Dealing with Materials Data: Collection, Analysis and Interpretation

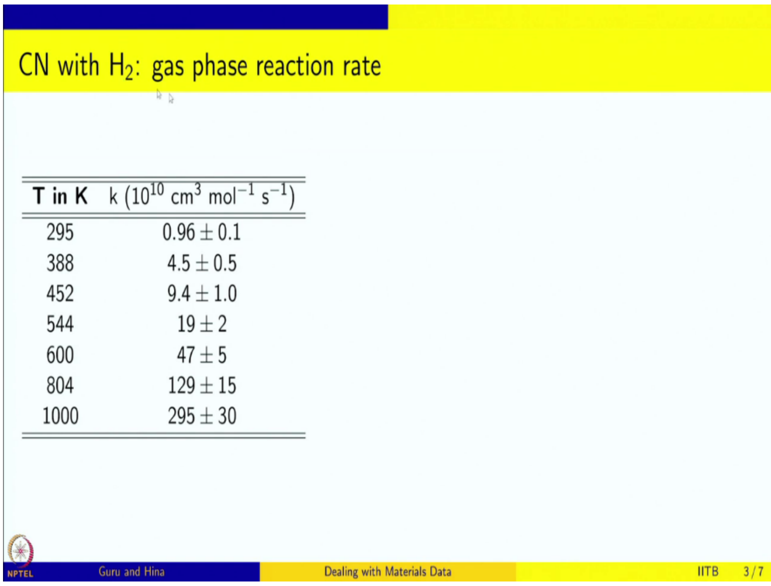
Professor M P Gururajan
Professor Hina A Gokhale
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

Lecture No. 95

Case study 2: Error analysis

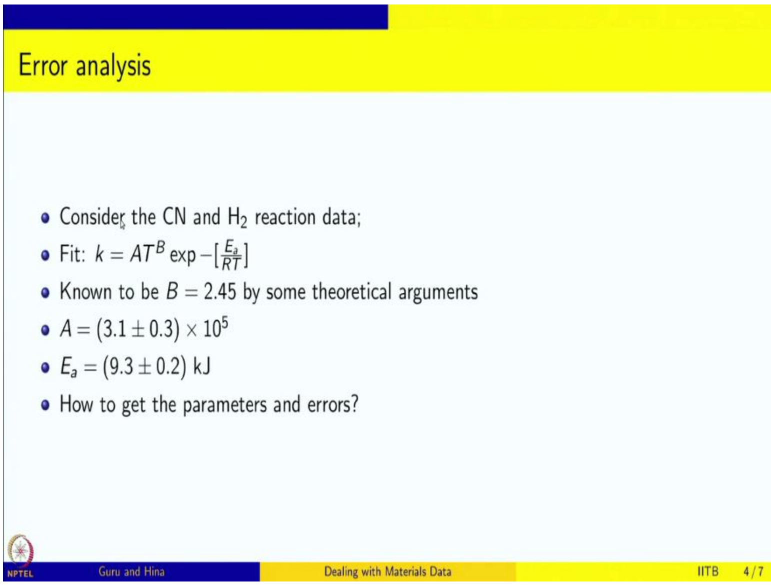
Welcome to Dealing with Materials Data. This is a course on Collection Analysis and Interpretation of Data.

(Refer Slide Time: 0:26)



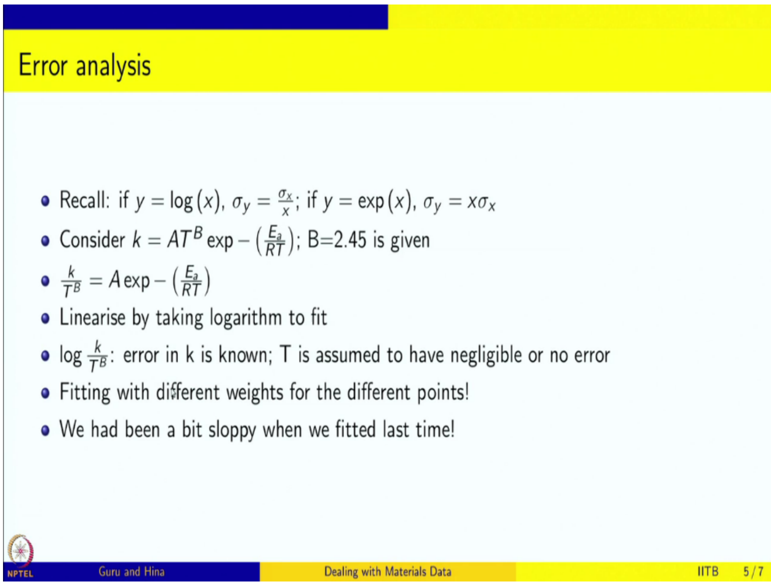
We are looking at module 6 and this is a second case study and this is a case study on Error Analysis.

