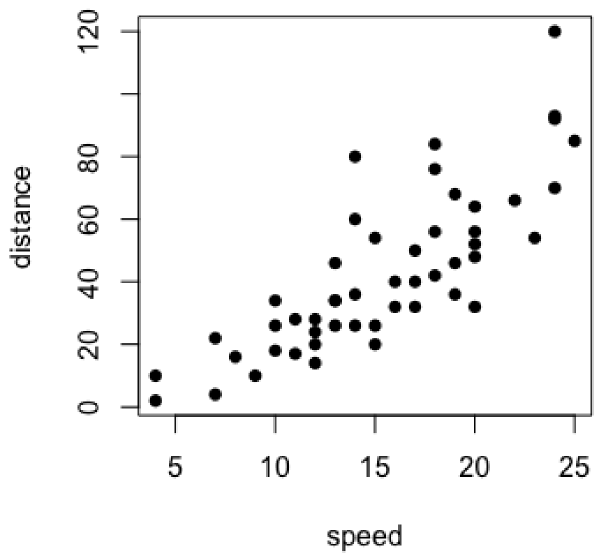
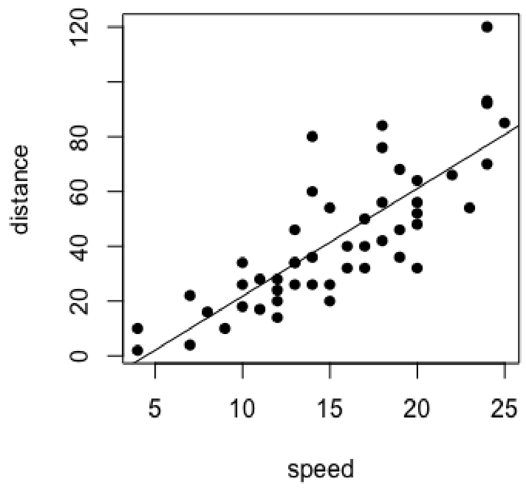
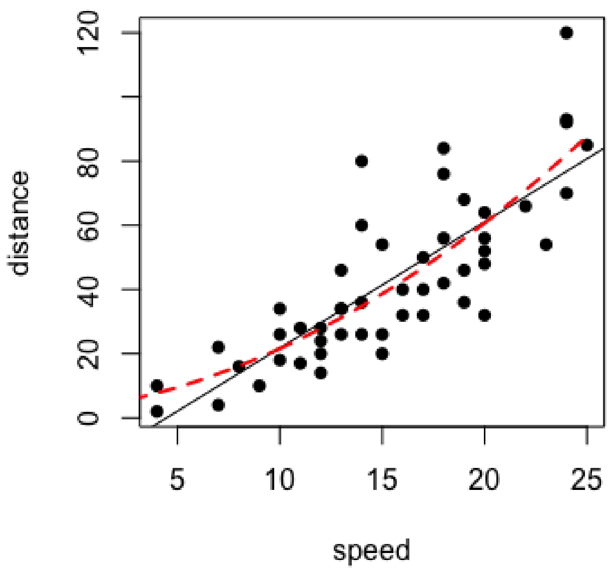
```
# title: "MSDS596 - HW3"
# author: "Diego Sarachaga"
# date: "10/23/2018"


#Problem 1
library(faraway)
library(MASS)

data(cars)
head(cars)

#part a
plot(dist ~ speed, data=cars, xlab = "speed", ylab = "distance",
pch=16)
```



