

절사층 총합추정을 위한 복합추정량

황종민^a, 신기일^{1,a}

“한국의국어대학교 통계학과

요약

절사표본 추출법은 절사층, 표본층, 전수층으로 모집단을 분리한 후 표본층과 전수층의 조사결과를 이용하여 전체 모집단의 총합을 추정하는 방법이다. 이 방법은 왜도가 심한 사업체조사에서 흔히 사용하는 방법이다. 절사층의 총합 추정은 전체 모집단 총합 추정에 영향을 미치므로 절사층 총합의 정확한 추정은 매우 중요하다. 최근 김지학과 신기일 (2011)은 절사층에서 소수의 표본을 추출하여 얻은 결과와 기존의 추정량에서 얻은 결과를 선형결합하는 복합추정법을 제안하였다. 본 논문에서는 최량선형불편예측(best linear unbiased predictor; BLUP)을 이용한 새로운 복합추정량을 제안하였으며 모의실험을 통하여 기존의 방법과 새로운 복합추정량의 우수성을 비교하였다.

주요용어: 표본조사, 비표본오차, 절사표본, 비추정.

1. 서론

모집단의 일부를 제외한 후 표본설계 하는 방법인 절사표본설계는 관심변수의 왜도가 큰 경우에 주로 사용되는 방법이다. 많은 경우 사업체조사의 왜도가 크기 때문에 사업체조사에서 흔히 사용하는 방법이다. 특히 사업체조사에서는 소규모 사업체의 휴업과 폐업이 자주 발생하여 모집단 또는 표본틀 관리가 어렵다. 또한 전체 총합에서 소규모 사업체가 차지하는 비중이 작기 때문에 이 부분을 조사에게 제외하여 조사의 효율성을 높이고 있다. 그러나 정확한 총합추정을 위해서는 조사에서 제외된 절사층 총합추정을 정확히 하는 것이 중요하다.

Hidiroglou (1986)는 수정절사법(modified cut-off sampling)을 제안하였으며 완전히 조사하지 않은 절사층(take-nothing stratum) 없이 표본층과 전수층 만을 고려한 방법을 제안하였다. 이 방법은 신민웅과 이상은 (2001)에 자세히 설명되어 있다. 최근 Lee (2011)은 수정절사법의 절사점 추정방법에 기초하여 주어진 허용오차를 만족하는 최소의 표본 수가 주어졌을 때 MSE(mean squared error)를 최소화 하는 절사점을 구하는 방법을 제안하였다. 또한 Lee와 Shin (2011)은 이에 관한 이론과 그 타당성을 제시하였다.

다른 한편에서는 절사층을 조사하지 않고 버려진 절사층의 총합추정을 위한 연구가 많이 진행되었다. 이러한 연구의 대부분은 행정자료나 다른 조사에서 이미 보조변수가 얻어졌다는 것을 가정하고 있다. 먼저 Elisson과 Elvers (2001)는 스웨덴의 기업체 매출액 조사에 절사법을 적용하였다. 이때 행정자료에서 얻어진 보조변수와 표본층의 관심변수와 보조변수를 이용한 비추정량을 이용하여 절사층의 총합을 추정하였다. Clark과 Kinyon (2007)은 미국 통계국(US. Census Bureau)의 2005년 소매업 조사(Annual Retail Trade Survey; ARTS)와 서비스업조사(Service Annual Survey; SAS)의 새로운 표본설계를 실시하면서 응답자 부담과 자료 수집비용을 절감하는 방안으로 절사법을 연구하였다. 또한

¹ 이 논문은 2011년도 한국의국어대학교 학술연구비 지원에 의해 이루어진 것임.

¹ 교신저자: (449-791) 경기도 용인시 모현면 산 89, 한국의국어대학교 통계학과, 교수. E-mail: keyshin@hufs.ac.kr

McDowney (2004)는 모형기반이론(model-based theory) 또는 예측이론을 통한 절사추정법을 이용하여 미국에너지정보국(Energy Information Administration; EIA)의 월별 전기 매출액 및 전기 발전 자료의 총 매출액을 추정하였다.

이와 같이 여러 연구에서 절사법이 사용되고 있으나 절사층의 총합을 추정하는데 있어 비록 조사 부담성 및 응답 부담성 등을 고려하더라도 완전히 조사를 하지 않은 것에 비해 소수의 표본을 절사층에 포함하여 조사하였을 경우 더 좋은 결과가 얻어질 수 있을 것으로 판단된다.

