한 국 통 계 학 회 논 문 집
2009, 16권, 5호, 753–762

$m \times m$ 분할표에서의 합치도 H

김진곤[a], 박미희[a], 박용규[1,a]

[a]가톨릭대학교 의학통계학과

요 약

평정자간 일치 정도를 나타내는 측도로 주로 사용되는 κ의 문제점을 해결하기 위해 박미희와 박용규 (2007)는 2×2 분할표에서 새로운 합치도 H를 제안하였다. 본 연구에서는 이를 확장하여 $m \times m$ 분할표에 대한 합치도 H와 그 분산을 구한다. 또한 κ의 역설과 주변분포와의 관계를 증명하고, 3×3 분할표 예제를 이용하여 기존의 합치도들과 비교한다.

주요용어: 합치도 H, κ의 역설, 주변분포.

1. 서론

합치도(measure of agreement)란 평정자들의 평가가 서로 어느 정도 일치하는가를 알아보는 신뢰도 평가의 측도로서, Cohen (1960)의 κ가 대표적이다. κ는 확률 모형에 의존하지 않고 계산이 간편하다는 장점이 있으나, 우연에 의해 나타나는 합치에 대한 정의가 불명확하고, 주변분포에 매우 민감한 측도라는 단점이 있다. κ 이외에도 여러 합치도들이 제시되었으나, 계산이 복잡하거나, 확률 모형에 의존하는 등의 여러 문제들이 있다.

Bennet 등 (1954)은 반응 범주의 역수가 우연에 의한 합치 비율의 가장 좋은 추정치라 하였고, Scott (1955)은 두 평정자가 주어진 범주에 같은 확률로 개체를 분류할 것이라는 가정 하에서 합치도를 추정하는 방법인 π를 제시하였다. Cicchetti 등 (1978)은 전체 합치만을 제시했을 때는 나타낼 수 없는 각 범주에 대한 평정자들의 합치정도를, 특정 범주 일치(specific category agreement)를 제시하여 각 범주별로 각각의 합치도를 따로 구해내는 방법을 제안하였다. 또한 Holley와 Guilford (1964)는 Bennet 등 (1954)의 우연에 의한 합치 비율을 사용한 G-지수를, Janson과 Vegelius (1979)는 2×2 분할표에서의 승적률(product-moment) 상관계수인 파이 계수(phi(φ) coefficient)를 $m \times m$분할표로 확장한 S를 제안하였다. Gwet (2001)은 평정자내 변동(intraobserver variation)을 나타내는 불확실성 계수를 이용한 AC_1을 제안하였다. 최근에는 박미희와 박용규 (2007)가 Feinstein과 Cicchetti (1990)가 제기한 2×2 분할표에서 나타나는 κ의 두 가지 역설을 보완하기 위해, 조화평균을 이용하여 평정자들의 평가 비율에 대한 비대칭성과 불균형성을 함께 보정한 새로운 합치도 H를 2×2 분할표에서 제안하였다.

본 연구에서는 κ의 역설과 주변분포와의 관계를 증명하고, 2×2 분할표에서 제안된 합치도 H를 3×3 분할표 이상으로 일반화하였다. 2절에서는 3×3 분할표에서의 여러 합치도를 소개하고 3×3 분할표에서의 κ의 역설을 예제 자료를 통하여 설명하였으며, 또한 κ의 역설과 주변분포와의 관계를 설명하였다. 3절에서는 3×3 분할표를 비롯해 $m \times m$ 분할표에서의 합치도 H를 제시하고, H의 최대우도 추정량과 분산을 유도하였다. 4절에서는 3×3 분할표의 예제를 통해 H와 다른 합치도들을 비교하였고, 5절에서는 본 연구의 결과와 의의에 대하여 정리하였다.

[1] 교신저자: (137-701) 서울시 서초구 반포4동 가톨릭대학교 성의교정 의학통계학과, 교수.
E-mail: ygpark@catholic.ac.kr

표 1: 평정자와 반응범주에 의한 전체 n개체의 분포

평정자 A	평정자 B 1	2	3	Total
1	n_{11}	n_{12}	n_{13}	n_{1+}
2	n_{21}	n_{22}	n_{23}	n_{2+}
3	n_{31}	n_{32}	n_{33}	n_{3+}
Total	n_{+1}	n_{+2}	n_{+3}	n

