Journal of the Korean Data & Information Science Society
2014, **25**(4), 799–805

http://dx.doi.org/10.7465/jkdi.2014.25.4.799
한국데이터정보과학회지

사영에 의한 제3종 제곱합

최재성[1]

[1]계명대학교 통계학과

접수 2014년 6월 11일, 수정 2014년 6월 26일, 게재확정 2014년 7월 8일

요 약

본 논문은 이원의 고정효과모형에서 사영에 근거한 제3종 제곱합을 구하는 방법을 다루고 있다. 제3종 제곱합의 모형적합 방식에 따른 분산분석에서 자료의 총제곱합은 요인별 제곱합으로 분해된 양과 일치하지 않는다. 변동량의 차이가 단순히 모형의 적합방식에 기인한다고 간주하는 고전적 해석과는 달리 자료의 분산분석과정에서 발생하는 변동량의 차이가 어디에서 어느 정도 발생하고 있는 가에 대해 사영을 이용하여 규명할 수 있음을 다루고 있다. 또한 사영공간의 중첩성으로 인한 변동량의 차에 대한 기하학적 해석과 함께 자료의 변동량 계산에 있어 고유근과 고유벡터가 어떻게 이용될 수 있는 가를 논의하고 있다.

주요용어: 고유근, 고유벡터, 균형자료, 사영, 사영행렬, 제3종 제곱합.

1. 서론

실험자료를 분석하기 위해 고정효과모형의 가정하에 자료분석하는 경우를 생각해 본다. 실험으로부터 비교하고자 하는 처리들의 효과와 관련된 가설의 검정방법으로 분산분석법은 선형모형의 가정하에 행해지는 분석방법이다. 이 방법은 자료의 전체변동량인 총제곱합을 변동요인으로 간주되는 성분들의 제곱합으로 분할하는 것으로 이루어진다. 분산분석의 기본개념 및 활용에 관한 문헌들은 Steel과 Torrey (1980), Cochran과 Cox (1957) 그리고 Montgomery (1976) 등에서 살펴볼 수 있다.

분산분석을 위한 요인별 제곱합의 계산을 위해 모형적합방식에 따른 다양한 기법을 이용할 수 있다. 즉, 변동요인별 제곱합의 계산에 총제곱합을 분할하는 방식의 모형비교를 통해 구하는 방법을 생각할 수 있다. 모형의 다양한 비교방식을 통해 변동요인별 제곱합을 구하는 방법들이 Milliken과 Johnson (1984) 그리고 Searle (1971) 등에서 구체적으로 논의되고 있다.

모형의 비교를 통해 성분들의 제곱합을 구할 때 모형의 적합방식에 따라 제1종 제곱합, 제2종 제곱합, 제3종 제곱합 등으로 구분하게 된다. 모형의 적합방식에 따른 제곱합은 자료가 균형자료일 때 변동요인별 제곱합은 동일한 값으로 주어진다. 그러나 처리간에 자료의 개수가 동일하지 않는 불균형자료인 경우에는 모형의 적합방식에 따라 변동요인별 제곱합은 달라질 수 있다. 제1종 제곱합은 자료가 균형이건 불균형이건 상관없이 총제곱합은 요인별 제곱합으로 분할되고 이들의 총합은 자료의 전체변동량과 동일하다. 이와 관련된 문헌으로 Choi (2012)는 제2종 제곱합에 대한 사영분석을 통하여 제1종 분석과 제2종 분석간의 차이를 규명하고 있음을 볼 수 있다. 그러나 제3종 제곱합은 자료가 불균형인 경우에 성분들의 제곱합은 자료의 전체변동량의 분할로 주어지지 않게 된다. 모형의 적합방식에 따른 제곱합을 구하는 방식에 대한 구체적 논의는 여러 문헌에서 제공되고 있으나 적합방식과 관련된 제곱합의

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

차이가 어디서 발생하고 있는 가에 대한 논의는 구체적이지 않다. 자료의 분산분석에서 발생하게 되는 분석량 간의 차이에 대한 규명을 위해 사영의 개념에 근거할 때 그러한 차이에 대한 해석이 가능하게 된다. Choi (2011, 2012)는 실험자료의 분산분석을 위해 사영을 이용하여 분석하는 기법을 다루고 있으며 Choi (2013)은 불완전 계수의 모형행렬을 갖는 선형모형에서 추정가능함수의 타당성을 검정하기 위해 사영행렬을 활용하는 방법을 논의하고 있다. 그러나 자료분석의 일반적인 ANOVA 방법에서 사영을 이용하여 활용하는 방법에 관한 연구는 많지 않은 편이다.

