# 한국아동·청소년패널조사자료에서 단위무응답의 실태 및 가중치 적용

이화정 $^1$  · 강석복 $^2$ 

<sup>1</sup> 경북대학교 생태환경대학·<sup>2</sup> 영남대학교 통계학과 접수 2014년 9월 24일, 수정 2014년 10월 21일, 게재확정 2014년 10월 27일

#### 요 약

설문조사에서는 대부분 단위무응답 또는 항목무응답이 발생한다. 무응답 발생비율이 높은 경우무응답을 무시한 분석은 잘못된 결과를 초래할 수도 있으므로 무응답에 대한 특성 파악도 필요하다. 횡단자료에서는 항목무응답 특성에 대한 연구는 가능하나 단위무응답의 특성을 파악하기가 어렵다. 단위무응답의 특성을 파악하기 위하여 본 논문에서는 종단자료인 한국아동·청소년패널조사 중 중1패 별자료를 이용하였다. 최근까지 공개된 한국아동·청소년패널조사자료를 이용한 논문들의 무응답 처리 실태를 파악하고, 그 중 단위무응답의 특성에 대하여 분석하였다. 이 자료를 이용한 논문에서는 대부분 단위무응답의 처리를 제거하는 방식을 취하고 있으며, 이 자료에서 제공하고 있는 가중치를 이용한 논문은 거의 없었다. 이에 단위무응답의 처리방법 중 가중치를 이용한 방법과 그 대상을 제외하는 두 가지 방법에 대해 여러 가지 특성들의 분석을 비교하였다. 많은 항목에서 두 가지 방법의 결과가 상이하게 나타났으며, 가중치를 이용한 경우가 제거법에 의한 결과에 비해 통계적으로 유의한 결과가더 많이 나타나 무응답에 대해 많은 논의가 필요할 것으로 생각된다.

주요용어: 가중치, 단위무응답, 한국아동·청소년패널조사, 항목무응답.

# 1. 서론

최근에 많은 공공기관에서 대규모 조사 자료를 공개하고 있는 추세이다. 이는 연구자가 개별적으로 하는 조사에 비해 객관적인 조사설계가 이루어지며 특정지역이 아닌 전국단위의 조사가 이루어지므로 좀더 양질의 데이터을 이용한 연구가 가능하게 되었다. 이로 인해 동일한 자료를 바탕으로한 다양한 연구가 쏟아져 나오고 있다. 개별적으로 실시한 설문조사를 이용한 연구논문들의 경우에는 연구자마다 조사설계 및 표본 추출방법에 따라 자료가 상이하여 논문들간의 비교가 이루어지기 어려웠으나, 동일한 원시자료를 이용한 분석들의 경우는 그 결과 비교가 한층 용이해지고 있다.

많은 기관에서 오랜기간에 걸쳐 자료조사가 이루어지고 있으나, 동일한 대상을 일정한 시간차이를 두고 반복적으로 측정하는 패널조사의 경우는 1993년 대우경제연구소의 한국가구패널조사 (KHPS; Korea Household Panel Study)가 우리나라 최초의 패널조사이었으며 공적 영역으로는 한국노동연구원에서 1998년 한국노동패널조사가 처음으로 시작되었다. 그 이후로는 고령화연구패널, 한구청소년패널조사, 한국교육고용패널조사, 한국교육종단연구조사, 여성가족패널조사, 국민노후보장패널조사, 한국복지패널조사, 청년패널조사, 한국아동패널조사 등이 있다. 공공부문의 패널조사는 2010년 5월 기준 19종에 이르며 노동패널을 제외한 18개 패널이 2001년 이후 시작되었다. 이와같이 여러 분야에서 수년간 패

<sup>&</sup>lt;sup>1</sup> (702-701) 대구 북구 대학로 80, 경북대학교 생태환경대학, 강사.

<sup>&</sup>lt;sup>2</sup> 교신저자: (712-749) 경북 경산시 대학로 280, 영남대학교 통계학과, 교수. E-mail: sbkang@yu.ac.kr

널조사가 진행되고 있고 패널조사의 많은 장점에도 무시할수 없는 문제점 중 하나가 패널 탈락 (panel attrition) 즉, 단위무응답의 발생이다. 조사가 반복될수록 1차년도 원표본에 대해 계속 조사를 하기가 어렵다. 특히, 표본 탈락이 비체계적으로 발생한 경우 남아 있는 패널의 대표성을 유지하기 어려운 점이 있다. 동일 집단이나 개인식별자를 적용하지 못하거나 조사기간 동안 측정변수가 추가되거나 수정되어 아직은 종단연구에 한계가 있다 (Lee 등,2008; Shin, 1998; Lee, 2005; Lee와 Kang, 2012).

각 분야마다 많은 종류와 양의 설문조사가 이루어지고 있어 조사자들의 응답거부가 점점 더 심해지고 있다. 과거에 비해 여론조사기관에서 발표되는 응답율이 최근 현저히 낮아지고 있는 추세이다. 무응답의 비중이 높아지고 오히려 응답자의 비율이 무응답자의 비율보다 낮은 경우가 많아 응답자들의 결과만으로 일반화하는 것은 잘못된 결론을 초래할 수 있으므로 무응답에 대한 논의와 연구가 필요하다. 패널조사자료의 경우는 횡단자료에 비해 단위무응답에 대한 연구가 가능함에도 불구하고 그 연구가 많지 않은 실정이다.

