Dealing with Materials Data: Collection, Analysis and Interpretation

Professor. M P Gururajan
Professor. Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

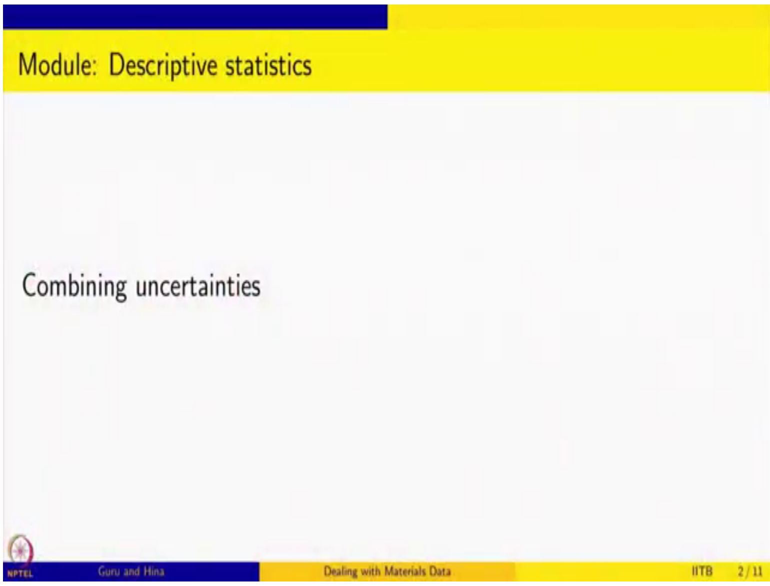
Lecture 29

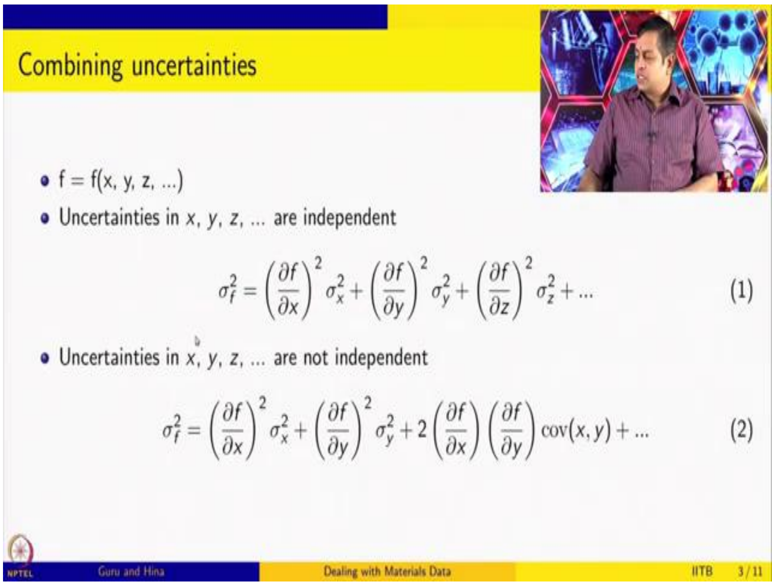
Combining Uncertainties

Welcome, to Dealing with Materials Data. In this, course we are going to learn about the Collection, Analysis and Interpretation of data form Material Science and Engineering. So, we have been looking at error and its propagation. In, the last session we looked at how error propagates, and we just looked at how the error in one quantity namely conductivity propagates when you do the conductivity measurements using Eddy current method.

In, terms of the thickness of the skin. So, we did this calculation in the previous session but, we want to deal with slightly more complicated propagation of error and that is what we are going to do in this session.

(Refer Slide Time: 01:05)

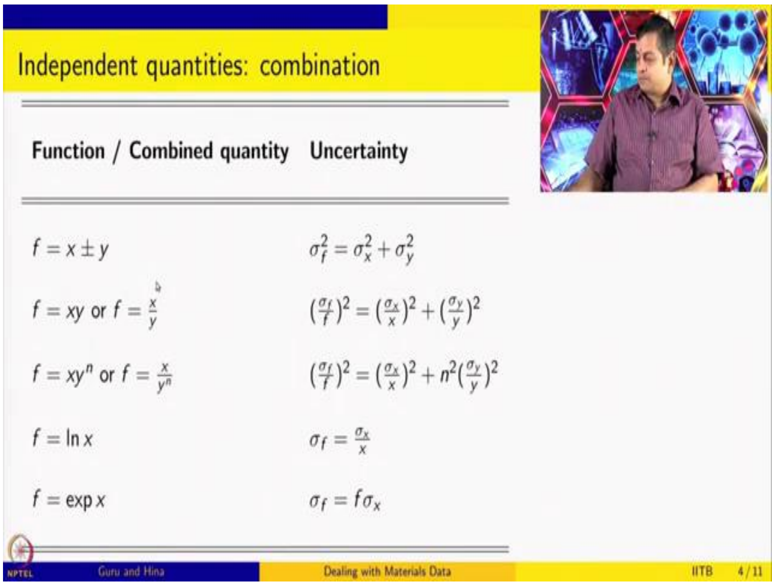




So, this is about combining the uncertainties and, how do we combined uncertainties, if, suppose we have a quantity f and this quantity f, depends on these variables x, y, z, etc and if, we assume that the uncertainties in x, y, z, etc, are independent then the variance in the quantity f is given by this formula where you take the partial derivative of, f with respect to x square it multiple by the variance and again take the partial derivative with respect to y and multiple by the variance and so on.

$$\sigma_f^2 = (\frac{\partial y}{\partial x})^2 \sigma_x^2 + (\frac{\partial y}{\partial y})^2 \sigma_y^2 + 2\left(\frac{\partial f}{\partial x}\right)\left(\frac{\partial f}{\partial y}\right) cov(x, y) + \cdots$$

