The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.4.729
29(4), 729–739

Nonparametric method in one-way layout based on joint placement

Kyoung-Ah Jeon[a] · Dongjae Kim[a,1]

[a]Department of Biomedicine · Health Science, The Catholic University of Korea

(Received April 6, 2016; Revised May 16, 2016; Accepted May 18, 2016)

Abstract

Kruskal and Wallis (1952) proposed a nonparametric method to test the differences between more than three independent treatments. This procedure uses rank in mixed sample combined with more than three unlike populations. This paper proposes a the new procedure based on joint placements for a one-way layout as extension of the joint placements described in Chung and Kim (2007). A Monte Carlo simulation study is adapted to compare the power of the proposed method with previous methods.

Keywords: one-way layout, joint placements, nonparametric

1. 서론

임상시험에서 류마티스 관절염 환자들에게 치료제인 비스테로이드항염제(NSAID), 항류마티스약제(DMARDs), 스테로이드(steroid)를 각각 처방했다고 가정하자. 치료제를 처방한 후 임상전문가는 환자가 느끼는 통증의 정도에 차이가 있는지 알아보기를 원할 것이다. 이때 세 개 이상의 처리 간에 효과차이를 검정하기 위한 방법으로 실험 계획법에는 일원배치법(one-way layout)이 있다. 일원배치법에서는 처리 간에 효과차이를 검정하기 위해 오차항은 서로 독립이고 정규분포 $N(0, \sigma^2)$를 따르는 확률변수라는 가정이 성립해야한다. 세 개 이상의 처리들이 모두 같다는 귀무가설을 모수적 방법인 분산분석법(ANOVA)으로 검정 할 수 있다. 그러나 오차항이 정규성 가정을 만족하지 못하게 되는 경우에는 제1종 오류를 제어할 수 없다. 이를 해결하는 하나의 방법으로 비모수적 검정법을 선택해야 한다.

일원배치법에서 대표적인 비모수 검정법은 Kruskal과 Wallis (1952)가 제안한 검정법이 있는데, 이 검정방법은 관측값 자체를 이용하는 대신에 이들 관측값의 통합순위를 매겨서 이용하는 것이 특징이다. Kim (1999)은 Orban과 Wolfe (1982)가 제안한 두 처리 중 어느 한 처리에 대한 상대적 위치정보를 이용해 처리효과 간의 차이를 검정하는 검정법을 일원배치모형으로 확장하여 Fixed Placement와 Updated Placement를 사용한 비모수적 검정방법을 제안하였다. 이 방법은 대조군과 처리군들 간의 상대적 위치정보를 이용한 Fixed Placement와 점차적으로 처리군들을 포함시켜가며 만든 대조군과 처리군 간의 상대적 위치정보를 이용한 Updated Placement를 이용하여 처리효과의 차이를 검정한다. 하지만 Kim (1999)이 제안한 검정법은 일원배치모형에서 모든 처리군에 대해 위치(placement)가 상대적인 위치정보를 이용하지 못한다는 단점이 있다. Chung과 Kim (2007)은 결합위치를 이용하여 일원배치모형

[1]Corresponding author: Department of Biomedicine · Health Science, The Catholic University of Korea, 222 Banpo-daero Seocho-gu, Seoul 06591, Korea. E-mail: djkim@catholic.ac.kr

에서 비모수 검정법을 제안하였다. 이 방법은 세 개 이상의 처리 중에서 자신이 포함되어 있는 처리군을 제외한 나머지 처리군을 대조군으로 보아 대조군의 관측값들에 대한 해당 관측값의 상대적인 위치정보를 이용하였고 소표본 분포를 구하여 처리효과간의 차이를 검정한다. 이 방법은 처리별 표본크기가 동일한 경우에 Kruskal-Wallis 검정법과 동일한 검정법이 된다.

본 논문에서는 일원배치모형에서 결합위치(joint placement)를 이용한 새로운 검정방법을 제안하였다. 귀무가설에 대한 일반 대립가설과 Chung과 Kim (2007)이 제안한 결합위치에 점수 함수를 적용시켜 만든 검정통계량을 제안하고, 그 검정통계량의 근사분포에 대해서도 언급하였다. 또한 결합위치를 이용해 새로이 제시한 검정법을 기존의 비모수 검정법인 Kruskal-Wallis 검정법 그리고 기존의 모수적 방법인 분산분석법(ANOVA)과 검정력(power)을 비교하였다.

2. 제안하는 방법

처리의 개수가 k인 일원배치법의 모형은

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad (i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, n_i)$$

이다. 여기서 Y_{ij}는 i번째 처리에서 j번째 추출한 확률표본이고 μ는 전체 평균을, α_i는 i번째 처리의 효과를 나타낸다. 또한 ϵ_{ij}는 오차항이며, 동일한 연속분포를 따르는 서로 독립인 확률변수로 가정하고 전체표본의 크기는 $N = \sum_{i=1}^{k} n_i$이라고 하자.

