

**한국아동·청소년패널조사(KCYPS)**  
**제 1-7 차 조사 데이터 유저 가이드**

**2017. 08**

■ 자료 업데이트 안내

- KCYPS 유저가이드는 연 1 회 이상 부정기적으로 업데이트됩니다.
- 다음 사이트에서 가장 최신의 유저가이드를 내려받을 수 있습니다.
- 한국아동·청소년 데이터 아카이브 : <http://archive.nypi.re.kr>

■ 자료 이용 관련 문의

- 데이터 이용 관련 :
- 정윤미 : ☎ 044-415-2188, [whandym@nypi.re.kr](mailto:whandym@nypi.re.kr)
- 패널 학술대회 관련 :
- 한지형 : ☎ 044-415-2244, [hjh@nypi.re.kr](mailto:hjh@nypi.re.kr)

# 차례

— 데이터 이용 관련 : .....	0
— 패널 학술대회 관련 : .....	0
I. 조사목적 .....	1
II. 조사모형 .....	3
1. 종단조사모형 .....	3
2. 조사모형의 특성 .....	5
III. 표집설계 .....	7
1. 표집설계 개요 .....	7
2. 패널별 표집절차 .....	9
3. KCYPS 원표본 .....	15
IV. 조사내용 .....	16
1. 조사영역·내용 구성 .....	16
2. 조사문항 구성 .....	19
3. 문항별 조사주기 .....	20
V. 조사실사 .....	21
1. 제1차년도 조사 .....	21
2. 제2~7차년도 조사 .....	22
VI. 데이터 구축 .....	23
1. 데이터 구축 현황 .....	23
2. 데이터 클리닝 개요 .....	27
VII. 데이터 가중치 .....	29
1. 초1 패널 가중치 .....	29
2. 초4 패널 가중치 .....	53
3. 중1 패널 가중치 .....	77
□ KCYPS 데이터 활용 유의사항 .....	100
참고문헌 .....	101

## 표 차례

<표 1> 초1 패널 모집단 구성(2009년).....	9
<표 2> 초1 패널 시·도별/도시규모별 표본학교 수.....	10
<표 3> 초4 패널 모집단 구성(2009년).....	11
<표 4> 초4 패널 시·도별/도시규모별 표본학교 수.....	12
<표 5> 중1 패널 모집단 구성(2009년).....	13
<표 6> 중1 패널 시·도별/도시규모별 표본학교 수.....	14
<표 7> KCYPS의 원표본.....	15
<표 8> KCYPS 조사문항 체계.....	18
<표 9> KCYPS 원표본 유지율(2010~2016).....	23
<표 10> 초1 패널 조사 데이터 구축 현황.....	24
<표 11> 초4 패널 조사 데이터 구축 현황.....	25
<표 12> 중1 패널 조사 데이터 구축 현황.....	26
<표 13> 초1 패널 2차년도 횡단면 가중치의 요약 통계량.....	31
<표 14> 초1 패널 2차년도 횡단면 가중치의 지역별/성별 분포.....	31
<표 15> 초1 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	32
<표 16> 초1 패널 2차년도 응답확률 추정값의 요약 통계량.....	32
<표 17> 초1 패널 2차년도 종단면 가중치의 요약 통계량.....	33
<표 18> 초1 패널 2차년도 종단면 가중치의 지역별/성별 분포.....	33
<표 19> 초1 패널 3차년도 횡단면 가중치의 요약 통계량.....	34
<표 20> 초1 패널 3차년도 횡단면 가중치의 지역별/성별 분포.....	35
<표 21> 초1 패널 3차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	36
<표 22> 초1 패널 3차년도 응답확률 추정값의 요약 통계량.....	36
<표 23> 초1 패널 3차년도 종단면 가중치의 요약 통계량.....	37
<표 24> 초1 패널 3차년도 종단면 가중치의 지역별/성별 분포.....	37
<표 25> 초1 패널 4차년도 횡단면 가중치의 요약통계량.....	38
<표 26> 초1 패널 4차년도 횡단면 가중치의 지역별/성별 분포.....	38
<표 27> 초1 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	39
<표 28> 초1 패널 4차년도 응답확률 추정값의 요약통계량.....	39
<표 29> 초1 패널 4차년도 종단면 가중치의 요약통계량.....	40
<표 30> 초1 패널 4차년도 종단면 가중치의 지역별/성별 분포.....	40

<표 31> 초1 패널 5차년도 횡단면 가중치의 요약통계량.....	41
<표 32> 초1 패널 5차년도 횡단면 가중치의 지역별/성별 분포 .....	42
<표 33> 초1 패널 5차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	43
<표 34> 초1 패널 5차년도 응답확률 추정값의 요약통계량 .....	43
<표 35> 초1 패널 5차년도 종단면 가중치의 요약통계량.....	44
<표 36> 초1 패널 5차년도 종단면 가중치의 지역별/성별 분포 .....	44
<표 37> 초1 패널 6차년도 횡단면 가중치의 요약통계량.....	45
<표 38> 초1 패널 6차년도 횡단면 가중치의 지역별/성별 분포 .....	45
<표 39> 초1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	46
<표 40> 초1 패널 6차년도 응답확률 추정값의 요약통계량 .....	47
<표 41> 초1 패널 6차년도 종단면 가중치의 요약통계량.....	47
<표 42> 초1 패널 6차년도 종단면 가중치의 지역별/성별 분포 .....	48
<표 43> 초1 패널 7차년도 횡단면 가중치의 요약통계량.....	49
<표 44> 초1 패널 7차년도 횡단면 가중치의 지역별/성별 분포 .....	49
<표 45> 초1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	50
<표 46> 초1 패널 6차년도 응답확률 추정값의 요약통계량 .....	51
<표 47> 초1 패널 7차년도 종단면 가중치의 요약통계량.....	51
<표 48> 초1 패널 7차년도 종단면 가중치의 지역별/성별 분포 .....	52
<표 49> 초4 패널 2차년도 횡단면 가중치의 요약 통계량 .....	55
<표 50> 초4 패널 2차년도 횡단면 가중치의 지역별/성별 분포 .....	55
<표 51> 초4 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	56
<표 52> 초4 패널 2차년도 응답확률 추정값의 요약 통계량.....	57
<표 53> 초4 패널 2차년도 종단면 가중치의 요약 통계량 .....	57
<표 54> 초4 패널 2차년도 종단면 가중치의 지역별/성별 분포 .....	58
<표 55> 초4 패널 3차년도 횡단면 가중치의 요약 통계량 .....	59
<표 56> 초4 패널 3차년도 횡단면 가중치의 지역별/성별 분포 .....	59
<표 57> 초4 패널 3차년도 종단면 가중치의 요약 통계량 .....	60
<표 58> 초4 패널 3차년도 종단면 가중치의 지역별/성별 분포 .....	61
<표 59> 초4 패널 4차년도 횡단면 가중치의 요약통계량.....	62
<표 60> 초4 패널 4차년도 횡단면 가중치의 지역별/성별 분포 .....	62
<표 61> 초4 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	63
<표 62> 초4 패널 4차년도 응답확률 추정값의 요약통계량 .....	64
<표 63> 초4 패널 4차년도 종단면 가중치의 요약통계량.....	64

<표 64> 초4 패널 4차년도 종단면 가중치의 지역별/성별 분포 .....	65
<표 65> 초4 패널 5차년도 횡단면 가중치의 요약통계량.....	66
<표 66> 초4 패널 5차년도 횡단면 가중치의 지역별/성별 분포 .....	66
<표 67> 초4 패널 5차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	67
<표 68> 초4 패널 5차년도 응답확률 추정값의 요약통계량 .....	67
<표 69> 초4 패널 5차년도 종단면 가중치의 요약통계량.....	68
<표 70> 초4 패널 5차년도 종단면 가중치의 지역별/성별 분포 .....	69
<표 71> 초4 패널 6차년도 횡단면 가중치의 요약통계량.....	70
<표 72> 초4 패널 6차년도 횡단면 가중치의 지역별/성별 분포 .....	70
<표 73> 초4 패널 6차년도 종단면 가중치의 요약통계량.....	72
<표 74> 초4 패널 6차년도 종단면 가중치의 지역별/성별 분포 .....	72
<표 75> 초4 패널 7차년도 횡단면 가중치의 요약통계량.....	73
<표 76> 초4 패널 7차년도 횡단면 가중치의 지역별/성별 분포 .....	73
<표 77> 초4 패널 7차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	74
<표 78> 초4 패널 7차년도 응답확률 추정값의 요약통계량 .....	75
<표 79> 초4 패널 7차년도 종단면 가중치의 요약통계량.....	75
<표 80> 초4 패널 7차년도 종단면 가중치의 지역별/성별 분포 .....	76
<표 81> 중1 패널 2차년도 횡단면 가중치의 요약 통계량 .....	78
<표 82> 중1 패널 2차년도 횡단면 가중치의 지역별/성별 분포 .....	79
<표 83> 중1 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	80
<표 84> 중1 패널 2차년도 응답확률 추정값의 요약 통계량.....	80
<표 85> 중1 패널 2차년도 종단면 가중치의 요약 통계량 .....	81
<표 86> 중1 패널 2차년도 종단면 가중치의 지역별/성별 분포 .....	81
<표 87> 중1 패널 3차년도 횡단면 가중치의 요약 통계량 .....	82
<표 88> 중1 패널 3차년도 횡단면 가중치의 지역별/성별 분포 .....	82
<표 89> 중1 패널 3차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	83
<표 90> 중1 패널 3차년도 응답확률 추정값의 요약 통계량.....	83
<표 91> 중1 패널 3차년도 종단면 가중치의 요약 통계량 .....	84
<표 92> 중1 패널 3차년도 종단면 가중치의 지역별/성별 분포 .....	84
<표 93> 중1 패널 4차년도 횡단면 가중치의 요약통계량.....	85
<표 94> 중1 패널 4차년도 횡단면 가중치의 지역별/성별 분포 .....	86
<표 95> 중1 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과.....	87
<표 96> 중1 패널 4차년도 응답확률 추정값의 요약통계량 .....	87

<표 97> 중1 패널 4차년도 종단면 가중치의 요약통계량.....	88
<표 98> 중1 패널 4차년도 종단면 가중치의 지역별/성별 분포 .....	88
<표 99> 중1 패널 5차년도 횡단면 가중치의 요약통계량.....	89
<표 100> 중1 패널 5차년도 횡단면 가중치의 지역별/성별 분포 .....	90
<표 101> 중1 패널 5차년도 종단면 가중치의 요약통계량.....	91
<표 102> 중1 패널 5차년도 종단면 가중치의 지역별/성별 분포 .....	92
<표 103> 중1 패널 6차년도 횡단면 가중치의 요약통계량.....	93
<표 104> 중1 패널 6차년도 횡단면 가중치의 지역별/성별 분포 .....	93
<표 105> 중1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과 .....	94
<표 106> 중1 패널 6차년도 응답확률 추정값의 요약통계량 .....	94
<표 107> 중1 패널 6차년도 종단면 가중치의 요약통계량.....	95
<표 108> 중1 패널 6차년도 종단면 가중치의 지역별/성별 분포 .....	95
<표 109> 중1 패널 7차년도 횡단면 가중치의 요약통계량.....	96
<표 110> 중1 패널 7차년도 횡단면 가중치의 지역별/성별 분포 .....	97
<표 111> 중1 패널 7차년도 종단면 가중치의 요약통계량.....	98
<표 112> 중1 패널 7차년도 종단면 가중치의 지역별/성별 분포 .....	99

## 그림 차례

[그림 1] KCYPS의 종단조사모형 .....	3
[그림 2] 표본대체가 없는 단일표본설계 .....	4
[그림 3] KCYPS 조사영역 기본 개념도.....	16
[그림 4] 데이터 클리닝 개념도.....	27





# I . 조사목적

초등학생에서 고등학생에 이르는 아동·청소년기는 생애단계에서 가장 많은 신체적·정신적 변화를 경험하는 시기이다. 특히 청소년들은 급속한 신체 변화와 인지 발달, 생활환경과 사회관계의 확대 등 변화된 상황 속에서 정체성의 혼란과 정서적 불안 등의 많은 문제에 부딪치게 된다. 이와 같은 발달상의 문제에 직면하여 청소년들이 주어진 발달과업을 성공적으로 완수하고 사회공동체의 건강한 자아상을 확립하도록 돕는 일은 청소년 연구의 중요한 과제라고 할 수 있다.

아동·청소년의 성장·발달에 관한 경험적 조사연구는 그동안 다양한 학문분과에서 지속적으로 추진되어 왔다. 이들 연구는 대부분 특정 시점의 조사자료를 활용한 횡단적 연구(cross-sectional study)로서, 패널 데이터를 토대로 한 종단적 연구(longitudinal survey)는 드물었다. 동일 대상을 시차를 두고 반복조사하는 패널 데이터는 변인간 인과관계의 설명이 용이하고 시간경과에 따른 개인·집단 변화의 동태적 분석이 가능하다는 장점이 있다. 따라서 아동·청소년을 대상으로 한 패널조사는 그들이 시간의 연속선상에서 어떤 성장과 발달 과정을 거치게 되는지, 또한 그 과정에서 겪게 되는 문제들과 그 요인은 무엇인지에 대한 학문적·정책적 시사점을 제공할 수 있다.

현재 국내에서 아동·청소년이 조사대상에 포함된 패널조사는 한국아동·청소년패널조사를 제외하고 대략 6종에 이른다. 하지만 이들 조사는 주로 후기 청소년을 대상으로 하거나, 노동·교육 등 특정 영역과 주제에 국한된 것으로서, 초·중·고교생을 포괄하는 아동·청소년 성장과 발달의 전체상을 파악하기에는 어려움이 있다.

이와 같은 측면에서 볼 때, 한국청소년정책연구원이 지난 2003~2008년 기간 동안 수행한 한국청소년패널조사(이경상, 임희진, 안선영, 2008)는 특정 주제나 생활영역에 국한되지 않고 아동·청소년 성장과 발달의 전체상을 파악하려 한 국내 최초의 종단조사라고 할 수 있다. 즉, 아동(초4 패널)과 청소년(중2 패널)을 모두 조사대상에 포함한 한편, 진로준비, 진로설정 등 선행 조사(노동패널, 청년패널 등)의 전통적인 주제와 더불어 청소년활동, 여가생활, 비행, 자아관, 심리특성 등 조사내용을 크게 확장하였다. 한국청소년패널조사는 지금까지 데이터 활용 논문(학위논문 및 연구논문)이 약 500여편에 달하고, 다양한 청소년정책 수립에 기초자료로 활용되는 등 아동·청소년 관련 학술연구와 정책 개발에 크게 기여한 것으로 평가받고 있다.

한국아동·청소년패널조사(Korean Children and Youth Panel Survey)는 한국청소년패널조사의 경험을 바탕으로 조사대상과 연구내용을 보다 확대하여 아동·청소년 성장과 발달의 다양한 양상을 종합적으로 파악하는 것을 목적으로 하고 있다. 2010년에 표집된 초등학교 1학년과 4학년, 중학교 1학년의 3개 패널

총 7,071명과 그 보호자를 대상으로 2016년까지 7개년에 걸쳐 추적조사를 실시하였으며, 매년 조사데이터의 일반 공개, 한국아동·청소년패널학술대회 및 대학원생논문 경진대회 개최, 데이터 분석보고서 발간 등을 통해 연구성과의 사회적 확산을 위한 사업을 전개하고 있다.

## Ⅱ. 조사모형

### 1. 종단조사모형

한국아동·청소년패널조사(이하 “KCYPS”)의 조사대상은 2010년 기준 초등학교 1학년과 4학년, 중학교 1학년의 3개 패널(이하 각각 “초1 패널” “초4 패널” “중1 패널”)로 구성되어 있다. 표본은 2009년 교육부 학교통계를 표집틀로 하여 층화다단계집락표집 방식으로 총 7,071명(초1 2,342명, 초4 2,378명, 중1 2,351명)을 선정하였다. 이들을 대상으로 2010년부터 2016년까지 7개년에 걸쳐 매년 추적조사를 실시함으로써, 초1에서 중1(초1 패널), 초4에서 고1(초4 패널), 중1에서 대1(중1 패널)에 이르는 아동·청소년기의 성장·발달 과정을 파악하게 된다.

초1, 초4, 중1의 3개 패널을 대상으로 7개년에 걸쳐 실시되는 한국아동·청소년패널의 종단조사모형은 다음 그림과 같다.

[그림 1] KCYPS의 종단조사모형

	제1차 (2010)	제2차 (2011)	제3차 (2012)	제4차 (2013)	제5차 (2014)	제6차 (2015)	제7차 (2016)
<b>초1 패널</b>	초1	초2	초3				
<b>초4 패널</b>	초4	초5	초6	초4	초5	초6	
<b>중1 패널</b>	중1	중2	중3	중1	중2	중3	중1
				고1	고2	고3	고1
							대1

이와 같은 종단조사 모형 구성을 통해 아동·청소년의 성장과 발달에 대한 다음 세가지 방향의 분석이 가능하다.

첫째, 횡단분석으로서 특정 시점에서 같은 연령집단 내부 또는 상이한 연령집단 간에는 어떤 유사성과 차이점이 나타나는가를 분석할 수 있다(그림 1)의 세로 방향 분석).

둘째, 종단비교 분석으로서 특정 연령집단의 특성과 성격이 시간의 경과에 따라 어떻게 변화해 가는가

를 분석할 수 있다(그림 1의 가로 방향 분석).

셋째, 시계열분석으로서 특정 연령집단이 시간의 경과에 따라 어떤 성장과 발달의 과정을 거치게 되는가를 분석할 수 있다(그림 1의 대각선 방향 분석).

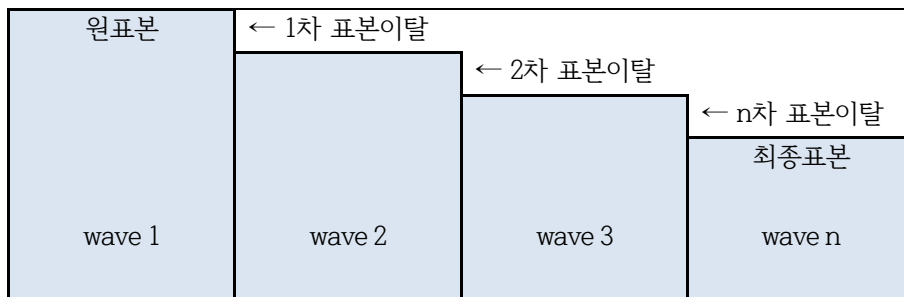
패널조사는 시계열적인 변화를 분석하는 종단연구의 일반적인 특성을 공유하지만, 그 구체적인 조사방법에 있어서는 다양한 갈래가 존재한다. KCYPS의 조사설계상 특성을 정리하면 다음과 같다.

첫째, KCYPS는 동일 대상에 대하여 여러 차례 조사를 반복하는 중다전망적 패널설계(multiple point prospective panel design)를 채택하고 있다. 2010년에서 2016년까지 총 7회에 걸쳐 조사가 이루어지는데, 국내·외의 다른 조사에 비해 조사기간이 짧은 단기 종단연구로 분류될 수 있다. 조사기간을 7년으로 설정한 것은 이 기간이 초등학교에서 중학교(초1 패널)와 고등학교(초4 패널), 중학교에서 대학교(중1 패널)로 이행하는 과정을 관찰할 수 있는 최적의 기간으로 볼 수 있기 때문이다.

둘째, KCYPS는 3개의 코호트를 대상으로 하는 중다패널설계(multiple panel design) 방식을 따르고 있다. 지금까지 국내 아동·청소년 관련 패널조사의 조사대상은 대부분 단일 표본이었고 중다패널의 사례는 많지 않았다. 한국청소년패널조사(KYPS, 2003~2008)는 초4와 중2 패널을 대상으로 하여 처음으로 중다패널 설계를 채택한 바 있다. 한국청소년패널조사의 후속연구인 KCYPS는 초4 패널, 중1 패널과 더불어 기존 조사에서 잘 다루지 않았던 초1 연령층을 조사대상에 포함하여 7개년간 조사함으로써 아동·청소년기 전반의 성장과 발달 과정을 포괄적으로 파악하고자 하였다.

셋째, 패널조사는 조사가 거듭될수록 불가피하게 표본 이탈(sample attrition)이 발생하는데, 이를 해결하기 위하여 순차적으로 표본을 투입하거나 표본을 대체하는 방법을 사용하기도 한다. KCYPS는 단기간의 조사이고 조사대상이 학생층이어서 상대적으로 표본 이탈이 적을 것으로 예상되어 표본 대체가 없는 단일(동일) 표본 설계 방식을 채택하였고, 표본 이탈로 인한 대표성의 문제를 데이터 가중치(weight) 설정을 통해 보정하고 있다.

[그림 2] 표본대체가 없는 단일표본설계



자료 : 박창남, 김희진, 김선업 (2001)의 자료를 수정·보완

## 2. 조사모형의 특성

다음으로, KCYPS 조사모형 설정의 근거와 배경을 패널 설계, 조사대상 연령 선정, 조사기간의 세가지 측면에서 기술하면 다음과 같다<sup>1</sup>.

### 중다패널 설계

KCYPS는 아동복지법상의 아동연령(만18세이하)과 청소년기본법상 청소년 연령(만 9세~24세), 그리고 교육법상의 아동연령(만 6세~12세) 모두에 해당되는 아동·청소년시기에 초점을 맞추어 초등학교에서 고등학교까지 학령기 전반에 있어서의 아동·청소년의 성장·발달에 주목하고자 하였다. 이는 학교에서 노동시장으로의 이행이나 진로·직업 등에 관심을 둔 다수의 선행 패널조사들과 차별성을 갖는다. 관련 패널 조사에서는 대체로 만15세 이상 30세 미만 연령에 해당하는 포괄적 연령층을 조사대상으로 하거나(노동패널, 청년패널 등), 하나의 패널(교육종단연구: 2005년 중1, 아동패널: 2008년 신생아)로 설계된 것이 다수이며, 소수의 패널조사에서 두 개의 패널(교육종단연구: 중3과 고3, 청소년패널: 중2와 초4)를 중단기로 추적조사한 사례가 있다.

KCYPS는 이와 같은 선행 패널조사들과 차별되는, 즉 만15세 이상 후기 청소년을 대상으로 하여 학교에서 노동시장으로의 이행이나 청소년기에서 성인기로의 이행에 초점을 두기보다는 아동·청소년 시기 그 자체에 집중하여 학령기 전반에 걸친 성장·발달을 동태적으로 관찰하고자 한다. 이와 같은 패널 구축의 목적을 달성하기 위하여 초등학생과 중학생을 조사대상으로 하여 학령기 전반에 관한 포괄적인 기초자료를 효율적으로 확보하고자 3개(초1, 초4, 중1)의 연령 패널을 설정하였다.

### 조사대상 연령 선정

아동·청소년을 조사대상으로 포함하는 국내 패널조사들의 대부분이 15세 이상 후기 청소년들을 조사대상으로 하고 있는 반면, 전문가들 사이에는 학령기로의 이행과 학교생활의 적응 등을 포함하는 초등학교 저학년의 전반적인 발달·성장 과정을 살펴볼 수 있는 자료에 대한 요구가 있어 왔다. 이와 같은 요구에 대응하는 실제적 종단자료의 생산과 제공은 이루어지지 않았다는 평가에 따라 초등학교 1학년이 조사대상 패널의 하나로 선정되었다.

선행 청소년패널조사에서도 포함되었던 초등학교 4학년 패널은 아동·청소년 관련 조사연구에서 가장 빈번하게 조사되어 온 연령이다. 이는 초등학교 4학년이 아동 전기에서 후기로 전환하는 과정에서 성장·발달의 변화양상을 가장 잘 포착할 수 있는 시기라고 평가되어 왔기 때문이다. 또한 초4 패널의 종단자료를 구축하게 되면, 청소년패널에서의 2004년 초4 패널 자료와의 패널 비교가 가능하게 되는 장점도 있다. 이에 본 연구에서는 초4 패널을 조사대상 연령패널로 설정하였다.

---

<sup>1</sup> 제1차년도 연구보고서(김지경 외, 2010, pp. 61-68)의 내용을 요약정리한 것임

초등학교 4학년과 더불어 중학교 2학년은 아동·청소년관련 조사연구에서 가장 빈번하게 조사대상이 되어 온 연령이다. 그러나 KCYPS의 2009년 기초연구에서 최근 아동·청소년의 신체적, 지적, 사회·정서적 발달이 이전 세대에 비해 빨라지면서 중학교 시기의 변화가 중학교 1학년과 2학년 사이에 급격하게 이루어지는 경우가 많다는 관계전문가의 의견이 있었다. 이러한 논의에 따르면 중학교 2학년보다는 중학교 1학년이 조사대상으로 더 적절한 것으로 판단되어 중1 패널을 조사대상에 포함하였다.

