실험계획 - 5주차 강의

5장: 분할법

시립대학교 통계학과

2021년 3월 29일

실험의 제약

예제: 온도와 시간 (교재 예제 5.4)

- 전자제품을 오븐 속에 넣고 몇 분간 고온으로 가열한 다음 부품이 정산적으로 작동하는 시간을 측정
- 온도: 오븐의 온도를 수시로 변경하는 것은 시간이 많이 걸린다. **수준 변경이 어렵다**.
- 가열 시간: 제품을 오븐에 넣고 가열하면서 시간에 맞추어 꺼내면 된다. **수준 변경이 쉽다**.
- 완전 랜덤화 이원배치 계획법을 적용하면 실험시간이 메우 오래 걸린다.

예제: 관개와 비료

- 관개(irrigation)은 작은 농지에 대하여 랜덤하게 적용하기 어렵다.
- 비료는 농지가 작아도 랜덤하게 적용하기 쉽다.

분할법의 개요

실험의 제약 조건

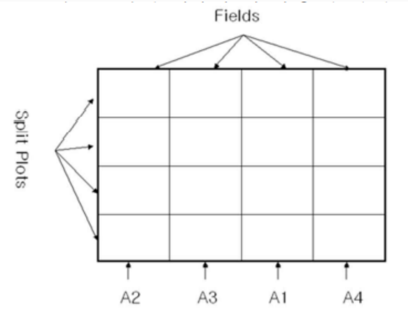
- 요인이 2개 (A, B)
- 요인 A (Hard-to-change factor)의 수준 변경이 어렵다(비용, 시간)
- 요인 B (Easy-to-change factor)의 수준 변경이 쉽다.

이러한 경우 실용적인 실험설계?

- 먼저 수준 변경이 어려운 A 의 수준을 랜덤하게 결정
- 결정된 요인 A 의 수준을 고정시키고 수준 변경이 용이한 요인 B 의 수준을 랜덤하게 순차적으로 배치

분할법의 의미

- Split-Plot design (설계의 이름은 농업 실험에서 유래)
- 교과서에서 분할법 II로 부른다.
  1. 큰 재배지(**main plot, field, 주구**)에 요인 A 를 랜덤하게 배치하고
  2. 주구 안에 작게 나누어진 작은 재배지(**sub plot, split plot, 분할구**)에 요인 B 를 랜덤하게 배치



분할법의 특성

- 실험단위의 내포성 (nested experimental units)

$$\text{sub plot} \subset \text{main plot}$$

- 랜덤화의 계층성 (hierachical ramdomization - two levels)
  1. 먼저 주구를 랜덤하게 배정
  2. 선택된 주구 안에서 분할구를 랜던하게 배정
- **랜덤화는 오차를 발생시킨다. 따라서 2개의 서로 다른 오차항이 나타난다.**
- 완전 랜덤화 이원배치는 랜덤화의 층이 1개

반복이 없는 분할법

반복이 없는 분할법

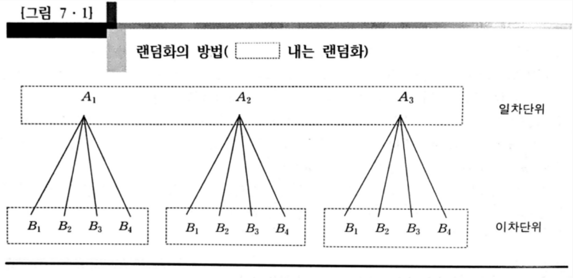


Figure 1: 그림출처: 현대실험계획법, 박성현 저

$$x_{ij} = \mu + \alpha_i + e_{1(i)} + \beta_j + (\alpha\beta)_{ij} + e_{2(ij)}$$

- $e_{1(i)}$: 요인 A의 랜덤화에 따른 주구에 의한 1차 오차항
- $e_{1(ij)}$: 요인 B의 랜덤화에 따른 분할구에 의한 2차 오차항

반복이 없는 분할법

$$x_{ij} = \mu + \underbrace{\alpha_i + e_{1(i)}}_{confounding} + \beta_j + \underbrace{(\alpha\beta)_{ij} + e_{2(ij)}}_{confounding}$$

- $e_{1(i)}$: 요인 A의 랜덤화에 따른 주구에 의한 1차 오차항
- $e_{1(ij)}$: 요인 B의 랜덤화에 따른 분할구에 의한 2차 오차항
- 요인 A 효과 α_i 와 1차 오차항 $e_{1(i)}$ 이 교락
- 상호작용도 2차 오차항 $e_{1(ij)}$ 과 교락
- **쓸모없는 모형!**
- 반복이 필요하다.

