11. 2수준 및 3수준계 요인설계

개요

요인의 수가 n이고 i번째 요인의 수준수가 k_i인 요인설계: $k_1 \times k_2 \times \cdots \times k_n$

$k_1 = k_2 = \cdots = k_n = k$이면 $k \times k \times \cdots \times k = k^n$으로 표기함.

k^n 요인설계의 처리조합수

k \ n	2	3	4	5	6	7
2	4	8	16	32	64	128
3	9	27	81	243	279	2187
4	16	64	256	1024	4096	16384
5	25	125	625	3125	15625	78125

k가 2 또는 3인 경우가 비용이나 시간 등의 제약을 고려할 때 가장 현실적인 선택임.

대비와 직교분해

a개의 처리조합의 평균 $\mu_1, \cdots, \mu_a$들을 비교

예) $a = 4$ 처리 1: 세제 A의 액체 제품
처리 2: 세제 A의 분말 제품
처리 3: 세제 B의 액체 제품
처리 4: 세제 B의 분말 제품

가설: 세제 A의 평균 = 세제 B의 평균
액체 세제 A와 B간의 평균 차이 = 분말세제 A와 B간의 평균 차이
액체 세제의 평균 = 분말 세제의 평균

이를 수식으로 표현하면

$$\mu_1 + \mu_2 = \mu_3 + \mu_4 \ (\text{즉, } \mu_1 + \mu_2 - \mu_3 - \mu_4 = 0)$$

$$\mu_1 - \mu_3 = \mu_2 - \mu_4 \ (\text{즉, } \mu_1 - \mu_2 - \mu_3 + \mu_4 = 0)$$

$$\mu_1 + \mu_3 = \mu_2 + \mu_4 \ (\text{즉, } \mu_1 - \mu_2 + \mu_3 - \mu_4 = 0)$$

이 되고, 이 식들을 연립해서 풀면 $\mu_1 = \mu_2 = \mu_3 = \mu_4$를 얻는다. 각 처리조합에서의 측정값의 합을 T_1, T_2, T_3, T_4라 하면 위의 세 가설들은 각각 다음의 통계량의 값을 사용하여 검정할 수 있다.

$$L_1 = T_1 + T_2 - T_3 - T_4,\ L_2 = T_1 - T_2 - T_3 + T_4,\ L_3 = T_1 - T_2 + T_3 - T_4$$

일반적으로 a개의 처리조합에서의 측정값의 합을 $T_1, \cdots, T_a$라 하면, $\sum_{i=1}^{a} c_i = 0$인 $c_1, \cdots, c_a$에 대하여

$$L = \sum_{i=1}^{a} c_i T_i$$

를 대비(contrast)라고 부른다. 특히 두 개의 대비 $L_1 = \sum_{i=1}^{a} c_i T_i$와 $L_2 = \sum_{i=1}^{a} d_i T_i$가

$$\sum_{i=1}^{a} c_i d_i = 0$$

인 조건을 만족시키면 직교대비라고 부른다.

(일반적으로 $\sum_i c_i \mu_i = 0$을 검정하기 위한 통계량과 기각역은

$$\frac{|\sum_i c_i T_i|}{\sqrt{r \cdot MS_E \sum_i c_i^2}} > t_{ar-a,\, \alpha/2} \Leftrightarrow \frac{(\sum_i c_i T_i)^2 / (r \sum_i c_i^2)}{MS_E} > F_{1,\, ar-a,\, \alpha}$$

이므로 $L = \sum_{i=1}^{a} c_i T_i$의 제곱합을

$$SS_L = \frac{L^2}{r \sum_{i=1}^{a} c_i^2}$$

으로 정의하면 가설 $\sum_i c_i \mu_i = 0$은 $SS_L / MS_E > F_{1,\, ar-a,\, \alpha}$일 때 기각된다.)

