응용통계연구 (2009)
22(1), 181–188

지분구조의 반복측정 자료에 대한 혼합모형

최재성[1]

[1]계명대학교 통계학과

(2008년 9월 접수, 2008년 10월 채택)

요 약

본 논문은 실험단위들의 구조적 특성으로 지분관계를 갖는 실험을 행해야 하는 경우를 가정한다. 지분계획하에서 처리를 구성하는 요인으로 반복측정 요인을 고려한다. 반복측정 요인의 수준들이 비확률화에 의해 지분구조의 실험단위들에 배정될 때, 비확률화에 따른 실험의 특성을 감안한 모형으로 복합대칭의 공분산 구조하에서 혼합효과 모형을 논의하고 있다. 처리의 일부 요인들이 시간 또는 공간상의 제약으로 인해 지분구조의 실험단위들에 임의적으로 배정될 수 없을 때, 지분구조의 실험단위들에 대한 반응값들은 어떤 구조적 상관관계를 나타내는 값들로 관측될 수 있음을 예상할 수 있다. 자료의 구조적 상관성을 고려한 공분산 구조하의 선형모형으로 확률요인과 고정요인을 포함하는 혼합효과의 모형을 제시하고 모형내 미지모수들에 대한 추론방법을 다루고 있다.

주요용어: 지분구조, 반복측정요인, 혼합효과, 복합대칭의 공분산 구조.

1. 서론

연구목적을 위한 실험 또는 관측조사에서 개체의 반응을 나타내는 실험단위들이 실험의 특성상 또는 표본추출설계로 인해 서로 다른 크기의 실험단위들로 구성되어 실험이 행해 질 때, 실험단위 또는 개체가 분할되거나 지분되어 있는 경우를 생각해 볼 수 있다. 상이한 크기의 실험단위를 갖는 실험계획들로 분할구 실험계획, 반복측정 계획 그리고 일부 지분계획 등을 들 수 있다. 이들의 특성은 개체의 반응에 영향을 미치는 처리들이 적어도 두 요인의 결합수준들로 이루어지고, 설계구조는 불완비 블럭으로 구성된다는 점이다. 실험 또는 관측조사에서 상이한 크기의 실험단위, 개체의 분할은 관측이 행해지는 작은 단위의 실험단위로 부터 얻어지는 오차성분의 분산이외에도 상이한 크기의 실험단위에 대한 변동성분을 추가적으로 구해야 함을 의미하고 있다. 서로 다른 크기의 실험단위들을 고려해야 하는 실험설계의 경우, 각 실험단위와 관련된 오차의 변동을 나타내는 분산을 추정함으로써 해당하는 효과에 대한 추론이 가능하게 된다. 서로 다른 크기의 실험단위들에 대한 오차들은 일반적으로 독립성의 가정하에 추론하게 된다. 상이한 크기를 갖는 실험계획들과 관련한 자료분석 모형과 분석방법들에 대한 논의는 Milliken과 Johnson (1984), Steel과 Torrie (1980)에서 보여진다. 둘 이상의 상이한 크기의 실험단위, 또는 개체로 실험 또는 관측조사가 행해질 때, 처리들은 적어도 두 요인의 결합수준들로 구성됨을 알 수 있다. 왜냐하면, 요인들의 수준결합으로 주어지는 일부처리들은 다른 크기의 실험단위에 배정되어야 하기 때문이다. 예로써, 두 요인의 수준결합으로 주어지는 처리의 분할구(split-plot) 실험은 두 요인들의 수준에 서로 다른 크기의 실험단위를 배정하는 실험계획이다.

