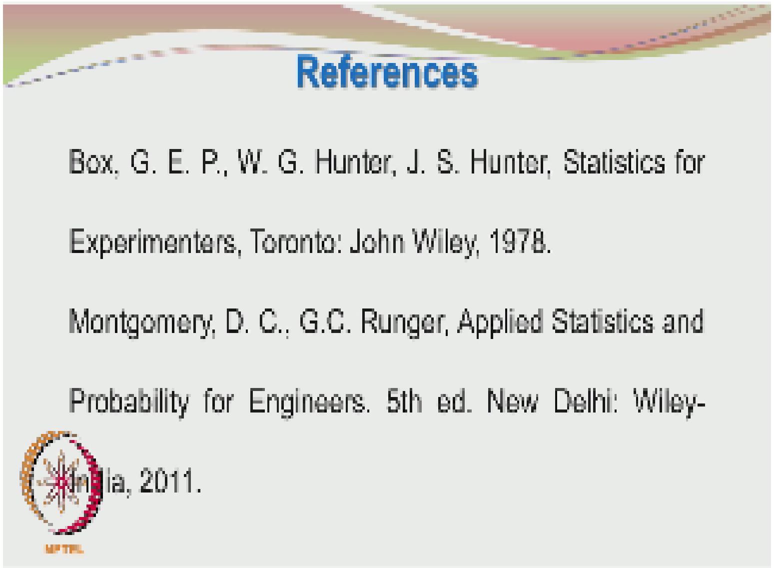
Statistics for Experimentalists
Prof. Kannan. A
Department of Chemical Engineering
Indian Institute of Technology – Madras

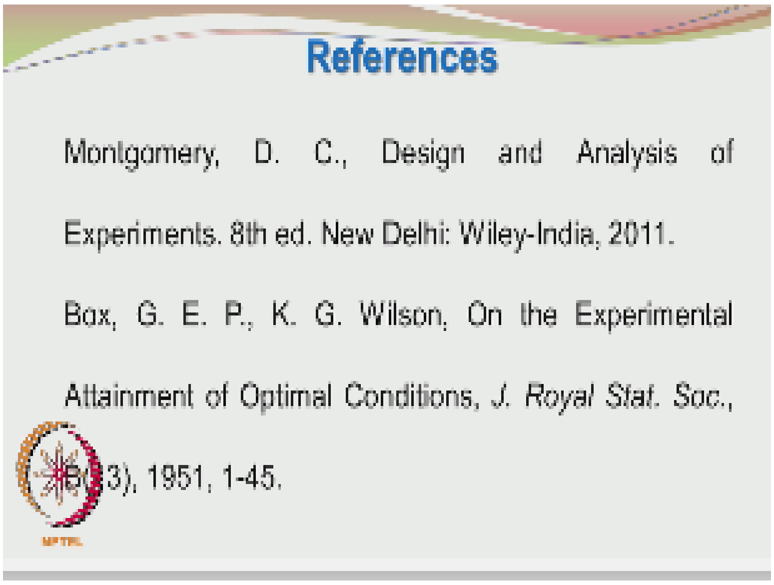
Lecture – 30
Factorial Design of Experiments - Part A

(Refer Slide Time: 00:27)



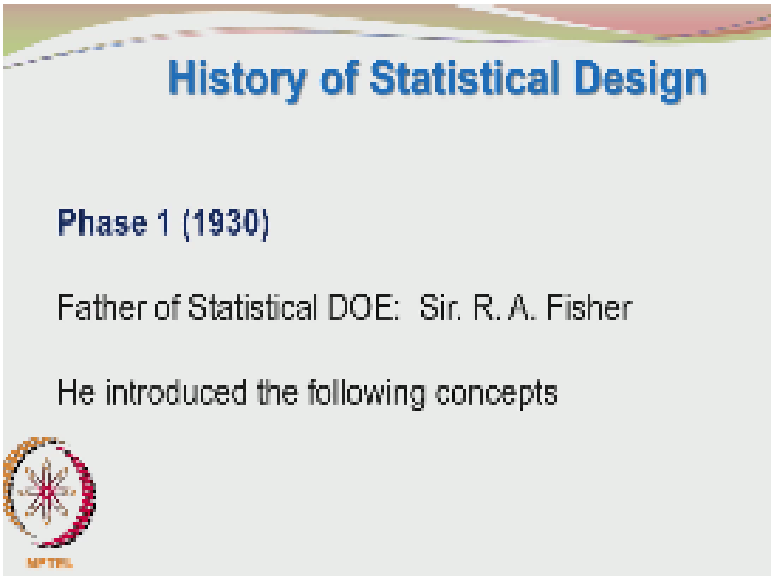
In today's series of lectures, we will be looking at factorial design of experiments. Some of the popular references for this topic are shown in this slide. The first one is the book by Box Hunter and Hunter, statistics for experimenters by John Wiley published in 1978. The prescribed textbook, Montgomery and Runger applied statistics and probability for engineers, 5th edition, Wiley India, 2011.

(Refer Slide Time: 01:02)



There is also the more detailed book on design of experiments by Montgomery, design and analysis of experiments, 8th edition, New Delhi Wiley India, 2011 and there is also the reference to the paper by Box and Wilson on the experimental attainment of optimal conditions, Journal of Royal Statistical Society, 1951.

(Refer Slide Time: 01:34)

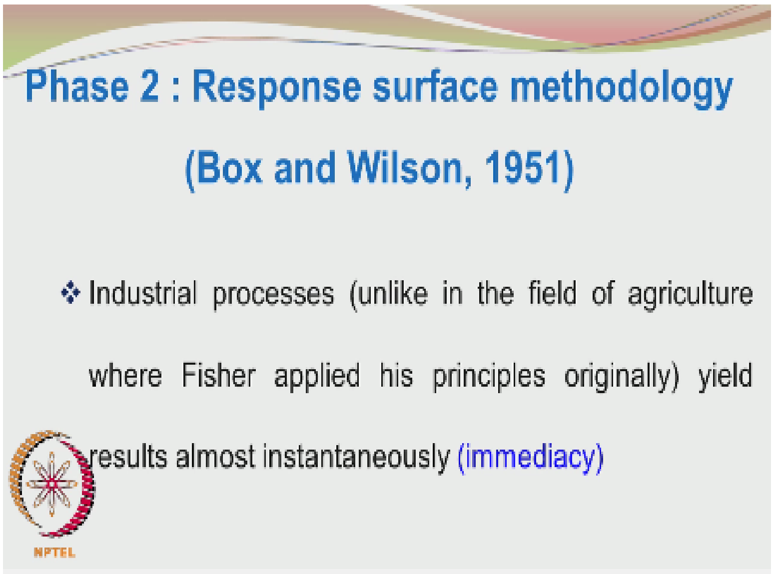


So, we will be looking at the history of statistical design, there was a phase in 1930, where Sir Fisher introduced some very interesting concepts. He was working in a Research Institute involving agriculture and he came up with some novel ideas.

(Refer Slide Time: 02:07)

He brought in the concepts of blocking, randomization and repetition, factorial design and analysis of variance, so these remain the backbone of statistical design of experiments, so we are really in depth to Fisher for bringing these concepts.

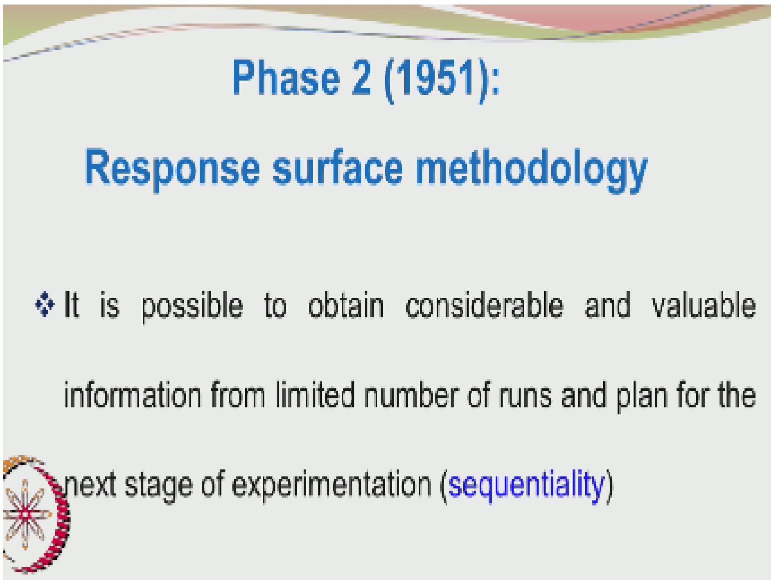
(Refer Slide Time: 02:33)



The next phase was the development of response surface methodology, we will be talking about this particular topic pretty soon and this was more suited for industries, where you do a set of experiments in a narrow range of operating conditions and you want to know in which direction you should proceed, in order to maximize the process yield or minimize the reaction time and so on.

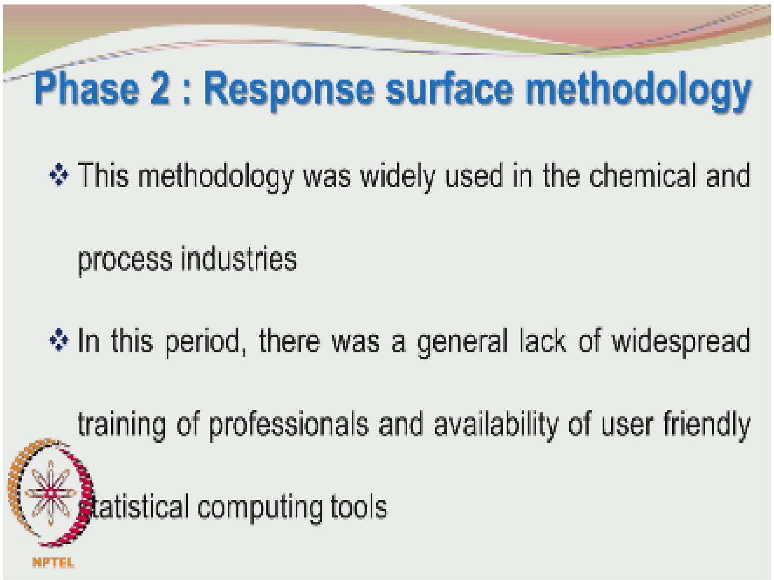
So, essentially this becomes an optimization exercise and you are looking at the particular search direction to find the optimum location, so this is a good combination of optimization and statistical design of experiments. Just imagine you are lost in the forest and you really do not know how to come out of it, what direction you should go. So, similarly the experimenter also faces this situation.

