Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp. 329 340

Confidence Interval for the Variance Component in a Unbalanced One-way Random Effects Model

Gyu Moon Song[1)]

Abstract

Two methods are proposed for constructing a confidence interval on the among group variance component in a unbalanced one-way random effects model. Computer simulation is used to compare these methods with alternative procedures. The results indicate that the method1 and methods2 perform well over small group size and large sample size respectively.

Keywords : 분산성분, 분광분해, 포함범위비율

1. 서론

변량모형의 응용에서 분산성분의 신뢰구간은 자주 요구된다. 본 연구에서는 불균형 일원변량모형에서 그룹 간 분산성분 σ_a^2에 대한 신뢰구간을 구성하는 문제를 생각한다. 이를 위해 먼저 불균형 모형에서 σ_a^2의 분산분석(Analysis of Variance, AOV)추정량의 분포를 유도하고, 이를 이용하여 새로운 신뢰구간을 제안한다.

오랫동안 변량모형에서 분산성분의 신뢰구간을 얻기 위한 방법이 여러 학자에 의해 제안되었다. 특히 불균형인 경우에 Thomas와 Hultquist(1978)은 Williams(1962)의 균형인 경우 σ_a^2에 대한 신뢰구간에서 S_1^2 대신에 hS_3^2을 대치하여 얻었고, 또한 Burdick과 Eickman (1986)은 Thomas와 Hultquist(1978)의 근사적인 방법과, Burdick, Maqsood 그리고 Graybill (1986)의 방법에 의해 새로운 신뢰구간을 제안하였다. 여기

---

1. Professor, Department of Statistics, Keimyung University.
E-mail: kms252@kmu.ac.kr

서 $S_1^2 = \frac{1}{t-1}\sum_i(\overline{y}_{i.} - \overline{y}_{..})^2$, $S_3^2 = \frac{1}{t-1}[\sum_i \overline{y}_{i.}^2 - \frac{1}{t}(\sum_i \overline{y}_{i.}^2)^2]$이고, $h = t/\sum_i(1/n_i)$ 이다. 본 연구에서는 $\chi^2(1)$의 선형결합인 $\hat{\sigma}_a^2$의 분포를 근사적인 χ^2분포로 유도하여 σ_a^2의 새로운 근사적인 신뢰구간을 제안한다. 여기서 $\chi^2(1)$은 자유도 1인 χ^2분포를 나타낸다.

2. 모형과 분산성분의 추정량의 분포

불균형 일원변량모형에서 i번째 그룹의 j번째 관측값 y_{ij}는 다음과 같이 나타낼 수 있다.

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \cdots, t,\ j = 1, \cdots, n_i,\ N = \sum n_i$$

여기서 $\alpha_i \sim N(0, \sigma_a^2)$, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$이고, α_i, ε_{ij}는 독립이라 가정한다.

이 모형을 행렬기호로 나타내면 다음과 같다.

$$\boldsymbol{y} = \boldsymbol{1}_N \mu + \boldsymbol{U}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \tag{2.1}$$

여기서 $\boldsymbol{\alpha} \sim N(\boldsymbol{0}, \sigma_a^2 \boldsymbol{I}_t)$, $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_e^2 \boldsymbol{I}_N)$ 이다. 그러면 $\boldsymbol{y}$는 평균 $\boldsymbol{1}_N\mu$, 분산-공분산 행렬 $\boldsymbol{V} = \sigma_a^2 \boldsymbol{U}\boldsymbol{U}' + \sigma_e^2 \boldsymbol{I}_N$인 정규분포를 따르는 확률변수의 벡터이고, $\boldsymbol{1}_N$은 모든 원소가 1인 $N \times 1$인 벡터를 나타내며, $\boldsymbol{U}$는 $diag\{\boldsymbol{1}_{n_1}, \cdots, \boldsymbol{1}_{n_t}\}$으로 $N \times t$ 행렬로 $\boldsymbol{1}_{n_i}$은 모든 원소가 1인 $n_i \times 1$인 벡터이다.

