[UCI]I804:24011-200000232935

*2006*年 *2*月

博士學位論文

블록 디자인에 의한 *cDNA Microarray* 실험에 관한 연구

朝鮮大學校 大學院

電算統計學科

崔 南 琇

목 차

ABSTRACT

A Study on cDNA Microarray Experiments based on Block Designs

BY Choi, Nam Soo.
Advisor : Prof. Choi, Kuey Chung
Department of Computer Science and Statistics,
Graduate School of Chosun University

Usually, can define that is action that raises change under several conditions about study subject who is fixed as important terminology that is used in society diapason such as experiment(experiment) physiotherapy, chemistry, biomedical engineering and observes and observe the phenomenon. Therefore, experimenter shall observes reason which make

change in result through experiment and examines closely. Causes that can explain what present state actually exists much innumerably, and also, various constraints with expense, time make perfect cause searching examination difficult. Accordingly, an actuality experiment between a lot of causes empirically or work which do to arrive in target that want illuminating attribute of leading persons handled in an experiment thus directly selecting some by dictionary information in several angles as purpose of an experiment speak can.

R if summarize basis principle.A.Fisher is scholar that is lead self-satisfaction which draft importance of experimental design in role and , adversely, statistics of statistics in experimental design while extends in the 1920s and the 30s and works in Rothamsted Experimental Station. He branch of advocated important principle those randomization(randomization), block anger (blocking), number of replication (replicates) be. And in formality of experimental design realization of problem and purpose establishment of an experiment, comprehension of constraint, suitable specification of the model etc. (that any leading person amounted interest when consider whether covariance(covariance) confirmed Itneun and model is crossing (crossed) model or is interest (nested) model because includes necessary leading persons out of leading persons and those leading persons reciprocal action in matrix to be included in matrix, and is interest model consider

Also, model whether fixed effect(fixed effect) model or probability (random) effect whether search and in each model whether normality despotic rule about error is sound process of consideration), decision (decision of number of replication In Accordance With path of confidence

interval or power of test required under constraint) of number of replication, randomization, selection of experimental design, a reserve experiment, source collection and analysis of variance follow.

In this paper, I studied desirable and reasonable balance block design for cDNA microarray design by an experiment once for all using traditional confounding method. Confirmed that traditional factor experimental design that is gotten from this paper is more useful than other what confounding method. Specially, suggested symmetrical block design for microarray experiment design than method that is suitable according to algorithm and circumstance, and generalized various microarray designs by latest literature.

Ⅰ. 서론(*Introduction*)

1.1 개요(*Outline*)

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 × 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 생물검정(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 만들기를 위해 마인드(mind)의 중요성의 대비에 대한 조화의 중요성을 강조했다. 그들은 중요성의 특별한 대비의 추정을 하는 한은 종합적인 능률에 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 추정 가능한 대비의 변동을 최소화하는 것을 찾는 것이고 중요성의 대비는 아니다 라고 명확히 제시했다. 그런데 요인 실험의 고전적인 설계법에서 모든 추정 가능한 대비는 동등한 중요성이 드물다는 것을 유념해야한다. 그러므로 추정가능 할 수도 또는 불가능할지도 모르는 다른 대비와 비교하여 더 높고 정확한 중요성에 따라 대비를 추정하는 균형요인배치법(balanced factorial design)을 선택하는 것은 당연한 것이다. 실제로 고전적인 교락법은 블록에 완전 또는 부분 교락한 좀 더 중요성이 적은 것을 희생시켜 중요성의 대비에 더 높고 정확하게 달성시키는 것을 확실하게 한다. 고전적인 교락법은 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 제안 한바 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 이용하므로서 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. 고전적인 교락된 설계법과 요인 microarrays에 일반화된 싸이클(cyclic) 설계법의 응용은 4절에서 소개한다. 이해를 돕기 위하여 2^2, 2^3 그리고 3×2 실험은 상세히 논의되었다. 그밖에 다른 실험에 대한 microarray 설계법은 유사하므로 생략 하였다.마지막으로 몇 5절에는 결론 및discussion을 넣었다.

Ⅱ. 기본 이론

2.1 확률화 블록계획법

2.1.1 선형 모형

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \ (i = 1 \cdots v,\ j = 1 \cdots b)$$

$$y_{11} = \mu + \tau_1 + \beta_1 + \epsilon_{11}$$

$$\vdots$$

$$y_{1b} = \mu + \tau_1 + \beta_b + \epsilon_{1b}$$

$$y_{21} = \mu + \tau_2 + \beta_1 + \epsilon_{21}$$

$$\vdots$$

$$y_{2b} = \mu + \tau_2 + \beta_b + \epsilon_{2b}$$

$$\vdots$$

$$y_{v1} = \mu + \tau_v + \beta_1 + \epsilon_{v1}$$

$$\vdots$$

$$y_{vb} = \mu + \tau_v + \beta_b + \epsilon_{vb}$$

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1b} \\ y_{21} \\ \vdots \\ y_{2b} \\ \vdots \\ y_{v1} \\ \vdots \\ y_{vb} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_v \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1b} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2b} \\ \vdots \\ \epsilon_{v1} \\ \vdots \\ \epsilon_{vb} \end{pmatrix} = X\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

2.1.2 최소자승법에 의한 정규방정식

1) $X^t X = \begin{pmatrix} vb & b1_v^t & v1_b^t \\ b1_v & bI_v & J_{v\times b} \\ v1_b & J_{b\times v} & vI_b \end{pmatrix}$

단 $1_v = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, $I_v = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & 0 & 1 \end{pmatrix}$

$J_{v \times b} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$

2) $\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{\epsilon}$,

정규방정식은 : $(X^t X)\hat{\boldsymbol{\theta}} = X^t \boldsymbol{y}$ 즉

$$\left.\begin{aligned} vb\hat{\mu} + b\sum_{i=1}^{v}\widehat{\tau_i} + v\sum_{i=1}^{b}\widehat{\beta_j} &= y_{..} = \sum_{i=1}^{v}\sum_{j=1}^{b} y_{ij} \\ b\hat{\mu} + b\widehat{\tau_i} + \sum_{j=1}^{b}\widehat{\beta_j} &= y_{i\cdot} = \sum_{j=1}^{b} y_{ij} \quad (i = 1 \cdots v) \\ v\hat{\mu} + \sum_{i=1}^{v}\widehat{\tau_i} + v\widehat{\beta_j} &= y_{\cdot j} = \sum_{i=1}^{v} y_{ij} \quad (j = 1 \cdots b) \end{aligned}\right\}$$

($v + b + 1$ 개의 연립방정식)

제약조건 :

$\sum_{i=1}^{v}\widehat{\tau_i} = 0$, $\sum_{j=1}^{b}\widehat{\beta_j} = 0$

을 주자.

$$\left.\begin{aligned} vb\hat{\mu} &= y_{..} \\ b\hat{\mu} + b\widehat{\tau_i} &= y_{i\cdot} \\ v\hat{\mu} + v\widehat{\beta_j} &= y_{\cdot j} \end{aligned}\right\}$$

이것의 해 (최소제곱추정량)는

$\hat{\mu} = \overline{y_{..}}$

$\widehat{\tau_i} = \overline{y_{i\cdot}} - \overline{y_{..}} \qquad (i = 1 \cdots v)$

$\widehat{\beta_j} = \overline{y_{\cdot j}} - \overline{y_{..}} \qquad (j = 1 \cdots b)$

이다. 단 $\overline{y_{\cdot\cdot}} = \frac{y_{\cdot\cdot}}{vb}$, $\overline{y_{i\cdot}} = \frac{y_{i\cdot}}{b}$, $\overline{y_{\cdot i}} = \frac{y_{\cdot j}}{v}$.

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\mu} \\ \hat{\tau_1} \\ \vdots \\ \hat{\tau_v} \\ \hat{\beta_1} \\ \vdots \\ \hat{\beta_b} \end{pmatrix} = \begin{pmatrix} \overline{y_{\cdot\cdot}} \\ \overline{y_{1\cdot}} - \overline{y_{\cdot\cdot}} \\ \vdots \\ \overline{y_{v\cdot}} - \overline{y_{\cdot\cdot}} \\ \overline{y_{\cdot 1}} - \overline{y_{\cdot\cdot}} \\ \vdots \\ \overline{y_{\cdot b}} - \overline{y_{\cdot\cdot}} \end{pmatrix}$$

2.1.3 모수함수 $\psi = C^t\boldsymbol{\theta}$ 의 추정가능 조건 :

$(vb \times 1)$ 행렬 a가 존재하여 다음 조건을 만족한다.

$a^t = (a_{11} \cdots a_{1b} a_{21} \cdots a_{2b} \cdots a_{v1} \cdots a_{vb})$라 하자.

$$C^t = a^t X = (\sum_{i=1}^{v}\sum_{j=1}^{b} a_{ij} \;\Big|\; \sum_{j=1}^{b} a_{1j}, \cdots \sum_{j=1}^{b} a_{vj} \;\Big|\; \sum_{i=1}^{v} a_{i1} \cdots \sum_{i=1}^{v} a_i b)$$

$\Leftrightarrow$ μ의 계수 : $\sum_{i=1}^{v}\sum_{j=1}^{b} a_{ij} = \sum_{j=1}^{b}(a_{1j} + \cdots + a_{vj})$: τ의 계수 합

$= \sum_{i=1}^{v}(a_{i1} + \cdots + a_{ib})$: β 계수 합

2.1.4 검정 :

1) $SSE(= SS_\Omega) = \sum_{i=1}^{v}\sum_{j=1}^{b} \epsilon_{ij}^2 = \sum_i\sum_j (y_{ij} - \hat{\mu} - \hat{\tau_i} - \hat{\beta_j})^2$

$= \sum_i\sum_j ((y_{ij} - \overline{y_{\cdot\cdot}}) - (\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}}) - (\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}}))^2$

$= \sum_i\sum_j (y_{ij} - \overline{y_{i\cdot}} - \overline{y_{\cdot j}} + \overline{y_{\cdot\cdot}})^2$

2) C : 처리대비로 구성된 행렬

귀무가설 H_0 : $C\theta = 0$: 검정가능

2-1) H_{01} : $\tau_1 = \tau_2 = \cdots = \tau_v = \tau$ 의 검정

· $C_1 = \left(\begin{array}{c|cccccc|ccc} 0 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{array}\right)_{(v-1)\times(v+b+1)}$

$rank C_1 = v-1 = q$

· 축소 모형 w_1 : $y_{ij} = \mu + \tau + \beta_j + \epsilon_{ij}$

$$= \mu^* + \beta_j + \epsilon_{ij}$$

· $SS_{w_1} = \sum_{j=1}^{b}\sum_{i=1}^{v}(y_{ij} - \overline{y_{\cdot j}})^2$

SS_{w_1} 의 자유도 = $v-1$

· 처리제곱합 :

$SS_{tr} = SS_{w_1} - SS_{\Omega}$

$$= \sum_i\sum_j(y_{ij} - \overline{y_{\cdot j}})^2 - \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}} - \overline{y_{\cdot j}} + \overline{y_{\cdot\cdot}})^2$$

$$= \sum_i\sum_j(y_{ij} - \overline{y_{\cdot j}})^2 - \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}})^2 + \sum_i\sum_j(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2$$

$$= \sum_i(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2\sum_{j=1}^{b}1 = b\sum_i(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2$$

2-2) H_{02} : $\beta_1 = \beta_2 = \cdots = \beta_b = \beta$ 의 검정

· $C_2 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & & & & & & & & & \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}_{(b-1)\times(v+b+1)}$

$rank C_2 = b-1$

· 축소모형 w_2 : $y_{ij} = \mu + \tau_i + \beta + \epsilon_{ij}$

$$= \mu^{**} + \tau_i + \epsilon_{ij} \qquad (4\text{-}14)$$

· $SS_{w2} = \sum_{i=1}^{v}\sum_{j=1}^{b}(y_{ij} - \overline{y_{i\cdot}})^2$

SS_{w2} 의 자유도 = $b-1$

· $SS_{blocks} = SS_{w2} - SS_{\Omega}$

$$= \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}})^2 - \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}} - \overline{y_{\cdot j}} + \overline{y_{\cdot\cdot}})^2$$

$$= \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}})^2 - \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}})^2 + \sum_i\sum_j(\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}})^2$$

$$= \sum_j(\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}})^2\sum_{i=1}^{v}1 = v\sum_j(\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}})^2$$

2-3) 총제곱합 (총변동) SST :

분해 $y_{ij} - \overline{y_{\cdot\cdot}} = (\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}}) + (\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}}) + (y_{ij} - \overline{y_{i\cdot}} - \overline{y_{\cdot j}} + \overline{y_{\cdot\cdot}})$ 로부터

$$SST = \sum_i\sum_j(y_{ij} - \overline{y_{\cdot\cdot}})^2$$

$$= \sum_i\sum_j(\overline{y_{i\cdot}} - \overline{y_{\cdot\cdot}})^2 + \sum_i\sum_j(\overline{y_{\cdot j}} - \overline{y_{\cdot\cdot}})^2 + \sum_i\sum_j(y_{ij} - \overline{y_{i\cdot}} - \overline{y_{\cdot j}} + \overline{y_{\cdot\cdot}})^2$$

$$= SS_{tr} + SS_{blocks} + SSE$$

2.2 교호작용의 의미

$\tau_1 \cdots \tau_a$: 첫째 요인 A의 a개 처리수준의 기대효과

$\beta_1 \cdots \beta_b$: 둘째 요인 B의 b개 처리수준의 기대효과

y_{ijk} : A의 i번째, B의 j번째 수준의 조합을 k번째 관찰치

μ_{ij} : A의 i번째, B의 j번째 수준의 조합에 대한 효과라 하자

선형모형 $y_{ijk} = \mu_{ij} + \epsilon_{ijk,}$ $1 \leqq i \leqq a, 1 \leqq j \leqq b, 1 \leqq k \leqq n$

이 아래와 같이 가법 모형일때

$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$

이 가법 모형이 아닐때 :

