Chapter 7: Multiple Regression II

This chapter covers some specialized topics unique to multiple regression.

- Extra sums of squares which are useful for conducting a variety tests about the regression coefficients
- The Standardized version of the multiple regression model
- Multicollinearity

7.1 Extra sums of squares

- Measurement of the marginal reduction in the error sums of squares when an independent variable or several variables are added to the model given a set of independent variables are already in the model.

- Measurement of the marginal increase in the regression sums of squares when an independent variable or several variables are added to the model given a set of independent variables are already in the model.

Review SSTO, SSE, and SSR

- SSTO = total sum of squares
- SSE = sum of squared errors (remaining variability of Y not explained by the $X_1,\ldots,X_{p-1}$)
- SSR = regression sum squares (variability of Y accounted for by the $X_1,\ldots,X_{p-1}$)

SSE

- $SSE(X_1)$ = Sum of squared errors using only X_1 in the model ($Y_i=\beta_0+\beta_1X_{1i}+\varepsilon_i$)
- $SSE(X_1,X_2)$ = Sum of squared errors using only X_1 and X_2 in the model ($Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\varepsilon_i$)
- $SSE(X_1, X_2, X_3)$ = Sum of squared errors using only X_1 and X_2 and X_3 in the model ($Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\varepsilon_i$)

⋮

Remember that as more variables are added to the model the corresponding SSE stays the same or is reduced. For example, $SSE(X_1) \geq SSE(X_1,X_2)$

SSR can be partitioned

$SSR(X_1)$ = regression sum squares with only X_1 in the model

$SSR(X_2|X_1)$ = reduction in error sum squared errors when X_2 is added to the model <u>given</u> that X_1 is already in the model

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1,X_2)$$

Or, equivalently:

$SSR(X_2|X_1) = SSR(X_1,X_2) - SSR(X_1)$

This is the "extra sum of squares" explained by using the regression model with addition of X_2

$SSR(X_3|X_1,X_2)$ = reduction in regression sum squares due when add X_3 is added to the model given that X_1 and X_2 are already in the model

$SSR(X_3|X_1,X_2) = SSE(X_1,X_2) - SSE(X_1,X_2,X_3)$

Or, equivalently:

$SSR(X_3|X_1,X_2)= SSR(X_1,X_2, X_3) - SSR(X_1,X_2)$

This is the "extra sum of squares" explained by using the regression model with addition of X_3

Suppose there are only three variables under consideration – X_1, X_2, and X_3. Then the "usual" SSR is:

$SSR(X_1,X_2,X_3)= SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1,X_2)$

ANOVA table containing decomposition of SSR:

Source of variation	**df**	**SS**	**MS**
Regression	3	SSR	MSR
X_1	1	$SSR(X_1)$	$MSR(X_1)$
$X_2|X_1$	1	$SSR(X_2|X_1)$	$MSR(X_2|X_1)$
$X_3|X_1,X_2$	1	$SSR(X_3|X_1,X_2)$	$MSR(X_3|X_1,X_2)$
Error	n-4	SSE	MSE
Total	n-1	SSTO	

where MSR = SSR/df

There are many other extra sums of squares that could be examined. For example, $SSR(X_3,X_2|X_1) = SSE(X_1) - SSE(X_1,X_2,X_3)$.

Exercise: When the regression model contains three X variables, a variety of decomposition of $SSR(X_1,X_2,X_3)$ can be obtained. Try to write out some using the extra sum of squares.

(1) $SSR(X_1,X_2,X_3)= SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1,X_2)$
(2)

An Example: Body Fat Example (Table 7.1) Table 7.1 contains a portion of the data for a study of the relationship of amount of body fat (Y) to several possible predictor variables, based on a sample of 20 health females 25-34 years old. The possible predictor variables are triceps skinfold thickness (X1), thigh circumference (X2), and midarm circumference (X3).

