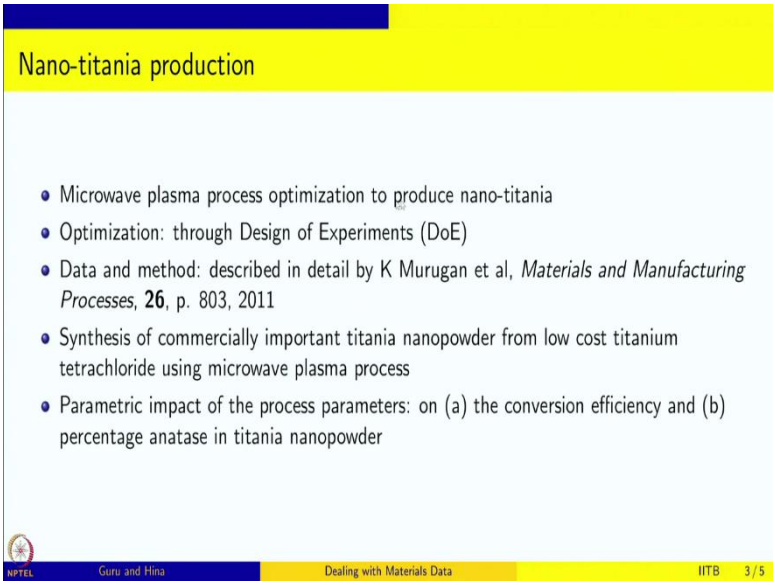
Dealing with Materials Data: Collection, Analysis and Interpretation

Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

Lecture No. 97
Case study 4: Design of experiment

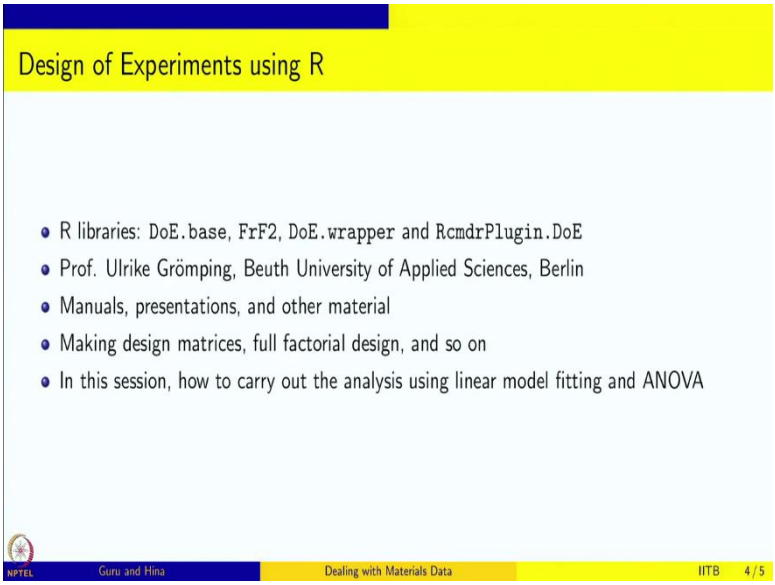
Welcome to Dealing with Materials Data. We are looking at the Collection, Analysis and Interpretation of Data from Material Science and Engineering.

(Refer Slide Time: 0:25)



We are in the module 6 which is on case studies and this is the fourth case study which is on Design of Experiments.

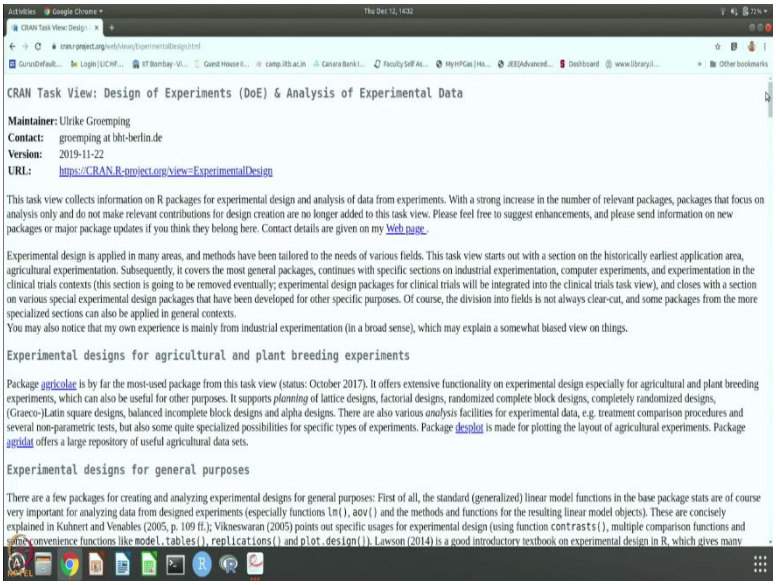
(Refer Slide Time: 0:36)

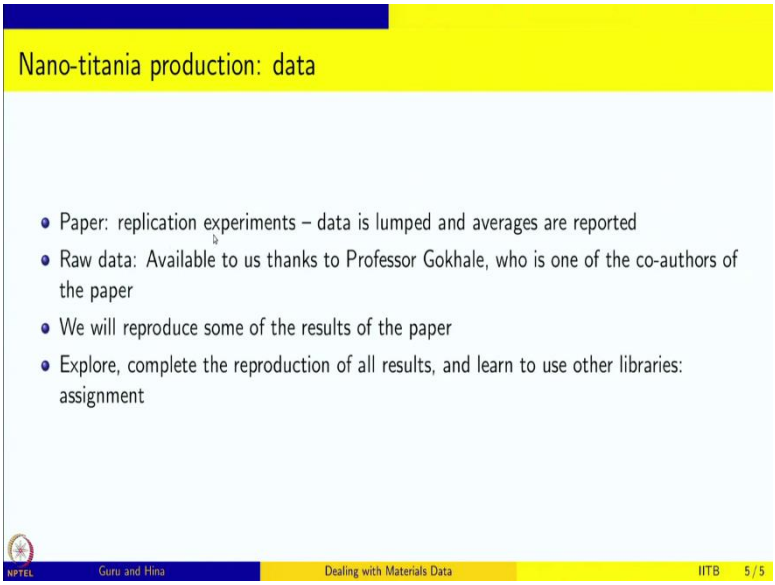


We are going to use this microwave plasma process optimisation to produce nano-titania as the example case to understand design of experiments. Here the optimisation is done through the design of experiments and the data and the method or describing detail in the paper by K Murugan et al, Materials and Manufacturing Processes, published in 2011.

This paper is about synthesis of commercially important titania nanopowder from low cast titanium tetrachloride using microwave plasma process and what they studied and tried to optimise is the parametric impact on process parameter of the process parameters on conversion efficiency and percentage anatase in titania powder. So they did the design of experiment, so that they can optimise the parameters to get maximum efficiency and best percentage and it is in titania nanopowder.

(Refer Slide Time: 1:30)



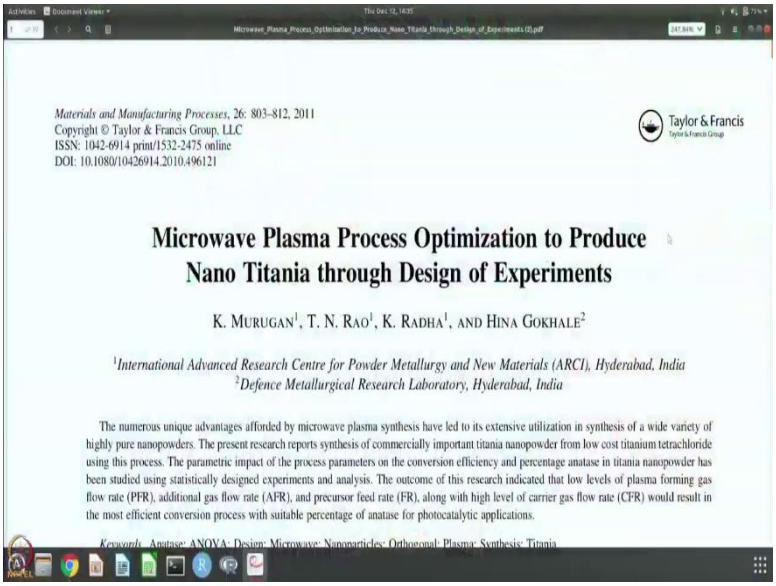


So there are lots of R libraries to do design of experiments DoE.base, FrF2, DoE.wrapper and there is also a plug in which will allow you to use GOI to do this analysis. Professor Ulrike Gromping has given lots of material including some slides and manuals etc. and there is also a CRAN page which is for design of experiments and analysis of experimental data which is what is shown here.

So it is at the CRAN R project experiment design of experiment's page. That there lots of libraries and material that is available and what you can do is given and they also refer to Gromping's work and so they have this fractional factorial tool level design which is very comprehensive R package and there are other packages also which you can explore for doing the design of experiments. However in this session, this is not what we are planning to do. We want to carry out the analysis using linear model fitting and ANOVA.

