

2007년 2월

석사학위논문

블록디자인에 의한 cDNA Microarray 실험에 관한 연구

조선대학교 대학원

전산통계학과

이 승 우

블록디자인에 의한 cDNA Microarray 실험에 관한 연구

cDNA Microarray Experimental Design based on Block
Designs

2007년 2월 23일

조선대학교 대학원

전산통계학과

이 승 우

블록디자인에 의한 cDNA Microarray 실험에 관한 연구

지도교수 최 규 정

이 논문을 이학 석사학위신청 논문으로 제출함

2006년 10월

조선대학교 대학원

전산통계학과

이 승 우

이승우의 석사학위논문을 인준함

위 원 장 조선대학교 교수 _____ 인

위 원 조선대학교 교수 _____ 인

위 원 조선대학교 교수 _____ 인

2006년 11월

조선대학교 대학원

목 차

ABSTRACT

I 서론	1
1.1 개요	1
II 교각 디자인	4
2.1 수준이 2인 교각 디자인	4
III 요인실험법	6
3.1 완전확률화계획에 의한 요인실험	6
3.1.1 일원배치법	6
3.1.2 이원배치법	6
3.1.3. 3요인 실험법	8
3.1.4 다요인 실험법	9
3.2 요인 실험법	11
3.2.1 2^2 -요인 실험	12
3.2.2 2^3 -요인 실험	14
3.2.3 2^k -요인 실험	15
3.2.4 s^k -요인실험법	17
3.3 교각법	20
3.3.1 2^3 요인실험	21
3.3.2 2^4 요인실험	22
3.4 일부실험법	25
3.4.1 2^n 형의 $\frac{1}{2}$ 실험법	26
3.4.2 2^n 형의 $\frac{1}{4}$ 실험법(실험)	27
3.4.3 3^n 형의 일부실험법	28

IV 요인 Microarrays	31
4.1 요인 Microarrays 실험	31
4.1.1 2^n 요인실험	31
4.1.2 2^3 요인실험	33
4.1.3 2^4 요인실험	35
4.1.4 3×2 요인 Microarrays	37
4.2 Visual C++를 이용한 프로그램 예제	38
V 결론 및 토론	46

표 목 차

<표 3.1> 요인 효과	12
<표 3.2> 2수준 요인 효과 계수표	13
<표 3.3> 3수준 요인효과 계수표	15
<표 4.1> 요인 모델	32
<표 4.2> 요인 모델	34
<표 4.3> 3×2요인을 위한 모델	37

그림 목 차

<그림 1> 프로그램에 의한 요인실험에 대한 output	44
---------------------------------------	----

ABSTRACT

cDNA Microarray Experimental Design based on Block Designs

BY Lee Seung Woo.

Advisor : Prof. Choi Kuey Chung

Department of Computer Science and Statistics,
Graduate School of Chosun University

We need more efficient designs when block design is not efficient. In general, Latin square and Youden square designs many times are not suitable because they require more replications which is expensive. If we do not need to estimate some interactions then we can confound them and lose all information on those interactions (many times high order interactions are zero so we don't need to estimate them). By losing information on some interactions, we gain on the information (or better precision) on main effects, two-factor interactions etc. So, single replication is good for economy and confounded row-column design is good for improving precision. Also, balanced factorial designs are introduced for cDNA microarray experiments. Single

replicate designs obtained using the classical method of confounding are shown to be particularly useful for deriving suitable balanced designs for cDNA microarrays. Classical factorial designs obtained using methods other than the method of confounding are also shown to be useful. The paper provides a systematic method of deriving designs for microarray experiments as opposed to algorithmic and ad-hoc methods, and generalizes several of the microarray designs given recently in literature. Also designs for factorial cDNA microarrays realized by programming.

I 서론

1.1 개요

Landgreber 외 (2005) 는 Yang 외 (2002)가 추천하는 표준 데이터 정상화 절차가 단지 dyes 때문에 전체적인 효력을 제거한다고 지적했다. 그러나 데이터가 정상화에 종속시킨 후에도 dye \times gene 상호 작용은 데이터에 생길지도 모른다. 그들은 특별한 방법을 사용하여 2^2 , 2^3 그리고 3×2 실험을 설계했다. Wit 외 (2004)는 simulated annealing 알고리즘을 사용하여 효율적인 설계법을 찾아냈고, dye-swap 설계법은 연구에 속하는 조건의 수에 증가로써 매우 비효율적이 되는 것을 보여줬다. 그들은 또한 simulated annealing 알고리즘을 사용하여 얻는 설계법과 비교할 때 확실히 잘 실행하는 혼합된 루프(loop) 설계법을 정의했다. Shah (1960) 와 Kshirsagar (1966) 에 의해 특성을 보기위한 균형요인배치법(balanced factorial design)은 농업 실험의 환경에서 주요하게 개발되어졌고, 균형 디자인(balanced design)은 다른 환경에서도 유용하게 사용하기 위하여 발견되었다. 예를 들어, Gupta 와 Mukerjee (1996), Kshirsagar 와 Wang (1996), 및 거기에 인용되는 것을 보면, 균형 설계법(balanced design)은 생물 검정(bioassay) 실험을 위해 또한 광범위하게 연구되었다. 이 논문의 목적은 cDNA microarray 실험을 위해 균형요인배치법(balanced factorial design)을 제공하기 위함이다. 2개의 실험 조건을 각각의 microarray 슬라이드에 교배시키기 때문에, 배열은 크기 2의 블록을 형성한다. 2개 이상 조건을 연구하면 블록은 불완전하다. Kerr 및 Churchill (2001a,b), 및 Churchill (2002) 은 ANOVA 모형을 사용하여 microarray 실험에 관련시켜 몇몇 설계법에 세부적인 논의를 준비했다. 그들은 "common reference" 설계법(microarray 실험을 위해 널리 쓰인 설계법)의 비효율적인 점을 보였다. 저자는 microarrays를 위해 루프(loop)설계법(또는 싸이클(cyclic) 설계법)을 소개하고, 루프(loop) 설계법이 상당한 개선 이상의 효율적인 점에서 "common reference" 설계법을 제공한 것을 보였다. 그런데, 루프(loop) 설계법은 많은 수의 처리 조합에서는 비효율적이다. Kerr와 Churchill (2001a)은 설계를 만들기 위해 A-optimality criterion을 이용하였다. Glonek과 Solomon (2003)은 효율적인 설계법을 만들기 위해 마인드(mind)의

중요성의 대비에 대한 조화의 중요성을 강조했다. 그들은 중요성의 특별한 대비의 추정하는 한은 종합적인 능력에 설계법이 대부분의 효율적인 설계법이 아닐지도 모른다는 것을 주장했다. 저자는 허용할 수 있는 설계법을 고려하고, 허용할 수 있는 설계법의 종류에 대해 그들 컴퓨터의 설계법의 조사를 제한했다.

예를 들면, Yang 과 Speed (2002) 에 의해 제안된 모든 comparisons 설계법은 실험자가 연구에 속하는 조건사이에 모든 pairwise 비교에서 동등하게 중요성을 갖을 때 유용하다. 특히 Glonek 과 Solomon (2003)은 고전적인 설계법은 단지 모든 추정가능한 대비의 변동을 최소화하는 것을 찾는 것이고 중요성의 대비는 아니다라고 명확히 제시했다. 그런데 요인 실험의 고전적인 설계법에서 모든 추정가능한 대비는 동등한 중요성이 드물다는 것을 유념해야한다. 그러므로 추정가능 할수도 또는 불가능 할지도 모르는 다른 대비와 비교하여 더 높고 정확한 중요성에 따라 대비를 추정하는 균형요인배치법(balanced factorial design)을 선택하는 것은 당연한 것이다. 실제로 고전적인 교각법은 블록에 완전 또는 부분 교각한 좀 더 중요성이 적은 것을 희생시켜 중요성의 대비에 더 높고 정확하게 달성시키는 것을 확실하게 한다. 고전적인 교각법은 균형요인배치법(balanced factorial designs)을 위해 매우 유용하지만, 요인 microarray 실험을 위해 적당한 설계법을 작성하는데, 그의 응용은 본 논문에서는 다루지 않았다. 최근 Glonek 와 Solomon (2004) 및 Wit 외 (2004) 는 monitoring 기법을 통한 설계법을 제안한 바 있다 . 또한 Landgrebe 외 (2005) 의 특별한 방법은 단지 요인의 수 그리고/또는 수준의 수가 작을 때 유용하다.

블록 크기 2인 균형요인배치법(balanced factorial designs)과 요인 microarray 설계법 사이에 연결을 본 논문에서 취급 하겠다. microarray 실험을 위한 균형설계법(balanced designs)은 연구에 유효한 균형요인설계법(balanced factorial designs)을 이용하므로써 쉽게 얻을 수 있음을 보였다. 그룹 분할가능하고 일반화된 싸이클(cyclic) 디자인과 같은 블록 크기 2인 균형요인설계법(balanced factorial designs) 역시 유용하다는 것을 보였다.

본 논문의 microarrays에 대한 역사적 배경은 Lee (2004) 를 참조하는 것이 좋을 것 같다. 전형적인 교각된 설계법과 요인 microarrays에 일반화된 싸이클(cyclic) 설계법의 응용은 제 4장에서 소개한다. 이해를 돕기 위하여 2^2 , 2^3 그리고 3×2 실험은 상세히 논의되었다. 그밖에 다른 실험에 대한 microarray 설계법은 유사하므로 생략 하였

다. 마지막으로 제 5장에는 결론 및 토론을 넣었다.

II 교락 디자인

2.1 수준이 2인 교락 디자인

2개 수준 각각에 m 개의 요인 F_1, F_2, \dots, F_m 을 관련시키는 요인 실험을 생각해보자. 처리조합은 m 개의 요소로 된 집합 a_1, a_2, \dots, a_m ($a_i = 0$ 또는 1) 으로 표시하고, i 번째 요인의 2개의 코드 수준으로 표시할 것이다, $i = 1, 2, \dots, m$. 예를 들어 Raghavarao (1971)을 보면, D 를 전형적인 교락법을 사용하여 얻은 m 개의 요인 실험을 위한 크기 2의 2^{m-1} 블록을 가지는 단순 반복 실험이라 한다. D 의 블록 사이에 교락된 2^{m-1} 요인효과(주효과, 교호작용)는 B_i , $i = 1, 2, \dots, 2^{m-1}$,로 나타내고, A_i , $i = 1, 2, \dots, 2^{m-1}$,는 교락되지 않은 요인효과를 나타낸다.