(Refer Slide Time: 0:35)



We have looked at this data earlier. This is cyanide interaction with hydrogen, it is a gas phase reaction and reaction rate is given with error and it is given in 10 to the power 10 centimetre cube mole inverse second inverse and the temperature is given in kelvin. We have done this exercise.

(Refer Slide Time: 0:56)

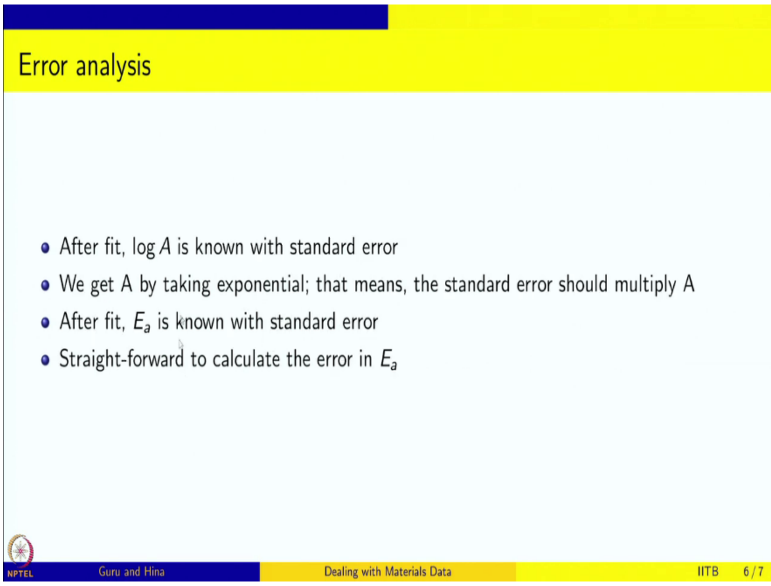


We have tried to, but we have not done the error analysis bit carefully, so what we want to do is that okay, so we will take this reaction rate data and we want to fit it to this form because we know that it should go as some constant temperature to the power B and by some theoretical arguments we know that these B is 2.45, exponential minus Ea by RT. So you can, we have done this exercise fitting.

We have taken k by T to the power B as the quantity and we have taken log on either side, so it gives log a minus Ea by RT, so by 1 by T versus log k by T to the power B, if you fit then you can get the intercept from which you can calculate a and from the slope you can calculate Ea by R. And the paper gives these values as 3.1 plus or minus 0. 3 into 10 power 5 and 9.3 plus or minus 0.2 kilo joules the A and Ea.

How do you get these parameters and we have done this exercise earlier and but we did not discuss in detail how the errors themselves got calculated. So in this exercise we are going to do this.

(Refer Slide Time: 2:09)

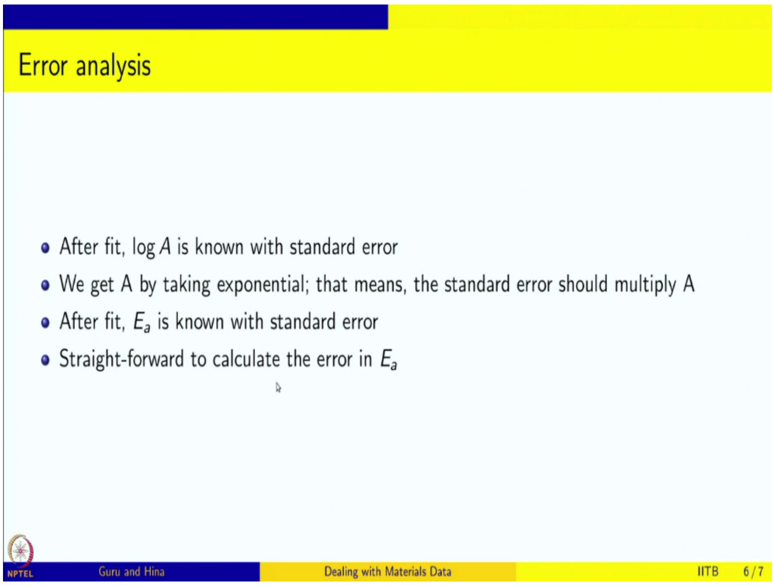


To do so, we have to recall that if why is log of x then the error in y is nothing but error in x by x, that comes because we take dho y by dho x and sigma x, so log x if you differentiate that it gives you 1 by x, so if you know the error in the x quantity, sigma x, then sigma x by x will be the error in log of that quantity, which is y and similarly if y is exponential x, the error in y will be just, the error in x multiplied by the x itself.

