Journal of the Korean Data & Information Science Society
2014, **25**(2), 423–429

http://dx.doi.org/10.7465/jkdi.2014.25.2.423
한국데이터정보과학회지

가중최소제곱법에 의한 제1종 사영제곱합

최재성[1]

[1]계명대학교 통계학과

접수 2014년 2월 24일, 수정 2014년 3월 11일, 게재확정 2014년 3월 19일

요 약

본 논문은 이원고정효과모형의 분산분석에서 오차의 독립성과 등분산성이 만족되지 않는 경우를 가정하고 있다. 자료분석을 위한 모수추정방법으로 가중최소제곱법을 가정하고 있으며 모수를 추정하기 위한 방법으로 모형의 순차적 적합방식을 이용하고 있다. 또한, 모형의 행렬표현식으로부터 벡터공간에서의 사영을 이용하여 자료를 분석하는 방법을 제시하고 있다. 모형의 순차적 적합에 해당하는 제1종 제곱합을 구하기 위하여 모형행렬에 의한 부분공간으로의 사영을 다루고 있다. 이 경우에 사영에 의한 제곱합을 사영제곱합으로 취급한다.

주요용어: 가중최소제곱법, 사영제곱합, 순차적 적합, 제1종 제곱합.

1. 서론

처리들의 비교를 위한 실험 자료를 분석하기 위해 선형모형을 가정할 때 자료의 행렬표현에서 주어지는 모형행렬은 일반적으로 불완전계수의 모형행렬이다. 불완전계수의 모형행렬을 계수행렬로 갖는 모수벡터의 추정값은 유일하게 결정되지 않고 무수히 많은 해를 갖게 된다. 선형모형의 가정하에 모형내 모수의 추정방법은 주로 최소제곱법에 근거하고 있다. 최소제곱법은 오차의 독립성과 분산의 동일성을 가정한다. 즉, $E(\boldsymbol{\epsilon}) = \mathbf{0}$이고 $Var(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$를 가정한다.

선형모형의 가정하에 모수를 추정하는 방법이 결정되면 실험분석과 관련된 처리들의 효과를 알아보기 위한 모형의 적합방식은 다양하다. 즉, 다양한 방법으로 모형을 적합시켜 처리효과의 추정값을 얻을 수 있다. 이 경우에 모형의 적합방식에 따른 여러 유형의 제곱합을 구할 수 있다. 이들 유형의 제곱합을 각기 제1종 제곱합, 제2종 제곱합, 제3종 제곱합 등으로 분류하게 된다.

모형의 행렬표현식으로부터 모형행렬에 의해 생성되는 벡터 부분공간으로의 사영을 생각할 수 있다. 모형행렬로 생성되는 벡터 부분공간은 모형적합 방식에 따라 상호직교하는 벡터 부분공간들로 분할되거나 그렇지 않거나 이다. 추정하고자하는 모수와 관련된 벡터공간으로의 사영을 구한 다음에 얻게 되는 사영까지의 거리제곱합은 그 모수와 관련된 변동량이 된다.

고정효과의 선형모형에서 모형의 적합방식에 따른 여러 유형의 제곱합에 관한 논의는 Milliken과 Johnson (1984) 등에서 살펴볼 수 있다. 또한, 선형모형의 가정하에 추정가능함수에 대한 정의와 추론에 대한 논의는 Searle (1971), Graybill (1976) 그리고 Milliken과 Johnson (1984) 등에서 살펴볼 수 있다.

선형모형의 행렬식표현에서 모형행렬이 완전계수의 열행렬이 아닌 경우에 모형내 미지모수들을 추정하는 방법으로 일부 모수들을 영으로 두는 제약식을 이용하여 모형행렬의 계수에 해당하는 모수들을 추

[1] (704-701) 대구광역시 달서구 달구벌대로 1095, 계명대학교 통계학과, 교수. E-mail: jschoi@kmu.ac.kr

정하거나 또는 모수들의 합을 영으로 두는 제약식을 이용하여 추정하게 된다. 이러한 제약식은 모수의 유일 해를 얻기 위한 한 방법으로 이용된다.

