Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.4 pp. 877~887

A Combined Procedure of Direct Question Method and Modified Randomized Response Technique for Estimating Population Proportion[1)]

Hyuk Joo Kim[2)]

Abstract

A two-stage procedure is proposed to estimate the population proportion of a sensitive group. The proposed procedure is obtained by combining the direct question method and a modified randomized response technique. It is verified that the proposed procedure is more efficient than existing methods under some mild conditions.

Keywords : Direct question method, Population proportion, Privacy protection, Randomized response technique.

1. 서 론

사회조사에서 응답자에게 주어지는 질문이 민감한 사안에 관한 것일 때 응답자는 대답을 하지 않거나 고의로 거짓대답을 하는 경우가 많다. 이때 발생하는 비표본오차는 조사 결과의 신뢰도를 떨어뜨리는 주요 원인이 된다.

Warner(1965)는 확률화응답기법(randomized response technique)을 도입하여, 응답자의 개인적 비밀을 보호하면서 모집단내의 민감집단의 비율을 추정하는 것을 가능하게 하였다. 이지(dichotomous)모집단의 경우 모비율을 추정하기 위하여 개발된 Warner의 기법과 원리는 이후 많은 학자들에 의하여 다지(polychotomous)모집단의 경우와 연속형 변수의 경우 등으로 응용되고 확장되었다. 또한 Fox와 Tracy(1986), Chaudhuri와 Mukerjee(1988)는 확률화응답기법을 체계적으로 정리하였으며, 국내에서는 류제복 등(1993)이 확률화응답기법에 관한 전문서적을 출판하였다.

모집단 비율을 추정하는 문제를 생각하자. 모집단 안의 개개인은 모두 민감집단(A)

---

1) This paper was supported by Wonkwang University in 2002.

2) Professor, Division of Mathematics and Informational Statistics and Institute of Basic Natural Sciences, Wonkwang University, Iksan, Jeonbuk 570-749, Korea.
E-mail: hjkim@wonkwang.ac.kr

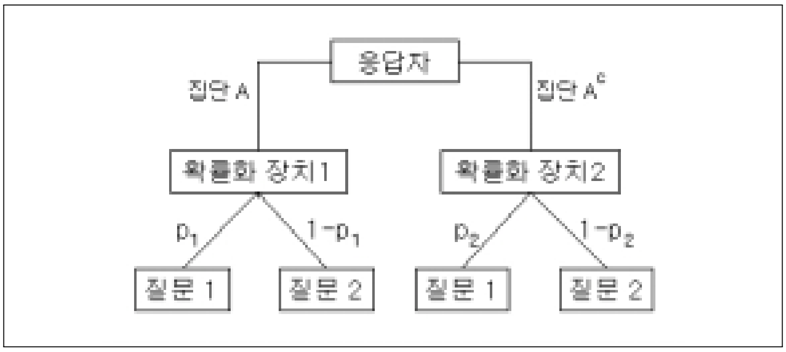
과 비민감집단(A^c) 중 하나에 속한다고 하고, 민감집단에 속하는 사람들의 모비율 π를 추정하고자 한다. Warner의 방법은 다음과 같다. 모집단으로부터 크기 n인 단순랜덤표본을 복원추출한다. 조사자는 다음과 같은 두 개의 질문을 준비한다.

질문 1: 당신은 민감집단(A)에 속하십니까?

질문 2: 당신은 비민감집단(A^c)에 속하십니까?

각 응답자에게는 질문 1과 질문 2중 하나만이 주어지는데, 질문 1이 주어질 확률은 p이고 질문 2가 주어질 확률은 $1-p$이다. 이 방법을 쓸 때 π의 추정량의 평균제곱오차가 직접질문법에서보다 작아지는 경우가 많다는 것이 밝혀졌다.