(Refer Slide Time: 04:21)



He is comfortable with the R and D results but where to move further especially, when there are large number of variables influencing the process, the response of his methodology is a very effective tool and unlike the field of Agriculture, the results in the industry are pretty much made available quite soon. So, we bring in the concept of immediacy and you also have to plan for the next stage of experimentation.

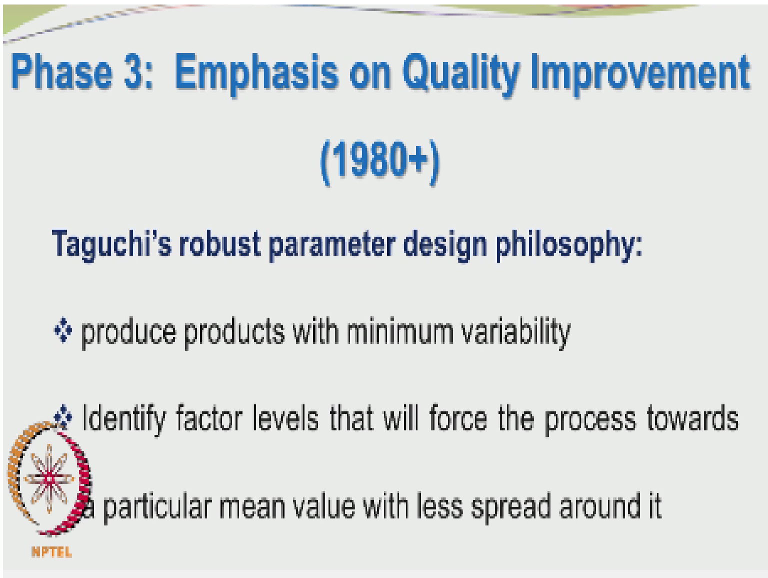
(Refer Slide Time: 04:33)



So, this brings in the concept of sequentiality. The response surface methodology was widely adopted in the chemical and process industries and during this period, there was a general lack of awareness of this kind of methodology. People were comfortable doing experiments in a very systematic manner, they did not really worry about the large number of experiments or even if they worried about it, they really did not know how to reduce the number of experiments.

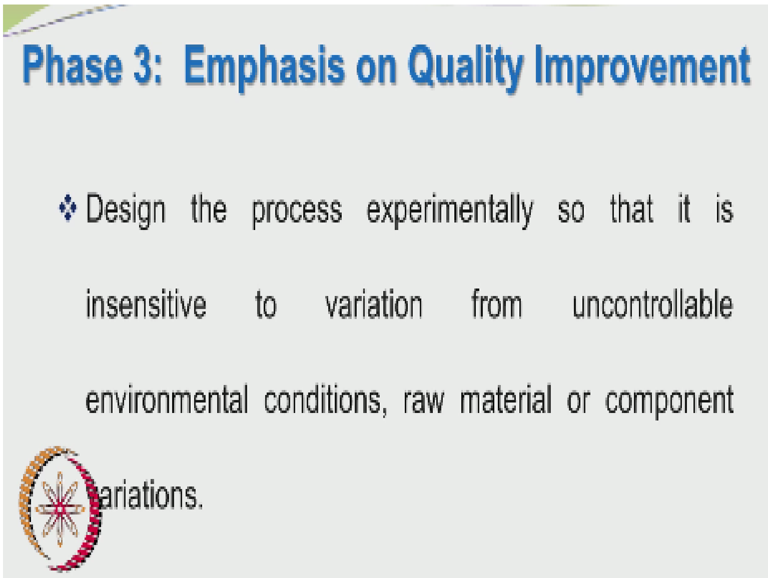
And they felt that by reducing the number of experiments, they would miss on valuable information and also there was a lack of user friendly statistical computing tools. Well, this is not really a drawback in my opinion because many of the statistical applications for real life problems involve reasonable mathematics, I would not say simple mathematics or complicated mathematics, reasonable mathematics I think, which should be done.

(Refer Slide Time: 06:01)



But I think to popularize this technique on a large scale especially to industrial people, there was a pressing need for software tools, where people could plug in the data and get the results. Then came in 1980s, the philosophy of Taguchi, this created quite a stir not only among the industrial people but also among the academicians, there was a big debate on the Taguchi principles.

(Refer Slide Time: 06:51)



And the basic idea in Taguchi's method is to have a robust experimental design, which will produce products with minimum variability and the technique involves the identification of levels of factors or settings of factors that will force the process towards a particular mean value with minimum spread around this particular value and also the experiment or rather the process has to be designed such that it becomes insensitive to variation from uncontrollable environmental conditions, raw material or component variations.

(Refer Slide Time: 07:24)

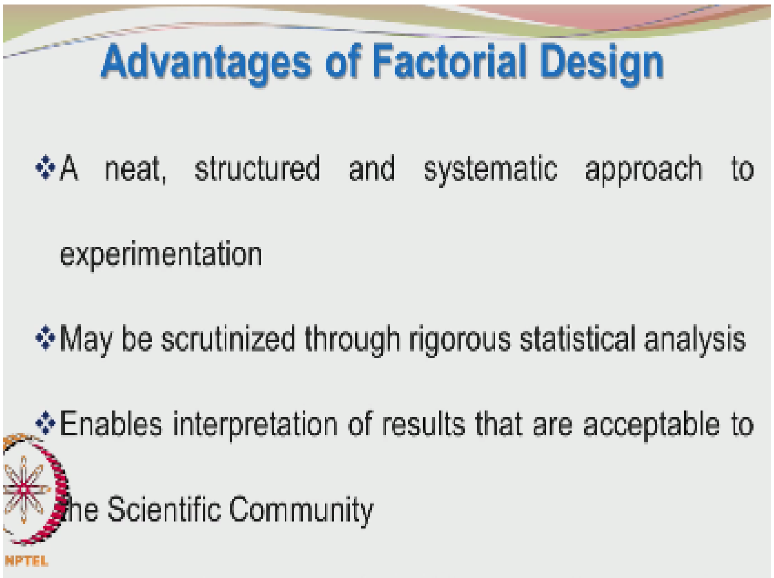
So, this became also popular and took the statistical analysis to the discrete parts industry like semiconductors, automotive, electronics, aerospace, manufacturing etc. and there was also a considerable research or flurry of research activity to find alternatives to the Taguchi method and quite importantly, the concept of statistical design was also implemented in the academic institutions.

(Refer Slide Time: 08:18)

So, this gives a hint that sufficient theory had developed by let us say, around 1970s; late 1970s for example, you saw the book by Box and Hunter, it was in 1978, so by 1970s and 1980s, it became a full-fledged academic course with sufficient clearing and Montgomery in his design of experiments book proclaims that the experimental design has to be integrated in engineering and science to attain industrial competitiveness.

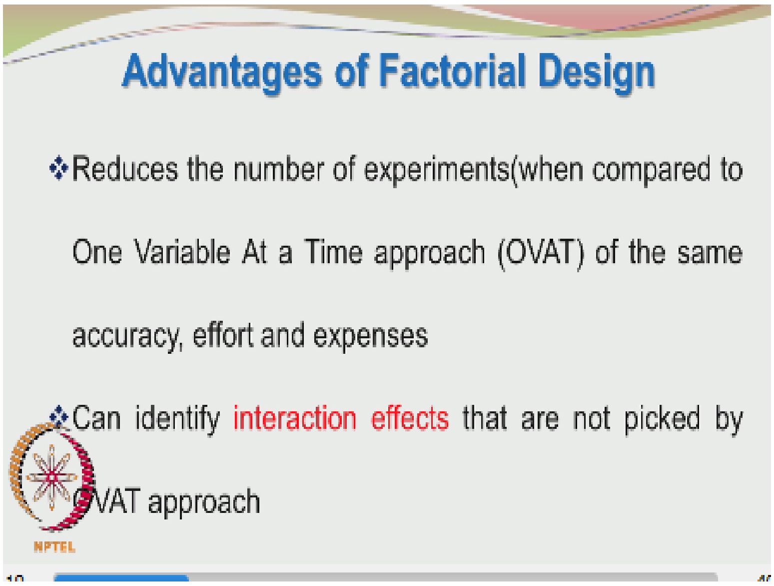
So, there must be a lot of industries, which can benefit by the application of design of experiments, also if students are trained in this area, they can contribute in an industrial environment to minimize or towards minimization of cost, time, manpower requirements and so on. So, look at this course rather than a theoretical subject as one which should be implemented for saving of time, money, manpower, energy and so on.

(Refer Slide Time: 09:17)



So, let us come down to the factorial design, what are the advantages? It is a neat structure and systematic approach to experimentation, it gives you a set of rules or procedures, which are simple to understand and easy to implement okay. There is nothing complicated, if it was complicated nobody would follow it and it can be subject to rigorous statistical analysis and the results interpreted through these statistical means are accepted by the scientific community.

