응용통계연구
제 4 권 제 1 호, 1991년

이원배치모형에서 순서대립가설에 대한 점근분포무관검정법에 관한 연구*

송 문 섭** 김 진 흠

〈요 약〉

본 논문에서는 이원배치모형에서 처리효과의 순서대립가설을 검정하기 위한 점근분포무관 검정법을 제안하고 제안한 통계량의 점근정규성과 일반화된 Puri의 통계량과의 점근상대효율을 살펴보았다. 또한 소표본에서 Monte Carlo연구를 통하여 제안된 통계량을 기존의 다른 방법들과 비교 연구하였다.

1. 서 론

우리가 다루고자 하는 이원배치모형은 다음과 같다.

$$X_{ijk} = \mu + \beta_i + \tau_j + \varepsilon_{ijk}$$
$$i=1, \cdots, I; \quad j=1, \cdots, J; \quad k=1, \cdots, n_{ij}$$

여기서 μ는 전체평균이고 β_i는 블록효과이며 τ_j는 처리효과이다. 그리고 ε_{ijk}는 오차항으로서 서로 독립이고 동일한 분포를 갖는 확률변수이다.

본 논문에서는 다음과 같은 순서대립가설의 검정문제에 관심이 있다.

귀무가설 $H_0 : \tau_1 = \tau_2 = \cdots = \tau_J$

대립가설 $H_1 : \tau_1 < \tau_2 < \cdots < \tau_J$ (적어도 한 부등식은 순증가)

이원배치모형에서 순서대립가설에 대한 비모수적 검정법들이 많이 제안되어 왔다. 대표적인 검정법들을 열거해보면 다음과 같다. $n_{ij}=1$인 경우로서는 Page(1963), Hollander(1967) 등의 검정법이 있고, $n_{ij}>1$인 경우로서는 Hettmansperger(1975), Govindarajulu and Mansouri-Ghiassi(1986) (이하의 G-M의 약자 사용) 등의 검정법을 들 수 있다. Kim, Song

* 본 연구는 한국학술진흥재단 연구비 89-21에 의하여 지원받았음

** (151-742) 서울시 관악구 신림동 서울대학교 계산통계학과 및 통계연구소

and Kim(1986)과 김동희, 엄동훈(1990)도 비슷한 내용을 다루었으며, 이 두 논문에서는 블록내의 순위에 대하여 가중순위합을 사용한 것이 특징이다.

Hettmansperger와 G-M의 검정법을 요약하면 다음과 같다.

(1) Hettmansperger 검정법

R_{ijk}를 i번째 블럭내에서 X_{ijk}의 순위라고 할 때, Hettmansperger 검정통계량은 다음과 같이 정의된다.

$$T = \sum_i \sum_j j \cdot (R_{ij\cdot}/n_{ij})$$

여기서 $R_{ij\cdot}$는 (i, j)번째 칸내의 순위들의 합이다.

(2) G-M 검정법

R_{ijk}를 혼합표본에서 $X_{ijk}-\hat{\beta}_i$의 순위라고 하자. 다만, $\hat{\beta}_i$는 β_i의 $\sqrt{n}$ – 일치추정량이다. G-M 검정통계량은 다음과 같이 정의된다.

$$L = \sum_j b_j \overline{R}_{\cdot j\cdot}$$

여기서 b_j는 $b_1 \leq b_2 \leq \cdots \leq b_J(b_1 \fallingdotseq b_J)$인 상수들이고 $\overline{R}_{\cdot j\cdot}$는 j번째 처리에 대한 순위들의 평균이다.

Hettmansperger 방법은 블록내의 순위를 사용하므로 분포무관 검정법이 되나 블록간의 정보를 이용하지 못한 단점이 있다. 이에 대하여 M-G방법은 혼합표본에서 잔차인 $X_{ijk}-\beta_i$의 순위를 사용하여 효율을 높이는 대신에 점근적으로만 분포무관인 방법이 된다.

