The Korean Communications
in Statistics Vol. 6, No. 1, 1999
pp. 117-129

Modified Nayak's Randomized Response Model

Gi-Sung Lee[1] Ki-Hak Hong[2]

Abstract

Nayak(1994) suggested a combined randomized response model that combined the Warner's model and Greenberg et al.'s model. In this paper, we extend Nayak's model to two sample case of including unknown unrelated character, also propose some combined models such W-M model and G-M model that modify the Nayak's model. We suggest the efficiency conditions of our models for Nayak's model, also find the efficiency condition of G-M model for the W-M model.

1. 서 론

Warner(1965)는 두 질문 중 하나는 민감한 질문으로서 다른 하나는 민감한 질문과 배반되는 질문으로 구성된 확률장치를 사용하여 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대해 정보를 이끌어 낼 수 있는 확률화응답모형(randomized response model ; RRM)을 처음으로 제시하였다. 그리고, Greenberg et al.(1969)은 Warner의 관련질문모형(related question model)에서 사용한 민감한 질문과 배반되는 질문 대신에 민감한 질문과 전혀 관계가 없는 무관한 질문을 사용하는 무관질문모형(unrelated question model)을 제안하였다. 또한, Morton(1976)은 민감한 질문과 "예"와 "아니오"라는 강요된 응답을 하도록 하는 3개의 설문으로 구성된 확률화응답모형을 제안하였으며, Nayak(1994)은 Warner의 모형과 Greenberg et al.의 모형을 합한 형태로 Warner의 모형에 무관한 질문을 추가하여 설문의 수가 3개인 확률화응답모형을 제안하였다.

본 논문에서는 Nayak의 확률화응답모형을 무관한 속성이 미지일 때도 사용할 수 있는 이표본 모형으로 확장하였다. 그리고, Nayak의 모형을 수정한 형태의 모형으로 Warner의 모형과 Morton의 모형에서 사용한 "예"라는 강요된 응답을 결합한 확률화응답모형(W-M 모형)과 Greenberg et al.의 모형과 Morton의 모형을 결합한 확률화응답모형(G-M 모형)을 제안하였다. 또한, 제안한 W-M 모형과 G-M 모형이 Nayak의 모형보다 효율적이 되는 조건을 제시하였고, G-M 모형이 W-M 모형보다 효율적이 되는 조건도 제시하였다.

2. Nayak의 모형과 그 확장

1) Assistant Professor, Department of Computer Science & Statistics, Woosuk University, Wanju-gun, Chonbuk, 565-701, KOREA.
2) Associate Professor, Department of Computer Science, Dongshin University, Daeho-dong, Naju, Chonnam, 520-714, KOREA.

이 장에서는 Warner의 모형과 Greenberg et al.의 모형을 결합한 Nayak의 확률화응답모형에 대하여 간략히 살펴보고, Nayak의 모형을 무관한 속성에 대한 모비율 π_y가 미지일 때도 사용할 수 있는 Nayak의 이표본 모형으로 확장하고자 한다.

2.1 Nayak의 모형

Nayak의 모형에서 사용하는 확률장치는 다음과 같이 3개의 설문으로 구성되어 있다.

설문 1 : 당신은 민감한 속성 A를 가지고 있습니까?
설문 2 : 당신은 민감한 속성 A를 가지고 있지 않습니까?
설문 3 : 당신은 무관한 속성 Y를 가지고 있습니까?

여기서, 설문 1이 선택될 확률은 p_1, 설문 2가 선택될 확률은 p_2, 설문 3이 선택될 확률 p_3이다 ($\sum_{i=1}^{3} p_i = 1$). 이 때, 응답자들은 확률장치에 의해서 선택된 설문에 대하여 "예" 또는 "아니오"라고 응답하게 된다.