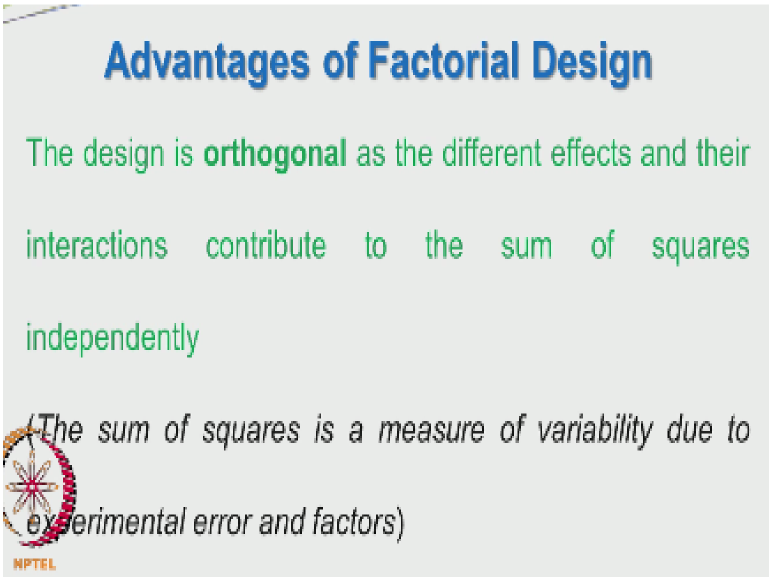
(Refer Slide Time: 10:10)



Or even if there where ambiguity or subjectivity that is quantified in terms of the p value which we saw yesterday. So, people were more comfortable doing one variable at a time experimentation, they will keep all other variables at their fixed values and then vary only one variable. After varying and completing the experiments with this particular variable, then they will keep this variable constant take up another variable vary that while all other remaining variables are kept at their fixed values.

So, this is a one variable at a time approach, it looks very logical but when you look at the factorial design of experiments, it is much more compact involves lesser number of runs for obtaining the same level of accuracy and another important thing, which will be encountering very frequently is the identification of the interaction effects through the factorial design of experiments.

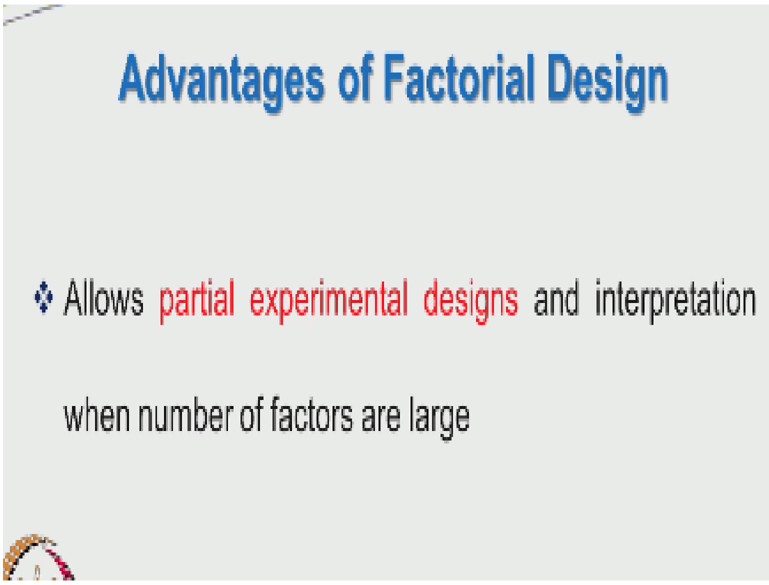
(Refer Slide Time: 11:22)



What is interaction effect, how does it influence the process, we will soon see. Another beautiful or elegant advantage of the factorial design is its orthogonal property or characteristics. I like this very much and what does orthogonality and how does it make the design elegant and simple, we will soon see. First my objective would be to give the introduction to the factorial design of experiments.

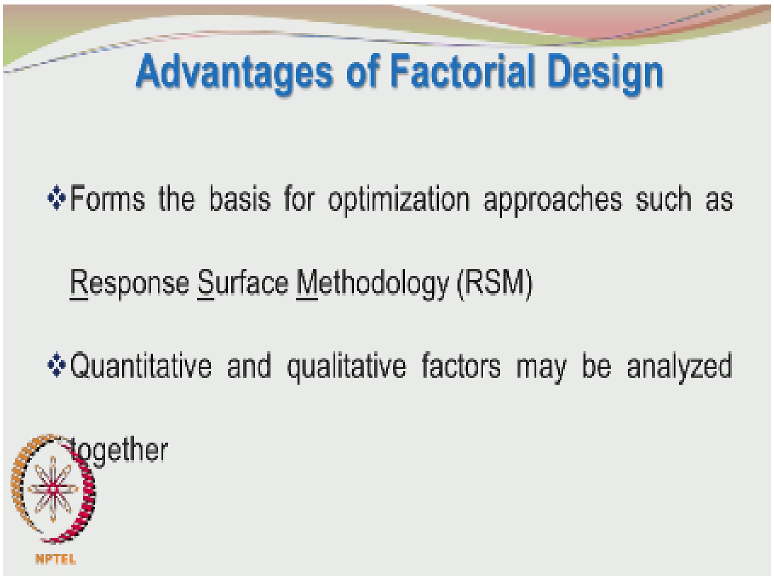
And then, we will look at the orthogonal designed experiments. Basic idea in the orthogonal designs is the different effects are sort of designed in such a way that their contribution of sum of squares to the overall sum of squares become independent of each other okay, this may be a bit difficult to understand at this moment in time but just let us keep it in the back of our minds and come back to it in a short while from now.

(Refer Slide Time: 12:25)



Even when you have a factorial design, which minimizes the number of experiments, sometimes you may have a large number of factors to study and even a factorial design would make things cumbersome. So, what you can do is; divide your factorial design into several fractions and carry out the fractions in a sequential manner, even fractions give you valuable information, so you do not have to complete all the fractions of the entire set.

(Refer Slide Time: 13:34)



But you may stop at any point you want, once you feel that you have got sufficient information. So, it is very flexible, so it does not force you in any way, it makes life simple when you have less number of factors and when you have large number of factors, it also helps you to proceed until you are satisfied. The factorial design of experiments, let me put to you bluntly is not implemented in its original form most of the time.

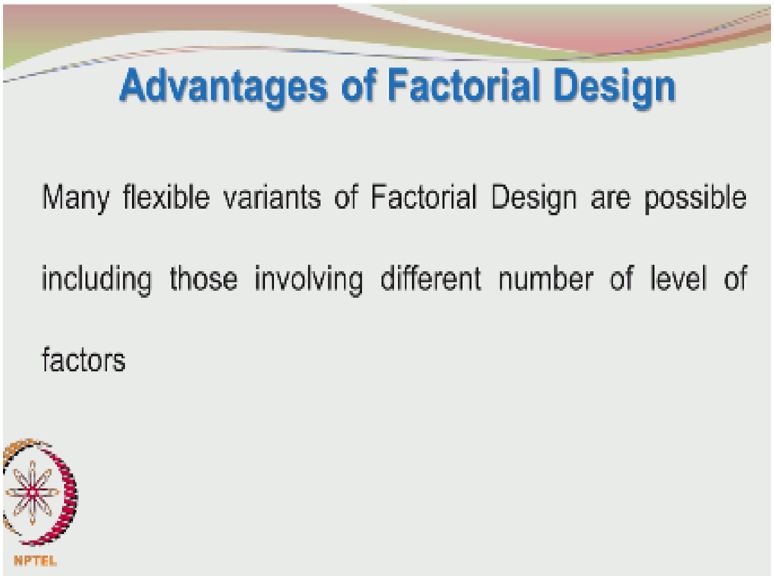
Because even factorial design of experiments may involve large number of runs and especially, when you have let us say, 5 variables that would be something like 32 runs and always you know that you have to carry out repeats to get an idea about the experimental error and in such situations, even 32 * 2 or 32 * 3 leads to 64 or 96 runs, which are probably too many.

And so it is important that we sort of streamline the entire methodology and designs are available to reduce the number of runs but they are all built around the factorial design concept okay, so it is very important for us to understand the factorial design and once we become familiar, understanding other designs become quite simple. So, it forms the basis for the optimization procedures like the response surface methodology, which I talked while earlier, the optimization of the process conditions.

Factorial designs play a very important role here and another beauty about factorial designs is it can handle both qualitative and quantitative variables together. What I mean is if you are carrying out experiments in a chemical reactor and you are looking at variables like field, flow rate, temperature, pressure, these are quantitative variables, you may also be looking at catalyst A and catalyst B.

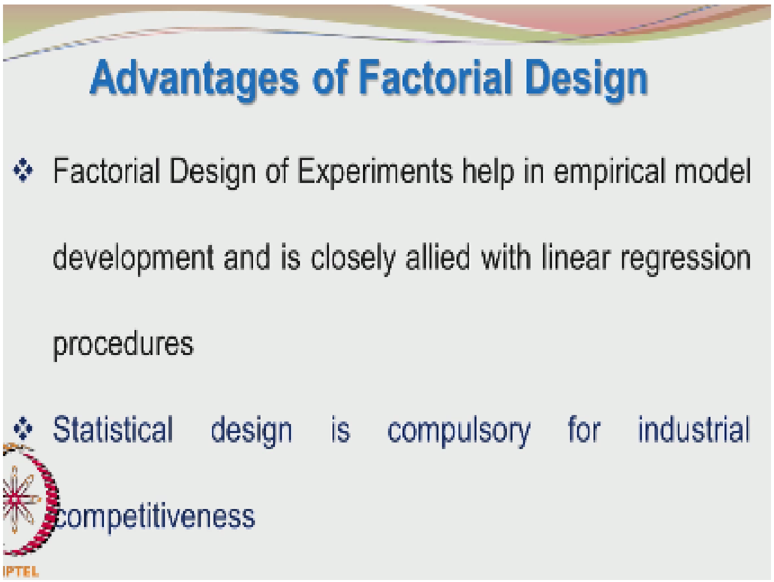
Unless you quantify the catalyst in a detailed fashion and incorporated into some mathematical model, we assume that these are qualitative variables, we want to see whether catalyst A is more effective than catalyst B in affecting the yield from the reaction or well, you can even do experiments with machine 1 and machine 2, so these 2 machines would then represent 2 levels of the qualitative factor machine type, okay.