본 논문은 고정효과모형의 가정하에 자료의 행렬식 표현으로부터 오차벡터에 대한 독립성과 등분산성의 가정이 성립하지 않는 $E(\boldsymbol{\epsilon}) = \mathbf{0}$이고 $Var(\boldsymbol{\epsilon}) = \Sigma$인 일반적인 경우를 가정한다. 따라서 모수추정방법으로 최소제곱법이 아닌 가중최소제곱법을 가정한다. 가중최소제곱법을 적용하여 모형의 적합방식에 따른 고정효과의 정도를 나타내는 제1종 제곱합을 논의하고자 한다. 제1종 제곱합은 가정된 모형내 모수들의 순차적 적합방식으로부터 계산되는 제곱합이다. 모형의 순차적 적합과 관련된 벡터공간으로의 사영과 사영에 의한 제1종 제곱합을 제공하고자 한다. 벡터공간에서의 사영에 대한 기본적 개념 및 제반성질이 Johnson과 Wichern (1988) 그리고 Graybill (1976) 등에서 자세히 논의된다.

2. 모형의 가정

개체의 반응을 나타내는 변수를 y라 둔다. 반응변수 y에 영향을 주는 요인으로 두 요인 A와 B를 가정한다. 요인 A는 a개의 수준을 갖고 요인 B는 b개의 수준을 갖는다 하자. 요인 A와 B의 두 수준결합 (i,j)에서 k번째 관측값을 y_{ijk}라 둔다. $i = 1,2,\cdots,a$, $j = 1,2,\cdots,b$이고 $k = 1,2,\cdots,l$이다. 자료의 개수를 n이라 두면 $n = abl$이다. 요인 A의 수준 i의 효과를 α_i, 요인 B의 수준 j의 효과를 β_j, (i,j)에서 두 수준의 교호작용에 따른 효과를 $(\alpha\beta)_{ij}$라 두자. 관측값 y_{ijk}에 대한 고정효과모형은

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \tag{2.1}$$

이다. 단, $i = 1,2,\cdots,a$, $j = 1,2,\cdots,b$, $k = 1,2,\cdots,l$이며 오차들은 독립이 아닐 수도 있고 $V(\epsilon_{ijk}) = \sigma_{ij}^2$으로 가정한다. 단일 관측값에 대한 고정효과모형은 자료의 행렬표현으로부터 다음과 같이 표현된다.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{\epsilon} \tag{2.2}$$

단, $\boldsymbol{y}$는 $n \times 1$벡터, $\boldsymbol{X}$는 계수가 $r\ (< p)$인 $n \times p$의 모형행렬이고 $\boldsymbol{\tau}$는 $p \times 1$인 모수벡터이며 $\boldsymbol{\epsilon}$은 $n \times 1$인 오차벡터이다. 오차벡터에 대한 가정은 $N(\mathbf{0}, \sigma^2 V)$이다. V는 양정치행렬로 주어진다. 식 (2.1)에 해당하는 모형행렬 $\boldsymbol{X}$와 모수벡터 $\boldsymbol{\tau}$의 형태를 살펴보면 $\boldsymbol{X} = (\boldsymbol{j},\ \boldsymbol{X}_A,\ \boldsymbol{X}_B,\ \boldsymbol{X}_{AB})$이고 $\boldsymbol{\tau}' = (\mu,\ \boldsymbol{\alpha}',\ \boldsymbol{\beta}',\ (\boldsymbol{\alpha\beta})')$으로 표시됨을 알 수 있다. 따라서 식 (2.2)는

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{X}_{AB}(\boldsymbol{\alpha\beta}) + \boldsymbol{\epsilon} \tag{2.3}$$

와 같이 표현된다. $\boldsymbol{j}$는 원소가 1인 $n \times 1$벡터이다. 자료의 행렬표현으로부터 $\boldsymbol{y}$는 다차원 공간상의 벡터로 간주되며 성분벡터들의 합으로 구해짐을 알 수 있다. 그러나 오차벡터의 성분들이 상관성을 가지며 분산이 동일하지 않다는 가정 때문에 각 부분공간으로의 사영과 사영까지의 거리 제곱합을 계산할 수가 없게 된다.