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

개체 또는 실험단위의 반응에 영향을 미치는 처리들이 요인들의 수준결합으로 주어질 때, 두 가지 유형의 요인들, 고정요인과 확률요인으로 구분하게 된다. 실험단위에 행해지는 처리들이 요인들의 수준결합으로 구성되고 이들 요인이 모두 상수의 고정효과를 갖는 것으로 간주될 때, 선형모형의 가정하에 고정효과를 알아보기 위한 모형을 고정효과 모형, 또는 모수모형이라 부른다. 처리구조에 있어서 고려된 모든 요인의 수준들이 각 요인의 수준들의 모집단으로 부터 취해진 확률표본으로 간주될 때, 요인들의 수준효과를 확률효과라 한다. 개체의 반응에 확률효과들이 어떻게 영향을 주고 있는 가를 알아 보기 위한 선형모형을 확률모형 또는 분산성분 모형이라 한다. 개체 또는 실험단위의 반응에 영향을 미치는 두 가지 유형의 요인들로 처리가 주어질 때, 선형모형의 가정하에서 고정효과와 확률효과를 함께 포함시키고 있는 모형을 혼합효과모형 또는 혼합모형이라 부른다. 실험계획의 일부로 동질적인 실험단위들로 그룹화하는 블럭요인도 확률요인으로 간주되고 그 수준들의 효과를 확률효과라 하나 이는 동질적인 실험단위들로 블럭화 하기 위한 계획구조하에서 발생하는 확률효과이므로 처리구조하에서 고려하는 확률효과와는 구분할 필요가 있다고 본다. 이와 관련된 논의는 Searle 등 (1992)과 Graybill (1976)에서 다양한 모형의 제시와 함께 분석방법을 알아볼 수 있다.

반복측정 계획의 특성은 분할구 실험계획과 유사하나 처리를 구성하는 한 요인의 수준들이 비확률화에 의해 실험단위에 배정된다는 점에서 차이가 있다. 비확률화에 의한 처치의 배정은 처치가 행해진 실험단위들의 오차에 대해 독립성을 가정할 수 없게 하고, 오차간에 어떤 상관성을 띄는 공분산 구조의 가능성을 제고하게 된다. 그러므로, 실험 또는 관측조사에서 반복측정 요인이 존재하고 반복측정 요인의 수준들이 시간 또는 공간상에서 연구자가 반복측정하게 되는 동일 실험단위 또는 동일 개체내에 임의로 배정할 수 없을 때, 반복측정되는 동일 개체내 반응값들 간에 종속성을 부여할 수 있게 된다.

본 논문은 반복측정요인으로 인한 상이한 크기의 실험단위들 간의 구조가 지분관계일 때, 고정요인과 혼합요인들의 효과를 추론하기 위한 혼합모형을 제시하고 추론방법을 논의하고자 한다. 반복측정 계획들은 반복측정 요인으로 주어지는 일부 요인의 수준들이 비확률화에 의해 실험단위에 배정됨으로 인해서 실험단위들 간의 오차에 독립성을 가정할 수 없다는 점이다. 지분(nesting)이 실험단위의 설계구조로 부터 야기될 때, 분산성분과 반복측정 요인의 배치에 따른 반응값간의 공분산 구조를 고려한 모형을 구체적으로 논의해 보기로 한다.

2. 모형의 가정

연구자의 실험환경에 적합한 모형설정을 위하여 처리에 대한 개체 또는 실험단위의 반응을 변수 Y라 둔다. 개체의 반응에 영향을 미치는 독립변수로 세 개의 요인 A, B 그리고 T를 생각한다. 요인 A는 확률요인으로 가정한다. 즉, 요인 A의 수준들은 모집단을 구성하고 모집단내 임의 한 수준에서의 효과는 확률분포를 따른다고 가정한다. 따라서, 확률요인의 수준효과들은 확률효과들로 간주되고 이들 효과에 대한 분포로 $N(0, \sigma_A^2)$을 가정한다. 요인 B는 고정요인으로 $j = 1, 2, \ldots, b$개의 고정된 수준들을 대상으로 각 수준이 개체의 반응에 미치는 효과에 관심을 두고 있으므로 수준들의 효과를 고정효과 또는 모수효과라 가정한다. 요인 T는 반복측정 요인으로 $k = 1, 2, \ldots, t$개의 수준을 갖는 고정요인이라 가정한다. T는 개체의 반응 Y를 T의 t개 수준에서 반복측정을 요하는 변수이므로 t개 수준들의 효과는 고정효과들이다. 또한, 반복측정요인 T는 시간 또는 공간상에서 연구자 임의로 실험단위에 임의로 배정할 수 없는 수준들을 갖는 요인으로 가정한다. 처리를 구성하는 한 요인의 수준들이 비확률화에 의해 분할된 실험단위 또는 분할된 개체에 배정될 때, 개체의 반응값들 간에 독립성을 가정할 수 없게 된다. 다시 말하면, 개체내 또는 실험단위내 관측값들 간에 종속성을 갖게 된다. 이는, 종속성을 감안한 자료의 공분산구조를 염두에 둬야함을 의미한다. 확률요인과 고정요인을 둘다 포함하고 있는 모형의 형태는 혼합모형이다. 따라서, 자료분석을 위한 모형으로 혼합모형의 가정하에 모형의 형태를 논의하기로 한다. 추

가적으로 실험에 이용되는 실험단위 또는 개체의 설계구조에서 지분구조의 실험단위들이 존재하는 경우를 가정하기로 한다. 지분구조의 실험단위를 갖는 지분계획에서의 선형모형은 실험단위의 지분구조에서 야기되는 분산성분을 고려하고 있는 혼합모형으로 제시되어야 함을 의미한다. 지분계획하의 일반 선형모형들과 관련된 분석방법들을 Montgomery (1976)와 Milliken과 Johnson (1984)에서 구체적으로 살펴볼 수 있다.

