■ 다원 배치법(multi-factor design)

- 관심의 요인이 3개 이상인 경우, 모든 요인의 수준조합에 대해 확률화를 적용하여 실험
- 요인의 수가 늘어나면, 실험횟수가 많아지고 이에 대해 랜덤화가 어려워짐
- 실험전체를 비슷한 관리 상태 하에서 수행하는데 여러 가지 어려움이 따름

  ⇨ 요인에 대한 충분한 기술적 검토를 거쳐 불필요한 요인라고 판단되면 과감히 요인의 수를 줄임

뒤에서 자세설명

□ 반복이 없는 삼원배치법 (고정효과모형)

- 요인 A, B, C가 각각 a, b, c 개의 수준을 가짐
- abc개의 모든 수준 조합에 대해 확률화를 적용하여 배치

○ 자료의 구조

		A_1	A_2	$\cdots$	A_a
B_1	C_1	Y_{111}	Y_{211}	$\cdots$	Y_{a11}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	C_c	Y_{11c}	Y_{21c}	$\cdots$	Y_{a1c}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_b	C_1	Y_{1b1}	Y_{2b1}	$\cdots$	Y_{ab1}
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	C_c	Y_{1bc}	Y_{2bc}	$\cdots$	Y_{abc}

○ 모형의 구조식

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}$$

반복이있는/없는 이원배치 차이점

- μ: 전체평균
- $\alpha_i, \beta_j, \gamma_k$: 요인의 주효과

→ 반복이있는경우 상호작용을 빼낼수있음.

- $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$: 두 요인의 상호작용
- $\varepsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- 3 요인의 상호작용 $(\alpha\beta\gamma)_{ijk}$는 오차항 ε_{ijk}에 교락되어 있어 별도로 검정할 수 없음

요인 C를 무시하면 각 조합별로 반복이 있는것으로 보임 (C의 개수만큼)

↳ 반복이 있다고 할 수 없음

○ 변동 분해

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

○ $TSS=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}Y_{ijk}^2-CT,\quad CT=\dfrac{Y_{...}^2}{abc}$ 원래정의는 $\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{...})^2$

○ $SSA=\dfrac{1}{bc}\sum_{i=1}^{a}Y_{i..}^2-CT,\ SSB=\dfrac{1}{ac}\sum_{j=1}^{b}Y_{.j.}^2-CT,\ SSC=\dfrac{1}{ab}\sum_{k=1}^{c}Y_{..k}^2-CT$

○ $SSAB=\dfrac{1}{c}\sum_{i=1}^{a}\sum_{j=1}^{b}Y_{ij.}^2-CT,\ SSAC=\dfrac{1}{b}\sum_{i=1}^{a}\sum_{k=1}^{c}Y_{i.k}^2-CT,$

$SSBC=\dfrac{1}{a}\sum_{j=1}^{b}\sum_{k=1}^{c}Y_{.jk}^2-CT$

$$\left\{\begin{array}{l}SSA=\sum\sum\sum(\overline{Y}_{i..}-\overline{Y}_{...})^2\\ SSB=\sum\sum\sum(\overline{Y}_{.j.}-\overline{Y}_{...})^2\\ SSC=\sum\sum\sum(\overline{Y}_{..k}-\overline{Y}_{...})^2\\ SSAB=\sum\sum\sum(\overline{Y}_{ij.}-\overline{Y}_{...})^2\\ SSAC=\sum\sum\sum(\overline{Y}_{i.k}-\overline{Y}_{...})^2\rightarrow ac-1\\ SSBC=\sum\sum\sum(\overline{Y}_{.jk}-\overline{Y}_{...})^2\end{array}\right.$$

공식이 없으므로 ( )는 전체에서 빼기.

