously, 3 and 4 would give us very theoretically very nice results, but practically their essence would be basically we lost because if you are able to do with the 2 factor model. So, we would stick to that in order to minimize our complications, but still get the best answer. So, we are willing to assume the 3 factor model. So, I am just reading it repeating it we are willing to assume the 3 factor and the higher interaction which are basically negligible this information suggests that resolution of 4 design would be appropriate based on which you will basically find all the effects.

So, this table which you just discussed shows that there are two resolutions of the 4 fractions. So, one is basically to the power 7 minus 2 another is which basically would

result in 32 runs and 2 to the power 7 minus 3 which will result in 16 runs. So, the appendix would contain the complete alias relationship for these 2 designs then aliases for 2 to the power n minus 3 which is basically 16 runs are given in the next table which is 10; one notice that we will notice that all the 7 main effects which is the 7 2 to the power 7 minus 2 which is our 7 minus that is 7 or the value of k ah. So, there.

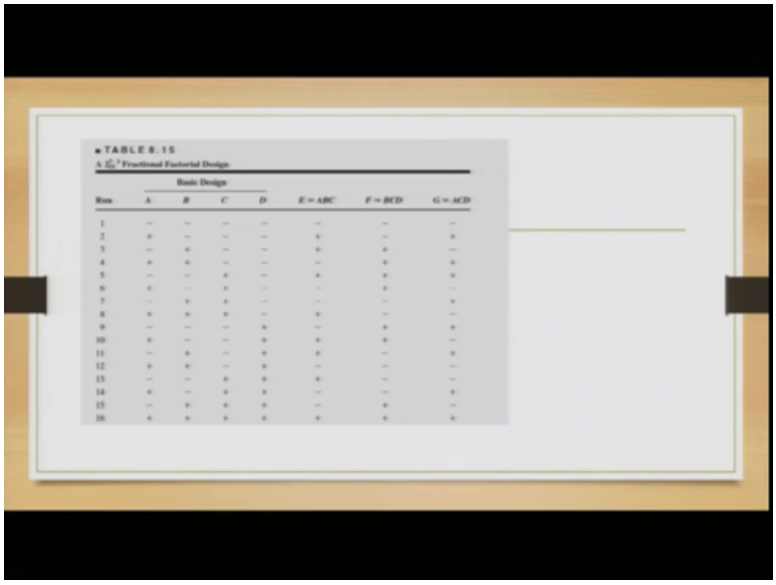
So, this notice that all the 7 main effects are aliased with the 3 factor interaction. So, k minus p; p is 3, therefore, the design will satisfy objectives that is it will allow the estimation on the main effects the secondary effects tertiary effects and; obviously, as you combine them they and the combined effect may decrease, but we will keep it maximum to a certain level means the combination such that we get the maximum effect or the maximum relationship which is coming out therefore, this design will satisfy our objectives that is it will allow the estimation on the main effects.

And it will give some insight regarding 2 factor interaction which is our main focus again I am saying it is 2 to the power that whatever k minus p is not 3 not 4, it is not necessary to run the 2 to the power 7 minus 2 design which is k minus p which would require about 32 runs. So, the appendix attend j shows that this design would allow the installation of all 7 main effects which is there and that this 15 of this twenty one runs for the 2 factor interaction could also be uniquely estimated and you can find out the results. So, this is probably more information. So, this is probably more information about interactions that is necessary would be obtains, but we will keep it to the minimum level because calculations is also on effect.

So, we notice that that it was this table was constructed considering the 2 to the power 4 design which is basically A B C D which you have. So, the combinations which will have would be basically combination A B C combination of B C D combination A C D which you are mentioning as E F G. So, the generators which you will have once you find out would be A B C E, then another generator will be B C D F another generator would be A C D G which you will see in the table the compliant and we have already checked in that table which was just finished in the last slide the complete defining relationship which is basically for I which is the indicators on the overall factor design which you are considering are A B E. So, combining them to find out the base effect another A B C D E F, one is A D E A D E F, one is A C D G, one is B D E G. So, these are the factors.