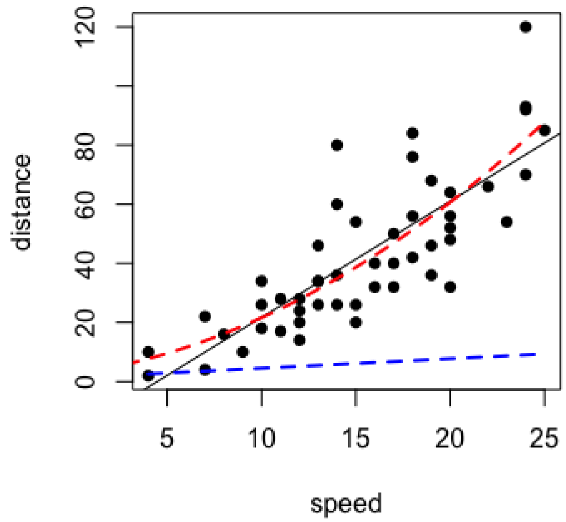

```
#part b
cars.lm <- lm(dist ~ speed, data=cars)
abline(cars.lm)
```



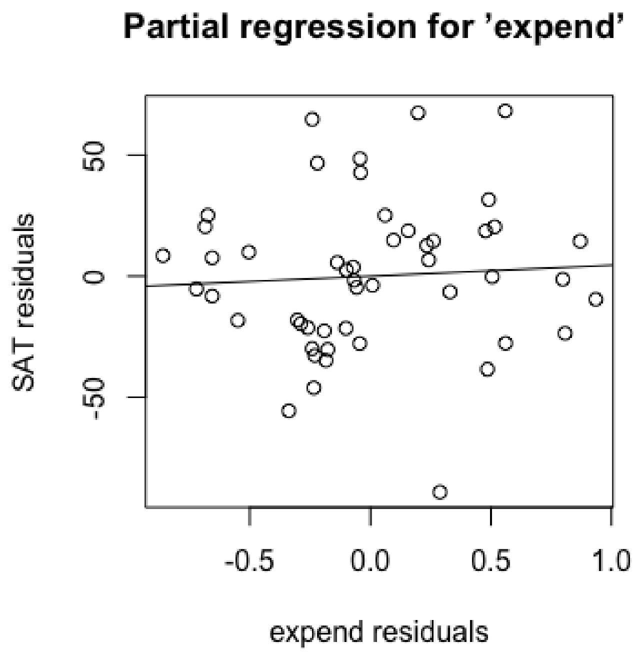

```
#part c
cars.lm2 <- lm(dist ~ speed + I(speed^2), data=cars)
xval <- seq(from=0, to=25, by=0.1)
hat2 <- predict(cars.lm2, newdata=list(speed=xval))
lines(xval, hat2, col="red", lty=2, lwd=2)
```




```
#part d
cars.lm3 <- lm(sqrt(dist) ~ speed, data=cars)
xval <- pretty(cars$speed, 50)
hat3 <- predict(cars.lm3, newdata=list(speed=xval))
lines(xval, hat3, col="blue", lty=2, lwd=2)
```



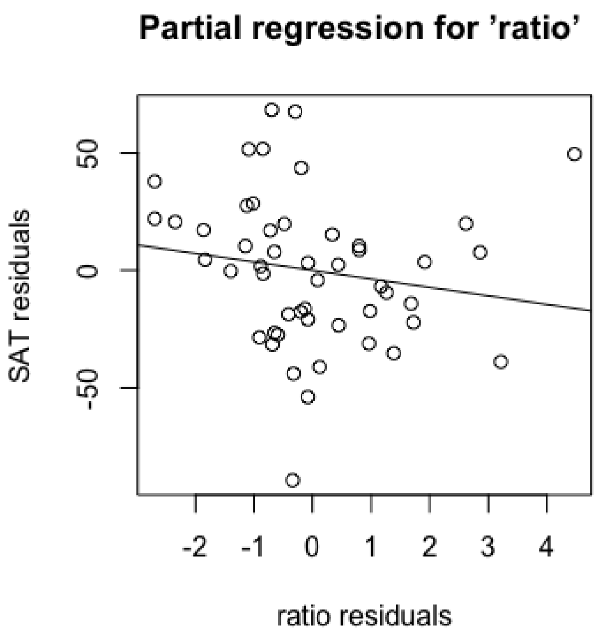

```
#Problem 2
library(faraway)
data(sat)
head(sat)

lmod <- lm(total ~ expend + salary + ratio + takers, sat)

#part a
#expend
d <- residuals(lm(total ~ salary + ratio + takers, sat))
m <- residuals(lm(expend ~ salary + ratio + takers, sat))
plot(m,d,xlab="expend residuals",ylab="SAT residuals",main="Partial
regression for 'expend' ")
abline(0,coef(lmod)['expend'])
```



