The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 119-128

# A Study On An Identification Of Interactions In A Nonreplicated Two-Way Layout With $L_1$-Estimation[1]

Ki Hoon Lee[2]

## Abstract

This paper proposes a method for detecting interactions in a two-way layout with one observation per cell. The identification of interactions in the model is not clear for they are confounding with error terms. The $L_1$-estimator is robust with respect to a y-direction outlier in linear model, so we are able to estimate main effects without affection of interactions. If an observation is classified as an outlier, we conclude it contains an interaction. An empirical study compared with a classical method is performed.

## 1. 서 론

본 연구에서 고려하는 이원배치자료의 수리적 모형은 다음과 같다.

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \quad ; i=1,\dots,a \ ; j=1,\dots,b,$$

여기서, $\mu$는 모평균, $\alpha_i$는 $i$블록효과, $\beta_j$는 $j$처리효과, $(\alpha\beta)_{ij}$는 $i$블록과 $j$처리의 교호작용, $\varepsilon_{ij}$는 오차항이다. 이때 교호작용 검정에 관한 가설은 다음과 같다.

$$H_o : (\alpha\beta)_{ij} = 0, \text{ for all } i, j .$$

그런데 반복이 없는 이원배치모형에서는 교호작용이 오차항과 교락되어 있어 이를 검정하는 것이 용이하지 않다. 또한 교호작용이 존재함에도 불구하고 이를 무시하고 주효과(main effect)에 대한 추론을 할 때, 오차항이 과대 추정되어 왜곡된 처리효과의 검정을 감수할 수밖에 없는 실정이다. 그래서 교호작용의 존재가 의심되면 각 셀에서 반복 측정된 자료를 얻어야하는데, 이러한 실험횟수의 증가는 경제적인 비용의 증가를 가져오므로 반복이 없는 이원배치에서 교호작용의 검출법의 제안은 이론적용의 오류와 경제성을 개선하는데 무척 필요하다고 할 수 있다.

Tukey(1949)는 교호작용의 형태가 승교호작용(product interaction: $(\alpha\beta)_{ij} = i \cdot \beta_j$)일 때, 자유도 1을 분리하여 교호작용을 위한 검정 통계량을 제안하였다. 그후로 Mandel(1961), Johnson과 Graybill(1972), Milliken과 Rasmuson(1977), Shukla(1982), Milliken과 Johnson(1989), Piepho (1994) 등이 Grubbs(1948)의 등분산 검정 아이디어를 이용한 갖가지 형태의 모수적 방법을 제안

1) 본 논문은 1998년도 전주대학교 학술연구비 연구소 지원과제에 의하여 연구되었음.
2) Associate Professor, School of Business, Jeonju University, Chonju, Chonbuk, 560-759, Korea

하고 있으나, 이들은 교호작용에 대한 가정이 없는 반면 여전히 몇가지 형태의 교호작용에 관해서는 검정할 수 없었다. 이기훈(1997)은 교호작용이 존재할 때 이와 무관하게 주효과를 검정하는 비모수적 검정법을 제안하였으나 교호작용의 검정은 불가능하였다.

본 연구에서는 오차항과 분리하여 교호작용을 검출하기 위해 로버스트 추론법의 아이디어를 사용한다. 2장에서는 실험계획 모형에서의 $L_1$-추정법의 소개를, 3장에서는 이에 의한 교호작용 검출방법을 제안하고, 모의실험에 의해 특성을 살펴보았다.

# 2. 실험계획모형에서 $L_1$-추정법

## 2.1 교호작용의 형태

일반적인 교호작용의 가정은 다음과 같은 식을 만족하도록 되어있다.

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0 \tag{2.1}$$

그러나 이러한 가정은 Terbeck과 Davies(1998), Daniel(1960)이 지적한 바와 같이 바람직한 가정은 아니다. 예를 들어 <표 1>과 같은 3×3 실험계획에서 주효과는 없고 단지 (1,3)셀에 교호작용이 존재한다고 가정하자. 그러나 이것이 식 (2.1)을 만족하기 위해서는 우측의 표와 같이 변해야 하는데, 단순한 구조의 교호작용 형태가 복잡하여짐은 물론이고, 최초에 의도했던 교호작용의 형태를 표현하지 못한다. 그래서 본 논문에서는 제약식 (2.1)을 가정하지 않은 더 일반적인 교호작용의 형태에서 추정법을 제안한다.

<표 1> 교호작용의 가정과 형태

| | | |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |

1) 한 셀에서 교호작용 존재

| | | |
|---|---|---|
| -2/9 | -2/9 | 4/9 |
| 1/9 | 1/9 | -2/9 |
| 1/9 | 1/9 | -2/9 |

2) 모든 셀에서 교호작용 존재((2.1) 가정)

## 2.2 $L_1$-추정법

앞절에서 소개한 이원배치 모형은 다음과 같이 변환할 수 있다.

