응용통계연구 제 13권 2호, 2000년, pp. 575-590

2단계 이표본 무관질문모형

이기성[1]) 홍기학[2])

요 약

Greenberg et al.(1969)은 무관질문모형에서 무관한 속성이 미지인 경우 두 개의 독립표본을 이용하여 민감한 속성에 대한 모비율을 추정해 내는 이표본 무관질문모형을 제안하였다. 본 논문에서는 Mangat-Singh(1990)의 모형을 개선한 형태의 김종호 외 2인(1992)이 제안한 2단계 무관질문모형과 이기성과 홍기학(1998)이 제안한 개선된 무관질문모형을 무관한 속성이 미지일 때 두 개의 독립표본을 이용하는 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형으로 확장하였다. 그리고, Greenberg et al.의 모형과 2단계 이표본 무관질문모형, 그리고 개선된 이표본 무관질문모형과의 효율성을 비교하였다.

주요용어: 2단계 무관질문모형, 개선된 무관질문모형, 이표본 무관질문모형.

1. 서론

응답자들의 신분이나 비밀을 보장함으로써 응답자들로부터 민감한 질문에 보다 더 정확한 정보를 얻을 수 있는 확률화응답모형(randomized response model ; RRM)은 Warner(1965) 이후 수많은 학자들에 의해 연구되고 발전되었다. 특히, Greenberg et al.(1969)은 무관질문모형(unrelated question model)을 제안하여 무관한 속성이 미지인 경우 두 개의 독립표본을 이용하는 이표본 무관질문모형(two sample unrelated question model)의 이론적 체계를 완성하였으며, Moors(1971)와 Folsom et al.(1973)은 이를 개선 보완하였다.

최근에 Mangat-Singh(1990)은 2단계 관련질문모형을 제안하였으며, 김종호 외 2인 (1992)은 이 모형을 2단계 무관질문모형과 강요질문모형으로 발전시켰다. 그리고, Mangat(1994)는 Mangat-Singh의 2단계 관련질문모형에서 사용한 2개의 확률장치를 1개로 줄여 그 사용 절차를 좀 더 단순화 한 개선된 관련질문모형을 제안하였다. 또한, 이기성과 홍기학(1998)은 Mangat의 절차를 이용하여 2단계 무관질문모형의 사용절차를 간소화 한 개선된 무관질문모형을 제안하였다.

한편, 김종호 외 2인이 제안한 2단계 무관질문모형과 이기성과 홍기학이 제안한 개선된 무관질문모형에서는 무관한 속성에 대한 모비율을 알고 있다고 가정하였으나, 일반적으로는 무관한 속성의 모비율이 알려져 있지 않는 경우가 많다.

1) (565-701) 전북 완주군 삼례읍 후정리 490, 우석대학교 전산통계학과, 부교수
E-mail: gisung@core.woosuk.ac.kr
2) (520-714) 전남 나주시 대호동 252, 동신대학교 컴퓨터학과, 부교수
E-mail: khhong@blue.dongshinu.ac.kr

따라서, 본 논문에서는 2단계 무관질문모형과 개선된 무관질문모형을 무관한 속성이 미지일 때 두 개의 독립표본을 이용하는 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형으로 확장하였다. 그리고, Greenberg et al.의 모형과 2단계 이표본 무관질문모형, 그리고 개선된 이표본 무관질문모형과의 효율성을 비교하였다.

2. 2단계 이표본 무관질문모형

민감한 질문과 전혀 관련이 없는 무관한 질문을 사용하여 민감한 속성에 대한 모비율 π를 추정하는 무관질문모형에서 무관한 속성에 대한 모비율 π_y가 알려져 있지 않은 경우가 많다. 따라서, π_y가 미지일 때 π를 추정하기 위해서는 최소한 두 개의 표본이 필요하게 되므로, 모집단으로부터 단순임의복원으로 크기가 n_1과 n_2인 두 개의 독립표본을 추출하여야 한다. 이 장에서는 Mangat-Singh의 관련질문모형을 발전시킨 2단계 무관질문모형과 2단계 무관질문모형의 사용절차를 단순화시킨 개선된 무관질문모형을 무관한 속성에 대한 모비율 π_y가 미지일 때 두 개의 표본을 이용하는 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형으로 확장해 보고자 한다.

2.1. 2단계 이표본 무관질문모형

단순임의복원추출된 서로 독립인 $n_i(i = 1, 2)$명의 응답자들은 다음과 같은 2개의 설문으로 구성된 확률장치 $R_1(i)$와 $R_2(i)$를 이용하게 된다.

<1단계의 확률장치 $R_1(i)$ >

	설문내용	확률
설문 1	당신은 민감한 속성 A를 가지고 있습니까?	T
설문 2	확률장치 $R_2(i)$로 가시오.	$1-T$

<2단계의 확률장치 $R_2(i)$ >

	설문내용	확률
설문 1	당신은 민감한 속성 A를 가지고 있습니까?	p_i
설문 2	당신은 무관한 속성 Y를 가지고 있습니까?	$1-p_i$

여기서, $T \neq 1$이고 $p_1 \neq p_2$ (즉, $(1-T)(p_1-p_2) \neq 0$)라고 가정한다.