각 처리의 효과가 모두 동일하다는 귀무가설과 일반 대립가설은 다음과 같다.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_k \quad \text{vs.} \quad H_1 : \alpha_i\text{들이 모두 같지는 않다.}$$

Chung과 Kim (2007)이 제안한 각 처리군에서 i번째 처리의 j번째 표본 Y_{ij}에 대한 결합위치 V_{ij}는

$$V_{ij} = \frac{1}{N - n_i} \sum_{\substack{h=1 \\ h \neq i}}^{k} \sum_{s=1}^{n_h} \chi(Y_{hs}, Y_{ij}), \quad \text{여기서} \quad \chi(x, y) = \begin{cases} 1, & x \le y\text{일 경우}, \\ 0, & \text{그 외} \end{cases} \tag{2.1}$$

이고 i번째 처리의 관측값을 제외한 혼합표본에서 Y_{ij}보다 작거나 같은 관측값의 개수를 이용한 확률변수이다.

식 (2.1)을 이용한 선형 위치 통계량은

$$S_N = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \phi(V_{ij}) \tag{2.2}$$

와 같이 정의된다. 이때 식 (2.2)에서 $\phi(\bullet)$는 $[0, 1]$의 범위에서 정의된 실변수 함수인 점수함수(score function)이다.

Hong과 Lee (2014)에 의해, 귀무가설 하에서 각각의 i에 대해서 $N \to \infty$, $n_i/N \to \lambda_i$ $(0 < \lambda_i < 1)$이면 표준화된 통계량 $\hat{S}_N$의 분포는 표준정규분포로 수렴하게 된다. 즉,

$$\hat{S}_N = \frac{S_N - E(S_N)}{\sqrt{\mathrm{Var}(S_N)}} \xrightarrow{N \to \infty} N(0, 1)$$

이다.

본 논문에서는 점수 함수로 $\phi(x) = |1 - 2x|$를 고려한다. 이때, 점수 함수에 따른 선형 위치 통계량은 다음과 같다.

$$S_N^{\phi} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} |1 - 2V_{ij}|$$

귀무가설 하에서 검정통계량 S_N^{ϕ}의 근사분포는

$$\hat{S}_N^{\phi} = \frac{S_N^{\phi} - E\left(S_N^{\phi}\right)}{\sqrt{\mathrm{Var}\left(S_N^{\phi}\right)}} \xrightarrow{N \to \infty} N(0,1)$$

이고, 이를 이용하여 주어진 가설을 검정할 수 있다.

3. 모의실험 및 결과

본 논문에서는 일원배치모형에서 결합위치를 이용하여 점수함수를 적용시킨 검정통계량에 근거해 새롭게 제시한 방법과 기존의 검정법들과의 비교를 위해 모수적 방법과 비모수적 방법과의 검정력을 비교해보았다. 모수적인 방법으로는 F-통계량을 이용한 분산분석법을 사용하고, 비모수적인 방법으로 Kruskal과 Wallis (1952)가 제안한 검정법을 사용하였다.

모집단의 분포로는 정규분포, 지수분포, Cauchy분포, 이중지수분포를 고려하였다. SAS를 이용해 정규분포는 RANNOR함수, 지수분포는 RANEXP함수, Cauchy분포는 RANCAU함수를 이용하여 난수를 생성하였다. 그리고 이중지수분포는 RANUNI함수를 이용해 역변환 방법으로 난수를 생성하였다. 또한 유의수준 α는 0.05로 하였다.

처리의 개수는 각각 4개와 6개를 선택하였고, 처리별 표본크기는 모두 다른 경우와 모두 같은 경우를 선택하였으며 유의수준 α를 0.05로 보정하기 위해 확률화 검정을 실시하였다. 위와 같은 조건에서 각 검정법들에 대한 검정력을 비교하기 위해서 10,000번 반복하는 Monte Carlo 모의실험을 실시하였다. 또한 모의실험의 결과를 처리의 개수가 4개이고 처리별 표본크기가 모두 다른 경우에는 Table 3.1, 처리의 개수가 4개이고 처리별 표본크기가 모두 같은 경우에는 Table 3.2, 처리의 개수가 6개이고 처리별 표본크기가 모두 다른 경우에는 Table 3.3, 처리의 개수가 6개이고 처리별 표본크기가 모두 같은 경우에는 Table 3.4에 정리하였다.

먼저 각 처리효과간의 차이가 동일할 때 제1종 오류를 잘 제어할 수 있는지 확인해 보면 분산분석법은 정규분포에서 처리의 수나 처리별 표본크기에 관계없이 0.05 근처의 값을 얻었다. 이중지수분포와 지수분포에서는 처리가 4개이고 처리별 표본크기가 모두 다른 경우와 처리가 6개이고 처리별 표본크기에 관계없이 0.05에 근사한 값을 얻었지만 처리가 4개이고 처리별 표본크기가 모두 같은 경우($n_1 = 16, n_2 = 16, n_3 = 16, n_4 = 16$)에는 0.0407, 0.0393의 값을 얻은 것을 확인 할 수 있다. 이는 정규분포보다 이중지수분포와 지수분포에서는 분산분석법이 제1종 오류를 제어하는데 어려움이 존재한다는 것을 알 수 있다. Cauchy분포에서는 처리가 4개이고 처리별 표본크기가 모두 다른 경우 0.0235, 0.0195, 처리가 4개이고 처리별 표본크기가 모두 같은 경우 0.0150, 0.0151이고 처리가 6개이고 처리별 표본크기가 모두 다른 경우 0.0367, 0.0223, 처리가 6개이고 처리별 표본크기가 모두 같은 경우 0.0160, 0.0175의 값을 얻었다. 이는 Cauchy분포에서는 제1종 오류를 제어하기 힘들다는 것을 알 수 있다. 다음으로 Kruskal-Wallis 검정법은 정규분포, 이중지수분포, Cauchy분포, 지수분포가 처리의 수와 처리별 표본크기에 관계없이 0.05에 가까운 값을 얻었다. 이중지수분포와 지수분포에서는 처리가 4개이고 처리별 표본크기