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij} \\ &= \mu + \alpha_i + \beta_j + E_{ij} \qquad ; \quad i = 1, \dots, a\ ; j = 1, \dots, b, \end{aligned}$$

여기서 $E_{ij}$는 오차항에 이상값이 더해진 오염된 분포(contaminated distribution)에서 뽑은 확률변수라 간주한다. 일반적으로 교호작용이 모든 셀에서 존재하는 것이 아니고, 일부 셀에서 발생한다고 가정하는 것이 합리적이다. 오염된 셀의 비율이 적다면 로버스트 회귀추정법에 의하여 오염에 무관하게 주효과를 추정할 수 있다.

본 연구에서는 점근붕괴점(asymptotic breakdown point)의 관점에서 로버스트한 방법을 고려하고자 한다. 즉 다음 식을 최소화하는 $L_1$-추정법을 이용하여 $\alpha_i$, $\beta_j$를 추정한다.

$$\sum_i \sum_j |y_{ij} - \alpha_i - \beta_j| \ .$$

일반적으로 $L_1$-회귀추정량의 붕괴점은 1/n이므로, $L_1$-추정법은 점근붕괴점 관점에서 로버스트하지 않다. 그러나 설명변수에 지렛대점(leverage point)이 존재하지 않을 때, $L_1$-회귀추정법이 종속변수(y-방향)의 이상값에 로버스트하다는 사실은 잘 알려져 있다. (Rousseeuw와 Leroy(1987), Kim(1995)) 본 논문에서 고려하는 실험계획 모형에는 지렛대점 설명변수가 존재하지 않으므로 $L_1$-추정법이 로버스트한 회귀추정법이라 할 수 있다. 즉, 오염된(교호작용을 포함한) 오차항을 갖는 셀의 비율이 붕괴율 미만이면, $L_1$-추정법으로 교호작용과 무관하게 $\alpha_i$와 $\beta_j$를 추정할 수 있다.

## 2.3 완전적합모형(Exact fit model)의 예제

이 절에서는 다음과 같이 오차항이 0인 완전적합 실험계획모형에서 $L_1$-추정법으로 주효과를 추정하는 예를 들어본다. 각 처리효과와 교호작용효과가 다음과 같다고 하자. (<표 2 참조>)

$$\mu = 0\,,\ \ \alpha_i = i-1,\ \ \beta_j = j-1,\ \varepsilon_{ij} = 0;\ \ i=1,\dots,5;\ j=1,\dots,4,$$
$$(\alpha\beta)_{41} = (\alpha\beta)_{51} = 2\,,\ \ (\alpha\beta)_{14} = (\alpha\beta)_{24} = 1,\ \text{나머지 교호작용은 0.}$$

여기서 일반적인 회귀계수 추정방법인 최소제곱법(LS)에 의한 모수추정과 최소절대오차법(LAD, $L_1$)에 의한 회귀계수의 로버스트 추정 결과는 <표 3>과 같다.

<표 2> 교호작용이 있는 이원배치

| 처리 / 블록 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 4 =3+1 |
| 2 | 1 | 2 | 3 | 5 =4+1 |
| 3 | 2 | 3 | 4 | 5 |
| 4 | 5 =3+2 | 4 | 5 | 6 |
| 5 | 6 =4+2 | 5 | 6 | 7 |

<표 3> 회귀계수(주효과) 추정값

| 모수 | 실제값 | LSE | LAD |
|---|---|---|---|
| $\mu$ | 0 | 0.75 | 0 |
| $\beta_1$ | 0 | 0 | 0 |
| $\beta_2$ | 1 | 0.20 | 1 |
| $\beta_3$ | 2 | 1.20 | 2 |
| $\beta_4$ | 3 | 2.60 | 3 |
| $\alpha_1$ | 0 | 0 | 0 |
| $\alpha_2$ | 1 | 1.00 | 1 |
| $\alpha_3$ | 2 | 1.75 | 2 |
| $\alpha_4$ | 3 | 3.25 | 3 |
| $\alpha_5$ | 4 | 4.25 | 4 |

<표 3>에서 알 수 있듯이 $L_1$-회귀추정법은 교호작용의 존재에 관계없이 주효과를 정확히 추정하였다. <표 3>에서 구한 추정값에 따라 각 자료의 잔차를 계산하면 다음 <표 4>, <표 5>와 같다. <표 4>의 모수적 추정은 교호작용의 영향 때문에 주효과의 추정이 정확히 이루어지지 않아 잔차가 정확하지 못하고, <표 5>에서는 $L_1$-추정법에 의한 잔차가 교호작용을 더한 오차와 일치함을 알 수 있다.

