The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.7.1399
29(7), 1399–1409

Nonparametric method in randomized block design for umbrella alternatives based on aligned method and placement

Jeonghyun Kim[a] · Dongjae Kim[a,1]

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

(Received September 30, 2016; Revised November 11, 2016; Accepted November 18, 2016)

Abstract

Nonparametric methods in randomized block design were suggested by Friedman (1937) for general alternatives and were also proposed by Page (1963) for ordered alternatives in one-way layout; in addition, K-sample rank tests for umbrella alternatives were suggested by Mack and Wolfe (1981). In this paper, we proposed a nonparametric method of umbrella alternatives for randomized block design using the aligned method proposed by Hodges and Lehmann (1962) to use block information and using placement suggested by Kim (1999). Monte Carlo simulation was also adapted to compare the power of the proposed procedure with previous methods.

Keywords: randomized block design, aligned method, placement, nonparametric method, umbrella alternatives

1. 서론

랜덤화 블록 계획법(randomized block design)은 동질적인 실험단위끼리 여러 개의 블록(block)으로 나눈 후 각 블록의 한 개에 실험 단위에 처리 1을 적용하고, 나머지 중 다른 한 개의 실험단위에 처리 2 등으로 적용하는 방법이다. 이와 같이 함으로써 다른 블록 간에는 매우 다르다 하더라도, 서로 동질적인 같은 블록 내에서는 효과적으로 처리효과를 비교할 수 있다. 랜덤화 블록 계획법(randomized block design)은 처리와 블록의 두 인자가 있는 이원배치법이며 그러나 분석의 주된 관심은 처리효과라 하겠다 (Song과 Kim, 1997). 처리 별 효과의 차이가 있는지를 검정할 때, 모분포의 형태가 가정한 분포의 형태를 만족한다면 모수적인 방법인 분산분석법을 사용하여 처리 효과들이 모두 같다는 귀무가설을 검정할 수 있다. 하지만 이러한 가정을 만족하지 않을 시 발생하는 오류의 가능성이 작고, 최소한의 가정하에 개발된 비모수적인 방법을 사용해야 한다.

의약연구에서 일반적으로 더 높은 수준의 처리수준이 적용되면, 처리효과는 더 증가할 것이다. 그러나 실제로 연구대상이 높은 용량에서 독성반응이 나타난다고 알려졌다면, 그로 인해 처리 효과가 감소하게 된다. 이렇게 되면 처리 효과는 어느 용량까지 서서히 증가하고 그 뒤로는 서서히 감소하는 것으로 예

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

측된다. 이러한 처리 효과의 형태를 우산형 패턴(umbrella pattern)이라 한다. 또한, 두 개의 다른 순서형 그룹으로 치료 효과가 분리되는 점을 우산형 패턴의 정점(peak)이라고 한다. 이와 같은 환경에서는 대조군과 몇몇 처리군을 비교하기 위해서 우산형 대립가설에 관한 정보를 이용하는 실험과정이 더 선호된다. 이러한 우산형 패턴에 대한 사전정보가 있는 경우에는 일반대립가설 또는 순서형 대립가설보다 우산형 대립가설을 사용하는 것이 더 효율적이다.

랜덤화 블록 계획법(randomized block design)에서 비모수적인 방법으로 일반대립가설의 대표적인 검정법으로는 각 블록 내 순위(rank)를 사용하여 처리 간의 효과 차이를 검정하는 Friedman (1937)이 제안한 검정법이 있고, 순서형대립가설에는 블록 내 순위(rank)를 사용하여 처리 별로 가중치를 적용하여 통계량을 계산하는 Page (1963)의 검정법이 있다. 하지만 두 가지 검정법 모두 블록 내 순위를 사용하여 블록 간의 정보를 사용하지 못하여 검정법의 효율이 떨어진다는 단점이 있다. 우산형 대립가설에 대한 검정법은 여러 가지 방법들이 제안되었다. 우산형 대립가설에 대한 대표적인 방법으로는 일원 배치 모형에서 k개의 표본 문제에 대하여 Mann과 Whitney (1947)가 제안한 U-통계량을 확장한 Mack과 Wolfe (1981)의 검정법이 있다. 그리고 k개의 표본 문제에 대해 순위통계량을 사용한 Hettmansperger와 Norton (1987)의 검정법이 있고, Mack과 Wolfe (1981)의 검정법을 확장한 Chen과 Wolfe (1990)의 검정법과 Chen과 Wolfe (1993), Alvo (2008), Bhat (2009), Salman (2010) 등이 제안한 방법들이 있다. 또한, Kim와 Kim (2016)은 랜덤화 블록 모형에서 Hodges와 Lehmann (1962)의 정렬방법과 Kim (1999)이 일원배치 모형에 대하여 제안한 대조군과 처리군 방법을 적용한 순서대립가설에서의 비모수적인 검정법을 제안하였다.

Orban과 Wolfe (1982)는 두 개의 처리 간 효과 차이를 검정하기 위해 위치(placement)를 이용한 비모수적 검정 방법을 제안하였다. 이는 두 처리 중 어느 한 처리에 대한 상대적 위치정보를 사용해 처리 효과 차이를 검정하는 방법으로써, 대조군의 표본 수가 처리군의 표본 수보다 클 때 더 유용하다고 알려져 있다. 또한, Kim (1999)은 이것을 확장하여 위치를 기반으로 한 대조군과 처리군의 비교 방법을 제안하였다 (Lee와 Kim, 2011).

