응용통계연구 제 11권 1호, 1998년, pp. 141-150

단순집락추출법에 의한 양적속성의 무관질문모형

이기성 † 홍기학 ‡

요 약

본 논문에서는 매우 민감한 조사에서 모집단이 양적속성을 갖는 여러 개의 집락으로 구성되어 있을 때, 집락을 추출단위로 하는 단순집락추출법에 양적속성의 무관질문모형을 적용하였다. 그리고, 일정한 비용 하에서 분산을 최소로 하는 집락의 크기와 표본집락의 수의 최적값을 구하여 최소분산의 형태를 도출하였다. 또한, 제안한 단순집락추출법에 의한 무관질문모형과 단순임의추출법에 의한 무관질문모형과의 효율성을 비교해 보았다.

1. 서론

Warner(1965)는 확률장치를 통한 간접응답으로 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대해 정보를 이끌어 낼 수 있는 확률화응답모형(randomized response model ; RRM)을 처음으로 제시하였다. 그 후 Greenberg et al.(1971)은 무관질문모형(unrelated question model)을 제안하여 양적속성(quantitative attribute)을 갖는 경우로 확장하였으며, 이러한 양적속성의 확률화응답모형은 많은 학자들에 의해 연구, 발전되어 왔다. 특히, 이기성과 홍기학(1995)은 조사하고자 하는 모집단이 양적속성을 갖는 여러 개의 층으로 구성되어 각 층의 크기를 모르는 경우에, Greenberg et al.이 제안한 무관질문모형을 층화이중추출법에 적용하였다.

본 논문에서는 모집단이 양적속성을 갖는 여러 개의 집락으로 구성되어 있을 때, 집락을 추출단위로 하는 단순집락추출법에 양적속성의 무관질문모형을 적용해 보고자 한다. 2절에서는 민감한 속성에 대한 모평균의 추정량과 분산식을 구해 보고, 3절에서는 일정한 비용 하에서 분산을 최소로 하는 집락의 크기와 표본집락의 수의 최적값을 구하여 최소분산의 형태를 도출해 보고자 한다. 또한, 4절에서는 단순집락추출법에 의한 양적속성의 무관질문모형과 기존의 단순임의추출법에 의한 양적속성의 무관질문모형과의 효율성을 민감한 속성이 선택될 확률 p의 변화에 따라 수치적으로 비교해 보고자 한다.

2. 단순집락추출법에 의한 양적속성의 무관질문모형

모집단이 N개의 집락으로 구성되어 있는 경우에 n개의 집락을 단순임의추출하는 단순집락추출법에 양적속성의 무관질문모형을 적용해 보고자 한다. 이 때, 각 집락은 $M_i(i =$

† (565-701) 전라북도 완주군 삼례읍 후정리 490, 우석대학교 전산통계학과 조교수
‡ (520-714) 전라남도 나주시 대호동 252, 동신대학교 컴퓨터학과 부교수

$1, 2, \cdots, N$)개의 조사단위로 구성되어 있다. 각 집락은 민감한 변수 X가 연속인 밀도함수 $g_i(\cdot)$ $(i = 1, 2, \cdots, N)$를 갖는다고 가정하고, Y를 밀도함수 $h_i(\cdot)$를 갖는 무관속성이라 하자. 그리고, X의 모평균 $\mu_{X(cl)}$를 추정하는데 있어서 무관한 변수 Y의 모평균 μ_Y를 알고 있다고 가정하자.

응답자들은 민감한 변수 X가 선택될 확률이 p이고 무관한 변수 Y가 선택될 확률이 $q = 1 - p$인 확률장치를 통해 선택된 변수에 대해 응답하게 된다.