(Refer Slide Time: 02:21)



$$f = x \pm y$$

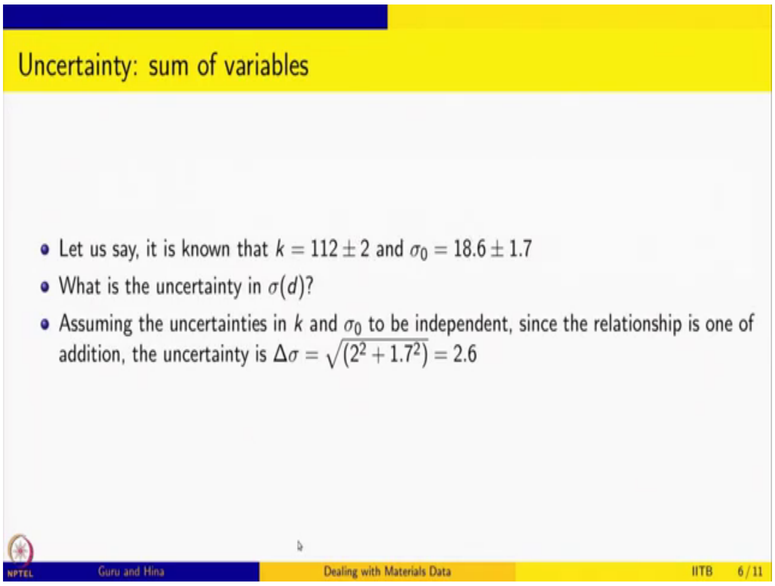
$$f = xy \; or \; \frac{x}{y}$$

$$f = xy^n or \; \frac{x}{y^n}$$

$$f = lnx$$

So, these are the formula that one can use but in all this we can assuming that the uncertainties in x and y is independent if it is not, so you have to also consider the co-variances.

(Refer Slide Time: 03:29)



$$\sigma(d) = \sigma_0 + kd^{-1/2}$$

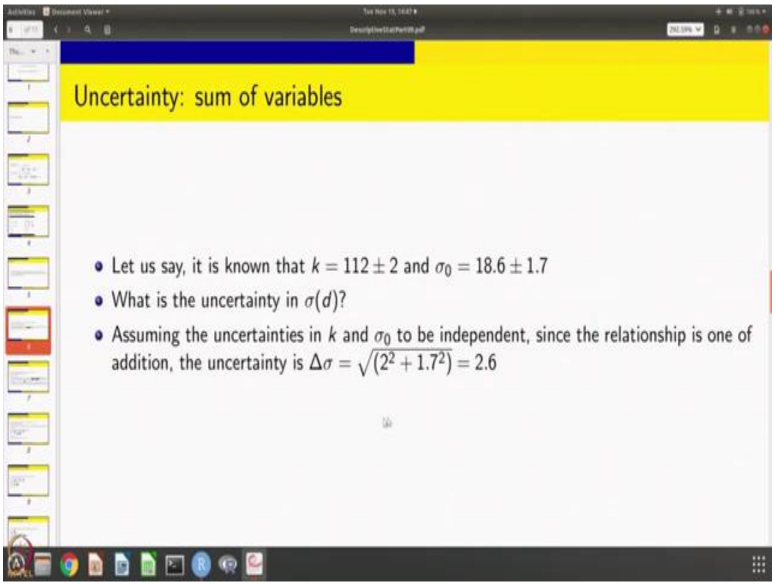
So, this constant that you will get if you fit to this function form is known as Hall-Petch co-efficient and σ_0 is the materials constant for initiating Dislocation Motion. Now, as you can see there could be uncertainty in many of these quantity. For, example that could be uncertainty that k that you have evaluated, there could be uncertainty in the σ_0 that you have value not able, obviously we also know that the grain size is not a single number.

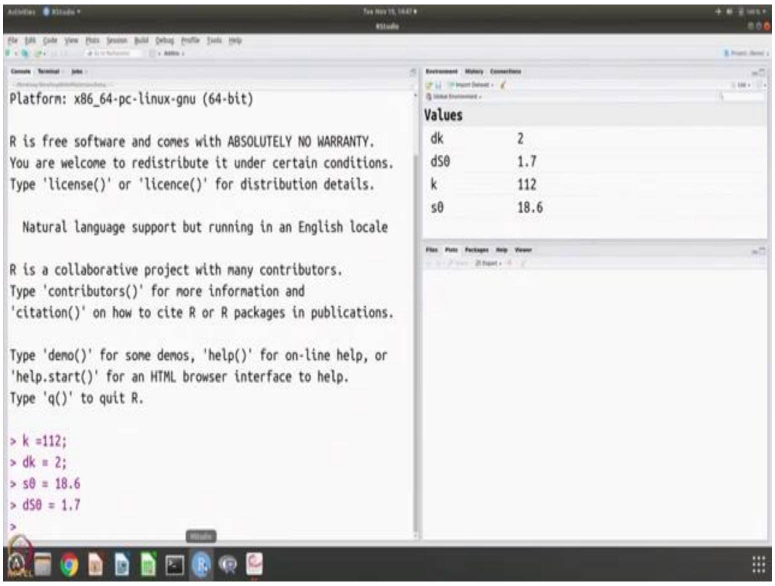
So, it has a uncertainty in addition we also know that it also has different distributions, I mean when we say, uncertainty in k and σ_0 for example, we are assuming that it is random noise. So, the distribution of the error is normal distribution. But, that need not be so like we have seen in one of the earlier sessions, that the grain size distribution can follow non-normal distribution and, in the case we saw it first beta.

So, in such cases how do we deal with and typically sometimes you will also see the grain size actually follows lognormal of distribution. And if so, how do we get the error that you get in the flow stress. This, session is about propagation of error. We have, error in these quantities, if we know how much the uncertainty can we, say anything about the uncertainty in this quantity. That is, what we are trying to calculate and let us do the first one.

Let us, say that the k value is this is for copper. The k value is 112 plus or minus 2 and this is calculated assuming that you are grain sizes are given in microns and sigma naught is 18.6 plus or minus 1.7. So, this is sigma for k and this is the for σ_0, and so we want to find out the uncertainty in the flow stress. We are going to, assume that these two uncertainties are independent and so it is one of addition. So, we can do this in R.

