정보통계학과 **백재욱** 교수

제4강 (4장)

이원배치법

정보통계학과 **백재욱** 교수

제4강 이원배치법

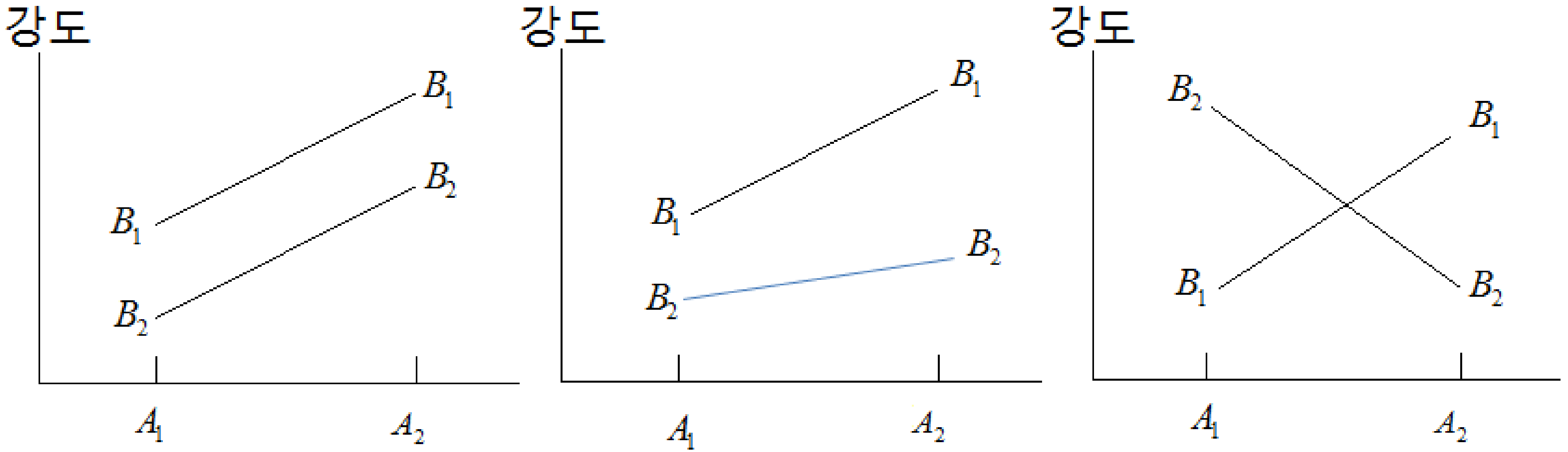
4.1 이원배치법 개요

4.1 이원배치법 개요

- 2개의 요인 A, B(독립변수)와 반응변수(종속변수)간의 관계를 살펴보기 위한 실험계획이다.
- 독립변수는 불연속적인 값을 갖고, 종속변수는 연속적인 값을 갖는다.
- 반복이 없는 경우와 있는 경우가 있다.
- 반복이 있는 경우에는 두 요인 간 상호작용효과를 검출할 수 있다.

4.1 이원배치법 개요

- 주효과(main effect): 요인 A의 수준 간 차이가 있는가?
- 상호작용효과(interaction effect, 교호작용효과): 요인 A의 서로 다른 수준에서 요인 B의 주효과가 다른가?



(a) 상호작용이 없는 경우 (b) 상호작용이 약한 경우 (c) 상호작용이 센 경우

<그림 4-1> 두 요인 간의 상호작용효과

제4강 이원배치법

4.2 실험의 랜덤화

4.2 실험의 랜덤화

◆완전 확률화 계획법

랜덤한 순서대로 두 요인의 수준의 조합조건에서 시험함

<표 4-1> 랜덤화를 위한 번호 부여

요인B / 요인A	B_1	B_2	B_3
A_1	1 ② 2 ③	3 ⑦ 4	5 ④ 6 ①
A_2	7 8 ⑤	9 10	11 12

R

sample(12)

[1] 6 1 2 5 8 9 3 12 10 7 4 11

정보통계학과 **백재욱** 교수

제4강 이원배치법

4.3 고정모형 (A, B 고정요인)

4.3 고정모형 (A, B 고정요인)

모수모형

<표 4-2> 이원배치법의 자료 구조

	B_1	B_2	…	B_b
A_1	x_{111} x_{112} $\vdots$ x_{11r}	x_{121} x_{122} $\vdots$ x_{12r}	… … …	x_{1b1} x_{1b2} $\vdots$ x_{1br}
A_2	x_{211} x_{212} $\vdots$ x_{21r}	x_{221} x_{222} $\vdots$ x_{22r}	… … …	x_{2b1} x_{2b2} $\vdots$ x_{2br}
…	…			
A_a	x_{a11} x_{a12} $\vdots$ x_{a1r}	x_{a21} x_{a22} $\vdots$ x_{a2r}	… … …	x_{ab1} x_{ab2} $\vdots$ x_{abr}

반복이 있는 경우의 장점

- 인자 조합의 효과 (교호작용 또는 상호작용)를 실험오차와 분리하여 구할 수 있다.
- 교호작용을 분리하여 검출할 수 있으므로 인자의 효과 (주효과)에 대한 검출이 좋아진다.
- 실험오차를 단독으로 구할 수 있다.