이 때, i번째 집락에서 j번째 응답자가 Z_{ij}라고 응답하면 Z_{ij}는 다음과 같은 확률밀도함수를 갖는다.

$$f(Z_{ij}) = pg_i(Z_{ij}) + qh_i(Z_{ij}), \ i = 1, 2, \cdots, N, \ j = 1, 2, \cdots, M_i. \tag{2.1}$$

그러므로, i번째 집락에서 기대할 수 있는 응답의 평균은 다음과 같다.

$$\mu_{Zi} = E_R(Z_{ij}) = p\mu_i + q\mu_{Yi}, \ i = 1, 2, \cdots, N. \tag{2.2}$$

여기서, μ_i는 i번째 집락에서의 민감한 그룹에 속하는 평균이고, μ_{Yi}는 i번째 집락에서의 무관한 변수 Y의 평균이다.

또한, i번째 집락에서의 응답의 분산은 다음과 같다.

$$\sigma_{Zi}{}^2 = p\sigma_i{}^2 + q\sigma_{Yi}{}^2 + pq(\mu_i - \mu_{Yi})^2. \tag{2.3}$$

여기서, $\sigma_i{}^2$은 i번째 집락에서의 X의 분산이고, $\sigma_{Yi}{}^2$은 i번째 집락에서의 Y의 분산이다.

i번째 집락에서 j번째 응답자가 응답한 확률화응답의 관찰치를 z_{ij}라 하면, i번째 집락에서의 민감한 그룹의 평균 μ_i의 추정량 $\hat{\mu}_i$는 다음과 같이 표현된다.

$$\hat{\mu}_i = \frac{1}{M_i p}\sum_{j=1}^{M_i}(z_{ij} - q\mu_{Yi}). \tag{2.4}$$

이 때, $\hat{\mu}_i$의 기대값과 분산은 다음과 같다.

$$\begin{aligned} E_R(\hat{\mu}_i) &= \mu_i \\ V_R(\hat{\mu}_i) &= \frac{1}{M_i p^2}\left[p\sigma_i{}^2 + q\sigma_{Yi}{}^2 + pq(\mu_i - \mu_{Yi})^2\right]. \end{aligned} \tag{2.5}$$

여기서, E_R과 V_R은 i번째 집락에서의 확률화응답에 대한 기대값과 분산의 표현이다.

단순집락추출법에 있어서 민감한 그룹에 속하는 조사단위당 모평균 $\mu_{X(cl)}$는 다음과 같다.

$$\mu_{X(cl)} = \frac{1}{N}\sum_{i=1}^{N}\mu_i. \tag{2.6}$$

따라서, 단순임의비복원추출된 표본 집락으로부터 민감한 그룹에 속하는 조사단위당 모평균 $\mu_{X(cl)}$의 추정량 $\hat{\mu}_{X(cl)}$는 다음과 같이 정의할 수 있다.

$$\begin{aligned}\hat{\mu}_{X(cl)} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{M_i p}\sum_{j=1}^{M_i}(z_{ij} - q\mu_{Yi})\right] \\ &= \frac{1}{np}\sum_{i=1}^{n}\left[\frac{1}{M_i}\sum_{j=1}^{M_i}(z_{ij} - q\mu_{Yi})\right]. \end{aligned} \tag{2.7}$$

만약 각 집락의 크기가 $M = M_i(i = 1, 2, \cdots, N)$로 일정하다면, 식 (2.7)로부터 추정량 $\hat{\mu}_{X(cl)}$는 다음과 같이 표현할 수 있다.

$$\hat{\mu}_{X(cl)} = \frac{1}{nMp}\sum_{i=1}^{n}\sum_{j=1}^{M}(z_{ij} - q\mu_{Yi}). \tag{2.8}$$

정리 2.1 추정량 $\hat{\mu}_{X(cl)}$는 모평균 $\mu_{X(cl)}$의 불편추정량이다.

