응용통계연구 제 5권 2호
1992년 pp 157-167

새로운 2단계 확률화응답모형

김종호1)·류제복2)·이기성1)

<요 약>

본 논문에서는 조사에 있어서 응답자들의 신분이나 비밀을 더 보장해 줌으로서 더욱 정확한 정보를 얻을 수 있는 새로운 2단계 확률화응답모형을 제안한다. 제안한 모형이 Warner모형, Liu-Chow의 반복시행모형, Mangat-Singh모형에 비하여 효율성이 있음을 보였다.

1. 서론

사회 여러분야의 조사에서 최근 연구의 관심은 응답을 회피하거나 고의적인 거짓응답으로 인한 비표본오차를 줄이는 데 있다. 비표본오차로서 발생되는 편의(bias)는 응답자들이 민감하거나 개인적인 이해와 관계되는 질문을 받았을 경우 매우 심하게 나타나고 있다. 이러한 편의를 줄이기 위해 1965년 Warner는 확률장치를 이용하여 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대해 정보를 이끌어 낼 수 있는 확률화응답기법 (randomized response technique : RRT)을 처음으로 제시하였다.

응답자의 신분이나 비밀을 보장함으로서 응답자로부터 민감한 질문에 보다 더 정확한 정보를 얻을 수 있는 이러한 확률화응답기법은 많은 학자들에 의해 연구, 발전 되어 왔다. 특히, Liu 와 Chow(1976)는 Warner모형의 확장으로서 반복시행모형을 제시하였고, Stem과 Steinhorst(1984)는 RRT를 우편조사나 전화조사등에 활용하는 방법을 제시하여 이의 사용을 실용화 시켰으며, Nathan(1988)은 1987년까지 발표된 RRT에 관한 250여종의 논문, 연구보고서, 책 등을 14가지 기준에 따라 분류한 분류목록을 작성하였다. 그리고, Chaudhuri 와 Mukerjee (1988)는 RRT에 대한 이론을 정리하여 체계화 시켰으며 Mangat 와 Singh (1990)은 두개의 확률장치를 이용하는 2단계 확률화응답모형을 제시하였다.

본 논문에서는 응답자의 신분이나 비밀을 보다 더 보장해주기 위하여 두개의 확률장치를 이용하는 새로운 2단계 확률화응답모형을 제시한다. 2장에서 Warner모형, Liu-Chow의 반복시행모형, Mangat-Singh모형에 대해 살펴보고, 3장에서는 새로운 2단

1) 100-715 서울 특별시 중구 필동 3가 26 동국대학교 통계학과
2) 360-764 충북 청주시 내덕동 36 청주대학교 응용통계학과

계 확률화응답모형을 제시하였다. 그리고 4장에서는 새로운 2단계 확률화응답모형이 Warner모형, Liu-Chow의 반복시행모형, 그리고 Mangat-Singh 모형보다 더 효율적이 되는 조건을 제시하고, 효율성이 있음을 보였다.

2. 확률화응답모형

본 장에서는 Warner모형과 Liu-Chow의 반복시행모형 그리고 Mangat-Singh모형에 대해 살펴보고자 한다.

2.1 Warner 모형

Warner(1965)가 처음 제시한 확률화 응답모형은 다음과 같다.

$$\begin{cases} \text{설문1 : "나는 민감한 그룹에 속한다"} \\ \text{설문2 : "나는 민감한 그룹에 속하지 않는다"} \end{cases}$$

단순임의복원추출된 n명의 응답자들이 설문1을 선택할 확률은 P 이고, 설문2를 선택할 확률이 $1-P$ 인 확률장치에 의해서 선택한 설문에 대해 "예" 또는 "아니오" 로만 응답한다.

