Math 4780 - Homework 3

Liam Fruzyna

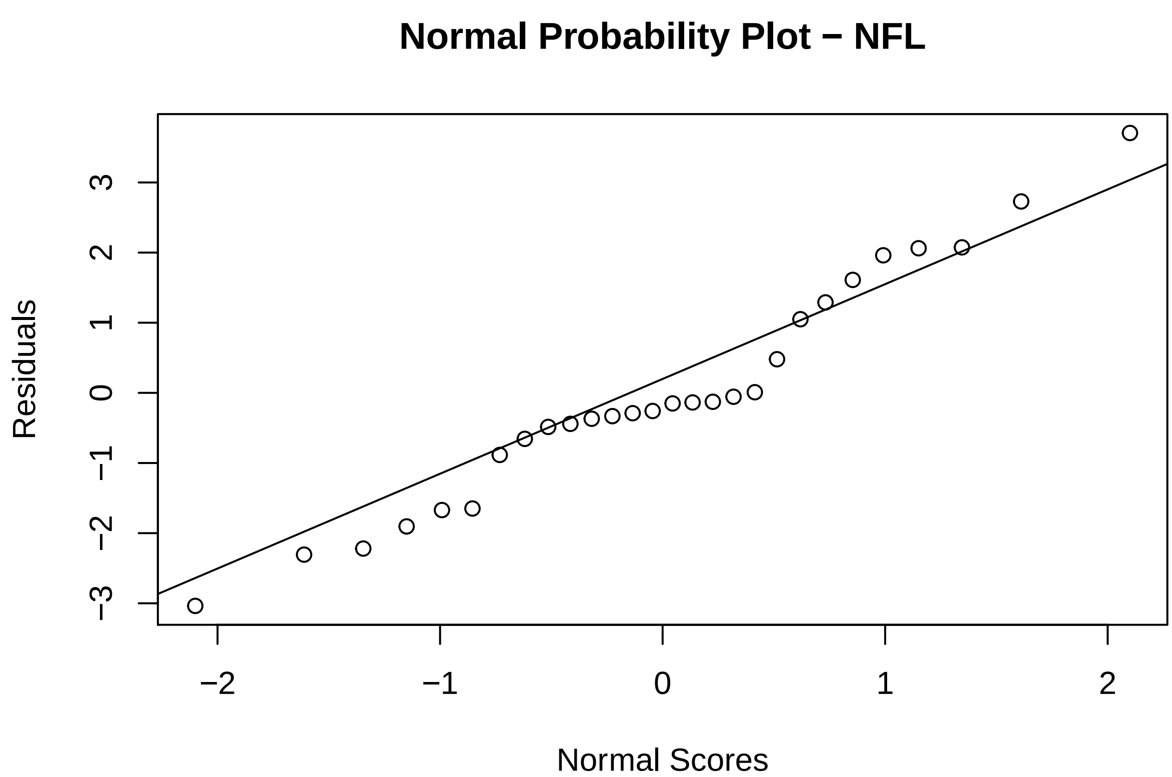
#4.2 Consider the previous problem 3.1

Previous work

```
library(MPV)
nfl = table.b1
model <- lm(y ~ x2 + x7 + x8, data=nfl)
```

a) Construct a normal probability plot of the residuals. Is there any problem with the normality assumption?

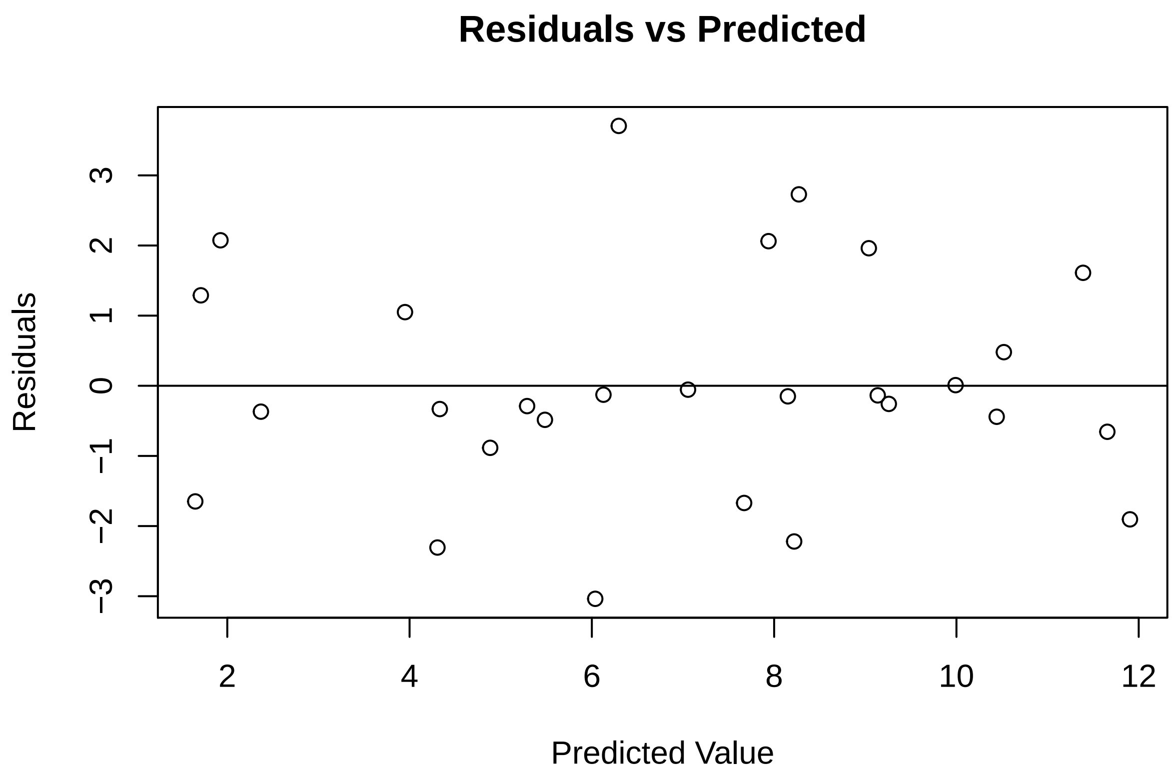
```
residuals = model$residuals
qqnorm(residuals, ylab='Residuals', xlab='Normal Scores', main='Normal Probability Plot - NFL')
qqline(residuals)
```



The plot has a slight ploblem with normality.

