

## An Identification of Outlying Cells in Contingency Tables via Correspondence Analysis Map

Chong Sun Hong<sup>1)</sup> and Jong Cheol Lee<sup>2)</sup>

### Abstract

When an appropriate model is fitted to explain a certain categorical data, outlying cell detection plays very important role to reduce the lack of fit. There exist many statistical methods to identify outlying cells in contingency table. In this paper, correspondence analysis is applied to identify one or two outlying cells. When corresponding relationships between categories of the row and columns are explored, we find that outlying cells could be identified via the correspondence analysis map.

*Keywords* : contingency table, correspondence analysis, cluster, exploratory data analysis, outlying cell.

### 1. 서 론

1930년대부터 연구가 시작되어 1980년에 많은 발전을 이룬 대응분석(correspondence analysis) 방법은 범주형 자료분석에 이용되어왔다. 대응분석은 분할표 자료의 행과 열을 저차원 공간상의 점들로 동시에 나타내어 행과 열의 범주간의 관계를 탐구하는 탐색적 자료분석(exploratory data analysis) 방법으로 행과 열 범주가 놓인 위치에 따라 하나의 군집(cluster)으로 고려하여 그 대응 관계를 나타낸다(Greenacre 1984, pp. 3-7; Greenacre와 Blasius 1994, pp. 3-8; 최용석 1993, 1장). 예를 들어 허명희(1989)는 13대 대통령 선거에 대해 7명의 후보와 15개 선거구역을 각각 행과 열로 구분하여 그 득표수를 대응분석한 결과인 대응분석그림(correspondence analysis map)을 제시하였다. 그 결과 D후보와 3개 지역, Y후보와 4개 지역이 한 군집을 이룬다는 것을 제시하였다. 이를 달리 표현하면 후보와 선거구역별 득표수가 다른 후보들과 달리 몇몇 지역에서 높은 득표수를 얻었다는 것을 설명해 주고 있다.

대응분석에 의해 행과 열의 일부 범주들이 하나의 군집을 이룬다는 것은 분할표 전체와 다른 특성을 갖는 군집으로 이해할 수 있다. 두 변수가 서로 독립인 분할표에서 하나의 칸이 매우 큰 경우는 이 칸에 해당하는 행과 열 범주가 서로 대응관계를 갖는 것으로 설명할 수 있다. 따라서

---

1) Professor, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea,  
E-mail : cshong@skku.ac.kr.

2) Lecturer, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea.

대응분석을 이용하여 분할표 자료에 대한 이상칸의 탐색을 할 수 있다.

분할표 자료에 대해 적합결여의 한 원인으로 고려할 수 있는 이상칸을 식별하기 위한 논의가 이루어져왔다. Simonoff(1988)가 제안한 후진단계방법(backwards-stepping method), Fuchs와 Kenett(1980)이 제안한 전진단계방법(forwards-stepping method) 등이 있다(홍종선과 최현집 1999, 5장). 또한 이종철(1999, pp. 31-40)은 후진단계방법을 응용하여 MOCI 방법(multiple outlying cells identification method)을 제안하였다. 이들 방법은 분할표 자료에 대해 이상칸을 식별한 결과를 수리적인 측면에서만 분석하고 있는데, 본 논문에서는 대응분석을 이용하여 시각적인 방법 즉, 통계적 도형에 구현할 수 있는 차트를 이용한 탐색적 자료분석으로 이상칸을 식별하는 연구 방법을 제안하고자 한다.

