Business Analytics for Management Decision
Prof. Rudra P Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture - 19
Inferential Analytics (Confidential Interval)

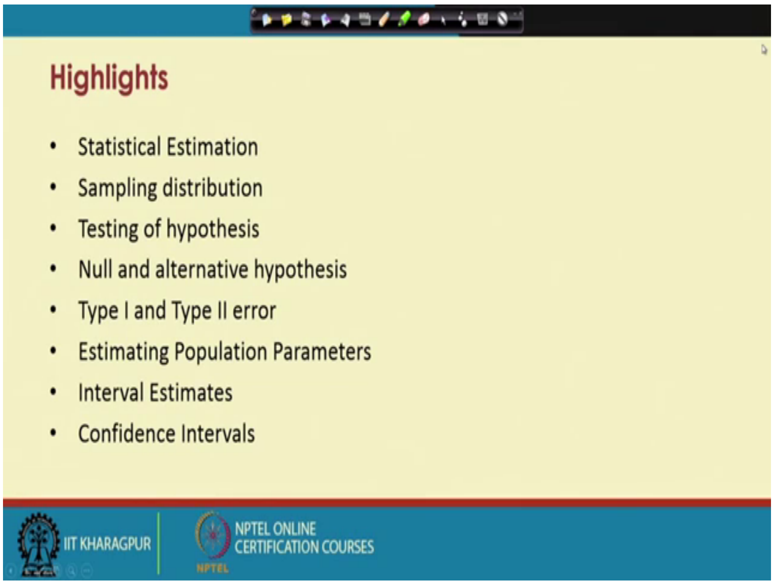
Hello everybody, this is Rudra Pradhan here. And welcome to BMD lecture series. Today we are going to discuss on inferential analytics, and that too the topic of discussion is on confidence intervals. In the earlier lectures we have already discussed hypothesis testings by using various test statistics.

Typically, we have discussed the problem related to z distributions, problem related to t distributions and problem related to chi square distributions. So, the idea behind all this discussion is with respect to comment on population parameters. So, we start with the game by fixing population parameter, and then with the help of you know sample statistics we have to validate the population parameters.

So, that means, the idea is here to check whether this sample statistic is unbiased estimator to population statistics. If not, what is the kind of you know bias biasness in the particular you know a systems or any kind of you know problems. So, the ideal rule is that your sample should be you know a sample information should be converged towards you know population informations. If there is a huge difference or these huge deviations then the decision-making process will be getting affected. So, sometimes you know if there is a huge deviation then the population parameter need to be changed.

So, obviously so, here the idea is a we have to create a kind of you know confidence interval with respect to population parameters. Earlier we our idea is a to test the you know kind of you know things, and then comment about the population parameters and heres we have to create a kind of you know concept called as you know confidence intervals. So, let us see how is this particular you know structures. So, in the process of you know these are all actually a lecture highlights.

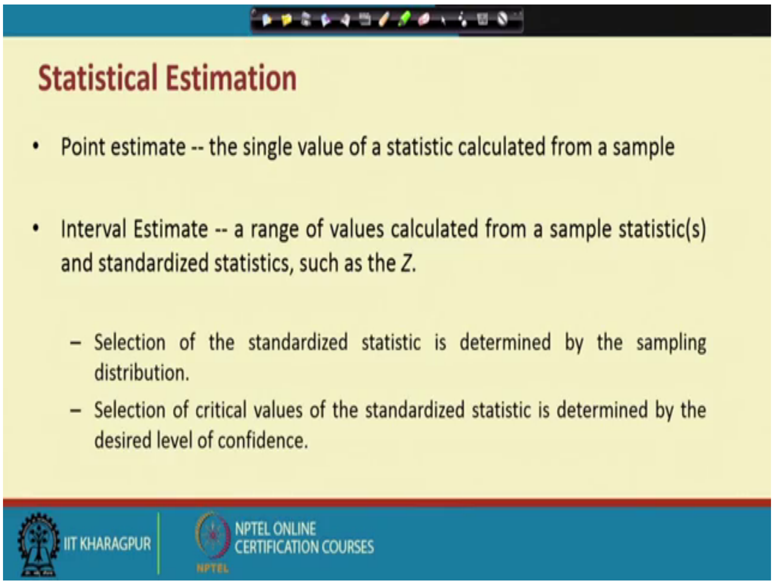
(Refer Slide Time: 02:20)



And this same kind of you know highlights, we will continue and because it is actually a part of the hypothesis testings. Earlier, structure is a like you know to validate the kind of you know concept, but here is with the given situation we have to create a kind of you know confidence interval for population parameter.

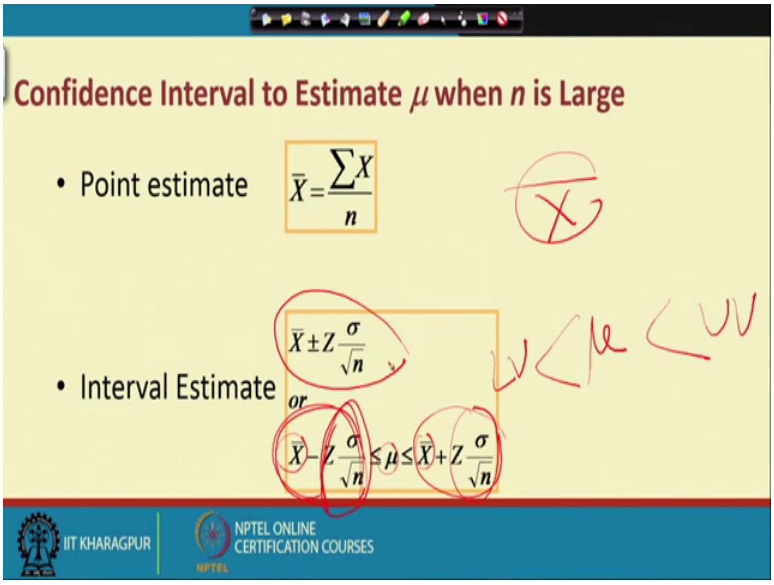
So, for instance a population parameter is nothing but called as you know mu. So, we like to know what is the lower range of the population parameter, and what is the upper range of the population parameter and by using different test statistics. So, that means, we will we will solve some of the problems by using a z distribution t distribution and chi square distributions. So, let us see by using all this test statistic how we can create a confidence interval for a population parameters. So, for that so, we first like to know you know the structure of you know estimations, and by the way there are 2 different estimation process; point estimation and interval estimations.

(Refer Slide Time: 03:16)



Point estimation is nothing but actually single value, which is obtained from the samples, but interval estimation is nothing but actually, it is a range of values calculated from a sample statistics. And so, the standard example is you know standardized statistic that is nothing but called as you know z statistics. So, here is using this you know standard distribution, we have to find out the kind of you know confidence interval for the population parameter.

(Refer Slide Time: 03:47)



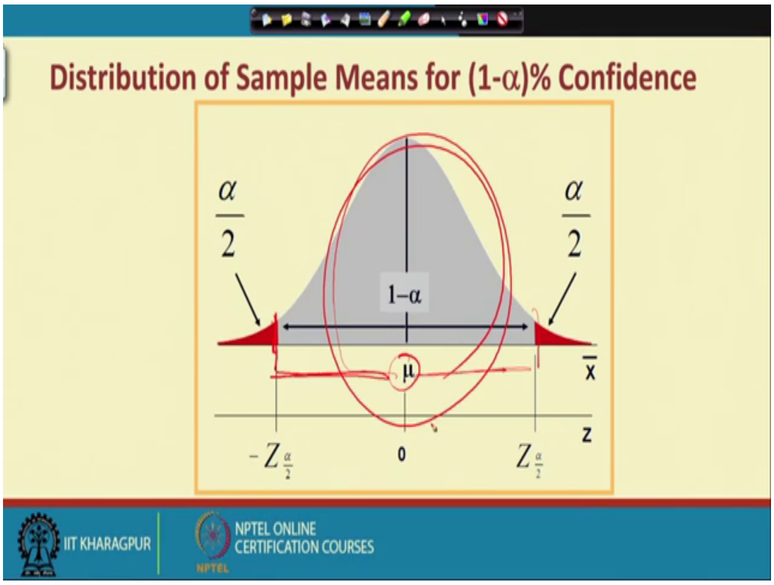
So now, corresponding to the structure of point estimation and interval estimation, the standard example is the mean of a particular sample say x bar. So, that is nothing but actually summation x by n. So, that is actually it is a single value that describes the kind of you know entire series, and in the case of you know interval estimations, ve to create a confidence interval for this particular you know mean value.

