The Korean Journal of Applied Statistics (2016) DOI: http://dx.doi.org/10.5351/KJAS.2016.29.5.897
29(5), 897–906

A multiplicative unrelated quantitative randomized response model

Gi-Sung Lee[a,1]

[a]Department of Children Welfare, Woosuk University

(Received May 23, 2016; Revised July 9, 2016; Accepted July 9, 2016)

Abstract

We augment an unrelated quantitative attribute to Bar-Lev *et al.*'s model (2004) which is composed of sensitive quantitative variable and scrambled one to present a multiplicative unrelated quantitative randomized response model(MUQ_RRM). We also establish theoretical grounds to estimate the sensitive quantitative attribute according to circumstances irrespective of known or unknown unrelated quantitative attribute. Finally, we explore the relationship among the suggested model, Eichhorn-Hayre model, Bar-Lev et al.'s model and Gjestvang- Singh's model, and compare the efficiency of our model with Bar-Lev *et al.*'s model.

Keywords: multiplicative model, scrambled variable, unrelated quantitative attribute, randomized response model, efficiency

1. 서론

Warner (1965)는 응답자들의 신분이나 사생활을 보호할 수 있는 확률장치를 통해 민감한 속성의 모비율을 추정할 수 있는 확률화응답모형을 처음으로 제안하였다. 그리고 Greenberg 등 (1971)은 무관질문모형(unrelated question model)을 제안하여 양적속성(quantitative attribute)을 갖는 경우로 확장하여 양적인 민감한 속성에 대한 정보를 얻을 수 있는 확률화응답모형을 제안하였다. 그 후 많은 학자들에 의해 확률화응답모형이 개선되고, 발전되어 왔다. Ahn과 Lee (2003)는 층화 무관질문모형을 제안하였으며, Ahn과 Lee (2004)는 층화추출법에 의한 이산 양적 확률화응답모형을 제안하였다. 그리고 Kim과 Warde (2004)는 층화 Warner 모형을 제안하였으며, Kim과 Elam (2005)은 최적할당을 이용한 2단계 층화 Warner 모형을 제안하기도 하였다. 또한 Lee 등 (2011)은 승법 무관질문모형을 제안하였으며, Lee와 Park (2015)은 층화 다지 확률화응답모형을 제안하기도 하였다. 한편, Eichhorn과 Hayre (1983)는 민감한 변수에 미리 정해진 값을 곱하여 변환된 응답을 하여 응답자들의 정보를 보호할 수 있는 승법모형(multiplicative model)을 제안하였으며, Bar-Lev 등 (2004)은 확률장치를 통해 선택된 민감한 변수 또는 변환된 변수(scrambled variable)에 대한 응답을 통해 민감한 양적인 정보를 얻어낼 수 있는 승법모형을 제안하였다. 최근 Gjestvang과 Singh (2007)은 Bar-Lev 등 (2004)의 모형에 강요된

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (2015R1A2A2A 01003699).
[1]Department of Children Welfare, Woosuk University, 443, Samnye-ro, Samnye-eup, Wanju-Gun, Jeollabuk-do, 55338, Korea. E-mail: gisung@woosuk.ac.kr

응답을 추가하여 민감한 변수에 대한 실제응답, 변환된 응답 그리고 강요된 응답 중에서 확률장치에 의해 선택된 질문에 응답하도록 하는 승법 강요된 양적 확률화응답모형을 제안하였다. 본 연구에서는 민감한 변수와 변환된 변수로 구성된 Bar-Lev 등 (2004)의 승법모형에 무관한 양적변수를 새롭게 추가한 승법 무관양적속성 확률화응답모형을 제안하고자 한다. 그리고 무관한 양적변수에 대한 정보를 알 때와 모를 때로 구분하여 민감한 양적속성 추정에 대한 이론적 체계를 마련하고자 한다. 또한, 제안한 승법 무관양적속성 확률화응답모형과 기존의 승법모형인 Eichhorn-Hayre 모형, Bar-Lev 등의 모형, 그리고 Gjestvang-Singh 모형과의 관계를 살펴보고, Bar-Lev 등의 모형과의 효율성을 비교해 보고자 한다.

2. 승법 무관양적속성 확률화응답모형

2.1. 승법 무관양적속성 확률화응답모형

이 절에서는 Gjestvang과 Singh (2007)의 승법 강요양적속성 모형을 승법 무관양적속성 모형으로 발전시킨 모형을 제안하고자 한다. 제안한 승법 무관양적속성 모형은 Gjestvang과 Singh의 모형에서 사용한 강요응답 대신에 민감한 변수와 무관한 질문에 대한 응답을 하도록 하는 형태를 취한다. 크기 N인 모집단으로부터 크기가 n인 표본을 단순임의복원으로 추출한다. i번째 응답자들은 다음과 같은 세 가지 형태의 질문들로 구성된 확률장치를 이용하여 응답하게 된다.

