The Korean Journal of Applied Statistics (2017) DOI: https://doi.org/10.5351/KJAS.2017.30.6.931
30(6), 931–941

Nonparametric method using linear placement statistics in randomized block design with replications

Aran Kim[a] · Dongjae Kim[a,1]

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

(Received September 14, 2017; Revised October 16, 2017; Accepted October 20, 2017)

Abstract

Typical Nonparametric methods for randomized block design with replications are two methods proposed by Mack (1981) and Mack and Skillings (1980). This method is likely to cause information loss because it uses the average of repeated observations instead of each repeated observation in the processing of each block. In order to compensate for this, we proposed a test method using linear placement statistics, which is a score function applied to the joint placement method proposed by Chung and Kim (2007). Monte Carlo simulation study is adapted to compare the power with previous methods.

Keywords: randomized block design with replications, linear placement statistics, norparametric methods

1. 서론

의학, 약학 분야의 실험에서는 같은 실험대상에 대하여 실험조건이나 처리를 달리하거나 여러 다른 시점에서 반복적으로 값을 측정하는 경우가 많다. 예를 들어, 연구자들은 같은 사람에게서 적출한 간세포를 여러 종류의 배양액에 넣고 일정 기간이 경과할 때마다 증가된 세포 수를 측정하여 각 배양액 간의 차이를 비교한다. 이러한 실험에서 얻은 자료를 반복측정 자료(repeated measurement data)라 한다. 같은 개체 또는 동질적인 개체들을 실험단위로 하여 여러 처리 간의 차이를 비교하는 실험계획 방법으로는 랜덤화 블록 계획법이 있다. 이 방법은 같은 개체들의 수치 또는 같은 블록의 개체들이 서로 동질적이기 때문에, 이질적인 개체들을 대상으로 실험할 때 생길 수 있는 개체들 간의 차이를 전체 실험오차에서 그만큼 줄일 수 있다는 장점이 있다 (Song과 Park, 1988). 이때, 하나의 처리 수준에 블록마다 두 명 이상의 연구대상이 할당되는 경우는 반복이 있는 랜덤화 블록 계획법이라고 한다.

반복이 있는 랜덤화 블록 계획법에서는 처리의 효과 차이를 알아보기 위하여 기존에 제안된 많은 방법이 있다. 모수적 방법으로는 분산분석법이 있고, 비모수적 방법에는 처리의 효과들이 모두 같지는 않다는 일반적인 대립가설에서 사용되는 Mack과 Skillings (1980)이 제안한 방법과 Mack (1981)이 제안한 방법이 있다. Mack과 Skillings (1980)이 제안한 방법은 Friedman (1937)의 방법을 적용한 것으로, 하나의 블록 안에서 관측된 값들의 순위를 사용하여 각 처리에서 블록들의 평균순위의 합을 이용해 검정통계량을 계산한 방법이다. Mack (1981)의 방법은 일원배치모형에서 관측된 값들의 순위를 사용한 Kruskal과 Wallis (1952)가 제안한 방법을 확장한 방법이다. 각각의 관측값 대신에 순위를 이용하여 검

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

정하기 때문에 모든 관측값의 정보를 사용하지 않는다. 그러므로 정보가 손실될 수 있다는 단점이 존재한다. 이를 보완하기 위해 Sim과 Kim (2013)은 Orban과 Wolfe (1982)가 제안한 위치와 Kim (1999)이 제안한 방법을 확장해 새로운 방법을 제안하였다.

여기서 Orban과 Wolfe (1982)가 제안한 방법은 두 처리 중에 어느 한 처리에 대한 상대적인 위치정보를 사용하여 처리 효과의 차이를 검정하는 것으로 대조군의 표본 수가 처리군의 표본 수보다 클 경우에 유용하다는 장점이 있다. 순서형 대립가설에서는 Hettmansperger (1957)가 제안한 방법과 Skillings와 Wolfe (1977, 1978)가 제안한 방법이 있다. Hettmansperger (1957)가 제안한 방법은 Page (1963)가 이원배치모형에서 블록 안에서의 순위를 이용해 제안한 방법을 확장한 것이고, Skillings와 Wolfe (1977, 1988)가 제안한 방법은 Jonckheere (1954)와 Terpstra (1952)가 제안한 방법을 확장한 것이다.

본 논문에서는 일원배치모형에서 Chung과 Kim (2007)이 제안한 결합위치(joint placement)를 확장하여 반복이 있는 랜덤화 블록 계획법에서 새로운 비모수 검정법을 제안하고자 한다. 이는 결합위치에 점수함수(score function)를 적용하여 만든 선형위치통계량(linear placement statistic)을 사용한 검정방법이다. Monte Carlo simulation study를 통하여 기존의 모수적 방법과 비모수적 방법 그리고 새롭게 제안된 방법을 비교하였다.