So, the factors would be named differently like it can be temperature it can be pressure it can be humidity depending on the example which you have to continue. Let me continue reading it, it will be B D E G, C E F G and A B F G depending on the level so; obviously, if you take the three factor model or the four factor model where k changes; obviously, the completions will come where the different combinations of the indicator so called variable or the combinations of factors would increase then; obviously, increase would result in complications in complications you need not in the answer, but in the calculations.

(Refer Slide Time: 12:40)



So, you have to consider them accordingly. So, once you have the 2 to the power 7 minus 3 where k 7 and p is 3 the factorial fractional factorial designs are. So, if you consider run starting from 1 to 16. So, the basic designs which you have considered what the factors was A B C D and the combinations we consider are given.

So, when you combine them. So, you have basically E which is combination a I am reading the last 3 columns first and then come to the combinations we are taking you will basically have A E which is combination A B C F which is combination B C B which you have found out and the last one is G which is a combination of A C D. Now if you consider the different factors or the levels of interactions which is taking place between these individual factors which is A B C D and; obviously, this k becomes 7 because you are considering the fifth one is E, sixth one is F and seventh one is G. So, if I combine all

of them through the combinations are say for example, I am arbitrary taking one the first row which is the first run.

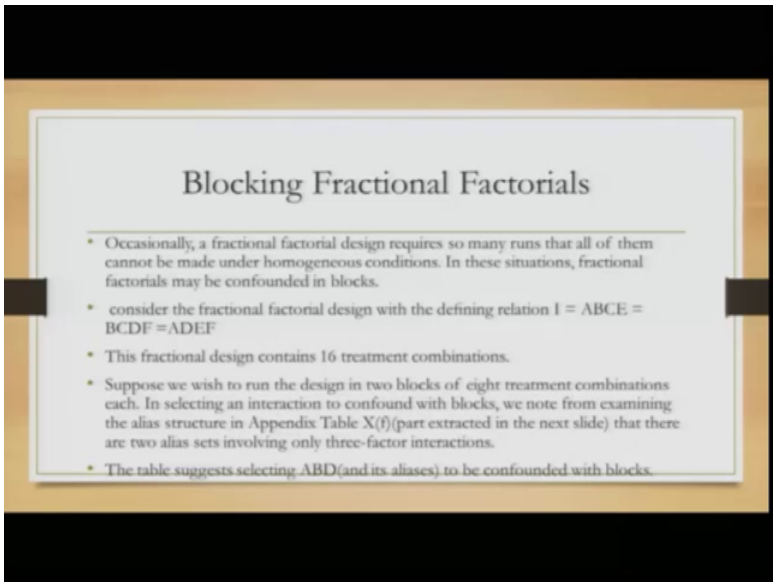
So, here the effects we are considering they are on the lower side that is why it is then if we consider the second row which is run to then I am considering the effects of a and E as positive and G is also as positive and we consider the effect of B C D and F as negative. Now, if you pay attention to this second row which is the second run, I have repeated that again and I am going to basically say that here if and an effect is positive so; obviously, it would mean D does not have C does not have and D does not have, but let us consider concentration of e.

So, what is he is a combination of a B C so; obviously, the effect of A is there. So, it will be positive. Now let us concentrate on the second last factor which is the sixth factor which we consider as F. So, what does F combined have it is B C D. So, does it effects your answer is no. So, hence the effect is on the lower side which is negative let us go to the last factor which is the seventh one which is G which is a combination of S C D does it have the effect of A, yes. So, it is positive. So, if you do this in these combinations. So, consider the say for example, the fourteenth run.

So, fourteenth run has A is positive B as negative C and D is positive. So, if we consider the combinations which is happening and depending on the positive and negative effect you will find out that a E F and G would basically have an effect on a positive side on the negative side; that means, higher effect lower effect similarly the 16th run. Now remember one thing the positive and negative which you are taking out are the are the effects on the higher side or lower side so; obviously, they can be 3 stages of effects also like high medium low or it can be 5 all. So, I am giving hypothetical examples it can be 5 also like very high moderate low very low.

So, these combinations; obviously, would result in a complicated table we are not going to consider that we have just mentioning that positive and negative every effect as a plus effect which is going to increase his influence negative means is going to decrease in the influence consider this that on those lines.