교호작용효과$(\tau\beta)_{ij} = \mu_{ij} - \mu - \tau_i - \beta_j \neq 0$을 포함 즉,

$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$

2.3 추정가능함수와 검정가능한 가설의 설정

2.3.1 모형의 행렬표시 :

$$\begin{pmatrix} y_{111} \\ \vdots \\ y_{11n} \\ y_{121} \\ \vdots \\ y_{12n} \\ \vdots \\ y_{1b1} \\ \vdots \\ y_{1bn} \\ y_{211} \\ \vdots \\ y_{21n} \\ \vdots \\ y_{ab1} \\ \vdots \\ y_{abn} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 1 & 0 & & 0 & 1 & 0 & & 0 & 1 & 0 & & 0 & 0 & & 0 \\ 1 & 1 & 0 & & 0 & 0 & 1 & & 0 & 0 & 1 & & 0 & 0 & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 1 & 0 & & 0 & 0 & 1 & & 0 & 0 & 1 & & 0 & 0 & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 1 & 0 & & 0 & 0 & 0 & & 1 & 0 & 0 & & 1 & 0 & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 1 & 0 & & 0 & 0 & 0 & & 1 & 0 & 0 & & 1 & 0 & & 0 \\ 1 & 0 & 1 & & 0 & 1 & 0 & & 0 & 0 & 0 & & 0 & 1 & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 0 & 1 & & 0 & 1 & 0 & & 0 & 0 & 0 & & 0 & 1 & & 0 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 0 & 0 & & 1 & 0 & 0 & & 1 & 0 & 0 & & 0 & 1 & & 1 \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots & \vdots & & & & & & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_a \\ \beta_1 \\ \vdots \\ \beta_b \\ (\tau\beta)_{11} \\ (\tau\beta)_{12} \\ \vdots \\ (\tau\beta)_{1b} \\ (\tau\beta)_{21} \\ \vdots \\ (\tau\beta)_{2b} \\ \vdots \\ (\tau\beta)_{a1} \\ \vdots \\ (\tau\beta)_{ab} \end{pmatrix} + \begin{pmatrix} \epsilon_{111} \\ \vdots \\ \epsilon_{11n} \\ \epsilon_{121} \\ \vdots \\ \epsilon_{12n} \\ \vdots \\ \epsilon_{1b1} \\ \vdots \\ \epsilon_{1bn} \\ \epsilon_{211} \\ \vdots \\ \epsilon_{21n} \\ \vdots \\ \epsilon_{ab1} \\ \vdots \\ \epsilon_{abn} \end{pmatrix}$$

즉 $y = X\theta + \epsilon$

단 $X : (abn) \times (1 + a + b + ab)$행렬

$rank\ X = ab$ ($\because \mu, \tau, \beta$의 계수의 모든 열은 $(\tau\beta)$계수 열을 이용 0으로 변화 가능)

2.3.2 모수함수 $\psi = c^t\theta$의 추정가능 조건 :

1) $q_{\cdot\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} q_{ijk}$, $q_{i\cdot\cdot} = \sum_j \sum_k q_{ijk}$, $q_{\cdot j \cdot} = \sum_i \sum_k q_{ijk}$

$q_{ij\cdot} = \sum_k q_{ijk}$라 하자(계속 이 기호는 사용됨)

$q^t = (q_{111} \cdots q_{11n} q_{121} \cdots q_{12n} \cdots q_{ab1} \cdots q_{abn})$가 존재하여 다음을 만족한다

$c^t = q^t X = (q_{...} q_{1..}, \cdots q_{a..} q_{.1.} \cdots q_{.b.} q_{11.}, \cdots q_{1b.} \cdots q_{ab.})$

$\leftrightarrow \mu$계수 $q_{...} = \sum_{i=1}^{a} q_{i..}$($\tau$계수 합)

$= \sum_{j=1}^{b} q_{.j.}$(β계수 합)

$= \sum_{i=1}^{a} \sum_{j=1}^{b} q_{ij.}$($(\tau\beta)$계수 합)

2) (예제 1) : $\psi = \tau_1 - \tau_2$: 추정불가

$q_{...} = 0 = q_{1..} + q_{2..} + q_{3..} + q_{a..}$

$= 1 + (-1) + 0 + \cdots + 0$

그러나

$q_{11.} = \cdots = q_{1b.} = 0$

$q_{1..} \neq \sum_{j=1}^{b} q_{1j.} = 0$, 한편 $q_{1..} = 1$

이므로 모순이다

따라서 ψ는 추정불가

(예제 2) : $\psi = (\tau\beta)_{11} - (\tau\beta)_{12}$: 추정불가

$q_{...} = 0 = q_{11.} + q_{12.} + q_{13.} + q_{ab.}$

$= 1 + (-1) + 0 + \cdots + 0$

그러나

$q_{.1.} = q_{.2.} = \cdots = q_{.b.} = 0$

$q_{.2.} = q_{12.} + q_{22.} + \cdots + q_{a2.}$

$= (-1) + 0 + \cdots + 0 = -1$

이므로 모순이다

따라서 ψ는 추정불가

2.3.3 추정가능함수 설정, 검정가능한 가설 설정

1) $q_{11\cdot} = 1 - \frac{1}{a} - \frac{1}{b} + \frac{1}{ab} \quad q_{12\cdot} = \cdots = q_{1b\cdot} = \frac{1}{ab} - \frac{1}{b}$

$q_{21\cdot} = q_{31\cdot} = \cdots = q_{a1\cdot} = \frac{1}{ab} - \frac{1}{a}$

$q_{ij\cdot} = \frac{1}{ab}$ (그 밖의 경우)

즉 $(q_{ij\cdot}) = \begin{pmatrix} 1 - \frac{1}{a} - \frac{1}{b} + \frac{1}{ab} & \left(\frac{1}{ab} - \frac{1}{b}\right) I_b^t \\ \left(\frac{1}{ab} - \frac{1}{a}\right) 1_a & \frac{1}{ab} J_{a \times b} \end{pmatrix}$인 경우 :

① $q_{1\cdot\cdot} = \sum_j \sum_k q_{1jk} = q_{11\cdot} + q_{12\cdot} + \cdots + q_{1b\cdot}$

$= \left(1 - \frac{1}{a} - \frac{1}{b} + \frac{1}{ab}\right) + \left(\frac{1}{ab} - \frac{1}{b}\right)(b-1) = 0$

$q_{a\cdot\cdot} = \sum_j \sum_k q_{ajk} = q_{a1\cdot} + q_{a2\cdot} + \cdots + q_{ab\cdot}$

$= \left(\frac{1}{ab} - \frac{1}{a}\right) + \frac{1}{ab} \times (b-1) = 0$

$q_{\cdot 1\cdot} = \sum_i \sum_k q_{i1k} = \sum_i q_{i1\cdot} = q_{11\cdot} + q_{21\cdot} + \cdots + q_{a1\cdot}$

$= \left(1 - \frac{1}{a} - \frac{1}{b} + \frac{1}{ab}\right) + \left(\frac{1}{ab} - \frac{1}{a}\right)(a-1) = 0$

$q_{\cdot b\cdot} = \sum_i \sum_k q_{ibk} = \sum_i q_{ib\cdot} = q_{1b\cdot} + q_{2b\cdot} + \cdots + q_{ab\cdot}$

$= \left(\frac{1}{ab} - \frac{1}{b}\right) + \frac{1}{ab}(a-1) = 0$

$q_{\cdot\cdot\cdot} = \sum_i q_{i\cdot\cdot} = \sum_{i=1}^{a} 0 = 0$

즉, $c^t = (q_{\cdot\cdot\cdot}\ q_{1\cdot\cdot} \cdots q_{a\cdot\cdot}\ q_{\cdot 1\cdot} \cdots q_{\cdot b\cdot}\ q_{11\cdot} \cdots q_{ab\cdot})$

$= (0 \;\; 0 \; \cdots \; 0 \;\; 0 \;\; \cdots \;\; 0 \; q_{11}. \; \cdots \; q_{ab}.)$

② $= \sum_i \sum_j q_{ij}.(\tau\beta)_{ij} = \sum_j q_{1j}.(\tau\beta)_{1j} + \sum_j q_{2j}. + \cdots + \sum_j q_{aj}.(\tau\beta)_{aj}$

$= (q_{11}.(\tau\beta)_{11} + q_{12}.(\tau\beta)_{12} + \cdots + q_{1b}.(\tau\beta)_{1b})$

$+ (q_{21}.(\tau\beta)_{21} + q_{22}.(\tau\beta)_{22} + \cdots + q_{2b}.(\tau\beta)_{2b}) + \cdots$

$+ (q_{a1}.(\tau\beta)_{ab} + q_{a2}.(\tau\beta)_{a2} + \cdots + q_{ab}.(\tau\beta)_{ab})$

$= \left(1 - \frac{1}{a} - \frac{1}{b} + \frac{1}{ab}\right)(\tau\beta)_{11} + \left(\frac{1}{ab} - \frac{1}{b}\right)((\tau\beta)_{12} + \cdots + (\tau\beta)_{1b})$

$+ \left(\frac{1}{ab} - \frac{1}{a}\right)(\tau\beta)_{21} + \frac{1}{ab}((\tau\beta)_{22} + \cdots + (\tau\beta)_{2b}) + \cdots$

$+ \left(\frac{1}{ab} - \frac{1}{a}\right)(\tau\beta)_{a1} + \frac{1}{ab}((\tau\beta)_{a2} + \cdots + (\tau\beta)_{ab})$

$= (\tau\beta)_{11} - \frac{1}{a}\sum_i (\tau\beta)_{i1} - \frac{1}{b}\sum_j (\tau\beta)_{1j} + \frac{1}{ab}\sum_i \sum_j (\tau\beta)_{ij}$

$= (\tau\beta)_{11} - \frac{1}{a}(\tau\beta._1 - \frac{1}{b}(\tau\beta)_{1.} + \frac{1}{ab}(\tau\beta)_{..}$

$= (\tau\beta)_{11} - (\overline{\tau\beta})._1 - (\overline{\tau\beta})_{1.} + (\overline{\tau\beta})_{..}$ 추정가능

$\because$ 앞 ①에서 c^t가 추정가능 조건 만족

즉 μ계수 = τ계수 합

= β계수 합

= $(\tau\beta)$계수 합

위의 q에서 $(q_{11.} \cdots q_{qb.})$을 $(q_{i1.} \cdots q_{ib.})$로 $(q_{21.} \cdots q_{q1})$을 $(q_{2j} \cdots q_{qj})$로

또한 $(\overline{\tau\beta})_{i.} = \frac{\sum_j (\tau\beta)_{ij}}{b}$, $(\overline{\tau\beta})._j = \frac{\sum_i (\tau\beta)_{ij}}{a}$,

$(\overline{\tau\beta})_{..} = \frac{\sum_i \sum_j (\tau\beta)_{ij}}{ab}$ 로 일반화하면

교호작용 효과 함수 r_{11}는 다음과 같이 r_{ij}로 일반화된다

$r_{ij} = (\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot}$: 추정 가능

③ 귀무 가설 $H_{01} : (\tau\beta)_{11} = \cdots = (\tau\beta)_{ab}$: 검정불가 (∵ 예제 2)

$H_{01} : r_{11} = \cdots = r_{ab}$: 검정가능 (∵ 앞 ②)

2) $q_{11\cdot} = \cdots = q_{1b\cdot} = \dfrac{1}{b}$

$q_{21\cdot} = \cdots = q_{2b\cdot} = -\dfrac{1}{b}$

$q_{ij} = 0$ (그 밖의 경우)

즉 $(q_{ij\cdot}) = \begin{pmatrix} \dfrac{1}{b}1_b^{\,t} \\ \hline -\dfrac{1}{b}1_b^{\,t} \\ \hline 0 \end{pmatrix}_{a\times b}$ 인 경우 :

① $q_{1\cdot\cdot} = \sum_j \sum_k q_{1jk} = q_{11\cdot} + \cdots + q_{1b\cdot} = \dfrac{1}{b} \times b = 1$

$q_{2\cdot\cdot} = \sum_j \sum_k q_{2jk} = q_{21\cdot} + \cdots + q_{2b\cdot} = \left(-\dfrac{1}{b}\right) \times b = -1$

$q_{i\cdot\cdot} = \sum_j \sum_k q_{ijk} = q_{i1\cdot} + \cdots + q_{ib\cdot} = 0 \times b = 0 \ (i \neq 1, 2)$

$q_{\cdot j\cdot} = \sum_j \sum_k q_{ijk} = q_{1j\cdot} + q_{2j\cdot} + q_{3j\cdot} + \cdots + q_{aj\cdot}$

$= \dfrac{1}{b} + \left(-\dfrac{1}{b}\right) + 0 + \cdots + 0 = 0 \ (j = 1 \cdots b)$

μ의 계수 $q_{\cdots} = \sum_{i=1}^{a} q_{i\cdot\cdot} = 1 + (-1) + 0 + \cdots + 0 = 0$ (τ 계수 합)

$= \sum_{j=1}^{b} q_{\cdot j\cdot} = 0 + \cdots + 0 = 0$ (β 계수 합)

$$= \sum_i \sum_j q_{ij}. \ ((\tau\beta)\text{계수 합})$$

이므로 추정 가능 조건 만족

② $i=1,2$에 대한 추정가능한 모수 함수는

$$\psi = \sum_i \sum_j q_{ij.}\tau_i + \sum_i \sum_j q_{ij.}(\tau\beta)_{ij}$$

$$= \left(\sum_j q_{1j.}\tau_1 + \sum_j q_{2j.}\tau_2 + \sum_j q_{3j.}\tau_3 + \cdots + \sum_j q_{aj.}\tau_a\right)$$

$$+ \left(\sum_j q_{1j.}(\tau\beta)_{1j} + \sum_j q_{2j.}(\tau\beta)_{2j} + \sum_j q_{3j.}(\tau\beta)_{3j} + \cdots + \sum_j q_{aj.}(\tau\beta)_{aj}\right)$$

$$= (q_{1..}\tau_1 + q_{2..}\tau_2 + q_{3..}\tau_3 + \cdots + q_{a..}\tau_a)$$

$$+ \left(q_{1..}\sum_j (\tau\beta)_{1j} + q_{2..}\sum_j (\tau\beta)_{2j} + q_{3..}\sum_j (\tau\beta)_{3j} + \cdots + q_{a..}\sum_j (\tau\beta)_{aj}\right)$$

$$= (\tau_1 - \tau_2 + 0 + \cdots + 0) + \left(\frac{\sum_j (\tau\beta)_{1j}}{b} + \frac{\sum_j (\tau\beta)_{2j}}{-b} + 0 + \cdots + 0\right)$$

$$= (\tau_1 - \tau_2) + \left(\frac{(\tau\beta)_{1.}}{b} - \frac{(\tau\beta)_{2.}}{b}\right)$$

$= (\tau_1 - \tau_2) + (\overline{\tau\beta})_{1.} - (\overline{\tau\beta})_{2.}$: 추정가능

$i=1,2$를 $i=l,k$로 일반화하면

추정가능 모수 함수는

$\psi = \tau_l - \tau_k + (\overline{\tau\beta})_{l.} - (\overline{\tau\beta})_{k.}$이다

③ 귀무가설 $H_{01} : \tau_1 = \tau_2 = \cdots = \tau_a$: 검정불가 ($\because$ 예제 1)

$$H_{01} : \left(\tau_1 + (\overline{\tau\beta})_{l.}\right) = \cdots = \left(\tau_a + (\overline{\tau\beta})_{a.}\right)$$

: 검정가능 ($\because$ 앞 ②에서 ψ은 추정가능)

<참고> ②에서 $i=l,k$대신 $j=m,n$으로 일반화하면 추정가능함수는

$\psi = \beta_m - \beta_n + (\overline{\tau\beta})_{.m} - (\overline{\tau\beta})_{.n}$이다.

