

한국종합사회조사 (Korean General Social Survey, KGSS) 2003-2021

연구책임자	김지범 (성균관대학교)
공동연구자	강정한 (연세대학교) 김석호 (서울대학교) 김창환 (캔사스대학교) 박원호 (서울대학교) 이윤석 (서울시립대학교) 최슬기 (KDI 국제정책대학원) 김솔이 (한국행정연구원)
연구조교	양동평 유광종 김규영 김혜린 정상운

성균관대학교 동아시아학술원 서베이리서치센터

누적자료 혹은 누적코드북 인용 표기 방법

김지범, 강정한, 김석호, 김창환, 박원호, 이윤석, 최슬기, 김솔이. (2022). 한국종합사회조사 2003-2021. 서울: 성균관대학교 출판부.

Kim, Jibum, Jeong-han Kang, Seok-ho Kim, Changhwan Kim, Wonho Park, Yun-Suk Lee, Seulki Choi, and Sori Kim. (2022). Korean General Social Survey 2003-2021. Seoul, Korea: Sungkyunkwan University.

성균관대학교 동아시아학술원 서베이리서치센터

성균관대학교 동아시아학술원 서베이리서치센터는 학술조사기관으로 2002년에 설립되었다.
서베이리서치센터에서는 한국종합사회조사를 2003년부터 2021년까지 수행했다.

서베이리서치센터

(03063) 서울특별시 종로구 성균관로 25-2 성균관대학교 호암관 605-7

(전화) 02-760-1270, 0709

EMAIL: jbk7000@skku.edu

kgss.src@gmail.com

이 저서는 2020년 대한민국 교육부와 한국연구재단의
지원(NRF-2020S1A3A2A03096777)과 성균관대학교의 지원을 받아 수행된 연구이다.

이 책은 한국사회변화 파악에 필수적인 한국종합사회조사의 2003-2021 누적자료집(cumulative codebook)이다. 2021년도 자료를 2003-2018 누적자료와 조화화(harmonized)하는 과정을 거쳐, 한국종합사회조사의 2003-2021 누적자료(cumulative data)를 만들었다. 누적자료집은 누적자료를 이용하는 연구자 및 학생들이 쉽게 자료를 이해할 수 있도록, 조사된 전 문항의 응답분포, 설문지, 조사원 지침서, 이용 문헌 등을 포함하였다. 연구자들은 KGSS 홈페이지(<http://kgss.skku.edu>)에서 누적원자료 및 관련 자료를 내려 받을 수 있다.

이 책은 많은 분들의 도움으로 출간 할 수 있었다. 첫째, 2003년부터 2018년까지 14번의 KGSS 조사에 협력해 주신 많은 응답자들(총 19,636명), 조사원으로 참여한 학생들(2,797명), 교수님들(116명)께 감사드린다(참조, 부록(5) KGSS 참여 교수·학생 명단). 특히 경상대 최태룡(13번), 청주대 이정환(13번), 부산대 김희제(11번), 충남대 박재목(10번), 전북대 김영정(10번), 남서울대 이주열(9번), 한림대 엄한진(9번) 교수님들로부터 과분한 도움을 받았다. 2021년 조사는 코로나19 상황과 조사비용 문제로 학생들과 함께 자료를 만들지 못했다. 둘째는, 직접 조사에 참여하지 않으셨지만, KGSS를 조사방법 혹은 통계 수업 시간에 이용하여 주신 교수님들과 KGSS를 분석하여 한국 사회과학에 기여해 주신 많은 연구자분들께 감사한다(참조, 부록(3) KGSS 이용 문헌 목록). 셋째는, 어려운 환경에서도 성균관대학교 서베이리서치센터를 이끌어 오신 석현호, 양중희, 김상욱 전임 센터장님들의 헌신에 감사드린다. 특히, KGSS 설립자이신 석현호 센터장님께서, 한국에서 많은 사회과학자들에게 자료의 질에 대한 중요성을 깨닫게 해주셨고, 조사방법론 및 조사 품질 향상 연구의 토대를 만들어 주셨다. 넷째는, 2021년 표본추출과 가중치를 만들어 주신 신희춘 박사님과 산업 직업 분류 작업을 해준 반미희씨께 감사드린다. 마지막으로, 서베이리서치센터의 조교로 활동한 양동평, 유광중, 김규영, 김혜린, 정상운 학생들의 꼼꼼한 일 처리 덕분에 코드북을 완성할 수 있었다.

