

다차원 범주형 자료에 대한 링차트 II : 조건부 링차트를 이용한 자료 분석

홍종선¹⁾ 이종철²⁾

요약

다차원 범주형 자료를 표준화된 링차트로 구현하면, 자료에 적합한 모형이 갖는 일차교호작용의 존재 유무를 파악할 수 있으며 또한 표준화된 조건부 링차트를 통하여 동시에 두개 이상의 일차교호작용의 존재유무를 발견할 수 있는데 3차원 자료에서는 최대 두개의 일차교호작용항을, 그리고 4차원 자료에서는 최대 4개의 일차교호작용항의 존재를 파악할 수 있다.

주요용어: 교차적비, 교호작용, 독립모형, 로그선형모형, 연관모형.

1. 서론

범주형 자료를 시각적으로 표현할 수 있는 구현 방법들 중 'four-fold circular display(Fienberg 1975)', 'two-way plot(Tukey 1977)', 그리고 모자이크 그림(mosaic plot)(Hartigan과 Kleiner 1984, Friendly 1994)은 3차원 이상도 가능하지만 주로 2차원 자료를 구현하여 준다. 다차원 범주형 자료에 대한 시각적인 표현과 변수들 간의 연관관계를 탐색할 수 있는 방법 중의 하나로 오민권, 홍종선, 이종철(1999)은 링차트(ring chart)를 제안하였는데 범주형 자료를 표현하는 링차트를 비롯하여 표준화된 링차트(standardized ring chart), 조건부 링차트(conditional ring chart), 표준화된 조건부 링차트(standardized conditional ring chart) 그리고 잔차 링차트(residual ring chart))를 제안하였다.

다차원 자료를 표현하는 그림들 중에서 four-fold circular display와 모자이크 그림은 두 변수들간의 연관 정도를 설명하여 준다. 그리고 모자이크 그림을 구현하고 활용하기 위한 소프트웨어로는 S-Plus 와 MANET(Hofmann 1999) 등이 있는데, 이들은 두 변수간의 독립적인 관계를 설명함에도 불구하고 세 변수 이상인 경우 동시에 연관관계를 살펴보는 데 어려움을 가지고 있다. 또한 다차원 자료를 표준화된 링차트로 표현하면, 구현된 변수의 순서에서 처음 고려된 두 변수의 교호작용(interaction)의 존재만을 파악할 수 있다. 그러나 본 논문에서는 표준화된 조건부 링차트를 통하여 고려되는 처음 고려된 두 변수뿐만 아니라 여러 종류의 교호작용항의 존재 유무를 동시에 파악할 수도 있다.

1) (110-745) 서울시 종로구 명륜동 3가 성균관대학교 경제학부 통계학전공, 교수
E-mail: cshong@skku.ac.kr

2) (110-745) 서울시 종로구 명륜동 3가 성균관대학교 경제학부 통계학전공, 강사
E-mail: jongskku@dragon.skku.ac.kr

본 논문의 이론은 오민권 등(1999)에서 제안한 링차트와 조건부 링차트를 기반으로 전개되므로 오민권 등(1999) 논문에서의 그림 2.1, 그림 2.2 그리고 그림 4.1을 정확하게 이해하여야 한다. 논문의 구성은 다음과 같다. 2차원 이상의 범주형 변수들의 연관관계를 나타내는 여러 일차교호작용항(first-order interaction term)의 존재성의 유무를 표준화된 조건부 링차트를 이용하여 동시에 파악할 수 있는 이론을 2절에서 설명하고, 3절에서는 실제 자료에 적합한 로그선형모형을 조건부 링차트로 구현하여 파악된 일차교호작용의 존재유무의 결과와 일치함을 서술하였다.

2. 변수들의 연관관계

2차원 $I \times J$ 분할표의 자료에서 두 변수의 교호작용이 존재하는 경우에는 이미 오민권 등(1999)의 논문에서 자세히 언급하였다. 단지 첫 번째 변수에 의하여 표준화된 링차트에서 임의의 $i = 1, \dots, I-1, j = 1, \dots, J-1$ 에 대하여 q_{ij} 와 $q_{i(j+1)}$ 의 경계선(구분선이라고 한다)과 $q_{(i+1)j}$ 와 $q_{(i+1)(j+1)}$ 의 구분선이 $2\pi/I$ 의 각도(예를 들어 분할표 자료에서는 180도)를 이루고 있으면 두 변수가 독립이라고 판명된다는 점을 상기한다. 여기에서 모든 i 와 j 에 대하여 $q_{ij} = p_{ij}/(I \times p_{i+})$ 으로 정의되는데 두 변수가 독립이면 교차적비(cross product ratio)가 1이며 이는 $q_{ij}q_{i'j'}/q_{i'j}q_{ij} = p_{ij}p_{i'j'}/p_{i'j}p_{ij} = 1$ ($i \neq i', j \neq j'$) 이다(오민권 등(1999)의 (2.4)식을 참조). 여기에서는 3차원과 4차원 범주형 자료를 구성하고 있는 변수들의 연관관계를 중심으로 토론하고자 한다.