2. 제안하는 방법

블록이 u개이고, 처리가 t개인 반복이 있는 랜덤화 블록 계획법의 모형은 다음과 같다.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \quad i = 1, 2, \ldots, t;\ j = 1, 2, \ldots, u;\ k = 1, 2, \ldots, n,$$

여기서 Y_{ijk}는 i번째 처리에서 j번째 블록의 k번째 값이고, μ는 전체평균, α_i는 i번째 처리의 효과이고 β_j는 j번째 블록의 효과를 나타낸다. 또한 ϵ_{ijk}는 오차항이며, 동일한 연속분포를 따르는 서로 독립인 확률 변수로 가정한다.

각 처리 효과가 모두 동일하다는 귀무가설과 처리효과들이 모두 같지는 않다는 대립가설은 다음과 같다.

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_t \quad \text{vs.} \quad H_1 : \text{처리효과가 적어도 하나는 다르다.} \tag{2.1}$$

Chung과 Kim (2007)은 일원배치모형에서 결합위치를 정의하였다. 이를 이용하여 블록이 u개인 반복이 있는 랜덤화 블록모형으로 확장하면, 일원배치모형 u개가 합해진 것으로 볼 수 있다. 따라서 j번째 블록($j = 1, 2, \ldots, u$)에서의 결합위치 V_{ijk}는

$$V_{ijk} = \frac{1}{nt - n} \sum_{\substack{h=1 \\ h \neq i}}^{t} \sum_{s=1}^{n} \chi(Y_{hjs}, Y_{ijk}), \quad \chi(x, y) = \begin{cases} 1, & x \leq y, \\ 0, & \text{otherwise}, \end{cases} \quad j = 1, 2, \ldots, u$$

이다. 이는 j번째 블록에서 i번째 처리의 관측값을 제외한 혼합표본에서 Y_{ijk}보다 작거나 같은 관측값의 개수를 이용한 $[0, 1]$에서의 확률변수이다.

또한, j번째 블록에서 정의된 결합위치를 이용하여 선형위치 통계량을 구하면 다음과 같다.

$$S_j = \sum_{i=1}^{t} \sum_{k=1}^{n} \phi\left(V_{ijk}\right), \quad j = 1, 2, \ldots, u.$$

이때, $\phi(\bullet)$는 $[0, 1]$의 범위에서 정의된 실변수 함수인 점수함수이다.

여기서 u개의 블록에서 정의된 선형위치통계량의 합 S를 이용하여 가설 (2.1)을 검정할 수 있다.

$$S = \sum_{j=1}^{u} S_j = \sum_{i=1}^{t}\sum_{j=1}^{u}\sum_{k=1}^{n} \phi\left(V_{ijk}\right).$$

한편, Hong과 Lee (2014)에 의해, 귀무가설 하에서 각각의 i에 대해, $n \to \infty$이면 S_j의 표준화된 통계량 $\hat{S}_j$의 분포는 표준정규분포로 수렴한다. 즉

$$\hat{S}_j = \frac{S_j - E\left(S_j\right)}{\sqrt{\mathrm{Var}\left(S_j\right)}} \longrightarrow N(0,1)$$

이다. 따라서

$$\hat{S} = \sum_{j=1}^{u} \hat{S}_j = \sum_{j=1}^{u} \frac{S_j - E\left(S_j\right)}{\sqrt{\mathrm{Var}\left(S_j\right)}} \longrightarrow N(0,u)$$

이므로, S의 표준화된 통계량 $\hat{S}$의 분포는 표준정규분포로 수렴한다.

$$\hat{S} = \frac{S - E\left(S\right)}{\sqrt{u\mathrm{Var}\left(S\right)}} \longrightarrow N(0,1).$$

본 논문에서 결합위치 V_{ijk}는 귀무가설을 참일 때 1/2의 값을 가지므로, $|1-2x|$는 귀무가설이 참일수록 0에 가까워지고 거짓일수록 1에 가까워지는 점수함수 $\phi(x) = |1-2x|$를 고려하였다. 이를 이용한 선형 위치 통계량은

$$S^{\phi} = \sum_{i=1}^{t}\sum_{j=1}^{u}\sum_{k=1}^{n} |1 - 2V_{ijk}|$$

이다. 또한 귀무가설 하에서 점수함수를 이용한 선형위치통계량 $\hat{S}^{\phi}$의 근사분포는 다음과 같다.

$$\hat{S}^{\phi} = \frac{S^{\phi} - E\left(S^{\phi}\right)}{\sqrt{u\mathrm{Var}\left(S^{\phi}\right)}} \longrightarrow N(0,1).$$

이때 N이 전체 표본 수라고 한다면, S^{ϕ}의 기댓값은 귀무가설 하에서 다음과 같다.

$$E(S^{\phi}) = \begin{cases} \dfrac{N(nt-n+2)}{2(nt-n+1)}, & nt-n\text{이 짝수일 때,} \\ \dfrac{N(nt-n+1)}{2(nt-n)}, & nt-n\text{이 홀수일 때.} \end{cases}$$

위 식을 이용하여 S^{ϕ}의 기댓값은 $nt-n$이 짝수와 홀수일 때 다르게 구할 수 있다. 또한 S^{ϕ}의 분산은 임의의 n, t, u에 대해서는 구체적인 식으로 표현할 수 없고, 특정 n, t, u에 대해서는 계산할 수 있다.