이에 본 연구에서는 절사층에서 소수의 표본을 조사한 경우를 살펴보았다. 따라서 절사층에서 일부의 표본을 추출하므로 더 이상 절사층이란 용어의 사용은 부적절하다. 이에 본 논문에서는 절사층과의 구분을 위해 절사표본층이란 용어를 사용하였다. 최근 김지학과 신기일 (2011)은 절사표본층에서 소수의 표본을 조사하여 얻은 결과와 표본층과 전수층에서 얻어진 자료를 이용하여 절사표본층을 추정하여 얻은 결과를 결합한 복합추정량을 제안하였다. 그러나 모의실험 결과 모집단 자료의 형태에 따른 우수성 비교에서 가장 우수한 결과를 주지 못하였다. 따라서 본 논문에서는 절사표본층에서 얻은 추정량의 분산을 줄여, 전체적으로 우수한 결과를 주는 새로운 복합추정량을 제안하였다. 또한 모의실험을 통하여 제안한 추정량과 기존에 제안된 추정량을 비교하여 그 우수성을 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 Benedetti 등 (2010)의 내용을 이용하여 절사표본 조사의 일반적 구조와 기호를 설명하고 기존에 사용되었던 절사층 추정법을 간단히 살펴보았다. 또한 기존의 절사층 총합 추정법의 내용은 박인호 (2007)와 김기학과 신기일 (2011)을 참조하였다. 3절에서는 모의실험을 통하여 김지학과 신기일 (2011)이 제안한 복합추정량, Sandal 등 (1992)이 제안한 SSW 추정량 그리고 본 논문에서 제안한 추정량의 우수성을 비교하였다. 4절에 결론이 있다.

2. 절사표본

2.1. 기존의 절사표본

2.1.1. 절사 표본 및 기호

절사표본조사의 방법들을 살펴보기 위해 Benedetti 등 (2010)의 설정과 기호를 사용하였다. 먼저 모집단은 3개의 층으로 구분된다. 크기가 N 인 모집단 U 는 모든 개체를 조사하는 전수층(U_C), 일부만을 포함하는 표본층(U_S) 그리고 표본조사에서 제외하거나 극히 일부의 표본을 추출하는 절사표본층(U_{SE})의 3가지 형태로 분할 또는 층화된다. 조사변수의 모총합은 각 층 총합을 더한 다음의 형태로 표시할 수 있다.

$$t_y = t_{yC} + t_{yS} + t_{ySE}. \quad (2.1)$$

또한 전수층과 표본층을 합하여 조사에 포함(inclusion)되는 층인 조사포함층의 총합 t_{yI} 는 $t_{yI} = t_{yC} + t_{yS}$ 이다. 여기서 t_y 는 모총합이고 t_{yC} , t_{yS} , t_{ySE} 는 전수층, 표본층, 절사표본층 각각의 층의 총합들이다. 그리고 표현을 간단히 하기 위하여 $U_I = U_C \cup U_S$, $S_I = U_C \cup S$ 라 정의하자. 여기서 S 는 표본층의 표본집합을 의미한다. 다음으로 모집단 각 개체에 대하여 보조변수 x 값의 모집단 및 분할층의 x -총합은 t_x , t_{xC} , t_{xS} , t_{xSE} , t_{xI} 로 표시하였다. 또한 절사표본추출에서 개체 k 가 표본에 포함될 확률(inclusion probability)은 $P(k \in U_C) = 1$, $0 < P(k \in U_S) < 1$ 이고 절사층을 사용한 경우에는 $P(k \in U_{SE}) = 0$ 이며 절사표본층인 경우에는 $0 < P(k \in U_{SE}) < 1$ 이다.

2.1.2. 모총합 추정전략

연구를 단순화하기 위해 본 논문에서는 조사상 무응답이 없다고 가정한다. 조사에 포함되지 않은 절사층에 대해서 사전에 알고 있는 정보, 예를 들면, 표본틀 상의 보조변수값 x_k 나 조사 이전에 파악한

x_k 와 y_k 의 관계 혹은 행정자료 등을 통하여 얻을 수 있는 보조 정보가 있으므로 이 정보를 이용하여 절사층 총합을 추정한다. 즉 관심변수와 보조변수를 이용하여 모형을 설정하고 이를 이용하여 각 층별 총합추정량을 구한 후 이를 종합하여 모집단 전체의 모총합을 추정한다. 이때 절사층의 총합을 직접적으로 추정하는 방식인

$$\hat{t}_y = t_{yC} + \hat{t}_{IS} + \hat{t}_{ySE} \quad (2.2)$$

을 사용하거나 절사층의 상대적 크기를 보정해 주는 방식인

$$\hat{t}_y = \hat{A}_y \hat{t}_{yI} \quad (2.3)$$

의 형태를 고려할 수 있다. 만일, $\hat{A}_y = (t_{yC} + \hat{t}_{IS} + \hat{t}_{ySE}) / (t_{yC} + \hat{t}_{IS})$ 이라 놓는다면 위의 두 방식은 결국 동일한 추정량이 된다. 여기서 \hat{A}_y 와 같이 \hat{t}_{yI} 에 일정한 수를 곱하여 전체를 추정하는 계수를 확대계수라고 부른다.