표 2: 다른 합치도들의 p_e

	3×3	$m \times m$
$p_e(\pi)$	${P_1}^2 + {P_2}^2 + {P_3}^2$	$\sum_{i=1}^{m} {P_i}^2$
$p_e(G)$	$\frac{1}{3}$	$\frac{1}{m}$
$p_e(\mathrm{AC}_1)$	$\frac{1}{2}[P_1(1-P_1) + P_2(1-P_2) + P_3(1-P_3)]$	$\frac{1}{m-1}\sum_{i=1}^{m} P_i(1-P_i)$

2. κ의 역설

2.1. 3×3 분할표와 합치도

두 명의 평정자 A, B가 전체 n개의 개체를 세 개의 가능한 명목형 범주 "1"과 "2"와 "3" 중 하나로 각각 평가할 때, 그 결과는 표 1과 같이 요약된다. 표 1에서 n_{ij}는 평정자 A가 i의 범주로, 평정자 B가 j의 범주로 분류한 개체의 수를 나타낸다. 그리고 n_{i+}는 평정자 A가 i범주로 분류한 총 개체 수를 나타내며, n_{+j}는 평정자 B가 j범주로 분류한 총 개체 수를 나타낸다($i, j = 1, 2, 3$).
A와 B 두 평정자가 개체들을 같은 범주로 분류한 비율인, 관찰된 합치 비율 p_o는 식 (2.1)과 같다.

$$p_o = \frac{n_{11} + n_{22} + n_{33}}{n}. \tag{2.1}$$

그러나 관찰된 합치 비율 중에는 우연에 의해 기대되는 합치 비율 p_e가 포함되어 있으므로, 식 (2.2)와 같이 이를 보정한 값을 합치도로 사용한다.

$$\frac{p_o - p_e}{1 - p_e}. \tag{2.2}$$

식 (2.2)에서 p_e를 어떻게 정의하느냐에 따라 여러 종류의 합치도가 제안되었다. Cohen의 κ는 p_e를 대응하는 주변확률들의 곱으로 표현한 합치도로, 3×3 분할표에서의 κ의 p_e는 식 (2.3)과 같다.

$$p_e(\kappa) = \left(\frac{n_{1+}}{n}\right)\left(\frac{n_{+1}}{n}\right) + \left(\frac{n_{2+}}{n}\right)\left(\frac{n_{+2}}{n}\right) + \left(\frac{n_{3+}}{n}\right)\left(\frac{n_{+3}}{n}\right). \tag{2.3}$$

2.2. 다른 합치도

Scott (1955)의 π는 각 범주에 대한 평균 평가 비율의 제곱의 합으로, Holley와 Guilford (1964)의 G는 반응 범주의 역수를 각각 우연에 의해 기대되는 합치비율 p_e로 정의하였다. 그리고 Gwet (2001)의 AC_1은 각 범주에 해당하는 분산의 합으로 표현된 불확실성 계수를 이용하여 p_e를 구한 것이다. 이 합치도들을 3×3 분할표와 $m \times m$ 분할표에 적용한 p_e는 표 2와 같다. 여기서 π와 AC_1에서의 P_i는 $\{(n_{i+}/n) + (n_{+i}/n)\}/2$이다.

표 3: κ의 첫 번째 역설에 대한 예제

(a) 대칭적 균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	30	2	2	**34**
2	1	29	1	**31**
3	2	2	31	**35**
Total	**33**	**33**	**34**	100

$p_o = 0.90$, $p_e(\kappa) = 0.33$, $\kappa = 0.85$

(b) 대칭적 불균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	85	1	2	**88**
2	3	3	1	**7**
3	2	1	2	**5**
Total	**90**	**5**	**5**	100

$p_o = 0.90$, $p_e(\kappa) = 0.80$, $\kappa = 0.51$

표 4: κ의 두 번째 역설에 대한 예제

(a) 대칭적 불균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	42	17	3	**62**
2	14	12	1	**27**
3	3	2	6	**11**
Total	**59**	**31**	**10**	100

$p_o = 0.60$, $p_e(\kappa) = 0.46$, $\kappa = 0.26$

(b) 비대칭적 불균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	17	1	32	**50**
2	2	24	3	**29**
3	1	1	19	**21**
Total	**20**	**26**	**54**	100

$p_o = 0.60$, $p_e(\kappa) = 0.29$, $\kappa = 0.44$

(c) 완전한 대칭적 불균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	75	7	2	**84**
2	7	4	1	**12**
3	2	1	1	**4**
Total	**84**	**12**	**4**	100

