한국통계학회논문집 제3권 1호
1996년 4월 pp. 121-130

신분보호를 고려한 의사 2단계 확률화응답기법[1)]

최 경 호[2)]

요 약

본 논문에서는 사회적 통념상 또는 개인적으로 매우 민감한 사안에 대한 조사시 무응답이나 거짓응답으로 인한 비표본오차를 줄여서 추정의 신뢰성을 높일 수 있도록 고안된 확률화응답기법을 이용하여 이지모집단내의 민감집단의 비율을 추정할 때 신분보호 정도가 같게되는 조건하에서 추정의 효율을 높일수 있는 의사 2단계 확률화응답기법을 제안하고 여러 형태의 확률화응답기법들과의 효율성 비교를 통하여 제안된 기법이 효율적일 수 있는 조건을 찾아 보았다.

1. 서 론

1965년 Warner는 사회적으로 또는 개인적으로 매우 민감한(sensitive)문제에 대한 조사시 비표본오차(non-sampling error)를 줄이어 추정의 신뢰성을 높이기 위한 방법으로 응답자에게 피해를 주지 않으면서 조사자가 얻고자 하는 정보를 얻을 수 있도록, 어떠한 민감한 질문에 대해서 응답자에게 직접적인 응답을 요구하는 것이 아니라 확률장치(random device)를 통한 간접적인 응답만을 요구 함으로써, 결과적으로 응답오차를 줄이고, 응답자의 호응을 높일 수 있는 확률화응답기법(randomized response technique)을 제시 하였다.

이후 Simmons(1967), Horvitz et al.(1976) 그리고 Mangat와 Singh(1990)등에 의하여 효율성 증대를 위한 확률화응답기법의 개선·보완이 이루어져 왔다. 그러나 확률화응답기법을 이용한 조사시 간과해서는 안될점 중의 하나는 신분보호의(privacy protection)의 문제이다. 그래서 응답자의 신분보호 정도를 측정하기 위한 노력도 행하여지고 있는데, 특히 Lanke(1976)는 조건부 확률을 이용하여 신분보호의 정도를 측정할 수 있는 측도를 제시하고 있다.

확률화응답기법을 이용한 조사에 있어서 획득되는 정보의 양과 신분보호(privacy protection) 정도 사이에는 상충되는 면이 있다. 따라서 이들 중의 어느 한 측면만을 지나치게 강조해서는 바람직한 조사결과를 얻지 못하게 된다. 그래서 이지(dichotomous)모집단내의 민감집단의 비율을 추정할 때 신분보호정도가 같게 되는 조건하에서 추정의 효율을 높일수 있는 의사(quasi) 2단계 확률화응답기법을 제안하고 여타의 확률화응답기법들과의 비교를 통하여 제안된 기법이 같은 신분보호정도 하에서 효율적일 수 있는 조건을 찾아 확률화응답기법을 이용한 조사시 효율성과 신분보호를 모두 고려한 조사를 행할 수 있는 토대를 마련해 보고자 한다.

---

1) 이 연구는 1995년도 전주대학교 학술연구비 지원에 의하여 수행되었습니다.
2) (560-759) 전주시 완산구 효자동 1200 전주대학교 통계학과 조교수.

2. 의사 2단계 확률화응답기법

Warner(1965)에 의하여 제시된 확률화응답기법에서는 모든 응답자가 확률 P로 민감한 질문에, 확률 $(1-P)$로 민감하지 않은 질문에 응답하도록 확률장치가 고안되어 있다. 이제 이를 응용한 다음의 방법을 제안하고 우리는 이 방법을 의사 2단계 확률화응답기법이라고 칭하도록 하겠다.

제안된 기법에서는 다음과 같은 두개의 확률장치가 필요하다.

$$\text{확률장치 I} \begin{cases} Q_1 : \text{당신은 민감한 집단에 속하는가 ? } (P_1) \\ Q_2 : \text{당신은 민감하지 않은 집단에 속하는가 ? } (1-P_1) \end{cases}$$

$$\text{확률장치 II} \begin{cases} Q_1 : \text{당신은 민감한 집단에 속하는가 ? } (P_2) \\ Q_2 : \text{당신은 민감하지 않은 집단에 속하는가 ? } (1-P_2) \end{cases}$$

이제 민감집단의 비율이 π인 이지모집단으로부터 단순임의복원추출에 의하여 n명의 표본을 추출한 다음 민감한 집단에 속하는 응답자는 확률장치 I을, 민감한 집단에 속하지 않는 응답자는 확률장치 II를 사용하여 Warner가 제안한 기법에서와 같은 방식으로 선택된 질문에 "예" 또는 "아니오"로 답하게 한다. 응답자 자신은 본인이 민감한 집단에 속하는지 혹은 속하지 않는지를 알고 있으므로 그에 따라 두개의 확률장치중 하나를 선택하면 된다. 이를 그림으로 나타내 보면 그림 2.1과 같다.