4.3 고정모형 (A, B 고정요인)

◆ 반복이 있는 이원배치 모수모형 (A, B 두 인자 모두 모수인자인 경우)

- 데이터의 구조 모형 : $x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$
- μ : 전체 모평균
- α_i : 인자 A의 효과로서 $\sum \alpha_i = 0$
- β_j : 인자 B의 효과로서 $\sum \beta_j = 0$
- $(\alpha\beta)_{ij}$: 인자 A, B의 교호작용효과로서

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0, \quad \varepsilon_{ijk} \sim N(0, \ \sigma^2{}_E)$$

4.3 고정모형 (A, B 고정요인)

◆ 검정하고자 하는 가설

(1) 인자 A에 대한 가설

$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$ (인자 A의 효과 간 차이가 없다)

H_1 : 적어도 하나의 α_i는 0 이 아니다. (인자 A의 효과 간 차이가 있다)

(2) 인자 B에 대한 가설

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$ (인자 B의 효과 간 차이가 없다)

H_1 : 적어도 하나의 β_i는 0 이 아니다. (인자 B의 효과 간 차이가 있다)

(3) 인자 A와 B의 교호작용에 대한 가설

H_0 : 모든 $(\alpha\beta)_{ij} = 0$ (교호작용이 없다)

H_1 : 적어도 $(\alpha\beta)_{ij}$ 중 하나는 0 이 아니다. (교호작용이 존재한다)

4.3 고정모형 (A, B 고정요인)

◆ 분산분석

$$(x_{ijk} - \bar{\bar{x}}) = (\bar{x}_{ij.} - \bar{\bar{x}}) + (x_{ijk} - \bar{x}_{ij.})$$

$$(x_{ijk} - \bar{\bar{x}}) = (\bar{x}_{i..} - \bar{\bar{x}}) + (\bar{x}_{.j.} - \bar{\bar{x}}) + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{\bar{x}}) + (x_{ijk} - \bar{x}_{ij.})$$

➡ $$\sum_i\sum_j\sum_k (\bar{x}_{ij.} - \bar{\bar{x}})^2 = \sum_i\sum_j\sum_k (\bar{x}_{i..} - \bar{\bar{x}})^2 + \sum_i\sum_j\sum_k (\bar{x}_{.j.} - \bar{\bar{x}})^2$$

$$+ \sum_i\sum_j\sum_k (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{\bar{x}})^2 + \sum_i\sum_j\sum_k (x_{ijk} - \bar{x}_{ij.})^2$$

➡ $$SS_T = SS_A + SS_B + SS_{A\times B} + SS_E$$

4.3 고정모형 (A, B 고정요인)

<표 4-4> 반복이 있는 이원배치법의 분산분석표

요 인	제곱합	자유도	평균제곱	F_0
요인 A	SS_A	$a-1$	MS_A	$MS_A/\ MS_E$
요인 B	SS_B	$b-1$	MS_B	$MS_B/\ MS_E$
상호작용 $A \times B$	$SS_{A\times B}$	$(a-1)(b-1)$	$MS_{A\times B}$	$MS_{A\times B}/\ MS_E$
E	SS_E	$ab(r-1)$	MS_E	
T	SS_T	$abr-1$		

(1) 인자 A에 대한 가설 검정

검정통계량 $F_0 = MS_A/MS_E > F(a-1,\ ab(r-1),\ \alpha)$

➡ 유의수준 α 에서 귀무가설 기각

(인자 A가 반응치에 유의한 영향을 준다)

4.3 고정모형 (A, B 고정요인)

(2) 인자 B에 대한 가설 검정

검정통계량 $F_0 = MS_B / MS_E > F(b-1,\ ab(r-1),\ \alpha)$

➡ 유의수준 α 에서 귀무가설 기각

(인자 B가 반응치에 유의한 영향을 준다)

(3) 교호작용 A×B에 대한 가설 검정

검정통계량 $F_0 = MS_{A\times B} / MS_E > F((a-1)(b-1),\ ab(r-1),\ \alpha)$

➡ 유의수준 α 에서 귀무가설 기각

(두 인자 A와 B 사이에 교호작용이 존재한다)

4.3 고정모형 (A, B 고정요인)

예제 4.1 4종류의 사료와 3종류의 돼지품종이 체중증가에 미치는 영향을 조사하라.

<표 4-5> 돼지 체중 증가량

사료 \ 품종	B_1	B_2	B_3
A_1	64, 66, 70	72, 81, 64	74, 51, 65
A_2	65, 63, 58	57, 43, 52	47, 58, 67
A_3	59, 68, 65	66, 71, 59	58, 39, 42
A_4	58, 41, 46	57, 61, 53	53, 59, 38

4.3 고정모형 (A, B 고정요인)

풀이 **(1) 가설의 설정**

① $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$

H_1 : 적어도 하나의 α_i 는 0 이 아니다.

② $H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$

H_1 : 적어도 하나의 β_i 는 0 이 아니다.

