The Korean Journal of Applied Statistics (2014) DOI: http://dx.doi.org/10.5351/KJAS.2014.27.2.239
27(2), 239–247

An Additive Stratified Quantitative Attribute Randomized Response Model

Gi-Sung Lee[a] · Seung-Chul Ahn[b] · Ki-Hak Hong[c,1] · Chang-Kyoon Son[d]

[a]Department of Children Welfare, Woosuk University
[b]Department of Mathematics Education, Woosuk University
[c]Department of Computer Science, Dongshin University
[d]Department of Information & Statistics Science, Dongguk University

(Received December 13, 2013; Revised January 22, 2014; Accepted January 22, 2014)

Abstract

For a sensitive survey in which the population is composed by several strata with quantitative attributes, we present an additive stratified quantitative attribute randomized response model which applied stratified random sampling instead of simple random sampling to the models of Himmelfarb-Edgell's additive quantitative attribute model and Gjestvang-Singh's. We also establish theoretical grounds to estimate the stratum mean of sensitive quantitative attributes as well as the over all mean. We deal with the proportional and optimal allocation problems in each suggested model and compare the relative efficiency of the suggested two models; subsequently, Himmelfarb-Edgell's model is more efficient than Gjestvang-Singh's model under the condition of stratified random sampling.

Keywords: Additive randomized response model, quantitative attribute, stratified sampling, sample allocation.

1. 서론

Warner (1965)가 응답자들의 신분이나 사생활을 보호할 수 있는 확률장치를 통해 민감한 속성의 모비율을 추정할 수 있는 확률화응답모형을 처음으로 제안한 이후 많은 연구자들이 모집단내 민감한 속성을 추정하기 위한 연구로서 Warner모형을 발전시켜 왔다.

특히, Himmelfarb와 Edgell (1980)은 민감한 양적인 변수의 정보를 얻기 위하여 가법 모형(additive model)을 제안하였다. 이와 같은 가법 모형은 응답자들에게 자신들의 응답에 변환된 변수를 사용하여 개인적인 민감한 정보를 보호하도록 하고 있다. 즉, 민감한 변수 X에 대한 모평균 μ_x를 추정하고자 할 때 변환된 변수 Z의 모평균 μ_z와 모분산 σ_z^2을 알고 있다고 가정한 후, 단순임의추출된 응답자들은 자신들의 진실된 응답에 변환된 변수의 분포로부터 얻어진 일정한 값을 더하여 응답하게 된다. 그리고 Gjestvang과 Singh (2009)은 두 개의 알고 있는 양의 실수 값 α와 β를 변환된 변수 Z에 곱한 후 민감한 변수 X에 더하거나 빼는 가법 모형을 제안하였다. 이 가법 모형은 단순임의추출된 i번째 응답자들에게

[1]Corresponding author: Department of Computer Science, Dongshin University, 252 Daeho-Dong, Naju, Chonnam 520-714, Korea. E-mail: khhong@dsu.ac.kr

$p = \beta/(\alpha+\beta)$의 확률로 $X_i + \alpha Z$라는 변환된 응답하도록 하고, $1-p = \alpha/(\alpha+\beta)$의 확률로 $X_i - \beta Z$라는 변환된 응답을 하도록 하여 모형의 효율성이 높이고 있다.

이와 같은 대표적인 두 가법 모형은 응답자들을 추출하는 데 있어서 모두 단순임의추출법을 사용하고 있다. 하지만 우리가 관심을 가지고 있는 모집단이 층으로 구성되어 있고, 층의 크기를 알고 있는 경우라면 단순임의추출법을 이용하는 것보다는 층화임의추출법을 사용하는 것이 보다 더 효율적일 것이다. 이러한 층화추출법을 확률화응답모형에 적용하는 연구는 Ahn과 Lee (2003, 2004), Kim과 Warde (2004), Son 등 (2008), Lee (2012) 등에 의해 이루어졌다.

