

제3장 순환표본조사 표본설계 및 가중치 작성방법 연구



최재혁·김경미

I. 서론

1. 연구배경 및 필요성

□ (연구배경) 2015년 인구주택총조사(센서스) 패러다임 전환으로 전수조사를 실시하지 않고 등록센서스(전수항목)와 표본조사(표본항목; 표본추출률 20%*)만을 실시하게 됨

* 일반조사구를 제외한 특수조사구는 전수조사함

- 등록센서스를 통해 인구주택총조사의 전수항목이 매년 작성되므로 표본항목에 대한 연간 작성방법의 검토가 필요함
- 과거 센서스**와의 시계열 단절을 방지하기 위해 읍·면·동 단위의 소지역 통계 작성 및 새로운 모수 추정방안 강구가 필요한 시점임

** 과거의 센서스는 전수, 표본항목 모두 읍·면·동 단위의 통계를 생산함

□ (연구 필요성) 2011년 인구총조사과의 “순환센서스(Rolling Census) 도입방안 연구” 과제 수요를 통해 2012년부터 관련 연구를 수행함

- 동 연구의 실효성을 높이기 위해 경상남도 거창군 전체와 인천광역시 연수구 5개동(선학동, 연수1, 2, 3동, 청학동)을 대상으로 시험조사 실시
- 순환센서스 전수항목이 아닌 표본항목에 대한 조사부문에서 실시되므로 순환표본조사(rolling sample survey)의 방법론의 전반적인 검토가 필요함
- 선행연구 및 해외사례 검토 등을 통해 순환센서스의 표본조사에 대한 실제적인 적용방법 검토가 필요함
- 순환표본조사의 표본설계 및 가중치 작성방법 등 기본적인 방법론 검토가 필요함

2. 연구목적

- 순환표본조사는 주기를 통해 표본단위가 중복되지 않게 추출되는 것이 중요하므로 설계단계에서 관리가 필요함
 - 순환표본조사의 표본설계를 위해서는 주기적으로 모집단 변동을 반영할 수 있는 모집단 리스트가 필수적임
 - 일정한 방법과 효율적인 표본설계로 순환표본조사를 실시해야 일정기간을 대상으로 하는 다년추정값 등 순환표본조사의 장점을 충분히 활용할 수 있음
- 순환표본조사의 결과는 조사시기별로 횡단추정값뿐만 아니라, 일정기간을 대상으로 하는 다년추정값도 작성할 수 있음
 - 다년추정값을 작성하는 방안은 일반적인 가정과 조건이 필요하며, 이를 적용하기 위해서는 심층적인 검토가 필요함
 - 다년추정값을 작성하는데 필요한 가중치 작성방안에 대한 방법론 정립이 필요함



3. 연구범위 및 한계

- 본 연구는 순환센서스 도입에 필요한 순환표본조사의 기본적인 방법론 검토만을 실시하며, 시험조사 자료 분석 및 활용 측면은 다루지 않음
- 순환표본조사는 표본추출틀(sampling frame)의 매 시점 갱신(수정·보완)을 전제로 출발하므로 이에 대한 가정을 통해 연구를 수행함

