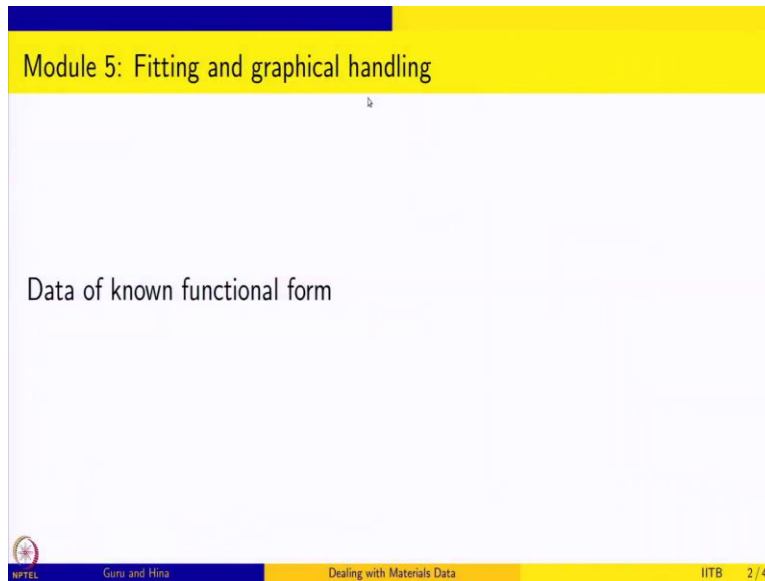


Dealing with Material Data: Collection, Analysis and Interpretation
Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 77
Data of Known Functional Form

(Refer Slide Time: 00:25)



Welcome to Dealing with Materials Data, we are looking at Collection Analysis and Interpretation of Data from Material Science and Engineering. We are in the module on fitting and graphical handling of data and we are going to look at data which has a known functional form in this session. So, we have looked at data which we know to be linear and we know how to fit that, we also know of data which has power law or exponential type of dependence and we can turn them into linear by doing logarithm a transformation on that expression.

And so you can use the transformed data at then do the analysis. And here is the third case which is also very common, where we know that there is a explicit functional form, and that functional form might not be a power law, it might not be exponential form or it might not be linear, now how do we deal with such data that is the exercise we are going to do.

(Refer Slide Time: 01:17)

Specific heat capacity

- Consider the variation of specific heat of copper with temperature in the low temperature regime (0–20K).
- Data is taken from Heat capacity of reference materials: Cu and W, G. K. White and S. J. Collocott, J. Phys. Chem. Ref. Data, 13, p. 1251, 1984.
- Plot the data: can you guess the functional dependence?

NPTEL Guru and Hina Dealing with Materials Data IITB 3 / 4

For this exercise, we are going to consider the variation of specific heat of copper with temperature in the low temperature regime, so we are going to consider between 0 and 20 Kelvin how the specific heat of copper varies, so we are going to take this data, it is from heat capacity of reference materials copper and tungsten by white and Collocott and this data we are going to take and first we are going to plot the data and look at whether we can guess the functional dependence.

(Refer Slide Time: 01:46)

