```
> library(foreign)
> library(MASS)
> cdata <- read.dta("http://www.ats.ucla.edu/stat/data/crime.dta")
> summary(cdata)

      sid          state               crime            murder
 Min.   : 1.0   Length:51          Min.   :  82.0   Min.   : 1.600
 1st Qu.:13.5   Class :character   1st Qu.: 326.5   1st Qu.: 3.900
 Median :26.0   Mode  :character   Median : 515.0   Median : 6.800
 Mean   :26.0                      Mean   : 612.8   Mean   : 8.727
 3rd Qu.:38.5                      3rd Qu.: 773.0   3rd Qu.:10.350
 Max.   :51.0                      Max.   :2922.0   Max.   :78.500
    pctmetro         pctwhite         pcths          poverty
 Min.   : 24.00   Min.   :31.80   Min.   :64.30   Min.   : 8.00
 1st Qu.: 49.55   1st Qu.:79.35   1st Qu.:73.50   1st Qu.:10.70
 Median : 69.80   Median :87.60   Median :76.70   Median :13.10
 Mean   : 67.39   Mean   :84.12   Mean   :76.22   Mean   :14.26
 3rd Qu.: 83.95   3rd Qu.:92.60   3rd Qu.:80.10   3rd Qu.:17.40
 Max.   :100.00   Max.   :98.50   Max.   :86.60   Max.   :26.40
     single
 Min.   : 8.40
 1st Qu.:10.05
 Median :10.90
 Mean   :11.33
 3rd Qu.:12.05
 Max.   :22.10

> # Oridinary Linear Regression
> summary(ols <- lm(crime ~ poverty + single, data = cdata))

Call:
lm(formula = crime ~ poverty + single, data = cdata)

Residuals:
    Min      1Q  Median      3Q     Max
-811.14 -114.27  -22.44  121.86  689.82

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1368.189    187.205  -7.308 2.48e-09 ***
poverty         6.787      8.989   0.755    0.454
single        166.373     19.423   8.566 3.12e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 243.6 on 48 degrees of freedom
```

```
Multiple R-squared:  0.7072,        Adjusted R-squared:  0.695
F-statistic: 57.96 on 2 and 48 DF,  p-value: 1.578e-13
```

```
> opar <- par(mfrow = c(2,2), oma = c(0, 0, 1.1, 0))
> plot(ols, las = 1)
> cdata[c(9, 25, 51), 1:2]
```

```
   sid state
9    9    fl
25  25    ms
51  51    dc
```

```
> d1 <- cooks.distance(ols)
> r <- stdres(ols)
> a <- cbind(cdata, d1, r)
> a[d1 > 4/51, ]
```

```
   sid state crime murder pctmetro pctwhite pcths poverty single        d1
1    1    ak   761    9.0     41.8     75.2  86.6     9.1   14.3 0.1254750
9    9    fl  1206    8.9     93.0     83.5  74.4    17.8   10.6 0.1425891
25  25    ms   434   13.5     30.7     63.3  64.3    24.7   14.7 0.6138721
51  51    dc  2922   78.5    100.0     31.8  73.1    26.4   22.1 2.6362519
           r
1  -1.397418
9   2.902663
25 -3.562990
51  2.616447
```

```
> rabs <- abs(r)
> a <- cbind(cdata, d1, r, rabs)
> asorted <- a[order(-rabs), ]
> asorted[1:10, ]
```

```
   sid state crime murder pctmetro pctwhite pcths poverty single         d1
25  25    ms   434   13.5     30.7     63.3  64.3    24.7   14.7 0.61387212
9    9    fl  1206    8.9     93.0     83.5  74.4    17.8   10.6 0.14258909
51  51    dc  2922   78.5    100.0     31.8  73.1    26.4   22.1 2.63625193
46  46    vt   114    3.6     27.0     98.4  80.8    10.0   11.0 0.04271548
26  26    mt   178    3.0     24.0     92.6  81.0    14.9   10.8 0.01675501
21  21    me   126    1.6     35.7     98.5  78.8    10.7   10.6 0.02233128
1    1    ak   761    9.0     41.8     75.2  86.6     9.1   14.3 0.12547500
31  31    nj   627    5.3    100.0     80.8  76.7    10.9    9.6 0.02229184
14  14    il   960   11.4     84.0     81.0  76.2    13.6   11.5 0.01265689
20  20    md   998   12.7     92.8     68.9  78.4     9.7   12.0 0.03569623
           r     rabs
25 -3.562990 3.562990
9   2.902663 2.902663
```

```
51  2.616447 2.616447
46 -1.742409 1.742409
26 -1.460885 1.460885
21 -1.426741 1.426741
1  -1.397418 1.397418
31  1.354149 1.354149
14  1.338192 1.338192
20  1.287087 1.287087
```

```
> summary(rr.huber <- rlm(crime ~ poverty + single, data = cdata))
```

```
Call: rlm(formula = crime ~ poverty + single, data = cdata)
Residuals:
    Min      1Q  Median      3Q     Max
-846.09 -125.80  -16.49  119.15  679.94

Coefficients:
            Value      Std. Error t value
(Intercept) -1423.0373   167.5899    -8.4912
poverty         8.8677     8.0467     1.1020
single        168.9858    17.3878     9.7186

Residual standard error: 181.8 on 48 degrees of freedom
```

```
> hweights <- data.frame(state = cdata$state, resid = rr.huber$resid, weight = rr.huber$w)
> hweights2 <- hweights[order(rr.huber$w), ]
> hweights2[1:15, ]
```

```
   state      resid    weight
25    ms -846.08536 0.2889618
9     fl  679.94327 0.3595480
46    vt -410.48310 0.5955740
51    dc  376.34468 0.6494131
26    mt -356.13760 0.6864625
21    me -337.09622 0.7252263