II. 순환표본조사의 표본설계

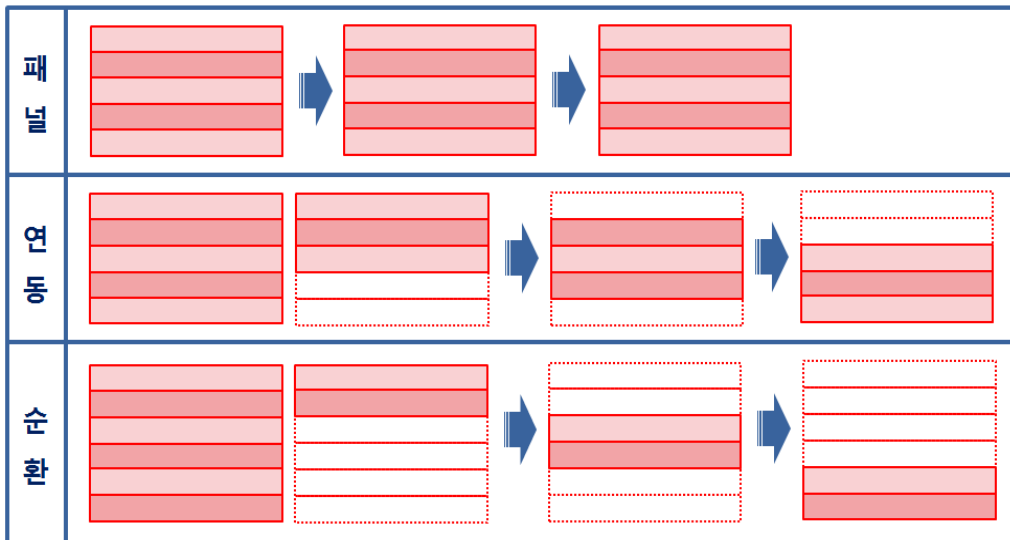
1. 순환센서스

- (개념) 인구주택총조사(센서스)를 일시(동일 기간, 동일 기준시점)에 실시하지 않고 일정기간, 일정주기로 나누어 조사하는 방법
 - 현재 등록센서스 전환에 따라 전수부문은 매년 공표될 계획이지만 표본항목은 5년마다 실시될 계획이며, 이를 5년 동안(매월 조사) 나누어 조사하는 개념임
- (장점) 소지역(시·군·구, 읍·면·동)의 통계를 공표수준에 따라 매년 생산할 수 있고 고정된 조사원 고용으로 비표본오차를 줄일 수 있으며, 매년 균등한 예산을 배정할 수 있는 등, 신뢰성 확보와 예산절감 부문에서 효율적인 조사운영이 가능함
- (활용) 프랑스가 2004년 순환센서스 체계로 이행하였으며, 미국은 2005년 long form(우리나라의 표본항목에 해당)에 대해서만 순환센서스로 운용
 - 미국 센서스의 long form은 지역사회조사(American Community Survey)로 변경되어 실시됨

2. 순환표본조사

- 순환표본조사는 연속조사인 패널조사(panel survey), 연동조사(rotation survey)와 유사한 방법임
 - 패널조사는 추출된 표본단위를 계속하여 추적·조사하는 방식임
 - 연동조사는 추출된 표본단위를 1차 조사하고 2차 조사 이후부터는 이중 일부를 교체하여 조사하는 방식임

[그림 3-1] 연속표본조사의 종류와 특성



- 일반적으로 가구에 대한 표본조사는 다원 층화 2단 집락추출방법(multi-way stratified two-stage cluster sampling)을 적용하고 있음
 - 가구원 조사인 경우는 가구 내 가구를 모두 조사하는 3단 집락추출 방법을 적용
- 층화 2단 집락추출방법*
 - * 필요에 따라 3단 집락추출로 변형될 수 있음



- 층화: 16개 시도(또는 시·군·구), 읍면/동 구분, 아파트/일반 구분 등의 층화변수 사용
- 1차 추출단위: 조사구
- 2차 추출단위: 가구
- 3차 추출단위: 가구원
- 2010년 인구주택총조사에서 표본항목 조사(일반조사구의 표본추출률 10%*)는 2원 층화 1단 집락추출방법(two-way stratified one-stage cluster sampling)을 적용하여 실시하였음
 - * 일반조사구를 제외한 특수조사구는 전수조사함
 - 층화: 시·군·구, 아파트/일반 구분
 - 추출단위: 조사구
 - 조사단위: 조사구 내 가구 및 가구원의 전수를 조사함
- 순환표본조사는 일반적인 표본조사의 방식을 그대로 유지하면서 진행되며, 각 조사시기의 표본설계가 독립적인 추정이 가능하도록 구성되어 있어야 함
 - 층화 및 추출방식(표본설계)은 전체 표본과 각 조사시기의 순환표본에 동일하게 적용되어야 함

