Journal of the Korean Data & Information Science Society
2013, **24**(6), 1401–1408

http://dx.doi.org/10.7465/jkdi.2013.24.6.1401
한국데이터정보과학회지

랜덤화 블록 계획법에서 위치를 이용한 비모수 검정법

심수진[1] · 김동재[2]

[12]가톨릭대학교 의학통계학과

접수 2013년 9월 25일, 수정 2013년 10월 29일, 게재확정 2013년 11월 11일

요 약

반복이 있는 랜덤화 블록 계획법에서 우산형 대립가설을 검정하기 위한 비모수 방법에는 Kim과 Kim (1992)이 제안한 방법이 있다. 본 논문에서는 Orban과 Wolfe (1982)가 제안한 위치와 Kim (1999)이 제안한 대조군과 처리군의 방법을 확장하여 랜덤화 블록 계획법에서 우산형 대립가설에서의 검정법을 제안하였다. 또한 여러 분포에 대한 모의실험 통하여 기존의 방법과의 검정력을 비교하였다.

주요용어: 비모수 방법, 우산형 대립가설, 위치.

1. 서론

일반적으로 약의 복용량이 증가하면 효과도 증가하는 추세를 보인다. 그러나 약의 효과에 대한 추세를 검정하다 보면 복용량이 늘어남에 따라 약의 효과도 같이 증가하다가, 일정 용량 수준부터는 복용량이 늘어날수록 약의 효과가 감소하는 경향을 살펴볼 수 있다. Mack과 Wolfe (1981)는 k개의 표본 문제에 대해 Mann과 Whitney 통계량을 사용하여 이러한 문제들을 다루었으며, 이를 우산형 대립가설이라 부르고 처리 효과가 증가의 추세를 보이다가 감소의 추세로 바뀌는 처리 용량 수준을 우산형 패턴의 정점이라 한다.

만일 모분포의 형태가 가정한 것과 같다면 모수적 통계방법을 사용하면 된다. 하지만 정규분포는 이론적인 분포로서, 실제로 어느 데이터가 정밀한 정규모집단에서 얻어지는 경우는 없다. 다만, 모집단의 분포가 정규분포에 가까운 경우가 많거나 또는 표본 통계량의 분포가 중심극한 정리에 의해 정규분포에 근사되어진다. 따라서 모집단에 대하여 구체적인 분포함수를 가정하는 것이 무리일 때에는 모집단의 분포에 대한 가정을 약화시키는 것이 오류의 가능성을 줄이고, 때로는 효율도 높일 수 있는 대안이 된다. 모집단의 분포함수에 대하여 특정형태를 가정하지 않는 통계적 방법이 비모수적 방법이다. 비모수적 검정법은 특정 분포함수의 모수에 대한 설명이 아닌 가설을 검정하는 것이다 (Song 등, 2007).

우산형 대립가설을 위한 비모수 검정법은 모형에 따라 많은 방법이 제안되었다. 일원배치모형에서는 Mann과 Whitney (1947)가 제안한 U통계량을 확장하여 우산형대립가설에 적용한 Mack과 Wolfe (1981)가 제안한 검정법이 있다. 그 외에도 Bhat (2009), Salman (2010) 그리고 Hettmansperger과 Norton (1987)이 제안한 방법이 있다. Kim과 Kim (1992)은 Mack과 Wolfe (1981)가 제안한 통계량을 랜덤화 블록 디자인으로 확장하여 제안한 방법이 있다.

[1] (137-701) 서울시 서초구 반포동 505번지, 가톨릭대학교 의학통계학교실, 대학원생.
[2] 교신저자: (137-701) 서울시 서초구 반포동 505번지, 가톨릭대학교 의학통계학교실, 교수.
E-mail: djkim@catholic.ac.kr

Orban과 Wolfe (1982)는 두 처리간 효과의 차이를 검정하기 위해 위치 (placement)를 사용한 비모수 검정법을 제안하였다. 이 방법은 두 처리중 어느 한 처리에 대한 상대적 위치정보를 이용하여 처리효과의 차이를 검정하는 방법이다. 또한 Kim (1999)은 이것을 확장하여 위치에 바탕을 둔 대조군과 처리군의 비교방법을 제안하였고 최근에 Lee와 Kim (2012)은 대조군과 처리군 방법을 확장하여 순서대립가설에 대한 비모수적 방법론을 제안하였다.