따라서, 이러한 Nayak의 모형에서 응답자가 "예"라고 응답할 확률을 구해 보면 다음과 같다.

$$\begin{aligned} \lambda &= \pi(p_1 + p_3\pi_y) + (1-\pi)(p_2 + p_3\pi_y) \\ &= \pi(p_1 - p_2) + (p_2 + p_3\pi_y)\,. \end{aligned} \tag{2.1}$$

여기서, π는 민감한 속성에 대한 모비율이며, π_y는 무관한 속성에 대한 모비율이며 알고 있다고 가정한다.

단순임의복원추출된 n명의 응답자들 중에서 "예"라고 응답한 사람들의 수를 n^*라 하면 $\hat{\lambda} = \frac{n^*}{n}$가 된다. 따라서, Nayak의 모형에서 민감한 속성에 대한 모비율 π의 추정량 $\hat{\pi}_n$와 그 분산은 다음과 같다.

$$\hat{\pi}_n = \frac{\hat{\lambda} - (p_2 + p_3\pi_y)}{p_1 - p_2}\,, \quad (p_1 \neq p_2)\,. \tag{2.2}$$

$$V(\hat{\pi}_n) = \frac{\pi(1-\pi)}{n} + \frac{\pi p_3(1-2\pi_y)(p_1 - p_2) + (p_2 + p_3\pi_y)(1 - p_2 - p_3\pi_y)}{n(p_1 - p_2)^2}\,, \quad (p_1 \neq p_2)\,. \tag{2.3}$$

2.2 이표본 모형으로의 확장

Nayak의 모형을 π_y가 미지일 때 두 개의 독립표본을 이용하여 민감한 속성에 대한 모비율 π를 추정할 수 있는 Nayak의 이표본 모형으로 확장하고자 한다.

단순임의복원추출된 서로 독립인 n_i $(i=1,2)$명의 응답자들은 앞에서 사용한 동일한 확률장치를 통해 선택된 설문에 대해 "예" 또는 "아니오"라고 응답한다.

이 때, 설문 1이 선택될 확률은 p_{i1}, 설문 2가 선택될 확률은 p_{i2}, 설문 3이 선택될 확률 p_{i3}이다 $(\sum_{j=1}^{3} p_{ij}=1,\ i=1,2)$.

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

$$\lambda_i=\pi(p_{i1}-p_{i2})+(p_{i2}+p_{i3}\pi_y),\ \ (i=1,2). \tag{2.4}$$

그리고, 표본 i에서 "예"라고 응답한 사람의 수를 n_i^*라 하면 $\hat{\lambda}_i=\dfrac{n_i^*}{n_i}$가 된다.

따라서, π와 π_y에 대한 추정량 $\hat{\pi}_{n2}$와 $\hat{\pi}_{yn2}$를 각각 구해 보면 다음과 같다.

$$\hat{\pi}_{n2}=\frac{p_{23}(p_{12}-\hat{\lambda}_1)-p_{13}(p_{22}-\hat{\lambda}_2)}{p_{13}(p_{21}-p_{22})-p_{23}(p_{11}-p_{12})}, \tag{2.5}$$

$$\hat{\pi}_{yn2}=\frac{(p_{21}-p_{22})\hat{\lambda}_1-(p_{11}-p_{12})\hat{\lambda}_2+p_{11}p_{22}-p_{12}p_{21}}{p_{13}(p_{21}-p_{22})-p_{23}(p_{11}-p_{12})} \tag{2.6}$$
$$,\ (p_{13}(p_{21}-p_{22})\neq p_{23}(p_{11}-p_{12})).$$

이 때, $E(\hat{\lambda}_i)=\lambda_i$이므로 $E(\hat{\pi}_{n2})=\pi$가 성립되어 추정량 $\hat{\pi}_{n2}$는 π의 불편추정량이 된다. 또한, $V(\hat{\lambda}_i)=\dfrac{\lambda_i(1-\lambda_i)}{n_i}$이고, $\hat{\lambda}_1$와 $\hat{\lambda}_2$가 독립이므로 $\hat{\pi}_{n2}$의 분산을 구해보면 다음과 같다.

$$V(\hat{\pi}_{n2})=\frac{\dfrac{p_{23}^2\lambda_1(1-\lambda_1)}{n_1}+\dfrac{p_{13}^2\lambda_2(1-\lambda_2)}{n_2}}{\{p_{13}(p_{21}-p_{22})-p_{23}(p_{11}-p_{12})\}^2} \tag{2.7}$$
$$,\ (p_{13}(p_{21}-p_{22})\neq p_{23}(p_{11}-p_{12}))$$

여기서, $\lambda_1=\pi(p_{11}-p_{12})+(p_{12}+p_{13}\pi_y)$, $\lambda_2=\pi(p_{21}-p_{22})+(p_{22}+p_{23}\pi_y)$이다.