연구자의 가정된 실험환경 조건에 부합하는 모형구축을 생각해 보기로 한다.

3. 모형에 관한 논의

조사모집단 또는 실험모집단에서 표본자료 또는 실험자료를 얻기 위한 개체 또는 실험단위의 표본추출계획으로 이원지분 계획을 가정한다. 이원 지분계획의 가정은 두 개의 서로 다른 크기의 실험단위들로 인한 오차 성분들에 대한 가정을 필요로 하게 한다. 예를 들어, 주택용 바닥재로 이용되는 다양한 재질의 몇개 상품들에 대해 일정강도하에서 내구력을 시간별로 측정하는 실험을 생각해 보기로 한다. 실험을 위한 실험단위들의 추출은 비교를 위한 바닥재의 각 상품에서 임의로 표본을 취하고 표본으로 취해진 상품을 실험에 필요한 일정크기로 분할하여 실험을 한다고 하자. 이때, 실험에 이용되는 한 특정재질의 소단위의 상품은 다른 재질의 상품의 생산단위에서 얻을 수 없는 작은 크기의 실험단위이다. 따라서, 각 재질의 소단위의 실험단위들은 해당재질의 생산단위에 지분되어 있음을 나타내고 있다. 이는 지분구조의 실험단위들로 행해지는 실험의 이원 지분계획은 서로 다른 크기의 실험단위를 제공함을 의미한다. 동일재질의 바닥재 상품중 실험용으로 추출된 상품들은 확률표본으로 주어지므로 상품간의 변이에 대해 독립성을 가정한다. 표본으로 추출된 상품의 분할로 주어지는 소단위의 실험단위들에 대한 변이 또한 변이에 대한 독립성을 가정하게 된다. 이원지분 계획으로 인한 상이한 크기의 실험단위의 인식은 이에 따른 변동을 예상하게 되고 상이한 변동은 서로 다른 크기의 실험단위에 대한 오차성분들로 표현된다. 이원 지분계획으로 구성된 서로 다른 크기의 실험단위들에 확률요인 A의 $i = 1, 2, \ldots, a$개 수준과 고정요인 B의 $j = 1, 2, \ldots, b$개 수준 그리고 반복측정 요인 T의 $k = 1, 2, \ldots, t$개 수준을 생각한다. y를 모집단내 개체 또는 실험단위의 한 특성을 나타내는 확률변수라 두자. 확률변수 y의 반응값은 세 요인들의 수준효과들인 확률효과와 고정효과들의 선형결합으로 주어진다고 가정한다. 요인 A의 수준 i에서의 수준효과를 α_i라 두면, a개의 α_i들은 확률효과들이다. 요인 B의 수준 j에서의 수준효과를 β_j라 두면, b개의 β_j들은 고정효과들을 나타낸다. 요인 T의 수준 k에서의 수준효과를 τ_k라 두면, t개의 τ_k 수준효과들도 고정효과들을 나타낸다. 요인 A의 수준 i, 요인 B의 수준 j 그리고 요인 T의 수준 k에서 개체의 반응변수 y에 대한 반응을 y_{ijk}라 두자. 이원 지분계획을 고려하고 있으므로 요인 A의 a개 수준과 요인 B의 b개 수준들은 큰 실험단위에서 측정된다고 가정하고 반복측정 요인 T의 t개 수준들은 작은 실험단위에서 측정된다고 가정한다. 두 개의 서로 다른 크기의 실험단위의 오차성분들을 포함하고 두 유형의 효과들을 고려한 혼합모형은 다음과 같다.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{j(i)} + \tau_k + (\alpha\tau)_{ik} + (\beta\tau)_{jk} + \epsilon_{k(ij)}. \tag{3.1}$$

