

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

Lecture – 15
ANOVA – Estimation of Model Parameters and Adequacy Test

Welcome, we continue ANOVA in this lecture we will discuss estimation of model parameters and test of assumptions test of assumptions.

(Refer Slide Time: 00:24)

© CET
I.I.T. KGP

ANOVA: Estimation of Model parameters
& test of assumptions

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Estimate: (1) Grand mean
 (2) Treatment effects
 (3) Residuals

	1	2	...	j	...	n	
1							
2							
...							
i				y_{ij}			$\bar{y}_{i.}$
...							
a							$\bar{y}_{..}$

$$\hat{\epsilon}_{ij} = \bar{y}_{.j} - \hat{\mu}$$

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\alpha}_i = \bar{y}_{i.} - \hat{\mu}$$

$$= \bar{y}_{i.} - \bar{y}_{..}$$

So, let us see the contents. So, first estimation of model parameters followed by confidence interval for model parameters.

(Refer Slide Time: 00:44)

Contents

- Estimation of model parameters
- Confidence interval of model parameters
- Simultaneous confidence interval of model parameters
- Test of assumptions
- Normality assumption
- Homoscedasticity assumption
- References

Source: This lecture is prepared based on "Design and Analysis of Experiments" by D C Montgomery, Wiley, 8th Edition

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

This estimation of model parameters followed by confidence interval does mean that (Refer Time: 00:55) interval, estimate when as there are many comparisons possible here.

So, simultaneous confidence interval another concept which is important, then the different tests and they are states test, test of assumptions. So, different assumptions and their tests like normality and homoscedasticity assumption and actually this again taken from the same book design analysis of experiment by Montgomery.

(Refer Slide Time: 01:29)

Estimation of the Model Parameter

Estimator for the parameters in the single-factor model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Then, overall mean, $\hat{\mu} = \bar{y}_{..}$

And treatment effects, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \quad i = 1, 2, 3, \dots, a$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Let us now see the model. A model is y_{ij} ANOVA model, $\mu + \tau_i + \epsilon_{ij}$. So, what are the model parameters here μ and τ_i in addition you require to compute also errors, which is basically after computation there is known residuals. So, what we want to estimate? We want to estimate; we want to make a grand mean, we want to estimate treatment effect, treatment effects, we want to estimate residuals.

So, already I have given you already because say if you recall the general data 1, 2 like I then it is A, and this side 1, 2 dot dot dot j, and then n; you will find out that somewhere y_{ij} is there and the corresponding \bar{y}_i and the grand average is \bar{y} . I have given you earlier. So, these μ estimate grand mean estimate is grand average from the data point.

And then individual mean estimate is individual label means, then τ_i estimate is μ_i minus μ minus μ estimate minus μ estimate which is basically \bar{y}_i minus \bar{y} , then another estimate will be residual. So, treatment third one is residual. Residual means that is basically if I say residual ϵ_{ij} cap, then this is nothing, but observed value minus predicted value I am writing it again because there is a there is some ϵ_{ij} cap is y_{ij} cap minus \hat{y}_{ij} .

(Refer Slide Time: 03:54)

$$\rightarrow \hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_i$$

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i = \mu + (\mu_i - \mu) = \mu_i = \bar{y}_i$$

So, these 2 this is basically. So, what is y_{ij} ? y_{ij} is $\mu + \tau_i + \epsilon_{ij}$ now what will be then y_{ij} cap this will be μ cap plus τ_i cap. So, again τ_i is basically that is μ_i cap plus τ_i is μ_i cap minus μ cap. So, this is nothing, but μ_i cap which

is nothing, but \bar{y}_i so; that means, the predicted value is the individual level average, then residual is individual observations minus \bar{y}_i then this will be y_{ij} minus \bar{y}_i .

Now this is estimate of residuals; this is what is known as estimation. Now we are primarily interested to on two things; one is that estimate of μ_i .

(Refer Slide Time: 05:19)

Estimate of μ_i : $\hat{\mu}_i \leftarrow \text{point estimate } \hat{\mu}_i = \bar{y}_i.$
 Estimate of σ^2 : $\hat{\sigma}^2$
 Interval estimate: $100(1-\alpha)\% \text{ CI}$
 $\bar{y}_i \sim N(\mu_i, \frac{\sigma^2}{n})$
 $y_{ij} \sim N(\mu_i, \sigma^2)$
 $\mu_i <$
 $\hat{\sigma}^2 = \frac{MSE}{n-1}$
 $-t_{N-a} \le \frac{\bar{y}_i - E(\bar{y}_i)}{\sqrt{V(\bar{y}_i)}} < +t_{N-a}.$

Similarly estimate of τ_i which is μ_i cap then your τ_i cap. So, then what do you require? You require to know the point estimate which is which is point estimate which is μ_i cap point estimate is \bar{y}_i dot and you also require to know interval estimate; that means, 100 into 1 minus alpha percent confidence interval.