본 논문에서는 Lee와 Kim (2012)가 제안한 방법을 확장하여 우산형 대립가설에 대한 새로운 비모수적 방법론을 제안하였다. 모의실험을 통하여 반복이 있는 랜덤화 블록 계획법에서 Kim과 Kim (1992)이 제안한 방법과, 모수적 검정법인 분산분석, 그리고 본 논문에서 제안한 검정방법의 검정력을 비교하였다.

2. 위치를 이용한 UCP 검정 방법

블록이 u개 있고 처리가 t개인 확률화 블록 계획법의 모형은

$$Y_{ijk} = \mu + \theta_i + \beta_j + \epsilon_{ijk} \quad (i = 1, 2, \ldots, t; \quad j = 1, 2, \ldots, u; \quad j = 1, 2, \ldots, n_{ij})$$

이다. 여기서 μ은 전체 평균을 나타내고 θ_i는 i번째 처리의 효과, β_j는 j번째 블록의 효과를 나타낸다. 또한 ϵ_{ijk}는 오차항이며, 동일한 연속분포를 따르는 서로 독립인 확률변수로 가정한다.

각 처리의 효과가 모두 동일하다는 귀무가설과 우산형 대립가설은

$$H_0 : \theta_0 = \theta_1 = \ldots = \theta_t \text{ vs } H_1 : \theta_0 \leq \theta_1 \leq \ldots \leq \theta_{p-1} \leq \theta_p \geq \theta_{p+1} \geq \ldots \geq \theta_{t-1} \geq \theta_t$$

이다.

위치를 이용한 UCP (updated control group) 검정 방법의 순서형 대립가설에서의 통계량을 우산형 대립가설로 확장하기 위해서 p (peak)를 기준으로 왼쪽 순서형 대립가설에 대한 위치 (placement) U^L_{ijk}는

$$\begin{aligned} N^L_{ij} U^L_{ijk} &= \sum_{h=1}^{n_{0j}} \chi(Y_{0jh}, Y_{ijk}) + \sum_{h=1}^{i-1} \sum_{s=1}^{n_{hj}} \chi(Y_{hjs}, Y_{ijk}) \\ &= [j\text{번째 블록에서 대조군 (0번째 처리군)과 } (i-1)\text{번째까지 처리군에서} \\ &\quad\ Y_{ijk}\text{보다 작거나 같은 표본의 갯수}] \end{aligned}$$

라고 정의한다. 여기서 $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, u$; $k = 1, 2, \ldots, n_{ij}$, $\chi(x, y)$는 $x \leq y$일 때 1, 그 외에는 0이고, $N^L_{ij} = n_{0j} + n_{1j} + \cdots + n_{i-1,j}$이다. 또한 오른쪽 순서형 대립가설에 대한 위치 (placement) U^R_{ijk}는

$$\begin{aligned} N^R_{t-i,j} U^R_{t-i,jk} &= \sum_{h=1}^{i-1} \sum_{s=1}^{n_{t-h,j}} \chi(Y_{t-h,js}, Y_{t-i,jk}) + \sum_{h=1}^{n_{tj}} \chi(Y_{tjh}, Y_{t-i,jk}) \\ &= [j\text{번째 블록에서 } (t-i+1)\text{번째 처리군과 대조군 } (t\text{번째 처리군})\text{까지 처리군에서} \\ &\quad\ Y_{t-i,jk}\text{보다 작거나 같은 표본의 갯수}] \end{aligned}$$

라고 정의한다. 여기서 $N^R_{t-i,j} = n_{tj} + n_{t-1,j} + \cdots + n_{t-i+1,j}$이다.

Kim (1999)이 제안한 블록에서 대조군과 처리군의 위치를 이용하는 방법을 확장하여 각 블록에서 대조군과 처리군의 위치를 이용하는 왼쪽과 오른쪽 순서형 대립가설에 대한 통계량은 각각

$$S^{L}_{j,up} = \sum_{i=1}^{p}\sum_{k=1}^{n_{ij}} \phi_{N_{ij}}(U^{L}_{ikj}),\ j = 1, 2, \ldots, u$$

$$S^{R}_{j,up} = \sum_{i=1}^{t-p}\sum_{k=1}^{n_{ij}} \phi_{N_{t-i,j}}(U^{R}_{t-i,jk}),\ j = 1, 2, \ldots, u$$

와 같이 정의된다. 이때 $\phi_{N_{ij}}$는 $[0, 1]$의 범위에서 정의된 실수값인 점수함수 (score function)이다.

