Influence Examples

In this example we use two small data sets to illustrate influence diagnostics.

The primary influence diagnostic is Cook's distance. We can get Cook's D and several other diagnostics using influence.measures(). A common rule of thumb says Cook'D > 1 (or > 0.5) indicated an influential observation.

Using the plot() function we can get diagnostic plots including Cook's distance and leverage.

Example 1

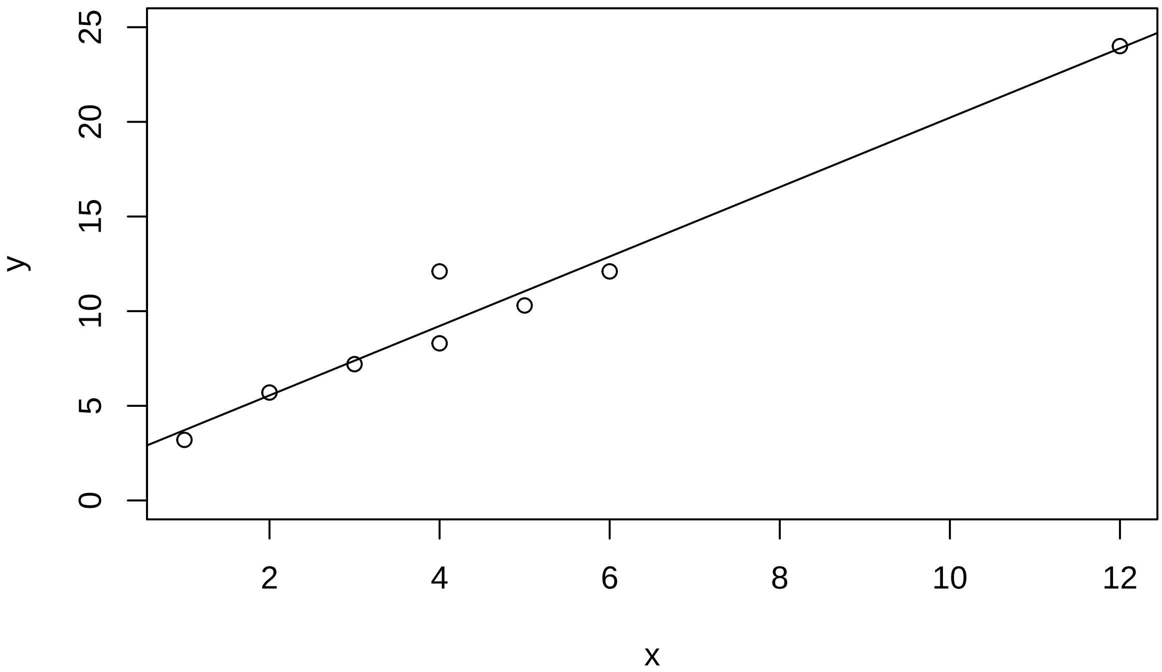
```
library(MASS)
library(car)
```

```
## Loading required package: carData
```

```
Ex1 <- read.csv("~/Dropbox/STAT512/Lectures/MultReg4/MR4_Influence1.csv")
Ex1
```

```
##    x    y
## 1  1  3.2
## 2  2  5.7
## 3  3  7.2
## 4  4  8.3
## 5  4 12.1
## 6  5 10.3
## 7  6 12.1
## 8 12 24.0
```