본 논문의 2절에서는 이상칸이 존재하는 이차원 분할표 자료를 대응분석 하였을 때 대응분석그림을 통하여 발견할 수 있는 이상칸의 존재성과 그 관계성을 설명하였다. 따라서 분할표 자료에 대해 대응분석 기법을 이용하여 이상칸을 탐색하기 위하여 이상칸 존재의 형태를 체계적으로 설정하여 살펴보고자 한다. 3절에서는 예를 들어  $4 \times 4$  분할표 자료에서 하나의 이상칸이 존재하는 경우와  $6 \times 5$  분할표 자료에서 두 개의 이상칸이 존재하는 경우로 구분하여 고려하였다. 하나의 이상칸이 존재하는 경우에는 관찰값이 기대값보다 큰 경우와 작은 경우로 나누어 살펴보았다. 두 개의 이상칸이 존재하는 경우에는 두 칸의 관찰값이 기대값보다 모두 큰 경우와 작은 경우 그리고 하나는 크고 다른 하나는 작은 경우인 세 종류로 나누어 설명하였다. 이러한 경우의 이상칸들이 동일한 행(또는 열)에 존재하는 경우와 서로 다른 행과 열에 존재하는 경우로 나누어 관찰하였다. 마지막으로 5절에는 대응분석그림을 통해 이상칸을 식별하는 한계와 어려움에 대하여 토론하고, 다차원 분할표인 경우를 포함한 다양한 로그선형모형에 적합할 때의 이상칸의 식별등을 향후 연구과제로 제안하였다.

## 2. 대응분석과 이상칸의 관계

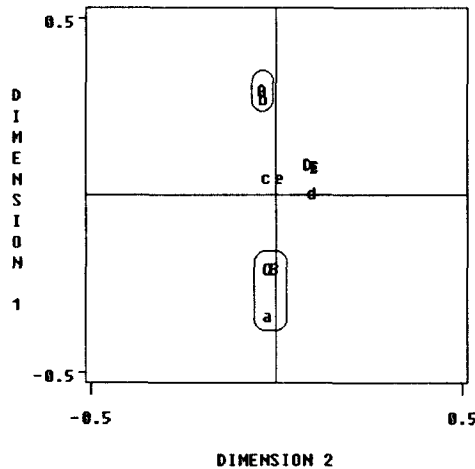
Simonoff(1988)가 이상칸의 식별을 위해 작성한 <표 1>의 2차원 분할표를 고려하자. 이 분할표에 대해 후진선택방법과 MOCI 방법을 적용하여 두 변수가 독립인 모형에 대한 이상칸을 식별하면 (a,B), (a,C), (b,A)칸이 이상칸으로 판별된다. <그림 1>은 <표 1>의 자료를 SAS의 대응분석에 의해 얻어진 대응분석그림이다. 본 연구에서 수행한 대응분석은 SAS/PROC CORRESP를 사용하였고 SAS/GPLOT을 이용하여 대응분석그림을 얻었다. 그리고 본 연구에서의 대응분석은 Greenacre(1984, pp. 60-66)의 대칭적 방법을 이용하였다.

행과 열의 각 범주의 위치를 살펴보면 행변수의 범주 a와 열변수의 범주 B, C가, 그리고 행변수의 범주 b와 열변수의 범주 A가 원점으로부터 멀리 떨어지고 나머지 범주들은 원점에 가까이 놓은 것을 볼 수 있다. 이러한 대응분석으로부터 얻어진 기하학적인 결과는 같은 거리에 위치한 행과 열 범주간에 대응관계가 있음을 보여준다. 이를 범주들간의 관계로 살펴보면 행변수의 a, b는 c, d, e와 다른 형태를 보이고, 열변수의 A, B, C는 D, E와 다른 형태를 보이고 있다. 이들 행과 열 범주를 하나씩 묶으면 (a,B), (a,C) 그리고 (b,A)로 표현할 수 있다. 이들 묶음은 <표 1>의 분할표에 대해 이상칸으로 식별된 것과 무관하지 않다. 따라서 이들 범주의 조합으로 얻어진 칸들을 이상칸으로 고려할 수 있다. 우선 (a,B)와 (a,C)칸값을 살펴보면 각각 41, 41로 동일 행과 열의 값들과 비교할 때 큰 값인 것을 알 수 있다. (b,A)칸값은 39로 동일 행과 열의 값들보다 큰 값임

을 알 수 있다. 따라서 이들 칸이 주어진 분할표 자료에 대해 행과 열 범주가 서로 대응관계를 가지고 있고 이를 이상칸으로 고려할 수 있다. 특히 <표 1>의 분할표 자료는 Simonoff(1988)가 이상칸의 식별하는 과정에서 가장효과(masking effect)를 우려하면서 생성한 자료인데, 대응분석그림을 통한 분석에서도 가장효과가 존재하는 자료의 이상칸을 탐색적으로 잘 식별하고 있다는 것을 파악할 수 있다.

