Homework 7

Yanyu Zheng yz2690

March 27, 2016

3. Chapter 10, problem 26

```
library("Sleuth3")
attach(ex1026)
fit11 = lm(Inhibit ~ UVB + Surface + UVB:Surface)
summary(fit11)
```

```
##
## Call:
## lm(formula = Inhibit ~ UVB + Surface + UVB:Surface)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.9722  -3.9444  -0.1806   1.4479  21.0278
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.181      4.292   0.275 0.787599
## UVB                 1226.389    232.773   5.269 0.000152 ***
## SurfaceSurface         1.278     11.066   0.115 0.909837
## UVB:SurfaceSurface -939.931    409.839  -2.293 0.039134 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.833 on 13 degrees of freedom
## Multiple R-squared:  0.7086, Adjusted R-squared:  0.6414
## F-statistic: 10.54 on 3 and 13 DF,  p-value: 0.000868
```

From the summary, we can see only UVB and UVB*Surface are significant. So we build a new model with only this two variables.

```
fit12 = lm(Inhibit ~ 0 + UVB + UVB:Surface)
summary(fit12)
```

```
##
## Call:
## lm(formula = Inhibit ~ 0 + UVB + UVB:Surface)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.2500  -3.2500   0.1016   1.6016  20.7500
##
```

```
## Coefficients: (1 not defined because of singularities)
##                    Estimate Std. Error t value Pr(>|t|)
## UVB                   363.3      103.3   3.516 0.003118 **
## UVB:SurfaceDeep       911.7      175.4   5.198 0.000108 ***
## UVB:SurfaceSurface       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.265 on 15 degrees of freedom
## Multiple R-squared:  0.8615, Adjusted R-squared:  0.843
## F-statistic: 46.63 on 2 and 15 DF,  p-value: 3.647e-07
```

```
detach(ex1026)
```

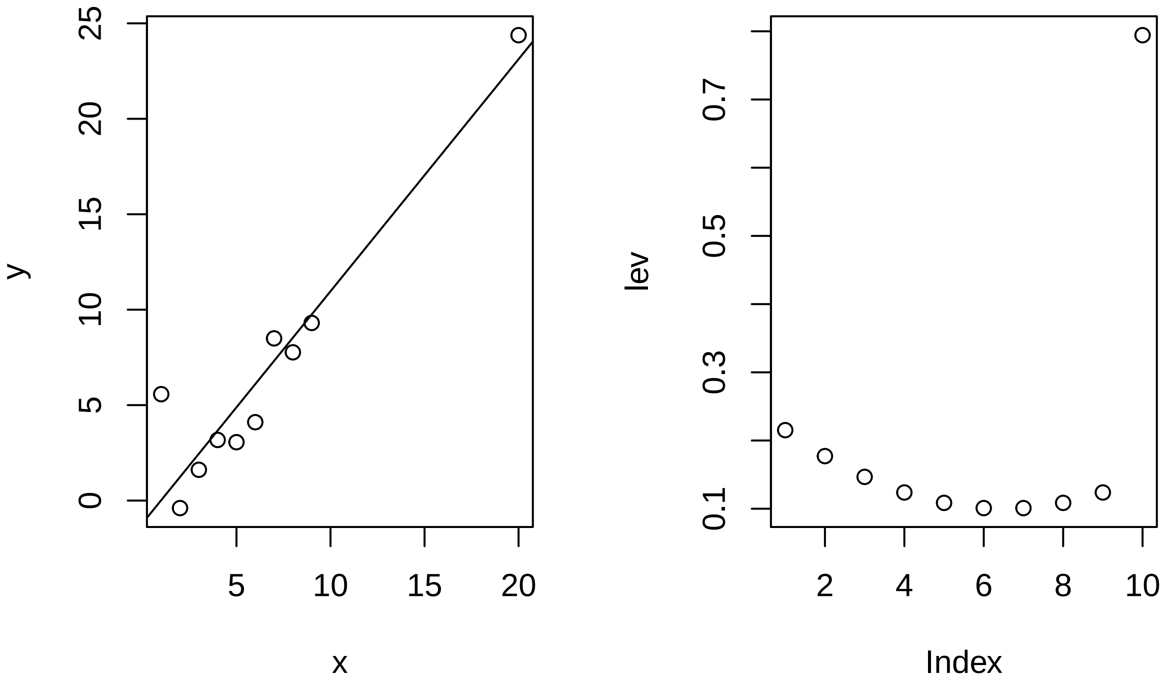
Yes, the effect is different. The difference is about 939.931, which means the UVB change for 1, the change in Inhibit is 911.7 higher in surface than in deep sea.

