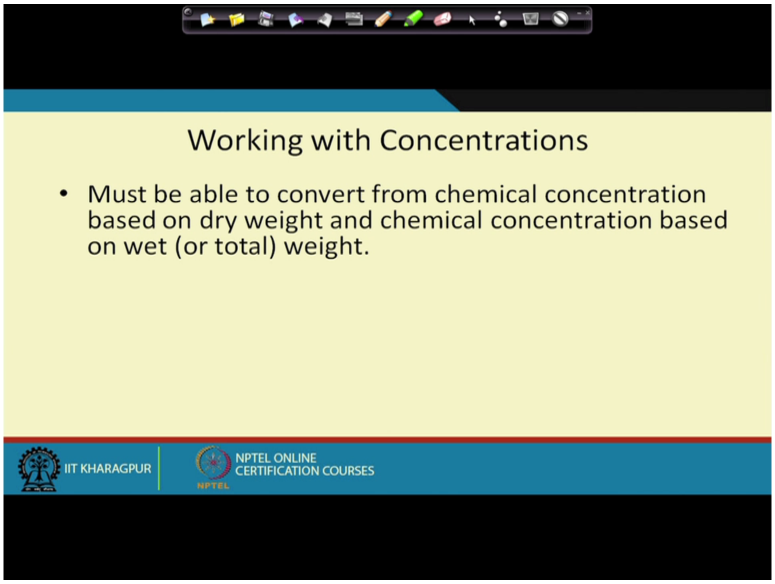
Integrated Waste Management for a Smart City
Professor Brajesh Kumar Dubey
Department of Civil Engineering
Indian Institute of Technology Kharagpur
Module-03 Lecture-14
Working with Data and Statistical Methods

Let us start our discussion from where we left in the previous video. We were talking about different analytical instruments and we have. So far if you looked at, we talked about how the data to be collected, what are the importance in terms, important aspect we need to consider in terms of the data collection and what is the important of having the good quality data. So now once we have the data, how we make sense out of those datas?

I will in this particular video, in this particular module, I will try to give you some example, some examples of how the data is used. Again, these are just illustrative purpose just to kind of get your thinking about how the data can be used in different ways. So let us get started.

(Refer Slide Time: 0:59)

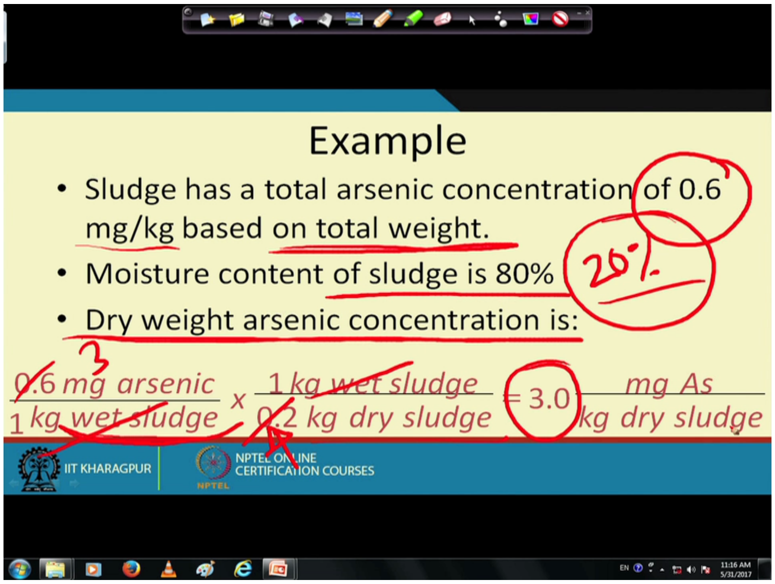


Working with Concentrations

- Must be able to convert from chemical concentration based on dry weight and chemical concentration based on wet (or total) weight.

So when you try to work with the concentration, as I said very early and I think couple of modules before, we should be able to convert the chemical concentration based on dry weight and the wet weight because that is the one thing we need to do because dry sample we like many times when we take the dry weight is based on the moisture, like moisture content has been figured in. Wet weight we are not, we have not factored the moisture content there. So it is a wet and dry. So that is, we will try to do some math here.

(Refer Slide Time: 1:29)



Example

- Sludge has a total arsenic concentration of 0.6 mg/kg based on total weight.
- Moisture content of sludge is 80%
- Dry weight arsenic concentration is:

20%

$$\frac{\overset{3}{\cancel{0.6}}\text{ mg arsenic}}{1\ \cancel{\text{kg wet sludge}}} \times \frac{1\ \cancel{\text{kg wet sludge}}}{\cancel{0.2}\text{ kg dry sludge}} = \boxed{3.0}\ \frac{\text{mg As}}{\text{kg dry sludge}}$$

So say for example, if in this particular case if you look at, we have a sludge which is, sludge has arsenic concentration of 0.6 milligram per kilogram based on total weight. So it is total weight means it is not the dry, it is the wet weight. And the moisture content of the sludge is 80 percent. So sludge is as typically we know has lot of moisture. So many times actually for material like sludge, we do not report moisture content, we report solids content. So solids content is like 100 minus- moisture content. So in this case, the solid contents will be 20 percent, as if you think up to the solid contents will be just 20 percent. So out of 100, say 100 unit of the sludge, 80 unit is just water and 20 unit is the solid present there.

So the dry weight, in terms of the dry weight arsenic concentrations, if the wet weight is 0.6, so that how much would be the dry weight? So if you, again most of the solid waste problems are not difficult problems. They are pretty straightforward. It is solid, that is one beauty thing about the solid waste course is things are, especially if you need to make sure you understand the units and you follow the units correctly.

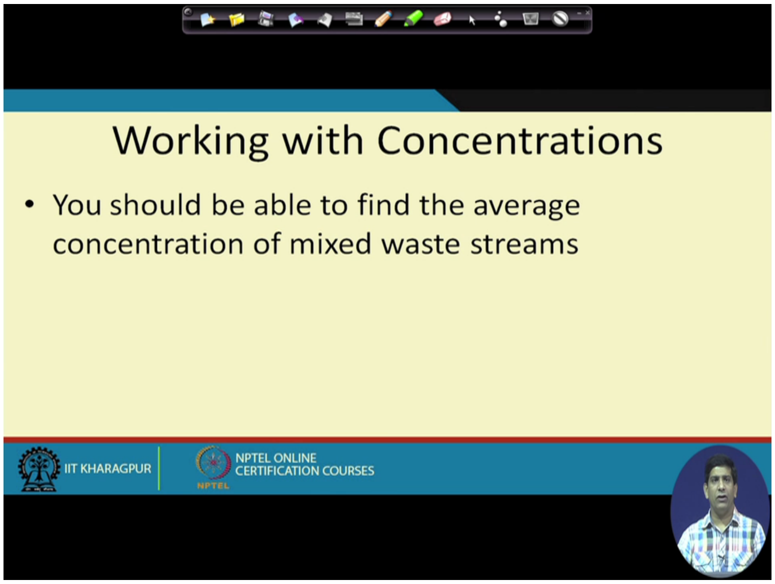
And as long as you do that and do a mass balance, most of the problems are kind of a mass balance problems. So if you do the mass balance, you will be okay. Say if you go for, so we had 0.6 milligram of arsenic per kg of the weight sludge, so that is we have a wet sludge here. That is the total, is not it? So we have a wet sludge but we have to calculate the dry weight arsenic concentration. So dry weight means in terms of the dry mass.

So if you think about the mass, we have 0.6 milligram of arsenic. That is the mass of arsenic in 1 kg of the wet sludge. Now if you want to convert that into the dry part, we need to know in 1 kg of wet how much is the dry? So for 1 kg of wet, we have 0.2 kg of dry, that I said like 20 percent will be the solid content. So in 1 kg of wet sludge, we have only 0.2 kg of dry sludge because the dry content or the solid content is only 20 percent.

So if you do this, your wet, that wet sludge and wet sludge will cross away. So whatever you left, it is 0.6 milligram of arsenic per 0.2 kg of dry sludge or if you do that, your 0.2 will cancel with 0.6 and you will get 3. So here you get 3 milligram of arsenic per kg of dry sludge. So pretty straightforward but many times I see students even making mistake in that. So you need to, so they get confused between the moisture content and solid contents and all that and they mess up.

So but this is, so this is how you kind of do that where if you have to convert from the wet and dry but why it is needed? Because many times you may, as I said earlier, most of the standards are based on the dry standards. Your sludge quality standard, your compost standard or soil standards, they are all based on dry because the moisture content may change. So based on dry, the numbers you can compare. But many times when you analyze certain samples, you will analyze them as it is. You will not, it will be the just wet sample, it will not be the dry samples. You should be able to convert between wet and dry. So this is and the example of how to do that. So that is one.

(Refer Slide Time: 4:54)



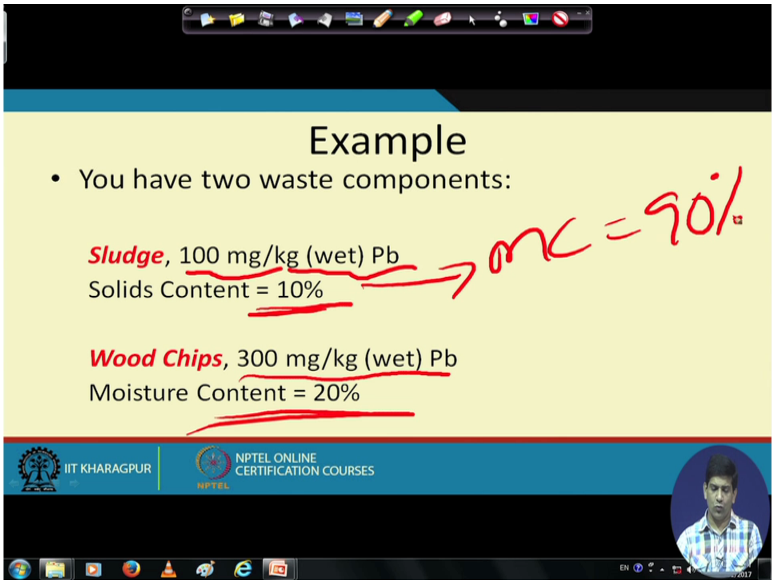
Working with Concentrations

- You should be able to find the average concentration of mixed waste streams

And then so you should be able to find the average of the mixed waste stream. So you will not have, for example, if you are working on a compost plant, you are trying to design a compost plant, you are getting waste from different sources. Waste may be coming from source A, source B, source C and for the individual source A, B and C, you know their moisture content, you know the arsenic concentration, you know the lead concentration. But for the mixed, you do not know it. And but you know in what ratio they are mixed.

