The Korean Journal of Applied Statistics (2015) DOI: http://dx.doi.org/10.5351/KJAS.2015.28.2.151
28(2), 151–158

Interblock Information from BIBD Mixed Effects

Jaesung Choi[a,1]

[a]Department of Statistics, Keimyung University

(Received February 10, 2015; Revised March 12, 2015; Accepted March 12, 2015)

Abstract

This paper discusses how to use projections for the analysis of data from balanced incomplete block designs. A model is suggested as a matrix form for the interblock analysis. A second set of treatment effects can be found by projections from the suggested interblock model. The variance and covariance matrix of two estimated vectors of treatment effects is derived. The uncorrelation of two estimated vectors can be verified from their covaraince structure. The fitting constants method is employed for the calculation of block sum of squares adjusted for treatment effects.

Keywords: Balanced, incomplete block, intrablock, interblock, projection.

1. 서론

실험에서 비교하고자하는 처리들의 비교에 대한 실험의 정도를 높이기 위해 실험단위들의 블록화가 행해진다. 실험에서의 블록설계는 실험에 이용되는 실험단위들이 동질적이 아닐 때 동질성을 갖는 실험단위들의 집단으로 분류하여 처리비교의 정도를 높이기 위한 처리들의 배치방법이다. 동질적 실험단위들의 집단 또는 군을 블록으로 정의한다. 블록설계에서 블록효과는 고정효과로 간주되기도 하고 확률효과로 간주되기도 한다. 블록의 크기가 비교하고자 하는 처리들을 모두 수용할 수 없는 크기의 블록으로 실험이 설계될 때 불완비블록설계라 한다. 불완비블록설계로 Youden 방격, 분할구 설계(split-plot designs)나 반복측정설계(repeated measures designs) 등을 생각할 수 있다. 불완비블록설계의 구성과 분석에 관한 논의는 Yates (1936), Davies (1956), Cochran과 Cox (1957) 그리고 John (1971) 등에서 살펴볼 수 있다.

균형불완비블록설계(balanced incomplete block design)는 임의의 두 처리가 동일 회수로 함께 나타나는 불완비블록설계이다. 균형불완비설계의 자료분석은 블록내 분석과 블록간 분석으로 구분된다. 블록내 분석은 블록효과가 상쇄된 처리효과에서의 모든 대비가 동일 블록내 관측값들의 비교로 표현될 수 있기 때문에 블록효과가 고정효과이든 확률효과이든 간에 영향을 받지 않는다. 균형불완비블록설계의 분석모형과 방법들에 관한 논의가 Hicks (1973)와 Montgomery (1976) 등에서 보여진다.

Yates (1940)는 균형불완비블록설계에서 블록효과들이 평균이 0이고 분산이 σ_β^2인 상관성이 없는 확률변수들이면 처리효과 $\tau_i \, (i = 1, 2, \ldots, t)$에 대한 추가정보를 획득할 수 있음에 주목하여 추가정보를 구하는 방법을 블록간 분석(interblock analysis)으로 표현했다. John은 블록간 분석모형으로 블록

[1]Department of Statistics, Keimyung University, 1095 Dalgubul-Daero, Dalseogu-Gu, Daegu 704-701, Korea.
E-mail: jschoi@kmu.ac.kr

합에 대한 선형모형에 근거하여 블록간 정보복구(recovery of interblock information)로 블록간 추정치(interblock estimate)를 구하는 과정을 다루고 있다.

본 논문은 Yates의 블록간 분석을 사영으로 다루어 보고자 한다. Yates는 블록효과가 확률효과로 간주될 때 처리효과에 대한 추가적 정보를 구하는 방법을 다루었다. 즉, 블록내 분석을 통한 처리효과들의 추정값에 대한 정보와는 별도로 블록간 분석을 통해 처리효과들의 추정값에 대한 정보를 추가적으로 얻을 수 있음을 의미하고 있다. 따라서 두개의 분석을 통하여 구해진 처리효과의 블록내 추정값과 블록간 추정값의 결합추정값을 구하는 문제와 결합추정량의 분산을 다루게 된다. 본 논문에서는 블록간 분석을 사영의 관점에서 행할 때 분석모형을 제시하고 John의 자료를 이용하여 동일한 결과를 갖는 사영분석방법을 제공하고자 한다. 블록간 분석에서 블록요인에 따른 변동량의 계산에 상수적합법(fitting constants method)을 적용할 수 있음을 다루고자 한다. 상수적합법은 Milliken과 Johnson (1984) 그리고 Searle 등 (1992) 등에서 논의된다. 사영에 의한 자료분석 방법은 Choi (2011, 2012, 2014)에서 다양한 선형모형의 가정하에 논의되고 있음을 살펴볼 수 있다.

