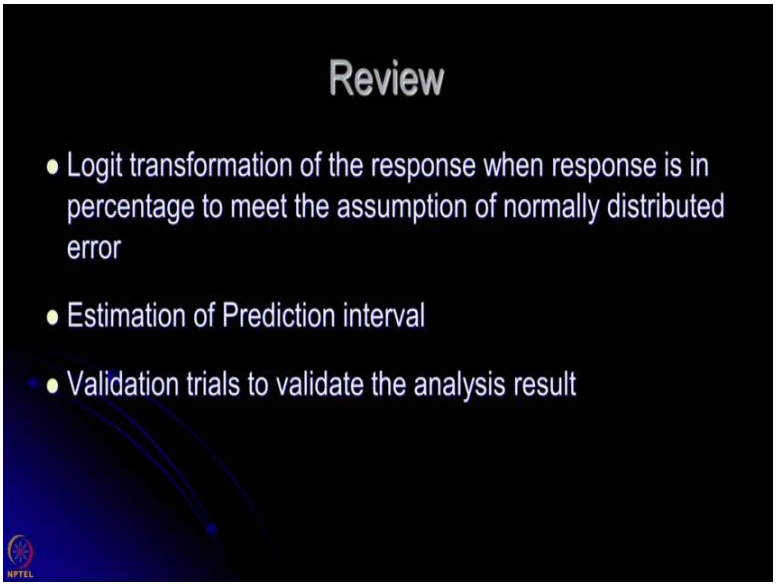
Dealing with Materials Data: Collection, Analysis and Interpretation
Professor. M. P. Gururajan
Professor Hina. A. Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 90
Design of Experiment IV

Hello and welcome to the course on Dealing with Materials Data. For past few sessions, we are going through the application of design of experiments for metallurgical applications and we are learning it through the case study optimization of production of Nano titania powder through micro plasma synthesis system.

(Refer Slide Time: 0:51)



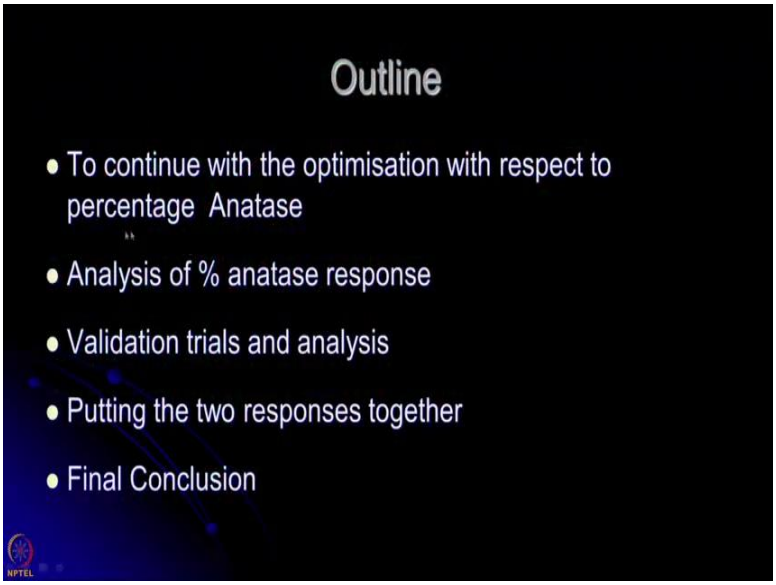
So, let us review, we found that when the response is in a percentage has a limited range in when it varies from 0 to 100 only, to meet the assumption of normality of error, we have to make a logit transformation of the response variable and carry out the analysis through with the logit transformed response. Then we talked about how to estimate the prediction interval, we of course worked out how to find the factors that are important by conducting a hypothesis testing that regression parameter beta i is equal to 0.

Then, looking at the T values and the P values we can decide whether we have to reject the hypothesis or accept the hypothesis. Here, rejected hypothesis is of importance because that says that, that factor is important in this predicting the efficiency of the system. So, now you see the importance of rejection region because rejecting the hypothesis gives us lot of

information compared to accepting the accepting the hypothesis says that, that particular factor has no effect, but when you reject the null hypothesis, you come to know that that particular factor has an effect on the response value.

