Journal of The Korea Society of Computer and Information
Vol. 20 No. 12, pp. 169-174, December 2015

www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2015.20.12.169

Permutation P-values for Inter-rater Agreement Measures

Yonghwan Um*

Abstract

Permutation p-values are provided for the agreement measures for multivariate interval data among many raters. Three agreement measures, Berry and Mielke's measure, Janson and Olsson's measure, and Um's measure are described and compared. Exact and resampling permutation methods are utilized to compute p-values and empirical quantile limits for three measures. Comparisons of p-values demonstrate that resampling permutation methods provide close approximations to exact p-values, and Berry and Mielke's measure and Um's measure show similar performance in terms of measuring agreement.

▸ Keyword : Permutation Method, agreement measure, Multivariate data

Ⅰ. Introduction

동일한 대상에 대해 두 명 또는 그 이상의 평가자들이 평가할 때 이들 사이의 평가가 얼마나 일치하는지를 나타내는 것이 일치도(measure of agreement)이며 이 개념은 의학, 정보기술 뿐만 아니라 교육학, 사회학, 심리학 등의 사회과학 분야에서 연구되어 온 중요한 통계적 관심사이다. 일치도는 평가자가 평가 대상을 동일한 범주에 분류하는 정도를 나타내고 이와 비슷한 척도인 연관성은 평가자들의 평가 결과에 대한 연관 정도를 나타내는 것으로서 일치도는 연관성의 특별한 케이스라 할 수 있다. 연관성을 나타내는 통계치로는 감마계수, Yule의 Q, Kendall의 tau, Φ 계수, 피어슨 상관계수, Intraclass 상관계수 등이 있으며, 가장 널리 사용되는 일치도는 Cohen이 제안한 k(카파)로서 명목

척도의 데이터에 대해 두 평가자 사이의 일

치 정도를 우연히 일치할(chance - corrected) 확률을 제거하여 측정한다[1]. 제안된 일치도가 보편적인 일치도로서 널리 사용되기 위해 필요한 특성은 높은 척도 수준의 다변량 데이터와 여러 평가자로의 확장 가능성이다. 평가자의 수가 여러 명으로 확장되고 순서 또는 거리척도를 갖는 다차원 데이터에 대해 정의될 수 있는 일치도는 그 사용도가 더욱 커질 것이다. 그동안 많은 연구자들이 Cohen의 k를 확대하는 시도를 하였는데, 예를 들어 여러 평가자들과 어느 한 특정 평가자 사이의 결합 일치도를 제안한 연구와 여러 평가자들 사이의 일치도를 모든 두 평가들 사이의 k의 평균으로부터 얻은 연구[2][3] 그리고 명목척도의 데이터에 대해서 Cohen의 k를 여러 평가자로 확장할 때의 문제점을 제기한 연구 등 다양하게 연구가 진행되었다[4][5]. 또한 Berry와 Mielke, Janson과 Olsson 그리고 엄용환은 각각 거리척도의 다변량 데이터에 대해 여러 평가자들 사이의 일치도를 제안하였다[6][7][8]. Berry와 Mielke는 두 평가자 사이의 일치정도를 나타내는 유클리디안 거리에 기초하여 거리척도의 다변량 데이터에 대해 여러 평가자 사이의 일치도를 제시하였고 Janson과 Olsson은 Berry와 Mielke의 일치도에서 유클리디안 거리 대신 제곱한 유클리디안 거리 (squared Euclidean distance)을 사용한 일치도를 제안하였다. 그리고 엄용환이 제안한 일치도는 데이터 포인트들에 의해 형성되는 다차원 심플렉스의 부피에 기초한 것이다.

본 연구에서는 퍼뮤테이션 검정에(permutation test) 의해 Berry와 Mielke, Janson과 Olsson 그리고 엄용환이 제안한 일치도에 대해 p값과 경험적인 분위수 한계(empirical quantile limit)를 산출하고자 한다. 퍼뮤테이션 검정은 Fisher가 최초로 제안한 이후로 꾸준히 발전해 왔는데[9], 이 검정은 분석을 위한 모든 정보가 관찰 데이터에 포함되어 있기 때문에 데이터에 의존하는(data-dependent) 검정이고 모

• First Author : Yonghwan Um , Corresponding Author: Yonghwan Um
*Yonghwan Um(uyh@sungkyul.ac.kr), Division of Industrial and Management Engineering, Sungkyul University
• Received: 2015. 11. 30, Revised: 2015. 12. 08, Accepted: 2015. 12. 16.