<표 4> LSE에 의한 잔차

| 처리 / 블록 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | -0.75 | 0.05 | 0.05 | 0.65 |
| 2 | -0.75 | 0.05 | 0.05 | 0.65 |
| 3 | -0.50 | 0.30 | 0.30 | -0.10 |
| 4 | 1.00 | -0.20 | -0.20 | -0.60 |
| 5 | 1.00 | -0.20 | -0.20 | -0.60 |

<표 5> $L_1$-추정법에 의한 잔차

| 처리 / 블록 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 |
| 2 | 0 | 0 | 0 | 1 |
| 3 | 0 | 0 | 0 | 0 |
| 4 | 2 | 0 | 0 | 0 |
| 5 | 2 | 0 | 0 | 0 |

## 3. $L_1$-추정법에 의한 교호작용 검출

### 3.1 $L_1$-추정량의 붕괴점

단순회귀의 M-추정량의 붕괴점은 실험점의 위치에 따라 다르지만 최대 29%까지 가능하다. 이는 설명변수가 원점을 중심으로 등간격으로 배열되어 있을 때 $1-2^{-1/2} \approx 0.29289$ 에 의해 구해진다.(Jureckova와 Sen(1996)) $L_1$-회귀추정량도 M-추정량의 형태이기 때문에 본 논문의 실험계획 모형에서는 점근 붕괴점의 값이 매우 작아진다. 실제로 앞의 <표 2>의 실험계획에서는 다음 정리에서 보이는 바와 같이 최소 1개, 최대 4개의 교호작용까지만 찾을 수 있다. 즉, 최대 가능 붕괴점이 20%(=4/20)인 셈이다.

일반적으로는 다음과 같은 정리에 의해 붕괴점을 유도할 수 있다. 우선 다음과 같이 표본의 수가 $n$인 다중회귀 모형을 가정하자.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} .$$

이때, $\boldsymbol{y}$-방향으로만 이상점이 존재한다고 가정하고, 고정된 $\boldsymbol{X}(n\times p)$에 관하여 $m^*$을 $\boldsymbol{\beta}$의 추정량 $\boldsymbol{T}_n$의 붕괴점이라 하자.( $n=ab, p=a+b-1$ ) 그리고 크기가 $m$인 $N=\{1,2,\ldots,n\}$의 부분집합 $M$에 대하여

$$\inf_{||\boldsymbol{b}||=1} \left\{ \sum_{i\in M} |\boldsymbol{x}_i' \boldsymbol{b}| \,/\, \sum_{i\in N} |\boldsymbol{x}_i' \boldsymbol{b}| \right\} > \frac{1}{2} \tag{3.1}$$

을 만족하는 최대 정수를 $m_*$라 하면 정리 3.1이 성립한다. 이 정리는 일반적인 독립변수의 형태의 점근 붕괴점에 관한 Ellis와 Morgenthaler(1992)의 정리를 Jureckova와 Sen(1996)이 고정된 독립변수 행렬에 관하여 변형한 것이다.

**정리 3.1** $T_n$을 다음 식을 만족하는 $\beta$에 관한 $L_1$-추정량이라 하자.

$$\sum_{i=1}^{n} |Y_i - x_i' t| := \min, \quad t \in R^p.$$

이때 $T_n$의 붕괴점 $m^*$에 관하여 다음 부등식이 성립한다.

$$m_* + 1 \le m^* \le m_* + 2 .$$ ■

예를 들어 반복이 없는 5×4 이원배치 실험계획의 경우, $X$는 상수열을 포함한 20×10 행렬이고 이들의 원소는 0 또는 1이다. 그리고 $y = (y_{11}, y_{12}, \ldots, y_{54})'$, $\beta = (\mu, \alpha_1, \ldots, \alpha_5, \beta_1, \ldots, \beta_4)'$이다. 만약 $y_{11}$이 이상값이라면 다음이 성립한다.

$$\begin{aligned}
&\inf \left\{ \sum_{i \in N \setminus M} |x_i' b| \,/\, \sum_{i \in N} |x_i' b| \right\} \\
&= \inf \frac{|\mu + \alpha_1 + \beta_2| + |\mu + \alpha_1 + \beta_3| + \cdots + |\mu + \alpha_5 + \beta_4|}{|\mu + \alpha_1 + \beta_1| + |\mu + \alpha_1 + \beta_2| + \cdots + |\mu + \alpha_5 + \beta_4|} \\
&= \inf \left( 1 - \frac{|\mu + \alpha_1 + \beta_1|}{|\mu + \alpha_1 + \beta_1| + |\mu + \alpha_1 + \beta_2| + \cdots + |\mu + \alpha_5 + \beta_4|} \right) > \frac{1}{2} .
\end{aligned}$$

