Journal of the Korean Data & Information Science Society
2015, **26**(1), 31–39
http://dx.doi.org/10.7465/jkdi.2015.26.1.31
한국데이터정보과학회지

사영에 의한 확률효과모형의 분석

최재성[1]

[1]계명대학교 통계학과

접수 2014년 11월 4일, 수정 2014년 12월 6일, 게재확정 2014년 12월 15일

요 약

본 논문은 확률효과모형에서 사영에 근거한 분산성분을 구하는 방법을 다루고 있다. 분산성분을 추정하기 위한 ANOVA방법에서 제곱합의 계산에 사영을 이용하는 방법을 제시하고 있다. 분산성분을 구하기 위한 사영의 이용은 모형행렬에 의한 사영공간을 분산성분별 제곱합을 얻기 위한 상호직교하는 부분공간들로 분할하게 된다. 부분공간들로 분할하기 위해 모형행렬 $\boldsymbol{X}$로의 사영에 단계별 방법 (stepwise procedure)을 적용하여 해당하는 공간으로의 사영행렬을 구하는 방법을 다루고 있다. 단계별 방법에 의해 주어지는 부분공간들의 직교성으로 인해 사영행렬의 곱은 영행렬로 주어지는 성질을 갖는다. 단계별 방법에 의한 순차적 사영은 해당하는 공간으로의 사영행렬에 대한 확인과 사영행렬의 구조를 파악할 수 있는 이점이 있다. 또한 분산성분의 추정을 위한 제1종 제곱합을 구하기 위한 방법으로 유용하다.

주요용어: 단계별 방법, 분산성분, 사영, 제1종 제곱합, 확률효과

1. 서론

실험자료를 분석하기 위한 선형모형이 확률모형인 경우를 가정한다. 이 경우에 실험단위의 반응 y에 영향을 주는 독립변수들은 모두 유한개의 수준을 갖는 요인으로 간주된다. 각 요인의 수준들이 모집단에서 임의로 추출된다고 가정할 때 수준효과의 변동을 나타내는 분산성분에 관심을 두게 된다. 다양한 유형의 확률모형과 분산성분의 추정방법에 관한 논의가 Milliken과 Johnson (1984), Montgomery (1976) 그리고 Searle (1992) 등에서 다루어지고 있다. Choi (2011, 2012, 2014)는 사영에 의한 분석방법을 제시하고 있다. 사영과 행렬에 관한 이론적 배경은 Graybill (1976)에서 구체적 논의를 살펴볼 수 있다.

확률효과모형에서 분산성분의 추정방법으로 적률법, 최우법과 MINQUE 방법 등을 이용할 수 있다. 균형자료의 경우에 세 방법들은 동일한 추정량을 나타내나 자료가 불균형일 때는 동일하지 않게 된다. 적률법은 불편추정량을 제공하며 자료가 불균형일 때 계산이 가장 쉬운 방법이나 다른 두 방법은 반복적인 수치연산과정을 필요로 한다. 적률법에 의한 분산성분의 추정방법은 분산성분의 추정량이 되는 평균제곱의 계산을 필요로 한다. 변동요인에 따른 제곱합을 구하는 방법으로 균형자료의 ANOVA 방법이 불균형자료에서도 동일하게 이용된다. Henderson (1953) 방법 I로도 불리우는 ANOVA 방법은 확률모형을 고정효과모형으로 가정하여 분산분석한 다음에 확률모형하에 평균평방의 기대값을 유도하게 된다. 균형자료의 경우 변동요인에 따른 제곱합을 구하는 방법은 전적으로 ANOVA 방법에 의존하나 불균형자료의 경우에 변동요인에 따른 제곱합을 구하는 방법은 다양하게 주어지고 방법에 따라 이차형식들도 다양하게 주어진다.

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

확률모형이 가정된 불균형자료로 부터 분산성분을 추정할 때 총제곱합을 분할하는 여러 방법들이 있다. 그러나, 분산성분을 추정하기 위한 목적으로 이들 중 하나를 이용해야 할 때 다른 방법에 비해 특정 하나를 선택하게 되는 기준이 없다. 따라서 분산성분을 추정하기 위해 최적으로 이용될 수 있는 분명한 제곱합 또는 이차형식의 집합이 유일하지 않기 때문에 이용될 수 있는 이차형식 집단의 다양성이 있게 된다.