집단의 분포에 대한 어떤 가정도 요구하지 않는 특징이 있다[10][11]. 특히 표본의 크기가 작을 경우, 퍼뮤테이션 검정은 큰 표본에 대해 사용되는 전통적인 근사 검정(asymptotic test)보다 더 정확한 p값을 제공한다. 더불어 퍼뮤테이션 기법은 전통적인 신뢰구간과 유사한 경험적인 분위수 한계를 제공한다.

본 논문의 제 2장에서는 다변량 거리척도 데이터에 대해 사용되는 평가자간 일치도를 간략히 소개하고 제 3장에서는 퍼뮤테이션 기법을 살펴보고자 한다. 제 4장에서는 실제 다변량 데이터를 이용하여 퍼뮤테이션에 의해 얻은 p값과 경험적인 분위수 한계 (empirical quantile limit)를 제시한다.

II. Inter-rater Agreement Measures for Multivariate Data

Berry와 Mielke는 평가자들을 블락으로 생각하는 랜덤화 블락계획법을 사용함으로써 b명의 평가자들이 n명의 대상을 평가할 때 평가자간의 일치도를 정의하였고 그 식은 다음과 같다.

$$R = 1 - \frac{\delta}{\mu_{\delta}} \tag{1}$$

여기서 δ는 관찰된 평가자간의 불일치 비율이고, μ_{δ} 는 기대되는 평가자간의 불일치 비율을 의미하며 다음의 수식에 의해 주어진다.

$$\delta = \left[n\binom{b}{2} \right]^{-1} \cdot \sum_{i=1}^{n} \sum_{s<t} \triangle(\mathbf{x}_{si}, \mathbf{x}_{ti})$$

$$\mu_{\delta} = \left[n^2\binom{b}{2} \right]^{-1} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{s<t} \triangle(\mathbf{x}_{si}, \mathbf{x}_{tj}).$$

여기서 $\triangle(\mathbf{x_{si}}, \mathbf{x_{tj}})$는 평가자 s와 평가자 t의 평가결과가 c차원 데이터로 주어질 때 두 평가자들 사이의 불일치 정도를 유클리디안 거리, 즉

$$\left[\sum_{k=1}^{c} (x_{sik} - x_{tjk})^2 \right]^{1/2}$$ 로 계산된다.

이 식에서 Berry와 Mielke는 일치도를 계산하기 위해 우선 b명의 평가자 중에서 얻을 수 있는 총 $\binom{b}{2}$개의 두 평가자간 일치도들을 구한 후 이들의 평균을 취하여 전체 평가자들 사이의 일치도를 산출하였다. 이 일치도 R 은 우연히 일치할 (chance-corrected) 확률을 보정한 것이고 거리척도의 다변량 데이터와 여러명의 평가자의 경우에 적용될 수 있다. 한편 Janson과 Olsson 그리고 엄용환은 각각 새로운 일치도를 제안하였는데 이들은 기본적으로 Berry와 Mielke의 방법을 따르고 있다. 다만 Janson과 Olsson은 두 평가자들 사이의 불일치 정도를 나타내는 값으로 유클리디안 거리 대신 제곱한 유클리디언 거리를 사용한 것이 다르다 하겠다. 즉

$$\triangle(\mathbf{x}_{si}, \mathbf{x}_{tj}) = \sum_{k=1}^{c} (x_{sik} - x_{tjk})^2$$

을 사용하였다. 그리고 엄용환은 일치도 R 을 구하기 위해 δ과 μ_{δ}를 다음의 새로운 방법을 제안한 바 있다[8].