이러한 확률화응답기법으로부터 응답자가 "예"라고 응답할 확률은 $\lambda = P\pi + (1-P)(1-\pi)$ 이다. 여기서, π 는 민감한 그룹에 속하는 모비율이다. n 명의 응답자중에서 "예" 라고 응답한 사람의 수를 n' 라고 하면 n' 는 $b(n, \lambda)$ 에 따른다. λ의 추정치는 $\hat{\lambda} = n'/n$ 가 되므로 π의 최우추정량 $\hat{\pi}_W$ 는 다음과 같다.

$$\hat{\pi}_W = \frac{n'}{(2P-1)n} + \frac{P-1}{2P-1}, \quad P \neq 1/2$$

그러나, Flinger 등(1977)은 $\hat{\pi}_W$가 특정구간에서는 음수값을 갖게 되므로 Warner의 $\hat{\pi}_W$ 를 제한된 최우추정량 (restricted MLE)이라 하고 실제적으로 π의 최우추정량이 다음과 같음을 지적하였다.

$$\hat{\pi}_W = \begin{cases} \dfrac{n'}{(2P-1)n} + \dfrac{P-1}{2P-1}, & P \neq 1/2 \text{ 이고 } \hat{\lambda} \in (1-P, P) \\ 1, & \hat{\lambda} \geq P > 1/2 \text{ 이거나 } \hat{\lambda} \leq P < 1/2 \\ 0, & \text{이외의경우} \end{cases}$$

최우추정량 $\hat{\pi}_W$의 기대값은

$$E(\hat{\pi}_W)=E\left(\frac{n'}{(2P-1)n}+\frac{P-1}{2P-1}\right)=\frac{\lambda}{(2P-1)}+\frac{P-1}{2P-1}=\pi$$

이 되므로 $\hat{\pi}_W$는 π의 불편추정량이 되며 분산은 다음과 같다.

$$V(\hat{\pi}_W)=\frac{\pi(1-\pi)}{n}+\frac{P(1-P)}{n(2P-1)^2}$$

2.2 Liu-Chow 의 반복시행모형

Liu-Chow(1976)가 제시한 반복시행모형은 기본적으로 Warner모형의 확장으로서 응답자가 Warner모형의 확률장치를 $m(m>1)$번 반복시행한 후 그때까지 "예"라고 응답한 횟수를 가지고 모비율 π를 추정하는 방법이다. 이 방법은 응답자로 하여금 시행을 여러번 반복하게 함으로서 모비율 π에 대한 정보를 더 많이 얻고, 또한 표본의 수를 증가시키지 아니하고도 분산을 감소시키므로서 효율을 높이고자 하였다. Liu-Chow는 반복시행시 응답자로 하여금 반복횟수를 많이 하지 않는 한 모비율 π를 추정하는데 매우 효율적인 방법임을 보였다.

Liu-Chow의 반복시행모형에서, j번째 응답자가 m번 시행에서 i번 "예"라고 응답한 횟수를 $X_j=i$ 로 나타내면

$$\Pr(X_j=i\,|\,m)=\binom{m}{i}[\pi P^i(1-P)^{m-i}+(1-\pi)P^{m-i}(1-P)^i]=W_i$$

$$j=1,2,\cdots,n \quad i=0,1,\cdots,m \quad \sum_{i=0}^{m}W_i=1$$

와 같이 된다.

단순임의복원추출된 n명의 응답자중에서 i번 "예"라고 응답한 사람들의 수를 n_i 라고 하면 $\sum_{i=0}^{m}n_i=n$ 이고 우도함수는 다음과 같다.

$$L=\prod_{i=0}^{m}W_i^{n_i}\quad,\quad \log L=\sum_{i=0}^{m}n_i\log W_i$$

우도함수가 위와 같이 복잡한 형태일 경우 $\frac{\partial \log L}{\partial \pi}=0$ 을 직접 계산하기가 곤란하므로 스코링방법(scoring method)을 이용하여 π의 최우추정량 $\hat{\pi}_L$ 을 구한다. π의 첫번째 근사치를 π_1 이라 하면

$$\pi_1=\pi_0+S(\pi_0)/I(\pi_0)=\pi_0+\delta_0$$

가 된다. 여기서, π_0 $(0 \le \pi_0 \le 1)$ 는 초기치이고 δ_0 는 수정항이며 $S(\pi_0)$ 와 $I(\pi_0)$ 는 $\pi=\pi_0$ 에서 스코어(score)와 정보(information)이다. 즉,

$$S(\pi_0)=\sum_{i=0}^{m}\frac{n_i}{W_i}\left(\frac{\partial W_i}{\partial \pi_0}\right)=\sum_{i=0}^{m}\frac{n_i[P^i(1-P)^{m-i}-P^{m-i}(1-P)^i]}{\pi_0 P^i(1-P)^{m-i}+(1-\pi_0)P^{m-i}(1-P)^i}$$

$$I(\pi_0)=\sum_{i=0}^{m}\frac{1}{W_i}\left(\frac{\partial W_i}{\partial \pi_0}\right)^2=n\sum_{i=0}^{m}\frac{\binom{m}{i}[P^i(1-P)^{m-i}-P^{m-i}(1-P)^i]^2}{\pi_0 P^i(1-P)^{m-i}+(1-\pi_0)P^{m-i}(1-P)^i}$$