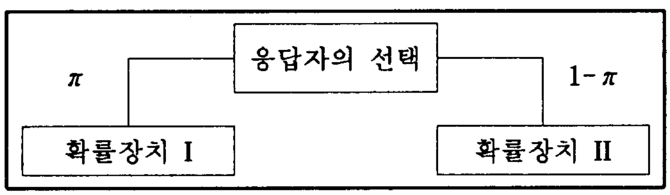

그림 2.1 의사 2단계 확률화응답기법의 과정

제안된 의사 2단계 확률화응답기법이 Warner기법과 다른점은 분명하다. 왜냐하면 Warner기법에서는 응답자들이 민감한 질문에 답할 확률이 P로서 동일한데 비해 제안된 기법에서는 선택하는 확률장치에 따라 민감한 질문에 답할 확률이 달라지게 된다. 특히 제안된 기법에서 $P_1 = P_2$인 경우가 바로 Warner기법이므로 결국 Warner기법은 제안한 기법의 한 특수한 경우로 여길수 있다. 제안된 기법을 이용한 조사과정에서도 Warner의 가정대로 모든 응답은 선택된 질문에 대하여 사실대로 이루어 진다고 하자.

확률변수 $Z_1, Z_2, \cdots, Z_n$과 Y를 다음과 같이 정의하자.

$$Z_i = \begin{cases} 1, & \text{만약 i번째 응답자가 "예"라고 응답하면} \\ 0, & \text{만약 i번째 응답자가 "아니오"라고 응답하면} \end{cases}$$

$$Y = \sum_{i=1}^{n} Z_i.$$

그러면 $Z_i(i=1,2,\cdots,n)$들은 서로 독립이며, Y는 n명중 "예"라고 응답한 응답자의 총 수로 이항분포를 따르게 된다.

[정리 2.1] 의사 2단계 확률화응답기법을 이용한 조사에서 모집단내의 민감집단의 비율 π의 불편추정량 $\widehat{\pi}_q$은 다음과 같다.

$$\widehat{\pi}_q = \frac{P_2-1}{P_1+P_2-1} + \frac{Y}{n(P_1+P_2-1)} \quad \text{단, } P_1+P_2 \neq 1 \tag{2.1}$$

(증명) 앞에서 정의된 확률변수 Z_i를 이용하자.

$$P_r(Z_i = 1) = \pi P_1 + (1-\pi)(1-P_2),$$
$$P_r(Z_i = 0) = \pi(1-P_1) + (1-\pi)P_2$$

이므로 $Y \sim b(n, \pi P_1 + (1-\pi)(1-P_2))$이다. 따라서, 우도함수는

$$L = \binom{n}{Y}[\pi P_1 + (1-\pi)(1-P_2)]^{Y} [\pi(1-P_1) + (1-\pi)P_2]^{n-Y}$$

로부터, π의 최우추정량 $\widehat{\pi}_q$은 식 (2.1)과 같다.

$\widehat{\pi}_q$의 기대값을 구해보면,

$$\begin{aligned} E(\widehat{\pi}_q) &= \frac{P_2-1}{P_1+P_2-1} + \frac{E(Y)}{n(P_1+P_2-1)} \\ &= \frac{P_2-1}{P_1+P_2-1} + \frac{n[\pi P_1 + (1-\pi)(1-P_2)]}{n(P_1+P_2-1)} \\ &= \pi \end{aligned}$$

이므로 $\widehat{\pi}_q$은 π의 불편추정량이 된다. ▤

이제 $\widehat{\pi}_q$의 분산을 구해보자.

[정리 2.2] 의사 2단계 확률화응답기법을 이용한 조사에서 식 (2.1)의 분산은 다음과 같다.

$$Var(\widehat{\pi}_q) = \frac{\pi(1-\pi)}{n} + \frac{\pi(P_2-P_1)}{n(P_1+P_2-1)} + \frac{P_2(1-P_2)}{n(P_1+P_2-1)^2} \tag{2.2}$$

(증명) 식 (2.1)로 부터

$$Var(\widehat{\pi}_q) = \frac{Var(Y)}{n^2(P_1+P_2-1)^2}$$

이며 $Y \sim b(n, \pi P_1 + (1-\pi)(1-P_2))$이므로 식 (2.2)가 성립한다. ▤

앞에서 언급한대로 식 (2.2)에서 $P_1 = P_2$이면 이는 Warner기법을 적용했을때의 π의 추정량, $\widehat{\pi}_w$, 의 분산인 $Var(\widehat{\pi}_w)$과 같게 된다.

