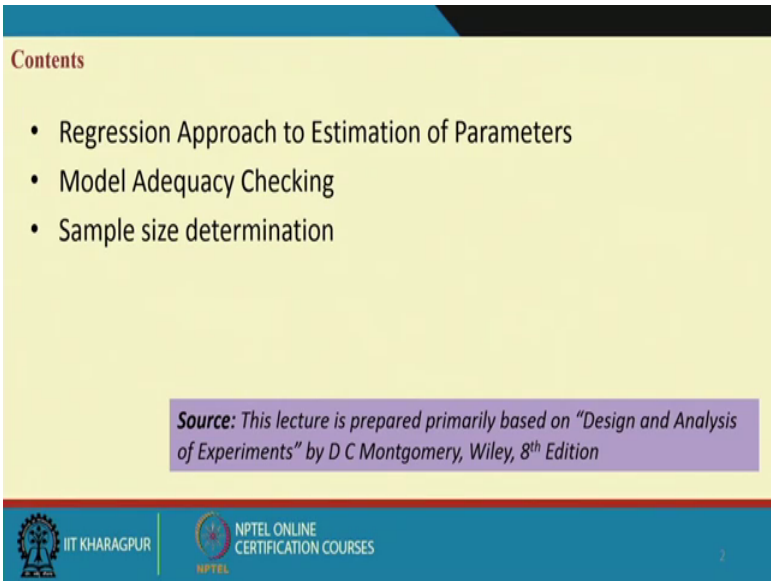
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

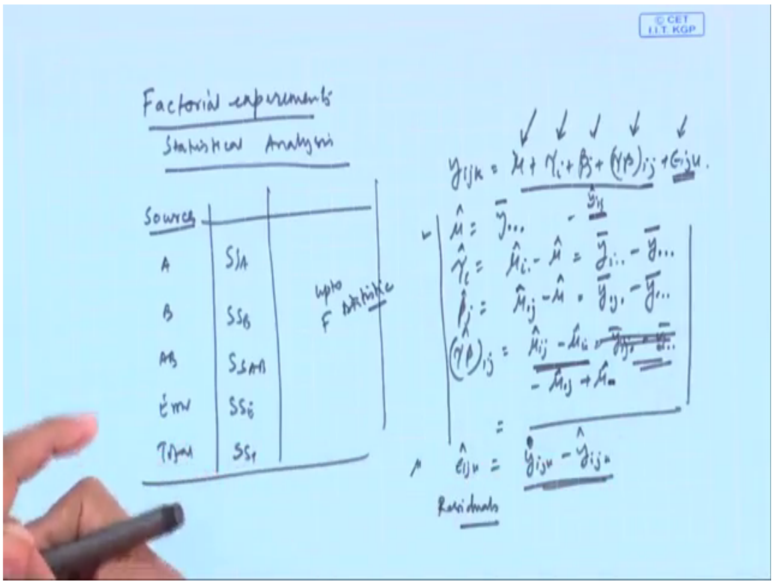
Lecture – 32
Estimation of Parameters And Model Adequacy Test for Factorial Experiments

Welcome. We will continue factorial experiment. Today's topic is estimation using Regression Approach, Model adequacy in terms of test of assumptions and Sample size calculation for Factorial Experiments.

(Refer Slide Time: 00:36)



(Refer Slide Time: 00:43)



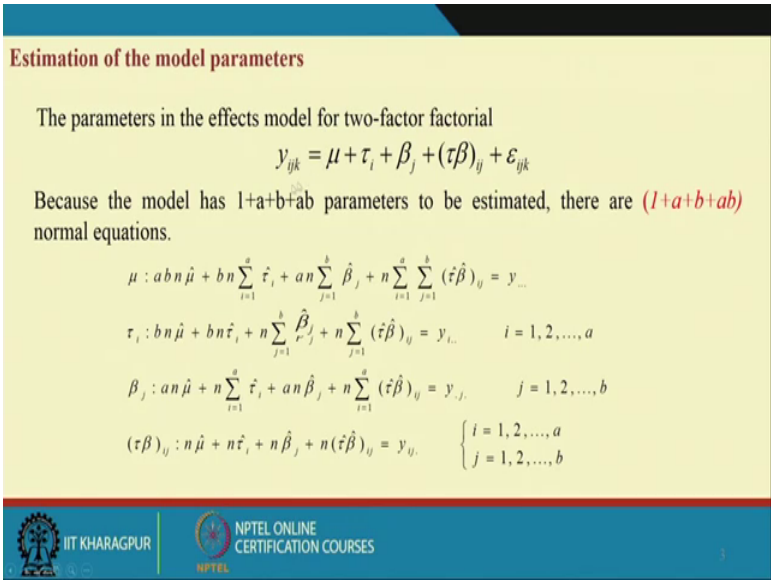
So, in last lecture, last 2 lectures we have talked about Factorial experiments. And in last lecture I have given you the Statistical Analysis of fact, factorial that is Statistical Analysis. And you have seen that, when there are multiple factors for example 2 factors, so we have computed the sources of variations A, B, A B interaction, Error then Total. And I have given you the formula for S S A, S S B, S S A B, S S Error, and S S Total. And then the other part like another table for up to, up to F statistics, and if you remember the model we consider is y i j k, then this will be mu plus tau i plus beta j plus tau beta i j plus epsilon i j k.

And the estimate of mu we have given as, that is y double, triple dot bar. Then tau i estimate is basically mu i cap minus mu cap which is basically y i dot, dot bar minus y dot dot dot triple dot bar in the same manner. Beta j cap equal to beta mu dot j cap minus mu cap which is basically y dot j dot bar minus y triple dot bar. And then, we have also computed tau beta i j cap which is mu i j cap minus mu cap then y i j dot bar minus y triple dot bar. No here formula is not this, here formula is minus mu i dot bar minus mu dot j cap plus mu dot the mu cap, and then corresponding y bar is taken and grand mean is this.

So that means, by estimation of parameters we talk about estimation for mean, grand mean, estimation for tau i, estimation for beta j, estimation for tau beta i j and also computation of residuals which will be i j k cap. So, once we compute all those things this is basically our y i j estimate, then this one will be y i j k minus y i j k estimate, this will give you the error terms or these are known as Residuals.

