■ 반복이 없는 이원배치법

일원배치법은 한 요인의 처리효과를 알아보기위한 실험방법

- 두 요인의 처리 효과를 알아보기 위한 실험 방법
- 교차설계(cross-over design) & 지분설계(nested design)

교차설계

		A					
		1	2	3	4	5	6
B	1	O	O	O	O	O	O
	2	O	O	O	O	O	O
	3	O	O	O	O	O	O

↑ 모든조합에대해 실험가능

지분설계

		A					
		1	2	3	4	5	6
B	1	O	O				
	2			O	O		
	3					O	O

↑ 각각의 경우에서만 실험가능 (한요인이 다른 요인 수준에서만)

ex) A의 1,2에 대해서만 B의 1이 나타나는경우

ex) 긴장(strain)을 측정하는 장비가 5종류고 각 장비는 4개의 헤드가 있을때
측정장비종류(요인), 헤드위치(요인)에 따른 측정에 차이가 있는지 알아본다면?
→ 장비 수준과 헤드수준의 모든결합 조건에 따른 실험 불가능!
헤드는 장비에 포함(nested)되어있다

□ 교차설계

○ 실험 설계

- ◦ 수준 수가 a인 요인 A, 수준 수가 b인 요인 B
- ◦ $a \times b$ 실험 전체를 완전 확률화

○ 자료구조

요인 B \ 요인 A	A_1	A_2	$\cdots$	A_a
B_1	Y_{11}	Y_{21}	$\cdots$	Y_{a1}
B_2	Y_{12}	Y_{22}	$\cdots$	Y_{a2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
B_b	Y_{1b}	Y_{2b}	$\cdots$	Y_{ab}

○ 구조식

- 1-요인설계의 구조식 (일원배치)

$$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + (\mu_i - \mu) + \varepsilon_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

  - τ_i: 요인의 처리효과

- 2-요인설계의 구조식

$$\Rightarrow Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \ j = 1, \ldots, b$$

- μ: 전체 평균

- α_i: 요인 A의 처리효과, $\sum_{i=1}^{a} \alpha_i = 0$ cf) $\Sigma \tau_i = 0$ 인것과유사

- β_j: 요인 B의 처리효과, $\sum_{j=1}^{b} \beta_j = 0$

- $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$: 오차항

○ 변동의 분해

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})$$

$$TSS = SSA + SSB + SSE$$

○ $TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{Y_{..}^2}{N}$: 자유도 $N-1$ $\because \sum\sum(Y_{ij} - \overline{Y}_{..}) = 0$

원래는 $\sum_{i}^{a}\sum_{j}^{b}(\overline{Y}_{i.} - \overline{Y}_{..})^2$

(j에대한제약없음: constant)

○ $SSA = b\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2 = \sum_{i=1}^{a}\frac{Y_{i.}^2}{b} - \frac{Y_{..}^2}{N}$: 자유도 $a-1$ $\because \sum\sum(\overline{Y}_{i.} - \overline{Y}_{..}) = 0$

○ $SSB = a\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2 = \sum_{j=1}^{b}\frac{Y_{.j}^2}{a} - \frac{Y_{..}^2}{N}$: 자유도 $b-1$ $\because \sum\sum(\overline{Y}_{.j} - \overline{Y}_{..}) = 0$

TSS-SSA-SSB ○ $SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$: 자유도 $(a-1)(b-1)$

- SSE 자유도: $N-1-(a-1)-(b-1) = (a-1)(b-1)$

○ 가설 검정

- 요인 A의 처리 효과의 동일성 검정

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

- 요인 B의 처리 효과의 동일성 검정

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

- 분산분석표

검정통계량 : 너무 크면 기각

변인	자유도	제곱합	평균제곱	F	
모형(처리A)	a-1	SSA	MSA = SSA/(a-1)	MSA/MSE	①
모형(처리B)	b-1	SSB	MSB = SSB/(b-1)	MSB/MSE	②
오차	(a-1)(b-1)	SSE	MSE = SSE/((a-1)(b-1))		
전체	N-1	TSS			

◉ 어느 화학공장에서 제품의 생산량에 영향을 미치는 것으로 예상되는 반응온도와 원료를 요인으로 생각하여 반복이 없는 이원배치의 실험 실시

○ 반응온도(A) = 180, 190, 200, 210

○ 원료(B) = 미국 M사, 일본 Q사, 국내 P사

12개의 실험구를 완전 확률화하여 실험한 결과

원료 \ 온도	180	190	200	210	합계
M	97.6	98.6	99.0	98.0	393.2
Q	97.3	98.2	98.0	97.7	391.2
P	96.7	96.9	97.9	96.5	388.0
합계	291.6	293.7	294.9	292.2	1172.4