응답자들은 확률장치에 의해서 선택된 설문에 대해 "예" 또는 "아니오"라고 응답하게 되며, 응답자들이 진실되게 응답한다고 가정하면 표본 i에서 "예"라고 응답할 확률은 다음과 같다.

$$\begin{aligned}\lambda_i &= T\pi + (1-T)\{p_i\pi + (1-p_i)\pi_y\} \\ &= \{T + (1-T)p_i\}\pi + (1-T)(1-p_i)\pi_y. \qquad (2.1)\end{aligned}$$

n_i명의 응답자 중에서 "예"라고 응답한 사람의 수를 n_{i1}이라 하면, 2단계 이표본 무관질문모형의 결합우도함수는

$$L \propto \lambda_1^{n_{11}}(1-\lambda_1)^{n_1-n_{11}}\lambda_2^{n_{21}}(1-\lambda_2)^{n_2-n_{21}}$$

이며, 로그우도함수는 다음과 같다.

$$logL = n_{11}log\lambda_1 + (n_1 - n_{11})log(1-\lambda_1) + n_{21}log\lambda_2 + (n_2 - n_{21})log(1-\lambda_2).$$

따라서, $\frac{\partial logL}{\partial \pi} = 0$와 $\frac{\partial logL}{\partial \pi_y} = 0$로부터 다음과 같은 두 우도방정식을 얻을 수 있다.

$$\{T + (1-T)p_1\}\frac{n_{11} - n_1\lambda_1}{\lambda_1(1-\lambda_1)} + \{T + (1-T)p_2\}\frac{n_{21} - n_2\lambda_2}{\lambda_2(1-\lambda_2)} = 0, \tag{2.2}$$

$$(1-T)(1-p_1)\frac{n_{11} - n_1\lambda_1}{\lambda_1(1-\lambda_1)} + (1-T)(1-p_2)\frac{n_{21} - n_2\lambda_2}{\lambda_2(1-\lambda_2)} = 0. \tag{2.3}$$

이 때, $\begin{vmatrix} T + (1-T)p_1 & T + (1-T)p_2 \\ (1-T)(1-p_1) & (1-T)(1-p_2) \end{vmatrix} = (1-T)(p_1 - p_2) \neq 0$이기 때문에, 식 (2.2)와 식 (2.3)의 두 우도방정식은 다음과 같은 두 조건을 만족할 때 해를 갖게 된다.

$$\frac{n_{11} - n_1\lambda_1}{\lambda_1(1-\lambda_1)} = 0, \qquad \frac{n_{21} - n_2\lambda_2}{\lambda_2(1-\lambda_2)} = 0. \tag{2.4}$$

식 (2.4)에서 λ_1과 λ_2가 0이나 1의 값을 갖지 않는다고 가정하면, 식 (2.1)의 λ_1과 λ_2를 식 (2.4)에 대입하여 다음과 같은 두 식을 얻을 수 있다.

$$\{T + (1-T)p_1\}\pi + (1-T)(1-p_1)\pi_y = \frac{n_{11}}{n_1}, \tag{2.5}$$

$$\{T + (1-T)p_2\}\pi + (1-T)(1-p_2)\pi_y = \frac{n_{21}}{n_2}. \tag{2.6}$$

따라서, 식 (2.5)와 식 (2.6)을 이용하여 π에 대한 최우추정량 $\hat{\pi}_{u_1}$를 구해 보면 다음과 같다.

$$\hat{\pi}_{u_1} = \frac{\frac{n_{11}}{n_1}(1-p_2) - \frac{n_{21}}{n_2}(1-p_1)}{p_1 - p_2}, \qquad p_1 \neq p_2. \tag{2.7}$$

정리 2.1 추정량 $\hat{\pi}_{u_1}$는 모비율 π의 비편향추정량이다.

증명:

$$
\begin{aligned}
E(\hat{\pi}_{u_1}) &= E\left[\frac{\frac{n_{11}}{n_1}(1-p_2)-\frac{n_{21}}{n_2}(1-p_1)}{p_1-p_2}\right] \\
&= \frac{\lambda_1(1-p_2)-\lambda_2(1-p_1)}{p_1-p_2} \\
&= \frac{\Big[T\pi+(1-T)\{p_1\pi+(1-p_1)\pi_y\}\Big](1-p_2)}{p_1-p_2} \\
&\quad -\frac{\Big[T\pi+(1-T)\{p_2\pi+(1-p_2)\pi_y\}\Big](1-p_1)}{p_1-p_2} \\
&= \pi.
\end{aligned}
$$

□

정리 2.2 추정량 $\hat{\pi}_{u_1}$의 분산은 다음과 같다.

$$
\begin{aligned}
V(\hat{\pi}_{u_1}) &= \frac{(1-p_2)^2}{n_1(p_1-p_2)^2}\Big[\pi(1-\pi)\{T+(1-T)p_1\}^2+(1-T)^2(1-p_1)^2\pi_y(1-\pi_y) \\
&\quad +(1-T)(1-p_1)\{T+(1-T)p_1\}(\pi+\pi_y-2\pi\pi_y)\Big] \\
&\quad +\frac{(1-p_1)^2}{n_2(p_1-p_2)^2}\Big[\pi(1-\pi)\{T+(1-T)p_2\}^2+(1-T)^2(1-p_2)^2\pi_y(1-\pi_y) \\
&\quad +(1-T)(1-p_2)\{T+(1-T)p_2\}(\pi+\pi_y-2\pi\pi_y)\Big]. \qquad (2.8)
\end{aligned}
$$