2.1. 3차원 자료

3차원 범주형 자료에서 가장 간단한 $2 \times 2 \times 2$ 분할표에 대하여 설명하고자 한다. 층변수(변수 C)의 범주수준에 따라 나누어진 2차원 분할표 자료에서 행변수에 대해 표준화시킨 자료를 조건부 링차트로 표현하면, 행변수(변수 A)와 열변수(변수 B)의 독립성뿐만 아니라 변수 A와 변수 B의 독립성이 만족한 상태에서 열변수(변수 B)와 층변수(변수 C)의 독립성도 파악할 수 있다.

(1) 변수 A와 B의 독립성

그림 2.1은 링차트를 구현할 때 고려하는 변수의 순서가 A-B 순서로 작성되고 변수 C를 조건으로 한 조건부 링차트이다. 즉 변수 C의 각 범주를 조건부로 하고 변수 A에 대해 표준화시킨 링차트이다. 그림 2.1의 조건부 링차트에서 변수 C의 범주수준인 C1, C2에서의 표준화된 링을 살펴보면, $\overline{a0}$ 와 $\overline{b0}$ 가 일직선($2\pi/2 = 180$ 도)을 이루고 동시에 $\overline{a'0}$ 와 $\overline{b'0}$ 도 일직선으로 표현된다. 이는 2차원 자료에 대하여 언급한 바와 같이 변수 A와 변수 B가 변수 C의 각 범주수준에서 독립임을 인지할 수 있다. 변수 C의 모든 범주수준에서 변수 A와 변수 B가 독립이므로 변수 A와 변수 B는 독립적인 관계로 설명된다.

일반적인 $I \times J \times K$ 인 3차원 분할표 자료로 확장시키면 2차원의 경우와 동일한 결과를 얻을 수 있다. 임의의 $i = 1, \dots, I-1, j = 1, \dots, J-1, k = 1, \dots, K$ 에서 q_{ijk} 와 $q_{i(j+1)k}$ 의 구분선과 $q_{(i+1)jk}$ 와 $q_{(i+1)(j+1)k}$ 의 구분선이 모두 $2\pi/I$ 의 각도를 이루고 있으면 변수 A와 B는

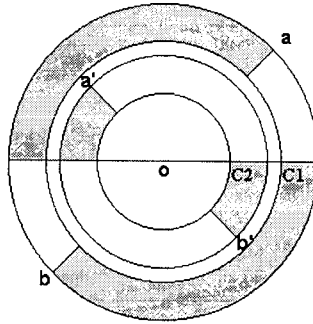


그림 2.1: 변수 A와 B가 독립

독립이라고 판단된다.

(2) 변수 B와 C의 독립성

그림 2-2에서는 두 개의 구분선 \overline{ao} 와 $\overline{a'o}$ 가 서로 중복되고, 동시에 \overline{bo} 와 $\overline{b'o}$ 도 서로 겹쳐서 $\overline{aa'o}$ 와 $\overline{bb'o}$ 가 일직선으로 나타난다. 이런 경우에는 각 조건부 링에서의 표준화된 칸 확률(standardized cell probability)의 크기가 일치한 경우이다. 우선 변수 C의 첫 번째 수준의 외부 링(outer ring)과 두 번째 수준의 내부 링(inner ring)에서 다음이 만족함을 알 수 있다.

$$q_{111} = q_{112}, \quad q_{121} = q_{122}. \quad (2.1)$$

식 (2.1)을 다시 표현하면

$$\frac{q_{111}/q_{121}}{q_{112}/q_{122}} = 1$$

이 되며, 이와같은 관계는 변수 A의 모든 범주수준인 $i=1,2$ 에서도 동일한 식을 유도할 수 있으므로 우리는 다음과 같은 교차적비의 관계를 유도할 수 있다.

$$\frac{q_{i11}q_{i22}}{q_{i12}q_{i21}} = 1, \quad (2.2)$$

여기서 $q_{ijk} = p_{ijk}/(I \times p_{i+k})$ 이다. 즉 변수 A의 모든 범주수준에서 식 (2.2)의 관계가 성립한다면, 변수 B와 C는 독립인 관계가 성립한다고 설명이 가능하다.

그림 2.1의 경우는 각 링의 구분선이 일직선을 나타내고는 있지만 중복되지 않으므로 변수 A와 B는 독립이지만 변수 B와 C가 독립이 아니며 그림 2.2에서는 변수 B와 C가 독립인 관계를 갖고 있다고 파악된다.

