

가구조사에서 층화정보를 이용한 사후가중치 보정에 관한 연구: 모의실험을 중심으로

신기일¹⁾, 이상은²⁾

요약

표본 조사에서는 정확한 모수 추정을 위해 층화추출법이 널리 사용되고 있다. 층화 추출법의 성패를 좌우하는 것은 표본틀에 포함된 층화정보의 양이다. 표본설계 당시에 층화를 위한 충분한 정보가 없는 경우에는 이중추출법(double sampling, two-phase sampling)을 이용하여 정보를 얻을 수 있다. 반면 실사 후에 층화와 관련된 추가 정보가 얻어진다면 사후가중치 보정을 이용하여 모수 추정의 정확성을 향상시킨다. 이와 같이 층화추출에서 층화변수와 관련된 정보는 추정의 정확성에 큰 영향을 미친다. 본 논문에서는 가구조사(household survey)에서 층화추출법을 사용할 때 표본틀에 포함된 층화정보의 사용 방법에 관하여 연구하였으며 모의실험을 통해 인구 가중치를 사용할 때 표본설계 당시에 얻어진 가구 가중치를 적용한 경우와 적용하지 않은 경우의 결과를 비교하였다. 또한 가구 정보 없이 추가로 사람 조사가 보완조사로 이루어진 경우에서의 가중치 적용 방법을 설명하고 모의실험을 통해 각 방법의 성능을 비교하였다.

주요 용어 : 층화추출법, 이중추출법, 층화 변수, 가구 가중치, 인구 가중치

1. 서론

표본 조사의 정확성 향상을 위한 많은 연구가 수행되었다. 정확한 표본 조사를 위해서는 표본틀이 갖고 있는 정보를 충분히 활용한 표본설계가 이루어져야 하며 표본설계는 실제 조사에서 실행할 수 있는 조사 방법과 추정방법을 제시하여야 한다. 표본틀 정보에서 매우 중요한 것은 층을 나눌 수 있는 층화정보이다. 층화정보를 이용한 대표적인 표본설계법이 층화추출법이며 이 방법을 사용하게 되면 정확한 모수 추정이 가능하다. Cochran (1977) 뿐 아니라 국내의 많은 보고서에 층화추출법의 특징이 잘 설명되어 있다. 만약 층을 나누는 층화정보가 표본틀에 충분히 존재하지 않는다면 층화정보를 얻기 위한 조사를 실시하는 것도 하나의 방법이다. 이렇게 층화정보를 얻기 위한 1차 조사와 관심변수를 조사하는 2차 조사(본 조사)를 나누어 실시할 수 있으며 이러한 표본설계 방법을 이중추출법(double sampling, two-phase sampling)이라 한다. 이중추출법은 기본적으로 층화정보를 얻기 위해 비용이 저렴한 방법을 이용하여 다수의 1차 조사를 실시한 후, 얻어진 1차 조사 자료 목록을 표본틀(master sampling frame)로 하여 2차 조사를 실시하는 방법이다. 이는 1차 조사에 비

1) 경기도 용인시 처인구 모현면 외대로 81. 한국외국어대학교 통계학과, 교수. E-mail: keyshin@hufs.ac.kr

2) 교신저자, 경기도 수원시 영통구 광교산로 154-42, 경기대학교 응용통계학과, 교수, E-mail: sanglee62@kgu.ac.kr

용이 소요되어 2차 조사의 표본 규모가 감소하더라도 2차 조사에서는 층화정보를 이용하기 때문에 정확한 모수 추정이 가능하기 때문이다. 이중추출법에 관한 내용은 Cochran (1977)에 자세히 설명되어 있다. 물론 이 외에도 다른 목적으로 이중추출법이 사용될 수 있다. 예를 들면 Fuller (2000)는 이중추출법에서 보조변수를 이용한 회귀추정을 연구하였으며 Hidiroglou (2001)는 비내포(non-nested) 이중추출법에서의 회귀추정량을 연구하였다. 그러나 층화정보를 얻어 추정량의 정확성을 향상시키는 이중추출법의 경우에는 분산 추정량 계산이 매우 복잡해지는 단점이 있다. 분산 추정과 관련된 연구로는 Kim and Yu (2011)이 있으며 Koyuncu and Kadilar (2009), Lee 등 (2015)은 이중추출법에서 사용하는 추정량의 정확성 향상을 위해 비추정량, 곱추정량 등을 연구하였다.

국내에서 대부분의 사람과 관련된 조사는 직접 사람을 추출하여 조사하지 않고, 먼저 가구를 추출한 후 가구에 포함된 사람(가구원)을 조사하고 있다. 이는 실질적으로 사람 표본들이 존재하지 않기 때문이다. 따라서 사람과 관련된 표본 조사는 승인 통계의 경우 통계청의 조사구를 이용하여 가구를 추출한 후 가구에 포함된 가구원을 조사하는 것이 흔한 방법이다. 대표적인 예로 국민독서실태조사(승인번호: 제 11318), 인터넷이용실태조사(승인번호: 제 12005), 인터넷중독실태조사(승인번호: 제12019) 등이 있다. 이를 위해 통계청의 인구주택총조사 표본들에서 지역별, 거처종류별, 가구원 수별로 층을 나눈 후, 층에 포함된 조사구를 추출하고, 조사구에서 가구를 추출한 후, 가구에 포함된 가구원을 조사한다. 따라서 표본설계 당시에, 지역별, 거처종류별, 가구원 수별로 층이 나누어지고, 각 층의 층 가중치를 이용하여 자료의 가중치가 결정된다. 그러나 많은 사람과 관련된 조사에서는 가구 가중치를 이용하여 모수를 추정하지 않고, 조사된 가구원 자료를 지역별, 성별, 연령별로 층화한 후 인구주택총조사의 지역별, 성별, 연령별 모집단 자료를 기반으로 표본가중치를 구하여 사용한다. 물론 이상점, 무응답 등의 보정은 필요한 경우에 추가로 이에 해당되는 가중치 보정을 실시한다. 많은 논문이 이상점과 무응답 보정을 이용한 최종 가중치 결정법을 연구하였으며 Oh and Shin (2016)은 이중추출법에서의 이상점 가중치 보정법에 관하여 연구하였다. 이중추출법에서는 추정에 사용된 층화정보를 얻기 위해 1차 조사와 같은 추가적인 노력을 한다. 따라서 표본설계 당시에 사용된 거처종류 및 가구원 수와 관련된 정보가 모수 추정에 사용되지 않는다면 모수 추정이 비효율적일 수 있다.