Say for example, if you have 100 kg of the fleet coming into a compost pile, say 40 kg coming from source A, 30 from source B, another 30 from source C. So based on their percentage you know that, you know individual mass. Again, you will do a mass balance because all the arsenic that is coming from different sources will be there. They are not going to go anywhere and the total mass will be in the denominator and the total analyte of interest will be in the numerator and that is how we get the average concentration. Essentially it is like weighted average and you, and we will look at an example for that matter as well over here.

(Refer Slide Time: 6:00)



Example

- You have two waste components:

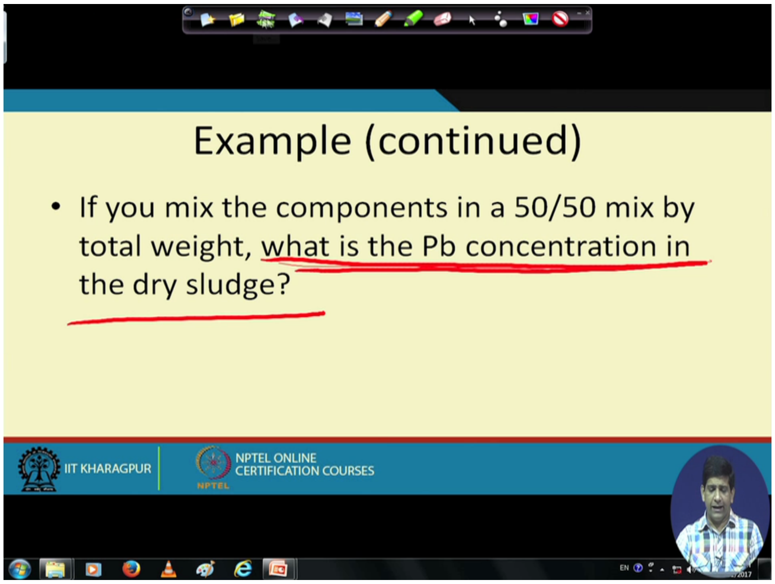
Sludge, 100 mg/kg (wet) Pb
Solids Content = 10%

Wood Chips, 300 mg/kg (wet) Pb
Moisture Content = 20%

So in terms of the concentration, if you have two waste components, we have a sludge and the wood chips. These are the two components we have. You have two waste component, there is a sludge and there is a wood chips. For sludge has a 100 milligram per kilogram, that is the lead and that is on the wet concentration. Sorry, so it is a 100 milligram per kilogram, that is in terms of wet. It is, and the solid content is 10 percent. In the earlier slide, we had that moisture content of 80 percent. Here, we have waste is, solid content is 10 percent. Just to, there are different ways you can see in different problems.

Wood chips again 300 milligram per kilogram of lead. Moisture content is 20 percent. So wood chips is not that moist but sludge is. Sludge, this particular sludge, the moisture content is 90 percent. So here the moisture content is 90 percent. So that is, so this is we have these two waste stream. No if I am, if you are mixing these two waste stream, so this is two waste like a waste stream coming in.

(Refer Slide Time: 7:09)



Example (continued)

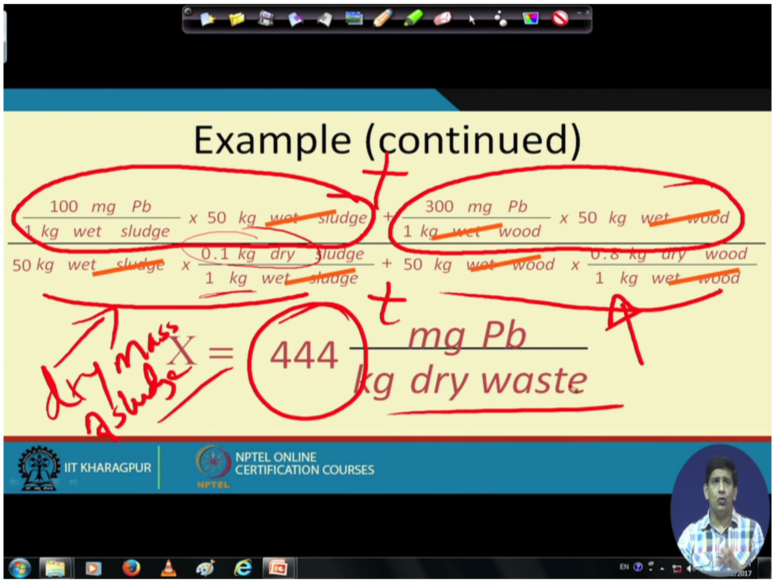
- If you mix the components in a 50/50 mix by total weight, what is the Pb concentration in the dry sludge?

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

And for say for some reason, we are mixing the components in a 50-50 mix by total weight. So we are taking, say if you are taking 100 kg, 50 kg of the sludge, 50 kg of the wood chips, we are mixing them together. And if we want, the question is asking us that what will be the lead concentration in the dry sludge. So what is the lead concentration in the dry sludge? So that is what we need to find out. So for that, what we will do?

(Refer Slide Time: 7:43)



Example (continued)

$$\frac{\frac{100\ mg\ Pb}{1\ kg\ wet\ sludge}\times 50\ kg\ \cancel{wet}\ sludge + \frac{300\ mg\ Pb}{1\ kg\ \cancel{wet}\ wood}\times 50\ kg\ \cancel{wet}\ wood}{50\ kg\ wet\ \cancel{sludge}\times\frac{0.1\ kg\ dry\ sludge}{1\ kg\ wet\ \cancel{sludge}} + 50\ kg\ \cancel{wet}\ wood\times\frac{0.8\ kg\ dry\ wood}{1\ kg\ \cancel{wet}\ wood}}$$

$$X = \boxed{444}\ \frac{mg\ Pb}{kg\ dry\ waste}$$

dry mass of sludge

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

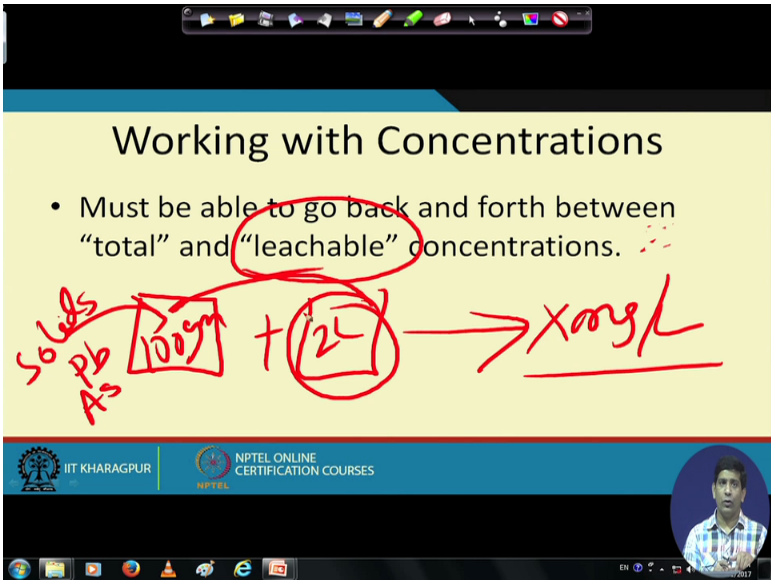
Again, it is a standard mass balance problem. Like where, so we have two sources: wet sludge and wood which we are calling it wet wood. We know that 100 milligram of lead per kg of wet sludge times 50 kg of wet sludge if you take total as 100 kg. So, and 300 milligram of lead per kg of wet wood, 50 kg of wet wood. So what we have on the, what we have on numerator? If you look at this carefully and it is pretty straightforward, so what is this? This is the mass of lead coming from the sludge. This is the mass of lead coming from the wood. If we add them together, that is the total mass of lead coming to the mixture.

And one assumption is there that mass of lead is there in the solid, it is not there in the liquid. So even if you take the moisture out, mass of lead is still there. So this is the mass of, this is two sources of lead. So source 1 and source 2, we add them together. That is the total amount of lead present in the mixture. Now, but we need to find out their concentration per dry waste. So we need to find out okay, what is the dry mass from this and what is the dry mass from this.

So how we will do that? We have 50 kg of wet sludge. We know that dry is 0.1 kg dry out of 1 kg, so it is only 10 percent solids. So this is the dry mass coming from dry for the sludge, dry mass of sludge. This is the dry mass from the wood. So we add them together, so that is the total dry mass. So what we are doing? We are doing a mass balance. Total lead on top, so this is the total lead. If we add these two together, that is the total lead on top.

Here if we add the things at the bottom, that is the total amount of dry mass and we get the number of 444 milligram of lead per kg of dry waste. We should, you go ahead and do that and you should get the same number and if you do not, let me know. But I do not think so there should be any problem here. So this is how we can do it. So this is just for two waste stream. We could have three, we could have four, we could have five but we can just extend this problem like that. So these are, these kind of problems that it is not as you can see it is not very difficult problem, you just have to make sure you do stepwise and follow the basic principles here and that should be okay. So this is on like how to calculate that.

(Refer Slide Time: 10:32)



Working with Concentrations

- Must be able to go back and forth between "total" and "leachable" concentrations.

And then if you have to go back between total and leachable, many times we are worried about what is the leachable concentration. If you remember from the TCLP test we talked about where we are, took the 100 gram of sample, added 2 liters, remember the TCLP test where how much? We had a 100 gram of sample, 2 liters of the leaching fluid, then you let it rotate and then you get the concentration of whatever x milligrams per liter.

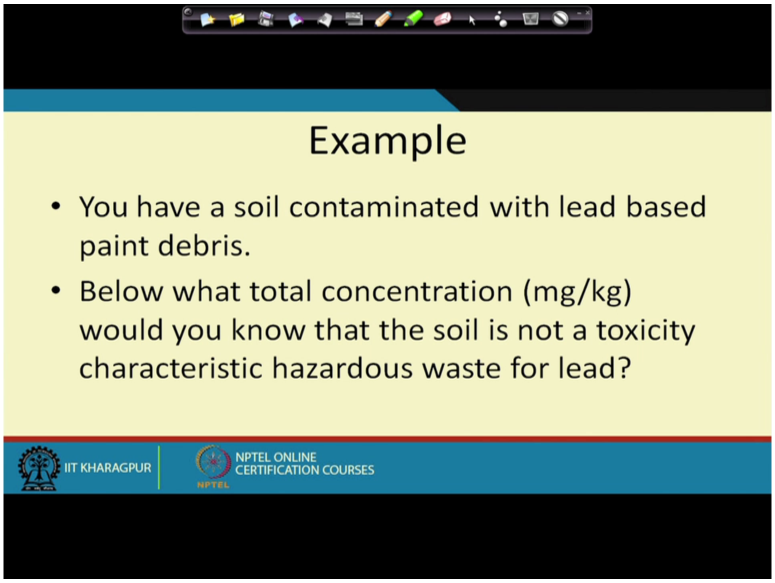
And this x milligrams per liter is whatever is the solid present here. Whatever is the solid or you can say even, say if you think about lead or arsenic or whatever, whatever lead and arsenic present in this solid phase, how much will go into this liquid phase? So once, if the things will move from the solid phase to the liquid phase, so how much will, it will move to the liquid phase, that is called the leachable concentration. How much is actually leached out? So when they are in the solid phase, they are in some sort of metrics.

