Life Cycle Assessment

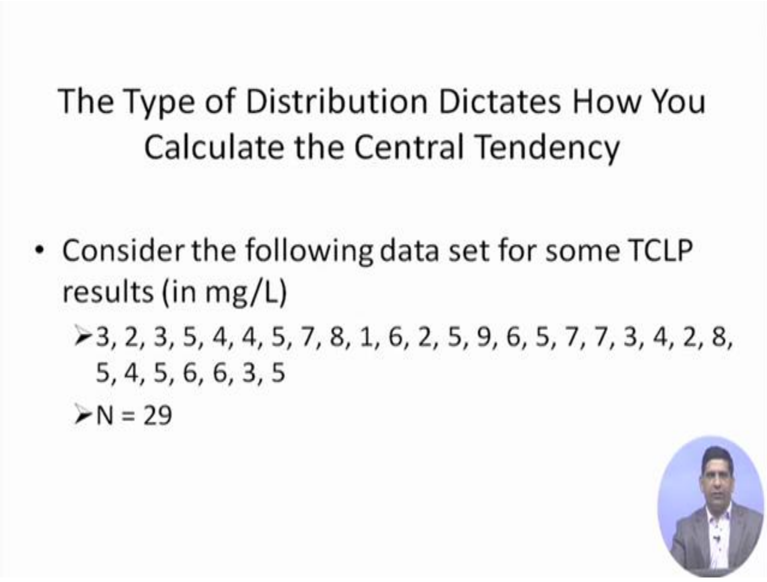
Prof. Brajesh Kumar Dubey
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture – 13
Environmental Data Collection and LCA Methodology (Contd.)

Welcome back, and we have been discussing about the data and we will continue this discussion in this module as well. So, I gave you example of the two data sets towards the end of the last module and will continue to work with those two data sets in this module and I will give you all kind of walk through an example of how the statistics plays important role. We also talked about the TCLP stuff which will kind of go back here and talk a bit about that as well.

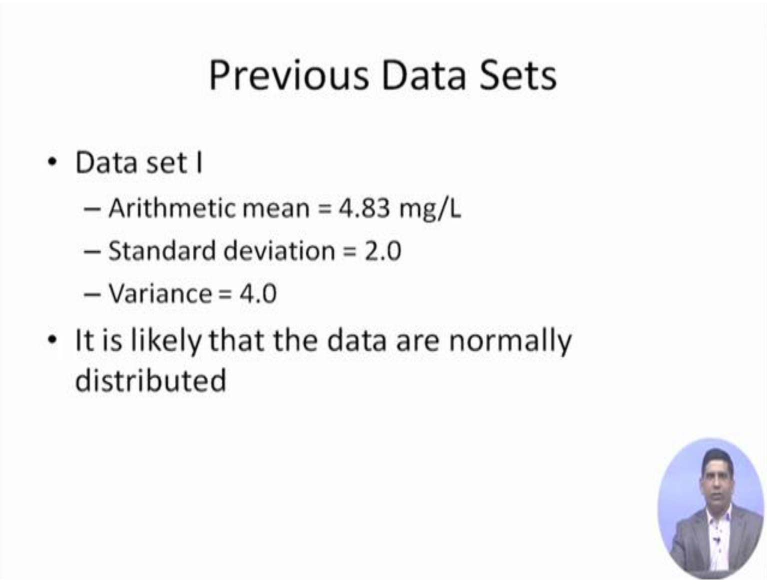
So, in this module though we will start with this statistical analysis for environmental data and then with the time permits will try to go little bit on how collecting all these data, how we start looking at LCA exercise. So, take up will it start getting into the life cycle analysis, if not in this then definitely in the next module will do that. So, in terms of the statistical analysis if you remember from the previous module we will look, I showed you this data set where we had 29 samples variability from 1 to 9 and if you have not reproduce this slide again over here, but if you remember we had a nice bell shaped curve which was the normal distribution.

(Refer Slide Time: 01:14)



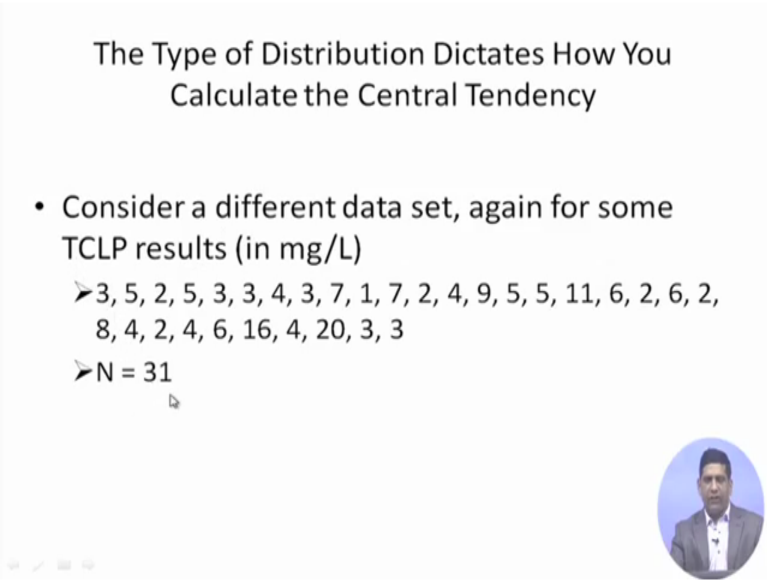
So, for the normal distribution this is we call this, will call this particular data as a data set one. So, this is our data set one and for this if you find out the arithmetic mean that comes at around 5.83 milligrams per litre and standard deviation is 2, variance is 4 and for this case we think that it is likely that the data is normally distributed.

