



제 5 강. 공분산분석

이재원
고려대학교 통계학과



주요내용



1. 공분산분석의 개념
2. 공변량이 하나인 경우에 대한 공분산 모형에서의 추정과 검정
3. 공변량이 둘 이상인 경우에 대한 공분산 모형에서의 추정과 검정
4. R 프로그램을 이용한 분석

공분산분석

(Analysis of Covariance : ANCOVA)

- 2개 이상의 처리에서 반응변수의 평균을 비교할 경우 반응변수에 영향을 주는 연속형 독립변수인 공변량(Covariate)을 함께 고려한 분석방법
- ANCOVA : ANOVA with covariate
- (예) 혈압강하제연구

공분산분석

(Analysis of Covariance : ANCOVA)

- 회귀분석과 분산분석의 결합
 - 각 처리안에서 공변량을 설명변수로 하여 회귀분석 시행
 - 추정의 정도 (precision)를 높임
- { 공변량이 하나인 경우
공변량이 두개인 경우



공변량이 하나인 경우에 대한 공분산분석

예제 5.1

◆ <표 5.3>은 당뇨병에 걸린 환자 20명에 대해 혈당을 낮추는 서로 다른 다섯 가지 치료법의 효능을 비교하고자 환자 20명을 랜덤하게 5그룹으로 나누어 각각의 치료법을 적용하여 한 달 후의 혈당량 수치를 측정한 자료이다. 그러나 한 달 후의 혈당량 수치가 초기 혈당량 수치에 영향을 받을 것으로 생각하여 초기 혈당량 수치도 함께 측정하였다.(p.183)

예제 5.1 (표5.3)

관측번호	치료법(trt)	x (초기 수치)	y (한 달 후 수치)
1	A	27.2	32.6
2	A	22.0	36.6
3	A	33.0	37.7
4	A	26.8	31.0
5	B	28.6	33.8
6	B	26.8	31.7
7	B	26.5	30.7
8	B	26.8	30.4
9	C	28.6	35.2
10	C	22.4	29.1
⋮	⋮	⋮	⋮

예제 5.1 설명

- 독립변수 : 치료법 (trt)
- 공변량 : x (초기 혈당 수치)
- 반응변수 : y (한 달 후 혈당 수치)
- 반응변수의 모평균에 영향을 끼칠 수 있는 또 다른 변수가 존재 \rightarrow 공변량의 영향 고려
- 초기 혈당량 수치를 보정한 상태에서 한달 후 혈당량 수치의 보정된 모평균에 차이가 있는지를 보아야 함

공변량이 하나인 경우

Trt 1		Trt 2		...	Trt I	
x_{11}	y_{11}	x_{21}	y_{21}	...	x_{I1}	y_{I1}
x_{12}	y_{12}	x_{22}	y_{22}	...	x_{I2}	y_{I2}
x_{13}	y_{13}	x_{23}	y_{23}	...	x_{I3}	y_{I3}
...
x_{1n_1}	y_{1n_1}	x_{2n_2}	y_{2n_2}	...	x_{In_I}	y_{In_I}

공변량이 하나인 경우

◆ 공분산분석을 하기 위해 필요한 두 가지 가정 ◆

1. 각 처리 안에서 반응변수에 미치는
공변량의 효과가 모두 동일해야 한다.

2. 공변량 효과가 0이 아니다.

공분산 모형

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \beta (x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad (i=1, \dots, I, \quad j=1, \dots, n_i) \\ &= \beta_0 + \alpha_i + \beta x_{ij} + \varepsilon_{ij} \\ &= \beta_{0i} + \beta x_{ij} + \varepsilon_{ij} \end{aligned} \quad (\text{식 5.1})$$
$$(\beta_0 = \mu - \beta \bar{x}_{..}, \quad \beta_{0i} = \beta_0 + \alpha_i, \quad \varepsilon_{ij} \sim iid N(0, \sigma^2))$$

공분산 모형 설명

y_{ij} : i 번째 처리에서 j 번째 개체의 반응값

x_{ij} : i 번째 처리에서 j 번째 개체의 공변량 값

μ : 반응 변수의 전체 평균

α_i : 처리의 효과

β : 모든 처리에 공통으로 작용하는 공변량의 효과

ε_{ij} : 등분산을 갖는 정규분포를 따른다고 가정

각 모수의 추정

- 모수 : μ, β, α_i
- 모수들의 최소제곱추정치(LSE)

$$\begin{aligned}\hat{\mu} &= \bar{y}_{..} \\ \hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..}) \\ \hat{\beta} &= \frac{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2}\end{aligned}\quad (\text{식 5.3})$$

따라서 β_{0i} 는 다음 식과 같이 추정됨

$$\hat{\beta}_{0i} = \hat{\beta}_0 + \hat{\alpha}_i = \bar{y}_{i.} - \hat{\beta} \bar{x}_{i.} \quad (\text{식 5.4})$$

각 모수의 검정

◆ 가설)

- $H_{01} : \beta = 0$

- 귀무가설은 처리효과를 제어한 상태에서 반응변수에 미치는 공변량 효과가 없다는 가정을 검정
- 귀무가설이 기각되지 않으면 분산분석 시행

- $H_{02} : \alpha_1 = \alpha_2 = \dots = \alpha_I$

- 공변량 효과를 제어한 상태에서 처리 간 반응변수의 차이가 있는지를 검정

각 모수의 검정

◆ 가설 - $H_{01}: \beta = 0$

$H_{02}: \alpha_1 = \alpha_2 = \cdots = \alpha_I$

Source	df	SS	MS	F
X	1	SSX	$MSX = SSX / 1$	F_1
Trt	$I - 1$	SS _{trt}	$MS_{trt} = SS_{trt} / (I - 1)$	F_2
Error	$N - I - 1$	SSE	$MSE = SSE / (N - I - 1)$	
Total	$N - 1$	SST		

$$F_1 = \frac{SSX / 1}{SSE / (N - I - 1)} \sim F(1, N - I - 1) \quad (\text{식 5.6})$$

$$F_2 = \frac{SS_{trt} / (I - 1)}{SSE / (N - I - 1)} \sim F(I - 1, N - I - 1) \quad (\text{식 5.7})$$