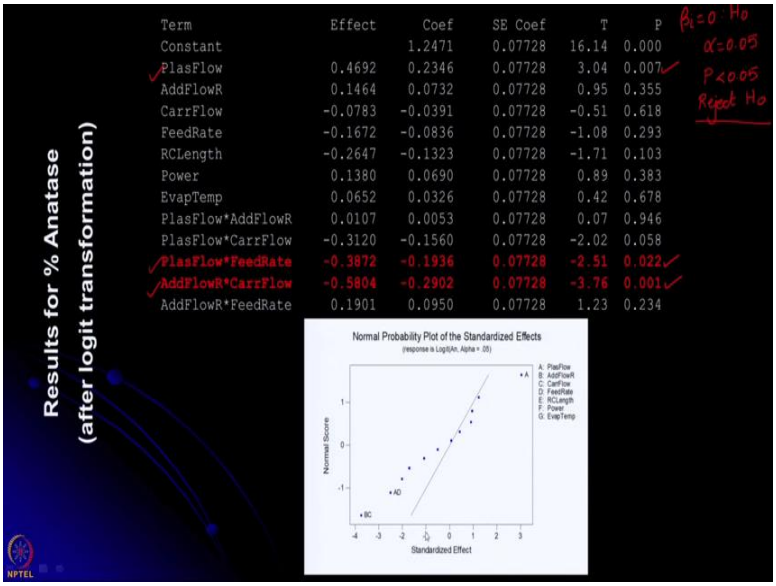
So, we look into the tables and the we also looked in the analysis of variance table to understand lack of fit, because of the 2 replica we could find out, if your the experiment, the model has any lack of fit, therefore again hypothesis was there is no lack of fit. And we, in this case, we accepted the hypothesis and therefore, we can say that there is no lack of fit in the hypothesis.

(Refer Slide Time: 3:12)



You remember there were two responses, so this time we want to continue the same process with a response of percentage anatase and remember now we are wiser people, so since anatase is also in a percentage value, we are going to take its logit transformation and then do the analysis and then we have the validation trials putting the two responses together and it will come to the final conclusion for this particular case study.

(Refer Slide Time: 3:40)



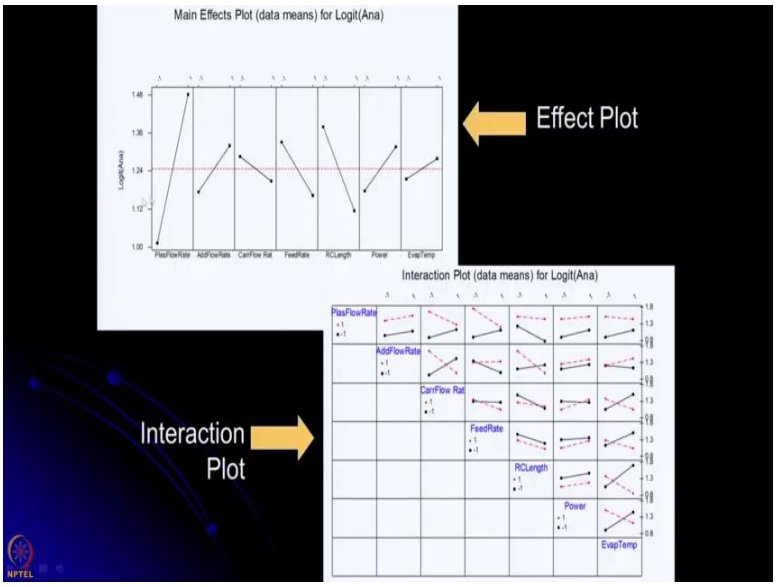
So, let us start, here this is the table for estimating the effect of each parameter using t-distribution for logit transformed anatase, percentage anatase. So, we are wiser now, we have already done the logit transformation of percentage anatase response and then we are doing the we are doing the analysis of a data. And here, what we find is that again we are giving the, this is an important, I am sorry, I should have made it in red, but it is not in red. So, here once again, we are, our hypothesis is that beta is equal to 0, this is our null hypothesis and alpha value is 0.05.

