

개선된 무관질문모형*

이기성¹⁾ 홍기학²⁾

요약

Mangat(1994)는 Mangat-Singh(1990)이 제안한 2단계 관련질문모형의 사용 절차를 좀 더 단순화시킨 개선된 관련질문모형을 제안하였다. 본 논문에서는 Mangat(1994)의 개선된 관련질문모형을 무관질문모형으로 확장하고자 한다. 또한, 제안한 무관질문모형이 Mangat의 개선된 관련질문모형과 Greenberg et al.(1969)의 무관질문모형보다 효율적임을 보였고, 김종호 외(1992) 2인이 제안한 2단계 무관질문모형보다 효율적이 되는 조건을 제시하였다.

1. 서론

민감한 속성 A 를 가지고 있는 모비율 π 를 추정하는데 있어서 좀 더 진실된 자료를 얻기 위하여 Warner(1965)는 확률화응답모형(randomized response model ; RRM)을 처음으로 제시하였다. 그 후 수많은 학자들에 의해 이에 대한 연구가 확대되고 발전되었으며, 특히 Greenberg et al.(1969)은 무관질문모형(unrelated question model)을 제안하였다. 최근에 Mangat-Singh(1990)은 2단계 관련질문모형을 제안하였으며, 김종호 외 2인(1992)은 이 모형을 2단계 무관질문모형으로 발전시켰다. 또한, Mangat(1994)는 Mangat-Singh의 2단계 관련질문모형에서 사용한 2개의 확률장치를 1개로 줄여 그 사용 절차를 좀 더 단순화한 개선된 관련질문모형을 제안하였다.

본 논문에서는 Mangat의 개선된 관련질문모형을 무관질문모형으로 발전시킨 개선된 무관질문모형을 제안하였다. 그리고, 응답자들이 확률장치를 사용함에도 불구하고 완전히 진실된 응답을 하지 않는 경우도 고려하였다. 또한, 제안한 개선된 무관질문모형이 Mangat의 개선된 관련질문모형과 Greenberg et al.의 무관질문모형보다 효율적임을 보였고, 김종호 외 2인이 제안한 2단계 무관질문모형보다 효율적이 되는 조건을 제시하였다.

2. 개선된 무관질문모형

단순임의복원추출된 n 명의 응답자들은 민감한 속성 A 를 가지고 있으면 “예”라고 응답하게 되고, 민감한 속성 A 를 가지고 있지 않으면 확률장치를 이용해서 다음과 같은 2개의 설문 중 선택된 질문에 대하여 응답한다.

설문 1 : 당신은 민감한 속성 A 를 가지고 있습니까?

설문 2 : 당신은 무관한 속성 Y 를 가지고 있습니까?

* 이 논문은 1998년도 우석대학교 학술연구 조성비에 의하여 연구되었음.

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

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

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. 응답자에 의해 이루어지는 이러한 모든 절차를 조사자는 관찰할 수 없다.

따라서, 이러한 개선된 무관질문모형에서 응답자가 “예”라고 응답할 확률을 구해 보면 다음과 같다.

$$\begin{aligned}\lambda &= \pi + (1-\pi)(1-p)\pi_y \\ &= \pi\{1 - (1-p)\pi_y\} + (1-p)\pi_y.\end{aligned}\quad (2.1)$$

여기서, π 는 민감한 속성에 대한 모비율이고, π_y 는 무관한 속성 Y 의 모비율이며, 알고 있다고 가정한다.

이러한 절차에서 “예”라는 응답은 민감한 속성 A 를 가지고 있는 응답자와 무관한 속성 Y 를 가지고 있는 응답자 모두에게 나올 수 있으므로 응답자 자신의 신분이나 프라이버시를 보호받게 된다.

