

**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 78**  
**Calibration, Fitting, Hypotheses Testing**

(Refer Slide Time: 00:25)

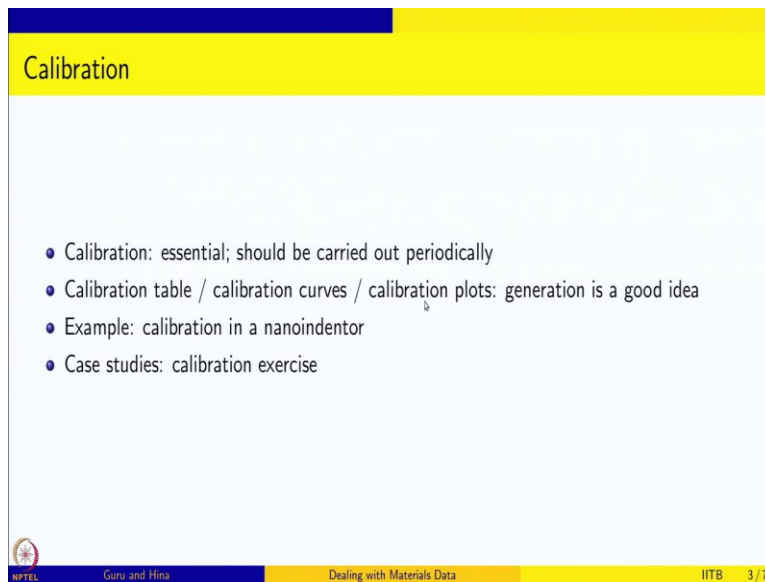
Module 5: Fitting and graphical handling

Calibration, Fitting, Hypotheses testing

NPTEL Guru and Hina Dealing with Materials Data IITB 2/7

Welcome to Dealing with Materials Data, we are looking at the Collection Analysis and Interpretation of Data from Material Science Engineering and we are in the module on fitting and graphical handling of data. We are using R to do fitting as well as plotting to understand the data better. In this session, we are going to look at calibration, fitting and hypothesis testing.

(Refer Slide Time: 00:44)



Calibration

- Calibration: essential; should be carried out periodically
- Calibration table / calibration curves / calibration plots: generation is a good idea
- Example: calibration in a nanoindenter
- Case studies: calibration exercise

NPTEL Guru and Mitra Dealing with Materials Data IITB 3/7

Calibration as we discussed sometime back is very essential and it should be carried out periodically, after calibration is done typically calibration table or curve or plot is generated and sometimes for doing this fitting is used, so you will have a material or some process for which you know what the reading should be and you will use your equipment and you will read what the equipment gives as the reading and the difference is to be corrected because we know the calibration material we will use that to correct.

So, when you work on a new material, then there is a difference then you can correct for the actual reading and for doing this typically one also has to look at the data and generate calibration curves or calibration plots. One example is for example calibration that is done in nanoindenter if you are doing in addition experiments. So, we will in our case studies look at a calibration exercise to give you an idea of how it works, typically it involves fitting and that is why it is in this topic.

(Refer Slide Time: 01:59)

Linearisation of functional forms

- $y = A \exp(-kx)$ : linearised by taking logarithm
- $y = A + B \exp(-kx)$ : linearise by taking logarithm of  $y - A$
- $y = (x - a)^p$ : linearise by taking logarithm
- $y = \frac{ax}{b+x}$ : linear in  $\frac{1}{x}$  versus  $\frac{1}{y}$  (Lineweaver-Burk plot);  $\frac{y}{x}$  versus  $y$  (Hanes method) and  $\frac{x}{y}$  versus  $x$  (Eadie-Hofstee method) also produce linear plots
- Hall-Petch relation:  $\sigma = \sigma_0 + \frac{k}{\sqrt{d}}$ : linear with the transformation  $\frac{1}{\sqrt{d}} \rightarrow x$

NPTEL Guru and Hina Dealing with Materials Data IITB 4/7

The second one is that we have been looking at fitting and we first looked at linear fitting and then we looked at cases where it can be turned into a linear form and fitted and also nonlinear, for example we fit it for specific heat which goes as  $aT$  plus  $bT$  cube. Now, there are functions that can be linearized, we have already seen some examples. For example, it is exponential you can take logarithm and if it is exponential with the

$$y = A + B \exp(-kx)$$

Linearise by taking logarithm of  $y - A$

And if it is power law, we saw one example that by taking logarithm you can make it a straight line, but there are also cases for example,

$$y = \frac{ax}{b+x} \text{ linear in } \frac{1}{x} \text{ versus } \frac{1}{y}$$

and that will be linear that is known as a Lineweaver-Burk plot, you can also plot  $y$  by  $x$  versus  $y$ , or  $x$  by  $y$  versus  $x$  and they also produce linear plots, the idea behind this is to just see that there is a linear relationship that exists after that you can actually do the analysis and find the parameters for the linear fit.