본 연구에서는 국내에서 흔히 사용하는 방법인 가구 추출 후 조사된 사람 수를 이용하여 인구 가중치를 구하여 사용하는 방법과 가구 추출 시에 계산된 가구 기반 가중치를 조사된 사람 수에 의해 계산된 인구 가중치로 보정한 경우의 결과를 비교하였다. 또한 본 연구에서는 실사에서 발생할 수 있는 현실적 상황인 추가 보완조사가 이루어질 경우의 최종 가중치가 모수 추정에 어떠한 영향을 미치는지 살펴보았다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 현행조사에서 흔히 모수 추정에 사용하는 방법을 설명하였고 보정가중치와 추가 보완조사 후 얻어진 보정 가중치를 이용하여 추정하는 방법을 설명하였다. 3절에서는 각 방법의 우수성을 모의실험을 통해 살펴보았으며 4절에 결론이 있다.

2. 가구조사에서의 표본설계 및 모수 추정 방법

2.1 인구주택총조사를 이용한 표본설계

국내에서 다수의 가구 조사가 실시되고 있으며 각 표본 조사에서 사람 또는 가구원을 조사하여 관심변수의 모수를 추정한다. 그러나 사람을 조사하기 위해서는 이에 해당되는 표본틀(sampling frame)이 존재해야 하지만 사람 표본틀은 존재하지 않는다. 따라서 현실적으로 사용 가능한 표본틀은 가구 표본틀이다. 통계청에서는 약 60 가구 씩 묶어 조사구를 만들었으며 이 조사구를 이용하여 가구를 추출한다. 즉 조사구를 추출한 후 조사구 내의 가구를 추출하는 이단계집락추출법이 흔히 사용된다. 이제 추출된 가구에서 가구원을 조사하며 이 가구원이 표본이 된다. 이단계집락추출법의 추정은 이미 잘 알려져 있다. 그러나 조사구의 수가 60개로 일정 수준 유지가 되어 있고, 단순랜덤추출로 조사구가 추출되며, 또한 조사구 내에서 같은 수의 가구가 단순랜덤추출 된다면 가구가 추출될 확률은 전국적으로 동일하다. 따라서 집락내상관계수가 $\rho = 0$ 을 만족한다면 단순임의추출로 가구가 추출되었다고 가정해도 모수 추정 및 분산 추정은 문제가 되지 않는다. 또한 가구 추출을 위해 사용되는 층화 변수는 지역, 주택형태 그리고 가구원 수이다. 이렇게 나누어진 층에서 배정된 가구 수에 해당되는 조사구를 추출하게 되고, 추출된 조사구에서 배정된 가구를 추출하게 된다. 이러한 방법은 직접 사람을 추출하고 조사할 수 없는 상황에서는 매우 흔히 사용되고 있다.

이제 사람과 관련된 조사를 위해 통계청의 인구총조사 결과를 모집단으로 사용할 수 있다. 즉 지역별, 성별, 연령별로 층을 나누고 나누어진 h 층에서 모집단 수 N_h 가 구해진다. 또한 가구를 추출한 후 얻어진 사람관련 조사결과에서 실제 사람조사 자료 수인 n_h 를 이용하여 단위무응답 보정 가중치인 $w_{hi} = w_h = N_h/n_h$ 가 얻어지며 이 가중치가 추정에 사용된다. 물론 조사 과정에서 발생할 수 있는 여러 문제로 인해 최종 결과로 얻어진 표본 수는 표본설계 당시에 예상했던 지역별, 연령별, 성별 표본 수 보다 많거나 적어질 수 있으므로 미리 표본설계 당시에 적정 또는 최소 응답 사람 표본 수를 정해줄 필요가 있다. 만약 층별로 정해진 목표 사람 표본 수에 비해 심하게 미달된다면 추가 보완조사를 통해 배정된 표본 수를 채우게 된다.

2.2 최종 가중치 계산 방법과 모수 추정 방법

2.1 절에서 설명하였듯이 현실적으로 가구관련 층화변수를 사용하여 층을 나누고 조사하였을 때 특정 연령층에서 실제 조사된 표본 수가 매우 적을 수 있으며 대표적으로 20대 연령층에서 이런 현상이 일어나고 있다. 즉 여러 가구 조사 결과를 살펴보면 20대 연령층의 표본 수가 표본설계 당시의 목표 표본 규모에 비해 매우 부족하게 된다. 이 경우 추가로 보완조사를 실시하게 되는데 이때 가구 기반 조사를 실시하게 되면 역시 20대 층의 목표 표본 수를 얻는 것은 현실적으로 매우 어렵다. 따라서 보

완조사는 20대 연령층과 같이 조사된 표본 수가 부족한 층을 기준으로 별도로 사람 기반의 추가 보완조사를 실시하게 된다. 결국 보완조사의 표본 조사 결과에는 가구 정보가 조사되지 않을 수도 있으며 이 경우에는 성별, 연령별 등 사람 기반 가중치만을 이용할 수밖에 없게 된다.