본 논문에서는 자료가 불균형일 때 제3종 제곱합과 관련된 모형의 적합방식에서 유도되는 모형행렬을 이용하여 사영의 관점에서 구하는 방법을 제공하고자 한다. 사영의 관점에서 자료분석이 행해질 때 기존의 분석방법에서 명확히 규명되지 않는 양에 대한 기하학적 해석과 설명이 간명하게 이루어질 수 있게 된다. 따라서 이 방법은 제3종 제곱합을 이용한 총제곱합의 분할에서 자료의 변동량이 일치하지 않는 부분에 대한 설명이 사영에 의해 구체적으로 이루어질 수 있음을 논의하고 있다. 사영의 관점에서 제곱합이 계산될 때 관측벡터의 벡터공간과 관련된 행렬의 여러 성질이 유도되고 활용되는 이점이 있게 된다. 특히, 사영공간에서의 고유근과 고유벡터를 이용한 제곱합의 계산과 더불어 이들의 활용에 관해 논의하고자 한다.

2. 모형의 가정

실험자료를 분석하기 위한 모형으로 이원 고정효과모형을 가정한다. 실험에서 고려되는 두 고정요인을 R과 C라 두자. R은 행에서의 수준들로 구성되는 행요인을 나타내고 C는 열에서의 수준들을 나타내는 열요인이라 간주한다. 행요인 R은 r개의 수준들로 구성되고 열요인 C는 c개의 수준들로 구성된다고 가정한다. R의 수준 i, $i = 1, 2, \cdots, r$와 요인 C의 수준 j, $j = 1, 2, \cdots, c$와의 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 고정요인이므로 실험단위의 관측반응에 대한 모형으로 고정효과모형을 가정한다. 요인 R의 수준 i에서의 수준효과를 τ_i라 두면, r개의 τ_i는 고정효과이다. 요인 C의 수준 j에서의 수준효과를 γ_j라 두면, c개의 γ_j도 고정효과를 나타낸다. 요인 R의 수준 i, 요인 C의 수준 j의 처리조합 (i, j)에서의 고정효과들의 교호작용을 δ_{ij}로 나타낸다. 자료를 분석하기 위한 고정효과모형은 다음과 같이 주어진다.

$$y_{ijk} = \mu + \tau_i + \gamma_j + \delta_{ij} + \epsilon_{ijk} \tag{2.1}$$

단, μ는 전체평균을 나타낸다. τ_i, $i = 1, 2, \cdots, r$는 요인 R의 i번째 수준효과이고 γ_j, $j = 1, 2, \cdots, c$는 요인 C의 j번째 수준효과를 나타낸다. δ_{ij}는 두 요인 R과 C의 결합수준 (i, j)에서의 교호작용이고 ϵ_{ijk}, $k = 1, 2, \cdots, n_{ij}$는 오차항을 나타낸다. 오차항의 분포로 $N(0, {\sigma_\epsilon}^2)$을 가정하고 오차들은 상호독립이라 가정한다. 두 요인의 수준결합으로 주어지는 rc개의 처리조합 (i, j)에서 수집된 자료를 크기 $n \times 1$인 벡터 $\boldsymbol{y}$로 나타내면 식 (2.1)에 해당하는 행렬모형식은 다음과 같이 표현된다.

$$\boldsymbol{y} = \boldsymbol{j}\mu + \boldsymbol{X}_R\boldsymbol{\tau} + \boldsymbol{X}_C\boldsymbol{\gamma} + \boldsymbol{X}_{RC}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{2.2}$$

