INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR

NPTEL
National Programme
on
Technology Enhanced Learning

Applied Multivariate Statistical Modeling

Prof. J. Maiti
Department of Industrial Engineering and Management
IIT Kharagpur

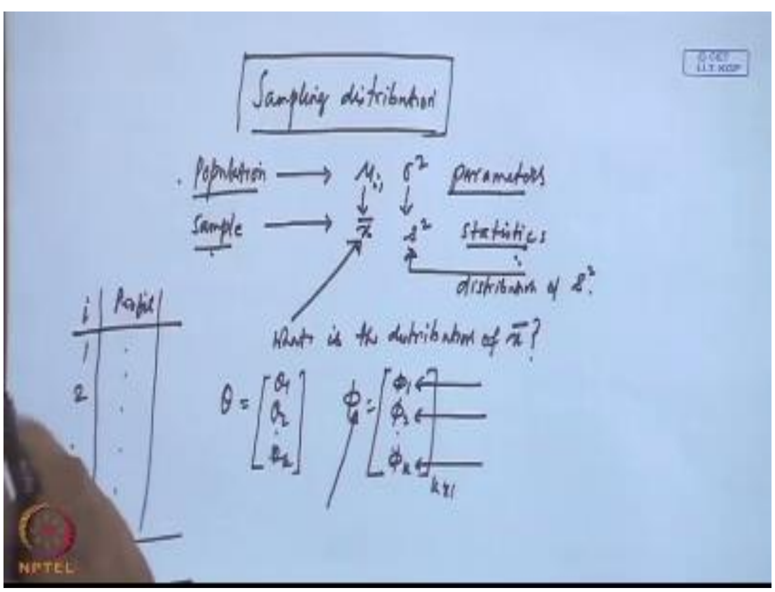
Lecture – 04

Topic

Sampling Distribution

Good afternoon in this lecture we will discuss sampling distribution.

(Refer Slide Time: 00:29)



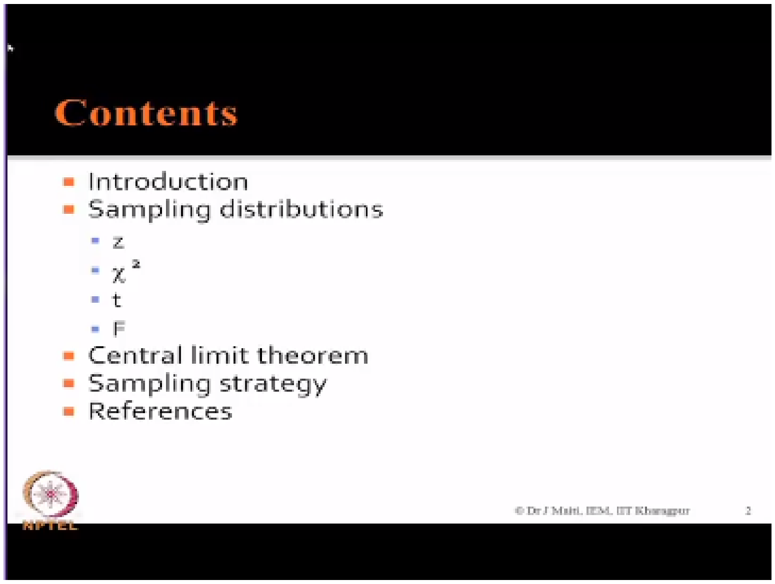
In the last lecture you have see that population each characterized by probability with certain proper parameters also we have discussed that sample can be collected from the population and then estimate of population parameters can be found out the in terms of the $\overline{x}$ in terms of a^2, $\overline{x}$ and a^2 each one is known as statistic okay and $\overline{x}$,a^2 collectively you can say statistics so population parameter sample statistic they are synonymous as I say when I talk about population I talk about population parameter. Parameters talk about sample the sample statistic is the estimate of parameter.

Okay now if I say that $\overline{x}$ is an estimate of population parameter μ then what is the distribution of what is the distribution of $\overline{x}$ similarly you see their s^2 is the sample variants got the variable x and it is an estimate of σ^2 that is population parameter you may be interested to know the distribution of s^2 getting me the distribution of a statistic is known as sample distribution so when we talk about sampling distribution.

We talk about the distribution of a statistic computed from the sample that distribution of that now here $\overline{x}$ is a sample statistic what is the distribution $\overline{x}$ similarly a^2 sample statistic what is the distribution of a^2 under sampling distribution we will discuss this two but pleases keep in mind in general if we say the θ a = let be θ_1, θ_2 like suppose θ_k there are k parameters per θ then definitely suppose your Ø is the sample statistic then corresponding to every population parameter there will be sample statistic will be there now by sampling distribution we say what is the distribution of this statistic what is the distribution of this statistic it may be innervate it may be multivariate so if you are interested no collectively what is the distribution of this statistic vector k_i cross one vector.

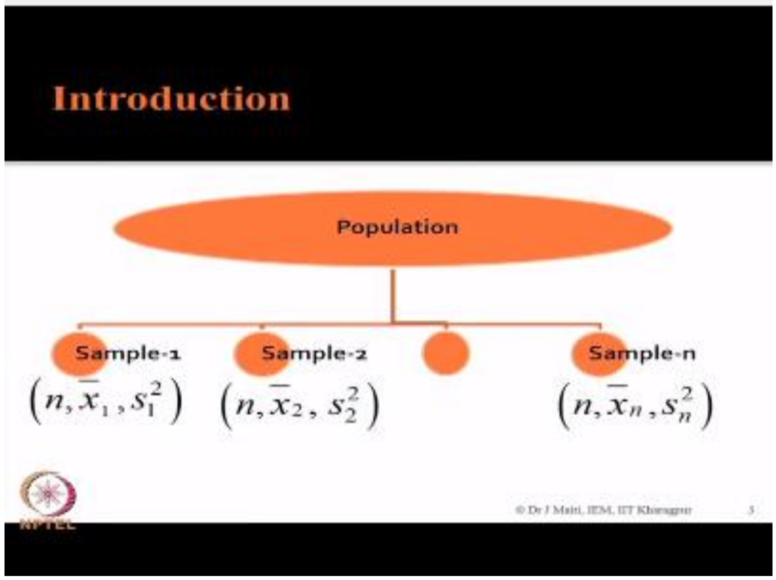
Then it will be multivariate distribution essentially by sampling distribution we were talking about probability distribution only but the reference is you are talking about the distribution o not the variable character in the population rather the statistic computed from sample distribution of that statistic clear it is for example we have taken the if we consider the same example again the profit if you see that 12 months data 1 to 12 so your profit is there different values of profit so you will calculate mean profit average per month my question here is what is the distribution of the $\overline{x}$. So with this line today's discussion.

(Refer Slide Time: 05:41)



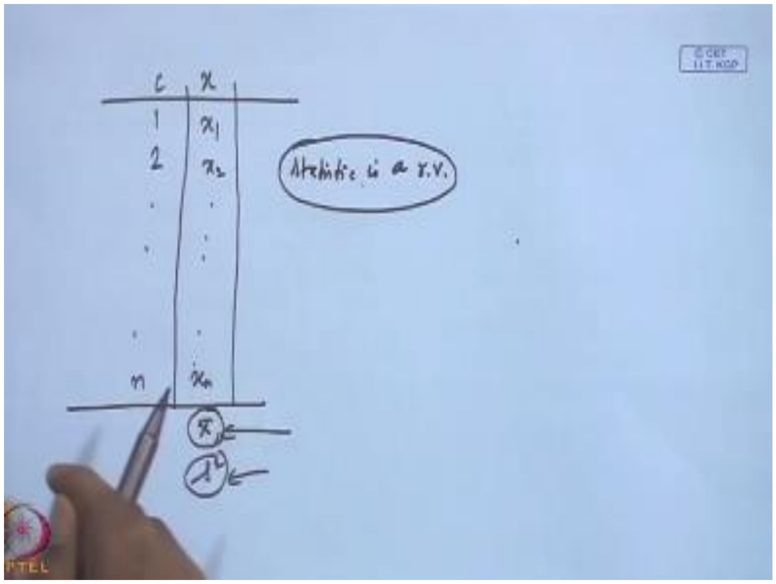
Is sampling distributions and you see that there are unit normal distribution z unit normal distribution then k^2 distribution t distribution F distribution these are the most popular mostly used widely used distribution for sample statistics then we will discuss center limit theorem finally sampling strategy because where how to collect the data it all depends on what strategy will be adopting because the data collection process if you wrong or faulty analysis will not give you good result. Okay so you may be wondering that where from this suppose you have collected one sample.

(Refer Slide Time: 06:37)



For example

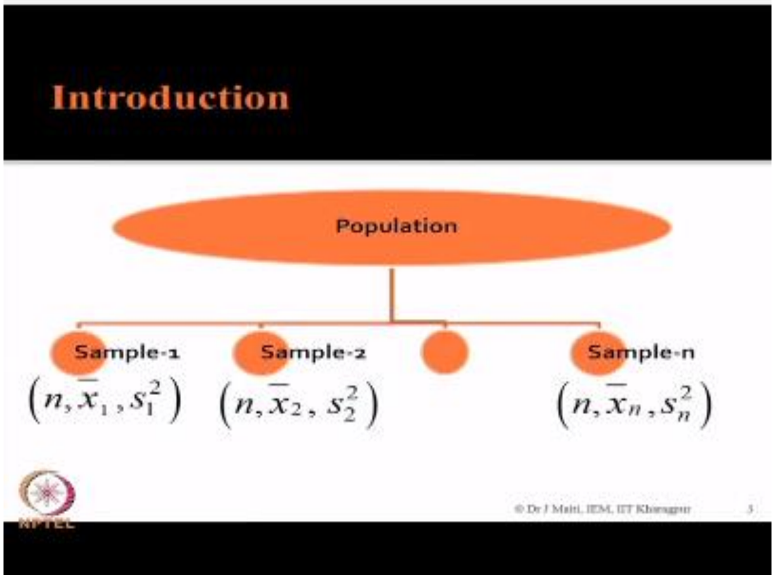
(Refer Slide Time: 06:41)



Collected the height of people for a particular community suppose n data points you have collected 1, 2….n your value x value is x_1, x_2….x_n you are think that I have computed $\bar{x}$ here and a^2 here and these are all because once you collect data this these are the fixed values so you are $\bar{x}$ is a fixed value a^2 is also a fixed value for a particular sample then where from this distribution is coming the distribution concept is coming because if you do the same sampling next time there is no guaranty that you will get the same value of $\bar{x}$ same value of $a^{2.}$

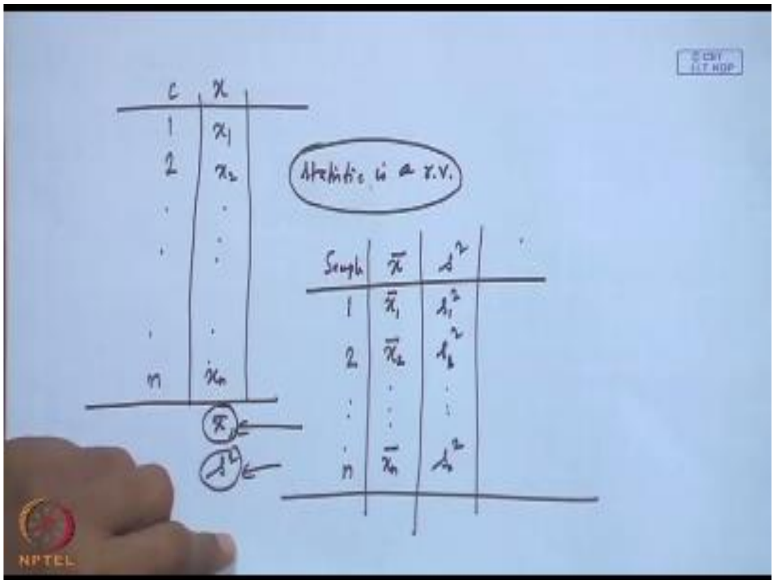
That is why we in statistics all statistics are otherwise each statistic is a random variable statistic is a random variable in this figure you see this figure.

(Refer Slide Time: 08:02)



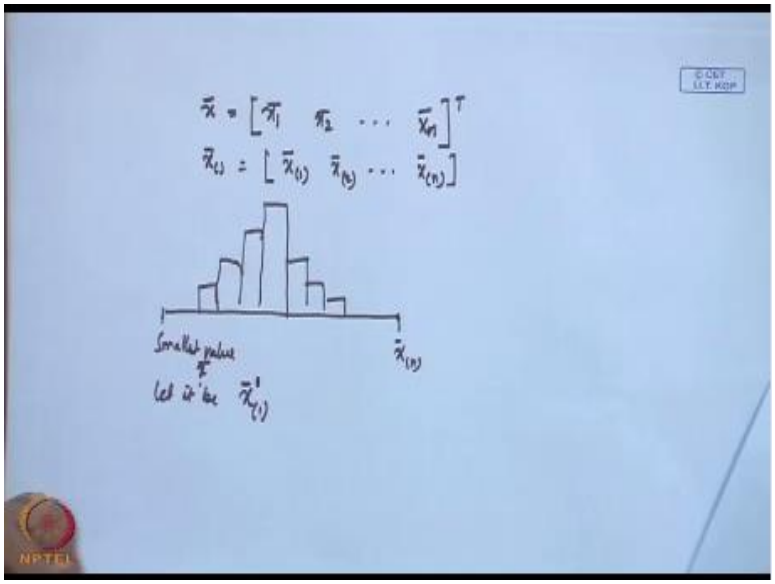
That sample n sample 1 to sample n collected from a particular population and the sample size n for all the cases sample size means the number of observations collected for sample and if you compute the mean and variants for each sample you will be getting for sample it is ($\bar{x}_1$, s_1^2), ($\bar{x}_2$, s_2^2) for second sample like this and it is obvious that all those values will be may not be same will be actually little different.

(Refer Slide Time: 08:45)



Suppose if I say that sample, sample 1, sample 2 like sample n's then if you calculate your mean and calculate variant $\overline{x}_1$ is the simple mean $\overline{x}_2$ is the second sample mean like this $\overline{x}_n$ is the n sample mean s_1^2 first sample standard variants, variants and s_2^2 is the second sample variants like this the variants s_n^2 now as these are not same value what will happen if you can draw histogram how you can find out the smallest.