무응답의 특성에 관련된 연구로 Hirschi와 Gottfredson (1993)은 자기통제력이 낮은 사람이 자기통제력이 높은 사람에 비해 부정확한 응답을 하는 경향이 있다고 하였으며, Watkins와 Melde (2007)은 자기통제력이 낮은 청소년들이 항목무응답을 발생하는 경향을 보인다고 하였다. Kim과 Chung (2012)은 항목무응답은 자기통제력이 떨어지는 대상에서 빈번하게 발생할 수 있으며 단위무응답은 자기통제보다는 권위에 대해 거부하는 비행성향과 관련될 수 있다고 하였다. 패널자료를 이용한 연구의 대부분이 단위무응답과 항목무응답이 발생되는 경우 그 항목 또는 대상을 제외시켜 분석하고 있다. 또한 무응답이 발생된 경우 결측이 무작위로 발생했다고 가정하고 결측치를 처리하는 방법은 문제가 될수 있다 (Watkins와 Melde, 2007; Kim과 Chung, 2012). 이에 정확한 연구를 위해서는 무응답의 특성을 연구하고 무응답 처리에 대한 적절한 처리에 대한 이해가 필요할 것으로 생각된다. 본 연구에서는 한국아동·청소년패널조사 (KCYPS; Korea Childeren and Youth Panel Survey)를 이용하여 무응답에 대한 처리에 대한 실태를 파악하고 무응답자들의 특성을 파악하고자 하였다. 한국아동·청소년패널조사와 같이 학교를 기반으로 하여 수집된 아동 및 청소년 자료의 장점으로는 성인에 비해 타당도가높고 사회적 바람직성 편향에 영향을 덜 받는다고 하였다 (Junger-Tas와 Marshall, 1999; Watkins와 Melde, 2007; Kim과 Chung, 2012). 2010년에 시작된 한국아동·청소년패널조사는 2016년까지 3개 연령코호트를 대상으로 시행될 예정이며 7개년에 걸쳐 현재 2012년까지 3차 자료까지 공개되어 있다.

2010년 초등학교 1학년, 초등학교 4학년, 중학교 1학년을 대상으로 층화다단계집락추출 방법으로 표본을 추출하였으며, 패널조사의 특성상 조사가 거듭될수록 발생되는 표본이탈로 인한 대표성 문제를 가중치로 보정하는 방식을 채택하였다 (National Youth Policy Institute, 2013). 1차 표본추출은 시도별, 도시규모별, 학제별 (중1패널의 경우)로 층으로 모집단의 인구구성에 따라 16개 시·도별로 비례배분하여 1단계 학교를 추출한 후 표본학교에서 최소한 2개 학급 이상이고 학생수가 50명 이상인 학교에서 1개 학급을 추출하는 방식으로 하였다.

## 2. 무응답 처리 실태

설문조사의 무응답은 크게 단위무응답 (unit nonresponse)과 항목무응답 (item nonresponse)으로 나 뉜다. 단위무응답은 조사 대상자로부터 전혀 정보를 얻지 못한 경우이며 항목무응답은 질문 문항 중 일부 문항에 대해서 응답하지 않은 경우를 말한다. 단위무응답의 처리 방법으로 가중값 조정 (weighting adjustment)과 무응답 자료를 분석에서 제외하는 방법이 있으며 항목무응답의 처리방법으로는 크게 완전제거법 (list-wise deletion), 단일대체방법 (single imputation)과 다중대체방법 (multiple imputation) 등이 있다.

많은 논문들에서 무응답에 대해 어떻게 처리하고 있는지를 파악하기 위해 최근에 공개한 KCYPS자

료를 이용하여 연구한 논문들을 대상으로 분석하였다. KCYPS의 1차 자료는 인구특성문항에 대해서는 모두 측정된 완전한 자료이며, 일부 설문문항에서 항목무응답이 발생되어 있다. 2차와 3차 자료에서 이 탈된 대상들로 인해 단위무응답이 발생되어 있으며, 1차 자료와 마찬가지로 항목무응답이 존재한다.

Table 2.1	Processing	condition	of nonres	ponse in	KCYPS
-----------	------------	-----------	-----------	----------	-------

content		frequency	%	frequency	%
year	2011	3	4.6		
	2012	17	26.2		
	2013	32	49.2		
	2014	13	20.0		
number of panels	1	61	93.8		
	2	2	3.1		
	3	2	3.1		
target	elementary school 1st grade	2	3.1		
	elementary school 4th grade	12	18.5		
	middle school 1st grade	47	72.3		
	elementary school 4th, Middle school 1st	2	3.1		
	elementary school 1st, 4th, Middle school 1st	2	3.1		
use panel	1st panel	31	47.7		
	2nd panel	23	35.4		
	1st,2nd panel	8	12.3		
	2nd,3rd panel	1	1.5		
	1st,2nd,3rd panel	2	3.1		
unit nonresponse	no mention	51	78.5	21	61.8
	mention	14	21.5	13	38.2
item nonresponse	no mention	60	92.3		
	mention	5	7.7		
total		65	100.0	34	100.0

KCYPS는 현재도 조사가 진행 중이며 현재 3차 자료까지만 공개된 상태이므로 이 자료를 이용한 논 문들에 대한 무응답 처리의 특성으로 일반화하기는 성급하다. 다만, 현재 많은 논문들이 무응답을 어떻 게 처리하고 있는지를 파악하기에는 충분하리라고 생각된다. Table 2.1은 KCYPS자료를 이용한 논문 들의 자료이용 현황을 분석한 결과이다. 분석에서 사용된 논문은 학술논문검색에서 '한국아동청소년패 널조사'와 'KCYPS'로 검색된 논문 중 2010년부터 실시되고 있는 자료를 이용한 논문이며, 연구재단의 등재 및 등재후보지의 게재된 논문을 그 대상으로 하였다. 논문의 수는 총 65개로 투고된 저널의 종류 는 29개로 나타났다. 2013년에 발표된 논문이 49.2%로 가장 많았으며 2014년 발표된 논문은 2014년 6월까지 자료인 것을 고려한다면 발표논문의 수는 매년 점점 그 수가 많아지고 있다고 볼 수 있다. 대 부분의 연구에서는 주로 1개 패널을 이용하였으며 그 중1패널 (middle school 1st grade)이 72.3%로 가장 많이 연구되었고 초4패널 (elementary school 4th grade)이 18.5%, 초1패널 (elementary school 1st grade)은 3.1%로 가 연구에 이용되었다. 단위무응답에 대한 언급은 전체의 21.5%만이 언급하였으 며 단위무응답의 경우 종단자료의 특성상 1차 자료는 단위무응답이 발생되지 않았으므로 2차 이상의 자 료를 이용한 34개 논문 중에서는 38.2%가 단위무응답 및 응답자에 대해 언급되었다. 항목무응답에 대 한 처리는 전체의 7.7%가 언급되어 연구대상과 관련된 단위무응답이 항목무응답에 비해서는 상대적으 로 많이 언급되었다. 단위무응답에 대한 언급이 있는 경우 분석에서 제외하는 방법이 적용되었으며 분 석에 사용된 자료에서는 가중치를 적용한 경우는 없었다.