확률모형의 분산성분을 추정하기 위한 ANOVA 방법 또는 Henderson 방법 I은 확률요인에 따른 제곱합을 구하기 위해 총제곱합을 변동요인에 따른 제곱합으로 분할하는 고정효과모형의 분산분석을 이용한다. 이들 제곱합의 기대값은 확률모형하에서 분산성분들의 선형함수로 구해진다. 분산성분들의 선형함수로 주어지는 제곱합의 기대값을 해당하는 제곱합과 같게 두는 선형방정식들의 해로 적률추정량을 구하게 된다. 변동요인에 따른 제곱합으로 Henderson 방법 I에 의한 제1종 제곱합의 이차형식들을 이용해 보기로 한다. 불균형자료에서 요인별 제1종 제곱합은 모형의 비교방식에서 모형내 모수들의 순서가 달라질 때 서로 다른 제곱합을 얻기 때문이다.

본 논문은 불균형자료에 대한 확률모형의 가정에서 적률법으로 분산성분을 추정하는 데 관심을 두고 있다.확률모형의 분산성분 추정에 이용되는 Henderson 방법 I은 확률효과를 고정효과로 간주하여 요인별 제곱합을 구하고 있으므로 동일한 양을 벡터공간에서의 사영을 이용하여 구할 수 있다는 점에 착안하고 있다. 불균형자료에 대한 Henderson 방법 I의 적용은 모형의 적합방식에 따라 서로 다른 이차형식들이 가능하므로 사영에 의한 제1종 제곱합을 이용하여 분산성분을 추정하고자 한다. 또한, Henderson 방법 I에서 제곱합의 기댓값 계산에 이용되는 Hartley의 합성법이 사영에서 어떻게 이용될 수 있는 가를 다루고자 한다.

2. p개 분산성분의 확률효과모형

실험자료의 분석을 위한 선형모형으로 다수의 확률요인을 갖는 확률모형을 가정한다. 실험단위의 반응을 y라 두고 반응에 영향을 미치는 확률요인들의 주효과와 교호작용을 나타내는 p개 확률효과를 α_1, α_2, $\cdots$, α_p라 두자. p개 확률효과를 갖는 확률모형을 행렬표현식으로 나타내면

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon} \tag{2.1}$$

이다. 단, $\boldsymbol{y}$는 크기 $n \times 1$인 관측벡터이고 모평균 μ의 계수벡터 $\boldsymbol{j}$는 n개의 원소가 모두 1인 열벡터이다. $\boldsymbol{X}_1$은 크기가 $n \times k_1$인 0과 1로 구성되는 계수행렬이다. $\boldsymbol{\alpha}_1$은 확률효과벡터이고 크기가 $k_1 \times 1$인 열벡터이다. $\boldsymbol{X}_2$는 크기가 $n \times k_2$인 계수행렬이다. $\boldsymbol{\alpha}_2$는 크기가 $k_2 \times 1$인 열벡터이다. $\boldsymbol{X}_p$는 $n \times k_p$인 계수행렬이다. $\boldsymbol{\alpha}_p$는 확률효과벡터이고 크기가 $k_p \times 1$인 열벡터이다. $\boldsymbol{\epsilon}$는 $n \times 1$인 오차벡터이다. $\boldsymbol{\alpha}_i, i = 1, 2, \cdots, p$는 확률효과벡터이므로 다변량 정규분포를 가정한다. 즉, $\boldsymbol{\alpha}_1$은 $N(\boldsymbol{0}, {\sigma_1}^2\boldsymbol{I}_{k_1})$, $\boldsymbol{\alpha}_2$는 $N(\boldsymbol{0}, \sigma_2^2\boldsymbol{I}_{k_2})$,$\cdots$, $\boldsymbol{\alpha}_p$는 $N(\boldsymbol{0}, \sigma_p^2\boldsymbol{I}_{k_p})$이다. $\boldsymbol{\epsilon}$은 $N(\boldsymbol{0}, \sigma_\epsilon^2\boldsymbol{I}_n)$인 분포를 따른다고 가정한다. 식 (2.1)의 행렬모형식에서 $\boldsymbol{\alpha}_i, i = 1, 2, \cdots, p$가 확률벡터이므로 관심의 모수들은 모평균 μ, 다변량 정규분포의 분산을 나타내는 σ_1^2, σ_2^2, $\cdots$, σ_p^2와 오차벡터의 분산 σ_ϵ^2이다.