- $SS(AB) = SSAB - SSA - SSB$, $SS(AC) = SSAC - SSA - SSC$, $SS(BC) = SSBC - SSB - SSC$
- $SSE = TSS - (SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC))$

① $H_0: \alpha_1 = \alpha_2 = \cdots \alpha_a = 0$
② $H_0: (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0$

- 분산분석표

변인		자유도	제곱합	평균제곱	F
주효과	A	a-1	SSA	MSA	MSA/MSE ①
	B	b-1	SSB	MSB	MSB/MSE
	C	c-1	SSC	MSC	MSC/MSE
상호작용	(AB)	(a-1)(b-1)	SS(AB)	MS(AB)	MS(AB)/MSE ②
	(AC)	(a-1)(c-1)	SS(AC)	MS(AC)	MS(AC)/MSE
	(BC)	(b-1)(c-1)	SS(BC)	MS(BC)	MS(BC)/MSE
오차		(a-1)(b-1)(c-1)	SSE	MSE	
전체		abc-1	TSS		

▣ 화학공장의 합성반응공정에서 합성율의 향상

abc-1 - (a-1 + b-1 + c-1 + (a-1)(b-1) + (a-1)(c-1) + (b-1)(c-1))

- 반응압력 : 8, 10, 12 (kg/cm^2)
- 반응시간 : 1.5, 2.0, 2.5 (hr)
- 반응온도 : 140, 150, 160 (^oC)

		A_1	A_2	A_3
B_1	C_1	74	61	50
	C_2	86	78	70
	C_3	76	71	60
B_2	C_1	72	62	49
	C_2	91	81	68
	C_3	87	77	64
B_3	C_1	48	55	52
	C_2	65	72	69
	C_3	56	63	60

삼원배치 데이터를 이원배치로.
(채울수 있어야됨)

	A_1	A_2	A_3
B_1	236	210	180
B_2	250	220	181
B_3	169	190	181

542 = $y_{3..}$

	A_1	A_2	A_3
C_1	194	178	151
C_2	242	231	207
C_3	219	211	184

614 = $y_{..3}$

A가 무시됨. → $y_{.21}$
72+62+49

	B_1	B_2	B_3
C_1	185	183	155
C_2	234	240	206
C_3	207	228	179

626 = $y_{.1.}$

C가 무시됨. → $y_{13.}$
48+65+56

$a=b=c=3$

- $CT = 1817^2/27 = 122277.37$
- $TSS = 74^2 + \cdots + 60^2 - CT = 3613.6$
- $SSA = \frac{1}{9}(655^2 + 620^2 + 542^2) - CT = 743.6$

  $bc \rightarrow 9$
- $SSB = \frac{1}{9}(626^2 + 651^2 + 540^2) - CT = 753.4$
- $SSC = \frac{1}{9}\underline{(523^2 + 680^2 + 614^2)} - CT = 1380.9$

  $\sum y_{..k}^2$
- $SSAB = \frac{1}{3}(236^2 + \cdots + 181^2) - CT = 2148.9$
- $SSAC = \frac{1}{3}(194^2 + \cdots + 184^2) - CT = 2133.6$
- $SSBC = \frac{1}{3}\underline{(185^2 + \cdots + 179^2)} - CT = 2190.9$

- $SS(AB) = SSAB - SSA - SSB = 651.9$
- $SS(AC) = SSAC - SSA - SSC = 9.1$
- $SS(BC) = SSBC - SSB - SSC = 56.6$
- $SSE = TSS - (SSAB + SSAC + SSBC - SSA - SSB - SSC^{*}) = 18.1$
- 분산분석표

변인	자유도	제곱합	평균제곱	F
A	a-1 2	743.6	371.8	164.5
B	2	753.4	376.7	166.7
C	2	1380.9	690.4	305.5
(AB)	(a-1)(b-1) 4	651.9	163.0	72.1
(AC)	4	9.1	2.3	기각 X 1.0
(BC)	4	56.6	14.2	6.3
오차	(a-1)(b-1)(c-1) 8	18.1	2.26	
전체	abc-1 26	3613.6		

모두 MSE로 나눔.