이러한 관계를 일반적인 $I \times J \times K$ 인 3차원 분할표 자료로 확장시키면 다음과 같다. 모든 i 와 j 에 대하여

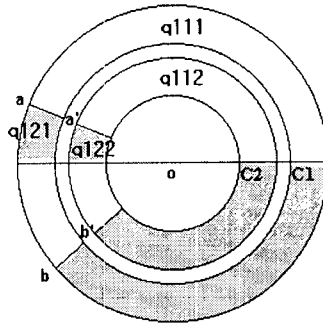


그림 2.2: 변수 B와 C가 독립

$$q_{ijk} = q_{ijk'} \quad (k \neq k')$$

이 만족하면 변수 B와 C의 관계는 독립이다.

(3) 결합관계

2.1절의 (1)과 (2)를 결합시킨 관계를 살펴보면 그림 2.3과 같은 표준화된 링차트를 얻을 수 있는데 구분선 \overline{ao} 와 $\overline{a'o}$ 그리고 \overline{bo} 와 $\overline{b'o}$ 가 각각 중복되어 $\overline{aa'o}$ 와 $\overline{bb'o}$ 로 겹치는 현상이 발생하며 그리고 $\overline{aa'o}$ 와 $\overline{bb'o}$ 는 180도 각도를 이루어 일직선으로 나타난다면, 변수 A와 B 그리고 변수 B와 C가 동시에 독립적인 관계임을 탐색할 수 있다.

이와 같이 3차원 범주형 자료에서 한 종류의 표준화된 조건부 링차트를 이용하여 최대 2개의 독립성의 관계를 발견할 수 있다는 사실을 유도하였다.

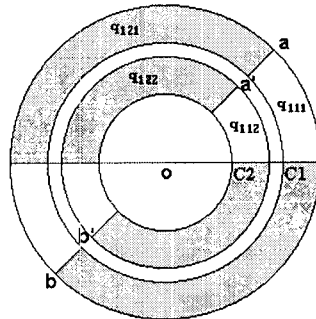


그림 2.3: 변수 A와 B, 그리고 변수 B와 C가 독립

2.2. 4차원 자료

예를 들어 4차원 자료인 $2 \times 2 \times 2 \times 2$ 분할표를 변수 D의 각 수준별로 분할한 뒤 표준화된 조건부 링차트를 작성하여 보기로 하자. 변수 D의 첫 번째 범주에 해당하는 외부 링

은 'D1'으로 표시하고 두 번째 범주에 해당하는 내부 링은 'D2'로 표시한다. 이러한 조건부 링차트를 변수 A-B-C의 순서로 구현하면 표준화된 조건부 링차트의 각 칸은 변수 A에 대해 표준화된 칸 확률(q_{ijkl})을 의미한다.

(1) 변수 A와 B의 독립성

그림 2.4에서도 그림 2.1에서와 동일하게 \overline{ao} 와 \overline{bo} 는 일직선을, 그리고 동시에 $\overline{a'o}$ 와 $\overline{b'o}$ 도 일직선으로 표현된다. 이는 표준화된 링차트에서 임의의 $l = 1, 2$ 에 대하여 q_{11+l} 과 q_{12+l} 의 구분선 \overline{ao} 와 \overline{bo} 그리고 q_{21+l} 과 q_{22+l} 의 구분선 $\overline{a'o}$ 와 $\overline{b'o}$ 가 $2\pi/2 = 180$ 도를 이루고 있다. 우선 외부 링 D1을 살펴보기로 하자. \overline{ao} 와 \overline{bo} 가 180도의 각도로 일직선을 이루면, $q_{11+l} = q_{21+l}$ 이고 또한 $q_{12+l} = q_{22+l}$ 이다. 이와 같은 칸확률들의 관계를 교차적비로 표현하면 다음과 같이 얻을 수 있다.

$$\frac{q_{11+l}q_{22+l}}{q_{12+l}q_{21+l}} = \frac{p_{11+l}p_{22+l}}{p_{12+l}p_{21+l}} = 1,$$

여기에서 $q_{ijkl} = p_{ijkl}/(I \times p_{i++})$ 이다.

내부 링 D2도 구분선 $\overline{a'o}$ 와 $\overline{b'o}$ 가 일직선을 이루므로 변수 D의 범주 구분에 상관없이 교차적비의 값이 모두 1이며, 이는 변수 A와 B가 독립이라는 것을 보여준다. 그림 2.3의 표준화된 조건부 링차트는 2.1절에서 언급한 바와 유사하게 변수 C에 대하여 고려하지 않은 상황에서 그리고 변수 D의 모든 범주수준에서 변수 A와 변수 B가 독립이므로 변수 A와 변수 B는 독립적인 관계를 갖고 있다고 설명된다.

