Two-way Analysis of Variance (two-way ANOVA, 二因子變異數分析)

- One-way ANOVA may be used to determine whether the means of a response under different levels of a factor are all the same.
- Two-way ANOVA may be used when there are two factors, and we want to determine
  - (i) whether the main effect for each factor exists and
  - (ii) whether there is an interaction effect between the two factors.
- Data assumption for two-way ANOVA. Let A and B be the two factors. Let $X_{i,j,m}$ be the m-th observed response when Factor A is of level i and Factor B is of level j. Suppose that

$$X_{i,j,m} = \mu_{i,j} + \varepsilon_{i,j,m}$$

for $1 \leq i \leq k$, $1 \leq j \leq b$, $1 \leq m \leq \ell$, where ℓ is the number of observations when Factor A is of level i and Factor B is of level j and $\varepsilon_{i,j,m}$'s are IID $N(0, \sigma^2)$. The mean $\mu_{i,j}$ is often expressed as

$$\mu_{i,j} = \mu + \alpha_i + \beta_j + \gamma_{i,j}, \tag{1}$$

where

$$\sum_{i=1}^{k} \alpha_i = 0 = \sum_{j=1}^{b} \beta_j,$$

$$\sum_{j=1}^{b} \gamma_{i,j} = 0 \text{ for each } i \text{ and } \sum_{i=1}^{k} \gamma_{i,j} = 0 \text{ for each } j.$$

- Testing problems.
  - Testing whether there is no main effect of Factor A.

$$H_0 : \alpha_1 = \cdots = \alpha_k = 0. \tag{2}$$

  - Testing whether there is no main effect of Factor B.

$$H_0 : \beta_1 = \cdots = \beta_b = 0 \tag{3}$$

  - Test whether there is no interaction effect between Factor A and Factor B.

$$H_0 : \gamma_{1,1} = \cdots = \gamma_{k,b} = 0 \tag{4}$$

- The decomposition of $\mu_{i,j}$ in (1) can be obtained using the following equations:

$$\mu = \frac{\sum_{i=1}^{k} \sum_{j=1}^{b} \mu_{i,j}}{kb},$$

$$\alpha_i = \frac{\sum_{j=1}^{b} \mu_{i,j}}{b} - \mu,$$

$$\beta_j = \frac{\sum_{i=1}^{k} \mu_{i,j}}{k} - \mu,$$

and

$$\gamma_{i,j} = \mu_{i,j} - \mu - \alpha_i - \beta_j.$$

Example 1. For $1 \le i \le 3$, $1 \le j \le 2$, let $\mu_{i,j}$ be as in the following table:

$\mu_{i,j}$	$i = 1$	$i = 2$	$i = 3$
$j = 1$	5	4	3
$j = 2$	7	5	3

Find μ, α_i, β_j and $\gamma_{i,j}$ for $1 \le i \le 3$ and $1 \le j \le 2$ so that

$$\mu_{i,j} = \mu + \alpha_i + \beta_j + \gamma_{i,j}$$

for $1 \le i \le 3$, $1 \le j \le 2$, where

$$\sum_{i=1}^{3} \alpha_i = 0 = \sum_{j=1}^{2} \beta_j,$$

$$\sum_{j=1}^{2} \gamma_{i,j} = 0 \text{ for each } i \text{ and } \sum_{i=1}^{3} \gamma_{i,j} = 0 \text{ for each } j.$$

Solution.

- Compute $\frac{\sum_{i=1}^{3} \mu_{i,j}}{3}$ and $\frac{\sum_{j=1}^{2} \mu_{i,j}}{2}$:

$\mu_{i,j}$	$i = 1$	$i = 2$	$i = 3$	$\frac{\sum_{i=1}^{3} \mu_{i,j}}{3}$
$j = 1$	5	4	3	$\frac{5+4+3}{3} = 4$
$j = 2$	7	5	3	$\frac{7+5+3}{3} = 5$
$\frac{\sum_{j=1}^{2} \mu_{i,j}}{2}$	$\frac{5+7}{2} = 6$	$\frac{4+5}{2} = 4.5$	$\frac{3+3}{2} = 3$	

- Compute μ:

$$\mu = \frac{5+4+3+7+5+3}{6} = \frac{4+5}{2} = 4.5.$$

- Compute $\alpha_i = \frac{\sum_{j=1}^{2} \mu_{i,j}}{2} - \mu$:

	$i=1$	$i=2$	$i=3$
$\frac{\sum_{j=1}^{2} \mu_{i,j}}{2}$	6	4.5	3
α_i	$6-4.5=1.5$	$4.5-4.5=0$	$3-4.5=-1.5$

- Compute $\beta_j = \frac{\sum_{i=1}^{3} \mu_{i,j}}{3} - \mu$:

	$\frac{\sum_{i=1}^{3} \mu_{i,j}}{3}$	β_j
$j=1$	4	$4-4.5=-0.5$
$j=2$	5	$5-4.5=0.5$

- Compute $\gamma_{i,j} = \mu_{i,j} - \mu - \alpha_i - \beta_j$: From previous calculation, we have $\mu = 4.5$ and

$\mu_{i,j}$	$i=1$	$i=2$	$i=3$	β_j
$j=1$	5	4	3	-0.5
$j=2$	7	5	3	0.5
α_i	1.5	0	-1.5	

Therefore,

$\gamma_{i,j}$	$i=1$	$i=2$	$i=3$	β_j
$j=1$	$5-4.5-1.5-(-0.5)$ $=-0.5$	$4-4.5-0-(-0.5)$ $=0$	$3-4.5-(-1.5)-(-0.5)$ $=0.5$	-0.5
$j=2$	$7-4.5-1.5-(0.5)$ $=0.5$	$5-4.5-0-0.5$ $=0$	$3-4.5-(-1.5)-0.5$ $=-0.5$	0.5
α_i	1.5	0	-1.5	

The main effects of Factors A and B exist, and there is an interaction effect between the two factors.

- Two-way ANOVA example.
  - Response: weight loss.
  - Factor A: receiving different drugs.

- Factor B: doing different exercises.

- Related statistics for testing problems in (2) – (4).
  - The grand mean (overall sample mean) is $\bar{X}_G = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{m=1}^{\ell} X_{i,j,m}$, where $n = kb\ell$.
  - Let $\bar{X}_{i\cdot}$ be the sample mean for the Factor A-level i group, then
$$\bar{X}_{i\cdot} = \frac{\sum_{j=1}^{b}\sum_{m=1}^{\ell} X_{i,j,m}}{b\ell}.$$
  - Let $\bar{X}_{\cdot j}$ be the sample mean and sample standard deviation for the Factor B-level j group, then
$$\bar{X}_{\cdot j} = \frac{\sum_{i=1}^{k}\sum_{m=1}^{\ell} X_{i,j,m}}{k\ell}.$$
  - Let $\bar{X}_{ij}$ be the sample mean for the Factor A-level i-Factor B-level j group, then
$$\bar{X}_{ij} = \frac{\sum_{m=1}^{\ell} X_{i,j,m}}{\ell}.$$
  - The total sum of squares, denoted by $\mathrm{SS}_{\text{total}}$, is
$$\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{m=1}^{\ell}(X_{i,j,m} - \bar{X}_G)^2.$$
  - The sum of squares due to Factor A. denoted by SSA, is
$$\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{m=1}^{\ell}(\bar{X}_{i\cdot} - \bar{X}_G)^2 = b\ell\sum_{i=1}^{k}(\bar{X}_{i\cdot} - \bar{X}_G)^2. \tag{5}$$
    Note that the SSA is the SST (sum of squares due to treatment) in one-way ANOVA using A as the treatment factor.
  - The sum of squares due to Factor B, denoted by SSB, is
$$\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{m=1}^{\ell}(\bar{X}_{\cdot j} - \bar{X}_G)^2 = k\ell\sum_{j=1}^{b}(\bar{X}_{\cdot j} - \bar{X}_G)^2. \tag{6}$$
  - The sum of squares due to interaction, denoted by SSI, is
$$\sum_{i=1}^{k}\sum_{j=1}^{b}\sum_{m=1}^{\ell}(\bar{X}_{i,j} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_G)^2$$

$$
\begin{aligned}
&= \ell \sum_{i=1}^{k} \sum_{j=1}^{b} (\bar{X}_{i,j} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_G)^2 \\
&= \ell \sum_{i=1}^{k} \sum_{j=1}^{b} (\bar{X}_{i,j} - \bar{X}_G)^2 - \mathrm{SSA} - \mathrm{SSB} \qquad (7) \\
&= \ell \left(\sum_{i=1}^{k} \sum_{j=1}^{b} \bar{X}_{i,j}^2 \right) - \ell k b (\bar{X}_G)^2 - \mathrm{SSA} - \mathrm{SSB}.
\end{aligned}
$$

