한국통계학회논문집 제2권 1호
1995년 4월 pp. 166-175

내부적 독립성에 대한 기하적 검정통계량[1)]

김기영[2)], 전명식[3)], 이광진[4)]

요 약

내부적 독립성 가설에 대해 전통적인 우도비원리 하에서 나온 검정통계량과 합교원리 하에서 나온 검정통계량들에 대한 자료분석적인 측면에서의 대안으로서 기하적 관점에서 유래된 하나의 heuristic 검정통계량이 제안된다. 아울러 기존 검정통계량들의 기하적 의미들도 살펴보았다. 나아가 제안된 검정통계량의 특성 및 점근분포를 유도하였으며, 모의실험을 통하여 기존 검정통계량들과의 검정력을 비교한다.

1. 내부적 독립성 검정

일단의 다변량자료들이 주어졌을 때 본격적인 다변량통계분석으로 들어가기에 앞서 효과적이고 경제적인 분석을 위한 예비단계로서 먼저 점점을 해 보아야 할 것들이 몇가지가 있다. 그 중에서 대표적인 것들은 고려되는 변수들 간의 '내부적 독립성(internal independence)', '선형성(linearity)', 그리고 '정규성(normality)'에 관한 진단이다. 이들 중에서 내부적 독립성의 진단은 다변량통계분석이 '다변량적'이기 위해 반드시 필요한 과정으로 판단된다. 즉 다변량자료가 주어진 경우, 독립성의 진단을 거쳐보아 극단적으로 모든 변수들이 서로 독립이라고 할 수 있다면 다변량분석을 할 필요없이 각 변수에 대해 일변량분석을 행하면 될 것이고, 다른 한 극단으로 변수들 간에 완전상관의 관계가 있다면 차원축약이 가능할 뿐만 아니라 이를 추가분석에서의 이론적 근거로 사용될 수 있을 것이다.

자료분석적 측면에서 내부적 독립성 검증의 첫단계는 각 변수들 사이의 표본상관계수들을 살펴보는 작업일 것이다. 만일 어느 두 변수간의 표본상관계수가 충분히 커서 서로 밀접히 상관되어 있는 것으로 나타난다면 전체적으로 변수들은 내부적 독립성을 유지하고 있다고 할 수 없을 것이다. 따라서 전체변수들에 대한 표본상관행렬의 비대각원소들이 거의 0에 가까운 경우 전체적인 변수들간의 무상관에 대한 가설(내부적 독립성)

$$\text{귀무가설 } H_0 : \text{모집단 공분산행렬이 대각행렬이다.} \qquad (1)$$

에 대한 검증은 자료분석적인 측면에서 매우 실제적이라 할 수 있다.

내부적 독립성 가설에 대한 검정통계량들로서는 우도비원리로부터 출발된 것으로는 Wilks (1935)의 통계량과 이를 향상시킨 것인 Box(1949)의 검정통계량, 그리고 합교원리에 따른 Roy &

---

1) 이 연구는 94년도 한국과학재단 연구비지원에 의한 결과임. 과제번호 941-0100-022-1
2) (136-701) 서울시 성북구 안암동 고려대학교 통계학과.
3) (136-701) 서울시 성북구 안암동 고려대학교 통계학과.
4) (301-729) 대전시 중구 목동 목원대학교 응용통계학과.

Bargmann (1958)의 step-down 검정절차와 Schuenemeyer & Bargmann(1978)의 단일 검정통계량이 있다. 그러나 Schuenemeyer & Bargmann의 단일 검정통계량은 내부적 독립성의 검정분야 이외에도 많은 응용성을 지닌 것으로 알려져 있지만 정규성의 가정하에서도 그의 귀무분포가 구해지지 않은 상태이다.

이 논문에서는 다변량정규분포 하에서 내부적 독립성 가설에 대한 기존의 몇가지 주요 검정통계량들의 기하적 의미들을 살펴보고(2절), 기하적 관점에서 유래된 역표본상관행렬에 근거한 새로운 검정통계량을 제안하고 자료분석적인 측면에서 그의 성질을 규명하고자 한다(3절). 나아가 모의실험을 통해 이의 활용가능성과 기존방법과의 비교를 꾀하고자 한다(4절).

