

## 새로운 무관확률화응답모형

이기성<sup>1)</sup> 홍기학<sup>2)</sup>

### 요약

본 논문에서는 응답자가 민감한 속성을 가지고 있지 않으면 직접 “예”라고 응답하고, 민감한 속성을 가지고 있으면 Greenberg et al.(1969)의 무관질문모형의 확률장치를 이용하여 선택된 질문에 응답을 하는 새로운 무관확률화응답모형을 제안하였다. 그리고, 제안한 모형이 Mangat(1994)의 관련질문모형보다 효율적이 되는 조건을 제시하였고, 수치적으로 효율성을 비교하였다. 또한, Leysieffer와 Warner(1976)의 위험함수와 Flinger et al.(1977)의 사생활 보호 측도를 이용하여 제안한 모형이 Mangat의 관련질문모형에 비하여 개인의 사생활을 보호해 주는 측면에서 더 효율적임을 보였다.

### 1. 서론

사회 여러 분야의 조사에서 응답자들이 민감한 질문을 직접적으로 받게 되면 무응답이나 거짓응답 또는 응답을 회피하는 경향이 나타나므로 조사자는 응답자들로부터 정확한 정보를 얻지 못하는 경우가 종종 발생하게 된다. Warner(1965)는 응답자들에게 민감한 질문과, 민감한 질문과 배반되는 질문에 대해 확률장치를 사용하여 응답하게 함으로써 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 낼 수 있는 확률화응답모형(randomized response model; RRM)을 처음으로 제시하였다. 그리고, Greenberg et al.(1969)은 민감한 질문과 배반되는 질문 대신에 민감한 질문과 전혀 관계가 없는 질문을 사용하는 무관질문모형(unrelated question model)을 제안하였다. 그 후 수많은 학자들에 의해 이에 대한 연구가 확대되고 발전되었으며, 최근 Mangat-Singh(1990)은 2단계 관련질문모형을 제안하였다. 또한, Mangat(1994)은 Mangat-Singh의 2단계 관련질문모형에서 사용한 2개의 확률장치를 1개로 줄여 그 사용 절차를 좀 더 단순화하여 응답자가 민감한 속성을 가지고 있으면 “예”라고 응답하고, 민감한 속성을 가지고 있지 않으면 Warner의 확률장치를 이용하여 선택된 질문에 대해 “예” 또는 “아니오”라고 응답하도록 하였다. 그러나, Mangat의 모형에서 민감한 속성을 가지고 있는 응답자가 직접 “예”라고 응답하고, 민감한 속성을 가지고 있지 않은 응답자가 확률장치를 통해 또 다시 민감한 질문과 그 배반 관계를 가지고 있는 질문을 받게 된다면 응답하는데 거부감을 느낄 수 있다. 그러므로, 민감한 속성을 가지고 있지 않은 응답자가 직접 “예”라고 응답하고, 민감한 속성을 가지고 있는 응답자에게는 민감한 질문과 무관한 질문으로 구성된 Greenberg et al.의 확률장치를 사용하도록 하면 응답자의 거부감을 줄일 수 있다.

따라서, 본 논문에서는 응답자가 민감한 속성을 가지고 있지 않으면 “예”라고 응답하고, 민감한 속성을 가지고 있으면 Greenberg et al.의 무관질문모형의 확률장치를 이용하

1) (565-701) 전북 완주군 삼례읍 후정리 490, 우석대학교 전산통계학과, 조교수

2) (520-714) 전남 나주시 대호동 252, 동신대학교 컴퓨터학과, 부교수

여 선택된 질문에 응답을 하는 새로운 무관확률화응답모형을 제안하였다. 2장 1절에서는 Mangat의 모형에 대하여 간략히 정리하였고, 2절에서는 새로운 무관확률화응답모형을 제안하여 3절에서는 민감한 속성에 대한 추정량과 그 분산을 구하였다. 4절에서는 Leysieffer와 Warner(1976)의 위험함수와 Flinger et al.(1977)의 사생활 보호 측도를 제안한 모형에 적용하였으며, 5절에서는 응답자들이 확률장치를 사용함에도 불구하고 진실로 응답하지 않는 경우를 다루었다. 3장 1절에서는 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적이 되는 조건을 제시하였고, 수치적으로 효율성을 비교하였다. 또한, 3장 2절에서는 제안한 무관확률화응답모형이 Mangat의 관련질문모형에 비하여 개인의 사생활을 보호해 주는 측면에서 더 효율적임 보였다.