③ H_0 : 모든 $(\alpha\beta)_{ij} = 0$

H_1 : 적어도 $(\alpha\beta)_{ij}$ 중 하나는 0 이 아니다.

4.3 고정모형 (A, B 고정요인)

풀이(계속) **(2) 수치변환** : $y_{ij} = x_{ij} - 60$

<표 4-6> 수치변환 후의 자료

사료 \ 품종	B_1	B_2	B_3
A_1	4, 6, 10	12, 21, 4	14, −9, 5
A_2	5, 3, −2	−3, -17, -8	−13, −2, 7
A_3	−1, 8, 5	6, 11, −1	−2, −21, −18
A_4	−2, −19, −14	−3, 1, −7	−7, −1, -22

4.3 고정모형 (A, B 고정요인)

풀이(계속) (3) T_{ij} 표의 작성

<표 4-7> T_{ij} 표

$A \backslash B$	B_1	B_2	B_3	$T_{i.}$
A_1	20	37	10	67
A_2	6	−28	−8	−30
A_3	12	16	−41	−13
A_4	−35	−9	−30	−74
$T_{.j}$	3	16	−69	$-50 = T$

4.3 고정모형 (A, B 고정요인)

풀이(계속) (4) 변동의 계산

$$CT = \frac{T^2}{abr} = \frac{2500}{36} = 69.44$$

$$SS_T = \sum_i \sum_j \sum_k x^2{}_{ijk} - CT = 3848 - 69.44 = 3778.56$$

$$SS_{AB} = \sum_i \sum_j \frac{T^2{}_{ij.}}{r} - CT = 2346.67 - 69.44 = 2277.23$$

$$SS_T = SS_T - SS_{AB} = 1501.33$$

$$SS_A = \sum_i \frac{T^2{}_{i..}}{br} - CT = \frac{1}{9}\{67^2 + (-30)^2 + (-13)^2 + (-74)^2\} - 69.44 = 1156.56$$

$$SS_B = \sum_j \frac{T^2{}_{.j.}}{ar} - CT = \frac{1}{12}\{3^2 + 16^2 + (-69)^2\} - 69.44 = 349.39$$

$$SS_{A\times B} = SS_{AB} - SS_A - SS_B = 771.28$$

4.3 고정모형 (A, B 고정요인)

풀이(계속) (5) 자유도의 계산

총 변동의 자유도 : $\phi_T = abr - 1 = 35$

요인 A : $\phi_A = a - 1 = 3$

요인 B : $\phi_B = b - 1 = 2$

상호작용 A × B : $\phi_{A\times B} = (a-1)(b-1) = 6$

잔차 : $\phi_E = ab(r-1) = 24$

4.3 고정모형 (A, B 고정요인)

풀이(계속) (6) 분산분석표의 작성

<표 4-8> 돼지 체중 증가량 자료 분산분석표

요인	제곱합	자유도	평균제곱	F_0
A (사료)	1156.56	3	385.52	6.16 *
B (품종)	349.39	2	174.70	2.79
$A \times B$ (상호작용)	771.28	6	128.55	2.05
E	1501.33	24	62.56	
T	3778.56	35		

- 두 요인 간 상호작용 효과에 대한 유의확률은 0.098이다(약간의 상호작용).
- 사료 종류에 따라 체중증가에 유의한 차이가 있다(유의한 주효과).
- 돼지품종에 대한 유의확률은 0.081로 약간의 영향력을 행사하는 듯하다(약간의 주효과).