Functional relationship

- It is known that $C_p(T) = aT + bT^3$
- Determine a and b

NPTEL Guru and Hina Dealing with Materials Data IITB 4 / 4

$$C_p(T) = aT + bT^3$$

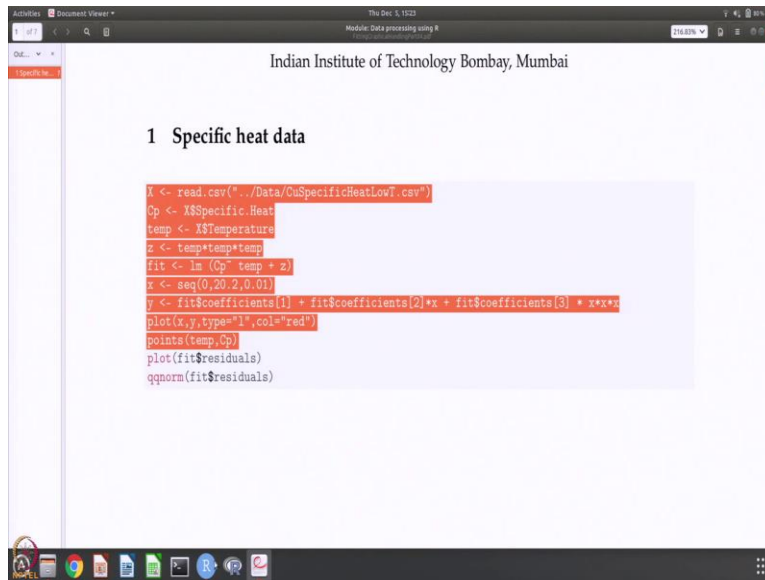
So, these are the two things that we are going to do with this data, so let us do the exercise.

(Refer Slide Time: 02:03)

Indian Institute of Technology Bombay, Mumbai

1 Specific heat data

```
X <- read.csv("../Data/CuSpecificHeatLowT.csv")
Cp <- X$Specific.Heat
temp <- X$Temperature
z <- temp*temp*temp
fit <- lm(Cp~temp+z)
x <- seq(0,20,2,0.01)
y <- fit$coefficients[1] + fit$coefficients[2]*x + fit$coefficients[3] * x*x*x
plot(x,y,type="l",col="red")
points(temp,Cp)
plot(fit$residuals)
qqnorm(fit$residuals)
```



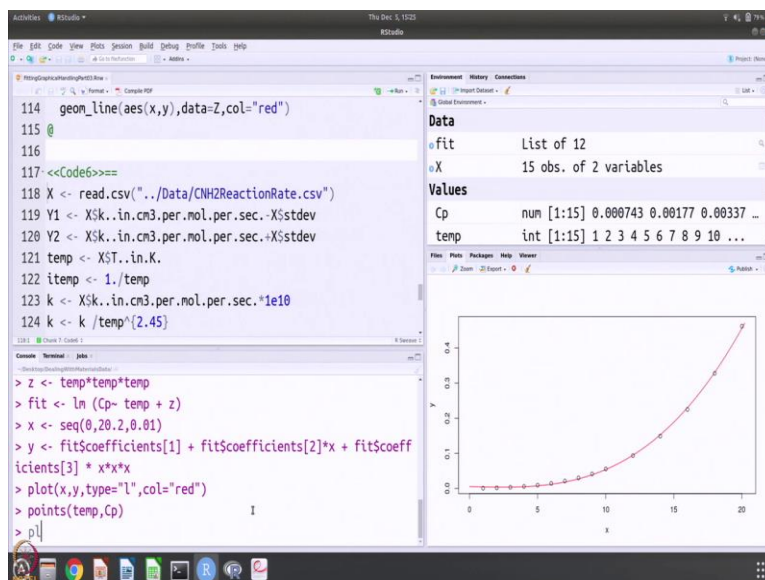
```
geom_line(aes(x,y),data=Z,col="red")
@
116
117 <<Code6>>===
118 X <- read.csv("../Data/CNH2ReactionRate.csv")
119 Y1 <- X$k..in.cm3.per.mol.per.sec.-X$stdev
120 Y2 <- X$k..in.cm3.per.mol.per.sec.-X$stdev
121 temp <- X$T..in.K.
122 itemp <- 1./temp
123 k <- X$k..in.cm3.per.mol.per.sec.*1e10
124 k <- k /temp^(2.45)
```