본 논문에서는 랜덤화 블록 계획법(randomized block design)에서 블록 간의 정보를 사용하기 위하여 Hodges와 Lehmann (1962)이 제안한 정렬방법을 이용하고, Kim (1999)의 방법을 사용하여 대조군과 처리군에 대한 우산형대립가설에서의 새로운 비모수적 검정법을 제안하였다. 그리고 Monte Carlo 모의실험을 시행하여 모수적 검정법인 분산분석법, 비모수적 검정법에서는 Friedman (1937) 방법과 제안한 검정법의 검정력(power)을 비교하였다.

2. 제안하는 방법

2.1. 모형

블록이 있고 처리가 k개인 랜덤화 블록 계획법의 모형

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

에서 Y_{ij}는 i번째 처리에서 j번째 반응값이고, μ는 전체평균, θ_i는 i번째 처리효과를 나타낸다. 또한, ϵ_{ij}는 오차항이며, 동일한 연속분포를 따르는 서로 독립인 확률 변수로 가정한다. Hodges와 Lehmann (1962)이 제안한 정렬 방법으로 블록간의 정보를 이용한 정렬자료는

$$Y_{ij}^* = Y_{ij} - \bar{Y}_{\cdot j}$$

이고, $\bar{Y}_{\cdot j} = \sum_{j=1}^{n} Y_{ij}/n$는 각 블록 효과인 블록의 평균이다. 우산형 대립가설의 형태는 다음과 같다.

$$H_1 : \theta_1 \le \theta_2 \le \cdots \le \theta_{p-1} \le \theta_p \ge \theta_{p+1} \ge \cdots \ge \theta_{k-1} \ge \theta_k$$
(최소한 하나는 부등호)

2.2. Updated Control Group 검정법

Updated Placement를 이용하여 우산형 대립가설을 검정하기 위해서는 정점(peak) p를 기준으로 하여 두 개의 대립가설인 왼쪽 순서형 대립가설($H_L : \theta_1 \le \theta_2 \le \cdots \le \theta_{p-1} \le \theta_p$)과 오른쪽 순서형 대립가설($H_R : \theta_p \ge \theta_{p+1} \ge \cdots \ge \theta_{k-1} \ge \theta_k$)으로 나누어 검정해야 한다. 왼쪽 순서형 대립가설의 각 처리 군에서 i번째 처리의 j번째 블록 표본 Y_{ij}의 Updated Placement U_{ij}^L은

$$N_i^L U_{ij}^L = \sum_{h=1}^{i-1}\sum_{s=1}^{n} \chi(Y_{hs}, Y_{ij}), \quad \chi(x,y) = \begin{cases} 1, & x \le y, \\ 0, & \text{otherwise} \end{cases}$$
$$= [(i-1)\text{번째까지 처리군에서 } Y_{ij}\text{보다 작거나 같은 표본의 갯수}]$$

라고 정의한다. 여기서 $i = 2, \ldots, p;\ j = 1, \ldots, n$이고, $N_i^L = n_1 + \cdots + n_{i-1} = n \times (i-1)$이다.

또한, 오른쪽 순서형 대립가설에 대한 $Y_{k-i+1,j}$의 Updated Placement $U_{k-i+1,j}^R$은

$$N_{k-i+1}^R U_{k-i+1,j}^R = \sum_{h=1}^{i-1}\sum_{s=1}^{n} \chi(Y_{k-h+1,s}, Y_{k-i+1,j}), \quad \chi(x,y) = \begin{cases} 1, & x \le y \\ 0, & \text{otherwise} \end{cases}$$
$$= [k\text{번째부터 } (k\text{-}i\text{+}2)\text{번째까지 처리군에서 } Y_{k-i+1,j}\text{보다 작거나 같은 표본의 갯수}]$$

라고 정의한다. 여기서 $i = 2, \ldots, k-p+1;\ j = 1, \ldots, n$이고, $N_{k-i+1}^R = n_k + \cdots + n_{k-i+2}$이다.

가설 검정을 위한 Kim (1999)이 제안한 방법 중에 Updated Placement를 이용한 방법을 왼쪽과 오른쪽 순서형 대립가설에 적용한 검정통계량은 각각

$$S_{UP}^L = \sum_{i=2}^{p}\sum_{j=1}^{n} \phi_{N_i}(U_{ij}),$$
$$S_{UP}^R = \sum_{i=2}^{k-p+1}\sum_{j=1}^{n} \phi_{N_{k-i+1}}(U_{k-i+1,j})$$

와 같이 정의된다. 이 때 ϕ_{N_i}는 [0, 1]의 범위에서 정의된 실수값인 점수함수(score function)이다. 서로독립인 U_{ij}^L과 $U_{k-i+1,j}^R$의 통계량을 이용하는 우산형 대립가설의 검정통계량은

$$S_{UP} = S_{UP}^L + S_{UP}^R = \sum_{i=2}^{p}\sum_{j=1}^{n} \phi_{N_i}(U_{ij}) + \sum_{i=2}^{k-p+1}\sum_{j=1}^{n} \phi_{N_{k-i+1}}(U_{k-i+1,j})$$

이며, 귀무가설 $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k$을 검정하기 위한 기각역은 $S_{UP} \ge S_\alpha$이다.