이 절에서는 가구 조사에서 사용할 수 있는 다양한 최종 가중치 계산법과 이를 이용한 모수 추정법을 설명하였으며 추가 보완조사에서 가구 층화정보가 없는 경우의 보정 가중치 계산 방법을 설명하였다. 흔히 인구주택총조사를 이용할 경우 지역, 주택종류, 가구원수가 가구와 관련된 층화 변수이고, 지역, 연령 그룹, 성별이 사람(가구원)과 관련된 층화변수이다. 이때 지역은 가구 기반 가중치와 사람 기반 가중치에 공통으로 사용된다. 따라서 가장 일반적인 층화 변수는 지역, 주택종류, 가구원수, 연령 그룹, 성별이다. 본 연구에서는 다음의 기호 r : 지역, t : 주택종류, w : 가구원수, a : 연령 그룹, g : 성별을 사용하였다. 이제 각 세부 층에 속한 i 번째 가구 또는 사람의 가중치를 w_{rtwagi} 라 하고, N_{rtwag} 를 모집단 수, n_{rtwag} 를 표본 수라 표시하였다.

2.2.1 가구 정보와 사람 정보가 동시에 있는 경우

2.2.1.1 가구 기준 가중치(M1)

가구와 관련된 층화변수만을 이용하여 구한 가구 기준 사람 가중치는 다음과 같다. 먼저 가구 수를 이용한 r 지역, t 주택종류, w 가구원수별 층별 가중치를 $w_{rtw}^{GAGU(1)}$ 라 하고, 최종 사람에게 주어진 r 지역, t 주택종류, w 가구원수, a 연령 그룹, g 성별 가구 기준 사람 가중치를 w_{rtwagi}^{GAGU} 라 하면 다음의 식을 이용하여 구한다.

$$w_{rtw}^{GAGU(1)} = \frac{N_{rtw}^{GAGU}}{n_{rtw}^{GAGU}}$$

$$w_{rtwagi}^{GAGU} = w_{rtw}^{GAGU(1)} \times \frac{N}{\sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{GAGU}} \quad (2.1)$$

여기서 N_{rtw}^{GAGU} 는 층별 모집단 가구 수이고, n_{rtw}^{GAGU} 는 층별 표본 가구 수이다. 따라서 성별, 연령별에 무관하게 사람의 가구 기준 가중치는 가구 특성인 주택종류와 가구원 수에 의해 결정된다. 이때 최종적으로 전체 모집단 사람 수와 가구 기준 가중치 합이 일치하도록 보정한다. 따라서 가구 기준 사람 가중치를 적용한 총합추정량은 다음과 같이 구해진다.

$$\hat{y}_{gagu} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{GAGU} y_{rtwagi} = \sum_r \sum_t \sum_w w_{rtw}^{GAGU} \sum_a \sum_g \sum_i y_{rtwagi}$$

이제 r 지역, t 주택종류, w 가구원수의 합을 y_{rtw} 라 하면, 즉 $y_{rtw} = \sum_a \sum_g \sum_i y_{rtwagi}$ 이라 하면 가구 기준 가중치를 이용한 총합 추정량은

$$\hat{Y}_{gagu} = \sum_r \sum_t \sum_w w_{rtw}^{GAGU} y_{rtw} \quad (2.2)$$

이 된다. 여기서 $w_{rtw}^{GAGU} = w_{rtwagi}^{GAGU}$ 이다.

2.2.1.2 사람 기준 가중치(M2)

사람과 관련된 층화변수만을 이용하여 구한 사람 기준 인구 가중치는 다음과 같다.

$$w_{rtwagi}^{PER} = w_{rag}^{PER} = \frac{N_{rag}^{PER}}{n_{rag}^{PER}} = \frac{\sum_t \sum_w N_{rtwag}^{PER}}{\sum_t \sum_w n_{rtwag}^{PER}} \quad (2.3)$$

N_{rag}^{PER} 는 r 지역, a 연령, g 성별인 층별 모집단 사람 수이고, n_{rag}^{PER} 는 각 층별 표본 사람 수이다. 따라서 사람 기준 인구 가중치를 적용한 총합추정량은 다음과 같이 구해진다.

$$\hat{y}_{per} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{PER} y_{rtwagi} = \sum_r \sum_a \sum_g w_{rag}^{PER} \sum_t \sum_w \sum_i y_{rtwagi}$$

이제 r 지역, a 연령, g 성별 합을 y_{rag} 라 하면, 즉 $y_{rag} = \sum_t \sum_w \sum_i y_{rtwagi}$ 이라 하면 사람 기준 인구 가중치를 이용한 총합 추정량은

$$\hat{Y}_{per} = \sum_r \sum_a \sum_g w_{rag}^{PER} y_{rag} \quad (2.4)$$

가 된다.

2.2.1.3 보정 가중치(M3)

보정 가중치는 먼저 가구와 관련된 층을 이용하여 가구 기반 사람 가중치를 부여하고, 최종적으로 사람과 관련된 세부 층으로 보정하는 방법을 사용한다. 여기서 보정의 의미는 가중치의 합을 일치시킨다는 의미이다. M1의 가구 기반 사람 가중치는 전국적으로 보정을 실시하는 반면 이 방법은 지역별, 연령별, 성별로 보정을 실시한다. 먼저 모든 가중치의 합은 그에 해당되는 모집단 인구수와 일치해야 하므로 다음 식이 만족되어야 한다.

- a. $\sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{GAGU} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{PER} = N^{PER}$
- b. $\sum_t \sum_w \sum_i w_{rtwagi}^{PER} = N_{rag}^{PER}$

이제 합계 순서를 바꾸면

$$N^{PER} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{GAGU} = \sum_r \sum_a \sum_g \sum_t \sum_w \sum_i w_{rtwagi}^{GAGU}$$

가 되고, $\sum_t \sum_w \sum_i w_{rtwagi}^{GAGU} = N_{rag}^{G-PER}$ 라 하면 이 값은 N_{rag}^{PER} 와 다른 값을 갖게 된다.