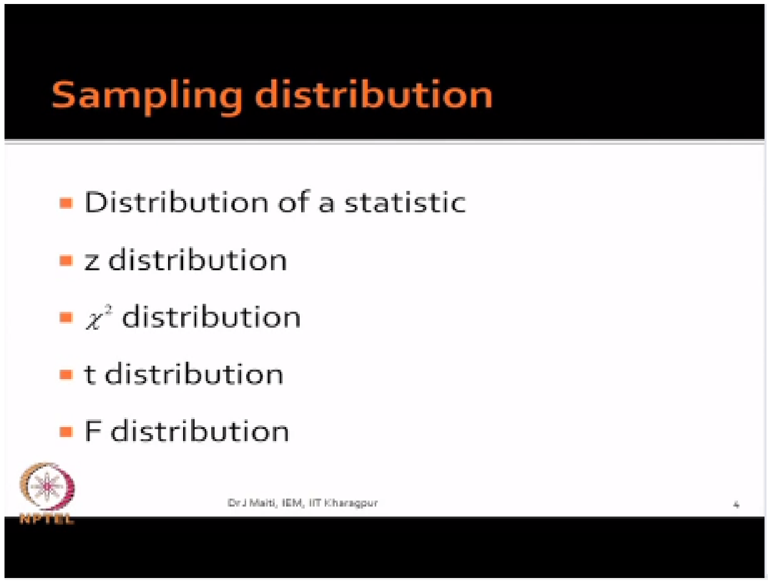
(Refer Slide Time: 09:36)



Smallest value of $\bar{x}$ let it be if I say x with brackets once bar originally what happen originally you have $\bar{x}_1$. $\bar{x}_2$ like this $\bar{x}_n$ if I say this is a vector $\bar{x}$ which is this one this T stands for transpose then you are what are we them from smallest to largest then the same thing what will happen if I say that some order we are given then this will be $\bar{x}1$, $\bar{x}2$, $\bar{x}n$ so this is the smallest one is this and the largest one is $\bar{x}n$.

So if you plot you will develop a histogram you may find out a distribution, it may be like this. That means what we are trying to say that if you collect sample one after another from the same population of the same size for the same variable you will get different values for that sample statistics and that sample that is why the sample statistics is random variable and you have the probability density function for the random variable. And that density function is the sampling distribution function, okay.

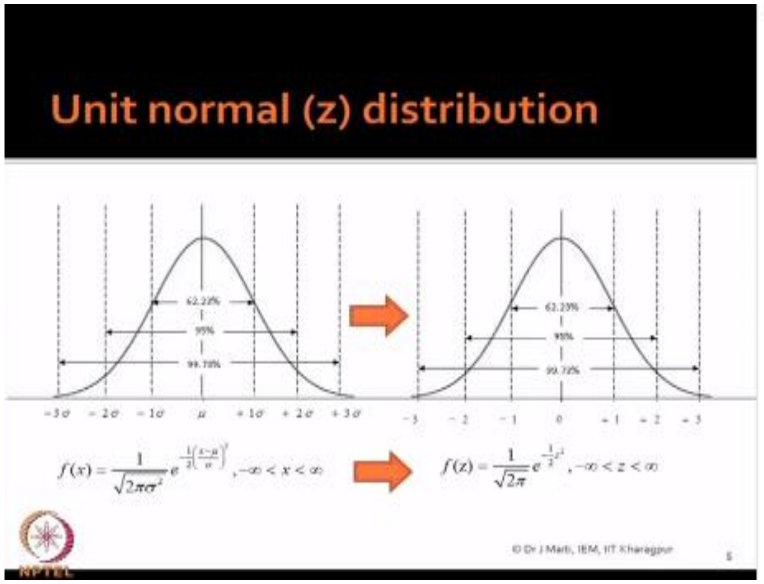
(Refer Slide Time: 11:33)



Now you have to know that what are the distribution that comes under sampling distribution, I told you earlier that there are four most popular sampling distribution like z distribution, x^2 distribution, t distribution and F distribution and these concept is very vital concept I am telling you and very important for later stage is when we talk about any multi variate model for example multiple regulation there are several bitter co-efficient you will defining out that regression co-efficient.

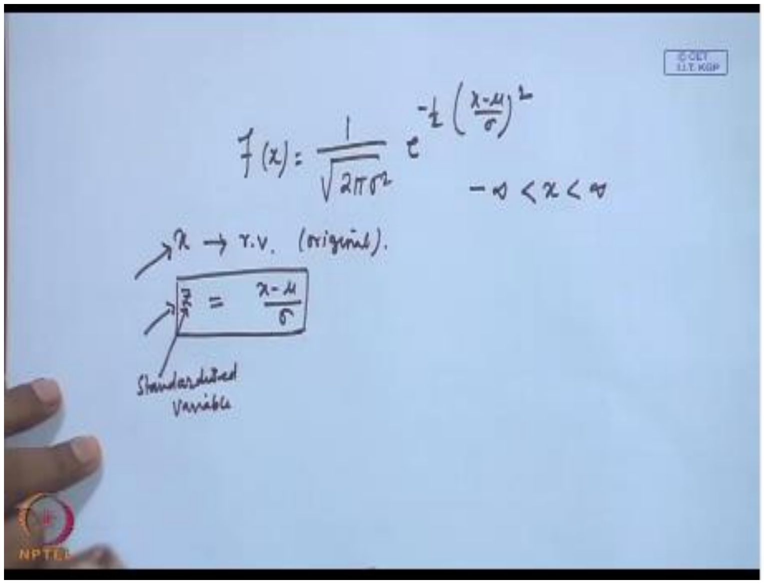
These regression co-efficient through data sample data you will be estimating they are basically the statistic, so each parameter will have the distribution, how do I know that what distribution it follows, if you do not understand these you will face problem there. So this is I can say that one of the most fundamental concept and it should be thought about this concept, okay. And now let us concentrate on what is Z distribution.

(Refer Slide Time: 12:58)



I am sure that you will not face it a difficult one because all of you know that this left hand side the figure if you see this is the normal distribution and the probability density function is f(x) = 1 by I am writing here once more.

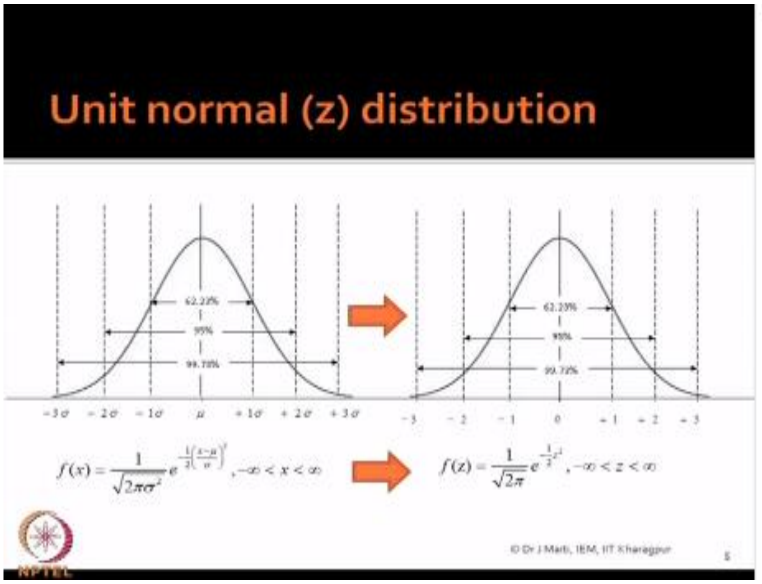
(Refer Slide Time: 13:20)



$F(x) = 1/ \sqrt{2}\Pi\ \sigma^2\ e^{½\ (x-\ \mu/\ \sigma)2}$ and x varies from - ∞ to + ∞ okay. Now where x is a random variable this is the original one original random variable original one what is origin, correct. Now let us transform x in this manner let JDJ transform variable of x which is x- μ/ σ so you have observed x you are creating one another variable z which using x as well as the population parameter like μ and σ and this is what is known as the transform standardize variable this is known an standardized variable.

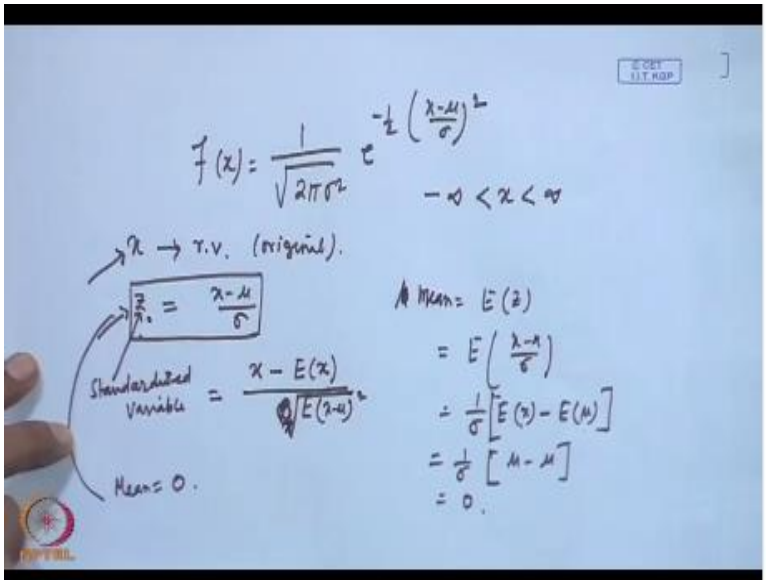
So when a variable is subtracted by its mean and divide and the resultant quantity is divided by the standard deviation that is known as standardized variable.

(Refer Slide Time: 14:50)



So that mean y standardized variable we mean.

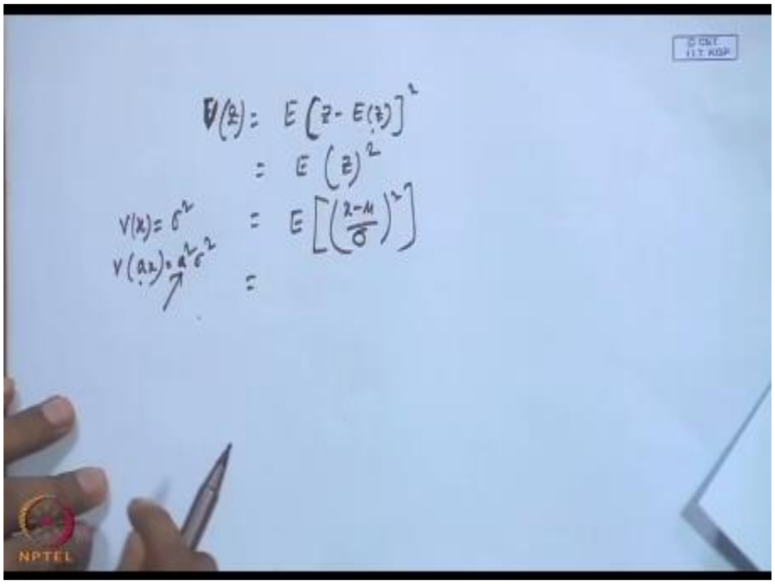
(Refer Slide Time: 14:52)



That the variable minus its expected value that is the mean divided by that what I can say sigma means basically $x - \mu$ expected value of this square. Square root, this square toot of this because we are considering standard deviation square root of this, correct. Now what will be the mean value of z, 0 how? Since mean is subtracted a brief case basically you can write in this manner suppose what is mean?

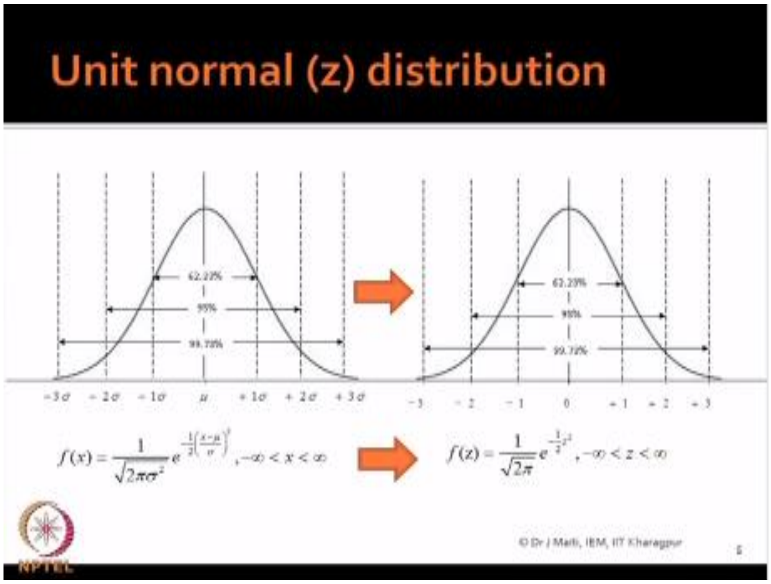
Mean is suppose if I write mean, mean is expected value of variable z here we are talking about μz so that mean expected value of $x - \mu / \sigma$ that mean $1/\sigma$ expected value of E(x) minus expected value of μ and c expected value of constant is constant so expected value of x is $\mu - \mu = 0$. So that is why the J, JD is a random variable whose mean value is 0, what will happen to its standard deviation? You will get one. In the same fashion you can find out that what is the standard deviation.

(Refer Slide Time: 16:35)



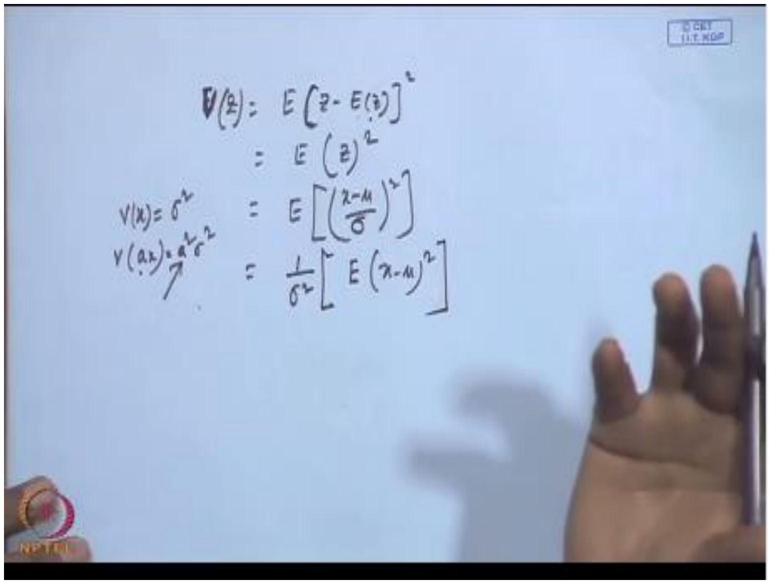
Standard deviation means expected value that is the variance you find out first variance of J which is the expected value of your JED minus expected value of JED that whole square, so expected value of JED already you got 0, so this basically expected value of JED square so it is expected value of JED is nothing but x- μ/ σ square so if the variance of x suppose variance of ax is σ^2 The variance of ax square in each constant it will be a square σ^2 so it will be σ^2 so that mean what you can wrote this one.

(Refer Slide Time: 17:39)



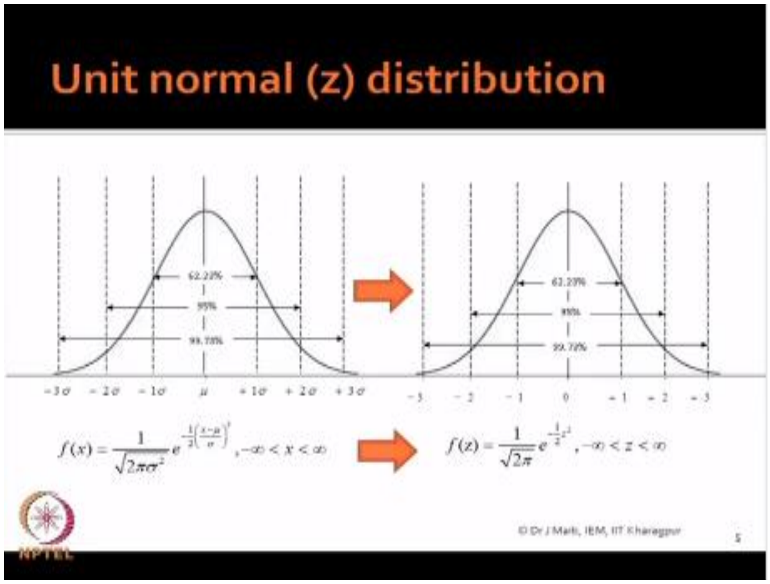
$1/\sigma^2$.