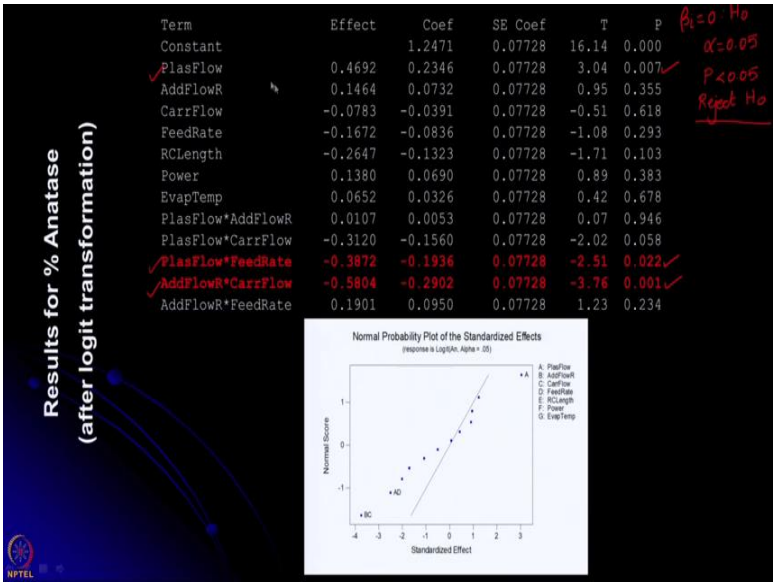
So, if P is less than 0.05, we say that reject null hypothesis. It means that for these values that is for plasma gas flow rate and the two interaction of plasma gas flow rate with feed rate and additional gas flow rate with carrier gas flow rate are the important or are the places where our H0 is rejected. It means that, the coefficients, the beta coefficients are not 0. In all other places we can say that beta coefficients are 0, it has no effect on the logit transformed percentage anatase and therefore, it reduces the model by having only these one main effect, one main factor and the two interaction factors.

Here, we have once again shown it by the graphical method that if there is a systematic error because of the regression equation and then there is a random error due to epsilon. This line shows if the model was only due to the random error, all data point should fall on this line, as further they go you realize that these are the points of importance.

Almost all points are far away here, but we have, from this table we find that when you consider alpha is equal to 0.05 that is your level of significance as 0.05 percent, sorry, 0.05 or 5 percent then these are the three values which are of importance.

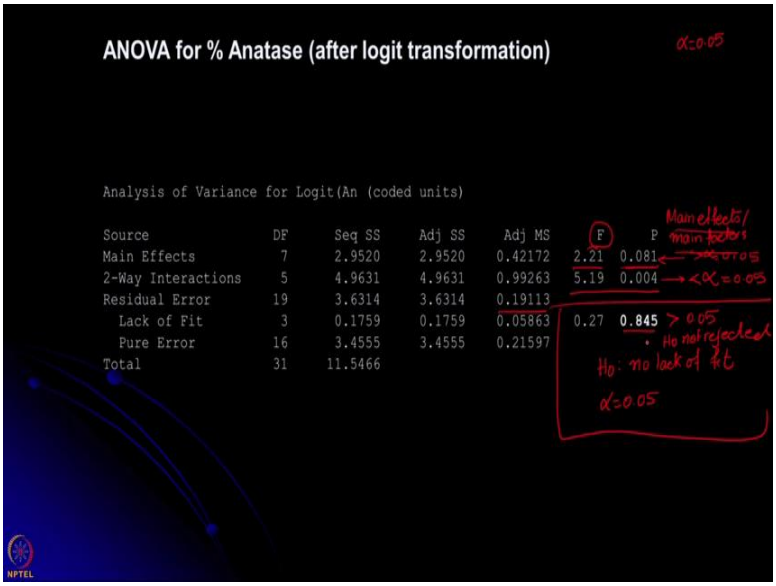
(Refer Slide Time: 6:35)