2.1.3. 모총합 추정량

이 절에서는 Sarndal 등 (1992)이 제안한 SSW 방법과 BLUP 방법 그리고 김지학과 신기일 (2011)이 제안한 방법을 간단히 살펴보았다. 여기서 SSW 방법과 BLUP 방법은 절사층을 사용하고 있다.

1. Sarndal-Swansson-Wretman(SSW) 방법

모집단과 조사포함층에서 두 변수 y 와 x 간의 총합비율 혹은 상대적 크기를 각각 $R_{yx} = t_y/t_x$ 과 $R_{yxI} = t_{yI}/t_{xI}$ 으로 정의하자. 만약 이 둘의 비율이 근사한 경우, 즉

$$R_{yxI} \approx R_{yx}$$

인 경우를 가정하여 Sarndal 등 (1992)은 다음의 추정량을 연구하였다.

$$\hat{t}_y^{SSW} = \hat{R}_{yxI} t_x, \quad (2.4)$$

여기서 $\hat{R}_{yxI} = \hat{t}_{yI}/\hat{t}_{xI}$ 는 표본으로부터 추정된 조사포함층에서의 총합비율을 나타낸다. 이는

$$\hat{t}_y^{SSW} = A'_x \hat{t}_{yI}$$

의 형태로도 표현될 수 있다. 확대계수 $A'_x = t_x/\hat{t}_{xI}$ 는 식 (2.3)의 A_x 와는 분모에서만 차이가 있다.

2. BLUP 방법

McDowney (2004)에서 사용한 방법으로 조사변수값 y_k 가 확률변수 Y_k 의 실현 값이며 모형 $y_k = \beta x_k + \sqrt{x_k} \epsilon_k$ 을 따른다고 가정하자. 여기서 y_k 는 관심변수, x_k 는 보조변수 그리고 $\epsilon_k \sim iid(0, \sigma^2)$ 이다. 그러면 모총합 t_y 의 최량선형불편추정량(best linear unbiased predictor; BLUP)은 다음과 같이 표현된다.

$$\hat{t}_y^{BLUP} = \hat{t}_{yI} + R_{YxS_I} \sum_{S_E} x_k, \quad (2.5)$$

여기서 $R_{YxS_I} = T_{yS_I}/T_{xS_I}$, $T_{yS_I} = \sum_{S_I} y_k$, $T_{xS_I} = \sum_{S_I} x_k$ 이다. 이 방법은 사업체조사에서 분산이 보조변수의 크기에 비례하는 경우에 주로 사용되는 방법이다. 그러나 박인호 (2007)의 모의실험 결과에서도 알 수 있듯이 관심변수와 보조변수와의 관계에 맞지 않는 모형을 사용할 경우 효율성이 매우 떨어지는 것으로 알려져 있다.

표 1: 모집단, 표본, 총합 분할

분할	전수층	표본층	절사표본층	전체
모집단	U_C	U_S	U_{SE}	U
모집단크기	N_C	N_S	N_{SE}	N
표본	U_C	S	S_{SE}	S_U
표본크기	N_C	n_S	n_{SE}	n
y 총합	t_{yC}	t_{yS}	t_{ySE}	t_y
x 총합	t_{xC}	t_{xS}	t_{xSE}	t_x

3. 수정절사법과 SSW 방법을 이용한 복합추정량

복합추정량은 절사층에서 전혀 조사를 하지 않는 방법을 사용하지 않고, 소수의 표본을 조사하여 이를 절사층의 총합추정에 사용하는 것이다. 따라서 복합추정량을 위해서는 절사층이 아닌 절사표본층을 사용한다. 표 1에 본 연구에서 사용하는 기호를 설명하였다.