한편 김혁주(1988)는 Warner의 방법을 변형한 방법을 제시하였다. 이 방법에서는 집단 A에 속하는 사람의 경우 질문 1이 주어질 확률이 p_1이고 질문 2가 주어질 확률은 $1-p_1$이며, 집단 A^c에 속하는 사람의 경우 질문 1이 주어질 확률은 p_2이고 질문 2가 주어질 확률은 $1-p_2$이다. 이러한 확률화는 두 개의 확률화 장치(randomizing device)를 준비함으로써 이룰 수 있다. 응답자가 어느 집단에 속하는지는 응답자 자신만이 알고 있으므로 응답자는 조사자가 보지 못하는 곳에서 자신에게 해당되는 확률화 장치를 사용하는 것이다. "예" 또는 "아니오"의 대답이 응답자가 어느 집단에 속하는지를 알려주지 않으므로 Warner의 방법에서와 마찬가지로 응답자의 신분을 보호하면서 π를 추정할 수 있다. 여기서 $p_1 = p_2$이면 이 방법은 Warner의 방법과 같은 것이 된다. <그림 1.1>은 김혁주의 방법을 나타내 본 것이다.



<그림 1.1> 김혁주(1988)의 방법

그런데 조사자가 민감한 속성이라고 생각한 속성도 응답자에 따라서는 민감한 내용이 아닌 것으로 생각하여 직접질문에도 사실대로 응답하는 경우도 있다. 이러한 응답자에 대해서는 굳이 복잡한 확률화응답기법을 쓸 필요가 없이 직접질문법을 써서 효율성을 기할 수 있다. 일반적으로 어떤 속성을 갖는 사람들의 모비율 π의 값이 작을수록 그 속성은 민감한 속성이라고 할 수 있으므로, π의 값이 아주 작지 않을 것으로

생각되는 경우($0.2 \le \pi \le 0.4$ 정도)라면 처음부터 확률화응답기법을 쓸 필요 없이 직접질문법을 병용할 수 있을 것이다. 이러한 의미에서 최경호(2003)는 직접질문법과 Warner의 확률화응답기법을 결합한 조사절차를 제안하여 추정량의 신뢰성과 효율을 높일 수 있도록 하였다.

본 논문에서는 최경호의 아이디어를 확장하여, 앞에서 소개한 김혁주의 방법과 직접질문법을 결합하여 모비율을 추정하는 방법을 제안하고 이 방법의 성질을 논하며 이 방법의 효율성을 김혁주 및 최경호의 방법들과 비교하고자 한다.

2. 결합 추정 절차와 추정량

모집단 안의 민감집단의 비율 π를 다음 절차에 의하여 추정한다. 먼저 모집단으로부터 크기 n인 표본을 단순랜덤복원추출한다. 뽑힌 n명의 응답자들에게 직접질문(내용은 질문 1과 같음)을 하여 “예” 또는 “아니오”의 응답을 얻는다. 다음 “아니오”라고 응답한 응답자들에 대해서만 김혁주(1988)의 확률화응답기법을 적용하여 다시 응답을 얻는다. 비민감집단에 속하는 응답자들은 직접질문에 대해 모두 “아니오”라고 사실대로 응답하며, 확률화 장치를 사용하여 선택한 질문에 응답하는 응답자들도 Warner(1965)에서와 같이 사실대로 응답한다고 가정한다.

민감집단에 속하는 응답자가 직접질문에 사실대로 응답할 확률을 T라 하고 확률변수 $X_i (i=1,2,\cdots,n)$와 Y를 다음과 같이 정의하자.

$$X_i = \begin{cases} 1 & i\text{번째 응답자가 “예”라고 대답할 때} \\ 0 & i\text{번째 응답자가 “아니오”라고 대답할 때} \end{cases}$$

$$Y = \sum_{i=1}^{n} X_i$$

그러면 $i=1,2,\cdots,n$에 대해 $P(X_i=1)$은

$$\theta = P(X_i=1) = \pi T + \pi(1-T)p_1 + (1-\pi)(1-p_2) \tag{2.1}$$

이고 $X_1, X_2, \cdots, X_n$은 서로 독립이며, Y는 시행횟수가 n이고 성공확률이 θ인 이항분포를 따르게 된다. 최우추정법(method of maximum likelihood estimation)에 의하여 π를 추정하면 π의 최우추정량은

$$\hat{\pi} = \frac{\frac{Y}{n} - (1-p_2)}{T(1-p_1) + (p_1+p_2-1)} \tag{2.2}$$

로 얻어진다.

