한국통계학회논문집 제4권 1호
1997년 4월 pp. 167-175

확률화 블럭 계획법에서 동위회귀를 이용한 우산형 대립가설의 비모수검정법[1)]

김 동 희[2)] , 김 영 철[3)]

요 약

확률화 블럭 계획법에서 동위회귀를 이용하여 우산형 대립가설에 대한 비모수검정법을 제안하고자 한다. 제안된 검정통계량은 Mack과 Wolfe (1981)의 통계량에서 처리들에 가중치를 준 형태가 되며, 동위회귀를 이용하여 확률변수인 가중치를 구하고 붓스트랩을 이용한 소표본에서의 모의 실험을 통하여 몇가지 사례 및 분포에 대해 제안된 통계량의 검정력을 알아본다.

제 1 장 서 론

n개의 관측치들을 특정 수준에 따라 t개의 집단으로 나누었을 때, 특정 수준에 따른 처리효과가 t개의 집단들에 대해 같은지 아니면 일정한 수준까지 처리효과가 증가하다가 감소하는지를 알아 보는데 관심을 둘 수 있다.

Mack과 Wolfe (1981)는 Jonckheere (1954) 통계량을 이용하여 k-표본에서 우산형 대립가설에 대한 분포무관 검정법을 제시하였고, Simpson과 Margolin (1986)은 약물의 복용 증가가 어느 시점에서는 약물효과의 감소로 나타나는 약물복용-반응 관계에 대한 반복적 비모수검정법을 제시하였다. 또한, Hettmansperger와 Norton (1987)은 선형 순위 통계량을 사용한 비모수적인 방법을 제시하고 Mack과 Wolfe (1981)의 통계량과 비교하였다. 김동희와 김영철 (1996)은 블럭 가중치를 이용한 검정통계량과 Mack과 Wolfe 형태의 검정통계량과의 점근상대효율이 최대가 되는 가중치 즉, 최적가중치를 이용한 우산형 대립가설에 대한 비모수검정법을 제시하고 Mack과 Wolfe의 통계량과 Hettmansperger와 Norton의 선형 순위 통계량과의 검정력을 비교하였다.

본 논문에서는 교호작용이 없는 확률화 블럭 모형

$$X_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}, \quad i = 1, \cdots, b;\ j = 1, \cdots, t;\ k = 1, \cdots, n_{ij}$$

에서 처리효과의 우산형 대립가설에 대한 비모수검정법을 다루고자 하는데, μ는 전체 평균, α_i는 블럭효과로서 장애모수로 취급하며, β_j는 처리효과이고, e_{ijk}는 서로 독립이고 동일한 분포를 갖

---

1) 이 논문은 1996년도 부산대학교 기성회 학술연구조성비 지원에 의하여 부산대학교 기초과학연구소에서 연구수행되었음.(RIBS-PNU-96-106).

2) 609-735 부산광역시 금정구 장전동 부산대학교 통계학과 교수
3) 609-735 부산광역시 금정구 장전동 부산대학교 통계학과

는 확률변수이며 이의 연속분포함수 F는 $\int f^2(x)dx < \infty$인 밀도함수 f를 갖는다.

우리는 다음과 같은 가설을 검정하고자 한다.

$$H_0: \beta_1 = \cdots = \beta_t$$

$$\text{vs } H_1: \beta_1 \le \cdots \le \beta_{h-1} \le \beta_h \ge \beta_{h+1} \ge \cdots \ge \beta_t \text{ (적어도 하나의 부등식을 만족).}$$

본 논문에서는 위와 같은 처리효과의 우산형 대립가설에 대한 비모수검정법을 제시하고자 하는데, 동위회귀를 이용하여 검정통계량에서 처리의 가중치를 구하고, 제안된 검정통계량의 정확한 분포를 구하기 어려우므로 붓스트랩을 이용하여 실험분포를 구하고 몇가지 사례 및 분포에 대하여 소표본 모의실험을 통한 실험검정력을 구하고자 한다. 2장에서는 본 논문에서 다루고자 하는 검정통계량을 제시하고, 동위회귀를 이용하여 검정통계량의 가중치를 구하는 방법을 설명하고 붓스트랩 기법을 사용하여 검정력을 구하는 알고리즘을 제시한다. 3장에서는 소표본에서의 모의실험을 수행하여 결과를 분석해 본다.