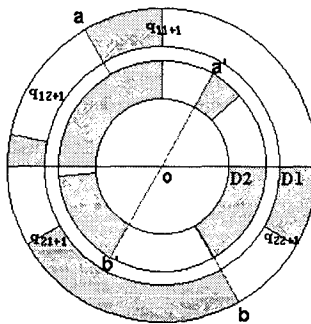


그림 2.4: 변수 A와 변수 B가 독립

일반적인 $I \times J \times K \times L$ 분할표에서도 임의의 $i = 1, \dots, I - 1, j = 1, \dots, J - 1, l = 1, \dots, L - 1$ 에 대하여 q_{ij+l} 과 $q_{i(j+1)+l}$ 의 구분선과 $q_{(i+1)j+l}$ 과 $q_{(i+1)(j+1)+l}$ 의 구분선이 모두 $2\pi/I$ 의 각도를 이루고 있으면 변수 A와 B는 독립적인 관계를 갖고 있다고 확장할 수 있다.

(2) 변수 A와 B의 독립과 동시에 변수 A와 C의 독립성

2.2절 (1)에서 논의한 변수 A와 B가 독립성을 만족하는 조건하에, 즉 구분선 \overline{ao} 와 \overline{bo} 그리고 $\overline{a'o'}$ 와 $\overline{b'o'}$ 가 일직선을 나타낼 때, 그림 2-5에서와 같이 다른 종류의 구분선 \overline{co} 와 \overline{do} 그리고 $\overline{c'o'}$ 와 $\overline{d'o'}$ (이는 임의의 k 와 l 에 대하여 q_{11kl} 과 q_{12kl} 의 구분선과 q_{21kl} 과 q_{22kl} 의 구분선을 말한다)가 추가적으로 180도 각도를 이루어 일직선으로 표현된다면, 각 칸확률을 이용하여 두 구분선이 일직선이라는 사실로 다음과 같은 관계를 유도할 수 있다. 우선 외부 링 D1에서

$$q_{1111} = q_{2111}, q_{1121} = q_{2121}$$

의 관계를 갖고있다. 이를 교차적비의 형태로 나타내면

$$\frac{q_{1111}q_{2121}}{q_{1121}q_{2111}} = 1$$

이며, 변수 B와 D의 임의의 범주수준 j, l 에 대해

$$\frac{q_{1j1l}q_{2j2l}}{q_{1j2l}q_{2j1l}} = \frac{p_{1j1l}p_{2j2l}}{p_{1j2l}p_{2j1l}} = 1 \tag{2.3}$$

이 성립하므로 변수 A와 C가 독립이라는 것을 보여준다. 즉 그림 2.5와 같이 표준화된 조건부 링차트에서 구분선 \overline{ao} 와 \overline{bo} 그리고 $\overline{a'o'}$ 와 $\overline{b'o'}$ 가 일직선을 나타냄과 동시에 추가적으로 다른 종류의 구분선 \overline{co} 와 \overline{do} 그리고 $\overline{c'o'}$ 와 $\overline{d'o'}$ 가 일직선으로 표현된다면, 변수 A와 B가 독립임과 동시에 변수 A와 C가 독립이라는 사실을 발견하였다.

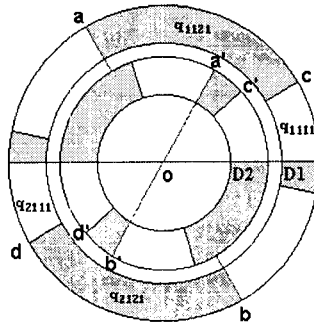


그림 2.5: 변수 A와 B, 변수 A와 C가 독립

4차원 분할표인 $I \times J \times K \times L$ 에서는 임의의 $i = 1, \dots, I - 1, j = 1, \dots, J - 1, l = 1, \dots, L - 1$ 에 대하여 q_{ij+l} 과 $q_{i(j+1)+l}$ 의 구분선과 $q_{(i+1)j+l}$ 과 $q_{(i+1)(j+1)+l}$ 의 구분선이 모두 $2\pi/I$ 의 각도를 이루고 있고 모든 $k = 1, \dots, K$ 에 대하여

$$q_{ijkl} = q_{(i+1)jkl}$$

이 만족한다면 변수 A와 B 그리고 변수 A와 C도 독립이라고 확장된다.

