

불균형 일원랜덤효과모형에서 EQDGs를 이용한 MIVQUE 추정량 비교

정병철¹⁾

요약

본 논문에서는 불균형 일원랜덤효과모형에서 분산성분에 대한 MIVQUE 추정량들의 효율에 대하여 연구하였다. MIVQUE 추정량에 대한 초기 추정치로 3가지 추정치가 사용되었으며 이들을 이용하여 얻어지는 MIVQUE 추정량의 효율성을 척도화된 추정량의 경험적 분위수를 이용하는 EQDGs 플롯을 이용하여 비교하였다. 모의실험 결과 집단간 분산의 초기치를 0으로 하고 집단내 분산의 초기치를 1로 사용한 MIVQUE 추정량의 효율이 추정량의 안정성 관점에서 다른 초기치를 사용한 MIVQUE 추정량에 비하여 약간 효율적인 것으로 나타났다.

주요용어: MIVQUE 추정량, 분산성분 모형, 초기치

1. 서론

분산성분모형은 농업실험, 공업실험 및 의학자료 등을 분석하는데 있어서 흔히 사용되는 모형으로, 이 모형의 주된 관심사항은 분산성분의 추정에 있다. 만일 균형자료일 경우 분산성분에 대한 ANOVA 추정량은 최소분산 불편추정량의 성질 등 추정량의 좋은 특성들을 갖는다. 그러므로 균형자료에서는 분산성분에 대한 ANOVA 추정량의 사용이 매우 타당하게 된다. 하지만 불균형 자료인 경우에 ANOVA 추정량은 불편추정량이라는 성질을 제외한 다른 성질을 잃어버리게 된다 (Searle, 1971; Searle 등, 1992). 이와 같은 이유때문에 불균형 자료를 갖는 분산성분모형에는 여러 추정방법들이 제안되었으며, 각 추정량들의 효율성에 대한 많은 연구가 진행되어 왔다 (Swallow와 Monahan, 1984; Corbeil과 Searle, 1976; Swallow와 Searle, 1978, Swallow, 1981 등).

본 연구에서는 불균형 일원랜덤효과모형에서 분산성분에 대한 여러 추정방법중에서 Rao (1971, 1973)에 의하여 제안된 최소노름 이차불편추정량(Minimum Norm Quadratic Unbiased Estimator; MINQUE) 추정량의 효율성에 대하여 알아보려고 한다. 만일 오차항이 정규분포를 따른다고 가정하면 MINQUE 추정량은 최소분산 이차불편추정량(Minimum Variance Quadratic Unbiased Estimator, MIVQUE)와 동일하므로 본 연구에서는 오차항의 정규분포를 가정한 MIVQUE 추정량에 대하여 다루었다. 그러나 분산성분에 대한 MIVQUE 추정량을 얻기 위해서는 항상 초기추정치가 필요하다. Swallow와 Monahan (1984)은 ANOVA 추정량을 초기추정치로 사용한 MIVQUE 추정량과 집단간 분산을 0, 집단내 분산을 1을 사

1) (136-742) 서울시 성북구 동선동 249-1 성신여자대학교 통계학과 초빙교수
E-mail: bcjung@empal.com

용한 초기추정치 등 두 가지 형태의 MIVQUE 추정량을 고려하였으며, Swallow와 Monahan (1978)은 초기추정치로 집단간 분산과 집단내 분산을 모두 1을 이용한 MIVQUE 추정량의 효율성에 대하여 연구하였다. 이들의 연구는 모두 추정량의 평균제곱오차(mean squared error, MSE)나 추정량의 분산-공분산행렬의 행렬식 (determinant)이나 트레이스 (trace)와 같은 단일 값 기준에 의해서 이루어졌다. 하지만 이들의 연구와 같은 단일 값에 의한 비교는 알려지지 않은 분산성분에 대한 사전지식을 요구한다.