2. 블록내 분석모형

Yates (1940)는 균형불완비블록설계의 블록효과와 관련된 어떤 조건하에서 블록간 추정값들로 호칭되는 처리효과의 다른방식으로 추정값들을 구할 수 있음을 다루었다. 균형불완비블록설계에서 Yates의 조건이 만족되는 블록간 분석을 위한 모형을 생각해 보기로 한다. 모형의 기술을 위해 균형불완비블록설계의 조건을 만족시키는 처리의 수를 t라 두자. 블록의 크기는 $k\,(k<t)$이며 각 처리는 b개 블록에서 r번 반복된다고 가정한다. 전체자료의 수를 n이라 둘 때 $n=tr=bk$이다. 각 처리 쌍이 동일블록에서 나타나는 횟수를 λ라 두면 $\lambda=r(k-1)/(t-1)$이다. 블록간 분석모형은 블록내 분석모형에서 추가적인 조건들이 만족되어야 하므로 블록내 분석모형을 살펴보기로 한다.

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

이다. y_{ij}는 j번째 블록에서 i번째 처리에 대한 관측값이 존재하는 경우의 값이다. μ는 전체 평균을 나타내고 τ_i는 i번째 처리효과이다. β_j는 j번째 블록효과를 나타내며 ϵ_{ij}는 $N(0,\sigma_\epsilon^2)$인 분포를 따르는 확률오차 성분이다. 블록효과와 오차성분은 상관성이 없다고 가정한다. 단, $i=1,2,\ldots,t$이고 $j=1,2,\ldots,b$이다. μ와 τ_i는 미지의 상수들이고 블록효과는 고정효과로 취급되기도 하고 확률효과로 간주되기도 한다. 식 (2.1)의 행렬표현식은

$$\boldsymbol{y}=\boldsymbol{j}\mu+\boldsymbol{X}_1\boldsymbol{\tau}+\boldsymbol{X}_2\boldsymbol{\beta}+\boldsymbol{\epsilon} \tag{2.2}$$

이다. $\boldsymbol{y}$는 크기가 $n\times 1$인 관측벡터이고 $n=rt=bk$이다. $\boldsymbol{j}$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 블록효과를 나타내는 열벡터이다. 블록효과와 오차성분은 상관성이 없음을 가정하고 있기 때문에 이들 간의 교호작용은 없다. 오차벡터 $\boldsymbol{\epsilon}$는 $N(\boldsymbol{0},\sigma_\epsilon^2\boldsymbol{I})$인 분포를 따른다고 가정한다.

블록내 분석모형 (2.2)를 이용하여 블록내 처리효과에 대한 추정을 사영으로 구해보기로 한다. 블록내 처리효과의 추정은 블록효과와 무관한 처리효과가 추정되어야 하므로 블록효과의 추정을 위한 단계별 모형은

$$\left(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^-\right)\boldsymbol{y}=\left(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^-\right)\boldsymbol{X}_2\boldsymbol{\beta}+\boldsymbol{\epsilon}_1 \tag{2.3}$$

