한국통계학회논문집 제1권1호
1994년 12월 pp. 157-164

# 확률화응답기법을 이용한 모비율의 추정시 층화표본의 최적할당에 관한 연구[1)]

최 경 호[2)] 김 연 형[3)]


## 요 약

본 연구에서는 확률화응답기법을 이용하여 모집단내의 민감집단의 비율을 추정함에 있어 조사의 효율성을 높이기 위한 층화표본의 최적할당방법을 제안한다. 확률화응답기법은 Warner(1965)에 의하여 제안된 방법으로 민감한 사안에 대한 조사시 무응답이나 거짓응답으로 인한 비표본오차를 줄일수 있는 기법으로 간접질문에 의한 조사방법이다. 여기에서 최적할당이란 베이즈위험을 최소로 하는 할당법을 의미하며, 이 과정에서 민감집단의 모비율에 대한 사전분포로는 베타분포를 취하였다.


## 1. 서 론

개인의 사생활과 관련된 문제(민감한 문제)에 대한 조사시 - 예컨대 A.I.D.S나 동성연애·약물중독·혼전성경험·낙태여부·탈세 등 - 직접질문방식을 사용할 경우 응답자는 응답을 회피하거나, 또는 거짓으로 응답하는 경향이 많아 비표본오차의 증대를 가져와 추정의 효율이 떨어진다.

그래서 민감한 문제에 대한 조사시 응답자의 신분보호의 정도를 높이며 응답자가 정직하게 응답하므로써 비표본오차의 발생을 줄일 수 있는 조사기법으로 1965년 Warner는 확률화응답기법을 제시 하였다. 이 기법은 확률장치를 통하여 간접적인 응답만을 요구하므로써 결과적으로 응답오차를 줄이고 응답자의 조사에 대한 호응을 높일 수 있는 조사 기법이다. 나아가 확률화응답기법을 이용한 조사시 사전정보를 이용하는 베이지안 접근법에 대해서는 Winkler와 Franklin(1979)에 의해서 연구 되었는데, 이들은 사전분포로 베타분포를 취하였다.

Warner는 그의 논문(1965)에서 고려되는 모집단을 상호배반적인 두 집단, 즉 민감집단 A와 민감하지 않는 집단 $\overline{A}$로 가정하고 확률화응답기법을 이용하여 민감집단의 비율 $\pi$의 최대우도추정량과 그의 분산을 유도하였다. 한편 그는 모집단으로부터 추출된 표본내의 응답자는 확률장치를 통한 응답시 거짓응답이 없으며, 표본의 추출법으로는 단순임의복원추출법을 가정하였다. 그런데 조사시 표본추출의 방법을 단순임의추출에 의했을때 보다 모집단을 몇개의 층으로 나누어 층화추출에 의했을때 추정의 효율이 증가됨을 알 수 있다.

본 연구에서는 확률화응답기법을 이용하여 이속성 모집단내의 민감집단의 비율추정시 $\pi$의 사전분포를 베타분포로 가정했을때 베이즈-위험을 최소로 하는 층화표본의 최적할당에 관해서 알아보고자 한다.

---

1) 이 연구는 1994년도 전주대학교 학술연구비 지원에 의하여 수행되었습니다. 건설적인 조언을 주신 공군사관학교 이계오교수님께 감사드립니다.
2)(560-759) 전주시 완산구 효자동 1200 전주대학교 통계학과
3)(560-759) 전주시 완산구 효자동 1200 전주대학교 통계학과

2절에서는 층화표본을 이용한 확률화응답기법의 적용시 베이지안 접근법을 통한 $\pi$의 베이즈 추정량에 대해 언급하고, 3절에서는 주어진 총조사비용하에서 베이즈 위험을 최소로 하는 최적할당방법을 찾았다. 제4절에서는 간단한 예제를 통하여 앞절에서 찾은 최적할당에 대해 살펴보고 이에따른 토의를 하였다.