One more example where you will see which can be turned into a linear form is the Hall-Petch relationship for the flow stress in materials, which is related to the grain size through this one by root  $d$  relationship. Of course if you do the transformation  $1$  by root  $d$  as  $x$  then you get  $\sigma$  is

equal  $\sigma_0 + kx$ , so this is the intercept and this is the slope. These two  $\sigma_0$  and  $k$  are known as the Hall-Petch parameters.

(Refer Slide Time: 03:36)

Hall-Petch coefficients for copper

- Data from NIST Monograph 177, Properties of Copper and Copper Alloys at Cryogenic Temperatures, N J Simon, E S Drexler, and R P Reed, 1992.
- Yield strength (MPa) as a function of grain size (in microns) at 295 K
- Hall-Petch:  $\sigma_y = 18.6(\pm 1.7) + 112(\pm 2)d^{-\frac{1}{2}}$
- Standard deviation in  $\sigma_y$ : 11 MPa

IITB 5/7

And we will do one exercise, so we will take the data from NIST monograph and which gives yield strength as a function of grain size at 295 Kelvin and the monograph also tells you that the Hall-Petch fitted parameters are 18.6 plus or minus 1.7 for the intercept and 112 plus or minus 2 for the slope of the Hall-Petch relationship.

And the standard deviation in the flow stress itself is 11 MPa of course once you know these two errors it should be easy for you to calculate what the error in  $\sigma_y$  should be, so this is something that we have already looked at, but how do these numbers come about? So, we will take the same data which was used to produce these numbers and generate these numbers for ourselves in order to understand how it works. So, it is a simple linear fitting that we are going to do.

(Refer Slide Time: 04:27)

The slide is titled "Fitting and hypotheses testing" and contains the following text:

- Fitting: do obtain parameters when functional form is known, to identify if there are correlations (and functional form if yes) : getting parameters from data and estimate the parameters of the underlying probability distribution
- Fitting: also to test if the data supports the given functional form: hypothesis testing
- If we just fit the copper strength data, will we get the exponent of  $d$  as  $-0.5$ ?
- Problem of hypothesis testing!

At the bottom of the slide, there is a footer with the IITB logo, the text "Guru and Mitra", "Dealing with Materials Data", and "IITB 6/7".

So, fitting as we will, we have seen is done for several reasons, one is to obtain parameters when functional form is known, for example we are assuming that Hall-Petch relationship holds and then we want to know what is  $\sigma_0$  is, so you can do that, or to identify if there are correlations, so there you can do all these transformations like take log or take  $y$  by  $x$  versus  $y$  or  $x$  by  $y$  versus  $x$  and things like that and plot.

And if they show a straight line, then you know that the functional relationship could be of the form  $ax + b$  for example, so you can do a fitting exercise and that is where plotting is very useful, graphical analysis is very useful, because you can just plot and see and if there is a relationship that you see then you can try to find the functional form, or if the functional form is known then you can obtain the parameters.

