

KEEP 패널 유지 실태와 가중치

변 중 석¹⁾

요약

한국교육고용패널(KEEP)은 2004년에 중학교 3학년, 일반고교 및 전문고교 3학년을 각각 코호트(cohort)로 선정하고, 각 코호트별 2000명씩 층화집락추출법(stratified-cluster sampling)으로 추출하여 코호트 패널조사로 최초 실시되었으며, 현재 5차 조사가 진행되고 있다. 현재 국내에는 20여종의 패널을 대상으로 종단연구자료 생성을 위한 조사가 진행되고 있지만, KEEP은 코호트를 대상으로 종단 연구되는 국내의 유일한 사례이다. KEEP의 코호트조사는 동일한 연령그룹을 대상으로 동일한 시점의 경험에 대한 다양한 횡단면 및 종단면 자료를 제공하게 되므로 조사 시점의 횡단면적인 분석 및 패널 변화를 분석하는 종단면적 분석이 가능하게 해 주는 유용한 자료를 제공하게 된다. 즉, 중3 코호트는 2차 연도 이후 고등학교에 재학하는 일반적인 진학 상황을 유지하지만, 고3 코호트는 대학 진학 및 취업/비진학으로 코호트의 상황이 변화하게 된다. 고3 코호트를 보면, 종단적으로는 고3 코호트가 진학과 취업 등으로의 상태 변화에 대한 유용한 자료를 제공하며, 횡단적으로는 해당 연령의 대학생에 대한 유용한 정보 및 취업/비진학에 대한 유용한 정보를 제공하게 된다. 그 결과, KEEP 자료는 동일 연령의 변화 및 조사 시점에서의 경험, 태도 등에 대한 다양한 분석이 가능한 횡단 및 종단적인 정보를 제공하게 된다.

본 연구에서는 3차년도(2006년)까지의 기초 자료 분석 및 조사보고서를 기준으로 최초 선정된 KEEP 패널이 3차 연도까지 진행된 조상 참여 및 탈락 현황을 정리하여 패널 관리 및 유지 실태를 살펴보고, 1차년도의 가중치 및 3차년도의 가중치를 분석하고, 3차년도 가중치를 비교하였으며, 3차년도 자료에 대해 가중치를 적용하여 가중치 효과를 검토해 보았다.

분석에서 나타난 중요한 결과로는 1) 3차년도의 패널유지율을 보면, 최초 패널 중 실제 유효 자료를 제공하는 유효 표본유지율이 79.2%, 남학생 중 군입대로 인한 잠재 표본유지율이 18.4%로 나타나 97.6%의 패널유지율을 보였고, 2) 패널이 3차연도까지 모두 참여한 패널의 완전응답율은 65.3%를 유지한 반면, 2차 이후 조사가 불가능한 표본탈락율은 14명(0.2%)이었다. 3) 중3 코호트를 대상으로 3차년도 가중치를 비교 분석해 본 결과, 일반계고교와 전문계고교의 대표 대표성은 거의 비슷하게 유지되고 있었으나 지역별 가중치 차이가 크게 나타나 향후 지역별 표본 규모를 점검할 필요가 있는 것으로 나타났다. 4) 가중치의 효과에서는 가중치가 모평균 추정에 미치는 영향을 살펴보았다. 특히, 모평균 추정에서는 가중치와 분석변수와의 상관관계에 따라 과소 및 과대 추정되며, 추정량의 분산도 증가될 수 있음을 확인하였다. 특히 사교육 지출비용은 모평균을 과대추정할 가능성이 있었으며, 분산도 증가할 수 있음을 보였다.

1) 한신대학교 정보통계학과 교수

I. 서론

한국교육고용패널(KEEP) 조사는 청소년의 교육 경험과 진학, 진로 및 직업으로의 이동 과정 및 경로 등을 파악하기 위해 2004년도 중3 코호트 2000명, 일반고3 코호트 2000명, 전문고3 코호트 2000명 등 6000명을 표본추출하여 동일한 표본을 매년 추적조사하여 인적 자료의 축적과 활용에 대해 유용한 자료의 구축 및 제공하기 위해 실시되는 조사이다. KEEP에서는 국가 인적자원개발 정책, 국가 교육정책, 진로 및 직업교육의 방안 마련 등의 노동 시장 정책 수립의 기초 자료와 인적자원개발 관련 기관의 연구 자료로 활용하기 위해 개인(학생), 가구(학부모), 학교행정이 및 담임교사를 대상으로 1차년도에 자료를 수집하였지만 2차년도 이후는 개인(학생)을 추적하여 자료를 수집한다.

최근 국내에서 널리 수행되는 패널조사는 동일한 패널을 추적하여 자료를 수집하는 반면, KEEP은 코호트 패널을 대상으로 하므로 일반적으로 널리 수행되는 국내의 다른 패널 조사와는 다르다고 생각된다. 코호트 조사는 동일한 연령대를 표본으로 선정하여 동일한 경험을 자료 수집하는 조사이다. 코호트 패널조사는 코호트조사를 위해 표본으로 선정한 후, 동일한 대상을 추적하여 자료를 수집하는 조사이므로 코호트에 대한 종단적인 자료뿐만 아니라 횡단적인 자료도 제공하게 된다. 즉, 동일한 연령은 대상으로 패널을 유지하기 때문에 코호트패널의 변화뿐만 아니라 조사시점에서의 자료까지 분석 가능하므로 종단 및 횡단의 유용한 정보가 제공하게 되는 것이다.

패널조사는 실제 자료 수집을 위해 패널을 관리하고 유지하면서 조사 활동을 진행하고, 현실의 제약된 여건에서 수집된 자료를 분석하므로 많은 어려움이 나타난다. 특히 패널은 표본 설계 당시의 모집단을 대표하도록 설계되어 있으므로 시간의 흐름에 따라 변화하는 모집단 변동을 반영하는 데 미흡한 현실이다. 특히, KEEP의 고3 코호트는 고교 졸업 후 대학을 진학하거나 취업을 한 후 2-3년에 남학생의 상당 수가 군입대를 하므로 코호트는 유지되지만 자료 수집의 어려움이 존재하는 현실이다. 군입대 패널은 표본은 관리되지만 자료 수집이 불가능하므로 군입대 기간 동안 표본이 훼손되어 패널의 표본 대표성을 심하게 왜곡시키고 있다. 이 외에도 KEEP과 같은 코호트 패널은 동일한 연령의 모집단을 대표하도록 설계되어 있으므로 시간의 흐름에 변화하는 모집단 변동을 어느 정도 커버하고 있지만, 추적시기가 오래될수록 패널의 이탈 및 탈락 등의 마모나 패널 무응답이 증가하는 현상으로 인해 표본의 대표성을 점검할 필요가 있다. 일반적으로 패널조사에서 표본의 대표성은 적절한 가중치의 조정을 통해 해결하고 있다.

KEEP 조사는 동일 연령층의 코호트를 패널 대상으로 선정하여 자료를 수집하므로 시간의 변화에 따라 패널의 신분(status)이 변화하기 때문에 조사 시점에서 패널의 현재 신분(status)을 파악하는 것이 향후 가중치 및 분석에서 중요한 기준을 제공하게 된다. 표본 설계 당시의 신분보다는 조사 시점에서 나타난 신분의 변화는 기존 패널의 유지, 관리 실태를 보여주는 표본 대표성을 나타낼 뿐만 아니라 패널이 시간의 변화에 따라 변화되는 모집단의 변화에 대한 자료로도 활용 가능하다. 즉, 현재 신분에 대한 가중치는 시간의 변화를

반영하는 모집단의 가중치로 사용해 1차년도 가중치와 비교해 보면 변동 모집단의 특성이 반영된 가중치가 되게 된다.

