7장 | 이원배치법 (TWO-WAY ANOVA)

SAS를 이용한 실험 계획과 분산 분석 (자유아카데미)

- treatment levels(처리수준) may consist of two factors(요인)

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, 3, 4$$

$$\left.\begin{aligned} \tau_1 &= (\text{A1 \& B1}) \\ \tau_2 &= (\text{A1 \& B2}) \\ \tau_3 &= (\text{A2 \& B1}) \\ \tau_4 &= (\text{A2 \& B2}) \end{aligned}\right\} \text{처리조합}$$

- 비료의 종류(A1,A2) + 농약의 종류(B1, B2)
- treatment combination(처리조합): 두 개 이상의 요인(factor)으로 구성된 실험 조건
- CRD with 4 treatment combinations

$$\boxed{y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, 2, 3, 4}$$

- 가능 $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$
- 불가능

$$H_0 \quad : \quad \text{fertilizer effect} = 0$$

or

$$H_0 \quad : \quad \text{pesticide effect} = 0$$

⇨ $$\boxed{y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, 2 \quad j = 1, 2}$$

- 가능

$$H_0 \quad : \quad \alpha_1 = \alpha_2 = 0$$ 비료효과

or

$$H_0 \quad : \quad \beta_1 = \beta_2 = 0$$ 농약효과

"Additive Model" (가법모형) : 상호작용 효과가 없는 모형

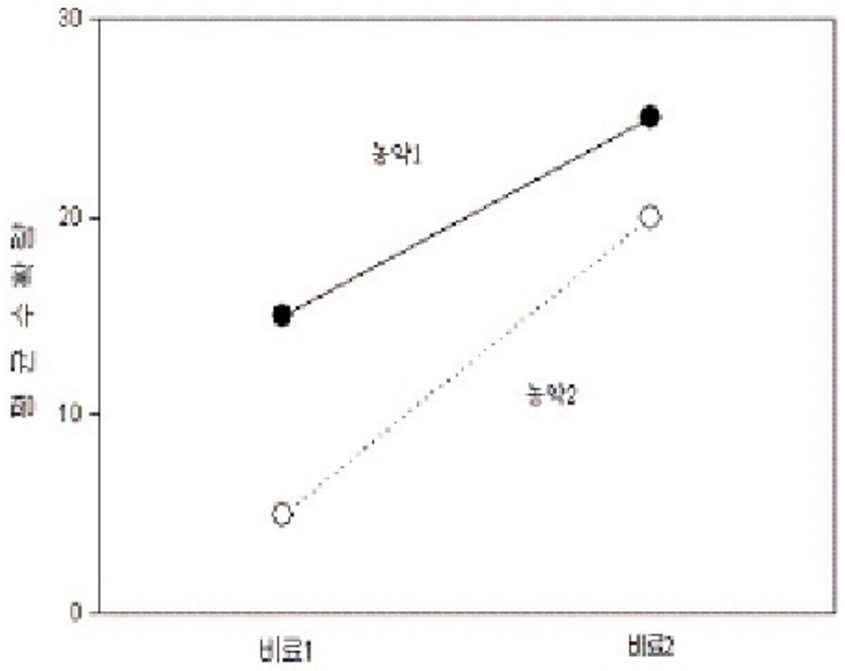
상호작용 효과

어떤 요인의 효과 크기가 다른 요인의 수준에 따라 달라지는 경우, 두 요인 사이에 상호작용 효과(interaction effect)가 존재한다고 한다.

(예) 비료효과는 존재하는가?

> 비료효과는 농약의 종류에 따라 다르다.

> 비료와 농약 사이에 상호작용 존재한다.



