

한국노동패널조사 98표본 원가구의 표본이탈 유형화 - 사건계열분석기법을 이용한 응답이력 분석 -

이혜정 | 한국노동연구원 전문위원(snp625@kli.re.kr)

KLI PANEL *Brief*

- 패널조사가 장기화되면 될수록 표본 유지는 중요하고, 이는 곧 표본이탈을 최소화하는 것을 의미함
- 본 연구는 98표본 원가구의 1~19차 응답이력을 사건계열분석기법으로 분석한 다음 이를 집단별로 유형화하고, 1차 조사 구축 당시 각 집단별 특성에 대해 살펴봄
- 집단은 잠재적 이탈 집단, 간헐적 이탈 집단, 일시적 이탈 집단, 확정적 이탈 집단, 계속 응답 집단으로 총5개로 구분함
- 최초 표본이탈 전 응답기간이 짧을수록, 그리고 최초 표본이탈 후 진입소요기간이 길어질수록 표본에서 이탈할 가능성이 높은 편임
- 가구주의 연령대가 낮을수록, 가구주의 학력이 높을수록 표본이탈의 가능성이 높은 편임
- 거주 지역이 서울이고, 주거 형태가 전세/월세/기타인 경우에 표본이탈의 가능성이 높은 편임
- 조사가 후기(9~10월)에 완료될수록, 가구 방문 횟수가 많을수록 표본이탈의 가능성이 높은 편임
- 패널 표본이탈률의 최소화를 위해 집단별 특성을 고려하여 패널 관리의 적용 범위를 달리 설정할 필요가 있음
- 패널 구축 시 대체 표본을 적절하게 허용하는 범위를 제고해 볼 필요가 있음
- 최초 이탈 후 조기 재복귀 노력은 이탈 방지 노력 못지않게 중요함
- 실사 단계에서 표본이탈에 대한 관리의 일시적 이탈과 간헐적 이탈 유형을 세분화해서 이루어질 필요가 있음
- 일시적 이탈 유형은 면접원과 응답자간의 유대관계가 중요하고, 간헐적 이탈 유형은 빈번한 주거이동이 있으므로 이사했을 때 계속 추적이 가능하도록 해야 함

1. 분석 목적 및 자료

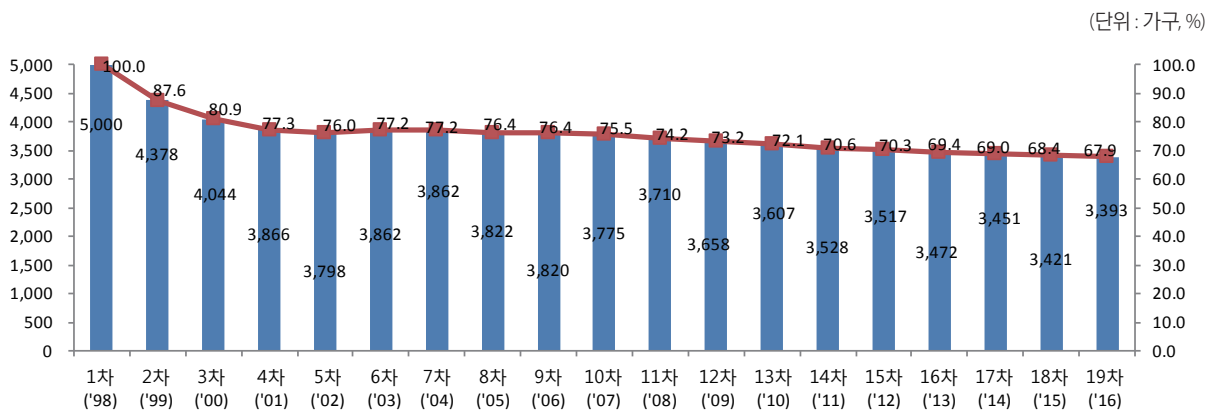
- 1998년부터 시작한 한국노동패널조사(KLIPS : Korean Labor and Income Panel Study)는 현재 20차 조사를 진행 중에 있으며 국내 패널조사에서 가장 오래된 조사인 만큼 지속적인 표본 유지에 대한 관심이 높음
 - 표본 유지는 곧 표본이 이탈하지 않도록 하는 것으로 볼 수 있음
 - 표본이탈이 무작위로 발생하는 것이라면 자료의 신뢰성에 문제가 없지만, 만약 그렇지 않다면 표본 대표성이 손실될 수 있을 뿐만 아니라 자료의 불균형도 야기하여 통계분석결과에 편의(bias)를 가져올 수 있음
- 본 연구는 98표본 원가구의 1~19차 응답이력을 가지고 응답형태가 유사한 것끼리 묶어 몇 개의 집단으로 유형화한 다음, 각 집단별 특성을 살펴봄
 - 분석 대상은 1차(1998년)에 구축한 5,000가구임