새로운 수정항 δ_1 은 π_1 에 의해 얻어지며, 이런 과정을 수정항이 무시될 수 있을 정도로 작아질 때까지 반복하여 추정량 $\hat{\pi}_L$ 을 얻게된다. 따라서 추정량 $\hat{\pi}_L$ 는 편의가 매우 작아지게 되며 점근적으로 불편추정량이 된다. 이 추정량의 근사분산은 정보의 역수가 된다. 즉

$$V(\hat{\pi}_L)\approx\frac{1}{I(\pi)}$$

이다.

2.3 Mangat-Singh모형

Mangat-Singh(1990)은 응답자의 신분을 보장하기 위하여 다음과 같은 2단계의 확률화응답기법을 제시하였다. 단순임의복원추출된 n 명의 응답자들이 두개의 확률장치 R_1 과 R_2 을 이용하게 되며 1단계로 확률장치 R_1에서 다음 두 설문에 대하여 응답한다.

$$R_1\begin{cases}\text{설문1 : “나는 민감한 그룹에 속한다”}\\ \text{설문2 : “확률장치 } R_2 \text{ 로 가시오.”}\end{cases}$$

이때 설문1이 선택될 확률을 T 로두면, 설문2가 선택될 확률은 $1-T$ 이다. 한편 설문2가 선택된 경우 2단계로 확률장치 R_2 에서는 다음 두 설문에 대하여 선택확률 P 와 $1-P$ 를 갖고 응답한다.

$$R_2\begin{cases}\text{설문1 : “나는 민감한 그룹에 속한다”}\\ \text{설문2 : “나는 민감한 그룹에 속하지 않는다.”}\end{cases}$$

이러한 확률화응답기법으로부터 응답자가 “예”라고 응답할 확률은

$$\lambda=T\pi+(1-T)\ \{P\pi+(1-P)(1-\pi)\}$$

이다. 여기서, π 는 민감한 그룹에 속하는 모비율이다. n 명의 응답자 중에서 "예" 라고 응답한 사람의 수를 n' 라고 하면, n' 는 $b(n, \lambda)$ 에 따른다. λ의 추정치는 $\hat{\lambda}=n'/n$ 가 되므로 π 의 최우추정량 $\hat{\pi}_M$ 은 다음과 같다.

$$\hat{\pi}_M=\frac{n'/n-(1-T)(1-P)}{2P-1+2T(1-P)}$$

최우추정량 $\hat{\pi}_M$ 의 기대값은

$$E(\hat{\pi}_M)=E\left(\frac{n'/n-(1-T)(1-P)}{2P-1+2T(1-P)}\right)=\frac{\lambda-(1-T)(1-P)}{2P-1+2T(1-P)}=\pi$$

이므로, $\hat{\pi}_M$ 는 π 의 불편추정량이 된다. 그리고, 추정량 $\hat{\pi}_M$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_M)=\frac{\pi(1-\pi)}{n}+\frac{(1-T)(1-P)\{1-(1-T)(1-P)\}}{n\{2P-1+2T(1-P)\}^2}$$

3. 새로운 2단계 확률화응답모형

기존의 확률화응답모형보다 응답자들의 신분이나 비밀을 더욱 보장해 주기 위한 새로운 2단계 확률화응답모형을 생각할 수 있다. 민감한 그룹에 속하는 모비율을 추정하기 위하여 단순임의복원추출된 n 명의 응답자에게 다음과 같은 2단계의 확률장치 R_1 과 R_2 을 사용하여 응답하는 방법으로 앞의 Mangat-Singh(1990)의 확률장치를 이용한다. 1 단계의 확률장치 R_1 에서 다음 두 설문에 대하여 응답한다.