일반적으로는 단위무응답에 대한 연구가 쉽지 않으나, KCYPS와 같이 측정된 종단자료는 2차와 3차에서 발생된 단위무응답을 1차 자료와 비교하면 단위무응답이 발생된 경우의 특성을 파악할 수 있다. 항목 무응답은 중1패널의 측정항목이 800여개에 이르는 방대한 양으로 항목무응답의 특성은 전체 대상에서 일반화되기는 어려울 것으로 판단되어 단위무응답의 특성을 파악하는데 주력하였다.

### 3. 단위 무응답 처리

#### 3.1. 단위무응답의 특성

Table 2.1의 결과에서 KCYPS자료 중 가장 많이 연구되고 있는 대상이 중1 패널인 것으로 나타나 본 연구에서는 3개 패널 중 중1패널의 단위무응답자들의 특성을 소개하고자 한다.

중1 패널의 1차 자료의 대상은 청소년 2351명과 그 보호자들이며 남학생 1176명 (50.0%)과 여학생 1175명 (50.0%)이다. 2차 자료는 청소년이 2280명 (97.0%), 보호자가 2220명 (94.4%)이며 3차 자료는 청소년이 2259명 (96.1%), 보호자가 2226명 (94.7%)로 단위무응답의 비율이 청소년은 3~4%이며 보호자는 5%를 초과한 것으로 나타났다. 단위무응답의 인구통계학적 특성을 파악하기 위해 조사참여 여부와 인구통계학적 특성에 대한 연관성 분석 결과를 Table 3.1에 제시하였다.

2011년에 조사된 2차 자료에서는 조사참여여부 즉, 단위무응답여부가 인구특성과의 연관성이 발생된 항목들을 살펴보면 청소년 응답에서는 성별 (gender), 자택 위치 (home address), 학교 위치 (schol address), 연간 전학경험 유무 (change of schools), 연간 결석 경험유무 (absence), 방과 후 교육 연간 경험유무 (after school education)였다. 보호자의 경우 청소년 응답과 마찬가지로 성별, 자택 위치,학교 위치에서는 동일하게 연관성이 발생하였다. 2012년의 3차 자료에서는 청소년 응답은 2차 자료와달리 학교제 (school)가 포함되었으며 성별, 자택 위치,학교 위치,연간 결석경험 유무,해외방문 경험유무 (overseas visit)였으며,보호자 응답은 성별,자택 위치,학교 위치에서만 연관성이 나타났다. KCYPS자료를 이용한 많은 연구에서 성별은 빠지지 않고 포함되었으며,부모 학력,가구소득,부모구성,형제자매 유무와 같은 특성이 상대적으로 많이 활용되었다. 이 중 성별은 2차와 3차 자료 모두 단위무응답의 발생여부와 연관성이 있으며 많은 연구에서 기본적으로 사용된 특성이다.

Table 3.1	Analysis	of survey	participation	(Y	/N	)

	J	J 1	\ / /	
	2nd panel(2011)		3rd panel (2012)	
content	youth	guardian	youth	guardian
	$\chi^2$ (p-value)	$\chi^2$ (p-value)	$\chi^2$ (p-value)	$\chi^2$ (p-value)
school (men, girl, coed)	.120 (.942)	2.604 (.272)	6.433(.040)	2.449(.294)
gender	7.703(.006)	4.297(.038)	4.543(.033)	6.184 (.013)
home address	27.517(.025)	61.312 (.000)	46.608(.000)	62.149(.000)
school address	27.391 (.026)	61.077(.000)	48.317(.000)	62.364(.000)
house style *	2.886 (.823)	15.908 (.000)	3.829(.700)	3.802(.703)
relationship of guardian	4.879 (.559)	2.489(.014)	5.634(.465)	4.127(.659)
eduation (father) *	4.479(.345)	3.616(.460)	4.288(.368)	2.934(.569)
eduation (mother) *	1.432 (.839)	2.277(.685)	8.217(.084)	7.736(.102)
work (father) *	.698 (.403)	.231 (.631)	.000(.984)	1.077(.299)
work (mother) *	.100 (.752)	.002 (.969)	.195(.659)	.349(.555)
family *	1.913 (.861)	7.717 (.179)	2.934(.710)	4.735(.449)
parent *	6.535 (.258)	18.136 (.003)	2.648(.754)	4.815(.439)
multicultural home *	1.636 (.201)	1.415(.234)	.009 (.924)	1.630(.202)
brother, sister	2.812(.094)	4.682(.030)	.173(.678)	.007(.932)
change of schools	5.159 (.023)	1.808 (.179)	2.660(.103)	.814(.367)
absence	4.828 (.028)	1.375(.241)	12.712(.000)	3.775(.052)
after school education	5.631 (.018)	3.558 (.059)	4.396(.036)	1.861(.172)
household economy burden	.807 (.848)	1.271 (.736)	2.881(.410)	6.051(.109)
overseas visit	.008 (.927)	2.177(.140)	8.329(.004)	8.498(.004)
study abroad *	.123 (.726)	.009 (.924)	2.554(.110)	.686(.408)
religion	6.791 (.147)	2.671 (.614)	9.782(.044)	.737(.947)
visit frequency (religion) *	1.383 (.710)	.259 (.967)	3.535 (.316)	2.437(.487)
* '1 '.				