Here it shows the main effect plots and the interaction plot. You can see that there are many interaction showing up, but we must remember that finally it is this process of rejecting the null hypothesis by comparing the P value with pre decided alpha is most important. This is the type 1 error, please remember P is a P is probability of type 1 error. Just recall your hypothesis testing, recall your hypothesis testing and regression analysis. This is an amalgamation of all of them, so that is why we are going through a case study where you can actually see how it gets applied, each and every aspects of it gets applied.

(Refer Slide Time: 7:37)



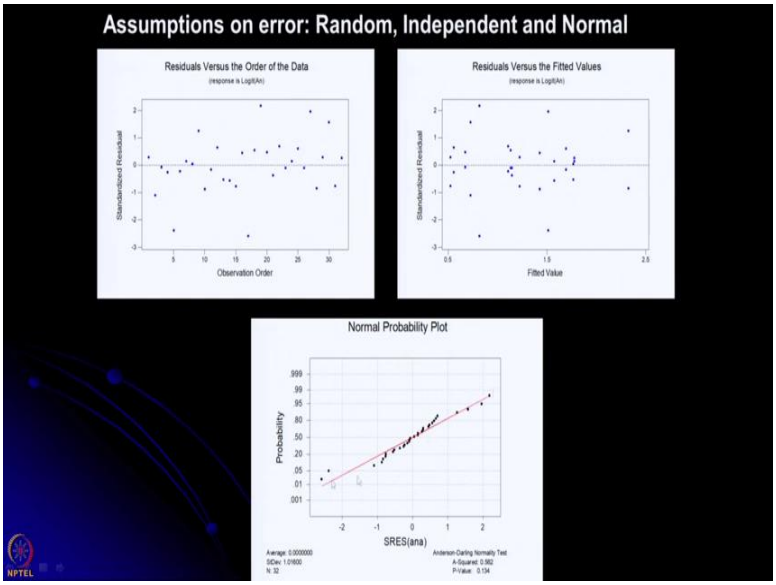
So, these are the plots. This is the analysis of variance table after logit transformation and as you can see here also it shows that there is the main effects are not important, this is not important. You see again here, alpha is equal to 0.05 and our testing of hypothesis is that main effects, the variance caused by the main effect and the variance caused by the error if this ratio is F which is large enough then you say that it has an effect.

The main that is there is an effect of the main factors the main factors. So, here what we are saying is that, though there we found that plasma gas flow rate is important, here we find that main effects have are not significant. So, main effects are not significant, however the interactions are significant because this value is less than alpha which is 0.05, this is so, because it is greater than alpha which is equal to 0.05, so this shows that main effects, this implies that main effects or factors do not change your variation in the value significantly compared to the residual error.

Whatever error is coming, the sigma that is variation caused by the error and the variation caused by the main effect are very similar, while this says that the error caused by the two way interaction is more significant is larger than the effect caused by the error. So, here again we are taking an F ratio, this is an ANOVA table. Again, if you look at our hypothesis is here is that H0 there is no lack of fit, alpha is 0.05 and this is greater than 0.05 and therefore, we say that H0 not rejected. These two are two different things, so please remember and therefore, this says that there is no lack of feed.

So, the model that we have chosen completely defines, completely gives the explanation of logit transformed percentage anatase.