So we rea going to do the most direct way possible so that you will understand it better and we will use this paper by Murugan et al to confirm that our analysis is okay and we are getting the same results as is reported in this paper. However, it might be a good idea to explore these and make some experimental design yourself, for example, the design matrix, how do you design and things like that. So you can use this libraries and generate them yourselves.

(Refer Slide Time: 3:25)



Materials and Manufacturing Processes, 26: 803–812, 2011

ISSN: 1042-6914 print/1532-2475 online
DOI: 10.1080/10426914.2010.496121

Taylor & Francis

Microwave Plasma Process Optimization to Produce Nano Titania through Design of Experiments

K. MURUGAN[1], T. N. RAO[1], K. RADHA[1], AND HINA GOKHALE[2]

[1]*International Advanced Research Centre for Powder Metallurgy and New Materials (ARCI), Hyderabad, India*
[2]*Defence Metallurgical Research Laboratory, Hyderabad, India*

The numerous unique advantages afforded by microwave plasma synthesis have led to its extensive utilization in synthesis of a wide variety of highly pure nanopowders. The present research reports synthesis of commercially important titania nanopowder from low cost titanium tetrachloride using this process. The parametric impact of the process parameters on the conversion efficiency and percentage anatase in titania nanopowder has been studied using statistically designed experiments and analysis. The outcome of this research indicated that low levels of plasma forming gas flow rate (PFR), additional gas flow rate (AFR), and precursor feed rate (FR), along with high level of carrier gas flow rate (CFR) would result in the most efficient conversion process with suitable percentage of anatase for photocatalytic applications.

Keywords Anatase; ANOVA; Design; Microwave; Nanoparticles; Orthogonal; Plasma; Synthesis; Titania

So the paper has replication experiments, so two levels and two sets of experiments had been done and what is reported in the paper is lump data only the averages are reported. But we have the raw data available to thanks to professor Gokhale who is one of the co-authors, so we are going to use the raw data and I am going to share that raw data also with you, so you can do the analysis on the data and confirm that the results that are reported in the paper is what we are getting.

And I am not going to reproduce all the results, I am going to leave some of them out so that you can try and do it yourself. It might take a bit of effort and little bit of reading and also some amount of practicing with R, and thinking about what the quantities are and how they are calculated and so on, but you have been thought design of experiments in the other part of the course and so you should be able to take all that knowledge and use R to solve whatever is left out.

(Refer Slide Time: 4:55)

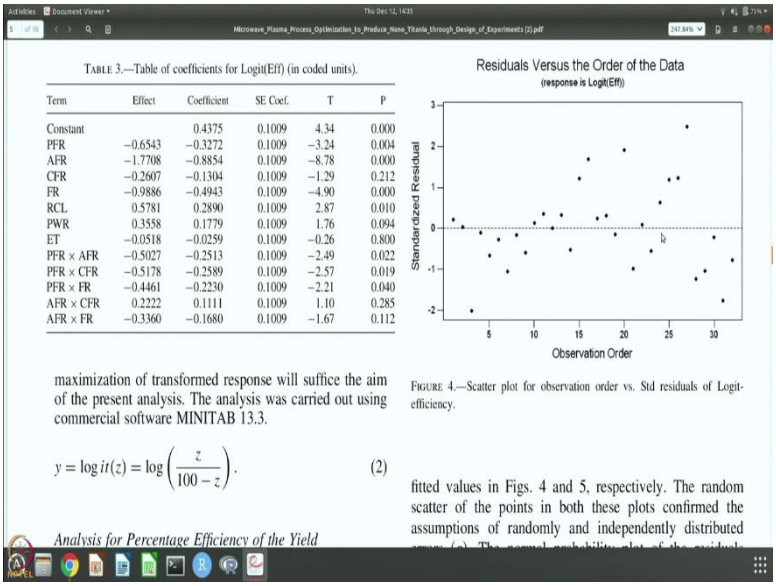


TABLE 3.—Table of coefficients for Logit(Eff) (in coded units).

Term	Effect	Coefficient	SE Coef.	T	P
Constant		0.4375	0.1009	4.34	0.000
PFR	−0.6543	−0.3272	0.1009	−3.24	0.004
AFR	−1.7708	−0.8854	0.1009	−8.78	0.000
CFR	−0.2607	−0.1304	0.1009	−1.29	0.212
FR	−0.9886	−0.4943	0.1009	−4.90	0.000
RCL	0.5781	0.2890	0.1009	2.87	0.010
PWR	0.3558	0.1779	0.1009	1.76	0.094
ET	−0.0518	−0.0259	0.1009	−0.26	0.800
PFR × AFR	−0.5027	−0.2513	0.1009	−2.49	0.022
PFR × CFR	−0.5178	−0.2589	0.1009	−2.57	0.019
PFR × FR	−0.4461	−0.2230	0.1009	−2.21	0.040
AFR × CFR	0.2222	0.1111	0.1009	1.10	0.285
AFR × FR	−0.3360	−0.1680	0.1009	−1.67	0.112

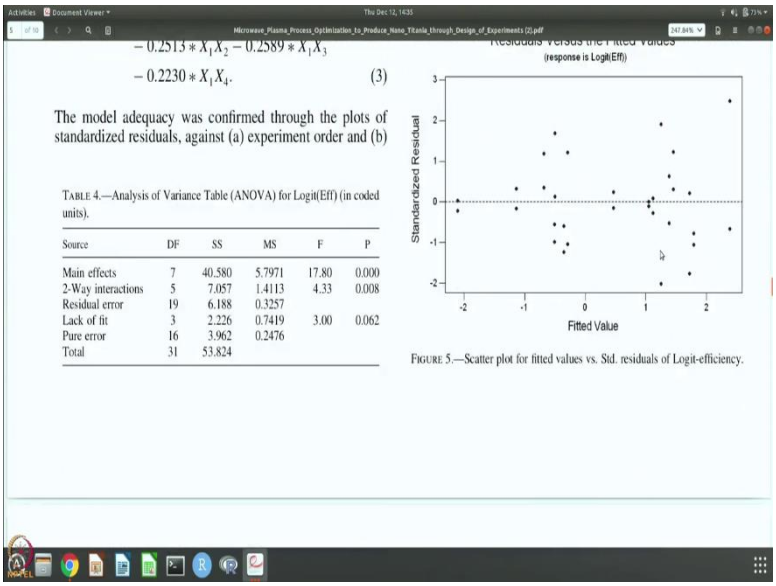
maximization of transformed response will suffice the aim of the present analysis. The analysis was carried out using commercial software MINITAB 13.3.

$$y = \log it(z) = \log\left(\frac{z}{100 - z}\right). \quad (2)$$

FIGURE 4.—Scatter plot for observation order vs. Std residuals of Logit-efficiency.

Analysis for Percentage Efficiency of the Yield

fitted values in Figs. 4 and 5, respectively. The random scatter of the points in both these plots confirmed the assumptions of randomly and independently distributed



$$- 0.2513 * X_1X_2 - 0.2589 * X_1X_3$$
$$- 0.2230 * X_1X_4. \quad (3)$$

The model adequacy was confirmed through the plots of standardized residuals, against (a) experiment order and (b)

TABLE 4.—Analysis of Variance Table (ANOVA) for Logit(Eff) (in coded units).

Source	DF	SS	MS	F	P
Main effects	7	40.580	5.7971	17.80	0.000
2-Way interactions	5	7.057	1.4113	4.33	0.008
Residual error	19	6.188	0.3257		
Lack of fit	3	2.226	0.7419	3.00	0.062
Pure error	16	3.962	0.2476		
Total	31	53.824			

FIGURE 5.—Scatter plot for fitted values vs. Std. residuals of Logit-efficiency.

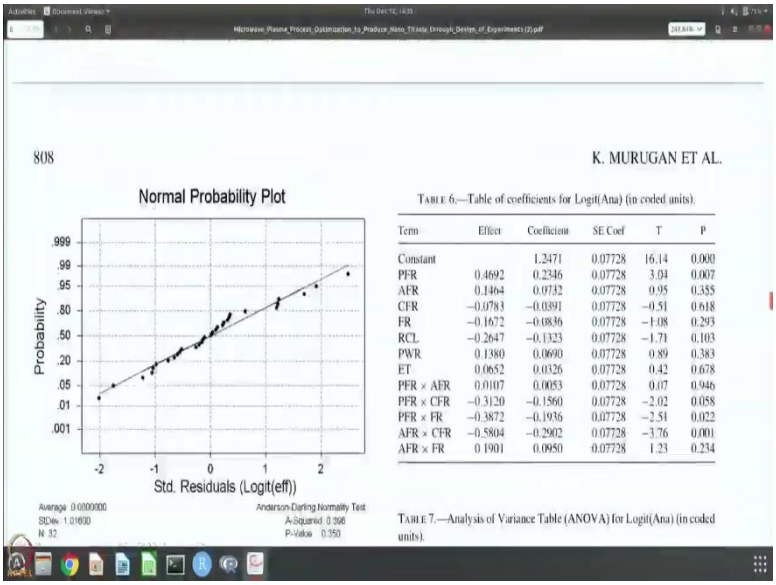
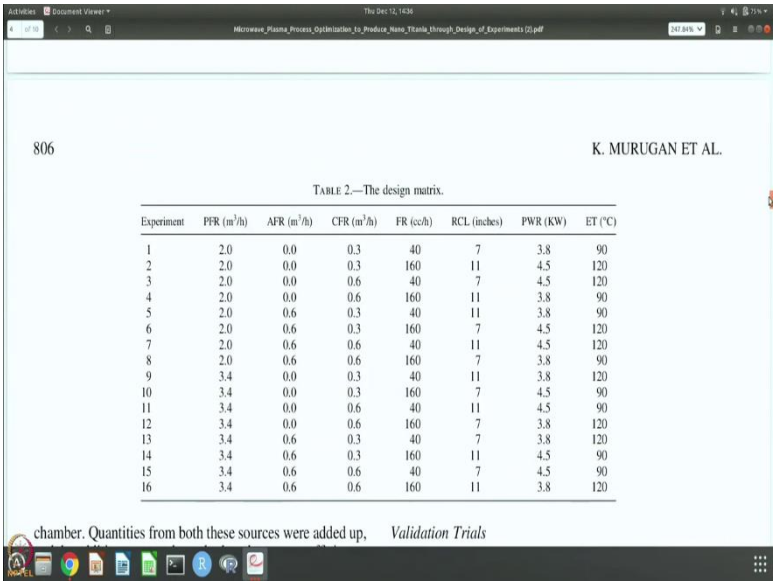