So we are going to take this quantity, and we are going to fit it for k by T to the power B and we are going to linearize by taking logarithm. And so log k by T to the power B, error in k is known and we are going to assume that the temperature has negligible or no error and in order to fit with different weights, we have to use the error that is given for k and when we fitted last time, we had been a bit sloppy because we just took the error value.

But we know that because we have done the log transformation, the error should also be transformed by making it sigma x by x, so the whatever error that was given for k, that is the delta k should be divided by k and that should be the error that should be used for fitting purposes. So we are going to do this exercise now.

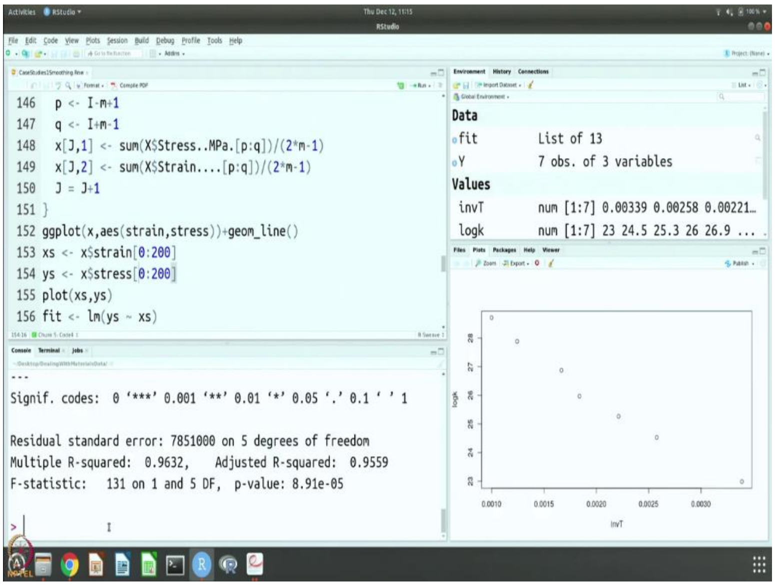
(Refer Slide Time: 3:42)