3. 순환표본조사의 표본크기

- 순환표본조사의 각 조사시점별 표본크기 배정은 균등배정(equal allocation)을 기본으로 함
 - 층별 배정방법 등을 통해 결정된 최종 표본크기를 조사시점별로 조사 시기 수(k)만큼 등분하여 각 차수의 최종 표본크기로 결정함
 - 각 조사시기별 층내 횡단추정값이 생산될 수 있도록(공표수준, 표본 오차 고려) 층별로 충분한 표본크기를 배정해야 함

- 일반적인 표본조사에서 추정에 필요한 표본크기가 각 조사시기별 표본 크기와 동일하게 적용되어야 함
 - 전체 순환표본조사의 표본크기가 추정의 정도(precision)를 확보해야 하는 수준으로 확보되어야 하는 것이 아니라 각 조사시기의 표본 크기가 추정의 정도를 확보할 수 있는 수준이어야 함
 - 예: 5차 주기의 순환표본조사에서 95% 신뢰수준, 목표오차(CV) 10%에 필요한 표본크기가 2,000개라고 가정하면,
 - 각 조사시기의 신뢰수준과 목표오차를 동일하게 적용하기 위해서는 2,000개의 순환표본조사 표본크기가 필요한 것이 아니라 각 조사시기에서 2,000개의 표본크기가 필요하며, 최종적으로 전체 순환 표본조사의 표본크기는 총 10,000개가 필요함

4. 순환표본조사의 총화

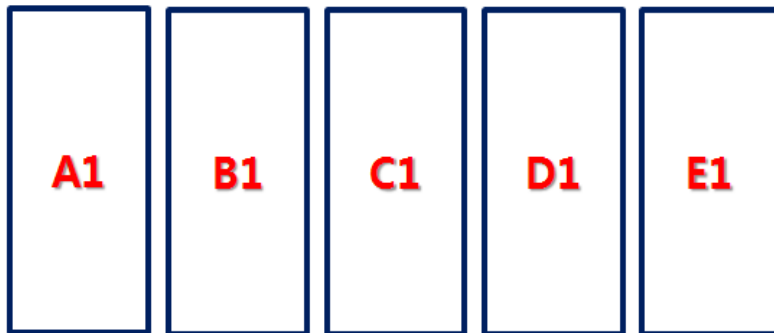
- 순환표본조사의 총화는 일반적인 표본조사의 총화와 동일한 방법으로 설계함
 - k 차 주기 순환표본조사는 모집단을 k 개로 분할(partition)하여 각 조사시기별로 표본을 추출하는 방법으로 구성되어 있음
 - 각 조사시기별로 총화기준이 달라지지 않으며, 전체 표본의 총화를 구성한 후 각 조사시기별로 구분하는 방법으로 진행됨
 - 즉, 각 k 차 주기 순환표본조사라고 가정하면, 전체 모집단이 k 개로 분할되어 있고 전체 모집단이 L 개 층으로 총화되었다면 k 개로 분할된 각 조사시기별 모집단도 L 개 층으로 총화되어 있음
- 내재적 총화변수(분류지표)는 각 차수에서 동일하게 적용해야 함
 - k 개로 분할된 모집단은 매 조사시기별로 조사시점에서 추출이 진행(수정·보완된 모집단 사용)됨

- 추출단계에서 적용될 내재적 층화변수는 수정·보완된 모집단에 대해 동일하게 적용함

5. 순환표본조사의 추출방법

- 최초 모집단을 k 개로 분할함
 - 모집단의 특성이 균등하게 배정되도록 구분하는 것이 가장 중요함
 - 예: 5개년 순환표본조사를 실시한다면, 전체 모집단을 A1, B1, C1, D1, E1그룹으로 구분함