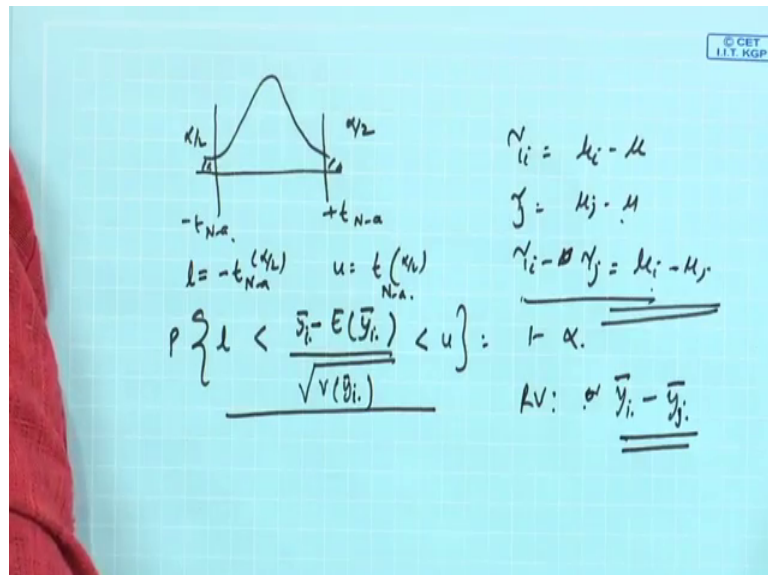
So, in order to know the interval estimate you require to know the distribution of \bar{y}_i cap what will be the distribution of \bar{y}_i cap? Here you have to you have to know hint what the time being we can say that this is basically μ_i and σ^2 by n . If y_{ij} is distributed and μ_i and σ^2 , then \bar{y}_i cap \bar{y}_i dot bar this will b mean value will be same mean value and variance part will be σ^2 by number of observations σ^2 by n .

If you know this then you can go for interval estimate now here what if you what interval estimate you want? You want interval estimate of this. So, as a result you require to know the σ^2 value. σ^2 that estimate is nothing, but here MSE by that is

MSE sigma square will be MSE or other way you can we can write like this m s suffix E you please keep in mind MSE is an estimate of the variability of y across different levels.

So, I hope that you understand it, now how to get it. So, if I create a my random variable y_i dot cap minus its expected value divided by square root of variance, then this will follow which distribution. Here now z and sigma is not known estimate t distribution. So, t distribution this will be minus t this will be plus t, what will be the degrees of freedom 40, what is available after calculation of the parameters like variance; that means, your availability will be N minus a here.

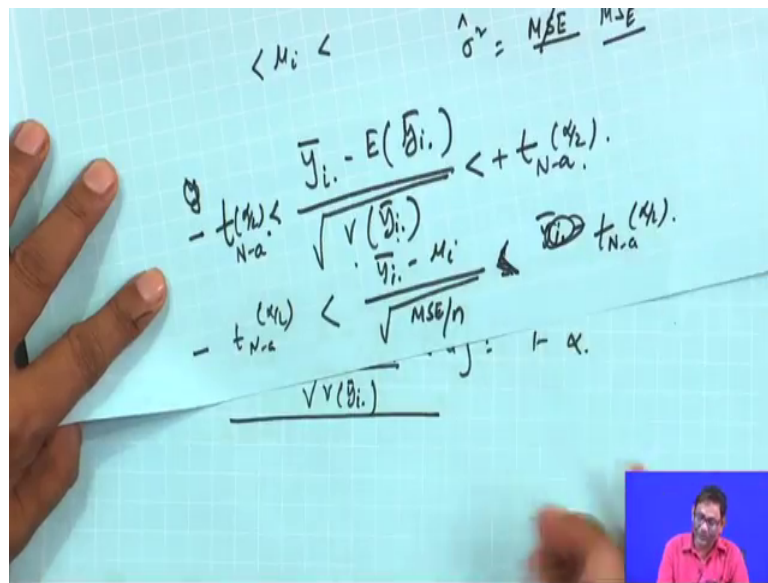
(Refer Slide Time: 08:45)



So, what do you do then your case is something like this t distribution this side plus t N minus a this is minus t N minus a.

Suppose you consider alpha by 2 this side and alpha by 2 to this side, then your l equal to minus t N minus a alpha by 2, and u equal to t alpha by 2 N minus a. So, probability 1 less than equal to that y bar minus expected value of y i dot bar minus expected value y i dot bar divided by square root of variance of y i dot bar this will be less than equal to u equal to 1 minus alpha you want this, this is what we have we have written here that 1 value is this will be alpha by 2 and this will become alpha by 2.

(Refer Slide Time: 09:48)



So, now you know this they what will happen then? That means, minus t n minus a alpha by 2 less than y i dot bar minus mu i divided by square root of MSE by n, because y i this sigma square is MSE by n less than less than or less than equal to y i dot bar minus sorry equal to plus t N minus a alpha by 2. So, the resultant quantity will be what you see resultant will be y i dot bar minus t alpha by 2 N minus a variance part mu i.

(Refer Slide Time: 10:30)

Point and Confidence Interval Estimate

Point estimate of μ_i would be, $\mu_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_i$ and confidence interval estimator for the i th treatment i.e. a $100(1-\alpha)$ percent confidence interval of μ_i is

$$\bar{y}_i - t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_i + t_{\alpha/2, N-a} \sqrt{\frac{MS_E}{n}}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

This is the confidence interval 1 into 1 minus un alpha percent confidence interval for mu i.