이 두 값을 일치시키는 것이 보정이며 이를 위해 $F_{rag} = \frac{N_{rag}^{PER}}{N_{rag}^{G-PER}}$ 를 구한 후 이 값을 가구 기반 가중치에 곱해주게 되면 보정 가중치가 얻어진다. 즉

$$\sum_r \sum_a \sum_g \sum_t \sum_w \sum_i w_{rtwagi}^{GAGU} \times F_{rag} = \sum_r \sum_a \sum_g N_{rag}^{G-PER} \times F_{rag} = \sum_r \sum_a \sum_g N_{rag}^{PER} = N^{PER}$$

가 된다. 따라서 가구 기반 사람 보정 가중치는

$$w_{rtwagi}^{ADJ} = w_{rtwagi}^{GAGU} \times F_{rag} = w_{rtwagi}^{GAGU} \times \frac{N_{rag}^{PER}}{N_{rag}^{G-PER}} \quad (2.5)$$

가 된다. 여기서 $N_{rag}^{PER} = \sum_t \sum_w \sum_i w_{rtwagi}^{PER}$ 이고 $N_{rag}^{G-PER} = \sum_t \sum_w \sum_i w_{rtwagi}^{GAGU}$ 이다. 결국 최종 보정 가중치를 이용한 총합 추정량은 다음과 같다.

$$\hat{Y}_{per-adj} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{ADJ} y_{rtwagi} \quad (2.6)$$

2.2.2 사람 기준 추가 보완 자료가 있는 경우

2.2.2.1 사람 기준 가중치(M4)

추가 보완조사에서 얻어진 자료가 지역별, 연령별, 성별로 얻어졌으나 가구 정보가 없는 경우가 실제 조사에서 흔히 발생하고 있다. 이제 추가로 조사된 자료가 r 지역, a 연령, g 성별로 얻어지고 크기가 n_{rag}^{PER} 이라 하자. 그러면 r 지역, a 연령, g 성별의 총 조사 자료 수는 $n_{rag}^{PER-t} = n_{rag}^{PER} + n_{rag}^{\prime PER}$ 가 된다. 따라서 방법 2의 n_{rag}^{PER} 대신에 n_{rag}^{PER-t} 를 대입하여 사람 기준 인구 가중치를 구한다. 즉

$$w_{rtwagi}^{PER-t} = w_{rag}^{PER-t} = \frac{N_{rag}^{PER}}{n_{rag}^{PER-t}} \quad (2.7)$$

이다. 따라서 추가 자료가 있는 경우의 사람 기준 인구 가중치를 적용한 총합추정량은 다음과 같이 구해진다.

$$\hat{Y}_{per-add} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{PER-t} y_{rtwagi} \quad (2.8)$$

2.2.2.2 보정 가중치(M5)

추가 보완조사가 있는 경우의 보정 가중치는 다음의 3 단계를 이용하여 구한다.

단계 [1] 가구 정보와 사람 정보가 있는 자료의 1차 가중치 구하기.

이 단계는 n_{rag}^{PER} 개의 자료에 해당되며 이때 가중치는 M3에서 사용한 방법으로 가중치를 구한다. 즉

$$w_{rtwagi}^{ADJ-O} = w_{rtwagi}^{GAGU} \times \frac{N_{rag}^{PER}}{N_{rag}^{G-PER}}$$

이고 여기서 $N_{rag}^{PER} = \sum_t \sum_w \sum_i w_{rtwagi}^{PER}$, $N_{rag}^{G-PER} = \sum_t \sum_w \sum_i w_{rtwagi}^{GAGU}$ 이다.

단계 [2] 추가로 얻어진 자료의 1차 가중치 구하기

이 단계는 n'_{rag} 개의 표본 자료가 있는 경우에 해당되며 이때 1차 가중치는 단계 [1]에서 얻어진 가중치인 w_{rtwagi}^{ADJ-O} 의 r 지역, a 연령, g 성별 평균을 이용한다. 이는 추가 자료의 경우 주택 종류와 가구원 수 정보가 없기 때문이다. 따라서 다음의 1차 가중치, $w_{rtwagi}^{ADJ-ADD}$ 를 사용한다.

$$w_{rtwagi}^{ADJ-ADD} = w_{ragi}^{ADJ-ADD} = w_{rag}^{ADJ-ADD} = \frac{1}{n_{rag}^{PER}} \sum_t \sum_w \sum_i w_{rtwagi}^{ADJ-O}$$

참고로 $\sum_t \sum_w \sum_i w_{rtwagi}^{ADJ-O} = N_{rag}^{PER}$ 이므로 $w_{rtwagi}^{ADJ-ADD} = \frac{N_{rag}^{PER}}{n_{rag}^{PER}} = w_{rag}^{PER}$ 이 되어 사람 기반 인구 가중치를 적용한 M2 결과와 일치한다.

단계 [3] 두 자료 합치기

각 단계에서 얻어진 1차 가중치를 기반으로 최종 가중치를 구한다. 여기서 최종 가중치는 얻어진 w_{rtwagi}^{ADJ-O} , $w_{rtwagi}^{ADJ-ADD}$ 의 r 지역, a 연령, g 성별 합이 해당 모집단 수 N_{rag}^{PER} 가 되도록 보정한다. 즉 $w_{rtwagi}^{ADJ-T} = \begin{cases} w_{rtwagi}^{ADJ-O}, & \text{1단계자료} \\ w_{rtwagi}^{ADJ-ADD}, & \text{2단계자료} \end{cases}$ 라 하면 최종 가중치는 다음과 같다.

$$w_{rtwagi}^{ADJ-F} = w_{rtwagi}^{ADJ-T} \times \frac{N_{rag}^{PER}}{\sum_t \sum_w \sum_i w_{rtwagi}^{ADJ-T}} \tag{2.9}$$

따라서 추가 자료가 있을 경우 보정 가중치를 적용한 총합추정량은 다음과 같이 구해진다.

$$\widehat{Y}_{per-add} = \sum_r \sum_t \sum_w \sum_a \sum_g \sum_i w_{rtwagi}^{ADJ-F} y_{rtwagi} \quad (2.8)$$

3. 모의실험

2절에서 설명한 5가지 추정량을 모의실험을 통하여 성능을 비교하였다. 특히 추가 보완조사가 없는 경우에 사용할 수 있는 방법인 M1, M2, M3가 상호 비교 가능하며 추가 자료가 있는 경우에 사용할 수 있는 방법인 M4, M5가 상호 비교 대상이다.

