Journal of The Korea Society of Computer and Information
Vol. 22 No. 7, pp. 133-139, July 2017
www.ksci.re.kr
https://doi.org/10.9708/jksci.2017.22.07.133

Permutation Analysis of Split-Half Reliability Coefficient

Yonghwan Um*

Abstract

In this paper, we describe a permutation procedure in which we compute a resampling probability value and empirical quantile limits for Split-Half measure of internal reliability. We use the Split-Half reliability coefficient given by two simple methods, the Spearman-Brown formula and the two-part coefficient alpha. The use of a permutation test for Split-Half reliability coefficient is highlighted as a valuable tool when the sample sizes are small and necessary assumptions cannot be met. The permutation tests for Split-Half reliability coefficient are illustrated with an example analysis of two survey data with a sample size of 15 and 35, respectively, and a hypothetical data with a sample size of 5.

▸Keyword: Permutation test, resampling, Split-Half reliability coefficient

I. Introduction

검사도구의 신뢰도는 검사도구가 측정하려는 것을 얼마나 정확하게, 일관성 있게 측정하였는지를 나타내는 통계수치로서 신뢰도가 높은 검사 도구를 사용하는 것은 연구의 성과를 결정짓는 중요한 요소이다. 신뢰도의 종류로는 재검사 신뢰도(test-retest reliability), 동형검사 신뢰도(parallei-form reliability), 문항간 내적 일관성 신뢰도(internal consistency reliability)가 있으며 동형검사나 재검사에 의한 신뢰도 측정이 가능하지 않을 때 연구자들은 거의 항상 문항간 내적 일관성 신뢰도를 널리 사용한다. 내적 일관성 신뢰도란 검사를 구성하고 있는 부분검사 또는 문항들이 측정하고자 하는 내용을 얼마나 일관성 있게 측정하고 있는지를 나타내는 정도이며, 크론바흐 α계수와 반분검사 신뢰도가 가장 널리 시용되는 지표이다. 크론바흐 α계수는 문항간의 내적 일관성을 측정하기 위하여 관찰점수 분산과 진점수 분산의 비율을 이용하는 반면에 반분검사 신뢰도(Split-Half Reliability)는 전체검사를 두 부분으로 양분하여 두개의 부분 검사간의 유사성으로 측정하는 방법이다. 반분검사 신뢰도를 측정하기 위해 전체 검사를 두 개로 양분하는 의미는 한 개의 검사를 두 개의 동형검사(검사 X 와 검사 Y)로 취급한다는 뜻이며, 이 때 동형검사 X 와 Y 사이에는 다음의 조건이 충족되어야 한다.

조건1 : 검사 X와 검사 Y의 진점수는 동일하다.

조건2 : 검사 X의 오차점수의 분산과 검사 Y의 오차점수의 분산은 동일하다.

조건3 : 검사 X의 진점수의 분산과 검사 Y의 진점수의 분산은 동일하다.

여기서 검사 X와 검사 Y 사이의 신뢰도를 피어슨 상관계수 r_{XY}에 의해 추정한다. 그러나 이 상관 계수는 문항 수가 반으로 줄어든 두 검사 점수간의 신뢰도이기 때문에 그대로 사용하면 검사의 신뢰도를 낮게 추정하는 문제가 있어 이를 교정하기 위해 Spearman-Brown이 제안한 반분검사 신뢰도 r_{SH}를 수식 (1)에 의해 구한다[1].

$$r_{SH} = \frac{2r_{XY}}{1+r_{XY}} \qquad (1)$$

수식 (1)에 의하면 $r_{SH} \geq r_{XY}$임을 알 수 있고, 검사 X와 검사 Y가 서로 독립적이면 $r_{SH} = 0$, 검사 X와 검사 Y 사이에 완

• First Author: Yonghwan Um, Corresponding Author: Yonghwan Um
*Yonghwan Um (uyh@sungkyul.ac.kr) ,Division of Industrial and Management Engineering, Sungkyul University
• Received: 2017. 05. 28, Revised: 2017. 06. 20, Accepted: 2017. 07. 04.

벽한 양의 상관성이 있으면 r_{SH} = 1이 된다. 그리고 일반적으로 $-\infty \le r_{SH} \le 1$ 이지만 0과 1 사이에 존재하는 값들만 유용한 것으로 여긴다.

그러나 실제적으로 동형검사로서의 조건 2를 만족하는 경우는 거의 없기 때문에 이 조건이 완화된 형태의 타우 동형검사(tau-equivalent form) 하에서 반분검사 신뢰도를 추정한다. 따라서 수식 (1)을 사용하기 보다는 크론바흐 α계수의 정의와 부합하는 형태의 새로운 추정치 $r_{2\alpha}$ (two-part coefficient alpha)가 반분검사 신뢰도로 제안되었다[2][3].

$$r_{2\alpha} = (4r_{XY}SD_XSD_Y)/(SD_X^2 + SD_Y^2 + 2r_{XY}SD_XSD_Y) \quad (2)$$

여기서 SD_X와 SD_Y는 각각 양분된 두 검사 X와 Y에 대한 표준편차이고 만일 SD_X = SD_Y이면 $r_{2\alpha}$는 r_{SH}와 같아진다. Charter의 연구에 의하면 r_{SH}는 $r_{2\alpha}$보다 반분검사 신뢰도를 과추정(overfitting) 하기 때문에 $r_{2\alpha}$가 r_{SH}보다 일반적으로 정확한 것으로 나타났고, 표준편차 SD_X와 SD_Y 사이에 차이가 있을 경우에는 $r_{2\alpha}$는 r_{SH}보다 작아지며, 이 차이가 커짐에 따라 신뢰계수 $r_{2\alpha}$는 더욱 크게 작아지는 것으로 나타났다. 실제적으로 SD_X와 SD_Y값이 일치하는 경우는 거의 없기 때문에 수식 (1) 보다는 수식 (2)에 의해 반분검사 신뢰계수를 추정하는 것이 더 바람직하다고 할 수 있다. 또한 Charter는 그의 다른 연구에서 $r_{2\alpha}$와 r_{SH}에 대한 신뢰구간을 제시한 바 있고 자세한 수식은 제 2절에서 소개한다[4].

본 연구는 퍼뮤테이션 검정(permutation test)을 이용하여 반분검사 신뢰도 r_{SH}와 $r_{2\alpha}$에 대한 p값과 경험적인 분위수 한계를(empirical quantile limits)를 산출하고 이것을 Charter가 제시한 r_{SH}와 $r_{2\alpha}$의 신뢰구간과 비교한다. 퍼뮤테이션 검정은 Fisher가 처음 제안한 이래로 Berry, Mielke, Johnston 등의 여러 연구자들에 의해 발전해온 통계절차로서 분석에 필요한 모든 정보가 관찰된 데이터 안에 들어 있다는 사실에 기초하고 있다. 특히 퍼뮤테이션 검정은 표본의 크기가 작을 경우나 모집단에 대한 정규성 가정이 만족되지 않을 때 요긴하게 사용될 수 있으며 모든 퍼뮤테이션 절차(permutation procedure)는 관찰 데이터의 모든 배열이 동일한 가능성을 갖고 발생할 수 있다는 가설 하에서 진행된다[5][6][7]. 퍼뮤테이션 검정의 내용은 제 3절에서 소개한다.

II. Confidence Intervals for r_{SH} and $r_{2\alpha}$

Charter(1997)는 재검사 신뢰도, 유사검사 신뢰도(alternate-form reliability), 타당도(validity coefficient), α계수, 반분검사 신뢰도 r_{SH}와 $r_{2\alpha}$에 대한 신뢰구간을 제안한 바 있다. Charter(1997)는 Zar가 제안한 상관계수의 신뢰구간 구축방법에 기초하여 r_{SH}의 신뢰구간을 구축하였고, Kristof가 (1970) 처음 제시한 방법에 의해 $r_{2\alpha}$에 대한 신뢰구간을 구축하였다[8][9][10]. Charter가 제안한 반분검사 신뢰도 r_{SH}와 $r_{2\alpha}$의 (1- α)*100% 신뢰구간은 각각 수식 (3)과 수식 (4)와 같다.

r_{SH}의 신뢰구간 =

$$[2r_L/(1+r_L),\ 2r_U/(1+r_U)] \quad (3)$$

$r_{2\alpha}$의 신뢰구간 =

$$[r_{2\alpha} - (D/2) - E,\ r_{2\alpha} - (D/2) + E] \quad (4)$$

수식 (3)의 r_L과 r_U은 각각 다음과 같다.

$$r_L = (10^A - 1)/(10^A + 1),$$
$$r_U = (10^B - 1)/(10^B + 1)$$

여기서 $A = z_L/1.1513$, $B = z_U/1.1513$ 이며 (z_L, z_U)은 r_{XY}의 신뢰구간으로서

$$z_L = z_r - z_{\alpha/2}\sqrt{\frac{1}{n-3}},$$
$$z_U = z_r + z_{\alpha/2}\sqrt{\frac{1}{n-3}}$$

이다. 또한 n은 표본의 크기, $z_{\alpha/2}$은 유의수준 α에 대한 z값, z_r은 r_{XY}의 Fisher 변환(Fisher Transformation)으로서 수식 (5)로 주어진다.

$$z_r = 1.1513 \cdot \log_{10}[(1+r_{XY})/(1-r_{XY})] \quad (5)$$

예를 들어 n=200 일 때 r_{XY} = 0.6이면 r_{SH} = 0.75이 되고, 이어서 z_r = 0.973, z_L = 0.833, z_U = 1.113, r_L = 0.682, r_U = 0.805 이 되어 r_{SH}의 95% 신뢰구간은 (0.682, 0.805)이 된다.

한편 수식(4)에서 D와 E는 각각 다음과 같고,

$$D = [(1 - r_{XY}^2)r_{2\alpha}^2 t_{\alpha/2}^2]/[(n-2)r_{XY}^2],$$
$$E = [D(1 + (D/4) - r_{2\alpha})]^{1/2}$$

여기서 $t_{\alpha/2}$는 자유도가 n-2인 student t를 뜻한다. 만일 n = 300이고 r_{XY} = 0.7, $r_{2\alpha}$ = 0.8이면 $t_{\alpha/2}$ = 1.9680, D = 0.008657와 E = 0.04183 이 되어 $r_{2\alpha}$의 95% 신뢰구간은 (0.7538, 0.8375)로 주어진다. 그리고 SD_x =SD_y가 되면 r_{SH}

의 신뢰구간과 $r_{2\alpha}$의 신뢰구간은 동일한 신뢰구간이 된다.

III. Permutation Test

퍼뮤테이션 검정은 완전히 데이터에 의존하는 (data-dependent) 절차로서 전통적인 통계검정에서 요구하는 가정으로부터 자유롭기 때문에 매우 유용한 방법이다. 퍼뮤테이션 검정은 정확한 검정(exact test)과 재표본 검정 (resampling test)으로 나눌 수 있는데, 정확한 검정은 관찰 데이터로부터 얻을 수 있는 모든 가능한 배열(arrangement)을 생성하고 이 각각의 배열에 대해서 검정통계량을 계산하는 것이며 재표본 검정은 모든 배열들 중의 일부만을 추출하여 사용한다. 만일 응답자의 수가 n이고 문항수가 k인 검사 도구에서 반분검사 신뢰도에 대한 정확한 검정을 실시할 때 동일한 확률로 만들어지는 가능한 모든 배열의 수는

$$M = (n!)^k$$

이 된다. 예를 들어 n = 10, k = 9이면 $M = (10!)^9 = 1.09\times10^{59}$ 개의 매우 많은 배열들이 만들어진다. 따라서 이 모든 배열들을 모두 고려하는 것은 실제적으로 가능하지 않기 때문에 이 중에서 L개(일반적으로 L=1,000,000)의 배열들을 임의로 추출하여 분석하는 재표본 검정을 사용한다.

L개의 배열들을 생성하기 위해서는 먼저 각 문항에 대해서 n개의 값들이 존재하므로 각 문항마다 문항 내에서 n개의 값들을 무작위로 섞어 전체적으로 하나의 배열($n\times k$ 데이터 행렬)을 생성한다. 그리고 이 과정을 L번 반복하여 L개의 배열을 만든 후 각 배열에 대해 반분검사 신뢰도 $(r_{SH})_i$ (또는 $(r_{2\alpha})_i$), (i = 1, 2, . . . , L) 를 계산한다. 이 때 L개의 반분검사 신뢰도들 중에서 원래의 관찰 데이터에서 계산된 반분검사 신뢰도 $(r_{SH})_o$ (또는 $(r_{2\alpha})_o$) 보다 크거나 같은 값을 갖는 반분검사 신뢰도들의 비율을 산출함으로써 p값을 계산한다.

$$p값 = \frac{1}{L}\sum_{i=1}^{L}\Phi_i(r_{SH})$$

또는

$$p값 = \frac{1}{L}\sum_{i=1}^{L}\Phi_i(r_{2\alpha})$$

여기서

$$\Phi_i(r_{SH}) = \begin{cases} 1 & r_{SH} \geq (r_{SH})_o 일때 \\ 0 & 아닐때 \end{cases}$$

$$\Phi_i(r_{2\alpha}) = \begin{cases} 1 & r_{2\alpha} \geq (r_{2\alpha})_o 일때 \\ 0 & 아닐때 \end{cases}$$

또한 퍼뮤테이션 검정은 모집단 분포에 의존하지 않기 때문에 전통적인 방법으로 $1-\alpha$ 신뢰구간을 구할 수 없으나 L개의 배열에서 얻은 L개의 반분검사 신뢰도 $(r_{SH})_i$ (또는 $(r_{2\alpha})_i$), (i = 1, 2, . . . , L) 로부터 $1-\alpha$ 경험적인 분위수 한계 (empirical quantile limits)를 구할 수 있다. 이를 위하여 먼저 재표본 기법을 사용하여 얻은 L개의 r_{SH}(또는 $r_{2\alpha}$)를 작은 값에서 큰 값으로 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 $(r_{SH})_1, (r_{SH})_2, . . . , (r_{SH})_L$ (또는 $(r_{2\alpha})_1, (r_{2\alpha})_2, \ldots, (r_{2\alpha})_L$) 에 대응되는 순서통계량을 $T_1 \leq T_2 \leq \cdots \leq T_L$ 이라 할 때 다음과 같이 주어진다.

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

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

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

IV. Examples

퍼뮤테이션 검정에 의한 p값 계산과 경험적인 분위수 한계를 예시하기 위해 두 개의 실제 데이터와 한 개의 가상데이터를 사용하였다. 이 실제 데이터들은 각각 인간의 내면성과 소비성향을 측정하는 검사도구들을 사용하여 수집된 것이고 가상데이터는 5점 척도의 4개 문항으로 구성되었다. 이 도구들로부터 반분검사 신뢰도 추정을 위한 동형검사를 만들기 위해 짝수번 문항과 홀수번 문항으로 양분하는 기우법을 사용하였다. 본 논문에서는 반분검사 신뢰도 추정, 퍼뮤테이션 p값 계산, 경험적인 분위수 한계 등의 모든 계산을 퍼뮤테이션 관련 R 프로그램에 의해 진행하였다.

1. Example 1

반분검사 신뢰도 연구를 위해 Prelog, Berry와 Mielke가 Levenson이 개발한 내면성 척도(Internality subscale)를 사용하여 얻은 데이터를 사용하였다(Table 1)[11][12]. 이 데이터는 총 n=15명의 참여자가 인간 내면성을 측정하는 6점 척도의 k=8개 문항에 응답한 결과이다. 이 데이터에 대해 퍼뮤테이션 검정을 할 때 고려해야하는 가능한 모든 배열의 수는 $M = (15!)^8 = 8.55\times10^{96}$ 이 되고 이 수는 매우 큰 값이므로 정확한 검정은 가능하지 않기 때문에 L = 1,000,000의 재표본 검정을 이용하였다. 이 데이터에서 관찰된 반분계수 신뢰도는 수식 (1)

Table 1. Raw scores for Internality subscale of Levenson Multi-Dimensional Locus of Control Scale

subject	item							
	1	2	3	4	5	6	7	8
1	6	5	4	5	6	4	5	6
2	4	3	5	3	4	6	5	5
3	6	6	2	4	5	3	4	6
4	3	6	6	3	3	4	3	4
5	2	6	2	4	3	5	6	6
6	5	6	4	5	6	4	3	2
7	2	4	3	5	2	3	3	6
8	6	4	5	5	6	6	6	6
9	4	3	4	4	6	6	3	4
10	5	4	4	5	5	5	4	6
11	6	4	6	2	3	5	6	3
12	3	6	4	6	2	5	5	2
13	4	3	2	4	3	4	5	2
14	3	2	6	3	4	3	3	3
15	4	6	2	5	2	2	3	2

Table 2. Permutation results and confidence intervals for internal subscale data

split-half reliability	resampling p-value	$1-\alpha$ empirical quantile limits		95% C.I.	99% C.I.
		$(Q_{0.025},\ Q_{0.975})$	$(Q_{0.005}\ ,\ Q_{0.995})$		
r_{SH}	0.2386	(-2.1264, 0.6798)	(-3.5899, 0.7820)	(-1.0708, 0.7846)	(-1.9550, 0.8490)
$r_{2\alpha}$	0.2370	(-1.9983, 0.6687)	(-3.3580, 0.7706)	(-1.0492, 0.7781)	(-1.9900, 0.8479)

과 (2)에 의해 각각 r_{SH} = 0.3321와 $r_{2\alpha}$ = 0.3257 로 계산되어 Charter의 연구에서처럼 $r_{2\alpha}$이 r_{SH}보다 작은 값을 보였고, 재표본 퍼뮤테이션 검정의 결과와 r_{SH}와 $r_{2\alpha}$에 대한 신뢰구간은 Table 2와 같다. Table 2에서 반분검사 신뢰도 r_{SH}에 대한 p값은 0.2386, 95% 경험적인 분위수 한계는 $Q_{0.025}$=-2.1264, $Q_{0.975}$=0.6798, 99% 경험적인 분위수 한계는 $Q_{0.005}$=-3.5899, $Q_{0.995}$=0.7820 이고, $r_{2\alpha}$에 대한 p값은 0.2370, 95% 경험적인 분위수 한계는 $Q_{0.025}$=-1.9983, $Q_{0.975}$=0.6687, 99% 경험적인 분위수 한계는 $Q_{0.005}$=-3.3580, $Q_{0.995}$=0.7706 이다. 재표본 퍼뮤테이션에 의한 r_{SH}와 $r_{2\alpha}$에 대한 p값은 비슷하게 나타났으나 r_{SH}와 $r_{2\alpha}$의 95%, 99% 경험적인 분위수 한계의 폭은 다소 넓게 나왔는데 이것은 표본의 크기가 작기 때문이다(n=15). 그리고 경험적인 분위수 한계를 비교할 때 $r_{2\alpha}$에 대한 95%, 99% 경험적인 분위수 한계의 폭은 각각 r_{SH}의 대응되는 경험적인 분위수 한계의 폭보다 좁게 나타났다. 경험적인 분위수 한계와 비교하기 위해 수식 (3)과 (4)에 의해 전통적인 신뢰구간을 계산하였으며, r_{SH}에 대한 95%, 99% 신뢰구간은 각각 (-1.0708, 0.7846), (-1.9550, 0.8490) 이고 $r_{2\alpha}$에 대한 95%, 99% 신뢰구간은 각각 (-1.0492, 0.7781), (-1.9900, 0.8479) 이다.

2. Example 2

Table 3은 어느 여자 대학교 학생들의 소비성향을 측정하기 위해 6점 척도의 k=8개 문항으로 구성된 검사 도구를 통해 얻은 데이터 중 일부분이다(n=35)[13]. 퍼뮤테이션 검정을 위해 고려해야 할 가능한 모든 배열의 수는 M = $(35!)^{10}$는 거의 무한대에 가까운 수이므로 L = 1,000,000의 재표본 검정을 이용하였다. 이 데이터로부터 관찰된 반분계수 신뢰도는 각각 r_{SH}= 0.6561 > $r_{2\alpha}$= 0.6554 이고 재표본 퍼뮤테이션 검정의 결과와 r_{SH}와 $r_{2\alpha}$에 대한 신뢰구간은 Table 4와 같다. Table 4에서 반분검사 신뢰도 r_{SH}에 대한 p값은 0.0015, 95% 경험적인 분위수 한계는 $Q_{0.025}$=-1.0003, $Q_{0.975}$=0.5008, 99% 경험적인 분위수 한계는 $Q_{0.005}$=-1.5042, $Q_{0.995}$=0.6020 이고, $r_{2\alpha}$에 대한 p값은 0.0014, 95% 경험적인 분위수 한계는 $Q_{0.025}$=-0.9828, $Q_{0.975}$=0.4967, 99% 경험적인 분위수 한계는 $Q_{0.005}$=-1.4751, $Q_{0.995}$=0.5968 이다. Example 1의 결과와 마찬가지로 $r_{2\alpha}$에 대한 95%, 99% 경험적인 분위수 한계의 폭은 각각 r_{SH}의 대응되는 경험적인 분위수 한계의 폭보다 좁게 나타났다. 또한 r_{SH}에 대한 95%, 99% 신뢰구간은 각각 (0.3123, 0.8280), (0.1450, 0.8617) 이고 $r_{2\alpha}$에 대한 95%, 99% 신뢰구간은 각각 (0.3109, 0.8277), (0.1377, 0.8623) 이다.

Table 3. Survey data on female college students' consumption propensity

subject	item									
	1	2	3	4	5	6	7	8	9	10
1	2	4	4	2	2	2	3	2	3	2
2	1	2	2	2	2	3	5	2	3	2
3	2	3	4	3	3	2	5	4	4	4
4	3	5	5	3	3	4	5	2	4	4
5	3	2	4	4	4	3	5	2	3	4
6	4	3	3	4	2	1	3	4	4	4
7	3	4	4	2	2	3	5	2	3	3
8	4	2	4	4	4	4	5	3	4	4
9	4	4	5	3	4	4	4	3	4	4
10	4	2	4	2	3	2	5	4	3	4
11	4	4	4	3	3	5	2	2	4	3
12	3	1	4	3	4	4	5	1	4	3
13	3	2	4	3	4	4	4	3	4	4
14	3	4	3	4	5	5	5	2	3	2
15	1	3	2	2	3	2	4	2	2	2
16	3	4	2	3	4	4	5	3	4	4
17	3	2	1	1	3	2	5	2	2	1
18	3	3	3	4	3	4	5	4	4	3
19	2	2	4	3	2	2	5	3	4	4
20	3	4	4	3	4	3	5	2	4	3
21	3	4	5	2	4	2	5	2	3	4
22	4	3	4	2	3	2	5	4	3	2
23	2	2	2	1	4	3	5	4	1	1
24	2	3	4	3	5	3	4	3	4	4
25	3	4	4	2	4	2	5	2	3	3
26	4	2	3	3	2	4	5	2	4	4
27	4	4	4	4	3	3	5	3	4	4
28	4	4	5	3	3	3	5	2	4	3
29	3	2	2	3	4	4	5	3	4	2
30	3	2	4	3	5	4	4	2	4	3
31	1	4	5	3	4	4	3	3	4	3
32	3	4	4	4	4	3	4	3	3	3
33	3	2	4	4	3	1	3	3	4	2
34	3	4	5	2	2	2	2	2	4	4
35	4	3	5	2	5	3	5	2	3	1

Table 4. Permutation results and confidence intervals for consumption propensity data

split-half reliability	resampling p-value	$1-\alpha$ empirical quantile limits		95% C.I.	99% C.I.
		$(Q_{0.025},\ Q_{0.975})$	$(Q_{0.005}\ ,\ Q_{0.995})$		
r_{SH}	0.0015	(-1.0003, 0.5008)	(-1.5042, 0.6020)	(0.3123, 0.8280)	(0.1450, 0.8617)
$r_{2\alpha}$	0.0014	(-0.9828, 0.4967)	(-1.4751, 0.5968)	(0.3109, 0.8277)	(0.1377, 0.8623)