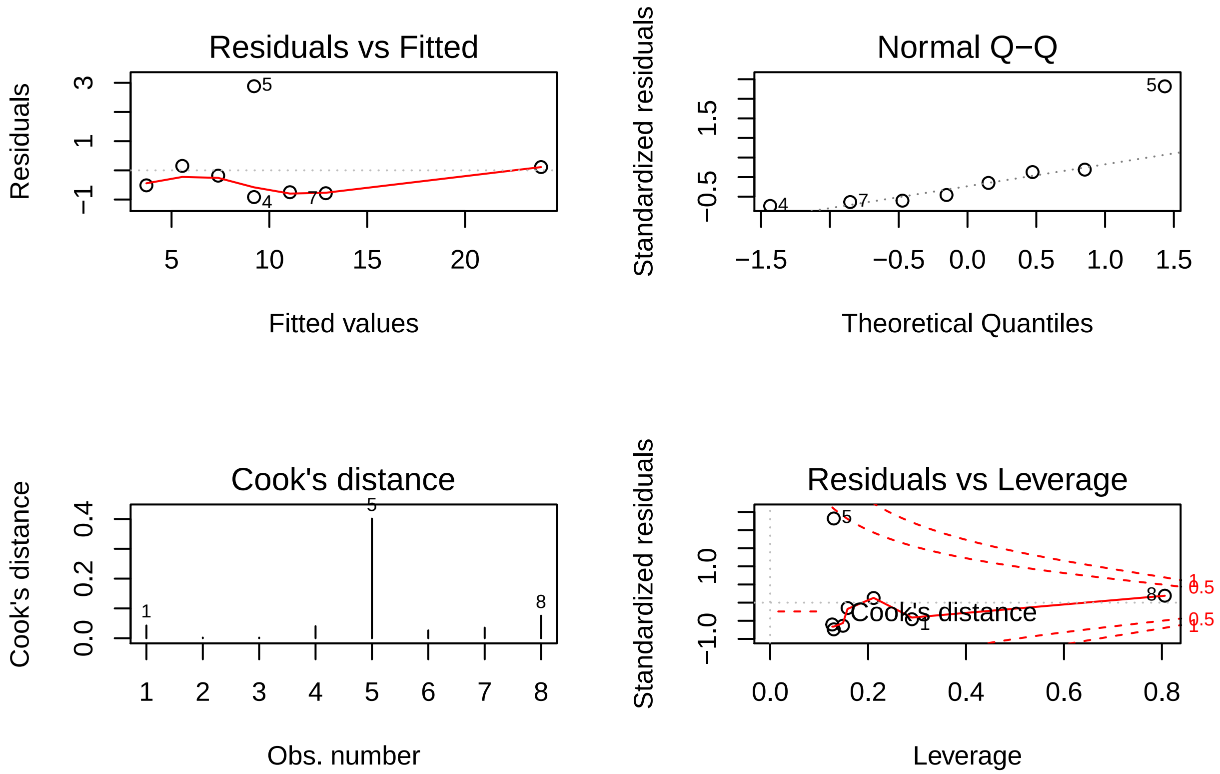
```
Model1 <- lm(y ~ x,data = Ex1)
plot(y ~ x,  ylim = c(0,25), data = Ex1)
abline(Model1)
```




```r
summary(Model1)
```

```
##
## Call:
## lm(formula = y ~ x, data = Ex1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9163 -0.7587 -0.3484  0.1220  2.8837
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.8804     0.8357    2.25   0.0654 .
## x             1.8340     0.1492   12.29 1.77e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 6 degrees of freedom
## Multiple R-squared:  0.9618, Adjusted R-squared:  0.9554
## F-statistic: 151.1 on 1 and 6 DF,  p-value: 1.766e-05
```

```r
par(mfrow=c(2,2))
plot(Model1, which = c(1:2,4:5))
```




```r
influence.measures(Model1)
```

```
## Influence measures of
##   lm(formula = y ~ x, data = Ex1) :
##
##    dfb.1_   dfb.x   dffit  cov.r  cook.d   hat inf
## 1 -0.2696  0.2047 -0.2715 1.8877 0.04268 0.290
## 2  0.0583 -0.0387  0.0606 1.8157 0.00220 0.211
## 3 -0.0519  0.0270 -0.0591 1.6977 0.00208 0.158
## 4 -0.1944  0.0529 -0.2725 1.3691 0.04051 0.130
## 5  1.8096 -0.4920  2.5358 0.0179 0.40126 0.130   *
## 6 -0.1000 -0.0255 -0.2161 1.4558 0.02632 0.127
## 7 -0.0473 -0.1005 -0.2518 1.4701 0.03548 0.149
## 8 -0.1911  0.3272  0.3560 7.3309 0.07558 0.806   *
```

```r
Resids1 <- data.frame(RawRes = resid(Model1), StdRes = stdres(Model1), RStdRes = rstudent(Model1))
Resids1
```