Table 3.1. Monte Carlo power estimates ($\alpha = 0.05$, $k = 4$, unequal sample size)

Distribution	n_1	n_2	n_3	n_4	α_1	α_2	α_3	α_4	F	K-W	P
Normal	5	7	9	11	0	0	0	0	0.0498	0.0424	0.0500
					0	0	0	1	0.5427	0.4927	0.5375
					1.5	1	0.8	0	0.6564	0.6047	0.5275
					1	0	0	1	0.5797	0.5392	0.4824
					1	0.8	0.8	0	0.4133	0.3714	0.4001
					1.5	1	1.5	1	0.1727	0.1462	0.1448
					0.8	0.8	0.8	0	0.3639	0.3246	0.2278
					1.5	2	1	0	0.9439	0.9205	0.8709
	13	15	17	19	0	0	0	0	0.0468	0.0426	0.0499
					0	0	0	1	0.8632	0.8392	0.7674
					1.5	1	0.8	0	0.9581	0.9432	0.8455
					1	0	0	1	0.9178	0.8980	0.7843
					1	0.8	0.8	0	0.7375	0.7022	0.6221
					1.5	1	1.5	1	0.3454	0.3199	0.2546
					0.8	0.8	0.8	0	0.6532	0.6187	0.5671
					1.5	2	1	0	0.9994	0.9987	0.9910
Double exponential	5	7	9	11	0	0	0	0	0.0459	0.0440	0.0535
					0	0	0	1	0.3096	0.3571	0.5048
					1.5	1	0.8	0	0.3938	0.4456	0.4817
					1	0	0	1	0.3438	0.3945	0.4718
					1	0.8	0.8	0	0.2367	0.2722	0.3801
					1.5	1	1.5	1	0.1072	0.1143	0.1465
					0.8	0.8	0.8	0	0.2803	0.2396	0.3604
					1.5	2	1	0	0.6849	0.7445	0.7761
	13	15	17	19	0	0	0	0	0.0482	0.0469	0.0534
					0	0	0	1	0.5585	0.6670	0.7398
					1.5	1	0.8	0	0.7099	0.8101	0.7960
					1	0	0	1	0.6305	0.7548	0.7931
					1	0.8	0.8	0	0.4371	0.5489	0.6116
					1.5	1	1.5	1	0.1857	0.2438	0.2892
					0.8	0.8	0.8	0	0.3759	0.4695	0.5636
					1.5	2	1	0	0.9510	0.9817	0.9744
Cauchy	5	7	9	11	0	0	0	0	0.0235	0.0424	0.0535
					0	0	0	1	0.0468	0.1604	0.2947
					1.5	1	0.8	0	0.0510	0.1975	0.2793
					1	0	0	1	0.0515	0.1759	0.2622
					1	0.8	0.8	0	0.0392	0.1292	0.2151
					1.5	1	1.5	1	0.0286	0.0692	0.0944
					0.8	0.8	0.8	0	0.0354	0.1134	0.2050
					1.5	2	1	0	0.0851	0.3385	0.4896
	13	15	17	19	0	0	0	0	0.0195	0.0462	0.0552
					0	0	0	1	0.0396	0.3034	0.4602
					1.5	1	0.8	0	0.0521	0.4102	0.5262
					1	0	0	1	0.0441	0.3508	0.4878
					1	0.8	0.8	0	0.0335	0.2418	0.3616
					1.5	1	1.5	1	0.0211	0.1142	0.1589
					0.8	0.8	0.8	0	0.0300	0.2058	0.3222
					1.5	2	1	0	0.0827	0.6543	0.8061

continues

Distribution	n_1	n_2	n_3	n_4	α_1	α_2	α_3	α_4	F	K-W	P
Exponential	5	7	9	11	0	0	0	0	0.0452	0.0440	0.0535
					0	0	0	1	0.5942	0.7571	0.8779
					1.5	1	0.8	0	0.6811	0.7908	0.7216
					1	0	0	1	0.6364	0.7693	0.6842
					1	0.8	0.8	0	0.4712	0.5742	0.5537
					1.5	1	1.5	1	0.1961	0.2876	0.2704
					0.8	0.8	0.8	0	0.4143	0.5197	0.5170
					1.5	2	1	0	0.9181	0.9634	0.9868
	13	15	17	19	0	0	0	0	0.0478	0.0469	0.0534
					0	0	0	1	0.8622	0.9831	0.9900
					1.5	1	0.8	0	0.9394	0.9915	0.9867
					1	0	0	1	0.9111	0.9854	0.9188
					1	0.8	0.8	0	0.7501	0.8935	0.7792
					1.5	1	1.5	1	0.3601	0.6187	0.4464
					0.8	0.8	0.8	0	0.6681	0.8349	0.7002
					1.5	2	1	0	0.9966	0.9998	1.0000

∗ F: 분산분석법(ANOVA); K-W: Kruskal-Wallis 검정법; P: 결합위치를 이용한 검정법.