또한 서로 독립인 각 블록의 $S^{L^\star}_{j,up}$와 $S^{R^\star}_{j,up}$의 통계량을 이용하는 우산형 대립가설의 검정통계량은

$$S_{up} = \sum_{j=1}^{u}(S^{L^\star}_{j,up} + S^{R^\star}_{j,up}) = \sum_{i=1}^{p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{\phi_{N_{ij}}(U^{L}_{ijk})}{\sqrt{n_{0j}}} + \sum_{i=1}^{t-p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{\phi_{N_{t-i,j}}(U^{R}_{t-i,jk})}{\sqrt{n_{tj}}}$$

이며 귀무가설 $H_0 : \theta_0 = \theta_1 = \ldots = \theta_t$를 검정하기 위한 기각역은 $S_{up} \geq s_\alpha$이다. 여기서, s_α는 귀무가설 하에 $P_0[S_{up} \geq s_\alpha] = \alpha$를 만족하는 상수이다.

본 논문에서는 점수함수로 정규점수함수 (normal score function) $\phi_{N_{ij}}(x)=\Phi^{-1}(x)$ ($\Phi^{-1}(x)$는 표준정규분포 누적분포함수의 역함수)와 지수점수함수 (exponential score function) $\phi_{N_{ij}}(x)=-ln(1-x)$를 이용한 방법을 제안한다. 각각의 점수함수에 따른 통계량은 다음과 같다.

$$S^{NS}_{up} = \sum_{i=1}^{p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{\Phi^{-1}\left(\dfrac{N^{L}_{ij}U^{L}_{ijk}+1}{N^{L}_{ij}+2}\right)}{\sqrt{n_{0j}}} + \sum_{i=1}^{t-p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{\Phi^{-1}\left(\dfrac{N^{R}_{t-i,j}U^{R}_{t-i,jk}+1}{N^{R}_{t-i,j}+2}\right)}{\sqrt{n_{tj}}}$$

$$S^{E}_{up} = -\sum_{i=1}^{p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{ln\left(1-\dfrac{N^{L}_{ij}U^{L}_{ijk}}{N^{L}_{ij}+1}\right)}{\sqrt{n_{0j}}} - \sum_{i=1}^{t-p}\sum_{j=1}^{u}\sum_{k=1}^{n_{ij}} \frac{ln\left(1-\dfrac{N^{R}_{t-i,j}U^{R}_{t-i,jk}}{N^{R}_{t-i,j}+1}\right)}{\sqrt{n_{tj}}}$$

검정통계량 S_{up}의 근사분포를 위해서 Kim 등 (2011)의 결과를 이용한다. 이때 표준화된 $S^{L^\star}_{j,up}, S^{R^\star}_{j,up}$의 근사분포는

$$S^{L^\star}_{j,up} = \frac{S^{L}_{j,up} - E_0(S^{L}_{j,up})}{\sqrt{n_{0j}}} \backsim N(0, \sigma^2_{j,up,L}),\ j = 1, 2, \ldots, u$$

이고, 이때 $n_{0j} \to \infty, n_{ij} \to \infty, \rho_{ij} = \dfrac{n_{ij}}{n_{0j}} \to \rho_i,\ \tilde{\rho}_i = \dfrac{\rho_i}{1+\rho_1+\rho_2+\cdots+\rho_{i-1}}$, $0 < \rho_i < \infty$ 이며, 분산 $\sigma^2_{j,up,L}$는

$$\sigma^2_{j,up,L} = \sum_{i=1}^{p}\tilde{\rho}_i(\tilde{\rho}_i+1)\sigma^2_\phi, \quad \sigma^2_\phi = \int_0^1 \phi^2_{N_{ij}}(x)\,dx - \left(\int_0^1 \phi_{N_{ij}}(x)\,dx\right)^2$$

이다. 또한

$$S^{R^\star}_{j,up} = \frac{S^{R}_{j,up} - E_0(S^{R}_{j,up})}{\sqrt{n_{tj}}} \backsim N(0, \sigma^2_{j,up,R}),\ j = 1, 2, \ldots, u$$