식(2.5)의 추정량 $\hat{\pi}_{n2}$는 특정구간에서 음수 값을 갖게 되므로 이를 보완하기 위하여 $p_{13}(p_{21}-p_{22}) > p_{23}(p_{11}-p_{12})$라고 가정하고 구한 수정된 추정량 $\tilde{\pi}_{n2}$는 다음과 같다.

$$\tilde{\pi}_{n2} = \begin{cases} 0, & p_{13}\lambda_2 - p_{23}\lambda_1 \leq p_{13}p_{22} - p_{12}p_{23} \\ \hat{\pi}_{n2}, & p_{13}p_{22} - p_{12}p_{23} < p_{13}\lambda_2 - p_{23}\lambda_1 < p_{13}p_{21} - p_{11}p_{23} \\ 1, & p_{13}p_{21} - p_{11}p_{23} \leq p_{13}\lambda_2 - p_{23}\lambda_1 \end{cases} \tag{2.8}$$

이 때, 수정된 추정량 $\tilde{\pi}_{n2}$은 불편추정량이 아니기 때문에 오차의 측도로서 π에 대한 평균제곱오차(MSE)를 사용해야 한다. 따라서,

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

로 두면, $\tilde{\pi}_{n2}$의 평균제곱오차가 $\hat{\pi}_{n2}$의 평균제곱오차보다 작게 됨을 다음과 같이 보일 수 있다.

$$\begin{aligned} MSE(\hat{\pi}_{n2}) &= E[(\hat{\pi}_{n2} - \pi)^2] \\ &= E[(\hat{\pi}_{n2} - \pi)^2 I(-\hat{\pi}_{n2})] + E[(\hat{\pi}_{n2} - \pi)^2 I(\hat{\pi}_{n2}) I(1-\hat{\pi}_{n2})] \\ &\quad + E[(\hat{\pi}_{n2} - \pi)^2 I(\hat{\pi}_{n2} - 1)] \\ &\geq E[\pi^2 I(-\hat{\pi}_{n2})] + E[(\hat{\pi}_{n2} - \pi)^2 I(\hat{\pi}_{n2}) I(1-\hat{\pi}_{n2})] \\ &\quad + E[(1-\pi)^2 I(\hat{\pi}_{n2} - 1)] \\ &= MSE(\tilde{\pi}_{n2}). \end{aligned}$$

3. 수정된 Nayak의 확률화응답모형

이 장에서는 Nayak의 확률화응답모형을 수정한 형태의 모형으로 W-M 모형과 G-M 모형을 제안하고자 한다

3.1 W-M 모형

이 절에서는 Warner의 모형과 Morton의 모형을 결합한 확률화응답모형(W-M 모형)을 제안하

고자 한다. 이 W-M 모형은 Nayak의 모형에서 사용한 설문 3의 무관한 질문 대신 Morton의 모형의 "예"라는 응답을 강요하는 형태의 모형이다. 따라서, W-M 모형은 Nayak의 모형에서 무관한 속성의 모비율이 $\pi_y = 1$인 특별한 경우의 모형임을 알 수 있다.

W-M 모형에서 사용하는 확률장치는 다음과 같이 3개의 설문으로 구성되어 있다.

설문 1 : 당신은 민감한 속성 A를 가지고 있습니까?
설문 2 : 당신은 민감한 속성 A를 가지고 있지 않습니까?
설문 3 : "예"라고 응답하세요.

