Multiple Regression II

Zhenisbek Assylbekov

Department of Mathematics

Regression Analysis

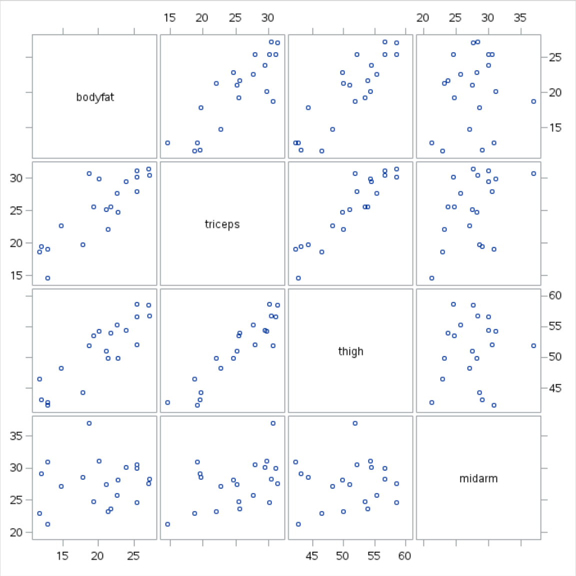
Chapter 7 example: Body fat

$n = 20$ healthy females 25–34 years old.

- x_1 = triceps skinfold thickness (mm)
- x_2 = thigh circumference (cm)
- x_3 = midarm circumference (cm)
- Y = body fat (%)

Obtaining Y_i, the percent of the body that is purly fat, requires immersing a person in water. Want to develop model based on simple body measurements that avoids people getting wet.

Scatterplot



Correlation coefficients

Pearson Correlation Coefficients, N = 20 Prob > |r| under H0: Rho=0			
	triceps	**thigh**	**midarm**
triceps	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.45778 0.0424	0.08467 0.7227	1.00000

Correlation coefficients

Pearson Correlation Coefficients, N = 20 Prob > |r| under H0: Rho=0			
	triceps	**thigh**	**midarm**
triceps	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.45778 0.0424	0.08467 0.7227	1.00000

There is high correlation among the predictors.

Correlation coefficients

Pearson Correlation Coefficients, N = 20 Prob > |r| under H0: Rho=0			
	triceps	**thigh**	**midarm**
triceps	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.45778 0.0424	0.08467 0.7227	1.00000

There is high correlation among the predictors. For example $r = 0.92$ for triceps and thigh. These two variables are *essentially carrying the same information*.

Correlation coefficients

Pearson Correlation Coefficients, N = 20 Prob > |r| under H0: Rho=0			
	triceps	**thigh**	**midarm**
triceps	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.45778 0.0424	0.08467 0.7227	1.00000

There is high correlation among the predictors. For example $r = 0.92$ for triceps and thigh. These two variables are *essentially carrying the same information*. Maybe only one or the other is really needed.

Correlation coefficients

Pearson Correlation Coefficients, N = 20 Prob > |r| under H0: Rho=0			
	triceps	**thigh**	**midarm**
triceps	1.00000	0.92384 <.0001	0.45778 0.0424
thigh	0.92384 <.0001	1.00000	0.08467 0.7227
midarm	0.45778 0.0424	0.08467 0.7227	1.00000

There is high correlation among the predictors. For example $r = 0.92$ for triceps and thigh. These two variables are *essentially carrying the same information*. Maybe only one or the other is really needed.

In general, one predictor may be essentially perfectly predicted by the remaining predictors (a high "partial correlation"), and so would be unnecessary if the other predictors are in the model.

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

Fact: As predictors are added, the SSE can only decrease.

