

**AI SUM Q5:** 가구조사 시 가구 내에서 1명을 왜 랜덤하게 추출해야 하는가?

**AI SUM Q5:** Why within-household selection?

▷ **질문 설명**

가구조사(household surveys) 시 표본으로 추출된 가구 내의 가구원들 중 1명을 왜 랜덤하게 추출해서 조사를 진행해야 하는가? 가구 내에 있는 가구원 모두를 조사하면 안 되는 것인가?

▶ **나침반 보기**

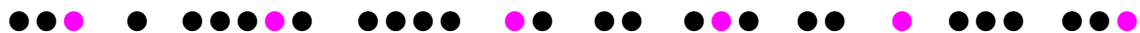
국내에서 진행되는 가구조사들을 잘 살펴보면, 어떤 조사는 표본가구 내 가구원 중 1명을 랜덤하게(확률적으로) 추출해서 조사를 진행하는 반면에 어떤 조사는 표본가구 내 가구원들을 모두 조사한다. 둘 중 어떤 방법이 올바른 것일까? 가구 내에서 1명을 랜덤하게 추출하는 것이 올바른 방법이다. 일반적으로 가구조사는 ‘가구’가 아니라 ‘개인’에 관한 통계를 얻기 위해 진행된다. 예를 들어, 실업률, 흡연율 등과 같은 개인들에 관한 국가 통계, 선거여론조사에서 유권자 개인들의 후보자 지지율 등을 말한다. 그런데 왜 가구 내에서 1명을 랜덤하게 추출해야 하는 것일까? 이 방법은 1949년 미국의 사회학자이며 통계학자인 Kish(키쉬)가 처음으로 제시한 것으로 오랜 역사를 가지고 있다.

이 방법에 대한 이해를 돕기 위해 다음 그림들(A1, B1)을 살펴보자. 각 그림에서 ‘원’은 ‘사람(개인)’을 나타내며 전체 29명이다. 여기서 ‘●’는 표본으로 추출된 사람들이다.

A1



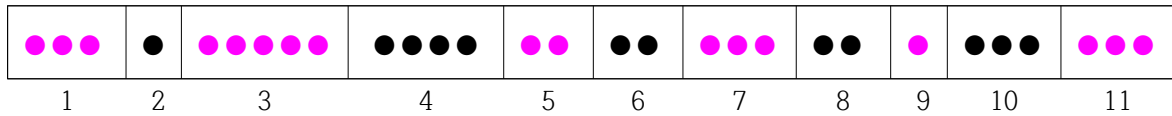
B1



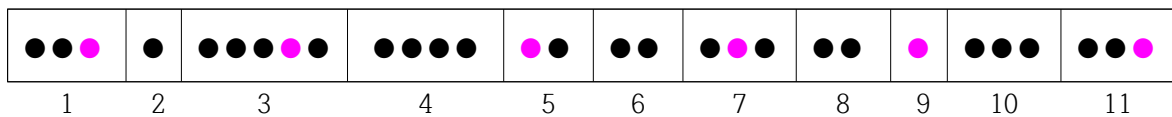
‘표본의 대표성’을 감안할 때, 여러분은 A1, B1의 표본 중 어떤 것을 선택하겠는가? ‘A1’이 비록 표본의 크기가 17명으로 더 크지만, 아마 여러분은 표본의 크기가 6명인 ‘B1’을 선택하게 될 것이다. 그 이유는 추출된 사람들이 거의 균일하게 넓게 퍼져있기 때문이다.