TABLE 6.—Table of coefficients for Logit(Ana) (in coded units).

Term	Effect	Coefficient	SE Coef	T	P
Constant		1.2471	0.07728	16.14	0.000
PFR	0.4692	0.2346	0.07728	3.04	0.007
AFR	0.1464	0.0732	0.07728	0.95	0.355
CFR	−0.0783	−0.0391	0.07728	−0.51	0.618
FR	−0.1672	−0.0836	0.07728	−1.08	0.293
RCL	−0.2647	−0.1323	0.07728	−1.71	0.103
PWR	0.1380	0.0690	0.07728	0.89	0.383
ET	0.0652	0.0326	0.07728	0.42	0.678
PFR × AFR	0.0107	0.0053	0.07728	0.07	0.946
PFR × CFR	−0.3120	−0.1560	0.07728	−2.02	0.058
PFR × FR	−0.3872	−0.1936	0.07728	−2.51	0.022
AFR × CFR	−0.5804	−0.2902	0.07728	−3.76	0.001
AFR × FR	0.1901	0.0950	0.07728	1.23	0.234

TABLE 7.—Analysis of Variance Table (ANOVA) for Logit(Ana) (in coded units).



806 K. MURUGAN ET AL.

TABLE 2.—The design matrix.

Experiment	PFR (m^3/h)	AFR (m^3/h)	CFR (m^3/h)	FR (cc/h)	RCL (inches)	PWR (KW)	ET (°C)
1	2.0	0.0	0.3	40	7	3.8	90
2	2.0	0.0	0.3	160	11	4.5	120
3	2.0	0.0	0.6	40	7	4.5	120
4	2.0	0.0	0.6	160	11	3.8	90
5	2.0	0.6	0.3	40	11	3.8	90
6	2.0	0.6	0.3	160	7	4.5	120
7	2.0	0.6	0.6	40	11	4.5	120
8	2.0	0.6	0.6	160	7	3.8	90
9	3.4	0.0	0.3	40	11	3.8	120
10	3.4	0.0	0.3	160	7	4.5	90
11	3.4	0.0	0.6	40	11	4.5	90
12	3.4	0.0	0.6	160	7	3.8	120
13	3.4	0.6	0.3	40	7	3.8	120
14	3.4	0.6	0.3	160	11	4.5	90
15	3.4	0.6	0.6	40	7	4.5	90
16	3.4	0.6	0.6	160	11	3.8	120

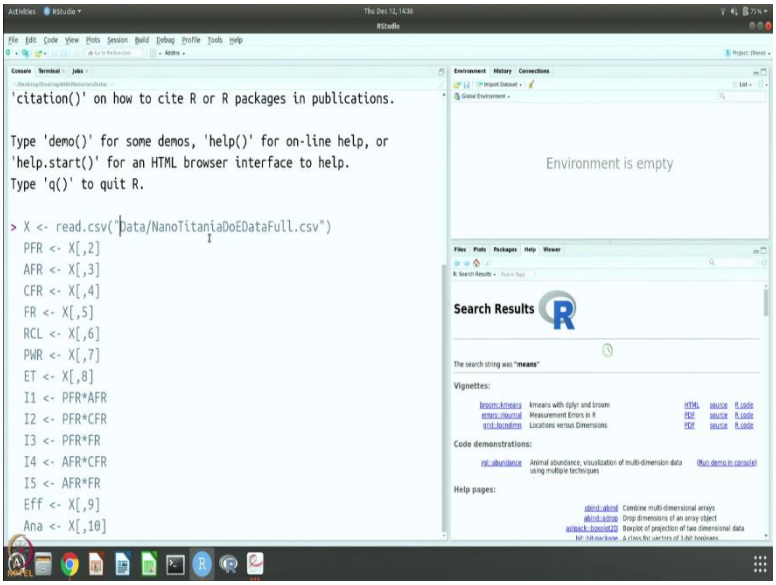
chamber. Quantities from both these sources were added up, *Validation Trials*

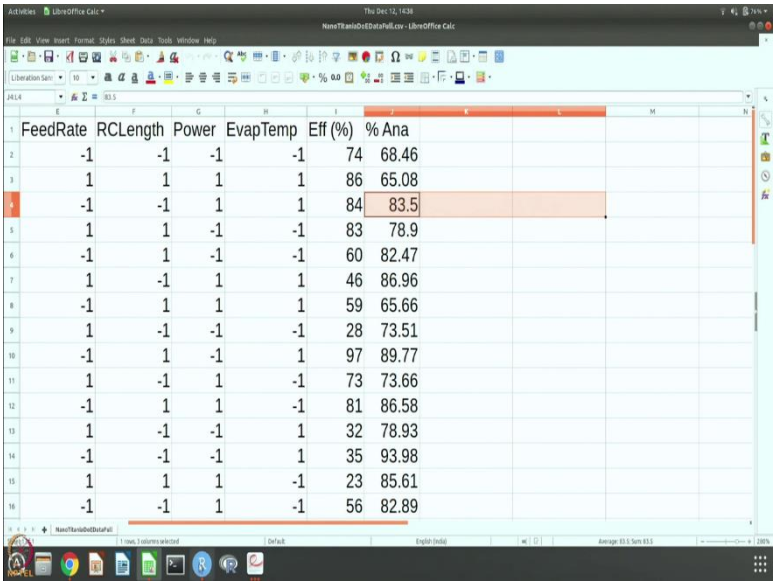
But in principal, you should be able to reproduce all the results in the paper using the lecture that you have heard on Design of Experiment and the code and the data that I am going to share with you. So let us do this, we will use; I am also going to keep this paper handy. So this is the paper by Murugan et al, so this is about microwave plasma process optimisation to produce nano titania through design of experiments and it also has some interesting conclusions which is worth going through.

And our aim is to actually take the data and produce the results. By results we mean that this table of coefficients and this analysis of variants table for example and these figures eight residuals figures and so on. So you should be able to produce this normal probability plot and the table of means plots.

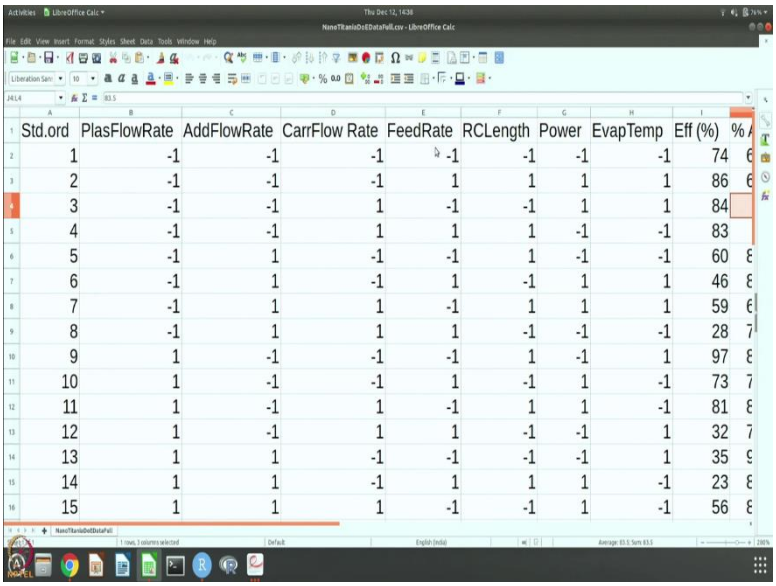
They might involve some effort but that will help you understand the methodology as well as R programming better. So I am going to do for one. So there are two exercises, they are repetitions of the same exercise, but one is done for percentage efficiency. The other one is done is for anatase percentage. So let us do this. So we are going to reproduce using R, okay.

