Journal of the Korean Data & Information Science Society
2015, **26**(2), 347–354

http://dx.doi.org/10.7465/jkdi.2015.26.2.347
한국데이터정보과학회지

균형불완비블럭설계의 사영분석

최재성[1]

[1]계명대학교 통계학과

접수 2015년 1월 9일, 수정 2015년 1월 26일, 게재확정 2015년 2월 9일

요 약

본 논문은 균형불완비블럭설계 (balanced incomplete block design)의 자료분석에 사영을 이용한 블럭내 분석 (intrablock analysis)방법을 다루고 있다. 블럭내 분석을 위해 단계별 방법 (stepwise procedure)에서 유도되는 분석모형을 이용하고 있다. 단계별 방법의 적용으로 인해 모형행렬로 주어지는 사영공간이 변동요인에 따른 부분공간들로 직교분할됨을 보여주고 있다. 단계별 과정에서 변동요인에 따른 변동량을 구하기 위해 해당하는 효과벡터의 계수행렬에 근거한 사영의 구조적 형태를 기술하고 있으며 상호직교하는 부분공간으로의 사영을 이용하여 블럭효과에 적합된 처리효과의 변동량를 구하는 과정을 구체적으로 다루고 있다. 또한, 사영에 의해 처리효과를 구하는 과정을 제시하고 있으며 단계별로 잔차벡터를 이용하여 모형설정하는 방법과 고유벡터에 의한 추정가능함수의 구성과 추정가능성을 논의하고 있다.

주요용어: 고유벡터, 균형불완비블럭, 단계별 방법, 블럭내 분석, 사영, 추정가능함수.

1. 서론

실험자료를 수집하기 위한 실험의 설계에서 실험단위들이 동질적이 아니면 동질적인 실험단위들의 그룹으로 분류하여 실험을 행하게 된다. 동질적인 실험단위들의 그룹은 블럭으로 정의되고 블럭화 (blocking)는 실험단위들 간의 변이효과를 줄이기 위해 이용되는 연구기법으로 간주된다. 실험단위들의 블럭화는 실험에서 고려되는 모든 처리들이 한 블럭내의 실험단위들에 임의로 배정되는 경우의 완비블럭설계 (randomized complete blocks design)와 블럭내의 실험단위들의 수가 비교하고자 하는 처리들의 수보다 적은 경우의 불완비블럭설계 (incomplete blocks design)를 고려하게 된다. 실험단위들의 블럭화와 관련된 다양한 설계 및 분석에 관한 논의가 Cochran과 Cox (1957), Hicks (1973), John (1971)과 Milliken과 Johnson (1984) 등에서 볼 수 있다.

실험단위들의 이질성으로 인해 동질적인 그룹의 블럭화가 필요하나 블럭내 실험단위의 수가 비교하고자 하는 처리의 수를 모두 수용할 수 없는 작은 크기의 블럭일 때 실험자는 불완비블럭설계를 이용하게 된다. 불완비블럭설계로 Youden 방격, 분할구 설계 (split-plot designs)나 반복측정설계 (repeated measures designs) 등을 생각할 수 있다. 실험에서 고려되는 모든 처리비교가 동일하게 중요할 때 각 블럭에서의 처리비교들은 균형적으로 선택되어야 한다. 즉, 처리들의 임의 쌍의 발생 수가 어떤 다른 쌍의 발생 수와 동일하게끔 균형을 맞추어 설계하는 것을 의미한다. 따라서 균형불완비블럭설계 (balanced incomplete block design)은 임의의 두 처리가 동일 회수로 함께 나타나는 불완비블럭설계이다. 불완비블럭설계의 구성과 분석에 관한 논의를 Yates (1936), Davies (1956) 그리고 John (1971) 등에서 살펴볼 수 있다.