Table 3.2. Monte Carlo power estimates ($\alpha = 0.05$, $k = 4$, equal sample size)

Distribution	n_1	n_2	n_3	n_4	α_1	α_2	α_3	α_4	F	K-W	P
Normal	8	8	8	8	0	0	0	0	0.0528	0.0467	0.0482
					0	0	0	1	0.4567	0.4026	0.2858
					1.5	1	0.8	0	0.6570	0.6089	0.4387
					1	0	0	1	0.5839	0.5408	0.4295
					1	0.8	0.8	0	0.3634	0.3200	0.2292
					1.5	1	1.5	1	0.1698	0.1485	0.1196
					0.8	0.8	0.8	0	0.3024	0.2633	0.1920
					1.5	2	1	0	0.9172	0.8852	0.7297
	16	16	16	16	0	0	0	0	0.0483	0.0463	0.0502
					0	0	0	1	0.8121	0.7810	0.6156
					1.5	1	0.8	0	0.9515	0.9360	0.8132
					1	0	0	1	0.9193	0.9024	0.7826
					1	0.8	0.8	0	0.7070	0.6717	0.5028
					1.5	1	1.5	1	0.3359	0.3108	0.2395
					0.8	0.8	0.8	0	0.6030	0.5631	0.4213
					1.5	2	1	0	0.9997	0.9996	0.9815
Double exponential	8	8	8	8	0	0	0	0	0.0479	0.0429	0.0418
					0	0	0	1	0.2646	0.2976	0.2536
					1.5	1	0.8	0	0.3950	0.4444	0.3806
					1	0	0	1	0.3392	0.3929	0.4235
					1	0.8	0.8	0	0.2081	0.2397	0.2175
					1.5	1	1.5	1	0.1091	0.1227	0.1274
					0.8	0.8	0.8	0	0.1766	0.2005	0.1875
					1.5	2	1	0	0.6546	0.6994	0.6121

continues

Distribution	n_1	n_2	n_3	n_4	α_1	α_2	α_3	α_4	F	K-W	P
Double exponential	16	16	16	16	0	0	0	0	0.0407	0.0436	0.0497
					0	0	0	1	0.5028	0.6131	0.5647
					1.5	1	0.8	0	0.7088	0.8096	0.7585
					1	0	0	1	0.6303	0.7488	0.7805
					1	0.8	0.8	0	0.4050	0.5113	0.4777
					1.5	1	1.5	1	0.1902	0.2424	0.2814
					0.8	0.8	0.8	0	0.3421	0.4270	0.4109
					1.5	2	1	0	0.9313	0.9693	0.9445
Cauchy	8	8	8	8	0	0	0	0	0.0150	0.0406	0.0402
					0	0	0	1	0.0365	0.1380	0.1535
					1.5	1	0.8	0	0.0440	0.1936	0.2171
					1	0	0	1	0.0379	0.1677	0.2194
					1	0.8	0.8	0	0.0310	0.1185	0.1352
					1.5	1	1.5	1	0.0206	0.0769	0.0839
					0.8	0.8	0.8	0	0.0257	0.1057	0.1183
					1.5	2	1	0	0.0715	0.3175	0.3452
	16	16	16	16	0	0	0	0	0.0151	0.0485	0.0536
					0	0	0	1	0.0351	0.2759	0.3359
					1.5	1	0.8	0	0.0492	0.3999	0.4804
					1	0	0	1	0.0417	0.3598	0.4884
					1	0.8	0.8	0	0.0340	0.2275	0.2776
					1.5	1	1.5	1	0.0227	0.1164	0.1501
					0.8	0.8	0.8	0	0.0272	0.1912	0.2324
					1.5	2	1	0	0.0796	0.6246	0.7181
Exponential	8	8	8	8	0	0	0	0	0.0451	0.0429	0.0418
					0	0	0	1	0.5051	0.6490	0.8200
					1.5	1	0.8	0	0.6931	0.7978	0.7370
					1	0	0	1	0.6338	0.7700	0.6451
					1	0.8	0.8	0	0.4256	0.5265	0.2812
					1.5	1	1.5	1	0.1945	0.3014	0.1947
					0.8	0.8	0.8	0	0.3566	0.4496	0.2071
					1.5	2	1	0	0.9036	0.9447	0.9204
	16	16	16	16	0	0	0	0	0.0393	0.0436	0.0497
					0	0	0	1	0.8232	0.9700	0.9852
					1.5	1	0.8	0	0.9365	0.9899	0.9643
					1	0	0	1	0.9084	0.9842	0.9243
					1	0.8	0.8	0	0.7227	0.8610	0.5654
					1.5	1	1.5	1	0.3641	0.6222	0.3798
					0.8	0.8	0.8	0	0.6311	0.7891	0.4491
					1.5	2	1	0	0.9955	0.9998	0.9982

$*$ F: 분산분석법(ANOVA); K-W: Kruskal-Wallis 검정법; P: 결합위치를 이용한 검정법.