질문 1: 당신의 민감한 변수 X_i의 값은 얼마입니까?

질문 2: 당신의 변환된 응답 X_iZ의 값은 얼마입니까?

질문 3: 당신의 무관한 변수 U_i의 값은 얼마입니까?

여기서 X_i (≥ 0)는 i번째 민감한 양적변수의 참값이며, Z는 $E_R(Z) = \mu_z$이고 $V_R(Z) = \sigma_z^2$인 분포를 따르는 변환변수로 양수값을 갖고 알고 있다고 가정하고, 무관한 변수 U_i의 모평균 μ_u도 알고 있다고 가정한다.

또한 확률장치에서 질문 1이 선택될 확률은 p_1이고, 질문 2가 선택될 확률은 p_2이며, 질문 3이 선택될 확률은 p_3이다 ($\sum_{i=1}^3 p_i = 1$). 즉, 응답자들의 응답의 분포는 다음과 같이 주어진다.

$$Y_i = \begin{cases} X_i, & (\text{선택확률} = p_1), \\ X_iZ, & (\text{선택확률} = p_2), \\ U_i, & (\text{선택확률} = p_3). \end{cases} \tag{2.1}$$

정리 2.1 민감한 양적속성의 모평균 μ_x의 비편향추정량 $\hat{\mu}_{x(L)}$는 다음과 같다.

$$\hat{\mu}_{x(L)} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i - p_3\mu_u}{p_1 + p_2\mu_z}. \tag{2.2}$$

증명: E_1을 모든 가능한 표본들에 대한 기대값이라 하고, E_2를 확률장치에 대한 기대값이라고 하면, $\hat{\mu}_{x(L)}$의 기대값은 다음과 같다.

$$\begin{aligned} E(\hat{\mu}_{x(L)}) &= E_1E_2\left[\frac{\frac{1}{n}\sum_{i=1}^n Y_i - p_3\mu_u}{p_1 + p_2\mu_z}\right] \\ &= E_1\left[\frac{\frac{1}{n}\sum_{i=1}^n E_2(Y_i) - p_3\mu_u}{p_1 + p_2\mu_z}\right] \end{aligned}$$

$$
\begin{aligned}
&= E_1\left[\frac{\frac{1}{n}\sum_{i=1}^{n}(X_ip_1+X_i\mu_zp_2+p_3U_i)-p_3\mu_u}{p_1+p_2\mu_z}\right] \\
&= \frac{p_1\mu_x+p_2\mu_z\mu_x+p_3\mu_u-p_3\mu_u}{p_1+p_2\mu_z} \\
&= \left(\frac{p_1+p_2\mu_z}{p_1+p_2\mu_z}\right)\mu_x \\
&= \mu_x.
\end{aligned}
$$

따라서 $\hat{\mu}_{x(L)}$는 모평균 μ_x의 비편향추정량이다. □

정리 2.2 추정량 $\hat{\mu}_{x(L)}$의 최소 분산은 다음과 같다.

$$
V_{\min}\left(\hat{\mu}_{x(L)}\right)=\frac{\mu_x^2}{n}\left[\frac{(1+C_x^2)\{p_1+p_2\mu_z^2(1+C_z^2)\}}{(p_1+p_2\mu_z)^2}-1+\frac{p_3C_u^2}{(1-p_3)^2}-\frac{p_3}{1-p_3}\right], \tag{2.3}
$$

여기서 $C_x^2=\sigma_x^2/\mu_x^2$, $C_z^2=\sigma_z^2/\mu_z^2$, $C_u^2=\sigma_u^2/\mu_u^2$이고, σ_x^2은 X_i의 모분산이며, σ_u^2은 U_i의 모분산이다.

증명: V_1을 모든 가능한 표본들에 대한 분산이라 하고, V_2를 확률장치에 대한 분산이라고 하면, $\hat{\mu}_{x(L)}$의 분산을 다음과 같이 표현할 수 있다.

$$
\begin{aligned}
V\left(\hat{\mu}_{x(L)}\right) &= E_1V_2(\hat{\mu}_{x(L)})+V_1E_2(\hat{\mu}_{x(L)}) \\
&= E_1\left[\frac{\frac{1}{n^2}\sum_{i=1}^{n}V_2(Y_i)}{(p_1+p_2\mu_z)^2}\right]+V_1\left[\frac{\frac{1}{n}\sum_{i=1}^{n}E_2(Y_i)-p_3\mu_u}{p_1+p_2\mu_z}\right],
\end{aligned} \tag{2.4}
$$

