

Combining Independent Permutation p-Values Associated with Multi-Sample Location Test Data

Yonghwan Um*

*Professor, Dept. of Industrial and Management Engineering, Sungkyul University, Anyang, Korea

[Abstract]

Fisher's classical method for combining independent p-values from continuous distributions is widely used but it is known to be inadequate for combining p-values from discrete probability distributions. Instead, the discrete analog of Fisher's classical method is used as an alternative for combining p-values from discrete distributions. In this paper, firstly we obtain p-values from discrete probability distributions associated with multi-sample location test data (Fisher-Pitman test and Kruskal-Wallis test data) by permutation method, and secondly combine the permutation p-values by the discrete analog of Fisher's classical method. And we finally compare the combined p-values from both the discrete analog of Fisher's classical method and Fisher's classical method.

▶ **Key words:** Discrete analog of Fisher's classical method, Combining p-values, Fisher-Pitman test, Kruskal-Wallis test, permutation

[요 약]

연속형 분포로부터 얻은 독립적인 p값들을 통합하는 Fisher의 고전적인 방법은 널리 사용되고 있지만 이산형 확률분포로부터 얻은 p값들을 통합하기에는 적절하지 않다. 대신에 유사 Fisher의 통합방법이 이산형 확률분포의 p값들을 통합하는 대안으로 사용된다. 본 논문에서는 첫째, 여러 표본들의 위치검정(Fisher-Pitman 검정과 Kruskal-Wallis 검정) 데이터와 관련된 이산형 확률분포로부터 퍼뮤테이션 방법에 의해 p값들을 구하고, 둘째로 이 p값들을 유사 Fisher의 통합방법을 이용하여 통합한다. 그리고 Fisher의 고전적인 방법과 유사 Fisher의 통합방법의 결과를 비교한다.

▶ **주제어:** 유사 Fisher의 고전적 방법, p값 통합, Fisher-Pitman 검정, Kruskal-Wallis 검정, 퍼뮤테이션

I. Introduction

특정 주제에 대해 연구된 데이터세트들이 존재하고, 전체 데이터세트들이 사용가능하다면 특정 통계가설을 지지해주는 통합된 증거가 존재하는지에 대한 질문이 자연적으로 생긴다. 이 문제를 연구하기 위해 메타분석을 실시하며, 메타분석의 한 방법으로서 가설검정의 신뢰도를 높이기 위해 연구결과들의 p값을 통합한다. 이 때 통합되는 p값들은 서로 독립이라고 가정하는데 이는 p값들 사이에 상관성이 존재하면 통합된 p값에 영향을 주기 때문이다. 통합 p값(combined p-value)은 여러 개별적인 연구들로부터 나온 p값들을 종합적으로 요약하여 정량적으로 산출한 결과이다.

지금까지 Tippett, Fisher, Stouffer, Wilkinson, Lunneborg, Edington, Simes, Pesarin, Zaykin 등의 많은 연구자들이 p값을 통합하기 위한 다양한 연구를 진행하여 왔다[1-9]. Pesarin은 연구들의 가설검정에서 제시된 p값들을 통합하는 다양한 방법들을 소개하였고, Fisher는 연속형 확률분포에서 얻은 독립적인 p값들(p_1, p_2, \dots, p_r)을 로그변환한 후 카이제곱분포를 이용하여 통합하는 고전적 방법(Fisher's classical method)을 제안하였으며 이 방법은 매우 우수한 근사적 성질을 나타내는 것으로 알려져 있다. Edington은 p값들의 합($p_1 + p_2 + \dots + p_r$)을 구하고 결과적으로 얻어지는 확률변수의 누적분포함수를 구하는 방식으로 p값을 통합하였으나 이 방법은 어느 한 개의 큰 p값이 압도적으로 통계량의 값을 결정하는 문제를 야기하였다. 그 밖에 p값 통합을 위해 Stouffer는 p값들의 표준정규점수로의 변환과 정규분포를 사용하였고 Tippett은 $\min(p_1, p_2, \dots, p_r)$ 을, Simes는 p값들의 순위 통계량을 사용하였으며 Zaykin 등은 어떤 기준값 이하의 p값들만을 선별하여 통합 p값을 구하는 Truncated Product Method를 제안하였다. 그 밖에 Kost와 McDermott는 Fisher 통계량의 근사적인 분포를 사용하여 서로 비독립적인 p값들을 통합한 바 있으며, 선정연과 김동재는 p값들에 대해 지수분포로 근사시키는 방법으로 p값을 통합하였다[10,11]