4. Chapter 11, problem 8

A case with large leverage has a residual with low variability. Because its explana- tory variable values are so unusual, it dictates the location of the estimated regression over the whole region in its vicinity; no other points in the region share the responsibility. Because its residual must be small, this case acts like a mag- net on the estimated regression surface. If, however, its response falls close to the regression surface (as determined by the remaining observations alone), it is not necessarily influential. Therefore, while a large leverage does not necessarily indi- cate that the case is influential, it does imply that the case has a high potential for influence.

plot for (b)

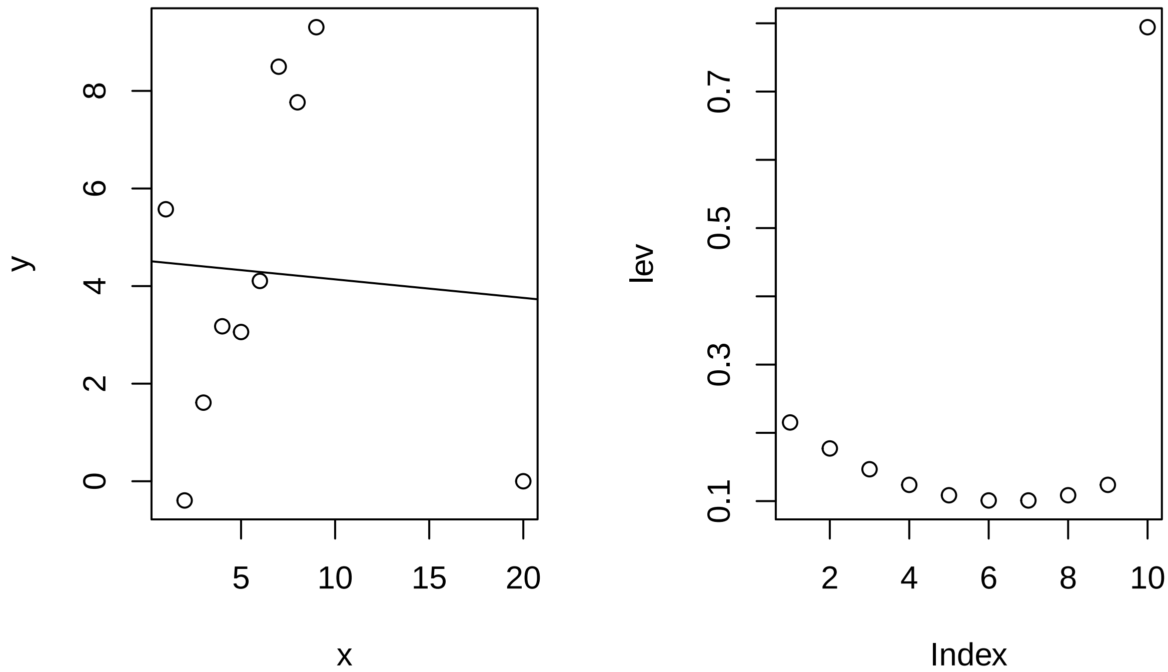
```
set.seed(7)
x = c(1:9, 20)
y = x + rnorm(10, sd = 2)
par(mfrow=c(1,2))
plot(x,y)
fit41 = lm(y~x)
abline(fit41)
lev = hat(model.matrix(fit41))
plot(lev)
```



From the plot we can see the last observation has high leverage but no substantial influence on the model.