제 2 장 제안된 검정통계량

2.1 동위회귀 함수

교호작용이 없는 확률화 블럭 계획법에서 처리효과의 우산형 대립가설을 검정하기 위해 Kim과 Kim (1992)은 다음과 같은 Mack과 Wolfe 형태의 통계량을 제안하였다.

$$A = \sum_{i=1}^{b} \left\{ \sum_{l=1}^{h-1} \sum_{m=l+1}^{h} U_{ilm} + \sum_{l=h}^{t-1} \sum_{m=l+1}^{t} U_{iml} \right\}$$

여기서, U_{ilm}은 (i, l)칸과 (i, m)칸에 있는 관측치들에 관한 Mann-Whitney 통계량이다.

본 논문에서는 Mann-Whitney 통계량에 l번째 처리와 m번째 처리에 임의의 상수를 부여하고 이 상수들의 차이를 가중치로 사용하여 다음과 같은 검정통계량을 제안하고자 한다.

$$AI = \sum_{i=1}^{b} \left\{ \sum_{l=1}^{h-1} \sum_{m=l+1}^{h} \frac{(c_m - c_l)}{n_{il} n_{im}} U_{ilm} + \sum_{l=h}^{t-1} \sum_{m=l+1}^{t} \frac{(c_l - c_m)}{n_{il} n_{im}} U_{iml} \right\}$$

여기서, n_{ij}는 칸의 크기이며 c_j는 우산형 상수이다.

Hettmansperger와 Norton (1987)은 c_j를 모르기 때문에

$$c_j = \begin{cases} j & , j = 1, \cdots, h \\ 2t - j & , j = h+1, \cdots, t \end{cases}$$

를 우산형 상수로서 사용하였고 본 논문에서는 동위회귀를 이용하여 c_j들을 구하고자 한다.

동위회귀에 대한 정의는 다음과 같다.

정의) $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_t\}$는 우산형 순서관계 $\beta_1 \le \cdots \le \beta_{h-1} \le \beta_h \ge \beta_{h+1} \ge \cdots \ge \beta_t$ 를 갖는 집합이라 하자. 벡터 $\boldsymbol{c} = (c_1, \cdots, c_t)$는 $\beta_l \le \beta_m$ 이면 $c_l \le c_m$ 이 되는 동위함수이다. 벡터

$\overline{X} = (\overline{X}_{\cdot 1 \cdot}, \cdots, \overline{X}_{\cdot t \cdot})$, $\overline{X}_{\cdot j \cdot} = \sum_{i=1}^{b} \sum_{k=1}^{n_{ij}} X_{ijk} / \sum_{i=1}^{b} n_{ij}$ 과 양의 가중치 $\boldsymbol{w} = (w_1, \cdots, w_t)$ 가 주어졌을 때, $\boldsymbol{\beta}$상에서의 동위함수군에 속하는 $\boldsymbol{c}$에 대해 $\sum_{j=1}^{t} w_j (\overline{X}_{\cdot j \cdot} - c_j)^2$ 을 최소화하는 동위함수 $\boldsymbol{c}$를 가중치 $\boldsymbol{w}$를 갖는 동위회귀라 한다.

본 논문에서는 편의상 $\boldsymbol{w} = (1, \cdots, 1)$인 경우에 대해서만 $\boldsymbol{c}$를 구하였다. Barlow *et al.* (1972)과 shi (1988)는 동위회귀에 대한 알고리즘(PAVA)을 제시하였고 Cran (1980)과 Geng과 Shi (1990)는 동위회귀를 계산하기 위한 프로그램을 제시하여 응용의 폭을 넓혔다.

예) 연령이 증가함에 따라 지적수준이 증가하다가, 어떤 연령 이후에는 지적수준이 감소하는 경향이 있는지를 알아보기 위하여 Pan과 Wolfe (1996)가 정리한 Wechsler의 성인 지적수준 (Wechsler Adult Intelligence Scale, WAIS) 자료를 이용한다.

