Journal of The Korea Society of Computer and Information
Vol. 24 No. 6, pp. 159-164, June 2019
www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.06.159

Permutation Test for the Equality of Several Independent Cronbach's Alpha Coefficients

Yonghwan Um*

Abstract

The statistical inference of Cronbach's alpha measure of internal reliability is known to be inaccurate when sample size is small and the assumption of normality is violated. In this paper, we describe the permutation method in which we compute resampling p-values for testing the difference between two or more independent Cronbach's alpha coefficients. When the over-all permutation test is significant, we also make pairwise post-hoc comparisons using permutaion method. The permutation tests for the equality of two independent Cronbach's alpha coefficients and three independent Cronbach's alpha coefficients are illustrated with an example analysis of survey data.

▸Keyword: Cronbach's Alpha, Permutation method, resampling p-values

I. Introduction

신뢰도 측정(reliability measurement)에 대한 연구는 심리학, 교육학, 의학, 사회학 등의 다양한 분야에서 많은 학자들이 관심 가져온 연구 주제이며 특히 크론바흐 알파 계수(Cronbach's alpha coefficient = $\hat{\alpha}$)는 가장 널리 사용되는 신뢰도 척도 중에 하나이다. 크론바흐 알파 계수는 검사 도구의 문항들이 측정하고자 하는 개념을 얼마나 일관성 있게 측정하는 지를 나타내는 내적 일관성 신뢰도(internal consistency reliability)의 하나로서 동형검사(parallel form)나 재검사(test-retest)에 의한 신뢰도 측정이 가능하지 않을 때 사용된다. 크론바흐 알파 계수는 문항 점수가 구간척도의 연속형 점수일 때 추정 가능하며 문항 점수의 분산을 고려하여 정의한다. 문항의 수가 k개($X_1, X_2, \cdots, X_k$) 일 때 크론바흐 알파 계수는 다음과 같이 정의된다.

크론바흐 알파 계수 = $\hat{\alpha}$ =

$$\frac{k}{k-1}\left(1-\frac{X_1\text{의 분산}+X_2\text{의 분산}+\cdots+X_k\text{의 분산}}{(X_1+X_2+\cdots+X_k)\text{의 분산}}\right)$$

크론바흐 알파 계수는 0과 1 사이의 값을 가지며 1에 가까울수록 신뢰도가 높다는 의미이고 대체로 0.7 이상일 때 신뢰도가 높다고 말할 수 있다[1]. 그러나 크론바흐 알파 계수값이 크게 나올지라도 이에 대한 통계적 검증이 요구되므로 그동안 여러 연구자들이 크론바흐 알파에 대한 다양한 추론 및 가설검정에 필요한 이론적인 토대를 제공해 왔다[2,3]. 연구자들은 크론바흐 알파 계수 추정을 위해 문항들의 오차는 서로 독립이고 동일한 분산(homogeneous error variance)을 갖는 정규분포를 따른다고 가정하였으며, 이러한 가정 하에 Feldt, Woodruff와 Salih는 $(1-\hat{\alpha})^{-1}$가 $F_{n-1,(k-1)(n-1)}$ 분포를 따른다고 보고하였다[4]. 또한 많은 연구자들이 두 개의 크론바흐 알파 계수(α_1과 α_2)가 서로 동일한지를 비교하는 연구를 진행하여 왔는데, 예를 들어 왼손잡이 학생들과 오른손잡이 학생들에게 각각 지각-운동 능력검사(perceptual-motor test)를 실시하여 얻은 두 개의 크론바흐 알파 계수가 동일한지의 여부를 검정하는 것이다. 이 검정을 위해 Feldt는 F검정을 제안한 바 있고, 표본의 수가 $m \geq 2$ 개로 확장된 경우에는 Hakstian과 Whalen이 제안한 검정 방법이나 Feldt, Woodruff와 salih가 제안한 카이제곱 검정을 사용하여 m개의 독립적인 크론바흐 알파 계수들(α_1, α_2, , , α_m)에 대해 동일성을 검정할 수 있다[5,6]. 또한 Feldt는 동일 표본을 대상으로 두 가지 형태의 검사도구를 사용하여 얻은 두 개의 비독립적인 크론바흐 알파 계수