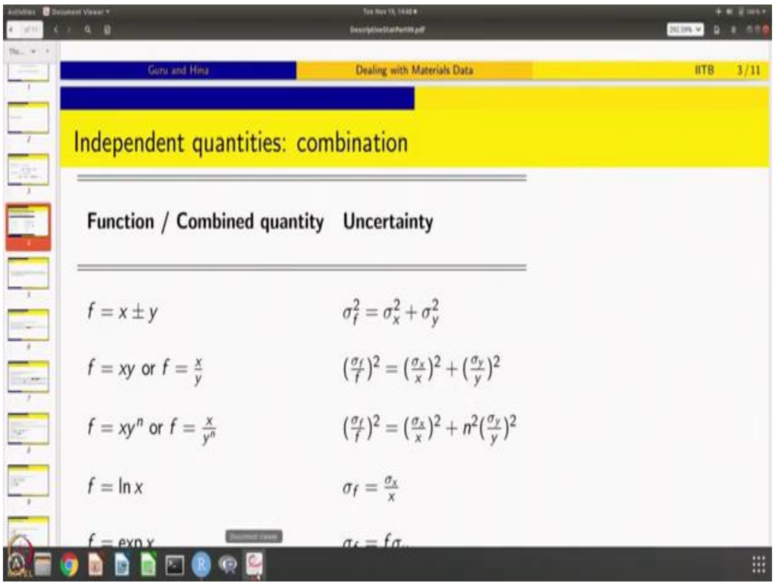
(Refer Slide Time: 06:02)

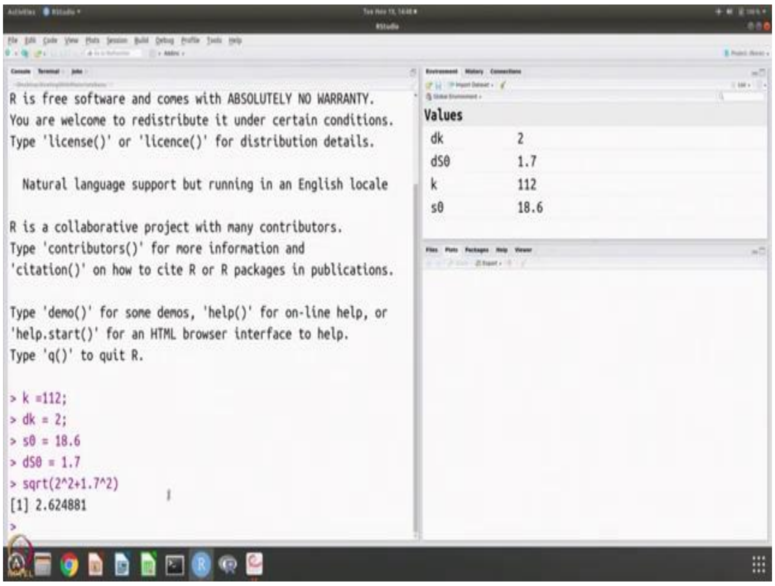




So, we want to say k is 112 and uncertainty in k is let us get this and s_0 is 18.6 and ds naught is 1.7. So, now we want to calculate the uncertainty in sigma and because, these quantities are in addition all you need to do is to look at the corresponding formula.

(Refer Slide Time: 06:59)



