Journal of the Korean Data & Information Science Society
2014, **25**(3), 547–554
http://dx.doi.org/10.7465/jkdi.2014.25.3.547
한국데이터정보과학회지

이원 분산성분의 사영분석

최재성[1]

[1]계명대학교 통계학과

접수 2014년 4월 2일, 수정 2014년 4월 19일, 게재확정 2014년 5월 1일

요 약

본 논문은 실험자료에 대한 분석모형으로 이원 분산분석모형을 가정한다. 확률효과 모형의 가정하에 분산성분의 추정량을 구하기 위한 방법으로 적률법을 가정하고 있다. 분산성분의 적률 추정방법인 Henderson의 방법 I과 방법 III을 다루고 있다. Henderson의 두 방법에서 소개되는 제곱합 대신에 벡터공간에서의 사영을 활용하는 방법을 제시하고 있다. 또한 제곱합의 기대값 계산을 위해 두 방법 모두 Hartley의 합성법을 제공하고 있으나 본 논문에서는 관련행렬의 고유근을 이용할 수 있음을 제시하고 있다. 분산성분의 해를 얻기 위한 방법의 차이에서 유도되는 연립방정식들은 같지 않으나 양수의 분산성분들에 대한 해는 유사함을 보여주고 있다.

주요용어: 고유근, 분산성분, 사영, 이차형식, 확률효과.

1. 서론

실험에 이용되는 처리가 요인들의 수준결합으로 주어지는 경우를 가정해 본다. 실험에서 고려되는 처리요인들이 모두 확률요인일 때 처리조합은 이들 요인의 수준결합으로 구성된다. 확률요인의 의미는 그 요인의 수준들의 모집단에서 수준이 임의로 추출됨을 의미한다. 임의로 추출된 수준의 효과는 고정효과가 아닌 확률효과로 간주되고 확률변수로 취급된다. 따라서 확률효과의 분포로 일반적으로 $N(0, \sigma^2)$을 가정하게 된다. 실험에 이용되는 처리조합이 확률요인들의 수준결합으로 구성되고 실험이 행해지는 경우를 생각한다. 실험에서 얻어진 자료를 분석하기 위한 선형모형으로 ANOVA 모형을 가정할 때 분산분석모형은 확률모형 또는 확률효과 모형으로 간주된다. 확률효과 모형의 가정하에서 자료를 분석할 때 관심모수들은 분산성분의 추론에 있다.

선형모형의 가정하에서 분산성분을 추론하는 방법들에 관한 논의는 Graybill (1976), Milliken과 Johnson (1984) 그리고 Searl 등 (1992)의 문헌에서 찾아볼 수 있다. 관심모수인 분산성분에 대한 추론방법은 적률법, 최대우도법과 Rao (1971)의 MINQUE방법 등을 이용할 수 있다. 최대우도법과 MINQUE방법은 계산이 힘들고 일반적으로 연산알고리즘을 이용해야 하는 어려움이 있는 반면에 적률법은 상대적으로 계산이 간단하게 행해지는 이점이 있다. 본 연구에서는 분산성분의 추정을 위한 방법으로 적률법에 의한 추정방법을 생각하고 있다. 적률법으로 분산성분을 추정할 때 요구되는 두 가지 계산은 변동요인에 따른 제곱합의 계산과 평균평방합의 기대값에 대한 계산이 필요하게 된다. 변동요인에 따른 제곱합의 계산을 위해 Henderson (1953)의 방법 I인 ANOVA 방법 또는 Henderson의 방법 III으로 불리우는 상수적합방법 (fitting constants method)을 적용할 수 있다. Henderson (1953)의 방법 I은 제곱합의 계산을 위해 분산분석법을 이용하고 있다. 처리간에 관측수가 동일한 균형자료인 경우에