So how much of that lead or arsenic will leach from the solid and go into the liquid, so that is called the leachable fraction. So that if we know the leachable fraction, we can calculate total as, if you know the leachable percent like at what fraction say 20 percent, 5 percent, 10 percent, usually you do not see much (leachable) leachability. Anything more than 20, you do not see over there, so that is over there. So you do not see this leachable concentration showing up over that part but any like you do, you will have not more than 10 percent of the leachable concentration.

So and if you know the total, you can assume certain leachable concentration. And the worst case scenario will be what? Worst case scenario will be the all the solids present. Say if you have certain amount of lead present here, the entire lead goes into the liquid. So that will be the worst scenario where the entire lead present in the solid phase will go into the liquid phase. Normally, it will not happen but worst case scenario that is there.

If you put a very, very strong acid and also depends on what kind of leaching fluid you have. For TCLP test, we use acetic acid and sodium hydroxide. That is not a very strong acid but if you put acquaragia, yes, most of it will go into the liquid phase. So depends on what kind of extraction fluid you use there as well. So we will, I think we have a problem on, we will some math relating to that as well.

(Refer Slide Time: 12:48)



Example

- You have a soil contaminated with lead based paint debris.
- Below what total concentration (mg/kg) would you know that the soil is not a toxicity characteristic hazardous waste for lead?

So if you have a soil contaminated with lead based paint debris, now lead based paint is, we have been, in India I think still some of the paint being sold is based on lead, the lead compounds are there. But most of the western world has banned lead based paint in 1976 if I remember correctly, so around that time. So around 1976, most especially in US and Europe, western European countries probably they did little bit earlier.

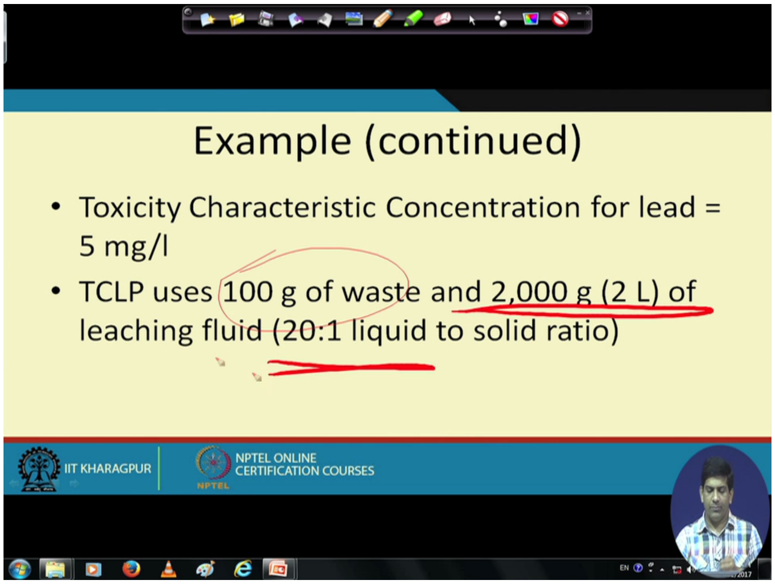
But this lead based paint is banned because remember that small kids, they will have, if you have a lead based paint on the wall, the kid has a tendency to like even pill things from the wall and then put them in the mouth, so lead will go into their body and then as we know lead is not good, especially for it affects the brain development. That is the reason I think I already told you that we have unleaded petrol now. Unleaded petrol is, that means what? We had leaded petrol at some point of time, so that is why it is unleaded now. So we do not use lead anymore in our petrol.

So similarly we have many countries do not use lead in their paints. Many, in India also there is some companies actually, they advertise saying that they do not have lead in their paints but I do not think it is really bad or it may be if I may be wrong. I am not 100 percent sure on that. But say if, but that is, that does not matter for this particular problem. Here we are talking about there is a soil contaminated with lead based paint debris which means the lead, the source of lead is the lead based paint coming from there.

And we want to find out that below what total concentration would you know that the soil is not a toxicity characteristic hazardous waste for lead. It is trying to trying to remind you about the TCLP test. So it is kind of a recap of the TCLP or toxicity characteristic leaching procedure, it is that is the TCLP test. Remember that we talked about that earlier. So if you have, if you do not remember, go back and look at that slides or video. So this TCLP test, we take 100 grams and use 2 liters of the leaching fluid and then whatever is the concentration comes out, we compare it with the regulatory standards for the TC standard.

And for lead, it is 5 milligrams per liter. We talked about that earlier. So for, this is 5 milligrams per liter. Now below what concentration, would you know for sure? So as I was just telling you few minutes back, so if you try to do this worst case leaching scenario where the entire lead present in the solid phase comes into the liquid phase, that will be the worst case leaching scenario. So if we have a concentration in which and say whatever is total lead present in the solid comes to the liquid, if we have a concentration less than that, that particular waste will never be hazardous waste.

(Refer Slide Time: 15:51)

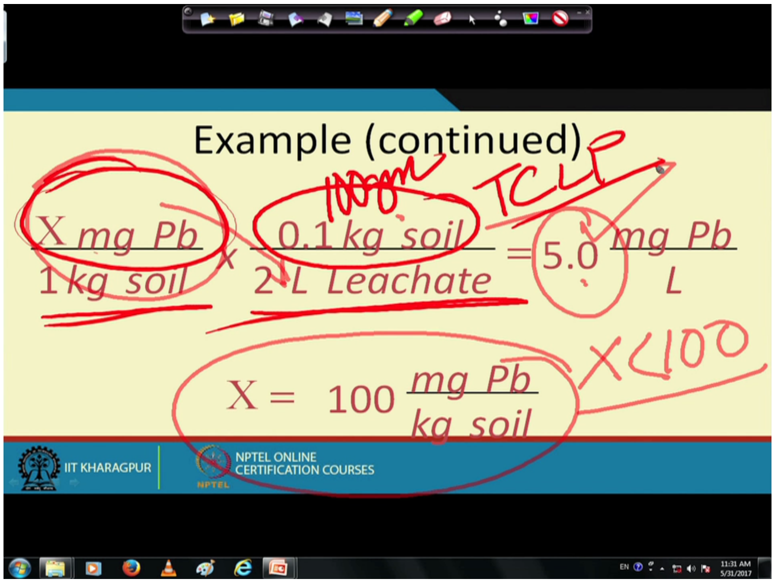


Example (continued)

- Toxicity Characteristic Concentration for lead = 5 mg/l
- TCLP uses 100 g of waste and 2,000 g (2 L) of leaching fluid (20:1 liquid to solid ratio)

So let us do the math and I will try to explain it again. And I think, and just try to follow it. So there is so you have, so toxicity characteristic concentration for lead is 5 milligrams per liter. We talked about that TCLP uses 100 grams of waste, it uses 100 grams of waste and 2 liters of the leaching fluid. So 20 is to 1 is the liquid to solid ratio, that is the typically used for TCLP. That is what is TCLP test requires.

(Refer Slide Time: 16:12)



Example (continued)

$$\frac{X\ \text{mg Pb}}{1\ \text{kg soil}} \times \frac{0.1\ \text{kg soil}}{2\ \text{L Leachate}} = 5.0\ \frac{\text{mg Pb}}{\text{L}}$$

$$X = 100\ \frac{\text{mg Pb}}{\text{kg soil}}$$

100gm

TCLP

X<100

So here say if we had x milligram of, but the question was, how much milligram of lead could be there in that lead based paint contaminated debris? Which the concentration, what is the maximum concentration that lead based paint debris soil sample can have so that it will not fail TCLP test? So that is what essentially the question was, so that it will not be a toxicity characteristic hazardous waste for lead.

So let us assume that number is x. So x is the amount, x milligram of lead is there, per kilogram, x amount of lead per kilogram of the soil. So we have x milligram of lead per kilogram of soil. So that is we have to find out this x. Now what is the second term here? We have used 0.1 kg of soil and 2 liters of leachate. Now where this is coming from? This is coming from the TCLP procedure. Remember the TCLP procedure, just in the previous slide also I viewed like you take 100 grams.

So 100 grams becomes 0.1 kg. Since here we have kilogram, I have put kilogram over here. So it is 100 grams, essentially it is a 100 gram. And then, 2 liter of the leaching fluid. So what we are assuming here that entire lead which is present on this, entire lead which coming from the, the entire lead that is coming from there is getting into the leachate. Entire lead which is present in the soil is getting into the leachate. That is no more lead left in the solid phase. Every lead that was present went into the liquid phase. That is the worst case leaching scenario.

So if that happens and if the concentration reaches 5 milligrams per liter, then there will be possibility in some scenarios, in the worst case leaching scenario, there will be possibility of this particular lead based paint soil debris to be a hazardous waste. So what is that concentration? That concentration if you do this math over here, and this 5 comes from the TC limit, this comes from the TCLP procedure. So we get x to be 100 milligram of lead per kilogram of soil. So this is the (max) this is the like amount of concentration. If the concentration is below this, if the concentration is less than 100, if your x is less than 100, that means that under no circumstances it will be hazardous waste. Because even if there is 100 percent leaching, the TC, it will not reach the TC limit even with the 100 percent leaching.

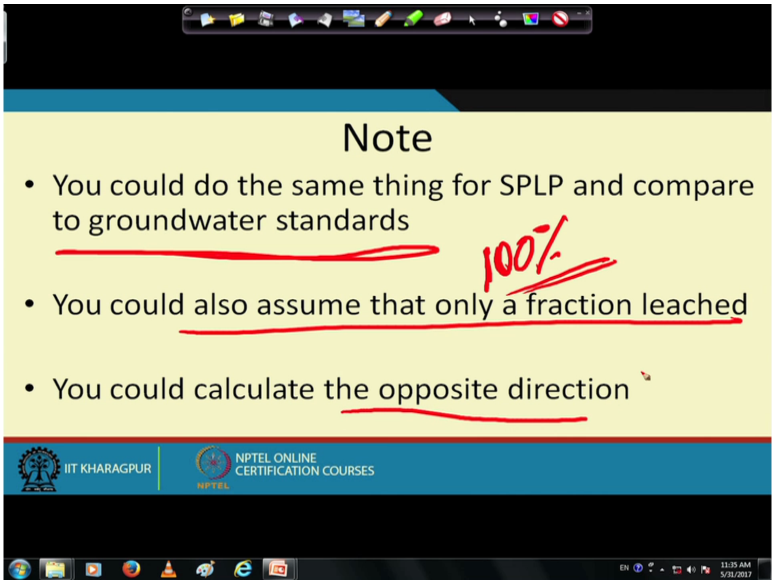
So but if it is concentration is above 100, there will be scenario, there could be a scenario. Theoretically, it is possible for it to become hazardous waste because the concentration will exceed, may exceed this 5 milligrams per liter. I hope I have made this clear. Again watch this part of the video. Again if you have to read, kind of go over that too. And you can always ask questions as well. So this is again important. These are very, very important concept. These are small small problems but lot of application based.