(t_{i1}, t_{i2}) 는 D 의 i 번째 블록에서 2개의 처리조합으로 나타내고, (y_{i1}, y_{i2}) 는 그것들에 대응시킨 관측치이다, $i = 1, 2, \dots, 2^{m-1}$. 그러면 블록 대비 합 $y_{i1} + y_{i2}$, $i = 1, 2, \dots, 2^{m-1}$,는 요인효과 B_i , $i = 1, 2, \dots, 2^{m-1}$,에 교락되어 블록효과를 추정한다. 그런데 교락되지 않은 요인효과 A_i , $i = 1, 2, \dots, 2^{m-1}$,는 블록 차 $y_{i1} - y_{i2}$, $i = 1, 2, \dots, 2^{m-1}$,를 이용하여 추정된다.

$$y_d = [(y_{11} - y_{12}), (y_{21} - y_{22}), \dots, (y_{b1} - y_{b2})]'$$

여기서 $b = 2^{m-1}$ 는 블록의 수를 나타낸다. $h_j = (h_{j1}, h_{j2}, \dots, h_{jb})'$ 이면, $E(h_j' y_d) = A_j$, $j = 1, 2, \dots, 2^{m-1}$,이다. 그러면 분명히 h_j , $j = 1, 2, \dots, 2^{m-1}$,는 크기 2^{m-1} 인 서로 직교하는 열벡터의 완비집합(complete set)의 형태이다. 또한 $Var(h_j' y_d) = h_j' h_j \sigma^2$ 은 $y_{i1} - y_{i2}$ 차의 상수변동(constant variance)으로 나타낸다. 2^{m-1} 배열에 관련시키는 microarray 실험의 경우에는 i 번째 배열에 교배시키는 2개의 처리 조합을 위한 \log_2 표현비율과 일치한다.

모든 2^{m-1} 개의 요인효과가 부분 교락된 설계법은 요인효과의 다른 집합을 교락한 그 이상의 반복을 추가해서 얻은 것으로 추정할지도 모른다. 우리가 2^m 개의 실험을 이

용하여 교각법을 설명했지만, 그 방법은 소수 또는 소수의 거듭제곱인 다수 수준의 요인에 일반적으로 적용할 수 있다. 실제로 유사한 결론으로 전형적인 교각법 등의 방법을 이용하여 얻은 블록 크기 2의 균형요인배치법(balanced factorial designs) 또한 사용되고 있다.

III 요인실험법

3.1 완전확률화계획에 의한 요인실험

3.1.1 일원배치법

실험설계의 기본원리인 랜덤화(randomization), 블록화(blocking), 반복(replication) 중 랜덤화의 한 방법이다. 이것은 랜덤화의 원리를 충실히 따른 원리이다. 선택된 변인 이외의 환경을 랜덤화의 방법을 통하여 평준화시킴으로써 실험의 분석력을 높이는데 목적이 있다.

$$\text{모형 : } y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad i = 1 \cdots a, \quad j = 1 \cdots n \quad (3.1)$$

이처럼 어떤 반응값에 대하여 하나의 변인의 영향만을 살피고자 할 때 이러한 실험설계를 일원배치법(one-way layout)이라 한다.

(1) 최소제곱 추정량 :

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$$

(2) 분산분석

$$\begin{aligned} SS_{\tau} &= \sum_{i=1}^a \sum_{j=1}^n \hat{\tau}_i^2 = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{또는} \quad \frac{N}{a} \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned}$$

$$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

$$SST = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

3.1.2 이원배치법

모형 $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$
 $i = 1 \cdots a, j = 1 \cdots b, k = 1 \cdots n$

$$N = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n 1 = abn$$

(1) 최소제곱 추정량 :

$$\hat{\mu} = \bar{y} \dots$$

$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y} \dots \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y} \dots$$

$$\text{교호작용 } (\tau\beta)_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots$$

(2) 분산분석

$$SS_{\tau} = \sum_i \sum_j \sum_k \hat{\tau}_i^2 = \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y} \dots)^2$$

$$= nb \sum_i (\bar{y}_{i..} - \bar{y} \dots)^2$$

(여기서 $nb = \frac{N}{a}$ 임에 유의)

$$SS_{\beta} = \sum_i \sum_j \sum_k \hat{\beta}_j^2 = \sum_i \sum_j \sum_k (\bar{y}_{.j.} - \bar{y} \dots)^2$$

$$= an \sum_j (\bar{y}_{.j.} - \bar{y} \dots)^2$$

(여기서 $an = \frac{N}{b}$ 임)

$$SS_{\tau\beta} = \sum_i \sum_j \sum_k (\hat{\tau\beta})_{ij}^2 = \sum_i \sum_j \sum_k (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2$$

$$= n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2$$

(여기서 $n = \frac{N}{ab}$ 임)

$$SSE = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot})^2$$

$$SST = SS_\tau + SS_\beta + SS_{\tau\beta} + SSE$$

$$= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{\dots})^2$$

3.1.3. 3요인 실험법

$$\begin{aligned} \text{모형 } y_{ijk\ell} = & \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} \\ & + (ABC)_{ijk} + \epsilon_{ijk\ell} \end{aligned}$$

$$i = 1 \cdots a_1, \quad j = 1 \cdots a_2, \quad k = 1 \cdots a_3, \quad \ell = 1 \cdots n$$

$$N = \sum_i \sum_j \sum_k \sum_\ell 1 = a_1 a_2 a_3 n$$

(1) 최소제곱 추정량

$$\hat{\mu} = \bar{y}_{\dots}$$

$$\hat{A}_i = \bar{y}_{i\cdot\cdot\cdot} - \bar{y}_{\dots}, \quad \hat{B}_j = \bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\dots}$$

$$\hat{C}_k = \bar{y}_{\cdot \cdot k \cdot} - \bar{y}_{\dots}$$

$$(\widehat{AB})_{ij} = \bar{y}_{ij\cdot\cdot} - \bar{y}_{i\cdot\cdot\cdot} - \bar{y}_{\cdot j \cdot \cdot} + \bar{y}_{\dots}$$

$$(\widehat{AC})_{ik} = \bar{y}_{i\cdot k \cdot} - \bar{y}_{i\cdot\cdot\cdot} - \bar{y}_{\cdot \cdot k \cdot} + \bar{y}_{\dots}$$

$$(\widehat{BC})_{jk} = \bar{y}_{\cdot j k \cdot} - \bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\cdot \cdot k \cdot} + \bar{y}_{\dots}$$

$$(\widehat{ABC})_{ijk} = \bar{y}_{ijk\cdot} - \bar{y}_{ij\cdot\cdot} - \bar{y}_{i\cdot k \cdot} - \bar{y}_{\cdot j k \cdot} + \bar{y}_{i\cdot\cdot\cdot} + \bar{y}_{\cdot j \cdot \cdot} + \bar{y}_{\cdot \cdot k \cdot} - \bar{y}_{\dots}$$

(2) 분산분석

$$\begin{aligned}
SS_A &= \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} \sum_{\ell=1}^n \hat{A}_i = \sum_i \sum_j \sum_k \sum_{\ell} (\bar{y}_{i\dots} - \bar{y}_{\dots})^2 \\
&= a_2 a_3 n \sum_i (\bar{y}_{i\dots} - \bar{y}_{\dots})^2 \\
SS_B &= \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} \sum_{\ell=1}^n \hat{B}_i = \sum_i \sum_j \sum_k \sum_{\ell} (\bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\dots})^2 \\
&= a_1 a_3 n \sum_{j=1}^{a_2} (\bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\dots})^2 \\
SS_C &= a_1 a_2 n \sum_{k=1}^{a_3} (\bar{y}_{\cdot \cdot k \cdot} - \bar{y}_{\dots})^2 \\
SS_{AB} &= \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} \sum_{\ell=1}^n (\hat{AB})_{ij} = a_3 n \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} (\bar{y}_{ij \cdot \cdot} - \bar{y}_{i \dots} - \bar{y}_{\cdot j \cdot \cdot} + \bar{y}_{\dots})^2 \\
SS_{AC} &= a_2 n \sum_{i=1}^{a_1} \sum_{k=1}^{a_3} (\bar{y}_{i \cdot k \cdot} - \bar{y}_{i \dots} - \bar{y}_{\cdot \cdot k \cdot} + \bar{y}_{\dots})^2 \\
SS_{BC} &= a_1 n \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} (\bar{y}_{\cdot j k \cdot} - \bar{y}_{\cdot j \cdot \cdot} - \bar{y}_{\cdot \cdot k \cdot} + \bar{y}_{\dots})^2 \\
SS_{ABC} &= n \sum_i \sum_j \sum_k (\bar{y}_{ijk \cdot} - \bar{y}_{ij \cdot \cdot} - \bar{y}_{i \cdot k \cdot} - \bar{y}_{\cdot j k \cdot} + \bar{y}_{i \dots} + \bar{y}_{\cdot j \cdot \cdot} + \bar{y}_{\cdot \cdot k \cdot} - \bar{y}_{\dots})^2 \\
SSE &= \sum_i \sum_j \sum_k \sum_{\ell} (\bar{y}_{ijkl} - \bar{y}_{ijk \cdot})^2 \\
SST &= SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SSE \\
&= \sum_i \sum_j \sum_k \sum_{\ell} (\bar{y}_{ijkl} - \bar{y}_{\dots})^2
\end{aligned}$$

3.1.4 다요인 실험법

(정의) $\alpha_1^{f_1} \times \alpha_2^{f_2} \times \dots \times \alpha_q^{f_q}$

여기서 a_1, a_2, \dots, a_q : level

q : 다른 수준의 수

f_i ($i = 1, \dots, q$) : 수준 a_i 에 대한 요인(factor)수

$k = \sum_{i=1}^q f_i$: 총요인 수

F_1, F_2, \dots, F_k : factor

$\prod_{i=1}^k a_i = \prod_{i=1}^q a_i^{f_i}$: treatment combination 수

모형 $y_{ijk\dots\ell} = \mu + A_i + B_j + C_k + \dots + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + \dots$
 $+ (ABC)_{ijk} + (3차\ 교호작용\ 나머지) + (4차\ 이상\ 교호작용)$
 $+ \epsilon_{ijk\dots\ell}$

$i = 1 \dots a_1, j = 1 \dots a_2, k = 1 \dots a_3, \dots, \ell = 1 \dots n, N = a_1 a_2 a_3 \dots n$

(1) 최소제곱 추정량

$$\hat{\mu} = \bar{y}_{\dots\dots}$$

$$\hat{A}_i = \bar{y}_{i\dots\dots} - \bar{y}_{\dots\dots}$$

$$\hat{B}_j = \bar{y}_{\cdot j \dots\dots} - \bar{y}_{\dots\dots}$$

$$\hat{C}_k = \bar{y}_{\cdot\cdot k \dots\dots} - \bar{y}_{\dots\dots}$$

$$(\widehat{AB})_{ij} = \bar{y}_{ij \dots\dots} - \bar{y}_{i \dots\dots} - \bar{y}_{\cdot j \dots\dots} + \bar{y}_{\dots\dots}$$

.....

$$(\widehat{ABC})_{ijk} = \bar{y}_{ijk \dots\dots} - \bar{y}_{ij \dots\dots} - \bar{y}_{i \cdot k \dots\dots} - \bar{y}_{\cdot jk \dots\dots} \\ + \bar{y}_{i \dots\dots} + \bar{y}_{\cdot j \dots\dots} + \bar{y}_{\cdot\cdot k \dots\dots} + \bar{y}_{\dots\dots}$$

.....