3. Henderson방법 I

식 (2.1)에서 확률벡터들의 분산성분을 추정하기 위한 Henderson방법 I을 먼저 살펴보기로 한다. Henderson방법 I에 대한 구체적 논의는 Searle (1971) 등에서 살펴볼 수 있다. Henderson방법 I은 행렬모형식 (2.1)의 확률벡터를 고정효과벡터로 간주하여 고정효과벡터에 따른 제곱합을 구한다. 일반선형모형에서 분산분석은 총제곱합 $\boldsymbol{y}'\boldsymbol{y}$를 $p+2$개 이차형식들의 합으로 분할하는 방법이다. 즉, $\boldsymbol{y}'\boldsymbol{y} =$

$\sum_{i=1}^{k} \boldsymbol{y}'\boldsymbol{A}_i\boldsymbol{y}$이다. $p+2$개 이차형식들의 합으로 분할하기 위한 방법은 제1종 제곱합을 구하는 모형비교방식을 이용하는 것이다. 왜냐하면, 제1종 제곱합은 불균형 자료에서도 모형의 순차적 적합을 통한 총제곱합의 분할을 제공하기 때문이다. 식 (2.1)에서 μ에 따른 제곱합을 계산하기 위한 분산분석법은 식 (2.1)의 축소모형인

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{\epsilon} \tag{3.1}$$

을 자료에 적합시켜 $R(\mu)$를 구한다. 여기서 $R(\beta)$는 오차분산 σ_ϵ^2의 추정을 위한 제곱합에서 모형내 모수벡터β에 따른 감소량을 나타낸다. R$(\cdot)$를 모수$(\cdot)$에 따른 제곱합에서의 감소량 (reduction in sum of squares)이라 부른다. $R(\mu)$는 μ가 포함되지 않은 모형 $\boldsymbol{y} = \boldsymbol{\epsilon}$에서의 오차제곱합 $\boldsymbol{y}'\boldsymbol{y}$에서 μ가 포함된 식 (3.1)의 적합으로부터 오차분산의 추정을 위한 잔차제곱합 $RSS(\mu)$=$\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}$을 뺀 양으로 구해진다. 여기서 $\boldsymbol{j}^-$는 $\boldsymbol{j}$의 Moore-Penrose의 일반화된 역행렬 (generalized inverse)을 나타내므로 $R(\mu)$=$\boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}$로 구해진다. 이 양은 μ의 계수벡터 $\boldsymbol{j}$로의 사영에 의한 사영제곱합과 일치한다. 이를 ssj라 두면

$$R(\mu) = \boldsymbol{y}'\boldsymbol{y} - (\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{y} - RSS(\mu) = \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y} = ssj \tag{3.2}$$

이다. $\boldsymbol{\alpha}_1$에 따른 제곱합은 식 (3.1)에 $\boldsymbol{\alpha}_1$이 추가된 모형

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{\epsilon} \tag{3.3}$$

을 자료에 적합시켜 오차분산을 추정하기위한 잔차제곱합 $RSS(\mu, \alpha_1)$을 구한다. $\boldsymbol{\alpha}_1$에 따른 제곱합은

$$\begin{aligned} R(\boldsymbol{\alpha}_1|\mu) &= RSS(\mu) - RSS(\mu, \boldsymbol{\alpha}_1) \\ &= (\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}) - \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1)(\boldsymbol{j}, \boldsymbol{X}_1)^-\boldsymbol{y}) \\ &= \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1)(\boldsymbol{j}, \boldsymbol{X}_1)^-\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y} \\ &= ss1 \end{aligned} \tag{3.4}$$

이다. $\boldsymbol{\alpha}_2$에 따른 제곱합은 식 (3.3)에 $\boldsymbol{\alpha}_2$를 추가한 모형

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \boldsymbol{\epsilon} \tag{3.5}$$

을 이용한다. 식 (3.5)의 적합으로부터 주어지는 잔차제곱합을 $RSS(\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$라 두면 $\boldsymbol{\alpha}_2$에 따른 제곱합 $R(\boldsymbol{\alpha}_2|\mu, \boldsymbol{\alpha}_1)$은

$$\begin{aligned} R(\boldsymbol{\alpha}_2|\mu, \boldsymbol{\alpha}_1) &= RSS(\mu, \boldsymbol{\alpha}_1) - RSS(\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1)(\boldsymbol{j}, \boldsymbol{X}_1)^-\boldsymbol{y}) - [\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)^-\boldsymbol{y}] \\ &= \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2)^-\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1)(\boldsymbol{j}, \boldsymbol{X}_1)^-\boldsymbol{y} \\ &= ss2 \end{aligned} \tag{3.6}$$

이다. $\boldsymbol{\alpha}_p$에 따른 제곱합은 식 (3.5)에 $\boldsymbol{\alpha}_p$를 추가한 모형

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_1\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon} \tag{3.7}$$