표1에서 정점이 3(20-24)인 경우, 남자는 16-17세에서 20-24세까지 지적수준이 증가하다가 그 이후부터는 감소하는 형태이므로 동위회귀 함수값들을 구할 필요가 없다. 반면에, 여자는 16-17세에서 20-24세까지 지적수준이 증가하지 않고 오히려 감소하고 있고 20-24세에서 25-34세까지도 감소하지 않고 증가하고 있음을 알수 있다. 따라서, 여자인 경우에 동위회귀함수값들을 구해보면 다음과 같다.

(1) 16-17세와 18-19세의 가중평균 $= (100 \times 9.97 + 100 \times 9.60)/(100 + 100) = 9.785$

연 령	16-17	18-19	20-24	25-34	35-44	45-54	55-64
표본크기	100	100	100	150	150	150	150
평 균	9.785	9.785	9.57	9.70	9.45	8.08	7.73

(2) 16-17세, 8-19세와 20-24세의 가중평균 $= (200 \times 9.785 + 100 \times 9.57)/(200 + 100) = 9.713$

연 령	16-17	18-19	20-24	25-34	35-44	45-54	55-64
표본크기	100	100	100	150	150	150	150
평 균	9.713	9.713	9.713	9.70	9.45	8.08	7.73

(3) 20-24세 이후는 평균값들이 감소를 하므로 2)에서 구한 평균값들이 동위회귀함수값들이다.

표 1. 연령과 성별에 따른 Wechsler의 성인 지적수준(WAIS) 자료의 표본크기와 평균값들

		연 령						
		16-17	18-19	20-24	25-34	35-44	45-54	55-64
남 자	표본크기	100	100	100	150	150	150	150
	평 균	9.79	9.87	10.41	10.25	10.09	9.15	8.34
여 자	표본크기	100	100	100	150	150	150	150
	평 균	9.97	9.60	9.57	9.70	9.45	8.08	7.73

2.2 붓스트랩 기법

확률변수 $c_j,\ j=1,\cdots,t$ 들이 포함된 검정통계량 AI의 정확한 분포를 구하는 일이 매우 어려울 뿐만 아니라 이론적 점근분포를 찾지 못하였기 때문에 붓스트랩 기법을 이용하여 유의수준 α에 대한 기각값을 결정하고자 한다.

붓스트랩을 이용한 검정통계량 AI의 기각역을 구하기 위한 알고리즘은 다음과 같다.

(1) 모분포로부터 각 칸에 $n_{ij},\ i=1,\cdots,b,\ j=1,\cdots,t$ 개의 확률표본을 생성한다.
(2) $X_{ijk},\ i=1,\cdots,b,\ j=1,\cdots,t,\ k=1,\cdots,n_{ij}$ 를 사용하여 동위회귀 함수 $c=(c_1,\cdots,c_t)$ 를 구하고, 검정통계량 AI 를 계산한다.
(3) (1) - (2) 과정을 R회(1000회) 반복하여 유의수준 α에서 $\Pr(AI\geq C_\alpha)\leq\alpha$ 를 만족하는 최소의 상수 C_α를 계산한다.
(4) 각 유의수준 α에서 (1) - (3) 과정을 S회(500회) 수행하여 구한 S개 기각값 $C_{\alpha1},\cdots,C_{\alpha S}$의 중앙값을 CV_α라 둔다. 이 CV_α를 붓스트랩 분포에 근거한 검정통계량 AI의 기각값으로 한다.

제 3 장 소표본 모의 실험 및 결론

본 논문에서 제시된 검정통계량의 실험검정력과 실험유의수준을 알아보기 위하여, 균일, 정규, 이중지수, 코시, 오염정규분포들에 대하여 모의실험을 수행하였다. ε-오염정규분포의 누적 분포함수는 $F(x)=(1-\varepsilon)\Phi(x)+\varepsilon\Phi(x/\sigma)$ 이고 Φ는 표준정규분포함수, $\varepsilon=0.1$ 과 $\sigma=3.0$ 이 사용되었다. 코시분포에서는 이차적률이 존재하지 않기 때문에, 표준정규분포의 표준편차에 대응하는 σ를 선택한다. 즉, $\int_{-\sigma}^{\sigma}(\pi(1+x^2))^{-1}dx=0.6827=\Phi(1)-\Phi(-1)$ 로 두고 구한 σ 값은 1.8326 이다.