#### 조사기간

KCYPS의 조사기간은 2010~2016년의 7년간으로 설계하였다. 7년간의 단기종단연구로 설계한 것은 조사대상 패널(초1, 초4, 중1)의 추적 조사를 통하여 초등학교에서 중학교로의 이행(transition)과정과 중학교에서 고등학교, 그리고 대학 진학까지의 이행과정을 관찰하는 데 가장 적절한 기간으로 평가되었기 때문이다. 즉 7년이라는 기간이 초1 패널이 중학교 1학년이 되기까지, 초4 패널이 고등학교 1학년이 되기까지, 그리고 중1 패널이 대학교 1학년이 되기까지의 이행과정을 제한적이기는 하나 단기종단조사로써 파악할 수 있는 가장 최적의 조사 기간으로 평가되기 때문이다. 이와 같은 조사기간의 설계는 초1 패널을 조사대상자에 포함하여 3개의 중다패널을 조사대상자로 설정하였을 때 가능한 것이기도 하다.

## Ⅲ. 표집설계

### 1. 표집설계 개요

KCYPS의 7개년간에 걸친 종단조사의 원표본<sup>2</sup> 표집설계와 표집절차를 설명하면 다음과 같다<sup>3</sup>. KCYPS의 모집단은 2010년 현재 전국의 초등학교 1학년과 4학년, 중학교 1학년에 재학중인 학생과 그 보호자(학부모)인 아동·청소년들로서 2009년 교육부의 전국 학교통계를 표집틀로 하였다.

표본의 크기는 대표성과 정밀성, 조사비용, 그리고 과거 본원이 수행했던 한국청소년패널조사(KYPS)의 원표본 유지율을 참고하여 전문가 자문과 논의를 거친 후 각 패널별 목표표본을 2,200명씩 총 6,600명으로 설정하였다. 표본의 추출은 다단계층화집락추출법(multi-stage stratified cluster sampling)을 사용하여 다음의 단계별로 이루어졌다.

#### 1단계 : 목표표본 할당

2010년 현재 16개 시·도의 초등학교 1학년과 4학년, 중학교 1학년 학생 수에 비례하여 지역별로 조사할 표본 수를 할당하였다. 조사대상 학교 수는 한 학교 당 한 학급씩, 학급 학생 전원을 조사하되 조사성공률(유효응답률)을 80%로 예상하여 산출하였다.

#### 2단계 : 표본학교 추출

조사대상 학교는 확률비례통계추출법(PPS sampling)에 의거하여 추출하였다. 초등학교 1학년과 4학년 표본은 시·도 및 도시규모<sup>4</sup>로, 중학교 1학년은 여기에 남녀공학 구분(남녀공학 학교/남자학교/여자학교)을 추가하여 전체 학교 목록을 작성한 뒤 조사대상 학교를 선정하였다. 1차 학교 선정 후 해당 학교의 조사대상 학년이 최소한 2개 학급 이상인지, 학생 수가 50명 이상인지를 확인하고 이 요건을 충족하지 않을 경우 학교 목록상 인접한 다른 학교로 대체하였다.

#### 3단계 : 표본학급 추출

조사대상 학급은 학교의 특성과 소재지를 확인하고, 해당 학년의 학급 수와 학급별 학생 수(중학교는 남녀별 학생 수 포함)에 대한 정보를 확보한 뒤 무작위로 선정하였다.

<sup>2</sup> “원표본”은 조사 1차년도에 표본으로 선정된 뒤 조사가 완료된 유효표본으로서, 이후 지속적인 추적조사의 대상이 되는 집단을 의미한다.

<sup>3</sup> 제1차년도 보고서(김지경 외, pp. 68-78)에 수록된 표본설계 내용을 요약·정리한 것임

<sup>4</sup> 16개 광역시·도와 도시규모(대도시/중소도시/군지역)를 교차하여 추출한 지역 집락은 27개(서울 2, 부산·대구·인천·광주·대전·울산 각 1, 경기·강원·충북·충남·전북·전남·경북·경남 각 2, 제주 1)이다. 지역 집락의 세부내역은 제1차년도 보고서(김지경 외, 2010, pp. 71) 참조.



#### 4단계: 대체표본 추출

표본학교 및 학급이 선정된 뒤 해당 학교에 조사의 취지를 설명하는 협조 요청 공문을 발송하였다. 1차 접촉에서 조사를 허락한 학교는 전체 271개교 중 174개교로서 64% (초1 패널 57%, 초4 패널 64%, 중1 패널 73%)였는데, 조사를 거절한 학교에 대해서는 학교 목록상에 인접한 다른 학교로 대체하였다.

## 2. 패널별 표집절차

### 1) 초1 패널

- 표집틀 : 교육부 전국 학교통계
- 표집방법 : 층화다단계집락표집(stratified multi-stage cluster sampling)
- 모집단 : 2010년 4월 기준 전국의 초등학교 1학년 재학생

<표 1> 초1 패널 모집단 구성(2009년)

지역		학교 수	학급 수	학생 수
계		6,046	16,707	466,268
서울		584	2,800	81,776
광역시	부산광역시	299	1,039	27,317
	대구광역시	214	793	23,028
	인천광역시	232	916	25,661
	광주광역시	146	529	15,815
	대전광역시	139	520	15,516
	울산광역시	118	373	11,064
	시군부	경기도	1,141	3,852
강원도		386	590	14,298
충청북도		273	560	14,866
충청남도		442	796	19,517
전라북도		413	752	17,992
전라남도		496	808	17,403
경상북도		540	948	23,196
경상남도		508	1,196	31,798
제주도		115	235	6,292

- 표집절차
  - 16개 시도를 시지역과 군지역으로 구분한 후 이들의 지역적 근접성을 기준으로 총 27개의 집락으로 나눔
  - 서울시는 강남과 강북으로, 경기도는 대도시와 중소도시(1,2,3)로, 기타 도지역은 중소도시와 군지역으로 집락화함(광역시는 한 개의 층으로 집락화).
  - 초1 패널 목표 표본수를 2,200명으로 정하고, 각 시도별 층별 배분은 학급수를 기준으로 비례배분함. 이에 따라 총 98개의 표본학교수 결정. 학교당 평균 80%의 유효응답률을 가정하여 전체 표본수 산정
  - 각 시도별 최소 표본 학교 수를 우선 할당한 후에 학생 수를 기준으로 비례 배분법으로 각 시·도별 할당. 27개 각 집락별로 각각 할당된 표본학교 추출
  - 학교별 초등학교 1학년 학생수를 반영한 규모비례확률표집법(PPS) 적용. 시·도별, 도시규모별 층 내

에서 학급 수를 기준으로 확률비례추출법에 따라 표본학교 선정(단, 2개 학급 이상이고 학생수가 50명 이상인 학교만 선정. 만약 위와 같은 조건을 충족하지 않을 경우에는 바로 인접한 학교를 표본학교로 대체)

- 추출된 학교의 초등학교 1학년 학급수에서 난수를 발생시켜 1개 학급 무작위 추출(사전 학교 섭외과정에서 우열반 등 특별편성이 되어 있는 학급은 제외)
- 추출된 학교 측의 조사협조 거절로 학교방문 조사가 불가능할 경우, 학교 리스트에서 추출된 학교의 아래, 위 학교순으로 대체(학교거절을 고려해 사전에 2순위 대체 학교 추출. 2순위 학교에서도 조사가 불가능할 경우 3순위 학교로 대체)

<표 2> 초1 패널 시·도별/도시규모별 표본학교 수

지역	규모	총구분	모집단		할당된 표본학교수	
			학교 수	학급 수		
	계		6,046	16,707	98	
서울	대도시	서울대1	300	1,415	6	
		서울대2	284	1,385	6	
부산	대도시	부산대1	299	1,039	6	
대구	대도시	대구대1	214	793	5	
인천	대도시	인천대1	232	916	6	
광주	대도시	광주대1	146	529	5	
대전	대도시	대전대1	139	520	5	
울산	대도시	울산대1	118	373	4	
경기	대도시	경기대1	288	1,138	4	
		중소도시	경기중1	273	739	3
			경기중2	229	939	4
			경기중3	351	1,036	4
강원	중소도시	강원중1	186	389	3	
	군지역	강원군1	200	201	1	
충북	중소도시	충북중1	123	341	3	
	군지역	충북군1	150	219	2	
충남	중소도시	충남중1	261	566	4	
	군지역	충남군1	181	230	1	
전북	중소도시	전북중1	276	598	4	
	군지역	전북군1	137	154	1	
전남	중소도시	전남중1	191	437	3	
	군지역	전남군1	305	371	2	
경북	중소도시	경북중1	363	749	5	
	군지역	경북군1	177	199	1	
경남	중소도시	경남중1	343	992	5	
	군지역	경남군1	165	204	1	
제주	중소도시	제주중1	115	235	4	

## 2) 초4 패널

- 표집틀 : 교육부 전국 학교통계
- 표집방법 : 층화다단계집락표집(stratified multi-stage cluster sampling)
- 모집단 : 2010년 4월 기준 전국의 초등학교 4학년 재학생

<표 3> 초4 패널 모집단 구성(2009년)

지역		학교 수	학급 수	학생 수
계		6,117	20,760	595,565
서울		584	3,429	102,001
광역시	부산광역시	299	1,308	36,040
	대구광역시	215	1,032	30,945
	인천광역시	233	1,171	33,179
	광주광역시	139	659	19,437
	대전광역시	147	684	20,820
	울산광역시	119	497	15,179
시군부	경기도	1,144	4,702	150,891
	강원도	396	735	18,057
	충청북도	276	701	18,838
	충청남도	446	948	24,424
	전라북도	419	912	23,098
	전라남도	517	998	22,940
	경상북도	545	1,203	30,549
	경상남도	523	1,505	41,342
	제주도	115	276	7,825

- 표집절차
  - 16개 시·도를 시지역과 군지역으로 구분한 후 이들의 지역적 근접성을 기준으로 총 27개의 집락으로 나눔
  - 서울시는 강남과 강북으로, 경기도는 대도시와 중소도시(1,2,3)로, 기타 도지역은 중소도시와 군지역으로 집락화함(광역시는 한 개의 층으로 집락화)
  - 초4 패널 목표 표본수를 2,200명으로 정하고, 각 시·도별 층별 배분은 학급수를 기준으로 비례배분함. 이에 따라 총 95개의 표본 학교 수 결정. 학교당 평균 80%의 유효응답률을 가정하여 전체 표본수 산정
  - 각 시·도별 최소 표본 학교 수를 우선 할당한 후에 학생 수를 기준으로 비례 배분법으로 각 시·도별 할당. 27개 각 집락별로 각각 할당된 표본학교 추출
  - 학교별 초등학교 4학년 학생수를 반영한 규모비례확률표집법(PPS) 적용. 시·도별, 도시규모별 층 내에서 학급 수를 기준으로 확률비례추출법에 따라 표본학교 선정(단, 2개 학급 이상이고 학생수가 50명 이상인 학교만 선정. 만약 위와 같은 조건을 충족하지 않을 경우에는 바로 인접한 학교를 표본학교로 대체)

- 추출된 학교의 초등학교 4학년 학급수에서 난수를 발생시켜 1개 학급 무작위 추출(사전 학교 섭외과정에서 우열반 등 특별편성이 되어 있는 학급은 제외)
- 추출된 학교 측의 조사협조 거절로 학교방문 조사가 불가능할 경우, 학교 리스트에서 추출된 학교의 아래, 위 학교 순으로 대체(학교거절을 고려해 사전에 2순위 대체 학교 추출. 2순위 학교에서도 조사가 불가능할 경우 3순위 학교로 대체)

<표 4> 초4 패널 시·도별/도시규모별 표본학교 수

지역	규모	층구분	모집단		할당된 표본학교수	
			학교 수	학급 수		
	계		6,117	20,760	95	
서울	대도시	서울대1	300	1,723	6	
		서울대2	284	1,706	5	
부산	대도시	부산대1	299	1,308	6	
대구	대도시	대구대1	215	1,032	5	
인천	대도시	인천대1	233	1,171	5	
광주	대도시	광주대1	147	684	5	
대전	대도시	대전대1	139	659	5	
울산	대도시	울산대1	119	497	4	
경기	대도시	경기대1	287	1,451	4	
		중소도시	경기중1	273	897	3
			경기중2	230	1,175	4
			경기중3	354	1,179	4
강원	중소도시	강원중1	188	477	3	
	군지역	강원군1	208	258	1	
충북	중소도시	충북중1	124	436	3	
	군지역	충북군1	152	265	2	
충남	중소도시	충남중1	264	674	4	
	군지역	충남군1	182	274	1	
전북	중소도시	전북중1	281	728	4	
	군지역	전북군1	138	184	1	
전남	중소도시	전남중1	196	539	3	
	군지역	전남군1	321	459	2	
경북	중소도시	경북중1	364	955	4	
	군지역	경북군1	181	248	1	
경남	중소도시	경남중1	355	1,261	5	
	군지역	경남군1	168	244	1	
제주	중소도시	제주중1	115	276	4	

### 3) 중1 패널

- 표집틀 : 교육부 전국 학교통계
- 표집방법 : 층화다단계집락표집(stratified multi-stage cluster sampling)
- 모집단 : 2010년 4월 기준 전국의 중학교 1학년 재학생

<표 5> 중1 패널 모집단 구성(2009년)

지역	남녀공학				남학교			여학교		
	학교수	학급수	남학생	여학생	학교수	학급수	남학생	학교수	학급수	여학생
계	2,320	13,789	258,223	230,953	376	2,847	101,267	421	2,527	88,907
서울	273	2,679	50,498	44,658	46	372	12,894	49	345	11,743
부산	95	775	14,168	12,462	33	299	10,247	42	258	8,751
대구	100	829	16,239	15,033	9	104	4,008	13	65	2,349
인천	66	485	9,887	9,446	27	259	10,528	31	230	9,141
광주	65	495	9,516	9,611	7	71	2,856	9	49	1,965
대전	64	505	9,778	8,793	10	65	2,289	12	55	1,850
울산	50	419	8,140	6,746	5	51	1,722	5	46	1,582
경기	476	3,746	76,013	68,744	35	264	10,142	33	244	9,086
강원	107	267	3,920	3,693	27	189	6,490	30	168	5,738
충북	96	421	7,294	6,394	16	117	4,050	17	107	3,710
충남	139	479	7,747	6,779	25	177	6,137	29	169	5,800
전북	165	635	10,935	9,120	19	90	2,851	20	103	3,415
전남	213	546	8,169	7,454	25	159	5,236	28	146	4,828
경북	200	576	9,068	7,858	44	273	9,088	53	224	7,573
경남	183	808	14,548	12,087	41	302	10,494	43	266	9,314
제주	28	124	2,303	2,075	7	55	2,235	7	52	2,062

- 표집절차
  - 16개 시·도를 시지역과 군지역으로 구분한 후 이들의 지역적 근접성을 기준으로 총 27개의 집락으로 나눔
  - 서울시는 강남과 강북으로, 경기도는 대도시와 중소도시(1,2,3)로, 기타 도지역은 중소도시와 군지역으로 집락화함(광역시에는 한 개의 층으로 집락화)
  - 중1 패널 목표 표본수를 2,200명으로 정하고, 각 시·도별 층별 배분은 학급수를 기준으로 비례배분함. 이에 따라 총 78개의 표본 학교 수 결정. 학교당 평균 80%의 유효응답률을 가정하여 전체 표본수 산정. 특히, 중1 패널의 경우 남녀공학과 여학교 및 남자학교로 구분하여 할당한 후에 표본학교 추출
  - 각 시·도별 최소 표본 학교 수를 우선 할당한 후에 학생 수를 기준으로 비례 배분법으로 각 시·도별 할당. 27개 각 집락별로 각각 할당된 표본학교 추출
  - 학교별 중학교 1학년 학생수를 반영한 규모비례확률표집법(PPS) 적용. 시·도별, 도시규모별, 학제별(남녀공학, 여학교와 남학교) 층 내에서 학급 수를 기준으로 확률비례추출법에 따라 표본학교 선정

(단, 2개 학급 이상이고 학생수가 50명 이상인 학교만 선정. 만약 위와 같은 조건을 충족하지 않을 경우에는 바로 인접한 학교를 표본학교로 대체)

- 추출된 학교의 중학교 1학년 학급수에서 난수를 발생시켜 1개 학급 무작위 추출(사전 학교 섭외과정에서 우열반 등 특별편성이 되어 있는 학급은 제외)
- 추출된 학교 측의 조사협조 거절로 학교방문 조사가 불가능할 경우, 학교 리스트에서 추출된 학교의 아래, 위 학교 순으로 대체(학교거절을 고려해 사전에 2순위 대체 학교 추출. 2순위 학교에서도 조사가 불가능할 경우 3순위 학교로 대체)

<표 6> 중1 패널 시·도별/도시규모별 표본학교 수

지역	규모	총구분	모집단 학교 수			할당된 표본학교 수				
			남녀공학	남학교	여학교	남녀공학	남학교	여학교	계	
	계		2,320	421	376	57	11	10	78	
서울	대도시	서울대1	127	29	31	3	1	0	4	
		서울대2	146	20	15	3	0	1	4	
부산	대도시	부산대1	95	42	33	3	1	1	5	
대구	대도시	대구대1	100	13	9	4	0	0	4	
인천	대도시	인천대1	66	31	27	3	1	1	5	
광주	대도시	광주대1	65	9	7	4	0	0	4	
대전	대도시	대전대1	64	12	10	3	1	0	4	
울산	대도시	울산대1	50	5	5	4	0	0	4	
경기	대도시	경기대1	130	10	14	2	1	0	3	
		중소도시	경기중1	110	7	7	2	0	0	2
			경기중2	110	4	5	3	0	0	3
			경기중3	126	12	9	1	0	1	2
강원	중소도시	강원중1	44	21	18	1	1	1	3	
	군지역	강원군1	63	9	9	1	0	0	1	
충북	중소도시	충북중1	42	10	10	2	1	0	3	
	군지역	충북군1	54	7	6	1	0	0	1	
충남	중소도시	충남중1	77	18	15	2	0	1	3	
	군지역	충남군1	62	11	10	1	0	0	1	
전북	중소도시	전북중1	98	13	14	2	1	0	3	
	군지역	전북군1	67	7	5	1	0	0	1	
전남	중소도시	전남중1	68	14	13	2	0	1	3	
	군지역	전남군1	145	14	12	1	0	0	1	
경북	중소도시	경북중1	116	38	31	2	1	1	4	
	군지역	경북군1	84	15	13	1	0	0	1	
경남	중소도시	경남중1	114	32	30	2	1	1	4	
	군지역	경남군1	69	11	11	1	0	0	1	
제주	중소도시	제주중1	28	7	7	2	1	1	4	

### 3. KCYPS 원표본

이상의 표집과정을 거쳐 2010년 제1차 조사를 완료한 표본은 다음의 표와 같다. 초1 패널이 2,342명, 초4 패널이 2,378명, 중1 패널이 2,351명으로서, 전체 원표본은 7,071명이다.

<표 7> KCYPS의 원표본

	계		초1 패널		초4 패널		중1 패널	
	표본수	학교수	표본수	학교수	표본수	학교수	표본수	학교수
전체	7,071	271	2,342	98	2,378	95	2,351	78
서울	766	31	258	12	274	11	234	8
광역시								
부산광역시	402	17	131	6	138	6	133	5
대구광역시	417	14	121	5	141	5	155	4
인천광역시	405	16	125	6	117	5	163	5
광주광역시	338	14	124	5	119	5	95	4
대전광역시	339	14	118	5	114	5	107	4
울산광역시	327	12	109	4	103	4	115	4
시군부								
경기도	1,130	40	394	15	390	15	346	10
강원도	316	12	92	4	110	4	114	4
충청북도	394	14	130	5	138	5	126	4
충청남도	330	14	114	5	122	5	94	4
전라북도	365	14	116	5	134	5	115	4
전라남도	344	14	122	5	107	5	115	4
경상북도	439	16	151	6	129	5	159	5
경상남도	434	17	142	6	147	6	145	5
제주도	325	12	95	4	95	4	135	4



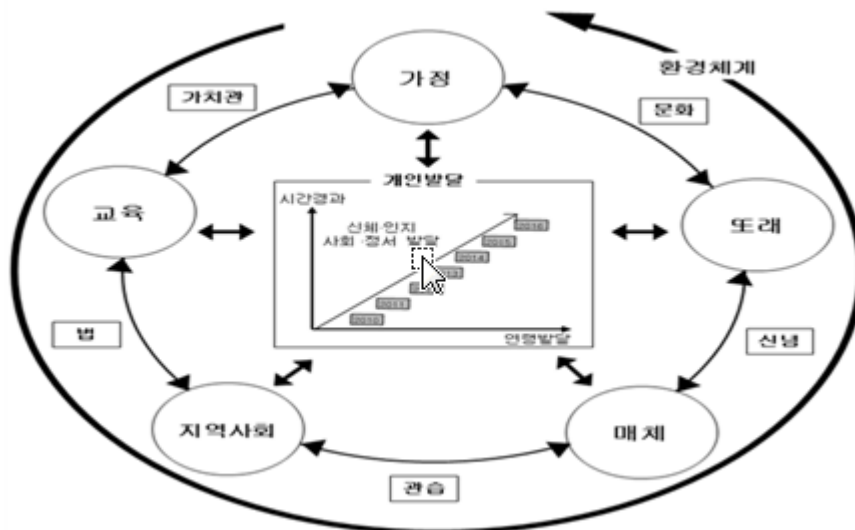
## IV. 조사내용

### 1. 조사영역 · 내용 구성

KCYPS는 아동·청소년 성장과 발달의 다양한 양상을 시간의 연속선상에서 종합적으로 파악하는 것을 목적으로 하고 있다. 청소년을 대상으로 한 기존의 패널조사가 대부분 10대 중후반기 청소년들의 교육과 진로 문제, 경제활동의 측면에 초점을 두고 있는 반면, KCYPS는 초등학생을 조사대상에 포함하여 연령층을 확대하였고, 신체발달과 심리/정서/행동 특성 변화 등을 다양한 주제를 다루고 있다. 이에 따라 다른 조사와 비교하여 조사영역이 포괄적이고 단일 조사로서는 조사문항이 상대적으로 많은 점이 특징적이다.

KCYPS의 조사내용은 인간은 자신을 둘러싼 복잡한 환경체계 속에서 체계와의 상호작용을 통해 성장한다는 브론펜브레너(Bronfenbrenner, 1979)의 생태학적 관점에 기초하여 구성되었다(김지경, 백혜정, 임희진, 이계오, 2010). 연구 1차년도에 설정한 KCYPS의 조사내용 구성의 개념모형은 [그림 II-2]와 같다.

[그림 3] KCYPS 조사영역 기본 개념도



조사영역은 크게 개인발달과 환경발달의 두 영역으로 구분되는데, 개인발달 영역은 아동·청소년이 자신을 둘러싼 주변환경의 영향을 받으면서 성장·발달해가는 양상을 파악하기 위한 것으로서, 신체 발달, 지적 발달, 사회·정서 발달의 세부 영역으로 구성되어 있다. 발달환경 영역은 아동·청소년의 사회화과정

에 중요한 영향을 미치는 다양한 환경요인으로서, 가정, 또래, 교육, 지역사회, 매체의 5가지 하위 영역으로 이루어졌다.

KCYPS의 조사영역과 내용은 이와 같은 연구 1차년도에 조사설계를 바탕으로 매년 청소년의 성장단계별 특성, 청소년을 둘러싼 사회환경 변화 등을 반영하여 부분적인 수정 보완 작업을 진행해 왔다. 2015년 제6차 조사의 내용을 반영한 전체 조사내용의 구성은 다음 <표 8>과 같다.

KCYPS의 조사내용은 대영역, 중영역, 소영역과 조사항목의 4단계이며, 이중 대영역은 개인발달과 발달환경, 배경변인 및 패널관리의 네 영역으로 구분된다.

개인발달 영역은 아동·청소년이 자신을 둘러싼 주변환경의 영향을 받으면서 성장·발달해가는 양상을 파악하기 위한 것으로서, 신체 발달, 지적 발달, 사회·정서적 발달, 진로계획, 비행, 생활시간의 6가지 중영역으로 구성된다. 연구 3차년도까지 비행은 사회·정서 발달의 하위영역으로, 생활시간은 개별 문항별로 개인발달과 발달환경 영역으로 나누어 분류되었으나, 두 영역 모두 행동(행위) 차원의 성장·발달 지표가 될 수 있음을 고려하여 4차년도부터는 개인발달의 독립된 영역으로 설정하였다. 아울러 3개 패널 모두 진학을 앞둔 시점인 6차년도에 진로계획 영역이 새롭게 추가되었다.