(Refer Slide Time: 17:41)



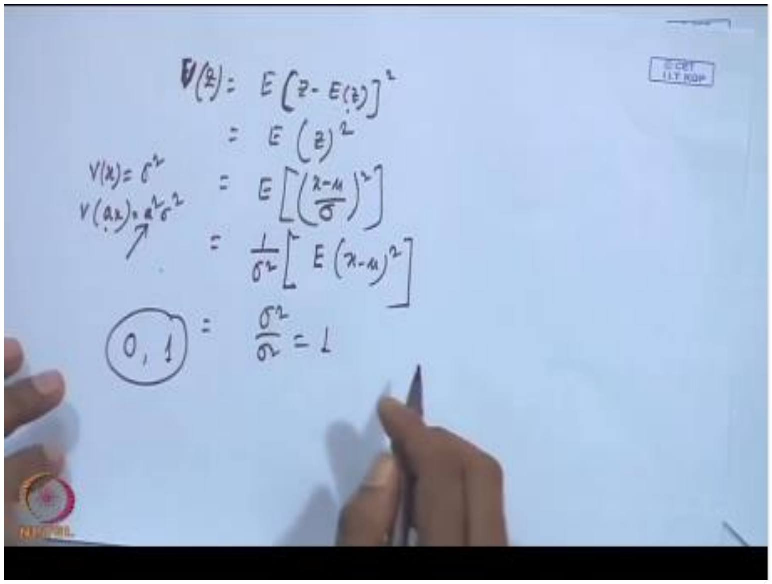
Into expected value of x – μ^2 what it is, expected value of x- μ^2 that is the σ^2 so your σ^2/ σ^2 this is 1, is this is the case now you put you see that in this equation.

(Refer Slide Time: 18:08)



Come back to this slide again what we have put we out JED = $x- \mu/ \sigma^2$ our resultant equation that probability density function for JED which is $1/ \sqrt{2\pi}\ \sigma^2$ is 1 so it is 2Π x 1 means 2Π square root, $e^{1/2z2}$ so this is the conversion of any variable for ensample this normal variable to its you need normal distribution. So Z distribution is also known as you need normal distribution because its mean value is 0 and standard deviation is 1.

(Refer Slide Time: 18:51)



Okay what is the use of why should we convert to unit normal distribution? Reason is that even if there are many variables but once you standardized those any variable it will be a unit normal, so you require only one normal unit normal distribution table and using that you are able to what I can say use that table to difference situation even with the original variable mean and standard deviation differs, okay.

(Refer Slide Time: 19:33)

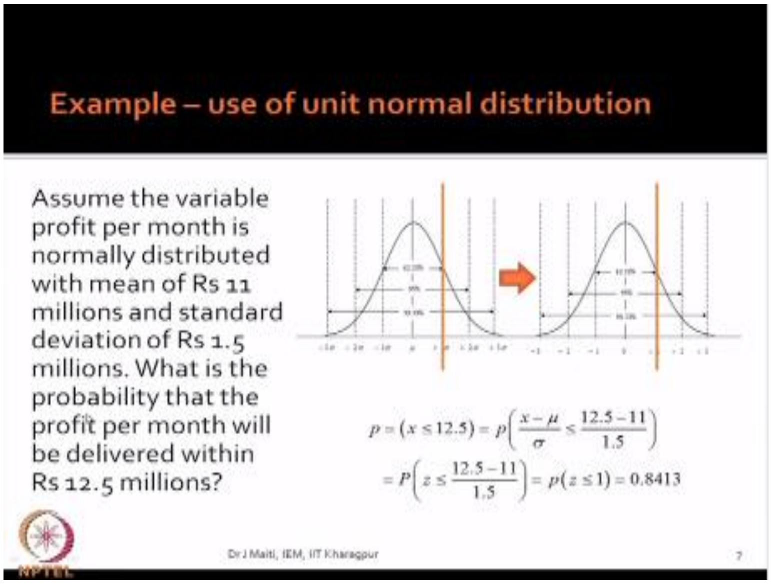
An example

Sl. No.	Months	Profit in Rs million	Sales volume in 1000	Absenteeism in %	Machine breakdown in hours	M-Ratio
1	April	10	100	9	62	1
2	May	12	110	8	58	1.3
3	June	11	105	7	64	1.2
4	July	9	94	14	60	0.8
5	Aug	9	95	12	63	0.8
6	Sep	10	99	10	57	0.9
7	Oct	11	104	7	55	1
8	Nov	12	108	4	56	1.2
9	Dec	11	105	6	59	1.1
10	Jan	10	98	5	61	1.0
11	Feb	11	105	7	57	1.2
12	March	12	110	6	60	1.2

 6

So this is our example and if I consider profit.

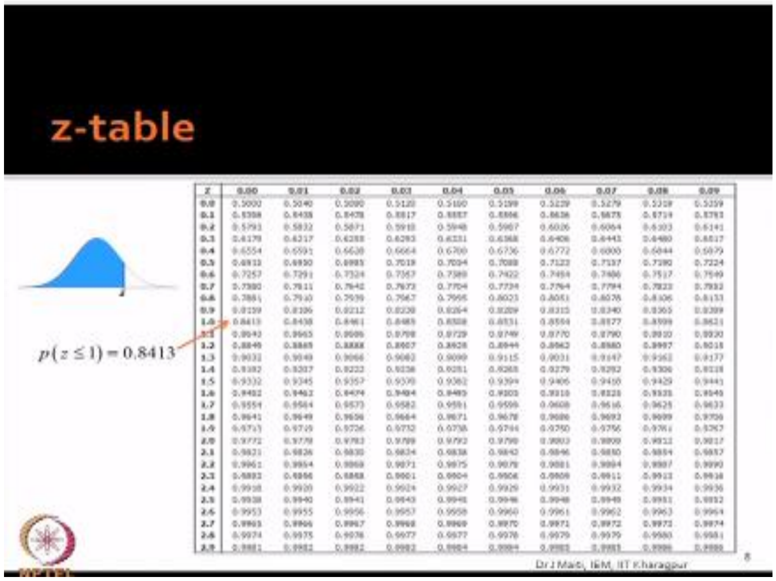
(Refer Slide Time: 19:39)



We are considering the profit I am showing you that what is the use of standard normal distribution here assume the variable profit for month is normally distributed with mean of rupees 11 millions and standard deviation for rupees 1.5 millions what is the probability that the profit of month will be now within 12.5 millions how do go about it in the left hand side this figure it is in the original variable right hand side is the in that immediate normal one now if you see these x axis the bottom what you see what is the mean is the little value and everywhere that one standard deviations two standard three standard deviation both side the d mark is in their now as $\sigma = 1$ that same will be now.

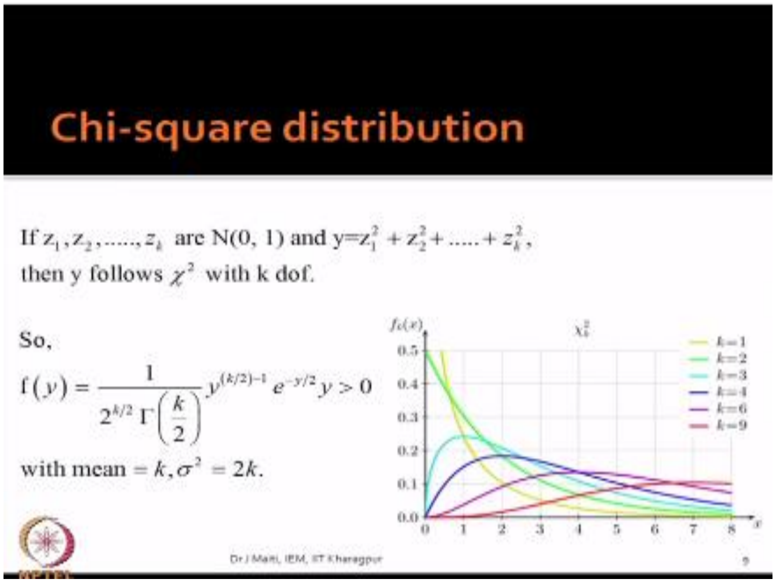
Men will be 0 and 1 x 1 that will be 1 and – 1 – 2- 3 like this so these line is what is the z line because you are introduced to know that your probability of profit less than equal to 12.5 Rs million and if I convert into z value it is coming one so probability z less than = 1 that men this one this is the z = 1 the left hand side values probability z = z less than = 1 will be the area under the normal distribution go from - ∞ to that z value because we are talking within this now you have standard normal table now what you will do once you get the z value you go for the table.

(Refer Slide Time: 21:37)



So our z value is 1 and you see that this is the sided portion is the probability this is the probability and which value is 0.8413 so you are able to find out the probability that your profit will be within this.

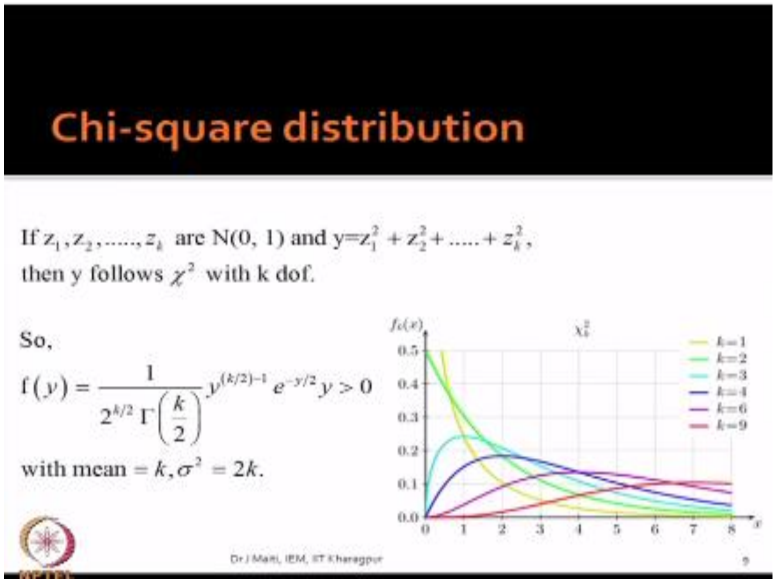
(Refer Slide Time: 22:02)



Now second distribution sampling distribution is chi – square okay do you have any idea that when you will use chi square distribution you find out that if you go through the standard book like very good [indiscernible][22:27] everywhere mean maybe when you talk about the constant density contours they are statistical this the hi square distribution is used but why chi- square distribution is used instead of t distribution or instead of z distribution or any other distribution what is the basis that.

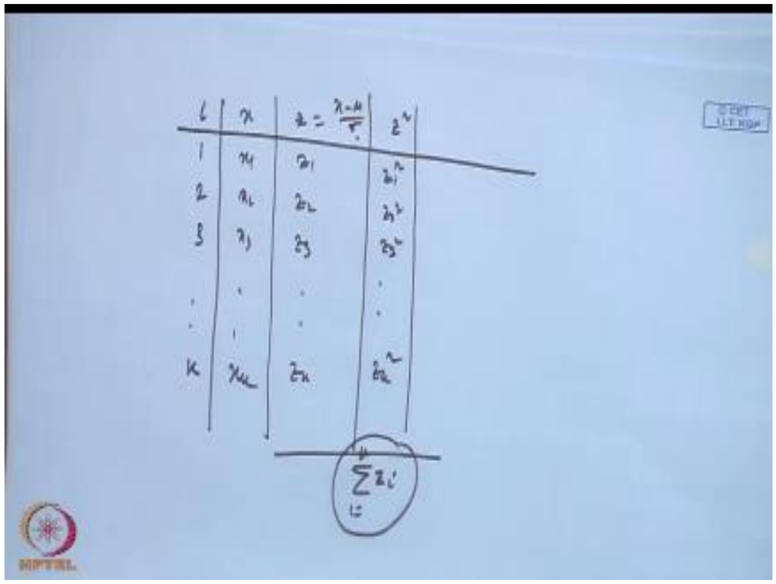
Men we must know that what is chi – square distribution how they generated and why it will be used now you see this slide here what we are seeing that if z_1 z_2 z_3.

(Refer Slide Time: 23:04)



What is z, z is normal distribution in the normal so you have collected suppose key of observations and those k unit that normal observations are the z_1 z_2 z_3 and getting one variable which is the sum of the normal variable unit normal variable square me what I mean to save here then suppose your collected.

(Refer Slide Time: 23:34)



In k data points 1 2 3 like k and you have x values for x1 x2 x3 it will xk and you know z values z = x – μ by σ as you mean there μ and σ are population parameter then you are getting z1 z2 z3 like zk now you are making z^2 $z1^2$ $z2^2$ $z3^2$ like this zk^2 now if you take some of these z_i i = 12k you get a quantity this is a statistical because this is linear some of the square of the variable values of the normal unit normal variable then this quantity that y what you created here y I show here it is here y any how you can change it to y no problem.

Okay so this quantity follows chi – square distribution with how many degrees of freedom k degrees of freedom so what is the essential learning here our learning is suppose I know that there is a normal variable x your collected data on it you convert it to standard normal each of the observations is squared and you have taken this sum and that sum you it for different purposes that sum will follow certain distribution that distributions chi- square distribution okay remember this one.

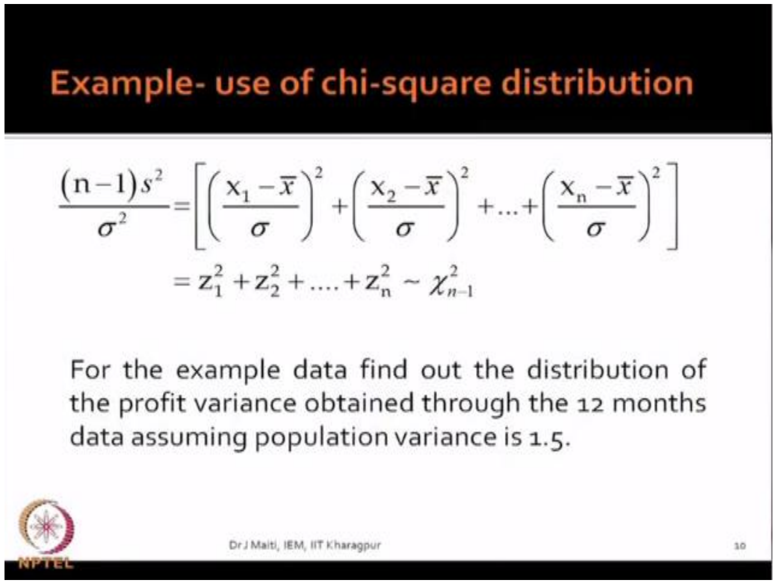
Suppose at usually y normal distribution we will started normal distribution is normal distribution plenty of things in the real world most of the things can be converted to normally distributed that are most of the cases so that is the starting point now for the purpose of your analysis purpose your model be link or the purpose for which you want to use it they are what is required we required the sum, sum of the square of the variables values then what you will do when you collect one simple you have to have the distribution of that value that is chi – square distribution any question understood fully or not so any question okay and the general form of chi – square distribution.