(Refer Slide Time: 10:47)

Confidence Interval Estimate

100(1- α) percent confidence interval on the difference in any two treatments means, say $\mu_i - \mu_j$, would be

$$\bar{y}_i - \bar{y}_j - t_{\alpha/2, N-d} \sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + t_{\alpha/2, N-d} \sqrt{\frac{2MS_E}{n}}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Now, now what will happen for the 2 difference between 2 treatments. So, we were interested to know $\mu_i - \mu_j$, because suppose what is the τ_i . τ_i is $\mu_i - \mu_j$ what is τ_j ? τ_j equal to $\mu_j - \mu_i$ then if I want $\tau_i - \tau_j$ this will be then $\mu_i - \mu_j$. So, if you want to know the difference between 2 treatment then what will happen the responds what is $\mu_i - \mu_j$ estimate, that will be $\bar{y}_i - \bar{y}_j$ this will be an our a random variable of interest and its standard deviation of the same way you will compute and you will find out this. That $\bar{y}_i - \bar{y}_j \pm t_{\alpha/2, N-d} \sqrt{\frac{2MS_E}{n}}$, less than equal to this less than equal to this is plus.

Sorry its correct $\bar{y}_i - \bar{y}_j \pm t_{\alpha/2, N-d} \sqrt{\frac{2MS_E}{n}}$ plus and this is minus $\sqrt{\frac{2MS_E}{n}}$ is coming, because you have $\sigma_i^2 + \sigma_j^2$ because of the difference between these 2. So, this formula, what will happen now?

(Refer Slide Time: 12:25)


Simultaneous Confidence Interval (SCI)

If, r such $100(1-\alpha)$ percent confidence intervals of interest, the probability that the r intervals will simultaneously be correct is at least $(1-r\alpha)$. The probability $r\alpha$ is often called the **experimentwise error rate** or overall confidence coefficient.

The SCI of i th treatment mean μ_i is

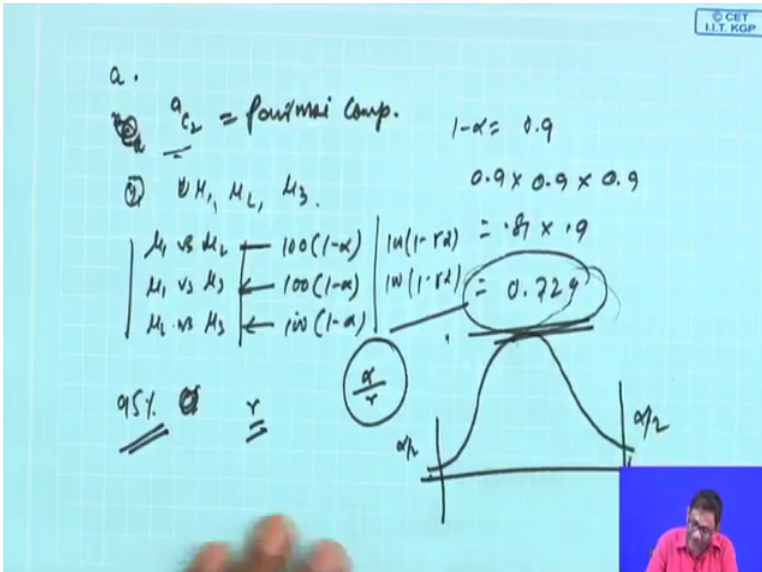
$$\bar{y}_i - t_{\alpha/2r, N-a} \sqrt{\frac{MS_E}{n}} \leq \mu_i \leq \bar{y}_i + t_{\alpha/2r, N-a} \sqrt{\frac{MS_E}{n}}$$

Bonferroni method



You are suppose you are comparing how many means? Suppose there are a means.

(Refer Slide Time: 12:30)



a.

$\alpha_{c2} = \text{pairwise comp.}$ $1-\alpha = 0.9$

① μ_1, μ_2, μ_3 $0.9 \times 0.9 \times 0.9$

μ_1 vs μ_2	$100(1-\alpha)$	$100(1-r\alpha) = .81 \times 100$
μ_1 vs μ_3	$100(1-\alpha)$	$100(1-r\alpha) = 0.729$
μ_2 vs μ_3	$100(1-\alpha)$	