여기서, 설문 1이 선택될 확률은 p_1, 설문 2가 선택될 확률은 p_2, 설문 3이 선택될 확률 p_3이다 $(\sum_{i=1}^{3} p_i = 1)$. 이 때, 응답자들은 확률장치에 의해서 선택된 설문 1과 2에 대하여 "예" 또는 "아니오"라고 응답하게 되고, 설문 3이 선택되면 "예"라고만 응답하게 된다.

따라서, 이러한 W-M 모형에서 응답자가 "예"라고 응답할 확률을 구해 보면 다음과 같다.

$$\begin{aligned} \lambda &= \pi(p_1 + p_3) + (1-\pi)(p_2 + p_3) \\ &= \pi(p_1 - p_2) + (1 - p_1) . \end{aligned} \tag{3.1}$$

여기서, π는 민감한 속성에 대한 모비율이다.

단순임의복원추출된 n명의 응답자들 중에서 "예"라고 응답한 사람들의 수를 n^*라 하면 $\hat{\lambda} = \frac{n^*}{n}$가 된다. 따라서, W-M 모형에서 민감한 속성에 대한 모비율 π의 추정량 $\hat{\pi}_{wm}$는 다음과 같다.

$$\hat{\pi}_{wm} = \frac{\hat{\lambda} - (1 - p_1)}{p_1 - p_2}, \quad (p_1 \neq p_2). \tag{3.2}$$

<정리 1> 추정량 $\hat{\pi}_{wm}$는 모비율 π의 불편추정량이다.

(증명)

$E(\hat{\lambda}) = \lambda$이므로, $\hat{\pi}_{wm}$의 기대값은 다음과 같다.

$$\begin{aligned} E(\hat{\pi}_{wm}) &= E\left[\frac{\hat{\lambda}-(1-p_1)}{p_1-p_2}\right] \\ &= \frac{\lambda-(1-p_1)}{p_1-p_2} \\ &= \frac{\pi(p_1-p_2)+(1-p_1)-(1-p_1)}{p_1-p_2} \\ &= \pi. \end{aligned}$$

■

<정리 2> 추정량 $\hat{\pi}_{wm}$의 분산은 다음과 같다.

$$V(\hat{\pi}_{wm}) = \frac{\pi(1-\pi)}{n} + \frac{(1-p_1)p_1 - \pi(p_1-p_2)p_3}{n(p_1-p_2)^2}. \tag{3.3}$$

(증명)

$V(\hat{\lambda}) = \dfrac{\lambda(1-\lambda)}{n}$ 이므로, $\hat{\pi}_{wm}$의 분산은 다음과 같다.

$$\begin{aligned} V(\hat{\pi}_{wm}) &= V\left[\frac{\hat{\lambda}-(1-p_1)}{p_1-p_2}\right] \\ &= \frac{\lambda(1-\lambda)}{n(p_1-p_2)^2} \\ &= \frac{[\pi(p_1-p_2)+(1-p_1)][1-\pi(p_1-p_2)-(1-p_1)]}{n(p_1-p_2)^2} \\ &= \frac{\pi(1-\pi)}{n} + \frac{(1-p_1)p_1 - \pi(p_1-p_2)p_3}{n(p_1-p_2)^2}. \end{aligned}$$

■

식(3.2)의 추정량 $\hat{\pi}_{wm}$는 특정구간에서 음수 값을 갖게 되므로 이를 보완하기 위하여 $p_1 > p_2$ 라고 가정하고 구한 수정된 추정량 $\tilde{\pi}_{wm}$는 다음과 같다.

$$\tilde{\pi}_{wm} = \begin{cases} 0, & \hat{\lambda} \le 1-p_1 \\ \hat{\pi}_{wm}, & 1-p_1 < \hat{\lambda} < 1-p_2 \\ 1, & 1-p_2 \le \hat{\lambda} \end{cases} \tag{3.4}$$

이 때, 수정된 추정량 $\tilde{\pi}_{wm}$는 불편추정량은 아니지만 $MSE(\tilde{\pi}_{wm}) < V(\hat{\pi}_{wm})$가 성립된다.