(Refer Slide Time: 16:50)



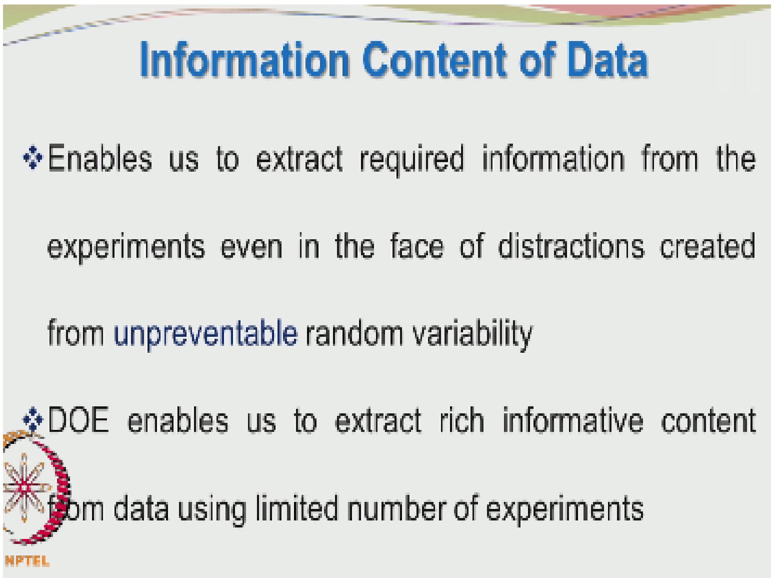
So, these are some examples of qualitative variables, you are cleaning clothes in 2 different washing machines, so they become qualitative variables. So, the advantage of factorial design is it helps you to consider even qualitative variables in addition to the usual quantitative variables. So, the factorial design is a very flexible one and many variants of the factorial design are available.

(Refer Slide Time: 17:10)



We will be looking at a few popular amongst them and it would be a simple matter for you to look up at the remaining ones and understand them. What do we get out of a factorial design? We get an empirical model, which gives the process response in terms of the different factors, their interactions tell us which factors are important, which interactions are important, again I am using the word interaction without really formally defining it, please be a bit patient I will come to it shortly.

(Refer Slide Time: 17:46)



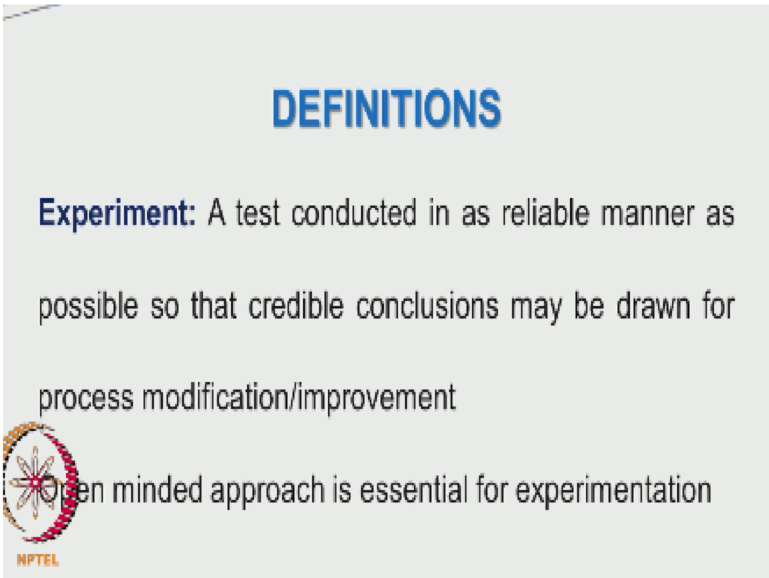
And as was cited from Montgomery's book, statistical design is compulsory for industrial competitiveness, this is a very important thing, so what I would like to tell the students especially, research scholars and students who are working in the laboratories, various undergraduate laboratories that it is okay to have scatter in your experimental data, you may not get exact reproducibility of your experimental runs.

Because there may be several factors, which may be affecting your runs, you do your best through blocking randomization to try to get rid of any systematic variability and even if you have variability beyond this, let us so be it, okay and you use proper statistical tools like the design of experiments to find out whether the factors in your experimental work or comparable with the experimental noise, in which case the factors are considered to be ineffective.

Or the factors are contributing way over the experimental noise effects, it sorts of helps us to segregate the contributions from different factors, even if the 2 factors are acting in a combined the manner that effect is also isolated and importantly all these effects are isolated from the experimental noise or the random errors and then it helps us to compare the effect of changing the different variables with the variation from experimental error.

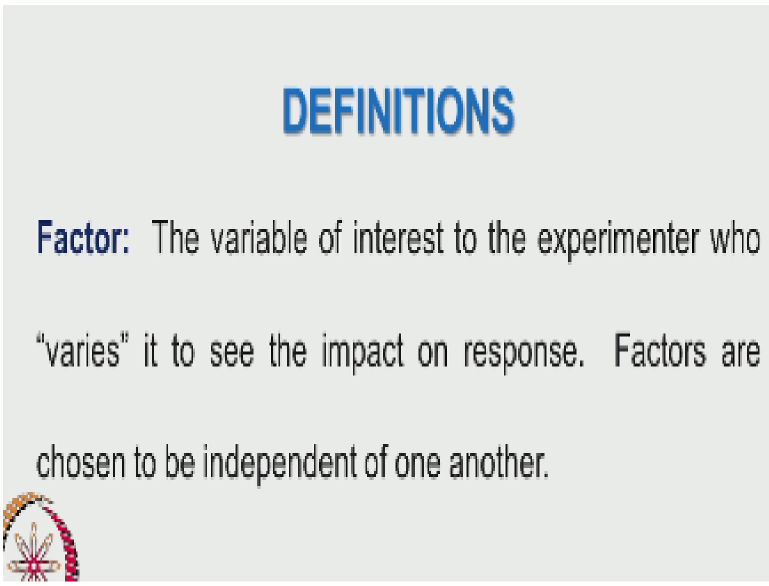
And helps us to make the necessary conclusions, again it involves hypothesis testing and we have to oscillate the null hypothesis and alternate hypothesis and based on the experimental evidence, we have to may come to a decision. So, it helps us to tap the rich informative content from the experiments using limited number of experiments that is a beauty of statistical design of experiments.

(Refer Slide Time: 19:48)



So, a few definitions are in order at present test conducted in as reliable manner as possible, so that credible conclusions may be drawn for process modification or improvement experiment, this is one such definition. Most important thing is an experimentalist or an experimenter should approach the experimental work with an open mind, he should not have preconceived notions that the experiment is going to behave in a certain way.

(Refer Slide Time: 20:38)



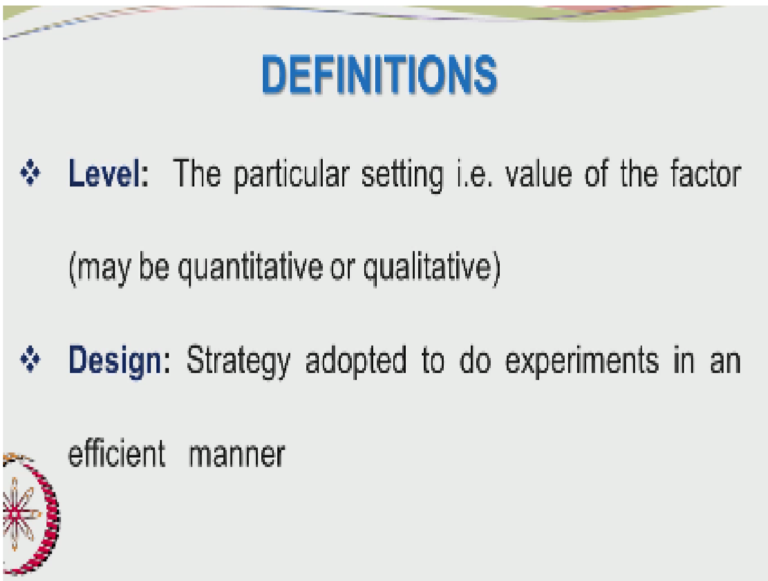
If that is there, then he will not have that curiosity or that open mindedness to accept results as they come and try to find reasons for that. Doing experiments also is an admission of the fact that our theoretical framework for analysing complex processes is not fully developed or it is highly complicated, they may not yield correct exact solutions or solutions in closed form, analytical expressions may not be obtainable.

The numerical simulations may be very expensive, so it helps us to fill in the gap also experimental work is essential to validate the high and numerical computations okay, so that the credibility may be built on those simulations, the model equations used and the assumptions made, they have to be validated with experimental data. So, inevitably wherever you go, you cannot avoid doing experiments.

Doing simulation work with computers is not the end of it, it is essential to do modelling of varying levels of detail but eventually, you have to back up the simulation results through proper experimentation. So, what is a factor? Factor is the variable of interest, which is being controlled by the experimenter and he can set it at different levels for example, if the experimenter has looking at a washing machines performance, he can alter the speed of the washing machine drum.

And so that would be a factor and he can set different speed levels for this particular factor. The factors are chosen to be independent of each other in other words, if I change one factor setting, if another factor is automatically affected by it, then the 2 factors are not independent. For example, if I increase the power to a machine and the machine speed increases as a result of it, then I cannot say the machine power and the machine speed are independent factors.