그러나 $y_{11}$, $y_{12}$가 이상값이라면

$$\begin{aligned}
&\inf \frac{|\mu + \alpha_1 + \beta_3| + |\mu + \alpha_1 + \beta_4| + \cdots + |\mu + \alpha_5 + \beta_4|}{|\mu + \alpha_1 + \beta_1| + |\mu + \alpha_1 + \beta_2| + \cdots + |\mu + \alpha_5 + \beta_4|} \\
&= \inf \left( 1 - \frac{|\mu + \alpha_1 + \beta_1| + |\mu + \alpha_1 + \beta_2|}{|\mu + \alpha_1 + \beta_1| + |\mu + \alpha_1 + \beta_2| + \cdots + |\mu + \alpha_5 + \beta_4|} \right) = \frac{1}{2}
\end{aligned}$$

이 되므로 붕괴점은 $m^* = 1$이 된다. 그러나 같은 행이나 열에서 이상값이 중복되어 나오지 않는다면, 예를 들어 $y_{11}$, $y_{22}$가 이상값이면 위와 같은 계산에 의해 식 (3.1)을 만족한다. 즉, 각 행이나 열에 이상값이 한번씩만 나올 때는 식 (3.1)을 만족하므로, 최대 4개의 이상값에도 $L_1$-추정법이 로버스트하다. 그러므로 위의 예제에서는 최대 가능 붕괴점은 4/20이고, 일반적으로 행과 열의 수가 같고 이상값이 각 행과 열에 이상적으로 배열된다면 최대 가능 점근 붕괴점은 25%에 달할 수 있다.

회귀모형추정에 있어서 M-추정량 형태의 추정량은 붕괴점이 낮고, 그나마 최적의 M-추정량 범함수(functional)의 형태도 유도되어 있지 않다. 그러나 오차항이 없는 완전적합 실험계획모형에서 $L_1$-추정법이 최적임이 Terbeck과 Davies(1998)에 의해 증명되었기 때문에 본 논문에서는 주효과를 추정하는데 $L_1$-추정법을 이용할 것을 제안한다.

## 3.2 교호작용의 검출

이원배치에서 이상값 또는 교호작용의 판정은 다음 절차를 따른다. $L_1$-추정량에 의하여 잔차

$$r_{ij} = y_{ij} - \widehat{y_{ij}}$$

를 구하고, 이를 표준화한 $r_{ij}/s$의 절대값을 일정한 값과 비교하여 이 값을 초과하면 그 셀에서

이상값, 즉 교호작용이 존재한다고 판단한다.

척도모수에 관한 추정량 $s$에 관한 최초 추정량은 다음 식 (3.2)를 사용한다.

$$s_0 = 1.4826\sqrt{\text{med}^* r_{ij}^2}\,, \tag{3.2}$$

여기서 $\text{med}^* r_{ij}^2$는 $r_{ij}^2$중에서 0이 되는 잔차를 제외한 나머지 값들의 중앙값이다.

흔히 척도추정량 초기값으로 $1.4826\sqrt{\text{med}\, r_{ij}^2}$ 을 사용하기도 하지만, 이는 중회귀모형에서 편차를 너무 작게 추정하는 경향이 있다. 그리고 Rousseeuw와 Leroy(1987)가 제안한 수정계수를 포함한 추정량

$$s_0 = 1.4826(1+5/(n-p))\sqrt{\text{med}\, r_{ij}^2} \tag{3.3}$$

도 모의실험에 의하면 좋은 추정값이 되지 못함을 확인 할 수 있었다.

$L_1$-회귀추정법의 특성상 회귀선이 p개의 관측값을 통과하므로 p개의 잔차가 0이 되어 잔차를 과소추정하는 경향이 있다. 그래서 본 논문에서는 과소추정하는 경향을 제거한 식 (3.2)를 제안하였는데, 실제 사용시에는 $\boldsymbol{X}$행렬의 종속성 때문에 추정하는 모수의 수가 ( $p=a+b-1$)개이므로 0이 되는 ( $a+b-1$)개의 잔차를 제외하고 나머지 잔차 제곱값들의 중앙값을 사용하였다. 모의실험 결과 (3.2)의 추정식이 다른 방법에 비해 우수한 것으로 판단되었다.