3. 사례

식이요법이 몸무게를 줄이는데 효과가 있는지를 알아보기 위해 40명의 비만 여성을 대상으로 실험하였다. 40명을 4그룹으로 랜덤하게 나누어 각 그룹에 다른 식이요법을 적용하였으며, 이때 각 그룹을 두 개의 블록으로 나누도록 하였다. Table 3.1은 실험 후 줄은 몸무게를 기록한 것이다 (Modified from Song과 Kim, 2015) ($\alpha = 0.05$).

Table 3.1. Weight reduction by diet therapy

	식이요법1	식이요법2	식이요법3	식이요법4
	15	25	19	22
	12	19	24	22
1	18	21	18	18
	16	22	16	19
	13	19	21	15
	9	13	13	33
	9	15	13	30
2	13	12	15	31
	7	15	18	27
	9	12	15	28

Table 3.2. Value of V_{ijk} and $\|1-2V_{ijk}\|$

	V_{ijk}				$\|1-2V_{ijk}\|$			
	식이요법1	식이요법2	식이요법3	식이요법4	식이요법1	식이요법2	식이요법3	식이요법4
	0.066667	1	0.666667	0.866667	0.866667	1	0.333333	0.733333
	0	0.733333	0.933333	0.866667	1	0.466667	0.866667	0.733333
1	0.266667	0.8	0.466667	0.466667	0.466667	0.6	0.066667	0.066667
	0.133333	0.933333	0.333333	0.666667	0.733333	0.866667	0.333333	0.333333
	0	0.733333	0.733333	0.2	1	0.466667	0.466667	0.6
	0	0.466667	0.533333	1	1	0.066667	0.066667	1
	0	0.6	0.533333	1	1	0.2	0.066667	1
2	0.333333	0.266667	0.666667	1	0.333333	0.466667	0.333333	1
	0	0.6	0.666667	1	1	0.2	0.333333	1
	0	0.266667	0.666667	1	1	0.466667	0.333333	1

Table 3.1의 자료를 이용하여 V_{ijk}와 $|1-2V_{ijk}|$를 구하면 Table 3.2와 같다. 이를 이용하여 구한 선형 위치 통계량 S^{ϕ}은

$$\begin{aligned} S^{\phi} &= \sum_{i=1}^{t}\sum_{j=1}^{u}\sum_{k=1}^{n}|1-2V_{ijk}| \\ &= 0.866667+1+0.466667+0.733333+1+\cdots+1+1+1+1+1 = 23.86667 \end{aligned}$$

이고, $E(S^{\phi})$와 $\mathrm{Var}(S^{\phi})$은 다음과 같다.

$$\begin{aligned} E(S^{\phi}) &= \frac{N(nt-n+1)}{2(nt-n)} = \frac{40(20-5+1)}{2(20-5)} = 21.333, \\ \mathrm{Var}(S^{\phi}) &= 0.3726 \end{aligned}$$

이때 표준화된 통계량 $\hat{S}^{\phi}$은 표준정규분포로 수렴한다.

$$\hat{S}^{\phi} = \frac{S^{\phi}-E\left(S^{\phi}\right)}{\sqrt{u\mathrm{Var}\left(S^{\phi}\right)}} = \frac{23.86667-21.333}{\sqrt{0.7452}} = 2.935.$$

따라서 2.935는 기각역 $z_{0.025}=1.96$보다 크므로 귀무가설을 기각하는 결과를 나타낸다.

Table 4.1. Monte Carlo power estimates: $\alpha = 0.05$, $t = 3$, block = 7, $n = 3$