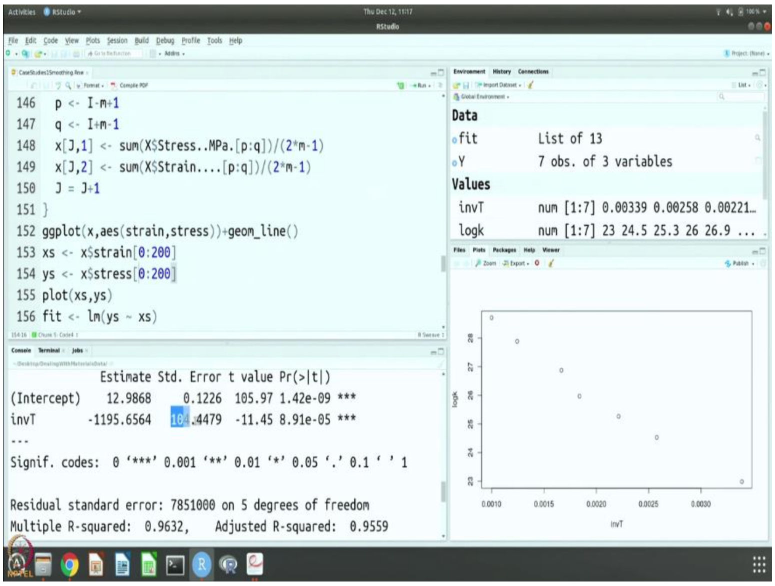


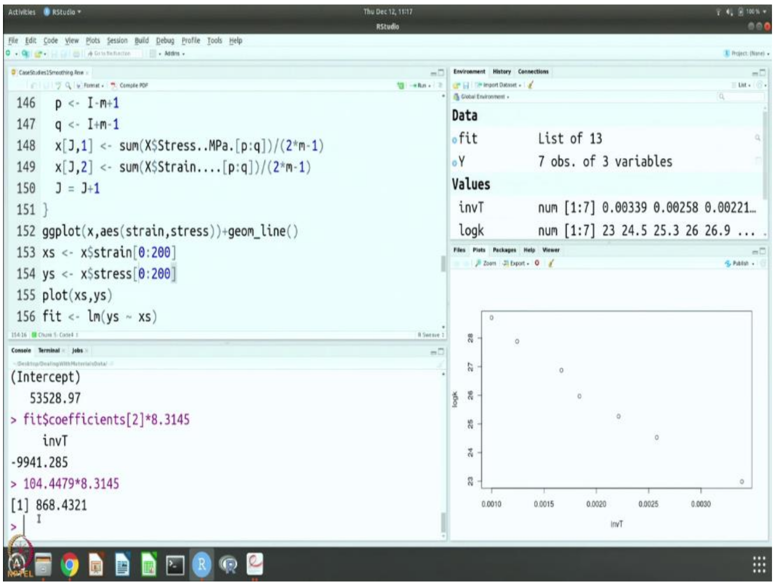
And then once a fit is known, of course you will know log a and the standard error in log A from which we are going to get the A value and the error in A value and because it is exponential, now you have to use the other formula, where you said that it is x into sigma x is the error in sigma y. And after fit, of course, Ea is known and it is straight forward to calculate the error in Ea.

So this is the exercise we are going to do and the paper that I have referred to also has the data for cyanide interaction, reaction with oxygen and I am going to leave that as an exercise for you to do and calculate the parameters as well as the errors in them, so that is what you will do. But for now we will take the cyanide hydrogen data and do the exercise as we did last time. But this time we will be little bit more careful with the way we calculate the errors.

(Refer Slide Time: 4:37)







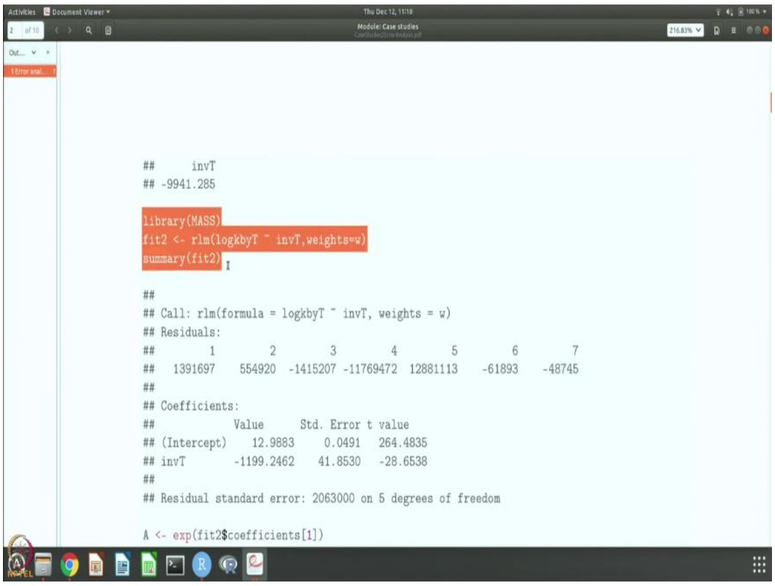
So, let us do this exercise. So first we are going to read the data and inverse T is nothing but 1 by T and log k is nothing but the k value and it is 10 power 10, I remember, so we can plot inverse T versus log k and log k by T is nothing but logarithm of K divided by T to the power 2.45 and now the standard deviation that we have for k has to be transformed because it has to be divided by the k itself.

And because this y, I mean T to the power 2.45 is a constant and we are going to consider it like some alpha and so we are going to just carry that constant. And so this is the standard deviation, and so the variance is squared and we are going to use 1 by standard variance as the weight for our fitting exercise. So this is the fitting exercise and this is the summary fit. So let us run this code. So we have this T versus log k and it has a slight curvature because it is also dependant on T through the other formula.