4개의 블럭과 5개의 처리를 갖고 있는 확률화 블럭 계획법에서 칸의 크기는 다음과 같은 6가지 경우에 대해 모의실험을 수행한다.

Case 1 : $n_{ij}=3,\ i=1,\cdots,4,\ j=1,\cdots,5.$

Case 2 : $n_{ij}=5,\ i=1,\cdots,4,\ j=1,\cdots,5.$

Case 3 : $n_{1j}=2,\ n_{2j}=4,\ n_{3j}=6,\ n_{4j}=8,\ j=1,\cdots,5.$

Case 4 : $n_{1j}=8,\ n_{2j}=6,\ n_{3j}=4,\ n_{4j}=2,\ j=1,\cdots,5.$

Case 5 : $n_{i1}=3,\ n_{i2}=4,\ n_{i3}=5,\ n_{i4}=6,\ n_{i5}=7,\ i=1,\cdots,4.$

Case 6 : $n_{i1}=7,\ n_{i2}=6,\ n_{i3}=5,\ n_{i4}=4,\ n_{i5}=3,\ i=1,\cdots,4.$

또한, 정점이 2, 3 과 4일 때를 고려하고 각 정점에 대해서 처리효과들은 다음과 같다.

Peak 2 : $(2\delta\sigma/3\ ,\ \delta\sigma,\ 2\delta\sigma/3\ ,\ \delta\sigma/3\ ,\ 0)$

Peak 3 : $(\delta\sigma/3\ ,\ 2\delta\sigma/3\ ,\ \delta\sigma,\ 2\delta\sigma/3\ ,\ \delta\sigma/3)$

Peak 4 : $(0\ ,\ \delta\sigma/3\ ,\ 2\delta\sigma/3\ ,\ \delta\sigma,\ 2\delta\sigma/3)$

여기서 δ는 0.0(0.2)1.0 이고 σ는 각 모집단에서의 표준편차이다.

블럭이 4이고 처리가 5인 확률화 블럭 계획법에서 Case 1은 각 칸의 크기가 모두 3인 실험을 계획하고 있고 Case 2는 각 칸의 크기를 5로 하여 실험을 수행하고자 한다. Case 3은 같은 블럭 내에서는 칸의 크기가 같으나 블럭번호가 증가할수록 칸의 크기가 증가하도록 계획하였고 Case 4는 Case 3와 대조가 되도록 계획하였다. Case 5는 처리번호가 증가할수록 칸의 크기가 증가하고 Case 6은 Case 5와 비교하기 위해 대조적으로 실험을 계획하였다.

붓스트랩을 이용해 얻은 기각값으로 위와 같은 실험 계획을 1,000회 수행한 결과에 의하면, 정점이 2와 4인 경우의 실험검정력이 정점이 3인 경우의 실험검정력보다 항상 높게 나타났으며, Case 2의 실험검정력이 Case 1의 실험검정력보다 전반적으로 높게 나타났다. 이는 칸의 크기가 클수록 실험검정력이 높게 나타난다는 것을 보여주고 있다. Case 3과 Case 4의 실험검정력이 유사하게 나타나서, 블럭 크기가 증가하거나 감소함에 따라 실험검정력에는 아무런 영향을 주지 않음을 알 수 있다. Case 5와 Case 6의 실험검정력도 유사하게 나타나서, 처리번호가 증가함에 따른 칸의 크기의 증가와 감소에 실험검정력이 영향을 받지 않음을 보여주고 있다. 전체적으로 Case 2가 다른 경우들에 비해 실험검정력이 다소 높음을 알 수 있다. 또한, 균일분포와 정규분포에서의 실험검정력보다 꼬리가 두꺼운 분포에서의 실험검정력이 좋음을 알 수 있다.