2. 기존 검정통계량들과 기하적 성질들

이 절에서는 내부적 독립성의 가설에 대한 우도비원리와 합교원리에 근거한 기존의 대표적 검정통계량들을 언급하며 이들이 가지는 기하적 의미를 살펴보고자 한다. 편의상 다음과 같은 표기를 사용하기로 한다. 다변량정규분포 $N_m(\mu, \Sigma)$를 따르는 m개의 변수 $X_1, X_2, \cdots, X_m$에 대하여, N개의 개체로부터 나온 N×m 자료행렬을 X, 그리고 이의 표본상관행렬을 R이라고 하자. 그리고 이에 대응되는 모상관행렬은 P로 표시될 것이다.

2.1. 우도비원리 하에서 나온 검정통계량

우도비원리 하에서 구축된 Wilks(1935)의 검정통계량 LR은

$$LR = -2 \log \Lambda = -N \log \{ \mathrm{Det}(R) \} \tag{2}$$

으로, 이 통계량은 (1)의 가설 하에서 점근적으로 자유도 m(m-1)/2인 χ^2 분포에 따른다. 이에 대해 Box(1949)는 Det(R)의 적률생성함수에 대한 Bernoulli 다항식 확장을 이용하여 식(2)의 통계량보다 χ^2(df=m(m-1)/2)분포에 점근적으로 더 접근한, 따라서 좀 더 향상된, 검정통계량을 제안하였는데 이는 식(2)의 승수 N 대신 다음과 같은 C를 사용한 바 있다.

$$LRm = -C \log \{ \mathrm{Det}(R) \}, \quad \text{여기서 } C = N-1-(2m+5)/6. \tag{3}$$

우도비원리하에서 유도된 검정통계량들은 Det(R) 만의 함수들인데, 소위 산포계수(scatter coefficient)라고 불리는 Det(R)의 기하적 의미는 m개의 R-구조벡터(R-structural vector)들에 의해 만들어지는 평행다면체(parallel-pipe)의 체적이다. 이것이 검정통계량으로 사용될 수 있는 이유는 다음과 같은 기하적 관계를 통해 설명될 수 있다. 즉 귀무가설 하에서는 모든 P-구조벡터(P-structural vector)들이 서로 수직이 되는데, 이 정립형태와 R-구조벡터들의 정립형태의 닮음정도를 이들 구조벡터들이 이루는 평행다면체의 체적으로도 잴 수 있을 것이기 때문이다. 다시 말해서 P-구조벡터들에 의해 만들어지는 평행다면체의 체적은 1이 되는 반면, R-구조벡터들에 의해 만들어지는 평행다면체에서의 최소높이가 0에 가까와 진다면(극단적으로 어느 하나의 표본다중상관계

수라도 1에 가까와 진다면) 이 통계량의 값은 0으로 접근하기 때문이다. 즉, Det(R)은 [0,1] 사이의 값을 가지면서, Det(R)=0은 완벽한 의존성(deterministic dependence)을, 그리고 Det(R)=1의 경우는 곧 내부적 독립성(internal independence)을 의미하기 때문이다.

그러나 이와 같은 우도비원리 하에서의 검정통계량은 단지 귀무가설에 대한 검정의 역할만 할 뿐으로, 만일 귀무가설이 채택되지 못했을 때의 추가적인 분석에서 나타날 수 있는 어떤 관심이 있는 모수들의 함수에 대한 결합신뢰영역의 추론 등에 관해서는 아무런 도움도 주지 못한다는 단점이 있다(Bargmann,1981).

2.2 Step-Down 합교검정방법

Roy & Bargmann(1958)은 다변량정규분포를 따르는 m개 확률변수들 간의 내부적 독립성 가설에 대해 합교원리 하에서의 한 검정방법을 유도하였다. 물론 이 검정방법은 두 변수집단 간의 독립성을 검정하는 문제로까지 확장될 수 있는 것이다. 이 방법의 주안점은 미리 정해진 값보다 같거나 더 큰 결합신뢰계수를 가지는, 모수들의 어떤 함수에 대한, 동시신뢰영역을 얻을 수 있다는 것이다. 이 검정방법은 다중독립성의 가설에 대해서는 우도비 검정통계량에 대한 대안으로 이용될 수 있을 뿐만 아니라, 나아가 그 당시까지는 이 분야에서 불가능했던 동시신뢰영역의 구축을 가능하게 하였다. 이 점이 큰 의미를 가지는 이유 중의 하나는 동시신뢰영역은 다중독립성의 가설이 기각되는 경우 'culprit'변수가 어떤 것인지를 탐지해 내는 역할을 할 수 있기 때문이다.