$$R_1=\begin{cases}\text{설문1 : "나는 민감한 그룹에 속한다."}\\ \text{설문2 : "확률장치 } R_2\text{로 가시오."}\end{cases}$$

설문1이 선택될 확률을 T로 두면, 설문2가 선택될 확률은 $1-T$이다. 한편 설문2가 선택된 경우 2단계의 확률장치 R_2에서는 다음 두 설문에 대하여 선택확률 P와 $1-P$를 갖고 응답한다.

$$R_2=\begin{cases}\text{설문1 : "나는 민감한 그룹에 속한다."}\\ \text{설문2 : "예" 라고 응답한다}\end{cases}$$

이러한 확률화 응답기법으로부터 응답자가 "예"라고 응답할 확률은

$$\lambda=T\pi+(1-T)\{P\pi+(1-P)\}$$

이다. 여기서, π는 민감한 그룹에 속하는 모비율이다. n명의 응답자 중에서 "예"라고

응답한 사람의 수를 n' 라고 하면, n' 는 $b(n, \lambda)$ 에 따르고, $\hat{\lambda}=n'/n$ 이므로 π 의 최우추정량 $\hat{\pi}_N$ 은 다음과 같다.

$$\hat{\pi}_N=\frac{n'/n-(1-T)(1-P)}{P+T(1-P)}$$

그리고, 최우추정량 $\hat{\pi}_N$ 의 기대값은

$$E(\hat{\pi}_N)=E\left[\frac{n'/n-(1-T)(1-P)}{P+T(1-P)}\right]=\frac{\lambda-(1-T)(1-P)}{P+T(1-P)}=\pi$$

이므로 추정량 $\hat{\pi}_N$ 는 π 의 불편추정량이 된다. 또한, 추정량 $\hat{\pi}_N$ 의 분산은 다음과 같다.

$$\begin{aligned} V(\hat{\pi}_N) &= \frac{\lambda(1-\lambda)}{n\{P+T(1-P)\}^2} \\ &= \frac{[\{1-P(1-T)\}\pi+(1-T)P][1-\{1-P(1-T)\}\pi-(1-T)P]}{n\{P+T(1-P)\}^2} \\ &= \frac{\pi(1-\pi)}{n}+\frac{(1-T)(1-P)(1-\pi)}{n\{P+T(1-P)\}} \end{aligned}$$

4. 효율성 비교

본 장에서는 제안한 새로운 2단계 확률화응답모형이 Warner모형, Liu-Chow의 반복시행모형, 그리고 Mangat-Singh모형보다 더 효율적이 되는 조건을 제시하고 그 효율성을 비교하고자 한다. 제안한 2단계 확률화응답모형은 아래조건에서 $V(\hat{\pi}_N)<V(\hat{\pi}_W)$ 이 성립하므로 Warner 모형보다 효율적이됨을 알 수 있다.

$$\frac{(1-T)(1-P)(1-3P)-PT}{(2P-1)^2(1-T)}<\pi\ ,\quad P\neq 1/2.$$

또한, Liu-Chow의 반복시행모형과의 비교에서는 제안한 2단계 확률화응답모형이 아래조건에서 효율적이다.

$$\frac{B-(B^2-4AC)^{1/2}}{2A}<\pi<\frac{B+(B^2-4AC)^{1/2}}{2A}\ ,\quad P\neq 1/2.$$

$$\text{단}, \; K = P + T(1-P)$$
$$A = 2P(1-2P)^2K^2$$
$$B = (1-2P)^2[(1-T)(2P^2-2P+1)+2KP]$$
$$C = (1-T)(1-2P)^2(2P^2-2P+1) - P^2(1-P)K$$

마찬가지로, 제안한 2단계 확률화응답모형이 아래조건에서 Mangat-Singh모형보다 효율적이 됨을 알 수 있다.