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 양적인 속성으로 되어 있는 경우에 Himmelfarb-Edgell의 가법 모형과 Gjestvang-Singh의 가법 모형에 단순임의추출법 대신에 층화추출법을 적용한 층화 가법 양적속성 확률화응답모형을 제안하고자 한다. 제안한 두 모형으로부터 각 층의 양적속성에 대한 모평균의 추정뿐만 아니라 모집단 전체 모평균에 대한 추정을 할 수 있는 이론적 체계를 마련하고자 한다. 그리고 층화추출에 있어서 각 층의 표본배분에 대하여 비례배분과 최적배분을 다루고자 한다. 마지막으로 두 층화 가법 양적속성 확률화응답모형들 간의 효율성을 비교해 보고자 한다.

2. Himmelfarb와 Edgell의 층화 가법 양적속성 확률화응답모형

이 장에서는 Himmelfarb와 Edgell (1980)의 가법 양적속성 모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용하여 각 층의 양적속성에 대한 모평균의 추정뿐만 아니라 모집단 전체 모평균을 추정할 수 있는 층화 가법 양적속성 확률화응답모형을 제안하고자 한다. 또한 제안한 층화 가법 양적속성 확률화응답모형에 있어서 비례배분과 최적배분의 표본배분 문제를 다루고자 한다.

크기 N인 모집단이 각 층의 크기가 N_h $(h = 1, 2, \ldots, L)$인 상호 배반인 L개의 층으로 구성되어 있으며 이때, 모집단의 각 층의 크기를 알고 있다고 가정한다. 각 층 N_h $(h = 1, 2, \ldots, L)$로부터 단순임의복원으로 크기가 n_h인 표본을 추출한다. 각 층에서 추출된 응답자들은

$$Y_{hi} = X_{hi} + Z_h$$

라는 변환된 응답(scrambled response)을 기록하도록 요구받게 되는데, 여기서 $X_{hi}\,(\geq 0)$는 h층의 i번째 민감한 양적변수의 참값이며, Z_h는 $E_R(Z_h) = \mu_{hz}$이고 $V_R(Z_h) = \sigma^2_{hz}$인 분포를 따르는 변환변수로 양수값을 갖고 알고 있다고 가정한다.

그러면, h층의 민감한 양적속성의 모평균 μ_{hx}의 추정량 $\hat{\mu}_{hx}$과 그 분산은 다음과 같이 얻어진다.

$$\hat{\mu}_{hx} = \frac{1}{n_h}\sum_{i=1}^{n_h}(Y_{hi} - \mu_{hz}),$$
$$V(\hat{\mu}_{hx}) = \frac{1}{n_h}\left(\sigma^2_{hz} + \sigma^2_{hx}\right),$$

여기서 σ^2_{hx}은 X_h의 모분산이다.

또한 모평균 μ_x의 추정량 $\hat{\mu}_{x(A)}$은 다음과 같다.

$$\hat{\mu}_{x(A)} = \sum_{h=1}^{L} W_h \hat{\mu}_{hx} = \sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(Y_{hi} - \mu_{hz}),$$

여기서 $W_h = N_h/N$이다.

정리 2.1 추정량 $\hat{\mu}_{x(A)}$는 민감한 양적속성의 모평균 μ_x의 비편향추정량이다.

증명:

$$\begin{aligned} E(\hat{\mu}_{x(A)}) &= E\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(Y_{hi}-\mu_{hz})\right] \\ &= E\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}\{(X_{hi}+Z_h)-\mu_{hz}\}\right] \\ &= \sum_{h=1}^{L}W_h\mu_{hx} \\ &= \mu_x. \end{aligned}$$

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

정리 2.2 서로 다른 층으로부터 응답자들을 독립적으로 복원추출한다면, 모평균 μ_x의 추정량 $\hat{\mu}_{x(A)}$의 분산은 다음과 같다.

$$V(\hat{\mu}_x(A)) = \sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sigma_{hz}^2+\sigma_{hx}^2\right).$$

증명:

$$\begin{aligned} V\left(\hat{\mu}_{x(A)}\right) &= V\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}(Y_{hi}-\mu_{hz})\right] \\ &= V\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}\{(X_{hi}+Z_h)-\mu_{hz}\}\right] \\ &= \sum_{h=1}^{L}\frac{W_h^2}{n_h}\left(\sigma_{hz}^2+\sigma_{hx}^2\right). \end{aligned}$$

□

다음은 층화추출에 있어서 표본의 크기 n을 각 층의 크기 n_h에 비례하여 배분하는 비례배분과 일정한 비용 하에서 분산을 최소화시키는 최적배분에 대해 살펴보자.

