



실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

<http://jupiter.hallym.ac.kr>

완전임의화 블록설계

세 개이상의 짝을 비교하는 문제

- A, B, C, D 4 종류의 타이어 마모도 비교
- 세 가지 콜라 P, C, E에 대한 맛 평가

와 같이 셋 이상 비교하는 경우 ⇨ 완전임의화 블록설계

블록(block)

어떤 요인의 수준 내에서 임의화가 이루어지면 이 요인을 **블록요인**이라고 하며 블록요인이 포함된 실험설계는 **블록설계(block design)**이라고 함.

블록요인의 각 수준내에서 완전하게 임의화가 된 경우 **완전임의화 블록설계(Completely Randomized Block Design: CRBD 또는 Randomized Complete Block Design: RCBD)**라고 함.

완전임의화 블록설계

네 가지 타이어에 대한 마모도

자동차 1:

A	C	B	D
---	---	---	---

자동차 2:

D	C	B	A
---	---	---	---

⋮

⋮

자동차 b:

A	B	D	C
---	---	---	---

각 자동차에서 타이어의 배정을 임의화 ⇨ 자동차가 블록요인

세 가지 콜라에 대한 평가

평가자 1:

P	C	E
---	---	---

평가자 2:

C	P	E
---	---	---

⋮

⋮

평가자 b:

E	C	P
---	---	---

각 평가자에게 콜라를 임의의 순서로 제공한 후 평가.

완전임의화 블록설계

5개의 블록요인 수준, 4개의 처리 수준인 경우 임의화

Block 1	Block 2	Block 3	Block 4	Block 5
T1	T2	T3	T2	T3
T3	T3	T4	T4	T1
T4	T1	T1	T1	T4
T2	T4	T2	T3	T2

참고: 이원배치법의 교차요인(요인 T와 B)에서 실험은 모든 수준 조합에서 임의배치(숫자는 임의 배정된 실험 순서)

T1B1(17)	T1B2(9)	T1B3(7)	T1B4(20)	T1B5(5)
T2B1(16)	T2B2(2)	T2B3(8)	T2B4(15)	T2B5(14)
T3B1(4)	T3B2(12)	T3B3(6)	T3B4(3)	T3B5(18)
T4B1(11)	T4B2(10)	T4B3(13)	T4B4(19)	T4B5(1)

완전임의화 블록설계

블록요인은

- 최초 실험목적에는 포함되지 않은 경우가 많음
 - 타이어 마모도를 비교하는 것이 목적이었으나 자동차에 따른 차이가 포함
 - 콜라 맛에 대한 비교가 목적이었으나 평가자라는 요인이 추가됨.
- 블록요인에 대한 차이는 있을 수 있으나 이 차이에 대한 관심은 별로 없으며 원래 실험목적에 대한 차이에만 관심.
- 블록에 의한 차이가 있는 경우 블록요인을 고려하지 않으면 오차제곱이 커져 잘못된 분석이 됨.
- 블록은 대개 임의효과인 경우가 많음.

완전임의화 블록설계

모형과 제곱합의 분해(반복이 없는 경우)

y_{ij} : i 번째 처리 j 번째 블록에서의 관측치

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad (i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b)$$

$$\sum \tau_i = 0, \quad \sum \beta_j = 0$$

$\epsilon_{ij} \sim N(0, \sigma^2)$ 이고 모두 독립

τ_i : 처리에 의한 차이

β_j : 블록에 의한 차이

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \quad \text{이므로}$$

양변을 제곱하여 합하면(교차항은 모두 0임이 증명)

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

완전임의화 블록설계

$$CT = \frac{y_{..}^2}{ab} = ab\bar{y}_{..}^2 \text{ 이라 하면}$$

$$\text{SST: } \sum_{i,j} (y_{ij} - \bar{y}_{..})^2 = \sum_{i,j} y_{ij}^2 - CT$$

$$\text{자유도: } ab - 1$$

$$\text{SSTrt: } \sum_{i,j} (\bar{y}_{i.} - \bar{y}_{..})^2 = b \sum_i \bar{y}_{i.}^2 - CT$$

$$\text{자유도: } a - 1$$

$$\text{SSBlock: } \sum_{i,j} (\bar{y}_{.j} - \bar{y}_{..})^2 = a \sum_j \bar{y}_{.j}^2 - CT$$