(3) 변수 B와 D의 독립성

그림 2.6에서와 같이 표준화된 확률 q_{11+1} 과 q_{12+1} 의 구분선(\overline{ao})과 q_{11+2} 과 q_{12+2} 의 구분선($\overline{a'o}$)은 180도 각도를 이루고 있지 않으나 (변수 A와 B는 연관관계이다) 원점에 대하여 동일한 각도를 이루고 있으며, 또한 q_{21+1} 과 q_{22+1} 의 구분선(\overline{bo})과 q_{21+2} 과 q_{22+2} 의 구분선($\overline{b'o}$)도 동일한 각도를 이루고 있어서 서로 중복되는 직선을 이루고 있는지 여부에 따라 변수 B와 D의 독립성 여부를 파악할 수 있다. 우선 구분선 \overline{ao} 와 $\overline{a'o}$ 가 중복되는 직선인 경우에 대하여 살펴보자. 만약 구분선 $\overline{aa'o}$ 이 직선이면 다음의 조건이 성립한다.

$$q_{11+1} = q_{11+2} , q_{12+1} = q_{12+2} .$$

이를 교차적비의 형태로 표현하면

$$\frac{q_{11+1} q_{12+2}}{q_{11+2} q_{12+1}} = 1$$

임을 얻을 수 있다. 또한 구분선 $\overline{bb'o}$ 가 일직선인 경우도 동일하게 구할 수 있으므로 변수 A의 범주수준 $i = 1, 2$ 에 대해

$$\frac{q_{i1+1}q_{i2+2}}{q_{i1+2}q_{i2+1}} = \frac{p_{i1+1}p_{i2+2}}{p_{i1+2}p_{i2+1}} = 1$$

이 성립하면 변수 B와 D가 독립임을 보여준다.

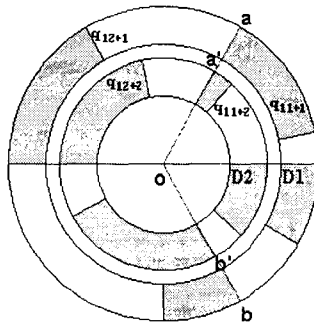


그림 2.6: 변수 B와 D 독립

일반적인 $I \times J \times K \times L$ 인 4차원 분할표에서는 임의의 $i = 1, \dots, I-1, j = 1, \dots, J-1$ 에 대하여

$$q_{ij+l} = q_{ij+l'} \quad (l \neq l')$$

이면 변수 B와 D는 독립이라고 확장된다.

(4) 변수 B와 D의 독립과 동시에 변수 C와 D의 관계

앞에서 논의한 변수 B와 D의 독립성은 구분선 \overline{ao} 와 $\overline{a'o}$ 가 그리고 \overline{bo} 와 $\overline{b'o}$ 가 각각 원점에 대하여 동일한 각도를 이루고 있을 때 만족한다는 것을 살펴보았다. 여기에서는 이러한 조건이 성립한다고 가정하고 추가적으로 다른 종류의 구분선들(예를 들어 q_{1111} 과 q_{1121} 의 구분선 \overline{co} 와 q_{1112} 과 q_{1122} 의 구분선 $\overline{c'o}$)이 원점에 대하여 동일한 각도를 이루어 일직선으로 표현된다면, 다음과 같은 관계가 만족한다.

$$q_{1111} = q_{1112}, q_{1121} = q_{1122}.$$

이러한 관계를 교차적비의 형태로 표현하면,

$$\frac{q_{1111} q_{1122}}{q_{1112} q_{1121}} = 1$$

을 얻을 수 있으며, 모든 구분선(예를 들어 $\overline{aa'o}$, $\overline{bb'o}$, $\overline{cc'o}$)들이 일직선인 경우에는 변수 A와 변수 B의 어떤 범주수준 i, j 에 대해

$$\frac{q_{ij11} q_{ij22}}{q_{ij12} q_{ij21}} = 1$$

이 성립한다는 것을 유도할 수 있다. 그러므로 그림 2.7와 같이 조건으로 하는 변수 D의 범주수준에 상관없이 모든 구분선들이 원점에 대하여 동일한 각도를 이루어 중복되게 표현된다면, 변수 B와 D가 독립임과 동시에 변수 C와 D도 독립적인 관계가 나타남을 발견하였다.

4차원 분할표인 $I \times J \times K \times L$ 자료에서는 임의의 $i = 1, \dots, I - 1, j = 1, \dots, J - 1$ 그리고 모든 k 에 대하여

$$q_{ijkl} = q_{ijk'l'} \quad (l \neq l')$$

이면 변수 B와 D 그리고 변수 C와 D도 독립이라고 확장된다.

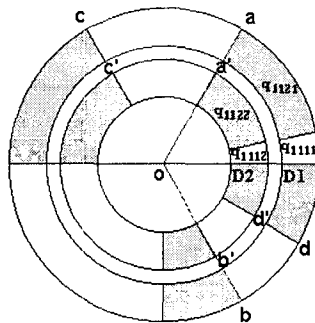


그림 2.7: 변수 B와 D, 변수 C와 D 독립

(5) 추가적인 관계들

이와 같이 표준화된 조건부 링차트를 이용하여 4차원 범주형 자료들에 대한 변수들간의 독립관계 4가지를 살펴보았다. 그리고 이런 관계들의 결합인 경우에 대하여 다음과 같은 3가지 경우를 추가적으로 고려하여 보자. 다음 경우들은 변수 D의 수준을 조건으로 하고 링차트를 구현할 때 고려되는 변수의 순서가 A-B-C이며 변수 A에 대하여 표준화된 조건부 링차트들을 작성한 것이다.