Is like this that f(y) is ½ 2 k/ 2 Γk/2 and you see that is and the mean value of chi – square is the k which is degrees of freedom and variance is two times degrees of freedom and there in this figure you see that this is the probability density function for a chi – square variable now that if x is chi – square then what is bring here 1 to 8 these are the values and ultimately you will be getting different shape of chi square density function when your k = 1 this is this as well as k = 2 it looks like exponential distribution but slowly that shape will change so degrees of freedom place in important role in chi – square distribution.

In z distribution what is the degree of freedom with nodding this of freedom we have not discussed anything related to degrees of freedom we get distribution when the no even normal distribution table you see there is no degrees of freedom column so the z distribution it is basically not affected by the degrees of freedom available with the data set chi – square distribution when you talk about chi- square distribution please keep in mind degrees of freedom is coming into coincidental.

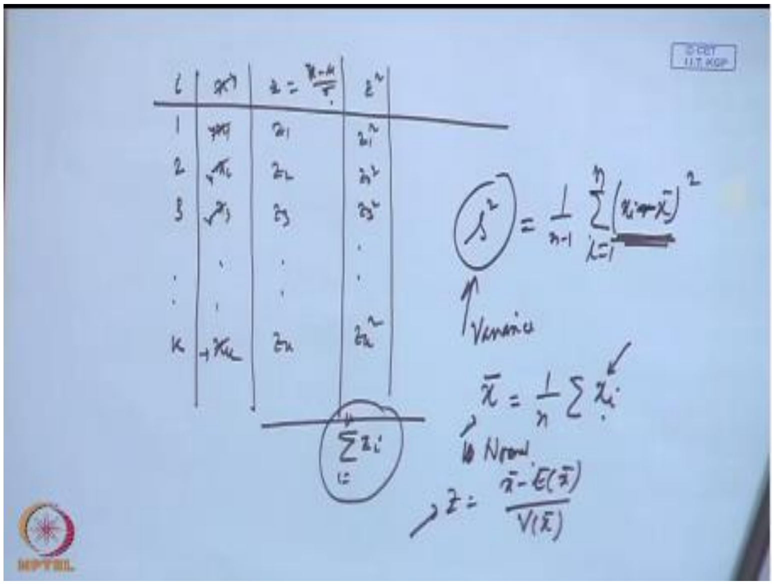
And chi - square is nothing but the normal square my x is normally distributed I making the square linear sum of the square of x that is my chi – square distribution so you may be wondering then where is the use now see I told you we want to find out the distribution of sample statics getting now one of the sample statics is a^2.

(Refer Slide Time: 28:39)



Variance.

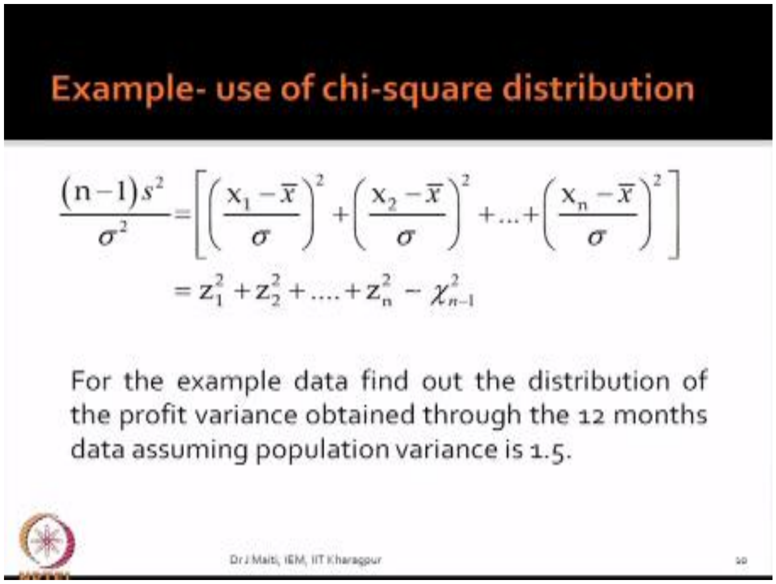
(Refer Slide Time: 28:43)



Yes or no very much what is the distribution of s^2 how do you know what is what will be the distribution of the $\bar{x}$ see $\bar{x}$ is 1/n sum total of xi x each normal distributed variable so $\bar{x}$ if x is normally distributed $\bar{x}$ also follow normal distribution we will create z by considering x bar – expected value of $\bar{x}$ by variance of $\bar{x}$ then you mean normal distribution I will not tell anything related to this computation later on I will tell you what is this from center limit theorem you will know is what will be the distribution but irrespective of what I mean to say that irrespective of the value that $\bar{x}$ value.

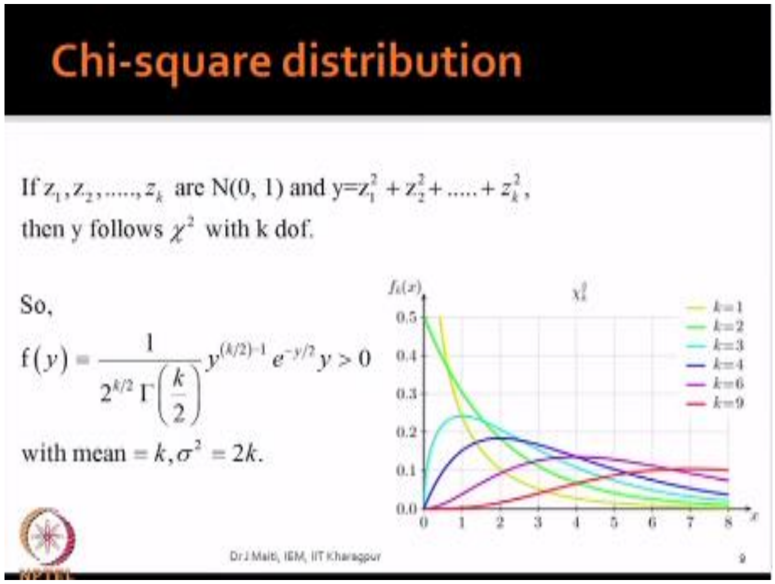
If you collect data for normal distribution that $\bar{x}$ will be normally distributed now what will be the a^2 how do what is a^2 s^2 is 1/n-1 sum total of i=1 to n $(x_i - \bar{x})^2$ can you find out any similarity here, what you have done. I say x is normally distributed $\bar{x}$ is also normally distributed differential also normally distributed. Then you have made the square in this, so when normal variable is squared and you take ∑ and do little bit of manipulation using the population variance what you will get, you will get standard normal z and ∑ of square of standard normal z, is it is not correct you see.

(Refer Slide Time: 30:53)



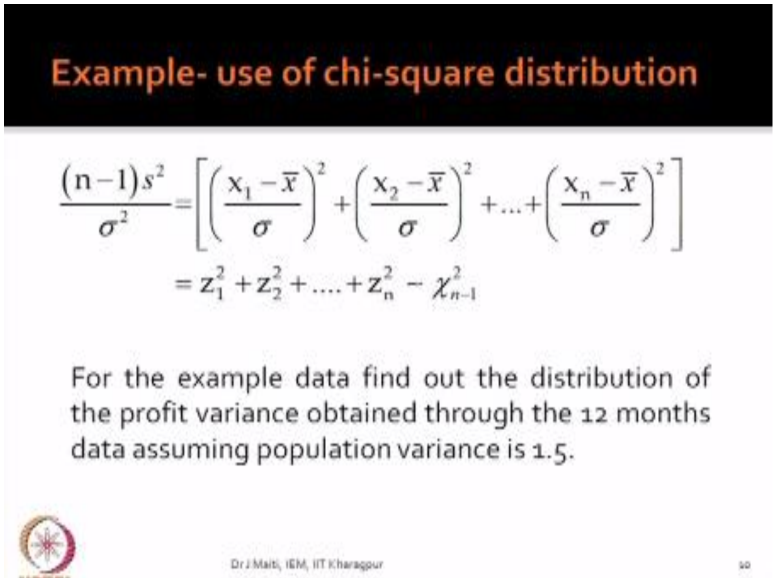
You see this one slide, what we have shown here s^2 it just, it is nothing but the formula we are giving $(x_1 - \bar{x})^2$ like this $+(x_2 - \bar{x})^2 + (x_3 - \bar{x})^2$ and σ^2 is divided this σ is the population variance. Now this quantity is coming like $z_1^2 + z_2^2 + z_n^2$, so it is the sum total of normal squares, then you variable square. So this chi square distribution that is why when we go for interval estimation or variance we use chi square distribution. So it should come to your mind that why $\bar{x}$ for $\bar{x}$ we are using normal distribution but for x^2 we are using chi square distribution, that mean the fraction the matter is here.

(Refer Slide Time: 31:56)



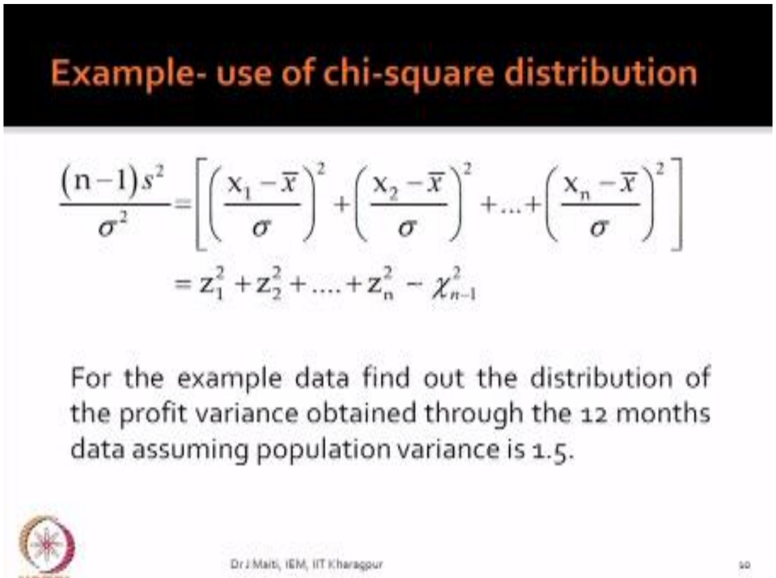
This is the development, fantastic.

(Refer Slide Time: 32:02)



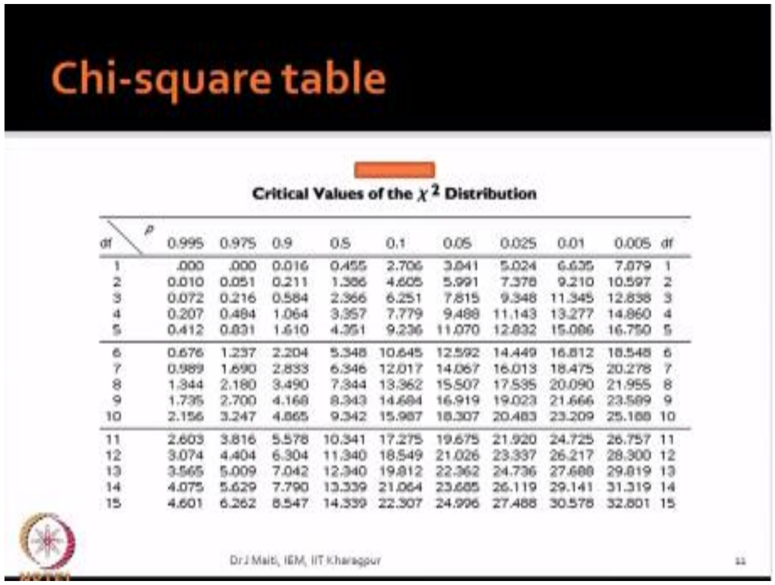
Then you may be wondering that why the element is 1 already we have seen that while calculating the s^2 one degree is lost then same thing continuous and ultimately our these $(n-1)s^2/\sigma^2$ these quantity follows chi square distribution with (n-1) degrees of freedom. Okay, any question for this, so any question, no question. I think it is sure, it is obvious now, so keep in mind this one because many times you will be using this type of derived units, you will add square and then add and chi square is equate but you do not know what distribution will be using just follow this concept and you use a chi square some other case it will be other distribution.

(Refer Slide Time: 33:07)



Okay, this is the huge for example, if we consider the data then profit variance obtained through the 12 months data assuming population variance you can find out the distribution of the variance component computed from the 12 months data, it is or the hugest of this chi square distribution will be revealed in subsequent lectures. But the concept remain same where we will use chi square distribution this data concept, okay.

(Refer Slide Time: 33:47)



Chi-square table

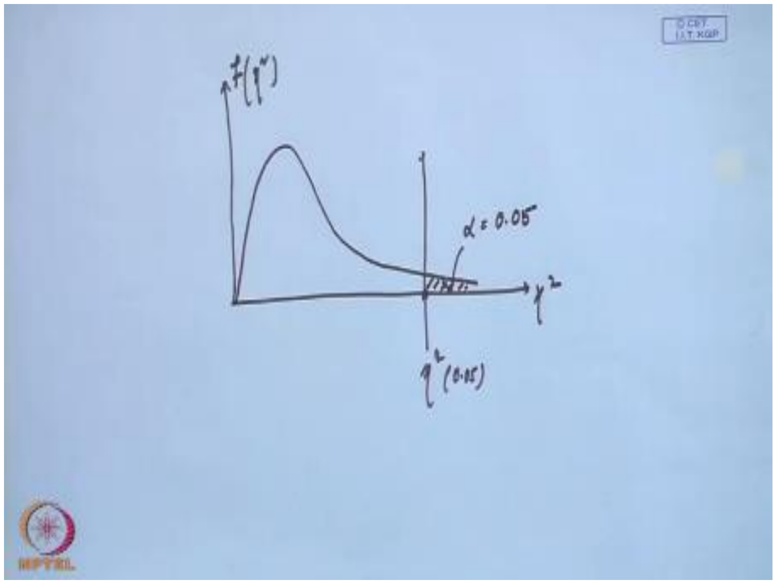
Critical Values of the χ^2 Distribution

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

Dr J Maiti, IEM, IIT Kharagpur

Chi square table, when you have gone for z table please in mind that in z table there is no degrees of freedom column. Chi square means there will be degrees of freedom, so they are the first column itself it is degrees of freedom 1 to 15 it will go, it will up to infinite because chi square value can go up to that level, then there are different probability values here and for different probability values what will be the your chi square value. So how to use this table any idea for example,

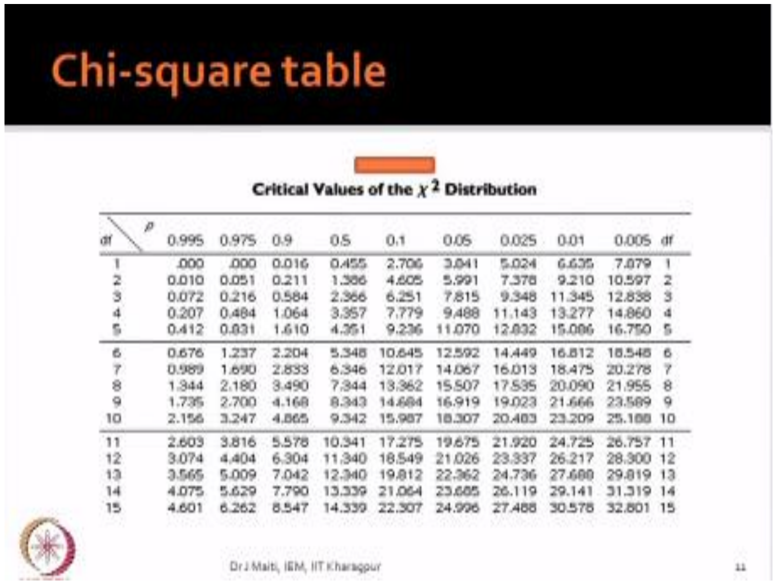
(Refer Slide Time: 34:28)



Let the chi square distribution pdf is like this chi square this is function of that chi square in pdf, let the distribution look like this, okay. Now I want to move what is the chi square value for the probability of this side let it be α which is 0., let it be 0.05 from this table can you find out this value, chi square value what will be this chi square 0.05 understand because these things you will be requiring later on, you have to see this chi square table frequently.