최초 추정량을 이용해 유도한 최종적인 척도추정량은 다음과 같이 정의한다.

$$s=\sqrt{\frac{\sum_i\sum_j w_{ij} r_{ij}^2}{(\sum_i\sum_j w_{ij}-p)}}\ ,$$

단, $w_{ij}=\begin{cases}1, & \text{만약 } \left|\dfrac{r_{ij}}{s_0}\right|\le 2.5\\ 0, & \text{기타의경우}\ .\end{cases}$

다음은 잔차를 최종 척도추정값으로 표준화한 뒤 2.5와 비교하는데, 즉 $|r_{ij}/s|\ge 2.5$이면 $(\alpha\beta)_{ij}\neq 0$이고 교호작용이 존재하는 것으로 판정한다. 여기서 기준값 2.5는 로버스트 추론에서 즐겨 쓰는 값이지만 임의적인 값임에는 틀림없다. 그러나 다음절의 모의실험에 의하면 척도추정과 교호작용이 존재하지 않는 상황의 경험적 유의수준을 조절하는데 매우 적절한 값임을 보여준다.

### 3.3 경험적 검정력(Empirical power) 비교

이 절에서는 제안한 $L_1$-방법과 모수적인 방법을 5×5 실험계획에서 비교하겠다. 모수적 방법은 적은 $a$, $b$에도 매우 우수한 근사를 갖는 Shukla(1982) 방법을 사용하였다. 이는 Bartlett 형태의 동질성 검정을 이용해 교호작용 검정을 제안하였는데, 다음과 같은 가설을 검정한다.

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2\ ,$$

여기서 $\sigma_k^2 \equiv \sigma^2 + \dfrac{a(a-1)\theta_k - \sum_i^a \theta_i}{(a-1)(a-2)(b-1)}$ , $\theta_k = \sum_{j=1}^{b}((\alpha\beta)_{kj} - (\overline{\alpha\beta})_{k.} - (\overline{\alpha\beta})_{.j} + (\overline{\alpha\beta})_{..})^2$ ,

$(\overline{\alpha\beta})_{k.} = \sum_j(\alpha\beta)_{kj}/b$, $(\overline{\alpha\beta})_{.j} = \sum_i(\alpha\beta)_{ij}/a$, $(\overline{\alpha\beta})_{..} = \sum_i\sum_j(\alpha\beta)_{ij}/ab$.

이때 $\sigma_k^2$의 불편추정량은 다음과 같다.

$$\widehat{\sigma_k}^2 = \frac{a(a-1)W_k - \sum_i^a W_i}{(a-1)(a-2)(b-1)}\ ,\ 단\ \ W_k = \sum_{j=1}^{b}(y_{kj} - \bar{y}_{k.} - \bar{y}_{.j} + \bar{y}_{..})^2,$$

$\bar{y}_{k.} = \sum_j y_{kj}/b,\quad \bar{y}_{.j} = \sum_i y_{ij}/a,\quad \bar{y}_{..} = \sum_i\sum_j y_{ij}/ab.$

여기서 각 행의 분산추정량으로 Milliken과 Rasmuson(1977)이 제안한

$$s_k^2 = \frac{\sum_j^b (y_{kj} - \bar{y}_{k.})^2}{b-1}\ ,\quad k=1,2,\dots,a$$

을 사용하지 않는 것은 이들이 서로 독립이 아니기 때문이다.

$H_0$를 검정하는 Shukla 통계량은 다음과 같다.

$$Q = a\ln(a^{-1}\sum_{i=1}^{a} W_i) - \sum_{i=1}^{a}\ln(W_i).$$

귀무가설하에서 $(a-1)(b-1)Q/d$는 $\chi^2(a-1)$을 따른다. 단,

$$d = a-1+\frac{a^2-1}{3a(b-1)} - \frac{2(a^4-1)}{15a^3(b-1)^3} - \frac{b-1}{(a-1)^2}\left\{\binom{a}{2}/(1+a(b-1)/2) + \frac{4}{a-1}\binom{a}{3}\left[(1+a(b-1))^{-1} - (2+a(b-1))^{-1}\right]\right\}.$$

모의실험에서 고려한 디자인은 각각 5×5, 7×7 실험계획 모형이며, 오차항의 분포는 표준정규분포인데 FORTRAN IMSL 패키지에서 난수를 얻어 각 경우에 10,000회 반복 실험하였다. 처리효과와 블록효과는 등간격이며($\alpha_i = i-1,\ \beta_j = j-1,\ i=1,\dots,(5\text{ or }7);\ j=1,\dots,(5\text{ or }7)$), 여러 교호작용의 조합에 대한 경험적 검정력이 <표 6>과 <표 7>에 주어져 있다. 여기서 Shukla 검정법의 유의 수준은 5%(괄호 안은 10%)이고, 참고적으로 식 (3.2)의 초기값을 사용한 척도추정값의 평균(10000회)값 s가 같은 칸에 표시되어 있다.