오차의 독립성과 분산의 동일성이 만족될 때, 고정효과모형내 모수의 추정을 위한 최소제곱법은 기하학적 측면에서 $\boldsymbol{y}$를 모형행렬 $\boldsymbol{X}$로의 직교사영을 의미한다. 왜냐하면 $\boldsymbol{X}$에 의해 생성되는 부분공간과 오차벡터 $\boldsymbol{\epsilon}$에 의해 생성되는 벡터공간의 직교성 때문이다. 이때 사영까지의 거리제곱합은 모수벡터에 기인한 변동량을 나타내므로 분산분석에 이용된다.

3. 가중최소제곱법에 의한 변동량

식 (2.3)의 오차벡터 $\boldsymbol{\epsilon}$에 대한 가정으로 부터 최소제곱법이 아닌 가중최소제곱법이 이용된다. 가중최소제곱법은 $\boldsymbol{\epsilon}'V^{-1}\boldsymbol{\epsilon}$를 최소로 하는 모수추정방법이다. 분산공분산 행렬 V가 양정치이므로 V^{-1}도 양정

치이다. 따라서 V^{-1}가 $\Lambda'\Lambda$로 표현되는 정칙행렬 Λ를 이용하여 $\boldsymbol{y}$를 선형변환하면 식 (2.2)는

$$\Lambda \boldsymbol{y} = \Lambda \boldsymbol{X\tau} + \Lambda \boldsymbol{\epsilon} \tag{3.1}$$

인 모형을 얻는다. $\boldsymbol{z} = \Lambda\boldsymbol{y}$, $\boldsymbol{Q} = \Lambda\boldsymbol{X}$, 그리고 $\boldsymbol{\epsilon}_\Lambda = \Lambda\boldsymbol{\epsilon}$으로 두면

$$\boldsymbol{z} = \boldsymbol{Q\tau} + \boldsymbol{\epsilon}_\Lambda \tag{3.2}$$

로 표현된다. 여기서 $\boldsymbol{\epsilon}_\Lambda$는 $N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$이다. 식 (2.2)에서의 가중최소제곱법은 변환된 모형식 (3.2)에서의 최소제곱법과 동일하다. 왜냐하면

$$\boldsymbol{\epsilon}' V^{-1}\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}\tau)' V^{-1}(\boldsymbol{y} - \boldsymbol{X}\tau) = \boldsymbol{\epsilon}_\Lambda'\boldsymbol{\epsilon}_\Lambda \tag{3.3}$$

이기 때문이다. $\boldsymbol{j}_\Lambda = \Lambda\boldsymbol{j}$, $\boldsymbol{Q}_A = \Lambda\boldsymbol{X}_A$, $\boldsymbol{Q}_B = \Lambda\boldsymbol{X}_B$, $\boldsymbol{Q}_{AB} = \Lambda\boldsymbol{X}_{AB}$라 두면

$$\boldsymbol{z} = \boldsymbol{j}_\Lambda\mu + \boldsymbol{Q}_A\boldsymbol{\alpha} + \boldsymbol{Q}_B\boldsymbol{\beta} + \boldsymbol{Q}_{AB}(\boldsymbol{\alpha\beta}) + \boldsymbol{\epsilon}_\Lambda \tag{3.4}$$

