The Korean Journal of Applied Statistics (2015) DOI: http://dx.doi.org/10.5351/KJAS.2015.28.6.1113
28(6), 1113–1120

A Stratified Multi-proportions Randomized Response Model

Gi-Sung Lee[a,1] · Kyung-Soon Park[a]

[a]Department of Children Welfare, Woosuk University

(Received September 8, 2015; Revised November 13, 2015; Accepted November 13, 2015)

Abstract

We propose a multi-proportions randomized response model by stratified simple random sampling for surveys of sensitive issues of a polychotomous population composed of several stratum. We also systemize a theoretical validity to apply multi-proportions randomized response model (Abul-Ela *et al.*' model, Eriksson's model) to stratified simple random sampling and derive the estimate and its dispersion matrix of the proportion of sensitive characteristic of population using the suggested model. Two types of sample allocations (proportional allocation and optimum allocation) are considered under the fixed cost. In efficiency, the Eriksson's model by stratified sampling are compared to the Abul-Ela *et al.*' model.

Keywords: multiproportions randomized response model, stratified random sampling, sample allocation, efficiency

1. 서론

사회 여러 분야의 조사에서 응답자들에게 개인적으로 민감한 질문을 공개적으로 혹은 직접적으로 하게 되면 무응답이나 거짓응답 또는 응답을 회피함으로써 응답자들로부터 정확한 정보를 얻을 수 없게 된다. 따라서 민감한 질문에 대해 보다 신뢰할 수 있는 정보를 얻기 위해서는 직접질문보다는 간접적인 대체 질문방식이 필요하게 된다. 이에 Warner (1965)는 응답자들에게 민감한 질문과, 민감한 질문과 배반되는 즉, 부의 관계를 갖는 질문에 대해 확률장치를 사용하여 응답하게 함으로써 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 낼 수 있는 확률화응답모형(randomized response model; RRM)을 처음으로 제시하였다. Abul-Ela 등 (1967)은 Warner의 이지모형을 다지모형으로 확장하였으며, Eriksson (1973)은 다지모집단에 대하여 2가지 대체 방법을 제시하였다. 그리고 Chaudhuri와 Mukerjee (1988)는 확률화응답모형에 대한 이론을 정리하여 체계화시켰다. 최근에는 이들 이론들의 실제적 활용에 많은 관심이 집중되고 있으며, 사회학, 경영학, 의학 등 여러 학문분야에서의 조사활동에도 이의 활용이 적극 모색되고 있다. Kim과 Warde (2004)는 층화추출법을 Warner모형에 적용하는 층화 확률화응답모형을 제안하였으며, Ahn과 Lee (2003)은 층화 무관질문모형을 제안하기도 하였다. 한편, Lee 등 (2007)은 2단계 집락추출법을 Mangat-Singh모형에 적용한 확률화응답모형을 제안하기도 하였다.

[1]Corresponding author: Department of Children Welfare, Woosuk University, 443, Samrye-Ro, Samrye-Eup, Wanju-Gun, Jeonbuk 55338, Korea. E-mail: gisung@woosuk.ac.kr

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 세대별, 연령별 또는 계층별에 따라 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 다지속성으로 되어 있는 경우에, Abul-Ela 등의 다지모형과 Eriksson의 다지무관모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용하여 각 층의 다지속성에 대한 모비율의 추정뿐만 아니라 모집단 전체 모비율에 대한 추정을 할 수 있는 층화 다지 확률화응답모형을 제안하고자 한다. 그리고 층화 다지모형에 있어서 각 층의 표본배분에 대하여 비례배분과 최적배분을 고려하여 다루고자 한다. 마지막으로 층화 다지 확률화응답모형들간의 효율성을 비교해 보고자 한다.