(Refer Slide Time: 01:33)



Especially when you have the arithmetic mean is greater that is standard deviation as well as variance the chances are that the data is normally distributed then we plotted this on a histogram which you saw in the previous towards the end of the previous module and you saw that it has a typical bell shaped normal distribution curve.

(Refer Slide Time: 02:21)



For the second data set let us look at the second data set which we had the earth in the previous slide towards the end of the previous module as well, this is again for some TCLP results where we have 31 samples greater variability, the data varies from 1 to 16 sorry 20, 1 to 20. So, lot of variability and this was a log normal this kind of showed that possibly it is a log normal distribution we do not know for that for sure yet, but it most it is not a normal distribution that we saw anyway be saw a long tail in this particular histogram that if you remember from the last module, if you do not remember you can look at the pdf, slides as well a few slides back it is there.

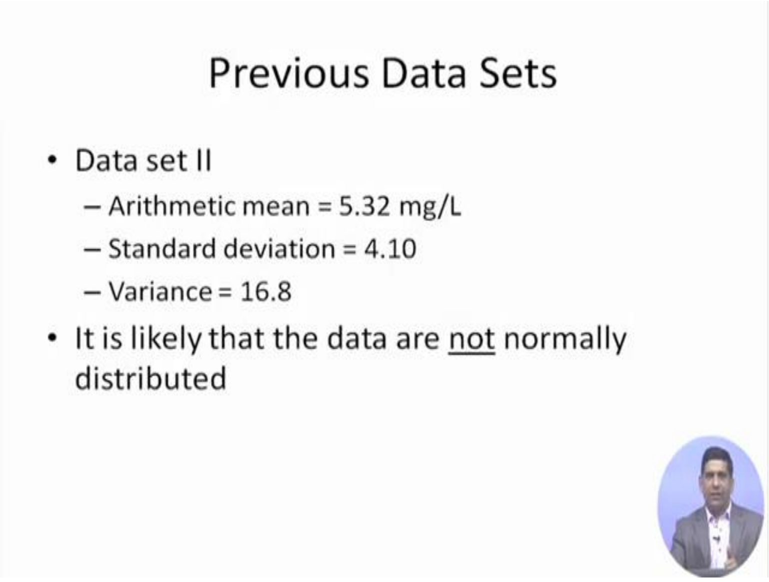
So, for 31 data sets this was the distribution and this is again both of them are coming from TCLP results and let us assume that these are the TCLP results for led coming from electronic waste, why led I will just explain that you. Let the TCLP is limit is 5 milligrams per litre. Now again what is TCLP what we talked about TCLP in the previous module. It is a toxicity characteristic leaching procedure is a test which is done to find out whether a waste is a hazardous waste or not whether a particular solid waste can be put in a MSW landfill or not. So, TCLP test is done simulating acidic conditions in MSW landfill where we are trying to simulate acetic acid and sodium hydroxide.

So, we have the acetic acid which is, why acetic acid? Because if you remember what is there in the municipal solid waste, municipal solid waste has lot of organic matter lot of food waste and the food waste when it degrades it forms acetic acid those high molecular organic acid gets broken down to smaller molecular organic acids and finally, they are to acetic acid and from the acetic acid they will get further to methane gas. So, that is why the acetic acid is used and sodium hydroxide to make it a more buffered solution pH of 4.93 and that is the pH we use for the TCLP test.

So, if you have electronic waste you run TCLP and you find try to find out if the if it fails for led if it comes out to be more than 5 milligrams per litre for led; that means, we cannot put this waste in a municipal solid waste landfill because if we put it in the municipal solid waste landfill in the case if there is a liner failure the leach it coming down and hitting the ground water, will exceed the led concentration that is the required concentration, that is the led limit for groundwater which is same as drinking water.

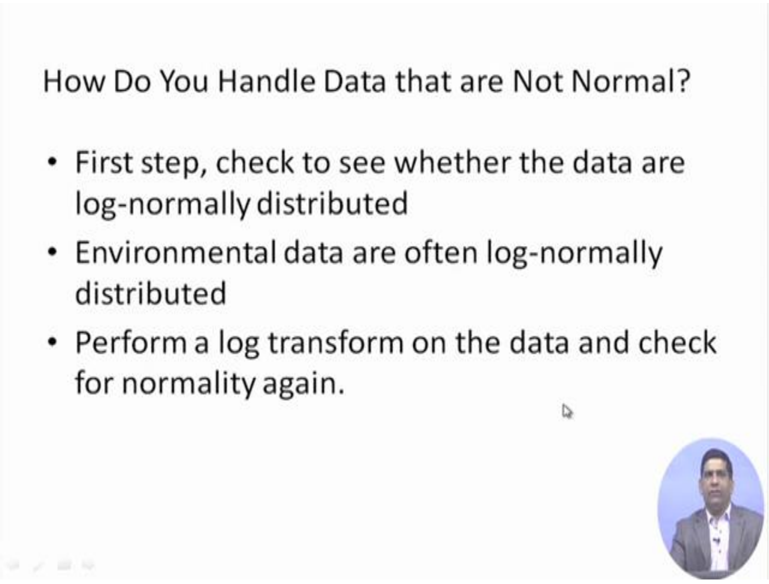
So, that is the rationale behind doing that is TCLP and all that, and we do not to just kind of give overview since we have been looking at this data. So, again two data points two data sets this is the second data.

