The Korean Communications
in Statistics Vol. 10 No. 1, 2003
pp. 125-134

A Comparison of Influence Diagnostics in Linear Mixed Models[1)]

Jang-Taek Lee[2)]

Abstract

Standard estimation methods for linear mixed models are sensitive to influential observations. However, tools and concepts for linear mixed model diagnostics are rudimentary until now and research is heavily demanded in linear mixed models.

In this paper, we consider two diagnostics to evaluate the effects of individual observations in the estimation of fixed effects for linear mixed models. Those are Cook's distance and COVRATIO. Results of our limited simulation study suggest that the Cook's distance is not good statistical quantity in linear mixed models. Also calibration point for COVRATIO seems to be quite conservative.

Keywords : Cook's distance, COVRATIO, Linear mixed models.

1. 서론

회귀모형이나 고정효과모형에 있어서 여러 가지 진단방법들에 대한 눈부신 연구들은 랜덤효과와 혼합모형에까지 확장되어 진행되어지고 있다. 1980년대 후반부터는 본격적으로 회귀모형에서 혼합모형으로 일반화되어서 진단의 문제가 언급되기 시작했는데, 혼합모형의 진단 중에서 제일 먼저 관심을 가졌던 분야는 분산성분의 점추정 이었으며, 실제로 분산성분추정치의 값이 음수가 되는 경우가 발생하는 중요한 원인이 영향력 관찰점이라고 생각되어지기 때문이었다. Hocking(1983)은 이러한 연구의 선구자로서 분산성분추정치가 음수가 되는 원인을 연구하였으며, 수상한 데이터를 규명하기 위한 진단도구들을 제안하였다. 그 후 Hocking과 Bremer(1987)는 균형혼합모형에서 AVE 추정량과 진단방법을 제안하였으나 불균형혼합모형에 있어서 AVE 추정량의 우수성은 입증되지 못했으며, 셀이 비어있는 경우에도 사용할 수가 없다. Fellner(1986)는 혼합모형에 있어서 분산성분을 추정하는데 이상점의 영향력을 제한최우추정량(REMLE)을 이용하여 구하였다. 또한 Beckman, Nachtsheim과 Cook(1987)은 혼합모형에 사용하는 일반적인 가정에 대한 타당성을 평가하는 우도원리에 입각한 새로운 방법을 제안하였다.

1990년대에는 보다 주목할 만한 논문들이 발표되었는데, Christensen, Pearson과 Johnson(1992)

1) This work was supported by Dankook University Research Fund in 2003
2) Professor, Division of Information and Computer Science, Dankook University, San 8, HanNam-Dong, YongSan-Gu, Seoul, Korea
E-mail : jtlee@dankook.ac.kr

은 REMLE를 이용한 혼합모형의 진단 문제를 다루었다. 그들은 특정 관측치를 제거한 후 나머지 관측치를 이용하여 영향력을 측정하는 소거법을 위한 편리한 계산방법 및 고정효과와 분산성분에 대한 진단도구들을 제안하였다. 그리고 Hurtado(1993)는 최우추정량(MLE)을 이용한 혼합모형의 진단에 사용되는 통계량을 제안하였는데, 분산성분의 비율이 기지인 경우에는 제안된 통계량을 이용하여 정확하게 한 개의 관측치가 영향력 관찰점인가를 알 수 있으며, 미지인 경우에는 몇 가지 가정 아래에서 근사적으로 적용 가능하다.

본 연구에서는 가장 보편화된 회귀모형에서의 영향력 측도들이 혼합모형에서도 적용 가능한 지를 구체적으로 모의실험을 통하여 살펴보고자 한다. 고려된 진단통계량은 회귀모형에서 기준치(calibration point)가 분명하게 제안되어 있는 쿡의 통계량(Cook, 1977)과 Belsley, Kuh와 Welsch (1980)가 제안한 COVRATIO를 사용하였다.

본 논문의 구성은 2절에서는 선형혼합모형과 고려된 영향력 측도에 대하여 간략하게 서술하고, 3절에서는 본 연구에서 사용된 분산성분추정량의 종류, 모의실험에 사용된 모형의 종류 및 조건, 모의실험의 과정과 추정된 쿡 통계량과 COVRATIO에 대한 결과를 다양한 영향력 관찰점의 발생 위치 및 개수별로 각각 나누어서 서술한다. 끝으로 4절에서는 본 연구의 결론이 주어진다.

2. 선형혼합모형과 영향력 측도

이 절에서는 선형혼합모형과 본 연구에서 사용한 영향력 측도를 간략하게 설명한다.

2.1 선형혼합모형

선형혼합모형은 다음과 같이 기술된다.

$$\underline{y} = X\beta + Z\underline{u} + \underline{e} \tag{2.1}$$

여기서 $\underline{y}$는 알려진 $n\times 1$ 관측치 벡터, X는 고정효과에 관련된 $n\times p$ 계획행렬, Z는 랜덤효과에 관련된 $n\times q$ 계획행렬, β는 고정효과로 불리는 $p\times 1$ 열벡터, $\underline{u}$는 랜덤효과로 불리는 $q\times 1$ 열벡터, 그리고 $\underline{e}$는 $n\times 1$ 오차벡터이다. 또한 $\underline{u}$에 포함된 랜덤효과들에 대응되는 분산성분들을 표시하기 위하여 $\underline{u}$를 c개의 부분벡터 $\underline{u}' = (\underline{u}_1' | \underline{u}_2' | \cdots | \underline{u}_c')$과 같이 분할하고 $\underline{u}_i$에 대응되는 계획행렬을 Z_i로 두면 식(2.1)은 다음과 같은 식(2.2)로 표현할 수 있다.

$$\underline{y} = X\beta + Z_1\underline{u}_1 + Z_2\underline{u}_2 + \cdots + Z_c\underline{u}_c + \underline{e} \tag{2.2}$$