2. 층화 다지 확률화응답모형

2.1. Abul-Ela 등의 층화 다지모형

이 절에서는 Abul-Ela 등이 다지모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용하여 각 층의 다지속성에 대한 모비율의 추정뿐만 아니라 모집단 전체 모비율에 대한 추정을 할 수 있는 층화 다지 확률화응답모형을 제안하고자 한다. 또한 제안한 층화 다지모형에 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 있다고 하자. 이 때, 모집단의 각 층의 크기를 알고 있다고 가정한다. 또한 각 층은 t (≥ 2)개의 상호 배반인 그룹으로 이루어져 있으며, 그 중에서 적어도 하나, 또는 최대로 $t-1$개의 그룹들을 민감한 그룹으로 분류할 수 있다고 하자. 그리고 추정하고자 하는 모비율을 $\pi_{h1}, \pi_{h2}, \ldots, \pi_{ht}$라고 하자. 여기서, $0 < \pi_{hj} < 1$ $(h = 1, 2, \ldots, L,\ j = 1, 2, \ldots, t)$이고 $\sum_{j=1}^{t} \pi_{hj} = 1$이다. 각 층 N_h $(h = 1, 2, \ldots, L)$로부터 단순임의복원으로 크기가 $n_{h1}, n_{h2}, \ldots, n_{hs}$인 s $(= t-1)$개의 독립표본을 추출한다. 각 층의 확률장치는 s개의 카드 덱으로 구성되어 있으며, 카드 덱 i $(i = 1, 2, \ldots, s)$는 i번째 표본을 의미한다. 각 층의 각각의 카드 덱은 t개의 다른 종류의 카드를 가지며, j $(j = 1, 2, \ldots, t)$번째 종류의 카드는 j번째 그룹을 나타낸다. 표본으로 추출된 응답자는 지정된 카드 덱으로부터 임의로 한 장의 카드를 뽑아, 조사자에게는 비공개된 카드에 적혀 있는 다음과 같은 설문에 응답하게 된다.

설문 i: "나는 j번째 그룹에 속한다."

만약 그 설문이 자신이 해당하는 그룹을 나타내면, 응답자는 "예"라고 응답하며 그렇지 않으면 "아니오"라고 응답하게 된다. h층의 i번째 표본에서 "나는 j번째 그룹에 속한다."라는 설문이 적혀 있는 카드의 비율을 p_{hij}라 하자. 여기서, $\sum_{j=1}^{t} p_{hij} = 1$이다.

h층의 i $(i = 1, 2, \ldots, s)$번째 표본에 속하는 응답자가 "예"라고 응답할 확률은 다음과 같다.

$$\lambda_{hi} = \sum_{j=1}^{t} p_{hij} \pi_{hj}$$

그런데, $s = t - 1$이고, $\sum_{j=1}^{t} \pi_{hj} = 1$이므로

$$\sum_{j=1}^{s} (p_{hij} - p_{hit}) \pi_{hj} = \lambda_{hi} - p_{hit}$$

이 된다. 위의 식을 행렬로 나타내면 다음과 같다.

$$\boldsymbol{p}_h \boldsymbol{\pi}_h = \boldsymbol{\xi}_h, \tag{2.1}$$

여기서

$$\boldsymbol{p}_h = \begin{bmatrix} p_{h11} - p_{h1t} & p_{h12} - p_{h1t} & \cdots & p_{h1s} - p_{h1t} \\ p_{h21} - p_{h2t} & p_{h22} - p_{h2t} & \cdots & p_{h2s} - p_{h2t} \\ \vdots & \vdots & \ddots & \vdots \\ p_{hs1} - p_{hst} & p_{hs2} - p_{hst} & \cdots & p_{hss} - p_{hst} \end{bmatrix},$$

$$\boldsymbol{\pi}_h = (\pi_{h1}, \ldots, \pi_{hs})',$$

$$\boldsymbol{\xi}_h = (\lambda_{h1} - p_{h1t}, \ldots, \lambda_{hs} - p_{hst})'.$$

만약, h층의 i번째 표본에서 "예"라고 응답한 수를 n_{hi*}이라 하면, λ_{h1}의 비편향추정량은 $\hat{\lambda}_{hi} = n_{hi*}/n_{hi}$이다. 또한 식 (2.1)로부터 만약 $\boldsymbol{p}_h$가 정칙행렬이라면 $\boldsymbol{\pi}_h$의 비편향추정량은 다음과 같이 얻어진다.

