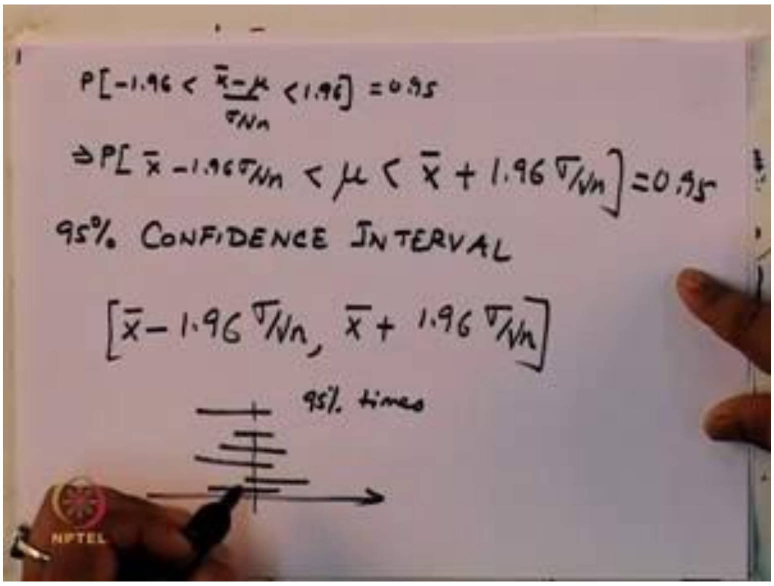
**Introduction to Biostatistics**
**Prof. Shamik Sen**
**Department of Bioscience and Bioengineering**
**Indian Institute of Technology, Bombay**

**Lecture - 30**
**Confidence intervals Part-II**

Hello and welcome to today's class. As a continuation from where we had left off in last class. We will start discussing about Confidence Intervals.

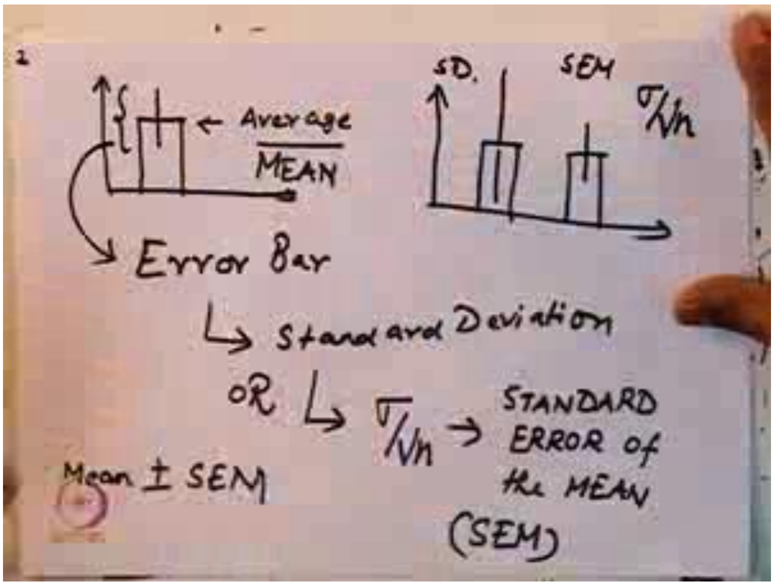
(Refer Slide Time: 00:36)




Just to remind you of what was covered in last class. We had shown that, given that we know that 95 percent of the data lies between plus minus 2 standard deviations. And given that for a sample mean for example, you know that x bar minus mu by sigma by root n is a confidence interval is the standard normal variable. So, you know that this probability is equal to 0.95. So, using some basic algebra you can show that mu lies between x bar minus 1.96 sigma by root n and x bar plus 1.96 sigma by root n; this probability is 95.

