

10장 | 임의 완전화 블록 설계


SAS를 이용한 실험 계획과 분산 분석 (자유아카데미)

서론

- 타이어 제조회사에서 개발된 두 종류의 타이어(A, B)가 마모도(wearing degree)에 따른 차이가 있는지 조사를 했다. 실험을 위해 동일한 자동차에 타이어A와 타이어B를 각각 앞바퀴에 한 개씩 설치하여 일정거리를 주행한 후, 마모도를 측정하였다.

car	타이어A	타이어B
1	y_{11}	y_{21}
2	y_{12}	y_{22}
3	y_{13}	y_{23}
4	y_{14}	y_{24}
5	y_{15}	y_{25}

서로 독립인 두 모집단의 모평균 비교?

• 두 모집단 간의 모평균 비교  $t_0 = \frac{\bar{y}_{1.} - \bar{y}_{2.}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$H_0 : \mu_A = \mu_B$$

$$y_{1j} \sim \text{i.i.d. } N(\mu_A, \sigma^2)$$

$$y_{2j} \sim \text{i.i.d. } N(\mu_B, \sigma^2)$$

y_{1j} and y_{2j} are independent ?

짜의 차 검정

- 같은 자동차 내의 상관관계 영향을 “짜”을 이용해 제거함
- 같은 자동차내의 유사성(운전자의 운전습관, 자동차 얼라인먼트 etc.) 이 높을 수록 “짜”의 효과는 높다.
- 짜 = BLOCK 의 개념
- 블록을 효율적으로 구성해서 실험설계를 하면 결과를 효율적으로 얻을 수 있다.

짝의 차 검정

car	타이어A	타이어B	$d_i = y_{1i} - y_{2i}$
1	y_{11}	y_{21}	d_1
2	y_{12}	y_{22}	d_2
3	y_{13}	y_{23}	d_3
4	y_{14}	y_{24}	d_4
5	y_{15}	y_{25}	d_5

$$d_i \sim \text{i.i.d. } N(\mu_d, \sigma_d^2)$$

$$H_0 : \mu_d = 0 \quad \longrightarrow \quad t_0 = \frac{\bar{d}}{s_d / \sqrt{n}}$$

독립 모평균 검정 vs. 짝의 차검정

$$t_0 = \frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10.24 - 9.76}{1.748 \sqrt{\frac{1}{5} + \frac{1}{5}}} = 0.58 < 2.306 = t_{0.025,8}$$

$$t_0 = \frac{\bar{d}}{s_d \sqrt{\frac{1}{n}}} = \frac{0.48}{0.0836 \sqrt{\frac{1}{5}}} = 12.8 > 2.776 = t_{0.025,4}$$

RCBD (RANDOMIZED COMPLETE BLOCK DESIGN)

- 임의화 완전블록설계
- 실험단위를 반응변수에 영향을 끼칠 것으로 기대되는 '외생변수의 값' (예: 자동차) 이 동일하게 짝을 지어 묶어 놓을 때, 이를 '블록'이라고 한다.
- 블록 안에서 원하는 실험조건이 모두 실행될 때, 이를 '완전설계(complete design)'이라 한다.
- RCBD 의 경우, 블록 내 반복 수는 1 이다.

RCBD (RANDOMIZED COMPLETE BLOCK DESIGN)

임의화완전블록설계(RCBD)의 특징

- 임의화(Randomized): 실험조건을 임의로 배정하는 과정은 블록 내에서 이루어진다.
- 완전(Complete): 한 블록 안에 모든 실험조건이 나타난다.
- 블록설계(Block Design): '블록화의 원리'를 이용한 실험설계이다.

RCBD 모형식

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

처리효과



블록효과



$$i = 1, 2, \dots, a$$

$$j = 1, 2, \dots, b$$

$$\sum_{i=1}^a \tau_i = 0$$

$$\sum_{j=1}^b \beta_j = 0$$

$$\epsilon_{ij} \sim \text{i.i.d. } N(0, \sigma^2)$$

$$\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$\text{SST} = \text{SStreat} + \text{SSblock} + \text{SSE}$$

RCBD 제곱합에 대한 간편한 공식

$$SST = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - CT, \quad CT = \frac{y_{..}^2}{ab}$$

$$SS_{\text{block}} = \sum_{j=1}^b \frac{y_{.j}^2}{a} - CT,$$

$$SS_{\text{treat}} = \sum_{i=1}^a \frac{y_{i.}^2}{b} - CT,$$

$$SSE = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \sum_{i=1}^a \frac{y_{i.}^2}{b} - \sum_{j=1}^b \frac{y_{.j}^2}{a} + CT$$