- The SSE (sum of squares due to error) is

$$\sum_{i=1}^{k} \sum_{j=1}^{b} \sum_{m=1}^{\ell} (X_{i,j,m} - \bar{X}_{i,j})^2.$$

- Note that one can show that

$$\mathrm{SS}_{\mathrm{total}} = \mathrm{SSA} + \mathrm{SSB} + \mathrm{SSI} + \mathrm{SSE}$$

by verifying (7) and using the result that

$$
\begin{aligned}
&\mathrm{SS}_{\mathrm{total}} - \mathrm{SSE} \\
&= \ell \sum_{i=1}^{k} \sum_{j=1}^{b} (\bar{X}_{i,j} - \bar{X}_G)^2 \\
&= \ell \left(\sum_{i=1}^{k} \sum_{j=1}^{b} \bar{X}_{i,j}^2 \right) - \ell k b (\bar{X}_G)^2.
\end{aligned}
$$

- For testing (2), the test statistic is

$$F = \frac{\mathrm{SSA}/(k-1)}{\mathrm{SSE}/(n-kb)}.$$

  - $F \sim F(k-1, n-kb)$ under H_0.
  - The F test rejects H_0 at level α if $F > f_{\alpha,k-1,n-kb}$.

- For testing (3), the test statistic is

$$F = \frac{\mathrm{SSB}/(b-1)}{\mathrm{SSE}/(n-kb)}.$$

  - $F \sim F(b-1, n-kb)$ under H_0.
  - The F test rejects H_0 at level α if $F > f_{\alpha,b-1,n-kb}$.

- For testing (4), the test statistic is

$$F = \frac{\text{SSI}/(k-1)(b-1)}{\text{SSE}/(n-kb)}.$$

  - $F \sim F((k-1)(b-1), n-kb)$ under H_0.
  - The F test rejects H_0 at level α if $F > f_{\alpha,(k-1)(b-1),n-kb}$.

Example 2. Suppose that we are interested in the effects of taking three weight loss drugs D_1, D_2 and D_3 while doing two types of exercises: jogging or walking, at the same time. 30 participants are assigned to receive one of the three drugs and required to jog or walk for 40 mininues 3 times per week. Their weight losses (in kilograms) are given below.

Jogging group				Walking group			
Drug	D_1	D_2	D_3	Drug	D_1	D_2	D_3
count	5	5	5	count	5	5	5
sample mean	5.4	4.6	6.7	sample mean	5.1	4.3	4.5
sample variance	0.04	0.01	0.09	sample variance	0.01	0.01	0.04

Determine whether the following conclusions can be made at the 0.05 significance level.

(a) Not all drugs have the same effect on weight loss.

(b) Not all exercises have the same effect on weight loss.

(c) There is a drug-exercise interaction effect on weight loss.

- Solution.

  (a) The grand mean:

$$\bar{X}_G = \frac{5(5.4+4.6+6.7+5.1+4.3+4.5)}{5\cdot 6} = \frac{30.6}{6} = 5.1.$$

  Let $\bar{X}_{i\cdot}$ be the sample mean for the D_i group, then

$$\bar{X}_{1\cdot} = \frac{5(5.4+5.1)}{5\cdot 2} = 5.25,$$

$$\bar{X}_{2\cdot} = \frac{5(4.6+4.3)}{5\cdot 2} = 4.45$$

  and

$$\bar{X}_{3\cdot} = \frac{5(6.7+4.5)}{5\cdot 2} = 5.6.$$

Let Factor A be the drug factor, then

$$\begin{aligned}
\text{SSA} &= 2 \cdot 5 \left((\bar{X}_{1\cdot} - \bar{X}_G)^2 + (\bar{X}_{2\cdot} - \bar{X}_G)^2 + (\bar{X}_{3\cdot} - \bar{X}_G)^2\right) \\
&= 10\left((5.25 - 5.1)^2 + (4.45 - 5.1)^2 + (5.6 - 5.1)^2\right) \\
&= 6.95
\end{aligned}$$

and

$$\begin{aligned}
\text{SSE} &= (5-1)(0.04 + 0.01 + 0.01 + 0.01 + 0.09 + 0.04) \\
&= 0.8.
\end{aligned}$$

The F statistic for testing the drug effect is

$$\frac{\text{SSA}/(3-1)}{\text{SSE}/(3 \cdot 2 \cdot 5 - (3 \cdot 2))} = \frac{6.95/2}{0.8/24} = 104.25.$$

From the table "0.95 quantiles for F distributions", $f_{0.05,2,24} = 3.40 < 104.25$, so we conclude that different drugs have different effects on weight loss at the 0.05 level.