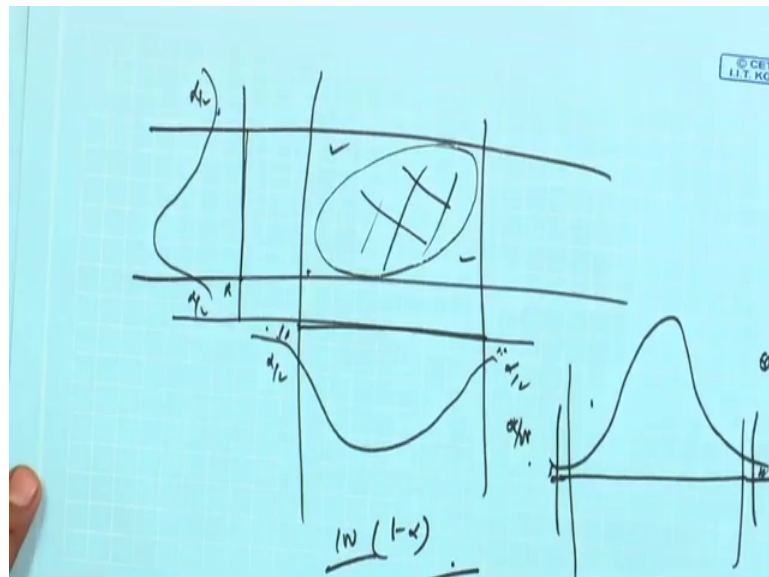
95% α γ $\frac{\alpha}{2}$

Because a level means a I means. So, if you compare twelve together in c a sorry ac 2 comparison pair wise comparison will be there pairwise comparison. So, if I have 3 that mean then y_1 bar suppose μ_1 μ_2 and μ_3 ; μ_1 versus μ_2 , μ_1 versus μ_3 μ_2 versus μ_3 ; 3 comparisons possible; now, every time if I consider here 100 into 1 minus alpha percent arbor a confident here 1 minus alpha percent confident here, also 1 minus alpha percent confidence. So, as all those things are basically, we are considering

simultaneously what will happen basically suppose if 1 minus alpha is 0.9 for 3 case then 0.9 into 0.9 into 0.9, I mean 0.9 into 0.9 that mean 0.81 into 0.9 that mean 0.729. So, that collectively what will happen your overall confidence level reduced to this, this will reduce to this one. So, what happen in order to put it that suppose you want to become 95 percent confident.

in order to put these the if you have p number of comparison or r number of comparisons, what do you do basically every time you the alpha, this alpha you do you equally divide it divide it by r. So, instead of 100 into 1 minus alpha you do 100 into 1 minus r alpha. So, you here it is 100 into 1 minus r alpha 100 into 1 minus r alpha like this. So, what do I mean then? Suppose I want this is some result t distribution you want this side alpha by 2 this side alpha by 2.

(Refer Slide Time: 14:52)



Suppose first comparison pair this, first this side and this side suppose this and this is this. So, this side alpha is coming this also alpha by 2 is coming, now the resultant will become less. So, as a result here instead you writing alpha by 2, here also alpha by 2, alpha by 2, alpha by 2, institute for 2 comparisons I am difference I am talking here instead of then what do you are basically doing you are making this portion this portion smaller.

So, you have this; so, for every individual you require the; so, everyone you now make it even smaller alpha by 2 r, this side also alpha by 2 r. So, if you make stricter confidence

then when you collectively multiply them if they are independent. So, ultimately you will result into 1 minus alpha into 100 percent see here.

So, as a result this is known as Bonferroni approach. So, this Bonferroni approach is giving this one is; so, everything is as it is only t alpha by 2 r; when you have r number of comparisons this is known as simultaneous confidence interval. This is the individual confidence interval, this is another individual confidence interval; simultaneous one maybe like this. So, I mean this portion this portion, this will be out simultaneously collectively it is a ellipse kind of thing.

So, in order to capture or take care of this portion, you are dividing by r and ultimately when you multiply. So, this means in this zone or the hyper the ellipse side level you will be able to do it, this is the formula. So, simultaneous two treatment means also similar case.

(Refer Slide Time: 17:06)

The slide is titled "Simultaneous Confidence Interval Estimate". Below the title, it states "The SCI of the difference in any two treatments means". The main content is a mathematical formula for the Bonferroni method:
$$\bar{y}_i - \bar{y}_j - t_{\alpha/2r, N-a} \sqrt{\frac{2MS_E}{n}} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + t_{\alpha/2r, N-a} \sqrt{\frac{2MS_E}{n}}$$
 Below the formula, it says "Bonferroni method". At the bottom, there are logos for IIT KHARAGPUR and NPTEL ONLINE CERTIFICATION COURSES.

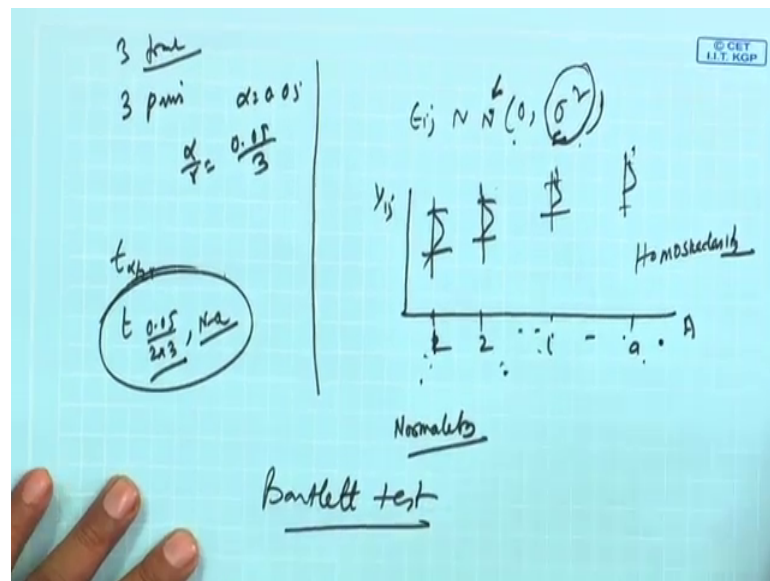
Here two different treatments, here I said that treatment difference or two different treatments first treatment and second treatment maybe two different two difference.