식 (3.1)에서 μ는 전체의 평균을 나타내고 α_i는 확률요인 A의 i번째 수준의 확률효과를 나타낸다. β_j는 고정요인 B의 수준 j의 고정효과이다. $(\alpha\tau)_{ik}$, $(\beta\tau)_{jk}$는 두 요인 A와 B의 교호작용을 나타낸다. $\delta_{j(i)}$는 큰 실험단위에서의 오차성분을 나타내고, $(\alpha\beta)_{ij}$와 교락되어(confounded) 있다. $\epsilon_{k(ij)}$는 작은 실험단위의 오차항을 나타낸다. 세 요인 교호작용을 나타내는 $(\alpha\beta\tau)_{ijk}$와 교락되어 있다고 가정한다.

두개의 서로 다른 실험단위들이 지분구조를 가질 때, 일반적인 가정은 분산성분을 나타내는 δ_{ij}와 $\epsilon_{k(ij)}$는 서로 독립이고 각기 $N(0, \sigma_\delta^2)$와 $N(0, \sigma_\epsilon^2)$인 분포를 따른다는 가정하에 해당하는 분산을 추론하

는데 목적이 있다. 그러나, 반복측정 요인 T의 t개 수준들이 비확률화(nonrandomiztation)에 의해 지분구조의 작은 실험단위에 행해짐으로 인해서 작은 실험단위에서의 관련된 t개 오차들은 오차간에 일정한 상관성을 갖는 오차들로 관측됨을 예상할 수 있다. 따라서, 이러한 현상들을 고려할 때, 지분구조의 설계구조와 관련된 모형의 일반적인 가정을 할 수 없게 된다. 즉, 작은 실험단위에서의 오차간의 공분산 구조하에 모형이 제안되어야 함을 의미하고 있다. 따라서, 오차벡터를 ϵ라 두고, $\epsilon=[\epsilon_1, \epsilon_2, \ldots, \epsilon_n]'$으로 정의한다. 여기서, n은 실험에서 관측된 반응값들의 수를 나타낸다. 모형 (3.1)에서 분산성분과 오차벡터의 공분산 구조를 고려한 가정으로 α_i는 $N(0, \sigma_\alpha^2)$, $\delta_{j(i)}$는 $N(0, \sigma_\delta^2)$, $(\alpha\tau)_{ik}$는 $N(0, \sigma_{\alpha\tau}^2)$, 상관성을 갖는 t개 오차들에 대한 분포로 $\text{MVN}(\mathbf{0}, \Sigma_t)$를 가정한다.

지분구조의 실험단위를 갖는 모형 (3.1)에서 큰 실험단위에 행해지는 두 요인 A와 B의 수준결합으로 주어지는 처리들 (a_i, b_j), $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, 각각에 다수의 실험단위들이 이용가능 할 때, 이에 해당하는 모형은 다음과 같이 표현된다.

$$y_{ijlk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_{l(ij)} + \tau_k + (\alpha\tau)_{ik} + (\beta\tau)_{jk} + (\alpha\beta\tau)_{ijk} + \epsilon_{k(ijl)}. \tag{3.2}$$

모형 (3.2)에서의 변화는 확률요인 A와 고정요인 B의 수준들의 결합으로 주어지는 개체단위의 처리들이 큰 실험단위인 개체에서 여러 번 관측될 때의 변이를 나타내는 오차항 $\delta_{l(ij)}$와 추정이 가능한 교호작용을 나타내는 항들이다. 즉, 실험의 반복수가 증가함으로써 모형 (3.1)과는 달리 오차성분과 교락된 교호작용의 항도 추가된다. 확률요인의 확률효과와 고정요인의 고정효과와의 교호작용은 확률효과로 간주되고 분산성분을 추정하게 된다. 그러나 확률효과와 오차항의 교호작용은 가능하지 않다. 왜냐하면, 교호작용은 개체의 반응에 영향을 미치는 처리를 구성하는 요인들의 효과만으로 정의되기 때문이다.