$$1-\left(\frac{K}{2K-1}\right)^2 < \pi \quad \text{단}, \; K = P + T(1-P).$$

그리고, Warner모형, Liu-Chow의 반복시행모형(반복수=2), Mangat-Singh모형과 새로운 2단계 확률화응답모형의 분산들을 n = 100 이고, π , P 와 T 를 변화시켜가면서 계산한 결과 각 모형의 분산들은 π 값에 별 영향을 받지 않는다. Warner모형, Liu-Chow의 반복시행모형과 새로운 2단계 확률화응답모형의 분산들은 작게 나타나고 있는 반면에, Mangat-Singh모형의 분산은 P 와 T 가 0.3~0.4 근처에서 다른 값들에 비해 매우 크게 나타나고 있음을 알 수 있다. Warner모형과 Liu-Chow의 반복시행모형의 분산들은 비슷한 모양을 하고 있으나 Liu-Chow의 반복시행모형의 분산이 Warner모형의 분산보다 전반적으로 작음을 알 수 있으며, 새로운 2단계 확률화응답모형의 분산은 P값이 작지 않고 T 값이 증가함에 따라 다른 세 모형에 비하여 분산이 작아짐을 알 수 있다.

이를 좀더 구체적으로 살펴보기 위하여, 새로운 2단계 확률화응답모형과 각 모형의 분산비를 계산한 결과 표 1를 얻었다. 표 1에서 1보다 큰 값은 새로운 2단계 확률화응답모형의 효율성이 다른 비교대상 모형에 비하여 좋음을 나타낸다.

표 1로부터 P값이 0.3보다 크고, T값이 증가함에 따라 전반적으로 새로운 2단계 확률화응답모형이 Warner모형보다 효율이 좋게 나타나고 있다. 특히, P값이 0.4에 가까워질수록 새로운 2단계 확률화응답모형의 효율이 월등함을 알 수 있다. 그리고, 새로운 2단계 확률화 응답모형은 Liu-Chow의 반복시행모형(반복수=2)에 대해서도 Warner모형과 비슷하게 P값이 0.3보다 크고, T 값이 증가함에 따라 전반적으로 효율이 좋게 나타나고 있다. 또한, P = 0.9인 경우에는 T가 크면 새로운 2단계 확률화응답모형이 Liu-Chow의 반복시행모형(반복수=2)보다 효율적이다. Mangat-Singh모형과의 효율성 비교에서는, 전반적으로 P값이 0.1보다 큰 경우 새로운 2단계 확률화응답모형이 Mangat-Singh 모형보다 효율이 좋게 나타나고 있다. 특히, T가 작고 P가 0.4에 가깝거나 P와 T가 0.3에 가까울수록 새로운 2단계 확률화응답모형의 효율이 상당히 좋게 나타나고 있다.

전반적으로 새로운 2단계 확률화응답모형이 P가 작지 않은 한 Warner모형, Liu-Chow 의 반복시행모형 및 Mangat-Singh모형보다 효율적이며 T가 증가함에 따라 그

표 1. 새로운 2단계 확률화응답모형과 Warner모형,Liu-Chow의 반복시행모형,Mangat-Singh모형의 효율성 비교

(W: Warner모형 L: Liu-Chow의 반복시행모형 M: Mangat-Singh모형)