RCBD ANOVA

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

Source	d.f.	MS	E(MS)	F_0
Block	$b - 1$	MSblock	$\sigma^2 + \frac{a}{b-1} \sum_{j=1}^b \beta_j^2$	
Treat	$a - 1$	MStreat	$\sigma^2 + \frac{b}{a-1} \sum_{i=1}^a \tau_i^2$	$\frac{\text{MStreat}}{\text{MSE}}$
Error	$(a - 1)(b - 1)$	MSE	σ^2	
Total	$ab - 1$			

SAS

```
proc glm data=a;  
    class treat block;  
    model y=block treat;  
run;
```

- block*treat 제외시킴
- block 을 treat 보다 먼저 입력 (ANOVA 윗 줄)
- block 에 대한 F-검정은 하지 않는다
- 반복없는 2원배치법과 동일 (except 랜덤화 과정)

RCBD 예

Example 타이어 마모도 실험을 위해 그림 10.1같은 RCBD로 4종류의 타이어 (A, B, C, D)를 자동차 4대에 장착하여 주행 후 마모도를 측정하였다. 그 결과 표 10.4를 얻었다. 타이어 종류에 따라 마모도가 다른지 조사해 보자.

타이어 종류	자동차1	자동차2	자동차3	자동차4
A	10.4	10.9	10.5	10.7
B	12.4	12.4	12.3	12.0
C	13.1	13.4	12.9	13.3
D	11.8	11.8	11.4	11.4

표 10.4: 4종류 타이어 마모도에 대한 RCBD 자료

RCBD 예

제곱합을 위해 다음과 같이 계산해 보았다.

$$CT = \frac{y_{..}^2}{ab} = \frac{1}{16} (10.4 + 12.4 + \cdots + 11.4)^2 = 2272.906$$

$$SST = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - CT = 2287.39 - 2272.906 = 14.4843$$

$$SS_{\text{block}} = \sum_{j=1}^b \frac{y_{.j}^2}{a} - CT = \frac{1}{4} (47.7^2 + 48.5^2 + \cdots + 47.4^2) - 2272.906 = 0.2718$$

$$SS_{\text{treat}} = \sum_{i=1}^a \frac{y_{i.}^2}{b} - CT = \frac{1}{4} (42.5^2 + 49.1^2 + \cdots + 46.4^2) - 2272.906 = 13.9218$$

$$SSE = SST - SS_{\text{block}} - SS_{\text{treat}} = 0.2906$$

그 결과 표 10.5의 분산분석표를 얻었다. 따라서

$$F_0 = 143.71 > 3.86 = F_{0.05, 3, 9} \quad (10.11)$$

이므로 타이어 종류에 따라 마모도가 유의하게 다르다.

Source	d.f.	SS	MS	F_0
자동차	4 - 1	0.2718	0.0906	
타이어	4 - 1	13.9218	4.6406	143.71
오차	(4 - 1)(4 - 1)	0.2906	0.0322	
전체	16 - 1	14.4843		

SAS CODE & OUTPUT

```
proc glm data=a;  
  class car tire ;  
  model wear = car tire ;  
  
run;
```