(2) 분산분석

$$SS_A = \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} \cdots \sum_{\ell=1}^n \hat{A}_i$$

$$= \sum_i \sum_j \sum_k \cdots \sum_{\ell} (\bar{y}_{i \dots \ell} - \bar{y}_{\dots \ell})^2$$

$$= \frac{N}{a_1} \sum_{i=1}^{a_1} (\bar{y}_{i \dots} - \bar{y}_{\dots})^2$$

$$SS_B = \frac{N}{a_2} \sum_{j=1}^{a_2} (\bar{y}_{\cdot j \dots} - \bar{y}_{\dots})^2$$

$$SS_C = \frac{N}{a_3} \sum_{k=1}^{a_3} (\bar{y}_{\dots k \dots} - \bar{y}_{\dots})^2$$

.....

$$SS_{AB} = \frac{N}{a_1 a_2} \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} (\bar{y}_{ij \dots} - \bar{y}_{i \dots} - \bar{y}_{\cdot j \dots} + \bar{y}_{\dots})^2$$

.....

$$SS_{ABC} = \frac{N}{a_1 a_2 a_3} \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} \sum_{k=1}^{a_3} (\bar{y}_{ijk \dots} - \bar{y}_{ij \dots} - \bar{y}_{i \cdot k \dots} - \bar{y}_{\cdot jk \dots} + \bar{y}_{i \dots} + \bar{y}_{\cdot j \dots} + \bar{y}_{\dots k \dots} - \bar{y}_{\dots})^2$$

.....

$$SSE = \sum_i \sum_j \sum_k \cdots \sum_{\ell} (\bar{y}_{ijk \dots \ell} - \bar{y}_{ijk \dots})^2$$

$$SST = SS_A + SS_B + \cdots + SS_{AB} + \cdots + SS_{ABC} + \cdots + SSE$$

$$= \sum_i \sum_j \sum_k \cdots \sum_{\ell} (\bar{y}_{ijk \dots \ell} - \bar{y}_{\dots \ell})^2$$

3.2 요인 실험법

(정의) S^k : (1) S : level 이고 1, 2, \dots s-1 사용

(2) k : factor 수 A, B, \dots K 등으로 요인을 표시

실험이 반복되지 않아도 S^k 개의 실험회수가 실시되며, 모든 요인효과(주효과와 교호작용)를 추정할 수 있다.

3.2.1 2^2 -요인 실험

(1) 수준

S=2 : level

K=2 : factor 수

두요인 $A = \{A_0, A_1\}$ $B = \{B_0, B_1\}$ (\because S=2 이므로)

(2) treatment combination :

1) 개수 = 2^2 or $2 \times 2 = 4$

2) 표준순서 (yates 순서)

00 10 01 11

<참고> 크기순서 00 01 10 11

(3) factor effect :

<표 3.1> 요인 효과

		B	
		B_0	B_1
A	A_0	00	01
	A_1	10	11

$$1) \hat{A}_0 = \frac{00 + 01}{2} - \frac{00 + 01 + 10 + 11}{4}$$

$$\hat{A}_1 = \frac{10 + 11}{2} - \frac{00 + 01 + 10 + 11}{4}$$

$$\hat{A} = A \text{의 효과} = \hat{A}_1 - \hat{A}_0 = \frac{1}{2} [(10 + 11) - (00 + 01)] \quad (3.2)$$

$$= \frac{1}{2^2 - 1} \quad (\text{대비 } A)$$

$$2) \hat{B}_0 = \frac{00 + 10}{2} - \frac{00 + 01 + 10 + 11}{4}$$

$$\hat{B}_1 = \frac{10 + 11}{2} - \frac{00 + 01 + 10 + 11}{4}$$

$$\hat{B} = B \text{의 효과} = \hat{B}_1 - \hat{B}_0 = \frac{1}{2} [(01 + 11) - (00 + 10)] \quad (3.3)$$

$$= \frac{1}{2^2 - 1} \quad (\text{대비 } B)$$

3) 교호작용 AB 의 효과

$$\widehat{AB} = AB \text{의 효과} = \frac{1}{2} [(B_1 \text{에서 } A \text{의 효과}) - (B_0 \text{에서 } A \text{의 효과})]$$

$$= \frac{1}{2} [(11 - 01) - (10 - 00)]$$

$$= \frac{1}{2} [(00 + 11) - (01 + 10)] \quad (3.4)$$

4) 요인 효과 계수표 ((3.2),(3.3),(3.4)를 부호로만 표시)

<표 3.2> 2수준 요인 효과 계수표

처리 조합	요인 효과		
	A	B	AB
00	-	-	-
10	+	-	+
01	-	+	-
11	+	+	+

3.2.2 2^3 -요인 실험

(1) level $S=2$

factor 수 $K=3$

factor $A = \{A_0, A_1\}$, $B = \{B_0, B_1\}$, $C = \{C_0, C_1\}$

(2) treatment combination

1) 개수 $= 2^3$ 또는 $2 \times 2 \times 2 = 8$

2) 표준순서 (yates 순서)

000 100 010 110 001 101 011 111

<참고> 크기 순서

000 001 010 011 100 101 110 111

(3) factor effect

A_0 000 001 010 011

A_1 100 101 110 111

1) A 의 효과 $= \frac{1}{4} [(100 + 101 + 110 + 111) - (000 + 001 + 010 + 011)]$

2) B 의 효과 $= \frac{1}{4} [(010 + 011 + 110 + 111) - (000 + 001 + 100 + 101)]$

3) C 의 효과 $= \frac{1}{4} [(001 + 011 + 101 + 111) - (000 + 010 + 100 + 110)]$

4) 교호작용 (AB , AC , BC , ABC)의 효과

앞의 2^2 -요인 실험 방법에 의하여 구할 수 있다. 그러나 계속 일반화하면 이 방법은 너무 복잡함으로 다음 요인 효과 계수표에 의하여 구한다.

<표 3.3> 3수준 요인효과 계수표

처 리 조 합	요 인 효 과						
	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>
000	—	—	+	—	+	+	—
100	+	—	—	—	—	+	+
010	—	+	—	—	+	—	+
110	+	+	+	—	—	—	—
001	—	—	+	+	—	—	+
101	+	—	—	+	+	—	—
011	—	+	—	+	—	+	—
111	+	+	+	+	+	+	+

3.2.3 2^K -요인 실험

(1) level $S = 2$

factor 수 = k 개

factor A, B, \dots, K (모두 K 개)

(2) treatment combination

1) 개수 = 2^k

2) 2^4 인 것의 예시

0000 1000 0100 1100 0010

1010 0110 1110 0001 1001

0101 1101 0011 1011 0111

1111 (모두 16개이며 표준순서로 썼음)

2^k 인 경우 0과 1을 k 자리로 구성하며 위 예시와 같이 일반화 하여 처리조합을 구성한다.

(3) factor effect(교호작용포함) : 반복수가 1인 경우

$$1) A \text{의 효과} = \frac{1}{2^{k-1}} [(\text{첫수가 1인 처리조합 합}) - (\text{첫수가 0인 처리조합 합})]$$

$$= \frac{1}{2^{k-1}} (\text{대비}_B)$$

2) B, C, ..., K인 경우도 유사함

$$3) AB \text{의 효과} = \frac{1}{2^{k-1}} (\text{대비}_{AB})$$

.....

$$AB \cdots K \text{의 효과} = \frac{1}{2^{k-1}} (\text{대비}_{AB \cdots K})$$

(4) 제곱합

(a) 2^2 -요인실험

$$\begin{aligned} SS_A &= \frac{N}{a_1} \sum_{i=1}^{a_1} (\bar{y}_{i \cdots \cdots} - \bar{y}_{\cdots \cdots})^2 \\ &= \frac{2^2}{2} \sum_{i=0}^1 (\bar{y}_{i \cdot} - \bar{y}_{\cdot \cdot})^2 \\ &= 2 [(\bar{y}_{0 \cdot} - \bar{y}_{\cdot \cdot})^2 + (\bar{y}_{1 \cdot} - \bar{y}_{\cdot \cdot})^2] \\ &= 2 \left[\left(\frac{00 + 01}{2} - \frac{00 + 01 + 10 + 11}{4} \right)^2 + \left(\frac{10 + 11}{2} - \frac{00 + 01 + 10 + 11}{4} \right)^2 \right] \\ &= \frac{2}{16} [((00 + 01) - (10 + 11))^2 + ((10 + 11) - (00 + 01))^2] \\ &= \frac{1}{8} [2((00 + 01) - (10 + 11))^2] (\because \text{위의 두 항은 같음}) \\ &= \frac{1}{4} (\text{대비}_{AB})^2 \\ SS_{AB} &= \frac{1}{2^2} (\text{대비}_{AB})^2 \end{aligned}$$

(b) 2^k -요인실험(반복수가 1인 경우)

위의(1)을 일반화하면

$$SS_A = \frac{1}{2^k} (\text{대비}_{AB})^2$$

.....

$$SS_{AB \cdots k} = \frac{1}{2^k} (\text{대비}_{AB \cdots K})^2$$

단 반복수가 n 인 경우는 위의 각 경우에 $\frac{1}{n}$ 배 하면 된다.

3.2.4 S^k -요인실험법

(1) $k=2$ 인 경우(3^2 -요인실험법)

(a) level : $s=3$

factor : A, B ($k=2$ 개)

(b) treatment combination

1) 개수 = 3^2 개

2) 00, 01, 02, 10, 11, 12, 20, 21, 22

(c) 직교분해 (직교다항식의 이용방법)

① 요인 A의 선형(1차)효과 : 직선의 기울기들의 평균

$$A_L = \frac{1}{2} [(\bar{y}_{1\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})] [(\bar{y}_{1\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})]$$

$$= \frac{1}{2} (-\bar{y}_{0\cdot} + \bar{y}_{2\cdot})$$

$$= \frac{1}{2} (-1 \ 0 \ 1) \begin{pmatrix} \bar{y}_{0\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{pmatrix}$$

대비 (-1, 0, 1) : 선형(1차) 다항계수라 함

② 요인 A의 2차효과 : 두 직선기울기의 차

$$A_Q = (\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) - (\bar{y}_{1\cdot} - \bar{y}_{0\cdot}) = \bar{y}_{2\cdot} - 2\bar{y}_{1\cdot} + \bar{y}_{0\cdot}$$

$$= (1 \ -2 \ 1) \begin{pmatrix} \bar{y}_{0\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{pmatrix}$$

대비(1 -2 1) : 2차 다항계수라 함

③ 요인 B의 선형효과

$$B_L = \frac{1}{2} [(\bar{y}_{\cdot 1} - \bar{y}_{\cdot 0}) + (\bar{y}_{\cdot 2} - \bar{y}_{\cdot 1})]$$

$$= \frac{1}{2} (-1 \ 0 \ 1) \begin{pmatrix} \bar{y}_{\cdot 0} \\ \bar{y}_{\cdot 1} \\ \bar{y}_{\cdot 2} \end{pmatrix}$$

④ 요인 B의 2차효과

$$B_Q = (\bar{y}_{\cdot 2} - \bar{y}_{\cdot 1}) - (\bar{y}_{\cdot 1} - \bar{y}_{\cdot 0}) = \bar{y}_{\cdot 0} - 2\bar{y}_{\cdot 1} + \bar{y}_{\cdot 2}$$

$$= (1 \ -2 \ 1) \begin{pmatrix} \bar{y}_{\cdot 0} \\ \bar{y}_{\cdot 1} \\ \bar{y}_{\cdot 2} \end{pmatrix}$$