Many times there will be a, what is a significance of this problem? Many times there might be a industry coming to you saying that like I have this particular waste, I have done this total analysis done and my, I do not have any other elements there or lead is only lead and arsenic are the issues or only lead is present or it could be any particular element. Lead was used as example here.

Say even it was only lead is present and the total concentration I got was, is less than, you get you look at the total concentration and you find that they are all less than 100 milligram per kilogram. And if the person is asking you, should I worry about this being a hazardous waste? Your answer will be no. Because it is, even in a worst case scenario, this is not going to fail TCLP. So it will not be hazardous waste at least for the lead. It could be hazardous waste for some other parameters but not for the lead.

And same thing you can do for other elements too. So that is, are the things you need to look at. So it is a very application oriented example problem. So these are the things that you will do when you are working for a particular company or you are working for say environmental protection agency, CPCB, SPCB and all those kind of stuff.

(Refer Slide Time: 20:42)



Note

- You could do the same thing for SPLP and compare to groundwater standards
- You could also assume that only a ~~fraction~~ leached 100%
- You could calculate the opposite direction

So you could do the same thing for SPLP. We have not talked much about SPLP. SPLP is a synthetic precipitation leaching procedure which we try to simulate artificial rainfall and that is done when we are trying to use something outside in the landfill. It is a beneficial reuse. When we go to the C&D part, do not worry too much about this SPLP right now. When we go to the C&D, only thing you remember about SPLP, that is a synthetic precipitation leaching procedure. It is used to simulate artificial rainfall.

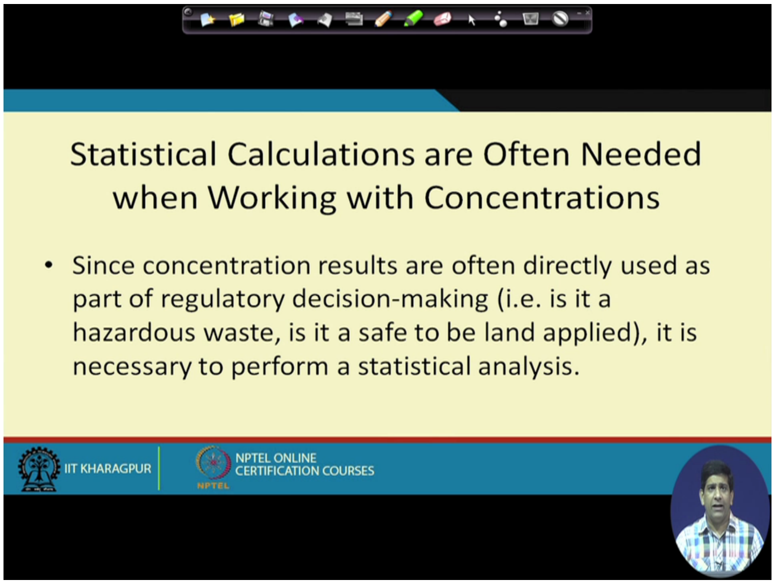
So whenever we are trying to use a waste material outside the landfill for any short of construction activity, this SPLP test is, we can predict using this SPLP test whether what will be the leaching and how this can be like how the leaching can happen. We can take the data, try to compare whether there could be a potential groundwater contamination and all those kind of stuff. So that is why when we do this SPLP test, we compare that with the groundwater standard. But we will come back and spend more time on this SPLP thing when we talk about construction and demolition waste.

Because the C&D landfill, they are mostly like when they are exposed to rainfall, we are trying to use C&D waste in, outside of the landfill. So we will talk about more in that context. And here remember that just now the problem that we did, we assumed that there is 100 percent leaching. That assumption was there, that we assumed it is a 100 percent leaching happening. But you can also assume that only a fraction leached, we can, because 100 percent leaching usually does not happen. So you can assume that 50 percent leached out, 40 percent leached out based on the experience that you have.

So you have, once you are working in industry or work for several years, you know that under normal circumstances, it does not really leach out. You never, you have never seen more than 50 percent leaching happening for that particular matrix, that particular waste type. To make it a more, I would say more realistic judgment, you can say that okay, I will assume it 50 percent leachability because that is the maximum we have ever seen happening. And then based on that, you can do the calculation as well. And we can do the calculation in the opposite direction too.

So it is once we know the leaching, we can know what is, what could be the maximum concentration happening and all that. So it can be done both ways. So these are some examples which is kind of relevant for that.

(Refer Slide Time: 22:58)



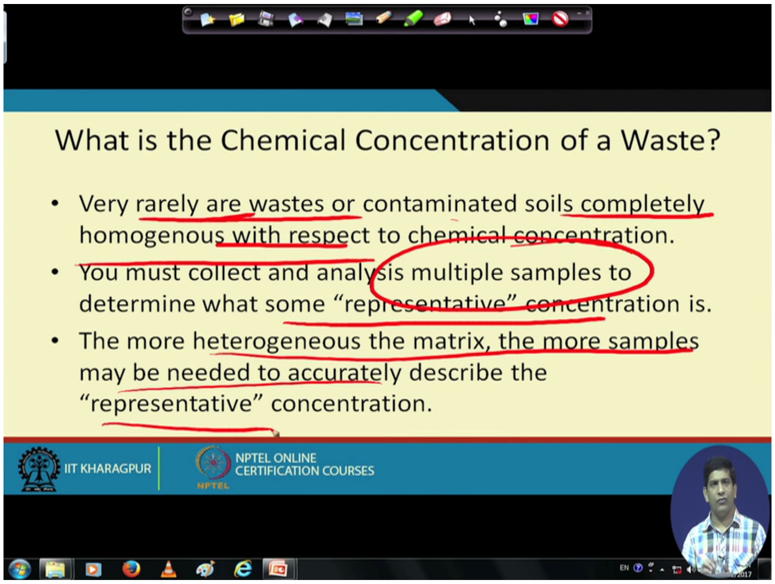
Statistical Calculations are Often Needed when Working with Concentrations

- Since concentration results are often directly used as part of regulatory decision-making (i.e. is it a hazardous waste, is it a safe to be land applied), it is necessary to perform a statistical analysis.

And then we will try to look at some of these statistics calculation. So this is actually very, very interesting part where you do the statistical calculation often needed when working with. So many times what we do, we go out. You are trying to characterize a waste. You go out and collect some samples. More the heterogeneity, more number of samples you should collect. You bring the samples to the lab, you analyze the samples and you get the data. Now you have to go from this sample mean to the population mean and that is where your confidence interval comes in picture as well. So we will look at an example of that.

And this dataset that I am going to present you, the two datasets, they are a hypothetical dataset and it is not the real dataset. It is just to illustrate some points but you will have, you may encounter something similar in a real world scenario as well. So here we are trying to do concentration. It is, whether it is a hazardous waste or a land applied, we do some statistical analysis. So let us look at them really quick.

(Refer Slide Time: 23:55)

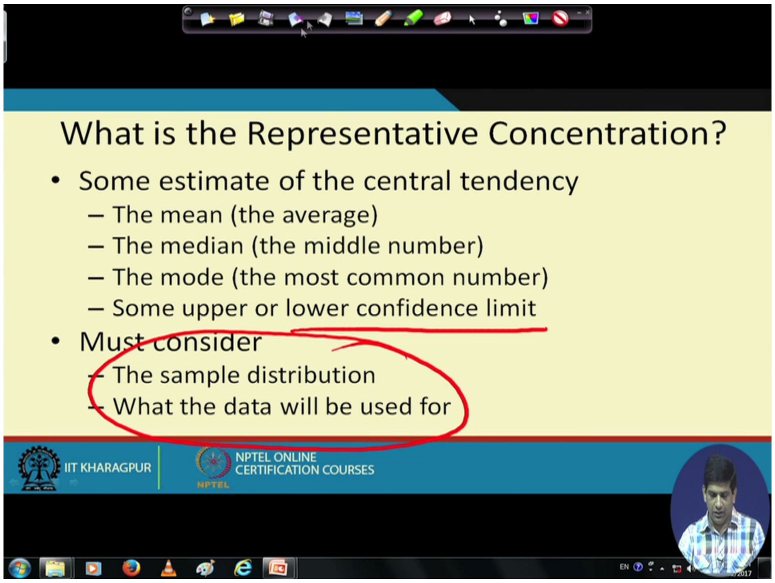


What is the Chemical Concentration of a Waste?

- Very rarely are wastes or contaminated soils completely homogenous with respect to chemical concentration.
- You must collect and analysis multiple samples to determine what some "representative" concentration is.
- The more heterogeneous the matrix, the more samples may be needed to accurately describe the "representative" concentration.

So when the chemical, as I said very rarely the waste will be homogeneous. They are not homogeneous, so they are not homogeneous. Very rarely you will, it is rarely you will see homogeneous sample. So what that means? You do multiple samples to determine the representative concentration. And more the heterogeneity, more samples needs to be collected to get the representative concentration. So we already talked about that just now.

(Refer Slide Time: 24:26)

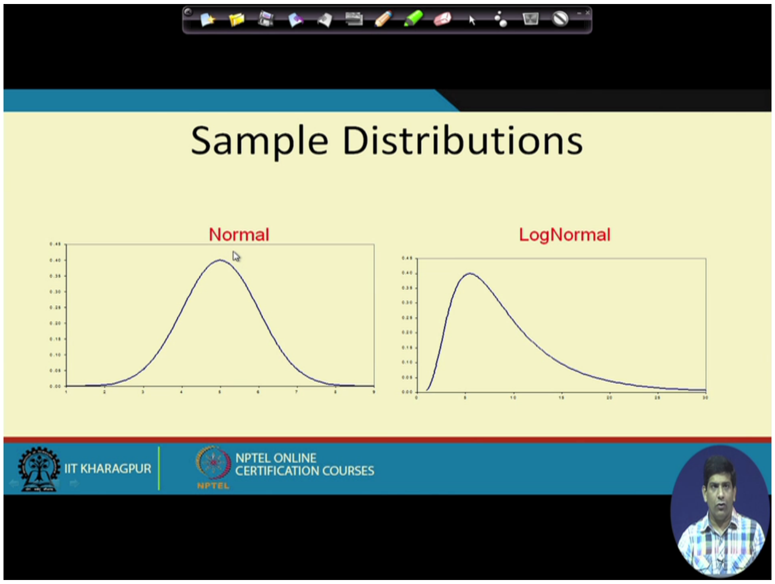


What is the Representative Concentration?