증명:

$$
\begin{aligned}
V(\hat{\pi}_{u_1}) &= V\left[\frac{\frac{n_{11}}{n_1}(1-p_2)-\frac{n_{21}}{n_2}(1-p_1)}{p_1-p_2}\right] \\
&= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1}+\frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right] \\
&= \frac{(1-p_2)^2}{n_1(p_1-p_2)^2}\Big[\pi(1-\pi)\{T+(1-T)p_1\}^2+(1-T)^2(1-p_1)^2\pi_y(1-\pi_y) \\
&\quad +(1-T)(1-p_1)\{T+(1-T)p_1\}(\pi+\pi_y-2\pi\pi_y)\Big] \\
&\quad +\frac{(1-p_1)^2}{n_2(p_1-p_2)^2}\Big[\pi(1-\pi)\{T+(1-T)p_2\}^2+(1-T)^2(1-p_2)^2\pi_y(1-\pi_y) \\
&\quad +(1-T)(1-p_2)\{T+(1-T)p_2\}(\pi+\pi_y-2\pi\pi_y)\Big].
\end{aligned}
$$

□

한편, 분산추정량 $\hat{V}(\hat{\pi}_{u_1})$는 다음과 같다.

$$\hat{V}(\hat{\pi}_{u_1}) = \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1-1} + \frac{(1-p_1)^2\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2-1}\right]. \tag{2.9}$$

여기서, $\hat{\lambda}_1 = n_{11}/n_1$이고 $\hat{\lambda}_2 = n_{21}/n_2$이다.

정리 2.3 분산추정량 $\hat{V}(\hat{\pi}_{u_1})$는 $V(\hat{\pi}_{u_1})$의 비편향추정량이다.

증명:

$$E\Big[\hat{V}(\hat{\pi}_{u_1})\Big] = \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2}{n_1-1}E[\hat{\lambda}_1(1-\hat{\lambda}_1)] + \frac{(1-p_1)^2}{n_2-1}E[\hat{\lambda}_2(1-\hat{\lambda}_2)]\right]$$

에서

$$\begin{aligned}
E[\hat{\lambda}_1(1-\hat{\lambda}_1)] &= E(\hat{\lambda}_1) - E(\hat{\lambda}_1^2) \\
&= E\left(\frac{n_{11}}{n_1}\right) - E\left(\frac{{n_{11}}^2}{{n_1}^2}\right) \\
&= \frac{1}{n_1}n_1\lambda_1 - \frac{1}{{n_1}^2}\Big[n_1\lambda_1(1-\lambda_1) + {n_1}^2{\lambda_1}^2\Big] \\
&= \frac{n_1-1}{n_1}\lambda_1(1-\lambda_1)
\end{aligned}$$

이고,

$$E[\hat{\lambda}_2(1-\hat{\lambda}_2)] = \frac{n_2-1}{n_2}\lambda_2(1-\lambda_2)$$

이므로,

$$\begin{aligned}
E\Big[\hat{V}(\hat{\pi}_{u_1})\Big] &= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2}{n_1-1}\frac{n_1-1}{n_1}\lambda_1(1-\lambda_1) + \frac{(1-p_1)^2}{n_2-1}\frac{n_2-1}{n_2}\lambda_2(1-\lambda_2)\right] \\
&= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1} + \frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right] \\
&= V(\hat{\pi}_{u_1})
\end{aligned}$$

이 된다. □

2.2. 개선된 이표본 무관질문모형

단순임의복원추출된 서로 독립인 $n_i(i=1,2)$명의 응답자들은 민감한 속성 A를 가지고 있으면

당신은 민감한 속성 A를 가지고 있습니까?

라는 설문에 대하여 직접 "예"라고 응답하게 되고, 민감한 속성 A를 가지고 있지 않으면 다음과 같은 2개의 설문으로 구성된 확률장치 $R(i)$를 이용하게 된다.

<확률장치 $R(i)$ >

	설문내용	확률
설문 1	당신은 민감한 속성 A를 가지고 있습니까?	p_i
설문 2	당신은 무관한 속성 Y를 가지고 있습니까?	$1-p_i$

여기서, $p_1 \neq p_2$라고 가정한다.

응답자들은 확률장치에 의해서 선택된 설문에 대해 "예" 또는 "아니오"라고 응답하게 되고, 응답자에 의해 이루어지는 이러한 모든 절차를 조사자는 관찰할 수 없으며, "예"라는 응답은 민감한 속성 A를 가지고 있는 응답자와 무관한 속성 Y를 가지고 있는 응답자 모두에게 나올 수 있으므로 응답자 자신의 신분이나 프라이버시를 보호받게 된다.

따라서, 응답자들이 진실되게 응답한다고 가정하면 표본 i에서 "예"라고 응답할 확률은 다음과 같다.

$$\begin{aligned}\lambda_i &= \pi + (1-\pi)(1-p_i)\pi_y \\ &= \{1-(1-p_i)\pi_y\}\pi + (1-p_i)\pi_y.\end{aligned} \tag{2.10}$$

n_i명의 응답자 중에서 "예"라고 응답한 사람의 수를 n_{i1}이라 하고, 개선된 이표본 무관질문모형에서 민감한 속성에 대한 모비율 π의 최우추정량 $\hat{\pi}_{u_2}$를 구해보면 다음과 같다.

$$\hat{\pi}_{u_2} = \frac{\frac{n_{11}}{n_1}(1-p_2) - \frac{n_{21}}{n_2}(1-p_1)}{p_1-p_2}, \quad p_1 \neq p_2. \tag{2.11}$$

정리 2.4 추정량 $\hat{\pi}_{u_2}$는 모비율 π의 비편향추정량이다.

