μ는 모집단의 전체평균

τ_i는 i번째 처리효과, $i=1,\ldots,p$

α_j는 j번째 행블록 효과, $j=1,\ldots,p$

β_h는 h번째 열위치 효과, $h=1,\ldots,c$

오차항 $\epsilon_{ijh} \sim iid\ N(0,\sigma^2)$

5.9 R을 이용한 라틴정방설계 분석

예제 5.3 4종류의 케이크(cake) a, b, c, d의 맛점수를 비교하고자 한다. 4명의 실험자가 반죽을 준비하고 한 개의 오븐에는 4개의 케이크를 구울 수 있다. 실험자와 오븐의 위치를 블록요인으로 고려하여 4×4 라틴정방설계로 실험을 하여 데이터를 얻었다. 통계적 분석으로 케이크 종류에 따라 맛이 다르다고 할 수 있는지 알아보고자 하며 블록효과에 대해서도 알아보고자 한다.

표 5.7 라틴정방 케이크 데이터

	실험자			
오븐	1	2	3	4
1	c 21	b 47	a 31	d 49
2	a 35	d 49	c 29	b 44
3	b 42	a 38	d 51	c 27
4	d 53	c 25	b 40	a 32

■

프로그램 5.1 케이크 데이터에 대한 라틴정방설계 분석

```
# Latin square design
a<-read.csv("d:\\13expd\\5.1cake.csv", header=TRUE)
a ; attach(a)
cake<-as.factor(a$cake)
mixer<-as.factor(a$mixer)
oven <-as.factor(a$oven)

by(y, cake, FUN=function(x){ c(mean(x), sd(x)) } )
by(y, mixer, FUN=function(x){ c(mean(x), sd(x)) } )
by(y, oven, FUN=function(x){ c(mean(x), sd(x)) } )
```

```
a1<-lm(y ~  oven + mixer + cake)     #순서: row+ col + trt
anova(a1)
pairwise.t.test(y, cake, p.adjust="none")

par(mfrow=c(1,3))
  plot(y~ oven + mixer + cake)    # 그림 5.1

par(mfrow=c(2,2))
  plot(a1)          # 그림 5.2 model diagnostic plots
detach(a)
```

```
> by(y, cake, FUN=function(x){ c(mean(x), sd(x)) } )
cake: a
[1] 34.000000  3.162278
------------------------------------------------------------
cake: b
[1] 43.250000  2.986079
------------------------------------------------------------
cake: c
[1] 25.50000  3.41565
------------------------------------------------------------
cake: d
[1] 50.500000  1.914854
```

케이크 종류(처리)별 평균과 표준편차를 보여 준다.

```
> by(y, oven, FUN=function(x){ c(mean(x), sd(x)) } )
oven: 1
[1] 37.00000 13.36663
------------------------------------------------------------
oven: 2
[1] 39.250000  8.958236
------------------------------------------------------------
oven: 3
[1] 39.500000  9.949874
------------------------------------------------------------
oven: 4
[1] 37.50000 12.01388
```

오븐(행)별 평균과 표준편차를 보여 준다.

```
> by(y, mixer, FUN=function(x){ c(mean(x), sd(x)) } )
mixer: 1
[1] 37.75000 13.40087
------------------------------------------------------------
mixer: 2
[1] 39.75000 10.93542
```

```
---------------------------------------------------------------
mixer: 3
[1] 37.00000 12.16553
---------------------------------------------------------------
mixer: 4
[1] 38.400000  8.905055
```