상호 작용 효과

주효과와 상호작용효과

- 주효과(main effect): 요인의 수준이 변하면서 기대값이 변하는 현상
- 상호작용효과(interaction effect): 어느 요인의 효과가 다른 요인의 수준에 따라 달라지는 현상

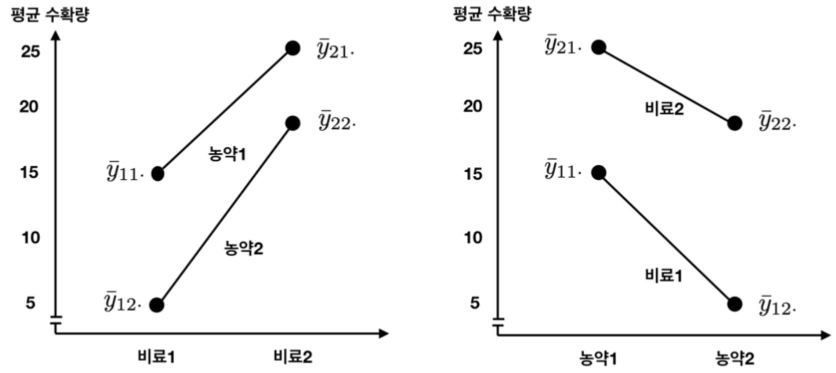


그림 7.2: 수확량에 대한 비료와 농약의 상호작용효과 그래프

가법모형 & 상호작용 모형

비료 농약

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1,2 \quad j = 1,2$$

가법모형
additive model

j=1 (농약1) 일 때, 비료효과?

$$E(y_{11}) - E(y_{21}) = \mu + \alpha_1 + \beta_1 - \mu - \alpha_2 - \beta_1 = \underline{\alpha_1 - \alpha_2}$$

j=2 (농약2) 일 때, 비료효과?

$$E(y_{12}) - E(y_{22}) = \mu + \alpha_1 + \beta_2 - \mu - \alpha_2 - \beta_2 = \underline{\alpha_1 - \alpha_2}$$

항상 동일하다

상호작용을 나타낼 수 없다 !!

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}, \quad i = 1,2 \quad j = 1,2$$

상호작용 모형
interaction model

$$E(y_{11}) - E(y_{21}) = \underline{\alpha_1 - \alpha_2 + (\alpha\beta)_{11} - (\alpha\beta)_{21}}$$

$$E(y_{12}) - E(y_{22}) = \underline{\alpha_1 - \alpha_2 + (\alpha\beta)_{12} - (\alpha\beta)_{22}}$$

반복있는 이원배치법

$$y_{ij} = \mu + \alpha_i + \beta_j + \underline{(\alpha\beta)_{ij} + \epsilon_{ij}}, \quad i = 1, 2 \quad j = 1, 2$$

- 반복없는 이원배치법
- "상호작용효과" "오차항" 구분할 수 없음

$$y_{ijk} = \mu + \alpha_i + \beta_j + \underline{(\alpha\beta)_{ij} + \epsilon_{ijk}}, \quad i = 1, 2 \quad j = 1, 2, \quad k = 1, 2, \cdots, n$$

- 반복있는 이원배치법
- "상호작용효과" "오차항" 구분가능

상호작용이 존재하면 '반복있는 이원배치' 사용해야 함.

이원배치법 모형식

상호작용이 없는 이원배치법(가법모형)의 모형식

$$
\begin{aligned}
&y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \\
&\quad i = 1,\, 2,\, \cdots,\, a, \\
&\quad j = 1,\, 2,\, \cdots,\, b \\
&\quad k = 1,\, 2,\, \cdots,\, n_{ij} \\
&\quad \sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0 \\
&\quad \epsilon_{ijk} \overset{iid}{\sim} N(0,\, \sigma^2)
\end{aligned}
$$

상호작용이 있는 이원배치법(상호작용모형)의 모형식

$$
\begin{aligned}
&y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \\
&\quad i = 1,\, 2,\, \cdots,\, a, \\
&\quad j = 1,\, 2,\, \cdots,\, b \\
&\quad k = 1,\, 2,\, \cdots,\, n_{ij} \\
&\quad \sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0 \\
&\quad \sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \quad j = 1,\, 2,\, \cdots,\, b \\
&\quad \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, \quad i = 1,\, 2,\, \cdots,\, a \\
&\quad \epsilon_{ijk} \overset{iid}{\sim} N(0,\, \sigma^2)
\end{aligned}
$$