모형(2.1)에서 분산성분의 AOV추정량은 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \hat{\sigma}_a^2 &= \frac{1}{k}[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} - \frac{t-1}{n-t}\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}] \\ &= \frac{t-1}{k}[S_1^2 - S_2^2] \end{aligned} \tag{2.2}$$

$$\begin{aligned} \hat{\sigma}_e^2 &= \frac{1}{N-t}\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y} \\ &= S_2^2 \end{aligned} \tag{2.3}$$

여기서 $\boldsymbol{A} = \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}' + (1/N)\boldsymbol{J}_N$, $\boldsymbol{B} = \boldsymbol{I}_N - \boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'$으로 $\boldsymbol{U}$는 식 (2.1)과 같고, $\boldsymbol{J}_N$은 모든 원소가 1인 $N \times N$ 행렬이다. 또한

$S_1^2 = \frac{1}{t-1}\sum_i(\bar{y}_{i.} - \bar{y}_{..})^2$, $S_2^2 = \frac{1}{N-t}\sum_i\sum_j(\bar{y}_{ij} - \bar{y}_{i.})^2$이다. 이때 $\hat{\sigma}_a^2$, $\hat{\sigma}_e^2$의 평균과 분산은 다음과 같다.

$$E(\hat{\sigma}_a^2) = \sigma_a^2$$

$$E(\hat{\sigma}_e^2) = \sigma_e^2$$

$$V(\hat{\sigma}_a^2) = \frac{2\sigma_e^4 N^2(N-1)(t-1)}{(N-t)(N^2 - \sum_i n_i^2)^2} + \frac{4\sigma_a^2\sigma_e^2 N}{N^2 - \sum_i n_i^2} + \frac{2\sigma_a^4[N^2(\sum_i n_i^2)^2 - 2N\sum_i n_i^3]}{(N^2 - \sum_i n_i^2)^2} \quad (2.4)$$

$$V(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{(N-t)}$$

모형 (2.1)에서 $\boldsymbol{y} \sim N(\boldsymbol{1}_N\mu, \boldsymbol{V})$이므로 $\boldsymbol{y} - \boldsymbol{1}_N\mu \sim N(\boldsymbol{0}, \boldsymbol{V})$이 되고,

$$\boldsymbol{z} = (1/\sigma_e)\,T(\boldsymbol{y} - \boldsymbol{1}_N\mu) \sim N(\boldsymbol{0}, \boldsymbol{I}_N) \quad (2.5)$$

이 된다. 여기서 V는 분광분해(Spectral Decomposition)에 의해 다음과 같이 나타낼 수 있다.

$$\boldsymbol{V} = \sigma_e^2 \boldsymbol{P}\boldsymbol{\lambda}\boldsymbol{P}' = \sigma_e^2 \boldsymbol{W} \quad (2.6)$$

그러면 $\boldsymbol{T}\boldsymbol{W}\boldsymbol{T}' = \boldsymbol{I}$가 되는 $\boldsymbol{T}$는 $\boldsymbol{T} = \boldsymbol{P}\boldsymbol{\lambda}^{-1/2}\boldsymbol{P}'$이 된다.

따라서 $\boldsymbol{y} = \sigma_e \boldsymbol{T}^{-1}\boldsymbol{z} + \boldsymbol{1}_N\mu$ 이고, $\boldsymbol{1}_N\boldsymbol{A} = \boldsymbol{0}$, $\boldsymbol{1}_N\boldsymbol{B} = \boldsymbol{0}$이 되어 $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$와 $\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}$는 다음과 같이 나타낼 수 있다.

$$\begin{aligned}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} &= \boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}\\ \boldsymbol{y}'\boldsymbol{B}\boldsymbol{y} &= \boldsymbol{z}'\boldsymbol{D}\boldsymbol{z}\end{aligned} \quad (2.7)$$

여기서 $\boldsymbol{C} = \sigma_e^2\boldsymbol{T}^{-1}\boldsymbol{A}\,\boldsymbol{T}^{-1}$, $\boldsymbol{D} = \sigma_e^2\boldsymbol{T}^{-1}\boldsymbol{B}\,\boldsymbol{T}^{-1}$이다.