[1] (704-701) 대구광역시 달서구 신당동 1000번지, 계명대학교 통계학과, 교수. E-mail: jschoi@kmu.ac.kr

Yates (1936)에 의해 소개된 균형불완비블럭설계는 b개 블럭에서 t개 처리에 대한 설계들이다. 개별 블럭은 k개 실험구를 포함한다. 단, $k < t$이다. 각 실험구는 하나의 처리를 포함하고 각 처리는 r개 실험구에서 나타난다. 어떤 처리도 블럭에서 한번 이상 나타나지 않는다. 처리의 각 쌍은 동일크기의 블럭에서 λ번 나타난다. 이러한 조건이 만족되는 실험설계를 BIB설계 또는 BIBD로 나타낸다. BIB설계에 요구되는 다섯 모수, b, t, r, k와 λ는

$$\begin{aligned}&(1)\ n = rt = bk\\&(2)\ \lambda(t-1) = r(k-1)\\&(3)\ b \geq t\end{aligned} \tag{1.1}$$

인 제한조건들이 만족되어야 하는 독립이 아닌 정수들로 주어진다. 단, n은 실험구의 전체수이다. $b = t$이고 $r = k$인 설계를 대칭적 (symmetric)이라 한다. 균형불완비블럭설계에 필수적인 다섯 모수들의 종속성으로 인해 BIB설계의 존재성과 구성에 관한 문제들과 함께 자료분석의 방법에 관한 연구가 많은 문헌들에서 논의되어 왔다.

본 논문은 균형불완비블럭설계를 통하여 수집된 자료의 블럭내 분석 (intrablock analysis)에서 구해진 분석결과와 동일한 결과를 사영의 관점에서 분석하는 방법을 제시하고 고전적 분석방법과의 비교를 통하여 사영이 변동량의 계산 및 추론에 있어서 효과적임을 입증해 보이고자 한다. 사영에 의한 자료분석 방법은 Choi (2011, 2012, 2014)에서 다양한 선형모형의 가정하에 논의되고 있음을 살펴볼 수 있다.

2. 블럭내 분석모형

블럭내 분석은 처리효과에서의 모든 대비 (contrast)의 추정값들이 동일블럭에서 실험구 간의 비교들로 표현되므로 블럭효과가 제거되어 분석이 이루어 진다. 블럭내 분석모형은

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{2.1}$$

이다. y_{ij}는 j번째 블럭에서 i번째 처리에 대한 관측값이 존재하는 경우의 값이다. μ는 전체 평균을 나타내고 τ_i는 i번째 처리효과이다. β_j는 j번째 블럭효과를 나타내며 $N(0, \sigma_b^2)$인 분포를 따른다고 가정한다. ϵ_{ij}는 $N(0, \sigma^2)$인 분포를 따르는 확률오차 성분이다. 블럭효과와 오차성분은 서로 독립이라고 가정한다. 단, $i = 1, 2, \cdots, t$ 이고 $j = 1, 2, \cdots, b$이다. 균형불완비블럭설계에 해당하는 블럭내 모형은 BIB설계의 특성들을 내포하고 있으나 모형만으로는 파악할 수 없다. BIB설계의 특성을 나타내는 행렬을 발생행렬 (incidence matrix)이라 한다. 발생행렬을 N으로 나타내면 N은 t개의 처리를 나타내는 행과 b개의 블럭열에서 i번째 처리가 j번째 블럭에서 발생하는 회수 n_{ij}의 행렬이다. BIB설계에서 n_{ij}는 0 또는 1이므로 $n_{ij}^2 = n_{ij}$이다. 행렬 NN'은 t개 행과 t개 열을 갖는다. NN'의 원소들을 ν_{ik}로 나타내면 $\nu_{ii} = \sum_{j=1}^t n_{ij}^2$이고 $\nu_{ik} = \sum_{j=1}^t n_{ij}n_{kj}$ $(i \neq k)$이다. BIBD의 경우에 $\nu_{ii} = r$이고 $\nu_{ik} = \lambda$ $(i \neq k)$이다. 따라서

$$\boldsymbol{NN'} = (r-\lambda)\boldsymbol{I}_t + \lambda\boldsymbol{J}_t \tag{2.2}$$

이다. 여기서 $\boldsymbol{I}_t$는 항등행렬이고 $\boldsymbol{J}_t$는 모든 원소가 1인 $t \times t$인 정방행렬이다. $\boldsymbol{NN'}$의 행렬식

$$|\boldsymbol{NN'}| = (r-\lambda)^{t-1}(r+(t-1)\lambda) = (r-\lambda)^{t-1}rk > 0 \tag{2.3}$$

을 이용하여 BIBD의 존재성을 확인할 수 있다. 따라서 가능한 BIB설계를 이용하여 수집된 자료의 분석모형으로 식 (2.1)이 가정되었을 때 행렬표현식은