In other words the 95 percent confidence interval lies between x bar minus 1.96 sigma by root n and x bar plus 1.96 sigma by root n. So, this is your 95 percent confidence interval. In other words, as we had shown in discussed in last class if this is your true population mean then if you do this sampling then bulk of the times you will have your population mean lie between your range; spin also does not work, let us write down. So, 95 percent of times this interval will contain your; mean that is what the 95 percent confidence intervals mean.

(Refer Slide Time: 02:39)

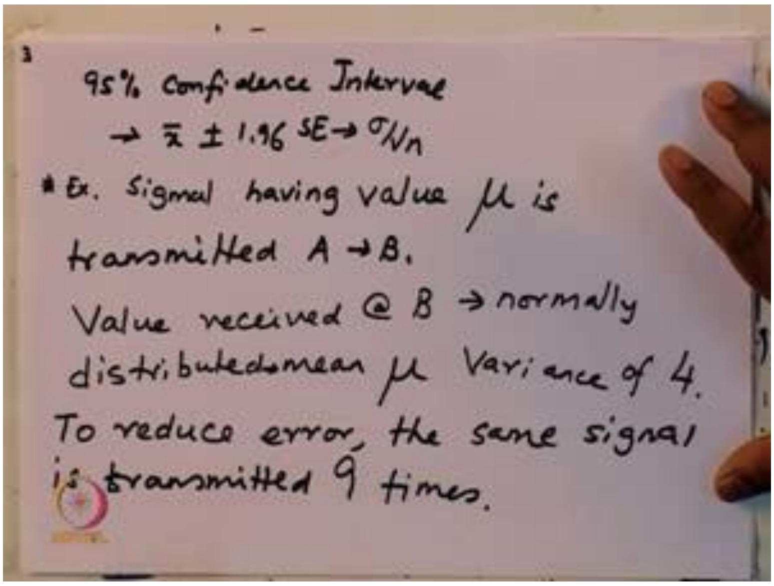



So, if you have seen experimental data been represented very often you will see data being plotted as; you have a bar and you have something which is ticking out on both sides. So, let us say hypothetically it is mid height of you know population of the class or so on and so forth. So, what does this mean, when data is plotted like this is your average or mean or mean value and this is called the Error Bar.

Now this error bar can be either your standard deviation or instead of standard deviation you can plot sigma by root n; this is also called the standard error of the mean. Or in short it is referred to as SEM. So, in papers or in research articles you will see data presented as mean plus minus SEM, this is called the standard error of the mean. So, if your sample size is small then your standard deviation is going to be significantly higher. So, in that way instead of plotting the standard deviation; so for the same data let us say if this is your mean let us say this is your standard deviation, this representation does not look nice.

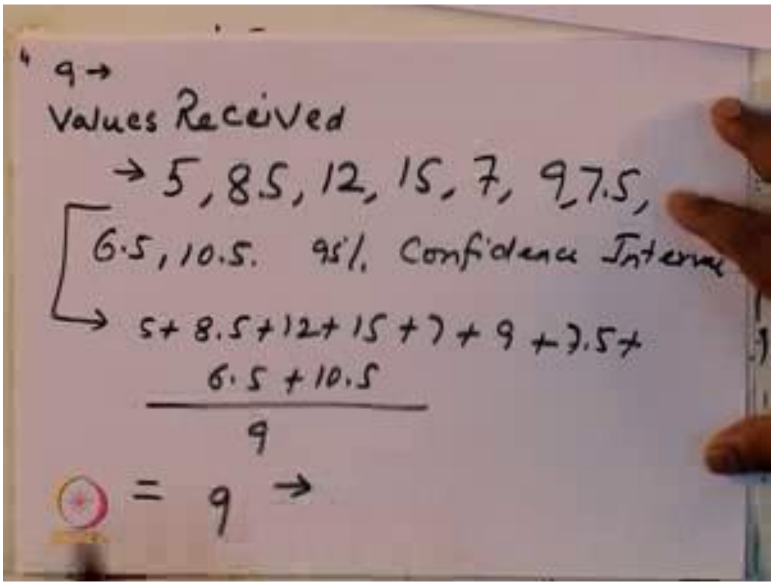
But if you plot the same data with the standard; so this is which standard deviation and this same data is plotted with standard error of the mean and you will see this is sigma (Refer Time: 04:39). And what is the reason for that because the standard error of the mean is sigma by root n. Even if you have n is equal to 4 then this is going to be a half its size; sigma by root 4, so this is going to be a half its size. So, that is why if you are have if n is reasonably large then standard deviation is also not so spread out it is less, but standard error of the mean is even lesser. So, it is sometimes it is helpful to present your data as mean plus minus SEM.