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

Fact: As predictors are added, the SSE can only decrease. The extra sums of squares is how much the SSE decreases:

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

Fact: As predictors are added, the SSE can only decrease. The extra sums of squares is how much the SSE decreases:

Definition. Let $x_1, x_2, \ldots, x_k$ be predictors in a model.

$$\begin{aligned} &\mathrm{SSR}(x_{j+1}, \ldots, x_k | x_1, x_2, \ldots, x_j) \\ &\quad = \mathrm{SSE}(x_1, x_2, \ldots, x_j) - \mathrm{SSE}(x_1, x_2, \ldots, x_j, x_{j+1}, \ldots, x_k), \end{aligned} \quad (1)$$

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

Fact: As predictors are added, the SSE can only decrease. The extra sums of squares is how much the SSE decreases:

Definition. Let $x_1, x_2, \ldots, x_k$ be predictors in a model.

$$\begin{aligned} &\mathrm{SSR}(x_{j+1}, \ldots, x_k | x_1, x_2, \ldots, x_j) \\ &\quad = \mathrm{SSE}(x_1, x_2, \ldots, x_j) - \mathrm{SSE}(x_1, x_2, \ldots, x_j, x_{j+1}, \ldots, x_k), \quad (1) \end{aligned}$$

the difference in the sums of squared errors from the reduced to the full model.

7.1 Extra sums of squares

"Extra" sums of squares are defined as the difference in SSE between a model with some predictors and a larger model that adds *additional* predictors.

Fact: As predictors are added, the SSE can only decrease. The extra sums of squares is how much the SSE decreases:

Definition. Let $x_1, x_2, \ldots, x_k$ be predictors in a model.

$$\begin{aligned} &\mathrm{SSR}(x_{j+1}, \ldots, x_k | x_1, x_2, \ldots, x_j) \\ &\quad = \mathrm{SSE}(x_1, x_2, \ldots, x_j) - \mathrm{SSE}(x_1, x_2, \ldots, x_j, x_{j+1}, \ldots, x_k), \quad (1) \end{aligned}$$

the difference in the sums of squared errors from the reduced to the full model.

This is how much of the total variation in SST is further explained by adding the new predictors.

Example with $k = 8$ predictors

The predictors under consideration are

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8.$$

Example with $k = 8$ predictors

The predictors under consideration are

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8.$$

There are two models

$$\begin{aligned} \text{Reduced} &: x_1, x_3, x_5, x_6, x_8 \\ \text{Full} &: x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \end{aligned}$$

Example with $k = 8$ predictors

The predictors under consideration are

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8.$$

There are two models

$$\begin{aligned} \text{Reduced} &: x_1, x_3, x_5, x_6, x_8 \\ \text{Full} &: x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \end{aligned}$$

$$\begin{aligned} \text{Extra SS} &= \mathrm{SSR}(x_2, x_4, x_7 | x_1, x_3, x_5, x_6, x_8) \\ &= \mathrm{SSE}(\text{reduced}) - \mathrm{SSE}(\text{full}) \\ &= \mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \\ &= \mathrm{SSR}(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) - \mathrm{SSR}(x_1, x_3, x_5, x_6, x_8) \end{aligned}$$

Example with $k = 8$ predictors

The predictors under consideration are

$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8.$$

There are two models

$$\begin{aligned} \text{Reduced} &: x_1, x_3, x_5, x_6, x_8 \\ \text{Full} &: x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \end{aligned}$$

$$\begin{aligned} \text{Extra SS} &= \mathrm{SSR}(x_2, x_4, x_7 | x_1, x_3, x_5, x_6, x_8) \\ &= \mathrm{SSE}(\text{reduced}) - \mathrm{SSE}(\text{full}) \\ &= \mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) \\ &= \mathrm{SSR}(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) - \mathrm{SSR}(x_1, x_3, x_5, x_6, x_8) \end{aligned}$$

This is how much *additional* total variability (SST) is explained by adding x_2, x_4, x_7 to a model that already has x_1, x_3, x_5, x_6, x_8.

7.2 Associated tests

We can formally test whether a certain set of predictors is useless, *in the presence* of other predictors in the model.

7.2 Associated tests

We can formally test whether a certain set of predictors is useless, *in the presence* of other predictors in the model. This is the *general linear test*.

7.2 Associated tests

We can formally test whether a certain set of predictors is useless, *in the presence* of other predictors in the model. This is the *general linear test*.