3.2 G-M 모형

이 절에서는 Greenberg et al.의 모형과 Morton의 모형을 결합한 확률화응답모형(G-M 모형)을 제안하고자 한다. 이 G-M 모형은 민감한 질문과 무관한 질문을 사용한 Greenberg et al.의 모형에 Morton의 모형의 "예"라는 응답을 강요하는 질문을 추가한 형태의 모형이다.

G-M 모형에서 사용하는 확률장치는 다음과 같이 3개의 설문으로 구성되어 있다.

설문 1 : 당신은 민감한 속성 A를 가지고 있습니까?
설문 2 : 당신은 무관한 속성 Y를 가지고 있습니까?
설문 3 : "예"라고 응답하세요.

여기서, 설문 1이 선택될 확률은 p_1, 설문 2가 선택될 확률은 p_2, 설문 3이 선택될 확률 p_3이다 ($\sum_{i=1}^{3} p_i = 1$). 이 때, 응답자들은 확률장치에 의해서 선택된 설문 1과 2에 대하여 "예" 또는 "아니오"라고 응답하게 되고, 설문 3이 선택되면 "예"라고만 응답하게 된다.

따라서, 이러한 G-M 모형에서 응답자가 "예"라고 응답할 확률을 구해 보면 다음과 같다.

$$\lambda = \pi p_1 + p_2 \pi_y + p_3 . \tag{3.5}$$

여기서, π는 민감한 속성에 대한 모비율이고, π_y는 무관한 속성에 대한 모비율이며 알고 있다고 가정한다.

단순임의복원추출된 n명의 응답자들 중에서 "예"라고 응답한 사람들의 수를 n^*라 하면 $\hat{\lambda} = \frac{n^*}{n}$가 된다. 따라서, G-M 모형에서 민감한 속성에 대한 모비율 π의 추정량 $\hat{\pi}_{gm}$는 다음과 같다.

$$\hat{\pi}_{gm} = \frac{\hat{\lambda} - p_2 \pi_y - p_3}{p_1} . \tag{3.6}$$

<정리 3> 추정량 $\hat{\pi}_{gm}$는 모비율 π의 불편추정량이다.

(증명)

$E(\hat{\lambda}) = \lambda$이므로, $\hat{\pi}_{gm}$의 기대값은 다음과 같다.

$$
\begin{aligned}
E(\hat{\pi}_{gm}) &= E\left[\frac{\hat{\lambda} - p_2\pi_y - p_3}{p_1}\right] \\
&= \frac{\lambda - p_2\pi_y - p_3}{p_1} \\
&= \frac{\pi p_1 + p_2\pi_y + p_3 - p_2\pi_y - p_3}{p_1} \\
&= \pi .
\end{aligned}
$$

■

<정리 4> 추정량 $\hat{\pi}_{gm}$의 분산은 다음과 같다.

$$
V(\hat{\pi}_{gm}) = \frac{\pi(1-\pi)}{n} + \frac{\pi p_1(1-p_1) + (p_2\pi_y + p_3)(1 - 2\pi p_1 - p_2\pi_y - p_3)}{np_1^{\,2}} . \tag{3.7}
$$

(증명)

$V(\hat{\lambda}) = \frac{\lambda(1-\lambda)}{n}$ 이므로, $\hat{\pi}_{gm}$의 분산은 다음과 같다.

$$
\begin{aligned}
V(\hat{\pi}_{gm}) &= V\left[\frac{\hat{\lambda} - p_2\pi_y - p_3}{p_1}\right] \\
&= \frac{\lambda(1-\lambda)}{np_1^{\,2}} \\
&= \frac{[\pi p_1 + p_2\pi_y + p_3][1 - \pi p_1 - p_2\pi_y - p_3]}{np_1^{\,2}} \\
&= \frac{\pi(1-\pi)}{n} + \frac{\pi p_1(1-p_1) + (p_2\pi_y + p_3)(1 - 2\pi p_1 - p_2\pi_y - p_3)}{np_1^{\,2}} .
\end{aligned}
$$