b) Construct and interpret a plot of the residuals versus the predicted result

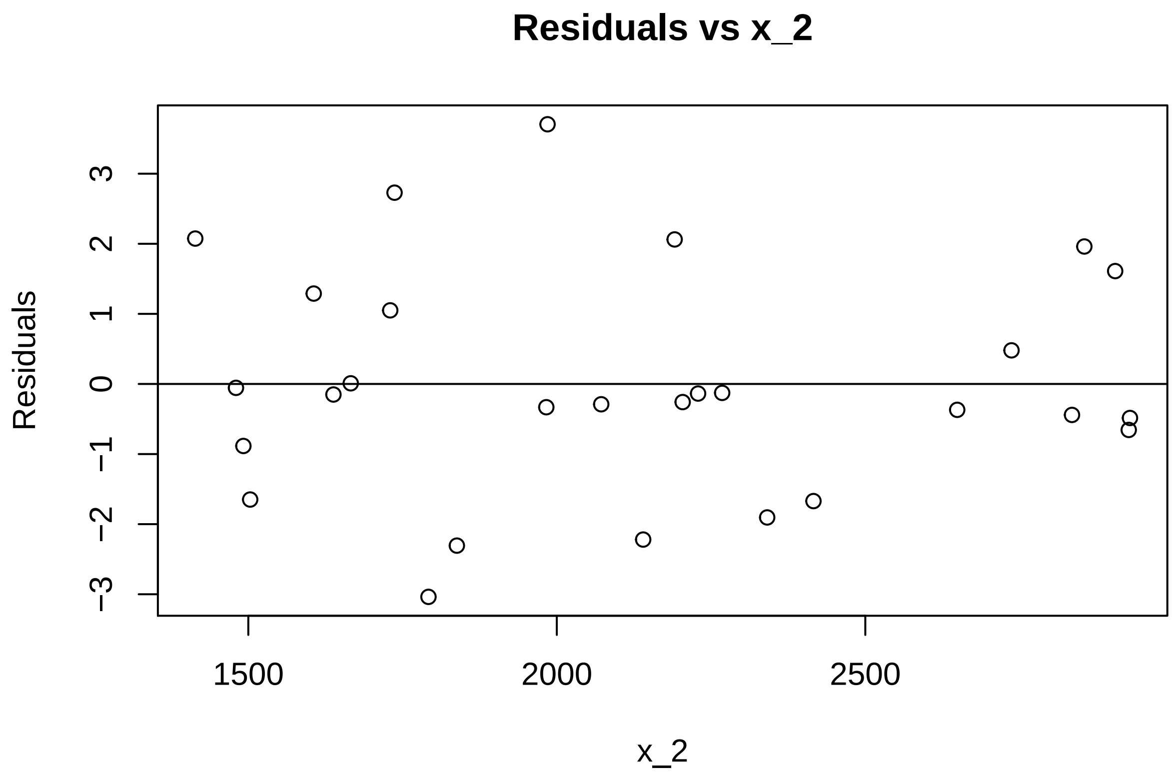
```r
predicted = predict(model)
plot(predicted, residuals, xlab='Predicted Value', ylab='Residuals', main='Residuals vs Predicted')
abline(0,0)
```



This plot is evenly distributed. It is good.

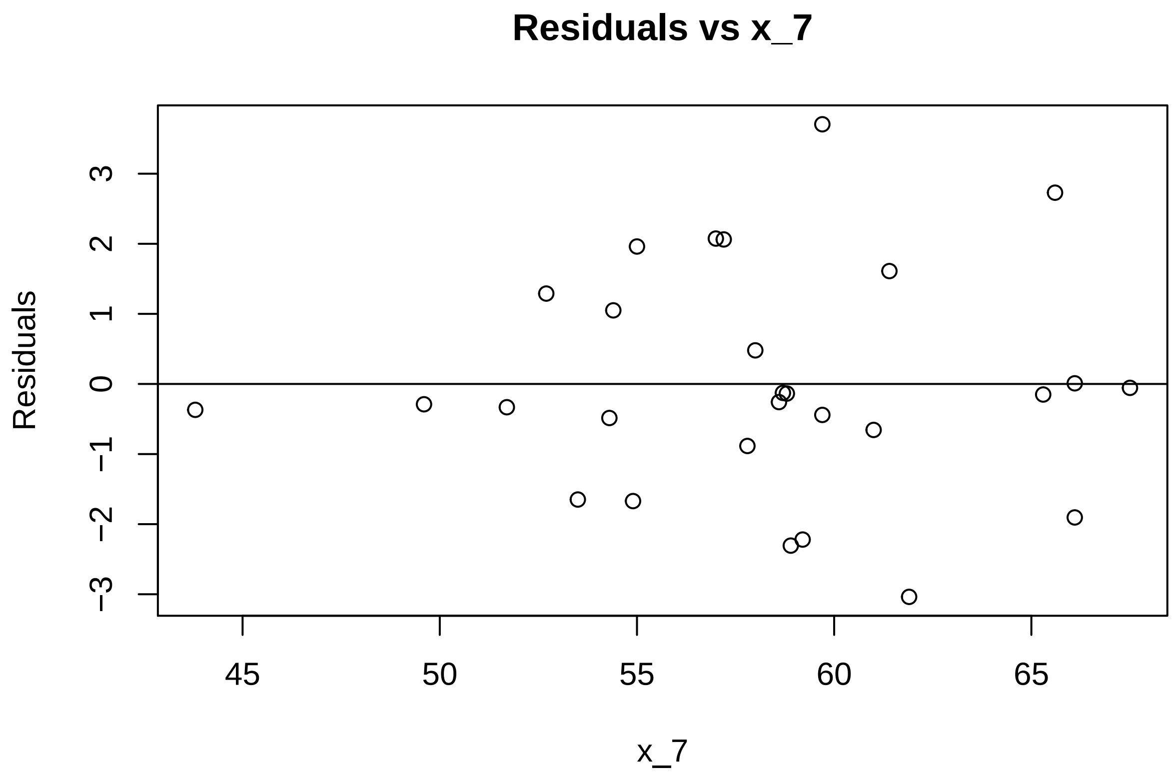
c) Construct plots of the residuals versus each of the regressors. Do these plots imply the regressors is specified correctly?

```r
plot(nfl$x2, residuals, xlab='x_2', ylab='Residuals', main='Residuals vs x_2')
abline(0,0)
```



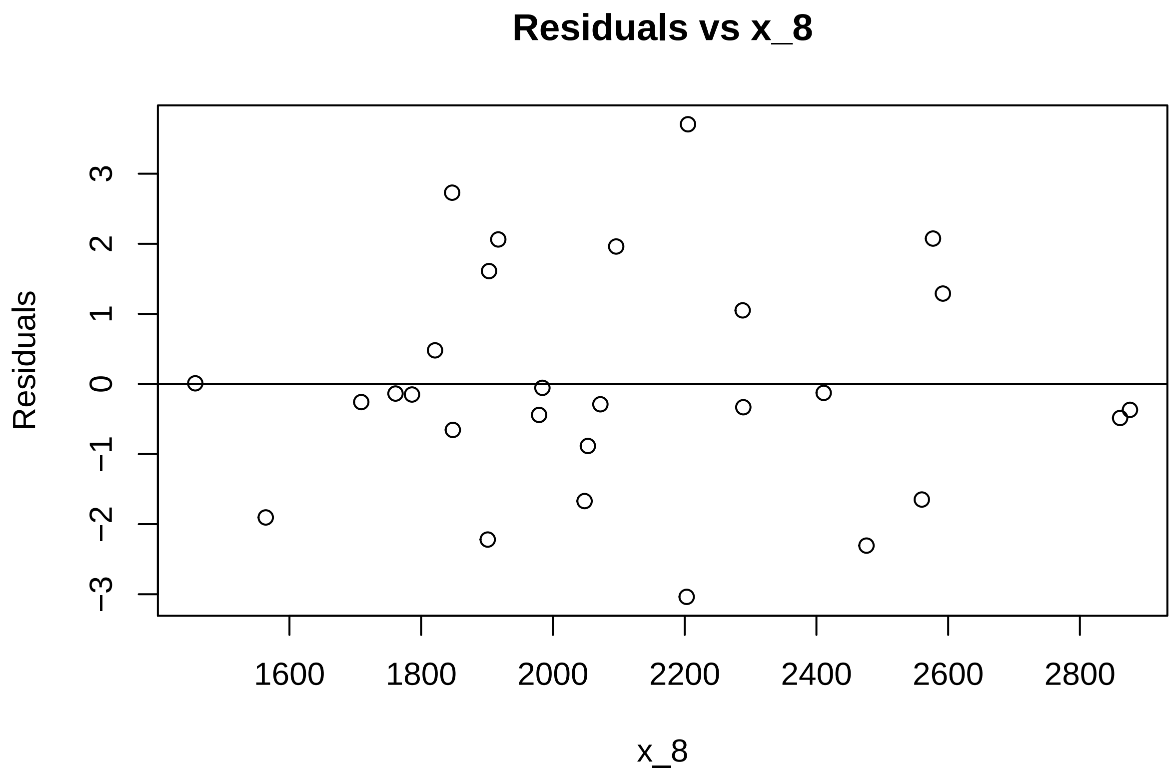
This plot is not even enough.

```
plot(nfl$x7, residuals, xlab='x_7', ylab='Residuals', main='Residuals vs x_7')
abline(0,0)
```



This plot definitely has nonconstant variance.