본 논문에서는 장애모수인 블럭효과 β_i들을 추정한 후에 잔차들에 대하여 Jonckheere 형태의 통계량을 적용한 점근분포무관검정법을 제안하고자 한다. 제안된 검정법의 모수적 검정법에 대한 점근상대효율은 특수한 경우에 Wilcoxon검정의 t검정에 대한 점근상대효율과 같으며, 소표본 Monte Carlo연구 결과에 의하면 꼬리가 두터운 분포에서는 제안된 검정법이 기존의 검정법보다 검정력이 우수한 것으로 나타났다.

2. 제안된 통계량과 그의 성질들

표현의 편의를 위하여 다음과 같은 기호를 사용하기로 한다.

$$n_{i\cdot} = \sum_j n_{ij}, \; n_{\cdot j} = \sum_i n_{ij}, \; n = \sum_i \sum_j n_{ij}$$

$$\overline{X}_{i\cdot\cdot} = \sum_j \sum_k X_{ijk} \,/\, n_{i\cdot}, \; \overline{X} = \sum_i \sum_j \sum_k X_{ijk}/n$$

$$Y_{i(j,\, j',\, j')} = \text{median}\{X_{ijk} - X_{i'j'k'} \mid 1 < k, \quad k' < n_{ij}\}$$

$$Z_{ijk}(\hat{\beta}) = X_{ijk} - \hat{\beta}_i$$

여기서, $\hat{\beta}_i$은 β_i의 $\sqrt{n}$ - 일치추정량이며, 이러한 일치추정량으로서는 최소제곱 추정량과 Lehmann(1963)에 의해서 제안된 로버스트 추정량이 있다. 전자는

$$\hat{\beta}_i = \overline{X}_{i\cdot\cdot} - \overline{X}$$

이고, 후자는 다음과 같이 주어진다.

$$\hat{\beta}_i = (IJ^2)^{-1} \sum_j \sum_{i'} \sum_{j'} Y_{i(j,i',j')}$$

$\mathrm{Suv}(\hat{\beta})$를 다음과 같이 정의하자.

$$\mathrm{Suv}(\hat{\beta}) = \sum_i \sum_{i'} \sum_s \sum_t \Psi(Z_{i'vt}(\hat{\beta}) - Z_{ius}(\hat{\beta}))$$

여기서 $\Psi(t)$는 $t>0$일 때 $\Psi(t)=1$, $t\leq 0$일 때 $\Psi(t)=0$으로 정의되는 지시함수이다. 본 논문에서 제안된 통계량은 다음과 같다.

$$S(\hat{\beta}) = \sum_{u<v} \mathrm{Suv}(\hat{\beta})$$

순서대립가설 H_1하에서는 각 $u<v$에 대하여 $S_{uv}(\hat{\beta})$의 값이 클 것을 기대할 수 있으며 따라서 검정통계량 $S(\hat{\beta})$의 값이 클 때 귀무가설 H_0를 기각한다. 이제 $S(\hat{\beta})$의 점근성질을 알아보자.

$S(\beta)$는 H_0하에서 분포무관이다. 따라서 적당한 조건하에서 $S(\hat{\beta})$와 $S(\beta)$가 점근적으로 동치임을 보여줌으로써 $S(\hat{\beta})$는 점근적으로 분포무관임을 보이고자 한다. 이를 위하여 잔차들에 근거한 통계량들의 점근 성질을 연구한 Randles(1984)의 결과를 사용하면 다음의 정리를 얻을 수 있다. 먼저 몇 가지 가정들을 정리하자.

가정 1 오차항의 확률밀도함수 f가 위로 유계이고 0을 중심으로 대칭이다.

가정 2 $\lim_{n\to\infty} n_{\cdot j}/n = \lambda_{\cdot j}$, $0<\lambda_{\cdot j}<1$, $j=1, \cdots, J$가 성립한다.