(Refer Slide Time: 05:26)



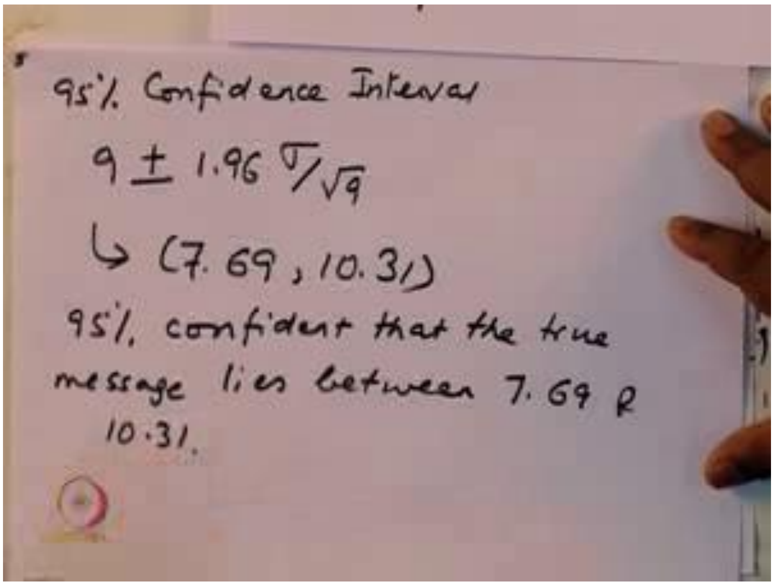

So, instead of having an interval estimate which gives the mode the minimum and the maximum you might have so; the 95 percent confidence interval gives us x bar plus minus 1.96 SE or 1.96. So, this is SE is sigma by root n. Let us take an example: imagine the signal having value mu is transmitted from A to B. And the value received at B is normally distributed with mean of mu and variance of 4. So, to reduce the error the same signal is sent or transmitted 9 times.

(Refer Slide Time: 07:16)