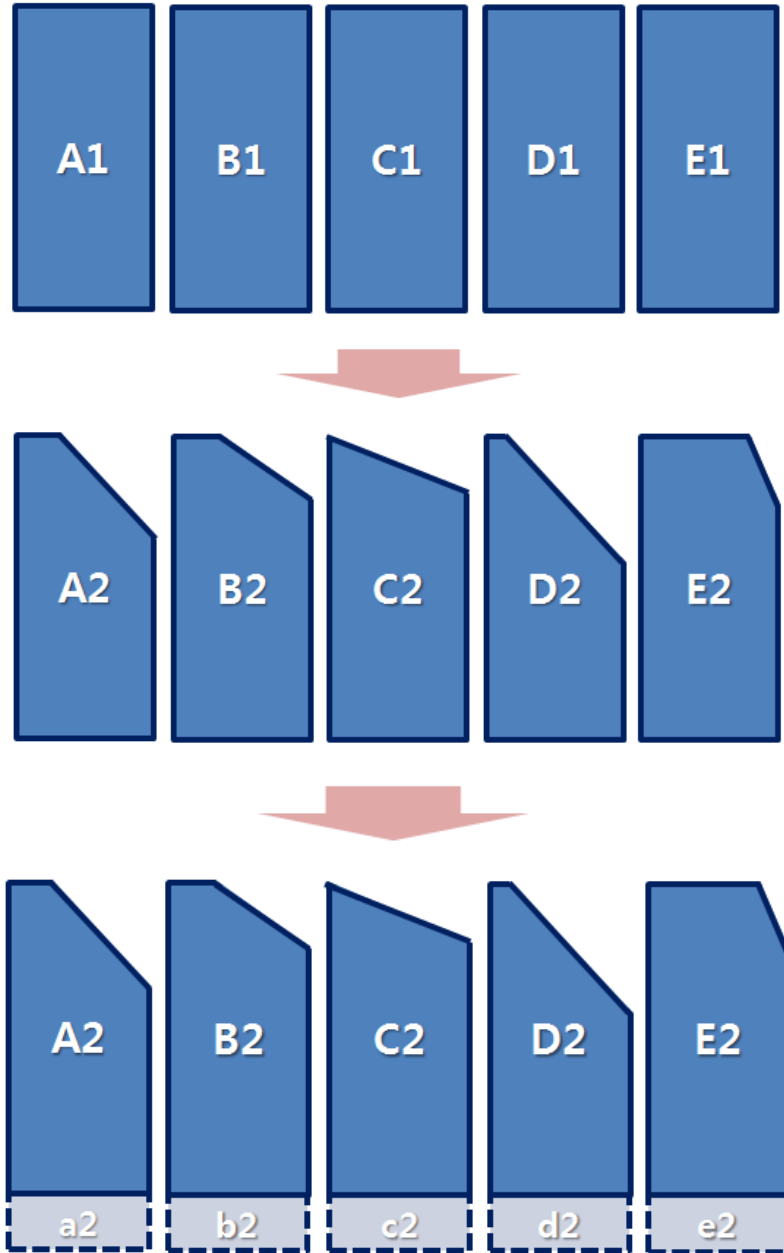
[그림 3-2] 순환표본조사의 모집단 분할(5개)



- 각 그룹은 동일한 기준으로 층화되어 있음
 - 각 차수별로 $n1$ 개의 표본을 추출하기로 하고 1차년도에 최초 A1그룹에서 $n1$ 개의 표본을 추출함
 - 최종표본은 $5 \times n1 = n1 + n2 + n3 + n4 + n5$ 로 구성함
- 2차 조사시기에서 모집단을 수정·보완하여 표본을 추출함
 - 모집단에서 제외되는 단위가 있는 경우 모든 그룹에서 단위를 제외하여 전체 모집단을 A2, B2, C2, D2, E2그룹으로 수정·구성함