So, typically the idea is here is. So, we have actually means mean value is x bar, and then corresponding to this mean value. So, we like to create a population parameter and that too lower limit of the lower value of the a population parameter and upper value of the population parameter. So, that is what the kind of you know structure which you want actually in this particular you know lectures. So, this is what the upper bound and this is what the lower bound. And this is sample statistics, and this is also sample statistic and this is population parameter. And this is what the actually z distribution, and this is what the z distributions.

So now in sum so, when you have actually x bar, by using this particular you know formula; so, you can you can able to find out the a you know lower value, and you can find out the upper value so; that means, the lower value will be x bar minus you know z sigma upon root n. And the upper bound will be a x bar plus z into sigma by root n. So, this is how we have to create a kind of you know confidence intervals. So, before we go for you know the kind of you know testing and the kind of you know example. So, let me give you the kind of you know snapshot what is exactly the confidence interval, and how is this particular you know structure altogether.

In the next slides so, I will highlight the kind of you know concept about the confidence interval.

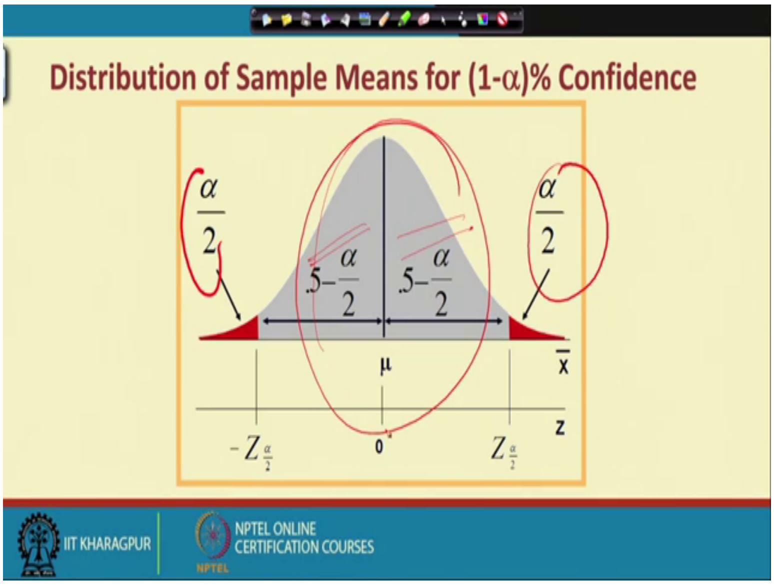
(Refer Slide Time: 05:59)



So, this is what actually confidence interval, and the what we are actually looking to know so, the this range and this range and this is the structure entirely called as you know confidence interval. And this is what the population parameter and population parameter, this is the left-hand side regions and this is the right-hand side regions. And now the with the help of you know hypothesis testing again. So, we can you can in a position to create a confidence interval for the population parameters. So now, earlier with respect to population parameters we like to test the process, and with the sample statistic we will give comment on the population parameter.

Here we are adding something extra that you know with respect to the population parameter. So, we have to find out the kind of you know confidence interval. So, that means, sample statistic will be used to create a confidence to create a kind of you know confidence interval for the population parameter. This is what the kind of you know objective in this particular you know lecture. So, let us see how it is actually coming a you know in a kind of you know some kind of you know example.

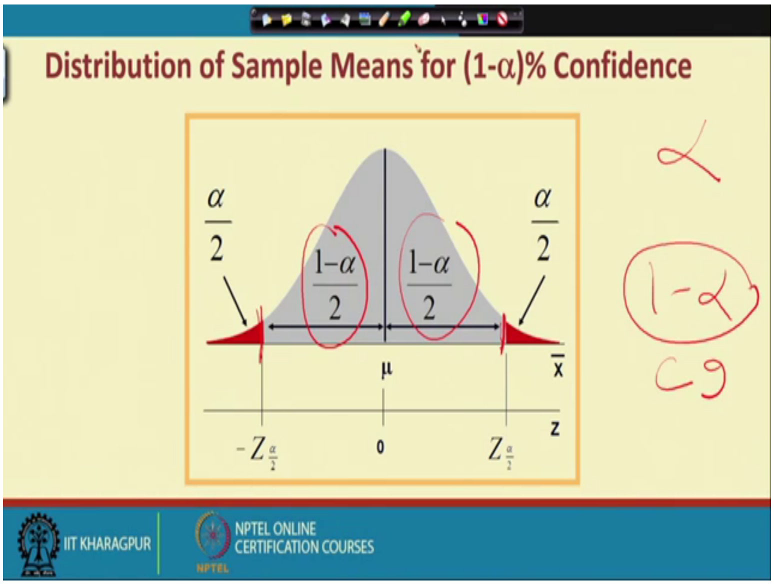
(Refer Slide Time: 07:12)



So, this is how the typical range. So, since we are you know giving emphasize on you know both the sides. So, you know with the fixing alpha is a type one error, then alpha by 2 will be coming this side and the remaining alpha by 2 coming this side. So, then this is the confidence interval the entire regions. And since it is a symmetrical distributions then you know 50 percent will come this side and 50 percent will come this side.

So, actually we have no business about the 50 percent observation and 50 percent observation. So, we can we can you know get to know you know more details once you take a you know standard examples right. So, let us see how is this particular you know structure altogether.

(Refer Slide Time: 08:00)

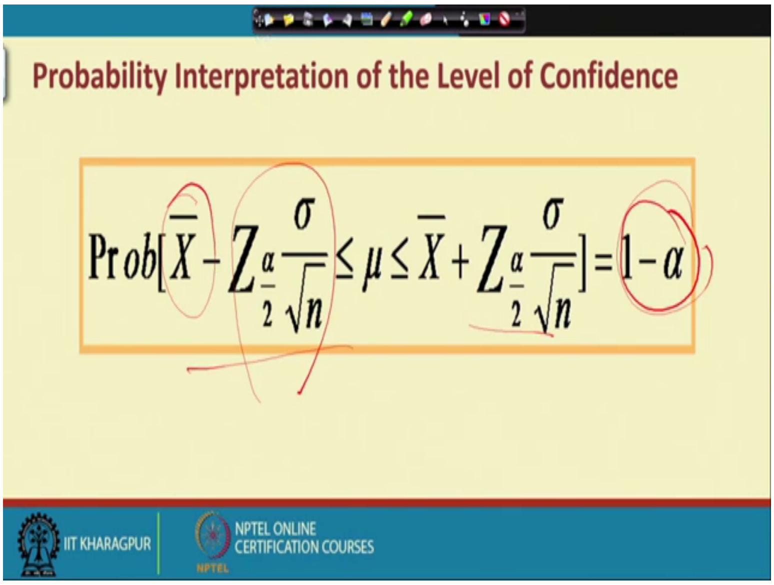


Then n this case it is a same things, but we are putting actually I mean say since you know total probability is exactly equal to 1. So, when you are putting alpha this sides; that means, you have 50 percent of alpha this side and 50 percent of alpha this side, then you know the remaining persons will be again divided into 2 equal parts because it is a bell-shaped curve and as a result. So, 1 minus alpha by 2 will become this side and 1 minus alpha by 2 will be this side.