<정리 1> 식 (2.2)의 $\hat{\pi}$은 π의 불편추정량(unbiased estimator)이다.

(증명) $Y \sim B(n, \theta)$이므로

$$
\begin{aligned}
E(\hat{\pi}) &= \frac{\theta-(1-p_2)}{T(1-p_1)+(p_1+p_2-1)} \\
&= \frac{\pi T+\pi(1-T)p_1-\pi(1-p_2)}{T(1-p_1)+(p_1+p_2-1)} \\
&= \pi
\end{aligned}
$$

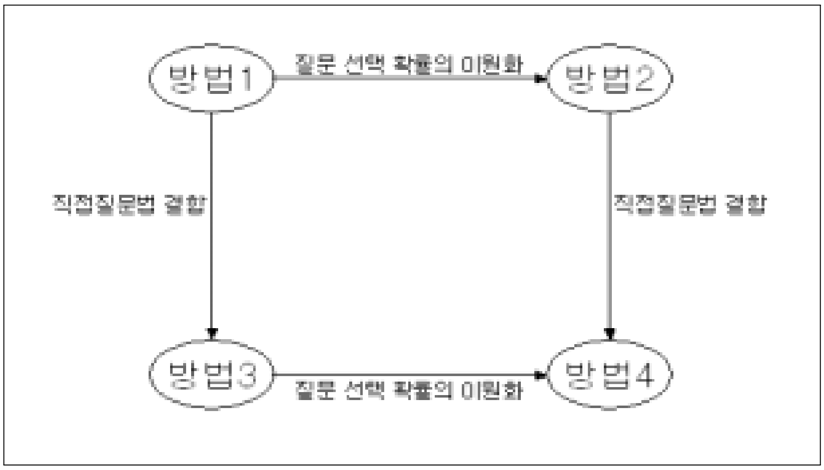
<정리 2> 식 (2.2)의 $\hat{\pi}$은 다음과 같은 분산을 갖는다.

$$
Var(\hat{\pi}) = \left\{\frac{2p_2-1}{T(1-p_1)+(p_1+p_2-1)}\right\}\left(\frac{\pi}{n}\right) - \frac{\pi^2}{n} + \frac{p_2(1-p_2)}{n\{T(1-p_1)+(p_1+p_2-1)\}^2} \tag{2.3}
$$

(증명)
$$
\begin{aligned}
Var(\hat{\pi}) &= \frac{Var(\frac{Y}{n})}{\{T(1-p_1)+(p_1+p_2-1)\}^2} \\
&= \frac{\theta(1-\theta)}{n\{T(1-p_1)+(p_1+p_2-1)\}^2}
\end{aligned}
$$
인데, 여기서 식 (2.1)을 대입하여 정리하면 식 (2.3)이 얻어진다.

3. 다른 방법들과의 비교

편의상 Warner(1965), 김혁주(1988), 최경호(2003)의 방법을 각각 방법 1, 방법 2, 방법 3이라 하고, 본 논문에서 제안된 방법을 방법 4라 하자. 그리고 방법 1, 방법 2, 방법 3, 방법 4에 의한 π의 추정량들을 각각 $\widehat{\pi_1}$, $\widehat{\pi_2}$, $\widehat{\pi_3}$, $\widehat{\pi_4}$로 표시하자. <그림 3.1>은 방법 1, 2, 3, 4의 관계를 나타낸 것이다. 이 절에서는 방법 4를 방법 2 및 방법 3과 비교한다. 방법 4는 방법 1에서 바로 나온 것이 아니고 한 단계(방법 2 또는 방법 3)를 거쳐서 만들어진 것이므로, 방법 1과 방법 4를 직접 비교하는 것은 필요성이 크지 않을 것이다. 마치 벡터를 합성하듯이, 방법 4와 방법 1의 차이는 방법 2와 방법 1의 차이에 방법 3과 방법 1의 차이를 합성한 것으로 생각하면 될 것이다.