이와 같은 단일값 비교의 한계를 극복하고자 Khuri (1997)는 분산성분 추정량의 분위수를 이용할 것을 제안하였다. 이는 QDGs (Quantile Dispersion Graphs)라 불리는 분위수 그래프를 이용하는 것으로 분산성분에 대한 사전정보를 요구하지 않는 비교방법이다. QDGs는 주어진 모수공간내에서 “0에서 1까지 각 확률값에 따른” 척도화된 추정량의 분위수의 최소값과 최대값을 그려봄으로써 추정방법들을 비교하는 방법론이다. 만일 분산성분의 추정량에 대한 정확한 분포 및 분위수가 알려지지 않는 경우라면 경험적 QDGs(Empirical QDGs, EQDGs)를 사용할 수 있다(Lee와 Khuri, 1999, 2000; 송석현과 정병철, 2002).

본 연구에서는 불균형 일원랜덤효과모형에서 초기값에 따른 3가지 MIVQUE 추정량의 효율성을 EQDGs를 이용하여 비교하고자 한다.

2. 모형 및 MIVQUE 추정량

다음과 같은 불균형 일원랜덤효과모형을 고려하자.

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \dots, k, j = 1, \dots, n_i, \quad (2.1)$$

여기서 μ 는 알려지지 않은 모수이고, α_i 는 *i.i.d.* $N(0, \sigma_\alpha^2)$ 을 따르는 i 번째 처리효과를 나타내며 e_{ij} 는 *i.i.d.* $N(0, \sigma_e^2)$ 을 따르며, α_i 와 e_{ij} 는 서로 독립이라 가정하자. 식 (2.1)을 벡터 형태로 표현하면 다음과 같다.

$$\mathbf{y} = \mu \mathbf{i}_N + \left(\bigoplus_{i=1}^k \mathbf{i}_{n_i} \right) \boldsymbol{\alpha} + \mathbf{e}, \quad (2.2)$$

여기서 $N = \sum_{i=1}^k n_i$, $\mathbf{y}' = (y_{11}, \dots, y_{1n_1}, \dots, y_{kn_k})$ 과 $\boldsymbol{\alpha}' = (\alpha_1, \dots, \alpha_k)$ 이며 \mathbf{e} 는 \mathbf{y} 와 유사하게 정의된다. \mathbf{i}_N 과 \mathbf{i}_{n_i} 는 각각 차수가 N 과 n_i 인 모든 원소가 1로 구성된 벡터를 나타내며 \bigoplus 는 직합 (direct sum)을 나타낸다. 이와 같은 정의하에서 \mathbf{y} 의 평균은 $\mu \mathbf{i}_N$ 이며 분산-공분산행렬은 다음과 같이 구해진다.

$$\Sigma = \sigma_\alpha^2 \left(\bigoplus_{i=1}^k \mathbf{J}_{n_i} \right) + \sigma_e^2 \mathbf{I}_N, \quad (2.3)$$

여기서 \mathbf{J}_{n_i} 는 모든 원소가 1로 이루어진 $n_i \times n_i$ 행렬을 나타내며 \mathbf{I}_N 은 차원이 N 인 단위 행렬을 나타낸다.

식 (2.1)과 같은 모형하에서 분산성분에 대한 MIVQUE 추정량을 얻기위하여 먼저 다음과 같은 값들을 정의하자 (Swallow, 1981).

$$b_i = \frac{n_i}{\sigma_{e0}^2 + n_i \sigma_{\alpha 0}^2}, \quad b = \frac{1}{\sum_i b_i}, \quad \bar{y}_i = \sum_j y_{ij} / n_i,$$

여기서 σ_{e0}^2 과 $\sigma_{\alpha 0}^2$ 은 각각 σ_e^2 과 σ_α^2 에 대한 초기추정량을 나타낸다. 이와 같은 정의하에서 오차항의 정규분포 가정하에서 얻어지는 σ_e^2 과 σ_α^2 에 대한 MIVQUE 추정량은 다음과 같이 얻어진다 (Swallow, 1981; Swallow와 Searle, 1978).

$$\begin{aligned}\hat{\sigma}_e^2 &= \frac{1}{s_{11}^0 s_{22}^0 - (s_{12}^0)^2} [-s_{12}^0 u_1^0 + s_{11}^0 u_2^0] \\ \hat{\sigma}_\alpha^2 &= \frac{1}{s_{11}^0 s_{22}^0 - (s_{12}^0)^2} [s_{22}^0 u_1^0 - s_{12}^0 u_2^0],\end{aligned}\quad (2.4)$$