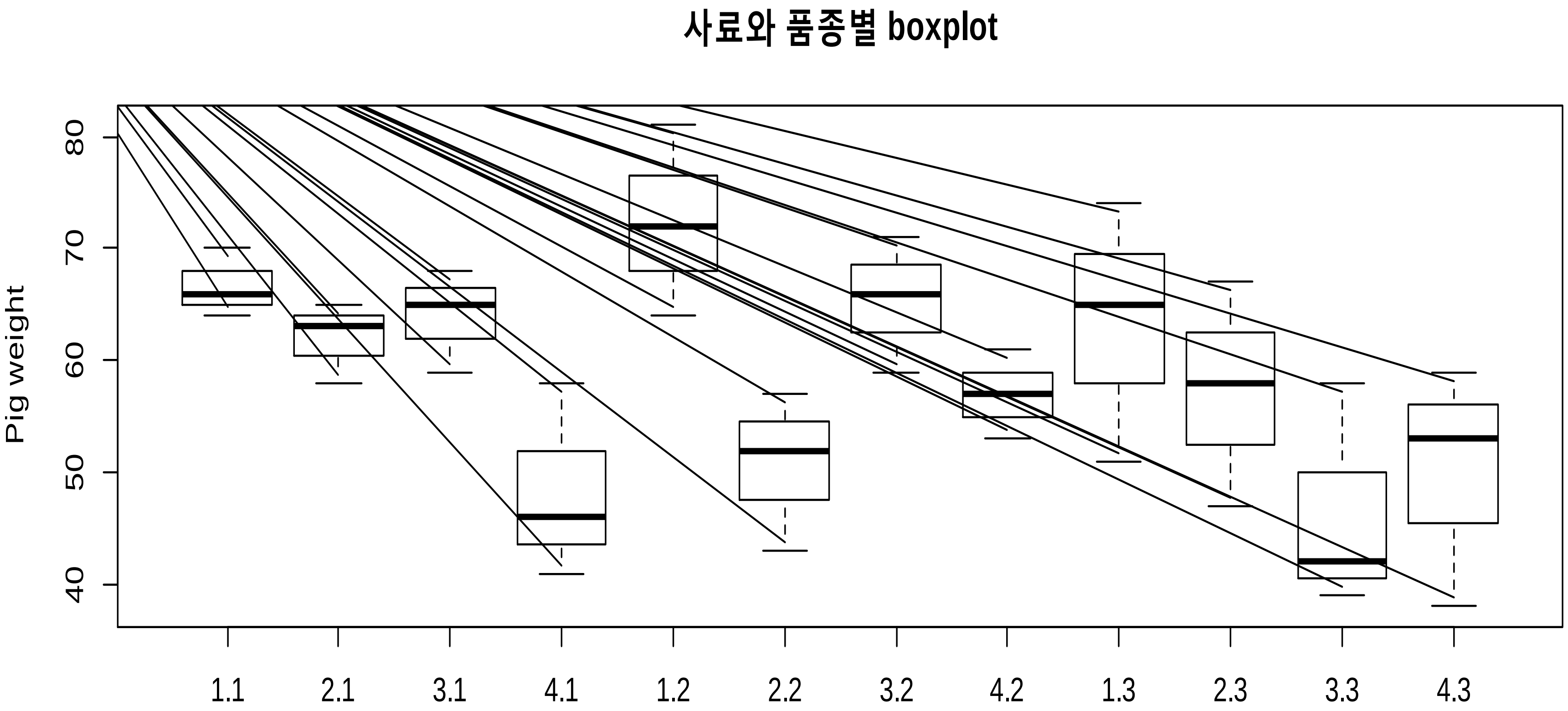
4.3 고정모형 (A, B 고정요인)

R 실습

```
wt = c(64, 66, 70, 72, 81, 64, 74, 51, 65, 65, 63, 58, 57, 43, 52, 47, 58, 67, 59, 68, 65, 66, 71,
59, 58, 39, 42, 58, 41, 46, 57, 61, 53, 53, 59, 38)
sa = c(1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4)
pum = c(1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3,1,1,1,2,2,2,3,3,3)
pig.wt = data.frame(sa, pum, wt)

pig.wt$sa = factor(pig.wt$sa, levels=c(1,2,3,4), labels=c("A1","A2","A3","A4"))
pig.wt$pum = factor(pig.wt$pum, levels=c(1,2,3), labels=c("B1","B2","B3"))
boxplot(wt ~ sa*pum, data=pig.wt, ylab="Pig weight", main="사료와 품종별 boxplot")
```

4.3 고정모형 (A, B 고정요인)

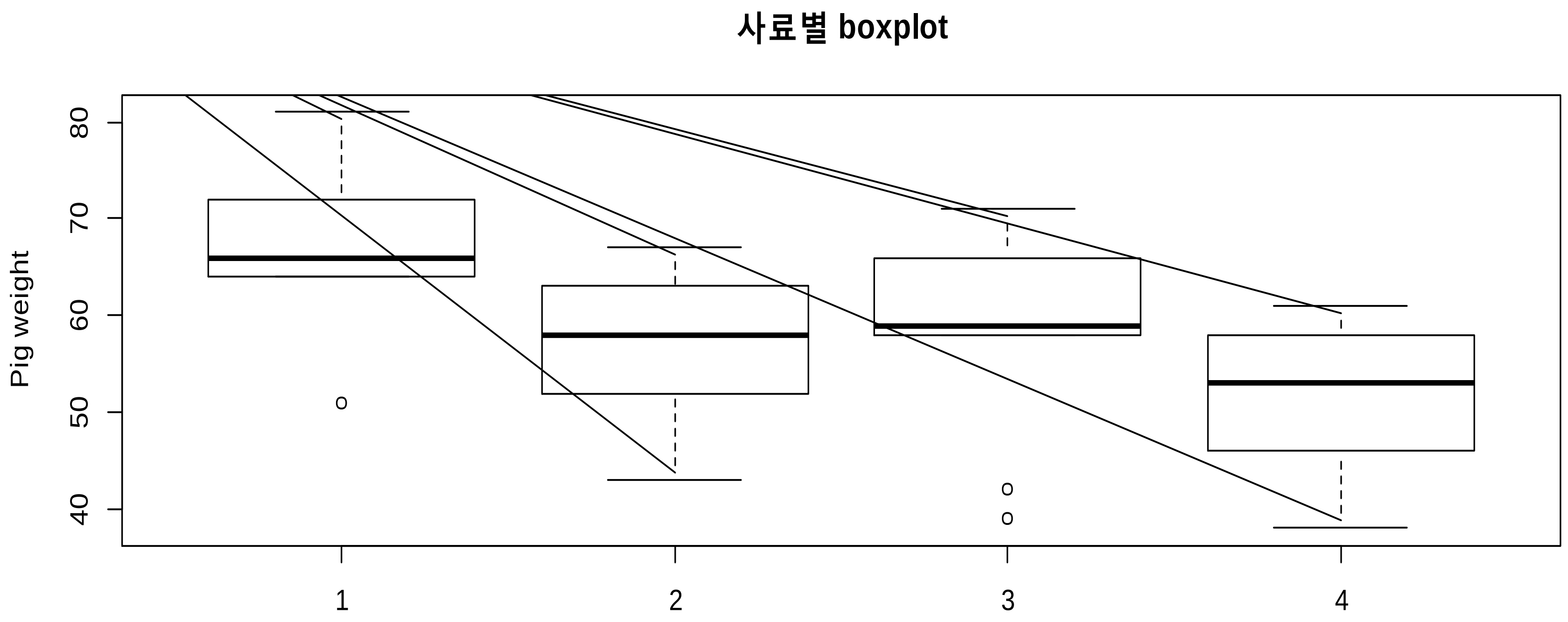
R 실습(계속)