(Refer Slide Time: 6:19)





FeedRate	RCLength	Power	EvapTemp	Eff (%)	% Ana
-1	-1	-1	-1	74	68.46
1	1	1	1	86	65.08
-1	-1	1	1	84	83.5
1	1	-1	-1	83	78.9
-1	1	-1	-1	60	82.47
1	-1	1	1	46	86.96
-1	1	1	1	59	65.66
1	-1	-1	-1	28	73.51
-1	1	-1	1	97	89.77
1	-1	1	-1	73	73.66
-1	1	1	-1	81	86.58
1	-1	-1	1	32	78.93
-1	-1	-1	1	35	93.98
1	1	1	-1	23	85.61
-1	-1	1	-1	56	82.89



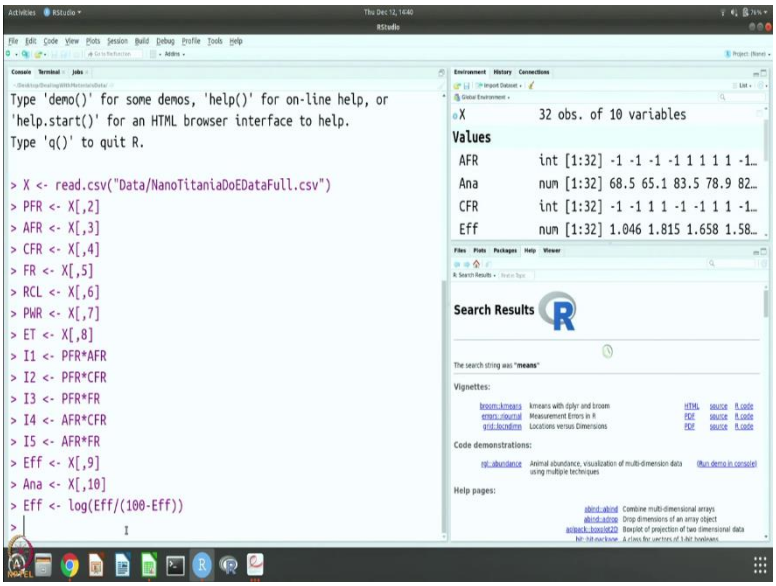
Std.ord	PlasFlowRate	AddFlowRate	CarrFlow Rate	FeedRate	RCLength	Power	EvapTemp	Eff (%)	% A
1	-1	-1	-1	-1	-1	-1	-1	74	6
2	-1	-1	-1	1	1	1	1	86	6
3	-1	-1	1	-1	-1	1	1	84	
4	-1	-1	1	1	1	-1	-1	83	
5	-1	1	-1	-1	1	-1	-1	60	8
6	-1	1	-1	1	-1	1	1	46	8
7	-1	1	1	-1	1	1	1	59	6
8	-1	1	1	1	-1	-1	-1	28	7
9	1	-1	-1	-1	1	-1	1	97	8
10	1	-1	-1	1	-1	1	-1	73	7
11	1	-1	1	-1	1	1	-1	81	8
12	1	-1	1	1	-1	-1	1	32	7
13	1	1	-1	-1	-1	-1	1	35	9
14	1	1	-1	1	1	1	-1	23	8
15	1	1	1	-1	-1	1	-1	56	8

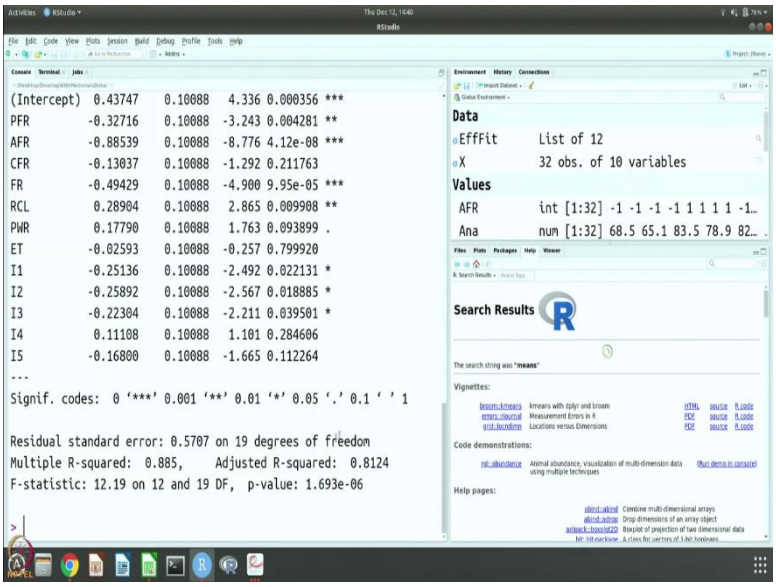
power requirement is adjustable and could be controlled manually through control on the anode current. Magnetron generated microwaves are transmitted to the gas filled water-cooled rectangular wave-guide. The plasmatron is a double walled, water cooled chamber with a central quartz tube of diameter 50 mm. The top of the plasmatron is fitted with a specially designed nozzle for feeding the plasma forming gas without turbulence. The powder synthesizing part includes reaction chamber, liquid feeding device with evaporator, heat exchanger and powder collection unit. The liquid or solid precursor is fed into the reaction chamber through a proper dosing device. One end of the water-cooled reaction chamber is connected to the plasmatron and the other end is connected with the heat exchanger. The reaction chamber length (RCL) between the plasmatron and the heat exchanger can be varied depending on the requirements. The seven controllable processing parameters are listed in Table 1.

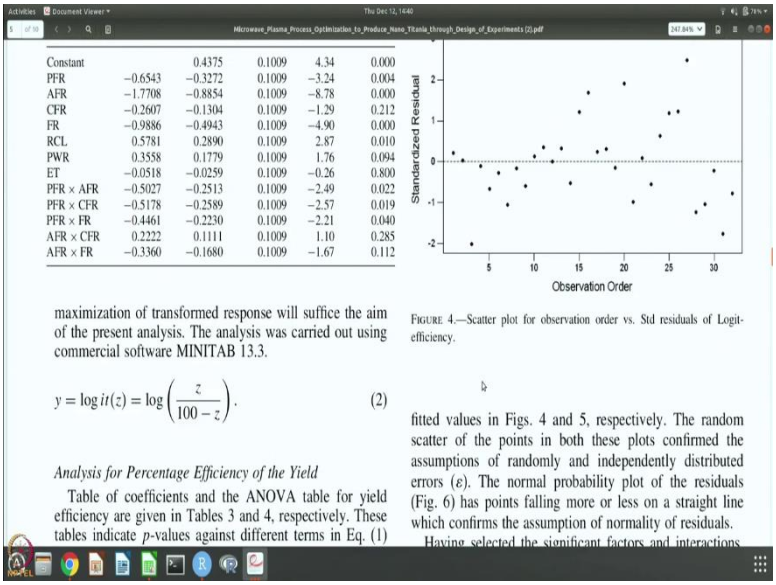
Among these, the first three factors affected the titania

TABLE 1.—Input factors, interactions, and responses.

Variable	Input or Response	Unit	Code
Plasma Forming Gas Flow Rate	Input	m^3/h (%)	PFR
Additional Gas Flow Rate	Input	m^3/h (%)	AFR
Carrier Gas Flow Rate	Input	m^3/h (%)	CFR
$TiCl_4$ Feed Rate	Input	cc/h	FR
Reaction Chamber Length	Input	inches	RCL
Magnetron Power	Input	kW	PWR
Evaporating Temperature	Input	°C	ET
PFR × AFR	Interaction		
PFR × CFR	Interaction		
PFR × FR	Interaction		
AFR × CFR	Interaction		
AFR × FFR	Interaction		
Percentage efficiency	Response		Eff
Percentage of anatase	Response		Ana







Constant		0.4375	0.1009	4.34	0.000
PFR	−0.6543	−0.3272	0.1009	−3.24	0.004
AFR	−1.7708	−0.8854	0.1009	−8.78	0.000
CFR	−0.2607	−0.1304	0.1009	−1.29	0.212
FR	−0.9886	−0.4943	0.1009	−4.90	0.000
RCL	0.5781	0.2890	0.1009	2.87	0.010
PWR	0.3558	0.1779	0.1009	1.76	0.094
ET	−0.0518	−0.0259	0.1009	−0.26	0.800
PFR × AFR	−0.5027	−0.2513	0.1009	−2.49	0.022
PFR × CFR	−0.5178	−0.2589	0.1009	−2.57	0.019
PFR × FR	−0.4461	−0.2230	0.1009	−2.21	0.040
AFR × CFR	0.2222	0.1111	0.1009	1.10	0.285
AFR × FR	−0.3360	−0.1680	0.1009	−1.67	0.112

maximization of transformed response will suffice the aim of the present analysis. The analysis was carried out using commercial software MINITAB 13.3.