증명:

$$\begin{aligned}E(\hat{\pi}_{u_2}) &= E\left[\frac{\frac{n_{11}}{n_1}(1-p_2) - \frac{n_{21}}{n_2}(1-p_1)}{p_1-p_2}\right] \\ &= \frac{\lambda_1(1-p_2) - \lambda_2(1-p_1)}{p_1-p_2} \\ &= \frac{\left[\pi\{1-(1-p_1)\pi_y\} + (1-p_1)\pi_y\right](1-p_2)}{p_1-p_2} \\ &\quad - \frac{\left[\pi\{1-(1-p_2)\pi_y\} + (1-p_2)\pi_y\right](1-p_1)}{p_1-p_2} \\ &= \pi.\end{aligned}$$

□

정리 2.5 추정량 $\hat{\pi}_{u_2}$의 분산은 다음과 같다.

$$\begin{aligned} V(\hat{\pi}_{u_2}) &= \frac{(1-p_2)^2}{n_1(p_1-p_2)^2}\Big[\pi(1-\pi)\{1-(1-p_1)\pi_y\}^2 \\ &\quad +\{(1-\pi)(1-p_1)\pi_y\}\{1-(1-p_1)\pi_y\}\Big] \\ &\quad +\frac{(1-p_1)^2}{n_2(p_1-p_2)^2}\Big[\pi(1-\pi)\{1-(1-p_2)\pi_y\}^2 \\ &\quad +\{(1-\pi)(1-p_2)\pi_y\}\{1-(1-p_2)\pi_y\}\Big]. \end{aligned} \tag{2.12}$$

증명:

$$\begin{aligned} V(\hat{\pi}_{u_2}) &= V\left[\frac{\frac{n_{11}}{n_1}(1-p_2)-\frac{n_{21}}{n_2}(1-p_1)}{p_1-p_2}\right] \\ &= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1}+\frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right] \\ &= \frac{(1-p_2)^2}{n_1(p_1-p_2)^2}\Big[\pi(1-\pi)\{1-(1-p_1)\pi_y\}^2+\{(1-\pi)(1-p_1)\pi_y\}\{1-(1-p_1)\pi_y\}\Big] \\ &\quad +\frac{(1-p_1)^2}{n_2(p_1-p_2)^2}\Big[\pi(1-\pi)\{1-(1-p_2)\pi_y\}^2+\{(1-\pi)(1-p_2)\pi_y\}\{1-(1-p_2)\pi_y\}\Big]. \end{aligned}$$

□

한편, 분산추정량 $\hat{V}(\hat{\pi}_{u_2})$는 다음과 같다.

$$\hat{V}(\hat{\pi}_{u_2}) = \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1-1}+\frac{(1-p_1)^2\hat{\lambda}_2(1-\hat{\lambda}_2)}{n_2-1}\right]. \tag{2.13}$$

여기서, $\hat{\lambda}_1 = n_{11}/n_1$이고 $\hat{\lambda}_2 = n_{21}/n_2$이다.

정리 2.6 분산추정량 $\hat{V}(\hat{\pi}_{u_2})$는 $V(\hat{\pi}_{u_2})$의 비편향추정량이다.

증명:

$$\begin{aligned} E[\hat{V}(\hat{\pi}_{u_2})] &= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2}{n_1-1}E[\hat{\lambda}_1(1-\hat{\lambda}_1)]+\frac{(1-p_1)^2}{n_2-1}E[\hat{\lambda}_2(1-\hat{\lambda}_2)]\right] \\ &= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2}{n_1-1}\frac{n_1-1}{n_1}\lambda_1(1-\lambda_1)+\frac{(1-p_1)^2}{n_2-1}\frac{n_2-1}{n_2}\lambda_2(1-\lambda_2)\right] \\ &= \frac{1}{(p_1-p_2)^2}\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1}+\frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right] \\ &= V(\hat{\pi}_{u_2}). \end{aligned}$$

□

3. 효율성 비교

이 장에서는 Greenberg et al.의 이표본 무관질문모형과 2단계 이표본 무관질문모형 그리고 개선된 이표본 무관질문모형과의 효율성을 비교해 보고자 한다.

Greenberg et al.의 이표본 무관질문모형에서 민감한 속성에 대한 모비율 π의 추정량 $\hat{\pi}_u$의 분산은 다음과 같다.

$$V(\hat{\pi}_u) = \frac{1}{(p_1 - p_2)^2}\left[\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{n_1} + \frac{(1-p_1)^2\lambda_2(1-\lambda_2)}{n_2}\right]. \tag{3.1}$$

여기서, $\lambda_i(i=1,2)$는 표본 i에서 "예"라고 응답할 확률로 다음과 같다.

$$\lambda_i = p_i\pi + (1-p_i)\pi_y. \tag{3.2}$$

비교하고자 하는 각 모형들의 추정량의 분산은 p_1과 p_2, π_y 그리고 n_1과 n_2의 선택에 따라 그 크기가 달라진다. 이 때, p_1과 p_2의 선택에서 Greenberg et al.이 제시한 $p_1 + p_2 = 1$의 조건보다 $p_2 = 0$의 조건이 추정량의 분산을 줄일 수 있으므로 더 바람직하다는 Moors의 제안을 따르기로 한다. 또한, Greenberg et al.이 코쉬-쉬바르쯔 부등식을 이용하여 구한 n_1과 n_2의 최적배분은 다음과 같다.