개선된 무관질문모형에서 민감한 속성에 대한 모비율 π 의 최대우도추정량 $\hat{\pi}_{un}$ 는 다음과 같다.

$$\hat{\pi}_{un} = \frac{n'/n - (1-p)\pi_y}{1 - (1-p)\pi_y}. \quad (2.2)$$

여기서, n' 은 단순임의복원추출된 n 명의 응답자들 중에서 “예”라고 응답한 사람의 수이며, n' 는 $b(n, \lambda)$ 를 따른다.

그러나, 추정량 $\hat{\pi}_{un}$ 는 특정구간에서 음수 값을 갖게 되는 제한된 최대우도추정량이므로 이를 보완한 최대우도추정량 $\hat{\pi}_{un}$ 는 다음과 같다.

$$\hat{\pi}_{un} = \begin{cases} 0, & n'/n < (1-p)\pi_y \\ \frac{n'/n - (1-p)\pi_y}{1 - (1-p)\pi_y}, & (1-p)\pi_y \leq n'/n \leq 1 \\ 1, & 1 > n'/n \end{cases} \quad (2.3)$$

한편, 최대우도추정량 $\hat{\pi}_{un}$ 의 기대값은

$$\begin{aligned}E(\hat{\pi}_{un}) &= E\left[\frac{n'/n - (1-p)\pi_y}{1 - (1-p)\pi_y}\right] \\ &= \frac{\lambda - (1-p)\pi_y}{1 - (1-p)\pi_y} \\ &= \frac{\pi\{1 - (1-p)\pi_y\} + (1-p)\pi_y - (1-p)\pi_y}{1 - (1-p)\pi_y} \\ &= \pi\end{aligned}$$

이 되므로 $\hat{\pi}_{un}$ 는 π 의 비편향추정량이다.

정리 2.1 개선된 무관질문모형에서 민감한 속성에 대한 모비율 π 의 추정량 $\hat{\pi}_{un}$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_{un}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)\pi_y}{n\{1 - (1-p)\pi_y\}}. \quad (2.4)$$

증명:

$$\begin{aligned} V(\hat{\pi}_{un}) &= V \left[\frac{n'/n - (1-p)\pi_y}{1 - (1-p)\pi_y} \right] \\ &= \frac{\lambda(1-\lambda)}{n\{1 - (1-p)\pi_y\}^2} \\ &= \frac{[\pi\{1 - (1-p)\pi_y\} + (1-p)\pi_y][1 - \pi\{1 - (1-p)\pi_y\} - (1-p)\pi_y]}{n\{1 - (1-p)\pi_y\}^2} \\ &= \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)\pi_y}{n\{1 - (1-p)\pi_y\}}. \end{aligned}$$

□

다음으로 민감한 속성을 가지고 있는 응답자가 완전히 진실된 응답을 하지 않은 경우에 대하여 살펴보자. 이 때, 민감한 속성을 가지고 있지 않은 응답자들이 거짓으로 응답할 이유는 없다고 가정한다. 따라서, 개선된 무관질문모형의 제안된 절차에 대하여 응답자들이 “예”라고 응답할 확률은 다음과 같다.

$$\lambda_1 = \pi\theta + (1-\pi)(1-p)\pi_y. \tag{2.5}$$

여기서, θ 는 민감한 속성을 가지고 있는 응답자가 진실로 응답할 확률이다.

이 때, 추정량 $\hat{\pi}_{un}$ 의 편의는 다음과 같다.

$$B(\hat{\pi}_{un}) = \frac{\pi(\theta - 1)}{1 - (1-p)\pi_y}. \tag{2.6}$$

따라서, 개선된 무관질문모형의 절차에서 민감한 속성을 가지고 있는 응답자가 진실로 응답하지 않을 경우에 식 (2.6)의 편의를 이용하여 추정량 $\hat{\pi}_{un}$ 의 평균제곱오차를 구해보면 다음과 같다.

$$MSE(\hat{\pi}_{un}) = \frac{\frac{\pi\theta(1-\pi\theta)}{n} + \frac{(1-\pi)(1-p)\pi_y\{1-(1-\pi)(1-p)\pi_y-2\pi\theta\}}{n} + \{\pi(\theta - 1)\}^2}{\{1 - (1-p)\pi_y\}^2}. \tag{2.7}$$

3. 효율성 비교

이 장에서는 제안한 개선된 무관질문모형이 Mangat의 개선된 관련질문모형과 Greenberg et al.의 무관질문모형, 그리고 김종호 외 2인이 제안한 2단계 무관질문모형보다 효율적이 되는 조건을 제시하고자 한다.