<표 1> 이차원 분할표

열변수 행변수	A	B	C	D	E	주변합
a	18	41	41	20	21	141
b	39	20	20	22	22	123
c	24	20	20	16	18	98
d	20	20	19	19	19	97
e	23	19	20	17	20	99
주변합	124	120	120	94	100	558



<그림 1> 대응분석그림

분할표의 전체 칸값에 대해 일부 칸값이 매우 작거나 큰 경우 이를 대응분석을 통해 대응관계를 설명할 수 있고 이상칸으로 고려할 수도 있다. 이에 본 논문에서는 이차원 분할표 자료에 대해 대응분석기법을 이용하여 이상칸을 탐색하여 그 형태를 체계적으로 살펴본다.

### 3. 사례 연구

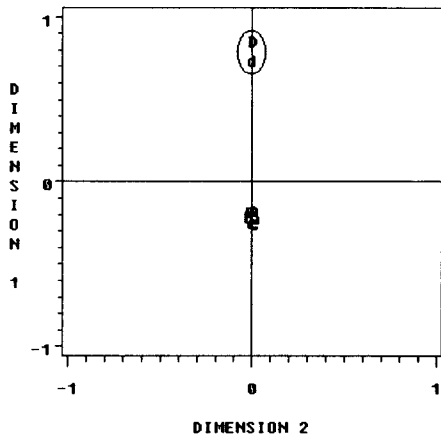
이차원 분할표에 대해 대응분석방법을 이용하여 이상칸을 식별하기 위해 우선 이상칸이 존재하는 분할표를 생성하기로 한다. 이상칸이 존재하는 분할표를 생성하기 위해 Simonoff(1988)가 제안한 칸 확률을 이용하여 다항확률표본추출을 한다. 생성된 분할표에 대해 MOCI방법으로 이상칸을 식별하고 대응분석으로 얻어진 결과와 비교하기로 한다.

### 3.1 하나의 이상칸이 존재하는 경우

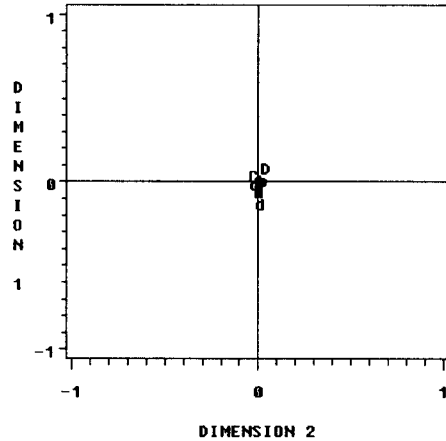
우선 이상칸이 하나인 <표 2>와 같은 이차원 분할표를 고려하자. 이 분할표에 대해 MOCI방법을 적용하여 이상칸을 식별하면, 행변수와 2가 독립인 모형에 식별된 이상칸은 (d,D)칸으로 이칸의 기대값은 4이다.

<표 2> (d,D)가 이상칸인 분할표

열변수 행변수	A	B	C	D	주변합
a	8	12	8	3	31
b	5	8	5	2	20
c	9	14	9	3	35
d	4	6	3	11	24
주변합	26	40	25	19	110