PARTITION OF SST

$$
\begin{aligned}
SST &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 \qquad \text{d.f. = abn-1} \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{y}_{.j.} - \bar{y}_{...})^2 \\
&+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2 \\
SSA &= bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 = \sum_{i=1}^{a}\frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn} \qquad \text{d.f. = a-1} \\
SSB &= an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 = \sum_{j=1}^{b}\frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn} \qquad \text{d.f. = b-1} \\
SSAB &= n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \qquad \text{d.f. = (a-1)(b-1)} \\
&= \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{y_{ij.}^2}{n} - \sum_{i=1}^{a}\frac{y_{i..}^2}{bn} - \sum_{j=1}^{b}\frac{y_{.j.}^2}{an} + \frac{y_{...}^2}{abn}
\end{aligned}
$$

간편한 공식

상호작용이 있는 이원배치법에서 제곱합의 간편한 공식

$$\mathrm{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \mathrm{CT}, \qquad \mathrm{CT} = \frac{y_{\cdots}^2}{abn} \tag{7.21}$$

$$\mathrm{SSA} = \sum_{i=1}^{a} \frac{y_{i\cdot\cdot}^2}{bn} - \mathrm{CT}, \tag{7.22}$$

$$\mathrm{SSB} = \sum_{j=1}^{b} \frac{y_{\cdot j\cdot}^2}{an} - \mathrm{CT}, \tag{7.23}$$

$$\mathrm{SSAB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij\cdot}^2}{n} - \sum_{i=1}^{a} \frac{y_{i\cdot\cdot}^2}{bn} - \sum_{j=1}^{b} \frac{y_{\cdot j\cdot}^2}{an} + \mathrm{CT}, \tag{7.24}$$

$$\mathrm{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij\cdot}^2}{n} \tag{7.25}$$

ANOVA

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$ d.f. = abn-ab = ab(n-1)

$$SST = SSA + SSB + SSAB + SSE$$

abn-1	a-1	b-1	(a-1)(b-1)	ab(n-1)

source	d.f.	S.S.	M.S.	F_0
A	a-1	SSA	MSA	MSA/MSE ?
B	b-1	SSB	MSB	MSB/MSE ?
AB	(a-1)(b-1)	SSAB	MSAB	MSAB/MSE ?
Error	ab(n-1)	SSE	MSE	
Total	abn-1	SST		

예

Example 공장에서 생산되는 제품의 불량품 개수에 영향을 끼치는 요인이 무엇인지 알아보려고 한다. 공장에서 사용하는 기계(Machine)효과와, 작업자(Employee) 효과를 이용한 이원배치법을 생각하고 있다. 따라서 현재 사용하는 기계 3 개(Machine1, Machine2, Machine3)와 작업자 2 명(Employee1, Employee2)에 대하여 전체 6 개의 처리조합을 생성하고 각 처리조합에서 생성되는 3 개의 배치[6]에 걸쳐 불량품 개수를 기록하였다. 분산분석을 행하여 불량품에 영향을 주는 요인을 규명하라(유의수준 0.05).