표 2. AI의 실험검정력 ($\alpha = 0.05$)

분포	정점	Case 1 δ 0.0	0.2	0.4	0.6	0.8	1.0	Case 2 δ 0.0	0.2	0.4	0.6	0.8	1.0
균일분포	2	0.048	0.106	0.206	0.315	0.444	0.577	0.055	0.124	0.275	0.439	0.608	0.734
	3	0.049	0.095	0.156	0.280	0.355	0.470	0.056	0.103	0.203	0.336	0.476	0.666
	4	0.043	0.097	0.206	0.311	0.438	0.578	0.067	0.122	0.266	0.438	0.588	0.756
정규분포	2	0.052	0.113	0.208	0.323	0.476	0.595	0.057	0.138	0.275	0.455	0.625	0.748
	3	0.048	0.103	0.154	0.273	0.380	0.489	0.059	0.094	0.214	0.343	0.497	0.683
	4	0.043	0.099	0.207	0.324	0.458	0.608	0.063	0.121	0.269	0.454	0.626	0.774
이중지수분포	2	0.047	0.134	0.235	0.387	0.563	0.684	0.057	0.153	0.334	0.527	0.701	0.820
	3	0.047	0.110	0.180	0.320	0.451	0.573	0.054	0.099	0.254	0.413	0.585	0.772
	4	0.037	0.115	0.261	0.382	0.554	0.684	0.062	0.141	0.335	0.535	0.697	0.825
코시분포	2	0.049	0.117	0.152	0.204	0.291	0.309	0.045	0.091	0.165	0.241	0.345	0.384
	3	0.048	0.084	0.118	0.182	0.226	0.294	0.046	0.074	0.129	0.215	0.301	0.362
	4	0.049	0.085	0.155	0.223	0.289	0.353	0.048	0.083	0.180	0.266	0.339	0.384
오염정규분포	2	0.049	0.110	0.243	0.405	0.572	0.709	0.049	0.135	0.337	0.498	0.690	0.831
	3	0.044	0.109	0.200	0.333	0.463	0.602	0.048	0.126	0.259	0.399	0.582	0.754
	4	0.062	0.120	0.250	0.405	0.538	0.722	0.044	0.157	0.344	0.527	0.689	0.833

표 3. AI의 실험검정력 ($\alpha = 0.05$)

분포	정점	Case 3 δ 0.0	0.2	0.4	0.6	0.8	1.0	Case 4 δ 0.0	0.2	0.4	0.6	0.8	1.0
균일분포	2	0.045	0.110	0.240	0.399	0.554	0.724	0.062	0.135	0.255	0.413	0.547	0.723
	3	0.060	0.097	0.184	0.342	0.485	0.652	0.047	0.114	0.199	0.344	0.477	0.639
	4	0.051	0.108	0.238	0.400	0.586	0.712	0.043	0.128	0.255	0.412	0.557	0.721
정규분포	2	0.060	0.108	0.235	0.405	0.577	0.747	0.057	0.126	0.271	0.426	0.571	0.734
	3	0.055	0.102	0.199	0.354	0.488	0.675	0.043	0.118	0.199	0.347	0.483	0.647
	4	0.049	0.115	0.243	0.430	0.598	0.733	0.049	0.122	0.256	0.427	0.585	0.744
이중지수분포	2	0.065	0.118	0.287	0.489	0.671	0.809	0.058	0.149	0.309	0.503	0.645	0.800
	3	0.051	0.121	0.236	0.411	0.585	0.764	0.043	0.127	0.249	0.402	0.554	0.746
	4	0.047	0.132	0.294	0.504	0.661	0.818	0.048	0.148	0.311	0.515	0.654	0.807
코시분포	2	0.058	0.081	0.163	0.236	0.289	0.397	0.054	0.092	0.146	0.263	0.291	0.383
	3	0.054	0.081	0.148	0.208	0.250	0.359	0.041	0.087	0.148	0.214	0.227	0.348
	4	0.044	0.086	0.135	0.257	0.298	0.384	0.040	0.101	0.166	0.256	0.266	0.371
오염정규분포	2	0.039	0.130	0.282	0.521	0.701	0.815	0.053	0.142	0.293	0.524	0.676	0.810
	3	0.044	0.126	0.225	0.404	0.593	0.718	0.039	0.114	0.208	0.405	0.569	0.765
	4	0.042	0.173	0.318	0.513	0.689	0.788	0.060	0.142	0.288	0.491	0.695	0.816

