STAT3008 Exercise 11 Solutions

(2011-2012 2nd Semester)

Q1. (9.2)

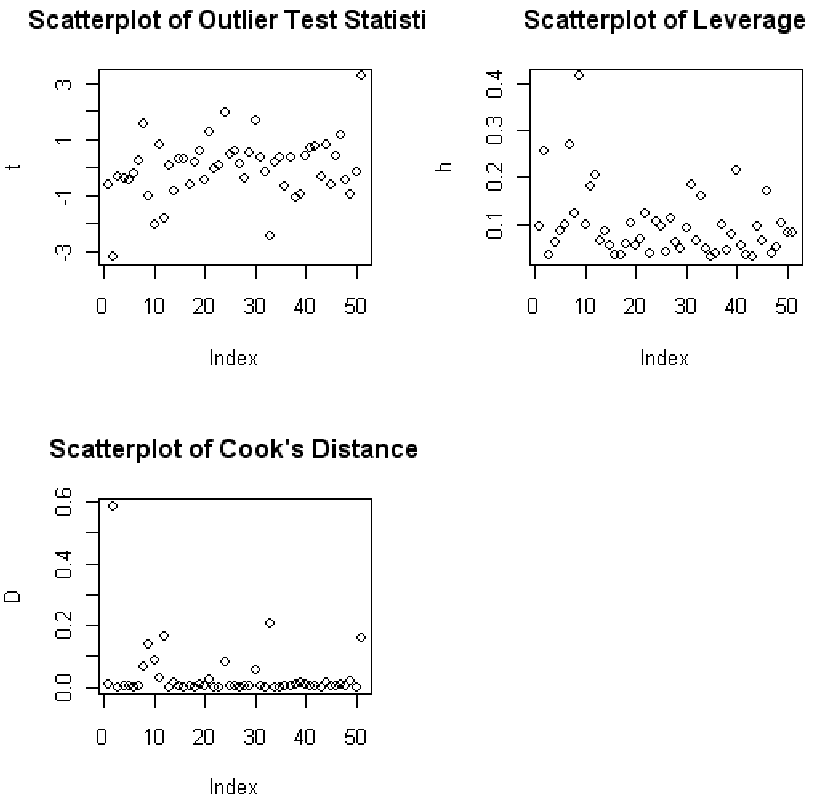
R Codes:

```
library(alr3)
f=c(2,33,12,51,9)
Fuel=1000*fuel2001$FuelC/fuel2001$Pop
Dlic=1000*fuel2001$Drivers/fuel2001$Pop
LogMiles=log2(fuel2001$Miles)
Tax=fuel2001$Tax
Income=fuel2001$Income
m=lm(Fuel~Tax+Dlic+Income+LogMiles)
D=cooks.distance(m)
h=hatvalues(m)
r=rstandard(m)
t=rstudent(m)
Names=c("Alaska","NewYork","Hawaii","Wyoming","D.C.")
Output=cbind(r,t,h,D)[f,]
row.names(Output)=Names
qt(1-0.025/51,45)
par(mfrow=c(2,2))
plot(t,main="Scatterplot of Outlier Test Statistic")
plot(h,main="Scatterplot of Leverage")
plot(D,main="Scatterplot of Cook's Distance")
```

	r	t	h	D
Alaska	-2.9149388	-3.1930222	0.25609617	0.5850260
New York	-2.3168723	-2.4382246	0.16237155	0.2081099
Hawaii	-1.7707873	-1.8143653	0.20572692	0.1624367
Wyoming	2.9542542	3.2460899	0.08378222	0.1596169
D.C.	-0.9962922	-0.9962102	0.41491327	0.1407798

$t_{\left(\frac{0.025}{51},45\right)} = 3.5270$

By using R, we can get the required values. As the critical value is 3.527 and all the t-statistics are less than 3.527, there are no outliers. Also, D.C. has the largest laverage, thus, D.C. has the largest influence in regression.



Q2. (9.3)

$$\left(X_{(i)}^T X_{(i)}\right) = X^T X - x_i x_i^T$$

$$\left[\left(X^T X\right)^{-1} + \frac{\left(X^T X\right)^{-1} x_i x_i^T \left(X^T X\right)^{-1}}{1-h_{ii}}\right]\left(X^T X - x_i x_i^T\right)$$

$$= I + \frac{\left(X^T X\right)^{-1} x_i x_i^T \left(X^T X\right)^{-1}}{1-h_{ii}}\left(X^T X - x_i x_i^T\right) - \left(X^T X\right)^{-1} x_i x_i^T$$

$$= I + \frac{1}{1-h_{ii}}\left[\left(X^T X\right)^{-1} x_i x_i^T - \left(X^T X\right)^{-1} x_i x_i^T \left(X^T X\right)^{-1} x_i x_i^T - (1-h_{ii})\left(X^T X\right)^{-1} x_i x_i^T\right]$$

$$= I + \frac{1}{1-h_{ii}}\left[\left(X^T X\right)^{-1} x_i x_i^T - \left(X^T X\right)^{-1} x_i h_{ii} x_i^T - (1-h_{ii})\left(X^T X\right)^{-1} x_i x_i^T\right]$$

$$= I + \frac{\left(X^T X\right)^{-1} x_i x_i^T}{1-h_{ii}}\left[1-h_{ii} - (1-h_{ii})\right] = I$$