In the example above, we can test whether x_2, x_4, x_7 are needed if x_1, x_3, x_5, x_6, x_8 are in the model.

7.2 Associated tests

We can formally test whether a certain set of predictors is useless, *in the presence* of other predictors in the model. This is the *general linear test*.

In the example above, we can test whether x_2, x_4, x_7 are needed if x_1, x_3, x_5, x_6, x_8 are in the model.

If full (with $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$) model has much lower SSE than reduced model (without x_2, x_4, x_7) then at least one of x_2, x_4, x_7 is needed.

F-test

Say we want to test whether we can drop q variables from a model that has $p = k + 1$ (including the intercept), $q < p$.

F-test

Say we want to test whether we can drop q variables from a model that has $p = k + 1$ (including the intercept), $q < p$.

To test H$_0$: $\beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_q} = 0$ in the full model

$$
\begin{aligned}
F^* &= \frac{[\mathrm{SSE}(\text{reduced}) - \mathrm{SSE}(\text{full})]/q}{\mathrm{SSE}(\text{full})/(n-p)} \\
&\sim F(q, n-p)
\end{aligned}
$$

F-test

Say we want to test whether we can drop q variables from a model that has $p = k + 1$ (including the intercept), $q < p$.

To test H_0: $\beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_q} = 0$ in the full model

$$\begin{aligned} F^* &= \frac{[\mathrm{SSE}(\text{reduced}) - \mathrm{SSE}(\text{full})]/q}{\mathrm{SSE}(\text{full})/(n-p)} \\ &\sim F(q, n-p) \end{aligned}$$

If H_0: $\beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_q} = 0$ is true; a p-value for the test is $\Pr(F^* > F_{q,n-p})$.

F-test

Say we want to test whether we can drop q variables from a model that has $p = k + 1$ (including the intercept), $q < p$.

To test H_0: $\beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_q} = 0$ in the full model

$$F^* = \frac{[\mathrm{SSE}(\text{reduced}) - \mathrm{SSE}(\text{full})]/q}{\mathrm{SSE}(\text{full})/(n-p)}$$
$$\sim F(q, n-p)$$

If H_0: $\beta_{j_1} = \beta_{j_2} = \ldots = \beta_{j_q} = 0$ is true; a p-value for the test is $\Pr(F^* > F_{q,n-p})$.

Can be done in R.

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

$$F^* = \frac{[\mathrm{SSE(reduced)} - \mathrm{SSE(full)}]/(\#\ \text{params in test})}{\mathrm{MSE(full)}}$$

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

$$\begin{aligned} F^* &= \frac{[\mathrm{SSE(reduced)} - \mathrm{SSE(full)}]/(\#\ \text{params in test})}{\mathrm{MSE(full)}} \\ &= \frac{[\mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_{1:8})]/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \end{aligned}$$

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

$$
\begin{aligned}
F^* &= \frac{[\mathrm{SSE(reduced)} - \mathrm{SSE(full)}]/(\#\ \text{params in test})}{\mathrm{MSE(full)}} \\
&= \frac{[\mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_{1:8})]/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \\
&= \frac{\mathrm{SSR}(x_2, x_4, x_7 | x_1, x_3, x_5, x_6, x_8)/3}{\mathrm{SSE}(x_{1:8})/(n-9)}
\end{aligned}
$$

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

$$
\begin{aligned}
F^* &= \frac{[\mathrm{SSE(reduced)} - \mathrm{SSE(full)}]/(\#\text{ params in test})}{\mathrm{MSE(full)}} \\
&= \frac{[\mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_{1:8})]/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \\
&= \frac{\mathrm{SSR}(x_2, x_4, x_7 | x_1, x_3, x_5, x_6, x_8)/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \\
&\sim F_{3,n-9}
\end{aligned}
$$