	T	0.1			0.3			0.5			0.7			0.9		
	P	W	L	M	W	L	M	W	L	M	W	L	M	W	L	M
π=0.1	0.1	0.05	0.03	0.12	0.14	0.08	2.18	0.27	0.15	30.05	0.57	0.30	2.41	1.28	0.70	1.18
	0.3	0.86	0.39	2.18	1.46	0.66	654.48	2.44	1.10	4.55	4.25	1.91	1.77	8.89	4.00	1.12
	0.4	5.31	2.57	33.93	8.24	3.97	12.95	12.80	6.19	2.95	21.17	10.23	1.57	41.30	19.96	1.10
	0.7	3.31	1.49	2.41	4.25	1.91	1.77	5.63	2.53	1.45	7.83	3.52	1.28	11.90	5.35	1.10
	0.9	1.28	0.70	1.18	1.46	0.79	1.12	1.67	0.91	1.08	1.96	1.06	1.04	2.32	1.26	1.01
π=0.3	0.1	0.10	0.09	0.19	0.25	0.19	2.60	0.44	0.35	31.88	0.74	0.58	2.43	1.25	0.97	1.18
	0.3	1.08	0.60	2.60	1.72	0.95	708.12	2.59	1.43	4.66	3.84	2.11	1.77	5.79	3.19	1.13
	0.4	6.01	3.08	37.82	8.66	4.43	13.57	12.17	6.22	2.99	17.07	8.73	1.57	24.38	12.46	1.11
	0.7	3.24	1.79	2.43	3.84	2.11	1.77	4.56	2.51	1.40	5.45	3.00	1.18	6.57	3.61	1.04
	0.9	1.25	0.97	1.18	1.33	1.03	1.13	1.42	1.10	1.08	1.51	1.17	1.04	1.61	1.25	1.01
π=0.5	0.1	0.16	0.13	0.27	0.35	0.29	3.35	0.59	0.49	37.93	0.89	0.74	2.71	1.30	1.07	1.24
	0.3	1.41	0.82	3.35	2.13	1.24	855.70	3.00	1.75	5.34	4.08	2.37	1.84	5.43	3.15	1.17
	0.4	7.46	3.88	46.67	10.21	5.31	15.96	13.46	7.00	3.37	17.37	9.03	1.70	22.16	11.53	1.15
	0.7	3.59	2.08	2.71	4.08	2.37	1.94	4.61	2.70	1.50	5.21	3.03	1.24	5.88	3.41	1.06
	0.9	1.30	1.07	1.24	1.35	1.11	1.17	1.41	1.16	1.11	1.47	1.21	1.06	1.53	1.26	1.02
π=0.7	0.1	0.23	0.18	0.40	0.48	0.38	5.07	0.76	0.60	54.80	1.09	0.85	3.55	1.46	1.13	1.38
	0.3	2.11	1.16	5.07	3.05	1.68	1254.34	4.09	2.25	7.36	5.25	2.89	2.42	6.54	3.60	1.28
	0.4	11.04	5.64	69.41	14.53	7.43	22.76	18.34	9.37	4.50	22.51	11.50	2.07	27.10	13.85	1.23
	0.7	4.74	2.61	3.55	5.25	2.89	2.42	5.79	3.18	1.78	6.35	3.49	1.38	6.94	3.82	1.10
	0.9	1.46	1.13	1.56	1.50	1.17	1.41	1.55	1.20	1.28	1.59	1.24	1.16	1.64	1.27	1.05
π=0.9	0.1	0.44	0.24	0.94	0.88	0.48	13.59	1.34	0.73	144.57	1.81	0.99	8.04	2.30	1.25	2.12
	0.3	5.38	2.42	13.59	7.53	3.39	3957.92	9.75	4.38	18.19	12.03	5.41	5.00	14.38	6.46	1.82
	0.4	29.36	14.19	187.58	37.49	18.11	59.14	45.83	22.15	10.56	54.39	26.28	4.08	63.18	30.53	1.69
	0.7	11.04	4.96	8.04	12.03	5.41	5.00	13.02	5.86	3.25	14.04	6.31	2.12	15.06	6.77	1.32
	0.9	2.30	1.25	2.12	2.36	1.28	1.21	2.42	1.31	1.56	2.47	1.34	1.32	2.53	1.38	1.10

효율은 증가함을 알 수 있다.

새로운 2단계 확률화응답모형에서 1단계의 확률장치 R_1은 그대로 이용하고 2단계의 확률장치 R_2의 설문2 대신 민감한 그룹과 전혀 관계가 없는 "나는 그룹 U에 속한다." 는 무관질문모형을 사용한 경우 그룹 U에 속하는 모비율 π_U 를 안다고 가정할때 π의 최우추정량 $\hat{\pi}_{UN}$은 다음과 같다.

$$\hat{\pi}_{UN}=\frac{n'/n-(1-T)(1-P)\pi_U}{P+T(1-P)}$$

이 추정량 $\hat{\pi}_{UN}$ 는 π 의 불편추정량이 되며 분산은 다음과 같다.

$$V(\hat{\pi}_{UN})=\frac{\pi(1-\pi)}{n}+\frac{K(1-T)(1-P)\{\pi(1-2\pi_U)+\pi_U\}+\pi_U(1-\pi_U)(1-P)^2(1-T)^2}{nK^2}$$

$$\text{단,}\quad K=P+T(1-P).$$

이 확률화응답모형은 $\pi_U=1$ 인 경우 앞장에서 제안한 새로운 2단계 확률화응답모형과 같은 형태가 된다. 마찬가지 방법으로 위의 2단계 확률화응답모형과 Warner모형, Liu-Chow의 반복시행모형, Mangat-Singh모형과의 효율성을 비교해 본 결과 π_U값에 약간의 영향을 받을 뿐 효율성에서는 새로운 2단계 확률화응답모형과 거의 비슷한 결과를 나타내고 있다.