## 2. 베이즈 추정량

크기가 N인 이속성 모집단내의 민감집단의 비율 $\pi$를 Winkler와 Franklin이 제시한 확률화응답기법을 이용한 베이지안 접근법을 통하여 추정함에 있어, 단순임의복원추출된 표본대신 층화표본을 이용하는 경우를 생각해 보자. 이때 모집단에 대해서는 다음을 가정한다.

첫째 : 모집단은 동질의 구성요소로 되어 있고,
둘째 : k개의 부모집단으로 모집단이 분할 될 수 있고,
세째 : 부모집단내의 민감집단의비율 $\pi_h(h=1,\cdots,k)$는 일정한 확률분포를 갖는다.

이제 우리의 목적은 위와같은 조건하에 있는 모집단에서 층화표본을 추출하여 모집단내의 민감집단의 비율을 추정함에 있어 제한된 조사비용하에서 베이즈-위험을 최소로 하는 할당법, 즉 최적할당법을 찾고자 함이다. 이를 행함에 있어 각 층으로 부터의 표본추출은 단순임의추출을 가정하고, h번째 층에 대하여 민감질문이 선택될 확률은 $P_h$이며 $n_h$명의 응답자중 "예"라고 응답한 응답자의 수는 $r_h$, 그리고 $\pi_h$의 사전분포는 $B(\alpha_h', \beta_h')$ 로 한다.

먼저 h번째층(h=1, 2, ⋯, k)에 대해서 생각해 보자. $\pi_h$의 사전분포가 베타분포임을 이용하여 이의 사후분포를 찾고, $\pi_h$의 베이즈 추정량은 사후분포의 평균임을 이용하면 h번째 층내의 민감집단의 비율 $\pi_h$에 대한 베이즈 추정량 $\widehat{\pi_h^B}$는 다음과 같다.

$$\begin{aligned}\widehat{\pi_h^B} = E(\pi_h|r_h,n_h) &= \int \pi_h f(\pi_h|r_h,n_h)d\pi_h \\ &= \int \pi_h \sum_{t_h=0}^{n_h} w_{t_h} f_\beta(\pi_h|\alpha_h'+t_h,\ \beta_h'+n_h-t_h)d\pi_h \\ &= \sum_{t_h=0}^{n_h} \frac{(\alpha_h'+t_h)w_{t_h}}{\alpha_h'+\beta_h'+n_h} \qquad (2.1)\end{aligned}$$

여기서, $w_{t_h} = w^*_{t_h} / \sum_{s_h=0}^{n_h} w^*_{s_h}$,

$$w^*_{t_h} = \binom{n_h}{t_h}\frac{B(\alpha'_h+t_h,\ \beta'_h+n_h-t_h)}{B(\alpha'_h,\ \beta'_h)}\sum_{j=0}^{\min(r_h,t_h)}\binom{t_h}{j}\binom{n_h-t_h}{r_h-j}P_h^{n_h-t_h-r_h+2j}(1-P_h)^{t_h+r_h-2j}$$

이다. 더우기, $w_{t_h}$는 $\pi_h$의 사후분포에서 나타나는 가중치이며 $t_h$는 관측될 수 없는 값으로 h번째층내의 응답자중 민감집단에 속하는 응답자의 수가된다.

한편 h번째 층의 크기를 $N_h$ 라면($\sum_{h=1}^{k} N_h = N$), 층화추출에 의한 모집단내의 민감 집단의비율 $\pi$의 베이즈 추정량, $\widehat{\pi}_{st}^{B}$, 은 다음과 같다.

$$\widehat{\pi}_{st}^{B} = \sum_{h=1}^{k}\left(\frac{N_h}{N}\right)\widehat{\pi}_{h}^{B}$$

$$= \sum_{h=1}^{k}\left(\frac{N_h}{N}\right)\sum_{t_h=0}^{n_h}\frac{(\alpha_h' + t_h)w_{t_h}}{\alpha' + \beta' + n_h} \tag{2.2}$$

만약 손실함수를 제곱형태로 택하면, h번째 층에서의 베이즈 위험, $\rho_h$,은 $\pi_h$의 사후분포의 분산에 대한 기대값과 같다. 즉,

$$\rho_h = E[V(\pi_h | r_h, n_h)] \tag{2.3}$$

이다.