Johnson과 Kotz(1970)에 의하면 다음이 성립한다.

$$\begin{aligned}\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z} &\sim \sum_{i=1}^{t-1}\lambda_i\,\chi^2(1)\\ \boldsymbol{z}'\boldsymbol{D}\boldsymbol{z} &\sim \sum_{i=1}^{N-t}\lambda_i^*\,\chi^2(1)\end{aligned} \quad (2.8)$$

여기서 λ_i는 $\boldsymbol{C}$의 고유값 $(\lambda_1 > \cdots > \lambda_{t-1})$, λ_i^*는 $\boldsymbol{D}$의 고유값 $(\lambda_1^* > \cdots > \lambda_{N-t}^*)$ 이고,

$AVB = \mathbf{0}$이 되어 $y'Ay$와 $y'By$는 독립이다. 따라서 $\hat{\sigma}_a^2$의 분포는 다음과 같다.

$$\hat{\sigma}_a^2 \sim \frac{1}{k}\left[\sum_{i=1}^{t-1}\lambda_i\ \chi^2(1) - \frac{t-1}{N-t}\sum_{i=1}^{N-t}\lambda_i^*\ \chi^2(1)\right] \sim \sum_{i=1}^{N-1} c_i\ \chi^2(1) \tag{2.9}$$

여기서 $c_1 = \frac{1}{k}\lambda_1, \cdots, c_{t-1} = \frac{1}{k}\lambda_{t-1}, c_t = -\frac{1}{k}\frac{t-1}{N-t}\lambda_1^*, \cdots, c_{N-1} = -\frac{1}{k}\frac{t-1}{N-t}\lambda_{N-t}^*$ 이다.

같은 방법으로

$$\hat{\sigma}_e^2 \sim \frac{1}{N-t}\sum_{i=1}^{N-t}\lambda_i^*\ \chi^2(1) \tag{2.10}$$

여기서 λ_i^*는 식(2.8)과 같다.

3.제안된 방법

Williams(1962)은 균형모형에서 σ_a^2에 대한 $100(1-\alpha)\%$ 신뢰구간을 다음과 같이 나타냈다.

$$\left[\frac{S_1^2 - S_2^2 F_2}{nF_1}\ ,\ \frac{S_1^2 - S_2^2 F_4}{nF_3}\right] \tag{3.1}$$

여기서 $n_1 = \cdots = n_t = n$이고, S_1^2과 S_2^2은 식(2.2)와 같다. 또 $F_1 = F_{\alpha_1, t-1, \infty}$, $F_2 = F_{\alpha_1, t-1, N-t}$, $F_3 = F_{1-\alpha_2, t-1, \infty}$, $F_4 = F_{1-\alpha_2, t-1, N-t}$, $\alpha_1 + \alpha_2 = \alpha$으로, F_{δ, ν_1, ν_2}는 자유도 ν_1, ν_2인 F분포의 상위 δ분위수 이다.

식(3.1)에서 n 대신 h, S_1^2 대신에 hS_3^2을 대치하면 다음과 같은 Thomas-Hultqist(1978)방법인 σ_a^2에 대한 $100(1-\alpha)\%$ 신뢰구간이 얻어진다.

$$\left[\frac{hS_3^2 - S_2^2 F_2}{hF_1}\ ,\ \frac{hS_3^2 - S_2^2 F_4}{hF_3}\right] \tag{3.2}$$

여기서 $S_3^2 = \frac{1}{t-1}\left[\sum_i \bar{y}_{i.}^2 - \frac{1}{t}(\sum_i \bar{y}_{i.})^2\right]$ 이고, $h = t/\sum_i (1/n_i)$이다. 이 방법을 Burdick과 Eickman(1986)에서 나타난 것처럼 USS방법으로 나타낸다.