이다. 단, $\boldsymbol{j}^{-}$는 Moore-Penrose의 일반화 역행렬(generalized inverse)을 나타낸다. 이 모형은 모평균 μ의 계수벡터 $\boldsymbol{j}$로의 사영이 감해진 잔차벡터를 이용하고 있다. 모형식 (2.3)에서 오차벡터 $\boldsymbol{\epsilon}_1$은 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})(\boldsymbol{X}_1\boldsymbol{\tau}+\boldsymbol{\epsilon})$으로 정의된다. 잔차벡터 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y}$를 $\boldsymbol{r}_1$이라 두자. 식 (2.3)을 이용한 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_2$로의 사영은 잔차벡터의 추정벡터를 나타낸다. $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_2$를 $\boldsymbol{X}_{22}$라 두고 블록효과의 추정벡터를 $\hat{\boldsymbol{\beta}}$이라 둘 때 잔차벡터의 추정벡터는 $\boldsymbol{X}_{22}\hat{\boldsymbol{\beta}}$이다. $\boldsymbol{r}_1$에서 $\boldsymbol{X}_{22}\hat{\boldsymbol{\beta}}$을 뺀 잔차벡터를 $\boldsymbol{r}_2$라 두면 $\boldsymbol{r}_2$를 이용한 처리효과의 모형은

$$\boldsymbol{r}_2 = \left(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-}\right)\boldsymbol{X}_1\boldsymbol{\tau}+\boldsymbol{\epsilon}_2 \tag{2.4}$$

이다. 식 (2.4)에서 $\boldsymbol{\tau}$의 계수행렬을 $\boldsymbol{X}_{11}$이라 두면 $\boldsymbol{X}_{11}=(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{X}_1$이다. 오차벡터 $\boldsymbol{\epsilon}_2$는 $(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})\boldsymbol{\epsilon}$으로 정의된다. $\boldsymbol{\tau}$의 추정은 $\boldsymbol{r}_2$를 $\boldsymbol{X}_{11}$로의 사영으로부터 구해진다. $\boldsymbol{\tau}$의 추정벡터를 $\hat{\boldsymbol{\tau}}$라 둘 때

$$\hat{\boldsymbol{\tau}} = \left[\left(\boldsymbol{I}-\boldsymbol{j}\boldsymbol{j}^{-}-\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-}\right)\boldsymbol{X}_1\right]^{-}\boldsymbol{r}_2 \tag{2.5}$$

로 구해진다. 식 (2.2)의 변동요인별 제곱합의 계산을 위한 방법으로 상수적합법이 가능하다. Henderson (1953)의 방법 III으로도 불리우는 상수적합법은 자료분석으로 가정된 모형이 고정효과와 확률효과를 포함하는 혼합모형으로 간주될 때 요인별 변동량을 구하기 위한 방법이다. $\boldsymbol{X}_f$를 $\boldsymbol{X}_f=(\boldsymbol{j},\boldsymbol{X}_1)$이라 둘 때 확률효과벡터 $\boldsymbol{\beta}$에 따른 제곱합의 계산을 위해 적합되는 모형은

$$\left(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{X}_f{}^{-}\right)\boldsymbol{y} = \left(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{X}_f{}^{-}\right)\boldsymbol{X}_2\boldsymbol{\beta}+\boldsymbol{\epsilon}_f \tag{2.6}$$

이다. 식 (2.6)에서 오차벡터 $\boldsymbol{\epsilon}_f$는 $(\boldsymbol{I}-\boldsymbol{X}_f\boldsymbol{X}_f{}^{-})\boldsymbol{\epsilon}$으로 정의된다.

3. 블록간 분석모형

Yates (1940)의 블록간 모형에 대한 조건은 식 (2.1)의 블록효과 β_j가 상수가 아닌 평균이 0이고 분산이 σ_b^2인 분포를 갖는 확률변수들로 가정한다. 또한, 블록효과 간의 상관성 뿐만아니라 블록효과와 블록내 오차 ϵ_{ij}간에도 상관성이 없음을 가정하고 있다. 이러한 가정하에서 블록 j에서 관측값들의 합은 블록합이므로 블록합을 $y_{\cdot j}$로 나타낼 때 블록합에 대한 John (1971)의 모형표현식을 따르면

$$\begin{aligned} y_{\cdot j} &= k\mu+\sum_{i=1}^{t} n_{ij}\tau_i+\left(k\beta_j+\sum_{i=1}^{t}\epsilon_{ij}\right) \\ &= k\mu+\sum_{i=1}^{t} n_{ij}\tau_i+\delta_j \end{aligned} \tag{3.1}$$