$$\hat{\boldsymbol{\pi}}_h = (\hat{\pi}_{h1}, \hat{\pi}_{h2}, \ldots, \hat{\pi}_{hs})' = \boldsymbol{p}_h^{-1} \boldsymbol{c}_h, \tag{2.2}$$

여기서 $\boldsymbol{c}_h = (\hat{\lambda}_{h1} - p_{h1t}, \ldots, \hat{\lambda}_{hs} - p_{hst})'$이다. 또한, π_{ht}는 $\hat{\pi}_{ht} = 1 - \sum_{j=1}^{s} \hat{\pi}_{hj}$로 추정될 수 있다.

s개의 표본들이 독립적으로 추출되었기 때문에, $n_{h1*}, n_{h2*}, \ldots, n_{hs*}$과 $\hat{\lambda}_{h1}, \hat{\lambda}_{h2}, \ldots \hat{\lambda}_{hs}$들은 독립이므로, 각각의 i에 n_{hi*}대해은 $b(n_{hi}, \lambda_{hi})$를 한다. 그러므로 $\boldsymbol{c}_h$의 산포행렬(dispersion matrix)은 다음과 같다.

$$\mathrm{disp}(\boldsymbol{c}_h) = \mathrm{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}),$$

여기서 $V_{hii} = \lambda_{hi}(1 - \lambda_{hi})/n_{hi}$이다.

식 (2.2)에 의해 $\hat{\boldsymbol{\pi}}_h$의 산포행렬은 다음과 같다.

$$\mathrm{disp}(\hat{\boldsymbol{\pi}}_h) = \boldsymbol{p}_h^{-1} \mathrm{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'. \tag{2.3}$$

h층의 각각의 i에 대해, V_{hii}의 비편향추정량은 다음과 같다.

$$\hat{V}_{hii} = \frac{\hat{\lambda}_{hi}\left(1 - \hat{\lambda}_{hi}\right)}{n_{hi} - 1}.$$

따라서 $\mathrm{disp}(\hat{\boldsymbol{\pi}}_h)$의 비편향추정량은 다음과 같다.

$$\widehat{\mathrm{disp}}(\hat{\boldsymbol{\pi}}_h) = \boldsymbol{p}_h^{-1} \mathrm{diag}\left(\hat{V}_{h11}, \ldots, \hat{V}_{hss}\right) \left(\boldsymbol{p}_h^{-1}\right)'. \tag{2.4}$$

다음으로 추정하고자 하는 모비율 $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_s)'$의 추정량 $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_s)'$는 다음과 같다.

$$\hat{\boldsymbol{\pi}} = \sum_{h=1}^{L} W_h \hat{\boldsymbol{\pi}}_h = \sum_{h=1}^{L} W_h \boldsymbol{p}_h^{-1} \boldsymbol{c}_h, \tag{2.5}$$

여기서 $W_h = N_h/N$이다.

각 층에 있어서 $\boldsymbol{\pi}_h$가 $\boldsymbol{\pi}_h$의 비편향추정량이면, $\hat{\boldsymbol{\pi}}$의 기대값은

$$E(\hat{\boldsymbol{\pi}}) = E\left(\sum_{h=1}^{L} W_h \hat{\boldsymbol{\pi}}_h\right) = \sum_{h=1}^{L} W_h E(\hat{\boldsymbol{\pi}}_h) = \sum_{h=1}^{L} W_h \boldsymbol{\pi}_h = \boldsymbol{\pi}$$

가 되므로, $\hat{\boldsymbol{\pi}}$는 $\boldsymbol{\pi}$의 비편향추정량이다.

