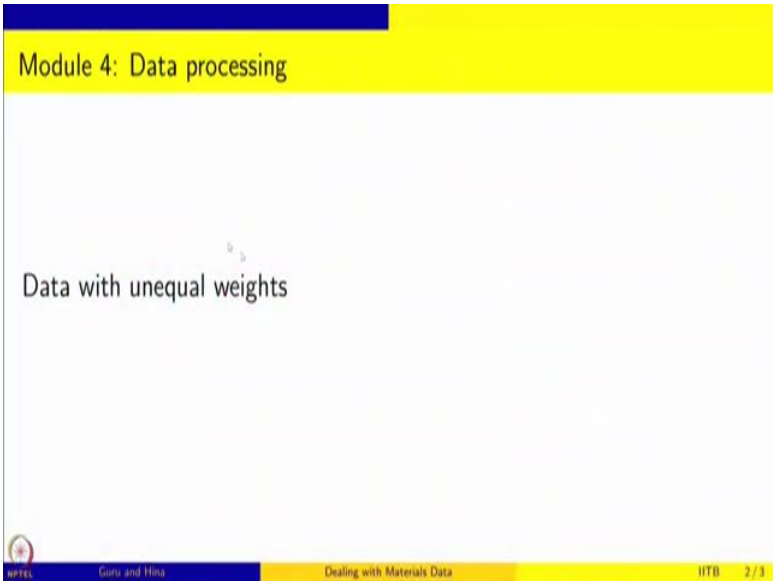
Dealing with Materials Data
Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 63
Data with Unequal Weights

Welcome to Dealing with Materials Data, we are looking at the collection, analysis and interpretation of data from Material Science and Engineering. We are in the module on Data Processing.

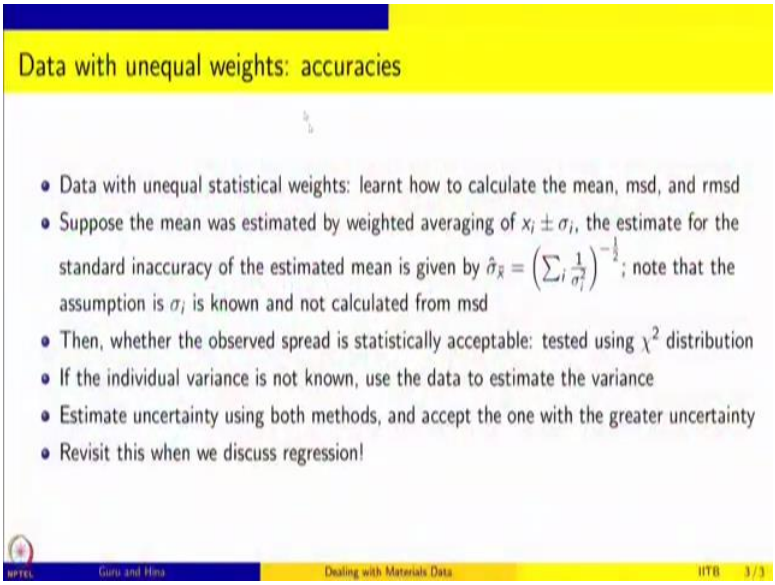
(Refer Slide Time: 00:26)



We have looked at data with equal statistical weights and how to do the analysis for mean and variance and to give confidence intervals and to tell what is the probability that the true mean will lie in some range and things like that. In all that we also assumed that it was normal distribution and we said the measurements are independent, then we gave the process of doing this.

But we also know that there are sometimes data with unequal weights we have seen one example in this module. And how do we deal with such data? How do we give accuracies for data with unequal weights.

(Refer Slide Time: 01:05)



So, that is what we want to discuss in this session. We know how to calculate mean and MSD and RMSD for data with unequal statistical weight. And suppose the mean was estimated by weighted averaging of $x_i \pm \sigma_i$, the estimate for the standard inaccuracy of the estimated mean right. We have calculated the mean by looking at some numbers, but they had this kind of standard deviation.