으로 주어진다. 단, n_{ij}는 처리 i가 블록 j에서 나타나면 1 아니면 0인 가변수이다. 식 (3.1)의 오차항은 블록효과 β_j에 대한 확률효과의 가정으로 인해 두 확률성분의 선형결합으로 주어지고 있다. 결합오차항 δ_j는 $E(\delta_j)=0$이고 $\mathrm{Var}(\delta_j)=k^2\sigma_\beta^2+k\sigma_\epsilon^2$인 분포를 따르게 된다. 모형식 (3.1)의 가정하에 최소제곱법을 적용하여 구해진 τ_i의 추정값을 $\tilde{\tau}_i$라 두면 $\tilde{\tau}_i$는 블록간 추정치이며 블록내 추정치(intrablock estimate) $\hat{\tau}_i$와는 독립적으로 구해진다. 식 (3.1)의 행렬표현식은 식 (2.2)의 $\boldsymbol{X}_2$를 이용하면

$$\boldsymbol{X}_2'\boldsymbol{y} = \boldsymbol{X}_2'\boldsymbol{j}\mu+\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{\tau}+\boldsymbol{X}_2'\boldsymbol{\delta} \tag{3.2}$$

로 표현된다. 식 (3.2)의 $\boldsymbol{X}_2'\boldsymbol{y}$는 식 (2.2)의 관측벡터 $\boldsymbol{y}$를 $\boldsymbol{X}_2'$에 의해 선형변환된 벡터이다. 행렬모형식 (3.2)는 BIBD의 블록간 분석을 위해 본 논문에서 제시하고 있는 모형이다. 식 (3.2)를 이용하여

블록간 처리효과벡터의 추정량을 사영으로 구해보자. 사영에 의한 추정방법으로 단계별 방법(stepwise procedure)를 적용하기로 한다. 단계별 방법에 의한 μ의 추정을 위한 모형은

$$\boldsymbol{X}_2'\boldsymbol{y} = \boldsymbol{X}_2'\boldsymbol{j}\mu + \boldsymbol{\delta}_1 \tag{3.3}$$

이다. 모형식 (3.3)의 가정하에 $\boldsymbol{X}_2'\boldsymbol{y}$를 μ의 계수행렬인 $\boldsymbol{X}_2'\boldsymbol{j}$로의 사영을 $\boldsymbol{p}_\mu$라 두고 $\boldsymbol{p}_\mu$를 구하면

$$\boldsymbol{p}_\mu = \boldsymbol{X}_2'\boldsymbol{j}\left(\boldsymbol{X}_2'\boldsymbol{j}\right)^{-}\boldsymbol{X}_2'\boldsymbol{y} \tag{3.4}$$

이다. 사영까지의 제곱거리를 ssp_μ라 둘 때 $\mathrm{ssp}_\mu = (\boldsymbol{p}_\mu)'\boldsymbol{p}_\mu$로 구해진다. $\boldsymbol{X}_2'\boldsymbol{y}$에서 $\boldsymbol{p}_\mu$를 뺀 양을 $\boldsymbol{r}$로 나타내면

$$\boldsymbol{r} = \left[\boldsymbol{I} - \boldsymbol{X}_2'\boldsymbol{j}\left(\boldsymbol{X}_2'\boldsymbol{j}\right)^{-}\right]\boldsymbol{X}_2'\boldsymbol{y} \tag{3.5}$$

이다. 잔차벡터 $\boldsymbol{r}$을 이용하여 $\boldsymbol{\tau}$를 추정하기 위한 단계별 적합모형은

$$\begin{aligned}\boldsymbol{r} &= \left[\boldsymbol{I} - \boldsymbol{X}_2'\boldsymbol{j}\left(\boldsymbol{X}_2'\boldsymbol{j}\right)^{-}\right]\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{\tau} + \boldsymbol{\delta}_2 \\ &= \boldsymbol{X}_\tau\boldsymbol{\tau} + \boldsymbol{\delta}_2\end{aligned} \tag{3.6}$$

로 주어진다. 단, $\boldsymbol{X}_\tau = [\boldsymbol{I} - \boldsymbol{X}_2'\boldsymbol{j}(\boldsymbol{X}_2'\boldsymbol{j})^{-}]\boldsymbol{X}_2'\boldsymbol{X}_1$이다. 식 (3.6)에서 잔차벡터 $\boldsymbol{r}$을 이용한 $\boldsymbol{X}_\tau$로의 사영을 $\boldsymbol{p}_\tau$라 두면

$$\boldsymbol{p}_\tau = \boldsymbol{X}_\tau\boldsymbol{X}_\tau^{-}\boldsymbol{r} \tag{3.7}$$