So, then how do you treat r? r will be number of treatment, when you are going for treatments simultaneously, when you are going for treatment different pairs. So, r will be number of pairs that is the issue. So, y i bar minus y i j and then again t by alpha by this one. Let me repeat here what happen? You are comparing finding out the confidence

interval for single mean and alpha by 2 is the level of significance, here you are saying there are r treatments you if your r different confidence level of interest, but they are occurring simultaneously you want simultaneous part.

So, alpha by r that is what you are considering, then here what happened that diff 2 difference in two treatments means, only one pair here what happened your considering all the simultaneously all the pairs. So, by r is coming.

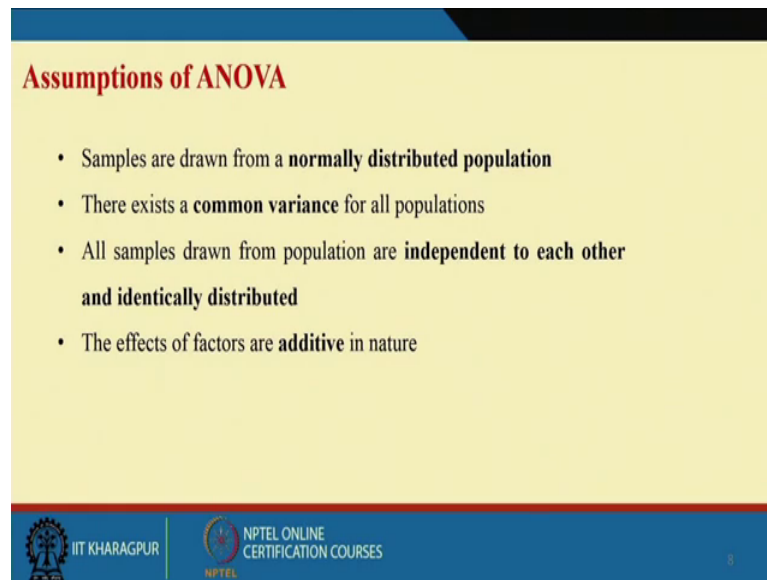
(Refer Slide Time: 18:29)



So, if I have 3 pairs or 3 treatment in the first case if I choose alpha equal to 0.05, then alpha by r will be 0.05 by 3, then when I say t alpha by 2 r then it will be nothing, but t 0.05 by 2 into 3. So, you have to add and then according and also N minus a is the degree of freedom you find out from table what is this value.

This approach is known as bonferroni approach.

(Refer Slide Time: 19:05)



Assumptions of ANOVA

- Samples are drawn from a **normally distributed population**
- There exists a **common variance** for all populations
- All samples drawn from population are **independent to each other and identically distributed**
- The effects of factors are **additive** in nature

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, this is what is confidence interval estimation. Now what happened I will tell you very quickly something an assumption of ANOVA now what are the assumptions? Say we ANOVA is normally distributed with mean 0 and variance sigma square (Refer Time: 19:33) suppose this side Y_{ij} and this side your fact A, A level one then level 2, level I level A then what will happen here if you take n observation suppose your data represent like this, here it should be like this, here may be it is like this here it may be like this.

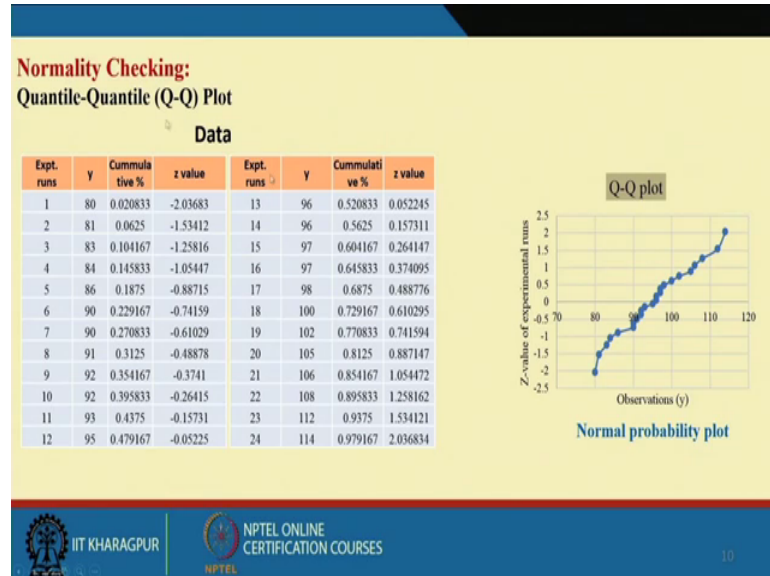
This is what is the variability part in y for different levels, this spread should be equal that is what we are saying sigma square this is known as Homoscedasticity. Now this is one of the assumptions here we have written common variance that mean the variance across different level equal. Another one I said which normally distributed also normally distributed that also you require to test. So, you also required to test normality.

So, I already told you how to do this kind of test, that be normality test and homoscedasticity test we will explain. Another assumption is that every observations are independent and they must be identical to identically distributed if they are coming from normal population all are coming from normal population with that particular parameters of normal distribution.

If they are coming for some other population it is true for everybody, but all are independent and identically distributed and the effects of factors are additive in nature

means if you increase factor level from 1 to 2, the effect in on response will be same if you increase from 2 to 3, that additive effects what we are talking about.