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

주로 ANOVA 방법을 이용하나 불균형자료의 경우 여러 가지 분석방법들이 이용된다. Searle (1971)은 균형자료의 경우에는 단일의 분산분석이 가능하나 불균형자료인 경우에는 다양한 유형의 분산분석이 가능하게 되어 분산의 추정값도 같지 않게 됨을 기술하고 있다. 본 논문에서는 Henderson의 방법 I 또는 방법 III에서 이용되는 제곱합 계산방식과는 달리 사영 (projection)에 의해 제곱합을 구하는 방법을 논의한다. 사영을 이용하여 자료를 분석하는 방법들에 관한 논의는 Choi (2011, 2012a, 2012b, 2013)에서 살펴볼 수 있다. 평균평방합의 기대값 계산을 위한 방법으로는 확률모형의 가정을 이용한 대수적방법과 Hartley (1967)의 합성법 (synthesis)이 있다. 본 연구에서는 분산성분의 계수계산을 위한 합성법 대신에 이차형식과 관련된 행렬의 고유근을 이용하는 방법을 다루고자 한다.

2. 이원 확률모형의 가정

실험의 처리조합을 구성하는 두 확률요인을 A와 B라 두자. 요인 A는 수준들의 모집단에서 임의로 추출된 $i = 1, 2, \cdots, a$개의 수준들로 구성되고 요인 B는 모집단에서 임의로 추출된 $j = 1, 2, \cdots, b$개의 수준들로 구성된다고 가정한다. 따라서, 요인 A의 수준 i와 요인 B의 수준 j와의 수준결합으로 주어지는 처리조합을 (i, j)로 나타낸다. 실험에 이용되는 실험단위들은 동질적이라 가정한다. 처리조합 (i, j)는 n_{ij}개의 실험단위에 임의로 배정된다고 가정한다. 처리조합 (i, j)가 행해진 실험단위에서의 관측반응을 y_{ijk}라 두자. 처리조합을 구성하는 두 요인이 확률요인이므로 실험단위의 관측반응에 대한 모형으로 확률효과 모형을 가정한다. 요인 A의 수준 i에서의 수준효과를 α_i라 두면, a개의 α_i들은 확률효과들 이다. 요인 B의 수준 j에서의 수준효과를 β_j라 두면, b개의 β_j들도 확률효과를 나타낸다. 요인 A의 수준 i, 요인 B의 수준 j의 처리조합 (i, j)에서의 확률효과들의 교호작용을 $(\alpha\beta)_{ij}$로 나타낸다. 자료를 분석하기 위한 확률모형은 다음과 같이 주어진다.

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

단, μ는 전체평균을 나타낸다. 요인들의 수준이 각 모집단에서 임의로 추출되므로 요인 A의 수준효과 $\alpha_i, i = 1, 2, \cdots, a$는 상호독립이고 $N(0, {\sigma_\alpha}^2)$을 따른다고 가정한다. 요인 B의 수준효과 $\beta_j, j = 1, 2, \cdots, b$의 분포로 $N(0, {\sigma_\beta}^2)$을 가정하고 효과들은 상호독립이라고 가정한다. ab개의 교호작용 $(\alpha\beta)_{ij}$는 상호독립이고 $N(0, {\sigma_{\alpha\beta}}^2)$인 분포로 부터의 확률표본으로 가정한다. 오차 ϵ_{ijk}, $k = 1, 2, \cdots, n_{ij}$에 대한 분포로 $N(0, {\sigma_\epsilon}^2)$을 가정하고 오차들은 상호독립이라 가정한다. 확률모형의 가정하에 추정해야 할 모수는 모평균 μ와 분산성분을 나타내는 σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$과 σ^2_ϵ이다. ab개의 처리조합에서 수집된 자료를 다차원상의 공간벡터를 $\boldsymbol{y}$라 두자. $\boldsymbol{y}$의 크기는 $n \times 1$이다. $n = \sum_{ij} n_{ij}$이다. 식 (2.1)에 해당하는 다차원상의 공간벡터 $\boldsymbol{y}$에 대한 모형식은 다음과 같이 주어진다.