• First Author : Yonghwan Um, Corresponding Author: Yonghwan Um
*Yonghwan Um (uyh@sungkyul.ac.kr), Division of Industrial and Management Engineering, Sungkyul University
• Received: 2019. 04. 19, Revised: 2019. 05. 31, Accepted: 2019. 05. 31.

에 대해서도 동일성 검정을 실시하였으며 Woodruff와 Feldt는 $m \geq 2$ 개의 비독립적인 크론바흐 알파 계수들 사이의 동일성을 검정하는 방법을 제시하였다[7,8].

본 연구는 퍼뮤테이션 검정(permutation test)을 이용하여 $m \geq 2$ 개의 독립적인 크론바흐 알파 계수들 사이의 동일성 검정 및 퍼뮤테이션 p값 산출을 실행하고 이것을 Feldt의 F 검정 그리고 Feldt, Woodruff와 Salih가 제안한 카이제곱 검정 결과와 비교한다. 퍼뮤테이션 검정은 Fisher가 최초로 제안한 통계 절차로서 모집단의 분포에 대해 어떤 가정도 요구하지 않는 일종의 비모수적 검정방법이며, 관찰된 데이터를 재배치하여 얻은 가능한 모든 통계량의 값을 고려하기 때문에 귀무가설(관찰 데이터의 모든 퍼뮤테이션이 동일한 확률로 발생할 수 있다는 가설) 하에서의 통계량의 분포를 정확하게 제공하는 장점이 있다[9]. 따라서 퍼뮤테이션 검정은 정규성 가정을 만족시키지 못할 때 널리 사용될 수 있으며, 크론바흐 알파 계수에 대한 추론이 정규성 가정이 위반되거나 표본의 크기가 작을 때 매우 부정확하다고 보고된 바 있어 퍼뮤테이션 검정은 적절한 대안이 될 수 있다[10]. 그동안 퍼뮤테이션 검정은 Fisher 이후로 여러 학자들에 의해 다양한 통계분야에서 사용되어 왔으며 최근에 Prelog, Berry와 Mielke는 크론바흐 알파에 대해 재표본(resampling) 퍼뮤테이션 p값을 산출한 바 있고 엄용환은 반분검사 신뢰도(split-half reliability)에 대한 재표본 퍼뮤테이션 p값과 경험적인 분위수 한계(empirical quantile limits)를 제시하였다[11.12].

본 논문의 제 2절에서는 크론바흐 알파 계수의 동일성 검정을 위해 Feldt가 제안한 F검정, Feldt, Woodruff와 Salih의 카이제곱 검정과 Hakstian과 Whalen의 카이제곱 검정을 소개하며, 제 3절에서는 퍼뮤테이션 검정을 소개한 후 제 4절에서는 이 검정들을 데이터에 적용하여 비교한다.

II. Test of independent alpha coefficients

1. Test of two independent alpha coefficients

Feldt는 각각 크기가 n_i과 n_j인 두 개의 독립적인 집단을 대상으로 측정한 크론바흐 알파 계수의 동일성($H_0: \alpha_i = \alpha_j$)을 검정하기 위해 다음의 통계량을 사용하였다.