첫째, 비례배분은 층화추출에 있어서 각 층의 크기 N_h를 알 수 있으나 층내변동에 관해서는 알 수 없는 경우, 표본의 크기 n을 각 층의 크기 N_h에 비례하여 배분하는 방법이다. 비례배분에 있어서 h층의 표본크기는 $n_h = n(N_h/N)$이므로 추정량 $\hat{\mu}_x$의 분산은 다음과 같다.

$$V(\hat{\mu}_{x(A)})_p = \frac{1}{n}\sum_{h=1}^{L}W_h\left(\sigma_{hz}^2+\sigma_{hx}^2\right).$$

둘째, 층화 최적배분은 표본의 크기 n을 각 층에 배분하는데 있어서 일정한 비용 하에서 $V(\hat{\mu}_{x(A)})$을 최소로 하거나, 일정한 분산 $V(\hat{\mu}_{x(A)})$하에서 비용을 최소로 하는 각 층의 표본크기 n_h를 결정하는 방법이다. 층화추출에 있어서 다음과 같은 비용함수를 사용한다.

$$C = c_0 + \sum_{h=1}^{L}c_h n_h,$$

여기서 c_0는 고정비용이고, c_h는 h층의 단위당 비용이다.

Cauchy-Schwarz의 부등식을 이용하여 고정된 비용 C에 대해 $V(\hat{\mu}_{x(A)})$을 최소로 하는 n_h는 다음과 같으므로

$$n_h = n \cdot \frac{W_h \sqrt{\sigma_{hz}^2 + \sigma_{hx}^2}/\sqrt{c_h}}{\sum_{h=1}^{L} W_h \sqrt{\sigma_{hz}^2 + \sigma_{hx}^2}/\sqrt{c_h}}.$$

최적배분의 경우에 추정량 $\hat{\mu}_{x(A)}$의 분산은 다음과 같다.

$$V\left(\hat{\mu}_{x(A)}\right)_o = \frac{\sum_{h=1}^{L} W_h \sqrt{\sigma_{hz}^2 + \sigma_{hx}^2}\sqrt{c_h}}{n} \cdot \sum_{h=1}^{L} \frac{W_h \sqrt{\sigma_{hz}^2 + \sigma_{hx}^2}}{\sqrt{c_h}}.$$

3. Gjestvang, Singh의 층화 가법 양적속성 확률화응답모형

이 장에서는 Gjestvang과 Singh (2009)의 가법 양적속성 모형에 층화추출법을 적용한 층화 가법 양적속성 확률화응답모형을 제안하고자 한다. 또한 제안한 층화 가법 양적속성 확률화응답모형에 있어서 비례배분과 최적배분의 표본배분 문제를 다루고자 한다.

크기 N인 모집단이 각 층의 크기가 $N_h\,(h = 1, 2, \ldots, L)$인 상호 배반인 L개의 층으로 구성되어 있으며 이때, 모집단의 각 층의 크기를 알고 있다고 가정한다. 각 층 $N_h\,(h = 1, 2, \ldots, L)$로부터 단순임의복원으로 크기가 n_h인 표본을 추출한다.

h층의 i번째 응답자들은 $p_h = \beta_h/(\alpha_h + \beta_h)$의 확률로 민감한 변수 $X_{hi} + \alpha_h Z_h$라는 변환된 응답하도록 하고, $1 - p_h = \alpha_h/(\alpha_h + \beta_h)$의 확률로 $X_{hi} - \beta_h Z_h$라는 변환된 응답을 하도록 한다. 이때, α_h, β_h는 h층의 양의 실수값으로 알고 있다고 가정한다.

