응용통계연구 제 **15**권 *1*호, *2002*년, *pp. 187–199*

# 정확검정들에 대한 고찰

강승호[1)]

요 약

표본의 크기가 작아 검정 통계량의 근사분포의 정확성이 의심스러울 때, 정확검정이 종종 사용된다. 정확검정의 장점은 1종의 오류 확률이 항상 유의수준보다 작거나 같음을 보장해 준다는 것이다. 본 논문에서는 정확검정을 만드는 여러 방법, 계산 알고리듬, 그리고 상업용 소프트웨어를 살펴보겠다. 그리고 정확검정에서 얻어지는 exact p-value와 원래 우도(true likelihood)에서 얻어지는 true p-value와의 관계도 살펴보겠다.

주요용어: 조건부 검정, 분할표, 소표본, 장애모수

## 1. 서론

최근에 분할표 분석에 대한 통계 방법이 활발히 연구되어지고 있다 (Kang 1997, Kim 1998, Kim et al 2000, Oh 1996 1998, Park 1998, Yang and Huh 1999, 고봉성 1997, 김지현 and 임현선 1998, 안윤옥 1993, 허명회 1997, 홍종선 and 임한승 1997). 전통적으로 분할표에서의 통계적 가설 검정은 표본 크기가 충분히 커질 때 검정통계량의 근사 분포에 의존하여 이루어져왔다. 하지만, 그러한 방법은 표본의 크기가 작으면, 정확검정에 의한 방법과 다른 결론을 내릴 수 있다. 예를 들어 다음의 $3 \times 9$ 분할표의 경우,

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 7 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

카이제곱 검정법의 asymptotic p-value와 exact p-value는 각각 0.1342, 0.0013으로 큰 차이가 있다. 카이제곱 검정법의 exact p-value를 계산하는 방법은 3.1절에 소개되어 있다. 표본의 크기가 작으면 가끔 자료의 일부를 버리거나, 또는 자료의 일부를 임의로 합치거나 하는데, 그럴 경우 파생되는 결과는 예측하기 어렵다. 예를 들어 다음의 $2 \times 5$ 분할표

| | | | | |
|---|---|---|---|---|
| 2 | 3 | 4 | 8 | 9 |
| 0 | 0 | 11 | 10 | 11 |

의 경우, 카이제곱 검정법의 asymptotic p-value와 exact p-value는 각각 0.086, 0.072이다. 그런데 처음 두 열을 없애버리면, asymptotic p-value는 0.48이 되고, 처음 두 열을 세 째 열에 합쳐버리면, asymptotic p-value는 0.92가 된다. 그러므로, 자료의 일부를 버리거나 또

1) (120-750) 서울특별시 서대문구 대현동, 이화여자대학교 통계학과, 조교수
E-mail:seungho@mm.ewha.ac.kr

는 자료의 일부를 임의로 합치는 방법은 바람직 하지 않다. 표본의 크기가 작을 때, 근사 분포를 이용하는 검정법의 문제점은 1종의 오류 확률이 유의수준보다 훨씬 더 커질 수 있는 위험을 가지게 된다는 것이다. 반면에 정확검정의 장점은 1종의 오류 확률이 항상 유의수준보다 작거나 같음을 보장해 준다는 것이다. 통계적 가설 검정에서 대부분의 경우 1종의 오류가 2종의 오류보다 더 심각한 오류이므로, 소표본 문제가 흔히 발생하는 의학, 역학(epidemiology)등에서 정확검정이 자주 사용된다. 정확검정 중에서는 아마도 $2 \times 2$ 분할표에서의 Fisher의 정확검정이 가장 널리 알려진 방법일 것이다. Fisher의 정확검정은 $r \times c$ 분할표에도 이론적으로는 확장될수 있으나, 계산량이 너무 많아 실현되지 못하고 있었는데, Mehta and Patel(1983)의 network algorithm으로 많은 경우 계산이 가능하게 되었다. 본 논문에서는 정확검정의 종류, 성질 그리고 소프트웨어들을 살펴보겠다. 정확검정에서 얻어진 exact p-value는 "정확"이라는 단어 때문에 종종 true p-value로 받아들여지기도 한다. 또한 본 논문에서는 exact p-value와 true p-value의 관련성을 살펴보겠다.