<그림 2(a)> 대응분석그림



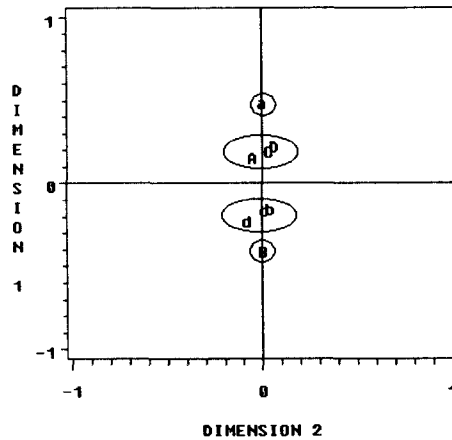
<그림 2(b)> 이상칸 수정한 분할표의 대응분석그림

<표 2>의 자료를 대응분석을 실시하여 얻어진 대응분석그림은 <그림 2(a)>에 표현하였다. 오른쪽 그림의 <그림 2(b)>는 (d,D)칸값을 기대값으로 수정하여 얻어진 그림이다. <그림 2(a)>를 살펴보면 행변수의 d와 열변수의 D가 하나의 군집을 이루는 것을 보여준다. 이는 두 범주가 교차해서 이루어지는 (d,D)칸값 11이 독립인 모형에 대해 기대값 4보다 큰 값을 가지고 있고 제1축에 멀리 위치한 하나의 군집을 이룬다. 따라서 하나의 칸값이 독립성 모형에 대해 기대값보다 큰 값을 가지는 경우 제1축에 대해 하나의 군집을 형성하여 이상칸의 위치를 보여준다. 제2축에 대해서는 변화를 보이지 않지만 제1축에만 많은 영향을 받는 것을 볼 수 있다. 이는 고유값과 고유벡터에 의해 두 개의 축으로 나타낸 것이므로 제1축이 가장 큰 고유값이 할당되기 때문에 나타난다. 참고로 제1축의 고유값은 0.40이며 전체고유값 합계에서 차지하는 비율을 나타내는 설명력(goodness of fit)은 99.93%이다. 반면에 <그림 2(b)>에서 제1축의 설명력은 86.76%이다.

<표 3> (a,B)가 이상칸인 분할표

열변수 행변수	A	B	C	D	주변합
a	8 0.38 / 0.31	2 0.10 / 0.07	8 0.38 / 0.32	3 0.14 / 0.33	21
b	5 0.25 / 0.19	8 0.40 / 0.27	5 0.25 / 0.20	2 0.10 / 0.22	20
c	9 0.26 / 0.35	14 0.40 / 0.47	9 0.26 / 0.36	3 0.09 / 0.33	35
d	4 0.29 / 0.15	6 0.43 / 0.20	3 0.21 / 0.12	1 0.07 / 0.11	14
주변합	26	30	25	9	90

※ 두번째 줄은  $(p_{ij}/p_{i+} / p_{ij}/p_{+j})$ 을 나타낸다.



<그림 3> 관찰값보다 칸 기대값이 큰 경우

<표 3>에 나타난 (a,B)칸의 행과 열에 대한 확률을 살펴보면 다른 행과 열과 다른 값을 갖는 것을 볼 수 있다. <그림 3>은 (a,B)칸이 이상칸으로 식별되는 <표 3>과 같은 분할표에 대한 대응분석그림이다. 여기에서 행변수의 범주 a와 열변수의 범주 B가 다른 범주에 비해 멀리 위치해 있다. (a,B)칸의 기대값 7에 비해 작은 칸값이므로 (a,B)칸의 행과 열 범주가 다른 범주와 구별되게 놓여있다. 즉 <그림 3>과 같이 나타난 대응분석그림으로 하나의 이상칸을 식별할 수 있다.

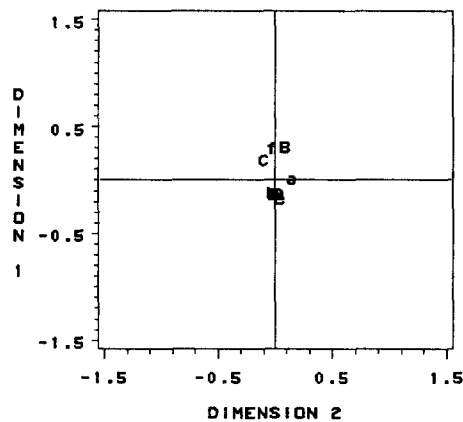
### 3.2 동일 행 또는 열에 두 개의 이상칸이 존재하는 경우

(1) 두 칸의 관찰값이 기대값보다 큰 경우

하나의 행에 두 개의 이상칸이 존재하는 <표 4>와 같은 분할표를 고려하자. MOCI방법에 의해 식별된 이상칸은 (f,B), (f,C)칸이다. 이 두 개의 칸에 대한 추정기대값은 각각 30, 39이다. 이 분할표에 대한 대응분석그림인 <그림 4>를 살펴보면, 행변수 f와 열변수 B, C가 나머지 범주를 나타내는 점들과 떨어진 군집을 이루고 있음을 발견할 수 있다. 따라서 식별방법에 의하여 이상칸으로 간주된 (f,B), (f,C)칸과 동일한 결론을 유도하였다.