식(2.2)에서 Z_i의 차수는 $n\times m_i$ 행렬, $\underline{u}_i$는 $m_i\times 1$ 열벡터이며, q와 m_i의 관계식은 $q = \sum_{i=1}^{c} m_i$와 같다. 아울러 다음과 같은 사실이 성립한다고 가정한다.

$$\underline{u}_i \sim N(\underline{0}, \sigma_i^2 I_{m_i}),\quad \underline{e} \sim N(\underline{0}, \sigma^2 I_n),\quad Cov(\underline{u}_i, \underline{e}') = 0,\quad Cov(\underline{u}_i, \underline{u}_j) = 0,\quad {}^{\forall} i \neq j. \tag{2.3}$$

여기서 I_{m_i}와 I_n은 각각 차수가 $m_i\times m_i$, $n\times n$인 항등행렬이다. 또한 γ_i를 $i=1,2,\cdots,c$에 대하여 $\gamma_i = \sigma_i^2/\sigma^2$으로 정의하면 관측치 벡터 $\underline{y}$의 분산행렬 V는 $V = \sigma^2 H$, $H = I_n + \sum_{i=1}^{c} \gamma_i Z_i Z_i'$으로 나타낼 수 있다. 따라서 식(2.2)는 다음과 같은 선형모형으로 기술된다.

$$\underline{y} = X\beta + \underline{e},\quad \underline{e} \sim N(\underline{0}, \sigma^2 H) \tag{2.4}$$

한편 H는 정칙행렬이며 대칭이므로 $H=TT'$의 관계를 만족하는 정칙행렬 T가 존재하며, 역행렬 T^{-1}을 식(2.4)의 좌변과 우변에 각각 곱하면, $\underline{e}^{*}=T^{-1}\underline{e}$의 분산행렬이 $\sigma^2 I_n$가 되어 $\underline{\beta}$를 추정하는데 BLUE인 최소제곱추정량을 사용할 수 있다. 그러므로 변환된 모형에 대한 여러 가지 회귀진단도구들을 사용하면, 혼합모형 $\underline{y}$에 대한 진단도구들을 구할 수 있다.

2.2 영향력 측도

관측치의 영향력을 판단하기 위해서는 영향력 측도가 필요하다. 지금까지 제안된 회귀모형에 대한 영향력 측도는 매우 많은 종류가 있으나 크게 나누어 소거법, 무한소교란법, 국소 영향력, 대치법과 같은 네 가지로 분류된다. 이 중에서 소거법이 가장 많이 사용되며 많은 측도들이 이 방법에 의해 유도되었다. 4가지 방법에 대한 상세한 설명과 관계식은 강근석과 김충락(1999) 등에 나와 있다. 본 연구에서는 소거법을 이용한 측도 중에서 쿡 통계량과 COVRATIO 두 가지 만을 고려하였다. 왜냐하면 두 가지 통계량 이외에도 앤드류-프레기본 통계량, DFBETAS, DFFITS와 같은 회귀진단통계량이 많이 있으나 기준치가 명확하게 제안되어 있지 않은 이유로 본 연구에서는 제외하였다.

(1) 쿡 통계량

소거법에 의한 측도 중에서는 쿡 통계량(Cook's distance)이 가장 많이 사용되고 있으며, 회귀모형인 경우에는 SAS, SPSS, MINITAB 등의 여러 통계패키지에서 간단한 명령어에 의해 쉽게 구해진다. 혼합모형에 있어서 i번째 고정효과의 영향력을 평가하기 위하여 고정효과에 대한 쿡 통계량은

$$D_{\beta}=(\underline{b}-\underline{b}_{[i]})'(X'V^{-1}X)(\underline{b}-\underline{b}_{[i]})/p \tag{2.5}$$

로 정의되며, 이 경우 $\underline{b}$는 $\underline{\beta}$의 일반화최소제곱추정량 $\underline{b}=(X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$이며, $\underline{b}_{[i]}$는 i번째 관측치를 제외한 $n-1$개의 관측치로 구한 $\underline{\beta}$의 일반화최소제곱추정량이다. 분산행렬 V가 알려져 있는 경우에는 D_{β}를 자유도가 p인 χ^2-분포를 이용하여 영향력 관찰점 여부를 판단할 수 있으며, 추정된 경우에는 근사적으로 사용할 수 있다(Christensen et al, 1992).

(2) COVRATIO

i번째 관측치의 $Cov(\underline{b})$의 추정치에 대한 영향력을 고려하기 위해 $Cov(\underline{b})=\sigma^2(X'H^{-1}X)^{-1}$와 $Cov(\underline{b}_{[i]})=\sigma^2_{[i]}(X_{[i]}'H^{-1}X_{[i]})^{-1}$를 비교하는 것이 바람직하다. 여기서 $X_{[i]}$는 i번째 관측치를 제거시킨 $(n-1)\times p$행렬이다. 그런데 이들은 모두 행렬이므로 스칼라로 바꾸어 주는 작업이 필요하며, Belsley, Kuh와 Welsch(1980)는 회귀모형에 대한 행렬식의 비인 COVRATIO를 제안하였는데, 혼합모형에 대해서는 σ^2의 추정치로 s^2, i번째 관측치를 제외하고 구한 $\sigma^2_{[i]}$의 추정치로 $s^2_{[i]}$을 사용하여 다음과 같이 정의할 수 있다.

$$\text{COVRATIO}_i=\frac{\det[s^2_{[i]}(X_{[i]}'H^{-1}X_{[i]})^{-1}]}{\det[s^2(X'H^{-1}X)^{-1}]}. \tag{2.6}$$

식(2.6)에서 det[·]는 행렬식을 의미하며, 회귀모형인 경우에 $COVRATIO_i$의 기준치로 Belsley, Kuh와 Welsch(1980)는 $|COVRATIO_i - 1| \geq 3p/n$을 제시한 바 있다.