(d) 직교다항계수행렬

$$\textcircled{1} \ S=3 \quad \begin{pmatrix} \text{1차다항계수} \\ \text{2차다항계수} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

$$\textcircled{2} \ S=4 \quad \begin{pmatrix} \text{1차 다항계수} \\ \text{2차 다항계수} \\ \text{3차 다항계수} \end{pmatrix} = \begin{pmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}$$

각행 합 = 0

$$\begin{aligned} \text{i) } A_L &= \frac{1}{6} [(\bar{y}_{1\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{2\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{0\cdot}) \\ &\quad + (\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{1\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{2\cdot})] \\ &= \frac{1}{6} (-3\bar{y}_{0\cdot} - \bar{y}_{1\cdot} + \bar{y}_{2\cdot} - 3\bar{y}_{3\cdot}) \end{aligned}$$

$$= (-3 - 113) \begin{pmatrix} \bar{y}_{0\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{3\cdot} \end{pmatrix}$$

$$\text{ii) } A_Q = [(\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) - (\bar{y}_{1\cdot} - \bar{y}_{0\cdot})] + [(\bar{y}_{3\cdot} - \bar{y}_{2\cdot}) - (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})]$$

$$= \bar{y}_{0\cdot} - \bar{y}_{1\cdot} - \bar{y}_{2\cdot} + \bar{y}_{3\cdot} = (1 - 1 - 11) \begin{pmatrix} \bar{y}_{0\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{3\cdot} \end{pmatrix}$$

$$\text{iii) } A_C = [(\bar{y}_{3\cdot} - \bar{y}_{2\cdot}) - (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})] - [(\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) - (\bar{y}_{1\cdot} - \bar{y}_{0\cdot})]$$

$$= -\bar{y}_{0\cdot} + 3\bar{y}_{1\cdot} - 3\bar{y}_{2\cdot} + \bar{y}_{3\cdot}$$

$$= (-13 - 31) \begin{pmatrix} \bar{y}_{0\cdot} \\ \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \\ \bar{y}_{3\cdot} \end{pmatrix}$$

$$\text{즉 } A_C = (\text{차}) - (\text{차})$$

(e) 제곱합

1) 처리대비에 대한 제곱합

$$\frac{\left(\sum_{i=1}^a C_i \bar{y}_{i\cdot} \right)^2}{\left(\sum_{i=1}^a \frac{C_i^2}{n} \right)} = \frac{\left(\sum_{i=1}^a C_i y_{i\cdot} \right)^2}{\left(\sum_{i=1}^a n C_i^2 \right)}$$

을 이용하여 인자 A 의 수준수가 a 이고 반복이 N 인 실험에서 요인 A 에 대한 직

$$\text{교대비 제곱합은 위로부터 } \frac{\left(\sum_{i=1}^a C_i y_{i\cdot} \right)^2}{\left(an \sum_{i=1}^a C_i^2 \right)} \text{이다.}$$

$$\begin{aligned}
2) &= \frac{(-y_{0..} + y_{2..})^2}{3 \times n((-1)^2 + 0^2 + 1^2)} = \frac{(y_{2..} - y_{0..})^2}{6n} \\
&= \frac{(-y_{0..} - 2y_{1..} + y_{2..})^2}{3 \times n(1^2 + (-2)^2 + 1^2)} = \frac{(y_{2..} - 2y_{1..} + y_{0..})^2}{18n}
\end{aligned}$$

$$SS_A = SS_{AL} + SS_{AQ}$$

$$\text{또는 } n \sum_{i=0}^2 (\bar{y}_{i..} - \bar{y} \dots)^2$$

$$\begin{aligned}
3) &= \frac{(y_{.2.} - y_{.0.})^2}{6n} \\
&= \frac{(y_{.2.} - 2y_{.1.} + y_{.0.})^2}{18n}
\end{aligned}$$

$$SS_B = SS_{BL} + SS_{BQ}$$

$$\text{또는 } n \sum_{j=0}^2 (\bar{y}_{.j.} - \bar{y} \dots)^2$$

$$\begin{aligned}
4) &= n \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 \\
&= n \sum_{i=1}^2 \sum_{j=1}^2 (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 \\
&= n \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} (y_{ij.} - \bar{y}_{ij.})^2 \\
&= n \sum_{i=1}^2 \sum_{j=1}^2 (y_{ij.} - \bar{y}_{ij.})^2
\end{aligned}$$

$$SST = SS_A + SS_B + SS_{AB} + SSE$$

$$\text{또는 } \sum_{i=1}^{a_1} \sum_{j=1}^{a_2} n (y_{ij.} - \bar{y} \dots)^2$$

3.3 교각법

실험회수를 늘리지 않고 실험전체를 몇 개의 블록으로 나누어 배치시킴으로서 동일 환경내의 실험회수를 적게하도록 고안해 낸 배치법으로서 실험오차가 작아지게 할 수 있으며, 실험의 정도를 향상시킨다. 보통 직교배열표를 사용하여 배치시키면 편하다.

3.3.1 2^3 요인실험

8가지의 수준조합: (1), a, b, c, ab, ac, bc, abc

block 1	block 2
(1)	a
ab	b
ac	c
bc	abc

$$\begin{aligned}
 ABC &= \frac{1}{4} \{ (a + b + c + abc) - ((1) + ab + ac + bc) \} \\
 &= (\text{block 2의 측정치 평균}) - (\text{block 1의 측정치 평균}) \\
 &\Rightarrow \text{block 2와 block 1의 차이}
 \end{aligned}$$

교호작용 ABC 와 block 효과가 서로 교락(confounding)되어 있다.

ABC 이외의 모든 효과 (A, B, C, AB, AC, BC)는 block과 교락되어 있지 않다.

교호작용 ABC 를 무시할 수 있는 경우에 이러한 교락법을 사용할 수 있다.

A 의 주효과:

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

A 의 주효과를 block 효과와 교락시키려면

block 1	block 2
a	(1)
ab	b
ac	c
abc	bc

A 와 C 의 교호작용:

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

교호작용 AC 를 block 효과와 교락시키려면

block 1	block 2
(1)	a
b	c
ac	ab
abc	bc

※ 일반적으로 2개의 block으로 나눌 때에는 최고차의 교호작용을 block과 교락시키는 것이 좋다.

3.3.2 2^4 요인실험

(1) 2개의 block으로 나누는 경우 (단독교락)

4인자 교호작용 $ABCD$ 를 block과 교락시키려면

$$\begin{aligned} ABCD &= \frac{1}{8} (a-1)(b-1)(c-1)(d-1) \\ &= \frac{1}{8} \{ (1) + ab + ac + ad + bc + bd + cd + abcd \} \\ &\quad - (a + b + c + d + abc + abd + acd + bcd) \end{aligned}$$

로부터 다음과 같이 나눈다.

⇒ block 1: (1), ab, ac, ad, bc, bd, cd, abcd → 주블록(principal block)((1)을 포함 하는 블록)

block 2: a, b, c, d, abc, abd, acd, bcd

(2) 4개의 block으로 나누는 경우 (2중교락)

block과 교락되는 요인이 $4-1=3$ 개이다.

이 중 2개의 요인이 정해지면 나머지 한 개는 자동적으로 정해진다.

먼저 3인자 교호작용 ABD 와 ACD 를 block과 교락시키기로 한다면

$$\begin{aligned}
ABD &= \frac{1}{8}(a-1)(b-1)(c+1)(d-1) \\
&= \frac{1}{8}\{(a+b+d+ac+bc+cd+abd+abcd)\} \\
&\quad -((1)+c+ab+ad+bd+abc+acd+bcd)
\end{aligned}$$

$$\begin{aligned}
ACD &= \frac{1}{8}(a-1)(b+1)(c-1)(d-1) \\
&= \frac{1}{8}\{(a+c+d+ab+bc+bd+acd+abcd)\} \\
&\quad -((1)+b+ac+ad+cd+abc+abd+bcd)
\end{aligned}$$

로부터 다음과 같이 4개의 block으로 나눈다.

block 1	block 2	block 3	block 4
$\begin{pmatrix} ABD에서 \oplus \\ ACD에서 \oplus \end{pmatrix}$	$\begin{pmatrix} ABD에서 \oplus \\ ACD에서 \ominus \end{pmatrix}$	$\begin{pmatrix} ABD에서 \ominus \\ ACD에서 \oplus \end{pmatrix}$	$\begin{pmatrix} ABD에서 \ominus \\ ACD에서 \ominus \end{pmatrix}$
<div style="border: 1px solid black; padding: 5px; text-align: center;"> a d bc abcd </div>	<div style="border: 1px solid black; padding: 5px; text-align: center;"> b ac cd abd </div>	<div style="border: 1px solid black; padding: 5px; text-align: center;"> c ab bd acd </div>	<div style="border: 1px solid black; padding: 5px; text-align: center;"> (1) ad abc bcd </div>
(BC에서 \oplus)	(BC에서 \ominus)	(BC에서 \ominus)	(BC에서 \oplus)

(이 네 block 중 block 4가 principal block.)

이 경우 ABD 와 ACD 를 곱해서 생기는 요인인 BC 도 block과 교락되어 있다.

$$ABD \times ACD = A^2 BCD^2 = BC \quad (2^n \text{형 실험에서 } A^2 = B^2 = C^2 = D^2 = \dots = 1)$$

$$\begin{aligned}
BC &= \frac{1}{8}(a+1)(b-1)(c-1)(d+1) \\
&= \frac{1}{8}\{((1)+a+d+ad+bc+abc+bcd+abcd)\} \\
&\quad - (b+c+ab+ac+bd+cd+abd+acd)
\end{aligned}$$

◇ Kempthorne의 방법 (2^4 형 실험에서)

(1) 단독교락

$I = ABCD$: 정의대비(defining contrast)

$$L = x_1 + x_2 + x_3 + x_4 \pmod{2} \quad (x_i : i \text{ 번째 인자의 수준})$$

block 1	block 2
(1)	a
ab	b
ac	c
ad	d
bc	abc
bd	abd
cd	acd
abcd	bcd
$L = 0$	$L = 1$

<예> $ac \Rightarrow L = x_1 + x_2 + x_3 + x_4 = 1 + 0 + 1 + 0 = 2 \pmod{2} = 0 \Rightarrow \text{block 1}$
 $abd \Rightarrow L = x_1 + x_2 + x_3 + x_4 = 1 + 1 + 0 + 1 = 3 \pmod{2} = 1 \Rightarrow \text{block 2}$

$$S_{Block} = S_{A \times B \times C \times D} \quad (S_{Block} = \frac{(\text{block 1의 합계})^2}{8} + \frac{(\text{block 2의 합계})^2}{8} - \frac{(\text{전체합계})^2}{16})$$