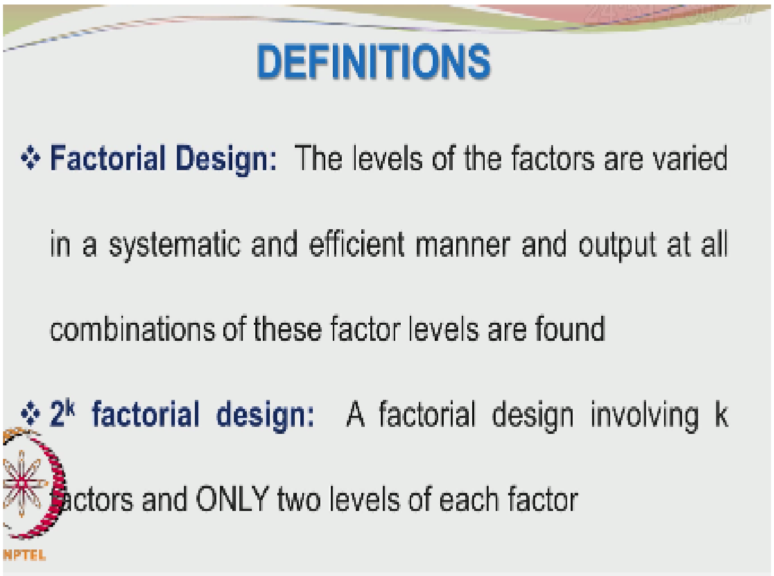
(Refer Slide Time: 23:19)



So, we have to look at factors, which may be varied independent of each other. Level; the particular setting or value of the factor, it may be quantitative or qualitative; quantitative means temperature is 30 degree centigrade, 50 degree centigrade and so on. Qualitative means machine A, machine B or catalyst A, catalyst B and so on. Design; what is the design? It is a strategy adopted to do experiments in an efficient manner, okay.

So, design sort of gives you the matrix of experimental conditions or settings, where the experiments have to be carried out, it is a kind of a blueprint for doing your experimental work and choosing an appropriate design is important, it is not as simple as it looks. Many times I have seen students, who are very enthusiastic about the statistical design of experiments often are at a loss as to what is the appropriate design for their experimental work okay.

(Refer Slide Time: 24:41)

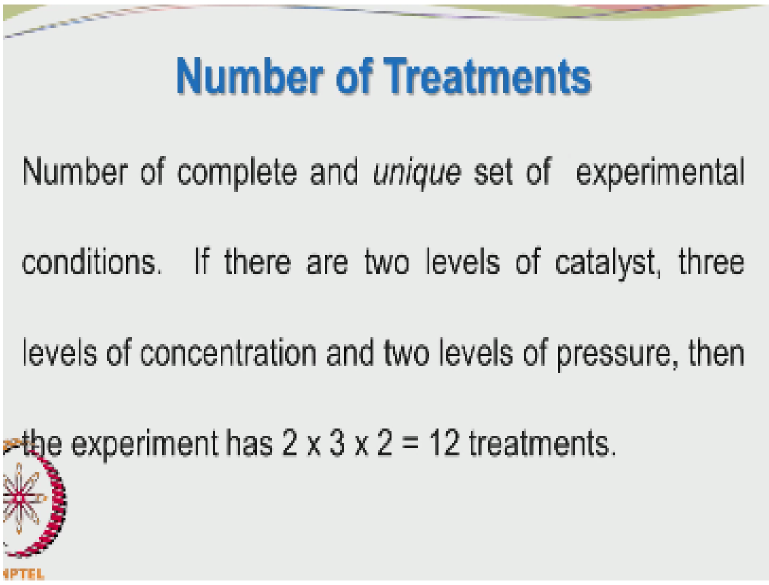


So, this is a very important feature and we have to look at the factors sorry; the reasons, which go into the choice of an appropriate strategy or an appropriate design. So, factorial design, the levels of the factors are varied in a systematic and efficient manner and output that all combinations of these factors are found. Now, we come to the important 2 power K factorial design, a factorial design involving K factors only and only 2 levels of each factor.

So, we represent a factorial design usually in the form of 2 power K, sometimes students may get confused, they think that they are only analysing 2 variables, this is not correct, you can analyse any number of variables you want, you are in fact, analysing K variables in the 2 power K design okay and what is the 2 doing there? The 2 represents the number of levels for each factor.

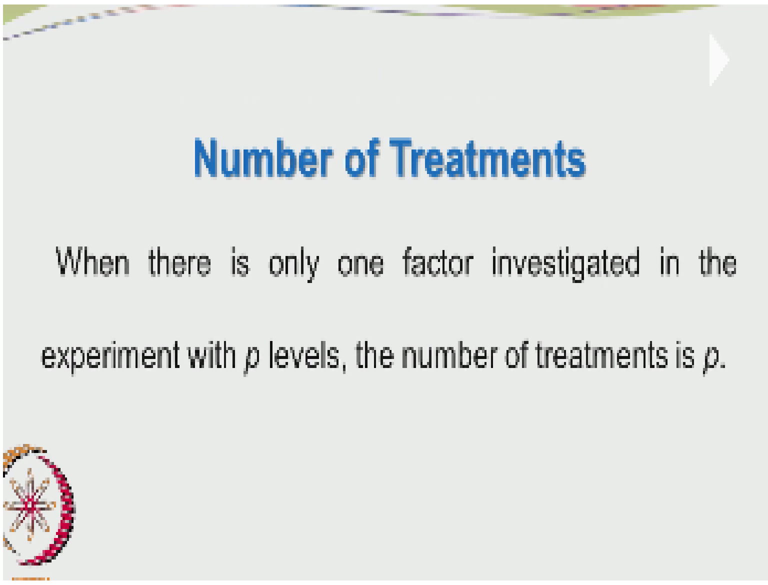
So, if you are having 2 that means each factor will have only 2 levels; a low level and a high level okay, if you are having a 3 power K design, again you can have any number of factors, you can call them K number of factors and the 3 represents the levels of each factor, each factor will have a low setting, medium setting and a high setting, so the 2 represents the number of levels in the factorial design.

(Refer Slide Time: 26:17)



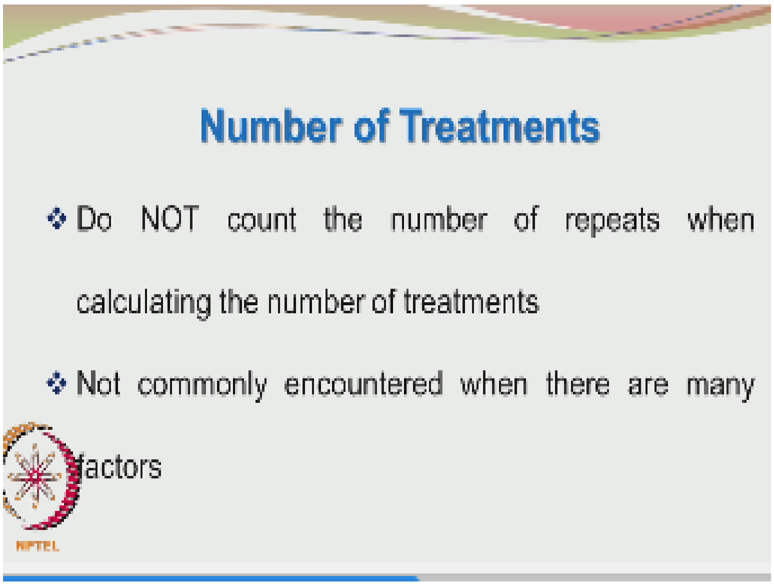
Here, it is referring to 2 levels of the factors. So, number of treatment, this is not a very common, we came across it quite frequently in the single variable experimentation. Now, we are looking at multi variable experimentation, so we can have any number of factors or variables and the number of treatments refers to the complete and unique set of experimental conditions.

(Refer Slide Time: 26:51)



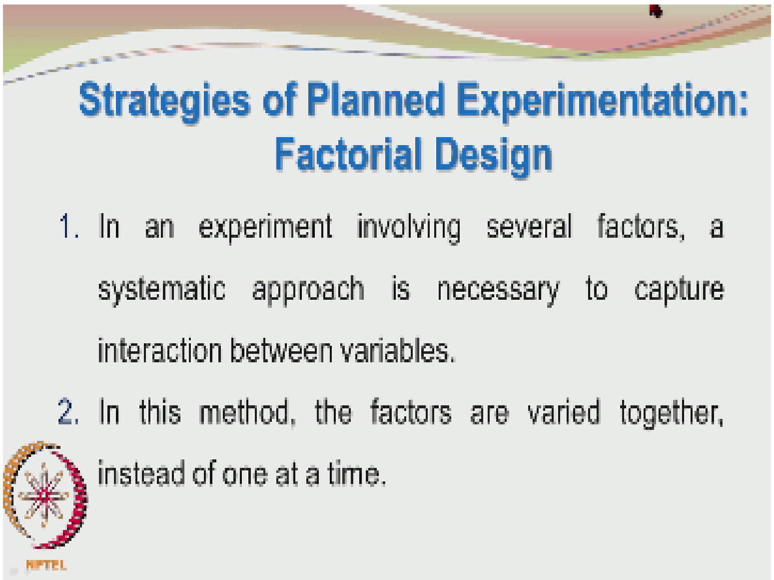
If there are 2 levels of catalyst, 3 levels of concentration and 2 levels of pressure, then the experiment has 2 * 3 * 2, which is 12 treatments. When there is only one factory investigated in the experiment with P levels, the number of treatments is P that we have seen.