$$\frac{n_1}{n_2} = \sqrt{\frac{(1-p_2)^2\lambda_1(1-\lambda_1)}{(1-p_1)^2\lambda_2(1-\lambda_2)}}. \tag{3.3}$$

이러한 조건들을 따르면, Greenberg et al.의 이표본 무관질문모형의 분산 식 (3.1)은 다음과 같이 표현된다.

$$V(\hat{\pi}_u) = \frac{1}{n{p_1}^2}\left[\sqrt{\lambda_1(1-\lambda_1)} + (1-p_1)\sqrt{\lambda_2(1-\lambda_2)}\right]^2. \tag{3.4}$$

여기서,

$$\begin{aligned}
\lambda_1(1-\lambda_1) &= \pi(1-\pi){p_1}^2 + (1-p_1)\Big[p_1(1-2\pi_y)\pi + \pi_y\{1-(1-p_1)\pi_y\}\Big],\\
\lambda_2(1-\lambda_2) &= \pi_y(1-\pi_y)
\end{aligned}$$

이다.

그리고, 동일한 조건하에서 2단계 이표본 무관질문모형의 분산 식 (2.8)은 다음과 같은 분산식으로 표현된다.

$$V(\hat{\pi}_{u_1}) = \frac{1}{n{p_1}^2}\left[\sqrt{\lambda_1(1-\lambda_1)} + (1-p_1)\sqrt{\lambda_2(1-\lambda_2)}\right]^2. \tag{3.5}$$

여기서,

$$\begin{aligned}
\lambda_1(1-\lambda_1) &= \pi(1-\pi)\{T+(1-T)p_1\}^2 + (1-T)^2(1-p_1)^2\pi_y(1-\pi_y)\\
&\quad + (1-T)(1-p_1)\{T+(1-T)p_1\}(\pi+\pi_y-2\pi\pi_y),\\
\lambda_2(1-\lambda_2) &= \pi(1-\pi)T^2 + (1-T)^2\pi_y(1-\pi_y) + (1-T)T(\pi+\pi_y-2\pi\pi_y)
\end{aligned}$$

이다.

마찬가지로, 동일한 조건하에서 개선된 이표본 무관질문모형의 분산 식 (2.12)는 다음과 같은 분산식으로 표현된다.

$$V(\hat{\pi}_{u_2}) = \frac{1}{n{p_1}^2}\left[\sqrt{\lambda_1(1-\lambda_1)} + (1-p_1)\sqrt{\lambda_2(1-\lambda_2)}\right]^2. \tag{3.6}$$

여기서,

$$\lambda_1(1-\lambda_1) = \pi(1-\pi)\{1-(1-p_1)\pi_y\}^2 + \{(1-\pi)(1-p_1)\pi_y\}\{1-(1-p_1)\pi_y\},$$
$$\lambda_2(1-\lambda_2) = \pi(1-\pi)(1-\pi_y)^2 + (1-\pi)\pi_y(1-\pi_y)$$

이다.

3.1. Greenberg et al.의 이표본 무관질문모형과 2단계 이표본 무관질문모형과의 효율성 비교

식 (3.4)와 식 (3.5)로부터 $V(\hat{\pi}_{u_1}) < V(\hat{\pi}_u)$ 를 만족하는 조건을 구해보면 다음과 같다.

$$\pi < \frac{B + \sqrt{B^2 + 4AC}}{2A}. \tag{3.7}$$

여기서,

$$A = {p_1}^2 - \{T + (1-T)p_1\}^2 - T^2,$$
$$B = 2\pi_y\Big[\{T + (1-T)p_1\}(1-T)(1-p_1) + T(1-T) - p_1(1-p_1)\Big]$$
$$\quad + \{p_1 - 2T - (1-T)p_1\},$$
$$C = (1-T)(1-p_1)\pi_y\{(1-T)(1-p_1)\pi_y - 1\} + (1-T)\pi_y\{(1-T)\pi_y - 1\}$$
$$\quad + (1-p_1)\pi_y\{1-(1-p_1)\pi_y\} + \pi_y(1-\pi_y)$$

이다.

또한, 식 (3.4)와 식 (3.5)를 이용하여 Greenberg et al.의 이표본 무관질문모형과 2단계 이표본 무관질문모형과의 효율성을 수치적으로 비교하기 위하여 $n = 100$, $T = 0.3$일 때 π와 π_y 및 p_1을 변화시켜 가면서 분산비 $V(\hat{\pi}_u)/V(\hat{\pi}_{u_1})$를 계산한 결과 다음 표 3.1을 얻었다.