(b) Let $\bar{X}_{\cdot 1}$ and $\bar{X}_{\cdot 2}$ be the sample means for the jogging group and the walking group respectively, then

$$\bar{X}_{\cdot 1} = \frac{5(5.4 + 4.6 + 6.7)}{3 \cdot 5} = \frac{16.7}{3}$$

and

$$\bar{X}_{\cdot 2} = \frac{5(5.1 + 4.3 + 4.5)}{3 \cdot 5} = \frac{13.9}{3}.$$

Let Factor B be the exercise factor, then

$$\begin{aligned}
\text{SSB} &= 3 \cdot 5 \sum_{j=1}^{2} (\bar{X}_{\cdot j} - \bar{X}_G)^2 \\
&= 15\left(\left(\frac{16.7}{3} - 5.1\right)^2 + \left(\frac{13.9}{3} - 5.1\right)^2\right) \\
&= \frac{19.6}{3} \approx 6.533333.
\end{aligned}$$

The F statistic for testing the exercise effect is

$$\frac{\text{SSB}/(2-1)}{\text{SSE}/(3 \cdot 2 \cdot 5 - (3 \cdot 2))} = \frac{19.6/3}{0.8/24} = 196.$$

From the table "0.95 quantiles for F distributions", $f_{0.05,1,24} = 4.26 < 196$, so we conclude that different exercises have different effects on weight loss at the 0.05 level.

(c) Since

$$\begin{aligned}
&\mathrm{SS}_{\text{total}} - \mathrm{SSE} \\
&= 5\sum_{i=1}^{3}\sum_{j=1}^{2}(\bar{X}_{i,j} - \bar{X}_G)^2 \\
&= 5\left(\sum_{i=1}^{3}\sum_{j=1}^{2}\bar{X}_{i,j}^2\right) - 5\cdot 3\cdot 2(\bar{X}_G)^2 \\
&= 5\left((5.4)^2 + (4.6)^2 + (6.7)^2 + (5.1)^2 + (4.3)^2 + (4.5)^2\right) - 30(5.1)^2 \\
&= 19.5,
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{SSI} &= \mathrm{SS}_{\text{total}} - \mathrm{SSE} - \mathrm{SSA} - \mathrm{SSB} \\
&= 19.5 - 6.95 - \frac{19.6}{3} \\
&= \frac{18.05}{3} \approx 6.016667
\end{aligned}$$

The F statistic for testing the exercise effect is

$$\frac{\mathrm{SSI}/((3-1)(2-1))}{\mathrm{SSE}/(3\cdot 2\cdot 5 - (3\cdot 2))} = \frac{(18.05/3)/2}{0.8/24} = 90.25.$$

From the table "0.95 quantiles for F distributions", $f_{0.05,2,24} = 3.40 < 90.25$, so we conclude that the drug-exercise interaction effect is significant at the 0.05 level.

- Fact 1 Given $d_1 \in \{1, 2, \ldots, 40\}$, $f_{0.05,d_1,d_2}$ and $f_{0.01,d_1,d_2}$ are strictly decreasing functions of d_2 when $d_2 \in \{1, 2, \ldots, 120\}$.

- Example 3. For the data in Example 2, suppose that we carry out one-way ANOVA using drug type as the treatment factor.

  (a) Find the sum of squares due to drug in the one-way ANOVA.

  (b) Find the total sum of squares in the one-way ANOVA.

  (c) Determine whether we can conclude that not all drugs have the same effect on weight loss at level 0.05.

  Solution.

  (a) The sum of squares due to drug is the same as that in two-way ANOVA, which is 6.95.

  (b) The total sum of squares is the same as that in two-way ANOVA, which is $19.5 + 0.8 = 20.3$.

(c) The SSE in one-way ANOVA is $20.3 - 6.95 = 13.35$ with degrees of freedom $29 - 2 = 27$. The F statistic is $(6.95/2)/(13.35/27) = 7.02809$. From Fact 1, $3.32 = f_{0.05,2,30} < f_{0.05,2,27} < f_{0.05,2,25} = 3.39$, $7.02809 > f_{0.05,2,27}$ and we can conclude that not all drugs have the same effect on weight loss at level 0.05.