$p_o = 0.80$, $p_e(\kappa) = 0.72$, $\kappa = 0.28$

(d) 불완전한 대칭적 불균형 주변분포

평정자 A	평정자 B 1	2	3	Total
1	75	13	4	**92**
2	1	4	2	**7**
3	0	0	1	**1**
Total	**76**	**17**	**7**	100

$p_o = 0.80$, $p_e(\kappa) = 0.71$, $\kappa = 0.31$

2.3. 3×3 분할표에서 κ의 역설

Feinstein과 Cicchetti (1990)가 제기한 2×2 분할표에서의 κ의 두 가지 역설은 모두 동일한 p_o값을 갖더라도, p_e값이 크면 식 (2.2)로 표현된 κ는 작게 나타나는 현상을 표현한 것으로, 이 때 p_e값의 크기는 분할표의 주변분포에 의존한다.

κ에 관한 첫 번째 역설은 주변분포가 대칭적인 경우에서, 두 평정자의 평가결과가 한 쪽으로 집중되는 경우가 집중되지 않은 경우보다, 즉, 불균형적일 때가 균형적일 때 보다 합치도가 높을 것이라 예상되는데도 불구하고 p_e값이 크게 나타나 오히려 κ의 값은 작게 된다는 것이다. 이에 대한 3×3 분할표에서의 예는 표 3과 같다.

표 3의 (a)와 (b) 모두 p_o의 값이 0.90으로 동일하지만, (b)의 $p_e(\kappa)$는 0.80으로 (a)의 0.33보다 상당히 큰 값을 가진다. 따라서 (b)의 κ값은 0.51로 (a)의 κ값인 0.85보다 매우 작게 나타났다.

κ에 관한 두 번째 역설은, 주변분포의 불균형이 대칭적으로 나타날 경우 비대칭적일 때보다, 또한 완전한 대칭적일 때가 불완전한 대칭적일 때보다 평정자들 간의 합치정도가 더 높으리라 예상되지만, κ값은 오히려 각각의 경우에 더 작게 나타난다는 것이다. 이에 대한 예는 표 4에 제시되었다.

표 4의 (a)와 (b) 모두 p_o의 값이 0.60으로 동일하지만, (a)의 $p_e(\kappa)$는 0.46으로 (b)의 0.29보다 크다. 따라서 (a)의 κ값은 0.26으로 (b)의 κ값인 0.44보다 작게 나타났다. 또한 (c)와 (d)의 경우도 p_o의 값이 0.80으로 동일하지만, (c)의 $p_e(\kappa)$는 0.72로 (d)의 0.71보다 크다. 따라서 (c)의 κ값은 0.28로서 (d)의 κ값인 0.31보다 더 작게 나타났다.

2.4. κ의 역설과 주변분포와의 관계

3×3 분할표의 각 주변확률은 식 (2.4)와 같이 표현할 수 있다.

$$n_{1+} = v_1 n, \quad n_{2+} = v_2 n, \quad n_{3+} = (1 - v_1 - v_2)n,$$
$$n_{+1} = w_1 n, \quad n_{+2} = w_2 n, \quad n_{+3} = (1 - w_1 - w_2)n, \tag{2.4}$$

여기서 v_1, v_2와 w_1, w_2는 0과 1 사이의 값을 갖는다.

균형적 주변분포란, 모든 v_i와 w_i $(i = 1, 2)$가 1/3에 근접한 값을 갖는 경우를 말하며, 그렇지 않으면 불균형적 주변분포라고 한다. 모든 주변확률의 비율이 1/3의 값을 가지는 경우는 완전한 균형이다. (이하에서는 3×3 분할표와 $m \times m$ 분할표의 κ를 각각 κ_3와 κ_m으로 표현한다.)