<표 4> 두 개의 이상칸이 존재하는 분할표 1

열변수 \ 행변수	A	B	C	D	E	주변합
a	22	19	14	29	26	110
b	58	31	46	91	74	300
c	28	14	21	42	35	140
d	52	25	40	78	65	260
e	7	3	4	9	8	31
f	34	47	55	51	43	230
주변합	201	139	180	300	251	1071



<그림 4> 두 개의 이상칸이 존재하는 대응분석그림 1

(2) 두 칸의 관찰값이 기대값보다 작은 경우

<표 5>에서 (c,C)칸과 (c,E)칸의 기대값은 각각 12와 21이다. <표 5>와 같이 (c,C)칸과 (c,E)칸의 관찰값보다 원래의 기대값이 큰 경우 대응분석을 수행하여 얻은 대응분석그림은 <그림 5>와 같다. 소문자 c와 대문자 C와 E도 다른 문자들에 비하여 멀리 떨어져 위치하고 있음을 파악할 수 있다. 이를 통하여 행범주와 열범주의 결합을 고려하면, 이상칸이 (c,C)칸과 (c,E)칸이라고 식별할 수 있다.

(3) 한 칸의 관찰값은 기대값보다 크고 다른 하나는 작은 경우

동일 열(행)에 두 칸의 기대값이 하나는 크고 작은 경우를 고려하자. 여기에서 (b,D)와 (e,D)칸의 기대값은 각각 66과 17이다. <표 6>과 같은 분할표에 대하여 대응분석한 <그림 6>을 살펴보자. 원점 주변에 물려있는 점들은 (A,B,C,E)와 (a,c,d,f)이다. 이들 행범주와 열범주들은 동일한 군집을 이루고 나머지 D와 b, e는 원점에서 멀리 위치한 것을 발견할 수 있다. 따라서 이들 행과 열의 범주에 의한 조합인 (b,D)와 (e,D)칸을 이상칸으로 식별한다.

이 절에서는 동일한 행 또는 열에 두 개의 이상칸이 존재하는 경우에 대하여 관찰값이 기대값보다 큰 경우와 작은 경우로 세분하여 고려하여 살펴보았다. 여기에서 우리는 이상칸에 대응하는

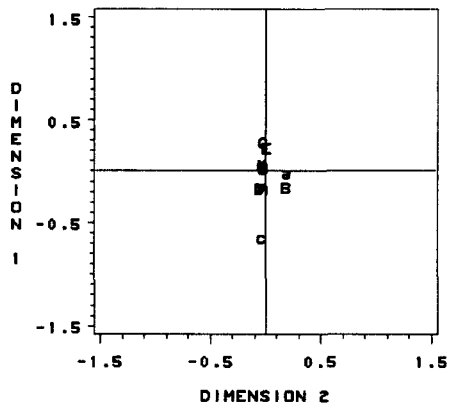
행범주와 열범주를 나타내는 문자들이 원점을 중심으로 어떤 방향으로 멀리 떨어져 있는지를 살펴보고 이에 대해서는 3.3절의 마지막 문단에서 토론하고자 한다.

<표 5> 두 개의 이상칸이 존재하는 분할표 2

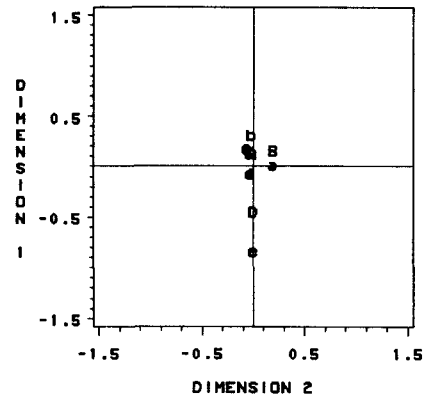
열범수 \ 행범수	A	B	C	D	E	주변합
a	22	19	14	29	26	110
b	58	31	46	91	74	300
c	28	14	1	42	5	90
d	52	25	40	78	65	260
e	7	3	4	9	8	31
f	34	17	25	51	43	170
주변합	201	109	130	300	221	961