이고, 이때 $n_{tj} \to \infty, n_{ij} \to \infty, \rho_{ij} = \dfrac{n_{ij}}{n_{tj}} \to \rho_i, \tilde{\rho}_i = \dfrac{\rho_{t-i}}{1+\rho_{t-1}+\rho_{t-2}+\cdots+\rho_{t-i+1}}$, $0 < \rho_i < \infty$ 이며, 분산 $\sigma^2_{j,up,R}$는

$$\sigma^2_{j,up,R} = \sum_{i=1}^{t-p}\tilde{\rho}_i(\tilde{\rho}_i+1)\sigma^2_\phi,\ \sigma^2_\phi = \int_0^1 \phi^2_{N_{t-i,j}}(x)\,dx - \left(\int_0^1 \phi_{N_{t-i,j}}(x)\,dx\right)^2$$

이다. 또 $S_{j,up}^{L^\star}$와 $S_{j,up}^{R^\star}$ 통계량이 각각 블록에서 독립이고 근사적으로 정규분포를 따르기 때문에

$$S_{up} = \sum_{j=1}^{u}(S_{j,up}^{L^\star} + S_{j,up}^{R^\star}) \backsim N\left(0, \sum_{j=1}^{u}(\sigma_{j,up,L}^2 + \sigma_{j,up,R}^2)\right)$$

이고 이를 이용하여 주어진 가설을 검정을 할 수 있다.

3. 모의실험의 계획 및 결과

우산형 대립가설에 대한 분석방법을 비교하기 위하여 모수적 검정방법으로 분산분석 (ANOVA), 비모수적인 방법으로는 Kim과 Kim (1992)을 사용하였다. 여기에 위치를 사용하는 방법을 제안하여 점수함수로 정규점수함수와 지수점수함수를 이용한 검정방법의 검정력을 비교해 보았다. 실행 프로그램은 SAS를 사용하였다. 모집단의 분포로는 정규분포, 지수분포, Cauchy분포, 이중지수분포 그리고 균일분포를 채택하였으며, 정규분포의 난수생성은 RANNOR 함수, 지수분포의 난수생성은 RANEXP 함수, 균일분포의 난수생성은 RANUNI 함수, Cauchy분포의 난수생성은 RANCAU 함수를 이용하였다. 이중지수분포는 RANUNI 함수를 이용하여 역변환 방법으로 난수를 생성하였다.

처리와 블록의 수는 각각 4개인 경우를 선택하였으며 표본의 크기는 같을 때를 고려하였다. 또한 검정통계량의 특성상 처리가 4개인 경우에는 정점 (peak)이 2일 때와 3일 때가 같은 통계량을 가지므로 정점이 2일때를 고려하였다. 그리고 표본의 크기는 3과 5일 때를 관찰하였으며 유의수준 α를 0.05로 보정하기 위해 확률화 검정을 사용하였다. 이러한 조건에서 각 검정통계량들이 기각역에 포함되는지를 판단하는 과정을 10,000번 반복하여 결과를 Table 3.1과 3.2 그리고 3.3으로 정리하였다.

각 처리의 효과가 모두 동일할 경우에 유의수준이 0.05를 만족하는지 살펴보면, 분산분석법의 실험유의수준은 표본의 크기가 3일 때 정규 분포와 지수분포인 Table 3.1에서 0.0514, 0.0549 그리고 Cauchy분포와 이중지수분포인 Table 3.2에서 0.0172, 0.0571, 균일분포인 Table 3.3에서 0.2221의 값들을 얻었다. 표본의 크기가 5일 때는 정규 분포와 지수분포인 Table 3.1에서 0.0414, 0.0547 그리고 Cauchy분포와 이중지수분포인 Table 3.2에서 0.0174, 0.0539, 균일분포인 Table 3.3에서 0.3459의 값들을 얻었다. 표본수에 따라 실험유의수준이 크게 차이나지는 않았고 정규성을 만족하지 않을 경우와 균일분포일 때, 분산분석법은 1종 오류를 제어하지 못함을 의미한다. 또한 Kim과 Kim (1992)이 제안한 분석법의 실험유의수준은 표본의 크기가 3일 때 정규 분포와 지수분포인 Table 3.1에서 0.0124, 0.0613, Cauchy분포와 이중지수분포인 Table 3.2에서 0.0410, 0.0613, 균일분포인 Table 3.3에서 0.1021의 값을 얻었다. 표본의 크기가 5일때는 정규 분포와 지수분포인 Table 3.1에서 0.0139, 0.0710, 이중지수분포와 Cauchy분포인 Table 3.2에서 0.0416, 0.0710, 균일분포인 Table 3.3에서 0.1230의 값을 얻었다. 모든 분포에서 표본수가 증가함에 따라 실험 유의수준이 약간 증가함을 볼 수 있고 비모수 방법임에도 불구하고 정규성을 만족하지 않을 경우, 1종오류를 제어하기 힘들다. 본 논문에서 제안한 방법들은 각 분포에서의 실험유의수준으로 1종 오류를 제어하는데 문제가 없음을 보여주고 있다.