5. 결 론

본 논문에서는 통계조사에 있어 응답자들의 신분이나 비밀을 보다 더 보장해 주기 위하여 새로운 2단계 확률화응답모형을 제안하였다. 새로운 2단계 확률화응답모형이 Warner모형, Liu-Chow의 반복시행모형, Mangat-Singh모형에 대해 효율적이 되는 조건을 제시하였고, 제안한 2단계 확률화응답모형과 Warner모형, Liu-Chow의 반복시행모형 그리고 Mangat-Singh모형들과의 효율성을 비교한 결과 P값이 작지않고 T값이 증가할수록 전반적으로 본 논문에서 제안한 2단계 확률화응답모형이 효율적임을 알 수 있다.

그리고, 새로운 2단계 확률화응답모형에서 2단계의 확률장치 R_2의 설문2 대신 무관질문모형을 적용해 본 결과 π_U 값에 약간의 영향을 받을 뿐 효율성에서는 거의 비슷한 결과를 나타내고 있다.

기존의 확률화응답모형이 응답자의 신분을 완전하게 보장하기 어려우므로 본 논문에서 제시한 바와 같이 일정 조건하에서는 응답자의 신분을 좀 더 보장해줄 수 있는 새로운 2단계 확률화응답모형이 바람직하다고 본다.

< 참고문헌 >

[1] Chaudhuri, A. and Mukerjee, R. (1988), "*Randomized Response: Theory and Techniques*," Marcel Dekker, Inc., New York.

[2] Flinger, M. A., Policello, G. E. and Singh, J. (1977), "A Comparision of Two RR Survey Methods with Consideration for the Level of Respondent Protection," *Communications in Statistics.- Theory Methods* 6, 1511-1524.

[3] Greenberg, B. G., Abul-Ela, Abdel-Latif A., Simmons, W. R. and Horvitz, D. G. (1969), "The Unrelated Question Randomized Response Model; Theoretical Framework," *Journal of the American Statistical Association*, 64, 520-539.

[4] Kendall, M. G. and Stuart, A. (1967), *The Advanced Theory of Statistics*, Vol. 2, Hafner Publishing Company, New York.

[5] Liu, P. T. and Chow, L. P. (1976, a), "A New Discrete Quantitative

Randomized Response Model," *Journal of the American Statistical Association*, 71, 72-73.

[6] Liu, P. T. and Chow, L. P. (1976, b), "The Efficiency of the Multiple Trial Randomized Response Technique," *Biometrics*, 32, 607-618.

[7] Mangat, N. S. and Singh, R. (1990), "An Alternative Randomized Response Procedures," *Biometrika*, 77, 439-42.

[8] Nathan, G. (1988), "A Bibliography on Randomized Response : 1965 -1987," *Survey Methodology, A Journal of Statistics Canada*, 14, 331-346.

[9] Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, John-Wiley, New York.

[10] Stem, D. E. and Steinhorst, R. K. (1984), "Telephone Interview and Mail Questionaire Applications of the Randomized Response Model," *Journal of the American Statistical Association*, 79, 555-564.

[11] Warner, S. L. (1965), "Randomized Response; A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association*, 60, 63-69.

A New Two - Stage Randomized Response Model

Jong-Ho Kim[1] · Jea-Bok Ryu[2] · Gi-Sung Lee[1]

Abstract

This paper presents a new two-stage randomized response model to protect greater privacy of respondents for the sensitive characters. The conditions when the proposed model will be more efficient than Warner model, Liu-Chow's multiple trial model and Mangat-Singh model have been obtained for the case when the respondents are truthful in their answer, and the efficiency of the proposed model is also compared with Warner model, Liu-Chow's multiple trial model and Mangat-Singh model.

1) Department of Statistics, Dongguk University, Pil-dong 3 ga, Jung-gu, Seoul, 100-715, Korea.

2) Department of Applied Statistics, Chongju University, 36 Naedok-dong, Chongju-si, Chungbuk, 360-764, Korea.