2.4 분산분석

2.4.1 $y = X\theta + \epsilon$에서 정규방정식 :

(1) X^tX

abn	$bn1_a^{\ t}$	$an1_b^{\ t}$	$n1_{ab}^{\ t}$
$bn1_a$	bnI_a	$nJ_{a\times b}$	$\begin{pmatrix} n & \cdots & n & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & n & \cdots & n & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & n & \cdots & n \end{pmatrix}_{a\times ab}$
$an1$	$nJ_{b\times a}$	anI_b	$\begin{pmatrix} n & \cdots & n & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & n & \cdots & n & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & n \end{pmatrix}_{b\times ab}$
$n1_{ab}$	$\begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & n & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{ab\times a}$	$\begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & n & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{ab\times b}$	nI_{ab}

여기서 $rank\ X = ab$이므로

$rank\ X^tX = ab$이고 X^tX 정방행렬이나 비정칙이다

(2) 정규방정식 $(X^tX)\hat{\theta} = X^ty$는 $(1+a+b+ab)$개의 방정식으로 다음과 같이 표시된다.

$$\left.\begin{aligned} abn\hat{\mu} + bn\sum_i \widehat{\tau_i} + an\sum_j \widehat{\beta_j} + n\sum_i \sum_j (\widehat{\tau\beta})_{ij} &= y_{\cdots}\left(= \sum_i \sum_j \sum_k y_{ijk}\right) \\ bn\hat{\mu} + bn\widehat{\tau_i} + n\sum_j \widehat{\beta_j} + n\sum_j (\widehat{\tau\beta})_{ij} &= y_{i\cdot\cdot}\left(= \sum_j \sum_k y_{ijk}\right) i = 1 \cdots a \\ an\hat{\mu} + n\sum_i \widehat{\tau_i} + an\widehat{\beta_j} + n\sum_i (\widehat{\tau\beta})_{ij} &= y_{\cdot j\cdot}\left(= \sum_i \sum_k y_{ijk}\right) j = 1 \cdots b \\ n\hat{\mu} + n\widehat{\tau_i} + n\widehat{\beta_j} + n(\widehat{\tau\beta})_{ij} &= y_{ij\cdot}\left(= \sum_k y_{ijk}\right) i = 1 \cdots a, j = 1 \cdots b \end{aligned}\right\}$$

2.4.2 정규방정식의 해 :

$rank\ \ (X^tX) = rank\ \ X = ab$이고 정규방정식은 $(1+a+b+ab)$개로 구성되었으나 정규방정식은 ab개로 구성된 부정방정식과 같으므로 다음과 같이 $(1+a+b+ab)-ab = 1+a+b$개의 제약식을 주면 $(X^tX)\hat{\theta} = X^ty$는 유일한 해를 갖는다

$$\left.\begin{array}{c} \sum_i \hat{\tau}_i = 0 \\ \sum_j \hat{\beta}_j = 0 \\ \sum_i (\widehat{\tau\beta})_{ij} = 0 \ \ j = 1 \cdots b \\ \sum_j (\widehat{\tau\beta})_{ij} = 0 \ \ i = 1 \cdots a \end{array}\right\}$$

(4-37) 조건하에 (4-36)해는 다음과 같다

$$\left.\begin{array}{c} \hat{\mu} = \bar{y}_{...} \\ \hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \\ (\widehat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \end{array}\right\}$$

단 $\bar{y}_{...} = \dfrac{y_{...}}{abn} = \dfrac{\sum_i \sum_j \sum_k y_{ijk}}{abn}$

$$\bar{y}_{i..} = \frac{y_{i..}}{bn} = \frac{\sum_j \sum_k y_{ijk}}{bn}$$

$$\bar{y}_{.j.} = \frac{y_{.j.}}{an} = \frac{\sum_i \sum_k y_{ijk}}{an}$$

$$\bar{y}_{ij.} = \frac{y_{ij.}}{n} = \frac{\sum_k y_{ijk}}{n}$$

2.4.3 완전 모형의 제곱합

$$\Omega : y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \text{ : 완전모형}$$

$$r_{ij} = (\tau\beta)_{ij} - (\overline{\tau\beta})_{i\cdot} - (\overline{\tau\beta})_{\cdot j} + (\overline{\tau\beta})_{\cdot\cdot}$$

$$\hat{\mu} + \hat{\tau} + \hat{\beta} + (\widehat{\tau\beta})_{ij} = (\bar{y}_{\cdots}) + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})$$

$$+ \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) = \bar{y}_{ij\cdot}$$

(1) $SSE = SS_{\Omega}$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - (\widehat{\tau\beta})_{ij}\right)^2$$

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

$$\Omega : y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

$$= \mu + \tau_i + \beta_j + \left(\gamma_{ij} + (\overline{\tau\beta})_{i\cdot} + (\overline{\tau\beta})_{\cdot j} - (\overline{\tau\beta})_{\cdot\cdot}\right) + \epsilon_{ijk}$$

$$= \left(\mu - (\overline{\tau\beta})_{\cdot\cdot}\right) + \gamma_{ij} + \left(\tau_i + (\overline{\tau\beta})_{i\cdot}\right) + \left(\beta_j + (\overline{\tau\beta})_{\cdot j}\right) + \epsilon_{ijk}$$

(2) 귀무가설 $H_{01} : \gamma_{11} = \cdots = \gamma_{ab} = \gamma$조건 하에서

완전모형의 축소모형 :

$$w_1 : y_{ijk} = \left(\mu - (\overline{\tau\beta})_{\cdot\cdot} + \gamma\right) + \left(\tau_i + (\overline{\tau\beta})_{i\cdot}\right) + \left(\beta_j + (\overline{\tau\beta})_{\cdot j}\right) + \epsilon_{ijk}$$

$$= \mu^* + \tau_i^* + \beta_i^* + \epsilon_{ijk}$$

이 모형은 교호작용이 없는 이원 배치형이고

$$SS_{w1} = \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right)^2$$

<참고> SS_w의 자유도는 ab이다 ($\because$ $rank\ (X^t X) = ab$)

그러나 SS_{w1}은 교호작용이 없어졌으므로 (4-37)에 의해 자유도가 $(a-1)+b$개 줄어지므로 SS_{w1}의 자유도는 $q = ab - (a-1) - b = (a-1)(b-1)$이다

(3) 귀무가설 $H_{02}:\left(\tau_1+(\overline{\tau\beta})_{1\cdot}\right)=\cdots=\left(\tau_a+(\overline{\tau\beta})_{a\cdot}\right)=\tau$조건하에서

완전모형 (4-40)의 축소모형은

$$w_2: y_{ijk}=(\mu-(\overline{\tau\beta})_{\cdot\cdot})+(\tau_i+(\overline{\tau\beta})_{i\cdot})+(\beta_j+(\overline{\tau\beta})_{\cdot j})+\gamma_{ij}+\epsilon_{ijk}$$

$$=(\mu-(\overline{\tau\beta})_{\cdot\cdot}+\tau)+(\beta_j+(\overline{\tau\beta})_{\cdot j})+\gamma_{ij}+\epsilon_{ijk}$$

$$=\mu^*+\beta_i{}^*+\gamma_{ij}+\epsilon_{ijk} \qquad \cdots$$

τ가 없어지고 γ가 추가된 모형으로 정규방정식을 구한 다음 추가 제약식

$$\sum_j \hat{\beta}_j{}^*=0 \qquad \sum_i (\widehat{\beta^*\gamma})_{ij}=0 \qquad (j=1\cdots b)$$

$$\sum_j (\widehat{\beta^*\gamma})_{ij}=0 \qquad (i=1\cdots a-1)$$

을 대입하여 해를 구하면

$$\left.\begin{array}{l}\widehat{\mu^*}=\overline{y}_{\cdots} \quad \widehat{\beta^*}=\overline{y}_{\cdot i\cdot}-\overline{y}_{\cdots} \\ (\widehat{\beta^*\gamma})_{ij}=\overline{y}_{ij\cdot}-\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdot j\cdot}+\overline{y}_{\cdots}\end{array}\right\}$$

이다. 따라서

$$SS_{w2}=\sum_i\sum_j\sum_k\left(y_{ijk}-\widehat{\mu^*}-\widehat{\beta^*}_j-(\widehat{\beta^*\gamma})_{ij}\right)^2$$

$$=\sum_i\sum_j\sum_k\left(y_{ijk}-\overline{y}_{ij\cdot}+\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdots}\right)^2$$

(4) 귀무가설 $H_{03}:\left(B_1+(\overline{\tau\beta})_{\cdot 1}\right)=\cdots=\left(B_b+(\overline{\tau\beta})_{\cdot b}\right)=B$조건하에서

완전모형의 축소모형은

$$w_3: y_{ijk}=(\mu-(\overline{\tau\beta})_{\cdot\cdot})+(\tau_i+(\overline{\tau\beta})_{i\cdot})+B+\gamma_{ij}+\epsilon_{ijk}$$

$$=(\mu-(\overline{\tau\beta})_{\cdot\cdot}+B)+(\tau_i+(\overline{\tau\beta})_{i\cdot})+\beta+\gamma_{ij}+\epsilon_{ijk}$$

$$=\mu^*+\tau_i{}^*+\gamma_{ij}+\epsilon_{ijk}$$

w_3에 따라 정규방정식을 만들고 추가 제약식

$$\left.\begin{array}{c} \sum_i \tau_i{}^* = 0 \\ \sum_i (\tau^* \gamma)_{ij} = 0 \ \ (j = 1 \cdots b) \\ \sum_j (\tau^* \gamma)_{ij} = 0 \ \ (i = 1 \cdots a-1) \end{array}\right\}$$

을 정규방정식에 대입하여 풀면 해는

$$\widehat{\mu^*} = \bar{y}_{...} \quad \widehat{\tau^*} = \bar{y}_{i..} - \bar{y}_{...}$$

$$(\widehat{\tau^* \gamma})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

이다. 따라서

$$SS_{w3} = \sum_i \sum_j \sum_k (y_{ijk} - \widehat{\mu^*} - \widehat{\tau^*}_i - (\widehat{\tau^* \gamma})_{ij})^2$$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.} + \bar{y}_{.j.} - \bar{y}_{...}\right)^2$$

(5) $SS_{AB} = SS_{w1} - SS_{\Omega}$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2 - \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

$$= \sum_i \sum_j \sum_k \left(\left(y_{ijk} - \bar{y}_{ij.}\right) + \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)\right)^2$$

$$- \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2 + \sum \sum \sum \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2$$

$$- \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

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

$$= n\sum_i \sum_j \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right)^2$$

(6) $SS_A = SS_{w2} - SS_\Omega$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij\cdot} + \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdots}\right)^2 - \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

$$= \sum_i \sum_j \sum_k \left(\bar{y}_{i\cdot\cdot} + \bar{y}_{\cdots}\right)^2 = \sum_i \left(\bar{y}_{i\cdot\cdot} + \bar{y}_{\cdots}\right)^2 \sum_j \sum_k 1$$

$$= bn\sum_i \left(\bar{y}_{i\cdot\cdot} + \bar{y}_{\cdots}\right)^2$$

(7) $SS_B = SS_{w3} - SS_\Omega$

$$= \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij\cdot} + \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right)^2 - \sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

$$= an\sum_j \left(\bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right)^2$$

(8) 편차의 분해

$$y_{ijk} - \bar{y}_{\cdots} = \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}\right) + \left(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}\right)$$
$$+ \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right) + \left(y_{ijk} - \bar{y}_{ij\cdot}\right)$$

로부터 다음을 얻는다

$$\sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{\cdots}\right)^2 = \sum \sum \sum \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}\right)^2 + \sum \sum \sum \left(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}\right)^2$$

$$+ \sum \sum \sum \left(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}\right)^2 + \sum \sum \sum \left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2$$

$\therefore SST = SS_A + SS_B + SS_{AB} + SS\Omega$

2.5 불완비 블록계획

2.5.1 불완비 블록계획의 모형

(1) 기호 :

μ : 전체 평균

$\tau_i (i = 1, 2, \cdots v)$: 처리 i의 효과

$\beta_j (j = 1, 2, \cdots b)$: 블록 j의 효과

k_j :block j에서의 처리의 수 즉 블록 size

r_i :처리 i의 반복수

n_{ij} : 블록j에서 이루어진 처리 i의 실험 횟수

n : 총 관찰치 즉 $n = \sum_i^v \sum_j^b n_{ij} = \sum_i^v r_i = \sum_j^b k_j$

$N = (n_{ij})$: $v \times b$ 빈도 행렬(incidenced matrix)