여기서

$$
\begin{aligned}
V_2(Y_i) &= E_2(Y_i^2)-[E_2(Y_i)]^2 \\
&= p_1X_i^2+p_2X_i^2(\mu_z^2+\sigma_z^2)+p_3U_i^2-(p_1X_i+p_2X_i\mu_z+p_3U_i)^2 \\
&= X_i^2[p_1+p_2(\sigma_z^2+\mu_z^2)-(p_1+p_2\mu_z)^2]+p_3(1-p_3)U_i^2-2p_3(p_1+p_2\mu_z)U_iX_i
\end{aligned}
$$

이므로

$$
\begin{aligned}
E_1\left[\frac{\frac{1}{n^2}\sum_{i=1}^{n}V_2(Y_i)}{(p_1+p_2\mu_z)^2}\right] &= \frac{1}{n(p_1+p_2\mu_z)^2}\left[\left\{p_1+p_2\left(\sigma_z^2+\mu_z^2\right)-(p_1+p_2\mu_z)^2\right\}\left(\sigma_x^2+\mu_x^2\right)\right. \\
&\quad \left.+p_3(1-p_3)\left(\sigma_u^2+\mu_u^2\right)-2p_3(p_1+p_2\mu_z)\mu_u\mu_x\right]
\end{aligned}
$$

이고,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}E_2(Y_i) &= \frac{1}{n}\sum_{i=1}^{n}(X_ip_1+X_i\mu_zp_2+p_3U_i) \\
&= (p_1+p_2\mu_z)\bar{X}_i+p_3\bar{U}_i
\end{aligned}
$$

이므로

$$
\begin{aligned}
V_1\left[\frac{\frac{1}{n}\sum_{i=1}^{n}E_2(Y_i)-p_3\mu_u}{p_1+p_2\mu_z}\right] &= V_1\left[\frac{(p_1+p_2\mu_z)\bar{X}_i+p_3\bar{U}_i-p_3\mu_u}{p_1+p_2\mu_z}\right] \\
&= \frac{\sigma_x^2}{n}+\left(\frac{p_3}{p_1+p_2\mu_z}\right)^2\frac{\sigma_u^2}{n}
\end{aligned}
$$

이다. 따라서 이들로부터 다음과 같은 분산 식을 얻을 수 있다.

$$V(\hat{\mu}_{x(L)}) = \frac{1}{n(p_1+p_2\mu_z)^2}\left[\left\{p_1+p_2\left(\sigma_z^2+\mu_z^2\right)\right\}\left(\sigma_x^2+\mu_x^2\right)-(p_1+p_2\mu_z)^2\mu_x^2 \right.$$
$$\left. +p_3(1-p_3)\mu_u^2+p_3\sigma_u^2-2p_3(p_1+p_2\mu_z)\mu_u\mu_x\right]. \quad (2.5)$$

식 (2.5)의 분산은 무관한 변수의 모평균 μ_u에 의존하고 있으므로 분산을 최소화시키는 μ_u를 선택할 필요가 있다. 식 (2.5)를 μ_u로 미분하여 0으로 두면 다음을 얻을 수 있다.

$$\mu_u = \frac{(p_1+p_2\mu_z)\mu_x}{1-p_3}. \quad (2.6)$$

식 (2.6)을 식 (2.5)에 대입하면, 다음과 같은 최소분산을 구할 수 있다.

$$V_{\min}(\hat{\mu}_{x(L)}) = \frac{\mu_x^2}{n}\left[\frac{(1+C_x^2)\{p_1+p_2\mu_z^2(1+C_z^2)\}}{(p_1+p_2\mu_z)^2}-1+\frac{p_3C_u^2}{(1-p_3)^2}-\frac{p_3}{1-p_3}\right].$$

□

식 (2.2)의 추정량 $\hat{\mu}_{x(L)}$는 μ_u에 의존하고, 식 (2.6)은 민감한 양적속성의 모평균 μ_x에 의존하므로 불행하게도 식 (2.6)을 실제로 사용할 수는 없다.

2.2. 승법 이표본 무관양적속성 확률화응답모형

이 절에서는 제안한 승법 무관양적속성 확률화응답모형에서 무관한 변수 U_i의 모평균 μ_u는 모를 때 두 개의 표본을 이용하여 민감한 변수에 대한 모평균 μ_x를 추정할 수 있는 승법 이표본 무관양적속성 확률화응답모형으로 확장해 보기로 하자.

표본의 크기가 n_1과 n_2인 서로 독립인 표본을 단순임의복원추출한다고 하자. 첫 번째 표본 n_1명의 응답자들은 다음과 같은 세 가지 형태의 질문들로 구성된 확률장치를 이용하여 응답하게 된다.

