Draper HW #1

Matthew Draper

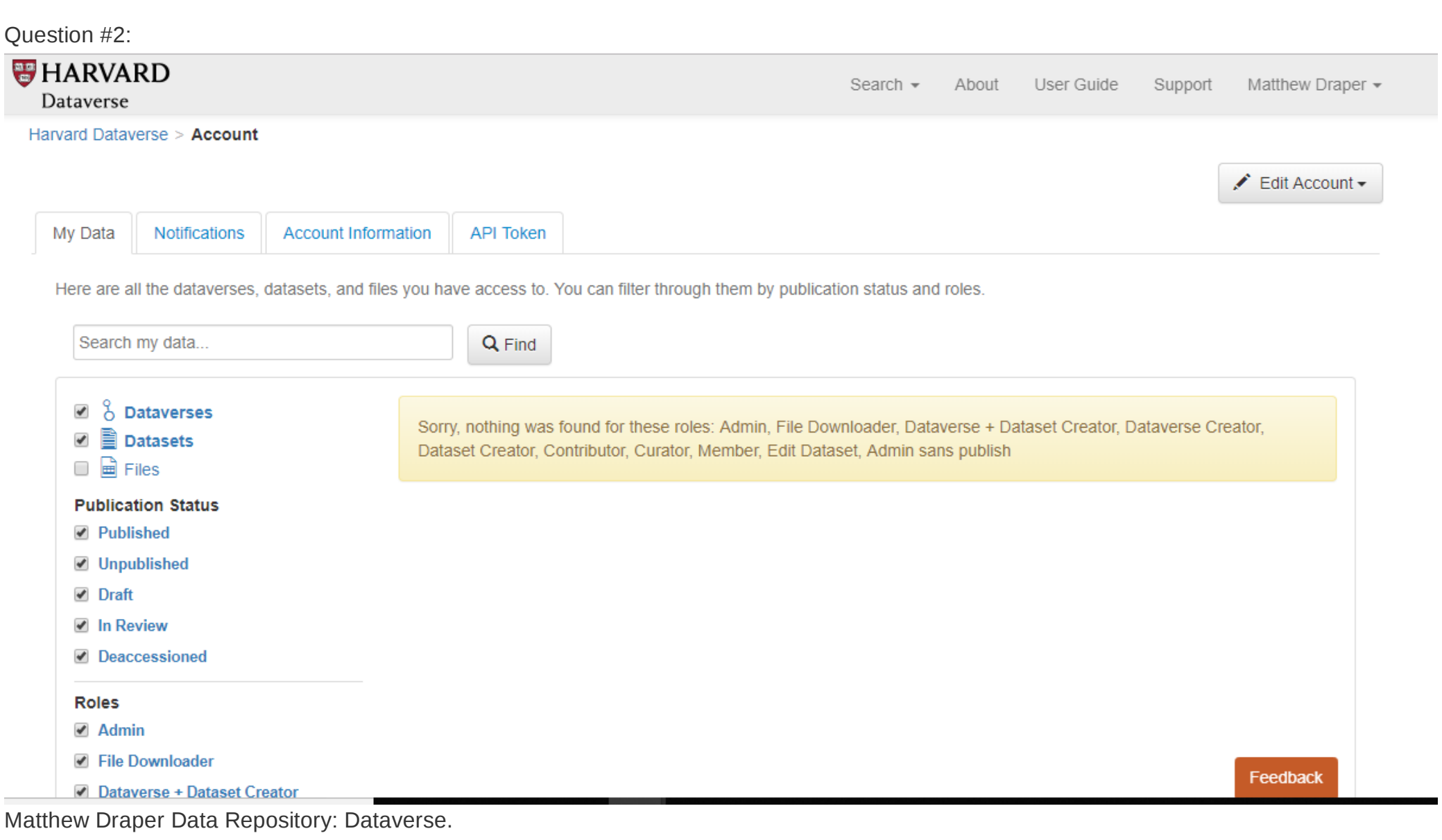
April 8, 2019

Question #1:

Shleifer, Andrei, and Robert W. Vishny. "Corruption." The Quarterly Journal of Economics, vol. 108, no. 3, 1993, pp. 599-617. JSTOR, www.jstor.org/stable/2118402.

Mauro, Paolo. "Corruption and Growth." The Quarterly Journal of Economics, vol. 110, no. 3, 1995, pp. 681-712. JSTOR, www.jstor.org/stable/2946696.

Question #2:



Here are all the dataverses, datasets, and files you have access to. You can filter through them by publication status and roles.