가 모두 다른 경우($n_1 = 13, n_2 = 15, n_3 = 17, n_4 = 19$)는 0.0469, 이 외의 값들은 대체로 0.04의 값을 갖기에 이중지수분포와 지수분포도 처리의 수와 처리별 표본크기에 관계없이 0.05에 근사한 값을 갖는다 할 수 있다. 또한 Cauchy분포에서는 0.05에 근사한 값을 얻어 분산분석법보다는 제1종 오류를 잘 제어하고 있다고 볼 수 있다. 마지막으로 결합위치를 이용한 검정법에서는 정규분포가 처리의 수와 처리별 표본크기에 관계없이 0.05에 가까운 값을 얻었다. 이중지수분포와 지수분포에서는 처리가

Table 3.3. Monte Carlo power estimates ($\alpha = 0.05$, $k = 6$, unequal sample size)

Distribution	n_1	n_2	n_3	n_4	n_5	n_6	α_1	α_2	α_3	α_4	α_5	α_6	F	K-W	P
Normal	5	7	9	11	13	15	0	0	0	0	0	0	0.0521	0.0415	0.0499
							0	1	1	1	1	0	0.7781	0.7341	0.4045
							0	0.5	0.5	0.5	0.8	1	0.3211	0.2761	0.1294
							1	0	0	0	0	1	0.7748	0.7366	0.4104
							1	0.8	0.5	0.5	0.5	0	0.3620	0.3211	0.1696
							1	1	1	1	0.5	0	0.6618	0.6170	0.3510
							0	0	0	0	0	1	0.6829	0.6373	0.4203
	10	12	14	16	18	20	0	0	0	0	0	0	0.0474	0.0423	0.0499
							0	1	1	1	1	0	0.9307	0.9157	0.4827
							0	0.5	0.5	0.5	0.8	1	0.5302	0.4830	0.1623
							1	0	0	0	0	1	0.9349	0.9180	0.4837
							1	0.8	0.5	0.5	0.5	0	0.5608	0.5175	0.1952
							1	1	1	1	0.5	0	0.8447	0.8161	0.4107
							0	0	0	0	0	1	0.8567	0.8247	0.4520
Double exponential	5	7	9	11	13	15	0	0	0	0	0	0	0.0488	0.0429	0.0505
							0	1	1	1	1	0	0.4544	0.5428	0.3929
							0	0.5	0.5	0.5	0.8	1	0.1760	0.2068	0.1407
							1	0	0	0	0	1	0.4544	0.5490	0.3864
							1	0.8	0.5	0.5	0.5	0	0.1971	0.2322	0.1659
							1	1	1	1	0.5	0	0.3624	0.4490	0.3367
							0	0	0	0	0	1	0.3872	0.4656	0.3909
	10	12	14	16	18	20	0	0	0	0	0	0	0.0458	0.0428	0.0537
							0	1	1	1	1	0	0.6543	0.7781	0.4663
							0	0.5	0.5	0.5	0.8	1	0.2743	0.3584	0.1726
							1	0	0	0	0	1	0.6446	0.7787	0.4623
							1	0.8	0.5	0.5	0.5	0	0.2995	0.3893	0.1974
							1	1	1	1	0.5	0	0.5162	0.6467	0.4036
							0	0	0	0	0	1	0.5327	0.6475	0.4148
Cauchy	5	7	9	11	13	15	0	0	0	0	0	0	0.0367	0.0454	0.0474
							0	1	1	1	1	0	0.0539	0.2287	0.2203
							0	0.5	0.5	0.5	0.8	1	0.0467	0.0987	0.0917
							1	0	0	0	0	1	0.0559	0.2293	0.2116
							1	0.8	0.5	0.5	0.5	0	0.0421	0.1054	0.1026
							1	1	1	1	0.5	0	0.0501	0.1835	0.1847
							0	0	0	0	0	1	0.0505	0.1952	0.2177
	10	12	14	16	18	20	0	0	0	0	0	0	0.0223	0.0491	0.0498
							0	1	1	1	1	0	0.0408	0.3576	0.2644
							0	0.5	0.5	0.5	0.8	1	0.0252	0.1515	0.1050
							1	0	0	0	0	1	0.0348	0.3590	0.2477
							1	0.8	0.5	0.5	0.5	0	0.0290	0.1646	0.1213
							1	1	1	1	0.5	0	0.0320	0.2828	0.2236
							0	0	0	0	0	1	0.0287	0.2704	0.2314
Exponential	5	7	9	11	13	15	0	0	0	0	0	0	0.0486	0.0429	0.0505
							0	1	1	1	1	0	0.7772	0.8934	0.4936
							0	0.5	0.5	0.5	0.8	1	0.3417	0.5792	0.1648
							1	0	0	0	0	1	0.7801	0.9398	0.8274
							1	0.8	0.5	0.5	0.5	0	0.3931	0.5634	0.3050
							1	1	1	1	0.5	0	0.6642	0.8338	0.5549
							0	0	0	0	0	1	0.7086	0.9127	0.8992
	10	12	14	16	18	20	0	0	0	0	0	0	0.0457	0.0428	0.0537
							0	1	1	1	1	0	0.9232	0.9827	0.5566
							0	0.5	0.5	0.5	0.8	1	0.5542	0.8219	0.2143
							1	0	0	0	0	1	0.9349	0.9970	0.9056
							1	0.8	0.5	0.5	0.5	0	0.5811	0.8291	0.3859
							1	1	1	1	0.5	0	0.8333	0.9547	0.6071
							0	0	0	0	0	1	0.8630	0.9880	0.9530

$*$ F: 분산분석법(ANOVA); K-W: Kruskal-Wallis 검정법; P: 결합위치를 이용한 검정법.