## 2. 새로운 무관확률화응답모형

### 2.1. MANGAT의 관련질문모형

단순임의복원추출된  $n$ 명의 응답자들은 민감한 속성  $A$ 를 가지고 있으면 “예”라고 응답하게 되고, 민감한 속성  $A$ 를 가지고 있지 않으면 다음과 같은 2개의 설문으로 구성된 Warner의 관련질문모형의 확률장치를 통해 선택된 질문에 대하여 응답한다.

설문 1 : 당신은 민감한 속성  $A$ 를 가지고 있습니까?

설문 2 : 당신은 민감한 속성  $A$ 를 가지고 있지 않습니까?

여기서, 설문 1이 선택될 확률은  $p$ 이고, 설문 2가 선택될 확률은  $1-p$ 이다. 이 때, 응답자들은 확률장치에 의해서 선택된 질문에 대해 “예” 또는 “아니오”라고 응답한다.

따라서, 이러한 Mangat의 관련질문모형에서 응답자가 “예”라고 응답할 확률을 구해 보면 다음과 같다.

$$\lambda = \pi + (1 - \pi)(1 - p), \quad (2.1)$$

여기서,  $\pi$ 는 민감한 속성에 대한 모비율이다.

단순임의복원추출된  $n$ 명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $n'$ 이라 하면  $\hat{\lambda} = n'/n$ 이 되므로 민감한 속성에 대한 모비율  $\pi$ 의 최우추정량  $\hat{\pi}_m$ 와 그 분산은 다음과 같다.

$$\hat{\pi}_m = \frac{\hat{\lambda} - (1 - p)}{p}, \quad (2.2)$$

$$V(\hat{\pi}_m) = \frac{\pi(1 - \pi)}{n} + \frac{(1 - \pi)(1 - p)}{np}. \quad (2.3)$$

### 2.2. 새로운 무관확률화응답모형

2장 1절에서 살펴본 Mangat의 관련질문모형에서는 민감한 속성을 가지고 있는 응답자에게 사실대로 “예”라고 직접 응답을 하게 하였고, 민감한 속성을 가지고 있지 않은 응답자는 확률장치를 통해 또 다시 민감한 질문과, 민감한 질문과 배반되는 질문을 받게 함으로

써 응답하는데 거부감을 느낄 수 있었다. 따라서, 이 절에서는 이러한 거부감을 줄이기 위하여 민감한 속성을 가지고 있지 않은 응답자에게 “예”라고 직접 응답을 하게 하고, 민감한 속성을 가지고 있는 응답자는 Greenberg et al.의 무관질문모형의 확률장치를 이용하여 선택된 질문에 응답을 하는 새로운 무관확률화응답모형을 제안하고자 한다.

단순임의복원추출된  $n$ 명의 응답자들은 민감한 속성  $A$ 를 가지고 있지 않으면

당신은 민감한 속성  $A$ 를 가지고 있지 않습니까?

라는 설문에 직접 “예”라고 응답하게 되고, 민감한 속성  $A$ 를 가지고 있으면 다음과 같은 2개의 설문으로 구성된 Greenberg et al.의 무관질문모형의 확률장치를 이용하게 된다.

설문 1 : 당신은 민감한 속성  $A$ 를 가지고 있습니까?

설문 2 : 당신은 무관한 속성  $Y$ 를 가지고 있습니까?

여기서, 설문 1이 선택될 확률은  $p$ 이고, 설문 2가 선택될 확률은  $1-p$ 이다. 이 때, 응답자들은 확률장치에 의해서 선택된 설문에 대해 “예” 또는 “아니오”라고 응답한다.

### 2.3. 제안한 모형에 대한 추정량과 그 성질

제안한 무관확률화응답모형에서 응답자가 “예”라고 응답할 확률을 구해 보면 다음과 같다.

$$\begin{aligned} \lambda &= (1-\pi) + \pi\{p + (1-p)\pi_y\} \\ &= 1 - \pi(1-p)(1-\pi_y), \end{aligned} \tag{2.4}$$

여기서,  $\pi_y$ 는 무관한 속성  $Y$ 의 모비율이며, 알고 있다고 가정한다.