<sup>\*:</sup> exit the item nonresponse

위의 분석자료에서 단위무응답의 발생여부와 연관성이 발생되고, 다른 논문에서도 많이 활용된 특성 중 항목 무응답이 발생되지 않은 것이 성별인 것으로 나타났다.

### 3.2. 가중치 적용 여부 비교

KCYPS 자료는 표본추출확률에 대한 기본가중치와 무응답 조정, 사후층화의 가중치를 반영하여 횡단적 가중치와 종단적 가중치를 제공하고 있다. 1차년도 자료는 횡단적 가중치만 제공이 되면 2차년도 이후 자료는 횡단적 가중치와 종단적 가중치가 제공되고 있다. 그러나, 상세한 가중치 사용에 대한 지침이 없으며 연구자에 일임하고 있어 주로 이 자료를 활용하고 있는 연구자들의 가중치 사용을 주저하게하는 요인인 것으로 판단된다. 실제 KCYPS를 이용한 무응답실태 결과에서 단위무응답의 경우 가중치를 이용한 경우는 없는 것으로 나타나 다른 분야에 비해 아직 KCYPS자료를 이용한 논문에서는 가중치를 거의 사용하고 있지 않다.

먼저 KCYPS자료에서 제공되고 있는 가중치 중 횡단적 가중치에 대해 간단히 소개하고자 한다 (National Youth Policy Institute, 2013). 1차년도 횡단적 가중치는 기본가중치, 표본의 무응답조정을 위한 가중치 그리고 사후층화 가중치의 3가지 가중치의 곱으로 산출된다. 표본추출률에 대한 기본 가중치는 추출률의 역수로 다음과 같다.

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

 $f_{hij}$ : 표본 추출률,  $K_h$ : h층에 속한 학교의 수,  $l_h$ : h층에서 추출된 학교의 수,  $M_{hi}$ : h층 i번째 학교의 학생 수,  $n_{hi}$ : 추출된 표본학교에서 조사한 학생 수

표본의 무응답 조정을 위한 가중치와 사후층화가중치는 각각 다음과 같다.

$$W_{2hij} = rac{n_{hi}}{r_{hi}}$$
r\_hi : 응답한 학생 수

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

 $PS_{hij}$ : 외부보조정보에 의한 셀의 빈도수

(표본조사 완료 후 연구변수에 영향을 미치는 변수들 중에서 외부보조정보를 확보한다.)

2차년도와 3차년도의 횡단적 가중치는 다음과 같은 방식으로 산출된다.

n차년도 기본가중치 = 1차년도가중치 × 1/(1차년도 대비 응답률)

n차년도 최종가중치= 기본가중치 imes (지역별,성별 모집단 크기)/(지역별,성별 기본 가중치의 합계)

위와 같은 방식으로 산출한 가중치들을 적용한 경우와 가중치를 적용하지 않은 2가지 경우를 여러 가지 특성에 대한 분석결과를 비교하였다.

가장 기본적으로 연구에 사용되는 특성인 성별에 따른 비교를 가중치 적용여부에 따라 분석한 결과를 Table 3.2에 제시하였다. Table 3.2의 결과는 성별에 따라 범주형 자료와 수치형 자료에 대해 검정과 t-검정으로 비교하였다. 범주형 자료로는 선호연애인/운동선수 유무 (preferred entertainment/athletes), 해외방문 (overseas visit), 전학 (change of schools), 결석경험 유무 (absence), 휴대전화 보유 (cell phone), 컴퓨터 사용여부 (using the computer), 그리고 여러 가지 체험활동 경험 여부 (activities of experience)를 사용하였다. 수치형 자료로는 방과 후 교육경험 수 (education experience)는 11개 과목의 합을 이용한 자료와 자아존중감, 자아탄력성, 삶의 만족도, 정서문제 등의 5점척도로 측정된 자료의 평균을 계산하여 사용하였다. 자아존중감의 10개 문항에서는 긍정적 자아존중감 5개

와 부정적 자아존중감 5개 문항이 혼용되어 있으며, 그 중 부정적 자아존중감 문항 중 신뢰도가 낮은 문항이 있다. 본 연구에서는 Chang과 Jung (2013)와 같이 신뢰도에 부정적 영향을 끼치는 문항을 제외한 9개 문항으로 부정적 자아존중감은 역코딩을 하여 사용하였다.

각 자료의 최초 측정년도의 자료에 대한 신뢰도 분석결과 자아존중감 (self esteem, 9개 문항)의 Cronbach  $\alpha$ 가 0.873, 자아탄력성 (self elasticity, 14개 문항)은 0.839, 삶의 만족도 (life satisfaction, 3개 문항)는 0.814이며 정서문제 중 주의집중 (attention, 7개 문항)은 0.790, 공격성 (aggression, 6개 문항)은 0.810, 신체증상 (physical symptoms, 8개 문항)은 0.862, 사회적 위축 (social withdrawal, 5개 문항)은 0.850 그리고 우울 (depression, 10개 문항)은 0.906으로 나타나 각 특성에 포함된 문항들의 신뢰도는 모두 높았다.

가중치의 경우, 2010년 1차 자료는 1차년도 횡단적 가중치를 2차 자료와 3차 자료는 각 년도의 횡단적 가중치를 적용하였다. 범주형 자료와 교육경험수의 경우는 1차, 2차, 3차년도 모두 조사되었으나 자아존중감, 자아탄력성, 삶의 만족도, 5가지 정서문제에 관련된 문항은 일부년도에서만 조사되었다. 조사된 년도나 분석의 종류에 상관없이 전체적으로 가중치를 적용한 경우는 가중치를 적용하지 않은 것에 비해 통계적으로 유의한 것으로 나타났다. 가중치를 적용하지 않은 경우 통계적으로 유의한 결과는 가중치를 적용 시모두 유의한 것으로 나타났으나, 가중치를 적용하지 않은 경우 통계적으로 유의하지 않은 것은 일부만 동일한 결과를 얻었으며 가중치 적용 시대부분 통계적으로 유의한 결과가 나타났다.