정리 1 가정 1을 가정하면 귀무가설 H_0하에서

$$\sqrt{n}\,(U_{uv}(\hat{\beta}) - U_{uv}(\beta)) \xrightarrow{P} 0$$

이다. 여기서, U_{uv}는 다음과 같이 정의되는 U-통계량이다.

$$U_{uv}(\gamma) = \sum_i \sum_{i'} (\sum_s \sum_t \Psi\,(Z_{i'vt}(\gamma) - Z_{ius}(\gamma))/n_{iu}\,n_{i'v})$$

정리 1에서 다음의 정리 2를 얻을 수 있으며, 정리 2로부터 $S(\hat{\beta})$는 $S(\beta)$와 점근적으로 동치이고 따라서 $S(\hat{\beta})$는 점근분포무관 검정법임을 알 수 있다.

정리 2 가정 1과 가정 2를 가정하면 귀무가설 H_0하에서 다음이 성립한다.

$$n^{-3/2}(S(\hat{\beta}) - S(\beta)) \xrightarrow{P} 0$$

다음의 정리 3은 $S(\hat{\beta})$의 점근정규성을 나타내는 정리이며, 이로부터 검정통계량의 기각값을 근사적으로 얻을 수 있다.

정리 3 가정 1과 가정 2를 가정하면 귀무가설 H_0하에서

$$[S(\hat{\beta}) - E_0(S(\beta))]/[\mathrm{Var}_0(S(\beta))]^{1/2} \xrightarrow{d} N(0, 1)$$

이다. 여기서, $E_0(S(\beta)$와 $\mathrm{Var}_0(S(\beta))$는 다음과 같다.

$$E_0(S(\beta) = (n^2 - \sum_j n_{.j}^2)/4$$

$$\mathrm{Var}_0(S(\beta)) = [n^2(2n+3) - \sum_j n_{.j}^2(2n_{.j} + 3)]/72$$

이제 $S(\hat{\beta})$에 기초한 검정법의 점근상대효율을 알아보기 위하여 다음과 같은 전이대립가설을 생각해보자.

$$H_{1n} : \tau_j = a_j \tau/\sqrt{n}, \quad j=1, \cdots, J$$

여기서 a_j는 $a_1 \leq a_2 \leq \cdots \leq a_J (a_1 \neq a_J)$인 상수들이고 $\tau(>0)$는 미지의 상수이다.

정리 4 가정 1, 가정 2 및 $\int f^2(t)dt < \infty$를 가정하면 전이대립가설 H_{1n}하에서

$$[S(\beta) - E(S(\beta))]/[\mathrm{Var}(S(\beta))]^{1/2} \xrightarrow{d} N(0, 1)$$

이다. 여기서, $E(S(\beta)$와 $\mathrm{Var}(S(\beta))$는 각각 다음과 같다.

$$E(S(\beta)) = \sum_{u<v} n_{.u} n_{.v} \int (1 - F(t - (a_v - a_u)\tau/\sqrt{n}))dF(t)$$

$$\mathrm{Var}(S(\beta)) = \mathrm{Var}_0(S(\beta))$$

제안된 검정법을 모수적 방벅과 비교하기 위하여 다음과 같은 검정통계량 A를 생각해보자. 검정통계량 A는 Puri(1965)가 생각한 모수적 방법을 일반화한 것으로서, 순서대립가설에 대한 합리적인 검정통계량이라고 할 수 있다.

$$A = \sum_i \sum_{u<v} \mathrm{d}_i \mathrm{n}_{iu} \mathrm{n}_{iv} (\overline{X}_{iv.} - \overline{X}_{iu.})$$

여기서, $d_1, \cdots, d_I$는 음이 아닌 상수이고 $\overline{X}_{ij}.$는 (i, j)번째 칸내의 관측값들의 평균이다.

정리 5 $\mathrm{Var}(X_{ijk}) = \sigma^2 < \infty$이면, 전이대립가설 H_{1n}하에서

$$n^{-3/2}(A - E(A))/[\mathrm{Var}(A)] \xrightarrow{d} N(0, 1)$$

이다. 여기서 $E(A)$와 $\mathrm{Var}(A)$는 각각 다음과 같다.

$$E(A) = \sum_i \sum_{u<v} \mathrm{d}_i \mathrm{n}_{iu} \mathrm{n}_{iv} (a_v - a_u) \ \tau/\sqrt{n}$$

$$\mathrm{Var}(A) = \sigma^2 \sum_i d_i^2 [(\sum_i n_{ij})^3 - \sum_i n_{ij}^3]/3$$