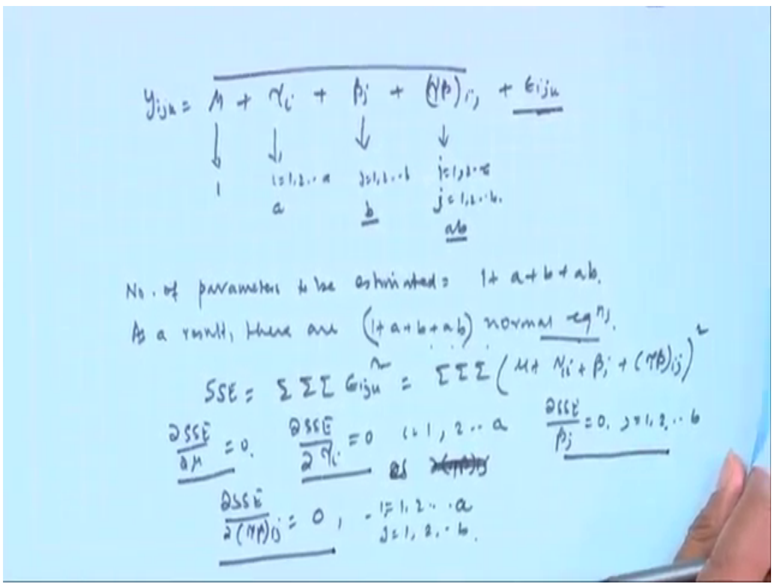
So, what, using regression approach we will show you that, how this same similar this formula are obtained. And then will go for plot, residual plots to check the Model adequacy and finally, will show you with how to compute Sample size given the threshold mean differences or given the parameter, population parameter values.

(Refer Slide Time: 04:54)



So, you see that y i j k, this is our 2 factor factorial fixed effect model. Here we have mu, tau i, beta j, tau beta i j these are the parameters to be estimated. So, we have 1 mu, a tau i, b beat beta j, and a b tau i beta j. So that means, we have 1 plus a plus b plus a b.

(Refer Slide Time: 05:29)

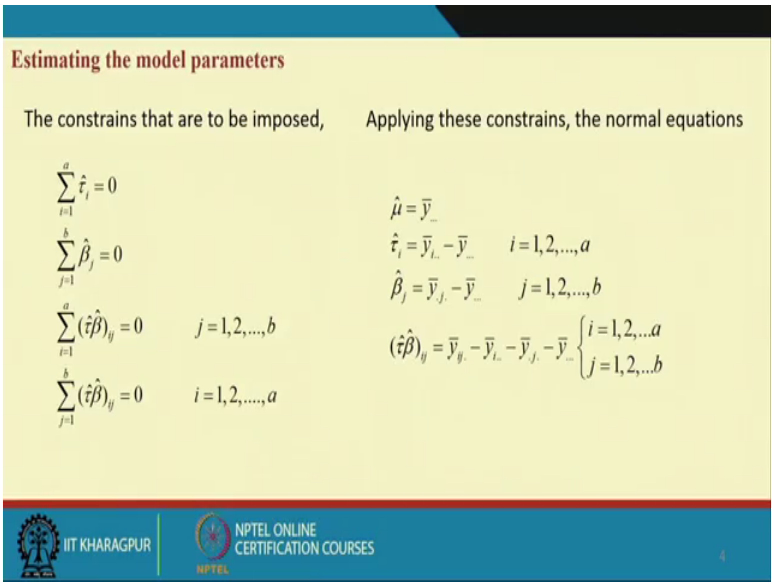


So, we have how many parameters to be estimated? We have mu that is 1 mu plus y i j k equal to mu plus tau i, we have i equal to 1, 2, a. So, a number of tau i plus beta j, we have j equal to 1, 2, b. So, b number of tau b. Now beta j plus tau beta i j were i equal to 1, 2, a, j equal to 1, 2, b. So, we have a b number of tau beta j, and then plus error is there y i j k. So, this will be estimated once this things are estimated.

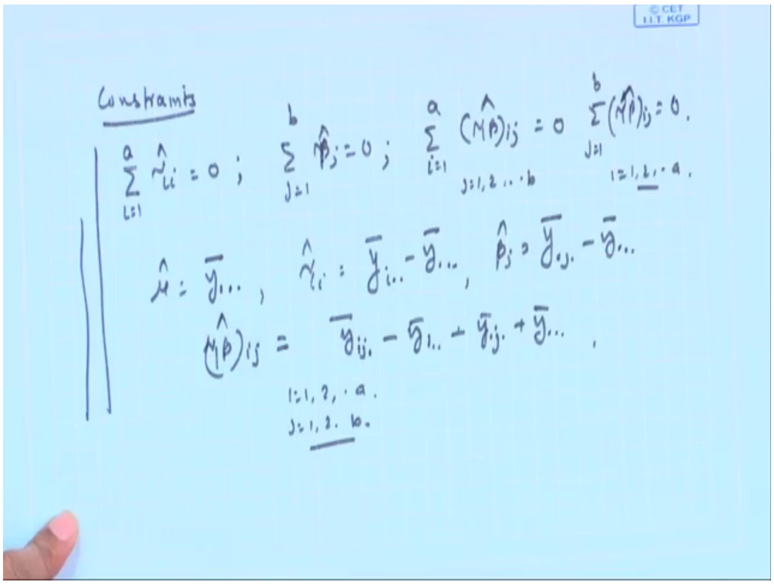
So that means, our number of parameters to be estimated, parameters to be estimated equal to 1 plus a plus b plus a b. And as a result, as a result, there are, there are how many equations? 1 plus a plus b plus a b normal equations. You have seen earlier that we found out S S E which is sum of epsilon i j k square and these will be basically sum of, triple sum of mu plus tau i plus beta j plus tau beta i j, this square and then you, we have done del S S E by del mu put 0, del S S E by del tau i equal to 0, i equal to 1, 2, a.