여기서

$$\begin{aligned}s_{11}^0 &= \sum b_i^2 - 2b \sum b_i^3 + b^2 (\sum b_i^2)^2, \\ s_{12}^0 &= \sum \frac{b_i^2}{n_i} - 2b \sum \frac{b_i^3}{n_i} + b^2 \sum b_i^2 \sum \frac{b_i^2}{n_i}, \\ s_{22}^0 &= \frac{N-k}{\sigma_{e0}^4} + \sum \frac{b_i^2}{n_i^2} - 2b \sum \frac{b_i^3}{n_i^2} + b^2 (\sum \frac{b_i^2}{n_i})^2, \\ u_1^0 &= \sum_i b_i^2 [\bar{y}_i - b \sum b_i \bar{y}_i]^2, \\ u_2^0 &= \frac{1}{\sigma_{e0}^4} \left[\sum_i \sum_j y_{ij}^2 - \sum_i n_i \bar{y}_i^2 \right] + \sum_i \frac{b_i^2}{n_i} [\bar{y}_i - b \sum b_i \bar{y}_i]^2.\end{aligned}\quad (2.5)$$

식 (2.4)에서 얻어진 MIVQUE 추정량은 항상 추정량의 초기값(σ_{e0}^2 와 $\sigma_{\alpha 0}^2$)에 의존하게 된다. 이때 주로 사용되는 초기값들로는 다음과 같은 3가지를 들 수 있다.

첫 번째로는 $\sigma_{e0}^2 = 1$ 과 $\sigma_{\alpha 0}^2 = 0$ 를 초기추정치로 사용하는 경우이다 (Swallow, 1984). 이는 SAS의 VARCOMP프로시저를 이용하여 MIVQUE 추정량을 구할때 디폴트로 사용되는 값으로 가장 간단하게 얻을 수 있는 초기치이다. 식 (2.5)와 (2.4)에 이들 값을 대입하여 얻어지는 MIVQUE 추정량을 MV1이라 정의하자.

두 번째 초기추정치는 $\sigma_{e0}^2 = 1$ 과 $\sigma_{\alpha 0}^2 = 1$ 를 사용하는 경우이다. Swallow와 Searle (1978)은 이들 초기값을 이용하는 것이 MIVQUE 추정량에 대한 분산의 계산이 편리하고 참 분산성분의 값을 모르는 경우 적절한 값이라는 관점에서 이러한 초기추정치를 사용하였다. 식 (2.5)와 (2.4)에 이들 값을 대입하여 얻어지는 MIVQUE 추정량을 MV2라 정의하자.

마지막으로 고려할 초기추정치는 분산성분에 대한 ANOVA 추정량을 초기추정치로 사용하는 경우이다. 이때 분산성분에 대한 ANOVA 추정량은 다음과 같다.

$$\begin{aligned}\hat{\sigma}_e^2(A) &= \frac{1}{N-k} \sum_i \sum_j (y_{ij} - \bar{y}_i)^2, \\ \hat{\sigma}_\alpha^2(A) &= \frac{1}{N - \sum_i n_i^2/N} \left[\sum_i n_i (\bar{y}_i - \bar{y}_..)^2 - (k-1) \hat{\sigma}_e^2(A) \right].\end{aligned}\quad (2.6)$$

식 (2.6)에 나타난 추정치를 식 (2.5)와 (2.4)에 이와 같은 값을 대입하여 얻어지는 MIVQUE 추정량을 MV3라 정의하자.

Swallow와 Monahan (1984)은 MV1과 MV3의 효율성에 대한 비교를 각 추정량의 평균 제곱오차 (Mean Square Error, MSE)와 편의 (bias)를 이용하여 실시하였다. 그들의 연구 결과 σ_α^2 가 작은 경우, 즉 $\sigma_\alpha^2/\sigma_\epsilon^2 \leq 1.0$ 인 경우, MV1과 MV3는 효율성에 큰 차이가 없었다. 하지만 $\sigma_\alpha^2/\sigma_\epsilon^2 > 1.0$ 인 경우 MV1의 효율성은 MV3에 비하여 떨어지는 것으로 나타났다. 또한 Swallow와 Searle (1978) 및 Swallow (1981)는 MV2의 분산과 ANOVA 추정량의 분산비를 이용하여 MV2의 효율성에 대한 연구를 실시하였다. 하지만 3가지 초기추정치를 이용하여 얻어지는 MIVQUE 추정량의 효율성에 대한 연구는 없었다. 더불어 MSE와 같은 단일값 비교의 한계를 극복하고자 본 연구에서는 추정량 비교에 그래프적 접근방법을 이용하고자 한다.