로 변환된다. 변환된 모형식 (3.2)에 최소제곱법의 적용은 $\boldsymbol{Q}'\boldsymbol{Q}\hat{\boldsymbol{\tau}} = \boldsymbol{Q}'\boldsymbol{z}$인 정규방정식으로 주어진다. 정규방정식으로부터 $\hat{\boldsymbol{\tau}} = (\boldsymbol{Q}'\boldsymbol{Q})^-\boldsymbol{Q}'\boldsymbol{z} = (\boldsymbol{X}'V^{-1}\boldsymbol{X})^-\boldsymbol{X}'V^{-1}\boldsymbol{y}$을 얻게 된다. $\hat{\boldsymbol{z}} = \boldsymbol{Q}\hat{\boldsymbol{\tau}} = \boldsymbol{QQ}^-\boldsymbol{z}$로 표현되므로 벡터공간에서 관측벡터 $\boldsymbol{z}$를 모형행렬 $\boldsymbol{Q}$로의 직교사영으로 간주할 수 있다. $\boldsymbol{Q}^-$는 Moore-Penrose의 일반화 역행렬이다.

모형식 (3.4)를 이용하면 요인별 제곱합을 사영에 근거하여 사영까지의 거리제곱합으로 구할 수 있다. Choi (2010, 2011)은 반복측정의 분할구 실험자료의 분석을 위한 혼합모형과 일원 확률모형의 분산성분을 사영에 의해 추론하는 방법을 다루었다. 그리고 Choi (2012)는 이원 고정효과모형에서 고정효과의 변동량을 구하는 방법으로 사영에 근거한 제2종 분석을 제시하고 있다.

모형식 (3.4)를 이용한 제1종 제곱합은 벡터공간에서 변환벡터 $\boldsymbol{z}$를 모형행렬 $\boldsymbol{Q}$로의 사영으로부터 구해진다. 즉, 모형행렬 $\boldsymbol{Q}$에 의해 생성된 벡터 부분공간이 상호 직교하는 벡터 부분공간으로 구성될 때 각 부분 벡터공간으로의 사영은 그 공간에서의 모수벡터에 기인하는 변동량을 제공하게 된다. 고정효과의 크기와 관련된 변동량을 계산하기 위한 모형의 순차적 적합방식이 벡터공간에서 상호 직교하는 부분 벡터공간으로의 한 순차적 사영을 의미한다는 점에 착안하여 동일한 변동량의 계산이 가능함을 논의하고 있다.

제1종 제곱합에 해당하는 모형적합방식은 고정효과모형내 모수의 순서대로 모수를 적합하게 된다. 행렬모형식 (3.4)에서 모형내 모수의 순서는 μ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $(\boldsymbol{\alpha}\beta)$의 순이다. 이들 모수와 관련된 벡터 부분공간으로의 사영을 생각한다. 모수 μ와 관련된 계수벡터는 $\boldsymbol{j}_\Lambda$이므로 $\boldsymbol{QQ}^-\boldsymbol{z}$를 $\boldsymbol{j}_\Lambda$로의 사영은 $\boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-\boldsymbol{z}$이고 사영까지의 제곱거리는 $\boldsymbol{z}'\boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-\boldsymbol{z}$가 된다. $\boldsymbol{y}$로 표현하면

$$\boldsymbol{z}'\boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-\boldsymbol{z} = \boldsymbol{y}'V^{-1}\boldsymbol{j}(\boldsymbol{j}'V^{-1}\boldsymbol{j})^{-1}\boldsymbol{j}'V^{-1}\boldsymbol{y} \tag{3.5}$$