을 이용한다. 식 (3.7)의 적합으로부터 주어지는 잔차제곱합을 $RSS(\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_p)$라 두면 $\boldsymbol{\alpha}_p$에

따른 제곱합 $R(\boldsymbol{\alpha}_p|\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{p-1})$은

$$\begin{aligned}
R(\boldsymbol{\alpha}_p|\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{p-1}) &= RSS(\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{p-1}) - RSS(\mu, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_p) \\
&= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{p-1})(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{p-1})^{-}\boldsymbol{y} \\
&- [\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}\prime(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)^{-}\boldsymbol{y}] \\
&= \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)^{-}\boldsymbol{y} \\
&- \boldsymbol{y}'(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{p-1})(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_{p-1})^{-}\boldsymbol{y} \\
&= ssp \qquad (3.8)
\end{aligned}$$

이다.

4. 사영행렬을 이용한 제곱합과 기댓값

불균형자료의 분산분석을 위한 제1종 제곱합은 모형의 순차적 적합방식을 이용하여 총제곱합을 요인별 제곱합으로 분할한다. 요인별 제곱합은 모형의 순차적 적합에서 주어지는 잔차제곱합의 차로 계산하게 된다. 이는 벡터 $\boldsymbol{y}$를 모형행렬 $\boldsymbol{X}$로의 사영에 따른 성분제곱합을 구하는 것과 동일하다. 여기서 $\boldsymbol{X}$=$(\boldsymbol{j}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_p)$인 분할행렬을 나타낸다. 분할행렬 $\boldsymbol{X}$로의 사영에서 사영까지의 거리제곱합은 $\boldsymbol{X}$의 일반화된 역행렬 $\boldsymbol{X}^{-}$로 나타낼 때 $\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$이다. 이 양을 성분제곱합으로 분해하기 위해 모형행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간을 상호직교하는 부분공간으로 분할하여 분할된 공간에서의 사영을 이용한다. 모형행렬 $\boldsymbol{X}$로의 사영은 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$이므로

$$\begin{aligned}
\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{X}^{-}(\boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon}) \\
&= \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{\epsilon} \qquad (4.1)
\end{aligned}$$

인 모형을 이용한다. 이는 고정효과모형에서 모형행렬 $\boldsymbol{X}$로의 사영인 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$를 이용하여 각 부분공간으로의 사영을 나타내는 사영행렬을 구하는 방식이다. 식 (4.1)을 이용하여 분할행렬 $\boldsymbol{X}$의 부분행렬과 관련된 공간으로의 사영행렬을 단계별 방법 (stepwise procedure)에 의해 구할 수 있다. 이는 잔차를 이용하는 모형으로부터 사영행렬에 의한 제1종 제곱합을 구할 수 있는 방법을 제공하고 있다. 즉, $\boldsymbol{y}$의 사영이 행해지는 $\boldsymbol{X}$로의 사영공간을 $p+1$개의 상호직교하는 부분공간으로 분할한다. 부분공간으로의 사영이 행해지고 각 부분공간에서 사영까지의 거리제곱합이 성분제곱합으로 이용된다. 모형행렬 $\boldsymbol{X}$에 의한 벡터공간은 사영행렬 $\boldsymbol{X}\boldsymbol{X}^{-}$에 의한 사영공간이므로 $\boldsymbol{X}\boldsymbol{X}^{-}$의 분할행렬이 직교하는 부분행렬들로 구성될 때 이를 이용하여 제1종 제곱합을 구한다. 식 (4.1)에서 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$를 모형내 계수행렬의 공간으로 순차적 사영을 생각해 보자. $\boldsymbol{j}$로의 사영은

$$\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y} = \boldsymbol{j}\boldsymbol{\mu} + \boldsymbol{\epsilon_1} \qquad (4.2)$$

인 모형으로부터 $\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$를 얻게 되므로 $\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y}$가 $\boldsymbol{j}$로의 사영이 된다. 이때의 사영공간을 나타내는 사영행렬은 $\boldsymbol{j}\boldsymbol{j}^{-}$임을 알 수 있다. $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$에서 $\boldsymbol{j}$로의 사영을 제외한 잔차는 $(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y}$이다. 이 잔차에 대한 모형은

$$(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y} = (\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})(\boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon}) \qquad (4.3)$$

이다. 잔차모형으로부터 $\boldsymbol{\alpha}_1$의 추정을 위한 벡터 부분공간으로의 사영을 위한 모형은

$$(\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y} = (\boldsymbol{X}\boldsymbol{X}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{\epsilon_2} \qquad (4.4)$$