단순임의복원추출된  $n$ 명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $n'$ 이라 하면  $\hat{\lambda} = n'/n$ 이 되므로 민감한 속성에 대한 모비율  $\pi$ 의 최우추정량  $\hat{\pi}_n$ 는 다음과 같다.

$$\hat{\pi}_n = \frac{1 - \hat{\lambda}}{(1-p)(1-\pi_y)}. \tag{2.5}$$

정리 2.1 추정량  $\hat{\pi}_n$ 는 모비율  $\pi$ 의 불편추정량이다.

증명:  $E(\hat{\lambda}) = \lambda$ 이므로,  $E(\hat{\pi}_n) = \frac{1 - \lambda}{(1-p)(1-\pi_y)} = \pi$ 가 성립된다. □

정리 2.2 제안한 무관확률화응답모형에서 민감한 속성에 대한 모비율  $\pi$ 의 추정량  $\hat{\pi}_n$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_n) = \frac{\pi(1-\pi)}{n} + \frac{\pi\{p + (1-p)\pi_y\}}{n(1-p)(1-\pi_y)}. \tag{2.6}$$

증명:  $V(\hat{\lambda}) = \frac{\lambda(1-\lambda)}{n}$ 이므로,  $V(\hat{\pi}_n) = \frac{\lambda(1-\lambda)}{n\{(1-p)(1-\pi_y)\}^2}$ 로부터 식 (2.6)을 얻을 수 있다. □

한편, 식 (2.5)의 추정량  $\hat{\pi}_n$ 는 특정구간에서 0과 1사이를 벗어나므로 이를 보완하여 구한 수정된 최우추정량  $\tilde{\pi}_n$ 는 다음과 같다.

$$\tilde{\pi}_n = \begin{cases} 0, & 1 \leq \hat{\lambda} \\ \hat{\pi}_n, & \{p + (1-p)\pi_y\} < \hat{\lambda} < 1 \\ 1, & \hat{\lambda} \leq \{p + (1-p)\pi_y\} \end{cases} \quad (2.7)$$

이 때, 수정된 최우추정량  $\tilde{\pi}_n$ 은 불편추정량이 아니기 때문에 오차의 측도로서  $\pi$ 에 대한 평균제곱오차(MSE)를 사용해야 한다.

정리 2.3  $\tilde{\pi}_n$ 의 평균제곱오차는  $\hat{\pi}_n$ 의 평균제곱오차보다 작다.

$$MSE(\tilde{\pi}_n) \leq MSE(\hat{\pi}_n) \quad (2.8)$$

증명:

$$I(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

로 두고,  $\tilde{\pi}_n$ 의 평균제곱오차와  $\hat{\pi}_n$ 의 평균제곱오차의 관계를 살펴보면

$$\begin{aligned} MSE(\hat{\pi}_n) &= E[(\hat{\pi}_n - \pi)^2] \\ &= E[(\hat{\pi}_n - \pi)^2 I(-\hat{\pi}_n)] + E[(\hat{\pi}_n - \pi)^2 I(\hat{\pi}_n) I(1 - \hat{\pi}_n)] + E[(\hat{\pi}_n - \pi)^2 I(\hat{\pi}_n - 1)] \\ &\geq E[\pi^2 I(-\hat{\pi}_n)] + E[(\hat{\pi}_n - \pi)^2 I(\hat{\pi}_n) I(1 - \hat{\pi}_n)] + E[(1 - \pi)^2 I(\hat{\pi}_n - 1)] \\ &= MSE(\tilde{\pi}_n) \end{aligned}$$

이므로, 식 (2.8)을 얻을 수 있다.  $\square$

## 2.4. 위험함수

제안한 무관확률화응답모형에 대하여 Leysieffer와 Warner가 제안한 응답자가 민감한 질문에 대하여 응답을 함으로써 입게 되는 사생활의 노출 위험 정도를 측정할 수 있는 위험함수를 구해보기로 하자. 또한, Flinger et al.의 사생활 보호 측도에 대해서 살펴보고, 제안한 모형에 이를 적용해 보고자 한다.