- 본 분석에서는 해당 조사차수에서 응답하지 않은 경우를 이탈로 보며, 그 이후 조사차수에서 응답을 다시 한 경우는 재복귀(재진입)한 것으로 간주함
- 한국노동패널 1~19차(1998~2016년 조사) 학술대회용 자료 사용

2. 98표본 원가구의 원표본유지율

- [그림 1]을 보면 원표본유지율은 1차에 구축한 다음 연도인 2차에서 87.6%로 크게 하락한 후 조사차수가 지남에 따라 조금씩 감소하고 있으나 안정적인 추세를 보임
- 2016년에 조사한 19차 조사는 67.9%로 나타남

[그림 1] 98표본 원가구의 원표본유지율



자료 : KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

3. 1~19차 응답이력의 형태 및 유형화

- 1~19차 응답이력은 총 862가지 형태이며, 가장 주된 형태는 1~19차를 모두 응답한 경우로 43.4%를 차지하고, 1차만 응답하고 나머지 차수는 모두 응답하지 않은 형태가 167가구(3.3%)로 다음으로 많음
- 응답이력은 어느 차수에서 무응답을 한 이후로 다시 응답하지 않거나, 응답과 무응답을 반복하거나, 조사가 진행한지 오래된 만큼 가구가 소멸¹⁾ 하는 등의 다양한 형태를 보이고 있음
- 1~19차의 응답이력을 최적일치법(OM : Optimal Matching)으로 거리를 계산하여 매트릭스로 구성(사건계열분석 : Event Sequence Analysis)한 후, 이를 가지고 집락분석을 실시하여 5개의 집단²⁾으로 구분함
- 사건계열분석을 할 때 1~19차를 모두 응답한 2,169가구는 제외하고, 나머지 2,831가구를 가지고 분석함
- 집단은 크게 단조 이탈(monotone attrition)과 비단조 이탈(nonmonotone attrition) 형태로 구분됨
- 단조 이탈은 다시 잠재적 단조 이탈과 확정적 단조 이탈로, 비단조 이탈은 간헐적 비단조 이탈과 일시적 비단조 이탈로 볼 수 있음

1) 가구의 원가구원이 사망, 합가, 이민 등의 이유로 가구가 더 이상 존재하지 않는 경우를 의미함

2) 집단 수의 결정은 Calinski/Harabasz pseudo-F가 가장 큰 값을 가지는 것으로 선정함

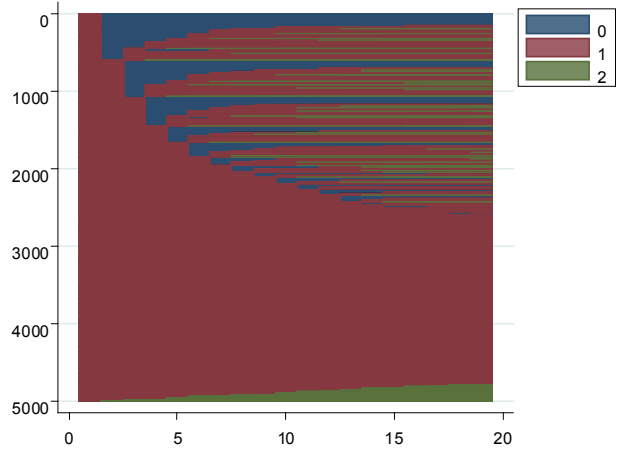
〈표 1〉 1~19차 응답이력 형태

(단위: 가구)