3. 모의실험

이 절에서는 쿡 통계량과 $COVRATIO_i$의 기준치들이 분산성분이 미지인 선형혼합모형에도 적용될 수 있는지를 모의실험을 통하여 알아본다.

3.1 사용된 분산성분추정량

분산성분들이 미지인 경우에는 데이터로부터 먼저 분산성분을 추정하고, 영향력 측도에 이를 대입하여 구하여만 한다. 혼합모형의 분산성분을 추정하기 위한 방법들은 첫째로 불편추정량의 원리에 입각한 추정량, 둘째는 우도원리에 입각한 추정량, 셋째는 베이지안 방법에 입각한 추정량들로 크게 분류된다. 이 중에서 본 연구에서는 대부분의 통계패키지가 제공하는 가장 보편화된 네 가지 추정량인 분산분석추정량(ANOVA 추정량), 최소분산이차불편추정량(MIVQUE), 최우추정량(MLE)과 제한최우추정량(REMLE)를 사용하였다. 여러 가지 분산성분추정량에 관한 보다 상세한 내용은 Rao와 kleffe(1988) 또는 Searle(1987)의 책에서 찾아볼 수 있다.

3.2 모의실험계획

모의실험에서는 가장 간단한 혼합모형인 일원변량모형을 선택하였다. 그리고 고려된 모의실험계획은 <표 1>과 같다. <표 1>에서 실험계획패턴의 괄호표기는 $(n_1, n_2, \cdots, n_p)$를 의미한다. 급내상관계수 $\rho = \sigma_a^2/(\sigma_a^2 + \sigma^2)$는 0.1부터 0.9까지 0.1 간격으로 선택되어졌으며 <표 1>에서 설명된 여러 가지 경우에 대하여 각각 100,000개의 자료가 생성되어졌다.

<표 1> 모의실험에 사용된 모형의 종류와 조건

모형	고려된 조건	사용된 조건의 값
$y_{ij} = \mu + a_i + e_{ij}$. $i = 1, 2, \cdots, p;\ j = 1, 2, \cdots, n_i$.	총관측치의 개수	$n = 15, 30$
	p의 값	$p = 3, 6$
	확률분포	$a_i \sim N(0, \sigma_a^2)$, $e_i \sim N(0, \sigma^2)$
	사용된 모수의 값	$\mu = 0$; $\sigma^2 = 1$
	급내상관계수	$\rho = 0.1, 0.2, \cdots, 0.9$
	독립성 가정	a_i, e_{ij} : 서로 독립
	실험계획패턴	P1 = (5,5,5), P2 = (2,5,8), P3 = (5,5,5,5,5,5), P4 = (2,2,5,5,8,8)

3.3 모의실험의 과정

일반적으로 영향력 관찰점이 이상점이 아닐 수도 있고, 이상점이 영향력 관찰점이 아닐 수도 있지만 실질적인 대부분의 경우에는 이상점이 영향력 관찰점이 되기 때문에 고정효과에 대한 영향력 관찰점의 후보로는 생성된 첫 번째 자료의 값에 5와 10을 각각 더한 값으로 정의를 하였다. 실

제로 정규분포이론에서 3배의 표준편차 밖으로 존재하는 관측치는 거의 존재하지 않기 때문에 급내상관계수 ρ의 값이 작으면 5를 더한 경우도 거의 발생하지 않는 경우임을 알 수 있다. 그러므로 적어도 이 자료들은 생성된 자료 내에서 가장 영향을 많이 주는 영향력 관찰점이라고 할 수 있으며, 이 경우 두 가지 측도가 이 자료를 영향력 관찰점이라고 지적할 수 있는 지를 확인하였다. 고려된 영향력 관찰점의 개수는 1개와 2개이고, 영향력 관찰점의 발생위치에 따른 변화를 확인하기 위하여 반복수가 적은 셀과 많은 셀에서 발생하는 두 가지 경우로 나누었다.

<표2> 영향력 관찰점의 발생위치 및 개수

영향력 관찰점의 개수	실험계획패턴의 종류			
	P1=(5,5,5)	P2=(2,5,8)	P3=(5,5,5,5,5,5)	P4=(2,2,5,5,8,8)
1개	C01=(5,★)	C02=(2,★)	C03=(5,★)	C04=(2,★)
		C05=(8,★)		C06=(8,★)
2개	C07=(5,★★)	C08=(8,★★)	C09=(5,★★)	C10=(8,★★)
	C11=(5,★|(5,★)	C12=(2,★|8,★)	C13=(5,★|5,★)	C14=(2,★|8,★)

<표 2>는 모의실험에서 사용된 영향력 관찰점의 개수 및 발생위치를 표시하며, 괄호 안의 왼쪽 숫자는 발생된 셀의 위치를 나타내며, 오른쪽의 ★의 개수는 영향력 관찰점의 개수이다. 예를 들어 C07은 반복수가 5개인 셀에 영향력 관찰점이 2개 발생한 경우이며, C12는 반복수가 2, 반복수가 8인 셀에 각각 1개씩의 영향력 관찰점이 발생한 경우를 나타낸다. 이상과 같이 불균형정도, 랜덤효과 수준의 수, 영향력 관찰점의 개수 및 위치를 고려한 모두 14가지 경우의 실험계획패턴을 이용하여 쿡 통계량의 경험 p-값과 COVRATIO 값이 $|COVRATIO_i - 1| < 3p/n$가 될 추정된 확률을 구하였다. 따라서 제시된 <표3>부터 표<5>는 만들어진 영향력 관찰점이 영향력 관찰점이 아니라고 판정할 추정된 근사확률 값이며, 다음과 같은 방법으로 구하였다.

먼저 모형에 대한 여러 가지 상이한 조건을 이용하여 정규난수를 만들고 분산성분을 4가지 방법을 이용하여 추정한다. 그리고 식(2.5)와 식(2.6)에 대입하여 계산한 뒤 두 통계량에 대한 기준치를 이용하여 영향력 관찰점 여부를 판가름하였다. 이 경우 쿡 통계량인 경우엔 모의실험에서 데이터를 가지고 구한 식(2.5)의 관측값을 이용하여 구한 p-값들의 전체 평균값이며, COVRATIO인 경우에는 $|COVRATIO_i - 1| < 3p/n$가 되는 전체회수를 전체반복수로 나누어서 구하였다.