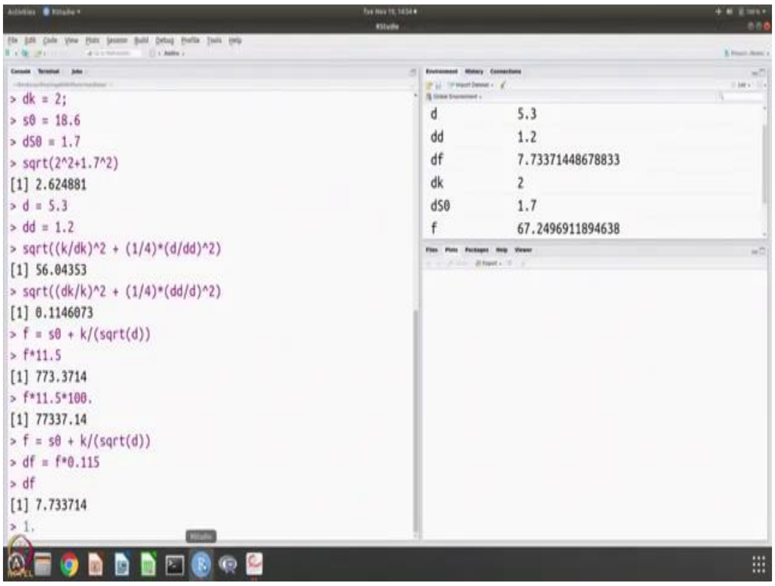


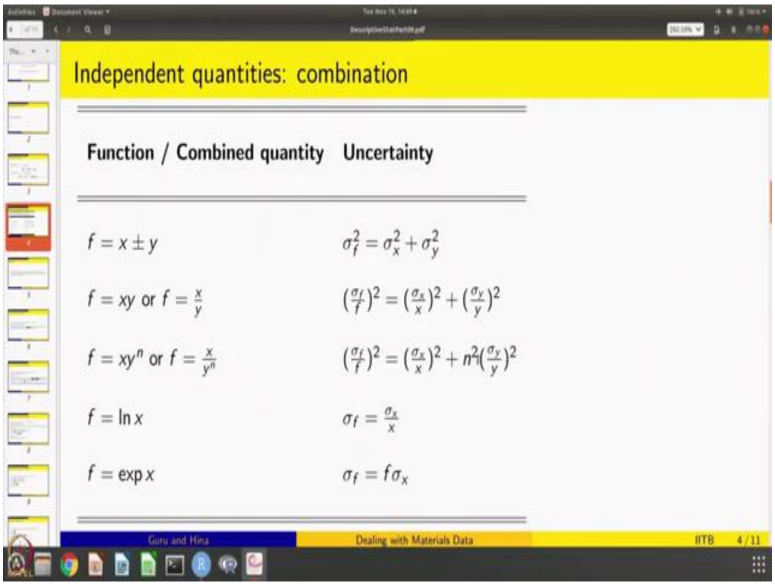
$$f = x \pm y$$

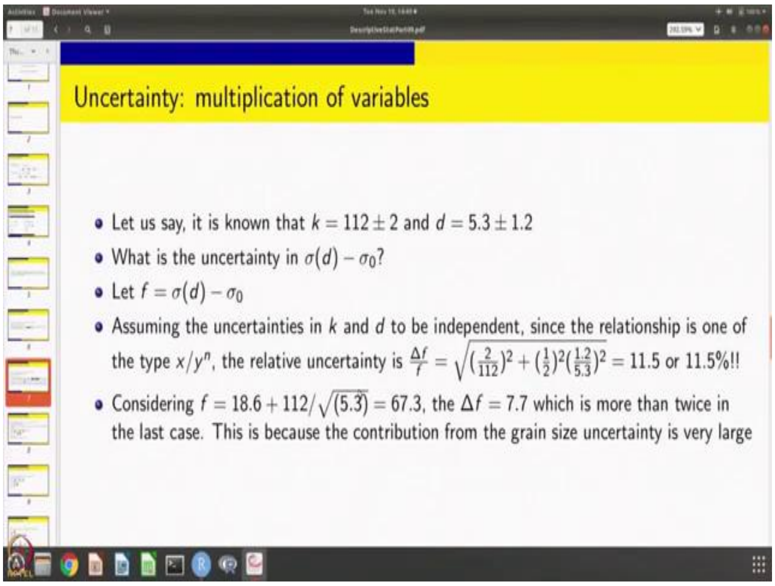
$$\sigma_f^2 = \sigma_x^2 + \sigma_y^2$$

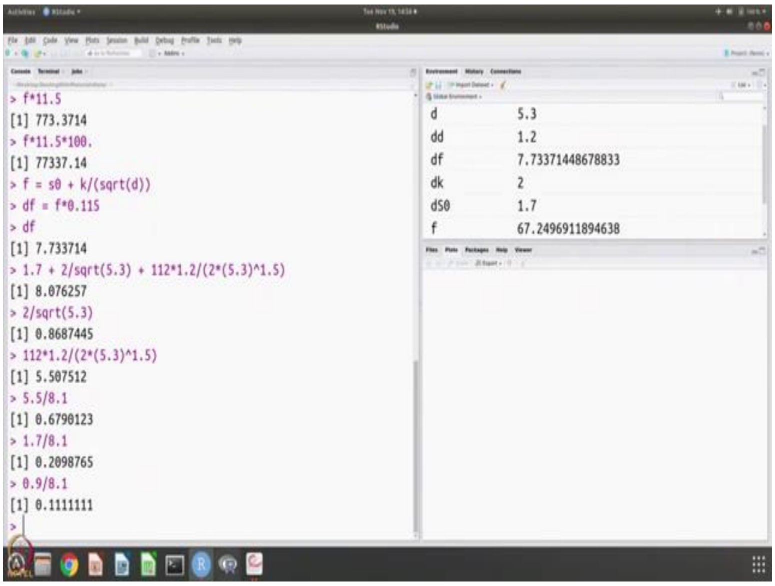
So, the uncertainty just gets added up. So, you will get square root of 2 square plus 1.7 square. So, that is of the order of 2.6. So, that is what we are finding here.

(Refer Slide Time: 07:40)









Now, what happens if suppose we have uncertainty in k and d let us, consider a quantity f which is the flow stress for a given grain size minus sigma 0. So, let us consider that quantity and let us look at the uncertainty. In this, case this is of the form x by y to the power n. So, and let us say the d is 5.3 and the uncertainty in d is 1.2. So, in this case we have found that the uncertainty has to be calculated using this formula. So, the relative uncertainty is square root of sigma x by x whole square, plus n square sigma y by y, whole square.

And sigma x happens to be in this case k and y happens to be d. So, it is k by dk whole square plus n is half. So, it is 1 by 4 multiplied by d by dd whole squared and the entire thing you have to take the square root. So, this is the quantity we have. There seems to be some problem. So, it is the other way around. It is dk by k, dt by d whole square. So, it is about 11.4 percent and this quantity is nothing but the delta f by f.

So, if you multiple this quantity by f and you can evaluate f for a given set of parameters. Then, this delta if happens to be about 7.7. So, you can calculate f. Let us, say that we know that it is s0 plus k, divided by square root of d. So, that is f. So, if you multiple f by 11.5 then you get 11.5 percent. So, 115 that is a, so it is 7.7 percent that is what is shown here. So, you get the error to be about 7.7.

And off course you can now consider the error in all the 3 quantities sigma naught, k and d. And can ask the question, what is the uncertainty in sigma, you can do the calculations in series, you can calculate first the uncertainty that is coming from here, and how does that uncertainty then add to the uncertainty that is coming from here, and then you can get the total uncertainty in sigma naught.

But, because the error that you are going to get from here is relative and the other one is just the sigma that you are going to get, it is going to become much more complicated. But, the easier to follow is the partial derivative formula. We have, seen that the error for example, R uncertainty in the flow stress should be doh sigma by doh sigma naught and the uncertainty in sigma naught, multiply plus doh sigma by doh k, into this should be uncertainty in k del k plus doh sigma doh d into del d.

So, if you do that then you know that in this case. For example, it is sigma 0, so this derivatives just gives the sigma naught itself and in this case it will give you the root d remind. So, k just gives you del k because 1 by root d del k, is what it be tells, and in this case k del d will remind and d will be give you minus half d to the power minus 3 by 2.