그러나 이러한 방법들은 연속형 확률분포로부터 나온 p값들을 통합하는데 사용되기 때문에 이산형 확률분포에 기초한 연구들에는 적용할 수 없다는 한계점이 있다. 따라서 이에 대한 대안으로서 Mielke, Johnston과 Berry는 유사 Fisher의 통합방법(discrete analog of Fisher's classical method)을 사용하여 이산형 확률분포로부터 나

온 독립적인 p값들을 통합하는 방법을 소개하였다[12]. 또한 이들은 모수검정의 제한점을 해결하기 위해 퍼뮤테이션 검정을 사용하였으며, 퍼뮤테이션 검정에서 생성되는 확률분포는 기본적으로 이산형이기 때문에 이로부터 얻은 퍼뮤테이션 p값을 통합하기 위해 기존의 Fisher의 고전적 방법이 아닌 유사 Fisher의 통합방법을 제안한 것이다. 퍼뮤테이션 검정은 최초로 Fisher에 의해 제안된 통계절차이며 모수검정에서 요구하는 가정 없이도 진행되고 주어진 데이터에만 의존하는 검정, 즉 모든 정보가 관측된 데이터 안에 포함되어 있는 검정으로서 정확한 p값을 제공하는 장점이 있다[13]. 퍼뮤테이션 검정은 정확한(exact) 퍼뮤테이션 검정과 재표본(resampling) 퍼뮤테이션 검정으로 나뉘는데, 정확한 퍼뮤테이션 검정은 주어진 데이터의 모든 가능한 배열을 구하고 각 배열마다 통계량을 산출하는 것이며, 재표본 퍼뮤테이션 검정은 표본의 크기가 커서 모든 배열을 고려하는 것이 불가능할 때 몇 개의(보통 1,000,000) 배열을 임의 추출하여 통계량을 구한다. Mielke, Johnston과 Berry는 퍼뮤테이션 방법에 의해 대응집단간의 평균 차이를 검정한 바 있으며 이 때 얻은 독립적인 p값들을 유사 Fisher의 통합방법에 의해 통합하여 Fisher의 고전적 방법보다 더 적합한 결과를 보였다[14]. 또한 엄용환은 Mann-Whitney 검정 데이터에서 얻은 독립적인 퍼뮤테이션 p값들에 유사 Fisher의 통합방법을 사용하여 정확한 결과를 제시하였다[15].

본 논문에서는 여러 집단 간의 중심을 비교하는 가설검정의 p값들을 통합한다. 이를 위해 첫째로 일원배치 분산분석 F 검정에 대응되는 Fisher-Pitman 검정과, 둘째로 집단들을 비모수적으로 순위를 이용하여 비교하는 Kruskal-Wallis 검정의 퍼뮤테이션 p값들을 각각 산출하고, 유사 Fisher의 통합방법과 Fisher의 고전적 통합방법에 의해 p값들을 통합하여 서로 비교하고자 한다[16,17].

II. Method

1. Fisher's classical method

Fisher는 연속형 확률분포하에서 r개의 연구로부터 얻은 r개의 독립적인 p값들을 통합하기 위해 수식(1)의 통계량을 제안하였다.

$$T = -2\ln \left[\prod_{i=1}^r p_i \right] = -2 \sum_{i=1}^r \ln p_i \sim \chi_{2r}^2 \quad (1)$$

이 때 통계량 T는 p_1, p_2, \dots, p_r 가 0과 1 사이의 연속형 균일분포를 갖는 확률변수라는 귀무가설 하에서 자유도 $2r$ 의 카이제곱분포를 따른다. 만일 통계량 T의 관찰값이 T_0 라 할 때 통합된 p값(P_C 로 표시)은 $P_C = P(T \geq T_0 | H_0)$ 로 계산된다. 결과적으로 이러한 Fisher의 고전적인 통합방법은 귀무가설 하에서는 정확한 통합 p값을 제시하지만, 유한의 확률공간에서 정의되는 이산형 확률분포를 따르는 독립 p값들을 통합하는 데는 적합하지 않고 오히려 통합 p값을 크게 만드는 보수적인 결과를 초래한다고 알려져 있다. Kincaid는 확률분포 안에 있을 수 있는 조그만 불연속성조차도 Fisher의 고전적인 방법에 의한 p값 통합에 심대한 영향을 준다고 보고한 바 있다 [18]. 따라서 만일 연속형 확률분포의 정규성과 같은 가정을 수용할 수 없는 경우에는 모든 검정이 본질적으로 이산형이기 때문에 Fisher의 고전적인 방법에 의한 p값 통합은 부정확하거나 근사적인 결과를 제공할 것이다.