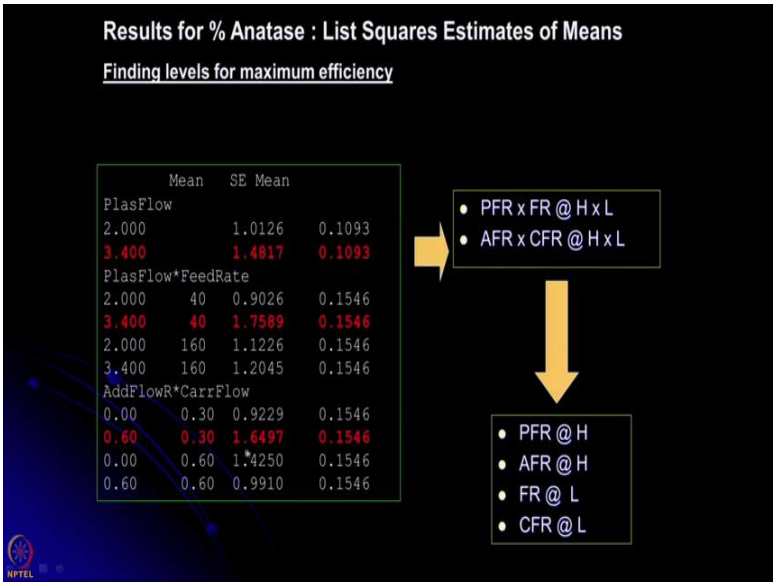
(Refer Slide Time: 10:57)



These are the confirmation to the assumption on error, so this the first plot shows that it is a random error, this also shows that it has no effect of the what fitted values it has taken, it is also random, this is random with respect to observational order, so there is no systematic changes because of the observations made. Here it says that there is no systematic change because of the values that we have fitted in the model.

And this shows that it is actually distributed as normal because we have plotted the standardised residuals, residual error for percentage anatase, of course logit transformed percentage anatase on the normal probability plot and it falls on the straight line and therefore, it is normally distributed. So, our assumptions are also correct.

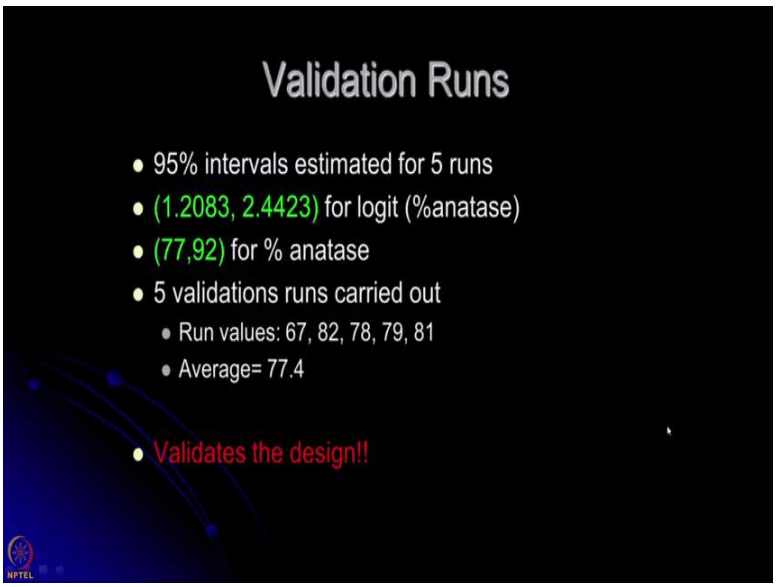
(Refer Slide Time: 12:04)



Then as we did in the past, we have to select the levels. So, now please remember that we have to actually select only the combination because this is not important, however it is coming up. It is plasma flow rate at a high level and feed rate at a low level will give you best percentage of anatase and the additional flow rate at a high level and carrier gas flow rate at a low level give you the best possible percentage anatase.

So, our selection is that the if you separate out the interaction level, it says that plasma flow rate, gas flow rate should be at high, additional gas flow rate should be also kept at high, feed rate should be kept at low and the carrier gas flow rate should be kept at low.