단, $n = \sum_{i,j} n_{ij}$이고 모평균 μ의 계수벡터 $\boldsymbol{j}$는 n개의 원소가 모두 1인 열벡터이다. $\boldsymbol{X_R}$은 크기가 $n \times r$인 0과 1로 구성되는 계수행렬이다. $\boldsymbol{\tau}$는 요인 R의 r개 고정효과를 나타내는 $r \times 1$인 열벡터이다. $\boldsymbol{X_C}$는 $n \times c$인 계수행렬이다. $\boldsymbol{\gamma}$는 요인 C의 c개 고정효과를 나타내는 $c \times 1$인 열벡터이다. $\boldsymbol{X_{RC}}$는 $n \times rc$인 계수행렬이다. 식 (2.2)의 행렬표현식은 $\boldsymbol{y}$가 다차원상의 성분벡터들의 합으로 구성됨을 의미한다. 이들 성분벡터는 자료의 전체변동량에 기여하는 변동요인과 관련된 부분공간으로의 사영에 해당

한다. 따라서 각 부분공간으로의 사영까지의 거리제곱합은 변동요인에 따른 제곱합을 나타내게 된다. 사영에 의한 거리제곱합은 모형의 적합방식에 따라 다양하게 계산되고 계산값들 간에 차이가 있게 된다.

식 (2.1)의 이원 고정효과모형에서 y_{ijk}는 다섯 개 성분들의 합으로 표현된다. 분산분석도 마찬가지로 관측값들의 총제곱합을 다섯 개의 성분제곱합으로 분해한다. 이들은 μ에 따른 제곱합, 행요인 R의 r개 수준효과 간의 변이에 따른 제곱합, 요인 C의 bc개 수준효과 간의 변이에 따른 제곱합, 행요인 R와 열요인 C의 rc개 결합수준에서의 교호작용 효과들의 변이에 따른 제곱합 그리고 오차 ϵ들의 변이에 따른 오차제곱합으로 구분된다. 여기서는 제3종 제곱합과 관련된 모형의 적합방식을 살펴보기로 한다.

3. 사영에 의한 제3종 제곱합

고정효과모형 (2.1)의 가정하에 제3종 제곱합의 계산방법은 관심효과만을 제외한 모형의 적합에 따른 변동량과 관심효과를 포함한 모형의 적합에서 주어지는 변동량간의 차이로서 구하게 된다. 이와같은 모형의 비교방식을 통하여 효과에 따른 제3종 제곱합은 가정된 모형내 효과들의 추정가능한 함수들의 가설을 검정하는 데 이용된다. 여러 유형의 모형비교 방식중 제3종 제곱합을 구하는 모형적합방식이 선호되는 이유로 Milliken과 Johnson (1984)은 제3종 가설들이 실험자가 관심을 갖는 의미있는 가설임을 나타내고 있다.

구체적 과정을 사영으로 언급하기 위하여 행렬모형식 (2.2)에서 모형행렬을 $\boldsymbol{X}$라 두면, $\boldsymbol{X} = (\boldsymbol{j}, \boldsymbol{X}_R, \boldsymbol{X}_C, \boldsymbol{X}_{RC})$이다. 행요인 R의 r개 수준효과에 따른 변동량을 SSR이라 두자. SSR은 완전효과모형 (full effects model)의 적합으로부터 구해지는 제곱합과 행요인 R의 주효과를 제외한 축소모형의 적합으로부터 얻어지는 제곱합과의 차이이다. 모형행렬 $\boldsymbol{X}$로부터 행효과와 관련된 계수행렬을 제외한 모형행렬을 $\boldsymbol{X}_r$이라 둘 때 사영을 이용한 제곱합은 다음과 같다.

$$SSR = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-)\boldsymbol{y} \tag{3.1}$$

단, $\boldsymbol{X}_r = (\boldsymbol{j}, \boldsymbol{X}_C, \boldsymbol{X}_{RC})$이다. 열요인 C의 c개 수준효과에 따른 변동량을 SSC라 두자. SSC는 완전효과모형의 적합으로부터 구해지는 제곱합과 행요인 C의 주효과를 제외한 축소모형의 적합으로부터 얻어지는 제곱합과의 차이이다. 모형행렬 $\boldsymbol{X}$로부터 열효과와 관련된 계수행렬을 제외한 모형행렬을 $\boldsymbol{X}_c$라 둘 때 사영을 이용한 제곱합은 다음과 같다.

$$SSC = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_c\boldsymbol{X}_c^-)\boldsymbol{y} \tag{3.2}$$