$N^t N = v \times v$ 일치행렬 (concurrence matrix)

또한 $r_i = \sum_{j=1}^b n_{ij}$

$k_j = \sum_{i=1}^v n_{ij}$

$n = \sum_{i=1}^v \sum_{j=1}^b n_{ij} = \sum_{i=1}^v r_i = \sum_{j=1}^b k_j$ (전체 실험 횟수)

(2) 모형

$$\left.\begin{array}{l} y_{ij\ell} = \mu + \tau_i + \beta_j + \epsilon_{ij\ell} \\ i = 1 \cdots v \ , \ j = 1 \cdots b \ , \ \ell = 1 \cdots n_{ij} \end{array}\right\}$$

$$
\left.
\begin{array}{l}
y_{111} = \mu + \tau_1 + \beta_1 + \epsilon_{111} \\
y_{112} = \mu + \tau_1 + \beta_1 + \epsilon_{112} \\
\cdots\cdots\cdots\cdots \\
y_{11n_{11}} = \mu + \tau_1 + \beta_1 + \epsilon_{11n_{11}} \\
y_{121} = \mu + \tau_1 + \beta_2 + \epsilon_{121} \\
\cdots\cdots\cdots\cdots \\
y_{12n_{12}} = \mu + \tau_1 + \beta_2 + \epsilon_{12n_{12}} \\
\cdots\cdots\cdots\cdots \\
y_{1b1} = \mu + \tau_1 + \beta_b + \epsilon_{1b1} \\
\cdots\cdots\cdots\cdots \\
y_{1bn_{1b}} = \mu + \tau_1 + \beta_b + \epsilon_{1bn_{1b}} \\
y_{211} = \mu + \tau_2 + \beta_1 + \epsilon_{211} \\
\cdots\cdots\cdots\cdots \\
y_{vb1} = \mu + \tau_v + \beta_1 + \epsilon_{vb1} \\
\cdots\cdots\cdots\cdots \\
y_{vbn_{vb}} = \mu + \tau_n + \beta_b + \epsilon_{vbn_{vb}}
\end{array}
\right\}
$$

$$
y = \begin{pmatrix} y_{111} \\ \vdots \\ y_{11n_{11}} \\ \vdots \\ y_{1b1} \\ \vdots \\ y_{1bn_{1b}} \\ y_{211} \\ \vdots \\ y_{21n_{21}} \\ \vdots \\ \vdots \\ y_{vb1} \\ \vdots \\ y_{vbn_{vb}} \end{pmatrix}
=
\begin{pmatrix}
1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 1 & 0 & & 0 & 1 & 0 & & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 1 & 0 & & 0 & 0 & 0 & & 1 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 1 & 0 & & 0 & 0 & 0 & & 1 \\
1 & 0 & 1 & & 0 & 1 & 0 & & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 0 & 1 & & 0 & 1 & 0 & & 0 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 0 & 0 & & 1 & 0 & 0 & & 1 \\
\vdots & \vdots & & & \vdots & \vdots & & & \vdots \\
1 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix} \mu \\ \tau_1 \\ \vdots \\ \tau_v \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_b \end{pmatrix}
+
\begin{pmatrix} \epsilon_{111} \\ \vdots \\ \epsilon_{11n_{11}} \\ \vdots \\ \epsilon_{1b1} \\ \vdots \\ \epsilon_{1bn_{1b}} \\ \epsilon_{211} \\ \vdots \\ \epsilon_{21n_{21}} \\ \vdots \\ \vdots \\ \epsilon_{vb1} \\ \vdots \\ \epsilon_{vbn_{vb}} \end{pmatrix}
$$

$$
= X\boldsymbol{\theta} + \boldsymbol{\epsilon}
$$

여기서 $X = (1_n \mid X_\tau \mid X_\beta)$: $n \times (1 + v + b)$ 行列 (계획行列)

$\theta = \begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix}$: 모수行列 (모수벡터)

$$X = (1_n \mid X_\tau \mid X_\beta)\begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix} + \epsilon = 1_n\mu + X_\tau\tau + X_\beta\beta + \epsilon$$

로도 표시된다.

(3) 관련기호

1) $r = \begin{pmatrix} r_1 \\ \vdots \\ r_v \end{pmatrix} = N\underline{1}_b = X_\tau^t\underline{1}_n \ [r^t = (X_\tau^t\underline{1}_n)^t = \underline{1}_n^t X_\tau]$

$k = \begin{pmatrix} k_1 \\ \vdots \\ k_b \end{pmatrix} = N^t\underline{1}_v = X_\beta^t\underline{1}_n \ [k^t = \underline{1}_n^t X_\beta]$

2) $K = \begin{pmatrix} k_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & k_b \end{pmatrix}_{b\times b}$, $R = \begin{pmatrix} r_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & r_v \end{pmatrix}_{v\times v}$

$\tau = \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_v \end{pmatrix}_{v\times 1}$, $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_b \end{pmatrix}_{b\times 1}$

$X_\tau^t X_\tau = R$

$X_\beta^t X_\beta = k$

$X_\tau^t X_\beta = N$

$\boldsymbol{T} = X_\tau^t\underline{y}$: 처리합 벡터

$\boldsymbol{\beta} = X_\beta^t\underline{y}$: 블록합 벡터

$G = \underline{1}_n^t\underline{y}$: 반응치 합

3) 선형모형

$$\left.\begin{array}{l}
y_{111} = 9 = \mu + \tau_1 + \beta_1 + \epsilon_{111} \\
y_{121} = 16 = \mu + \tau_1 + \beta_2 + \epsilon_{121} \\
y_{211} = 10 = \mu + \tau_2 + \beta_1 + \epsilon_{211} \\
y_{231} = 8 = \mu + \tau_2 + \beta_3 + \epsilon_{231} \\
y_{321} = 14 = \mu + \tau_3 + \beta_2 + \epsilon_{321} \\
y_{331} = 12 = \mu + \tau_3 + \beta_3 + \epsilon_{331} \\
y_{411} = 8 = \mu + \tau_4 + \beta_1 + \epsilon_{411}
\end{array}\right\}$$

이 모형을 行列式 으로 표시하면 다음과 같다.

$$= \mu + \tau_1 + \tau_2 + \tau_3 + \tau_4 + \beta_1 + \beta_2 + \beta_3 + \epsilon_{111}$$

$$= \mu + \tau_1 + 。.\tau_2 + 。.\tau_3 + 。.\tau_4 + \beta_1 + 。.\beta_2 + 。.\beta_3 + \epsilon_{121}$$

$$= \mu + 。.\tau_1 + \tau_2 + 。.\tau_3 + 。.\tau_4 + 。.\beta_1 + \beta_2 + 。.\beta_3 + \epsilon_{211}$$

. . . . . .

$$y = \begin{pmatrix} y_{111} \\ y_{121} \\ y_{211} \\ y_{231} \\ y_{321} \\ y_{331} \\ y_{411} \end{pmatrix} = \begin{pmatrix} 9 \\ 16 \\ 10 \\ 8 \\ 14 \\ 12 \\ 8 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \epsilon$$

$$y = X\theta + \epsilon$$

$$= (1_n \mid X_\tau \mid X_\beta)\boldsymbol{\theta} + \epsilon$$

$$X_\tau = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; X_\tau = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \; \boldsymbol{\theta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix}$$

2.5.2 모수의 추정과 연결된 계획

모형 $y_{ij\ell} = \mu + \tau_i + \beta_j + \epsilon_{i\ell}$, $i = 1, \cdots, v$, $j = 1, \cdots, b$, $\ell = 1, \cdots, n_{ij}$ 즉

$y = X\theta + \epsilon$, $X = (1_n | X_\tau | X_\beta)$에서

(1) $X^t X = (1_n | X_\tau | X_\beta)^t (1_n | X_\tau | X_\beta)$

$$= \begin{pmatrix} 1_n^t \\ X_\tau^t \\ X_\beta^t \end{pmatrix} (1_n \ X_\tau \ X_\beta) = \begin{pmatrix} 1_n^t 1_n & 1_n^t X_\tau & 1_n^t X_\beta \\ X_\tau^t 1_n & X_\tau^t X_\tau & X_\tau^t X_\beta \\ X_\beta^t 1_n & X_\beta^t X_\tau & X_\beta^t X_\beta \end{pmatrix}$$

$$= \begin{pmatrix} n & r^t & k^t \\ r & R & N \\ k & N^t & K \end{pmatrix}$$

(2) 정규방정식 $(X^t X)\hat{\theta} = X^t y$ 계산

$$\begin{pmatrix} n & r^t & k^t \\ r & R & N \\ k & N^t & K \end{pmatrix} \begin{pmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} 1_n^t \\ X_\tau^t \\ X_\beta^t \end{pmatrix} y$$

$$\begin{pmatrix} n\hat{\mu} + r^t\hat{\tau} + k^t\hat{\beta} \\ r\hat{\mu} + R\hat{\tau} + N\hat{\beta} \\ k\hat{\mu} + N^t\hat{\tau} + K\hat{\beta} \end{pmatrix} = \begin{pmatrix} 1_n^t y \\ X_\tau^t y \\ X_\beta^t y \end{pmatrix}$$

$$\begin{aligned} n\hat{\mu} + r^t\hat{\tau} + k^t\hat{\beta} &= 1_n^t y = G \\ r\hat{\mu} + R\hat{\tau} + N\hat{\beta} &= X_\tau^t y = T \\ k\hat{\mu} + N^t\hat{\tau} + K\hat{\beta} &= X_\beta^t y = B \end{aligned}$$

$$(r - NK^{-1}k)\hat{\mu} + (R - NK^{-1}N^t)\hat{\mu} + (N - NK^{-1}K)\hat{\beta} = T - NK^{-1}B$$

여기서

$$\left.\begin{aligned} r - NK^{-1}k &= r - N1_b = r - r = 0 \\ N - NK^{-1}K &= N - N = 0 \end{aligned}\right\rangle \text{이므로}$$

$R - NK^{-1}N^t = C$, $T - NK^{-1}B = q$ 라 두면

$C\hat{\tau} = q$

이다. 이식은 블록효과 ($\hat{\beta}$항)가 제거된 축소정규방정식(reduced normal equation)

이고

$C: v \times v$ 정보행렬

q : 조정된 처리 합)

C^{-} : C 의 $G-inverse$ 라면 해는

$\hat{\tau} = C^{-}q$

$\hat{\beta} = K^{-1}B - K^{-1}N^{t}\hat{\tau} - K^{-1}K\hat{\mu}$

$= K^{-1}B - K^{-1}N^{t}(C^{-}q) - 1_{b}\hat{\mu}$

(정리 2-1) < 정보행렬 $C = R - NK^{-1}N^{t}$에 관한 정리 >

↳ $R = \begin{pmatrix} r_1 & & 0 \\ & \ddots & \\ 0 & & r_v \end{pmatrix}, k = \begin{pmatrix} k_1 & & 0 \\ & \ddots & \\ 0 & & k_b \end{pmatrix}, N = (n_{ij})$

(a) C: 대칭행렬

(b) $i)$ C의 행의 합 = 0

$ii)$ C의 열의 합 = 0

$iii)$ $rank\, C \leqq v - 1$

모수함수 $\psi = a^{t}\tau$: 추정가능(estimate) $a^{t} = a^{t}C^{-}C$ 즉

(c) $a^{t} = a^{t}C^{-}C$ 이면 모수함수 $\psi = a^{t}\tau$는 추정가능

(d) 모수함수 $\psi = a^{t}\tau$가 추정 가능한 경우 다음식이 성립한다.

$i)$ $Var(a^{t}\hat{\tau}) = a^{t}C^{-}a\sigma^{2}$

$ii)$ $Var(a^{t}\hat{\tau})$은 C^{-}의 선택에 무관 (독립)

(증명)

(a) $C^{t} = (R - NK^{-1}N^{t} = R^{t} - NK^{-1}N^{t}$ $(R^{t} = R)$

$= R - NK^{-1}N^t = C$

$\therefore$ C : 대칭行列

(b) $i)$ $C: v \times v$ 行列이므로

행의 合 : $C1v = (R - NK^{-1}N^t)1v$

$= R1v - NK^{-1}N^t1v$

$= r - NK^{-1}k = r - N1_b$

$= r - r = 0$ ($\because$ $N1_b = r$)

$ii)$ 列의 合 : $C^t1v = (R - NK^{-1}N^t)^t1v$

$= (R - NK^{-1}N^t)1_v$

$= r - r = 0$ ($\because$ 앞 $i)$)

$(= R1_v - NK^{-1}\underbrace{N^t1_v}_{k} = r - N\underbrace{K^{-1}k}_{1_b} = r - r)$

$iii)$ C에서 1행에 나머지 행을 모두 代入하면 1행은

앞 $i)$에 의해 모두 0이 되고 이 행렬을 C_1이라 하자

$rank\ C = rank\ C_1 \le v - 1$ ($\because$ C: $v \times v$ 行列)

모수함수 $\Psi = a^t\tau$: estimate $E(a^t\hat{\tau}) = a^t\tau$

(c) Markov 정리에 의해 $E(a^t\hat{\tau}) = a^t\tau$이면 $\Psi = a^t\tau$ 는

$E(a^t\hat{\tau}) = a^tE(\hat{\tau}) = a^tE(C^-q)$

$= a^tC^-E(T - NK^{-1}B)$ ($\because$ $q = T - NK^{-1}B$)

$= a^tC^-E(X_\tau^ty - NK^{-1}X_\beta^ty)$ ($\because$ $T = X_\tau^ty$, $B = X_\beta^ty$)

$= a^tC^-(X_\tau^ty - NK^{-1}X_\beta^ty)E(y)$

$= a^tC^-(X_\tau^t - NK^{-1}X_\beta^t)(X\theta)$

$$= a^t C^- (X_\tau^t - NK^{-1}X_\beta^t)(1n \mid X_\tau \mid X_\beta)\begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix}$$

$$= a^t C^- (X_\tau^t 1n - NK^{-1}X_\beta^t 1n \mid X_\tau^t X_\tau - NK^{-1}X_\beta^t X_\tau \mid X_\tau^t X_\beta - NK^{-1}X_\beta^t X_\beta)\begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix}$$

여기서

$$X_\tau^t 1n - NK^{-1}X_\beta^t 1n = r - NK^{-1}k = r - N1_b = r - r = 0$$

$$X_\tau^t X_\tau - NK^{-1}X_\beta^t X_\tau = R - NK^{-1}(X_\beta^t X_\tau)^t$$

$$= R - NK^{-1}N^t = C$$

$$X_\tau^t X_\beta - NK^{-1}X_\beta^t X_\beta = N - NK^{-1}K = N - N = 0$$

을 이용하면

$$E(a^t\hat{\tau}) = a^t C^{-1}(0 \mid C \mid 0)\begin{pmatrix} \mu \\ \tau \\ \beta \end{pmatrix}$$

$$= a^t C^{-1} - C\tau = a^t\tau \qquad (\because a^t = a^t C^{-1}C \text{ 인 가정})$$

따라서 $\psi = a^t\tau$ 는 추정가능이다.