## 2. 장애모수(nuisance parameter)가 없는 경우의 정확검정

귀무가설에서 모수의 값이 정해지는 경우에는 장애모수가 존재하지 않는다. 장애모수가 존재하지 않으면, 정확검정은 검정통계량의 분포를 이용하여 쉽게 만들 수 있다. 일표본 이항분포문제가 장애모수가 없는 경우의 대표적인 예이다. 일표본 이항분포 문제는 다음과 같이 기술된다. 확률변수 $X$가 이항분포 $B(n,p)$를 따를 때, 귀무가설은 $H_0 : p = p_0$이다.여기서 $p_0$는 알려진 값이다. 검정통계량을 T라고 표시하고, T의 값이 클수록 유의하다고 하면, $X$의 관측된 값을 $x^0$라고 할때,

$$\text{exact p-value} = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p_0^x (1-p_0)^{n-x} I_{[T(x) \geq T(x^0)]}$$

여기서 $I_{[.]}$는 조건이 만족되면 1, 아니면 0을 취하는 지수함수 (indicator function) 이다. 여기서는 장애모수가 없어서 exact p-value는 확률변수 $X$의 원래 분포(true distribution)을 사용하여 얻어졌다. 그러므로 이 경우 exact p-value와 true p-value는 일치한다. 하지만, 장애모수가 있는 경우(3절)에는, exact p-value는 true p-value와 같지 않다.

## 3. 장애모수가 있는 경우의 정확검정

### 3.1 조건부 정확검정

우리가 알고 있는 많은 정확검정은 장애모수(nuisance parameter)가 있는 경우의 정확검정이며, Fisher의 정확검정도 이 범주에 속한다. 여기서는 이표본 이항분포문제를 예로 들겠다. 확률변수 $X_k$가 $(k = 1, 2)$ 이항분포 $B(n_k, p_k)$를 따르며, $X_1$과 $X_2$는 서로 독립일 때, 귀무가설은 $H_0 : p_1 = p_2 = p, 0 < p < 1$, 대립가설은 $H_1 : p_1 \neq p_2$이다. 여기서 유의할 점은 $p$는 값이 미지인 장애모수라는 점이다. 귀무가설에서의 결합분포 $(X_1, X_2)$는 다음과 같다.

$$P(X_1 = x_1, X_2 = x_2 | H_0) = \frac{n_1!}{x_1!(n_1 - x_1)!} \frac{n_2!}{x_2!(n_2 - x_2)!} p^{x_1+x_2} (1-p)^{n_1+n_2-x_1-x_2}$$

여기서 우리는 귀무가설에서의 $(X_1, X_2)$의 분포를 알지만, $p$값을 모르므로, 그 분포를 이용하여 정확검정을 할 수는 없다. 조건부 정확검정에서는 미지인 장애모수 $p$를 제거하기 위해, 귀무가설에서 $p$의 충분통계량인 $S = X_1 + X_2$ 주어졌을 때의 $(X_1, X_2)$의 조건부 분포를 이용하여 검정하게 된다. 간단한 계산에 의하면, $S = X_1 + X_2$이 주어졌을 때의 $(X_1, X_2)$의 조건부 분포는 기하분포임이 알려져 있다.

$$P(X_1 = x_1, X_2 = x_2 | S, H_0) = \frac{n_1!n_2!S!(n-S)!}{n!x_1!x_2!(n_1-x_1)!(n_2-x_2)!}$$