F-test example with $k = 8$ predictors

To test H_0: $\beta_2 = \beta_4 = \beta_7 = 0$,

$$
\begin{aligned}
F^* &= \frac{[\mathrm{SSE(reduced)} - \mathrm{SSE(full)}]/(\#\text{ params in test})}{\mathrm{MSE(full)}} \\
&= \frac{[\mathrm{SSE}(x_1, x_3, x_5, x_6, x_8) - \mathrm{SSE}(x_{1:8})]/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \\
&= \frac{\mathrm{SSR}(x_2, x_4, x_7 | x_1, x_3, x_5, x_6, x_8)/3}{\mathrm{SSE}(x_{1:8})/(n-9)} \\
&\sim F_{3,n-9}
\end{aligned}
$$

if H_0: $\beta_2 = \beta_4 = \beta_7 = 0$ is true. ($x_{1:8} = x_1, x_2, \ldots, x_8$)

Bodyfat example

https://github.com/zh3nis/MATH440/blob/main/chp07/bodyfat_extra_ss.R

```
> anova(m_full, m_reduced)
Analysis of Variance Table

Model 1: bodyfat ~ triceps + thigh + midarm
Model 2: bodyfat ~ triceps + midarm
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     16  98.405
2     17 105.934 -1   -7.5293 1.2242 0.2849
```

Bodyfat example

https://github.com/zh3nis/MATH440/blob/main/chp07/bodyfat_extra_ss.R

```
> anova(m_full, m_reduced)
Analysis of Variance Table

Model 1: bodyfat ~ triceps + thigh + midarm
Model 2: bodyfat ~ triceps + midarm
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     16  98.405
2     17 105.934 -1   -7.5293 1.2242 0.2849
```

FTR $H_0 : \beta_2 = 0$ in

$$\text{bodyfat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \beta_3 \times \text{midarm}_i + \epsilon_i$$

Type I (sequential) sums of squares

Say you have $k = 4$ predictors. Then the SSR for the full model can be written

$$\begin{aligned} \mathrm{SSR} &= \mathrm{SSR}(x_1, x_2, x_3, x_4) \\ &= \mathrm{SSR}(x_1) + \mathrm{SSR}(x_2|x_1) + \mathrm{SSR}(x_3|x_1, x_2) + \mathrm{SSR}(x_4|x_1, x_2, x_3). \end{aligned}$$

These are called **sequential sums of squares**, or Type I sums of squares. They explain how much variability is absorbed by adding predictors sequentially to a model. There are four corresponding hypothesis tests with these sequential sums of squares:

Model	Hypothesis	F-statistic
$Y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$	$\mathrm{H}_0 : \beta_1 = 0$	$\frac{\mathrm{MSR}(x_1)}{\mathrm{MSE}(x_1)}$
$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$	$\mathrm{H}_0 : \beta_2 = 0$	$\frac{\mathrm{MSR}(x_2\|x_1)}{\mathrm{MSE}(x_1,x_2)}$
$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$	$\mathrm{H}_0 : \beta_3 = 0$	$\frac{\mathrm{MSR}(x_3\|x_1,x_2)}{\mathrm{MSE}(x_1,x_2,x_3)}$
$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$	$\mathrm{H}_0 : \beta_4 = 0$	$\frac{\mathrm{MSR}(x_4\|x_1,x_2,x_3)}{\mathrm{MSE}(x_1,x_2,x_3,x_4)}$

Sequential SS in R — Bodyfat example

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

Response: bodyfat
          Df Sum Sq Mean Sq F value    Pr(>F)
triceps    1 352.27  352.27 57.2768 1.131e-06 ***
thigh      1  33.17   33.17  5.3931   0.03373 *
midarm     1  11.55   11.55  1.8773   0.18956
```

Sequential SS in R — Bodyfat example

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

Response: bodyfat
          Df Sum Sq Mean Sq F value    Pr(>F)    
triceps    1 352.27  352.27 57.2768 1.131e-06 ***
thigh      1  33.17   33.17  5.3931   0.03373 *  
midarm     1  11.55   11.55  1.8773   0.18956    
```

- Reject $H_0 : \beta_1 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \epsilon_i$