(d) $i)$ • $Cov(q) = Cov(T - NK^{-1}B) = Cov(X_\tau^t y - NK^{-1}X_\beta^t y)$

$$= Cov((X_\tau^t - NK^{-1}X_\beta^t)y)$$

$$= (X_\tau^t - NK^{-1}X_\beta^t)Cov(y)(X_\tau^t - NK^{-1}X_\beta^t)^t$$

$$= (X_\tau^t - NK^{-1}X_\beta^t)(\sigma^2 I_v)(X_\tau - X_\beta K^{-1}N^t)$$

$$= \sigma^2(X_\tau^t X_\tau - X_\beta^t K^{-1}N^t - NK^{-1}X_\beta^t X_\tau^t + NK^{-1}X_\beta^t K^{-1}N^t)$$

$$= \sigma^2(R - NK^{-1}N^t - NK^{-1}N^t + NK^{-1}N^t K^{-1}N^t)$$

$$= \sigma^2(R - NK^{-1}N^t) = \sigma^2 C \qquad \cdots\cdots \qquad ①$$

• $Cov(\hat{\tau}) = Cov(C^- q) = C^- Cov(q)(C^-)^t$

$= C^{-}(\sigma^2 C)(C^{-})^t$ ($\because$ ①)

$= C^{-}CC^{-}\sigma^2$ ($\because$ C가 대칭이므로 C^{-}도 대칭이므로 $(C^{-})^t = C^{-}$이다.) $\cdots$ ②

- $Var(a^t\hat{\tau}) = a^t Cov(\hat{\tau})a$

$= a^t(C^{-}CC^{-}\sigma^2)a$ ($\because$ ②)

$= (a^t C^{-}C)C^{-}\sigma^2 a$

$= a^t C^{-}\sigma^2 a$ ($\because a^t\tau$ 가 추정가능이므로 앞(c)에서 $a^t = a^t C^{-}C$)

$= (a^t C^{-}a)\sigma^2$

$ii)$ $\psi = a^t\tau$가 추정 가능한 모수함수이므로 (c)에서 $a^t = a^t C^{-}C$ 이다.

만약 $\ell^t = a^t C^{-}$ 라면 $a^t = \ell^t C$ 이다.

$Var(a^t\hat{\tau}) = Var(\ell^t C\hat{\tau})$

$= (\ell^t C)Cov(\hat{\tau})(\ell^t C)^t$

$= \ell^t C(\underbrace{C^{-}CC^{-}\sigma^2}_{①})\underbrace{C\ell}_{②}$

($\because$ ① : 앞 ②식)

($\because$ ② : C: 대칭行列)

$= \ell^t(CC^{-}C)C^{-}C \cdot \ell \cdot \sigma^2$

$= \ell^t CC^{-}C\ell\sigma^2$ ($\because$ $C = CC^{-}C$ 즉 C^{-} : C의 $G-inverse$)

$= \ell^t \cdot C \cdot \ell \cdot \sigma^2$

즉 $\ell^t = a^t C^{-}$ 일때

$Var(a^t\hat{\tau}) = Var(\ell^t C\hat{\tau})$

$= \ell^t C\ell\sigma^2$ (일정)

이것은 C^- (여러 개 있음) 의 선택에 관계없이 $Var(a^t\hat{\tau})$는 일정함을 보여준다.

(정의) 연결 계획 : $rank\ C = v-1$ (혹은 $rank\ X = v+b-1$) 인 블록계획

<참고> 정보행렬 $C = R - NK^{-1}N^t$에서

$rank\ C \leqq v-1$ 이고 $rank\ X \leqq v+b-1$임.

(정리 2-2) $C = R - NK^{-1}N^t$가 블록계획의 정보행렬인 경우

모수함수 $\Psi = a^t\tau$ 가 추정 가능 $\Leftrightarrow$ $rank\ C = v-1$

(증명) $C\hat{\tau} = q$ 에서 $G,\ H$를 C의 임의의 두 $G-inverse$라 하자.

그러면 $\hat{\tau} = C^-q = Gq\ \ or\ \ Gq$

$q = C\hat{\tau} = CGq = CHq$

$C(G-H)q = 0$

이것은 $(G-H)q = t1_v$ 를 의미한다.

2.6. 분산분석

2.6.1 분 산 분 석

(1) 모형 $w_1 : y_{ij} = \mu + \beta_j + \epsilon_{ij}$ $(\hat{y} = \hat{\mu} + \hat{\beta}_j)$

$i = 1, 2, \cdots, v\ ,\ j = 1, 2, \cdots, b$

이것은 처리효과를 무시한 이원배치 축소모형이다.

$((\beta_i) \neq \beta_j)$블록 합 벡터인 $B = X_\beta^t y = \begin{pmatrix} B_1 \\ \dot{B}_j \\ \dot{B}_b \end{pmatrix}$ 의 j번째 값이라 하자.

그러면

$$\hat{\mu} = \overline{y_{\circ\circ}} = \frac{G}{n} \quad (G = 1_n^t y)$$

$$\hat{\beta}_j = \overline{y_{\circ j}} - \overline{y_{\circ\circ}} = \frac{B_i}{k_j} - \frac{G}{n}$$

(2) $SSw_1 = \sum_i \sum_j (y_{ij} - \hat{\mu} - \hat{\beta}_j)^2$

$$= \sum_i \sum_j (y_{ij} - \frac{B_i}{k_j})^2 \qquad (\because \ \hat{\mu} + \hat{\beta}_j = \frac{B_i}{k_j} \ : (5-32), (5-33) \text{ 에서})$$

$$= y^t y - \beta^t k^{-1} B$$

(3) $SSblocks = SST - SSw_1$

$$= \sum_i \sum_j (y_{ij} - \overline{y_{\circ\circ}})^2 - SSw_1$$

$$= \sum \sum y_{ij}^2 - \sum_i^v \sum_j^b \overline{y_{\circ\circ}}^2 - SSw_1$$

$$= y^t y - n\overline{y_{\circ\circ}}^2 - (y^t y - B^t k^{-1} B)$$

$$= B^t k^{-1} B - \frac{G^2}{n} \qquad (\because \ \overline{y_{\circ\circ}} = \frac{G}{n})$$

<참고> $SSblock = B^t k^{-1} B - \frac{G^2}{n} \qquad \left(\begin{array}{c} B = X_\beta \cdot t_y, B^t = y^t \cdot X_\beta \\ k = X_\beta^t \cdot X_\beta, k^{-1} = ? \end{array} \right)$

$$= \sum_j \frac{B_j^2}{k_j} - n\left(\frac{G}{n}\right)^2 = \sum_j k_j \left(\frac{B_j}{k_j}\right)^2 - \sum_j k_j \left(\frac{G}{n}\right)^2$$

$$(\because \sum_j^b k_j = n)$$

$$= \sum_j k_j \overline{y_{\circ j}}^2 - \sum_j k_j \overline{y_{\circ\circ}}^2 = \sum_j k_j (\overline{y_{\circ j}} - \overline{y_{\circ\circ}})^2$$

(4) $SSE = \sum_i \sum_j (y_{ij} - \hat{\mu} - \hat{\tau_i} - \hat{\beta_j})^2$

$$= (y - X\hat{\theta})^t (y - X\hat{\theta}) = (y^t - \hat{\theta}^t X^t)(y - X\hat{\theta})$$

$$= y^t y - \hat{\theta}^t X^t y - y^t X\hat{\theta} + \hat{\theta}^t X^t X\hat{\theta}$$

$$= y^t y + \hat{\theta}^t X^t X\hat{\theta} - 2\hat{\theta}^t X^t y$$

$$(\because\ y^t x\hat{\theta} = (y^t X\hat{\theta})^t = \hat{\theta}^t X^t y)$$

$$= y^t y + \hat{\theta}^t X^t y - 2\hat{\theta}^t X^t y$$

$$(\because\ (X^t X)\hat{\theta} = X^t y\ :\ \text{정규방정식})$$

$$= y^t y - \hat{\theta}^t X^t y$$

$$= y^t y - (\hat{\mu} \mid \hat{\tau}^t \mid \hat{\beta}^t)\begin{pmatrix} G \\ T \\ B \end{pmatrix} \quad \Rightarrow \quad \because \quad \hat{\theta}^t = \begin{pmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\beta} \end{pmatrix}^t = (\hat{\mu} \mid \hat{\tau}^t \mid \hat{\beta}^t)$$

$$= y^t y - (\hat{\mu} G + \hat{\tau}^t T + \hat{\beta}^t B) \qquad X^t y = (1_n \mid X_\tau \mid X_\beta)^t$$

$$= y^t y - \hat{\mu} G - \hat{\tau}^t . T \qquad = \begin{pmatrix} 1_n^t y \\ X_\tau^t y \\ X_\beta^t y \end{pmatrix} = \begin{pmatrix} G \\ T \\ B \end{pmatrix}$$

$$- (K^{-1}B - K^{-1}N^t\hat{\tau} - 1_b\hat{\mu})^t B$$

$$= y^t y - \hat{\mu} G - \hat{\tau}^t T - B^t K^{-1} B + \hat{\tau}^t N K^{-1} B + \hat{\mu} 1_b^t B$$

$$= y^t y - B^t K^{-1} B + \hat{\tau}^t (T - NK^{-1}B) - \hat{\mu}\ (G - 1_b^t B)$$

$$= y^t y - B^t K^{-1} B - \hat{\tau}^t q$$

$$(\ \because\ q = T - NK^{-1}B, \quad G = 1_b^t B = 1_v^t T\ )$$

(5) $SS_{tr} = SSw_1 - SSE$

$$= (y^t y - B^t K^{-1} B) - (y^t y - B^t K^{-1} B - \hat{\tau}^t q)$$

$= \hat{\tau}\,^{t}q$

2.6.2 블록의 효율성

처리가 v 인 모든 계획은 연결계획임을 가정 (∴계획d의 정보 C_d 가 $v \times v$ 일때 $rank\ C_d = v-1$)

<정의> 처리효과와 블록효과가 직교 ⇆
두 효과가 독립(처리 효과는 블록 효과의 존재 여부에 무관함)

<참고> 불완비 블록계획은 직교 블록계획이 아님

* 불완비 블록계획의 모형은 다음과 같다
$y_{ijl} = \mu + \gamma_i + \beta_j + \epsilon_{ijl}$
$i = 1,2 \cdots, v$, $j = 1,2 \cdots, b$, $l = 1,2 \cdots, n_{ij}$
μ : 전체 평균
γ_i : 처리 i의 효과
β_j : 블록 j의 효과

1. 기호 :

$D(v, b, r, k)$: v개의 처리 , b개의 블록, 각 처리의 반복수가

$r_1 = r_2 = \cdots = r_n = r$로 같고, 각 블록의 크기가

$k_1 = k_2 = \cdots = k_b = k$로 같은 블록 계획의 집합

2. 평균 효율의 정의

임의의 $d \in D(v, b, r, k)$ 에 대하여

$$E_d = \frac{\text{확률화 블록 계획에 있어서 } \widehat{\gamma_i} - \widehat{\gamma_j}\text{의 분산의 평균}}{\text{불완비 블록 계획에 있어서 } \widehat{\gamma_i} - \widehat{\gamma_j}\text{의 분산의 평균}} : d\text{의 평균 효율}$$

E_d: average Efficiency in $d \in D$ based on pairwise treatment differences.