로 주어진다. $\boldsymbol{X}_\tau$로의 사영 $\boldsymbol{p}_\tau$로부터 처리효과의 블록간 추정벡터 $\tilde{\boldsymbol{\tau}}$를 구하면

$$\tilde{\boldsymbol{\tau}} = \boldsymbol{X}_\tau^{-}\boldsymbol{r} \tag{3.8}$$

이다. $\tilde{\boldsymbol{\tau}}$와 처리효과의 블록내 추정벡터$\hat{\boldsymbol{\tau}}$는 상관성이 없는 추정벡터이다.

4. 블록간 정보의 복구

Yates (1940)는 BIBD의 분석에서 한 처리대비(treatment contrast)의 추정치로 블록내 추정치와 블록간 추정치라 불리우는 두 독립적인 불편추정치를 논의하고 일반적인 블록내 추정치보다 정확도가 높은 처리대비를 추정하기 위하여 이들 두 추정치를 결합시키는 방법을 다루었다. 블록간 정보의 복구(recovery of interblock information)라 불리우는 이러한 분석을 2절에서 제시한 행렬모형식을 이용하여 다루어 보기로 한다. $\boldsymbol{\tau}$의 두 추정벡터 $\hat{\boldsymbol{\tau}}$와 $\tilde{\boldsymbol{\tau}}$의 관련성을 알아보기 위해 $\mathrm{Cov}\binom{\hat{\boldsymbol{\tau}}}{\tilde{\boldsymbol{\tau}}}$를 구해보면

$$\begin{aligned}\mathrm{Cov}\begin{pmatrix}\hat{\boldsymbol{\tau}}\\ \tilde{\boldsymbol{\tau}}\end{pmatrix} &= \mathrm{Cov}\begin{pmatrix}\boldsymbol{X}_{11}{}^{-}\boldsymbol{r}_2\\ \boldsymbol{X}_\tau^{-}\boldsymbol{r}\end{pmatrix} \\ &= \mathrm{Cov}\begin{pmatrix}\boldsymbol{X}_{11}{}^{-}\left(\boldsymbol{I} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-}\right)\left(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^{-}\right)\boldsymbol{y}\\ \boldsymbol{X}_\tau{}^{-}\left[\boldsymbol{I} - \boldsymbol{X}_2'\boldsymbol{j}\left(\boldsymbol{X}_2'\boldsymbol{j}\right)^{-}\right]\boldsymbol{X}_2'\boldsymbol{y}\end{pmatrix} \\ &= \begin{pmatrix}\sigma_\epsilon^2\boldsymbol{R}_1\boldsymbol{R}_1' & \boldsymbol{0}\\ \boldsymbol{0} & \sigma_\beta^2\boldsymbol{R}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^2\boldsymbol{R}_2' + \sigma_\epsilon^2\boldsymbol{R}_2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)\boldsymbol{R}_2'\end{pmatrix}\end{aligned} \tag{4.1}$$

으로 계산된다. 단, $\boldsymbol{R}_1 = \boldsymbol{X}_{11}{}^{-}(\boldsymbol{I} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^{-})(\boldsymbol{I} - \boldsymbol{j}\boldsymbol{j}^{-})$이고 $\boldsymbol{R}_2 = \boldsymbol{X}_\tau{}^{-}[\boldsymbol{I} - \boldsymbol{X}_2'\boldsymbol{j}(\boldsymbol{X}_2'\boldsymbol{j})^{-}]$이다. 식