응답이력	빈도
10000000000000000000	167
11000000000000000000	92
11100000000000000000	86
11011111111111111111	68
11110000000000000000	57
10111111111111111111	52
⋮	⋮
11111110111111111111	22
11111111111111000000	22
⋮	⋮
11111222222222222222	19
11112222222222222222	14
⋮	⋮
11100001111111111111	4
11111111111111111111	2,169
전체	5,000

자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

〔그림 2〕 1~19차 응답이력 그래프

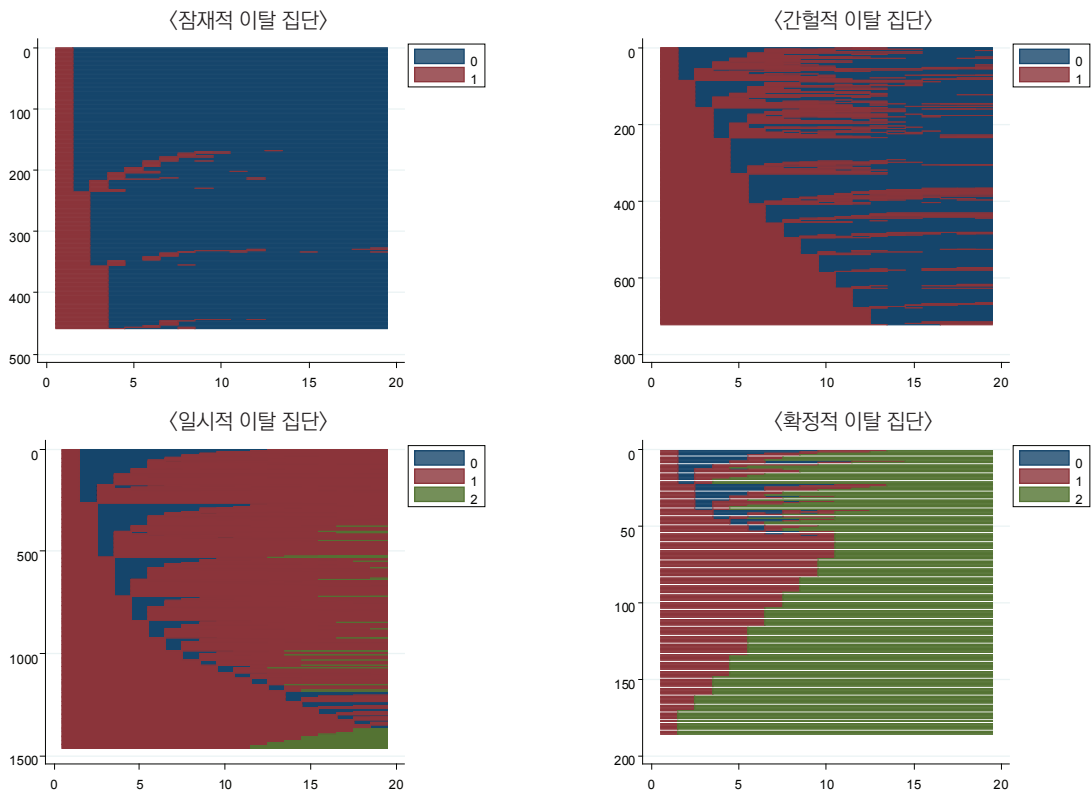


주: 1) 0 해당 차수에서 무응답, 1 해당 차수에서 응답, 2 해당 차수에서 소멸
2) X축은 조사차수, Y축은 가구를 의미함

자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

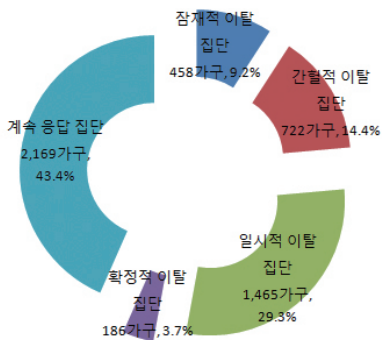
- 각 집단의 명칭은 (1) 잠재적 이탈 집단 (2) 간헐적 이탈 집단 (3) 일시적 이탈 집단 (4) 확정적 이탈 집단 (5) 계속 응답 집단으로 부여함(그림 3 참고). (1)~(4)는 사건계열분석을 통해 구분된 집단이고, (5)는 1~19차를 모두 응답한 집단임
 - 잠재적 이탈 집단: 3차 이전까지 응답하였으나 다시 복귀하는 경우는 거의 없음. 초기 이탈이나 소멸되지 않는 한 항상 재복귀 가능성은 존재함
 - 간헐적 이탈 집단: 꾸준히 응답하다가 최초 이탈 후 진입과 재진입을 반복하는 형태이거나, 아니면 최초 이탈 후 계속 무응답 형태(재진입 없음)를 나타내고 있음