DIST	α_1	α_2	α_3	ANO	MACK	MS	AR
Normal	0	0	0	0.0494	0.0187	0.0474	0.0503
	0	0.1	0	0.0585	0.0188	0.0552	0.0561
	0	0.1	0.2	0.0797	0.0209	0.0698	0.0607
	0	0.5	0.5	0.3500	0.0547	0.3013	0.1106
	0	0.7	0.8	0.6944	0.1395	0.6223	0.2092
	0	0.5	1	0.8142	0.2108	0.7445	0.2819
	0	1.2	1.7	0.9990	0.7889	0.9977	0.7751
Exponential	0	0	0	0.0425	0.0181	0.0449	0.0512
	0	0.1	0	0.0582	0.0230	0.0727	0.0572
	0	0.1	0.2	0.0819	0.0316	0.1235	0.0699
	0	0.5	0.5	0.3869	0.1498	0.5666	0.1628
	0	0.7	0.8	0.7026	0.3517	0.8421	0.3195
	0	0.5	1	0.8103	0.4831	0.9403	0.5750
	0	1.2	1.7	0.9947	0.9262	0.9994	0.9000
Double exponential	0	0	0	0.0449	0.0177	0.0461	0.0506
	0	0.1	0	0.0568	0.0200	0.0561	0.0600
	0	0.1	0.2	0.0626	0.0225	0.0628	0.0641
	0	0.5	0.5	0.2028	0.0485	0.2408	0.1094
	0	0.7	0.8	0.4071	0.1046	0.4825	0.1989
	0	0.5	1	0.5139	0.1410	0.5893	0.2459
	0	1.2	1.7	0.9383	0.5488	0.9619	0.6412
Cauchy	0	0	0	0.0190	0.0204	0.0435	0.0528
	0	0.1	0	0.0153	0.0224	0.0527	0.0606
	0	0.1	0.2	0.0174	0.0209	0.0570	0.0622
	0	0.5	0.5	0.0266	0.0300	0.1149	0.0824
	0	0.7	0.8	0.0341	0.0466	0.2119	0.1139
	0	0.5	1	0.0413	0.0554	0.2684	0.1328
	0	1.2	1.7	0.0856	0.1684	0.6222	0.3216

ANO = ANOVA for randomized block design; MACK = method proposed by Mack; MS = method proposed by Mack and Skillings; AR = linear placement statistics.

4. 모의실험 및 결과

본 논문에서 새롭게 제안한 선형위치통계량을 이용한 검정법과 기존의 검정법들과의 차이를 비교하였다. 여기서 기존의 방법들은 모수적인 방법으로는 분산분석법(ANOVA)을 사용하였고, 비모수적 방법에는 Mack과 Skillings (1980)이 제안한 방법과 Mack (1981)이 제안한 방법을 사용하였다. 모집단의 분포는 정규분포, 지수분포, 이중지수분포, Cauchy분포를 고려하였다.

기존의 방법들을 새롭게 제안된 방법의 검정력과 비교하기 위하여 SAS프로그램을 사용해 모의실험을 실행하였다. 정규분포는 RANNOR함수, 지수분포는 RANEXP함수, Cauchy분포는 RANCAU함수, 이중지수분포는 RANUNI함수를 역변환 방법을 적용해 난수를 생성하였다. 생성된 난수를 사용하여 계산된 통계량의 검정력을 비교하기 위해서 10,000번을 반복하는 Monte Carlo simulation study를 실시하였다.

처리의 수는 3개와 5개, 블록의 수는 7개와 10개인 경우로 선택하였고 각 블록 내에서 처리별 반복의

Table 4.2. Monte Carlo power estimates: $\alpha = 0.05$, $t = 3$, block = 10, $n = 3$

DIST	α_1	α_2	α_3	ANO	MACK	MS	AR
Normal	0	0	0	0.0496	0.0183	0.0496	0.0483
	0	0.1	0.2	0.0943	0.0237	0.0908	0.0562
	0	0.3	0.5	0.3867	0.0518	0.3426	0.0942
	0	0.5	0.6	0.5779	0.0857	0.5202	0.1347
	0	0.7	0.7	0.7887	0.1551	0.07289	0.2084
	0	0	1	0.9823	0.4114	0.9612	0.4337
	0	1.1	1.6	1.0000	0.8515	0.9996	0.8058
Exponential	0	0	0	0.0505	0.0204	0.519	0.0515
	0	0.1	0.2	0.1002	0.0356	0.1601	0.0653
	0	0.3	0.5	0.4004	0.1333	0.6516	0.1593
	0	0.5	0.6	0.6047	0.2383	0.8086	0.2144
	0	0.7	0.7	0.7887	0.3955	0.9208	0.3069
	0	0	1	0.9772	0.7522	0.9994	0.9033
	0	1.1	1.6	0.9995	0.9676	1.0000	0.9391
Double exponential	0	0	0	0.0487	0.0200	0.0474	0.0483
	0	0.1	0.2	0.0712	0.0250	0.0826	0.0575
	0	0.3	0.5	0.2163	0.0434	0.2618	0.0910
	0	0.5	0.6	0.3409	0.0693	0.4122	0.1315
	0	0.7	0.7	0.4969	0.1065	0.5943	0.1836
	0	0	1	0.8055	0.2615	0.8773	0.3671
	0	1.1	1.6	0.9754	0.6194	0.9896	0.6760
Cauchy	0	0	0	0.0156	0.0196	0.0503	0.0485
	0	0.1	0.2	0.0158	0.0215	0.0640	0.0539
	0	0.3	0.5	0.0221	0.0308	0.1316	0.0714
	0	0.5	0.6	0.0305	0.0374	0.1859	0.0829
	0	0.7	0.7	0.0338	0.0460	0.2805	0.1054
	0	0	1	0.0486	0.0844	0.4892	0.1739
	0	1.1	1.6	0.0830	0.1876	0.7582	0.3288

ANO = ANOVA for randomized block design; MACK = method proposed by Mack; MS = method proposed by Mack and Skillings; AR = linear placement statistics.