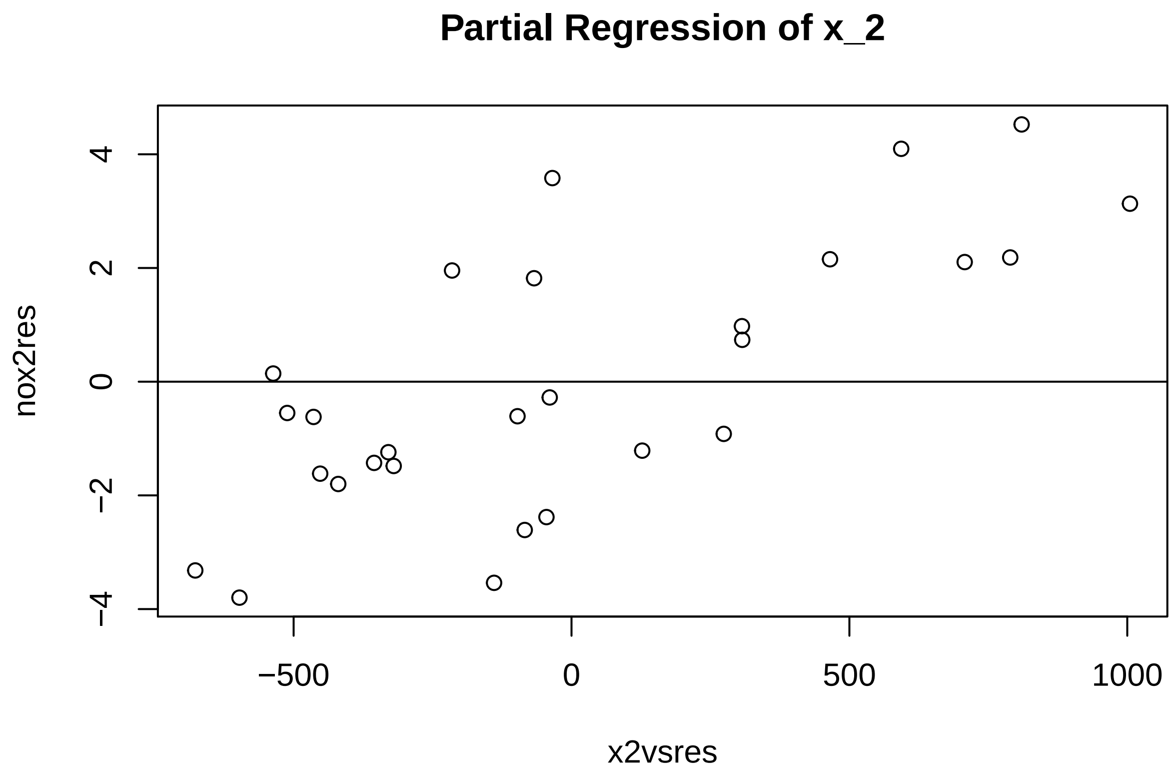
```
plot(nfl$x8, residuals, xlab='x_8', ylab='Residuals', main='Residuals vs x_8')
abline(0,0)
```



This plot is pretty good.

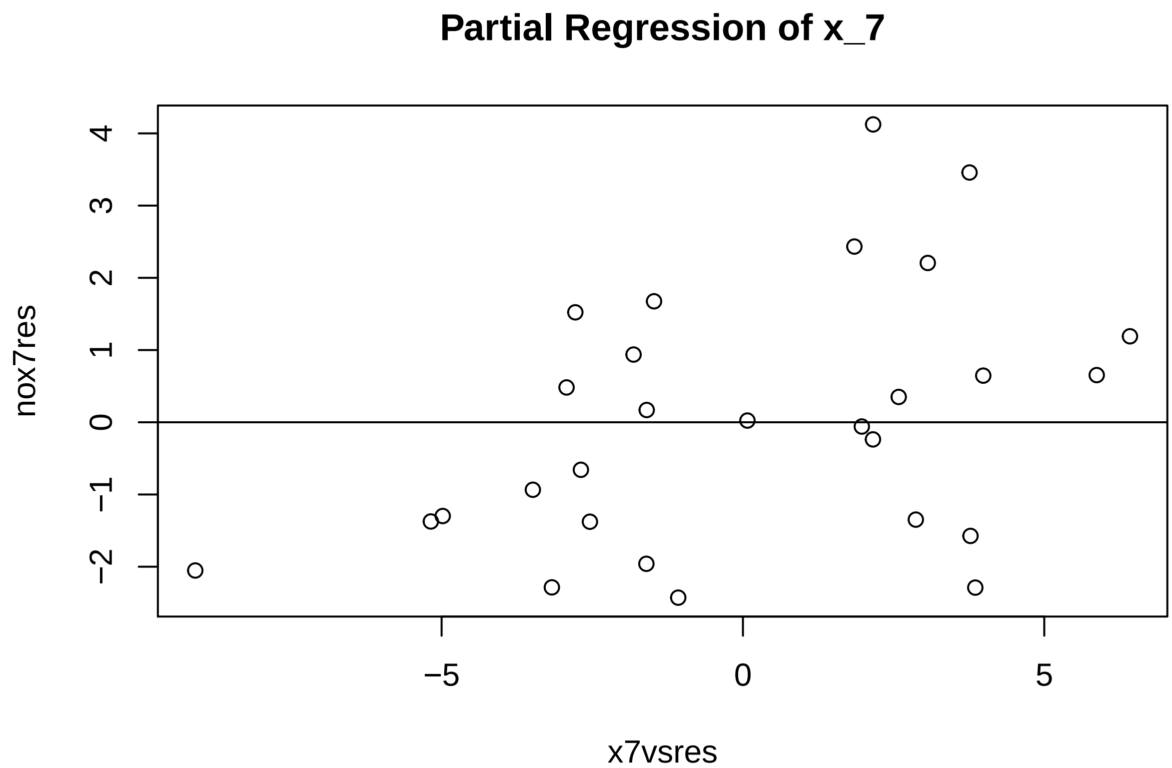
d) Construct the partial regression plots for the model. Compare the plots to those of part c. Discuss the types of information the plots provide.

```
nox2 <- lm(y ~ x7 + x8, data=nfl)
nox2res <- nox2$residuals
x2vs <- lm(x2 ~ x7 + x8, data=nfl)
x2vsres <- x2vs$residuals
plot(x2vsres, nox2res, main='Partial Regression of x_2')
abline(0,0)
```



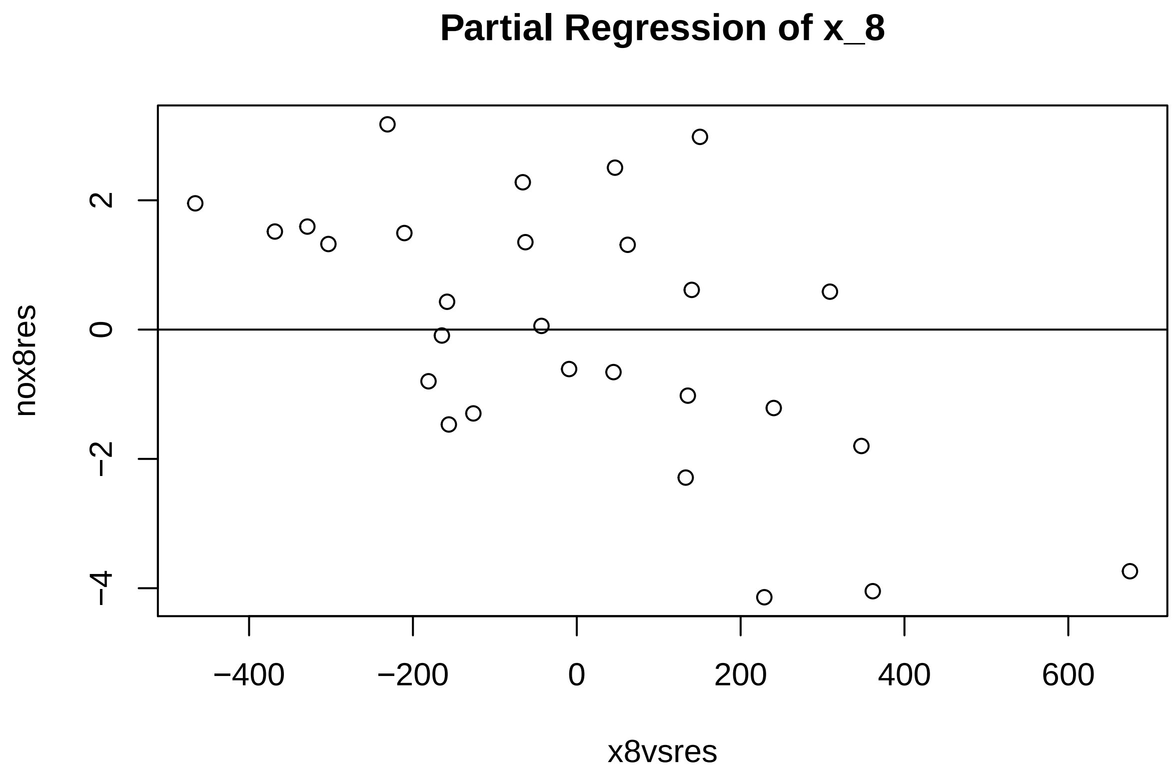
x_2 has a good linear relationship.

```
nox7 <- lm(y ~ x2 + x8, data=nfl)
nox7res <- nox7$residuals
x7vs <- lm(x7 ~ x2 + x8, data=nfl)
x7vsres <- x7vs$residuals
plot(x7vsres, nox7res, main='Partial Regression of x_7')
abline(0,0)
```