단, $\boldsymbol{X}_c = (\boldsymbol{j}, \boldsymbol{X}_R, \boldsymbol{X}_{RC})$이다. 두 요인 R과 C의 교호작용 벡터 $\boldsymbol{\delta}$에 따른 제곱합을 $SSRC$라 두자. $SSRC$는 완전효과모형의 적합으로부터 구해지는 제곱합과 두 요인 R과 C의 교호작용효과를 제외한 축소모형의 적합으로부터 얻어지는 제곱합과의 차이이다. 모형행렬 $\boldsymbol{X}$로부터 교호작용과 관련된 계수행렬을 제외한 모형행렬을 $\boldsymbol{X}_{rc}$라 둘 때 사영을 이용한 제곱합은 다음과 같다.

$$SSRC = \boldsymbol{y}'(\boldsymbol{X}\boldsymbol{X}^- - \boldsymbol{X}_{rc}\boldsymbol{X}_{rc}^-)\boldsymbol{y} \tag{3.3}$$

단, $\boldsymbol{X}_{rc} = (\boldsymbol{j}, \boldsymbol{X}_R, \boldsymbol{X}_C)$이다. $\boldsymbol{y}$의 이차형식으로 주어지는 제곱합들은 전체변동량의 가능한 한 분할이나 이들의 합은 총제곱합과 일치하지 않게 된다. 그 이유는 모형행렬 $\boldsymbol{X}$에 의해 생성되는 사영공간이 평균을 포함하여 네 성분을 나타내는 부분공간들로 구성되어야 하나 제3종 제곱합을 구하는 모형의 적합방식으로 인해 단순히 직교하는 두 부분공간으로 나누어지기 때문이다.

변동요인의 성분을 나타내는 부분공간들에서 계산된 양은 이들 공간의 중복성으로 인해 각 부분공간으로의 사영제곱합은 전체변동량의 분할로 주어지지 않는다. 이 경우에 각 변동성분과 관련된 중복량을 구하기 위한 방법은 모형행렬 $\boldsymbol{X}$에 의한 사영공간이 중복되지 않는 성분벡터로 구성되는 부분공간으로의 사영을 이용하는 것이다. 모형행렬 $\boldsymbol{X}$에 의한 사영공간이 행렬 $\boldsymbol{X}_{rc}$에 의한 공간과 그렇지 않은 두

공간으로 분리될 수 있다. 또한, $\boldsymbol{X}_{rc}$는 요인 R과 C의 주효과에 따른 변동량 계산에서 부분공간을 나타내는 사영행렬에 공통적으로 포함되는 행렬 $\boldsymbol{X}_{RC}$를 포함하고 있지 않다. 그러므로 $\boldsymbol{X}_{RC}$를 포함하고 있지 않는 $\boldsymbol{X}_{rc}$에 의해 생성되는 공간으로의 사영을 이용하여 행요인 R의 주효과와 열요인 C의 주효과에 따른 변동량을 계산할 수 있게 된다. $\boldsymbol{X}_{rc}$로의 사영은 $\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}^{-}\boldsymbol{y}$이다. $\boldsymbol{X}_{rc}$가 세 성분벡터들로 구성되므로 모형의 순차적 적합에 해당하는 제곱합을 계산하게 된다. $\boldsymbol{X}_{rc}$로의 사영에서 사영까지의 거리제곱합을 SS_T라 두면

$$SS_T = \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{y} \tag{3.4}$$

이다. 식 (3.4)의 두 번째 항을 SS_t라 두자. SS_t는 두 성분들 SS_{tr}과 SS_{tc}의 합으로 표시될 수 있다. SS_{tr}은 다음과 같이 구해진다.

$$SS_{tr} = \boldsymbol{y}'[(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_R][(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_R]^{-}\boldsymbol{y} \tag{3.5}$$

SS_{tr}은 $\boldsymbol{y}$의 이차형식임을 알 수 있다. 이 이차형식과 관련된 행렬을 $\boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-}$이라 두면

$$\boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-} = [(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_R][(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-})\boldsymbol{X}_R]^{-} \tag{3.6}$$

이다. SS_{tc}는 다음과 같이 구해진다.

$$SS_{tc} = \boldsymbol{y}'[(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-} - \boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-})\boldsymbol{X}_C][(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-} - \boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-})\boldsymbol{X}_C]^{-}\boldsymbol{y} \tag{3.7}$$