먼저, Leysieffer와 Warner는 위험함수를 다음과 같이 정의하였다.

$$g(R, A) = \frac{p(R|A)}{p(R|A^c)}, \quad (2.9)$$

$$g(R, A^c) = \frac{p(R|A^c)}{p(R|A)}, \quad (2.10)$$

여기서,  $p(R|A)$ 는 민감한 속성  $A$ 를 가지고 있는 응답자가  $R$ 이라고 응답할 확률이고,  $p(R|A^c)$ 는 민감하지 않은 속성  $A^c$ 를 가지고 있는 응답자가  $R$ 이라고 응답할 확률이다. 이 때, 식

(2.9)와 식 (2.10)에서  $g > 1$ 인 경우에 각각  $A$ 와  $A^c$ 에 속한 응답자의 신분노출 위험을 나타낸다.

따라서, 제안한 무관확률화응답모형에서 “예”라는 응답을  $Y$ 로 표시하고, “아니오”라는 응답을  $N$ 으로 표시하면, 다음을 얻을 수 있다.

$$\begin{aligned} p(Y|A) &= p + (1-p)\pi_y, & p(N|A) &= (1-p)(1-\pi_y), \\ p(Y|A^c) &= 1, & p(N|A^c) &= 0, \\ p(Y) &= (1-\pi) + \pi\{p + (1-p)\pi_y\}, \\ p(N) &= \pi(1-p)(1-\pi_y). \end{aligned}$$

이 때, 응답자들이  $Y$ 나  $N$ 으로 응답을 했을 때, 자신의 민감한 속성  $A$ 를 드러낼 확률은 각각 다음과 같다.

$$p(A|Y) = \frac{p(A)p(Y|A)}{p(Y)} = \frac{\pi\{p + (1-p)\pi_y\}}{1 - \pi(1-p)(1-\pi_y)}, \quad (2.11)$$

$$p(A|N) = \frac{p(A)p(N|A)}{p(N)} = \frac{\pi\{(1-p)(1-\pi_y)\}}{\pi(1-p)(1-\pi_y)} = 1. \quad (2.12)$$

제안한 무관확률화응답모형을 Leysieffer와 Warner의 위험함수를 이용하여 표현해 보면 다음과 같다.

$$g(Y, A) = \frac{p(Y|A)}{p(Y|A^c)} = p + (1-p)\pi_y, \quad (2.13)$$

$$g(N, A^c) = \frac{p(N|A^c)}{p(N|A)} = 0. \quad (2.14)$$

다음으로 Flinger et al.은 다음과 같은 식을 사생활 보호 측도로 사용하였다.

$$J_1 = \frac{1 - \max\{p(A|Y), p(A|N)\}}{1 - \pi} \quad (2.15)$$

제안한 무관확률화응답모형에 대하여 식 (2.11)과 식 (2.12)를 식 (2.15)에 대입하면

$$\max\{p(A|Y), p(A|N)\} = 1$$

이므로,  $J_1 = 0$ 를 얻을 수 있다. 이는 제안한 무관확률화응답모형이 응답자들의 사생활을 잘 보호해 주고 있음을 나타내는 것이다.

### 2.5. 응답자가 진실된 응답을 하지 않았을 경우

이 절에서는 민감한 속성을 가지고 있는 응답자가 완전히 진실된 응답을 하지 않은 경우에 대하여 살펴보기로 하자. 이 때, 민감한 속성을 가지고 있지 않은 응답자들이 거짓으로 응답할 이유는 없다고 가정한다. 따라서, 제안한 무관확률화응답모형에서 응답자들이 “예”라고 응답할 확률은 다음과 같다.

$$\lambda_1 = (1-\pi) + \pi\{\theta p + (1-p)\pi_y\}, \quad (2.16)$$

여기서,  $\theta$ 는 민감한 속성을 가지고 있는 응답자가 진실로 응답할 확률이다.

