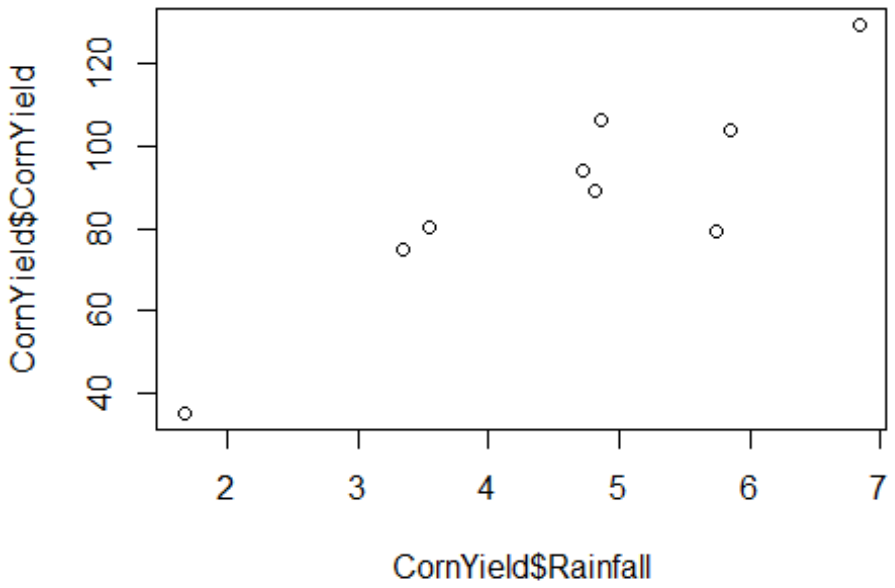
Morgan Exam 1

Isaiah Morgan

February 24, 2018

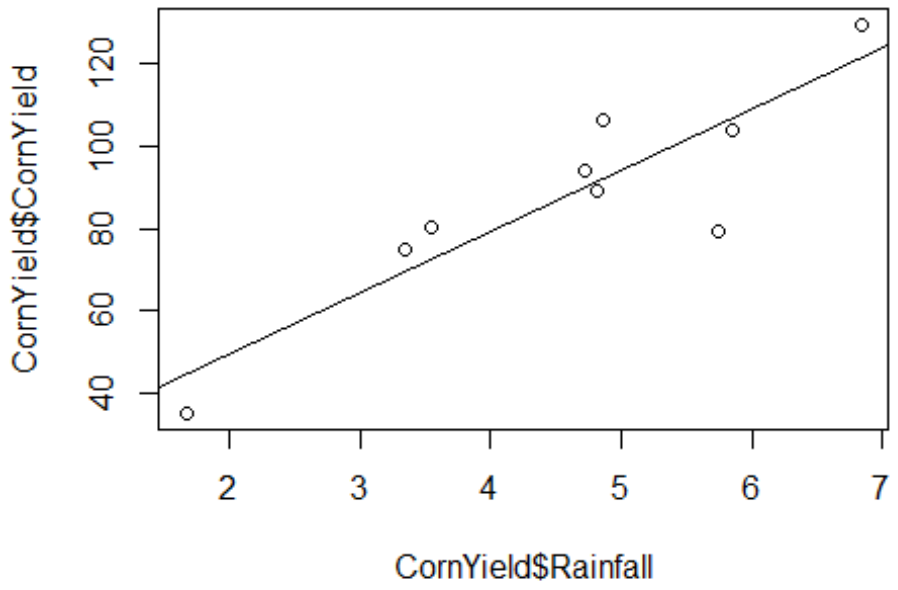
```
#1
CornYield <- read.table('C:/Users/imoe9/Documents/School Work/STAT PROG/R
Files/STAT 330/Exam 1/CornYield.txt', header = TRUE)
plot(CornYield$Rainfall, CornYield$CornYield)
```