귀무가설하에서 즉, 교호작용이 없을 때($(\alpha\beta)_{ij}=0,\ \forall i,j$), 제안한 검출법이 경험적 유의수준(empirical significance level)을 0.02~0.035 등으로 5%보다 보수적으로 맞추고 있고, s도 참값과 매우 근사함을 볼 수 있다. 5×5 실험계획의 경우, 교호작용의 수가 적으면 그 크기가 $3\sigma$를 벗어날 때 검출비율, 즉 검출력이 매우 우수하고, 생략된 결과에서도 $2\sigma$ 크기의 작은 교호작용을 찾는 검출력도 0.17 정도로 상당함을 알 수 있었다. <표 6>의 4번째 형태인 (2.1)의 가정을 만족하는 코너(corner) 교호작용에서는 Shukla 검정법이 우수하지만 각 행이나 열에 균등하게 위치한 교호작용은 찾아내지 못하여 등분산 검정형태의 Shukla 검정통계량의 한계를 보여주고 있다. 7×7 계획의 경우에서도 경험적 유의수준을 잘 맞추고 있고, Shukla 검정과 비교하여 그 우위를 더 확연히 보여주고 있다.

교호작용 또는 이상값의 수가 많아지면 식 (3.2)에 의한 척도 추정량이 과대추정되는 경향이 있어 결과적으로 검출력이 낮아진다. 이 경우 식 (3.3)에 의하여 매우 근사한 추정값을 얻을 수 있지만, 이 또한 귀무가설하에서는 심한 과소추정을 하기 때문에 여전히 식 (3.2)에 의한 추정방법을 제안한다. 제안한 통계량 값이 척도 추정값으로 표준화해준 값이므로 좀더 우수한 척도모수 추정방법이 유도된다면 검출력을 더 개선할 여지가 있다. 기타 4×5, 6×7의 경우에도 모의검정력실험을 행하였는데, 비슷한 검출력 행태를 보여서 이의 제시는 생략하였다.

<표 6> 5×5 실험계획에서 $L_1$-추정법의 검출력과 Shukla 방법의 검정력 비교
(반복횟수=10,000)

| 교호작용 형태 | $L_1$의 교호작용 검출횟수 | | | | | 교호작용 형태 | $L_1$의 교호작용 검출횟수 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Shukla 검정력(×10000) $\alpha$=5%(10%) | | | | | | Shukla 검정력(×10000) $\alpha$=5%(10%) | | | | |
| 0 0 0 0 0 | 311 | 276 | 265 | 284 | 289 | 3 0 0 0 0 | 3760 | 257 | 248 | 286 | 245 |
| 0 0 0 0 0 | 287 | 271 | 274 | 296 | 301 | 0 0 0 0 0 | 307 | 207 | 238 | 209 | 217 |
| 0 0 0 0 0 | 292 | 285 | 303 | 271 | 294 | 0 0 0 0 0 | 294 | 212 | 230 | 211 | 210 |
| 0 0 0 0 0 | 282 | 275 | 336 | 282 | 309 | 0 0 0 0 0 | 259 | 206 | 210 | 230 | 219 |
| 0 0 0 0 0 | 245 | 273 | 299 | 290 | 283 | 0 0 0 0 0 | 251 | 223 | 221 | 222 | 248 |
| | 486 (997) | | | | s= .966 | | 948 (1708) | | | | s= 1.053 |
| 5 0 0 0 0 | 8099 | 283 | 257 | 262 | 297 | 3 0 0 0 -3 | 3558 | 101 | 91 | 59 | 3871 |
| 0 0 0 0 0 | 288 | 197 | 233 | 189 | 202 | 0 0 0 0 0 | 89 | 70 | 81 | 79 | 76 |
| 0 0 0 0 0 | 263 | 198 | 197 | 205 | 208 | 0 0 0 0 0 | 75 | 80 | 86 | 78 | 77 |
| 0 0 0 0 0 | 246 | 230 | 219 | 225 | 219 | 0 0 0 0 0 | 70 | 70 | 78 | 68 | 87 |
| 0 0 0 0 0 | 252 | 220 | 189 | 217 | 214 | -3 0 0 0 3 | 3889 | 100 | 80 | 88 | 3837 |
| | 2509 (4053) | | | | s= 1.054 | | 6146 (7541) | | | | s= 1.355 |
| 3 0 0 0 0 | 3431 | 154 | 171 | 182 | 318 | 3 0 0 0 0 | 1937 | 45 | 54 | 51 | 46 |
| 0 0 0 0 0 | 179 | 146 | 161 | 151 | 167 | 0 3 0 0 0 | 52 | 2120 | 43 | 50 | 35 |
| 0 0 0 0 0 | 179 | 133 | 155 | 156 | 174 | 0 0 3 0 0 | 47 | 58 | 2147 | 38 | 38 |
| 0 0 0 0 0 | 175 | 179 | 178 | 141 | 182 | 0 0 0 3 0 | 47 | 40 | 40 | 2173 | 45 |
| 0 0 0 0 3 | 290 | 161 | 150 | 137 | 3752 | 0 0 0 0 3 | 57 | 46 | 45 | 44 | 2273 |
| | 1034 (1883) | | | | s= 1.155 | | 48 (114) | | | | s= 1.593 |