본 논문에서는 이용한 점수함수로는 정규점수함수(normal score function) $\phi_{N_i}(\mathrm{x})=\Phi^{-1}(x)$ ($\Phi^{-1}(x)$는 표준정규분포 누적분포함수의 역함수와 지수점수함수(exponential score function) $\phi_{N_i}(x) = -\ln(1-x)$를 이용한 방법을 제안한다.

정규점수함수(normal score function)와 지수점수함수(exponential score function)를 이용한 통계량은

$$S_{UP}^{NS} = \sum_{i=2}^{p}\sum_{j=1}^{n}\Phi^{-1}\left(\frac{N_i^L U_{ij}^L + 1}{N_i^L + 2}\right) + \sum_{i=2}^{k-p+1}\sum_{j=1}^{n}\Phi^{-1}\left(\frac{N_{k-i+1}^R U_{k-i+1,j}^R + 1}{N_{k-i+1}^R + 2}\right)$$

$$S_{UP}^{E} = -\sum_{i=2}^{p}\sum_{j=1}^{n}\ln\left(1 - \frac{N_i^L U_{ij}^L}{N_i^L + 1}\right) - \sum_{i=2}^{k-p+1}\sum_{j=1}^{n}\ln\left(1 - \frac{N_{k-i+1}^R U_{k-i+1,j}^R}{N_{k-i+1}^R + 1}\right)$$

와 같이 정의되며, 여기서 Kim 등 (2011)이 제안한 표준화된 S_{UP} 통계량인 S_{UP}^{L*}와 S_{UP}^{R*}의 근사분포 중 왼쪽은

$$S_{UP}^{L*} = \frac{S_{UP}^L - E\left(S_{UP}^L\right)}{\sqrt{n}} \sim N\left(0, \sigma_{UP,L}^2\right)$$

이고, 이때 $n \to \infty$, $\rho_n = 1/i \to \rho_i$, $0 < \rho_i < \infty$, $\tilde{\rho}_i$=$1/(1+\rho_1+\cdots+\rho_{i-1})$이고, 분산 $\sigma_{UP,L}^2$는

$$\sigma_{UP,L}^2 = \sum_{i=2}^{p}\tilde{\rho}_i(\tilde{\rho}_i+1)\sigma_\phi^2,\ \sigma_\phi^2 = \int_0^1 \phi_{N_i}^2(x)dx - \left(\int_0^1 \phi_{N_i}^2(x)\right)^2$$

이다. 또한, 오른쪽은

$$S_{UP}^{R*} = \frac{S_{UP}^R - E\left(S_{UP}^R\right)}{\sqrt{n}} \sim N\left(0, \sigma_{UP,R}^2\right)$$

이고, 이때 $n \to \infty$, $\rho_n = 1/i \to \rho_i$, $0 < \rho_i < \infty$, $\tilde{\rho}_i$=$1/(1+\rho_k+\cdots+\rho_{k-i+2})$이고, 분산 $\sigma_{UP,R}^2$는

$$\sigma_{UP,R}^2 = \sum_{i=2}^{k-p+1}\tilde{\rho}_i(\tilde{\rho}_i+1)\sigma_\phi^2, \quad \sigma_\phi^2 = \int_0^1 \phi_{N_{k-i+1}}^2(x)dx - \left(\int_0^1 \phi_{N_{k-i+1}}^2(x)\right)^2$$

이다. 또 S_{UP}^{L*}과 S_{UP}^{R*} 통계량이 각각 독립이고 근사적으로 정규분포를 따르기 때문에

$$S_{UP} = S_{UP}^{L*} + S_{UP}^{R*} \sim N\left(0, \left(\sigma_{UP,L}^2 + \sigma_{UP,R}^2\right)\right)$$

이고, 이를 이용하여 검정할 수 있다.

3. 모의실험 및 결과

본 논문에서는 랜덤화 블록 계획법(randomized block design)에서 정렬방법과 위치를 이용한 정규점수함수(normal score function)와 지수점수함수(exponential score function)의 검정통계량에 근거하여 새롭게 제안한 방법과 기존의 모수적인 방법인 분산분석법, 비모수적 방법인 Friedman (1937)이 제안한 방법과의 차이를 비교했다. 모집단의 분포로는 정규분포, 비대칭적인 지수분포, 정규분포보다 꼬리가 두꺼운 Cauchy 분포, 그리고 정규분포보다 꼬리가 약간 두꺼운 이중지수분포를 사용하였다. SAS를 이용하여 정규분포의 난수 생성은 RANNOR 함수를 이용하였고, 지수분포의 난수 생성은 RANEXP 함수, Cauchy 분포의 난수 생성은 RANCAU 함수, 그리고 이중지수분포의 난수 생성은 RANUNI 함수를 이용하여 역변환 방법으로 난수를 생성하였다. 그리고 유의수준 α는 0.05로 하였다.