Sequential SS in R — Bodyfat example

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

Response: bodyfat
          Df Sum Sq Mean Sq F value    Pr(>F)    
triceps    1 352.27  352.27 57.2768 1.131e-06 ***
thigh      1  33.17   33.17  5.3931   0.03373 *  
midarm     1  11.55   11.55  1.8773   0.18956    
```

- Reject $H_0 : \beta_1 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \epsilon_i$
- Reject $H_0 : \beta_2 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \epsilon_i$

Sequential SS in R — Bodyfat example

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

Response: bodyfat
          Df Sum Sq Mean Sq F value    Pr(>F)
triceps    1 352.27  352.27 57.2768 1.131e-06 ***
thigh      1  33.17   33.17  5.3931   0.03373 *
midarm     1  11.55   11.55  1.8773   0.18956
```

- Reject $H_0 : \beta_1 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \epsilon_i$
- Reject $H_0 : \beta_2 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \epsilon_i$
- FTR $H_0 : \beta_3 = 0$ in
  $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \beta_3 \times \text{midarm}_i + \epsilon_i$

Sequential SS in R — Bodyfat example

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

Response: bodyfat
          Df Sum Sq Mean Sq F value    Pr(>F)
triceps    1 352.27  352.27 57.2768 1.131e-06 ***
thigh      1  33.17   33.17  5.3931   0.03373 *
midarm     1  11.55   11.55  1.8773   0.18956
```

- Reject $H_0 : \beta_1 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \epsilon_i$
- Reject $H_0 : \beta_2 = 0$ in $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \epsilon_i$
- FTR $H_0 : \beta_3 = 0$ in
  $\text{fat}_i = \beta_0 + \beta_1 \times \text{triceps}_i + \beta_2 \times \text{thigh}_i + \beta_3 \times \text{midarm}_i + \epsilon_i$
- Order entered (triceps, thigh, midarm) matters!

ANOVA table & decomposing the SSR(F)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.9846118	132.3282039	21.52	<.0001
Error	16	98.4048882	6.1503055		
Corrected Total	19	495.3895000			

The sequential extra sums of squares are given on the previous slide:

$\mathrm{SSR}(x_1) = 352.3$; $\mathrm{SSR}(x_2|x_1) = 33.2$, and $\mathrm{SSR}(x_3|x_1, x_2) = 11.5$.

ANOVA table & decomposing the SSR(F)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.9846118	132.3282039	21.52	<.0001
Error	16	98.4048882	6.1503055		
Corrected Total	19	495.3895000			

The sequential extra sums of squares are given on the previous slide:

$SSR(x_1) = 352.3$; $SSR(x_2|x_1) = 33.2$, and $SSR(x_3|x_1, x_2) = 11.5$.

Almost all of the $SSR(x_1, x_2, x_3) = 397.0$ is explained by x_1 (triceps) alone.

ANOVA table & decomposing the SSR(F)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	396.9846118	132.3282039	21.52	<.0001
Error	16	98.4048882	6.1503055		
Corrected Total	19	495.3895000			

The sequential extra sums of squares are given on the previous slide:

$\mathrm{SSR}(x_1) = 352.3$; $\mathrm{SSR}(x_2|x_1) = 33.2$, and $\mathrm{SSR}(x_3|x_1, x_2) = 11.5$.

Almost all of the $\mathrm{SSR}(x_1, x_2, x_3) = 397.0$ is explained by x_1 (triceps) alone.

Also note, as required,

$$\begin{aligned} \mathrm{SSR}(x_1, x_2, x_3) &= 397.0 = 352.3 + 33.2 + 11.5 \\ &= \mathrm{SSR}(x_1) + \mathrm{SSR}(x_2|x_1) + \mathrm{SSR}(x_3|x_1, x_2). \end{aligned}$$

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model.

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model. Examples:

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model. Examples:

- $R^2_{Y2|1} = \mathrm{SSR}(x_2 \mid x_1)/\mathrm{SSE}(x_1)$

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model. Examples:

- $R^2_{Y2|1} = \mathrm{SSR}(x_2 \mid x_1)/\mathrm{SSE}(x_1)$
- $R^2_{Y3|12} = \mathrm{SSR}(x_3 \mid x_1, x_2)/\mathrm{SSE}(x_1, x_2)$

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model. Examples:

- $R^2_{Y2|1} = \mathrm{SSR}(x_2 \mid x_1)/\mathrm{SSE}(x_1)$
- $R^2_{Y3|12} = \mathrm{SSR}(x_3 \mid x_1, x_2)/\mathrm{SSE}(x_1, x_2)$
- $R^2_{Y23|1} = \mathrm{SSR}(x_2, x_3 \mid x_1)/\mathrm{SSE}(x_1)$

7.4 Coefficients of partial determination

We can *standardize* extra sums of squares to be between 0 and 1.

The **coefficient of partial determination** is the fraction by which the SSE is reduced when adding new predictor(s) to an existing model. Examples:

- $R^2_{Y2|1} = \mathrm{SSR}(x_2 \mid x_1)/\mathrm{SSE}(x_1)$
- $R^2_{Y3|12} = \mathrm{SSR}(x_3 \mid x_1, x_2)/\mathrm{SSE}(x_1, x_2)$
- $R^2_{Y23|1} = \mathrm{SSR}(x_2, x_3 \mid x_1)/\mathrm{SSE}(x_1)$

For example, if $R^2_{Y3|12} = 0.5$ then *50% of the remaining variability* is explained by adding x_3 to a model that already had x_1 and x_2.

Partial R^2 in R — Bodyfat example