$F_{0.01,2,8}$, $F_{0.01,4,8}$ 과 비교

α에 따라 달라지는 결과

	α=0.05	α=0.01
2,8	4.46	8.65
4,8	3.84	7.01

$* = \cancel{SSA} + \cancel{SSB} + SS(AB) + SSA + \cancel{SSC} + SS(AC) + SSB + SSC + SS(BC) - \cancel{SSA} - \cancel{SSB} - \cancel{SSC} = SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC)$

$F_{0.10\,(4,8)} = 2.81$ 에서도 기각불가능

○ 분산분석표상에서 (AC)는 유의수준 $\alpha = 0.10$에서 기각시키지 못하기 때문에 오차항에 포함시켜 재작성

굳이 모델에 넣을 필요 X 심플하게.

변인	자유도	제곱합	평균제곱	F
A	2	743.6	371.8	163.8
B	2	753.4	376.7	165.7
C	2	1380.9	690.4	304.1
(AB)	4	651.9	163.0	71.8
(BC)	4	56.6	14.2	6.3
오차	12	27.2	2.27	
전체	26	3613.6		

MSE가 업데이트됨.

update!

	$\alpha=0.05$	$\alpha=0.01$
2,12	3.89	6.93
4,12	3.26	5.41

모든 F가 유의함.

○ 분산 분석후 추정

아까한거

- ○ 일차적으로 분산분석표에 의한 F-검정이 끝나면, 유의하지 않은 상호작용은 오차항에 흡수시켜 다시 F-검정을 실시

- 주효과만 유의한 경우 → ab, bc, ca 모두유의하지 않는경우 a, b, c만 남음
  - ○ 각 요인수준에서의 모평균 추정
    - 점추정 : $\hat{\mu}(A_i) = \overline{Y}_{i..}$
    - 구간추정 : $\overline{Y}_{i..} \pm t_{\alpha/2,\nu}\sqrt{MSE^*}/\sqrt{bc}$

$Var(\overline{Y}_{i..}) = Var\left(\frac{\sum_j\sum_k Y_{ijk}}{bc}\right) = \frac{bc\sigma^2}{b^2c^2} = \frac{\sigma^2}{bc}$

*글붙인건 최종모형의 MSE라는뜻.

  - ○ 수준조합에 대한 모평균 추정
    - 점추정 : $\hat{\mu}(A_iB_jC_k) = \overline{Y}_{i..} + \overline{Y}_{.j.} + \overline{Y}_{..k} - 2\overline{Y}_{...}$
    - 구간추정 : $\overline{Y}_{i..} + \overline{Y}_{.j.} + \overline{Y}_{..k} - 2\overline{Y}_{...} \pm t_{\alpha/2,\nu}\sqrt{MSE^*}/\sqrt{n_e}$
    - $\frac{1}{n_e} = \frac{1}{bc} + \frac{1}{ac} + \frac{1}{ab} - \frac{2}{abc} \rightarrow n_e = \frac{abc}{a+b+c-2}$

incentive number
: 실질적인 ne는 n보다 대부분작음
중복되는정보가 제외됨

Var 구하기 어려움
독립이 아니므로 Cov ≠ 0, 계산해야함.

역수로 먼저외우면쉬움!

$\hat{\mu}(A_iB_jC_k) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k = \underbrace{(\hat{\mu} + \hat{\alpha}_i)}_{\overline{Y}_{i..}} + \underbrace{(\hat{\mu} + \hat{\beta}_j)}_{\overline{Y}_{.j.}} + \underbrace{(\hat{\mu} + \hat{\gamma}_k)}_{\overline{Y}_{..k}} - 2\underbrace{\hat{\mu}}_{\overline{Y}_{...}}$

- 주효과와 일부 상호작용만 유의한 경우

(예) $A, B, C, (AC)$ 만 유의하다면,

- ○ 수준조합에 대한 모평균 추정
  - 점추정 : $\hat{\mu}(A_i B_j C_k) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\gamma)}_{ik}$

$$= \hat{\mu} + \hat{\alpha}_i + \hat{\gamma}_k + \widehat{(\alpha\gamma)}_{ik} + \hat{\beta}_j = \overline{Y}_{i.k} + \overline{Y}_{.j.} - \overline{Y}_{...}$$