- ☑ Dataverses
- ☑ Datasets
- ☐ Files

Publication Status

- ☑ Published
- ☑ Unpublished
- ☑ Draft
- ☑ In Review
- ☑ Deaccessioned

Roles

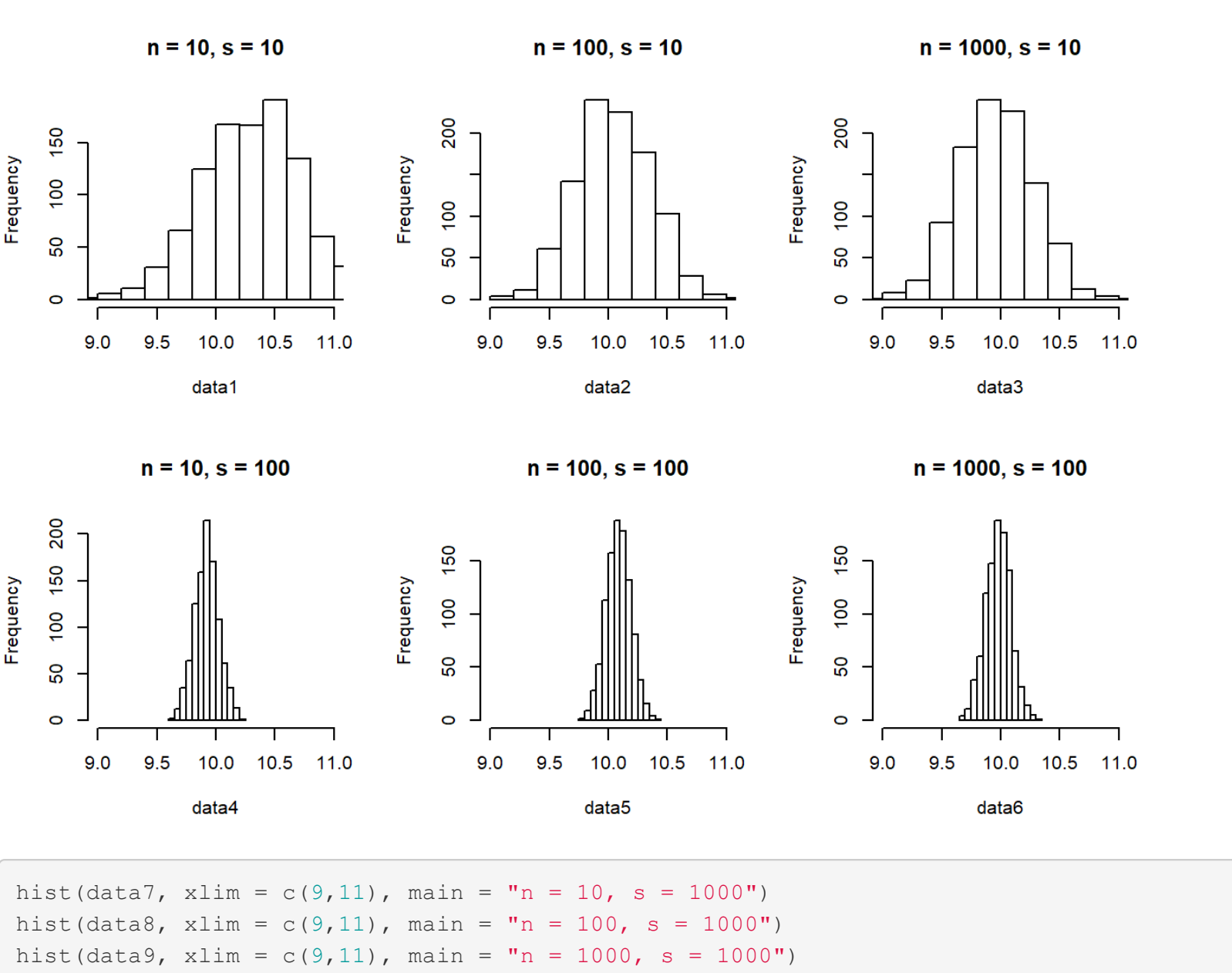
- ☑ Admin
- ☑ File Downloader
- ☑ Dataverse + Dataset Creator

Sorry, nothing was found for these roles: Admin, File Downloader, Dataverse + Dataset Creator, Dataverse Creator, Dataset Creator, Contributor, Curator, Member, Edit Dataset, Admin sans publish

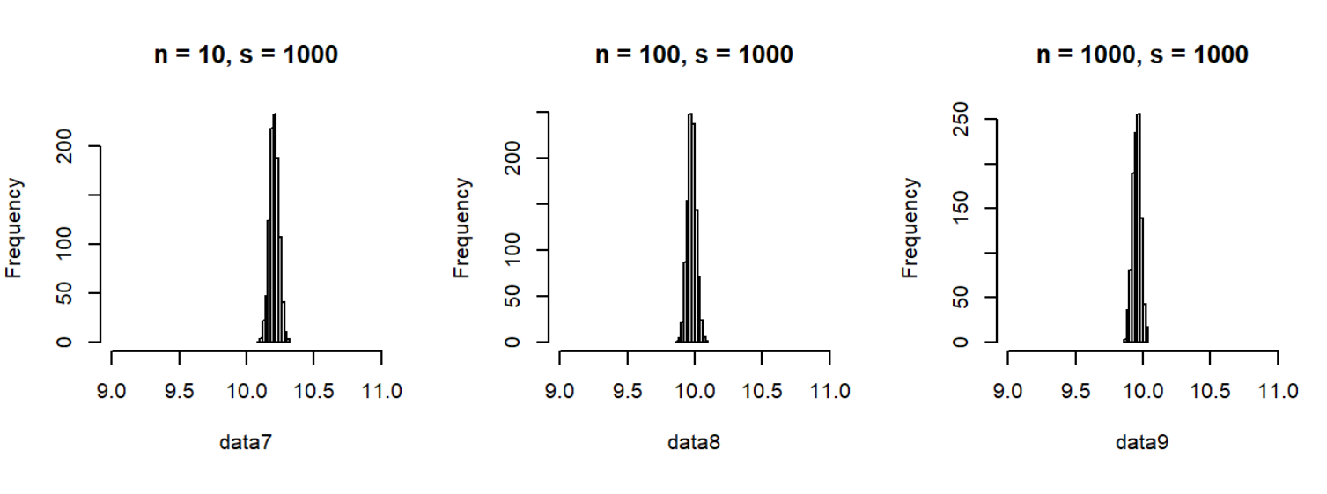
Matthew Draper Data Repository: Dataverse.