<표 7> 7×7 실험계획에서 $L_1$-추정법의 검출력과 Shukla 방법의 검정력 비교
(반복횟수=10,000)

| 교호작용 형태 | $L_1$의 교호작용 검출횟수 | | | | | | | 교호작용 형태 | $L_1$의 교호작용 검출횟수 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Shukla 검정력(×10000) $\alpha$=5%(10%) | | | | | | | | Shukla 검정력(×10000) $\alpha$=5%(10%) | | | | | | |
| 0 0 0 0 0 0 0 | 230 | 225 | 228 | 224 | 192 | 221 | 221 | 3 0 0 0 0 0 0 | 4864 | 231 | 209 | 218 | 232 | 225 | 217 |
| 0 0 0 0 0 0 0 | 217 | 258 | 253 | 220 | 193 | 234 | 218 | 0 0 0 0 0 0 0 | 229 | 210 | 208 | 184 | 196 | 208 | 192 |
| 0 0 0 0 0 0 0 | 215 | 233 | 229 | 228 | 201 | 231 | 220 | 0 0 0 0 0 0 0 | 218 | 188 | 215 | 177 | 188 | 195 | 195 |
| 0 0 0 0 0 0 0 | 218 | 241 | 235 | 247 | 228 | 250 | 239 | 0 0 0 0 0 0 0 | 212 | 178 | 201 | 212 | 192 | 207 | 207 |
| 0 0 0 0 0 0 0 | 238 | 243 | 214 | 224 | 251 | 252 | 245 | 0 0 0 0 0 0 0 | 233 | 203 | 195 | 213 | 206 | 192 | 199 |
| 0 0 0 0 0 0 0 | 258 | 235 | 229 | 225 | 237 | 233 | 231 | 0 0 0 0 0 0 0 | 219 | 183 | 202 | 228 | 199 | 222 | 168 |
| 0 0 0 0 0 0 0 | 240 | 204 | 232 | 247 | 219 | 223 | 253 | 0 0 0 0 0 0 0 | 212 | 188 | 184 | 202 | 196 | 203 | 210 |
| | 534 (1015) | | | | | | s= .978 | | 1036 (1904) | | | | | | s= 1.012 |
| 3 0 0 0 0 0 3 | 2122 | 192 | 225 | 196 | 158 | 168 | 2168 | 3 0 0 0 0 0 0 | 3473 | 68 | 74 | 70 | 58 | 71 | 62 |
| 0 0 0 0 0 0 0 | 213 | 98 | 118 | 104 | 105 | 113 | 190 | 0 3 0 0 0 0 0 | 66 | 3479 | 65 | 67 | 72 | 78 | 61 |
| 0 0 0 0 0 0 0 | 181 | 108 | 102 | 104 | 114 | 104 | 195 | 0 0 3 0 0 0 0 | 62 | 65 | 3587 | 53 | 73 | 73 | 67 |
| 0 0 0 0 0 0 0 | 193 | 112 | 124 | 114 | 124 | 101 | 191 | 0 0 0 3 0 0 0 | 66 | 63 | 63 | 3584 | 67 | 69 | 58 |
| 0 0 0 0 0 0 0 | 183 | 109 | 112 | 112 | 118 | 126 | 156 | 0 0 0 0 3 0 0 | 63 | 63 | 67 | 62 | 3595 | 67 | 62 |
| 0 0 0 0 0 0 0 | 203 | 116 | 92 | 106 | 124 | 117 | 188 | 0 0 0 0 0 3 0 | 65 | 53 | 59 | 66 | 56 | 3540 | 71 |
| 3 0 0 0 0 0 3 | 2150 | 180 | 167 | 171 | 162 | 170 | 2205 | 0 0 0 0 0 0 3 | 58 | 65 | 65 | 62 | 54 | 63 | 3539 |
| | 870 (1680) | | | | | | s= 1.155 | | 57 (154) | | | | | | s= 1.306 |

## 4. 결 론

본 논문에서 제안한 $L_1$-추정법에 의한 교호작용 검출법은 반복이 없는 이원배치 실험계획에서 교호작용이 개수가 적고 각 블록과 처리 수준에 고루 분포되었을 때, 교호작용을 검출하는데 매우 우수한 방법이다.