검정력을 살펴보면 Cauchy분포에서는 모든 대립가설에서 점수함수로 정규점수함수를 사용한 UCP 검정법의 효율이 가장 좋았다. 하지만 다른 분포에서는 대립가설의 형태에 따라 다름을 알 수 있다. 우선 정규분포에서는 전체적으로 Cauchy분포와 같이 점수함수로 정규점수함수를 사용한 UCP검정법의 효율이 가장 좋지만 초반부터 커지다가 평행을 이루는 대립가설의 형태에서 평균의 차이가 클수록 분산분석의 효율이 좋다. 또한 중간 정도까지 평행하다가 작아져서 또다시 평행하는 대립가설에서도 평균의 차이가 클수록 분산분석의 효율이 좋음을 알 수 있다.

지수분포에서는 대립가설의 형태에 따라 효율성이 많이 달라지는것을 알 수 있다. 초반부터 커지다가 평행을 이루는 대립가설의 형태에서는 정규분포와 같이 평균의 차이가 클수록 분산분석의 효율이 좋

음을 알 수 있다. 또한 중간정도까지 평행하다가 작아져서 또다시 평행하는 대립가설에서는 평균의 차이가 작을때는 Kim과 Kim (1992)이 제안한 방법의 효율이 좋지만 평균의 차이가 클수록 분산분석의 효율이 좋아짐을 알 수 있다. 초반에 평행을 이루다가 마지막에 작아지는 대립가설과 처음에 작았다가 중간부분에서 커져 평행을 이루고 마지막에 다시 작아지는 대립가설에서는 정규점수함수를 사용한 UCP검정법의 효율이 항상 좋았다. 그리고 정점부분에서만 큰 대립가설에서는 평균의 크기가 작을때는 정규점수함수를 사용한 UCP검정법의 효율이 좋지만 평균의 차이가 커질수록 Kim과 Kim (1992)이 제안한 방법의 효율이 좋았다. 하지만 처음부터 커져 평행을 이루다가 점점에 작아지는 대립가설과 처음에 커졌다가 점점 작아지는 대립가설에서는 대체적으로 Kim과 Kim (1992)의 효율이 좋음을 알 수 있다.

이중지수분포에서도 지수분포와 같이 대립가설의 형태에 따라 효율성이 많이 달라지는 것을 알 수 있는데 초반부터 커지다가 평행을 이루는 대립가설의 형태에서는 평균의 차이가 클수록 정규점수함수를 사용한 UCP방법의 효율성이 좋음을 알 수 있다. 또한 중간정도까지 평행하다가 작아져서 또다시 평행하는 대립가설과 처음부터 커져 평행을 이루다가 점점에 작아지는 대립가설에서는 Kim과 Kim (1992)가 제안한 방법의 효율이 좋았지만 초반에 커져서 평행을 이루다가 마지막에 작아지는 대립가설과 처음에 커졌다가 점점 작아지는 대립가설에서는 정규점수함수를 사용한 UCP검정법의 효율이 좋음을 알 수 있었다. 그리고 점점부분에서만 큰 대립가설과 초반에 평행을 이루다가 마지막에 작아지는 대립가설에서는 평균의 차이가 작을때는 정규점수함수를 사용한 UCP방법이 효율적이지만 평균의 차이가 클수록 Kim과 Kim (1992)의 방법이 효율적임을 알 수 있었다. 마지막으로 균일분포에서는 모든 대립가설에서 점수함수로 정규점수함수를 이용한 UCP방법보다 지수점수함수를 이용한 방법이 좋음을 알 수 있다.