식 (2.3)을 계산하기 위하여 $\pi_h$의 사후분포를 Winkier와 Franklin(1979)이 제안한 근사형태로 취하면 이는 다음의 $\alpha_h''$ 과 $\beta_h''$ 을 모수로 갖는 베타분포 이므로

$$V(\pi_h | r_h, n_h) = \frac{\alpha_h''\beta_h''}{(\alpha_h'' + \beta_h'')^2(\alpha_h'' + \beta_h'' + 1)} \tag{2.4}$$

이다. 여기서, $\alpha_h''$ 과 $\beta_h''$ 은 다음과 같다.

$$\alpha_h'' = \left[\left(\frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h} - 1 + P_h\right)^2\left(P_h - \frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h}\right)\right.$$
$$\left. - \left(\frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h} - 1 + P_h\right)\left(\frac{(\alpha_h^* + r_h)(\beta_h^* + n_h - r_h)}{(\alpha_h^* + \beta_h^* + n_h)^2(\alpha_h^* + \beta_h^* + n_h + 1)}\right)\right]$$
$$\times\left[\frac{1}{(2P_h - 1)}\cdot\frac{(\alpha_h^* + \beta_h^* + n_h)^2(\alpha_h^* + \beta_h^* + n_h + 1)}{(\alpha_h^* + r_h)(\beta_h^* + n_h - r_h)}\right] \tag{2.5}$$

$$\beta_h'' = \left[\left(\frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h} - 1 + P_h\right)\left(P_h - \frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h}\right)^2\right.$$
$$\left. - \left(P_h - \frac{\alpha_h^* + r_h}{\alpha_h^* + \beta_h^* + n_h}\right)\left(\frac{(\alpha_h^* + r_h)(\beta_h^* + n_h - r_h)}{(\alpha_h^* + \beta_h^* + n_h)^2(\alpha_h^* + \beta_h^* + n_h + 1)}\right)\right]$$
$$\times\left[\frac{1}{(2P_h - 1)}\cdot\frac{(\alpha_h^* + \beta_h^* + n_h)^2(\alpha_h^* + \beta_h^* + n_h + 1)}{(\alpha_h^* + r_h)(\beta_h^* + n_h - r_h)}\right] \tag{2.6}$$

한편, h번째 층에서 $n_h$명의 응답자에 대해서 "예"라는 응답을 얻을 확률을 $\lambda_h = P_h\pi_h + (1 - P_h)(1 - \pi_h)$라면,

$$\alpha_h^* = E^2(\lambda_h)[1-E(\lambda_h)][V(\lambda_h)]^{-1} - E(\lambda_h)$$
$$\beta_h^* = E(\lambda_h)[1-E(\lambda_h)]^2[V(\lambda_h)]^{-1} - [1-E(\lambda_h)]$$

이며,

$$E(\lambda_h) = (2P_h-1)E(\pi_h)+(1-P_h)$$
$$V(\lambda_h) = (2P_h-1)^2V(\pi_h)$$

이다. 또한

$$E(\pi_h) = \frac{\alpha_h'}{\alpha_h'+\beta_h'}$$

$$V(\pi_h) = \frac{\alpha_h'\beta_h'}{(\alpha_h'+\beta_h')^2(\alpha_h'+\beta_h'+1)}$$

따라서, 식 (2.5)와 (2.6)으로부터

$$V(\pi_h|r_h,n_h) = \frac{\alpha_h^*(\beta_h^*+n_h)+(\beta_h^*+n_h-\alpha_h^*)r_h-r_h^2}{(2P_h-1)^2(\alpha_h^*+\beta_h^*+n_h)^2(\alpha_h^*+\beta_h^*+n_h+1)} \tag{2.7}$$

이다.