What I mean to say, I know the probability right hand side probability here for which I want to know what will be the chi square value. If you are giving like this chi square with probability 0.05 you cannot calculate, you cannot find out the value from here, which degrees of freedom.

(Refer Slide Time: 35:49)



Chi-square table

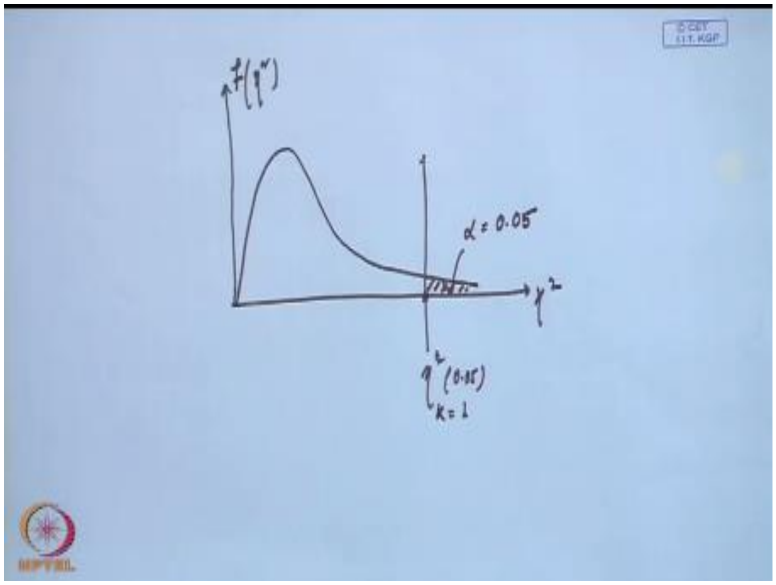
Critical Values of the χ^2 Distribution

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

Dr J Maiti, IEM, IIT Kharagpur

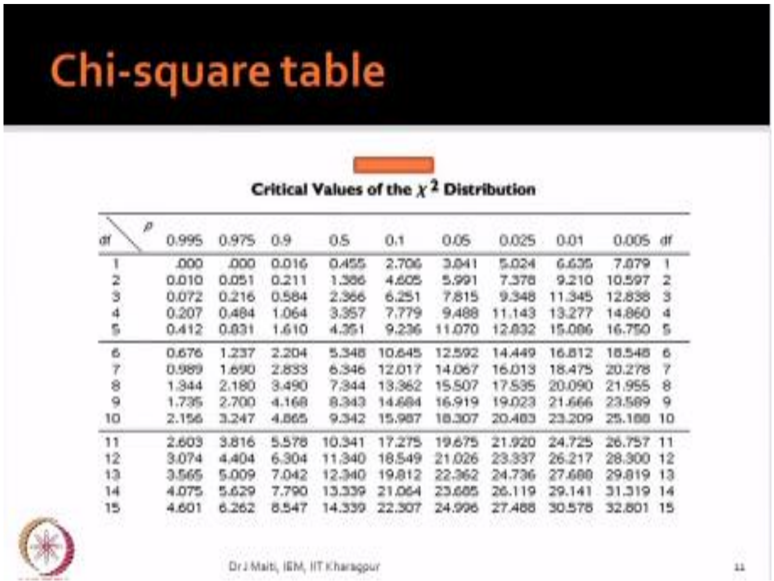
Is it 1 degree of freedom or 15 degree of freedom or 120 degrees of freedom.

(Refer Slide Time: 36:00)



So that mean another quantity should be your, which is known as degree of freedom. So suppose k=1, then what is this value.

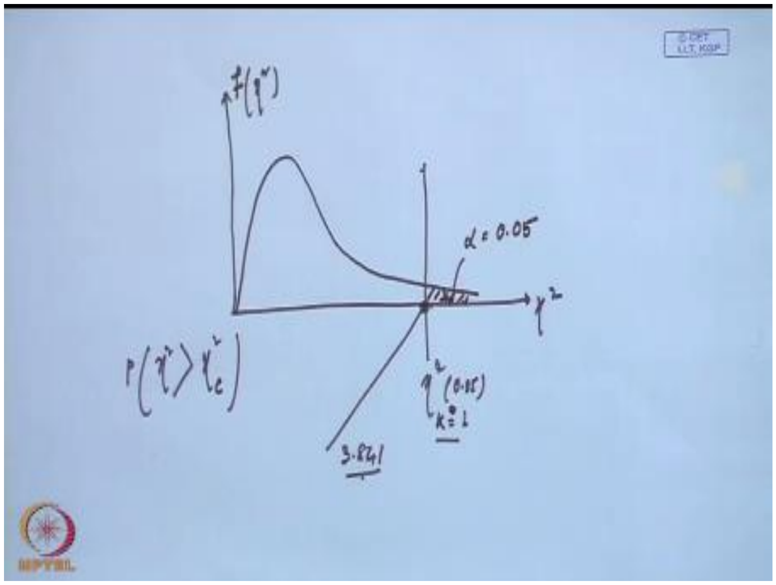
(Refer Slide Time: 36:08)



df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

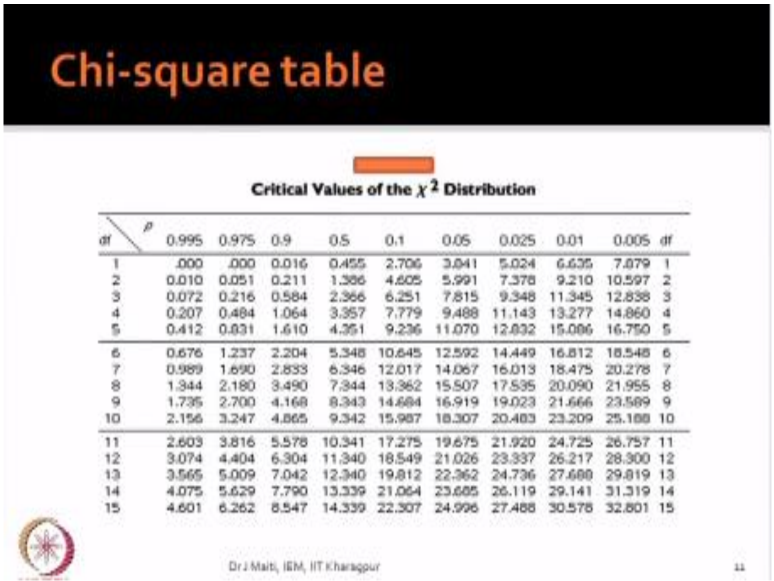
k=1 probability is 0.05 so your chi square value is 3.841, so my this value is 3.841. Suppose require a value where the probability.

(Refer Slide Time: 36:26)



Probability, so what is this, this is, this one is probability that chi square value is this probability that chi square value this value that will be greater than some value, let it be. So let any value we have basically k=1 some value these three epidermis, but that value will say that chi square computed value, getting me. Probability that, that this is make chi square is in chi square is computed this value this will be greater than this, this is the value, this side less than this will greater side.

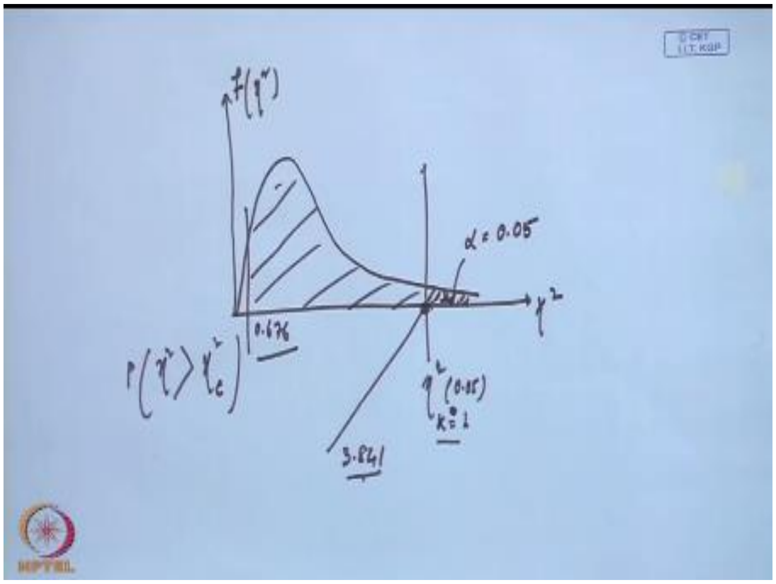
(Refer Slide Time: 37:22)



df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

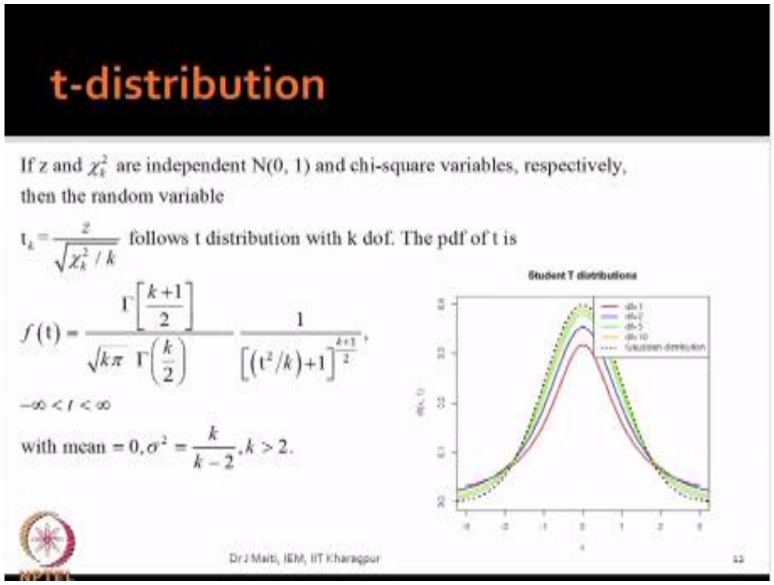
Now if your degree of freedom is 6 and you want the probability that the value that 0.995 then your value chi square value will be 0.676 which will be somewhere here.

(Refer Slide Time: 37:34)



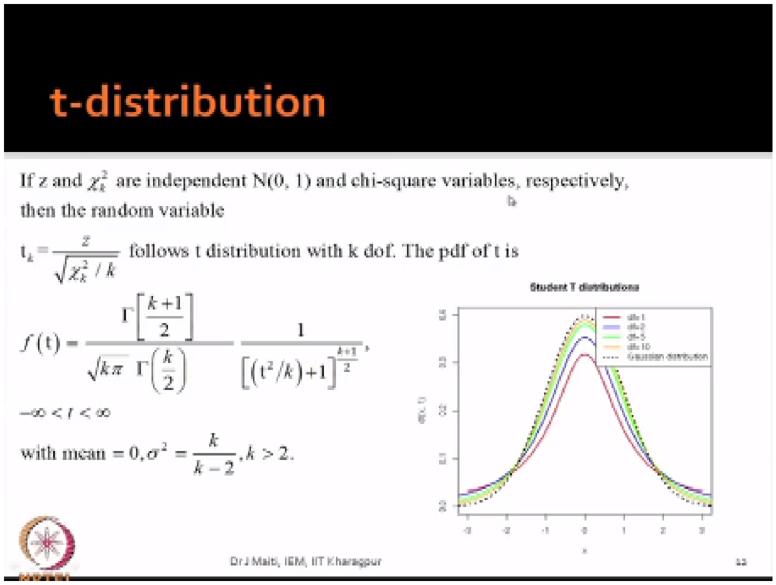
So you are talking about the total probability this side, okay.

(Refer Slide Time: 37:48)



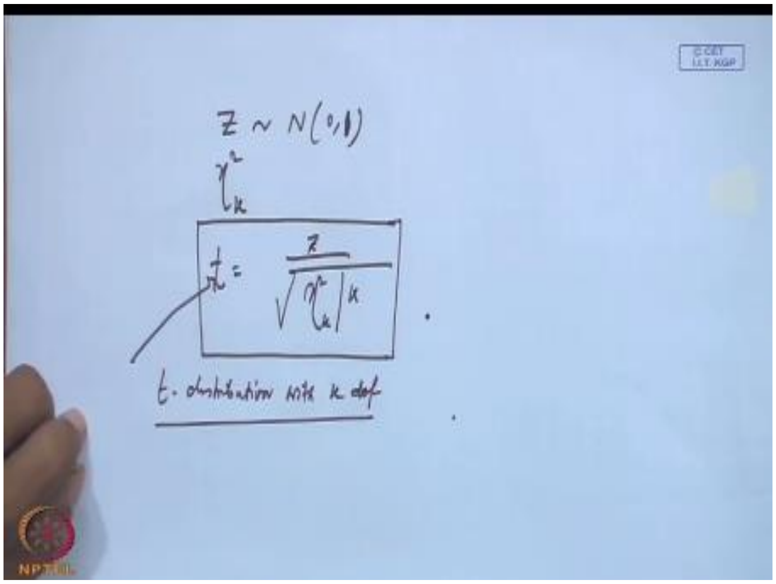
Now let us see t-distribution, t-distribution when you will use t-distribution, first you have to understand when you will be using t-distribution then we will see the uses of t-distribution then again I will show you the table how to use table to find out the different critical values for t-distribution, you see here.

(Refer Slide Time: 38:20)



The z and x_k are independent, normal and chi square variable so we are now considering two variables.