처리의 수는 3개와 6개의 두 가지 경우를 선택하였다. 처리가 3개인 경우에서는 블록의 크기가 5개인 경우와 15개인 경우를 고려하였다. 처리의 수가 6개인 경우에서는 정점(peak)이 3인 경우와 4인 경우로 두 가지를 선택하였고, 각각 블록의 크기는 5개인 경우와 10개인 경우를 고려하였다. 그리고 유의수준 α를 0.05로 보정하기 위하여 확률화검정을 실시하였고, 이러한 조건에서 각 통계량에 대한 검정력을 비교하기 위하여 10,000번을 반복하는 Monte Carlo 모의실험을 시행하였다. 모의실험의 결과는

Table 3.1. Monte Carlo power estimates: $\alpha = 0.05$, $k = 3$, block $= 5, 15$

DIST	n	θ_1	θ_2	θ_3	ANO	FRI	UNS	UE
Normal	5	0.0	0.0	0.0	0.0491	0.0405	0.0501	0.0495
		0.0	1.7	1.7	0.6314	0.4548	0.2364	0.2568
		1.5	1.5	0.0	0.5119	0.3745	0.2144	0.2309
		0.0	1.6	0.0	0.5761	0.4198	0.8264	0.8211
	15	0.0	0.0	0.0	0.0495	0.0457	0.0500	0.0502
		0.0	1.1	1.1	0.8474	0.6979	0.4543	0.5106
		1.0	1.0	0.0	0.7649	0.6114	0.4061	0.4503
		0.0	0.8	0.0	0.5616	0.4267	0.7878	0.7717
Exponential	5	0.0	0.0	0.0	0.0358	0.0388	0.0498	0.0490
		0.0	1.7	1.7	0.6781	0.5785	0.2308	0.2556
		1.5	1.5	0.0	0.5898	0.5095	0.2212	0.2415
		0.0	1.6	0.0	0.6400	0.4835	0.8502	0.8257
	15	0.0	0.0	0.0	0.0425	0.0440	0.0495	0.0512
		0.0	1.1	1.1	0.8409	0.8468	0.4835	0.5279
		1.0	1.0	0.0	0.7755	0.7999	0.4433	0.4749
		0.0	0.8	0.0	0.5952	0.6947	0.8515	0.7968
Cauchy	5	0.0	0.0	0.0	0.0172	0.0395	0.0489	0.0490
		0.0	1.7	1.7	0.0745	0.1465	0.1243	0.1275
		1.5	1.5	0.0	0.0611	0.1194	0.1148	0.1141
		0.0	1.6	0.0	0.0704	0.1285	0.2633	0.2601
	15	0.0	0.0	0.0	0.0171	0.0481	0.0508	0.0500
		0.0	1.1	1.1	0.0510	0.2371	0.1301	0.1290
		1.0	1.0	0.0	0.0431	0.2055	0.1199	0.1171
		0.0	0.8	0.0	0.0317	0.1558	0.2045	0.1987
Double Exponential	5	0.0	0.0	0.0	0.0427	0.0440	0.0493	0.0493
		0.0	1.7	1.7	0.3930	0.3167	0.1979	0.2208
		1.5	1.5	0.0	0.3197	0.2557	0.1781	0.1975
		0.0	1.6	0.0	0.3566	0.2866	0.6218	0.6307
	15	0.0	0.0	0.0	0.0465	0.0453	0.0496	0.0498
		0.0	1.1	1.1	0.5597	0.5404	0.3365	0.3699
		1.0	1.0	0.0	0.4796	0.4669	0.2987	0.3304
		0.0	0.8	0.0	0.3274	0.3245	0.5883	0.5820

ANO = ANOVA for randomized block design; FRI = Friedman이 제안한 방법; UNS = 정규점수함수를 이용한 선형위치통계량; UE=지수점수함수를 이용한 선형위치통계량.

처리가 3개이고 블록이 5개인 경우와 10개인 경우의 실험결과는 Table 3.1에 제시하였고, 처리가 6개이고 정점(peak)이 3이며, 블록이 5개와 10개인 경우의 실험결과는 Table 3.2에, 처리가 6개이고 정점(peak)이 4이며, 블록이 5개와 10개인 경우의 실험결과는 Table 3.3에 제시하였다.

각 처리의 효과가 모두 같을 경우에 유의수준 α가 0.05를 만족하는지 보면, 분산분석법의 실험유의수준이 처리가 3개이고 블록의 수가 5개, 15개일 때, 처리가 6개이고 정점(peak)이 3, 4이며 블록의 크기가 5, 10개인 경우에 정규분포에서는 모두 0.05에 가깝고, 지수분포에서는 0.04와 0.045 근방의 값이어서 정규분포보다는 대체로 낮았고, 처리가 3이고 블록이 5개인 경우일 때 0.0358로 더욱 낮았다. 이중지수분포에서는 0.045 근방의 값을 얻어 정규분포보다는 낮은 것을 알 수 있다. 특히 Cauchy 분포에서는 Table 3.1, Table 3.2, Table 3.3 순으로 0.0172, 0.0171, 0.0170, 0.0177, 0.0175, 0.0177의 값으로 1종

Table 3.2. Monte Carlo power estimates: $\alpha = 0.05$, $k = 6$, $p = 3$, block $= 5, 10$