Try the following four models where body fat (Y) is regressed

1) On triceps skinfold thickness (X1) alone;
2) On thigh circumference (X2) alone;
3) On X1 and X2 only;
4) On all three predictor variables.

```
> Data <- read.table('C:/Ch07TA01.txt', header=FALSE)
> colnames(Data) <- c("X1", "X2", "X3", "Y")
>  # X1 = triceps skinfold thickness;
>  # X2 = thigh circumference;
>  # X3 = midarm circumference;
>  # Y = body fat
>
> print('Model 1: on X1 alone')
[1] "Model 1: on X1 alone"
> anova(lm(Y~X1, data=Data))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27  44.305 3.024e-06 ***
Residuals 18 143.12    7.95
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> print( 'Model 2: on X2 alone')
[1] "Model 2: on X2 alone"
> anova(lm(Y~X2, data=Data))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value  Pr(>F)
X2         1 381.97  381.97  60.617 3.6e-07 ***
Residuals 18 113.42    6.30
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> print( 'Model 3: on X1 and X2 only')
[1] "Model 3: on X1 and X2 only"
> anova(lm(Y~X1+X2, data=Data))
Analysis of Variance Table
```

```
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27 54.4661 1.075e-06 ***
X2         1  33.17   33.17  5.1284    0.0369 *
Residuals 17 109.95    6.47
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> print( 'Model 4: on X1, X2 and X3')
[1] "Model 4: on X1, X2 and X3"
> anova(lm(Y~X1+X2+X3, data=Data))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27 57.2768 1.131e-06 ***
X2         1  33.17   33.17  5.3931   0.03373 *
X3         1  11.55   11.55  1.8773   0.18956
Residuals 16  98.40    6.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> anova(lm(Y~X2+X3+X1, data=Data))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X2         1 381.97  381.97 62.1052 6.735e-07 ***
X3         1   2.31    2.31  0.3762    0.5483
X1         1  12.70   12.70  2.0657    0.1699
Residuals 16  98.40    6.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(lm(Y~X1+X3+X2, data=Data))
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27 57.2768 1.131e-06 ***
X3         1  37.19   37.19  6.0461   0.02571 *
X2         1   7.53    7.53  1.2242   0.28489
Residuals 16  98.40    6.15
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Using the R outputs and find or compute the following:

1) SSE(X1) = 143.12
2) SSR(X1) = 352.27
3) SSE(X2) = 113.42
4) SSR(X2) = 381.97
5) SSE(X1, X2) = 109.95
6) SSR(X1, X2) = 385.44
7) SSR(X2 | X1) = 33.17
8) SSR(X1 |X2) = 3.47
9) SSE(X1, X2, X3) = 98.41
10) SSR(X1, X2, X3) = 396.98
11) SSR(X3 | X1, X2) = 11.54
12) SSR(X2, X3 | X1) = 44.71

Notes:

From the Type I SS:

SSR	Value
$SSR(X_1)$	352.2698
$SSR(X_2\|X_1)$	33.1689
$SSR(X_3\|X_1,X_2)$	11.5459

From the Type II SS:

SSR	Value
$SSR(X_1\|X_2,X_3)$	12.70489
$SSR(X_2\|X_1,X_3)$	7.52928
$SSR(X_3\|X_1,X_2)$	11.5459

Check the following using the R outputs:

(1) $SSR(X_1,X_2,X_3) = SSR(X_1)+ SSR(X_2|X_1)+ SSR(X_3|X_1,X_2)$
(2) Why are $SSR(X_1)$ and $SSR(X_1|X_2,X_3)$ so different? Look at the correlation matrix.