식 (2.4)의 v_1, v_2, w_1, w_2를 식 (2.3)에 대입하면 다음과 같다.

$$\begin{aligned} p_e(\kappa_3) &= \frac{v_1 w_1 n^2 + v_2 w_2 n^2 + (1 - v_1 - v_2)(1 - w_1 - w_2)n^2}{n^2} \\ &= v_1 w_1 + v_2 w_2 + (1 - v_1 - v_2)(1 - w_1 - w_2). \end{aligned} \tag{2.5}$$

또한 $v_1 = (1/3) + v_1'$, $v_2 = (1/3) + v_2'$, $w_1 = (1/3) + w_1'$, $w_2 = (1/3) + w_2'$로 표현하면, v_1', v_2', w_1', w_2'가 0에 가까운 값을 가질 때 균형적 주변분포이다. v_1'과 v_2' 그리고 w_1'과 w_2'이 각각 동일한 부호를 가질 때는 대칭적 주변분포에 해당된다. 이를 식 (2.5)에 대입하면 다음과 같다.

$$p_e(\kappa_3) = \frac{1}{3} + \{v_1' w_1' + v_2' w_2' + (v_1' + v_2')(w_1' + w_2')\} \tag{2.6a}$$
$$= \frac{1}{3} + (2v_1' w_1' + v_1' w_2' + v_2' w_1' + 2v_2' w_2'). \tag{2.6b}$$

동일한 방법으로 $m \times m$ 분할표의 $p_e(\kappa_m) = \sum_{i=1}^{m} (n_{i+}/n)(n_{+i}/n)$는 다음과 같이 표현된다.

$$p_e(\kappa_m) = \sum_{i=1}^{m-1} v_i w_i + \left(1 - \sum_{i=1}^{m-1} v_i\right)\left(1 - \sum_{i=1}^{m-1} w_i\right) \tag{2.7a}$$
$$= \frac{1}{m} + \sum_{i=1}^{m-1} v_i' w_i' + \left(\sum_{i=1}^{m-1} v_i'\right)\left(\sum_{i=1}^{m-1} w_i'\right) \tag{2.7b}$$
$$= \frac{1}{m} + 2\sum_{i=1}^{m-1} v_i' w_i' + \sum_{i,j=1',}^{m-1} \sum_{i \neq j}^{m-1} v_i' w_j'. \tag{2.7c}$$

κ의 첫 번째 역설과 대칭적 <균형, 불균형> 주변분포

주변분포가 대칭적이면 식 (2.6a)의 우변의 대괄호속의 각 항들은 모두 동일한 부호를 갖는 값들의 곱이므로 0보다 큰 값이 된다. 따라서 v_1', v_2', w_1', w_2'이 모두 0에 가까운 값일수록 $p_e(\kappa_3)$의 값이 작아지며, $v_1' = v_2' = w_1' = w_2' = 0$ 일 때, 즉 $v_1 = v_2 = w_1 = w_2 = 1/3$인 완전한 균형일 때 최소가 된다.

$m \times m$ 분할표에서도 주변분포가 대칭적일 때 식 (2.7b)에서 두개의 합을 구성하는 곱의 항들이 모두 동일한 부호를 가지므로, 모든 v_i'와 w_i'가 0에 가까운, 즉 균형적 주변분포일수록 $p_e(\kappa_m)$의 값이 작아진다. 따라서 주변분포가 대칭적인 경우 불균형적 주변분포일 때가 균형적 주변분포일 때 보다 더 큰 값을 $p_e(\kappa_m)$갖고, 더 작은 κ_m값을 갖는다.

κ의 두 번째 역설과 <대칭, 비대칭> 불균형 주변분포

대칭적 불균형 주변분포일 때에는, 식 (2.6a)의 우변의 대괄호속의 각 항들이 모두 동일한 부호를 갖는 값들의 곱이므로, 모두 양의 값을 갖는다. 그러나 비대칭적 불균형 주변분포일 때는 세 개의 곱의 항 중 적어도 하나는 음의 값을 갖는다. 즉, 주변분포의 불균형의 정도가 같다면($|v'_i| = |w'_i|$, $i = 1, 2$), 대칭적 주변분포와 비대칭적 주변분포는 세 개의 곱의 절대값은 동일하지만, 대칭적일 경우에는 모든 곱이 양의 값을, 비대칭적일 경우에는 적어도 하나의 곱이 음의 값을 가지게 된다. 따라서 비대칭적 불균형 주변분포 보다 대칭적 불균형 주변분포가 더 큰 $p_e(\kappa_3)$값을 갖고, 더 작은 κ_3값을 가지게 된다.