(2) 2중교락

$$\begin{array}{ll} I = ABD & L_1 = x_1 + x_2 + x_4 \pmod{2} \\ I = ACD & L_2 = x_1 + x_3 + x_4 \pmod{2} \\ I = BC & L_3 = x_2 + x_3 \pmod{2} \end{array}$$

	x_1	x_2	x_3	x_4	L_1	L_2	L_3
(1)	0	0	0	0	0	0	0
a	1	0	0	0	1	1	0
b	0	1	0	0	1	0	1
c	0	0	1	0	0	1	1
d	0	0	0	1	1	1	0
ab	1	1	0	0	0	1	1
ac	1	0	1	0	1	0	1
ad	1	0	0	1	0	0	0
bc	0	1	1	0	1	1	0
bd	0	1	0	1	0	1	1
cd	0	0	1	1	1	0	1
abc	1	1	1	0	0	0	0
abd	1	1	0	1	1	0	1
acd	1	0	1	1	0	1	1
bcd	0	1	1	1	0	0	0
abcd	1	1	1	1	1	1	0

block 1	block 2	block 3	block 4
<div> (1) ad abc bcd </div>	<div> c ab bd acd </div>	<div> b ac cd abd </div>	<div> a d bc abcd </div>
$L_1 = L_2 = 0$	$L_1 = 0, L_2 = 1$	$L_1 = 1, L_2 = 0$	$L_1 = L_2 = 1$
$(L_3 = 0)$	$(L_3 = 1)$	$(L_3 = 1)$	$(L_3 = 0)$

$$S_{Block} = S_{A \times B \times D} + S_{A \times C \times D} + S_{B \times C}$$

$$(S_{Block} = \sum_{i=1}^4 \frac{(block\ i\ 의\ 합계)^2}{4} - \frac{(전체합계)^2}{16})$$

3.4 일부실시법

인자의 수가 많아지면 인자의 처리조합의 수는 급격히 증가하고 1반복의 실험만 실시하더라도 실험의 회수는 기하급수적으로 늘어나게 되어 실험의 실시가 곤란하게 된다. 그러므로 필요로 하는 요인에 대한 정보를 얻기 위하여 의미가 적은 고차의 교호

작용을 희생시켜서 실험의 회수를 적게하자는 것이 일부실험시법(fractional factorial design)이다. 일부실험시법은 1반복의 일부만을 실시하는 실험배치를 의미한다고도 할 수 있다.

3.4.1 2^n 형의 $\frac{1}{2}$ 실험시법

예로 2^4 형의 $\frac{1}{2}$ 실험시법(2^{4-1} 실험)

2^4 형 실험에서 $I=ABCD$ 로 하면

block 1: (1), ab , ac , ad , bc , bd , cd , $abcd$

block 2: a , b , c , d , abc , abd , acd , bcd

1) block 1을 사용하는 경우 ($I=+ABCD$)

$$\begin{aligned} A &= \frac{1}{4} \{ (A\text{인자의 수준1의 데이터의 합}) - (A\text{인자의 수준0의 데이터의 합}) \} \\ &= \frac{1}{4} \{ (ab+ac+ad+abcd) - ((1)+bc+bd+cd) \} \end{aligned}$$

$$\begin{aligned} BCD &= \frac{1}{4} \{ (B, C, D\text{인자의 수준합이 홀수인 데이터의 합}) \\ &\quad - (B, C, D\text{인자의 수준합이 짝수인 데이터의 합}) \} \\ &= \frac{1}{4} \{ (ab+ac+ad+abcd) - ((1)+bc+bd+cd) \} \end{aligned}$$

$$\therefore A = BCD$$

마찬가지로 $B = ACD$, $C = ABD$, $D = ABC$,

$$AB = CD, AC = BD, AD = BC$$

$$A \times (ABCD) = A^2 BCD = BCD$$

$$AB \times (ABCD) = A^2 B^2 CD = CD$$

요 인	ϕ
$A(=BCD)$	1
$B(=ACD)$	1
$C(=ABD)$	1
$D(=ABC)$	1
$AB(=CD)$ $AC(=BD)$ $AD(=BC)$	1 1 1
T	7

2) block 2를 사용하는 경우 ($I = -ABCD$)

$$A = \frac{1}{4} \{ (a + abc + abd + acd) - (b + c + d + bcd) \} = -BCD$$

$$B = -ACD, \quad C = -ABD, \quad D = -ABC, \\ AB = -CD, \quad AC = -BD, \quad AD = -BC$$

3.4.2 2^n 형의 $\frac{1}{4}$ 실시법 (2^{n-2} 실험)

이중교락을 통하여 네 개의 block으로 만들어 이 중 하나를 택한다.

<예> 2^5 형 실험에서 $I = ABCD = ACE = BDE$

요 인	ϕ
A	1
B	1
C	1
D	1
E	1
AB AD	1 1
T	7

(2차 이상의 교호작용을 무시할 수 있는 경우)

2^{n-p} 실험 ($p \geq 3$)도 있음

<예> 2^{7-3} 실험: 인자의 수가 7개일 때, $2^7 = 128$ 회의 $\frac{1}{8}$ 인 16회만 실험

3.4.3 3ⁿ형의 일부 실시법

(1) 3ⁿ⁻¹ 실험 (3ⁿ형의 $\frac{1}{3}$ 실시)

<예> 3³⁻¹ 실험

3³형 실험에서 $I = ABC$ 를 block과 교락시키는 경우

$$L = x_1 + x_2 + x_3 \pmod{3}$$

block 1 (L=0): 000, 012, 021, 102, 111, 120, 201, 210, 222

block 2 (L=1): 001, 010, 022, 100, 112, 121, 202, 211, 220

block 3 (L=2): 002, 011, 020, 101, 110, 122, 200, 212, 221

이 세 block 중 한 개만 택하여 실험한다.

별명관계:

$$\begin{aligned} A &= A(ABC) = A^2 BC = (A^2 BC)^2 = A^4 B^2 C^2 = AB^2 C^2 \\ A &= A(ABC)^2 = A^3 B^2 C^2 = B^2 C^2 = (B^2 C^2)^2 = B^4 C^4 = BC \\ B &= B(ABC) = AB^2 C \\ B &= B(ABC)^2 = AC \\ C &= C(ABC) = ABC^2 \\ C &= C(ABC)^2 = AB \end{aligned}$$

$$\begin{aligned} AB^2 &= AB^2(ABC) = AC^2 \\ AB^2 &= AB^2(ABC)^2 = BC^2 \end{aligned}$$

$$\Rightarrow \begin{cases} A = AB^2 C^2 = BC \\ B = AB^2 C = AC \\ C = ABC^2 = AB \\ AB^2 = AC^2 = BC^2 \end{cases}$$

요 인	ϕ
$A(AB^2C^2, BC)$	2
$B(AB^2C, AC)$	2
$C(ABC^2, AB)$	2
$AB^2(AC^2, BC^2)$	2
T	8

(2) 3^{n-2} 실험 (3^n 형의 $\frac{1}{9}$ 실시)

<예>

1) 3^{3-2} 실험

3^3 형 실험에서 $I = \frac{AB^2C}{(X)} = \frac{AC^2}{(Y)} = \frac{AB}{(XY)} = \frac{BC}{(XY^2)}$ 로 하여 한 block만 택하면

$$AX = A(AB^2C) = A^2B^2C = (A^2B^2C)^2 = A^4B^4C^2 = ABC^2$$

$$AY = A(AC^2) = A^2C^2 = (A^2C^2)^2 = A^4C^4 = AC$$

$$A(XY) = A(AB) = A^2B = (A^2B)^2 = A^4B^2 = AB^2$$

$$A(XY^2) = A(BC) = ABC$$

$$AX^2 = A(AB^2C)^2 = A^3B^4C^2 = BC^2$$

$$AY^2 = A(AC^2)^2 = A^3C^4 = C$$

$$A(XY)^2 = A(AB)^2 = A^3B^2 = B^2 = (B^2)^2 = B^4 = B$$

$$A(XY^2)^2 = A(BC)^2 = AB^2C^2$$

별명관계:

$$\begin{array}{cccccccc} & (AX) & (AY) & (A(XY)) & (A(XY^2)) & (AX^2) & (AY^2) & (A(XY)^2) & (A(XY^2)^2) \\ A & = ABC^2 & = AC & = AB^2 & = ABC & = BC^2 & = C & = B & = AB^2C^2 \\ (B & = AC & = ABC^2 & = AB^2 & = BC^2 & = ABC & = AB^2C^2 & = A & = C) \\ \vdots & & & & & & & & \\ (BC^2 & = A & = ABC & = AB^2C^2 & = B & = ABC^2 & = AB^2 & = AC & = C) \\ \vdots & & & & & & & & \end{array}$$

요 인	ϕ
$A(ABC^2, AC, AB^2, ABC, BC^2, C, B, AB^2 C^2)$	2
T	2

실제로 사용할 수 없음.

- 2) 3^{4-2} 실험: 인자의 수가 4개 일 때, $3^4 = 81$ 회의 $\frac{1}{9}$ 인 9회만 실험

IV 요인 Microarrays

4.1 요인 Microarrays 실험

4.1.1 2^2 요인실험

2^2 실험에서 우리는 2개 수준 각각에 2개의 요인 F_1 과 F_2 를 가지고 있다. 전과같이 각 요인의 두개 수준은 0과 1로 코드를 한다. F_1 과 F_2 에 더하여 2개 이상의 요인은 있다. 2개의 방해요인이 있다. 두 수준의 색깔과 배열 또는 b수준의 슬라이드, 여기서 b는 실험에서 사용된 전체수의 슬라이드를 표시한다. 따라서, 우리는 F_d 요인의 두 개 수준의 색깔을 0(cy5:빨강)과 1(cy3:녹색)로 된 코드로 표시할 것이다. 데이터의 모델은 아래와 같이 쓰인다.

$$(y_{i1} - y_{i2}) = (\delta_1 - \delta_2) + (\tau_{t_{i1}} - \tau_{t_{i2}}) + \epsilon_{i1i2}$$

i 번째 블록으로부터의 블록내 비교 $y_{i1} - y_{i2}$ 는 i 번째 배열에서 교배시킨 두 처리조합 t_{i1} 과 t_{i2} 의 \log_2 표현비와 일치한다. $\delta_1 - \delta_2$ 는 색깔 효과, τ_j 는 j 번째 처리조합의 효과, 그리고 ϵ_{i1i2} 은 평균이 0, 분산이 σ^2 를 갖는 독립적으로 분포된 확률오차이다. \log_2 표현비는 표준화와 원인수정방법에 종속된다고 추정된다. 블록효과는 블록내 비교 $y_{i1} - y_{i2}$ 로 배출되어진다는 것을 주의해라. 색깔 효과 $\delta_1 - \delta_2$ 는 처리들이 색깔에 직교한다면 주효과와 교호작용의 요인 추정으로부터 자동적으로 배출되어진다. 다시 말하면, 색깔의 두 수준이 각각의 처리조합과 함께 동등하게 나타난다면, 위의 모델은 각 유전자가 개별적으로 적용할 수 있음을 주의해라. 따라서 처리 효과, 색깔효과는 특별한 유전자이다.

각 반복에서 교락된 효과에 따라 교락의 방법을 사용하여 얻어진 단일 반복 모형은 아래의 표4.1에 주어졌다.