$\begin{cases} H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0, \quad H_1: \text{not } H_0 \\ H_0: \beta_1 = \beta_2 = \beta_3 = 0 \end{cases}$

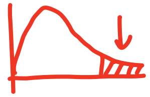
○ $TSS = \sum_{i=1}^{4}\sum_{j=1}^{3} y_{ij}^2 - \frac{y_{..}^2}{12} = 97.6^2 + \cdots + 96.5^2 - \frac{1172.4^2}{12} = 6.22$

○ $SSA = \frac{291.6^2 + 293.7^2 + 294.9^2 + 292.2^2}{3} - \frac{1172.4^2}{12} = 2.22$

○ $SSB = \frac{393.2^2 + 391.2^2 + 388^2}{4} - \frac{1172.4^2}{12} = 3.44$

○ $SSE = TSS - SSA - SSB = 6.22 - 2.22 - 3.44 = 0.56$

변인	자유도	제곱합	평균제곱	F
모형(처리A)	3	2.22	0.74	7.96 $= 0.74/0.093$
모형(처리B)	2	3.44	1.72	18.49 $= 1.72/0.093$
오차	$11-3-2=$ 6	0.56	0.093	
전체	11	6.22		



○ $F_{0.05}(3,6)=4.76,\ F_{0.05}(2,6)=5.14$ ⇨ 유의수준 5%에서 두 요인 모두 유의함 (H_0 두개 모두 기각)

⇒ 반응온도와 원료의 종류에 따라 생산량의 차이가 있다고 할 수 있음

유의한 차이가 없다고 나오는경우?

⇒ 처리를 바꿔도 같다는뜻!

일원배치법으로 재구성가능

```
chemistry <- scan(what=list("","",1))
1 1 97.6 2 1 98.6 3 1 99.0 4 1 98.0
1 2 97.3 2 2 98.2 3 2 98.0 4 2 97.7
1 3 96.7 2 3 96.9 3 3 97.9 4 3 96.5

names(chemistry) <- c("temp","material","amount")
chemistry <- data.frame(chemistry)

result <- lm(amount~temp+material, data=chemistry)
anova(result)
```

문자

길이가같은경우 행 - observation
열 - 변수

linear model

y

a indicator

p-value 직접제공
작을수록 강력한 반증: H_0기각. 차이가있다고 할수있음

$Var(\overline{Y}_{i.}) = Var\left(\frac{\sum_j y_{ij}}{b}\right) = \frac{b\sigma^2}{b^2} = \frac{\sigma^2}{b} \rightarrow$ MSE로 추정 / $Var(\overline{Y}_{.j}) = Var\left(\frac{\sum_i y_{ij}}{a}\right) = \frac{a\sigma^2}{a^2} = \frac{\sigma^2}{a} \rightarrow$ MSE로 추정

○ $\mu(A_i)$와 $\mu(B_j)$의 추정

추정했으니까 z가아니라 t

- $\mu(A_i)$의 구간추정: $\overline{Y}_{i.} \pm t_{\alpha/2,(a-1)(b-1)}\sqrt{MSE/b}$
- $\mu(B_j)$의 구간추정: $\overline{Y}_{.j} \pm t_{\alpha/2,(a-1)(b-1)}\sqrt{MSE/a}$

◉ 95% 신뢰구간

- $t_{\alpha/2,(a-1)(b-1)}\sqrt{MSE/b} = 2.447\sqrt{0.093/3} = 0.43$
- $t_{\alpha/2,(a-1)(b-1)}\sqrt{MSE/a} = 2.447\sqrt{0.093/4} = 0.37$
- $\mu(A_1)$의 95% 신뢰구간 = 97.2 ± 0.43 = [96.77, 97.63]
- $\mu(B_1)$의 95% 신뢰구간 = 98.3 ± 0.37 = [97.93, 98.67]

$\overline{Y}_{1.} \pm t_{\alpha/2,3\times2} \cdot \sqrt{MSE/b}$

$\overline{Y}_{1.} = 291.6/3 = 97.2$
$t_{0.025,6} = 2.447$
$MSE = 0.093$

$\overline{Y}_{.1} \pm t_{\alpha/2,6} \cdot \sqrt{MSE/a}$

$\overline{Y}_{.1} = 393.2/4 = 98.3$
$t_{0.025,6} = 2.447$
$MSE = 0.093$

ex) ① 고흡(약 / 가짜약) → 체질무시
② 사람마다 (약먹기전 / 후) → 체질고려!