Table 3.2 Result of differences in the use of the weights

St					
1st preferred entertainment/athletes					
year         overseas visit change of schools (Y/N)         (Y/N)         .63 (.83)         .12 (.800 (.000) (.					$\chi^2/\mathrm{t}\ (p\text{-value})$
Change of schools absence (Y/N)	1st	preferred entertainment/athletes	(Y/N)	46.034 (.000)	15492.656 (.000)
absence cell phone cell phone (Y/N)   3.266 (.627)   259.593 (.000 cell phone (Y/N)   32.688 (.000)   7488.066 (.000 using the computer (Y/N)   10.623 (.001)   2812.481 (.000 activities   Health/public health   6.533 (.011)   1677.129 (.000 fexperience   Science/information   7.408 (.006)   3566.231 (.000 (Y/N)   280.000   280.000 (Y/N)   280.000   280	year	overseas visit	(Y/N)	.745 (.388)	12.800 (.000)
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		change of schools	(Y/N)	.063 (.803)	.810 (.368)
using the computer		absence	(Y/N)	.236 (.627)	259.593 (.000)
Company		cell phone	(Y/N)	32.688 ( (.000)	7488.066 (.000)
Of experience (Y/N)		using the computer	(Y/N)	10.623 (.001)	2812.481 (.000)
(Y/N)		activities		6.533 (.011)	1677.129 (.000)
Adventure/pioneer   .010 (.919)   .49.741 (.000)		of experience		7.408 (.006)	3566.231 (.000)
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		(Y/N)			11.658 (.001)
Service   21.023 (.000)   5383.130 (.000   experience of job   34.584 (.000)   6480.195 (.000   environmental conservation   .005 (.945)   16.039 (.000)   6480.195 (.000   environmental conservation   .005 (.945)   16.039 (.000)   6480.195 (.000   17.810 (.000)   5493.311 (.000   5493.311 (.0					49.741 (.000)
Experience of job environmental conservation development of humanity   0.005 (.945)   16.039 (.000)   17.810 (.000)   5493.311 (.000)   5493.311 (.000)   17.810 (.000)   5493.311 (.000)   17.810 (.000)   5493.311 (.000)   17.810 (.000)   5493.311 (.000)   17.810 (.000)   5493.311 (.000)   17.810 (.0				25.698 (.000)	6954.314 (.000)
environmental conservation development of humanity   17.810 (.000)   5493.311 (.000)   6493.311 (.00				21.023 (.000)	5383.130 (.000)
$\begin{array}{ c c c c c }\hline & development of humanity \\ education \\ experience \\ (number) & group tutoing \\ visit the classes \\ internet \\ & after school education \\ & & & & & & & & & & & & & & & & & & $					6480.195 (.000)
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$					16.039 (.000)
experience (number)   Individual tutoring (number)   group tutoing visit the classes   -1.52 (.879)   4.405 (.000)   internet   .028 (.978)   .643 (.520)   after school education   .198 (.843)   -29.252 (.000)   self esteem   -3.271 (.001)   -62.252 (.000)   self elasticity   .086 (.931)   -1.112 (.266)   life satisfaction   -6.584 (.000)   -114.566 (.000)   year   aggression   2.408 (.016)   51.286 (.000)   social withdrawal   .114 (.909)   20.527 (.000)   depression   4.481 (.000)   76.481 (.000)   year   life satisfaction   -1.691 (.091)   -40.482 (.000)   year   attention   -1.691 (.091)   -40.482 (.000)   year   attention   -1.691 (.091)   -40.482 (.000)   year   attention   -4.571 (.000)   -94.445 (.000)   year   attention   -4.327 (.000)   -61.479 (.000)   aggression   aggression   1.334 (.182)   24.098 (.000)   physical symptoms   4.340 (.000)   74.800 (.000)   social withdrawal   -0.167 (.868)   15.280 (.000)   social withdrawal   -0.167 (.868)   15.280 (.000)   .5280 (.				17.810 (.000)	5493.311 (.000)
Commons   Group tutoing   2.446 (.015)   40.465 (.000)   visit the classes   -1.52 (.879)   4.405 (.000)   internet   .028 (.978)   .643 (.520)   after school education   .198 (.843)   -29.252 (.000)   self esteem   .3.271 (.001)   -62.252 (.000)   self elasticity   .086 (.931)   -1.112 (.266)   life satisfaction   -6.584 (.000)   -114.566 (.000)   year   aggression   .1.709 (.088)   -8.637 (.000)   year   aggression   2.408 (.016)   51.286 (.000)   year   social withdrawal   .114 (.909)   20.527 (.000)   depression   6.567 (.000)   117.418 (.000)   year   life satisfaction   -1.691 (.091)   -40.482 (.000)   year   attention   -4.571 (.000)   -94.445 (.000)   year   life satisfaction   -4.571 (.000)   -94.445 (.000)   year   life satisfaction   -4.327 (.000)   -61.479 (.000)   aggression   1.334 (.182)   24.098 (.000)   physical symptoms   4.340 (.000)   74.800 (.000)   social withdrawal   -0.167 (.868)   15.280 (.000)   social withdrawal   -0.167 (.868)   15.280 (.000)   .5000   .				2.812 (.005)	41.192 (.000)
Visit the classes internet   .028 (.978)   .643 (.520)   .643 (.520)   .643 (.520)   .643 (.520)   .643 (.520)   .643 (.520)   .643 (.520)   .643 (.520)   .6225 (.000)				.629 (.530)	2.087 (.037)
internet after school education   .028 (.978)   .643 (.520)     self esteem   .198 (.843)   .29.252 (.000)     self elasticity   .086 (.931)   .1.112 (.266)     life satisfaction   .6.584 (.000)   .114.566 (.000)     year   aggression   .1.709 (.088)   .8.637 (.000)     year   aggression   .2.408 (.016)   .51.286 (.000)     physical symptoms   .114 (.909)   .20.527 (.000)     depression   .114 (.909)   .20.527 (.000)     depression   .6.567 (.000)   .117.418 (.000)     year   life satisfaction   .1.691 (.091)   .40.482 (.000)     year   aggression   .1.691 (.091)   .40.482 (.000)     year   life satisfaction   .4.571 (.000)   .94.445 (.000)     aggression   .1.334 (.182)   .24.098 (.000)     physical symptoms   .1.334 (.182)   .24.098 (.000)     physical symptoms   .4.340 (.000)   .74.800 (.000)     social withdrawaL   .0.167 (.868)   .15.280 (.000)		(number)		2.446 (.015)	40.465 (.000)
After school education   1.98 (.843)   -29.252 (.000     self esteem   -3.271 (.001)   -62.252 (.000     self elasticity   .086 (.931)   -1.112 (.266)     life satisfaction   -6.584 (.000)   -114.566 (.000     2nd   attention   -1.709 (.088)   -8.637 (.000)     year   aggression   2.408 (.016)   51.286 (.000     physical symptoms   4.481 (.000)   76.481 (.000     social withdrawal   .114 (.909)   20.527 (.000     depression   6.567 (.000)   117.418 (.000     year   life satisfaction   -1.691 (.091)   -40.482 (.000     year   life satisfaction   -4.571 (.000)   -94.445 (.000     attention   -4.327 (.000)   -61.479 (.000     aggression   physical symptoms   4.340 (.000)   74.800 (.000)     physical symptoms   4.340 (.000)   74.800 (.000)     social withdrawal   -0.167 (.868)   15.280 (.000)				152 (.879)	4.405 (.000)