<그림 3.1> 방법 1, 2, 3, 4의 관계

3.1 방법 2와 방법 4의 비교

방법 2에 의한 π의 추정량인 $\widehat{\pi_2}$의 분산은 다음과 같으며, 이것은 김혁주(1988, pp.51~52)에 주어져 있다.

$$Var(\widehat{\pi_2}) = \left(\frac{2p_2-1}{p_1+p_2-1}\right)\left(\frac{\pi}{n}\right) - \frac{\pi^2}{n} + \frac{p_2(1-p_2)}{n(p_1+p_2-1)^2} \tag{3.1}$$

식 (2.3)에서 $T=0$인 경우 이는 식 (3.1)과 같으며, 식 (2.2) 즉 $\widehat{\pi_4}$도 $T=0$인 경우 $\widehat{\pi_2}$와 같은 것임을 김혁주(1988, 식 (5))로부터 알 수 있다.

식 (2.3)과 식 (3.1)로부터 명백히 다음의 정리를 얻는다.

<정리 3> $T>0$, $p_1<1$이고 $p_2 \geq \frac{1}{2}$, $p_1+p_2>1$이면 $Var(\widehat{\pi_4}) < Var(\widehat{\pi_2})$, 즉 방법 4는 방법 2보다 효율적이다.

조건 $T>0$과 $p_1<1$이 성립하지 않는 경우는 무의미한 경우이므로, 방법 4가 방법 2보다 효율적이기 위한 실질적인 조건은 $p_2 \geq \frac{1}{2}$과 $p_1+p_2>1$이며, 우리는 이러한 조건을 충족하는 방법 4를 얼마든지 만들 수 있다.

π와 T의 몇 개 값에 대하여, 그리고 <정리 3>에 해당하는 p_1과 p_2의 몇 개 값에 대하여 $Var(\widehat{\pi_2}) / Var(\widehat{\pi_4})$의 값을 구한 결과가 부록의 <표 1>과 <표 2>에 정리되어 있다. 민감집단에 속하는 응답자가 직접질문에 사실대로 응답할 확률보다는

사실대로 응답하지 않을 확률이 높은 것이 현실적이라고 보아 $T \le 0.5$인 경우만 고려하였다. 이 표들로부터 $\pi = 0.2$일 때와 $\pi = 0.3$일 때 다음 사실들을 관찰할 수 있다.

(1) p_1과 p_2가 주어진 상태에서는 T가 클수록 방법 4가 방법 2에 비해 효율적이다.
(2) p_1과 T가 주어진 상태에서는 p_2가 작을수록 방법 4가 방법 2에 비해 효율적이다.
(3) p_2와 T가 주어진 상태에서는 p_1이 작을수록 방법 4가 방법 2에 비해 효율적이다.

π와 T의 값이 고정된 상태에서 방법 2에 대한 방법 4의 상대효율이 큰 값을 갖기 위한 p_1과 p_2의 조합을 찾는 문제를 생각해 보자. <표 1>과 <표 2>에서 보면 p_2가 크고 p_1은 작아서 $p_1 + p_2$가 1보다 조금 더 큰 정도일 때 상대효율이 커지는 경향이 대체적으로 존재하나, 반드시 그런 것은 아니다(예컨대 <표 1>에서 $T = 0.5$일 때, $p_1 = 0.2$, $p_2 = 0.9$이면 상대효율이 16.500인데, $p_1 = 0.3$, $p_2 = 0.8$이면 상대효율이 16.877). 식 (2.3)과 식 (2.4)를 근거로 한 상대효율이 복잡한 번분수식으로 표현되기 때문에, 일정한 패턴이 성립하는 것이 아니므로 표를 이용하여 상황에 따라 판단해야 할 것이다.