 - 일시적 이탈 집단: 전반적으로 응답이력이 높은 편이고, 잠시 무응답 구간(이탈)이 존재하나 재진입 후에는 지속적인 응답 형태를 보임
 - 확정적 이탈 집단: 조사 중반 전 까지 지속적인 응답이력을 보였으나 최종적으로는 소멸 집단임
 - 계속 응답 집단: 1~19차를 모두 응답한 집단임
- 계속 응답 집단은 43.4%로 집단 중에서 가장 많으며, 다음으로 일시적 이탈 집단이 29.3%를 차지하고 있음(그림 4 참고)
 - 확정적 이탈 집단은 3.7%로 가장 비율이 낮은 집단임
- 간헐적 이탈 집단의 평균 응답기간은 8.1년으로 일시적 이탈 집단(15.7년)에 비해 절반 정도 짧음(그림 5 참고)
 - 확정적 이탈 집단은 5.3년이고 잠재적 이탈 집단은 2년인 것으로 나타남

[그림 3] 집단별 사건연쇄 그래프



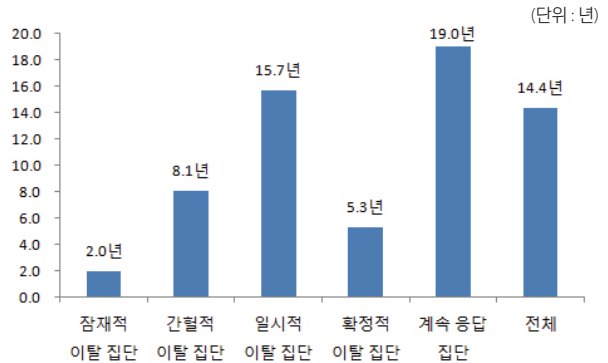
주: 1) 0 해당 차수에서 무응답, 1 해당 차수에서 응답, 2 해당 차수에서 소멸
 2) X축은 조사차수, Y축은 가구를 의미함
 자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

[그림 4] 집단별 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

[그림 5] 집단별 평균 응답기간



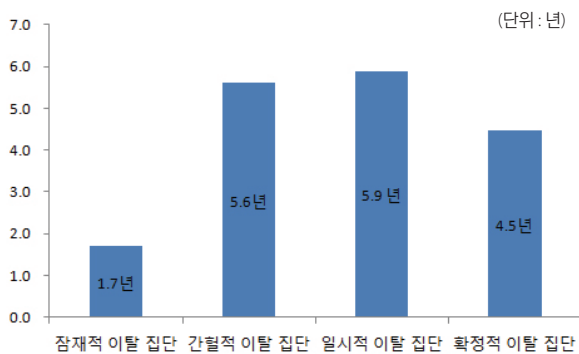
자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

4. 집단별 표본이탈 특징

- 조사 기간 중 응답을 한 번이라도 하지 않는 경우를 보는 것이므로 계속 응답 집단은 제외하고 잠재적 이탈 집단, 간헐적 이탈 집단, 일시적 이탈 집단, 확정적 이탈 집단의 표본이탈 특징을 살펴봄
- 최초 표본이탈 전 응답기간이 길수록 표본에서 이탈할 경향이 낮은 편임