2. Fisher-Pitman test and Kruskal-Wallis test

여러 집단들 사이의 평균을 비교하는 일원배치 분산분석(one-way ANOVA)에서 F검정은 분포의 정규성과 집단 사이의 등분산성을 요구한다. 그러나 이러한 가정을 충족할 수 없을 때 F검정의 대안으로 사용할 수 있는 퍼뮤테이션 검정이 Fisher-Pitman 검정이다. 일원배치 분산분석에서 F 통계량은 $F = (\text{집단간 평균제곱})/(\text{집단내 평균제곱})$ 으로 계산하지만 Fisher와 Pitman은 F값 계산 없이도 검정할 수 있는 Fisher-Pitman 통계량을 제안하였다. 비교하려는 집단의 수를 g , 각 집단의 크기를 n_k (즉 $k=1, \dots, g$)라 할 때 Fisher-Pitman 통계량은 수식(2)로 정의된다.

$$S = \sum_{k=1}^g n_k \bar{x}_k^2 \tag{2}$$

여기서

$$\bar{x}_k = \frac{1}{n_k} \sum_{l=1}^{n_k} x_{kl}$$

이고 x_{kl} 는 k 번째 집단의 l 번째 측정값이다. 통계량 F와 S 사이에는 수식(3)과 같은 관련성이 있어 F와 S는 동일한 검정을 수행하지만 Fisher-Pitman 통계량(S)은 F 통계량보다 더 쉽게 계산할 수 있고 분산추정치에 의존하지 않는 장점이 있다.

$$F = \frac{\text{집단간 평균 제곱}}{\text{집단내 평균 제곱}} = \frac{\sum_{k=1}^g n_k (\bar{x}_k - \bar{\tilde{x}})^2 / (g-1)}{\sum_{k=1}^g \sum_{l=1}^{n_k} (x_{kl} - \bar{x}_k)^2 / (n-g)} = \frac{(S - n\bar{\tilde{x}}^2) / (g-1)}{(Y - S) / (n-g)} \tag{3}$$

여기서

$$\bar{\tilde{x}} = \frac{1}{n} \sum_{k=1}^g n_k \bar{x}_k, \quad Y = \sum_{k=1}^g \sum_{l=1}^{n_k} x_{kl}^2$$

이다.

또한 Kruskal-Wallis 검정은 집단 간의 중심을 비교하는 비모수통계방법이다. 집단의 분포에 대한 가정을 하지 않고 데이터를 순위화하여 통계량을 정의한다. 따라서 Kruskal-Wallis 통계량의 확률분포는 이산형이 되므로 Fisher의 고전적인 방법에 의한 p값 통합은 적절치 않다. Kruskal-Wallis 통계량은 수식(4)와 같다.

$$H = \frac{12}{n(n+1)} \sum_{k=1}^g \frac{1}{n_k} \left[U_k - \frac{n_k(n+1)}{2} \right]^2 \tag{4}$$

여기서 U_k 는 g 개의 집단 안의 모든 측정값들을 작은 값부터 큰 값으로 순위를 1, 2, ...로 부여할 때 k 번째 집단에 속하는 측정값들의 순위합이고 $n = n_1 + n_2 + \dots + n_g$ 이다. 만일 데이터 중에 동일한 값이 있을 경우에는 순위들의 평균치를 부여한다. 수식(4)에서 $n_k(n+1)/2$ 는 모든 집단들의 중심은 동일하다는 귀무가설 하에서 i 번째 집단의 순위들의 기대합(expected sum)이다. 수식(4)를 정리하면 계산이 좀 더 간편한 형태의 수식(5)가 된다. 그리고 $g=2$ 이면 Kruskal-Wallis 검정은 Mann-Whitney 검정과 동일한 검정이 된다.

$$H = \frac{12}{n(n+1)} \sum_{k=1}^g \frac{U_k^2}{n_k} - 3(n+1) \tag{5}$$

3. Fisher's exact discrete analog of Fisher's classical method

Mielke, Johnston과 Berry가 제안한 유사 Fisher의 통합방법(exact discrete analog of Fisher's classical

method)을 Fisher-Pitman 퍼뮤테이션 검정과 Kruskall-Wallis 검정의 퍼뮤테이션 p값 통합에 적용한다.