- The data in Example 2 are available at

  `https://stat.walkup.tw/teaching/statistics/data/anova_data.txt`

  To read the data into R, open R and run

  ```
  getwd()
  ```

  Then it shows a directory. Move the data file to the directory and run

  ```
  data <- read.table("anova_data.txt", header=TRUE)
  ```

  Then the data set is read into R. It is called `"data"` in R. Enter the command `"data"` in R and the content of the data set will appear:

  ```
  > data
     weight.loss drug exercise
  1          5.6    1        1
  2          5.6    1        1
  .
  .
  .
  30         4.3    3        2
  ```

  Here the first column (`weight.loss`) gives observations for weight loss, the second column (`drug`) indicates the type of drug used (i for the i-th type) and the third column (`exercise`) gives the type of exercise.

  We can check the sample mean and sample standard deviation for the group with drug level i and exercise level j for each (i, j).

  ```
  weight.loss <- data[,1]
  drug <-  data[,2]
  exercise <-  data[,3]
  j <- 1 #jogging group
  for (i in 1:3){
    data.ij <- weight.loss[(drug==i)&(exercise==j)]
    print(mean(data.ij))
    print(var(data.ij))
  }
  j <- 2 #walking group
  ```

```
for (i in 1:3){
  data.ij <- weight.loss[(drug==i)&(exercise==j)]
  print(mean(data.ij))
  print(var(data.ij))
}
```

The sample means and sample variances are the same as in the table in Example 2.

To use R to carry out the two-way ANOVA directly, we can run the following commands:

```
weight.loss <- data[,1]
drug <- as.factor(data[,2])
exercise <- as.factor(data[,3])
anova(lm(weight.loss ~ drug + exercise + drug*exercise))
```

The R output is:

```
Analysis of Variance Table

Response: weight.loss
              Df Sum Sq Mean Sq F value    Pr(>F)
drug           2 6.9500  3.4750  104.25 1.464e-12 ***
exercise       1 6.5333  6.5333  196.00 4.829e-13 ***
drug:exercise  2 6.0167  3.0083   90.25 6.827e-12 ***
Residuals     24 0.8000  0.0333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Example 4. Suppose that we carry out two-way ANOVA by running the R command

```
anova(lm(weight.loss ~ drug + exercise + drug*exercise))
```

where the R vectors `drug` and `exercise` consist of observations of two factor variables respectively and `weight.loss` consists of observations of a continuous random variable. The output ANOVA table is

```
Analysis of Variance Table

Response: weight.loss
              Df Sum Sq Mean Sq F value    Pr(>F)
drug           2 6.9500  3.4750  104.25 1.464e-12 ***
exercise       1 6.5333  6.5333  196.00 4.829e-13 ***
```

```
drug:exercise  2 6.0167  3.0083   90.25 6.827e-12 ***
Residuals     24 0.8000  0.0333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer the following questions.

(a) Is there a significant interaction effect between `drug` and `exercise` on `weight.loss` at 0.05 level?

(b) Can we conclude that the main effect of `drug` on `weight.loss` is significant at level 0.05?

(c) Can we conclude that the main effect of `exercise` on `weight.loss` is significant at level 0.05?

(d) How many levels does the factor `drug` have?

(e) How many levels does the factor `exercise` have?

(f) What is the sample size?

Solutions.

(a) Yes (p-value $= 6.827 \times 10^{-12} < 0.05$).

(b) Yes (p-value $= 1.464 \times 10^{-12} < 0.05$).

(c) Yes (p-value $= 4.829 \times 10^{-13} < 0.05$).

(d) The sum of squares due to `drug` has $2 = k - 1$ degrees of freedom, so the factor `drug` has $k = 3$ levels.

(e) The sum of squares due to `exercise` has $1 = b - 1$ degree of freedom, so the factor `exercise` has $b = 2$ levels.

(f) The sum of squares due to error (SSE) has $24 = n - kb$ degrees of freedom, where $k = 3$ and $b = 2$ from Parts (d) and (e). Therefore, the sample size is $n = 24 + 6 = 30$.

- Note. If we are given data for running two-way ANOVA with Factors A and B, then we can also carry out a one-way ANOVA using the same data. In the two cases, SSA values are the same, and SS_{total} values are the same.

- Example 5. In Example 4, suppose we also run the R command

```
anova(lm(weight.loss ~ drug))
```

to carry out one-way ANOVA for the same data. Answer the following questions based on the ANOVA table in Example 4.

(a) Find the sum of squares due to `drug` in one-way ANOVA.

(b) Find the total sum of squares in one-way ANOVA.