Table 3.1 Monte carlo power estimates (α=0.05)

Dist	n	θ_1	θ_2	θ_3	θ_4	UNS	UES	F	A
Normal	3	0	0	0	0	0.0500	0.0501	0.0514	0.0124
		0	0.2	0.2	0.2	0.0880	0.0736	0.0639	0.0192
		0.4	0.4	0	0	0.1711	0.1141	0.1414	0.0672
		0.6	0.6	0.6	0.1	0.3619	0.2230	0.2561	0.1316
		0	0.8	0	0	0.7048	0.2970	0.3884	0.4451
		0	0.6	0.6	0	0.6845	0.4153	0.3387	0.3054
		0.8	0.8	0.4	0.1	0.4701	0.3345	0.2764	0.2279
		0	0.2	0.1	0	0.1740	0.1237	0.0621	0.0503
		0.1	0.8	0.4	0	0.8359	0.5410	0.3598	0.4512
	5	0	0	0	0	0.0500	0.0500	0.0414	0.0139
		0	0.2	0.2	0.2	0.0965	0.0802	0.0663	0.0169
		0.4	0.4	0	0	0.2070	0.1251	0.2350	0.1032
		0.6	0.6	0.6	0.1	0.5165	0.3080	0.4372	0.2340
		0	0.8	0	0	0.8449	0.6478	0.5941	0.6844
		0	0.6	0.6	0	0.8739	0.5947	0.6113	0.5579
		0.8	0.8	0.4	0.1	0.6011	0.3841	0.5014	0.3300
		0	0.2	0.1	0	0.2102	0.1374	0.0618	0.0369
		0.1	0.8	0.4	0	0.9632	0.7277	0.6231	0.6811
Exponen-tial	3	0	0	0	0	0.0500	0.0501	0.0549	0.0613
		0	0.2	0.2	0.2	0.0730	0.0570	0.0636	0.0606
		0.4	0.4	0	0	0.1162	0.0674	0.1150	0.1290
		0.6	0.6	0.6	0.1	0.2594	0.1260	0.1817	0.2350
		0	0.8	0	0	0.5421	0.2552	0.2933	0.7011
		0	0.6	0.6	0	0.4780	0.2306	0.2292	0.4283
		0.8	0.8	0.4	0.1	0.3981	0.1138	0.2890	0.4051
		0	0.2	0.1	0	0.1207	0.0595	0.0640	0.1105
		0.1	0.8	0.4	0	0.6374	0.2878	0.2793	0.6694
	5	0	0	0	0	0.0501	0.0501	0.0547	0.0710
		0	0.2	0.2	0.2	0.0795	0.0644	0.0769	0.0693
		0.4	0.4	0	0	0.1677	0.0945	0.1837	0.2063
		0.6	0.6	0.6	0.1	0.3626	0.1826	0.3053	0.3575
		0	0.8	0	0	0.7423	0.3851	0.4782	0.8099
		0	0.6	0.6	0	0.6544	0.3222	0.3829	0.6501
		0.8	0.8	0.4	0.1	0.4701	0.1831	0.4628	0.7154
		0	0.2	0.1	0	0.1770	0.0884	0.0745	0.1672
		0.1	0.8	0.4	0	0.8184	0.4186	0.4549	0.8311

** UNS=UCP method using normal score function, UES=UCP method using exponential score function
F=ANOVA, A= The method proposed by Kim and Kim

Table 3.2 Monte carlo power estimates (α=0.05)