Fisher-Pitman 검정의 통계량을 S_{ij} , Kruskall-Wallis 검정의 통계량을 H_{ij} 라 하자. 여기서 i 는 Fisher-Pitman 통계량 또는 Kruskall-Wallis 통계량을 사용한 전체 r 개의 연구 중에 i 번째 연구를 말하고 j 는 i 번째 개별 연구에서 퍼뮤테이션에 의해 고려할 수 있는 m_i 개의 배열의 수이다. 즉, $j=1,2, \dots, m_i$, 이고 $i = 1,2, \dots, r$ 이다. 그리고 i 번째 개별 연구에서 집단의 수는 g 이고 각 집단의 표본크기를 각각 $n_{i1}, n_{i2}, \dots, n_{ig}$ 이라 할 때 가능한 배열의 수는

$$m_i = \frac{(n_{i1} + n_{i2} + \dots + n_{ig})!}{n_{i1}!n_{i2}! \dots n_{ig}!}$$

로 계산되며 각각의 배열은 동일한 확률로 발생한다. 이 때 집단들 사이에는 차이가 없다는 귀무가설 하에서, i 번째 연구 데이터로부터 계산되는 퍼뮤테이션 p값($=p_i$)은 전체 m_i 개의 통계량 중에서 원래의 관측된 데이터로부터 계산된 통계량보다 크거나 같은 값을 갖는 통계량들의 비율이고 수식(6)으로 계산한다.

$$p_i = \sum_{j=1}^{m_i} \frac{\phi_{ij}}{m_i} \tag{6}$$

여기서

$$\phi_{ij} = \begin{cases} 1 & S_{ij} \geq S_{i0} \text{ 일때,} \\ 0 & \text{아닐때} \end{cases}$$

또는

$$\phi_{ij} = \begin{cases} 1 & H_{ij} \geq H_{i0} \text{ 일때,} \\ 0 & \text{아닐때} \end{cases}$$

이며 S_{i0} 와 H_{i0} 는 각각 i 번째 연구에서 관측된 Fisher-Pitman 통계량과 Kruskall-Wallis 통계량이다.

동일한 방법으로 r 개의 연구로부터 r 개의 퍼뮤테이션 p값(p_1, p_2, \dots, p_r)을 구하고, p_1, p_2, \dots, p_r 은 서로 독립적인 관계가 있으므로 수식(7)를 이용하여 정확한 통합 p값, 즉 유사 Fisher의 통합방법에 의한 p값($=P_D$)을 구한다.

$$P_D = \frac{1}{M} \sum_{j_1=1}^{m_1} \dots \sum_{j_k=1}^{m_k} \omega_{j_1, \dots, j_k} \tag{7}$$

여기서 M 은 p_1, p_2, \dots, p_r 을 통합하기 위해 고려해야 할 전체 배열의 수로서 $M = \prod_{i=1}^r m_i$ 이고,

$$\omega_{j_1, \dots, j_k} = \begin{cases} 1 & \sum_{i=1}^r S_{ij_i} \geq \sum_{i=1}^r S_{i0} \text{ 일때,} \\ 0 & \text{아닐때} \end{cases}$$

또는

$$\omega_{j_1, \dots, j_k} = \begin{cases} 1 & \sum_{i=1}^r H_{ij_i} \geq \sum_{i=1}^r H_{i0} \text{ 일때,} \\ 0 & \text{아닐때} \end{cases}$$

이다. 또한 $\sum_{i=1}^r S_{ij_i}$ 와 $\sum_{i=1}^r H_{ij_i}$ 은 각각 M 개의 각 배열에서 계산되는 r 개의 Fisher-Pitman 통계량들의 합과 Kruskall-Wallis 통계량들의 합이며, $\sum_{i=1}^r S_{i0}$ 과 $\sum_{i=1}^r H_{i0}$ 은 각각 r 개의 연구에서 관측된 Fisher-Pitman 통계량들의 합과 Kruskall-Wallis 통계량들의 합이다.

III. Examples and Results

집단들 간의 중심의 차이를 검증하는 Fisher-Pitman 검정과 Kruskall-Wallis 검정과 관련하여 두 종류의 데이터(가상 데이터와 실제 데이터)에 대해 퍼뮤테이션 p값을 계산하고 이 p값들을 통합하였다. 유사 Fisher의 통합방법에 의해 통합된 p값($=P_D$)을 제시하였고 이를 Fisher의 고전적 방법에 의해 통합된 p값($=P_C$)과 비교하였다. 본 논문에서는 퍼뮤테이션 생성, Fisher-Pitman 통계량 계산, Kruskall-Wallis 통계량 계산, 퍼뮤테이션 p값 계산, 통합 p값 계산 등 모든 계산을 R 프로그램에 의해 진행하였다.