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X\gamma}\\&= \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{X}_2\boldsymbol{\beta} + \boldsymbol{\epsilon}\end{aligned} \tag{2.4}$$

$\boldsymbol{y}$는 크기가 $n \times 1$인 관측벡터이다. 여기서 n=rt=bk이다. $\boldsymbol{X}$는 크기가 $n \times (1+t+b)$인 모형행렬이다. $\boldsymbol{X}$는 $\boldsymbol{X} = (\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)$로 구성된 분할행렬이고 $\boldsymbol{\gamma}'$은 $(\mu, \boldsymbol{\tau}', \boldsymbol{\beta}')$으로 구성된 벡터이다. $\boldsymbol{j}$는 모든 원소가 1인 $n \times 1$인 열벡터이다. μ는 전체 평균을 나타내는 모수이다. $\boldsymbol{X}_1$은 크기가 $n \times t$이고 모수벡터 $\boldsymbol{\tau}$의 계수행렬이며 $\boldsymbol{\tau}$는 t개 처리효과들로 구성된 열벡터이다.$\boldsymbol{X}_2$는 크기가 $n \times b$이고 모수벡터 $\boldsymbol{\beta}$의 계수행렬이다. $\boldsymbol{\beta}$는 b개 블럭효과를 나타내는 열벡터이다. 블럭효과와 오차성분은 서로 독립으로 가정하고 있기 때문에 이들 간의 교호작용은 없다.

식 (2.4)를 이용하여 사영에 의한 블럭내 분석을 논의해 보기로 한다. 다차원상의 관측벡터 $\boldsymbol{y}$를 모형행렬 $\boldsymbol{X}$로의 사영 (projection)을 $\boldsymbol{y}_p$라 둘 때, $\boldsymbol{y}_p = \boldsymbol{X}\boldsymbol{X}^-\boldsymbol{y}$로 구해진다. 여기서 $\boldsymbol{X}\boldsymbol{X}^-$는 $\boldsymbol{X}$로의 사영행렬이고 $\boldsymbol{X}^-$는 Moore-Penrose의 일반화역행렬 (generalized inverse)을 나타낸다. 모형행렬 $\boldsymbol{X}$가 세 부분행렬 $\boldsymbol{j}$, $\boldsymbol{X}_1$과 $\boldsymbol{X}_2$로 구성되는 분할행렬이므로 모형행렬 $\boldsymbol{X}$로의 사영은 $\boldsymbol{j}$로의 사영성분과 $\boldsymbol{j}$로의 사영과 직교하는 $\boldsymbol{X}_1$으로의 사영성분 그리고 $\boldsymbol{j}$와 $\boldsymbol{X}_1$으로의 사영성분들과 상호 직교하는 $\boldsymbol{X}_2$로의 사영성분의 합으로 구해질 수 있다. 이는 다차원의 벡터공간에서 사영의 관점에서 분석이 행해질 때 자료전체의 고정된 변동량을 다양하게 분석할 수 있는 방법을 제공하여 주는 분석측도로 이용됨을 의미한다.

BIBD의 블럭내 분석을 위해 모형의 단계별 적합방식 (stepwise procedure)을 이용하기로 한다. 먼저 $\boldsymbol{X}$의 부분행렬인 $\boldsymbol{j}$로의 사영을 생각해 보자. $\boldsymbol{j}$로의 사영을 위한 모형식은 식 (2.4)로부터

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{\epsilon}_1 \tag{2.5}$$

인 모형의 적합을 가정한다. 식 (2.5)의 적합으로부터 구해진 $\boldsymbol{j}$로의 사영을 $\boldsymbol{j}_p$라 둘 때

$$\boldsymbol{j}_p = \boldsymbol{j}\hat{\mu} = \boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y} \tag{2.6}$$

로 구해진다. μ에 따른 변동량은 $\boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}$로 구해진다. $\boldsymbol{y}$에서 $\boldsymbol{j}$로의 사영을 제외한 잔차벡터를 $\boldsymbol{r}$이라 두자. $\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{y}$로 구해진다. $\boldsymbol{j}$로의 사영 $\boldsymbol{j}\hat{\mu}$에 직교하는 블럭효과벡터와 관련된 사영성분을 얻기 위한 모형으로

$$\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2\boldsymbol{\beta} + \boldsymbol{\epsilon}_2 \tag{2.7}$$