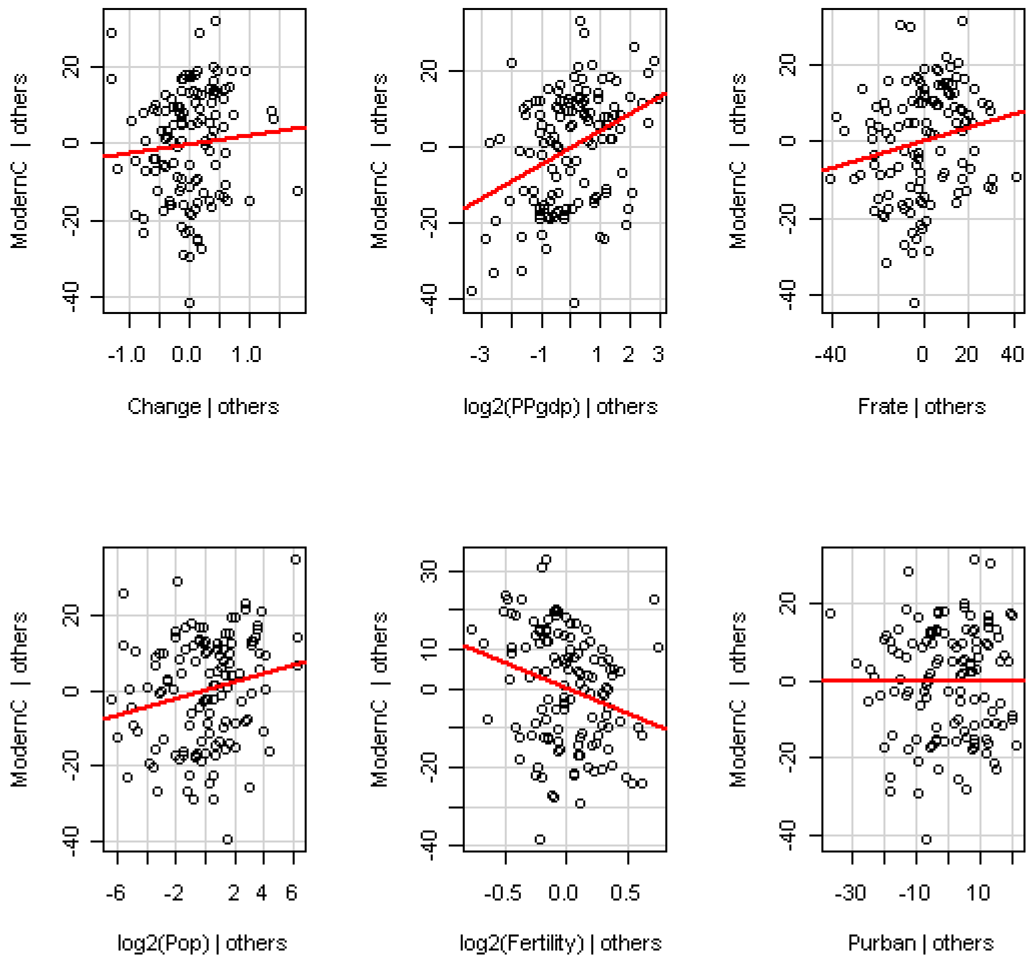
$$\therefore \left(X_{(i)}^T X_{(i)}\right)^{-1} = \left(X^T X\right)^{-1} + \frac{\left(X^T X\right)^{-1} x_i x_i^T \left(X^T X\right)^{-1}}{1-h_{ii}}$$

Q3. (9.9)

9.9.1

R Codes:

```
library(car)
library(alr3)
m1=lm(ModernC~Change+log2(PPgdp)+Frate+log2(Pop)+log2(Fertility)+Purban,data=UN3)
par(mfrow=c(2,3))
for (name in attr(m1$terms,"term.labels")){
av.plot(m1,name,identify=FALSE)}
plot(cooks.distance(m1))
```



Separated cases at the right or left of an added-variable plot would indicate influence. However, no such cases appear in these plots. None of the localilies is overly influential.

9.9.2

R Codes:

```
T=rstudent(m1)
A=(abs(T)> qt(1-0.025/210,210-6-2))
```

There are no observations greater than the critical value so there is no outlier.