- Some estimate of the central tendency
  - The mean (the average)
  - The median (the middle number)
  - The mode (the most common number)
  - Some upper or lower confidence limit
- Must consider
  - The sample distribution
  - What the data will be used for

So once the sample is done, then you try to find out, you analyze the sample. Then you do some sort of and central tendency. Again from your, if you have not taken a statistics or if you have taken statistics long time ago, go and revise it. So there is a term like mean, median, mode, upper or lower confidence. And then you look at the sample distribution, whether it is a normal distribution, lognormal distribution and what the data will be used for those kind of things comes in picture as well.

(Refer Slide Time: 24:52)



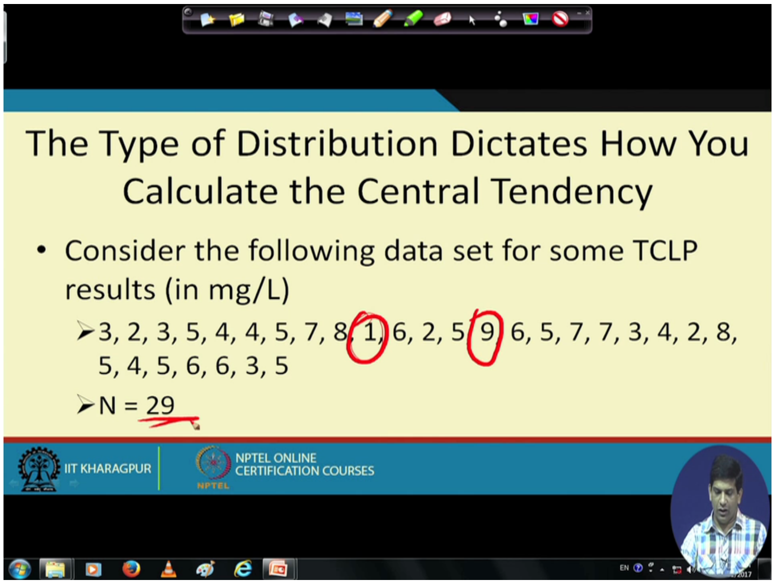
Sample Distributions

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

So in terms of sample distribution, we know this is a typical log distribution that the bell shaped curve and this is a lognormal where you have a tail extending on one side. So that is a lognormal distribution. So this is our the lognormal part. This is a normal distribution which you all know.

(Refer Slide Time: 25:07)



The Type of Distribution Dictates How You Calculate the Central Tendency

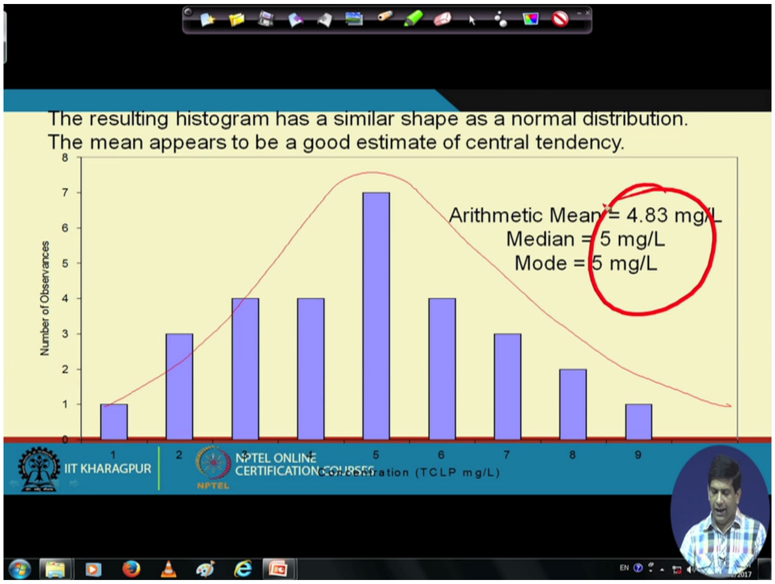
- Consider the following data set for some TCLP results (in mg/L)
  - 3, 2, 3, 5, 4, 4, 5, 7, 8, 1, 6, 2, 5, 9, 6, 5, 7, 7, 3, 4, 2, 8, 5, 4, 5, 6, 6, 3, 5
  - N = 29

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

So now if you look at this data, so there is a, we say we have done a TCLP test. Since TCLP is a very common test used, so we have done some TCLP test. And we collected some sample, did some TCLP test and we found these numbers. So as you can see there is a lot of variability in the number. I think the minimum concentration is 1 and the maximum is 9 and so between 1 and 9 is the value that we have. And total number of samples is 29. So now if you want to, try to look at how these samples when we try to do some analysis part of that.

(Refer Slide Time: 25:43)



The resulting histogram has a similar shape as a normal distribution.
The mean appears to be a good estimate of central tendency.

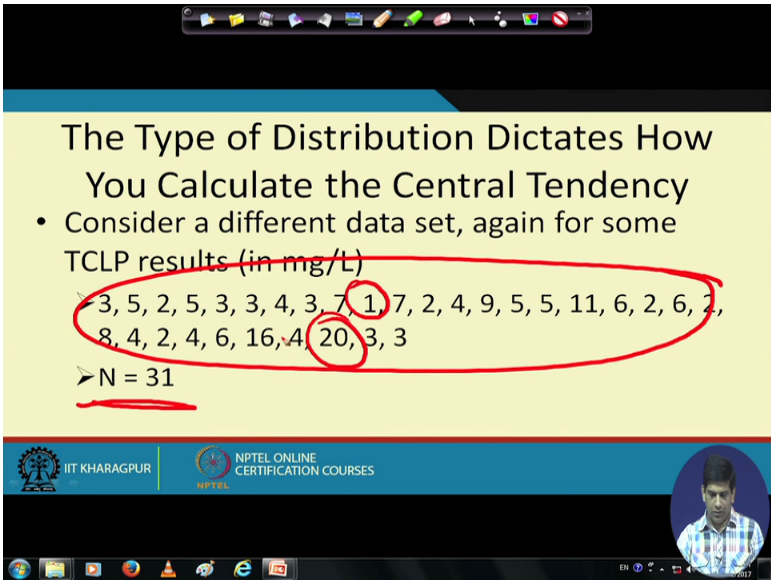
Arithmetic Mean = 4.83 mg/L
Median = 5 mg/L
Mode = 5 mg/L

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

So if you make a histogram of that, we see a similar shape as a normal distribution. If you kind of think about this as kind of something going like this and kind of coming down that, so that is like a normal distribution, is not it? We just looked at it earlier. So it is a normal distribution kind of curve. And then you can calculate arithmetic mean, your median, your mode. So those things here, you get those numbers in terms of the central tendency. And there is a way to find out whether it is a normal distribution or not. Like I would say it is, there is a (())(26:18) thumb rule to find out and we will talk about that.

(Refer Slide Time: 26:20)

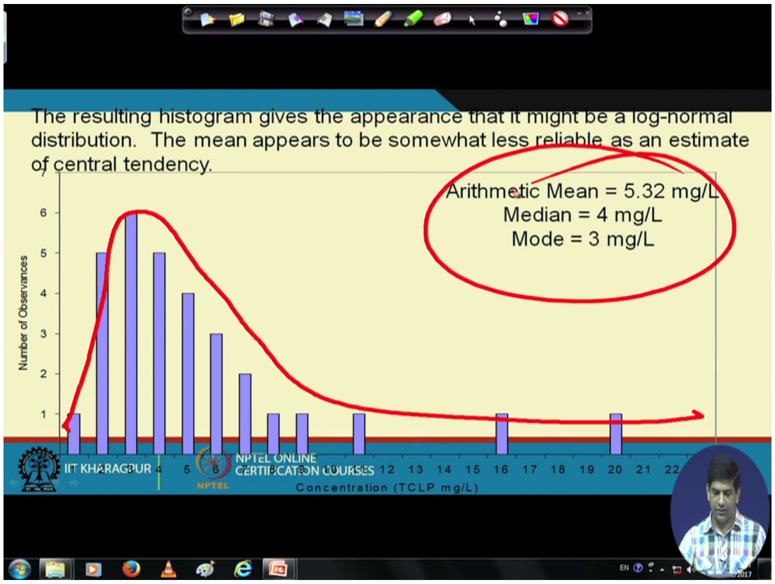


The Type of Distribution Dictates How You Calculate the Central Tendency

- Consider a different data set, again for some TCLP results (in mg/L)
  - 3, 5, 2, 5, 3, 3, 4, 3, 7, 1, 7, 2, 4, 9, 5, 5, 11, 6, 2, 6, 2, 8, 4, 2, 4, 6, 16, 4, 20, 3, 3
  - N = 31

And then you have a second set of dataset where we had some samples, samples were done where we have number of samples is 31. And these are the data we got. And in terms of the data that we got, the minimum value here is again 1 but the maximum value goes all the way up to 20. So it lot more variability than what we saw in the previous like previous distribution.

(Refer Slide Time: 27:02)

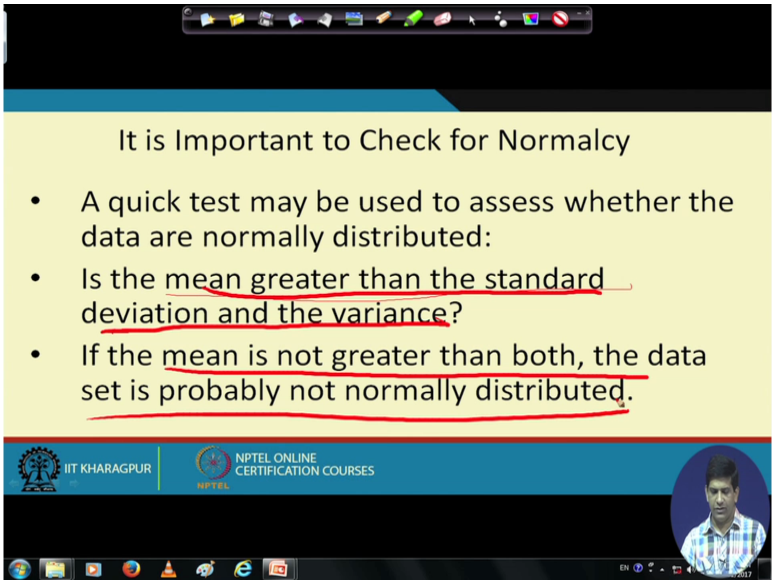


The resulting histogram gives the appearance that it might be a log-normal distribution. The mean appears to be somewhat less reliable as an estimate of central tendency.

Arithmetic Mean = 5.32 mg/L
Median = 4 mg/L
Mode = 3 mg/L

So if we do the same histogram for this sample, we see the histogram more like a lognormal distribution that you saw. Like you have something like this and then you have a long tail. So that is a typical lognormal distribution and then you can do this arithmetic mean, median and mode for that too.