3.1 자료 생성

먼저 지역별(4개), 주택형태별(3개), 가구원수별(4개)로 총 48개 층을 만든 후 각 층에서 10,000개의 자료를 생성하였다. 따라서 총 48만 가구가 생성되었으며 각 가구는 1인 가구, 2인 가구, 3인 가구, 4인 가구가 있으므로 총 120만 사람이 생성되었다. 각 사람에게 5세에서 60세로 일정(uniform)하게 연령을 주었으며 성별도 베르누이(0.5)에서 랜덤으로 생성하였다. 성별을 5세 이상으로 한 것은 일부 실태조사에서 5세 이상의 자료만을 사용하기 때문으로 초기 연령을 특정 값으로 정하는 것은 모의실험 결과에 전혀 영향을 주지 않는다. 이렇게 만들어진 자료에 계수를 곱하여 최종 관심 자료가 생성되었다. 이를 위한 최종 모형 식은 다음과 같다.

모형 :

$$y_{rtwagi} = \mu + \alpha \times region_r + \beta \times htype_t + \gamma \times ngaguwon_w + \tau \times age_a + \delta \times gender_g + \epsilon_{rtwagi}$$

여기서 오차 ϵ_{rtwagi} 는 $N(0, 3^2)$ 과 $Gamma(1, 10)$ 을 사용하여 생성하였다. 다음으로 $region_r=1, 2, 3, 4$ 인 4개 지역별로, $htype_t=1, 2, 3$ 인 3개 주택종류별로, $ngaguwon_w=1, 2, 3, 4$ 인 4개 가구원수별로, age_a 는 5세-10세, 이후 10세 단위로 60세까지 생성하였으며, $gender_g=1, 2$ 인 이항변수로 생성한 후 계수인 $(\alpha, \beta, \gamma, \tau, \delta)$ 에 다양한 값을 주어 각 층별 평균의 차이에 따른 추정의 정확성을 살펴보았다. 또한 표본 규모에 따른 추정법의 효과를 살펴보기 위해 $n^{GAGU} = 10,000$ 과 $20,000$ 을 사용하여 결과를 살펴보았다. 반복은 $R=1,000$ 을 사용하였으며 반복할 때 마다 모집단을 새롭게 구축하여 특정 모집단에 의해 발생할 수 있는 영향력을 제거하였다.

사용된 비교통계량은 편향(bias), 절대편향(absolute bias), 절대상대편향(absolute relative bias) 그리고 제곱근 MSE(mean square error)이며 정의는 다음과 같다.

$$Bias = \frac{1}{R} \sum_{r=1}^R (\widehat{Y}^{(r)} - \overline{Y}^{(r)})$$

$$Absolute\ Bias = \frac{1}{R} \sum_{r=1}^R |\widehat{Y}^{(r)} - \overline{Y}^{(r)}|$$

$$Absolute\ Relative\ Bias = \frac{1}{R} \sum_{r=1}^R |\widehat{Y}^{(r)} - \overline{Y}^{(r)}| / \overline{Y}^{(r)}$$

$$Root\ Mean\ Square\ Error = \sqrt{\frac{1}{R} \sum_{r=1}^R (\widehat{Y}^{(r)} - \overline{Y}^{(r)})^2}$$

3.2 모의실험 결과

모의실험 결과에서 <표 3.1>에서 <표 3.8>은 오차가 정규분포인 경우의 결과이며 <표 3.9>에서 <표 3.16>은 오차의 분포가 감마분포인 경우의 결과이다. 각 분포에서 표본 수 $n^{GAGU} = 10,000$ 과 $20,000$ 인 경우의 결과가 수록되었고 비교 통계량인 편향, 절대편향, 절대상대편향 그리고 제공된 MSE의 결과가 수록되었다.

3.2.1 오차가 정규 분포를 따르는 경우

$n = 10,000$ 의 결과인 <표 3.1>에서 <표 3.4>의 비교 통계량 결과에서 편향 결과를 살펴보면 가구 층화 변수와 사람 층화 변수의 층간 평균의 차이가 없는 경우, 즉 $(\alpha, \beta, \gamma, \tau, \delta) = (5, 0, 0, 0, 0)$ 인 경우, 매우 일정한 결과를 주는 것을 확인할 수 있다.

모든 결과에서 편향을 기준으로 살펴보면 M3와 M1과의 비교, M3와 M2와의 비교에서 M3가 우수한 결과를 주지 않고 있다. 이러한 경향은 M5와 M4에서도 나타난다. 따라서 편향을 기준으로 할 경우에는 제안된 방법이 우수하다고 판단할 수 없다. 그러나 절대상대편향을 보면 그 비율이 매우 작기 때문에 편향은 없다고 판단하는 것이 타당하다. 결국 모든 방법이 불편추정량이라고 판단할 수 있어 편향을 비교하는 것은 큰 의미가 없다.