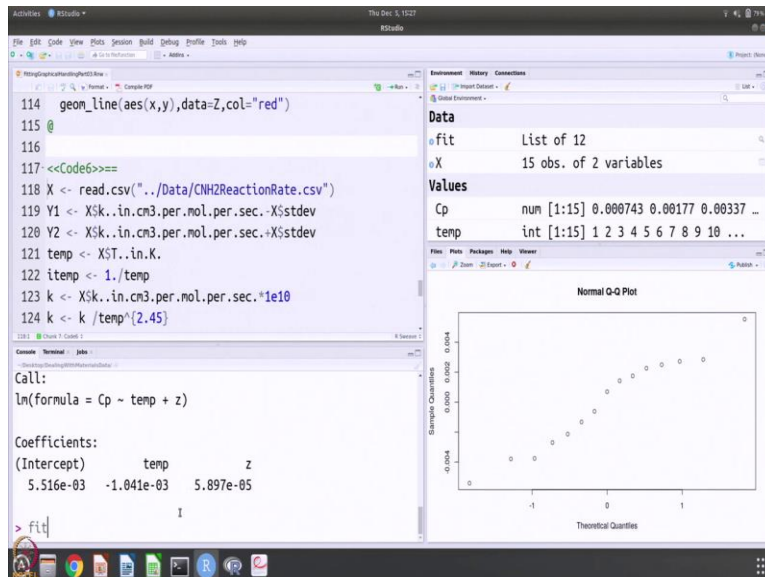
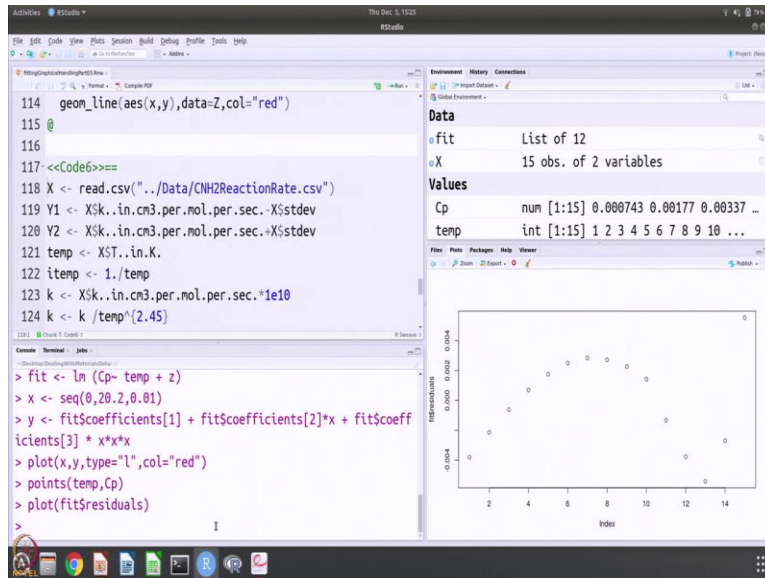
Console

```
> z <- temp*temp*temp
> fit <- lm(Cp~temp+z)
> x <- seq(0,20,2,0.01)
> y <- fit$coefficients[1] + fit$coefficients[2]*x + fit$coefficients[3] * x*x*x
> plot(x,y,type="l",col="red")
> points(temp,Cp)
> p[
```

Data

Variable	Type	Description
fit	List of 12	Linear model fit
X	15 obs. of 2 variables	Data matrix
Cp	num [1:15]	Specific heat values
temp	int [1:15]	Temperature values





So, let us start R and so we are in the write directory, so let us do the first exercise, so we read the data which is the specific heat, copper specific heat at low temperatures data, it is in a csv format, so I am going to store the specific heat in the variable Cp and temperature in the variable temp and because we know that the data fits to $aT + bT^3$, so I am going to make the T^3 as transformed into another variable z, so in terms of z this becomes linear.