서로 다른 층으로부터 응답자들을 독립적으로 복원추출한다면, 모비율 $\boldsymbol{\pi}$의 추정량 $\hat{\boldsymbol{\pi}}$의 산포행렬은 다음과 같다.

$$\operatorname{disp}(\hat{\boldsymbol{\pi}}) = \sum_{h=1}^{L} W_h^2 \left[\boldsymbol{p}_h^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'\right]. \tag{2.6}$$

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 층내 변동에 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 $\hat{\boldsymbol{\pi}}$의 산포행렬은 다음과 같다.

$$\operatorname{disp}_p(\hat{\boldsymbol{\pi}}) = \frac{1}{n} \sum_{h=1}^{L} n_h W_h \left[\boldsymbol{p}_h^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss})(\boldsymbol{p}_h^{-1})'\right]. \tag{2.7}$$

둘째, 층화 최적배분은 표본의 크기 n을 각 층에 배분하는데 있어서 일정한 비용 하에서 $\operatorname{disp}(\hat{\boldsymbol{\pi}})$을 최소로 하거나, 일정한 $\operatorname{disp}(\hat{\boldsymbol{\pi}})$ 하에서 비용을 최소로 하는 각 층의 표본크기 n_h를 결정하는 방법이다. 층화추출에 있어서 다음과 같은 비용함수를 사용한다.

$$C = c_0 + \sum_{h=1}^{L} c_h n_h, \tag{2.8}$$

여기서 c_0는 고정비용이고, c_h는 h층의 단위당 비용이다.

비용 C를 일정하게 고정시키고 $\operatorname{disp}(\hat{\boldsymbol{\pi}})$을 최소로 하는 n_h을 Cauchy-Schwarz의 부등식을 이용하여 구해 보면 다음과 같다.

$$n_h = n \cdot \frac{W_h^2 \left[{\boldsymbol{p}_h}^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'\right] / c_h}{\sum_{h=1}^{L} W_h^2 \left[\boldsymbol{p}_h^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'\right] / c_h}. \tag{2.9}$$

그러므로 최적배분을 했을 때 $\hat{\boldsymbol{\pi}}$의 산포행렬은 다음과 같다.

$$\begin{aligned}\operatorname{disp}_{\min}(\hat{\boldsymbol{\pi}}) = {} & \frac{\sum_{h=1}^{L} W_h^2 \left[\boldsymbol{p}_h^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'\right] c_h}{n} \\ & \cdot \sum_{h=1}^{L} \frac{W_h^2 \left[\boldsymbol{p}_h^{-1} \operatorname{diag}(V_{h11}, V_{h22}, \ldots, V_{hss}) \left(\boldsymbol{p}_h^{-1}\right)'\right]}{c_h}.\end{aligned} \tag{2.10}$$

2.2. Eriksson의 층화 다지무관모형

이 절에서는 Eriksson이 다지무관모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용한 층화 다지무관모형을 제안하고자 한다.

2.1절과 동일하게, 각 층 $N_h\,(h = 1, 2, \ldots, L)$로부터 단순임의복원으로 크기가 $n_{h1}, n_{h2}, \ldots, n_{hs}$인 $s\ (= t-1)$개의 독립표본을 추출한다. 이 때 각 층은 $t\ (\geq 2)$개의 상호 배반인 그룹으로 이루어져

있다. h층의 i $(i=1,2,\ldots,s)$번째 표본에 속하는 각각의 응답자들은 다음과 같은 2개의 설문에 응답하게 된다. 이 모형은 각 층에 Greenberg 등 (1969)의 무관질문모형을 다지모형으로 확장한 형태이다. 이 때 각 층은 민감한 속성에 따라 t개의 그룹으로 나누었다고 가정하자.

설문 1: "나는 그룹 i에 속한다." 설문 2: "나는 그룹 Y에 속한다."

여기서, 설문 1이 선택될 확률은 p_{hij} $(j=1,2,\ldots,t-1)$이고, 설문 2가 선택될 확률은 p_{hit}이며, Y는 민감한 속성과는 무관한 이지속성이다.

h층의 i번째 표본에서 응답자들이 "예"라고 응답할 확률은 다음과 같다.

$$\lambda_{hi}=\sum_{j=1}^{t-1}p_{hij}\pi_{hj}+p_{hit}\pi_{hy},\quad (i=1,2,\ldots,t),$$

여기서 $\sum_{j=1}^{t}p_{hij}=1$이고, π_{hy}는 h층의 무관한 속성 Y의 모비율로 알고 있다고 가정한다.