```
> library(rsq)
> m_midarm_only = update(m_full, . ~ . - triceps - thigh)
> m_midarm_triceps = update(m_full, . ~ . - thigh)
> rsq.partial(m_midarm_triceps, m_midarm_only)
$adjustment
[1] FALSE

$variables.full
[1] "triceps" "midarm"

$variables.reduced
[1] "midarm"

$partial.rsq
[1] 0.7817311
```

7.6 Multicollinearity

```
> summary(m_full)

Call:
lm(formula = bodyfat ~ triceps + thigh + midarm, data = bodyfat_data)
...

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190

Multiple R-squared:  0.8014,Adjusted R-squared:  0.7641
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

7.6 Multicollinearity

```
> summary(m_full)

Call:
lm(formula = bodyfat ~ triceps + thigh + midarm, data = bodyfat_data)
...

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190

Multiple R-squared:  0.8014,Adjusted R-squared:  0.7641
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06
```

We reject $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, BUT we fail to reject $H_0 : \beta_j = 0$ individually!

7.6 Multicollinearity

The set x_1, x_2, x_3 are useful for explaining body fat, but none of the three are useful in the presence of the other two.

Why?

7.6 Multicollinearity

The set x_1, x_2, x_3 are useful for explaining body fat, but none of the three are useful in the presence of the other two.

Why? The predictors are measuring similar phenomena; their sample values are highly correlated.

7.6 Multicollinearity

The set x_1, x_2, x_3 are useful for explaining body fat, but none of the three are useful in the presence of the other two.

Why? The predictors are measuring similar phenomena; their sample values are highly correlated. For example, $r = 0.924$ between triceps thickness x_1 and thigh circumference x_2.

7.6 Multicollinearity

The set x_1, x_2, x_3 are useful for explaining body fat, but none of the three are useful in the presence of the other two.

Why? The predictors are measuring similar phenomena; their sample values are highly correlated. For example, $r = 0.924$ between triceps thickness x_1 and thigh circumference x_2.

This is known as multicollinearity among the predictors.

Effects of multicollinearity

Effects of multicollinearity

- Model may still provide a good fit and precise prediction/estimation of the response.

Effects of multicollinearity

- Model may still provide a good fit and precise prediction/estimation of the response.
- Several estimated regression coefficients b_1, b_2, ..., b_k will have large standard errors (pp. 281–283), leading to conclusions that individual predictors are *not significant* although overall F-test may be *highly* significant.

Effects of multicollinearity

- Model may still provide a good fit and precise prediction/estimation of the response.
- Several estimated regression coefficients b_1, b_2, . . . , b_k will have large standard errors (pp. 281–283), leading to conclusions that individual predictors are *not significant* although overall F-test may be *highly* significant.
- Concept of "holding all other predictors constant" doesnt make sense in practice.