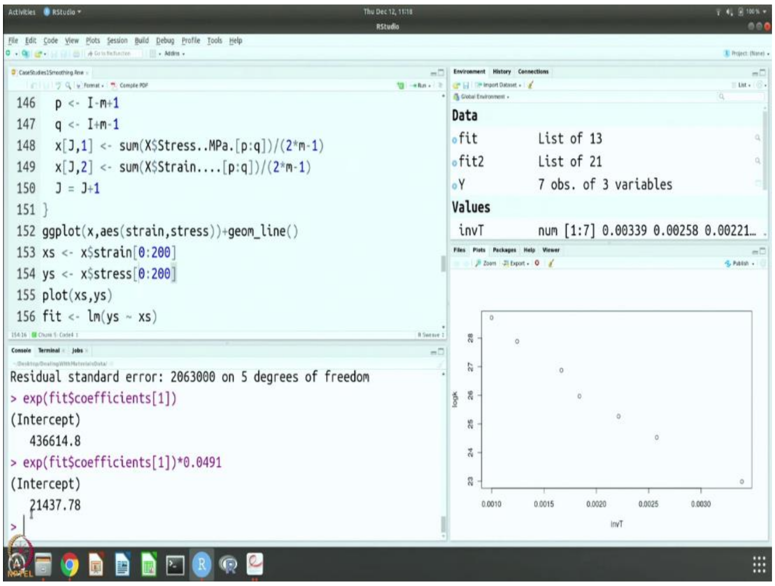
Now we have the estimated value for intercept and standard error and so you can calculate, so you can, this is the standard error. So if you have the exponential of, so you get 4.366148 into 10 power 5 as the intercept and to calculate the error, you have to multiply by the standard error because that is, this is a transformation where we are taking exponential. So the correct answer is now 4.36 plus or minus 0.5 into 10 power 5.

Similarly, you have the fit coefficient 2 and that is Ea by r, so we have to multiply by the universal gas constant and that gives you 9.9 kilo joules and to calculate error in this quantity, we just have to take this value and multiply it by… So we have minus 9.9 plus or minus 0.9 kilo joules as the fitting parameter. So all this we did by using the formula linear model, so we fitted log k by T versus 1 by T.

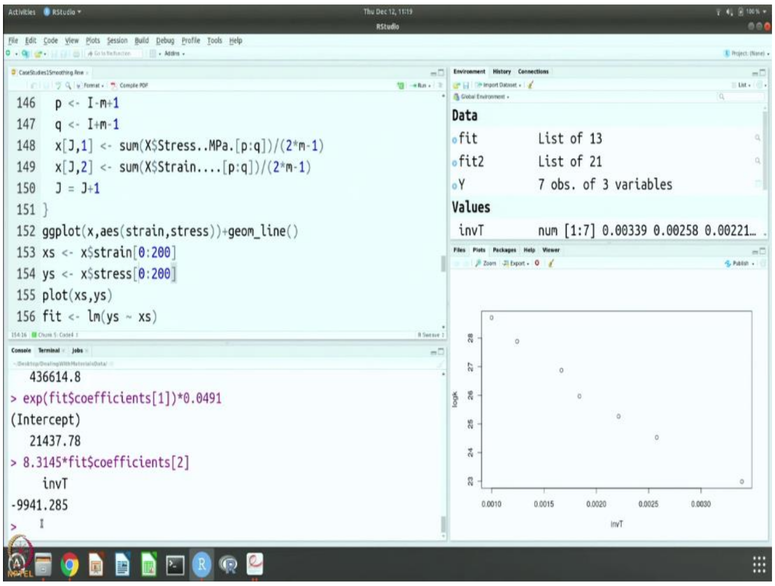
(Refer Slide Time: 8:12)



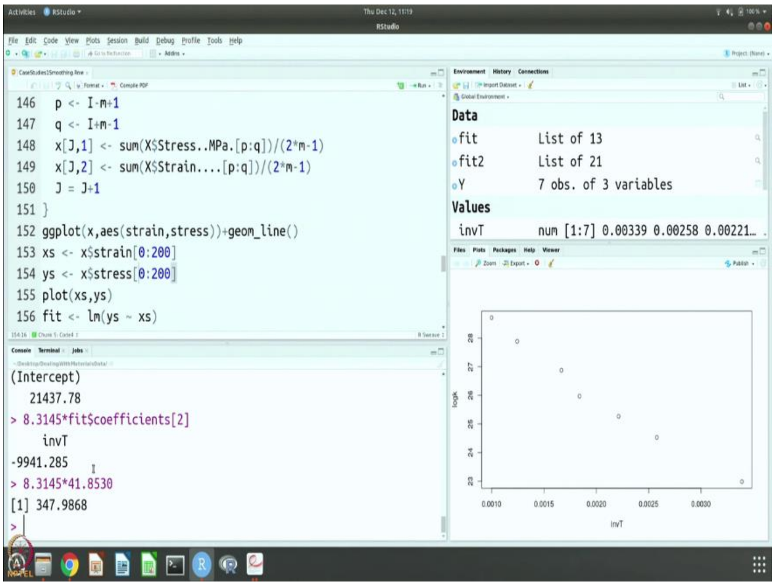
Module: Case studies
invT
-9941.285
library(MASS)
fit2 <- rlm(logkbyT ~ invT,weights=w)
summary(fit2)
##
Call: rlm(formula = logkbyT ~ invT, weights = w)
Residuals:
1 2 3 4 5 6 7
1391697 554920 -1415207 -11769472 12881113 -61893 -48745
##
Coefficients:
Value Std. Error t value
(Intercept) 12.9883 0.0491 264.4835
invT -1199.2462 41.8530 -28.6538
##
Residual standard error: 2063000 on 5 degrees of freedom
A <- exp(fit2$coefficients[1])