(Refer Slide Time: 26:59)



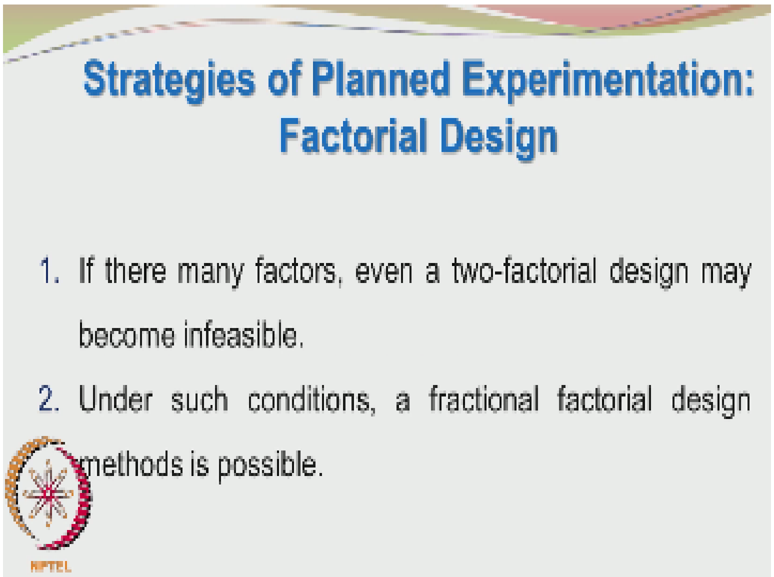
So, when you want to find the number of treatments, do not count the number of repeats when calculating the number of treatments and it is not commonly used when there are several factors.

(Refer Slide Time: 27:09)



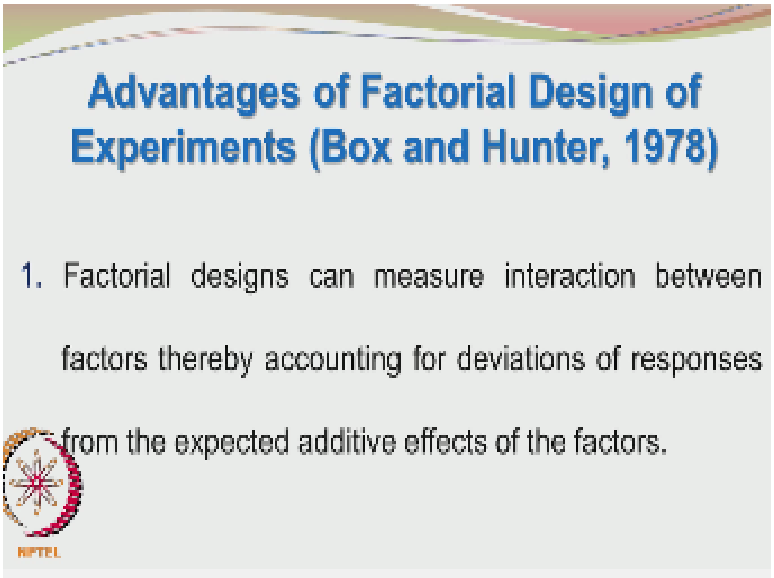
So, in factorial design what we are doing is; we are not varying the variables one at a time, sometimes the variables may be changed together okay, so in doing so, you are reducing the number of experiments without losing on the information aspect.

(Refer Slide Time: 27:36)



So, as I said earlier, if there are several factors, then even a two factorial design may become infeasible, so a fractional factorial design is required.

(Refer Slide Time: 27:47)

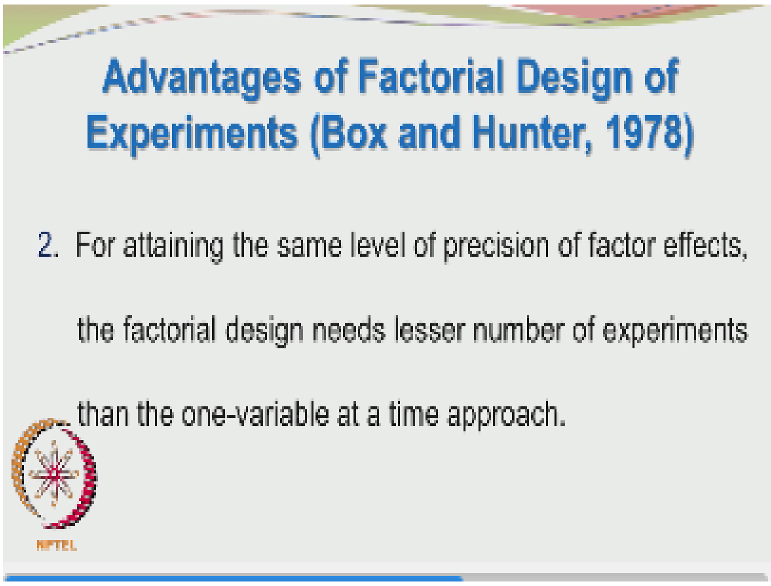


So, I think it is time for a small break now, please sort of think of those difficulties you were facing when doing experimental work and also reflect on how the stated advantages of the factorial design may help you in your experimental work, we will continue further with the discussion on the advantages of factorial design in a short while from now. Fine, so the factorial designs can measure interaction between factors.

Thereby accounting for deviation of responses from the expected additive effects of the factors, sometimes even though you are having independent factors, when they act together they may act either independently or in a combined fashion, it may be a synergetic fashion, if you want to put it like that. What is meant by a synergetic fashion? The net outcome of the process would be over and above the independent or additive action of these 2 factors.

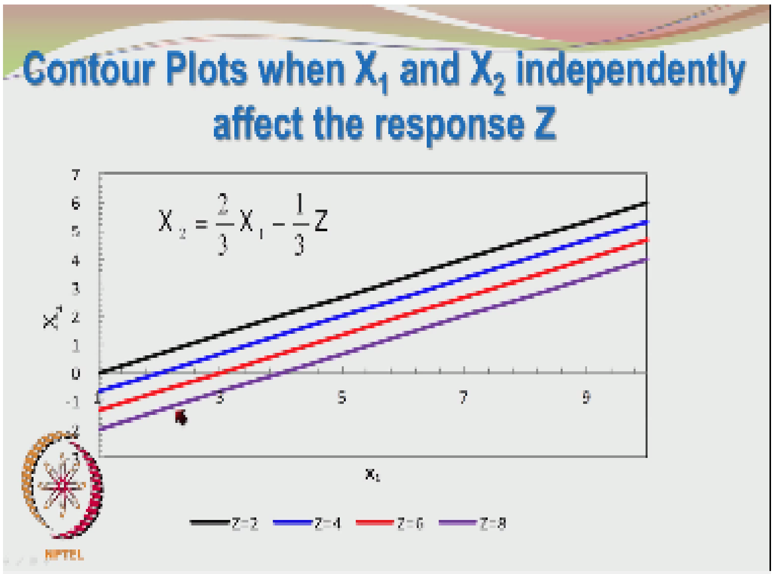
For example, you are having a factor A and factor B, which are completely independent in all respects and if you vary them in a factorial design, the response would be a constant + beta 1 XA + beta 2 XB, so this shows that over an average value, there is contribution from factory A and contribution from factor B but on top of it, if you also have a contribution like beta 0 + beta 1 XA + beta 2 XB + beta 1 2 XA XB.

(Refer Slide Time: 30:18)



That means, in addition to acting independently, the 2 factors are also combining into XA XB and influencing the process, so this is termed as interaction and for attaining the same level of precision of the effects of the factors, the factorial design requires less number of runs than the one variable at a time approach.

(Refer Slide Time: 30:32)



So, you can see that this is a simple figure, where the process output is z, so z is given by 2X1 - 3X2, okay, so I am plotting X2 and X1 and these are values of constant z, so z is = 2, z = 2, z = 4, z =6 and z = 8, so this is a kind of a contour plot, where you are shown the variation of both the variables on the X1, X2 axis and these lines represent values of constant response and these are linear.

So, if you look at this; is z is = 2X1 – 3X2, so z is affected twice by X1 and 3 times by X2 and X1 and X2 are acting in opposite senses, increase in X1, increases the z value but increase in X2 actually, is found to reduce the z value but the important thing is the 2 factors X1 and X2 are acting independent of each other and this is another example, where you have z is = 3X2 + 2X1.

So, you can see that again X1 and X2 are acting independent of each other, z is a constant and you can see that there is a negative slope, okay. So, when X1 increases X2 decreases but when X1 increases and X2 increases, you can see now that the z values increasing, so you can have different types of relationship between the factors and the process response and in these 2 examples, we have shown the 2 factors to be acting independent of each other.

(Refer Slide Time: 33:42)

Scores of a Batsman

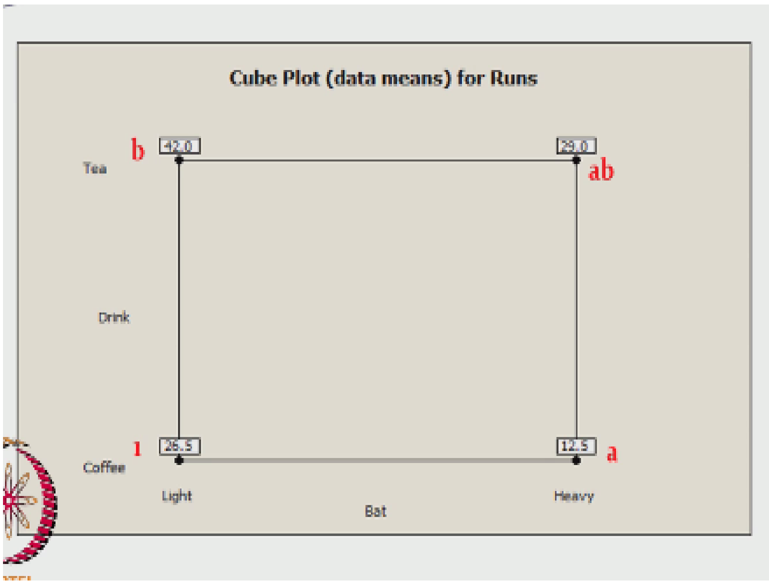
Innings	Bat	Drink	Runs
1	Light	Coffee	25
2	Light	Coffee	28
3	Light	Tea	40
4	Light	Tea	44
5	Heavy	Coffee	15
6	Heavy	Coffee	10
7	Heavy	Tea	30
8	Heavy	Tea	28

But in many processes, there may be considerable interaction involving the 2 factors, which also has to be accounted for. So, I will just give an example to demonstrate interaction between factors using the Minitab software. Let us look at the scores of a batsman in different innings okay, let us consider 8 innings, these may be for 4 Test matches or more depending on how many innings were completed in a given Test match.