	Employee1	Employee2
Machine1	20, 18, 14	19, 20, 20
Machine2	14, 18, 14	12, 12, 9
Machine3	13, 16, 13	9, 4, 4

표 7.3: 기계와 작업자에 따른 불량품 개수 자료

예

y_{ijk} 를 i 번째 기계에 j 번째 작업자가 작업한 k 번째 배치의 불량품 개수라고 한다면 다음과 같은 모형식이 성립한다.

$$
\begin{aligned}
& y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \\
& \quad i = 1,\ 2,\ 3, \\
& \quad j = 1,\ 2, \\
& \quad k = 1,\ 2,\ 3 \\
& \quad \sum_{i=1}^{3} \alpha_i = 0, \quad \sum_{j=1}^{2} \beta_j = 0 \\
& \quad \sum_{i=1}^{3} (\alpha\beta)_{ij} = 0, \quad j = 1,\ 2, \\
& \quad \sum_{j=1}^{2} (\alpha\beta)_{ij} = 0, \quad i = 1,\ 2,\ 3, \\
& \quad \epsilon_{ijk} \overset{iid}{\sim} N(0,\ \sigma^2)
\end{aligned}
\tag{7.38}
$$

예

제곱합을 계산하면

$$\text{SST} = \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3} y_{ijk}^2 - \frac{y_{\cdots}^2}{18} = 428.500$$

$$\text{SS(Mach)} = \sum_{i=1}^{3} \frac{y_{i\cdot\cdot}^2}{6} - \frac{y_{\cdots}^2}{18} = 229.333$$

$$\text{SS(Empl)} = \sum_{j=1}^{2} \frac{y_{\cdot j\cdot}^2}{9} - \frac{y_{\cdots}^2}{18} = 53.388$$

$$\text{SS(Machi*Empl)} = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{y_{ij\cdot}^2}{3} - \sum_{i=1}^{3} \frac{y_{i\cdot\cdot}^2}{6} - \sum_{j=1}^{2} \frac{y_{\cdot j\cdot}^2}{9} + \frac{y_{\cdots}^2}{18} = 87.111$$

$$\text{SSE} = \text{SST} - \text{SS(Mach)} - \text{SS(Empl)} - \text{SS(Mach*Empl)} = 58.666$$

이를 이용하여 분산분석표(ANOVA table)를 구하면 표 7.4가 되며 Mach*Empl 상호작용효과가 유의하게 나왔으므로 제품의 품질에 영향을 끼치는 요인은 기계와 작업자 모두 해당되며 자세한 분석은 조건부 주효과를 살펴봐야 한다.

Source	d.f.	SS	MS	F_0	p-value
Machine	2	229.333	114.666	$\frac{114.666}{4.888} = 23.45$	0.0001
Employee	1	53.388	53.388	$\frac{53.388}{4.888} = 10.92$	0.0063
Mach*Empl	2	87.111	43.555	$\frac{43.555}{4.888} = 8.91$	0.0042
Error	12	58.666	4.888		
Total	$abn - 1$	428.500			

표 7.4: 기계와 작업자에 따른 불량품 개수의 분산분석표

SAS CODE

```
proc glm data=a;
  class machine employee;
  model defects=machine|employee;
run;
```