한편 $r_h$의 주변확률분포는 근사적으로 Beta-Binomial분포, BB($\alpha_h^*, \beta_h^*$),를 따르므로 BB분포의 모멘트로 부터,

$$E(r_h) = n_h\frac{\alpha_h^*}{\alpha_h^*+\beta_h^*} \tag{2.8}$$

$$E(r_h^2) = n_h(n_h-1)\frac{(\alpha_h^*+1)\alpha_h^*}{(\alpha_h^*+\beta_h^*+1)(\alpha_h^*+\beta_h^*)}+n_h\frac{\alpha_h^*}{\alpha_h^*+\beta_h^*} \tag{2.9}$$

이다. 따라서 식 (2.7), (2.8) 그리고 (2.9)으로부터 식 (2.3)은 다음과 같다.

$$\begin{aligned}\rho_h = &\frac{1}{(2P_h-1)^2(\alpha_h^*+\beta_h^*+n_h)^2(\alpha_h^*+\beta_h^*+n_h+1)}\\ &\times\left[\alpha_h^*(\beta_h^*+n_h)+(\beta_h^*+n_h-\alpha_h^*)\frac{\alpha_h^*}{\alpha_h^*+\beta_h^*}n_h\right.\\ &\left.-\frac{(\alpha_h^*+1)\alpha_h^*}{(\alpha_h^*+\beta_h^*+1)(\alpha_h^*+\beta_h^*)}n_h(n_h-1)-\frac{\alpha_h^*}{\alpha_h^*+\beta_h^*}n_h\right]\end{aligned} \tag{2.10}$$

이제, $r_h$( $h=1$ , … , $k$ )가 서로 독립이라면 유한실수 $a_h$에 대해 이차(quadratic) 제곱형태의 손실함수를 사용하여 $D = \sum_{h=1}^{k}a_h\pi_h$(단, $a_h=N_h/N$는 유한실수)를 추정하는데 있어서 베이즈

위험 $\rho$은 다음과 같다.

$$\rho = \sum_{h=1}^{k} a_h^2 \rho_h \tag{2.11}$$

## 3. 최 적 할 당

h번째 층에서의 표본조사 단위비용을 $c_h$라 하고, C를 표본조사 총비용이라 하자. 그리고 층화추출에서 각 층에 할당된 표본의 크기가 $n_1, n_2, \cdots, n_k$일때 다음의 선형관계를 가정하자. 즉,

$$C = \sum_{h=1}^{k} c_h n_h \tag{3.1}$$

그러면 식 (2.11)과 (3.1)을 이용하여 우리가 찾고자 하는 최적할당은 다음의 라그랑쥐 승수법(Largrange multiplier method)으로 부터 구할 수 있다.

$$\phi(n_h) = \sum_{h=1}^{k} a_h^2 \frac{1}{(2P-1)^2(\alpha_h^* + \beta_h^* + n_h)^2(\alpha_h^* + \beta_h^* + n_h + 1)} [\alpha_h^*(\beta_h^* + n_h)$$

$$+ n_h \frac{\alpha_h^*}{\alpha_h^* + \beta_h^*}(\beta_h^* + n_h - \alpha_h^*) - n_h(n_h - 1)\frac{\alpha_h^*(\alpha_h^* + 1)}{(\alpha_h^* + \beta_h^*)(\alpha_h^* + \beta_h^* + 1)}$$

$$- n_h \frac{\beta_h^*}{\alpha_h^* + \beta_h^*}] + \triangle(\sum_{h=1}^{k} c_h n_h - C)$$

$$\frac{\partial\phi(n_h)}{\partial n_h} = - \frac{a_h^2 \alpha_h^* \beta_h^*}{(2P_h - 1)^2(\alpha_h^* + \beta_h^*)(\alpha_h^* + \beta_h^* + 1)(\alpha_h^* + \beta_h^* + n_h)^2} + \triangle c_h$$