Similarly del S S E by beta j equal to 0, j equal to 1, 2, b and del tau beta i j sorry, del S S E by, del S S E by del tau beta i j this we put 0, and it will be i equal to 1, 2, a, j equal to 1, 2, b. So, as a result this one give you one equation, this gives a equation, this give b equations, and this give a plus a into b equations. So, total you will be having 1 plus a plus b a b normal equations now, using those normal equations and the constraint.

(Refer Slide Time: 08:16)



(Refer Slide Time: 08:18)

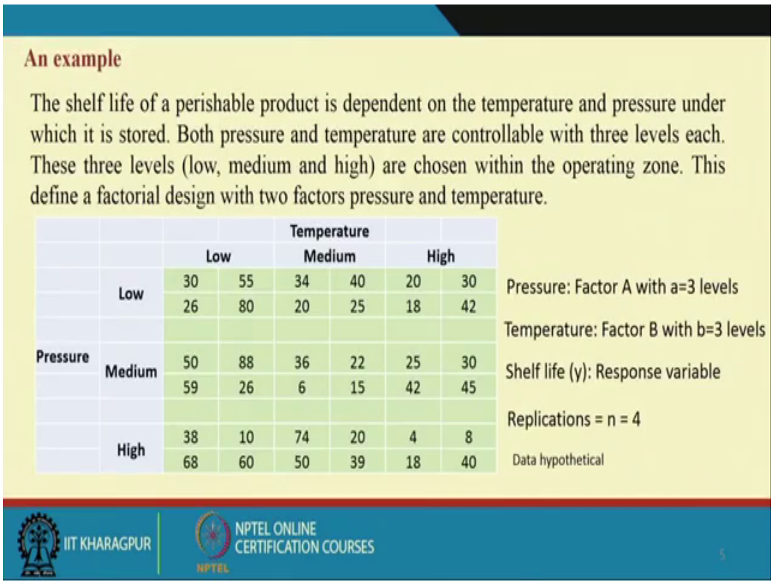


So, there are certain constraints. What are those constrains? Constraint are sum total of tau i, i equal to 1 to a. This estimate will become 0, similarly j equal to 1 to b, sum total of beta j cap. This will become 0, similarly tau j i equal to 1 to a, tau beta i j this cap, this will become 0 and sum total of j equal to 1 to b, tau beta, tau beta i j that also will become zero; obviously, everywhere you please remember i k. So, here if I put i equal to 1 to a, then j equal to here 1 to b, here i equal to 1 to a. In the totals into mu cap plus b n these we are into these plus a n into this sum of these plus these equal to this. Now we have seen that sum of this is 0, sum of this beta j is 0, and this also will become 0. So, a b and mu cap will become y triple dot. So that mean mu cap will be y triple dot by a b in Sample size.

Similarly here, these become 0 and this quantity, this quantity already known. So, this quantity you will calculate. So, in this manner if you go, you will calculate this values like this. So, what is this value? mu cap will be y bar triple do, tau i will cap will be y i dot dot bar minus y triple dot bar, then your beta j cap will be y dot j dot bar minus y dot dot dot bar, and tau beta that i j that cap will be y bar i j dot minus y i dot dot bar minus y dot j dot bar plus y triple dot bar. So, i equal to 1, 2, a, j equal to 1, 2, b,

This is what is the parameter estimation using Regression approach. So, as I told you in the first we started with this, I told this one that y i i dot bar minus this. The same thing, this, these equations, this equation what you have written here, these equations using the Regression approach we are also getting the same equations ok.

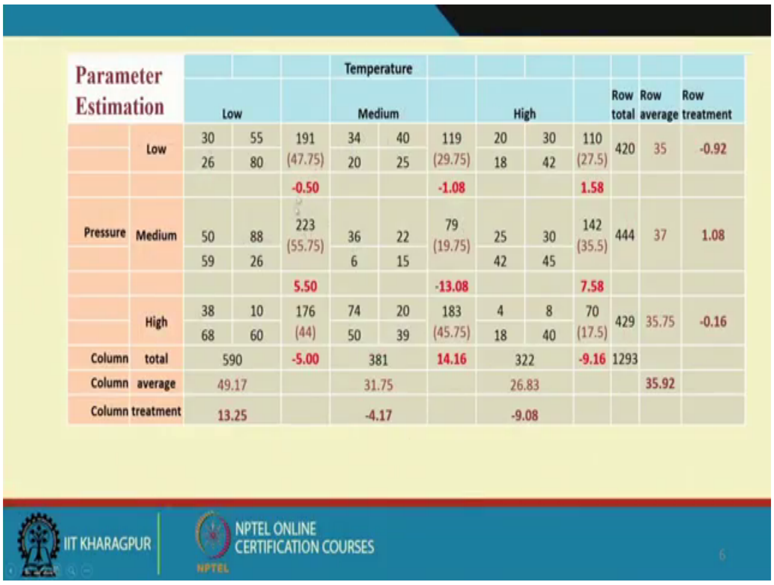
(Refer Slide Time: 11:30)



Now, I will show you one example and how do you compute all those parameters as well as their errors. So, this is what is the example which we have discussed earlier. Here basically, shelf life is a response variable pressure with 3 levels, temperature with 3 levels, contour level variables, 4 replicates in each experimental setting and then these are the data, hypothetical data we optic.

Now, using this data set and, and the formula what we have discussed here in Regression approach, this formula using this formulas with the data set, will be, will be in a position to compute all the parameters for this 2 factor Factorial experiments.