수는 3개로 표본의 크기가 모두 같은 경우를 고려하였다. 그리고 유의수준 α를 0.05로 보정하기 위해서 확률화 검정을 실시하였다. 처리가 3개이고 블록이 7개인 경우는 Table 4.1, 처리가 3개이고 블록이 10개인 경우는 Table 4.2, 처리가 5개이고 블록이 7개인 경우는 Table 4.3, 마지막으로 처리가 5개이고 블록이 10개인 경우에는 Table 4.4로 나타내었다.

각 처리의 효과가 모두 동일할 때 검정방법이 제 1종 오류를 잘 제어하는지 살펴보면, 분산분석법의 경우 정규분포에서는 Table 4.1에서 0.0494, Table 4.2에서 0.0496, Table 4.3에서 0.0502, Table 4.4에서 0.0505로 0.05에 근사한 값을 얻었다. 지수분포에서는 Table 4.1과 Table 4.3에서 0.0425와 0.0462로 비교적 낮았다. 이중지수분포에서는 Table 4.1과 Table 4.4에서 0.0449와 0.0461로 비교적 낮았고 Cauchy분포에서는 모든 경우에서 0.02이하로 다소 낮게 나왔다. 이를 통해 분산분석법은 정규분포에서는 제 1종 오류를 잘 제어하지만 지수분포와 이중지수에서는 제 1종 오류를 제어하는데 어려움이 존재하고, Cauchy분포에서는 제 1종오류를 제어하는 것이 힘들다는 것을 알 수 있다. Mack (1981)의 방법은 모든 경우에서 0.02에 근사한 값들을 얻었다. Mack과 Skillings (1980)에서는 지수분포의 경

Table 4.3. Monte Carlo power estimates : $\alpha = 0.05$, $t = 5$, block = 7, $n = 3$

DIST	α_1	α_2	α_3	α_4	α_5	ANO	MACK	MS	AR
Normal	0	0	0	0	0	0.0502	0.0173	0.0471	0.0487
	0	0	0.1	0	0	0.0542	0.0214	0.0515	0.0394
	0	0	0	0	0.1	0.0594	0.0202	0.0560	0.0397
	0	0.5	0.5	0.5	0.5	0.3197	0.0489	0.2783	0.0855
	0	0.2	0.4	0.6	0.8	0.5962	0.1001	0.5430	0.1393
	1	1	1	0	0	0.9848	0.4854	0.9734	0.4656
	0	0	0	1.3	1.3	0.9999	0.8347	0.9996	0.7728
Exponential	0	0	0	0	0	0.0462	0.0204	0.0488	0.0482
	0	0	0.1	0	0	0.0536	0.0204	0.0701	0.0489
	0	0	0	0	0.1	0.0554	0.0233	0.0688	0.0420
	0	0.5	0.5	0.5	0.5	0.3378	0.1204	0.5496	0.0991
	0	0.2	0.4	0.6	0.8	0.6164	0.2838	0.8753	0.2448
	1	1	1	0	0	0.9795	0.8055	0.9970	0.9997
	0	0	0	1.3	1.3	0.9991	0.9693	1.0000	0.9981
Double exponential	0	0	0	0	0	0.0496	0.0208	0.0507	0.0498
	0	0	0.1	0	0	0.0501	0.0221	0.0488	0.0545
	0	0	0	0	0.1	0.0524	0.0222	0.0483	0.0493
	0	0.5	0.5	0.5	0.5	0.1753	0.0449	0.2178	0.0713
	0	0.2	0.4	0.6	0.8	0.3313	0.0727	0.4210	0.1444
	1	1	1	0	0	0.8014	0.3091	0.8905	0.4283
	0	0	0	1.3	1.3	0.9653	0.5848	0.9868	0.6807
Cauchy	0	0	0	0	0	0.0181	0.0205	0.0450	0.0502
	0	0	0.1	0	0	0.0185	0.0212	0.0490	0.0422
	0	0	0	0	0.1	0.0162	0.0191	0.0422	0.0505
	0	0.5	0.5	0.5	0.5	0.0196	0.0262	0.1033	0.0670
	0	0.2	0.4	0.6	0.8	0.0222	0.0399	0.1839	0.0918
	1	1	1	0	0	0.0379	0.0933	0.4689	0.2008
	0	0	0	1.3	1.3	0.0554	0.1680	0.6934	0.3331

ANO = ANOVA for randomized block design; MACK = method proposed by Mack; MS = method proposed by Mack and Skillings; AR = linear placement statistics.