위에서 예로 든 세 개의 대비 L_1, L_2, L_3는 직교대비들로서 자유도 3인 처리제곱합 SS_{Trt}를 각각 자유도 1인 3개의 대비제곱합들로 분해한다. 즉,

$$\begin{aligned}
& SS_{L_1} + SS_{L_2} + SS_{L_3} \\
&= \frac{L_1^2}{4r} + \frac{L_2^2}{4r} + \frac{L_3^2}{4r} \\
&= \frac{1}{4r}[3(T_1^2 + T_2^2 + T_3^2 + T_4^2) + 2(T_1T_2 - T_1T_3 - T_1T_4 - T_2T_3 - T_2T_4 + T_3T_4) \\
&\qquad + 2(-T_1T_2 - T_1T_3 + T_1T_4 + T_2T_3 - T_2T_4 - T_3T_4) \\
&\qquad + 2(-T_1T_2 + T_1T_3 - T_1T_4 - T_2T_3 + T_2T_4 - T_3T_4)] \\
&= \frac{1}{4r}[3(T_1^2 + T_2^2 + T_3^2 + T_4^2) - 2(T_1T_2 + T_1T_3 + T_1T_4 + T_2T_3 + T_2T_4 + T_3T_4)] \\
&= \frac{1}{4r}[4(T_1^2 + T_2^2 + T_3^2 + T_4^2) - (T_1 + T_2 + T_3 + T_4)^2] \\
&= \frac{1}{r}\sum_i T_i^2 - \frac{1}{4r}(\sum_i T_i)^2 = SS_{Trt}
\end{aligned}$$

이 성립함을 알 수 있다. 일반적으로 수준의 수가 a인 요인 A의 변동 SS_A는 각각 자유도가 1인 $(a-1)$개의 직교대비 $L_1, \cdots, L_{a-1}$에 의한 변동들로 분해할 수 있다.

(예 11.1) 3^2 요인설계 A: 제조회사 (A_0, A_1, A_2) B: 성형온도 (B_0, B_1, B_2)

1) A와 B는 유의한가?

2) A가 유의하다면 국산과 외제의 차이 때문인가? 혹은 자회사와 국내 타회사의 차이 때문인가?

3) B가 유의하다면 그 효과가 1차적인가 혹은 2차적인가?

L_1 = 국산과 외제의 차이 = $\frac{1}{6}(T_{0.} + T_{1.}) - \frac{1}{3}T_{2.} = -0.5$

L_2 = 자회사와 국내 타회사의 차이 = $\frac{1}{3}T_{0.} - \frac{1}{3}T_{1.} = 11$

L_1과 L_2는 직교대비로서

$$SS_{L_1} = \frac{L_1^2}{r\sum_i c_i^2} = \frac{(-0.5)^2}{3\left(\frac{1}{36} + \frac{1}{36} + \frac{1}{9}\right)} = 0.5$$

$$SS_{L_2} = \frac{(11)^2}{3\left(\frac{1}{9} + \frac{1}{9}\right)} = 181.5$$

$\therefore\ SS_A = SS_{L_1} + SS_{L_2} = 182$

L_l = B의 1차 효과 = $(T_{.1} - T_{.0}) + (T_{.2} - T_{.1}) = T_{.2} - T_{.0} = 39$

L_q = B의 2차 효과 = $(T_{.2} - T_{.1}) - (T_{.1} - T_{.0}) = T_{.0} - 2T_{.1} + T_{.2} = -3$

L_l과 L_q는 직교대비로서

$$SS_l = \frac{L_l^2}{r\sum_i c_i^2} = \frac{(39)^2}{3(1+1)} = 253.5$$

$$SS_q = \frac{(-3)^2}{3(1+4+1)} = 0.5$$

$\therefore\ SS_B = SS_L + SS_q = 254$ (p. 266 분산분석표 오류)