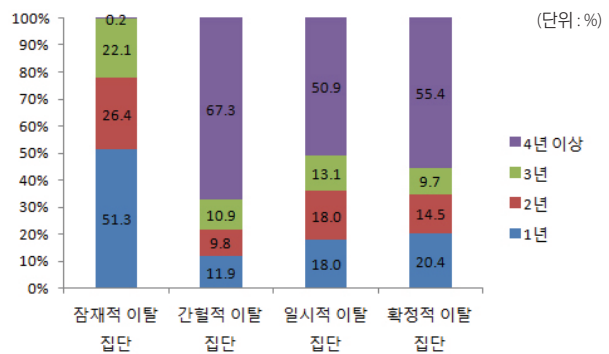
- [그림 6]을 보면 잠재적 이탈 집단(1.7년)의 응답기간이 가장 짧고, 일시적 이탈 집단(5.9년)과 간헐적 이탈 집단(5.6년)의 응답기간이 상대적으로 긴 편에 속함
- 응답기간의 분포를 살펴보면 최초 표본이탈 전 응답기간이 1년인 경우가 잠재적 이탈 집단은 51.3%로 가장 높으며, 이에 비해 간헐적 이탈 집단과 일시적 이탈 집단은 모두 4년 이상인 경우가 67.3%와 50.9%로 상반된 결과를 보여줌(그림 7 참고)

[그림 6] 최초 표본이탈 전 평균 응답기간



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

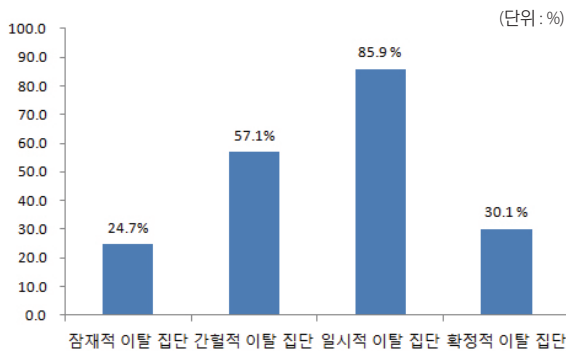
[그림 7] 최초 표본이탈 전 응답기간의 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

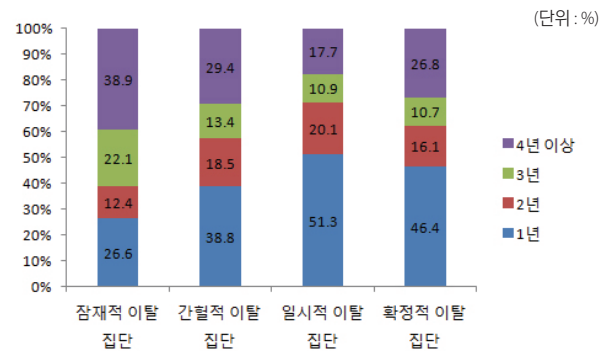
- 최초 표본이탈 후 다시 진입하는데 걸리는 기간이 길어질수록 표본이탈의 가능성이 높음
 - 일시적 이탈 집단의 경우 다른 집단에 비해 가장 짧게 걸리고, 한 번 이탈 후 진입한 가구의 비율 역시 가장 높음
 - 이에 반해 잠재적 이탈 집단은 정반대로 나타남
 - [그림 8]을 보면 최초 표본이탈 후 복귀하는 집단은 잠재적 이탈 집단(24.7%)과 확정적 이탈 집단(30.1%)이 매우 낮은 편에 속하고, 일시적 이탈 집단이 85.9%로 가장 높음
 - 재진입하는데 걸리는 기간의 분포를 보면 일시적 이탈은 재진입까지 기간이 1년인 경우가 절반 이상(51.3%)이고 4년 이상이 17.7%로 가장 낮은 비율인 반면에, 간헐적 이탈 집단은 1년이 38.8%이고 4년 이상이 29.4%로 나타남
 - 잠재적 이탈 집단은 4년 이상이 38.9%로 집단 중에서 가장 비율이 높음(그림 9 참고)
 - 처음 표본이탈이 발생했을 때 최대한 빨리 재복귀할 수 있도록 하는 것이 표본유지에 도움이 됨

[그림 8] 최초 표본이탈 후 재진입 비율



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

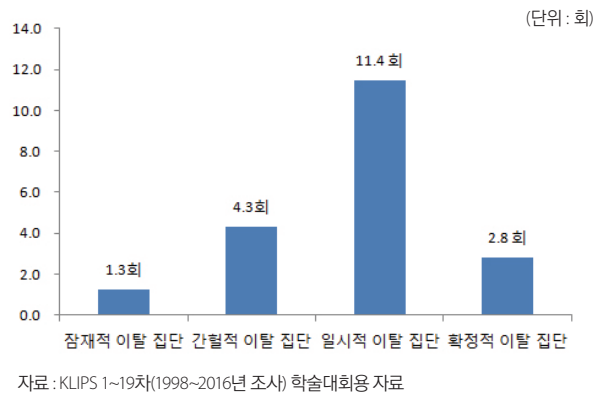
[그림 9] 재진입하는데 걸리는 기간의 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