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{X}_{AB}(\boldsymbol{\alpha\beta}) + \boldsymbol{\epsilon} \tag{2.2}$$

단, 모평균 μ의 계수벡터 $\boldsymbol{j}$는 n개의 원소가 모두 1인 열벡터이다. $\boldsymbol{X}_A$는 크기가 $n \times a$인 0과 1로 구성되는 계수행렬이다. $\boldsymbol{\alpha}$는 요인 A의 a개 확률효과를 나타내는 $a \times 1$인 열벡터이다. $\boldsymbol{X}_B$는 $n \times b$인 계수행렬이다. $\boldsymbol{\beta}$는 요인 B의 b개 확률효과를 나타내는 $b \times 1$인 열벡터이다. $\boldsymbol{X}_{AB}$는 $n \times ab$인 계수행렬이다. $(\boldsymbol{\alpha\beta})$는 요인 A와 요인 B의 ab개 교호작용을 나타내는 $ab \times 1$인 열벡터이다. 다차원상의 관측벡터 $\boldsymbol{y}$에 대한 일반적인 가정은 MVN$(\boldsymbol{j}\mu, \boldsymbol{\Sigma})$로 주어진다. 처리조합 (i, j)의 실험모집단에 대한 분포로 각기 평균 μ이고 분산 σ^2_{ij}인 정규분포를 가정할 때, 크기 n인 실험자료를 나타내는 벡터 $\boldsymbol{y}$는 MVN$(\boldsymbol{j}\mu, \boldsymbol{\Sigma})$인 분포로 부터의 한 관측벡터로 간주된다. 식 (2.2)의 모형행렬을 $\boldsymbol{X}$라 둘 때, $\boldsymbol{X}=(\boldsymbol{j}, \boldsymbol{X}_A, \boldsymbol{X}_B, \boldsymbol{X}_{AB})$이다. 모평균 μ와 두 요인들의 확률효과 벡터를 포함하는 벡터를 $\boldsymbol{\theta}$라 두면 $\boldsymbol{\theta}'=(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta}))$임을 나타

낸다. 모형행렬 $\boldsymbol{X}$는 크기가 $n \times p$인 계수행렬이고 $\boldsymbol{\theta}$는 $p \times 1$인 모수벡터이다. 단, $p = (1+a+b+ab)$개로 주어진다.

확률모형의 가정하에 분산성분을 구하는 방법으로 적률법을 가정한다. 적률법에 의한 추정은 두 가지 계산과정을 필요로 한다. 하나는 확률요인의 변동에 따른 제곱합의 계산과 다른 하나는 제곱합의 기대값을 계산하는 과정이다. 확률모형의 가정하에 적률법으로 분산성분을 추정할 때 Henderson (1953) 방법 I인 ANOVA방법이나 Henderson 방법 III을 적용할 수 있다. Henderson의 방법 III 또는 상수적합법 (fitting constants method)은 완전모형 (full model)과 부분모형 (sub model)의 적합에 따른 제곱합의 차로 해당요인의 변동량을 구한 다음에 Hartley (1967)의 합성법 (synthesis)을 이용하여 분산성분의 계수를 구한다. 그러나, 본 논문에서는 확률요인의 변동에 따른 제곱합의 계산을 위해 확률벡터의 계수행렬로 생성된 벡터공간에서 사영행렬을 이용한다. 변동요인에 따른 평균평방합의 기대값을 계산하기 위해 고유근을 어떻게 활용할 수 있는 가를 구체적으로 살펴보기로 한다.

3. 사영을 이용한 확률요인의 제곱합

이원 확률모형의 행렬모형식 (2.2)의 가정하에 사영을 이용한 요인별 제곱합을 구해본다. 분산성분의 추정에 필요한 제곱합을 Henderson의 방법 III은 완전모형의 적합에 따른 제곱합과 부분모형의 적합에 따른 제곱합간의 차로 확률요인별 변동에 따른 제곱합을 계산하게 된다. Henderson의 방법 III을 적용할 때, 변동요인별 제곱합 계산은 완전모형과 부분모형의 적합에 따른 변동량간의 차에 대한 계산으로 행해진다. 이러한 번거로운 계산과정은 사영행렬의 이용으로 단순화 될 수 있다.