DIST	n	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	ANO	FRI	UNS	UE
Normal	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0504	0.0326	0.0499	0.0501
		0.0	0.5	1.0	0.7	0.4	0.0	0.2230	0.1495	0.5792	0.5584
		0.0	1.0	2.0	1.5	1.0	0.0	0.7797	0.6322	0.9786	0.9697
		0.0	1.0	1.8	1.5	1	0.5	0.5760	0.4249	0.8786	0.8622
		0.0	0.5	1.0	0.0	0.0	0.0	0.2423	0.1627	0.4436	0.4740
		0.0	1.5	1.5	1.5	1.5	0.0	0.7425	0.5918	0.8956	0.8464
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0504	0.0463	0.0499	0.0501
		0.0	0.4	0.8	0.3	0.1	0.0	0.3077	0.2336	0.6340	0.6227
		0.0	0.5	1.0	0.7	0.4	0.0	0.4873	0.3886	0.8537	0.8239
		0.0	0.7	1.4	1.0	0.8	0.4	0.6918	0.5709	0.9268	0.9097
		0.0	0.5	1.0	0.0	0.0	0.0	0.5466	0.4238	0.7065	0.7321
		0.0	0.5	0.5	0.5	0.5	0.0	0.2241	0.1749	0.4566	0.4176
Exponential	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0428	0.0337	0.0500	0.0501
		0.0	0.5	1.0	0.7	0.4	0.0	0.2495	0.2839	0.7316	0.6235
		0.0	1.0	2.0	1.5	1.0	0.0	0.7935	0.7871	0.9892	0.9669
		0.0	1.0	1.8	1.5	1.0	0.5	0.6149	0.6267	0.9402	0.8841
		0.0	0.5	1.0	0.0	0.0	0.0	0.2771	0.2727	0.5375	0.5100
		0.0	1.5	1.5	1.5	1.5	0.0	0.7689	0.7156	0.9168	0.8315
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0400	0.0454	0.0502	0.0511
		0.0	0.4	0.8	0.3	0.1	0.0	0.3264	0.4697	0.7670	0.6537
		0.0	0.5	1.0	0.7	0.4	0.0	0.5269	0.6892	0.9358	0.8457
		0.0	0.7	1.4	1.0	0.8	0.4	0.7070	0.8304	0.9699	0.9157
		0.0	0.5	1.0	0.0	0.0	0.0	0.5795	0.7300	0.8010	0.7448
		0.0	0.5	0.5	0.5	0.5	0.0	0.2408	0.3587	0.5819	0.4594
Cauchy	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0170	0.0344	0.0503	0.0506
		0.0	0.5	1.0	0.7	0.4	0.0	0.0248	0.0615	0.1830	0.1722
		0.0	1.0	2.0	1.5	1.0	0.0	0.0500	0.1597	0.3764	0.3565
		0.0	1.0	1.8	1.5	1.0	0.5	0.0364	0.1131	0.2810	0.2624
		0.0	0.5	1.0	0.0	0.0	0.0	0.0223	0.0591	0.1413	0.1384
		0.0	1.5	1.5	1.5	1.5	0.0	0.0459	0.1522	0.2970	0.2705
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0177	0.0448	0.0494	0.0507
		0.0	0.4	0.8	0.3	0.1	0.0	0.0211	0.0881	0.1608	0.1535
		0.0	0.5	1.0	0.7	0.4	0.0	0.0239	0.1279	0.2277	0.2133
		0.0	0.7	1.4	1.0	0.8	0.4	0.0293	0.1846	0.2713	0.2552
		0.0	0.5	1.0	0.0	0.0	0.0	0.0238	0.1375	0.1789	0.1776
		0.0	0.5	0.5	0.5	0.5	0.0	0.0208	0.0781	0.1335	0.1286
Double Exponential	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0445	0.0325	0.0498	0.0502
		0.0	0.5	1.0	0.7	0.4	0.0	0.1304	0.1103	0.4305	0.4146
		0.0	1.0	2.0	1.5	1.0	0.0	0.4759	0.4207	0.8825	0.8634
		0.0	1.0	1.8	1.5	1.0	0.5	0.3256	0.2857	0.7054	0.6868
		0.0	0.5	1.0	0.0	0.0	0.0	0.1473	0.1226	0.3338	0.3487
		0.0	1.5	1.5	1.5	1.5	0.0	0.4458	0.3936	0.7371	0.6734
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0472	0.0435	0.0503	0.0502
		0.0	0.4	0.8	0.3	0.1	0.0	0.1659	0.1753	0.4606	0.4396
		0.0	0.5	1.0	0.7	0.4	0.0	0.2681	0.2860	0.6699	0.6257
		0.0	0.7	1.4	1.0	0.8	0.4	0.3927	0.4174	0.7699	0.7317
		0.0	0.5	1.0	0.0	0.0	0.0	0.2992	0.3112	0.5291	0.5222
		0.0	0.5	0.5	0.5	0.5	0.0	0.1271	0.1400	0.3404	0.3051

ANO = ANOVA for randomized block design; FRI = Friedman이 제안한 방법; UNS = 정규점수함수를 이용한 선형위치통계량; UE=지수점수함수를 이용한 선형위치통계량.

Table 3.3. Monte Carlo power estimates: $\alpha = 0.05$, $k = 6$, $p = 4$, $\text{block} = 5, 10$