각 REMLE와 MLE를 구하는 경우에 수렴조건을 만족하기 위한 최대반복횟수는 20번으로 제한하였고, 추정량이 수렴되지 않는 경우에는 제외하였다. 이 경우 MIVQUE를 구하는 경우에는 가장 많이 사용되는 사전추측값 1을 사용하였다. 난수생성은 통계패키지 SAS를 이용하였는데, 정규분포난수는 난수발생함수 RANNOR을 사용하여 생성되어졌다.

5를 더하여 만든 영향력 관찰점인 <표3>과 <표4>를 통하여 알 수 있는 중요한 사실들을 요약하면 다음과 같다.

1. 균형디자인 P1인 경우, C01를 살펴보면 쿡의 통계량에 대한 p-값을 살펴보면 영향력 관찰점을 잘 지적하지 못한다. 이러한 경향은 급내상관계수의 값이 커질수록 더욱 심해지며, 반면에 COVRATIO는 쿡 통계량보다 영향력 관찰점을 잘 지적한다. 영향력 관찰점이 2개인 C07을 살펴보면 쿡의 통계량이나 COVRATIO 모두 영향력 관찰점을 지적하지 못한다. 각각 다른 셀에

<표3> 5를 더하여 만든 영향력 관찰점인 경우

	ρ	Cook				Covratio			
		ANOVA	MIVQUE	REMLE	MLE	ANOVA	MIVQUE	REMLE	MLE
C01	0.1	0.28718	0.28718	0.28718	0.25108	0.05439	0.05258	0.05269	0.02929
	0.2	0.30970	0.30970	0.30970	0.26873	0.05136	0.04899	0.04904	0.02866
	0.3	0.33435	0.33434	0.33434	0.28859	0.04611	0.04387	0.04388	0.02760
	0.4	0.36140	0.36140	0.36140	0.31086	0.04153	0.03904	0.03904	0.02647
	0.5	0.39567	0.39566	0.39566	0.34106	0.03536	0.03338	0.03337	0.02351
	0.6	0.43628	0.43626	0.43626	0.37848	0.02862	0.02684	0.02687	0.01968
	0.7	0.48824	0.48822	0.48822	0.42782	0.02311	0.02147	0.02144	0.01623
	0.8	0.55579	0.55577	0.55577	0.49527	0.01687	0.01576	0.01577	0.01248
	0.9	0.65900	0.65898	0.65898	0.60448	0.01070	0.00980	0.00978	0.00834
C02	0.1	0.23407	0.25267	0.25961	0.22355	0.15768	0.14080	0.12166	0.08400
	0.2	0.24026	0.25958	0.26621	0.22820	0.15220	0.14174	0.12647	0.09290
	0.3	0.24797	0.26765	0.27407	0.23444	0.14723	0.14212	0.12932	0.09932
	0.4	0.25748	0.27719	0.28349	0.24241	0.14000	0.14099	0.13279	0.10281
	0.5	0.27063	0.28994	0.29607	0.25263	0.13203	0.14227	0.13428	0.10791
	0.6	0.28860	0.30659	0.31266	0.26698	0.12338	0.13584	0.12941	0.10750
	0.7	0.31332	0.32890	0.33464	0.28555	0.11362	0.13232	0.12656	0.10628
	0.8	0.35809	0.37053	0.37555	0.32214	0.10215	0.11824	0.11507	0.10091
	0.9	0.44851	0.45658	0.45968	0.40194	0.08488	0.10026	0.09666	0.08800
C03	0.1	0.44168	0.44169	0.44169	0.42116	0.05371	0.05056	0.05056	0.04039
	0.2	0.47639	0.47639	0.47639	0.45197	0.04749	0.04483	0.04482	0.03749
	0.3	0.51315	0.51313	0.51313	0.48600	0.03896	0.03660	0.03658	0.03272
	0.4	0.55396	0.55394	0.55394	0.52561	0.03017	0.02809	0.02813	0.02600
	0.5	0.59693	0.59690	0.59690	0.56846	0.02341	0.02097	0.02091	0.01988
	0.6	0.64330	0.64327	0.64327	0.61586	0.01673	0.01502	0.01502	0.01464
	0.7	0.69707	0.69704	0.69704	0.67211	0.01092	0.00938	0.00930	0.00924
	0.8	0.75580	0.75578	0.75578	0.73454	0.00719	0.00585	0.00586	0.00591
	0.9	0.82968	0.82967	0.82967	0.81411	0.00433	0.00341	0.00338	0.00340