2^2 요인설계

A 요인의 효과(A) ≡ A_1 수준에서의 평균 − A_0 수준에서의 평균

$= \frac{1}{2r}(ab + a) - \frac{1}{2r}(b + (1))$ (r: 반복수)

$= \frac{1}{2r}(ab + a - b - (1))$

B 요인의 효과(B) ≡ B_1 수준에서의 평균 − B_0 수준에서의 평균

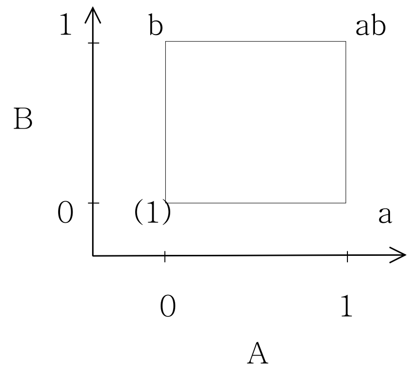
$= \frac{1}{2r}(ab + b) - \frac{1}{2r}(a + (1))$

$= \frac{1}{2r}(ab + b - a - (1))$

A×B 상호작용(AB) ≡ (A_1 수준에서의 B의 효과 − A_0 수준에서의 B의 효과)/2

$$= \frac{1}{2}\left[\frac{1}{r}(ab-a)-\frac{1}{r}(b-(1))\right]$$

$$= \frac{1}{2r}(ab+(1)-a-b)$$



	처리조합	A	B	AB	
$T_{00.}$	(1)	−1	−1	+1	μ_1
$T_{10.}$	a	+1	−1	−1	μ_2
$T_{01.}$	b	−1	+1	−1	μ_3
$T_{11.}$	ab	+1	+1	+1	μ_4

(A열 × B열 = AB열)

$$\left.\begin{aligned} &1\text{열: } -\mu_1+\mu_2-\mu_3+\mu_4=0 \\ &2\text{열: } -\mu_1-\mu_2+\mu_3+\mu_4=0 \\ &3\text{열: } \ \ \mu_1-\mu_2-\mu_3+\mu_4=0 \end{aligned}\right\} \Rightarrow \mu_1=\mu_2=\mu_3=\mu_4$$

$L_A = ab+a-b-(1)$, $L_B = ab+b-a-(1)$, $L_{AB} = ab+(1)-a-b$는 직교대비이므로 그 제곱합은

$$SS_A = \frac{L_A^2}{r\sum_i c_i^2} = \frac{1}{4r}(ab+a-b-(1))^2 = r(A)^2$$

$$SS_B = \frac{L_B^2}{r\sum_i c_i^2} = \frac{1}{4r}(ab+b-a-(1))^2 = r(B)^2$$

$$SS_{A\times B} = \frac{L_{AB}^2}{r\sum_i c_i^2} = \frac{1}{4r}(ab+(1)-a-b)^2 = r(AB)^2$$

이고

$$SS_T = \sum_i\sum_j\sum_k x_{ijk}^2 - \frac{T^2}{4r} = SS_{Model} + SS_E = SS_A + SS_B + SS_{A\times B} + SS_E$$

↖ A_iB_j수준에서의 k번째 관측값

(단, $T = \sum_i \sum_j \sum_k x_{ijk}$)

가 된다. (표 11.3) ($r=1$이면 SS_E의 자유도는 0임에 유의할 것!)

(예 11.2)

A	B	AB	
-1	-1	+1	10
+1	-1	-1	0
-1	+1	-1	10
+1	+1	+1	-10

$$A = \frac{1}{4}(-10+0-10-10) = -7.5$$

$$B = \frac{1}{4}(-10-0+10-10) = -2.5$$

$$AB = \frac{1}{4}(10-0-10-10) = -2.5$$

$$SS_A = 2(-7.5)^2 = 112.5$$

$$SS_B = 2(-2.5)^2 = 12.5$$

$$SS_{A\times B} = 2(-2.5)^2 = 12.5$$

$$SS_T = \sum_i \sum_j \sum_k x_{ijk}^2 - \frac{T^2}{8} = 157.5$$ (p.271 분산분석표)