3. $d \in D(v, b, r, k)$의 정보행렬:

$$C_d = R - NK^{-1}N^t$$

$$= \underbrace{rI_v}_{①} - \underbrace{\frac{1}{k}NN^t}_{②} \quad (\because r_1 = \cdots = r_v = r,\ k_1 = \cdots k_b = k)$$

<참고> ① $R = \begin{pmatrix} r_1 & & \cdots & 0 \\ & r_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & r_v \end{pmatrix} = \begin{pmatrix} r & & \cdots & 0 \\ & r & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & r \end{pmatrix} = rI_v$, N : 빈도행렬

$$②NK^{-1}N^t = N\begin{pmatrix} \frac{1}{k_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{k_b} \end{pmatrix}N^t = N\begin{pmatrix} \frac{1}{k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{k} \end{pmatrix}N^t$$

$$= \frac{1}{k}NI_bN^t = \frac{1}{k}NN^t$$

4. 평균효율 E_d

(1) E_d 의 분자 :

i) 정보행렬 : d 가 확률화 블록 계획이라면 $r = b, k = v$ 이고 빈도행렬은 $N = J_{v \times b}$이다. ($J_v = J_{v \times v}$) (r:각 처리의 반복이 일정하게 k:각 블록에서 나타난 처리의 수)

$$C_d = rI_v - \frac{1}{k}NN^t$$

$$= rI_v - \frac{1}{v} J_{v \times b} J_{v \times b}{}^t \quad (\text{since } r = b, k = v; N = J_{v \times b})$$

$$= rI_v - \frac{r}{v} J_v$$

(2) E_d의 분모

임의의 A가 K차 대칭행렬에, $\lambda_i (i = 1, \cdots, k)$를 A의 i번째 고유치, $P_i (i = 1, \cdots, k)$를 λ_i 에 대응한 A의 정규직교 고유 vector라 하자

(즉, $P_i{}^t P_i = 1, P_i{}^t P_j = 0 \ (i \neq j)$

$$A = \sum_{i=1}^{k} \lambda_i P_i P_i^t \text{ 이다}$$

⑤

i)정보행렬

정보행렬 C_d 는 v차 대칭행렬이므로 다음과 같이 분해된다

$$C_d = \sum_{i=1}^{v} \lambda_i P_i P_i{}^t$$

또한 d가 연결계획 즉 $rank\ C_d = v - 1$이므로 λ_i중 0 인 것은 제거하면 다음과 같이 표시된다

$$C_d = \sum_{i=1}^{v-1} \lambda_i P_i P_i{}^t$$

ii) G-inverse C_d^- :

$$C_d^- = \sum_{i=1}^{v} \frac{1}{\lambda_i} P_i P_i{}^t \quad (\because rank\ \ C_d = v \text{ 인 경우})$$

⑥

$$= \sum_{i=1}^{v-1} \frac{1}{\lambda_i} P_i P_i{}^t \quad (d\text{가 연결계획인 경우})$$

iii) 분산 :

$$\qquad\qquad\qquad\qquad j \qquad\qquad\qquad\qquad k$$

$1 \times v$ 행렬에 $a_{jk}^{t} = (0, \cdots, 0, 1, 0, \cdots, 0, -1, 0, \cdots, 0)$에 대하여

$$Var(a_{jk}^{t}\hat{\gamma}) = Var(\widehat{\gamma_j} - \widehat{\gamma_k}) = a_{jk}^{t} \bar{C_d} a_{jk} \sigma^2$$

$$= a_{jk}^{t} \left(\sum_{i=1}^{v-1} \frac{1}{\lambda_i} P_i P_i^{t} \right) a_{jk}^{t} \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{v-1} \frac{1}{\lambda_i} (a_{jk}^{t} P_i)(a_{jk}^{t} P_i)^t$$

$$= \sigma^2 \sum_{i=1}^{v-1} \frac{1}{\lambda_i} (a_{jk}^{t} P_i)^2$$

iv) $\frac{1}{r} C_d$ 의 고유치 및 분해 :

$C_d = \sum_{i=1}^{v} \lambda_i P_i P_i^t$ 이므로

$$\frac{C_d}{r} = \frac{1}{r} \sum_{i=1}^{v} \lambda_i P_i P_i^t = \sum_{i=1}^{v} \frac{\lambda_i}{r} P_i P_i^t = \sum_{i=1}^{v} e_i P_i P_i^t$$

$e_i = \frac{\lambda_i}{r}$ 는 $\frac{C_d}{r}$의 고유치이다.

또한 $e_i = \frac{\lambda_i}{r} \leq 1$이라 제한한 $\frac{C_d}{r}$의 고유치를 표준 효율인자라 한다

v) E_d의 분모 :

$$\text{분모} = \frac{1}{{}_vC_2} \sum_{j<k}^{v} Var(\widehat{\gamma_j} - \widehat{\gamma_k})$$

$$= \frac{2}{v(v-1)} \sum_{j<k} \left(\sigma^2 \sum_{i=1}^{v-1} \frac{1}{\lambda_i} (a_{jk}^{t} p_i)^2 \right)$$

$$= \frac{2\sigma^2}{v(v-1)} \frac{1}{r} \sum_{i=1}^{v-1} \frac{1}{e_i} \left(\sum_{j<k} (a_{jk}^{t} p_i)^2 \right) \quad (\because e_i = \frac{\lambda_i}{r})$$

$$= \frac{2}{rv(v-1)} \sum_{i=1}^{v-1} \frac{1}{e_i} \underbrace{\left(vP_i^t P_i - (P_i^t 1v)^2\right)}_{⊛}$$

$$= \frac{2\sigma^2}{r} \sum_{i=1}^{v-1} \frac{e_i^{-1}}{v-1} \quad \underbrace{\left(\because \sum_{i=1}^{v-1} \frac{(P_i^t 1v)^2}{e_i} = 0\right)}_{⑦}$$

※ 주어진 A가 정칙이므로 A^{-}가 아니고 $A^{-1} = \sum_{i}^{v} \frac{1}{\lambda_i} p_i p_i^{\,t}$

※ 주어진 C_d가 비정칙이면 $C_d^{-} = \sum_{i}^{v-1} \frac{1}{\lambda_i} p_i p_i^{\,t}$

vi) 평균 효율

$$E_d = \frac{\frac{2}{r}\sigma^2}{\frac{2\sigma^2}{r}\sum_{i=1}^{v-1}\frac{e_i - 1}{v-1}} = \frac{v-1}{\sum_{i=1}^{v-1} e_i^{-1}}$$

5. <정의> 최적 계획

계획 $d \in D(v, b, r, k)$ 에서

(1) A_d-최적 계획 ⇆ 평균효율 E_d 을 최소로

(2) E_d-최적 계획 ⇆ 최소의 e_i를 최대로 ($\hat{\gamma}_i - \hat{\gamma}_j$의 최대분산을 최소화) 하는계획

(3) D_d-최적 계획 ⇆ e_i를 최대로 하는 계획(신뢰 구역의 체적(Volum)을 최소로)

(4) 총체적 최적 계획 ⇆ A_d-, E_d-, D_d- 최적 계획을 동시에 만족하는 계획

(5) (M · S)최적 계획 ⇆ $\sum_{i=}^{v-1} e_i$가 최대인 계획 중 $\sum_{i=}^{v-1} e_i^{\,2}$ 를 최소로 하는 계획

2.7 요인실험법

2.7.1 완전확률화계획에 의한 요인실험

1) 다요인 실험법

(정의) $a_1^{f_1} \times a_2^{f_2} \times \cdots \times a_q^{f_q}$

여기서 $a_1,\ a_2,\ \cdots\ a_q$: level

q : 다른 수준의 수

$f_i\ (i=1,\ \cdots q)$: 수준 a_i에 대한 요인(factor)수

$k = \sum_{i=1}^{q} f_i$: 총요인 수

$F_1,\ F_2,\ \cdots\ F_k$: factor

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

treatment combination은 너무 복잡함으로 다음 예제로 대체함

1. 일원배치법

모형 $y_{ij} = \mu + \tau_i + \epsilon_{ij}\ \ i = 1 \cdots a\ ,\ j = 1 \cdots n$

$$N = \sum_{i=1}^{a}\sum_{j=1}^{n} 1 = an$$

(1) 최소제곱 추정량 :

$\hat{\mu} = \bar{y}_{..}\ ,\ \hat{\tau_i} = \bar{y}_{i.} - \bar{y}_{..}$

(2) 분산분석

$$SS_\tau = \sum_{i=1}^{a}\sum_{j=1}^{n} \hat{\tau_i} = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$= n\sum_{i=1}^{a}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \text{ 또는 } \frac{N}{a}\sum_{i}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$SSE = \sum_{i}\sum_{j}(y_{ij} - \bar{y}_{i\cdot})^2$$

$$SST = \sum_{i}\sum_{j}(y_{ij} - \bar{y}_{\cdot\cdot})^2$$

2. 이원배치법(두 요인 실험법)

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 :

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

$\hat{\tau}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$ $\hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$

교호작용 $\overline{(\tau\beta)}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}$

(2) 분산분석

$$SS_\tau = \sum_{i}\sum_{j}\sum_{k}\hat{\tau}_i = \sum_{i}\sum_{j}\sum_{k}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$$

$$= nb\sum_{i}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$$

(여기서 $nb = \frac{N}{a}$ 임에 유의)

$$SS_\beta = \sum_{i}\sum_{j}\sum_{k}\hat{\beta}_j = \sum_{i}\sum_{j}\sum_{k}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$$

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

(여기서 $an = \dfrac{N}{b}$ 임)

$$SS_{\tau\beta} = \sum_i\sum_j\sum_k(\widehat{\tau\beta})_{ij} = \sum_i\sum_j\sum_k(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2$$

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

(여기서 $n = \dfrac{N}{ab}$ 임)

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

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

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

3. 3요인 실험법

모형 $y_{ijk\ell} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk}$

$$+ (ABC)_{ijk} + \epsilon_{ijk\ell} \quad (6.4)$$

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

$$N = \sum_i\sum_j\sum_k\sum_\ell 1 = a_1a_2a_3n$$

(1) 최소제곱 추정량

$\hat{\mu} = \bar{y}_{\cdots\cdot}$

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

$$\widehat{C}_k = \bar{y}_{..k.} - \bar{y}_{....}$$

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

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

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

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

(2) 분산분석

$$SS_A = \sum_{i=1}^{a_1}\sum_{j=1}^{a_2}\sum_{k=1}^{a_3}\sum_{\ell=1}^{n}\widehat{A}_i = \sum_i\sum_j\sum_k\sum_\ell(\bar{y}_{i...} - \bar{y}_{....})^2$$

$$= a_2a_3n\sum_i(\bar{y}_{i...} - \bar{y}_{....})^2$$

$$SS_B = \sum_{i=1}^{a_1}\sum_{j=1}^{a_2}\sum_{k=1}^{a_3}\sum_{\ell=1}^{n}\widehat{B}_i = \sum_i\sum_j\sum_k\sum_\ell(\bar{y}_{.j..} - \bar{y}_{....})^2$$

$$= a_1a_3n\sum_{j=1}^{a_2}(\bar{y}_{.j..} - \bar{y}_{....})^2$$

$$SS_C = a_1a_2n\sum_{k=1}^{a_3}(\bar{y}_{..k.} - \bar{y}_{....})^2$$

$$SS_{AB} = \sum_{i=1}^{a_1}\sum_{j=1}^{a_2}\sum_{k=1}^{a_3}\sum_{\ell=1}^{n}(\widehat{AB})_{ij} = a_3n\sum_{i=1}^{a_1}\sum_{j=1}^{a_2}(\bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....})^2$$

$$SS_{AC} = a_2n\sum_{i=1}^{a_1}\sum_{k=1}^{a_3}(\bar{y}_{i.k.} - \bar{y}_{i...} - \bar{y}_{..k.} + \bar{y}_{....})^2$$

$$SS_{BC} = a_1n\sum_{j=1}^{a_2}\sum_{k=1}^{a_3}(\bar{y}_{.jk.} - \bar{y}_{.j..} - \bar{y}_{..k.} + \bar{y}_{....})^2$$

$$SS_{ABC} = n\sum_i\sum_j\sum_k(\bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{i.k.} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{..k.} - \bar{y}_{....})^2$$

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

$$SST = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SSE$$

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

4. 다요인 실험법

모형 $y_{ijk\cdots\ell} = \mu + A_i + B_j + C_k + \cdots + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + \cdots$

$+ (ABC)_{ijk} +$(3차 교호작용 나머지) + (4차 이상 교호작용)

$+ \epsilon_{ijk\cdots\ell}$

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

(1) 최소제곱 추정량

$\hat{\mu} = \bar{y}_{\cdots\cdot}$

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

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

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

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

……

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

$+ \bar{y}_{i\cdots\cdot} + \bar{y}_{\cdot j\cdot\cdots\cdot} + \bar{y}_{\cdot\cdot k\cdot\cdots\cdot} + \bar{y}_{\cdots\cdot}$

……

(2) 분산분석

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

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

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

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

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

……

$$SS_{AB} = \frac{N}{a_1a_2}\sum_{i=1}^{a_1}\sum_{j=1}^{a_2}(\bar{y}_{ij\cdots\cdot} - \bar{y}_{i\cdots\cdot} - \bar{y}_{\cdot j\cdots\cdot} + \bar{y}_{\cdots\cdot})^2$$

……

$$SS_{ABC} = \frac{N}{a_1a_2a_3}\sum_{i=1}^{a_1}\sum_{j=1}^{a_2}\sum_{k=1}^{a_3}(\bar{y}_{ijk\cdot\cdots} - \bar{y}_{ij\cdot\cdots} - \bar{y}_{i\cdot k\cdot\cdots}$$

$$- \bar{y}_{\cdot jk\cdot\cdots} + \bar{y}_{i\cdots\cdot} + \bar{y}_{\cdot j\cdots\cdot} + \bar{y}_{\cdot\cdot k\cdot\cdots}$$

$$- \bar{y}_{\cdots\cdot})^2$$

……

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

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

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

2.7.2 2^k-요인 실험

1. (정의) S^k : (1) S : level 이고 1, 2, ⋯ s-1 사용

(2) k : factor 수 A, B, ⋯ K 등으로 요인을 표시

2. 2^2-요인 실험 :

(1) S=2 : level

K=2 : factor 수

두요인 $A=\{A_0, A_1\}$ $B=\{B_0, B_1\}$ (∵ 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 :

		B	
		B_0	B_1
A	A_0	00	01
	A_1	10	11

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

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

$\hat{A}=A$의 효과 $=\hat{A}_1-\hat{A}_0=\dfrac{1}{2}[(10+11)-(00+01)]$ (4.1)

$=\dfrac{1}{2^2-1}$ (대비 A)

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

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

$\hat{B} = B$의 효과 $= \hat{B}_1 - \hat{B}_0 = \frac{1}{2}[(01+11)-(00+10)]$ (4.2)

$= \frac{1}{2^2-1}$ (대비 B)

3) 교호작용 AB의 효과

$\widehat{AB} = AB$의 효과 $= \frac{1}{2}[(B_1$에서 A의 효과) - (B_0에서 A의 효과)]

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

$= \frac{1}{2}[(00+11)-(01+10)]$ (4.3)

4) 요인 효과 계수표 ((4.1),(4.2),(4.3)를 부호로 만 표시)

처 리 조 합	요 인 효 과		
	A	B	AB
00	−	−	−
10	+	−	+
01	−	+	−
11	+	+	+

3. 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 $2\times2\times2=8$

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

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

3) C의 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 일반화 하면 이 방법은 너무 복잡함으로 다음 요인 효과 계수표에 의하여 구한다.