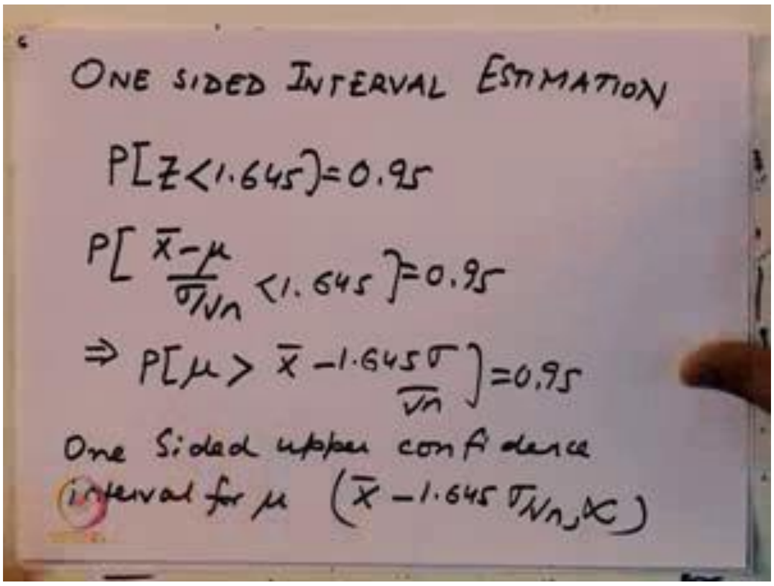

Let us say- the 9 times you have send it and what the signals that the values received as follows: it is 5, 8.5, 12, 15, 7, 9, 7.5, 6.5 and 10.5. We want to find out the 95 percent confidence interval. So what do we have? We have 95 confidence interval, so your average is mu. So, average of this is 5 plus 8.5 plus 12 plus 15 plus 7 plus 9 plus 7.5 plus 6.5 plus 10.5 by 3 4 5 6 7 by 9 and this should come out to be. So, you plug in the values you should come out to be 9.

(Refer Slide Time: 08:39)



So, this would mean my 95 percent confidence interval has to be 9 plus minus 1.96 into sigma by root of 9. So, if you plug in the values this comes out to be 7.69 to 10.31. So, this tells is that we have 95 percent confident that the true message lies between 7.69 and 10.31.

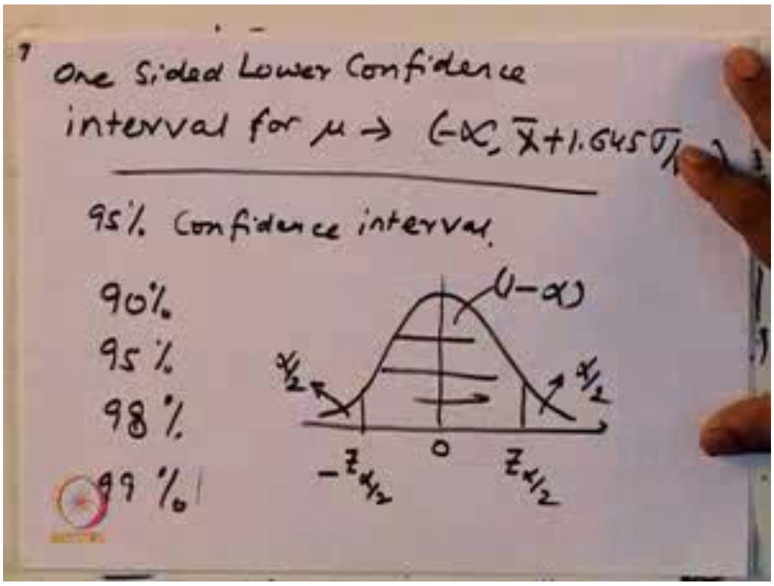
(Refer Slide Time: 09:54)




So instead of having a lower and an upper limit you might also be interested in finding out a lower limit. So, we can have what is called a One Sided Interval Estimation. So for doing that; for standard normal variable we know that Z less than one, so 1.645 is equal to 0.95. So, we can put x bar minus mu by sigma by root n less than 1.645 equal to 0.95 and this would give us probability of mu greater than x bar minus 1.645 sigma by root n equal to 0.95. Again the 95 percent confidence interval, so this is the one sided upper confidence interval for mu will be given by x bar minus 1.645 sigma by root n comma infinity, because this is the lowest value of possible was possible for mu.

Similarly, you can calculate an absolute upper value.