<u>2^3 요인설계</u>

A = A_1 수준에서의 평균 - A_0 수준에서의 평균

$$= \frac{1}{4r}(a+ac+ab+abc) - \frac{1}{4r}((1)+c+b+bc)$$

마찬가지 방법으로

$$B = \frac{1}{4r}(b+bc+ab+abc) - \frac{1}{4r}((1)+a+c+ac)$$

$$C = \frac{1}{4r}(c+ac+bc+abc) - \frac{1}{4r}((1)+a+b+ab)$$

$$AB = \frac{1}{2}(C_0 \text{ 수준에서의 } AB + C_1 \text{ 수준에서의 } AB)$$

$$= \frac{1}{2}\left[\frac{1}{2r}(ab+(1)-a-b) + \frac{1}{2r}(abc+c-ac-bc)\right]$$

$$= \frac{1}{4r}((1)+c+ab+abc) - \frac{1}{4r}(a+b+ac+bc)$$

마찬가지 방법으로

$$AC = \frac{1}{4r}((1)+b+ac+abc) - \frac{1}{4r}(a+c+ab+bc)$$

$$BC = \frac{1}{4r}((1)+a+bc+abc) - \frac{1}{4r}(b+c+ab+ac)$$

$$ABC = \frac{1}{2}(C_1 \text{ 수준에서의 } AB - C_0 \text{ 수준에서의 } AB)$$

$$= \frac{1}{2}\left[\frac{1}{2r}(abc+c-ac-bc) - \frac{1}{2r}(ab+(1)-a-b)\right]$$

$$= \frac{1}{4r}(a+b+c+abc) - \frac{1}{4r}((1)+ab+ac+bc)$$

계수표: 표 11.5

각 요인의 제곱합 $= \frac{(\text{대비})^2}{8r} = 2r(\text{요인효과})^2$

(예컨대 $L_A = a+ab+ac+abc-(1)-b-c-bc$)

$$SS_T = \sum_i \sum_j \sum_k \sum_m x_{ijkm}^2 - \frac{T^2}{8r}$$

$$= SS_A + SS_B + SS_C + SS_{A\times B} + SS_{A\times C} + SS_{B\times C} + SS_{A\times B\times C} + SS_E$$

분산분석표: 표 11.6

(예 11.3) $A = \frac{1}{8}(4+1+1+5+1+3-2+11) = 3$

$SS_A = 4(3)^2 = 36 \ \cdots$

$SS_T = 78$

분산분석표: p.276

주효과 그림: 그림 11.3

최적 조건: $A_0B_0C_1$, $A_1B_0C_0$, $A_0B_1C_0$ $\rightarrow$ 제조원가와 생산량을 고려할 때 $A_0B_0C_1$이 바람직함. (표 11.7)

Note: 1) 일반적으로 2^n 요인설계에 있어 요인효과의 계산은 계수표를 이용하여 간단히 할 수 있다.

요인효과(effect) = $\frac{\text{대비}}{2^{n-1}\cdot r}$

요인제곱합(SS_{effect}) = $\frac{(\text{대비})^2}{2^n \cdot r} = 2^{n-2}\cdot r\,(\text{요인효과})^2$

이때 요인효과의 분산은 $N = 2^n \cdot r$이라 할 때

$$Var(effect) = Var(\bar{x}_+ - \bar{x}_-) = Var(\bar{x}_+) + Var(\bar{x}_-)$$

$$= \frac{\sigma^2}{N/2} + \frac{\sigma^2}{N/2} = \frac{4}{N}\sigma^2 = \frac{\sigma^2}{2^{n-2} \cdot r}$$

$$\Rightarrow \widehat{s.e.}\,(\text{effect}) = \sqrt{\frac{MS_E}{2^{n-2} \cdot r}}$$

$$\therefore\ \frac{SS_{effect}}{MS_E} = \frac{2^{n-2} \cdot r(\text{effect})^2}{MS_E} = \left[\frac{\text{effect}}{\widehat{s.e.}(\text{effect})}\right]^2 \sim^{H_0} F_{1,\, 2^n(r-1)}$$

⇒ 분산분석에 의한 요인효과의 검증과 t 검증은 동일함!