따라서 본 연구에서는 먼저, 3차년도의 패널을 대상으로 하여 KEEP 패널의 표본 유지 및 응답 패턴을 살펴보고, 1차년도와 3차년도의 가중치를 분석하여 모집단 변화를 가중치로 살펴볼 것이며, 분석 관점에서는 가중치가 모평균 추정 및 상관계수 추정에 미치는 효과를 살펴보고자 한다.

II. 패널의 유지 실태 및 응답 패턴

패널조사에서는 시간이 경과하면서 패널의 이탈이나 추적 실패 등으로 인해 조사가 불가능한 표본의 마모가 발생하게 된다. 패널조사에서는 조사 불가능하거나 조사에 참여하지 않은 무응답 표본의 특성을 살펴보는 것도 표본의 유지 및 관리에 필요한 정보를 제공하게 된다.

KEEP은 2004년 중3, 일반계 고3, 전문계 고3 등 3개 코호트의 표본을 선정하여 동일한 대상을 매년 추적하여 국가인적자원개발 정책, 국가 교육정책 및 노동시장 정책 수립을 위한 기초 자료 제공, 그리고 국가 인적자원개발 정책 수립이나 관련 분야의 연구 자료로 활용 등을 목적으로 수행되는 조사이다. KEEP은 1차년도 조사에서는 개인(학생), 가구(학부모), 학교행정가 및 담임교사를 대상으로 자료를 수집하였지만 2차년도 이후는 개인(학생)을 추적하여 자료를 수집한다. 본 연구에서는 계속 추적조사의 대상이 되는 학생을 분석의 관심 대상으로 하여 한다. 본 절에서는 3차 까지 진행된 KEEP 패널의 유지 실태 및 응답 패턴을 정리해 봄으로써 KEEP 패널의 표본 유지 현황을 살펴보고자 한다.

1. 패널 유지 실태

KEEP 패널은 동일 연령층이 같은 시기에 겪는 교육, 진로 및 취업, 생활 등 동일한 시대의 경험을 자료 수집하는 코호트 패널조사이다. KEEP은 코호트별 층화집락추출법으로 지역별로 학교 및 학급을 추출하고, 표본 학급 내에서 학생을 패널로 추출하여 각 코호트별로 2000명씩 6000명을 대상으로 2004년도에 처음으로 실시되어 2008년도에 5차까지 조사가 완료되었다. 현재 수집된 패널 조사 자료는 4차까지 제공되고 있으나 가중치자료는 3차까지 완료된 상황이다.

〈표 1〉 KEEP 조사의 패널 유지 현황

패널 유지 실패		1차년도 (2004)		2차년도 (2005)		3차년도 (2006)				
유효 표본 (명)	조사 성공	중학생	6000	6000	5817	5256	4752	-		
		일반고						2000	1287	1273
		전문고						2000	477	467
		대학생						-	2456	1791
		취업/비진학						-	1033	*557
조사실패		-	-	561	-	664				
조사 불능	군입대(명)	-	-	183	88	1248	1106			
	기타(명)	-	-	-	95	-	142			
표본 유지	유효표본유지율(%)	100	100	98.4	97.0	97.6	79.2			
	군입대율(%)	-	-	-	1.4	-	18.4			
조사 성공율(%)	전체패널기준	100		87.6		68.1				
	유효표본기준	100		90.4		86.0				

주) 1) 조사불능 : '기타' 는 유학, 가출, 해외연수, 이민, 원양어선, 보호소, 사망 등

2) 유효표본기준의 조사성공율 = (조사성공 표본) / (유효표본) X100

*) 조사성공 : '취업/비진학' 은 중3 코호트 취업/비진학자 12명 포함

〈표 1〉에서 보듯이 2차년도 코호트별 학생 신분의 변화를 보면, 중3 코호트는 일반계고교나 전문계고교로 진학하는 비율이 88.2%(1764명, 일반고 1287명, 전문고 477명), 일반계 및 전문계의 고3 코호트는 대학 진학이 61.4%(2456명), 취업 및 비진학(재수)이 25.9%(1033명)로 87.6%가 신분의 변화가 파악되었다. 유효표본으로 분류되었지만 일시적으로 조사 참여를 거절하거나 조사기간 동안 패널 추적에 실패한 조사실패 패널은 전체의 9.4%(561명)이었고, 이를 유효표본으로 포함해 살펴 본 2차년도 전체 패널의 유효표본유지율은 97.0%나 되었다. 반면, 고3 코호트 중에서 유학, 가출, 해외연수, 이민 및 군입대 등의 사유로 현실적으로 조사가 불가능하여 조사 불능으로 분류된 패널이 3.1%(183명)이었다. 군 입대자는 군 제대 후 다시 패널로 자료를 수집할 수 있으므로 표본으로 유지 가능하므로 KEEP 2차년도의 표본 유지율은 98.4%이었다. 2차년도의 전체패널기준 조사 성공률은 87.6%, 조사실패를 포함한 유효 표본기준 조사성공율은 90.4%이었다.

3차년도 코호트별 학생 신분의 변화를 보면, 중3 코호트는 고교 재학 비율이 87.0%(1740명), 일반계 및 전문계의 고3 코호트는 대학 재학이 44.8%(1791명), 취업 및 비진학(재수)이 13.6%(545명, 12명은 중3 코호트)로 68.1%가 신분이 파악되었다. 유효표본으로 분류되었지만 조사를 일시적으로 거절하거나 조사기간 동안 패널 추적에 실패한 패널은 전체의 11.1%(664명)이었고, 이를 유효표본으로 포함해 살펴 본 3차년도 전체 패널의 유효표본유지율은 79.2%이었다. 반면, 유학, 가출, 해외연수, 이민 등과 군입대로 인해 조사 불능으로 분류된 패널이 20.8%(1248명)나 되었으나, 군입대가 18.4%(1106명)로 많은 반면, 조사 자체

가 현실적은 불가능한 비율은 2.4%에 불과했다. 군 입대자는 군 제대 후 다시 패널로 자료를 수집할 수 있으므로 표본으로 유지 가능하므로 KEEP 2차년도의 표본 유지율은 97.6%이었다. 2차년도의 전체패널기준 조사 성공률은 68.1%, 조사실패를 포함한 유효 표본기준 조사성공율은 86.0%이었다.

1-3차년도 사이에 나타난 패널 유지 및 관리 실태를 보면, 1) 유효표본 자료 크기는 전체 패널기준으로 보면, 2005년 97.0%, 2006년 79.2%로 나타나 해마다 감소하고 있으며, 2) 조사를 일시적으로 거절하거나 조사기간 동안 패널 추적에 실패한 조사 실패 패널 규모는 2005년 9.4%, 2006년 11.1%로 나타나 해마다 약 10% 정도 발생하였고, 3) 유학, 가출, 해외연수, 이민 및 군입대 등으로 인한 조사 불능으로 분류되는 비율은 2005년에 3.0%, 2006년에 20.8%로 2006년(3차년도)에 급격히 증가하였는데, 특히 고3 코호트에서 2006년 이후 군입대 표본이 매우 많았고, 4) 조사성공율은 유효표본 기준 2005년에 90.4%, 2006년에 86.0%로 해마다 85%이상 조사자료 수집에 성공하였고, 5) 잠재 패널인 군입대 표본을 포함하는 전체 패널의 표본 유지비율은 2005년 98.4%, 2006년 97.6%로 매우 높았다.