$$y = \log it(z) = \log\left(\frac{z}{100 - z}\right). \quad (2)$$

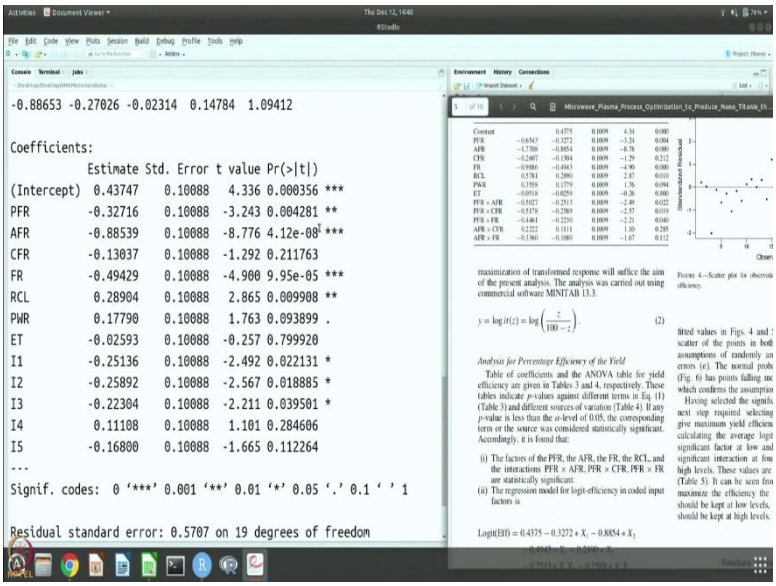
Analysis for Percentage Efficiency of the Yield

Table of coefficients and the ANOVA table for yield efficiency are given in Tables 3 and 4, respectively. These tables indicate p-values against different terms in Eq. (1)

FIGURE 4.—Scatter plot for observation order vs. Std residuals of Logit-efficiency.

fitted values in Figs. 4 and 5, respectively. The random scatter of the points in both these plots confirmed the assumptions of randomly and independently distributed errors (ε). The normal probability plot of the residuals (Fig. 6) has points falling more or less on a straight line which confirms the assumption of normality of residuals.

Having selected the significant factors and interactions



```
-0.88653 -0.27026 -0.02314  0.14784  1.09412

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.43747    0.10088   4.336 0.000356 ***
PFR         -0.32716    0.10088  -3.243 0.004281 **
AFR         -0.88539    0.10088  -8.776 4.12e-08 ***
CFR         -0.13037    0.10088  -1.292 0.211763
FR          -0.49429    0.10088  -4.900 9.95e-05 ***
RCL          0.28904    0.10088   2.865 0.009908 **
PWR          0.17790    0.10088   1.763 0.093899 .
ET          -0.02593    0.10088  -0.257 0.799920
I1          -0.25136    0.10088  -2.492 0.022131 *
I2          -0.25892    0.10088  -2.567 0.018885 *
I3          -0.22304    0.10088  -2.211 0.039501 *
I4           0.11108    0.10088   1.101 0.284606
I5          -0.16800    0.10088  -1.665 0.112264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5707 on 19 degrees of freedom
```

So I am going to take, the first step is to read the data. So let us do that. So we want to read this data and we, what the paper reports is only the, so this is the design matrix and this is what is shown in coded form in the data that I will show you and the data on efficiency and percentage anatase is needed that is not here, so I am going to show you the data. So this might be familiar to you from the presentation and in the presentation what was shown for these two is the average of efficiency in anatase of sixteen experiments.

But here we are showing you the full data, so I am going to show for thirty two experiments and so the first sixteen and the next sixteen is basically reputation of the first sixteen. And in each case what is the percentage anatase obtained and efficiency percentage is what is given. So this is the data that we are going to use. So, the first step is to read the data and we know that the different columns of the table give you this PFR, AFR, CFR, FR, RC, Power, ET etc.

So I am going to use the same symbols because these were the once that are used in the paper, so I am going to say second column is PFR, third column is AFR, fourth column is CRF etc. And then there are interactions and the interactions are between PFR and the AFR, CFR, FR and AFR with CFR and FR. So these were the five interactions which were decided to be important for this study and that is given here in this table.

So PFR with AFR, CFR, FR and AFR with CFR, FR; so these were decoded to be important and these are the other seven parameters. So seven plus five there are twelve parameters and the response is percentage efficacy and percentage anatase which is what is given. So once we have this, so let us read and get this data. So once we have this data in place, then we can start our analysis.

The first thing we want to do is that we want to do the logit transformation, which is very important and I as you have learnt, if you do not do in this case, you might get wrong results like you might get percentages which go above hundred and things like that. So after we do the logit transformation, we want to fit and we want to look at how the fitted parameters look like.

So this is the, so I am saying that this logit efficiency is a function PFR, AFR, CFR, FR, RCL, PWR, ET and these 5 interaction parameters and there is a constant that will show up anyway. So now you can see that the fit actually gives you the fitted parameters along with their standard error and now let us compare what we have in the paper.

Okay, so here is what is there in the paper and you can see that this 4375 is this and 3272 is here and 8854 is here and so on. So basically, this column corresponds to the estimate of the parameter and this is the standard error. So it is 10088 which is 1009 up to 4 decimal places and if you take it up to second decimal place, the T value is given for 0.336 minus 3.24, 8.78 etc. and these are the P values.

And in the paper alpha was taken to be 0.05, so anything less than 0.05 here was considered to be important and the conclusion that was drawn was PFR, FR, FR, RCL and the 3 interactions of PFR with the other quantities where statistically significant. And here you can see we have the statistical significance marked by 3 stars, 2 stars and 1 star. So anything up to 1 star is 0.05, so point is actually 0.1, so it is a 10 percent significance level.

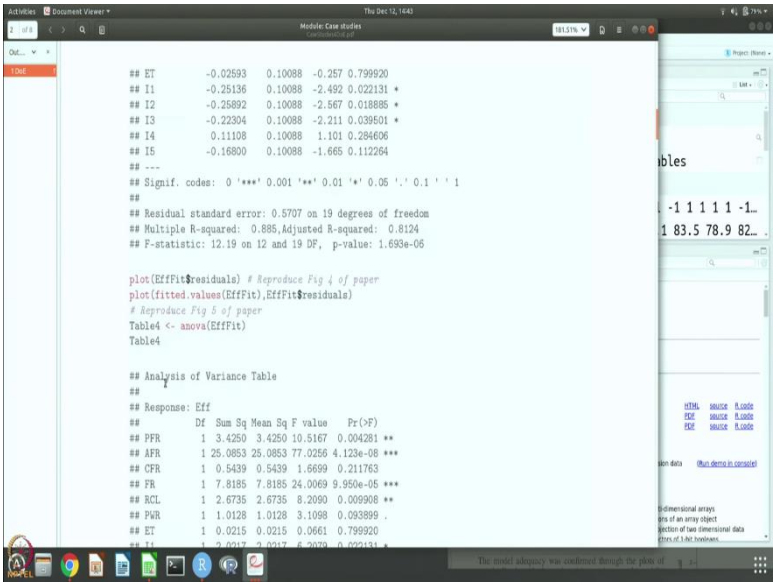
So the other ones, so you can see that I1, I2, I3 which corresponds to this PFR, AFR, PFR, CFR, PFR, FR are considered to be important here also, so significance code indicated that

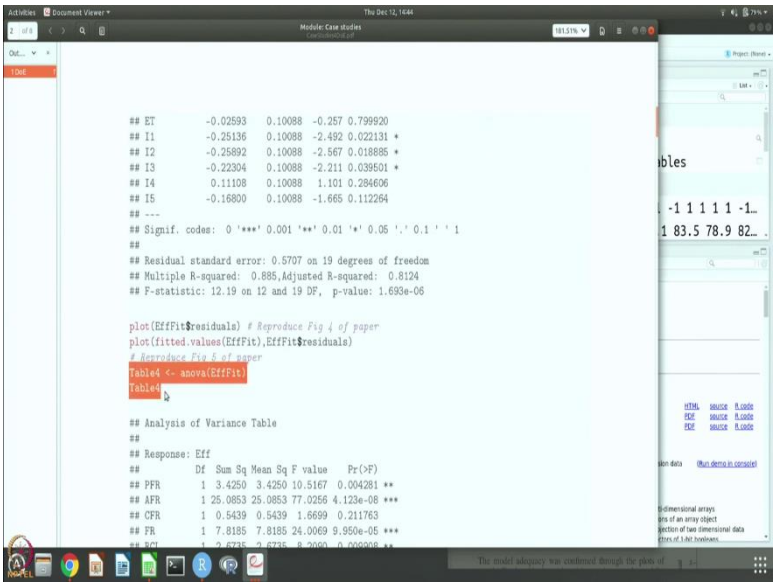
with alpha 0.05 these 3 are significant. These 3 are significant, these 2 are significant even with alpha 0.01, so obviously for 0.5 they are significant, so that is the RCL and PFR and then with 0.001 level of significance, these 2 Fr and the intercept and AFR are important.