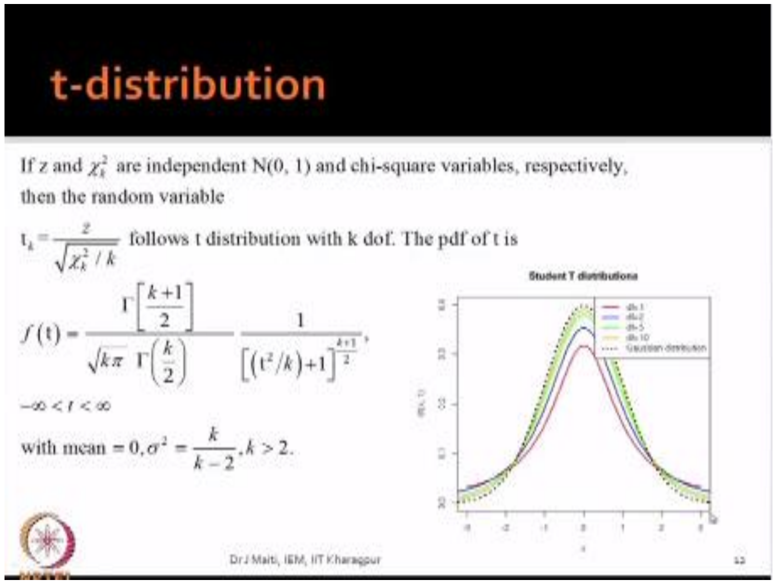
(Refer Slide Time: 38:29)



One is normal variable that you did normal that is z, another one we are talking about chi square let it be chi square with k degrees of freedom, fine. Now you create another variable which is t which is z by square of chi square k/k. Now in your development process suppose when you are developing model if you find that you have created some variable which is of this form that normal by square root of chi square divided by its degrees of freedom, then these quantity the t will follow t-distribution with k degrees of freedom.

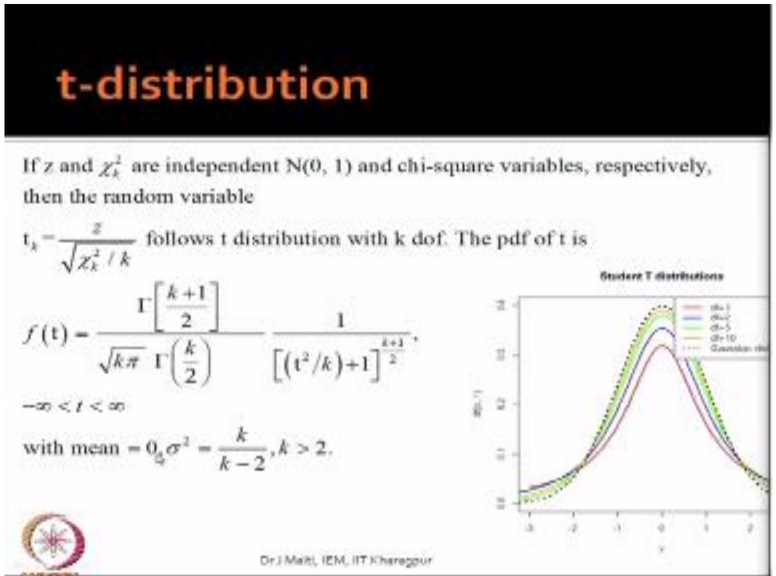
So that mean t-distribution k also, degree of freedom will come degrees of freedom is very important for chi square distribution for t-distribution also.

(Refer Slide Time: 39:50)



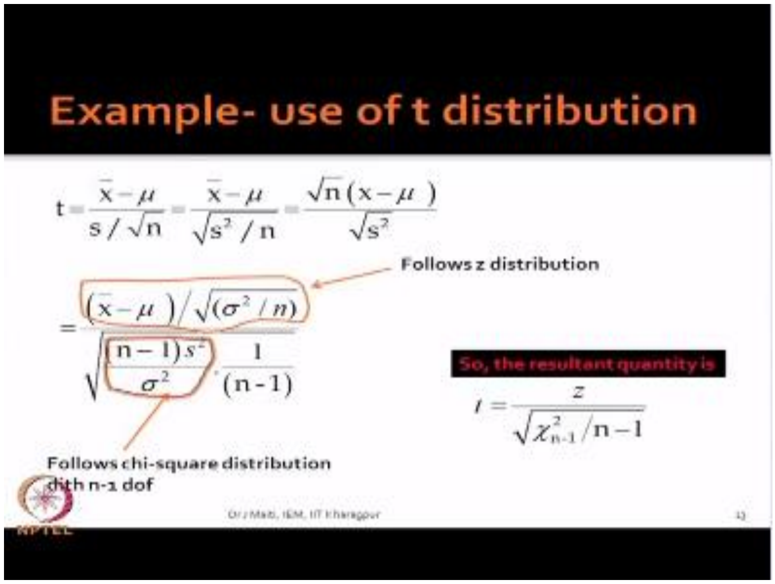
And you see the shape of t-distribution it is similar to normal distribution. When your this chi, the degree of freedom will be infinite that is exactly match with normal distribution what is the mean value of t distribution you are not getting see z what is the main value of z that main value divided by same thing 0 is coming so what is happening here that t distribution main value is 0.

(Refer Slide Time: 40:33)



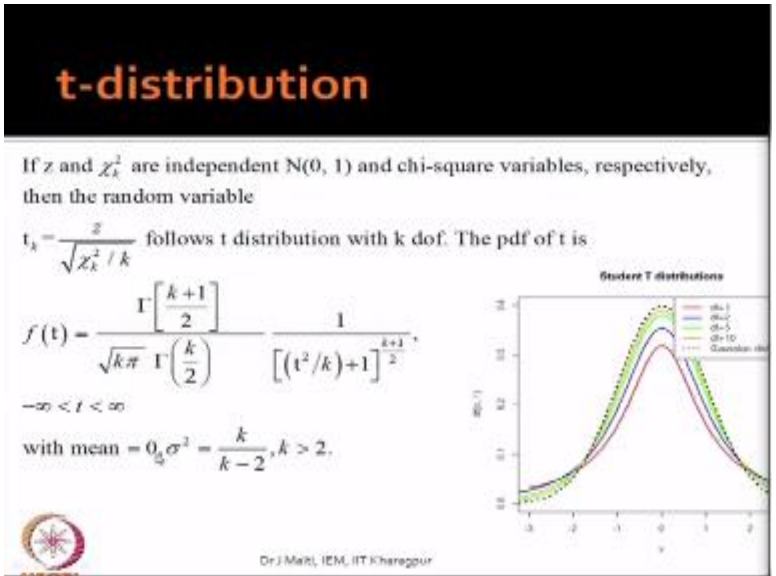
And standard deviation value is √k/k-2 and k must be ≥ 2.And you see that this one here the diagram itself the main value 0 what it is each what it is huge.

(Refer Slide Time: 40:56)



It is there basically it is takes what is the parameter of t distribution what is the parameter of the normal distribution main when σ^2 or σ what is the parameter of t distribution that is the base. What is the variant?

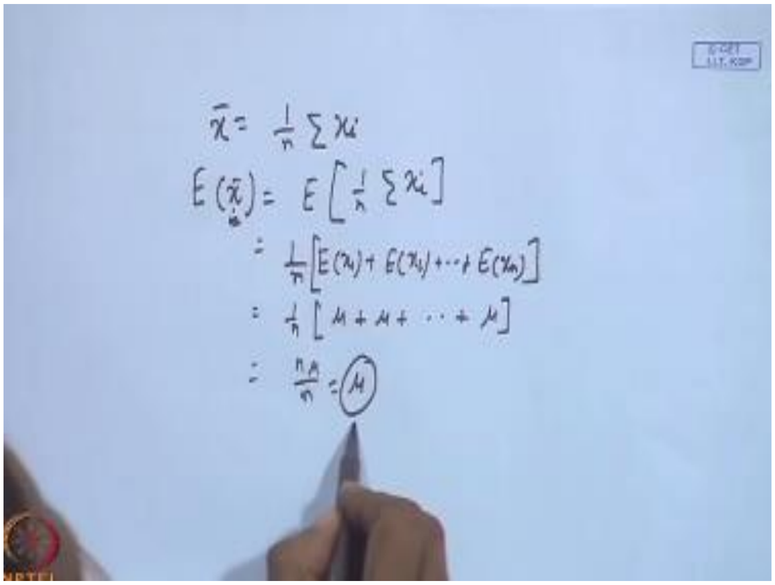
(Refer Slide Time: 41:36)



See everywhere k there you see the function γ k +1/2 k here the k k is there so that mean the parameter of this distribution t distribute the k degrees of freedom and you are getting that why mean it is 0 and σ^2 in terms of k getting me so you will know the parameter from the pdf only in t distribution pdf all the all other values are constant tr definitely the random variable otherwise for other things you see that π is constant 1 and that is that π only k is there everywhere.

The γ distribution is given getting me now come back to this huge what is the huge how do you used this distribution now let us see that.

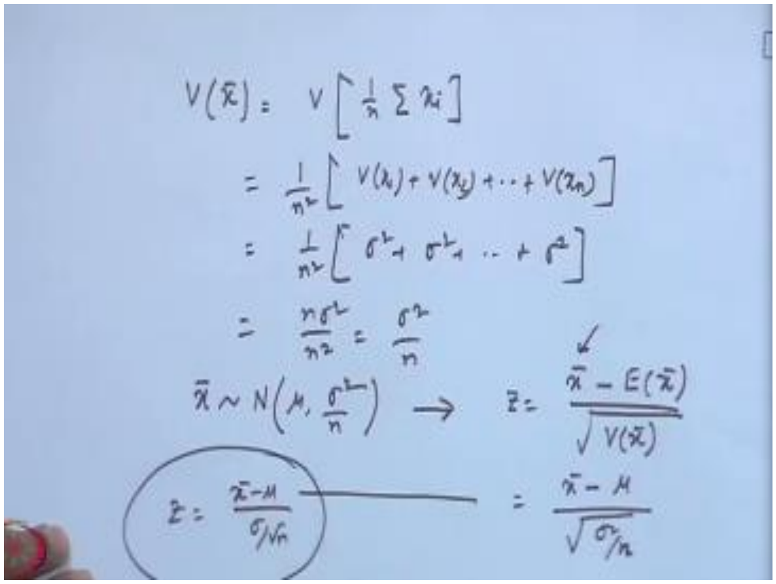
(Refer Slide Time: 42:42)



T is like this that $\bar{x}$- μ / s/ $\sqrt{n}$ you created this type of composition how it will why it will or how it will come I am showing you one thing that when we talk about the distribution of $\bar{x}$ so $\bar{x}$ is 1/n sum total of x_i then I ask you what is the expected value of $\bar{x}$ you said that expected value of 1/ n sum total of x_i so that mean 1/n expected value of x_1 + expected value of x_2 + expected value of x_n what is the expected value of x1 all are μ.

If x is a normal distributed every observation is will be normal distributed with respected mean so it is basically n μ / n that is μ so we say that expected value of sample average is μ now what is will be the variants component.

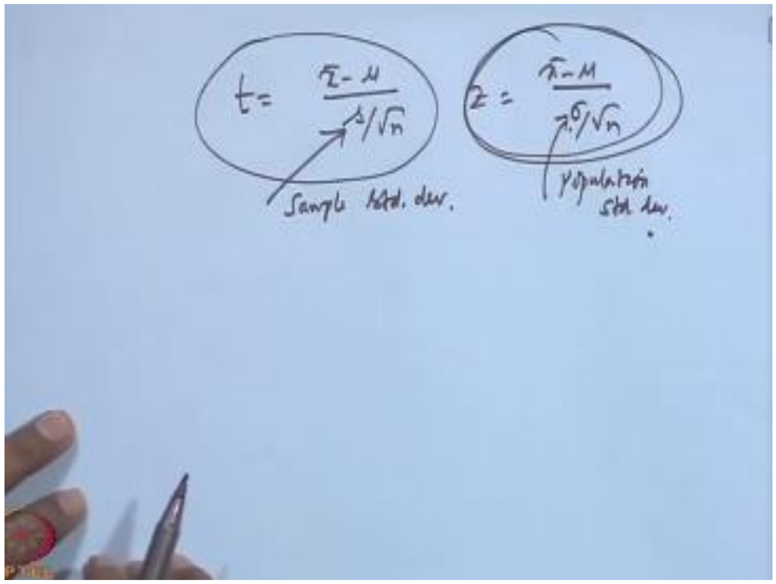
(Refer Slide Time: 44:11)



If I want to know variants of $\bar{x}$ that means this is nothing but variants of 1/ n sum total of again x_i so as I said that variants if you that 1/ n^2 then variants of x_1 + variants of x_2 + variants of x_n x_1 x_2 x_n all are normal random variable with variants σ^2 so 1/ n^2 then σ^{2+} σ^2 like + σ^2 which will be n σ^2/ n^2 that is σ^2/n .So that means using that if I create that $\bar{x}$ I told you if normal a distributed with mean of this mean will be μ and $x\sigma^2$ / n now you are creating a z variable here if you create a z variable here.

So z is nothing but that $\bar{x}$ this is the normal variable minus it expected value of $\bar{x}$ / √ v $\bar{x}$ if any variable random variable is accepted by it is mean and divided by the standard deviation it is z so what is these $\bar{x}$ - excepted value of $\bar{x}$ each μ we have already proved and your variants each σ^2 / √n so it is $\sigma^{2/}$ variants is σ^2/ n. For this is the quantity or other way I can say z is $\bar{x}$- μ/σ/√n so this will follow z distribution, now here in tk what is happening here in this tk you can see that what we are saying.

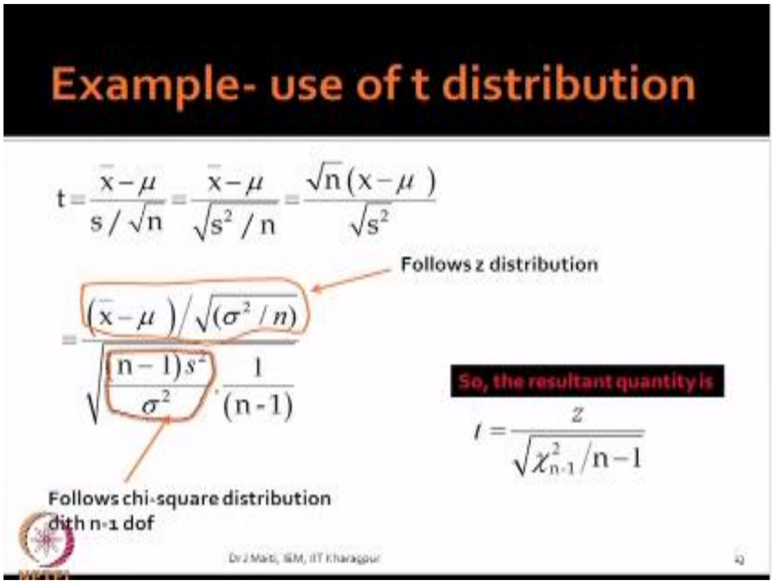
(Refer Slide Time: 46:30)



Suppose the $\bar{x}$ / μ is there but σ is not known you are using s in z case $\bar{x}$ / μ / $\sigma\sqrt{}$ n the σ is the population standard deviation but this is not known instead you are using sample standard deviation in the tks so s is a random variable but here σ is here constant. Now depending on the situation ultimately what level there are different the condition of each sample suggest very large then the same this quantity can be able to normal also but most general case is this quantity each t why?