의사 2단계 확률화응답기법을 이용함으로써 얻을 수 있는 장점은 응답을 각 집단별로 분리하여 얻게 되는 효과가 있다는 점이다. 즉 Warner기법을 포함한 기존의 확률화응답기법에서는 민감집단에 속하는 응답자들과 민감집단에 속하지 않는 응답자들의 응답이 혼합된 형태로 얻어지지만 본 논문에서 제안하고 있는 의사 2단계 확률화응답기법을 이용하면 각 집단별로 분리된 형태의

응답을 얻게 되는 효과가 있어 추정의 효율을 높일 수 있게 된다. 한편, 식 (2.2)에서 (P_1, P_2)는 $(1-P_1, 1-P_2)$일때와 동일하며, P_1+P_2가 1에서 멀리 떨어질수록 $\widehat{\pi_q}$의 분산은 작아진다. 그래서 분산만을 고려한다면 P_1은 민감집단에 속하는 응답자가 사용하는 확률장치내의 민감질문이 선택될 확률이고, P_2는 비민감집단에 속하는 응답자가 사용하는 확률장치내의 민감질문이 선택될 확률이므로 작은값의 P_1과 P_2를 선택하는 것이 좋겠으나 P_2가 너무 작거나 크면 $\widehat{\pi_q}$이 0과 1을 벗어나므로 P_1+P_2가 1보다 크게되지 않는 범위내에서 적절히 선택하는 것이 바람직 하겠다.

제안한 기법을 의사 2단계 확률화응답기법이라고 부르는 이유는 그림 2.1에서 알 수 있듯이 이 기법에 있어서 응답을 얻는 과정이 2단계의 응답과정으로 간주될 수 있으며 특히 첫번째 단계의 응답과정은 응답자 자신에 의해서 선택되지만 조사자에게 주어지는 응답이 없으며, 두번째 단계의 응답과정은 Warner과 같아 Mangat와 Singh(1990)의 2단계 확률화응답기법과 및 Warner(1965)의 기법과는 다르기 때문이다.

3. 신분보호 측도

확률화응답기법을 이용하여 이지모집단내의 민감속성의 비율을 추정함에 있어서 지금까지 발표된 많은 기법들에서는 Warner기법과의 분산비를 통하여 추정의 효율이 좋게되는 조건을 찾고자 노력하였다. 즉 이들에서는 추정하고자 하는 민감집단의 모비율에 대한 정보를 응답자로부터, 어떻게 하면 많이 얻을 수 있는 가에 문제의 핵심을 두고있다. 부언하면, 더욱 효율적인 추정량을 찾는데 있어 통계적인 관점에서만 접근하고 있다. 반면에 응답자의 입장에서는 자신이 정직하게 응답하여도 자신의 신분이 노출되지 않기를 바란다. 따라서 확률화응답기법을 이용한 조사시 우리가 간과해서는 안될 중요한 사항중의 하나는 응답자가 응답을 함으로써 느끼는 신분보호의 정도이다. 즉, 응답자가 자신의 응답결과 자신이 민감집단에 속하는 것으로 의심을 받을 확률이 클수록 응답자는 정직한 응답을 하지 않을 것으로 기대된다.

Lanke(1976)는 신분보호정도를 재는 측도(measure)로 조건부 확률을 이용한 다음의 방법을 제시하고 있다. 모집단내의 민감집단을 A라 했을때, 응답자의 "r"(r= "예" 또는"아니오")이라는 응답에 대하여 그 응답자가 민감집단에 속하는 것으로 여겨질 확률, 즉

$$P_r[\text{응답자가 민감집단에 속한다} \mid \text{응답자가 "r"이라는 응답을 한다}] \equiv P_r[\,A \mid \text{"r"}] \qquad (3.1)$$

을 계산하여 $\max[\,P_r(A \mid \text{예}),\ P_r(A \mid \text{아니오})]$ 이 작은 방법일수록 더욱 신분보호의 정도가 잘 되는 확률화응답기법이라 하였다. 이는 응답자가 본인의 응답으로 인하여 민감집단에 속하는 것으로 의심받을 확률이 작을수록 신분보호측면에서 더욱 좋은 확률화응답기법임을 의미한다. 따라서 통계적인 측면과 응답자의 입장을 모두 고려해 볼때, 확률화응답기법의 효율성에 대한 언급시에 분산뿐만 아니라 신분보호에 대해서도 함께 고려하는 것이 바람직 하겠다.