■

식(3.6)의 추정량 $\hat{\pi}_{gm}$는 특정구간에서 음수 값을 갖게 되므로 이를 보완하여 구한 수정된 추정량 $\tilde{\pi}_{gm}$는 다음과 같다.

$$
\tilde{\pi}_{gm} = \begin{cases} 0, & \hat{\lambda} \le p_2\pi_y + p_3 \\ \hat{\pi}_{gm}, & p_2\pi_y + p_3 < \hat{\lambda} < p_1 + p_2\pi_y + p_3 \\ 1, & p_1 + p_2\pi_y + p_3 \le \hat{\lambda} \end{cases} \tag{3.8}
$$

이 때, 수정된 추정량 $\tilde{\pi}_{gm}$ 는 불편추정량은 아니지만 $MSE(\tilde{\pi}_{gm}) < V(\hat{\pi}_{gm})$가 성립된다.

다음으로 G-M 모형을 π_y가 미지일 때 두 개의 독립표본을 이용하여 민감한 속성에 대한 모비율 π를 추정할 수 있는 G-M 이표본 모형으로 확장해 보기로 하자.

단순임의복원추출된 서로 독립인 $n_i\ (i=1,2)$명의 응답자들은 G-M 모형에서 사용한 동일한 확률장치를 통해 선택된 설문에 대해 "예" 또는 "아니오"라고 응답한다.

이 때, 설문 1이 선택될 확률은 p_{i1}, 설문 2가 선택될 확률은 p_{i2}, 설문 3이 선택될 확률 p_{i3}이다 $(\sum_{j=1}^{3} p_{ij} = 1,\ i=1,2)$.

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

$$\lambda_i = \pi p_{i1} + p_{i2}\pi_y + p_{i3},\ (i=1,2). \tag{3.9}$$

그리고, 표본 i에서 "예"라고 응답한 사람의 수를 n_i^*라 하면 $\hat{\lambda}_i = \dfrac{n_i^*}{n_i}$ 가 된다.

따라서, π와 π_y에 대한 추정량 $\hat{\pi}_{gm2}$와 $\hat{\pi}_{ygm2}$를 각각 구해 보면 다음과 같다.

$$\hat{\pi}_{gm2} = \frac{p_{22}(p_{13}-\hat{\lambda}_1) - p_{12}(p_{23}-\hat{\lambda}_2)}{p_{12}p_{21} - p_{11}p_{22}}, \tag{3.10}$$

$$\hat{\pi}_{ygm2} = \frac{p_{11}(p_{23}-\hat{\lambda}_2) - p_{21}(p_{13}-\hat{\lambda}_1)}{p_{12}p_{21} - p_{11}p_{22}},\ (p_{12}p_{21} \neq p_{11}p_{22}). \tag{3.11}$$

<정리 5> 추정량 $\hat{\pi}_{gm2}$는 모비율 π의 불편추정량이다.

(증명)

$E(\hat{\lambda}_i) = \lambda_i$이므로, $\hat{\pi}_{gm2}$의 기대값은 다음과 같다.

$$
\begin{aligned}
E(\hat{\pi}_{gm2}) &= E\left[\frac{p_{22}(p_{13}-\hat{\lambda}_1)-p_{12}(p_{23}-\hat{\lambda}_2)}{p_{12}p_{21}-p_{11}p_{22}}\right] \\
&= \frac{p_{22}(p_{13}-\lambda_1)-p_{12}(p_{23}-\lambda_2)}{p_{12}p_{21}-p_{11}p_{22}} \\
&= \frac{p_{22}(p_{13}-\pi p_{11}-p_{12}\pi_y-p_{13})-p_{12}(p_{23}-\pi p_{21}-p_{22}\pi_y-p_{23})}{p_{12}p_{21}-p_{11}p_{22}} \\
&= \pi .
\end{aligned}
$$

<정리 6> 추정량 $\hat{\pi}_{gm2}$의 분산은 다음과 같다.