개선된 무관질문모형의 분산 식 (2.4)와 Mangat의 개선된 관련질문모형의 분산식을 이용하여 다음과 같은 정리를 얻을 수 있다.

정리 3.1 개선된 무관질문모형의 추정량 $\hat{\pi}_{un}$ 는 Mangat의 개선된 관련질문모형의 추정량 $\hat{\pi}_m$ 보다 항상 효율적이다.

증명: Mangat의 개선된 관련질문모형의 추정량 $\hat{\pi}_m$ 의 분산은

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

이므로, $V(\hat{\pi}_m) - V(\hat{\pi}_{un}) > 0$ 을 이용하여 $V(\hat{\pi}_{un}) < V(\hat{\pi}_m)$ 를 만족하는 조건을 구해보면 $\pi_y < 1$ 임을 알 수 있다. 따라서, 이는 늘 만족되는 조건이므로 개선된 무관질문모형이 Mangat의 개선된 관련질문모형보다 항상 효율적이다. \square

그리고, 개선된 무관질문모형의 분산 식 (2.4)와 김종호 외 2인의 2단계 무관질문모형의 분산식을 이용하여 다음과 같은 정리를 얻을 수 있다.

정리 3.2 개선된 무관질문모형의 추정량 $\hat{\pi}_{un}$ 는 다음과 같은 조건에서 김종호 외 2인의 2단계 무관질문모형의 추정량 $\hat{\pi}_{ut}$ 보다 더 효율적이다.

$$\pi > \frac{[A^2 - B(1-T)^2(1-p)(1-\pi_y) - AB(1-T)]\pi_y}{A[B(1-T)(1-2\pi_y) + A\pi_y]}. \quad (3.1)$$

여기서, $A = p + T(1-p)$, $B = 1 - (1-p)\pi_y$ 이고, T 는 2단계 무관질문모형에서 1단계 확률장치 R_1 으로부터 “당신은 민감한 속성 A 를 가지고 있습니까?”라는 설문 선택될 확률이며, $1-T$ 는 “확률장치 R_2 로 가시오.”라는 설문 선택될 확률이다.

증명: 김종호 외 2인의 2단계 무관질문모형의 추정량 $\hat{\pi}_{ut}$ 의 분산은

$$V(\hat{\pi}_{ut}) = \frac{\pi(1-\pi)}{n} + \frac{(1-T)^2(1-p)^2\pi_y(1-\pi_y) + (1-T)(1-p)\{p + T(1-p)\}(\pi + \pi_y - 2\pi\pi_y)}{n\{p + T(1-p)\}^2}$$

이므로, $V(\hat{\pi}_{ut}) - V(\hat{\pi}_{un}) > 0$ 을 이용하여 $V(\hat{\pi}_{un}) < V(\hat{\pi}_{ut})$ 를 만족하는 조건을 구해보면 식 (3.1)을 얻을 수 있다. \square

개선된 무관질문모형과 2단계 무관질문모형과의 효율성을 수치적으로 비교하기 위하여 n , π , π_y , T , p 를 변화시켜 가면서 분산비를 계산해 본 결과, p 값이 작고 T , π_y 값이 작을 수록 개선된 무관질문모형이 2단계 무관질문모형보다 더 효율이 좋게 나타나고 있었으며, 특히 π 값이 클 때 개선된 무관질문모형이 더욱 효율적임을 알 수 있었다.