4. 효율성 비교

본 절의 목적은 이지모집단의 경우에 있어서 우리가 제안한 의사 2단계 확률화응답기법과 여러형태의 확률화응답기법들과의 효율성 비교시 단순히 분산을 비교하는 대신 Lanke(1976)가 제안한 측도를 이용하여 신분보호의 정도가 같게 되는 조건을 찾은 다음, 그러한 조건하에서 분산을 비교함으로써 같은 신분보호하에서 제안한 기법이 효율적인 조건을 찾고자 하는데 있다.

4.1 Warner기법과의 비교

본 논문에서 제안된 의사 2단계 확률화응답기법에 의한 추정량과 분산을 각각 $\widehat{\pi_q}$, $Var(\widehat{\pi_q})$이라 하고 Warner기법에 의한 추정량과 분산을 $\widehat{\pi_w}$, $Var(\widehat{\pi_w})$이라 했을 때 두 방법에 따른 조사시 신분보호의 정도가 같게 되는 조건을 찾아보자. Warner기법에서 민감집단이 선택될 확률을 P_w라 했을때 일반성의 상실없이, $P_w<1/2$을 가정할 수 있다. 그러면 Warner에서 $\max[\, P_{r_w}(A|예), P_{r_w}(A|아니오)] = P_{r_w}(A|아니오)$이고, 3.1절로부터 의사 2단계 확률화응답기법에서는 $P_1+P_2<1$일때 $\max[\, P_{r_q}(A|예), P_{r_q}(A|아니오)] = P_{r_q}(A|아니오)$이다.

[보조정리 4.1] 의사 2단계 확률화응답기법에서 $P_1+P_2<1$을 만족하는 모든 P_1과 P_2에 대하여 Warner기법과 제안한 기법의 신분보호정도가 같게 되는, 즉 $P_{r_w}(A|아니오) = P_{r_q}(A|아니오)$가 되는, P_w는 모든 π에 대하여 다음과 같다.

$$P_w = \frac{P_2}{1-P_1+P_2} \tag{4.1}$$

(증명) 먼저 Warner기법에서,

$$\begin{aligned} P_{r_w}(A|아니오) &= \frac{\pi P_{r_w}(아니오|A)}{P_{r_w}(아니오)} \\ &= \frac{\pi(1-P_w)}{P_w+\pi(1-2P_w)} \end{aligned} \tag{4.2}$$

이며, 의사 2단계 확률화응답기법에서,

$$\begin{aligned} P_{r_q}(A|아니오) &= \frac{\pi P_r(아니오|A)}{P_r(아니오)} \\ &= \frac{\pi(1-P_1)}{P_2+\pi(1-P_1-P_2)} \end{aligned} \tag{4.3}$$

이다. 따라서 식 (4.2)과 (4.3)로부터

$$\begin{aligned} P_w &= \frac{P_2(\pi-1)}{(P_1-P_2-1)-\pi(P_1-P_2-1)} \\ &= \frac{P_2}{1-P_1+P_2} \end{aligned}$$

이다. ▥

식 (4.1)의 P_w는 $P_w=\frac{1}{2}-\frac{1-P_1-P_2}{2(1-P_1+P_2)}$로서 가정한대로 $P_w<\frac{1}{2}$를 만족하며, Warner 기법에서 P_w가 주어지면 식(4.1)로 부터 신분보호정도가 같게되는 의사 2단계 확률화응답기법에서의 계획확률 P_1과 P_2를 찾을 수 있다.

이제 식 (4.1)의 조건하에서 의사 2단계 확률화응답기법이 Warner기법에 비해 효율적일 조건을 찾아보자.

[정리 4.1] $P_1+P_2<1$를 만족하는 P_1과 P_2에 대해 $P_1\le P_2$면 π에 관계없이 신분보호정도가 같은 조건하에서 의사 2단계 확률화응답기법이 Warner기법에 비해 항상 효율적이다. 즉, $Var(\widehat{\pi_q})\le Var(\widehat{\pi_w})$이다.