$$\frac{1-\hat{\alpha}_i}{1-\hat{\alpha}_j} \sim F_{n_j-1, n_i-1} \qquad (1)$$

이 통계량은 자유도가 $n_j - 1$, $n_i - 1$인 F분포를 따르며 (n_j는 더 큰 알파 계수와 관련된 표본의 크기이고 n_i는 더 작은 알파 계수와 관련된 표본의 크기), 두 개의 $\hat{\alpha}_i$과 $\hat{\alpha}_j$ 중 더 작은 알파 계수를 분자에, 더 큰 알파 계수를 분모에 위치하도록 정의한 것이다. 따라서 귀무가설이 사실이 아닐 때 $(1-\hat{\alpha}_i) > (1-\hat{\alpha}_j)$은 항상 1보다 큰 값을 갖으며 이로 인해 양측검정($H_1: \alpha_i \neq \alpha_j$)을 실시할 때 보통의 F분포에서 제공하는 것과 다른 임계값을 사용해야 한다. 이를테면 10% 유의수준에서 양측검정을 실시할 때는 해당되는 F분포의 95% 분위수를, 5% 유의수준일 때는 F분포의 97.5% 분위수를 임계점으로 사용해야 한다. 또한 이 때 p값은 다음과 같이 계산한다.

$$p값 = 2 \cdot P\left(F_{n_j-1, n_i-1} > \frac{1-더\ 작은\ \hat{\alpha}}{1-더\ 큰\ \hat{\alpha}}\right) \qquad (2)$$

F검정의 p 값 산출을 예시하기 위해 n_1=31, $\hat{\alpha}_1$=0.84, n_2=61, $\hat{\alpha}_2$=0.75 로 주어졌다고 할 때, F 통계량 = (1-0.75)/(1-0.84) =1.5625이 되고 F분포의 자유도는 30과 60이 된다. 만일 5% 유의수준에서 가설검정 하면 F=1.5625이 임계점(= $F_{30,60}$ 분포의 97.5 분위수=1.815)보다 작으므로 귀무가설을 기각하지 못하며 두 개의 알파 계수 사이에 유의한 차이가 없다고 말할 수 없다.

2. Test of more than two independent alpha coefficients

표본의 크기가 각각 $n_1, n_2, \ldots, n_m$ 인 m개의 독립적인 집단을 대상으로 측정한 크론바흐 알파 계수들의 동일성 ($H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_m$)을 검정하기 위해 Feldt, Woodruff와 Salih 그리고 Hakstian과 Whalen은 각각 자유도가 m-1인 카이제곱 분포를 따르는 통계량을 제안하였다. 우선 Feldt, Woodruff와 Salih가 제안한 통계량은 다음과 같다.

$$\chi^2 = \sum_{i=1}^{m} [1/(1-\hat{\alpha}_i)^{1/3} - D]^2 / C \sim \chi^2_{m-1} \qquad (3)$$

여기서

$$D = \sum_{i=1}^{m} [1/(1-\hat{\alpha}_i)^{1/3}]/m$$

$$C = \sum_{i=1}^{m} B_i/m$$

$$B_i = 2/[9(A_i-1)(1-\hat{\alpha}_i)^{2/3}]$$

$$A_i = n_i(k_i-1)/(k_i+1)$$

이고 k_i는 i번째 검사도구의 문항 수이다.

또한 Hakstian과 Whalen은 카이제곱 분포를 따르는 수식 (4)의 통계량 HW를 제안했다.

$$HW = \sum_{i=1}^{m}\left[\frac{(1-\hat{\alpha_i})^{1/3}-\mu}{S_i}\right]^2 \sim \chi^2_{m-1} \quad (4)$$

여기서 μ는 가중치=$1/S_i^2$을 이용하여 $(1-\hat{\alpha_i})^{1/3}$을 가중 평균한 값이고 S_i^2은 $(1-\hat{\alpha_i})^{1/3}$의 분산이다.

$$\mu = \frac{\sum_{i=1}^{m}(1-\hat{\alpha_i})^{1/3}/S_i^2}{\sum_{i=1}^{m}S_i^2},$$

$$S_i^2 = \frac{18(n_i-1)(1-\hat{\alpha_i})^{2/3}}{(9n_i-11)^2}\left(\frac{k_i}{k_{i-1}}\right)$$ 이다.