$$= 0$$

이로부터,

$$(\alpha_h^* + \beta_h^* + n_h) = \frac{1}{(2P_h - 1)}\sqrt{\frac{1}{\triangle}}\frac{|a_h|}{\sqrt{c_h}}\sqrt{\frac{\alpha_h^* \beta_h^*}{(\alpha_h^* + \beta_h^*)(\alpha_h^* + \beta_h^* + 1)}}$$

$$\therefore n_h = \frac{1}{(2P_h - 1)}\sqrt{\frac{1}{\triangle}}\frac{|a_h|}{\sqrt{c_h}}\sqrt{\frac{\alpha_h^* \beta_h^*}{(\alpha_h^* + \beta_h^*)(\alpha_h^* + \beta_h^* + 1)}} - (\alpha_h^* + \beta_h^*) \tag{3.2}$$

한편, $C = \sum_{h=1}^{k} c_h n_h$임을 이용하면

$$C = \sum_{h=1}^{k} \frac{1}{(2P_h - 1)}\sqrt{\frac{1}{\triangle}}\sqrt{c_h}\,|a_h|\sqrt{\frac{\alpha_h^* \beta_h^*}{(\alpha_h^* + \beta_h^*)(\alpha_h^* + \beta_h^* + 1)}} - \sum_{h=1}^{k} c_h(\alpha_h^* + \beta_h^*)$$

이다. 그래서

$$C+\sum_{h=1}^{k}c_h(\alpha_h^*+\beta_h^*)=\frac{1}{(2P_h-1)}\sqrt{\frac{1}{\triangle}}\sum_{h=1}^{k}\sqrt{c_h}\,|a_h|\sqrt{\frac{\alpha_h^*\beta_h^*}{(\alpha_h^*+\beta_h^*)(\alpha_h^*+\beta_h^*+1)}}$$

이다. 따라서

$$\frac{1}{(2P_h-1)}\sqrt{\frac{1}{\triangle}}=\frac{C+\sum_{h=1}^{k}c_h(\alpha_h^*+\beta_h^*)}{\sum_{h=1}^{k}\sqrt{c_h}\,|a_h|\sqrt{\frac{\alpha_h^*\beta_h^*}{(\alpha_h^*+\beta_h^*)(\alpha_h^*+\beta_h^*+1)}}}$$

이며, 이를 식 (3.2)에 대입하여 다음의 최적할당 $n_h$를 찾을수 있다.

$$n_h=\frac{[C+\sum_{h=1}^{k}c_h(\alpha_h^*+\beta_h^*)]}{\sum_{h=1}^{k}\sqrt{c_h}\,|a_h|\sqrt{\frac{\alpha_h^*\beta_h^*}{(\alpha_h^*+\beta_h^*)(\alpha_h^*+\beta_h^*+1)}}}\frac{|a_h|}{\sqrt{c_h}}\sqrt{\frac{\alpha_h^*\beta_h^*}{(\alpha_h^*+\beta_h^*)(\alpha_h^*+\beta_h^*+1)}}-(\alpha_h^*+\beta_h^*) \tag{3.3}$$

단, 여기서

$$\alpha_h^*=\left[(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}+(1-P_h)\right]^2\left[P_h-(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}\right]\times\left[(2P_h-1)^2\alpha_h'\frac{\beta_h'}{(\alpha_h'+\beta_h')^2(\alpha_h'+\beta_h'+1)}\right]^{-1}-\left[(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}+(1-P_h)\right] \tag{3.4}$$

$$\beta_h^*=\left[(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}+(1-P_h)\right]\left[P_h-(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}\right]^2\times\left[(2P_h-1)^2\alpha_h'\frac{\beta_h'}{(\alpha_h'+\beta_h')^2(\alpha_h'+\beta_h'+1)}\right]^{-1}-\left[P_h-(2P_h-1)\frac{\alpha_h'}{\alpha_h'+\beta_h'}\right] \tag{3.5}$$

## 4. 예제 및 토의

예를 통하여 살펴보자. $N_1=100, N_2=200, N_3=300, N_4=500, N_5=1000$이고, 총조사비용이 2000이며 단위당 조사비용이 Ⅰ$(C_1=5, C_2=4, C_3=3, C_4=2, C_5=1)$, Ⅱ$(C_1=10, C_2=8, C_3=6, C_4=4, C_5=2)$, Ⅲ$(C_1=50, C_2=40, C_3=30, C_4=20, C_5=10)$, Ⅳ$(C_1=75, C_2=60, C_3=45, C_4=30, C_5=15)$일때, $\alpha_h'=\alpha'$, $\beta_h'=\beta'$ 그리고 $P_h=P$에 따른 최적할당을 찾아보면 표1과 같다.