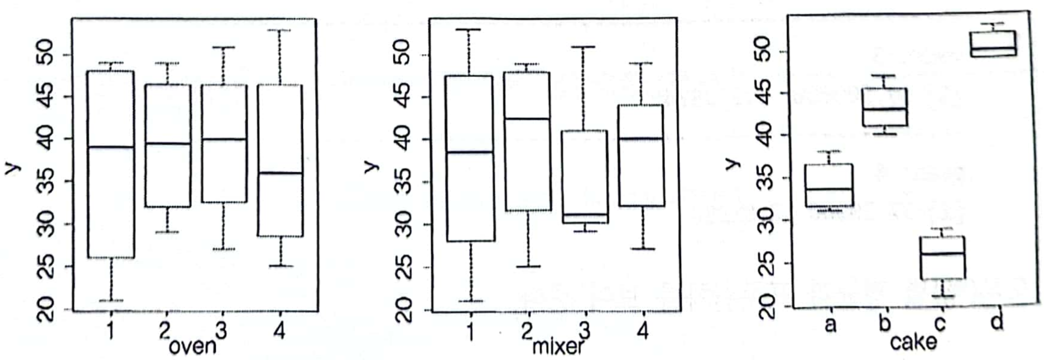
실험자(열)별 평균과 표준편차를 보여 준다.

```
> a1<-lm(y ~  oven + mixer + cake)    #순서: row+ col + trt
> anova(a1)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
oven       3   18.69    6.23  0.5227 0.682359
mixe       3   18.06    6.02  0.5053 0.692710
cake       3 1417.19  472.40 39.6415 0.000238 ***
Residuals 6   71.50   11.92
```

분산분석표 결과를 보여 주며 케이크(처리)효과에 대한 다음의 가설

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \text{에 대해 } H_1 : not\ H_0$$

에 대한 검정으로 검정통계량 $F_t = 39.6415$이고 p-값$= 0.000238 < 0.05 = \alpha$이므로 유의수준 0.05에서 H_0를 기각한다. 즉, 4개 케이크 맛 평균이 모두 동일하다고 할 수 없다. 〈그림 5.1〉의 상자그림을 보면 각 요인의 수준별 퍼짐성과 수준 간의 차이를 볼 수 있다. 모형에 대한 검토로 잔차에 대한 그림은 〈그림 5.2〉에서 보여 준다.



(a) 오븐별 상자그림 (b) 실험자별 상자그림 (c) 케이크별 상자그림

그림 5.1 케이크 데이터에 대한 상자그림

```
> pairwise.t.test(y, cake, p.adjust="none")
        Pairwise comparisons using t tests with pooled SD
data:  y and cake
  a       b       c
b 0.00076 -       -
c 0.00145 1.8e-06 -
d 3.9e-06 0.00435 4.5e-08
```

유의수준 0.05에서 LSD 다중비교 결과를 보면 a, b, c, d 모두 서로 다른 그룹으로 나뉜다.