(Refer Slide Time: 12:13)



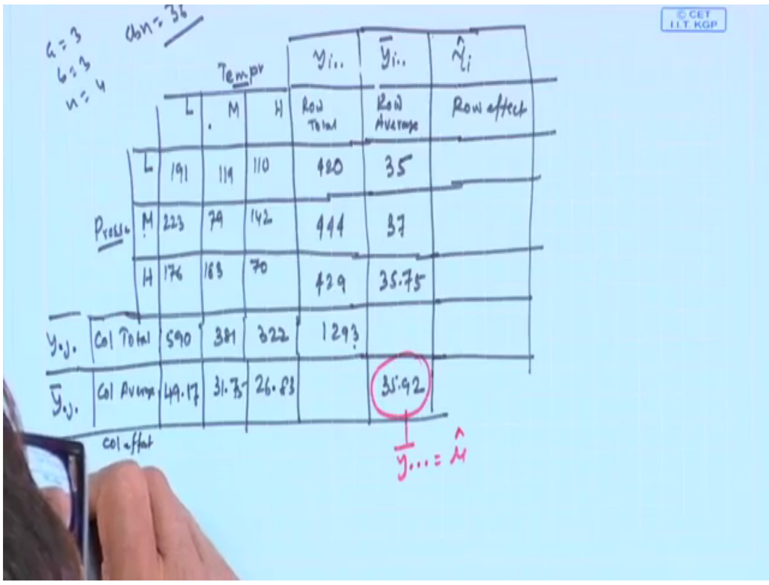
Parameter Estimation					Temperature								
		Low			Medium			High			Row total	Row average	Row treatment
	Low	30	55	191	34	40	119	20	30	110	420	35	-0.92
		26	80	(47.75)	20	25	(29.75)	18	42	(27.5)			
				-0.50			-1.08			1.58			
Pressure	Medium	50	88	223 (55.75)	36	22	79 (19.75)	25	30	142 (35.5)	444	37	1.08
		59	26		6	15		42	45				
				5.50			-13.08			7.58			
	High	38	10	176	74	20	183	4	8	70	429	35.75	-0.16
		68	60	(44)	50	39	(45.75)	18	40	(17.5)			
Column	total	590		-5.00	381		14.16	322		-9.16	1293		
Column	average	49.17			31.75			26.83				35.92	
Column	treatment	13.25			-4.17			-9.08					

Now, see what happen here. This rod, the experimental data you know. What we have done here, we have created, row total, row average, row treatment column total, column average, column treatment. So, row total you see, ultimately you see the data set here. Data set here row with respect to low with respect to medium with respect to high, column with respect to low respect to medium respect to high for temperature, then row low row total for low will be total of 1, 2, 3, 4 into 3, 12 observations.

Similarly column total will be for another 12 observations. So, we have computed here 30 plus 55 plus 26 plus 80, this is what is the cell total which is 191, second cell total is 119, third cell total is 110, you sum up 191, 119 and 131, 110, what you will get? You will get 420. So, that means, this 420 is sum of all the values related to pressure low, similarly second one is sum of all the values related to pressure medium, all the values related to pressure high. So, these are all row total. Now row total 420 divided by 12 because there are 12 observation will give you 35.

So, then 35 is the row average for n. So, that mean what happened, I will, I will come back later on. Then similarly you find out the column total, column total will be this, this cell total, this cell total, plus this cell total. So 191 plus 223 plus 176 this will be 590. Similarly here 119 plus 79 plus 183 this is 381, and then 110, 142, and 70, 322.

(Refer Slide Time: 14:34)



So, essentially what happen, essentially what happen? You have pressure low, medium, and high, this is pressure. Then we have temperature also, temperature is our, low, then our medium, then high, so this is temperature. Now, what I said that we are computing, we are computing, here we are computing row total. For the first row, for the first row this one is 420, for the second row this is 444, for the third row 429. Similarly we got column total, column total is here we got 590 then 381 then 322. Then what is the grand total? Grand total is sum of row total or sum of column total this is 1293.

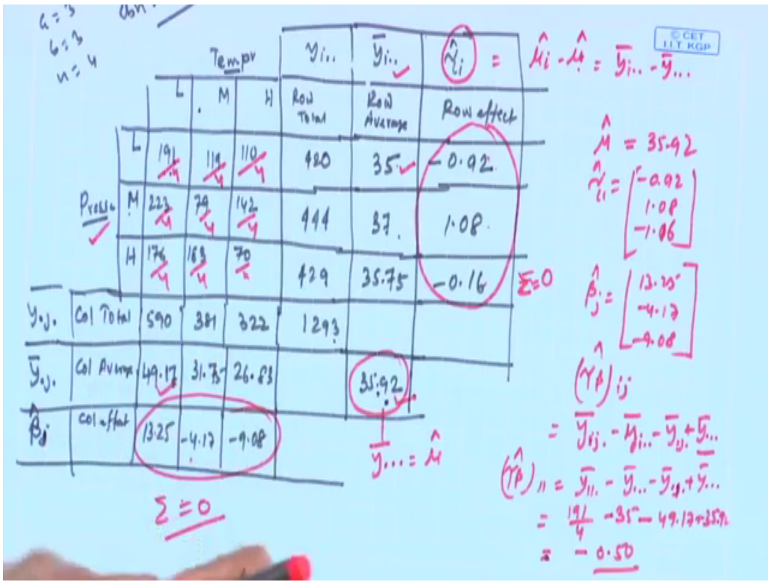
So, in addition we have computed cell total, means there are 4 observations, one the total of that 4, total of these 4, total of these 4 like these. So, cell totals are, here it is 191, then here it is 119, and here it is 110, then similarly here second one, cell total is 223, 2, 2, 3 then your 79, then your 142, and third one is 176 then 183 then your 70. So, 191, 223 plus 176, this 3 plus 1, 4, plus 6, 10 this total gives you this, these total gives you this, these total gives you this, these total will give you these, like this.

So, first you calculate row total, this row total is y i double dot, that is what is row total similarly, column total is nothing but, y dot j dot this is our column total. Then what I am saying, you find out row average y bar i double dot. So, this is your row average. So, how do compute the row average? Similarly find out row, column average these equal to column average. So, how do you find out? So, row total divided by number of observation will be 12.

So, this will give you 420 divided by 12 will give you 35, 444 by 12 will give you 37, then your 429 by 35 will 12, will give you 35.75.

Similarly, column average you will find out, 590 by 12 is 49.7, then 31.75, then your 26.83, 26.83. Then you find out the grand average, grand average means 1293 divided by 36 because you have a equal to 3, b equal to 3, n equal to 4. So, a b n equal to 36 all total is to this, divided by 36 this will give you 35.92. So, from parameter estimation point of view, this one is your y triple dot bar, which is an estimate of mu cap. Now what happened, we will find we try to find out Row effect, Row effect, Row effect means what? That is basically tau i cap ok.