충분통계량의 정의에 의하여 위의 분포는 더 이상 $p$에 의존하지 않게 된다. 검정통계량을 T라 표시하고, T의 값이 클수록 유의하다고 하자. 그러면,$(X_1, X_2)$의 관측된 값을 $(x_1^0, x_2^0)$라고 하고 $s^0 = x_1^0 + x_2^0$일 때, 조건부 정확검정에 의한 exact p-value는 다음과 같다.

$$\text{exact p-value} = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} P(X_1 = x_1, X_2 = x_2 | S = s^0, H_0) I_{[T(x_1,x_2) \geq T(x_1^0, x_2^0)]}$$

식에서 보듯이 exact p-value는 true p-value가 아니며, true p-value는 다음과 같이 주어지며, $p$ 때문에 알 수 없는 값이다.

$$\text{true p-value} = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} P(X_1 = x_1, X_2 = x_2 | H_0) I_{[T(x_1,x_2) \geq T(x_1^0, x_2^0)]}$$

우리는 위의 exact p-value정의에서, 여러 가지 검정통계량 T를 선택함으로써, 여러 가지 정확검정법을 만들어 낼 수 있다. 예를 들면,

$$T(X_1 = x_1, X_2 = x_2) = \frac{1}{P(X_1 = x_1, X_2 = x_2 | S = s^0, H_0)}$$

인 경우가 Fisher의 정확검정이고, 카이제곱 검정통계량을 사용한다면 정확 카이제곱 검정이 된다.

귀무가설에 있는 장애모수의 충분통계량에 대한 조건부 분포를 이용하여 정확검정을 하는 조건부 정확검정에 대해, 과거에는 일부 통계학자의 비판이 있기도 했으나, 현재에는 대부분의 통계학자가 동의하고 있다 (Barnard 1979, Berkson 1978, Greenland 1991). 정확검정의 장점은 1종의 오류 확률을 항상 유의수준 $\alpha$보다 작거나 같음을 보장해 준다는 것인데, 조건부 정확검정에서는 이를 다음과 같이 쉽게 증명할 수 있다.

$$P(\text{reject } H_0 | H_0) = \sum_s P(\text{reject } H_0 | S = s, H_0) P(S = s | H_0) \leq \sum_s \alpha P(S = s | H_0) = \alpha$$

## 3.2 무조건부 정확검정

무조건부 정확검정에서는 미지의 장애변수를 없애기 위해 maximization principle을 사용한다. 예를 들어 이표본 이항분포 문제의 경우, exact p-value는 다음과 같이 주어진다.

$$\text{exact p-value} = max_{0<p<1} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} P(X_1 = x_1, X_2 = x_2 | H_0) I_{[T(x_1,x_2) \geq T(x_1^0, x_2^0)]}$$

실제로 최대값을 계산하기는 어려우므로, $p$의 값을 0.001부터 0.999까지 0.001씩 증가시켜가면서 두 이항분포 확률의 곱의 합을 계산한 후, 그 중 가장 큰 것을 최대값으로 취한다. 무조건부 정확검정은 미지의 장애모수에 대한 간단한 충분통계량이 존재하지 않아, 조건부 정확검정을 사용할 수 없을 때 종종 사용된다. 무조건부 정확검정의 단점은 최대값을 갖는 $p$의 값과 실제 $p$의 참값이 멀리 떨어져 있다면, 무조건부 정확검정은 매우 보수적인 방법, 즉 주어진 유의수준 $\alpha$보다 훨씬 작은 유의수준을 사용하는 방법이 될 수도 있다. 무조건부 정확검정은 미지의 장애모수가 여러 개이면 사용하기 곤란하다는 또 하나의 단점이 있다.

### *3.3 순열검정 (permutation tests)*

순열검정은 관측된 확률변수 $X$에 대한 모수적 확률분포 가정 없이 정확검정을 만드는데 흔히 사용되는 방법이다. 본 절에서는 이표본 문제를 예를 들어 설명하겠다. 이표본 문제에서는 확률변수 $X_{ij}, (i = 1, 2 : j = 1, 2, , , n_i)$가 누적 확률분포 $F_i$를 따를 때, 관심 있는 귀무가설은 $H_0 : F_1 = F_2$이다. 이때 순열검정은 다음의 순서로 만들어진다.