단순임의복원추출된  $n$ 명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $n'$ 이라 하면  $\hat{\lambda}_1 = n'/n$ 이 되므로 민감한 속성에 대한 모비율  $\pi$ 의 최우추정량  $\hat{\pi}'_n$ 는 다음과 같다.

$$\hat{\pi}'_n = \frac{1 - \hat{\lambda}_1}{1 - \theta p - (1-p)\pi_y}. \quad (2.17)$$

식 (2.17)의 추정량  $\hat{\pi}'_n$ 는 모비율  $\pi$ 의 불편추정량이다. 그러나, 실제적으로  $\theta$ 의 값을 알지 못하므로  $\hat{\pi}'_n$ 를  $\pi$ 의 추정치로 사용할 수가 없다. 따라서, 모비율  $\pi$ 를 추정하기 위하여 식 (2.5)의  $\hat{\pi}_n$ 를 추정량으로 사용하고자 한다. 이 때, 추정량  $\hat{\pi}_n$ 는 다음과 같은 편의를 갖는다.

$$B(\hat{\pi}_n) = \frac{\pi p(1-\theta)}{(1-p)(1-\pi_y)}.$$

그러므로, 제안한 무관확률화응답모형의 절차에서 민감한 속성을 가지고 있는 응답자가 진실로 응답하지 않을 경우에  $\hat{\pi}_n$ 의 평균제곱오차를 구해보면 다음과 같다.

$$\begin{aligned} &MSE(\hat{\pi}_n) \\ &= \frac{\lambda_1(1-\lambda_1)}{n\{(1-p)(1-\pi_y)\}^2} + \frac{\{\pi p(1-\theta)\}^2}{\{(1-p)(1-\pi_y)\}^2} \\ &= \left\{ \frac{1}{n} \left[ \pi\theta(1-\pi\theta) + \pi \{1-\theta p - (1-p)\pi_y\} \{1-\pi\{1-\theta p - (1-p)\pi_y\}\} - \theta(1-\pi\theta) \right] \right. \\ &\quad \left. + \{\pi p(1-\theta)\}^2 \right\} \cdot \frac{1}{\{(1-p)(1-\pi_y)\}^2}. \end{aligned} \quad (2.18)$$

### 3. 두 모형의 효율성 비교

#### 3.1. 효율성 비교

이 절에서는 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적이 되는 조건을 제시하고, 수치적으로 효율성을 비교해 보고자 한다.

제안한 무관확률화응답모형의 분산 식 (2.6)과 Mangat의 관련질문모형의 분산 식 (2.3)을 이용하여  $V(\hat{\pi}_n) < V(\hat{\pi}_m)$ 를 만족하는 조건을 구해보면 식 (3.1)을 얻을 수 있다.

$$\pi < \frac{1}{1 + \frac{p\{p + (1-p)\pi_y\}}{(1-p)^2(1-\pi_y)}}. \quad (3.1)$$

식 (3.1)의 우변 항은  $p$ 값이 0에 가까워지면 1에 가까운 값을 가지게 되며,  $p$ 값이 1에 가까워지면 0에 가까운 값을 가지게 된다. 따라서,  $p$ 값이 작을 경우  $\pi$ 값에 관계없이 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적인 반면,  $p$ 값이 클 경우에는  $\pi$ 값이 작을 때에 한해서 제안한 무관확률화응답모형이 더 효율적임을 알 수 있다. 물론,  $\pi_y$ 값에도 영향을 받기는 하지만 그 영향력은 상대적으로 작은 편이다.

이러한 사실을 구체적으로 살펴보기 위하여 수치적인 비교를 통하여 효율성을 비교해 보고자 한다. 따라서,  $n = 100$ 일 때,  $\pi$ 와  $p$  및  $\pi_y$ 를 변화시켜 가면서  $\hat{\pi}_m$ 에 대한  $\hat{\pi}_n$ 의 상대 효율(relative efficiency)

$$RE = V(\hat{\pi}_m)/V(\hat{\pi}_n)$$

를 계산한 결과 다음 표 3.1을 얻었다. 표 3.1에서 1보다 큰 값은 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 더 효율적임을 나타낸다. 식 (3.1)에서 살펴보았듯이 표 3.1에서도 제안한 무관확률화응답모형은  $p$ 값이 작을 경우  $\pi$ 값에 관계없이 Mangat의 관련질문모형보다 효율이 좋게 나타나고 있으며,  $p$ 값이 클 경우에는  $\pi$ 값이 작을 때에 한해서 효율적으로 나타나고 있다.