발달환경 영역은 아동·청소년의 사회화과정에 중요한 영향을 미치는 다양한 환경요인으로서, 가정환경, 친구관계, 교육환경, 지역사회환경, 매체환경, 활동·문화환경의 5가지 중영역으로 구성된다. 활동·문화환경은 연구 3차년도까지는 지역사회환경으로 분류되었는데, 아동·청소년 활동이 지역사회에만 한정되지 않고 가정, 학교 등 다양한 영역에서 전개된다는 점에서 4차년도부터 별도의 독립된 영역으로 분류하였다.

이밖에 배경변인 영역의 조사내용은 성별, 교급, 보호자 직업 등 데이터 분석에 필요한 청소년과 보호자의 기본 속성 관련 항목들로 구성되어 있으며, 패널관리 영역은 소속학교명, 조사성공 여부, 횡/종단면가중치 등 데이터 관리 관련 항목들을 포함하고 있다. 연구 3차년도까지 배경변인과 데이터관리 영역 항목들은 다른 영역에 포함되었는데, 내용상 다른 조사문항과의 구분이 필요한 것으로 판단되어 4차년도부터 각각 독립된 별개의 영역으로 설정하였다.

<표 8> KCYPS 조사문항 체계

대영역	조사영역		조사항목
	중영역	소영역	
개인발달	신체발달	체격	키, 몸무게
		신체변화	[남자] 몽정 시작 시기, [여자] 초경 시작 시기
		건강	건강상태 평가, 체육시간 중 운동시간
		건강문제	건강문제 유무, 지장받는 정도
	지적발달	성적	과목별 성적평가 여부, 과목별 점수, 과목별 성적 주관적 평가 등
		학습습관	성취가치, 숙달목적 지향성, 행동통제, 학업시간 관리
	사회정서 발달	정서문제	주의집중, 공격성, 신체증상, 사회적 위축, 우울
		자아인식	자아존중감, 자아탄력성, 자아정체감, 진로정체감
		삶의 목표·만족도	삶의 목표, 삶의 만족도
	진로계획	진로계획	장래 희망 교육수준, 장래 직업 결정 여부, 고교 졸업 후 진로계획
		진로관	진로관련 대화상대/빈도, 직업선택 요인, 대학 진학/취업 이유
	비행	비행	연간 경험 유무, 연간 경험 횟수
		비행 피해	연간 경험 유무, 연간 경험 횟수
		비행 친구	연간 비행경험 친구수
	생활시간	수면 시간	취침 시간, 기상 시간
		학업관련 시간	학원/과외 시간, 학교숙제 시간, 학원/과외숙제 시간 등
		여가시간	독서 시간, 컴퓨터/게임기 오락시간, TV/비디오 시청시간 등
	발달환경	가정환경	가족구성
양육방식			방임, 학대, 감독, 애정, 비밀관성, 과잉기대, 과잉간섭, 합리적 설명
보호자 부재			방과후 보호자 부재 일수/ 부재 시간
용돈			월평균 용돈, 용돈 주는 사람
친구관계		또래애착	의사소통, 신뢰, 소외
		부모-친구관계	부모님의 친구 인지도/친구를 만난 경험/친구 선호도
		이성친구	이성친구 유무
교육환경		전학·결석 경험	연간 전학경험 유무, 연간 결석경험 유무, 연간 결석일수
		학교생활 적응	학습활동, 학교규칙, 교우관계, 교사관계
		방과후교육	연간 경험 유무, 방과후교육 받은 이유/받지 않은 이유, 월평균 교육비, 방과후 교육비 가계부담 정도
지역사회 환경		지역사회	지역사회 인식, 공동체의식
		다문화	다문화 수용도
매체환경		컴퓨터	컴퓨터 사용여부, 사용시간, 사용장소, 이용빈도
		휴대전화	휴대전화 보유여부/종류, 이용빈도, 의존도
		사이버비행	연간 경험여부, 연간 경험횟수
		성인용 매체	이용빈도, 몰입도
활동·문화 환경		체험활동	연간 참여경험 유무, 연간 참여시간, 만족도, 희망 체험활동 등
		동아리활동	연간 참여경험 유무, 연간 참여시간, 만족도
	여행·문화활동	가족/단체 여행 연간 경험횟수, 문화활동 연간 경험횟수	
	해외방문 경험	해외방문 연간 경험유무, 해외 교육·연수 연간 경험유무 등	
	종교활동	종교 유무/종류, 종교시설 방문빈도	
	팬덤활동	선호 연예인/운동선수 유무, 팬덤활동 빈도	
	배경변인	성별, 지역, 교급, 학년, 보호자 학력/직업, 가정경제수준 등	
	패널관리	D, 패널 구분, 조사참여 여부, 가중치 등	

\* 중1 패널 제7차년도 조사의 “고졸 후 진로현황” 영역의 문항은 표에서 생략하였음.

## 2. 조사문항 구성

KCYPS의 조사문항은 관련 학문분과에서 널리 활용되어 신뢰도와 타당도가 검증된 문항 또는 척도가 있을 경우 이를 우선적으로 활용하는 것을 원칙으로 하였다. 단, 조사대상 집단의 연령을 고려하여 난이도가 높거나 문항 수가 과도한 척도는 전문가 자문을 통해 문장 수정, 문항 수 조정 작업 등을 거친 후에 예비조사를 통해 최종 문항을 확정하였다<sup>5</sup>. 선행연구의 척도를 활용한 KCYPS 조사척도의 출처는 KCYPS 코드북에 제시되어 있다.

조사문항의 선정이 완료된 이후에는 조사대상 연령층을 고려한 추가적인 수정·보완 작업이 이루어졌다. KCYPS는 상이한 연령대의 세 집단을 조사대상으로 하여 동일한 문항에 대해서도 조사대상 집단별로 이해수준에 있어서 차이가 날 수 있다. 따라서 초1 패널의 1~3차년도 조사(초1에서 초3 학령기)에서 초4 또는 중1 패널과 동일한 문항을 조사해야 할 경우, 해당 문항의 난이도 검토를 통해 이해하기 쉬운 문장으로 바꾸거나, 질문방식을 변경하는 등의 작업을 진행하였다<sup>6</sup>.

한편, KCYPS의 조사표는 청소년용과 보호자용으로 구분된다. 본 연구의 주된 조사대상이 청소년이기 때문에 모든 문항에 대하여 청소년들에게 직접 질문하는 것을 원칙으로 하되, 연령 특성상 청소년이 정확하게 응답하기 어려운 문항과 부모의 직업, 소득 등 구체적인 정보가 필요한 문항은 보호자 조사표에 반영하였다. 그러나, 조사차수가 거듭되어 청소년이 직접 응답 가능한 연령대에 도달하면 청소년 조사문항으로 전환하여 보호자 조사문항은 점진적으로 줄여나가는 것을 원칙으로 하였다.

---

<sup>5</sup> 예컨대 Rosenberg(1965)의 자아존중감 척도의 경우, 조사시점 현재 초4 학령 이상은 10개 조사문항을 모두 사용한 반면, 초3 학령(초1 패널 3차년도 조사)에 대해서는 5개 문항의 간이척도를 사용하였다.

<sup>6</sup> 학교생활 적응 조사항목을 예시하면, 초4 학령 이상은 “학교생활이 재미있다”는 진술문에 대하여 “매우 그렇다”~ “전혀 그렇지 않다”의 4점 척도로 응답하게 한 반면, 초1~3 학령기에는 “학교생활이 재미있나요?”로 질문하고 “아주 재미있다”~“전혀 재미없다”의 서술형 응답을 제시하였다.

### 3. 문항별 조사주기

KCYPS는 특정 영역과 주제에 대한 조사가 아니라 아동·청소년 생활·행동·의식의 다양한 영역을 포괄하고 있어 전체적으로 다른 조사에 비해 문항 수가 많은 편이다. 따라서 모든 문항을 매년 조사하기 어렵기 때문에 문항별 특성과 중요도 등을 고려하여 조사 간격(주기)을 다르게 설정하였다.

문항별 조사주기는 크게 ① 매년 조사와 ② 2~3년 간격의 주기조사, ③ 특정 시점 조사 등으로 구분되는데, 조사주기 설정의 기본원칙은 다음과 같다.

첫째, 가변성이 높은 항목(속성), 즉 시간의 경과에 따라 변화 가능성이 높은 항목은 매년 조사한다. 주로 행동 차원의 조사항목으로서 생활시간, 학교생활 적응, 비행 경험 유무 및 횟수, 컴퓨터/휴대전화 이용빈도, 체험활동/동아리 활동, 해외방문 경험과 더불어 체격, 건강상태 평가, 학업성적 등이 여기에 해당한다.

둘째, 가변성이 낮고 상대적으로 안정적인 항목(속성)에 대한 조사는 2년 또는 3년을 주기로 조사한다. 정서문제, 자아존중감, 자아탄력성, 자아정체감, 학습습관, 부모의 양육방식, 삶의 목표 등의 조사항목이 여기에 해당한다. 단, 부모의 양육방식 중 방임과 학대, 삶의 만족도 등의 조사항목은 학술적·정책적 활용도가 높은 주요 변인임을 고려하여 매년 조사항목으로 결정하였다.

셋째, 주기적 조사항목의 조사시점은 해당 항목의 내용 검토를 통해 ① 조사 차수 또는 ② 조사 학령을 기준으로 설정한다. 조사 차수를 기준으로 한 조사는 동일한 시점에 세 패널을 동시에 조사하는 방식이며, 학령을 기준으로 한 조사는 각 패널이 특정 학령(예: 초4, 중1, 고1 등)에 도달했을 때 조사하는 방식을 의미한다. 예컨대, 진로정체감은 10대 중반기에 형성되는 것이 일반적임을 고려하여 중학교 3학년이 되는 시기부터 격년으로 조사하는 것으로 결정하였다.

넷째, 모든 문항은 시계열 비교가 가능하도록 조사 기간 중 한 패널당 2회 이상 조사하고, 집단간 비교가 가능하도록 세 패널 모두를 대상으로 조사하는 것을 원칙으로 한다.

## V. 조사실사

### 1. 제1차년도 조사

- 조사기간 : 2010년 10월 ~ 11월
- 자료수집 방법
  - － 청소년 : 학교방문을 통한 집단면접조사. 복수(4인)의 면접원을 투입하여 그룹별로 집단면접조사관리 (self-reported + 면접조사원 응답관리 + 면접조사원 editing + 조사사례품 지급의 방식으로 수행)
  - － 보호자 : 유치조사
- 조사실사 절차
  - － 학교 섭외
    - 추출된 학교에 협조 공문, 안내자료 및 한국청소년정책연구원장 개인서신 발송
    - 전화를 이용해 학교 담당자(교장, 교감선생님, 교무부장 등)와 통화
    - 공문 수신 확인 후 조사 취지 설명, 협조 요청
    - 추출된 학급의 방문 조사 가능일 파악, 방문 약속(조사 학급 선정 및 담임교사 연락처 확보)
    - 전화 접촉 실패시 담당 수퍼바이저, 담당 연구원이 전화 또는 직접 방문하여 설득
    - 최종 거절시 다른 학교로 대체
  - － 청소년 조사
    - 조사대상 학급 담임 선생님께 전화를 걸어 사전에 정한 약속시간을 재확인 후 학교 방문
    - 해당 학급 청소년 전원에게 청소년조사 진행
    - 면접원(4인 1조) 조사 보조 및 현장 에디팅
    - 보호자에게 전달할 설문지와 안내자료 배부 후 보호자 조사 협조 설명
    - 응답사례품 지급
  - － 보호자 조사
    - 학생을 통해 전달받은 설문지를 가정에서 작성 후 봉투에 밀봉하여 학생편에 설문지 전달
    - 학급 담임교사가 설문지 수거
    - 면접원이 학교를 재방문하여 담임교사로부터 조사표 수합
- 실사 담당기관 : (주)밀워드브라운 미디어리서치

## 2. 제2~7차년도 조사

- 조사기간 : 매년(2011~2016) 10월 ~ 12월
- 자료수집 방법
  - 청소년 : 개인별 추적을 통한 개별면접조사. 이탈 변수의 측정을 위해 self-reported + 면접조사원 응답관리 + 면접조사원 editing + 조사사례품 지급의 방식으로 수행
  - 보호자(부모) : 전화면접조사 (전화면접조사에 적합하게 구조화된 조사표로 SES 등 측정)
- 조사실사 절차
  - 조사예고 및 면접약속
    - 패널 가구에 조사예고문 및 가정통신문 발송
    - 패널 부모(보호자) 및 패널 청소년과의 전화 접촉 후 협조 요청
    - 패널 청소년이 편한 장소와 시간으로 면접약속
  - 청소년 개별 면접조사
    - 약속된 시간에 약속된 장소를 방문, 조사진행
    - 본인 여부 확인, 음료나 간식 등 제공
    - 조사취지 및 조사표 작성 방법에 대한 간략한 설명
    - 조사표 전달 후 패널이 자기기입식으로 조사표 직접 작성
  - 면접조사원 응답관리
    - 면접조사원 현장 editing
    - 오류가 있을 경우 조사표 수정 요청
  - 응답사례품 지급 및 패널 관리
    - 약속을 지키지 않은 패널은 다시 전화해 재약속 후 조사
    - 연락을 회피하거나 참여거절의사를 밝힌 이탈위험 패널은 특별관리
  - 부모 전화조사
    - 조사가 완료된 학생의 부모(보호자)에게 전화
    - 가정통신문 수령 및 숙지 여부 확인
    - 조사에 대한 간단한 설명 후 조사 협조 요청
    - 부모용 질문에 대한 전화조사 진행
  - 감사 전화
    - 조사 완료된 모든 학생의 가정에 전화 통화
    - 참여에 대한 감사 인사 및 지속적인 협조 부탁
- 실사 담당기관 : (주)밀워드브라운 미디어리서치

## Ⅵ. 데이터 구축

### 1. 데이터 구축 현황

패널조사는 동일한 조사대상에 대하여 주기를 두고 반복조사를 하는 특성 때문에 추적조사 과정에서 조사대상자의 사망, 질병, 조사 거부 등으로 인한 표본 이탈(sample attrition)이 발생하게 된다. 표본 이탈은 조사 차수가 거듭될수록 증가하게 되는데, 표본 이탈이 과다할 경우 조사결과의 대표성, 신뢰성에 문제가 야기될 수도 있다. 이에 따라 조사 첫 해에 선정된 표본을 어느 정도 유지하고 있는가를 나타내는 원표본 유지율은 해당 조사가 얼마나 성실하게 수행되었는지, 조사결과가 얼마나 신뢰할만한지를 판단하는 지표로 간주되기도 한다.

KCYPS는 2010년에 확정된 원표본 7,071명을 대상으로 지금까지 총 7차례에서 걸쳐 조사가 실시되었다. 조사 마지막 해인 제7차 조사에서 조사에 성공한 표본은 5,862명으로 최종 원표본유지율은 82.9%이다. 패널별로는 초1 패널의 원표본유지율이 85.5%로서 가장 높고, 초4 패널이 83.2%, 중1 패널이 80.0%로 나타났다.

<표 9> KCYPS 원표본 유지율(2010~2016)

		전체	초1 패널	초4 패널	중1 패널
제1차 조사	원표본	7,071	2,342	2,378	2,351
제2차 조사	조사성공 표본 (원표본 유지율)	6,808 (96.3%)	2,264 (96.7%)	2,264 (95.2%)	2,280 (97.0%)
제3차 조사	조사성공 표본 (원표본 유지율)	6,678 (94.4%)	2,200 (93.9%)	2,219 (93.3%)	2,259 (96.1%)
제4차 조사	조사성공 표본 (원표본 유지율)	6,319 (89.4%)	2,119 (90.5%)	2,092 (88.0%)	2,108 (89.7%)
제5차 조사	조사성공 표본 (원표본 유지율)	6,271 (88.7%)	2,110 (90.1%)	2,070 (87.1%)	2,091 (88.9%)
제6차 조사	조사성공 표본 (원표본 유지율)	6,184 (87.5%)	2,067 (88.3%)	2,061 (86.7%)	2,056 (87.5%)
제7차 조사	조사성공 표본 (원표본 유지율)	5,862 (82.9%)	2,002 (85.5%)	1,979 (83.2%)	1,881 (80.0%)

KCYPS의 지역별 원표본 유지율은 다음 <표 10> ~ <표 11>과 같다.



<표 10> 초1 패널 조사 데이터 구축 현황

	1차조사	2차 조사(2011)			3차 조사(2012)			4차 조사(2013)			5차 조사(2014)			6차 조사(2015)			7차 조사(2016)		
	(2010) 원표본	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율
전체	2,342	2,342	2,264	96.7%	2,342	2,200	93.9%	2,342	2,119	90.5%	2,342	2,110	90.1%	2,342	2,067	88.3%	2,342	2,002	85.5%
서울	258	251	242	96.4%	253	230	90.9%	251	215	85.7%	248	211	85.1%	248	199	80.2%	242	178	73.6%
부산	131	132	127	96.2%	133	130	97.7%	131	123	93.9%	134	129	96.3%	131	124	94.7%	131	124	94.7%
대구	121	122	119	97.5%	122	118	96.7%	121	113	93.4%	120	114	95.0%	121	111	91.7%	121	107	88.4%
인천	125	129	127	98.4%	132	125	94.7%	130	119	91.5%	132	119	90.2%	137	115	83.9%	133	110	82.7%
광주	124	126	120	95.2%	127	114	89.8%	124	108	87.1%	124	104	83.9%	123	101	82.1%	122	101	82.8%
대전	118	117	116	99.1%	120	118	98.3%	121	118	97.5%	126	122	96.8%	122	117	95.9%	118	109	92.4%
울산	109	109	103	94.5%	109	101	92.7%	109	100	91.7%	128	114	89.1%	107	92	86.0%	107	89	83.2%
경기	394	397	376	94.7%	393	349	88.8%	397	331	83.4%	374	303	81.0%	402	316	78.6%	411	318	77.4%
강원	92	92	88	95.7%	92	83	90.2%	93	83	89.2%	107	95	88.8%	90	79	87.8%	92	74	80.4%
충북	130	129	123	95.3%	127	124	97.6%	126	118	93.7%	113	109	96.5%	121	117	96.7%	120	111	92.5%
충남*	114	115	111	96.5%	115	110	95.7%	116	104	89.7%	109	103	94.5%	115	109	94.8%	122	114	93.4%
전북	116	116	113	97.4%	116	113	97.4%	116	108	93.1%	124	117	94.4%	117	110	94.0%	115	107	93.0%
전남	122	121	121	100.0%	120	116	96.7%	121	115	95.0%	130	123	94.6%	121	115	95.0%	121	111	91.7%
경북	151	150	148	98.7%	149	145	97.3%	149	141	94.6%	150	143	95.3%	147	138	93.9%	149	140	94.0%
경남	142	141	138	97.9%	139	132	95.0%	141	130	92.2%	128	112	87.5%	145	131	90.3%	142	129	90.8%
제주	95	95	92	96.8%	95	92	96.8%	96	93	96.9%	95	92	96.8%	95	93	97.9%	96	80	83.3%

\* 세종시는 충남에 포함

<표 11> 초4 패널 조사 데이터 구축 현황

	1차조사	2차 조사(2011)			3차 조사(2012)			4차 조사(2013)			5차 조사(2014)			6차 조사(2015)			7차 조사(2016)		
	(2010) 원표본	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율
전체	2,378	2,378	2,264	95.2%	2,378	2,219	93.3%	2,378	2,092	88.0%	2,378	2,070	87.0%	2,378	2,061	86.7%	2,378	1,979	83.2%
서울	274	275	261	94.9%	275	251	91.3%	273	240	87.9%	276	234	84.8%	282	241	85.5%	275	222	80.7%
부산	138	139	131	94.2%	143	138	96.5%	143	122	86.0%	142	124	87.3%	146	124	84.9%	143	127	88.8%
대구	141	141	138	97.9%	139	137	98.6%	134	126	94.0%	136	123	90.4%	133	122	91.7%	134	119	88.8%
인천	117	117	107	91.5%	114	95	83.3%	118	88	74.6%	119	89	74.8%	120	92	76.7%	119	85	71.4%
광주	119	121	114	94.2%	120	107	89.2%	121	103	85.1%	120	105	87.5%	121	105	86.8%	123	106	86.2%
대전	114	114	112	98.2%	114	110	96.5%	118	111	94.1%	117	109	93.2%	119	109	91.6%	117	103	88.0%
울산	103	103	98	95.1%	103	100	97.1%	102	90	88.2%	102	86	84.3%	99	87	87.9%	101	80	79.2%
경기	390	389	357	91.8%	393	354	90.1%	395	318	80.5%	400	335	83.8%	393	322	81.9%	402	313	77.9%
강원	110	111	107	96.4%	109	102	93.6%	108	102	94.4%	105	98	93.3%	105	96	91.4%	104	94	90.4%
충북	138	138	138	100.0%	139	138	99.3%	135	130	96.3%	134	129	96.3%	133	128	96.2%	137	127	92.7%
충남*	122	122	121	99.2%	129	127	98.4%	130	121	93.1%	126	118	93.7%	126	121	96.0%	130	122	93.8%
전북	134	134	130	97.0%	132	126	95.5%	131	121	92.4%	130	121	93.1%	130	119	91.5%	127	112	88.2%
전남	107	105	94	89.5%	104	85	81.7%	104	84	80.8%	106	81	76.4%	105	80	76.2%	104	75	72.1%
경북	129	127	119	93.7%	124	116	93.5%	127	117	92.1%	128	117	91.4%	126	115	91.3%	123	102	82.9%
경남	147	147	144	98.0%	144	139	96.5%	143	131	91.6%	142	126	88.7%	145	130	89.7%	146	126	86.3%
제주	95	95	93	97.9%	96	94	97.9%	96	87	90.6%	95	75	78.9%	95	70	73.7%	93	66	71.0%

\* 세종시는 충남에 포함

<표 12> 중1 패널 조사 데이터 구축 현황

	1차조사	2차 조사(2011)			3차 조사(2012)			4차 조사(2013)			5차 조사(2014)			6차 조사(2015)			7차 조사(2016)		
	(2010) 원표본	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율	원표본	조사 표본	원표본 유지율
전체	2,351	2,351	2,280	97.0%	2,351	2,259	96.1%	2,351	2,108	89.7%	2,351	2,091	88.9%	2,351	2,056	87.5%	2,351	1,881	80.0%
서울	234	238	223	93.7%	236	213	90.3%	233	178	76.4%	238	191	80.3%	240	186	77.5%	249	176	70.7%
부산	133	134	129	96.3%	132	129	97.7%	134	123	91.8%	136	125	91.9%	132	120	90.9%	134	112	83.6%
대구	155	154	153	99.4%	155	153	98.7%	153	145	94.8%	150	139	92.7%	154	143	92.9%	156	126	80.8%
인천	163	163	153	93.9%	161	150	93.2%	164	131	79.9%	162	127	78.4%	168	136	81.0%	168	119	70.8%
광주	95	93	91	97.8%	93	88	94.6%	93	79	84.9%	93	77	82.8%	93	77	82.8%	93	69	74.2%
대전	107	104	100	96.2%	104	100	96.2%	102	93	91.2%	110	97	88.2%	106	92	86.8%	104	86	82.7%
울산	115	115	109	94.8%	114	109	95.6%	116	111	95.7%	110	102	92.7%	116	102	87.9%	114	88	77.2%
경기	346	340	331	97.4%	348	334	96.0%	349	316	90.5%	349	311	89.1%	347	310	89.3%	356	294	82.6%
강원	114	114	111	97.4%	115	110	95.7%	114	102	89.5%	119	105	88.2%	117	96	82.1%	107	72	67.3%
충북	126	127	126	99.2%	129	127	98.4%	130	125	96.2%	125	121	96.8%	124	119	96.0%	125	110	88.0%
충남*	94	96	91	94.8%	92	91	98.9%	96	90	93.8%	99	92	92.9%	95	88	92.6%	97	84	86.6%
전북	115	115	113	98.3%	114	112	98.2%	117	110	94.0%	113	104	92.0%	116	97	83.6%	116	101	87.1%
전남	115	118	115	97.5%	119	110	92.4%	116	103	88.8%	124	110	88.7%	112	95	84.8%	113	94	83.2%
경북	159	161	158	98.1%	160	158	98.8%	159	152	95.6%	154	147	95.5%	157	148	94.3%	159	133	83.6%
경남	145	144	143	99.3%	144	142	98.6%	142	129	90.8%	135	123	91.1%	141	133	94.3%	141	122	86.5%
제주	135	135	134	99.3%	135	133	98.5%	133	121	91.0%	134	120	89.6%	133	114	85.7%	119	95	79.8%

\* 세종시는 충남에 포함

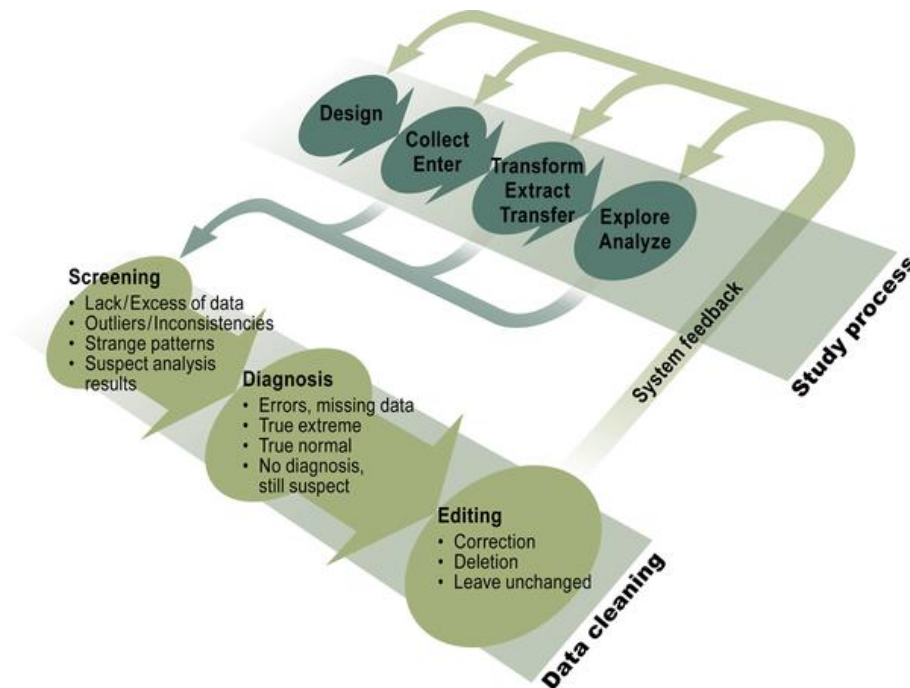
## 2. 데이터 클리닝 개요

데이터 클리닝(data cleaning)은 조사결과의 비표본오차(nonsampling error)를 최소화하기 위한 노력의 일환으로서, 데이터의 입력 에러와 고의적/비고의적 오답, 결측치, 이상치 등에 대한 보정을 통해 데이터의 품질을 향상시키는 작업을 말한다.