DIST	n	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	ANO	FRI	UNS	UE
Normal	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0512	0.0352	0.0501	0.0500
		0.0	0.1	0.3	0.8	0.4	0.0	0.1461	0.0918	0.3944	0.3922
		0.0	0.4	0.7	1.0	0.5	0.0	0.2195	0.1425	0.5881	0.5578
		0.0	0.6	1.0	1.5	0.8	0.3	0.3925	0.2627	0.7967	0.7725
		0.0	0.0	0.0	1.0	0.5	0.0	0.2470	0.1617	0.4487	0.4755
		0.0	0.5	0.5	0.5	0.5	0.0	0.1157	0.0770	0.2691	0.2389
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0501	0.0448	0.0500	0.0500
		0.0	0.1	0.3	0.8	0.4	0.0	0.3008	0.2325	0.6057	0.5978
		0.0	0.4	0.7	1.0	0.5	0.0	0.4915	0.3938	0.8383	0.8111
		0.0	0.6	1.0	1.3	0.8	0.5	0.6181	0.4931	0.8844	0.8589
		0.0	0.0	0.0	1.0	0.5	0.0	0.5508	0.4275	0.6873	0.7204
		0.0	0.5	0.5	0.5	0.5	0.0	0.2002	0.1672	0.4362	0.3963
Exponential	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0463	0.0356	0.0495	0.0510
		0.0	0.1	0.3	0.8	0.4	0.0	0.1561	0.1674	0.5097	0.4331
		0.0	0.4	0.7	1.0	0.5	0.0	0.2587	0.2944	0.7222	0.6156
		0.0	0.6	1.0	1.5	0.8	0.3	0.4422	0.4620	0.8903	0.8098
		0.0	0.0	0.0	1.0	0.5	0.0	0.2855	0.2789	0.5414	0.5082
		0.0	0.5	0.5	0.5	0.5	0.0	0.1195	0.1511	0.3700	0.2963
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0464	0.0440	0.0498	0.0501
		0.0	0.1	0.3	0.8	0.4	0.0	0.3274	0.4677	0.7575	0.6235
		0.0	0.4	0.7	1.0	0.5	0.0	0.5234	0.6830	0.9318	0.8251
		0.0	0.6	1.0	1.3	0.8	0.5	0.6350	0.7690	0.9550	0.8660
		0.0	0.0	0.0	1.0	0.5	0.0	0.5750	0.7223	0.7834	0.7241
		0.0	0.5	0.5	0.5	0.5	0.0	0.2435	0.3596	0.5857	0.4333
Cauchy	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0175	0.0331	0.0509	0.0495
		0.0	0.1	0.3	0.8	0.4	0.0	0.0192	0.0481	0.1328	0.1225
		0.0	0.4	0.7	1.0	0.5	0.0	0.0236	0.0631	0.1701	0.1592
		0.0	0.6	1.0	1.5	0.8	0.3	0.0278	0.0839	0.2270	0.2122
		0.0	0.0	0.0	1.0	0.5	0.0	0.0250	0.0644	0.1360	0.1294
		0.0	0.5	0.5	0.5	0.5	0.0	0.0194	0.0464	0.1001	0.0931
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0177	0.0438	0.0494	0.0503
		0.0	0.1	0.3	0.8	0.4	0.0	0.0217	0.0943	0.1614	0.1546
		0.0	0.4	0.7	1.0	0.5	0.0	0.0233	0.1304	0.2304	0.2178
		0.0	0.6	1.0	1.3	0.8	0.5	0.0276	0.1449	0.2480	0.2349
		0.0	0.0	0.0	1.0	0.5	0.0	0.0252	0.1377	0.1829	0.1779
		0.0	0.5	0.5	0.5	0.5	0.0	0.0183	0.0708	0.1263	0.1251
Double Exponential	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0440	0.0338	0.0499	0.0499
		0.0	0.1	0.3	0.8	0.4	0.0	0.0999	0.0864	0.3005	0.2927
		0.0	0.4	0.7	1.0	0.5	0.0	0.1361	0.1194	0.4381	0.4147
		0.0	0.6	1.0	1.5	0.8	0.3	0.2131	0.1905	0.6259	0.5991
		0.0	0.0	0.0	1.0	0.5	0.0	0.1469	0.1244	0.3293	0.3405
		0.0	0.5	0.5	0.5	0.5	0.0	0.0810	0.0663	0.2045	0.1849
	10	0.0	0.0	0.0	0.0	0.0	0.0	0.0487	0.0439	0.0489	0.0494
		0.0	0.1	0.3	0.8	0.4	0.0	0.1721	0.1877	0.4399	0.4283
		0.0	0.4	0.7	1.0	0.5	0.0	0.2617	0.2874	0.6462	0.6076
		0.0	0.6	1.0	1.3	0.8	0.5	0.3392	0.2548	0.7115	0.6758
		0.0	0.0	0.0	1.0	0.5	0.0	0.2966	0.3156	0.4947	0.5097
		0.0	0.5	0.5	0.5	0.5	0.0	0.1276	0.1360	0.3109	0.2861

ANO = ANOVA for randomized block design; FRI = Friedman이 제안한 방법; UNS = 정규점수함수를 이용한 선형위치통계량; UE=지수점수함수를 이용한 선형위치통계량.

오류를 제어하지 못하는 것을 확인할 수 있다. 다음으로 Friedman (1937)의 검정법을 살펴보면 처리가 3개이거나, 처리가 6개이고 정점(peak)이 3, 4인 경우 0.04에서 0.045 근방의 값을 가졌으며, 모두 처리마다 블록의 수가 커질수록 Friedman (1937)의 방법이 0.05에 가까워졌다. 본 논문에서 제안한 정규점수함수(normal score function)와 지수점수함수(exponential score function)를 이용한 UNS와 UE는 모든 분포와 처리, 블록의 크기에 상관없이 0.05 근방에 형성되어 큰 문제는 없다는 것을 보여주고 있다.