(증명) 식 (4.1)로 부터,

$$Var(\widehat{\pi_w}) = \frac{\pi(1-\pi)}{n} + \frac{P_2(1-P_1)}{n(P_1+P_2-1)^2}$$

이다. 따라서 식 (3.2)을 이용하면,

$$n[\ Var(\widehat{\pi_w})-Var(\widehat{\pi_q})] =\frac{1}{(1-P_1-P_2)^2}\{[\ \pi(1-P_1-P_2)+P_2]\ (P_2-P_1)\}$$

이므로 $P_1\le P_2$면 $Var(\widehat{\pi_q})\le Var(\widehat{\pi_w})$이 된다. 즉 제안된 기법에서 민감집단에 속하는 응답자들보다 비민감집단에 속하는 응답자들의 "아니오"라고 응답할 기회가 커질수록 제안된 의사 2단계 확률화응답기법은 Warner기법보다 같은 신분보호정도 하에서 효율적이다. ▥

표 4.1로부터 고정된 P_1과 P_2에 대해서는 π가 증가할 수록, 또 고정된 π와 P_1에 대해서는 P_2가 0.5에 가까울수록 의사 2단계 확률화응답기법이 Warner기법에 비해 신분보호정도는 같으면서 효율성이 향상되는 경향을 보이며, 고정된 π와 P_2에 대해서는 P_1이 증가 할수록 효율이 감소함을 알 수 있다.

표 4.1 신분보호정도가 같은 조건하에서 $Var(\widehat{\pi_q})/Var(\widehat{\pi_w})$

π = 0.1					π = 0.2				
P_1	P_2	$Var(\widehat{\pi_q})$	$Var(\widehat{\pi_w})$	$Var(\widehat{\pi_q})/Var(\widehat{\pi_w})$	P_1	P_2	$Var(\widehat{\pi_q})$	$Var(\widehat{\pi_w})$	$Var(\widehat{\pi_q})/Var(\widehat{\pi_w})$
0.2	0.3	.9100	.9300	.9785	0.2	0.3	.9600	1.0000	.9600
	0.4	1.5400	1.5900	.9686		0.4	1.5600	1.6600	.9398
	0.5	2.7678	2.8678	.9651		0.5	2.7378	2.9378	.9319
	0.6	5.8900	6.0900	.9672		0.6	5.7600	6.1600	.9351
	0.7	20.5900	21.0900	.9763		0.7	20.1600	21.1600	.9527
0.3	0.4	2.7233	2.7567	.9879	0.3	0.4	2.7600	2.8267	.9764
	0.5	6.2400	6.3400	.9842		0.5	6.2100	6.4100	.9688
	0.6	23.7900	24.0900	.9875		0.6	23.5600	24.1600	.9752
0.4	0.5	24.9900	25.0900	.9960	0.4	0.5	24.9600	25.1600	.9921

실질적으로, P_1은 민감집단에 속하는 응답자들이 사용하는 확률장치내의 민감질문이 선택될 확률이므로 작은 값으로 선택해주는 것이 바람직하며 이에 대하여 P_2는 0.5근처의 값을 선택해 줌으로써 같은 신분보호정도 하에서 Warner기법에 비하여 제안한 기법의 효율이 증가하게 된다. 그런 의미에서 이지모집단내의 민감집단의 비율을 추정하기 위한 조사시, 의사 2단계 확률화응답기법을 이용하면 계획확률 P_1과 P_2의 조정을 통하여 Warner기법에 비해 신분보호도 같은 수준으로 유지하면서 추정의 효율을 높일 수 있을 것이다.