응답자들의 응답의 분포는 다음과 같이 주어진다.

$$Y_{hi} = \begin{cases} X_{hi} + \alpha_h Z_h, & \text{선택확률} : p_h = \dfrac{\beta_h}{\alpha_h + \beta_h}, \\ X_{hi} - \beta_h Z_h, & \text{선택확률} : 1 - p_h = \dfrac{\alpha_h}{\alpha_h + \beta_h}. \end{cases}$$

h층의 민감한 양적속성의 모평균 μ_{hx}의 추정량 $\hat{\mu}_{hx}$과 그 분산은 다음과 같이 얻어진다.

$$\hat{\mu}_{hx} = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{hi},$$
$$V(\hat{\mu}_{hx}) = \frac{1}{n_h} \left[\sigma_{hx}^2 + \alpha_h \beta_h \left(\sigma_{hz}^2 + \mu_{hz}^2\right)\right].$$

다음으로 추정하고자 하는 모평균 μ_x의 추정량 $\hat{\mu}_{x(G)}$은 다음과 같다.

$$\hat{\mu}_{x(G)} = \sum_{h=1}^{L} W_h \hat{\mu}_{hx} = \sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} Y_{hi},$$

여기서 $W_h = N_h/N$이다.

정리 3.1 추정량 $\hat{\mu}_{x(G)}$는 민감한 양적속성의 모평균 μ_x의 비편향추정량이다.

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

$$\begin{aligned} E\left(\hat{\mu}_{x(G)}\right) &= E_1 E_2 \left[\sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} Y_{hi}\right] \\ &= E_1 \left[\sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} E_2(Y_{hi})\right] \end{aligned}$$

에서

$$\begin{aligned} E_2(Y_{hi}) &= p_h E_2(X_{hi} + \alpha_h Z_h) + (1 - p_h) E_2(X_{hi} - \beta_h Z_h) \\ &= p_h X_{hi} + (1 - p_h) X_{hi} + \alpha_h p_h \mu_{hz} - \beta_h (1 - p_h) \mu_{hz} \\ &= X_{hi} + \frac{\alpha_h \beta_h \mu_{hz}}{\alpha_a + \beta_h} - \frac{\mu_{hz} \alpha_h \beta_h}{\alpha_a + \beta_h} \\ &= X_{hi} \end{aligned}$$

이므로

$$\begin{aligned} E(\hat{\mu}_{x(G)}) &= E_1 \left[\sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} X_{hi}\right] \\ &= \sum_{h=1}^{L} W_h \mu_{hx} \\ &= \mu_x \end{aligned}$$

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

정리 3.2 서로 다른 층으로부터 응답자들을 독립적으로 복원추출한다면, 모평균 μ_x의 추정량 $\hat{\mu}_{x(G)}$의 분산은 다음과 같다.

$$V\left(\hat{\mu}_{x(G)}\right) = \sum_{h=1}^{L} \frac{W_h^2}{n_h} \left[\sigma_{hx}^2 + \alpha_h \beta_h \left(\sigma_{hz}^2 + \mu_{hz}^2\right)\ \right]. \tag{3.1}$$

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

$$\begin{aligned} V(\hat{\mu}_{x(G)}) &= E_1 V_2(\hat{\mu}_{x(G)}) + V_1 E_2(\hat{\mu}_{x(G)}) \\ &= E_1 \left[\sum_{h=1}^{L} \frac{W_h^2}{n_h^2} \sum_{i=1}^{n_h} V_2(Y_{hi})\right] + V_1 \left[\sum_{h=1}^{L} \frac{W_h}{n_h} \sum_{i=1}^{n_h} E_2(Y_{hi})\right]. \end{aligned} \tag{3.2}$$