x_7 does not have a strong linear relationship.

```
nox8 <- lm(y ~ x7 + x2, data=nfl)
nox8res <- nox8$residuals
x8vs <- lm(x8 ~ x7 + x2, data=nfl)
x8vsres <- x8vs$residuals
plot(x8vsres, nox8res, main='Partial Regression of x_8')
abline(0,0)
```



x_8 has a good linear relationship.

e) Compute the studentized residuals and the R-student residuals for the model. What information is converyed by these scaled residuals?

```
rstandard(model)
```

```
##            1            2            3            4            5
##  2.231851618  1.225616368  1.702625305  1.029767789  0.006124483
##            6            7            8            9           10
## -0.418876221 -1.206836995  0.299328499  1.338032316 -1.441760607
##           11           12           13           14           15
## -0.036468456  1.251090093 -0.083851688 -0.160668820 -1.335367350
##           16           17           18           19           20
##  0.644990078 -0.196937383 -0.365011749 -0.078998342 -0.206464327
##           21           22           23           24           25
## -1.869940122  0.817274105 -0.551056514 -0.276544687 -1.018586104
##           26           27           28
## -0.094055761 -0.262130195 -1.048746774
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:MPV':
##
##     cement
```

```r
studres(model)
```

```
##            1            2            3            4            5 
##  2.454354223  1.239218310  1.777586702  1.031123075  0.005995537 
##            6            7            8            9           10 
## -0.411563960 -1.218993620  0.293574644  1.361631132 -1.476806719 
##           11           12           13           14           15 
## -0.035701602  1.266752172 -0.082098218 -0.157370596 -1.358701256 
##           16           17           18           19           20 
##  0.636954384 -0.192946834 -0.358322410 -0.077345090 -0.202296957 
##           21           22           23           24           25 
## -1.980521136  0.811437522 -0.542899513 -0.271154408 -1.019417881 
##           26           27           28 
## -0.092092392 -0.256979177 -1.051031132
```

These scaled residuals help find important points as well as outliers. The fist point may be an outliear as it has a relatively high values.

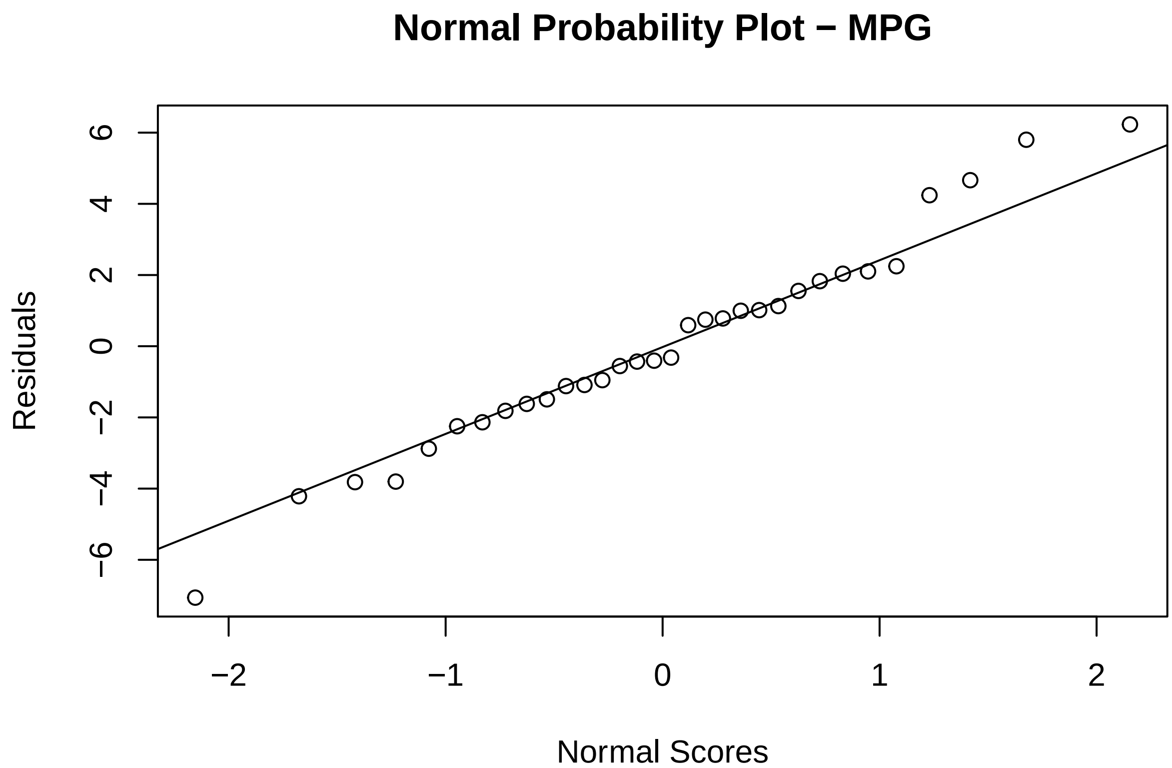
#4.4 Using 3.5

Previous work

```r
mpg <- table.b3
model <- lm(y ~ x1 + x6, data=mpg)
```

a) Construct a normal probability plot of the residuals. Is there any problem with the normality assumption?