이러한 내용을 식 (2.6b)의 괄호 속 네 개의 항으로 설명하면 다음과 같다.

$$① \; 2v'_1w'_1, \quad ② \; v'_1w'_2, \quad ③ \; v'_2w'_1, \quad ④ \; 2v'_2w'_2$$

우선 ②와 ③은 불균형의 정도가 같으면 $v'_1v'_2(= w'_1w'_2)$가 되며, 대칭적 불균형 주변분포와 비대칭적 불균형 주변분포일 때 모두 동일한 값을 갖는다. 그러나 ①과 ④를 정리하면 다음과 같다.

$$\begin{aligned} 2v'_1w'_1 + 2v'_2w'_2 &= v_1'^2 + w_1'^2 - (v'_1 - w'_1)^2 + v_2'^2 + w_2'^2 - (v'_2 - w'_2)^2 \\ &= \left(v_1'^2 + v_2'^2 + w_1'^2 + w_2'^2\right) - \left\{(v'_1 - w'_1)^2 + (v'_2 - w'_2)^2\right\}. \end{aligned} \tag{2.8}$$

식 (2.8)의 첫 번째 항은 대칭적 불균형 주변분포와 비대칭적 불균형 주변분포에서 동일한 값을 갖지만, 두 번째 항이 대칭적 불균형 주변분포일 때는 0, 비대칭적 불균형 주변분포일 때는 양의 값을 갖게 된다. 따라서 비대칭적 불균형 주변분포일 때의 $p_e(\kappa_3)$가 대칭적 불균형 주변분포일 때의 $p_e(\kappa_3)$보다 작은 값을 갖고, 비대칭적 불균형 주변분포일 때 더 큰 κ_3값을 갖는다.

$m \times m$ 분할표에 대해서도 동일한 방법, 즉 주변분포의 불균형의 정도가 같다면, 식 (2.7c)의 세 번째 항의 값은 대칭적 불균형 주변분포일 때와 비대칭적 불균형 주변분포일 때가 동일하다. 그리고 두 번째 항을 정리하면 다음과 같다.

$$2\sum_{i=1}^{m-1} v'_iw'_i = \sum_{i=1}^{m-1} \left\{\left(v_i'^2 + w_i'^2\right) - (v'_i - w'_i)^2\right\}. \tag{2.9}$$

식 (2.9)에서도 불균형의 정도가 같다면, $(v_i'^2 + w_i'^2)$은 대칭적 불균형 주변분포와 비대칭적 불균형 주변분포에서 동일한 값을 갖지만, $(v'_i - w'_i)^2$은 대칭적 불균형 주변분포일 때 모두 0의 값을 가지고, 비대칭적 불균형 주변분포인 경우에는 적어도 하나 이상은 양의 값을 갖게 된다. 그러므로 대칭적 불균형 주변분포가 비대칭적 불균형 주변분포 보다 더 큰 $p_e(\kappa_m)$값을 가지고, 더 작은 κ_m값을 갖는다.

κ의 두 번째 역설과 <완전한, 불완전한> 대칭적 불균형 주변분포

대칭적 주변분포이면 v'_i와 w'_i는 동일한 부호를 갖는다. 또한 불균형의 정도가 같다는 것은 완전한 대칭적 불균형 주변분포와 불완전한 대칭적 불균형 주변분포 모두 i번째 범주의 불균형의 정도가 $(v'_i + w'_i)/2$로 동일하다는 것을 의미한다. 여기서 $v'_i = w'_i (i = 1, \ldots, m)$이면 완전한 대칭적 불균형 주변분포이고, 적어도 하나의 i번째 범주가 $v'_i \neq w'_i$이면 불완전한 대칭적 불균형 주변분포이다.

$n_{3+} = v_3n$, $n_{+3} = w_3n$으로 두면 $v_3 = (1/3) + v'_3$, $w_3 = (1/3) + w'_3$이다. 그리고 식 (2.6a)의 대괄호 속의 세 번째 항은 $v'_1 + v'_2 + v'_3 = 0$과 $w'_1 + w'_2 + w'_3 = 0$인 조건을 이용하면 $(-v'_3)(-w'_3)$가 되므로, 대괄호 속의 항들은 다음과 같이 표현할 수 있다.

$$\sum_{i=1}^{3} v'_iw'_i = \frac{1}{4}\sum_{i=1}^{3} \left\{(v'_i + w'_i)^2 - (v'_i - w'_i)^2\right\},$$

여기서 대괄호 속의 첫 번째 항 $(v_i' + w_i')^2$은 불균형의 정도가 같을 경우, 완전한 대칭적 불균형 주변분포와 불완전한 대칭적 불균형 주변분포일 때 동일하지만, 두 번째 항 $(v_i' - w_i')^2$은 완전한 대칭적 불균형 주변분포 일 때는 모든 v_i'와 w_i'가 동일하므로 0의 값을 갖고, 불완전한 대칭적 불균형 주변분포일 때는 적어도 하나의 양의 값을 갖는다. 이에 따라 완전한 대칭적 불균형 주변분포가 불완전한 대칭적 불균형 주변분포 보다 더 큰 $p_e(\kappa_3)$값을 가지고, 더 작은 κ_3값을 갖는다.