증명:

$$\begin{aligned}E(\hat{\mu}_{X(cl)}) &= E_i E_R\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i\right) \\ &= E_i\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\mu_i \\ &= \mu_{X(cl)}. \end{aligned}$$

여기서, E_i는 단순임의비복원추출법으로 추출한 집락에 대한 기대값의 표현이다. □

정리 2.2 각 집락의 크기가 M_i인 N개의 집락에서 n개의 집락을 단순임의비복원추출한다. 이 때, 민감한 그룹에 속하는 조사단위당 모평균 $\mu_{X(cl)}$의 추정량 $\hat{\mu}_{X(cl)}$의 분산은 식 (2.9)와 같다.

$$\begin{aligned}V(\hat{\mu}_{X(cl)}) &= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2 \\ &\quad + \frac{1}{nNp^2}\sum_{i=1}^{N}\frac{1}{M_i}\left[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\right]. \end{aligned} \tag{2.9}$$

증명:

$$V(\hat{\mu}_{X(cl)}) = V_i E_R(\hat{\mu}_{X(cl)}) + E_i V_R(\hat{\mu}_{X(cl)}).$$

여기서, V_i는 단순임의비복원추출법으로 추출한 집락에 대한 분산의 표현이다.

$$\begin{aligned}
V_i E_R(\hat{\mu}_{X(cl)}) &= V_i E_R\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i\right) \\
&= V_i\left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right) \\
&= \frac{N-n}{N}\frac{1}{n(N-1)}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2 \\
&= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2,
\end{aligned}$$

$$\begin{aligned}
E_i V_R(\hat{\mu}_{X(cl)}) &= E_i V_R\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_i\right) \\
&= E_i\left[\frac{1}{n^2}\sum_{i=1}^{n}V_R(\hat{\mu}_i)\right] \\
&= E_i\left[\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{M_i p^2}\{p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\}\right] \\
&= \frac{1}{n}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M_i p^2}[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2] \\
&= \frac{1}{nNp^2}\sum_{i=1}^{N}\frac{1}{M_i}[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2].
\end{aligned}$$

□

또한, 각 집락의 크기가 M_i인 N개의 집락에서 n개의 집락을 단순임의복원추출할 때, 모평균 $\mu_{X(cl)}$의 추정량 $\hat{\mu}_{X(cl)}$의 분산은 식 (2.10)과 같다.

$$\begin{aligned}
V(\hat{\mu}_{X(cl)}) &= \frac{1}{nN}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2 \\
&+ \frac{1}{nNp^2}\sum_{i=1}^{N}\frac{1}{M_i}\left[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\right]. \qquad (2.10)
\end{aligned}$$

만약 식 (2.10)으로부터 각 집락의 크기가 $M = M_i (i = 1, 2, \cdots, N)$로 일정하다면, 추정

량 $\hat{\mu}_{X(cl)}$의 분산은 다음과 같이 표현할 수 있다.

$$V(\hat{\mu}_{X(cl)}) = \frac{1}{nN}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2 + \frac{1}{nMNp^2}\sum_{i=1}^{N}\left[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\right]. \tag{2.11}$$

3. M과 n의 최적값

식 (2.11)로부터 집락의 크기 M과 표본집락의 수 n을 증가시키면 분산은 감소하지만 M과 n의 증가에 따라 조사비용은 증가하게 된다. 따라서, 일정한 비용 하에서 표본의 정도를 최대로 하는 M과 n의 값을 결정해 보고자 한다.

먼저 비용함수를 고려해야 하는데, 조사비용은 집락의 크기와 표본집락의 수에 의하여 정하여지므로 비용함수는 대개 다음과 같은 형태를 취한다.

$$C = c_0 + nc_1 + nMc_2. \tag{3.1}$$

여기서, C는 총비용이고, c_0는 고정비용으로 조사행정비, 표본설계비용 등을 포함하며 표본의 크기와는 관계없이 소요되는 비용이다. c_1은 집락당 소요비용으로 표본집락의 선정, 집락간의 여비, 집락내 조사단위의 명부작성 등에 필요한 비용을 포함한다. c_2는 조사단위당 소요비용으로 확률장치를 이용한 면접 또는 실측비용, 조사단위간의 여비, 조사자료의 집계분석비용 등을 포함한다.