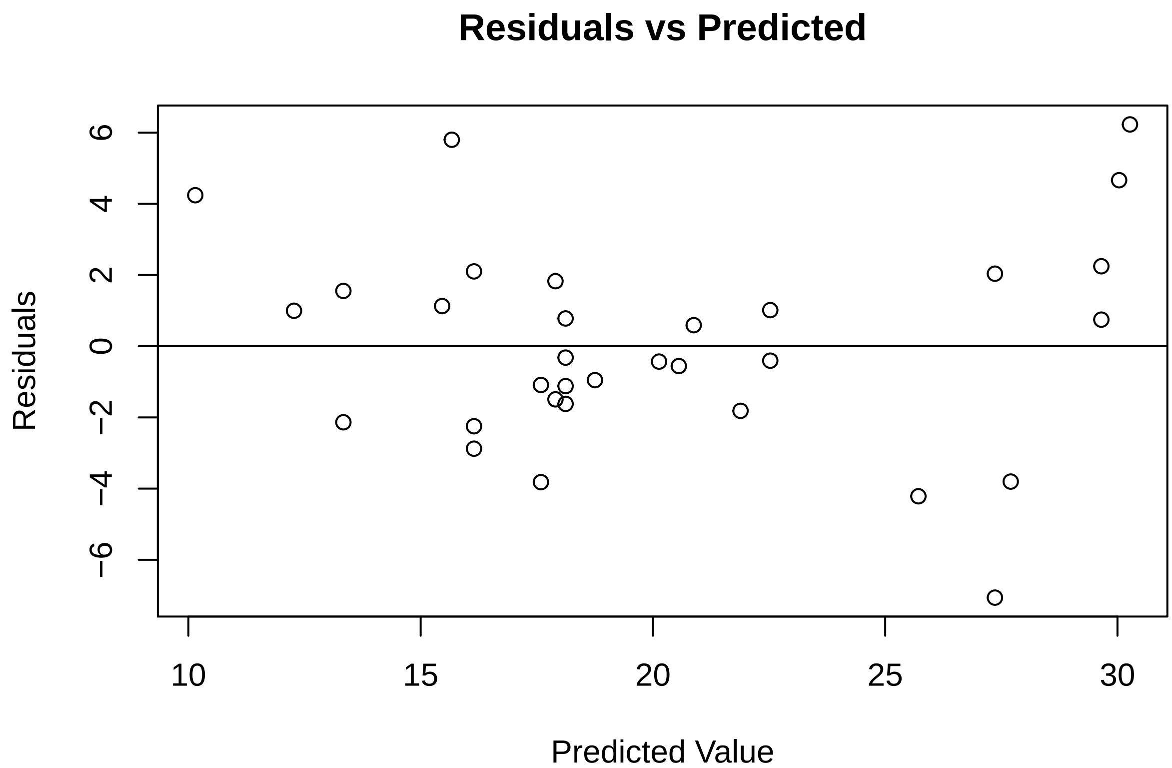
```r
residuals = model$residuals
qqnorm(residuals, ylab='Residuals', xlab='Normal Scores', main='Normal Probability Plot - MPG')
qqline(residuals)
```



This plot appears normal.

b) Construct and interpret a plot of the residuals versus the predicted response

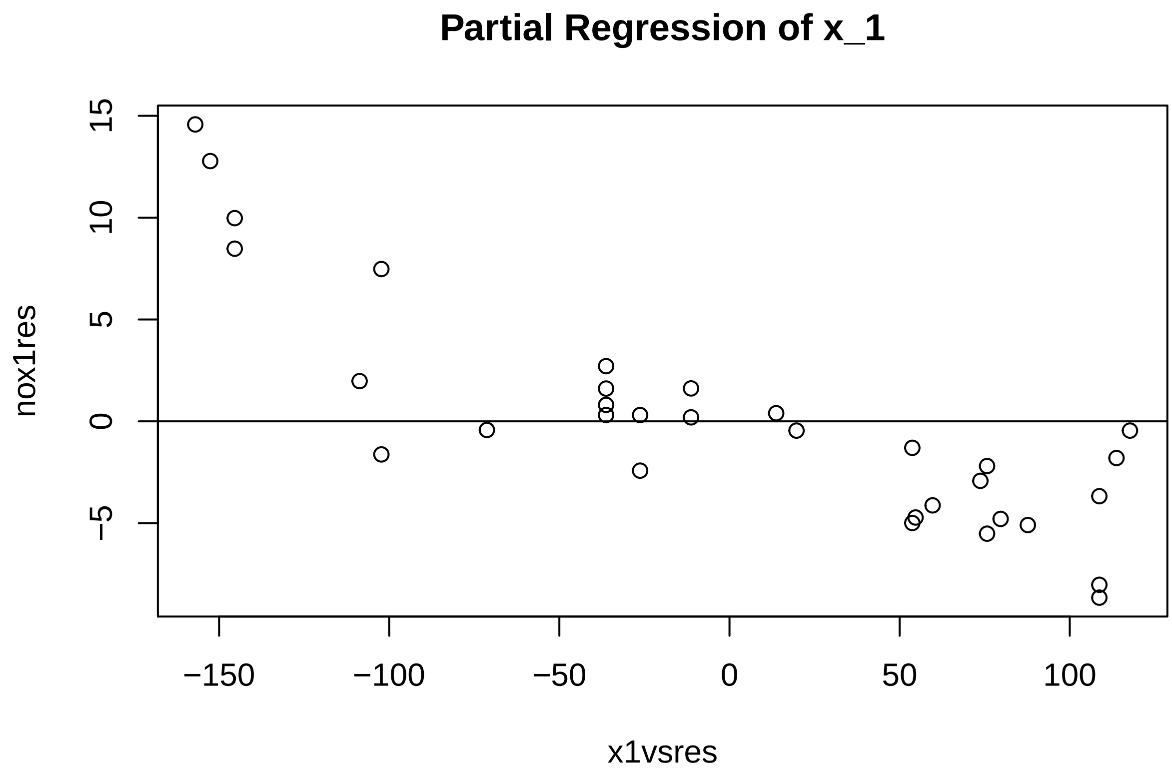
```
predicted = predict(model)
plot(predicted, residuals, xlab='Predicted Value', ylab='Residuals', main='Residuals vs Predicted')
abline(0,0)
```



The pattern does not appear to be linear.

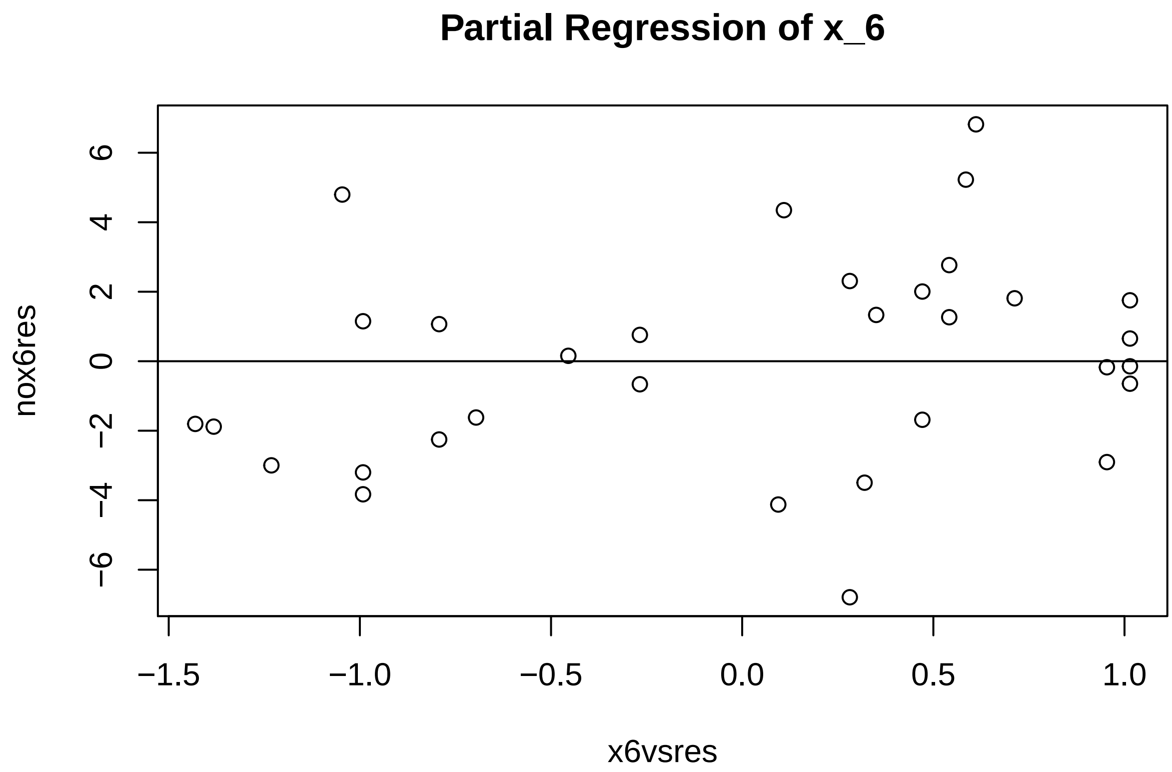
c) Construct the partial regression plots for this model.

```
nox1 <- lm(y ~ x6, data=mpg)
nox1res <- nox1$residuals
x1vs <- lm(x1 ~ x6, data=mpg)
x1vsres <- x1vs$residuals
plot(x1vsres, nox1res, main='Partial Regression of x_1')
abline(0,0)
```



x_1 appears to show a linear pattern.

```
nox6 <- lm(y ~ x1, data=mpg)
nox6res <- nox6$residuals
x6vs <- lm(x6 ~ x1, data=mpg)
x6vsres <- x6vs$residuals
plot(x6vsres, nox6res, main='Partial Regression of x_6')
abline(0,0)
```



x_6 does not appear to show a pattern, it may be unnecessary.