표 1에서 $n_{SE} = 0$ 이면 절사층이 되고, $n_{SE} > 0$ 이면 절사표본층이 된다. 일반적으로 절사표본의 기본 가정이 절사층의 표본관리가 어렵고 무응답 비율이 매우 높기 때문이므로 소수의 표본이 절사표본층에 배정되는 것이 타당하다. 만약 절사표본층에 충분한 수의 표본이 배정된다면 수정절사법을 사용하면 된다. 이제 절사표본층에서 표본이 추출되었다면 절사표본층에서 얻어진 결과와 표본층에서 얻어진 결과를 결합한 선형결합 추정량 또는 복합 추정량을 만들 수 있다. 본 논문에서는 분산이 크지만 상대적으로 편향이 작은 절사표본층의 총합 추정량으로 다음의 추정량을 사용하였다. 이 추정량은 사업체조사에서 흔히 사용되고 있는 비추정량(ratio estimator)이다.

$$\hat{t}_{ySE}^{MODI} = \frac{\hat{t}_{ySE}}{\hat{t}_{xSE}} t_{xSE}. \quad (2.6)$$

김지학과 신기일 (2011)은 SSW 방법과 식 (2.6)을 이용한 복합추정량을 제안하였다. 즉 식 (2.6)을 이용하여 절사표본층의 총합을 추정하고 SSW 방법에서 쉽게 구할 수 있는 절사표본층의 총합추정량 $\hat{t}_{ySE}^{SSW} = (\hat{t}_{yI}/\hat{t}_{xI})t_{xSE}$ 을 사용하여 절사표본층의 총합을 추정한 후 이를 선형결합한다. 즉 절사표본층의 총합 추정을 위해 \hat{t}_{ySE}^{SSW} 와 \hat{t}_{ySE}^{MODI} 를 이용한 복합추정량은 다음과 같다.

$$\hat{t}_{ySE}^{MODI-SSW} = \left(\alpha^{[1]} \frac{\hat{t}_{ySE}}{\hat{t}_{xSE}} + (1 - \alpha^{[1]}) \frac{\hat{t}_{yI}}{\hat{t}_{xI}} \right) t_{xSE}. \quad (2.7)$$

이 추정량을 사용하기 위해서는 가중치 α 의 추정값이 필요하다. 흔히 복합추정량을 위해 사용하는 가중치 추정량은 다음과 같다.

$$\hat{\alpha} = \frac{\text{MSE}(R_{SE}^{[SSW]})}{\text{MSE}(R_{SE}^{[MODI]}) + \text{MSE}(R_{SE}^{[SSW]})} \approx \frac{\text{Var}(R_{SE}^{[SSW]})}{\text{Var}(R_{SE}^{[MODI]}) + \text{Var}(R_{SE}^{[SSW]})}, \quad (2.8)$$

여기서 $R_{SE}^{[MODI]} = \hat{t}_{ySE}/\hat{t}_{xSE}$ 이고 $R_{SE}^{[SSW]} = \hat{t}_{yI}/\hat{t}_{xI}$ 이다. 또한 $\alpha^{[1]} = 1$ 이면 식 (2.6)의 추정량이 되고, $\alpha^{[1]} = 0$ 이면 SSW 방법이 된다. 가중치 추정에 관한 자세한 내용은 Rao (2003)를 살펴보기 바란다.

2.2. 제안된 복합추정량

김지학과 신기일 (2011)에서 제안한 방법은 절사표본층 총합추정을 위해 식 (2.6)을 이용한다. 이 방법은 일반적인 비추정량이다. 또한 SSW 방법을 사용하였다. 그러나 김지학과 신기일 (2011)에서는 모집단 형태에 따라 추정량의 우수성이 달리 나왔다. 즉 비율형에서는 SSW 방법이 우수한 결과를

주었으며 다른 세 가지 형태에서는 제안한 복합추정량이 우수하였다. 물론 이 방법들의 비교에서 사용한 모의실험에서는 각각의 자료에서 가중치 α 를 추정하지 않고, 전체적으로 하나의 값을 사용하였으므로 정확한 비교라고 할 수는 없다. 또한 박인호 (2007)의 결과를 보면 비율형에서는 BLUP 방법이 가장 우수한 결과를 주고 있다.

이에 본 논문에서는 BLUP 방법과 식 (2.6)을 결합한 복합추정량을 제안하였다. 또한 추정량의 정확한 비교를 위하여 본 논문에서는 붓스트랩 방법을 사용하여 가중치 α 를 추정하였다.

SSW 방법이나 BLUP 모두 표본층과 전수층을 이용하여 구한 비추정량을 사용한다. 그러나 일반적으로 전수층은 매우 큰 대규모 사업체이고, 표본층은 이보다 작은 사업체이다. 따라서 사업체 특성을 비교하면 절사표본층과 표본층의 특성이 더욱 유사할 수 있다. 이에 본 논문에서는 전수층과 표본층을 모두 이용한 방법과 표본층만을 이용한 방법 등 두 가지 방법을 제안하였다.