So, we are looking at the batsman using a light bat and heavy bat, I do not know how many of you would recollect sometime back, maybe 10 years back, there was lot of discussion on a star batsman using a heavy bat and what was the impact on his shoulders and arms, anyway you can have a light bat and you can also have a heavy bat. Nowadays, heavy bats are becoming more popular and let us also look at the type of a drink, the batsman may have had before coming out to bat, its coffee or tea.

So, he is either using a light bat or a heavy bat or may have drunk coffee or tea before coming out to bat okay, it is a very fictitious example, just to drive home the concept of interaction. So, we want to see the influence of the bat and the effect of drink on his runs scoring well, the runs are not looking very impressive overall, the highest is only 44 well, it may be a T20 match also, so we cannot say anything.

(Refer Slide Time: 35:51)



But these are the runs scored anyway, so we have to see the influence of the type of bat he has used and the type of drink he had taken before coming out to play and see the impacts on the runs he is scoring. Unquantifiable or qualifiable factors like form, all these things may not be considered, they are random effects or uncontrollable factors. So, we will represent them in a cube plot, it is not a cube but it is a general term, it is a square plot.

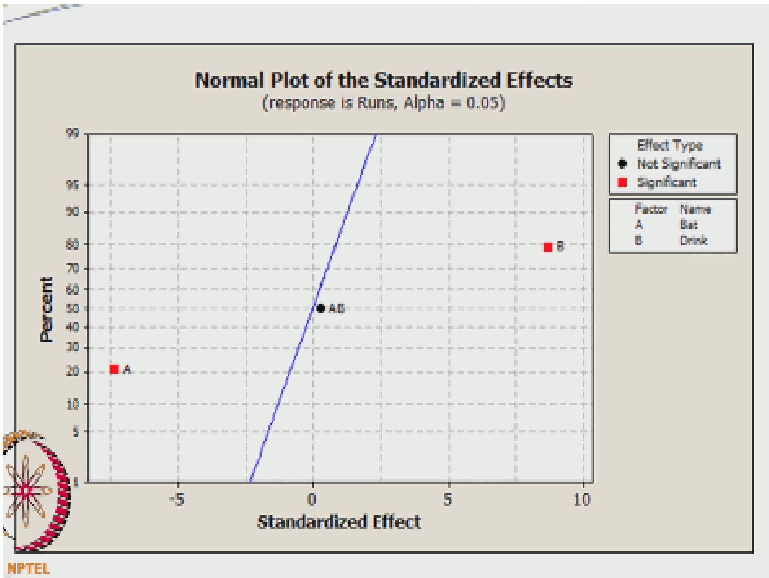
So, you are having bat on the x axis and drink on the y axis, you are having a light bat here and a heavy bat, the average runs scored with light bat and coffee is 26.5, light bat and coffee, so 25 + 28 is 53; 53/2 is 26.5 and that is what we have as the average here. So, we are saying that light bat is a lower setting, coffee is a lower setting. It is completely arbitrary you can set light bat as a higher setting and coffee as a higher setting.

It is completely arbitrary for convenience, we are just putting light bat as low level and coffee as low level, tea as high level and heavy bat as high level okay, so when you have 2 factors and these 2 factors are at their low levels, we represent it by 1, then here the drink is still coffee but the bat is heavy, so factor A, which is the weight of the bat or type of the bat is now at a higher level.

So, factor A is at a higher level, factor B is at a lower level and so we call it as small a, on the other hand if you go like this, you are now moving in from a lower level of drink to a higher level of drink, from coffee you are moving on to tea but the bat is still at a lower level, a light bat okay. Lower level and higher level I am just using it without implying that lower level is bad and the higher level is good, I am not saying that.

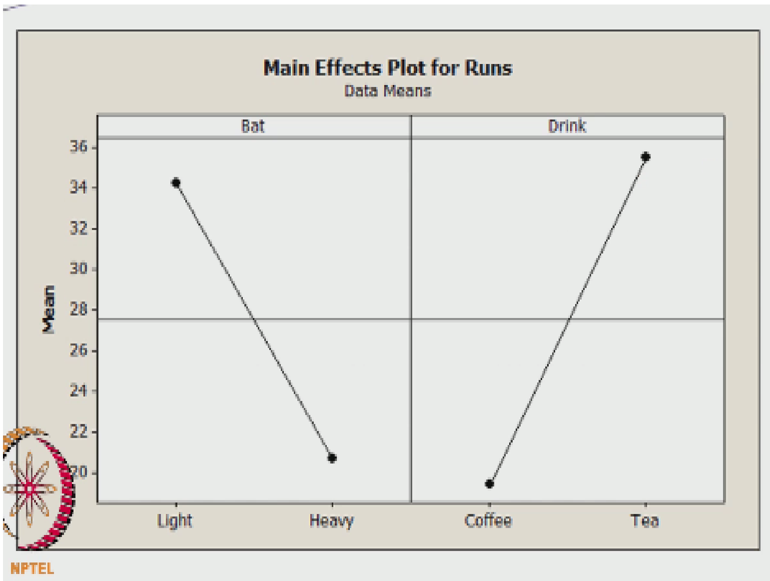
It is just 2 settings of the experimental variables, so you are having coffee at lower level and tea at higher level but the bat is still lower level of light bat, so you are having the label as small b. In this case, you would easily understand that both the settings are at their higher levels, bat is at the heavy setting and drink is at a tea setting and so you are having a * b, so this is the general nomenclature.

(Refer Slide Time: 38:52)



We will be following in design of experiments and calculations of the different effects. So, we plot a normal probability graph and we can see that A and B are very far from the normal probability line saying that these 2 are having an effect on the runs scored, they cannot be at dismissed as random effects. AB is the interaction between the bat and the type of drink and it can be seen that it is not having an effect.

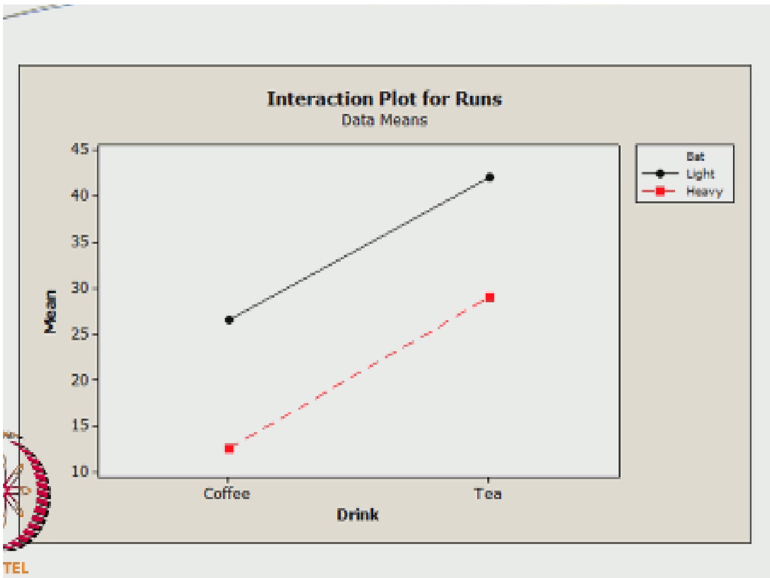
(Refer Slide Time: 39:54)



You can also see that the type of bat is having a negative influence, whereas the type of drink is having a positive influence for example, if I am going in from a light bat to a heavy bat, the performance is getting lower. If I am going from coffee to tea, the performance is getting better, so you can see a lot of information given in one simple diagram. So, this is the effect of the main effects.

Or this is the diagram, which shows the main effects, it shows that, when I am going from a light bat to a heavy bat, the mean runs scored is reducing. If I am going from coffee drink to tea drink, the average or means runs scored is increasing, so it looks like heavy bat is not good and coffee is not good. If you want to score more runs, you have to go for a light bat and drink tea before coming out to play.

(Refer Slide Time: 40:36)



If I am also looking at the interaction plots, they are parallel what it means is; if I am going from coffee to tea using a heavy bat, my performance increases by a certain extent. The performance increases by the same extent, when I am using a light bat also of course, when I am using a light bat my performance, is higher but with the light bat, if I am changing my drink from coffee to tea, my increase in runs scored is equal or comparable to the runs scored when I am going from coffee to tea with the heavy bat.

So, the weight of the bat or whether I am using a light bat or a heavy bat, it really does not affect the performance enhancement by changing the drink okay. So, the same change is observed, when I am changing the variable from one level to another level at different settings of the other variable okay that means the first variable or first factor is acting independent of the other factor.