여기서

$$\begin{aligned} V_2(Y_{hi}) &= E_2\left(Y_{hi}^2\right) - \left[E_2(Y_{hi})\right]^2 \\ &= p_h E_2(X_{hi} + \alpha_h Z_h)^2 + (1 - p_h) E_2(X_{hi} - \beta_h Z_h)^2 - X_{hi}^2 \\ &= \left(\sigma_{hz}^2 + \mu_{hz}^2\right)\left[p_h \alpha_h^2 + (1 - p_h)\beta_h^2\right] + 2X_{hi}\mu_{hz}(p_h \alpha_h - (1 - p_h)\beta_h) \\ &= \left(\sigma_{hz}^2 + \mu_{hz}^2\right)\left[\frac{\alpha_h^2 \beta_h}{\alpha_h + \beta_h} + \frac{\alpha_h \beta_h^2}{\alpha_h + \beta_h}\right] + 2X_{hi}\mu_{hz}\left[\frac{\alpha_h \beta_h}{\alpha_h + \beta_h} - \frac{\alpha_h \beta_h}{\alpha_h + \beta_h}\right] \\ &= \alpha_h \beta_h \left(\sigma_{hz}^2 + \mu_{hz}^2\right) \end{aligned}$$

이다. 따라서

$$E_1\left[\sum_{h=1}^{L}\frac{W_h^2}{n_h^2}\sum_{i=1}^{n_h}V_2(Y_{hi})\right] = E_1\left[\sum_{h=1}^{L}\frac{W_h^2}{n_h^2}\sum_{i=1}^{n_h}\left\{\alpha_h\beta_h\left(\sigma_{hz}^2+\mu_{hz}^2\right)\right\}\right] \tag{3.3}$$

$$= \sum_{h=1}^{L}\frac{W_h^2}{n_h}\left[\alpha_h\beta_h\left(\sigma_{hz}^2+\mu_{hz}^2\right)\right],$$

$$V_1\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}E_2(Y_{hi})\right] = V_1\left[\sum_{h=1}^{L}\frac{W_h}{n_h}\sum_{i=1}^{n_h}X_{hi}\right] \tag{3.4}$$

$$= \sum_{h=1}^{L}\frac{W_h^2}{n_h}\sigma_{hx}^2$$

이므로 식 (3.3)과 식 (3.4)를 식 (3.2)에 대입하면, 분산 식 (3.1)을 얻을 수 있다. □

다음은 층화추출에 있어서 표본의 크기 n을 각 층의 크기 n_h에 비례하여 배분하는 비례배분과 일정한 비용 하에서 분산을 최소화시키는 최적배분에 대해 살펴보자.

첫째, 비례배분에 있어서 h층의 표본크기는 $n_h = n(N_h/N)$이므로 추정량 $\hat{\mu}_{x(G)}$의 분산은 다음과 같다.

$$V(\hat{\mu}_{x(G)})_p = \frac{1}{n}\sum_{h=1}^{L}W_h\left[\sigma_{hx}^2+\alpha_h\beta_h\left(\sigma_{hz}^2+\mu_{hz}^2\right)\right].$$

둘째, 층화 최적배분에 대한 비용함수식 (2.1)으로부터 Cauchy-Schwarz의 부등식을 이용하여 고정된 비용에 대해 $V(\hat{\mu}_{x(G)})$을 최소로 하는 n_h을 구해 보면 다음과 같다.

$$n_h = n\cdot\frac{W_h\sqrt{\sigma_{hx}^2+\alpha_h\beta_h(\sigma_{hz}^2+\mu_{hz}^2)}/\sqrt{c_h}}{\sum_{h=1}^{L}W_h\sqrt{\sigma_{hx}^2+\alpha_h\beta_h(\sigma_{hz}^2+\mu_{hz}^2)}/\sqrt{c_h}}.$$

그러므로 최적배분을 했을 때 $\hat{\mu}_{x(G)}$의 분산은 다음과 같다.