질문 1: 당신의 민감한 변수 X_i의 값은 얼마입니까?

질문 2: 당신의 변환된 응답 X_iZ_1의 값은 얼마입니까?

질문 3: 당신의 무관한 변수 U_i의 값은 얼마입니까?

이때 질문 1이 선택될 확률은 p_{11}, 질문 2가 선택될 확률은 p_{12}, 질문 3이 선택될 확률은 p_{13}이다 ($\sum_{i=1}^{3} p_{1i}=1$). 즉 첫 번째 표본으로부터의 응답 분포는 다음과 같이 주어진다.

$$Y_{1i} = \begin{cases} X_i, & (\text{선택확률} = p_{11}), \\ X_iZ_1, & (\text{선택확률} = p_{12}), \\ U_i, & (\text{선택확률} = p_{13}). \end{cases} \quad (2.7)$$

또한 두 번째 표본 n_2명의 응답자들은 다음과 같은 세 가지 형태의 질문들로 구성된 확률장치를 이용하여 응답하게 된다.

질문 1: 당신의 민감한 변수 X_i의 값은 얼마입니까?

질문 2: 당신의 변환된 응답 X_iZ_2의 값은 얼마입니까?

질문 3: 당신의 무관한 변수 U_i의 값은 얼마입니까?

이때 질문 1이 선택될 확률은 p_{21}, 질문 2가 선택될 확률은 p_{22}, 질문 3이 선택될 확률은 p_{23}이다 $(\sum_{i=1}^{3} p_{2i} = 1)$. 즉, 두 번째 표본으로부터의 응답 분포는 다음과 같이 주어진다.

$$Y_{2i} = \begin{cases} X_i, & (\text{선택확률} = p_{21}), \\ X_i Z_2, & (\text{선택확률} = p_{22}), \\ U_i, & (\text{선택확률} = p_{23}). \end{cases} \tag{2.8}$$

이 때

$$E(Y_{1i}) = p_{11}\mu_x + p_{12}\mu_x\mu_{z1} + p_{13}\mu_u, \tag{2.9}$$

$$E(Y_{2i}) = p_{21}\mu_x + p_{22}\mu_x\mu_{z2} + p_{23}\mu_u \tag{2.10}$$

이므로, 식 (2.9)와 식 (2.10)으로부터 다음을 얻을 수 있다.

$$p_{23}E(Y_{1i}) - p_{13}E(Y_{2i}) = (p_{11}p_{23} - p_{21}p_{13})\mu_x + (p_{12}p_{23}\mu_{z1} - p_{22}p_{13}\mu_{z2})\mu_x. \tag{2.11}$$

정리 2.3 민감한 양적속성의 모평균 μ_x의 비편향추정량 $\hat{\mu}_{x(T)}$는 다음과 같다.

$$\hat{\mu}_{x(T)} = \frac{p_{23}\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i} - p_{13}\frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}. \tag{2.12}$$

증명: E_1을 모든 가능한 표본들에 대한 기대값이라 하고, E_2를 확률장치에 대한 기대값이라고 하면, $\hat{\mu}_{x(T)}$의 기대값은 다음과 같다.

$$\begin{aligned}
E(\hat{\mu}_{x(T)}) &= E_1E_2\left[\frac{p_{23}\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i} - p_{13}\frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}\right] \\
&= E_1\left[\frac{p_{23}\frac{1}{n_1}\sum_{i=1}^{n_1} E_2(Y_{1i}) - p_{13}\frac{1}{n_2}\sum_{i=1}^{n_2} E_2(Y_{2i})}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}\right] \\
&= E_1\left[\frac{p_{23}\frac{1}{n_1}\sum_{i=1}^{n_1}(p_{11}X_i + p_{12}\mu_{z1}X_i + p_{13}U_i) - p_{13}\frac{1}{n_2}\sum_{i=1}^{n_2}(p_{21}X_i + p_{22}\mu_{z2}X_i + p_{23}U_i)}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}\right] \\
&= \frac{p_{23}(p_{11}\mu_x + p_{12}\mu_x\mu_{z1} + p_{13}\mu_u) - p_{13}(p_{21}\mu_x + p_{22}\mu_x\mu_{z2} + p_{23}\mu_u)}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})} \\
&= \left[\frac{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})}\right]\mu_x \\
&= \mu_x.
\end{aligned}$$

따라서 $\hat{\mu}_{x(T)}$는 모평균 μ_x의 비편향추정량이다. □

정리 2.4 추정량 $\hat{\mu}_{x(T)}$의 최소 분산은 다음과 같다.