Question # 3:

```
## We wish to illustrate the Central Limit Theorem with a histogram of the means of 1,000 random samples of siz
e 10 drawn from a population with mean of 10 and variance of 1.

## n = 10, s = 10
data1 <- c()
n <- rnorm(10,10,1)
s <- 10
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data1[i] <- meandraw
}
## n = 100, s = 10
data2 <- c()
n <- rnorm(100,10,1)
s <- 10
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data2[i] <- meandraw
}
## n = 1000, s = 10
data3 <- c()
n <- rnorm(1000,10,1)
s <- 10
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data3[i] <- meandraw
}
## n = 10, s = 100
data4 <- c()
n <- rnorm(10,10,1)
s <- 100
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data4[i] <- meandraw
}
## n = 100, s = 100
data5 <- c()
n <- rnorm(100,10,1)
s <- 100
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data5[i] <- meandraw
}
## n = 1000, s = 100
data6 <- c()
n <- rnorm(1000,10,1)
s <- 100
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data6[i] <- meandraw
}
## n = 10, s = 1000
data7 <- c()
n <- rnorm(10,10,1)
s <- 1000
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data7[i] <- meandraw
}
## n = 100, s = 1000
data8 <- c()
n <- rnorm(100,10,1)
s <- 1000
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data8[i] <- meandraw
}
## n = 1000, s = 1000
data9 <- c()
n <- rnorm(1000,10,1)
s <- 1000
for(i in 1:1000){
  draw <- sample(n, s, replace = TRUE, prob = NULL)
  meandraw <- mean(draw)
  data9[i] <- meandraw
}


par(mfrow=c(2,3))
hist(data1, xlim = c(9,11), main = "n = 10, s = 10")
hist(data2, xlim = c(9,11), main = "n = 100, s = 10")
hist(data3, xlim = c(9,11), main = "n = 1000, s = 10")
hist(data4, xlim = c(9,11), main = "n = 10, s = 100")
hist(data5, xlim = c(9,11), main = "n = 100, s = 100")
hist(data6, xlim = c(9,11), main = "n = 1000, s = 100")
```



```
hist(data7, xlim = c(9,11), main = "n = 10, s = 1000")
hist(data8, xlim = c(9,11), main = "n = 100, s = 1000")
hist(data9, xlim = c(9,11), main = "n = 1000, s = 1000")
```



```
## Note: this code returns the error "figure margins too large"
par(mfrow=c(3,3))
hist(data1, xlim = c(9,11), main = "n = 10, s = 10")
hist(data2, xlim = c(9,11), main = "n = 100, s = 10")
hist(data3, xlim = c(9,11), main = "n = 1000, s = 10")
hist(data4, xlim = c(9,11), main = "n = 10, s = 100")
hist(data5, xlim = c(9,11), main = "n = 100, s = 100")
hist(data6, xlim = c(9,11), main = "n = 1000, s = 100")
hist(data7, xlim = c(9,11), main = "n = 10, s = 1000")
hist(data8, xlim = c(9,11), main = "n = 100, s = 1000")
hist(data9, xlim = c(9,11), main = "n = 1000, s = 1000")
```