RStudio
146 p <- I-m+1
147 q <- I+m-1
148 x[J,1] <- sum(X$Stress..MPa.[p:q])/(2*m-1)
149 x[J,2] <- sum(X$Strain....[p:q])/(2*m-1)
150 J = J+1
151 }
152 ggplot(x,aes(strain,stress))+geom_line()
153 xs <- x$strain[0:200]
154 ys <- x$stress[0:200]
155 plot(xs,ys)
156 fit <- lm(ys ~ xs)
Residual standard error: 2063000 on 5 degrees of freedom
> exp(fit$coefficients[1])
(Intercept)
436614.8
> exp(fit$coefficients[1])*0.0491
(Intercept)
21437.78
>
Data
fit List of 13
fit2 List of 21
Y 7 obs. of 3 variables
Values
invT num [1:7] 0.00339 0.00258 0.00221…
logk
invT



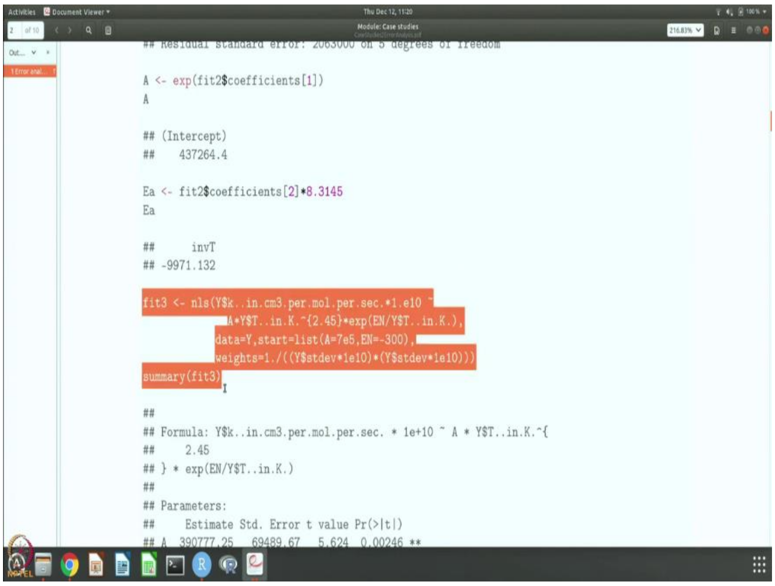
RStudio
146 p <- I-m+1
147 q <- I+m-1
148 x[J,1] <- sum(X$Stress..MPa.[p:q])/(2*m-1)
149 x[J,2] <- sum(X$Strain....[p:q])/(2*m-1)
150 J = J+1
151 }
152 ggplot(x,aes(strain,stress))+geom_line()
153 xs <- x$strain[0:200]
154 ys <- x$stress[0:200]
155 plot(xs,ys)
156 fit <- lm(ys ~ xs)
436614.8
> exp(fit$coefficients[1])*0.0491
(Intercept)
21437.78
> 8.3145*fit$coefficients[2]
invT
-9941.285
>
Data
fit List of 13
fit2 List of 21
Y 7 obs. of 3 variables
Values
invT num [1:7] 0.00339 0.00258 0.00221…
logk
invT