3. EQDGs를 이용한 추정량 비교

Lee와 Khuri (1999)는 불균형 일원랜덤효과모형에서 ANOVA 추정량과 ML추정량의 효율성에 대한 연구를 QDGs와 EQDGs를 이용하여 수행하였다. 이 절에서는 초기값에 따른 MIVQUE 추정량의 효율성을 Lee와 Khuri (1999)에 의한 EQDGs를 이용하여 파악하고자 한다. 모형 (2.1)에 대한 분산성분의 추정량은 분산성분의 초기값 뿐아니라 자료의 불균형의 정도에도 의존하게 되므로 불균형의 정도가 서로 다른 3가지 형태의 디자인을 사용하여 그 효율성을 비교하고자 한다. 먼저 일반성을 잃지 않고 $\mu = 0$ 라 하면 (2.1)의 모형은 다음과 같이 다시 쓸 수 있다.

$$y_{ij} = \alpha_i + e_{ij}, \quad i = 1, \dots, k, \quad t = 1, \dots, n_i. \quad (3.1)$$

만일 식(3.1)에서 $\phi = \sigma_\alpha^2/\sigma_\epsilon^2$ 이라 놓는다면 $\alpha_i \sim iid N(0, \phi\sigma_\epsilon^2)$ 이고 $e_{ij} \sim iid N(0, \sigma_\epsilon^2)$ 이다.

2절에서 설명한 초기추정량에 따른 MIVQUE 추정량들의 효율성을 비교하기 위하여 총 표본의 수 N 을 25로 고정된 상태에서 거의 균형된 디자인(D_1), 약간 불균형된 디자인(D_2) 및 아주 심하게 불균형된 디자인(D_3) 등 다음과 같은 3가지 형태의 불균형 디자인을 사용하였다 (Lee와 Khuri, 1999; Searle 등, 1992).

$$\begin{aligned} D_1 &= \{4, 5, 5, 5, 6\} \\ D_2 &= \{1, 1, 1, 11, 11\} \\ D_3 &= \{1, 1, 1, 1, 21\}. \end{aligned} \quad (3.2)$$

식 (3.2)에서 볼 수 있듯이 고려된 3가지 디자인은 모두 그룹의 수 $k = 5$ 이며 총 표본수 $N = 25$ 로 고정되었으므로 각 디자인과 추정방법에 대한 비교가 가능하다. 이와 같은 디자인에서 Ahrens and Pincus (1981)에 의하여 제안된 불균형의 정도에 대한 척도, 즉, $\nu(D) = \left(\sum_{i=1}^k n_i\right)^2 / k \sum_{i=1}^k n_i^2$ 의 값을 구해보면 $\nu(D_1) = 0.984$, $\nu(D_2) = 0.510$ 과 $\nu(D_3) = 0.281$ 을 각각 얻게되므로 D_3 디자인의 불균형 정도가 가장 심하게 나타나는 것을 알 수 있다.

모형 (3.1)에서 ϕ 와 σ_ϵ^2 이 갖을 수 있는 전체 모수공간은 U 라 정의하고 다음과 같은 영

역으로 선택하였다.

$$U = \{(\phi, \sigma_e^2) \mid 0.2 \leq \phi \leq 10, 0.2 \leq \sigma_e^2 \leq 2\}. \quad (3.3)$$

더불어 모수공간 U 에서 모의실험을 위하여 선택된 (ϕ, σ_e^2) 의 집합을 T 라 정의하면, 이는 다음과 같이 선택되었다.

$$T = \{(\phi, \rho, \sigma_e^2) \mid \phi = 0.2(0.2)10, \sigma_e^2 = 0.2(0.2)2.0\}. \quad (3.4)$$