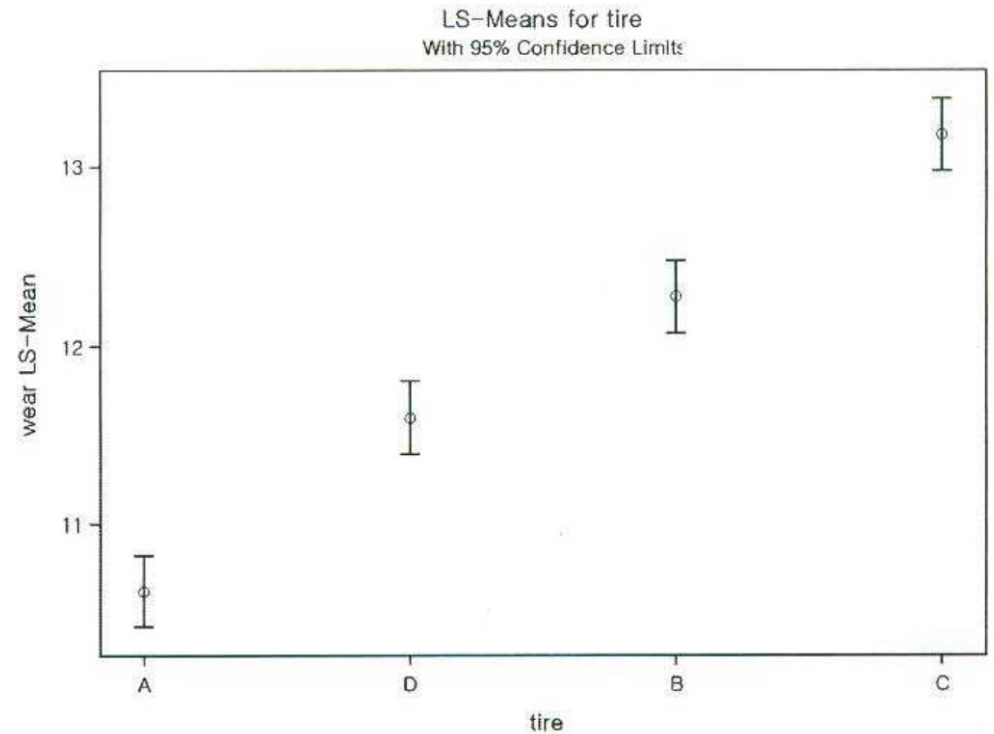
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	14.19375000	2.36562500	73.26	<.0001
Error	9	0.29062500	0.03229167		
Corrected Total	15	14.48437500			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
car	3	0.27187500	0.09062500	2.81	0.1005
tire	3	13.92187500	4.64062500	143.71	<.0001

SAS CODE & OUTPUT

```
proc glm data=a plots=( diffplot(center)
                        meanplot(cl ascending) );

  class car tire ;
  model wear = car tire ;
  lsmeans tire / adjust=tukey lines ;
run;
```



RCBD 안의 이원배치법

예를 들어 타이어 마모도에 대한 실험에서 브랜드(Brand)로는 ‘외국산(foreign)’과 ‘국산(domestic)’을, 타이어 종류(Type)로는 ‘사계절(all season)’과 ‘겨울용(winter)’을 비교하려고 한다. 따라서 전체 처리조합은 $2 \times 2 = 4$ 개가 되고 4대의 자동차에 4 개의 처리조합을 랜덤하게 배치함으로써 RCBD 를 수행하였다. 주행 후 마모도를 측정한 결과 표 10.7을 얻게 되었다.

Brand	Type	자동차1	자동차2	자동차3	자동차4
domestic	all	10.4	10.9	10.5	10.7
domestic	winter	12.4	12.4	12.3	12.0
foreign	all	11.8	11.8	11.4	11.4
foreign	winter	13.1	13.4	12.9	13.3

모형식

$$y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk},$$

$$i = 1, 2, 3, 4, \quad j = 1, 2, \quad k = 1, 2,$$

$$\sum_{i=1}^4 \rho_i = 0, \quad \sum_{j=1}^2 \alpha_j = 0, \quad \sum_{k=1}^2 \beta_k = 0,$$

$$\sum_{j=1}^2 (\alpha\beta)_{jk} = 0, \quad \sum_{k=1}^2 (\alpha\beta)_{jk} = 0,$$

$$\epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$$

여기서 ρ_i 는 블록효과이고 α_j 는 브랜드(외국산, 국산)효과, β_k 는 타이어 종류(사계절, 겨울용)효과를 의미한다. 블록과 처리 간 상호작용은 RCBD 정의상 존재하지 않는데, 이는 자동차 내 브랜드나 타이어 종류의 차이가 자동차가 바뀌어도 일정하게 유지된다는 의미로 해석된다. $(\alpha\beta)_{jk}$ 는 브랜드와 타이어의 상호작용효과