(Refer Slide Time: 12:56)

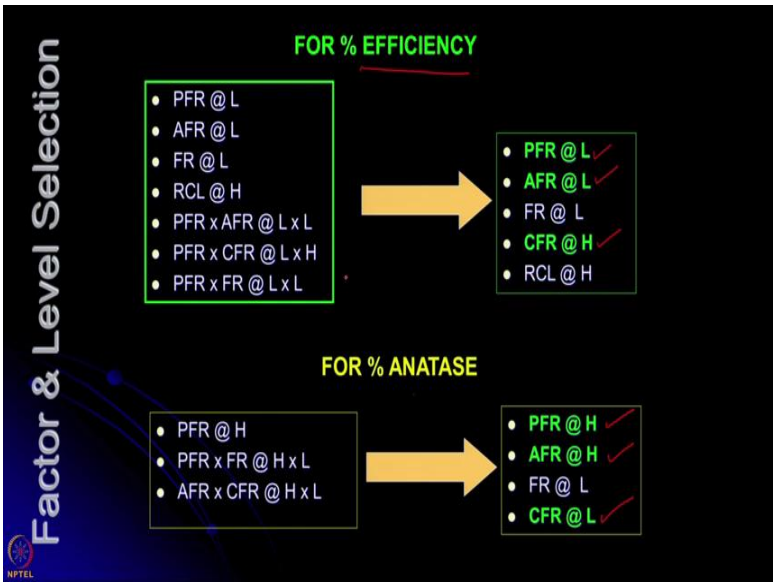


With this, we perform the validation runs, we find the interval for 5 validation run trials.

Please recall what we have done the calculations in the past, the same applies here and for the logit transformed variable, we find that this is the interval, the predicted interval, 95 percent it did interval estimation for 5 runs with this combination of the, the factors. Then if you revert back, it will take the inverse logit transformation, it says that it falls from 77 to 92 percent interval where the interval anatase should fall. The 5 validation trials are carried out and the average is 77.4 and therefore it falls here. And therefore, it validates the design. So, as such our job is done.

Again, we have to ask ourselves a question, is it everything okay? It should be okay, we did the, we found the best combination of factor and level to find highest efficiency and we found the best combination of factor and level to find the highest percentage anatase. What are they?

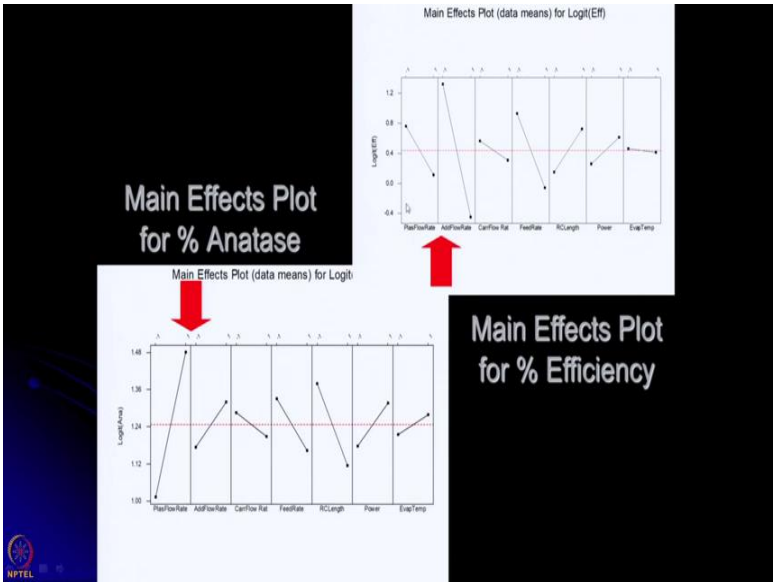
(Refer Slide Time: 14:26)



They are given here, but if you look at it, there is a problem, look at the problem. Here, in the first case for the percentage efficiency, we are saying that keep plasma flow gas flow rate at low while in percentage anatase, we say that you keep it at high.

Additional gas flow rate, here you say here that you keep it at low, here you say that you keep it at high. And carrier gas flow rate, it says, you keep it at high, here we say that, you keep it at low, so what does it mean? It says that either you have highest efficiency of the micro plasma synthesis system or you have the maximum percentage anatase, you cannot have both. This is what it really says.