$$V(\hat{\mu}_{x(G)})_o = \frac{\sum_{h=1}^{L}W_h\sqrt{\sigma_{hx}^2+\alpha_h\beta_h(\sigma_{hz}^2+\mu_{hz}^2)}\sqrt{c_h}}{n}\cdot\sum_{h=1}^{L}\frac{W_h\sqrt{\sigma_{hx}^2+\alpha_h\beta_h(\sigma_{hz}^2+\mu_{hz}^2)}}{\sqrt{c_h}}.$$

4. 효율성 비교

두 모형 간의 효율성을 비교하기 위하여 Himmelfarb-Edgell의 층화 가법 모형에 대한 Gjestvang-Singh의 층화 가법 모형의 상대효율을 구해보면 다음과 같다.

$$\mathrm{RE} = \frac{V(\hat{\mu}_{x(A)})}{V(\hat{\mu}_{x(G)})} = \frac{\sum_{h=1}^{L}(\sigma_{hz}^2+\sigma_{hx}^2)}{\sum_{h=1}^{L}\left[\sigma_{hx}^2+\alpha_h\beta_h(\sigma_{hz}^2+\mu_{hz}^2)\right]} \tag{4.1}$$

$$= \frac{\sum_{h=1}^{L}(1+VR_h)}{\sum_{h=1}^{L}\left[\alpha_h\beta_h(1+C_{hz}^{-2})+VR_h\right]},$$

Table 4.1. Efficiency Comparison

C_{1z}	C_{2z}	$VR_1 = 0.5$		$VR_1 = 1.0$		$VR_1 = 1.5$	
		$VR_2 = 0.5$	$VR_2 = 1.0$	$VR_2 = 1.0$	$VR_2 = 1.5$	$VR_2 = 1.5$	$VR_2 = 2.0$
0.5	0.5	2.7522	2.2012	1.9138	1.7374	1.6181	1.5320
	1.0	2.8222	2.2392	1.9389	1.7557	1.6323	1.5436
	1.5	2.8355	2.2464	1.9436	1.7591	1.6350	1.5458
	2.0	2.8402	2.2489	1.9452	1.7603	1.6359	1.5465
1.0	0.5	2.8222	2.2392	1.9389	1.7557	1.6323	1.5436
	1.0	2.8957	2.2786	1.9646	1.7744	1.6469	1.5554
	1.5	2.9097	2.2860	1.9694	1.7779	1.6496	1.5576
	2.0	2.9147	2.2887	1.9711	1.7791	1.6505	1.5584
1.5	0.5	2.8355	2.2464	1.9436	1.7591	1.6350	1.5458
	1.0	2.9097	2.2860	1.9694	1.7779	1.6496	1.5576
	1.5	2.9239	2.2935	1.9743	1.7814	1.6523	1.5598
	2.0	2.9289	2.2962	1.9760	1.7827	1.6533	1.5606
2.0	0.5	2.8402	2.2489	1.9452	1.7603	1.6359	1.5465
	1.0	2.9147	2.2887	1.9711	1.7791	1.6505	1.5584
	1.5	2.9289	2.2962	1.9760	1.7827	1.6533	1.5606
	2.0	2.9339	2.2988	1.9777	1.7839	1.6542	1.5613

여기서 $VR_h = \sigma^2_{hx}/\sigma^2_{hz}$, $C_{hz} = \sigma_{hz}/\mu_{hz}$이다.

두 모형간의 효율성을 수치적으로 비교하기 위하여 모집단이 2개의 층으로 구성되어 있고, $\alpha_1 = \alpha_2 = 0.09$, $\beta_1 = \beta_2 = 0.1$이라 가정하였다. 그리고 $h\,(h = 1, 2)$층의 분산비(VR_h)와 C_{hz}를 0.5에서 2.0까지 변화시켜가면서 상대효율을 구해보면 다음 Table 4.1과 같다.

Table 4.1에서 1보다 큰 값은 Gjestvang-Singh의 층화 가법 모형이 Himmelfarb-Edgell의 층화 가법 모형보다 효율성이 좋음을 나타낸다. Table 4.1로부터 전반적으로 Gjestvang-Singh의 층화 가법 모형이 Himmelfarb-Edgell의 층화 가법 모형보다 효율적인 것으로 나타났다. 또한 식 (4.1)에서 알 수 있듯이 $\alpha_h\beta_h$값이 작을수록 즉, 제시한 모형의 특성이 직접질문에 가까워질수록 Gjestvang-Singh의 층화 가법 모형이 효율적임을 알 수 있었다.