식 (3.4)에서 ϕ 의 경우 0.2에서 10까지 0.2단위씩 증가하며 50개의 점이 선택되고 σ_e^2 는 0.2에서부터 2까지 0.2단위씩 증가하며 10개의 점이 선택되므로 전체 모의실험 집합 T 는 모두 500 개의 점으로 이루어졌다. 주어진 공간 T 의 각 (ϕ, σ_e^2) 값과 주어진 각 디자인 D_i , ($i = 1, 2, 3$)에서 랜덤벡터 \mathbf{y} 가 발생되었으며, 발생된 \mathbf{y} 를 이용하여 σ_α^2 에 대한 척도화된 추정량 $\widehat{W}_\alpha = \widehat{\sigma}_\alpha^2 / \sigma_\alpha^2$ 이 MV1, MV2와 MV3에 대하여 계산하였다. 이때 척도화된 \widehat{W}_α 를 $\widehat{\sigma}_\alpha^2$ 대신 사용한 이유는 $\widehat{\sigma}_\alpha^2$ 의 경우 σ_α^2 의 실제값에 의존하게 되므로 이의 효과를 제거하기 위함이다. 또한 MIVQUE 추정량은 σ_α^2 에서 음의 추정치를 얻는 경우가 존재한다. 보통 음의 추정치가 나타난 경우에는 0으로 대체하는 것이 일반적이지만 본 연구에서는 이를 0으로 대체하지 않고 추정량의 분위수값을 계산할 때 얻어진 값 그대로 이용하였다.

총 2000번의 반복에 의해서 \widehat{W}_α 값이 실험공간 T 에서 만들어졌으며 이들을 통해 경험적인 분위수가 얻어졌다. 먼저 주어진 디자인 D 와 추정방법 m 에서 \widehat{W}_α 의 p^{th} 경험적 분위수를 $\widehat{Q}_\alpha^m(p, D, \phi, \sigma_e^2)$ 라 정의하면 실험공간 T 에서 p^{th} 경험적 분위수의 최대값과 최소값은 다음과 같이 정의된다.

$$\begin{aligned} \widehat{Q}_\alpha^m(p, D) &= \max_{(\phi, \sigma_e^2) \in T} \widehat{q}_\alpha(p, D, \phi, \sigma_e^2), \\ \widehat{R}_\alpha^m(p, D) &= \min_{(\phi, \sigma_e^2) \in T} \widehat{q}_\alpha(p, D, \phi, \sigma_e^2), \end{aligned}$$

여기서 $D = D_1, D_2, D_3$ 이고 $m = MV1, MV2, MV3$ 이다. 그러므로 척도화된 추정량 \widehat{W}_α 에 대한 EQDGs는 $\widehat{Q}_\alpha^m(p, D)$ 과 $\widehat{R}_\alpha^m(p, D)$ 를 p 에 대하여 플롯함으로써 얻어진다.

이와 같이 얻어진 EQDGs에 대한 해석은 다음과 같은 두 가지 관점에서 이루어진다. 만일 어떤 추정량의 $[\widehat{Q}_\alpha^m(p, D) - \widehat{R}_\alpha^m(p, D)]$ 가 $p(0 \leq p \leq 1)$ 의 영역내에서 다른 추정량에 비하여 작은 값을 갖는다면 그 추정량은 덜 변동된다는 관점에서 효율적인 추정량이다. 또한 어떤 추정량의 EQDGs가 다른 추정량에 비하여 좀 더 균일한 정도 (EQDGs가 평평한 모습)를 보인다면 그 추정량은 다른 추정량에 비하여 좀 더 안정적인 추정량임을 나타낸다.

다음 표 3.1-표 3.3은 각 디자인 및 추정방법에 의하여 얻어진 \widehat{W}_α 의 최대값과 최소값인 $\widehat{Q}_\alpha^m(p, D)$ 과 $\widehat{R}_\alpha^m(p, D)$ 값을 나타내며, 그림 3.1은 표 3.1-표 3.3의 $\widehat{Q}_\alpha^m(p, D)$ 과 $\widehat{R}_\alpha^m(p, D)$ 을 그래프로 그린 EQDGs를 나타낸다.

