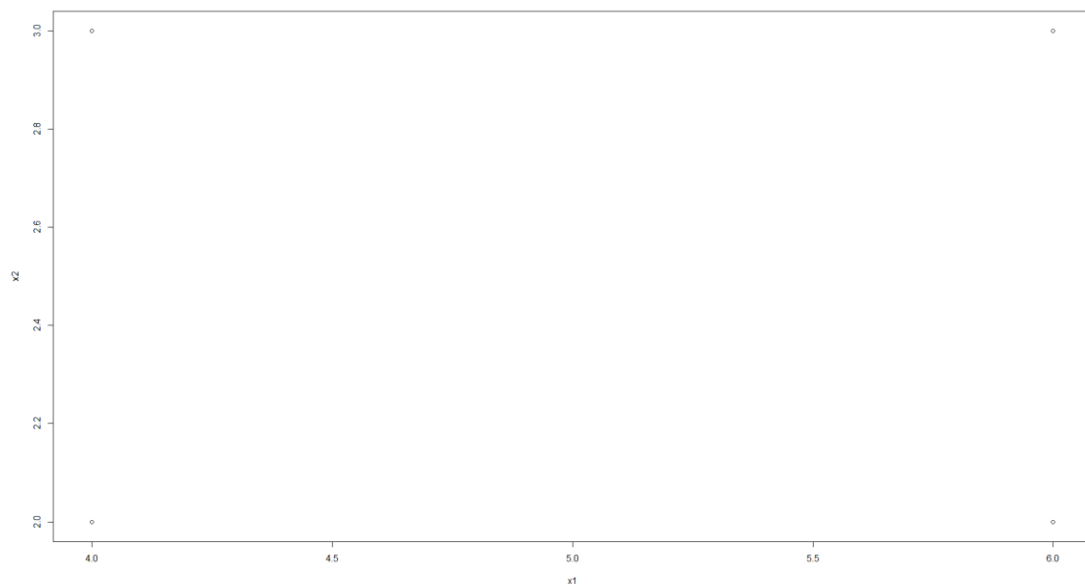


```
#####
##### Multicollinearity #####
#####
##### extreme case when the predictor variables are uncorrelated #####
> ex.data1<-read.table(file="C:/jenn/teaching/stat440540/data/CH7/CH07TA06.txt")
> x1 <- ex.data1$V1
> x2 <- ex.data1$V2
> y <- ex.data1$V3
> ex.data
> ex.data1
  V1 V2 V3
1  4  2 42
2  4  2 39
3  4  3 48
4  4  3 51
5  6  2 49
6  6  2 53
7  6  3 61
8  6  3 60
>
```



```
> cor(cbind(y, x1, x2)) ##notice that x1 and x2 are uncorrelated
      y      x1      x2
y 1.0000000 0.7419309 0.6384057
x1 0.7419309 1.0000000 0.0000000
x2 0.6384057 0.0000000 1.0000000
```

```
> ## Fit the model  $y = b_0 + b_1*x_1 + b_2*x_2$ 
> myfit1 <- lm(y ~ x1 + x2, data=ex.data1)
> summary(myfit1)
```

Call:

```
lm(formula = y ~ x1 + x2, data = ex.data1)
```

```
Residuals:
```

```
 1    2    3    4    5    6    7    8
1.625 -1.375 -1.625 1.375 -2.125 1.875 0.625 -0.375
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.3750    4.7405   0.079 0.940016
x1           5.3750    0.6638   8.097 0.000466 ***
x2           9.2500    1.3276   6.968 0.000937 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.877 on 5 degrees of freedom
Multiple R-squared:  0.958,    Adjusted R-squared:  0.9412
F-statistic: 57.06 on 2 and 5 DF,  p-value: 0.000361
```

```
> anova(myfit1)
```

```
Analysis of Variance Table
```

```
Response: y
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
x1     1  231.125  231.125   65.567 0.0004657 ***
x2     1  171.125  171.125   48.546 0.0009366 ***
Residuals 5  17.625    3.525
```

```
      Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
>
```

```
> vif(myfit1)
```

```
x1 x2
 1  1
```

```
> ## Fit the model y = b0 + b1*x2 + b2*x1
```

```
> myfit1_1 <- lm(y ~ x2 + x1, data=ex.data1)
```

```
> summary(myfit1_1)
```

```
Call:
```

```
lm(formula = y ~ x2 + x1, data = ex.data1)
```

```
Residuals:
```

```
 1    2    3    4    5    6    7    8
1.625 -1.375 -1.625 1.375 -2.125 1.875 0.625 -0.375
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.3750    4.7405   0.079 0.940016
x2           9.2500    1.3276   6.968 0.000937 ***
x1           5.3750    0.6638   8.097 0.000466 ***
```