(Refer Slide Time: 21:30)



So, we have already take seen this quantile. Quantile plot in normal distribution lecture, I have showed you how to develop the quantile quantile plot, an that plot I was actually shown to you. So, quantile qauntile plot means you find out cumulative probability then corresponding z value, then observed value what are observed value, but what are z value you plot you will get a straight line like plot and if it is a straight line then it is normally distributed.

If it is deviating from the straight line, significantly then it is not normal.

(Refer Slide Time: 22:08)

Homoskedasticity checking:

Bartlett's test

- Bartlett's test is used to check the equality of variance among treatments
- Hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

$$H_1 : \text{above not true for at least one } \sigma_i^2$$
- Sampling distribution of the relevant statistic is closely approximated by the chi-square distribution with $(a-1)$ degrees of freedom when the a random samples are from independent normal populations.

$\chi_0^2 = 2.3026 \frac{q}{c}$

where



$q = (N-a) \log_{10} S_p^2 - \sum_{i=1}^a (n_i-1) \log_{10} S_i^2$
 $c = 1 + \frac{1}{3(a-1)} \left(\sum_{i=1}^a (n_i-1)^{-1} - (N-a)^{-1} \right)$
 $S_p^2 = \frac{\sum_{i=1}^a (n_i-1) S_i^2}{(N-a)}$

→

$\chi_0^2 > \chi_{a(a-1)}^2$

↓

Reject H_0
- Bartlett's test is very sensitive to the normality assumption.


IIT KHARAGPUR

NPTEL ONLINE
CERTIFICATION COURSES

And for equality of co variance or homoscedasticity purpose, you require to do Bartlett test. So, what is Bartlett test? In Bartlett test we assume that null hypothesis is the variances are equal. Sigma 1 square equal to sigma 2 square equal to sigma s square; alternate hypothesis is at least one of the variance is different from the others. So, then what you require, you require a appropriate statistics that will ultimately talks about the null hypothesis case under null hypothesis, some appropriate statistics and that statistics is chi square what is what is basically developed by Bartlett.

So, what do you Bartlett has done that chi square is 2.3026 onto q by c, this is the quantity which follows chi square distribution with a minus 1 degrees of freedom. So, if you ask me how what it is derived is not your job, it is the statistician who has made it Bartlett has done it. So, these quantity now q and c formula is given.

So, my request to all of you go to the chapter 3 of Montogomery and find out this formula and this is the formula for q and c because other values are known if data is available to you. And then these quantity follows chi square distribution with a with a minus 1 degrees of freedom, find out the threshold r alpha and accordingly you do the test the way normality test is a sorry the way hypothesis test is done thank you very much.

(Refer Slide Time: 24:04)

Bartlett's test (Contd.)
 Conduct Bartlett's test for the following data:

Factor Ground clutter	Observations (Replications)								Total	Average
Low (1)	90	86	96	84	100	92	92	81	721	90.125
Medium (2)	102	97	106	90	105	97	96	80	773	96.625
High (3)	114	93	112	91	108	95	98	83	794	99.25
									2288	95.33333

Hypotheses are:
 $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$
 $H_1 : \text{above not true for at least one } \sigma_i^2$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

But we have some more points to tell some examples like the Bartlett test for the first example, reader scope example here what happened using Bartlett test what you have found out that.

(Refer Slide Time: 24:15)

Bartlett's test (Contd.)
 From the data after computing, we get

$$S_1^2 = 39.55357; S_2^2 = 72.55357; S_3^2 = 121.0714$$

Then,

$$S_p^2 = \frac{7 \times 39.55357 + 7 \times 72.55357 + 7 \times 121.0714}{21} = 77.72618$$

$$q = 21 \times \log_{10}(77.72618) - 7 \times [\log_{10}(39.55357) + \log_{10}(72.55357) + \log_{10}(121.0714)] = 0.91$$

$$c = 1 + \frac{1}{3 \times 2} \left(\frac{3}{7} - \frac{1}{21} \right) = 1.0635$$

And, the test statistic is $\chi_0^2 = 2.3026 \times \frac{0.91}{1.0635} = 1.97$

From table, $\chi_{0.05,2}^2 = 5.99$

Therefore, Null hypothesis cannot be rejected.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Chi square computed one is 1.97, but tabulated one is 5.99 so; that means, the tabulated one is much more than the computed one. So, you cannot reject null hypothesis. So, that means the variability across the levels are not different, there is equality of variances across different levels of ground clutter.

(Refer Slide Time: 24:45)

Bartlett's test – Example 2 (Power data)

Apply Bartlett's test to the etch rate data.