5. 결론

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 양적인 속성으로 되어 있는 경우에 Himmelfarb-Edgell의 가법 모형과 Gjestvang-Singh의 가법 모형에 단순임의추출법 대신에 층화추출법을 적용한 층화 가법 양적속성 확률화응답모형을 제안하였다. 제안한 두 모형으로부터 각 층의 양적속성에 대한 모평균의 추정뿐만 아니라 모집단 전체 모평균에 대한 추정을 할 수 있는 이론적 체계를 마련하였다. 그리고 제안한 두 모형에서 비례배분과 최적배분 문제를 다루었으며, 각 배분법에 따른 분산식을 도출하였다. 마지막으로 두 층화 가법 양적속성 확률화응답모형들 간의 효율성을 비교해 본 결과 Gjestvang-Singh의 층화 가법 모형이 Himmelfarb-Edgell의 층화 가법 모형보다 효율적으로 나타났고, 특히 $\alpha_h\beta_h$값이 작을수록 즉, 제시한 모형의 특성이 직접질문에 가까워질수록 Gjestvang-Singh의 층화 가법 모형의 효율성이 커짐을 알 수 있었다.

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.

Gjestvang, C. R. and Singh, S. (2009). An improved randomized response model: Estimation of mean, *Journal of Applied Statistics*, **36**, 1361–1367.

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.

Himmelfarb, S. and Edgell, S. E. (1980). Additive constants model: A randomized response technique for eliminating evasiveness to quantitative response questions, *Journal Psychological Bulletin*, **87**, 525–530.

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. (2012). Unrelated question randomized response model by stratified replicated systematic sampling, *Journal of the Korean Data Analysis Society*, **14**, 781–791.

Son, C. K., Hong, K. H. and Lee, G. S. (2008). Estimation for nonresponse bias of randomized response data using post-stratification, *Journal of the Korean Data Analysis Society*, **10**, 1507–1516.

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

층화 가법 양적속성 확률화응답모형

이기성[a] · 안승철[b] · 홍기학[c,1] · 손창균[d]

[a]우석대학교 아동복지학과, [b]우석대학교 수학교육학과, [c]동신대학교 컴퓨터학과
[d]동국대학교 정보통계학과

(2013년 12월 13일 접수, 2014년 1월 22일 수정, 2014년 1월 22일 채택)

요 약

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 양적인 속성으로 되어 있는 경우에 Himmelfarb-Edgell의 가법 모형과 Gjestvang-Singh의 가법 모형에 단순임의추출법 대신에 층화추출법을 적용한 층화 가법 양적속성 확률화응답모형을 제안하였다. 제안한 두 모형으로부터 각 층의 양적속성에 대한 모평균의 추정뿐만 아니라 모집단 전체 모평균에 대한 추정을 할 수 있는 이론적 체계를 마련하였다. 그리고 제안한 두 모형에서 비례배분과 최적배분 문제를 다루었으며, 각 배분법에 따른 분산식을 도출하였다. 마지막으로 두 층화 가법 양적속성 확률화응답모형들 간의 효율성을 비교해 본 결과 Gjestvang-Singh의 층화 가법 모형이 Himmelfarb-Edgell의 층화 가법 모형보다 효율적으로 나타났고, 특히 $\alpha_h\beta_h$값이 작을수록 즉, 제시한 모형의 특성이 직접질문에 가까워질수록 Gjestvang-Singh의 층화 가법 모형의 효율성이 커짐을 알 수 있었다.

주요용어: 가법 확률화응답모형, 양적속성, 층화추출법, 표본배분.

[1]교신저자: (520-714) 전남 나주시 대호동 252, 동신대학교 컴퓨터학과. E-mail: khhong@dsu.ac.kr