가구 층화 변수 그리고 사람 층화 변수의 층간 평균 차이가 없는 경우는 모든 방법의 결과가 매우 유사하다. 이는 층의 평균에 차이가 없는 경우에는 층 정보를 사용하는 것이 큰 의미가 없다는 것을 의미하며 예상된 결과이다. 그러나 사람 층화 변수의 층간 평균이 일정하지 않은 경우, 즉 $\tau > 0, \delta > 0$ 인 경우에는 M2가 M1에 비해 매우 우수한 결과를 주고 또한 $\beta > 0, \gamma > 0$ 인 경우, 즉 가구 층화 변수의 층간 평균에 차이가 있는 경우에는 M2에 비해 M1이 우수한 결과를 주고 있다. 반면 M3는 어떤 경우에도 다른 두 방법의 우수한 결과를 모두 보여 주고 있다. 특히 가구 층화 변수와 사람 층화 변수의 모든 층간 평균의 차이가 큰 경우에는 M3가 매우 우수한 결과를 주고 있다. 결론적으로 층의 평균 간에 차이가 없다는 확실한 증거가 없다면 모든 층의 정보를 사용하여 만든 가중치를 사용하는 방법인 M3가 안전하면서도 우수한 결과를 준다.

또한 추가로 보완조사가 이루어진 경우에는 인구 가중치만을 사용하는 M4에 비해 M5가 매우 우수한 결과를 주고 있으며 이는 예상된 결과이기도 하다. 특히 주택형태, 가구원 수 층의 평균에 큰 차이가 있는 자료의 경우 반드시 M5 방법을 사용해야 한다. 이러한 결과는 표본 규모가 $n^{GAGU} = 10,000$ 인 경우와 $20,000$ 인 경우에서도 같은

패턴의 결과를 준다. 또한 <표 3.5>에서 <표 3.8>을 살펴보면 표본 규모와는 무관하게 각 방법의 성능 비교 결과가 유지된다고 판단된다.

<표 3.1> 정규분포에서의 편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	편향				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	379.42	336.61	398.63	377.91	434.90
(5, 0, 0, 10, 3)	311.78	12637.11	12436.41	10075.25	9878.20
(10, 0, 0, 50, 5)	-93415.94	236.56	18937.23	341.24	18037.66
(5, 5, 3, 0, 0)	-114.75	-2751.98	-285.49	-2335.92	44.12
(10, 10, 5, 0, 0)	-384.47	1111.41	85.55	2463.57	1425.48
(5, 5, 3, 10, 3)	97712.46	-1059.55	-1756.37	-7657.68	-8347.16

<표 3.2> 정규분포에서의 절대편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Absolute bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	18845.42	18888.86	18966.37	18192.51	18241.18
(5, 0, 0, 10, 3)	1129251.00	188952.00	183655.50	177542.00	173473.80
(10, 0, 0, 50, 5)	5079756.86	941735.80	916794.59	903865.68	883632.76
(5, 5, 3, 0, 0)	18625.18	53674.54	19079.14	52955.02	21858.15
(10, 10, 5, 0, 0)	18544.01	108007.80	18609.79	104915.10	26545.23
(5, 5, 3, 10, 3)	1060230.96	200416.51	188697.87	189269.77	180458.95

<표 3.3> 정규분포에서의 절대상대편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Relative bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	1.55E-05	1.55E-05	1.56E-05	1.50E-05	1.50E-05
(5, 0, 0, 10, 3)	6.85E-04	1.15E-04	1.11E-04	1.08E-04	1.05E-04
(10, 0, 0, 50, 5)	1.50E-03	2.78E-04	2.70E-04	2.66E-04	2.60E-04
(5, 5, 3, 0, 0)	1.50E-05	4.34E-05	1.54E-05	4.28E-05	1.77E-05
(10, 10, 5, 0, 0)	1.46E-05	8.49E-05	1.46E-05	8.25E-05	2.09E-05
(5, 5, 3, 10, 3)	6.34E-04	1.20E-04	1.13E-04	1.13E-04	1.08E-04

〈표 3.4〉 정규분포에서의 RMSE 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	RMSE				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	23754.72	23757.73	23874.33	22726.93	22820.88
(5, 0, 0, 10, 3)	1425823.00	237752.70	235436.60	223783.80	221319.60
(10, 0, 0, 50, 5)	6273324.88	1154453.41	1133433.00	1116937.37	1099739.87
(5, 5, 3, 0, 0)	22908.85	67912.31	23440.66	66269.29	26859.49
(10, 10, 5, 0, 0)	22990.68	134657.80	23377.84	129520.60	33579.14
(5, 5, 3, 10, 3)	1346575.20	254540.45	243420.75	243269.59	233101.28

〈표 3.5〉 정규분포에서의 편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	편향				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	-315.08	-314.68	-304.21	8.13	18.00
(5, 0, 0, 10, 3)	-92847.73	-1616.13	-4445.39	-1803.95	-4567.02
(10, 0, 0, 50, 5)	-154327.46	-45408.57	-38052.00	-41700.35	-34560.19
(5, 5, 3, 0, 0)	1762.34	3667.46	1862.14	3571.22	1897.61
(10, 10, 5, 0, 0)	1293.13	-1855.04	1235.45	50.23	3075.98
(5, 5, 3, 10, 3)	219608.10	2160.56	4062.47	-283.46	1573.08

〈표 3.6〉 정규분포에서의 절대편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Absolute bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	12879.72	12880.03	12823.27	12550.06	12499.05
(5, 0, 0, 10, 3)	789036.09	131979.22	127383.36	128902.85	123444.73
(10, 0, 0, 50, 5)	3609813.28	663797.29	651860.50	661331.84	644065.09
(5, 5, 3, 0, 0)	20146.29	55473.27	19788.71	54091.49	22430.49
(10, 10, 5, 0, 0)	18592.85	98600.30	19231.04	95764.98	30614.52
(5, 5, 3, 10, 3)	999612.70	183160.96	176630.77	175933.12	171547.26