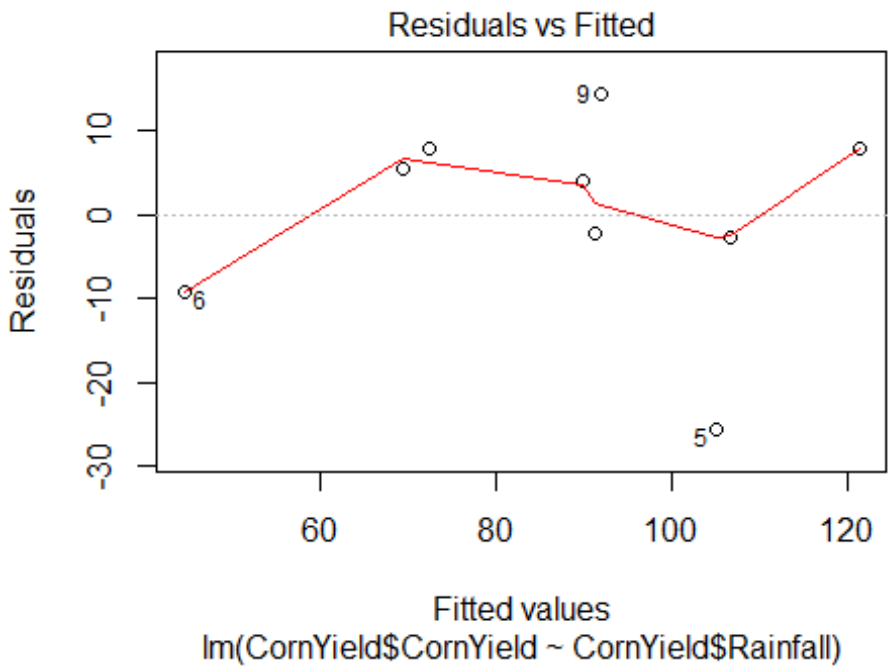

```
#2
fitcy <- lm(CornYield$CornYield ~ CornYield$Rainfall)
coef(fitcy)

##        (Intercept) CornYield$Rainfall
##           19.76071           14.83463

#3 Yes, I would reject the null hypothesis that the slope of the line is
equal to zero because the p-value is equal to 0.00134
summary(fitcy)

##
## Call:
## lm(formula = CornYield$CornYield ~ CornYield$Rainfall)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.560  -2.743   3.920   7.870  14.343
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          19.761     13.933   1.418  0.19905
## CornYield$Rainfall   14.835      2.886   5.140  0.00134 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.68 on 7 degrees of freedom
## Multiple R-squared:  0.7905, Adjusted R-squared:  0.7606
## F-statistic: 26.42 on 1 and 7 DF,  p-value: 0.001339
```

```
#4
plot(CornYield$Rainfall, CornYield$CornYield)
abline(coef(fitcy))
```