<표 2> 국내외 패널조사의 패널 유지 현황

패널명	KLIPS (한국/1998)	복지패널 (한국/2005)	대우패널 (한국/1993)	PSID (미국/1968)	BHPS (영국/1991)	GSOEP (독일/1984)
2차년도	87.6	92.1	79.0	89.0	87.7	89.9
3차년도	80.9	86.7	66.0	86.3	81.5	86.0

(주: 한국복지패널 가중치 관련 회의 자료 인용, 2009. 2. 10)

참고로 국내외 주요 패널조사의 표본 유지율에 대한 자료는 <표 2>와 같다. 표본 유지율을 국내외 패널조사 결과와 비교해 보면, KEEP의 유효표본유지율은 2차년도에 97.0%, 3차년도에 79.2%로 나타나 국내외 패널에 비해 2차년도의 표본 유지 규모가 가장 높은 반면, 3차년도는 국내외 패널의 표본 유지율과 비슷한 수준으로 나타났다. 그러나 KEEP은 3차년도에 군입대자를 조사 불능으로 분류하여 계산했으므로 이를 감안한다면 3차년도의 표본 유지 규모도 국내외 타 패널조사에 비해 높은 수준으로 표본을 유지하고 있다고 할 수 있다.

KEEP 패널의 표본 관리에서 나타난 주요한 특징으로는 1) 동일 연령층의 코호트 패널로 표본을 관리하며, 2) 코호트 패널로 동일한 대상을 매년 추적조사하며, 3) 전반적으로 패널의 이탈 비율이 적어 표본유지율이 95%이상으로 나타나 타 패널조사에 비해 높게 나타나고 있으며, 4) 유효표본 기준 조사 성공률이 86%이상으로 비교적 높으며, 5) 남학생의 경우 20세-24세 정도까지 군입대로 인해 자료 수집이 불가능한 상황이 존재하여 패널 초기에 코호트별로 유효 표본크기가 작게 나타난다는 점 등이다.

2. 패널 응답 패턴

패널조사에서는 설계 당시 선정된 표본이 대표성을 갖고 있지만 계속된 조사에서 표본의 이탈 등 표본 마모와 무응답이 증가하므로 표본 대표성의 유지가 어렵게 된다. 따라서 패널조사에서는 표본 마모 및 무응답 발생에 따른 조정이 필요하게 되므로 무응답 패턴을 파악하는 것도 필요하다.

KEEP은 아직 패널 시작 초기라 무응답 패턴을 발견하기가 쉽지 않기 때문에 3차년도까지의 패널의 응답 패턴을 살펴봄으로써 향후 계속된 조사에서의 무응답 패턴 파악을 위한 기초자료로 검토해 보고자 한다.

<표 3> KEEP 패널의 응답 패턴

응답 패턴		1차	2차	3차	유효 표본	(%)
완전응답		○	○	○	3919	65.3
부분 응답	조사 불능	○	X	○	169	2.8
		○	○	X	1234	20.6
	조사 실패	○	-	-*	561	9.4
		○	○	-	103	1.7
소멸패턴		○	X	X	14	0.2
응답 패널 규모		6000	5256	4088	6000	100.0

주) ○ : 응답, X : 무응답

* : 조사 시점별 조사실패 패널을 표현한 것으로 두 기간 조사 실패를 의미하는 것은 아님.

<표 3>은 KEEP의 3차년도까지의 조사년도마다 패널에 대한 응답여부를 확인하여 응답 패턴을 정리해 보았다. KEEP은 설계 당시 중3 및 고3 학생들을 추출하였기에 패널의 이탈이나 신분 변화 등으로 인한 표본 탈락이 높지 않으며, 또한 추출된 학생들을 매년 추적하여 조사를 하므로 패널의 조사 참여 및 응답을 높일 수 있다. 실제로 중3 코호트는 3차년도에 대부분 고교에 재학 중이며, 고3 코호트는 많은 패널이 대학에 재학하거나 취업, 재수를 하므로 학생 패널의 추적이 비교적 용이하다. 하지만 KEEP 패널 중 3차년도에 일시적으로 군입대, 해외유학 및 연수, 원양어선 등 일시적으로 조사가 불가능한 규모가 높게 나타나지만 이는 일시적인 무응답 패널일 뿐 표본 이탈 혹은 탈락 등의 마모 패널이 아님을 유념할 필요가 있고, 또한 조사 참여를 일시적으로 거절하거나 조사기간 동안 자료수집에 실패한 조사실패 패널도 향후 추적을 통해 자료를 수집할 수 있어 부분응답 패턴으로 분류하였다. 그러나 조사불능 패널 중 조사 참여를 강력하게 거절하거나 사망 및 추적 완전 실패 등의 사유로 자료 수집이 실제로 완전하게 불가능한 패널을 소멸패턴으로 분류하였다.

3차년도까지 진행된 KEEP의 응답 패턴 특성을 보면, 1) 3차년도까지 조사에 성공한 완전 응답한 패널은 65.3%(3919명)로 나타났다. 참고로 KLIPS는 3차년도까지의 완전응답은

최초 표본 규모 기준 77.9%이었다(김영원의, 2005). 특히, 완전 응답 패턴을 코호트별로 살펴보면, 중3 코호트는 83.6%(1671명), 일반고3 코호트는 61.3%(1225명), 전문고3 코호트는 51.6%(1023명)이었다. 이러한 현상은 고3 코호트가 군입대로 인해 낮은 것이다. 2) 부분응답 패턴을 보면, 유학, 가출, 해외연수, 이민 및 군입대 등으로 인한 조사가 불가능한 패널의 부분응답 패턴이 23.4%이었고, 조사 참여를 일시적으로 거절하거나 조사기간 동안 자료수집에 실패한 조사실패 패널의 부분응답 패턴은 11.1%로 나타나 34.5%가 부분응답 패턴으로 나타났다. 3) 조사 참여를 강력하게 거절하거나 사망 및 추적 완전 실패 등의 사유로 표본 탈락이 예상되는 응답 패턴인 소멸 패턴은 0.2%로 매우 작게 나타났다. 유효 표본 기준 응답율은 2차년도에 90.4%, 3차년도에 86.0%로 3차년도에 약간 감소하였다.

패널조사에서는 조사가 진행될수록 표본 패널이 안정화되기 전까지 표본의 무응답이나 조사 참여 거절이 높게 나타날 수 있으며, 또한 KEEP은 남학생의 군복무 기간 일시적인 조사 참여가 불가능하여 표본이 일시적으로 훼손되므로 표본의 대표성 유지 및 확보를 위해 표본을 효율적으로 관리하는 방안 마련이 요구되며, 또한 조사 참여 패널을 독려하여 무응답 감소를 줄여 대표성을 유지해야 하며, 패널조사가 성공적으로 수행되기 위해 조사시점간 무응답율의 변화를 검토할 필요가 있다.

III. KEEP 가중치

패널조사에서 가중치는 확률추출, 무응답 및 사후층화 등으로 인한 불균등 추출확률을 보정하기 위해 널리 사용된다. 표본은 모집단의 크기나 분석 목적 상 표본 추출 과정에서 표본단위마다 서로 다른 추출 확률을 가지고 표본으로 추출하게 된다. KEEP은 중학교, 일반계고교, 전문계고교의 학교급을 코호트로 하여, 각 코호트마다 기본적으로 층화집락추출법(stratified-cluster sampling)으로 표본을 추출하였다. 층화지표로는 지역 및 학교 유형(1차)을 사용하였으며, 학생 수 비율 등급에 따라 표본 추출 학교 및 학급 수를 선정하여 패널 학생을 선정하였다. 따라서 코호트마다, 각 지역별 학교 및 학급 규모별, 학생의 성별에 따라 서로 다른 추출 확률로 패널이 선정되었다.