d) Compute the studentized residuals and the R-student residuals for the model. What information is converyed by these scaled residuals?

```r
rstandard(model)
```

```
##          1          2          3          4          5          6 
##  0.2717882 -0.3900557 -0.1980105  0.7313729 -0.6404817 -0.7469016 
##          7          8          9         10         11         12 
## -0.1376161  0.2007271  1.6672464  0.2656477 -0.5642251  2.2343937 
##         13         14         15         16         17         18 
## -1.4445983 -0.1539895 -2.4499447 -0.3253417  1.5186561  0.5431086 
##         19         20         21         22         23         24 
## -0.1113846 -0.5121263  0.3438369  2.0277182  0.3917877  0.7990680 
##         25         26         27         28         29         30 
##  0.7068771  0.3504291 -1.3231450  0.6274283 -0.7822915 -1.0015118 
##         31         32 
## -1.3281597 -0.3785506
```

It appears that 12 and 15 are close to being outliers, 22 is also high, but not as high.

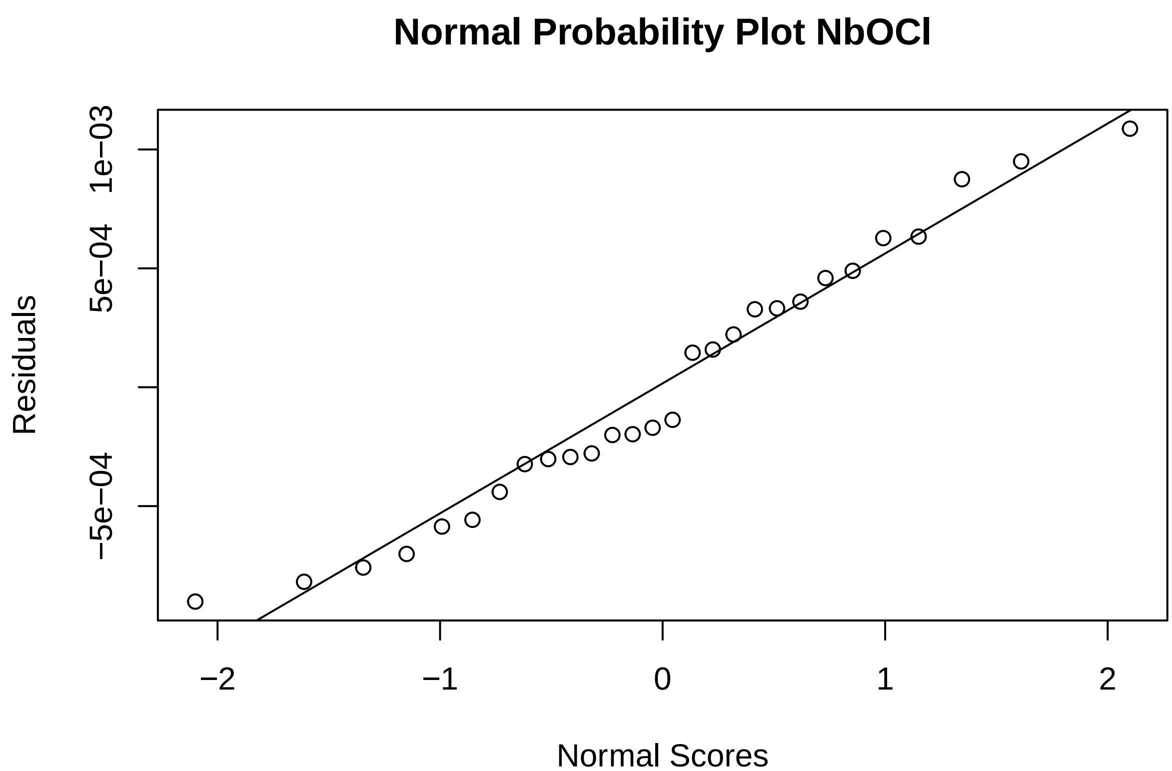
#4.15 Using 3.9

Previous work

```r
NbOCl <- table.b6
model <- lm(y ~ x1 + x4, data=NbOCl)
```

a) Construct a normal probability plot of the residuals. Is there any problem with the normality assumption?

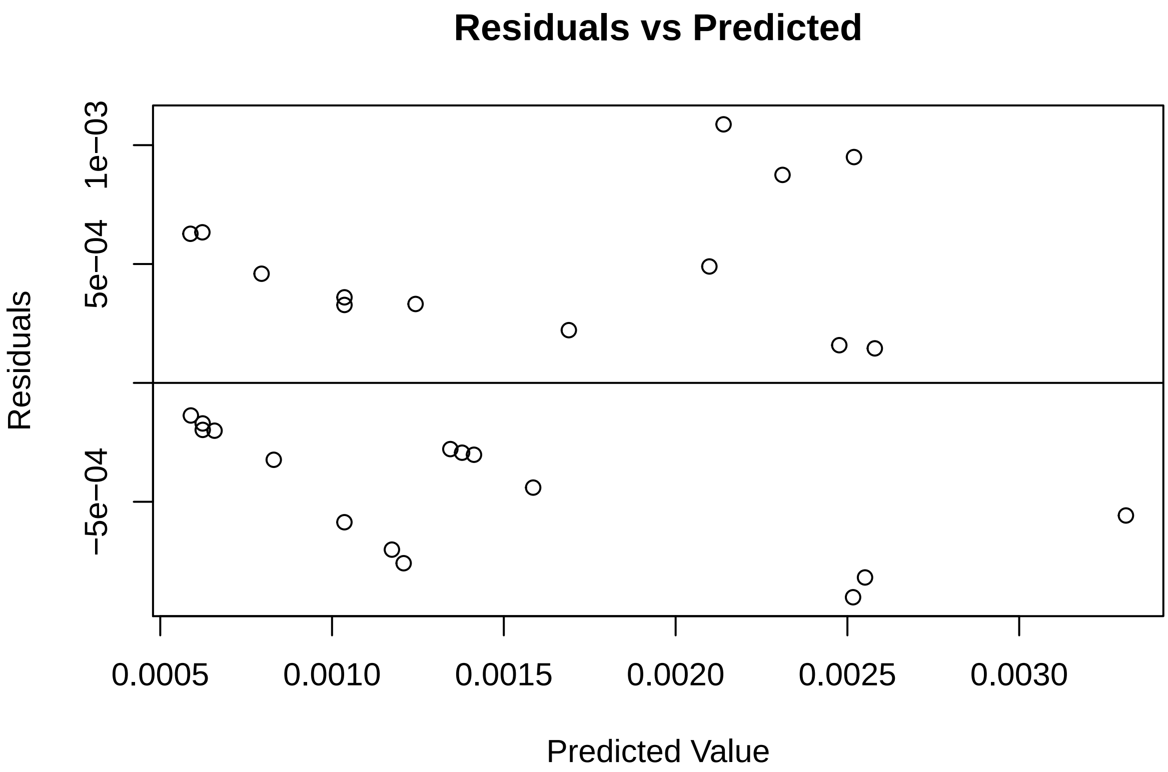
```r
residuals = model$residuals
qqnorm(residuals, ylab='Residuals', xlab='Normal Scores', main='Normal Probability Plot NbOCl')
qqline(residuals)
```