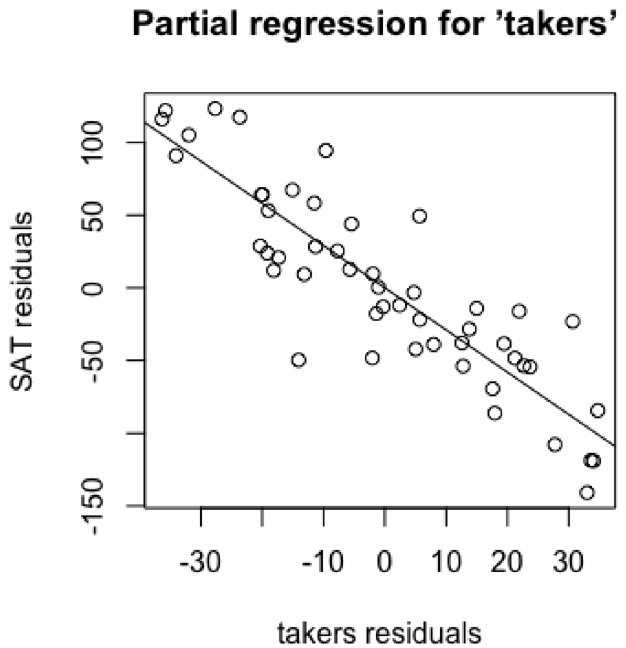

```
#salary
d <- residuals(lm(total ~ expend + ratio + takers, sat))
```

```
m <- residuals(lm(salary ~ expend + ratio + takers, sat))
plot(m,d,xlab="salary residuals",ylab="SAT residuals",main="Partial
regression for 'salary' ")
abline(0,coef(lmod)['salary'])
```


```
#ratio
d <- residuals(lm(total ~ expend + salary + takers, sat))
m <- residuals(lm(ratio ~ expend + salary + takers, sat))
plot(m,d,xlab="ratio residuals",ylab="SAT residuals",main="Partial
regression for 'ratio' ")
abline(0,coef(lmod)['ratio'])
```

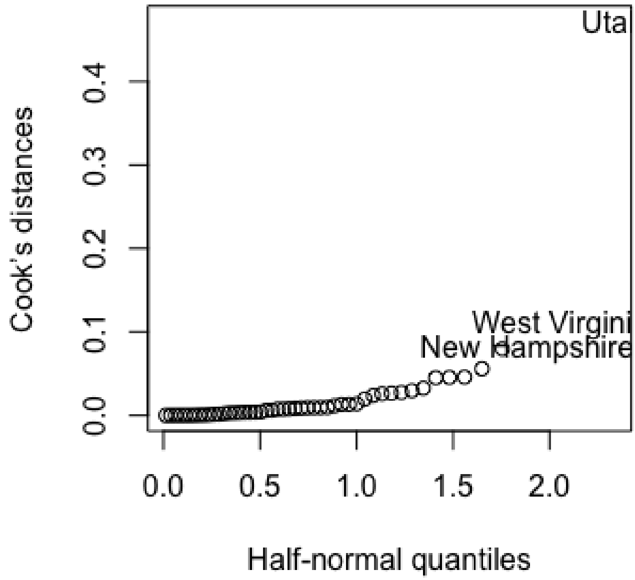


```
#takers
d <- residuals(lm(total ~ expend + salary + ratio, sat))
m <- residuals(lm(takers ~ expend + salary + ratio, sat))
plot(m,d,xlab="takers residuals",ylab="SAT residuals",main="Partial
regression for ’takers’ ")
abline(0,coef(lmod)['takers'])
```