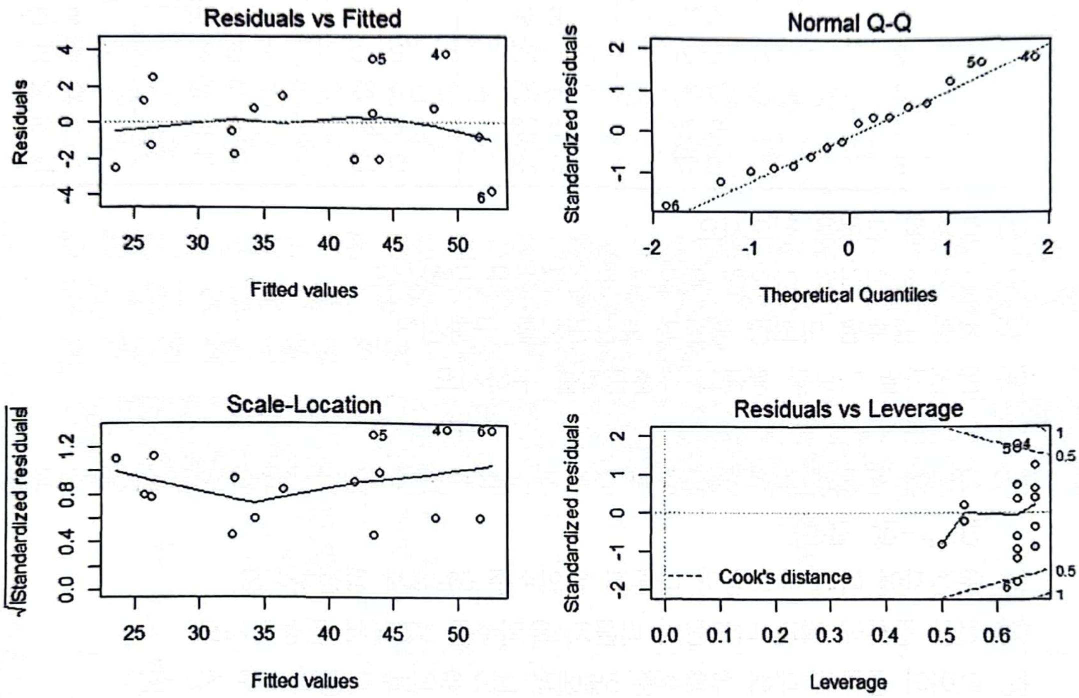


그림 5.2 케이크 데이터에 대한 모형진단그림: 위 왼쪽부터 (a) 잔차그림 (b) 잔차에 대한 정규 Q-Q 그림, 아래 왼쪽부터 (c) 적합값에 대한 제곱근 표준화 잔차그림 (d) 그룹별 표준화 잔차그림

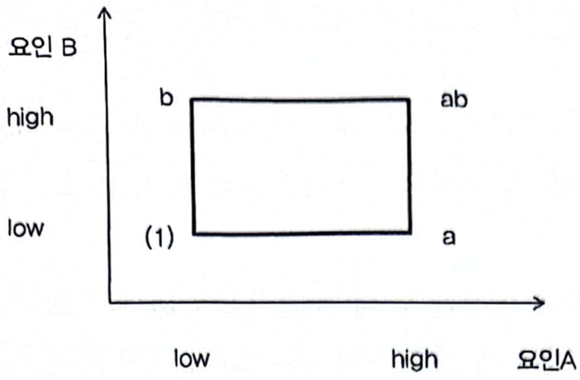


그림 7.1 이요인설계에서의 처리조합

요인이 2개 수준을 가지므로 효과에 대한 계산식이 간단하다. 요인의 주효과와 상호작용효과에 대한 계산식을 구해보자.

A의 주효과 = A의 1 수준평균 - A의 0 수준평균

$$A=\frac{a_1b_0+a_1b_1}{2}-\frac{a_0b_0+a_0b_1}{2}=\frac{1}{2}(a_1b_1+a_1b_0-a_0b_1-a_0b_0)$$

B의 주효과 = B의 1 수준평균 - B의 0 수준평균

$$B=\frac{a_1b_1+a_0b_1}{2}-\frac{a_1b_0+a_0b_0}{2}=\frac{1}{2}(a_1b_1-a_1b_0+a_0b_1-a_0b_0)$$

AB 상호작용효과 = B의 1 수준에서 A의 수준 간 평균차-B의 0 수준에서 A의 수준 간 평균차

$$AB=\frac{a_1b_1-a_0b_1}{2}-\frac{a_1b_0-a_0b_0}{2}=\frac{1}{2}(a_1b_1-a_1b_0-a_0b_1+a_0b_0)$$ ■

효과 계산식을 보면 A의 주효과를 구할 경우 A의 1 수준이 포함된 처리조합에 대한 계수의 부호는 "+", A의 0 수준이 포함된 처리조합에 대한 계수의 부호는 "−"로 대비(contrast)가 됨을 알 수 있다. 마찬가지로 B의 주효과를 구할 경우 B의 1 수준이 포함된 처리조합은 "+", B의 0 수준이 포함된 처리조합은 "−"로 대비가 된다. 여기서 a_ib_j는 a_ib_j 처리조합에서의 데이터를 나타낸다.

검정통계량 $F_A = 36$, p-값$= 0.105 > 0.05$로 유의수준 5%에서 H_0를 기각하지 못한다. 즉, 압력효과는 유의하지 않으므로 압력 수준 간 차이가 있다고 할 수 없다.

(2) 요인 B 온도효과에 대한 가설검정

$$H_0:\ \beta_0 = \beta_1 = 0 \text{에 대해 } H_1 : not\ H_0$$

검정통계량 $F_B = 25$이고 p-값$= 0.126 > 0.05$로 유의수준 5%에서 H_0를 기각하지 못한다. 즉, 온도효과는 유의하지 않으므로 온도 수준 간 차이가 있다고 할 수 없다.