표 4. AI의 실험검정력 ($\alpha = 0.05$)

분포	정점	Case 5 δ 0.0	0.2	0.4	0.6	0.8	1.0	Case 6 δ 0.0	0.2	0.4	0.6	0.8	1.0
균일분포	2	0.051	0.136	0.272	0.434	0.617	0.731	0.050	0.116	0.229	0.395	0.561	0.685
	3	0.049	0.090	0.184	0.348	0.480	0.626	0.050	0.101	0.207	0.290	0.493	0.614
	4	0.049	0.108	0.220	0.398	0.562	0.695	0.047	0.135	0.304	0.433	0.634	0.729
정규분포	2	0.057	0.139	0.270	0.446	0.609	0.747	0.049	0.116	0.232	0.408	0.568	0.706
	3	0.052	0.095	0.195	0.340	0.500	0.669	0.058	0.096	0.187	0.328	0.495	0.625
	4	0.053	0.112	0.229	0.401	0.575	0.723	0.047	0.129	0.302	0.468	0.645	0.737
이중지수분포	2	0.059	0.155	0.323	0.535	0.684	0.803	0.047	0.129	0.281	0.497	0.657	0.798
	3	0.048	0.103	0.241	0.402	0.578	0.746	0.056	0.111	0.243	0.397	0.582	0.707
	4	0.056	0.117	0.258	0.480	0.647	0.805	0.052	0.157	0.351	0.521	0.696	0.790
코시분포	2	0.066	0.106	0.139	0.258	0.322	0.381	0.048	0.098	0.133	0.245	0.322	0.376
	3	0.042	0.069	0.121	0.214	0.277	0.336	0.050	0.064	0.135	0.190	0.281	0.356
	4	0.050	0.076	0.138	0.229	0.310	0.382	0.052	0.112	0.172	0.243	0.324	0.386
오염정규분포	2	0.042	0.147	0.319	0.498	0.713	0.791	0.050	0.151	0.297	0.479	0.665	0.806
	3	0.057	0.130	0.250	0.380	0.592	0.745	0.048	0.122	0.236	0.418	0.586	0.725
	4	0.042	0.148	0.328	0.508	0.658	0.809	0.043	0.143	0.322	0.538	0.705	0.808

참고 문헌

[1] 김동희, 김영철 (1996). 확률화 블럭 계획법에서 최적 가중치를 이용한 우산형 대립가설의 비모수검정법, 「응용통계연구」, 제9권 1호.

[2] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference under Order Restrictions*, New York : Wiley.

[3] Cran, G.W. (1980). Algorithm AS 149 : Amalgamation of Means in the Case of Simple Ordering, *Applied Statistics*, **29**, 209-211.

[4] Geng, Z. and Shi, N.Z. (1990). Isotonic Regression for Umbrella Orderings, *Applied Statistics*, **39**, 397-424.

[5] Hettmansperger, T.P. and Norton, R.M. (1987). Tests for Patterned Alternatives in K-sample Problems, *Journal of the American Statistical Associations*, **82**, 292-299.

[6] Jonckheere, A.P. (1954). A Distribution-Free K-sample Test against Ordered Alternatives, *Biometrika*, **41**, 133-145.

[7] Kim, D.H. and Kim, Y.C. (1992). Distribution-Free Tests for Umbrella Alternatives in a Randomized Block Design, *Journal of Nonparametric Statistics*, **1**, 277-285.

[8] Mack, G.A. and Wolfe, D.A. (1981). K-sample Rank Tests for Umbrella Alternatives, *Journal of the American Statistical Associations*, **76**, 175-181.

[9] Pan, G. and Wolfe, D.A. (1996). Comparing Groups with Umbrella Orderings, *Journal of the American Statistical Associations, 91, 311-317.*

[10] Shi, N.Z. (1988). A Test of Homogeneity for Umbrella Alternatives and Tables of the Level Probability, *Communications in Statistics-Theory and Methods*, **A17(3)**, 657-670.

[11] Simpson, D.G. and Margolin, B.H. (1986). Recursive Nonparametric Testing for Dose-Response Relationship Subject to Downturns at High Doses, *Biometrika*, **73**, 589-596.