(c) Find the sum of squares due to error in one-way ANOVA.

Sol.

(a) The sum of squares due to `drug` in the two-way ANOVA table given in Example 4 is 6.95, so the sum of squares due to `drug` in one-way ANOVA is 6.95.

(b) The total sum of squares in two-way ANOVA is $6.95 + 6.5333 + 6.0167 + 0.8 = 20.3$, so the total sum of squares in one-way ANOVA is 20.3.

(c) The sum of squares due to error in one-way ANOVA is $20.3 - 6.95 = 13.35$.

- For the data in Example 4, the one-way ANOVA table obtained by running the R command

```
anova(lm(weight.loss ~ drug))
```

is

```
Analysis of Variance Table

Response: weight.loss
          Df Sum Sq Mean Sq F value  Pr(>F)
drug       2   6.95  3.4750  7.0281 0.00349 **
Residuals 27  13.35  0.4944
---
Signif. codes:  0  ‘***’  0.001  ‘**’  0.01  ‘*’  0.05  ‘.’  0.1  ‘ ’  1
```

Note that the sum of squares due to `drug` and the sum of squares due to error are the same as those computed in Example 5.

- Two-way ANOVA without interaction. When the interaction effect is not significant, one might want to carry out a two-way ANOVA without interaction. In such case, SSA and SSB are still defined according to (5) and (6) and

$$SS_{\text{total}} = \text{SSA} + \text{SSB} + \text{SSE}.$$

- Example 6. Suppose that the experiment in Example 2 is carried again for another group of participants with more types of drugs and exercises, and the results are stored in three variables in R: `weight.loss`, `drug`, `exercise`. The variable `weight.loss` stores the weight loss amounts for all participants, the variable `drug` stores the drug types, and the variable `exercise` stores the exercise types. Both `drug` and `exercise` are factors in R. Suppose we run

```
anova(lm(weight.loss ~ drug + exercise + drug*exercise))
```

and obtain the following R output:

```
Analysis of Variance Table

Response: weight.loss
               Df  Sum Sq Mean Sq F value   Pr(>F)
drug            2 15.0540  7.5270  7.9707 0.002215 **
exercise        1  7.3013  7.3013  7.7317 0.010386 *
drug:exercise   2  0.9687  0.4843  0.5129 0.605182
Residuals      24 22.6640  0.9443
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It appears that the drug-exercise interaction effect is not significant at level 0.05. Suppose that we run

```
anova(lm(weight.loss ~ drug + exercise))
```

to fit Model II: the two-way ANOVA model without interaction effect.

(a) Find the total sum of squares.

(b) Find the SSE in Model II.

(c) Consider the F test for testing H_0: all drugs have the same effect on weight loss in Model II. Find the F statistic for the test. What can be said about the p-value of the test?

Sol.

(a) The total sum of squares is $15.0540 + 7.3013 + 0.9687 + 22.6640 = 45.988$.

(b) The SSE is $15.0540+7.3013+0.9687+22.6640-(15.0540+7.3013) = 0.9687 + 22.6640 = 23.6327$.

(c) The F statistic is

$$\frac{15.0540/2}{23.6327/(2+24)} = 8.280984.$$

From the table of the 0.99 quantiles for F-distributions, we have $5.39 = f_{0.01,2,30} < f_{0.01,2,26} < f_{0.01,2,25} = 5.57$, so $8.280984 > f_{0.01,2,26}$. The p-value is

$$P(F(2,26) > 8.280984) < P(F(2,26) > f_{0.01,2,26}) = 0.01.$$

- For the data in Example 6, we can carry out two-way ANOVA model without interaction by running the R command

```
anova(lm(weight.loss ~ drug + exercise ))
```

and the R output is

```
Analysis of Variance Table

Response: weight.loss
          Df  Sum Sq Mean Sq F value   Pr(>F)
drug       2 15.0540  7.5270  8.2810 0.001650 **
exercise   1  7.3013  7.3013  8.0327 0.008767 **
Residuals 26 23.6327  0.9089
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note. The sum of squares due to error is the same as the SSE computed in Example 6, and the total sum of squares is $15.0540 + 7.3013 + 23.6327 = 45.988$, which is the same as the total sum of squares computed in Example 6.

- In a two-way ANOVA with interaction, if the interaction effect is not significant, then we can carry out a two-way ANOVA without interaction. If the main effect for one of the two factors is not significant, then it suggests that a one-way ANOVA is sufficient.