2) 만약 $r=1$이면 SS_E의 자유도가 0이 되므로 요인효과의 검증을 위해서는 고차의 상호작용 (예를 들어 2^4 요인설계의 경우 3차 이상의 상호작용 ABC, ABD, ACD, BCD, ABCD)에 대한 제곱합을 더해서 이를 SS_E로 간주하여 검증하거나 요인효과들에 대한 정규점수그림을 그려 유의한 요인효과를 찾아낸다.

3) 모든 요인효과를 한 번에 효율적으로 계산하기 위해서 계수표 대신에 Yates의 알고리즘을 사용할 수 있다. 예 11.3의 경우에 Yates의 알고리즘을 적용하면 다음과 같다.

A	B	C	처리합	(1)	(2)	(3)	분모	효과	
-	-	-	-4	-3	1	16	8r	1.00	평균
+	-	-	1	4	15	24	4r	3.00	A
-	+	-	-1	2	11	18	4r	2.25	B
+	+	-	5	13	13	6	4r	0.75	AB
-	-	+	-1	5	7	14	4r	1.75	C
+	-	+	3	6	11	2	4r	0.25	AC
-	+	+	2	4	1	4	4r	0.50	BC
+	+	+	11	9	5	4	4r	0.50	ABC

3^2 요인설계

요인 A의 수준수가 3이므로 자유도는 2가 되고, A가 양적 요인이며 등간격일 때는 SS_A를 예 11.1에서와 같이 1차 효과 L_l과 2차 효과 L_q에 의한 대비의 변동 SS_l과 SS_q로 분할할 수 있다. 즉,

$$L_l = T_{2..} - T_{0..}$$

$$L_q = T_{0..} - 2\,T_{1..} + T_{2..}$$

$$SS_l = \frac{L_l^2}{r'\sum_i c_i^2} = \frac{L_l^2}{6r} \quad (\because\ r' = 3\times r,\ \sum_i c_i^2 = 2)$$

$$SS_q = \frac{L_q^2}{r'\sum_i c_i^2} = \frac{L_q^2}{18r} \ (\because \ \sum_i c_i^2 = 6)$$

$$SS_A = SS_l + SS_q$$

가 되며, SS_B에 대해서도 같은 관계가 성립한다.

(예 11.4) SS_{AB} = 9개 처리조합간의 변동 (자유도 8)

$$= \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} T_{ij.}^2 - \frac{T^2}{18} = 75.781$$

$SS_{A\times B}$ = A와 B간의 상호작용 (자유도 2×2=4)

$$SS_{AB} = SS_A + SS_B + SS_{A\times B}$$

$$SS_T = \sum_i\sum_j\sum_k x_{ijk}^2 - \frac{T^2}{18} = 78.296$$

$$SS_E = SS_T - SS_{AB} = 2.515$$

$$A: L_l = 37.6 - 10.4 = 27.2$$

$$L_q = 10.4 - 2(26.3) + 37.6 = -4.6$$

$$SS_l = \frac{L_l^2}{6r} = \frac{(27.2)^2}{12} = 61.653^{**}$$

$$SS_q = \frac{L_q^2}{18r} = \frac{(-4.6)^2}{36} = 0.588$$

$$SS_A = SS_l + SS_q = 62.241^{**}$$

$$B: L_l = 23.0 - 19.8 = 3.2$$

$$L_q = 19.8 - 2(31.5) + 23.0 = -20.2$$

$$SS_l = \frac{(3.2)^2}{12} = 0.853$$

$$SS_q = \frac{(-20.2)^2}{36} = 11.334^{**}$$

$SS_B = SS_l + SS_q = 12.188^{**}$ (p. 282 분산분석표)