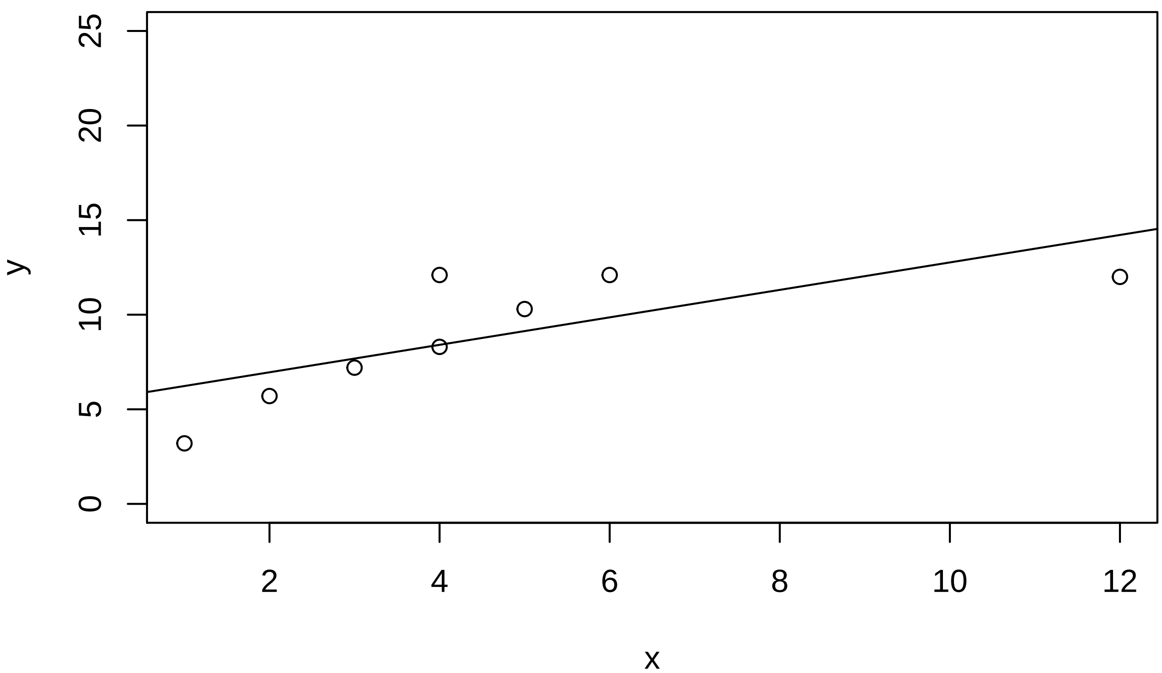
```
##       RawRes     StdRes    RStdRes
## 1 -0.5143975 -0.4577025 -0.4253142
## 2  0.1516432  0.1280619  0.1170641
## 3 -0.1823161 -0.1490206 -0.1362890
## 4 -0.9162754 -0.7367175 -0.7051787
## 5  2.8837246  2.3186154  6.5631678
## 6 -0.7502347 -0.6021330 -0.5670699
## 7 -0.7841941 -0.6374358 -0.6026606
## 8  0.1120501  0.1907716  0.1746804
```

Example 2

```r
Ex2 <- read.csv("~/Dropbox/STAT512/Lectures/MultReg4/MR4_Influence2.csv", header=TRUE)
Ex2
```

```
##    x    y
## 1  1  3.2
## 2  2  5.7
## 3  3  7.2
## 4  4  8.3
## 5  4 12.1
## 6  5 10.3
## 7  6 12.1
## 8 12 12.0
```

```r
Model2 <- lm(y ~ x, data = Ex2)
plot(y ~ x, ylim = c(0,25), data = Ex2)
abline(Model2)
```