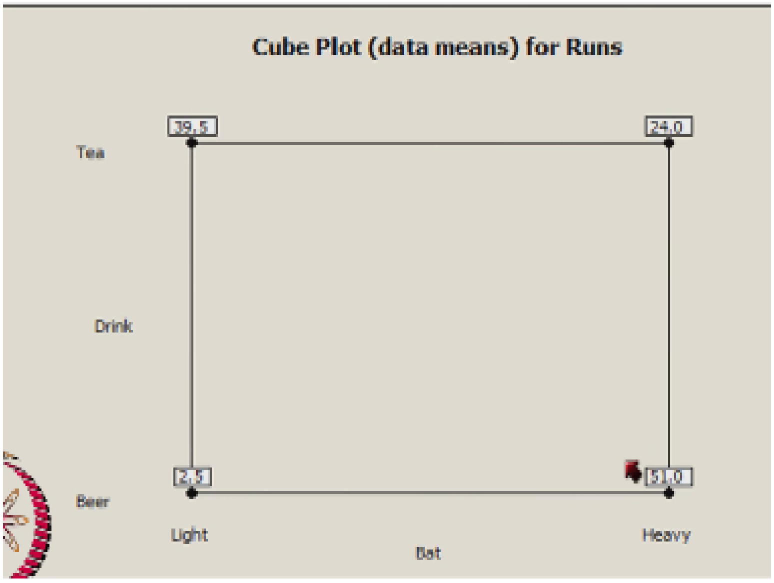
(Refer Slide Time: 42:09)

Innings	Bat	Drink	Runs
1	Light	Beer	5
2	Light	Beer	0
3	Light	Tea	40
4	Light	Tea	39
5	Heavy	Beer	50
6	Heavy	Beer	52
7	Heavy	Tea	25
8	Heavy	Tea	23

I request you to think about this a bit. Now, we want to demonstrate interaction. Let us now substitute beer for coffee, again this is a completely fictitious example, I do not say that batsmen have beer before coming out to play, just to drive home the point of interaction and coming up with this example. So, you are having light bat again as usual and then heavy bat, then instead of coffee, the batsman may either drink tea or may drink beer before coming out to play.

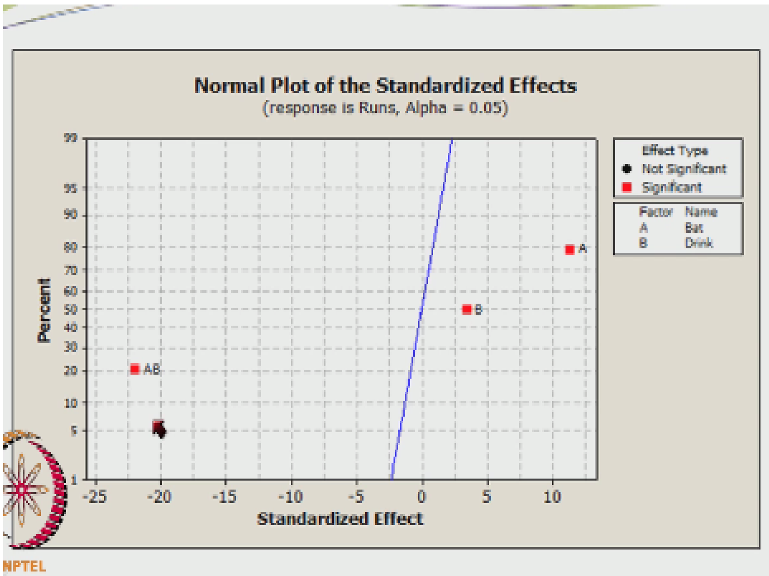
I do not think he will drink both and come out to play either drink beer or drink tea and come out to play and a couple of things are noticeable in the previous table of runs scored, the runs were pretty close to one another not too high not too low but if you look at this table of runs scored, you can see some very bad performances and you can see some very good performances, so even he has scored a half century a couple of times.

(Refer Slide Time: 43:28)



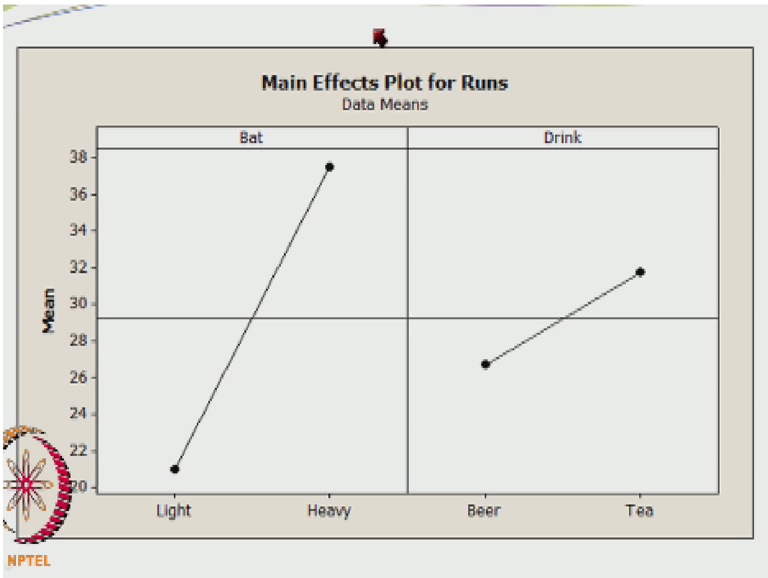
So, let us analyse the effect of the light or heavy bat or beer or tea on the batsman's performance. So, again we are putting this in a plot, where for each setting, the average value is reported. For beer and light bat, the average runs code is only 2.5, so 5 + 0 is 5; 5/2 is 2.5, so the average runs scored is only 2.5. Similarly, with a heavy bat and beer, the average performance is 51.

(Refer Slide Time: 44:07)



From beer and tea 39.5, heavy bat and tea, it is 24, so what do these numbers really mean? So, now when we plot the normal chart or graph, we can see something very interesting. Earlier this AB factor was close to this line and we dismissed it as random effect, now you can see that A and B are lying on the same side of the graph, they are having a positive influence on the batting or run scored.

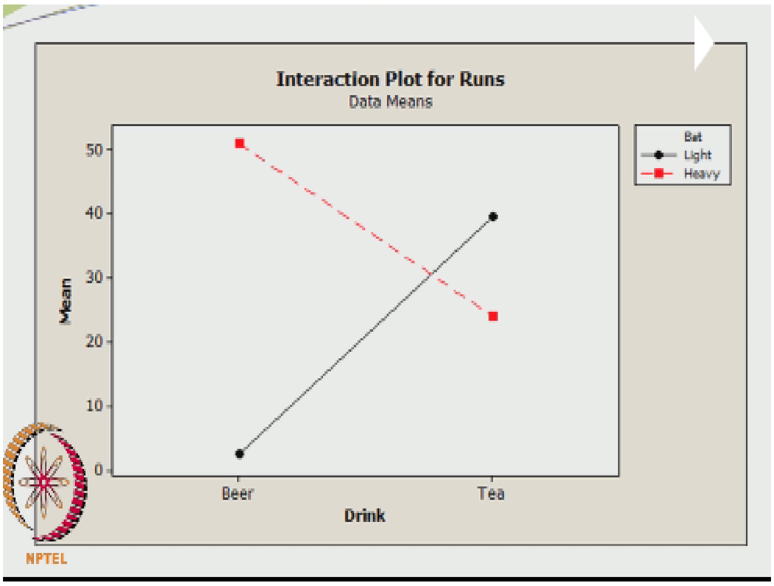
(Refer Slide Time: 44:54)



Whereas, this AB combination is having a negative influence on the runs scored, very interesting, now AB is also significant. A is the type of bat used and B is the drink taken by the batsman before coming out to play, AB is the interaction between the 2. So, this main effects plot tells that heavy bat is good, if I am going from a light bat to a heavy bat, I am able to score more runs.

And if I am going from beer to tea, again I am scoring more runs, so it looks like the batsman should go in for a heavy bat and drink tea rather than beer before going out to play but these main effect plots are misleading, when considerable interaction effects are present. So, let us look at the interaction effects before discussing or comparing between main effects and interaction effects.

(Refer Slide Time: 45:35)



This graph is very striking and completely different from the earlier interaction graph we saw previously, the earlier interaction graph was parallel, when you went from one drink to another drink for a given bat, the batsman's performance increased by a certain extent and the batsman's performance increased by a same or similar extent when he changed the drink at a different setting of the bat.

Earlier, he was using heavy bat and even after changing to light bat, the performance up gradation did not change, when he changes the drink but if you look at this, if the batsman takes beer and uses a light bat, his performance is very poor. When he goes for tea and still with the light bat, his performance increases dramatically, the batsman using a heavy bat and tea, the performance is bad okay.

Heavy bat and tea is still better than beer and light bat okay, so tea has helped him somewhat but this is very interesting when he goes for a heavy bat and has taken beer, the performance improves like anything okay, one may imagine that the batsman may be in extremely good spirits with the beer and also he may be swinging the bat merrily even though, it was heavy connecting and making a lot of runs okay.

So, it may be seen that if a batsman is using a light bat, his performance changes by a certain extent, when he goes from beer to tea but when he is using the heavy bat, the performance changes in a completely different manner. So, the effect of one factor on the runs scored depends on the setting of the other factor, when the interactions were not present, the effect of one factor on the response was independent of the setting of the other factor.

Now, when we have interactions, this is very, very important, when we have interactions, the effect of one factor on the response will depend upon the settings of the other factor or other factors. So, depending upon the level of the second factor, the first factor shows the effect on the process, so this clearly shows that there is interaction between these factors. A simple way to detect interaction is to see the interaction plot given by the statistical software.

If the relationships are parallel, then there is no interaction or very little interaction and if the plots are approaching each other or they have different slopes or they are even intersecting and crossing each other, it shows that interaction is present and it has to be accounted for in the model, simple one variable at a time model will not be able to detect interaction and you have to go in for a statistical design of experiments.

So, you can see that things became a bit complicated, I would not say complicated, a bit more tricky even with 2 variables; the interaction between 2 variables can slightly complicate matters. Now, imagine the situation, when you are having many variables some of them interacting and some of them are not interacting and then what will you do? You will have to resort to factorial design of experiments not only to identify the main effects.

But also to identify the interactions, some authors even go as far as to say that main effects are useless, you go and look at the interaction effects first. So, we will see after a small break. Thanks for your attention.