〈표 3.7〉 정규분포에서의 절대상대편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Relative bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	1.06E-05	1.06E-05	1.06E-05	1.03E-05	1.03E-05
(5, 0, 0, 10, 3)	4.79E-04	8.00E-05	7.73E-05	7.82E-05	7.49E-05
(10, 0, 0, 50, 5)	1.06E-03	1.96E-04	1.92E-04	1.95E-04	1.90E-04
(5, 5, 3, 0, 0)	1.61E-05	4.44E-05	1.58E-05	4.33E-05	1.79E-05
(10, 10, 5, 0, 0)	1.45E-05	7.68E-05	1.50E-05	7.46E-05	2.38E-05
(5, 5, 3, 10, 3)	5.94E-04	1.09E-04	1.05E-04	1.04E-04	1.02E-04

〈표 3.8〉 정규분포에서의 RMSE 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	RMSE				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	16159.95	16162.03	16097.21	15839.94	15773.89
(5, 0, 0, 10, 3)	997672.47	164278.31	158634.02	160871.91	155020.74
(10, 0, 0, 50, 5)	4514901.36	840582.35	818602.75	824139.24	804568.23
(5, 5, 3, 0, 0)	25703.45	68459.98	25364.37	66890.99	28401.63
(10, 10, 5, 0, 0)	23537.57	123054.39	24216.66	121373.63	38511.57
(5, 5, 3, 10, 3)	1271911.00	231640.53	221153.29	226121.99	216388.10

3.2.2 오차가 감마 분포를 따르는 경우

이 절에서는 오차가 정규 분포를 따르지 않고, 비대칭인 감마 분포를 따르는 경우의 결과를 〈표 3.9〉에서 〈표 3.16〉에 수록하였다. 결과를 살펴보면 오차가 정규분포인 경우와 같은 패턴의 결과를 보여주고 있다. 따라서 층간 평균에 차이가 있는 경우 층에 해당되는 층화정보를 이용하여 가중치를 구해야 하며 확실하게 층간 평균에 차이가 없다는 증거가 없는 한 모든 층화정보를 사용하는 M3와 M5를 사용하는 것이 타당하다.

〈표 3.9〉 감마분포에서의 편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	편향				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	499.44	543.11	518.23	944.48	920.90
(5, 0, 0, 10, 3)	122761.14	-1446.00	-1761.43	4450.36	4177.12
(10, 0, 0, 50, 5)	482931.60	116503.72	119885.83	123233.25	126184.08
(5, 5, 3, 0, 0)	1762.34	3667.46	1862.14	3571.22	1897.61
(10, 10, 5, 0, 0)	1293.13	-1855.04	1235.45	50.2253	3075.98
(5, 5, 3, 10, 3)	219608.10	2160.56	4062.47	-283.460	1573.08

〈표 3.10〉 감마분포에서의 절대편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Absolute bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	19790.06	19951.77	19874.15	19614.76	19658.57
(5, 0, 0, 10, 3)	1141174.10	193118.76	195245.07	186565.88	189603.51
(10, 0, 0, 50, 5)	5368717.72	995317.03	974126.74	955662.05	942217.36
(5, 5, 3, 0, 0)	20146.29	55473.27	19788.71	54091.49	22430.49
(10, 10, 5, 0, 0)	18592.85	98600.30	19231.04	95764.98	30614.52
(5, 5, 3, 10, 3)	999612.70	183160.96	176630.77	175933.12	171547.26

〈표 3.11〉 감마분포에서의 절대상대편향 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Relative bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	1.61E-05	1.63E-05	1.62E-05	1.60E-05	1.60E-05
(5, 0, 0, 10, 3)	6.87E-04	1.16E-04	1.18E-04	1.12E-04	1.14E-04
(10, 0, 0, 50, 5)	1.58E-03	2.92E-04	2.86E-04	2.81E-04	2.77E-04
(5, 5, 3, 0, 0)	1.61E-05	4.44E-05	1.58E-05	4.33E-05	1.79E-05
(10, 10, 5, 0, 0)	1.45E-05	7.68E-05	1.50E-05	7.46E-05	2.38E-05
(5, 5, 3, 10, 3)	5.94E-04	1.09E-04	1.05E-04	1.04E-04	1.02E-04

〈표 3.12〉 감마분포에서의 RMSE 결과, $n = 10,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	RMSE				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	24920.28	24960.21	25007.90	24525.44	24591.79
(5, 0, 0, 10, 3)	1438849.70	240579.23	241190.64	232008.28	233261.72
(10, 0, 0, 50, 5)	6934102.45	1235806.73	1207896.67	1196082.11	1169579.64
(5, 5, 3, 0, 0)	25703.45	68459.98	25364.37	66890.99	28401.63
(10, 10, 5, 0, 0)	23537.57	123054.39	24216.66	121373.63	38511.57
(5, 5, 3, 10, 3)	1271911.00	231640.53	221153.29	226121.99	216388.10

〈표 3.13〉 감마분포에서의 편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	편향				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	-897.70	-957.23	-901.83	-912.035	-858.19
(5, 0, 0, 10, 3)	-9984.36	-3029.90	-2222.47	-971.391	-173.30
(10, 0, 0, 50, 5)	288352.62	-47054.84	-32208.80	-57345.15	-42854.77
(5, 5, 3, 0, 0)	-84.22	307.52	-73.03	477.990	112.52
(10, 10, 5, 0, 0)	-34.43	-1250.52	72.48	-3238.19	-1940.80
(5, 5, 3, 10, 3)	239.35	7592.18	12414.36	6278.59	10998.83

〈표 3.14〉 감마분포에서의 절대편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Absolute bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	13670.93	13640.50	13656.42	13146.93	13171.47
(5, 0, 0, 10, 3)	729761.94	129712.18	128929.56	125684.86	125224.54
(10, 0, 0, 50, 5)	3531763.00	640660.50	618241.10	640235.10	618030.40
(5, 5, 3, 0, 0)	12832.08	35922.59	12929.25	35573.46	13711.56
(10, 10, 5, 0, 0)	12611.09	66429.19	13382.18	65738.91	17075.69
(5, 5, 3, 10, 3)	793763.40	132731.70	129525.50	133076.60	129104.20