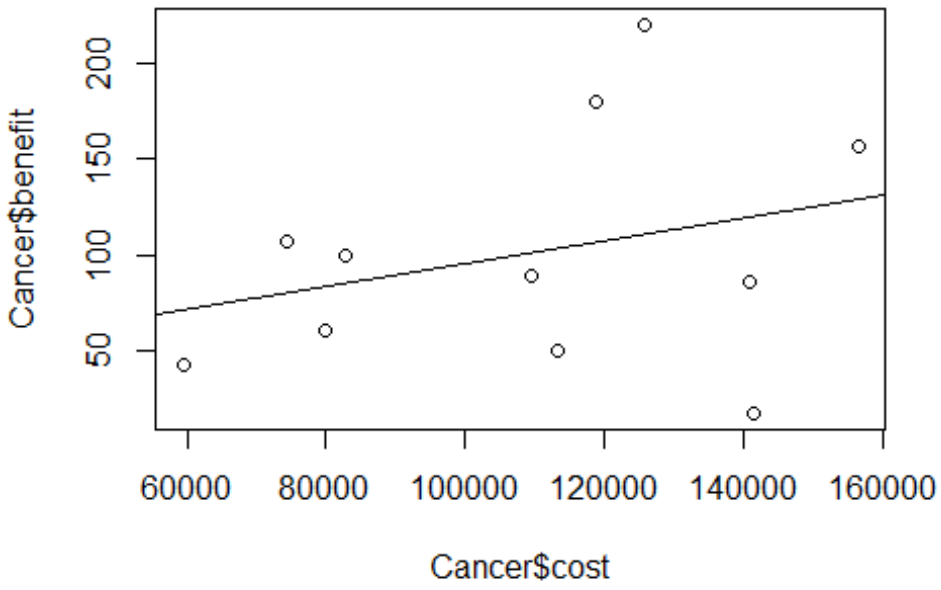



```
#5
plot(fitcy, which = 1)
```




```
#6
c4 <- cbind(1,4)
bhat <- coef(fitcy)
c4%*%bhat

##          [,1]
## [1,] 79.09922

#7
vhatbhat <- vcov(fitcy)

uci <- c4%*%bhat + qt(.975,493)*sqrt(c4%*%vhatbhat%*%t(c4))
lci <- c4%*%bhat - qt(.975,493)*sqrt(c4%*%vhatbhat%*%t(c4))
c(lci,c4%*%bhat,uci)

## [1] 70.12382 79.09922 88.07463

#8
s2hat <- 12.68^2

upi <- c4%*%bhat + qt(.975,493)*sqrt(c4%*%vhatbhat%*%t(c4) + s2hat)
lpi <- c4%*%bhat - qt(.975,493)*sqrt(c4%*%vhatbhat%*%t(c4) + s2hat)
c(lpi,c4%*%bhat,upi)

## [1]  52.61827  79.09922 105.58017

#9
rm(list=ls())
```

```r
Cancer <- read.table('C:/Users/imoe9/Documents/School Work/STAT PROG/R Files/STAT 330/Exam 1/cancer.txt', header = TRUE)
plot(Cancer$cost,Cancer$benefit)

#10
fit <- lm(Cancer$benefit~Cancer$cost)
abline(coef(fit))
```




```r
#11 If hii > 4/n

#12
abs(hatvalues(fit)) > 4/ length(hatvalues(fit))
```

```
##     1     2     3     4     5     6     7     8     9    10    11
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
#13
rstudent(fit)
```

```
##          1          2          3          4          5          6
## -0.5764752  0.5132572  0.4702014 -0.3789665  2.2486157 -0.1921946
##          7          8          9         10         11
##  0.2468625 -0.8849901 -0.5443293  1.2864676 -2.1768906
```

```r
#14 Drugs 5 and 11 are both potential outliers because they are greater than 2.
```

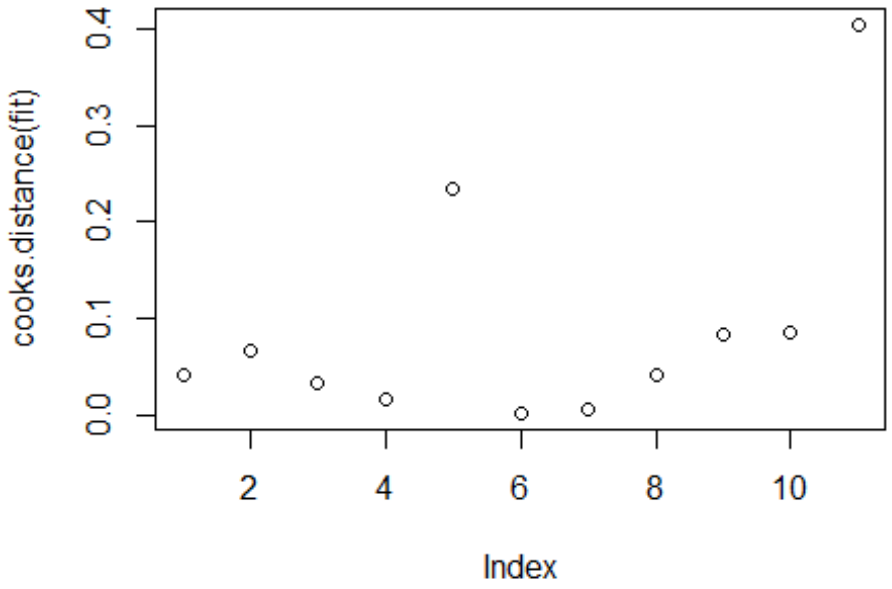
```r
#15 If Cook's distance exceeds 4/(n-2), it is potentially an outlier.

#16 None of the points are flagged by this test.
cooks.distance(fit) > (4/9)
```

```
##     1     2     3     4     5     6     7     8     9    10    11
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
#17 None of the points are flagged by this test
pf(cooks.distance(fit),1,9) > .5
```