(Refer Slide Time: 19:47)



So, similarly find out the column effect, this is nothing but beta j dot, sorry beta j cap. So, your tau i cap will be what? That is mu i cap minus mu cap, mu i cap minus mu cap, it is nothing but mu i cap is this one, minus mu cap is this one. So, this is basically y triple, double dot bar minus y triple dot bar. So, in that sense what will happen? 35 minus 35.92. So, this will be minus 0.92, similarly 37 minus this will be 1.08, similarly then the other one will be 0.16. So, these are the treatment effect for pressure.

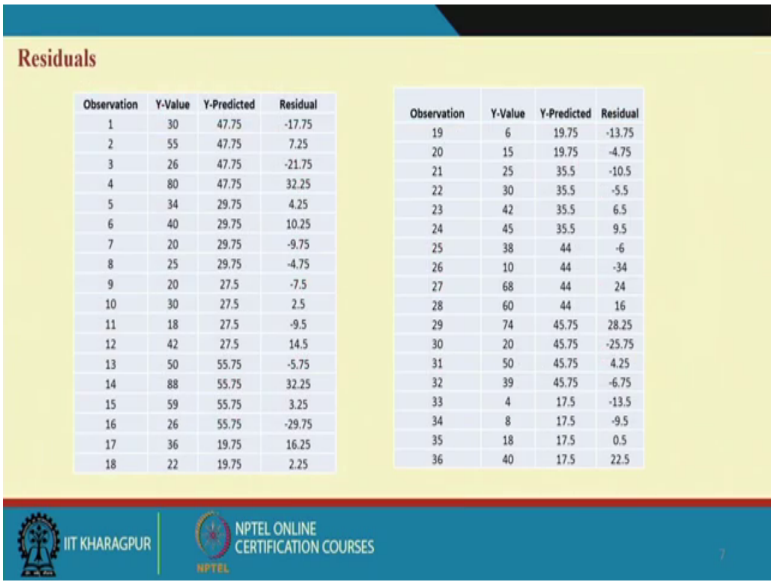
In the same manner, if you subtract the column average from the grand average, what you are getting? Here you are getting 13.25, here you are getting minus 4.17, and here you are getting minus 9.08. These are nothing but, these are nothing but the beta j cap, that is temperature effect, interestingly, if you sum up these tau i, this plus this plus this, sum will give you, give you equal to 0. Here also if you sum up, this will give you 0, so that is the con, restriction.

So, from parameter estimation point of view mu cap is known now, that is your 35.92 then tau i cap 4 values are known, a 3 values are known, minus 0.92, 1.08, minus1.16 so this is what is tau i. And beta j cap point of view you have 13.25, minus 4.17, minus 9.08. What more you want? You want tau beta cap i j. So, what is this computation? this will be y i j dot bar minus y i triple double dot bar minus y dot j dot bar plus y triple dot bar, it indicates that, what is y i j bar, you have to, how many observations are there in each of the cell? 4 observation, this is the total divided by 4, this total divided by 4, this total divided by 4, so every total divided by 4 will give you, give you the cell average.

Now, this minus row average minus column average plus grand average will give you, that mean, tau beta cap 1 1, then it will be y 1 1 dot bar minus y 1 dot dot bar minus y 1 dot bar plus y dot dot dot bar. So, this will be 191 by 4 minus 35 minus 49.17 plus 35.92. This will give you a value called minus 0.50; in the same manner you compute other values. Now see the slide, so we have these are the treatment of a row treatment pressure effect, column treatment temperature effect, and these are the cell treatment effect, interaction effects, interaction effects.

So, there will be how many interaction effects? 3 cross 3 9, 1, 2, 3, 4, 5, 6, 7, 8, 9. Now you sum up across rows, these plus these plus this will be 0, across column these plus these plus this will be 0, these are the constraints are satisfied. So, this is what is your parameter estimation in 2 factor factorial design with different levels ok.

(Refer Slide Time: 24:21)

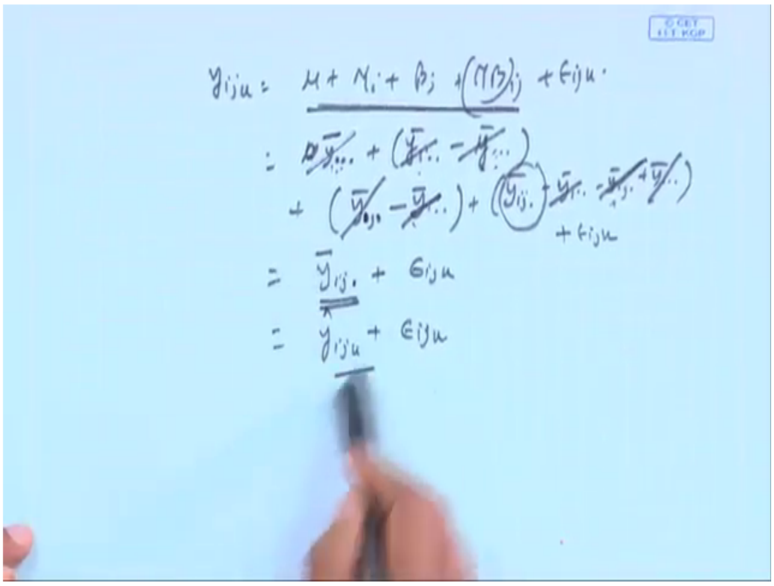


Observation	Y-Value	Y-Predicted	Residual
1	30	47.75	-17.75
2	55	47.75	7.25
3	26	47.75	-21.75
4	80	47.75	32.25
5	34	29.75	4.25
6	40	29.75	10.25
7	20	29.75	-9.75
8	25	29.75	-4.75
9	20	27.5	-7.5
10	30	27.5	2.5
11	18	27.5	-9.5
12	42	27.5	14.5
13	50	55.75	-5.75
14	88	55.75	32.25
15	59	55.75	3.25
16	26	55.75	-29.75
17	36	19.75	16.25
18	22	19.75	2.25

Observation	Y-Value	Y-Predicted	Residual
19	6	19.75	-13.75
20	15	19.75	-4.75
21	25	35.5	-10.5
22	30	35.5	-5.5
23	42	35.5	6.5
24	45	35.5	9.5
25	38	44	-6
26	10	44	-34
27	68	44	24
28	60	44	16
29	74	45.75	28.25
30	20	45.75	-25.75
31	50	45.75	4.25
32	39	45.75	-6.75
33	4	17.5	-13.5
34	8	17.5	-9.5
35	18	17.5	0.5
36	40	17.5	22.5

Now, what will be your predicted value? Predicted value will be, you just see, predicted value will be nothing but the cell average.