검정력을 살펴보면 대립가설의 형태에 따라 다르다는 것을 알 수 있다. 먼저 Table 3.1을 살펴보면, 처리가 3이고 정규분포에서 블록이 5개인 경우에는 점점 커져서 정점(peak) 이후에 평행하거나 평행하다가 정점(peak) 이후에 작아지는 대립가설에서 분산분석법의 검정력이 가장 높았고, 정점(peak)까지 커지다가 다시 작아지는 뚜렷한 우산형 대립가설에서는 제안한 검정력이 가장 높았다. 지수분포에서는 대립가설의 형태에 따라 정규분포와 유사한 결과를 볼 수 있다. Cauchy 분포에서는 점점 커져서 정점(peak) 이후에 평행하거나 평행하다가 정점(peak) 이후에 작아지는 대립가설에서 Friedman (1937)이 제안한 방법이 가장 높았고, 우산형 대립가설형태에서는 제안한 방법이 더 좋았다. 이중지수분포에서는 정규분포에서와 유사한 결과가 나왔다. 그리고 정규분포에서의 블록이 15개인 경우에는 블록이 5개인 경우와 유사한 결과가 나왔고, 지수분포에서는 점점 커져서 정점(peak) 이후에 평행하거나 평행하다가 정점(peak) 이후에 작아지는 대립가설에서는 Friedman (1937)이 제안한 방법이 가장 높았고, 우산형 대립가설에서는 제안한 방법들이 더 좋았다. Cauchy 분포와 이중지수분포에서 모두 블록의 크기가 5개인 경우와 비슷한 결과를 볼 수 있다.

Table 3.2를 보면, 처리가 6개, 정점(peak)이 3이고 정규분포에서 블록이 5개인 경우에는 정점(peak)까지 커지다가 다시 작아지는 뚜렷한 우산형 대립가설에서는 제안한 방법이 가장 높은 것으로 나왔으며, 평균의 차이가 클수록 더 높았다. 그리고 정점(peak)까지 커지다가 정점(peak) 이후로 작아지지만 조금 덜 작아지는 대립가설, 정점(peak)까지 커지다가 이후로는 작아져서 평행하는 대립가설 그리고 커지다가 평행을 유지하다가 다시 작아지는 대립가설에서는 제안한 검정법의 검정력이 가장 높은 것으로 나왔다. 지수분포에서는 모든 대립가설에서 제안한 방법의 검정력이 가장 높은 것을 알 수 있다. Cauchy 분포와 이중지수분포에서는 실험한 대립가설에서 모두 제안한 검정법의 검정력이 가장 높은 것을 알 수 있다. 마찬가지로 Table 3.2에서 블록이 10개인 경우를 살펴보면, 정규분포와 지수분포, Cauchy 분포 그리고 이중지수분포에서 모두 블록인 5개인 경우와 비슷한 결과가 나왔다는 것을 알 수 있다.

마지막으로 Table 3.3을 보면, 처리가 6개, 정점(peak)이 4이고 정규분포에서 블록이 5개인 경우에는 정점(peak)까지 커지다가 이후로 작아지는 뚜렷한 우산형 대립가설에서 제안한 검정법이 높았으며 정점(peak)이 3인 경우와 마찬가지로 평균의 차이가 클수록 더 높다는 것도 알 수 있다. 정점(peak)까지 커지다가 정점(peak) 이후로 작아지지만 조금 덜 작아지는 대립가설, 평행하다가 정점(peak)에서 제일 크고 다시 작아지는 대립가설 그리고 커져서 평행하다가 다시 작아지는 대립가설에서는 제안한 방법의 검정력이 가장 높았다. 지수분포에서는 실험한 모든 대립가설에서는 제안한 방법의 검정력이 가장 높았다. Cauchy 분포와 이중지수분포에서도 제안한 방법의 검정력이 가장 높다는 것을 볼 수 있다. 마찬가지로 Table 3.3의 블록이 10개인 경우의 정규분포에서는 점점 커지다가 정점(peak) 이후 다시 작아지는 우산형 대립가설과 평행하다가 정점(peak)에서 가장 크고 다시 작아지는 대립가설에서는 제안한 방법의 검정력이 가장 높았다. 지수분포에서는 블록인 5개인 경우와 유사한 결과가 나왔다. Cauchy 분포와 이중지수분포에서도 나머지 실험한 모든 대립가설의 경우에서 블록이 5개인 경우와 같은 결과가 나왔다는 것을 볼 수 있다.

4. 결론 및 고찰

본 논문에서는 랜덤화 블록 계획법(randomized block design)에서 우산형 대립가설을 검정하기 위

한 비모수 검정법을 제안하였다. 이 통계량은 블록 간의 정보를 이용하기 위한 Hodges와 Lehmann (1962)의 정렬방법과 Orban과 Wolfe (1982)의 논문에서 사용된 선형 위치(linear placement)에 정규점수함수(normal score function)와 지수점수함수(exponential score function)를 이용하여 만들어졌다. 그리고 제 1종 오류를 살펴보면, 분산분석법의 경우 정규분포 이외에 제 1종 오류를 제어하지 못하고, 특히 Cauchy 분포에서서는 더욱 제어하지 못하였다. 제안한 방법들은 모두 0.05 근방의 값으로 잘 제어되었다. Friedman (1937)이 제안한 검정법은 0.04의 근방으로 분산분석이나 제안한 방법들보다 더 보수적인 검정법임을 알 수 있다.