이다. 모형행렬 $(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{X}_1$으로의 사영은 $[(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{X}_1][(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{X}_1]^-(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{y}$이다. $[(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{X}_1]$을 $\boldsymbol{X}_{11}$이라 두자. 식 (3.12)에서 $\boldsymbol{\alpha}_1$을 추정하기 위한 사영공간에서의 사영행렬은 $\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-$이다. $\boldsymbol{X}_2$와 관련된 공간으로의 사영을 구하기 위해 $\boldsymbol{XX}^-\boldsymbol{y}$에서 $\boldsymbol{j}$로의 사영과 $\boldsymbol{X}_1$과 관련된 공간으로의 사영을 제외한 잔차를 $\boldsymbol{r}_{11}\boldsymbol{y}$라 두자. $\boldsymbol{r}_{11}$은 $(\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-(\boldsymbol{XX}^- - \boldsymbol{jj}^-))$이다. 잔차모형은

$$\boldsymbol{r}_{11}\boldsymbol{y} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-(\boldsymbol{XX}^- - \boldsymbol{jj}^-))(\boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon}) \tag{4.5}$$

로 주어진다. 식 (4.5)에서 $(\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-(\boldsymbol{XX}^- - \boldsymbol{jj}^-))$와 $\boldsymbol{j}$와의 곱은 $\boldsymbol{0}$이고 $\boldsymbol{X}_1$과의 곱도 $\boldsymbol{0}$임을 알 수 있다. 또한 $(\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-(\boldsymbol{XX}^- - \boldsymbol{jj}^-)) = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-)$이므로 $\boldsymbol{\alpha}_2$의 추정을 위한 모형은

$$\boldsymbol{r}_{11}\boldsymbol{y} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-)\boldsymbol{X}_2\boldsymbol{\alpha}_2 + \boldsymbol{\epsilon_{11}} \tag{4.6}$$

로 표현된다. $\boldsymbol{X}_{22} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-)\boldsymbol{X}_2$라 두면 $\boldsymbol{X}_{22}$로의 사영은 $\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_{11}\boldsymbol{y}$이다. 식 (4.6)에서 사영행렬은 $\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-$이고 $\boldsymbol{jj}^-$와 $\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-$와는 직교한다. $\boldsymbol{r}_{11}\boldsymbol{y}$에서 $\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_{11}\boldsymbol{y}$를 뺀 $(\boldsymbol{r}_{11} - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_{11})\boldsymbol{y}$를 $\boldsymbol{r}_{22}\boldsymbol{y}$라 두자. 잔차모형은

$$\begin{aligned}\boldsymbol{r}_{22}\boldsymbol{y} &= (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^- - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-) \\ &\quad \times (\boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha} + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_p\boldsymbol{\alpha}_p + \boldsymbol{\epsilon})\end{aligned} \tag{4.7}$$

이다. $\boldsymbol{r}_{22}$와 $\boldsymbol{j}$와의 곱, $\boldsymbol{r}_{22}$와 $\boldsymbol{X}_1$과의 곱 그리고 $\boldsymbol{r}_{22}$와 $\boldsymbol{X}_2$와의 곱이 모두 $\boldsymbol{0}$이 되므로 $\boldsymbol{\alpha}_3$의 추정을 위한 모형은

$$\boldsymbol{r}_{22}\boldsymbol{y} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^- - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-)\boldsymbol{X}_3\boldsymbol{\alpha}_3 + \boldsymbol{\epsilon}_{22} \tag{4.8}$$

이다. $\boldsymbol{X}_{33} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^- - \boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{r}_{11})\boldsymbol{X}_3$라 두면 $\boldsymbol{X}_{33}$로의 사영은 $\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{r}_{22}\boldsymbol{y}$이다. $\boldsymbol{\alpha}_3$의 추정을 위한 사영공간에서의 사영행렬은 $\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-$이다. $\boldsymbol{\alpha}_p$를 추정하기 위한 잔차모형은

$$\begin{aligned}\boldsymbol{r}_{(p-1)(p-1)}\boldsymbol{y} &= (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^- - \cdots - \boldsymbol{X}_{(p-1)(p-1)}\boldsymbol{X}_{(p-1)(p-1)}{}^-)\boldsymbol{X}_p\boldsymbol{\alpha}_p \\ &\quad + \boldsymbol{\epsilon}_{(p-1)(p-1)}\end{aligned} \tag{4.9}$$