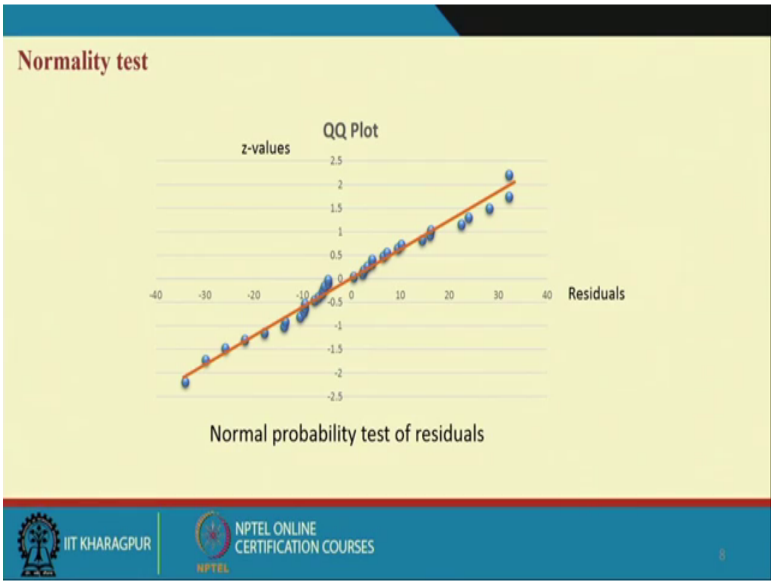
(Refer Slide Time: 24:41).



Cell average will be the predicted value, because if you, if you see that the formula. So, y i j k equal to mu plus tau i plus beta j your tau beta i j plus epsilon i j k, this is what is the predicted value. Now if you put this suppose, y triple dot bar plus this one y i double dot bar minus y triple dot bar plus y dot j dot bar minus y triple dot bar plus you put, y i j dot bar minus y i dot dot bar minus y dot j dot bar plus i dot dot dot bar, this one and plus error plus epsilon i j k. So, this one, these will be cancelled out, these and this will be cancelled out, these and this will be cancelled out, cancelled out, one, y i dot bar these and these and this, these and these will be cancelled out, sorry, this is y dot j dot bar so that means, y dot j dot bar, this will be cancelled out, and this is cancelled with this, but this one will be cancelled out with this. So, it will be remaining like these.

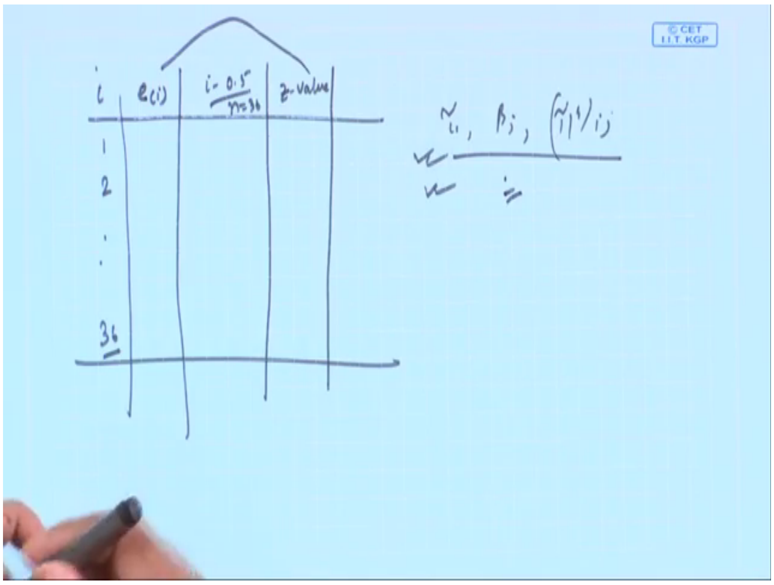
So, y i j dot bar plus epsilon i j k, and this is what is our y i j k predicted value plus error y i j k. So, this is nothing but the cell average. So, all cell averages are put here. So, when observation, this are there will be there will be 36 observations, and you are getting y value and y predictable or residuals you are getting, this is what is the computation of residuals ok.

(Refer Slide Time: 26:48)



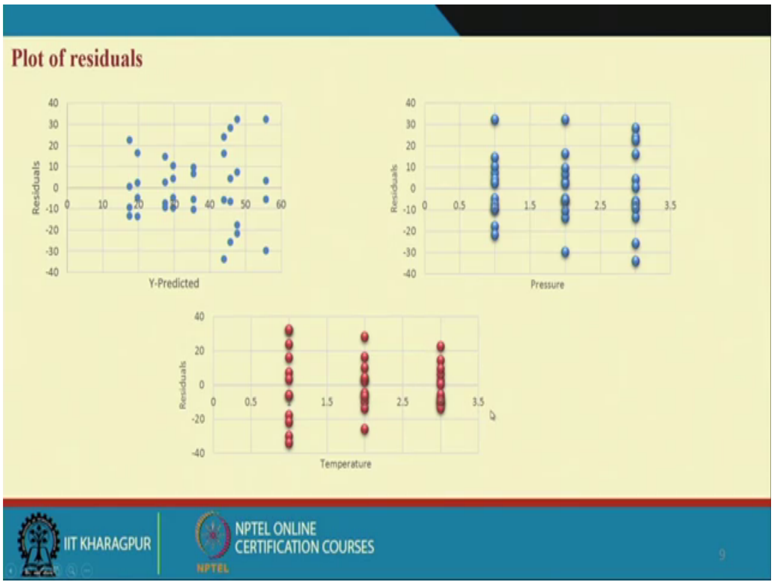
Now, next is our adequacy test. In adequacy test, we will show you the residual plots first one is the QQ plot quantile quantile plot.