분산 식 (2.11)에서

$$S_1^{\,2} = \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu_{X(cl)})^2$$
$$S_2^{\,2} = \frac{1}{Np^2}\sum_{i=1}^{N}\left[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\right]$$

로 두면, $V(\hat{\mu}_{X(cl)})$는 다음과 같이 표현된다.

$$V(\hat{\mu}_{X(cl)}) = \frac{S_1^{\,2}}{n} + \frac{S_2^{\,2}}{nM}. \tag{3.2}$$

일정한 비용 하에서 분산을 최소로 하는 M과 n의 값을 식 (3.1)의 비용함수와 분산 식 (3.2)를 이용하여 구해 보기로 하자.

비용함수 식 (3.1)을 n의 함수로 표현해 보면 다음과 같다.

$$n = \frac{C - c_0}{c_1 + Mc_2}. \tag{3.3}$$

이 때, 식 (3.3)을 식 (3.2)에 대입하여 분산식을 다음과 같이 구할 수 있다.

$$V(\hat{\mu}_{X(cl)}) = \frac{c_1 + Mc_2}{C - c_0}\left({S_1}^2 + \frac{{S_2}^2}{M}\right)$$
$$= \frac{c_1 {S_1}^2}{C - c_0}\left[\left(M\frac{c_2}{c_1} + \frac{1}{M}\frac{{S_2}^2}{{S_1}^2}\right) + \left(1 + \frac{c_2}{c_1}\frac{{S_2}^2}{{S_1}^2}\right)\right]. \tag{3.4}$$

식 (3.4)를 M으로 미분하여 M의 최적값 M_0를 구하면 다음과 같다.

$$M_0 = \sqrt{\frac{{S_2}^2}{{S_1}^2}\frac{c_1}{c_2}}. \tag{3.5}$$

또한, 식 (3.5)의 M_0값을 식 (3.3)에 대입하여 n의 최적값 n_0를 구하면 다음과 같다.

$$n_0 = \frac{C - c_0}{c_1 + M_0 c_2}$$
$$= (C - c_0)\frac{\sqrt{{S_1}^2/c_1}}{\sqrt{{S_1}^2 c_1} + \sqrt{{S_2}^2 c_2}}. \tag{3.6}$$

따라서, 식 (3.5)와 식 (3.6)에서 구한 M_0와 n_0의 값을 식 (3.2)에 대입하여 최소분산 $V_{min}(\hat{\mu}_{X(cl)})$ 를 다음과 같이 구할 수 있다.

$$V_{min}(\hat{\mu}_{X(cl)}) = \frac{1}{n_0}\left({S_1}^2 + \frac{{S_2}^2}{M_0}\right)$$
$$= \frac{c_1 + M_0 c_2}{C - c_0}\left({S_1}^2 + \frac{{S_2}^2}{M_0}\right)$$
$$= \frac{1}{C - c_0}\left(\sqrt{{S_1}^2 c_1} + \sqrt{{S_2}^2 c_2}\right)^2. \tag{3.7}$$

4. 효율성 비교

단순집락추출법과 단순임의추출법과의 양적속성의 무관질문모형의 효율성을 비교해 보고자 한다.

크기가 nM개인 단순임의복원추출에 있어서 민감한 그룹에 속하는 모평균 μ_X의 추정량 $\hat{\mu}_X$의 분산은 다음과 같다.

$$V(\hat{\mu}_X) = \frac{1}{nMp^2}\left[p{\sigma_X}^2 + q{\sigma_Y}^2 + pq(\mu_X - \mu_Y)^2\right]. \tag{4.1}$$

여기서, ${\sigma_X}^2$은 민감한 변수 X의 모분산이고, ${\sigma_Y}^2$은 무관한 변수 Y의 모분산이다.