<표 6> 두 개의 이상칸이 존재하는 분할표 3

열범수 \ 행범수	A	B	C	D	E	주변합
a	22	19	14	29	26	110
b	58	31	46	31	74	240
c	28	14	21	42	35	140
d	52	25	40	78	65	260
e	7	3	4	39	8	61
f	34	17	25	51	43	170
주변합	201	109	150	270	251	981



<그림 5> 대응분석그림 2



<그림 6> 대응분석그림 3

### 3.3 서로 다른 행과 열에 두 개의 이상칸이 존재하는 경우

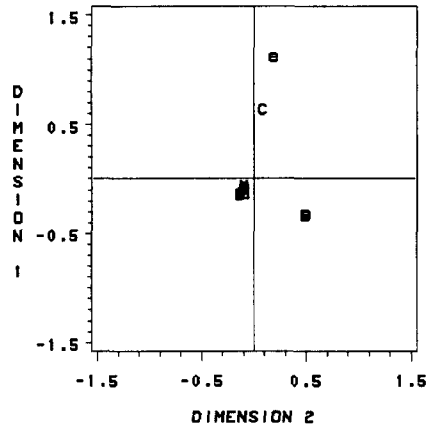
(1) 칸값이 기대값보다 큰 경우

<표 7>에서 서로 다른 행과 열에 존재하는 (a,B)와 (e,C)칸의 기대값은 각각 16과 13이며, 두

개의 칸값이 기대값보다 모두 큰 경우의 자료이다. 이 자료를 대응분석한 그림을 <그림 7>에 제시하였다. <그림 7>을 살펴보면 (a,B)와 (e,C)를 제외한 나머지 문자들이 가까이 위치하는데 이는 두 칸이 서로 다른 군집을 이루는 이상칸임을 보여준다(<그림 7>의 남동쪽에 a와 B점이 중복되어 나타나 있다).

<표 7> 서로 다른 행과 열에 존재하는 분할표 4

열변수 \ 행변수	A	B	C	D	E	주변합
a	22	41	14	29	26	132
b	58	31	46	91	74	300
c	28	14	21	42	35	140
d	52	25	40	78	65	260
e	7	3	44	9	8	71
f	34	17	25	51	43	170
주변합	201	131	190	300	251	1073



<그림 7> 대응분석그림 4

(2) 칸값이 기대값보다 작은 경우

<표 8>에서 서로 다른 행과 열에 존재하는 (b,A)와 (d,D)의 기대값은 각각 45, 57인데, 표에 나열되어있는 칸값은 기대값보다 모두 작은 경우이며, 대응분석한 그림은 <그림 8>과 같이 나타난다. (a,c,e,f)와 (B,C,E)가 하나의 군집을 이루고 b, d, A, D가 이 군집으로 떨어져 나타나있다. 따라서 이들 문자의 조합으로 이상칸을 식별할 수 있게 된다. 가능한 조합은 (b,A)와 (d,D) 또는 (b,D)와 (d,A)가 된다. 이와 같이 칸값이 기대값보다 작은 경우에는 두 가지 조합을 고려할 수 있는데, 이런 경우에는 해당되는 칸값을 살펴보고 이상칸을 식별해야 한다. 즉 칸 기대값을 살펴보고 이상칸의 식별을 판단하여야 하겠다.

(3) 한 칸이 관찰값은 기대값보다 크고 다른 하나는 작은 경우

<표 9>와 같이 두 개의 이상칸이 서로 다른 행과 열에 존재하는 경우 대응분석을 통하여 이상칸을 식별하도록 하자. (a,A)칸과 (d,D)칸의 기대값은 각각 33과 53이다 (a,A)칸은 기대값보다 큰



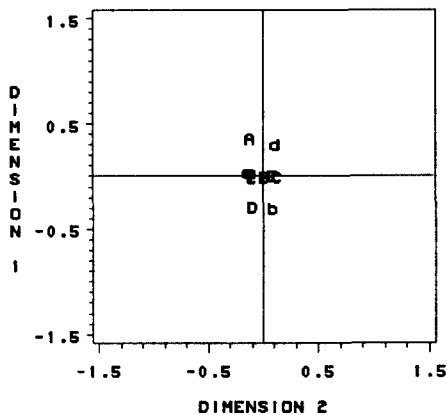
칸값을 가졌고 (d,D)칸은 작은 경우이다. 대응분석그림인 <그림 9>를 살펴보면 (b,c,e,f) 그리고 (B,C,E)가 하나의 군집을 이루고 나머지 문자들이 동떨어져 나타나있다. 즉 행범주 a와 d 그리고 열범주 A와 D가 군집과 차이가 나타난다. 그러므로 이들 문자의 조합인 (a,A)와 (d,D) 또는 (a,D)와 (d,A)가 이상칸으로 고려할 수 있는데, 이들의 칸 기대값을 살펴봄으로써 (a,A)와 (d,D)를 이상칸으로 식별할 수 있다.