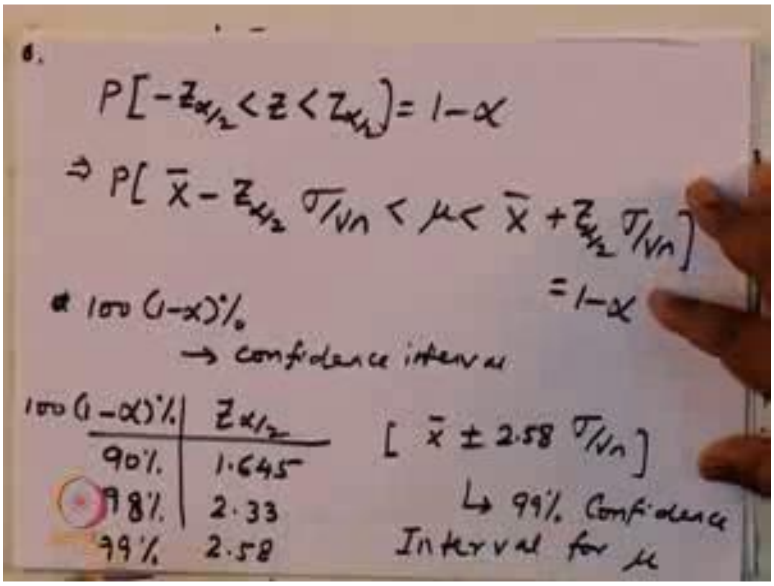
(Refer Slide Time: 11:21)




So, the one sided lower confidence interval for mu turns out to be between minus infinity to x bar plus. So, this is what you have the confidence intervals created for one sided. So, this set stipulates the lower bound for x bar of and this stipulates x bar 1 plus 6.5 sigma by (Refer Time: 12:07) n stipulates the upper bound for x bar or mu sorry mu. Now for all these cases we have assumed the 95 percent confidence interval, but theoretically we can calculate the confidence 50 percent confidence interval or 90 percent confidence interval so on so forth. And statistically what people typically use maybe the 90 percent, the 95 percent, the 98 percent or the 99 percent confidence intervals. So, how do we calculate the 90 percent on all these other confidence intervals?

So, what you need to note ok. So, let us say we have this interval this is 0, you have this normal distribution is symmetric; this value is minus Z alpha by 2 and this is Z alpha by 2. So, this corresponds to that the area under this curve is alpha by 2.The area under this curve is alpha by 2. So, what is in the center this is going to be 1 minus alpha. So, we can calculate.

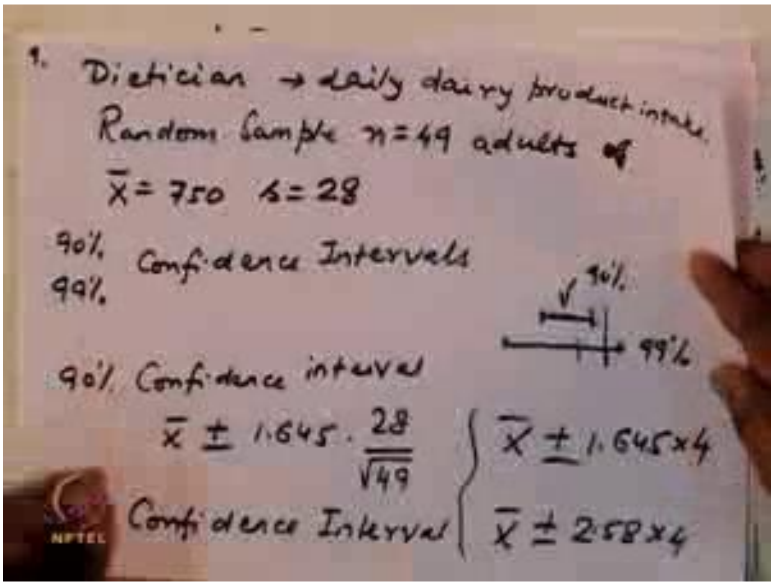
(Refer Slide Time: 13:42)




So, we can find out probability of minus Z alpha by 2 less than less than Z alpha by 2 equal to 1 minus alpha. And from this equation we can again write x bar minus Z alpha by 2 into sigma by root n less than mu less than x bar plus Z of alpha by 2 into sigma by root n is equal to 1 minus alpha. So, for 100 into minus 1 minus alpha is the confidence interval.