그리고 위의 식을 행렬로 나타내면 다음과 같다.

$$\boldsymbol{\lambda}_h=\boldsymbol{p}_h\boldsymbol{\pi}_h^* \tag{2.11}$$

여기서

$$\boldsymbol{p}_h=\begin{bmatrix} p_{h11} & p_{h12} & \cdots & p_{h1t}\\ p_{h21} & p_{h22} & \cdots & p_{h2t}\\ \vdots & \vdots & \ddots & \vdots\\ p_{ht1} & p_{ht2} & \cdots & p_{htt}\end{bmatrix},$$

$$\boldsymbol{\lambda}_h=(\lambda_{h1},\ldots,\lambda_{ht})',$$

$$\boldsymbol{\pi}_h^*=(\pi_{h1},\ldots,\pi_{ht-1},\pi_{hy})'.$$

$\hat\lambda_{h1},\ldots,\hat\lambda_{ht}$가 $\lambda_{h1},\ldots,\lambda_{ht}$에 대응하는 표본비율이라 하고, $\hat{\boldsymbol{\lambda}}_h=(\hat\lambda_{h1},\ldots,\hat\lambda_{ht})'$이라 하면, $\boldsymbol{\pi}_h^*$의 비편향추정량은 $\boldsymbol{p}_h$가 정칙행렬일 때 다음과 같이 나타낼 수 있다.

$$\hat{\boldsymbol{\pi}}_h^*=(\hat\pi_{h1},\hat\pi_{h2},\ldots,\hat\pi_{ht-1},\hat\pi_{hy})=\boldsymbol{p}_h^{-1}\hat{\boldsymbol{\lambda}}_h. \tag{2.12}$$

또한, π_{hy}의 비편향추정량은 $\hat\pi_{hy}=1-\sum_{j=1}^{t-1}\hat\pi_{hj}$가 된다.

식 (2.3)과 식 (2.4)와 같이 $\hat{\boldsymbol{\pi}}_h^*$의 산포행렬과 이 산포행렬의 비편향추정량은 다음과 같이 주어진다.

$$\begin{aligned}\mathrm{disp}(\hat{\boldsymbol{\pi}}_h^*)&=\boldsymbol{p}_h^{-1}\mathrm{diag}(V_{h11},\ldots,V_{htt})\left(\boldsymbol{p}_h^{-1}\right)',\\ \widehat{\mathrm{disp}}(\hat{\boldsymbol{\pi}}_h^*)&=\boldsymbol{p}_h^{-1}\mathrm{diag}(\hat V_{h11},\ldots,\hat V_{htt})\left(\boldsymbol{p}_h^{-1}\right)',\end{aligned} \tag{2.13}$$

여기서 $V_{hii}=\lambda_{hi}(1-\lambda_{hi})/n_{hi}$, $\hat V_{hii}=\hat\lambda_{hi}(1-\hat\lambda_{hi})/(n_{hi}-1)$이다.

다음으로 추정하고자 하는 모비율 $\boldsymbol{\pi}$의 추정량 $\hat{\boldsymbol{\pi}}$는 다음과 같다.

$$\hat{\boldsymbol{\pi}}=\sum_{h=1}^{L}W_h\hat{\boldsymbol{\pi}}_h^*=\sum_{h=1}^{L}W_h\boldsymbol{p}_h^{-1}\hat{\boldsymbol{\lambda}}_h. \tag{2.14}$$

각 층에 있어서 $\hat{\boldsymbol{\pi}}_h^*$가 $\boldsymbol{\pi}_h$의 비편향추정량이면, $\hat{\boldsymbol{\pi}}$는 $\boldsymbol{\pi}$의 비편향추정량이 된다.