1. Example 1

Table1은 p값 통합을 위해 사용된 네 개의 연구($r=4$)와 3개 집단($g=3$)의 반응값들로 이루어진 가상 데이터이다. 즉 $n_{11} = 5, n_{12} = 6, n_{13} = 5, n_{21} = 4, n_{22} = 5, n_{23} = 6, n_{31} = 7, n_{32} = 4, n_{33} = 5, n_{41} = 4, n_{42} = 6, n_{43} = 6$ 이다. 각 연구의 데이터 세트에서 생성되는 퍼뮤테이션의 수는 각각 $m_1 = 16! / (5!6!6!) = 2018016, m_2 = 15! / (4!5!6!) = 630630, m_3 = 16! / (7!4!6!) = 1441440, m_4 = 16! / (4!6!6!) = 1681680$ 로 계산되고, 관측된 Fisher-Pitman 검정통계량과 Kruskall-Wallis 검정통계량은 각각 $S_{10} = 3116.367, S_{20} = 2876.217, S_{30} = 3217.2, S_{40} = 3365.917, H_{10} = 3.1618, H_{20} = 2.7951, H_{30} = 5.4371, H_{40} = 1.5569$ 이다(Table2). 이 값들과 수식 (6)을 이용하여 얻은 Fisher-Pitman 검정에서의 퍼뮤테이션 p값은 각각 $p_1 = 0.1681, p_2 = 0.4757, p_3 = 0.1401, p_4 = 0.2692$ 이며, Kruskall-Wallis 검정에서의 퍼

뮤테이션 p값은 각각 $p_1=0.2131$, $p_2=0.2637$, $p_3=0.0598$, $p_4=0.4802$ 이다(Table3). 통합된 p값을 계산하기 위해 수식 (7)을 이용하였으며 이 때 $\sum_{i=1}^4 S_{i0} = 12575.7$, $\sum_{i=1}^4 H_{i0} = 12.9509$ 이고 가능한 배열의 수 $M = \prod_{i=1}^k m_i = 2018016 \times 630630 \times 1441440 \times 1681680$ 는 무한대에 가까운 수이므로 이 중에서 1,000,000개의 배열을 임의로 추출하여 각 배열에 대응되는 Fisher-Pitman 검정통계량과 Kruskal-Wallis 검정통계량을 계산하였다. 결과적으로 총 1,000,000개의 배열 중에 $\sum_{i=1}^4 S_{i0} = 12575.7$ 보다 크거나 같은 Fisher-Pitman 통계량들의 합인 수는 124889개 이고 $\sum_{i=1}^4 H_{i0} = 12.9509$ 보다 크거나 같은 Kruskal-Wallis 검정통계량들의 합인 수는 97939개로 나타나 유사 Fisher의 통합방법에 의한 통합 p값(P_D)은 각각 0.1249와 0.0979로 계산되었다(Table4와 Table5). Table4와 Table5에서는 P_D 와 더불어 Fisher의 고전적 통합방법에 의한 p값(P_C)을 제시하였으며 P_D 와 P_C 를 비교하기 위해 다음과 같이 정의된 표준화 퍼센트 차이 (standardized % difference) 값을 이용하였다.

$$\text{표준화 퍼센트 차이} = 100 \times (P_C - P_D) / P_D$$

표준화 퍼센트 차이는 P_C 가 P_D 와 비교하여 상대적으로 얼마나 큰 차이가 있는지를 퍼센트로 나타낸 것으로서 이 수치가 양의 값일 때는 Fisher의 고전적 통합방법이 통합된 p값을 크게 만드는 보수적인 결과를 낳는다는 것을 의미한다. 따라서 Fisher-Pitman 검정의 경우에는 35.79%, Kruskal-Wallis 검정에서는 19.31%의 표준화 퍼센트 차이를 나타냄으로써 Fisher의 고전적 통합방법은 유사 Fisher의 통합방법보다 더 큰 p값을 제시함을 알 수 있다. 참고로 Table4와 Table5의 P_C 는 수식(1)에 의해 다음과 같이 계산된다. Fisher-Pitman 검정의 경우에는

$$\begin{aligned} -2 \sum_{i=1}^r \ln p_i &= -2[\ln(0.1681) + \ln(0.4757) \\ &\quad + \ln(0.1401) + \ln(0.2692)] \\ &= 11.6079 \end{aligned}$$

이므로 $P_C = P(\chi_8^2 \geq 11.6079 \mid H_0) = 0.1696$ 이고, Kruskal-Wallis 검정에서는

$$\begin{aligned} -2 \sum_{i=1}^r \ln p_i &= -2[\ln(0.2131) + \ln(0.2637) \\ &\quad + \ln(0.0598) + \ln(0.4802)] \\ &= 12.8598 \end{aligned}$$

이므로 $P_C = P(\chi_8^2 \geq 12.8598 \mid H_0) = 0.1168$ 이다.