본 절에서는 먼저, KEEP의 가중치 부여 방법을 간단히 정리한 후, 각 코호트별로 1차년도 가중치와 3차년도 가중치를 계산하여 KEEP 가중치를 살펴보고, 3차년도 가중치를 이용하여 가중치의 효과를 검토하고자 한다.

1. KEEP의 가중치 부여 방법

일반적으로 가중치는 표본설계단계에서의 불균등 추출확률 보정, 자료수집 결과를 반영

하는 무응답 보정 및 추정 후 모집단과의 차이를 보정하는 사후층화 보정 등의 3단계로 구분하여 가중치를 부여한다. 또한 패널조사와 같은 종단연구에서는 횡단면 가중치뿐만 아니라 종단면 가중치를 동시에 고려하며, 1차년도에는 횡단면 가중치만을 사용하지만 2차년도 이후에는 횡단면 가중치 이외에 종단면 가중치를 고려한다.

KEEP에서 부여하는 가중치를 보면, 1) 1단계 가중치로 학교, 학급 및 학생의 표본 추출 확률을 보정하기 위한 불균등 추출확률에 대한 설계 가중치를 부여하고 있으며, 2) 2단계 가중치로는 무응답을 보정하기 위한 무응답 가중치를 부여하고 있으며, 3) 3단계 가중치로는 2단계까지 부여한 가중치의 총합이 모집단과 일치하도록 보정하기 위한 사후층화 보정 가중치로 구분하여 부여하고 있다. KEEP의 사후 층화 보정을 위한 자료로 교육통계연보의 지역, 성별, 학교 유형 및 진학 상황을 고려하고 있다.

2. 1차년도 가중치 : 횡단면 가중치

1차년도 가중치는 KEEP 설계 시점의 가중치로 설계 당시 표본의 대표성이 유지되는 가중치를 의미하며, 불균등 추출확률을 나타내는 설계 가중치와 모집단과의 차이를 보정하는 사후층화 보정 가중치를 결합하여 최종 가중치를 계산된다. 1차년도에는 패널에 참여하는 모든 학생의 자료가 수집되므로 무응답이 발생하지 않기 때문에 무응답 조정 가중치는 모두 “1”로 동일하게 부여한다.

KEEP에서 학생의 설계(혹은 불균등 추출확률 보정) 가중치는 학생의 추출 확률의 역수로 계산하게 되는데, 학생 추출확률은 지역별(중3 및 일반계 고3 코호트) 또는 학교유형별(전문계 고3 코호트)로 학교의 추출확률, 추출된 표본학교에서의 학급 추출확률, 추출된 표본학급에서의 학생 추출확률의 곱으로 계산한다. 학생들의 설계 가중치 총합은 중3 및 일반계 고3 코호트는 (지역별 X 성별)로, 전문계 고3 코호트는 (학교유형별 X 성별)로 계산한다. 이 때 사후층화 보정 가중치는 지역별 또는 학교유형별 교육통계연보의 학생 수와 가중치 총합의 상대적인 비율로 계산한다. 따라서 1차년도 KEEP의 가중치는 학생의 설계 가중치와 사후층화 보정 가중치를 곱하여 계산한다.

KEEP의 1차년도 가중치 계산에 사용되는 학생의 표본 추출 확률, 설계 가중치, 가중치 총합 및 사후층화 보정 가중치는 코호트별로 다음과 같이 계산한다. 중3 코호트와 일반계 고3 코호트는 지역별 모집단 특성을 반영하기에 동일한 과정으로 계산되지만, 전문계 고3 코호트는 학교유형별 모집단 특성으로 반영하여 계산된다.

- 중3 및 일반계 고3 코호트 : 학생의 표본 추출확률과 설계 및 사후층화 보정 가중치

$$\begin{aligned} \text{학생의 표본추출확률} &= (\text{지역별 표본 학교 추출확률}) \\ &\quad \times (\text{표본 학급 추출확률} | \text{추출된 표본학교}) \\ &\quad \times (\text{표본 학생 추출확률} | \text{추출된 표본학급}) \end{aligned}$$

$$\text{학생의 설계 가중치} = \frac{1}{\text{학생의 표본추출확률}}$$

$$\text{가중치 총합} = \sum_{\text{지역별} \times \text{성별}} \text{학생 설계 가중치}$$

$$\text{학생의 사후층화 보정 가중치} = \frac{\text{지역별 전체 학생 수}}{\text{지역별 가중치 총합}}$$

- 전문계 고3 코호트 : 학생의 표본 추출확률과 설계 및 사후층화 보정 가중치

$$\begin{aligned} \text{학생의 표본추출확률} &= (\text{학교유형별 표본 학교 추출확률}) \\ &\times (\text{표본 학급 추출확률} | \text{추출된 표본 학교}) \\ &\times (\text{표본 학생 추출확률} | \text{추출된 표본 학급}) \end{aligned}$$

$$\text{학생의 설계 가중치} = \frac{1}{\text{학생의 표본추출확률}}$$

$$\text{가중치 총합} = \sum_{\text{학교유형별} \times \text{성별}} \text{학생 설계 가중치}$$

$$\text{학생의 사후층화 보정 가중치} = \frac{\text{학교유형별 전체 학생 수}}{\text{학교유형별 가중치 총합}}$$

코호트별로 계산된 학생의 1차년도 최종 가중치는 학생의 설계 가중치와 사후층화 보정 가중치를 이용하여 다음과 같이 계산하면 된다.

$$\text{학생의 1차년도 가중치} = \text{학생의 설계 가중치} \times \text{사후층화 보정 가중치}$$

3. 2차년도 이후 가중치

KEEP은 동일한 연령층인 코호트 내에서 패널을 선정하여 조사하기 때문에 설계 가중치는 동일 코호트를 대표하는 대표성이 적절하게 유지되므로 1차년도의 설계 가중치는 2차년도 이후에도 고정된 기본 가중치로 사용하고 있다. 실제로 2차년도 이후 KEEP 가중치는 전년도 가중치를 기본 설계 가중치로 사용하며, 무응답 보정 및 사후층화 보정 가중치를 추가로 계산하여 3가지 유형의 가중치를 결합하여 2차년도 이후 각 조사년도마다 최종 가중치를 계산하고 있다.

먼저 2차년도 이후 KEEP에서의 무응답 보정 가중치는 2차년도 이후 조사 시점에서 응답률을 계산하여 무응답을 보정하고 있다. 응답률은 각 코호트의 지역별, 성별 및 학교유형별로 응답률을 계산하는 데, 중3 및 일반계 고3 코호트는 (1차년도 학교지역×성별)을 기준으로, 전문계 고3 코호트는 (1차년도 학교유형×성별)을 기준으로 응답률을 계산하며, 무응답 보정 가중치는 응답률의 역수로 다음과 같다.

$$\text{응답률} = \frac{\text{2차년도 이후의 조사년도에서 조사성공 표본 수}}{\text{2차년도 이후의 조사년도에서 유효 표본 수}}$$

$$\text{무응답 보정 가중치} = \frac{1}{\text{응답률}}$$

2차년도 이후 사후층화 보정 가중치도 1차년도와 마찬가지로 조사년도의 가중치 총합과 지역별 학생 수와의 차이를 보정하고 있다. 여기서 조사년도의 가중치 총합은 1차년도 가중치와 조사년도의 무응답 보정 가중치의 곱으로 계산되는 데, 코호트별로 1차년도와는 달리 신분의 변화가 있으므로 조사시점의 신분을 고려하여 계산하고 있다.