3.2 방법 3과 방법 4의 비교

방법 3에 의한 π의 추정량인 $\widehat{\pi_3}$의 분산은 다음과 같으며, 이것은 최경호(2003, p.277)에 주어져 있다.

$$Var(\widehat{\pi_3}) = \left\{ \frac{2p-1}{T(1-p)+(2p-1)} \right\} \left(\frac{\pi}{n} \right) - \frac{\pi^2}{n} + \frac{p(1-p)}{n\{T(1-p)+(2p-1)\}^2} \qquad (3.2)$$

본 논문에서 제안된 방법 4에서 $p_1 = p_2$인 특수한 경우가 방법 3이다. 부록의 <표 3>~<표 6>에는 $\pi = 0.3$인 경우 T의 여러 값에 대하여 $p = 0.6, 0.7, 0.8, 0.9$인 방법 3과 $p_1 + p_2 = 2p$인 방법 4의 효율성을 비교한 결과가 나와 있다. 표에 수록되어 있는 값들은 식 (2.3)과 식 (3.2)를 근거로 하여 계산된 $Var(\widehat{\pi_3})/Var(\widehat{\pi_4})$의 값들이다. 우리는 이 표들로부터 p_2가 크고 p_1이 작을수록 방법 3에 대한 방법 4의 상대효율이 커지는 경향이 있음을 볼 수 있으나, 방법 2와 비교할 때와 마찬가지로 이것 역시 항상 성립하는 것은 아니므로 그때그때 상황에 따라 판단해야 할 것이다.

3.3 조사비용의 비교

직접질문법에서 소요되는 단위당 조사비용을 C_D라 하고 확률화응답기법에서 소요되는 단위당 조사비용을 C_R이라 하자. 방법 2와 방법 4의 비교는 최경호(2003,

pp.277~278)에 나와 있는 방법 1과 방법 3의 비교와 매우 유사한 내용이 된다. 먼저 방법 2에서 필요한 총조사비용은 $C_2 = nC_R$이다. 방법 4의 경우를 보면, 먼저 n명에 대해서 1인당 조사비용이 C_D인 직접질문이 이루어지고 이들 중 "아니오"라고 응답하는 사람들에 대해 1인당 조사비용이 C_R인 확률화응답기법에 의한 조사가 이루어지는데, "아니오"라고 응답하는 사람의 평균수는 $n\pi(1-T)+n(1-\pi)$명, 즉 $n(1-\pi T)$명이므로, 방법 4에서 필요한 총조사비용의 기대값은 $E(C_4) = nC_D + n(1-\pi T)C_R$이다. 따라서 같은 크기의 표본에 대하여 생각할 때 방법 2에 대한 방법 4의 조사비용의 비의 기대값은

$$E\left(\frac{C_4}{C_2}\right) = \frac{C_D}{C_R} + 1 - \pi T \tag{3.3}$$

가 되며, $C_D/C_R < \pi T$이면 방법 4가 방법 2에 비해 평균 조사비용이 절감된다.

한편 방법 3과 방법 4의 조사비용을 비교하는 문제는 아주 간단하다. 직접질문에서 "아니오"라고 응답한 응답자가 방법 3에서는 모두 하나의 확률화 장치를 사용하는데, 방법 4에서는 두 개의 확률화 장치 중 한 개를 사용한다. 따라서 방법 4에서는 방법 3에 비하여 하나의 확률화 장치를 추가로 준비하는 비용만이 더 들 것이다.

4. 결 론

민감한 사안에 관한 조사에서 거짓응답이나 무응답으로 인해 발생하는 비표본오차는 조사의 신뢰도를 떨어뜨리는 큰 문제점이다. 응답자들의 개인적 비밀을 보호하여 협조를 얻음으로써 직접질문이 갖는 단점을 보완하여 신뢰성 있는 추정을 하기 위하여 고안된 방법이 확률화응답기법이다.

그런데 조사하는 사안의 민감도가 아주 높지 않은 경우 일부 응답자들은 직접질문을 해도 거짓응답이나 무응답을 하지 않고 사실대로 응답하므로, 표본으로 추출된 모든 응답자들에 대해서 처음부터 확률화응답기법을 적용할 필요가 없는 경우도 있다. 이러한 의미에서 최경호(2003)는 직접질문법과 Warner(1965)의 확률화응답기법을 결합하여 모비율을 추정하는 2단계 절차를 제안하였다.