1. 각 관측치에 점수(score)를 준다. 점수의 선택은 관심이 있는 대립가설에 따라 달라질 수 있다.
2. 사용할 검정 통계량을 선택한다.
3. 점수들의 모든 가능한 조합에 대해, 검정 통계량을 구하여, 순열분포 (permutation distribution)를 구한다.
4. 얻어진 순열분포를 이용하여 p-value를 구한다.

## 4. 알고리듬과 소프트웨어

본 절에서는 $r \times c$ 분할표에서 독립성 검정을 예제로 사용하면서 조건부 정확검정에서 사용되는 network algorithm에 대해서 설명하도록 하겠다. $r \times c$ 분할표에서 독립성 검정의 귀무가설은 $H_0 : p_{ij} = p_{i+}p_{+j} (i = 1, , , r, j = 1, , , c)$로 주어지며, $p_{i+} (i = 1, , , , r)$과 $p_{+j} (j = 1, , , c)$가 미지인 장애모수가 된다. 이 장애모수들에 대한 귀무가설에서의 충분통계량은 행합 $(X_{i+}, i = 1, , , r)$과 열합 $(X_{+j}, j = 1, , , c)$이 된다. 조건부 정확검정의 정의에 의하면, 일단 관측된 $r \times c$ 분할표의 행합과 열합이 같은 모든 $r \times c$ 분할표를 고려해야 한

다. 그런 분할표의 집합은 종종 reference set이라 불리우며, $(X^0_{i+}, X^0_{+j})$가 관측된 행합, 열합일때, 구체적으로는 다음과 같이 주어진다.

$$\{X|X\text{는 } r \times c\text{분할표}, \sum_{j=1}^{c} X_{ij} = X^0_{i+}, i = 1,,, r, \sum_{i=1}^{r} X_{ij} = X^0_{+j}, j = 1,,, c\}$$

reference set의 크기는 분할표의 차원이 증가하고, 총돗수가 증가함에 따라 기하급수적으로 늘어나게 되는데, Gail and Mantel (1977)은 정규근사를 이용하여 reference set의 근사크기를 추정하는 방법을 제안하였다. 표 1에서 보면 알 수 있듯이, 분할표의 크기가 조금만 커져도, reference set의 크기가 너무 커져, 아무리 빠른 컴퓨터에 의지한다고 해도, 완전조사 (complete enumeration)에 의한 exact p-value를 구하는 일은 불가능하게 된다. Mehta and Patel(1983)은 network algorithm으로 이 문제를 해결하였다. Network algorithm에

표1. Fisher의 정확검정과 근사검정

| 분할표 | exact p-value | asymp. p-value | reference set 크기 |
|---|---|---|---|
| 2 0 1 2 6 5<br>1 3 1 1 1 2<br>1 0 3 1 0 0<br>1 2 1 2 0 0 | 0.0454 | 0.0666 | $10 \times 10^6$ |
| 1 2 2 1 1 0 1<br>2 0 0 2 3 0 0<br>0 1 1 1 2 7 3<br>1 1 2 0 0 0 1<br>0 1 1 1 1 0 0 | 0.0393 | 0.1213 | $64 \times 10^9$ |