Table 3.1. Efficiency comparison

$\pi_{11}=0.1,\ \pi_{12}=0.2,\ \pi_{13}=0.7,\ \pi_{21}=0.2,\ \pi_{22}=0.3,\ \pi_{23}=0.5$								
π_{1y}	π_{2y}	$\frac{V_a(\hat{\pi}_{11})}{V_e(\hat{\pi}_{11})}$	$\frac{V_a(\hat{\pi}_{12})}{V_e(\hat{\pi}_{12})}$	$\frac{V_a(\hat{\pi}_{13})}{V_e(\hat{\pi}_{13})}$	$\frac{V_a(\hat{\pi}_{21})}{V_e(\hat{\pi}_{21})}$	$\frac{V_a(\hat{\pi}_{22})}{V_e(\hat{\pi}_{22})}$	$\frac{V_a(\hat{\pi}_{23})}{V_e(\hat{\pi}_{23})}$	$\frac{V_a(\hat{\pi})}{V_e(\hat{\pi})}$
	0.1				4.427	6.550	2.187	9.361
0.1	0.3	10.972	15.119	8.302	4.020	6.013	1.688	8.898
	0.5				3.726	5.618	1.411	8.535
	0.1				4.427	6.550	2.187	5.557
0.3	0.3	6.077	7.363	4.118	4.020	6.013	1.688	5.390
	0.5				3.726	5.618	1.411	5.255
	0.1				4.427	6.550	2.187	4.131
0.5	0.3	4.415	5.165	2.858	4.020	6.013	1.688	4.038
	0.5				3.726	5.618	1.411	3.961
$\pi_{11}=0.05,\ \pi_{12}=0.15,\ \pi_{13}=0.8,\ \pi_{21}=0.25,\ \pi_{22}=0.35,\ \pi_{23}=0.4$								
π_{1y}	π_{2y}	$\frac{V_a(\hat{\pi}_{11})}{V_e(\hat{\pi}_{11})}$	$\frac{V_a(\hat{\pi}_{12})}{V_e(\hat{\pi}_{12})}$	$\frac{V_a(\hat{\pi}_{13})}{V_e(\hat{\pi}_{13})}$	$\frac{V_a(\hat{\pi}_{21})}{V_e(\hat{\pi}_{21})}$	$\frac{V_a(\hat{\pi}_{22})}{V_e(\hat{\pi}_{22})}$	$\frac{V_a(\hat{\pi}_{23})}{V_e(\hat{\pi}_{23})}$	$\frac{V_a(\hat{\pi})}{V_e(\hat{\pi})}$
	0.1				5.244	7.372	2.200	7.514
0.1	0.3	7.444	11.150	5.860	4.646	6.620	1.613	7.197
	0.5				4.228	6.082	1.310	6.943
	0.1				5.244	7.372	2.200	4.924
0.3	0.3	4.936	6.426	3.465	4.646	6.620	1.613	4.786
	0.5				4.228	6.082	1.310	4.672
	0.1				5.244	7.372	2.200	3.816
0.5	0.3	3.851	4.767	2.556	4.646	6.620	1.613	3.732
	0.5				4.228	6.082	1.310	3.663

또한 서로 다른 층으로부터 응답자들을 독립적으로 복원추출한다면, 모비율 $\boldsymbol{\pi}$의 추정량 $\hat{\boldsymbol{\pi}}$의 산포행렬은 다음과 같다.

$$\mathrm{disp}(\hat{\boldsymbol{\pi}}) = \sum_{h=1}^{L} W_h^2 \left[\boldsymbol{p}_h^{-1}\mathrm{diag}(V_{h11}, V_{h22}, \ldots, V_{htt})\left(\boldsymbol{p}_h^{-1}\right)'\right]. \tag{2.15}$$

3. 효율성 비교

Abul-Ela 등의 층화 다지모형과 Eriksson의 층화 다지무관모형과의 효율성을 두 모형의 산포행렬 식들을 이용하여 수치적으로 비교해 보고자 한다.