요인효과 계수표

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

4. 2^K-요인 실험 :

(1) level $S=2$

factor 수 = k개

factor A, B, ⋯ K (모두 K개)

(2) treatment combination

1) 개수 $= 2^k$

2) 2^4인것의 예시

0000 1000 0100 1100 0010

1010 0110 1110 0001 1001

0101 1101 0011 1011 0111

1111 (모두 16개이며 표준순서로 썼음)

2^k인 경우 0과 1을 k자리로 구성하며 위 예시와 같이 일반화 하여 처리조합을 구성한다.

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

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

$=\frac{1}{2^{k-1}}$(대비$_{\mathrm{B}}$)

2) B, C, ⋯, K인 경우도 유사함

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

. . . . . .

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

5. 제곱 합

(1) 2^2-요인실험

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

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

$$= 2\left[(\bar{y}_{0\cdot} - \bar{y}_{\cdot\cdot})^2 + (\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot})^2\right]$$

$$= 2\left[\left(\frac{00+01}{2} - \frac{00+01+10+11}{4}\right)^2 + \left(\frac{10+11}{2} - \frac{00+01+10+11}{4}\right)^2\right]$$

$$= \frac{2}{16}\left[((00+01) - (10+11))^2 + ((10+11) - (00+01))^2\right]$$

$$= \frac{1}{8}\left[2((00+01) - (10+11))^2\right] (\because \text{위의 두 항은 같음})$$

$$= \frac{1}{4}(\text{대버}_{\text{AB}})^2$$

$$SS_{AB} = \frac{1}{2^2}(\text{대버}_{\text{AB}})^2$$

(2) 2^k − 요인실험(반복수가 1인 경우)

위의(1)을 일반화하면

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

. . . . . .

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

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

<참고>

1. 효과

(1) A의 효과$= \frac{1}{2^{3-1}\times 3}\left[(100+110+101+111) - (000+010+001+011)\right]$

$$= \frac{1}{12}\left[(222+216+228+220) - (258+240+240+216)\right]$$

$$= -5.67$$

(2) B의 효과$= \dfrac{1}{2^{3-1} \times 3}[(010+110+011+111)-(000+100+001+101)]$

$= \dfrac{1}{12}[(240+216+216+220)-(258+222+240+228)]$

$=-4.67$

(3) C의 효과 $= -2.67$

(4) AB의 효과 $= \dfrac{1}{12}[(000+110+001+111)-(100+010+101+011)]$

$= \dfrac{1}{12}[(258+216+240+220)-(222+240+228+216)]$

$= 2.33$

AC의 효과 $= 4.33$

BC의 효과 $= -0.67$

(5) ABC의 효과 $= \dfrac{1}{12}[(100+010+001+111)-(000+110+101+011)]$

$= \dfrac{1}{12}[(222+240+240+220)-(258+216+228+216)]$

$=0.33$

2. 제곱합

(1) $SS_A = \dfrac{1}{n2^k}(\text{대비}_{\text{A}})^2$

$= \dfrac{1}{3 \times 2^3}[(100+110+101+111)-(000+010+001+011)]^2$

$= \dfrac{1}{24}[(222+216+228+220)-(258+240+240+216)]^2$

$= \dfrac{1}{24} \times 68^2 = \dfrac{1}{24} \times 4624 = 192.67$

(2) $SS_B = \dfrac{1}{3 \times 2^3}[(240+216+216+220)-(258+222+240+228)]^2$

$= \dfrac{1}{24} \times 3136 = 130.67$

(3) $SS_C = \frac{1}{24}[(240+228+216+220)-(258+222+240+216)]^2$

$= \frac{1}{24} \times 32^2 = 42.67$

(4) $SS_{AB} = \frac{1}{24}[(258+216+240+220)-(222+240+228+216)]^2$

$= \frac{1}{24} \times 28^2 = \frac{1}{24} \times 784 = 32.67$

(5) $SS_{AC} = \frac{1}{24}[(258+240+228+220)-(222+216+240+216)]^2$

$= \frac{1}{24} \times 52^2 = \frac{1}{24} \times 2704 = 112.67$

(6) $SS_{BC} = \frac{1}{24}[(258+222+216+220)-(240+216+240+228)]^2$

$= \frac{1}{24} \times 8^2 = \frac{1}{24} \times 64 = 2.67$

(7) $SS_{ABC} = \frac{1}{24}[(222+240+240+220)-(258+216+228+216)]^2$

$= \frac{1}{24} \times 4^2 = \frac{1}{24} \times 16 = 0.67$

(8) $SST = \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\sum_{l=1}^{3}(y_{ijkl} - \bar{y}_{....})^2$

$= \sum_i \sum_j \sum_k \sum_l (y_{ijkl} - \frac{230}{3})^2$

$= (84-\frac{230}{3})^2 + (89-\frac{230}{3})^2 + (85-\frac{230}{3})^2 + (78-\frac{230}{3})^2$

$+ (79-\frac{230}{3})^2 + (83-\frac{230}{3})^2 + (80-\frac{230}{3})^2 + (82-\frac{230}{3})^2$

$+ (78-\frac{230}{3})^2 + (75-\frac{230}{3})^2 + (70-\frac{230}{3})^2 + (71-\frac{230}{3})^2$

$+ (75-\frac{230}{3})^2 + (75-\frac{230}{3})^2 + (72-\frac{230}{3})^2 + (75-\frac{230}{3})^2$

$$+(80-\frac{230}{3})^2+(73-\frac{230}{3})^2+(76-\frac{230}{3})^2+(69-\frac{230}{3})^2$$

$$+(71-\frac{230}{3})^2+(70-\frac{230}{3})^2+(74-\frac{230}{3})^2+(76-\frac{230}{3})^2$$

$$=\frac{5772}{9}=641.3333\cdots$$

(9) $SSE = SST-SS_A-SS_B-SS_C-SS_{AB}-SS_{AC}-SS_{BC}-SS_{ABC}$

$= 641.3333-192.6667-130.6667-42.6667-32.6667-112.6667$

$-2.6667-0.6667 \fallingdotseq 126.6664 \fallingdotseq 126.67$

2.7.3 S^k-요인실험

1. k=2인 경우(3^2-요인실험법)

(1) level : s=3

factor : A, B (k=2개)

(2) treatment combination

1) 개수 = 3^2개

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

(3) 직교분해 (직교다항식의 이용방법)

1) 요인 A의 선형(1차)효과 : 직선의 기울기들의 평균

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

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

$$=\frac{1}{2}(-101)\begin{pmatrix}\bar{y}_{0\cdot}\\ \bar{y}_{1\cdot}\\ \bar{y}_{2\cdot}\end{pmatrix}$$

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

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

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

4) 요인 B의 2차효과

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

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

5) 직교다항계수행렬

① S=3인 경우 $\begin{pmatrix} \text{1차다항계수} \\ \text{2차다항계수} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}$ (앞 1) 2) 참조)

②S=4인 경우 $\begin{pmatrix} \text{1차 다항계수} \\ \text{2차 다항계수} \\ \text{3차 다항계수} \end{pmatrix} = \begin{pmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix}$ 각행 합 = 0

(나머지 교과서 S=5, S=6, S=7인 경우 참조)

i) $A_L = \frac{1}{6}\Big[(\bar{y}_{1\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{2\cdot} - \bar{y}_{0\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{0\cdot})$

$+ (\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{1\cdot}) + (\bar{y}_{3\cdot} - \bar{y}_{2\cdot})\Big]$

$$= \frac{1}{6}(-3\bar{y}_{0\cdot} - \bar{y}_{1\cdot} + \bar{y}_{2\cdot} - 3\bar{y}_{3\cdot})$$

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

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

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

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

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

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

즉 $A_C =$ (차)-(차)

(4) 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 A의 수준수가 a이고 반복이 N인 실험에서 요인 A에 대한 직

교대비 제곱합은 위로부터

$$\frac{\left(\sum_{i=1}^{a} C_i y_{i..}\right)^2}{\left(an\sum_{i=1}^{a} C_i^{\,2}\right)}$$ 이다.

2) $= \dfrac{(-y_{0..}+y_{2..})^2}{3\times n\left((-1)^2+0^2+1^2\right)} = \dfrac{(y_{2..}-y_{0..})^2}{6n}$

$= \dfrac{(-y_{0..}-2y_{1..}+y_{2..})^2}{3\times n\left(1^2+(-2)^2+1^2\right)} = \dfrac{(y_{2..}-2y_{1..}+y_{0..})^2}{18n}$

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

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

3) $= \dfrac{(y_{.2.}-y_{.0.})^2}{6n}$

$= \dfrac{(y_{.2.}-2y_{.1.}+y_{0..})^2}{18n}$

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

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

4) $n\sum_{i=1}^{a_1}\sum_{j=1}^{a_2}(\overline{y}_{ij.}-\overline{y}_{i..}-\overline{y}_{.j.}+\overline{y}_{...})^2$

$n\sum_{i=1}^{2}\sum_{j=1}^{2}(\overline{y}_{ij.}-\overline{y}_{i..}-\overline{y}_{.j.}+\overline{y}_{...})^2$

$n\sum_{i=1}^{a_1}\sum_{j=1}^{a_2}(y_{ij.}-\overline{y}_{ij.})^2$

$n\sum_{i=1}^{2}\sum_{j=1}^{2}(y_{ij.}-\overline{y}_{ij.})^2$

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

또는 $\sum_{i=1}^{a_1} \sum_{j=1}^{a_2} n\,(y_{ij\cdot} - \overline{y}_{\cdots})^2$

Ⅲ. 교락 디자인

3.1 크기*2*인 교락 디자인

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

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

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

여기서 $b = 2^{m-1}$는 블록의 수를 나타낸다. $h_j = (h_{j1}, h_{j2}, \cdots, h_{jb})'$이면, $E(h_j' y_d) = A_j$, $j = 1, 2, \cdots, 2^{m-1}$,이다. 그러면 분명히 h_j, $j = 1, 2, \cdots, 2^{m-1}$,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 i번째 배열에 교배시키는 2개의 처리 조합을 위한 $\log_2$ 표현비율과 일치한다.

모든 2^{m-1}개의 요인효과가 부분 교락된 설계법은 요인효과의 다른 집합을 교락한 그이상의 반복을 추가해서 얻은 것으로 추정할지도 모른다. 우리가 2^m개의 실험을 이용하여 교락법을 설명했지만, 그 방법은 소수 또는 소수의 거듭제곱인 다수 수준의 요인에 일반적으로 적용할 수 있다. 실제로 유사한 결론으로 고전적인 교락법이외의 방법을 이용하여 얻은 블록 크기 2의 균형요인배치법(balanced factorial designs)을 위해 또한 지속된다.

Ⅳ. 요인 Microarrays

4.1 2^2 실험

2^2 실험에서 우리는 2개 수준 각각에 2개의 요인 F_1 과 F_2를 가지고 있다. 전과같이 각 요인의 두개 수준은 0과 1로 코드를 한다. F_1 과 F_2에 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_d요인의 두 개 수준의 색깔을 0(cy5:빨강)과 1(cy3:녹색)로 된 코드로 표시할 것이다. 데이터의 모델은 아래와 같이 쓰여진다.

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

i번째 블록으로부터의 블록내 비교 $y_{i1} - y_{i2}$는 i번째 배열에서 교배시킨 두 처리조합 t_{i1}과 t_{i2}의 $\log_2$ 표현비와 일치한다. $\delta_1 - \delta_2$는 색깔 효과, τ_j는 j번째 처리조합의 효과, 그리고 ϵ_{i1i2}은 평균이 0, 분산이 σ^2를 갖는 독립적으로 분포된 랜덤 오차이다. $\log_2$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 표1에 주어졌다.

표1 2^2요인 모델

모델	블럭	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 2^2실험에서 반복모형 모델의 표현을 다른 방법으로 접근해보면

ⓧ를 요인별 부호곱이라 하고, 0을 -, 1을 +부호로한다. 주효과는 블록별 부호곱하여 "+"인 것이고, 교호작용은 블록별로 부호를 곱한 결과에서 2요인 이상 부호를 곱하여 "+"인 것으로 한다. 예를 들어 아래와 같다.

D_{11} : 00ⓧ01 = (−−ⓧ−+) = (+,−) ⇒F_1

D_{12} : 00ⓧ10 = (−−ⓧ+−) = (−,+) ⇒F_2

D_{13} : 00ⓧ11 = (−−ⓧ++) = (−,−) ⇒F_1F_2

위의 결과는 표1과 같다.

microarray 모델은 행요인 F_g(색깔)과 열요인 블록(슬라이드)을 갖는 행렬 모델이다. F_g의 수준 0(빨강)과 1(녹색)은 첫 번째와 두 번째열에 각각 해당한다. D_{11}과 D_{12}를 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_1과 F_2을 추정한다. 그에 반해, 교호작용 F_1F_2는 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 취함으

로써 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_2의 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_2를 추정한다. 모델의 선택은 바라는 효율 또는 대상의 대조 추정의 정확에 의존할 것이다.

4.2 2^3실험

교락의 방법을 사용하여 얻어진 2^3실험을 위한 모든 가능한 단일 반복 모델은 표2에 표현됐다.

표2 2^3요인을 위한 모델

모델	블럭	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, $F_1F_2F_3$

2^3실험 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⊕ 001)(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ⓧ100) = (−−−ⓧ+−−) = (− + +)

F_2, F_3, F_2F_3

D_{24} : (000ⓧ111) = (−−−ⓧ+++) = (− − −)

F_1F_2, F_1F_3, F_2F_3

D_{27} : (000ⓧ110) = (−−−ⓧ++−) = (− − +)

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 것이

다. 대개 3인자와 더 높은 서열의 교호작용들은 결여된 것으로 가정할 것이다. 그러나, 적절한 microarray 모델은 3인자 교호작용 또한 효율적으로 추정되어졌다면 표2의 모델로부터 더 쉽게 구성할 수 있다.

4. 3 3×2 실험

3×2실험을 위한 요인 모델은 Lewis and Tuck(1985)에 의해서 작성되어졌다. 일반화된 순환 방법의 구성에 기본을 둔 이러한 모델들은 표3에 표현됐다.