<표 4.1> 2^2 요인 모델

모델	블록	교락된 효과
D_{11}	[00, 01],[10, 11]	F_1
D_{12}	[00, 10][01, 11]	F_2
D_{13}	[00, 11],[01, 10]	F_1F_2

위와 같은 2^2 실험에서 반복모형 모델의 표현을 다른 방법으로 접근해보면 \otimes 를 요인 별 부호곱이라 하고, 0을 -, 1을 +부호로 한다. 주효과는 블록별 부호곱하여 “+”인 것이고, 교호작용은 블록별로 부호를 곱한 결과에서 2요인 이상 부호를 곱하여 “+”인 것으로 한다. 예를 들어 아래와 같다.

$$D_{11}: 00\otimes 01 = (--\otimes -+) = (+, -) \Rightarrow F_1$$

$$D_{12}: 00\otimes 10 = (--\otimes +-) = (-, +) \Rightarrow F_2$$

$$D_{13}: 00\otimes 11 = (--\otimes ++) = (-, -) \Rightarrow F_1F_2$$

위의 결과는 표1과 같다.

microarray 모델은 행요인 F_g (색깔)과 열요인 블록(슬라이드)을 갖는 행렬 모델이다. F_g 의 수준 0(빨강)과 1(녹색)은 첫 번째와 두 번째열에 각각 해당한다. D_{11} 과 D_{12} 을 사용함으로써 우리는 열에 4개의 슬라이드가 해당하는 뒤따라오는 microarray 모델을 얻는다.

00	11	10	01
01	10	00	11

첫 번째 행에서의 모든 처리조합은 cy5로 분류하고 두 번째 행은 cy3로 분류한다. 예를 들면, 첫 번째 열 [00, 01]은 각각 빨강과 녹색색깔로 구분된 처리조합 00과 01로 교배된 하나의 슬라이드를 나타낸다. 열에서의 처리조합은 행들의 처리조합에 직교로 배열되어졌다. 오직 사용된 슬라이드의 수는 처리조합수가 복합적일 때 행과 처리의 직교는 얻어졌다. 위의 microarray 모델은 50%의 효율을 가진 주효과 F_1 과 F_2 을 추정한다. 그에 반해, 교호작용 F_1F_2 는 D_{11} 과 D_{12} 에서 동시에 교락되어지지 않았기 때문에 완전한 효율을 가지고 추정한다. 사실 이 모델은 Kerr and hurchill(2001a)에 의해서

주어진 common loop(CL)모델이다. 뒤따라오는 모델은 D_{13} 의 두 번의 반복을 취함으로써 Landgrebe et al.(2005)에 의해서 언급된 cross-swap(CS)모델이다.

00	01	11	10
11	10	00	01

Landgrebe et al.(2005)는 16개의 슬라이드에서 두개의 모델을 고려했다. 한 모델은 CL의 4개의 반복을 사용하는 것과 또 하나는 각각 CL과 CS를 두 번씩 반복해 사용하는 것이다. 사실, D_{11} , D_{12} , D_{13} 의 적절한 결합은 F_1 , F_2 그리고 F_1F_2 의 추정의 원하는 효율과 함께 추가적인 모형을 얻을 수 있다.

예를 들면, D_{11} 의 3번 반복, D_{12} 의 2번 반복 그리고 D_{13} 의 3번 반복을 취함으로써 16개의 슬라이드를 갖는 뒤따라오는 microarray 모델을 낳는다.

00	10	01	11	00	10	10	01	00	11	01	11	00	01	11	10
01	11	00	10	01	11	00	11	10	01	10	00	11	10	00	01

이 모델은 각각 62.5%, 75% 그리고 62.5% 효율을 가진 F_1 , F_2 그리고 F_1F_2 를 추정한다. 모델의 선택은 바라는 효율 또는 대상의 대조 추정의 정확에 의존할 것이다.

4.1.2 2^3 요인실험

교각의 방법을 사용하여 얻어진 2^3 실험을 위한 모든 가능한 단일 반복 모델은 표4.2에 표현됐다.

<표 4.2> 2^3 요인 모델

모델	블록	교락된 효과
D_{21}	[000,100],[001,101],[010,110],[011,111]	F_2, F_3, F_2F_3
D_{22}	[000,010],[001,011],[100,110],[101,111]	F_1, F_3, F_1F_3
D_{23}	[000,001],[010,011],[100,101],[110,111]	F_1, F_2, F_2F_3
D_{24}	[000,111],[001,110],[010,100],[011,100]	F_1F_2, F_1F_3, F_2F_3
D_{25}	[000,011],[001,010],[101,110],[100,111]	$F_1, F_2F_3, F_1F_2F_3$
D_{26}	[000,101],[001,100],[011,110],[010,111]	$F_2, F_1F_3, F_1F_2F_3$
D_{27}	[000,110],[010,100],[011,101],[001,111]	$F_3, F_1F_2, F_1F_2F_3$

2^3 실험 역시 반복모형 모델의 표현을 다른 방법으로 접근해보면 main 블록을 000과 나머지 7개중에서 1개를 선택하고 나머지 블록은 main 블록의 요인합을 mod2로 하여 main 블록과 동치가 되도록 나머지 6개를 2개씩 조합한다. 예를 들어 D_{23} 의 main 블록은 [000, 001]이고 나머지 블록은 $(000 \oplus 001)(\text{mod } 2) = 001$ 되게 2개씩 조합하여 [010, 011] [100, 101] [110, 111] 3개의 블록을 얻을 수 있다. 이것을 표1에서 예를 들었던 것처럼 곱하게 되면 표2와 같은 결과를 얻을 수 있다.

$$D_{21} : (000 \otimes 100) = (--- \otimes +--) = (- + +)$$

$$F_2, F_3, F_2F_3$$

$$D_{24} : (000 \otimes 111) = (--- \otimes +++) = (- - -)$$

$$F_1F_2, F_1F_3, F_2F_3$$

$$D_{27} : (000 \otimes 110) = (--- \otimes ++-) = (- - +)$$

$$F_3, F_1F_2, F_1F_2F_3$$

D_{24}, D_{25}, D_{26} 그리고 D_{27} 을 사용함으로써 우리는 16개의 슬라이드를 가진 뒤따라오는 모델을 얻을 수 있다.

111	110	010	011	000	001	101	100	000	001	011	010	110	100	101	111
000	001	101	100	011	010	110	111	101	100	110	111	000	010	011	001

세 개의 주효과들은 75% 효율성으로 추정되어졌고, 세 개의 2인자 교호작용들은 50% 효율성으로 추정되어졌다. 게다가 microarray 모형은 적절한 결합으로 표2의 모델들은 얻어질 수 있다. 예를 들면, 만약 2인자 교호작용들이 주효과와 비교에서 더 작은 분산을 가지고 추정되어졌다면 D_{21} , D_{22} 그리고 D_{23} 는 종종 더 많은 결합을 사용했을 것이다. 대개 3인자와 더 높은 서열의 교호작용들은 결여된 것으로 가정할 것이다. 그러나, 적절한 microarray 모형은 3인자 교호작용 또한 효율적으로 추정되어졌다면 표2의 모델로부터 더 쉽게 구성할 수 있다.

4.1.3 2^4 요인실험

교락법을 사용하여 얻어진 2^4 실험에서의 교락요인

- D_{61} (0000, 0001) (0010, 0011) (0100, 0101) (0110, 0111) F_1, F_2, F_3, F_1F_2
 (1000, 1001) (1010, 1011) (1100, 1101) (1110, 1111) $F_1F_3, F_2F_3, F_1F_2F_3$
 요인별 부호곱 = (+ + + -)
- D_{62} (0000, 0010) (0001, 0011) (0100, 0110) (0101, 0111) F_1, F_2, F_4, F_1F_2
 (1000, 1010) (1001, 1011) (1100, 1110) (1101, 1111) $F_1F_4, F_2F_4, F_1F_2F_4$
 요인별 부호곱 = (+ + - +)
- D_{63} (0000, 0011) (0001, 0010) (0100, 0111) (0101, 0110) F_1, F_2, F_1F_2, F_3F_4
 (1000, 1011) (1001, 1010) (1100, 1111) (1101, 1110) $F_1F_3F_4, F_2F_3F_4, F_1F_2F_3F_4$
 요인별 부호곱 = (+ + - -)
- D_{64} (0000, 0100) (0001, 0101) (0010, 0110) (0011, 0111) F_1, F_3, F_4, F_1F_3
 (1000, 1100) (1001, 1101) (1010, 1110) (1011, 1111) $F_1F_4, F_3F_4, F_1F_3F_4$
 요인별 부호곱 = (+ - + +)
- D_{65} (0000, 0101) (0001, 0100) (0010, 0111) (0011, 0110) F_1, F_3, F_1F_3, F_2F_4
 (1000, 1101) (1001, 1100) (1010, 1111) (1011, 1110) $F_1F_2F_4, F_2F_3F_4, F_1F_2F_3F_4$

요인별 부호곱 = (+ - + -)

D_{66} (0000, 0110) (0001, 0111) (0010, 0100) (0011, 0101) F_1, F_4, F_1F_4, F_2F_3
 (1000, 1110) (1001, 1111) (1010, 1100) (1011, 1101) $F_1F_2F_3, F_2F_3F_4, F_1F_2F_3F_4$

요인별 부호곱 = (+ - - +)

D_{67} (0000, 0111) (0001, 0110) (0010, 0101) (0011, 0100) $F_1, F_2F_3, F_2F_4, F_3F_4$
 (1000, 1111) (1001, 1110) (1010, 1101) (1011, 1100) $F_1F_2F_3, F_1F_2F_4, F_1F_3F_4$

요인별 부호곱 = (+ - - -)

D_{68} (0000, 1000) (0001, 1001) (0010, 1010) (0011, 1011) F_2, F_3, F_4, F_2F_3
 (0100, 1100) (0101, 1101) (0110, 1110) (0111, 1111) $F_2F_4, F_3F_4, F_2F_3F_4$

요인별 부호곱 = (- + + +)

D_{69} (0000, 1001) (0001, 1000) (0010, 1011) (0011, 1010) F_2, F_3, F_1F_4, F_2F_3
 (0100, 1101) (0101, 1101) (0110, 1111) (0111, 1110) $F_1F_2F_4, F_1F_3F_4, F_1F_2F_3F_4$

요인별 부호곱 = (- + + -)

D_{610} (0000, 1010) (0001, 1011) (0010, 1000) (0011, 1001) F_2, F_4, F_1F_3, F_2F_4
 (0100, 1110) (0101, 1111) (0110, 1100) (0111, 1101) $F_1F_2F_3, F_1F_3F_4, F_1F_2F_3F_4$

요인별 부호곱 = (- + - +)

D_{611} (0000, 1011) (0001, 1010) (0010, 1001) (0011, 1000) $F_2, F_1F_3, F_1F_4, F_3F_4$
 (0100, 1111) (0101, 1110) (0110, 1101) (0111, 1100) $F_1F_2F_3, F_1F_2F_4, F_2F_3F_4$

요인별 부호곱 = (- + - -)

D_{612} (0000, 1100) (0001, 1101) (0010, 1110) (0011, 1111) F_3, F_4, F_1F_2, F_3F_4
 (0100, 1000) (0101, 1001) (0110, 1010) (0111, 1011) $F_1F_2F_3, F_1F_2F_4, F_1F_2F_3F_4$

요인별 부호곱 = (- - + +)

D_{613} (0000, 1101) (0001, 1100) (0010, 1111) (0011, 1110) $F_3, F_1F_2, F_1F_4, F_2F_4$
 (0100, 1001) (0101, 1000) (0110, 1011) (0111, 1010) $F_1F_2F_3, F_1F_3F_4, F_2F_3F_4$

요인별 부호곱 = (- - + -)

D_{614} (0000, 1110) (0001, 1111) (0010, 1100) (0011, 1101) $F_4, F_1F_2, F_1F_3, F_2F_3$
 (0100, 1010) (0101, 1011) (0110, 1000) (0111, 1001) $F_1F_2F_4, F_1F_3F_4, F_2F_3F_4$

요인별 부호값 = (- - - +)

D_{615} (0000, 1111) (0001, 1110) (0010, 1101) (0011, 1100) $F_1F_2, F_1F_3, F_1F_4, F_2F_3$
 (0100, 1011) (0101, 1010) (0110, 1001) (0111, 1000) $F_2F_4, F_3F_4, F_1F_2F_3F_4$

요인별 부호값 = (- - - -)

위의 main design을 적당히 조합하여 배열함으로써 2^3 실험에서와 같은 모델들을 얻을 수 있다.