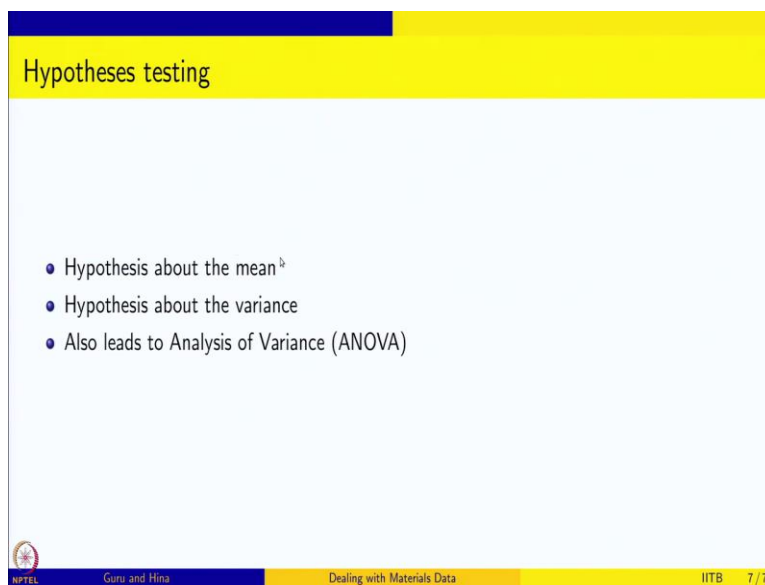
And getting parameters from data and estimating the parameters of the underlying probability distribution is something that we have already seen, so this is done and so it is part of the fitting exercise. But fitting can also be done to test if the data supports a given functional form, so this is the hypothesis testing. For example, you can take the data and you can ask the question is it true that the flow stress is related to the grain size as power minus half, so that is a question that you can ask and to test whether the given data supports this hypothesis is the hypothesis testing.

And so for example, you can ask the question fit the copper strength data will be get an exponent of  $d$  as minus 0.5. And this problem is the problem of hypothesis testing, so we are going to deal

with this in greater detail not just for copper, we will take lots of materials for which data has been collected by one of the researchers, it is available in the literature available for everybody in raw form to download and do the analysis.

So, that is something that we are also going to do as part of our case study in the next module. But for now, we will at least see how to fit and if you assume that there is a given form can we get the parameters, that is the part that we are doing now and we will continue doing in this session and the next.

(Refer Slide Time: 06:50)



The slide is titled "Hypotheses testing" and contains a bulleted list of topics. The slide has a yellow header and a blue footer. The footer contains the NPTEL logo, the name "Guru and Hina", the course title "Dealing with Materials Data", and the slide number "IITB 7/7".

- Hypothesis about the mean<sup>3</sup>
- Hypothesis about the variance
- Also leads to Analysis of Variance (ANOVA)

So, hypothesis testing is something that we have already seen at some level, for example we said, what is the probability that the mean will lie within some range, so you can also have a hypothesis that the true mean is this or it is in this range and then you can test whether the hypothesis is true, you can also make hypotheses about variance and test.

And you can test the hypothesis about means from different experiments, so it naturally leads to analysis of variance, which is something that we are going to look at in this module in one of the sessions and from there it also leads to design of experiments and so on. So, we will do a design of experiments case study also in the next module.

(Refer Slide Time: 07:48)

The image shows two screenshots from an RStudio environment. The top screenshot is a presentation slide titled "Module: Data processing using R" by M P Gururajan and Hina A Gokhale at IIT Bombay. It features a section titled "1 Copper: Hall-Petch" and a block of R code. The bottom screenshot shows the RStudio interface with the same code executed in the console, displaying the R help text and a plot of the Hall-Petch relationship.

**Module: Data processing using R**

M P Gururajan and Hina A Gokhale  
Indian Institute of Technology Bombay, Mumbai

### 1 Copper: Hall-Petch

```
X <- read.csv("../Data/CuStrengthGrainSize295K.csv")
D <- X$d..micron
YS <- X$YS..M.Pa
d <- 1./sqrt(D)
plot(d,YS)
fit <- lm(YS~d)
abline(fit$coefficients)
plot(fit$residuals)
qqnorm(fit$residuals)
```

Console Terminal Jobs

You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

```
> X <- read.csv("../Data/CuStrengthGrainSize295K.csv")
> D <- X$d..micron
> YS <- X$YS..M.Pa
> d <- 1./sqrt(D)
> plot(d,YS)
> fit <- lm(YS~d)
> abline(fit$coefficients)
>
```

Environment History Connections

Data

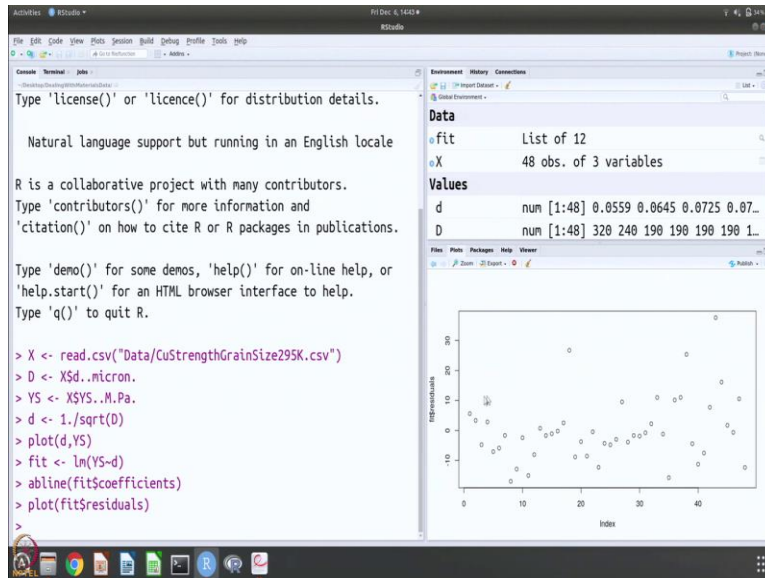
- fit List of 12
- X 48 obs. of 3 variables

Values

Variable	Values
d	num [1:48] 0.0559 0.0645 0.0725 0.07...
D	num [1:48] 320 240 190 190 190 190 1...

The plot shows a scatter of data points with a linear regression line. The x-axis is labeled 'd' and ranges from 0 to 4. The y-axis is labeled 'YS' and ranges from 0 to 4000. The residuals plot below shows the distribution of residuals, which appears to be roughly normal.





So, for now what we are trying to do is to do the fitting exercise. So, let us do that. So let us start R version 3.6.1 and what we want to do is the Hall-Petch for copper. So, let us just try to do this, so what is the process? We first want to read the data and it is in csv format, it is copper strength grain size at 295 Kelvin in csv format, this data is taken from the NIST monograph and then we are going to take the grain size and store it in the variable capital D, we were to take the yield strength and store it in the variable YS.

And the small d is 1 by square root of capital D, so this is D to the power minus half and then we are going to plot d versus yield strength and we are also going to fit yield strength as a function of d, so the intercept and the slope should give us the Hall-Petch parameters. So, you can see that the data the d versus yield strength is like this, so this is actually square root of the 1 by square root of the grain size.

And so you can see that it is like this, so we can try to see where the fitted line is, so the fitted line goes through all these points and so it does seem to follow this relationship and of course you can also plot the residuals and you can see that about 0 on either side the data is spread and there are a few data points which are slightly away, but most of the data is between minus 10 and or between minus 20 and plus 20, so you can see that the data is that the residuals is random or it looks random.

(Refer Slide Time: 10:02)

Natural language support but running in an English locale

R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

```
> X <- read.csv("Data/CuStrengthGrainSize295K.csv")
> D <- X$d..micron.
> YS <- X$YS..M.Pa.
> d <- 1./sqrt(D)
> plot(d,YS)
> fit <- lm(YS~d)
> abline(fit$coefficients)
> plot(fit$residuals)
> qqnorm(fit$residuals)
>
```

**Data**

fit	List of 12
X	48 obs. of 3 variables
Values	
d	num [1:48] 0.0559 0.0645 0.0725 0.07...
D	num [1:48] 320 240 190 190 190 190 1...

**Normal Q-Q Plot**

Type 'q()' to quit R.

```
> X <- read.csv("Data/CuStrengthGrainSize295K.csv")
> D <- X$d..micron.
> YS <- X$YS..M.Pa.
> d <- 1./sqrt(D)
> plot(d,YS)
> fit <- lm(YS~d)
> abline(fit$coefficients)
> plot(fit$residuals)
> qqnorm(fit$residuals)
> fit
```

**Call:**  
lm(formula = YS ~ d)

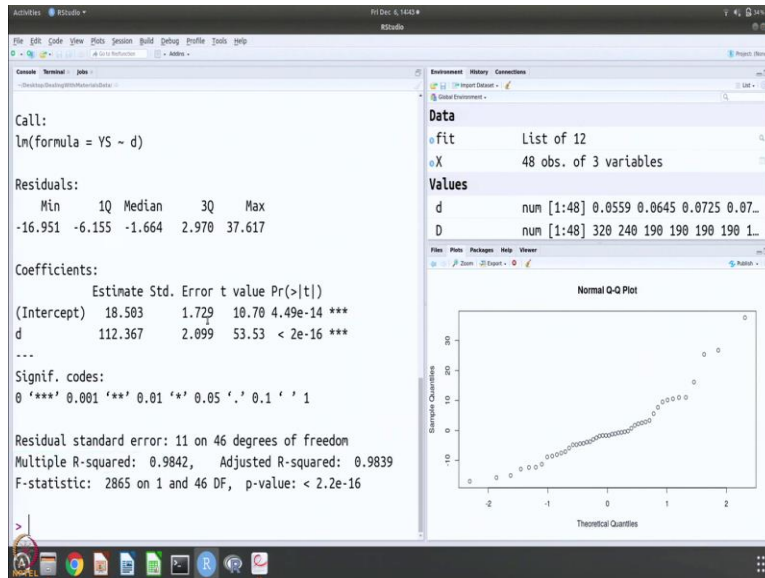
**Coefficients:**

(Intercept)	d
18.5	112.4

**Data**

fit	List of 12
X	48 obs. of 3 variables
Values	
d	num [1:48] 0.0559 0.0645 0.0725 0.07...
D	num [1:48] 320 240 190 190 190 190 1...

**Normal Q-Q Plot**



## Hall-Petch coefficients for copper

- Data from NIST Monograph 177, Properties of Copper and Copper Alloys at Cryogenic Temperatures, N J Simon, E S Drexler, and R P Reed, 1992.
- Yield strength (MPa) as a function of grain size (in microns) at 295 K
- Hall-Petch:  $\sigma_y = 18.6(\pm 1.7) + 112(\pm 2)d^{-\frac{1}{2}}$
- Standard deviation in  $\sigma_y$ : 11 MPa



However, if you plot the qqnorm for example, I do not see a straight line, so there is, there seems to be some problem, I mean it is not, the error does not seem to be normally distributed, you can also look at the fit and of course you get 18.5 and 112.4 and you can call for summary fit, so it gives you 18.5, 1.7 and 112.367, 2, so if you look at the fitting parameters that was given by the monograph, so it is 18.6 plus or minus 1.7 112 plus or minus 2.

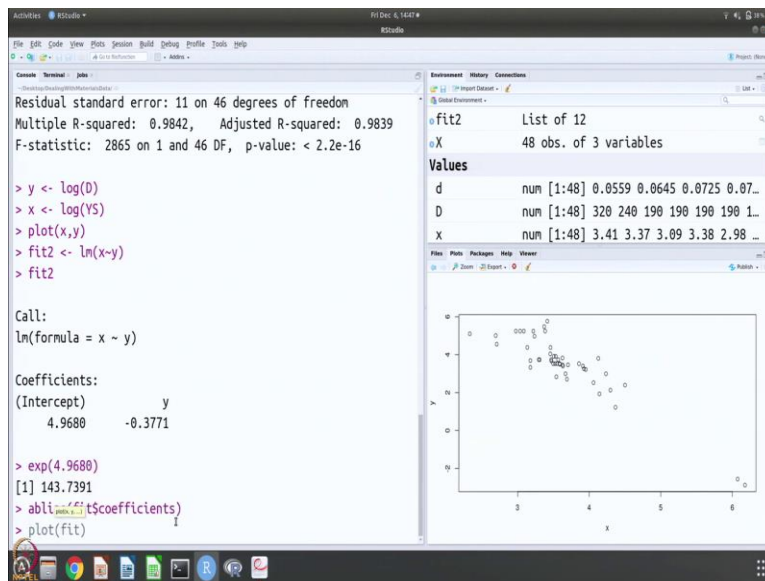
And in this case, we see that so we get 18.5 instead of 18.6 the error is 1.7 and that is right and 112 so that is 112 and 2 is the error in the slope of d to the power minus half that is, so we see that we get the parameters that is described in Hall-Petch, but there is a problem if you want to know if it