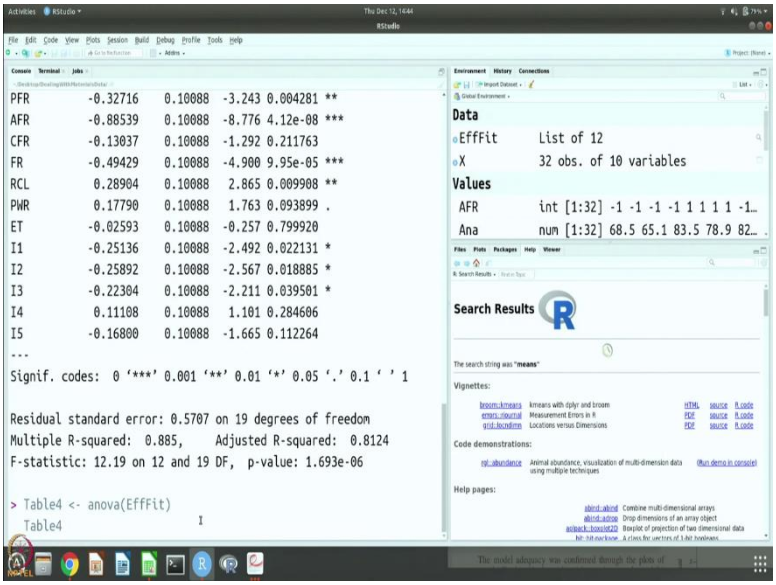
So these are these 3 are important. That is what is given now, so 0.4375 also that is this quantity plus minus 0.3272 PFR that is X1 and minus 0.8854 AFR. And the next one CFR is not significant, so X3 is kept and then it is X4 which is given 4943, so that is given with X4 and then the 5th one RCL is important.

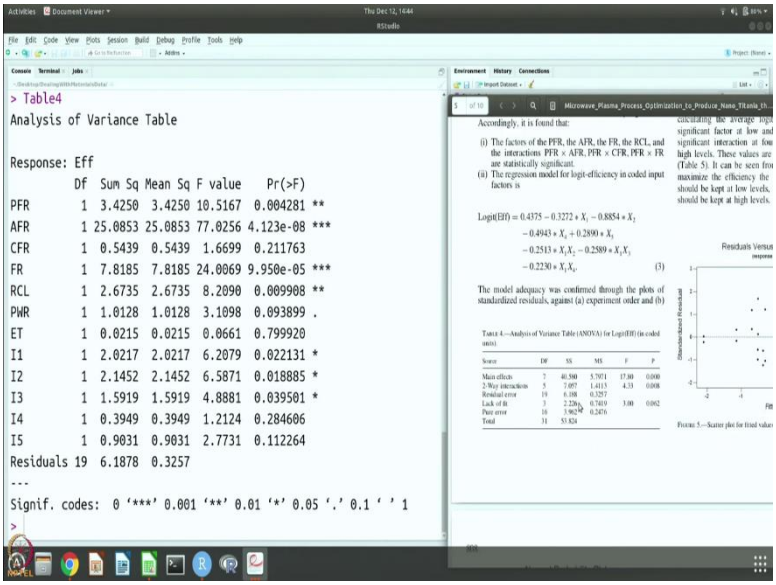
So 2890 X5 is given but X6 is not important, so we have left it out and then 0.2513 X1, X2, so 2513 X1, X2, 2590, 2589 that is X1, X3 and 2230 X1, X4, so these are the parameters which are important. So it is a same information which is here and same table from which same conclusion is being drawn. Of course the next step is to do ANOVA on this linear fitting that we have done. So let us do that.

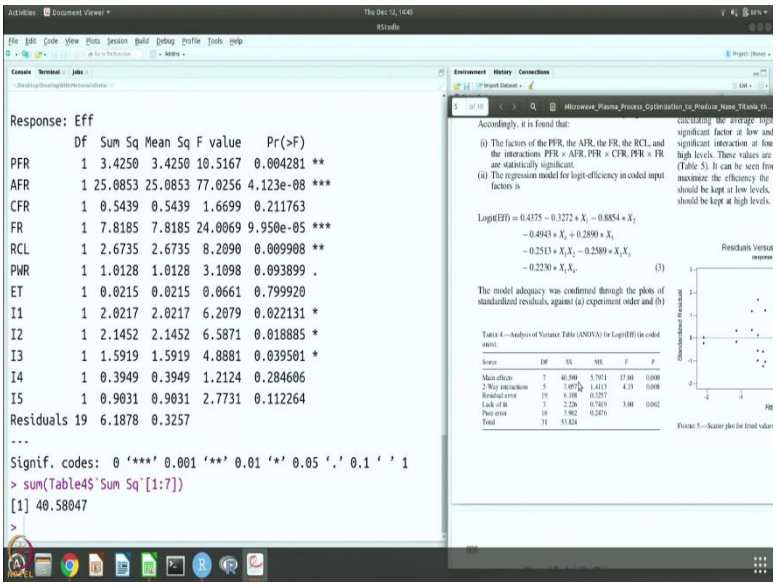
(Refer Slide Time: 13:30)









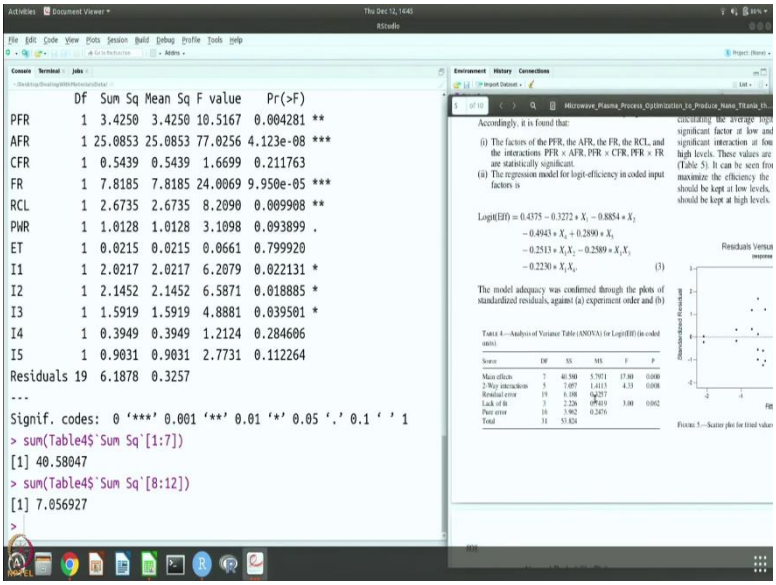


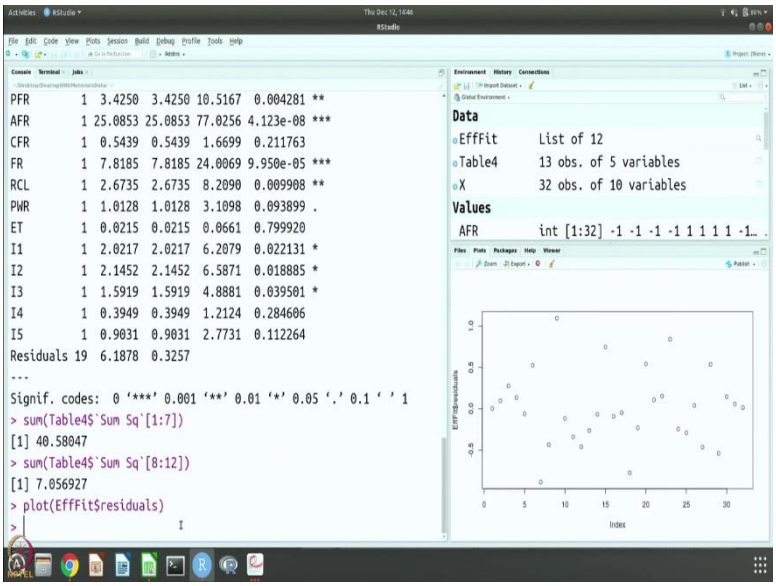
So you get these values here and you can see that this is partly a reproduction of a this table which is the next table and for example residual say 6.188, so that is the information here and they have also included the main effects and two way interactions separately with some of squares you can do that too, so here is the command to do that. Let us do. So this is 40.5804 which is the same quantity, which is given 40.580.

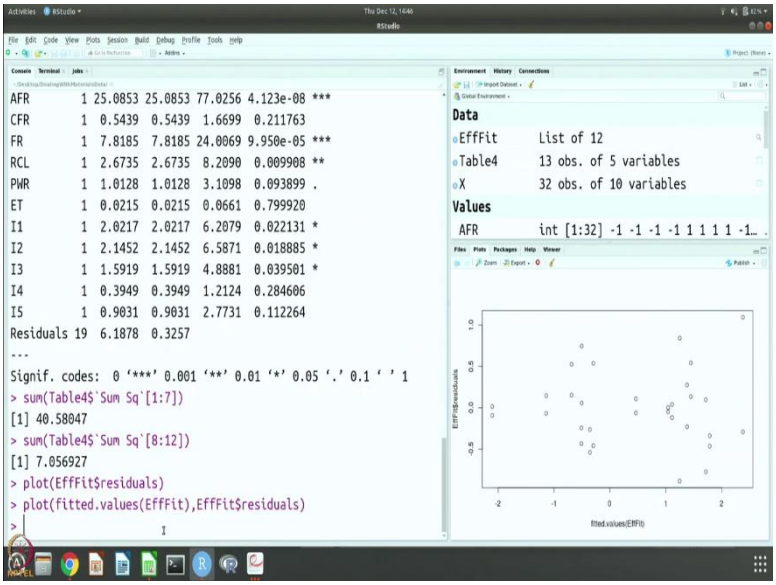
And if you go from 8 to 12 which are the two way interactions, so you will get 7.057 which is the quantity that you get here, two way interactions 7.057. So basically the ANOVA table can also be reproduced and off course we also want to reproduce the figures 4 and 5 which is for scatter plot of observation versus residuals, of course, I do not have the data and observation order, so this will be slightly jumbled up but you can look at the presentation where observation

order is also available. So you can get it in that format and here is the fitted value versus residuals.