(Refer Slide Time: 26:59)



You all know, what you do basically, you write down the errors from in ascending order so there will be 36 errors. So, you write down from ascending order mean suppose error if I say, error if I write the residuals is, e i and ascending order these and then you 36 is this one, and then and you find out i minus 0 .5 by n, here n equal to 36, then you will be getting cumulative probability and then find out the corresponding z value assuming that, Errors are normal distributed, now plot these 2. So, z value and residual. So, it is this is what is it is plot here, and we found that, there is a fit lines straight line fitting to the origin. So, it is basically it is normally a data normally distributed.

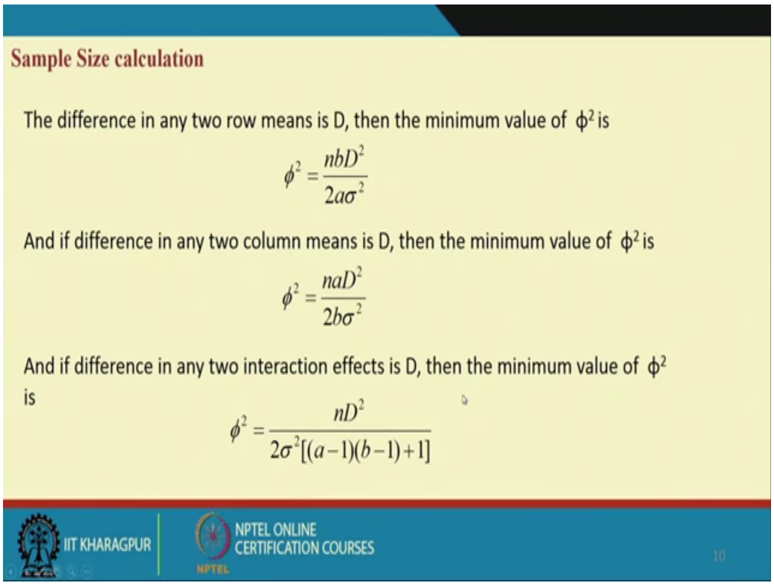
(Refer Slide Time: 27:55)



Similarly, plot of residuals you see the residual verses y predicted. So, you have the y predicted value, you have the residual values. So, you plot residual y predicted if there is no pattern. So, it gives us many things for example, that basically talk about that, that data independent and those kind of things are this is, this one is done ok.

So, now here a, if I go for residual with individual that x values so, it will, it will give you the across x whether the residual is showing constraint variants or not. So, from the plot here we cannot say that there is a there is a change in variability operative aqua pressure, but in case of aqua temperature there is little change, but may be that change is not that much significant change may be, but you require some other kind of test to check whether the, there is the variation violation here or not, but from the plot there is slight difference.

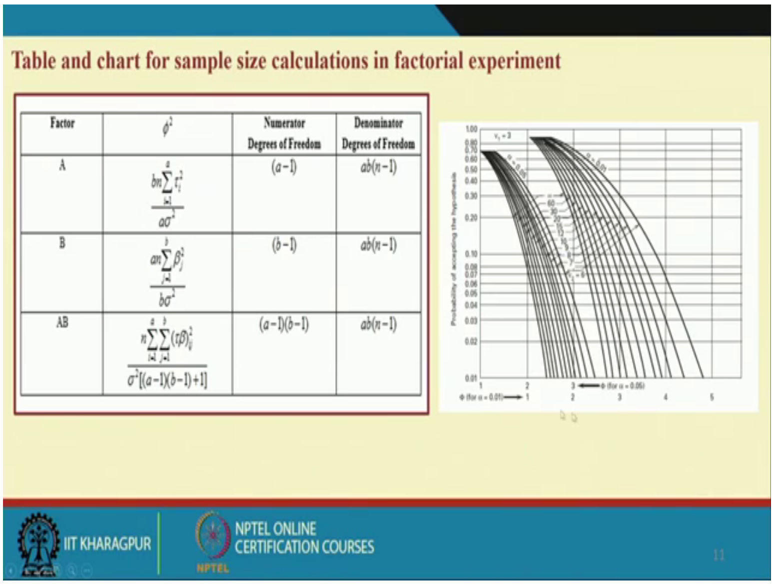
(Refer Slide Time: 29:08)



Then another important thing is that Sample Size calculation. So, in like 1 way ANOVA we have, we have a b nu the sample size calculation, when the 2 issues, when if you know the tau i is, tau i beta j and tau beta, tau beta i j, this values are known. Then you will use one kind of formula, if the values are not known then you have to and all and an and the difference between 2 treatments or 2 means are given either row means or column means or the interaction effects, the difference given, then you will using another kind of formula. First, I will, let us see that if we consider that the difference between any 2 row means is d then the phi square will be computed like this, what is this phi square? n b D square by 2 a sigma square.

Now, if the difference between in column means is D, then it will be n a D square by 2 b sigma square, and if it is any 2 treatment difference of any 2 interaction effect is D then the minimum value of phi square will be, n D square by 2 sigma square into a minus 1 b minus 1 plus 1. This is what is the, what is the results obtain from statistical theory. So, then once we know phi square, we know the phi, then we know the what is the sample, your Degrees of Freedom, numerator denominator Degrees of Freedom, and then using the o c curve we will be able to find out the beta value.