(Refer Slide Time: 04:58)



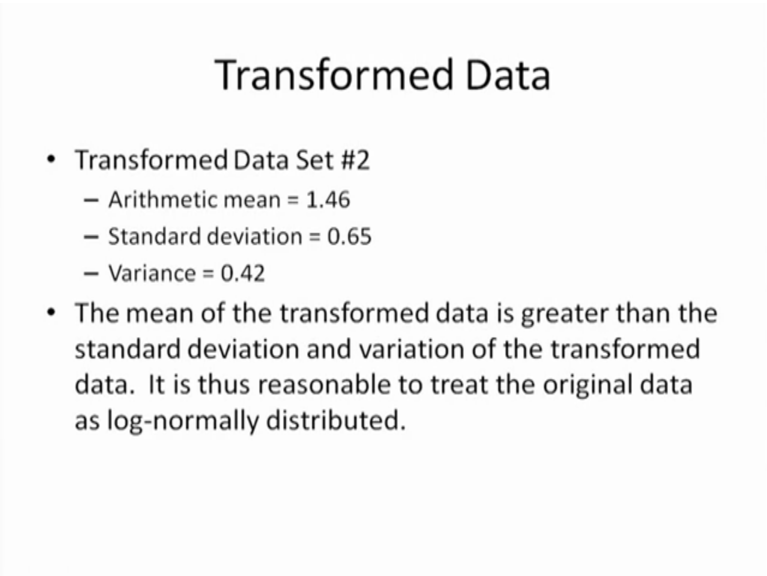
Now, if you look at the second data set arithmetic mean is 5.32, standard deviation is 4.1, variance is 16.8. Since the arithmetic mean is although it is greater than the standard deviation, but it is less than variance. So, it is likely to the data is not normally distributed and we saw that in the histogram as well which we it has a long tail.

(Refer Slide Time: 05:21)



So, if it is not normally distributed what we can do with that, if the data is not normal then what we do? The first thing we do is we check whether the data is log normal distributed. So, what is the meaning of log normal distribution we will take the log value of that you like transform; you took a log transform of the data and check for normal see again. So, the data points if you do the log transformation they should show up this, the log transform data should show a normal distribution. So, environmental data most of the time log normal distributed as I have mentioned earlier as well.

(Refer Slide Time: 05:59)

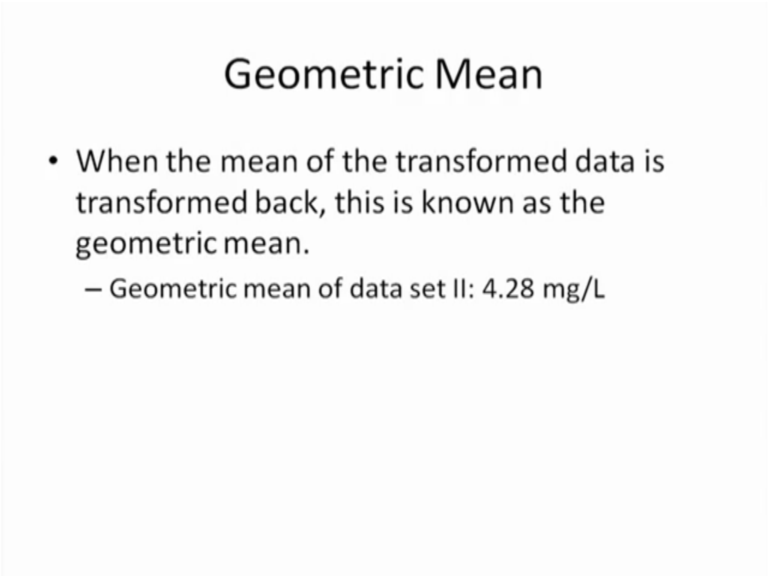


So, if you do the log transform of the data of the second data set which was not normal distributed we find arithmetic mean to be 1.46 standard deviation is 0.65 in the variance is 0.42.

So, these values came after we remember that particular. So, if you go back to this data set we took this data set we took the log value of this then for this 31 data set we have the for the individual data we have the log value and then we took the log, log value and then we have taken the arithmetic mean and the standard deviation of the log transform values from there. So, these are not the actual data it is the log transform data. So, the mean of the transform data is greater than the standard deviation in the variance of the transform data, so that is reasonable to treat the original data as log normal distribution because the transform data source to be normally distributed so; that means, the original data is log normal distribution which we kind of saw in that histogram as well typically that long tails it basically signifies that it is a log normal distribution.

So, when we have the mean of the transform data back this if you have looked at some of this literature you may have seen arithmetic mean and geometric mean.