또한, 식(3.1)에서 n 대신 k_0로 대치하면 다음과 같은 σ_a^2에 대한 $100(1-\alpha)\%$ 신뢰구간이 얻어진다.

$$\left[\frac{S_1^2 - S_2^2 F_2}{k_0 F_1}\ ,\ \frac{S_1^2 - S_2^2 F_4}{k_0 F_3}\right] \tag{3.3}$$

여기서 $k_0 = \dfrac{N - \sum(n_i^2/N)}{t-1}$이다. 이 방법을 (3.2)에서와 같이 ANOVA방법으로 나타낸다.

식 (3.2)와 (3.3)은 (3.1)의 변형으로 얻어진 방법들이다. 그러나 Burdick과 Eickman(1986)은 Thomas-Hultquist방법을 근사시키고, Burdick, Maqsood 그리고 Graybill(1986)에 의해 다음과 같은 σ_a^2에 대한 $100(1-\alpha)\%$ 신뢰구간을 얻었다.

$$\left[\frac{hS_3^2 L_1}{F_1(1+hL_1)}\ ,\ \frac{hS_3^2 U_1}{F_3(1+hU_1)}\right] \tag{3.4}$$

여기서 $L_1 = \dfrac{S_3^2}{F_2 S_2^2} - \dfrac{1}{m}$, $U_1 = \dfrac{S_3^2}{F_4 S_2^2} - \dfrac{1}{M}$, $m = Min(n_1, \cdots, n_t)$ $M = Max(n_1, \cdots, n_t)$ 이다. 이 방법을 (3.2)에서와 같이 INT방법으로 나타낸다.

식 (2.4)의 $V(\hat{\sigma}_a^2)$에서 σ_a^2 대신 $\hat{\sigma}_a^2$, σ_e^2 대신 $\hat{\sigma}_e^2$으로 대치하여 $\widehat{V}(\hat{\sigma}_a^2)$을 얻을 수 있다. Ames와 Webster(1991)는 분산성분의 선형결합의 분포에서 근사적인 자유도를 구하였으나 본 연구에서는 식(2.9)에서 $\hat{\sigma}_a^2$의 분포와 Graybill(1961)의 (정리17.1)을 이용하여 $\hat{\sigma}_a^2$의 근사적인 분포 $\chi^2(d)$를 얻는다. 이때 d는 다음과 같이 얻어진다.

$$d^* = \frac{\widehat{V}(\hat{\sigma}_a^2)}{\sum c_i^2} = \frac{\widehat{V}(\hat{\sigma}_a^2)}{(1/k^2)\sum\lambda_i^2 + [(t-1)^2/k^2(N-t)^2]\sum\lambda_i^{*2}}$$

$$d \simeq round(d^* + t/3) \tag{3.5}$$

여기서 $t/3$는 자유도를 얻는데 사용된 수정항으로 t는 그룹의 수이고, $round$는 괄호 안의 값을 소숫점에서 반올림하여 정수가 얻어지는 함수이다.

따라서 Graybill의 (정리17.2)에 의하여 σ_a^2에 대한 근사적인 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이 얻을 수 있다.

$$\frac{d\ \hat{\sigma}_a^2}{\chi^2_{\alpha/2}(d)} \le \sigma_a^2 \le \frac{d\ \hat{\sigma}_a^2}{\chi^2_{1-\alpha/2}(d)} \tag{3.6}$$

이 방법을 제안된 방법1로 나타낸다.

또 다른 방법으로 식(2.2)에서 $\hat{\sigma}_a^2 = \frac{t-1}{k}[S_1^2 - S_2^2]$이므로

$$S_1^2 \sim \frac{1}{t-1}\sum \lambda_i \chi^2(1)$$
$$S_2^2 \sim \frac{1}{N-t}\sum \lambda_i^* \chi^2(1) \quad (3.7)$$