2차년도에 고등학생인 패널은 2차년도 이후 (조사년도의 학교지역별×성별)로, 대학생인 패널은 (1차년도 학교지역별×성별×대학진학자), 취업자 혹은 비진학자인 패널은 (1차년도 학교지역별×비진학자)별로 가중치 총합을 계산한다. 2차년도의 신분별 가중치 총합과 무응답 보정 가중치 및 최종 가중치는 다음과 같다.

- 2차년도 고등학생의 가중치 총합

$$\text{가중치 총합} = \sum_{\text{2차년도 학교지역별} \times \text{성별}} (\text{1차년도 가중치} \times \text{2차년도 무응답 보정 가중치})$$

- 2차년도 대학생 가중치 총합

$$\text{가중치 총합} = \sum_{\substack{\text{1차년도 학교지역별} \\ \times \text{성별} \times \text{대학진학자}}} (\text{1차년도 가중치} \times \text{2차년도 무응답 보정 가중치})$$

- 2차년도 취업자 및 비진학자의 가중치 총합

$$\text{가중치 총합} = \sum_{\substack{\text{1차년도 학교지역별} \\ \times \text{비진학자}}} (\text{1차년도 가중치} \times \text{2차년도 무응답 보정 가중치})$$

- 2차년도의 사후층화 보정 가중치 : 모두 적용

$$\text{학생의 사후층화 보정 가중치} = \frac{\text{지역별 전체 학생 수}}{\text{지역별 가중치 총합}}$$

2차년도 가중치는 학생들의 신분 변화를 반영하여 계산하게 되는 데, 최종 가중치는 1차년도 가중치, 무응답 보정 가중치와 사후층화 보정 가중치를 이용하여 다음과 같이 계산하면 된다.

$$\text{학생의 2차년도 가중치} = \text{1차년도 가중치} \times \text{2차년도 무응답 보정 가중치} \times \text{2차년도 사후층화 보정 가중치}$$

3차년도 이후 가중치를 보면, 고등학생은 2차년도와 동일하게 계산하여 적용하지만, 대학생 및 취업자 등 비진학자에 대한 모집단 정보가 없거나 불확실하다고 판단하여 사후층화 보정 가중치를 사용하고 있지 않다. 3차년도 이후 가중치를 정리하면 다음과 같다.

- 3차년도이후 고등학생의 가중치 총합 : 2차년도와 동일

$$\text{가중치 총합} = \sum_{\text{조사시점의 학교지역별} \times \text{성별}} (\text{1차년도 가중치} \times \text{조사시점 무응답 보정 가중치})$$

- 3차년도이후 대학생 및 취업자, 비진학자 가중치 총합 : 계산하지 않음

- 3차년도 이후의 사후층화 보정 가중치 : 고등학생에게만 적용

$$\text{학생의 사후층화 보정 가중치} = \frac{\text{지역별 전체 학생 수}}{\text{지역별 가중치 총합}}$$

3차년도 이후 가중치는 2차년도와 마찬가지로 코호트 패널의 신분 변화를 반영하여 계산하게 되는 데, 고등학생의 최종 가중치는 1차년도 가중치, 무응답 보정 가중치와 사후층화 보정 가중치를 이용하여 계산하지만, 대학생 및 취업, 비진학자는 1차년도 가중치와 무응답 보정 가중치로 계산하고 있다.

- 3차년도이후 고등학생의 가중치

$$\text{학생의 3차년도이후 가중치} = \text{1차년도 가중치} \times \text{조사년도 무응답 보정 가중치} \times \text{조사년도 사후층화 보정 가중치}$$

- 3차년도이후 대학생, 취업 및 비진학자 가중치

$$\text{학생의 3차년도이후 가중치} = \text{1차년도 가중치} \times \text{조사년도 무응답 보정 가중치}$$

4. 코호트별 KEEP의 가중치 분포

앞서 살펴본 바와 같이 KEEP의 가중치는 코호트별, 조사시점에서의 신분 상태별 가중치가 계산되고 있다. 본 절에서는 중3코호트의 1차년도의 가중치와 3차년도의 가중치 분포를 통해 KEEP의 가중치 특성을 살펴보고자 한다. 중3 코호트는 2005년(2차년도) 이후 고등학생으로 진학하게 되므로 1차년도 가중치와 달리 3차년도 가중치는 고등학생의 가중치가 새롭게 계산된다.

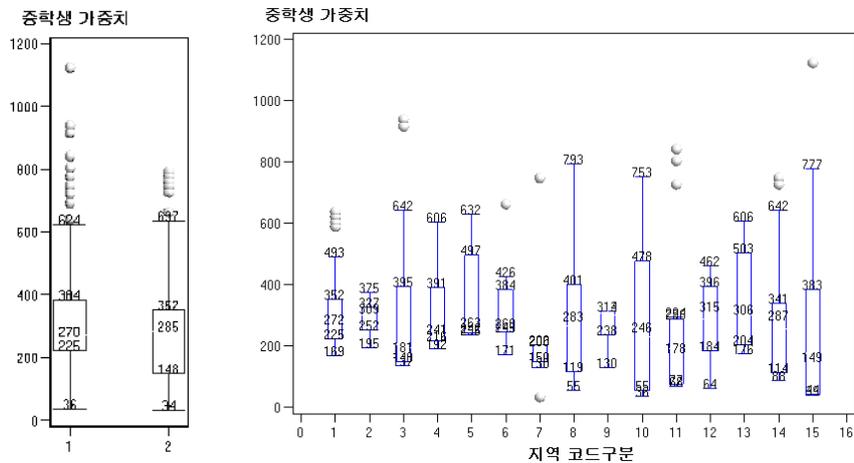
가. 1차년도 가중치 분포

먼저, 중3 코호트의 1차년도 가중치 분포를 성별 및 지역별로 요약한 결과는 <표 4>와 같으며, 가중치의 분포 특성 및 요약 통계를 나타낸 상자그림은 <그림 1>과 같다. 중3 코호트의 1차년도 가중치를 성별로 보면, 남학생이 319.7로 여학생 280.6보다 약간 크지만 평균 가중치나 변동계수 등을 감안해 본다면 성별 가중치의 차이는 크지 않았다. 하지만 남학생의 개별 가중치가 1123으로 아주 큰 값이 있었다.