이 접근법을 요약하면 다음과 같다. 내부적 독립성의 가설(1)은 다음과 같이 분할시킬 수 있다.

$$H_0 = \bigcap_{\text{for all } \underline{a}} H_{0\underline{a}} \tag{4}$$

여기서 $H_{0\underline{a}}:\ Corr(X_j,\, a_1X_1+\cdots+a_{j-1}X_{j-1})=0,\ \ \underline{a}=(a_1,\cdots,a_{j-1})'\neq 0',\quad j=2,\cdots,m$

이 사실을 이용하여 제안된 검정규칙은 다음과 같다. 만일

$$\bigcap_{j=2}^{m}\{r^2_{j.(1,2,\cdots,j-1)}\leq k\} \text{ 이라면 귀무가설 } H_0\text{를 채택} \tag{5}$$

$$\bigcup_{j=2}^{m}\{r^2_{j.(1,2,\cdots,j-1)}> k\} \text{ 이라면 귀무가설 } H_0\text{를 기각}$$

여기서 $r^2_{j.(1,2,\cdots,j-1)}$은 X_j와 $\{X_1, X_2, \cdots, X_{j-1}\}$의 표본다중상관계수의 제곱이며, k는 주어진 유의수준(α) 하에서 다음을 만족하는 것이다.

$$\prod_{j=2}^{m} P[\ r^2_{j.(1,2,\cdots,j-1)}\leq k \mid \rho^2_{j.(1,2,\cdots,j-1)}=0] = 1-\alpha \tag{6}$$

그런데 $r^2_{j.(1,2,\cdots,j-1)}$들은 $\rho^2_{j.(1,2,\cdots,j-1)}=0$ 하에서 Beta((j-1)/2,(n-j+1)/2)분포를 따르며, 모든 j (j=2,3,···,m) 에 대하여 서로 독립이므로, k의 결정은 incomplete beta function의 표들을 이용하면

된다.

다음으로 step-down 합교원리하에서 나온 이 검정통계량이 R-구조공간에서 가지는 기하적 의미를 살펴보면 다음과 같다. Lee(1992)에 의하면 Det(R)은

$$Det(R) = \prod_{j=2}^{m} (1 - r^2_{j,(1,2,\cdots,j-1)}) \tag{7}$$

으로 표현될 수 있으므로 log{Det(R)}은 (m-1)개의 독립적인 성분들의 합으로 분할됨을 볼 수 있다. 이로 보아 우도비원리 하에서는 m개의 P-구조벡터들에 의해서 만들어지는 평행다면체의 체적이 1이 된다는 사실에 비추어 R-구조벡터들에 의해 만들어지는 평행다면체의 체적 하나만으로 R-구조벡터들이 서로 수직인 P-구조벡터에 전체적으로 얼마나 가까운가를 본 반면, 합교원리 하에서의 step-down적 접근방법은 각각의 R-구조벡터들이 해당 P-구조벡터들에 얼마나 가까운지에 대한 독립적인 측도 m개를 모두 축차적으로 사용한다는데 그 의미를 부여할 수 있다.

2.3 합교원리하에서의 단일검정통계량

내부적 독립성에 대한 검정법에는 2.1절과 2.2절에서 언급된 통계량들이 있음에도 불구하고, 합교원리 하에서 나온 단일(single) 검정통계량의 필요성과 그것이 가지는 의미에 관해서는 Bargmann(1981)에 의해 토론된 바 있다.