(4.1)에서 $\mathrm{Cov}(\hat{\boldsymbol{\tau}}, \tilde{\boldsymbol{\tau}}) = \mathbf{0}$를 나타내므로 $\hat{\boldsymbol{\tau}}$와 $\tilde{\boldsymbol{\tau}}$는 독립인 추정벡터임을 보여준다. 결합추정벡터를 $\hat{\boldsymbol{\psi}}$라 두자. i번째 성분을 $\hat{\psi}_i$라 두면

$$\hat{\psi}_i = w_1\hat{\tau}_i + w_2\tilde{\tau}_i \tag{4.2}$$

이다. 단, w_1과 w_2는 $w_1 + w_2 = 1$이 되는 가중치를 나타낸다. 두 추정량 $\hat{\tau}_i$과 $\tilde{\tau}_i$의 분산추정치를 이용하여 가중값을 구하게 된다. 두 추정량의 분산추정값은 식 (4.1)로부터 구해진다. 처리효과 τ의 결합추정치 $\hat{\psi}_i$를 얻기 위해 w_1과 w_2의 값을 계산해야 한다. w_1과 w_2의 계산은 각 추정량의 분산추정값에 근거하고 있으므로 각 추정량의 분산을 구해야 한다. 개별 추정량의 분산은 블록내 오차변동을 나타내는 분산성분 σ_ϵ^2과 확률효과로 가정된 블록효과의 분산 σ_β^2의 선형함수로 주어지기 때문에 이들 분산성분을 추정할 필요가 있다. 분산성분들인 σ_ϵ^2과 σ_β^2의 추정을 위한 식 (2.2)의 모형행렬을 $\boldsymbol{X} = (\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)$라 두자. 분산성분 σ_β^2을 추정하기 위한 모형은 식 (2.6)으로부터

$$\begin{aligned}\boldsymbol{r}_f &= \left(\boldsymbol{I} - \boldsymbol{X}_f\boldsymbol{X}_f^{\,-}\right)\boldsymbol{y}\\ &= \left(\boldsymbol{I} - \boldsymbol{X}_f\boldsymbol{X}_f^{\,-}\right)\boldsymbol{X}_2\boldsymbol{\beta} + \boldsymbol{\epsilon}_f\\ &= \boldsymbol{X}_b\boldsymbol{\beta} + \boldsymbol{\epsilon}_f\end{aligned} \tag{4.3}$$

이다. 처리효과에 적합된 블록변동량은 $\boldsymbol{r}_b'\boldsymbol{X}_b\boldsymbol{X}_b^{\,-}\boldsymbol{r}_b$로 계산되고 $\boldsymbol{X}_b$는 $(\boldsymbol{I} - \boldsymbol{X}_f\boldsymbol{X}_f^{\,-})\boldsymbol{X}_2$이다. 오차성분 σ_ϵ^2의 추정은 $\boldsymbol{y}$를 $\boldsymbol{X}$로의 사영을 뺀 잔차벡터의 제곱합을 이용할 수 있다. 즉, 오차제곱합은 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$로 계산된다. 처리효과에 적합된 블록제곱합과 오차성분의 잔차제곱합은 $\boldsymbol{y}$의 2차형식으로 주어지므로 적률법에 의한 연립방정식의 해로 분산성분들의 추정치를 얻을 수 있다. 가중치 w_1과 w_2는 처리효과 τ의 블록내 추정량 $\hat{\tau}$의 분산 $\sigma_{\hat{\tau}}^2$와 블록간 추정량 $\tilde{\tau}$의 분산 $\sigma_{\tilde{\tau}}^2$를 이용하여 구한다. 이들 분산은 식 (4.1)을 이용하여 계산할 수 있다. 가중치의 계산은

$$w_1 = \frac{\left(\sigma_{\hat{\tau}}^2\right)^{-1}}{\left(\sigma_{\hat{\tau}}^2\right)^{-1} + \left(\sigma_{\tilde{\tau}}^2\right)^{-1}}, \qquad w_2 = \frac{\left(\sigma_{\tilde{\tau}}^2\right)^{-1}}{\left(\sigma_{\hat{\tau}}^2\right)^{-1} + \left(\sigma_{\tilde{\tau}}^2\right)^{-1}} \tag{4.4}$$

으로 주어진다.

5. BIBD 분석자료의 예

균형불완비블록설계의 자료로부터 처리효과의 블록내 추정치와 블록간 추정치의 결합추정치를 구하는 과정을 살펴보기로 한다. 다음 Table 5.1의 자료는 John (1961)의 식기세척 실험으로부터 취해진 자료이다.