그런데 이것이 가구 내에서 1명을 랜덤하게 추출하는 것과 관련이 있을까? A1, B1은 원래 아래의 A2, B2와 같이 각 칸(cell)으로 구성된 것인데, 이 칸을 보이지 않도록 한 것이 A1, B1이다. 여기서 각 칸은 ‘가구(household)’를 나타내며, 여기서는 1인 가구, 2인 가구, 3인 가구 등 총 11개 가구로 구성된다(한국의 실제 일반가구의 수는 2021년 기준으로 21,448,463개이다). 이때 A2, B2는 1, 3, 5, 7, 9, 11번 가구가 추출된 것이며, 이들 6개 표본가구들은 추출 간격(sampling interval)이 ‘2’인 계통추출법(systematic sampling)으로 추출되었다(추출간격이 ‘3’인 경우, 1, 4, 7, 10번 가구가 추출된다). 따라서 A1은 표본가구 내 가구원 모두가 표본으로 추출된 것이고, B1은 표본가구 내에서 1명이 랜덤하게 추출된 것이다(가구원이 1명인 경우 그 1명을 조사한다). 결과적으로 B1과 같이 균일하게 넓게 추출되도록 하기 위해서는 B2에서처럼 가구 내에서 1명을 랜덤하게 추출해야 한다. ‘가구’는 최종 목적지가 아니라 개인을 추출하기 위한 중간 과정에 불과한 것이다.

### A2



### B2



만약 보다 빠른 시간 안에 목표로 하는 표본 크기(예: 17명)를 편리하고 쉽게 확보함으로써 조사비용을 절약하고자 한다면, 표본가구 내 가구원 모두를 조사하는 ‘A2’와 같은 방법을 사용할 수 있다. 하지만 이와 같은 방법을 사용하면 부정적인 영향이 반드시 따른다. 이는 ‘부정확한 통계’이다. 이렇게 되는 근본적인 이유는 동일 가구 내의 가구원들은 서로 유사한 특성을 가지고 있으므로 유사한 응답을 할 것이기 때문이다. 또한 동일 가구 내 가구원들은 응답하는

것을 옆에서 서로 지켜볼 수 있기 때문에 참조를 하게 될 것이고 특히, 사적이고 민감한 질문을 하게 되면 옆에 다른 사람이 있어 정확한 응답을 꺼릴 수 있기 때문이다. 아울러 이런 사실에도 불구하고, 가구원들을 전부 조사하는 것과 1명을 랜덤 추출하는 것이 조사결과에 차이가 없다고 누군가 주장을 한다면, 이는 조사비용을 아끼거나 편리하게 조사를 하기 위한 편법에 지나지 않는다.

사실, 가구 내에서 1명을 추출하느냐, 아니면 가구원 모두를 조사하느냐 하는 것은 일반적인 서베이방법론(survey methodology)에서 선택 사항이 아니다. 선택 사항은 가구 내에서 1명을 랜덤하게 추출하기 위해 다양한 방법들 중 어떤 방법을 사용하느냐 하는 것이다. 예를 들어, 표본으로 추출된 가구의 가구원 목록을 간단하게 작성한 후 난수를 이용하여 랜덤하게 1명을 추출하는 방법, (가구원 목록을 작성하지 않고) 조사 시점을 기준으로 가장 생일이 가까운 1명을 가구 내에서 조사하는 방법(최근 생일자 방법) 등을 선택할 수 있다.

이렇게 가구 내에서 1명을 랜덤하게 추출하는 방법은 ‘방문조사(대면조사, in-person surveys, face-to-face surveys)’ 뿐만 아니라 집전화(일반전화)를 이용한 ‘전화조사(telephone surveys)’에서도 동일하게 적용되어야 한다. 집전화의 경우, 가구원들이 같이 사용하지만, 면접원이 전화를 했을 때 전화를 받는 사람만을 대상으로 조사를 하면 응답자들이 특정 계층의 사람들에게 집중될 수 있다(이것은 방문조사 시 면접원이 가구를 방문했을 때 집 문을 열어주는 사람을 조사하는 것과 같은 상황이다). 예를 들어, 낮 시간에 집전화조사를 하면 고연령층, 주부, 직업이 없는 사람만이 주로 조사가 될 수 있다. 이로 인하여 조사결과가 부정확해질 수밖에 없다. 이것을 피하는 유일한 방법은 가구원 중 1명을 랜덤하게 추출하는 것이다.