(Refer Slide Time: 15:40)



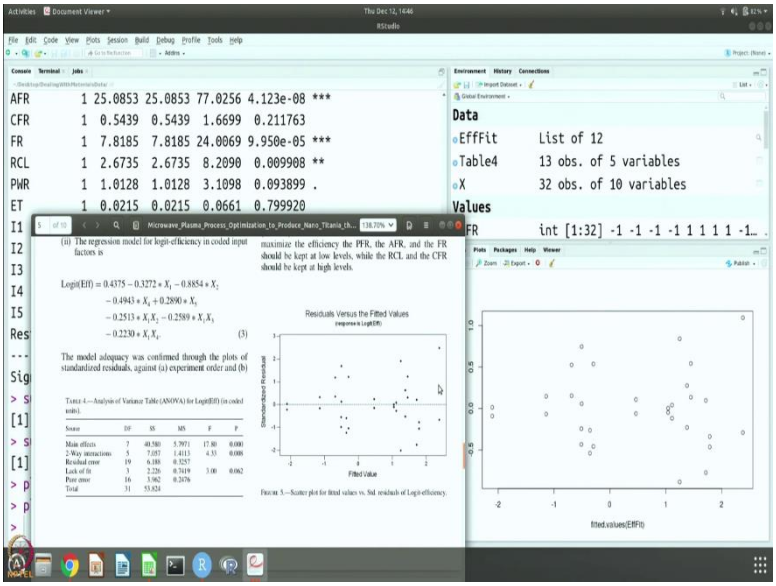




So, let us do those two plots now, they can also be done rather straight forward, so this is for plotting the residual so you get the plot, the other one is for plotting the fitted values versus residuals, that is also a scatter plot. So let us do that and this is the plot and you can compare it with the plot from the papers, so this is the plot form the paper.

So you can see that we get the same plot and you can of course draw line at zero separate this data and look at how it looks. So this is for doing the up to residuals and I am going to leave the normal probability plot for you to explore as well as the total means for logit efficiency for you to explore. Once we have done this exercise, we can repeat the same thing with the anatase percentage, right.

(Refer Slide Time: 17:15)



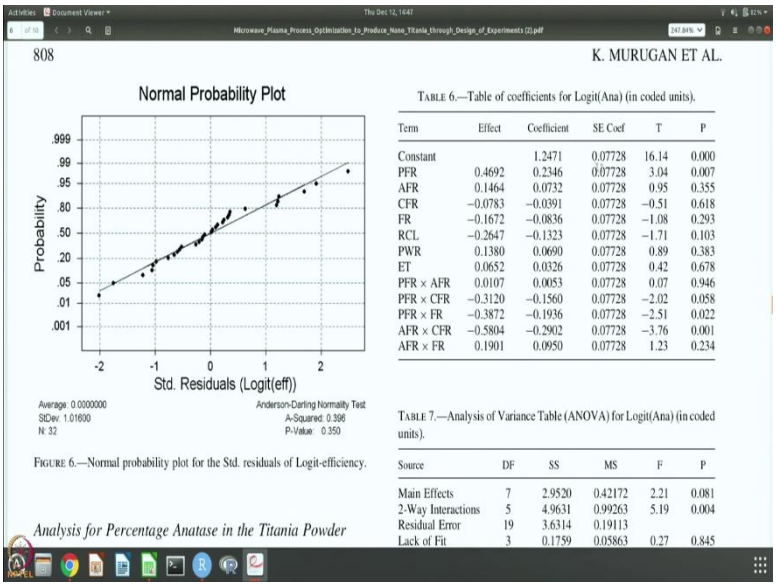


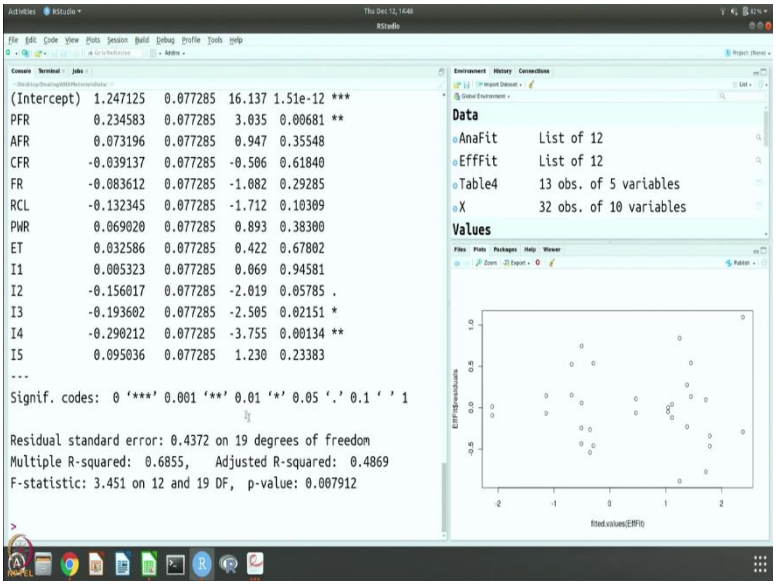
TABLE 6.—Table of coefficients for Logit(Ana) (in coded units).

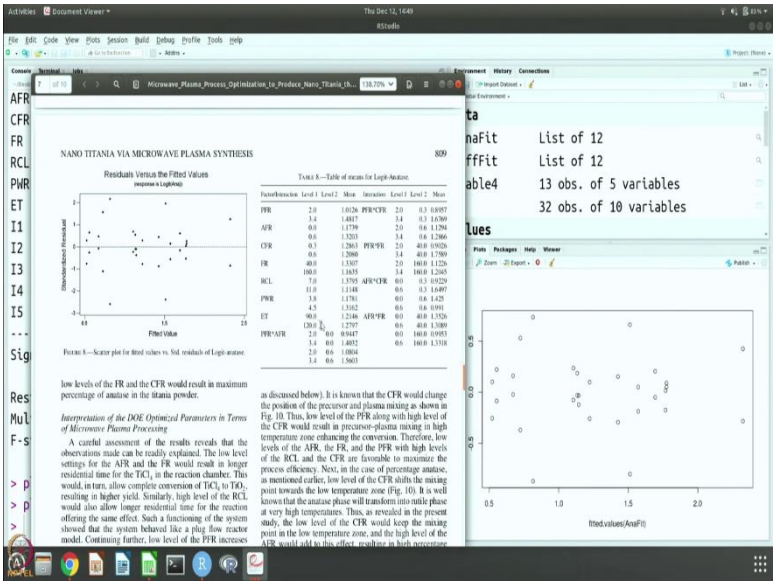
Term	Effect	Coefficient	SE Coef	T	P
Constant		1.2471	0.07728	16.14	0.000
PFR	0.4692	0.2346	0.07728	3.04	0.007
AFR	0.1464	0.0732	0.07728	0.95	0.355
CFR	−0.0783	−0.0391	0.07728	−0.51	0.618
FR	−0.1672	−0.0836	0.07728	−1.08	0.293
RCL	−0.2647	−0.1323	0.07728	−1.71	0.103
PWR	0.1380	0.0690	0.07728	0.89	0.383
ET	0.0652	0.0326	0.07728	0.42	0.678
PFR × AFR	0.0107	0.0053	0.07728	0.07	0.946
PFR × CFR	−0.3120	−0.1560	0.07728	−2.02	0.058
PFR × FR	−0.3872	−0.1936	0.07728	−2.51	0.022
AFR × CFR	−0.5804	−0.2902	0.07728	−3.76	0.001
AFR × FR	0.1901	0.0950	0.07728	1.23	0.234

FIGURE 6.—Normal probability plot for the Std. residuals of Logit-efficiency.

TABLE 7.—Analysis of Variance Table (ANOVA) for Logit(Ana) (in coded units).