M-추정량 형태 중에서 특히 $L_1$-추정법을 사용한 이유는 최적의 점근 붕괴점을 갖는 M-범함수(M-functional)는 유도되어 있지 않지만 오차항이 없는 완전적합모형에서 $L_1$-추정법이 최적임이 이론적으로 증명되어 있기 때문이다. 본 논문에 그 결과는 수록하지 않았지만 실제로 오차가 없는 경우에 $L_1$-추정법에 의하면 이론적인 붕괴점 비율을 넘는 개수의 이상값도 검출할 수 있었다. 또한 교호작용에 식 (2.1)의 제약이 있고 각 행이나 열에서 이상값(교호작용)의 부호의 개수가 일치할 때, $L_1$-추정법은 80% 정도의 이상값도 검출하는 매우 특이한 행태를 보여주었다. 물론 완전적합모형에 오차가 더해졌을 때 그러한 현상은 사라지지만 실험계획 모형에서 $L_1$-추정법 특성에 관한 이론적 연구가 계속되어야 함을 시사하고 있다.

그리고 본문에서 언급한 바와 같이 식 (3.2)보다 우수한 로버스트 척도모수 초기값을 유도할 수 있다면 검정력이 개선될 수 있는 가능성이 있어 이에 대한 연구도 계속 될 수 있다.

교호작용 검출에 M-추정량 형태가 아닌 Roussecuw(1984)가 제안한 LMS(least median of squares) 방법 등을 사용할 수 있는데, 이는 단순 선형회귀모형에서는 우수한 붕괴점을 갖지만 본 논문에서 제안하는 다중회귀모형의 형태에서는 붕괴점이 급속히 감소하기 때문에 여기서 고려하지 않았다. 그러나 블록 수와 처리 수가 매우 많을 때, 이론적으로는 붕괴점이 본 논문에서 제안한 방법보다 개선될 수 있기 때문에 M-추정량 형태가 아닌 기타 로버스트 추정법에 의한 교호작용 검출법의 유도도 가능하다고 할 수 있다.

본 연구는 교호작용 검정법보다는 검출법의 제안에 제한되어 있다. 각 셀별로 교호작용을 검출하기 때문에 이를 유의성 검정으로 확장하기에는 총체적 유의수준(overall significance level)을 제어하지 못하는 약점을 갖고 있는 것이 본 연구의 한계이다.

## 참 고 문 헌

[1] 이기훈 (1997). 반복이 없는 이원배치에서 분포의 동일성 검정에 대한 비모수적 검정법, 「한국통계학회논문집」, 4권 3호, 765-774

[2] Daniel, C.(1960). Locating Outliers in Factorial Designs, *Technometrics,* 2, 149-156.

[3] Ellis, S. P. and Morgenthaler, S.(1992). Leverage and Breakdown in $L_1$ Regression, *Journal of the American Statistical Associations*, 87, 143-148.

[4] Grubbs, F. E.(1948). On Estimation of Precision of Measuring Instruments and Product Variability, *Journal of the American Statistical Associations*, 43, 243-264.

[5] Johnson, D. E. & Graybill, F. A.(1972). An Analysis of Two-way Model with Interaction and No Replications, *Journal of the American Statistical Associations*, 67, 878-888.

[6] Jureckova, J. and Sen, P. K.(1996). *Robust Statistical Procedures*, John Wiley & Sons, New York.

[7] Kim, B.(1995). On the Robustness of $L_1$-estimator in Linear Regression Models, *The*

*Korean Communications in Statistics*, Vol. 2, No. 2, 277-287.

[8] Mandel, J.(1961). Non-additivity in Two-way Analysis of Variance, *Journal of the American Statistical Associations*, 56, 878-888.

[9] Milliken, G. A. and Johnson, D. E.(1989). *Analysis of Messy Data. Vol. 2 : Nonreplicated Experiments*, New York; Van Nostrand Reinhold.

[10] Milliken, G. A. and Rasmuson, D.(1977). A Heuristic Technique for Testing for the Presence of Interaction in Nonreplicated Factorial Experiments, *Australian Journal of Statististics*, 19, 32-38.

[11] Piepho, H.(1994). On Tests for Interaction in a Nonreplicated Two-way Layout, *Australian Journal of Statististics*, 26, 363-369.

[12] Rousseeuw, P. J.(1984). Least Median of Squares Regression, *Journal of the American Statistical Associations*, 79, 871-880.

[13] Rousseeuw, P. J. and Leroy, A. M.(1987). *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.

[14] Shukla, G. K.(1982). Testing the Heterogeneity of Variances in a Two-way Classification, *Biometrika*, 69, 411-416.

[15] Terbeck, W. and Davies, P. L.(1998). Interaction and Outliers in the Two-way Analysis of Variance, *The Annals of Statistics*, 26, 1279-1305.

[16] Tukey, J. W.(1949). One Degrees of Freedom for Non-additivity, *Biometrics*, 5, 232-242.