두 변수집단 간의 독립성에 대한 단일 합교검정통계량은 정준상관계수이다. 그리고 m개의 변수들 간의 내부적 독립성에 대한 단일의 합교검정통계량은 1978년에서야 Schuenemeyer & Bargmann에 의해서 고안되었다. 그 이전까지만 해도 2.2절에서 설명된 Roy & Bargmann(1958)에 의해 제안된 m개의 'step-down' 상관계수들이 단순상관계수나 다중상관계수들에 대한 신뢰영역을 구하는데 그 역할을 다 하였다. 그러나 이 검정법에는 여러 개의 검정통계량이 사용된다는 이유때문에 인자분석등에서 나타나는 문제들, 예를 들어 인자적재행렬의 추정, 에는 적용이 불가능 하였다. 또한 범주형 변수들을 'normalizing'하는 문제에 있어서도 Lancaster의 원리를 3차원 이상으로 확장하는 데에는 단 하나의 'super-canonical'상관계수가 관심의 대상이 되고, m개의 변수들을 두 개의 집단으로 나누었을 때 나타날 수 있는 가장 큰 정준상관계수의 상한을 얻을 필요가 있을 수도 있다. 이런 문제들의 합교원리적 해결을 위해 내부적 독립성에 대한 단일 합교검정통계량이 필요했던 이유가 된다.

Schuenemeyer(1975)은 기하적 관점에서 출발하여 내부적 독립성에 대한 합교원리 하에서의 단일 검정통계량이 아래의 Q가 됨을 보였다.

$$Q = (\lambda_m - \lambda_1) \;/\; (\lambda_m + \lambda_1) \tag{8}$$

여기서 λ_1, λ_m은 각각 정칙인 크기 m x m인 표본상관행렬의 가장 작은, 큰 고유값이다.

이 검정통계량 Q는 $(\lambda_1/\lambda_m)^{1/2}$의 함수인데, 이의 기하적 의미는 상관행렬에 의해 만들어지는 다음과 같은 m차원 타원체 $x^t R^{-1} x = 1$를 원점을 지나는 평면으로 잘랐을 때 나타나는 타원들

중에서 가장 뾰족한(elongated) 타원의 장축의 길이를 단축의 길이로 나눈 값이다. 이 값이 1 근처에 있다는 것은 m차원 타원체 자체가 초구에 가깝다는 것을 의미하고 이는 결국 내부적 독립성을 지지해 주는 상황으로 간주될 수 있다. 그리고 이 때의 Q의 값은 0에 가깝다. 한편 이 값이 클수록 m차원의 타원체는 어느 한쪽으로는 좁게 되는데 이는 변수들 간에 유의적인 상관관계가 존재함을 의미하게 되며 이 때의 Q의 값은 1로 향하게 된다.

3. 통계량 $L = \underline{1}^t R^{-1} \underline{1}$과 이의 제 성질

이 절에서는 내부적 독립성의 가설에 대해서 2절에서 설명된 우도비원리나 합교원리 하에서의 검정통계량들과는 달리 순수히 기하적 사고로부터 출발된 검정통계량 L을 제안한다. 이는 앞에서도 언급되었듯이 우도비원리하에서의 검정통계량들이 가지는 약점과 합교원리에서 나온 Schuenemeyer의 단일 검정통계량의 분포가 아직까지 알려지지 않았다는 점 등을 고려하여 고안된 것이다.

자료분석적인 면에서 표본상관행렬의 모든 비대각원소들이 무상관을 보이는 상황에서 전체 변수들간의 무상관에 관한 가설검증을 위해 제안하는 통계량은 다음과 같다.

$$L = \underline{1}^t R^{-1} \underline{1} \quad \text{여기서 } \underline{1}\text{은 모든 원소가 1인 행벡터이다.} \tag{9}$$

검정통계량 L이 가지는 기하적 의미는 다음의 정리로부터 고찰될 수 있다.

[정리 3.1] <Lee,1991>
원점에서부터 모든 R-구조벡터들의 종점(end point)들을 지나는 초평면(hyperplane)까지의 최소거리는 $(\underline{1}^t R^{-1} \underline{1})^{-1/2}$이다.

이 통계량 L이 검정통계량으로 사용되어질 수 있는 이유는 다음과 같다. 즉, L은 $1<L<\infty$ 의 값을 가지면서, 내부적 독립이라는 귀무가설 하에서는 L = m 이 된다. 또한 고려되는 변수들이 모두 밀접한 양(+)의 상관관계를 가질수록 L은 1로 접근하고, 어느 하나의 단순상관계수가 -1에 가까와 짐에 따라 L은 ∞로 접근한다. L은 R만의 함수이므로 원래의 변수들에 대한 affine transformation하에 불변성(invariance)을 가진다.