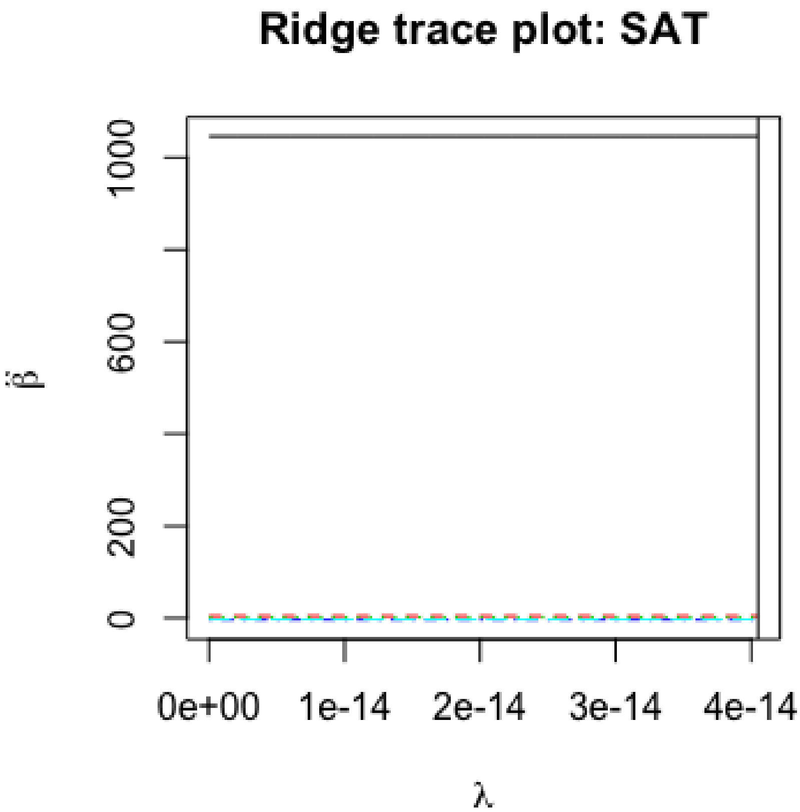

```
#part b
hatv <- hatvalues(lmod)
head(sort(hatv,decreasing=T))
#       Utah   | California | Connecticut   | New Jersey |   New York  |
Alaska
#   0.2921128 |  0.2821179 | 0.2254519    | 0.2220978  |   0.1915752 |
0.1803061

sum(hatv)
#5

#part c
stud <- rstudent(lmod)
jackres <- stud*(44/(45-stud^2))^0.5
head(jackres[order(abs(stud),decreasing=T)])
# West Virginia |    Utah      |  North Dakota    | New Hampshire |
Nevada     |     Iowa
# -3.491336     |    2.700696  |     2.318848     |    2.291067    |
-1.772759    |  1.586454


#part d
```

```
states <- row.names(sat)
cook <- cooks.distance(lmod)
halfnorm(cook,3,labs=states,ylab="Cook's distances")
```