정리 4와 정리 5로부터 A에 대한 $S(\hat{\beta})$의 점근상대효율은 다음과 같이 주어진다.

$$\mathrm{ARE}(S(\hat{\beta}), A) = \lim_{n\to\infty} \left[\frac{\sigma(A) \sum n_{\cdot u} n_{\cdot v} (a_v - a_u) \int f^2(t)dt}{\sigma(S(\hat{\beta})) \sum_i \sum_{u<v} d_i n_{iu} n_{iv} (a_v - a_u)} \right]^2$$

여기서, $\sigma(A)$와 $\sigma(S(\beta))$는 각각 정리 5의 $\mathrm{Var}(A)$와 정리 3의 $\mathrm{Var}_0(S(\beta))$와 같다.

특히, $n_{ij} = n/IJ$로서 각 칸의 크기가 같고 $a_u = j$, $d_i = 1$일 때

$$\mathrm{ARE}(S(\beta), A) = 12\sigma^2 (\int f^2(t)dt)^2$$

이며, 이것은 Wilcoxon검정의 t검정에 대한 점근상대효율과 같은 결과이다.

3. 소표본 Monte Carlo 연구

소표본에서, 제안된 통계량 $S(\hat{\beta})$를 G-M통계량 L, Hettmanperger 통계량 T, 모수적 통계량 A와 비교하기 위하여 Monte Carlo연구를 하였다. 모의실험에서는 블록효과 4개 수준과 처리효과 3개 수준을 선택했다. 각 칸내의 반복횟수와 처리효과의 형태는 다음과 같다.

$$n_{i1} = 3, \quad n_{i2} = 5, \quad n_{i3} = 7\ ; \quad (\tau_1, \tau_2, \tau_3) = (-\delta\sigma, 0, \delta\sigma)$$

여기서 δ는 0.0에서 0.6까지 0.2씩 증가하며 σ는 각 모집단의 표준편차이다.

본 소표본 비교연구에서 고려한 분포는 균일분포, 정규분포, 오염정규분포, Cauchy분포이다. 여기서 ϵ-오염정규분포는 분포함수가

$$F(x) = (1-\epsilon)\Phi(x) + \epsilon\Phi(x/\sigma)$$

로 주어지는 분포로서, ϵ은 오염의 정도를 나타낸다.

각 모형과 분포함수에 대하여 주어진 크기의 표본을 IMSL에 있는 부프로그램을 사용하여 생성하고, 이 표본에 대하여 모든 검정통계량의 값을 계산한 다음에, 검정통계량의 값과 유의수준 5%에서 기각값을 비교한다. 이러한 실험을 500번 반복하여 검정통계량의 값이 기각값을 초과한 횟수를 세어서 그 결과를 500으로 나눈 것이 경험적 검정력이며, 그 결과가 표 1에 수록되어 있다. 표에서 S와 L에 사용된 상위첨자 *와 **는 각각 β의 추정에 최소제곱 추정량과 로버스트 추정량을 사용한 경우를 뜻하며, S는 제안된 통계량, L은 G-M 통계량을 뜻한다.

표 1에서 볼 수 있듯이 꼬리가 짧은 분포인 균일분포와 정규분포의 경우에는 통계량 A에 기초한 모수적 방법의 검정력이 다른 비모수적 방법보다 다소 우수하다. 그러나, 꼬리가 두

터운 분포인 오염정규분포와 Cauchy분포의 경우에는 비모수적 방법의 검정력이 모수적 방법보다 월등히 우수한 것으로 나타났다. 비모수적 방법들에서는 제안된 통계량 S**(또는 S*)가 거의 모든 경우에 G-M통계량 L**(또는 L*)와 Hettmansperger 통계량 T보다 조금씩 좋은 것으로 나타났다.

$\delta=0$인 경우는 경험적 유의수준을 뜻하며, 전체적으로는 비슷한 경향을 나타내고 있다. 다만, Cauchy분포에서는 S*와 L*가 보수적인 결과를 보이며, 오염정규분포에서는 모든 경우에 경험적 유의수준이 크게 나타났다.