반복이 있는 분할법

반복이 있는 분할법

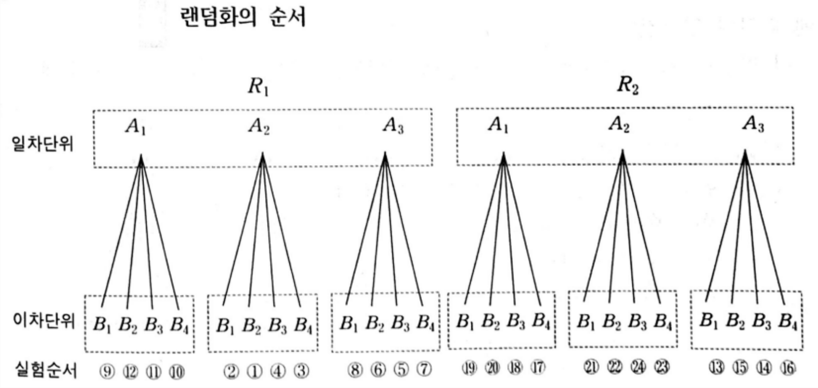
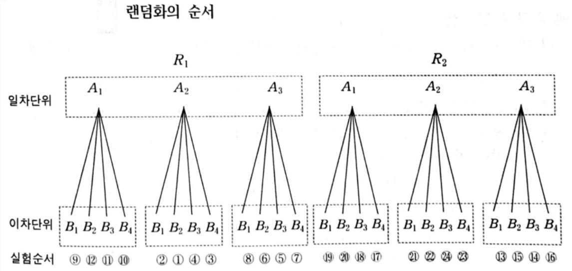


Figure 2: 그림출처: 현대실험계획법, 박성현 저

반복이 있는 분할법



- 반복은 단일 분할법을 여러 번 반복한다.

$$x_{ijk} = \mu + r_k + \alpha_i + e_{1(ik)} + \beta_j + (\alpha\beta)_{ij} + e_{2(ijk)}$$

- 일반적으로 반복의 효과는 r_k은 임의효과(블럭)로 놓는다.

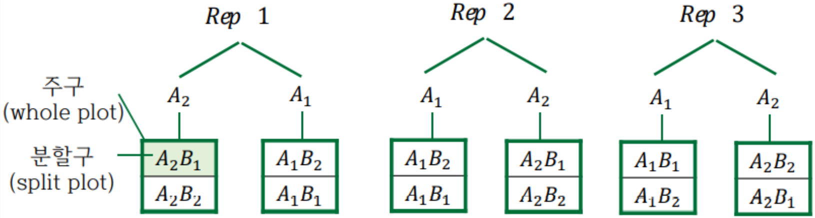
반복이 있는 분할법

$$x_{ijk} = \underbrace{\mu + r_k + \alpha_i + e_{1(ik)}}_{\text{1st level}} + \underbrace{\beta_j + (\alpha\beta)_{ij} + e_{2(ijk)}}_{\text{2nd level}}$$

- 모형을 2개의 레벨로 나눌 수 있다.
- 첫 번째 레벨
  - 반복이 없는 이원배치와 같다.
  - 반복 R 과 요인 A 의 상호작용은 서로 교락. 따라서 추론 불가
- 두 번째 레벨
  - 반복이 있는 이원배치와 같다.
  - 요인 A 과 요인 B 의 상호작용 추론 가능

교과서 예제

■ **공업시험 사례**:
2가지 방법의 공정(A)과 2가지 온도수준(B)이 제품의 강도에 미치는 영향을 조사하기 위해 반복 3회의 실험. A의 수준변경이 어려움. B의 수준변경 쉬움.