$$\begin{aligned}
V(\hat{\mu}_{x(T)}) = {} & \frac{1}{\{p_{23}(p_{11} + p_{12}\mu_{z1}) - p_{13}(p_{21} + p_{22}\mu_{z2})\}^2} \\
& \times \left[p_{23}^2\left\{\frac{A_1\mu_x^2 + p_{13}(1 - p_{13})\mu_u^2(1 + C_u^2) + p_{13}^2\mu_u^2C_u^2 - 2p_{13}(p_{11} + p_{12}\mu_{z1})\mu_u\mu_x}{n_1}\right\}\right. \\
& \left. + p_{13}^2\left\{\frac{A_2\mu_x^2 + p_{23}(1 - p_{23})\mu_u^2(1 + C_u^2) + p_{23}^2\mu_u^2C_u^2 - 2p_{23}(p_{21} + p_{22}\mu_{z2})\mu_u\mu_x}{n_2}\right\}\right].
\end{aligned} \tag{2.13}$$

여기서

$$A_1 = \left\{p_{11} + p_{12}\mu_{z1}^2\left(1+C_{z1}^2\right)\right\}\left(1+C_x^2\right) - (p_{11}+p_{12}\mu_{z1})^2,$$
$$A_2 = \left\{p_{21} + p_{22}\mu_{z2}^2\left(1+C_{z2}^2\right)\right\}\left(1+C_x^2\right) - (p_{21}+p_{22}\mu_{z2})^2$$

이다.

증명:

$$V(\hat{\mu}_{x(T)}) = \frac{p_{23}^2 V\left(\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right) + p_{13}^2 V\left(\frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}\right)}{\{p_{23}(p_{11}+p_{12}\mu_{z1}) - p_{13}(p_{21}+p_{22}\mu_{z2})\}^2} \tag{2.14}$$

에서 V_1을 모든 가능한 표본들에 대한 분산이라 하고, V_2를 확률장치에 대한 분산이라고 하면,

$$\begin{aligned} V\left(\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right) &= E_1V_2\left[\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right] + V_1E_2\left[\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right] \\ &= E_1\left[\frac{1}{n_1^2}\sum_{i=1}^{n_1} V_2(Y_{1i})\right] + V_1\left[\frac{1}{n_1}\sum_{i=1}^{n_1} E_2(Y_{1i})\right] \end{aligned}$$

로부터

$$\begin{aligned} E_1\left[\frac{1}{n_1^2}\sum_{i=1}^{n_1} V_2(Y_{1i})\right] &= \frac{1}{n_1}\left[\left\{p_{11} + p_{12}\left(\sigma_{z1}^2+\mu_{z1}^2\right) - (p_{11}+p_{12}\mu_{z1})^2\right\}\left(\sigma_x^2+\mu_x^2\right)\right. \\ &\quad \left. + p_{13}(1-p_{13})(\sigma_u^2+\mu_u^2) - 2p_{13}(p_{11}+p_{12}\mu_{z1})\mu_u\mu_x\right] \\ &= \frac{1}{n_1}\left[\left\{p_{11}+p_{12}\mu_{z1}^2\left(1+C_{z1}^2\right)\right\}\mu_x^2\left(1+C_x^2\right) - (p_{11}+p_{12}\mu_{z1})^2\mu_x^2\left(1+C_x^2\right)\right. \\ &\quad \left. + p_{13}(1-p_{13})\mu_u^2\left(1+C_u^2\right) - 2p_{13}(p_{11}+p_{12}\mu_{z1})\mu_u\mu_x\right], \\ V_1\left[\frac{1}{n_1}\sum_{i=1}^{n_1} E_2(Y_{1i})\right] &= V_1\left[\frac{1}{n_1}\sum_{i=1}^{n_1}(p_{11}X_i + p_{12}\mu_{z1}X_i + p_{13}U_i)\right] \\ &= V_1\left[(p_{11}+p_{12}\mu_{z1})\bar{X}_i + p_{13}\bar{U}_i\right] \\ &= (p_{11}+p_{12}\mu_{z1})^2\frac{\sigma_x^2}{n_1} + p_{13}^2\frac{\sigma_u^2}{n_1} \\ &= \frac{1}{n_1}\left[(p_{11}+p_{12}\mu_{z1})^2C_x^2\mu_x^2 + p_{13}^2C_u^2\mu_u^2\right] \end{aligned}$$

이므로

$$\begin{aligned} V\left(\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right) &= \frac{1}{n_1}\left[\left\{p_{11}+p_{12}\mu_{z1}^2\left(1+C_{z1}^2\right)\right\}\mu_x^2\left(1+C_x^2\right) - (p_{11}+p_{12}\mu_{z1})^2\mu_x^2\right. \\ &\quad \left. + p_13(1-p_13)\mu_u^2\left(1+C_u^2\right) + p_{13}^2\mu_u^2C_u^2 - 2p_{13}(p_{11}+p_{12}\mu_{z1})\mu_u\mu_x\right] \\ &= \frac{1}{n_1}\left[A_1\mu_x^2 + p_{13}\mu_u^2\left\{1 - p_{13}\left(1-C_u^2\right)\right\} - 2p_{13}(p_{11}+p_{12}\mu_{z1})\mu_u\mu_x\right] \end{aligned} \tag{2.15}$$