만일 알파 계수들의 동일성 검정이 유의하게 되면 계속해서 알파 계수를 쌍으로 비교하는 사후분석(pairwise post-hoc comparison, $H_0: \alpha_i = \alpha_j$)을 F 검정에 의해 실시한다. 본 연구에서는 사후분석을 위해 가설검정에서 발생하는 오류율을 일정수준(이를테면 5%) 이하로 관리하는 실험별 오류율(experimentwise error rate)의 하나인 Bonferroni 방식을 이용하였다. Bonferroni 방식에서는 전체 유의수준을 쌍비교의 수로 나눈 값을 개별 쌍비교의 유의수준으로 사용한다. 만일 전체 오류율이 5%이고 6개의 알파 계수들을 비교한다면 가능한 모든 쌍비교의 수는 ${}_6C_2$ =6(6-1)/2=15 개가 되므로 각각의 쌍 비교는 0.05/15=0.003의 유의수준에서 진행된다.

카이제곱 검정에 의해 3개의 알파 계수들(α_1, α_2, α_3) 사이의 동일성을 검정하는 예를 들어보자. 만일 $\hat{\alpha_1}$=0.80, n_1=30, $\hat{\alpha_2}$=0.87, n_2=25, $\hat{\alpha_3}$=0.93, n_3=40이고 $k_1 = k_2 = k_3$=30라 할 때 수식 (3)에 의해 A_1=28.0645, A_2=23.3871, A_3=37.4194, B_1=0.0240, B_2=0.0380, B_3=0.0359, C=0.0329, D=2.0368이 되어 Feldt, Woodruff와 Salih에 의한 카이제곱 통계량은 자유도=2를 갖는 χ^2=7.988이 되고 p값=0.0184으로 계산된다. 마찬가지로 수식 (4)에 의해 μ=0.4695, S_1^2=0.002753, S_2^2=0.002504, S_3^2= 0.001013이 되어 Hakstian과 Whalen의 통계량 HW=8.6284이 되고 p값=0.01338이 된다. 그러므로 유의수준 5%에서 3개의 알파 계수들 사이에 유의한 차이가 있다고 말 할 수 있으며 쌍 비교의 사후분석을 진행한다. 수식 (1)에 의해 알파 계수 α_1과 α_2를 쌍 비교할 때 $F_{24,29}$= 1.538 (p값=0.267)이 되고, α_1과 α_3의 경우에는 $F_{39,29}$=2.857 (p값=0.0043), α_2와 α_3의 경우에는 $F_{39,24}$=1.857 (p값 =0.111)이다. 전체 오류율을 5%라 하면 각각의 쌍 비교에 사용되는 유의수준은 0.05/3=0.0167 이므로 p값=0.0043 < 0.0167을 보인 알파 계수 α_1과 α_3 사이에서만 유의한 차이가 존재한다.

III. Method

퍼뮤테이션 검정은 모수 검정에서 요구되는 가정 없이 전적으로 데이터에만 의존하는 검정이므로 전통적인 검정보다 유용하게 사용된다. 기존의 모수 검정들은 표본의 크기가 작거나 근사적인 분포를 가정할 경우에 부정확한 p값을 제공하는 반면에 퍼뮤테이션 검정은 정확한 p값을 제공하는 것으로 알려져 있다 [13-15]. 퍼뮤테이션 검정은 관찰 데이터를 재배치하여 생성되는 모든 배열의 각각으로부터 통계량을 계산하여 진행하는 정확한 검정(exact test)과 모든 배열들 중에 일부 배열만을 추출하여 검정에 사용하는 재표본 검정(resampling test)으로 나뉜다. 표본의 크기가 각각 n_1, n_2, , , , n_m이고 $k_1 = k_2 = \cdots = k_m = k$인 독립적인 m개의 크론바흐 알파 계수들의 동일성에 대해 정확한 퍼뮤테이션 검정을 실시할 때 동일한 확률로 생성되는 모든 배열의 수는