모형 (3.1)의 행렬표현은 다음과 같이 주어진다.

$$\boldsymbol{y} = \boldsymbol{X_1\beta} + \boldsymbol{X_2u_1} + \boldsymbol{X_3u_2} + \boldsymbol{Zv} + \boldsymbol{\epsilon}. \tag{3.3}$$

식 (3.3)에서 $\boldsymbol{X_1}$은 $n \times p$인 모형행렬을 나타낸다. $\boldsymbol{\beta}$는 $p \times 1$개의 모수들을 나타내는 모수벡터이다. $\boldsymbol{X_2}$는 확률요인 A의 a개 수준들의 확률효과들과 관련된 $n \times a$ 계수행렬이다. $\boldsymbol{u_1}$은 확률요인 A의 a개 수준효과들을 성분으로 갖는 확률벡터이다. 모형 (3.1)에서의 $\delta_{j(i)}$는 $\boldsymbol{u_1}$을 구성하고 있다. $\boldsymbol{X_3}$는 확률요인 A와 고정요인 B의 교호작용을 나타내는 확률효과들과 관련된 $n \times ab$ 계수행렬이다. $\boldsymbol{u_2}$는 확률요인 A와 고정요인 B의 ab개 수준결합에서 주어지는 교호작용을 나타내는 성분들의 확률벡터이다. $\boldsymbol{Z}$는 큰 실험단위에서의 오차성분을 나타내는 $n \times b$ 계수행렬이다. $\boldsymbol{v}$는 b개의 성분을 갖는 확률오차들의 벡터이다. 모형 (3.1)에서의 $\epsilon_{k(ij)}$들로 구성되는 $\boldsymbol{\epsilon}$은 n개의 오차성분을 나타내는 오차벡터이다. 행렬표현에서 확률요인 A의 a개 수준들의 확률효과들을 성분으로 갖는 확률벡터 $\boldsymbol{u_1}$과 확률요인 A와 고정요인 B의 교호작용을 나타내는 확률효과들의 벡터 $\boldsymbol{u_2}$가 주어졌을 때,

$$E(\boldsymbol{y}) = E[E(\boldsymbol{y}|\boldsymbol{u_1}, \boldsymbol{u_2})] = E(\boldsymbol{X_1\beta} + \boldsymbol{X_2u_1} + \boldsymbol{X_3u_2}) = \boldsymbol{X_1\beta}$$

로 주어지며,

$$\text{Var}(\boldsymbol{y}) = \boldsymbol{X_2}(\sigma_\alpha^2 \boldsymbol{I}_{a\times a})\boldsymbol{X_2'} + \boldsymbol{X_3}(\sigma_{(\alpha\tau)}^2 \boldsymbol{I}_{ab\times ab})\boldsymbol{X_3'} + \boldsymbol{Z}(\sigma_\delta^2 \boldsymbol{I}_{b\times b})\boldsymbol{Z'} + \boldsymbol{R}$$

로 표현된다.

모형 (3.1)에서의 가정들을 행렬식으로 표현한 식 (3.3)에서의 가정들로 표현하면, $\boldsymbol{u_1}$은 $\text{MVN}(\mathbf{0}, \sigma_\alpha^2 \boldsymbol{I}_{a\times a})$, $\boldsymbol{u_2}$는 $\text{MVN}(\mathbf{0}, \sigma_{(\alpha\tau)}^2 \boldsymbol{I}_{ab\times ab})$, $\boldsymbol{v}$는 $\text{MVN}(\mathbf{0}, \sigma_\delta^2 \boldsymbol{I}_{b\times b})$이고 $\boldsymbol{\epsilon}$은 $\text{MVN}(\mathbf{0}, \boldsymbol{R})$로 나타낼 수 있게 된다. 단, R은 오차벡터 $\boldsymbol{\epsilon}$의 공분산행렬을 의미한다. 반복측정 요인 T의 t개 수준들이 지분구조의 작은

실험단위들에 임의로 배정될 때 오차간의 관련성은 없다는 일반적인 가정하에 ϵ에 대한 다변량 확률분포로 MVN($\mathbf{0}$, $\sigma_\epsilon^2 \boldsymbol{I}_n$)을 가정할 수 있으나 비확률화에 의해 실험단위에 배정될 때 가정은 성립하지 않게 되고 비확률화로 인해 발생하게 되는 오차간의 상관성을 고려한 공분산 구조의 행렬로 표현된다. 다시 말하면, 지분구조의 실험단위를 갖는 실험에 의해 자료를 분석할 때, 단순히 상이한 크기의 실험단위로부터 발생하는 분산성분들을 고려한 분석이 행해지게 되나, 처리구조에 있어 반복측정 요인의 존재는 공분산 구조하의 좀더 복잡한 분석을 필요로 하게 된다. 이는 단순히 혼합모형에 내포된 확률효과들의 분산성분이나 상이한 크기의 실험단위들의 분산성분들에 대한 추론보다는 오차간에 내재한 상관성을 고려한 공분산 구조하에 처리효과들을 비교할 때, 추론에 정확성을 더할 수 있음을 예상할 수 있다. 그러므로, 설계구조의 지분계획이 반복측정 요인에 의한 반복측정 계획으로 행해질 때, 야기될 수 있는 공분산 구조를 감안한 모형제시가 필요하게 된다.이러한 공분산 구조는 지분구조의 다른 크기의 실험단위에 행해지는 처리들에 대해서도 실험의 다른 요인에 의해 주어질 수 있음에 유의할 필요가 있다.