self esteem         -3.271 (.001)         -62.252 (.000)           self elasticity         .086 (.931)         -1.112 (.266)           life satisfaction         -6.584 (.000)         -114.566 (.000)           2nd         attention         -1.709 (.088)         -8.637 (.000)           year         aggression         2.408 (.016)         51.286 (.000)           physical symptoms         4.481 (.000)         76.481 (.000)           social withdrawal         .114 (.909)         20.527 (.000)           depression         6.567 (.000)         117.418 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           year         life satisfaction         -4.327 (.000)         -61.479 (.000)           attention         -4.327 (.000)         -61.479 (.000)         -61.479 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)					
self elasticity     .086 (.931)     -1.112 (.266)       life satisfaction     -6.584 (.000)     -114.566 (.000)       2nd     attention     -1.709 (.088)     -8.637 (.000)       year     aggression     2.408 (.016)     51.286 (.000)       physical symptoms     4.481 (.000)     76.481 (.000)       social withdrawal     .114 (.999)     20.527 (.000)       depression     6.567 (.000)     117.418 (.000)       3rd     self esteem     -1.691 (.091)     -40.482 (.000)       year     life satisfaction     -4.571 (.000)     -94.445 (.000)       attention     -4.327 (.000)     -61.479 (.000)       aggression     1.334 (.182)     24.098 (.000)       physical symptoms     4.340 (.000)     74.800 (.000)       social withdrawaL     -0.167 (.868)     15.280 (.000)			after school education		-29.252 (.000)
life satisfaction         -6.584 (.000)         -114.566 (.000)           2nd         attention         -1.709 (.088)         -8.637 (.000)           year         aggression         2.408 (.016)         51.286 (.000)           physical symptoms         4.481 (.000)         76.481 (.000)           social withdrawal         .114 (.909)         20.527 (.000)           depression         6.567 (.000)         117.418 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           attention         -4.327 (.000)         -61.479 (.000)           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)					-62.252 (.000)
2nd         attention         -1.709 (.088)         -8.637 (.000)           year         aggression         2.408 (.016)         51.286 (.000)           physical symptoms         4.481 (.000)         76.481 (.000)           social withdrawal         .114 (.909)         20.527 (.000)           depression         6.567 (.000)         117.418 (.000           3rd         self esteem         -1.691 (.091)         -40.482 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           attention         -4.327 (.000)         -61.479 (.000)           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)		v		.086 (.931)	-1.112 (.266)
year     aggression     2.408 (.016)     51.286 (.000)       physical symptoms     4.481 (.000)     76.481 (.000)       social withdrawal     .114 (.909)     20.527 (.000)       depression     6.567 (.000)     117.418 (.000       3rd     self esteem     -1.691 (.091)     -40.482 (.000)       year     life satisfaction     -4.571 (.000)     -94.445 (.000)       attention     -4.327 (.000)     -61.479 (.000)       aggression     1.334 (.182)     24.998 (.000)       physical symptoms     4.340 (.000)     74.800 (.000)       social withdrawaL     -0.167 (.868)     15.280 (.000)				-6.584 (.000)	-114.566 (.000)
physical symptoms       4.481 (.000)       76.481 (.000)         social withdrawal       .114 (.909)       20.527 (.000)         depression       6.567 (.000)       117.418 (.000         3rd       self esteem       -1.691 (.091)       -40.482 (.000         year       life satisfaction       -4.571 (.000)       -94.445 (.000         attention       -4.327 (.000)       -61.479 (.000         aggression       1.334 (.182)       24.998 (.000)         physical symptoms       4.340 (.000)       74.800 (.000)         social withdrawaL       -0.167 (.868)       15.280 (.000)	2nd			-1.709 (.088)	-8.637 (.000)
social withdrawal depression         .114 (.909)         20.527 (.000)           3rd         self esteem         -1.691 (.091)         -40.482 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           attention         -4.327 (.000)         -61.479 (.000)           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)	year			2.408 (.016)	51.286 (.000)
depression         6.567 (.000)         117.418 (.000)           3rd         self esteem         -1.691 (.091)         -40.482 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           attention         -4.327 (.000)         -61.479 (.000)           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)				4.481 (.000)	76.481 (.000)
3rd         self esteem         -1.691 (.091)         -40.482 (.000)           year         life satisfaction         -4.571 (.000)         -94.445 (.000)           attention         -4.327 (.000)         -61.479 (.000)           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)				.114 (.909)	20.527 (.000)
year         life satisfaction         -4.571 (.000)         -94.445 (.000           attention         -4.327 (.000)         -61.479 (.000           aggression         1.334 (.182)         24.098 (.000)           physical symptoms         4.340 (.000)         74.800 (.000)           social withdrawaL         -0.167 (.868)         15.280 (.000)				6.567 (.000)	117.418 (.000)
attention       -4.327 (.000)       -61.479 (.000)         aggression       1.334 (.182)       24.098 (.000)         physical symptoms       4.340 (.000)       74.800 (.000)         social withdrawaL       -0.167 (.868)       15.280 (.000)	3rd			-1.691 (.091)	-40.482 (.000)
aggression       1.334 (.182)       24.098 (.000)         physical symptoms       4.340 (.000)       74.800 (.000)         social withdrawaL       -0.167 (.868)       15.280 (.000)	year			-4.571 (.000)	-94.445 (.000)
physical symptoms 4.340 (.000) 74.800 (.000) social withdrawaL -0.167 (.868) 15.280 (.000)				-4.327 (.000)	-61.479 (.000)
social withdrawaL -0.167 (.868) 15.280 (.000)		aggression		1.334 (.182)	24.098 (.000)
				4.340 (.000)	74.800 (.000)
depression $7.250 (.000)$ $136.025 (.000)$				-0.167 (.868)	15.280 (.000)
		depression		7.250 (.000)	136.025 (.000)