(Refer Slide Time: 16:07)



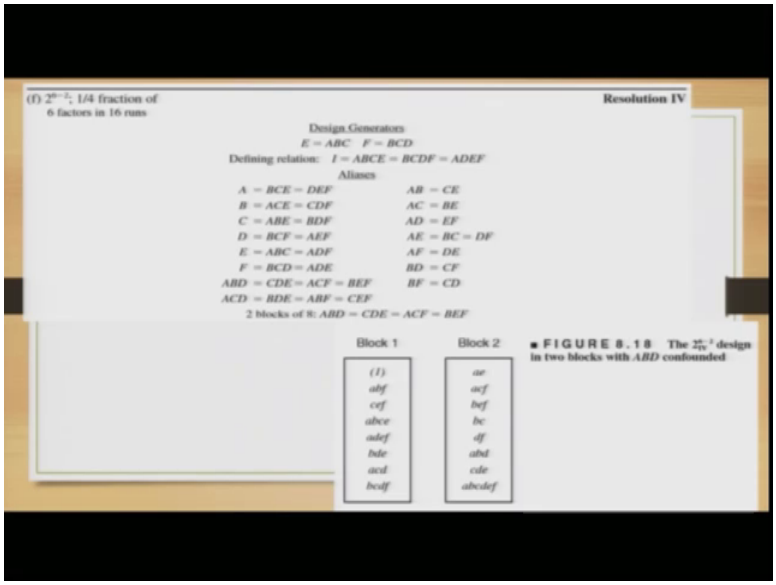
So, now blocking the fractional factorials we will consider it further. So, occasion occasionally a fractional factorial design requires so many runs that all of them cannot be made under homogeneous conditions with the condition. So, that of say for example, you are taking the factors as temperature humidity pressure. So, if they are homogeneous the effects of them we are going to consider on the same level that may not be possible in these situations fractional factorials may be compounded in blocks to find out the combined effects.

So, consider the fractional factorial design with the defining relationship with I the indicator factors of the combine where you have A B C D C E. So, you are combining the factors individually you have a you individually a individually of C individually you have E. So, you combine them to find out their combined effect similarly other one is B C D F; again we combine them individually you have B C D F. So, you take their effects in a collective manner so; obviously, the positive and negative would be countered accordingly and the last being ad E F; you will do the same thing.

So, this fractional design contains 16 treatment combinations as we have already seen and we can find it out. So, suppose we wish to run the design in 2 blocks of 8 treatment combinations. So, 16 combinations would be broken into 8 and 80 can we other combinations also, but we are taking a 8 and 8 in selecting an interaction to come to confirmed with blocks we note from examining the alias structure as given that there are

2 alias sets involving only 3 factor interactions. So, even though the combinations are A B C D, but in that case we find out that dividing 16 runs into 2 blocks of 8 8; each you have 3 aliases which were the factor interactions are only 3 in number. So, the table suggests selecting A B D to be confounded with the blocks such that we can find out the combined effect accordingly.

(Refer Slide Time: 18:18)