일반적으로 원데이터 파일이 작성되기까지는 ① 응답자 또는 조사원에 의한 조사표의 기입, ② 응답이 완료된 조사표의 점검(editing), ③ 응답 데이터의 코드화(coding), ④ 코딩된 데이터 입력(punching) 등의 과정을 거치게 된다. 이 과정에서 여러 가지 오류가 수정되지만, 일부 오류는 잔존하게 되고 때로는 새로운 오류가 덧붙여지기도 한다. 이와 같은 응답·입력상의 오류를 수정·보완하는 것은 정확한 조사결과를 얻기 위해 선행되어야 할 필수적인 작업이라 할 수 있다. 특히 KCYPS의 조사결과는 학술적·정책적인 용도로 널리 활용될 수 있도록 일반에 공개되기 때문에 높은 수준의 데이터 품질을 유지하는 것이 매우 중요한 과제이다.

KCYPS의 데이터 클리닝은 관련 선행연구(Van den Broeck, Cunningham, Eeckels, & Herbst, 2005 등)의 권고사항을 참고하여 각 변인별 데이터 스크리닝(screening) → 진단(diagnosis) → 에디팅(editing)의 순으로 진행하였으며, 횡단면 클리닝과 종단면 클리닝을 병행 실시하였다.

[그림 4] 데이터 클리닝 개념도



자료: Van den Broeck, Cunningham, Eeckels, Herbst (2005, p. 967)

데이터 클리닝의 주된 대상의 하나인 개방형 응답문항의 이상치(outlier)에 대해서는, ① 평균값과 표준편차를 기준( $M \pm 2SD$  또는  $M \pm 3SD$ )으로 의심 사례를 검출한 뒤, ② 조사기관을 통해 해당 사례의 응답자에게 직접 확인하는 것을 원칙으로 하되, 불가능할 경우 ③ 관련된 다른 문항과의 교차분석, 또는 ④ 일반적인 상식에 기초한 허용 가능성 등을 검토하여 판정하였다. 단, 분석방법에 따라 필요할 경우 연구자가 별도의 기준을 설정하여 이상치를 지정할 수 있게 하였고, 이 경우 이상치 설정의 기준을 해당 저술에 명시하여야 한다.

## Ⅶ. 데이터 가중치

KCYPS는 동일 표본에 대하여 여러 차례 조사를 실시하므로 조사된 표본들의 조사년도간의 순수변동은 정확하게 추정할 수 있으나, 패널 효과와 모집단의 시간에 따른 변동에 의한 특성은 추정하기가 쉽지 않다. 조사 시점마다 패널 유지율이 상이하므로 설계 당시에 자체가중설계를 했을지라도 조사 차수 진행되면서 불균등확률표본설계가 되기 때문에 매 조사차수마다 가중치를 계산해야 한다. KCYPS의 패널별 횡단면/종단면 가중치 산출방식을 설명하면 다음과 같다<sup>7</sup>.

### 1. 초1 패널 가중치

#### 1) 1차년도 횡단면 가중치

표본추출확률에 대한 가중치(기본 가중치)

KCYPS의 표본추출방식은 일종의 층화2단집락추출법으로서, 시·도와 도시규모를 층으로 간주할 수 있다. 각 층에서 1차추출단위인 학교는 학생수를 기준으로 확률비례추출법으로 선정하고, 추출된 표본학교에서는 학급을 랜덤하게 선정한 후에 학생을 조사하였다. 따라서 추출된 표본학교에서 학생을 랜덤하게 선정하여 조사하는 2단집락추출법으로 볼 수 있으므로 표본추출률을 다음의 식으로 나타낼 수 있다.

$$f_{hij} = l_h \times \frac{M_{hi}}{\sum_{i=1}^{K_h} M_{hi}} \times \frac{n_{hi}}{M_{hi}} = \frac{l_h \times n_{hi}}{\sum_{i=1}^{K_h} M_{hi}}$$

여기서  $K_h$ 는  $h$ 층에 속한 학교의 수,  $l_h$ 는  $h$ 층에서 추출된 학교의 수,  $M_{hi}$ 는  $h$ 층  $i$ 번째 학교의 학생 수,  $n_{hi}$ 는 추출된 표본학교에서 조사한 학생수를 나타낸다.

표본추출률에 대한 가중치는 추출률의 역수이므로 아래와 같은 식으로 표현할 수 있다.

$$W_{1hij} = \frac{\sum_{i=1}^{K_h} M_{hi}}{l_h \times n_{hi}} = \frac{M_h}{l_h \times n_{hi}}$$

여기서  $M_h$ 는  $h$ 층의 전체 학생수를 나타낸다.

<sup>7</sup> 제1~6차 조사의 가중치 산출방식을 수록하였으며, 7차 조사의 데이터 가중치 산출방식은 추후 업데이트 예정임.

## 표본의 무응답 조정을 위한 가중치

표본으로 선정된 조사단위들 중 응답거부로 조사에 성공하지 못하거나 조사된 내용이 부실하여 무응답처리한 조사 단위에 대한 조정이 무응답 조정의 가중치로서, 다음의 식으로 계산할 수 있다.

$$W_{2hij} = \frac{n_{hi}}{r_{hi}}$$

여기서  $r_{hi}$ 와  $n_{hi}$ 는 각각 완전 응답한 표본크기와 추출된 표본크기이다.

## 사후층화 가중치

표본설계시에는 층화변수로 고려하지 않았으나 표본조사 완료 후에 연구변수에 영향을 미치는 변수들 중에서 정확한 외부보조정보를 확보하게 되면, 외부보조정보 데이터의 구조와 표본조사 데이터의 구조를 분석하여 층으로 구분하고 표본조사의 각 셀에 가중치의 합계를 계산한 결과와 외부보조정보의 층별 합계를 비교하여 가중치를 계산한다.

$$BF_{hij} = \frac{PS_{hij}}{\sum W_{1hii} \cdot W_{2hii}}$$

여기서  $PS_{hii}$ 는 외부보조정보에 의한 셀의 빈도수이고, 분모의 내용은 설계가중치와 무응답 조정을 위한 가중치를 곱한 후에 합계를 계산결과이다. KCYPS에서는 16개 광역시·도×성별 사후층화를 실시하였다. 최종 가중치는 세가지 가중치를 곱하여 아래와 같은 식으로 산출한다.

$$W_{thij} = W_{1hii} \cdot W_{2hii} \cdot BF_{hii}$$

## 2) 2차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,264/2,342=96.7%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과

같이 계산되었다(2011년의 초등학교 2학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

초1 패널 2차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 13> 초1 패널 2차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,264	55	603	210	475,425
남자	1,166	55	603	212	247,424
여자	1,098	57	546	208	228,001

<표 14> 초1 패널 2차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,166	51.5	1,098	48.5	2,264	100.0	247,424	52.0	228,001	48.0	475,425	100.0
서울	131	54.1	111	45.9	242	10.7	42,483	51.7	39,729	48.3	82,212	17.3
부산	66	52.4	60	47.6	126	5.6	14,523	52.4	13,216	47.6	27,739	5.8
대구	60	50.0	60	50.0	120	5.3	12,426	52.5	11,245	47.5	23,671	5.0
인천	68	54.4	57	45.6	125	5.5	13,708	51.5	12,922	48.5	26,630	5.6
광주	65	54.2	55	45.8	120	5.3	8,745	52.8	7,827	47.2	16,572	3.5
대전	56	48.3	60	51.7	116	5.1	8,251	51.7	7,712	48.3	15,963	3.4
울산	52	50.5	51	49.5	103	4.5	6,131	53.7	5,294	46.3	11,425	2.4
경기	188	50.0	188	50.0	376	16.6	65,166	51.8	60,660	48.2	125,826	26.5
강원	47	53.4	41	46.6	88	3.9	7,260	52.0	6,701	48.0	13,961	2.9
충북	60	48.8	63	51.2	123	5.4	7,816	52.3	7,115	47.7	14,931	3.1
충남	48	42.5	65	57.5	113	5.0	10,334	51.9	9,589	48.1	19,923	4.2
전북	63	55.8	50	44.2	113	5.0	9,105	51.3	8,636	48.7	17,741	3.7
전남	65	53.7	56	46.3	121	5.3	8,761	51.7	8,169	48.3	16,930	3.6
경북	90	53.3	79	46.7	169	7.5	12,202	52.7	10,951	47.3	23,153	4.9
경남	60	51.3	57	48.7	117	5.2	17,123	53.1	15,153	46.9	32,276	6.8
제주	47	51.1	45	48.9	92	4.1	3,390	52.4	3,082	47.6	6,472	1.4

### 3) 2차년도 종단면 가중치

#### 기본 가중치

2차년도 가중치 산출을 위해서 1차년도 조사를 통해 수집된 정보와 2차년도 응답여부 변수를 이용한



적절한 통계적 모형을 설정하고 이 모형에 근거한 2차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 2차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 2차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 15>는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 2차년도 응답확률을 추정하기 위하여 사용된 변수들은 '초등학교 입학 전 6개월 이상 다닌 보육 및 교육기관-유치원-이용기간(년)'(EDU3C01A), '이용한 사교육-한글/국어-시작연령'(EDU3D01), '함께 살고 있는 부모님 구성'(FAM1B), '교육수준-부'(HAK2A), '컴퓨터나 게임기를 가지고 노는 시간-학교 가는 날-분'(TIM7B), '수업 시간에 배운 내용을 잘 알고 있나요?'(EDU2A03) 등이다.

<표 16>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.001~1.000의 범위 내에 존재하며, 평균과 표준편차는 각각 0.967과 0.046 정도이다.

2차년도 종단면 기본 가중치는 1차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 15> 초1 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
EDU3C01A	2	8105.2744	0.0001
EDU3D01	4	16.6014	0.0023
FAM1B	3	14.0595	0.0028
HAK2A	3	15.0355	0.0018
TIM7B	3	283.5542	0.0001
EDU2A03	2	8827.5272	0.0001

<표 16> 초1 패널 2차년도 응답확률 추정값의 요약 통계량

N	최소값	최대값	평균	표준편차
2,342	0.0001	1.0000	0.9667	0.0462

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2011년의 초등학교 2학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

초1 패널 2차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 17> 초1 패널 2차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,264	54	658	210	475,425
남자	1,166	54	658	212	247,424
여자	1,098	54	595	208	228,001

<표 18> 초1 패널 2차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,166	51.5	1,098	48.5	2,264	100.0	247,424	52.0	228,001	48.0	475,425	100.0
서울	131	54.1	111	45.9	242	10.7	42,483	51.7	39,729	48.3	82,212	17.3
부산	66	52.4	60	47.6	126	5.6	14,523	52.4	13,216	47.6	27,739	5.8
대구	60	50.0	60	50.0	120	5.3	12,426	52.5	11,245	47.5	23,671	5.0
인천	68	54.4	57	45.6	125	5.5	13,708	51.5	12,922	48.5	26,630	5.6
광주	65	54.2	55	45.8	120	5.3	8,745	52.8	7,827	47.2	16,572	3.5
대전	56	48.3	60	51.7	116	5.1	8,251	51.7	7,712	48.3	15,963	3.4
울산	52	50.5	51	49.5	103	4.5	6,131	53.7	5,294	46.3	11,425	2.4
경기	188	50.0	188	50.0	376	16.6	65,166	51.8	60,660	48.2	125,826	26.5
강원	47	53.4	41	46.6	88	3.9	7,260	52.0	6,701	48.0	13,961	2.9
충북	60	48.8	63	51.2	123	5.4	7,816	52.3	7,115	47.7	14,931	3.1
충남	48	42.5	65	57.5	113	5.0	10,334	51.9	9,589	48.1	19,923	4.2
전북	63	55.8	50	44.2	113	5.0	9,105	51.3	8,636	48.7	17,741	3.7
전남	65	53.7	56	46.3	121	5.3	8,761	51.7	8,169	48.3	16,930	3.6
경북	90	53.3	79	46.7	169	7.5	12,202	52.7	10,951	47.3	23,153	4.9
경남	60	51.3	57	48.7	117	5.2	17,123	53.1	15,153	46.9	32,276	6.8
제주	47	51.1	45	48.9	92	4.1	3,390	52.4	3,082	47.6	6,472	1.4

#### 4) 3차년도 횡단면 가중치

##### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,200/2,342=93.9%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

##### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 초등학교 3학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

##### 가중치의 분포

초1 패널 3차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 19> 초1 패널 3차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,200	49	663	216	474,629
남자	1,134	49	663	218	247,027
여자	1,066	57	577	214	227,602

<표 20> 초1 패널 3차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,134	51.5	1,066	48.5	2,200	100.0	247,027	52.0	227,602	48.0	474,629	100.0
서울	126	55.0	103	45.0	229	10.4	42,031	51.7	39,245	48.3	81,276	17.1
부산	67	51.5	63	48.5	130	5.9	14,468	52.4	13,128	47.6	27,596	5.8
대구	58	49.6	59	50.4	117	5.3	12,485	52.6	11,229	47.4	23,714	5.0
인천	69	55.6	55	44.4	124	5.6	13,784	51.3	13,096	48.7	26,880	5.7
광주	72	52.6	65	47.4	137	6.2	8,800	52.7	7,884	47.3	16,684	3.5
대전	57	47.9	62	52.1	119	5.4	8,280	51.7	7,744	48.3	16,024	3.4
울산	52	51.5	49	48.5	101	4.6	6,121	53.6	5,297	46.4	11,418	2.4
경기	170	48.7	179	51.3	349	15.9	65,110	51.8	60,580	48.2	125,690	26.5
강원	44	53.0	39	47.0	83	3.8	7,256	51.9	6,726	48.1	13,982	2.9
충북	62	49.6	63	50.4	125	5.7	7,779	52.2	7,116	47.8	14,895	3.1
충남	45	41.3	64	58.7	109	5.0	10,451	52.0	9,648	48.0	20,099	4.2
전북	65	57.5	48	42.5	113	5.1	9,144	51.4	8,637	48.6	17,781	3.7
전남	62	53.4	54	46.6	116	5.3	8,756	51.9	8,101	48.1	16,857	3.6
경북	79	54.5	66	45.5	145	6.6	12,078	52.5	10,942	47.5	23,020	4.9
경남	58	52.7	52	47.3	110	5.0	17,065	53.0	15,140	47.0	32,205	6.8
제주	48	51.6	45	48.4	93	4.2	3,419	52.5	3,089	47.5	6,508	1.4

### 5) 3차년도 종단면 가중치

#### 기본 가중치

3차년도 종단면 가중치는 1, 2, 3차년도에 모두 응답한 학생들(2,166명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,166/2,342=92.5%, 2차년도 대비 응답률 2,166/2,264 =95.7%).

3차년도 가중치 산출을 위해서 2차년도 조사를 통해 수집된 정보와 3차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 3차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 3차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 3차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거

후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 21>은 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 3차년도 응답확률을 추정하기 위하여 사용된 변수들은 ‘교육수준-모(HAK2B)’, ‘근로일수-부-주중(JOB2A01)’, ‘근로일수-모-주중(JOB2B01)’, ‘학교숙제(분)-비등교일(TIM3D)’, ‘친구들과 놀기(분)-비등교일(TIM9D)’, 전학경험 유무(EDU1A) 등이다.

<표 22>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.500~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.965와 0.094 정도이다.

3차년도 종단면 기본 가중치는 3차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{2차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 21> 초1 패널 3차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
HAK2B	3	24.9	<.0001
JOB2A01	3	171.1	<.0001
JOB2B01	3	18.0	0.0004
TIM3D	8	16.5	0.0631
TIM9D	6	81.6	<.0001
EDU1A	1	22.3	<.0001

<표 22> 초1 패널 3차년도 응답확률 추정값의 요약 통계량

N	최소값	최대값	평균	표준편차
2,351	0.5000	0.9999	0.9647	0.0939

#### 최종 가중치 (사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 초등학교 3학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

#### 가중치의 분포

초1 패널 3차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 23> 초1 패널 3차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,166	49	1,088	219	474,629
남자	1,115	49	739	222	247,027
여자	1,051	54	1,088	217	227,602

<표 24> 초1 패널 3차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,115	51.5	1,051	48.5	2,166	100.0	247,027	52.0	227,602	48.0	474,629	100.0
서울	125	55.3	101	44.7	226	10.4	42,031	51.7	39,245	48.3	81,276	17.1
부산	66	52.4	60	47.6	126	5.8	14,468	52.4	13,128	47.6	27,596	5.8
대구	57	49.1	59	50.9	116	5.4	12,485	52.6	11,229	47.4	23,714	5.0
인천	66	54.5	55	45.5	121	5.6	13,784	51.3	13,096	48.7	26,880	5.7
광주	68	51.5	64	48.5	132	6.1	8,800	52.7	7,884	47.3	16,684	3.5
대전	55	47.4	61	52.6	116	5.4	8,280	51.7	7,744	48.3	16,024	3.4
울산	52	51.5	49	48.5	101	4.7	6,121	53.6	5,297	46.4	11,418	2.4
경기	169	49.0	176	51.0	345	15.9	65,110	51.8	60,580	48.2	125,690	26.5
강원	44	53.7	38	46.3	82	3.8	7,256	51.9	6,726	48.1	13,982	2.9
충북	59	48.8	62	51.2	121	5.6	7,779	52.2	7,116	47.8	14,895	3.1
충남	45	42.1	62	57.9	107	4.9	10,451	52.0	9,648	48.0	20,099	4.2
전북	62	56.4	48	43.6	110	5.1	9,144	51.4	8,637	48.6	17,781	3.7
전남	62	53.4	54	46.6	116	5.4	8,756	51.9	8,101	48.1	16,857	3.6
경북	79	54.9	65	45.1	144	6.6	12,078	52.5	10,942	47.5	23,020	4.9
경남	58	52.7	52	47.3	110	5.1	17,065	53.0	15,140	47.0	32,205	6.8
제주	48	51.6	45	48.4	93	4.3	3,419	52.5	3,089	47.5	6,508	1.4

## 6) 4차년도 횡단면 가중치

### 기본 가중치

표본탈락률을 보정하기 위하여 1차년도 대비 응답률의 역수를 1차년도 횡단면 가중치에 곱하여 계산하였다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산하였다.

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

초1 패널 4차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 25> 초1 패널 4차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,119	51	694	224	474,126
남자	1,093	55	694	226	246,700
여자	1,026	51	618	222	227,426

<표 26> 초1 패널 4차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,093	51.6	1,026	48.4	2,119	100.0	246,700	52.0	227,426	48.0	474,126	100.0
서울	119	55.1	97	44.9	216	10.2	41,471	51.7	38,762	48.3	80,233	16.9
부산	63	51.2	60	48.8	123	5.8	14,503	52.4	13,174	47.6	27,677	5.8
대구	56	49.6	57	50.4	113	5.3	12,486	52.6	11,267	47.4	23,753	5.0
인천	66	55.9	52	44.1	118	5.6	13,943	51.4	13,201	48.6	27,144	5.7
광주	60	55.6	48	44.4	108	5.1	8,810	52.7	7,894	47.3	16,704	3.5
대전	54	45.8	64	54.2	118	5.6	8,250	51.7	7,704	48.3	15,954	3.4
울산	51	51.0	49	49.0	100	4.7	6,154	53.7	5,301	46.3	11,455	2.4
경기	164	49.5	167	50.5	331	15.6	65,102	51.8	60,561	48.2	125,663	26.5
강원	44	53.0	39	47.0	83	3.9	7,259	51.8	6,749	48.2	14,008	3.0
충북	58	49.2	60	50.8	118	5.6	7,753	52.2	7,112	47.8	14,865	3.1
충남	44	42.3	60	57.7	104	4.9	10,632	51.8	9,888	48.2	20,520	4.3
전북	64	59.3	44	40.7	108	5.1	9,175	51.4	8,665	48.6	17,840	3.8
전남	61	53.0	54	47.0	115	5.4	8,672	51.8	8,072	48.2	16,744	3.5
경북	75	53.2	66	46.8	141	6.7	12,019	52.6	10,841	47.4	22,860	4.8
경남	67	51.5	63	48.5	130	6.1	17,018	53.0	15,118	47.0	32,136	6.8
제주	47	50.5	46	49.5	93	4.4	3,453	52.6	3,117	47.4	6,570	1.4

### 7) 4차년도 종단면 가중치

#### 기본 가중치

4차년도 종단면 가중치는 제1~4차년도에 모두 응답한 청소년들(2,083명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,083/2,342=88.9%, 2차년도 대비 응답률 2,083/2,264 =92.0%, 3차년도 대비 응답률

2,083/2,166=96.2%).

4차년도 가중치 산출을 위해서 3차년도 조사를 통해 수집된 정보와 4차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 4차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 4차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 4차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 27>은 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 4차년도 응답확률을 추정하기 위하여 사용된 변수들은 '등교일 학원/과외 시간(TIM2A)', '전학경험 유무(EDU1A)' 등이다.

<표 28>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.500~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.962와 0.037 정도이다.