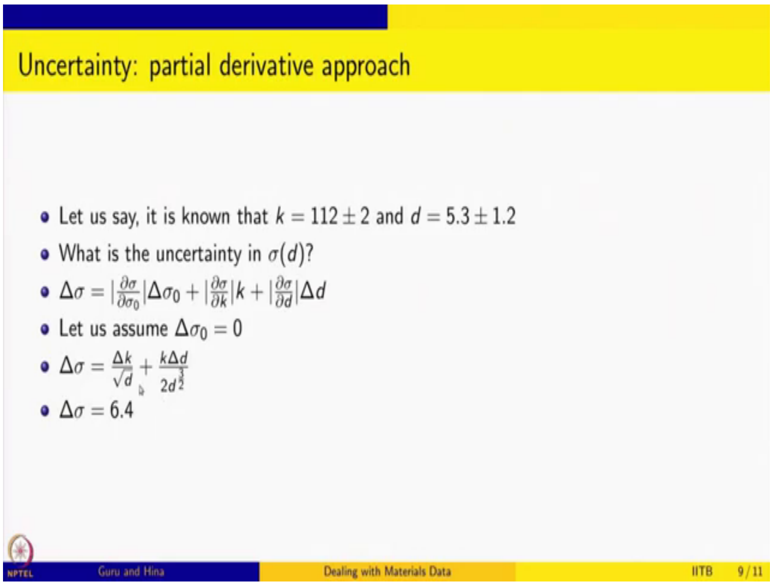
So, that is comes down, so 2 d to the power 3 by 2. Because, we have taken mod. So, you can now add all these uncertainties we know that this is 1.7 for example, and we know that this is 2 divided by square root of 5.3 and this is for example 112 multiplied by 1.2 divided by 2, 5.3 to the power 3 by 2. So, we can evaluate this quantity so let us do that and find out how much is the error.

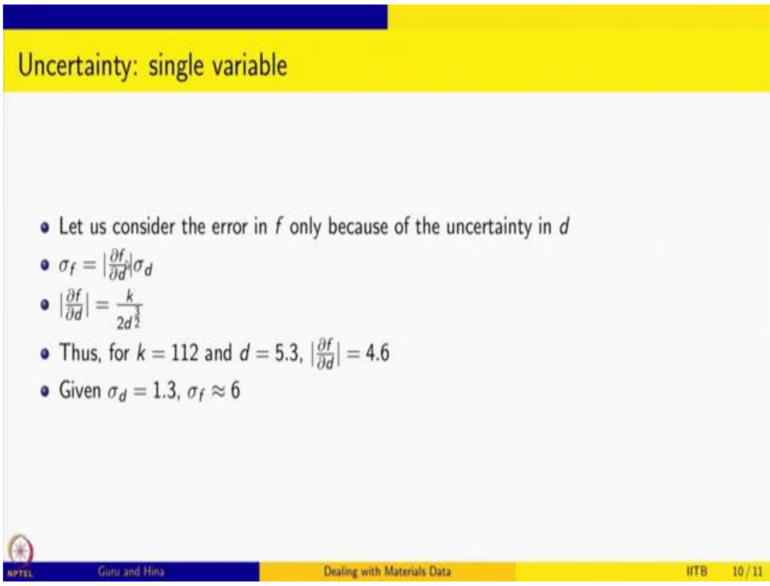
So, we want to do the so 1.7 plus the uncertainty in k that is 2 divided by square root of 5.3, 2 divided by square root of 5.3 and then we have k which is 112, and multiplied by the uncertainty in d divided by 2 into 5.3 to the power 1.5. So, you get about 8 and now you also know the relative values. For example, of this 8.1, 1.7 comes from sigma 0 and 2 by square root of 5.3 that is some about 0.8 comes from this uncertainty which is in k and remaining. So, if you have about 1.7 plus 0.8 so about 2.5.

So, the remaining 5.5 comes from the other quantity. So, we can calculate this quantity and this is what the 5.5 comes to. So, you can also know that the uncertainty is in d or giving you the contribution. So, it is 5.5 divided by some 8.1. So, it is about 67, 68 percent contribution is coming from here and then. So, 1.7 divided by 8.1, so that is another 21. So, it is about 68 and 21 so 89. So, remaining 11 is what is coming from this.

So, that is 0.9 divided by 8.1 so that is 11 percent. So, you can see the relative contribution to the error also by doing this. So, what we are doing in this case is to use the formula and calculate the uncertainty.

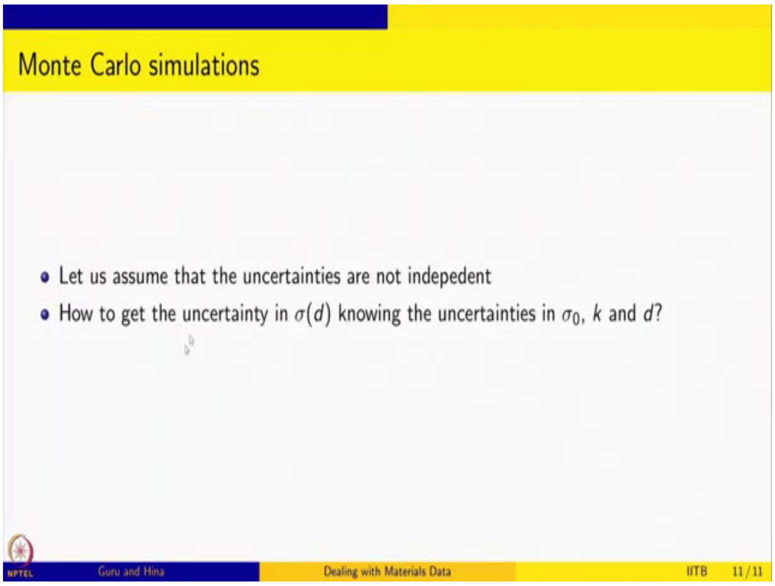
(Refer Slide Time: 15:57)





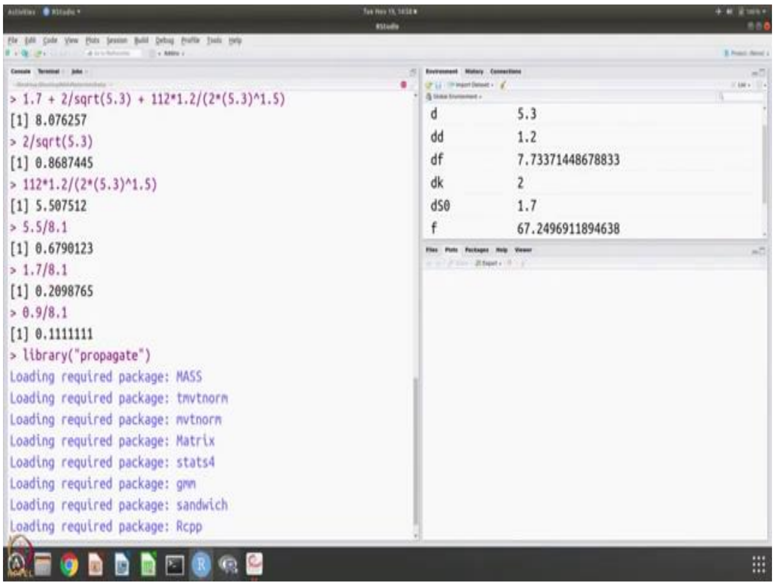
Of course, so we can now combine different things you can assume one of the uncertainties to be 0 and you can find out what is the error that is coming. So, we know that it is 5.5 and this is about 11 about 0.8. So, that is 6.4 and so you can calculate and if you assume that only uncertainty is coming from del d then you can calculate and so on, so forth.