Henderson의 방법 III이 고정효과모형의 요인별 변동량을 구하는 방법중 순차적 적합에 따른 제1종 제곱합과 일치함을 살펴본다. 식 (2.2)에 Henderson 방법 III을 적용할 때 모형에서 추정되어야 할 분산성분들은 σ^2_α, σ^2_β, $\sigma^2_{(\alpha\beta)}$과 σ^2_ϵ이다. 분산성분들의 선형함수로 주어지는 제곱합을 $\mathrm{R}(\cdot)$로 표기한다. $\mathrm{R}(\cdot)$는 $(\cdot)$로 표시된 효과들을 포함한 모형의 적합에 따른 제곱합에서의 감소 (reduction in sum of squares)를 나타내는 표기이다. 추정해야 할 분산성분이 넷이므로 관련된 제곱합도 네개이다. 네개의 제곱합을 구하기 위해 모형행렬 $\boldsymbol{X}$에 따른 제곱합을 SST라 두자. SST=$\mathrm{R}(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta}))$이다. SST=$\boldsymbol{y}'\boldsymbol{X}\boldsymbol{X}^-\boldsymbol{y}$이다. 평균 μ의 적합에 따른 제곱합의 감소를 SSM이라 두자. SSM을 $\mathrm{R}(\mu)$로 나타내면 $\mathrm{R}(\mu)=\boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^-\boldsymbol{y}$이다. 따라서, 평균에 따른 적합량을 제외한 제곱합의 감소를 $\mathrm{R}(\boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta})|\mu)$로 표현하면 $\mathrm{R}(\boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta})|\mu)$=SST-SSM를 나타낸다. 관련된 사영행렬로 표현하면

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta})|\mu) = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{y} \tag{3.1}$$

이다. 식(3.1)의 2차형식과 관련된 사영행렬은 모형행렬 $\boldsymbol{X}$에 의해 생성된 벡터공간 χ에서 $\boldsymbol{j}$에 의해 생성된 부분공간과 직교하는 부분공간으로의 사영을 나타내는 행렬이다. 행렬 $\boldsymbol{X}_1$을 $\boldsymbol{X}_1 = (\boldsymbol{j}, \boldsymbol{X}_A)$라 두자. $\boldsymbol{X}_1$에 의해 생성된 부분공간으로의 사영까지의 제곱거리는 $\mathrm{R}(\boldsymbol{\mu}, \boldsymbol{\alpha})=\boldsymbol{y}'\boldsymbol{X}_1\boldsymbol{X}_1{}^-\boldsymbol{y}$이다. 벡터공간 χ에서 $\boldsymbol{X}_1$에 의해 생성된 부분공간과 직교하는 부분 공간으로의 사영에 이르는 제곱거리는 다음과 같다.

$$R(\boldsymbol{\beta}, (\boldsymbol{\alpha\beta})|\mu, \boldsymbol{\alpha}) = \boldsymbol{y}\prime(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_1\boldsymbol{X}_1{}^-)\boldsymbol{y} \tag{3.2}$$

식 (3.2)의 2차형식과 관련된 사영행렬 $(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_1\boldsymbol{X}_1{}^-)$는 $\boldsymbol{X}_1$의 사영행렬 $\boldsymbol{X}_1\boldsymbol{X}_1{}^-$과 직교함을 보여주고 있다. 식 (3.2)는 두 확률효과 벡터 $\boldsymbol{\beta}$와 $(\boldsymbol{\alpha\beta})$의 변동에 따른 제곱합을 나타내고 있다. 행렬 $\boldsymbol{X}_2$를 $\boldsymbol{X}_2 = (\boldsymbol{j}, \boldsymbol{X}_A, \boldsymbol{X}_B)$라 두자. $\boldsymbol{X}_2$에 의해 생성된 벡터공간으로의 사영까지의 제곱거리는 $\mathrm{R}(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta})=\boldsymbol{y}'\boldsymbol{X}_2\boldsymbol{X}_2{}^-\boldsymbol{y}$이다. 벡터공간 χ에서 $\boldsymbol{X}_2$에 의해 생성된 벡터공간과 직교하는 부분 벡터공간으로의 사영에 이르는 제곱거리는 다음과 같다.

$$R((\boldsymbol{\alpha\beta})|\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{y}\prime(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_2\boldsymbol{X}_2{}^-)\boldsymbol{y} \tag{3.3}$$

식 (3.3)은 행렬 $\boldsymbol{X_2}$에 의한 벡터공간과 직교하는 χ의 부분 벡터공간으로의 사영을 하였을 때 사영까지의 거리제곱합을 나타내고 있다. 오차벡터 $\boldsymbol{\epsilon}$에 의해 생성된 벡터공간으로의 사영행렬은 $\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-}$이다. 사영까지의 거리제곱합을 SSE라 두자.