This appears mostly normal but there is a slight problem.

b) Construct and interpret a plot of the residuals versus the predicted response

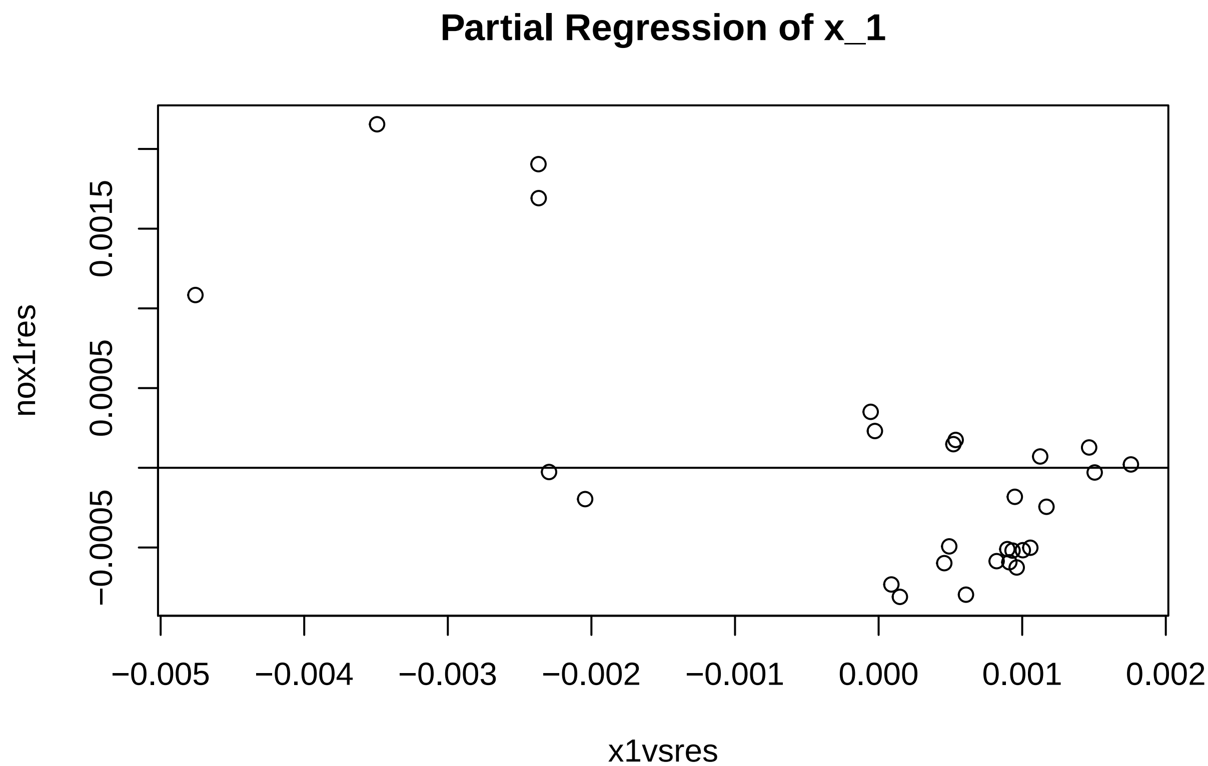
```r
predicted = predict(model)
plot(predicted, residuals, xlab='Predicted Value', ylab='Residuals', main='Residuals vs Predicted')
abline(0,0)
```



There does not appear to be a linear pattern.

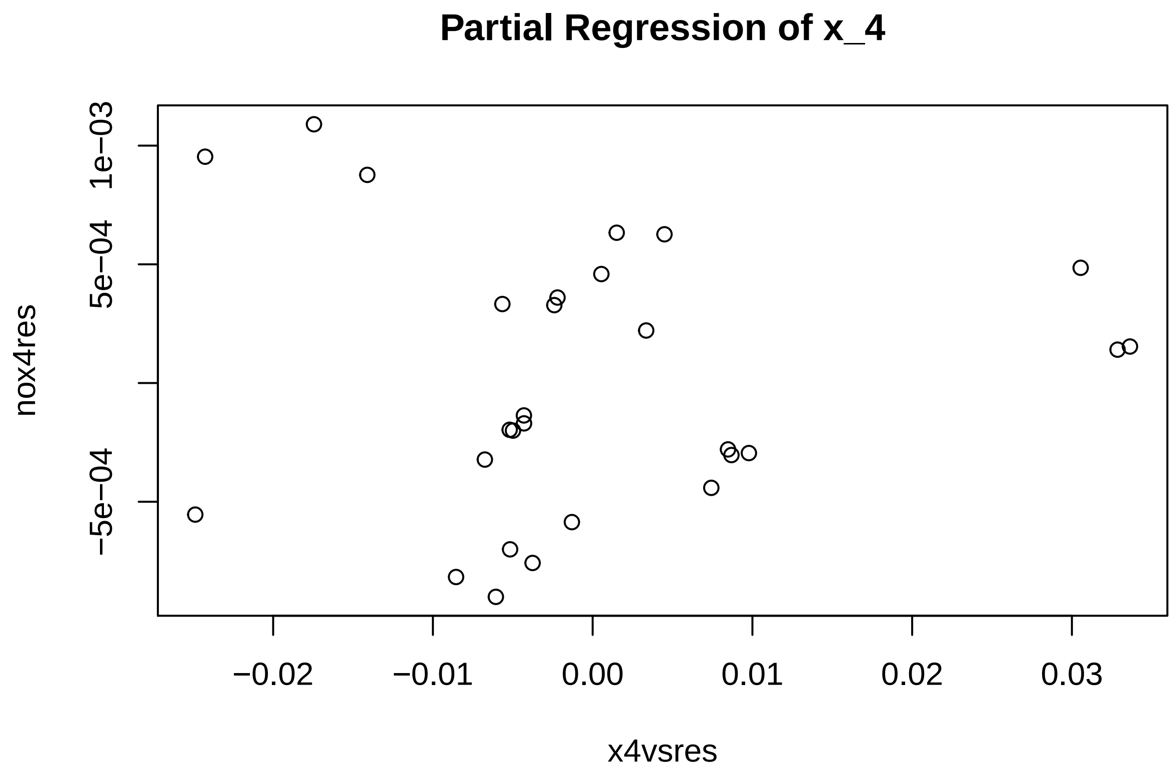
c) Construct the partial regression plots for this model. Are some variables currently in the model not necessary?

```
nox1 <- lm(y ~ x4, data=NbOCl)
nox1res <- nox1$residuals
x1vs <- lm(x1 ~ x4, data=NbOCl)
x1vsres <- x1vs$residuals
plot(x1vsres, nox1res, main='Partial Regression of x_1')
abline(0,0)
```



x_1 may show a linear pattern.

```
nox4 <- lm(y ~ x1, data=NbOCl)
nox4res <- nox4$residuals
x4vs <- lm(x4 ~ x1, data=NbOCl)
x4vsres <- x4vs$residuals
plot(x4vsres, nox4res, main='Partial Regression of x_4')
```



x_4 does not show a linear pattern.