으로 나타낼 수 있다. $V(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}) = 2tr(\boldsymbol{A}\ \boldsymbol{V})^2 + 4\boldsymbol{\mu}'\boldsymbol{A}\ \boldsymbol{V}\boldsymbol{A}\ \boldsymbol{\mu}$ 이므로

$$V(S_1^2) = \frac{1}{(t-1)^2} V(\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}) = \frac{2}{(t-1)^2}[tr(\boldsymbol{A}\ \boldsymbol{V})^2 + 2\boldsymbol{\mu}'\boldsymbol{A}\ \boldsymbol{V}\boldsymbol{A}\ \boldsymbol{\mu}] \quad (3.8)$$

또 식(3.7)로부터 제안된 방법1에서와 같이 $V(S_1^2) = V(\frac{1}{t-1}\sum \lambda_i \chi^2(d_1))$이라 가정하자. 그러면

$$V(S_1^2) = \frac{1}{(t-1)^2}\sum \lambda_i^2\, 2d_1 \quad (3.9)$$

이 되어, 식(3.8)과 (3.9)로부터

$$d_1' = \frac{tr(\boldsymbol{A}\ \boldsymbol{V})^2 + 2\boldsymbol{\mu}'\boldsymbol{A}\ \boldsymbol{V}\boldsymbol{A}\ \boldsymbol{\mu}}{\sum \lambda_i^2}$$

$$d_1 \simeq round(d_1') \quad (3.10)$$

같은 방법으로,

$$d_2' = \frac{tr(\boldsymbol{B}\ \boldsymbol{V})^2 + 2\boldsymbol{\mu}'\boldsymbol{B}\ \boldsymbol{V}\boldsymbol{B}\ \boldsymbol{\mu}}{\sum \lambda_i^{*2}}$$

$$d_2 \simeq round(d_2') \quad (3.11)$$

Graybill의 (정리17.2)에 의하여 식(3.10)과 (3.11)로 부터

$$d'' = \frac{\left[\frac{t-1}{k}S_1^2 - \frac{t-1}{k}S_2^2\right]^2}{\frac{\left(\frac{t-1}{k}\right)^2 S_1^4}{d_1} + \frac{\left(\frac{t-1}{k}\right)^2 S_2^4}{d_2}} = \frac{(S_1^2 - S_2^2)^2}{(S_1^4/d_1) + (S_2^4/d_2)}$$

$$d' = round(d'' + t/3) \tag{3.12}$$

여기서 $t/3$는 식(3.5)에서와 같다.

식(3.10)과 (3.11)로부터 σ_a^2에 대한 근사적인 $100(1-\alpha)\%$ 신뢰구간은 다음과 같이 얻을 수 있다.

$$\left[\frac{d'\hat{\sigma}_a^2}{\chi^2_{\alpha/2}(d')}, \frac{d'\hat{\sigma}_a^2}{\chi^2_{1-\alpha/2}(d')}\right] \tag{3.13}$$

이 방법을 제안된 방법2로 나타낸다.

4. 모의 실험

실험에 사용된 불균형 형태는 Swallow와 Searle(1984)에 나타난 13가지 형태를 이용하여 그룹의 수가 적고, 보통이고, 많은 경우를 각각 t=3, 6, 9로 나타내고, 또한 표본의 수가 적고, 보통이고, 많은 경우를 N=15, 30, 45로 나타내었다. 이것을 나타내면 <표 4.1>과 같다. 또한 각 방법들이 균형인 경우에 어떻게 적용되는 지를 알아보기 위해 <표 4.1>을 변형한 9가지 형태를 나타내었다. 이것은 <표 4.2>와 같다

<표 4.1> 실험에 사용한 불균형 형태

형태	t	N	n_i
P1	3	15	3, 5, 7
P2	3	15	1, 5, 9
P3	3	15	1, 7, 7
P4	6	30	3, 3, 5, 5, 7, 7
P5	6	30	1, 1, 5, 5, 9, 9
P6	6	30	1, 1, 7, 7, 7, 7
P7	6	30	1, 1, 1, 1, 13, 13
P8	9	45	3, 3, 3, 5, 5, 5, 7, 7, 7
P9	9	45	1, 1, 1, 5, 5, 5, 9, 9, 9
P10	9	45	1, 1, 1, 7, 7, 7, 7, 7, 7
P11	9	45	1, 1, 1, 1, 1, 1, 1, 19, 19
P12	3	30	2, 10, 18
P13	3	45	3, 15, 27