우 Table 4.1에서 0.0449, 이중지수분포의 경우 Table 4.1에서 0.0461, 코시분포의 경우 Table 4.1에서 0.0435, Table 4.3에서 0.045, Table 4.4에서 0.0461을 제외하고는 제 1종 오류를 제어하는 데 어려움이 없다고 할 수 있다. 결합위치를 이용한 검정방법에서는 거의 모든 경우 0.05에 근사해 제 1종 오류를 제어하는데 큰 문제가 없음을 볼 수 있다.

다음으로 검정력을 보면, 정규분포에서는 처리와 블록 수에 관계없이 모든 형태에서 분산분석법의 검정력이 가장 높게 나타났다. 대부분의 경우에 Mack과 Skillings (1980)의 검정력이 두 번째로 높지만, 새로 제안한 방법이 Table 4.1에서 대립가설이 낮게 증가했다가 감소하는 형태와 Table 4.4에서 마지막에 낮게 증가하는 형태의 경우 각각 0.0561과 0.0577로 Mack과 Skillings (1980)의 검정력보다 높아 분산분석법 다음으로 높다는 것을 알 수 있다. Mack (1981)이 제안한 검정력은 Table 4.1과 Table 4.2에서 대립가설이 높게 증가하다 덜 증가하는 형태의 경우 각각 0.7889와 0.8515, Table 4.3에서 완만했다가 감소하는 형태와 완만했다가 높게 증가하는 형태는 각각 0.4854, 0.8347, Table 4.4에서 증가 후 완만했다가 감소하는 형태는 0.6362로 세 번째로 높게 나타났는데, 이를 제외하고는 새롭게 제안한 검정방법

Table 4.4. Monte Carlo power estimates: $\alpha = 0.05$, $t = 5$, block = 10, $n = 3$

DIST	α_1	α_2	α_3	α_4	α_5	ANO	MACK	MS	AR
Normal	0	0	0	0	0	0.0505	0.0196	0.0482	0.0506
	0	0	0.1	0	0	0.0621	0.0210	0.0612	0.0543
	0	0	0	0	0.1	0.0600	0.0187	0.0551	0.0577
	0	0.1	0.2	0.3	0.4	0.2391	0.0344	0.2127	0.0751
	0	0	0.5	0.5	0.5	0.6550	0.0915	0.5927	0.1452
	0	0	0.3	0.6	0.9	0.9384	0.2391	0.9095	0.2767
	0	1	1	1	0	0.9995	0.6362	0.9985	0.6127
Exponential	0	0	0	0	0	0.0491	0.0193	0.0478	0.0510
	0	0	0.1	0	0	0.0580	0.0233	0.0763	0.0577
	0	0	0	0	0.1	0.0611	0.0225	0.0784	0.0563
	0	0.1	0.2	0.3	0.4	0.2685	0.0830	0.5435	0.1302
	0	0	0.5	0.5	0.5	0.6598	0.2760	0.9063	0.2312
	0	0	0.3	0.6	0.9	0.9407	0.6297	0.9978	0.6581
	0	1	1	1	0	0.9976	0.9164	0.9999	0.8007
Double exponential	0	0	0	0	0	0.0461	0.0206	0.0483	0.0505
	0	0	0.1	0	0	0.0502	0.0207	0.0542	0.0595
	0	0	0	0	0.1	0.0524	0.0228	0.0554	0.0589
	0	0.1	0.2	0.3	0.4	0.1299	0.0323	0.1709	0.0800
	0	0	0.5	0.5	0.5	0.3458	0.0710	0.4544	0.1487
	0	0	0.3	0.6	0.9	0.6488	0.1610	0.7792	0.2754
	0	1	1	1	0	0.9327	0.4153	0.9780	0.5662
Cauchy	0	0	0	0	0	0.0146	0.0204	0.0461	0.0515
	0	0	0.1	0	0	0.0179	0.0216	0.0525	0.0608
	0	0	0	0	0.1	0.0145	0.0204	0.0511	0.0583
	0	0.1	0.2	0.3	0.4	0.0160	0.0268	0.0890	0.0670
	0	0	0.5	0.5	0.5	0.0225	0.0359	0.2007	0.0962
	0	0	0.3	0.6	0.9	0.0274	0.0593	0.3775	0.1492
	0	1	1	1	0	0.0347	0.1209	0.6548	0.2758

ANO = ANOVA for randomized block design; MACK = method proposed by Mack; MS = method proposed by Mack and Skillings; AR = linear placement statistics.

이 Mack (1981)이 제안한 방법보다 높게 나타났다.