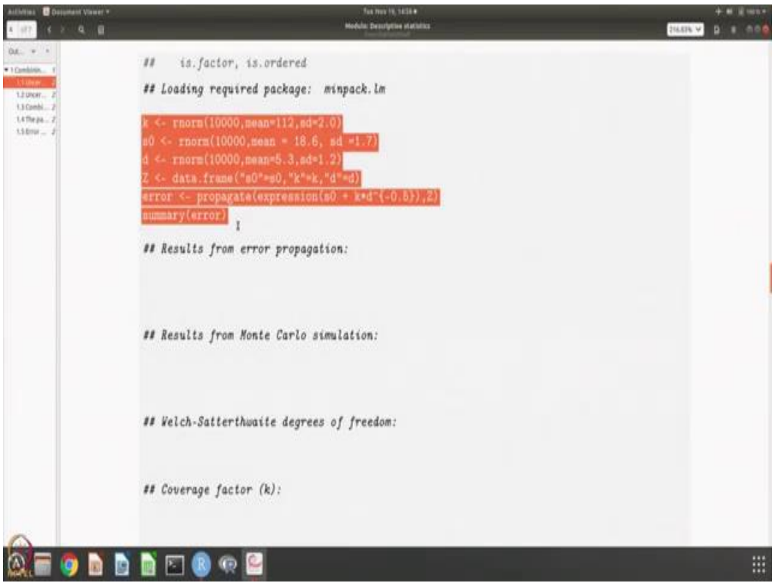
(Refer Slide Time: 16:27)

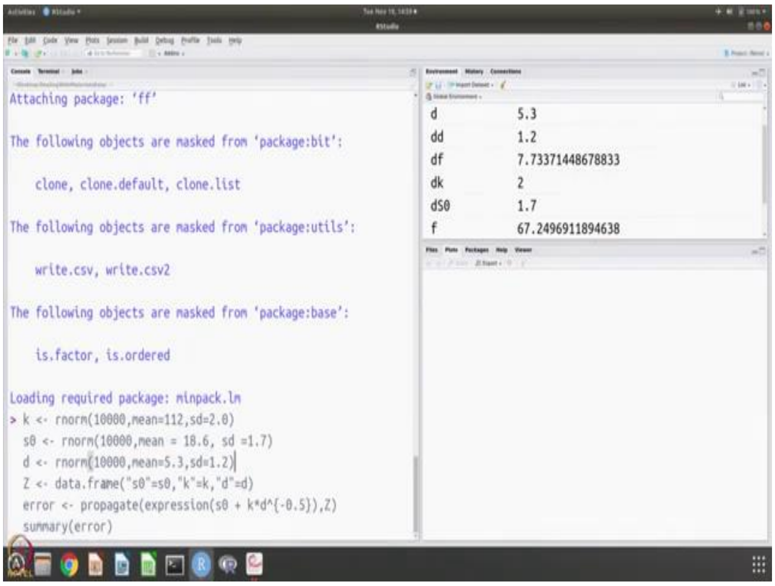


Of course you can also, assume that the uncertainties are not independent and then if you know the uncertainty in sigma 0 k and d how do we get the uncertainty in sigma and in that case you need to know the co-variance that is contributing and for that we can do the Monte Carlo simulations. So, that is what I will show next.

(Refer Slide Time: 17:13)