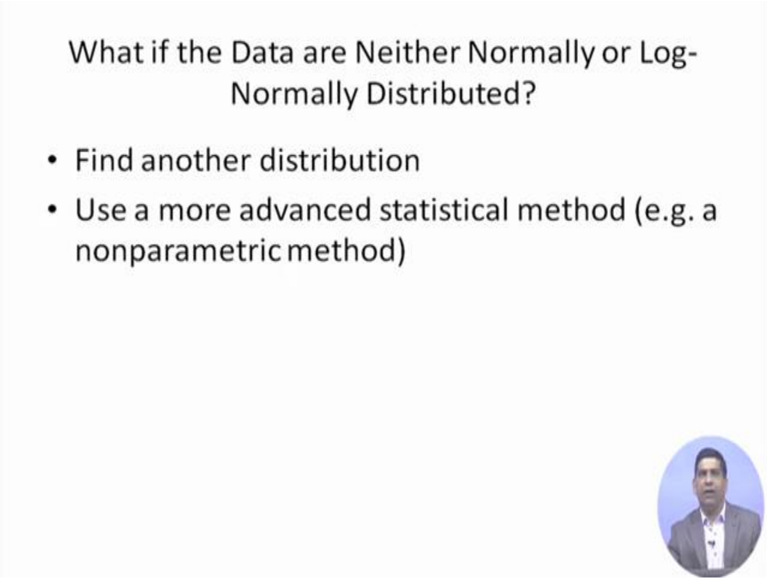
(Refer Slide Time: 07:03)



So, this geometric mean means when you have a log transform data when you have a log normal distribution the mean of the transform data is transform back. So, here the mean was if you look at the previous slide the arithmetic mean of the transform data is 1.46. So, the geometric mean of the original data is 4.28. Now what is this 4.28? 4.28 is essentially e to the power of 1.46 because you have taken in the log. So, now, we are taking anti log. So, this is e to the power of 1.46 will give you 4.28 milligrams per litre.

If you a calculator in hand you go and check it should match. So, e to the power of 1.65 will give you 4.28 milligrams per litre. So, that is what the geometric mean that you will see that this is called geometric mean just to distinguish between arithmetic mean and this is because the arithmetic mean usually we use it for the normal distributed, normal distribution, for the log normal distribution we call it a geometric mean. So, this is the geometric mean is 4.28.

(Refer Slide Time: 08:17)

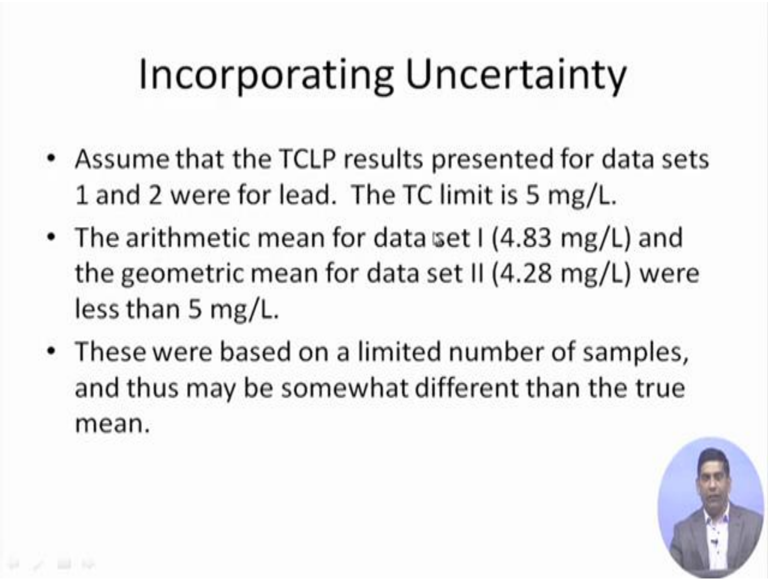


Now, if the data is neither normal nor log-normal distribution; then what you will do? So, in that case you can find another distribution, you can use more advanced statistical methods that is non parametric method. In environmental field this environmental statistics is itself a very big area of work research as well as teaching some places there is a core start at the PhD level or at the M. Tech level as a elective for the M. Tech level on environmental statistics where people go in more detail about in terms of different distributions there is a if you Google, I call it a professor Google actually. If you Google and you will go and you can find out e p a document on around I think around 550 to 600 pages document, just focused on how to handle different types of environmental data. So, statistical methods have been proposed for advanced statistical methods to deal with different kinds of data. So, that is very much important in terms of the data analysis.

So, for this class, for this particular course let us just do not worry too much about the other statistical methods will not go there, we will assume like for most of the data since most of the data is either normal or log normal. So, we have kind of covered well for some of the very specific cases where are the neither or normal log normal that is we will leave it for our some PhD students and other stuff to deal with that in the research, but for this course we will not worry too much about those aspect in there.

So, if you have just for your information that the slide was there just to kind of give you some information that in case this is not the normal distribution be there is a way to do it. So, just for your, to take care of your curiosity if you have the curiosity which you should have.

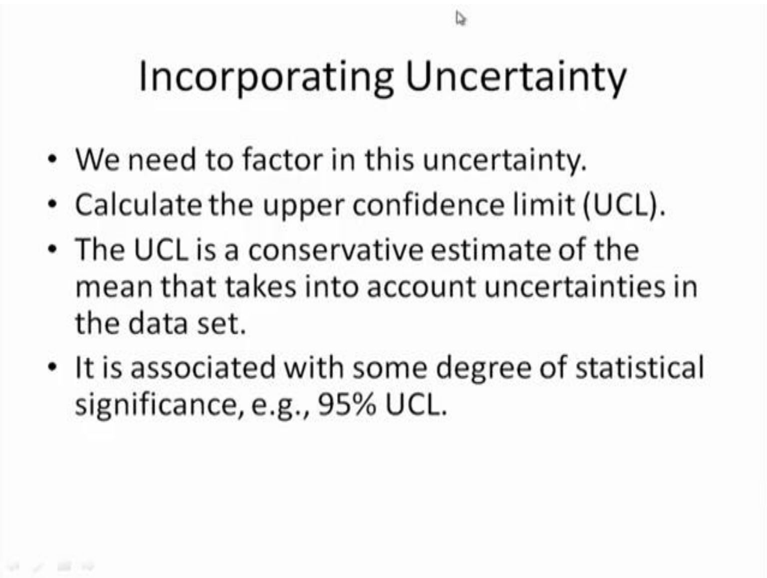
(Refer Slide Time: 10:01)