$m \times m$ 분할표에서도 식 (2.7b)의 세 번째 항을 $(-v_m')(-w_m')$로 두면, 두 번째와 세 번째 항은 $\sum_{i=1}^m v_i'w_i' = 1/4 \sum_{i=1}^m \{(v_i' + w_i')^2 - (v_i' - w_i')^2\}$으로 표현되어, 3×3 분할표에서와 동일한 방법으로 증명된다.

3. 3×3 이상 분할표에서의 합치도 H

3.1. 3×3 분할표에서의 H

비대칭성의 보정: 두 평정자 A, B의 각 범주로의 평가 비율인 n_{i+}/n와 n_{+i}/n의 평균을 구한다. 여기서 각 범주의 평가 비율$(n_{i+}/n, n_{+i}/n)$은 두 평정자가 같은 전체 개체 수(n)에서 구한 값으로서 실험에서 변하지 않는 값이므로 평균 평가 비율은 산술평균으로 구한다. 각 범주의 평균 평가 비율은 식 (3.1)과 같이 표현한다.

$$P_i = \frac{n_{i+} + n_{+i}}{2n}, \quad i = 1, 2, 3, \tag{3.1}$$

여기서 P_i은 i번째 범주에 해당하는 두 평정자의 평균 평가 비율이다.

불균형성의 보정: 각 범주에 같은 수의 개체가 분류되었다는 조건하에서 평균이 계산되면 균형성이 유지되므로, 불균형성을 보정하기 위하여 조화평균을 구하면 식 (3.2)와 같다.

$$\frac{3}{\frac{1}{P_1} + \frac{1}{P_2} + \frac{1}{P_3}} = \frac{3P_1P_2P_3}{P_1P_2 + P_2P_3 + P_1P_3}. \tag{3.2}$$

이를 이용하여 구한 기대 합치 비율은 식 (3.3)과 같다.

$$p_e(H_3) = 3\left(\frac{3P_1P_2P_3}{P_1P_2 + P_2P_3 + P_1P_3}\right)^2. \tag{3.3}$$

식 (2.2)의 p_e에 식 (3.3)을 대입하여 구한 3×3 분할표의 합치도 H_3는 다음과 같다.

$$H_3 = \frac{p_o - p_e(H_3)}{1 - p_e(H_3)}. \tag{3.4}$$

3.2. 3×3 분할표에서의 합치도 H의 최대우도 추정량과 분산

식 (3.4)에 p_{ij}의 최대우도추정량 $\widehat{p_{ij}} = n_{ij}/n$을 대입하여 얻은 3×3 분할표에서의 합치도 H_3의 최대우도추정량은 식 (3.5)와 같다.

$$\hat{H}_3 = \frac{4n(n_{11} + n_{22} + n_{33}) - 27A^2}{4n^2 - 27A^2}, \tag{3.5}$$

여기서

$$A = \frac{(n_{1+} + n_{+1})(n_{2+} + n_{+2})(n_{3+} + n_{+3})}{(n_{1+} + n_{+1})(n_{2+} + n_{+2}) + (n_{2+} + n_{+2})(n_{3+} + n_{+3}) + (n_{1+} + n_{+1})(n_{3+} + n_{+3})}.$$

3×3 분할표의 합치도 H_3의 점근적 대표본 분산은, 관찰된 합치 빈도 $\sum n_{ii}$가 성공의 확률 p_o인 이항확률실험을 n번 시행했을 때, 성공의 횟수를 나타내는 이항확률변수라는 사실을 이용하여 다음과 같이 얻어진다.

$$\begin{aligned}\mathrm{Var}(H_3) = \mathrm{Var}\left\{\frac{p_o - p_e(H_3)}{1 - p_e(H_3)}\right\} &= \frac{1}{\{1 - p_e(H_3)\}^2}\mathrm{Var}(p_o) \\ &= \frac{1}{\{1 - p_e(H_3)\}^2}\frac{p_o(1 - p_o)}{n}. \end{aligned} \tag{3.6}$$

분산의 최대우도 추정량은 식 (3.6)에 식 (3.1)–(3.3)에서 정의된 P_i를 이용하여 구한 $p_e(H_3)$의 최대우도 추정량과 식 (3.5)를 대입하면 얻을 수 있다.