이고 $\boldsymbol{X}_{pp} = (\boldsymbol{XX}^- - \boldsymbol{jj}^- - \boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^- - \cdots - \boldsymbol{X}_{(p-1)(p-1)}\boldsymbol{X}_{(p-1)(p-1)}{}^-)\boldsymbol{X}_p$라 두면 $\boldsymbol{X}_p$로의 사영행렬은 $\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-$이다. 따라서 $\boldsymbol{X}$로의 사영공간을 나타내는 사영행렬 $\boldsymbol{XX}^-$의 순차적 분할을 이용한 제1종 제곱합은

$$\boldsymbol{y}'\boldsymbol{XX}^-\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{jj}^-\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{y} + \cdots + \boldsymbol{y}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{y} \tag{4.10}$$

로 구해진다. 분산성분을 추정하기 위한 이들 이차형식 $\boldsymbol{y}'\boldsymbol{Q}^-\boldsymbol{y}$의 기대값은

$$E(\boldsymbol{y}'\boldsymbol{Qy}) = E[tr(\boldsymbol{y}'\boldsymbol{Qy})] = tr[(\boldsymbol{\Sigma} + \boldsymbol{\mu\mu}')\boldsymbol{Q})] = tr(\boldsymbol{Q\Sigma}) + tr(\boldsymbol{\mu}'\boldsymbol{Q\mu}) \tag{4.11}$$

이다. $\boldsymbol{\Sigma}$는 $\boldsymbol{y}$의 분산공분산 행렬이고 $\boldsymbol{\Sigma} = \sigma_1^2\boldsymbol{X}_1\boldsymbol{X}_1{}' + \sigma_2^2\boldsymbol{X}_2\boldsymbol{X}_2{}' + \cdots + \sigma_p^2\boldsymbol{X}_p\boldsymbol{X}_p{}' + \sigma_\epsilon^2\boldsymbol{I}_n$로 구해지며 $\boldsymbol{\mu}$는 $\boldsymbol{y}$의 평균벡터이다. 식 (4.10)에서 $\boldsymbol{y}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{y}$의 기대값은

$$\begin{aligned}E(\boldsymbol{y}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{y}) &= tr(\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{\Sigma}) \\ &= \sigma_1^2 tr(\boldsymbol{X}_1{}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{X}_1) + \sigma_2^2 tr(\boldsymbol{X}_2{}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{X}_2) \\ &\quad + \sigma_3^2 tr(\boldsymbol{X}_3{}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{X}_3) + \cdots \\ &\quad + \sigma_p^2 tr(\boldsymbol{X}_p{}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{X}_p) + \sigma_\epsilon^2 tr(\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-)\end{aligned} \tag{4.12}$$

이다. 여기서 tr은 정방행렬의 대각원소들의 합을 나타내는 대각합 (trace)이다. tr(A)는 정방행렬 A의 대각원소들의 합이고 A가 멱등행렬일 때 tr(A)는 A의 계수 (rank)와 일치한다. $\boldsymbol{y}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{y}$의 기대값은

$$\begin{aligned}E(\boldsymbol{y}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{y}) &= tr(\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{\Sigma})\\ &= \sigma_1^2 tr(\boldsymbol{X}_1{}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{X}_1) + \sigma_2^2 tr(\boldsymbol{X}_2{}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{X}_2)\\ &+ \sigma_3^2 tr(\boldsymbol{X}_3{}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{X}_3) + \cdots\\ &+ \sigma_p^2 tr(\boldsymbol{X}_p{}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{X}_p) + \sigma_\epsilon^2 tr(\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-)\end{aligned} \tag{4.13}$$

이다. $\boldsymbol{y}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{y}$의 기대값은

$$\begin{aligned}E(\boldsymbol{y}\prime\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{y}) &= tr(\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{\Sigma})\\ &= \sigma_1^2 tr(\boldsymbol{X}_1{}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{X}_1) + \sigma_2^2 tr(\boldsymbol{X}_2{}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{X}_2)\\ &+ \sigma_3^2 tr(\boldsymbol{X}_3{}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{X}_3) + \cdots\\ &+ \sigma_p^2 tr(\boldsymbol{X}_p{}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{X}_p) + \sigma_\epsilon^2 tr(\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-)\end{aligned} \tag{4.14}$$