SS_{tc}는 $\boldsymbol{y}$의 이차형식임을 알 수 있다. 이 이차형식과 관련된 행렬을 $\boldsymbol{X}_{tc}\boldsymbol{X}_{tc}{}^{-}$이라 두면

$$\boldsymbol{X}_{tc}\boldsymbol{X}_{tc}{}^{-} = [(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-} - \boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-})\boldsymbol{X}_C][(\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} - \boldsymbol{j}\boldsymbol{j}^{-} - \boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-})\boldsymbol{X}_C]^{-} \tag{3.8}$$

이다. 즉, $\boldsymbol{X}_{rc}$로의 사영을 나타내는 사영행렬 $\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-}$은 직교하는 사영행렬들의 합으로 표현된다.

$$\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^{-} = \boldsymbol{j}\boldsymbol{j}^{-} + \boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-} + \boldsymbol{X}_{tc}\boldsymbol{X}_{tc}{}^{-} \tag{3.9}$$

따라서 식 (3.4)에서 SS_T의 계산은

$$SS_T = \boldsymbol{y}'\boldsymbol{j}\boldsymbol{j}^{-}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_{tr}\boldsymbol{X}_{tr}{}^{-}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{X}_{tc}\boldsymbol{X}_{tc}{}^{-}\boldsymbol{y} \tag{3.10}$$

로 구해진다. 식 (3.10)은 요인별 제3종 제곱합이 자료전체의 변동량과 일치하지 않는 부분에 대한 차이를 설명하여 주고 있다. 이는 요인의 변동량을 계산하기 위한 모형의 적합방식에 해당하는 사영공간이 직교분할될 때 상호직교하는 사영행렬들의 합으로 표시됨을 나타낸다. 이 행렬들을 이용하여 성분변동량의 합과 전체변동량의 차이가 어디에서 발생하고 있는 가를 확인할 수 있게 되고 그 차이가 어느 정도인 가를 알 수 있게 된다.

4. 제3종 제곱합의 예

제3종 제곱합을 계산하기 위한 자료의 예로 Milliken과 Johnson (1984)의 불균형 자료를 이용하기로 한다. Table 4.1은 두 처리요인 R과 C의 수준들이 각각 R_1, R_2와 C_1, C_2, C_3일 때 이들의 결합수준에서 행해진 실험자료를 나타내고 있다.

Table 4.1 Unbalanced two-way experimental data

R C	C_1	C_2	C_3
R_1	19, 20, 21	24, 26	22, 25, 25
R_2	25, 27	21, 24, 24	31, 32, 33

행렬모형식 (2.2)의 가정하에 변동요인에 따른 제3종 제곱합을 사영에 의해 구해보기로 한다. 식 (3.1)을 이용한 $SSR = 61.714$로 주어진다. SSR은 또한 식 (3.1)의 이차형식과 관련된 사영행렬의 고유벡터를 이용하여 구할 수도 있다. 즉, 식 (3.1)의 사영행렬 $\boldsymbol{XX}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-$의 0아닌 고유근을 갖는 고유벡터를 구한다. 왜냐하면 $\boldsymbol{XX}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-$는 멱등행렬로 고유근은 0 또는 1인 근을 갖기 때문이다. 행요인 R은 수준이 둘이므로 자유도는 1이다. 따라서 사영행렬 $\boldsymbol{XX}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-$는 1차원의 부분공간으로의 사영을 나타내는 행렬이고 0 아닌 고유근 1을 하나만 갖게 된다. 고유근 1에 해당하는 고유벡터를 $\boldsymbol{E}_t'$로 나타내면 $\boldsymbol{E}_t' = (\boldsymbol{E}_{t1}, \boldsymbol{E}_{t2})$로 주어진다. $\boldsymbol{E}_{t1}$과 $\boldsymbol{E}_{t2}$는 다음과 같다.

$\boldsymbol{E}_{t1} = (-0.2182, -0.2182, -0.2182, -0.3273, -0.3273, -0.2182, -0.2182, -0.2182)$
$\boldsymbol{E}_{t2} = (0.3273, 0.3273, 0.2182, 0.2182, 0.2182, 0.2182, 0.2182, 0.2182)$

따라서, 고유벡터 $\boldsymbol{E}_t'$를 이용한 SSR의 계산은 $\boldsymbol{y}'(\boldsymbol{E}_t\boldsymbol{E}_t')\boldsymbol{y}$로 행해지고 식 (3.1)과 거의 동일한 값 61.704를 구하게 된다. 0.01의 차이는 고유벡터의 성분표시에서 반올림한 자리수로 인한 오차이다. 고유근과 고유벡터는 변동요인에 따른 변동량의 계산에 수월할 뿐만아니라 자유도의 수, 벡터공간의 종속성을 파악하는 데 도움이 됨을 알 수 있다. 열요인 C의 수준변화에 따른 변동량은 식 (3.2)로 부터 $SSC = 77.169$로 구해진다. 교호작용에 따른 평방합 $SSRC$는 식 (3.3)에 의해 71.631로 주어진다.