제안한 검정법과 모수적인 검정방법인 분석분석법과 비모수적인 방법인 Friedman (1937)이 제안한 방법을 정규분포, 지수분포, Cauchy 분포 그리고 이중지수분포에서 비교한 모의실험을 통하여 분포에 따라 검정력이 상이하다는 것을 알 수 있었다. 모의실험에서 대체로 우산형 대립가설이 아닌 경우에서는 분산분석법과 Friedman (1937)이 제안한 검정법이 좋았지만 정점(peak)이 명확한 우산형 대립가설과 커지다가 평행을 유지하다가 다시 작아지는 대립가설에서는 제안한 방법의 검정력이 가장 높게 나왔다. 그리고 정점(peak)까지 커지다가 이후에는 작아져서 평행하는 대립가설, 평행하다가 정점(peak)에서 가장 커졌다가 작아지는 대립가설, 커져서 평행하다가 다시 작아지는 대립가설에서는 대체로 제안한 방법의 검정력이 높게 나왔지만, 블록의 크기, 처리의 크기 등에 따라 기존의 방법이 더 좋은 경우도 있었다.

따라서 본 논문에서 제안한 방법은 뚜렷한 우산형 대립가설의 형태가 아닐 시에는 기존의 방법들보다 검정력이 떨어지거나 블록의 크기, 처리의 수 등의 조건에 따라 다르다는 것을 알 수 있다. 하지만 정점(peak)이 명확한 우산형 대립가설과 커지다가 평행을 유지하다가 다시 작아지는 대립가설에서는 어떤 분포이든 상관없이 제안한 검정법을 사용하면 분산분석이나 Friedman (1937)이 제안한 검정법보다 효율적인 분석을 할 수 있다. 이런 경우에는 제안한 방법을 사용하는 것이 더욱 효율적일 것이다. 하지만 미지의 블록효과가 존재할 수 있으므로 비모수적 검정법의 분포무관의 성질을 유지하며 블록 간 정보를 추출하는 것에 어려움이 따를 수 있다는 문제가 있다.

References

Alvo, M. (2008). Nonparametric tests of hypotheses for umbrella alternatives, *Canadian Journal of Statistics*, **36**, 143–156.

Bhat, S. V. (2009). Simple k-sample rank tests for umbrella alternatives, *Research Journal of Mathematics and Statistics*, **1**, 27–29.

Chen, Y. I. and Wolfe, D. A. (1990). A study of distribution-free tests for umbrella alternatives, *Biometrical Journal*, **32**, 47–57.

Chen, Y. I. and Wolfe, D. A. (1993). Nonparametric procedures for comparing umbrella pattern treatment effects with a control in a one-way layout, *Biometrics*, **49**, 455–465.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *Journal of the American Statistical Association*, **32**, 675–701.

Hettmansperger, T. P. and Norton, R. M. (1987). Tests for patterned alternatives in k-sample problems, *Journal of the American Statistical Association*, **82**, 292–299.

Hodges, J. L. and Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance, *The Annals of Mathematical Statistics*, **33**, 482–497.

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

Kim, H. and Kim, D. (2016). Nonparametric procedures based on aligned method and placement for ordered alternatives in randomized block design, *Journal of the Korean Data and Information Science Society*, **29**, 707–717.

Kim, D., Lee, S., and Wang, W. (2011). The asymptotic behavior of linear placement statistics, *Statistics and Probability Letters*, **81**, 326–336.

Lee, S. and Kim, D. (2011). Nonparametric procedures using placement in randomized block design with replications, *Journal of the Korean Data and Information Science Society*, **22**, 1105–1112.

Mack, G. A. and Wolfe, D. A. (1981). K-sample rank tests for umbrella alternatives, *Journal of the American Statistical Association*, **76**, 175–181.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other, *The Annals of Mathematical Statistics*, **18**, 50–60.

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

Page, E. B. (1963). Ordered hypotheses for multiple treatments: a significance test for linear ranks, *Journals of the American Statistical Association*, **58**, 216–230.

Salman A. S. (2010). *A new nonparametric test for the umbrella alternative* (Master's Thesis), University of North Dakota, North Dakota.

Song, H. and Kim, D. (1997). *Statistics*, Cheongmoongak, Seoul.

랜덤화 블록 계획법에서 우산형 대립가설에 대한 정렬방법과 위치를 이용한 비모수 검정법

김정현[a] · 김동재[a,1]

[a]가톨릭대학교 의생명· 건강과학과

(2016년 9월 30일 접수, 2016년 11월 11일 수정, 2016년 11월 18일 채택)

요 약

랜덤화 블록 계획법(randomized block design)에서 대립가설형태에 따라 많은 비모수적인 방법들이 제안되었다. 일반대립가설에서 대표적으로 Fridman (1937)의 검정법이 있고, 순서형 대립가설에서는 Page (1963)의 검정법이 있다. 우산형 대립가설에 대한 비모수적 방법으로는 일원 배치 모형에서 k개의 표본 문제에 대하여 Mack과 Wolfe (1981)의 검정법이 있다. 본 논문에서는 랜덤화 블록 계획법(randomized block design)에서 우산형대립가설에 대하여 블록 간의 정보를 이용한 Hodges와 Lehmann (1962)의 정렬방법과 위치를 이용한 Kim (1999)의 검정법을 이용하여 검정법을 제안하였다. 또한, Monte carlo 모의실험을 통하여 제안된 검정법과 기존의 검정법을 비교하였다.

주요용어: 랜덤화 블록 계획법, 정렬방법, 위치, 비모수 방법, 우산형 대립가설

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