So, now if you remember for both the samples as I said earlier if you said that both the data set if it is for lead where the TCLP limit is 5 milligrams per litre, now the arithmetic mean for data set one is 4.83 geometric mean for data set two is 4.28 and both are less than 5 milligrams per litre and these are based on 29 samples in 31 sample. So, they are a good number of samples, but are still they are limited number of samples too because depends on if you remember we have got said earlier is more the variability more the sample more the samples we need. So, since we have used 29 and 31. So, that is a good number of sample, but still it is may not be we have to kind of look at going from this now sample mean to the population mean.

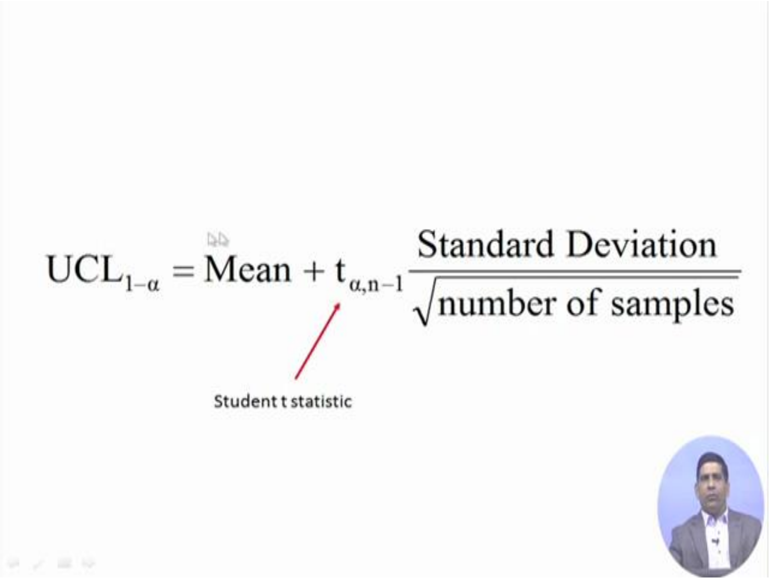
So, if you remember from where statistics class, at the very beginning of the statistics class it is whatever course when which every year you took the statistics you had this concept of sample mean to population mean. So, we will try to go there and we will see whether this sample with the 29 or 31 that we have collected for based on which we are getting the numbers is 4.83 milligrams per litre for data set one and 4.28 milligrams per litre for data set two, in this case both the data set both the scenarios is passing TCLP for led. But what about if we incorporate in uncertainty and incorporate our like if we do the data like a nice statistical analysis on that, for this try to calculate the population mean.

(Refer Slide Time: 11:35)



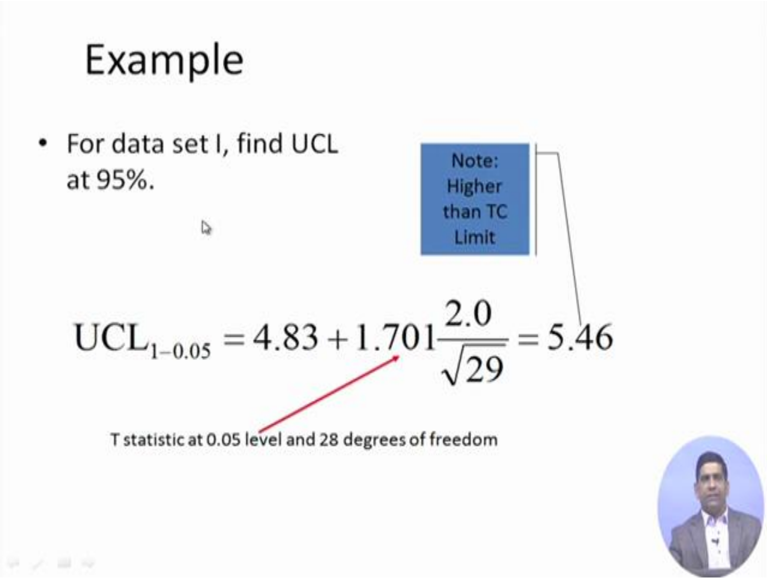
So, we need to factor the when we go from sample mean to population mean we incorporate what is known as uncertainty and we calculate upper confidence limit and the lower confidence limit. Now what are these upper confidence lower conferences, if it might be reminding you your statistics class whenever you took that I am pretty sure you must have covered this concept. So, this UCL is a conservative estimate, UCL is a conservative estimate of the mean that takes into account uncertainties in the data set. So, since the data set may have certain uncertainties like how you sample it how the sample was analysed and what are the how good the sample in terms of population like a representing the real scenario. So, far that it is associated with some degree of a statistical significance. So, we look at what is the statistical significance.