So, that means, you know alpha in. So, there are 2 sides actually alpha and 1 minus alpha 1 minus alpha actually called as you know confidence intervals. And then alpha is the kind of you know rejected zone. So, since it is A 2 say you know 2 tailed kind of you know game. So, both the sides the appearance will be alpha by 2 and alpha by 2, and then the remaining 1 minus alpha. So, it will be 50 percent left to the mean mu, and 50 percent you know right of the mean mu. So, that means, corresponding to the z distribution 0. So, 1 minus alpha by 2 will be in the right side and it will continue up to. So, these this range and again this sides it will be continue up to this range.

So, as a result. So, it will be give you a kind of you know confidence intervals. So, let us see here is you know how it is exactly coming a.

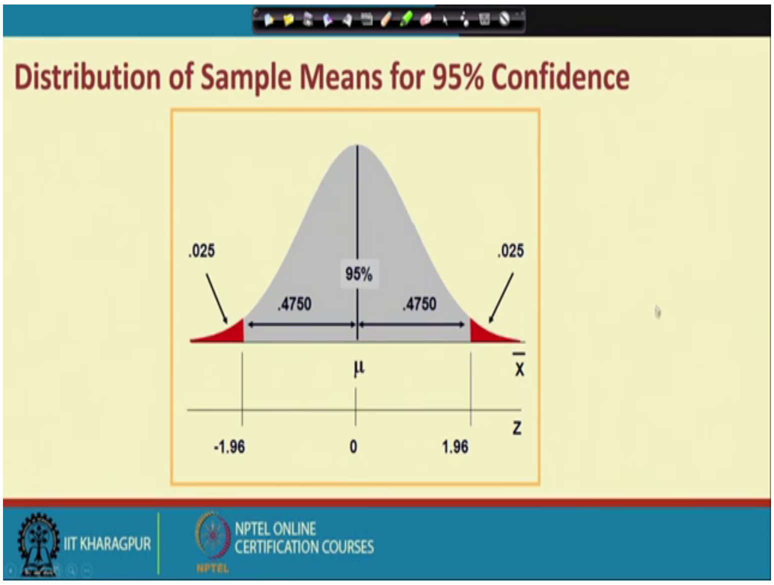
(Refer Slide Time: 19:30)



So, this is what actually the kind of you know structure. So, since it is a all together you know total probability equal to 1, and corresponding to alpha. So, the probability of this particular confidence interval will give you the you know kind of you know concept called as you know 1 minus alpha. So, that means, the 1 minus alpha all together it is nothing but actually probability of the entire you know zones.

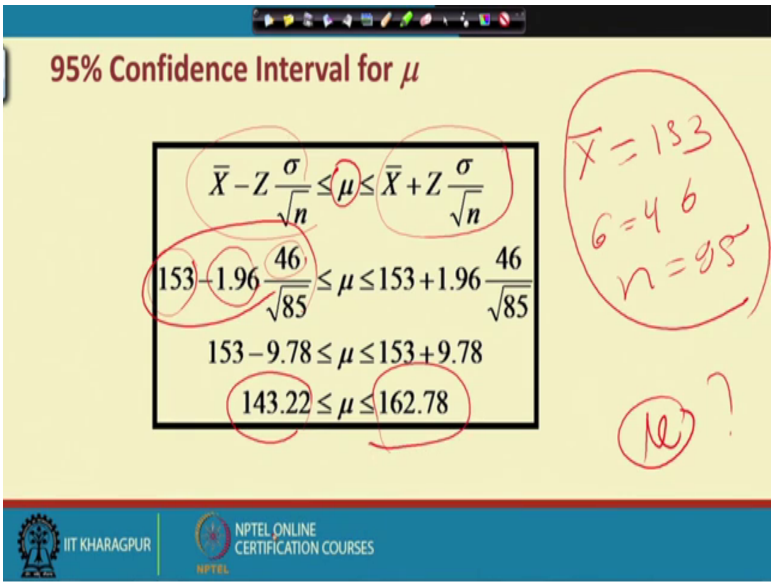
So that means, here is. So, once you find out the you know mean value of the you know sample statistics and you know follow the z distribution, then the lower bound and the upper bound with the entire confidence interval will be represented by 1 minus alpha. So, in order to understand you know better. So, let us take an you know live examples right. So, we will take a live examples and then we can discuss the details right.

(Refer Slide Time: 10:22)



So, this is same things. So, in this case actually. So, what we will what we will do here in this particular examples, we are fixing actually 95 percent of confidence interval, but as a result by default alpha will be represented by 5 percent. And since you know both the sides we are targeting then left side will be 0.25 and right side will be 0.25. And then out of you know 95 percent, 0.475 will be in the left side again 0.475 will be in the right side.

(Refer Slide Time: 10:54)



Then it will be giving you in a you know kind of you know structures. See here is so, the so far as a confidence interval is concerned. So, this is actually population parameter. This is population parameters, and then this is population parameters. And then so, the lower bound will be this much, and the upper bound will be this much. So, that means, what we need actually to create a confidence interval, we need first sample statistics, and then the z value and the sigma of this particular series and provided a sample size, right. So, that means, having let us in this particular examples, we are fixing let us assume that x bar equal to 153 and standard deviation equal to 46 and n equal to 85.

So now having this this much of informations, then our target is you what should be the range of you know population parameter mu. So, that means, you know with the given sample information we are trying to comment on the population parameter that to create a kind of you know confidence interval. In the earlier lectures, what we have done actually.

So, with respect to population parameters we obtained actually sample statistic, and then we get the (Refer Time: 12:09) you know test statistics, then we like to validate the test statistic, whether it will be in the rejected zone or you know accepted zone that to acceptance of you know null hypothesis and acceptance of you know alternative hypothesis; that since both cannot go up simultaneously one will go at a times, and that will give you the a kind of you know status about the population parameter.

But here the game is little bit you know different. So, that too with a sample statistic and the choice of a particular test statistic we have to create a confidence interval for the population parameter.

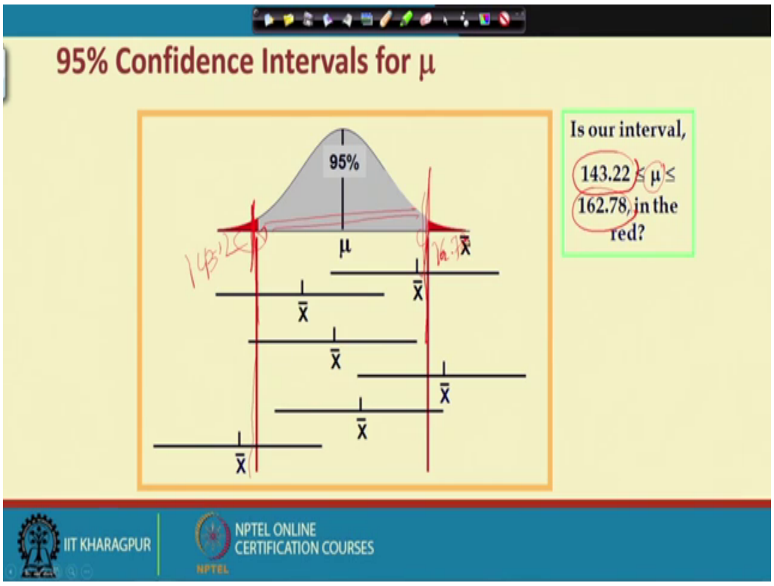
So, the standard example is here. So now, in these particular examples we are using a you know z statistics, and that too for this confidence intervals the requirement is you must have a sample statistics; that is, x bar and then the z since it is a 95 percent confidence intervals. So, you go to the z tables at the 95 percent you know I means when confidence interval is 95 percent. Then the z value will be coming actually 1.96, then we are assuming sigma equal to 46 and n equal to 85.