So, if I plot the values of alpha 100 into 1 minus alpha in percentage you can have 95 percent, you can have 90 percent, you can have 95 percent or you can have 99, percent 98 percent or 99 percent. So, the corresponding values of Z alpha by 2 for 90 percent it is 1.645, for 98 percent it is 2.33, and for 99 percent it is 2.58. So, this would mean just so if you want to calculate the 99 percent confidence interval then we have the bound; the upper and lower bound becomes x bar plus minus 2.58 sigma by root n. So, this is the 99 percent confidence interval for mu.

(Refer Slide Time: 15:48)

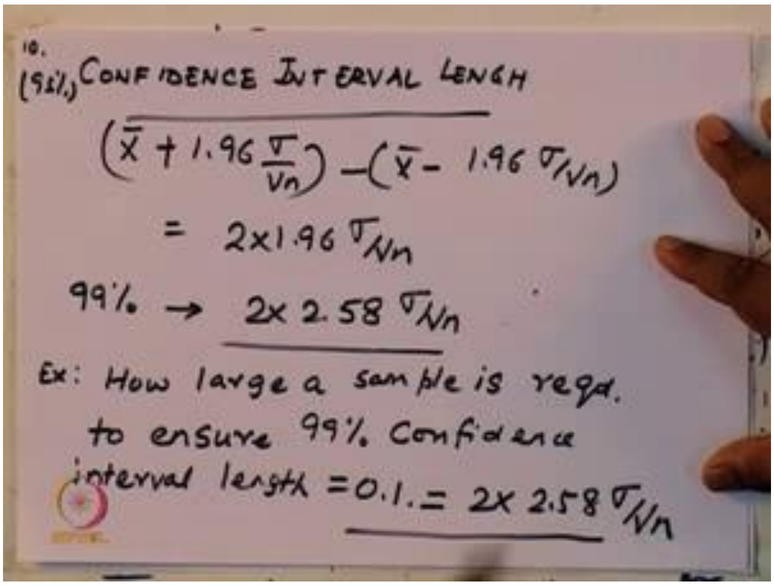



So let us take a sample example. Let us say a dietitian, selects a random sample of n equal to 49 adults. So, what the dietitian is probing is the daily dairy product intake. And what it finds is on an average x bar is equal to 750 with a standard deviation s equal to 28. So, we want to calculate the 90 percent and the 99 percent confidence intervals. So, as per this calculation for 90 percent; so the confidential interval for 90 percent confidence interval will be where x bar plus minus 1.645 into s which is 28 and n is 49 so root of 49. So, it is x bar plus minus 1.645. So, this is into 4. And the 99 percent confidence interval will be x bar plus minus 2.58 into 4.

So, what you can clearly see is this magnitude is a tighter limit. In other words if this is the interval length for 90 percent, the interval length for 99 percent is larger; the interval length is larger. So, this increases the chance that the population mean will lie within it. Versus, if you choose a tighter range it is likely it is possible there is a greater chance that your population will be lie outside this.

So, these were the population means. Once again would like to reiterate that for greater confidence interval range we always have a bigger interval and for smaller range; so for 90 percent you have a smaller interval. So, there is a greater chance of error that your population mean would not lie within this limit. So, that also brings us to another point.

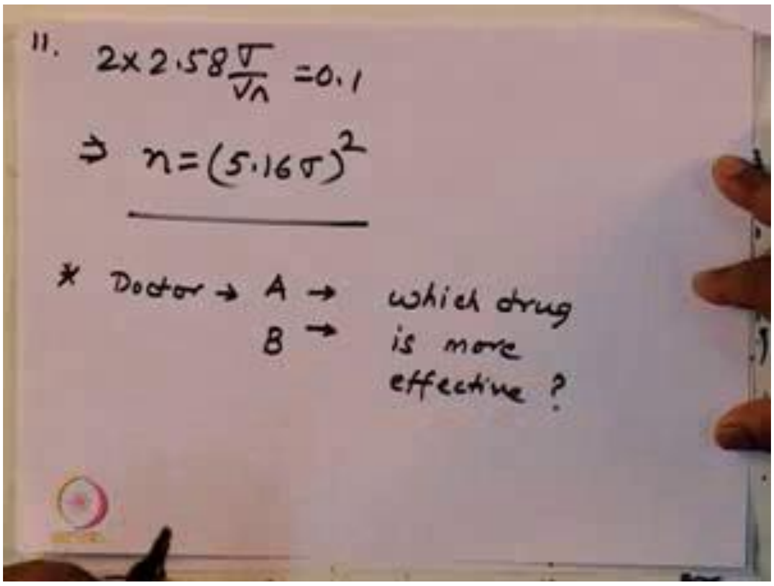
(Refer Slide Time: 19:07)