표 3.1에서 1보다 큰 값은 2단계 이표본 무관질문모형이 Greenberg et al.의 이표본 무관질문모형보다 더 효율적임을 나타낸다. 표 3.1에 나타난 결과들은 상대적으로 p_1값 보다는 π값이나 π_y값에 더 많은 영향을 받고 있는 것으로 나타나고 있는데, 이는 식 (3.4)와 식 (3.5)의 $\lambda_2(1-\lambda_2)$의 값이 π값과 π_y값으로 이루어져 있으며, p_1값에 의존하고 있지 않기 때문이다. 표 3.1로부터 2단계 이표본 무관질문모형은 π값이 0.1로 작은 경우에 π_y값이 0.9인 경우를 제외하고는 p_1값에 관계없이 Greenberg et al.의 이표본 무관질문모형보다 더 효율이 좋게 나타나고 있었다. 그리고, $\pi_y = 0.5$일 때는 π값과 p_1값에 관계없이 Greenberg et al.의 이표본 무관질문모형보다 그 효율성이 높으므로 2단계 이표본 무관질문모형을 사용

표 3.1: Greenberg et al.의 이표본 무관질문모형과
2단계 이표본 무관질문모형과의 효율성 비교

π	p_1 \ π_y	0.1	0.3	0.5	0.7	0.9
0.1	0.1	1.0000	1.1519	1.0756	0.8840	0.5107
	0.3	1.0000	1.1497	1.1006	0.9575	0.6760
	0.5	1.0000	1.1402	1.1182	1.0189	0.8099
	0.7	1.0000	1.1159	1.1202	1.0656	0.9284
	0.9	1.0000	1.0577	1.0764	1.0684	1.0194
0.2	0.1	0.8128	1.0637	1.0411	0.8810	0.5192
	0.3	0.8473	1.0587	1.0531	0.9372	0.6694
	0.5	0.8839	1.0506	1.0593	0.9830	0.7905
	0.7	0.9246	1.0375	1.0550	1.0157	0.8957
	0.9	0.9722	1.0160	1.0292	1.0216	0.9803
0.3	0.1	0.7040	1.0000	1.0178	0.8858	0.5337
	0.3	0.7676	1.0000	1.0226	0.9277	0.6695
	0.5	0.8288	1.0000	1.0244	0.9625	0.7803
	0.7	0.8919	1.0000	1.0216	0.9893	0.8781
	0.9	0.9612	1.0000	1.0106	1.0026	0.9645
0.4	0.1	0.6346	0.9537	1.0043	0.8988	0.5557
	0.3	0.7209	0.9620	1.0055	0.9282	0.6767
	0.5	0.7991	0.9710	1.0058	0.9542	0.7779
	0.7	0.8757	0.9811	1.0050	0.9766	0.8699
	0.9	0.9563	0.9930	1.0023	0.9941	0.9575
0.5	0.1	0.5879	0.9209	1.0000	0.9209	0.5879
	0.3	0.6928	0.9391	1.0000	0.9391	0.6928
	0.5	0.7836	0.9568	1.0000	0.9568	0.7836
	0.7	0.8691	0.9741	1.0000	0.9741	0.8691
	0.9	0.9552	0.9913	1.0000	0.9913	0.9552

할 때에는 π_y값이 0.5가 되도록 해 주는 것이 가장 바람직함을 알 수 있었다. 또한, $\pi = \pi_y$인 경우에 두 모형의 효율성은 p_1값의 영향을 받지 않고 동일한 것으로 나타나는데 이는 식 (2.1)과 식 (3.2)가 같은 값을 갖게되어 두 모형의 분산이 각각 서로 같아지기 때문이다. 한편, T값을 0.1에서 0.9까지 변화시켜 가면서 효율성을 비교해 본 결과 전반적으로 표 3.1과 비슷한 경향을 나타내고 있었으나, T값이 클수록 2단계 이표본 무관질문모형의 효율이 좋게 나타나는 결과를 얻을 수 있었다. 이러한 결과는 T값이 1에 가까워질수록 직접질문의 형태를 취하기 때문에 그 만큼 분산이 작아지기 때문이다.

3.2. Greenberg et al.의 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성 비교

식 (3.4)와 식 (3.6)으로부터 $V(\hat{\pi}_{u_2}) < V(\hat{\pi}_u)$를 만족하는 조건을 구해보면 다음과 같다.

$$\pi < \frac{B}{A}. \tag{3.8}$$

표 3.2: Greenberg et al.의 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성 비교

π	π_y / p_1	0.1	0.3	0.5	0.7	0.9
0.1	0.1	0.5963	0.8976	1.0014	1.0501	1.0644
	0.3	0.6271	0.8933	0.9867	1.0281	1.0363
	0.5	0.6736	0.8939	0.9754	1.0121	1.0216
	0.7	0.7489	0.9050	0.9688	0.9999	1.0112
	0.9	0.8845	0.9458	0.9759	0.9935	1.0026
0.2	0.1	0.4712	0.8497	1.0253	1.1187	1.1456
	0.3	0.5295	0.8494	0.9983	1.0717	1.0823
	0.5	0.6056	0.8584	0.9786	1.0381	1.0495
	0.7	0.7125	0.8837	0.9686	1.0133	1.0264
	0.9	0.8776	0.9432	0.9781	0.9987	1.0078
0.3	0.1	0.4224	0.8403	1.0750	1.2127	1.2508
	0.3	0.5010	0.8457	1.0352	1.1354	1.1419
	0.5	0.5943	0.8608	1.0055	1.0799	1.0855
	0.7	0.7145	0.8918	0.9874	1.0380	1.0462
	0.9	0.8837	0.9512	0.9876	1.0083	1.0143
0.4	0.1	0.4103	0.8657	1.1582	1.3439	1.3917
	0.3	0.5074	0.8766	1.1028	1.2275	1.2213
	0.5	0.6136	0.8942	1.0582	1.1423	1.1332
	0.7	0.7388	0.9221	1.0243	1.0755	1.0721
	0.9	0.8987	0.9670	1.0034	1.0220	1.0226
0.5	0.1	0.4256	0.9313	1.2906	1.5338	1.5895
	0.3	0.5431	0.9473	1.2139	1.3633	1.3321
	0.5	0.6610	0.9626	1.1458	1.2353	1.1990
	0.7	0.7859	0.9776	1.0842	1.1312	1.1072
	0.9	0.9239	0.9925	1.0271	1.0413	1.0334

여기서,

$$A = {p_1}^2 - \{1 - (1 - p_1)\pi_y\}^2 - (1 - \pi_y)^2,$$
$$B = 2\pi_y(1 - p_1)[\{1 - (1 - p_1)\pi_y\} - p_1] + (1 - \pi_y)(p_1 + 2\pi_y - 2)$$

이다.