〈표 4〉 중3 코호트의 1차년도 성별, 지역별 가중치 요약 통계

성별	표본크기	Sum	Mean	StdErr	CV	Max	Min
남자(1)	996	318463.12	319.74	5.71	56.31	1123.56	35.71
여자(2)	1004	281690.54	280.57	5.27	59.51	793.1	34.31
중학생	2000	600153.66	300.08	3.91	58.21	1123.56	34.31

지역	표본크기	Sum	Mean	StdErr	CV	Max	Min
1 서울	400	120596.48	301.49	5.48	36.36	637.45	168.51
2 부산	160	47721.97	298.26	3.77	15.97	375.03	194.97
3 대구	120	36501.75	304.18	20.80	74.92	939.82	136.76
4 인천	120	35504.91	295.87	10.20	37.78	605.84	191.61
5 광주	60	20513.38	341.89	17.81	40.36	631.71	236.25
6 대전	60	20273.98	337.90	20.94	48.00	662.63	171.40
7 울산	60	16371.67	272.86	31.71	90.02	748.30	34.31
8 경기	440	131939.97	299.86	8.19	57.28	793.10	55.04
9 강원	60	16224.37	270.41	8.67	24.84	313.52	129.94
10 충북	60	18313.45	305.22	32.11	81.50	753.27	35.71
11 충남	80	22579.20	282.24	29.06	92.09	842.35	67.77
12 전북	80	23112.29	288.90	15.23	47.16	461.62	63.62
13 전남	60	20677.99	344.63	18.87	42.42	606.07	176.49
14 경북	100	30121.20	301.21	21.16	70.27	748.51	88.03
15 경남	140	39701.05	283.58	24.91	103.92	1123.56	39.49



〈그림 1〉 중3 코호트의 1차년도 가중치

지역별로 살펴보면, 지역별 평균 가중치가 207에서 345정도로 비교적 고르게 나타났으나 상대적인 가중치의 변동은 지역별로 큰 차이가 있었으며, 지역내에서도 개인별 가중치의 차이가 심한 것으로 나타나 표본의 자료 수집 결과에 따라 지역별 모수 추계에 미치는 영향이 매우 큼을 보여주고 있다. <그림 1>에서 보듯이 남학생 가중치 분포에서 지나치게 큰 가중치를 갖는 패널이 있었고, 지역내에서도 서울, 대구, 대전, 울산, 충남, 경남 등에서 개인별 가중치가 큰 패널이 있었다.

나. 3차년도 가중치 분포

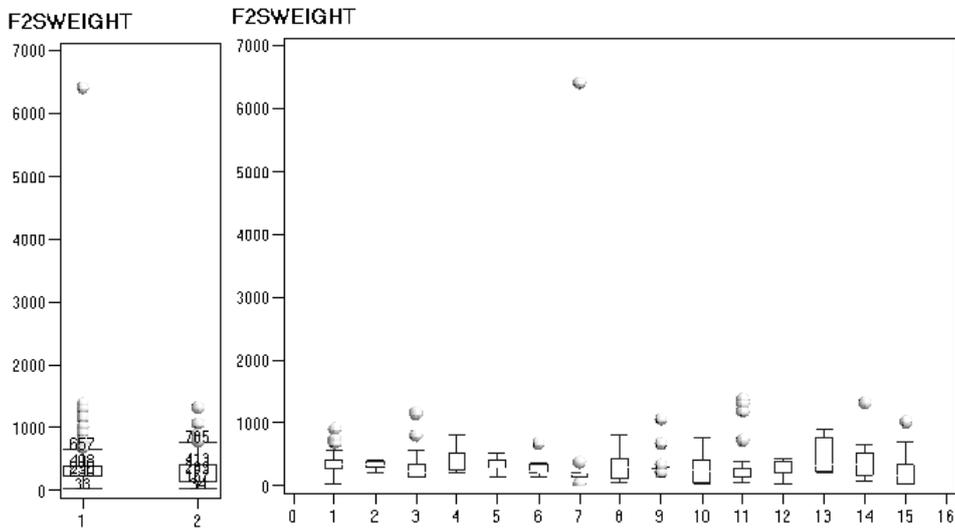
중3 코호트는 3차년도에 패널의 신분이 고등학생, 취업 및 비진학 등으로 신분이 변화하게 된다. 실제로 3차년도에 중3 코호트의 패널 중 87.0%(1740명)가 고교에 재학 중이며, 12명이 취업 혹은 비진학으로 나타났다. 중3 코호트의 패널은 2차년도 이후 일반계 고교 혹은 전문계 고교로 진학하므로 이를 구분하여 가중치를 살펴보기로 한다. 중3 코호트의 3차년도 가중치는 앞서 살펴본 바와 같이 2차년도의 고등학생에 대한 가중치 부여 방법 및 계산 과정에 따라 최종 가중치가 계산된 것이다.

중3 코호트의 일반계 고교에 재학중인 패널의 3차년도 가중치 분포를 성별 및 지역별로 요약한 결과는 <표 5>와 같으며, 가중치의 분포 특성 및 요약 통계를 나타낸 상자그림은 <그림 2>와 같다. 중3 코호트의 일반계 고교 3차년도 가중치를 성별로 보면, 남학생이 359.5로 여학생 309.2보다 약간 크지만 평균 가중치나 변동계수 등을 감안해 본다면 성별 가중치의 차이는 크지 않았다. 하지만 남학생의 개별 가중치가 6407로 아주 큰 값이 있었다. 지역별로 살펴보면, 지역별 평균 가중치가 277에서 438정도로 지역별로 차이가 약간 나타났으며, 특히 전남 지역의 가중치가 크게 나타났으며, 변동계수를 보더라도 지역별로 상대적인 가중치의 변동이 매우 큰 것으로 나타났다. 지역내에서도 개인별 가중치의 차이가 지역별로 매우 심한 것으로 나타나 표본의 자료 수집 결과에 따라 지역별 모수 추계에 미치는 영향이 매우 큼을 보여주고 있다. 특히 울산 지역은 최소 개인 가중치가 34.3인데 반해 최대 가중치는 6406으로 패널 개인간 가중치의 차이 상당히 심한 것으로 나타났다. <그림 2>를 보면 대부분의 지역에서 지역내 개별 가중치가 큰 특이값이 존재하고 있었으며, 특히 울산 지역은 지역내뿐만 아니라 지역간에서도 개인별 가중치가 큰 패널이 있었다.

<표 5> 중3 코호트의 일반계고교의 3차년도 성별, 지역별 가중치 요약 통계

성별	N	Sum	Mean	StdErr	CV	Max	Min
1 남자	622	223584.26	359.46	12.77	88.62	6406.96	33.40
2 여자	651	201265.94	309.16	7.52	62.06	1328.88	34.30
일반고교	1273	424850.20	333.74	7.36	78.70	6406.96	33.40

고교지역	N	Sum	Mean	StdErr	CV	Max	Min
1 서울	262	93359.04	356.33	8.32	37.80	905.68	44.56
2 부산	89	30688.1	344.81	5.87	16.06	424.02	223.18
3 대구	80	26839.04	335.49	31.74	84.62	1177.45	150.72
4 인천	67	24424.46	364.54	18.73	42.06	809.89	223.02
5 광주	41	14694.05	358.39	14.27	25.49	533.85	152.25
6 대전	49	15098.08	308.12	24.04	54.61	676.55	146.50
7 울산	33	11916.00	361.09	189.25	301.07	6406.96	34.30
8 경기	282	90766.74	321.87	10.96	57.18	822.99	60.23
9 강원	30	9829.7	327.66	29.15	48.72	1064.91	222.70
10 충북	42	11764.72	280.11	39.65	91.74	785.05	33.40
11 충남	48	16299.41	339.57	53.68	109.52	1376.69	70.02
12 전북	56	14553.91	259.89	17.46	50.26	448.96	52.45
13 전남	33	14463.10	438.28	42.57	55.80	918.96	205.48
14 경북	58	21633.70	372.99	34.86	71.18	1328.88	90.06
15 경남	103	28520.15	276.89	26.6	97.48	1020.97	34.70
일반고교	1273	424850.2	333.74	7.36	78.7	6406.96	33.40