이다. 모수벡터 $\boldsymbol{\alpha}$의 계수행렬은 $\boldsymbol{Q}_A$이다. $\boldsymbol{QQ}^-\boldsymbol{z}$에서 $\boldsymbol{j}_\Lambda$로의 사영 $\boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-\boldsymbol{z}$에 적합된 직교성분은 $(\boldsymbol{QQ}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-)\boldsymbol{z}$이다. $\boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-$와 $(\boldsymbol{QQ}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-)$에 의해 생성되는 벡터공간들은 직교한다. $\boldsymbol{Q}_1 = [(\boldsymbol{QQ}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-)\boldsymbol{Q}_A]$로 두자. $(\boldsymbol{QQ}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^-)\boldsymbol{z}$를 $\boldsymbol{Q}_A$로의 사영은 $\boldsymbol{Q}_1\boldsymbol{Q}_1^-\boldsymbol{z}$이다. 사영까지의 제곱거리는 $\boldsymbol{z}'\boldsymbol{Q}_1\boldsymbol{Q}_1^-\boldsymbol{z}$이다. $\boldsymbol{Q}_1$을 Λ, $\boldsymbol{X}$, $\boldsymbol{X}_A$로 나타내면

$$\boldsymbol{Q}_1 = \Lambda[\boldsymbol{X}(\boldsymbol{X}'V^{-1}\boldsymbol{X})^-\boldsymbol{X}' - \boldsymbol{j}(\boldsymbol{j}'V^{-1}\boldsymbol{j})^{-1}\boldsymbol{j}']V^{-1}\boldsymbol{X}_A \tag{3.6}$$

이다. $\boldsymbol{Q}_1$의 일반화된 역행렬 $\boldsymbol{Q}_1^-$는 다음과 같이 구해진다.

$$\boldsymbol{Q_1}^- = \Big(\boldsymbol{X_A}'V^{-1}[\boldsymbol{X}(\boldsymbol{X}'V^{-1}\boldsymbol{X})^{\boldsymbol{-}}\boldsymbol{X}' - \boldsymbol{j}(\boldsymbol{j}'V^{-1}\boldsymbol{j})^{-1}\boldsymbol{j}']V^{-1}\boldsymbol{X_A}\Big)^- \tag{3.7}$$

제곱거리를 $\boldsymbol{y}$로 나타내면

$$\boldsymbol{z}'\boldsymbol{Q_1}\boldsymbol{Q_1^-}\boldsymbol{z} = \boldsymbol{y}'\Lambda'\boldsymbol{Q_1}\boldsymbol{Q_1^-}\Lambda\boldsymbol{y} \tag{3.8}$$

로 구해진다. 모수벡터 $\boldsymbol{\beta}$의 계수행렬은 $\boldsymbol{Q}_B$이다. $(\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^- - \boldsymbol{Q}_1\boldsymbol{Q}_1^-)\boldsymbol{z}$를 $\boldsymbol{Q}_B$로의 사영을 생각해 보자. $\boldsymbol{Q}_2$를 $\boldsymbol{Q}_2 = (\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^- - \boldsymbol{Q}_1\boldsymbol{Q}_1^-)\boldsymbol{Q}_B$라 두면 사영은 $\boldsymbol{Q}_2\boldsymbol{Q}_2^-\boldsymbol{z}$이다. 제곱거리를 $\boldsymbol{y}$로 나타내면

$$\boldsymbol{z}'\boldsymbol{Q}_2\boldsymbol{Q}_2^-\boldsymbol{z} = \boldsymbol{y}'\Lambda'\boldsymbol{Q}_2\boldsymbol{Q}_2^-\Lambda\boldsymbol{y} \tag{3.9}$$

로 구해진다.

식 (3.4)에서 모수들의 순서대로 교호작용 $(\boldsymbol{\alpha}\beta)$의 모형적합과정으로부터 계수행렬 $\boldsymbol{Q}_{AB}$로의 사영에 이용되는 벡터는 $(\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^- - \boldsymbol{Q}_1\boldsymbol{Q}_1^- - \boldsymbol{Q}_2\boldsymbol{Q}_2^-)\boldsymbol{z}$로 주어짐을 확인할 수 있다. $\boldsymbol{Q}_3$를 $\boldsymbol{Q}_3 = (\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda\boldsymbol{j}_\Lambda^- - \boldsymbol{Q}_1\boldsymbol{Q}_1^- - \boldsymbol{Q}_2\boldsymbol{Q}_2^-)\boldsymbol{Q}_{AB}$로 둘 때 $\boldsymbol{Q}_{AB}$로의 사영은 $\boldsymbol{Q}_3\boldsymbol{Q}_3^-\boldsymbol{z}$로 구해진다. 사영까지의 거리제곱을 $\boldsymbol{y}$로 나타내면

