Chapter 7

Multiple Regression II

Instructor: Li C.X.

Outline

- Extra sums of squares
- General linear test
- Partial determination and partial correlation
- Standardized version of the multiple regression model
- Multicollinearity

7.1 Extra Sums of Squares

- For a given dataset, the total sum of squares remains the same, no matter what predictors are included
- As we include more predictors, the regression sum of squares (SSR) increases (technically does not decrease), and the error sum of squares (SSE) decreases
- SSR + SSE = SSTO, regardless of predictors in model
- An extra sum of squares measures the marginal increase in the regression sum of squares, or decrease in the error sum of squares, when one or several predictor variables are added.

Example

- Output: Body fat percentage via underwater weighing
  - Underwater weighing is expensive/difficult
- Input:

1. triceps skin fold thickness(X_1) (肱三头肌皮褶厚度)
2. thigh circumference (X_2)
3. midarm circumference (X_3)

Subject i	Triceps Skinfold Thickness X_{i1}	Thigh Circumference X_{i2}	Midarm Circumference X_{i3}	Body Fat Y_i
1	19.5	43.1	29.1	11.9
2	24.7	49.8	28.2	22.8
3	30.7	51.9	37.0	18.7
...	...	...	...	...
18	30.2	58.6	24.6	25.4
19	22.7	48.2	27.1	14.8
20	25.2	51.0	27.5	21.1

(a) Regression of Y on X_1

$$\hat{Y} = -1.496 + .8572X_1$$

Source of Variation	SS	df	MS
Regression	352.27	1	352.27
Error	143.12	18	7.95
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = .8572$	$s\{b_1\} = .1288$	6.66

(b) Regression of Y on X_2

$$\hat{Y} = -23.634 + .8565X_2$$

Source of Variation	SS	df	MS
Regression	381.97	1	381.97
Error	113.42	18	6.30
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_2	$b_2 = .8565$	$s\{b_2\} = .1100$	7.79

(c) Regression of Y on X_1 and X_2

$$\hat{Y} = -19.174 + .2224X_1 + .6594X_2$$

Source of Variation	SS	df	MS
Regression	385.44	2	192.72
Error	109.95	17	6.47
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = .2224$	$s\{b_1\} = .3034$	.73
X_2	$b_2 = .6594$	$s\{b_2\} = .2912$	2.26

(d) Regression of Y on X_1, X_2, and X_3

$$\hat{Y} = 117.08 + 4.334X_1 - 2.857X_2 - 2.186X_3$$

Source of Variation	SS	df	MS
Regression	396.98	3	132.33
Error	98.41	16	6.15
Total	495.39	19	

Variable	Estimated Regression Coefficient	Estimated Standard Deviation	t^*
X_1	$b_1 = 4.334$	$s\{b_1\} = 3.016$	1.44
X_2	$b_2 = -2.857$	$s\{b_2\} = 2.582$	−1.11
X_3	$b_3 = -2.186$	$s\{b_3\} = 1.596$	−1.37

Extra Sums of Squares

- When a model contains just X_1, denote: $SSR(X_1)$, $SSE(X_1)$
- Model Containing X_1, X_2: $SSR(X_1, X_2)$, $SSE(X_1, X_2)$

For the example,

- $SSR(X_1) = 352.27$ $\quad$ $SSE(X_1) = 143.12$
- $SSR(X_1, X_2) = 385.44$ $\quad$ $SSE(X_1, X_2) = 109.95$
- Extra sum of squares