- 모집단이 갱신되어 추가되는 모집단 단위에 대해서는 최초 그룹을 분할하는 기준과 동일한 방법으로 5개의 그룹(a2, b2, c2, d2, e2그룹)으로 분할함
- 2차 조사시기에서 최종적인 모집단 그룹은 $A2+a2$, $B2+b2$, $C2+c2$, $D2+d2$, $E2+e2$ 로 구성되게 됨
 - 여기서 b2는 추가된 모집단 단위 중 B그룹에 배정된 단위이고 B2는 B1에서 모집단 단위의 변동 중 삭제된 단위가 제거된 그룹임
- 2차 조사시기의 표본은 $B2+b2$ 그룹에서 동일한 표본크기 $n2$ 에 대해 동일한 내재적 층화변수, 동일한 표본추출방법으로 표본을 추출함
- 1차 조사시기의 표본은 A1에서 추출되었지만 2차 조사시기에서 그 그룹이 $A2+a2$ 로 변경되며, 이는 추후 추정단계에서 사용됨
- 3차 조사시기에서도 추가적으로 모집단을 수정·보완하여 표본을 추출함
 - 모집단에서 제외되는 단위가 있는 경우 모든 그룹에서 단위를 제외하여 전체 모집단을 $A3$, $B3$, $C3$, $D3$, $E3$ 그룹으로 수정·구성함
 - 여기서 C3는 2차 조사시기에서 수정·보완된 그룹 $C2+c2$ 에서 모집단 단위의 변동 중 삭제된 단위가 제거된 그룹임
 - 3차 조사시기에서 갱신되어 추가되는 모집단 단위에 대해서는 1차, 2차 조사시기에서 그룹을 분할하는 기준과 동일한 방법으로 5개의 그룹(a3, b3, c3, d3, e3그룹)으로 분할함
 - 3차 조사시기에서 최종적인 모집단 그룹은 $A3+a3$, $B3+b3$, $C3+c3$, $D3+d3$, $E3+e3$ 로 구성되게 됨

[그림 3-3] 순환표본조사 모집단의 1, 2차 조사단계에서의 변화



□ 각 조사시기에서 동일한 방법으로 표본을 추출하되, 이미 추출된 그룹의 모집단의 변동(수정·보완)은 각 차수별 분석 및 최종분석단계에서 사용됨

① 1차 시점 추출프레임: A1

- 모집단 구성: A1, B1, C1, D1, E1

② 2차 시점 추출프레임: B2+b2

- 모집단 구성: A2+a2, B2+b2, C2+c2, D2+d2, E2+e2

③ 3차 시점 추출프레임: C3+c3

- 모집단 구성: A3+a3, B3+b3, C3+c3, D3+d3, E3+e3

④ 4차 시점 추출프레임: D4+d4

- 모집단 구성: A4+a4, B4+b4, C4+c4, D4+d4, E4+e4

⑤ 5차 시점 추출프레임: E5+e5

- 모집단 구성: A5+a5, B5+b5, C5+c5, D5+d5, E5+e5

□ 최초 모집단은 A1, B1, C1, D1, E1로 시작되었고, 5차 조사시기의 최종 조사가 완료된 시점의 모집단은 A그룹(A5+a5), B그룹(B5+b5), C그룹(C5+c5), D그룹(D5+d5), E그룹(E5+e5)으로 구성됨

○ 최초 모집단 구성(1차 조사시기)에서 각 그룹의 크기는 동일하게 구성되나 모집단이 수정·보완된 후에는 모집단의 각 그룹의 크기는 동일하지 않으며, 이는 추정단계에서 보정하여 해결함

○ 표본크기는 최초 설계단계에서 결정하기 때문에 각 조사시기별로 동일한 표본크기를 가지게 됨

□ 각 차수의 조사가 종료된 시점의 모집단 구성

① 1차 조사시기 모집단 구성

- A그룹(A1), B그룹(B1), C그룹(C1), D그룹(D1), E그룹(E1)