프로그램 7.1 접착제 데이터 반복 없는 2^2 요인설계 분석

```
# example 7.2
press<-c(100,100, 150,150)
temp<-c(50, 100, 50,100)
y<-c(9,13, 14,20)
  press<-as.factor(press)
  temp<-as.factor(temp)
dd<-aov(y~ press + temp)
summary(dd)

par(mfrow=c(1,2))
  plot(y~ temp);   plot(y~press)    # 그림 7.2
interaction.plot(temp, press, y, bty="l", main="interaction plot")
interaction.plot(press, temp, y, bty="l", main="interaction plot")
# 그림 7.3
```

■

예제 7.3 접착제의 강도가 접착제를 사용할 때의 압력과 온도에 영향을 받는다고 생각되어 이요인실험을 하여 주어진 압력(A)과 온도(B)하에서 2개씩의 데이터를 다음과 같이 얻었다. 〈프로그램 7.2〉를 이용하여 결과를 얻는다.

표 7.7 반복 2^2 요인설계 접착제 데이터

압력[A]	온도(℃)[B] 50		100	
100	9	11	11	13
150	14	16	14	20

(1) 〈표 7.7〉의 데이터에 대한 2^2 요인설계에 대한 통계적 모형(고정효과모형)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}, \quad i = 0,1, \quad j = 0,1, \quad k = 1,2(\text{반복})$$

여기서

μ는 모집단의 전체평균

α_i는 요인 A의 i번째 수준 압력효과

β_j는 요인 B의 j번째 수준 온도효과

$\alpha\beta_{ij}$는 요인 A의 i번째 수준과 요인 B의 j번째 수준의 상호작용효과

ϵ_{ijk}는 오차항 $\sim iid\ N(0, \sigma^2)$

(2) 분산분석표와 가설검정

```
> summary(dd)
            Df Sum Sq Mean Sq F value Pr(>F)
press        1     50      50   8.333 0.0447 *
temp         1      8       8   1.333 0.3125
press:temp   1      0       0   0.000 1.0000
Residuals    4     24       6
```

(a) 요인 A 압력효과에 대한 가설검정

$$H_0 : \alpha_1 = \alpha_2 = 0 \text{ (요인 A 압력효과 없다)에 대해 } H_1 : \text{not } H_0$$

검정통계량 $F_A = 8.333$, p-값 $= 0.0447 < 0.05$로 유의수준 5%에서 H_0를 기각한다. 압력 수준 간 차이가 있다고 할 수 있다.

(b) 요인 B 온도효과에 대한 가설검정

$$H_0 : \beta_0 = \beta_1 = 0 \text{ (요인 B 온도효과 없다)에 대해 } H_1 : \text{not } H_0$$

검정통계량 $F_B = 1.333$, p-값 $= 0.3125 > 0.05$로 유의수준 5%에서 H_0를 기각하지 못한다. 즉, 온도효과는 유의하지 않으므로 압력 수준 간 차이가 있다고 할 수 없다.

(c) 요인 A와 요인 B 상호작용효과에 대한 가설검정

H_0: $\alpha\beta_{00} = \alpha\beta_{01} = \alpha\beta_{10} = \alpha\beta_{11} = 0$ (상호작용효과 없다)에 대해 H_1 : not H_0

검정통계량 $F_{AB} = 0.000$, p-값$= 1.0000 > 0.05$로 유의수준 5%에서 H_0를 기각하지 못한다. 즉, 온도와 압력의 상호작용효과는 있다고 할 수 없다.

(3) 요인효과에 대한 모수추정

```
> model.tables(dd)
Tables of effects
press
 100  150
-2.5  2.5
temp
 50 100
 -1   1
 press:temp
     temp
press 50         100
  100 -9.108e-17  9.108e-17
  150  9.108e-17 -9.108e-17
```

프로그램 7.2 접착제 데이터 2 반복 2^2 요인설계 분석

```
# example 7.3: 2 반복 이요인설계
press<-c(100,100, 150,150, 100,100, 150,150)
temp<-c( 50, 50,   50, 50, 100,100, 100,100)
   y<-c( 9,11,     14, 16,  11, 13,  14, 20)
press<-as.factor(press)
temp<-as.factor(temp)
dd<-aov(y~ press*temp)
summary(dd) ;
model.tables(dd)
```

■

```
> L<-aov(y ~ rep+rep:block+ A*B*C )
> anova(L)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value  Pr(>F)
rep        1   36.00   36.00  0.1689 0.69812
A          1    6.25    6.25  0.0293 0.87075
B          1  441.00  441.00  2.0690 0.20985
C          1 1681.00 1681.00  7.8865 0.03762 *
rep:block  2    6.25    3.13  0.0147 0.98549
A:B        1   21.12   21.12  0.0991 0.76562
A:C        1  676.00  676.00  3.1715 0.13504
B:C        1    0.25    0.25  0.0012 0.97401
A:B:C      1  990.12  990.12  4.6452 0.08369 .
Residuals  5 1065.75  213.15
```

분산분석 결과를 보면 유의수준 $\alpha = 0.05$ 에서 C 효과만 유의하다. ■

예제 7.9 (교락법) A, B, C 세 개 요인설계를 2개 블록에 나누어서 실험을 했다.