$$SSE = \boldsymbol{y}\prime\boldsymbol{y} - R((\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, (\boldsymbol{\alpha\beta})) = \boldsymbol{y}\prime(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{y} \tag{3.4}$$

식 (2.1)의 확률모형의 가정에 따른 Henderson 방법 III의 적용에서 R(·)로 표기된 해당제곱합은 2차형식과 관련된 사영행렬로 표시될 수 있음을 보여주고 있다. 또한, 식 (3.1)에서 식 (3.4)의 제곱합은 모형행렬 $\boldsymbol{X}$의 벡터공간 χ가 상호직교하는 벡터공간의 합공간으로 구성될 때, 부분 벡터공간에서의 사영행렬로 주어지는 2차형식임을 보여준다.

4. 고유근을 이용한 기대값

적률법으로 확률모형에 내포된 분산성분을 추정하기 위한 또 하나의 계산과정은 변동요인에 따른 제곱합의 기대값 계산이다. 요인별 제곱합의 기대값을 구하기 위한 방법으로 Hartley의 합성법을 이용할 수 있다. 먼저, 2차형식의 기대값을 살펴보기로 한다. 행렬모형식 (2.2)에서 $V(\boldsymbol{y})$를 구해보자.

$$V(\boldsymbol{y}) = V(\boldsymbol{j}\mu + \boldsymbol{X}_A\boldsymbol{\alpha} + \boldsymbol{X}_B\boldsymbol{\beta} + \boldsymbol{X}_{AB}(\boldsymbol{\alpha\beta}) + \boldsymbol{\epsilon}) \tag{4.1}$$

이다. 확률모형의 가정을 이용할 때,

$$\boldsymbol{\Sigma} = \sigma_\alpha^2\boldsymbol{X}_A\boldsymbol{X}_A' + \sigma_\beta^2\boldsymbol{X}_B\boldsymbol{X}_B' + \sigma_{(\alpha\beta)}^2\boldsymbol{X}_{AB}\boldsymbol{X}_{AB}' + \sigma_\epsilon^2\boldsymbol{I} \tag{4.2}$$

이다. 분산성분을 구하기 위한 제곱합은 $\boldsymbol{y}$의 이차형식으로 주어진다. 이차형식을 $\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}$로 나타내면 M은 양반정치행렬이며 멱등행렬이다.

$$\begin{aligned} E(\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y}) =& tr(\boldsymbol{M\Sigma}) + \mu^2\boldsymbol{j}'\boldsymbol{M}\boldsymbol{j} \\ =& \sigma_\alpha^2 tr(\boldsymbol{M}\boldsymbol{X}_A\boldsymbol{X}_A') + \sigma_\beta^2 tr(\boldsymbol{M}\boldsymbol{X}_B\boldsymbol{X}_B') + \sigma_{(\alpha\beta)}^2 tr(\boldsymbol{M}\boldsymbol{X}_{AB}\boldsymbol{X}_{AB}') + \sigma_\epsilon^2 tr(\boldsymbol{M}) \end{aligned} \tag{4.3}$$

으로 구해진다. $\boldsymbol{y}'M\boldsymbol{y}$의 M은 사영공간으로의 사영행렬로 주어지고 $\boldsymbol{j}$와의 곱은 $M\boldsymbol{j} = \boldsymbol{0}$이다. 식 (4.3)에 의한 제곱합들의 기대값을 구해 보기로 한다. 식 (3.1)의 제곱합을 $\boldsymbol{Q}_1$이라 두면

$$\begin{aligned} E(\boldsymbol{Q}_1) =& \sigma_\alpha^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_A\boldsymbol{X}_A') + \sigma_\beta^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_B\boldsymbol{X}_B') + \\ & \sigma_{(\alpha\beta)}^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{j}\boldsymbol{j}^-)\boldsymbol{X}_{AB}\boldsymbol{X}_{AB}') + \sigma_\epsilon^2 tr(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{j}\boldsymbol{j}^-) \end{aligned} \tag{4.4}$$