4차년도 종단면 기본 가중치는 4차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 3\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 27> 초1 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
TIM2A	5	13.1	0.0226
EDU1A	1	38.6	<.0001

<표 28> 초1 패널 4차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
2,166	0.5000	0.9999	0.9619	0.0372



최종 가중치 (사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 초등학교 4학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

4차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 29> 초1 패널 4차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,083	47	1,195	228	474,126
남자	1,074	54	853	230	246,700
여자	1,009	47	1,195	225	227,426

<표 30> 초1 패널 4차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,074	51.6	1,009	48.4	2,083	100.0	246,700	52.0	227,426	48.0	474,126	100.0
서울	118	55.1	96	44.9	214	10.3	41,471	51.7	38,762	48.3	80,233	16.9
부산	61	51.7	57	48.3	118	5.7	14,503	52.4	13,174	47.6	27,677	5.8
대구	56	49.6	57	50.4	113	5.4	12,486	52.6	11,267	47.4	23,753	5.0
인천	64	55.2	52	44.8	116	5.6	13,943	51.4	13,201	48.6	27,144	5.7
광주	56	53.8	48	46.2	104	5.0	8,810	52.7	7,894	47.3	16,704	3.5
대전	53	45.7	63	54.3	116	5.6	8,250	51.7	7,704	48.3	15,954	3.4
울산	50	51.0	48	49.0	98	4.7	6,154	53.7	5,301	46.3	11,455	2.4
경기	161	49.7	163	50.3	324	15.6	65,102	51.8	60,561	48.2	125,663	26.5
강원	44	54.3	37	45.7	81	3.9	7,259	51.8	6,749	48.2	14,008	3.0
충북	55	48.2	59	51.8	114	5.5	7,753	52.2	7,112	47.8	14,865	3.1
충남	44	43.1	58	56.9	102	4.9	10,632	51.8	9,888	48.2	20,520	4.3
전북	62	58.5	44	41.5	106	5.1	9,175	51.4	8,665	48.6	17,840	3.8
전남	61	53.0	54	47.0	115	5.5	8,672	51.8	8,072	48.2	16,744	3.5
경북	75	53.6	65	46.4	140	6.7	12,019	52.6	10,841	47.4	22,860	4.8
경남	67	51.9	62	48.1	129	6.2	17,018	53.0	15,118	47.0	32,136	6.8
제주	47	50.5	46	49.5	93	4.5	3,453	52.6	3,117	47.4	6,570	1.4

## 8) 5차년도 횡단면 가중치

### 기본가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,070/2,378=87.0%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산하였다(2014년의 초등학교 5학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

5차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 31> 초1 패널 5차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,110	52	697	224	473,280
남자	1,086	55	697	227	246,287
여자	1,024	52	640	222	226,993

<표 32> 초1 패널 5차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,086	51.5	1,024	48.5	2,110	100.0	246,287	52.0	226,993	48.0	473,280	100.0
서울	115	54.2	97	45.8	212	10.0	41,197	51.8	38,345	48.2	79,542	16.8
부산	64	51.2	61	48.8	125	5.9	14,503	52.4	13,168	47.6	27,671	5.8
대구	59	50.9	57	49.1	116	5.5	12,484	52.5	11,276	47.5	23,760	5.0
인천	64	54.2	54	45.8	118	5.6	13,968	51.3	13,248	48.7	27,216	5.8
광주	58	55.8	46	44.2	104	4.9	8,797	52.7	7,905	47.3	16,702	3.5
대전	55	47.0	62	53.0	117	5.5	8,245	51.8	7,686	48.2	15,931	3.4
울산	49	51.6	46	48.4	95	4.5	6,120	53.5	5,313	46.5	11,433	2.4
경기	166	50.8	161	49.2	327	15.5	64,995	51.8	60,543	48.2	125,538	26.5
강원	43	53.8	37	46.2	80	3.8	7,226	51.8	6,728	48.2	13,954	2.9
충북	59	48.8	62	51.2	121	5.7	7,724	52.1	7,102	47.9	14,826	3.1
충남	43	41.0	62	59.0	105	5.0	10,776	51.9	9,993	48.1	20,769	4.4
전북	64	57.7	47	42.3	111	5.3	9,143	51.4	8,655	48.6	17,798	3.8
전남	59	51.8	55	48.2	114	5.4	8,677	51.9	8,051	48.1	16,728	3.5
경북	75	53.2	66	46.8	141	6.7	11,970	52.6	10,782	47.4	22,752	4.8
경남	66	50.4	65	49.6	131	6.2	16,978	53.0	15,048	47.0	32,026	6.8
제주	47	50.5	46	49.5	93	4.4	3,484	52.5	3,150	47.5	6,634	1.4

## 9) 5차년도 종단면 가중치

### 기본 가중치

5차년도 종단면 가중치는 1, 2, 3, 4, 5차년도에 모두 응답한 학생들(2,028명)만을 대상으로 계산되었다 (1차년도 대비 응답률 2,028/2,342=86.6%, 4차년도 대비 응답률 2,028/2,083 =97.4%).

5차년도 가중치 산출을 위해서 4차년도 조사를 통해 수집된 정보와 5차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 5차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 5차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 5차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거

후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 33>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 5차년도 응답확률을 추정하기 위하여 사용된 변수들은 ‘학원/과외숙제 시간-비등교일(TIM4C)이다.

<표 34>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.946~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.978과 0.007 정도이다.

5차년도 종단면 기본 가중치는 4차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 4\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 33> 초1 패널 5차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
TIM4C	4	115	<.0001

<표 34> 초1 패널 5차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
2,028	0.94643	0.99999	0.97808	0.00652

#### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 초등학교 4학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

#### 가중치의 분포

5차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 35> 및 <표 36>에 제시되어 있다.

<표 35> 초1 패널 5차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,028	48	1,268	233	473,280
남자	1,044	54	876	236	246,287
여자	984	48	1,268	231	226,993

<표 36> 초1 패널 5차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,044	51.5	984	48.5	2,028	100.0	246,287	52.0	226,993	48.0	473,280	100.0
서울	110	54.7	91	45.3	201	9.9	41,197	51.8	38,345	48.2	79,542	16.8
부산	59	51.3	56	48.7	115	5.7	14,503	52.4	13,168	47.6	27,671	5.8
대구	57	50.4	56	49.6	113	5.6	12,484	52.5	11,276	47.5	23,760	5.0
인천	62	54.4	52	45.6	114	5.6	13,968	51.3	13,248	48.7	27,216	5.8
광주	53	54.6	44	45.4	97	4.8	8,797	52.7	7,905	47.3	16,702	3.5
대전	51	45.5	61	54.5	112	5.5	8,245	51.8	7,686	48.2	15,931	3.4
울산	49	51.6	46	48.4	95	4.7	6,120	53.5	5,313	46.5	11,433	2.4
경기	158	50.6	154	49.4	312	15.4	64,995	51.8	60,543	48.2	125,538	26.5
강원	42	54.5	35	45.5	77	3.8	7,226	51.8	6,728	48.2	13,954	2.9
충북	56	48.3	60	51.7	116	5.7	7,724	52.1	7,102	47.9	14,826	3.1
충남	43	42.2	59	57.8	102	5.0	10,776	51.9	9,993	48.1	20,769	4.4
전북	61	58.1	44	41.9	105	5.2	9,143	51.4	8,655	48.6	17,798	3.8
전남	57	50.9	55	49.1	112	5.5	8,677	51.9	8,051	48.1	16,728	3.5
경북	74	53.2	65	46.8	139	6.9	11,970	52.6	10,782	47.4	22,752	4.8
경남	65	51.6	61	48.4	126	6.2	16,978	53.0	15,048	47.0	32,026	6.8
제주	47	51.1	45	48.9	92	4.5	3,484	52.5	3,150	47.5	6,634	1.4

## 10) 6차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,067/2,342=88.3%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2015년의 초등학교 6학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

(3) 가중치의 분포

6차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 37> 초1 패널 6차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,067	52	718	229	473,120
남자	1,066	57	718	231	246,199
여자	1,001	52	647	227	226,921

<표 38> 초1 패널 6차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,066	51.6	1,001	48.4	2,067	100.0	246,199	52.0	226,921	48.0	473,120	100.0
서울	112	56.0	88	44.0	200	9.7	40,909	51.7	38,201	48.3	79,110	16.7
부산	63	50.8	61	49.2	124	6.0	14,528	52.4	13,199	47.6	27,727	5.9
대구	55	49.5	56	50.5	111	5.4	12,524	52.6	11,299	47.4	23,823	5.0
인천	62	54.9	51	45.1	113	5.5	14,020	51.4	13,255	48.6	27,275	5.8
광주	56	56.0	44	44.0	100	4.8	8,865	52.7	7,951	47.3	16,816	3.6
대전	55	47.0	62	53.0	117	5.7	8,104	51.7	7,564	48.3	15,668	3.3
울산	46	50.5	45	49.5	91	4.4	6,100	53.5	5,307	46.5	11,407	2.4
경기	159	50.0	159	50.0	318	15.4	64,845	51.8	60,429	48.2	125,274	26.5
강원	44	55.0	36	45.0	80	3.9	7,200	51.8	6,709	48.2	13,909	2.9
충북	56	47.9	61	52.1	117	5.7	7,684	52.2	7,038	47.8	14,722	3.1
충남	47	43.1	62	56.9	109	5.3	11,157	51.9	10,337	48.1	21,494	4.5
전북	62	57.4	46	42.6	108	5.2	9,195	51.5	8,659	48.5	17,854	3.8
전남	60	51.7	56	48.3	116	5.6	8,668	51.8	8,053	48.2	16,721	3.5
경북	76	54.7	63	45.3	139	6.7	11,918	52.6	10,719	47.4	22,637	4.8
경남	67	51.1	64	48.9	131	6.3	16,971	53.1	15,008	46.9	31,979	6.8
제주	46	49.5	47	50.5	93	4.5	3,511	52.4	3,193	47.6	6,704	1.4

## 11) 6차년도 종단면 가중치

### 기본 가중치

6차년도 종단면 가중치는 1, 2, 3, 4, 5, 6차년도에 모두 응답한 학생들(1,972명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,972/2,342=84.2%, 5차년도 대비 응답률 1,972/2,028=97.2%).

6차년도 가중치 산출을 위해서 5차년도 조사를 통해 수집된 정보와 6차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 6차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 6차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 6차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 39>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 6차년도 응답확률을 추정하기 위하여 사용된 변수들은 '학교 지역: 시도'(ARA1A)이다.

<표 40>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.950~0.990의 범위 내에 존재하며, 평균과 표준편차는 각각 0.973과 0.013 정도이다.

6차년도 종단면 기본 가중치는 5차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 5\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 39> 초1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
ARA1A	4	11.0	0.0508

<표 40> 초1 패널 6차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
1,972	0.95025	0.98997	0.97255	0.01265

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2015년의 초등학교 6학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

5차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 41> 및 <표 42>에 제시되어 있다.

<표 41> 초1 패널 6차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,972	49	1,287	240	473,120
남자	1,015	56	916	243	246,199
여자	957	49	1,287	237	226,921



<표 42> 초1 패널 6차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,015	51.5	957	48.5	1,972	100.0	246,199	52.0	226,921	48.0	473,120	100.0
서울	107	56.6	82	43.4	189	9.6	40,909	51.7	38,201	48.3	79,110	16.7
부산	58	51.3	55	48.7	113	5.7	14,528	52.4	13,199	47.6	27,727	5.9
대구	54	49.5	55	50.5	109	5.5	12,524	52.6	11,299	47.4	23,823	5.0
인천	61	56.0	48	44.0	109	5.5	14,020	51.4	13,255	48.6	27,275	5.8
광주	49	53.8	42	46.2	91	4.6	8,865	52.7	7,951	47.3	16,816	3.6
대전	50	45.5	60	54.5	110	5.6	8,104	51.7	7,564	48.3	15,668	3.3
울산	45	50.0	45	50.0	90	4.6	6,100	53.5	5,307	46.5	11,407	2.4
경기	151	49.8	152	50.2	303	15.4	64,845	51.8	60,429	48.2	125,274	26.5
강원	41	54.7	34	45.3	75	3.8	7,200	51.8	6,709	48.2	13,909	2.9
충북	54	47.8	59	52.2	113	5.7	7,684	52.2	7,038	47.8	14,722	3.1
충남	46	43.4	60	56.6	106	5.4	11,157	51.9	10,337	48.1	21,494	4.5
전북	59	58.4	42	41.6	101	5.1	9,195	51.5	8,659	48.5	17,854	3.8
전남	55	49.5	56	50.5	111	5.6	8,668	51.8	8,053	48.2	16,721	3.5
경북	75	54.7	62	45.3	137	6.9	11,918	52.6	10,719	47.4	22,637	4.8
경남	64	51.6	60	48.4	124	6.3	16,971	53.1	15,008	46.9	31,979	6.8
제주	46	50.5	45	49.5	91	4.6	3,511	52.4	3,193	47.6	6,704	1.4

## 12) 7차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,002/2,342=85.5%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2016년의 중학생 1학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### (3) 가중치의 분포

7차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 43> 초1 패널 7차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,002	55	729	236	473,196
남자	1,039	55	729	237	246,639
여자	963	64	722	235	226,557

<표 44> 초1 패널 7차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,033	51.9	959	48.1	1,992	100.0	245,310	52.1	225,574	47.9	470,884	100.0
서울	103	57.5	76	42.5	179	9.0	40,438	52.0	37,258	48.0	77,696	16.5
부산	59	48.8	62	51.2	121	6.1	14,534	52.4	13,212	47.6	27,746	5.9
대구	53	50.0	53	50.0	106	5.3	12,568	52.7	11,300	47.3	23,868	5.1
인천	60	55.0	49	45.0	109	5.5	13,945	51.4	13,169	48.6	27,114	5.8
광주	56	56.0	44	44.0	100	5.0	8,711	52.5	7,870	47.5	16,581	3.5
대전	51	47.2	57	52.8	108	5.4	7,960	51.7	7,440	48.3	15,400	3.3
울산	44	50.0	44	50.0	88	4.4	6,053	53.3	5,305	46.7	11,358	2.4
경기	157	49.7	159	50.3	316	15.9	64,680	51.8	60,232	48.2	124,912	26.5
강원	40	53.3	35	46.7	75	3.8	7,188	51.8	6,678	48.2	13,866	2.9
충북	53	48.2	57	51.8	110	5.5	7,630	52.2	6,980	47.8	14,610	3.1
충남	50	43.9	64	56.1	114	5.7	11,441	51.7	10,668	48.3	22,109	4.7
전북	63	58.9	44	41.1	107	5.4	9,260	51.5	8,721	48.5	17,981	3.8
전남	57	51.8	53	48.2	110	5.5	8,644	51.9	8,003	48.1	16,647	3.5
경북	77	55.0	63	45.0	140	7.0	11,833	52.8	10,574	47.2	22,407	4.8
경남	67	51.9	62	48.1	129	6.5	16,861	53.0	14,960	47.0	31,821	6.8
제주	43	53.8	37	46.3	80	4.0	3,564	52.7	3,204	47.3	6,768	1.4

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 10명은 사후층화의 과정이 생략됨.

### 13) 7차년도 종단면 가중치

#### 기본 가중치

7차년도 종단면 가중치는 1, 2, 3, 4, 5, 6, 7차년도에 모두 응답한 학생들(1,879명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,879/2,342=80.2%, 6차년도 대비 응답률 1,879/1,972=95.3%).

7차년도 가중치 산출을 위해서 6차년도 조사를 통해 수집된 정보와 7차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 7차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 7차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 7차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 45>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 7차년도 응답확률을 추정하기 위하여 사용된 변수들은 '학교 지역: 시도'(ARA1A)이다.

<표 46>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.846~0.996의 범위 내에 존재하며, 평균과 표준편차는 각각 0.954과 0.034 정도이다.

7차년도 종단면 기본 가중치는 6차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{6차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 45> 초1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
ARA1A	5	40.7	<.0001

<표 46> 초1 패널 6차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
1,879	0.84615	0.99565	0.95413	0.03421

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2016년의 중학교 1학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

7차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 47> 및 <표 48>에 제시되어 있다.

<표 47> 초1 패널 7차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,879	54	1,184	252	473,421
남자	970	54	967	255	246,869
여자	909	61	1,184	249	226,552

<표 48> 초1 패널 7차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	964	51.6	905	48.4	1,869	100.0	245,310	52.1	225,575	47.9	470,885	100.0
서울	95	57.2	71	42.8	166	8.9	40,438	52.0	37,258	48.0	77,696	16.5
부산	54	50.0	54	50.0	108	5.8	14,534	52.4	13,212	47.6	27,746	5.9
대구	52	50.5	51	49.5	103	5.5	12,568	52.7	11,300	47.3	23,868	5.1
인천	57	57.6	42	42.4	99	5.3	13,945	51.4	13,169	48.6	27,114	5.8
광주	45	52.3	41	47.7	86	4.6	8,711	52.5	7,870	47.5	16,581	3.5
대전	46	45.5	55	54.5	101	5.4	7,960	51.7	7,440	48.3	15,400	3.3
울산	42	49.4	43	50.6	85	4.5	6,053	53.3	5,305	46.7	11,358	2.4
경기	143	48.6	151	51.4	294	15.7	64,680	51.8	60,232	48.2	124,912	26.5
강원	38	54.3	32	45.7	70	3.7	7,188	51.8	6,679	48.2	13,867	2.9
충북	50	47.6	55	52.4	105	5.6	7,630	52.2	6,980	47.8	14,610	3.1
충남	48	43.6	62	56.4	110	5.9	11,441	51.7	10,668	48.3	22,109	4.7
전북	59	59.6	40	40.4	99	5.3	9,260	51.5	8,721	48.5	17,981	3.8
전남	53	50.0	53	50.0	106	5.7	8,644	51.9	8,003	48.1	16,647	3.5
경북	75	54.7	62	45.3	137	7.3	11,833	52.8	10,574	47.2	22,407	4.8
경남	64	52.5	58	47.5	122	6.5	16,861	53.0	14,960	47.0	31,821	6.8
제주	43	55.1	35	44.9	78	4.2	3,564	52.7	3,204	47.3	6,768	1.4

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 10명은 사후층화의 과정이 생략됨.

## 2. 초4 패널 가중치

### 1) 1차년도 횡단면 가중치

표본추출확률에 대한 가중치(기본 가중치)

KCYPS의 표본추출방식은 일종의 층화2단집락추출법으로서, 시도와 도시규모를 층으로 간주할 수 있다. 각 층에서 1차추출단위인 학교는 학생수를 기준으로 확률비례추출법으로 선정하고, 추출된 표본학교에서는 학급을 랜덤하게 선정한 후에 학생을 조사하였다. 따라서 추출된 표본학교에서 학생을 랜덤하게 선정하여 조사하는 2단집락추출법으로 볼 수 있으므로 표본추출률을 다음의 식으로 나타낼 수 있다.

$$f_{hij} = l_h \times \frac{M_{hi}}{\sum_{i=1}^{K_h} M_{hi}} \times \frac{n_{hi}}{M_{hi}} = \frac{l_h \times n_{hi}}{\sum_{i=1}^{K_h} M_{hi}}$$

여기서  $K_h$ 는  $h$ 층에 속한 학교의 수,  $l_h$ 는  $h$ 층에서 추출된 학교의 수,  $M_{hi}$ 는  $h$ 층  $i$ 번째 학교의 학생 수,  $n_{hi}$ 는 추출된 표본학교에서 조사한 학생수를 나타낸다.

표본추출률에 대한 가중치는 추출률의 역수이므로 아래와 같은 식으로 표현할 수 있다.

$$W_{1hij} = \frac{\sum_{i=1}^{K_h} M_{hi}}{l_h \times n_{hi}} = \frac{M_h}{l_h \times n_{hi}}$$

여기서  $M_h$ 는  $h$ 층의 전체 학생수를 나타낸다.

표본의 무응답 조정을 위한 가중치

표본으로 선정된 조사단위들 중 응답거부로 조사에 성공하지 못하거나 조사된 내용이 부실하여 무응답처리한 조사 단위에 대한 조정이 무응답 조정의 가중치로서, 다음의 식으로 계산할 수 있다.

$$W_{2hij} = \frac{n_{hi}}{r_{hi}}$$

여기서  $r_{hi}$ 와  $n_{hi}$ 는 각각 완전 응답한 표본크기와 추출된 표본크기이다.

사후층화 가중치

표본설계시에는 층화변수로 고려하지 않았으나 표본조사 완료 후에 연구변수에 영향을 미치는 변수들 중에서 정확한 외부보조정보를 확보하게 되면, 외부보조정보 데이터의 구조와 표본조사 데이터의 구조

를 분석하여 층으로 구분하고 표본조사의 각 셀에 가중치의 합계를 계산한 결과와 외부보조정보의 층별 합계를 비교하여 가중치를 계산한다.

$$BF_{hij} = \frac{PS_{hij}}{\sum W_{1hij} \cdot W_{2hij}}$$

여기서  $PS_{hij}$ 는 외부보조정보에 의한 셀의 빈도수이고, 분모의 내용은 설계가중치와 무응답 조정을 위한 가중치를 곱한 후에 합계를 계산결과이다. KCYPS에서는 16개 광역시·도×성별 사후층화를 실시하였다. 최종 가중치는 세가지 가중치를 곱하여 아래와 같은 식으로 산출한다.

$$W_{thi} = W_{1hii} \cdot W_{2hii} \cdot BF_{hi}$$

## 2) 2차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,264/2,378=95.2%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2011년의 초등학생 5학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

초4 패널 2차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 49> 초4 패널 2차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,264	65	751	266	603,245
남자	1,183	65	744	267	315,755
여자	1,081	76	751	266	287,490

<표 50> 초4 패널 2차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,183	52.3	1,080	47.7	2,263	100.0	315,755	52.3	287,410	47.7	603,165	100.0
서울	135	51.7	126	48.3	261	11.5	53,541	52.1	49,281	47.9	102,822	17.0
부산	65	49.6	66	50.4	131	5.8	19,081	52.9	16,980	47.1	36,061	6.0
대구	75	54.3	63	45.7	138	6.1	16,361	53.2	14,377	46.8	30,738	5.1
인천	54	50.9	52	49.1	106	4.7	17,238	51.9	15,948	48.1	33,186	5.5
광주	57	50.4	56	49.6	113	5.0	11,280	52.3	10,275	47.7	21,555	3.6
대전	60	53.6	52	46.4	112	4.9	10,248	51.7	9,555	48.3	19,803	3.3
울산	49	50.0	49	50.0	98	4.3	7,913	52.7	7,090	47.3	15,003	2.5
경기	186	51.8	173	48.2	359	15.9	80,811	52.1	74,346	47.9	155,157	25.7
강원	57	53.3	50	46.7	107	4.7	9,625	52.8	8,589	47.2	18,214	3.0
충북	67	48.6	71	51.4	138	6.1	10,023	52.5	9,082	47.5	19,105	3.2
충남	65	53.7	56	46.3	121	5.3	13,355	52.3	12,159	47.7	25,514	4.2
전북	70	53.8	60	46.2	130	5.7	12,060	51.8	11,217	48.2	23,277	3.9
전남	47	50.0	47	50.0	94	4.2	11,903	52.3	10,841	47.7	22,744	3.8
경북	70	58.8	49	41.2	119	5.3	16,066	53.1	14,193	46.9	30,259	5.0
경남	78	54.2	66	45.8	144	6.4	22,077	53.1	19,509	46.9	41,586	6.9
제주	48	52.2	44	47.8	92	4.1	4,173	51.3	3,968	48.7	8,141	1.3

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 1명은 사후층화의 과정이 생략됨.

### 3) 2차년도 종단면 가중치

#### 기본 가중치

2차년도 가중치 산출을 위해서 1차년도 조사를 통해 수집된 정보와 2차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 2차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 2차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼



저 각 설명변수에 대하여 2차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha = 0.15$ ).

<표 51>는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 2차년도 응답확률을 추정하기 위하여 사용된 변수들은 ‘살고 있는 곳(시·도)’(ARA2A), ‘그 이외에 방과 후 친구들과 노는 시간-학교 가는 날-시간’(TIM9B), ‘학교 숙제를 하는 시간-학교가지않는날-분’(TIM3D), ‘공부 시간에 딴 짓을 한다’(EDU2A05), ‘부모님(보호자)께서는 보통 아이들이 할 수 있는 일도 불안해 하시며 내게는 못 하게 하신다’(FAM2E01), ‘교과서나 참고서 이외의 책을 읽는 시간-학교가지않는날-시간’(TIM6C), ‘컴퓨터 사용-전자우편’(MDA1D05), ‘컴퓨터 사용-개인 홈페이지 활동’(MDA1D07), ‘휴대폰 사용-친구와의 통화’(MDA2B03), ‘휴대폰 사용-사진 촬영’(MDA2B06), ‘휴대전화로 한참 동안 아무에게서도 연락이 오지 않으면 불안하다’(MDA2C03), ‘방과 후 교육의 가계에 부담되는 정도’(EDU5E) 등이다.

<표 52>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.001~1.000의 범위 내에 존재하며, 평균과 표준편차는 각각 0.955와 0.072 정도이다.