<표 8> 서로 다른 행과 열에 존재하는 분할표 5

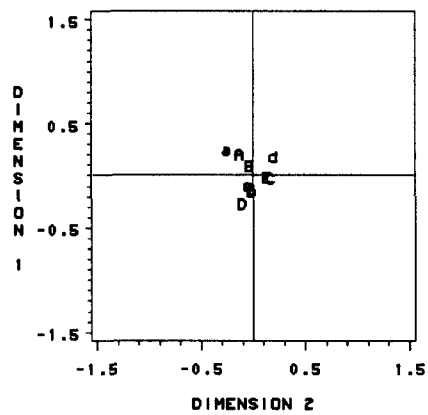
열변수 \ 행변수	A	B	C	D	E	주변합
a	22	19	14	29	26	110
b	18	31	46	91	74	260
c	28	14	21	42	35	140
d	52	25	40	28	65	210
e	7	3	4	9	8	31
f	34	17	25	51	43	170
주변합	161	109	150	250	251	921

<표 9> 서로 다른 행과 열에 존재하는 분할표 6

열변수 \ 행변수	A	B	C	D	E	주변합
a	52	19	14	29	26	140
b	58	31	46	91	74	300
c	28	14	21	42	35	140
d	52	25	40	28	65	210
e	7	3	4	9	8	31
f	34	17	25	51	43	170
주변합	231	109	150	250	251	991



<그림 8> 대응분석그림 5



<그림 9> 대응분석그림 6

이상의 8가지 사례를 살펴보면 2차원 분할표에서 이상칸을 식별하기 위해 대응분석을 이용할 수 있음을 알 수 있다. 이상칸의 기대값이 칸값보다 크거나 작은 경우에 따라 범주들의 위치가 달

리 표현되어, 하나 또는 두 개의 군집으로 표현되지 않고 멀리 산재되어 나타나므로 이들 범주의 조합으로 이상칸을 식별할 수 있게 된다.

두 개의 이상칸이 존재하는 3.2절과 3.3절에 나열되어 있는 대응분석그림들을 살펴보면, <그림 4>부터 <그림 6>은 <그림 7>부터 <그림 9>보다 제2축에 대한 산포도가 크다는 것을 발견하였다. 이를 통하여 두 개의 이상칸이 서로 다른 행과 열에 존재할 때는 제1축의 설명력이 같은 행(또는 열)에 존재할 때보다 적어지고 제2축의 설명력이 증가한다는 사실을 유도할 수 있다.

3.2(1)절과 3.3(1)절은 두 칸값이 기대값보다 큰 경우인데, 여기에서는 각 칸의 행범주와 열범주를 나타내는 점들이 원점을 중심으로 동일한 방향으로 퍼져있다(<그림 4>와 <그림 7> 참조). 반면에 두 칸값이 기대값보다 모두 작은 경우인 3.2(2)절과 3.3(2)절에서는 각 칸의 행범주와 열범주를 나타내는 점들이 원점을 중심으로 서로 다른 방향으로 퍼져있다(<그림 5>와 <그림 8> 참조). 한 칸의 관찰값은 기대값보다 크고 다른 하나는 작은 경우인 3.2(3)절과 3.3(3)절에서는 대응되는 행범주와 열범주를 나타내는 점들이 한 쌍은 원점을 중심으로 같은 방향으로 그리고 다른 하나는 다른 방향으로 산재하고 있음을 발견할 수 있다(<그림 6>과 <그림 9> 참조). 따라서 대응분석그림에서 군집에서부터 멀리 떨어진 점들에 해당되는 행과 열범주의 방향을 고려하면서 조합하면, 이상칸을 탐색적으로 식별하는데 도움이 된다. 그러나 여러 개의 이상칸이 존재하는 경우에 대하여는 위에서 언급한 이상칸 식별 방법과 일치된 방향성에 대한 결론을 유도하는 데는 한계가 있음을 발견하였다.