1. 전수층과 표본층 모두를 이용하는 방법

이 방법은 절사표본층에서 얻은 자료를 기반으로 얻어진 식 (2.6)을 이용하여 절사표본층의 총합추정량을 구하고, BLUP 방법에서 사용한 절사표본층의 총합추정량을 선형결합하여 복합추정량을 얻는 것이다. 먼저 BLUP의 절사표본층 총합추정량은 다음과 같다.

$$\hat{t}_{ySE}^{BLUP} = R_{YxS_I} \sum_{SE} x_k = R_{YxS_I} t_{xSE} = \frac{T_{yS_I}}{T_{xS_I}} t_{xSE}. \quad (2.9)$$

따라서 식 (2.6)과 식 (2.9)를 선형결합한 비추정량을 제안할 수 있다.

$$\hat{t}_{ySE}^{MODI-BLUP} = \hat{R}_{SE}^{[MODI-BLUP]} t_{xSE} = \left(\alpha^{[2]} \frac{\hat{t}_{ySE}}{\hat{t}_{xSE}} + (1 - \alpha^{[2]}) \frac{T_{yS_I}}{T_{xS_I}} \right) t_{xSE}. \quad (2.10)$$

2. 표본층 만을 이용하는 방법

절사표본층에서 구한 비추정량과 표본층만을 이용하여 구한 비추정량을 선형결합한 복합비추정량을 사용하여 절사표본층의 총합을 추정한다. 이는 표본층이 절사층과 유사할 것으로 판단되기 때문이며 또한 일반적으로 표본층의 자료수가 충분하므로 비추정값을 추정하기에는 충분하다고 판단할 수 있기 때문이다. 따라서 다음의 절사표본층 총합추정량을 얻을 수 있다.

$$\hat{t}_{ySE}^{MODI-BLUPA} = \hat{R}_{SE}^{[MODI-BLUPA]} t_{xSE} = \left(\alpha^{[3]} \frac{\hat{t}_{ySE}}{\hat{t}_{xSE}} + (1 - \alpha^{[3]}) \frac{\hat{t}_{yS}}{\hat{t}_{xS}} \right) t_{xSE}. \quad (2.11)$$

또한 최적의 $\alpha^{[2]}$, $\alpha^{[3]}$ 는 식 (2.8)을 사용하여 구할 수 있다.

3. 모의실험을 통한 모총합 추정량 비교

3.1. 모의실험 설계

2절에서 살펴본 절사표본 하에서 모총합 추정량들의 효율성을 비교하기 위해 모의실험을 수행하였다. 모의실험은 Lee 등 (1995)과 김지학과 신기일 (2011)이 사용한 방법을 사용하여 크기 $N = 10,000$ 인 모집단을 다음과 같이 생성하였다. 먼저 보조자료 x_k 는 평균 48이고 분산 768을 갖는 감마분포로부터 생성하였다. 주어진 x_k 값에 대하여 모두 네 종류의 조사자료 y_k 를 발생시켰으며 각각 평균 $\mu(x) = a + bx + cx^2$ 이고 분산 $\sigma^2(x) = d^2 x^{2g}$ 을 갖는 감마분포를 가정하였다. 표 2는 선택된 상수 a, b, c, d, g 의 값과 생성된 모집단 자료의 평균 및 표준편차를 나타낸다. 첫 번째로 생성된 자료는 보

표 2: 몬테카를로 모의실험에 사용된 조사변수

조사변수형태	a	b	c	d	g	평균	표준오차
비례형	0	1.50	0.00	5.13	0.50	71.9604513	53.8997110
선형	20	1.50	0.00	13.79	0.25	92.4053023	54.9977009
볼록형	0	0.25	0.01	4.91	0.50	43.0610470	56.6092963
오목형	0	3.00	-0.01	5.60	0.50	113.5912253	63.1183669

표 3: 모의실험을 위한 모집단 크기, 표본 크기 및 추출 비율

절사점 기준	N_C	N_S	N_E	n_{SE}	N_E	n_{SE}/N_S	n_E/N_E
80%	167	5806	4027	333	0	0.0574	0.0000
				328	5	0.0565	0.0012
				323	10	0.0556	0.0025
90%	167	7283	2550	333	0	0.0457	0.0000
				328	5	0.0450	0.0020
				323	10	0.0443	0.0039
95%	167	8241	1592	333	0	0.0404	0.0000
				328	5	0.0398	0.0031
				323	10	0.0392	0.0063

조변수와의 관계가 원점을 지나는 비례적 형태(ratio)이고, 두 번째 자료는 양의 절편 값을 갖는 선형관계(regression)이고, 세 번째 자료는 볼록한 형태(convex), 마지막 자료는 오목한 형태(concave)를 갖도록 하였다. 보조변수와의 피어슨 선형상관계수는 대략 0.75 정도이다.