표3 3×2요인을 위한 모델

모델	블럭	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) gerneralized 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) ⊕ 10(mod(3, 2)) = (10, 11) ⊕ 10 mod(3, 2)

= (20, 21) ⊕ 10 mod(3, 2) = (00, 01) : main block

D'_{41} (00, 01) ⊕ 20 mod(3, 2) = (20, 21) ⊕ 20 mod(3, 2)

= (10, 11) ⊕ 20 mod(3, 2) = (00, 01) 순환됨

∴ $D_{41} \cong D'_{41}$

2) D_{42} (00, 11) ⊕ 10(mod(3, 2)) = (10, 21) ⊕ 10 mod(3, 2)

= (20, 01) ⊕ 11 mod(3, 2) = (01, 10) ⊕ 10 mod(3, 2)

= (11, 20) ⊕ 10 mod(3, 2) = (21, 00) ⊕ 11 mod(3, 2)

= <u>(00, 11)</u> : 순환

main block

(2) 다른 main design

i) D_{44} (00, 20) (10,00) (20,10)

[처음부터 ⊕ 10 mod(3, 2) ,⊕ 10 mod(3, 2),⊕ 11 mod(3, 2) 순환]

(01, 21) (11,01) (21,11) (00,20) [순환]

[처음부터 ⊕ 10 mod(3, 2), ⊕ 10 mod(3, 2), ⊕ 11 mod(3, 2)]

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

[처음부터 ⊕10, ⊕10, ⊕11, ⊕10, ⊕10, ⊕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 외.(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.

4.4 2^4실험

교락의 방법을 사용하여 얻어진 2^4실험을 위한 모든 가능한 단일 반복 모델은 다음과

같이 표현됐다

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 부호곱 = (－ － － ＋)

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 부호곱 = (－ － － －)

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

Ⅴ. 결론 및 토론(*Concoluding and Dicussion*)

문헌에 의하면 바람직한 통계적인 성질을 갖는 microarray 디자인 블록 구성을 위하여 이용 가능한 블럭 균형요인디자인 소개 되어 있다, 특히, 2수준 요인에 대한 고전적인 교락 디자인들이 microarray 디자인으로는 효율적이라는 사실을 쉽게 알 수 있다.
Lewis와 Tuck(1985) 그리고 Gupta(1987)에 의해 표(table1, table2, table3)로 만들어진 블록크기2인 블럭 모델들이 효과적이라는 것을 입증 하였다.따라서 본 논문에서는 요인의 수와 수준이 많아 지면 상상할 수 없을만큼 아주 복잡해지므로 단일 실험에 의한 블록을 만드는 방법과 고전적인 교락행렬계획법에 의한 단일실험 요인모델에 의한 행렬 계획과 주로 고전적인 디자인과 연결 하여 그의 관계 나타내어 수준과 요인이 동시에 많거나 수준과 요인이 독립적으로 많을 때를 고려하여 균형 블록 디자인을 사용하였다. 아울러 향후 본 디자인에 의한 실험분석은 4개의 카테고리(classical methods, tree methods, machine learning method, 그리고 최근 일반화된 LDA(Linear discriminant analysis) 및 PAM(Predictive Analysis of Microarrays) method)를 연구하여 비교분석하는데 역점을 두어야 할 것이다.

참고 문헌

Abraham, B., H. Chipman, and K. Vijayan. (1999). "Some Risks in the Construction and Analysis of Supersaturated Designs." Technometrics, Vol. 41, pp. 135-141.

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

Barnett, V., and T. Lewis (1994). Outliers in Statistical Data, 3rd edition. Wiley, New York.

Bingham, D., and R. R. Sitter. (1999). "Minimum Aberration Two-Level Fractional Factorial Split-Plot Designs." Technometrics, Vol. 41, pp. 62-70.

Bisgaard, S. (1998-1999). "Conditional Inference Chart for Small Unreplicated Two-Level Factorial Experiments." Quality Engineering, Vol. 11, pp. 267-271.

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.

Box, G. E. P. (1992-1993). "Sequential Experimentation and Sequential Assembly of Designs." Quality Engineering, Vol. 5, No. 2, pp. 321-330.

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.

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.

Burdick, R. K., and F. A. Graybill (1992). Confidence Intervals on Variance Components. Dekker, New York.

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

Cornfield, J. and Tukey, J.W. (1956). Average Values of Mean Squares in Factorials, Annals of Mathematical Statistics, 27, 907-949.

Daniel, C. and Wilcoxon, F. (1966). Factorial 2^{p-q} Plans Robust Against Linear and Quadratic Trends, Technometrics, 8, 259-278.

Fries, A. and Hunter, W.G. (1980). Minimum Aberration 2^{k-p} Designs, Technometics, 22, 601-608.

Jones, B. (1980). Algorithm AS 156: Combining two component designs form a row-and-column design. Applied Statistics, 29, 334-337.

Montgomery, D.C. (1976). Design and Analysis of Experiments, John Wiely & Sons, New York.

Paired comparison designs for factorial experiments. Applied Statistics. 34:227-234.

Kshirsagar, A.M. (1966). Balanced factorial designs. Journal of the royal Statistical Society. B28:59-67

Raghavarao, D. (1971). Construcions and Combinatorial Problems in Design of Experiments. New York:Wiley.

Shah, B.V. (1960). Balanced factorial experiments. Annals of Mathematical Statistics.

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.

Cornell, J. A. (1990). Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data. 2nd edition. Wiley, New York.

Hamada, M., and N. Balakrishnan (1998). "Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals" (with Discussion). Statistica Sinisa, Vol. 8, pp. 1-41.

Hamada, M., and C. F. J. Wu (1992). "Analysis of Designed Experiments with Complex Aliasing." Journal of Quality Technology, Vol. 24, No. 3, pp. 130-137.

Hines, W. W., D. C. Montgomery, D. M. Goldsman, and C. M. Borror. (2003). Probability and Statistics in Engineering. 4th edition. Wiley, New York.

Holcomb, D. R., D. C. Montgomery, and W. M. Carlyle. (2003). "Analysis of Supersaturated Designs." Journal of Quality Technology, Vol. 35, No. 1, pp. 13-27.

Huang, P., D. Chen, and J. Voelkel. (1999), "Minimum Aberration Two-level Split-Plot Designs." Technometrics, Vol. 41, pp. 314-326.

Khuri, A. I., and J. A. Cornell (1996). Response Surfaces: Designs and Analyses. 2nd edition. Dekker, New York.

Larntz, K., and P. Whitcomb (1998). "Use of Replication in Almost Unreplicated Factorials." Presented at the Fall Technical Conference, Corning, N.Y.

Li, W. W., and C. F. J. Wu. (1997). "Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs." Technometrics, Vol. 39, pp. 171-179.

Lin, D. K. J. (1993). "A New Class of Supersaturated Designs." Technometrics, Vol. 35, pp. 28-31.

Lin, D. K. J. (1995). "Generating Systematic Supersaturated Designs." Technometrics, Vol. 37, pp. 213-223.

Lin, D. K. J. (2000). "Recent Developments in Supersaturated Designs."

Chapter 18 in Statistical Process Monitoring and Optimization," S. H. Park and G. G. Vining, eds. Marcel Dekker, New York, pp. 305-319.

Loughin, T. M. (1998). "Calibration of the Lenth Test for Unreplicated Factorial Designs." Journal of Quality Technology, Vol. 30, pp. 171-175.

Loughin, T. M., and W. Noble (1997). "A Permutation Test for Effects in an Unre plicated Factorial Design." Technometrics, Vol. 39, pp. 180-190.

Mee, R. W.(2004). "Efficient Two-Level Designs for Estimating All Main Effects and Two-Factor Interactions." Journal of Quality Technology, Vol. 36, pp. 400-412.

Mee, R. W., and Peralta, M. (2000). "semifolding 2^{k-p} Designs." Technometrics, Vol. 42, No. 2, pp.122-143

Miller, R. G.(1991). Simultaneous Statisical Inference. Springer-Verlag, New York.

Montgomery, D. C.(2001). Introduction to Statistical Quality Control. 4th edition. Wiley, New York.

[Montgomery, D. C.(1999). "Experimental Design for Product and Process Design and Development." Journal of the Royal Statistical Society, D, Vol. 48, pp. 159-177

Montgomery, D. C., C. M. Borror, and J. D. Stanley (1997-1998). "Some

Cautions in the Use of Plackett-Burman Designs." Quality Engineering, Vol. 10, pp. 371-381

Montgomery, D. C., E.A. Peck, and G. G. Vining(2001). Introduction to Linear Regression Analysis. 3rd edidion. Wiley, New York.

Searle, S. R., G. Casella, and G. E. McCulloch (1992). Variance components. Wiley, New York.

Taguchi, G. (1991). Introduction to Quality Engineering. Asian Productivity Organization, UNIPUB, White Plains, N.Y.

Ting, N., R. K. Burdick, F. A. Graybill, S. Jeyaratnam, and T.-F. C. Lu (1990). "Confidence Intervals on Linear Combinations of Variance Components That Are Unrestricted in Sign." Journal of Statistical Computation and Simulation, Vol. 35, pp. 135-143.

Wu, C. F. J.(1993). "Construction of Supersaturated Designs Through Partially Aliased Interactions." Biometrika, Vol. 80, pp. 661-669.

Ye, K., and M. Hamada (2000). "Critical Values of the Lenth Method for Unreplicated Factorial Designs." Journal of Quality Technology, Vol. 32, pp. 57-66.

Montgomery, D. C., and G. C. Runger (1996). "Foldovers of 2^{k-p} Resolution IV Experimental Designs." Journal of Quality Technology, Vol. 28, pp. 446-450.

Montgomery, D. C., and G. C. Runger (2003). Applied Statistics and Probability

for Engineers. 3rd edition. Wiley, New York.

Myers, R. H. (1990). Classical and Modern Regression with Applications. 2nd edition. PNS-Kent, Boston.

Myers, R. H., and D. C. Montgomery (2002). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. 2nd edition. Wiley, New York.

Myers, R. H., and D. C. Montgomery (1997). "A Tutorial on Generalized linear Models."Journ al of Quality technology, Vol. 29, pp. 274-291.

Myers, R. H., and D. C. Montgomery, and G. G. Vining. (2002). Generalized Linear Models with Applications in Engineering and the Sciences. John Wiley & Sons, New York.

Myers, R. H., and D. C. Montgomery, and G. G. Vining, C. M. Borror, and S. M. Kowalski. (2004). "Response Surface Methodology: A Retrospective and Literature Survey." Journal of Quality Technology, Vol. 36, pp. 53-77.

Nair, V. N., et al., (eds.) (1992). "Taguchi's Parameter Design: A Panel Discussion." Technometrics, Vol. 34, pp. 127-161.

Nelson, L. S. (1995a). "Using Nested Designs Ⅰ: Estimation of Stansard Deviations." Journal of Quality Technology, Vol. 27, No. 2, pp. 169-171.

Nelson, L. S. (1995b). "Using Nested Designs Ⅱ: Confidence Limits for

Standard Deviations." Journal of Quality Technology, Vol. 27, No. 3, pp. 265-267.

국문초록

블록 디자인에 의한 *cDNA Microarray* 실험에 관한 연구

최 남 수
지도교수 : 최 규 정 교수님
학 과 : 전산 통계학과
조선대학교 대학원

일반적으로 실험(experiment)이란 물리, 화학, 의공학 등 사회 전분야에서 사용되는 중요한 용어로서 일정한 연구 대상에 대하여 여러 조건하에서 변화를 일으키게 하여 그 현상을 관찰 및 관측하는 행위라고 정의할 수 있다. 따라서 실험자는 한 실험을 통하여 결과물에 변화를 주는 이유를 관찰, 규명하게 된다. 현실적으로 어떤 현상을 설명해 줄 수 있는 원인은 무수히 많이 존재하며, 또한 비용, 시간 등 여러 가지 제약조건들은 완전한 원인규명을 더욱 어렵게 만든다. 이에 따라 실제 실험은 많은 원인들 중 경험적으로 혹은 사전 정보에 의해 몇가지를 선택하게 되는데 이와 같이 직접 실험에 취급된 요인들의 속성을 여러 각도에서 조명하여 원하는 목표에 이르도록 하는 일이 실험의 목적이라 할 수 있다.

기본 원칙을 요약하면 R.A.Fisher는 1920년대와 30년대에 걸쳐서 Rothamsted Experimental Station에서 일하는 동안 실험계획법에서의 통계학의 역할과, 역으로, 통계학에서의 실험계획법의 중요성을 기초한 선구자적인 학자이다. 그는 세 가지의 중요한 원리를 주창하였는데 그것들은 확률화(randomization), 블록화(blocking), 반복수(replicates)이다. 그리고 실험계획법의 절차로 문제의 인식 및 실험의 목적 설정, 제약조건의 이해, 적절한 모형의 설정(모형에 포함될 요인들과 그 요인들의 교호작용 중 필요한 요인들을 모형에 포함시켜 공변성(covariance)이 있음를 확인하였으며 모형이 교차(crossed) 모형인지 지분(nested) 모형인지를 고려하며, 지분모형인 경우 어느 요인이 지분되었는가를 고려하였다. 또한, 모형이 고정 효과(fixed effect) 모형인지 확률(random) 효과 모형인지를 탐색하여 각 모형에 있어서 오차에 관한 정규성 가정이 타당한지를 고려),반복수의 결정(반복수의 결정은 제약 조건 아래 요구되는 신뢰구간의 길이나 검정력에 의거), 확률화, 실험계획법의 선택, 예비실험, 자료 수집과 분산분석의 과정을 따랐다.

본 논문에서는 고전적인 교락법을 써서 단 한번의 실험으로 cDNA microarray디자인을 위한 바람직하고 타당한 균형 블록디자인을 연구하였다. 본 논문에서 얻어진 고전적인 요인실험계획법이 다른 어떤 교락법 보다도 유용함을 확인하였다. 특히 알고리즘과 상황에 따라 적당한 방법보다는 microarray 실험디자인을 위한 대칭적인 블록 디자인을 제안하였으며, 최근문헌에 의한 여러 가지 microarray 디자인들을 일반화하였다.