3.3. $m \times m$ 분할표에서의 합치도 H

$m \times m$ 분할표에서의 A와 B 두 평정자가 개체들을 같은 범주로 분류한 비율인 관찰된 합치 비율 p_o는 식 (3.7)과 같다.

$$p_o = \frac{\sum_{i=1}^m n_{ii}}{n}. \tag{3.7}$$

m개의 각 범주에 해당하는 두 평정자의 평균 평가 비율 $P_i(i = 1, \ldots, m)$는 다음과 같다.

$$P_i = \frac{\frac{n_{i+}}{n} + \frac{n_{+i}}{n}}{2} = \frac{n_{i+} + n_{+i}}{2n}. \tag{3.8}$$

이때 기대 합치도는 m개 P_i의 조화평균인 식 (3.9)와 같다.

$$p_e(H_m) = m^3\left(\frac{1}{\sum_{i=1}^m \frac{1}{P_i}}\right)^2. \tag{3.9}$$

H_m의 최대우도 추정량은 식 (3.10)과 같으며, 점근적 대표본 분산은 식 (3.6)과 동일한 모양으로 표현된다.

$$\hat{H}_m = \frac{4n\sum_{i=1}^m n_{ii} - m^3\left(\frac{1}{\sum_{i=1}^m \frac{1}{n_{i+}+n_{+i}}}\right)^2}{4n^2 - m^3\left(\frac{1}{\sum_{i=1}^m \frac{1}{n_{i+}+n_{+i}}}\right)^2}. \tag{3.10}$$

4. 다른 합치도들과의 비교

2.3절에서 제시되었던 κ의 두 가지 역설에 대한 예제 표 3과 4에서 구한 관찰된 합치 비율 p_o와 각 합치도들의 우연에 기대되는 합치 비율 p_e, 합치도 값을 표 5에 제시하였다.

κ에 대한 첫 번째 역설부분인 표 3에서, (a)와 같이 주변확률이 0.33으로 대칭적이고 균형적인 경우에는 모든 합치도의 값이 0.85로 동일하다. 그러나 (b)와 같이 주변확률이 대칭적이고 불균형적인 경우에는 (a)와 동일한 $p_o(= 0.90)$임에도 불구하고 p_o에 비하여 κ, π는 매우 작게 나타난다. 이에 비해 AC_1과 H는 p_e값이 작게 나타나서 p_o에 가까운 값을 가지며, (a)와 (b)간의 합치도 차이가 크게 나지 않는다.

표 5: κ의 역설에 대한 예제에서의 각 합치도들의 p_e와 합치도 값

표	p_o	p_e					κ	G	π	AC_1	H
		κ	G	π	AC_1	H					
3. (a)	0.90	0.33	1/3	0.33	0.33	**0.33**	0.85	0.85	0.85	0.85	**0.85**
(b)		0.80		0.80	0.10	**0.02**	0.51		0.50	0.89	**0.90**
4. (a)	0.60	0.46	1/3	0.46	0.27	**0.13**	0.26	0.40	0.26	0.45	**0.54**
(b)		0.29		0.34	0.33	**0.32**	0.44		0.40	0.40	**0.41**
(c)	0.80	0.72	1/3	0.72	0.14	**0.02**	0.28	0.70	0.28	0.77	**0.80**
(d)		0.71		0.72	0.14	**0.02**	0.31		0.28	0.77	**0.80**

두 번째 역설부분인 표 4에서는 동일한 $p_o(=0.60)$값을 갖지만, (b)와 같이 비대칭적 불균형 분포인 경우보다 (a)와 같은 대칭적 불균형 분포일 때, 합치도 값이 κ와 π는 상대적으로 더 작게 나타나지만, AC_1과 H는 다소 높게 나타나 반대의 결과를 보인다. (a)의 경우 H는 0.54로 p_o에 가장 가까운 값을 가진다. 또한 (c)의 완전한 대칭적 불균형 주변분포와 (d)의 불완전한 대칭적 불균형 주변분포에서, κ는 합치도의 값이 κ의 두 번째 역설에 따라 차이를 보이고 있지만, AC_1와 H는 역설과는 달리 합치도의 값이 차이가 없으며 p_o에 근접한 값을 가진다. 특히 H는 합치도의 값이 0.80으로서 p_o와 동일한 값을 가진다. 같은 크기의 범주에서 주변분포와 무관하게 동일한 p_e의 값을 갖는 G는 p_o값에 비해 상대적으로 AC_1과 H와 비교하여 더 작은 값을 갖는다.