이다. $\boldsymbol{y}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{y}$의 기대값은

$$\begin{aligned}E(\boldsymbol{y}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{y}) &= tr(\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{\Sigma})\\ &= \sigma_1^2 tr(\boldsymbol{X}_1{}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{X}_1) + \sigma_2^2 tr(\boldsymbol{X}_2{}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{X}_2)\\ &+ \sigma_3^2 tr(\boldsymbol{X}_3{}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{X}_3) + \cdots\\ &+ \sigma_p^2 tr(\boldsymbol{X}_p{}'\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-\boldsymbol{X}_p) + \sigma_\epsilon^2 tr(\boldsymbol{X}_{pp}\boldsymbol{X}_{pp}{}^-)\end{aligned} \tag{4.15}$$

이다. σ_ϵ^2을 추정하기 위한 제곱합 $\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$의 기대값은

$$\begin{aligned}E(\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}) &= \sigma_1^2 tr(\boldsymbol{X}_1{}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{X}_1) + \sigma_2^2 tr(\boldsymbol{X}_2{}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{X}_2)\\ &+ \sigma_3^2 tr(\boldsymbol{X}_3{}'(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{X}_3) + \cdots\\ &+ \sigma_p^2 tr(\boldsymbol{X}_p{}'(I-\boldsymbol{X}\boldsymbol{X}^-)(I-\boldsymbol{X}\boldsymbol{X}^{--})\boldsymbol{X}_p) + \sigma_\epsilon^2 tr(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\\ &= \sigma_\epsilon^2 tr(\boldsymbol{I}-\boldsymbol{X}\boldsymbol{X}^-)\end{aligned} \tag{4.16}$$

이다. $p+1$개의 미지모수를 포함하는 연립방정식을 이용하여 분산성분에 대한 추정치를 얻게 된다. 이차형식의 기대값을 계산하기 위한 컴퓨터방법을 Hartley (1967)의 합성법 (synthesis)이라 한다. Hartley의 합성법을 이용하여 분산성분의 계수를 구하게 된다.

5. 자료의 예

불균형자료의 분산성분을 추정하기 위한 자료의 예로 실험단위의 반응 y에 영향을 주는 두 요인으로 A와 B를 가정한다. 실험에 이용되는 각 요인의 수준들이 수준들의 모집단에서 임의로 추출된다고 가정한다. Table 5.1은 요인 A_1의 세 수준이 a_1, a_2, a_3이고 요인 A_2의 세 수준이 b_1, b_2, b_3일 때 이들의 결합수준에서 행해진 실험자료를 나타내고 있다.

Table 5.1의 자료를 분석하기 위한 모형으로 행렬표현식 (2.1)을 이용한다. A와 B가 확률요인일 때 해당하는 행렬표현식은 다음과 같다.

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_3\boldsymbol{\alpha}_3 + \boldsymbol{\epsilon} \tag{5.1}$$

Table 5.1 Unbalanced two-way data

A B	b_1	b_2	b_3
a_1	8,10	12,9,10	13,15
a_2	14,16,18	7,9	15,17
a_3	11,13	9,12	11,8,12

여기서, $\boldsymbol{\alpha}_1$은 요인 A의 임의로 추출된 세 수준효과들의 확률벡터, $\boldsymbol{\alpha}_2$는 요인 B의 임의로 추출된 세 수준효과들의 확률벡터이고 $\boldsymbol{\alpha}_3$는 A와 B의 교호작용을 나타내는 확률효과들의 벡터로 간주된다.

행렬표현식 (2.1)의 분산성분들로 주어지는 σ_1^2, σ_2^2, σ_3^2과 σ_ϵ^2을 추정하기 위해 4절에서 논의된 방법을 활용하기로 한다. 식 (4.12)로부터

$$E(\boldsymbol{y}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{y}) = 14\sigma_1^2 + 0.2857143\sigma_2^2 + 4.857143\sigma_3^2 + 2\sigma_\epsilon^2 \tag{5.2}$$

를 얻게 된다. 식 (4.13)으로부터

$$E(\boldsymbol{y}'\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{y}) = 13.71429\sigma_2^2 + 4.642857\sigma_3^2 + 4\sigma_\epsilon^2 \tag{5.3}$$

를 구한다. 식 (4.14)로부터

$$E(\boldsymbol{y}'\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{y}) = 9.071429\sigma_3^2 + 4\sigma_\epsilon^2 \tag{5.4}$$

를 얻는다. 그리고 식 (4.16)으로부터

$$E(\boldsymbol{y}'(I - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}) = 12\sigma_\epsilon{}^2 \tag{5.5}$$