2차년도 종단면 기본 가중치는 1차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 51> 초4 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
ARA2A	15	26.7104	0.0312
TIM9B	11	627.2789	<.0001
TIM3D	14	19.4941	0.1469
EDU2A05	2	8.4257	0.0148
FAM2E01	2	10.8922	0.0043
TIM6C	2	10.1745	0.0062
MDA1D05	3	5.6252	0.1313
MDA1D07	1	4.8933	0.0270
MDA2B03	2	13.1146	0.0014
MDA2B06	2	5.4870	0.0643
MDA2C03	2	5.8762	0.0530
EDU5E	3	12.6293	0.0055

<표 52> 초4 패널 2차년도 응답확률 추정값의 요약 통계량

N	최소값	최대값	평균	표준편차
2,342	0.0001	1.0000	0.9546	0.0721

최종 가중치 (사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2011년의 초등학생 5학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

초4 패널 2차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 53> 초4 패널 2차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,264	63	922	266	603,242
남자	1,183	63	922	267	315,755
여자	1,081	74	881	266	287,487

<표 54> 초4 패널 2차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,183	52.3	1,080	47.7	2,263	100.0	315,755	52.3	287,410	47.7	603,165	100.0
서울	135	51.7	126	48.3	261	11.5	53,541	52.1	49,281	47.9	102,822	17.0
부산	65	49.6	66	50.4	131	5.8	19,081	52.9	16,980	47.1	36,061	6.0
대구	75	54.3	63	45.7	138	6.1	16,361	53.2	14,377	46.8	30,738	5.1
인천	54	50.9	52	49.1	106	4.7	17,238	51.9	15,948	48.1	33,186	5.5
광주	57	50.4	56	49.6	113	5.0	11,280	52.3	10,275	47.7	21,555	3.6
대전	60	53.6	52	46.4	112	4.9	10,248	51.7	9,555	48.3	19,803	3.3
울산	49	50.0	49	50.0	98	4.3	7,913	52.7	7,090	47.3	15,003	2.5
경기	186	51.8	173	48.2	359	15.9	80,811	52.1	74,346	47.9	155,157	25.7
강원	57	53.3	50	46.7	107	4.7	9,625	52.8	8,589	47.2	18,214	3.0
충북	67	48.6	71	51.4	138	6.1	10,023	52.5	9,082	47.5	19,105	3.2
충남	65	53.7	56	46.3	121	5.3	13,355	52.3	12,159	47.7	25,514	4.2
전북	70	53.8	60	46.2	130	5.7	12,060	51.8	11,217	48.2	23,277	3.9
전남	47	50.0	47	50.0	94	4.2	11,903	52.3	10,841	47.7	22,744	3.8
경북	70	58.8	49	41.2	119	5.3	16,066	53.1	14,193	46.9	30,259	5.0
경남	78	54.2	66	45.8	144	6.4	22,077	53.1	19,509	46.9	41,586	6.9
제주	48	52.2	44	47.8	92	4.1	4,173	51.3	3,968	48.7	8,141	1.3

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 1명은 사후층화의 과정이 생략됨.

#### 4) 3차년도 횡단면 가중치

##### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,219/2,378=93.3%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

##### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 초등학생 6학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

## 가중치의 분포

초4 패널 3차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 55> 초4 패널 3차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,219	62	774	272	602,761
남자	1,169	62	746	270	315,566
여자	1,050	76	774	274	287,195

<표 56> 초4 패널 3차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,168	52.7	1,049	47.3	2,217	100.0	315,423	52.3	287,115	47.7	602,538	100.0
서울	130	51.8	121	48.2	251	11.3	53,240	52.1	48,955	47.9	102,195	17.0
부산	68	49.6	69	50.4	137	6.2	19,018	52.8	16,979	47.2	35,997	6.0
대구	74	54.0	63	46.0	137	6.2	16,412	53.2	14,453	46.8	30,865	5.1
인천	51	53.7	44	46.3	95	4.3	17,300	51.9	16,009	48.1	33,309	5.5
광주	55	51.4	52	48.6	107	4.8	11,373	52.4	10,339	47.6	21,712	3.6
대전	61	55.5	49	44.5	110	5.0	10,281	51.7	9,594	48.3	19,875	3.3
울산	51	50.5	50	49.5	101	4.6	7,945	52.9	7,086	47.1	15,031	2.5
경기	185	52.3	169	47.7	354	16.0	80,626	52.1	74,232	47.9	154,858	25.7
강원	53	52.0	49	48.0	102	4.6	9,619	52.9	8,560	47.1	18,179	3.0
충북	66	47.8	72	52.2	138	6.2	10,117	52.6	9,123	47.4	19,240	3.2
충남	66	52.4	60	47.6	126	5.7	13,375	52.3	12,205	47.7	25,580	4.2
전북	72	57.1	54	42.9	126	5.7	12,067	51.7	11,251	48.3	23,318	3.9
전남	45	52.9	40	47.1	85	3.8	11,847	52.4	10,783	47.6	22,630	3.8
경북	68	58.6	48	41.4	116	5.2	16,001	53.1	14,117	46.9	30,118	5.0
경남	74	53.2	65	46.8	139	6.3	22,026	53.1	19,453	46.9	41,479	6.9
제주	49	52.7	44	47.3	93	4.2	4,176	51.2	3,976	48.8	8,152	1.4

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 2명은 사후층화의 과정이 생략됨.

## 5) 3차년도 종단면 가중치

### 기본 가중치

3차년도 종단면 가중치는 1, 2, 3차년도에 모두 응답한 학생들(2,170명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,170/2,378=91.3%, 2차년도 대비 응답률 2,170/2,263 =95.9%).

3차년도 가중치 산출을 위해서 2차년도 조사를 통해 수집된 정보와 3차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 3차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 3차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 3차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다. 범주가 새롭게 정의된 변수들의 범주 값과 빈도표 및  $\chi^2$ -검정의 결과는 부록에 제시되어 있다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

초등학교 4학년 패널의 경우, 위와 같은 과정을 적용하였으나 통계적으로 유의하게 선택된 변수가 없었다. 따라서 3차년도 종단면 기본 가중치는 2차년도 종단면 가중치와 동일한 값이 적용되었다.

기본 가중치 = 2차년도 가중치

#### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 초등학교 6학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

#### 가중치의 분포

초4 패널 3차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 57> 초4 패널 3차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,170	60	1,012	278	602,675
남자	1,139	60	1,012	277	315,560
여자	1,031	76	920	278	287,115

<표 58> 초4 패널 3차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,138	52.5	1,031	47.5	2,169	100.0	315,423	52.3	287,115	47.7	602,538	100.0
서울	129	51.8	120	48.2	249	11.5	53,240	52.1	48,955	47.9	102,195	17.0
부산	66	50.0	66	50.0	132	6.1	19,018	52.8	16,979	47.2	35,997	6.0
대구	74	54.4	62	45.6	136	6.3	16,412	53.2	14,453	46.8	30,865	5.1
인천	49	52.7	44	47.3	93	4.3	17,300	51.9	16,009	48.1	33,309	5.5
광주	54	51.4	51	48.6	105	4.8	11,373	52.4	10,339	47.6	21,712	3.6
대전	59	55.1	48	44.9	107	4.9	10,281	51.7	9,594	48.3	19,875	3.3
울산	49	50.5	48	49.5	97	4.5	7,945	52.9	7,086	47.1	15,031	2.5
경기	171	51.2	163	48.8	334	15.4	80,626	52.1	74,232	47.9	154,858	25.7
강원	53	52.0	49	48.0	102	4.7	9,619	52.9	8,560	47.1	18,179	3.0
충북	66	48.2	71	51.8	137	6.3	10,117	52.6	9,123	47.4	19,240	3.2
충남	66	52.4	60	47.6	126	5.8	13,375	52.3	12,205	47.7	25,580	4.2
전북	69	56.1	54	43.9	123	5.7	12,067	51.7	11,251	48.3	23,318	3.9
전남	44	52.4	40	47.6	84	3.9	11,847	52.4	10,783	47.6	22,630	3.8
경북	67	58.3	48	41.7	115	5.3	16,001	53.1	14,117	46.9	30,118	5.0
경남	73	53.3	64	46.7	137	6.3	22,026	53.1	19,453	46.9	41,479	6.9
제주	49	53.3	43	46.7	92	4.2	4,176	51.2	3,976	48.8	8,152	1.4

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 1명은 사후층화의 과정이 생략됨.

## 6) 4차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,092/2,378=88.0%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

사후층화 가중치는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 중학생 1학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

## 가중치의 분포

4차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 59> 초4 패널 4차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,092	70	890	287	601,268
남자	1,100	71	805	287	315,186
여자	992	70	890	288	286,082

<표 60> 초4 패널 4차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,098	52.6	990	47.4	2,088	100.0	314,862	52.4	285,796	47.6	600,658	100.0
서울	117	49.0	122	51.0	239	11.4	52,988	52.4	48,115	47.6	101,103	16.8
부산	61	49.6	62	50.4	123	5.9	19,136	52.8	17,117	47.2	36,253	6.0
대구	70	54.7	58	45.3	128	6.1	16,441	53.2	14,484	46.8	30,925	5.1
인천	45	51.7	42	48.3	87	4.2	17,276	51.8	16,081	48.2	33,357	5.6
광주	50	48.1	54	51.9	104	5.0	11,303	52.3	10,326	47.7	21,629	3.6
대전	58	52.3	53	47.7	111	5.3	10,349	51.9	9,586	48.1	19,935	3.3
울산	48	53.3	42	46.7	90	4.3	7,959	53.1	7,040	46.9	14,999	2.5
경기	175	54.7	145	45.3	320	15.3	80,448	52.1	74,009	47.9	154,457	25.7
강원	54	53.5	47	46.5	101	4.8	9,586	53.0	8,487	47.0	18,073	3.0
충북	62	47.7	68	52.3	130	6.2	9,981	52.3	9,086	47.7	19,067	3.2
충남	64	53.3	56	46.7	120	5.7	13,398	52.1	12,337	47.9	25,735	4.3
전북	71	57.7	52	42.3	123	5.9	12,135	51.7	11,345	48.3	23,480	3.9
전남	44	53.7	38	46.3	82	3.9	11,782	52.5	10,673	47.5	22,455	3.7
경북	68	59.1	47	40.9	115	5.5	15,870	53.3	13,895	46.7	29,765	5.0
경남	68	51.9	63	48.1	131	6.3	22,009	53.4	19,243	46.6	41,252	6.9
제주	43	51.2	41	48.8	84	4.0	4,201	51.4	3,972	48.6	8,173	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 4명은 사후층화의 과정이 생략됨

## 7) 4차년도 종단면 가중치

### 기본 가중치

4차년도 종단면 가중치는 1, 2, 3, 4차년도에 모두 응답한 학생들(2,012명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,012/2,378=84.6%, 2차년도 대비 응답률 2,012/2,263 =88.9%, 3차년도 대비 응답률

2,012/2,170=92.7%).

4차년도 가중치 산출을 위해서 3차년도 조사를 통해 수집된 정보와 4차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 4차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 4차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 4차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 61>는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 4차년도 응답확률을 추정하기 위하여 사용된 변수들은 '방과후 보호자 부재 일수'(FAM4A), '모친의 최종학력'(HAK2B), '보호자의 삶의 만족도'(PAR2), '등교일 학원/과외 시간'(TIM4A), '책, CD, 비디오테이프, DVD 구입'(ACT6B08), '인터넷 사이트에 동영상 올리기'(ACT6B09), '나는 공부가 지루해도 계획한 것은 마친다'(INT2C03), '장래가 희망적이지 않은 것 같다'(PSY1E09), '다른 사람 헐박하기'(DLQ1A09), '나는 우리 동네 사람들 대부분을 알고 있다'(COM1A01), '주변에 어려움에 처해있는 친구가 있다면 적극적으로 도울 수 있다'(COM2A01), '형제자매 유무'(FAM1D), '부끄럼을 많이 탄다'(PSY1D02) 등이다.

<표 61> 초4 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
FAM4A	1	762.6	<.0001
HAK2B	4	7.9	0.0949
PAR2	2	9	0.0108
TIM4A	3	98.3	<.0001
ACT6B08	1	298.1	<.0001
ACT6B09	1	8.8	0.0029
INT2C03	2	13.6	0.0011
PSY1E09	1	55.4	<.0001
DLQ1A09	2	542.7	<.0001
COM1A01	2	7.6	0.0215
COM2A01	2	15.8	0.0004
FAM1D	2	8	0.0181
PSY1D02	3	9.7	0.0207



<표 62>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.500~0.998의 범위 내에 존재하며, 평균과 표준편차는 각각 0.932와 0.085 정도이다.

<표 62> 초4 패널 4차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
2,170	0.5000	0.9981	0.9320	0.0853

4차년도 종단면 기본 가중치는 3차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{3차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

최종 가중치(사후층화 가중치)

사후층화 가중치는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 중학생 1학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

4차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 63> 초4 패널 4차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,012	64	1,334	299	601,163
남자	1,057	64	1,019	298	315,170
여자	955	73	1,334	299	285,992

<표 64> 초4 패널 4차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,055	52.5	954	47.5	2,009	100.0	314,862	52.4	285,796	47.6	600,658	100.0
서울	115	49.6	117	50.4	232	11.5	52,988	52.4	48,115	47.6	101,103	16.8
부산	58	50.0	58	50.0	116	5.8	19,136	52.8	17,117	47.2	36,253	6.0
대구	70	54.7	58	45.3	128	6.4	16,441	53.2	14,484	46.8	30,925	5.1
인천	40	50.6	39	49.4	79	3.9	17,276	51.8	16,081	48.2	33,357	5.6
광주	48	48.5	51	51.5	99	4.9	11,303	52.3	10,326	47.7	21,629	3.6
대전	56	52.3	51	47.7	107	5.3	10,349	51.9	9,586	48.1	19,935	3.3
울산	47	54.0	40	46.0	87	4.3	7,959	53.1	7,040	46.9	14,999	2.5
경기	159	54.1	135	45.9	294	14.6	80,448	52.1	74,009	47.9	154,457	25.7
강원	51	52.6	46	47.4	97	4.8	9,586	53.0	8,487	47.0	18,073	3.0
충북	62	48.1	67	51.9	129	6.4	9,981	52.3	9,086	47.7	19,067	3.2
충남	64	53.3	56	46.7	120	6.0	13,398	52.1	12,337	47.9	25,735	4.3
전북	68	57.6	50	42.4	118	5.9	12,135	51.7	11,345	48.3	23,480	3.9
전남	43	53.1	38	46.9	81	4.0	11,782	52.5	10,673	47.5	22,455	3.7
경북	65	58.0	47	42.0	112	5.6	15,870	53.3	13,895	46.7	29,765	5.0
경남	66	51.6	62	48.4	128	6.4	22,009	53.4	19,243	46.6	41,252	6.9
제주	43	52.4	39	47.6	82	4.1	4,201	51.4	3,972	48.6	8,173	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 3명은 사후층화의 과정이 생략됨

## 8) 5차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,070/2,378=87.0%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

사후층화 가중치는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2014년의 중학생 2학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

5차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 65> 및 <표 66>에 제시되어 있다.

<표 65> 초4 패널 5차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,070	69	812	290	599,300
남자	1,089	90	795	288	313,957
여자	981	69	812	291	285,343

<표 66> 초4 패널 5차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,088	52.6	979	47.4	2,067	100.0	313,705	52.4	284,651	47.6	598,356	100.0
서울	117	50.6	114	49.4	231	11.2	52,636	52.4	47,854	47.6	100,490	16.8
부산	66	52.8	59	47.2	125	6.0	19,007	52.7	17,056	47.3	36,063	6.0
대구	67	53.6	58	46.4	125	6.0	16,410	53.2	14,429	46.8	30,839	5.2
인천	48	55.2	39	44.8	87	4.2	17,286	51.7	16,120	48.3	33,406	5.6
광주	51	48.1	55	51.9	106	5.1	11,296	52.3	10,294	47.7	21,590	3.6
대전	55	50.9	53	49.1	108	5.2	10,262	51.8	9,545	48.2	19,087	3.3
울산	48	55.8	38	44.2	86	4.2	7,926	52.9	7,044	47.1	14,970	2.5
경기	177	52.5	160	47.5	337	16.3	80,084	52.1	73,551	47.9	153,635	25.7
강원	51	52.0	47	48.0	98	4.7	9,499	52.9	8,444	47.1	17,943	3.0
충북	62	48.1	67	51.9	129	6.2	9,947	52.4	9,032	47.6	18,979	3.2
충남	63	53.4	55	46.6	118	5.7	13,449	52.1	12,371	47.9	25,820	4.3
전북	72	58.5	51	41.5	123	6.0	12,108	51.8	11,281	48.2	23,389	3.9
전남	41	51.9	38	48.1	79	3.8	11,781	52.6	10,607	47.4	22,388	3.7
경북	67	58.3	48	41.7	115	5.6	15,801	53.3	13,827	46.7	29,628	5.0
경남	67	53.2	59	46.8	126	6.1	22,015	53.4	19,235	46.6	41,250	6.9
제주	36	48.6	38	51.4	74	3.6	4,198	51.5	3,961	48.5	8,159	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 3명은 사후층화의 과정이 생략됨

## 9) 5차년도 종단면 가중치

### 기본 가중치

5차년도 종단면 가중치는 1, 2, 3, 4, 5차년도에 모두 응답한 학생들(1,913명)만을 대상으로 계산되었다 (1차년도 대비 응답률 1,913/2,378=80.4%, 4차년도 대비 응답률 1,913/2,012 =95.1%).

5차년도 가중치 산출을 위해서 4차년도 조사를 통해 수집된 정보와 5차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 5차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 5차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 5차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 67>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 5차년도 응답확률을 추정하기 위하여 사용된 변수들은 '월평균 용돈'(FAM5A)이다.

<표 68>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.500~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.949와 0.026 정도이다.

<표 67> 초4 패널 5차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
FAM5A	14	70.7	<.0001

<표 68> 초4 패널 5차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
1,913	0.5	0.99999	0.94962	0.02557

5차년도 종단면 기본 가중치는 4차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 4\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

#### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2014년의 중학생 2학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

#### 가중치의 분포

5차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 69> 및 <표 70>에 제시되어 있다.

<표 69> 초4 패널 5차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,913	74	1,360	313	598,960
남자	1,005	87	1,053	312	313,954
여자	908	74	1,360	314	285,006

<표 70> 초4 패널 5차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,004	52.5	907	47.5	1,911	100.0	313,705	52.4	284,651	47.6	598,356	100.0
서울	108	50.0	108	50.0	216	11.3	52,636	52.4	47,854	47.6	100,490	16.8
부산	56	52.8	50	47.2	106	5.5	19,007	52.7	17,056	47.3	36,063	6.0
대구	67	54.0	57	46.0	124	6.5	16,410	53.2	14,429	46.8	30,839	5.2
인천	37	51.4	35	48.6	72	3.8	17,286	51.7	16,120	48.3	33,406	5.6
광주	47	48.0	51	52.0	98	5.1	11,296	52.3	10,294	47.7	21,590	3.6
대전	51	50.5	50	49.5	101	5.3	10,262	51.8	9,545	48.2	19,807	3.3
울산	45	57.0	34	43.0	79	4.1	7,926	52.9	7,044	47.1	14,970	2.5
경기	153	53.5	133	46.5	286	15.0	80,084	52.1	73,551	47.9	153,635	25.7
강원	48	51.1	46	48.9	94	4.9	9,499	52.9	8,444	47.1	17,943	3.0
충북	62	48.4	66	51.6	128	6.7	9,947	52.4	9,032	47.6	18,979	3.2
충남	63	53.8	54	46.2	117	6.1	13,449	52.1	12,371	47.9	25,820	4.3
전북	67	58.8	47	41.2	114	6.0	12,108	51.8	11,281	48.2	23,389	3.9
전남	40	52.6	36	47.4	76	4.0	11,781	52.6	10,607	47.4	22,388	3.7
경북	62	56.9	47	43.1	109	5.7	15,801	53.3	13,827	46.7	29,628	5.0
경남	63	52.5	57	47.5	120	6.3	22,015	53.4	19,235	46.6	41,250	6.9
제주	35	49.3	36	50.7	71	3.7	4,198	51.5	3,961	48.5	8,159	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답)2명은 사후층화의 과정이 생략됨

## 10) 6차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,061/2,378=86.7%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2015년의 중학생 3학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

6차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 71> 및 <표 72>에 제시되어 있다.

<표 71> 초4 패널 6차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,061	72	869	290	597,309
남자	1,091	90	812	287	313,318
여자	970	72	869	293	283,991

<표 72> 초4 패널 6차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,088	52.9	969	47.1	2,057	100.0	312,452	52.4	283,615	47.6	596,067	100.0
서울	122	51.3	116	48.7	238	11.6	52,293	52.4	47,565	47.6	99,858	16.8
부산	64	51.6	60	48.4	124	6.0	18,925	52.7	16,998	47.3	35,923	6.0
대구	66	53.2	58	46.8	124	6.0	16,373	53.2	14,417	46.8	30,790	5.2
인천	54	58.1	39	41.9	93	4.5	17,265	51.7	16,113	48.3	33,378	5.6
광주	50	46.3	58	53.7	108	5.3	11,227	52.2	10,284	47.8	21,511	3.6
대전	56	51.9	52	48.1	108	5.3	10,170	52.0	9,396	48.0	19,566	3.3
울산	49	56.3	38	43.7	87	4.2	7,923	53.0	7,018	47.0	14,941	2.5
경기	171	53.4	149	46.6	320	15.6	79,722	52.1	73,224	47.9	152,946	25.7
강원	50	52.1	46	47.9	96	4.7	9,442	53.0	8,378	47.0	17,820	3.0
충북	62	47.7	68	52.3	130	6.3	9,893	52.4	8,976	47.6	18,869	3.2
충남	65	53.3	57	46.7	122	5.9	13,664	52.0	12,600	48.0	26,264	4.4
전북	72	59.5	49	40.5	121	5.9	12,115	51.9	11,231	48.1	23,346	3.9
전남	39	52.0	36	48.0	75	3.6	11,717	52.7	10,530	47.3	22,247	3.7
경북	66	58.9	46	41.1	112	5.4	15,725	53.3	13,759	46.7	29,484	4.9
경남	69	53.1	61	46.9	130	6.3	21,792	53.2	19,157	46.8	40,949	6.9
제주	33	47.8	36	52.2	69	3.4	4,206	51.4	3,969	48.6	8,175	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답 등) 4명은 사후층화의 과정이 생략됨

## 11) 6차년도 종단면 가중치

### 기본 가중치

6차년도 종단면 가중치는 1, 2, 3, 4, 5, 6차년도에 모두 응답한 학생들(1,855명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,855/2,378=78.0%, 5차년도 대비 응답률 1,855/1,913=97.0%).

6차년도 가중치 산출을 위해서 5차년도 조사를 통해 수집된 정보와 6차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 6차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 6차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 6차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

결과적으로 6차년도 응답여부 변수에 예측력이 있는 유용한 변수는 선택되지 않았으며, 따라서 모든 개체에 대하여 응답확률 추정값(1,855/1,913=0.970)이 동일하게 적용되었다.

6차년도 종단면 기본 가중치는 5차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 5\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2015년의 중학생 3학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

6차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 73> 및 <표 74>에 제시되어 있다.



<표 73> 초4 패널 6차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,855	75	1,475	322	597,037
남자	978	97	1,118	320	313,420
여자	877	75	1,475	323	283,617

<표 74> 초4 패널 6차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	975	52.6	877	47.4	1,852	100.0	312,450	52.4	283,616	47.6	596,066	100.0
서울	105	49.5	107	50.5	212	11.4	52,293	52.4	47,565	47.6	99,858	16.8
부산	55	53.9	47	46.1	102	5.5	18,925	52.7	16,998	47.3	35,923	6.0
대구	65	53.7	56	46.3	121	6.5	16,373	53.2	14,417	46.8	30,790	5.2
인천	36	51.4	34	48.6	70	3.8	17,265	51.7	16,113	48.3	33,378	5.6
광주	47	47.5	52	52.5	99	5.3	11,226	52.2	10,284	47.8	21,510	3.6
대전	51	51.0	49	49.0	100	5.4	10,170	52.0	9,396	48.0	19,566	3.3
울산	45	58.4	32	41.6	77	4.2	7,923	53.0	7,018	47.0	14,941	2.5
경기	142	54.0	121	46.0	263	14.2	79,722	52.1	73,224	47.9	152,946	25.7
강원	46	50.5	45	49.5	91	4.9	9,442	53.0	8,378	47.0	17,820	3.0
충북	61	47.7	67	52.3	128	6.9	9,893	52.4	8,976	47.6	18,869	3.2
충남	64	54.7	53	45.3	117	6.3	13,664	52.0	12,600	48.0	26,264	4.4
전북	67	59.3	46	40.7	113	6.1	12,115	51.9	11,232	48.1	23,347	3.9
전남	37	52.1	34	47.9	71	3.8	11,717	52.7	10,530	47.3	22,247	3.7
경북	61	57.5	45	42.5	106	5.7	15,725	53.3	13,759	46.7	29,484	4.9
경남	63	53.4	55	46.6	118	6.4	21,792	53.2	19,157	46.8	40,949	6.9
제주	30	46.9	34	53.1	64	3.5	4,205	51.4	3,969	48.6	8,174	1.4

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답 등)3명은 사후층화의 과정이 생략됨

## 12) 7차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(1,979/2,378=83.2%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2016년의 고등학생 1학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

6차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 75> 및 <표 76>에 제시되어 있다.