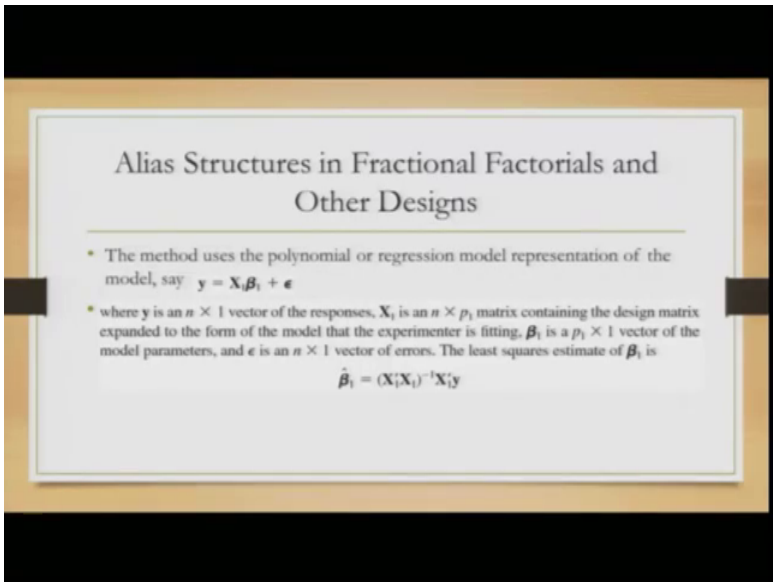
So, now we take the factory real factor design as 2 to the power 6 minus 2. So, if it is 6 minus 2 the design generators are you are considering E as A B C and F as B C D. So, defining relationship the indicator which you get are as I mentioned a B C E B C D F and A D F based on the we proceed. So, aliases when we find out and the combinations would be a would be clubbed with B C E and D E F; that means, the effect factors are being combined together in order to reduce the overall composition of calculations. Similarly, you will have B you have and we and in that case it will be A, C is equal to B. So, these equality means the effects; we are going to combine them in that manner the third one is C which is A B E and B D F similarly the fourth one is D which is equal to B C F and A E F considering that we basically continue. So, once we have the blocks are defined accordingly.

So, the blocks are in block; one you have I will read it the combined effects are for the factors considering that you have initially started with A B C D are and; obviously, the later on only combined those factors under the headings or give the name as E in

different combinations of A B C D, then F different combinations of a B C D G different combinations of A B C D. So, once the blocks are there the first block would basically have I am reading some are I repeating some of the factors which is A B F C F A B C E so on and so forth will be C D F and from block 2; you will basically have A E A C F so and so forth; in the last one which is the combination of A B C D E F.

So, this gives you the combinations of 2 to the power 6 minus 2 factors where k 6, again, I am repeating why 6 because you have a man you have basically consider them at one level lower and p is basically and that p value is basically 2. Now we will consider the alias structures or the factorial designs for the regression model.

(Refer Slide Time: 20:38)



So, let us first consider the polynomial regression on the simple regression concept and then explain accordingly. So, the method used the polynomial of the regression model representation of the model which is y. So, this bolt means they are vectors and matrices. So, you have basically l say for example, a k number of independent variables which are basically denoted by capital X 1, capital X 2, capital X till capital X k. So, these are the factors which affect independent variables and; obviously, the combination of these independent variables considering the multiple linear regression to be true you consider the overall effect is there such that you are able to predict forecast whatever what is use the dependent variable which is y.

So; obviously, when you are trying to predict there would be errors. So, if you plot the errors along with the best fit line. So, in regression you what you do want to do is you want to basically fit the best fit line. So, the best fit line would definitely not go through all the points because we draw on join all the points they are not a best fit straight line. So, they would be differences between all consider the distance would be there between the actual coordinates and considered in this let us consider.

So, even though I did mention that you have basically k number of variables which are independent, but consider that you have to make our life simple and I can only explain that in the 2 dimensional in the 3 dimensional case consider that you are without drawing. So, if you if you read if you understand what I am saying this will definitely make sense and; obviously, I will explicitly mention that later on consider that you have the very simple concept of y is equal to m X plus C which you have all done in basic class nine or ten now what you have is y is basically the value which you want to predict X is basically the readings which you have taken.

And there is only one X 1 X means it is a scalar; obviously, X would have different values and m is the slope; that means, slope that is the rate of change of y which is taking place of X and C the intercept. So, if we have in a 2 man dimension what you have you have in the X axis all the X values which are plotted in the y axis you have the y value is plotted and if you basically fit a best fit line. So, where it intercepts the y axis that is the height is C and the tan of this angle of y by X whatever y and y and whatever X length you take that tan of that angle that is equal to m. Now if you go to 2 dimension 1.

So, you will basically consider that X 1 is being measured from my I am consider that where I am standing is the origin it is going towards you I am measuring X 1 and if I consider say for example, X 2 being measured from my point again it is basically the origin and it is going to the right and I am measuring the y values along the vertical direction. So, if you again plot the X 1 and X 1 and X 2 as coordinates, you will basically be able to predict y.