So, we are going to fit a linear model for Cp, it depends on temperature and the temperature cube, but that we have turned into a linear variable now, so in terms of z it is linear, but in terms of temperature, it is actually cube. And we are going to generate a sequential, a sequence of numbers from 0 to 20.2 and 0.01 and use this fitted coefficients to calculate this is a plus bT plus cT^3

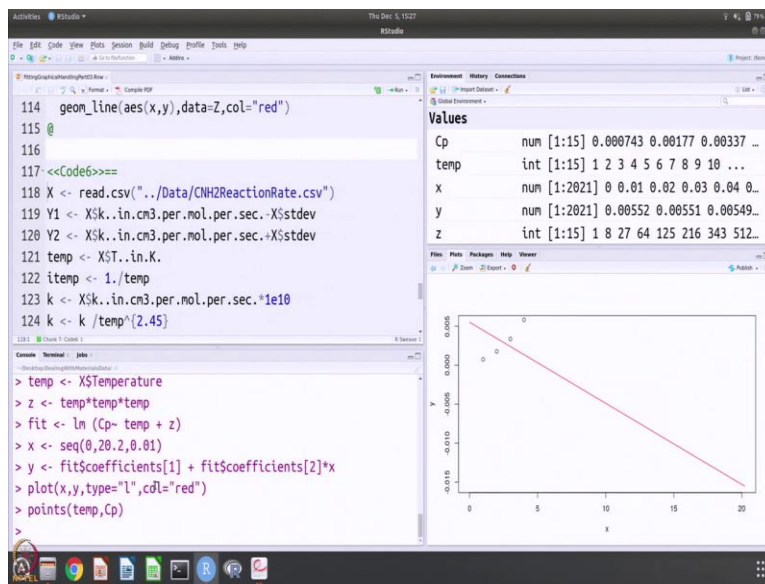
and once we have that we are going to plot that in red colour and we are going to plot the points that we have got from the experiment.

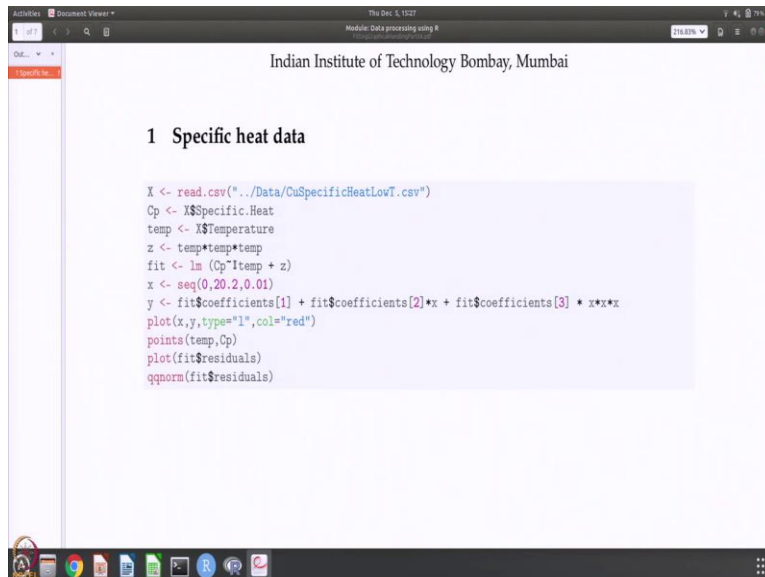
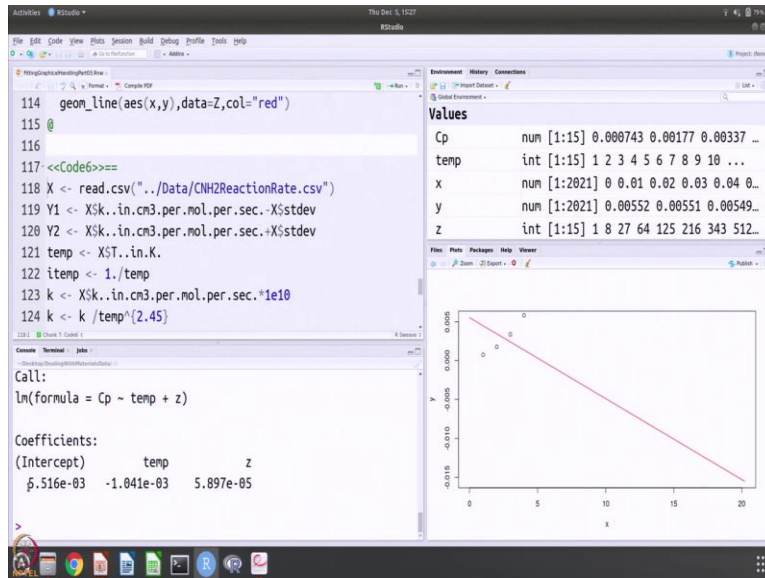
So, you can see that you have the data and the curve goes through very nicely and you can do the plotting the residuals, so you can see that the residuals are really not random and they have some nice functional dependence, so the error in this data is not normally distributed of course you can look at qq norm also and you will find that that is the case, so you will see that it is really not a straight line.