< 그림 5.6> 반복 3회의 분할법 II 의 두 단계의 렌덤화의 과정

교과서 예제

< 그림 5.7 > 분할법 II 의 두 단계의 랜덤화의 과정

반복 R	가열시간 B(분)	온도 A (°F)			
		580	600	620	640
I	5	217	158	229	223
	10	233	138	186	227
	15	175	152	155	156
II	5	188	126	160	201
	10	201	130	170	181
	15	195	147	161	172
III	5	162	122	167	182
	10	170	185	181	201
	15	213	180	182	199

- 반복이 없는 3원배치법과 동일한 자료의 구조. 실험의 랜덤화는 다름

분할법에서의 분산분석

분할법에서의 분산분석

요인	제곱합	자유도	평균제곱합	F_0
A	SS_A	$a-1$	MS_A	MS_A/MS_{E1}
R	SS_R	$r-1$	MS_R	
$E1$	SS_{E1}	$(a-1)(r-1)$	MS_{E1}	
B	SS_B	$b-1$	MS_B	MS_B/MS_{E2}
$A \times B$	$SS_{A\times B}$	$(a-1)(b-1)$	$MS_{A\times B}$	$MS_{A\times B}/MS_{E2}$
$E2$	SS_{E2}	$a(b-1)(r-1)$	MS_{E2}	
T	SS_T	$abr-1$		

주의:

- 요인 A의 효과를 위한 F-통계량은 1차 주구 오차 제곱합 SS_{E1} 으로 계산
- 요인 A, 상호작용의 효과를 위한 F-통계량은 2차 분할구 오차 제곱합 SS_{E2} 으로 계산

분할이 불가능한 경우

두 인자가 분할이 불가능한 경우

- 두개의 처리 모두 수준의 변경이 어려운 경우: 화로의 온도와 중간원료의 종류(교과서 예제 5.3)

1. 이원배치법에서 ab개의 처리를 먼저 1차 랜덤화
2. 추출된 처리를 고정시키고 반복 측정 (2차 랜덤화)

- 추출된 처리가 블럭이 되는 경우
- 모형

$$x_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

여기서

$$\gamma_{ij} \sim N(0, \sigma_1^2), \quad e_{ijk} \sim N(0, \sigma_2^2)$$

두 인자가 분할이 불가능한 경우

- 모형을 다음과 같이 쓰기도 한다.

$$x_{ijk} = \mu + \alpha_i + \beta_j + e_{1(ij)} + e_{2(ijk)}$$

- 1차 오차항 $e_{1(ij)}$ 와 상호작용 $(\alpha\beta)_{ij}$는 교락!
- 따라서 자료의 구조는 반복이 있는 이원배치이다.
- 하지만 **랜덤화의 절차가 달라서** 상호작용에 대한 추론이 불가능하다.
- 반복이 있는 이원배치에 블럭 효과가 추가된 경우
- 교과서에서 분할법 I

두 인자가 분할이 불가능한 경우

- 완전 랜덤화 이원배치법

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

- 두 처리의 조합을 먼저 랜덤화

$$x_{ijk} = \mu + \alpha_i + \beta_j + \underbrace{(\alpha\beta)_{ij} + e_{1(ij)}}_{confounding} + e_{2(ijk)}$$

두 인자가 분할이 불가능한 경우 분산분석

요인	제곱합	자유도	평균제곱합	F_0
A	SS_A	$a-1$	MS_A	MS_A/MS_{E1}
B	SS_B	$b-1$	MS_B	MS_B/MS_{E1}
$E1(+A \times B)$	SS_{E1}	$(a-1)(b-1)$	MS_{E1}	
$E2$	SS_{E2}	$ab(r-1)$	MS_{E2}	
T	SS_T	$abr-1$		

주의:

- 요인 A와 요인 B 의 효과를 위한 F-통계량은 1차 오차 제곱합 SS_{E1} 으로 계산
- 1차 오차 제곱합 SS_{E1} 에는 1차 오차와 상호작용이 교락

결론

결론

동일한 자료 구조를 가져도 실험의 구조(랜덤화 방법)에 따라서 분석 방법이 변한다!!

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