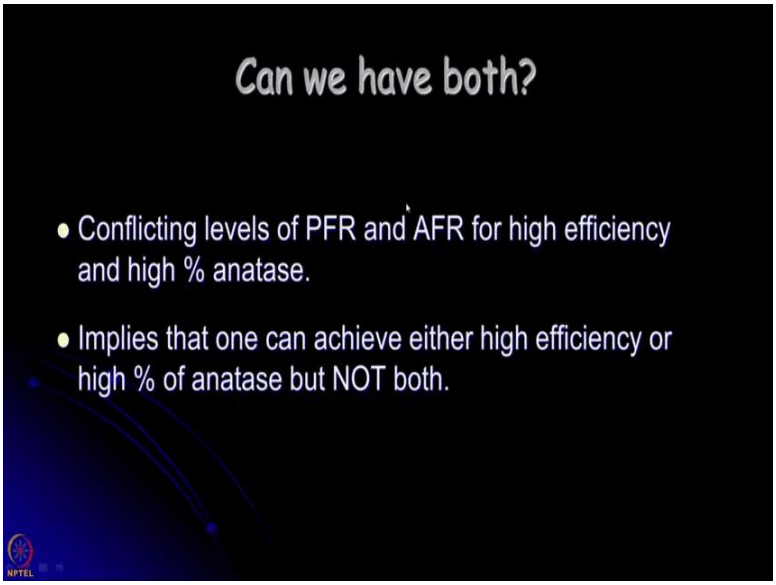
(Refer Slide Time: 15:24)



And that you can see here because here if you look at the main effect plot also, you see that the slopes are in different directions.

So, if you are choosing maximum on both sides, your level selections are exactly opposite.

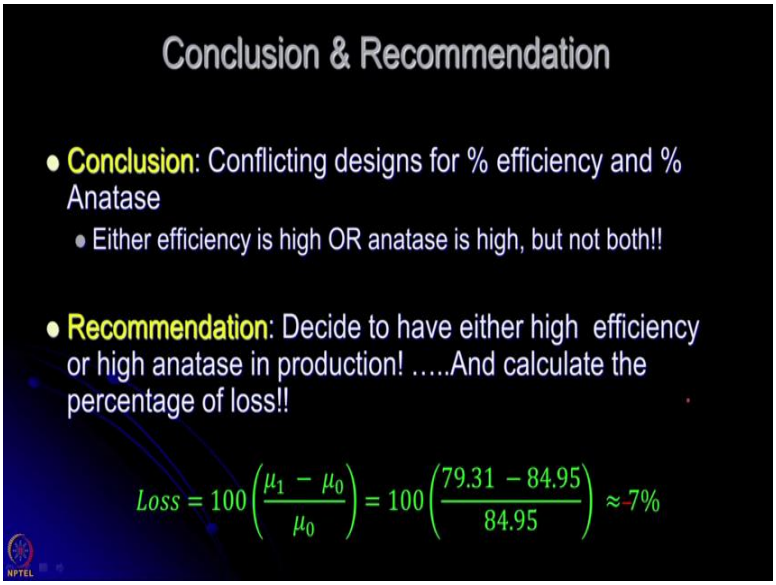
(Refer Slide Time: 15:39)



So, can we have both? We cannot conflicting level of plasma gas flow rate and additional gas flow rate, I am not even talking about carrier gas flow rate because in both the cases it has come only as an interaction.

So, if this is the case then you can have high efficiency but you cannot have a high percentage of anatase. So, you have to, you cannot achieve both. So, what to do in this case?

(Refer Slide Time: 16:12)



The conclusion as such is that conflicting designs for percentage, highest percentage efficiency and highest percentage anatase, so either efficiency is high or anatase is high, but not both. This is your conclusion, but in reality when you want to implement it, you have to give some solution, at least you must give some recommendation.