② 2차 조사시기 모집단 구성

- A그룹(A_2+a_2), B그룹(B_2+b_2), C그룹(C_2+c_2), D그룹(D_2+d_2), E그룹(E_2+e_2)

③ 3차 조사시기 모집단 구성

- A그룹(A_3+a_3), B그룹(B_3+b_3), C그룹(C_3+c_3), D그룹(D_3+d_3), E그룹(E_3+e_3)

④ 4차 조사시기 모집단 구성

- A그룹(A_4+a_4), B그룹(B_4+b_4), C그룹(C_4+c_4), D그룹(D_4+d_4), E그룹(E_4+e_4)

⑤ 5차 조사시기 모집단 구성

- A그룹(A_5+a_5), B그룹(B_5+b_5), C그룹(C_5+c_5), D그룹(D_5+d_5), E그룹(E_5+e_5)



Ⅲ. 순환표본조사의 가중치 부여방법

1. 추출가중치

- 순환표본조사는 각 차수별로 추출이 진행되므로, 각 추출단계에서 추출가중치를 작성하게 됨
 - 각 조사시기에서 모집단의 변동(수정·보완)이 반영되므로 해당 조사시기별 추출틀(A_1 , B_2+b_2 , C_3+c_3 , D_4+d_4 , E_5+e_5)의 정보를 사용하여 설계가중치를 계산함
 - 모집단 정보(모집단 크기 및 인구·사회학적 분포)가 변경되었다 하더라도 최초 결정한 표본크기는 변경하지 않고 그대로 적용함
 - 따라서 각 차수별 추정값의 오차는 변동이 발생할 수 있음

□ 각 차수별 추출가중치는 다음과 같이 계산됨

① 1차 조사시기 추출가중치: $\frac{A1}{n1}$

② 2차 조사시기 추출가중치: $\frac{B2+b2}{n2}$

③ 3차 조사시기 추출가중치: $\frac{C3+c3}{n3}$

④ 4차 조사시기 추출가중치: $\frac{D4+d4}{n4}$

⑤ 5차 조사시기 추출가중치: $\frac{E5+e5}{n5}$

□ 각 조사시기별 추출가중치는 층별로 진행되며, 2단 이상 집락추출로 설계된 경우는 가구추출확률(조건부 확률)을 적용하여 계산함

- 2단 집락추출에서 모집단 및 표본은 조사구 기준으로 구성되는 것을 기본으로 함

2. 최종가중치 보정

□ 최종가중치의 보정은 추정량의 종류에 따라 선택하여 결정함

- 횡단추정량의 경우는 추정량이 작성되는 조사시기에서의 모집단 정보를 기준으로 보정함

① 1차 조사시기 모집단 보정기준: $A1+B1+C1+D1+E1$

② 2차 조사시기 모집단 보정기준: $A2+a2+B2+b2+C2+c2+D2+d2+E2+e2$

③ 3차 조사시기 모집단 보정기준: $A3+a3+B3+b3+C3+c3+D3+d3+E3+e3$

④ 4차 조사시기 모집단 보정기준: $A4+a4+B4+b4+C4+c4+D4+d4+E4+e4$

⑤ 5차 조사시기 모집단 보정기준: $A5+a5+B5+b5+C5+c5+D5+d5+E5+e5$



- 다년 추정치를 작성하는 경우, 최종가중치 작성방법이 추정방법에 따라 달라질 수 있음
- 최종 시점의 가중치를 사용
 - 전 시점 가중치의 평균을 사용
 - 최종가중치의 보정은 횡단추정량과 마찬가지로 추정량이 작성되는 조사시기의 모집단 정보를 기준으로 보정함
 - 다년 추정치 작성 단계에서 각 추정치의 생성시점(조사시기)의 모집단 정보를 기준으로 보정한다면, 조사시기별 최종가중치의 변동이 커 추정량이 안정적이 못하기 때문에 사용이 제한적임
- 각 조사시기별 최종가중치의 보정은 다음과 같이 계산됨