표1. $\alpha'$, $\beta'$ 그리고 표본조사 단위비용에 따른 최적할당.

| 단위비용 \ 할당 | | $\alpha'$ = 0.5 | 2 | 10 | 20 |
|---|---|---|---|---|---|
| | | $n_1$ $n_2$ $n_3$ $n_4$ $n_5$ | $n_1$ $n_2$ $n_3$ $n_4$ $n_5$ | $n_1$ $n_2$ $n_3$ $n_4$ $n_5$ | $n_1$ $n_2$ $n_3$ $n_4$ $n_5$ |
| $\beta'$ = 2 | I | 31 71 123 252 | 32 72 125 255 722 | 34 76 132 270 764 | 36 81 141 289 817 |
| | II | 16 36 63 128 363 | 16 37 64 131 371 | 18 41 71 146 413 | 20 46 80 164 466 |
| | III | 3 8 14 29 83 | 4 9 15 32 91 | 5 13 23 47 133 | 8 18 32 65 185 |
| | IV | 2 5 10 21 59 | 3 6 11 23 67 | 4 10 19 38 109 | 7 16 28 57 162 |
| 4 | I | 32 72 125 256 | 32 73 127 359 733 | 34 77 134 274 775 | 37 82 143 292 827 |
| | II | 16 37 64 132 374 | 17 38 66 135 382 | 18 42 73 150 424 | 21 47 82 168 477 |
| | III | 4 9 16 33 93 155 | 4 10 17 35 101 | 6 14 24 50 143 | 8 19 34 69 196 |
| | IV | 3 7 12 24 70 | 3 7 13 27 78 | 5 12 20 42 120 | 7 17 29 61 173 |
| 10 | I | 33 75 131 267 | 34 76 132 270 764 | 36 80 139 285 806 | 38 85 148 303 859 |
| | II | 18 40 70 143 406 | 18 41 74 146 413 | 20 45 78 161 456 | 22 50 88 179 508 |
| | III | 5 12 21 44 125 | 5 13 23 47 133 | 7 17 30 62 175 | 10 22 39 80 228 |
| | IV | 4 10 17 36 102 | 4 10 19 39 109 | 6 15 26 53 152 | 9 20 35 72 204 |
| 20 | I | 80 140 286 809 | 36 81 141 289 817 | 38 85 148 303 859 | 40 91 157 322 912 |
| | II | 20 45 79 162 458 | 20 46 80 164 466 | 22 50 88 179 508 | 25 56 97 198 561 |
| | III | 7 17 30 62 178 | 8 18 32 65 185 | 10 22 39 80 228 | 12 28 48 99 280 |
| | IV | 6 15 26 54 154 | 7 16 28 57 162 | 9 20 35 72 204 | 11 25 44 90 257 |