- 경우 1 : (2.2절의 (1), (2), (3)의 내용 결합)
변수 A와 B, A와 C, 그리고 B와 D는 독립이고 변수 C와 D는 연관인 경우
- 경우 2 : (2.2절의 (1), (3)의 내용 결합)
변수 A와 B, B와 D는 독립이고 변수 A와 C, C와 D는 연관인 경우
- 경우 3 : (2.2절의 (1), (2), (3), (4)의 내용 결합)
변수 A와 B, A와 C, B와 D, 그리고 C와 D가 독립인 경우

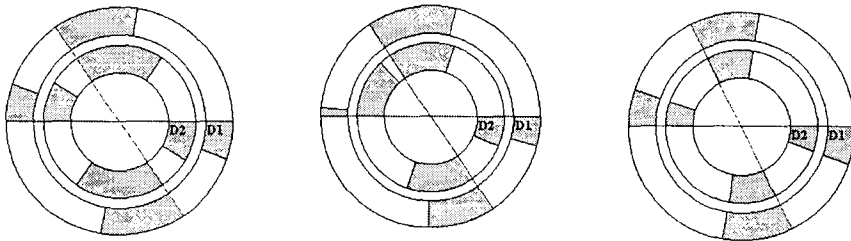


그림 2.8: 경우 1, 2, 3의 조건부 링차트

여기에서 논의한 3가지 경우를 포함해서 2.2절에서 토론한 표준화된 조건부 링차트를 구현하여 자세히 관찰하면, 4차원 범주형 자료를 구성하고 있는 변수들의 연관성과 독립성 관계를 탐색할 수 있다는 사실을 발견하였다.

2.1 (1)절과 2.2 (1) 절의 내용은 모두 링차트를 작성하면서 고려하는 변수순으로 볼때 처음 두 변수가 독립적인 관계임을 파악할 수 있고, 2.1 (2)절과 2.2 (3) 절의 내용도 동일하게 고려하는 변수순의 마지막 두 변수가 독립적임을 나타낸다. 그리고 2.2 (2)절과 2.2 (4)절에서 논의한 독립성은 2.2 (1)절과 2.2 (3)절의 연결선상에서 고려할 수 있다. 이와 같은 이론은 5차원 이상의 자료에도 확장하여 정리할 수 있으나 독자에게 남겨두기로 한다.

3. 실증적 예제

3.1. $4 \times 2 \times 2$ 인 3차원 자료

덴마크의 Welfare Study 에서 조사한 자료로 사회적 지위와 냉장고 보유에 따라 거주 주택이 자가 또는 임대인지를 조사한 자료이다(Andersen, 1991, p174).

표 3.1: $4 \times 2 \times 2$ 인 3차원 자료

사회적 지위 (A)	임대/자가 (B)	냉장고 보유(C)	
		예	아니오
I	임대	304	38
	자가	92	64
II	임대	666	85
	자가	174	113
III	임대	894	93
	자가	379	321
IV	임대	720	84
	자가	433	297

이 자료에 적합한 로그선형모형은 $[AB][BC](G^2 = 7.38, p\text{-값}=0.2873)$ 이다. 이를 조건부 링차트로 구현하여 변수들간의 연관관계를 살펴보기로 하자. 먼저 변수 B의 조건하에 변수 A-C의 순서로 구현된 표준화된 링차트를 살펴보자.

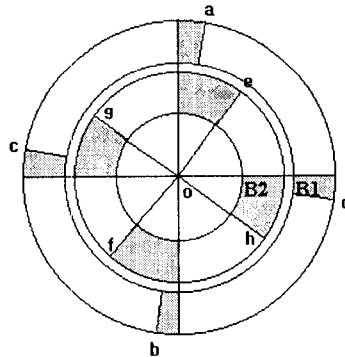


그림 3.1: 변수 A-C 순으로 구현된 표준화된 조건부 링차트

그림 3.1을 살펴보면 조건부로 사용된 변수 B의 각 수준별로 나타난 두 개의 링에서 변수 A의 범주 수가 4개이므로 각 링이 4등분($360\text{도}/4 = 90\text{도}$)됨을 알 수 있다. 우선 외곽 링(B1)에서 살펴보면 구분선 \overline{aob} , \overline{cod} 가 직교하고 안쪽 링(B2)에서의 구분선 \overline{eof} 와 \overline{goh} 도 직교한다는 것을 알 수 있다. 이것은 2.1 (1)절과 그림 2.2에서 언급한 바와 같이 변수 A와 C가 독립적인 관계를 갖고 있다고 파악할 수 있지만 2.1 (3)절과 그림 2.3의 형태와는 달리 구분선 \overline{ao} 와 \overline{eo} , \overline{bo} 와 \overline{fo} 는 일직선을 이루지 않으므로 변수 A와 B는 연관성이 존재한다고 파악된다.