이다. 마찬가지 방법으로 다음을 얻을 수 있다.

$$V\left(\frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}\right) = \frac{1}{n_2}\left[A_2\mu_x^2 + p_{23}\mu_u^2\left\{1-p_{23}\left(1-C_u^2\right)\right\} - 2p_{23}(p_{21}+p_{22}\mu_{z2})\mu_u\mu_x\right]. \tag{2.16}$$

식 (2.15)와 식 (2.16)을 식 (2.14)에 대입하면, 분산 식 (2.13)을 얻을 수 있다. □

3. 기존 승법모형들과의 비교

3.1. Eichhorn과 Hayre의 모형과의 비교

이 절에서는 제안한 승법 무관양적속성 확률화응답모형과 Eichhorn과 Hayre (1983)의 모형과의 관계를 살펴보고자 한다. Eichhorn과 Hayre의 모형으로부터 민감한 양적속성의 모평균 μ_x의 추정량 $\hat{\mu}_{x(E)}$과 그 분산은 다음과 같다.

$$\hat{\mu}_{x(E)} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{\mu_z}, \tag{3.1}$$

$$V(\hat{\mu}_{x(E)}) = \frac{\mu_x^2}{n}\left[C_x^2 + C_z^2\left(1 + C_x^2\right)\right], \tag{3.2}$$

여기서 $C_z^2 = \sigma_z^2/\mu_z^2,\ \mu_x = X/N,\ X = \sum_{i=1}^{N} X_i,\ C_x^2 = \sigma_x^2/\mu_x^2$이다.

따라서 만약 제안한 승법 무관양적속성 확률화응답모형에서 $p_1 = 0$, $p_2 = 1$이고 $p_3 = 0$이면, Eichhorn과 Hayre의 모형이 된다.

3.2. Bar-Lev 등의 모형과의 비교

이 절에서는 제안한 승법 무관양적속성 확률화응답모형과 Bar-Lev 등 (2004)의 모형과의 관계 및 효율성을 비교하고자 한다. Bar-Lev 등의 모형으로부터 민감한 양적속성의 모평균 μ_x의 추정량 $\hat{\mu}_{x(B)}$과 그 분산은 다음과 같다.

$$\hat{\mu}_{x(B)} = \frac{1}{n\{(1-p)\mu_z + p\}}\sum_{i=1}^{n} Y_i, \tag{3.3}$$

$$V(\hat{\mu}_{x(B)}) = \frac{\mu_x^2}{n}\left[C_x^2 + \left(1 + C_x^2\right)C_z^2(p)\right]. \tag{3.4}$$

여기서 $C_z^2(p) = \{(1-p)\mu_z^2(1+C_z^2) + p\}/[(1-p)\mu_z + p]^2 - 1$이다.

따라서 만약 제안한 승법 무관양적속성 확률화응답모형은 $p_1 = p$, $p_2 = 1 - p$이고 $p_3 = 0$이면, Bar-Lev 등의 모형이 된다.

두 모형 간의 효율성을 비교하기 위하여 Bar-Lev 등의 모형에 대한 승법 무관양적속성 확률화응답모형의 상대효율을 구해보면 다음과 같다.

$$\mathrm{RE} = \frac{V(\hat{\mu}_{x(B)})}{V_{\min}(\hat{\mu}_{x(L)})} = \frac{C_x^2 + (1 + C_x^2)C_z^2(p)}{\left[\frac{(1+C_x^2)\{p_1 + p_2\mu_z^2(1+C_z^2)\}}{(p_1+p_2\mu_z)^2} - 1 + \frac{p_3 C_u^2}{(1-p_3)^2} - \frac{p_3}{1-p_3}\right]}. \tag{3.5}$$

두 모형간의 효율성을 수치적으로 비교하기 위하여 $p_1 = p = 0.6$, $p_2 = 0.2$, $p_3 = 0.2$와 $p_1 = p = 0.7$, $p_2 = 0.2$, $p_3 = 0.1$ 그리고 $p_1 = p = 0.8$, $p_2 = 0.1$, $p_3 = 0.1$로 가정하였고, $\mu_z = 1, 0.5$, $\mu_x = \mu_u = 0.5$, $C_u = 0.1$로 두고 C_x와 C_z를 0.1에서 0.4까지 0.1씩 변화시켜가면서 상대효율을 구해보면 다음 Table 3.1과 같다.