또한, 개선된 무관질문모형의 분산 식 (2.4)와 Greenberg et al.의 무관질문모형의 분산식을 이용하여 다음과 같은 정리를 얻을 수 있다.

정리 3.3 개선된 무관질문모형의 추정량 $\hat{\pi}_{un}$ 는 Greenberg et al.의 무관질문모형의 추정량 $\hat{\pi}_u$ 보다 항상 효율적이다.

증명: Greenberg et al.의 무관질문모형의 추정량 $\hat{\pi}_u$ 의 분산은

$$V(\hat{\pi}_u) = \frac{\pi(1-\pi)}{n} + \frac{(1-p)\{p\pi(1-2\pi_y) - (1-p)\pi_y^2 + \pi_y\}}{np^2}$$

이므로, $V(\hat{\pi}_u) - V(\hat{\pi}_{un})$ 를 구해 보면 다음과 같다.

$$\begin{aligned} V(\hat{\pi}_u) - V(\hat{\pi}_{un}) &= \frac{(1-p)\{p\pi(1-2\pi_y) - (1-p)\pi_y^2 + \pi_y\}}{np^2} - \frac{(1-\pi)(1-p)\pi_y}{n\{1-(1-p)\pi_y\}} \\ &= \frac{(1-p)\{[1-(1-p)\pi_y]\{p\pi(1-2\pi_y) - (1-p)\pi_y^2 + \pi_y\} - p^2(1-\pi)\pi_y\}}{np^2\{1-(1-p)\pi_y\}} \\ &= \frac{(1-p)[Bp\pi(1-2\pi_y) - \pi_y B^2 - p^2(1-\pi)\pi_y]}{np^2 B} \\ &> 0 \end{aligned}$$

여기서, $B = 1 - (1-p)\pi_y$ 이다.

따라서, $V(\hat{\pi}_{un}) < V(\hat{\pi}_u)$ 라는 관계가 성립되므로 개선된 무관질문모형이 Greenberg et al.의 무관질문모형보다 항상 효율적임을 알 수 있다. \square

4. 결론

본 논문에서는 Mangat의 개선된 관련질문모형을 무관질문모형으로 발전시킨 개선된 무관질문모형을 제안하였으며, 응답자들이 확률장치를 사용함에도 불구하고 완전히 진실된 응답을 하지 않는 경우도 고려하였다. 또한, 제안한 개선된 무관질문모형이 Mangat의 개선된 관련질문모형과 Greenberg et al.의 무관질문모형보다 항상 효율적임을 보였고, 김종호 외 2인이 제안한 2단계 무관질문모형보다 효율적이 되는 조건을 제시하였다.

참고문헌

- [1] 류 제복, 홍 기학, 이 기성(1993). <확률화응답모형>, 자유아카데미, 서울.
- [2] 김 종호, 류 제복, 이 기성(1992). 새로운 2단계 확률화응답모형, <응용통계연구>, 제5권 2호, 157-167.
- [3] Greenberg, B. G., Abul-Ela, Abdel-Elatif 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.
- [4] Mangat, N. S. (1994). An Improved Randomized Response Strategy, *Journal of the Royal Statistical Society : Series B*, Vol. 56, 93-95.
- [5] Mangat, N. S. and Singh, R. (1990). An Alternative Randomized Response Procedure, *Biometrika*, Vol. 77, 439-442.

- [6] 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 3월 접수, 1998년 5월 최종수정]

Improved Unrelated Question Model*

Gi-Sung Lee¹⁾ Ki-Hak Hong²⁾

ABSTRACT

In this paper, we proposed improved unrelated question model which has the benefit of simplicity the Kim et al.'s two-stage unrelated question model(1992). Conditions are obtained under which the proposed model is more efficient than the Greenberg et al.'s model(1971) and Kim et al.'s two-stage unrelated question model.

* This paper was supported by academic research fund of woosuk university, 1998.

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