Dist	n	θ_1	θ_2	θ_3	θ_4	UNS	UES	F	A
		0	0	0	0	0.0501	0.0501	0.0172	0.0410
		0	0.2	0.2	0.2	0.0659	0.0595	0.0169	0.0427
		0.4	0.4	0	0	0.0921	0.0719	0.0213	0.0613
		0.6	0.6	0.6	0.1	0.1461	0.1086	0.0228	0.0835
	3	0	0.8	0	0	0.3314	0.1941	0.0314	0.1641
		0	0.6	0.6	0	0.2445	0.1639	0.0260	0.1354
		0.8	0.8	0.4	0.1	0.2014	0.1544	0.0314	0.1010
		0	0.2	0.1	0	0.0940	0.0727	0.0169	0.0555
Cauchy		0.1	0.8	0.4	0	0.3591	0.2114	0.0344	0.1847
		0	0	0	0	0.0515	0.4950	0.0174	0.0416
		0	0.2	0.2	0.2	0.0729	0.0620	0.0178	0.0435
		0.4	0.4	0	0	0.1052	0.0773	0.0208	0.0746
		0.6	0.6	0.6	0.1	0.1990	0.1310	0.0241	0.1103
	5	0	0.8	0	0	0.4011	0.2611	0.0330	0.3071
		0	0.6	0.6	0	0.3484	0.2163	0.0276	0.2040
		0.8	0.8	0.4	0.1	0.2258	0.1547	0.0278	0.1801
		0	0.2	0.1	0	0.1078	0.0805	0.0180	0.0618
		0.1	0.8	0.4	0	0.4011	0.3077	0.0297	0.2176
		0	0	0	0	0.0501	0.0500	0.0571	0.0613
		0	0.2	0.2	0.2	0.0627	0.0550	0.0625	0.6350
		0.4	0.4	0	0	0.0908	0.0642	0.0768	0.0994
		0.6	0.6	0.6	0.1	0.1759	0.0995	0.1064	0.1469
	3	0	0.8	0	0	0.3995	0.1914	0.1351	0.5286
		0	0.6	0.6	0	0.3514	0.1903	0.1129	0.2836
		0.8	0.8	0.4	0.1	0.2126	0.1084	0.1340	0.3308
		0	0.2	0.1	0	0.0890	0.0658	0.0594	0.0822
Double Exponen -tial		0.1	0.8	0.4	0	0.6676	0.2585	0.1239	0.5460
		0	0	0	0	0.0535	0.0600	0.0539	0.0710
		0	0.2	0.2	0.2	0.0784	0.0801	0.0591	0.0773
		0.4	0.4	0	0	0.1123	0.0925	0.1115	0.1254
		0.6	0.6	0.6	0.1	0.2262	0.1693	0.1549	0.1941
	5	0	0.8	0	0	0.5118	0.3504	0.2314	0.6430
		0	0.6	0.6	0	0.4482	0.3202	0.1858	0.3684
		0.8	0.8	0.4	0.1	0.2716	0.1853	0.2293	0.4023
		0	0.2	0.1	0	0.1180	0.1067	0.0622	0.1043
		0.1	0.8	0.4	0	0.6679	0.4026	0.2130	0.6439

** UNS=UCP method using normal score function, UES=UCP method using exponential score function
F=ANOVA, A= The method proposed by Kim and Kim

Table 3.3 Monte carlo power estimates (α=0.05)

Dist	n	θ_1	θ_2	θ_3	θ_4	UNS	UES	F	A
		0	0	0	0	0.0573	0.0531	0.2221	0.1021
		0	0.2	0.2	0.2	0.0783	0.0937	0.5286	0.1413
		0.4	0.4	0	0	0.0140	0.0464	0.9528	0.5317
		0.6	0.6	0.6	0.1	0.0076	0.0839	0.9944	0.7225
	3	0	0.8	0	0	0.0659	0.2546	1.0000	1.0000
		0	0.6	0.6	0	0.0367	0.3188	1.0000	0.9834
		0.8	0.8	0.4	0.1	0.0030	0.0333	1.0000	0.9625
		0	0.2	0.1	0	0.0118	0.0125	0.4981	0.6021
Uniform		0.1	0.8	0.4	0	0.0311	0.4427	1.0000	1.0000
		0	0	0	0	0.0557	0.0515	0.2359	0.1230
		0	0.2	0.2	0.2	0.0948	0.1198	0.6434	0.2000
		0.4	0.4	0	0	0.0192	0.0667	0.9972	0.7360
		0.6	0.6	0.6	0.1	0.0353	0.0995	1.0000	0.9353
	5	0	0.8	0	0	0.0838	0.2816	1.0000	1.0000
		0	0.6	0.6	0	0.0556	0.3372	1.0000	1.0000
		0.8	0.8	0.4	0.1	0.0240	0.0481	1.0000	0.9980
		0	0.2	0.1	0	0.0156	0.0334	0.6415	0.6453
		0.1	0.8	0.4	0	0.0590	0.5061	1.0000	1.0000

** UNS=UCP method using normal score function, UES=UCP method using exponential score function
F=ANOVA, A= The method proposed by Kim and Kim

4. 결론 및 고찰

본 논문에서는 확률화 블록 계획법에서 우산형 대립가설에 대한 분포무관 검정법을 제안하였다. 이 통계량은 Orban과 Wolfe (1982)의 논문에서 사용된 위치 (placement)를 사용하여 만들어졌다.