$$M = [(n_1+n_2+\cdots+n_m)!]^k \quad (5)$$

이 된다. 예를 들어 n_1=7, n_2=8, k=4 일 때 M = 2.92×10^{48}개의 배열이 생성된다. 이 값은 매우 큰 무한대에 가까운 수이므로 실제적으로 정확한 퍼뮤테이션 검정을 실시하는 것이 불가능하기 때문에 이 배열들 중에 일부(L개, 일반적으로 L=1,000,000)를 임의 추출하여 분석하는 재표본 검정을 사용한다. 재표본 퍼뮤테이션 검정의 진행 과정은 첫째로 k개의 각 문항 내 있는 N(=$n_1+n_2+\cdots+n_m$)개의 문항값(X_{ij})들로부터 크기가 각각 n_1, n_2, , , n_m인 m개의 그룹들을 랜덤으로 만들어 하나의 N×k 데이터 행렬 형태의 임의 배열을 생성하고 둘째로 이 데이터 행렬을 m개의 데이터 행렬($n_1\times k$, $n_2\times k$, , , $n_m\times k$)로 순차적으로 분할하여 각 행렬에 해당되는 알파 계수 $\hat{\alpha_i}$, i =1, 2, , , ,m, 를 산출한 후 통계량 χ_i^2(또는 F_i)를 계산하며 이를 L번 반복 시행하여 L개의 통계량을 얻는다. 이 때 L개의 통계량 중에서 원래의 관찰된 데이터로부터 계산된 통계량 χ_0^2(또는 F_0) 보다 크거나 같은 값을 갖는 통계량의 비율이 p값이다.

$$p값 = \frac{1}{L}\sum_{i=1}^{L}\Psi_i(\chi^2) \quad (6)$$

또는

$$p값 = \frac{1}{L}\sum_{i=1}^{L}\Psi_i(F) \quad (7)$$

여기서

$$\Psi_i(\chi^2) = \begin{cases}1 & \chi^2 \ge \chi_o^2 일 때\\ 0 & 아닐 때\end{cases}$$

$$\Psi_i(F) = \begin{cases} 1 & F \geq F_o \text{일때} \\ 0 & \text{아닐때} \end{cases}$$

IV. Examples

퍼뮤테이션 검정에 의한 p값 산출을 예시하기 위해 두 개의 실제데이터를 사용하였다. 두 여학교 학생들의 소비성향을 측정하여 두 개의 알파 계수 사이의 동일성 검정을 실시하였고, 세 개 학년(2,3,4학년) 학생들의 사교성을 측정하여 세 알파 계수 사이의 동일성 검정을 실시하였다[16,17]. 본 논문에서는 퍼뮤테이션 p값 계산 및 F 검정과 χ^2검정에 의한 p값 계산 등의 모든 계산을 R 프로그램에 의해 진행하였다.

1. Example 1

Table 1. Survey data on female college students' consumption propensity

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

Table 1은 어느 두 개(A와 B)의 여자 대학교 학생들의 소비성향을 측정한 데이터(m=2, n_1=15, n_2=15, k=10)의 일부분이다. 퍼뮤테이션 검정을 위해 고려해야 할 전체 배열의 수는 $(15+15)^{10}$은 거의 무한대에 가까운 매우 큰 수 이므로 L=1,000,000의 재표본 검정을 실시하였다. 이 데이터로부터 관측된 각 대학 학생들의 크론바흐 알파 계수는 $\widehat{\alpha_A}$ = 0.3780, $\widehat{\alpha_B}$ = 0.7308 로서 B 대학 학생들의 신뢰도가 더 높았으며(Table 2) 두 알파 계수의 동일성 검정의 결과는 Table 3과 같다. F 검정의 통계치는 2.3108($F_{14,14}$=2.3108), p값=0.1290이며 재표본 퍼뮤테이션 검정의 p값=0.1485이므로 두 학교 학생들에 대한 소비성향에 신뢰도는 유의한 차이가 없다고 말할 수 있다.