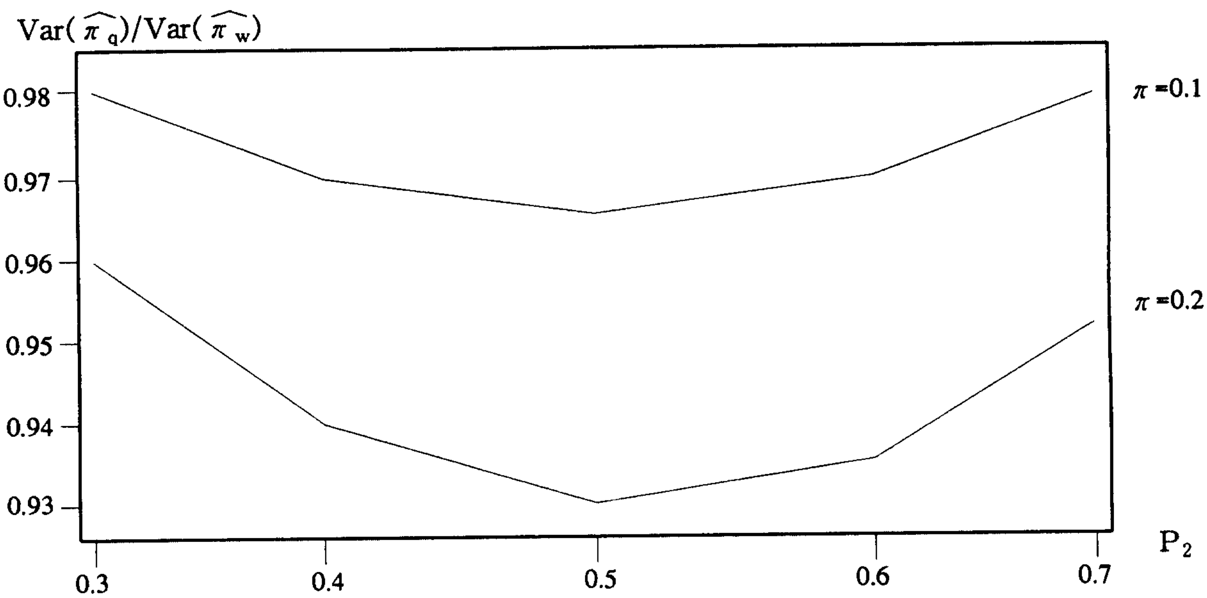

그림 4.1 고정된 π와 $P_1(0.2)$에 대하여 P_2의 변화에 따른 효율성 비교

4.2 Simmons기법과의 비교

Simmons(1967)의 기법은 Warner기법에 비하여 효율성을 증가시킬 수 있도록 고안된 조사기법이다. 이제 이 기법과 본 논문에서 제안된 의사 2단계 확률화응답기법과의 효율을 같은 신분보호정도하에서 비교하여보자. Simmons의기법에서 $\max[\,P_{r_s}(A \mid 예), P_{r_s}(A|아니오)] = P_{r_s}(A|예)$이다. 따라서, Simmons의 기법에 의한 추정량과 분산을 각각 $\widehat{\pi_s}$, Var($\widehat{\pi_s}$)이라 하면 의사 2단계 확률화응답기법과 Simmons의 기법과의 신분보호의 정도가 같게되는 조건은 다음과 같다.

[보조정리 4.2] Simmons의 기법에서 무관질문과 관련된 모집단 비율을 π_u라면, $P_1+P_2<1$를 만족하는 의사 2단계 확률화응답기법과의 신분보호의 정도가 같게되는 P_s 는 π에 관계없이 다음과 같다. 단, P_s는 Simmons의 기법에서 민감질문이 선택될 확률이다.

$$P_s = \frac{(1-P_1-P_2)\pi_u}{P_2+(1-P_1-P_2)\pi_u} \tag{4.4}$$

(증명) Simmons의 기법에서,

$$P_{r_s}(A|예)=\frac{\pi[\ P_s+(1-P_s)\pi_u]}{P_s\pi+(1-P_s)\pi_u} \tag{4.5}$$

이다. 따라서 식 (4.3)과 (4.5)로부터 $P_{r_q}(A|아니오)$ = $P_{r_s}(A|예)$를 만족하는 P_s는 식 (4.4)이다. ▥

Simmons의 기법에서 P_s가 주어지고 π_u가 기지(known)이면 식 (4.4)로부터 신분보호정도가 같게되는 의사 2단계 확률화응답기법에서의 계획확률 P_1과 P_2를 찾을 수 있다.

이제 식 (4.4)의 조건하에서 의사 2단계 확률화응답기법이 Simmons의 기법에 비해 효율적인 조건을 찾아보자.

[정리 4.2] 의사 2단계 확률화응답기법에서 $P_1+P_2<1$와 $P_1 \le P_2$를 만족하는 P_1과 P_2에 대해서, π에 관계없이 Simmons의 기법과 의사 2단계 확률화응답기법의 신분보호정도가 같으면 $Var(\widehat{\pi_q}) \le Var(\widehat{\pi_s})$일 조건은 다음과 같다.

$$\pi_u \le \frac{P_2}{P_1+P_2} \tag{4.6}$$

단, π_u는 Simmons의 기법에서 무관질문과 관련된 모집단의 비율이다.