Table 3.4. Monte Carlo power estimates ($\alpha = 0.05$, $k = 6$, equal sample size)

Distribution	n_1	n_2	n_3	n_4	n_5	n_6	α_1	α_2	α_3	α_4	α_5	α_6	F	K-W	P
Normal	10	10	10	10	10	10	0	0	0	0	0	0	0.0563	0.0491	0.0485
							0	1	1	1	1	0	0.7764	0.7333	0.3528
							0	0.5	0.5	0.5	0.8	1	0.3798	0.3380	0.1436
							1	0	0	0	0	1	0.7767	0.7344	0.3310
							1	0.8	0.5	0.5	0.5	0	0.3788	0.3318	0.1350
							1	1	1	1	0.5	0	0.5649	0.5109	0.2070
							0	0	0	0	0	1	0.5310	0.4659	0.1673
	15	15	15	15	15	15	0	0	0	0	0	0	0.0467	0.0406	0.0496
							0	1	1	1	1	0	0.9389	0.9199	0.4220
							0	0.5	0.5	0.5	0.8	1	0.5675	0.5237	0.1864
							1	0	0	0	0	1	0.9370	0.9205	0.4259
							1	0.8	0.5	0.5	0.5	0	0.5709	0.5305	0.1912
							1	1	1	1	0.5	0	0.7899	0.7504	0.2703
							0	0	0	0	0	1	0.7661	0.7111	0.2360
Double exponential	10	10	10	10	10	10	0	0	0	0	0	0	0.0455	0.0436	0.0476
							0	1	1	1	1	0	0.4481	0.5525	0.3102
							0	0.5	0.5	0.5	0.8	1	0.2059	0.2478	0.1404
							1	0	0	0	0	1	0.4514	0.5458	0.3169
							1	0.8	0.5	0.5	0.5	0	0.2051	0.2555	0.1443
							1	1	1	1	0.5	0	0.3001	0.3674	0.1945
							0	0	0	0	0	1	0.2898	0.3374	0.1433
	15	15	15	15	15	15	0	0	0	0	0	0	0.0463	0.0422	0.0527
							0	1	1	1	1	0	0.6525	0.7773	0.4184
							0	0.5	0.5	0.5	0.8	1	0.3031	0.3955	0.1845
							1	0	0	0	0	1	0.6529	0.7758	0.4177
							1	0.8	0.5	0.5	0.5	0	0.3034	0.3958	0.1881
							1	1	1	1	0.5	0	0.4676	0.5812	0.2611
							0	0	0	0	0	1	0.4437	0.5338	0.1996
Cauchy	10	10	10	10	10	10	0	0	0	0	0	0	0.0160	0.0404	0.0428
							0	1	1	1	1	0	0.0301	0.2263	0.1659
							0	0.5	0.5	0.5	0.8	1	0.0252	0.1164	0.0903
							1	0	0	0	0	1	0.0288	0.2298	0.1729
							1	0.8	0.5	0.5	0.5	0	0.0191	0.1173	0.0914
							1	1	1	1	0.5	0	0.0247	0.1574	0.1158
							0	0	0	0	0	1	0.0255	0.1491	0.0970
	15	15	15	15	15	15	0	0	0	0	0	0	0.0175	0.0426	0.0562
							0	1	1	1	1	0	0.0301	0.3666	0.2359
							0	0.5	0.5	0.5	0.8	1	0.0221	0.1687	0.1233
							1	0	0	0	0	1	0.0324	0.3582	0.2263
							1	0.8	0.5	0.5	0.5	0	0.0201	0.1701	0.1217
							1	1	1	1	0.5	0	0.0252	0.2390	0.1481
							0	0	0	0	0	1	0.0249	0.2239	0.1264
Exponential	10	10	10	10	10	10	0	0	0	0	0	0	0.0454	0.0436	0.0475
							0	1	1	1	1	0	0.7765	0.8922	0.3522
							0	0.5	0.5	0.5	0.8	1	0.4096	0.6178	0.2357
							1	0	0	0	0	1	0.7810	0.9443	0.7542
							1	0.8	0.5	0.5	0.5	0	0.4112	0.6217	0.2388
							1	1	1	1	0.5	0	0.5955	0.7476	0.2261
							0	0	0	0	0	1	0.5575	0.7454	0.4265
	15	15	15	15	15	15	0	0	0	0	0	0	0.0437	0.0422	0.0527
							0	1	1	1	1	0	0.9228	0.9840	0.4787
							0	0.5	0.5	0.5	0.8	1	0.5917	0.8440	0.3093
							1	0	0	0	0	1	0.9319	0.9959	0.8648
							1	0.8	0.5	0.5	0.5	0	0.5970	0.8444	0.3160
							1	1	1	1	0.5	0	0.7884	0.9186	0.3001
							0	0	0	0	0	1	0.7752	0.9609	0.6239

$*$ F: 분산분석법(ANOVA); K-W: Kruskal-Wallis 검정법; P: 결합위치를 이용한 검정법.

4개이고 처리별 표본크기가 모두 다른 경우, 처리가 4개이고 처리별 표본크기가 모두 같은 경우($n_1 = 16, n_2 = 16, n_3 = 16, n_4 = 16$), 그리고 처리가 6개이고 처리별 표본크기에 관계없이 0.05 근처의 값을 얻었다. 그러나 처리가 4개이고 처리별 표본크기가 모두 같은 경우($n_1 = 8, n_2 = 8, n_3 = 8, n_4 = 8$)인 경우에는 두 분포에서 모두 0.0418의 값을 얻었지만 Cauchy분포에서도 0.05에 근사한 값을 얻었기에 실험유의수준으로 제1종 오류를 제어하는데 큰 문제가 없음을 알 수 있다.