C04	0.1	0.36133	0.36830	0.37615	0.36601	0.11342	0.10074	0.07046	0.05527
	0.2	0.37078	0.37957	0.38654	0.37187	0.11825	0.10589	0.08767	0.07465
	0.3	0.38384	0.39448	0.40094	0.38240	0.11838	0.11138	0.09876	0.08803
	0.4	0.40105	0.41292	0.41894	0.39776	0.11504	0.11352	0.10504	0.09505
	0.5	0.42383	0.43637	0.44193	0.41795	0.10678	0.11619	0.11005	0.10150
	0.6	0.45426	0.46650	0.47129	0.44561	0.09432	0.11102	0.10652	0.09928
	0.7	0.49576	0.50672	0.51073	0.48376	0.08172	0.10633	0.10481	0.09852
	0.8	0.55646	0.56527	0.56825	0.54089	0.06626	0.09413	0.09352	0.08863
	0.9	0.65639	0.66253	0.66414	0.63904	0.04855	0.07132	0.07109	0.06869
C05	0.1	0.31218	0.32350	0.31934	0.26530	0.01179	0.01594	0.01452	0.00755
	0.2	0.35521	0.36312	0.35949	0.29755	0.01218	0.01524	0.01397	0.00789
	0.3	0.39908	0.40407	0.40103	0.33176	0.01061	0.01313	0.01234	0.00759
	0.4	0.44416	0.44639	0.44397	0.36987	0.00982	0.01126	0.01045	0.00689
	0.5	0.49351	0.49307	0.49150	0.41428	0.00865	0.01014	0.00990	0.00657
	0.6	0.54133	0.53959	0.53919	0.46018	0.00734	0.00769	0.00767	0.00536
	0.7	0.60194	0.59933	0.59972	0.52337	0.00637	0.00631	0.00652	0.00521
	0.8	0.66783	0.66615	0.66738	0.59873	0.00493	0.00504	0.00495	0.00403
	0.9	0.75638	0.75594	0.75710	0.70524	0.00308	0.00318	0.00319	0.00251
C06	0.1	0.46900	0.48222	0.47830	0.44705	0.01782	0.03160	0.02711	0.02187
	0.2	0.52110	0.53027	0.52790	0.49099	0.01629	0.02462	0.02249	0.01953
	0.3	0.57420	0.57895	0.57775	0.53948	0.01330	0.01740	0.01639	0.01479
	0.4	0.62555	0.62675	0.62663	0.58935	0.01095	0.01225	0.01185	0.01173
	0.5	0.67642	0.67563	0.67622	0.64093	0.00756	0.00832	0.00821	0.00803
	0.6	0.72572	0.72401	0.72528	0.69411	0.00540	0.00552	0.00565	0.00565
	0.7	0.77588	0.77446	0.77583	0.75019	0.00394	0.00340	0.00352	0.00360
	0.8	0.82719	0.82677	0.82816	0.80912	0.00272	0.00201	0.00208	0.00227
	0.9	0.88421	0.88503	0.88580	0.87404	0.00150	0.00117	0.00116	0.00123
C07	0.1	0.45107	0.45104	0.45101	0.37325	0.37789	0.38741	0.38746	0.41743
	0.2	0.46783	0.46792	0.46790	0.39285	0.39697	0.40730	0.40752	0.42733
	0.3	0.48319	0.48335	0.48334	0.41083	0.40618	0.41669	0.41684	0.42638
	0.4	0.50278	0.50298	0.50297	0.43342	0.41032	0.42012	0.42036	0.42670
	0.5	0.52545	0.52566	0.52566	0.45846	0.40781	0.41765	0.41781	0.42292
	0.6	0.55315	0.55337	0.55337	0.48940	0.39673	0.40633	0.40665	0.41128
	0.7	0.58823	0.58842	0.58842	0.52792	0.37849	0.38706	0.38722	0.39019
	0.8	0.63732	0.63746	0.63746	0.58133	0.34881	0.35660	0.35667	0.35788
	0.9	0.71579	0.71584	0.71584	0.66781	0.30580	0.31381	0.31374	0.31340