```
> # Correlation structure
> cor(Data)
          X1        X2        X3         Y
X1 1.0000000 0.9238425 0.4577772 0.8432654
X2 0.9238425 1.0000000 0.0846675 0.8780896
X3 0.4577772 0.0846675 1.0000000 0.1424440
Y  0.8432654 0.8780896 0.1424440 1.0000000
>
```

7.2 Uses of extra sums of squares in tests for regression coefficients

Model A is nested within model B

Model B has all the terms of model A and at least an additional term

Reduced Model: Model A
Complete Model: Model B

Example:

Reduced Model: $E(Y_i)=\beta_0 + \beta_1 X_{i1}$
Complete Model: $E(Y)=\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$

Example:

Reduced Model: $E(Y)=\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$
Complete Model: $E(Y)=\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$
Complete Model: $E(Y)=\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}$

Hypothesis test steps for a partial (nested) F-test

(That is, test whether some β_k=0.)

1) H_o: Reduced Model: $E(Y)=\beta_0 + \beta_1 X_1 + \cdots + \beta_g X_g$

   H_a: Complete Model: $E(Y) = \beta_0 + \beta_1 X_1 + \cdots + \beta_g X_g + \beta_{g+1} X_{g+1} + \cdots + \beta_{p-1} X_{p-1}$

   Restated another way:

   H_o: $\beta_{g+1} = \cdots = \beta_{p-1} = 0$ (Note: there are p-1-g β's =0)

   H_a: At least one of the β's in H_o are not 0

2) Test statistic

   The general linear test statistic (2.70):

$$F^* = \frac{\frac{SSE(R) - SSE(F)}{df_R - df_F}}{\frac{SSE(F)}{df_F}}$$

$$F^* = \frac{\left(SSE(X_1,\ldots,X_g) - SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})\right)/(p-1-g)}{SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})/(n-p)}$$

$$= \frac{SSR(X_{g+1},\ldots,X_{p-1} \mid X_1,\ldots,X_g)/(p-1-g)}{SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})/(n-p)}$$

$$= \frac{MSR(X_{g+1},\ldots,X_{p-1} \mid X_1,\ldots,X_g)}{MSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})}$$

   Note that p-1-g = number of β's in H_0

Note:

- $SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})$ measures the prediction error in the complete model.
- $SSE(X_1,\ldots,X_g)$ measures the prediction error in the reduced model.
- Since $SSE(X_1,\ldots,X_g)$ measures the error of a model with less variables than the complete model,
  $SSE(X_1,\ldots,X_g) \geq SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1})$
- $SSE(X_1,\ldots,X_g) - SSE(X_1,\ldots,X_g,X_{g+1},\ldots,X_{p-1}) = SSR(X_{g+1},\ldots,X_{p-1}|X_1,\ldots,X_g)$ measures how much prediction error is reduced by removing the p-1-g variables from the complete model (remember there are p-1-g β=0 in H_a).
- If $SSR(X_{g+1},\ldots,X_{p-1}|X_1,\ldots,X_g)$ is small, then F^* will be small leading to a "don't reject H_o" result. This suggests that since the prediction error is not affected much by the removal of variables, the reduced model may be better than the complete model.

- If $SSR(X_{g+1},\ldots,X_{p-1}|X_1,\ldots,X_g)$ is large, then F^* will be large leading to a "Reject H_o" result. This suggests that the prediction error has increased by a lot when the group of independent variables are removed. Therefore, use the complete model.

3) $F(1-\alpha, p-1-g, n-p)$

   p-1-g = numerator D.F.
   n-p = denominator D.F.

4) Conclusion
   - Reject H_o: The ____ variables are important in predicting Y (complete model is better).
   - Don't reject H_o: There is not sufficient evidence to show that _____ variables are important in predicting Y (reduced model may be better).

Suppose there are only three variables under consideration – X_1, X_2, and X_3. Special cases of the partial F test:

1) Test $\beta_{g+1}=0$ where g+1=3 for this example

$$F^* = \frac{MSR(X_3 \mid X_1, X_2)}{MSE(X_1, X_2, X_3)}$$ is equivalent to doing a t-test for $\beta_3=0$

2) Test $\beta_1=\beta_2=\beta_3=0$

$$F^* = \frac{MSR(X_1, X_2, X_3 \mid \ )}{MSE(X_1, X_2, X_3)}$$ is equivalent to doing a overall F test (Chapter 6)

Example (Body fat example continued):
For the following questions, write out the null and alternative hypotheses and perform the formal test:

1) We wish to test for the model with all three predictor variables whether midarm circumference (X3) can be dropped from the model **(This is example of testing whether a single parameter is 0; the F-test is equivalent to t-test. Check this out).**

2) We wish to test for the model with all three predictor variables whether both thigh circumference (X2) and midarm circumference (X3) can be dropped from the model **(This is example of testing whether several parameters are 0).**

Note: We can do these two tests using R directly (Check the R output).