는 reference set에 있는 모든 분할표를, 한쪽 마디 (node)에서 시작하여 중간이 있는 여러마디를 거친 후, 다른 쪽 마디로 이어지는 모든 경로(path)와 일대일 대응시키고, 각 경로의 길이를 해당되는 분할표에서 얻어지는 검정통계량의 값과 일치하도록 만든다. 이 때 exact p-value를 구하는 일은 관측된 분할표에 대응하는 경로의 길이보다 작거나 같은 모든 경로를 세는 일과 같아지게 된다. Network algorithm은 각 중간 마디마다 맨 끝마디까지 가는 경로들의 길이의 최대값과 최소값을 구한다. 만약 최대값이 관측된 분할표에 해당되는 경로의 길이보다 작다면, 그 중간 마디에서 파생되는 모든 경로는 exact p-value에 포함되어야 하므로, 더 이상 각각의 경로를 조사할 필요없이, 그 중간 마디에서 파생되는 모든 경로들의 확률을 한번에 exact p-value에 넣는다. 만약 최소값이 관측된 분할표에 해당되는 경로의 길이보다 크다면, 그 중간 마디에서 파생되는 모든 경로는 exact p-value에 포함되지 않을 것이므로, 그 중간 마디에서 파생되는 모든 경로들은 더 이상 고려할 필요가 없어지게 된다. 만약 위의 조건 둘중 어느 것도 만족되지 않으면, 그 중간 마디에서 다시 하위 단계에 있는 마디들을 생성한 후, 다시 위의 조건 둘을 조사한다. Netowrk algorithm은 이처럼 implicit enumeration을 사용하므로, 계산시간을 엄청나게 줄일 수 있다. Network algorithm은 $r \times c$

분할표에서 독립성 검정 문제이외에도 여러가지 문제에 조금씩 변형되면서 적용되었고 (Agresti et al 1990, Hirji et al 1987 1988, Mehta et al 1984 1985 1988 1992 1998), 그러는 과정에서 만들어진 프로그램은 상업용 소프트웨어 StatXact (http://www.cytel.com)로 판매되고 있다. StatXact는 분할표에서의 정확검정만이 아니라, 비모수 검정법에서의 정확검정도 가능하며, StatXact version 4.0의 기능들은 표2에 요약되어 있다. 비록 network algorithm이 성공적이고, 컴퓨터가 아무리 빨라졌다해도, 분할표의 차원이 아주 커지면, 여전히 exact p-value의 계산은 불가능하다. 이러한 경우 다음의 세 가지 방향으로 연구가 진행되어져 왔다. 첫 번째 방법은 reference set에 있는 모든 분할표를 조사하는 대신, 그 일부만 조사하여 exact p-value와 신뢰구간을 구하는 방법이다. 이렇게 얻어진 p-value는 Monte Carlo p-value라고 불리운다 (Agresti et al 1979, Senchaudhuri et al 1995). 두 번째 방법은 분할표의 일부분에는 돗수가 많고, 나머지 부분에는 돗수가 적은 분할표에 사용되는 방법이다. 돗수가 많은 부분에서는 asymptotic p-value를 구하고, 돗수가 적은 부분에서는 exact p-value를 구하여 합치는 방법이다 (Baglivo et al 1988). 세 번째 방법은 기존의 대표본 이론과는 다른 가정하에서 얻어지는 대표본 이론의 결과를 이용하는 방법이다. $k$를 분할표의 칸수라고 할 때, 기존의 대표본 이론에서는 $k$는 고정되어 있고, 표본 크기 $n$만 증가한다. 하지만, 새로운 가정에서는 $n$이 증가함에 따라 $k$도 적절한 속도로 증가하게 하여 얻어지는 새로운 대표본 결과를 이용한다 (Koehler 1986).

## 5. 맺음말

정확검정은 이산형 분포를 이용함으로 인하여 유의수준 $\alpha$를 모두 다 사용하지 못하고, 실제로는 $\alpha$보다 적은 유의수준을 사용하게 된다. 때로는 실제 정확검정이 사용한 유의수준이 $\alpha$보다 훨씬 작은 경우도 생기는데, 이점 때문에 정확검정이 종종 보수적(conservative)이라고 비난 받곤 하였다 (Suissa and Shuster 1985). 하지만, Upton (1992)은 얻어진 p-value를 보고하는 것으로 충분하며, 임의의 유의수준과 비교할 필요는 없다고 주장하였다. 독립성 가정을 만족하는 자료에 대한 연구는 많이 이루어졌고, 최근에는 상관자료에서의 정확검정에 대한 연구가 발표되고 있다 (Fay and Gennings 1996, Kang and Park 2000). 그 외에도 최근의 연구 동향으로는 조건부 분포가 복잡한 경우 Markov chain Monte Carlo를 이용하는 방법(Smith et al 1996), 포아송 분포의 정규 근사에 기초한 importance sampling을 이용하는 방법이 연구되어지고 있다 (Booth and Butler 1999)

## 참고문헌

[1] 고봉성 (1997). 반복조사를 통한 범주형 자료의 오분류 탐색, *The Korean Communications in Statistics*, Vol 4, 75-90.

[2] 김지현, 임현선 (1998). 분할표를 이용한 조건부 독립성 검정, The *Korean Journal of Applied Statistics*, Vol 11, 257-268.

[3] 안윤옥 (1993). 의학 연구자료 분석과 통계적 기법, *The Korean Journal of Applied Statistics*, Vol 6, 183-190.

[4] 허명회 (1997). 2원 분할표의 소표본 검증법, *The Korean Communications in Statistics*, Vol 10, 339-352.

[5] 홍종선, 임한승 (1997). 범주형 자료에서 연관성 측도들의 비교 연구, *The Korean Communications in Statistics*, Vol 4, 645-662.

[6] Agresti, A, Mehta, C. R. and Patel N. R. (1990). Exact Inference for Contingency Tables with Ordered Categories, *Journal of the American Statistical Association*, Vol 85, 453-458.

[7] Agresti, A, Wackerly, D. and Boyett, J. (1979). Exact Conditional Tests for Cross-Classifications : Approximations of Attained Significance Level, *Psychometrika*, Vol 44, 75-83

[8] Baglivo, J., Oliver, D. and Pagano, M. (1988). Methods for the Analysis of Contingency Tables with Large and Small Cell Counts, *Journal of the American Statistical Association*, Vol 83, 1006-1013

[9] Barnard, G. A. (1979). In Contradiction to J. Berkson's Dispraise : Conditional Tests can be more Efficient. *Journal of Statistical Planning and Inference*, Vol 3, 181-188.

[10] Berkson, J. (1978). In Dispraise of the Exact Test, *Journal of Statistical Planning and Inference*, Vol 2, 27-42

[11] Booth, J. and Butler, R. (1999). Monte Carlo approximation of exact conditional tests for log-linear models, *Biometrika*, Vol 86, 321-332.

[12] Fay, M. P. and Gennings, C. (1996). Non-Parametric Two-Sample Tests for Repeated Ordinal Responses, *Statistics in Medicine*, Vol 15, 429-442.

[13] Gail, M. and Mantel, N. (1977). Counting the Number of Contingency Tables with Fixed Margins, *Journal of the American Statistical Association*, Vol 72, 859-862.

[14] Greenland, S. (1991). On the Logical Justifications of Conditional Tests for Two-by-Two Contingency Tables, *The American Statistician*, Vol 45, 248-251.

[15] Hirji, K. F., Mehta, C. R. and Patel, N. R. (1987). Computing Distributions for Exact Logistic Regression, *Journal of the American Statistical Association*, Vol 82, 1110-1117.

[16] Hirji, K. F., Mehta, C. R. and Patel, N. R. (1988). Exact Inference for Matched Case-Control Studies, *Biometrics*, Vol 44, 803-814.

[17] Kang, S. H. (1997). The Bahadur Efficiency of the Power-Divergence Statistics Condi-

tional on Margins for Testing Homogeneity with Equal Sample Size, *Journal of the Korean Statistical Society*, Vol 26, 453-465.

[18] Kang, S. H. and Park, S. (2000). Exact Likelihood Ratio Test of Independence of Binary Responses within Clusters, *Computational Statistics and Data Analysis*, Vol 33, 15-23.

[19] Kim, H. (1998). A Study on Cell Influences to Chi-square Statistic in Contingency Tables, *The Korean Communications in Statistics*, Vol 5, 35-42.

[20] Kim, D. Hong, C. S. and Oh, M. G. (2000). The Saddlepoint Approximation Methods for Statistical Inference in Contingency Tables, *The Korean Communications in Statistics*, Vol 7, 313-326.

[21] Koehler, K. J. (1986). Goodness-of-Fit Tests for Log-Linear Models in Sparse Contingency Tables, *Journal of the American Statistical Association*, Vol 81, 483-493.

[22] Mehta, C. R. and Patel, N. R. (1983). A Network Algorithm for Performing Fisher's Exact Test in r x c Contingency Tables, *Journal of the American Statistical Association*, Vol 78, 427-434.

[23] Mehta, C. R. Patel, N. R. and Gray, R. (1985). Computing an Exact Confidence Interva for the Common Odds Ratio in Several 2 x 2 Contingency Tables, *Journal of the American Statistical Association*, Vol 80, 969-973.

[24] Mehta, C. R. Patel, N. R. and Senchaudhuri, P. (1992). Exact Stratified Linear Rank Tests for Ordered Categorical and Binary Data, *Journal of Computational and Graphical Statistics*, Vol 1, 21-40.

[25] Mehta, C. R. Patel, N. R. and Senchaudhuri, P. (1998). Exact Power and Sample-Size Computations for the Cochran-Armitage Trend Tests, *Biometrics*, Vol 54, 1615-1621.

[26] Mehta, C. R. Patel, N. R. and Tsiatis, A. (1984). Exact Significance Testing to Establish Treatment Equivalence with Ordered Categorical Data, *Biometrics*, Vol 40, 819-825.

[27] Mehta, C. R. Patel, N. R. and Wei, L. J. (1988). Computing Exact Significance Tests with Restricted Randomization Rules, *Biometrika*, Vol 75, 295-302.

[28] Oh, M. (1996). Inference for Order Restrictions on Odds in Contingency Tables, *Journal of the Korean Statistical Society*, Vol 25, 381-392.

[29] Oh, M. (1998). Tests For and Against a Positive Dependence Restrictions in Two-Way Ordered Contingency Tables, *Journal of the Korean Statistical Society*, Vol 27, 205-220.

[30] Park, C. (1998). The Chi-squared Test of Independence for a Multi-way Contingency Table with All Margins Fixed, *Journal of the Korean Statistical Society*, Vol 27, 197-203.

[31] Senchaudhuri, P., Mehta, C. R. and Patel, N. (1995). Estimating Exact P-values by the Method of Control Variates or Monte Carlo Rescue, *Journal of the American Statistical Association*, Vol 90, 640-648.

[32] Smith, P. W. F, Forster, J. J and McDonald, J. W. (1996). Monte Carlo exact tests for square contingency tables, *Journal of the Royal Statistical Society, Series A*, Vol 159, 309-321.

[33] Suissa, S. and Shuster, J. J. (1985). Exact Unconditional Sample Size for the 2 x 2 Binomial Trials, *Journal of the Royal Statistical Society Series A*, Vol 148, 317-327.

[34] Upton, G. J. G. (1982). A Comparison of Alternative Tests for the 2 x 2 Comparative Trial, *Journal of the Royal Statistical Society Series A*, Vol 145, 86-105.

[35] Yang, K. S. and Huh, M. H. (1999). Correspondence Analysis of Two-way Contingency Tables with Ordered Column Categories, *Journal of the Korean Statistical Society*, Vol 28, 347-358.

[ 2001년 10월 접수, 2002년 1월 채택 ]

표2. StatXact이 처리할 수 있는 정확검정(연속형 자료)

| Inference for Continuous Data | |
|---|---|
| One-Sample Goodness-of-Fit Inference | Chi-Square Goodness-of-Fit Test<br>Kolmogorov Goodness-of-Fit Test<br>Shapiro-Wilk Test, Lilliefors Test<br>Runs Test |
| Two-Sample Inference: Paired Samples | Wilcoxon Signed-Rank Test, Sign Test<br>Hodges-Lehmann Estimation<br>Permutation Test, McNemar's Test<br>Marginal Homogeneity Test |
| Two-Sample Inference:<br>Independent Samples | Wilcoxon-Mann-Whitney Test<br>Hodges-Lehmann Estimation<br>Generalized Wilcoxon-Gehan Test<br>Normal Scores Test, Savage Score Test<br>Logrank Test, Siegel-Tukey Test<br>Ansari-Bradley Test, Klotz Test<br>Mood Test, Conover Test<br>Permutation Test with General Scores<br>Kolmogorov-Smirnov Test<br>Wald-Wolfowitz Runs Test |
| K-Sample Inference: Repeated Samples | Friedman Test, Kendall's W<br>Cochran Q's Test, Quade Test, Page Test |
| K-Sample Inference: Independent Samples | Median Test, Kruskall-Wallis Test<br>Normal Scores Test, Savage Scores Test<br>One-Way ANOVA with General Scores<br>K-Sample Logrank Test<br>K-Sample Wilcoxon-Gehan Test<br>Jonckheere-Terpstra Test<br>Linear by Linear Association Test<br>K-Sample Trend Test with Censoring |

표3. StatXact이 처리할 수 있는 정확검정(범주형 자료)

| Inference for Categorical Data | |
|---|---|
| One-Sample | Binomial Test, Multinomial Estimates<br>Poisson Test |
| Two Binomial Samples | Fisher's Exact Test<br>Confidence Interval on the Difference<br>Confidence Interval on the Ratio<br>Unconditional Exact Test |
| Two Stratified Poisson Samples | Test of Homogeneity of Relative Risks<br>Test of the Coomon Relative Risk |
| Stratified $2 \times 2$ Contingency Tables | Test of Homogeneity of Odds-Ratio<br>Mantel-Haenszel Test |
| Stratified $2 \times C$ Contingency Tables | Wilcoxon Rank Sum Test<br>Normal Scores Test, Savage Test<br>Trend Test, Permutation Test |
| Unordered $R \times C$ Contingency Table | Pearson's Chi-Square Test<br>Likelihood Ratio Test<br>Fisher's Exact Test |
| Singly Ordered $R \times C$ Contingency Tables | Kruskall-Wallis Test<br>Normal Scores Test<br>One-Way ANOVA with General Scores |
| Doubly Ordered $R \times C$ Contingency Tables | Jonckheere-Terpstra Test<br>Linear-by-Linear Association Test |
| Exact Power | Comparing Two Binomial<br>Comparing K Ordered Binomial<br>Comparing Two Ordered Multinomial |

표4. StatXact이 처리할 수 있는 정확검정(연관성 측도)

| Inference for Measure of Association | |
|---|---|
| Ordinal Data | Pearson's Correlation<br>Spearman's Correlation<br>Kendall's Coefficient of Concordance<br>Kendall's Tau, Somer's D Coefficient<br>Gamma Coefficient |
| Nominal Data | Goodman-Kruskal Tau<br>Uncertainty Coefficient |
| Measure of Agreement | Cohen's Kappa, Weighted Kappa |

# Investigation on Exact Tests

Seung-Ho Kang[1)]

ABSTRACT

When the sample size is small, exact tests are often employed because the asymptotic distribution of the test statistic is in doubt. The advantage of exact tests is that it is guaranteed to bound the type I error probability to the nominal level. In this paper we review the methods of constructing exact tests, the algorithm and commercial software. We also examine the difference between exact p-values obtained from exact tests and true p-values obtained from the true underlying distribution.

*Keywords* : conditional test, contingency table, small sample, nuisance parameter

1) Assistant Professor, Department of Statistics, Ewha Womans University
E-mail:seungho@mm.ewha.ac.kr