■ 확률화 블록설계법 (randomized complete block design)

- 확률화 완비(complete) 블록설계법 : 앞내용과 모양은 비슷한데 개념이 완전 다름...

처리의개수가2개!

- 쌍을 이룬 비교의 일반화
- 블록(block) : 요인의 처리효과 비교의 정확도를 높이기 위해 예비지식을 활용하여 나눈 동질적인 실험단위
  - (예제) 처리: 운동화의 두 상표 block: 운동화를 신은 사람
  - (예제) 처리: 옥수수 품종 block: 지역

○ 실험 설계

- a개의 수준(처리)과 b개의 블록가 있다고 가정
- 각 블록 안에 처리에 대해 관측값은 하나
- **각 블록 안에 처리의 배열은 확률적으로 결정**

- Weight of Chickens – Snee (1985)
  - 사료에 성장촉진제 추가
    - Control (추가하지 않음), Low dose, High dose
  - 크기가 유사한 것으로 블록
  - 성숙기의 평균 무게(단위: pound)

Block	Control	Low dose	High dose	합계
1	3.93	3.99	3.96	11.88
2	3.78	3.96	3.94	11.68
3	3.88	3.96	4.02	11.86
4	3.93	4.03	4.06	12.02
5	3.84	4.10	3.94	11.88
6	3.75	4.02	4.09	11.86
7	3.98	4.06	4.17	12.21
8	3.84	3.92	4.12	11.88
합계	30.93	32.04	32.30	95.27

○ 실험설계

```
trt <- 3
block <- 8
design <- NULL
for (i in 1:block)
  design <- c(design,sample(1:trt,trt))
design <- data.frame(matrix(design,block,trt,byrow=T))
Block <- 1:block
design <- cbind(Block,design)
```

벡터에서 같은길이, 비복원으로 뽑기 ⇒ shuffle! 순서바꿔서 정리

Block	X_1	X_2	X_3
1	o	a	a
2	a	a	a
3		⋮	
⋮			

이런모양으로 실험설계
seed에 따라 다르게 나옴

○ 통계적 모형

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1,\ldots,a, \;\; j = 1,\ldots,b.$$

- Y_{ij}: 블록 j에서 처리 i를 한 반응변수
- μ: 전체 평균
- α_i: 처리효과, $\sum_{i=1}^{a} \alpha_i = 0$
- β_j: 블록 효과, $\sum_{j=1}^{b} \beta_j = 0$
- $\varepsilon_{ij} \sim$ iid $N(0, \sigma^2)$

○ 가설 검정

이원배치랑 수학적으로는 같음ㅎ.ㅎ

- ○ 처리효과의 동일성 검정
  - $H_0 : \alpha_1 = \cdots = \alpha_a = 0$ vs H_1 : 최소한 하나 이상의 α_i는 0이 아님

- ○ 변동분해: $TSS = SSA + SSBL + SSE$
  - $TSS = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - \frac{Y_{..}^2}{N}$: 자유도$=N-1$
  - $SSA = b\sum_{i=1}^{a}(\overline{Y}_{i.} - \overline{Y}_{..})^2 = \sum_{i=1}^{a}\frac{Y_{i.}^2}{b} - \frac{Y_{..}^2}{N}$: 자유도$=a-1$
  - $SSBL = a\sum_{j=1}^{b}(\overline{Y}_{.j} - \overline{Y}_{..})^2 = \sum_{j=1}^{b}\frac{Y_{.j}^2}{a} - \frac{Y_{..}^2}{N}$: 자유도$=b-1$

- $SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$:

  자유도 $= N - (a-1) - (b-1) - 1 = (a-1)(b-1)$

○ 분산분석표

변인	자유도	제곱합	평균제곱	F
모형(처리A)	a-1	SSA	MSA = SSA/(a-1)	MSA/MSE
블록	b-1	SSBL	MSBL = SSBL/(b-1)	MSBL/MSE
오차	(a-1)(b-1)	SSE	MSE = SSE/((a-1)(b-1))	
전체	N-1	TSS		

무슨소리..