표 3.1: D_1 디자인에서 $\widehat{W}_\alpha = \widehat{\sigma}_\alpha^2/\sigma_\alpha^2$ 에 대한 MV1, MV2 및 MV3의 비교

p	MV1		MV2		MV3	
	minimum	maximum	minimum	maximum	minimum	maximum
0.01	-1.20310	0.06727	-1.20807	0.07047	-1.20310	0.07125
0.05	-0.79338	0.17946	-0.79960	0.18552	-0.79338	0.18350
0.1	-0.54478	0.27087	-0.55863	0.27901	-0.54478	0.27898
0.2	-0.24788	0.42586	-0.24263	0.43497	-0.24788	0.43602
0.3	0.04464	0.57768	0.05542	0.58462	0.05689	0.58358
0.4	0.33186	0.72368	0.33902	0.73593	0.33412	0.73159
0.5	0.61538	0.88922	0.63174	0.90892	0.62854	0.91323
0.6	0.93101	1.07793	0.93836	1.08644	0.94724	1.08963
0.7	1.15982	1.50323	1.15696	1.51236	1.15858	1.50990
0.8	1.39700	2.06374	1.42352	2.08575	1.42778	2.08883
0.9	1.86063	3.03035	1.82878	3.05424	1.83301	3.03182
0.95	2.28316	4.09659	2.24223	3.99207	2.24580	4.00187
0.99	3.15239	6.09835	3.02346	6.14639	3.03122	6.17913

표 3.2: D_2 디자인에서 $\widehat{W}_\alpha = \widehat{\sigma}_\alpha^2/\sigma_\alpha^2$ 에 대한 MV1, MV2 및 MV3의 비교

p	MV1		MV2		MV3	
	minimum	maximum	minimum	maximum	minimum	maximum
0.01	-0.78851	-0.00009	-3.02586	0.04408	-0.78851	0.02911
0.05	-0.57629	0.01916	-2.00294	0.15062	-0.57220	0.13497
0.1	-0.46127	0.04301	-1.58798	0.22648	-0.45084	0.21927
0.2	-0.32194	0.10863	-0.99883	0.37861	-0.27441	0.37869
0.3	-0.18614	0.20566	-0.52873	0.52440	-0.08157	0.53185
0.4	-0.00386	0.35251	-0.05208	0.67761	0.22983	0.68743
0.5	0.23597	0.53490	0.38172	0.84291	0.59348	0.84828
0.6	0.55658	0.81825	0.90058	1.19404	0.82238	1.12081
0.7	0.94716	1.20900	1.11604	1.82176	1.10621	1.58216
0.8	1.41895	1.98541	1.41644	2.80018	1.42729	2.28765
0.9	2.35088	3.52780	1.93415	4.18743	1.87729	3.61769
0.95	3.37899	5.38774	2.43963	5.69998	2.31252	5.30509
0.99	5.34302	9.24375	3.57417	9.06119	3.13294	9.54920

표 3.3: D_3 디자인에서 $\widehat{W}_\alpha = \widehat{\sigma}_\alpha^2/\sigma_\alpha^2$ 에 대한 MV1, MV2 및 MV3의 비교

p	MV1		MV2		MV3	
	minimum	maximum	minimum	maximum	minimum	maximum
0.01	-2.35125	0.0009	-4.39529	0.0283	-2.35125	0.0207
0.05	-1.72572	0.0468	-3.16851	0.1303	-1.72572	0.1234
0.1	-1.38490	0.0865	-2.41849	0.2198	-1.38233	0.2159
0.2	-0.96663	0.1739	-1.60842	0.3692	-0.95353	0.3759
0.3	-0.64064	0.2698	-0.95544	0.5256	-0.58966	0.5358
0.4	-0.31055	0.3973	-0.32593	0.6748	-0.10513	0.7040
0.5	0.06883	0.5720	0.33827	0.8485	0.51480	0.8829
0.6	0.56315	0.8289	0.88960	1.2603	0.88108	1.3791
0.7	0.95845	1.4251	1.13178	2.1676	1.12724	2.1449
0.8	1.43908	2.5214	1.44796	3.3522	1.44779	3.1656
0.9	2.33767	4.7219	1.92593	5.2337	1.91573	5.1478
0.95	3.22475	7.0552	2.44085	7.1631	2.30683	7.3426
0.99	5.18141	12.5582	3.41757	11.3302	3.12169	12.5477