를 얻는다. 분산성분의 계수행렬을 $\boldsymbol{C}$라 두고 분산성분벡터를 $\boldsymbol{\tau}$라 두자. 여기서 $\boldsymbol{\tau}$는 $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_\epsilon^2)$이다. 제곱합벡터를 $\boldsymbol{q}$라 두면 $\boldsymbol{q}$는 $(\boldsymbol{y}'\boldsymbol{X}_{11}\boldsymbol{X}_{11}{}^-\boldsymbol{y}, \boldsymbol{y}\prime\boldsymbol{X}_{22}\boldsymbol{X}_{22}{}^-\boldsymbol{y}, \boldsymbol{y}\prime\boldsymbol{X}_{33}\boldsymbol{X}_{33}{}^-\boldsymbol{y}, \boldsymbol{y}\prime(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y})'$ 이다. 적률법으로 분산성분을 추정하기 위한 방정식을 행렬로 표현하면

$$\boldsymbol{C\tau} = \begin{pmatrix} 14 & 0.2857143 & 4.857143 & 2 \\ 0 & 13.71429 & 4.642857 & 4 \\ 0 & 0 & 9.071429 & 4 \\ 0 & 0 & 0 & 12 \end{pmatrix} \begin{pmatrix} \mu \\ \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \\ \sigma_\epsilon^2 \end{pmatrix} = \begin{pmatrix} 36.28571 \\ 44.83929 \\ 77.6131 \\ 35.83333 \end{pmatrix} = \boldsymbol{q} \tag{5.6}$$

이다. 연립방정식의 해로 σ_1^2=-0.34519884, σ_2^2=-0.05214219, σ_3^2=7.23906417와 $\sigma_\epsilon^2 = 2.98611083$를 얻게 된다.

6. 결론

본 논문은 확률효과모형의 가정하에 분산성분을 사영에 근거하여 구하는 방법을 다루고 있다. 사영이 관측벡터의 벡터공간에서 정의되므로 모형행렬로 생성되는 사영공간을 분할하는 방법은 다양할 수 있다. 본 논문은 불균형자료에 대한 분산성분의 추정방법으로 ANOVA방법이라 불리우는 Henderson 방법 I을 이용하고 있다. Henderson 방법 I에는 분산성분의 추정을 위해 제곱합과 제곱합의 기댓값이 요구된다. 분산성분의 추정을 위한 제곱합 계산에 모형행렬 $\boldsymbol{X}$로의 사영인 $\boldsymbol{X}\boldsymbol{X}^-\boldsymbol{y}$에 단계별 방법 (stepwise procedure)을 제안하고 있다. 이 방법은 모형행렬 $\boldsymbol{X}$에 의한 사영공간이 상호직교하는

부분공간으로의 사영행렬을 구하는 데 이용될 수 있음을 나타내고 있다. 총제곱합 $\boldsymbol{y}'\boldsymbol{y}$의 분할은 다양한 모형적합방식과 다양한 기법의 사영을 통하여 행해질 수 있다. 그러나 변동요인별 제곱합으로 분해되는 총제곱합의 분할은 제1종 제곱합을 구하는 분산분석법으로 모형의 순차적 적합방식에 따른 잔차제곱합 간의 차를 이용하여 구하거나 해당하는 사영공간의 분할을 통하여 얻어진다. 사영공간을 관심의 직교하는 부분공간들로 분할하기 위해 $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$를 단계별로 확률벡터에 적합시키는 방법 (stepwise procedure)을 구체적으로 논의하고 있다. 또한, 분산성분의 계수를 구하기 위한 Hartley의 합성법과 구해지는 방정식의 구조를 예로써 자세히 설명하고 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.

Corbeil, R. R. and Searle, S. R. (1976). A comparison of variance component estimators. *Biometrics*, **32**, 779−791.

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

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

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.

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

Journal of the Korean Data & Information Science Society
2015, **26**(1), 31–39

http://dx.doi.org/10.7465/jkdi.2015.26.1.31
한국데이터정보과학회지

The analysis of random effects model by projections

Jaesung Choi[1]

[1]Department of Statistics,Keimyung University

Received 4 November 2014, revised 6 December 2014, accepted 15 December 2014

Abstract

This paper deals with a method for estimating variance components on the basis of projections under the assumption of random effects model. It discusses how to use projections for getting sums of squares to estimate variance components. The use of projections makes the vector subspace generated by the model matrix to be decomposed into subspaces that are orthogonal each other. To partition the vector space by the model matrix stepwise procedure is used. It is shown that the suggested method is useful for obtaining Type I sum of squares requisite for the ANOVA method.

Keywords: Projection, random effects, stepwise procedure, Type I sums of squares.

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