Table 1. Four data sets consisting of three groups each.

data set 1			data set 2			data set 3			data set 4		
grp.1	grp.2	grp.3	grp.1	grp.2	grp.3	grp.1	grp.2	grp.3	grp.1	grp.2	grp.3
13	13	10	14	11	14	12	14	11	13	14	10
11	12	14	13	12	9	13	18	16	14	18	15
12	15	16	12	13	11	11	21	19	12	20	15
10	16	21	14	13	19	9	9	15	10	10	14
14	10	18		14	15	10		18		12	17
	17				23	14				21	16
						15					

Table 2. m_i , S_{i0} and H_{i0} for each data set in Table 1.

	data set 1	data set 2	data set 3	data set 4
m_i	2,018,016	630,630	1,441,440	1,681,680
S_{i0}	3116.367	2876.217	3217.2	3365.917
H_{i0}	3.1618	2.7951	5.4371	1.5569

Table 3. Permutation p Values for each data set in Table 1.

	data set 1	data set 2	data set 3	data set 4
perm. p value (Fisher-Pitman Test)	0.1681	0.4757	0.1401	0.2692
perm. p value (Kruskal-Wallis Test)	0.2131	0.2637	0.0598	0.4802

Table 4. Combined p Values for Fisher-Pitman Test

method	combined p value	standardized % difference
discrete analog of Fisher's classical method	0.1249(=P _D)	35.79
Fisher's classical method	0.1696(=P _C)	

Table 5. Combined p Values for Kruskal-Wallis Test

method	combined p value	standardized % difference
discrete analog of Fisher's classical method	0.0979(=P _D)	19.31
Fisher's classical method	0.1168(=P _C)	

2. Example 2

Table6은 통합 p값을 구하기 위해 Mielke와 Berry에 의해 사용된 이 집단(two-sample)의 실제데이터로서 세 개의 연구에서 두 개의 처리를 적용해 얻은 반응값이다. (r=3, g=2, n₁₁=5, n₁₂=4, n₂₁=3, n₂₂=7, n₃₁=4, n₃₂=4)[19]. 각 데이터 세트에서 생성되는 퍼뮤테이션의 수는 m₁ = 9!/(5!4!) =126, m₂ = 10!/(3!7!) = 120, m₃ = 8!/(4!4!) =70이고 관측된 Fisher-Pitman 검정통계량과 Kruskal-Wallis 검정통계량은 각각 S₁₀=2905.05, S₂₀=4801.286, S₃₀=2938, H₁₀=0.7412, H₂₀=1.5714, H₃₀=1.7078 이다(Table7). 그리고 수식 (6)에 의해 계산되는 Fisher-Pitman 검정에서의 퍼뮤테이션 p값은 각각 p₁=0.3968, p₂=0.2, p₃=0.2286, Kruskal-Wallis 검정에서의 퍼뮤테이션 p값은 각각 p₁=0.4603, p₂=0.2667, p₃=0.2286이다(Table8). 계속해서 통합 p값을 계산하는 수식(7)에서 $\sum_{i=1}^4 S_{i0} = 9924.336$, $\sum_{i=1}^4 H_{i0} = 4.0204$ 이고, 가능한 배열의 수 $M = \prod_{i=1}^r m_i = 126 \times 120 \times 70 = 1,058,400$ 중에서 $\sum_{i=1}^4 S_{i0} = 9924.336$ 보다 크거나 같은 Fisher-Pitman 통계량들의 합의 수는 200784개이고, $\sum_{i=1}^4 H_{i0} = 12.9509$ 보다 크거나 같은 Kruskal-Wallis 검정통계량들의 합의 수는 285124개로 나타나 유사 Fisher의 통합방법에 의한 통합 p값(=P_D)은 각각 200784/1058400=0.1898와 285124/1058400=0.2694로 계산된다. (Table9과 Table10). 이 통합 p값(=P_D)들은 전체 배열 1058400개를 모두 사용하여 산출된 비율이기 때문에 정확한 p값이라고 말할 수 있다. 반면에 Fisher-Pitman 검정과 Kruskal-Wallis 검정에서의 Fisher의 고전적 방법에 의한 계산된 통합 p값은 각각

$$\begin{aligned}
 -2 \sum_{i=1}^r \ln p_i &= \\
 &= 2[\ln(0.3968) + \ln(0.2) + \ln(0.2286)] \\
 &= 8.0192, \\
 P_C &= P(\chi_6^2 \geq 8.0192 | H_0) = 0.2367
 \end{aligned}$$