5. 결론

합치도로 널리 사용되는 Cohen의 κ가 주변분포에 민감하다는 단점을 보완하기 위해 박미희와 박용규 (2007)는 2×2 분할표에서 주변분포의 비대칭성과 불균형성을 보정한 합치도 H를 제안하였다. 본 연구에서는 이를 일반화하여 3×3 분할표를 비롯해 $m\times m$ 분할표에서의 합치도 H와 최대우도추정량, 분산을 제시하였다. 또한 Feinstein과 Cicchetti (1990)가 제기한 2×2 분할표에서의 κ의 역설을, 3×3 이상의 분할표에 적용하여 κ의 역설과 주변분포와의 관계를 증명하였다.

3×3 분할표의 예제를 통하여 κ를 비롯한 다른 합치도들과 H를 비교한 결과, κ와 π는 합치정도가 높게 예상되는 주변분포를 가진 예제에서 낮은 합치정도를 보였으나, AC_1과 H는 상대적으로 더 큰 합치도의 값을 보였다. 또한 H는 κ, π, G, AC_1 보다 관찰된 합치 비율(p_o)에 가까운 값을 가져, 주변분포의 불균형성과 대칭성으로 인해 나타나는 κ의 문제점을 해결할 수 있는 측도임을 확인할 수 있었다. 향후 연구에서는 기존에 제기된 κ의 역설이 주로 대칭적 주변분포를 대상으로 했다는 점을 감안하여, 비대칭적 주변분포들의 비교, 즉 비대칭적 균형과 불균형, 완전한 비대칭적 불균형과 불완전한 비대칭적 불균형 주변분포들의 비교를 통해, κ의 주변분포에 대한 민감성을 추가로 밝히고, H를 적용하여 해결이 가능한지와 확률 시뮬레이션을 통해 H의 검정력 및 표본수 결정에 대한 내용을 다루어 볼 필요가 있다.

참고 문헌

박미희, 박용규 (2007). COHEN의 합치도의 두 가지 역설을 해결하기 위한 새로운 합치도의 제안, <응용통계연구>, **20**, 117–132.

Bennet, E. M. and Alpert, R. and Goldstein, A. C. (1954). Communications through limited response questioning, *Public Opinion Quarterly*, **18**, 303–308.

Cicchetti, D. V., Lee, C., Fontana, A. F. and Dowds, B. N. (1978). A computer program for assessing specific category rater agreement for qualitative data, *Educational and Psychological Measurement*, **38**, 805–813.

Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, **20**, 37–46.

Feinstein, A. R. and Cicchetti, D. V. (1990). High agreement but low kappa: 1. The problems of two paradoxes, *Journal of Clinical Epidemiology*, **43**, 543–549.

Gwet, K. (2001). Handbook of inter-rater reliability. STATAXIS Publishing company, Gaithersburg.

Holley, J. W. and Guilford, J. P. (1964). A note on the G index of agreement, *Educational and Psychological Measurement*, **24**, 749–753.

Janson, S. and Vegelius, J. (1979). On generalizations of the G index and the phi coefficient to nominal scales, *Multivariate Behavioral Research*, **14**, 255–269.

Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding, *Public Opinion Quarterly*, **19**, 321–325.

2009년 6월 접수; 2009년 8월 채택

Measure of Agreement H in $m \times m$ Contingency Table

Jin-Gon Kim[a], Mi-Hee Park[a], Yong-Gyu Park[1,a]

[a]Department of Biostatistics, The Catholic University of Korea

Abstract

A measure of agreement H in 2×2 contingency table was proposed by Park and Park (2007) to resolve the two paradoxes of κ. In this study, we generalize H to where the number of categories is greater than two and derive its asymptotic large-sample variance. We also explain the relationships between κ's paradoxes and marginal distributions. Using some examples of 3×3 contingency tables, the behaviors of H and other measures of agreement are compared.

Keywords: Measure of agreement H, paradoxes of κ, marginal distributions.

[1] Corresponding Author: Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea.
E-mail: ygpark@catholic.ac.kr