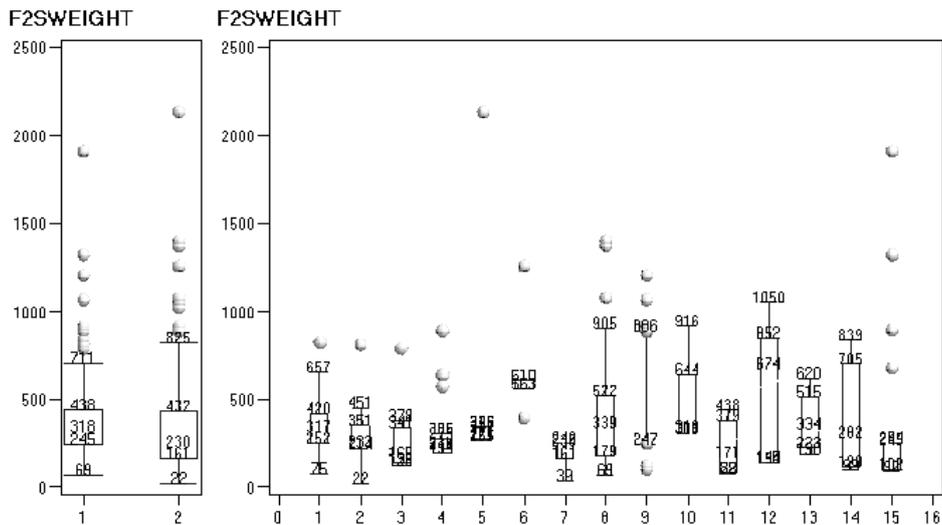
〈그림 2〉 중3 코호트의 일반계고교 패널의 3차년도 가중치

중3 코호트의 전문계 고교에 재학중인 패널의 3차년도 가중치 분포를 성별 및 지역별로 요약한 결과는 <표 6>과 같으며, 가중치의 분포 특성 및 요약 통계를 나타낸 상자그림은 <그림 3>과 같다.

〈표 5〉 중3 코호트의 일반계고교의 3차년도 성별, 지역별 가중치 요약 통계

성별	N	Sum	Mean	StdErr	CV	Max	Min
1 남자	246	88614.14	360.22	14.04	61.13	1908.86	69.04
2 여자	221	75492.32	341.59	20.46	89.03	2133.14	22.09
전문고교	467	164106.46	351.41	12.18	74.88	2133.14	22.09

고교지역	N	Sum	Mean	StdErr	CV	Max	Min
1 서울	69	23229.73	336.66	15.30	37.75	825.36	75.30
2 부산	54	14884.35	275.64	15.96	42.54	810.41	22.09
3 대구	27	8020.24	297.05	43.42	75.95	790.67	125.10
4 인천	29	9364.05	322.90	38.43	64.09	886.95	194.20
5 광주	11	5110.30	464.57	167.11	119.30	2133.14	264.71
6 대전	7	4515.58	645.08	105.35	43.21	1257.68	394.32
7 울산	20	3771.05	188.55	18.54	43.98	249.19	39.42
8 경기	93	37812.61	406.59	29.90	70.91	1397.76	69.04
9 강원	22	8249.66	374.98	67.61	84.57	1204.36	102.65
10 충북	13	6211.32	477.79	73.62	55.56	915.50	308.11
11 충남	31	6339.95	204.51	25.39	69.12	438.45	78.22
12 전북	16	8914.60	557.16	91.54	65.72	1050.34	133.87
13 전남	21	7855.08	374.05	32.45	39.75	619.71	189.83
14 경북	27	9800.74	362.99	51.09	73.13	838.56	99.49
15 경남	27	10027.22	371.38	85.92	120.22	1908.86	91.98
전문고교	467	164106.46	351.41	12.18	74.88	2133.14	22.09



〈그림 3〉 중3 코호트의 전문계고교 패널의 3차년도 가중치

중3 코호트의 전문계 고교 3차년도 가중치를 성별로 보면, 남학생이 360.2로 여학생 341.6으로 평균 가중치나 변동계수 등을 감안해 본다면 성별 가중치의 차이는 거의 없었으나 여학생의 가중치 분포가 남학생보다 다소 개별 가중치 차이가 심한 것으로 나타났다. 지역별로 살펴보면, 지역별 평균 가중치가 205에서 645 정도로 지역별로 차이가 크게 나타났으며, 특히 대전, 전북, 충북, 광주, 경기도 타 지역에 비해 가중치가 크게 나타났으며, 변동계수를 보더라도 지역별로 상대적인 가중치의 변동이 매우 큰 것으로 나타났다. 지역내에서도 개인별 가중치의 차이가 지역별로 매우 심한 것으로 나타나 표본의 자료 수집 결과에 따라 지역별 모수 추계에 미치는 영향이 매우 큼을 보여주고 있다. 특히 광주 지역의 패널은 개인 가중치가 2133, 경남에서 1908로 나타나 지역별 및 지역내 패널 개인간 가중치의 차이 상당히 심한 것으로 나타났다. <그림 3>을 보면 대부분의 지역에서 지역내 개별 가중치가 큰 특이값이 존재하고 있었다.

다. 중3 코호트의 3차년도 가중치 비교

중3 코호트는 3차년도에 일반계고교와 전문계고교에 재학하게 되므로 중3 코호트의 3차년도 가중치를 일반계고교 패널과 전문계 고교 패널 가중치를 비교해 봄으로써 중3 코호트의 개별 가중치 변화 비교 및 3차년도의 표본 대표성을 검토하는 자료가 될 것이다.

중3 코호트의 일반계 고교 패널의 평균 가중치는 351, 전문계 고교 패널의 가중치는 333.7로 거의 비슷하게 나타나 평균 가중치 및 변동계수를 감안해 본다면 중3 코호트는 3차년도에도 표본 대표성을 동일하게 유지하는 것으로 보인다. 성별로 일반계고교와 전문계고교의 가중치를 비교해 보더라도 거의 비슷한 결과를 보여주고 있지만 지역별 일반계고교와 전문계고교의 가중치 차이가 지역에 따라 크게 나타나고 있다. 이는 중3 코호트가 고교로 진학하는 비율이 지역에 따라 차이를 보여주는 결과로 나타난 것이다.

가중치의 비교는 향후 패널을 추가할 때 지역별 패널의 특성을 감안하여 표본을 추가하는 자료로 활용할 수 있을 것이다.

IV. 가중치 효과

일반적으로 표본조사에서는 불균등 추출확률 등의 가중치를 감안하여 모수를 추정하게 된다. 표본이론 측면에서 가중치를 반영하여 추정하면 추정치의 비편향성(unbiasness) 보장 및 분산의 과소 추정 문제를 해결해 준다고 알려져 있다. 중3 코호트의 3차년도 가중치를 이용하여 가중치의 효과를 살펴보았다.

일반적으로 모수 추정시 가중치를 사용하여 가중평균으로 비편향된 결과를 쉽게 얻을 수 있지만 가중치와 분석변수와의 상관관계에 따라 과소 혹은 과대 추정하기도 한다. 즉, 가중

치와 분석 변수사이에 상관관계가 존재하지 않을 때 표본 크기가 증가하면 가중평균은 비편향 결과를 얻을 수 있지만 양의 상관관계를 갖는다면 과대추정, 음의 상관관계는 갖는다면 과소추정하는 경향을 보인다고 한다(김규성, 2004). 또한 가중치가 분산에 미치는 영향은 파악하기 어렵지만 특성치의 분포가 동일한 평균(μ)과 분산(σ_y^2)을 갖는 독립적인 분포에서 선정되었다면 다음과 같은 결과를 갖는다고 알려져 있다.

$$Var(\bar{y}_w | sample) = \frac{\sum w_i^2}{(\sum w_i)^2} \sigma_y^2 = (1 + CV_w^2) \sigma_y^2$$

여기서 $\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$, CV_w 는 가중치의 변동계수이다.