한편 P=0.6인 경우 표본조사 단위비용과 사전분포 즉 $\alpha'$, $\beta'$ 에 따른 베이즈 위험을 살펴보면 그림 1과 같다. 이로부터 우리는 총조사비용이 고정된 상태에서 표본조사 단위비용이 증가하면 베이즈 위험이 커진다는 것을 알 수 있다. 특히 $\alpha'=\beta'=2$ 일때와 $\alpha'=\beta'=20$ 일때의 베이즈 위험을 보면 $\alpha'=\beta'=2$ 일때의 베이즈 위험량이 훨씬 크다는 것을 알 수 있다. 우리는 보통, 확률화응답기법을 이용하여 모집단내의 민감집단의 비율을 추정할때 사전정보의 이용없이 이를 추정하는데 이보다는 사전정보를 이용하는 베이지안 접근법을 통한 추정이 더 의미가 있겠고 나아가 층화표본을 이용할 경우 본 연구에서 제시한 최적할당을 통한 층화표본을 이용함으로써 추정의 효율을 높일 수 있을것으로 기대된다. 더우기 이차 손실함수에 대해 Warner의 가정대로 조사과정에서 거짓응답이 없다면, 베이즈위험을 급격히 줄일 수 있도록 하기 위해서는 P를 증가시키면 되는데 이는 신분보호(privacy protection)의 정도를 떨어뜨릴 위험성을 내포하게 된다.

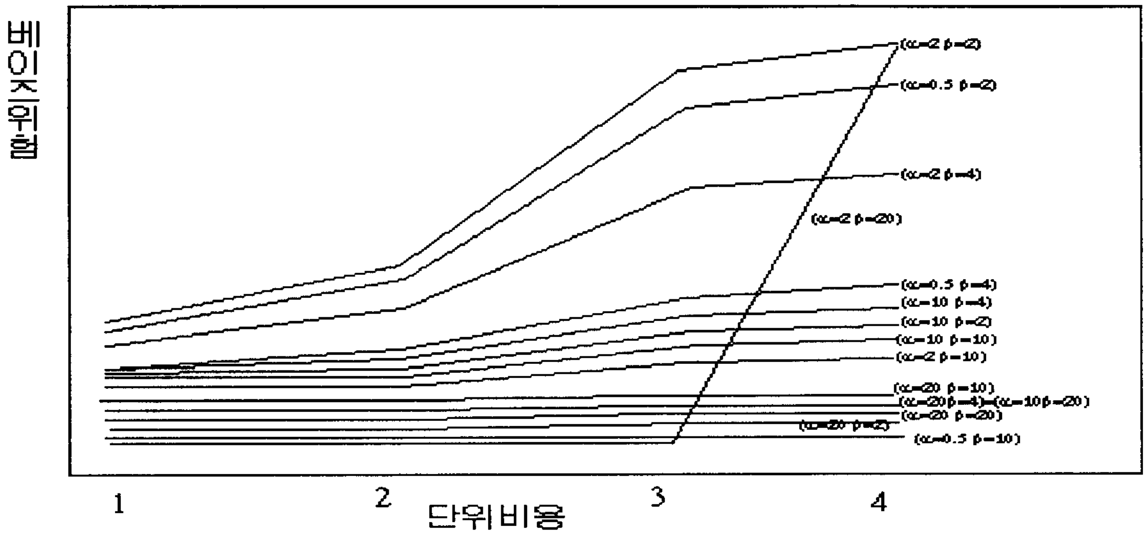


그림1. P=0.6일때 $\alpha'$, $\beta'$ 그리고 표본조사 단위비용에 따른 베이즈위험

## 참고 문헌

[1] Chatterjee, S. (1972), "A study of optimum allocation in multivariate stratified surveys. "Srand. Aktuar Tidskr, 73-80.

[2] Gunel, G. (1985). "A bayesian comparison of randomized and voluntary response sampling model." Communications in statistics - theory and methods, 14(10), 2411-2435.

[3] Hansen, M. H., Hurwitz, W. N., Madow, W. G. (1953). Sample Survey Methods and Theory, vol. I & II, John Wiley & sons, New York.

[4] Ishii, G., Hayakawa, R. (1960). "On the compound binomial distribution, "Institute of Statistical Mathmatics, (Tokyo) Annals, vol. 12, 69-80.

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

[6] Warner, S. L. (1965), "Randomized response : A study technique for eliminating evasive answer bias," Journal of the American Statistical Association, 60, 63-69.

[7] Winker, R. L., Franklin, L. A. (1979), "Warner's randomized response model : A bayesian approach," Journal of the American Statistical Association, 74, 207-214.