<표4> 5를 더하여 만든 영향력 관찰점인 경우 (계속)

	ρ	Cook				Covratio			
		ANOVA	MIVQUE	REMLE	MLE	ANOVA	MIVQUE	REMLE	MLE
C08	0.1	0.37926	0.40189	0.39542	0.32011	0.31129	0.37454	0.37272	0.26935
	0.2	0.40401	0.42048	0.41512	0.34067	0.31766	0.36629	0.36665	0.28340
	0.3	0.43089	0.44193	0.43756	0.36233	0.31085	0.34820	0.35040	0.28449
	0.4	0.45966	0.46640	0.46254	0.38640	0.30111	0.32872	0.33118	0.27962
	0.5	0.49411	0.49687	0.49383	0.41597	0.28951	0.30876	0.31124	0.27486
	0.6	0.53568	0.53588	0.53360	0.45376	0.27105	0.28029	0.28443	0.26085
	0.7	0.58799	0.58636	0.58492	0.50539	0.24755	0.25152	0.25495	0.24209
	0.8	0.65392	0.65242	0.65212	0.57818	0.21797	0.21659	0.21982	0.21308
	0.9	0.74548	0.74510	0.74621	0.68800	0.17415	0.16963	0.17146	0.16970
C09	0.1	0.56582	0.56542	0.56534	0.52977	0.22370	0.22805	0.22838	0.23783
	0.2	0.59316	0.59307	0.59303	0.55840	0.25537	0.26356	0.26400	0.26391
	0.3	0.62101	0.62112	0.62110	0.58794	0.27566	0.28528	0.28566	0.28265
	0.4	0.65109	0.65131	0.65131	0.62019	0.28410	0.29377	0.29425	0.28932
	0.5	0.68220	0.68246	0.68247	0.65373	0.27687	0.28661	0.28696	0.28222
	0.6	0.71728	0.71753	0.71754	0.69163	0.25870	0.26827	0.26866	0.26464
	0.7	0.75621	0.75642	0.75642	0.73383	0.22944	0.23690	0.23704	0.23504
	0.8	0.80141	0.80154	0.80154	0.78293	0.19338	0.19749	0.19747	0.19603
	0.9	0.85988	0.85993	0.85993	0.84667	0.15203	0.15176	0.15165	0.15101
C10	0.1	0.53306	0.56441	0.55933	0.51830	0.21209	0.27341	0.28022	0.25573
	0.2	0.56760	0.58689	0.58449	0.54513	0.21467	0.25494	0.25877	0.24397
	0.3	0.60216	0.61272	0.61197	0.57344	0.20415	0.22518	0.22892	0.22186
	0.4	0.64133	0.64558	0.64543	0.60819	0.18789	0.19415	0.19833	0.19542
	0.5	0.68244	0.68260	0.68302	0.64718	0.16114	0.15887	0.16350	0.16438
	0.6	0.72713	0.72480	0.72571	0.69328	0.13438	0.12565	0.12912	0.13225
	0.7	0.77428	0.77149	0.77278	0.74545	0.10827	0.09568	0.09789	0.10050
	0.8	0.82583	0.82418	0.82548	0.80486	0.08097	0.06788	0.06906	0.07091
	0.9	0.88288	0.88314	0.88408	0.87135	0.05456	0.04584	0.04636	0.04678
C11	0.1	0.26671	0.26668	0.26669	0.25238	0.10027	0.09260	0.09291	0.07517
	0.2	0.28045	0.28043	0.28044	0.26001	0.11095	0.10386	0.10427	0.08364
	0.3	0.29563	0.29559	0.29560	0.26921	0.11813	0.11181	0.11225	0.09044
	0.4	0.31558	0.31552	0.31552	0.28277	0.12523	0.12030	0.12059	0.09835
	0.5	0.34204	0.34196	0.34196	0.30245	0.13010	0.12657	0.12665	0.10606
	0.6	0.37652	0.37642	0.37642	0.32954	0.13202	0.12939	0.12945	0.11319
	0.7	0.42054	0.42043	0.42043	0.36765	0.13414	0.13076	0.13081	0.11579
	0.8	0.48814	0.48803	0.48803	0.42981	0.12678	0.12501	0.12509	0.11503
	0.9	0.59619	0.59611	0.59611	0.53831	0.10587	0.10218	0.10216	0.09674
C12	0.1	0.66281	0.67048	0.64540	0.47076	0.01424	0.02078	0.02658	0.03553
	0.2	0.67066	0.67689	0.65532	0.49568	0.01561	0.02239	0.02730	0.03397
	0.3	0.67408	0.67935	0.66165	0.51853	0.01899	0.02577	0.02929	0.03287
	0.4	0.68062	0.68515	0.67045	0.54494	0.02136	0.02717	0.02963	0.03233
	0.5	0.68907	0.69258	0.68108	0.57412	0.02404	0.02883	0.03038	0.03138
	0.6	0.69626	0.69864	0.69033	0.60048	0.02660	0.03049	0.03179	0.03060
	0.7	0.70892	0.70962	0.70370	0.63221	0.02714	0.02935	0.03020	0.02814
	0.8	0.72817	0.72694	0.72414	0.66768	0.02726	0.02848	0.02915	0.02647
	0.9	0.77030	0.76822	0.76798	0.72430	0.02533	0.02521	0.02561	0.02338
C13	0.1	0.43065	0.43067	0.43067	0.41490	0.07209	0.06631	0.06656	0.05249
	0.2	0.45549	0.45550	0.45551	0.43550	0.07815	0.07284	0.07294	0.06168
	0.3	0.48336	0.48335	0.48336	0.45995	0.07801	0.07365	0.07373	0.06412
	0.4	0.51685	0.51682	0.51682	0.49085	0.07718	0.07409	0.07424	0.06761
	0.5	0.55477	0.55472	0.55472	0.52722	0.06923	0.06633	0.06635	0.06257
	0.6	0.60013	0.60006	0.60006	0.57226	0.06120	0.05882	0.05873	0.05645
	0.7	0.65299	0.65292	0.65292	0.62622	0.05110	0.04876	0.04878	0.04741
	0.8	0.71498	0.71493	0.71493	0.69119	0.03895	0.03625	0.03618	0.03549
	0.9	0.79913	0.79911	0.79911	0.78112	0.02427	0.02168	0.02166	0.02147
C14	0.1	0.45784	0.46874	0.46608	0.44124	0.02990	0.04768	0.04152	0.03521
	0.2	0.49565	0.50496	0.50277	0.47307	0.03674	0.05195	0.04819	0.04265
	0.3	0.53553	0.54185	0.54040	0.50738	0.03970	0.05056	0.04833	0.04430
	0.4	0.57959	0.58285	0.58230	0.54728	0.03881	0.04612	0.04528	0.04335
	0.5	0.62532	0.62570	0.62596	0.59034	0.03792	0.04046	0.04080	0.04064
	0.6	0.67638	0.67521	0.67619	0.64230	0.03292	0.03216	0.03368	0.03379
	0.7	0.73085	0.72943	0.73087	0.70082	0.02606	0.02425	0.02465	0.02532
	0.8	0.79128	0.79060	0.79226	0.76889	0.01966	0.01612	0.01685	0.01747
	0.9	0.85954	0.86048	0.86161	0.84673	0.01197	0.00953	0.00973	0.00997