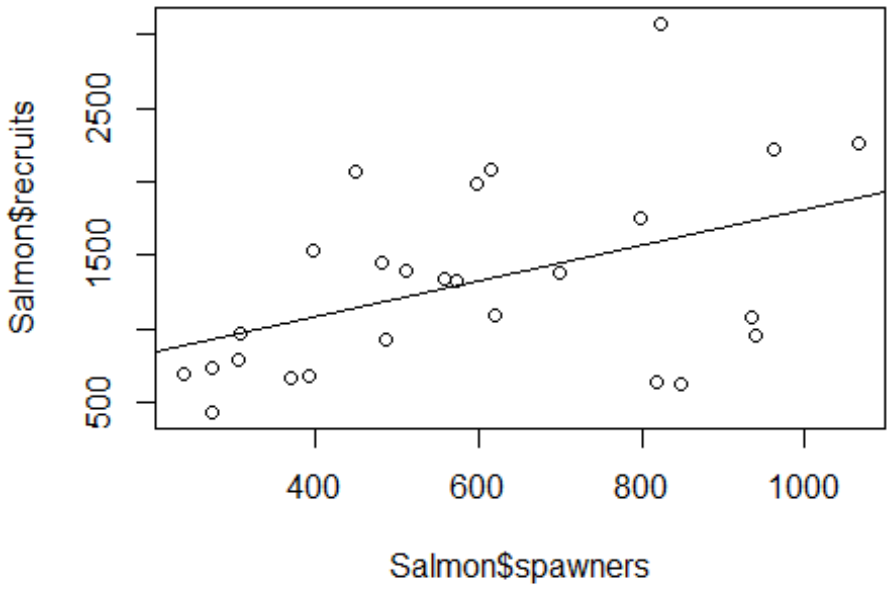
```
##     1     2     3     4     5     6     7     8     9    10    11
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```r
#18 Both Point 5 and 11 have a Cook's Distance that is significantly greater than the rest
plot(cooks.distance(fit))
```

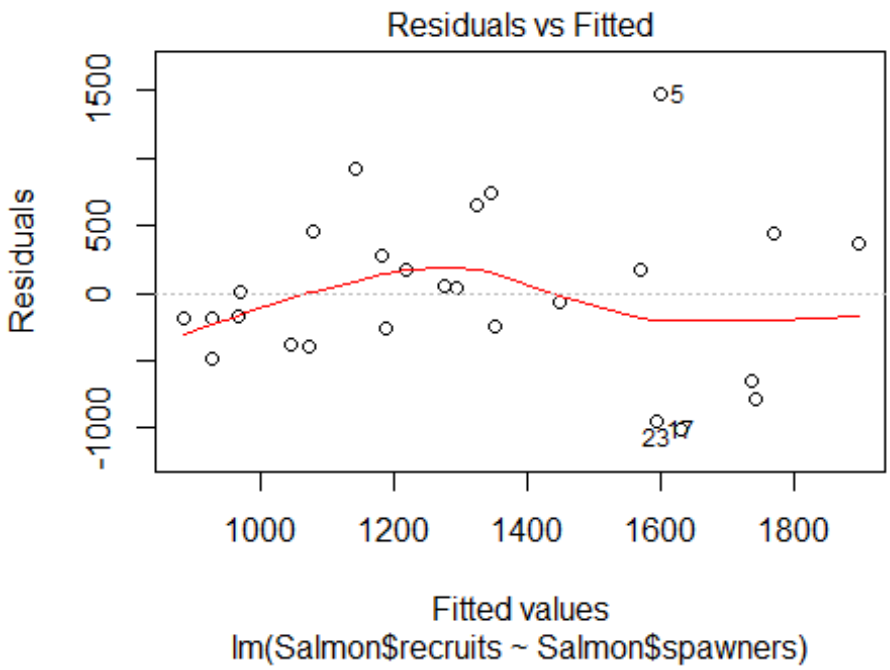



```r
#19 These points don't appear to be as much as outliers as they are high influence points, so the high cook's distance of both is probably more the result of being high inluence. If they were outliers they would have failed more of the diagnostic tests.

#20
rm(list=ls())
Salmon <- read.table('C:/Users/imoe9/Documents/School Work/STAT PROG/R Files/STAT 330/Exam 1/salmon.txt', header = TRUE)
```