Power (W)	Observations					Total	Averages
	1	2	3	4	5		
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707

Calculations

$S_1^2 = 400.7; S_2^2 = 280.3; S_3^2 = 421.3; \text{ and } S_4^2 = 232.5$

$$S_p^2 = \frac{4(400.7) + 4(280.3) + 4(421.3) + 4(232.5)}{16} = 333.7$$

$$q = 16 \log_{10}(333.7) - 4[\log_{10}400.7 + \log_{10}280.3 + \log_{10}421.3 + \log_{10}232.5] = 0.21$$

$$c = 1 + \frac{1}{3(3)} \left(\frac{4}{4} - \frac{1}{16} \right) = 1.10$$

$$\chi_0^2 = 2.3026 \frac{(0.21)}{(1.10)} = 0.43$$

Summary

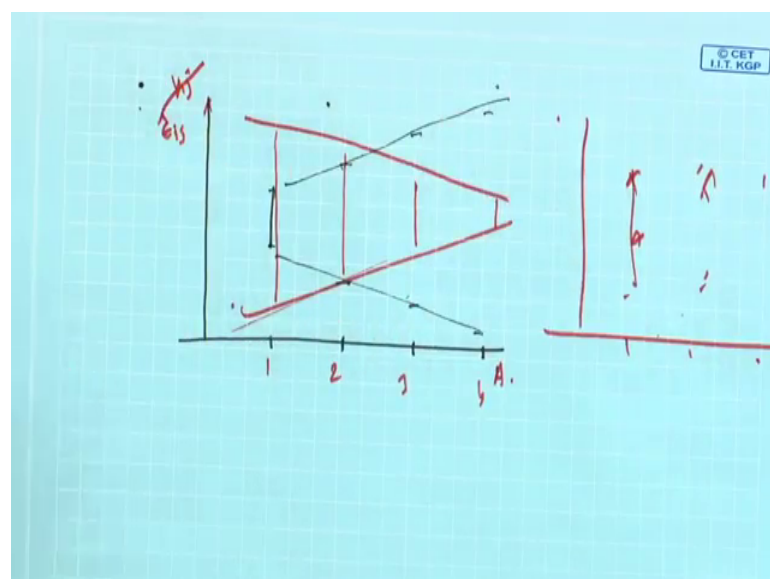
We cannot reject the null hypothesis.

There is no evidence to counter the claim that all five variances are the same. This is the same conclusion reached by analyzing the plot of residuals versus fitted values

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Now, the in the second case power factor level case also we found out that chi square is 0.43 and computed one will be will be much more, because you know the computed one will be chi square a minus 1. So, 4 minus 1 is 3 and alpha and then like this like you will calculate for certain alpha point 0.05 with 3 some value will be that value will be there, but here all though I have not given in this, but montgomery book, that is available what is this value from table you see it is a you have to see it there is no evidence that claim that all 5 that 1 2 3 4. 4 variances are the same this is the same conclusion it is mainly in the plot of the residuals these are the some plots later on.

(Refer Slide Time: 25:41)



Any how you must see that computed chi square, tabulated chi square, confidential square less than tabulated fine the variances are equal. In fact, there is another method for most (Refer Time: 25:57) is checking that is known as Modified Levene method.

(Refer Slide Time: 26:00)

Homoskedasticity cheking:
Modified Levene's test

- Modified Levene's test is robust to departures from normality.
- To test the hypothesis of equal variances in all treatments, the it uses the absolute deviation of the observations y_{ij} in each treatment from the treatment median, say, \tilde{y}_i .

$$d_{ij} = |y_{ij} - \tilde{y}_i| \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n_i \end{cases}$$

- The modified Levene's test evaluates whether or not the means of these deviations are equal for all treatments.
- The test statistic for Levene's test is simply the usual ANOVA F statistic for testing equality of means applied to the absolute deviations.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

What is there in Modified Levene method? So, you find out the median from each form each level of data for example.

(Refer Slide Time: 26:09)

Modified Levene's test - Example

A civil engineer is interested in determining whether four different methods of estimating flood flow frequency produce equivalent estimates of peak discharge when applied to the same watershed. Each procedure is used six times on the watershed, and the resulting discharge data (in cubic feet per second) are shown in the upper panel of the table below. The analysis of variance for the data, summarized, implies that there is a difference in mean peak discharge estimates given by the four procedures. The plot of residuals versus fitted values, shown in figure, is disturbing because the outward-opening funnel shape indicates that the constant variance assumption is not satisfied.

Peak Discharge Data

Estimation Method	Observations						\bar{y}_i	s_i	S_i
1	0.34	0.12	1.23	0.70	1.75	0.12	0.71	0.520	0.66
2	0.91	2.94	2.14	2.36	2.86	4.55	2.63	2.610	1.09
3	6.31	8.37	9.75	6.09	9.82	7.24	7.93	7.805	1.66
4	17.15	11.82	10.95	17.20	14.35	16.82	14.72	15.59	2.77

Deviations d_{ij} for the Modified Levene Test

Method	Deviations d_{ij}					
1	0.18	0.40	0.71	0.18	1.23	0.40
2	1.70	0.33	0.47	0.25	0.25	1.94
3	1.495	0.565	1.945	1.715	2.015	0.565
4	1.56	3.77	4.64	1.61	1.24	1.23

Analysis of Variance for Peak Discharge Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Methods	708.3471	3	236.1157	76.07	<0.001
Error	62.0811	20	3.1041		
Total	770.4282	23			

Summary
Null hypothesis of equal variances is rejected

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Suppose this is the data set, this is my level one and these are the observation find out the median of it.

So, second level out of this 5 observation 6 observations find out the median, third level find out the median, fourth level find out the median, then subtract each row data by the respective median not mean and then you will be getting this kind of data. This data is nothing, but this minus median and the mod of that data only; that means, if you go back d_{ij} is equal to mod of y_{ij} minus $y_{i \text{ median}}$. So, then you are from this original data you are getting the deviation data. Now you use traditional ANOVA with the deviation data.