또한, 식 (3.4)와 식 (3.6)을 이용하여 Greenberg et al.의 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성을 수치적으로 비교하기 위하여 $n = 100$일 때 π와 π_y 및 p_1을 변화시켜 가면서 분산비 $V(\hat{\pi}_u)/V(\hat{\pi}_{u_2})$를 계산한 결과 다음 표 3.2를 얻었다.

표 3.2에서 1보다 큰 값은 개선된 이표본 무관질문모형이 Greenberg et al.의 이표본 무관질문모형보다 더 효율적임을 나타낸다. 표 3.2에 나타난 결과들은 상대적으로 p_1값보다는 π값이나 π_y값에 더 많은 영향을 받고 있는 것으로 나타나고 있는데, 이는 식 (3.4)와 식 (3.6)의 $\lambda_2(1 - \lambda_2)$의 값이 π값과 π_y값으로 이루어져 있으며, p_1값에 의존하고 있지 않기 때문이다. 표 3.2로부터 개선된 이표본 무관질문모형은 π값이 0.5보다 큰 경우에 p_1값이 작을수록 π_y값이 클수록 Greenberg et al.의 이표본 무관질문모형보다 더 효율이 좋게 나타나고 있었다. 특히, $\pi_y = 0.9$일 때는 π값과 p_1값에 관계없이 Greenberg et al.의 이표본 무관질문모형보다 그 효율성이 높으므로 개선된 이표본 무관질문모형을 사용할 때에는 π_y값이 0.9가 되도록 해 주는 것이 가장 바람직함을 알 수 있었다. 이 결과는 π_y값이 클 때에는 식 (3.4)와 식 (3.6)에 포함되어 있는 $\lambda_i(1 - \lambda_i)$값이 Greenberg et al.의 이표본 무관질문모형에 비해 개선된 이표본 무관질문모형이 더 작아지게 되어 개선된 이표본 무관질문모형의 분산이 작아지기 때문이다.

3.3. 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성 비교

식 (3.5)와 식 (3.6)으로부터 $V(\hat{\pi}_{u_2}) < V(\hat{\pi}_{u_1})$를 만족하는 조건을 구해보면 다음과 같다.

$$\pi < \frac{B + \sqrt{B^2 + 4AC}}{2A} \tag{3.9}$$

여기서,

$$A = \{T + (1 - T)p_1\}^2 + T^2 - \{1 - (1 - p_1)\pi_y\}^2 - (1 - \pi_y)^2,$$
$$B = \{T + (1 - T)p_1\}\{1 - 2(1 - T)(1 - p_1)\pi_y\} + T\{1 - 2(1 - T)\pi_y\}$$
$$+\{1 - (1 - p_1)\pi_y\}\{2(1 - p_1)\pi_y - 1\} + (1 - \pi_y)(2\pi_y - 1)\},$$
$$C = (1 - T)(1 - p_1)\pi_y\{1 - (1 - T)(1 - p_1)\pi_y\} + (1 - T)\pi_y\{1 - (1 - T)\pi_y\}$$
$$+(1 - p_1)\pi_y\{(1 - p_1)\pi_y - 1\} + \pi_y(\pi_y - 1)$$

이다.

또한, 식 (3.5)와 식 (3.6)을 이용하여 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성을 수치적으로 비교하기 위하여 $n = 100, T = 0.3$일 때 π와 π_y 및 p_1을 변화시켜 가면서 분산비 $V(\hat{\pi}_{u_1})/V(\hat{\pi}_{u_2})$를 계산한 결과 다음 표 3.3을 얻었다.

표 3.3: 2단계 이표본 무관질문모형과 개선된 이표본 무관질문모형과의 효율성 비교

π	π_y / p_1	0.1	0.3	0.5	0.7	0.9
0.1	0.1	0.5963	0.7792	0.9310	1.1879	2.0842
	0.3	0.6271	0.7769	0.8965	1.0736	1.5329
	0.5	0.6736	0.7840	0.8723	0.9932	1.2614
	0.7	0.7489	0.8110	0.8648	0.9383	1.0891
	0.9	0.8845	0.8941	0.9066	0.9299	0.9835
0.2	0.1	0.5797	0.7988	0.9848	1.2697	2.2063
	0.3	0.6249	0.8023	0.9479	1.1434	1.6167
	0.5	0.6851	0.8170	0.9238	1.0560	1.3275
	0.7	0.7706	0.8517	0.9181	0.9976	1.1459
	0.9	0.9026	0.9284	0.9503	0.9776	1.0280
0.3	0.1	0.5999	0.8403	1.0561	1.3690	2.3433
	0.3	0.6526	0.8457	1.0123	1.2239	1.7055
	0.5	0.7171	0.8608	0.9814	1.1219	1.3911
	0.7	0.8011	0.8918	0.9665	1.0491	1.1914
	0.9	0.9193	0.9512	0.9772	1.0056	1.0515
0.4	0.1	0.6466	0.9076	1.1531	1.4952	2.5040
	0.3	0.7038	0.9111	1.0968	1.3224	1.8046
	0.5	0.7678	0.9208	1.0520	1.1971	1.4567
	0.7	0.8436	0.9397	1.0191	1.1012	1.2324
	0.9	0.9397	0.9738	1.0010	1.0280	1.0679
0.5	0.1	0.7239	1.0113	1.2906	1.6655	2.7037
	0.3	0.7839	1.0087	1.2139	1.4516	1.9226
	0.5	0.8435	1.0061	1.1458	1.2911	1.5301
	0.7	0.9042	1.0036	1.0842	1.1612	1.2738
	0.9	0.9672	1.0012	1.0271	1.0504	1.0819