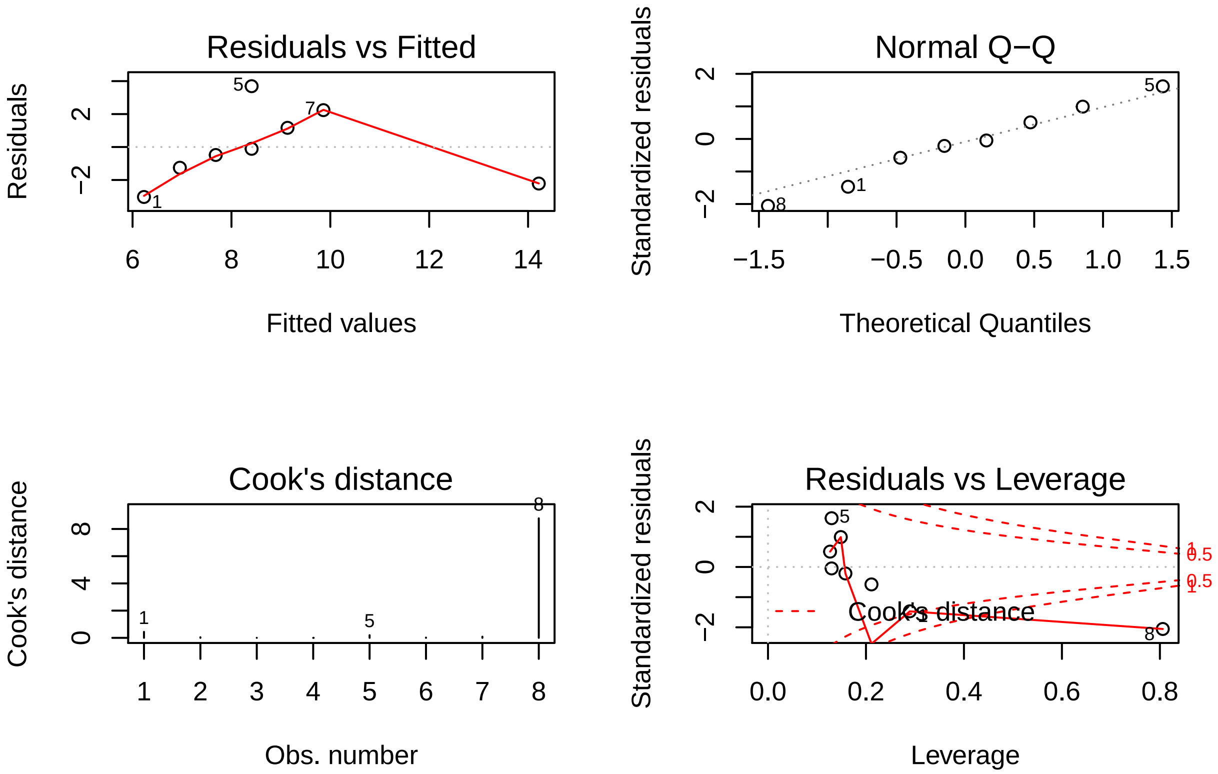



```r
summary(Model2)
```

```
##
## Call:
## lm(formula = y ~ x, data = Ex2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0308 -1.4968 -0.2958  1.4338  3.6912
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.5049     1.5335   3.590   0.0115 *
## x             0.7260     0.2738   2.652   0.0379 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.447 on 6 degrees of freedom
## Multiple R-squared:  0.5396, Adjusted R-squared:  0.4628
## F-statistic: 7.032 on 1 and 6 DF,  p-value: 0.03794
```

```r
par(mfrow=c(2,2))
plot(Model2, which = c(1:2,4:5))
```




```r
influence.measures(Model2)
```

```
## Influence measures of
##   lm(formula = y ~ x, data = Ex2) :
##
##    dfb.1_    dfb.x   dffit cov.r   cook.d   hat inf
## 1 -1.0630  0.80687 -1.0704 0.830 0.440003 0.290   *
## 2 -0.2704  0.17969 -0.2812 1.628 0.044800 0.211
## 3 -0.0751  0.03905 -0.0854 1.684 0.004340 0.158
## 4 -0.0120  0.00326 -0.0168 1.654 0.000169 0.130
## 5  0.5420 -0.14737  0.7595 0.527 0.195229 0.130
## 6  0.0839  0.02136  0.1812 1.509 0.018851 0.127
## 7  0.0777  0.16512  0.4138 1.182 0.085904 0.149
## 8  3.7803 -6.47311 -7.0422 0.646 8.782265 0.806   *
```

```r
Resids2 <- data.frame(RawRes = resid(Model2), StdRes = stdres(Model2), RStdRes = rstudent(Model2))
Resids2
```

```
##       RawRes      StdRes     RStdRes
## 1 -3.0308294 -1.46955159 -1.67679736
## 2 -1.2568075 -0.57836868 -0.54333921
```

```
## 3 -0.4827856 -0.21503762 -0.19706243
## 4 -0.1087637 -0.04765383 -0.04351003
## 5  3.6912363  1.61728188  1.96575742
## 6  1.1652582  0.50963159  0.47563628
## 7  2.2392801  0.99188254  0.99028264
## 8 -2.2165884 -2.05648471 -3.45554911
```

```r
#Fit the model with Obs 8 omitted
Model3 <- update(Model2, subset = -8)
summary(Model3)
```

```
##
## Call:
## lm(formula = y ~ x, data = Ex2, subset = -8)
##
## Residuals:
##        1        2        3        4        5        6        7
## -0.63548  0.08387 -0.19677 -0.87742  2.92258 -0.65806 -0.63871
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.055      1.353   1.519  0.18920
## x              1.781      0.346   5.147  0.00362 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.456 on 5 degrees of freedom
## Multiple R-squared:  0.8412, Adjusted R-squared:  0.8095
## F-statistic: 26.49 on 1 and 5 DF,  p-value: 0.003625
```

```r
beta0hat <- Model3$coef[[1]]
beta1hat <- Model3$coef[[2]]
beta0hat; beta1hat
```

```
## [1] 2.054839

## [1] 1.780645
```

```r
lht(Model2, rbind(c(1,0), c(0,1)), rhs=c(beta0hat, beta1hat))
```

```
## Linear hypothesis test
##
## Hypothesis:
## (Intercept) = 2.05483870967742
## x = 1.78064516129032
##
## Model 1: restricted model
## Model 2: y ~ x
##
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1      8 141.077
## 2      6  35.921  2    105.16 8.7823 0.01651 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow=c(1,1))
plot(y ~ x, ylim = c(0,25), data = Ex2)
```

```r
abline(Model2, lty = 1)
abline(Model3, lty = 2)
legend("topleft", lty = c(1,2), c("All Data", "Remove Obs #8"))
```