<표 4.2> 실험에 사용한 균형 형태

형태	t	N	n
Q1	3	9	3, 3, 3
Q2	3	18	6, 6, 6
Q3	3	27	9, 9, 9
Q4	6	18	3, 3, 3, 3, 3, 3
Q5	6	36	6, 6, 6, 6, 6, 6
Q6	6	54	9, 9, 9, 9, 9, 9
Q7	9	27	3, 3, 3, 3, 3, 3, 3, 3, 3
Q8	9	54	6, 6, 6, 6, 6, 6, 6, 6, 6
Q9	9	81	9, 9, 9, 9, 9, 9, 9, 9, 9

사용된 자료는 SAS(Statistical Analysis System)의 난수를 사용하여 얻었고, SAS IML을 이용하여 모의실험이 수행되었다. 본 연구에서는 $\rho = \sigma_a^2/\sigma_a^2 = 1$인 경우에 국한하여 실험이 이루어졌고, 10,000번의 실험을 통하여 앞서 논의한 5가지 방법에 대한 동일 꼬리($\alpha_1 = \alpha_2 = \alpha/2$)를 갖는 σ_a^2에 대한 양측 95% 신뢰구간을 구하여 각 방법의 포함범위비율(coverage probability)과 신뢰구간의 평균길이를 계산하였다.

5. 실험결과

실험의 결과가 <표 5.1>에서 <표 5.4>까지 주어진다. <표 4.1>에 대해 모의실험을

통하여 σ_a^2의 95% 신뢰구간에 대한 포함범위비율과 신뢰구간의 평균길이가 각각 <표 5.1>과 <표 5.2>에 주어지고, <표 4.2>에 대해서도 모의실험에 의한 σ_a^2의 95% 신뢰구간에 대한 포함범위비율과 신뢰구간의 평균길이가 각각 <표 5.3>과 <표 5.4>에 주어진다.

<표 5.1> 불균형인 경우 95% 신뢰구간에 대한 포함범위비율

형태	제안된 방법1	제안된 방법2	ANOVA	USS	INT
P1	0.9337	0.9215	0.9719	0.9749	0.8804
P2	0.9113	0.9301	0.9635	0.9735	0.7221
P3	0.9122	0.9336	0.9628	0.9742	0.7324
P4	0.9555	0.9302	0.9725	0.9755	0.9416
P5	0.9359	0.9273	0.9634	0.9693	0.6320
P6	0.9438	0.9261	0.9634	0.9703	0.6443
P7	0.8925	0.9063	0.9391	0.9690	0.7864
P8	0.9648	0.9208	0.9570	0.9562	0.9590
P9	0.9477	0.9077	0.9417	0.9520	0.6013
P10	0.9528	0.9153	0.9468	0.9598	0.5957
P11	0.8728	0.8811	0.9034	0.9615	0.9601
P12	0.9153	0.9415	0.9559	0.9737	0.8608
P13	0.9270	0.9518	0.9598	0.9747	0.9147

<표 5.2> 불균형인 경우 95% 신뢰구간에 대한 평균길이

형태	제안된 방법1	제안된 방법2	ANOVA	USS	INT
P1	47.7533	43.7251	55.7642	56.4458	56.6444
P2	50.9115	73.4293	60.5080	68.8257	68.3420
P3	50.6158	74.6117	59.6435	68.1346	59.7233
P4	14.0368	5.2860	7.3275	7.1387	7.5006
P5	14.1677	6.5082	7.4778	8.4596	9.6960
P6	14.1144	6.4439	7.4150	8.4216	7.4400
P7	15.0625	11.0912	8.1478	10.3157	10.3488
P8	7.9009	3.2128	4.1271	3.9866	4.1439
P9	7.9728	3.6545	4.1865	4.7892	6.3726
P10	7.9756	3.5963	4.1776	4.7210	4.1691
P11	8.2273	6.1050	4.5287	6.0610	6.7080
P12	46.1133	52.1203	51.0064	54.0358	54.2844
P13	43.0793	45.8099	46.3584	48.8040	49.0127