성별 뿐만 아니라 다른 항목에 대해서 분석한 결과도 크게 다르지 않았다. 가중치 사용여부에 따라 결 과의 차이가 심하게 나타났다. 이 결과는 가중치 계산에 표본추출방법이 고려되었으나 분석에서는 단순 히 가중치만 반영하였기 때문에 가중치만큼 자료의 수가 많은 것으로 계산되어 표준오차가 매우 작아지 므로 가중치를 적용한 경우가 그렇지 않은 경우보다 상대적으로 많은 항목에서 유의한 것으로 나타나는 경향이 있다. Heo와 Chung (2012)은 피어슨의 카이제곱 검정은 측정자료가 서로 독립이라는 가정하 에서 분석이 되므로 복합표본설계로 수집된 자료로 일반적인 피어슨의 카이제곱 검정을 실시한다면 잘 못된 결론을 얻을 수도 있다고 지적하였으며 Chung (2012)은 국민건강영양조사자료를 근거로 하여 층 화집락계통추출의 표본을 분석할 경우에는 반드시 층화, 집락, 가중치를 고려한 복합표본설계를 이용하 여야 한다고 하였으며, SUDAAN, Stata, Westat, SAS의 복합표본설계 프로시져, SPSS의 Complex Samples 등의 프로시져을 이용한 방법을 언급하였다. Park과 Choi (2014)는 국민건강영양조사자료를 복합표본을 이용하여 복합표본 교차분석, 복합표본회귀분석을 적용하였다. 최근 조사자료들이 대부분 복합표본으로 추출되므로 복합표본설계를 적용한 분석이 쉽게 이루어질 수 있도록 표본설계에 따른 계 층이나 군집에 대한 변수도 함께 제공하는 필수적이다. 종단적 가중치의 경우 2차년도의 가중치와 3차 년도의 가중치가 각각 제공되고 있으나, 1차 자료와 2차 자료를 이용한 연구, 2차와 3차 자료를 이용한 연구, 1차, 2차, 3차 자료를 모두 이용한 연구에서 종단적 가중치를 어떻게 적용이 되어야 하는지에 대 해 판단하기가 어렵다. 따라서, KCYPS자료를 이용한 많은 연구가 좀 더 정확한 결론을 얻기 위해서는 단순한 가중치가 아닌 다양한 가중치가 필요하며, 연구자들이 쉽게 적용할 수 있는 가중치 적용방법에 대한 소개가 요구된다.

# 4. 결론 및 논의

최근 체계적으로 잘 설계된 대규모 조사자료를 이용한 많은 연구들이 진행되고 있다. 과거에 비해 점 점 분석방법도 다양해지고 정교하게 이루어지고 있다. 조사대상이나 항목의 무응답에 대해서 뿐만 아니 라 객관적인 표본설계와 조사로 인해 모집단의 특성을 제대로 반영할 수 있는 표본추출에 대한 가중치 적용도 많이 이루어지고 있는 실정이다. 여러 분야에서 무응답에 대한 처리를 여러 가지 방법으로 논의 되고 있으며, 무응답을 포함한 가중치도 많은 분야에서 적용되고 있다. 그 중 KCYPS자료에서 무응답 에 대한 처리실태를 파악하고, 그 중 단위무응답의 특성 및 처리에 대해 분석하였다. 단위무응답은 연구 자들의 연구대상에 해당되는 부분으로 언급은 1, 2, 3차 자료의 분석논문 중 21.5%가 언급하였으며, 단 위무응답이 발생되지 않은 1차 자료를 제외한 경우에서는 38.2%가 단위무응답에 대해 언급하였다. 이 에 비해 항목무응답에 대한 처리는 전체의 7.7%만이 언급되었다. 일반적인 조사자료에서는 단위무응답 의 특성 파악이 용이하지 않으나 종단자료는 추적조사로 단위무응답의 특성연구가 가능하다. 이에 단위 무응답의 처리에 앞서 KCYPS자료의 단위무응답자들의 특성을 파악하였다. 2011년에 조사된 2차 자 료에서 단위무응답여부가 인구특성과의 연관성이 발생된 항목이 청소년 응답에서는 성별, 자택 위치, 학 교 위치, 연간 전학경험 유무, 연간 결석 경험유무, 방과 후 교육 연간 경험유무이며, 3차 자료에서는 학 교제, 성별, 자택 위치, 학교 위치, 연간 결석 경험유무, 해외방문 경험유무이었다. 보호자 응답은 2차 3차 자료 모두 성별, 자택 위치, 학교 위치에서만 연관성이 나타났다. 단위무응답의 처리에서 표본추출 을 고려한 가중치를 제공하고 있어, 단순히 단위무응답자의 처리여부에 따른 비교가 아니라 전체 모집단 의 특성과 조사대상만으로 분석한 경우의 결과 차이를 비교한 결과 거의 많은 부분에서 통계적인 차이 를 보였다. 조사된 년도나 분석의 종류에 상관없이 전체적으로 가중치를 적용한 경우는 가중치를 적용 하지 않은 것에 비해 통계적으로 유의한 것으로 나타났다. 가중치를 적용하지 않은 경우 통계적으로 유 의한 결과는 가중치를 적용 시 모두 유의한 것으로 나타났으나 가중치를 적용하지 않은 경우 통계적으로 유의하지 않은 것의 상당부분이 가중치 적용 시 통계적으로 유의한 결과가 나타났다. 이는 각각의 분석