를 가정한다. $\boldsymbol{r}$을 식 (2.7)의 모형행렬 $(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2$로의 사영을 이용하여 $\boldsymbol{j}\hat{\mu}$과 직교인 사영성분을 구한다. 이는 블럭효과에 영향을 받지 않는 처리효과를 구하기 위한 모형의 적합방식으로 $\boldsymbol{y} - \boldsymbol{j}\hat{\mu}$을 이용한 $(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2$로의 사영을 의미하고 있다. 식 (2.7)의 적합으로부터 $(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2$로의 사영을 $\boldsymbol{X}_{2p}$라 두면

$$\begin{aligned}\boldsymbol{X}_{2p} &= [(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2]\hat{\boldsymbol{\beta}} = [(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2][(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2]^-\boldsymbol{r} \\ &= \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}\end{aligned} \tag{2.8}$$

로 구해진다. 단, $\boldsymbol{X}_{22} = (\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_2$를 나타낸다. $\boldsymbol{\beta}$에 따른 변동량은 $\boldsymbol{r}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}$로 구해진다. 이 양은 블럭내 분석에서 처리효과에 적합되지 않은 블럭변동량을 나타낸다. $\boldsymbol{r}$에서 식 (2.8)의 $\boldsymbol{X}_{2p}$를 제외한 잔차벡터를 $\boldsymbol{r}_2$라 두면 $\boldsymbol{r}_2 = (\boldsymbol{I} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-)\boldsymbol{r}$로 구해진다. 블럭효과와 무관한 처리효과를 구하기 위한 모형으로

$$\boldsymbol{r}_2 = (\boldsymbol{I} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-)(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{\epsilon}_3 \tag{2.9}$$

을 이용하게 된다. $(\boldsymbol{I} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-)(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_1$을 $\boldsymbol{X}_{11}$이라 두면 가정된 모형은 다차원상의 관측벡터 $\boldsymbol{r}_2$를 모형행렬 $\boldsymbol{X}$의 부분행렬들인 $\boldsymbol{j}$와 $\boldsymbol{X}_2$에 의해 생성된 벡터 부분공간으로의 사영을 제외한 잔차벡터를 $\boldsymbol{X}_1$으로의 사영을 구하기 위한 모형임을 보여준다. 식 (2.9)의 적합으로부터 구해지는 사영을

$\boldsymbol{X}_{1p}$라 두면

$$\begin{aligned}\boldsymbol{X}_{1p} &= [(\boldsymbol{I}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_1][(\boldsymbol{I}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_1]^{-}\boldsymbol{r}_2 \\ &= \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-}\boldsymbol{r}_2 \end{aligned} \tag{2.10}$$

로 구해진다. 단, $\boldsymbol{X}_{11} = (\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{X}_1$이다. $\boldsymbol{\tau}$에 따른 변동량은 $\boldsymbol{r}_2'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-}\boldsymbol{r}_2$로 구해진다. 이 변동량은 블럭효과에 적합된 처리효과의 변동량을 나타낸다. 식 (2.9)의 적합으로부터 주어지는 잔차벡터를 $\boldsymbol{r}_1$이라 두면 $\boldsymbol{r}_1 = (\boldsymbol{I}-\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-})\boldsymbol{r}_2$로 구해진다. 식 (2.10)에서 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-}-\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-})$을 $\boldsymbol{X}_r$로 나타낼 때 $\boldsymbol{\epsilon}$에 따른 변동량은 $\boldsymbol{r}_1'\boldsymbol{X}_r\boldsymbol{X}_r{}^{-}\boldsymbol{r}_1$으로 구해진다. 자료의 전체변동량 $\boldsymbol{y}'\boldsymbol{y}$를 변동요인에 따른 성분변동량으로 나타내면

$$\boldsymbol{y}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y} + \boldsymbol{r}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-}\boldsymbol{r} + \boldsymbol{r}_2'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-}\boldsymbol{r}_2 + \boldsymbol{r}_1'\boldsymbol{X}_r\boldsymbol{X}_r{}^{-}\boldsymbol{r}_1 \tag{2.11}$$

로 주어지며 이는 BIBD에서의 블럭내 분석과 일치하고 있음을 보여준다.