Plot for (c)

```
set.seed(7)
x = c(1:9, 20)
y = x + rnorm(10, sd = 2)
y[10] = 0
par(mfrow=c(1,2))
plot(x,y)
fit42 = lm(y~x)
abline(fit42)
lev = hat(model.matrix(fit42))
plot(lev)
```



From the plot we can see the last observation has high leverage and completely changed the model.

5. Chapter 11, problem 16

```
attach(case1101)
fit5 = lm(Metabol ~ Sex * Gastric)
lev = hat(model.matrix(fit5))
stud = rstudent(fit5)
cook = cooks.distance(fit5)
detach(case1101)
```

For case 32, the leverage is 0.2528749, studentized residual is 5.1205163, cook's distance is 1.1672546

6. Chapter 11, problem 20

a

```
attach(ex1120)
fit61 = lm(Calcite ~ Carbonate)
data6 = ex1120[-which.min(Carbonate),]
fit62 = lm(Calcite ~ Carbonate, data = data6)
data6 = data6[-which.min(data6$Carbonate),]
fit63 = lm(Calcite ~ Carbonate, data = data6)
```

Model with all obs

```r
summary(fit61)
```

```
##
## Call:
## lm(formula = Calcite ~ Carbonate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46796 -0.64104 -0.04927  0.67301  1.55856
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.4984     3.1766  -0.472    0.644
## Carbonate     1.0703     0.1156   9.259 7.93e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9959 on 16 degrees of freedom
## Multiple R-squared:  0.8427, Adjusted R-squared:  0.8329
## F-statistic: 85.73 on 1 and 16 DF,  p-value: 7.929e-08
```

Model delete the smallest X

```r
summary(fit62)
```

```
##
## Call:
## lm(formula = Calcite ~ Carbonate, data = data6)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2799 -0.4816 -0.1364  0.7184  1.4871
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.6727     4.6247   0.578    0.572
## Carbonate     0.9217     0.1663   5.541 5.65e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9807 on 15 degrees of freedom
## Multiple R-squared:  0.6718, Adjusted R-squared:  0.6499
## F-statistic:  30.7 on 1 and 15 DF,  p-value: 5.653e-05
```