$W_1 = 0.7, W_2 = 0.3,\ n_{11} = n_{12} = n_{13} = n_{21} = n_{22} = n_{23} = 100$이라고 가정하고 π_{1y}, π_{2y}와 $\pi_{11}, \pi_{12}, \pi_{13}, \pi_{21}, \pi_{22}, \pi_{23}$를 변화시켜 가면서 두 다지모형의 산포행렬 비를 계산하였다. 이 때,

$$p_{111} = p_{211} = 0.375,\ p_{112} = p_{212} = 0.375,\ p_{113}p_{213} = 0.250$$
$$p_{121} = p_{221} = 0.375,\ p_{122} = p_{222} = 0.250,\ p_{123}p_{223} = 0.375$$
$$p_{131} = p_{231} = 0.250,\ p_{132} = p_{232} = 0.375,\ p_{133}p_{233} = 0.375$$

로 두었으며, Abul-Ela 등의 층화 다지모형의 산포행렬을 V_a로, Eriksson의 층화 다지무관모형의 산포행렬을 V_e로 표현하였다.

Table 3.1에서 1보다 큰 값은 Eriksson의 층화 다지무관모형이 Abul-Ela 등의 층화 다지모형보다 효율성이 좋음을 나타낸다. Eriksson의 층화 다지무관모형이 Abul-Ela 등의 층화 다지모형보다 효율적임을 알 수 있다. 특히, π_{1y}, π_{2y}값이 작을수록 Eriksson의 층화 다지무관모형의 효율성이 증가됨을 볼 수 있었다.

4. 결론

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 세대별, 연령별, 또는 계층별에 따라 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 다지속성으로 되어 있는 경우에 사용할 수 있는 층화 다지 확률화응답모형을 제안하였다. Abul-Ela 등의 다지모형과 Eriksson의 다지무관모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용하여 각 층의 다지속성에 대한 모비율 및 모집단 전체 모비율에 대한 추정과 분산 및 분산추정량을 도출함으로서 층화 다지 확률화응답모형의 이론적 체계를 구축하였다. 그리고 층화 다지모형에 있어서 각 층의 표본배분에 대하여 비례배분과 최적배분 문제를 다루었다. 또한 층화 다지 확률화응답모형들간의 효율성을 비교해 본 결과 Eriksson의 다지무관모형이 Abul-Ela 등의 다지모형보다 효율적임을 알 수 있었다. 특히 π_{1y}, π_{2y}값이 작을수록 Eriksson의 층화 다지무관모형의 효율성이 증가됨을 볼 수 있었다.

References

Abul-Ela, A.-L. A., Greenberg, B. G. and Horvitz, D. G. (1967). A multi-proportions randomized response model, *Journal of the American Statistical Association*, **62**, 990–1008.

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

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

Eriksson, S. (1973). *Randomized interviews for sensitive questions* (Ph. D. thesis), University of Gothenburg.

Greenberg, B. G., Abul-Ela, A. 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.

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., Ryu, J. B., Hong, K. H. and Son, C. K. (2007). A study on two-stage cluster Mangat-Singh model, *Journal of The Korean Data Analysis Society*, **9**, 1801–1810.

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

층화 다지 확률화응답모형

이기성[a,1] · 박경순[a]

[a]우석대학교 아동복지학과

(2015년 9월 8일 접수, 2015년 11월 13일 수정, 2015년 11월 13일 채택)

요 약

본 논문에서는 사회적으로나 개인적으로 매우 민감한 조사에서 세대별, 연령별 또는 계층별에 따라 조사하고자 하는 모집단이 여러 개의 층으로 구성되어 있고, 각 층이 다지속성으로 되어 있는 경우에, Abul-Ela 등의 다지모형과 Eriksson의 다지무관모형에서 사용한 단순임의추출법 대신에 층화추출법을 적용하여 각 층의 다지속성에 대한 모비율의 추정뿐만 아니라 모집단 전체 모비율에 대한 추정을 할 수 있는 층화 다지 확률화응답모형을 제안하였다. 그리고 층화 다지모형에 있어서 각 층의 표본배분에 대하여 비례배분과 최적배분을 고려하여 다루었다. 또한 층화 다지 확률화응답모형들간의 효율성을 비교해 본 결과 Eriksson의 다지무관모형이 Abul-Ela 등의 다지모형보다 효율적임을 알 수 있었다.

주요용어: 다지 확률화응답모형, 층화추출법, 표본배분, 효율성

[1]교신저자: (565-701) 전북 완주군 삼례읍 삼례로 443, 우석대학교 아동복지학과. E-mail: gisung@woosuk.ac.kr