- 블록효과의 동일성 검정 → $H_0: \beta_1 = \cdots = \beta_b = 0$
    - 설계에 있어 ab개의 처리 조합은 실험 단위의 집합에 대해 확률적으로 배치된 것이 아님
    - 블록은 실험단위이고 확률화는 각 단위안에서 제한되어짐

    - 만약 두 개의 요인에 대해 관심이 있는 경우에는 다른 설계법을 설계
    - 이원설계의 상대적 효율성을 평가하는데 사용

- F_b가 1보다 크면 클수록 블록화의 효과가 좋음 블록화하길 잘함! 일원설계안돼

  ⇨ 이원설계가 일원설계에 비해 효율적임

- F_b가 1보다 작으면 실험을 다시 수행하는 경우 블록화에 주의 또는 블록화 포기 ⇨ 완전확률화 설계 실시

$\frac{MSBL}{MSE} < 1$ ⇒ 블록의 변동이 MSE보다 작다, 효과가 미미하다

◉ Weight of Chickens

서로독립 ⇒ Cov=0 $Y_{\cdot 1}$

Block	Control	Low dose	High dose	합계	평균
1	3.93	3.99	3.96	11.88	3.960
2	3.78	3.96	3.94	11.68	3.893
3	3.88	3.96	4.02	11.86	3.953
4	3.93	4.03	4.06	12.02	4.007
5	3.84	4.10	3.94	11.88	3.960
6	3.75	4.02	4.09	11.86	3.953
7	3.98	4.06	4.17	12.21	4.070
8	3.84	3.92	4.12	11.88	3.960
합계	30.93	32.04	32.3	95.27	
평균	3.866	4.005	4.038		3.970

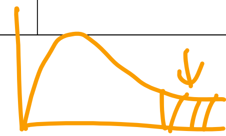
$Y_{1\cdot}$ $Y_{2\cdot}$ $Y_{n\cdot}$

- $TSS = \sum_{i=1}^{3}\sum_{j=1}^{8} y_{ij}^2 - \frac{y_{..}^2}{24} = 3.93^2 + \cdots + 4.12^2 - \frac{95.27^2}{24} = 0.2533$
- $SSA = \frac{30.93^2 + 32.04^2 + 32.3^2}{8} - \frac{95.27^2}{24} = 0.1324$
- $SSBL = \frac{11.88^2 + \cdots + 11.88^2}{3} - \frac{95.27^2}{24} = 0.0542$
- $SSE = TSS - SSA - SSBL = 0.2533 - 0.1324 - 0.0542 = 0.0667$

○ 분산분석표

변인	자유도	제곱합	평균제곱	F	p-값
촉진제 (처리)	a-1 2	SSA 0.1324	MSA 0.0662	13.889 ① F_a	0.0005
블록	b-1 7	SSBL 0.0542	MSBL 0.0077	1.626 ② F_β	0.2077
오차	(a-1)(b-1) 14	SSE 0.0667	MSE 0.0048		
전체	N-1 23	TSS 0.2533			

- 5% 유의수준에서 $F_{0.05}(2,14) = 3.739$ < 13.889



⇨ 성장촉진제 양에 따라 병아리 성장에 차이가 있음

처리

처리효과동일성검정

$\boxed{F_a > F(\alpha,\ a-1,\ (a-1)(b-1))}$ → H_0 기각

⊕ 블록효과동일성검정

$F_\beta = \frac{MSBL}{MSE} > 1$ 인경우 : 블록효과有, 완전설계이득.

$F_\beta < 1$ 인경우: 블록효과↓ ⇒ 완전확률화설계로다시!

$\boxed{F_b > 1}$ → H_0 기각

◉ 4가지 옥수수 품종(A, B, C, D)의 생산량을 비교하기 위해 4곳의 지역에서 파종하여 옥수수 생산량을 조사

확률화블록설계

처리: 옥수수품종 (A.B.C.D)
블록: 지역 (1.2.3.4)

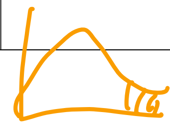
지역 1	지역 2	지역 3	지역 4
D	B	C	A
C	A	B	B
A	D	A	D
B	C	D	C

○ 실험결과

품종	지역 1	지역 2	지역 3	지역 4
A	9.3	9.4	9.6	10.0
B	9.4	9.3	9.8	9.9
C	9.2	9.4	9.5	9.7
D	9.7	9.6	10.0	10.2