한편 정규분포하에서 표본상관행렬의 미분가능한 함수로 된 일반적인 통계량의 점근(asymptotic)분포에 관한 Konishi(1979)의 결과를 이용하면, 표본의 수가 무한대로 갈 때 귀무가설하에서의 L의 점근분포를 구할 수 있다. 따라서 검정법에 필요한 귀무분포의 근사방법을 구한 셈이다.

[정리 3.2] <Lee,1991>
정규분포와 내부적 독립이라는 귀무가설하에서 L의 점근분포는 다음과 같다.

$$\Pr\{ n*(L-m)/\tau < x \} = \Phi(x) - 1/(n*)^{1/2} \{ a_1 \Phi^{(1)}(x)/(2\tau) + a_3 \Phi^{(3)}(x)/(2\tau)^3 \} + 0((n*)^{-1}) \quad (10)$$

여기서 n*는 n* = N-1-2d 이고 d는 근사방법에 필요한 수정인자(correction factor)이다. 그리고 Φ(j)(x)는 표준정규분포함수 Φ(x)의 j차 도함수이고, τ2 (>0)과 a1, a3의 계산과정은 Konishi(1979)에 주어져 있다.

4. 모의실험

기하적 배경과 함께 새로이 제안된 검정통계량 L이 가지는 자료분석적인 측면에서의 성능을 모의실험을 통해 살펴보고 그의 활용가능성을 알아보고자 한다.

우선 L통계량의 점근분포의 효율을 보기 위하여 다음과 같은 모의실험 작업을 수행하여 보았다. 내부적 독립성을 만족하는 다변량정규분포(변수의 개수가 2, 3, 4, 5)에서 표본의 크기가 50, 100, 200, 300인 표본상관행렬을 독립적으로 10,000개씩을 반복생성하여, 귀무분포로서 식(10)을 이용한 검정통계량 L이 유의수준(0.05, 0.10)을 잘 유지하고 있는지를 여러 가지의 수정인자에 대해 살펴본 결과는 아래 <표1>과 같다. <표1>에서는 지면상 검정통계량 L에 대해서만 유의수준의 유지여부를 제시하고 있지만, 검정통계량 LR에 대해서도 같이 점검해본 바 이도 잘 유지되고 있음을 알 수 있었다. 필요한 표본상관행렬의 생성은 IMSL을 사용하여 구한 Wishart행렬의 Bartlett decomposition을 활용하였다.

<표1>를 보면 유의수준 0.05와 0.10하에서 적절한 수정인자를 결정하면 유의수준이 대체로 잘 만족되고 있음을 알 수 있다. 그러나 수정인자의 값의 결정방법은 Konishi(1979)에도 언급되어 있지 않을 뿐만 아니라, Konishi와의 개인교신에 의하면 Konishi(1993)는 "It was, however, difficult to give a clear description on the optimal choice for d(수정인자) theoretically"라고 답하였듯이, 저자들도 여러가지 모의실험을 통하여 결정규칙을 얻을려고 노력해 보았으나 불행히도 결정규칙을 찾을 수는 없을 것이라는 결론을 얻었다. 그러나 표를 통해서도 볼 수 있듯이 표본의 크기와 유의수준이 커질수록 수정인자의 결정폭은 상당히 줄어들고 있다는 사실을 확인할 수 있다. 또한 변수의 개수가 많아질수록 수정인자의 값은 커지는 경향이 있다. 표에서 배경색이 주어진 부분은 주어진 유의수준에 가장 가까운 것을 표시한 것으로 이에 해당하는 수정인자의 값이 다른 통계량과의 검정력비교에서 고려될 것이다.