그리고

$$\begin{aligned}
 -2 \sum_{i=1}^r \ln p_i &= \\
 &= 2[\ln(0.4603) + \ln(0.2667) + \ln(0.2286)] \\
 &= 7.1470, \\
 P_C &= P(\chi_6^2 \geq 7.1470 | H_0) = 0.3075
 \end{aligned}$$

이며 표준화 퍼센트 차이는 각각 24.71%, 14.14%이다.

Table 6. Three data sets consisting of three treatments each.

data set 1		data set 2		data set 3	
treat.1	treat.2	treat.1	treat.2	treat.1	treat.2
14	20	16	21	16	22
19	27	22	18	12	29
15	14	13	24	21	16
23	18		15	17	19
11			28		
			25		
			19		

Table 7. m_i, S_{i0} and H_{i0} for each data set

	data set 1	data set 2	data set 3
m _i	126	120	70
S _{i0}	2905.05	4081.286	2938
H _{i0}	0.7412	1.5714	1.7078

Table 8. Permutation p Values for each data set

	data set 1	data set 2	data set 3
perm. p value (Fisher-Pitman Test)	0.3968	0.2	0.2286
perm. p value (Kruskal-Wallis Test)	0.4603	0.2667	0.2286

Table 9. Combined p Value for Fisher-Pitman Test

method	combined p value	standardized % difference
exact discrete analog of Fisher's classical method	0.1898(=P _D)	24.71
Fisher's classical method	0.2367(=P _C)	

Table 10. Combined p Value for Kruskal-Wallis Test

method	combined p value	standardized % difference
exact discrete analog of Fisher's classical method	0.2694(=P _D)	14.14
Fisher's classical method	0.3075(=P _C)	

IV. Conclusion

연구결과의 p값들을 통합하는 Fisher의 고전적인 통합 방법은 연속형 확률분포라는 가정하에서 사용되어 왔다. 또한 모수적 가설검정의 연구는 정규성, 등분산성 같은 가정을 요구하므로 이 가정이 충족되지 못할 때 퍼뮤테이션 검정은 하나의 대안으로 사용된다. 그러나 퍼뮤테이션 검정으로 말미암은 확률분포는 기본적으로 이산형 확률분포이므로 퍼뮤테이션 검정에 의해 얻은 p값들을 통합하기 위해 Fisher의 고전적인 통합방법이 아닌 유사 Fisher의 통합방법을 사용한다. 본 논문은 여러 집단들의 중심의 차이가 있는지를 검정하는 Fisher-Pitman 검정과 Kruskal-Wallis 검정을 적용한 데이터들로부터 퍼뮤테이션 p값들을 산출하고, 이 p값들을 유사 Fisher의 통합방법과 Fisher의 고전적인 통합방법에 의해 통합하고 비교한 것이다.

본 연구에서는 r=4, g=3인 가상데이터와 r=3, g=2인 실제데이터를 사용하여 통합 p값(P_D와 P_C)를 산출하였다. 여기서 가상 데이터의 경우에는 전체 배열의 수가 매우 크므로(M≃∞) 통합 p값을 근사적으로 산출한 반면에, 실제 데이터의 경우에는 M=1058400 개의 퍼뮤테이션을 모두 사용할 수 있어 정확한 통합 p값을 계산할 수 있었다. 결과적으로 Fisher-Pitman 검정과 Kruskal-Wallis 검정에서 모두 Fisher의 고전적인 방법이 유사 Fisher의 통합방