3. BIBD에서의 추정가능함수

블럭효과에 무관한 처리효과를 추정하기 위한 방법으로 식 (2.8)을 이용한다. 식 (2.8)에서 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{y}$를 $\boldsymbol{X}_1$으로의 사영 $\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^{-}\boldsymbol{y}$는 사영공간내에서 최소제곱거리를 나타내는 점이므로 최소제곱법으로 추정벡터를 구하는 것과 동일하게 된다. 사영에 의한 $\boldsymbol{\tau}$의 추정벡터 $\hat{\boldsymbol{\tau}}$는

$$\hat{\boldsymbol{\tau}} = [(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{X}_1]^{-}\boldsymbol{r}_2 \tag{3.1}$$

로 구해진다. $\boldsymbol{\tau}$의 한 선형함수인 $\boldsymbol{l}'\boldsymbol{\tau}$가 추정가능함수인 가는 추정가능함수의 정의에 따른 $l'\boldsymbol{\tau}$의 추정값이 어떤 최소제곱해의 선택에도 불변 (invariant)인 가를 확인함으로서 알 수 있다. 또는 $\boldsymbol{l}'$이 추정가능함수의 한 성질인

$$\boldsymbol{l}' = \boldsymbol{l}'[(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{X}_1]^{-}[(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{X}_1] \tag{3.2}$$

를 만족시키고 있는 가를 확인함으로 알 수 있다. 단, $\boldsymbol{l}$은 t개의 성분을 갖는 실수값 벡터이며 $\boldsymbol{l} = (l_1, l_2, \cdots, l_t)'$이다. 비교하고자 하는 모든 처리들이 수용되지 못하는 불완비블럭의 실험설계에서는 일부처리들의 비교는 불가능하게 된다. 따라서 처리들의 비교를 위해 어떤 함수들이 추정가능한 가를 파악하는 것은 추론을 위해 필요한 과정이다. BIB설계에서 추론가능함수는 임의 개수의 처리에 대한 대비 (contrast) 또는 대비들의 선형함수는 추론가능함을 보여준다. t개 처리들의 한 비교를 나타내는 추정함수를 고정효과들의 선형함수로 나타내면

$$\boldsymbol{l}'\boldsymbol{\tau} = l_1\tau_1 + l_2\tau_2 + \cdots + l_t\tau_t \tag{3.3}$$

이다. t개 처리들의 한 비교에서 $\sum_{i=1}^{t} l_i = 0$일 때 비교는 대비로 정의된다. $\boldsymbol{l}'\boldsymbol{\tau}$가 추정함수인 가의 판단은 $\boldsymbol{\tau}$를 나타내는 벡터공간내의 한 모수점인 가를 확인함으로서 알 수 있다. 즉, $\boldsymbol{l}$이 추정가능함수이면 모수공간으로의 사영에서

$$\boldsymbol{l}' = \boldsymbol{l}'\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11} \tag{3.4}$$

으로 주어져야 한다. BIBD에서 추정가능함수를 구성하는 문제를 생각해 보자. 블럭효과가 제외된 처리효과들의 비교를 위한 추정가능함수는 $\boldsymbol{\tau}$의 함수이며 최소제곱해의 선택에 불변인 추정치를 갖는 모수이다. 따라서, 그러한 추정치를 갖는 한 방법으로 $\boldsymbol{X}_{11}'\boldsymbol{X}_{11}$의 고유벡터를 이용할 수 있다. $\boldsymbol{X}_{11}'\boldsymbol{X}_{11}$는

계수(rank) k ($< t$)인 비음정치(nonnegative definite)행렬이므로 $\boldsymbol{X}'_{11}\boldsymbol{X}_{11}$의 고유벡터들을 $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\cdots$, $\boldsymbol{p}_k$라 두면 추정가능함수 $\boldsymbol{l}' = (l_1, l_2, \cdots, l_t)$는

$$\boldsymbol{l} = r_1\boldsymbol{p}_1 + r_2\boldsymbol{p}_2 + \cdots + r_k\boldsymbol{p}_k \tag{3.5}$$

로 구해진다. 단, r_i ($i = 1, 2, \cdots, k$)는 실수이다.