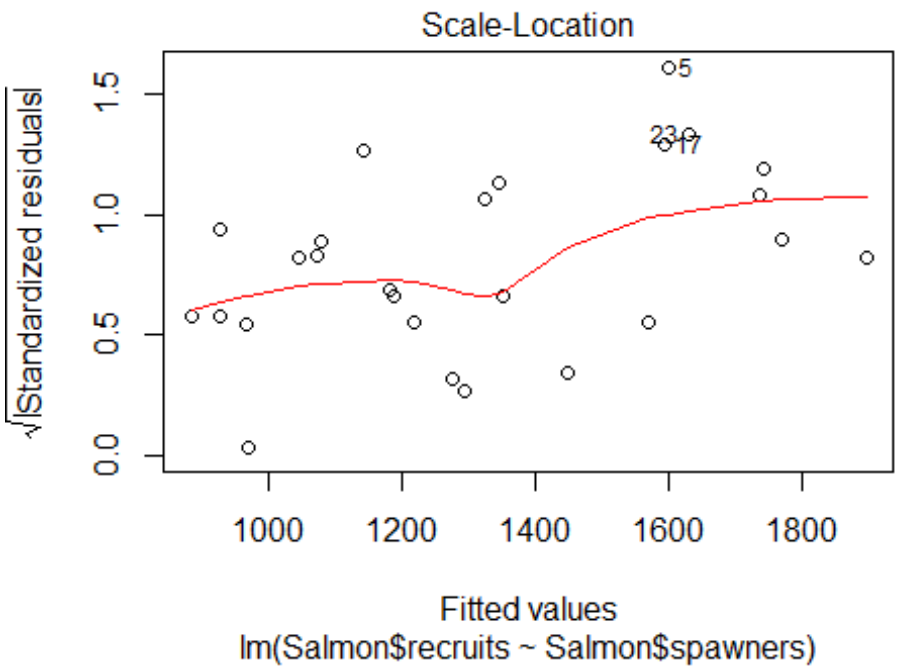
```r
plot(Salmon$spawners, Salmon$recruits)

#21
fit <- lm(Salmon$recruits ~ Salmon$spawners)
abline(coef(fit))
```




```r
#22
plot(fit, which = 1)
```

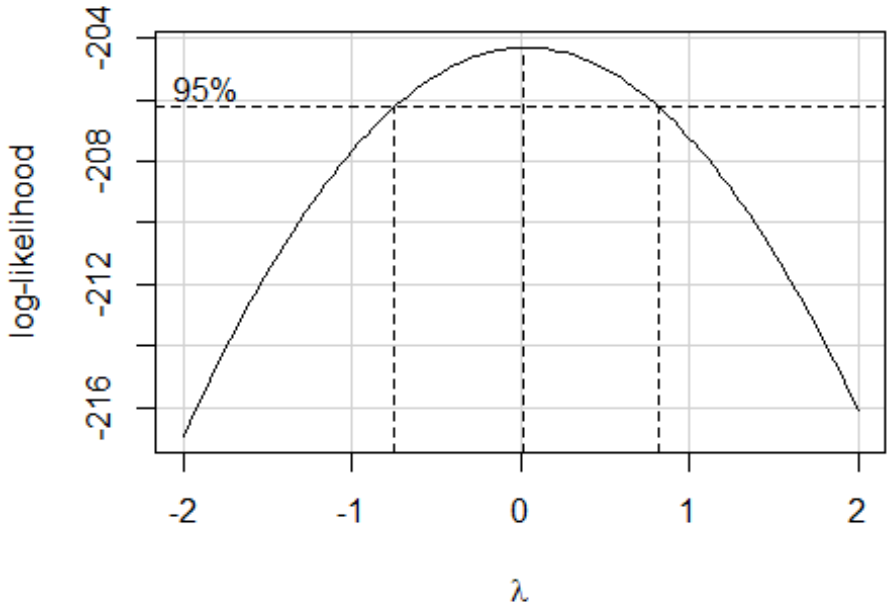



```r
#23 No, though the plot does not have a slope of zero, it doesn't deviate in a way that is overly concerning

#24
plot(fit, which = 3)

#25
abline(fit)
```