A(온도)는 1차 효과가, B(촉매)는 2차 효과가 유의함. (그림 11.5)

합성수율을 높이는 조건: A_2B_1 ⇒ 최적조건을 구하기 위해서는 EVOP(16장)을 사용(표 11.8)

3^3 요인설계

모든 요인이 양적 요인이며 등간격일 경우는 각 요인에 의한 변동은 3^2 요인설계

와 마찬가지로 1차 효과 L_l과 2차 효과 L_q에 의한 대비의 변동 SS_l과 SS_q로 분할할 수 있다. 예컨대 SS_A는

$$L_l = T_{2...} - T_{0...} \qquad SS_l = \frac{L_l^2}{18r}$$

$$L_q = T_{0...} - 2\,T_{1...} + T_{2...} \qquad SS_q = \frac{L_q^2}{54r}$$

$$SS_A = SS_l + SS_q$$

로 분할된다. 또한 2요인 및 3요인 상호작용에 의한 변동은 3원배치법에서와 마찬가지 방법으로 계산할 수 있다. 예컨대 A×B에 의한 변동은

$$SS_{A\times B} = SS_{AB} - SS_A - SS_B \quad \left(SS_{AB} = \frac{1}{3r}\sum_i\sum_j T_{ij..}^2 - \frac{T^2}{27r}\right)$$

으로 구할 수 있고

$$SS_{A\times B\times C} = SS_{ABC} - (SS_A + SS_B + SS_C + SS_{A\times B} + SS_{A\times C} + SS_{B\times C})$$

$$\left(SS_{ABC} = \frac{1}{r}\sum_i\sum_j\sum_k T_{ijk.}^2 - \frac{T^2}{27r}\right)$$

으로 구할 수 있다. 이때 분산분석표는 다음과 같다.

변인	제곱합	자유도
A	SS_A	2
B	SS_B	2
C	SS_C	2
A×B	$SS_{A\times B}$	4
A×C	$SS_{A\times C}$	4
B×C	$SS_{B\times C}$	4
A×B×C	$SS_{A\times B\times C}$	8
E	SS_E	27(r-1)
Total	SS_T	27r-1

Note: 모든 요인효과를 한 번에 효율적으로 계산하기 위해서 2^k 요인설계와 마찬가지로 3^k 요인설계에도 Yates의 알고리즘을 사용할 수 있다. 예 11.4의 경우에 Yates의 알고리즘을 적용하면 다음과 같다.

			(1)	(2)	effect	divisor	SS(=(2)^2/divisor)
0	0	1.2	19.8	74.3	-	-	
1	0	7.3	31.5	27.2	A_l	$2^1\times3^1\times2$	61.653
2	0	11.3	23.0	-4.6	A_q	$2^1\times3^2\times2$	0.588
0	1	6.0	10.1	3.2	B_l	$2^1\times3^1\times2$	0.853
1	1	11.6	7.9	-0.9	A_lB_l	$2^2\times3^0\times2$	0.101
2	1	13.9	9.2	2.9	A_qB_l	$2^2\times3^1\times2$	0.350
0	2	3.2	-2.1	-20.2	B_q	$2^1\times3^2\times2$	11.334
1	2	7.4	-3.3	3.5	A_lB_q	$2^2\times3^1\times2$	0.510
2	2	12.4	0.8	5.3	A_qB_q	$2^2\times3^2\times2$	0.390

(divisor = $2^a\times3^b\times r$ 여기서 a는 해당효과의 요인수를 나타내고 b는 총 요인수 – 해당효과의 1차 요인수를 나타냄.)

H.W. #3

p.284 6, 9, 12