$$\boldsymbol{z}'\boldsymbol{Q}_3\boldsymbol{Q}_3^-\boldsymbol{z} = \boldsymbol{y}'\Lambda'\boldsymbol{Q}_3\boldsymbol{Q}_3^-\Lambda\boldsymbol{y} \tag{3.10}$$

이다. 고정효과모형의 가정하에 최소제곱법이 적용될 수 없는 조건에서 대수적 방법으로 요인에 따른 변동량의 계산은 쉽지가 않다. 이러한 문제를 해결하기 위한 방법으로 자료의 행렬표현으로부터 벡터공간에서 사영의 개념을 이용하면 요인별 변동량을 상대적으로 간편하게 구할 수 있음을 제시하고 있다.

4. 자료의 예

개체의 반응에 영향을 주는 두 요인 A와 B의 수준결합에서 실험이 행해졌다고 가정한다. 요인 A는 a_1, a_2의 두 수준이고 요인 B는 b_1, b_2, b_3의 세 수준이라 하자. 요인들의 수준결합으로 주어지는 처리가 동수의 개체들에 임의로 배정되어 수집된 자료가 Table 4.1에 나타나 있다.

Table 4.1 Two-way experimental data

A	B		
	b_1	b_2	b_3
a_1	19.0, 17.0, 18.0	20.2, 17.8, 19.0	16.0, 14.0, 15.0
a_2	14.5, 12.5, 13.0	15.6, 12.4, 14.0	11.0, 9.0, 10.0

Table 4.1의 자료분석을 위한 행렬모형식으로

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.1}$$

을 가정한다. 이는 행렬모형식 (2.3)에서 교호작용 항을 제외한 모형이다. 오차벡터에 대한 가정은 $N(\boldsymbol{0}, \sigma^2 V)$이고 V는 대각원소가 같지 않는 대각행렬이라고 가정한다. V의 i번째 대각원소를 d_i, $i = 1, 2, \cdots, 18$로 두자. 이 때 d_i의 값은 다음과 같다고 가정한다.

$$\begin{aligned} d_1 = d_2 = d_3 = 1,\ d_4 = d_5 = d_6 = 1.44,\ d_7 = d_8 = d_9 = 1, \\ d_{10} = d_{11} = d_{12} = 1,\ d_{13} = d_{14} = d_{15} = 2.56,\ d_{16} = d_{17} = d_{18} = 1 \end{aligned} \tag{4.2}$$

오차벡터 $\boldsymbol{\epsilon}$의 가정으로부터 식 (2.2)의 모수벡터 $\boldsymbol{\tau}$의 추정방법으로 가중최소제곱법을 적용하게 된다. 가중분산을 나타내는 양정치행렬 V로부터 $V^{-1} = \Lambda'\Lambda$되는 Λ를 이용하여 $\Lambda\boldsymbol{y}$를 구한다.

$\Lambda \boldsymbol{y} = (19, 17, 18, 16.83, 14.83, 15.83, 16, 14, 15, 14.5, 12.5, 13, 9.75, 7.75, 8.75, 11, 9, 10)$이다.
식 (4.1)의 양변에 Λ를 곱하여 $\boldsymbol{j}_\Lambda = \Lambda \boldsymbol{j}$, $\boldsymbol{Q}_A = \Lambda \boldsymbol{X}_A$, $\boldsymbol{Q}_B = \Lambda \boldsymbol{X}_B$, $\boldsymbol{z} = \Lambda \boldsymbol{y}$로 두면

$$\boldsymbol{z} = \boldsymbol{j}_\Lambda \mu + \boldsymbol{Q}_A \boldsymbol{\alpha} + \boldsymbol{Q}_B \boldsymbol{\beta} + \boldsymbol{\epsilon}_\Lambda \tag{4.3}$$