모형행렬 $\boldsymbol{X}$로의 사영은 $\boldsymbol{XX}^-\boldsymbol{y}$이므로 사영까지의 거리제곱합은 $\boldsymbol{y}'\boldsymbol{XX}^-\boldsymbol{y}$이 되고 이 양은 10,189이다. 전체평균 μ에 따른 제곱합은 $\boldsymbol{y}'\boldsymbol{jj}^-\boldsymbol{y}$이고 9950.062이다. 따라서 요인 R과 C의 주효과와 교호작용에 따른 변동량은 $\boldsymbol{y}'(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{y}$이므로 238.938이다. 이 양은 요인 R의 주효과에 따른 평방합 61.714, 요인 C의 주효과에 따른 평방합 77.169와 교호작용에 따른 평방합 71.631의 합으로 주어지지 않는다. 이들 세 성분의 합은 210.514이다. 이들 변동량 간의 차이는 모형적합방식에 따른 사영공간의 분할이 상호 직교하는 부분공간들로 구성되지 않기 때문이다.

다시말하면, 요인 R의 주효과에 따른 평방합 61.714를 구하기 위한 사영공간은 식 (3.1)에서 사영행렬 $(\boldsymbol{XX}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-)$에 의해 생성되나 요인 C의 주효과에 따른 평방합을 구하기 위한 식 (3.2)의 사영행렬 $(\boldsymbol{XX}^- - \boldsymbol{X}_c\boldsymbol{X}_c^-)$에 의한 사영공간과는 직교하지 않음을 알 수 있다. 이들 두 부분공간의 직교성은 사영행렬을 이용하여 확인할 수 있다. 요인 R의 주효과에 따른 제곱합을 구하기 위한 일차원의 사영공간은 요인 C의 주효과에 따른 제곱합을 구하기 위한 이차원의 사영공간과는 직교하지 않는 부분공간들이다.

사영공간의 중첩성을 고려하지 않는 제3종 제곱합의 계산은 모형행렬에 의해 생성되는 사영공간에서의 변동량과 일치하지 않게 된다. 따라서 변동량의 차이를 알아보기 위해 식 (3.1)의 SSR과 식 (3.2)의 SSC는 사영공간의 중첩을 허용할 때의 계산방법임을 알 수 있다. $\boldsymbol{X}_{rc}$에 의한 사영공간은 세 성분의 직교하는 부분공간들로 분할될 수 있는 유일한 사영공간이므로 $\boldsymbol{X}_{rc}\boldsymbol{X}_{rc}{}^-$의 직교분해를 이용하여 차이를 계산한다.

식 (3.5)의 SS_{tr}은 76.563으로 구해진다. SSR의 61.714와의 차이를 dR이라 두면 $dR = 14.849$이다. 식 (3.7)로부터 SS_{tc}는 90.744로 구해진다. SSC는 77.169이므로 SSC와의 차이를 dC라 두면 $dC = 13.575$로 주어진다. SS_{tr}과 SS_{tc} 그리고 $SSRC$의 합은 $\boldsymbol{y}'(\boldsymbol{XX}^- - \boldsymbol{jj}^-)\boldsymbol{y}$의 238.938이 됨을 알 수 있다. 변동요인에 따른 제3종 제곱합의 계산결과를 관련된 2차형식으로 나타내면 Table 4.2와 같이 주어진다.