  - 구간추정 : $\overline{Y}_{i.k} + \overline{Y}_{.j.} - \overline{Y}_{...} \pm t_{\alpha/2,\nu} \sqrt{MSE^*} / \sqrt{n_e}$
  - $\dfrac{1}{n_e} = \dfrac{1}{b} + \dfrac{1}{ac} - \dfrac{1}{abc} \rightarrow n_e = \dfrac{abc}{ac + b - 1}$

- 모든 요인이 유의한 경우 (더이상 줄일게 없음)
  - 수준조합에 대한 모평균 추정
    - 점추정 : $\hat{\mu}(A_iB_jC_k) = \overline{Y}_{ij.} + \overline{Y}_{i.k} + \overline{Y}_{.jk} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + \overline{Y}_{...}$

      $= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk}$
    - 구간추정 : $\hat{\mu}(A_iB_jC_k) \pm t_{\alpha/2,\nu}\sqrt{MSE^*}\,/\,\sqrt{n_e}$
    - $n_e = \dfrac{abc}{ab+ac+bc-a-b-c+1}$

$$\frac{1}{n_e} = \frac{1}{c} + \frac{1}{b} + \frac{1}{a} - \frac{1}{bc} - \frac{1}{ac} - \frac{1}{ab} + \frac{1}{abc}$$

$$\left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij}\right) + \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\gamma)}_{ik}\right) + \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\beta\gamma)}_{jk}\right)$$

$$- \left(\hat{\mu} + \hat{\alpha}_i\right) - \left(\hat{\mu} + \hat{\beta}_j\right) - \left(\hat{\mu} + \hat{\gamma}_k\right) + \hat{\mu}$$

예제에서는 (AC) 하나만 유의하지 않았음.

▣ 수준조합 $A_1B_2C_2$의 모평균의 점추정값과 95% 신뢰구간

○ $\hat{\mu}(A_1B_2C_2) = \bar{y}_{12.} + \bar{y}_{.22} - \bar{y}_{.2.} = \frac{250}{3} + \frac{240}{3} - \frac{651}{9} = 91$

○ $91 \pm t_{0.025}(12)\sqrt{2.27/1.8} = 91 \pm 2.179 \times 1.123 = 91 \pm 2.4$ ⇨ (88.6%, 93.4%)

→ 최종자유도

○ $\frac{1}{n_e} = \frac{1}{3} + \frac{1}{3} - \frac{1}{9} \rightarrow n_e = 1.8$

$\left(\frac{1}{c} + \frac{1}{a} - \frac{1}{ac}\right)$

$\hat{\mu}(A_iB_jC_k) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + (\widehat{\alpha\beta})_{ij} + (\widehat{\beta\gamma})_{jk}$

$\rightarrow \left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\widehat{\alpha\beta})_{ij}\right) + \left(\hat{\mu} + \hat{\beta}_j + \hat{\gamma}_k + (\widehat{\beta\gamma})_{jk}\right) - \left(\hat{\mu} + \hat{\beta}_j\right)$

$= \bar{Y}_{ij.} + \bar{Y}_{.jk} - \bar{Y}_{.j.}$

□ 반복이 있는 삼원배치법 (고정효과모형)

- ○ 요인 A, B, C가 각각 a, b, c 개의 수준을 가짐
- ○ 반복수가 r일 때 $N = abcr$개의 모든 수준 조합에 대해 확률화를 적용하여 배치