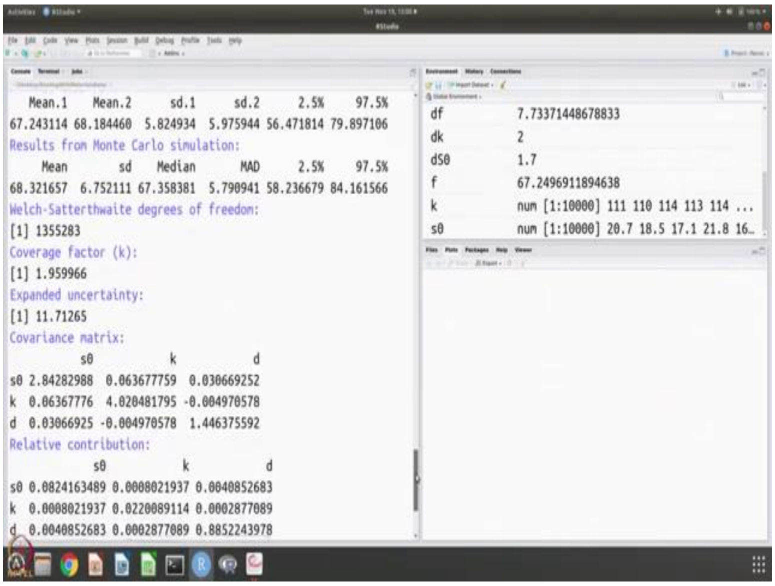


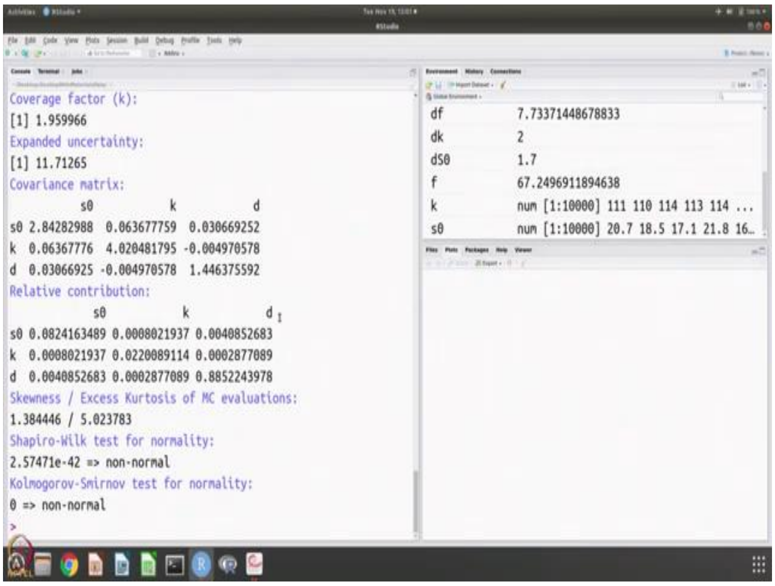
Let us consider the and for Monte Carlo simulations if we are going to use the library propagate. We have the library propagate and then we are going to let us, consider these. So, what is this we are going to assume that the error in k is a normal distribution and the mean is 112 an standard deviation is 2 that we know, and in sigma naught again we are going to take a mean of 18.6 and standard deviation of 1.7, and again we are assuming that is normally distributed.

For starters, we are also going to assumes that the grain size is also normally distributed it need not be, we will do lognormal, for example, as one more case and but, for starters let us assume that even, d is normally distributed. So, the other calculations that we have done so far we were, assuming implicitly that this is also an error and so that has mean 5.3 and standard deviation 1.2.

So, we take all these random variants that we have generated for these quantity and then we are going to use a Monte Carlo simulations using them and this is the expression s 0 plus k by d to the power minus 0.5 and we are then going to find out how much is the error that you get in resultant quantity because of these variations.

(Refer Slide Time: 18:40)



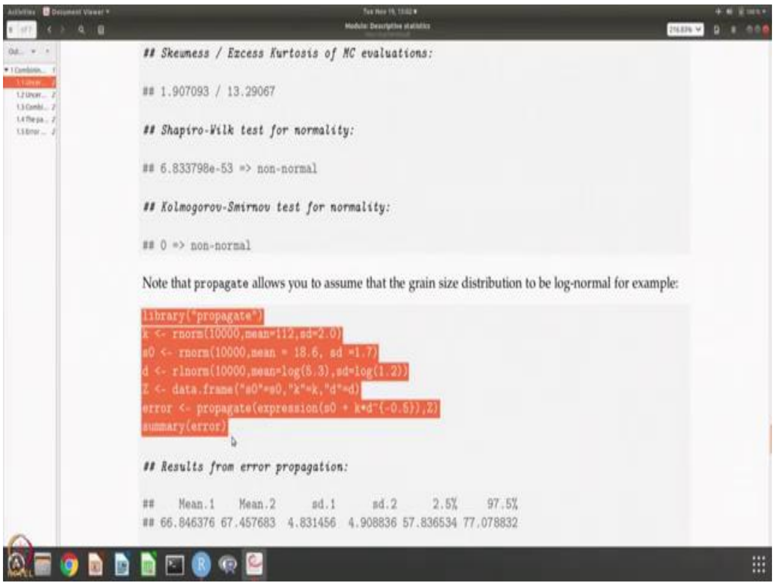


So, here lot of information that propagate gives. So, it gives you the means and the standard deviations so which is not very different from what we have found and from Monte Carlo simulation it gives you the means and standard deviations. So, that is also a not very different from what you see and what is this degrees of freedom coverage factor etc, we will come back uncertainty we have already found that it is of the order of 11 we have found and the co-variance matrix is given.

So, it tells you relative importance of these of diagonal terms. So, you can see that k sigma 0 and d sigma 0 these are the co-variance values and k,d for example. So, they are all relatively small so compare to these quantity they are not very big and you can also see the relative contribution off course this we have seen it is from these simulations you find that 88 percent comes from here, about 8 percent comes from sigma naught and k contributes about 2 percent.

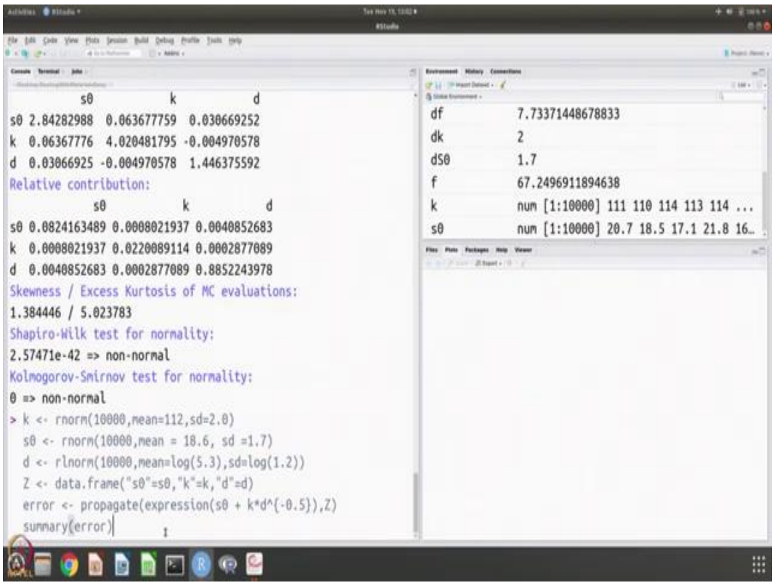
So, in terms of relative contribution again we see that maximum contribution comes from d, the next one comes from s naught sigma naught and the third is form the k parameter and the skewness and excess kurtosis is given here and some of these test for normality. These are also things that have not discuss yet like Kolmogorov-Smirnov we will come back to this at a later point but using such simulation again you can get the error propagation.