So, as a result. So, the lower after simplifications, the lower bound will be coming you know this much that is 143.222 and the upper bound will be coming actually 162.78. So, that means, actually since you know many you know managerial kind of you know problems or you know business problems. So, targeting to a kind of you know single value is a very difficult.

And so, far as you know decision making process is concerned. Let us say we will we like to have a kind of you know confidence interval, and within that confidence intervals. You know, the decision can be more effective and more efficient that is how you know most of the problems. So, corresponding you know sample informations we have to create a kind of you know confidence interval for the population parameter.

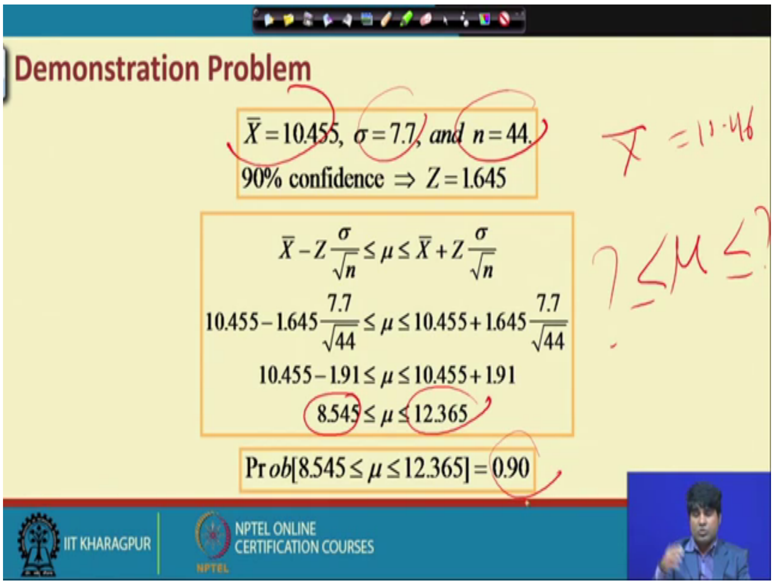
Likewise, we can have a another examples.

(Refer Slide Time: 14:18)



And then so, this is how the you know looks likes the 95 percent of confidence interval. And here is corresponding to the previous examples. So, the lower limit will be this much and the upper limit will be this much. And then you know your you know mu will be as usual it is given there. So, accordingly so, the red in the red area, this is what actually the you know critical value for the lower range, and this is the critical value for the higher range. And this is what actually confidence interval. So, this persons is nothing but called as a 143.22. And this person is nothing but called as a 162.78 right. So, this is what the a kind of you know confidence intervals.

(Refer Slide Time: 15:08)



So, we will take another example here. And earlier examples we have taken 90, 95 percent of the confidence interval, and here we are taking in 90 percent of the confidence interval. So now, if you go to the z tables, then the you know for 90 percent of confidence interval z statistic will give you 1.645.

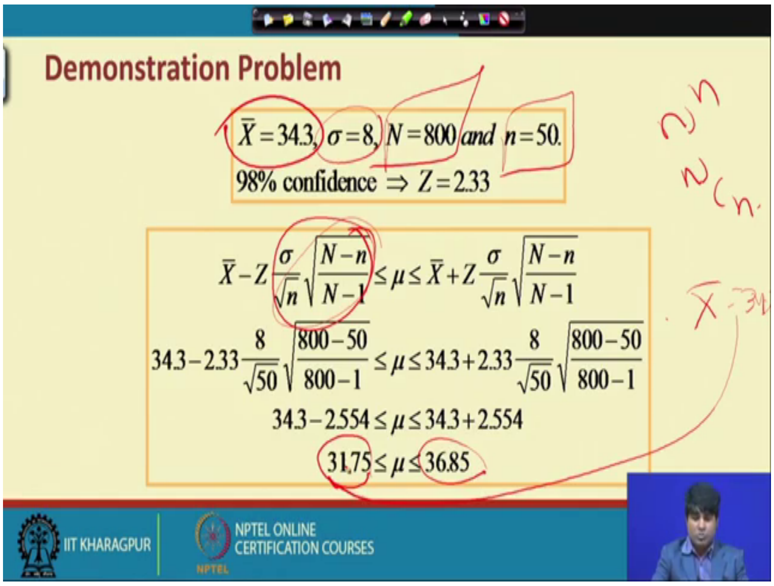
So now, again assuming actually mean value standard deviations and n, and then again you will find actually what is the population you know parameters confidence interval so that means, actually so, having x bar equal to 10.46. So, what is actually confidence interval for you know population parameters and that to the lower limit and the upper limit.

Again same structures, by using the sample statistics, you are actually predicting the population parameter. That too you are creating a kind of you know confidence interval. Because you know sometimes you know to capture the entire population is very difficult. But with the help of you know sample statistics, we can create a kind of you know range that to fix a you know lower range and you know upper range, and then accordingly the decision-making process can be more efficient and you know can be more effective.

So, that means, actually after you know after putting all these value, you can you can calculate the lower limit is this much and upper limit will be this much, then the probability between this particular you know series will be nothing but actually 0.90.

So, that means, actually we are we are going in a reverse kind of you know directions. So, here actually we are fixing first you know confidence interval, and obtaining the test statistics; by using the test statistics and the sample statistics we are creating a kind of you know interval for the population parameter. That is the objective behind this particular you know discussions.

(Refer Slide Time: 17:05)



So now so, a this is another kind of you know examples. And here say you know we are putting actually a kind of you know things like you know another example you are putting x bar equal to 34.3 and standard deviation is equal to say 8, and then n capital n equal to 8 800, and small n equal to 50. And a ninth we are putting 98 percent confidence interval and at the 98 percent confidence interval z is coming 2.33.

That is actually from the z table and here something actually you know corresponding to previous 2 problems. This problem is a little bit you know different. Here actually is instead of you know sample n equal to 50 here, there are 2 samples that is n equal to capital n equal to 200. And small n equal to you know 50. So, I will I will and let me first you know give you the kind of you know clarity here. So, our actually structure of you know testing is like this.

Sample sampling distribution and then population then you know most of the instances we you know touch sample to populations. But you know sometimes you know we can actually compare with the sampling distribution to the populations; when you are you know doing actually sampling distribution to populations, then you know the drop this sample from the population may be finite and may be infinite.

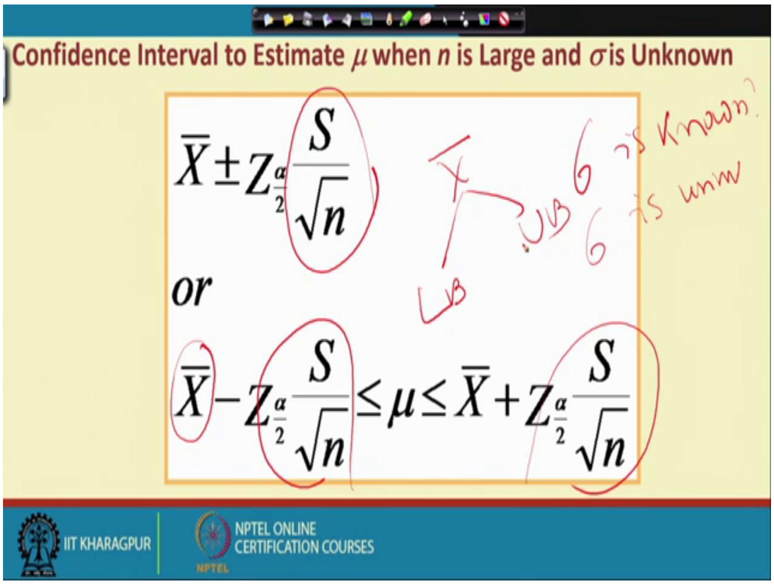
But again, under the finite; the structure is actually with replacement and without replacement. So now, in the case of you know without replacement, the standard a be a standard you know means a population standard deviations will be actually a effecting. So, that means, in the in the in the case of you know sampling distribution to the population. So, mean of the sampling distributions sampling distribution will always converge to the population mean. But you know a variance of the sampling distribution will not actually exactly equal to population variance.