So, let us say if I want; so your total confidence interval length is what, it is basically x bar plus let us say if you want to 95 percent confidence interval length 1.96 sigma by root n minus x bar minus 1.96 into sigma by root n. This is 2 into 1.96 into sigma by root n. So, this is your confidence interval length.

Accordingly, for 99 percent confidence interval length this will be 2 into 2.58 into sigma by root n. Let us consider a simple example and the question is as follows; how large a sample is required to ensure 99 percent confidence interval length equal to 0.1. So, what we have to do is we have to set 0.1 is equal to 2 into 2.58 into sigma by root n. So, based on this you have an equation where you have root n in this equation.

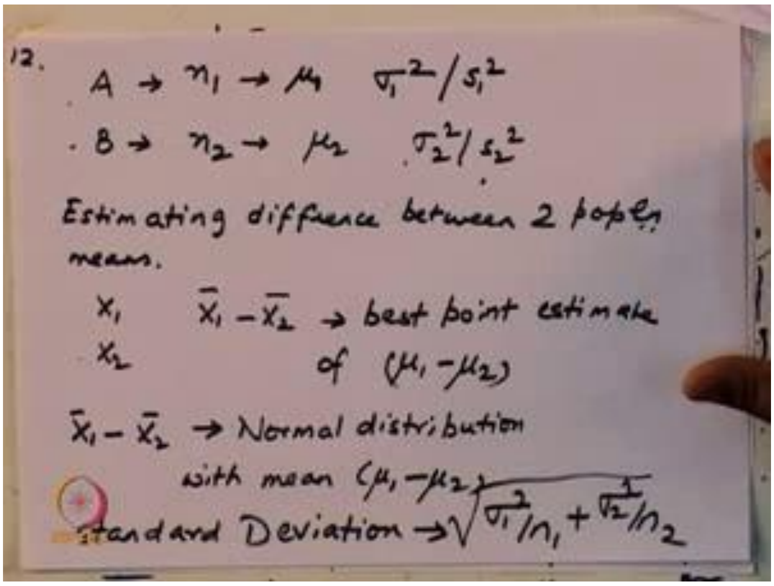
(Refer Slide Time: 20:59)




So, we can from this equation, we have 2 into 2.58 into sigma by root n is equal to 0.1 so from this I can calculate n is equal to 5.16 into sigma whole square. So, what you see? Your n has a sigma square dependence. So, if you want the 99 percent confidence interval to be a certain range then your n has to be much larger. So, till now we have discussed about calculating the confidence intervals for one particular (Refer Time: 40:41), but in experimental cases, in the general case what we are often interested is to compare the effect of two or two different situations to probe the effect of a particular molecule or some particular measurable metric.

In other words let us say you have a patient (Refer Time: 22:02) patient and the doctor is wants to understand the doctor is considering of administrating drug A or drug B and he wants to understand which of these two drugs will have a we much more effective for a given population.

(Refer Slide Time: 22:49).