분명히, 이 책은 한국연구재단, 성균관대학교 재단, 그리고 성균관대학교 동아시아학술원의 KGSS 자료 수집에 대한 재정적 지원의 결과물이다. 한국종합사회조사는 2003년부터 2014년까지 한국연구재단 지원 아래 가능하였다. 2016년부터 2018년은 한국연구재단의 지원이 중단되었지만, 성균관대학교, 서울대학교 사회발전연구소, 보건사회연구원, 한국행정연구원 등의 많은 기관과 연구자들의 물심양면 지원으로 KGSS를 지속할 수 있었다. 다행스럽게도 KGSS는 2020년부터 한국연구재단의 한국사회과학연구(SSK)의 장기구축데이터 연구를 통하여 지원받게 되었다. 2021년 KGSS는 한국연구재단의 지원과 성균관대학교 재단의 지원을 통해 수행되었다.

석현호 명예교수님께서 첫 번째 통계자료집인 2003 한국종합사회조사 머리말에서 한국종합사회조사가 “사회과학 자료 부실을 극복하는 데 획기적으로 기여하게 될 것”이라는 예측은 20여년이 지난 지금 틀리지 않았다고 생각한다. 앞으로도, 한국종합사회조사를 만들어 오신 많은 분들의 희생과 노력이 헛되지 않도록, 조사의 기본 원칙을 충실히 지키며 전반적 조사과정을 공개하면서 한국 사회과학에 기여 할 수 있는 좋은 자료를 생산할 수 있도록 최선을 다하겠다.

2022년 5월

2021년 KGSS 연구진과 자문위원을 대표하여
서베이리서치센터 센터장 김 지 범

한국종합사회조사 자문위원 및 운영위원 명단

국내 자문위원

서울: 이재열 자문위원장(서울대학교 사회학과), 계봉오, 신동준(국민대학교 사회학과), 김두환(덕성여자대학교 사회학과), 김선웅(동국대학교 통계학과), 김영원(숙명여자대학교 통계학과), 김주연(서울시립대학교 도시사회학과), 김중백, 송재룡(경희대학교 사회학과), 유은혜(숭실대학교 사회학과), 이건(한양대학교 행정학과), 이민아(중앙대학교 사회학과), 이재묵(한국외국어대학교 정치외교학과), 임동균(서울대학교 사회학과), 김옥대(한국방송통신대학교), 박찬웅, 한준(연세대학교 사회학과), 이명진, 심재만(고려대학교 사회학과), 이용모(건국대학교 행정학과), 최셋별, 김현수(이화여자대학교 사회학과), 하상웅(서강대학교 정치학과)

경기도: 송인덕(중부대학교 신문방송학과), 신준옥(성결대학교 사회복지학과), 신희주(가톨릭대학교 사회학과), 여정희(인하대학교 소비자학과), 윤상길(신한대학교 언론학과), 한지은(평택대학교 사회복지학과), 김정인(수원대학교 행정학과), 변종석(한신대학교 응용통계학과), 유우현(인천대학교 신문방송학과), 차승은(수원대학교 아동청소년학과), 호정화(아주대학교 사회학과)

강원도: 엄한진(한림대학교 사회학과), 이태정(연세대학교 경제학과), 김여진(강원대학교 사회학과)

충청북도: 염지혜(중원대학교 사회복지학과), 이해진(충북대학교 사회학과), 이정환(청주대학교 사회학과)

충청남도: 이주열(남서울대학교 보건행정학과), 최준혁(순천향대학교 미디어컨텐츠학과), 황명진(고려대학교 사회학과), 황선재(충남대학교 사회학과), 이원재(카이스트 문화기술대학원), 이혜경(배재대학교 공공정책학과), 박정선(경찰대학교 행정학과)

경상북도: 류성진(대구대학교 미디어커뮤니케이션학과), 박중규(경북대학교 심리학과), 신형진, 조우은(경북대학교 사회학과), 윤우석(계명대학교 경찰행정학과), 정용교(영남대학교 사회학과)

경상남도: 김희재(부산대학교 사회학과), 박철현(동의대학교 경찰행정학과), 송유진, 윤상우(동아대학교 사회학과), 예동근(부경대학교 국제학부), 이건혁(창원대학교 신문방송학과), 이성균(울산대학교 사회복지학과), 김주호(경상대학교 사회학과), 김동완(경남대학교 사회학과)

전라북도: 설동훈, 김재우(전북대학교 사회학과), 백종일(원광대학교 빅데이터·금융통계학부)

전라남도: 김영태(목포대학교 정치언론홍보학과), 신지원(전남대학교 사회학과), 조선일(순천대학교 행정학과), 지병근(조선대학교 정치외교학과)