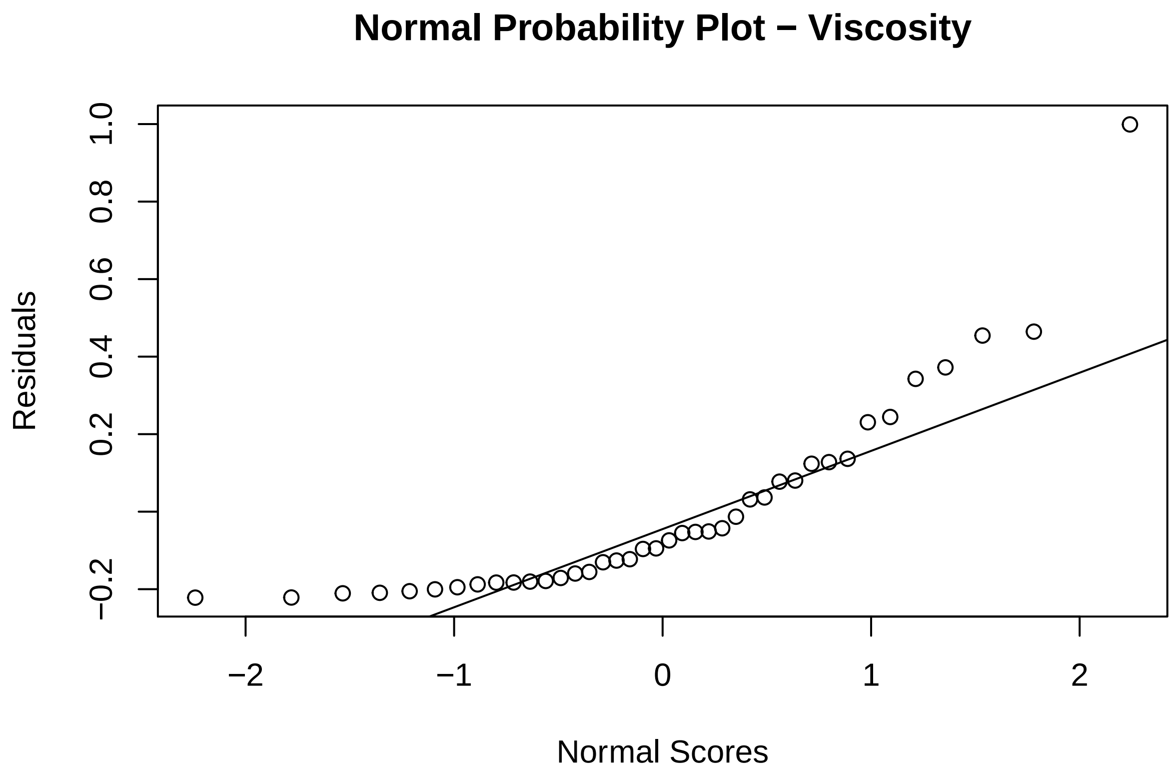
#4.17 Using 3.14

Previous work

```
viscosity <- table.b10
model <- lm(y ~ x1 + x2, data=viscosity)
```

a) Construct a normal probability plot of the residuals. Is there any problem with the normality assumption?

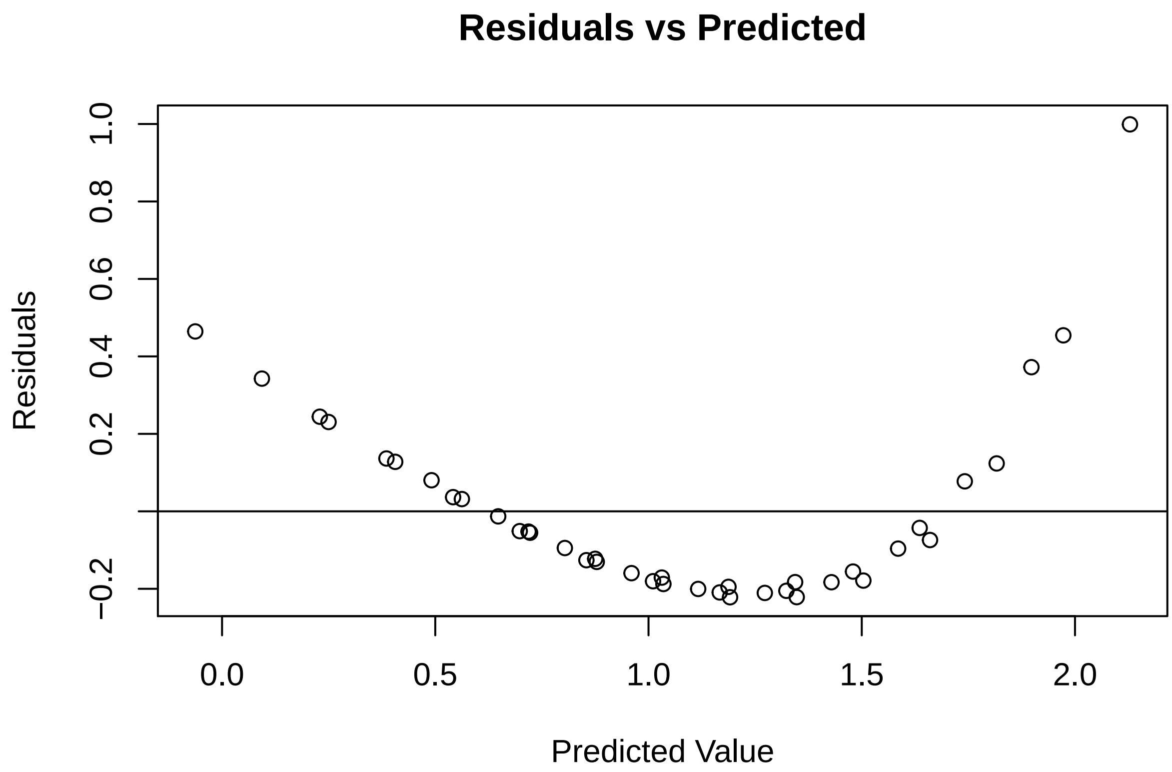
```
residuals = model$residuals
qqnorm(residuals, ylab='Residuals', xlab='Normal Scores', main='Normal Probability Plot - Viscosity')
qqline(residuals)
```



The plot does not appear to be normal.

b) Construct and interpret a plot of the residuals versus the predicted response

```
predicted = predict(model)
plot(predicted, residuals, xlab='Predicted Value', ylab='Residuals', main='Residuals vs Predicted')
abline(0,0)
```



There is definitely a nonlinear pattern.

c) Compute the PRESS statistic for both this model and the second in 3.14. Based on this statistic, which model is most likely to provide better predictions of new data?

```
residuals/(1-lm.influence(model)$hat)
```

```
##           1           2           3           4           5           6 
##  1.14548800  0.50686229  0.13519800 -0.07991364 -0.19165067 -0.23743967 
##           7           8           9          10          11          12 
## -0.23932862 -0.20511994 -0.14559578 -0.06332083  0.41013482  0.08313029 
##          13          14          15          16          17          18 
## -0.10134736 -0.19021131 -0.21764758 -0.20701915 -0.16592174 -0.09972347 
##          19          20          21          22          23          24 
## -0.01398824  0.08844560 -0.04704130 -0.16649957 -0.21552954 -0.21708411 
##          25          26          27          28          29          30 
## -0.18595670 -0.13000876 -0.05295577  0.03846460  0.14605610  0.26854172 
##          31          32          33          34          35          36 
## -0.21124569 -0.21921726 -0.18850670 -0.13330409 -0.05651167  0.03420574 
##          37          38          39          40 
##  0.13890183  0.25411460  0.38521494  0.53676844
```

```
tempmodel <- lm(y ~ x2, data=viscosity)
tempres <- tempmodel$residuals
tempres/(1-lm.influence(tempmodel)$hat)
```

```
##           1           2           3           4           5           6
```

```
##  1.50652587  0.88679402  0.52401886  0.31324930  0.20382068  0.15997994
##           7           8           9          10          11          12
##  0.16065962  0.19845626  0.26317334  0.35256169  0.56742139  0.23852262
##          13          14          15          16          17          18
##  0.05229144 -0.03792441 -0.06613266 -0.05557216 -0.01379108  0.05390475
##          19          20          21          22          23          24
##  0.14204895  0.24792488 -0.17357363 -0.28926414 -0.33575927 -0.33580235
##          25          26          27          28          29          30
## -0.30406112 -0.24833733 -0.17233568 -0.08280206  0.02199079  0.14066119
##          31          32          33          34          35          36
## -0.64640945 -0.64272132 -0.60551204 -0.54784930 -0.47229440 -0.38608538
##          37          38          39          40
## -0.28925587 -0.18530602 -0.06949225  0.06196468
```

The model without x_1 appears to have greater peaks and valleys so the original model with both x_1 and x_2 is prefered.