$$\text{자유도: } b - 1$$

$$\text{SSE: } \sum_{i,j} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \text{SST} - \text{SSTrt} - \text{SSBlock}$$

$$\text{자유도: } ab - a - b + 1 = (a - 1)(b - 1)$$

완전임의화 블록설계

따라서 분산분석표는

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
블록	SSBlock	$b - 1$	$MSBlock = \frac{SSBlock}{b - 1}$		
처리	SSTrt	$a - 1$	$MSTrt = \frac{SSTrt}{a - 1}$	$F_0 = \frac{MSTrt}{MSE}$	$\Pr[F_{a-1, (a-1)(b-1)} > F_0]$
오차	SSE	$(a - 1)(b - 1)$	$MSE = \frac{SSE}{(a - 1)(b - 1)}$		
전체	SST	$ab - 1$			

가 된다.

처리에 따른 차이여부에 대한 귀무가설 $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$ 대 대립가설 $H_1: \text{적어도 하나는 0이 아님}$ 에 대한 검정은 $F_0 > F_{a-1, (a-1)(b-1); \alpha}$ 이면 귀무가설 기각

\Leftrightarrow

유의확률이 유의수준보다 작으면 귀무가설 기각

완전임의화 블록설계

- 관례적으로 분산분석표에서 블록요인은 제일 위에 씀.
- 블록요인에 따른 차이여부는 일반적으로 관심이 없으므로 검정통계량 등 계산은 안함.
- 통계프로그램에서는 블록요인을 따로 지정하지 않으므로 이들 값이 나옴.

블록요인은 실험목적에 포함되지 않은 요인으로 실험에서 임의화를 완전하게 할 수 없어서 발생하는 경우가 대부분이라 블록설계인 것을 놓치기 쉬움

- 네 가지 타이어의 차이: 각 자동차에서 임의로 타이어 A, B, C, D를 배정(어느 요인의 각 수준에서 임의화) 자동차에 의한 차이가 있을 수 있다는 사실을 무시하고 타이어의 차이만 보는 일원배치 분산분석을 하는 오류
- 세 가지 콜라에 대한 평가: 각 평가요원에게 임의의 순서로 콜라 P, C, E를 주고 평가(어느 요인의 각 수준에서 임의화) 평가요원에 따른 차이를 무시하고 콜라의 차이만 보는 일원배치 분산분석을 하는 오류

완전임의화 블록설계

블록을 무시한 경우와 블록이 포함된 분산분석표의 비교

요인	제곱합	자유도
블록	SSBlock	$b - 1$
처리	SSTrt	$a - 1$
오차	SSE	$(a - 1)(b - 1)$
전체	SST	$ab - 1$

요인	제곱합	자유도
처리	SSTrt	$a - 1$
오차	SSE	$ab(a - 1)$
전체	SST	$ab - 1$

$a(b - 1)$

- 블록을 무시한 경우 블록설계의 블록제곱합과 오차제곱합의 합이 일원배치의 오차제곱합이 됨.
- 자유도도 마찬가지
- 블록에 따른 차이가 있는 경우 오차제곱합이 과대 계상되어 유의함에도 유의하지 않다는 결론에 도달할 가능성이 큼

완전임의화 블록설계

보기: 네 종류의 타이어 A, B, C, D의 마모도의 차이가 있는지 알아보기 위해 4대의 자동차에 타이어를 장착하여 일정 거리를 주행한 후 마모도의 차이를 조사하였다. 각 자동차 네 바퀴에 4종류의 타이어를 임의로 선택하여 장착하여 다음과 같은 자료를 얻었다.

타이어	자동차1	자동차2	자동차3	자동차4	평균
A	10.4	10.9	10.5	10.7	10.625
B	12.4	12.4	12.3	12.0	12.275
C	13.1	13.4	12.9	13.3	13.175
D	11.8	11.8	11.4	11.4	11.600
평균	11.925	12.125	11.775	11.850	11.91875

$\sum y_{ij} = 190.7$, $\sum y_{ij}^2 = 2287.39$ 이므로 $CT = 190.7^2/16 = 2272.906$ 이다.