Question #4:

```
library(foreign)
cox<-read.dta("coxappend.dta")

x1 <- cox$eneth
x2 <- cox$ml
y <- cox$enps
z <- c(rep(1,54))

## Calculation of XprimeX matrix
X <- cbind(z,x1,x2)
tX <- t(X)
tXX <- tX%*%X
itXX <- solve(tXX)

## Calculation of XprimeY matrix
tXY <- tX%*%y

## Calculation of betahat matrix
betahat <- itXX%*%tXY
print("Beta Hat Matrix")
```

```
## [1] "Beta Hat Matrix"
```

```
betahat
```

```
##          [,1]
## z  2.30999406
## x1 0.28205730
## x2 0.01460861
```

```
## Interpretation: z = constant

## Calculation of standard errors
yhat <- X%*%betahat
e <- y-yhat
esq <- e^2
RSS <- sum(esq)
vcov <- itXX*(RSS/(54-3-1))
print("Variance-Covariance Matrix")
```

```
## [1] "Variance-Covariance Matrix"
```

```
vcov
```

```
##               z            x1            x2
## z   0.260333888 -0.1331912304 -1.101199e-03
## x1 -0.133191230  0.0828370950  2.405192e-04
## x2 -0.001101199  0.0002405192  5.867236e-05
```

```
rawerror<-rbind(vcov[1,1],vcov[2,2],vcov[3,3])
stderror<-sqrt(rawerror)
regtable<-cbind(betahat,stderror)
print("Estimate and Standard Errors")
```

```
## [1] "Estimate and Standard Errors"
```

```
regtable
```

```
##          [,1]        [,2]
## z  2.30999406 0.510229250
## x1 0.28205730 0.287814341
## x2 0.01460861 0.007659788
```

```
## Up to this point, I have reproduced the model without the interaction term (this is all we did in 204B). I'm
 not completely sure how to incorporate the interaction term, so I wanted to verify this part of the regression
 first. As you can see, the results are the same:

reg1<-(lm(enps ~ eneth + ml, data = cox))
summary (reg1)
```

```
##
## Call:
## lm(formula = enps ~ eneth + ml, data = cox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9182 -0.7541 -0.4245  0.5149  5.3157
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.309994   0.505202   4.572  3.1e-05 ***
## eneth       0.282057   0.284979   0.990   0.3270
## ml          0.014609   0.007584   1.926   0.0597 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.407 on 51 degrees of freedom
## Multiple R-squared:  0.07818,    Adjusted R-squared:  0.04203
## F-statistic: 2.163 on 2 and 51 DF,  p-value: 0.1255
```

```
vcov(reg1)
```

```
##              (Intercept)         eneth            ml
## (Intercept)  0.255229302 -0.1305796377 -1.079607e-03
## eneth       -0.130579638  0.0812128382  2.358032e-04
## ml          -0.001079607  0.0002358032  5.752192e-05
```

```
## Now I'll attempt to include the interaction term. My hypothesis is that it will be purely multiplicative:

x1 <-cox$eneth
x2 <-cox$ml
x3 <-(cox$eneth*cox$ml)
y <- cox$enps
z <- c(rep(1,54))

## Calculation of XprimeX matrix
X<-cbind(z,x1,x2,x3)
tX<-t(X)
tXX<-tX%*%X
itXX<-solve(tXX)

## Calculation of XprimeY matrix
tXY<-tX%*%y

## Calculation of betahat matrix
betahat<-itXX%*%tXY
print("Beta Hat Matrix")
```

```
## [1] "Beta Hat Matrix"
```

```
betahat
```

```
##           [,1]
## z   3.22123292
## x1 -0.42343000
## x2 -0.09311689
## x3  0.08715612
```

```
## Interpretation: z = constant

## Calculation of standard errors
yhat <- X%*%betahat
e <- y-yhat
esq <- e^2
RSS <- sum(esq)
vcov <- itXX*(RSS/(54-3-1))
print("Variance-Covariance Matrix")
```

```
## [1] "Variance-Covariance Matrix"
```

```
vcov
```

```
##               z           x1            x2            x3
## z   0.276294130 -0.159059370 -0.0088533135  0.0064479995
## x1 -0.159059370  0.106872832  0.0063632281 -0.0049920850
## x2 -0.008853313  0.006363228  0.0009892514 -0.0007622743
## x3  0.006448000 -0.004992085 -0.0007622743  0.0006167237
```