지수분포에서는 Table 4.3일 때 대립가설이 완만했다가 감소하는 형태에서 새롭게 제안한 방법의 검정력이 0.9997로 가장 높게 나타난 것을 알 수 있다. 이 형태를 제외하고는 Mack과 Skillings (1980)의 방법의 검정력이 가장 높게 나타난 것을 알 수 있다. 대립가설 형태와 상관없이 분산분석법은 Mack과 Skillings (1980) 다음으로 검정력이 높았다. Mack (1981)이 제안한 검정력은 Table 4.1에서 증가 후 낮게 증가하는 형태와 높게 증가하다 덜 증가하는 형태의 경우 각각 0.3517, 0.9262, Table 4.2에서 증가 후 낮게 증가하는 형태의 경우 0.2383와 증가 후 완만한 형태의 경우 0.3955, 높게 증가하다 덜 증가하는 형태의 경우 0.9676, Table 4.3에서 증가 후 완만한 형태와 일정하게 증가하는 형태의 경우 각각 0.1204, 0.2838, Table 4.4에서 증가 후 완만해진 형태와 증가 후 완만했다가 감소하는 형태의 경우 각각 0.2760, 0.9164로 세 번째로 높게 나타났다. 이를 제외하고는 새롭게 제안한 검정방법이 Mack (1981)이 제안한 방법보다 높게 나타났다.

이중지수분포에서는 새롭게 제안한 방법의 검정력이 Table 4.1일 때 대립가설이 낮게 증가했다가 감소

하는 형태와 일정하게 낮게 증가하는 형태의 경우 각각 0.0600, 0.0641, Table 4.3일 때 중간에 증가 후 다시 감소하는 형태의 경우 0.545, Table 4.4일 때 중간에 증가 후 다시 감소하는 형태와 완만하다가 낮게 증가하는 형태의 경우 각각 0.0595, 0.0589로 가장 높게 나타났다. 그 외의 형태에서는 Mack과 Skillings (1980)이 제안한 검정방법, 분산분석법, 새롭게 제안한 검정방법, 마지막으로 Mack (1981)이 제안한 방법 순으로 검정력이 높게 나타났다.

Cauchy분포에서는 앞서 3개의 분포에 비해 검정력이 낮아졌는데, 특히 분산분석법의 검정력이 많이 낮아진 것을 볼 수 있다. 새롭게 제안한 검정력이 Table 4.1에서 대립가설이 낮게 증가했다가 감소하는 형태와 일정하게 낮게 증가하는 형태의 경우 각각 0.0606과 0.0622, Table 4.3에서 완만하다가 낮게 증가하는 형태의 경우 0.0505, Table 4.4에서 중간에 증가 후 다시 감소하는 형태와 완만하다가 낮게 증가하는 형태의 경우 각각 0.0608, 0.0583으로 가장 높게 나타났다. 그 외에 형태에서는 Mack과 Skillings (1980)이 제안한 검정방법, 새롭게 제안한 검정방법, Mack (1981)의 검정방법, 분산분석법순으로 검정력이 나타났다.

Table 4.1에서 대부분의 경우 정규분포에서는 분산분석법, Mack과 Skillings (1980)의 방법, 제안한 검정법, Mack (1981)의 방법순으로 검정력이 높았고 지수분포에서는 Mack과 Skillings (1980)의 방법, 분산분석법, 제안한 방법, Mack (1981)의 방법순으로 검정력이 높았다. 또한 이중지수분포에서는 Mack과 Skillings (1980)의 방법, 분산분석법, 제안한 방법, Mack (1981)의 방법순이고 Cauchy분포에서는 Mack과 Skillings (1980)의 방법, 제안한 방법, Mack (1981)의 방법, 분산분석법순으로 높게 나타났다. 이때 대립가설 형태에 따라 우위가 엇갈리는 결과를 보였지만, 대립가설이 낮게 증가했다가 감소하는 형태에서 다른 형태에서보다 제안하는 방법의 검정력이 비교적 높은 것을 특징으로 볼 수 있다.

5. 결론 및 고찰

본 논문에서는 반복이 있는 랜덤화 블록 모형을 검정하기 위해 비모수적 방법인 결합위치를 이용한 새로운 검정방법을 제안하였다. Chung과 Kim (2007)이 제안한 결합위치에 $[0,1]$의 범위에서 정의된 점수함수를 이용하여 선형위치통계량을 만들었다. Monte Carlo simulation study를 통하여 새롭게 제안된 방법의 검정력을 정규분포, 지수분포, 이중지수분포, Cauchy분포에서 모수적 방법인 분산분석법, 비모수적 방법인 Mack과 Skillings (1980)이 제안한 방법과 Mack (1981)이 제안한 방법을 사용하여 얻은 검정력들과 비교하였다.