만약, 각 집락에서의 민감한 그룹에 속하는 평균이 전체 모평균과 비슷하다면, 분산 식 (2.11)의 첫 번째 항은 거의 0에 가깝게 되며, 식 (4.2)와 같이 표현될 수 있다.

$$V(\hat{\mu}_{X(cl)}) \doteq \frac{1}{nMNp^2}\sum_{i=1}^{N}\left[p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\right]. \tag{4.2}$$

따라서, 단순집락추출법에 의한 분산 식 (2.11)과 단순임의추출법에 의한 분산 식 (4.1)로부터 분산의 차이를 구해 보면 다음과 같다.

$$\begin{aligned} V(\hat{\mu}_X(cl)) - V(\hat{\mu}_X) &= \frac{1}{nMp^2}\left[\frac{1}{N}\sum_{i=1}^{N}\{p{\sigma_i}^2 + q{\sigma_{Yi}}^2 + pq(\mu_i - \mu_{Yi})^2\}\right. \\ &\qquad \left. -\{p{\sigma_X}^2 + q{\sigma_Y}^2 + pq(\mu_X - \mu_Y)^2\}\right] \\ &= \frac{1}{nMp^2}\left[p\left(\frac{1}{N}\sum_{i=1}^{N}{\sigma_i}^2 - {\sigma_X}^2\right) + q\left(\frac{1}{N}\sum_{i=1}^{N}{\sigma_{Yi}}^2 - {\sigma_Y}^2\right)\right. \\ &\qquad \left. +pq\left(\frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu_{Yi})^2 - (\mu_X - \mu_Y)^2\right)\right]. \end{aligned} \tag{4.3}$$

이 때, 식 (4.3)에 의한 해석학적인 비교가 어려우므로 수치적인 비교를 통하여 효율성을 비교해 보고자 한다.

p값의 변화에 따른 단순임의추출법과 단순집락추출법과의 효율성을 비교하기 위하여, $M = 100$, $N = 5$, $n = 2$일 때 $\mu_X = \mu_i$, $\mu_Y = \mu_{Yi}$라고 가정하고, ${\sigma_i}^2 = 10$ $(i = 1,2,3,4,5)$인 경우와 ${\sigma_i}^2 = 8$ $(i = 1,2,3,4,5)$인 경우로 나누어 ${\sigma_X}^2$을 변화시켜 가면서 분산비 $V(\hat{\mu}_X)/V(\hat{\mu}_{X(cl)})$를 계산하였다. 이 때, 표 4.1은 ${\sigma_Y}^2 = {\sigma_{Yi}}^2 = 10$ $(i = 1,2,3,4,5)$이라고 가정하였고, 표 4.2는 ${\sigma_{Yi}}^2 = 10$일 때, ${\sigma_Y}^2 = 11$, ${\sigma_Y}^2 = 9$라고 가정하여 작성하였다.

표 4.1과 표 4.2에서 1보다 큰 값은 단순집락추출법에 의한 양적속성의 무관질문모형이 단순임의추출법에 의한 양적속성의 무관질문모형보다 더 효율적임을 나타낸다.

표 4.1과 표 4.2에서 일반적으로 ${\sigma_i}^2$의 값이 ${\sigma_X}^2$의 값보다 작을수록, 또한 p값이 커질수록 단순집락추출법이 단순임의추출법보다 더 효율적인 경향을 띠고 있음을 알 수 있다. 이는 집락의 크기가 같고 집락간 분산들이 서로 동질적이면서 작을수록 단순집락추출법이 단순임의추출법보다 효율적이라는 일반적인 표본이론과 일치하고 있다.

표 4.1: ${\sigma_Y}^2 = {\sigma_{Yi}}^2 = 10$일 때, p값의 변화에 따른 단순임의추출과 단순집락추출과의 분산비 비교

p	${\sigma_i}^2 = 10$			${\sigma_i}^2 = 8$		
	${\sigma_X}^2 = 9$	${\sigma_X}^2 = 11$	${\sigma_X}^2 = 13$	${\sigma_X}^2 = 9$	${\sigma_X}^2 = 11$	${\sigma_X}^2 = 13$
0.1	0.99	1.01	1.03	1.01	1.03	1.05
0.2	0.98	1.02	1.06	1.02	1.06	1.10
0.3	0.97	1.03	1.09	1.03	1.10	1.16
0.4	0.96	1.04	1.12	1.04	1.13	1.22
0.5	0.95	1.05	1.15	1.06	1.17	1.28
0.6	1.94	1.06	1.18	1.07	1.20	1.34
0.7	0.93	1.07	1.21	1.08	1.24	1.41
0.8	0.92	1.08	1.24	1.10	1.29	1.48
0.9	0.91	1.09	1.27	1.11	1.33	1.55