So, they would be a 3 dimensional points and in the in the 3 dimensional case they would be a plane so; obviously, if you go to the higher dimension there we use the word a hyper plane. So, I am not going to go into such details or such complications. So, when you consider the regression model what you have is basically some axis these are the actual

values which you observe you can find out the reading consider you want to predict say for example, the rainfall.

And you think the pressure temperature and humidity at certain place are important to predict the rainfall. So, hands pressure temperature humidity are X 1, X 2, X 3 correspondingly and the rainfall is y. Now when you get the actual values of the humidity pressure and temperature and you predict it; obviously, they would not be exact values. So, they would be any difference which is the error. So, if you check this slide it is basically used in an error that error term which is the white noise in the regression concept that would have some properties.

So, the technically the properties which we generally consider in the regression model are I will mention one by one and I am only mentioning there I am not going to go into the actual theoretical either proofs or the mathematical equation as of now I will come to that later on. So, basically what you have the distribution of each of the axis are considered to be normal hence the y distribution is also normal and also the fact that the error term which you have will also be normally distributed with the mean value of 0 and a standard deviation which is fixed which is sigma square or at the best is one we can basically consider that the variance of the errors as one depending on the problem formulation point one point number 2 these values of beta one to beta k which you are taking for each individual independent variable. So, for the case of temperature humidity and pressure we consider the 3 values which is k is equal to 3.

So, these values of beta one beta 2 beta 2 v which we have are technically the rate of change of y with respect to X 1, X 2, X 3 considering the other 2 are fixed which means that if beta 1 is the rate of change of y with respect to X 1, we are considering the X 2 and X 3 are fixed. So, if you are considering the beta 2 as the rate of change of y with respect to X 2, then we are considering X 1 and X 3 fixed. So, it is basically the. So, called marginal rate changes of the dependent variable with respect to the independent variable we also consider the relationship or the covariance or the correlation existing between all the combinations of X 1, X 2, X 3 is nonexistent which is 0, we will also consider the covariance existing between the error and these X s, X 1 to X k are 0. So, there are different very simple mathematical notation which are which I said I am skipping, but just I am telling on a very for the for the general understanding.

So, once you listen to this you will understanding understand how the regression model is done. So, if you come back to what I said for the last 3 4 minutes, I come back to the first bullet point it mentions that y which is the dependent variable is equal to X 1 now X 1 is basically here is a matrix because y values you are taking from y 1 to y n what is n, you will ask that question n is the number of readings which you have so; obviously, if you have y 1 y 2 y 3 y 4 till with the y n values.

So, correspondingly you will have X 1 values it readings are from X 1 1 to X 1 n where the second suffix would basically mean the number of readings visit the first period reading second period reading third period reading till the n th period reading similarly you will have X suffix 2 1 2 X of X 2 k 2 n.

Then similarly you will have for the k one x suffix k one till x suffix k n and the errors would always; obviously, for each term there would be m for each reading the set of reading. So, X 1 1, X 2 1, X 3 1 till X k 1, you have y one and you have an error term also similarly in the second row if you are imagining it will be y 2 then the x values would be X 1 2, X 2 2, X 3 2 till X k 2 and similarly we will go to the k n th one. So, if you consider the X values it is a matrix of size n cross k the errors are the size of n cross one beta is the size of k cross one because there are k number of betas and y value is basically n cross one.

So, if you consider that. So, why is n cross one vector of responses which I am again repeating X 1 is the size of n cross p one matrix. So, here I am considering p as the number of reading the axis which you have it what I have been mentioning was key continuing the design matrix expanded to the form and beta is basically of size p cross one which I mentioned k cross one vectors of the model parameters and epsilon is the errors which is also n cross; obviously, I mentioned the errors. So, the least square estimate is given by beta hat.

So obviously, that means the estimator is given by the inverse of the manipulation of the transpose of x factor into the x matrix into is x multiplied by the transpose of x into y. So, with this again I would; obviously, I will come to the effect of y; how we are trying to find out the beta again it, I will repeat it in the in the very simple qualitative sense not in the quantitative sense.

So, with this, I will end the 27th lecture and continue more discussion about this factor of models in the concept of digression later on. Have a nice day.

Thank you very much.