<표 5.3> 균형인 경우 95% 신뢰구간에 대한 포함범위비율

형태	제안된 방법1	제안된 방법2	ANOVA	USS	INT
Q1	0.9213	0.9071	0.9819	0.9819	0.9819
Q2	0.9377	0.9356	0.9771	0.9771	0.9771
Q3	0.9435	0.9414	0.9738	0.9738	0.9738
Q4	0.9473	0.9146	0.9808	0.9807	0.9807
Q5	0.9613	0.9358	0.9706	0.9698	0.9698
Q6	0.9724	0.9383	0.9622	0.9609	0.9609
Q7	0.9533	0.9115	0.9808	0.9783	0.9783
Q8	0.9725	0.9174	0.9561	0.9540	0.9540
Q9	0.9812	0.9159	0.9516	0.9497	0.9497

<표 5.4> 균형인 경우 95% 신뢰구간에 대한 평균길이

형태	제안된 방법1	제안된 방법2	ANOVA	USS	INT
Q1	51.9975	75.1747	64.0439	64.3142	64.3142
Q2	44.7495	35.2919	51.1462	51.2328	51.2328
Q3	43.6334	24.6993	47.9460	47.9968	47.9968
Q4	14.7974	6.5465	8.8409	8.1463	8.1463
Q5	13.9132	4.9668	6.9799	6.7338	6.7338
Q6	13.6804	4.4903	6.4527	6.3058	6.3058
Q7	8.1219	3.5573	4.9888	4.5243	4.5243
Q8	7.9251	2.9189	3.9598	3.7894	3.7894
Q9	7.8354	2.7151	3.6601	3.5576	3.5576

불균형인 경우인 <표 5.1>에 나타난 결과로부터 제안된 방법1은 기존의 방법에 비해 불균형이 약한 경우 즉 P1, P4, P8 형태와 불균형이 보통이고, 그룹의 수 t와 표본의 수 N이 클 때 즉, P5, P6, P8, P9, P10에서 좋은 포함범위비율을 보이며, 제안된 방법2는 표본의 수가 적은 경우 즉, P2, P3, P12, P13에서 좋은 포함범위비율을 보이고 있다. <표 5.2>에서는 제안된 방법1은 t가 작은 경우 즉 P2, P3, P12, P13에서 기존 방법에 비해 평균길이가 짧고, 제안된 방법2는 t가 클 때 즉 P4, P5, P6, P8, P9, P10에서 평균길이가 가장 짧게 나타났다.

한편 균형인 경우인 <표 5.3>에 나타난 결과는 제안된 방법1이 대부분의 형태에서 좋은 포함범위비율을 보이고 있으며, <표 5.4>는 제안된 방법2가 신뢰구간의 평균길이는 가장 짧게 나타났다.

이상을 종합하면 제안된 방법1은 불균형이 심하지 않은 경우, 제안된 방법2는 그룹의 수가 적은 경우 포함범위비율과 평균길이가 기존의 방법에 비하여 잘 유지하고 있음을 알 수 있다.

6. 수치 예

Burdick, Maqsood 그리고 Graybill(1986)에 나타난 예를 가지고 3절에 나타난 방법들을 나타낸다. <표 6.1>의 자료는 생산라인에서 랜덤하게 추출된 식물기름으로 채워진 5그룹의 병의 무게를 나타낸다.