로 구해진다.식 (3.2)의 제곱합을 $\boldsymbol{Q}_2$라 두면

$$\begin{aligned} E(\boldsymbol{Q}_2) =& \sigma_\alpha^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_1}\boldsymbol{X_1}^-)\boldsymbol{X}_A\boldsymbol{X}_A') + \sigma_\beta^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_1}\boldsymbol{X_1}^-)\boldsymbol{X}_B\boldsymbol{X}_B') + \\ & \sigma_{(\alpha\beta)}^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_1}\boldsymbol{X_1}^-)\boldsymbol{X}_{AB}\boldsymbol{X}_{AB}') + \sigma_\epsilon^2 tr(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_1}\boldsymbol{X_1}^-) \end{aligned} \tag{4.5}$$

이다. 식 (3.3)의 제곱합을 $\boldsymbol{Q}_3$라 두면

$$\begin{aligned} E(\boldsymbol{Q}_3) =& \sigma_\alpha^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_2}\boldsymbol{X_2}^-)\boldsymbol{X}_A\boldsymbol{X}_A') + \sigma_\beta^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_2}\boldsymbol{X_2}^-)\boldsymbol{X}_B\boldsymbol{X}_B') + \\ & \sigma_{(\alpha\beta)}^2 tr((\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_2}\boldsymbol{X_2}^-)\boldsymbol{X}_{AB}\boldsymbol{X}_{AB}') + \sigma_\epsilon^2 tr(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X_2}\boldsymbol{X_2}^-) \end{aligned} \tag{4.6}$$

로 구해진다. 식 (3.4)의 제곱합을 $\boldsymbol{Q}_\epsilon$이라 두면

$$E(\boldsymbol{Q}_\epsilon) = \sigma_\epsilon^2 tr(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^-) \tag{4.7}$$

로 구해진다. Hartley (1967)는 식 (4.3)에서 분산성분들의 계수들로 주어지는 대각합를 구하기 위한 방법으로 합성법을 제공하고 있으나 계산과정이 간편하지 않다. Hartley의 합성법 대신에 제곱합의 기대값과 관련된 행렬의 성질을 이용하면 분산성분의 계수는 간편하게 구해진다. 즉, 대각합내 행렬은 정방행렬이므로 대각합은 고유근의 합이 되는 성질을 이용할 수 있다. 고유근을 이용할 때 식 (4.3)은 다음과 같은 형태로 주어진다.

$$E(\boldsymbol{y}'\boldsymbol{M}\boldsymbol{y})=\sigma_\alpha^2\sum_{i=1}^{n}\lambda_{i(\boldsymbol{M}\boldsymbol{X}_A\boldsymbol{X}_{A'})}+\sigma_\beta^2\sum_{j=1}^{n}\lambda_{j(\boldsymbol{M}\boldsymbol{X}_B\boldsymbol{X}_{B'})}+\sigma_{(\alpha\beta)}^2\sum_{k=1}^{n}\lambda_{k(\boldsymbol{M}\boldsymbol{X}_{AB}\boldsymbol{X}_{AB'})}+\sigma_\epsilon^2\sum_{l=1}^{n}\lambda_{l(\boldsymbol{M})} \quad (4.8)$$

단, $\lambda_{i(\cdot)}$는 $(\cdot)$내 행렬의 고유근을 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$으로 나타낼 때 i번째 고유근이다. 불균형자료의 경우에 균형자료와는 달리 총제곱합을 분해하는 다양한 방법들로 인해 동일한 모형의 분산성분들에 대한 해는 일정하지 않다. Searle (1971)은 분산성분을 추정하기 위한 목적으로 다양한 방법중 하나를 이용하게 될 때 그 중 하나를 선택할 수 있는 기준이 없기 때문에 다양한 이차형식들이 이용될 수 있음을 논의하고 있다. 다시말하면, Henderson의 방법 I 또는 Henderson의 방법 III으로 부터 주어지는 이차형식의 집합들이 동일 모형의 분산성분추정에 이용될 수 있음을 의미하고 있다.