3. Example 3

Table 5는 n=5, k=4로 이루어진 가상데이터로서 각 문항은 5점 척도(1부터 5)로 측정된다. 퍼뮤테이션 검정을 위해 고려해야 할 가능한 모든 배열의 수는 $M = (5!)^4 = 207360000$이고 이 수는 큰 수이지만 정확한 검정과 재표본 검정을 모두 이용하였다. 이 데이터로부터 관찰된 반분계수 신뢰도는 각각 $r_{SH} = 0.9154 > r_{2\alpha} = 0.8532$ 이고 재표본 퍼뮤테이션 검정의 결과와 r_{SH}와 $r_{2\alpha}$에 대한 신뢰구간은 Table 6, Table 7과 같다. 정확한 검정과 재표본 검정은 모두 거의 동일한 p값과 분위수 한계를 나타냈다. (r_{SH}의 경우 정확한 검정과 재표본 검정의 p값은 각각 0.0362과 0.0363이고 $r_{2\alpha}$의 경우 정확한 검정과 재표본 검정의 p값은 각각 0.0353과 0.0358). 그리고 Example 1과 Example 2의 결과와 마찬가지로 $r_{2\alpha}$에 대한 95%, 99% 경험적인 분위수 한계의 폭은 각각 r_{SH}의 대응되는 경험적인 분위수 한계의 폭보다 좁게 나타났다.

Table 5. Hypothetical data (n=5 and k=4)

subject	item			
	1	2	3	4
1	4	5	5	4
2	2	3	3	2
3	5	4	5	4
4	1	3	2	3
5	5	4	4	5

Table 6. Exact permutation results and confidence intervals for hypothetical data

split-half reliability	resampling p-value	$1-\alpha$ empirical quantile limits		95% C.I.	99% C.I.
		$(Q_{0.025}, Q_{0.975})$	$(Q_{0.005}, Q_{0.995})$		
r_{SH}	0.0362	(-14.000, 0.9371)	(-45.0783, 0.9787)	(-0.3532, 0.9947)	(-2.2331, 0.9978)
$r_{2\alpha}$	0.0353	(-7.5652, 0.8811)	(-18.0000, 0.9474)	(-0.2697, 0.9830)	(-2.6330, 0.9941)

Table 7. Resampling permutation results and confidence intervals for hypothetical data

split-half reliability	resampling p-value	$1-\alpha$ empirical quantile limits		95% C.I.	99% C.I.
		$(Q_{0.025}, Q_{0.975})$	$(Q_{0.005}, Q_{0.995})$		
r_{SH}	0.0363	(-14.000, 0.9372)	(-45.0783, 0.9787)	(-0.3532, 0.9947)	(-2.2331, 0.9978)
$r_{2\alpha}$	0.0358	(-7.4444, 0.8811)	(-18.0000, 0.9474)	(-0.2697, 0.9830)	(-2.6330, 0.9941)

V. Conclusion

본 논문은 내적 일관성 신뢰도의 하나인 반분검사 신뢰도 (r_{SH}와 $r_{2\alpha}$)에 대해 퍼뮤테이션 검정을 이용하여 p값과 $1-\alpha$ 경험적인 분위수 한계를 산출하는 방법을 소개한 것이다. 퍼뮤테이션 검정에서 이 방법은 이론적인 근사 분포가 아닌 통계치의 실제 이산형 분포에 기초한 확률값을 이용하여 진행된다. 퍼뮤테이션 검정은 모집단의 분포를 가정하지 않고, 정규성과 분산의 동질성(homogeneity),임의표본(random sample) 같은 전통적인 모수검정에서 요구되는 가정에 의존하지 않는다는 장점을 갖고 있어 전통적인 근사검정보다 더 선호하는 검정으로 인식되고 있다. 또한 전통적인 근사검정은 표본의 크기가 클 경우에 가능하기 때문에 표본의 크기가 작거나 특히 검사의 오차점수들이 서로 상관되어 있을 때는 종종 퍼뮤테이션 검정을 사용한다. 퍼뮤테이션 검정은 정확한 퍼뮤테이션 검정과 재표본 퍼뮤테이션 검정으로 나뉘는데 정확한 퍼뮤테이션 검정은 표본의 크기가 작은 경우에 유용하게 사용되고 재표본 퍼뮤테이션 검정은 표본의 크기에 제한 없이 사용된다.