$$\delta = \left[n\binom{b}{c+1} \right]^{-1} \cdot \sum_{i=1}^{n} \sum_{1 \le s_1 < \cdots < s_{c+1} \le b} \triangle(\mathbf{x}_{s_1 i}, \mathbf{x}_{s_2 i}, \ldots, \mathbf{x}_{s_{c+1} i})$$

$$\mu_{\delta} = \left[n^{c+1} \binom{b}{c+1} \right]^{-1} \cdot \sum_{i_1=1}^{n} \cdots \sum_{i_{c+1}=1}^{n} \sum_{1 \le s_1 < \cdots < s_{c+1} \le b} \triangle(\mathbf{x}_{s_1 i_1}, \mathbf{x}_{s_2 i_2}, \ldots, \mathbf{x}_{s_{c+1} i_{c+1}})$$

여기서 $\triangle(\mathbf{x}_{s_1 i_1}, \mathbf{x}_{s_2 i_2}, \ldots, \mathbf{x}_{s_{c+1} i_{c+1}})$는 c변량 데이터 $\mathbf{x}_{s_1 i_1}, \mathbf{x}_{s_2 i_2}$, , $\mathbf{x}_{s_{c+1} i_{c+1}}$ 로 구성되는 심플렉스의 부피이고 이 값은 $(c+1) \times (c+1)$ 데이터 행렬의 디터미넌트(determinent)로 주어진다. 예를 들어 이변량 데이터의 경우 $\triangle(\mathbf{x}_{s_1 i_1}, \mathbf{x}_{s_2 i_2}, \mathbf{x}_{s_3 i_3})$은 3 ×3 데이터 행렬로부터 다음과 같이 계산 된다.

$$\triangle(\mathbf{x}_{s_1 i_1}, \mathbf{x}_{s_2 i_2}, \mathbf{x}_{s_3 i_3}) = \frac{1}{2!} \mathrm{abs}\left(\begin{vmatrix} 1 & 1 & 1 \\ x_{s_1 i_1 1} & x_{s_2 i_2 1} & x_{s_3 i_3 1} \\ x_{s_1 i_1 2} & x_{s_2 i_2 2} & x_{s_3 i_3 2} \end{vmatrix} \right)$$

Berry와 Mielke의 일치도, Janson과 Olsson의 일치도, 그리고 엄용환의 일치도는 모두 0과 1 사이의 값을 가지며, 1에 가까울수록 평가자간 일치도가 큼을 뜻한다.

III. Permutation Test

퍼뮤테이션 검정은 관찰 데이터의 모든 가능한 배열에 기초하여 이루어진다. 예를 들어 두 집단(크기가 각각 m_1, m_2)의 평균을 비교하는 2-표본 검정의 경우, 귀무가설 하에서 관측값들이 동일한 모집단에서 나온 것으로 볼 수 있기 때문에 관측값들은 표본들 사이에서 서로 상호 교환이 가능하며 관측값들이 어느 집단에 배치되느냐는 완전 무작위라고 생각할 수 있다. 따라서 관찰 데이터의 모든 가능한 배열의 수는

$\frac{(m_1+m_2)!}{(m_1!)(m_2!)}$이 된다.

퍼뮤테이션 검정은 정확한 검정(exact test)과 재표본 검정(resampling test)으로 나뉜다. 정확한 검정에서는 관찰값들로부터 얻을 수 있는 모든 가능한 배열들을 생성하고 이 배열 하나 하나에 대해서 검정통계량을 계산한다. 이 때 p값은 검정통계량들의 모든 값들 중에서 실제 관찰값에 대한 검정통계량의 값과 같거나 보다 더 극단적인 값을 갖는 검정통계량의 비율로 정의된다. 그러나 표본의 크기가 증가함에 따라 고려해야 할 배열의 수가 크게 증가하기 때문에 전체로부터 일부 배열만을 무작위 추출하여 p값을 구하는데 이것을 재표본 검정이라 한다. 이 때 추출된 L개의 배열 중에서 실제 관찰값에 대한 검정통계량의 값과 같거나 보다 더 극단적인 값을 갖는 검정통계량의 비율이 재표본 기법에 의한 p값이 된다. 이 때 사용하는 L값으로 5,000~10,000이면 충분하다고 주장하는 학자들도 있지만[12], 많은 학자들은 p값이 작을 경우에는 보다 정확한 추론을 위해 보통 큰 값의 L을 사용하는 것을 제안한다. (예를 들어 L = 1,000,000)[13].

본 논문에서는 b명의 평가자가 n명의 대상을 평가하여 얻은 데이터에 대한 퍼뮤테이션을 고려한다. 이 때 b명의 평가자를 b개의 블록으로 간주하고 각 블록 안에서 랜덤화가 이루어지며, 얻을 수 있는 총 가능한 배열의 수는 귀무가설(H_0) 하에서 $M = (n!)^b$ 이 된다. 이 때 H_0은 각 배열이 1/M의 동일한 확률로 발생한다는 가설이다. 예를 들어 n=8, b=4 일 때 고려해야 할 배열의 수는 모두 $M = (8!)^4 = 2.64 \times 10^{18}$ 이다. 그런데 이것은 매우 큰 값이므로 정확한 p값 계산이 가능하지 않기 때문에 근사적으로 p값을 산출하기 위해 재표본 검정을 사용한다.

식 (1)에서 일치도 R과 δ은 서로 선형적인 관계가 있으므로 R에 대한 검정은 곧 δ에 대한 검정과 같다고 볼 수 있다. 그러므로 H_0 하에서 p값은 다음과 같이 주어진다.

$P\left(R \geq 1 - \frac{\delta_0}{\mu_\delta} | H_0\right) = P(\delta \leq \delta_0 | H_0)$, 여기서

δ_0은 관찰된 δ이다.

따라서 H_0 하에서 정확한 검정에 의한 p값은

$P(\delta \leq \delta_0 | H_0) = \frac{1}{M}\sum_{i=1}^{M} \Psi_i(\delta)$ 이며,

여기서 $\Psi_i(\delta) = \begin{cases} 1 & \delta \leq \delta_0 \text{일 때} \\ 0 & \text{아닐 때} \end{cases}$

이다. 그리고 재표본 검정에 의한 p값은

$P(\delta \leq \delta_0 | H_0) = \frac{1}{L}\sum_{i=1}^{L} \Psi_i(\delta)$ 이 된다.

퍼뮤테이션 기법은 모집단 분포에 의존하지 않기 때문에 전통적인 방법으로 $1-\alpha$

신뢰구간을 얻을 수 없다. 그러나 검정 통계량의 퍼뮤테이션 분포로부터 전통적인 신뢰구간과 유사한 $1-\alpha$ 경험적 분위수 한계(empirical quantile limit)를 구할 수 있다[14]. 경험적인 q번째 분위수 (q^{th} empirical quantile)란 L개의 δ가 w_q보다 크거나 같을 확률이 최소한 q이고 w_q 보다 작거나 같을 확률이 최소한 1-q 이 됨을 만족하는 w_q를 일컫는다. 경험적 분위수 한계를 구하려면 먼저 재표본 기법을 사용하여 얻은 L개의 δ를 작은 값에서 큰값으로 정렬한 후 하한값($Q_{\alpha/2}$)과 상한값($Q_{1-\alpha/2}$)을 구한다. 이 $Q_{\alpha/2}$과 $Q_{1-\alpha/2}$은 각각 $\delta_1, \delta_2, , , \delta_L$ 에 대응되는 순서통계량을 $W_1 \leq W_2 \leq \cdots \leq W_L$ 이라 할 때 다음과 같이 주어진다.

$$Q_{\alpha/2} = W_{\text{최대}[1, \text{정수}[(\alpha/2)L + 0.5]]}$$

$$Q_{1-\alpha/2} = W_{\text{최소}[L, \text{정수}((1-\alpha/2)L + 0.5)]}$$

여기서 정수[$(\alpha/2)L+0.5$]은 괄호([ ])안의 계산결과에서 정수만을 취한다는 의미이다.

IV. Example

1. Example1

퍼뮤테이션 기법에 의해 Berry와 Mielke, Janson과 Olsson 그리고 엄용환의 일치도를 비교하기 위해 Janson과 Olsson이 사용한 가상적인 거리척도의 이변량 데이터를 이용하였다(표 1). 이 데이터는 3명의 평가자가(b=3) 사진을 보고 5명의 사진 속 인물에(n=5) 대해 몸무게와 키(c=2)를 평가한 것이다. 이로부터 생각할 수 있는 모든 퍼뮤테이션에 의한 배열의 수는 $M = (5!)^3 = 1{,}728{,}000$ 이 되고 이 수는 큰 값이 아니므로 정확한 검정이 가능하며 L=1,000,000을 이용한 재표본 검정도 함께 실시하였다. 표 2는 표 1의 데이터에 대해 관찰된 평가자간의 불일치 비율 δ와 정확한 p값, 재표본에 의한 p값을 구한 결과이다. 각각의 벙법에서 정확한 검정과 재표본 검정 모두 비슷한 p값을 제공하고 있다. Berry와 Mielke의 경우 δ_0=8.768, 정확한 p값 = 0.05604, 재표본에 의한 p값 = 0.05603이고, Janson과 Olsson의 경우 δ_0=48.20, 정확한 p값 = 0.00006944, 재표본에 의한 p값 = 0.000067이며, 엄용

환의 경우 δ_0=58.60, 정확한 p값 = 0.05931, 재표본에 의한 p값 = 0.05917 이다. 이 결과 Berry와 Mielke 그리고 엄용환이 제시한 통계량은 매우 유사한 크기의 p값을 나타낸 반면 Janson과 Olsson의 통계량은 현저히 작은 p값을 제시하였다. 이러한 현상은 엄용환이 발표한 논문에서 Berry와 Mielke의 일치도와 엄용환의 일치도가 매우 유사하게 기능하는 반면에 Janson과 Olsson의 통계량은 평가자간 일치도를 과대하게 부풀리는 특징이 있음을 지적한 결과와 일치한다고 말할 수 있다. 한편 표 3은 재표본 기법에 의한 $1-\alpha$ 경험적 분위수 한계를 제시하고 있는데 0.95 경험적 분위수 한계는 각각 $Q_{0.025}$=8.2와 $Q_{0.975}$=13.4 (Berry와 Mielke의 경우), $Q_{0.025}$=105.13와 $Q_{0.975}$=262.6 (Janson과 Olsson의 경우) 그리고$Q_{0.025}$=46.4와 $Q_{0.975}$=202.2 (엄용환의 경우) 이며. 0.99 경험적 분위수 한계는 각각 $Q_{0.025}$=7.133와 $Q_{0.975}$=13.8 (Berry와 Mielke의 경우), $Q_{0.025}$=83.4와 $Q_{0.975}$=271.933 (Janson과 Olsson의 경우) 그리고$Q_{0.025}$=30.8와 $Q_{0.975}$=228.2 (엄용환의 경우) 이다. 이처럼 $1-\alpha$ 경험적 분위수 한계가 넓게 나오는 것은 표본의 크기가 작기 때문이다 (n=5).

Table 1. Three observers' ratings of weight and height

object	rater 1		rater 2		rater 3	
	weight	height	weight	height	weight	height
1	71	166	76	171	74	171
2	73	160	80	170	80	165
3	86	187	93	174	101	185
4	59	161	66	163	62	162
5	71	172	77	182	83	181

Table 2. Permutation results for example 1

measures	δ_0	exact p-value	p-value by resampling
Berry & Mielke	8.768	0.05604	0.05603
Janson & Olsson	48.20	0.00006944	0.000067
Um	58.60	0.05931	0.05917

Table 3. $1-\alpha$ empirical quantile limits for example 1

measures	$1-\alpha$ empirical quantile limits	
	$(Q_{0.025}, Q_{0.975})$	$(Q_{0.005}, Q_{0.995})$
Berry & Mielke	(8.2, 13.4)	(7.133, 13.8)
Janson & Olsson	(105.13, 262.6)	(83.4, 271.933)
Um	(46.4, 202.2)	(30.8, 228.2)

2. Example 2

표 4는 4명의 평가자가(b=4) 5명의 초등 학생들(n=5)의 성격(사교성(S), 창의성(C), 적극성(P), 즉 c=3)을 최저 1점에서 최고 10점의 리커트 형태의 척도로 평가한 가상 데이터이다. 퍼뮤테이션에 의한 모든 배열의 수는 $M=(5!)^4$ = 207,360,000 이므로 정확한 검정에 의한 p값 계산은 가능하지 않으므로 L=1,000,000 의 재표본 검정을 이용하였다. 표 5는 재표본 기법에 의해 얻은 p값과 $1-\alpha$ 경험적 분위수 한계이다. Berry와 Mielke의 경우 δ_0=1.001, p값 =0.001028, Janson과 Olsson의 경우 δ_0=1.511, p값= 0.000079, 엄용환이 제안한 일치도에서는 δ_0=0.8, p값 = 0.006449 이다. 예제 데이터1에서와 마찬가지로 Janson과 Olsson의 통계량에서는 매우 작은 p값을 보였다. 그리고 0.95 경험적 분위수 한계는 각각 $Q_{0.025}$=1.1556와 $Q_{0.975}$=1.5778 (Berry와 Mielke의 경우), $Q_{0.025}$=2.4와 $Q_{0.975}$= 4.0889 (Janson과 Olsson의

경우) 그리고$Q_{0.025}$=1.4와 $Q_{0.975}$=9.6 (엄용환의 경우) 이며 0.99 경험적 분위수 한계는 각각 $Q_{0.025}$=1.0889와 $Q_{0.975}$=1.6 (Berry와 Mielke의 경우), $Q_{0.025}$=2.0667와 $Q_{0.975}$=4.2 (Janson과 Olsson의 경우) 그리고$Q_{0.025}$=0.6와 $Q_{0.975}$=11.6 (엄용환의 경우) 이다.

Table 4. Four observers' rating of sociability, creativity and positiveness

object	rater 1			rater 2			rater 3			rater 4		
	S	C	P	S	C	P	S	C	P	S	C	P
1	9	8	7	8	7	6	8	7	6	7	8	9
2	10	8	9	9	8	7	9	7	8	8	8	6
3	7	6	8	6	5	7	7	6	8	6	7	8
4	6	8	7	5	7	6	4	6	8	5	6	7
5	5	7	8	4	7	8	6	8	9	4	6	7

Table 5. Permutation results for example 2

measures	δ_0	p-value by resampling	$1-\alpha$ empirical quantile limits	
			$(Q_{0.025}, Q_{0.975})$	$(Q_{0.005}, Q_{0.995})$
Berry & Mielke	1.001	0.001028	(1.1556, 1.5778)	(1.0889, 1.6)
Janson & Olsson	1.511	0.000079	(2.4, 4.0889)	(2.0667, 4.2)
엄용환	0.8	0.006449	(1.4, 9.6)	(0.6, 11.6)

V. Conclusion

본 논문은 여러 명의 평가자가 평가 대상을 평가한 결과가 거리척도의 다변량 데이터로 주어질 때 이 평가자들 사이의 일치도를 퍼뮤테이션 검정에 의해 비교한 것이다.

퍼뮤테이션 검정은 전통적인 모수적 검정에 비해 몇 가지 장점을 갖고 있다. 첫째, 퍼뮤테이션 검정은 데이터에 의존하는데 이것은 분석에 필요한 모든 정보가 관측 데이터에 있다는 것이며, 둘째로 모집단에 대한 어떤 이론적인 분포를 가정하지 않는다는 것이다. 셋째로 퍼뮤테이션 검정은 정규성, 등분산성 같은 전통적인 모수적 검정에서 주로 요구하는 가정을 필요로 하지 않으며, 넷째로 퍼뮤테이션 검정은 근사 분포가 아닌 이산형 퍼뮤테이션 분포에 근거한 p값을 제시한다. 끝으로 모집단에서 추출한 임의 표본 뿐 아니라 비임의 표본에 대해서도 퍼뮤테이션 검정은 사용될 수 있다

본 논문은 일치도 비교를 위해 Berry와 Mielke, Janson과 Olsson 그리고 엄용환이 제안한 일치도 통계량을 사용하였다. 퍼뮤테이션 검정은 정확한 검정과 재표본에 의한 근사 검정으로 나누어 실시하였으며 퍼뮤테이션 배열의 수가 클 경우에는 재표본에 의한 근사 검정을 실시하였다. 일치도 비교를 위해 예제 데이터를 사용하여 퍼뮤테이션 검정을 통해 δ에 대한 p값과 $1-\alpha$ 경험적 분위수 한계를 구하였으며 그 결과를 다음과 같이 요약할 수 있다.

● 정확한 검정에 의한 p값과 재표본의 근사 검정에 의한 p값의 크기가 유사하게 나타났다. 따라서 퍼뮤테이션 배열의 수가 클 때는 재표본 기법에 의한 근사적인 p값이 정확한 검정에 의한 p값을 대신할 수 있으며 재표본 기법에 의한 근사 검정은 표본 크기가 작거나 클 때에 모두 사용될 수 있다. 특히 작은 크기의 표본에서의 퍼뮤테이션 겁정은 표본 크기가 크다는 가정하에 전통적으로 실시하는 근사 검정보다 정밀한 결과를 제공할 수 있다.

● Janson과 Olsson의 통계량에 대한 p값이 Berry와 Mielke의 통계량과 엄용환의 통계량에 대한 p값과 크게 다른 이유는 Janson과 Olsson의 통계량이 Berry와 Mielke의 통계량 그리고 엄용환의 통계량과 다르게 일치도를 과대하게 부풀리기 때문이다.

REFERENCES

[1] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement, Vol. 20, pp. 37-46, 1960.

[2] G. W. Willam, Comparing the joint agreement of several raters with another rater, Biometrics, Vol. 32, pp. 619-627, 1976.

[3] R. J. Light, Measures of response agreement for aualitative data: some generalizations and alternatives, Psychological Bulletin,Vol. 76, pp. 365-377, 1971.

[4] L. Hubert, Kappa revisited, Psychological Bulletin, Vol. 36, pp. 207-216, 1983.

[5] A. J. Cogner, Integration and generalization of kappas for multiple raters, Psychological Bulletin, Vol. 88, pp. 322-328, 1980.

[6] K. J. Berry, and P. W. Mielke Jr. A generalization of Cohen's kappa agreement measure to interval measurement and multiple raters. Educational and Psychological Measurement, Vol. 48, pp. 921-933, 1988.

[7] H. Janson, and U. Olsson, A measure of agreement for interva or nominal multivariate observations, Educational and Psychological Measurement, Vol. 61, No. 2,pp. 277-289. 2001.

[8] Y. H. Um, A new agreement measure for interval multivariate observations, Journal of Korean Data & Information Science Society, Vol. 15, pp. 263-271, 2004.

[9] E. J. G. Pitman, Significance tests which may be applied to sample from any populations, III. The analysis of variance test. Biometrika, Vol. 29, pp. 322-335, 1938.

[10] A. F. Hayes, Permustat: randomization tests for the Machintosh, Behavior Research Methods, Instruments, & Computers, Vol 28, pp. 473-475, 1996.

[11] R. S. Chen and W. P. Dunlap, SAS procedures for approximate randomizatio tests, Behavior Research Methods, Instruments, & Computers, Vol. 25, pp. 406-409, 1993.

[12] P. S. Maxim, Quantative research Methods in the Social Sciences. New York: Oxford University Press, 1999.

[13] J. E. Johnston, K. J. Berry, and P. W. Mielke, Permutation tests: precision in estimating probability values., Perceptual and Motor Skills, Vol. 105, pp. 915-920, 2007.

[14] P. W. Mielke Jr. and K. J. Berry, Permutation methods: a distance function approach. (2nd ed.) New York: Springer-Verlag, 2007.

Authors

Yonghwan Um received the B.S. and M.S. in Chemistry M.S. from Yonsei University, Korea, in 1981, 1983, M.S. in Biostatistics from Emory University, in 1990 and Ph.D. in Statistics from University of Florida, U.S.A. in 1995.

Dr. Um joined the faculty of the Department of Computational Statistics at Sungkyul University, Anyang, Korea, in 1996. He is currently a Professor in the Division of Industrial and Management Engineering, Sungkyul University. He is interested in reliability measure, data-mining, statistical inference.