따라서

$$SST: \sum_{i,j} y_{ij}^2 - CT = 2287.39 - 2272.906 = 14.484$$

완전임의화 블록설계

$$\text{SSTrt: } b \sum_i \bar{y}_{i.}^2 - CT = 4 \cdot (10.625^2 + 12.275^2 + 13.175^2 + 11.6^2) - 2272.906 = 13.9215$$

$$\text{SSBlock: } a \sum_j \bar{y}_{.j}^2 - CT = 4 \cdot (11.925^2 + 12.125^2 + 11.775^2 + 11.850^2) - 2272.906 = 0.2715$$

$$\text{SSE: } \text{SST} - \text{SSTrt} - \text{SSBlock} = 14.484 - 13.9215 - 0.2715 = 0.291$$

이를 요약한 분산분석표

요인	제곱합	자유도	평균제곱(MS)	F	유의확률
블록	0.2715	3	0.0905		
처리	13.9215	3	4.6405	143.52	0.0000
오차	0.291	9	0.0323		
전체	14.484	15			

으로 얻는다.

$F_0 = 143.52 > F_{3,9;0.05} = 3.86$ 이므로 귀무가설을 기각한다. 따라서 타이어에 따른 마모도는 차이가 있다. (위 계산에 반올림 오차 있음에 유의)

완전임의화 블록설계

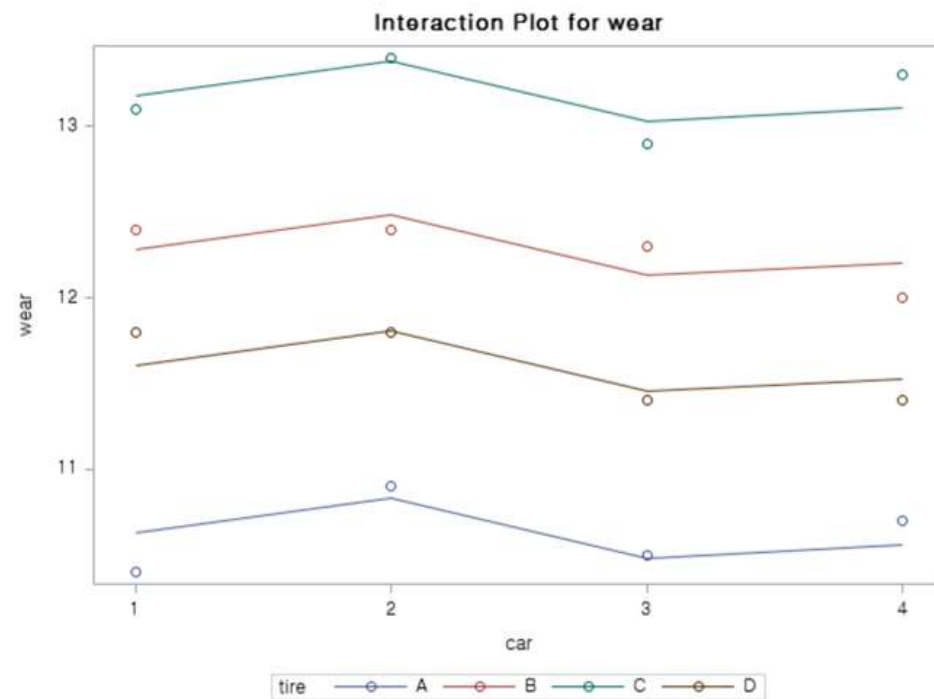
SAS를 사용한 블록설계분석

- 통계 패키지에는 블록요인을 따로 지정하지 않으므로 이원배치법과 같은 방법으로 분석.
- 위 자료는 반복이 없는 이원배치법과 같은 SAS 명령

```
data a /* crbd1.sas */;  
  input tire $ car wear @@;  
  datalines;  
A 1 10.4  A 2 10.9  A 3 10.5  A 4 10.7  
B 1 12.4  B 2 12.4  B 3 12.3  B 4 12.0  
C 1 13.1  C 2 13.4  C 3 12.9  C 4 13.3  
D 1 11.8  D 2 11.8  D 3 11.4  D 4 11.4  
;  
proc glm data = a;  
  class car tire;  
  model wear = car tire;  
run;
```

완전임의화 블록설계

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



수고하셨습니다^^