반복 1				반복 2				반복 3				반복 4			
블록1		블록2		블록3		블록4		블록5		블록6		블록7		블록8	
(1)	10	a	17	(1)	11	a	8	(1)	6	b	9	a	17	(1)	9
ab	17	b	12	b	9	ab	9	a	15	ab	14	b	13	ab	15
c	9	ac	19	ac	16	c	6	bc	8	c	7	c	9	ac	17
abc	10	bc	11	abc	16	bc	2	abc	1	ac	14	abc	16	bc	14

(1) 반복별로 블록과 교락된 효과를 찾아보자.

같은 블록에서 동일한 부호를 갖는 효과가 블록과 교락되어 있으므로 계수표 〈표 7.14〉를 참조하여 교락된 효과를 찾을 수 있다.(〈표 7.8〉과 동일한 계수표이다)

표 7.14 2^3 **요인설계에서 제곱합을 구하기 위한 계수표**

	처리조합							
효과	(1) $A_0B_0C_0$	c $A_0B_0C_1$	b $A_0B_1C_0$	bc $A_0B_1C_1$	a $A_1B_0C_0$	ac $A_1B_0C_1$	ab $A_1B_1C_0$	abc $A_1B_1C_1$
A	−	−	−	−	+	+	+	+
B	−	−	+	+	−	−	+	+
C	−	+	−	+	−	+	−	+
AB	+	+	−	−	−	−	+	+
AC	+	−	+	−	−	+	−	+
BC	+	−	−	+	+	−	−	+
ABC	−	+	+	−	+	−	−	+

각 반복에서 블록과 교락된 효과는 다음과 같이 찾을 수 있다.

Rep 1: AB

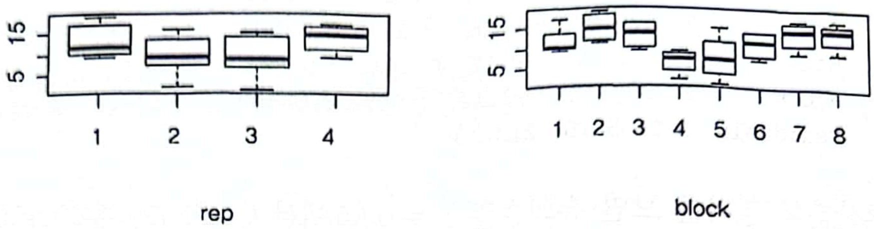
Rep 2: AC

Rep 3: BC

Rep 4: ABC

(2) 상자그림

(a)



(b)

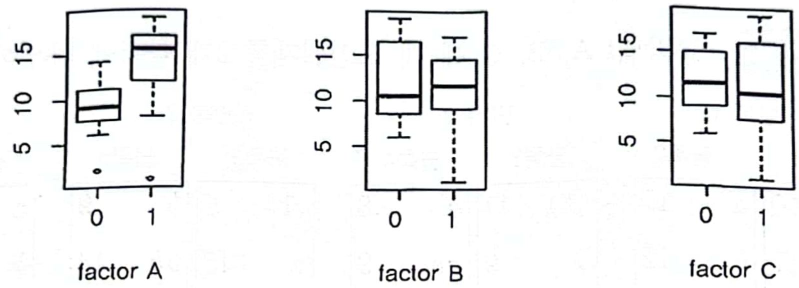


그림 7.8 (a) 삼요인 교락법 반복과 블록에 대한 상자그림
(b) 삼요인 교락법 요인 수준별 상자그림

(3) 분산분석

프로그램 7.4 반복 교락법 분석

```
da<- read.csv("D:\\13expd\\예제7.9.csv", header=T)   # 예제 7.9 데이터
da
attach(da)
rep<-as.factor(Rep)                        # 반복
  block<-as.factor(block)                  # 블록
  A<-as.factor(A)
  B<-as.factor(B)
  C<-as.factor(C)

par(mfrow=c(1,2))                          # 그림 7.8(a)
 boxplot(y~rep, sub="rep") ; boxplot(y~block, sub="block")

op=par(mfrow=c(1,3))                       # 그림 7.8(b)
  boxplot(y~A, sub="factor A") ; boxplot(y~B, sub="factor B")
```

```
 boxplot(y~C, sub="factor C")
par(op)

L<-aov(y ~ rep+ block +A+B+C )
anova(L)
model.tables(L, type="means")      # 모든 요인 수준평균
model.tables(L, type="effects")    # 모든 요인 수준효과

L1<-aov(y ~ rep +A+B+C+ A:B + A:C )
anova(L1)
model.tables(L1, type="means")     # 모든 요인 수준평균
model.tables(L1, type="effects")   # 모든 요인 수준효과
```

```
> da
   Rep block A B C  y
1    1     1 0 0 0 10
2    1     1 1 1 0 17
3    1     1 0 0 1  9
4    1     1 1 1 1 10
5    1     2 1 0 0 17
6    1     2 0 1 0 12
7    1     2 1 0 1 19
8    1     2 0 1 1 11
9    2     3 0 0 0 11
10   2     3 0 1 0  9
11   2     3 1 0 1 16
12   2     3 1 1 1 16
13   2     4 1 0 0  8
14   2     4 1 1 0  9
15   2     4 0 0 1  6
16   2     4 0 1 1  2
17   3     5 0 0 0  6
18   3     5 1 0 0 15
19   3     5 0 1 1  8
20   3     5 1 1 1  1
21   3     6 0 1 0  9
22   3     6 1 1 0 14
23   3     6 0 0 1  7
24   3     6 1 0 1 14
25   4     7 1 0 0 17
26   4     7 0 1 0 13
27   4     7 0 0 1  9
28   4     7 1 1 1 16
29   4     8 0 0 0  9
30   4     8 1 1 0 15
31   4     8 1 0 1 17
32   4     8 0 1 1 14
```

```
> L<-aov(y ~ rep+ block +A+B+C )
> anova(L)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
rep        3  130.125  43.375  5.3463 0.0067874 **
block      4  136.750  34.187  4.2139 0.0116663 *
A          1  180.500 180.500 22.2480 0.0001173 ***
B          1    6.125   6.125  0.7550 0.3947350
C          1    8.000   8.000  0.9861 0.3320121
Residuals 21 170.375   8.113
```

반복, 블록, 요인의 주효과만 포함한 모형에 대한 분산분석표를 보여 준다. 요인 A 효과가 유의하다. 반복과 블록효과 또한 유의하다.

```
> model.tables(L, type="means")        # 모든 요인 수준평균
Tables of means
Grand mean
11.4375
rep
     1      2      3      4
13.125  9.625  9.250 13.750
block
     1      2      3      4      5      6      7      8
 9.813 13.063 14.813  8.063  9.688 13.188 11.438 11.438
A
     0      1
 9.063 13.813
B
     0      1
11.875 11.000
C
     0      1
11.938 10.938

> model.tables(L, type="effects")      # 모든 요인 수준효과
Tables of effects
rep
      1       2       3       4
 1.6875 -1.8125 -2.1875  2.3125
block
     1      2      3      4      5      6      7      8
-1.625  1.625  3.375 -3.375 -1.750  1.750  0.000  0.000
A
     0      1
-2.375  2.375
B
     0      1
 0.4375 -0.4375
```

```
C
    0    1
 0.5 -0.5
```

결과를 보면, 요인 A의 수준효과는 $\alpha_0 = -2.375$, $\alpha_1 = 2.375$이다. 요인 B의 수준효과는 $\beta_0 = -0.4375$, $\beta_1 = 0.4375$이다. 요인 C의 수준효과는 각각 $\gamma_0 = 0.5$, $\gamma_1 = -0.5$이다. ■

주효과와 관심 있는 2차 상호작용효과를 포함한 모형에 대한 분산분석표이다. 유의수준 $\alpha = 0.05$에서 요인효과는 모두 유의하지 않다.

```
> model.tables(faov2)  # model parameters
Tables of effects
A
    0     1
-2.75  2.75
B
    0     1
-2.75  2.75
C
    0     1
-0.25  0.25
A:B
   B
A   0     1
  0 -3.75  3.75
  1  3.75 -3.75
A:C
   C
A   0     1
  0  1.75 -1.75
  1 -1.75  1.75
B:C
   C
B   0     1
  0 -2.25  2.25
  1  2.25 -2.25
```

faov2 모형에 대한 모수추정값을 보여 준다. ■

예제 8.5 [그림 8.3]의 데이터에 대해 R을 이용하여 요인효과에 대한 통계적 분석을 하고자 한다. 〈프로그램 8.2〉를 활용한다.

표 8.11 엑셀 데이터셋 그림 8.3 데이터.csv

A	B	C	D	y
0	0	0	0	45
0	1	1	0	80
1	1	0	0	65
1	0	1	0	60
1	0	0	1	75
0	1	0	1	45
1	0	0	1	100
1	1	1	1	96

프로그램 8.2 부분 사요인설계 분석

```
frac<-read.csv("D:\\13expd\\그림8.3데이터.csv", header=T)
  frac ;  attach(frac)
A<-as.factor(A)
B<-as.factor(B)
C<-as.factor(C)
D<-as.factor(D)

op=par(mfrow=c(1,4))     # 그림 8.5
  boxplot(y~A, sub="A") ;  boxplot(y~B, sub="B")
  boxplot(y~C, sub="C") ;  boxplot(y~D, sub="D")
par(op)

tapply(y, A, mean)  ;  tapply(y, B, mean)
tapply(y, C, mean)  ;  tapply(y, D, mean)

faov1<-aov(y~A+B+C+D)
  summary(faov1)
model.tables(faov1)  # model parameters
```

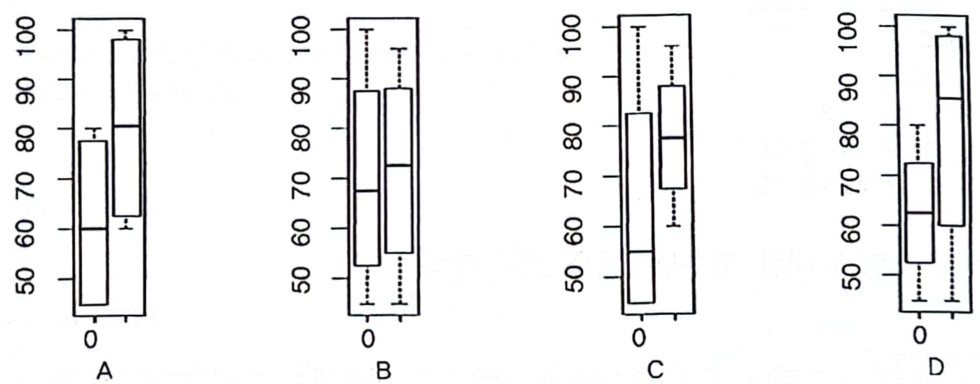


그림 8.5 부분 사요인설계 요인별 상자그림

```
> tapply(y, A, mean)
    0     1
61.25 80.25
> tapply(y, B, mean)
   0    1
70.0 71.5
> tapply(y, C, mean)
    0     1
63.75 77.75
> tapply(y, D, mean)
   0    1
62.5 79.0
```

각 요인수준별 평균을 보여 준다.

```
> faov1<-aov(y~A+B+C+D)
>   summary(faov1)
            Df Sum Sq Mean Sq F value Pr(>F)
A            1  722.0   722.0   1.538  0.303
B            1    4.5     4.5   0.010  0.928
C            1  392.0   392.0   0.835  0.428
D            1  544.5   544.5   1.160  0.360
Residuals    3 1408.5   469.5
```

주효과만 포함한 모형에 대한 분산분석표이다. 유의수준 $\alpha = 0.05$에서 요인효과는 모두 유의하지 않다.

```
> model.tables(faov1)  # model parameters
Tables of effects
A
   0    1
-9.5  9.5
B
    0     1
-0.75  0.75
C
 0  1
-7  7
D
    0     1
-8.25  8.25
```

모형에서 효과에 대한 추정값을 보여 준다. ■