So, the standard adjust adjustment between a variance of the sampling distribution to the population distributions will be like this. So, this is the kind of you know adjustment factor, and that is the case with you know the there will be difference with replacement and without replacement. And in the case of a 3 replacement the drop the sample total sample point will be n, to the power n, and with replacement without replacement it will be n c n. So, this is the structure adjustment factors you know while you are you know connecting with the variance of the sampling distribution to the a populations.

So, as a result after this adjustment then you have to actually connect with the values of the variables that is x bar. And then standard deviations and then putting the value of you know population sample and you know small n equal to 50, then you will be find the lower bound will be this much and the upper bound will be this much. So, that means, corresponding to mean x bar equal to 34.3. So, we are in a position to say population parameter can vary actually between 31.75 to 36.85. So, that is actually the kind of you know structure we are predicting. So, that means, here the idea is in this in this particular confidence interval, using sample statistics you have to create a confidence interval for the population parameter.

So, that means, technically population parameter is not actually givens. So, with sample statistics we are you know predicting the population structures. So, if population is givens then; obviously, our job is just to you know give comment to the population parameter subject to you know sample information. But sometimes with the help of you know sample informations, we can actually create a kind of you know confidence interval for the population parameter. That is what we are actually doing here.

(Refer Slide Time: 21:14)



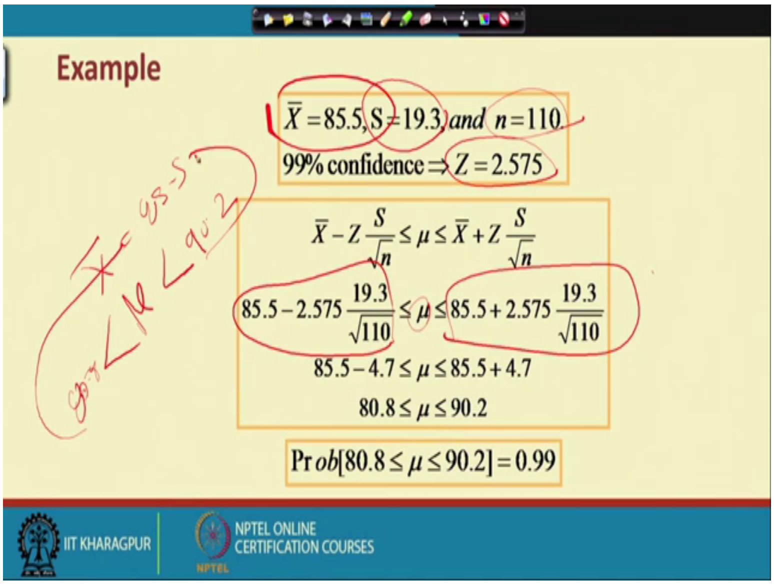
Now so, this is actually the standard structure about the confidence interval actually, there are you know 2 different ways when you are using you know z statistic to create a confidence interval.

So, in one case here you know standard deviation you know sigma is known to you; in one case sigma is known, and the previous 2 examples where you we have actually discussed the issue were sigma is known to you. But there is a you know sometimes some you know you will find the issue you know where you have to calculate the confidence interval for the population parameter where sigma is actually unknowns sigma is unknown. So, this is the case where sigma is unknown, then you know this the same sample information can be used to get the get the actual sigma value. So, in that case the standard adjustment will be s by root n.

So, since sigma is not you know available. So, from the sample informations you have to find out the sample variance, and that sample variance will be used as a proxy for you know this you know sigma components. So, rest of the things are you know more or less sames. So, again. So, the once you get the sample statistics, and then with the help of this particular adjustment you will be find the lower bound and you will be find the upper bound. So, that means, again having x bar x bar. So, what is the lower bound for you know population parameter mu and what is the upper bound for this you know population parameter mu. So, it is a same structure. So, here there are 2 different you know situation, in one situation you know sigma is known to you another situation sigma is unknown to you, and both the cases when you apply z statistic, then assuming that you know n is a sufficiently large.

If n is not sufficiently large then you will be follow the t distributions. So, that is in the in the next slides. So, I will give you the example about the a small sample case, but in the mean times. So, let us take an examples to justify this particular you know z statistic; where you know sigma is a not known to you.

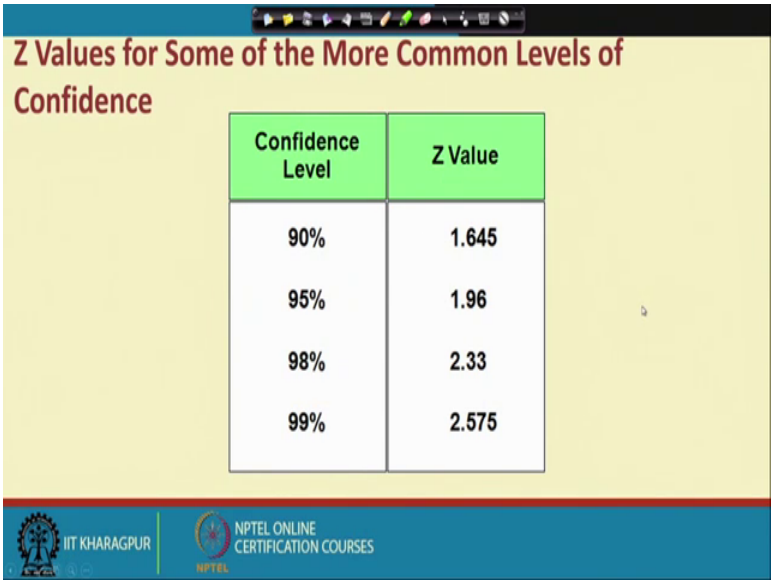
(Refer Slide Time: 23:17)



So, that means, in the sample statistic you know sample information. So, here first you have to calculate the mean, and then you have to calculate the kind of you know variance. And then s n is also given. So, again you have to fix the confidence intervals whether 995 99 or 97, then corresponding to the confidence intervals. You have to you have to first collect actually z information, again this the structure is almost all same.

So, once you have a x bar and z value and the standard deviations and the size of n, and accordingly you are in a position to find out the lower bound for the population parameter mu, and find out the you are in a position to find out the upper bound for the population parameter mu. So, as a result again having x bar 85.5, your population parameter will be ranging between's a you know 0.80 0.8 to 90, 0.90 0.2. So, it will be the lower range and it will be the higher range. So, this is what actually the kind of you know structure called as you know confidence interval.