① 1차 조사시기 최종가중치: $(A1+B1+C1+D1+E1) \times \frac{A1/n1}{\sum A1/n1}$

② 2차 조사시기 최종가중치:

$$(A2+a2+B2+b2+C2+c2+D2+d2+E2+e2) \times \frac{B2+b2/n2}{\sum B2+b2/n2}$$

③ 3차 조사시기 최종가중치:

$$(A3+a3+B3+b3+C3+c3+D3+d3+E3+e3) \times \frac{B3+b3/n3}{\sum B3+b3/n3}$$

④ 4차 조사시기 최종가중치:

$$(A4+a4+B4+b4+C4+c4+D4+d4+E4+e4) \times \frac{B4+b4/n4}{\sum B4+b4/n4}$$

⑤ 5차 조사시기 최종가중치:

$$(A5+a5+B5+b5+C5+c5+D5+d5+E5+e5) \times \frac{B5+b5/n5}{\sum B5+b5/n5}$$

IV. 결론 및 제언

1. 순환표본조사 표본설계

- 순환표본조사의 표본추출은 각 기준시점에 맞추어 수행됨
 - 순환표본조사의 장점은 공표주기가 상대적으로 짧지 않은 조사의 단점을 보완할 수 있다는 점이 가장 크며, 이를 위해 모집단의 변동이 반영될 리스트 확보가 필요함
 - 최초 모집단 리스트를 k 등분(예. 5개라면, 통계적인 기준으로 A1, B1, C1, D1, E1그룹으로 분할)하여 분할된 최초 그룹에서 표본을 추출함
 - 2차 조사시기에서 모집단의 변동이 발생하게 되면(단위 삭제 및 추가; 모집단 수정·보완), 추가된 모집단 단위에 대해 1차 조사시기 기준과 동일한 방법으로 분할하고 1차 조사시기 모집단이 수정된 모집단과 추가된 모집단의 두 번째 그룹에서 표본을 추출함
 - 3차 이후 조사시기부터는 동일한 방법으로 모집단을 수정·보완한 후 표본을 추출하되, 추출하지 않은 그룹(이미 추출하였거나 이후 추출할 그룹)의 모집단 변동은 각 조사시기별 분석 및 최종분석 단계에서 가중치 보정값으로 적용됨

2. 순환표본조사의 가중치 작성방법

- 순환표본조사는 각 조사시기별로 추출이 진행되므로, 각 조사시기인 추출단계에서 설계가중치를 작성하게 됨
 - 각 조사시점에서 모집단의 변동이 반영되므로 조사시점별 추출틀 정보를 사용하여 설계가중치를 계산함
 - 모집단 정보(모집단 크기 및 인구·사회학적 분포)가 변경되었다 하더라도 최초 결정한 표본크기는 변경하지 않고 그대로 적용함



- 따라서 각 조사시기별 추정값의 오차는 변동이 발생할 수 있음
- 최종가중치의 보정은 추정량의 종류에 따라 선택하여 결정함
 - 횡단추정치(橫斷推定値)의 경우는 추정치가 작성되는 시점(조사시기)에서의 모집단 정보를 기준으로 보정함
 - 다년 추정치를 작성하는 경우는 최종 조사시기의 가중치를 사용하거나 각 조사시기별 가중치의 평균을 사용하므로 횡단추정량과 마찬가지로 추정량이 작성되는 조사시기를 대표하는 그룹으로 보정함
 - 다년 추정치 작성 단계에서 각 추정치의 생성시점(조사시기)의 모집단 정보를 기준으로 보정한다면, 조사시기별 최종가중치의 변동이 커 추정량이 안정적이 못하기 때문에 사용이 제한적임