표에서 처리는 세제(detergent)이고 관측반응은 표준시험조건하에서 세척된 접시의 수를 나타낸다. 블록은 세 가지 세제를 동시에 검사하는 세 명의 세척조이다. 비교하고자 하는 처리의 수 $t = 9$이고 블록의 수 $b = 12$이다. 각 처리의 반복수 $r = 4$이고 블록의 크기 $k = 3$이며 각 처리쌍이 동일 블록에서 나타나는 횟수는 $\lambda = 1$이다.

Table 5.2의 자료에 대한 블록내 분석모형식 (2.2)와 블록간 분석모형식 (3.2)의 가정하에서 블록간 정보의 복구과정을 살펴보기로 한다. 우선 블록내 분석모형으로 처리효과벡터 $\boldsymbol{\tau}$의 값을 추정하면 식 (2.5)로부터 $\hat{\boldsymbol{\tau}}' = (0.33, -2.22, -6.22, -12.89, 5.89, 3.55, 1.67, -0.22, 10.11)$로 구해진다. 블록내 분석에서 구해지는 분산분석표는 Table 5.2이다. 표로부터 식 (3.8)의 $\hat{\sigma_\epsilon}^2$이 0.82임을 알 수 있다. 블록간 분석모형에 의한 블록간 처리효과들의 추정값은 식 (3.2)의 단계별 적합으로 부터 구해지게 된다. 식 (3.4)의 적합으로부터 구해지는 처리효과의 블록간 추정벡터는 $\tilde{\boldsymbol{\tau}}' = (0.33, -4.00, -6.00, -13.00,$

Table 5.1. Balanced incomplete block data

Block (Set of 3 operators)	Treatment (number of plates)		
1	$A(19)$	$B(17)$	$C(11)$
2	$D(6)$	$E(26)$	$F(23)$
3	$G(21)$	$H(19)$	$J(28)$
4	$A(20)$	$D(7)$	$G(20)$
5	$B(17)$	$E(26)$	$H(19)$
6	$C(15)$	$F(23)$	$J(31)$
7	$A(20)$	$E(26)$	$J(31)$
8	$B(16)$	$F(23)$	$G(21)$
9	$C(13)$	$D(7)$	$H(20)$
10	$A(20)$	$F(24)$	$H(19)$
11	$B(17)$	$D(6)$	$J(29)$
12	$C(14)$	$E(24)$	$G(21)$

Table 5.2. Analysis of variance for intrablock

Source of variation	Sum of squares	Degrees of Freedom	Mean Square
Blocks (unadjusted)	412.75	11	
Blocks (adjusted)	10.06	11	0.91
Treatments (unadjusted)	1489.50	8	
Treatments (adjusted)	1086.81	8	135.85
Error	13.19	16	0.82
Total	1512.75	35	

$6.67, 4.67, 0.33, 0.00, 11.00)$이다. 처리효과의 블록내 추정치와 블록간 추정치는 상관성이 없으므로 처리효과의 결합추정치를 구할 수 있다. 블록내 추정치와 블록간 추정치의 무상관성은 $\mathrm{Cov}(\hat{\boldsymbol{\tau}}, \tilde{\boldsymbol{\tau}}) = \mathbf{0}$로 부터 알 수 있다. 처리효과의 결합추정치는 식 (4.2)로부터 구해지나 블록효과가 확률효과이므로 분산성분인 σ_β^2을 추정해야 한다. 블록내 분석에서 주어지는 분산성분들은 처리효과에 적합된 블록변동량과 오차에 따른 변동량이다. 이들 두 분산성분은 사영제곱합으로 구해진다. 처리효과에 적합된 블록변동량 $\boldsymbol{y}'\boldsymbol{X}_b\boldsymbol{X}_b{}^-\boldsymbol{y} = 10.06$으로 계산된다. 오차제곱합 $\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y} = 13.19$로 구해진다. 분산성분의 추정을 위해 적률법을 이용할 때 연립방정식은

$$\begin{aligned} 10.06481 &= 27\sigma_b^2 + 11\sigma_\epsilon^2, \\ 13.18519 &= 16\sigma_\epsilon^2 \end{aligned} \tag{5.1}$$

