Collinearity Example

In this example we use a small data set to illustrate collinearity diagnostics.

The simplest (and most common) type of collinearity is correlation between two predictors. This can be detected using pairwise correlations and scatterplots. But additional diagnostics are also considered here.

When collinearity is detected, the easiest solution is to simply remove one or more of the violating predictors from the regression model!

```
library(car)
```

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

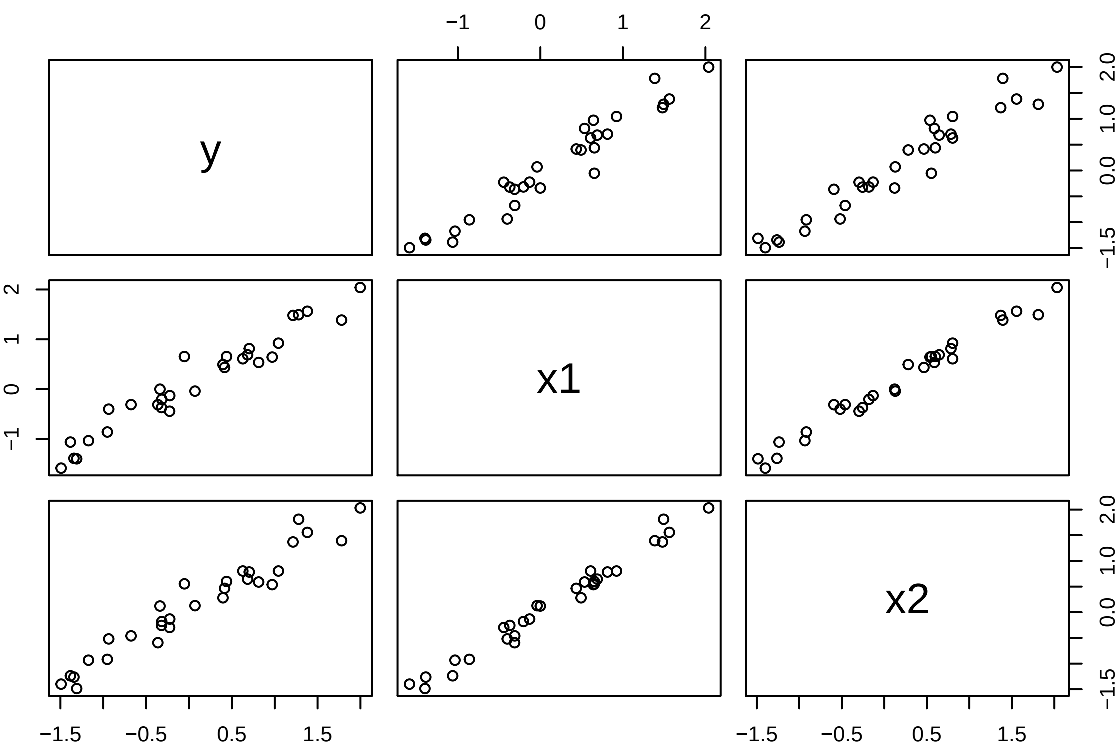
```
library(perturb)
InData <- read.csv("~/Dropbox/STAT512/Lectures/MultReg4/MR4_Collin.csv")
str(InData)
```

```
## 'data.frame':    30 obs. of  3 variables:
##  $ y : num  1.78 0.813 1.043 -1.492 -0.363 ...
##  $ x1: num  1.386 0.536 0.923 -1.585 -0.312 ...
##  $ x2: num  1.394 0.589 0.803 -1.399 -0.593 ...
```

```
cor(InData)
```

```
##            y        x1        x2
## y  1.0000000 0.9697501 0.9688815
## x1 0.9697501 1.0000000 0.9899724
## x2 0.9688815 0.9899724 1.0000000
```

```
pairs(InData)
```



```r
Model1 <- lm(y ~ x1, data = InData)
summary(Model1)
```

```
##
## Call:
## lm(formula = y ~ x1, data = InData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63626 -0.13556  0.00135  0.15229  0.46513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.07262    0.04549  -1.596    0.122
## x1           1.00129    0.04763  21.022   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2456 on 28 degrees of freedom
## Multiple R-squared:  0.9404, Adjusted R-squared:  0.9383
## F-statistic: 441.9 on 1 and 28 DF,  p-value: < 2.2e-16
```

```r
Model2 <- lm(y ~ x2, data = InData)
summary(Model2)
```

```
##
## Call:
## lm(formula = y ~ x2, data = InData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53511 -0.09575 -0.01236  0.13131  0.50497
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.06926    0.04611  -1.502    0.144
## x2           0.99487    0.04803  20.712   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.249 on 28 degrees of freedom
## Multiple R-squared:  0.9387, Adjusted R-squared:  0.9365
## F-statistic:   429 on 1 and 28 DF,  p-value: < 2.2e-16
```

```r
Model3 <- lm(y ~ x1 + x2, data = InData)
summary(Model3)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = InData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59239 -0.11943  0.01742  0.09999  0.45795
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.07189    0.04478  -1.606    0.120
## x1           0.54766    0.33185   1.650    0.110
## x2           0.45569    0.33002   1.381    0.179
##
## Residual standard error: 0.2417 on 27 degrees of freedom
## Multiple R-squared:  0.9443, Adjusted R-squared:  0.9402
## F-statistic: 229.1 on 2 and 27 DF,  p-value: < 2.2e-16
```

```
vif(Model3)
```

```
##       x1       x2
## 50.11382 50.11382
```

```
colldiag(Model3, add.intercept = FALSE)
```

```
## Condition
## Index    Variance Decomposition Proportions
##           x1    x2
## 1   1.000 0.005 0.005
## 2  14.289 0.995 0.995
```