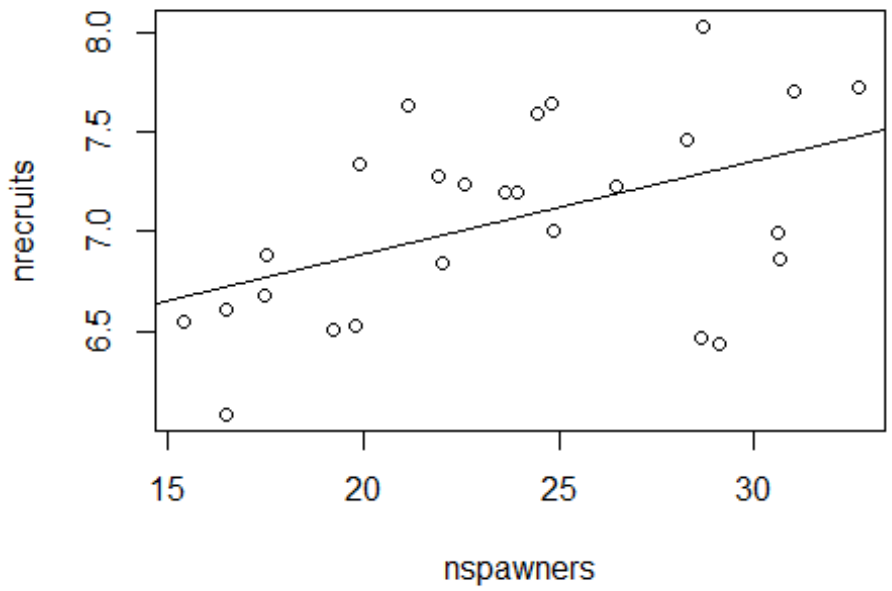



```
#26 The Line is significantly different from zero with a p-value at 0.01839
summary(fit)

##
## Call:
## lm(formula = Salmon$recruits ~ Salmon$spawners)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1003.49  -348.27   -34.25   337.79  1469.79
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     596.0436   307.1241   1.941   0.0641 .
## Salmon$spawners   1.2199     0.4822   2.530   0.0184 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 590.9 on 24 degrees of freedom
## Multiple R-squared:  0.2105, Adjusted R-squared:  0.1776
## F-statistic:   6.4 on 1 and 24 DF,  p-value: 0.01839

#27 From the Boxcox it is clear that a log transform should be applied to the
data to achieve normality as the peak of the curve is approximately zero.
library(alr3)

## Warning: package 'alr3' was built under R version 3.4.3
```

```
## Loading required package: car

## Warning: package 'car' was built under R version 3.4.2

boxCox(fit)
```

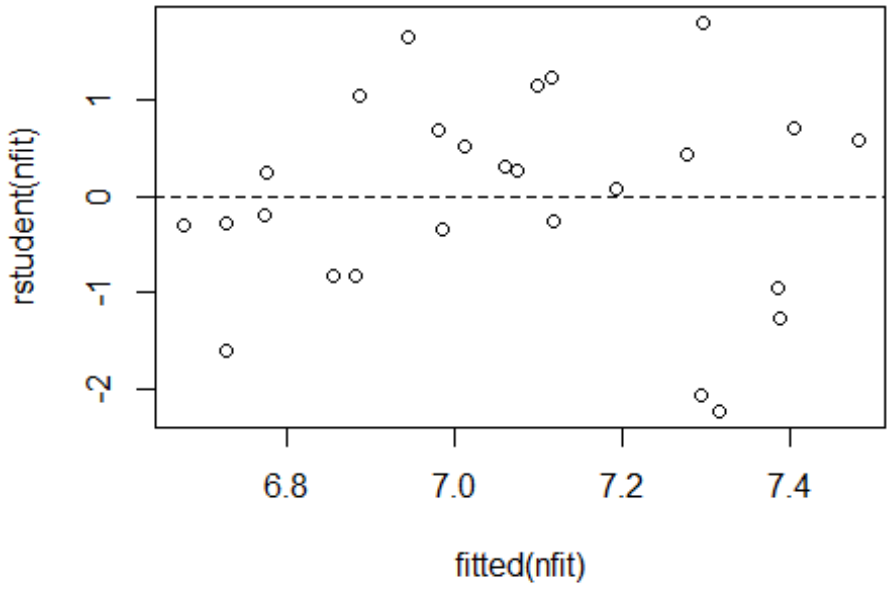



```
#28 If I were to transform this variable, I would cubic root it as the the
powerTransform value is approximately 1/3.
powerTransform(Salmon$spawners)

## Estimated transformation parameters
## Salmon$spawners
##       0.3227392

#29
nrecruits <- log(Salmon$recruits)
nspawners <- sqrt(Salmon$spawners)
nfit <- lm(nrecruits ~ nspawners)
coef(nfit)

## (Intercept)   nspawners
##  5.95626686  0.04669339

#30
plot(nspawners, nrecruits)
abline(coef(nfit))
```