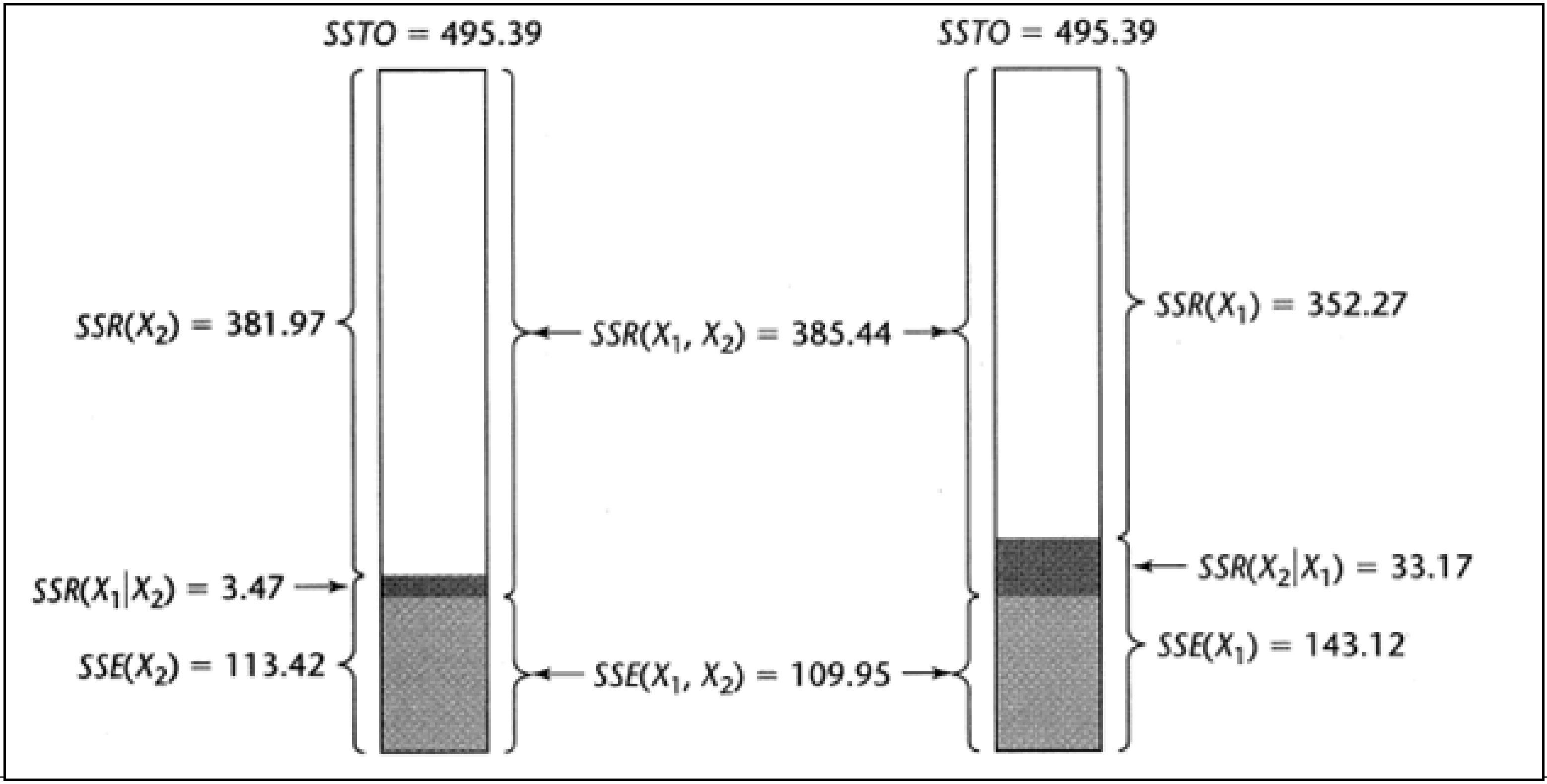
$$SSR(X_2|X_1) = SSR(X_1,X_2) - SSR(X_1) = SSE(X_1) - SSE(X_1,X_2) = 33.17$$

- The extra sum of squares $SSR(X_2 \mid X_1)$ measure the marginal effect of adding X_2 to the regression model when X_1 is already in the model