(Refer Slide Time: 12:35)



Usually we use 95 percent UCL or 99 percent UCL those things are used in terms of calculating this like upper confidence limit. So, if you look at here this is example of this is a like a equation we typically used for UCL. So, UCL 1 minus alpha which is your like a say 95 percent. So, alpha will be 0.05 and it is defined as mean plus the t value from the student t test remember the student t test we have in every statistics books towards the end of the book we can find the t table, where you have alpha is a degree alpha is your significance n minus 1 is the degree of freedom where the n is number of sample and then is standard deviation divided by the square root of the number of sample.

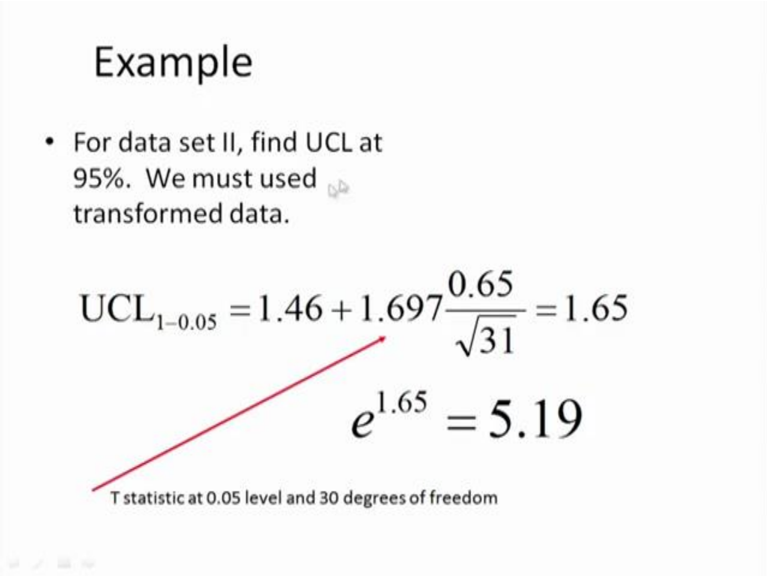
(Refer Slide Time: 13:15)



So, now if you have to calculate for the data set one if you are calculating what is the 95 percent that UCL 95 percent, UCL 95 percent means 1 minus this 0.05, which is the like that is 1 minus that will give us 95 percent because 1 minus 0.05 is 0.95. So, for 95 percent we if you remember the equation mean, mean was 4.83 and t statistic came out to be at 0.05 level which is at 28 degrees of freedom, it came out to be 1.701 this standard deviation was 2, number of samples was 29. So, the square root of 29 and then if you do this upper this calculation at 95 percent limit for the upper confidence you get a value of 5.46. Now this 5.46 is greater than 5 milligrams per litre.

So, all though your arithmetic mean, so that is like a 4.83 which was meeting TCLP standard, but when you went to incorporating uncertainty, when you went to the population mean at 95 percent confidence interval for upper confidence in limit you see that the value actually comes out to be higher than 5. So, now, let us look at the second one. So, if we and then we will discuss what does that mean in terms of values like a failing TCLP now when we go to the population mean how to handle that.

(Refer Slide Time: 14:49)



So, let us look at the second set. So, in the second data set as you can see on the slide over here for the same 95 percent, like a for 95 percent confidence limit we have trying to calculate it here your mean was here remember this these equations for t test, g test, upper confidence limit, lower confidence limit they assumes that the data set is normally distributed. So, that is why when you see the number on this particular slide over here we are using this value that we are using in this particular equation is 1.46. So, which was if you remember that geometric mean was 4.28, but that 4.28 came because it was e to the power of e to the power of 1.46.

So, here we are using the values of the transform data. So, the standard deviation value is we are using it for the transform data, the mean value we are using it for transform data because that transform data is normally distributed and as I said for t test, g test, upper confidence limit all this calculation is based on normal distribution. So, that is why we are using the log transform data. So, you need to be very very clear about that that is that there should not be any confusion on that said why we are using log transform data because the equation the basics of the equation assumes that we are working with a normal distribution.

So, the log transform data that we had from this second data set was normally distributed that we looked at earlier. So, that is why so come back to this equation now. So, here if you look at 1.46 that is the log transform mean we get the t statistic value at 0.05 level in thirty degrees of freedom because 1 minus 0.05. So, that is a 95 percent. So, this is alpha is point 0.05 and we get the value of 1.697 you can get it from any t test look at up any statistics books and you will find this number is standard deviation n was 31.