(증명) 식 (4.4)를 이용하면,

$$nVar(\widehat{\pi_s}) = \pi(1-\pi)+\frac{P_2[\ (1-P_1-P_2)(1-2\pi_u)\pi+P_2]}{(1-P_1-P_2)^2\pi_u} + \frac{P_2(1-P_1)}{(1-P_1-P_2)^2}$$

이다. 따라서 식 (3.2)로부터,

$$n(Var(\widehat{\pi_s})-Var(\widehat{\pi_q})) = \frac{1}{(1-P_1-P_2)^2}\left[\ \frac{\pi(1-P_1-P_2)[\ P_2-\pi_u(P_1+P_2)]\ +P_2^2}{\pi_u}+P_2(P_2-P_1)\right]$$

이다. 따라서 식 (4.6)이 성립된다. ▤

표 4.2로부터 고정된 P_1, P_2와 π_u에 대해서는 π가 감소할 수록, 또 고정된 π, π_u와 P_1에 대해서는 P_2가 증가 할 수록 의사 2단계 확률화응답기법을 이용한 조사가 Simmons의 기법을 이용한 조사보다 신분보호정도를 같게한 경우에서효율이 좋아짐을 알 수 있다.

표 4.2 신분보호정도가 같은 조건하에서 $Var(\widehat{\pi_q})/Var(\widehat{\pi_s})$

$\pi = 0.1$						$\pi = 0.2$					
π_u	P_1	P_2	$Var(\widehat{\pi_q})$	$Var(\widehat{\pi_s})$	$Var(\widehat{\pi_q})/Var(\widehat{\pi_s})$	π_u	P_1	P_2	$Var(\widehat{\pi_q})$	$Var(\widehat{\pi_s})$	$Var(\widehat{\pi_q})/Var(\widehat{\pi_s})$
0.4	0.2	0.3	.9100	1.8600	.4892	0.4	0.2	0.3	.9600	1.9600	.4898
		0.4	1.5400	4.1400	.3720			0.4	1.5600	4.2600	.3662
		0.5	2.7678	9.8956	.2797			0.5	2.7378	10.0489	.2724
		0.6	5.8900	28.7400	.2049			0.6	5.7600	28.9600	.1989
		0.7	20.5900	143.9400	.1430			0.7	20.1600	144.3600	.1397
	0.3	0.4	2.7233	7.2678	.3747		0.3	0.4	2.7600	7.4044	.3727
		0.5	6.2400	22.0900	.2825			0.5	6.2100	22.2850	.2787
		0.6	23.7900	114.3901	.2080			0.6	23.5600	114.7601	.2053
	0.4	0.5	24.9900	87.8400	.2845		0.4	0.5	24.9600	88.1600	.2831
0.6	0.2	0.3	.9100	1.5100	.6026	0.6	0.2	0.3	.9600	1.5600	.6154
		0.4	1.5400	3.2233	.4778			0.4	1.5600	3.2600	.4785
		0.5	2.7678	7.4419	.3719			0.5	2.7378	7.4563	.3672
		0.6	5.8900	20.9900	.2806			0.6	5.7600	20.9600	.2748
		0.7	20.5900	102.5233	.2008			0.7	20.1600	102.3600	.1970
	0.3	0.4	2.7233	5.6752	.4799		0.3	0.4	2.7600	5.7007	.4841
		0.5	6.2400	16.6733	.3743			0.5	6.2100	16.6600	.3727
		0.6	23.7900	83.8901	.2836			0.6	23.5600	83.7601	.2813
	0.4	0.5	24.9900	66.5900	.3753		0.4	0.5	24.9600	66.4933	.3754

또한 정리 4.2로부터, 의사 2단계 확률화응답기법에서 계획확률 P_1과 P_2를 적절히 선택해 주면 모비율 π에 관계없이, 거의 모든 무관집단의 비율 π_u에 대하여 Simmons의 기법보다 신분보호정도를 같게 유지하면서 항상 효율적인 조사를 실시할 수 있음을 알 수 있다.