4. 블럭내 분석자료의 예

다음은 블럭내 분석을 위한 Montgomery (1976)의 자료를 나타낸다. 화학공정의 반응시간에 영향을 주는 4종류의 촉매에 대한 실험으로 BIB설계를 이용한 자료이다. 처리의 수 $t = 4$이고 블럭의 수도 $b = 4$이다. 블럭의 크기 $k = 3$이고 각 처리는 블럭에서 $r = 3$회 발생한다. 따라서 $n = tr = bk = 12$이다.

Table 4.1 Balanced incomplete block data

Treatment (Catalyst)	Block (Batch of raw material)			
	1	2	3	4
1	73	74	–	71
2	–	75	67	72
3	73	75	68	–
4	75	–	72	75

Table 4.1의 불완비 블럭설계에서 $\lambda = \frac{r(k-1)}{(t-1)} = 2$로 구해진다. 이는 임의 쌍의 처리가 2번 동일하게 발생하는 불완비 블럭설계이므로 균형불완비블럭설계임을 나타내고 있다. 균형불완비블럭설계의 존재성은 식 (2.3)에서 계산된 $|\boldsymbol{NN}'|$의 값 9로 확인된다. 블럭내 분석을 위한 모형으로 식 (2.4)를 가정한다. 모형의 단계별 적합방식에 따른 식 (2.5)의 적합으로 부터 $\boldsymbol{j}_p = \boldsymbol{jj}^-\boldsymbol{y} = 72.5\boldsymbol{j}$로 구해진다. $\boldsymbol{y}$를 $\boldsymbol{j}$로의 사영에 의한 변동량은 $\boldsymbol{y}'\boldsymbol{jj}^-\boldsymbol{y} = 63075$이다. $\boldsymbol{y}'\boldsymbol{y} = 63156$이므로 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{jj}^-)\boldsymbol{y} = 81$로 구해진다. 단계별 적합의 두번 째 단계에서 식 (2.7)의 모형을 가정한다. $(\boldsymbol{I} - \boldsymbol{jj}^-)\boldsymbol{y}$로 주어지는 잔차벡터 $\boldsymbol{r}$을 계산하면 $\boldsymbol{r} = (0.5, 0.5, 2.5, 1.5, 2.5, 2.5, -5.5, -4.5, -0.5, -1.5, -0.5, 2.5)'$로 구해진다. $\boldsymbol{r}$을 $\boldsymbol{X}_{22}$로 표시되는 $(\boldsymbol{I} - \boldsymbol{jj}^-)\boldsymbol{X}_2$로의 사영은 식 (2.8)의 $\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}$로 계산되며 그 값은 $(1.167, 1.167, 1.167, 2.167, 2.167, 2.167, -3.500, -3.500, -3.500, 0.167, 0.167, 0.167)'$이다. 사영까지의 거리제곱합 $\boldsymbol{r}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r} = 55$로 계산된다. 블럭효과를 제외한 처리효과를 구하기 위한 단계별 적합모형은 식 (2.9)이다. 식 (2.9)의 $\boldsymbol{r}_2$를 구해보면 $\boldsymbol{r}_2 = (-0.667, -0.667, 1.333, -0.667, 0.333, 0.333, -2.000, -1.000, 3.000, -1.667, -0.667, 2.333)'$ 으로 구해진다. 모형식 (2.9)의 적합으로부터 구해지는 사영 $\boldsymbol{X}_{1p}$는 $\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{r}_2$로 계산되고 그 값은 $(-1.417, -0.792, 2.208, -0.292, -0.042, 0.333, -1.250, -0.875, 2.125, -1.292, -1.042, 2.333)'$이다. $\boldsymbol{r}'_2\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_2$로 계산되는 사영까지의 거리제곱합은 22.75로 구해진다. 처리효과벡터 $\boldsymbol{\tau}$의 추정벡터를 $\hat{\boldsymbol{\tau}}$라 두면 $\hat{\boldsymbol{\tau}} = (-1.125, -0.875, -0.500, 2.500)$으로 구해진다.

오차성분에 대한 잔차제곱합은 식 (2.9)의 적합으로부터 구해지는 잔차벡터 $\boldsymbol{r}_1$의 내적으로 구해진다. 이 값은 $\boldsymbol{r}'_1\boldsymbol{r}_1 = 3.25$로 계산된다. 잔차제곱합은 또한 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{jj}^-)\boldsymbol{y}$에서 블럭효과에 따른 제곱합 $\boldsymbol{r}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}$과 처리효과에 따른 제곱합 $\boldsymbol{r}'_2\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_2$를 뺀 양으로도 구해진다.