SAS OUTPUT

The GLM Procedure

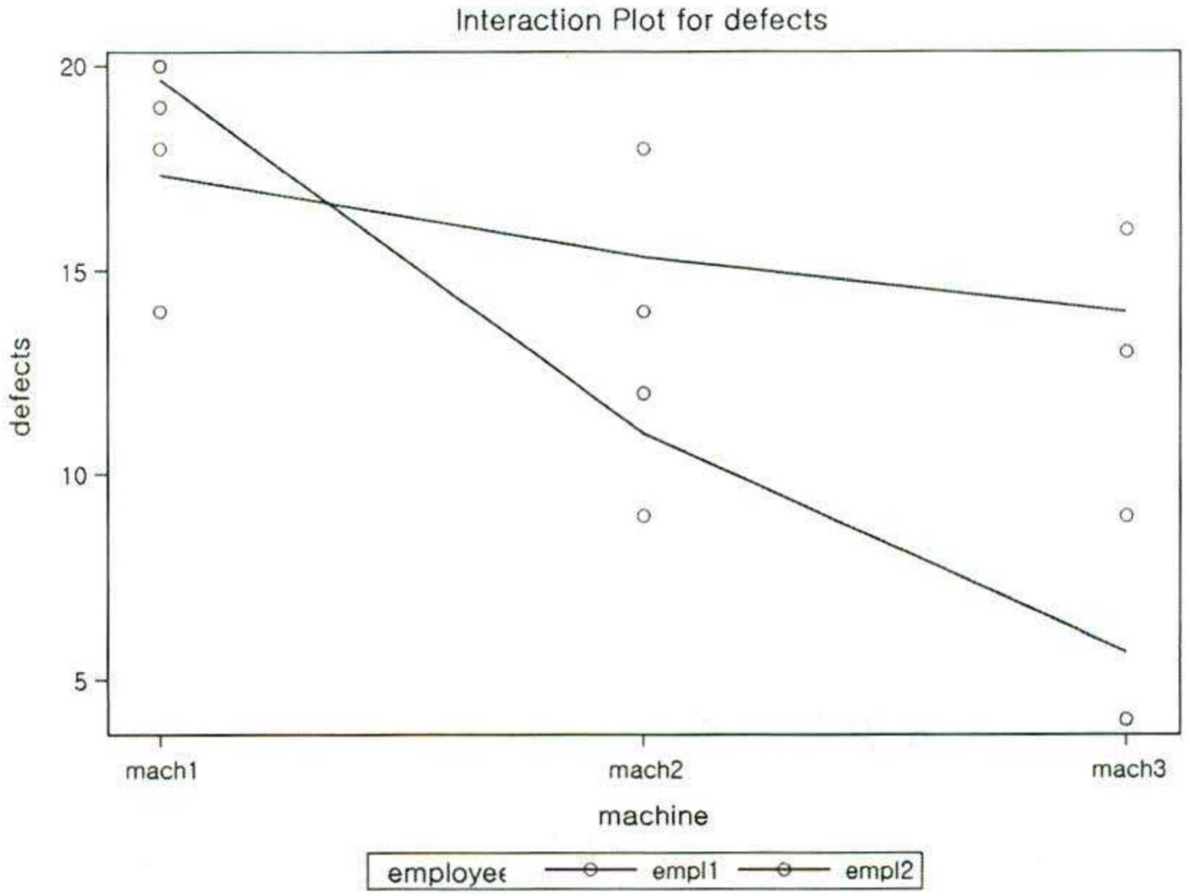
Dependent Variable: defects

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	369.8333333	73.9666667	15.13	<.0001
Error	12	58.6666667	4.8888889		
Corrected Total	17	428.5000000			

R-Square	Coeff Var	Root MSE	defects Mean
0.863088	15.98373	2.211083	13.83333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
machine	2	229.3333333	114.6666667	23.45	<.0001
employee	1	53.3888889	53.3888889	10.92	0.0063
machine*employee	2	87.1111111	43.5555556	8.91	0.0042

Source	DF	Type III SS	Mean Square	F Value	Pr > F
machine	2	229.3333333	114.6666667	23.45	<.0001
employee	1	53.3888889	53.3888889	10.92	0.0063
machine*employee	2	87.1111111	43.5555556	8.91	0.0042



SAS CODE

```
proc glm data=a;
  class machine employee;
  model defects=machine|employee;
  lsmeans machine*employee / slice=employee;
  lsmeans machine*employee / slice=machine;
run;
```

SAS OUTPUT

The GLM Procedure
Least Squares Means

machine*employee Effect Sliced by employee for defects					
employee	DF	Sum of Squares	Mean Square	F Value	Pr > F
empl1	2	16.888889	8.444444	1.73	0.2192
empl2	2	299.555556	149.777778	30.64	<.0001

The GLM Procedure
Least Squares Means

machine*employee Effect Sliced by machine for defects					
machine	DF	Sum of Squares	Mean Square	F Value	Pr > F
mach1	1	8.166667	8.166667	1.67	0.2205
mach2	1	28.166667	28.166667	5.76	0.0335
mach3	1	104.166667	104.166667	21.31	0.0006