5. 자료분석의 예

이원 확률모형의 분산성분을 구하기 위한 방법으로 사영과 고유근의 활용을 Milliken과 Johnson (1984)의 자료를 통해 구체적으로 살펴보기로 한다. Milliken과 Johnson (1984)은 Henderson 방법 I의 제곱합과 Henderson 방법 III의 제1종 제곱합을 이용하여 분산성분을 추정하는 방법으로 다음의 이원 자료를 들고 있다.

Table 5.1 Two way data

$A \diagdown B$	1	2	3
1	10,12,11	13,15	21,19
2	16,18	13,19,14	11,13

Table 5.1의 자료에 대한 확률모형으로 식 (2.1)을 가정하고 있다. Henderson 방법 I의 제곱합은 식 (2.1)의 가정하에 분산분석법에 의해 계산된다. 3절에서 Henderson 방법 III의 제곱합을 사영을 이용한 제곱합으로 나타내고 있다. Henderson 방법 III은 효과에 따른 제곱합을 계산하기 위해 완전모형 (full model)과 부분모형의 적합에 따른 제곱합에서의 감소량을 이용하고 있다. 행렬모형식 (2.2)의 적합은 $R(\cdot)$으로 나타낼 때 $R(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\alpha\beta})$이고 이는 $\boldsymbol{y}$를 $\boldsymbol{X}$로의 사영을 생각할 때 사영까지의 거리제곱인 $\boldsymbol{y}'(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$가 된다.

식 (3.1)의 제곱합 $\boldsymbol{Q}_1 = 124.5476$, 식 (3.2)의 제곱합 $\boldsymbol{Q}_2 = 123.9048$, 식 (3.3)의 제곱합 $\boldsymbol{Q}_3 = 109.14150$이고 식 (3.4)의 제곱합 $\boldsymbol{Q}_\epsilon = 30.6667$로 구해진다. 식 (4.4), (4.5), (4.6)과 (4.7)로부터 구한 기대값은 다음과 같이 얻어진다.

$$\begin{aligned}
E(\boldsymbol{Q}_1) &=7{\sigma_\alpha}^2 + 9.2857{\sigma_\beta}^2 + 11.5714\sigma_{(\alpha\beta)}^2 + 5\sigma_\epsilon^2 \\
E(\boldsymbol{Q}_2) &=9.1429\sigma_{(\alpha\beta)}^2 + 4\sigma_\epsilon^2 \\
E(\boldsymbol{Q}_3) &=4.5176\sigma_{(\alpha\beta)}^2 + 2\sigma_\epsilon^2 \\
E(\boldsymbol{Q}_\epsilon) &=8\sigma_\epsilon^2
\end{aligned} \quad (5.1)$$

적률법을 이용하는 상수적합법에서 사영을 이용한 연립방정식의 해는 $\sigma_\alpha^2 = -2.9026$, $\sigma_\beta^2 = -3.8503$, $\sigma_{(\alpha\beta)}^2 = 13.9465$이고 $\sigma_\epsilon^2 = 3.8475$이다. Henderson의 방법 I에 의한 Milliken과 Johnson의 해를 살펴보면 $\sigma_\alpha^2 = -8.2523$, $\sigma_\beta^2 = -10.8817$, $\sigma_{(\alpha\beta)}^2 = 15.8593$이고 $\sigma_\epsilon^2 = 3.8333$ 이다. 사영을 이용하는 방법에서의 연립방정식과 Hartley의 합성법에서의 연립방정식은 서로 다른 식을 제공하므로 해의 값은 같지 않게 된다. 왜냐하면 사영을 이용하는 방법에서는 고유근을 활용하고 있기 때문이다. 그러나 방법 간의 차이가 있으나 양수의 분산성분에 대한 추정값들은 거의 유사함을 알 수 있다.

Henderson의 방법 III에서 제1종 제곱합을 이용한 경우의 제1종 추정치는 $\sigma_\alpha^2 = -8.0329$, $\sigma_\beta^2 = -10.5877$, $\sigma_{(\alpha\beta)}^2 = 22.4620$이고 $\sigma_\epsilon^2 = 3.8333$으로 구해짐을 Milliken과 Johnson (1984)에서 볼 수 있다. Henderson의 방법 III에서 제1종 제곱합을 사영으로 구한 경우의 추정값들은 $\sigma_\alpha^2 = -8.0414$, $\sigma_\beta^2 = -10.1684, \sigma_{(\alpha\beta)}^2 = 22.4627$이고 $\sigma_\epsilon^2 = 3.8333$으로 거의 동일하게 구해짐을 알 수 있다.