모의실험의 전체적인 결과를 살펴보면, Mack (1981)이 제안한 방법과 Cauchy분포를 제외하고는 대부분에서 제 1종 오류를 잘 제어했다고 볼 수 있다. 이때, Cauchy분포에서는 Mack과 Skillings (1980)이 제안한 방법과 새롭게 제안한 방법이 분산분석법과 Mack (1981)이 제안한 방법보다 제 1종 오류를 잘 제어했다고 볼 수 있다. 검정력에서는 정규분포일 경우 거의 모든 대립가설에서 분산분석법의 검정력이 가장 높았다. 하지만 새롭게 제안된 방법은 정규분포에서 낮은 검정력을 보이지만 일반적으로 정규분포를 따르는 자료에는 모수적 방법을 사용하므로 큰 문제가 되지 않는다고 볼 수 있다. 지수분포, 이중지수분포, Cauchy분포일때 대부분의 대립가설에서 Mack과 Skillings (1980)이 제안한 방법의 검정력이 가장 높았다. 다만 지수분포에서 완만하다가 감소하는 형태의 경우 Mack과 Skillings (1980)이 제안한 방법의 검정력보다 새롭게 제안한 방법의 검정력이 가장 높게 나타났다. 또한 이중지수분포와 Cauchy분포에서는 검정력이 낮은 형태일 때, 새롭게 제안한 방법의 검정력이 비교적 높게 나왔으므로 효과차이가 작은 그룹에서의 검정을 할 때 효율적이라고 할 수 있다.

본 논문에서 제안한 방법인 선형위치통계량은 표준정규분포로 수렴한다는 장점이 있기 때문에 기존의 방법들보다 비모수 검정법에서 유용할 것이라고 기대된다. 또한 기존의 방법들은 각 관측값 대신 관측

값의 순위의 평균을 사용하기 때문에 정보의 손실이 발생할 수 있지만 제안한 방법에서는 관측치의 평균이 아닌 결합위치를 사용하기 때문에 의미 있는 차별성이 있다고 생각할 수 있다. 추후에 선형위치통계량에 대한 연구가 진행된다면 본 논문에서 이용한 점수함수 외에 처리 별 표본크기가 다른 경우에서도 기존 방법 중 가장 높은 검정력을 도출할 수 있는 새로운 점수함수를 이용한 선형위치통계량에 대한 연구가 필요할 것으로 생각된다. 또한 새롭게 제안한 방법은 처리군의 용량수준이 대조군과의 효과 차이가 작은 그룹이 존재하는 경우에서 제안한 방법의 검정력이 가장 높지만, 전체적으로는 낮은 검정력을 가지기 때문에 추후에 더 보완해야 할 것으로 보인다.

References

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

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. (1957). Non-parametric inference for ordered alternatives in a randomized block design, *Psychometrika*, **40**, 53–62.

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

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika*, **41**, 133–145.

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.

Mack, G. A. (1981). A quick and easy distribution-free test for main effects in a two-factor ANOVA, *Communications in Statistics - Simulation and Computation*, **10**, 571–591.

Mack, G. A. and Skillings, J. H. (1980). A Friedman-type rank test for main effects in a two-factor ANOVA, *Journal of the American Statistical Association*, **75**, 947–951.

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.

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

Sim, S. and Kim, D. (2013). Nonparametric method using placement in a randomized complete block design, *Journal of the Korean Data Information Science Society*, **24**, 1401–1408.

Skillings, J. H. and Wolfe, D. A. (1977). Testing for ordered alternatives by combining independent distribution-free block statistics, *Communications in Statistics-Theory and Methods*, **6**, 1453–1463.

Skillings, J. H. and Wolfe, D. A. (1978). Distribution-free tests for ordered alternatives in a randomized block design, *Journals of the American Statistical Association*, **73**, 427–431.

Song, H. and Kim, D. (2015). *Understanding Statistics*, Chenog Moon Gak Publisher, Seoul.

Song, H. and Park, D. (1998). *Analysis of Repeated Measures and Cross-over Design*, Free Academy, Seoul.

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking, *Indagationes Mathematicae*, **14**, 327–333.

반복이 있는 랜덤화 블록 계획법에서 선형위치통계량을 이용한 비모수 검정법

김아란[a] · 김동재[a,1]

[a]가톨릭대학교 의생명·건강과학과

(2017년 9월 14일 접수, 2017년 10월 16일 수정, 2017년 10월 20일 채택)

요 약

반복이 있는 랜덤화 블록 계획법(randomized block design with replications)에서의 대표적인 검정법은 Mack이 제안한 방법과 Mack과 Skillings이 제안한 방법이 있다. 이 방법은 각 블록의 처리에서 반복된 각 관측값 대신에 반복된 관측값들의 평균을 이용하여 순위를 매기기 때문에 정보의 손실이 발생할 가능성이 있다. 이를 보완하기 위해 본 논문에서는 Chung과 Kim (2007)이 제안한 결합위치(joint placement) 방법에 점수함수(score function)를 적용한 선형위치통계량(linear placement statistics)을 이용한 검정방법을 제안하였다. 또한 Monte Carlo simulation study를 통해 기존의 방법들과 검정력을 비교하였다.

주요용어: 반복이 있는 랜덤화 블록 계획법, 비모수적 방법, 선형위치통계량

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