

4. 가중치 산정 방법

- 본 조사가 확률적 표본추출 방법론으로 설계되었지만, 조사되어진 결과를 표본 추출에 따른 추출률의 차이와 응답률 및 모집단에 대한 정보 등을 이용하여 모집단의 구조와 표본 구조를 가중치를 사용하여 맞추으로써 추정의 정확도를 높임
- 추정단계에서 가중치를 이용하면 모집단에 대한 특성치인 모수에 대한 비편향 추정량(unbiased estimator)을 얻을 수 있음
 - 만약, 통계분석 과정에서 가중치를 무시하고 분석한 추정치는 심각한 편향(bias)이 발생할 수 있음
 - 표본의 크기가 큰 대규모 조사에서 문제가 되는 것은 추정량의 편향이기 때문에 추정과정에서 반드시 가중치를 이용해야 함
- 본 조사에서는 ㉠ 설계가중치 산정, ㉡ 무응답에 대한 조정, ㉢ 모집단 정보를 이용한 조정 등의 세 가지를 고려하였고, 가중치 조정단계에서 이용한 모집단 정보는 2023년 「통계청 추계인구」임
- 표본추출은 층화 2단 확률비례 계통추출법을 적용하였는데, 1차 추출단위는 조사구이고, 2차 추출단위는 가구 및 만3세 ~ 69세 이하 가구원임
- 각 층에서 배정된 표본 조사구 수만큼을 조사구 내의 가구 수에 비례하는 확률비례 계통추출법에 따라 추출
- 본 조사에서 설계가중치는 표본추출 과정에서 층화 2단 확률비례 계통추출법을 적용함에 따라 나타나는 조사구 내의 가구별 추출확률의 차이를 반영할 수 있도록 산정
- 가중치 작성 과정에서 사용될 기호들을 정리하면 다음과 같음
 - L : 층의 수
 - N_h : 층 h 의 모집단 조사구 수
 - n_h : 층 h 의 표본 조사구 수
 - S_{hi} : 층 h 의 i 번째 조사구에 대한 크기의 측도(해당 조사구의 총 가구 수)

- $S_h = \sum_{i=1}^{N_h} S_{hi}$: 층 h 에서 크기의 측도에 대한 총합
- M_{hi} : 층 h 의 i 번째 조사구 내 가구 수
- m_{hi} : 층 h 의 i 번째 조사구 내 조사착수 가구 수(응답+거절)
- r_{hi} : 층 h 의 i 번째 표본조사구 내 조사완료 가구 수(응답)

5. 최종가중치

○ 설계가중치

- 본 조사의 설계가중치는 각 표본 조사구에 대한 표본 추출률의 역수와 표본 조사구에서 가구조사 착수율의 역수를 곱하여 다음과 같이 산출

$$\text{설계가중치} = \frac{S_h}{n_h S_{hi}} \times \frac{M_{hi}}{m_{hi}}$$

- 원칙적으로 각 표본 조사구에서는 10가구씩을 표본으로 조사하였기 때문에 $m_{hi} = 10$ 임

- $S_{hi} \approx M_{hi}$ 를 가정할 수 있는 경우(표본추출률 상의 조사구 내 가구 수와 실제 가구 수에 차이가 작은 경우) 각 층에서 설계가중치는 해당 층 내에서 일정한 값이 되어 설계가중치 $= \frac{S_h}{n_h \times 10}$ 으로 표현할 수 있음

- 이때, 각 지역 내 층에서 표본 가구들은 모두 동일한 설계가중치를 갖게 됨
- 실제 조사과정에서는 표본추출률의 조사구 내 가구 수와 실제 가구 수에 차이가 있어 각 층 내에서 설계가중치는 동일하지는 않지만 비슷한 값을 갖게 됨

○ 무응답 조정 가중치 반영

- 해당 지역 내 세부 층에서 조사 가구에 대한 설계가중치는 원칙적으로 동일
- 본 조사에서 무응답 조정은 표본 조사구 단위로 진행되었으며 무응답 조정계수는 다음 식에 따라 구함

$$\text{무응답 조정계수} = \frac{m_{hi}}{r_{hi}}$$

○ 2022년 추계인구를 이용한 가중조정

- 모집단 정보를 이용한 가중치 조정은 모집단에 대한 정보를 이용하여 모집단의 구조와 표본 구조를 유사하게 맞추으로써 추정의 정확도를 높이는 것을 목적으로 함
- 가중치 작성단계에서 사용한 모집단에 대한 정보는 2021년 등록센서스 표본추출틀 자료임
- 본 조사에서는 2022 스마트폰 이용률을 이용하여 2023년 현재 스마트폰 이용자 분포를 추정하였으므로 이를 레이킹 비 접근법(Raking Ratio Method)을 적용하여 가중치를 조정
- 레이킹 비 접근법(Raking Ratio Method)에 활용한 변수는 17개시도 * 성별 (2) * 연령별(6) 변수들이며 <표 3>은 2023년 스마트폰 이용자에 대한 지역별 * 성별 * 연령별 분포임
- 최종 가중치는 다음 식에 따라 산정

$\text{최종 가중치} = \text{설계가중치} \times \text{무응답 조정계수} \times \text{모집단 정보 조정계수}$