So, there is some systematic error in the data that you can make out by doing this analysis and you can look at fit and you see that there is an intercept that it is giving, so this is the coefficient for T, so that is the a and this is the coefficient for z, which is the coefficient for T cube, so it is aT plus bT cube, a is some minus 1.041 10 power minus 3 and b is 5.897 10 power minus 5, this coefficient is very small because it is multiplying T cube. For example, at 20 T cube is like 20 into 20 into 20, so that is of the order of 8000.

So, if you multiply by this small number, so this number looks small but because it is multiplying a large number it will still give you a contribution, so that is what is happening. And you can do this exercise so you can when you are plotting this line that we were plotting, you can see by leaving out this for example so fit coefficient 3 is very small.

(Refer Slide Time: 05:34)

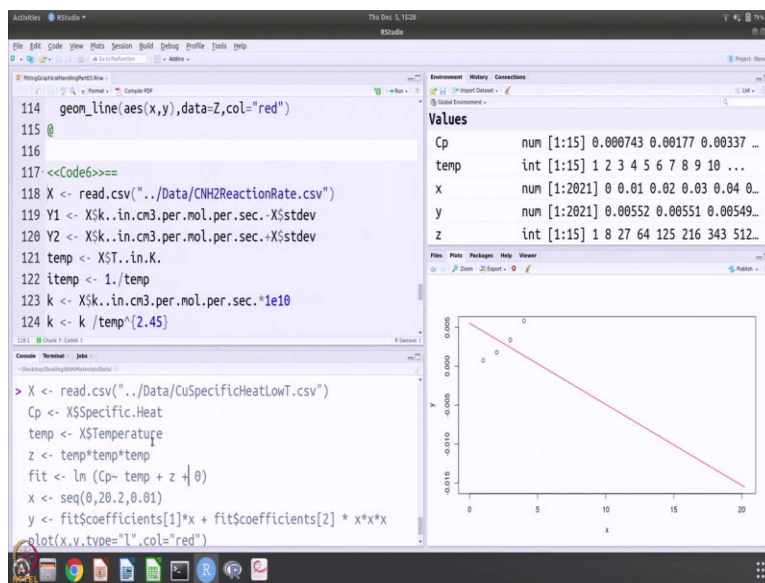
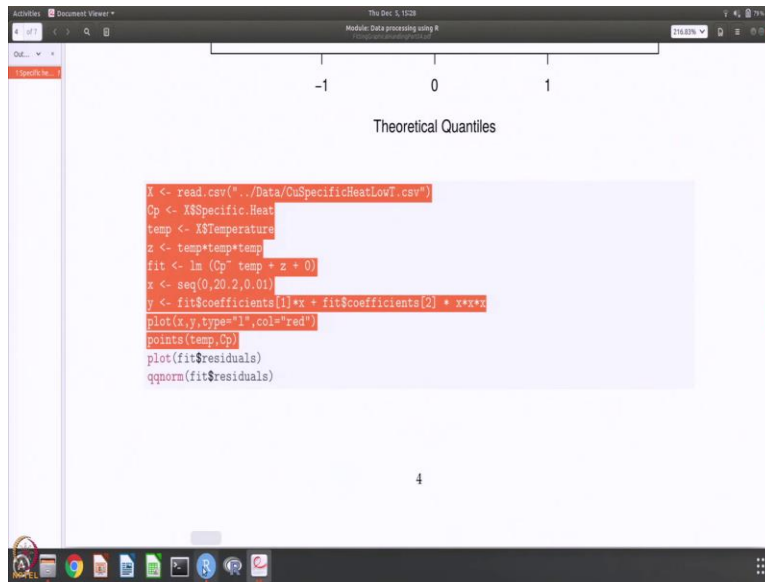