3. 향후 연구 및 제언

- 순환표본조사(순환센서스 포함)를 실시하기 위해서는 표본추출틀의 구축 및 수정·보완 방법에 대한 선행적인 연구가 필요함
 - 2015년 이후 인구주택총조사가 등록기반 방식으로 전환되어 매년 주택DB, 가구DB, 가구원DB 구축이 가능함에 따라 가구(가구원) 단위 표본추출틀의 포함수준(coverage)을 향상시킬 수 있는 기반이 마련 되어야 함
 - 인구주택총조사 표본항목을 순환표본조사 체계로 개편하기 위한 연구가 이루어지고 있으며 이에 병행하여 순환 표본조사 체계에 맞는 표본추출틀 구축 및 보완 방안에 대한 연구가 필요함
- 인구주택총조사의 가구(조사구 포함)에 대한 표본추출틀 구축방안에 대한 추가적인 연구와 검토가 필요함

- 가장 최신의 주택공시자료, 건축물대장 및 주민등록자료를 연계한 등록상의 주택 및 가구 명부에 2014년 11월 가구주택기초조사를 통한 등록되지 않은 주택 외 거처와 가구를 보완하여 가구DB를 구축해야 함
 - 각 가구의 주소와 공간정보를 활용하여 조사구번호를 부여하는 방안을 검토해야 함
 - 기 구축된 표본추출틀을 순환표본조사 조사시기에 맞춰 각각의 그룹이 모집단을 대표하도록 균등하게 분할하는 방법이 필요함
- 인구주택총조사의 가구에 대한 표본추출틀 수정·보완 방안에 대한 추가적인 연구와 검토가 필요함
- 각종 등록상의 거처 및 가구자료를 통해 구축된 가구DB와 전년도 표본추출틀을 연계하여 표본추출틀 내 가구의 삭제, 추가 및 변경이 필요하며, 각 가구의 주소와 공간정보를 활용하여 조사구번호를 갱신해야 함
 - 조사구 번호는 가능한 기존 번호를 유지하되 가구 수에 따라 분할, 삭제 및 신규 부여함
 - 순환표본조사는 조사시기별 분할된 그룹을 이용하여 표본설계가 이루어져야 하므로 매년 발생하는 신규가구를 각 그룹에 균등하게 배분해야 하며, 각 그룹에서 삭제된 단위를 고려하여 차등 배분할지에 대한 추후 연구가 필요함
 - 기 분할된 그룹별로 삭제 및 신규가구를 보완하는 체계로 관리
- 순환표본조사를 등록센서스의 표본항목 조사에서 적용하기 위해 향후 연구가 지속적으로 진행되어야 함
- 미국 사회조사를 중심으로 순환표본조사의 다년 추정량의 작성 및 활용방안에 대한 검토

- 프랑스 순환표본조사의 사례검토를 통해 전반적인 조사설계를 검토
- 최종적으로 순환센서스(표본항목 조사를 순환표본조사로 실시: 표본 추출률 20%) 시험조사 자료분석을 통해 소지역(시·군·구 및 읍·면·동) 추정방안 및 통계 작성방안에 대한 심층 분석이 필요함



【참고문헌】

- 김중익(2013), 가구 추출틀 구축 및 보완 방법
등록센서스과(2014.2), 행정자료 정비를 위한 주소표준화 매뉴얼
인구총조사과(2012), 2010년 인구주택 총조사 표본결과 심층분석 연구
통계개발원(2014), 순환센서스 적용을 위한 표본틀 구축 및 갱신방안 연구
_____ (2013), 순환센서스 도입검토를 위한 시험조사 자료분석 결과보고
_____ (2009), 인구센서스 방법론 연구Ⅱ
_____ (2008), 인구센서스 방법론 연구Ⅰ
_____ (2007), 2010년 인구주택총조사 방법론 연구
통계청(2011), 2015 인구주택총조사 표본조사방안 연구(부처 간 공동 과제 학습 과정)
표본과(2013.12), 가구 표본 추출틀 작성 및 관리방안 검토 보고
Uhl(2011), Building and Maintaining the Master Address File
US Census Bureau(2009), ACS Design and Methodology