표 3.1: 제안한 무관확률화응답모형과 Mangat의 관련질문모형과의 효율성 비교

$\pi$	$p \backslash \pi_y$	0.1	0.3	0.5	0.7	0.9
		0.1	0.1	72.1860	55.0661	38.5916
	0.3	14.7246	11.2839	7.9430	4.6976	1.5438
	0.5	4.6649	3.5906	2.5384	1.5076	0.4974
	0.7	1.3200	1.0204	0.7244	0.4320	0.1431
	0.9	0.1725	0.1339	0.0954	0.0571	0.0190
0.3	0.1	23.2192	16.8569	11.2890	6.3754	2.0071
	0.3	4.7731	3.5296	2.4028	1.3770	0.4393
	0.5	1.5780	1.1862	0.8198	0.4764	0.1539
	0.7	0.4994	0.3810	0.2670	0.1572	0.0514
	0.9	0.0887	0.0685	0.0486	0.0290	0.0096
0.5	0.1	12.9327	8.7372	5.5161	2.9653	0.8952
	0.3	2.6058	1.8388	1.2020	0.6648	0.2055
	0.5	0.8709	0.6363	0.4285	0.2432	0.0769
	0.7	0.2898	0.2178	0.1505	0.0875	0.0282
	0.9	0.0575	0.0443	0.0313	0.0186	0.0061
0.7	0.1	7.7766	4.6851	2.7309	1.3840	0.3992
	0.3	1.4651	0.9695	0.6026	0.3200	0.0956
	0.5	0.4786	0.3377	0.2207	0.1221	0.0377
	0.7	0.1610	0.1190	0.0810	0.0464	0.0148
	0.9	0.0333	0.0255	0.0180	0.0106	0.0035
0.9	0.1	3.2878	1.6004	0.8319	0.3923	0.1077
	0.3	0.5227	0.3149	0.1835	0.0930	0.0268
	0.5	0.1596	0.1078	0.0681	0.0366	0.0110
	0.7	0.0526	0.0382	0.0255	0.0144	0.0045
	0.9	0.0110	0.0083	0.0058	0.0034	0.0011

그리고,  $\pi_y$  값이 작을수록 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적이다. 좀 더 구체적으로 살펴보면  $\pi$  값이 0.5보다 작고  $p$ 와  $\pi_y$  값이 0.3보다 작을 때는 제안한 무관확률화응답모형의 효율이 좋게 나타나고 있다. 그리고,  $\pi$  값이 0.7보다 클 때  $p$  값이 0.1 정도로 아주 작은 경우를 제외하고는 Mangat의 관련질문모형의 효율이 좋게 나타나고 있다.

### 3.2. 비밀보장의 정도 비교

제안한 무관확률화응답모형과 Mangat의 관련질문모형에 대하여 Leysieffer와 Warner가 제안한 위험함수와 Flinger et al.의 사생활 보호 측도를 이용하여 사생활의 노출 위험 정도를 비교해 보고자 한다.

Mangat의 관련질문모형에서

$$\begin{aligned} p(Y|A) &= 1, & p(N|A) &= 0, \\ p(Y|A^c) &= 1-p, & p(N|A^c) &= p, \\ p(Y) &= \pi + (1-\pi)(1-p), \\ p(N) &= (1-\pi)p \end{aligned}$$

이므로, 응답자들이  $Y$ 나  $N$ 으로 응답을 했을 때, 자신의 민감한 속성  $A$ 를 드러낼 확률은 각각 다음과 같다.

$$\begin{aligned} p(A|Y) &= \frac{p(A)p(Y|A)}{p(Y)} = \frac{\pi}{\pi + (1-\pi)(1-p)}, \\ p(A|N) &= \frac{p(A)p(N|A)}{p(N)} = \frac{0}{(1-\pi)p} = 0. \end{aligned}$$