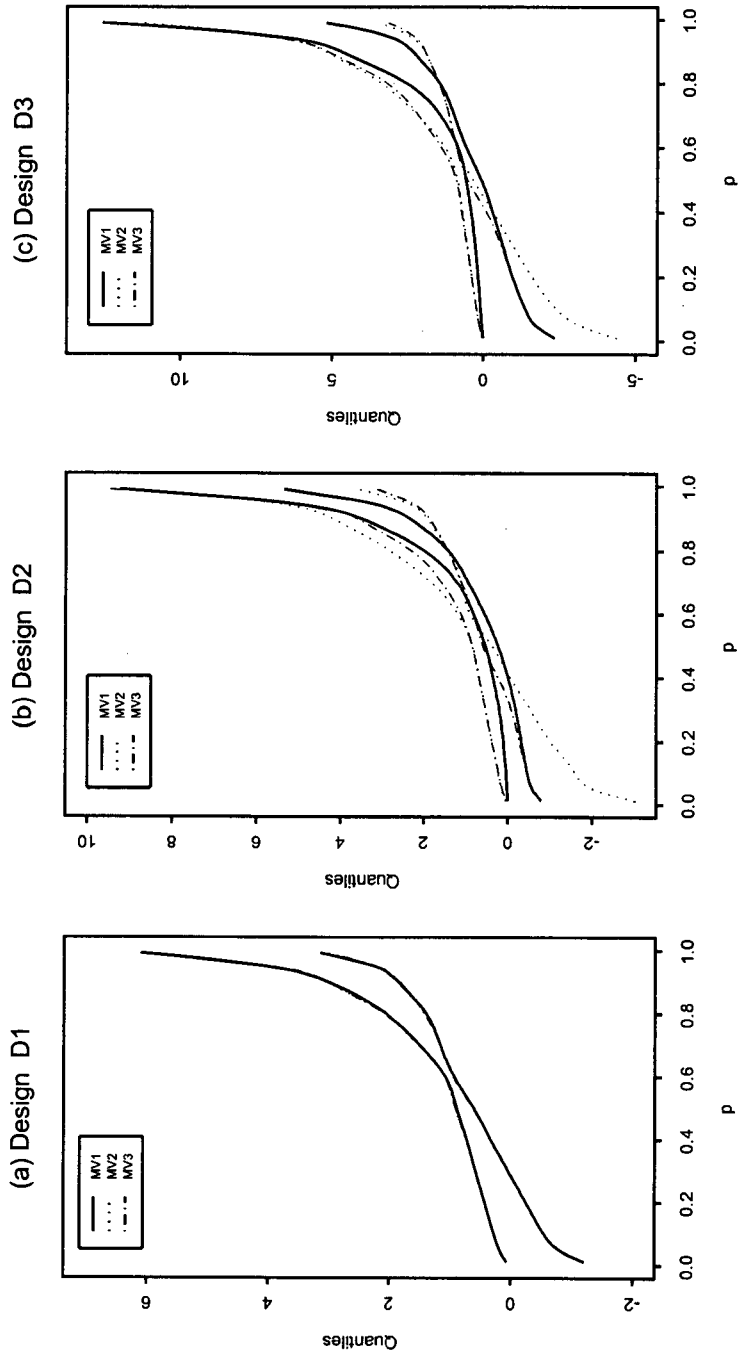


그림 3.1: 각 디자인에서 $\widehat{W}_\alpha = \widehat{\sigma}_\alpha^2 / \sigma_\alpha^2$ 에 대한 MV1, MV2 및 MV3의 비교

표 3.1에서 표 3.3과 그림 3.1을 살펴보면 다음과 같은 결과들을 알 수 있다. 먼저 거의 균형된 디자인 D_1 에서는 3가지 초기값에 따른 MV1, MV2와 MV3 추정량의 \widehat{W}_α 에 대한 분위수의 최대값과 최소값이 거의 비슷하게 나타나 추정량간 효율성의 차이가 없었다. D_2 디자인의 경우 MV1의 \widehat{W}_α 에 대한 분위수의 최대값과 최소값의 차이가 MV2나 MV3에 비하여 작게 나타나 MV1이 σ_α^2 를 추정하는데 있어 좀 더 안정적인 추정량으로 나타났다. 더불어 MV2 추정량은 $p = 0.99$ 인 경우를 제외한 나머지 모든 p 의 영역에서 MV3에 비하여 약간 큰 변동을 보여주어 추정량의 안정성 관점에서 가장 좋지 않은 추정량으로 나타났다. D_3 디자인의 경우에도 MV1이 σ_α^2 를 추정하는데 있어 좀 더 안정적인 추정량으로 나타났으며 MV2추정량은 다른 추정량에 비하여 약간 큰 변동을 보이는 것으로 나타났다.