4.3 고정모형 (A, B 고정요인)

R 실습(계속)

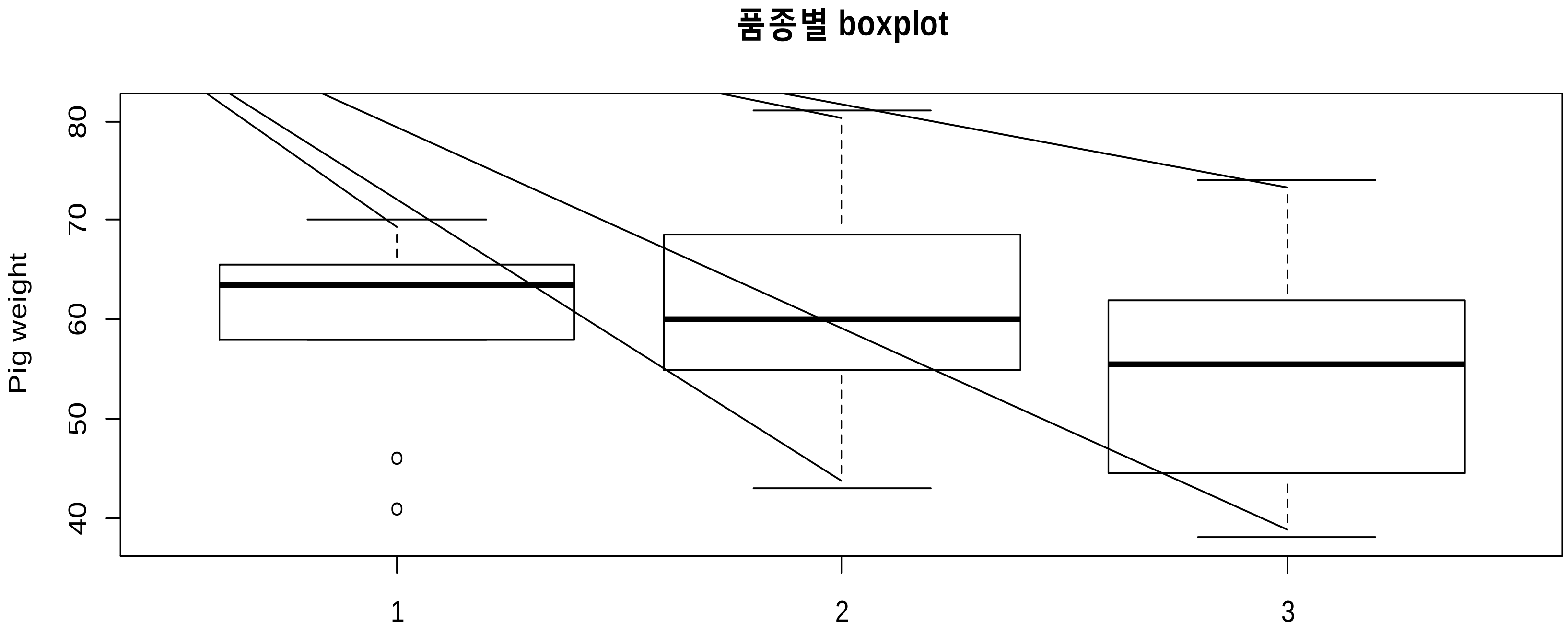
boxplot(wt ~ sa, data=pig.wt, ylab="Pig weight", main="사료별 boxplot")



4.3 고정모형 (A, B 고정요인)

R 실습(계속)