Extra sum of squares

We can switch the order of X_1 and X_2 in these expressions

- $SSR(X_2) = 381.97$ $SSE(X_2) = 113.42$
- $SSR(X_1, X_2) = 385.44$ $SSE(X_1, X_2) = 109.95$
- $SSR(X_1 \mid X_2) = SSE(X_2) - SSE(X_1, X_2) = 3.47$



Extra sum of squares

- Definition of extra sum of squares

$$\begin{aligned} SSR(X_2 \mid X_1) &= SSR(X_1, X_2) - SSR(X_1) \\ &= SSE(X_1) - SSE(X_1, X_2) \end{aligned}$$

- Extends to any number of Predictors

$$\begin{aligned} SSR(X_3 \mid X_1, X_2) &= SSR(X_1, X_2, X_3) - SSR(X_1, X_2) \\ &= SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \end{aligned}$$

$$SSR(X_1, X_2) \;=\; SSR(X_1) \;+\; SSR(X_2 | X_1)$$

$$\boxed{\sigma^2\chi^2(2,\ \delta_{R2})} \qquad \boxed{\sigma^2\chi^2(1,\ \delta_{R1})} \qquad \boxed{\sigma^2\chi^2(1,\ \delta_{R2} - \delta_{R1})}$$

$$\delta_{R2} = \frac{1}{\sigma^2}\sum_{k=1}^{2}\sum_{l=1}^{2} SS_{kl}\beta_k\beta_l,\ \ \delta_{R1} = \frac{1}{\sigma^2} SS_{XX}\beta_1^2 \quad \Rightarrow \delta_{R2} - \delta_{R1} = 0, \text{if } \beta_2 = 0$$

Decomposition of SSR

- Similarly

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2 \mid X_1) + SSR(X_3 \mid X_1, X_2)$$

$$SSTO = SSR(X_1, X_2, X_3) + SSE(X_1, X_2, X_3)$$

Source	SS	df	MS
Regression	SSR(X1,X2,X3)	3	MSR(X1,X2,X3)
X1	SSR(X1)	1	MSR(X1)
X2|X1	SSR(X2|X1)	1	MSR(X2|X1)
X3|X1,X2	SSR(X3|X1,X2)	1	MSR(X3|X1,X2)
Error	SSE(X1,X2,X3)	n-4	MSE(X1,X2,X3)
Total	SSTO	n-1	

7.2 Use Extra Sums of Squares In Tests

- General linear test for Single β_k
- Test whether a single $\beta_k = 0$
- Example: First order model with three predictor variables

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

- To test $H_0 : \beta_3 = 0$ vs $H_1 : \beta_3 \neq 0$

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$

Reduced model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$

General linear test for Single β_k

- For the full model we have SSE(F) = SSE(X_1,X_2, X_3)
- For the reduced model we have SSE(R) = SSE(X_1,X_2)
- General linear test

$$F^* = \frac{SSE(R)-SSE(F)}{df_R-df_F} \Big/ \frac{SSE(F)}{df_F} \overset{H_0}{\sim} F(1, n-4)$$

$$F^* = \frac{SSE(X_1,X_2)-SSE(X_1,X_2,X_3)}{(n-3)-(n-4)} \Big/ \frac{SSE(X_1,X_2,X_3)}{n-4}$$

$$\text{SSE}(X_1,X_2) - \text{SSE}(X_1,X_2, X_3) = \text{SSR}(X_3 \mid X_1,X_2)$$

$$F^* = \frac{SSR(X_3|X_1,X_2)}{1} \Big/ \frac{SSE(X_1,X_2,X_3)}{n-4} = \frac{MSR(X_3|X_1,X_2)}{MSE(X_1,X_2,X_3)}$$

Rejection Region: $F^* \geq F(1-\alpha; 1, n-4)$ P-value $= P\left(F(1;4) \geq F^*\right)$

Body fat example

- Body fat: Can X_3 (midarm circumference) be dropped from the model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$