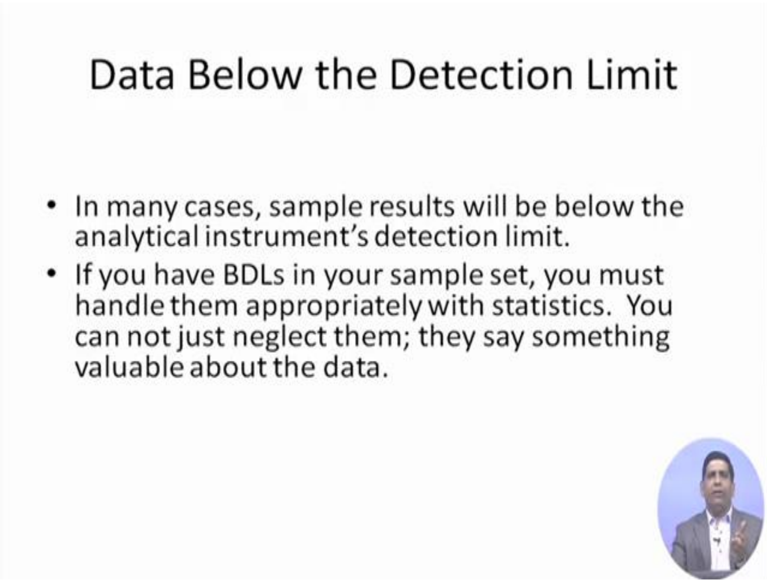
So, square root of n again the same equation and then you get a value of 1.65. So, this 1.65 value of UCL 95 percent is of the log transform data and to get it the actual value to the geometric mean for the population will take the anti log. So, antilog will be e to the power of 1.65. So, that is why you see over e to the power 1.65 and that we have say value of 5.19. So, in both the cases for the data set one as well as for the data set two we are getting a value of the population mean at 95 percent upper confidence limit to be higher than 5 milligrams per litre.

So, we are in a dilemma now, when we look at the sample mean is sample mean was passing TCLP, but when you go to the population mean they are failing TCRP because the TC limit was 5 milligrams per litre. So, what you will do? What you will do in this scenario? Say if you are a regulator or if you have to make the decision what you will do for them this scenario, what does that mean? That again these data sets that we have showed you they are not I want to tell you that these are not actual values, these are the values that I have cooked up because just to illustrate some point here just that these are the scenarios that you will deal with when you are looking at the environmental sample where things are not things will be very tricky and this situation do so up.

So, here what option we have? One option we have is if you are a regulator if you say work for the state pollution control board, central pollution control board you just tell that industry the industry with whoever is producing this waste this is your hazardous waste because when I do the statistics it came out to be a hazardous waste. So, you have to manage a hazardous waste. The problem with that is the hazardous waste is very very expensive to manage sometimes 100 times more expensive to manage than a non hazardous waste because of the strict regulation and so in that case what as a industry I may come up with an argument say if I am running the industry I will come up with an argument that let me go back and do more sample because that is the problem is not it.

So, if because there is so much difference between the population mean and the sample mean; that means, the sample is not really able to like predict the population mean that nicely because of the heterogeneity of the sample. So, to deal with the heterogeneity of the sample if you go back and do more and more analysis we may probably be closer to the let the population mean and in that case even if it comes out to be able than 5 then of course, we have to deal with that as a hazardous waste. So, that is kind of some idea about how things are managed in terms of like a hazardous waste or not in terms of environmental sample. So, next let us look at what are the other issues in terms of the environmental sample like if you have a detection limit to if you remember is the previous module will talk about detection limit.

(Refer Slide Time: 19:42)



So, detection limits, what is detection limit? It is the lowest concentration that an instrument will give you with a 95 percent confidence that yes the number is correct number is above 0 and it is correct say if you are running a particular instrument, any instrument if you run when you run the sample it will give you a number, but whether the number makes sense or not it is you as a researcher you as a M. Tech student or a PhD student or a B. Tech student whoever is you, whoever you are watching this video you it is your responsibility to make sure that the data is correct do not come and tell your supervisor or your professor that this is what I got from the computer.

See the computer is not getting a masters degree computer is not getting a PhD degree or a B.Tech degree, it is you who are getting the degree. So, as a researcher, as a researcher in training you need to make sure that the number that you are reporting in your thesis, in your report, in your dissertation, is it correct number because they many times when you run some samples you may even see negative concentration, negative concentration does not mean anything is not it.

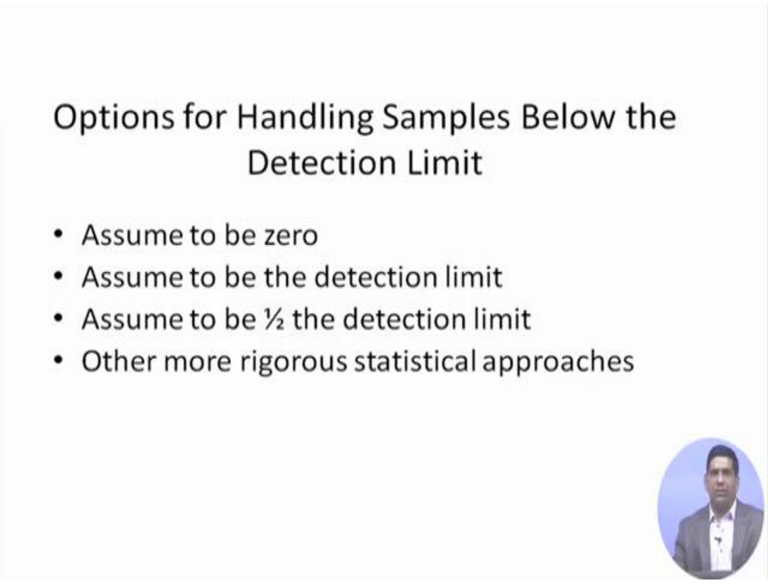
So, you have to make sure you know what is the detection limit of the instrument that you are using and then if you go very closer to the detection limit you need to make sure that the data is correct and then if you have something which is below the detection limit we already talked about how we calculate detection limit and which in the last module, but when you have the data which is below the detection limit and they do come out in many cases the sample results will be below the analytical instruments detection limit.