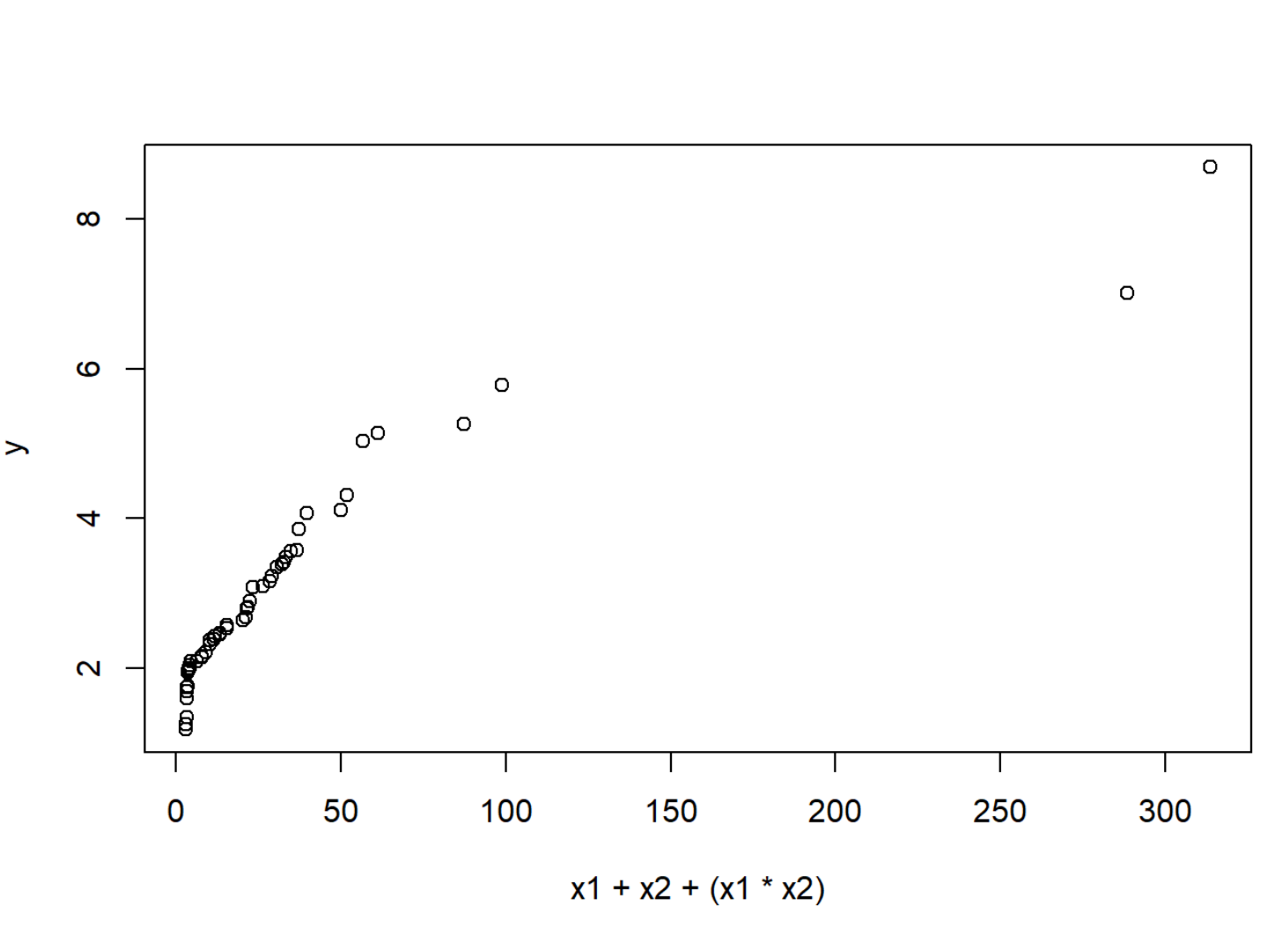
```
rawerror<-rbind(vcov[1,1],vcov[2,2],vcov[3,3],vcov[4,4])
stderror<-sqrt(rawerror)
regtable<-cbind(betahat,stderror)
print("Estimate and Standard Errors")
```

```
## [1] "Estimate and Standard Errors"
```

```
regtable
```

```
##           [,1]       [,2]
## z   3.22123292 0.52563688
## x1 -0.42343000 0.32691411
## x2 -0.09311689 0.03145237
## x3  0.08715612 0.02483392
```

```
qqplot(x1+x2+(x1*x2),y)
```



```
## We observe curvature in the qq plot, so I haven't added the qq line.


## My hypothesis was correct: the interaction is purely multiplicative. We obtain the same results using the lm
command:

reg2<-lm(enps ~ eneth + ml + (eneth:ml),data=cox)
summary(reg2)
```

```
##
## Call:
## lm(formula = enps ~ eneth + ml + (eneth:ml), data = cox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1665 -0.7586 -0.3471  0.4543  3.8924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.22123    0.52564   6.128 1.38e-07 ***
## eneth       -0.42343    0.32691  -1.295  0.20119
## ml          -0.09312    0.03145  -2.961  0.00469 **
## eneth:ml     0.08716    0.02483   3.510  0.00096 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.273 on 50 degrees of freedom
## Multiple R-squared:  0.2604, Adjusted R-squared:  0.216
## F-statistic: 5.867 on 3 and 50 DF,  p-value: 0.001633
```

```
vcov(reg2)
```

```
##              (Intercept)        eneth            ml      eneth:ml
## (Intercept)  0.276294130 -0.159059370 -0.0088533135  0.0064479995
## eneth       -0.159059370  0.106872832  0.0063632281 -0.0049920850
## ml          -0.008853313  0.006363228  0.0009892514 -0.0007622743
## eneth:ml     0.006448000 -0.004992085 -0.0007622743  0.0006167237
```