그림 3.2는 변수 C를 조건으로 하고 변수 B-C의 순서로 구현된 표준화된 링차트이다. 2.1 (1)절에서의 설명을 바탕으로 변수 A의 범주 수가 4개이므로 구분선 \overline{aob} , \overline{cod} , \overline{eof} 가

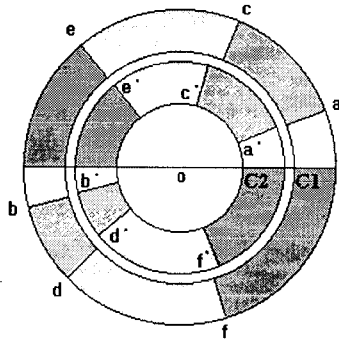


그림 3.2: 변수 B-A 순으로 구현된 표준화된 조건부 링차트

일직선인지 여부를 통하여 변수 A와 B가 독립적인 관계를 가짐을 살펴볼 수 있다. 그러나 모두 일직선에 가깝지 않으므로 변수 A와 B가 연관이 존재함을 보여줄 뿐만 아니라 $\overline{aa'o}$, $\overline{cc'o}$, $\overline{ee'o}$, $\overline{bb'o}$, $\overline{dd'o}$, $\overline{ff'o}$ 는 모두 일직선에 가까우므로 2.1 (2)절과 그림 2.2의 설명에 의해 변수 A와 C가 독립임을 파악할 수 있다.

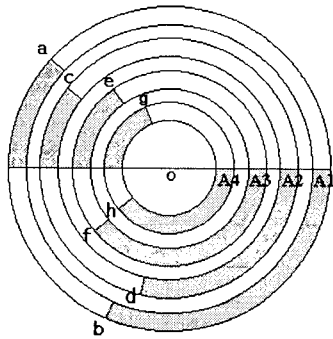


그림 3.3: 변수 C-B 순으로 구현된 표준화된 조건부 링차트

그림 3.3의 조건부 링차트는 변수 A를 조건으로 하고 변수 C-B의 순서로 구현되었고 \overline{aob} , \overline{cod} , \overline{eof} , \overline{goh} 가 일직선을 이루지 않으므로 변수 C와 B가 연관을 가짐을 보여준다. 또한 \overline{ao} , \overline{co} , \overline{eo} , 그리고 \overline{go} 가 일직선을 이루지 않으므로 변수 A와 B도 연관을 가짐을 보여주고 있다.

자료에 적합한 로그선형모형을 그림 3.1, 그림 3.2, 그리고 그림 3.3의 조건부 링차트를 통해서 변수 A와 C만이 독립적인 관계를 나타내고 A와 B, B와 C는 서로 연관을 가짐을 살펴볼 수 있다. 그러므로 자료에 가장 적합한 모형 $[AB][BC]$ 와 일치한 분석결과를 살펴볼 수 있다.

3.2. $2 \times 2 \times 2 \times 2$ 인 4차원 자료

1983년에 Danish Institute for Building Research에서 16세 이상의 사람에게 가정집에서 실내 환경에 따라 두통여부에 대해 조사한 자료이다(Andersen 1991, p202).

표 3.2: $2 \times 2 \times 2 \times 2$ 인 4차원 자료

흡연습관 (A)	연령 (B)	성별 (c)	두통횟수 (D)	
			일주일내 1회이상	일주일내 1회미만
흡연	40세 이하	남성	11	142
		여성	45	83
	40세 이상	남성	11	145
		여성	15	76
비흡연	40세 이하	남성	8	117
		여성	29	89
	40세 이상	남성	7	113
		여성	8	80

위 자료에 적합한 로그선형모형은 $[AD][BC][BD][CD](G^2 = 9.98, p\text{-값}=0.1895)$ 로 이 조건부 링차트를 이용하여 변수들간의 연관여부를 살펴보기로 하자.

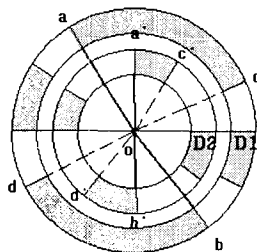


그림 3.4: 변수 A-B-C 순으로 구현된 표준화된 조건부 링차트

그림 3.4는 변수 D를 조건으로 하여 변수 A-B-C의 순서로 구현된 표준화된 링차트이다. 이는 2.2 (2)절의 그림 2.5와 유사한 형태를 이루고 있다. 즉 \overline{aob} , $\overline{a'ob'}$ 가 일직선을 이루므로 변수 A와 B가 독립적인 관계임을 파악할 수 있다. 또한 \overline{cod} , $\overline{c'od'}$ 가 일직선을 이루므로 변수 A와 C가 독립적인 관계임을 살펴볼 수 있다.