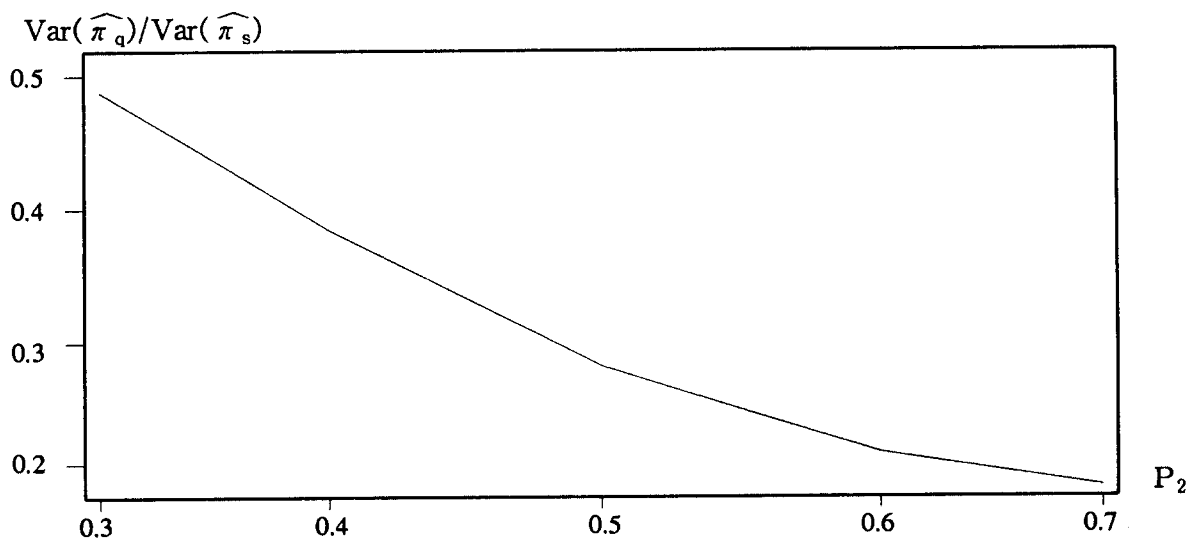

그림 4.2 고정된 $\pi(0.1)$, $\pi_u(0.4)$ 그리고 $P_1(0.2)$에 대하여

P_2의 변화에 따른 효율성 비교

5. 결 론

확률화응답기법은 사회정의 구현이나 정책상 필요한 조사임에도 불구하고 개인의 사생활과 매우 밀접하게 관련되어 있어서 직접조사방법인 일반적인 표본조사방법으로는 원하는 자료에 대한 정보수집이 곤란한 경우 효율적인 정보수집을 위하여 1965년 Warner에 의하여 개발된 간접조사방법이다. 본 논문에서는 이의 개선된 방법으로 의사 2단계 확률화응답기법을 제안하였다. 또한 확률화응답기법을 이용한 조사시 간과해서는 안될 점 중의 하나인 신분보호측면을 다루어 제안된 기법과 여러 형태의 확률화응답기법과의 신분보호정도가 같게되는 조건을 찾은 다음, 효율성비교시 단순히 분산을 비교하는 대신 신분보호정도가 같게되는 조건하에서 효율성을 비교해 보았다. 그 결과 Warner기법과의 비교에서는 $P_1+P_2<1$을 만족하는 P_1과 P_2에 대하여 $P_1 \le P_2$이면 의사 2단계 확률화응답기법이 같은 신분보호정도 하에서 항시 효율적임을 알 수 있었고, Simmons기법과의 비교에서는 $P_1+P_2<1$와 $P_1 \le P_2$를 만족하는 P_1과 P_2에 대하여 π_u가 $\pi_u \le P_2/(P_1+P_2)$를 만족하면 제안된 의사 2단계 확률화응답기법이 같은 신분보호정도 하에서 효율적임을 알수 있었다. 따라서 본 논문에서 제안된 의사 2단계 확률화응답기법을 이용하면 확률장치에서 P_1과 P_2의 적절한 조정을 통하여 기존의 여러형태의 확률화응답기법보다 신분보호정도가 같은 조건에서 더욱 추정의 효율이 좋은 조사 방법을 적용할 수 있겠다.

참 고 문 헌

[1] Horvitz, D. G., Shah, B.V. and Simmons, W.R.(1967). The unrelated question randomized responsemodel, *American Statistical Association Proceedings of social Statistics*, 65-72.

[2] Horvitz, D. G., Greenberg, B. G. and Abernathy, J. R. (1976). RR : a data gathering device for sensitive question, *International Statistical Review*, Vol. 44, 181-196.

[3] Lanke, J. (1976). On the degree of protection inrandomized interviews, *International Statistical Review*, Vol. 44, 197-203.

[4] Mangat, N. S. and Singh, R. (1990). An alternative randomized response procedure, *Biometrika*, Vol. 77, 439-442.

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

[6] Warner, S. L.(1965). Randomized response : A Servey technique for eliminating evasive answer bias, *Journal of the American Statistical Associatioin*, Vol. 60, 63-69.