행렬식 (3.3)으로의 표현은 실험 모집단 또는 관측 모집단에서 처리가 행해진 실험단위, 혹은 개체의 특성에 대한 추론을 위해 일반 선형모형으로 혼합모형의 근거하에 분석을 하는 경우 자료의 표현으로 행렬식의 표현은 간편하고 효율적인 방법으로 분석할 수 있음을 보여주고 있다.

오차벡터 ϵ의 공분산 구조로 Huynh-Feldt(H-F)의 조건을 만족시키는 복합대칭(compound symmetry)의 구조를 가정한다. Huynh과 Feldt (1970)는 공분산 구조의 좀 더 일반적인 조건을 다루고 있으나 복합대칭 구조와 오차의 독립성은 Huynh-Feldt(H-F)의 조건이 성립하는 경우들이다. 상관성을 갖는 반복측정 요인 T의 t개 오차들에 대한 복합대칭 구조의 공분산 행렬은 다음과 같다.

$$\boldsymbol{\Sigma}_t = \sigma_\epsilon^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}$$

따라서, 공분산 행렬 R은 $\boldsymbol{\Sigma}_t$를 대각원소로 갖는 블럭 대각행렬로 주어진다. 즉, 개체간의 변이는 독립이고 개체내의 오차는 가정된 상관성을 갖게 됨을 나타내고 있다. 행렬식 (3.3)에서 추정되어야 할 모수들은 고정효과 벡터 $\boldsymbol{\beta}$, 확률효과 벡터들의 분산성분들 σ_α^2, $\sigma_{(\alpha\tau)}^2$과 오차성분 σ_δ^2 그리고 공분산 구조의 σ_ϵ^2이다. 이들 모수들을 추론하기 위한 방법으로 Corbeil과 Searle (1976)에 의한 REML(restricted maximum likelihood)방법을 이용하기로 한다. REML 방법은 관측값들의 우도함수를 고정효과를 나타내는 부분과 고정효과를 제외한 확률효과들 만을 포함하는 두 부분으로 나누고 있다. 고정효과가 제외된 확률효과만의 주변우도는 모형의 고정효과들에 종속되지 않는 제한된 우도함수(restricted likelihood function)를 나타낸다. 이 우도함수를 최대로 하는 모수추정량들이 분산성분들의 REML 추정량들이다.

4. 방염제 자료의 예

어떤 가연성 소재에 화기를 억제할 수 있는 다수의 방염제가 이용될 수 있다고 가정하자. 가연성 소재에 이용된 방염제(A)의 효과는 큰 차이가 없다고 가정한다. 그러나 방염제의 접착기술(B)에 따른 효과는 일정한 차이가 있다고 가정할 때, 시간(T)별 방염효과의 정도를 관측하는 실험을 행한다고 가정한다. 연구대상의 가연성 소재 또한 다양한 재질로 제조된다고 가정한다. 실험을 위해 이용되는 다양한 재질의 일정 크기의 가연성 소재를 이용한다고 하자. 다수의 방염제중 임의로 추출한 세 종류의 방염제를 이용하여 비교하고자 하는 두 가지 접착기술로 가연성 소재에 방염처리를 한다. 방염처리된 일정크기의

표 4.1. 방염제 자료의 생성표

소재	요인A	요인B	요인T		
			t_1	t_2	t_3
1	$a1$	$b1$	32.0	33.0	45.0
2	$a1$	$b2$	10.8	12.5	20.1
3	$a2$	$b1$	25.0	35.8	48.0
4	$a2$	$b2$	23.0	28.0	35.0
5	$a3$	$b1$	19.8	25.8	35.0
6	$a3$	$b2$	24.0	25.8	29.1
7	$a1$	$b1$	25.0	28.0	39.8
8	$a1$	$b2$	19.8	25.8	33.1
9	$a2$	$b1$	22.0	30.8	43.1
10	$a2$	$b2$	20.0	22.0	29.8
11	$a3$	$b1$	18.8	29.8	33.1
12	$a3$	$b2$	12.0	22.8	26.0