모의실험을 통하여 이 검정방법을 정규분포, Cauchy분포, 지수분포, 이중지수분포 그리고 균일분포

에서 모수적 검정방법과 여러 점수함수를 사용한 경우와 비교한 결과 분포에 따라 효율성이 다름을 알 수 있었다. 또한 균일분포를 제외하고 점수함수로 지수점수함수를 이용했을때보다는 정규점수함수를 이용한 경우가 효율성이 좋음을 알 수 있다. 모의실험의 전체적인 결과를 살펴보면 점수함수로 지수점수함수를 이용했을때는 정규점수함수를 이용했을때보다 낮은 검정력을 보여주었다. 또한 정규분포와 Cauchy분포를 가정했을 때는 본 논문에서 제시한 위치를 이용하여 정규점수함수를 사용한 방법의 검정력이 Kim과 Kim (1992)가 제안한 방법보다 검정력이 높았고 심지어 대부분 분산분석법의 검정력보다 높았다. 하지만 모의실험 결과에서 이중지수분포와 지수분포를 가정했을때는 검정력이 대립가설의 형태에 따라 달라짐을 알 수 있다. 또한 균일분포일 때는 지수점수함수를 사용한 방법의 검정력이 높음을 알 수 있다. 이러한 경향의 원인은 실험 유의수준의 차이와 각 검정방법에 차이가 있기 때문이라고 볼 수 있다. 효율성은 표본의 크기가 커질수록 차츰 차이가 줄어드는 것을 볼 수 있다.

따라서 분포에 따라서 본 논문에서 제시한 위치를 이용한 검정방법을 사용하는 것이 효율적인 분석이 될 수 있다. 또한 정규분포를 포함한 여러 분포에서도 다른 검정 방법의 검정력보다 위치를 이용한 검정방법의 검정력이 크게 낮아지지 않는 것으로 나타났다. 하지만 미지의 블록 효과가 존재하기 때문에 비모수 방법의 장점인 분포 무관의 성질은 유지하면서 블록간의 정보를 추출해 내는 것은 쉽지 않다는 문제가 있다. 또한 대립가설의 형태마다 검정력이 다른데 반복이 있는 랜덤 블록화 계획법에서 모든 경우에 검정력이 높은 검정법의 연구가 필요하다.

대부분의 자료에서 정점이 알려져 있지 않은 경우가 많은데 이러한 경우에는 연관성을 나타내는 상관계수의 방향성을 이용해 우산형 패턴의 정점을 찾은 Shin과 Kim (2005)을 참고하여 정점을 구하여 검정을 할 수 있다.

References

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

Hettmansperger, T. P. and Norton, R. M. (1987). Tests for patterned alternatives in k-sample problems. *The Korean Journal of Applied Statistics*, **82**, 292-299.

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, D. and Kim, Y. (1992). On the distribution-free tests for umbrella alternatives in a randomized block design. *The Korean Journal of Applied Statistics*, **5**, 41-57.

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. (2012). Nonparametric procedures using placement in randomized block design with replications. *Journal of the Korean Data & Information Science Society*, **22**, 1105-1112.

Mack, G. A. and Wolfe, D. A. (1981). K-sample 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 the 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 placenents. *Jornals of the American Statistical Association*, **77**, 666-671.

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

Shin, M and Kim, D. (2005). *Test of umbrella pattern with ordinal response data*, Master Thesis, University of Catholic, Seoul.

Song, M., Park, C and Lee, J. (2007). *Nonparametric statistical methods*, 1st Ed., Freeacademy, Seoul.

Journal of the Korean Data & Information Science Society
2013, **24**(6), 1401–1408

http://dx.doi.org/10.7465/jkdi.2013.24.6.1401
한국데이터정보과학회지

Nonparametric method using placement in a randomized complete block design

Sujin Sim[1] · Dongjae Kim[2]

[12]Department of Biostatistics, The Catholic University of Korea

Received 25 September 2013, revised 29 October 2013, accepted 11 November 2013

Abstract

Kim and Kim (1992) proposed typical nonparametric method for umbrella alternative in randomized block design with replications. In this paper, We consider a test procedure for umbrella alternatives in a randomized block design using extension of the two sample placement tests described in Orban and Wolfe (1982) and treatment tests described in Kim (1999). We perform a Monte Carlo study to compare the empirical powers of the test statistics for underlying distributions.

Keywords: Nonparametric, placement, umbrella alternatives.

[1] Researcher, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea.
[2] Corresponding author: Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701, Korea. E-mail: djkim@catholic.ac.kr