- 표본에 재진입한 이후의 응답횟수는 잠재적 이탈 집단이 가장 적고 일시적 이탈 집단이 가장 많은 편임
 - 집단 중에서 잠재적 이탈 집단이 1.3회로 응답횟수가 가장 적고, 확정적 이탈 집단이 2.8회, 간헐적 이탈 집단이 4.3회, 그리고 일시적 이탈 집단이 11.4회 순으로 나타남(그림 10 참고)

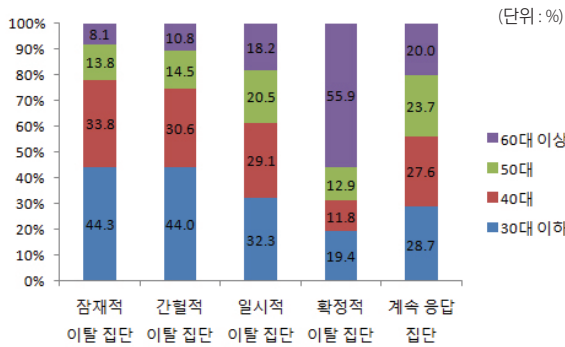
[그림 10] 재진입한 이후 응답횟수



5. 집단별 인구학적 및 실사 관련 특성

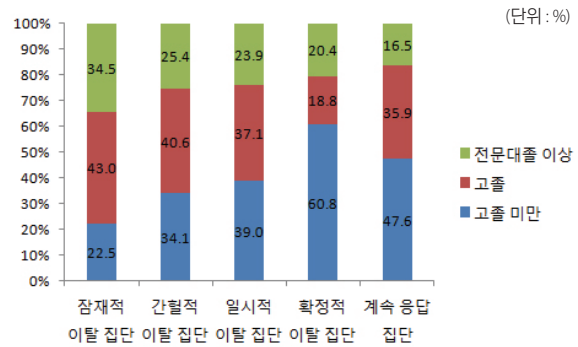
- 1차 구축 당시의 인구학적 정보를 가지고 집단별 특성을 살펴보았으며, 인구학적 정보로는 가구주의 연령대, 가구주의 학력, 거주지역과 주거형태임
- 가구주의 연령대가 낮을수록 표본이탈의 가능성이 높은 편임(그림 11 참고)
 - 30대 이하의 비율은 잠재적 이탈 집단(44.3%) 및 간헐적 이탈 집단(44%)로 다른 집단에 비해 가장 높음
 - 확정적 이탈 집단은 60대 이상의 비율이 55.9%로 가장 높으며, 이는 조사차수의 장기화로 가구주가 고령 또는 노환 등의 이유로 사망하여 소멸가구가 된 것에 영향을 미침
 - 계속 응답 집단은 가구주의 연령대가 고르게 분포되어 있음
- 가구주의 학력이 높을수록 표본이탈의 가능성이 높은 편임(그림 12 참고)
 - 계속 응답 집단은 가구주의 학력이 고졸 미만인 경우가 47.6%로 가장 높음
 - 가구주의 학력이 전문대졸 이상인 경우는 잠재적 이탈 집단(34.5%), 간헐적 이탈 집단(25.4%), 그리고 일시적 이탈 집단(23.9%)로 계속 응답 집단(16.5%)에 비해 높은 편임
- 거주 지역이 서울인 경우 표본이탈의 가능성이 높은 편임(그림 13 참고)
 - 잠재적 이탈 집단, 간헐적 이탈 집단과 일시적 이탈 집단은 서울의 비율이 다른 지역에 비해 가장 높음
 - 계속 응답 집단은 광역시(32.4%)와 그 외(30.2%)의 비율이 높음
- 주거 형태가 자가인 경우 표본이탈의 가능성이 낮은 편임(그림 14 참고)
 - 계속 응답 집단은 자가인 경우가 63.1%로 가장 높은 반면에, 잠재적 이탈 집단과 간헐적 이탈 집단은 절반 이상의