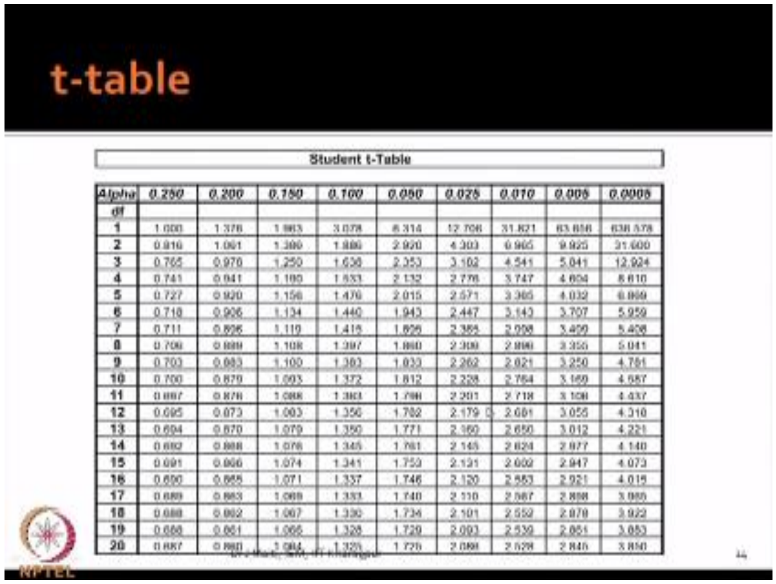
Because this quantity follows to distribution how can you justify that this quantity follows t distribution if you want to justify this then I just return this one the same thing I will written in this manner $\bar{x}$ / μ / $\sigma\sqrt{}$ n and then the within the $\sqrt{}$ n component is n-1 s^2/ $\sigma^2$1/n-1 just menu pollution what is that the nominator and denominator is manipulated with some constants. Then what is this $\bar{x}$ / μ / $\sigma\sqrt{}$ n this is z already we have seen this is z top version what is the bottom position earlier I have shown you that.

(Refer Slide Time: 48:10)



$(n-1)s^2/\sigma^2$ follow chi square distribution, you back you will see what is this one, chi square distribution so you are resultant in variable is that z vale z random by the square root of chi square by its degrees of freedom, so it is t-distribution that is the huge y_A will use t-distribution in this case. Okay, so you see this formula and appreciated because if you can understand this huge problem will solved, that is the huge follows chi square distribution. So the resultant quantity is this, now you will require to use chi square table.

(Refer Slide Time: 49:06)

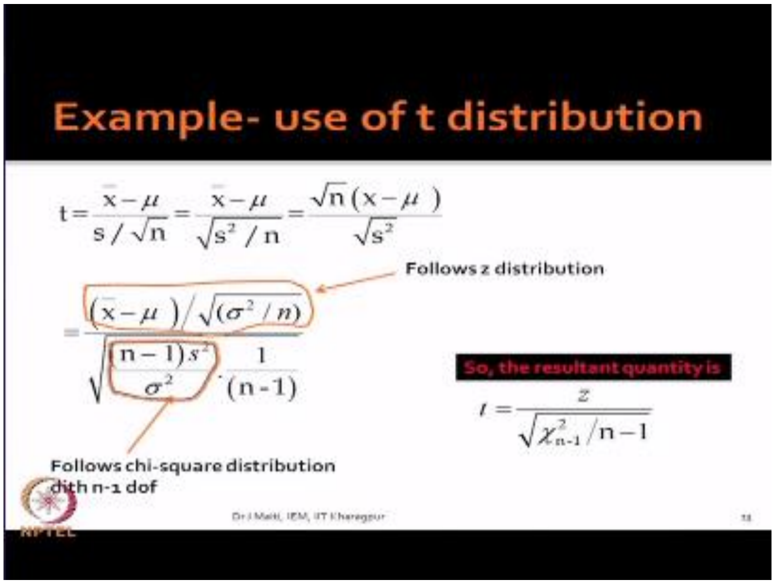


Alpha df	0.250	0.200	0.150	0.100	0.050	0.025	0.010	0.005	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.656	636.578
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850

This is our chi square table, so please keep in mind and in chi square table also there will be degrees of freedom. Degrees of freedom coming into consideration become the parameter is degrees of freedom the parameter of the distribution which will appear. Okay, I think I, you will be able to find out suppose if I say that in my t-distribution with 11 degrees of freedom and what is the probability that means that 10025 if we consider then you will be getting value of 2.228 but if you see the z distribution for the same thing 0.25 there is no need of any degrees of freedom 1.96 will be the z value, okay.

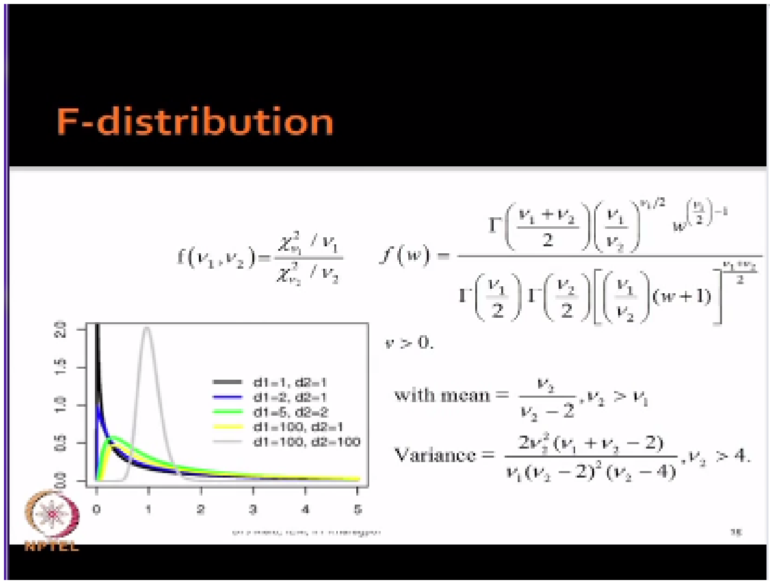
Then come to f distribution, sir one question okay pre distribution that is a what student is appreciated. No that is developed by that gauss it I think I just yes and it developed value as a student so for I know this one student that t distribution is. If that is a instead of using the population variations we are using the sampling, sample variants but population variation is still there that n-1 a^2/σ^2 okay but question is that we but when you compute the t there is know.

(Refer Slide Time: 50:53)



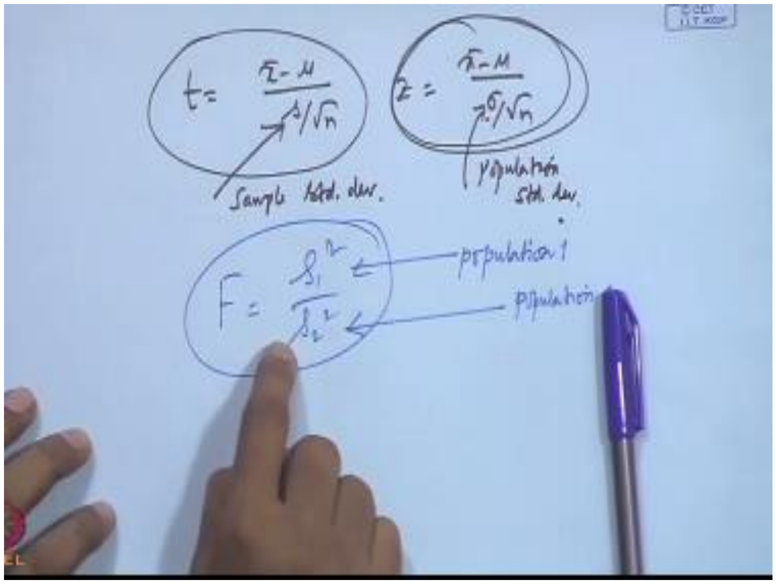
Population variants when you compute the there is no population variants s is there that everything is can be calculated so now I want to know what is the distribution of this and we found that the distribution is t distribution okay.

(Refer Slide Time: 51:13)



F distribution now come to a ratio major what I am trying to we will see here suppose you have two population for the same variable population 1 and population 2 are characterizing variable and you have computed you have taken sample from both the population you have computed your standard derivation or variants for population 1 as well as population 2 you want to compare whether variability in population 1 is different and the variability popluation2 are not what they are basically equal that mean you're basically creating a variable.

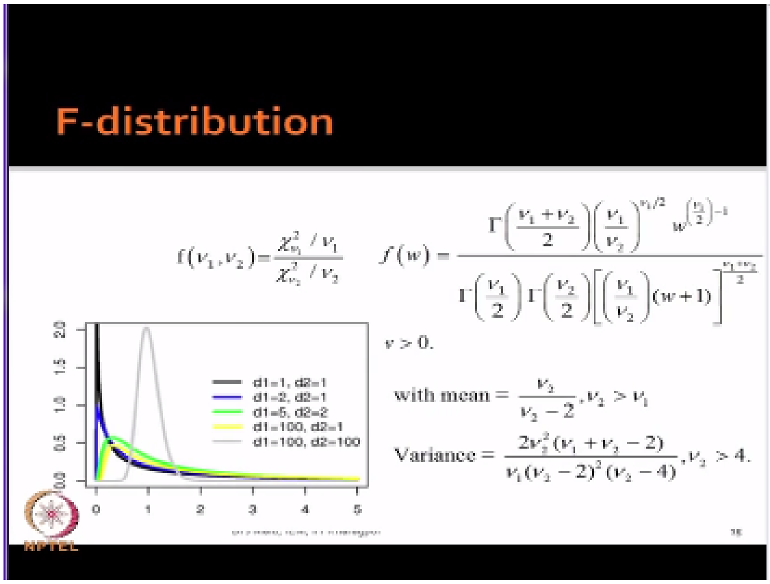
(Refer Slide Time: 51:57)



Which is suppose s_1^2 / s_2^2 this is coming from population 1 this is coming from population 2 weather 2 population are having same variants or not in many models we assume that the population variants are equal for example when you do Enova even in menu work is also we will be seeing that one of the condition for some sometimes we use that population where uses are equal.

That mean this ratio we want to require to know in Enova you will be finding what the use of this enormous use of the distribution in regression also you will be seeing so the ratio.

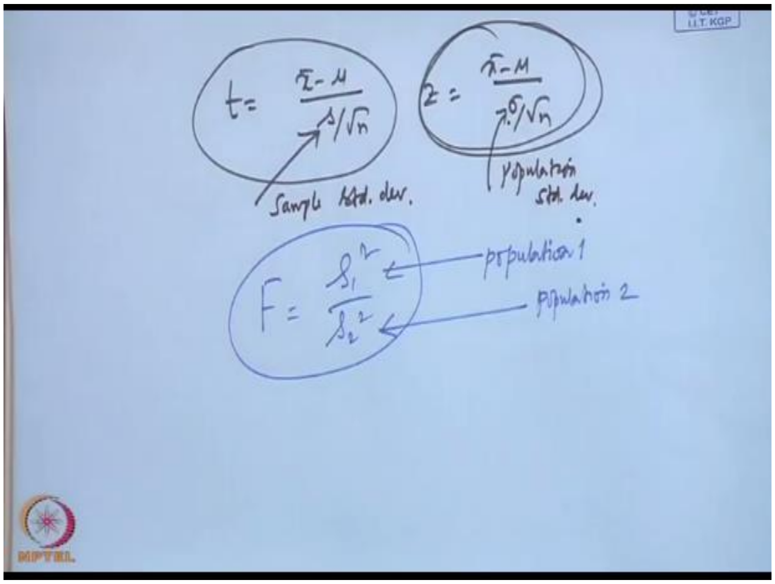
(Refer Slide Time: 52:54)



When your are some quantity you are finding out which is basically which is basically the ratio of 2 k^2 variable divided by their respective degrees of freedom then the quantity basically this is what I mean to say if this is the w this w if we cleared like this the quainter follows f distribution with 2 degrees of freedom enumerative degrees of freedom and d enumerative degrees of freedom why you see the distribution here v 1 and v2 every one is there register gamma function basically are the cases.

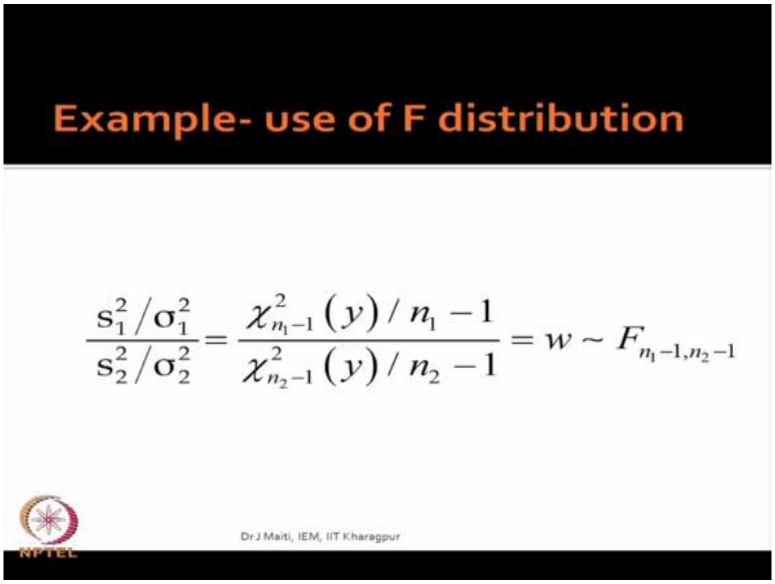
So getting me so these distribution is characterized by numerated degrees of prism and denominated degrees of freedom and when do you use f distribution when you drive any quantity which is the receive of $2k^2$ variables and definitely that ratio and innovated ratio these ratio weighted ratio of variable $2k^2$ when weight is nothing but the degrees of freedom 1/ that degrees of freedom you are getting me for example what I say which I.

(Refer Slide Time: 54:18)



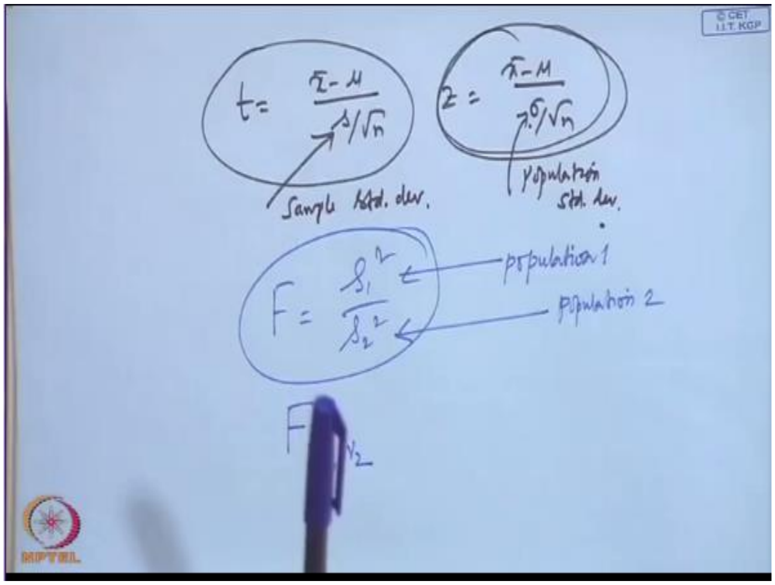
Let us see the use.