<표 75> 초4 패널 7차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,979	71	897	302	597,131
남자	1,036	106	810	302	312,465
여자	943	71	897	302	284,666

<표 76> 초4 패널 7차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,028	52.4	934	47.6	1,962	100.0	309,967	52.4	281,879	47.6	591,846	100.0
서울	112	50.0	112	50.0	224	11.4	52,191	51.7	48,706	48.3	100,897	17.0
부산	66	50.8	64	49.2	130	6.6	18,649	52.9	16,574	47.1	35,223	6.0
대구	62	53.4	54	46.6	116	5.9	16,334	53.3	14,289	46.7	30,623	5.2
인천	47	54.7	39	45.3	86	4.4	16,898	51.9	15,671	48.1	32,569	5.5
광주	50	49.5	51	50.5	101	5.1	10,969	52.2	10,055	47.8	21,024	3.6
대전	56	52.3	51	47.7	107	5.5	9,985	51.9	9,239	48.1	19,224	3.2
울산	44	55.7	35	44.3	79	4.0	7,854	53.2	6,921	46.8	14,775	2.5
경기	164	53.6	142	46.4	306	15.6	77,866	51.9	72,071	48.1	149,937	25.3
강원	48	51.6	45	48.4	93	4.7	9,600	53.6	8,317	46.4	17,917	3.0
충북	52	43.3	68	56.7	120	6.1	9,711	52.4	8,805	47.6	18,516	3.1
충남	66	53.7	57	46.3	123	6.3	14,447	52.6	13,028	47.4	27,475	4.6
전북	69	60.5	45	39.5	114	5.8	12,434	52.3	11,335	47.7	23,769	4.0
전남	37	48.7	39	51.3	76	3.9	11,837	52.6	10,675	47.4	22,512	3.8
경북	61	58.1	44	41.9	105	5.4	16,161	53.8	13,868	46.2	30,029	5.1
경남	63	53.8	54	46.2	117	6.0	20,949	53.2	18,456	46.8	39,405	6.7
제주	31	47.7	34	52.3	65	3.3	4,082	51.3	3,869	48.7	7,951	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답 등) 17명은 사후층화의 과정이 생략됨

### 13) 7차년도 종단면 가중치

#### 기본 가중치

7차년도 종단면 가중치는 1, 2, 3, 4, 5, 6, 7차년도에 모두 응답한 학생들(1,768명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,768/2,378=74.3%, 6차년도 대비 응답률 1,768/1,855=95.3%).

7차년도 가중치 산출을 위해서 6차년도 조사를 통해 수집된 정보와 7차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 7차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 7차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 7차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 77>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 7차년도 응답확률을 추정하기 위하여 사용된 변수들은 '정서문제: 신체증상-자주 열이 나는 것처럼 느껴진다'(PSY1C08), '기상 시간-비등교일'(TIM1A03)이다.

<표 78>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.598~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.954와 0.024 정도이다.

7차년도 종단면 기본 가중치는 6차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{6차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 77> 초4 패널 7차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
PSY1C08	2	7.9	0.0194
TIM1A03	4	154.1	<.0001

<표 78> 초4 패널 7차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
1,768	0.59789	0.99999	0.95413	0.02370

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2016년의 고등학생 1학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

7차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 79> 및 <표 80>에 제시되어 있다.

<표 79> 초4 패널 7차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,768	72	1,556	337	595,937
남자	927	108	1,174	336	311,077
여자	841	72	1,556	339	284,860

<표 80> 초4 패널 7차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	924	52.6	833	47.4	1,757	100.0	309,966	52.4	281,879	47.6	591,845	100.0
서울	101	50.0	101	50.0	202	11.5	52,191	51.7	48,706	48.3	100,897	17.0
부산	55	53.9	47	46.1	102	5.8	18,649	52.9	16,574	47.1	35,223	6.0
대구	60	53.1	53	46.9	113	6.4	16,334	53.3	14,289	46.7	30,623	5.2
인천	34	51.5	32	48.5	66	3.8	16,897	51.9	15,671	48.1	32,568	5.5
광주	47	50.5	46	49.5	93	5.3	10,969	52.2	10,055	47.8	21,024	3.6
대전	53	52.0	49	48.0	102	5.8	9,985	51.9	9,239	48.1	19,224	3.2
울산	41	57.7	30	42.3	71	4.0	7,854	53.2	6,921	46.8	14,775	2.5
경기	130	53.3	114	46.7	244	13.9	77,866	51.9	72,071	48.1	149,937	25.3
강원	43	49.4	44	50.6	87	5.0	9,600	53.6	8,317	46.4	17,917	3.0
충북	52	43.7	67	56.3	119	6.8	9,711	52.4	8,805	47.6	18,516	3.1
충남	65	56.0	51	44.0	116	6.6	14,447	52.6	13,028	47.4	27,475	4.6
전북	65	60.7	42	39.3	107	6.1	12,434	52.3	11,335	47.7	23,769	4.0
전남	36	49.3	37	50.7	73	4.2	11,837	52.6	10,675	47.4	22,512	3.8
경북	56	56.6	43	43.4	99	5.6	16,161	53.8	13,868	46.2	30,029	5.1
경남	58	55.8	46	44.2	104	5.9	20,949	53.2	18,456	46.8	39,405	6.7
제주	28	47.5	31	52.5	59	3.4	4,082	51.3	3,869	48.7	7,951	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답 등) 11명은 사후층화의 과정이 생략됨

### 3. 층1 패널 가중치

#### 1) 1차년도 횡단면 가중치

표본추출확률에 대한 가중치(기본 가중치)

KCYPS의 표본추출방식은 일종의 층화2단집락추출법으로서, 시도와 도시규모를 층으로 간주할 수 있다. 각 층에서 1차추출단위인 학교는 학생수를 기준으로 확률비례추출법으로 선정하고, 추출된 표본학교에서는 학급을 랜덤하게 선정한 후에 학생을 조사하였다. 따라서 추출된 표본학교에서 학생을 랜덤하게 선정하여 조사하는 2단집락추출법으로 볼 수 있으므로 표본추출률을 다음의 식으로 나타낼 수 있다.

$$f_{hij} = l_h \times \frac{M_{hi}}{\sum_{i=1}^{K_h} M_{hi}} \times \frac{n_{hi}}{M_{hi}} = \frac{l_h \times n_{hi}}{\sum_{i=1}^{K_h} M_{hi}}$$

여기서  $K_h$ 는  $h$ 층에 속한 학교의 수,  $l_h$ 는  $h$ 층에서 추출된 학교의 수,  $M_{hi}$ 는  $h$ 층  $i$ 번째 학교의 학생 수,  $n_{hi}$ 는 추출된 표본학교에서 조사한 학생수를 나타낸다.

표본추출률에 대한 가중치는 추출률의 역수이므로 아래와 같은 식으로 표현할 수 있다.

$$W_{1hij} = \frac{\sum_{i=1}^{K_h} M_{hi}}{l_h \times n_{hi}} = \frac{M_h}{l_h \times n_{hi}}$$

여기서  $M_h$ 는  $h$ 층의 전체 학생수를 나타낸다.

표본의 무응답 조정을 위한 가중치

표본으로 선정된 조사단위들 중 응답거부로 조사에 성공하지 못하거나 조사된 내용이 부실하여 무응답처리한 조사 단위에 대한 조정이 무응답 조정의 가중치로서, 다음의 식으로 계산할 수 있다.

$$W_{2hij} = \frac{n_{hi}}{r_{hi}}$$

여기서  $r_{hi}$ 와  $n_{hi}$ 는 각각 완전 응답한 표본크기와 추출된 표본크기이다.

사후층화 가중치

표본설계시에는 층화변수로 고려하지 않았으나 표본조사 완료 후에 연구변수에 영향을 미치는 변수들 중에서 정확한 외부보조정보를 확보하게 되면, 외부보조정보 데이터의 구조와 표본조사 데이터의 구조

를 분석하여 층으로 구분하고 표본조사의 각 셀에 가중치의 합계를 계산한 결과와 외부보조정보의 층별 합계를 비교하여 가중치를 계산한다.

$$BF_{hij} = \frac{PS_{hij}}{\sum W_{1hij} \cdot W_{2hij}}$$

여기서  $PS_{hij}$ 는 외부보조정보에 의한 셀의 빈도수이고, 분모의 내용은 설계가중치와 무응답 조정을 위한 가중치를 곱한 후에 합계를 계산결과이다. KCYPS에서는 16개 광역시·도×성별 사후층화를 실시하였다. 최종 가중치는 세가지 가중치를 곱하여 아래와 같은 식으로 산출한다.

$$W_{thij} = W_{1hij} \cdot W_{2hij} \cdot BF_{hij}$$

## 2) 2차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,280/2,351=97.0%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2011년의 중학생 2학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

중1 패널 2차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 81> 중1 패널 2차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,280	43	1,203	283	644,485
남자	1,152	43	1,203	291	334,919
여자	1,128	76	863	274	309,565

<표 82> 중1 패널 2차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,150	50.5	1,126	49.5	2,276	100.0	334,442	52.0	309,190	48.0	643,632	100.0
서울	96	43.6	124	56.4	220	9.7	58,224	52.3	53,143	47.7	111,367	17.3
부산	72	55.8	57	44.2	129	5.7	21,296	52.3	19,392	47.7	40,688	6.3
대구	77	50.0	77	50.0	154	6.8	18,426	53.0	16,321	47.0	34,747	5.4
인천	45	29.6	107	70.4	152	6.7	18,716	51.4	17,691	48.6	36,407	5.7
광주	38	40.4	56	59.6	94	4.1	12,047	51.6	11,281	48.4	23,328	3.6
대전	53	52.0	49	48.0	102	4.5	11,182	52.1	10,271	47.9	21,453	3.3
울산	72	66.1	37	33.9	109	4.8	8,838	52.6	7,959	47.4	16,797	2.6
경기	156	46.7	178	53.3	334	14.7	81,926	51.5	77,295	48.5	159,221	24.7
강원	72	64.9	39	35.1	111	4.9	9,916	51.3	9,400	48.7	19,316	3.0
충북	60	48.8	63	51.2	123	5.4	10,588	51.1	10,117	48.9	20,705	3.2
충남	47	51.6	44	48.4	91	4.0	13,492	51.7	12,610	48.3	26,102	4.1
전북	67	59.3	46	40.7	113	5.0	12,975	52.2	11,894	47.8	24,869	3.9
전남	41	37.3	69	62.7	110	4.8	12,486	51.8	11,636	48.2	24,122	3.7
경북	87	56.1	68	43.9	155	6.8	16,812	52.4	15,265	47.6	32,077	5.0
경남	82	56.6	63	43.4	145	6.4	23,103	52.4	20,965	47.6	44,068	6.8
제주	85	63.4	49	36.6	134	5.9	4,415	52.8	3,950	47.2	8,365	1.3

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 4명은 사후층화의 과정이 생략됨.

### 3) 2차년도 종단면 가중치

#### 기본 가중치

2차년도 가중치 산출을 위해서 1차년도 조사를 통해 수집된 정보와 2차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 2차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 2차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 2차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다. 여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).



<표 83>은 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 2차년도 응답확률을 추정하기 위하여 사용된 변수들은 ‘살고 있는 곳(시·도)’(ARA2A), ‘학생이 일주일에 컴퓨터를 사용하는 시간-주중 하루평균(분)’(MDA1B01), ‘방과 후 교육을 받은 주된 이유-1순위’(EDU5B01), ‘하고싶은 체험 활동-2순위’(ACT1D02), ‘하고싶은 체험 활동-3순위’(ACT1D01), ‘나는 나에게 만족한다’(PSY2A01), ‘부모님(보호자)께서는 나의 의견을 존중해 주신다’(FAM2B01), ‘부모님(보호자)께서는 내 친한 친구들을 만나 보신 적이 있습니까?’(FRN2B) 등이다.

<표 84>에는 응답확률 추정값에 대한 요약통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.001~1.000의 범위내에 존재하며, 평균과 표준편차는 각각 0.970과 0.055정도이다.

2차년도 종단면 기본 가중치는 1차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 83> 중1 패널 2차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
ARA2A	15	24.8300	0.0523
MDA1B01	12	22.2378	0.0349
EDU5B01	4	16.5380	0.0024
ACT1D02	4	6.8078	0.1464
ACT1D01	3	9.7425	0.0209
PSY2A01	2	7.0577	0.0293
FAM2B01	3	11.9174	0.0077
FRN2B	3	6.1916	0.1026

<표 84> 중1 패널 2차년도 응답확률 추정값의 요약 통계량

N	최소값	최대값	평균	표준편차
2,351	0.0001	1.0000	0.9698	0.0547

#### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2011년의 중학생 2학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

## 가중치의 분포

중1 패널 2차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 85> 중1 패널 2차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,280	42	2,024	283	644,539
남자	1,152	42	1,327	291	334,949
여자	1,128	75	2,024	274	309,590

<표 86> 중1 패널 2차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,150	50.5	1,126	49.5	2,276	100.0	334,442	52.0	309,190	48.0	643,632	100.0
서울	96	43.6	124	56.4	220	9.7	58,224	52.3	53,143	47.7	111,367	17.3
부산	72	55.8	57	44.2	129	5.7	21,296	52.3	19,392	47.7	40,688	6.3
대구	77	50.0	77	50.0	154	6.8	18,426	53.0	16,321	47.0	34,747	5.4
인천	45	29.6	107	70.4	152	6.7	18,716	51.4	17,691	48.6	36,407	5.7
광주	38	40.4	56	59.6	94	4.1	12,047	51.6	11,281	48.4	23,328	3.6
대전	53	52.0	49	48.0	102	4.5	11,182	52.1	10,271	47.9	21,453	3.3
울산	72	66.1	37	33.9	109	4.8	8,838	52.6	7,959	47.4	16,797	2.6
경기	156	46.7	178	53.3	334	14.7	81,926	51.5	77,295	48.5	159,221	24.7
강원	72	64.9	39	35.1	111	4.9	9,916	51.3	9,400	48.7	19,316	3.0
충북	60	48.8	63	51.2	123	5.4	10,588	51.1	10,117	48.9	20,705	3.2
충남	47	51.6	44	48.4	91	4.0	13,492	51.7	12,610	48.3	26,102	4.1
전북	67	59.3	46	40.7	113	5.0	12,975	52.2	11,894	47.8	24,869	3.9
전남	41	37.3	69	62.7	110	4.8	12,486	51.8	11,636	48.2	24,122	3.7
경북	87	56.1	68	43.9	155	6.8	16,812	52.4	15,265	47.6	32,077	5.0
경남	82	56.6	63	43.4	145	6.4	23,103	52.4	20,965	47.6	44,068	6.8
제주	85	63.4	49	36.6	134	5.9	4,415	52.8	3,950	47.2	8,365	1.3

\* 주: 기타(국제학교, 대안학교, 학교 안다님, 무응답) 4명은 사후층화의 과정이 생략됨.

## 4) 3차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,259/2,351=96.1%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 중학생 3학년 기준).

$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

중1 패널 3차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 87> 중1 패널 3차년도 횡단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,259	42	1,256	283	640,357
남자	1,140	42	1,256	292	332,982
여자	1,119	77	883	275	307,375

<표 88> 중1 패널 3차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,139	50.4	1,119	49.6	2,258	100.0	332,676	52.0	307,375	48.0	640,051	100.0
서울	91	42.7	122	57.3	213	9.4	57,649	52.3	52,675	47.7	110,324	17.2
부산	72	56.2	56	43.8	128	5.7	21,193	52.4	19,250	47.6	40,443	6.3
대구	79	51.0	76	49.0	155	6.9	18,364	53.0	16,293	47.0	34,657	5.4
인천	42	28.0	108	72.0	150	6.6	18,652	51.4	17,609	48.6	36,261	5.7
광주	38	41.8	53	58.2	91	4.0	12,025	51.6	11,266	48.4	23,291	3.6
대전	51	49.5	52	50.5	103	4.6	11,128	52.2	10,196	47.8	21,324	3.3
울산	74	67.9	35	32.1	109	4.8	8,824	52.6	7,942	47.4	16,766	2.6
경기	152	45.5	182	54.5	334	14.8	81,464	51.5	76,730	48.5	158,194	24.7
강원	71	64.5	39	35.5	110	4.9	9,901	51.4	9,366	48.6	19,267	3.0
충북	61	49.2	63	50.8	124	5.5	10,545	51.2	10,066	48.8	20,611	3.2
충남	48	52.7	43	47.3	91	4.0	13,516	51.8	12,593	48.2	26,109	4.1
전북	67	59.8	45	40.2	112	5.0	12,914	52.1	11,857	47.9	24,771	3.9
전남	41	38.3	66	61.7	107	4.7	12,425	51.7	11,603	48.3	24,028	3.8
경북	87	55.8	69	44.2	156	6.9	16,724	52.4	15,183	47.6	31,907	5.0
경남	80	56.3	62	43.7	142	6.3	22,966	52.5	20,805	47.5	43,771	6.8
제주	85	63.9	48	36.1	133	5.9	4,386	52.7	3,941	47.3	8,327	1.3

\* 주: 기타(고등학교, 국제학교, 대안학교, 학교 안다님, 무응답) 1명은 사후층화의 과정이 생략됨.

## 5) 3차년도 종단면 가중치

### 기본 가중치

3차년도 종단면 가중치는 1, 2, 3차년도에 모두 응답한 학생들(2,226명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,226/2,351=94.7%, 2차년도 대비 응답률 2,226/2,276 =97.8%).

3차년도 가중치 산출을 위해서 2차년도 조사를 통해 수집된 정보와 3차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 3차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 3차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 3차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 89>은 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 3차년도 응답확률을 추정하기 위하여 사용된 변수들은 '다른 사람 집단따돌림 시키기(DLQ3A06)', '학교숙제(시간)-비등교일(TIM3C)' 등이다.

<표 90>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.5000~1.0000의 범위 내에 존재하며, 평균과 표준편차는 각각 0.9842와 0.0746 정도이다.

3차년도 종단면 기본 가중치는 1차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{2차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 89> 중1 패널 3차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
DLQ3A06	7	14.7	<0.0406
TIM3C	2	78.0	<0.0001

<표 90> 중1 패널 3차년도 응답확률 추정값의 요약 통계량

N	최소값	최대값	평균	표준편차
2,276	0.5000	1.0000	0.9842	0.0746

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2012년의 중학생 3학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

중1 패널 3차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 91> 중1 패널 3차년도 종단면 가중치의 요약 통계량

	N	최소값	최대값	평균	합계
전체	2,226	42	3,084	288	640,051
남자	1,128	42	1,374	295	332,676
여자	1,098	76	3,084	280	307,375

<표 92> 중1 패널 3차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,128	50.7	1,098	49.3	2,226	100.0	332,676	52.0	307,375	48.0	640,051	100.0
서울	91	43.3	119	56.7	210	9.4	57,649	52.3	52,675	47.7	110,324	17.2
부산	72	56.7	55	43.3	127	5.7	21,193	52.4	19,250	47.6	40,443	6.3
대구	78	51.0	75	49.0	153	6.9	18,364	53.0	16,293	47.0	34,657	5.4
인천	42	29.0	103	71.0	145	6.5	18,652	51.4	17,609	48.6	36,261	5.7
광주	38	42.2	52	57.8	90	4.0	12,025	51.6	11,266	48.4	23,291	3.6
대전	50	49.0	52	51.0	102	4.6	11,128	52.2	10,196	47.8	21,324	3.3
울산	73	69.5	32	30.5	105	4.7	8,824	52.6	7,942	47.4	16,766	2.6
경기	151	45.8	179	54.2	330	14.8	81,464	51.5	76,730	48.5	158,194	24.7
강원	70	64.2	39	35.8	109	4.9	9,901	51.4	9,366	48.6	19,267	3.0
충북	60	48.8	63	51.2	123	5.5	10,545	51.2	10,066	48.8	20,611	3.2
충남	46	52.9	41	47.1	87	3.9	13,516	51.8	12,593	48.2	26,109	4.1
전북	67	59.8	45	40.2	112	5.0	12,914	52.1	11,857	47.9	24,771	3.9
전남	40	38.1	65	61.9	105	4.7	12,425	51.7	11,603	48.3	24,028	3.8
경북	85	55.6	68	44.4	153	6.9	16,724	52.4	15,183	47.6	31,907	5.0
경남	80	56.3	62	43.7	142	6.4	22,966	52.5	20,805	47.5	43,771	6.8
제주	85	63.9	48	36.1	133	6.0	4,386	52.7	3,941	47.3	8,327	1.3

## 6) 4차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,108/2,351=89.7%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 고등학생 1학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

4차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 93> 중1 패널 4차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,108	46	1,441	305	642,924
남자	1,075	46	1,441	311	334,522
여자	1,033	80	1,010	299	308,402

<표 94> 중1 패널 4차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,057	50.9	1,020	49.1	2,077	100.0	329,858	52.0	304,118	48.0	633,976	100.0
서울	77	41.2	110	58.8	187	9.0	56,852	51.5	53,600	48.5	110,452	17.4
부산	71	57.7	52	42.3	123	5.9	20,972	52.8	18,720	47.2	39,692	6.3
대구	74	51.4	70	48.6	144	6.9	18,050	52.9	16,055	47.1	34,105	5.4
인천	41	32.8	84	67.2	125	6.0	18,296	51.9	16,967	48.1	35,263	5.6
광주	32	39.5	49	60.5	81	3.9	11,695	51.3	11,121	48.7	22,816	3.6
대전	48	53.3	42	46.7	90	4.3	10,980	52.2	10,056	47.8	21,036	3.3
울산	72	66.1	37	33.9	109	5.2	8,734	52.8	7,819	47.2	16,553	2.6
경기	143	46.6	164	53.4	307	14.8	80,372	51.6	75,490	48.4	155,862	24.6
강원	66	66.7	33	33.3	99	4.8	10,107	52.1	9,299	47.9	19,406	3.1
충북	55	46.2	64	53.8	119	5.7	10,316	51.6	9,691	48.4	20,007	3.2
충남	45	52.3	41	47.7	86	4.1	14,225	52.5	12,860	47.5	27,085	4.3
전북	64	57.7	47	42.3	111	5.3	13,169	52.7	11,829	47.3	24,998	3.9
전남	34	35.4	62	64.6	96	4.6	12,539	51.9	11,627	48.1	24,166	3.8
경북	85	55.6	68	44.4	153	7.4	17,061	53.1	15,055	46.9	32,116	5.1
경남	77	59.7	52	40.3	129	6.2	22,241	52.5	20,085	47.5	42,326	6.7
제주	73	61.9	45	38.1	118	5.7	4,249	52.5	3,844	47.5	8,093	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 31명은 사후층화의 과정이 생략됨

## 7) 4차년도 종단면 가중치

### 기본 가중치

4차년도 종단면 가중치는 1, 2, 3, 4차년도에 모두 응답한 학생들(2,074명)만을 대상으로 계산되었다(1차년도 대비 응답률 2,074/2,351=88.2%, 2차년도 대비 응답률 2,074/2,276 =91.1%, 3차년도 대비 응답률 2,074/2,226=93.2%).

4차년도 가중치 산출을 위해서 3차년도 조사를 통해 수집된 정보와 4차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 4차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 4차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 4차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로

지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 95>은 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 4차년도 응답확률을 추정하기 위하여 사용된 변수들은 ‘무단결석\_경험 여부’(DLQ1A03), ‘모친의 종사상 지위’(JOB5B), ‘대다수 사람들이 하는 대로 그저 따라가는 것이 최선이다’(PSY2C04), ‘성폭행이나 성희롱’(DLQ1B13), ‘나와 문화적 배경이 다른 사람을 이웃으로 받아들일 수 있다’(COM3A01), ‘주소1(시/도)’(ARA2A), ‘학원/과외 숙제시간’(TIM3A) 등이다.

<표 95> 중1 패널 4차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
DLQ1A03	2	14.3	0.0008
JOB5B	2	15.8	0.0004
PSY2C04	3	18.0	0.0004
DLQ1B13	2	28.9	<.0001
COM3A01	3	13.7	0.0033
ARA2A	15	63.5	<.0001
TIM3A	7	26.5	0.0004

<표 96>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.5000~0.9999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.834와 0.201 정도이다.