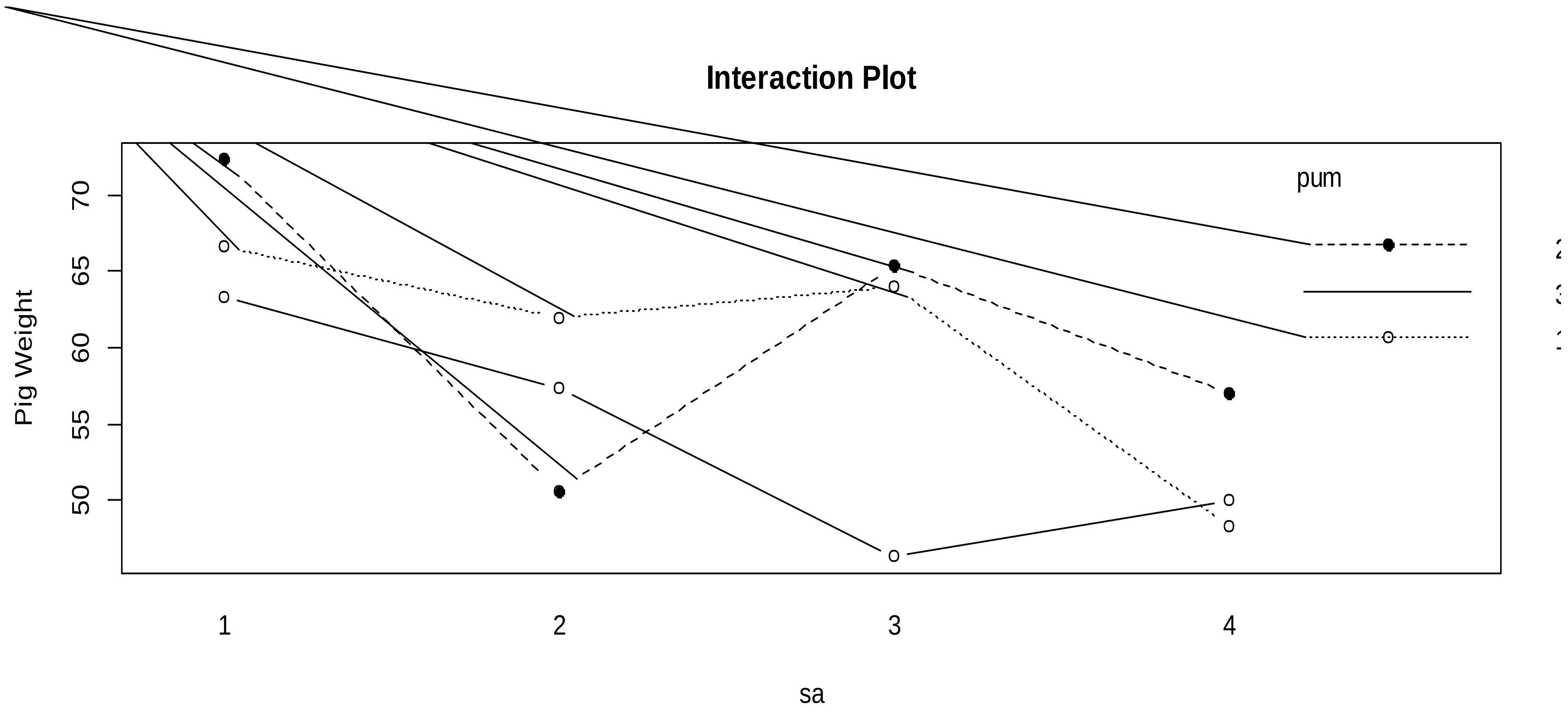
boxplot(wt ~ pum, data=pig.wt, ylab="Pig weight", main="품종별 boxplot")



4.3 고정모형 (A, B 고정요인)

R 실습(계속)

```
with(pig.wt, interaction.plot(x.factor=sa, trace.factor=pum, response=wt, fun=mean, type="b",
legend=T, ylab="Pig Weight", main="Interaction Plot", pch=c(1,19)))
```



4.3 고정모형 (A, B 고정요인)

R 실습(계속)

bartlett.test(wt ~ pum * sa, data=pig.wt) # Bartlett's test for homogeity of variances

Bartlett test of homogeneity of variances

data: wt by pum by sa

Bartlett's K-squared = 0.6876, df = 2, p-value = 0.7091

aov.out = aov(wt ~ sa * pum, data=pig.wt)