한편, 조사 목적에 따라 가구 내에서 1명이 아닌 2명 이상(예: 미성년 1명, 성인 1명)을 랜덤하게 추출할 수도 있다. 가구 내 랜덤 추출(within-household selection)에 관한 보다 상세한 방법론은 다음 참고문헌들에서 확인할 수 있다. 첫 번째 참고문헌은 본 센터에서 진행한 국내 전화조사방법론에 관한 연구 논문으로, 가구 내 랜덤 추출(within-household selection)을 사용하였다.

Kim, S. W., Lee, S. K., Hong, S. J., and Park, S. H. (2012). “List-assisted RDD sampling in Korea: testing the feasibility of national survey under [within-household selection](#),” *International Journal of Public Opinion Research*, **24**, 79-92.

Groves, Fowler, Couper, Lepkowski, Singer, and Tourangeau (2009). *Survey Methodology*, pp. 134-137. Wiley

### ♣ 일화(에피소드)

2008년 본 센터에서는 국내 한 대형 조사회사를 대상으로 전화조사방법론 (telephone survey methodology)을 자문하였고, 조사의 정확성을 높이기 위해 일반전화(집전화)조사 시 가구 내에서 랜덤 추출(within-household selection) 방법을 사용하도록 권고를 하였다. 하지만 이 조사회사의 전화조사 총책임자는 난색을 보였다. 그런 방법은 미국에서는 사용이 가능할지 모르지만, 자신의 10년이 넘는 국내 전화조사 경험에 비추어 볼 때, 한국에서는 안 된다는 것이 그 사람의 기본적인 생각이었고, 시도조차 하려들지 않았다. 장기간의 반복적인 설명과 설득 끝에 시험조사를 해보기로 했다. 컴퓨터를 이용한 전화조사(CATI) 시스템을 가구 내 랜덤 추출을 할 수 있도록 보완하였으며, 이와 함께 가구 내 랜덤 추출을 할 수 있도록 면접원 교육을 진행하였다. 시험 조사는 어떻게 되었을까?

조사를 시작한지 불과 1시간도 채 안되어서, 전화면접원 중 1명이 전화를 걸은 가구 내 성인들의 목록을 컴퓨터 화면을 통해 입력을 하는데 성공을 했고, 이어서 바로 컴퓨터 프로그램을 이용해서 성인 1명을 랜덤하게 추출한 뒤, 그 추출된 성인을 대상으로 조사를 하기 시작했다. 그 전화조사 총책임자는 놀라움을 금치 못했다. “미국에서는 되고, 한국에서는 안 된다”는 공식은 무참히 깨졌다. 이윽고 얼마 안 되어 다른 면접원들도 같은 방식으로 조사를 진행하기 시작했다. 한국에서 전화조사의 새로운 역사가 시작되는 순간이었다.

참고로 2020년대 현재까지도 국내에서는 일반전화조사 시 가구 내 랜덤 추출을 하는 조사회사는 극히 드물다(그나마 좀 다행스러운 것은 요즘은 일반전화조사보다는 휴대전화조사를 더 많이 진행한다). 하지만 미국의 경우 일반전화조사 시 ‘가구 내 랜덤 추출’은 일반적으로 사용하는 방법론이다. ‘가구 내 랜덤 추출’은 책에 나와 있는 탁상공론적 이론이 아니고 널리 통용되는 기법이다.



---

‘참고문헌’으로 사용하시는 경우 다음과 같이 넣으시기 바랍니다.

동국대학교 서베이앤헬스폴리시리서치센터(2023). AISUM Q5: 가구조사 시 가구 내에서 1명을 왜 랜덤하게 추출해야 하는가? pp1-4. Retrieved from <https://shprc.dongguk.edu/article/aisum/list>