다음으로 전수층과 표본층을 분리하기 위해 먼저 보조변수값 x_k 를 기준으로 내림차순으로 정리하였다. 전수층은 x_k -누적합이 5% 이상을 차지하는 167개의 개체들로 구성하였고, 절사표본층은 x_k -누적합이 100 p_x % 이상을 차지하고 남은 개체들로 구성하였으며, 나머지 개체들은 표본층을 구성하였다. 이때, 절사점으로 다음의 세 기준점들, 80%, 90%, 95% (즉, $p_x = 0.80, 0.90, 0.95$)를 고려하였다. 총표본 크기는 $n = 500$ 이고 표본추출률은 $f = n/N = 0.05$ 을 사용하였다. 또한 새롭게 제안된 복합추정량의 모의실험을 위하여 절사표본층에서 $n_{SE} = 5, 10$ 을 추출하였다. α 를 추정하기 위해서는 각 추정량의 MSE가 필요하다. MSE를 계산하기 위해 붓스트랩 방법이 사용되었다. 이때 사용한 붓스트랩 반복수 $B = 500$ 을 사용하였다. 또한 얻어진 복합추정량의 비교를 위해 사용한 반복은 $R = 3,000$ 을 사용하였으며 각 표본에서 편향(bias), 상대편향(relative bias, rbias), 제곱근 평균제곱오차(root mean square error; rmse)를 구하여 비교 통계량으로 사용하였다.

$$\text{bias} = \bar{\hat{t}}_y - t_y,$$

$$\text{rbias}(\%) = \frac{100(\bar{\hat{t}}_y - t_y)}{t_y},$$

$$\text{rmse} = \sqrt{\frac{1}{R} \sum_{r=1}^R [\hat{t}_y(r) - t_y]^2},$$

여기서 평균 $\bar{\hat{t}}_y = \sum_{r=1}^R \hat{t}_y(r)/R$ 이다.

본 논문의 모의실험은 박인호 (2007)에서 우수한 결과를 주는 것으로 알려진 SSW 방법과 김지학과 신기일 (2011)이 제안한 복합추정량 $\hat{t}_{ySE}^{MODI-SSW}$, 그리고 본 논문에서 제안한 $\hat{t}_{ySE}^{MODI-BLUP}$ 과 $\hat{t}_{ySE}^{MODI-BLUPA}$ 이다. 박인호 (2007)와 김지학과 신기일 (2011)의 모의실험 결과를 요약하면 비율형에서는 BLUP이 가장 우수하고, 다른 모집단 자료 형태에서는 $\hat{t}_{ySE}^{MODI-SSW}$ 가 가장 우수한 결과를 주는 것을

표 4: 80% 절사점 하에서 모총합 추정량 비교

형태	$n_{SE} = 5$				$n_{SE} = 10$				
	SSW	M-S	M-B	M-BA	SSW	M-S	M-B	M-BA	
비율	bias	-4245	-3857	-4126	-3827	-4245	-3257	-3603	-3215
	rbias	-0.59	-0.54	-0.57	-0.53	-0.59	-0.45	-0.50	-0.45
	rmse	15685	14808	13722	14950	15685	14666	13657	14795
선	bias	-48027	-28869	-30288	-28723	-48027	-21942	-21858	-22033
	rbias	-5.20	-3.12	-3.28	-3.11	-5.20	-2.37	-2.37	-2.38
	rmse	50355	35070	37817	34711	50355	29039	30548	28856
블록	bias	46528	6594	6093	7001	46528	7718	6612	8146
	rbias	10.81	1.53	1.41	1.63	10.81	1.79	1.54	1.89
	rmse	49303	27323	36589	25503	49303	25612	31660	24304
오목	bias	-43179	-20937	-21748	-20429	-43179	-16150	-14934	-16174
	rbias	-3.80	-1.84	-1.91	-1.80	-3.80	-1.42	-1.31	-1.42
	rmse	46568	31072	39398	29467	46568	27564	33316	26380

표 5: 90% 절사점 하에서 모총합 추정량 비교

형태	$n_{SE} = 5$				$n_{SE} = 10$				
	SSW	M-S	M-B	M-BA	SSW	M-S	M-B	M-BA	
비율	bias	-1011	-1661	-1861	-1641	-1011	-1385	-1593	-1364
	rbias	-0.14	-0.23	-0.26	-0.23	-0.14	-0.19	-0.22	-0.19
	rmse	16422	15710	15097	15778	16422	15501	14956	15561
선	bias	-33806	-18546	-19405	-18476	-33806	-14188	-14391	-14200
	rbias	-3.66	-2.01	-2.10	-2.00	-3.66	-1.54	-1.56	-1.54
	rmse	37526	25216	26467	25087	37526	21469	22149	21407
블록	bias	22918	4843	6100	4581	22918	5828	6533	5630
	rbias	5.32	1.12	1.42	1.06	5.32	1.35	1.52	1.31
	rmse	28632	19291	23731	18775	28632	18644	21724	18285
오목	bias	-20555	-9414	-12169	-8818	-20555	-8695	-10337	-8248
	rbias	-1.81	-0.83	-1.07	-0.78	-1.81	-0.77	-0.91	-0.73
	rmse	28184	20947	24772	20507	28184	20810	23442	20518

확인할 수 있다.