<표5> 10을 더하여 만든 영향력 관찰점인 경우

	ρ	Cook				Covratio			
		ANOVA	MIVQUE	REMLE	MLE	ANOVA	MIVQUE	REMLE	MLE
C01	0.1	0.03359	0.03360	0.03360	0.01915	0.00001	0.00001	0.00001	0.00000
	0.2	0.03962	0.03963	0.03963	0.02245	0.00002	0.00002	0.00002	0.00000
	0.3	0.04783	0.04785	0.04785	0.02713	0.00000	0.00000	0.00000	0.00001
	0.4	0.05728	0.05730	0.05730	0.03294	0.00003	0.00004	0.00004	0.00001
	0.5	0.06979	0.06983	0.06983	0.04115	0.00002	0.00002	0.00002	0.00000
	0.6	0.08722	0.08726	0.08726	0.05311	0.00001	0.00002	0.00002	0.00000
	0.7	0.11303	0.11307	0.11307	0.07198	0.00000	0.00000	0.00000	0.00000
	0.8	0.15834	0.15838	0.15838	0.10731	0.00000	0.00000	0.00000	0.00000
	0.9	0.25678	0.25681	0.25681	0.19154	0.00000	0.00000	0.00000	0.00000
C02	0.1	0.03074	0.04576	0.05018	0.02355	0.20448	0.17395	0.18054	0.14968
	0.2	0.03238	0.04784	0.05236	0.02518	0.17486	0.14943	0.15616	0.12737
	0.3	0.03474	0.05098	0.05555	0.02737	0.15282	0.12998	0.13562	0.11124
	0.4	0.03736	0.05438	0.05888	0.02987	0.12593	0.11021	0.11394	0.09333
	0.5	0.04125	0.05917	0.06362	0.03357	0.10548	0.09249	0.09437	0.07962
	0.6	0.04702	0.06592	0.07020	0.03849	0.08318	0.07339	0.07472	0.06162
	0.7	0.05576	0.07574	0.07977	0.04587	0.06277	0.05736	0.05693	0.04673
	0.8	0.07188	0.09233	0.09621	0.05864	0.04203	0.03880	0.03863	0.03119
	0.9	0.11368	0.13240	0.13618	0.09045	0.02246	0.02020	0.01925	0.01651
C04	0.1	0.08927	0.10510	0.11896	0.09635	0.12525	0.10026	0.09437	0.07662
	0.2	0.09326	0.11098	0.12411	0.10097	0.09899	0.08330	0.08511	0.07087
	0.3	0.09864	0.11850	0.13086	0.10672	0.07418	0.07109	0.07200	0.06167
	0.4	0.10609	0.12825	0.13959	0.11460	0.05421	0.05563	0.05893	0.05090
	0.5	0.11614	0.14068	0.15108	0.12493	0.03793	0.04085	0.04543	0.03864
	0.6	0.13160	0.15815	0.16721	0.13973	0.02216	0.02877	0.03218	0.02776
	0.7	0.15411	0.18199	0.18954	0.16050	0.01255	0.01744	0.01999	0.01746
	0.8	0.19666	0.22415	0.23000	0.19859	0.00573	0.00858	0.00954	0.00804
	0.9	0.29215	0.31440	0.31799	0.28326	0.00148	0.00248	0.00241	0.00227
C07	0.1	0.16575	0.16541	0.16536	0.08710	0.43530	0.43997	0.43918	0.53899
	0.2	0.17132	0.17117	0.17113	0.09298	0.46480	0.46973	0.46936	0.52518
	0.3	0.17837	0.17838	0.17836	0.10040	0.48824	0.49352	0.49320	0.51806
	0.4	0.18731	0.18745	0.18743	0.10960	0.50141	0.50400	0.50379	0.51018
	0.5	0.19874	0.19900	0.19899	0.12136	0.50589	0.50492	0.50482	0.49293
	0.6	0.21366	0.21400	0.21400	0.13669	0.49607	0.48880	0.48898	0.46759
	0.7	0.23566	0.23605	0.23605	0.15900	0.46734	0.45468	0.45457	0.43124
	0.8	0.27220	0.27261	0.27262	0.19594	0.40615	0.38863	0.38864	0.36719
	0.9	0.34891	0.34923	0.34923	0.27294	0.29318	0.27268	0.27252	0.25878
C08	0.1	0.08164	0.11410	0.10373	0.04148	0.22111	0.35646	0.32229	0.07892
	0.2	0.09427	0.12084	0.11239	0.04931	0.23502	0.33283	0.31156	0.09554
	0.3	0.10673	0.12781	0.12107	0.05751	0.23670	0.30714	0.29406	0.10422
	0.4	0.11958	0.13588	0.13060	0.06608	0.23440	0.28236	0.27474	0.11054
	0.5	0.13606	0.14744	0.14357	0.07781	0.22217	0.25642	0.25386	0.11502
	0.6	0.15664	0.16328	0.16071	0.09238	0.20289	0.22877	0.22768	0.11194
	0.7	0.18810	0.18921	0.18789	0.11494	0.17867	0.19415	0.19490	0.10643
	0.8	0.24195	0.23793	0.23771	0.15599	0.13916	0.14688	0.15012	0.08950
	0.9	0.36257	0.35621	0.35779	0.26379	0.08417	0.08582	0.08811	0.06005
C09	0.1	0.29281	0.29127	0.29098	0.24100	0.22536	0.21793	0.21626	0.25313
	0.2	0.30631	0.30537	0.30518	0.25528	0.27786	0.27055	0.26924	0.29184
	0.3	0.32329	0.32286	0.32275	0.27325	0.32091	0.31030	0.30943	0.31523
	0.4	0.34241	0.34243	0.34239	0.29363	0.34209	0.32835	0.32750	0.32554
	0.5	0.36644	0.36677	0.36675	0.31913	0.33154	0.31234	0.31150	0.30703
	0.6	0.39804	0.39857	0.39858	0.35256	0.29230	0.26648	0.26559	0.26126
	0.7	0.43965	0.44023	0.44025	0.39654	0.22024	0.19204	0.19144	0.18802
	0.8	0.50170	0.50222	0.50223	0.46206	0.13539	0.10860	0.10785	0.10730
	0.9	0.60776	0.60805	0.60805	0.57460	0.05057	0.03379	0.03359	0.03408
C10	0.1	0.21404	0.29341	0.27620	0.21244	0.20198	0.30822	0.30588	0.20689
	0.2	0.23412	0.29521	0.28354	0.22455	0.19459	0.25398	0.24913	0.18077
	0.3	0.25641	0.30195	0.29453	0.23919	0.17380	0.20626	0.20063	0.15529
	0.4	0.27972	0.31184	0.30789	0.25507	0.14275	0.16206	0.15647	0.12788
	0.5	0.31058	0.32956	0.32859	0.27763	0.10835	0.11764	0.11285	0.09712
	0.6	0.34957	0.35779	0.35874	0.30829	0.07548	0.08139	0.07966	0.07135
	0.7	0.40534	0.40394	0.40608	0.35562	0.04462	0.04781	0.04691	0.04475
	0.8	0.49277	0.48529	0.48795	0.43935	0.01978	0.02063	0.02143	0.02208
	0.9	0.63134	0.62525	0.62829	0.58963	0.00477	0.00455	0.00472	0.00548

영향력 관찰점이 2개 존재하는 경우인 C11인 경우는 같은 셀에 영향력 관찰점이 2개 존재하는 C07보다는 양호하나 여전히 영향력 관찰점을 잘 지적하지 못한다.