$$
V(\hat{\pi}_{gm2}) = \frac{\dfrac{p_{22}^2\lambda_1(1-\lambda_1)}{n_1}+\dfrac{p_{12}^2\lambda_2(1-\lambda_2)}{n_2}}{(p_{12}p_{21}-p_{11}p_{22})^2} . \tag{3.12}
$$

(증명)

$V(\hat{\lambda}_i)=\dfrac{\lambda_i(1-\lambda_i)}{n}$ 이고, $\hat{\lambda}_1$ 와 $\hat{\lambda}_2$ 가 독립이므로 $\hat{\pi}_{gm2}$ 의 분산은 다음과 같다.

$$
\begin{aligned}
V(\hat{\pi}_{gm2}) &= V\left[\frac{p_{22}(p_{13}-\hat{\lambda}_1)-p_{12}(p_{23}-\hat{\lambda}_2)}{p_{12}p_{21}-p_{11}p_{22}}\right] \\
&= \frac{p_{22}^2V(\hat{\lambda}_1)+p_{12}^2V(\hat{\lambda}_2)}{(p_{12}p_{21}-p_{11}p_{22})^2} \\
&= \frac{\dfrac{p_{22}^2\lambda_1(1-\lambda_1)}{n_1}+\dfrac{p_{12}^2\lambda_2(1-\lambda_2)}{n_2}}{(p_{12}p_{21}-p_{11}p_{22})^2} .
\end{aligned}
$$

여기서, $\lambda_1=\pi p_{11}+p_{12}\pi_y+p_{13}$, $\lambda_2=\pi p_{21}+p_{22}\pi_y+p_{23}$ 이다.

■

식(3.10)의 추정량 $\hat{\pi}_{gm2}$는 특정구간에서 음수 값을 갖게 되므로 이를 보완하기 위하여 $p_{12}p_{21}>p_{11}p_{22}$라고 가정하고 구한 수정된 추정량 $\tilde{\pi}_{gm2}$는 다음과 같다.

$$
\tilde{\pi}_{gm2} = \begin{cases} 0, & p_{12}\hat{\lambda}_2-p_{22}\hat{\lambda}_1 \le p_{12}p_{23}-p_{13}p_{22} \\ \hat{\pi}_{gm2}, & p_{12}p_{23}-p_{13}p_{22} < p_{12}\hat{\lambda}_2-p_{22}\hat{\lambda}_1 < p_{12}(p_{21}+p_{23})-p_{22}(p_{11}+p_{13}) \\ 1, & p_{12}(p_{21}+p_{23})-p_{22}(p_{11}+p_{13}) \le p_{12}\hat{\lambda}_2-p_{22}\hat{\lambda}_1 \end{cases} \tag{3.13}
$$

이 때, 수정된 추정량 $\tilde{\pi}_{gm2}$ 는 불편추정량은 아니지만 $MSE(\tilde{\pi}_{gm2}) < V(\hat{\pi}_{gm2})$ 가 성립된다.

4. 효율성 비교

각 모형들의 효율성을 비교하는 데 있어서 평균제곱오차를 이용하는 것이 바람직할 것이라고 여겨지나, 구간들이 서로 다른 관계로 비교가 곤란하여 이 장에서는 분산 식만을 이용하여 효율성을 비교해 보고자 한다.

4.1 W-M 모형과 Nayak의 모형의 비교

이 절에서는 W-M 모형의 추정량 $\hat{\pi}_{wm}$ 가 Nayak의 모형의 추정량 $\hat{\pi}_n$ 보다 효율적이 되는 조건을 제시하고자 한다.

식(2.3)과 식(3.3)으로부터

$$V(\hat{\pi}_n) - V(\hat{\pi}_{wm}) = \frac{\pi p_3(1-2\pi_y)(p_1 - p_2) + (p_2 + p_3\pi_y)(1 - p_2 - p_3\pi_y)}{n(p_1 - p_2)^2} - \frac{(1-p_1)p_1 - \pi(p_1 - p_2)p_3}{n(p_1 - p_2)^2} > 0$$

을 만족하는 조건을 구해보면 다음과 같다.

$$\pi > \frac{p_1(1-p_1) - (p_2 + p_3\pi_y)(1 - p_2 - p_3\pi_y)}{2(p_1 - p_2)p_3(1-\pi_y)}, \quad (p_1 > p_2,\ \pi_y \neq 1). \tag{4.1}$$

그러므로, W-M 모형은 식(4.1)의 조건에서 $V(\hat{\pi}_{wm}) < V(\hat{\pi}_n)$ 를 만족하게 된다.