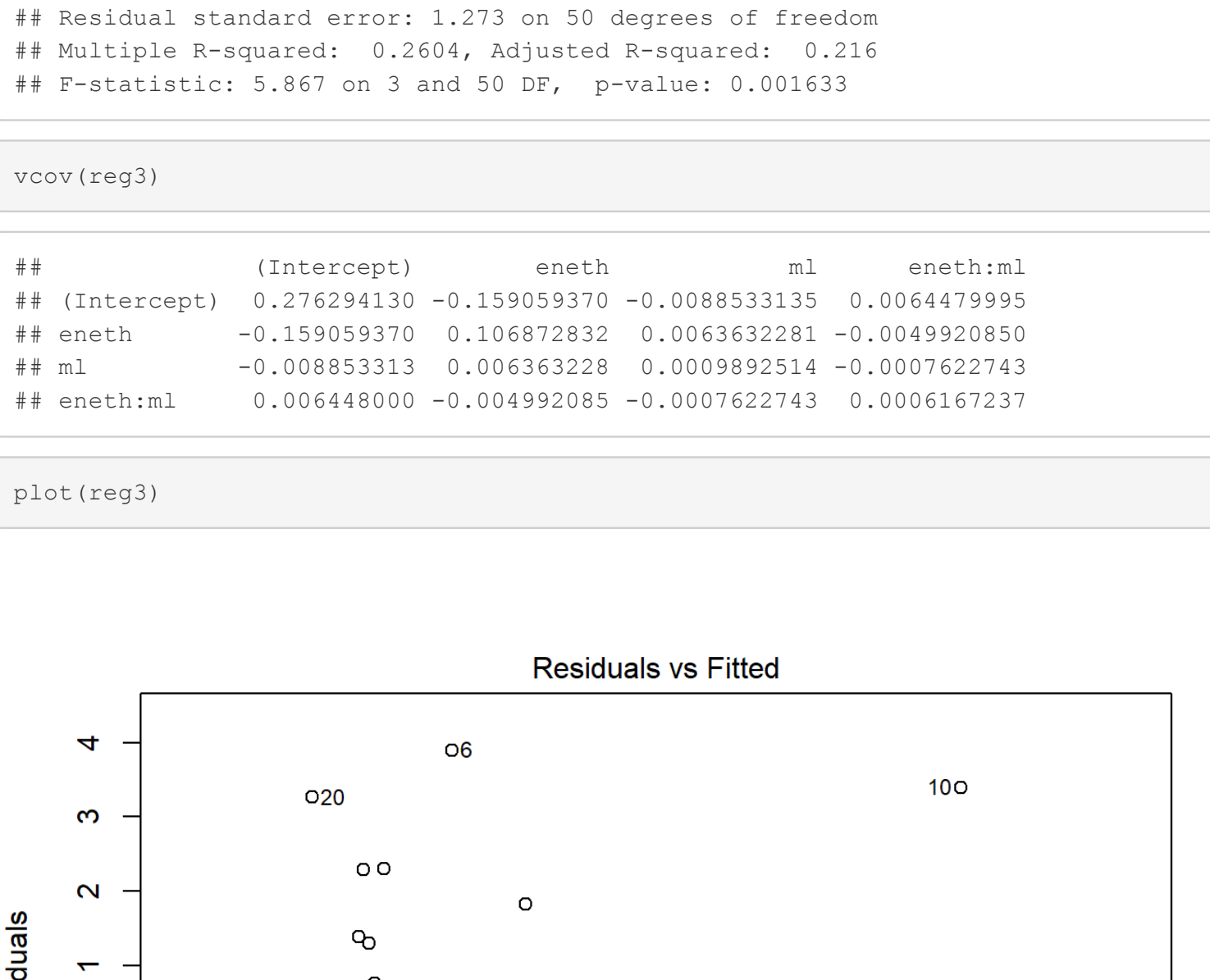
```
## Note: there are two possible ways to enter the interaction

reg3<-lm(enps ~ eneth * ml, data = cox)
summary (reg3)
```

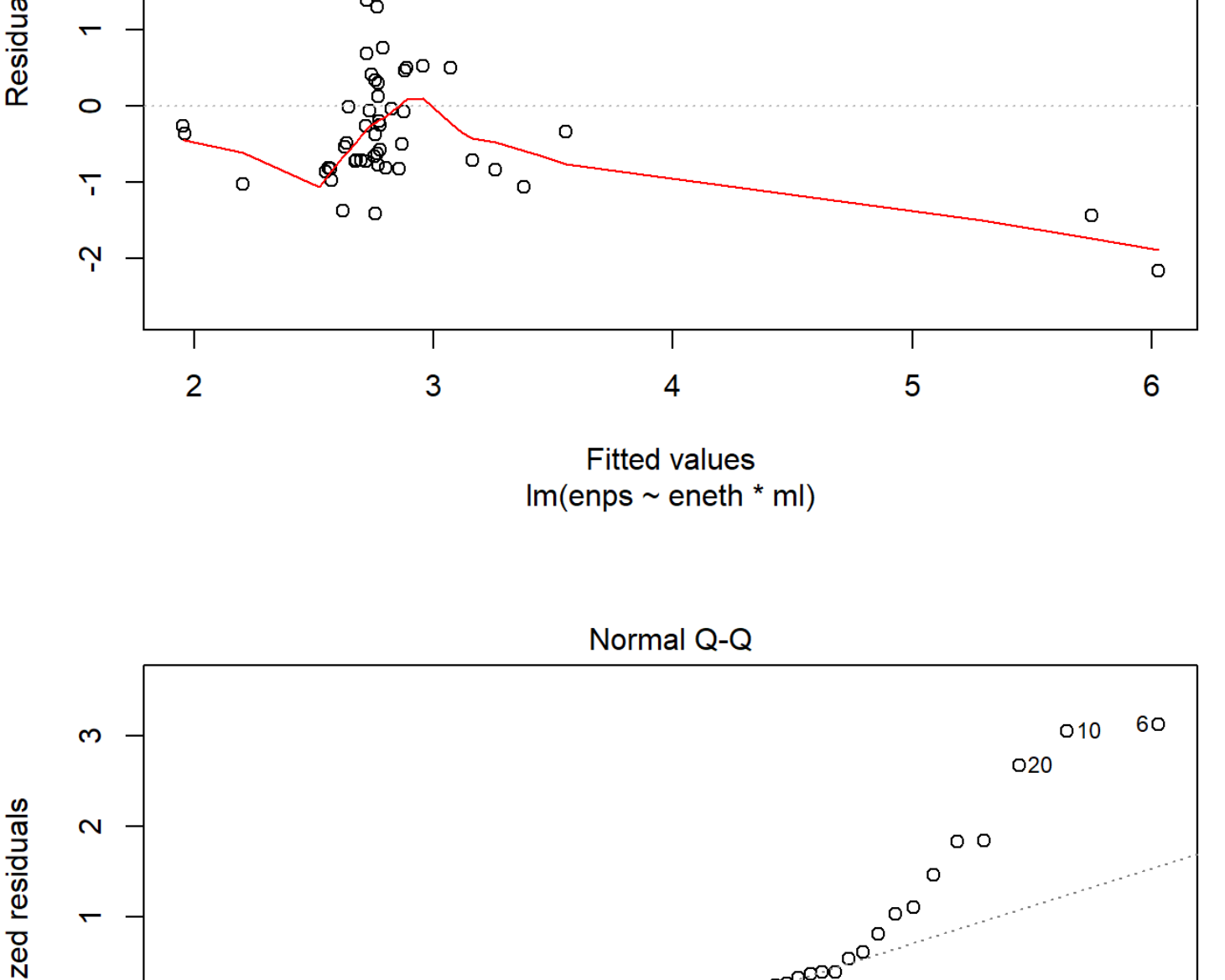
```
##
## Call:
## lm(formula = enps ~ eneth * ml, data = cox)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1665 -0.7586 -0.3471  0.4543  3.8924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.22123    0.52564   6.128 1.38e-07 ***
## eneth       -0.42343    0.32691  -1.295  0.20119
## ml          -0.09312    0.03145  -2.961  0.00469 **
## eneth:ml     0.08716    0.02483   3.510  0.00096 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.273 on 50 degrees of freedom
## Multiple R-squared:  0.2604, Adjusted R-squared:  0.216
## F-statistic: 5.867 on 3 and 50 DF,  p-value: 0.001633
```