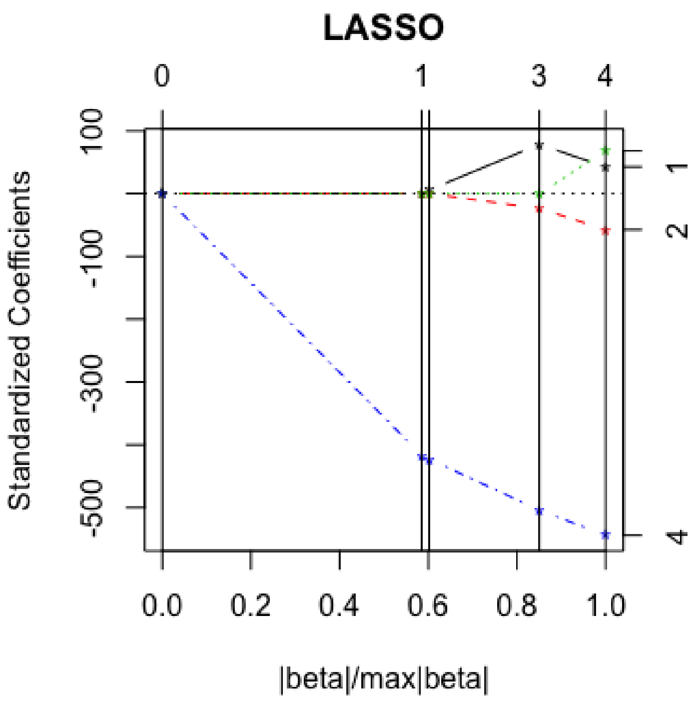


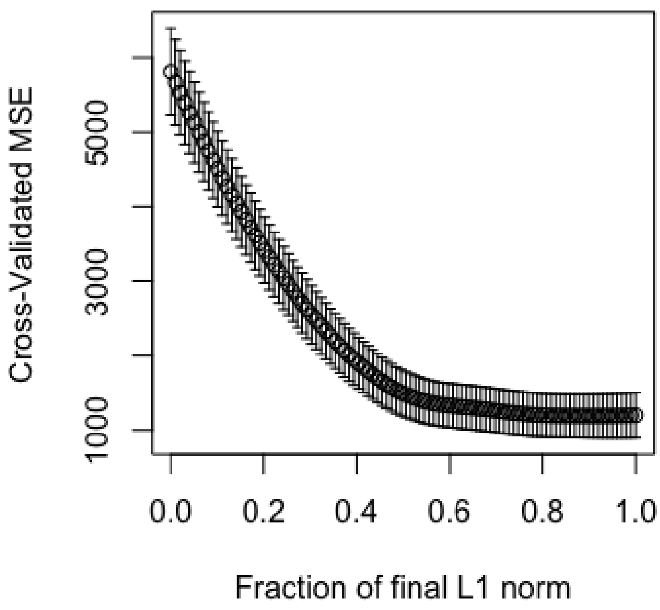

```
#part e
x <- model.matrix(lmod)[, -1]
vif(x)
# expend    |  salary    | ratio      | takers
# 9.465320 |  9.217237  | 2.433204  | 1.755090


#part f
rgmod <- lm.ridge(total ~ expend + salary + ratio + takers, sat,
lambda = seq(0, 4.05e-14, len=21))
select(rgmod)
#modified HKB estimator is 0.3496881
#modified L-W estimator is 0.4728102
#smallest value of GCV  at 4.05e-14

coef(rgmod)
#                      expend     salary     ratio    takers
# 0.0000e+00 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 2.0250e-15 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 4.0500e-15 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 6.0750e-15 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 8.1000e-15 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 1.0125e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 1.2150e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 1.4175e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 1.6200e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 1.8225e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 2.0250e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
```

```
# 2.2275e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 2.4300e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 2.6325e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 2.8350e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 3.0375e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 3.2400e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 3.4425e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 3.6450e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 3.8475e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481
# 4.0500e-14 1045.972 4.462594 1.637917 -3.624232 -2.904481


matplot(rgmod$lambda, coef(rgmod), type="l",
        xlab=expression(lambda),ylab=expression(hat(beta)),
        main = 'Ridge trace plot: SAT')
abline(v=4.05e-14)
```



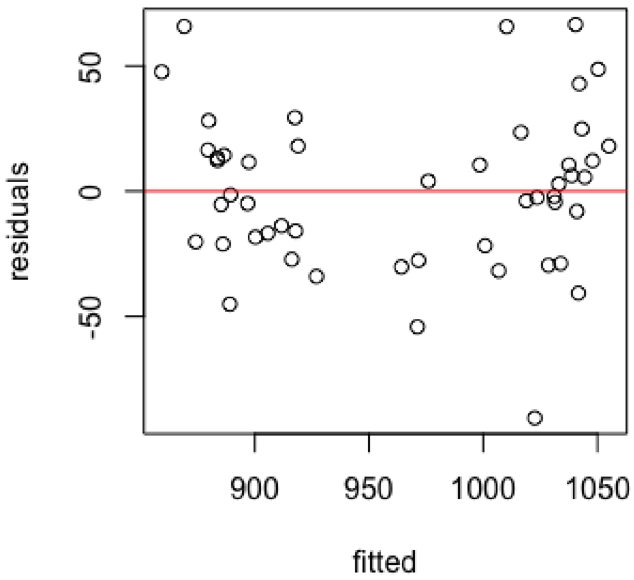

```
#part g
#install.packages("lars")
library(lars)

lmod <- lars(as.matrix(sat[1:4]), sat$total); plot(lmod)
cvlmod <- cv.lars(as.matrix(sat[1:4]), sat$total)
cvlmod$index[which.min(cvlmod$cv)]
# 0.9494949
```