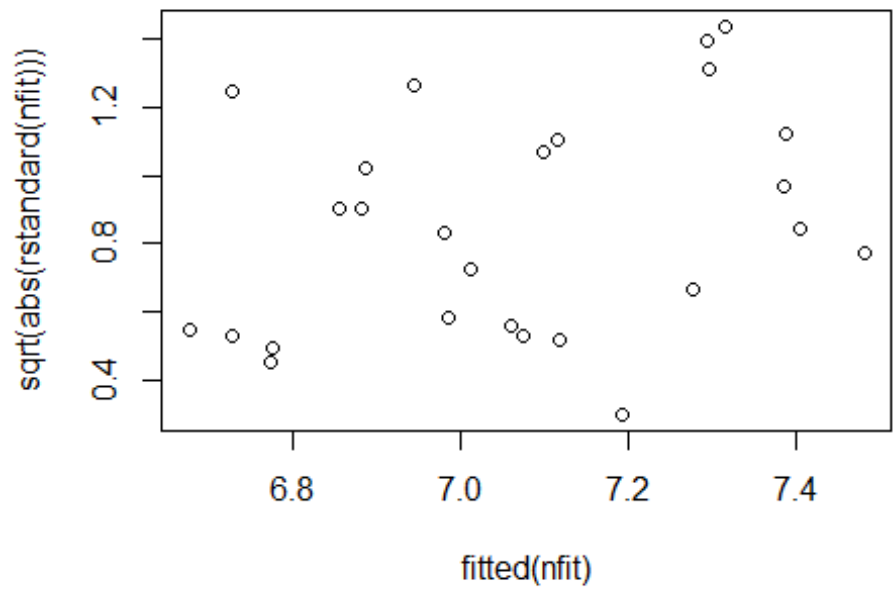



```
#31 Yes, the slope is significantly different from zero at a = 0.05 with a p
= 0.0128.
summary(nfit)

##
## Call:
## lm(formula = nrecruits ~ nspawners)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87505 -0.29915  0.07142  0.28356  0.73314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.95627    0.42147  14.132 3.94e-13 ***
## nspawners    0.04669    0.01735   2.691   0.0128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4432 on 24 degrees of freedom
## Multiple R-squared:  0.2317, Adjusted R-squared:  0.1997
## F-statistic:  7.24 on 1 and 24 DF,  p-value: 0.01278

#32
plot(fitted(nfit),rstudent(nfit))
```