(Refer Slide Time: 54:20)



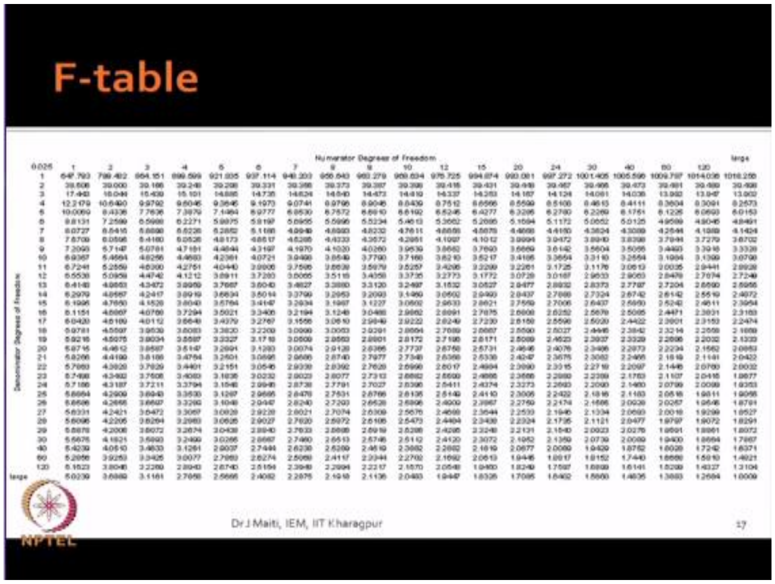
See S_1^2, S_2^2 from to population that population variations and σ1 and σ_2^2 I can write this like this because of all of you we have already seen that n-1 s_1^2/n-1 S^2/ σ^2 follows k^2 distribution so that means this quantity will be k^2 divided by degrees of freedom the denominator quantity will be again k^2/ that respective degrees of freedom correct so this is w and this one is F distributed so in F distribution please keep in mind.

(Refer Slide Time: 55:01)



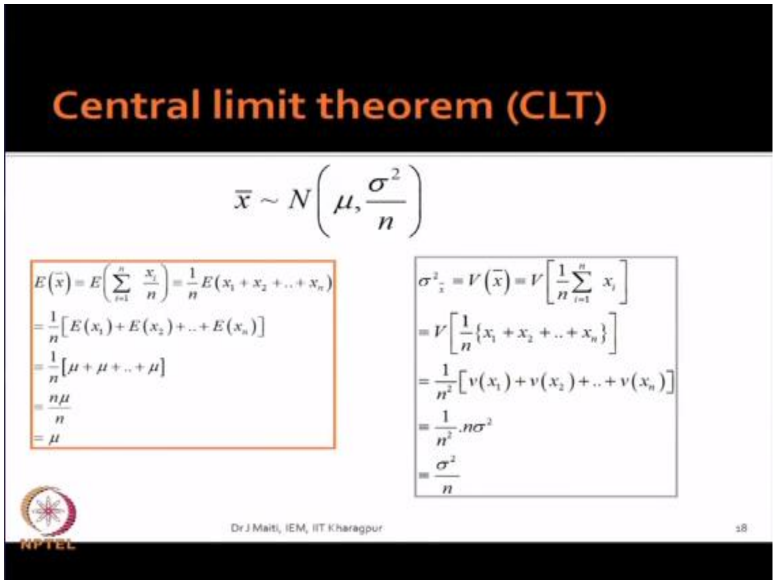
When we talk about f distribution there will be numerated degrees of freedom denominated degrees of freedom so when you go to see the f distribution table.

(Refer Slide Time: 55:11)



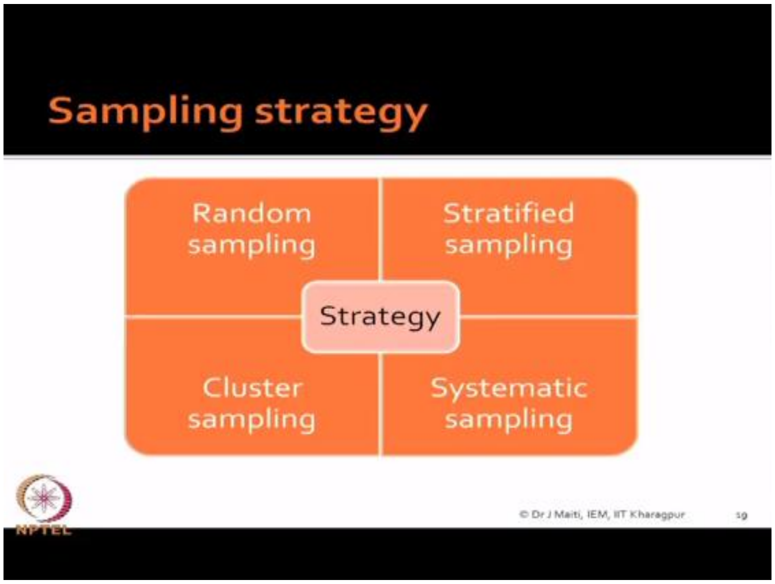
We have to see the two distribute to different degrees of freedom for example if you are interested to know for five numerated degrees of freedom and three degree that denominated degrees of three denominator and five numerated degrees of freedom with a probability value 0.025 then you find out this value you will be getting this value 0.7 that is 7.76 okay so this way it will be used.

(Refer Slide Time: 55:46)



And center limit theorem is the final one for us today again we will discuss next class we will discuss sampling strategy we will finish by this center limit theorem says that if you sample form normal population or other population when the sample size is large then the distribution of $\bar{x}$ the statistical distribution that is the sampling distribution of $\bar{x}$ will be normal $\bar{x}$ is normally distributed with mean μ variants σ^2 / n okay.

(Refer Slide Time: 56:32)



So there are many sampling strategy first one is random sampling, stratified sampling, cluster sampling, systematic sampling, there are some other sampling combination bring all those things basically right we talk about the sample statistic there be you have collected data how you collected data strategy means what method you have adopted while collecting the sample random means you will randomize the collection process during such amount that each and every observation is equally likely to come.

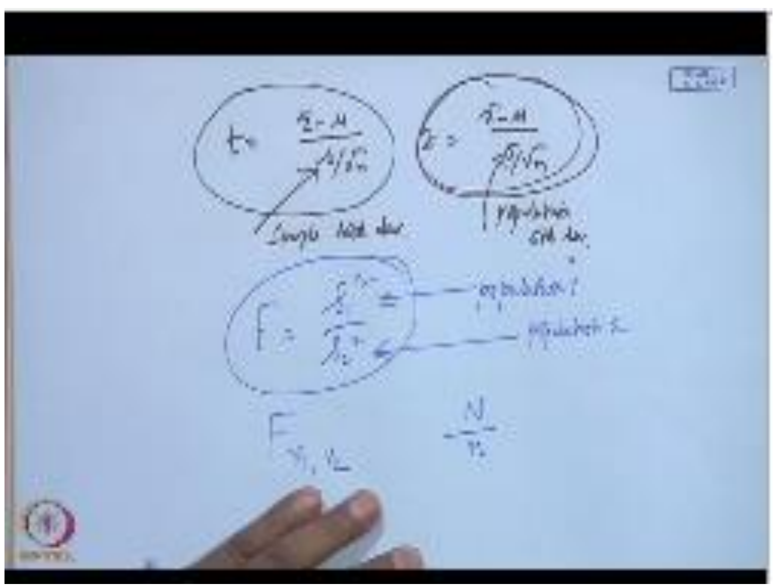
Okay study were resembling is sometimes required what happen suppose you want to see that at different age groups what is the pattern suppose for a particular suppose pattern then you know young people middle age people old people three state you will create and every data you randomly select observation clustor resembling each what clustor resembling is suppose using that the our exit poll this time for.

Punch isolation in Bengal there are in most Bengal there are so many district in each district is a clustor suppose you cannot go for every district and collect sample form is district you can randomize the cluster so that means you will select these two randomization experiment you will see the some of the district suppose you select one district based on randomization and every borders on that district is sampled then that is single state cluster sampling now it may so happen that you may go for three four or more.

District at random and again each of the district you will collect data from individuals selecting randomly getting me first one is district number of district is here you select one randomly one district sample everybody that is single state cluster sampling second is first randomize the selection of the district take few district and again in each district you randomize the selection of individual that is two cluster now if you again in district level you can go for city also they are one more randomization is possible.

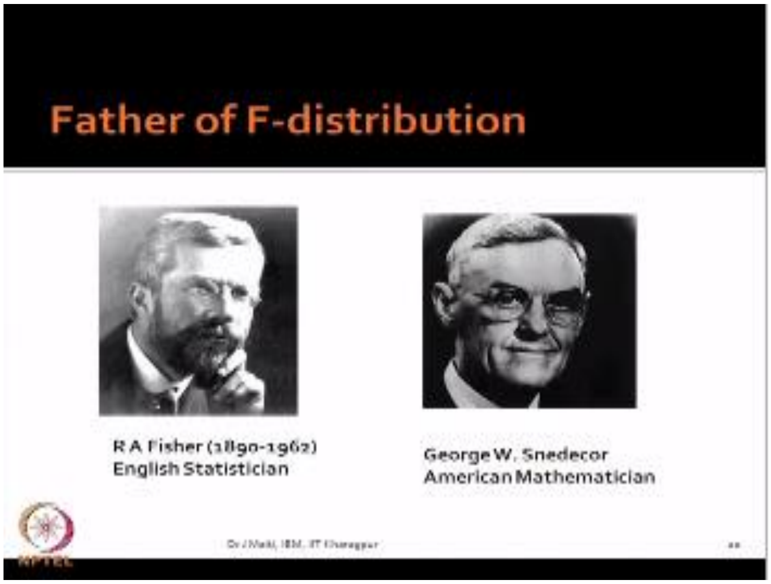
So that multi state cluster will come last one is systematic sampling systematic sampling means you basically follow an order. For example, suppose the 10^{th} day will set a target for example that k^{th} will collect in one sample correct suppose the population size is N and sample size in n.

(Refer Slide Time: 59:19)



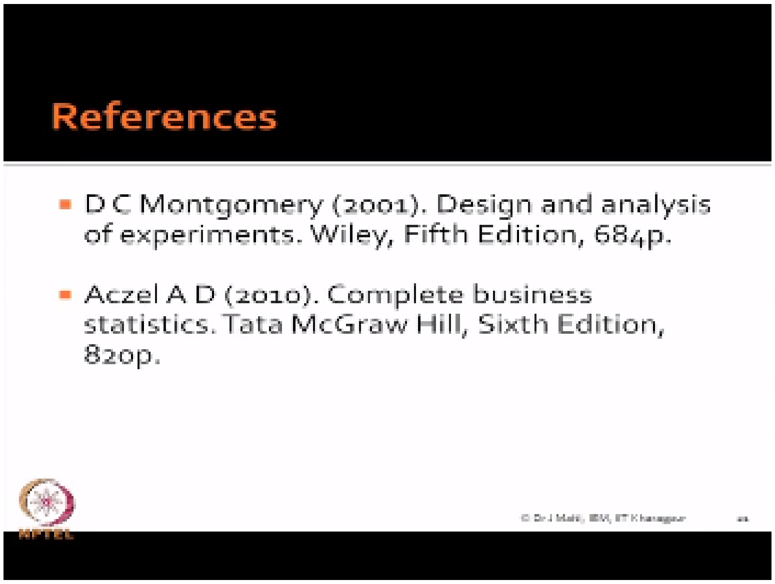
So that u and 2 suppose this quantity is k you want to collect to collect a sample of n so what you will do from the first N observations you may be at the L^{th} point you observe then you go on adding K the second observation will after you will select the $l+k^{th}$ item then l+2k item with that way you will collect the n sample, getting me. So these are the strategies you go through some books and you know that as I told you that the distribution will be heavily used an F distribution all of you know that.

(Refer Slide Time: 01:00:03)



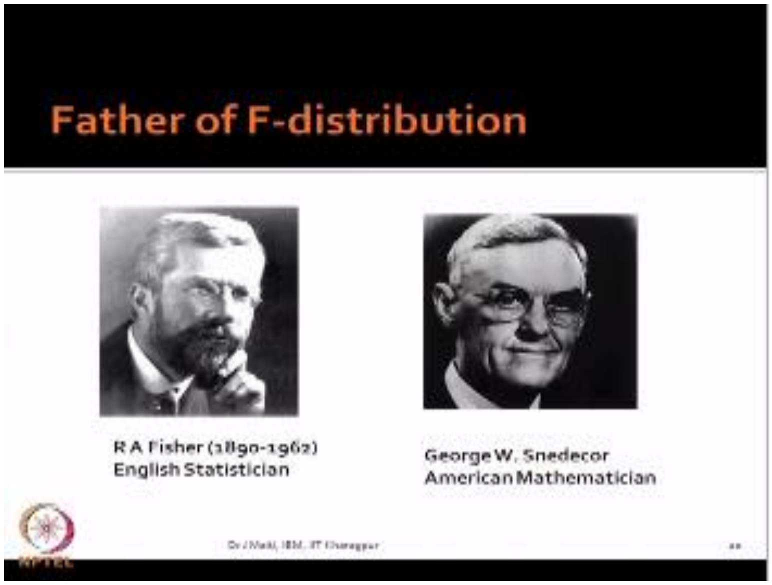
R A fisher that English statistician is contributed a lot in agricultural sector and George W. Snedecor American mathematician they are the pioneer in the developing the F distribution ok and next class I will bring that student distribution particular person

(Refer Slide Time: 01:00:27)



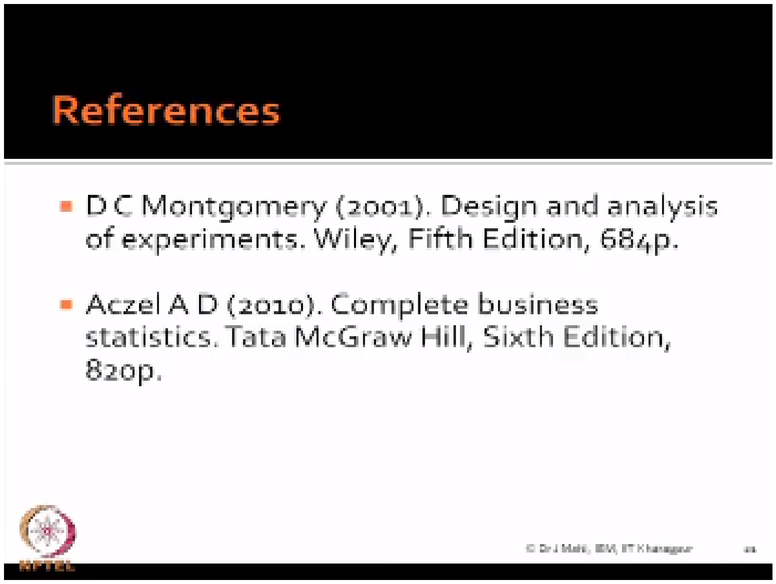
This will you can remember throughout as you know fisher and Snedecor that they have basically developed the F distribution.

(Refer Slide Time: 01:00:39)



Snedecor and fisher RA fisher is only one name.

(Refer Slide Time: 01:00:46)



Montgomery and this Aczel AD complete business statistics this both books are very, very available ok and for multi variate statics Johnson and Witcher and here at all that you told in the beginning finish ok ,so next class we will discuss estimation particularly confidence in terminal. So I think again it will be on Tuesday coming Tuesday 7'o clock evening Tuesday 7 to 9 okay, okay thank you very much.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in