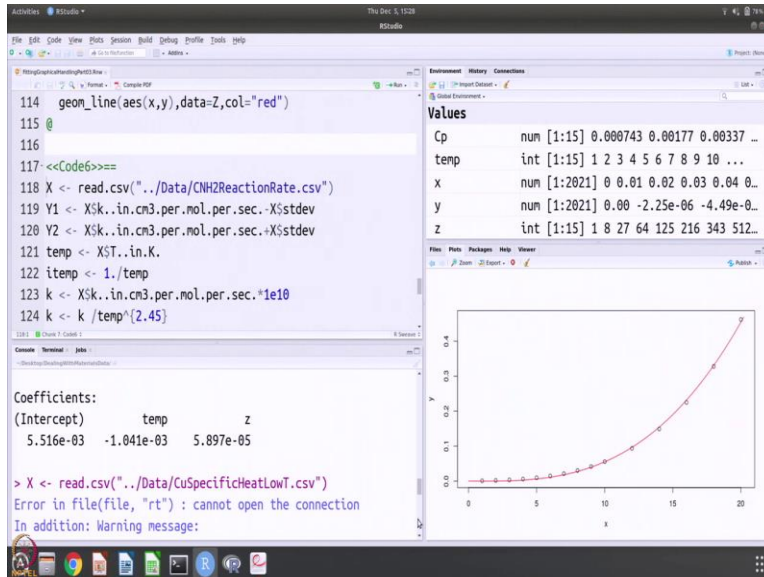
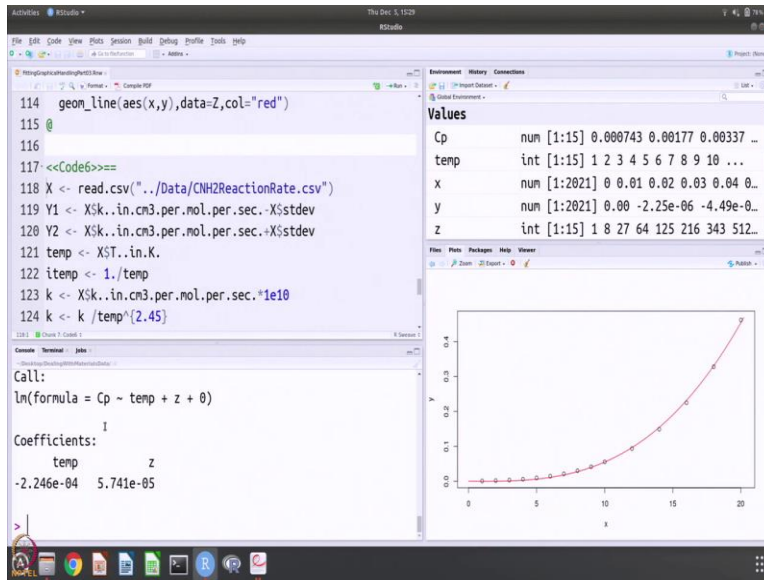


And so if you plot how does it look? So, you can see that it looks very different, so that is obviously because even though the coefficient is small it does contribute because the multiplying number is very large. So, you can play around like this to just to get a feel for this, but we do see that it gives an intercept, but we know from our theory that the specific heat should fit to aT plus bT cube at T is equal to 0 we should not get any intercept.