3.2. 결과분석

모의실험 결과는 표 4에서 표 6에 나와 있다. 여기서 SSW는 $\hat{\mu}_{ySE}^{SSW}$ 를 사용한 추정량이고, M-S는 $\hat{\mu}_{ySE}^{MODI-SSW}$ 를 M-B는 $\hat{\mu}_{ySE}^{MODI-BLUP}$ 를 M-BA는 $\hat{\mu}_{ySE}^{MODI-BLUPA}$ 를 사용한 추정량을 의미한다.

먼저 표본의 크기와 절사점 그리고 모집단의 형태에 상관없이 SSW에 비해 M-S, M-B, M-BA가 모두 우수한 결과를 주는 것을 확인할 수 있다. 특히 비율형인 경우 M-B가 가장 우수한 결과를 주고 있는 반면 다른 모집단 형태에서는 M-S와 M-BA가 우수한 결과를 주고 있다. M-S와 M-BA를 비교하면 비율형에서는 유사한 결과를 주고 있는 반면 다른 모집단 형태에서는 M-BA가 우수한 결과를 주고 있다.

4. 결론

각국의 국가통계기관에서는 조사의 효율성과 편의성으로 인해 절사표본조사를 많이 사용하고 있다. 박인호 (2007)는 기존의 여러 추정방법들 중 모의실험을 통해 SSW가 우수한 결과를 주는 것을 확인하였다. 물론 비율형만을 고려한다면 SSW보다 BLUP이 우수한 결과를 주는 것을 확인하였다. 또

표 6: 95% 절사점 하에서 모총합 추정량 비교

형태		$n_{SE} = 5$				$n_{SE} = 10$			
		SSW	M-S	M-B	M-BA	SSW	M-S	M-B	M-BA
비율	bias	-1667	-1525	-1624	-1515	-1667	-1013	-1128	-1002
	rbias	-0.23	-0.21	-0.23	-0.21	-0.23	-0.14	-0.16	-0.14
	rmse	17395	16256	15978	16284	17395	16821	16548	16848
선	bias	-22511	-11628	-12433	-11556	-22511	-9872	-10462	-9823
	rbias	-2.44	-1.26	-1.35	-1.25	-2.44	-1.07	-1.13	-1.06
	rmse	28434	19540	19916	19511	28434	18588	18680	18590
블록	bias	11514	3551	5484	3270	11514	3806	5851	3508
	rbias	2.67	0.82	1.27	0.76	2.67	0.88	1.36	0.81
	rmse	21338	17591	18335	17586	21338	17667	17863	17710
오목	bias	-12206	-5262	-8209	-4871	-12206	-4924	-7432	-4585
	rbias	-1.07	-0.46	-0.72	-0.43	-1.07	-0.43	-0.65	-0.40
	rmse	23289	19678	20496	19627	23289	19742	19884	19764

한 김지학과 신기일 (2011)은 비율형을 제외한 다른 모집단 형태에서 $\hat{t}_{ySE}^{MODI-SSW}$ 가 우수한 결과를 주는 것을 확인하였다. 그러나 이 논문에서 가중치 α 는 각각의 반복된 자료에 따라 계산된 것이 아니라 최적의 α 를 고정한 후 비교한 결과이다. 반면에 본 논문에서는 붓스트랩을 사용하여 실제 자료가 얻어진 경우를 가정한 상태에서의 최적의 α 를 구한 후 이 결과를 이용하여 추정량의 우수성을 비교하였다. 비교 결과 김지학과 신기일 (2011)이 제안한 M-S 방법이 비율형에서도 SSW에 비해 우수한 결과를 주고 있음을 확인하였다.