에서 가중치를 적용함으로써 가중치 계산에서 표본추출과 무응답은 반영이 되어졌으나 가중치 적용으로 인한 발생된 표준오차의 문제점은 반영되지 않았다. 또한, KCYPS자료는 각 년도별로 횡단적 가중치과 종단적 가중치 2가지 가중치만 제공하고 있으며 그 가중치 사용에 대해서는 연구자에게 일임하고 있다. 그러나, 가중치 적용을 하기 위해서는 표본에 대한 이해와 통계적 지식이 요구되므로 연구자들이 충분히 적용할 수 있는 가중치 적용에 대한 상세한 지침 이 필요하며 복합표본설계를 적용한 분석이 가능하도록 계층과 집락에 해당되는 변수를 구분하여 제공하는 것이 필요하다. 또한, 단순한 가중치가 아니라 년도 별 선택 수 및 사용패널의 수와 종단 및 횡단연구를 병행한 경우 등 여러가지 자료 사용에 따른 적절한 가중치가 제시되어야 할 것으로 생각된다.

#### References

- Chang, H. L. and Chung, I. J. (2013). The effects of parenting attitudes on adolescent prosocial behavior and aggression mediated by self-esteem. *Korean Journal of Family Social Work*, **40**, 33-61.
- Chung, C. E. (2012). Complex sample design effects and inference for Korea National Health and Nutrition Examination Survey data. Journal of Nutrition and Health, 45, 600-612.
- Heo, S. and Chung, Y. (2012). Effect of complex sample design on Pearson test statistic for homogeneity.

  Journal of the Korean Data & Information Science Society, 23, 757-764.
- Hirschi, T. and Gottfredson, M. R. (1993). Commentary: Testing the general theory of crime. *Journal of Research in Crime and Delinquency*, **30**, 47-54.
- Junger-Tag, J. and Marshall, I. H. (1999). The self-report methodology in crime research. In *Crime and Justice; A Review of Research*, edited by M. Torny, University of Chicago Press, Chicago.
- Kim, S. W. and Chung, L. J. (2012). Factors influencing item nonresponse and unit nonresponse in a youth panel survey. *Studies on Korean Youth*, **23**, 27-50.
- Lee, H. G. (2008). Status analysis of Korea panel research (How to increase the accuracy and efficiency of the survey statistics: Based on the field survey and sampling), Statistical research institute, Daejeon.
- Lee, J. Y. (2005). Longitudinal study of household survey data (Statistical studies 05-03), Statistics Korea, Daejeon.
- National Youth Policy Institute (2013). 1st 3rd survey data user's guide in Korea children and youth panel survey (KCYPS), National Youth Policy Institute, Seoul.
- Park, C. and Choi H. S. (2014). A study of the factors influential on a health-related quality of life using complex sample design. Journal of the Korean Data & Information Science Society, 25, 829-846.
- Shin, D. K. (1998). Status and implications of United States panel data Focus on PSID, NLSY, KHPS, KLIPS, Korean Labor Institude, Seoul.
- Watkins, A. K. and Melde, C. (2007). The effectof self-control on unit and item nonresponse in an adolescent sample. *Journal of Research in Crime and Delinquency*, 44, 267-294.

# The unit-nonresponse status and use of weight in the KCYPS

Hwa-Jung Lee $^1$  · Suk-Bok Kang $^2$ 

<sup>1</sup>College of Ecology and Environmental Science, Kyungpook National University

<sup>2</sup>Department of Statistics, Yeungnam University

#### Abstract

Usually unit-nonresponse or item-nonresponse occurs in the survey. In case the rate of nonresponse is high, the analysis ignoring nonresponse may cause the wrong effect. The characterization of nonresponse is required. In a cross-sectional data, it is possible to study the characteristics of item-nonresponse but it is hard to study the characteristics of the unit-nonresponse. In order to identify the characteristics of the unit-nonresponse, this study used the first-year student of middle schools in the Korea children and youth panel survey (KCYPS) data. We investigated the handling situation of nonresponse and analyzed the characteristics of the unit-nonresponse. Most of the papers applied the way of getting rid of nonresponse, so that there was little paper using weights. In this paper, we compared the results of the analyses depending on whether the weight is used or not. The method of using weights showed statistically significant results much more than that of removing. More discussion will be needed.

Keywords: Item-nonresponse, the Korea children and youth panel survey, unit-nonresponse, weights.

<sup>&</sup>lt;sup>1</sup> Instructor, College of Ecology and Environmental Science, Kyungpook National University, Daegu 702-701, Korea.

<sup>&</sup>lt;sup>2</sup> Corresponding author: Professor, Department of Statistics, Yeungnam University, Kyeongbuk 712-749, Korea. E-mail: sbkang@yu.ac.kr