본 논문에서는 최경호의 방법을 변형하여, 직접질문법과 김혁주(1988)의 확률화응답기법을 결합한 2단계 절차를 제안하였다. 제안된 방법은 $p_2 \geq \frac{1}{2}$이고 $p_1 + p_2 > 1$이면 김혁주의 방법보다 효율적인 것으로 밝혀졌다. 또한 부록의 <표 3>~<표 6>에서 볼 수 있듯이, $p_1 + p_2 = 2p$를 만족하는 것으로서 최경호의 방법보다 효율적인 제안된 방법이 항상 존재한다. p_1과 p_2의 차이가 너무 크면 응답자들의 완전한 협력을 얻기 어려울 것이다. 하지만 p_1과 p_2의 차이가 극히 작으면서도 최경호의 방법보다 제안된 방법이 효율적인 경우가 항상 존재한다는 것은 의미있는 결과라 할 수 있다.

본 논문에서는 T가 알려져 있다고 가정하였는데, 이는 최경호에 언급된 바와 같이

예비조사를 통하여 추측하거나 사전 지식, 이를테면 직접질문으로 수행된 비슷한 성격의 선행조사로부터 알아낼 수 있을 것이며, 만일 이것이 불가능할 경우에는 2표본을 이용하는 방안을 생각해 볼 만하다.

부 록

<표 1> $\pi=0.2$인 경우 방법 2에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.6	0.5	1.963	3.252	4.870	6.823	9.117
0.7	0.5	1.325	1.698	2.118	2.586	3.104
0.8	0.5	1.140	1.290	1.449	1.619	1.798
0.9	0.5	1.052	1.105	1.160	1.217	1.275
0.5	0.6	2.236	3.955	6.152	8.826	11.980
0.6	0.6	1.435	1.947	2.536	3.203	3.949
0.7	0.6	1.208	1.435	1.682	1.950	2.237
0.8	0.6	1.102	1.209	1.321	1.438	1.560
0.9	0.6	1.040	1.082	1.124	1.167	1.211
0.4	0.7	2.512	4.667	7.441	10.816	14.778
0.5	0.7	1.541	2.191	2.948	3.810	4.775
0.6	0.7	1.272	1.576	1.909	2.273	2.667
0.7	0.7	1.149	1.307	1.476	1.654	1.841
0.8	0.7	1.078	1.159	1.243	1.329	1.419
0.9	0.7	1.032	1.065	1.098	1.132	1.166
0.3	0.8	2.767	5.300	8.529	12.400	16.877
0.4	0.8	1.635	2.403	3.295	4.305	5.429
0.5	0.8	1.329	1.697	2.102	2.545	3.023
0.6	0.8	1.190	1.393	1.610	1.841	2.085
0.7	0.8	1.111	1.227	1.348	1.474	1.605
0.8	0.8	1.061	1.123	1.187	1.253	1.320
0.9	0.8	1.026	1.052	1.079	1.105	1.132
0.2	0.9	2.912	5.538	8.734	12.407	16.500
0.3	0.9	1.684	2.489	3.399	4.401	5.488
0.4	0.9	1.359	1.754	2.183	2.643	3.133
0.5	0.9	1.214	1.441	1.682	1.935	2.201
0.6	0.9	1.133	1.271	1.415	1.565	1.719
0.7	0.9	1.082	1.165	1.251	1.340	1.430
0.8	0.9	1.046	1.093	1.141	1.189	1.238
0.9	0.9	1.020	1.040	1.061	1.081	1.102