<표1>

변수의 개수=2		수정인자						
유의 수준	표본의 크기	0	1	2	3	4	5	6
0.05	50	0.0611	0.0568	0.0511	0.0475	0.0438	0.0384	0.0333
	100	0.0565	0.0536	0.0511	0.0478	0.0461	0.0439	0.0417
	200	0.0498	0.0488	0.0468	0.0449	0.0434	0.0423	0.0410
	300	0.0535	0.0524	0.0517	0.0510	0.0503	0.0496	0.0489
0.10	50	0.1087	0.1016	0.0942	0.0885	0.0810	0.0738	0.0666
	100	0.1030	0.0986	0.0948	0.0912	0.0868	0.0839	0.0812
	200	0.1075	0.1058	0.1032	0.1016	0.0992	0.0977	0.0959
	300	0.1048	0.1034	0.1028	0.1018	0.1005	0.0994	0.0981

변수의 개수=3		수정인자						
유의 수준	표본의 크기	2	3	4	5	6	7	8
0.05	50	0.0682	0.0632	0.0577	0.0517	0.0460	0.0411	0.0373
	100	0.0570	0.0546	0.0526	0.0502	0.0481	0.0458	0.0426
	200	0.0540	0.0524	0.0509	0.0498	0.0487	0.0479	0.0470
	300	0.0537	0.0534	0.0524	0.0514	0.0509	0.0501	0.0493
0.10	50	0.1129	0.1049	0.0969	0.0900	0.0816	0.0731	0.0658
	100	0.1064	0.1029	0.0993	0.0957	0.0915	0.0874	0.0839
	200	0.1055	0.1034	0.1023	0.1006	0.0988	0.0965	0.0946
	300	0.1019	0.1009	0.0999	0.0987	0.0981	0.0976	0.0966

변수의 개수=4		수정인자						
유의 수준	표본의 크기	7	8	9	10	11	12	13
0.05	50	0.0561	0.0496	0.0427	0.0386	0.0334	0.0287	0.0237
	100	0.0565	0.0548	0.0529	0.0502	0.0480	0.0457	0.0424
	200	0.0526	0.0515	0.0505	0.0494	0.0484	0.0473	0.0462
	300	0.0503	0.0500	0.0492	0.0485	0.0482	0.0475	0.0469
		4	5	6	7	8	9	10
0.10	50	0.1114	0.1035	0.0964	0.0884	0.0807	0.0739	0.0674
	100	0.1078	0.1041	0.1007	0.0977	0.0950	0.0910	0.0870
	200	0.1051	0.1030	0.1015	0.0996	0.0976	0.0959	0.0942
	300	0.1044	0.1039	0.1031	0.1017	0.1006	0.0987	0.0981

변수의 개수=5		수정인자						
유의수준	표본의 크기	8	9	10	11	12	13	14
0.05	50	0.0718	0.0669	0.0625	0.0581	0.0543	0.0492	0.0455
	100	0.0626	0.0602	0.0583	0.0551	0.0532	0.0504	0.0471
	200	0.0578	0.0566	0.0548	0.0534	0.0519	0.0507	0.0495
	300	0.0553	0.0547	0.0538	0.0531	0.0524	0.0515	0.0505
		6	7	8	9	10	11	12
0.10	50	0.1196	0.1109	0.1029	0.0981	0.0911	0.0834	0.0774
	100	0.1056	0.1019	0.0980	0.0951	0.0910	0.0872	0.0844
	200	0.1088	0.1068	0.1057	0.1047	0.1032	0.1015	0.0993
	300	0.1014	0.1005	0.0988	0.0981	0.0973	0.0958	0.0951

우도비원리 하에서 나온 Box의 개량된 검정통계량과의 검정력비교를 위해서 아래의 몇가지 대립가설의 상황을 선택하였다. 참고로 합교원리 하에서 유도된 통계량 U의 귀무분포는 아직까지 알려지지 않은 상태이므로 이는 비교대상에서 제외되었다.

아래 <표3>는 귀무분포의 모집단 상관행렬이 등상관행렬인 경우에 대한 검정력비교에 대한 결과이고, <표2>은 대립분포의 모집단 상관행렬 P가 아래와 같은 형식을 가지는 경우에 대한 검정력비교에 관한 것이다.

$$P=\begin{bmatrix} 1.000 & & & & \\ 0.010 & 1.000 & & \text{대칭} & \\ 0.015 & 0.020 & 1.000 & & \\ 0.020 & 0.025 & 0.030 & 1.000 & \\ 0.025 & 0.030 & 0.035 & 0.040 & 1.000 \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