Table 2. Observed alpha coefficients($\widehat{\alpha_A}$, $\widehat{\alpha_B}$)

college	alpha coefficients
A	$\widehat{\alpha_A}$ = 0.3780
B	$\widehat{\alpha_B}$ = 0.7308

Table 3. p values for testing equality of two alpha coefficients

tests	statistics	p values
F test	$F_{14,14}$=2.3108	0.1290
permutation test	-	0.1485

2. Example 2

Table 4는 어느 초등학교 2,3,4학년 학생들의 사교성을 측정하기 위해 5점 척도의 검사도구로부터 얻은 데이터이다 (m=3, n_1=12, n_2=10, n_3=10, k=5). 퍼뮤테이션 검정을 위해 고려해야 할 전체 배열의 수 $(12+10+10)^5$은 매우 큰 수 이므로 L=1,000,000의 재표본 검정을 실시하였다. 각 학년별로 관측된 알파 계수는 각각 $\widehat{\alpha_A}$ = 0.8858, $\widehat{\alpha_B}$ = 0.4369, $\widehat{\alpha_C}$ = 0.1844로서 2학년의 사교성이 가장 높고 4학년이 가장 낮은 신뢰도를 보였다(Table 5). 세 학년의 알파 계수 사이의 동일성에 대한 Feldt, Woodruff와 Salih의 카이제곱 검정의 통계치는 χ_2^2=7.2648, p값=0.02645이고, Hakstian과 Whalen에 의한 통계량 HW=8.3282, p값=0.01554 이므로 세 알파 계수 사이에 유의한 차이가 있는 것으로 나타났다. 계속해서 F 검정에 의한 사후분석을 실시했을 때 2학년과 3학년 사이에서 신뢰도는 차이가(p값=0.02362>0.05/3=0.0167) 없고, 3학년과 4학년 사이에서도 신뢰도는 차이가(p값=0.5899> 0.0167) 없으나 2학년과 4학년 사이에서만 신뢰도의 유의한 차이(p값=0.00640<0.0167)가 나타났다. 재표본 퍼뮤테이션에 의한 전체 검정에서는 p값=0.01004이므로 세 알파 계수 사이에 유의한 차이가 있으며 퍼뮤테이션 사후 분석은 2학년과 4학년 사이에서만 신뢰도의 유의한 차이(p값 = 0.01046 <0.0167)가 있다 (2학년과 3학년 비교에 대한 p값 = 0.03512>0.0167, 2학년과 3학년 비교에 대한 p값 = 0.6321>0.0167).

Table 4. Survey data on primary school students'(2^{nd},3^{rd},4^{th} grade) sociality

subject (2^{nd} grade)	item 1	2	3	4	5	subject (3^{rd} grade)	item 1	2	3	4	5	subject (4^{th} grade)	item 1	2	3	4	5
1	5	5	5	3	5	1	2	4	2	4	4	1	2	1	2	2	2
2	5	2	4	4	5	2	2	4	4	4	3	2	2	3	3	3	3
3	2	4	4	2	3	3	3	4	5	3	4	3	3	4	2	2	3
4	1	3	1	4	1	4	1	4	2	2	1	4	1	3	3	1	2
5	1	1	4	1	2	5	2	1	4	3	4	5	3	4	2	1	4
6	4	5	5	5	4	6	2	4	4	4	2	6	2	1	5	1	1
7	1	1	1	1	1	7	3	4	4	4	2	7	3	2	3	3	2
8	2	2	3	1	1	8	4	2	4	4	4	8	2	2	2	1	1
9	2	3	4	2	3	9	4	2	4	3	5	9	2	4	2	2	2
10	4	4	4	4	4	10	2	1	4	2	2	10	2	2	4	2	3
11	2	2	4	1	2												
12	3	2	2	3	3												