Now, how do we tell or that it should not put any coefficients. Because we tried T plus T cube, but it is adding the constant, so I have to explicitly say that the constant should be 0, well it is very easy, you just have to add a 0 to say that okay if there is a constant that is 0.

(Refer Slide Time: 06:29)





So, if you do that and that is the only difference between the 2 scripts that we have written. So, this script is exactly the same as our earlier script except that now it is temp plus z plus 0 to indicate that though intercept should be 0 and so there are only 2 coefficients one is for temp, other one is for temp cube, which is called z in our case. So, you can do and of course you find that it fits and fit parameter you can look at that there is only coefficient for temp.

And these coefficients are different from what we got in the earlier case see the temperature coefficient now is very different from what we had when it is 1.04 10 power minus 3, which has become 2.2 10 power minus 4 and that is because there was an error that was coming from the intercept which should have been 0 which was not 0.

On the other hand, I think the parameter that multiplies z it is 5.9 here and it is about 5.7 here, so it is relatively small but because again this number multiplies T cube even small changes here can make a difference so this is a much better fit. So, we now know exactly how to fit to a given polynomial form, so we have fitted to aT plus the bT cube.

And sometimes for example, there are other functional forms that are known to fit for specific heat at other temperature ranges, so sometimes it is $aT + bT^2 + cT^3$ and things like that. So, we will do one more exercise and we will see in all these cases, if the data is perfect when you fit it, for example, even you can fit it to a third order polynomial and it should give 0 for the intercept and these temperature squared term, but that is not going to happen because there are errors in data.

So, if you try to fit it will generically try to fit and it will give you some numbers for those quantities also which should in principle be 0. And in those cases it is useful to explicitly put the functional form, this is something that we have seen in the other case also where the theoretically we know that it should be T to the power 2.45, but if you just try to fit for the generic b it will give you a number and that number was something like 2.9, so it was very different from 2.45 20 percent off.

So, in order to make sure that you do not make such errors then you fit it exactly to that function like this case we said intercept should be 0, so you can demand that it should be 0 and if you do that, then you will get the right parameters. So, to summarize we are looking at fitting, this is to take the data and there are independent parameters that we change and we look at the parameter that we measure and there are errors in this process and taking this into account can we say anything about the functional relationship between the independent variables and the measurements is what we are trying.

And we have looked at several cases linear and quantities that can be turned into linear form and general functional dependence and how to do the fitting and there are more data that we are going to use like I mentioned at the beginning of this module there is lots of data on copper in terms of its grain size and temperature and cold work percentage et cetera. And what is their effect on the strength of copper for example.

So, we are going to look at such data, we are going to do more of this analysis, more of this fitting and in general the fitting itself as we have mentioned, could be because we know that there is a

functional form so we want to get the parameters, could be because we do not know the functional form so we, can we guess what the functional form is going to be from the data and sometimes we have not even interested in actually getting a functional form of this thing, it is just a way of making a simple mathematical model which mimics what happens in the real experiment.

So, this you can use it like a tool that I have done some twenty experiments, I know the data based on this, I am going to predict if the input parameter comes something what is going to be my output parameter. In that sense also this is very powerful and these are the techniques that are useful for building statistical models for making predictions, which we are not going to get into in this course, but when you get into those problems like machine learning or artificial intelligence and things like that, they are all based on these ideas.

So, there is this idea of fitting a statistical model and that fitting is done in some sense by optimizing the error or the deviation of the data from the fitted line. So, it is a combination of this statistical tools and optimization tools that actually makes it possible for us to build such predictive models.

So, regression or this fitting exercise happens to be the very basic statistical concept that we need to understand to better understand this model. So, we are going to be spending some more time, we will do few more exercises to look at this fitting and graphical handling of data before we move onto some case studies to close this course. Thank you.