결론적으로 제안된 검정통계량 S**는 계산이 복잡한 단점이 있으나 검정력이 다른 방법보다 우수한 것으로 나타났다.

〈표 1〉 검정통계량들의 실험검정력 ($\alpha=0.05$)

분 포	δ	검 정 통 계 량 S*	S**	L*	L**	T	A
균 일 분 포	0.0	0.058	0.054	0.058	0.052	0.046	0.052
	0.2	0.346	0.340	0.308	0.308	0.267	0.346
	0.4	0.726	0.724	0.720	0.712	0.670	0.784
	0.6	0.950	0.948	0.948	0.952	0.938	0.972
정 규 분 포	0.0	0.050	0.050	0.056	0.060	0.048	0.052
	0.2	0.352	0.356	0.326	0.346	0.286	0.340
	0.4	0.756	0.766	0.734	0.738	0.672	0.770
	0.6	0.948	0.948	0.956	0.956	0.940	0.964
오염정규분포 ($\epsilon=0.10$, $\sigma=5.0$)	0.0	0.072	0.086	0.080	0.080	0.066	0.072
	0.2	0.560	0.594	0.540	0.584	0.530	0.388
	0.4	0.966	0,966	0.944	0.956	0.944	0.750
	0.6	1.000	1.000	1.000	1.000	1.000	1.000
Cauchy 분포	0.0	0.026	0.048	0.028	0.054	0.048	0.038
	0.2	0.162	0.298	0.150	0.292	0.280	0.100
	0.4	0.424	0.648	0.386	0.622	0.642	0.150
	0.6	0.690	0.900	0.676	0.888	0.886	0.274

*는 β의 추정량으로 최소제곱추정량을 사용한 경우이고, **는 Lehmann의 로버스트 추정량을 사용한 경우이다.

◇ 참 고 문 헌 ◇

(1) 김동희, 임동훈(1990), 다요인실험계획에서 순서대립가설에 대한 비모수검정법의 연구, 응용통계연구 3, 11-25

(2) Hettmansperger, T. P.(1975), "Non-parametric inference for ordered alternatives in a randomized block design," *Psychometrika*, **40**, 53-62.

(3) Hollander, M. (1967), "Rank tests for randomized blocks when the alternatives have a priori ordering," *Annals of Mathematical Statistics*, **38**, 867-877.

(4) Kim, D. H., Song, M. S. and Kim, W. C. (1986), "A distribution-free rank test for ordered alternatives in randomized complete block design," *Journal of the Korean Statistical Society*, **15**, 9-25.

(5) Lehmann, E. L. (1963), "Robust estimation in analysis of variance," *Annals of Mathematical Statistics*, **34**, 957-966.

(6) Govinderajulu, Z. and Mansouri-Ghiassi, S. H. (1986), "On nonparametric tests for ordered alternatives in two-way layouts," *Advances in Order Restricted Statistical Inference, Lecture Notes in Statistics*, **37**, Springer-Verlag, 153-168.

(7) Page, E. B. (1963), "Ordered hypotheses for multiple treatments : a significance test for linear ranks," *Journal of American Statistical Association*, **58**, 216-230.

(8) Puri, M. L. (1965), "Some distribution-free k-sample rank tests of homogeneity against ordered alternatives," *Communications on Pure and Applied Mathematics*, **18**, 51-63.

(9) Randles, R. H. (1984), "On tests applied to residuals," *Journal of the American Statistical Association*, **79**, 349-354.

On Asymptotically distribution-Free Test for Ordered Alternatives in Two-Way Layouts

Moon Sup Song* and Jin-Heum Kim

⟨Abstract⟩

An Asymptotically distribution-free test is considered for testing parallelism against ordered alternatives in two-way layouts. The test procedure is based on a statistic which uses Jonckheere's idea after adjusting the block effects. Large-sample properties including the efficiency and the limiting distribution of the test statistic are obtained. The proposed test is compared with other tests through a small-sample Monte Carlo study.

* Department of Computer Science and Statistics, Seoul National University, Seoul, 151-742 Korea.