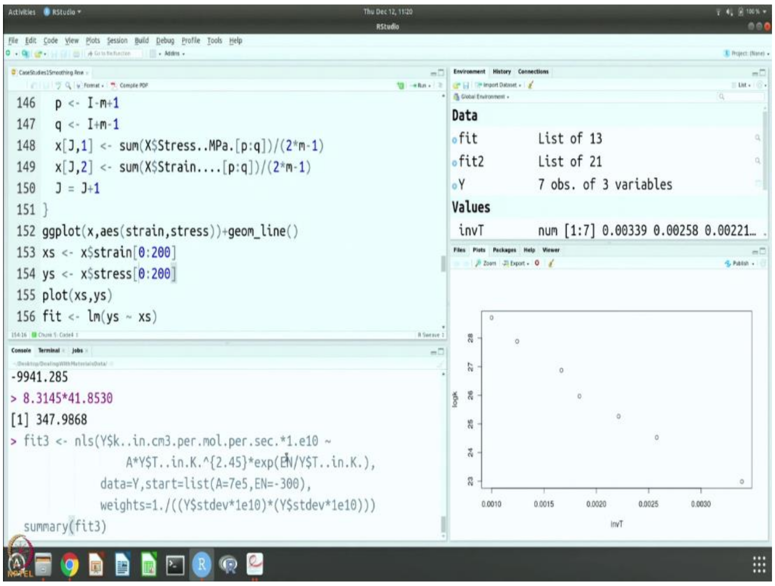
And of course there are other ways of fitting, so let us try some other ways of fitting also. So we will use the library mass and we will do robust linear model, right. So we have done the robust linear model and same formula we have fitted and using the same weights. Now you can see that this is the value. So it is 4.3 again and if you multiply this quantity by…you will get…so it is 4.3 plus or minus 0.2 into 10 to the power 5, right. So that is what we get.

And in this case again, you have to multiply by 8.3145 into fit dollar coefficients 2. So it gives 9.9 and if you want to calculate the error in this quantity, off course we have to multiply by the error because this is standard, so it is just audition, right. So it is 9.9 plus or minus 0.3 kilo joules per mole per second is the, kilo joules per mole per kelvin is the activation energy that you get from this fitting.

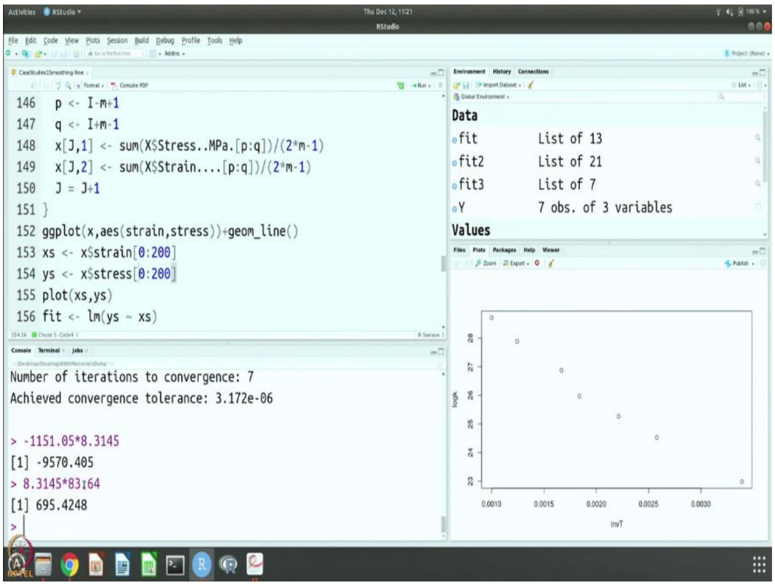
(Refer Slide Time: 10:13)



Residual standard error: 2063000 on 5 degrees of freedom
A <- exp(fit2$coefficients[1])
A
(Intercept)
437264.4
Ea <- fit2$coefficients[2]*8.3145
Ea
invT
-9971.132
fit3 <- nls(Y$k..in.cm3.per.mol.per.sec.*1.e10 ~
A*Y$T..in.K.^{2.45}*exp(EN/Y$T..in.K.),
data=Y,start=list(A=7e5,EN=-300),
weights=1./((Y$stdev*1e10)*(Y$stdev*1e10)))
summary(fit3)
##
Formula: Y$k..in.cm3.per.mol.per.sec. * 1e+10 ~ A * Y$T..in.K.^{
2.45
} * exp(EN/Y$T..in.K.)
##
Parameters:
Estimate Std. Error t value Pr(>|t|)
A 390777.25 69489.67 5.624 0.00246 **