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.877 on 5 degrees of freedom
Multiple R-squared: 0.958, Adjusted R-squared: 0.9412
F-statistic: 57.06 on 2 and 5 DF, p-value: 0.000361

```
> anova(myfit1_1)
Analysis of Variance Table
```

```
Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
x2     1  171.125 171.125  48.546 0.0009366 ***
x1     1  231.125 231.125  65.567 0.0004657 ***
Residuals 5  17.625  3.525
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
>
>
> ## Fit the model y = b0 + b1*x1
> myfit2 <- lm(y ~ x1, data=ex.data1)
> summary(myfit2)
```

```
Call:
lm(formula = y ~ x1, data = ex.data1)
```

```
Residuals:
    Min     1Q   Median     3Q     Max
-6.750 -3.750  0.125  4.500  6.000
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.500     10.111   2.324  0.0591 .
x1           5.375      1.983   2.711  0.0351 *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

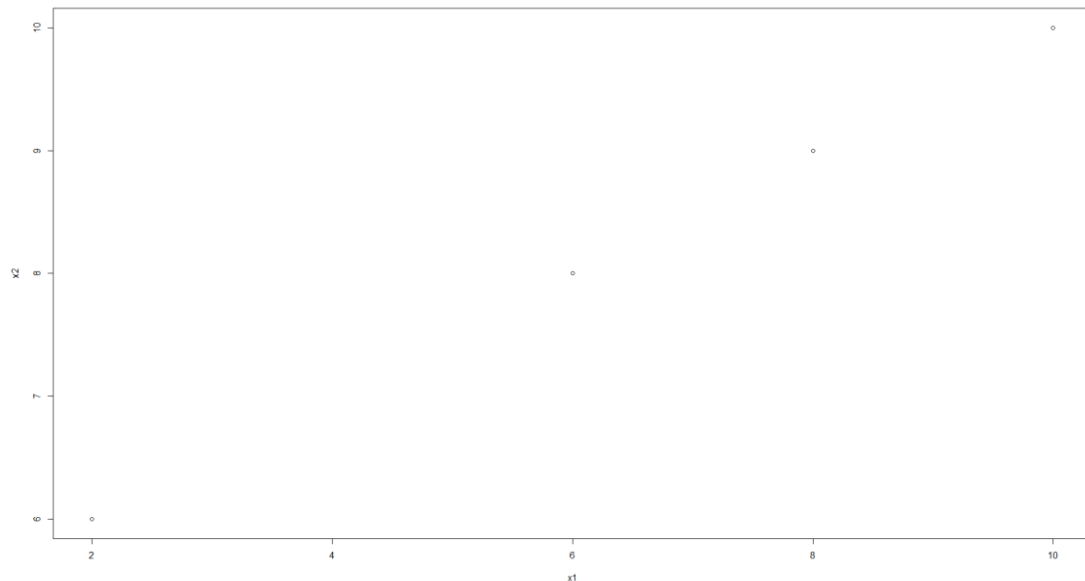
Residual standard error: 5.609 on 6 degrees of freedom
Multiple R-squared: 0.5505, Adjusted R-squared: 0.4755
F-statistic: 7.347 on 1 and 6 DF, p-value: 0.03508

```
> anova(myfit2)
Analysis of Variance Table
```

```
Response: y
      Df    Sum Sq  Mean Sq F value Pr(>F)
x1     1    231.12  231.125  7.347 0.03508 *
Residuals 6  188.75   31.458
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
##### extreme case when the predictor variables are perfectly
#####correlated#####
> ex.data2<-read.table(file="C:/jenn/teaching/stat440540/data/CH7/CH07TA08.txt")
> ex.data2
  V1 V2 V3
1  2  6 23
2  8  9 83
3  6  8 63
4 10 10 103
>
plot(x1,x2) ##notice that x1 and x2 are perfectly correlated by x2= 5+.5 x1
```



```
>
> ## Fit the model y = b0 + b1*x1 + b2*x2
> myfit4 <- lm(y ~ x1 + x2, data=ex.data2)
> summary(myfit4)
```

```
Call:
lm(formula = y ~ x1 + x2, data = ex.data2)
```

```
Residuals:
    1     2     3     4 
1.255e-15 -6.097e-15  5.385e-16  4.304e-15
```

```
Coefficients: (1 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.000e+00  6.476e-15  4.633e+14 <2e-16 ***
x1           1.000e+01  9.068e-16  1.103e+16 <2e-16 ***
x2           NA         NA         NA         NA
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 5.365e-15 on 2 degrees of freedom
Multiple R-squared: 1, Adjusted R-squared: 1
F-statistic: 1.216e+32 on 1 and 2 DF, p-value: < 2.2e-16

Warning message:

In summary.lm(myfit4) : essentially perfect fit: summary may be unreliable
> anova(myfit4)

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	3500	3500 1.2	161e+32	< 2.2e-16 ***
Residuals	2	0	0		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Warning message:

In anova.lm(myfit4) :

ANOVA F-tests on an essentially perfect fit are unreliable

>

>