Table 5. Observed alpha coefficients($\widehat{\alpha_A}$, $\widehat{\alpha_B}$, $\widehat{\alpha_C}$)

grade	alpha coefficients
2^{nd}	$\widehat{\alpha_A}$ = 0.8858
3^{rd}	$\widehat{\alpha_B}$ = 0.4369
4^{th}	$\widehat{\alpha_C}$ = 0.1844

Table 6. p values for over-all test and post-hoc test

tests		statistics	p values
over-all test(Feldt, Woodruff & Salih)		χ_2^2=7.2648	0.0265
over-all test((Hakstian & Whalen)		HW=8.3282	0.0155
post-hoc test	2^{nd} grade vs. 3^{rd}	$F_{11,9}$=4.9296	0.0236
	2^{nd} grade vs. 4^{th} grade	$F_{11,9}$=7.1403	0.0064
	3^{rd} grade vs. 4^{th} geade	$F_{9,9}$=1.4484	0.5899

Table 7. Resampling permutation p values for over-all test and post-hoc test

tests		p values
over-all permutation test		0.0100
post-hoc permutation test	2^{nd} grade vs. 3^{rd} grade	0.0351
	2^{nd} grade vs. 4^{th} grade	0.0105
	3^{rd} grade vs. 4^{th} grade	0.6321

V. Conclusion

본 연구는 퍼뮤테이션 검정에 의해 2개 이상의 독립적인 크론바흐 알파 계수들 사이의 동일성 여부를 검정하고 p값을 산출하는 방법을 소개한 것이며 이 결과를 대응되는 F 검정 또는 카이제곱 검정과 비교한 것이다. 고전적인 근사 검정에서 정규성을 가정하여 얻은 p값은 일반적으로 부정확할 수 있으나 퍼뮤테이션 검정에 의하면 정확한 p값 계산이 가능하다. 특히 고려해야 할 배열의 수가 너무 많아서 정확한 퍼뮤테이션 검정(exact test)이 가능하지 않을 경우에도 재표본 퍼뮤테이션 검정(resampling test)은 표본의 크기에 관계없이 재표본의 횟수가 충분하다면(즉 큰 값의 L) 매우 정확한 p값을 제공한다.

본 논문에서는 퍼뮤테이션 검정을 실시하기 위해 표본의 크기가 각각 n_1=15, n_2=15인 소비성향 데이터(m=2)와 n_1=12, n_2=10, n_3=10인 사교성 데이터(m=3)를 예제 데이터로 사용하였으며 각 데이터에서 생성되는 배열의 수가 매우 크므로 L = 1,000,000의 재표본 검정을 실시하였다. 소비성향 데이터의 경우에는 퍼뮤테이션 p값(=0.1485) > F 검정의 p값(=0.1290) > α=0.05 이므로 두 개의 크론바흐 알파 계수 사이에 차이가 없는 것으로 나타났다. 사교성 데이터에서는 퍼뮤테이션 p값(=0.01004) < χ_2^2 검정의 p값(=0.02645 또는 0.01554) < α=0.05 이므로 세 개의 크론바흐 알파 계수 사이에 차이가 있는 것으로 나타났고, Bonferroni 방식에 의한 사후분석 결과는 2학년과 4학년간의 크론바흐 알파 계수의 차이에 대한 퍼뮤테이션 p값(=0.01046)과 F검정의 p값(=0.00640)이 모두 유의수준 0.05/3=0.0167보다 작으므로 크론바흐 알파 계수의 유의한 차이가 있는 것으로 나타났다. 퍼뮤테이션 검정에 의한 p값은 대응되는 F 검정(또는 χ_2^2 검정)의 p값보다 크거나 작은 값을 보이므로 검정결과가 보수적(conservative)이거나 리버럴(liberal)하다고 말할 수 있으나 퍼뮤테이션 검정이 모집단의 분포에 의존하지 않는다는 것과, 표본의 크기가 작거나 정규성 가정을 충족하지 못할 때 모수검정의 대안이 된다는 것이 큰 장점이라 할 수 있다. 특히 표본의 크기가 작을 때 근사적인 분포에 근거한 부정확한 통계추론에 비해 퍼뮤테이션 검정은 더 정확한 결과를 제공하고 전통적인 검정에서 요구하는 가정으로부터 자유로울 수 있어 그 활용도가 높다고 할 수 있다. 더욱이 퍼뮤테이션 검정은 표본의 크기가 작은 경우가 다반사인 심리학, 사회학, 교육학 등의 사회과학 분야의 연구에서 널리 사용되는 추론방법이 될 수 있다. 게다가 두 예제 데이터에서 보듯이 재표본 퍼뮤테이션 p값을 계산하는데 걸리는 시간이 717.68초(소비성향 데이터)와 633.99초(사교성 데이터)에 불과하다는 것은 퍼뮤테이션 검정이 실제적인 대안으로 사용될 수 있음을 보여준다.