Source of Variation	SS	df	MS
Regression	396.98	3	132.33
X_1	352.27	1	352.27
$X_2 \mid X_1$	33.17	1	33.17
$X_3 \mid X_1, X_2$	11.54	1	11.54
Error	98.41	16	6.15
Total	495.39	19	

$$F^* = \frac{SSR(X_3|X_1,X_2)}{1} \Big/ \frac{SSE(X_1,X_2,X_3)}{n-4} = 1.88$$

- For α=0.01 we require $F(0.99; 1,16) = 8.53 > 1.88$
- We conclude $H_0 : \beta_3 = 0$

Test whether several $\beta_k = 0$

- For example: $H_0 : \beta_2 = \beta_3 = 0$ vs H_1 : not both are zero
- The general linear test can be used again

$$F^* = \frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{(n-2)-(n-4)} \Big/ \frac{SSE(X_1, X_2, X_3)}{n-4}$$

- But $\text{SSE}(X_1) - \text{SSE}(X_1, X_2, X_3) = \text{SSR}(X_2, X_3 \mid X_1)$

$$\Rightarrow \quad F^* = \frac{\left[\dfrac{SSR(X_2, X_3 \mid X_1)}{2}\right]}{\left[\dfrac{SSE(X_1, X_2, X_3)}{n-4}\right]} = \frac{MSR(X_2, X_3 \mid X_1)}{MSE(X_1, X_2, X_3)}$$

Test whether several $\beta_k = 0$

Full Model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \quad \varepsilon_i \sim NID\left(0, \sigma^2\right)$

$H_0 : \beta_q = \ldots = \beta_{p-1} = 0 \quad H_A :$ At least one of $\beta_q \ \ldots \ \beta_{p-1} \neq 0$

$\Rightarrow$ Reduced Model: $Y_i = \beta_0 + \beta_1 X_{i1} \ldots + \beta_{q-1} X_{i,q-1} + \varepsilon_i \quad (q < p)$

Full Model: $SSE(F) = SSE\left(X_1, X_2 \ldots, X_{p-1}\right) \qquad df_F = n - p$

Reduced Model: $SSE(R) = SSE\left(X_1, X_2 \ldots, X_{q-1}\right) \quad df_R = n - q$

General Linear Test: $F^* = \dfrac{\left[\dfrac{SSE(R) - SSE(F)}{df_R - df_F}\right]}{\left[\dfrac{SSE(F)}{df_F}\right]} \overset{H_0}{\sim} \quad F\left(p - q, n - p\right)$

$$\Rightarrow \quad F^* = \frac{\left[\dfrac{SSR\left(X_q, \ldots, X_{p-1} \mid X_1, X_2 \ldots, X_{q-1}\right)}{p - q}\right]}{\left[\dfrac{SSE\left(X_1, X_2 \ldots, X_{p-1}\right)}{n - p}\right]} = \frac{MSR\left(X_q, \ldots, X_{p-1} \mid X_1, X_2 \ldots, X_{q-1}\right)}{MSE\left(X_1, X_2 \ldots, X_{p-1}\right)}$$

7.3 Summary of Tests

- Test whether all $\beta_k = 0$
- Test whether a single $\beta_k = 0$
- Test whether some $\beta_k = 0$

- Test involving relationships among coefficients, for example,
  - $H_0 : \beta_1 = \beta_2$ vs. $H_a : \beta_1 \neq \beta_2$
  - $H_0 : \beta_1 = 3, \beta_2 = 5$ vs. H_a : otherwise
  - H_0: $\beta_1 - 2\beta_2 + \beta_3 = 0$ vs H_a: $\beta_1 - 2\beta_2 + \beta_3 \neq 0$

- Key point in all tests: form the full model and the reduced model, then calculate the SSE(F) and SSE(R).