#### 4. 결 론

분할표 자료에 대해 모형을 설정하여 변수들간의 관계를 설명하는 것이 중요한 부분이나 적합결여의 한 원인인 이상칸을 식별하여 그 관계를 정확히 설명할 필요가 있다. 이러한 이상칸을 식별하기 위한 많은 방법들이 제안되었는데, 본 논문에서는 이상칸의 식별을 위한 탐색적인 방법으로 대응분석을 적용하였다. 2차원 평면에 표현되는 대응분석의 결과에서 행 또는 열의 일부 범주들의 위치가 대부분의 범주에 대해 달리 나타나는 것을 이용하면, 이상칸을 식별할 수 있음을 발견하였다.

칸 기대값에 비하여 칸값의 크기(또는 비율)의 증감과 이상칸의 수에 따라 이상칸을 탐색하는데 어려움이 존재한다. 본 논문에서는 분할표 자료에 대해 소수의 칸이 이상칸인 경우 통계적 도형을 이용한 탐색적인 방법으로 고려할 수 있는 방법을 제안하였다. 이상칸을 정확히 식별하기 위해서는 기존의 방법인 Simonoff의 후진단계방법, Fuchs와 Kenett의 전진단계방법, 그리고 이중철의 MOCI방법을 이용하여 구체적으로 식별이 가능하지만 하나 또는 두 개의 이상칸이 존재하는 경우에는 대응분석 방법으로 탐색적인 이상칸의 식별이 가능해짐을 보인 것이 본 연구의 특징이다.

여러 개의 이상칸이 존재할 때 기존의 식별방법들에는 가장효과(masking effect)와 편승효과(swamping effect)가 존재하였는데 본 논문에서 제안한 탐색적인 방법과의 연계성에 대하여도 향후 연구가 진행되어야 하겠다. 뿐만 아니라 다차원 분할표에 대해서도 대응분석을 통하여 이상칸의 식별이 가능할 것이다. 또한 MOCI 방법과 마찬가지로 분할표의 다양한 로그선형모형(log-linear model)과 순위변수(ordinal variable)를 포함하는 모형인 선형대선형 모형(linear by linear model)과 행 또는 열 효과 모형(row or column effect model)에 대해 이상칸을 탐색할 수 있는 탐색적인 방법으로 대응분석을 이용하여 이상칸을 식별할 수 있음을 향후 연구 과제로 남겨 놓고자 한다.

## References

- [1] 이종철(1999). 다차원 분할표에서의 다중이상칸 식별, 「박사학위논문」, 성균관대학교 중앙도서관, 서울.
- [2] 최용석(1993). 「SAS 대응분석」, 자유아카데미, 서울.
- [3] 허명희(1989). 대응분석의 이론과 응용, 「응용통계」, 4권 2호, pp. 23-31.
- [4] 홍종선, 최현집(1999). 「로그선형모형을 이용한 범주형 자료분석」, 자유아카데미, 서울.
- [5] Fuchs, C. and Kenett, R. (1980). A test for outlying cells in the multinomial distribution and two-contingency tables, *Journal of the American Statistical Association*, Vol. 75, pp. 395-398.
- [6] Greenacre, Michael J. (1984). *Theory and Applications of Correspondence Analysis*, Academic Press, New York.
- [7] Greenacre, Michael J. and Blasius, Joerg (1994). *Correspondence Analysis in the Social Science*, Academy Press, New York.
- [8] SAS Institute (1996). *SAS/STAT Change and Enhancement, User's Guide, Version 6.12*, Cary, N.C. : SAS Institute Inc.
- [9] SAS Institute (1996). *SAS/GRAPH Change and Enhancement, User's Guide, Release 6.07 Edition*, Cary, N.C. : SAS Institute Inc.
- [10] Simonoff, J. S. (1988). Detecting outlying cells in two-way contingency table via backwards-stepping, *Technometrics*, Vol. 30, pp 339-345.