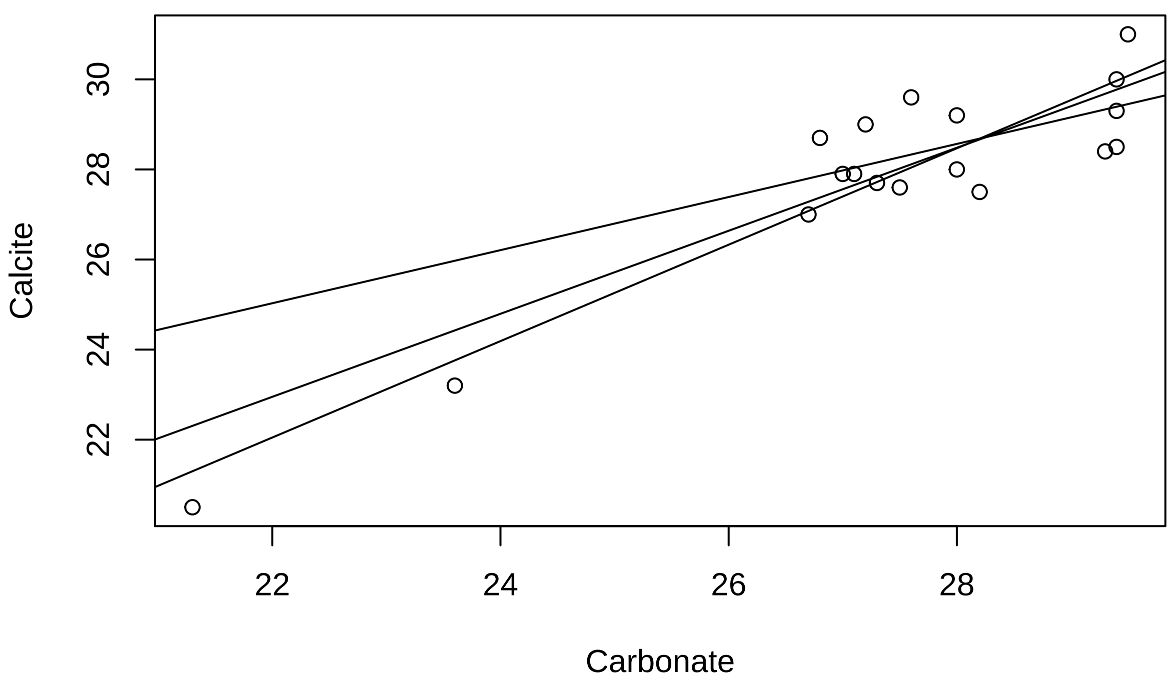
Model delete the smallest two X

```r
summary(fit63)
```

```
##
## Call:
## lm(formula = Calcite ~ Carbonate, data = data6)
##
## Residuals:
##      Min       1Q  Median       3Q      Max
## -1.1844 -0.7038 -0.1139  0.6854  1.5492
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.0589     6.1592   1.958   0.0705 .
## Carbonate     0.5896     0.2196   2.684   0.0178 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8875 on 14 degrees of freedom
## Multiple R-squared:  0.3398, Adjusted R-squared:  0.2926
## F-statistic: 7.205 on 1 and 14 DF,  p-value: 0.0178
```

b

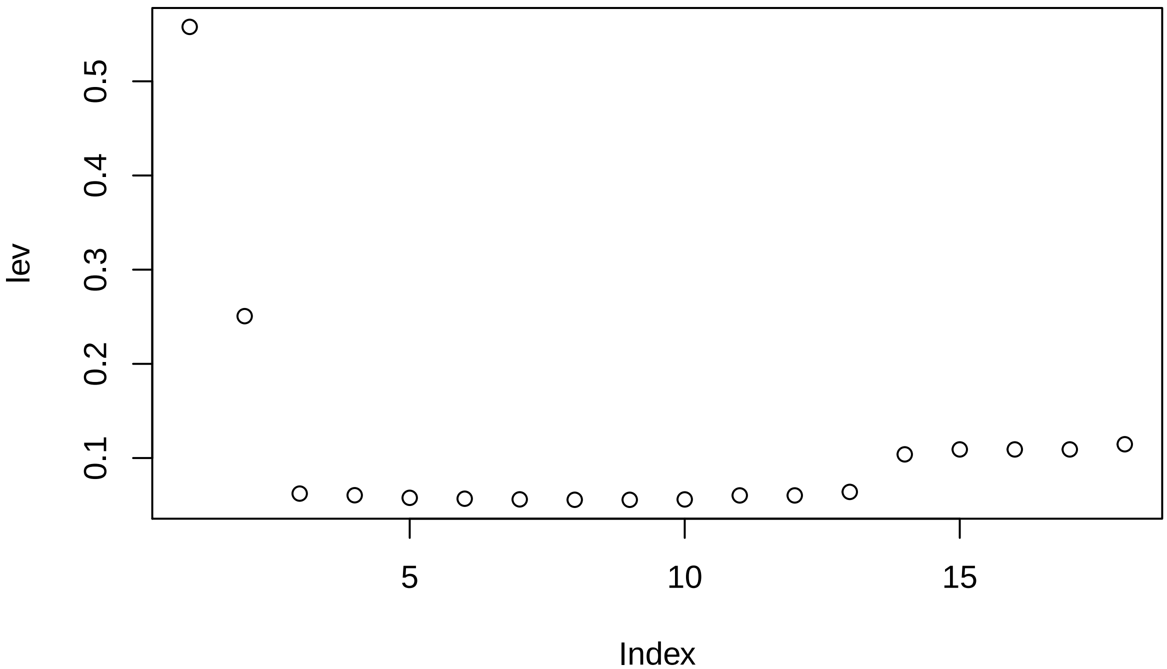
```r
par(mfrow=c(1,1))
plot(Carbonate, Calcite)
abline(fit61)
abline(fit62)
abline(fit63)
```