summary(aov.out)

```
          Df  Sum Sq  Mean Sq  F value  Pr(>F)
sa         3  1156.6  385.5    6.163    0.00294 **
pum        2  349.4   174.7    2.793    0.08121 .
sa:pum     6  771.3   128.5    2.055    0.09712 .
Residuals 24 1501.3 62.6
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

4.3 고정모형 (A, B 고정요인)

이원배치에서 반복이 없는 경우

모형: $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

총변동의 분해: $SS_T = SS_A + SS_B + SS_E$

<표 4-9> 반복수가 1인 이원배치법의 분산분석표

요 인	제곱합	자유도	평균제곱	F_0
요인 A	SS_A	$a-1$	MS_A	MS_A / MS_E
요인 B	SS_B	$b-1$	MS_B	MS_B / MS_E
E	SS_E	$(a-1)(b-1)$	MS_E	
T	SS_T	$ab-1$		

정보통계학과 **백재욱** 교수

제4강 이원배치법

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

01 데이터의 구조모형 (반복이 없는 경우)

- $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$
- μ : 전체 모평균
- α_i : 인자 A의 효과로서 $\sum \alpha_i = 0$,
- β_j : 인자 B의 효과로서 $N(0,\ {\sigma^2}_B)$ $(\sum \beta_j \neq 0)$
- $\varepsilon_{ij} \sim N(0,\ {\sigma^2}_E)$

우리의 관심사

(1) 인자 B는 블록인자이므로 ${\sigma^2}_B$를 추정한다.

$$\widehat{\sigma^2}_B = \frac{MS_B - MS_E}{a}, \quad a : \text{인자 A의 수준수}$$

(2) A 인자는 모수인자이므로 각 수준에서 모평균?

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

02 데이터의 구조모형 (반복이 있는 경우)

- $x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$
- μ : 전체 모평균
- α_i : 인자 A의 효과로서 $\sum\alpha_i = 0$,
- β_j : 인자 B의 효과로서 $N(0,\ {\sigma^2}_B)$을 따름 $(\sum\beta_j \neq 0)$,
- $(\alpha\beta)_{ij}$: 인자 A, B의 교호작용효과로서 서로 독립인 $N(0,\ {\sigma^2}_{A\times B})$ 을 따름
- $\sum_i (\alpha\beta)_{ij} = 0, \quad \sum_j (\alpha\beta)_{ij} \neq 0$

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

02 데이터의 구조모형 (반복이 있는 경우)

<표 4-10> 분산분석표

요인	제곱합	자유도	평균제곱	F −값
인자 $\boldsymbol{A}$	SS_A	$a-1$	MS_A	$MS_A/MS_{A\times B}$
인자 $\boldsymbol{B}$	SS_B	$b-1$	MS_B	MS_B/MS_E
교호작용 $\boldsymbol{A}\times\boldsymbol{B}$	$SS_{A\times B}$	$(a-1)(b-1)$	$MS_{A\times B}$	$MS_{A\times B}/MS_E$
잔차 $\boldsymbol{E}$	SS_E	$ab(r-1)$	MS_E	
합 $\boldsymbol{T}$	SS_T	$abr-1$		

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

예제 4.2 많은 로트(B)에서 3개의 로트를 택하고, 첨가량 A를 3개의 수준에서 실험하며, 각 인자 수준의 조합마다 2번씩 반복 실험했다. 각 인자의 효과와 교호작용효과를 구하라.

<표 4-11> 화학제품의 불순물 양

첨가량 \ 로트	B_1		B_2		B_3	
A_1	1.0	0.3	3.2	2.6	1.3	19
A_2	4.2	3.3	6.1	5.3	3.1	4.1
A_3	5.3	6.2	6.6	7.1	6.0	6.4

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

풀이

<불순물의 양에 대한 분산분석표>

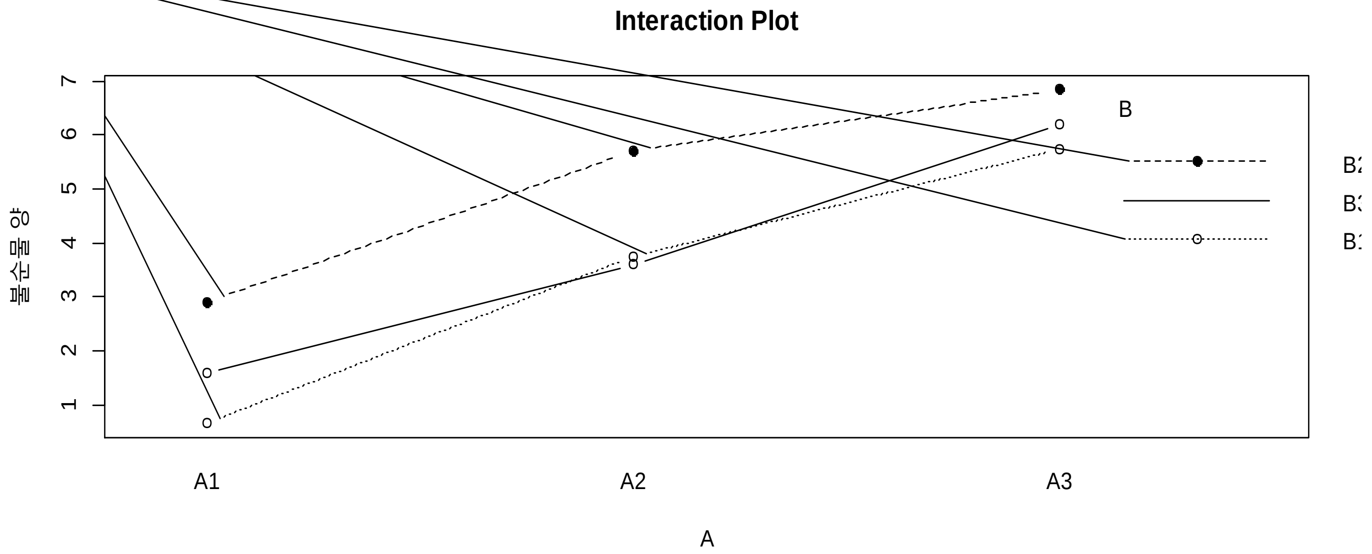
Source	DF	SS	MS	F	P
A	2	62.621	31.311	79.26	0.000
B	2	10.234	5.117	18.88	0.001
Interaction	4	1.582	0.396	1.46	0.292
Error	9	2.440	0.271		
Total	17	76.878			

- 두 인자 A와 B간 교호작용은 유의하지 않으나
  인자 A와 B의 주효과는 매우 유의함
  (유의하지 않은 상호작용효과를 오차에 pooling시킬 수 있음)

4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

R 실습

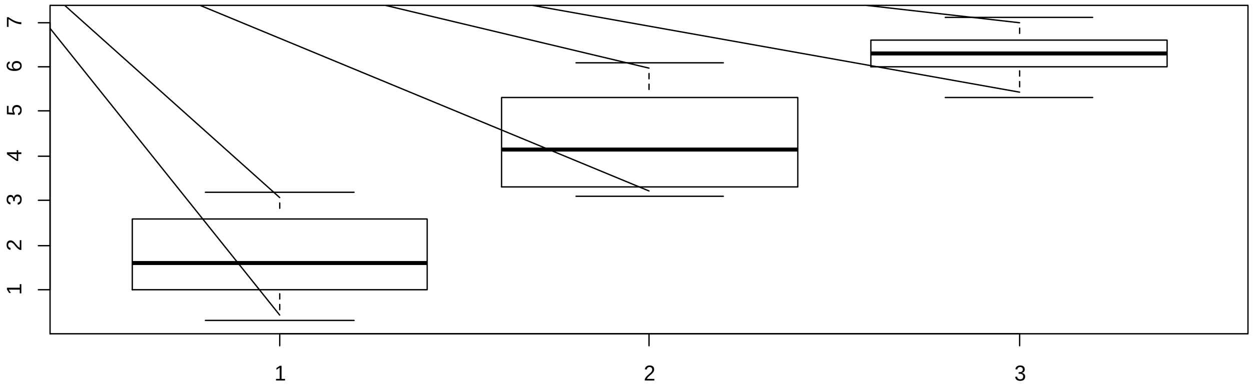
```
y <- c(1.0, 0.3, 3.2, 2.6, 1.3, 1.9, 4.2, 3.3, 6.1, 5.3, 3.1, 4.1, 5.3, 6.2, 6.6, 7.1, 6.0, 6.4)