이와 같은 EQDGs의 결과들을 통하면 어떤 형태의 불균형 자료라 할지라도 초기추정치를 $\sigma_{\alpha_0}^2 = 0$ 과 $\sigma_{\epsilon_0}^2 = 1$ 을 사용하는 MV1이 다른 초기치를 사용하는 MIVQUE 추정량에 비하여 추정량의 안정성 관점에서 약간 나은 추정량인 것으로 나타났다.

4. Conclusions

본 논문에서는 불균형 일원랜덤모형에서 분산성분에 대한 MIVQUE 추정량의 효율성을 척도화한 추정량의 EQDGs를 이용하여 비교하였다. 모의실험 결과, 거의 균형된 디자인에서는 3가지 초기추정량을 사용한 MIVQUE 추정량은 효율성에 차이를 보이지 않았다. 반면 자료가 약간 불균형된 디자인이나 매우 불균형된 디자인인 경우, 비교적 간단하게 계산되는 MV1 추정량 ($\sigma_{\alpha_0}^2 = 0$ 과 $\sigma_{\epsilon_0}^2 = 1$)이 다른 초기치를 이용하는 MIVQUE 추정량에 비하여 추정량의 안정성 관점에서 약간 나은 추정량인 것으로 나타났다.

참고문헌

- 송석현, 정병철 (2002). 불균형 선형혼합모형에서 추정량, <응용통계연구>, 15, 337-354.
- Ahrens, H. and Pincus, R. (1981). On two measures of unbalancedness in a one-way model and their relation to efficiency, *Biometrical Journal*, 23, 227-235.
- Corbeil, R.R. and Searle, S.R. (1976). A comparison of variance component estimators, *Biometrics*, 32, 779-791.
- Khuri, A.I. (1997). Quantile dispersion graphs for analysis of variance estimates of variance components, *Journal of Applied Statistics*, 24, 711-722.
- Lee, J. and Khuri, A.I. (1999). Graphical technique for comparing designs for random models, *Journal of Applied Statistics*, 26, 933-947.
- Lee, J. and Khuri, A.I. (2000). Quantile dispersion graphs for the comparison of designs for a random two-way model, *Journal of Statistical Planning and Inference*, 91, 123-137.
- Searle, S.R. (1971). *Linear Models*, John Wiley & Sons, New York.
- Searle, S.R., Casella G. and McCulloch, C.E. (1992). *Variance Components*, Wiley, New York.

- Swallow, W.H. (1981). Variance of locally minimum variance quadratic unbiased estimators ("MIVQUE's") of variance components, *Technometrics*, **23**, 271-283.
- Swallow, W.H. and Monahan, J.F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components, *Technometrics*, **26**, 47-57.
- Swallow, W.H. and Searle, S.R. (1978). Minimum variance quadratic unbiased estimation (MIVQUE) of variance components, *Technometrics*, **20**, 265-272.

[2005년 2월 접수, 2005년 4월 채택]

Comparison of MIVQUE Estimators Using EQDGs for the One-way Random Model with Unbalanced Data

Byoung Cheol Jung ¹⁾

ABSTRACT

In this study, the MIVQUE estimators of variance components for the one-way random model with unbalanced data are investigated. In order to compare the efficiency of MIVQUE estimators obtained by using three priori estimates, the Empirical Quantile Dispersion Graphs (EQDGs) are used. From the results of Monte-Carlo study, the MIVQUE estimator using $\sigma_\alpha^2 = 0$ and $\sigma_e^2 = 1$ as the priori estimate performs well relative to other estimators.

Keywords: MIVQUE, Variance components model, Priori value

1) Contract Professor, Dept. of Statistics, Sungshin University, 249-1 dongsun-dong, sungbuk-gu, Seoul 136-742, Korea
E-mail : bcjung@empal.com