법보다 더 큰 p값을 나타내었다 (가상 데이터에서 Fisher-Pitman 검정 : P_D = 0.1249 < P_C = 0.1696, 가상 데이터에서 Kruskal-Wallis 검정 : P_D = 0.0979 < P_C = 0.1168, 실제 데이터에서 Fisher-Pitman 검정 : P_D = 0.1898 < P_C = 0.2367, 실제 데이터에서 Kruskal-Wallis 검정 : P_D = 0.2694 < P_C = 0.3075). 이 결과는 확률분포가 연속성으로부터 벗어날 경우 Fisher의 고전적인 방법에 의한 통합은 p값을 크게 하기 때문에 퍼뮤테이션 p값들의 통합은 유사 Fisher의 통합방법에 의해 진행되어야함을 보여주는 것이다. 실제로 이는 Fisher의 고전적인 방법에 의해 통합된 p값은 확률분포상에 존재하는 미세한 불연속성에 의해서도 큰 영향을 받는다는 연구결과와 부합한다고 말할 수 있고 또한 Fisher의 고전적인 통합방법은 연구결과의 p값들(p₁, p₂, ..., p_r)이 0과 1 사이에서 독립적인 균일분포를 갖는다는 무리한 가정에 기초하고 있기 때문에 생긴 결과라고 할 수 있다.

따라서 표본의 크기가 작을 때 모수적 검정 대신 퍼뮤테이션 검정을 사용해야 하는 연구들의 p값들은 유사 Fisher의 통합방법에 의해 통합할 때 정확한 결과를 얻을 수 있다. 실제로 교육학, 심리학, 사회학 등의 사회과학 분야의 많은 연구에서는 작은 크기의 표본을 이용한 연구가 많아 본 연구의 p값 통합은 유용한 연구방법이 될 수 있으며 또한 p값 통합이 서로 일치하지 않은 여러 연구결과들을 객관적으로 평가, 종합하는 메타분석의 필요한 통계처리 과정의 하나이므로 메타분석의 정확성을 높이는 데도 기여한다고 볼 수 있다.

REFERENCES

- [1] I. H. Tippett, The methods of statistics, Williams and Norgate, London, 1931.
- [2] R. A. Fisher, "Statistical methods for research workers", (5th ed.) Edinburgh, Scot.: Oliver & Boyd, 1934.
- [3] S. A. Stouffer, E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. William, The American Soldier, *Adjustment During Army Life*, Princeton Univ. Press, Princeton, 1949.
- [4] B. Wilkinson, "A Statistical Consideration in Psychological Research", *Psychological Bull*, 48, 156-158, 1951.
- [5] C. E. Lunneborg, "Analysis by resampling: concepts and applications". Pacific Grove, CA: Duxbury, 2000.
- [6] E. S. Edington, "An additive model for combining probability values from independent experiments", *Journal of Psychology*, 80, 351-363, 1972.
- [7] R. J. Simes, "An improved Bonferroni procedure for multiple tests

of significance”, *Biometrics*, 73, 751-754, 1986.

- [8] E. Pesarin, “Multivariate permutation tests: with applications in biostatistics”, Chichester, UK: Wiley, 2001.
- [9] D. Zaykin, L. Zhivotovsky, P. H. Westfall, and B. S. Weir, “Truncated product method for combining p-values”, *Genetic Epidemiology*, 22, 170-185, 2002.
- [10] J. T. Kost and M. P. McDermott, “Combining dependent P-values”, *Statistics & Probability Letters*, 60, 183-190, 2002.
- [11] J. Seon and D. Kim, “New Method for Combining P-values in Metha-Analysis”, *The Korean Journal of Applied Statistics*, 26, 5, 797-806, 2013.
- [12] P. W. Jr. Mielke, J. E. Johnston and K. J. Berry, “Combining probability value from independent permutation tests: a discrete analog of Fisher’s classical method”, *Psychological Reports*, 95, 449-458, 2004.
- [13] R. A. Fisher, *A design of experiment*, Oliver & Boyd, Edinburgh, 1935.
- [14] P. W. Jr. Mielke, J. E. Johnston and K. J. Berry, “Comparisons of continuous and discrete methods for combining probability values associated with matched-pairs t-test data”, 100, 799-805, 2005.
- [15] Y. H. Um, “Combining independent permutation p Values associated with Mann-Whitney test data”, *Journal of the Korean Society of Computer and Information*, 23, 7, 99-104, 2018.
- [16] K. J. Berry and P. W. Mielke, “The Fisher-Pitman Permutation Test: An Alternative to the F Test”, *Psychological Reports*, 90, 495-502, 2002.
- [17] G. W. Corder, and D. I. Foreman. *Nonparametric Statistics for Non-Statisticians*. Hoboken: John Wiley & Sons. pp. 99-105, 2009.
- [18] W. M. Kincaid, “The combination of tests based on discrete distributions”, *Journal of the American Statistical Association*, 57, 10-19, 1962.
- [19] 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 Department of Industrial and Management Engineering, Sungkyul University. He is interested in reliability measure, data-mining, statistical inference.