처리효과들의 선형함수로 추정가능함수를 생각해 보자. 블럭효과에 적합된 처리효과벡터의 계수행렬 $\boldsymbol{X}_{11}$으로의 사영을 나타내는 사영행렬은 $\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-$이므로 처리효과벡터 $\boldsymbol{\tau}$의 공간을 나타내는

사영행렬은 $\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11}$이다. Table 4.1의 자료분석을 위한 모형식 (2.4)의 적합으로 부터 구해지는 $\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11}$을 구해보면

$$\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11} = \begin{pmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{pmatrix} \tag{4.1}$$

이다. 식 (3.3)의 $\boldsymbol{l}'\boldsymbol{\tau}$에서 $\boldsymbol{l}'$을 $(1,-1,-1,1)$로 두자. 이때 $\boldsymbol{l}'\boldsymbol{\tau}$를 $\boldsymbol{l}_1'\boldsymbol{\tau}$이라 두면 $\tau_1 - \tau_2 - \tau_3 + \tau_4$이다. $\boldsymbol{l}_1'\boldsymbol{\tau}$가 추정가능함수이기 위해서는 식 (3.4)가 만족되어야 한다. 행렬 (4.11)을 이용하여 $(1,-1,-1,1)\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11}$을 계산하면 $(1,-1,-1,1)$로 구해진다. 따라서 $\boldsymbol{l}_1'\boldsymbol{\tau}$은 추정가능함수임을 알 수 있다. 한편 τ_1이 추정가능함수인 가의 확인은 $\boldsymbol{l}'$을 $\boldsymbol{l}_2' = (1,0,0,0)$로 취하여 $\boldsymbol{l}_2'\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11}$을 계산한다. 계산된 벡터는 $(0.75,-0.25,-0.25,-0.25)$로 주어지므로 τ_1이 추정가능하지 않음을 알 수 있다. 추정가능한 함수들은 $\boldsymbol{X}_{11}'\boldsymbol{X}_{11}$의 고유벡터를 이용하여 구성할 수 있다. $\boldsymbol{X}_{11}'\boldsymbol{X}_{11}$의 음아닌 양수의 세 고유근에 해당하는 고유벡터들은

$$\boldsymbol{p}_1 = \begin{pmatrix} 0.000 \\ -0.809 \\ 0.497 \\ 0.313 \end{pmatrix}, \ \boldsymbol{p}_2 = \begin{pmatrix} 0.866 \\ -0.289 \\ -0.289 \\ -0.289 \end{pmatrix}, \ \boldsymbol{p}_3 = \begin{pmatrix} 0.000 \\ -0.106 \\ -0.648 \\ 0.754 \end{pmatrix} \tag{4.2}$$

로 주어진다. 세 고유벡터들의 한 선형결합으로 $\boldsymbol{r}' = (1,0,-1)$을 이용한 벡터를 $\boldsymbol{l}_r$이라 두자. $\boldsymbol{l}_r$은

$$\begin{aligned} \boldsymbol{l}_r &= \boldsymbol{p}_1 - \boldsymbol{p}_3 \\ &= \begin{pmatrix} 0.000 \\ -0.809 \\ 0.497 \\ 0.313 \end{pmatrix} - \begin{pmatrix} 0.000 \\ -0.106 \\ -0.648 \\ 0.754 \end{pmatrix} = \begin{pmatrix} 0.000 \\ -0.703 \\ 1.145 \\ -0.442 \end{pmatrix} \end{aligned} \tag{4.3}$$

로 주어진다. $\boldsymbol{l}_r'\boldsymbol{\tau}$의 추정가능성은 식 (3.4)의 $\boldsymbol{l}_r'\boldsymbol{X}_{11}{}^{-}\boldsymbol{X}_{11}$에서 동일한 계수벡터로 주어지므로 추정가능한 함수임을 보여준다. Table 4.1의 자료에 대한 사영분석은 Montgomery의 자료분석과 동일한 결과를 보여주고 있으며 사영이 추정가능함수의 확인과 검토에도 유용하게 이용될 수 있음을 보여주고 있다.