(Refer Slide Time: 24:42)

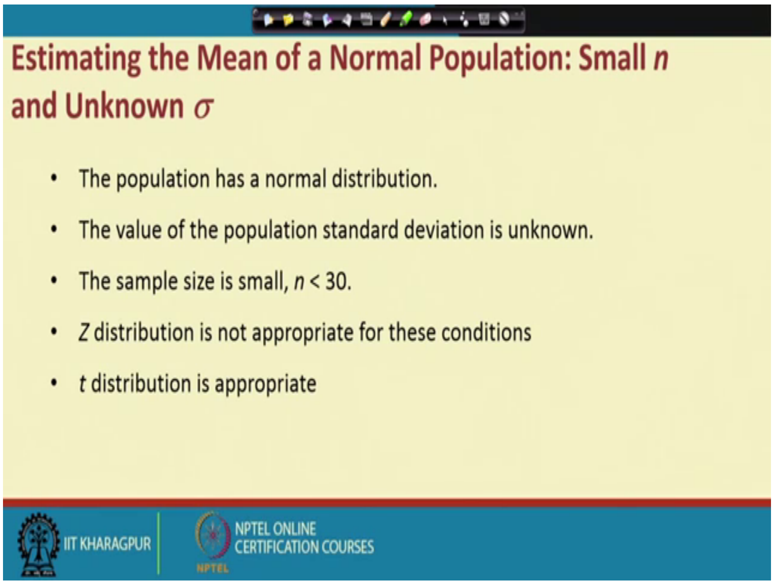

Z Values for Some of the More Common Levels of Confidence

Confidence Level	Z Value
90%	1.645
95%	1.96
98%	2.33
99%	2.575

So, again this is how the standard structure and you can take confidence interval 90 percent you can take 95 percent you can take 97 percent 98 percent 99 percent, then corresponding to your confidence intervals you have to first find out the z value. So, once you find out the z value, then with the help of you know sample statistics x bar and standard deviations and the size of the samples, that is n then you are in a position to calculate the confident confidence interval for the population parameter.

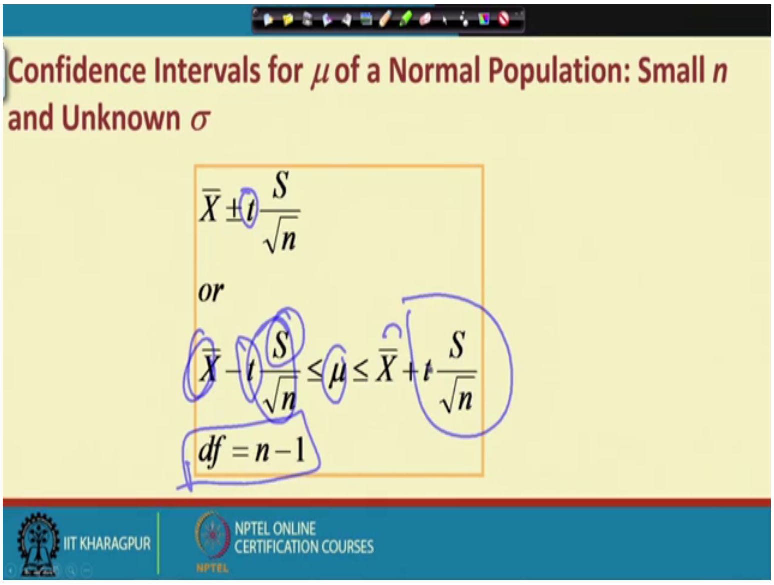
And when you have no information about the standard deviations sigma; that is population parameter then in that case sample can be used as a proxy for representing the you know standard deviations, then accordingly you will calculate the a confidence interval for the population parameters.

(Refer Slide Time: 25:38)



And then in the third case, when you are when your sample size is actually small, and a standard deviation is unknown, then in that case we strictly you know use t distributions. And in that case z distribution is not appropriate; because it is a small sample case. So, but the procedure is almost all same it is like this.

(Refer Slide Time: 26:00)



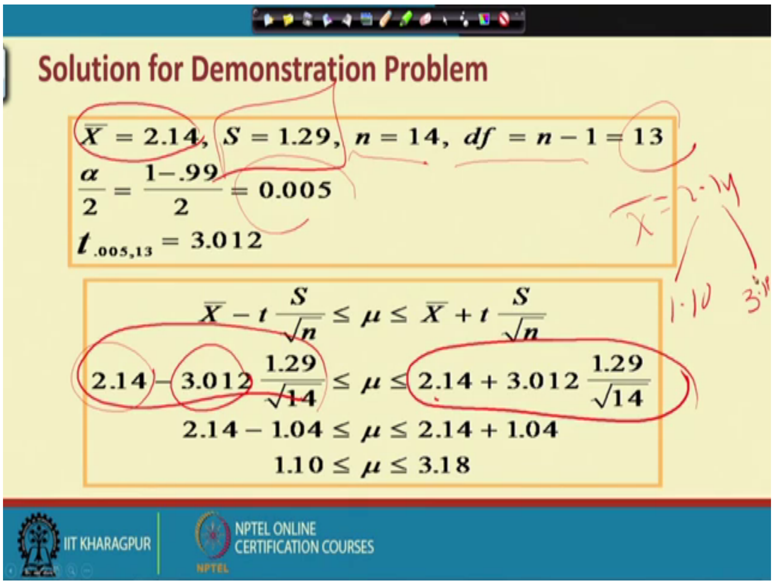
So, in the earlier case. So, we are putting actually a z here. So now, in this case we are putting actually t.

So, the rest of the things will be more or less same. So, this is what the x bar, and against minus of this much will give you the lower bound, and you know plus of this much of will give you the upper bound. So, only replacement is a t in place of you know z. And in that when we will applying actually t distributions. So, one of the standard requirement is actually degree of freedom.

So, that means, to get the t value you know t criticals. So, you need to know the alpha value that is you know either you know 90 percent 95 percent or 97 percent, and then again you need to know the degree of freedom; which is nothing but actually difference between n minus k, but in this case, you know k equal to 1 because we are going for you know one sample case.

So, knowing the degree of freedom and alpha, you can get the t critical. So, the t critical is the one of the input to this particular confidence interval, then from the samples you will get x bar, and you will get you know s, and then n will be by default number of counts, and as a result you can in a position to find out you know lower bound for the population parameter mu, and again population upper bound for the population parameter mu. So, you know let us take an examples, and then we will highlight here. So, this is what the standard examples.

(Refer Slide Time: 27:33)

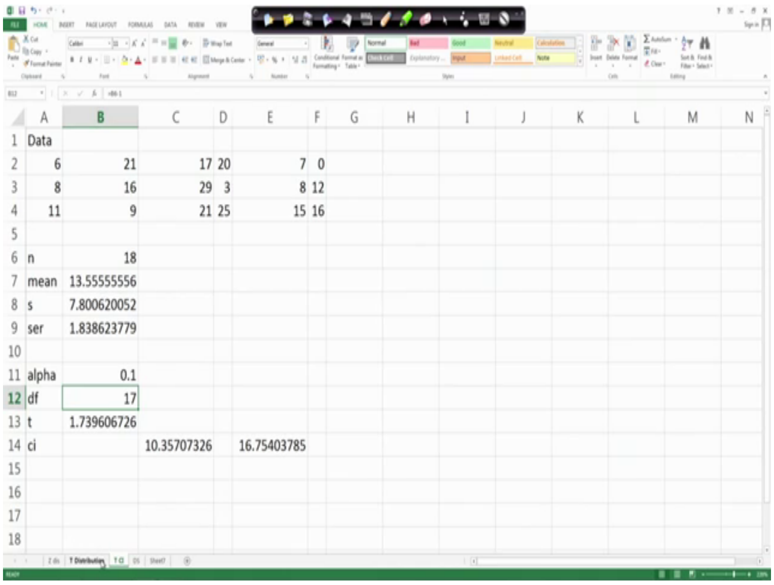


So now, from the sample statistic sample informations, you are having actually x bar equal to this much. And you know variance equal to this much. And n is 14's and degree of freedom by default will be n minus 1's because it is a one sample case.

So, this will be 13 and alpha by 2 both the sides you are targeting, then by default it will be coming 0.005. So, again same structures. So, fixing actually x bar 2.14 and the a then the t value. So, for that actually you have to go to the t tables. So, the t tables and you know corresponding to degree of degree of freedom 13 and you know, 0.005. So, the t critical will be 3.012. And then so, 1.29 that is standard deviation is givens. So now, corresponding to the n, your lower limit persons will be this much, and against corresponding to this mean 2.14 and the t 3.012 and the standard deviations 1.29 and n 14. So, this will be coming actually upper bound.