이제 검정력을 살펴보면 다음과 같다. 정규분포에서 처리의 수와 처리별 표본의 크기와는 상관없이 분산분석법의 검정력이 가장 높다는 것을 알 수 있다. 그렇지만 처리가 4개이고 처리별 표본크기가 5, 7, 9, 11일 때 완만한 형태를 보이다가 증가하는 패턴과 계단형태를 보이며 감소하는 패턴의 대립가설은 제안한 검정법이 Kruskal-Wallis 검정법보다 검정력이 높게 나타났다. 또한 일정한 크기로 감소, 증가하는 패턴의 대립가설에서 제안한 검정법이 Kruskal-Wallis 검정법과 비슷한 검정력을 보였다. 처리가 6개일 때는 처리별 표본크기에 관계없이 분산분석법, Kruskal-Wallis 검정법, 제안한 검정법 순으로 검정력이 낮다는 것을 확인할 수 있다. 이중지수분포에서 처리가 4개이고 처리별 표본크기가 5, 7, 9, 11일 때 7개 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높았고 처리별 표본크기가 13, 15, 17, 19일 때 점차 감소하는 패턴과 증가하다 정점을 찍고 일정한 크기로 감소하는 패턴의 대립가설에서 제안한 검정법의 검정력이 Kruskal-Wallis 검정법 다음으로 높게 나온 것을 제외하고 나머지 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높다. 처리가 4개이고 처리별 표본크기가 8, 8, 8, 8일 때 감소하여 완만한 형태를 보이다가 증가하는 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높았고 일정한 크기로 감소, 증가하는 패턴의 대립가설에서 제안한 검정법이 Kruskal-Wallis 검정법과 비슷하지만 분산분석법보다는 검정력이 높다. 처리별 표본크기가 16, 16, 16, 16일 때 감소하여 완만한 형태를 보이다가 증가하는 패턴과 일정한 크기로 감소, 증가하는 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높고 나머지 패턴의 대립가설에 대해서는 Kruskal-Wallis 검정법이 검정력이 가장 높았고 다음으로 제안한 검정법의 검정력이 높았다. 처리가 6개이고 처리별 표본크기가 5, 7, 9, 11, 13, 15에서 완만한 형태를 보이다 증가하는 패턴의 대립가설이 Kruskal-Wallis 검정법보다는 검정력이 낮지만 분산분석법보다는 검정력이 높다. 처리가 6개이고 처리별 표본크기가 모두 같은 경우일 때 검정력은 모든 패턴의 대립가설에 대해 분산분석법, Kruskal-Wallis 검정법, 제안한 검정법 순으로 낮다. Cauchy분포에서 처리가 4개이고 처리별 표본크기에 관계없이 제안한 검정법이 7개 패턴의 대립가설이 모두 검정력이 가장 높게 나왔다. 처리가 6개이고 5, 7, 9, 11, 13, 15일 때 완만한 형태를 보이다가 증가하는 패턴의 대립가설은 제안한 검정법의 검정력이 가장 높지만 나머지 패턴의 대립가설에 대해서는 Kruskal-Wallis 검정법과 제안한 검정법의 검정력이 비슷하고 높게 나온 것을 알 수 있다. 또한 처리가 6개이고 처리별 표본크기가 모두 같은 경우일 때 모든 패턴의 대립가설에 대해 제안한 검정법이 Kruskal-Wallis 검정법 다음으로 검정력이 높다. 지수분포에서 처리가 4개이고 처리별 표본크기에 관계없이 완만한 형태를 보이다 증가하는 패턴과 증가하다 정점을 찍고 일정한 크기로 감소하는 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높다. 처리별 표본크기가 5, 7, 9, 11일 때 완만한 형태를 보이다가 감소하는 패턴의 대립가설에서 Kruskal-Wallis 검정법과 제안한 검정법의 검정력이 비슷한 것을 제외하고는 나머지 패턴의 대립가설에서는 Kruskal-Wallis 검정법보다는 검정력이 낮지만 분산분석법보다는 제안한 검정법의 검정력이 높다. 처리가 4개이고 처리별 표본크기가 8, 8, 8, 8과 16, 16, 16, 16일 때 완만한 형태를 보이다 증가하는 패턴의 대립가설에서 제안한 검정법의 검정력이 가장 높다. 그리고 점차 감소하는 패턴, 감소하여 완만한 형태를 보이다가 증가하는 패턴, 증가하다 정점을 찍고 일정한 크기로 감소하는 패턴의 대립가설에서 Kruskal-Wallis 검정법보다는 검정력이 낮지만 분산분석법보다는 제안한 검정법의 검정력이 높다. 또한 처리별 표본크기가 16, 16, 16, 16일 때에는 일정한 크기로 감소,증가하는 패턴의 대립가설에서 역시나 Kruskal-Wallis 검정법보다는 검정력이 낮지만 분산분석법보다는 제안한 검정법의 검정력이 높다. 처리가 6개이고 처리별 표본크기가 5, 7, 9, 11, 13, 15와 10, 12, 14, 16, 18,

20일 때 감소하여 완만한 형태를 보이다가 증가하는 패턴의 대립가설에서 분산분석법, 제안한 검정법, Kruskal-Wallis 검정법 순으로 검정력이 높다. 처리가 6개이고 처리별 표본크기가 10, 10, 10, 10, 10, 10와 15, 15, 15, 15, 15, 15일 때 모든 패턴에서 Kruskal-Wallis 검정법, 분산분석법, 제안한 검정법순으로 검정력이 낮아지는 것을 알 수 있다.