<표 96> 중1 패널 4차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
2,276	0.5000	0.9999	0.8341	0.2014

4차년도 종단면 기본 가중치는 3차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{3차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2013년의 고등학생 1학년 기준).



$$\text{최종 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

4차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포는 다음의 표와 같다.

<표 97> 중1 패널 4차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,074	40	2,805	310	643,315
남자	1,062	40	2,285	316	335,183
여자	1,012	75	2,805	304	308,132

<표 98> 중1 패널 4차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,047	51.1	1,001	48.9	2,048	100.0	329,858	52.0	304,118	48.0	633,976	100.0
서울	74	40.9	107	59.1	181	8.8	56,852	51.5	53,600	48.5	110,452	17.4
부산	71	58.2	51	41.8	122	6.0	20,972	52.8	18,720	47.2	39,692	6.3
대구	74	52.1	68	47.9	142	6.9	18,050	52.9	16,055	47.1	34,105	5.4
인천	41	33.6	81	66.4	122	6.0	18,296	51.9	16,967	48.1	35,263	5.6
광주	32	39.5	49	60.5	81	4.0	11,695	51.3	11,121	48.7	22,816	3.6
대전	48	53.3	42	46.7	90	4.4	10,980	52.2	10,056	47.8	21,036	3.3
울산	72	69.2	32	30.8	104	5.1	8,734	52.8	7,819	47.2	16,553	2.6
경기	142	47.0	160	53.0	302	14.7	80,372	51.6	75,490	48.4	155,862	24.6
강원	64	66.0	33	34.0	97	4.7	10,107	52.1	9,299	47.9	19,406	3.1
충북	54	45.8	64	54.2	118	5.8	10,316	51.6	9,691	48.4	20,007	3.2
충남	44	52.4	40	47.6	84	4.1	14,225	52.5	12,860	47.5	27,085	4.3
전북	64	57.7	47	42.3	111	5.4	13,169	52.7	11,829	47.3	24,998	3.9
전남	33	34.7	62	65.3	95	4.6	12,539	51.9	11,627	48.1	24,166	3.8
경북	84	55.3	68	44.7	152	7.4	17,061	53.1	15,055	46.9	32,116	5.1
경남	77	59.7	52	40.3	129	6.3	22,241	52.5	20,085	47.5	42,326	6.7
제주	73	61.9	45	38.1	118	5.8	4,249	52.5	3,844	47.5	8,093	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 26명은 사후층화의 과정이 생략됨

## 8) 5차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,091/2,351=88.9%)의 역수를 1차

년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

최종 가중치(사후층화 가중치)

최종 가중치는 해당 표본이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2014년의 고등학생 2학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

가중치의 분포

5차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 다음의 <표 99> 및 <표 100>에 제시되어 있다.

<표 99> 중1 패널 5차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,091	41	1,326	304	635,250
남자	1,067	41	1,326	310	330,976
여자	1,024	85	986	297	304,274

<표 100> 중1 패널 5차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,033	50.8	1,002	49.2	2,035	100.0	320,558	51.9	297,218	48.1	617,776	100.0
서울	82	42.7	110	57.3	192	9.4	55,368	51.3	52,463	48.7	107,831	17.5
부산	72	59.5	49	40.5	121	5.9	20,316	52.7	18,252	47.3	38,568	6.2
대구	70	50.7	68	49.3	138	6.8	17,627	52.9	15,717	47.1	33,344	5.4
인천	40	32.0	85	68.0	125	6.1	17,733	51.6	16,614	48.4	34,347	5.6
광주	29	38.2	47	61.8	76	3.7	11,437	51.4	10,822	48.6	22,259	3.6
대전	48	52.7	43	47.3	91	4.5	10,537	52.0	9,724	48.0	20,261	3.3
울산	71	67.6	34	32.4	105	5.2	8,542	52.7	7,672	47.3	16,214	2.6
경기	135	45.9	159	54.1	294	14.4	78,190	51.4	73,850	48.6	152,040	24.6
강원	64	64.0	36	36.0	100	4.9	9,758	51.9	9,039	48.1	18,797	3.0
충북	53	46.5	61	53.5	114	5.6	9,990	51.4	9,432	48.6	19,422	3.1
충남	45	52.3	41	47.7	86	4.2	13,693	52.2	12,555	47.8	26,248	4.2
전북	62	58.3	44	41.5	106	5.2	12,849	52.6	11,588	47.4	24,437	4.0
전남	29	32.6	60	67.4	89	4.4	12,206	51.7	11,381	48.3	23,587	3.8
경북	86	55.8	68	44.2	154	7.6	16,519	53.0	14,652	47.0	31,171	5.0
경남	73	56.6	56	43.4	129	6.3	21,654	52.3	19,722	47.7	41,376	6.7
제주	74	64.3	41	35.7	115	5.7	4,139	52.6	3,735	47.4	7,874	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 56명은 사후층화의 과정이 생략됨.

## 9) 5차년도 종단면 가중치

### 기본 가중치

5차년도 종단면 가중치는 1, 2, 3, 4, 5차년도에 모두 응답한 학생들(1,995명)만을 대상으로 계산되었다 (1차년도 대비 응답률 1,995/2,351=84.9%, 4차년도 대비 응답률 1,995/2,074 =96.2%).

5차년도 가중치 산출을 위해서 4차년도 조사를 통해 수집된 정보와 5차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고 이 모형에 근거한 5차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 5차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 5차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로

지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

결과적으로 5차년도 응답여부 변수에 예측력이 있는 유용한 변수는 선택되지 않았으며, 따라서 모든 개체에 대하여 응답확률 추정값( $1,995/2,074=0.968$ )이 동일하게 적용되었다.

5차년도 종단면 기본 가중치는 4차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 4\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

#### 최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2014년의 고등학생 2학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

#### 가중치의 분포

5차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 101> 및 <표 102>에 제시되어 있다.

<표 101> 중1 패널 5차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,995	37	2,928	318	633,947
남자	1,024	37	2,326	323	330,356
여자	971	78	2,928	313	303,591

<표 102> 중1 패널 5차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	996	51.1	954	48.9	1,950	100.0	320,558	51.9	297,218	48.1	617,776	100.0
서울	70	41.2	100	58.8	170	8.7	55,368	51.3	52,463	48.7	107,831	17.5
부산	71	60.2	47	39.8	118	6.1	20,316	52.7	18,252	47.3	38,568	6.2
대구	69	51.9	64	48.1	133	6.8	17,627	52.9	15,717	47.1	33,344	5.4
인천	39	33.1	79	66.9	118	6.1	17,733	51.6	16,614	48.4	34,347	5.6
광주	28	37.8	46	62.2	74	3.8	11,437	51.4	10,822	48.6	22,259	3.6
대전	48	53.9	41	46.1	89	4.6	10,537	52.0	9,724	48.0	20,261	3.3
울산	70	70.0	30	30.0	100	5.1	8,542	52.7	7,672	47.3	16,214	2.6
경기	131	46.6	150	53.4	281	14.4	78,190	51.4	73,850	48.6	152,040	24.6
강원	59	63.4	34	36.6	93	4.8	9,758	51.9	9,039	48.1	18,797	3.0
충북	52	46.0	61	54.0	113	5.8	9,990	51.4	9,432	48.6	19,422	3.1
충남	44	52.4	40	47.6	84	4.3	13,693	52.2	12,555	47.8	26,248	4.2
전북	62	59.0	43	41.0	105	5.4	12,849	52.6	11,588	47.4	24,437	4.0
전남	28	31.8	60	68.2	88	4.5	12,206	51.7	11,381	48.3	23,587	3.8
경북	83	55.3	67	44.7	150	7.7	16,519	53.0	14,652	47.0	31,171	5.0
경남	72	58.5	51	41.5	123	6.3	21,654	52.3	19,722	47.7	41,376	6.7
제주	70	63.1	41	36.9	111	5.7	4,139	52.6	3,735	47.4	7,874	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 45명은 사후층화의 과정이 생략됨.

## 10) 6차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(2,056/2,351=87.5%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = 1\text{차년도 가중치} \times \frac{1}{1\text{차년도 대비 응답률}}$$

### 최종 가중치(사후층화 가중치)

최종 가중치는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를 고려하여 다음과 같이 계산되었다(2015년의 고등학생 3학년 기준).

$$\text{사후층화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

## 가중치의 분포

6차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 다음의 <표 103> 및 <표 104>에 제시되어 있다.

<표 103> 중1 패널 6차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	2,056	46	1,323	304	624,373
남자	1,041	46	1,323	311	324,148
여자	1,015	85	998	296	300,225

<표 104> 중1 패널 6차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	1,014	50.5	995	49.5	2,009	100.0	315,710	51.8	293,434	48.2	609,114	100.0
서울	81	42.6	109	57.4	190	9.5	54,648	51.3	51,889	48.7	106,537	17.5
부산	69	58.0	50	42.0	119	5.9	19,935	52.6	18,000	47.4	37,935	6.2
대구	70	50.7	68	49.3	138	6.9	17,406	52.8	15,552	47.2	32,958	5.4
인천	41	31.3	90	68.7	131	6.5	17,399	51.4	16,449	48.6	33,848	5.6
광주	28	37.3	47	62.7	75	3.7	11,266	51.4	10,670	48.6	21,936	3.6
대전	47	51.1	45	48.9	92	4.6	10,332	51.9	9,578	48.1	19,910	3.3
울산	68	68.0	32	32.0	100	5.0	8,390	52.5	7,582	47.5	15,972	2.6
경기	138	47.1	155	52.9	293	14.6	77,135	51.4	73,010	48.6	150,145	24.6
강원	61	64.9	33	35.1	94	4.7	9,613	51.8	8,928	48.2	18,541	3.0
충북	51	45.1	62	54.9	113	5.6	9,834	51.4	9,293	48.6	19,127	3.1
충남	45	51.7	42	48.3	87	4.3	13,499	52.1	12,391	47.9	25,890	4.3
전북	57	58.2	41	41.8	98	4.9	12,676	52.6	11,435	47.4	24,111	4.0
전남	30	33.7	59	66.3	89	4.4	11,948	51.7	11,177	48.3	23,125	3.8
경북	85	56.3	66	43.7	151	7.5	16,258	52.9	14,451	47.1	30,709	5.0
경남	72	56.2	56	43.8	128	6.4	21,299	52.4	19,346	47.6	40,645	6.7
제주	71	64.0	40	36.0	111	5.5	4,072	52.5	3,683	47.5	7,755	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 56명은 사후층화의 과정이 생략됨.

## 11) 6차년도 종단면 가중치

### 기본 가중치

6차년도 종단면 가중치는 1, 2, 3, 4, 5, 6차년도에 모두 응답한 학생들(1,926명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,926/2,351=81.9%, 5차년도 대비 응답률 1,926/1,995=96.5%).

6차년도 가중치 산출을 위해서 5차년도 조사를 통해 수집된 정보와 6차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 6차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 6차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 6차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거 후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

<표 105>에는 변수선택 과정을 통해 선택된 변수들과 로지스틱 회귀분석의 결과를 보여주고 있다. 6차년도 응답확률을 추정하기 위하여 사용된 변수들은 '비행: 연간 행동경험 친구수-무단결석'(DLQ3A03)이다.

<표 106>에는 응답확률 추정값에 대한 요약 통계량이 제시되어 있다. 응답확률의 추정값은 대략 0.914~0.999의 범위 내에 존재하며, 평균과 표준편차는 각각 0.966과 0.009 정도이다.

6차년도 종단면 기본 가중치는 5차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = 5\text{차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

<표 105> 중1 패널 6차년도 로지스틱 회귀분석을 이용한 변수선택 결과

변수	자유도	Wald Chi-Square	p-value
DLQ3A03	2	46.4	<0.0001

<표 106> 중1 패널 6차년도 응답확률 추정값의 요약통계량

N	최소값	최대값	평균	표준편차
1,926	0.91429	0.99999	0.96598	0.00856

최종 가중치(사후층화 가중치)

최종 가중치(사후층화 가중치)는 해당 학생이 속한 지역별/성별 기본 가중치의 합계와 모집단 크기를

고려하여 다음과 같이 계산되었다(2015년의 고등학생 3학년 기준).

$$\text{사후총화 가중치} = \text{기본 가중치} \times \frac{\text{지역별/성별 모집단 크기}}{\text{지역별/성별 기본 가중치의 합계}}$$

### 가중치의 분포

6차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 107> 및 <표 108>에 제시되어 있다.

<표 107> 중1 패널 6차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,926	38	3001	323	621,968
남자	983	38	2404	328	322,897
여자	943	80	3001	317	299,071

<표 108> 중1 패널 6차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	963	50.9	928	49.1	1,891	100.0	315,710	51.8	293,434	48.2	609,144	100.0
서울	67	40.6	98	59.4	165	8.7	54,648	51.3	51,889	48.7	106,537	17.5
부산	68	59.6	46	40.4	114	6.0	19,935	52.6	18,000	47.4	37,935	6.2
대구	69	51.9	64	48.1	133	7.0	17,406	52.8	15,552	47.2	32,958	5.4
인천	40	33.9	78	66.1	118	6.2	17,399	51.4	16,449	48.6	33,848	5.6
광주	26	36.6	45	63.4	71	3.8	11,266	51.4	10,670	48.6	21,936	3.6
대전	47	54.0	40	46.0	87	4.6	10,332	51.9	9,578	48.1	19,910	3.3
울산	67	72.0	26	28.0	93	4.9	8,390	52.5	7,582	47.5	15,972	2.6
경기	130	47.3	145	52.7	275	14.5	77,135	51.4	73,010	48.6	150,145	24.6
강원	56	64.4	31	35.6	87	4.6	9,613	51.8	8,928	48.2	18,541	3.0
충북	50	44.6	62	55.4	112	5.9	9,834	51.4	9,293	48.6	19,127	3.1
충남	44	51.8	41	48.2	85	4.5	13,499	52.1	12,391	47.9	25,890	4.3
전북	57	58.2	41	41.8	98	5.2	12,676	52.6	11,435	47.4	24,111	4.0
전남	28	32.2	59	67.8	87	4.6	11,948	51.7	11,177	48.3	23,125	3.8
경북	82	55.8	65	44.2	147	7.8	16,258	52.9	14,451	47.1	30,709	5.0
경남	66	57.9	48	42.1	114	6.0	21,299	52.4	19,346	47.6	40,645	6.7
제주	66	62.9	39	37.1	105	5.6	4,072	52.5	3,683	47.5	7,755	1.3

\* 주 : 기타(국제학교, 대안학교, 학교 안다님, 무응답) 35명은 사후총화의 과정이 생략됨.



## 12) 7차년도 횡단면 가중치

### 기본 가중치

기본 가중치는 표본탈락률을 보정하기 위하여 1차년도 대비 응답률(1,881/2,351=80.0%)의 역수를 1차년도 횡단면 가중치에 곱하여 계산되었다.

$$\text{기본 가중치} = \text{1차년도 가중치} \times \frac{1}{\text{1차년도 대비 응답률}}$$

### 가중치의 분포

7차년도 횡단면 가중치의 요약 통계량 및 지역별/성별 분포가 다음의 <표 109> 및 <표 110>에 제시되어 있다.

<표 109> 중1 패널 7차년도 횡단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,881	53	1,468	346	650,616
남자	927	53	1,468	363	336,243
여자	954	95	1,001	330	314,373

<표 110> 중1 패널 7차년도 횡단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	927	49.3	954	50.7	1,881	100.0	336,243	51.7	314,373	48.3	650,616	100.0
서울	77	46.4	89	53.6	166	8.8	57,924	57.0	43,770	43.0	101,694	15.6
부산	62	56.9	47	43.1	109	5.8	22,508	54.1	19,116	45.9	41,624	6.4
대구	57	45.2	69	54.8	126	6.7	17,171	48.5	18,256	51.5	35,427	5.4
인천	33	27.5	87	72.5	120	6.4	16,541	49.0	17,218	51.0	33,759	5.2
광주	26	37.7	43	62.3	69	3.7	9,548	47.8	10,409	52.2	19,957	3.1
대전	41	49.4	42	50.6	83	4.4	9,613	46.7	10,963	53.3	20,576	3.2
울산	55	63.2	32	36.8	87	4.6	8,354	51.1	8,000	48.9	16,354	2.5
경기	131	47.3	146	52.7	277	14.7	84,046	52.6	75,873	47.4	159,919	24.6
강원	24	44.4	30	55.6	54	2.9	4,165	32.4	8,673	67.6	12,838	2.0
충북	51	48.1	55	51.9	106	5.6	11,340	51.1	10,873	48.9	22,213	3.4
충남	40	48.8	42	51.2	82	4.4	13,890	48.6	14,669	51.4	28,559	4.4
전북	58	58.0	42	42.0	100	5.3	13,564	50.0	13,543	50.0	27,107	4.2
전남	35	37.2	59	62.8	94	5.0	13,088	52.2	11,995	47.8	25,083	3.9
경북	72	54.5	60	45.5	132	7.0	17,256	50.7	16,764	49.3	34,020	5.2
경남	66	54.5	55	45.5	121	6.4	23,604	51.1	22,569	48.9	46,173	7.1
제주	58	69.0	26	31.0	84	4.5	3,736	57.8	2,728	42.2	6,464	1.0
결측	41	57.7	30	42.3	71	3.8	9,895	52.5	8,954	47.5	18,849	2.9

### 13) 7차년도 종단면 가중치

#### 기본 가중치

7차년도 종단면 가중치는 1, 2, 3, 4, 5, 6, 7차년도에 모두 응답한 학생들(1,723명)만을 대상으로 계산되었다(1차년도 대비 응답률 1,723/2,351=73.3%, 6차년도 대비 응답률 1,723/1,926=89.5%).

7차년도 가중치 산출을 위해서 6차년도 조사를 통해 수집된 정보와 7차년도 응답여부 변수를 이용한 적절한 통계적 모형을 설정하고, 이 모형에 근거한 7차년도 조사대상자의 응답확률 추정값을 계산하였다. 이를 위해 7차년도의 응답여부 변수를 응답의 경우 '1' 그리고 무응답의 경우 '0'을 갖는 이항변수로 정의하고 로지스틱 회귀모형을 이용한 응답확률 예측방법을 적용하였다.

로지스틱 회귀모형의 적합에 앞서 주요 설명변수의 응답률 예측에 대한 효율성을 알아보기 위하여 먼저 각 설명변수에 대하여 7차년도 응답자와 무응답자 간의 차이를 비교하는 것이 필요하다. 응답자 그룹과 무응답자 그룹 간의 차이를 변수별로 분석하기 위해 근사  $\chi^2$ -검정을 시행하였다. 각 범주별 빈도수가 적어서  $\chi^2$ -검정의 결과를 신뢰할 수 없는 경우 빈도수가 적은 범주를 묶은 후에 검정을 실시하였다.

여러 설명변수들의 동시적 예측력을 반영하며 동시에 최대우도 추정량의 유일성을 살피기 위하여 로지스틱 회귀모형의 변수선택 방법을 고려하였다. 변수선택을 위한 방법으로 각 변수들의 추가 또는 제거

후 모형의 설명력에 대한 Wald의 적합도 검정을 바탕으로 한 단계적 방법(stepwise method)을 사용하였다( $\alpha=0.15$ ).

결과적으로 7차년도 응답여부 변수에 예측력이 있는 유용한 변수는 선택되지 않았으며, 따라서 모든 개체에 대하여 응답확률 추정값( $1,723/1,926=0.895$ )이 동일하게 적용되었다.

7차년도 종단면 기본 가중치는 6차년도 종단면 가중치에 응답확률 추정값의 역수를 곱하여 계산되었다.

$$\text{기본 가중치} = \text{6차년도 가중치} \times \frac{1}{\text{응답확률 추정값}}$$

### 가중치의 분포

7차년도 종단면 가중치의 요약 통계량 및 지역별/성별 분포가 <표 111> 및 <표 112>에 제시되어 있다.

<표 111> 중1 패널 7차년도 종단면 가중치의 요약통계량

	N	최소값	최대값	평균	합계
전체	1,723	43	3,354	365	629,612
남자	853	43	2,687	378	322,804
여자	870	90	3,354	353	306,808

<표 112> 중1 패널 7차년도 종단면 가중치의 지역별/성별 분포

	가중치 부여 이전						가중치 부여 이후					
	남자		여자		전체		남자		여자		전체	
	N	%	N	%	N	%	N	%	N	%	N	%
전체	853	49.5	870	50.5	1723	100.0	322,804	51.3	306,808	48.7	629,612	100.0
서울	62	44.6	77	55.4	139	8.1	57,375	55.9	45,349	44.1	102,724	16.3
부산	58	57.4	43	42.6	101	5.9	19,476	51.2	18,595	48.8	38,071	6.0
대구	54	46.2	63	53.8	117	6.8	14,529	46.1	17,000	53.9	31,529	5.0
인천	30	28.8	74	71.2	104	6.0	15,085	45.4	18,150	54.6	33,235	5.3
광주	23	35.9	41	64.1	64	3.7	10,508	48.7	11,086	51.3	21,594	3.4
대전	39	50.6	38	49.4	77	4.5	9,682	48.0	10,495	52.0	20,177	3.2
울산	54	71.1	22	28.9	76	4.4	7,423	50.2	7,358	49.8	14,781	2.3
경기	121	47.1	136	52.9	257	14.9	82,748	52.3	75,570	47.7	158,318	25.1
강원	23	44.2	29	55.8	52	3.0	4,510	33.1	9,135	66.9	13,645	2.2
충북	48	46.6	55	53.4	103	6.0	10,409	583.6	9,028	46.4	19,437	3.1
충남	40	49.4	41	50.6	81	4.7	13,213	49.4	13,525	50.6	26,738	4.2
전북	53	57.0	40	43.0	93	5.4	13,249	51.0	12,715	49.0	25,964	4.1
전남	30	34.5	57	65.5	87	5.0	14,270	54.9	11,743	45.1	26,013	4.1
경북	70	53.8	60	46.2	130	7.5	16,404	51.5	15,438	48.5	31,842	5.1
경남	61	58.1	44	41.9	105	6.1	22,239	52.6	20,080	47.4	42,319	6.7
제주	51	67.1	25	32.9	76	4.4	3,610	57.8	2,635	42.2	6,245	1.0
결측	36	59.0	25	41.0	61	3.5	8,074	47.6	8,906	52.4	16,980	2.7

## □ KCYPS 데이터 활용 유의사항

1. KCYPS 데이터에 대한 저작권과 지적소유권은 한국청소년정책연구원이 보유하고 있으며, 이 권리는 대한민국의 저작권법과 국제저작권 조약에 의하여 보호됩니다.
2. KCYPS 데이터는 학술 연구 및 정책 개발 등 비영리적인 목적으로만 이용 가능하며, 타인에게 판매·양도할 수 없습니다.
3. KCYPS 데이터를 활용하여 연구논문을 집필할 경우 다음 사항에 유의하여 주시기 바랍니다.
  - 1) 데이터 출처 표기 : “한국청소년정책연구원”(기관명)에서 수행한 “한국아동·청소년패널조사”(조사명) 데이터를 활용하였음을 명시
  - 2) 최신 데이터 활용 : 데이터는 부정기적으로 업데이트되므로 항상 최신의 데이터를 내려받아 사용
  - 3) 개인 정보 보호 : 조사대상 청소년과 관련된 구체적인 신상정보가 연구내용에 포함되어서는 안 됨
  - 4) 조사명 약어 사용 : 조사명을 축약하여 사용할 경우, 영문은 “KCYPS”, 국문은 “아동·청소년패널”로 표기
  - 5) 주제어 표기 : 학술논문의 경우 주제어에 “KCYPS” “한국아동·청소년패널조사” 중 하나를 포함
4. KCYPS 데이터를 활용한 연구결과가 출간되었을 경우, 그 내용(저자명, 논문제목, 학술지명 등)을 본원 담당자에게 알려 주시기 바랍니다.
  - ▶ 연락처 : 정윤미 (☎ 044-415-2188, whandym@nypi.re.kr)

## 참고문헌

- 김지경, 백혜정, 임희진, 이계오. (2010). 한국아동·청소년패널조사 2010 I. 서울: 한국청소년정책연구원.
- 박창남, 김희진, 김선업. (2001). 청소년패널조사 기초연구. 서울: 한국청소년정책연구원.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*.  
Harvard University Press.
- Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data Cleaning: Detecting,  
Diagnosing, and Editing Data Abnormalities. *PLoS Med*, 2(10), 966-970.