146 p <- I-m+1
147 q <- I+m-1
148 x[J,1] <- sum(X$Stress..MPa.[p:q])/(2*m-1)
149 x[J,2] <- sum(X$Strain....[p:q])/(2*m-1)
150 J = J+1
151 }
152 ggplot(x,aes(strain,stress))+geom_line()
153 xs <- x$strain[0:200]
154 ys <- x$stress[0:200]
155 plot(xs,ys)
156 fit <- lm(ys ~ xs)
Data
fit List of 13
fit2 List of 21
Y 7 obs. of 3 variables
Values
invT num [1:7] 0.00339 0.00258 0.00221…
-9941.285
> 8.3145*41.8530
[1] 347.9868
> fit3 <- nls(Y$k..in.cm3.per.mol.per.sec.*1.e10 ~
A*Y$T..in.K.^{2.45}*exp(EN/Y$T..in.K.),
data=Y,start=list(A=7e5,EN=-300),
weights=1./((Y$stdev*1e10)*(Y$stdev*1e10)))
summary(fit3)
invT



146 p <- I-m+1
147 q <- I+m-1
148 x[J,1] <- sum(X$Stress..MPa.[p:q])/(2*m-1)
149 x[J,2] <- sum(X$Strain....[p:q])/(2*m-1)
150 J = J+1
151 }
152 ggplot(x,aes(strain,stress))+geom_line()
153 xs <- x$strain[0:200]
154 ys <- x$stress[0:200]
155 plot(xs,ys)
156 fit <- lm(ys ~ xs)
Data
fit List of 13
fit2 List of 21
fit3 List of 7
Y 7 obs. of 3 variables
Values
Number of iterations to convergence: 7
Achieved convergence tolerance: 3.172e-06
> -1151.05*8.3145
[1] -9570.405
> 8.3145*83.64
[1] 695.4248
>
invT

Of course you can do one more where we are going to do the non linear fitting. Let us do that and see how it is different from our… See in this case, we are going to do the fitting by just taking k and saying that it is A T to the power 2.45 in to exponential some activation energy which is scaled by the universal gas constant times temperature and the data is y and we have to give some initial values for the A values and the EN values.

And the weights are, of course, in this case we have not done any transformations, so we are just going to take the variant squared. So now you can see that we have fitted and you can see, in this case also it is 3.9 plus or minus 0.7 and here again you have to just multiply it by 8.3145 to get the values. So it is 9.5 and the error, so it is 9.5 plus or minus 0.7 kilo joules per mole per kelvin is the activation energy.

So the good news is that we are getting the same value of about 4, 4.3, 3.9 as the estimate for the A, the pre-exponential constant and we are getting the same 9.5 kilo joules for the 9.5, 9.9 and 9.9, so which is the activation energy we are getting. So to summarise, we have learnt about error propagation in the error analysis, so we know how the error should go and knowing that if a transformation is log or exponential, we know how the error should be calculated.

So when you do fitting and for fitting when you do transformation on the variables, when the error is given for those variables, they should also be accordingly transformed and in this case we were doing weighted fitting, by giving different weights to different points which is based on the variants of that particular data measurement, so when we do the transformation, we have to keep track of the transformation on the error also.

If you do that, then you get the fitting and once you had fitted, from the standard error that you get for the fitted parameters, you can also give the error in the quantity that you are trying to estimate and again you have to use the error propagation formulae that we have learned in terms of exponentials just adding and constants and things like that and once you do that you can give these parameters.

So when you have some experimental data and when you have done some fitting, when you want to report its always a good idea to also report the error and for doing that you can use the information that you get while fitting as the standard error and you can report it so that people have a clearer idea about the accuracy to which you are reporting your values that you have estimated.

So there is two more data sets that is available in from the same paper and we are going to share those data also in CSV format with you, so I strongly recommend that you do similar exercise and that is a far more easier exercise because there is no T to the power 2.45 in those cases, it is straight forward Arrhenius fit, k which is the reaction rate is some A times exponential minus Ea by RT.

So your task is to evaluate the A and the Ea. And you can do that by linearizing and fitting, but while linearizing, because you are taking a logarithm and because the error is given in k, you should also calculate the error accordingly and use that to weight when you are doing the fitting, so that is the part that you need to do and then you can get those parameters of fitting and you can compare with what is given in the paper.

And in this case of course what is given in the paper does not match with what we are getting, but consistently the different methods are giving values which are comparable, so I assume that what we are doing is okay, but of course you are welcome to dig deeper and convince yourself that what you are doing is okay. And if not, you should let us know. Thank you.