So, as a result so, having x bar equal to 2.14. So, your lower limit value will be 1.01, and upper limit value will be 3.18. So, this is how you can create a confidence interval for actually a mu subject to availability of you know sample statistic, and the kind of you know natural stress statistics. So, this is another kind of you know examples. And in that to. So, what I will do? I will take you to the example. And let us take an example in the excel sheet, and I will show you how it is exactly the particular case.

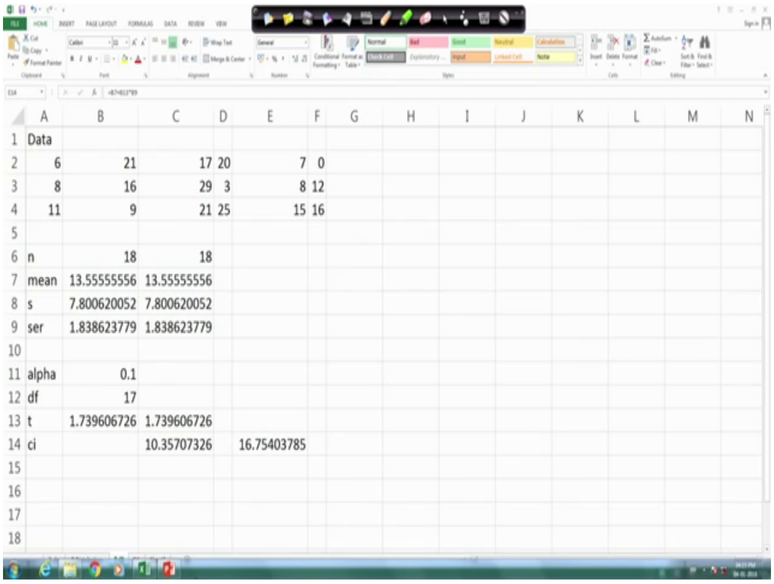
(Refer Slide Time: 29:28)



And let us see here is and this is actually t distribution. And so, this is this is the standard examples.

So, these are all actually sample points. And I have taken actually a 18 sample points, and then so, here actually these are the parame these are the items which you are supposed to know, and our actual objective is to create a confidence interval, and that to the sample mean which will be derived from this particular you know samples. So, first requirement of this problem is to check how much you know sample size; in this case which you have already reported n equal to 8. So, what will you do? Just put actually equal to sign.

(Refer Slide Time: 30:12)



and then you put the command about you know count. And count, so, obviously so, you will just give the command about the count. So, automatically you just select the range, and the range will be giving you the idea that you know it is coming 18.

So, that means, there are totally you know eight 18 samples, then we need to calculate mean of this particular series, again put equal to sign. And then you go to the menu box. And put the command about the average. Again, you what you will do? You have to select the range; that is, starting from a to 2 f 4, and then put. And you will find the mean of this particular series is coming 13.6. And that is one of the requirement for this confidence interval, and against we need actually is, that is the kind of you know standard deviation.

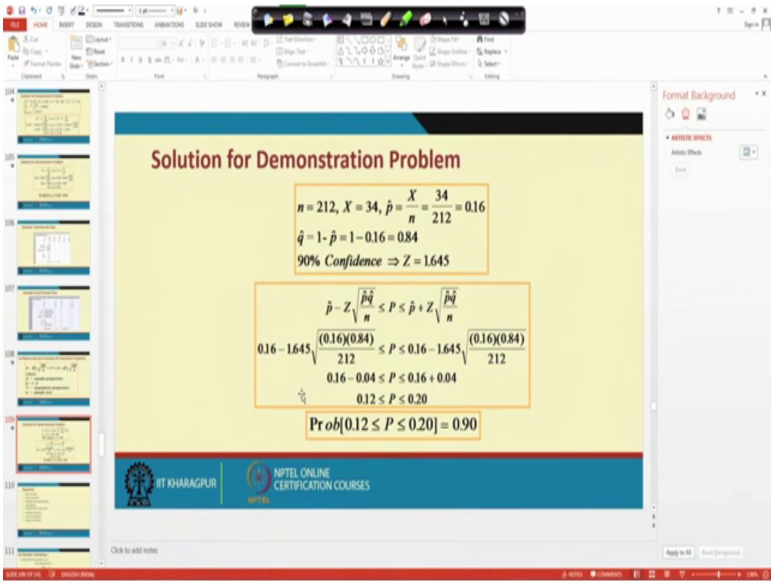
So, which is nothing but actually you know you just put you know equal to sign, again then what we will do? You put actually standard deviation, and standard deviation for samples. So, so click here, and again's you select the range. That is again A 2 to A 4, and then you just you know close the loop, and you will get actually standard deviation equal to 7.80, and then you will get standard error.

Standard error is nothing but actually equal to so, the standard deviations divided by the samples that is the count. So, you know it is actually square root of this particular samples. So, square root of this samples. So, square root of this sample. So, you put this. So, this is actually standard error, and then alpha we have a fixed directories you know 10 percent. So, it is coming 0.1. And as a result, degree of freedom will be 17 because the sample size is 18. So, 18 minus 1 the degree of freedom is coming 17. Now you have to go to the actually first you know calculate t value. And there is no point to go to the actuality t tables. So, here actually excel spreadsheet by default will give you the t value. So, what will you do? So, you just you know put you know equal to sign, and then ask for actually t information.

So, this is actually t information. So, it will ask for you know alpha value and the degree of freedom. So, that is actually this ones and then and this ones. So, then you close the loop and put the enter. So, this will create you know t values. So now, go back to our you know structure. So, we need actually the lower limit will be x bar, and minus t of 1 the adjustment factors, and then upper bound will be mean plus the adjustment factors. So, as a result the confidence interval will be like this. So, here what we will do I have already calculated. So, that means, actually so the this is actually coming mean the mean value t value.

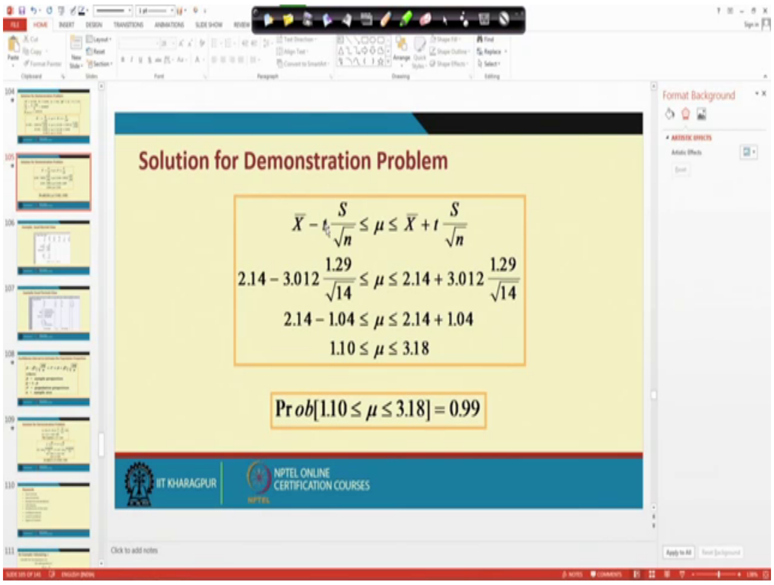
So, mean minus this adjustment factors will give you the lower limit. And mean plus this adjustment factor will give you the upper limit. So, as a result so, your mean is coming 13's this one is the 13s and then the lower bound will be 10.36, and the upper bound will be 16.75. So, as a result if you go to this actually is go to these actually.

(Refer Slide Time: 34:10)



So, see here is so, what we have actually in the t case?