(Refer Slide Time: 27:20)

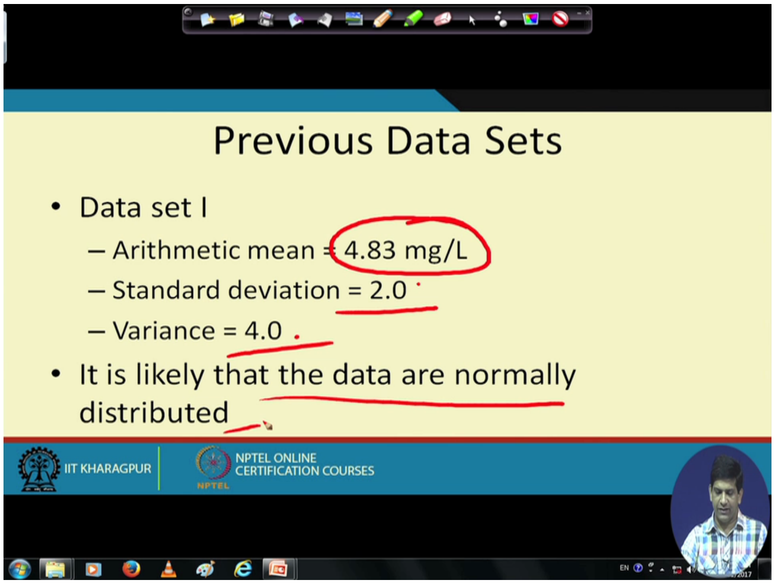


It is Important to Check for Normalcy

- A quick test may be used to assess whether the data are normally distributed:
- Is the mean greater than the standard deviation and the variance?
- If the mean is not greater than both, the data set is probably not normally distributed.

So in terms of the thumb rule what we do, first we try to check for the normalcy. So quick test is used to assess whether the data is normally distributed or not. So for that, we do whether the mean is greater than the standard deviation, that the mean is greater than the standard deviation and the variance. So that is the one way of doing it. If the mean is greater than the standard deviation or the variance, we do that. If mean is not greater than both, the dataset is probably not normally distributed. If the mean is not greater than both, if not greater than standard deviation and variance, that means the data is not normally distributed. So that is the like the part for that part.

(Refer Slide Time: 27:56)

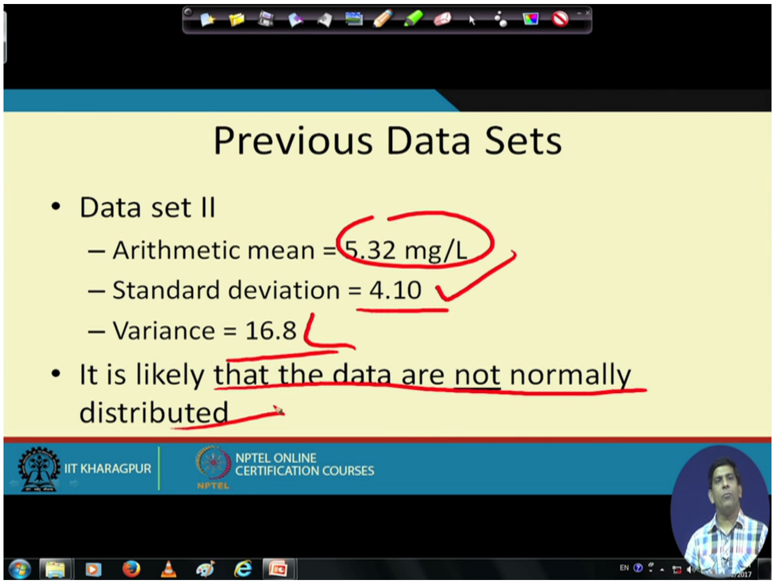


Previous Data Sets

- Data set I
  - Arithmetic mean = 4.83 mg/L
  - Standard deviation = 2.0
  - Variance = 4.0
- It is likely that the data are normally distributed

So like let us look at the dataset. So Dataset I, again you go back and do this math yourself. I am just, you can do this, this is a simple math. So you can do this, arithmetic mean is 4.83, standard deviation is 2, variance is 4. Since the mean is greater than both 2 and 4, so it is likely that the data is normally distributed. So that is for the Dataset I and we saw that when we plotted that histogram as well.

(Refer Slide Time: 28:23)

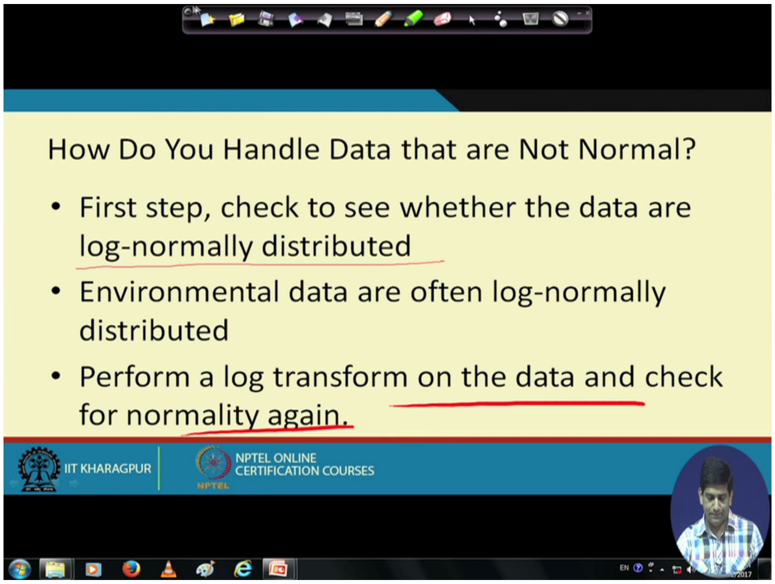


Previous Data Sets

- Data set II
  - Arithmetic mean = 5.32 mg/L
  - Standard deviation = 4.10
  - Variance = 16.8
- It is likely that the data are not normally distributed

For the second one, arithmetic mean is 5.32. Sorry, so arithmetic mean is 5.32, standard deviation is 4.10, variance is 16. So since the arithmetic mean, although it is greater than the standard deviation but it is less than the variance. So that means it is likely that the data is not normally distributed. So once the data is not normally distributed, we have to kind of find out okay, so what kind of distribution it may have. With the histogram that we did, it kind of felt like it is a lognormal distribution with the when we plotted the histogram.

(Refer Slide Time: 28:57)



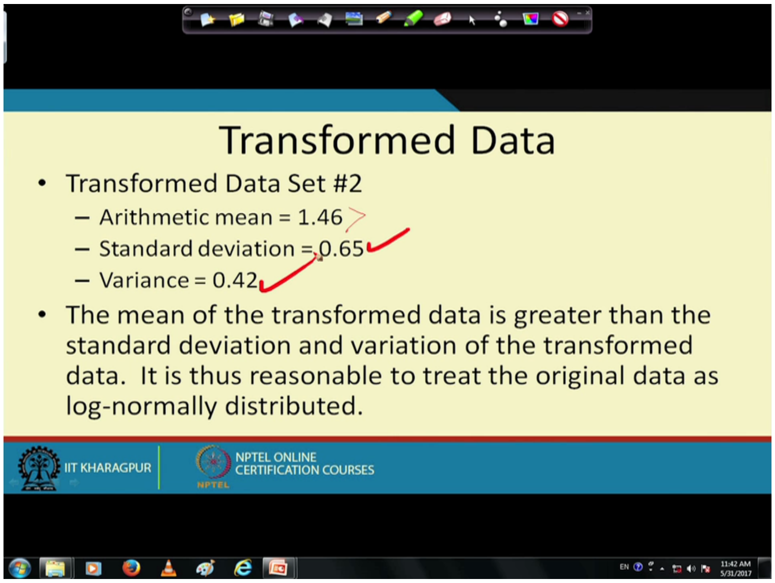
How Do You Handle Data that are Not Normal?

- First step, check to see whether the data are log-normally distributed
- Environmental data are often log-normally distributed
- Perform a log transform on the data and check for normality again.

So once you do the Handle Data which is not normal because remember that T-test, G-test, Chi-square and all those that confidence interval and everything that we do, it is based on assuming normal distribution. So that you need to make sure the data is normal distributed like the one, the equations that we are using. So how do you handle the data which is not normal? First thing is check whether the data is lognormal. So that is the first step we do, whether the data is lognormal distributed. Most of the environmental samples, they are oftenly log-normally distributed, so you do not see them normal distributed.

So how you do that? You log transform the data, you take the log of the data and do this histogram again and look for mean, mode, median, again for check for normalcy, normality or normalcy again, you do that.

(Refer Slide Time: 29:47)

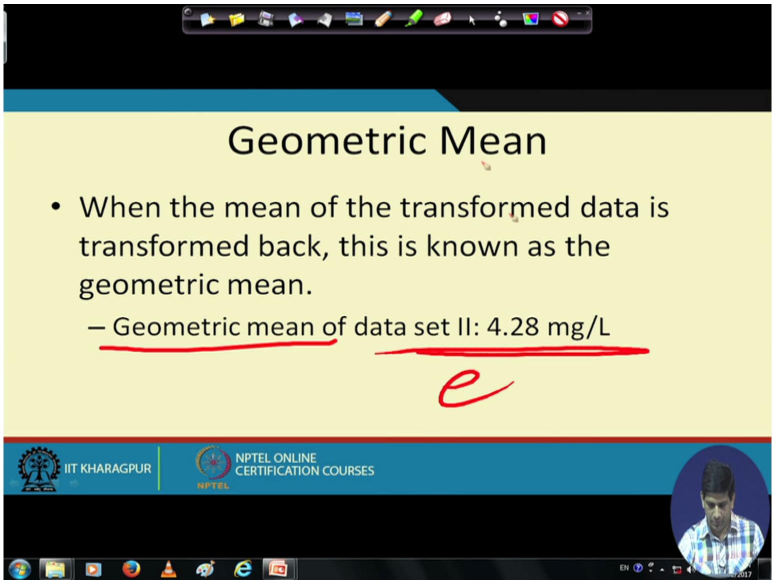


Transformed Data

- Transformed Data Set #2
  - Arithmetic mean = 1.46 >
  - Standard deviation = 0.65 ✓
  - Variance = 0.42 ✓
- The mean of the transformed data is greater than the standard deviation and variation of the transformed data. It is thus reasonable to treat the original data as log-normally distributed.

And so if you do the transformed dataset for the number two, we found the arithmetic mean to be 1.46 which is greater than both standard deviation as well as the variance. So since 1.46 is greater than this as well as this, so that means we are okay. They are most likely it is, we can treat this data as, original data is lognormal distributed and the log transformed data is normal distributed. So that is typically that you see in most of the environmental samples as well.