2. 불균형디자인 P2인 경우, 영향력 관찰점의 개수가 1개인 경우에 C02와 C05를 C01과 비교하면 불균형정도보다는 영향력 관찰점이 존재하는 셀의 위치에 더욱 민감함을 살펴볼 수 있다. 쿡의 통계량은 셀의 도수가 적은 곳에 영향력 관찰점이 위치하는 경우가 영향력 관찰점을 좀 더 잘 지적하나 COVRATIO는 그 반대이다. 영향력 관찰점이 2개인 C08과 C12를 비교하면 C12의 COVRATIO만이 영향력 관찰점을 지적하고 나머지 경우는 지적을 못한다.
3. 균형디자인 P3인 경우, C03과 C01, C09와 C07, C13과 C11를 비교하면 셀의 개수가 늘어나면 쿡의 통계량은 영향력 관찰점을 지적할 가능성은 더 떨어지나 COVRATIO는 그 반대이다. 또한 불균형디자인 P4인 경우도 셀의 개수가 늘어나면 C04과 C06을 C02와 C05와 각각 비교하여 살펴볼 때 균형디자인과 마찬가지 결론이 성립한다.
4. 영향적 관찰점을 조사하는 경우에 분산성분추정량의 선택은 ANOVA추정량, MIVQUE와 REMLE는 비슷한 효율성을 가지나 MLE는 세 가지 추정량보다 쿡의 통계량 및 COVRATIO에 대하여 영향력관찰점을 잘 지적하기 때문에 네 가지 추정량 중에서 영향력 측도에 사용될 가장 바람직한 추정량이라고 판단되어진다.

10을 더하여 만든 영향력 관찰점의 결과인 <표5>를 <표3>과 <표4>와 비교하면 쿡 통계량은 상대적으로 모든 실험계획패턴에서 영향력 관찰점을 잘 지적한다. 따라서 쿡 통계량은 영향력 관찰점의 크기에 영향을 많이 받는다고 할 수 있다. 반면에 COVRATIO는 빈도수가 적은 셀에 영향력 관찰점이 존재하는 C02와 C04, 또 같은 셀에 영향력 관찰점이 2개가 존재하는 C07에서 C10에는 영향력 관찰점의 크기의 변화가 큰 영향을 미치지 않음을 확인할 수 있었다.

4. 결론

본 연구에서는 영향력 측도로 회귀모형에서 기준치가 명확하게 제안되어 있는 쿡 통계량과 COVRATIO를 사용하여 선형혼합모형에도 적용가능한 지를 모의실험을 통하여 살펴보았다. 제한된 모의실험의 결과, 두 통계량의 값은 영향력 관찰점의 개수 및 발생 셀의 위치에 매우 큰 영향을 받는 것으로 판명되었으며, 회귀모형의 기준치를 사용한 COVRATIO는 전반적으로 너무 보수적으로 판정하며 또한 영향력 관찰점이 많고, 같은 셀에 많이 발생하는 경우에는 적당하지 않은 것으로 간주되어진다. 그리고 χ^2-분포를 이용한 쿡 통계량의 기준치는 혼합모형에서 영향력 관찰점이 매우 특이한 극단값인 경우를 제외한 대부분의 경우에 적당하지 않은 것으로 판명되었다. 또한 본 연구는 선형혼합모형의 고정효과에 대한 영향력 관찰점의 지적에 대한 논의였지만 Hurtado (1993)의 연구에 의하면 고정효과에 대한 영향력 관찰점은 랜덤효과의 영향력 관찰점이며, 오차항의 분산성분추정에 같이 영향을 미친다고 알려져 있으며, 따라서 랜덤효과와 분산성분에의 추정에도 비슷한 결론에 도달될 것으로 간주되어진다.

본 연구에서는 논의되지 않았지만, 혼합모형에 있어서 두 통계량의 타당한 기준치를 정하는 문제는 향후 논의되어야 할 중요한 주제가 될 것이다. 또한 본 연구에서 사용한 랜덤효과의 공분산 구조가 특정구조가 아닌 일반적인 구조를 가질 때 쿡 통계량과 COVRATIO에 대한 보다 폭넓은

접근도 반드시 도전하여야 할 중요과제라고 생각되어진다.

참 고 문 헌

[1] 강근석, 김충락 (1999). 「회귀분석」, 교우사.

[2] Beckman, R. J., Nachtsheim, C. J., and Cook, R. D. (1987). Diagnostics for Mixed-Model Analysis of Variance, *Technometrics,* 29, 413-426.

[3] Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics,* New York: John Wiley.

[4] Christensen, R., Pearson, L. M. and Johnson, W. (1992). Case-Deletion Diagnostics for Mixed Models, *Technometrics,* 34, 38-44.

[5] Cook, R. D. (1977). Detection of Influential Observations in Linear Regression, *Technometric* , 19, 15-18.

[6] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression,* New York: Chapman and Hall.

[7] Fellner, W. H. (1986). Robust Estimation of Variance Components, *Technometrics,* 28, 51-60.

[8] Hocking, R. R. (1983). A Diagnostic Tool for Mixed Models with Application to Negative Estimates of Variance Components, *SUGI 8,* 711-716.

[9] Hocking, R. R. and Bremer, R. H. (1987). Estimation of Variance Components in Mixed Factorial Models Including Model-Based Diagnostics, *SUGI 12,* 1162-1167.

[10] Hurtado, G. (1993). *Detection of Influential Observations in Linear Mixed Models,* Ph.D. dissertation, North Carolina State University.

[11] Rao, C. R. and Kleffe, J. (1988). *Estimation of Variance Components and Applications,* Amsterdam: North-Holland.

[12] Searle, S. R. (1987). *Linear Models for Unbalanced Data,* John Wiley & Sons, New York.

[2002년 12월 접수, 2003년 3월 채택]