그림 3.5를 살펴보면 2.2 (4)절의 그림 2.7과 유사한 형태를 보이고 있다. 즉 \overline{ao} 와 $\overline{a'o}$, 그리고 \overline{bo} 와 $\overline{b'o}$ 도 거의 일직선이므로 변수 A와 B가 독립적인 관계를 가짐을 나타내고 \overline{co} 와 $\overline{c'o}$, 그리고 \overline{do} 와 $\overline{d'o}$ 도 일직선에 가까우므로 변수 C와 A도 독립적인 관계에 있음을 알 수 있다.

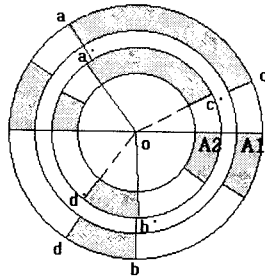


그림 3.5: 변수 D-B-C 순으로 구현된 표준화된 조건부 링차트

따라서 그림 3.4 또는 그림 3.5의 조건부 링차트를 통하여 $[AD][BC][BD][CD]$ 인 모형과 일치한 결과를 파악할 수 있다.

4. 결론

본 논문에서는 다차원 범주형 자료를 다양한 종류의 링차트 중에서 표준화된 조건부 링차트를 이용하여 여러 개의 연관관계항의 존재여부를 동시에 살펴볼 수 있음을 토론했다. 또한 실제 자료를 이용하여 적합한 로그선형모형이 갖는 여러 개의 연관관계항의 존재를 살펴볼 수 있다.

구현하는 링차트에서의 각 칸은 링의 각도로 표현되기 때문에 칸의 수가 360개 이상인 경우에는 칸의 크기에 해당되는 각도가 1도 보다 작기 때문에 링차트로 표현하는데 한계가 있다. 따라서 2^8 인 8차원이상의 자료이거나 모든 범주수가 256개이상인 자료에 대해 링차트를 이용하여 변수들간의 연관관계를 살펴보는 데 어려움을 갖고 있다. 또한 링차트로는 변수들의 독립적인 관계 또는 일차교호작용의 관계만을 파악할 수 있을 뿐이다. 여기서 제안한 링차트를 이용하여 이차이상의 교호작용관계를 설명하는데 한계가 있으므로 향후 연구과제로 남겨 두기로 한다.

조건부 링차트와 표준화된 조건부 링차트를 포함하여 다차원범주형 자료를 표현할 수 있는 다양한 링차트를 구현하는 시스템은 JAVA로 작성되었으며 MS Internet Explorer(4.0 버전 이상)을 사용하여 <http://stat.skku.ac.kr/omg0906/Ring.html>을 접속하면 이용할 수 있다.

참고문헌

- [1] 오민권, 홍종선, 이종철 (1999). 다차원 범주형 자료에 대한 링차트. <응용통계연구>, 제12권 1호, 225-239.
- [2] Agresti, A. (1990). *Categorical Data Analysis*, John Wiley & Sons.
- [3] Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). *Discrete Multivariate Analysis, Theory and Practice*, the MIT Press.
- [4] Christensen, R. (1997). *Log-linear Models and Logistics Regression*, Springer-Verlag.
- [5] Fienberg, S.E. (1975). Perspective Canada as a social report, *Social Indicators Research* 2, 154-174.
- [6] Friendly, M. (1994). Mosaic displays for multi-way contingency tables, *Journal of the American Statistical Association*, Vol. 89, 190-200.
- [7] Hartigan, J.A. and Kleiner, B. (1984). A mosaic of the television ratings, *The American Statistician*, Vol. 38, 32-35.
- [8] Heike Hofmann (1999). Simpson on board the Titanic? Interactive Methods for Dealing with Multivariate Categorical Data, *Statistical Computing & Statistical Graphics Newsletter*, Vol. 9, 16-19
- [9] Tukey, J.W. (1977). *Exploratory Data Analysis*, Addison-Wesley Publishing Company.

[1999년 7월 접수, 1999년 12월 채택]

Ring Chart II for Multidimensional Categorical Data Analysis using Conditional Ring Charts

Chong Sun Hong ¹⁾ Jong Cheol Lee ²⁾

ABSTRACT

In this paper, using the standardized conditional ring charts, the existences of more than two first-order interaction terms that the appropriate log-linear model have could be represented simultaneously. At most 2 (4) first-order interaction terms can be found for three (four) dimensional data with the standardized conditional ring charts.

Keywords: Association model; Cross-product ratio; Independent model; Interaction; Log-linear model.

1) Professor, Department of Statistics, SungKyunKwan University.

E-mail: cshong@skku.ac.kr

2) Lecturer, Department of Statistics, SungKyunKwan University.

E-mail: jongskku@dragon.skku.ac.kr