4.2 G-M 모형과 Nayak의 모형의 비교

이 절에서는 G-M 모형의 추정량 $\hat{\pi}_{gm}$ 가 Nayak의 모형의 추정량 $\hat{\pi}_n$ 보다 효율적이 되는 조건을 제시하고자 한다.

식(2.3)과 식(3.7)로부터

$$V(\hat{\pi}_n) - V(\hat{\pi}_{gm}) = \frac{\pi p_3(1-2\pi_y)(p_1 - p_2) + (p_2 + p_3\pi_y)(1 - p_2 - p_3\pi_y)}{n(p_1 - p_2)^2} - \frac{\pi p_1(1-p_1) + (p_2\pi_y + p_3)(1 - 2\pi p_1 - p_2\pi_y - p_3)}{n{p_1}^2} > 0$$

을 만족하는 조건을 구해보면 다음과 같다.

$$\pi > \frac{(1-p_2-p_3\pi_y)[(p_1-p_2)^2(p_2\pi_y+p_3)-p_1^2(p_2+p_3\pi_y)]}{p_1(p_1-p_2)[p_1p_3(1-2\pi_y)+(p_1-p_2)\{2(p_2\pi_y+p_3)+p_1-1\}]}, \ (p_1 \neq p_2). \quad (4.2)$$

그러므로, G-M 모형은 식(4.2)의 조건에서 $V(\hat{\pi}_{gm}) < V(\hat{\pi}_n)$를 만족하게 된다.

4.3 W-M 모형과 G-M 모형의 비교

이 절에서는 G-M 모형의 추정량 $\hat{\pi}_{gm}$가 W-M 모형의 추정량 $\hat{\pi}_{wm}$보다 효율적이 되는 조건을 제시하고자 한다.

식(3.3)과 식(3.8)로부터

$$V(\hat{\pi}_{wm}) - V(\hat{\pi}_{gm}) = \frac{(1-p_1)p_1 - \pi(p_1-p_2)p_3}{n(p_1-p_2)^2} - \frac{\pi p_1(1-p_1)+(p_2\pi_y+p_3)(1-2\pi p_1-p_2\pi_y-p_3)}{np_1^2} > 0$$

을 만족하는 조건을 구해보면 다음과 같다.

$$\pi > \frac{(p_1-p_2)^2(p_2\pi_y+p_3)(1-p_2\pi_y-p_3)-p_1^3(1-p_1)}{p_1(p_1-p_2)[(p_1-p_2)\{2(p_2\pi_y+p_3)-(1-p_1)\}-p_1p_3]}, \quad (p_1 > p_2). \quad (4.3)$$

그러므로, G-M 모형은 식(4.3)의 조건에서 $V(\hat{\pi}_{gm}) < V(\hat{\pi}_{wm})$를 만족하게 된다.

참고문헌

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

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

[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*, *A*6(15), 1511-1524.

[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*, 64, 520-539.

[5] Horvitz, D. G., Greenberg, B. G., Abernathy J. R. (1976). Randomized Response : a Data-Gathering Device for Sensitive Question, *International Statistical Review*, 44, 181-196.

[6] Mangat, N. S. (1994). An Improved Randomized Response Strategy, *Journal of the Royal Statistical Society : Series B*, 56, 93-95.

[7] Mangat, N. S. and Singh, R. (1990). An Alternative Randomized Response Procedure, *Biometrika*, 77, 439-442.

[8] Nayak, T. K. (1994). On Randomized Response Surveys for Estimating a Proportion, *Communications in Statistics - Theory and Methods,* 23(11), 3303-3321.

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