And the if there is difference, then mean of this deviations you just test. If the mean of differences across different levels are different, than the variance are different otherwise this is equal. So, here with another interesting example, we found out that using Modified Levene test, we found out that p value is less than 001, computed value is seven point 76.07 and then tabulated will be much lower.

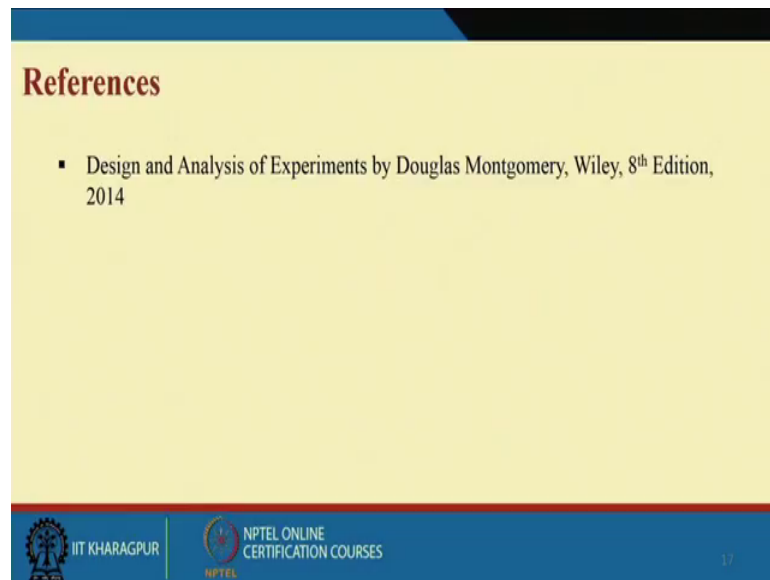
So, what I can say that there is there is difference for example, in the first level this s the this is the spread, second level this s the field, third level this is the field fourth level this is spread. So, what is the spread here then? So, arrow spread is also sigma square. So, level 1, 2, 3, 4 first level spread is like this you just see the minimum to maximum.

This is minimum to maximum speed like this second one second one like this third one third one like this, and in fourth one fourth one is also like this. So, it is if you add this, it is basically making a finagling of it. So, when your data represent this kind of finagling effect either towards left finagling towards right or it may.

So, happen here is the maximum here is lower, here is lower, here is lower. So, maybe finagling to the left, when you your data suppose this is y_{ij} and these are the factor level 1 2 3 4 different level. So, if your data represent like this then finagling right or finagling left, then there is change in these is variability difference in variability.

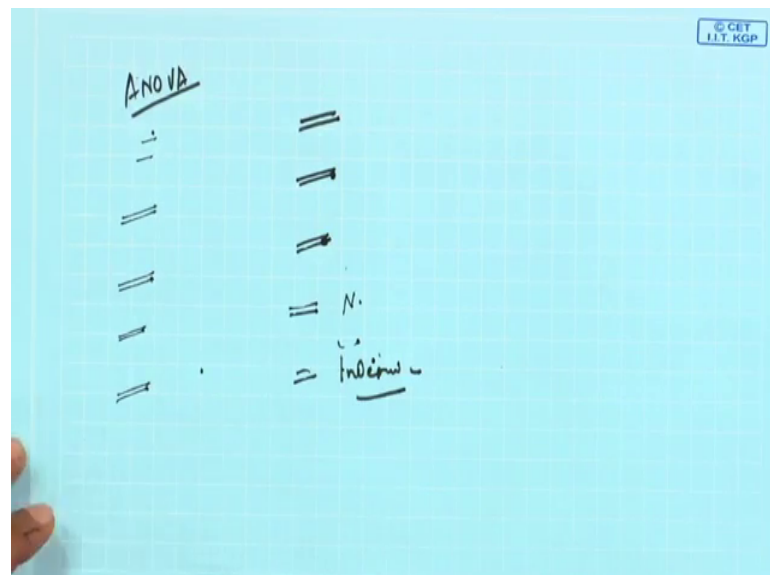
Usually we not use y_{ij} we use the residual ϵ_{ij} , residual in the residual plot you will you will come to know all those things. So, if your data represent random pattern something like this whatever may be the level, but these are all random here here here here there is no finagling left or finagling right, then this is homoscedasticity data otherwise it is a heterogeneous, heteroscedasticity a they are variances across different layers are not same very important one.

(Refer Slide Time: 29:37)



So, then I hope that you now you are comfortable, you will become comfortable with ANOVA in the previous lecture and this lecture last tool. In fact, last three lectures we talked about ANOVA and we have given starting from ANOVA model then partitioning the observation.

(Refer Slide Time: 29:53)



Then partitioning the sum squares then ANOVA table, then f test, that is hypothesis test, and then what happened we estimated the parameters, then the we have seen the mean and point estimation and interval estimation, we introduced a concept called

simultaneous confidence interval using (Refer Time: 30:25) approach and then the test of assumptions, test of assumption, normality test, variance test, and then there will be there will be test like independent test, independent observation test; that means, the data are not related to each other; so, there are so many things.

Thank you very much.