<표3>에서 변수의 개수가 2인 경우는 L통계량이나 LR통계량의 검정력은 아주 비슷함을 볼 수 있는데, 이는 변수의 개수가 2인 경우는 L통계량과 LR통계량이 함수적으로 일대일 대응관계를 가지기 때문이다. 여기에서의 경미한 차이는 모의실험에서의 오차한계와 두 통계량의 점근적 귀무분포함수의 차이에 귀속된다. 즉, L통계량의 점근적 귀무분포의 누적분포함수는 Konishi(1979)의 결과를 적용하여 누적정규분포함수들의 항으로 표현되었고, 이에 대해 LR통계량의 점근적 귀무분포는 통상적인 카이제곱분포가 사용되었다. 등상관구조를 지니는 대립분포에 대해서 L통계량이 LR통계량보다 일률적으로 모두 검정력이 높음을 볼 수 있다. 지면상 이 논문에서는 생략되었지만 음,양의 다양한 등상관계수에 대해서도 비교를 해 본 결과 또한 여기에 제시된 표의 내용과 유사한 패턴을 지니고 있었다. 마찬가지로 <표2>에서도 L통계량이 LR통계량보다 일률적으로 모두 검정력이 높음을 보이고 있다.

이런 현상들은 LR통계량은 아무런 정보가 없는 상황에서 이용되는 것을 전제로 해서 만들어진 통계량이지만, L통계량은 실제 자료분석적인 입장에서 일차적으로 표본상관행렬의 각 원소들이 아주 작아서 개별 모상관계수들이 0이라는 것이 확인된 상황을 전제로 하고 있다는 사실을 반영하는 것으로 해석될 수 있을 것이다. 만일 L통계량의 전제조건을 그대로 반영할 수 있는 LR통계량의 유도과정이 존재한다면, 이로부터 유도될 LR통계량보다 L통계량이 우수하다고는 주장할 수 없겠지만 현재까지는 이런 현실적인 상황을 고려한 LR통계량은 존재하지 않는 것으로 알고 있다.

<표2> <참고: 표에서 100, 200, 300은 표본의 개수임>

P의 주대각원소 아래의 행렬	(위)유의수준 0.05 / (아래) 유의수준 0.10					
	L			LR		
	100	200	300	100	200	300
.01 .015 .02	0.0820 0.1433	0.1120 0.1846	0.1496 0.2323	0.0653 0.1260	0.0812 0.1503	0.1084 0.1777
.01 .015 .02 .02 .025 .03	0.1615 0.2524	0.2802 0.3823	0.3923 0.5051	0.0934 0.1669	0.1493 0.2391	0.2059 0.3241
.01 .015 .02 .02 .025 .03 .025 .03 .035 .04	0.3454 0.4518	0.5754 0.6744	0.7381 0.8272	0.1448 0.2433	0.2816 0.3957	0.4198 0.5529

<표3>

표본의 크기	변수의 개수	등상관 계수	유의수준 0.05		유의수준 0.10	
			L	LR	L	LR
100	2	0.01	0.0550	0.0536	0.1000	0.1004
		0.03	0.0907	0.0914	0.1597	0.1610
		0.05	0.1658	0.1675	0.2665	0.2708
	3	0.01	0.0651	0.0609	0.1180	0.1084
		0.03	0.1775	0.1158	0.2615	0.1988
		0.05	0.3971	0.2600	0.5022	0.3642
	4	0.01	0.0813	0.0609	0.1524	0.1174
		0.03	0.3012	0.1471	0.4158	0.2391
		0.05	0.6234	0.3580	0.7365	0.4759
	5	0.01	0.1091	0.0629	0.1774	0.1132
		0.03	0.4516	0.1863	0.5658	0.2806
		0.05	0.8173	0.4575	0.8822	0.5768
200	2	0.01	0.0604	0.0605	0.1104	0.1126
		0.03	0.1269	0.1280	0.2197	0.2234
		0.05	0.2794	0.2830	0.4000	0.4036
	3	0.01	0.0769	0.0650	0.1454	0.1209
		0.03	0.3026	0.1936	0.4233	0.3057
		0.05	0.6569	0.4867	0.7644	0.6192
	4	0.01	0.1091	0.0704	0.1721	0.1249
		0.03	0.5254	0.2773	0.6291	0.3890
		0.05	0.8940	0.6625	0.9384	0.7689
	5	0.01	0.1473	0.0747	0.2355	0.1363
		0.03	0.7165	0.3589	0.8036	0.4735
		0.05	0.9782	0.8125	0.9889	0.8885
300	2	0.01	0.0677	0.0680	0.1097	0.1141
		0.03	0.1769	0.1781	0.2678	0.2753
		0.05	0.4031	0.4045	0.5365	0.5451
	3	0.01	0.0937	0.0711	0.1569	0.1312
		0.03	0.4256	0.2961	0.5525	0.4105
		0.05	0.8256	0.6783	0.8919	0.7794
	4	0.01	0.1339	0.7560	0.2120	0.1461
		0.03	0.6879	0.4073	0.7919	0.5319
		0.05	0.8892	0.6581	0.9862	0.9190
	5	0.01	0.1981	0.0869	0.3019	0.1587
		0.03	0.8711	0.5239	0.9235	0.6436
		0.05	0.9971	0.9511	0.9993	0.9734