로 변환된다. 관측벡터 $\boldsymbol{y}$의 벡터공간에서 제1종 제곱합에 해당하는 변동량을 구하기 위해 사영의 개념을 이용한다. 사영의 이용은 벡터공간에서의 행렬을 이용하므로 자료표현의 단순화와 계산의 편리성 및 벡터공간의 여러 유용한 성질을 활용할 수 있다는 점에서 자료분석의 단순화를 유도할 수 있는 이점이 있다.

식 (4.2)에서 모형행렬 $\boldsymbol{Q} = (\boldsymbol{j}_\Lambda, \boldsymbol{Q}_A, \boldsymbol{Q}_B)$에 의해 생성되는 벡터공간에서 $\boldsymbol{j}_\Lambda$로의 사영은 $\boldsymbol{j}_\Lambda \boldsymbol{j}_\Lambda^- \boldsymbol{z}$이고 $\hat{\mu} = 14.75$이다. 사영까지의 제곱거리는 $\boldsymbol{z}' \boldsymbol{j}_\Lambda \boldsymbol{j}_\Lambda^- \boldsymbol{z} = 3318.23$이다. 모수벡터 $\boldsymbol{\alpha}$를 추정하기 위한 벡터공간은 $\boldsymbol{j}_\Lambda$에 의해 생성된 벡터공간과 직교하는 공간이다. 이 공간은 행렬 $\boldsymbol{Q}_1 = [(\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda \boldsymbol{j}_\Lambda^-)\boldsymbol{Q}_A]$의 열벡터에 의해 생성되므로 사영은 $\boldsymbol{Q}_1 \boldsymbol{Q}_1^- \boldsymbol{z}$가 된다.

$\widehat{\boldsymbol{\alpha}}' = (2.55, -2.55)$로 추정되고 사영까지의 제곱거리는 $\boldsymbol{z}' \boldsymbol{Q}_1 \boldsymbol{Q}_1^- \boldsymbol{z} = 98.70$이다. 모수벡터 $\boldsymbol{\beta}$를 추정하기 위해 $\boldsymbol{z}$를 $\boldsymbol{Q}_2 = (\boldsymbol{Q}\boldsymbol{Q}^- - \boldsymbol{j}_\Lambda \boldsymbol{j}_\Lambda^- - \boldsymbol{Q}_1 \boldsymbol{Q}_1^-)\boldsymbol{Q}_B$로의 사영을 이용한다. $\boldsymbol{Q}_2$로의 사영은 $\boldsymbol{Q}_2 \boldsymbol{Q}_2^- \boldsymbol{z}$이다. $\widehat{\boldsymbol{\beta}}' = (0.77, 1.62, -2.39)$이고 사영까지의 거리제곱은 $\boldsymbol{z}' \boldsymbol{Q}_2 \boldsymbol{Q}_2^- \boldsymbol{z} = 45.07$이다. $\boldsymbol{z}$를 $\boldsymbol{Q}$로의 사영은 $\boldsymbol{Q}\boldsymbol{Q}^- \boldsymbol{z}$이므로 사영까지의 거리제곱은 $\boldsymbol{z}' \boldsymbol{Q}\boldsymbol{Q}^- \boldsymbol{z} = 3462.00$이다.