본 논문에서는 표본의 크기가 각각 n=15, 35인 실제 데이터와 n=5인 가상데이터를 예제로 사용하였으며 각 데이터에서 생성되는 배열의 수가 매우 크기 때문에 n=15, 35인 데이터에서는 정확한 퍼뮤테이션 검정은 가능하지 않아 L=1,000,000의 재표본 퍼뮤테이션 검정을 이용하였다. 결과적으로 Example 1과 2의 데이터에서 r_{SH}와 $r_{2\alpha}$에 대한 p값은 거의 같았으나, $r_{2\alpha}$의 경험적인 분위수 한계의 폭이 r_{SH}의 경험적인 분위수 한계의 폭보다 좁게 나왔는데, 이것은 Charter의 연구에서 언급되었듯이 $r_{2\alpha}$이 r_{SH}보다 신뢰도를 더 정확히 추정하기 때문이라 볼 수 있다. 더불어 재표본 퍼뮤테이션 검정을 위해 L=1,000,000 개의 배열을 생성하고 p값을 계산하는데 걸리는 시간은 단지 132.2초(example 1의 경우), 165.6초(example 2의 경우)에 불과하였는데 이러한 점은 재표본 퍼뮤테이션을 사용하는 장점이라 말 할 수 있다. Example 3의 가상데이터에서도 Example 1과 2에서와 같은 결과가 나왔으며, 특히 정확한 퍼뮤테이션 검정과 재표본 퍼뮤테이션 검정이 거의 일치하는 결과를 제공함으로써 재표본 퍼뮤테이션 검정이 정확한 퍼뮤테이션 검정을 대체할 수 있는 것으로 나타났다. 이 결과는 재표본 퍼뮤테이션 검정이 정확한 퍼뮤테이션 검정에 대한 좋은 근사적인 방법이 된다는 연구 결과와 같다고 할 수 있다[6]. 비교적 작은 문항 수와 (k=4) 작은 표본 크기의(n=5) 가상 데이터에서조차도 정확한 퍼뮤테이션 검정에 의해 p값을 계산하는데 걸린 시간이 5시간 13분 21초인데 비해 재표본 퍼뮤테이션 검정에 의한 시간은 불과 90.67초라는 사실은 재표본 퍼뮤테이션 검정이 실제적인 대안이 될 수 있음을 보여준다.

REFERENCES

[1] C. Spearman, Correlations calculated from faulty data, British Journal of Psychology, Vol. 3, pp. 271-295, 1910.
[2] R. A. Charter, Note on the underrepresentation of the split-half reliability formula for unequal standard deviations, Perceptual and Motor Skills, Vol. 82, pp401-402, 1996.
[3] P. J. Rulon, A simplified procedure for determining the reliability of a test by split halves, Harvard Educational Review, Vol. 9, 99-103, 1939.
[4] R. A. Charter, Testing the difference between two or more independent alpha coefficients: an example, Perceptual and Motor Skills, Vol. 84, pp. 464-466, 1997.

[5] R. A. Fisher, A design of experiment, Oliver & Boyd, Edinburgh, 1935.
[6] P. W. Mielke Jr. and K. J. Berry, Permutation methods: a distance function approach. (2nd ed.) New York: Springer-Verlag, 2007.
[7] 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.
[8] R. A. Charter, Confidence Interval formulas for split-half reliability coefficients, Perceptual and Motor Skills, Vol. 86, pp. 1168-1170, 2000.
[9] J. H. Zar, Biostatistical Analysis. (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall, 1984.
[10] W. Kristof, On the sampling theory of reliability estimation, Journal of Mathematical Psychology, Vol. 7, 371-377, 1970.
[11] A. J. Prelog, K. J. Berry and P. W. Mielke, Resampling permutation probability values forCronbach's alpha, Perceptual and Motor Skills, Vol. 108, pp. 431-438, 2009.
[12] H. Levenson, Differentiating among internality, powerful others, and chance. In H. M. Lefcourt (Ed.), Research with the locus of control construct, Vol. 1, New York: Academic Press, pp. 15-63, 1981.
[13] M. H. Huh, SPSS Questionnaire Survey Methods: from Basic to Practical Use. Hannnare Publishing Col, 2010.

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.