4. 결론 및 고찰

본 논문에서는 일원배치모형을 검정하기 위한 비모수적 방법으로 결합위치를 이용한 검정통계량을 제안하였다. 이 통계량은 Chung과 Kim (2007)의 논문에서 제안된 결합위치를 $[0,1]$의 범위에서 정의된 점수함수를 이용하여 만들었다. 모의실험을 통해 결합위치를 이용한 검정법의 검정력을 정규분포, 이중지수분포, Cauchy분포, 지수분포에서 모수적 검정법인 분산분석법, 비모수적 검정법인 Kruskal-Wallis 검정법을 사용하여 얻은 검정력과 비교하였다.

모의실험의 전체적인 결과를 살펴보면, 처리가 4개이고 처리별 표본크기가 다르고 총 표본크기가 작은 경우에는 정규분포에서 분산분석법의 검정력이 높지만 결합위치를 이용한 검정법도 분산분석법의 검정력과 비슷한 수준으로 나왔다. 나머지 3개의 분포에 대해서는 결합위치를 이용한 검정법이 대부분의 대립가설 형태에서 검정력이 높게 나왔다. 또한 처리별 표본크기가 동일한 경우에는 일부 대립가설 형태에서 제안한 검정법이 Kruskal-Wallis 검정법과 검정력이 비슷하거나 보다 높게 나왔음을 알 수 있다. 처리가 6개이고 총 표본크기가 매우 큰 경우에는 정규분포에서는 F-통계량을 이용한 분산분석법의 검정력이 높았다. 그리고 이중지수분포와 지수분포에서 Kruskal-Wallis 검정법의 검정력이 높았으며 Cauchy분포에서는 Kruskal-Wallis 검정법과 논문에서 제안한 검정법의 검정력이 비슷하게 나타났음을 보여주고 있다.

모의실험의 결과로 나타난 검정력의 경향을 생각해 보면 본 논문에서 제안한 결합위치를 이용한 검정방법은 처리의 수가 4개이고 처리별 표본크기가 모두 다른 경우에 Kruskal-Wallis 검정법보다 더 효율적이라고 말할 수 있을 것이다. 본 논문에서 처리가 6개인 경우 unequal sample size와 equal sample size 모두 결합위치를 이용한 검정법의 검정력이 대체로 다른 두 검정법보다 낮은 문제점을 보이고 있는데, 이는 처리의 수가 늘어남에 따라 제안한 검정법의 검정력이 낮은 것을 알 수 있다.

본 논문에서 제안한 점수함수를 이용한 선형 위치 통계량 이외에도 이차 함수 형태를 띄는 점수함수를 적용할 수 있을 것으로 보인다. 이차 함수 형태를 띄는 점수함수 외에도 적용할 수 있는 함수형태가 더 존재하는지 연구하여 결과를 살펴 볼 필요가 있다. 검정통계량이 표준정규분포로 수렴한다는 점을 토대로 본 논문에서 제안한 결합위치를 이용한 검정법은 일원배치법 이외에도 특정한 실험설계를 위한 유용한 비모수 검정방법이 될 것으로 기대된다.

References

Chung, T. and Kim, D. (2007). Nonparametric method using placement in one-way layout, *The Korean Communications in Statistics*, **14**, 551–560.

Hong, I. and Lee, S. (2014). Kruskal-Wallis one-way analysis of variance based on linear placements, *Korean Mathematical Society*, **51**, 701–716.

Kim, D. (1999). A class of distribution-free treatments versus control tests based on placements, *Far East Journal of Theoretical Statistics*, **3**, 19–33.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association*, **47**, 583–621.

Orban, J. and Wolfe, D. A. (1982). A class of distribution-free two-sample tests based on placements, *Journals of the American Statistical Association*, **77**, 666–671.

일원배치법에서 결합위치를 이용한 비모수 검정법

전경아[a] · 김동재[a,1]

[a]가톨릭대학교 의생명 · 건강과학과

(2016년 4월 6일 접수, 2016년 5월 16일 수정, 2016년 5월 18일 채택)

요 약

독립된 세 개 이상의 처리 간에 차이 유무를 검정하는 비모수적 방법에는 Kruskal과 Wallis (1952)가 제안한 검정법이 있다. 세 개 이상의 다른 모집단으로부터 결합된 표본관측 값들의 순위를 이용한 검정기법이다. 본 논문에서는 Chung과 Kim (2007)이 제안한 결합위치 방법을 확장하여 일원배치모형에서 새로운 방법을 제안하였다. 또한 모의실험(Monte Calro simulation study)를 통하여 기존의 검정법과 제안한 방법의 검정력을 비교하였다.

주요용어: 일원배치모형, 결합위치, 비모수 방법

[1]교신저자: (06591) 서울 서초구 반포대로 222, 의생명·건강과학과 가톨릭대학교.
E-mail: djkim@catholic.ac.kr