소재를 세 부분으로 나누어 세 수준의 일정시간 동안 화염에 두고 방염처리된 소재의 내연성 대한 측정 값으로 마모정도(y)를 관측한다고 하자. 이 경우에 실험은 두 개의 서로 다른 크기의 실험단위를 이용하게 된다. 즉, 실험에 이용되는 일정크기의 가연성 소재와 시간별 측정용으로 이용하기 위해 작은 크기로 나누어진 소재이다. 실험에 이용되는 가연성 소재는 다양한 재질로 제조되기 때문에 작은 단위의 소재와 큰 단위의 소재는 지분관계의 실험단위들로 간주된다. 방염제의 세 수준은 방염제의 모집단에서 임의로 추출된 세 수준, $a_i\,(i = 1,2,3)$는 확률변수들로 간주되므로 이들의 효과 $\alpha_i\,(i = 1,2,3)$들은 확률효과들로 간주된다. 두 가지 접착기술은 고정요인의 두 수준 $b_j\,(j = 1,2)$를 나타내므로 이들 수준효과 $\beta_j\,(j = 1,2)$는 고정효과를 나타낸다. 반복측정 요인인 시간 T의 세 수준 $t_k\,(k = 1,2,3)$의 효과를 나타내는 $\tau_k\,(k = 1,2,3)$는 고정효과들이다. 일정크기의 가연성 소재는 다양한 재질로 주어지기 때문에 작은 크기로 나누어진 소재들도 해당하는 재질의 특성을 갖는 지분구조의 실험단위들 이다. 반복측정 요인 시간 T의 세 수준이 비확률화에 의해 소단위의 소재에 배정됨으로 인해서 재질의 특성을 갖는 소 단위의 소재간에 일정한 상관성을 띄는 공분산 구조를 갖는다고 예상할 수 있다. 확률요인 A와 고정요인 B의 수준결합으로 주어지는 처치가 임의로 배정된 일정 크기의 소재에 행해지고 고정요인 T의 세 수준들이 비확률화에 의해 배정된다고 하자. 다음은 위의 실험으로 부터 수집된 생성자료표를 나타내고 있다.

위 자료를 분석하기 위한 모형으로 식 (3.2)를 가정할 때, 자료에 해당하는 모형은 다음과 같이 주어진다.

$$y_{ijlk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_{l(ij)} + \tau_k + (\alpha\tau)_{ik} + (\beta\tau)_{jk} + (\alpha\beta\tau)_{ijk} + \epsilon_{k(ijl)}, \qquad (4.1)$$

단, $i = 1,2,3$, $j = 1,2$, $l = 1,2,\ldots,12$이고 $k = 1,2,3$이다. y_{ijlk}는 확률요인 A의 수준 i와 고정요인 B의 수준 j의 결합으로 주어지는 처리 (i,j)가 행해진 가연성 소재 j내 시간요인 T의 수준 t_k에서 관측된 값을 나타낸다.

모형내 모수들을 추정하기 위해 REML 방법을 이용한다. 관측벡타 $\mathbf{y}$의 우도함수는 적절한 변수변환을 이용할 때, 고정효과를 포함하는 부분과 고정효과와 상관없는 두 부분의 결합우도함수로 주어진다. 고정효과를 포함하지 않는 변환벡타의 주변우도함수는 고정효과와 무관한 분산성분들만의 제한우도함수(restricted likelihood function)이고 모형내 분산성분들의 REML 추정값은 제한우도함수를 이용하여 구해진다. 모형 (4.1)내 분산성분들의 REML 추정값은 다음과 같다.

$\hat\sigma^2_\alpha = 0.5670$, $\hat\sigma^2_{\alpha\beta} = 0.5127$, $\hat\sigma^2_{\alpha\beta\tau} = 2.8435$, $\hat\sigma^2_\delta = 19.4111$이고 $\hat\sigma^2_\epsilon = 3.9556$이다.

모형내 고정효과를 나타내는 모수들의 추정값과 추정오차는 다음과 같이 주어진다.