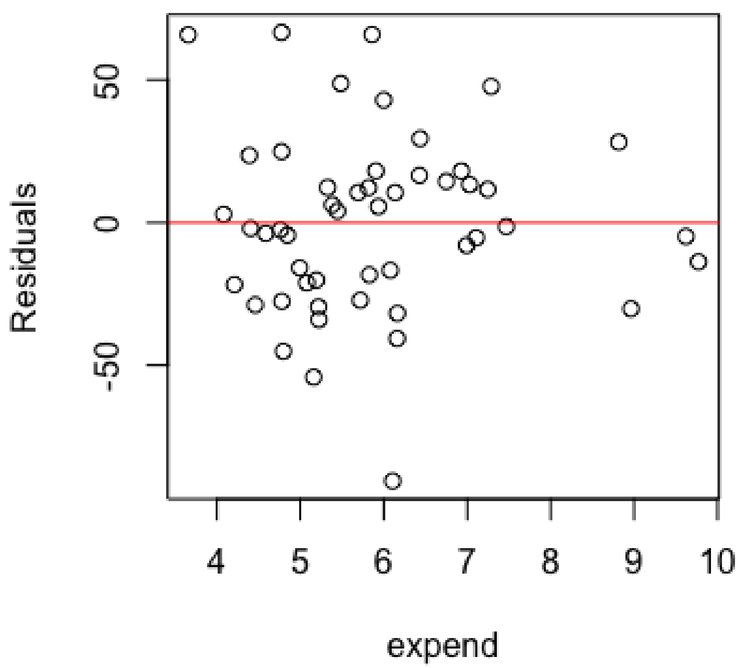



```
predict(lmod,s=0.9494949,type="coef",mode="fraction")$coef
# expend     | ratio     | salary    | takers
# 5.732724   | -2.900450 | 1.084311  | -2.831963

#Problem 3
data(sat)
lmod <- lm(total ~ expend + salary + ratio + takers, sat)

#a
plot(fitted(lmod), residuals(lmod),xlab="fitted",ylab="residuals")
abline(h=0, col="red")
```

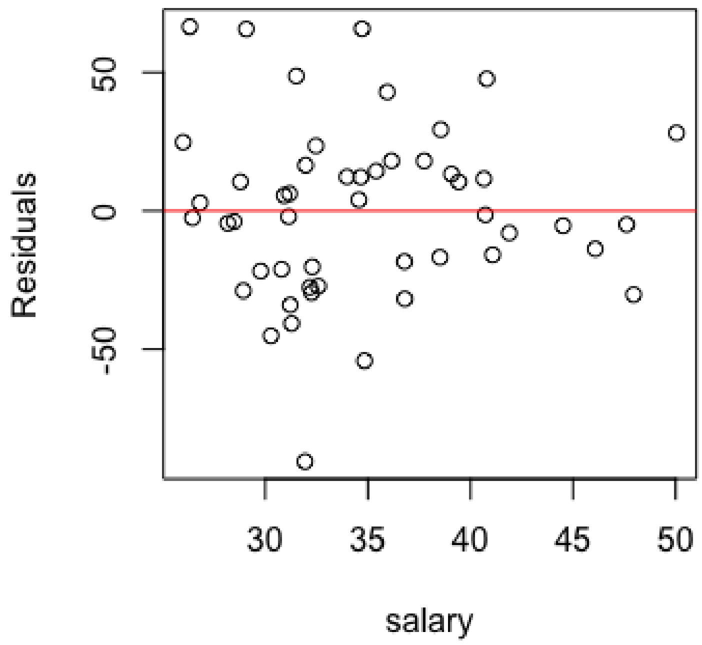
```
#It looks like a nonlinearity pattern

#part b
plot(fitted(lmod),sqrt(abs(residuals(lmod))),
xlab="Fitted",ylab=expression(sqrt(hat(epsilon))))
```