퍼뮤테이션 검정방법은 앞으로도 신뢰도 뿐 아니라, 평정자간 일치도(agreement measure), 상관(association measure) 등의 여러 척도들에 적용 가능하며, 특히 퍼뮤테이션 방법에 의한 두 개 이상의 비독립적인 알파계수들 사이의 동일성 검정이 추후 과제가 될 것이다.

REFERENCES

[1] R. A. Peterson, A meta-analysis of Cronbach's alpha. The Journal of Consumer Research, 21, 381-391. 1994.

[2] W. Kristof, The statistical theory of stepped-up reliability coefficients when a test has been divided into deveral equivalent parts, Psychometrika, 28, 221-238, 1963.

[3] L. S. Feldt, The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty, Psychometrika, 30, 357-370, 1965.

[4] L. S. Feldt, D. J. Woodruff, and F. A. Salih, Statistical inference for coefficient alpha, Applied Psychological Measurement, 11, 93-103, 1987.

[5] L. S. Feldt, A test of the hypothesis for Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests, Psychometrika, 34, 363-373, 1969.

[6] A. R. Hakstian and T. E. Whalen, A k-sample significance test for independent alpha coefficients, Psychometrika, 41, 219-231, 1976.

[7] L. S. Feldt, A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample, Psychometrika, 45, 99-105, 1980.

[8] D. J. Woodruff and L. S. Feldt, Tests for equality of several alpha coefficients when their sample estimates are dependent, Psychometrika, 51, 393-413, 1986.

[9] R. A. Fisher, A design of experiment, Oliver & Boyd, Edinburgh, 1935.

[10] E. S. Edington and P. Onghena, Randomization tests, (4th ed.) Boca Raton, FL: Chapman & Hall/CRC, 2007.

[11] A.J. Prelog, K.J. Berry, and P.W. Mielke, Resampling Permutation Probability Values For Cronbach's Alpha, Perceptual and motor skills, 108, 2, 431-439, 2009.

[12] Y. Um, Permutation Analysis of Split-Half Reliability Coefficient, Journal of the Korea society of computer and information, 22, 7, 133-139, 2017.

[13] C. B. Holms, Sample size in Psychological research, Percepyual and Motor Skills, 49, 283-288, 1979.

[14] P. W. Mielke and K. J. Berry, Permutation methods : A distance function approach, Springer-Verlag, New York. 2001.

[15] P. I. Good, Permutation tests: A practical guide to data analysis, 2nd ed. Birkhauser, Massachusetts, 2001.

[16] M. H. Huh, SPSS Questionnaire Survey Methods: from Basic to Practical Use. Hannnare Publishing Col, 2010.

[17] E. H. Suh, SPSS Statistical Analysis, Free Academy, 2005.

Authors

Yonghwan Um received the B.S. and M.S. in Chemistry from Yonsei University, Korea, in 1981, 1983, respectively, 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.