```
> # Test if X3 can be dropped
> Full <- lm(Y~X1+X2+X3, data=Data)
> DropX3 <- lm(Y~X1+X2, data=Data)
> anova(DropX3, Full)
Analysis of Variance Table

Model 1: Y ~ X1 + X2
Model 2: Y ~ X1 + X2 + X3
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     17 109.951
2     16  98.405  1    11.546 1.8773 0.1896
>
> # test if both X2 and X3 can be dropped
> DropX2X3 <- lm(Y~X1, data=Data)
> anova(DropX2X3, Full)
Analysis of Variance Table

Model 1: Y ~ X1
Model 2: Y ~ X1 + X2 + X3
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     18 143.120
2     16  98.405  2    44.715 3.6352 0.04995 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
’ 1
>
```

7.3 Summary of tests concerning regression coefficients

1) Overall F-test – Tests the importance of all variables at once. The null hypothesis is $H_o:\beta_1=\ldots=\beta_{p-1}=0$

2) T-test - Tests the importance of only one variable at a time. The null hypothesis is $H_o:\beta_g=0$

3) Partial F-test – Tests the importance of a group of variables at the same time. The null hypothesis is $H_o:\beta_{g+1}=\ldots=\beta_{p-1}=0$ where there are $g \leq p-1$ independent variables in the reduced model.

The overall F-test and the t-test are special cases of the partial F-test.

4) Other tests:

For example, the full model containing 3 X variables:

$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\varepsilon_i$

a) Test $H_o:\beta_1=\beta_2=\beta_c$

$H_a:\beta_1\neq\beta_2$

The reduced model under is H_o: $Y_i=\beta_0+\beta_c(X_{1i}+X_{2i})+\beta_3X_{3i}+\varepsilon_i$

Use F test with 1 and n-4 degrees of freedom.

b) Test $H_o:\beta_1=3, \beta_2=5$

H_a: not both equalities in H_o hold

The reduced model under is H_o: $Y_i-(3X_{1i}+5X_{2i})=\beta_0+\beta_3X_{3i}+\varepsilon_i$

Use F test with 2 and n-4 degrees of freedom.

7.4 Coefficients of partial determination

Extra sums of squares can be used in calculating a R^2 given other variables are in the model.

> R^2 - measures the proportion reduction in the variation of Y achieved by using all of the independent variables in the model.
>
> $R^2_{Y3|12}$ - measures the proportion reduction in SSE by adding X_3 to the model given that X_1 and X_2 are already in the model (this is one specific example).

The "coefficient of partial determination" is calculated the following way for these examples:

$$R^2_{Y3|12} = \frac{SSR(X_3 \mid X_1, X_2)}{SSE(X_1, X_2)} = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{SSE(X_1, X_2)}$$

$$R^2_{Y3|2} = \frac{SSR(X_3 \mid X_2)}{SSE(X_2)} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)}$$

$$R^2_{Y1|234} = \frac{SSR(X_1 \mid X_2, X_3, X_4)}{SSE(X_2, X_3, X_4)} = \frac{SSE(X_2, X_3, X_4) - SSE(X_1, X_2, X_3, X_4)}{SSE(X_2, X_3, X_4)}$$

Notes:

1) The coefficient of partial determination is between 0 and 1. The closer to 1, the more the reduction in SSE.

2) The coefficient of partial determination is often used in "model building" procedures. To determine what independent variables should be added to a model, the coefficient of partial determination can be examined to see which independent variable reduces SSE the most.

Example: (Body fat example continued):

Find

$$R^2_{Y2|1} = \frac{SSR(X_2 \mid X_1)}{SSE(X_1)} = \frac{33.17}{143.12} = 0.232$$

$$R^2_{Y3|12} = \frac{SSR(X_3 \mid X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.54}{109.95} = 0.105$$

$$R^2_{Y1|2} = \frac{SSR(X_1 \mid X_2)}{SSE(X_2)} = 113.42 = 0.031$$

When X_2 is added to the regression model containing X_1 here, the error sum of squares SSE(X_1) is reduced by 23.2 percent.
The error of sum squares for the model containing X_1 and X_2, is only reduced by another 10.5 percent when X_3 is added to the model.
If the regression model already contains X_2, adding X_1 reduces by only 3.1 percent.

7.5 Standardized multiple regression model

- Control roundoff errors in normal equations calculation
  - When the number of X variables is small (say 3 or less) roundoff effects can be controlled by carrying a sufficient number of digits in intermediate calculations. Serious roundoff effects can arise with a large number of X variables.
- Permit comparisons of the estimated regression coefficients in common units
  - X variables have substantially different magnitudes
- The correlation transformation:

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \overline{Y}}{s_Y}\right)$$

$$X_{ik}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \overline{X}_k}{s_k}\right), k = 1, \cdots, p-1$$

- Standardized regression model (notice **no intercept, why**):

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_1^* X_{i2}^* + \cdots \beta_1^* X_{i,p-1}^* + \varepsilon_i^*$$

- In this case, the X'X matrix is the **correlation matrix of the X variables.** All elements are between -1 and 1.

- The relationship between the standardized regression coefficients and the original regression coefficients:

$$\beta_k = \frac{s_Y}{s_k}\beta_k^* \qquad k = 1, \cdots, p-1$$

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}_1 - \beta_2 \overline{X}_2 \cdots - -\beta_{p-1}\overline{X}_{p-1}$$

Example: (Figure 6.5 data) Dwaine Studios
Y: sales in a community, expressed in thousands of dollars
X1: number of persons aged 16 or younger in the community, expressed in thousands of persons
X2: per capita disposable personal income in the community, expressed in thousands of dollars

Part of the data set:

X1	X2	Y
68.5	16.7	174.4
45.2	16.8	164.4
91.3	18.2	244.2

R codes and output:

```
> # Chapter 7: Standardized regression model
> # Read in the data
> Data <- read.table(file="CH06FI05.txt", header=FALSE)
> colnames(Data) <- c("X1", "X2", "Y")
> # number of observations
> n <- dim(Data)[1]
> # correlation matrix
> cor(Data)
          X1        X2         Y
X1 1.0000000 0.7812993 0.9445543
X2 0.7812993 1.0000000 0.8358025
Y  0.9445543 0.8358025 1.0000000
>
> # Correlation transformation
> X1 <- Data$X1
> X1.Star <- 1/sqrt(n-1)*(X1-mean(X1))/sd(X1)
> X2 <- Data$X2
> X2.Star <- 1/sqrt(n-1)*(X2-mean(X2))/sd(X2)
> Y <- Data$Y
> Y.Star <- 1/sqrt(n-1)*(Y-mean(Y))/sd(Y)
> # the design matrix X corresponds to the transformed varaibles
> X <- cbind(X1.Star, X2.Star)
>
> # look at X'X and X'Y
> t(X)%*%X
          X1.Star   X2.Star
X1.Star 1.0000000 0.7812993
X2.Star 0.7812993 1.0000000
> t(X)%*%Y.Star
             [,1]
X1.Star 0.9445543
X2.Star 0.8358025
>
> # Estimate the regression using the transformed varaibles
> b <- solve(t(X)%*%X) %*% t(X)%*%Y.Star
> b
             [,1]
X1.Star 0.7483670
X2.Star 0.2511039
>
> # Transform back to original regression coefficients
> b2 <- sd(Y)/sd(X2)*b[2]
> b1 <- sd(Y)/sd(X1)*b[1]
> b0 <- mean(Y)-b1*mean(X1) - b2*mean(X2)
> print(c(b0,b1,b2))
[1] -68.857073   1.454560   9.365500
```

7.6 Multicollinearity and its effects

When two predictor variables X_1 and X_2 are uncorrelated, we have:
$SSR(X_2) = SSR(X_2|X_1)$ and $SSR(X_1) = SSR(X_1|X_2)$

When independent variables are highly correlated with each other, "intercorrelation" or "multicollinearity" is said to exist. This can cause estimates of the β's to be "unstable" (meaning that from sample-to-sample-to-sample, the $\hat{\beta}$'s have a lot of variability).

Therefore, interpretation of how an independent and dependent variable are related by examining the $\hat{\beta}$ may not give good results.

Example 1: Table 7.6 Work crew productivity example. Investigate the effect of work crew size (X1) and level of bonus pay (X2) on crew productivity (Y).

X1	X2	Y
4	2	42
4	2	39
4	3	48
4	3	51
6	2	49
6	2	53
6	3	61
6	3	60

(a) Regression of Y on X1 and X2
Yhat = 0.375+5.375X1+9.250X2

(b) Regression of Y on X1
Yhat = 23.5+5.375X1

(c) Regression of Y on X2
Yhat = 27.25+9.250X2

Notice that X1 and X2 are uncorrelated. Pay attention to the estimates of the parameters in the three modes.

Example 2: Table 7.8 (page 281)

X1	X2	Y
2	6	23
8	9	83
6	8	63
10	10	103

Two fitted response functions

(1) Yhat = -87+X1+18X2

(2) Yhat = -7+9X1+2X2

Check in fact X2 = 5 + 0.5X1. So X1 and X2 are perfectly correlated. The two fitted response functions are entirely different response surfaces. They have the same fitted values only when they intersect.