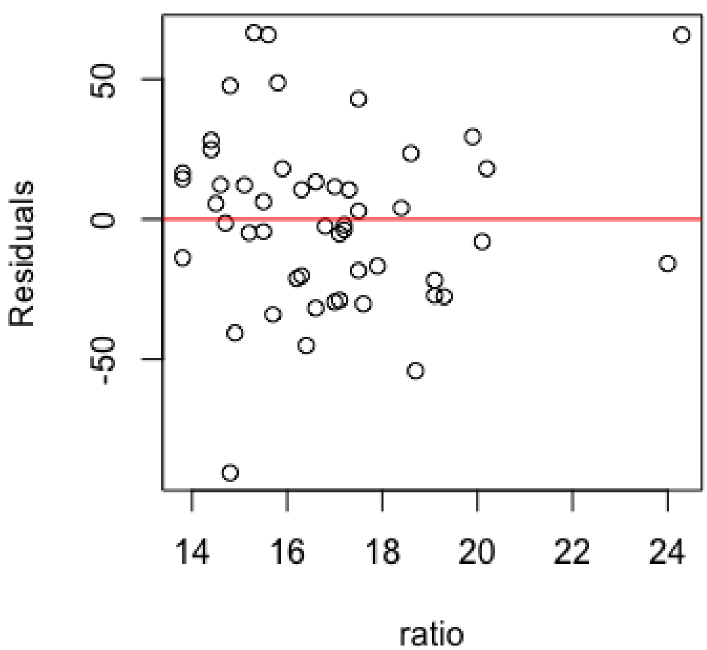

```
plot(sat$expend,residuals(lmod),xlab="expend",ylab="Residuals")
abline(h=0, col="red")
```



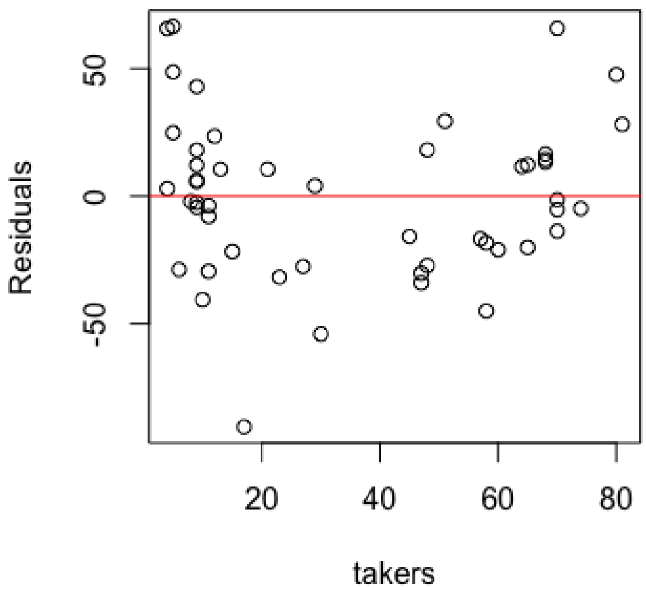
```
plot(sat$salary,residuals(lmod),xlab="salary",ylab="Residuals")
abline(h=0, col="red")
```




```
plot(sat$ratio,residuals(lmod),xlab="ratio",ylab="Residuals")
abline(h=0, col="red")
```