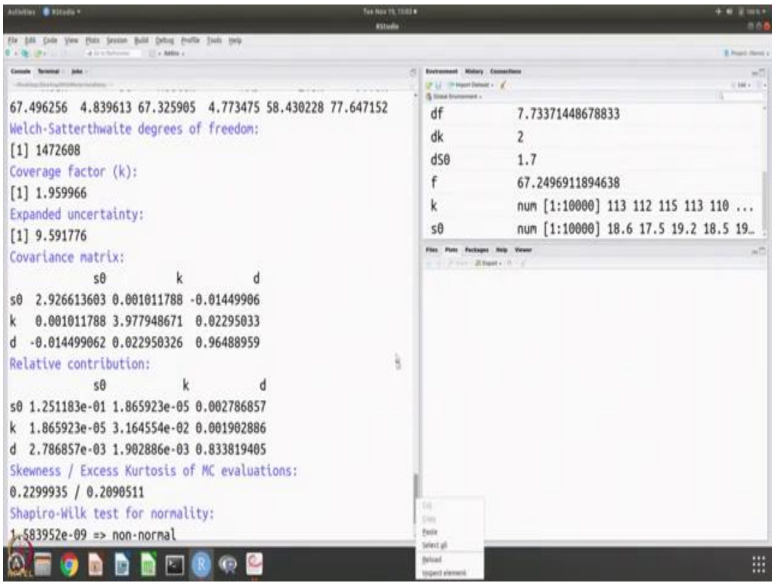
(Refer Slide Time: 20:44)



So, this is what the information that we have gotten. Now, you can do one more thing you can make sure that the de-distribution is not normal but log normal. Suppose that is the case then what happens to your error.

(Refer Slide Time: 21:09)



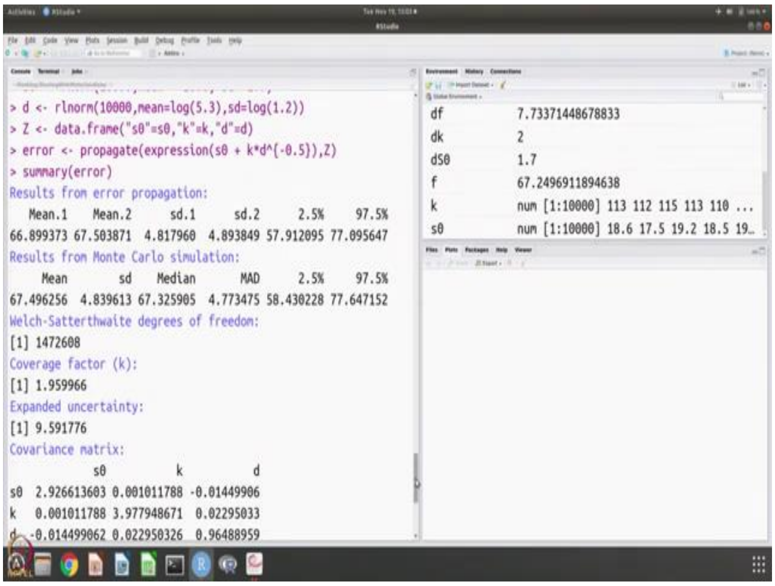


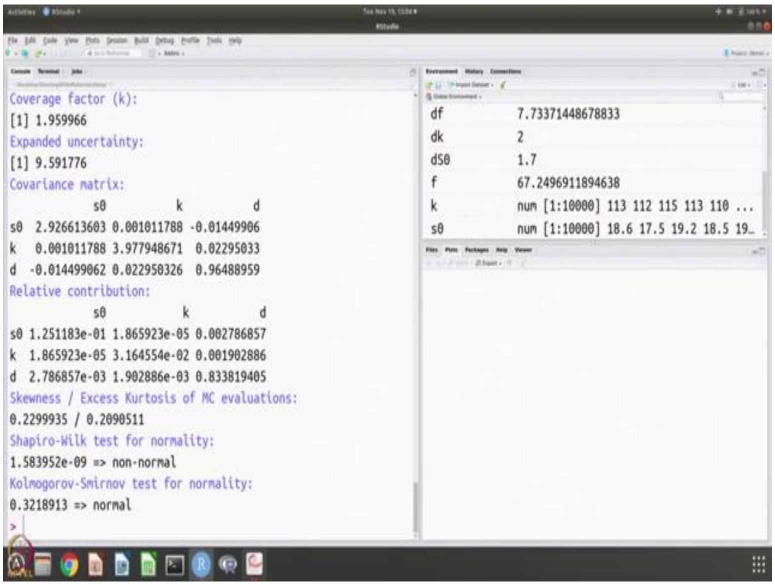
So, you can do that again the simulations are very helpful because, then you can assume any distribution for the quantities and you can generate the variants and then use that in the simulations. So, the other quantities are the same k and s naught I am assuming that they are normally distributed with the given mean values and the standard deviations.

But the grain size I am going to assume log normal distribution with a mean at log 5.3, and standard deviation of log 1.2 and using that as the parameters then, we are going to do Monte Carlo simulation run and we are going to get the information.

Again, we see that majority of the contribution comes from d about 83 percent, and this is about 12 percent and this is about 0.031. So, it is about 3 percent. So, and the co-variance matrix again, you can see that they are not contributing much most of the contribution is coming from the variances. So, you can assume that they are independent that is a fairly good approximation.

(Refer Slide Time: 22:30)





And but, you can see that the values that you get for example, mean is not changing much the standard deviation is certainly very different if you assume that it is log normal distribution for the grain size. So, which is more meaningful but we do not know in this case. Were these parameters were obtained s0, k etc. What was a exact distribution of the grain size we do not know but, it might be good approximation to assume that it is log normal. If so, this simulation then tells you how much is the error.

So, to summarize what is said that we have done. We, have found out how to calculate error, when the function depends on more than one variable and we have calculated the error assuming that the uncertainties in these quantities are independent, when you do that there are formula that you can evaluate or you can use the partial derivative formula directly evaluate the quantity.

In some, cases you get the error to be the uncertainty in some cases you get the relative error or relative uncertainty delta f by f is the quantity that you get but, in either case you can find them out. But if you want to also include the co-variances or in-corporate distributions which are not normal for doing the error propagation you can use the library propagate and you can carry out Monte Carlo simulations to get the error.

So, this sort of completes your session on descriptive data analysis. We have looked at how to describe data, how to plot data with error bugs and also understand how the error propagations happens and with summary session we are going to conclude this module and then we will move on to the probability distributions modules thank you.