RCBD 제곱합에 대한 간편한 공식

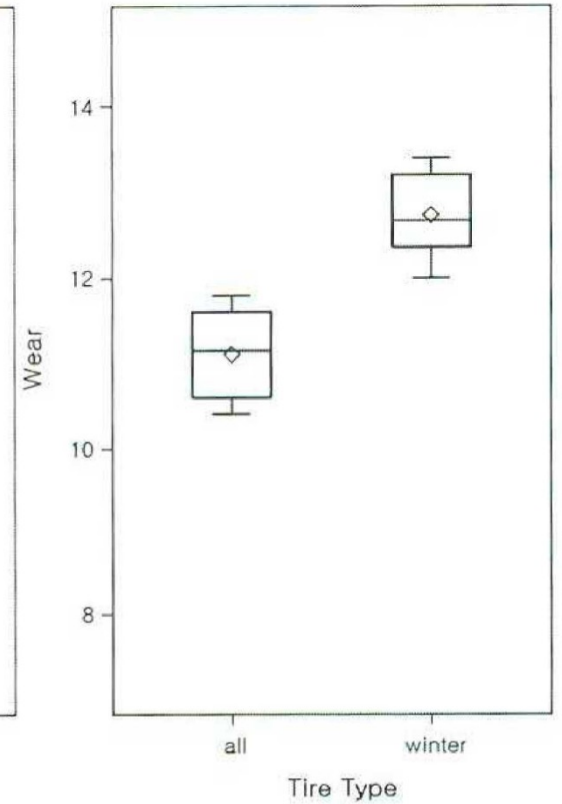
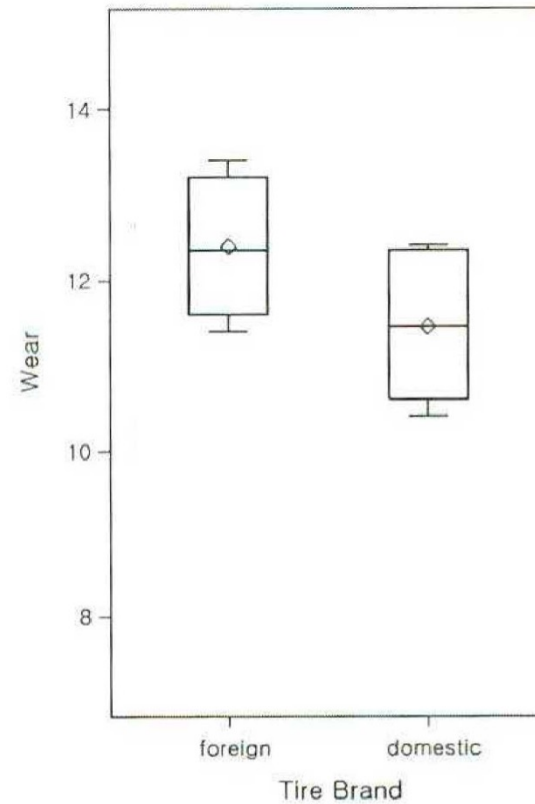
Source	d.f.	SS
Car	$4 - 1$	$\sum_{i=1}^4 \frac{y_{i..}^2}{4} - CT, \quad CT = \frac{y_{...}^2}{16}$
Brand	$2 - 1$	$\sum_{j=1}^2 \frac{y_{.j.}^2}{8} - CT$
Type	$2 - 1$	$\sum_{k=1}^2 \frac{y_{..k}^2}{8} - CT$
Brand*Type	$(2-1)(2-1)$	$\sum_{j=1}^2 \sum_{k=1}^2 \frac{y_{.jk}^2}{4} - \sum_{j=1}^2 \frac{y_{.j.}^2}{8} - \sum_{k=1}^2 \frac{y_{..k}^2}{8} + CT$
Error	$3+3+3$	$SST - SS_{car} - SS_{brand} - SS_{type} - SS_{brand*type}$
Total	$16 - 1$	$\sum_{i=1}^4 \sum_{j=1}^2 \sum_{k=1}^2 y_{ijk}^2 - CT$

표 10.8: RCBD 안의 이원배치법에 대한 자유도와 제곱합

SAS CODE & OUTPUT

```
proc glm data=a;  
  class car brand type ;  
  model wear = car brand type brand*type ;  
run;
```

Source	d.f.	SS	MS	F_0	p-value
Car	3	0.2718	0.0906	2.81	0.1005
Brand	1	3.5156	3.5156	108.87	< 0.0001
Type	1	10.4006	10.4006	322.08	< 0.0001
Brand*Type	1	0.0056	0.0056	0.17	0.6862
Error	9	0.2906	0.0322		
Total	15	14.484			



SUMMARY

- RCBD 는 블록안에서 랜덤화
- 블록을 이용해서 외생변수의 영향을 제어함
- 블록과 블록사이의 SSblock 클수록 SSE 줄어듬
- 블록 수가 증가하면 SSE 의 d.f. 줄어듬
- SSE 는 작을수록, d.f. 는 클수록 좋음
- 적은 블록 수+효과적인 블록킹(blocking) 요구