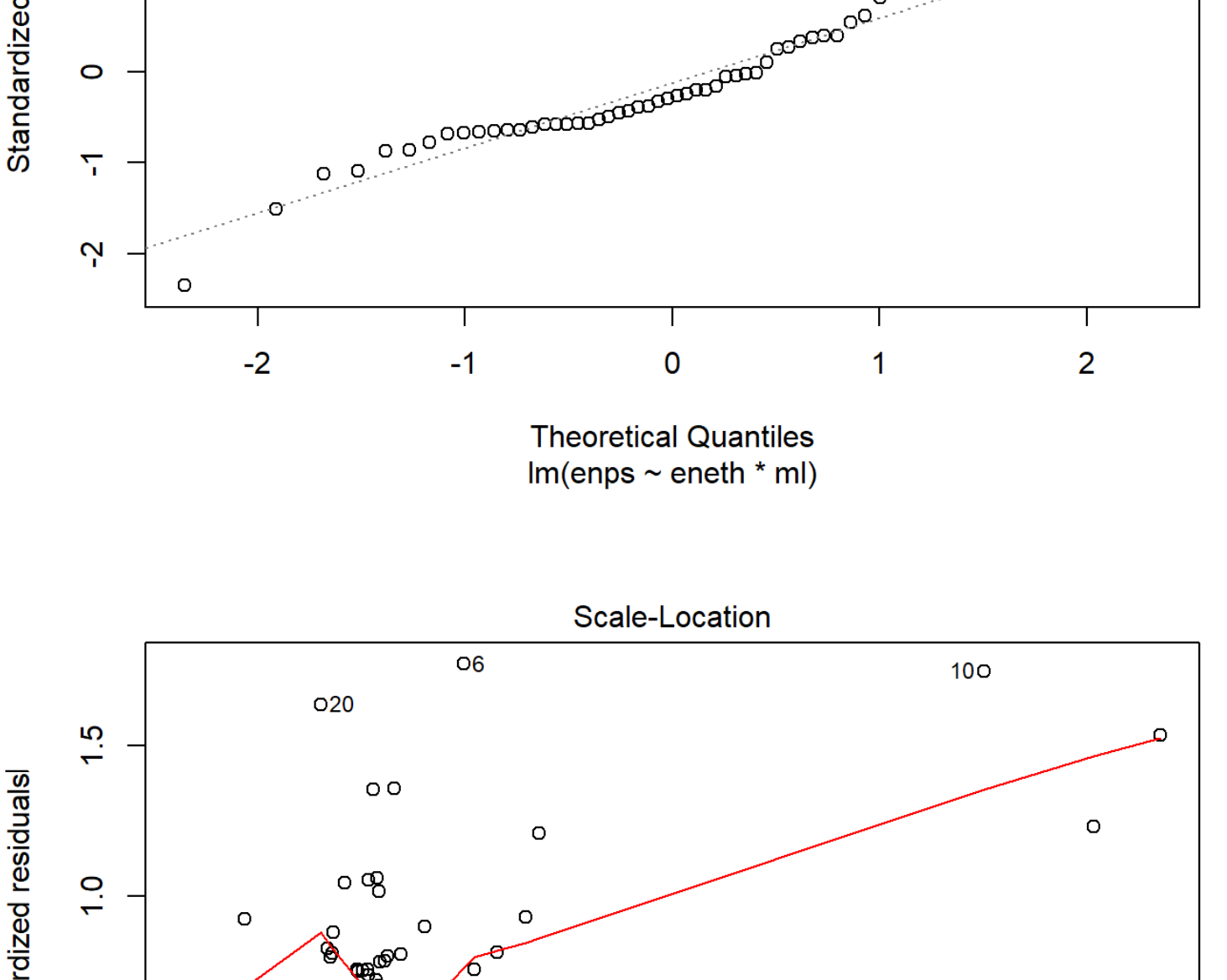
```
vcov(reg3)
```

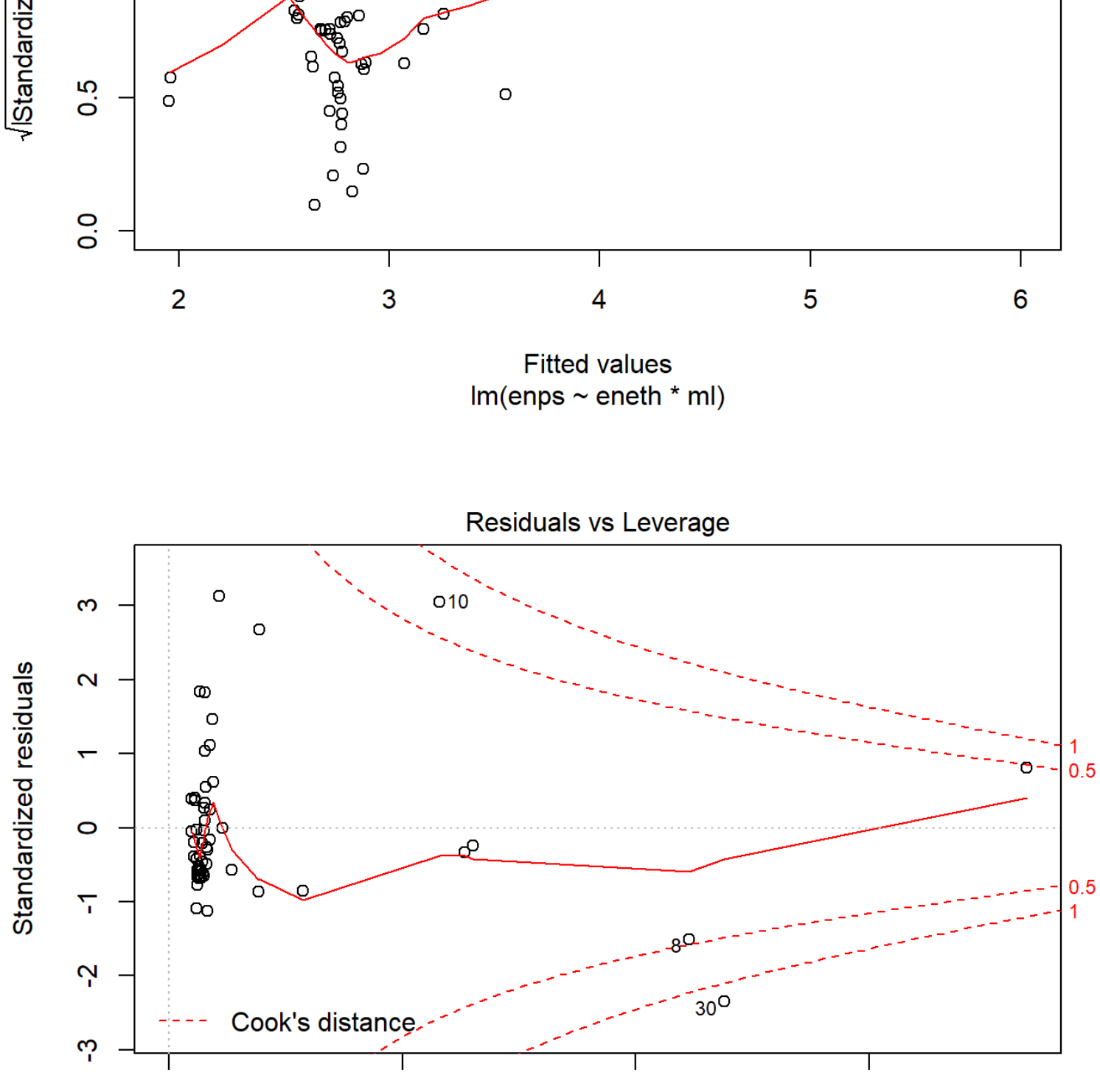
```
##              (Intercept)        eneth            ml      eneth:ml
## (Intercept)  0.276294130 -0.159059370 -0.0088533135  0.0064479995
## eneth       -0.159059370  0.106872832  0.0063632281 -0.0049920850
## ml          -0.008853313  0.006363228  0.0009892514 -0.0007622743
## eneth:ml     0.006448000 -0.004992085 -0.0007622743  0.0006167237
```

```
plot(reg3)
```









```
## In the residuals vs. fitted plot, we observe either curvature or discontinuous linearity, with the strongest
curvature in the region where we have the most data. We also see a wedge emerging in the data as we move from l
eft to right. This indicates that there may be a nonlinear relationship that was not explained by the model (an
d is therefore showing up in our residutals). The scale location plot exhibits similar discontinuity and nonlin
earity, meaning that residuals are not evenly spread along our range of predictors. The equal variance (homoske
dasticity) assumption appears to be violated. There is also significant clustering around the fitted value of
 3. Finally, the residuals vs. leverage plot indicates that outlying cases may be driving our results. We see s
everal observations (#10, #30) beyond the Cook's Distance line, indicating that our results would change if the
se observations were omitted. All of this indicates that a linear model may not be the best way to understand t
hese data. Alternatively, omitted variables or collection bias could be skewing the results. At a minimum, we w
ould want to check the results without the influential outliers and further investigate the discontinuity and n
onlinearity in the residual plots.
```