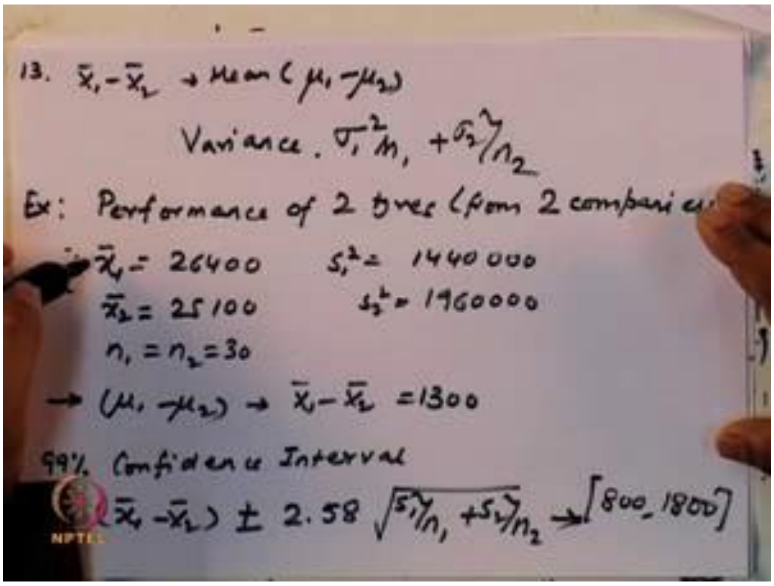



So, the question is which drug is more effective. So, how do we address this question? This brings us to the concept; so imagine you had this drug A you administrate it with n 1 number of in n 1 sample in the population. And from this you have measured the response in terms of a mean response mu 1 and a mean variance sigma 1 square. So, if it was the whole population this is sigma 1 square reviews for a sample it is s 1 square.

Similarly for drug B you had n 2, you sampled n 2 and you know the population means can be either mu 2 and sigma 2 square or s 2 square if it is for sample this is for a population. So, this brings us to the question of estimating difference between two population means. So, what we want to know that; if you have two random variables x 1 and x 2; this is corresponding to the drug response of A, this is corresponding to the drug response of B. So, we want to calculate x 1 bar minus x 2 bar, and this you would intuitively think that this is the best point estimate of mu 1 minus mu 2.

And I can say that this random variable x 1 minus x 2 this would follow a normal distribution with mean mu 1 minus mu 2, and what is the standard deviation? And the standard deviation which has to be sigma 1 square by n plus n 1 plus sigma 2 square by n 2; standard deviation this is a variance, so this is a standard deviation. So, the standard deviation will be this one.

(Refer Slide Time: 25:30)




Let us take a simple case: once again your x 1 bar minus x 2 bar, we have a mean of mu 1 minus mu 2, and variance which is going to be sigma square by n 1 plus sigma 2 square by n 2. For example take a simple case: let us say you are comparing the f you know; you are comparing the performance of 2 tyres from 2 companies. And you have been given x 1 bar as 26400, x 2 bar and both n 1; so x 2 bar as 25100 and n 1 equal to n 2 equal to 30, s 1 square as 14,40,000 and s 2 square as 19,60,000.

So, what you will do? You will calculate mu 1 minus mu 2, point estimate of mu 1 minus mu 2 is x 1 bar minus x 2 bar and this is 1300. And you can also calculate; so for this you can calculate this quantity square root of this as a standard deviation. And you can calculate the 99 percent confidence interval which is given by x 1 bar minus x 2 bar plus minus 2.58 times root of s 1 square by n 1 plus s 2 square by n 2. And if you plug in the values and you see this is always positive then this would mean that this tyre from company 1 is going to be better than tyre from company 1.

So, let us say- I do not know the exact value for this. Let us say it comes out to be between 800 and 1800. So, if this was the range then you know this is always positive. So, you can comment that with 95 percent confidence you can state that tyre A is better than tyre 2. So, this is how you can make use of confidence intervals to either get an estimate of the population mean or even compare different populations to say which is better which is worse.

With that, I thank you for your attention and I look forward to next discuss.