Table 3.1에서 1보다 큰 값은 승법 무관양적속성 확률화응답모형이 Bar-Lev 등의 모형보다 효율성이 좋음을 나타낸다.

Table 3.1로부터 C_x값이 작을수록 그리고 C_x값이 클수록 제안한 승법 무관양적속성 확률화응답모형이 Bar-Lev 등의 모형보다 효율성이 좋게 나타남을 알 수 있었다. 그리고 제안한 승법 무관양적속성 모형은 $p_1 = p$값이 커질수록 또한 $\mu_z = 1$일 때 보다 $\mu_z = 0.5$일 때가 더 효율적인 것으로 나타났다.

Table 3.1. Efficiency comparison

		C_z	C_x			
			0.1	0.2	0.3	0.4
$\mu_z = 1$	$p_1 = p = 0.6$, $p_2 = 0.2$, $p_3 = 0.2$	0.1	0.7650	0.7893	0.7956	0.7979
		0.2	0.9403	0.8621	0.8340	0.8221
		0.3	1.0632	0.9450	0.8861	0.8573
		0.4	1.1361	1.0178	0.9414	0.8985
	$p_1 = p = 0.7$, $p_2 = 0.2$, $p_3 = 0.1$	0.1	0.8808	0.8945	0.8978	0.8990
		0.2	0.9930	0.9387	0.9207	0.9133
		0.3	1.0727	0.9903	0.9523	0.9344
		0.4	1.1202	1.0369	0.9866	0.9594
	$p_1 = p = 0.8$, $p_2 = 0.1$, $p_3 = 0.1$	0.1	0.8873	0.8969	0.8990	0.8997
		0.2	1.0458	0.9517	0.9263	0.9165
		0.3	1.1980	1.0267	0.9675	0.9427
		0.4	1.3122	1.1073	1.0176	0.9764
$\mu_z = 0.5$	$p_1 = p = 0.6$, $p_2 = 0.2$, $p_3 = 0.2$	0.1	1.1263	1.0369	0.9667	0.9219
		0.2	1.1391	1.0483	0.9760	0.9291
		0.3	1.1588	1.0663	0.9907	0.9409
		0.4	1.1833	1.0893	1.0101	0.9566
	$p_1 = p = 0.7$, $p_2 = 0.2$, $p_3 = 0.1$	0.1	1.1292	1.0608	1.0103	0.9793
		0.2	1.1353	1.0667	1.0151	0.9830
		0.3	1.1447	1.0758	1.0227	0.9891
		0.4	1.1564	1.0876	1.0327	0.9973
	$p_1 = p = 0.8$, $p_2 = 0.1$, $p_3 = 0.1$	0.1	1.3794	1.1776	1.0685	1.0130
		0.2	1.3889	1.1858	1.0744	1.0173
		0.3	1.4037	1.1991	1.0841	1.0243
		0.4	1.4226	1.2168	1.0972	1.0340

3.3. Gjestvang과 Singh의 모형과의 비교

이 절에서는 제안한 승법 무관양적속성 확률화응답모형과 Gjestvang과 Singh (2007)의 모형과의 관계를 살펴보고자 한다.

Gjestvang과 Singh의 모형으로부터 민감한 양적속성의 모평균 μ_x의 추정량 $\hat{\mu}_{x(G)}$과 그 분산은 다음과 같다.

$$\hat{\mu}_{x(G)} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - p_3 F}{p_1 + p_2\mu_z}, \tag{3.6}$$

$$V_{\min}(\hat{\mu}_{x(G)}) = \frac{\mu_x^2}{n}\left[\frac{\left(1+C_x^2\right)\left\{p_1 + p_2\mu_z^2\left(1+C_z^2\right)\right\}}{(p_1+p_2\mu_z)^2} - 1 - \frac{p_3}{1-p_3}\right]. \tag{3.7}$$

따라서 만약 확률장치의 질문 3에서 $U_i = F$이면, 제안한 승법 무관양적속성 확률화응답모형은 Gjestvang과 Singh의 모형이 된다. 여기서 F는 미리 정해진 강요된 응답값이다.