[그림 11] 집단별 가구주의 연령대 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

[그림 12] 집단별 가구주의 학력 분포



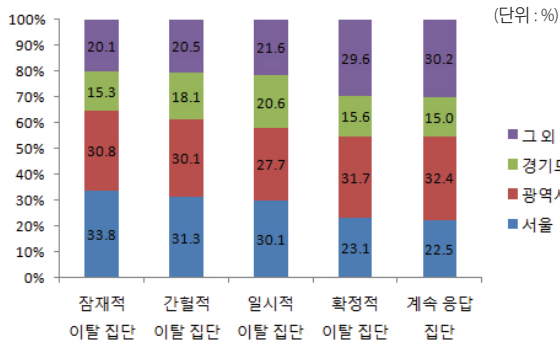
주: 학력을 응답하지 않은 1명 제외 후 분석

자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

비율이 전세/월세/기타인 것으로 나타남

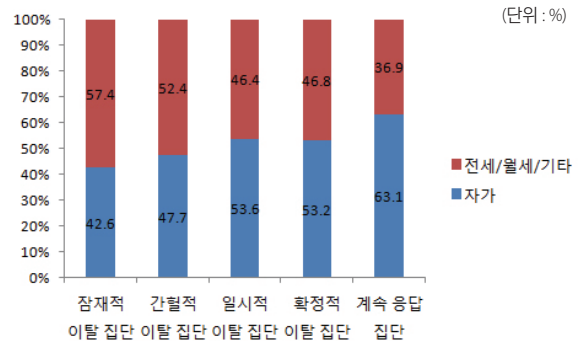
- 이는 잠재적 이탈 집단과 간헐적 이탈 집단은 주거 형태가 불안정하여 빈번한 이사가 표본이탈과 관련이 있을 수 있음
- 일시적 이탈 집단은 자가가 53.6%로 전세/월세/기타(46.4%)보다 7.2%p 높음

[그림 13] 집단별 거주 지역 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

[그림 14] 집단별 주거 형태 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

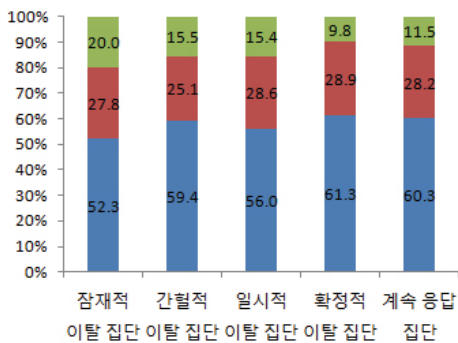
- 실사 관련 정보인 조사 시기 및 가구 방문 횟수를 집단별로 분석하였으며, 이 때 사용한 자료 또한 1차 조사임
- 조사가 후기(9~10월)에 완료될수록 표본이탈의 가능성이 높은 편임(그림 15 참고)
 - 모든 집단에서 절반 이상이 대부분 조사 초기에 조사가 완료됨
 - 특히 계속 응답 집단은 60.3%로 아주 높은 편임
 - 조사가 후기(9~10월)에 완료된 비율은 잠재적 이탈 집단이 20%로 가장 높고, 간헐적 이탈 집단과 일시적 이탈 집단이 각각 15.5%와 15.4%이며, 계속 응답 집단은 11.5% 순으로 낮음
- 가구 방문 횟수가 많을수록 표본이탈의 가능성이 높은 편임(그림 16 참고)
 - 1회 방문하여 조사를 완료한 비율이 가장 높은 집단은 확정적 이탈 집단(54.3%)이고 다음으로 높은 집단은 계속 응답 집단으로 40.2%로 나타남

- 3회 이상 방문한 경우는 잠재적 이탈 집단이 47.8%로 가장 높고, 일시적 이탈 집단(38.1%)과 간헐적 이탈 집단(35.5%)도 계속 응답 집단(32.7%)에 비해 높은 편임