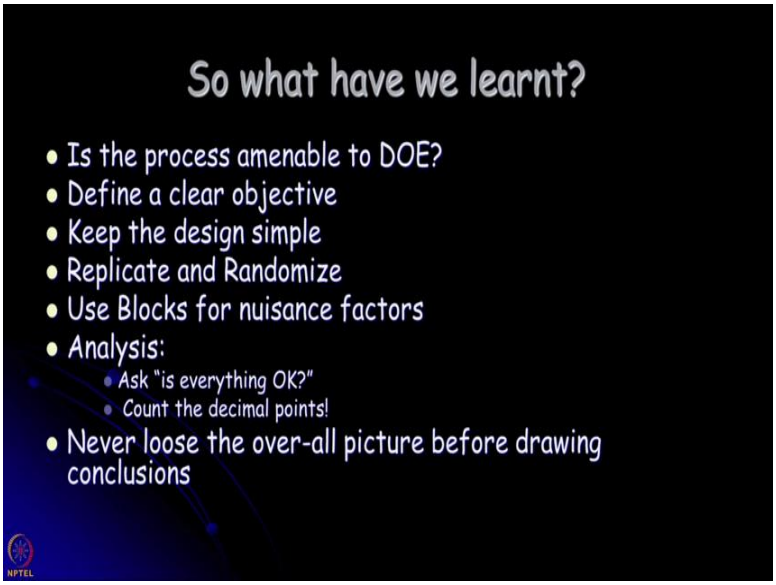
So, the recommendation would be let the experimenter decide. In your case, if you are the experimenter, you have to take a call whether you want a high efficiency or you want a high percentage anatase in production. And then decide what is your loss, what is your percentage loss. Generally, percentage loss is defined in this manner. Suppose, you decide that you want the high efficiency then you have to know what is your loss with respect to percentage anatase, had the percentage anatase been kept at the highest, your mean would have been mu 0.

So, had your percentage anatase would have been kept at highest, its mean would have been 84.95 and but now that you have decided that you want a high efficiency, you are losing your mean value from 84.95 percent to 79.31 percent. And this comes to, I am sorry, it is a negative 7 percent, so I must write it down, it is negative, so it is roughly negative 7 percent.

So, if you are able to lose the 7 percent of high percentage anatase in your powder then you will have the highest efficiency. This is what should be your recommendation, this is what should be the recommendation at the end of the complete analysis, design and analysis of

experiment. So, we come to this next stage, let us conclude the whole session of design of experiment case study.

(Refer Slide Time: 18:41)



We went through it in four sessions totally. First question is that, which process is amenable to design of experiment? We discussed that. Then we say that we must have a clear objective. We must keep the design simple, we while making the objective clear, we must know what is response variable and what are the controllable factors, what are uncontrollable factors of which what is an error and what is an uncontrollable nuisance factor.

We must use randomization, replication and blocking in order to control this. Randomization makes the experiment looks like a whatever is happening naturally in life. Replication makes sure that the error between the sets of experiment is nullified or it is averaged out. Then the blocking says that if you know that there are nuisance factors such as shift or a person working in the, a person defect affect affecting the process, those things you block it out and keep a systematic, count it also as an error fact, a controlled factor in it. Use blocks for the nuisance factor.

Carry out the analysis by using the analysis of regression model and the analysis of variance together. But every time keep asking yourself, is everything okay? Is everything found fine? Also, I would say that it is very common mistake to, when you do this analysis using any of this readymade software then you have to control your decimal points, you have to control your digit, significant digits. And finally, never lose the overall sight of the overall picture before drawing the final conclusion.

Like in this particular case, we could have drawn the conclusion in two cases, but then it does not really help the person who wants to really commercialize this method of microwave plasma synthesis of Titania Nano Titania, because you said that you will have maximum efficiency for certain levels and exactly opposite levels would give you the maximum percentage anatase, in that case, you have to ask him, “What is your preference”?

If it is maximum efficiency, then this is your loss towards percentage anatase. If it is maximum percentage anatase then this is your loss with respect to efficiency. This is what we discuss. I hope you have basically learned the application of hypothesis testing, regression analysis and analysis of variance during this sessions on design of experiment and I hope you enjoyed it. Thank you.