So, you get a value between 0 and the detection limit say in the previous if you remember from the previous module the detection limit for that analyte was 0.65 milligrams per litre. Now if you have a value between 0 and 0.65. So, can we trust that number I do not think? So, because it is deduction limit is 0.65 anything below 0.65 we cannot trust that number. So, what does that mean? It means it is below detected; it does not mean the value is 0. So, if you have BDL in your sample set you must handle them appropriately with statistics. So, if you just put a value that it is a BDL BDL BDL many times you here if you look at the environmental report, environment analysis report people just reporte as BDL, but BDL does not mean anything BDL unless you provide the detection limit information means nothing. So, you said it below detection limit. So, what is the detection limit you have to tell me, what is the detection limit then only I will say it is below this number?

So, for example, in the example that we did for me for the metal detection limit we found 0.65 milligrams per litre. So, if the correct way of writing those are less than 0.65 milligrams per litre, we do not know the value it is less than 0.65 milligrams per litre and it could be anything between 0 and 0.65 milligrams per litre, but we cannot say for sure. But if you use this kind of like less than 0.65 milligrams per litre, you cannot use that information in your statistics you do not, you do not because that is if you put a like a text value in your excel table that text value will not be taken it as a number. So, or if you have a less than BDL or BDL and you provide BDL information separately then again you are not using these data as an statistics, in your statistical analysis.

So, BDL values also mean something is not it say it means that there are certain number of samples which is actually below say in this case 0.65 milligrams per litre. But how to incorporate these samples this sample results in statistical analysis because for statistical analysis you need to have numbers, numbers are important there. So, you need to use numbers to do that in that case what you do.

(Refer Slide Time: 23:39)



There are different options which are off sense for handling the samples below the detection limit; some studies suggest that you assume it to be 0. Now what is the problem of assuming it to be 0? The problem of assuming it to be 0 is that you are being too what say conservative or you are because if something is below detection say in that case the value if the detection limit is 0.65 as you saw in the previous example, if something is below 0.65 milligrams per litre and we assume it all of the values say there were out of 50 data points I have I have maybe 5 or 6 data points which came out to be below detection and if I take all of these to be 0; that means, actually we are being not true to the data because not all of them will be 0 because it is we do not know the number, but they may not be 0 there might be some concentration present there, but which is since we cannot measure it with 95 percent confidence limit we are not reporting those numbers because it is still there could be some concentration.

So, other option is assume it to be same as the detection limit, which is again you are being otherwise assume it to be 0 is your being too liberal because you are making your cases, you are showing that look better in terms of the environmental performance and for assuming to be the detection limit your making becoming too conservative because you are assuming that everything is 0.65 milligrams per litre, which is not true because they will not be 0.65 milligrams per litre it will be somewhere between 0 and 0.65 milligrams per litre.

One option is you assume it to be half of the detection limit which is the most common approach that you see in many of the research articles or research reports that you will see people will assume it to be half of the detection limit and to some extent I agree with that aspect too like you can, like I will put forward the argument why I agree because at least what you are doing say again if we have working with a analyte which is the detection limit of points 6 5 milligrams per litre as we saw earlier in one example. So, if you are work, if you are taking it half which would be 0.32 or 0.33 milligrams per litre, say you had 6 or 7 data points below detection you what you are saying that yes sir some of them could be higher than half detection some of them could be lower than half, but you are taking it kind of in the middle.

So, this is your 0 number then this is your like detection limit number. So, in between you are taking a line in the middle say that just assuming in to a middle some will be about some will be lower, it will kind of I will off suit like I will take care of the impact of higher number as well as the lower numbers were taking at the middle ground. So, that can I makes sense and where you can use this and it because again why we need all this because if you just put the BDL values those cells will be ignored when you do the statistical analysis, statistics requires number they do not work with test values, they need the numeric value. So, for that if you put half of the detection limit at least it will take that information and give you the values in terms of the t test, g test, chi square test and know and whatever you do in terms of the statistical analysis.

Then there are other more regressive statistical approaches which you can also take for handling sample below the detection limit. So, that is also done, but in terms of like a in terms of like for our case scenario here again as I said earlier this statistical methods for environmental engineers or environmental statistics it itself is a course which goes in great detail about all these aspect. The goal of introducing you a set of maybe 10 20 slides over here on statistics is just to kind of get you thinking about statistical aspect that is involved, you most of the students I assume that you had a statistics class which should be true whenever I teach in a classroom environment I find and you most of the students especially at the final year B. Tech level or the M. Tech and PhD level they had taken some sort of statistics course.

So, you do know some of this stuff, but you may not know from environmental analysis perspective. So, I just wanted you to start thinking in more in the environmental line because what kind of data analysis you will need because in terms of the environmental sample and these are just a very brief overview with a brief in few examples illustrating few key points and how to handle the samples below detection, how to calculate method detection limit, how to go for 95 percent confidence level, how to go from sample mean to population mean. So, these are just like an overview of basic environmental it in like environmental analysis for like a statistical analysis for environmental sample, so environmental data. So, this will give you a good kind of background when going when we start looking at data in terms of LCA method, LCA analysis. So, life cycle assessment calculations.

So, with that we will kind of wrap up this particular module and from next module will start looking at LCA methodology how we do LCA and how kind of some of the overview of doing LCA. So, hope you have been enjoying the course. So, far and you with this we have kind of covered the more requisite background of that you need to go into understanding of LCA, so that from next module you will see some LCA examples.

Thank you and keep watching.