Effects of multicollinearity

- Model may still provide a good fit and precise prediction/estimation of the response.
- Several estimated regression coefficients $b_1, b_2, \ldots, b_k$ will have large standard errors (pp. 281–283), leading to conclusions that individual predictors are *not significant* although overall F-test may be *highly* significant.
- Concept of "holding all other predictors constant" doesnt make sense in practice.
- Signs of regression coefficients may be "opposite" of intuition (or what we might think *marginally* they might be based on a scatterplot).

Bodyfat example

```
lm(formula = bodyfat ~ triceps + thigh + midarm,
   data = bodyfat_data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190
```

Bodyfat example

```
lm(formula = bodyfat ~ triceps + thigh + midarm,
   data = bodyfat_data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190
```

Two of the three regression effects are *negative*.

Bodyfat example

```
lm(formula = bodyfat ~ triceps + thigh + midarm,
   data = bodyfat_data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190
```

Two of the three regression effects are *negative*. Holding midarm and triceps constant, increasing the thigh circumference 1 mm *decreases* bodyfat.

Bodyfat example

```
lm(formula = bodyfat ~ triceps + thigh + midarm,
   data = bodyfat_data)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
triceps        4.334      3.016   1.437    0.170
thigh         -2.857      2.582  -1.106    0.285
midarm        -2.186      1.595  -1.370    0.190
```

Two of the three regression effects are *negative*. Holding midarm and triceps constant, increasing the thigh circumference 1 mm *decreases* bodyfat. Does this make sense?

Detecting multicollinearity

Predictor x_j has a *variance inflation factor* of

$$VIF_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is the R^2 from regressing x_j on the remaining predictors $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$.

Detecting multicollinearity

Predictor x_j has a *variance inflation factor* of

$$VIF_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is the R^2 from regressing x_j on the remaining predictors $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$.

High R_j^2 (near 1) $\Rightarrow$ x_j is linearly associated with other predictors $\Rightarrow$ high VIF_j.

Detecting multicollinearity

Predictor x_j has a *variance inflation factor* of

$$VIF_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is the R^2 from regressing x_j on the remaining predictors $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$.

High R_j^2 (near 1) $\Rightarrow$ x_j is linearly associated with other predictors $\Rightarrow$ high VIF_j.

- $VIF_j \approx 1 \Rightarrow x_j$ is not involved in any multicollinearity.

Detecting multicollinearity

Predictor x_j has a *variance inflation factor* of

$$VIF_j = \frac{1}{1 - R_j^2},$$

where R_j^2 is the R^2 from regressing x_j on the remaining predictors $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$.

High R_j^2 (near 1) $\Rightarrow x_j$ is linearly associated with other predictors $\Rightarrow$ high VIF_j.

- $VIF_j \approx 1 \Rightarrow x_j$ is not involved in any multicollinearity.
- $VIF_j > 10 \Rightarrow x_j$ is involved in severe multicollinearity.

VIF$_j$'s in R

```
> library(car)
> vif(m_full)
 triceps    thigh   midarm
708.8429 564.3434 104.6060
```

What do you conclude?

Remedies of multicollinearity

- Drop one or more predictors from the model. We'll discuss this in Chapter 9.

Remedies of multicollinearity

- Drop one or more predictors from the model. We'll discuss this in Chapter 9.
- More advanced: *principle components regression* uses new predictors that are linear combinations of the original predictors as predictors in a new model. The new predictors are selected to be uncorrelated. Disadvantage: the new predictors might be hard to interpret.

Remedies of multicollinearity

- Drop one or more predictors from the model. We'll discuss this in Chapter 9.
- More advanced: *principle components regression* uses new predictors that are linear combinations of the original predictors as predictors in a new model. The new predictors are selected to be uncorrelated. Disadvantage: the new predictors might be hard to interpret.
- More advanced: *ridge regression* (Section 11.2).