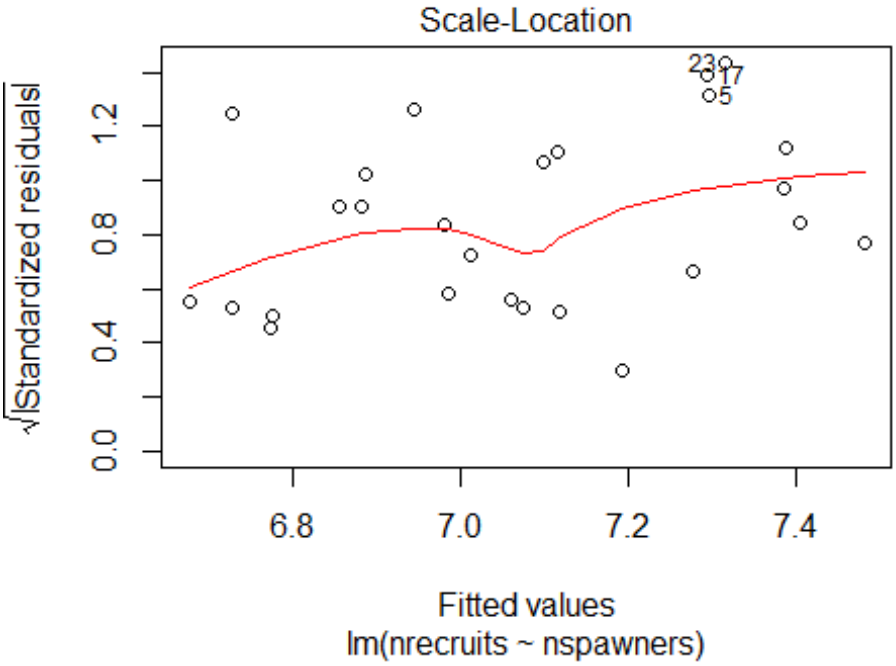
```r
#33
abline(h=0, lty = 2)
```



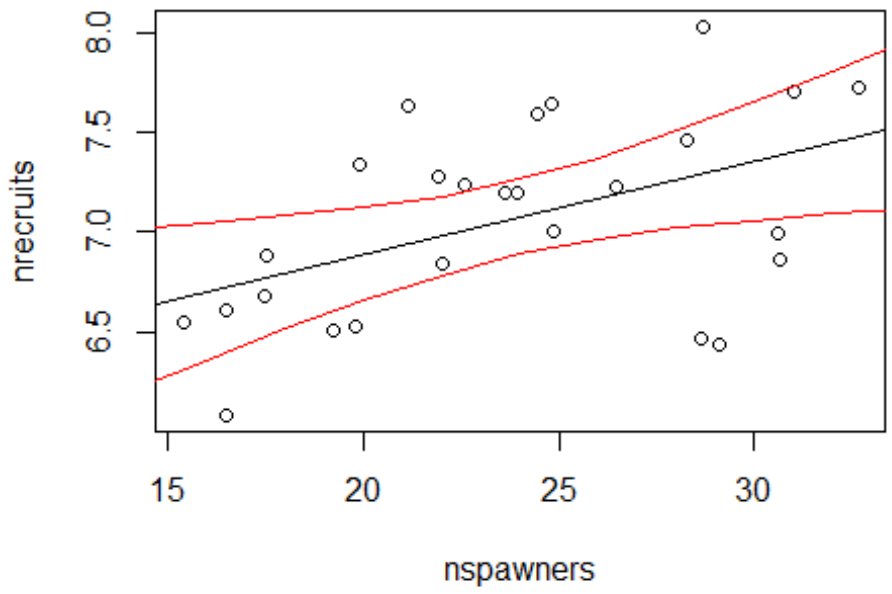

```r
#34
plot(fitted(nfit),sqrt(abs(rstandard(nfit))))
```




```r
#35
plot(nfit, which = 3)
```



```r
#36 Yes, there is sufficient evidence to conclude that the slope of the line
is not equal to zero.

#37
plot(nspawners, nrecruits)
abline(coef(nfit))
bhat <- coef(nfit)
vhatbhat <- vcov(nfit)

xx <- seq(12,42,by=2)
newxx <- cbind(1,xx)

uci <- newxx%*%bhat + qt(.975,14)*sqrt(diag(newxx%*%vhatbhat%*%t(newxx)))
lci <- newxx%*%bhat - qt(.975,14)*sqrt(diag(newxx%*%vhatbhat%*%t(newxx)))

lines(xx,uci,col='red')
lines(xx,lci,col='red')
```




```r
#38
s2hat <- .4432^2

upi <- newxx%*%bhat + qt(.975,493)*sqrt(diag(newxx%*%vhatbhat%*%t(newxx)) +
s2hat)
lpi <- newxx%*%bhat - qt(.975,493)*sqrt(diag(newxx%*%vhatbhat%*%t(newxx)) +
s2hat)
```

```
plot(nspawners, nrecruits)
abline(coef(nfit))
lines(xx,upi,col='blue')
lines(xx,lpi,col='blue')
```