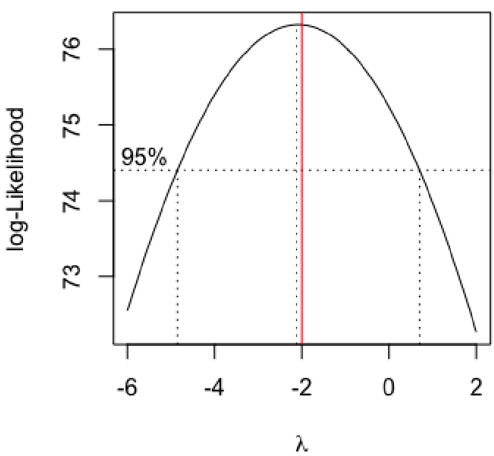


```
plot(sat$takers,residuals(lmod),xlab="takers",ylab="Residuals")
abline(h=0, col="red")
```

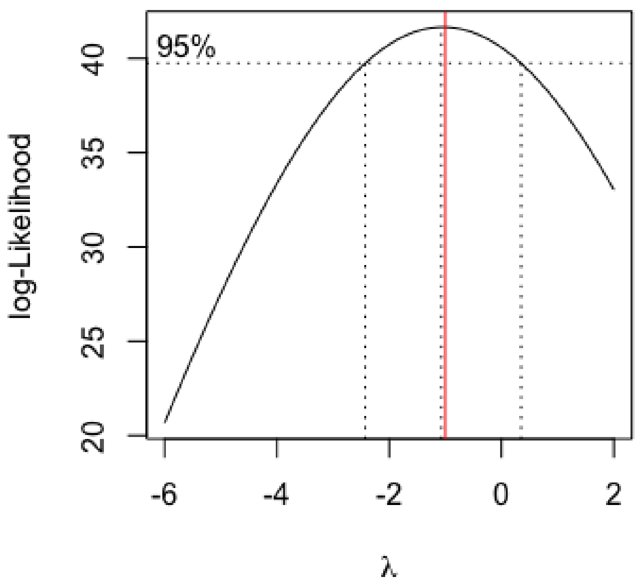



```
#salary is the most appropiate for these data acording to it's plot.
```

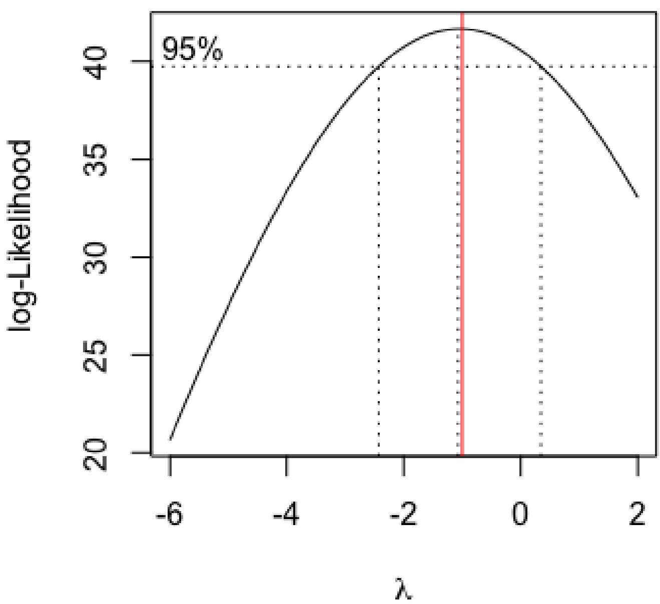
```
#c
boxcox(lmod,lambda=seq(-6,2,.1))
abline(v=-2, col = "red")
```



```
lmod1 <- lm(total^2 ~ expend + salary + ratio + takers, sat)
boxcox(lmod1,lambda = seq(-6,2,.1), plotit = TRUE);
abline(v=-1, col = "red")
```




```
lmod2 <- lm(sqrt(total) ~ expend + salary + ratio + takers, sat)
boxcox(lmod1,lambda = seq(-6,2,.1), plotit = TRUE);
abline(v=-1, col = "red")
```



95%
log-Likelihood
20
25
30
35
40
-6
-4
-2
0
2
λ