4.1.4 3×2 요인 Microarrays

3×2실험을 위한 요인 모델은 Lewis and Tuck(1985)에 의해서 작성되어졌다. 일반화된 순환 방법의 구성에 기본을 둔 이러한 모델들은 표4.3에 표현됐다.

<표 4.3> 3×2요인을 위한 모델

모델	블록	추정의 효율		
		F_1	F_2	F_1F_2
D_{41}	[00,01],[10,11],[20,21]	0	1	1
D_{42}	[00,11],[10,21],[20,01],[01,10],[11,20],[21,00]	0.75	1	0.25
D_{43}	[00,10],[10,20],[20,00],[01,11],[11,21],[21,01]	0.75	0	0.75

<참고>

(1) generalized cyclic method

- i) 10씩 합하여 처음으로 순환될 때까지 계속
 - ii) 10 10 11씩 순서대로 합을 순환될 때까지 계속
- 위에서 10을 20, 11을 21로 하여도 좋음

$$1) D_{41} (00, 01) \oplus 10(\text{mod}(3, 2)) = (10, 11) \oplus 10 \text{ mod}(3, 2)$$

$$= (20, 21) \oplus 10 \text{ mod}(3, 2) = (00, 01) : \text{main block}$$

$$D'_{41} (00, 01) \oplus 20 \text{ mod}(3, 2) = (20, 21) \oplus 20 \text{ mod}(3, 2)$$

$$= (10, 11) \oplus 20 \text{ mod}(3, 2) = (00, 01) \text{ 순환됨}$$

$$\therefore D_{41} \cong D'_{41}$$

$$\begin{aligned}
2) \quad D_{42} \quad (00, 11) \oplus 10(\text{mod}(3, 2)) &= (10, 21) \oplus 10 \text{ mod}(3, 2) \\
&= (20, 01) \oplus 11 \text{ mod}(3, 2) = (01, 10) \oplus 10 \text{ mod}(3, 2) \\
&= (11, 20) \oplus 10 \text{ mod}(3, 2) = (21, 00) \oplus 11 \text{ mod}(3, 2) \\
&= (00, 11) : \text{순환}
\end{aligned}$$

main block

(2) 다른 main design

$$i) \quad D_{44} \quad (00, 20) \quad (10, 00) \quad (20, 10)$$

[처음부터 $\oplus 10 \text{ mod}(3, 2)$, $\oplus 10 \text{ mod}(3, 2)$, $\oplus 11 \text{ mod}(3, 2)$ 순환]

(01, 21) (11, 01) (21, 11) (00, 20) [순환]

[처음부터 $\oplus 10 \text{ mod}(3, 2)$, $\oplus 10 \text{ mod}(3, 2)$, $\oplus 11 \text{ mod}(3, 2)$]

$$ii) \quad D_{45} \quad (00, 21) \quad (10, 01) \quad (20, 11) \quad (01, 20) \quad (11, 00) \quad (21, 10) \quad (00, 21) \text{ [순환]}$$

[처음부터 $\oplus 10$, $\oplus 10$, $\oplus 11$, $\oplus 10$, $\oplus 10$, $\oplus 11$ 순환]

* 여기서 $D_{44} \cong D_{43}$ $D_{45} \cong D_{42}$ 이다.

따라서 3 block으로 된 BS design은 5개가 아니고 3개이며

6 block으로된 XL, AL design은 5개가 아니고 3개임을 알 수 있다.

예를 들면, D_{42} 는 6개의 슬라이드를 가진 뒤따라오는 모델이다.

00	10	20	01	11	21
11	21	01	10	20	00

Landgrebe et al.(2005)는 D_{41} 는 BS모델과 같이 그리고 D_{42} , D_{43} 는 각각 XL, AL과 같은 모델의 두 번의 반복을 언급했다. 게다가 모델들은 D_{41} , D_{42} , D_{43} 의 적절한 결합에 의해서 얻어질 수 있다.

4.2 Visual C++를 이용한 프로그램 예제

<예> 2^5 -요인실험

아래 코딩부분은 main block를 표시해주는 부분이다.


```

if(num == 5)
{
    int flag = 0;
    m_adata = new int[temp];
    m_bdata = new int[temp];
    m_cdata = new int[temp];
    m_ddata = new int[temp];
    m_edata = new int[temp];

    for(int i=0; i<temp; i++)
    {
        if(i<16) m_adata[i] = 0;
        else m_adata[i] = 1;

        if(i%2) m_edata[i] = 1;
        else m_edata[i] = 0;
    }
    for(int b=0; b<temp; b+=8)
    {
        if(flag == 0)
        {
            m_bdata[b] = 0;
            m_bdata[b+1] = 0;
            m_bdata[b+2] = 0;
            m_bdata[b+3] = 0;
            m_bdata[b+4] = 0;
            m_bdata[b+5] = 0;
            m_bdata[b+6] = 0;
            m_bdata[b+7] = 0;

            flag = 1;
        }
        else
        {
            m_bdata[b] = 1;
            m_bdata[b+1] = 1;
            m_bdata[b+2] = 1;
            m_bdata[b+3] = 1;
            m_bdata[b+4] = 1;
            m_bdata[b+5] = 1;
            m_bdata[b+6] = 1;
            m_bdata[b+7] = 1;

            flag = 0;
        }
    }
}

```

```

}

    }
    for(int c=0; c<temp; c+=4)
    {
        if(flag == 0)
        {
            m_cdata[c] = 0;
            m_cdata[c+1] = 0;
            m_cdata[c+2] = 0;
            m_cdata[c+3] = 0;
            flag = 1;
        }
        else
        {
            m_cdata[c] = 1;
            m_cdata[c+1] = 1;
            m_cdata[c+2] = 1;
            m_cdata[c+3] = 1;
            flag = 0;
        }
    }
    for(int d=0; d<temp; d+=2)
    {
        if(flag == 0)
        {
            m_ddata[d] = 0;
            m_ddata[d+1] = 0;
            flag = 1;
        }
        else
        {
            m_ddata[d] = 1;
            m_ddata[d+1] = 1;
            flag = 0;
        }
    }
}

```

아래 코딩부분은 인자들을 비교하여 요인함(\oplus)하여 mod2를 해준 결과부분을 나타내는 부분이다.

```
if(num == 5)
{
    m_totcounter = new int[itemp];
    for(int j=0; j<itemp; j++)
    {
        m_totcounter[j] = 0;
    }

    for(int i=0; i<itemp; i++)
    {
        for(int a=0; a<total; a++)
        {
            if(m_adata[i] == m_adata5[a] && m_bdata[i] == m_bdata5[a]
            && m_cdata[i] == m_cdata5[a] && m_ddata[i] == m_ddata5[a]
            && m_edata[i] == m_edata5[5])
            {
                m_totcounter[i] = m_totcounter[i] + 1;
            }
        }
    }
}
```

아래 코딩부분은 Experiment result부분으로 교락된 인자들을 찾아주는 부분이다.

```
if(num == 5)
{
    m_DataCreate = new CString[itemp];
    m_DataCreate_1 = new CString[itemp];
    for(int a=0; a<temp; a++)
    {
        m_DataCreate[a] = "";
        m_DataCreate_1[a] = "";
    }
    for(int i=1; i<temp; i++)
    {
        if(m_adata[i] == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(a) ";
        if(m_bdata[i] == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(b) ";
    }
}
```

```

if(m_cdata[i] == 0)

        m_DataCreate[i] = m_DataCreate[i] + "(c) ";
if(m_ddata[i] == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(d) ";
if(m_edata[i] == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(e) ";

////////////////////////////////////
if((m_adata[i]+m_bdata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,b) ";
if((m_adata[i]+m_cdata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,c) ";
if((m_adata[i]+m_ddata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,d) ";
if((m_adata[i]+m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,e) ";
if((m_bdata[i]+m_cdata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(b,c) ";
if((m_bdata[i]+m_ddata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(b,d) ";
if((m_bdata[i]+m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(b,e) ";
if((m_cdata[i]+m_ddata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(c,d) ";
if((m_cdata[i]+m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(c,e) ";
if((m_ddata[i]+m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(d,e) ";

////////////////////////////////////
if((m_adata[i] + m_bdata[i] + m_cdata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,b,c) ";
if((m_adata[i] + m_cdata[i] + m_ddata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,c,d) ";
if((m_adata[i] + m_ddata[i] + m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,d,e) ";
if((m_adata[i] + m_bdata[i] + m_ddata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,b,d) ";
if((m_adata[i] + m_bdata[i] + m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,b,e) ";
if((m_adata[i] + m_cdata[i] + m_edata[i])%2 == 0)
        m_DataCreate[i] = m_DataCreate[i] + "(a,c,e) ";

```

```

////////////////////////////////-----
        if((m_bdata[i] + m_cdata[i] + m_ddata[i])%2 == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(b,c,d) ";
        if((m_bdata[i] + m_cdata[i] + m_edata[i])%2 == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(b,c,e) ";
        if((m_bdata[i] + m_ddata[i] + m_edata[i])%2 == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(b,d,e) ";

//-----
        if((m_cdata[i] + m_ddata[i] + m_edata[i])%2 == 0)
            m_DataCreate[i] = m_DataCreate[i] + "(c,d,e) ";

////////////////////////////////
        if((m_adata[i] + m_bdata[i] + m_cdata[i] + m_ddata[i])%2 == 0)
            m_DataCreate_1[i] = m_DataCreate_1[i] + "(a,b,c,d) ";
        if((m_adata[i] + m_bdata[i] + m_cdata[i] + m_edata[i])%2 == 0)
            m_DataCreate_1[i] = m_DataCreate_1[i] + "(a,b,c,e) ";
        if((m_adata[i] + m_cdata[i] + m_ddata[i] + m_edata[i])%2 == 0)
            m_DataCreate_1[i] = m_DataCreate_1[i] + "(a,c,d,e) ";
        if((m_bdata[i] + m_cdata[i] + m_ddata[i] + m_edata[i])%2 == 0)
            m_DataCreate_1[i] = m_DataCreate_1[i] + "(b,c,d,e) ";

////////////////////////////////
if((m_adata[i] + m_bdata[i] + m_cdata[i] + m_ddata[i] + m_edata[i])%2 == 0)
    m_DataCreate_1[i] = m_DataCreate_1[i] + "(a,b,c,d,e) ";
    //MessageBox(m_DataCreate[i]);

    }

}

```

<그림 1> 프로그램에 의한 2^5 요인실험에 대한 output

Properties

수준 : 0, 1 → 2수준

ex) 2ⁿ승

요인 : 5

Input Number

Exit

Data Init

5개의 Data를 요인 비교하였습니다.