$\hat{\mu} = 28.8500(2.2808)$, $\hat{\beta}_1 = 11.8167(3.1664)$, $\hat{\tau}_1 = -10.5833(1.7928)$, $\hat{\tau}_2 = -6.0333(1.7928)$, $(\widehat{\beta\tau})_{11} = -6.3167(2.5354)$이고 $(\widehat{\beta\tau})_{12} = -4.1000(2.5354)$이다.
그러나, 위 자료를 분석하기 위해 모형 (4.1)을 적합시켰을 때, 오차항의 분산성분을 제외한 나머지 네개의 분산성분들에 대한 Wald Z 검정통계량의 관측값들은 유의수준 0.05에서 모두 유의하지 않으므로 아래의 축소된 모형을 적합시켜 보기로 한다.

$$y_{jlk} = \mu + \beta_j + \delta_{l(j)} + \tau_k + (\beta\tau)_{jk} + \epsilon_{k(jl)}. \tag{4.2}$$

위 모형의 적합성에 대한 우도비 검정통계량의 관측값은 ${\chi^2}_{(1)} = 19.68$로 관측되고 p값은 0.0001 미만이므로 모형의 유의성을 나타내고 있다. 모형 (4.2)에 따른 분산성분들과 모수에 대한 추정값들은 다음과 같다.
두 분산성분들에 대한 추정값들은 각기 $\hat{\sigma}^2_\delta = 20.2748$이고 $\hat{\sigma}^2_\epsilon = 6.2304$이다. 고정효과들에 대한 모수 추정값들은 $\hat{\mu} = 28.88500(2.1018)$, $\hat{\beta}_1 = 11.8167(2.9724)$, $\hat{\tau}_1 = -10.5833(1.4411)$, $\hat{\tau}_2 = -6.0333(1.4411)$, $(\widehat{\beta\tau})_{11} = -6.3167(2.0380)$이고 $(\widehat{\beta\tau})_{12} = -4.1000(2.0380)$으로 구해진다.

5. 결론

본 논문은 지분구조의 실험단위들에 처리로 구성되는 요인들 중 일부요인이 반복측정 요인이고, 반복측정 요인의 수준이 지분구조의 실험단위들에 비확률화로 배정될 때 발생할 수 있는 관측반응들 간의 상관성에 착안하고 있다. 개체의 반응에 대한 독립성의 가정이 성립하지 않을 때, 선형모형에 대한 일반적인 가정은 성립하지 않게 된다. 따라서, 실험단위들에 반복측정 요인을 포함한 여러 요인들의 수준결합이 처리로 행해질 때, 야기될 수 있는 관측값들의 상관성을 다룬 공분산 구조가 논의되고 있다. 또한, 실험자료를 얻기 위한 실험계획이 서로 다른 크기의 실험단위를 필요로 하는 실험설계일 때, 이로 인한 오차성분들을 다루고 있다. 그리고 처리요인으로 고정요인뿐만 아니라 확률요인도 고려함으로 인해서 고정효과및 확률효과를 포함하는 혼합효과 모형을 가정하고 있다. 가정된 실험환경의 조건하에서 혼합효과 모형구축에 관한 논의와 함께 제시된 모형으로 자료분석할 수 있는 적합한 예를 통하여 구체적으로 모형구축 과정과 모형내 미지모수들을 추론하는 방법을 설명하고 있다. 관측값들의 상관성에 따른 공분산 구조로 복합대칭의 공분산 구조를 논의하고 있다. 더불어, 실험단위의 지분구조로 실험을 행하는 지분계획하에서 반복측정 요인이 이용될 때, 실험계획은 공분산 구조를 갖는 반복측정 계획으로 인식되어야 함을 밝히고 있다.

참고문헌

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 & Sons, New York.

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

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

A Mixed Model for Nested Structural Repeated Data

Jaesung Choi[1]

[1]Dept. of Statistics, Keimyung University

(Received September 2008; accepted October 2008)

Abstract

This paper discusses the covariance structures of data collected from an experiment with a nested design structure, where a smaller experimental unit is nested within a larger one. Due to the nonrandomization of repeated measures factors to the nested experimental units, compound symmetry covariance structure is assumed for the analysis of data. Treatments are given as the combinations of the levels of random factors and fixed factors. So, a mixed-effects model is suggested under compound symmetry structure. An example is presented to illustrate the nesting in the experimental units and to show how to get the parameter estimates in the fitted model.

Keywords: Nested structure, repeated measures factor, mixed effect, covariance structure.

[1]Professor, Dept. of Statistics, Keimyung University, 1000 Shindang-Dong, Dalseo-Gu, Daegu 704-701, Korea. E-mail: jschoi@kmu.ac.kr