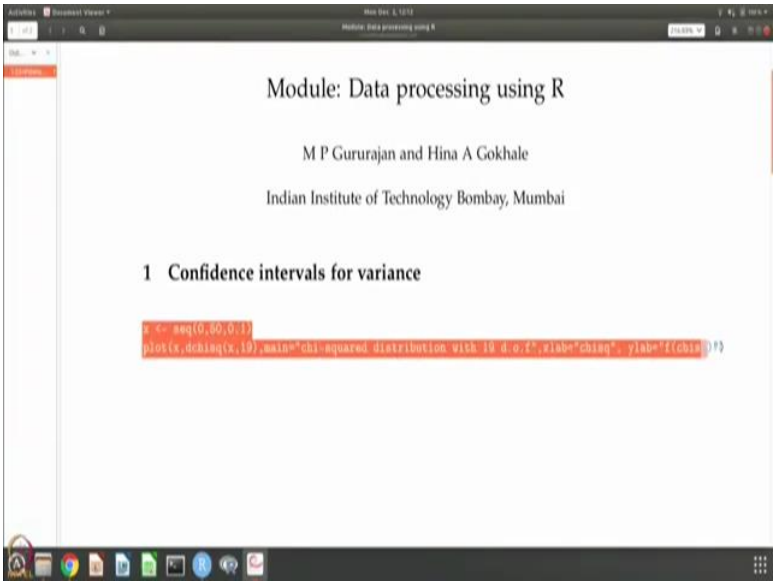
$$\widehat{\sigma_x} = (\sum_i \frac{1}{\sigma_i^2})^{-\frac{1}{2}}$$

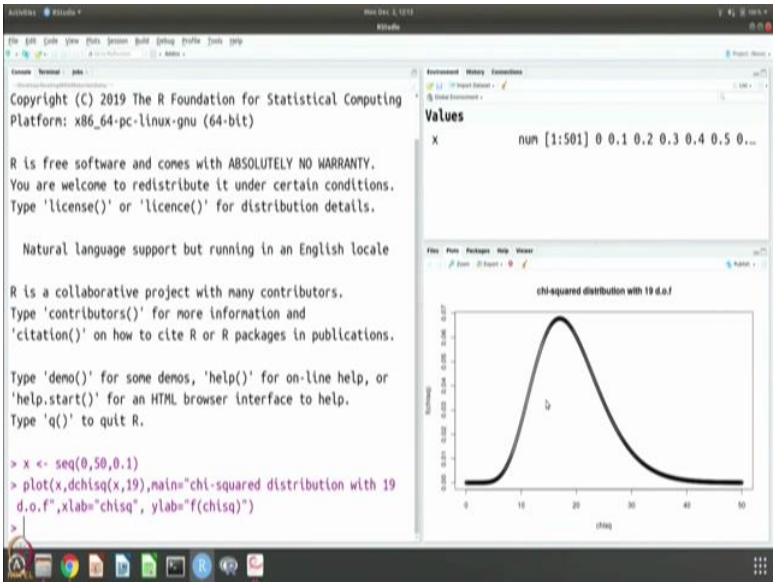
Because we know that, because it is a sum of square of random variables the observed spread will follow chi-squared distribution. So, should follow chi-squared distribution if it truly follows. So, we can compare the data spread with the chi-squared distribution and decide the weather what we are seeing is acceptable or not. So, that is how we do this test. If the individual variance is not known, of course, you can use the data to estimate the variance.

And so, there are two ways of estimating the uncertainty. Which one should you accept, you should actually do both and accept the one with greater uncertainty, just to be on the safer side. So, we will have more to say on this when we discuss regression. So the analysis of variance and its errors, etc so we will do a little bit more in detail in regression.

For this session I am just going to show how the chi squared distribution looks. So, we will plot that and that will be the end of this session.

(Refer Slide Time: 03:29)





So, we know chi-squared distribution is chisq is the command. So, let us plot it. So, this is the sequence of 0 to 50 and we want to plot the chi square distribution function and it has 19 degrees of freedom, because remember we are dealing with conductivity data which had some 20 data points.

So, we have been dealing with a degree of freedom of 19 for the t distribution. So far the same 19 degrees of freedom, I am plotting here the chi squared distribution and this is how the chi squared distribution looks. Okay, so we will come back to chi-square distribution and analysis the variance a little bit in detail when we do regression and analysis of variance. Thank you.