6. 결론

본 논문은 실험자료의 분석을 위한 이원 분산분석모형으로 확률효과 모형을 가정하고 있다. 모형에 내포된 모수가 확률변수로 간주되기 때문에 분산성분의 추정에 관심을 두고 있다. 분산성분의 추정을 위해 적률법을 이용하는 Henderson의 두 방법인 분산분석법과 상수적합방법을 다루고 있다. Henderson의 방법 I과 III으로도 불리우는 이들 방법은 분산성분의 적률추정량을 얻기 위해 필요한 제곱합과 기대값의 계산방법을 제공하고 있다. Henderson의 방법 I은 제곱합의 계산을 위해 고정효과의 모형에서 이용되는 분산분석법을 이용하여 구하고 있다. Henderson의 방법 III은 제곱합을 모형의 적합을 통해 구하게 된다. 이때 모형의 적합은 완전모형과 완전모형의 부분인 부분모형의 적합에 따른 제곱합간의 차로 구해야 하는 방식을 따르고 있다. 두 방법 모두 기대값의 계산은 Hartley의 합성법을 이용하여 구하게 된다.

본 논문의 주안점은 Henderson의 방법들에서 제공되는 제곱합의 계산방식을 벡터공간에서의 사영을 이용하여 효율적으로 계산할 수 있음을 논의하고 있다. 특히 상수적합법 또는 Henderson의 방법 III은 다양한 선형모형들을 자료에 적합시켜 해당하는 제곱합을 구하는 방식을 이용하고 있다. 이 방법은 완전모형과 부분모형들의 적합에 따른 제곱합에서의 감소를 이용하고 있다. 본 논문에서는 Henderson의 방법 III에 따른 제곱합과 일치하는 사영공간으로의 사영제곱합 뿐만아니라 다양한 방법들에서 사영이 이용될 수 있음을 보여주고 있다. 또한, 제곱합의 기대값 계산을 위해 Hartley의 합성법 대신에 관련된 행렬의 고유근을 이용하는 방법을 제공하고 있다. 사영에 의한 제곱합의 계산과 관련행렬의 고유근의 활용에 대한 타당성을 입증하기 위해 Milliken과 Johnson (1984)에서의 자료결과와 거의 동일하게 주어짐을 보여주고 있다.

References

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

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

Choi, J. S. (2012b). Esimable functions of fixed-effects model by projections. *Journal of the Korean Data & Information Science Society*, **23**, 487-494.

Choi, J. S. (2013). Estimable functions of less than full rank linear model. *Journal of the Korean Data & Information Science Society*, **23**, 333-339.

Corbeil, R. R. and Searle, S. R. (1976). A comparison of variance component estimators. *Biometrics*, **9**, 226-252.

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

Hartley, H. O. (1967). Expectations, variances and covariances of ANOVA mean squares by "synthesis". *Biometrics*, **23**, 105-114.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, **32**, 779-791.

Huynh, H., and Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact F-distributions. *Journal of the American Statistical Association*, **65**, 1582-1589.

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

Rao, C. R. (1971). Minimum variance quadratic unbiased estimation of variance components. *Journal of Multivariate Analysis*, **1**, 445-456.

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

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

Journal of the Korean Data & Information Science Society
2014, **25**(3), 547–554

http://dx.doi.org/10.7465/jkdi.2014.25.3.547
한국데이터정보과학회지

Projection analysis for two-way variance components

Jaesung Choi[1]

[1]Department of Statistics,Keimyung University

Received 2 April 2014, revised 19 April 2014, accepted 1 May 2014

Abstract

This paper discusses a method of estimating variance components for random effects model. Henderson's method I and III are discussed for the esimation of variance components. This paper shows how to use projections instead of using Henderson's methods for the calculation of sums of squares which are quadratic forms in the observations. It also discusses that eigenvalues can be used for getting the expectations of sums of squares in place of using the method of Hartley's synthesis. It shows the suggested method is much more effective than those methods.

Keywords: Eigenvalue, projection, quadratic form, random effects, variance components.

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