A <- c(rep(1, 6), rep(2, 6), rep(3, 6))

B <- c(rep(c(1, 1, 2, 2, 3, 3), 3))

c <- data.frame(y, A, B)

c$A <- factor(c$A, levels=c(1, 2, 3), labels=c("A1", "A2", "A3"))

c$B <- factor(c$B, levels=c(1, 2, 3), labels=c("B1", "B2", "B3"))

with(c, interaction.plot(x.factor=A, trace.factor=B, response=y, fun=mean, type="b", legend=T, ylab="불순물 양",

main="Interaction Plot", pch=c(1,19)))
```



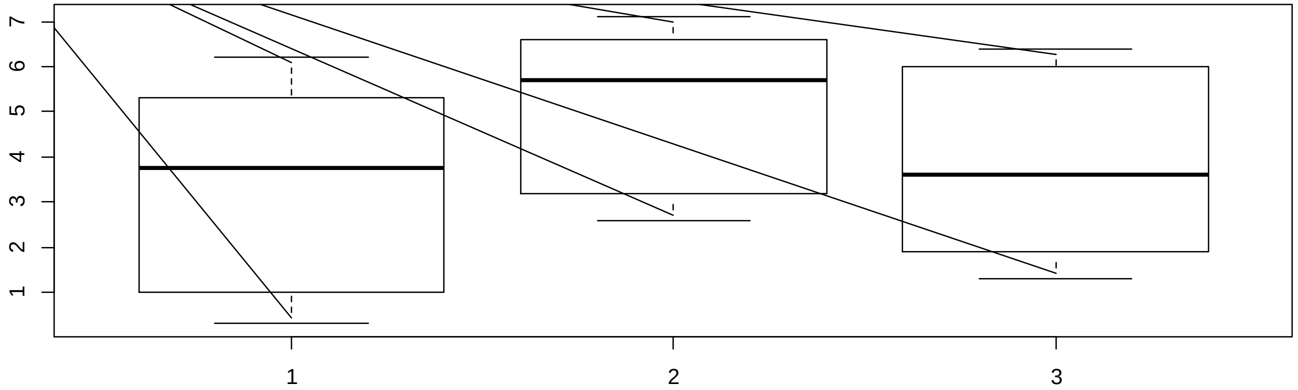
4.4 혼합모형 (A: 고정요인, B: 랜덤요인)

R 실습(계속)

boxplot(y ~ A)



boxplot(y ~ B)



anova1 <- aov(y ~ A*B, data=c) # 요인 A와 B가 모두 고정요인인 경우

anova2 <- aov(y ~ A*B+Error(A/B), data=c) # 요인 A가 고정요인이고 요인 B가 랜덤요인인 경우

summary(anova2)

Error: A

	Df	Sum Sq	Mean Sq
A	2	62.62	31.31

Error: A:B

	Df	Sum Sq	Mean Sq
B	2	10.234	5.117
A:B	4	1.582	0.396

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	9	2.44	0.2711		

정보통계학과 **백재욱** 교수

다음 시간 안내

제5강 (5장)

랜덤화블록계획과 라틴정방계획