• 1차 조사 때 조사 성공이 상대적으로 어려웠던 가구는 표본을 유지하는데 어려운 편임

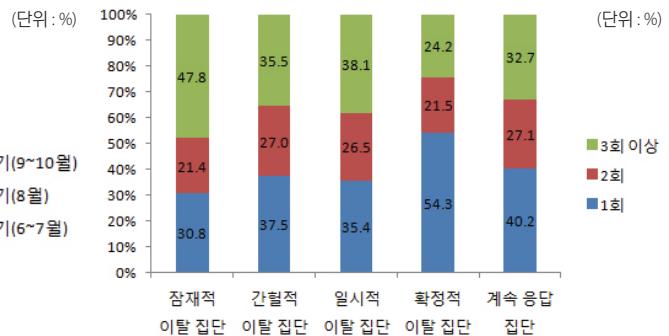
- 모집단의 특성에 따라 추출된 최초 표본을 구축하는 것이 원칙이나, 첫 조사부터 조사 협조도가 좋지 않은 가구를 설득하여 조사를 성공하더라도 조사차수가 길어지면 결국 먼저 이탈할 가능성이 높음
- 표본 유지 측면에서 보면 패널 구축 시 대체 표본을 적절하게 허용하는 범위를 제고해 볼 필요가 있음

[그림 15] 집단별 조사 시기에 따른 분포



주: 조사 날짜(월)가 결측인 482가구 제외 후 분석
 자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

[그림 16] 집단별 가구 방문 횟수에 따른 분포



자료: KLIPS 1~19차(1998~2016년 조사) 학술대회용 자료

6. 요약 및 시사점

• 패널조사가 장기화되면 될수록 표본 유지는 중요하고, 이는 곧 표본이탈을 최소화하는 것을 의미함

- 표본이탈이 발생하면 표본의 대표성이 손실될 수 있을 뿐만 아니라, 자료의 불균형을 야기하여 통계분석결과에 편의(bias)를 가져올 수 있음
- 조사 초반에 표본이탈이 단조 형태로 가장 많이 발생하나, 패널조사가 오래될수록 비단조 형태로 더 많이 발생하므로 표본이탈의 유형을 고려해야 함
 - 본 연구의 분석 결과 역시 단조 형태(12.9%)보다 비단조 형태(43.7%)의 비율이 더 높게 나타남

• 본 연구는 98표본 원가구의 1~19차 응답이력을 사건계열분석기법으로 분석한 다음 이를 집단별로 유형화하고, 1차 조사 구축 당시 각 집단별 특성에 대해 살펴봄

- 집단은 잠재적 이탈 집단, 간헐적 이탈 집단, 일시적 이탈 집단, 확정적 이탈 집단, 계속 응답 집단으로 총5개로 구분함
- 최초 표본이탈 전 응답기간이 짧을수록, 그리고 최초 표본이탈 후 진입소요기간이 길어질수록 표본에서 이탈할 가능성이 높은 편임
- 가구주의 연령대가 낮을수록, 가구주의 학력이 높을수록 표본이탈의 가능성이 높은 편임
- 거주 지역이 서울이고, 주거 형태가 전세/월세/기타인 경우에 표본이탈의 가능성이 높은 편임
- 조사가 후기(9~10월)에 완료될수록, 가구 방문 횟수가 많을수록 표본이탈의 가능성이 높은 편임

- 패널 표본이탈률의 최소화를 위해 집단별 특성을 고려하여 패널 관리의 적용 범위를 달리 설정할 필요가 있음
 - 패널 구축 시 대체 표본을 적절하게 허용하는 범위를 제고해 볼 필요가 있음
 - 최초 이탈 후 조기 재복귀 노력이 이탈 방지 노력 못지않게 중요함
 - 실사 단계에서 표본이탈에 대한 관리는 일시적 이탈과 간헐적 이탈 유형 등으로 세분화해서 이루어질 필요가 있음
 - 일시적 이탈 유형은 면접원과 응답자간의 유대관계가 중요하고, 간헐적 이탈 유형은 빈번한 주거이동이 있으므로 이사했을 때 계속 추적이 가능하도록 해야 함

 - 표본이탈에 따른 자료의 분석 및 통계적 측면
 - 분석적 측면에서는 표본이탈 통제 방식(균형패널만 사용, 전체 표본 사용, 장기 무응답자 제외 등)에 따라 이질적인 결과가 나타날 수 있기 때문에 주의 필요
 - 통계적 측면에서는 표본이탈에 따른 세분화된 검증이 필요하며, 필요시 가중치 부여와 임putation 등을 통해 이를 보완하여 개선하는 노력 필요
-