Table 4.2 Type III sums of squares evaluated by quadratic forms for data in Table 4.1

source of variaton	df	SS	MS
R	1	$\boldsymbol{y}'(\boldsymbol{XX}^- - \boldsymbol{X}_r\boldsymbol{X}_r^-)\boldsymbol{y}$= 61.714	61.714
C	2	$\boldsymbol{y}'(\boldsymbol{XX}^- - \boldsymbol{X}_c\boldsymbol{X}_c^-)\boldsymbol{y}$= 77.169	38.585
RC	2	$\boldsymbol{y}'(\boldsymbol{XX}^- - \boldsymbol{X}_{rc}\boldsymbol{X}_{rc}^-)\boldsymbol{y}$=71.631	35.815
Error	10	$\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{XX}^-)\boldsymbol{y}$ =20.0	2.0
corrected total	15	$\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{jj}^-)\boldsymbol{y}$=258.938	

5. 결론

본 논문은 고정효과모형의 가정하에 제3종 제곱합을 사영에 근거하여 구하는 방법을 다루고 있다. 사영에 의한 제3종 제곱합은 모형의 적합방식으로부터 유도되는 사영행렬을 구하여 이용하게 된다. 즉, 성분별 모수벡터에 따른 변동량의 계산은 해당하는 부분공간으로의 사영을 이용하여 사영까지의 거리제곱합으로 계산하게 된다. 이때 사영까지의 거리제곱합은 $\boldsymbol{y}$의 이차형식이다. $\boldsymbol{y}$의 이차형식에서의 행렬은 부분공간에서의 사영행렬로 주어지며 멱등행렬이다. 모형행렬 $\boldsymbol{X}$로의 사영공간이 제3종 제곱합의 계산을 위한 모형의 적합방식에 따른 부분공간들로 분할될 때, $\boldsymbol{X}$로의 사영공간은 상호 직교하지 않는 중첩되는 부분공간들로 구성된다. 사영이 행해지는 부분공간들의 중첩성으로 인해 부분공간에서의 제곱합은 전체변동량의 성분들로 주어지지 않음을 다루었다. 부분공간의 중첩성에 대한 확인은 각 공간에서의 고유벡터들을 활용하여 규명할 수 있음을 논의하였다.

사영을 이용한 제3종 제곱합의 계산방식에서 사영이 행해지는 부분공간들의 비직교성은 모형의 적합방식과 관련되어 있음을 논의하였으며 부분공간들의 중첩성은 해당하는 사영행렬의 고유벡터를 이용하여 입증할 수 있음을 밝히고 있다. 또한, 모형의 적합방식에 기인한 제3종 제곱합의 계산을 위한 각 부분공간에서의 사영행렬의 형태를 파악하고 이들 사영행렬이 각 성분의 제곱합을 나타내는 이차형식의 관련 행렬임을 보여주고 있다. 변동요인별 제3종 제곱합이 해당하는 부분공간의 사영행렬을 포함하는 이차형식이기 때문에 그 행렬의 고유근과 고유벡터를 활용할 수 있음을 다루었고 제곱합의 계산에 어떻게 활용하는 가를 보여주고 있다. 변동요인에 따른 제3종 제곱합이 자료의 전체 변동량과 일치하지 않는 부분에 대해 사영이 행해지는 부분공간의 중첩성으로 설명하고 이들 중첩성을 구체적으로 규명하는데 사영행렬의 고유벡터가 활용될 수 있음을 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. (2013). Estimable functions of less than full rank linear model *Journal of the Korean Data & Information Science Society*, **24**, 333−339.

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

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

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

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.

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. (1992). *Variance components*, John Wiley and Sons, Inc., New York.

Steel, R. G. and Torrie, J. H. (1980). *Principles and procedures of statistics*, McGraw-Hill, Inc., New York.

Journal of the Korean Data & Information Science Society
2014, **25**(4), 799–805

http://dx.doi.org/10.7465/jkdi.2014.25.4.799
한국데이터정보과학회지

Type III sums of squares by projections

Jaesung Choi[1]

[1]Department of Statistics, Keimyung University

Received 11 June 2014, revised 26 June 2014, accepted 8 July 2014

Abstract

This paper deals with a method for getting the Type III sums of squares on the basis of projections under the assumption of two-way fixed effects model. For unbalanced data in general total sum of squares is not equal to the sum of componentwise Type III sums of squares. There are some differencies between two quantities. The suggested method using projections can detect where the differences occur and how much they are different. The traditional ANOVA method could not explain clearly the differences. It also discusses how eigenvectors and eigenvalues of the projection matrices can be used to get the Type III sums of squares.

Keywords: Eigenvalue, eigenvector, projection, projection matrix, Type III sums of squares.

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