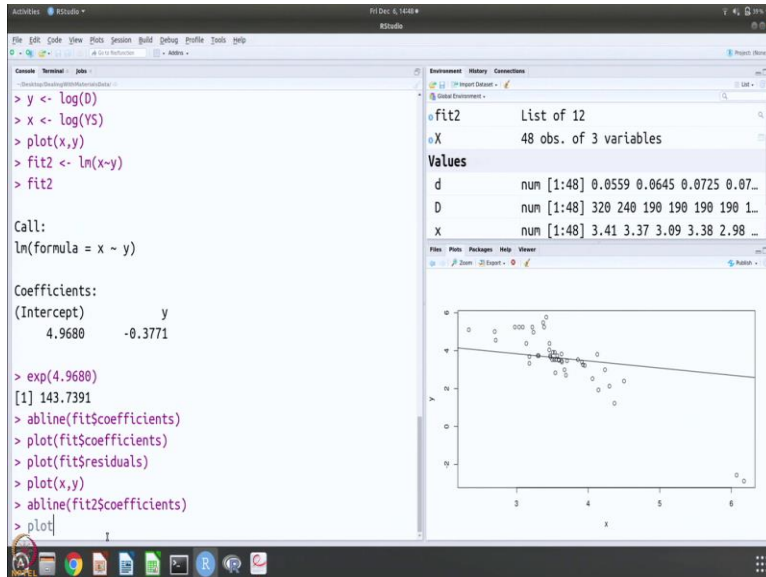
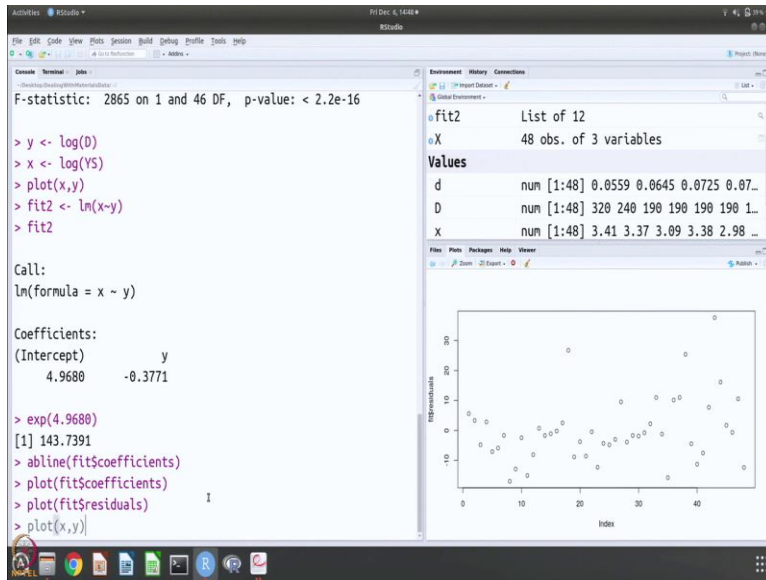
is actually  $d$  to the power of minus half because the Hall-Petch relationship is like this, so the Hall-Petch relationship is like this.

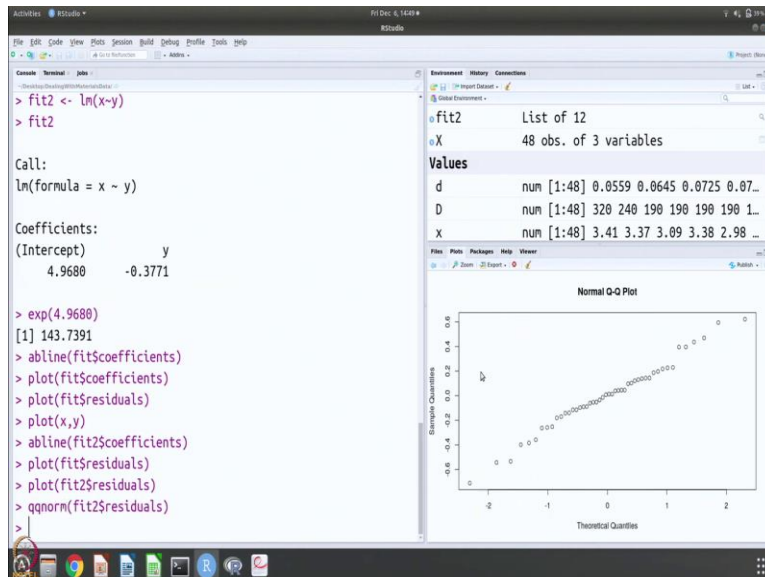
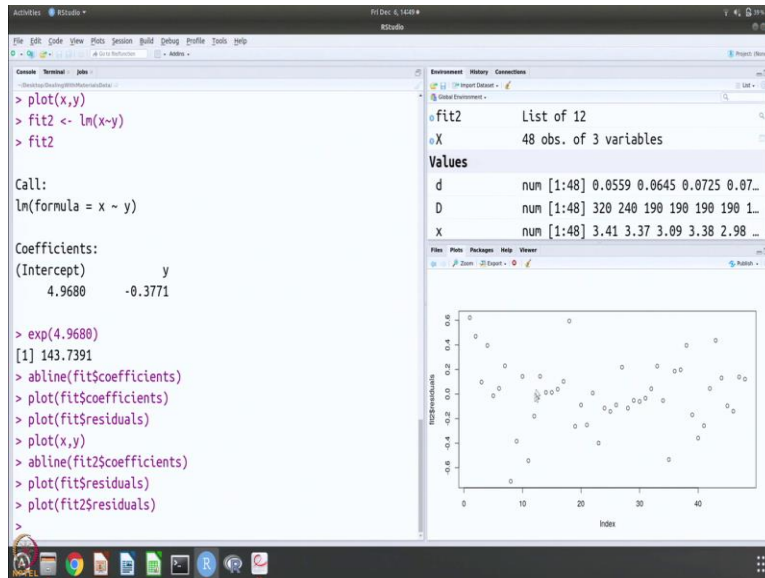
And so if you want to know whether it is  $d$  to the power minus half, one of the things that you can do is to take logarithm of  $\sigma - \sigma_0$  and logarithm of  $d$  and try to see if that gives you a factor of half and logarithm of  $k$  then will be the it will be  $\sigma - \sigma_0$  so you take log, so you will have  $\log k$  and  $d$  to the power some minus  $n$ , so if you, so you will have  $\log k$  minus  $n$  times  $\log d$ , so that  $n$  should be half or minus half and  $k$  should be given by the intercept.