Element Compare

Element Create

Element Total

Information

Number : 31

Experiment Result :

(a,b) (a,c) (a,d) (a,e) (b,c) (b,d) (b,e) (c,d) (c,e) (d,e)

(a,b,c,d) (a,b,c,e) (a,c,d,e) (b,c,d,e)

Experiment Result

Num	a	b	c	d	e	f	Total
0	0	0	0	0	0		0
1	0	0	0	0	1		16
2	0	0	0	1	0		16
3	0	0	0	1	1		16
4	0	0	1	0	0		16
5	0	0	1	0	1		16
6	0	0	1	1	0		16
7	0	0	1	1	1		16
8	0	1	0	0	0		16
9	0	1	0	0	1		16
10	0	1	0	1	0		16
11	0	1	0	1	1		16
12	0	1	1	0	0		16
13	0	1	1	0	1		16
14	0	1	1	1	0		16
15	0	1	1	1	1		16
16	1	0	0	0	0		16
17	1	0	0	0	1		16
18	1	0	0	1	0		16
19	1	0	0	1	1		16
20	1	0	1	0	0		16
21	1	0	1	0	1		16
22	1	0	1	1	0		16
23	1	0	1	1	1		16
24	1	1	0	0	0		16

Num	a'	b'	c'	d'	e'	f'
0	0	0	0	0	1	
1	0	0	0	0	0	
2	0	0	0	1	1	
3	0	0	1	0	0	
4	0	0	1	0	1	
5	0	0	1	1	0	
6	0	0	1	1	1	
7	0	1	0	0	0	
8	0	1	0	0	1	
9	0	1	0	1	0	
10	0	1	0	1	1	
11	0	1	1	0	0	
12	0	1	1	0	1	
13	0	1	1	1	0	
14	0	1	1	1	1	
15	1	0	0	0	0	
16	1	0	0	0	1	
17	1	0	0	1	0	
18	1	0	0	1	1	
19	1	0	1	0	0	
20	1	0	1	0	1	
21	1	0	1	1	0	
22	1	0	1	1	1	
23	1	1	0	0	0	
24	1	1	0	0	1	

Num	Element Create
0	
1	(a) (b) (c) (d) (a,b) (a,c) (a,d) (b,c) (b,d) (c,d) (a,b,c) (a,c,d)
2	(a) (b) (c) (e) (a,b) (a,c) (a,e) (b,c) (b,e) (c,e) (a,b,c) (a,b,e)
3	(a) (b) (c) (a,b) (a,c) (b,c) (d,e) (a,b,c) (a,d,e) (b,d,e) (c,d,e)
4	(a) (b) (d) (e) (a,b) (a,d) (a,e) (b,d) (b,e) (d,e) (a,d,e) (a,b,d)
5	(a) (b) (d) (a,b) (a,d) (b,d) (c,e) (a,b,d) (a,c,e) (b,c,e) (c,d,e)
6	(a) (b) (e) (a,b) (a,e) (b,e) (c,d) (a,c,d) (a,b,e) (b,c,d) (c,d,e)
7	(a) (b) (a,b) (c,d) (c,e) (d,e) (a,c,d) (a,d,e) (a,c,e) (b,c,d) (b,c,e)
8	(a) (c) (d) (e) (a,c) (a,d) (a,e) (c,d) (c,e) (d,e) (a,c,d) (a,d,e)
9	(a) (c) (d) (a,c) (a,d) (b,e) (c,d) (a,c,d) (a,b,e) (b,c,e) (b,d,e)
10	(a) (c) (e) (a,c) (a,e) (b,d) (c,e) (a,b,d) (a,c,e) (b,c,d) (b,d,e)
11	(a) (c) (a,c) (b,d) (b,e) (d,e) (a,d,e) (a,b,d) (a,b,e) (b,c,d) (b,c,e)
12	(a) (d) (e) (a,d) (a,e) (b,c) (d,e) (a,b,c) (a,d,e) (b,c,d) (b,c,e)
13	(a) (d) (a,d) (b,c) (b,e) (c,e) (a,b,c) (a,b,e) (a,c,e) (b,c,d) (b,c,e)
14	(a) (e) (a,e) (b,c) (b,d) (c,d) (a,b,c) (a,c,d) (a,b,d) (b,c,e) (b,c,e)
15	(a) (b,c) (b,d) (b,e) (c,d) (c,e) (d,e) (a,b,c) (a,c,d) (a,d,e) (a,b,e)
16	(b) (c) (d) (e) (b,c) (b,d) (b,e) (c,d) (c,e) (d,e) (b,c,d) (b,c,e)
17	(b) (c) (d) (a,e) (b,c) (b,d) (c,d) (a,d,e) (a,b,e) (a,c,e) (b,c,d)
18	(b) (c) (e) (a,d) (b,c) (b,e) (c,e) (a,c,d) (a,d,e) (a,b,d) (b,c,e)
19	(b) (c) (a,d) (a,e) (b,c) (d,e) (a,c,d) (a,b,d) (a,b,e) (a,c,e) (b,c,e)
20	(b) (d) (e) (a,c) (b,d) (b,e) (d,e) (a,b,c) (a,c,d) (a,c,e) (b,d,e)
21	(b) (d) (a,c) (a,e) (b,d) (c,e) (a,b,c) (a,c,d) (a,d,e) (a,b,e) (b,c,e)
22	(b) (e) (a,c) (a,d) (b,e) (c,d) (a,b,c) (a,d,e) (a,b,d) (a,c,e) (b,c,e)
23	(b) (a,c) (a,d) (a,e) (c,d) (c,e) (d,e) (a,b,c) (a,b,d) (a,b,e) (b,c,e)
24	(c) (d) (e) (a,b) (c,d) (c,e) (d,e) (a,b,c) (a,b,d) (a,b,e) (b,c,e)

V 결론 및 토론

문헌에 의하면 바람직한 통계적인 성질을 갖는 microarray 디자인 블록 구성을 위하여 이용가능 한 블록 균형요인 디자인이 소개 되어 있다, 특히, 2수준 요인에 대한 전형적인 교각 모델들이 microarray 디자인으로는 효율적이라는 사실을 쉽게 알 수 있다.

Lewis와 Tuck(1985), Gupta(1987)에 의해 표로 만들어진 블록크기 2인 블록 모델들이 효과적이라는 것을 입증 하였다. 요인의 수와 수준이 많아지면 실험의 회수가 늘어나며 실험의 실시가 어려워지는데 이러한 실험의 회수를 줄이는 방법에는 요인실험, 교각법, 일부실험법 등 여러 가지 방법들이 있다. 본 논문에서는 실험의 회수를 줄여보고자 교각법을 사용하고자 하였으며, 이러한 이론을 바탕으로 프로그래밍하여 나타내보았다. 추후 프로그램을 더 활성화시켜 일반화시켜 보도록 노력하겠다.

참고 문헌

- [1] Box, G.E.P. and Wilson, K.G. (1951). On the Experimental Attainment of Optimum Conditions, Journal of Royal Statistical Society, B, 13, 1-45
- [2] Chen, J. and Wu, C.F.J. (1991). Some Results on s^{n-k} Fractional Factorial Designs with Minimum Aberration or Optimal Moments, Annals of statistics, 19, 1028-1041
- [3] Cornfield, J. and Tukey, J.W. (1956). Average Values of Mean Squares in Factorials, Annals of Mathematical Statistics, 27, 907-949.
- [4] Daniel, C. and Wilcoxon, F. (1966). Factorial 2^{p-q} Plans Robust Against Linear and Quadratic Trends, Technometrics, 8, 259-278.
- [5] Fries, A. and Hunter, W.G. (1980). Minimum Aberration 2^{k-p} Designs, Technometrics, 22, 601-608.
- [6] Jones, B. (1980). Algorithm AS 156: Combining two component designs form a row-and-column design. Applied Statistics, 29, 334-337.
- [7] Montgomery, D.C. (1976). Design and Analysis of Experiments, John Wiely & Sons, New York.
- [8] Paired comparison designs for factorial experiments. Applied Statistics. 34:227-234.
- [9] Kshirsagar, A.M. (1966). Balanced factorial designs. Journal of the royal Statistical Society. B28:59-67
- [10] Raghavarao, D. (1971). Construcions and Combinatorial Problems in Design of Experiments. New York:Wiley.
- [11] Shah, B.V. (1960). Balanced factorial experiments. Annals of Mathematical Statistics.
- [12] Abraham, B., H. Chipman, and K. Vijayan. (1999). "Some Risks in the Construction and Analysis of Supersaturated Designs." Technometrics, Vol. 41, pp. 135-141.

- [13] Barnett, V., and T. Lewis (1994). Outliers in Statistical Data, 3rd edition. Wiley, New York.
- [14] Bingham, D., and R. R. Sitter. (1999). "Minimum Aberration Two-Level Fractional Factorial Split-Plot Designs." *Technometrics*, Vol. 41, pp. 62-70.
- [15] Bisgaard, S. (1998-1999). "Conditional Inference Chart for Small Unreplicated Two-Level Factorial Experiments." *Quality Engineering*, Vol. 11, pp. 267-271.
- [16] Bisgaard, S. (2000). "The Design and Analysis of $2^{k-p} \times 2^{q-r}$ Split Plot Experiments." *Journal of Quality Technology*, Vol. 32, pp. 39-56.
- [17] Box, G. E. P. (1992-1993). "Sequential Experimentation and Sequential Assembly of Designs." *Quality Engineering*, Vol. 5, No. 2, pp. 321-330.
- [18] Box, G. E. P. (1999). "Statistics as a Catalyst to Learning by Scientific Method Part II – A Discussion" (with Discussion), *Journal of Quality Technology*, Vol. 31, pp. 16-29.
- [19] Burdick, R. K., C. M. Borror, and D. C. Montgomery. (2003). "A Review of Methods for Measurement Systems Capability Analysis." *Journal of Quality Technology*, Vol. 35, No. 4, pp. 342-354.
- [20] Burdick, R. K., and F. A. Graybill (1992). *Confidence Intervals on Variance Components*. Dekker, New York.
- [21] Coleman, D. E., and D. C. Montgomery (1993). "A Systematic Approach to Planning for a Designed Industrial Experiment" (with Discussion). *Technometrics*, Vol. 35, pp. 1-27.
- [22] Cornell, J. A. (1990). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. 2nd edition. Wiley, New York.
- [23] Hamada, M., and N. Balakrishnan (1998). "Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals" (with Discussion). *Statistica Sinica*, Vol. 8, pp. 1-41.