표 3.3에서 1보다 큰 값은 개선된 이표본 무관질문모형이 2단계 이표본 무관질문모형보다 더 효율적임을 나타낸다. 표 3.3에 나타난 결과들은 상대적으로 p_1값이나 π값보다는 π_y값에 더 많은 영향을 받고 있는 것으로 나타났다. 표 3.3로부터 개선된 이표본 무관질문모형은 π값이 0.5보다 큰 경우에 p_1값이 작을수록 π_y값이 클수록 2단계 이표본 무관질문모형보다 더 효율이 좋게 나타나고 있었다. 특히, π_y값이 0.9와 같이 매우 클 때 2단계 이표본 무관질문모형보다 그 효율성이 높으므로 개선된 이표본 무관질문모형을 사용할 때에는 π_y값이 0.9가 되도록 해 주는 것이 가장 바람직함을 알 수 있었다. 이 결과는 표 3.2에서 설

명했던 것과 마찬가지로 π_y값이 클 때에는 식 (3.5)와 식 (3.6)에 포함되어 있는 $\lambda_i(1-\lambda_i)$값이 2단계 이표본 무관질문모형에 비해 개선된 이표본 무관질문모형이 더 작아지게 되어 개선된 이표본 무관질문모형의 분산이 작아지기 때문이다. 한편, T값을 0.1에서 0.9까지 변화시켜 가면서 효율성을 비교해 본 결과 전반적으로 표 3.3과 비슷한 경향을 나타내고 있었으나, T값이 클수록 2단계 이표본 무관질문모형의 효율이 향상되는 결과를 얻을 수 있었다. 이러한 결과는 역시 T값이 1에 가까워질수록 직접질문의 형태를 취하기 때문에 그 만큼 분산이 작아지기 때문이다.

4. 결론

본 논문에서는 2단계 무관질문모형과 개선된 무관질문모형에 대하여 무관한 속성이 미지일 때 두 개의 독립표본을 이용하는 이표본 무관질문모형으로 확장하였다. 그리고, Greenberg et al.의 이표본 무관질문모형과 2단계 이표본 무관질문모형, 그리고 개선된 이표본 무관질문모형과의 효율성을 비교해 보았다. 그 결과, 2단계 이표본 무관질문모형은 π값이 매우 작을 때, 개선된 이표본 무관질문모형은 π_y값이 클수록 Greenberg et al.의 이표본 무관질문모형보다 효율이 좋게 나타나고 있었다. 또한, 2단계 이표본 무관질문모형은 T값이 클수록, 개선된 이표본 무관질문모형은 π_y값이 클수록 효율적인 것으로 나타났다. 이러한 사실을 이용하게 되면 좀 더 효율적으로 이표본 무관질문모형을 실제조사에 적용할 수 있으리라고 여겨진다.

참고문헌

[1] 김종호, 류제복, 이기성 (1992). 새로운 2단계 확률화응답모형, <응용통계연구>, 제5권 2호, 157-167.

[2] 이기성, 홍기학 (1998). 개선된 무관질문모형, <응용통계연구>, 제11권 2호, 415-421.

[3] Abdel-Latif A., Abul-Ela, Greenberg, B.G. and Horvitz, D.G. (1967). A Multiproportions Randomized Response Model, *Journal of the American Statistical Association*, vol. 62, 990-1008.

[4] Folsom, R.E., Greenberg, B.G., Horvitz, D.G. and Abernathy, J.R. (1973). The Two Alternate Questions Randomized Response Model for Human Surveys, *Journal of the American Statistical Association*, vol. 68, 525-530.

[5] 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.

[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] Moors, J.J.A. (1971). Optimization of the Unrelated Question Randomized Response Model, *Journal of the American Statistical Association*, vol. 66, 627-629.

[9] 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 12월 접수, 1999년 11월 채택]

Two-Stage Two Sample Unrelated Question Model

Gi–Sung Lee [1)] Ki–Hak Hong [2)]

ABSTRACT

In this paper, we extended the Kim et al.'s two-stage unrelated question model(1992) and the Lee et al.'s improved unrelated question model(1998) to two sample unrelated question model of using two independent samples in the case of unknown π_y.

Keywords: Two-stage unrelated question model; Improved unrelated question model; Two sample unrelated question model.

1) Associate Professor, Department of Computer Science & Statistics, Woosuk University.
E-mail: gisung@core.woosuk.ac.kr
2) Associate Professor, Department of Computer Science, Dongshin University.
E-mail: khhong@blue.dongshinu.ac.kr