식 (4.2)에 근거하여 상호직교하는 벡터 부분공간으로의 사영은 각 벡터 부분공간에서의 모수추정값을 제공할 뿐만 아니라 제1종 제곱합에 해당하는 사영까지의 거리제곱합을 제공하고 있음을 알 수 있다. 즉,

$$\boldsymbol{z}' \boldsymbol{Q}\boldsymbol{Q}^- \boldsymbol{z} = \boldsymbol{z}' \boldsymbol{j}_\Lambda \boldsymbol{j}_\Lambda^- \boldsymbol{z} + \boldsymbol{z}' \boldsymbol{Q}_1 \boldsymbol{Q}_1^- \boldsymbol{z} + \boldsymbol{z}' \boldsymbol{Q}_2 \boldsymbol{Q}_2^- \boldsymbol{z} \tag{4.4}$$

임을 나타낸다. 관측벡터 $\boldsymbol{y}$가 공분산행렬을 갖는 벡터공간에서의 벡터일 때, 양정치인 공분산행렬의 성질로부터 직교공간의 벡터 $\boldsymbol{z}$로 변환하여 자료분석에 이용할 수 있음을 보여주고 있다.

5. 결론

본 논문은 이원 고정효과모형의 가정에서 요인들의 수준효과에 따른 변동량의 분석을 다루고 있다. 선형모형의 가정하에 개체의 반응값이 동일한 분산을 갖지 않는 경우의 분석방법으로 벡터공간에서 사영을 이용하는 방법을 제시하고 있다.

처리간 실험자료의 변동량이 동일하지 않으므로 모형내 모수의 추정방법으로 가중최소제곱법을 이용하고 있으며 가중최소제곱법에 의한 모수의 추정을 위해 모형의 순차적 적합을 적용하고 있다. 자료의 행렬표현으로부터 벡터공간에서 사영이 어떻게 활용되어야 하는 가를 구체적 논의를 통해서 그 과정을 설명하고 있다. 즉, 모형의 순차적 적합과정에 해당하는 벡터 부분공간들의 직교화와 관련된 행렬들이 어떻게 구해지는 가를 다루고 있다.

관심모수에 따른 변동량은 모수의 추정과 관련된 사영공간으로의 사영이 취해졌을 때 그 사영까지의 거리제곱임을 보여주고 있다. 또한, 관측벡터 $\boldsymbol{y}$가 공분산을 갖는 벡터공간으로부터의 벡터로 간주될 때 직교하는 벡터공간의 벡터로 변환시켜 사영을 이용하게 되면 분산분석의 간편성뿐만 아니라 행렬의 성질과 벡터공간의 성질들을 유용하게 이용할 수 있음을 나타내고 있다.

References

Choi, J. S. (2010). A mixed model for repeated split-plot data. *Journal of the Korean Data & Information Science Society*, **21**, 1-9.

Choi, J. S. (2011). Variance components in one-factor random model by projections. *Journal of the Korean Data & Information Science Society*, **22**, 381-387.

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

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

Johnson, R. A. and Wichern, D. W. (1988). *Applied multivariate statistical analysis*, 2nd edition, Prentice Hall, Inc., Englewood Cliffs.

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

Searle, S. R. (1971). *Linear models*, John Wiley and Sons, Inc., New York.

Journal of the Korean Data & Information Science Society
2014, **25**(2), 423–429

http://dx.doi.org/10.7465/jkdi.2014.25.2.423
한국데이터정보과학회지

Type I projection sum of squares by weighted least squares

Jaesung Choi[1]

[1]Department of Statistics, Keimyung University

Received 24 February 2014, revised 11 March 2014, accepted 19 March 2014

Abstract

This paper discusses a method for getting Type I sums of squares by projections under a two-way fixed-effects model when variances of errors are not equal. The method of weighted least squares is used to estimate the parameters of the assumed model. The model is fitted to the data in a sequential manner by using the model comparison technique. The vector space generated by the model matrix can be composed of orthogonal vector subspaces spanned by submatrices consisting of column vectors related to the parameters. It is discussed how to get the Type I sums of squares by using the projections into the orthogonal vector subspaces.

Keywords: Projection sum of squares, sequential manner, Type I sum of squares, weighted least squares.

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