7.4 Coefficients of Partial Determination

Regression of Y on X_1: $\quad Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$

Variation Explained: $SSR(X_1)$ $\quad$ Unexplained: $SSE(X_1) = SSTO - SSR(X_1)$

Regression of Y on X_2: $\quad Y_i = \beta_0 + \beta_2 X_{i2} + \varepsilon_i$

Variation Explained: $SSR(X_2)$ $\quad$ Unexplained: $SSE(X_2) = SSTO - SSR(X_2)$

Regression of Y on X_1, X_2: $\quad Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

Variation Explained: $SSR(X_1, X_2)$ $\quad$ Unexplained: $SSE(X_1, X_2) = SSTO - SSR(X_1, X_2)$

Proportion of Variation in Y, Not Explained by X_1, that is Explained by X_2:

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

Proportion of Variation in Y, Not Explained by X_2, that is Explained by X_1:

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

Coefficients of Partial Determination

- Partial determination measures the marginal contribution of one X variable when others are already in the model.
- Coefficient of partial determination between Y and X_1 given X_2 in the model is denoted as

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

  - Measures ***additional*** information in X_1 helping predict Y
- Similarly,

$$R^2_{Y2|1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)} = \frac{SSR(X_2|X_1)}{SSE(X_1)}$$

Coefficients of Partial Determination

- General case :Consider model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- Coefficient of Partial Determination

$$R^2_{Y1|2:(p-1)} = \frac{SSR\left(X_1 \mid X_2, \cdots, X_{p-1}\right)}{SSE\left(X_2, \cdots, X_{p-1}\right)}$$

- Define two variables:
  - residuals of predicting Y as function of $X_2, \cdots, X_{p-1}$

$$e_i(Y|X_2, \cdots, X_{p-1}) = Y_i - \hat{Y}_i(X_2, \cdots, X_{p-1})$$

  - residuals of predicting X_1 as function of $X_2, \cdots, X_{p-1}$

$$e_i(X_1|X_2, \cdots, X_{p-1}) = X_{i1} - \hat{X}_{i1}(X_2, \cdots, X_{p-1})$$

- $R^2_{Y1|2:(p-1)}$ equals to R^2 for regressing $e_i(Y|X_2, \cdots, X_{p-1})$ on $e_i(X_1|X_2, \cdots, X_{p-1})$

Coefficients of Partial Correlation

- Coefficients of Partial Determination is between 0 and 1.
- Coefficients of Partial Correlation:
  - square root of a coefficient of partial determination, following the same sign with the regression coefficient.

$$R_{Y\,2|1} = \text{sgn}\{\beta_2\}\sqrt{R^2_{Y\,2|1}}$$

7.5 Standardized Regression Model

- Numerical precision errors can occur when $(X'X)^{-1}$ is poorly conditioned near singular :
  - colinearity
  - when the predictor variables have substantially different magnitudes
- Standardized multiple regress
- Makes easier comparison of magnitude of effects of predictors measured on different measurement scales

Standardized Regression Model

- First, transformed variables

$$\frac{Y_i - \bar{Y}}{s_y} \qquad s_y = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{n-1}}$$

$$\frac{X_{ik} - \bar{X}_k}{s_k}, k = 1, \ldots, p-1 \qquad s_k = \sqrt{\frac{\sum (X_{ik} - \bar{X}_k)^2}{n-1}}, k = 1, \ldots, p-1$$

- Correlation Transformation

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{s_y}\right) \qquad X_{ik}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ik} - \bar{X}_k}{s_k}\right), k = 1, \ldots, p-1$$

- The regression model using the transformed variables:

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

Standardized Regression Model

- Let
$$X^* = \begin{pmatrix} X^*_{11} & \cdots & X^*_{1,p-1} \\ X^*_{21} & \cdots & X^*_{2,p-1} \\ \cdots & & \\ X^*_{n1} & \cdots & X^*_{n,p-1} \end{pmatrix} \quad r_{XX} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{21} & 1 & \cdots & r_{2,p-1} \\ \cdots & & & \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{pmatrix}$$
- Then $(X^*)'X^* = r_{XX} \qquad (X^*)'Y^* = r_{XY}$
  - Note that
$$\begin{aligned} \sum x^*_{i1}x^*_{i2} &= \sum\left(\frac{X_{i1}-\bar{X}_1}{\sqrt{n-1}s_1}\right)\left(\frac{X_{i2}-\bar{X}_2}{\sqrt{n-1}s_2}\right) \\ &= \frac{1}{n-1}\frac{\sum(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)}{s_1 s_2} \\ &= \frac{\sum(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)}{[\sum(X_{i1}-\bar{X}_1)^2\sum(X_{i2}-\bar{X}_2)^2]^{1/2}} \end{aligned}$$
- Makes all entries in $(X'X)$ matrix for the transformed variables fall between -1 and 1 inclusive

Standardized Regression Model

- The regression model using the transformed variables:

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

- Coefficients represent changes in Y in standard deviation (SD) units as each predictor increases 1 SD (holding all others constant)
- Then the LSE or MLE estimators satisfy

$$\mathbf{r}_{XX}\mathbf{b}^* = \mathbf{r}_{XY} \;\Rightarrow\; \mathbf{b}^* = \left(b_1^*, b_2^*, \cdots, b_{p-1}^*\right)^T = \mathbf{r}_{XX}^{-1}\mathbf{r}_{XY}$$

$$\Rightarrow \quad \begin{aligned} b_k &= \left(\tfrac{s_Y}{s_k}\right) b_k^*,\ k = 1, \ldots, p-1 \\ b_0 &= \bar{Y} - b_1\bar{X}_1 - \ldots - b_{p-1}\bar{X}_{p-1} \end{aligned}$$

7.6 Multicollinearity

- Consider model with 2 predictors (this generalizes to any number of predictors) $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$
- When X_1 and X_2 are uncorrelated, the regression coefficients b_1 and b_2 are the same whether we fit simple regressions or a multiple regression, and:

$$\mathrm{SSR}(X_1) = \mathrm{SSR}(X_1 \mid X_2) \qquad \mathrm{SSR}(X_2) = \mathrm{SSR}(X_2 \mid X_1)$$

- When X_1 and X_2 are highly correlated, their regression coefficients become unstable, and their standard errors become larger (smaller t-statistics, wider CI^{s}), leading to strange inferences when comparing simple and partial effects of each predictor
- Estimated means and predicted values are not affected

Perfectly Correlated Predictor Variables

Regress Y on both X_1 and X_2. If X_1 and X_2 are perfectly correlated (say $X_2 = 5 + .5X_1$), then

- We have infinitely many possible solutions which fits the model equally well (have the same SSE).
- The perfect relation between X_1 and X_2 does not inhibit our ability to obtain a good fit.
- Usually, we still have good fit of the data, in addition, we still have good prediction.
- The estimated regression coefficients tends to have large sampling variability when the predictor variables are highly correlated.

R Code

```
dat = read.table('fat.txt')
X1 = dat[,1];  X2 = dat[,2]; X3 = dat[,3];Y = dat[,4]
fit1 = lm(Y~X1);   fit2 = lm(Y~X2)
fit12 = lm(Y~X1+X2); fit = lm(Y~X1+X2+X3)
SSE1 = deviance(fit1);  SSE2 = deviance(fit2)
SSE12 = deviance(fit12);  SSE123 = deviance(fit)
SSR1.2 = deviance(fit2)-deviance(fit12)
SSE2 = deviance(fit2);   RY1.2 = SSR1.2/SSE2
###another way of calculating
e1 = residuals(lm(Y~X2))
e2 = residuals(lm(X1~X2))
cor(e1,e2)^2
```

Homework

- P290

  7.3 7.10 7.12 7.16 7.24 7.30

  7.31 State the reduced models and give the tests for testing whether or not: (1),(2),(3)