Mangat의 관련질문모형을 Leysieffer와 Warner의 위험함수를 이용하여 표현해 보면

$$g(Y, A) = \frac{1}{1-p}, \quad (3.2)$$

$$g(N, A^c) \approx \infty \quad (3.3)$$

가 된다. 다음으로 Flinger et al.의 사생활 보호 측도를 Mangat의 관련질문모형에 적용해 보면

$$\max\{p(A|Y), p(A|N)\} = \frac{\pi}{\pi + (1-\pi)(1-p)}$$

이므로, 다음을 얻을 수 있다.

$$J_1 = \frac{(1-\pi)(1-p)}{(1-\pi)\{\pi + (1-\pi)(1-p)\}} \quad (3.4)$$

따라서, 식 (2.13)과 식 (3.2), 그리고 식 (2.14)와 식 (3.3)으로부터 제안한 무관확률화응답모형이 Mangat의 관련질문모형에 비하여 개인의 사생활을 보호해 주는 측면에서 더욱 효율적임을 알 수 있다. 또한, Flinger et al.의 사생활 보호 측도를 이용하여 비교하여도 마찬가지로 결과를 얻을 수 있다.



#### 4. 결론

본 논문에서는 Mangat의 관련질문모형을 사용하는데 응답자들이 느끼는 거부감을 줄일 수 있는 새로운 무관확률화응답모형을 제안하였다. 이 모형은 민감한 속성을 가지고 있지 않은 응답자에게 “예”라고 직접 응답을 하게 하고, 민감한 속성을 가지고 있는 응답자는 Greenberg et al.의 무관질문모형의 확률장치를 이용하여 선택된 질문에 응답을 하게 된다. 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적이 되는 조건을 제시하였고, 수치적으로 효율성을 비교하였다. 그 결과, 제안한 무관확률화응답모형이  $p$ 값이 작을 경우  $\pi$ 값에 관계 없이 Mangat의 관련질문모형보다 효율적이었고,  $p$ 값이 클 경우에는  $\pi$ 값이 작을 때에 한해서 효율적으로 나타나고 있음을 알 수 있었다. 그리고,  $\pi_y$ 값이 작을 수록 제안한 무관확률화응답모형이 Mangat의 관련질문모형보다 효율적으로 나타났다. 또한, 제안한 무관확률화응답모형이 Mangat의 관련질문모형에 비하여 개인의 사생활을 보호해 주는 측면에서 더 효율적임을 알 수 있었다.

#### 참고문헌

- [1] 류 제복, 홍 기학, 이 기성(1993).〈확률화응답모형〉 자유아카데미, 서울.
- [2] Chang, H. J. and Liang, D. H. (1996). A Two-Stage Unrelated Question Randomized Response Procedure, *Australian Journal of Statistics*, vol. 38(1), 43-51.
- [3] Flinger, M. A., Policello, G. E., and Singh, J. (1977). A Comparison of Two Randomized Response Survey Methods with Consideration for the Level of Respondent Protection, *Communications in Statistics - Theory and Methods*, vol. 6, 1511-1526.
- [4] 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*, vol. 64, 520-539.
- [5] Leysieffer, F. W. and Warner, S. L. (1976). Respondent Jeopardy and Optimal Designs in RR Models, *Journal of the American Statistical Association*, vol. 72, 649-656.
- [6] Mangat, N. S. (1994). An Improved Randomized Response Strategy, *Journal of the Royal Statistical Society : Series B*, vol. 56, 93-95.
- [7] Mangat, N. S. and Singh, R. (1990). An Alternative Randomized Response Procedure, *Biometrika*, vol. 77, 439-442.
- [8] Warner, S. L. (1965). Randomized Response ; A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association*, vol. 60, 63-69.

[ 1998년 5월 접수, 1999년 2월 최종수정 ]

# New Unrelated Question Randomized Response Model

Gi-Sung Lee<sup>1)</sup> Ki-Hak Hong<sup>2)</sup>

## ABSTRACT

In this paper, we proposed a new unrelated question randomized response model that has improved the Mangat's model(1994) in view of the variance under the some conditions. We showed that our model is more protective than the Mangat's model in view of the protection of privacy.

---

1) Department of Computer Science & Statistics, Woosuk University, Wanju-gun, Chonbuk, 565-701, Korea

2) Department of Computer Science, Dongshin University, Daeho-dong, Naju, Chonnam, 520-714, Korea