<표 2> $\pi=0.3$인 경우 방법 2에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.6	0.5	1.967	3.266	4.908	6.904	9.268
0.7	0.5	1.329	1.707	2.137	2.620	3.158
0.8	0.5	1.143	1.297	1.462	1.638	1.825
0.9	0.5	1.054	1.109	1.167	1.226	1.287
0.5	0.6	2.233	3.947	6.143	8.827	12.011
0.6	0.6	1.435	1.949	2.544	3.220	3.981
0.7	0.6	1.210	1.440	1.691	1.965	2.260
0.8	0.6	1.104	1.213	1.328	1.449	1.576
0.9	0.6	1.042	1.084	1.128	1.173	1.219
0.4	0.7	2.495	4.612	7.328	10.632	14.522
0.5	0.7	1.536	2.179	2.928	3.783	4.746
0.6	0.7	1.271	1.573	1.907	2.272	2.670
0.7	0.7	1.149	1.309	1.479	1.660	1.851
0.8	0.7	1.079	1.161	1.247	1.336	1.428
0.9	0.7	1.033	1.067	1.101	1.136	1.172
0.3	0.8	2.722	5.152	8.216	11.872	16.098
0.4	0.8	1.620	2.364	3.224	4.199	5.286
0.5	0.8	1.322	1.682	2.080	2.514	2.986
0.6	0.8	1.187	1.389	1.604	1.834	2.078
0.7	0.8	1.111	1.227	1.349	1.476	1.609
0.8	0.8	1.061	1.125	1.190	1.257	1.326
0.9	0.8	1.027	1.053	1.081	1.108	1.136
0.2	0.9	2.813	5.225	8.111	11.405	15.080
0.3	0.9	1.652	2.410	3.260	4.197	5.218
0.4	0.9	1.345	1.724	2.136	2.579	3.054
0.5	0.9	1.208	1.430	1.666	1.915	2.179
0.6	0.9	1.131	1.268	1.412	1.562	1.718
0.7	0.9	1.082	1.166	1.253	1.343	1.436
0.8	0.9	1.047	1.095	1.144	1.194	1.245
0.9	0.9	1.021	1.042	1.063	1.085	1.106

<표 3> $\pi=0.3$인 경우 $p=0.6$인 방법 3에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.3	0.9	2.128	2.285	2.368	2.409	2.423
0.4	0.8	1.457	1.565	1.637	1.684	1.714
0.5	0.7	1.154	1.205	1.241	1.267	1.285
0.55	0.65	1.064	1.089	1.107	1.120	1.129
0.65	0.55	0.957	0.933	0.915	0.902	0.891
0.7	0.5	0.933	0.883	0.847	0.820	0.800

<표 4> $\pi=0.3$인 경우 $p=0.7$인 방법 3에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.5	0.9	1.492	1.551	1.599	1.638	1.671
0.6	0.8	1.157	1.188	1.214	1.237	1.257
0.65	0.75	1.064	1.079	1.093	1.104	1.114
0.75	0.65	0.959	0.943	0.929	0.917	0.907
0.8	0.6	0.937	0.905	0.877	0.852	0.831
0.9	0.5	0.948	0.877	0.816	0.764	0.719

<표 5> $\pi=0.3$인 경우 $p=0.8$인 방법 3에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.7	0.9	1.191	1.212	1.231	1.249	1.266
0.75	0.85	1.077	1.088	1.098	1.108	1.116
0.85	0.75	0.949	0.938	0.927	0.917	0.908
0.9	0.7	0.919	0.896	0.874	0.853	0.834

<표 6> $\pi=0.3$인 경우 $p=0.9$인 방법 3에 대한 방법 4의 상대효율

p_1	p_2	T				
		0.1	0.2	0.3	0.4	0.5
0.85	0.95	1.098	1.106	1.114	1.122	1.129
0.88	0.92	1.035	1.038	1.041	1.044	1.048
0.92	0.88	0.971	0.967	0.964	0.960	0.957
0.95	0.85	0.934	0.926	0.917	0.909	0.901

참고문헌

1. 김혁주 (1988), Warner의 확률화 응답모형의 한 변형에 관하여, *품질관리학회지*, 제16권 제1호, 49-58.
2. 류제복, 홍기학, 이기성 (1993), <확률화응답모형>, 자유아카데미, 서울.
3. 최경호 (2003), Combined procedure of direct question and randomized response technique, *Journal of the Korean Data & Information Science Society*, 14, 275-278.
4. Chaudhuri, A. and Mukerjee, R. (1988), *Randomized Response: Theory and Techniques*, Marcel Dekker, Inc., New York.
5. Fox, J. A. and Tracy, P. E. (1986), *Randomized Response: A Method for Sensitive Surveys*, Sage Publications, Inc., Newbury Park.
6. Warner, S. L. (1965), Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.

[2003년 9월 접수, 2003년 11월 채택]