Source	DF	SS	MS	F	P
Main Effects	7	2.9520	0.42172	2.21	0.081
2-Way Interactions	5	4.9631	0.99263	5.19	0.004
Residual Error	19	3.6314	0.19113		
Lack of Fit	3	0.1759	0.05863	0.27	0.845





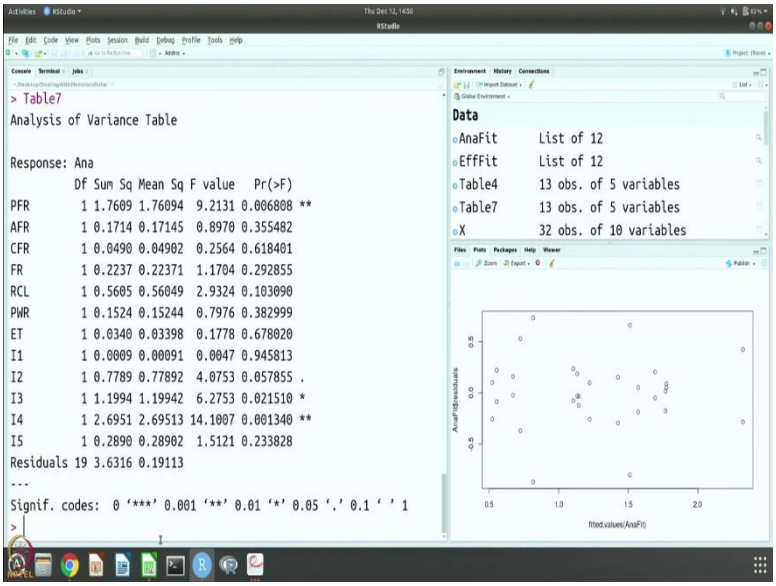
So let us do that. So we take logit on anatase percentage and we again fit it for a this thing and you get the fitted parameters and you get the significant ones to be I3, I4, PFR and the intercept and that is the conclusion that is also drawn from this namely that for anatase you see that constant PFR, the rest of them are greater than 0.05 except for PFR, FR, DFR, CFR, so these are the only for which are important.

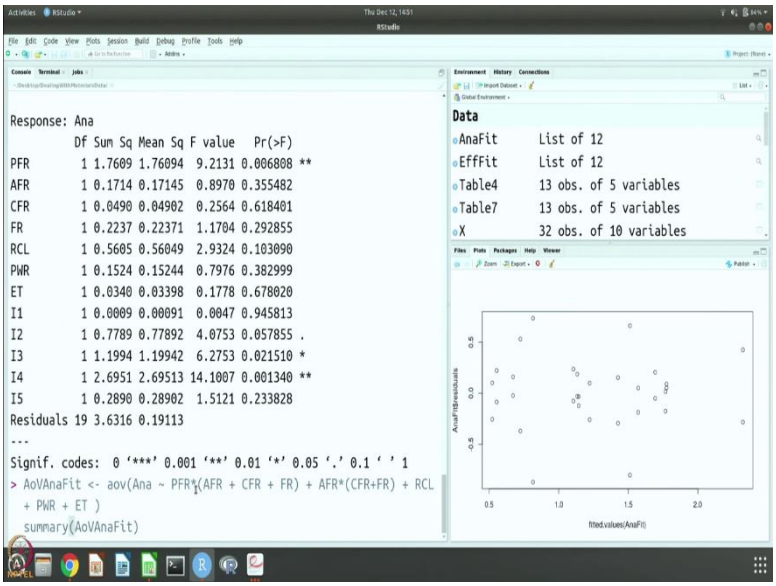
And that is the same conclusion you draw from here, so up to 0.01significance alpha 0.01 it is I4 and PFR and 0.05 significance it is I3 and 0.001 significance it is the intercept. So that is what is given here in the in the paper. So they look at it and then they say that constant X1, X4, X1, X3 and PFR was already there. Okay. So then you can again do the same plots and do the ANOVA and so on.

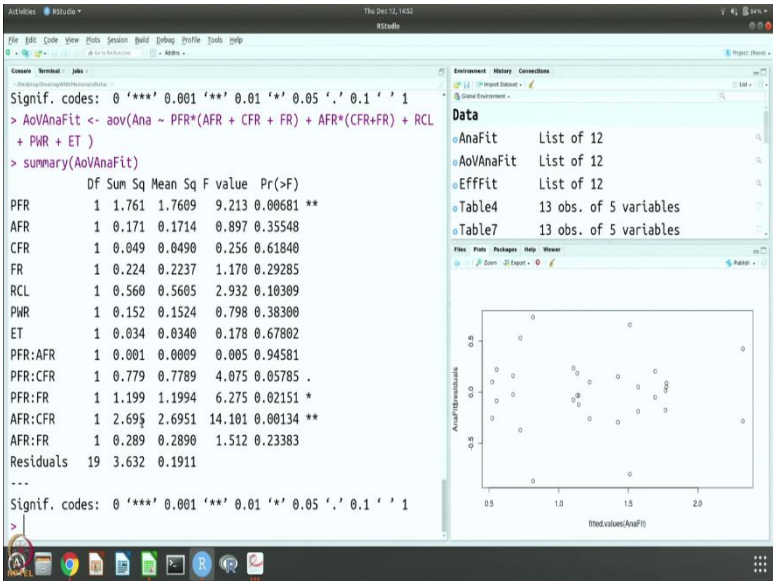
So those are the commands that are given here after you have the fit, you can make those plots, so you can make the residual plot you can also make the fitted values versus residuals plot and you can compare this figure with what is given in the paper. So this is observation order. So you can see here how this figure compares with this, so we are getting same results and off course the other quantity to reproduce is the table of ANOVA and you can do that by reproducing the table 7 or at least parts of table 7 by doing the ANOVA.

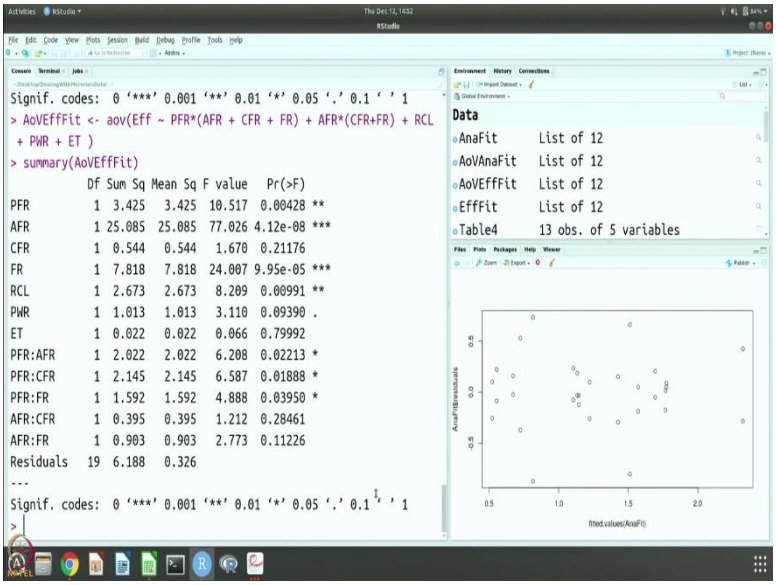
So you draw the same conclusions and again you can just add up the sums of squares for these are adapt the sums up squares for these and you will get the numbers which are the same, so we did and you can add up. So you will get 2.952 and 4.963 and you see that it is 2.952 and 4.963, so that is the way you get and there is also of course residual which is 3.6314 and here residual is 3.6316, so it is the same result, so we can see that we have reproduced.

(Refer Slide Time: 20:40)









Now this is not the only command, there is also another way of doing this analysis, which is what I want to show here, so let us do this. So this is the command AOV so anatase the logit transformed is functioned of, so this star symbol means that individual that is PFR plus AFR plus PFR colon AFR which is the cross term and any duplicate will be removed. So PFR, CFR means PFR plus CFR and PFR, CFR but PFR is already there so that will be removed and so here this is another way of writing the same interaction.

So PFR has an interaction with all these AFR has an interaction with these two, so those are the 5 interactions and the remaining terms without interaction and with interaction terms. So you can do and then you can do a fitting and you get the same result, I mean the results are not at all different, which is expected except that now instead of I1, I2 etc., you can clearly see what the interactions are also.

So this is another one line command for you to do the same analysis and get the same sort of results and you can also do this not just for ANOVA, you can also do it for efficiency, so lets do that also. So you get the same conclusions namely that these 3 interactions are important and these 4 are important and these are not important from the point of view of the significance, right.

So this is the paper and so we are able to reproduce most of the results and the remaining ones also you will be able to reproduce yourself and its also a good idea to understand these significance levels and the table 9 validation trails, but I am going to leave that to you to explore on your own and because all the data will be available and the presentation of design of

experiments is available and this script will be available to you so it should not be very difficult for you to reproduce them on your own.

So to summarize design of experiments is very important and you can optimize process parameters by carefully setting up the experiments and then analysing them and such statistically planned experiments and the statistics that you get from them will make life easy for you in terms of optimisation otherwise there are too many parameters and you need to take a call on how many experiments you will do and how you will change the parameters and so on.

So this is a nice way of doing it and for doing that off course there are lots of libraries in R which you can use but I have also shown you that with whatever we have learned, we have learned the linear model fitting and ANOVA, using just these 2 commands or combinations of such commands you can get all the information you want.

So if you want to set up a new set of experiments for some optimisation, here is a way to explore and I also strongly urge you to go through the material on design of experiments and the libraries that are available and the other ways of doing things. For example, I have not shown how to make the design matrix but you can generate those things also using R, which will help you set up your experiments.

So we have looked at one two factor experiment and reproduce the design of experiments analysis that was done in the paper. I hope this will help you set up more experiments of your own along these slides. Thank you.