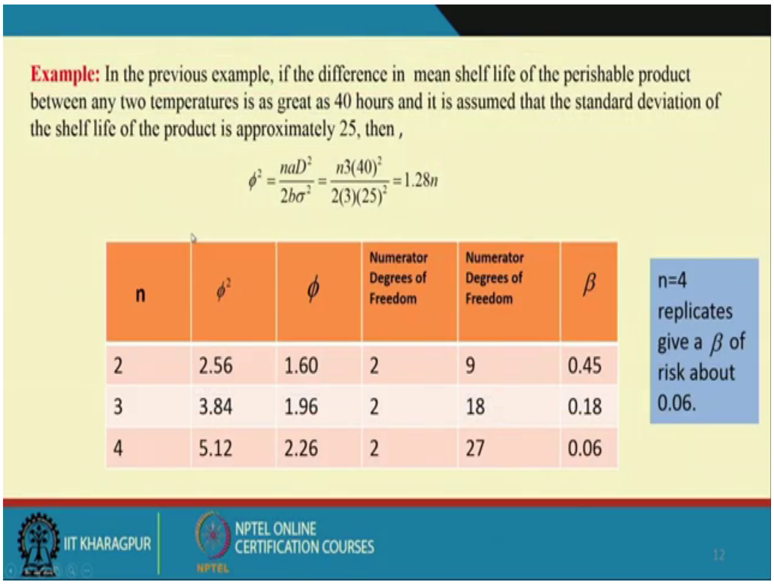
(Refer Slide Time: 30:52)



Factor	ϕ^2	Numerator Degrees of Freedom	Denominator Degrees of Freedom
A	$\frac{bn\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$	$(a-1)$	$ab(n-1)$
B	$\frac{an\sum_{j=1}^{b}\beta_j^2}{b\sigma^2}$	$(b-1)$	$ab(n-1)$
AB	$\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{\sigma^2[(a-1)(b-1)+1]}$	$(a-1)(b-1)$	$ab(n-1)$

For example, if you know, if you choose a particular alpha and know the beta, phi value, and then knowing the Degrees of Freedom nu 1 and nu 2, you will be knowing that, what is the probability of accepting a false hypothesis that is beta that can be computed. So, in this table we have given the, suppose the tau y beta j and tau beta j this value, these population values parameter values are known in that case you use this formulation and use the chart also, but it is hardly, it is hardly known. Because, it is difficult to know from the data what we know that we know the estimate of these and then we will choose a particular difference.

(Refer Slide Time: 31:32)



Now, in this example I will show you only the, suppose in, if you go back to the parameter estimates, you see the treatment effects, temperature treatment effects are very large, and also you have seen in the, in the 2 peg analysis of variance the temperature effect is significant. So, we are now, we are now you finding out the, that, this one the Sample Size, we, in example considering the temperature level, and it is levels. So, suppose 40 hour that mean that if the difference between mean life of a perishable product for between 2 temperatures is this 40 hours is the acceptable one, and if we assume that the shelf standard deviation is 25. So, this is not from the, that data we have computed the standard deviation, we have assumed that it is 25, let it be it may not be.

So, if this is the situation, then phi square equal to this, and then putting a b and D value sigma square value, it will be 1.28 n. And then for different values of n, you compute the phi square value and then phi value, and you know the Degrees of Freedom all those, the all cases, basically that degrees of numerator Degrees of Freedom, and denominator Degrees of Freedom.

So, all those case you will be computing, once you know the a n values, and this values are b values are known. So, now, if numerator Degrees of Freedom is 2, denominator, this one is numerator, another denominator, Degrees of Freedom is 9, that is nu 1 and nu 2, and then using your phi, a phi, is 1.60, and then you have to you have to use appropriate one, here 2 3 and different values of nu 2 is given. So, this plot will not help you, you require nu 1 equal to

2, and nu 2 equal to 9, that plot, that plot then, you will find out, you will find out that the beta value will be 0.45.

Similar thing we have given in 1 way analysis of variants, when you talk about this Sample Size calculation. So, you require 2 Degrees of Freedom that is nu 1 and nu 2. And you require phi, and then you will be able to find out the beta value, this beta value is computed here. So, if we consider that beta 1, power should be 90 per cent or more, or beta values at max may be 0.1, in that case you require 4 number of 4 sample size, sample size will be 4, then you will be getting the desired level of beta value. Now in this example, we have chosen 4 replication is 4, in this example replication is 4.

So, this is what is your sample size calculation, only 1 mistake here is that, we talk about both case numerator, one is this numerator this is denominator, or nu 1 and nu 2, first Degrees of Freedom and second Degrees of Freedom you just write down, and second thing is that this is the sigma square we have assumed that 25, and it may not be 25. So, it actual values are known then this calculation will be different. And you have to, once you know the nu 1 and nu 2, numerator and denominator Degrees of Freedom, then and you know the phi, then you have to use, you have to find out the alpha value, alpha may, will be usually 0.05 or 0.01, if you took alpha actually 0.05 and phi value is given like phi values are known, and if the nu 1 and nu 2 matches, then from there you will be able to find out the beta values ok.

So, this is what is our Sample Size calculation and I hope that you will be able to find out the sample sizes. So, very quickly I will just recap, recall all the things what is thought in this lecture now, that we first thought, told you the estimation part, and estimation means basically the parameter estimation. Now what are those parameters? mu, tau I, beta j, tau beta i j, and I have given you, how from this experimental data you will compute all those things, and then the predicted value of every observation or other way I can say the what will be the fitted values, and then from fitted values when you subtract it from the original value you will get the residuals, this is what first thing we have given to you.

And you have compute, with an example we have shown, how this computation will be done. And also what we have said that the, these formula that calculation comb, estimation of mu, tau I, beta j, tau beta j, all those thing using the regression approach by that minimising the sum square errors.

How it is to be computed that is also shown. Please remember, a number of parameters to be estimated is equal to the number of normal equations or, number of normal equations you will be using, will be equal to the number of parameters to be estimated. There will be lot of constraint depending on the, what kind of factor model, it is 2 factor factorial model or 3 or more, but the procedure remains same.

There will be lot of constraint which is to be satisfied, putting all those things you will be able to estimate the parameters. After you estimate the parameter, parameters then the sample, model adequacy test in terms of normality, in terms of mostcuracity, in terms of independents, or uncalculated errors all those things you require to be, require to be satisfied, require to be test, and you see that they are satisfied or not. And finally, sample size is very important, hence so, what I want that, you use the o c curve concept what you have discussed here for sample size computation.

So, thank you very much.