(Refer Slide Time: 30:16)

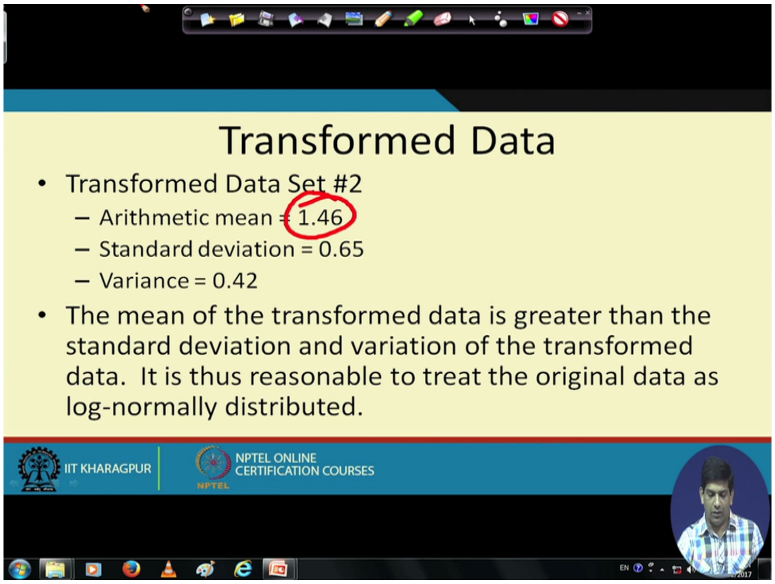


Geometric Mean

- When the mean of the transformed data is transformed back, this is known as the geometric mean.
  - Geometric mean of data set II: 4.28 mg/L

e

(Refer Slide Time: 30:27)



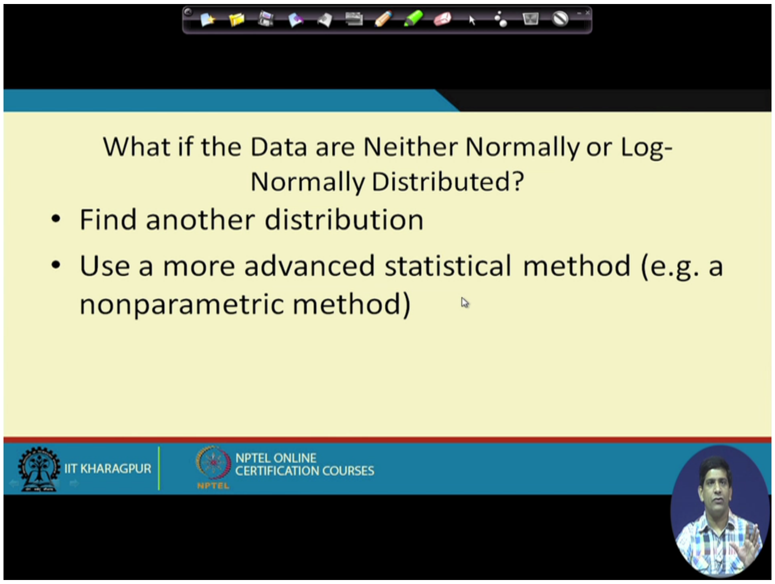
Transformed Data

- Transformed Data Set #2
  - Arithmetic mean = 1.46
  - Standard deviation = 0.65
  - Variance = 0.42
- The mean of the transformed data is greater than the standard deviation and variation of the transformed data. It is thus reasonable to treat the original data as log-normally distributed.

So once we have done that, we, but there is a term called geometric mean, you may have heard about this term. Whenever the data is lognormal distributed, when we take that e to the power of that data, like earlier if you remember we had the arithmetic mean of 1.46. So the arithmetic mean for this particular, this is 1.46. But when we, so that is the arithmetic mean for the transformed data.

So when we try to go for the (lognorm) like the original data, we call it a geometric mean. And that is this geometric mean is your, this geometric mean data is like is 4.28 which is e to the power of (1.) whatever was the previous number. Let us see what was that number. It was, it is 1.46. If you do e to the power of 1.46, you will get 4.28. So that is, and this since it is a lognormal, we call them geometric mean so that we can differentiate between the dataset which is normally distributed and the lognormally distributed.

(Refer Slide Time: 31:21)

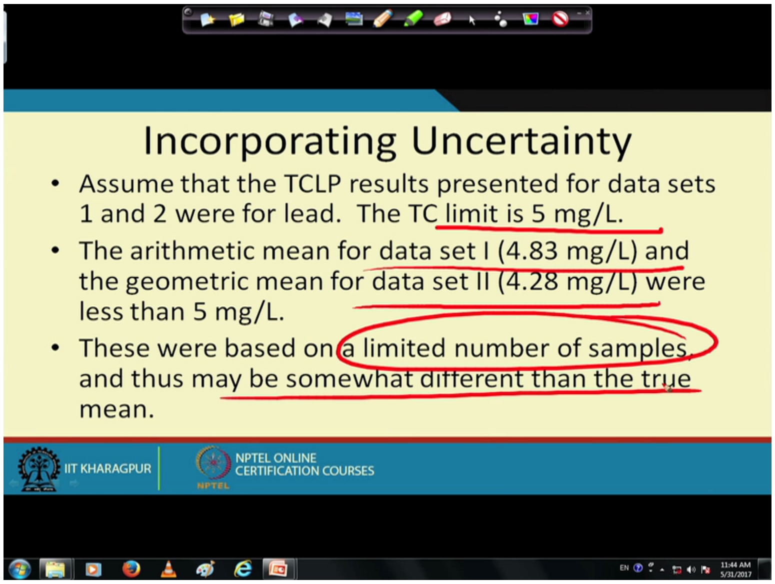


What if the Data are Neither Normally or Log-Normally Distributed?

- Find another distribution
- Use a more advanced statistical method (e.g. a nonparametric method)

So if the data is neither lognormal or lognormal, we can find another distribution, you can use more advanced statistical method. Nonparametric method will not go there in this particular class but there are ways to do it. There are ways to handle it. You can talk to your statistics professor.

(Refer Slide Time: 31:35)



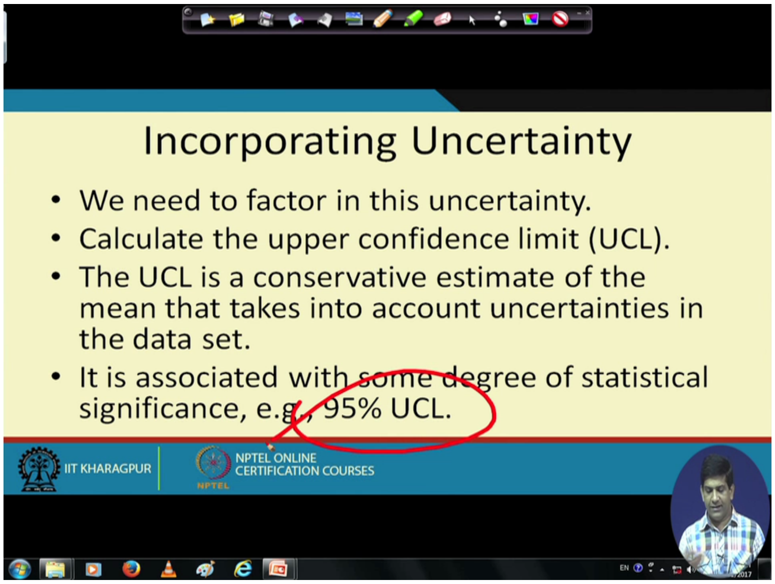
Incorporating Uncertainty

- Assume that the TCLP results presented for data sets 1 and 2 were for lead. The TC limit is 5 mg/L.
- The arithmetic mean for data set I (4.83 mg/L) and the geometric mean for data set II (4.28 mg/L) were less than 5 mg/L.
- These were based on a limited number of samples, and thus may be somewhat different than the true mean.

So how we incorporate uncertainties? Say let us assume that this whole data, we said that this data was for TCLP result and let us assume it was for lead because we have been using with lead just for our like for our argument sake. And we know the TC limit for lead is 5 milligrams per liter. So if we can take this data, arithmetic mean for dataset was 4.83. The geometric mean for dataset is 4.28. So they are both below 5 milligrams per liter, that means it is not a hazardous waste. So it is not a hazardous waste.

So based on the limited number of samples that we have, we can somewhat, we can say that it is not a hazardous waste. But there although this is a limited number of sample and the data is very close to 5, so maybe somewhat we are not sure whether this the mean that we got is very same as that will be the true mean. Again from the sample mean to the population mean, so that is what we are talking about here. So for that we try to incorporate what is the uncertainty. So let us try to do that.

(Refer Slide Time: 32:39)

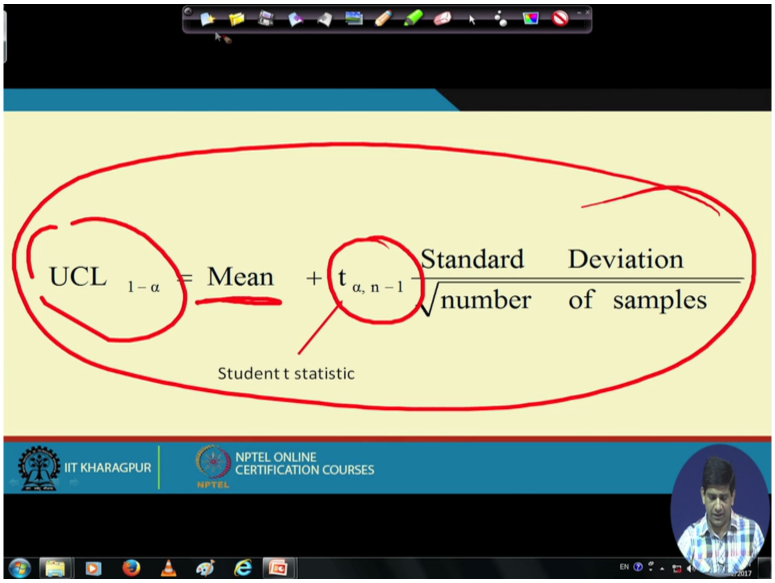


Incorporating Uncertainty

- We need to factor in this uncertainty.
- Calculate the upper confidence limit (UCL).
- The UCL is a conservative estimate of the mean that takes into account uncertainties in the data set.
- It is associated with some degree of statistical significance, e.g., 95% UCL.

And when we try, when we do this uncertainty, we need to do this upper confidence limit. So UCL is the conservative estimate to take about the dataset. So let us, and there is always some degree of statistical significance. So if you want, if you are interested into find out 95 percent UCL and again there is a formula for that. I will not go into derivation of that formula. I will just use that formula. Derivation will be covered in your statistics class, should be covered in that.