<표 6.1> 병의 무게

	그룹				
	1	2	3	4	5
	15.70	15.69	15.75	15.68	15.65
	15.68	15.71	15.82	15.66	15.60
	15.64		15.75	15.59	
	15.60		15.71		
			15.84		
$\bar{y}_{.j}$	15.655	15.700	15.774	15.643	15.625

3절에 나타난 방법들을 보이기 위해 σ_a^2에 대한 95% 양측 신뢰구간을 구하기 위한 값들을 나타내면 $m = 2$, $M = 5$, $h = 2.804$, $k_0 = 3.094$이고, 요구되는 양쪽 꼬리가 동일한 F값은 $F_1 = 2.786$, $F_2 = 4.275$, $F_3 = 0.121$, $F_4 = 0.114$ 이다. <표 6.2>는 각 방법에 대해 계산된 95% 신뢰구간이다. 먼저, 제안된 방법1을 구하기 위하여 $d^* = 1$이므로 $d = 3$이 되어 $\chi^2_{0.025}(3) = 0.2158$, $\chi^2_{0.975}(3) = 9.3484$ 얻어진다. 따라서 제안된 방법1에 대한 신뢰구간의 하한은 0.001212, 상한은 0.052517 이다. 또한 제안된 방법2를 구하기 위하여 $d_1' = 3.7315$, $d_2' = 11$이므로 $d_1 = 4$, $d_2 = 11$이 되고, 이것을 이용한 $d'' = 2.8331$이므로 $d' = 4$이 되어 $\chi^2_{0.025}(4) = 0.2158$, $\chi^2_{0.975}(4) = 9.3484$이다. 따라서 제안된 방법2에 대한 신뢰구간의 하한은 0.001356, 상한은 0.031193이 된다.

<표 6.2> <표 6.1>의 자료에 대한 95% 신뢰구간

방법	신뢰구간	구간길이
제안된 방법1	[0.001212, 0.052517]	0.05131
제안된 방법2	[0.001356, 0.031193]	0.02984
ANOVA	[0.000543, 0.038609]	0.03807
USS	[0.000106, 0.028657]	0.02855
INT	[-0.000578, 0.028665]	0.02924

σ_a^2의 AOV추정량은 $\hat{\sigma}_a^2$=0.0037776이다. 또한 위 자료는 그룹의 수가 5인 경우로 그룹의 크기가 보통인 경우이다. 제안된 방법1과 2는 AOV추정량에 바탕을 둔 신뢰구간으로 <표 6.2>에서 제안된 방법1은 구간길이가 가장 길게 나타나고, 제안된 방법2는 ANOVA 보다 짧고, USS나 INT와 비슷하며, INT는 하한이 음수로 나타나고 있다.

참고문헌

1. Ames, M.H. and Webster, J.H. (1991). On Estimating Approximate Degrees of Freedom. *The American Statistician*, Vol. 45, 45-50.
2. Burdick, R.K. and Eickman, J. (1986). Confidence Intervals on the Among Group Varinace Component in the Unbalanced One-Fold Nested Design, *Journal of Statistical Computation and Simulation*. Vol. 26, 205-219.
3. Burdick, R.K. and Maqsood, F. (1986). Confidence Intervals on the Intraclass Correlation in the Unbalanced One-Way Classification. *Communications in Statistics-Theory and Method*, 15(11), 3353-3378.
4. Burdick, R.K. and Graybill, F.A. (1988). The Present Status of Confidence Interval Estimation on Variance Components in balanced and Unbalanced Random Models. *Communications in Statistics-Theory and Method,* 17(4), 1165-1195.
5. Graybill, F.A. (1961). *An Introduction to Linear Statistical Models*, Vol I, McGRAW-Hill Book Company, Inc.
6. Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, Inc.
7. Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components,* Wiley, New York.
8. Swallow, W.H. and Monahan, J.F. (1984). Monte Carlo Comparison of ANOVA, MINQUE, REML, and ML Estimators of Variance Components. *Technometrics*, Vol. 26, 47-57.
9. Thomas, J.D. and Hultquist, R.A. (1978). Interval estimation for the unbalanced case of the one-way random effects medel. *The Annals of Statistics*, 6, 582-587.
10. Williams, J.S. (1962). A confidence interval for variance components. *Biometrika,* 49, 278-281.

[2002년 10월 접수, 2002년 10월 채택]