〈표 3.15〉 감마분포에서의 절대상대편향 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	Relative bias				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	1.11E-05	1.11E-05	1.11E-05	1.07E-05	1.07E-05
(5, 0, 0, 10, 3)	4.39E-04	7.81E-05	7.76E-05	7.57E-05	7.54E-05
(10, 0, 0, 50, 5)	1.04E-03	1.88E-04	1.82E-04	1.88E-04	1.82E-04
(5, 5, 3, 0, 0)	1.03E-05	2.87E-05	1.03E-05	2.85E-05	1.10E-05
(10, 10, 5, 0, 0)	9.82E-06	5.17E-05	1.04E-05	5.12E-05	1.33E-05
(5, 5, 3, 10, 3)	4.71E-04	7.88E-05	7.69E-05	7.90E-05	7.67E-05

〈표 3.16〉 감마분포에서의 RMSE 결과, $n = 20,000$

parameter ($\alpha, \beta, \gamma, \tau, \delta$)	RMSE				
	가구 조사			가구와 사람 추가조사	
	M1	M2	M3	M4	M5
(5, 0, 0, 0, 0)	17194.00	17178.29	17191.03	16562.59	16583.04
(5, 0, 0, 10, 3)	894349.48	161310.83	159307.82	157875.90	156301.18
(10, 0, 0, 50, 5)	4413944.91	801353.70	769728.10	796661.60	766536.50
(5, 5, 3, 0, 0)	16012.67	45428.03	16160.02	44531.59	17149.47
(10, 10, 5, 0, 0)	16090.07	84561.46	17079.34	84057.22	21228.10
(5, 5, 3, 10, 3)	997138.59	164765.61	158819.00	163846.80	158562.90

4. 결론

본 연구에서는 승인통계 중 가구 조사에서 인구 기반 가중치만을 사용한 경우와 표본설계 당시에 고려되었던 가구 기반 층화정보와 인구 기반 층화정보를 모두 이용하여 만든 보정 가중치를 사용한 경우에서 추정의 정밀성을 모의실험을 통하여 비교하였다. 또한 실사 중에 가구 정보를 얻지 못하고 연령 및 성별과 같은 사람관련 층화정보만을 얻은 추가 보완조사가 실시된 경우에 사용하는 가중치 보정 방법의 성능을 모의실험을 통해 비교하였다.

모의실험 결과 표본설계 당시에 사용된 가구관련 층화정보와 인구관련 정보를 모두 사용하는 방법이 추정의 정밀성을 보장하는 매우 우수한 결과를 주고 있음을 확인하였으며 이는 추가 보완조사가 이루어진 경우에도 모든 정보를 이용할 때 우수한 결과를 주고 있다. 물론 이 결과들은 가구 층화 변수의 층 평균 간에 차이가 있을 때 얻어진 결과로 어느 정도 예상된 결과이기도 하다. 모의실험에서 오차의 분포는 결과와 큰 차이를 보이지 않았으며 표본 규모도 결과에 큰 영향을 미치지 않았다. 가장 큰 영향을 주는 것은 층화 변수의 층간 평균의 차이이며 만약 층간 평균의 차이가 있거나 혹은 차이가 없다는 확실한 증거가 없다면 모든 층화정보를 사용하여 최종 가중치를 결정하는 것이 타당하다고 판단된다.

(2017년 7월 26일 접수, 2017년 9월 12일 수정, 2017년 10월 12일 채택)

감사의 글

이 연구는 2017년 한국외국어대학교 교내연구비 지원을 받아 수행되었음.

참고문헌

- Cochran, W. G. (1977). *Sampling techniques (3rd edition)*, John Wiley & Sons, New York.
- Fuller, W. A. (2000). Two-Phase sampling, SSC Annual meeting. *In proceedings of the survey methods sections*, 23-30, Ottawa, Canada.
- Hidiroglou, M. A. (2001). Double sampling, *Survey methodology*, Vol. 27, 143-154.
- Kim, J. K. and Yu, C. L. (2011), Replication variance estimation under two-phase sampling, *Survey Methodology*, Vol. 37, 67-74.
- Koyuncu, N., Kadilar, C. (2009). Family of estimators of population mean using two auxiliary variables in stratified random sampling, *Communications in statistics-Theory and Methods*, Vol. 38, 791-802.
- Lee, S. E., Jin, Y. Shin, K.-I. (2015). A note on complex two-phase sampling with different sampling units of each phase. *Communications for Statistical Applications and Methods*, Vol. 22, No. 5, 435-443.
- Oh, J.-T. and Shin, K.-I. (2016). An outlier weight adjustment using generalized ratio-cum-product method for two phase sampling, *The Korean Journal of Applied Statistics*, Vol. 29, No. 7, 1185-1199.

A Study on Post-Weight Adjustment Method Using Informations of Stratification in Household Survey

Shin, Key-Il¹⁾ · Lee, Sang Eun²⁾

Abstract

In sample survey, stratified sampling is widely used for accurate parameter estimation. It is the amount of information contained in the sampling frame that determines the success or failure of the stratified sampling method. If there is not enough information for stratification at the time of sampling design, information can be obtained using double sampling (two-phase sampling). On the other hand, if additional information related to stratification is obtained after the survey, post weight adjustment is used to improve the accuracy of the parameter estimation. In this way, information related to stratification variables in stratified sampling has a great influence on the accuracy of estimation. In this paper, we investigate the use of information for stratification contained in the sampling frame when using stratified sampling in household survey. Simulation results were used to compare the results of applying the existing household weights when using person weights at the time of sample design. In addition, we applied the weight adjustment method in the case where the supplementary survey was made by the additional individual survey without household information, and compared the performance of each method through simulation.

Key word : stratified sampling, double sampling, household weight, person weight

1) Professor, Dept. of Statistics, Hankuk University of Foreign Studies

2) Professor, Dept. of Applied Statistics, Kyonggi University