(Refer Slide Time: 34:14)

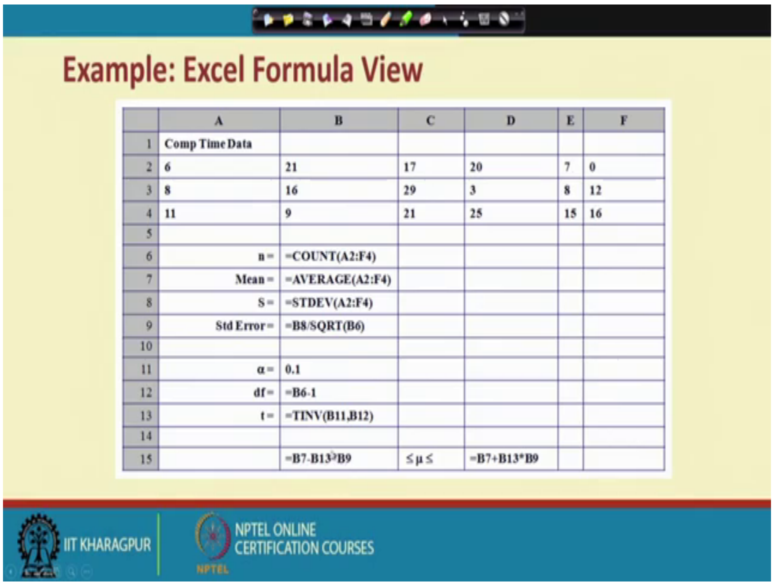


So, t value which you have already reported, and then x bar we have calculated. So, once you have connect with it t with you know variance and at the adjustment factor that is actually standard error. So, this person will give you the lower limit and again against this for x bar plus this much of adjustment will be factor will give you the upper limit.

So, this is what actually excel spreadsheet is actually helping you to get this you know lower limit and to get this you know upper limit. So, this is nothing but actually mean minus. So, the t upon the standard errors, and this is actually t value plus this adjustment factor of you know standard error. So, as a result so, this clears this clearly gives the signal that you know having the sample information and sample statistics and the choice of test statistic will help you lot to get a confidence interval for population parameters. And this is one of the typical requirement for any kind of you know managerial problem so far as you know management decision is concerned.

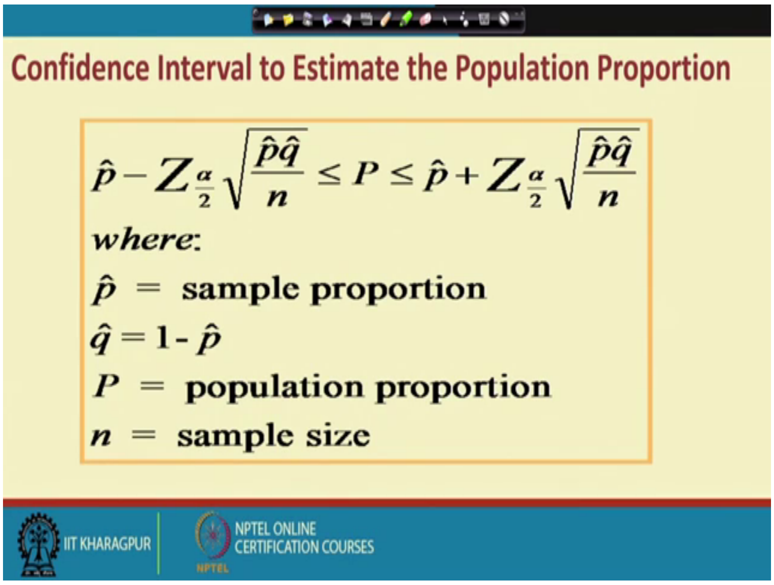
So now you know with the help of with the help of this particular you know indications. So, you may you may predict the particular you know environment, and then you can give you a fair decisions as per the management is concerned, right. So, in likewise we have actually couple of other examples which you have already highlighted, and this is what we have already discussed after calculating the t value. So, you have to just you know mean minus upon t adjustment vectors, and mean plus t adjustment vector, then you will find the lower limit value and the upper limit value.

(Refer Slide Time: 35:46)



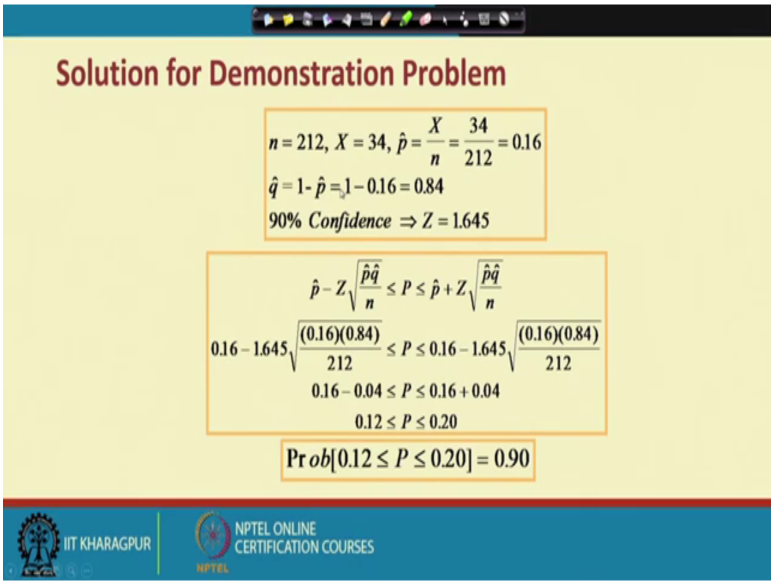
So, again similar similarly you can also go for you know population proportions.

(Refer Slide Time: 36:02).



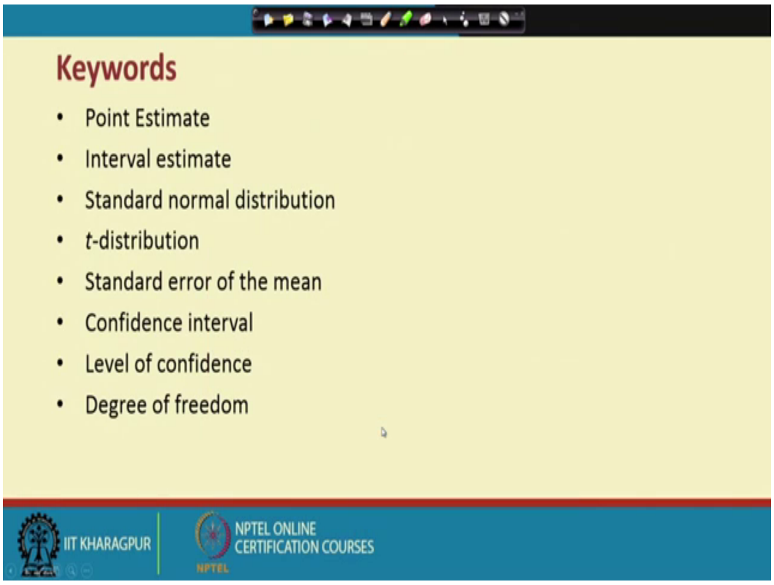
So, this is also same things. So, here actually this is actually population parameters confidence inter you know figures, and this is the lower limit for you know this is the mean value, and sample proportions, and then and this part is the lower limit and this part is the upper limit.

(Refer Slide Time: 36:21)



So, likewise actually if you take you know assume the p value is and the estimated p value, then your confidence intervals for 90 percent of confidence intervals, then the a kind of you know lower limit will be 0.12, and upper limit will be 0.20. So, here p is given actually the kind of you know 0.16. So, having 0.16 sample proportions the population proportions will be in between 0.1 to 0.20. So, whatever may be the shape of the problems. So, you may be in a position to calculate the confidence interval for the population parameter.

(Refer Slide Time: 36:51)



And these are the things we have already discussed, and with this we will stop here today.

And thank you very much.