표 4.2: $\sigma_{Yi}^2 = 10(\sigma_Y^2 = 11,\ \sigma_Y^2 = 9)$일 때, p값의 변화에 따른 단순임의추출과 단순집락추출과의 분산비 비교

p	$\sigma_i^2 = 10,\ \sigma_Y^2 = 11$			$\sigma_i^2 = 8,\ \sigma_Y^2 = 9$		
	$\sigma_X^2 = 9$	$\sigma_X^2 = 11$	$\sigma_X^2 = 13$	$\sigma_X^2 = 9$	$\sigma_X^2 = 11$	$\sigma_X^2 = 13$
0.1	1.08	1.1	1.12	0.92	0.94	0.96
0.2	1.06	1.1	1.14	0.94	0.98	1.02
0.3	1.04	1.1	1.16	0.96	1.02	1.09
0.4	1.02	1.1	1.18	0.98	1.07	1.15
0.5	1.00	1.1	1.20	1.00	1.11	1.22
0.6	0.98	1.1	1.22	1.02	1.16	1.30
0.7	0.96	1.1	1.24	1.05	1.21	1.37
0.8	0.94	1.1	1.26	1.07	1.26	1.45
0.9	0.92	1.1	1.28	1.10	1.32	1.54

5. 결론

모집단이 양적속성을 갖는 여러 개의 집락으로 구성되어 있을 때, 집락을 추출단위로 하는 단순집락추출법에 양적속성의 무관질문모형을 적용하여 민감한 속성에 대한 모평균의 추정량과 분산식을 구하였다. 그리고, 일정한 비용 하에서 분산을 최소로 하는 집락의 크기와 표본집락의 수의 최적값을 구하여 최소분산의 형태를 도출하였다. 또한, 단순집락추출법에 의한 양적속성의 무관질문모형과 기존의 단순임의추출법에 의한 양적속성의 무관질문모형의 효율성을 민감한 속성이 선택될 확률 p의 변화에 따라 수치적으로 비교하였다. 그 결과, 일반적으로 σ_i^2의 값이 σ_X^2의 값보다 작을수록, 또한 p값이 커질수록 단순집락추출법이 단순임의추출법보다 더 효율적인 경향을 띠고 있음을 알 수 있었다.

참고문헌

[1] 류제복, 홍기학, 이기성(1993). ⟨확률화응답모형⟩, 자유아카데미.

[2] 박홍래(1989). ⟨통계조사론⟩, 영지문화사.

[3] 이기성, 홍기학(1995). 층화이중추출법에 의한 양적속성의 무관질문모형, ⟨응용통계연구⟩, 제 8권 1호, 27-38.

[4] Chauduri, A. and mukerjee, R.(1988). *Randomized Response : Theory and Techniques*, Marcel Dekker, Inc., New York.

[5] Cochran, W. G.(1977). *Sampling Techniques*, 3rd ed. John Wiley & Sons, New York.

[6] 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*, vol. 66, 243-250.

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

[1997년 4월 접수, 1997년 9월 최종수정]

Unrelated Question Model with Quantitative Attribute by Simple Cluster Sampling

Gi-Sung Lee[†], Ki-Hak Hong[‡]

ABSTRACT

In this paper, we developed one-stage cluster randomized response model for obtaining quantitative data by using the Greenberg et al. model(1971) when the population was made up of sensitive quantitative clusters. We obtained the minimum variance by calculating the cluster's size and the optimum number of sample clusters under the some given constant cost. We compared the efficiency of our model with the Greenberg et al. model by simple random sampling.

[†] Department of Computer Science and Statistics, Woosuk University, Wanju-gun, Chonbuk 565-701, Korea.

[‡] Department of Computer Science, Dongshin University, Daeho-dong, Naju, Chonnam 520-714, Korea.