5. 결론

다변량자료분석에 수반되는 내부적 독립성가설에 대한 검정방법으로 표본상관행렬의 역행렬에 근거한 새로운 검정통계량을 제안하고 그의 기하적 의미와 검정통계량으로서의 질을 모의실험을 통해 비교해 보았다.

모의실험의 결과를 종합하면, 유의수준을 지킨다는 점에서 적절한 수정인자를 사용한 Konishi의 결과로부터 유도한 점근분포가 검정통계량의 귀무분포로 사용가능하다는 것을 알 수 있다. 다행스러운 점은 수정인자의 선택에 그 결과가 크게 민감하지 않으며 또한 표본의 크기가 커짐에 따라 그 영향이 작아진다는 점이다.

나아가 우도비 원리에 바탕을 둔 Box의 수정된 검정통계량과의 검정력비교는 대립가설이 등상관행렬인 경우 등상관계수가 높아지고 변수의 개수가 커짐에 따라 제안된 검정통계량 L의 우월성이 두드러졌다. 또한 그 외의 다른 형태의 대립가설 경우에도 제안된 검정통계량의 검정력은 균일하게 우위를 보였다.

이와 같은 사실을 종합해 볼 때, 제안된 검정통계량은 표본상관행렬의 각 원소들이 0에 가까운 경우 전체적인 변수들의 무상관의 가설에 대한 검증에 자료분석적인 면에서의 활용이 기대된다고 하겠다.

참 고 문 헌

[1] Bargmann, R.E.(1981). A Single Union-Intersection Statistic for Testing Internal Independence. Principles and Applications, A presentation on the occasion of the Golden Jubille Celebration Conference of the Indian Statistical Institute.

[2] Bartlett, M.S.(1954). A Note on Multiplying Factors for Various Chi-squared Approximations, *Journal of the Royal Statistical Society*, Series B, 16, 296-298.

[3] Box, G.E.P.(1949). A General Distribution Theory for a class of Likelihood Criteria, *Biometrika*, 36, 317-349.

[4] Konishi, S.(1979). Asymptotic Expansions for the Distributions of Functions of a Correlation Matrix, *Journal of Multivariate Analysis*, 9, 259-266.

[5] Konishi, S.(1993). Personal letter with S. Konishi at the Institute of Statistical Mathematics(Tokyo, Japan).

[6] Lee, Kwangjin(1991). A Study on the Geometric Characteristics of Inter-correlation Structure and the Eigen Transformation of a Correlation Matrix, Unpublished Ph.D. Dissertation, Korea University.

[7] Roy. S.N. and Bargmann, R.E.(1958). Test of Multiple Independence and the Associated Confidence Bounds, *Annals of Mathematical Statistis*, 29, 491-503.

[8] Shuenemeyer, J.H. and Bargmann, R.E.(1978). Maximum Eccentricity as a Union-Intersection Test Statistic in Multivariate Analysis, *Journal of Multivariate Analysis*, 8, 268-273.

[9] Wilks, S.S.(1935), On the Independence of k Sets of Normally Distributed Statistical Variables, *Econometrika*, 3, 309-326.