(Refer Slide Time: 33:05)

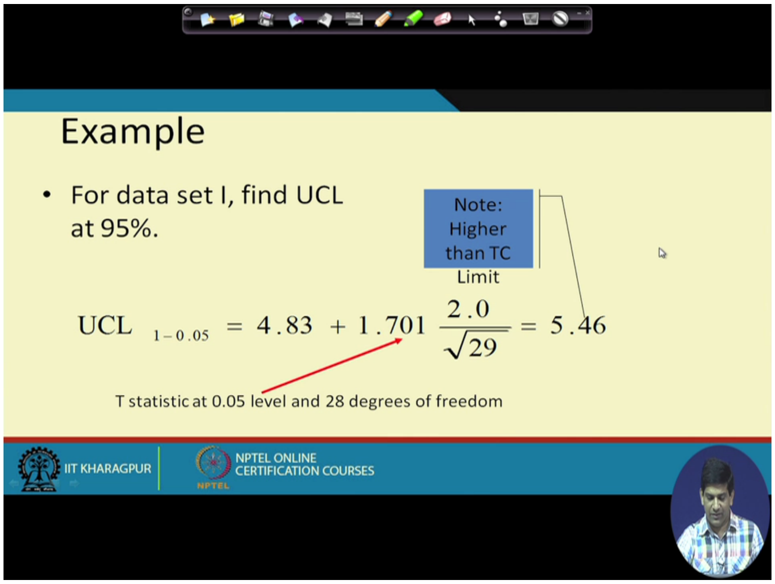


$$\text{UCL}_{1-\alpha} = \text{Mean} + t_{\alpha,\, n-1} \frac{\text{Standard Deviation}}{\sqrt{\text{number of samples}}}$$

Student t statistic

So this is the typical formula that we use where the upper confidence limit at 95 percent interval or 90 percent interval whatever you want. And so that is, this is the what we are trying to find out. It is the mean plus+ the t value where alpha is the your what the degree of significance, n minus- 1 is the degree of freedom. Then you have the standard deviation divided by the number of, square root of the number of samples. So that is the formula that is used for that, so to find out the UCL.

(Refer Slide Time: 33:44)



Example

- For data set I, find UCL at 95%.

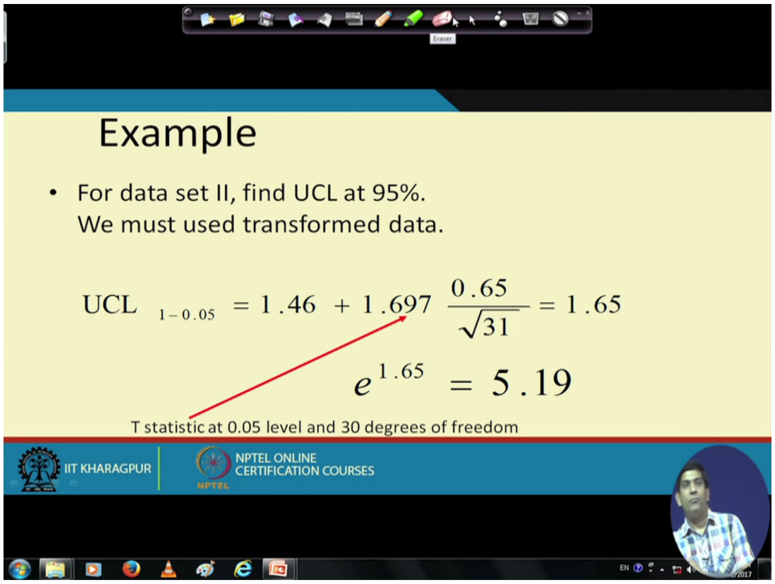
Note: Higher than TC Limit

$$UCL_{1-0.05} = 4.83 + 1.701\frac{2.0}{\sqrt{29}} = 5.46$$

T statistic at 0.05 level and 28 degrees of freedom

So if you want to find this UCL, so for these two dataset that we had, for the Dataset I, we find the UCL at 95 percent. So we just plug in the number, 4.83. This is the t value that we got from the t table, we plug in all these number, 29, number of samples and we get a value of 5.46. Note that this value is higher than 5 milligrams per liter. So that is the interesting part here.

(Refer Slide Time: 34:07)



Example

- For data set II, find UCL at 95%.
  We must used transformed data.

$$UCL_{1-0.05} = 1.46 + 1.697 \frac{0.65}{\sqrt{31}} = 1.65$$

$$e^{1.65} = 5.19$$

T statistic at 0.05 level and 30 degrees of freedom

And same thing if you do for the second set, again you do, again we have to do this UCL on the log transformed data. So that is why you see this data over there. And then you do e to the power of that 1.65 to find out the actual and then again you get 5.19.

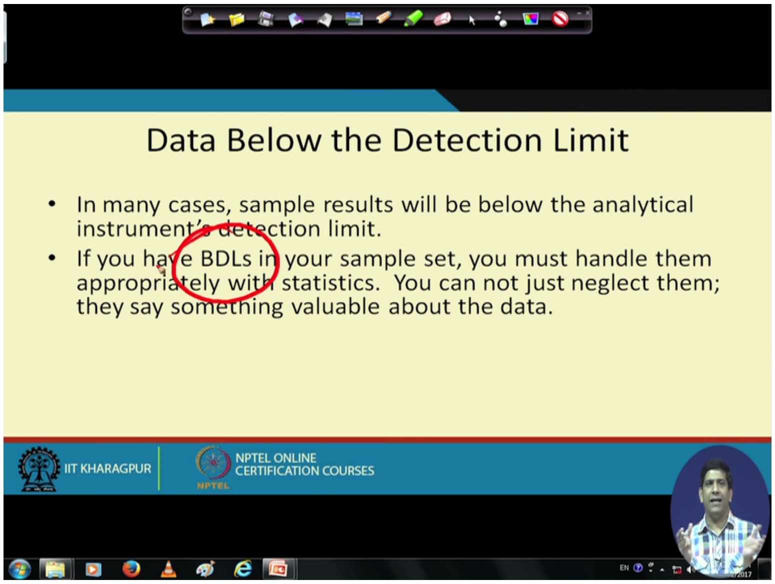
So both of these dataset although when we took the sample mean they were, when we took the sample mean, their values were less than 5 milligrams per liter. But when we took this UCL, the values are more than 5 milligrams per liter. And the 5 milligrams per liter was the TC limit. So now we are in dilemma whether it is a hazardous waste or not a hazardous waste. So let us discuss that little bit more. So that is like how you handle that.

So if you, so in that case, what you will do? Because managing anything as, any non-hazardous waste as a hazardous waste becomes too much of costly. So as a regulator, so if you are a regulator and you are looking at this data, how will you respond? What will be your, what decision you will take? One thing could be that, that means there is a lot of heterogeneity in the sample. So probably we need to go back and do more sampling.

If you are a highly conservative regulator, you will say that okay I am just, this is a hazardous waste because I did the statistics on it and I found that your data is above 5 although it is not too much above 5. That is actually more critical case and so it is, I will consider it hazardous waste. If I am the industry person who is, who has produced this waste, I will say since the data is so close and it is just above 5 and if you do the sample minute is below 5, I think that it should not be considered as a hazardous waste.

Or I should be given more time or more I will go back and do more sampling, maybe rather than doing 30 sample that we did, we will do 50 samples or 60 samples or even 100 samples and that may come out to be cheaper than managing it as a hazardous waste. And then try to convince the regulator that this is not a hazardous waste or ultimately we may find that it is a hazardous waste. So this thing of the, this kind of things will come up when you are in a practically doing things in the field. So that is the reason I wanted to illustrate that example. And I hope you like that. This is really interesting part of like how things may come up.

(Refer Slide Time: 36:21)



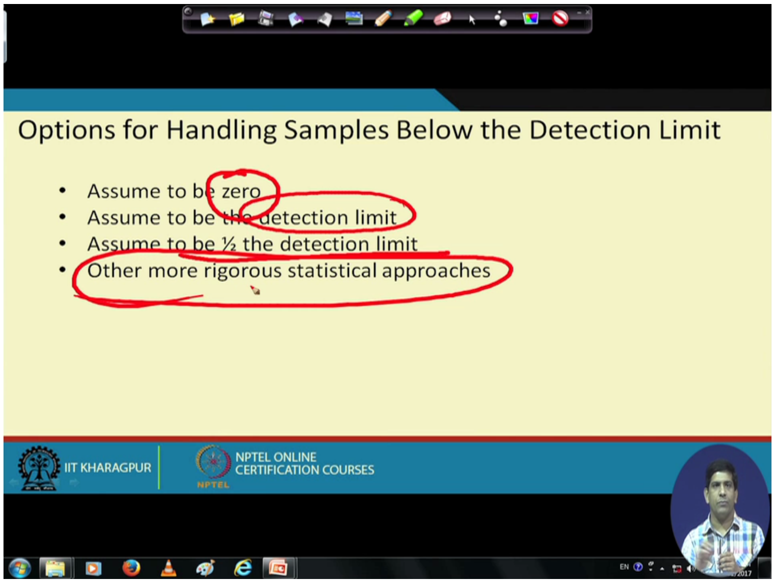
Data Below the Detection Limit

- In many cases, sample results will be below the analytical instrument's detection limit.
- If you have BDLs in your sample set, you must handle them appropriately with statistics. You can not just neglect them; they say something valuable about the data.

So with saying that just a small thing before we close up, so we talked about this detection limit earlier. So if we have the data below the detection limit, remember we talked about that you do the detection limit and so many cases the sample results will be below the detection limit because the numbers will come and they are many times reported as BDLs which is the below detection limit. So how you handle them in the statistics? So if you are doing this kind of statistics like UCL calculation, how you handle them?

They, because below detection limit also is some information. So we cannot put at a zero because that is not correct, we talked about that. We cannot like if we just ignore those dataset because when you do the statistics, statistics wants numbers. We cannot just write BDL, it is, in a text and that will not be taken in Excel spreadsheet when you are trying to do the statistical calculations. So we need to put some numbers there. So what numbers we should put? So there are different arguments for different people. And so there are essentially three things people do.

(Refer Slide Time: 37:20)



Options for Handling Samples Below the Detection Limit

- Assume to be zero
- Assume to be the detection limit
- Assume to be ½ the detection limit
- Other more rigorous statistical approaches

Sometimes, they assume it to be zero, they put the zero number there which is not really true and you would see rarely being used for that. And you do not, some will do it as a detection limit that again kind of goes little bit higher side because it may not be same as a detection limit if the value is below that. And many places, most common thing is to use as half the detection limit. So if the detection limit is 0.65, you use as 0.32 or 0.33. So that is half means some data may be higher, some data may be lower. So you are kind of in the middle.

So that is like a more like we can say more realistic kind of approach. So that is how it is handled. But there are different and otherwise there are more rigorous statistical approaches there to do that. So you can use those approaches depending on how important that aspect is.

So with that, we will kind of close this particular discussion and next, in the next module, we will start talking about, remember I told you that in this particular week, we will also try to talk about waste management rules. So in the next video, we will start talking about the new rule that has come up and how it is different from the previous rule and do some discussion associated with that. So I hope you are enjoying this course and thank you.