When multicollinearity does not exist, the estimated values of the parameters should not greatly change when variables are added or removed from the model.

Notes:

1) In Chapter 10, other methods are introduced to detect multicollinearity.
2) In Chapter 11, remedial measures are introduced to lessen the effect of multicollinearity.
3) Multicollinearity has effects on regression coefficients and the precision of the estimates.
4) Multicollinearity has effects on extra sums of squares.
5) Multicollinearity generally does not affect the C.I.s for E(Y) or P.I.s for Y.
6) Multicollinearity has effects on simultaneous tests of parameters.

Example (body fat example) can be used to demonstrate points 3)-6):

We have the following facts:

1) X_1 and X_2 are highly correlated (coefficient of simple correlation is 0.924)
2) X_3 is not highly related to X_1 and X_2 individually (0.458 and 0.085)
3) X_3 is highly correlated with X_1 and X_2 together (the coefficient of multiple determination when X_3 is regressed on X_1 and X_2 is 0.9980)

T

The effects of multicollinearity on

- Regression coefficients

Variables in model	b_1	b_2
X_1	0.8572	-
X_2	-	0.8565
X_1, X_2	0.2224	0.6594
X_1, X_2, X_3	4.334	-2.857

- Extra Sums of squares
  $SSR(X_1) = 352.27$; $SSR(X_1 \mid X_2) = 3.47$

- $s\{b_k\}$

Variables in model	$s\{b_1\}$	$s\{b_2\}$
X_1	0.1288	-
X_2	-	0.1100
X_1, X_2	0.3034	0.2912
X_1, X_2, X_3	3.016	2.582

- Fitted values and predictions
  Suppose $X_{h1}=25.0$, $X_{h2}=50.0$, $X_{h3}=29.0$

Variables in model	MSE	$\hat{Y}_h$	$s\{\hat{Y}_h\}$
X_1	7.95	19.93	0.632
X_1, X_2	6.47	19.36	0.624
X_1, X_2, X_3	6.15	19.19	0.621

- Simultaneous tests of β's
  Look at the model containing X_1 and X_2 only.

  Suppose significance level α=0.01.Using F-test, we reject the H_0: $\beta_1 = \beta_2 = 0$, and conclude that not both coefficients equal 0.

Analysis of Variance Table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	385.43871	192.71935	29.80	<.0001
Error	17	109.95079	6.46769		
Corrected Total	19	495.38950			

Controlling the family level of significance at 0.01, using two separate t-tests, we conclude that $\beta_1 = 0$ and $\beta_2 = 0$.

```
> summary(lm(Y~X1+X2, data=Data))

Call:
lm(formula = Y ~ X1 + X2, data = Data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9469 -1.8807  0.1678  1.3367  4.0147

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.1742     8.3606  -2.293   0.0348 *
X1            0.2224     0.3034   0.733   0.4737
X2            0.6594     0.2912   2.265   0.0369 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.543 on 17 degrees of freedom
Multiple R-squared: 0.7781,     Adjusted R-squared: 0.7519
F-statistic:  29.8 on 2 and 17 DF,  p-value: 2.774e-06
```