○ 자료의 구조

		A_1	$\cdots$	A_a
B_1	C_1 $\vdots$ C_c	$Y_{1111} \cdots Y_{111r}$ $\vdots$ $Y_{11c1} \cdots Y_{11cr}$	$\cdots$ $\ddots$ $\cdots$	$Y_{a111} \cdots Y_{a11r}$ $\vdots$ $Y_{a1c1} \cdots Y_{a1cr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
B_b	C_1 $\vdots$ C_c	$Y_{1b11} \cdots Y_{1b1r}$ $\vdots$ $Y_{1bc1} \cdots Y_{1bcr}$	$\cdots$ $\ddots$ $\cdots$	$Y_{ab11} \cdots Y_{ab1r}$ $\vdots$ $Y_{abc1} \cdots Y_{abcr}$

○ 모형의 구조식

반복이있으면상호작용이있음

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \underline{(\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

- μ: 전체평균
- $\alpha_i, \beta_j, \gamma_k$: 요인의 주효과
- $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$: 두 요인의 상호작용
- $(\alpha\beta\gamma)_{ijk}$: 세 요인의 상호작용
- $\varepsilon_{ijk} \sim$ iid $N(0, \sigma^2)$

○ 변동 분해

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

$$+(\overline{Y}_{i.k.}-\overline{Y}_{i...}-\overline{Y}_{..k.}+\overline{Y}_{....})+(\overline{Y}_{.jk.}-\overline{Y}_{.j..}-\overline{Y}_{..k.}+\overline{Y}_{....})$$

$$+(Y_{ijk.}-\overline{Y}_{ij..}-\overline{Y}_{i.k.}-\overline{Y}_{.jk.}+\overline{Y}_{i...}+\overline{Y}_{.j..}+\overline{Y}_{..k.}-\overline{Y}_{....})+e_{ijkl}$$

○ $TSS=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{r}Y_{ijkl}^2-CT$, $CT=\frac{Y_{....}^2}{N}$

○ $SSA=\frac{1}{bcr}\sum_{i=1}^{a}Y_{i...}^2-CT$, $SSB=\frac{1}{acr}\sum_{j=1}^{b}Y_{.j..}^2-CT$,

$SSC=\frac{1}{abr}\sum_{k=1}^{c}Y_{..k.}^2-CT$

○ $SSAB=\frac{1}{cr}\sum_{i=1}^{a}\sum_{j=1}^{b}Y_{ij..}^2-CT$, $SSAC=\frac{1}{br}\sum_{i=1}^{a}\sum_{k=1}^{c}Y_{i.k.}^2-CT$,

$$SSBC = \frac{1}{ar}\sum_{j=1}^{b}\sum_{k=1}^{c} Y_{.jk.}^2 - CT$$

- $$SSABC = \frac{1}{r}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} Y_{ijk.}^2 - CT$$
- $SS(AB) = SSAB - SSA - SSB$, $SS(AC) = SSAC - SSA - SSC$,

  $SS(BC) = SSBC - SSB - SSC$
- $$\begin{aligned} SS(ABC) &= SSABC - (SSA + SSB + SSC + SS(AB) + SS(AC) + SS(BC)) \\ &= SSABC - (SSAB + SSAC + SSBC - SSA - SSB - SSC) \end{aligned}$$
- $SSE = TSS - SSABC$

분산분석표

변인		자유도	제곱합	평균제곱	F
주효과	A	a-1	SSA	MSA	MSA/MSE
	B	b-1	SSB	MSB	MSB/MSE
	C	c-1	SSC	MSC	MSC/MSE
상호 작용	(AB)	(a-1)(b-1)	SS(AB)	MS(AB)	MS(AB)/MSE
	(AC)	(a-1)(c-1)	SS(AC)	MS(AC)	MS(AC)/MSE
	(BC)	(b-1)(c-1)	SS(BC)	MS(BC)	MS(BC)/MSE
	(ABC)	(a-1)(b-1)(c-1)	SS(ABC)	MS(ABC)	MS(ABC)/MSE
오차		abc(r-1)	SSE	MSE	
전체		abcr-1	TSS		

전체자유도 - 앞에있는거 전부

$$= (abcr-1) - ((a-1)+(b-1)+(c-1)+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1)+(a-1)(b-1)(c-1))$$

$$= abc(r-1)$$