From the plot, we can see the points deleted have substantial influence on the regression line. Thus the change in R squared.

c

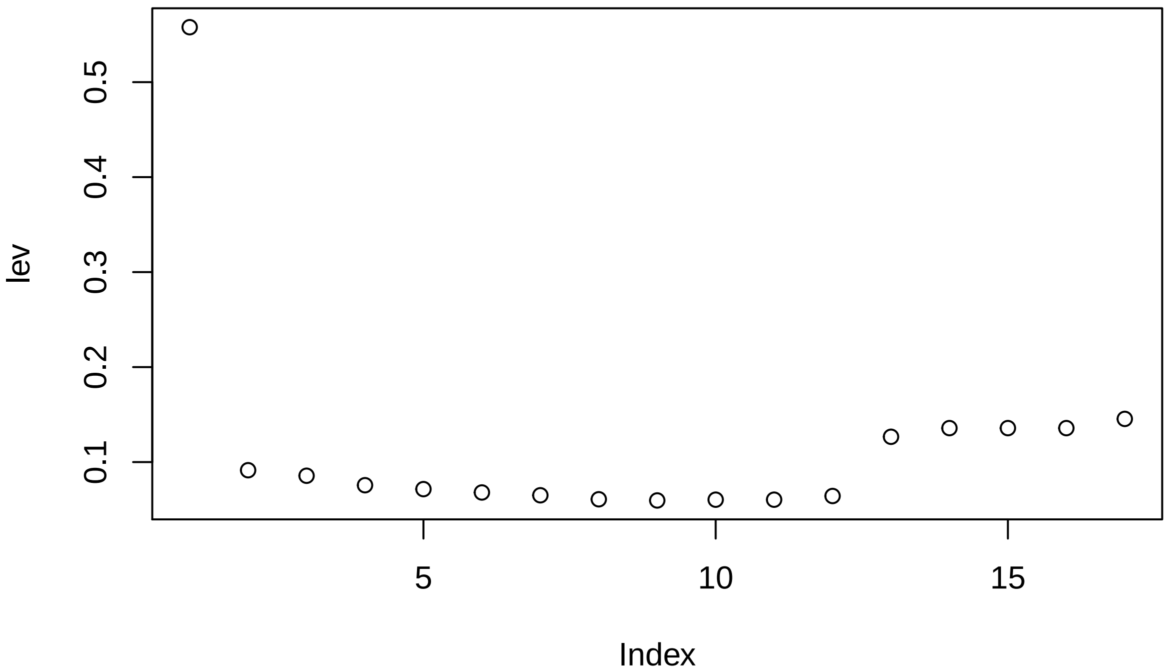
```
lev = hat(model.matrix(fit61))
plot(lev)
```



From the plot, we can see the two smallest X have high leverage 0.56 and 0.25.

d

```
lev = hat(model.matrix(fit62))
plot(lev)
```



From the plot, we can see the leverage of the second smallest X became 0.56 instead of 0.25.

e

From the above computation, we can see one single observation with high would make all other points' leverage smaller. So pairs of influential observations are not commanly found since the most influential observation tend to "mask" other less influential ones.

f

The two cases have substantial influence on the model.