5. 결론

본 논문은 균형불완비블럭설계 (balanced incomplete block design)의 블럭내 분석 (intrablock analysis)으로 사영을 이용하는 방법을 논의하고 있다. 균형불완비블럭의 설계로부터 블럭효과에 영향을 받지 않는 처리효과 간의 비교를 위해 벡터공간을 분할하는 방법을 제시하고 있다. 벡터공간을 분할하기 위한 방법으로 모형의 단계별 적합방법 (stepwise procedure)를 이용하고 있다. 즉, 블럭효과가 제외된 처리효과만의 비교를 위한 사영공간은 단계별 모형의 가정으로부터 주어지는 모형행렬로 생성되는 공간이다. 처리효과만의 비교를 위한 사영공간에서의 사영과 처리효과에 따른 변동량의 계산과정을 구체적으로 다루고 있다. 또한, BIB설계의 가정된 모형으로부터 모형내 모수효과를 추정하기 위한 잔차모형의 설정에서 잔차에 대한 구성을 자세히 다루고 있다.

BIB설계에서 추정가능함수가 어떻게 확인되고 구성되어야 하는 문제도 다루어지고 있다. 처리효과 또는 처리효과들의 어떤 선형함수가 추정가능함수이기 위해서는 모수공간으로의 사영을 이용하여 확인

할 수 있음을 다루고 있다. 추정가능함수를 구성하기 위한 방법으로 단계별 모형의 가정에서 주어지는 처리효과벡터의 계수행렬의 고유벡터를 이용하는 방법을 제시하고 있다. 변동요인에 따른 모형의 단계별 설정과정과 모형행렬로의 사영을 구하는 과정을 균형불완비블럭설계의 Montgomery (1976) 자료 예를 통하여 구체적으로 보여주고 있다. 또한, 사영이 추정가능함수의 구성과 판단에도 효과적으로 이용될 수 있다는 점에서 고전적 분석방법과는 뚜렷한 차이가 있음을 보여준다.

References

Choi, J. S. (2011). Type I analysis by projections. *The Korean Journal of Applied Statistics*, **24**, 373−381.

Choi, J. S. (2012). Type II analysis by projections. *Journal of the Korean Data & Information Science Society*, **23**, 1155−1163.

Choi, J. S. (2014). Projection analysis for two-way variance components. *Journal of the Korean Data & Information Science Society*, **23**, 547−554.

Cochran, W. G. and Cox, G. M. (1957). *Experimental designs*, John Wiley and Sons, Inc., New York.

Davies, O. L. (1956). *Design and analysis of industrial esperiments*, Second edition, Hafner Publishing Company, New York.

Graybill, F. A. (1976). *Theory and application of the linear model*, Wadsworth, Inc., California.

Hicks, C. R. (1973). *Fundamental concepts in the design of experiments*, Holt, Rinehart and Winston, Inc., New York.

John, P.W. M. (1971). *Statistical design and analysis of experiments*, The Macmillan Company, New York.

Milliken, G. A. and Johnson, D. E. (1984). *Analysis of messy data*, Van Nostrand Reinhold, New York.

Montgomery, D. C. (1976). *Design and analysis of experiments*, John Wiley and Sons, Inc., New York.

Searle, S. R., Casella, G. and McCulloch, C. E. (1971). *Linear models*, John Wiley and Sons, Inc., New York.

Yates, F. (1936). Incomplete randomized blocks. *Annals of Eugenics*, **7**, 121−140.

Journal of the Korean Data & Information Science Society
2015, **26**(2), 347–354

http://dx.doi.org/10.7465/jkdi.2015.26.2.347
한국데이터정보과학회지

Projection analysis for balanced incomplete block designs

Jaesung Choi[1]

[1]Department of Statistics, Keimyung University

Received 9 January 2015, revised 26 January 2015, accepted 9 February 2015

Abstract

This paper deals with a method for intrablock anlaysis of balanced incomplete block designs on the basis of projections under the assumption of mixed effects model. It shows how to construct a model at each step by the stepwise procedure and discusses how to use projection for the analysis of intrablock. Projections are obtained in vector subspaces orthogonal to each other. So the estimates of the treatment effects are not affected by the block effects. The estimability of a parameter or a function of parameters is discussed and eigenvectors are dealt for the construction of estimable functions.

Keywords: Eigenvectors, estimability, intrablock analysis, projection, stepwise procedure.

[1] Professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr