Journal of the Korean Data & Information Science Society
2012, **23**(4), 757–764

http://dx.doi.org/10.7465/jkdi.2012.23.4.757
한국데이터정보과학회지

복합표본자료에서 동질성검정을 위한 피어슨 검정통계량의 효과

허순영[1] · 정영애[2]

[1]창원대학교 통계학과 · [2]창원대학교 유아교육과

접수 2012년 6월 25일, 수정 2012년 7월 13일, 게재확정 2012년 7월 19일

요약

복합표본설계에 기초한 범주형 조사자료는 통상적인 피어슨 카이제곱검정에 필요한 조건을 만족하지 못한다. 그러나 많은 조사연구에서 복잡한 표본설계 방법을 적용하고 있지만, 종래의 피어슨 검정결과를 제시하고 있다. 본 연구는 복합표본설계에 의한 범주형자료의 동질성검정에 대한 실증분석을 통해, 종래의 피어슨 검정과 불편검정인 왈드검정, 표본설계를 반영한 비율추정치를 사용하는 피어슨 검정을 비교하였다. 분석결과, 종래의 피어슨검정은 표본설계를 반영하는 검정들에 비해 통계량 값이 매우 크고, 유의확률이 심각하게 작게 나타나는 것을 확인하였다. 복합표본설계를 반영하되 추정량의 분산을 아는 경우와 모르는 경우의 비교에서는 범주수, 설계효과행렬의 고유치들의 평균과 표준편차에 영향을 받는 것을 확인하였다.

주요용어: 동질성검정, 복합표본설계, 설계효과, 왈드 검정, 피어슨 검정.

1. 서론

2개 이상의 범주를 갖고 있는 범주형자료의 적합도검정, 동질성검정, 독립성검정에서는 관찰치들이 서로 독립이라는 가정 아래 피어슨 카이제곱검정을 실시한다. 그러나 복합표본설계에 의한 조사자료는 일반적 피어슨 카이제곱검정이 기초한 가정을 만족하지 못한다. 복합표본설계 (complex sample design)는 층화 (stratification), 집락 (clustering), 다단계 (multi-stage) 또는 다상 (multi-phase), 불균등확률 (unequal probability), 다중틀 (multi-frame) 등을 복합적으로 사용하여 표본을 추출한다 (Lavrakas, 2008). 오늘날 많은 조사연구는 이러한 복합표본설계에 기초하여 수행되고 있다 (예로, Kim 등, 2009; Kim 등, 2010). 이러한 복합표본설계에 의해 조사된 범주형자료는 종래의 피어슨 카이제곱검정을 실시할 경우 잘못된 검정결과를 얻을 수 있다.

Holt 등 (1980)은 피어슨 카이제곱검정에 미치는 표본설계의 효과를 분석하였는 데, 독립성검정에서보다 적합도검정과 동질성검정에서 그 효과가 훨씬 심각하다는 실증분석결과를 제시하였다. Rao와 Scott (1981, 1984, 1987)과 Tomas와 Rao (1987)는 복합표본설계에 의해 조사된 2차자료의 분석시에 표본설계를 고려한 확률추정치를 사용한 피어슨 검정통계량의 편의성을 보완하는 검정통계량을 제시하였다.

두 개 이상의 검정방법들을 비교하는 방법 중의 하나가 그것들의 검정력을 비교하는 것이고, 최근에 검정력을 계산하여 비교하려는 시도들이 늘고 있다 (Lee 등, 2012). Heo (2006)는 동질성검정에서

[1] 교신저자: (641-773) 경남 창원시 의창구 퇴촌로92, 창원대학교 통계학과, 부교수.
E-mail: syheo@changwon.ac.kr
[2] (641-773) 경남 창원시 의창구 퇴촌로92, 창원대학교 유아교육과, 교수.

Rao-Scott이 제시한 조정된 피어슨 검정과 불편검정인 왈드검정 (Wald test)의 검정력을 비교하는 실증분석을 통해 Rao-Scott의 조정된 피어슨검정은 제1종오류가 증가하고 검정력곡선이 상향하는 경향을 띠고 있음을 보여준다.

현재 국내에서는 많은 조사연구가 실시되고 있고, 그 중 많은 경우 복잡한 표본설계 방법을 적용하고 있어 일반적 피어슨 카이제곱검정이 기초한 가정을 만족하지 않는다. 그럼에도 불구하고 조사결과를 제시할 때, 단순히 피어슨 카이제곱검정결과를 제시하는 경우를 종종 볼 수 있다. Holt 등 (1980)은 두 모집단 동질성검정에서, 시뮬레이션을 통해 왈드검정과 비율에 대한 불편추정량을 사용한 피어슨 검정을 비교하였고, 표본추출법을 고려한 비율에 대한 불편추정량을 사용한 피어슨 검정은 표본추출방법에 따라 차이는 있으나 대체로 설정된 유의수준보다 훨씬 큰 값을 갖는 것을 보여주었다. Holt 등이 비교에 사용한 피어슨 검정통계량은 비율에 대한 불편추정량을 사용하는 것으로 표본설계를 반영한 추정량이나, 국내에서는 다항분포를 가정한 단순비율추정량을 사용하는 경우가 여전히 종종 있다.

본 연구는 복합표본설계에 의해 조사된 실증자료분석을 통해, 범주형자료의 동질성검정에서 종래의 피어슨 카이제곱검정, 불편검정인 왈드검정, 그리고 표본설계를 반영한 비율에 대한 불편추정량을 사용한 피어슨 검정에 대해 각각 검정통계량과 유의확률을 계산하여 실증비교하였다. 또 왈드검정과 불편추정량을 사용한 피어슨 검정의 설계효과분석을 통해, Holt 등 (1980)의 시뮬레이션에서 다루어지지 않은 복합표본설계를 적용한 조사자료에서 설계효과행렬의 고유값들의 크기가 미치는 효과를 실증분석하였다. 본 연구의 목적은 이러한 실증분석을 통해 종래의 피어슨 카이제곱검정을 위한 가정을 만족하지 못하는 조사자료분석에서 표본설계가 반영된 분석의 필요성과 종래의 피어슨 카이제곱검정의 사용시에 주의의 필요성을 환기시키고자 한다. 실증분석을 위해 2009년 경상남도 시·군 교육청 고객만족도 조사자료를 사용하였다.

본 논문의 2절에서는 두 모집단 동질성 검정을 위한 기존의 세 가지 검정통계량과 연구결과들을 요약하였고, 3절에서는 복합표본설계에 의해 조사된 실증자료분석을 통해 세 가지 검정통계량의 검정결과와 설계효과행렬에 따른 검정결과의 차이들을 비교분석하였고, 4절에는 결론을 제시하였다.

2. 동질성검정

두 모집단으로부터 크기가 n_1과 n_2인 독립표본을 추출한 후, 상호배반인 K개의 범주로 구성된 범주형 변수에 대해 표본원소들이 각 범주에 속할 확률 $(p_{ij}, i = 1, 2;\ j =, 2, \cdots, K)$이 두 모집단에서 서로 동일한지를 검정하는 동성질검정에서 귀무가설은

$$H_0 : p_1 = p_2\ (= p)$$

이다. 여기서, $p_i = (p_{i1}, p_{i2}, \cdots, p_{i,K-1})^T$이고 $p = (p_1, p_2, \cdots, p_{K-1})^T$다. 이 때, 주어진 표본추출방법 하에서 검정통계량은

$$X^2 = (\widehat{p}_1 - \widehat{p}_2)^T \widehat{V}^{-1} (\widehat{p}_1 - \widehat{p}_2) \tag{2.1}$$

의 형태를 갖는다. 여기서, $\widehat{V} = \widehat{V}ar(\widehat{p}_1 - \widehat{p}_2)$으로 $Var(\widehat{p}_1 - \widehat{p}_2)$의 점추정량이다.

각 모집단으로부터 단순임의복원추출법에 의해 표본이 추출된 경우, p_{ij}는 $\widehat{p}_{ij} = n_{ij}/n_i$로 추정되고, 귀무가설 하에서 j번째 범주에 속할 확률 (p_j)에 대한 추정량은 $\widehat{p}_j = (n_{1j} + n_{2j})/(n_1 + n_2)$로

$$\widehat{Var}(\widehat{p}_1 - \widehat{p}_2) = \frac{n_1 + n_2}{n_1 n_2} \widehat{P} \tag{2.2}$$

이 된다. 여기서, $\widehat{P} = \mathrm{diag}(\widehat{p}) - \widehat{p}\widehat{p}^T$ 이고, $\widehat{p} = (\widehat{p}_1, \widehat{p}_2, \cdots, \widehat{p}_{K-1})^T$ 이다. 귀무가설이 참이고 n_1과 n_2가 충분히 클 때, (2.1)식의 검정통계량 (X^2_{srs}라 하자)은 근사적으로 자유도가 $K-1$인 카이제곱분포 χ^2_{K-1}를 따른다.

한편, 각 모집단으로부터 층화, 집락, 불균등 추출방법 등을 복합적으로 사용해 표본을 추출한 복합표본의 경우, X^2_{srs}의 근사분포는 더 이상 χ^2_{K-1}가 아니다. 모집단이 L개의 층으로 층화되고 각 층으로부터 n_{ih} (≥ 2)개의 집락이 확률비례에 의해 추출되고, 각 집락으로부터 최종단위들 ($n_{ihl} \geq 1$)이 표본으로 추출되었다고 하자. 이 때, p_{ij}는

$$\widehat{p}_{ij} = \widehat{A}_{ij}/\widehat{N}_i \tag{2.3}$$

에 의해 추정되고, 여기서 $\widehat{A}_{ij} = \sum_{t \in s_i} w_{it}\delta_{jt}$, $\widehat{N}_i = \sum_j \widehat{A}_{ij}$이다. 이 때, s_i는 i번째 표본에 속하는 최종단위들의 집합이고, i번째 표본의 t번째 최종단위가 j번째 범주에 속하면 $\delta_{jt} = 1$이고 아니면 $\delta_{jt} = 0$이다. w_{it}는 i번째 표본에서 t번째 최종단위의 표본가중치이다. 여기서는 표기의 간소화를 위해 첨자 t를 층, 집락, 최종단위를 망라하는 것으로 사용하였다. 식 (2.3)의 $\widehat{p}_{ij}$ 들로 구성된 $\widehat{p}_i = (\widehat{p}_{i1}, \cdots, \widehat{p}_{i,K-1})^T$가 p_i의 일치추정량이고, $n_i^{-1/2}(\widehat{p}_i - p_i)$의 분포가 근사적으로 정규분포 $N_{K-1}(0, V_i)$를 따른다면, 식 (2.1)의

$$\widehat{V} = \widehat{Var}(\widehat{p}_1 - \widehat{p}_2) = \frac{\widehat{V}_1}{n_1} + \frac{\widehat{V}_2}{n_2} \tag{2.4}$$

이 된다. $\widehat{V}_i$이 V_i의 일치통계량일 때, 식 (2.4)를 식(2.1)에 대입한 통계량은 왈드검정통계량으로 표본크기가 n_i가 클 때, 귀무가설 하에서 자유도가 $K-1$이 카이제곱분포 χ^2_{K-1}을 따른다 (Rao와 Scott, 1981; Shao, 1996). 3절에서는 이 검정통계량을 X^2_w으로 표기하였다.

복합표본의 경우, X^2_{srs}의 분포는 근사적으로

$$X^2_{srs} \approx \sum_{j=1}^{K-1} d_j Z_j^2 \tag{2.5}$$

이다. 여기서, $Z_j \sim N(0,1)$이고 d_j는

$$D = \frac{n_2 D_1 + n_1 D_2}{n_1 + n_2} \tag{2.6}$$

의 고유치들로 $D_i = P^{-1}V_i$는 i번째 모집단에 대한 설계효과행렬 (design effect matrix)이다 (Holt 등, 1980). $D_i = P^{-1}V_i = I$이면 모든 j에 대해 $d_j = 1$이 되어 χ^2_{K-1}이지만, $P \neq V_i$ $(i = 1, 2)$이므로 X^2_{srs}의 근사분포는 χ^2_{K-1}를 따르지 못하고 실제 유의수준은 설정된 유의수준보다 대체로 더 큰 값을 갖는다. 식 (2.5)의 $\sum_j d_j Z_j^2$에서 Z_j^2은 자유도가 1인 카이제곱 확률변수이므로 $\sum_j d_j Z_j^2$의 평균은 $\sum_j d_j$이고 분산은 $2\sum_j d_j^2$이다. 따라서 X^2_P의 검정력은 d_j들의 크기와 분산에 영향을 받게 된다 (Rao와 Scott, 1981, 1984, 1987; Tomas와 Rao, 1987).

조사자료의 2차자료 분석시에 추정량의 분산을 종종 알지 못한다. 이 때, 식 (2.3)의 $\widehat{p}_{ij}$를 식 (2.1)과 (2.2)에 적용해 검정통계량 (이라 함)을 계산한 후, X^2_P의 근사분포가 자유도 $K-1$인 제이제곱분포를 따른다고 가정하여 검정하기도 한다. 그 경우, 귀무가설이 참일 때 j번째 범주에 속할 확률은 $\widehat{p}_j = (n_{1j}\widehat{p}_{1j} + n_{2j}\widehat{p}_{2j})/(n_1 + n_2)$로 추정된다. 이 X^2_P의 근사분포도 식 (2.5)로 표현할 수 있다. 그 경우, P의 추정량은 식 (2.3)을 사용하여 구하고, 이 경우도 실제 유의수준은 설정된 유의수준보다 큰 값을 갖는다 (Holt 등, 1980; Heo, 2006). Heo (2006)는 두 검정통계량 (X^2_P와 X^2_w)의 검정력곡선의 비교를 통해 검정통계량 X_p의 검정력곡선은 왈드검정통계량 X^2_P의 검정력곡선에 비해 전반적으로 상향하는 경향이 있음을 보여준다.

3. 실증분석

3.1. 자료

경상남도 도교육청은 2009년 산하 시·군교육청에 대한 학생, 학부모의 교육만족도와 교사의 업무만족도를 파악하기 위해 표본조사를 실시하였다. 표본추출을 위해, 20개 시·군과 학교급 (초등학교와 중학교)으로 층화한 후, 각 층 내에서 학교별 학급수에 비례하여 표본학교를 선정하였다. 이 후, 표본학교에서 조사대상 학년별로 1개 학급을 선정하였고, 표본학급 내에서 최종조사대상 학생을 할당된 수만큼 조사하였는데, 이 때 학교별 1개 학급선정과 최종조사대상의 선정은 지침을 작성하여 각 표본학교에서 지침에 따라 무작위로 선정하도록 하였다. 학부모는 표본으로 선정된 학생의 학부모를 조사대상으로 하였고, 교사는 표본으로 선정된 학교당 10명 내외를 조사대상으로 하였다. 표본가중치는, 표본설계 시에 결정된 기본가중치에 단위 무응답을 조정한 후, 사후층화에 의해 재조정하였다. 만족도 문항의 경우 무응답은 각 표본학교별 응답자의 평균값으로 결측치를 보정하였다. 표본설계에 대한 자세한 내용은 Chung 등 (2009)와 Heo와 Chang (2010)을 참고할 수 있다.

학생, 학부모 교육만족도와 교사 업무만족도는 각각 7개와 6개의 지표영역에 대해 조사하였고, 각 영역별로 다수의 질문들 5점 척도로 측정하였다. Table 3.1은 학부모, 교사용 설문지에서 각 영역별로 조사한 문항수와, 동질성검정을 위해 범주화한 후의 영역별 응답범주수이다. 영역별 범주화를 위해 각 영역별 문항들에 대한 개별 응답자들의 응답평균을 계산한 후, 이 평균점수를 1점 간격으로 범주화하여 5개 범주를 구성하였다. 범주화한 후, 각 범주에 포함된 응답자수가 작아서 일반적 피어슨 카이제곱검정의 대표본근사조사건을 만족시키지 못하는 범주는 인접범주에 묶어서 재범주화하였다. Table 3.1의 범주수는 학부모와 교사에 대해 영역별 재범주화한 후의 범주수이다.

Table 3.1 The numbers of items and response categories by satisfaction domains and survey subjects

No.	Satisfaction domain	Parent		Teacher	
		No. of items	No. of categories	No. of items	No. of categories
1	School operation	8	4	11	3
2	Diversity of school education	10	4	10	3
3	Advancement of curriculum	8	4	10	4
4	Assistance for health and further continuing education	4	4	4	4
5	Building infrastructure for improving teacher's ability	3	4	3	4
6	Region's specialized business	2	5	3	4
7	Ability obtained from school education	5	5	-	-
	Total	40		41	

3.2. 분석결과

동질성검정을 위한 모집단은 10개 시지역과 10개 군지역으로 하였다. 여기서는 학부모와 교사 자료만을 분석대상으로 하였다. 학부모 응답자는 총 3,456명으로 시지역 응답자가 2,453명이고 군지역 응답자는 1,003명이다. 교사응답자는 총 1,347명이고 시지역 응답자가 834명, 군지역 응답자는 513명이다. Table 3.2 3.5는 각 영역별로 시지역과 군지역 간에 만족도분포에 차이가 있는지를 검정하는 동질성검정의 결과이다. Table 3.2와 Table 3.3은 학부모와 교사에 대해, 단순임의복원추출을 가정한 검정 (검정통계량 X^2_{srs}), T표본설계를 반영한 왈드검정 (검정통계량 X^2_w), 표본설계를 반영하였지만

분산을 모르고 2차자료를 사용하는 검정 (검정통계량 X_P^2)의 검정결과를 보여준다. Table 3.2와 Table 3.3의 p-value에서 <.0001은 유의확률이 0.001보다 작은 값임을 나타낸다. 두 표 모두 통계량의 값은 $X_w^2 < X_P^2 < X_{srs}^2$순서이고 단순임의추출을 가정하는 X_{srs}^2은 모든 영역에서 매우 큰 값을 갖는다. 그 결과 X_{srs}^2은 모든 영역에 대해 통계적으로 매우 크게 유의한 차이가 있다. 그러나 Table 3.2의 학부모의 경우에서 X_w^2은 상대적으로 유의확률이 크게 증가하였고, 영역 5의 경우는 통계적으로 유의한 차이가 없다. Table 3.3의 교사 경우는 두 통계량간의 검정결과 차이가 더 커서, X_w^2는 6개 영역 중 4개 영역이 유의수준 5%에서 유의한 차이가 없고, 나머지 두 영역도 유의수준 1%에서 유의한 차이가 없다.

Table 3.2 The test statistics and p-values for homogeneous test between city and county (parent)

Domain No.	No. of Cat.	X_{srs}^2		X_w^2		X_P^2	
		Test statistic	p-value	Test statistic	p-value	Test statistic	p-value
1	4	57.4359	<.0001	13.5290	0.0090	24.9034	<.0001
2	4	102.5195	<.0001	21.7877	0.0002	46.7309	<.0001
3	4	64.0638	<.0001	18.1679	0.0011	27.0988	<.0001
4	4	79.5358	<.0001	18.4772	0.0010	39.0144	<.0001
5	4	14.2800	0.0025	4.4075	0.3537	4.6000	0.3309
6	5	79.3581	<.0001	19.6276	0.0015	35.0680	<.0001
7	5	48.7765	<.0001	11.2001	0.0476	20.9270	0.0008

Table 3.3 The test statistics and p-values for homogeneous test between city and county (teacher)

Domain No.	No. of Cat.	X_{srs}^2		X_w^2		X_P^2	
		Test statistic	p-value	Test statistic	p-value	Test statistic	p-value
1	3	34.1850	<.0001	7.8260	0.0497	38.6285	<.0001
2	3	34.3671	<.0001	7.6931	0.0528	37.0577	<.0001
3	4	21.5058	<.0001	6.9697	0.1375	23.1665	0.0001
4	4	37.1121	<.0001	11.4525	0.0219	37.7469	<.0001
5	4	28.9220	<.0001	7.2105	0.1252	31.4921	<.0001
6	4	45.9348	<.0001	8.0737	0.1522	45.0003	<.0001

Table 3.4와 Table 3.5는 식 (2.6)에서 주어진 X_P^2에 대한 X_w^2의 설계효과 행렬의 고유치들을 계산하여 평균과 표준편차를 구한 결과이다. Table 3.4와 Table 3.5의 마지막 열은 검정통계량값 X_P^2을 X_w^2 값으로 나눈 백분비이다. Table 3.4와 Table 3.5에서, 범주수가 큰 경우 (예를 들면, 4)는 상대적으로 범주수가 작은 경우 (예를 들면, 3)에 비해 고유치들의 평균과 표준편차가 작더라도 검정통계값의 상대크기가 더 큰 것을 알 수 있다. 또한 동일한 범주수에서 고유치들의 평균은 작아도 표준편차가 큰 경우, 증가분이 더 크다. 학부모의 경우, 영역 5와 6은 고유치들의 평균이 1보다 작고 표준편차는 1보다 약간 크고, 상대크기도 다른 경우에 비해 작다. 학부모의 영역 5는 학부모와 교사를 망라하여 고유치들의 평균과 표준편차가 가장 작고 두 검정통계량 값들 간의 차이도 무시할 만하다.

Table 3.4 Mean and standard deviation of eigenvalues of design effect matrix for comparing X_P^2 and X_w^2 for homogeneous test between city and county by satisfaction domains (parent)

Domain No.	No. of Cat.	Design effect		Test statistics		Relative size
		$\bar{d}$	s_d	X_w^2	X_P^2	
1	4	1.0802	1.6287	13.5290	24.9034	184.1%
2	4	1.3631	2.6198	21.7877	46.7309	214.5%
3	4	1.3916	2.6102	18.1679	27.0988	149.2%
4	4	1.0976	1.6985	18.4772	39.0144	211.1%
5	4	0.9280	1.2473	4.4075	4.6000	104.4%
6	5	0.9470	1.3067	19.6276	35.0680	178.7%
7	5	1.0011	1.4665	11.2001	20.9270	186.8%

Table 3.5 Mean and standard deviation of eigenvalues of design effect matrix for comparing X_P^2 and X_w^2 for homogeneous test between city and county by satisfaction domains (teacher)

Domain No.	No. of Cat.	Design effect		Test statistics		Relative size
		$\bar{d}$	s_d	X_w^2	X_P^2	
1	3	4.2162	3.8202	7.8260	38.6285	493.6%
2	3	4.3191	3.6673	7.6931	37.0577	481.7%
3	4	3.9944	1.9768	6.9697	23.1665	332.4%
4	4	2.6727	2.1411	11.4525	37.7469	329.6%
5	4	3.7719	2.6442	7.2105	31.4921	436.8%
6	4	3.5520	2.7177	8.0737	45.0003	557.4%

4. 결론

복합표본설계에 기초한 조사자료는, 범주형자료 분석에서 통상적으로 사용하는 피어슨 카이제곱검정이 기초한 가정을 만족하지 않는다. 그러나 국내에서 실시되는 많은 조사연구에서, 표본설계에서는 복잡한 표본추출방법을 적용하고 있지만, 자료분석과정에서는 종래의 피어슨 카이제곱검정결과를 그대로 제시하는 경우가 많이 있다. 본 연구는 복합표본설계에 의해 조사된 자료의 실증적 분석을 통해, 두 모집단의 동질성검정에서 종래의 피어슨 카이제곱검정과 복합표본설계를 고려한 왈드검정, 표본설계를 반영한 비율추정량을 사용한 피어슨 검정을 비교 분석하였다.

실증분석을 통해서 단순임의복원추출을 가정하는 종래의 피어슨 카이제곱검정은 표본설계를 반영한 왈드검정에 비해 통계량이 매우 크게 나타나고, 그 결과 유의확률은 심각하게 작은 것을 확인하였다. 복합표본설계를 반영하되 추정량의 분산을 아는 경우와 모르는 경우의 비교에서는, 첫째, 설계효과행렬의 고유치들의 평균과 표준편차가 더 작아도 범주수가 더 크면, 분산을 모르는 경우의 통계량 값이 더 크고, 둘째, 동일한 범주수의 경우 고유치들의 평균이 작더라도 표준편차가 크면, 분산을 모르는 경우의 통계량 값이 더 크고, 셋째, 평균이 1보다 작고 표준편차가 1에 가까우면 범주수가 크더라도 분산을 아는 경우와 모르는 경우의 통계량 값이 차이가 작은 것도 함께 확인하였다.

본 연구의 실증분석은 2009년 경남교육청 고객만족도조사자료 분석을 통해 이루어진 것으로, 복합표본조사의 범주형자료분석에서 기존의 피어슨 카이제곱검정 사용은 주의가 필요함을 경험적으로 확인하였다. 본 연구의 실증분석결과는 특정표본설계에 기초하여 두 개 모집단과 일부 변수들에 대해서 얻은 제한적인 결과이고, 복합표본설계에 의한 조사자료분석에서 분산을 모르는 경우 비율에 대한 불편추정량을 사용한 피어슨검정을 사용해도 되는 지에 대한 결정은 각 조사연구의 설계효과를 고려한 판단이 필요하다.

참고문헌

Chung, Y., Jung, D. and Heo, S. (2009). *2009 Customer satisfaction evaluation survey to the service from Gyeongsangnam-do regional offices of education*, Changwon National Univeristy, Gyeongnam.

Heo. S. and Chang. D. (2010). A sample survey design for service satisfaction evaluation of regional education offices. *Journal of the Korean Data & Information Science Society*, **21**, 671-678.

Holt, D., Scott, A. J. and Ewings, P. D. (1980). Chi-squared tests with survey data. *Journal of the Royal Statistical Society A*, **143**, 302-320.

Heo, S. (2006). Power analysis of the Rao-Scott first-order adjustment to the Pearson test for homogeneity. *Proceedings of Joint Statistical Meeting*, Seattle, U.S.A., 3126-3129.

Kim, D. H., Cho, K. H., Hwang, J. S. and Jung, K. H. (2009). A sample design for life and attitude survey of Gyeongbuk people. *Journal of the Korean Data & Information Science Society*, **20**, 1165-1167.

Kim, D. H., Hwang, J. S. and Kwak, S. G. (2010). A sample design for the survey on actual state of SMEs. *Journal of the Korean Data & Information Science Society*, **21**, 1021-1029.

Lee, C., Kang, H. and Sim, S. (2012). An implementation of the sample size and the power for testing mean and proportion. *Journal of the Korean Data & Information Science Society*, **23**, 53-61.

Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*, Vol.2, SAGE Publication, Inc., London.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit the independence in two-way tables. *Journal of the American Statistical Association*, **76**, 221-230.

Rao, J. N. K. and Scott, A. J. (1984). On chi-squared test for multiway contingency tables with cell proportions estimated from survey data. *The Annals of Statistics*, **12**, 46-60.

Rao, J. N. K. and Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. *The Annals of Statistics*, **15**, 385-397.

Shao, J. (1996). Resampling methods in sample surveys (with discussion). *Statistics*, **27,** 203-254.

Journal of the Korean Data & Information Science Society
2012, **23**(4), 757–764

http://dx.doi.org/10.7465/jkdi.2012.23.4.757
한국데이터정보과학회지

Effect of complex sample design on Pearson test statistic for homogeneity

Sunyeong Heo[1] · Young-Ae Chung[2]

[1]Department of Statistics, Changwon National University
[2] Department of Childhood Education, Changwon National University

Received 25 June 2012, revised 13 July 2012, accepted 19 July 2012

Abstract

This research is for comparison of test statistics for homogeneity when the data is collected based on complex sample design. The survey data based on complex sample design does not satisfy the condition of independency which is required for the standard Pearson multinomial-based chi-squared test. Today, lots of data sets ara collected by complex sample designs, but the tests for categorical data are conducted using the standard Pearson chi-squared test. In this study, we compared the performance of three test statistics for homogeneity between two populations using data from the 2009 customer satisfaction evaluation survey to the service from Gyeongsangnam-do regional offices of education: the standard Pearson test, the unbiased Wald test, and the Pearson-type test with survey-based point estimates. Through empirical analyses, we fist showed that the standard Pearson test inflates the values of test statistics very much and the results are not reliable. Second, in the comparison of Wald test and Pearson-type test, we find that the test results are affected by the number of categories, the mean and standard deviation of the eigenvalues of design matrix.

Keywords: Complex sample design, design effect, homogeneity test. Pearson test, Wald test.

[1] Corresponding author: Associate professor, Department of Statistics, Changwon National University, Changwon 641-773, Korea. E-mail: syheo@changwon.ac.kr
[2] Professor, Department of Childhood Education, Changwon National University, Changwon 641-773, Korea.