이 결과는 가중치의 산포가 크면 가중평균의 변동계수가 증가하게 되어 가중평균의 신뢰도가 저하된다는 것을 의미하는 것이다(김규성외, 2005).

KEEP은 상당히 많은 조사 항목을 자료 수집하므로 패널의 개인 가중치는 조사 결과의 추정에 많은 영향을 미치게 될 것이므로 가중치의 효과가 다르게 나타날 것이다. 본 연구에서는 KEEP 자료 중 일주일 공부시간, 본인 월평균 용돈, 평균여가시간, 사교육 지출 비용 및 키와 몸무게 자료를 이용하여 가중치의 효과를 분석하였다(<표 6>).

<표 6>을 보면, 일주일 공부시간, 본인 월평균 용돈, 평균여가시간, 사교육 지출 비용 및 키와 몸무게 자료와 가중치와의 상관계수를 나타내었다. 상관계수를 보면, 사교육 지출비용은 가중치와 양의 상관계수(0.14)가 큰 반면 일주일 공부시간은 거의 '0'에 가깝게 작은 값을 가지고 있으며, 본인 월평균 용돈, 평균여가시간 등은 음의 상관관계로 나타나고 있다. 따라서 이들 자료는 가중치와의 상관관계를 통해 본다면 가중평균으로 모평균을 추정하는 경우 과소 혹은 과대 추정될 수 있음을 보여 주고 있으며, 분산도 가중치의 변동계수에 의해 증가될 수 있음을 보여주고 있다.

실제로 가중치를 사용하여 사교육 지출비용을 가중평균으로 추정하는 경우 과대 추정할 가능성이 있으며, 분산도 11.2%(0.334²)만큼 증가할 수 있음을 보여주고 있다.

<표 6> 중3 코호트의 일반계고교의 3차년도 기초 요약 통계와 상관계수

변수	N	평균	변동계수	최소값	최대값
일주일공부시간	1272	10.24	99.47	0	90
본인월평균용돈	1270	53543	92.37	2000	600000
평균여가시간	1270	2.43	64.11	0	10
평균여가시간(주말)	1269	6.99	46.10	0	20
사교육지출비용	772	343772	84.34	0	2000000
키	1271	167.73	4.86	145	192
체중	1267	59.63	18.48	36	103
가중치	1273	333.74	78.70	33.40	6407

상관계수	일주일 공부시간	본인(월) 평균용돈	평균(평일) 여가시간	평균(휴일) 여가시간	사교육 지출비용	키	몸무게
	0.0085	-0.0284	-0.0125	-0.0476	0.1438	0.0663	0.0618
가중치	0.761	0.3113	0.6573	0.0903	<.0001	0.0182	0.0277
	1272	1270	1270	1269	772	1271	1267

V. 결론

KEEP은 코호트 패널조사이며, 동일한 대상을 매년 추적하여 국가인적자원개발 정책, 국가 교육정책 및 노동시장 정책 수립을 위한 기초 자료 제공, 그리고 국가 인적자원개발 정책 수립이나 관련 분야의 연구 자료로 활용 등을 목적으로 수행되는 조사로 종단적인 가치뿐만 아니라 횡단적인 가치를 갖는 매우 의미있는 자료를 제공하고 있다.

본 연구에서는 3차까지 완료된 가중치를 살펴보기 위해 KEEP 패널의 표본 유지 실태, 응답 패턴 및 가중치 특성 파악 및 비교, 그리고 가중치가 추정에 미치는 효과를 검토해 보았다.

앞서 논의한 바와 같이 본 연구에서 나타난 중요한 결과로는 1) 3차년도 패널유지율을 보면, 최초 패널 중 실제 유효 자료를 제공해 주는 유효 표본유지율이 79.2%, 남학생 중 군입대로 인한 잠재 표본유지율이 18.4%로 나타나 97.6%의 패널유지율을 보였고, 2) 패널이 3차년도까지 모두 참여한 패널의 완전응답율은 65.3%를 유지한 반면, 2차 이후 조사가 불가능한 표본탈락율은 14명(0.2%)이었다. 3) 중3 코호트를 대상으로 3차년도 가중치를 비교 분석해 본 결과, 일반계고교와 전문계고교의 대표 대표성은 거의 비슷하게 유지되고 있었으나 지역별 가중치 차이가 크게 나타나 향후 지역별 표본 규모를 점검할 필요가 있는 것으로 나타났다. 4) 가중치의 효과에서는 가중치가 모평균 추정에 미치는 영향을 살펴보았다. 특히, 모평균 추정에서는 가중치와 분석변수와의 상관관계에 따라 과소 및 과대 추정되며, 추정량의 분산도 증가될 수 있음을 확인하였다. 특히 사교육 지출비용은 모평균을 과대추정할 가능성이 있었으며, 분산도 증가할 수 있음을 보였다.

결론적으로 KEEP의 표본 유지 수준은 여타 국내의 패널조사와 비슷한 수준을 보여주지만 2-3년 후 패널의 군제대로 다시 조사에 참여한다면 높은 표본 유지율을 보여주게 될 것으로 기대한다. 또한 군입대자에 대한 효율적인 패널 관리 방안 마련과 군입대로 인한 패널 자료의 공백을 어떻게 해결할 것인가 하는 문제도 고려해 볼 필요가 있을 것이다. 그리고 가중치는 추정량의 비편향성을 유지하기 위해 부여되기 때문에 분석 변수와의 상관관계를 검토하여 추정할 필요가 있으며, 가중치의 산포가 클수록 가중평균의 분산도 증가하여 추정 효율이 떨어짐을 확인하여 추정할 필요가 있음을 유념할 필요가 있다.

참고문헌

- 강석훈(2003), KLIPS의 가중치 부여방안 연구, 한국노동패널연구 2003-4, 한국노동연구원.
- 김규성(2004), 표본조사에서 가중치 부여 효과에 대한 연구, 한국조사연구학회 추계학술발표대회 논문집, 3-12.
- 김규성, 황영은, 박진우(2005), 패널조사에서 가중치 부여 방법 및 효과에 관한 연구, 제6회 한국노동패널 학술대회 논문집, 311-322.
- 김영원, 김재광, 이기재, 조유미(2005), 한국노동패널 표본의 대표성과 가중치 보정 방법, 제6회 한국노동패널 학술대회 논문집, 323-342.
- 한국교육고용패널조사(2005), 한국직업능력개발원.
- 한국교육고용패널조사(2006), 한국직업능력개발원.
- 한국교육고용패널조사(2007), 한국직업능력개발원.
- 한국교육고용패널조사 1차-3차년도 조사 사용자 지침서(2007), 한국직업능력개발원.
- 한국복지패널 가중치 회의 자료(2009), 한국보건사회연구원.

Abstract

Panel Maintenance Patterns and Weights at KEEPSurvey

Jong-Seok Byun(Hanshin University)

Using KEEP 1-3 data, we analyse the panel maintenance patterns and weights at KEEP survey. The results are reported as follows.

- 1) At third-wave, the survey on 79.2% out of new panel was completed, but the survey maintains 97.6% of panel including the potential panel.
- 2) To third wave, Keep successfully surveyed 65.3% of panels.
- 3) In case of Middle student cohort, representativeness of sample guarantee fully to third-wave. But the panel size of the region will need to inspect in the future because the weights is greatly different by the region and the panel.
- 4) We verify the effects of the weight on the estimation of the parameters and the variance estimation.