본 논문에서는 BLUP에서 얻어진 결과를 선형결합한 새로운 복합추정량 $\hat{t}_{ySE}^{MODI-BLUP}$ 과 $\hat{t}_{ySE}^{MODI-BLUPA}$ 을 제안하였다. 모의실험 결과 $\hat{t}_{ySE}^{MODI-BLUPA}$ 가 매우 우수한 결과를 주고 있다. 따라서 절사층에서 흔히 발생할 수 있는 무응답 및 조사 부담성과 응답 부담성 등의 조사현실을 고려한 후, 절사표본층에서 소수의 표본을 추출할 수 있는 경우에 본 논문에서 제안한 복합추정량인 $\hat{t}_{ySE}^{MODI-BLUPA}$ 을 사용하게 되면 매우 우수한 총합추정을 할 수 있을 것으로 판단된다. 물론 모집단 자료 형태가 비율형이라는 것이 확인되면 $\hat{t}_{ySE}^{MODI-BLUP}$ 을 사용하는 것이 좋다. 그러나 모집단 형태에 관한 정보가 확실하지 않은 경우에는 $\hat{t}_{ySE}^{MODI-BLUPA}$ 을 사용하는 것이 우수한 결과를 줄 수 있을 것이다. 또한 표본층이 여러 층으로 나누어져 있는 경우에는 절사표본층과 유사한 형태를 주는 표본층만을 사용하여 얻은 $\hat{t}_{ySE}^{MODI-BLUPA}$ 의 사용을 고려하는 것이 타당하다.

참고 문헌

- 김지학, 신기일 (2011). 복합추정량을 이용한 절사표본 총합 추정에 관한 연구, <응용통계연구>, **24**, 1115-1128.
- 박인호 (2007). 절사표본조사에 관한 연구, <국민계정>, **30**, 81-106.
- 신민웅, 이상은 (2001). <표본조사를 위한 표본설계>, 교우사
- Benedetti, R., Bee, M. and Espa, G. (2010). A framework for cut-off sampling in business survey design, *Journal of Official Statistics*, **26**, 651-671.
- Clark, K. K. and Kinyon, D. L. (2007). *Can We Continue to Exclude Small Single-Establishment Businesses from Data Collection in the Annual Retail Trade Survey and the Service Annual Survey?*, Presented at the Third International Conference on the Establishment Surveys in Montral, Quebec, Canada.
- Elisson, H. and Elvers, E. (2001). Cut-off sampling and estimation, In *Proceedings of Statistics Canada Symposium on Achieving Data Quality in a Statistical Agency: A Methodological Perspective*, 2001.

- Hidioglou, M. A. (1986). The construction of a self-representing stratum of large units in survey design, *The American Statistician*, **4**, 27–31.
- Lee, H., Rancourt, E. and Sarndal, C.-E. (1995). Experiment with variance estimation from survey data with imputed value, *Journal of Official Statistics*, **10**, 231–243.
- Lee, S. E. (2011). The cut-off point based on MSE in modified cut-off sampling, *Journal of the Korean Official Statistics*, **16**, 82–94
- Lee, S. E. and Shin, K.-I. (2011). Alternative determination of cut-off point based on MSE, In *Proceedings of ISI 2011*, Dublin.
- McDowney, P. (2004). Simulation Result of Probability Proportional Size Sampling for EIA's Monthly Natural Gas Production Survey, a summary note of the Fall 2004 meeting of the American Statistical Association committee on Energy Statistics. <http://circa.europa.eu/irc/dsis/nacecpacon/info/data/en/handbook>
- Rao, J. N. K. (2003). *Small Area Estimation*, Wiley Interscience, A John Wiley and Sons, New York.
- Sarndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

2011년 12월 접수; 2011년 12월 수정; 2012년 1월 채택

An Alternative Composite Estimator for the Take-Nothing Stratum of the Cut-Off Sampling

Jong-Min Hwang^a, Key-Il Shin^{1,a}

^aDepartment of Statistics, Hankuk University of Foreign Studies

Abstract

Cut-off sampling that discards a part of the population from the sampling frame, is a widely used method for a business survey. Usually, to the estimate of population total, an accurate estimate of the total of the take-nothing stratum is required. Many estimators have been developed to estimate the total of the take-nothing stratum. Recently Kim and Shin (2011) suggested a composite estimator and showed the superiority of that estimator. In this paper, we suggest an alternative composite estimator obtained by combining BLUP estimator and a ratio estimator obtained by the small samples from the take-nothing stratum. Small simulation studies are performed for a comparison of the estimators and we confirm that the new suggested estimator is superior.

Keywords: Sampling survey, non-sampling error, cut-off sampling, ratio estimator.

This research was supported by the research fund of Hankuk University of Foreign Studies(2011).

¹ Corresponding author: Professor, Department of Statistics, Hankuk University of Foreign Studies, San 89, Wangsan, Mohyun, Yongin, Kyonggi Do 449-791, Korea. E-mail: keyshin@hufs.ac.kr