31    nj  331.11603 0.7383578
14    il  319.10036 0.7661169
1     ak -313.15532 0.7807432
20    md  307.19142 0.7958154
19    ma  291.20817 0.8395172
18    la -266.95752 0.9159411
2     al  105.40319 1.0000000
3     ar   30.53589 1.0000000
4     az  -43.25299 1.0000000
```

```
> #bisquare weights
> rr.bisquare <- rlm(crime ~ poverty + single, data=cdata, psi = psi.bisquare)
> summary(rr.bisquare)
```

```
Call: rlm(formula = crime ~ poverty + single, data = cdata, psi = psi.bisquare)
Residuals:
    Min      1Q  Median      3Q     Max
-905.59 -140.97  -14.98  114.65  668.38

Coefficients:
            Value      Std. Error t value
(Intercept) -1535.3338   164.5062    -9.3330
poverty        11.6903     7.8987     1.4800
single        175.9303    17.0678    10.3077

Residual standard error: 202.3 on 48 degrees of freedom
```

```
> biweights <- data.frame(state = cdata$state, resid = rr.bisquare$resid, weight = rr.bisqua
> biweights2 <- biweights[order(rr.bisquare$w), ]
> biweights2[1:15, ]
```

```
   state     resid      weight
25    ms -905.5931 0.007652565
9     fl  668.3844 0.252870542
46    vt -402.8031 0.671495418
26    mt -360.8997 0.731136908
31    nj  345.9780 0.751347695
18    la -332.6527 0.768938330
21    me -328.6143 0.774103322
1     ak -325.8519 0.777662383
14    il  313.1466 0.793658594
20    md  308.7737 0.799065530
19    ma  297.6068 0.812596833
51    dc  260.6489 0.854441716
50    wy -234.1952 0.881660897
5     ca  201.4407 0.911713981
10    ga -186.5799 0.924033113
```

```
> #Confidence Interval Example
> library(DAAG)
> roller.lm <- lm(depression~weight, data=roller)
> summary(roller.lm)
```

```
Call:
lm(formula = depression ~ weight, data = roller)

Residuals:
   Min     1Q Median     3Q    Max
-8.180 -5.580 -1.346  5.920  8.020

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.0871     4.7543  -0.439  0.67227
weight        2.6667     0.7002   3.808  0.00518 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.735 on 8 degrees of freedom
Multiple R-squared:  0.6445,        Adjusted R-squared:  0.6001
F-statistic:  14.5 on 1 and 8 DF,  p-value: 0.005175

> #Confidence Interval Calculation
> SEb <- summary(roller.lm)$coefficients[2, 2]
> coef(roller.lm)[2] + qt(c(0.025,.975), 8)*SEb

[1] 1.051984 4.281508

> #fitted values and standard Errors
> fit.with.se <- predict(roller.lm, se.fit=TRUE)
> fit.with.se$se.fit

 [1] 3.614297 2.976896 2.880798 2.308147 2.197133 2.130050 2.142445 2.384221
 [9] 3.370270 4.917728

> # SE
> print(sqrt(fit.with.se$se.fit^2+fit.with.se$residual.scale^2))

 [1] 7.643943 7.364009 7.325689 7.119990 7.084781 7.064265 7.068012 7.145014
 [9] 7.531629 8.339710

> #plot
> plot(depression~weight, data=roller, xlab = "Weight of Roller (tonnes)",
+ ylab = "Depression in Lawn (mm)", pch = 16)
> roller.lm <- lm(depression~weight, data = roller)
> abline(roller.lm$coef, lty = 1)
> #Cross Validation Example
> rand <- sample(1:15)%%3 + 1
> (1:15)[rand == 1] # Observation numbers for the first group

[1]  2  4  7 10 12

> (1:15)[rand == 2] # Observation numbers for the second group

[1]  1  8 11 14 15

> (1:15)[rand == 3] # Observation numbers for the third group

[1]  3  5  6  9 13
```

```
> houseprices.lm <- lm(sale.price~area, data=houseprices)
> CVlm(houseprices, houseprices.lm, plotit=TRUE)

Analysis of Variance Table

Response: sale.price
          Df Sum Sq Mean Sq F value Pr(>F)
area       1  18566   18566       8  0.014 *
Residuals 13  30179    2321
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


fold 1
Observations in test set: 5
             11  20    21     22   23
area        802 696 771.0 1006.0 1191
cvpred      204 188 199.3  234.7  262
sale.price  215 255 260.0  293.0  375
CV residual  11  67  60.7   58.3  113

Sum of squares = 24351    Mean square = 4870    n = 5

fold 2
Observations in test set: 5
             10   13    14      17     18
area        905  716 963.0 1018.00 887.00
cvpred      255  224 264.4  273.38 252.06
sale.price  215  113 185.0  276.00 260.00
CV residual -40 -112 -79.4    2.62   7.94

Sum of squares = 20416    Mean square = 4083    n = 5

fold 3
Observations in test set: 5
                9   12     15    16     19
area        694.0 1366 821.00 714.0 790.00
cvpred      183.2  388 221.94 189.3 212.49
sale.price  192.0  274 212.00 220.0 221.50
CV residual   8.8 -114  -9.94  30.7   9.01

Sum of squares = 14241    Mean square = 2848    n = 5

Overall (Sum over all 5 folds)
  ms
3934
```

```
> #model-based estimate in the regression output for the total data
> summary(houseprices.lm)$sigma^2

[1] 2321
```