이다. 연립방정식의 해로 $\sigma_b^2 = 0.04$와 $\sigma_\epsilon^2 = 0.82$를 구한다. 식 (4.4)에 의한 가중치를 계산하면 $w_1 = 0.991$로 주어지고 $w_2 = 0.009$로 구해진다. 이들 가중치를 이용하여 구하면 $\hat{\psi_1} = w_1\hat{\tau_1} + w_2\tilde{\tau_1} = 0.991(0.33) + 0.09(0.33) = 0.33$이다. $\hat{\psi_2} = -2.62$, $\hat{\psi_3} = -6.17$, $\hat{\psi_4} = -12.91$, $\hat{\psi_5} = 6.07$, $\hat{\psi_6} = 3.80$, $\hat{\psi_7} = 1.37$, $\hat{\psi_8} = -0.17$이고 $\hat{\psi_9} = 10.31$로 구해진다.

6. 결론

본 논문은 균형불완비블록설계(balanced incomplete block design)의 블록내 분석(intrablock analysis)과 블록간 분석(interblock analysis)에 사영을 이용하는 방법을 논의하고 있다. 블록내 분석에서 행렬표현식의 모형으로 부터 처리효과의 블록내 추정벡터를 사영으로 구하는 과정을 모형의 단계별 적합

방식으로 설명하고 있다. 또한, 블록내 분석모형에서 처리효과에 적합된 블록변동량의 계산에 상수적합법(fitting constants method)을 적용할 때 이용되는 모형의 형태를 다루고 있다. 블록간 분석을 위한 블록합의 모형으로 행렬표현식의 모형을 제시하고 있으며 제시된 행렬표현식을 이용하여 처리효과의 블록간 추정벡터를 얻는 방법을 제공하고 있다.

처리효과의 블록내 추정벡터와 블록간 추정벡터는 서로 상관성이 없음을 두 추정벡터의 분산공분산행렬을 통하여 규명하고 있으며 처리효과의 결합추정치를 얻기 위한 가중치가 두 추정벡터의 분산공분산행렬로 부터 어떻게 구해지는 가를 구체적으로 논의하고 있다. 또한, 처리효과에 적합된 블록제곱합은 확률효과로 간주된 블록효과의 변동량이며 상수제곱법으로 할 수 있음을 보여주고 있다. 그리고 적률법으로 분산성분을 구하는 과정을 John (1971)의 자료 예를 통해 상세히 다루고 있다.

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, New York.

Davies, O. L. (1956). *Design and Analysis of Industrial Experiments*, Second edition, Hafner Publishing Company, New York.

Henderson, C. R. (1953). Estimation of variance and covariance components, *Biometrics*, **9**, 226−252.

Hicks, C. R. (1973). *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart and Winston, New York.

John, P. W. M. (1961). An application of a balanced incomplete block design, *Technometrics*, **3**, 51−54.

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, New York.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*, John Wiley and Sons, New York.

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

Yates, F. (1940). The recovery of interblock information in balanced incomplete block designs, *Annals of Eugenics*, **10**, 317−325.

균형불완비블록설계의 혼합효과에서 블록간 정보

최재성[a,1]

[a]계명대학교 통계학과

(2015년 2월 10일 접수, 2015년 3월 12일 수정, 2015년 3월 12일 채택)

요 약

본 논문은 균형불완비블록설계(balanced incomplete block design)에서 사영에 근거한 블록내(intrablock) 분석과 블록간(interblock) 분석을 다루고 있다. 블록간 분석을 위한 행렬모형을 제시하고 블록간 추정벡터를 구하는 방법을 다루고 있다. 처리효과의 블록내 추정벡터와 블록간 추정벡터의 분산공분산행렬을 규명하고 공분산행렬의 구조적 특성으로 두 추정벡터 간에 상관성이 없음을 보여주고 있다. 처리효과의 상관성없는 두 추정벡터를 이용한 결합추정에서 가중치를 구하는 방법으로 공분산행렬을 이용할 수 있음을 다루고 있다. 또한 처리효과에 적합된 블록변동량의 계산은 상수적합법을 이용한 블록제곱합과 일치함을 보여주고 있다.

주요용어: 균형불완비블록설계, 결합추정, 블록내 분석, 블록간 분석, 사영.

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