모형제곱합



◆ 제 1 종 제곱합(Type I SS)

- $\text{Model SS} = \text{SS}(\text{trt}, x)$
 $= \underline{\text{SS}(\text{trt}) + \text{SS}(x|\text{trt})}$

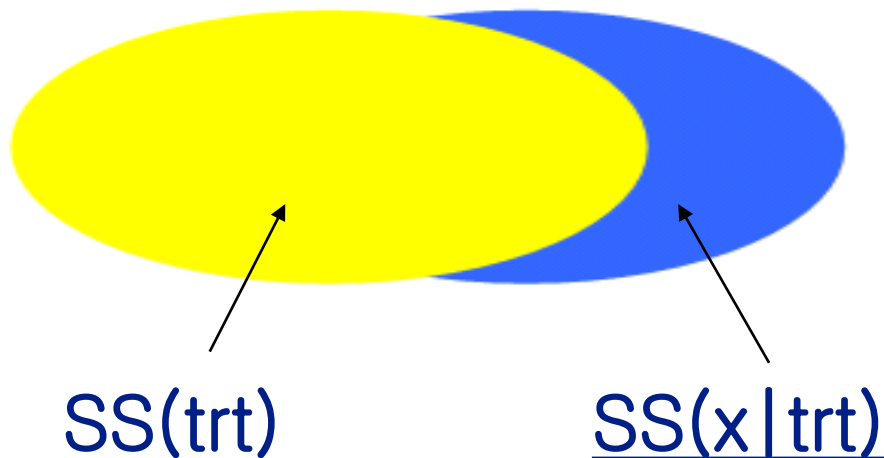
◆ 제 3 종 제곱합(Type III SS)

- $\underline{\text{SS}(\text{trt}|x) + \text{SS}(x|\text{trt})}$

모형제곱합

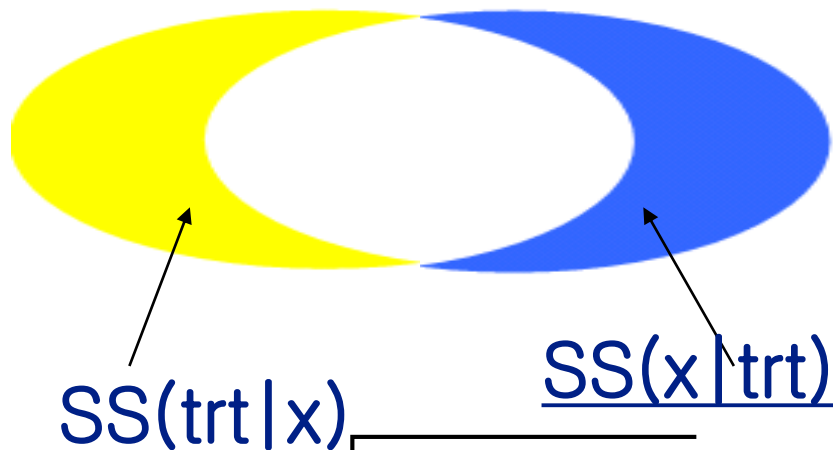
◆ TYPE I SS

$$SS(\text{trt}, x) =$$



◆ TYPE III SS

$$SS(\text{trt}|x) + \underline{SS(x|\text{trt})}$$

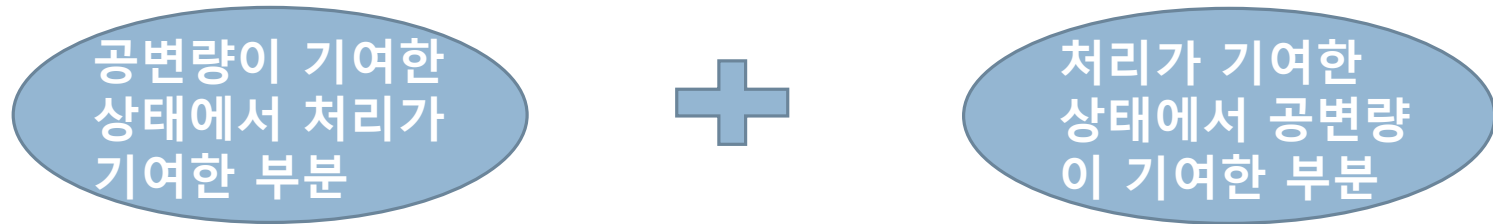


모형제곱합

◆ 제 1 종 제곱합(Type I SS)



◆ 제 3 종 제곱합(Type III SS)



초기혈당수치가 고려된 후 치료법간의 차이를 확인한다는
분석의 목적에 따라 **TYPE3 SS**사용

R program

```
library(HH); library(lsmmeans); library(car);  
혈당량자료=read.table("c:\\WORK\\혈당량자료.txt",header=T)  
혈당량자료  
attach(혈당량자료)  
options(contrasts=c("contr.sum", "contr.poly"))  
  
# 회귀계수의 동일성 검정(교호작용 존재 확인)  
모형1=lm(치료후혈당량 ~ 치료법 * 초기혈당량, data=혈당량자료)  
Anova(모형1, type="3")
```

- **lm 함수** : 공분산분석에는 lm 함수를 사용. 공분산분석의 가정 중 하나인 처리 간 회귀계수의 동일성을 확인하기 위해 처리와 공변량 사이의 교호작용의 유무를 검정. ~ 우변에 치료법*공변량과 같이 입력하면, 처리와 공변량의 교호작용 유무 검정.
- **Anova 함수 (Car 라이브러리)** : SS type에 대한 옵션을 지정해서 결과를 출력

R program (계속)

```
# 일원공분산분석(One-way ANCOVA)
```

```
모형2=lm(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료)
```

```
summary(모형2)
```

```
# xy plot 그리기
```

```
ancovaplot(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료)
```

- **lm 함수** : ~를 중심으로 좌변에는 반응변수, 우변에는 치료법과 공변량을 입력.
- **ancovaplot 함수(HH 라이브러리)** : 식의 구조는 lm 함수와 동일. xy plot을 그려줌.

R program (계속)

```
# 공변량 효과 제어시 치료법의 효과 검정
```

```
anova(모형2)
```

```
Anova(모형2, type="3")
```

```
# Type 1 SS와 Type 3 SS
```

```
summary(aov(치료후혈당량 ~ 초기혈당량 + 치료법), data=혈당량자료))
```

```
summary(aov(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료))
```

- **anova 함수** : model fitting function(ex. lm)에 사용함. lm 함수에서 ~ 우변에 입력하는 변수의 순서에 따라 출력 결과에서 그 변수의 SS type이 다름
- **Anova 함수(type="3" 옵션 지정)** : SS type에 대한 옵션을 지정해서 결과를 출력
- **aov 함수** : ~ 우변에 입력하는 변수의 순서에 따라 출력 결과에서 그 변수의 SS type이 다름

R program (계속)

```
# LSMEANS(Adjusted means) 계산
```

```
lsmeans(모형2, ~치료법)
```

```
# Tukey(HSD) 검정(다중비교)
```

```
모형2.lsm=lsmeans(모형2, pairwise ~ 치료법, glhargs=list())print(model2.lsm, omit = 1)
```

```
print(모형2.lsm, omit=1)
```

```
plot(모형2.lsm[[2]])
```

```
# 모형적합성 검토
```

```
par(mfrow = c(2,2))
```

```
plot(모형2)
```

- **lsmeans 함수 (lsmeans 라이브러리)** : ~우변에 처리를 입력하고, 좌변에는 변수를 입력하지 않으면 처리의 보정된 평균(adjusted mean)을 계산. ~ 우변에 처리를 입력하고, 좌변에는 pairwise라는 명령어를 입력하면 Tukey(HSD) 검정을 시행.

교호작용 효과 검정

공변량 효과가 같아야 한다는 가정



공변량과 처리 사이에 교호작용
(interaction)이 없다



교호작용 검정을 위해 lm 함수에서
~ 우변에 치료법*초기혈당량 입력

회귀계수의 동일성 검정

```
> # 회귀계수의 동일성 검정 (교호작용 존재 확인)
> 모형1=lm(치료후혈당량 ~ 치료법 * 초기혈당량, data=혈당량자료)
> Anova(모형1, type="3")
Anova Table (Type III tests)
```

Response: 치료후혈당량

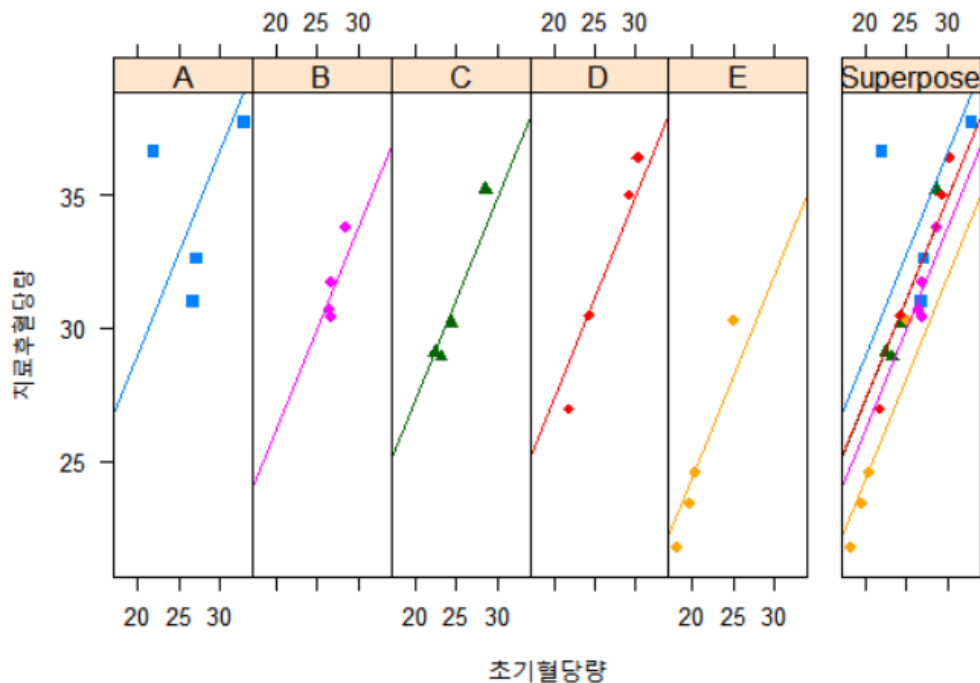
	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2.663	1	0.8521	0.377700
치료법	44.456	4	3.5556	0.047202 *
초기혈당량	51.981	1	16.6295	0.002222 **
치료법:초기혈당량	36.476	4	2.9173	0.077290 .
Residuals	31.258	10		

“H0 : 치료법과 초기혈당량의
교호작용이 없다”라는 귀무가설에
대해 제 3종 제곱합 (Type 3 SS) 검정

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

회귀계수의 동일성 확인 xyplot (그림 5.1)

```
> # xy plot 그리기  
> ancovaplot(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료)
```



처리효과 검정

```
> # 일원 공분산분석 (One-way ANOVA)
> 모형2=lm(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료)
> summary(모형2)
```

Call:

```
lm(formula = 치료후혈당량 ~ 치료법 + 초기혈당량, data = 혈당량자료)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.1360	-1.0024	-0.2827	0.7257	6.0806

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.8130	4.3797	2.697	0.017350 *
치료법1	2.1306	1.0417	2.045	0.060093 .
치료법2	-0.6378	1.0375	-0.615	0.548570
치료법3	0.4646	0.9893	0.470	0.645847
치료법4	0.5022	1.0039	0.500	0.624662
초기혈당량	0.7534	0.1723	4.373	0.000637 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.2 on 14 degrees of freedom

Multiple R-squared: 0.8112, Adjusted R-squared: 0.7437

F-statistic: 12.03 on 5 and 14 DF, p-value: 0.0001164

‘H0 : 처리의 효과와 공변량의 효과가 없다’ 라는 귀무가설에 대한 검정결과

처리효과검정

```
> # 공변량 효과 제어시 치료법의 효과 검정
```

```
> anova(모형2)
```

Analysis of Variance Table

Response: 치료후혈당량

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
치료법	4	198.407	49.602	10.252	0.0004301 ***
초기혈당량	1	92.528	92.528	19.125	0.0006369 ***
Residuals	14	67.734	4.838		

제 1종 제곱합(Type I SS)
: SS(치료법)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> Anova(모형2, type="3")
```

Anova Table (Type III tests)

Response: 치료후혈당량

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	35.198	1	7.2751	0.0173502 *
치료법	34.188	4	1.7666	0.1916720
초기혈당량	92.528	1	19.1248	0.0006369 ***
Residuals	67.734	14		

제 3종 제곱합(Type III SS)
: SS(치료법 | 초기혈당량)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

처리효과검정

```
> summary(aov(치료후혈당량 ~ 초기혈당량 + 치료법, data=혈당량자료))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
초기혈당량	1	256.75	256.75	53.067	4e-06 ***
치료법	4	34.19	8.55	1.767	0.192
Residuals	14	67.73	4.84		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

제 3종 제곱합(Type III SS)
: SS(치료법 | 초기혈당량)

```
> summary(aov(치료후혈당량 ~ 치료법 + 초기혈당량, data=혈당량자료))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
치료법	4	198.41	49.60	10.25	0.000430 ***
초기혈당량	1	92.53	92.53	19.12	0.000637 ***
Residuals	14	67.73	4.84		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

제 1종 제곱합(Type I SS)
: SS(치료법)

공분산 모형식 (식 5.11)

$$\hat{y}_{1j} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta} x_{1j} = 11.8130 + 2.1306 + 0.7534x_{1j}$$

$$\hat{y}_{2j} = \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta} x_{2j} = 11.8130 - 0.6378 + 0.7534x_{2j}$$

$$\hat{y}_{3j} = \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta} x_{3j} = 11.8130 + 0.4646 + 0.7534x_{3j}$$

$$\hat{y}_{4j} = \hat{\beta}_0 + \hat{\alpha}_4 + \hat{\beta} x_{4j} = 11.8130 + 0.5022 + 0.7534x_{4j} \quad (5.11)$$

$$\hat{y}_{5j} = \hat{\beta}_0 + \hat{\alpha}_5 + \hat{\beta} x_{5j} = 11.8130 - 2.4596 + 0.7534x_{5j}$$

○○○ 보정된 평균 (Adjusted mean) ○○○

- 각 처리마다 공변량 효과에 대해서 보정(adjust)한 상태에서의 반응변수의 평균

```
> # LSMEANS(Adjusted means) 계산  
> lsmeans(모형2, ~치료법)
```

치료법	lsmean	SE	df	lower.CL	upper.CL
A	33.0	1.15	14	30.5	35.4
B	30.2	1.15	14	27.7	32.7
C	31.3	1.10	14	28.9	33.7
D	31.3	1.12	14	28.9	33.7
E	28.4	1.34	14	25.5	31.3

Confidence level used: 0.95

처리별 보정된 평균 (식 5.12)

$$\bar{y}_{ad1} = \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta} \bar{x}_{..} = 11.8130 + 2.1306 + 0.7534 \times 25.26 = 32.975$$

$$\bar{y}_{ad2} = \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta} \bar{x}_{..} = 11.8130 - 0.6378 + 0.7534 \times 25.26 = 30.207 \quad (5.12)$$

$$\bar{y}_{ad3} = \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta} \bar{x}_{..} = 11.8130 + 0.4646 + 0.7534 \times 25.26 = 31.309$$

$$\bar{y}_{ad4} = \hat{\beta}_0 + \hat{\alpha}_4 + \hat{\beta} \bar{x}_{..} = 11.8130 + 0.5022 + 0.7534 \times 25.26 = 31.347$$

$$\bar{y}_{ad5} = \hat{\beta}_0 + \hat{\alpha}_5 + \hat{\beta} \bar{x}_{..} = 11.8130 - 2.4596 + 0.7534 \times 25.26 = 28.385$$

처리 간 다중비교

```
> # Tukey(HSD) 검정 (다중비교)
> 모형2.lsm=lsmmeans(모형2, pairwise ~ 치료법, glhargs=list())
> print(모형2.lsm, omit=1)
```

```
$lsmmeans
```

치료법	lsmean	SE	df	lower.CL	upper.CL
A	33.0	1.15	14	30.5	35.4
B	30.2	1.15	14	27.7	32.7
C	31.3	1.10	14	28.9	33.7
D	31.3	1.12	14	28.9	33.7
E	28.4	1.34	14	25.5	31.3

Confidence level used: 0.95

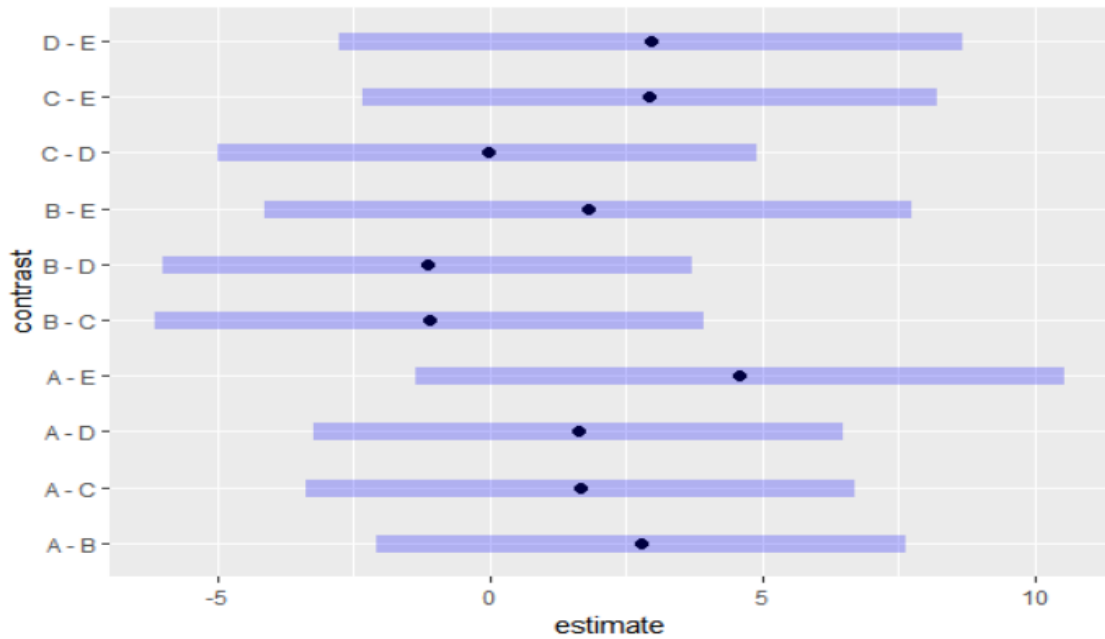
```
$contrasts
```

contrast	estimate	SE	df	t.ratio	p.value
A - B	2.7685	1.56	14	1.780	0.4216
A - C	1.6660	1.62	14	1.029	0.8378
A - D	1.6284	1.56	14	1.043	0.8316
A - E	4.5903	1.91	14	2.401	0.1718
B - C	-1.1024	1.62	14	-0.683	0.9570
B - D	-1.1401	1.56	14	-0.730	0.9457
B - E	1.8218	1.90	14	0.957	0.8696
C - D	-0.0376	1.59	14	-0.024	1.0000
C - E	2.9242	1.69	14	1.729	0.4485
D - E	2.9619	1.83	14	1.616	0.5115

P value adjustment: tukey method for comparing a family of 5 estimates

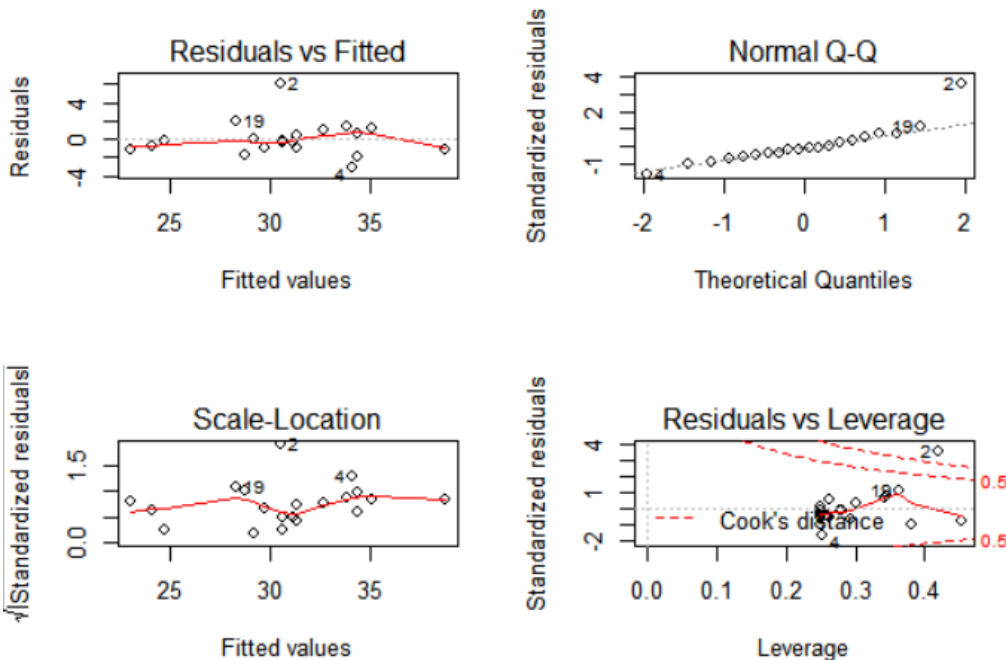
○○○ Tukey 신뢰구간 (그림5.2) ○○○

```
> plot(모형2.lsm[[2]])
```



모형적합성 검토 (그림 5.3)

```
> # 모형적합성 검토  
> par(mfrow=c(2,2))  
> plot(모형2)
```





공변량이 둘 이상인 경우에 대한 공분산분석

예제 5.2

- ▶ 〈표 5.5〉는 어떤 약물에 대한 체내 배출연구에서 얻은 자료이다. 연구자는 약의 형태에 따라 체내로부터 배출되는 약물의 양이 달라지는지를 알고자 한다. 그런데 배출되는 약물의 양은 약의 형태뿐만 아니라 배출된 약물을 측정할 시간과 각 개체의 항진대사 점수에도 영향을 받을 것으로 생각한다. 이러한 경우에는 측정시간과 항진대사 점수를 2개의 공변량으로 하여 이들을 제어한 약의 형태에 대한 효과를 공분산분석을 통해 알 수 있다. (p.194)

예제 5.2 (표5.5)

관측 번호	약의 형태 (type)	항진진대사 점수(x_1)	소요시간 (x_2)	약물량 (y)
1	1	37	61	11.3208
2	2	37	37	12.9151
3	3	45	53	18.8947
4	4	41	41	14.6739
5	5	57	41	8.6493
6	6	49	33	9.5238
7	1	49	49	7.6923
8	2	53	53	0.0017
⋮	⋮	⋮	⋮	⋮

공분산 모형

$$\begin{aligned}y_{ij} &= \mu + \alpha_i + \beta_1(x_{1ij} - \overline{x_{1..}}) + \beta_2(x_{2ij} - \overline{x_{2..}}) + \varepsilon_{ij} \\&= \beta_0 + \alpha_i + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij} \\&= \beta_{0i} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}\end{aligned}\quad (\text{식 5.13})$$

$$(\beta_0 = \mu - \beta_1 \bar{x}_{1..} - \beta_2 \bar{x}_{2..}, \beta_{0i} = \beta_0 + \alpha_i, \varepsilon_{ij} \sim iid N(0, \sigma^2))$$

공분산 모형 설명

- ▶ y_{ij} : i 번째 약의 상태에서 j 번째 개체의 배출된 약물량
- ▶ x_1 : 항진진대사 점수, x_2 : 소요시간
- ▶ μ : 반응변수의 전체평균
- ▶ α_i : 약의 형태의 효과
- ▶ β_1 : 모든 처리에 공통으로 작용하는 첫번째 공변량인 항 신진대사 점수의 효과
- ▶ β_2 : 두번째 공변량인 소요시간의 효과
- ▶ $\bar{x}_{1..}, \bar{x}_{2..}$: 각각 두 공변량의 전체평균
- ▶ ε_{ij} : 등분산을 갖는 정규분포를 따른다고 가정

공분산 모형

공변량이 하나인 경우에서와 마찬가지로 두 개의 공변량이
모든 약의 형태에 대해서 같은 효과를 가져야한다는 가정을 만족



공분산분석 실시

공변량이 두 개인 경우

◆ 공분산분석에서의 검정

1. $H_0 : \beta_1 = 0$: 항진대사의 효과가 없음
2. $H_0 : \beta_2 = 0$: 소요시간의 효과가 없음
3. $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_6$:

항진대사의 효과와 소요시간의 효과를 제어한 후 배출된 약물량의 모평균이 약의 형태에 따라 차이가 없음

R program

```
library(lsmmeans); library(car);  
약물배출량자료=read.table("약물배출량자료.txt",header=T)  
약물배출량자료$약형태 = factor(약물배출량자료$약형태)  
attach(약물배출량자료)  
options(contrasts=c("contr.sum", "contr.poly"))  
  
# 회귀계수의 동일성 검정(교호작용 존재 확인)  
공분산모형1=lm(약물량 ~ 약형태 + 약형태*항신진대사점수 +  
약형태*소요시간, data=약물배출량자료)  
Anova(공분산모형1, type='3')
```

R program (계속)

이원공분산분석(Two-way ANOVA)

```
공분산모형2=lm(약물량 ~ 약형태 + 항진진대사점수 + 소요시간,  
data=약물배출량자료)  
summary(공분산모형2)
```

공변량 효과 제어 시 약형태의 효과 검정

```
anova(공분산모형2)  
Anova(공분산모형2, type='3')
```

LSMEANS(Adjusted means) 계산

```
lsmeans(공분산모형2, ~약형태)
```

R program (계속)

Tukey(HSD) 검정(다중비교)

```
공분산모형2.lsm=lsmmeans(공분산모형3, pairwise ~ 약형태,  
glhargs=list())  
print(공분산모형2.lsm, omit=1)  
plot(공분산모형2.lsm[[2]])
```

모형적합성 검토

```
par(mfrow=c(2,2))  
plot(공분산모형2)
```

회귀계수의 동일성 검정

```
> 공분산모형1=lm(약물량 ~ 약형태 + 약형태*항신진대사점수 + 약형태*소요시간, data=약물배출량자료)
> Anova(공분산모형1, type='3')
Anova Table (Type III tests)
```

Response: 약물량

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	0.95	1	0.0397	0.8443
약형태	99.22	5	0.8325	0.5434
항신진대사점수	6.71	1	0.2815	0.6022
소요시간	8.58	1	0.3599	0.5561
약형태: 항신진대사점수	172.69	5	1.4488	0.2550
약형태: 소요시간	42.33	5	0.3551	0.8722
Residuals	429.10	18		

약물배출량자료의 각 모수 추정치

```
> # 이원공분산분석(Two-way ANOVA)
> 공분산모형2=lm(약물량 ~ 약형태 + 항신진대사점수 + 소요시간, data=약물배출량자료)
> summary(공분산모형2)
```

Call:

```
lm(formula = 약물량 ~ 약형태 + 항신진대사점수 + 소요시간, data = 약물배출량자료)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-7.2182	-2.6338	-0.3798	1.4762	9.6464

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	36.6379	8.9567	4.091	0.000329	***
약형태2	0.9011	2.7816	0.324	0.748388	
약형태3	7.9145	2.7677	2.860	0.007925	**
약형태4	3.0769	2.8239	1.090	0.285185	
약형태5	9.5484	2.8526	3.347	0.002340	**
약형태6	5.8434	2.8383	2.059	0.048929	*
항신진대사점수	-0.7607	0.1375	-5.533	6.47e-06	***
소요시간	0.1647	0.1061	1.552	0.131810	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.751 on 28 degrees of freedom

Multiple R-squared: 0.6131, Adjusted R-squared: 0.5163

F-statistic: 6.337 on 7 and 28 DF, p-value: 0.0001631

'H₀: 약 형태의 효과와 공변량들의 효과가 없다' 라는 귀무가설에 대한 검정결과

공분산분석 결과

```
> # 공변량 효과 제어 시 약형태의 효과 검정
```

```
> anova(공분산모형2)
```

Analysis of Variance Table

Response: 약물량

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
약형태	5	250.40	50.08	2.2182	0.0805 .
항진진대사점수	1	696.73	696.73	30.8610	6.095e-06 ***
소요시간	1	54.40	54.40	2.4098	0.1318
Residuals	28	632.14	22.58		

제 1종 제곱합(Type I SS)
: SS(약형태)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> Anova(공분산모형2, type='3')
```

Anova Table (Type III tests)

Response: 약물량

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	377.77	1	16.7328	0.0003294 ***
약형태	422.22	5	3.7403	0.0101773 *
항진진대사점수	691.24	1	30.6179	6.47e-06 ***
소요시간	54.40	1	2.4098	0.1318104
Residuals	632.14	28		

제 3종 제곱합(Type III SS)
: SS(약형태 | 항진진대사점수, 소요시간)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

공분산 모형식 (식 5.14)

$$\begin{aligned}\hat{y}_{1j} &= \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}_1 x_{11j} + \hat{\beta}_2 x_{21j} \\ &= 41.1853 - 4.5474 - 0.761x_{11j} + 0.165x_{21j} \\ \hat{y}_{2j} &= \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta}_1 x_{12j} + \hat{\beta}_2 x_{22j} \\ &= 41.1853 - 3.6463 - 0.761x_{12j} + 0.165x_{22j} \\ \hat{y}_{3j} &= \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta}_1 x_{13j} + \hat{\beta}_2 x_{23j} \\ &= 41.1853 + 3.3671 - 0.761x_{13j} + 0.165x_{23j} \\ \hat{y}_{4j} &= \hat{\beta}_0 + \hat{\alpha}_4 + \hat{\beta}_1 x_{14j} + \hat{\beta}_2 x_{24j} \\ &= 41.1853 - 1.4705 - 0.761x_{14j} + 0.165x_{24j} \\ \hat{y}_{5j} &= \hat{\beta}_0 + \hat{\alpha}_5 + \hat{\beta}_1 x_{15j} + \hat{\beta}_2 x_{25j} \\ &= 41.1853 + 5.0010 - 0.761x_{15j} + 0.165x_{25j} \\ \hat{y}_{6j} &= \hat{\beta}_0 + \hat{\alpha}_6 + \hat{\beta}_1 x_{16j} + \hat{\beta}_2 x_{26j} \\ &= 41.1853 + 1.2961 - 0.761x_{16j} + 0.165x_{26j}\end{aligned}\tag{5.14}$$

○○○ 보정된 평균 (Adjusted mean) ○○○

```
> # LSMEANS(Adjusted means) 계산
```

```
> lsmeans(공분산모형2, ~약형태)
```

약형태	lsmean	SE	df	lower.CL	upper.CL
1	5.67	1.99	28	1.59	9.75
2	6.57	1.95	28	2.58	10.56
3	13.58	1.97	28	9.55	17.62
4	8.75	1.95	28	4.74	12.75
5	15.22	1.99	28	11.15	19.29
6	11.51	1.98	28	7.46	15.57

```
Confidence level used: 0.95
```

처리별 보정된 평균 (식 5.15)

$$\begin{aligned}\bar{y}_{ad1} &= \hat{\beta}_0 + \hat{\alpha}_1 + \hat{\beta}_1 \bar{x}_{1..} + \hat{\beta}_2 \bar{x}_{2..} \\ &= 41.1853 - 4.5474 - 0.761 \times 51.22 + 0.165 \times 48.56 = 5.670\end{aligned}$$

$$\begin{aligned}\bar{y}_{ad2} &= \hat{\beta}_0 + \hat{\alpha}_2 + \hat{\beta}_1 \bar{x}_{1..} + \hat{\beta}_2 \bar{x}_{2..} \\ &= 41.1853 - 3.6463 - 0.761 \times 51.22 + 0.165 \times 48.56 = 6.571\end{aligned}$$

$$\begin{aligned}\bar{y}_{ad3} &= \hat{\beta}_0 + \hat{\alpha}_3 + \hat{\beta}_1 \bar{x}_{1..} + \hat{\beta}_2 \bar{x}_{2..} \\ &= 41.1853 + 3.3671 - 0.761 \times 51.22 + 0.165 \times 48.56 = 13.584\end{aligned} \tag{5.15}$$

$$\begin{aligned}\bar{y}_{ad4} &= \hat{\beta}_0 + \hat{\alpha}_4 + \hat{\beta}_1 \bar{x}_{1..} + \hat{\beta}_2 \bar{x}_{2..} \\ &= 41.1853 - 1.4705 - 0.761 \times 51.22 + 0.165 \times 48.56 = 8.747\end{aligned}$$

처리 간 다중비교 (표5.17)

```
> # Tukey(HSD) 검정 (다중비교)
> 공분산모형2.lsm=lsmmeans(공분산모형2, pairwise ~ 약형태, glhargs=list())
> print(공분산모형2.lsm, omit=1)
```

\$lsmmeans

약형태	lsmean	SE	df	lower.CL	upper.CL
1	5.67	1.99	28	1.59	9.75
2	6.57	1.95	28	2.58	10.56
3	13.58	1.97	28	9.55	17.62
4	8.75	1.95	28	4.74	12.75
5	15.22	1.99	28	11.15	19.29
6	11.51	1.98	28	7.46	15.57

Confidence level used: 0.95

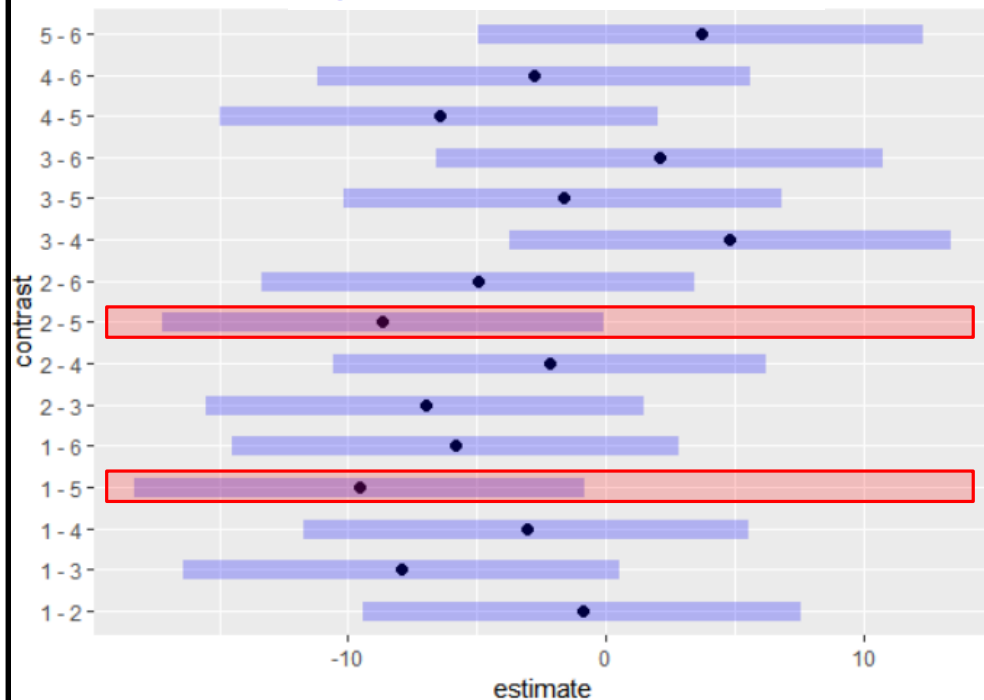
\$contrasts

contrast	estimate	SE	df	t.ratio	p.value
1 - 2	-0.901	2.78	28	-0.324	0.9995
1 - 3	-7.914	2.77	28	-2.860	0.0767
1 - 4	-3.077	2.82	28	-1.090	0.8814
1 - 5	-9.548	2.85	28	-3.347	0.0256
1 - 6	-5.843	2.84	28	-2.059	0.3367
2 - 3	-7.013	2.78	28	-2.521	0.1522
2 - 4	-2.176	2.75	28	-0.790	0.9669
2 - 5	-8.647	2.80	28	-3.086	0.0467
2 - 6	-4.942	2.76	28	-1.792	0.4866
3 - 4	4.838	2.80	28	1.727	0.5263
3 - 5	-1.634	2.78	28	-0.587	0.9911
3 - 6	2.071	2.84	28	0.729	0.9765
4 - 5	-6.471	2.78	28	-2.326	0.2176
4 - 6	-2.767	2.75	28	-1.005	0.9124
5 - 6	3.705	2.83	28	1.309	0.7778

P value adjustment: tukey method for comparing a family of 6 estimates

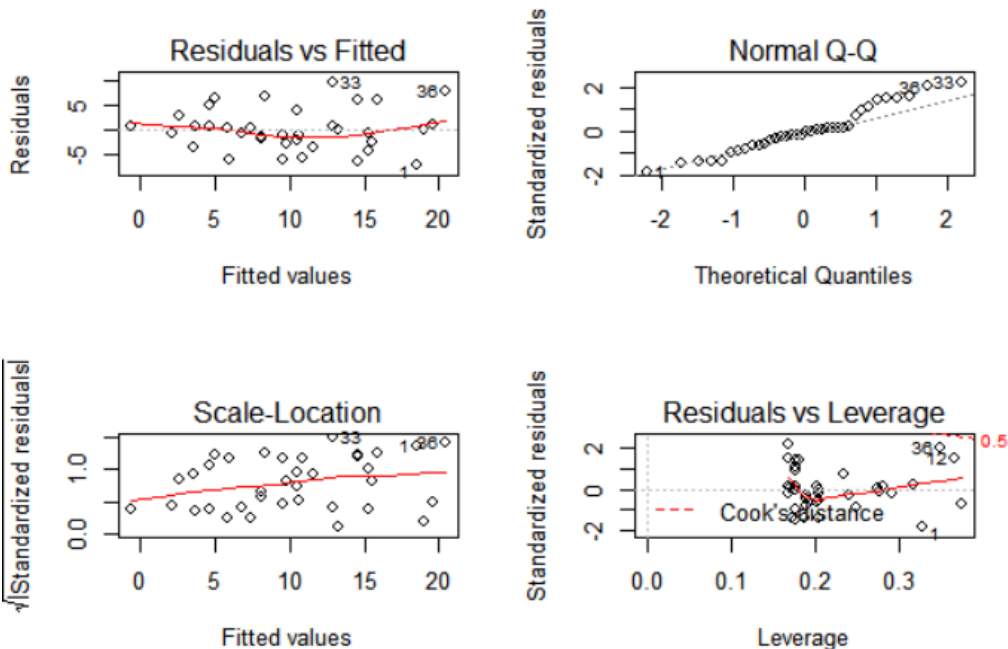
Tukey 신뢰구간 (그림 5.4)

```
> plot(공분산모형2.lsm[[2]])
```



모형적합성 검토 (그림 5.5)

```
> # 모형적합성 검토  
> par(mfrow=c(2,2))  
> plot(공분산모형2)
```



핵심정리

1. 공분산분석(ANCOVA) – ANOVA with covariate

2. 공분산분석을 하기 위해 필요한 두 가지 가정

- 1) 각 처리 안에서 반응변수에 미치는 공변량의 효과가 모두 동일해야 한다.
- 2) 공변량 효과가 0 이 아니다.

3. 공분산분석을 하기 위해 검정할 두 가지 귀무가설

$H_0 : \beta = 0$ [공변량이 하나일때]

$\beta_1 = \beta_2 = 0$ [공변량이 둘일때]

$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_l$

(공변량들의 효과가 없다)

(처리 간 차이가 없다)

핵심정리

4. 공변량이 하나인 경우

- 1) $\text{trt} \times x$ 를 모형에 넣어, 각 처리마다 공변량 효과가 같은가를 검정
- 2) Type III SS를 사용하여 처리효과 검정

5. 공변량이 둘인 경우

- 1) $\text{trt} \times x_1, \text{trt} \times x_2$ 를 모형에 넣어, 각 처리마다 공변량들의 효과가 같은가를 검정
- 2) Type III SS를 사용하여 처리효과 검정



제5강

수고하셨습니다!