But we do not know  $\sigma_0$ , so we assume that it is a constant, so we just take, so we can just take the data and it is not quite complete, but one can try and do this exercise, so let us say that we have  $d$ , so we take logarithm of  $d$  and store it as  $y$  and we store as  $x$  the logarithm of yield strength.

(Refer Slide Time: 13:17)







So, we can see whether we can plot this and does it follow a straight line? So, let us try to do the fitting and the fitting is  $y$ , sorry  $x$  as a function of  $y$ , so  $fit2$ , so the parameters are like this, so if you take exponential 4.9680 so that is 143, when we fitted we got 112 so the  $k$  happens to be 143 and  $y$  happens to be minus 0.3 so it is not minus 0.5, so it is minus 0.4.

So, there seems to be slight deviation so you can also look at the line that you get so we can look at the error, so again the error, so we called it  $fit2$ , so let us do this again and then so you can see that this is a line and the data is different, so this is slightly off, of course but if you look at the, so  $fit2$ , so you see that about 0 on either side this is again normally distributed we can probably normally distributed we can look at  $qqnorm$ .

So, this certainly seems to be the error now seems to be normally distributed, but we so we assumed that  $d$  goes as power minus half and we fitted and we got some parameters. And as far as fitting goes that is good enough, because if you look at the NIST data, this is from several different sources the data is not even from a single source, so if you look at the grain size versus yield strength and all the existing data can be put in one single form and within some 11 MPa plus or minus you can actually predict the yield stress, which is very good.

However, if you ask a different question namely that is the power minus half and try to do the fitting you see that it is not so surely approximate because there is still a problem  $\sigma - \sigma_0$  and we have not taken that  $\sigma_0$  into account or if you assume that the  $\sigma_0$  is 18.5 like we got earlier you can do this exercise, you can try subtract it and then for the remaining quantity you can try to do  $k \log d$  kind of fitting a  $\log k + m \log d$  kind of fitting and find out what the  $m$  and  $\log k$  values turned out to be.

And see if that actually gives you this functional form. But we are going to do it in greater detail in the next module when we look at some case study so we will take Hall-Petch as a specific case study and check if there is enough statistical evidence and any physical reason why we expected to be minus half or could be something else. So, that is a question that we are going to look at slightly later.

But this is the data that is available and there are, there is lots more in this monograph, it also gives the yield strength as a function of percentage cold work and temperature and so on and so forth and impurities, concentration of impurities and so on and so forth. So, it is a good idea to explore the data and try to see if there are trends.

And for example, if the two data are correlated and questions like that one can ask. And one can use fitting as the starting point for exploring the data further. So, in the next session, we are going to look at analysis of variance and that will bring us to the end of this module and then we will move on to the case studies module. Thank you.