4. 결론

본 연구에서는 승법모형에 강요된 응답을 추가한 Gjestvang과 Singh (2007)이 제안한 모형에서 강요

된 응답 대신 무관한 양적변수에 대한 응답을 하도록 하는 승법 무관양적속성 확률화응답모형을 제안하였다. 그리고 무관한 양적변수에 대한 정보를 알 때와 모를 때로 구분하여 민감한 양적속성 추정에 대한 이론적 체계를 마련하였다. 또한, 제안한 승법 무관양적속성 확률화응답모형과 기존의 승법모형인 Eichhorn-Hayre 모형, Bar-Lev 등의 모형, 그리고 Gjestvang-Singh 모형과의 비교를 통해 기존의 승법모형들이 제안한 승법 무관양적속성 확률화응답모형의 특별한 경우임을 확인할 수 있었고, 제안한 모형과 Bar-Lev 등의 모형과의 효율성을 수치적으로 비교한 결과 $C_x\,(=\sigma_x/\mu_x)$값이 작을수록 그리고 $C_z\,(=\sigma_z/\mu_z)$값이 클수록 제안한 승법 무관양적속성 확률화응답모형이 Bar-Lev 등의 모형보다 효율성이 좋게 나타남을 알 수 있었다. 그리고 제안한 승법 무관양적속성 모형은 $p_1 = p$값이 커질수록 또한 $\mu_z = 1$일 때 보다 $\mu_z = 0.5$일 때가 더 효율적인 것으로 나타났다.

References

Ahn, S. C. and Lee, G. S. (2003). A stratified unrelated question model, *Journal of the Korean Data Analysis Society*, **5**, 853–864.

Ahn, S. C. and Lee, G. S. (2004). A stratified discrete quantitative randomized response model, *Journal of The Korean Data Analysis Society*, **6**, 181–191.

Bar-Lev, S. K., Bobovitch, E., and Boukai, B. (2004). A note on randomized response models for quantitative Data, *Metrika*, **60**, 255–260.

Eichhorn, B. H. and Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data, *Journal of Statistical Planning and Inference*, **7**, 307–316.

Gjestvang, C. R. and Singh, S. (2007). Forced quantitative randomized response model: a new device, *Metrika*, **66**, 243–257.

Greenberg, B. G., Kubler, R. R., Abernathy, J. R., and Horvitz, D. G. (1971). Applications of the RR technique in obtaining quantitative data, *Journal of the American Statistical Association*, **66**, 243–250.

Kim, J. M. and Elam, M. E. (2005). A two-stage stratified Warner's randomized response model using optimal allocation, *Metrika*, **61**, 1–7.

Kim, J. M. and Warde, W. D. (2004). A stratified Warner's randomized response model, *Journal of Statistical Planning and Inference*, **120**, 155–165.

Lee, G. S., Hong, K. H., and Son, C. K. (2011). A multiplicative unrelated quantitative randomized response model. In *2011 Proceedings of the Spring Conference Korean Statistical Society*, **169**.

Lee, G. S. and Park, K. S. (2015). A stratified multiproportions randomized response model, *The Korean Journal of Applied Statistics*, **28**, 1113–1120.

Warner, S. L. (1965). Randomized response; a survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, **60**, 63–69.

승법 무관양적속성 확률화응답모형

이기성[a,1]

[a]우석대학교 아동복지학과

(2016년 5월 23일 접수, 2016년 7월 9일 수정, 2016년 7월 9일 채택)

요 약

본 연구에서는 민감한 변수와 변환된 변수로 구성된 Bar-Lev 등 (2004)의 승법모형에 무관한 양적변수를 새롭게 추가한 승법 무관양적속성 확률화응답모형을 제안하였다. 그리고 무관한 양적변수에 대한 정보를 알 때와 모를 때로 구분하여 민감한 양적속성 추정에 대한 이론적 체계를 마련하고자 하였다. 또한 제안한 승법 무관양적속성 확률화응답모형과 기존의 승법모형인 Eichhorn-Hayre 모형, Bar-Lev 등의 모형, 그리고 Gjestvang-Singh 모형과의 관계를 살펴보았고, Bar-Lev 등의 모형과의 효율성을 비교하였다. 그 결과, 기존의 승법모형들이 제안한 승법 무관양적속성 확률화응답모형의 특별한 경우임을 확인할 수 있었고, 제안한 모형과 Bar-Lev 등의 모형과의 효율성을 수치적으로 비교한 결과 $C_x\,(=\sigma_x/\mu_x)$값이 작을수록 그리고 $C_z\,(=\sigma_z/\mu_z)$값이 클수록 제안한 승법 무관양적속성 확률화응답모형이 Bar-Lev 등의 모형보다 효율성이 좋게 나타남을 알 수 있었다. 그리고 제안한 승법 무관양적속성 모형은 $p_1 = p$값이 커질수록 또한 $\mu_z = 1$일 때 보다 $\mu_z = 0.5$일 때가 더 효율적인 것으로 나타났다.

주요용어: 승법모형, 변환된 변수, 무관양적속성, 확률화응답모형, 효율성

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2015R1A2A2A01003699).

[1](55338) 전북 완주군 삼례읍 삼례로 443번지, 우석대학교 아동복지학과. E-mail: gisung@woosuk.ac.kr