○ 분산분석표 a=4, b=4

변인	자유도	제곱합	평균제곱	F
품종	a-1 3 4-1	0.385	0.1283	14.42 → F_α
블록	b-1 3 4-1	0.825	0.2750	
오차	(a-1)(b-1) 9 16-3-3	0.080	0.0089	
전체	N-1 15 16-1	1.290		



- 5% 유의수준에서 $F_{0.05}(3,9) = 3.86$ < 14.42

⇨ 옥수수 품종에 따라 옥수수 생산량에 차이가 있음

블록고려X

○ 만약 이 실험을 완전 확률화 설계법으로 생각하고 분석을 했다면

변인	자유도	제곱합	평균제곱	F
품종	3	0.385	0.1283	1.70
오차	12	0.905	0.0754	
전체	15	1.290		

- 5% 유의수준에서 품종에 따라 옥수수 생산량에 차이가 있다고 할 수 없음 ⇨ 앞에서 블록에 의해 설명되는 변동이 모두 오차의 변동으로 포함됨

$F_{0.05}(3, 12) = 3.45$

주의

$F(\alpha, p-1, N-p)$

○ 처리효과에 대한 다중비교

- $H_0 : \mu_{i.} = \mu_{k.}$ vs $H_1 : \mu_{i.} \neq \mu_{k.}$ 또는 $\mu_{i.} - \mu_{k.}$의 신뢰구간
- $\overline{Y}_{i.} - \overline{Y}_{k.} \pm c\sqrt{MSE}\sqrt{2/b}$
  - 최소유의차: $c = t_{\alpha/2,(a-1)(b-1)}$

    chap02의 N-p
  - Bonferroni: $c = t_{\alpha/(2k),(a-1)(b-1)}$, $k =$ 비교검정의 경우의 수
  - Scheffe: $c = \sqrt{(a-1)F_{\alpha,a-1,(a-1)(b-1)}}$

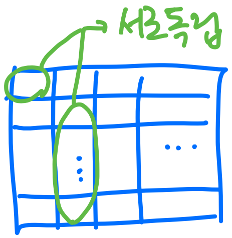
    chap02의 p-1
  - Tukey: $\frac{1}{\sqrt{2}} q_{\alpha,a,(a-1)(b-1)}$

$$Var(\overline{Y}_{i.} - \overline{Y}_{k.}) = Var\left(\frac{\sum_{j=1}^{b} y_{ij}}{b} - \frac{\sum_{j=1}^{b} y_{kj}}{b}\right)$$

$$= Var\left(\frac{y_{i1} + \cdots + y_{ib}}{b} - \frac{y_{k1} + \cdots + y_{kb}}{b}\right)$$

$$= \frac{b\sigma^2}{b^2} + \frac{b\sigma^2}{b^2} - 2Cov\left(\frac{y_{i1} + \cdots + y_{ib}}{b}, \frac{y_{k1} + \cdots + y_{kb}}{b}\right) = \frac{2\sigma^2}{b}$$

$$Cov(\mu + \alpha_i + \beta_j + \varepsilon_{ij}, \mu + \alpha_k + \beta_k + \varepsilon_{kj})$$
$$= Cov(\varepsilon_{ij}, \varepsilon_{kj})$$
$$= 0$$



10/16

review) chap02 완전확률화설계 다중비교

다중비교
- 전체처리비교
- 처리-대조비교

1. F검정기반 전체처리비교

① Fisher(LSD)

$|\bar{y}_{i.} - \bar{y}_{j.}| > t(\frac{\alpha}{2}, N-p)\sqrt{MSE}\sqrt{\frac{1}{n}+\frac{1}{n}}$

② Bonferroni

$|\bar{y}_{i.} - \bar{y}_{j.}| > t(\frac{\alpha}{2k}, N-p)$, 이때 $k = {}_pC_2$

③ Scheffe

$|\bar{y}_{i.} - \bar{y}_{j.}| > \sqrt{(p-1)F(\alpha, p-1, N-p)}$

2. 평균범위기반 전체처리비교

① Tukey(HSD)

$|\bar{y}_{i.} - \bar{y}_{j.}| > \frac{q(\alpha, p, N-p)}{\sqrt{2}} \cdot$

② Duncan

3. 처리대조비교

Dunnett

$|\bar{y}_c - \bar{y}_i| > t'(\alpha, p-1, N-p)$