제주도: 최현(제주대학교 사회학과)

해외자문위원

Tom W. Smith (NORC, Univ. of Chicago), Yang-chih Fu (Academia Sinica), Noriko Iwai (Osaka Univ. of Commerce), Yanjie Bian (University of Minnesota), Weidong Wang (Renmin Univ. of China), Chinfen Chang (Academia Sinica), Hee-Choon Shin (National Center for Health Statistics), Hyeyoung Woo (Portland State University), Jinkook Lee (USC & RAND), Joonmo Son (National University of Singapore), Chaeyoon Lim (University of Wisconsin-Madison), Jaesok Son (NORC, Univ. of Chicago), Min-Woong Sohn (University of Kentucky)

서베이리서치센터 운영위원

강민석(성균관대학교 심리학과), 김기현(성균관대학교 사회복지학과), 유홍준(성균관대학교 사회학과), 정성은(성균관대학교 신문방송학과), 조민호(성균관대학교 행정학과), 조원빈(성균관대학교 정치학과)

머리말 · 5

제1부 KGSS의 소개

개요 · 10

자료/조사 내용 및 방법 · 11

- 1) 조사 내용 · 13
- 2) 조사 방법 · 20
- 3) 자료 이용 · 28

제2부 통계표

1. 2003-2021 KGSS Cumulative Codebook · 35

2. ISSP Module · 399

- 1) ISSP 2003/2013 National Identity · 399
- 2) ISSP 2004/2014 Citizenship · 418
- 3) ISSP 2005 Work Orientations · 436
- 4) ISSP 2006/2016 Role of Government · 453
- 5) ISSP 2007 Leisure Time and Sports · 468
- 6) ISSP 2008/2018 Religion · 484
- 7) ISSP 2009 Social Inequality · 501
- 8) ISSP 2010/2020 Environment · 521
- 9) ISSP 2011 Health and Health Care · 537
- 10) ISSP 2012 Family and Changing Gender Roles · 548

3. EASS Module · 564

- 1) EASS 2006/2016 Family · 564
- 2) EASS 2008/2018 Globalization & Culture · 607
- 3) EASS 2010/2020 Health · 625
- 4) EASS 2012 Network Social Capital · 645

4. 2003-2021 KGSS 가구원 특성 · 661

부록1. 18-21 한국종합사회조사 변수명 설문지 · 687

부록2. 03-21 KGSS 조사원 지침서 · 754

부록3. KGSS 이용 문헌 목록 · 756

부록4. KGSS ISSP/EASS/제안설문모듈 현황 · 758

부록5. KGSS 참여 교수·학생 명단 · 760

부록6. 03-21 KGSS 설문구성 현황 · 768

부록7. KGSS 색인 · 770

개요

2003년에 태동된 한국종합사회조사(Korean General Social Survey, KGSS)(설립자: 석현호 성균관대학교 사회학과 명예교수)는 사업체가 후원한 상업적 조사, 정부 및 공공기관에서 발주하여 조사업계에서 수행하는 많은 조사와 달리 전반적 조사오류가 적은 자료를 만들고자 하는 발상의 전환으로 시작된 “공익적인 사회과학 자료 창출 사업”¹⁾이다. 2002년 한국학술진흥재단의 <사회기초예비조사연구>라는 기초학문지원사업이 그 출발점이었다. 성균관대학교 동아시아학술원 서베이리서치센터 한국종합사회조사 예비연구(KGSS Pilot Study)는 “학술적 목적”으로 “사회과학의 제반 핵심적 주제”들을 포함하고 “방법론적 원칙들을 엄정히 준수”하며, “주기적이고 반복적인 방식”으로 “확률표집의 이상적 기준을 최대한 충족”시키면서 “국제비교연구”를 가능케 하고 무상으로 제공되어야 한다는 원칙을 가지고 시작했다²⁾.

한국종합사회조사는 미국의 시카고대학교 NORC(NORC at the University of Chicago)에서 1972년 최초로 창안되어 지난 50년 동안 전 세계적으로 엄청난 파급효과를 발휘하고 있는 미국의 종합사회조사(General Social Survey, GSS)의 ‘한국판’ 조사이다.³⁾ 이 조사는 위에서 언급한 이상적 요건들을 빠짐없이 충족시킬 목적으로 설계된 공익적 성격의 대단위 학술사회조사 DB 창출 및 확산 사업이다. 한국종합사회조사는 한국사회의 기초를 이루는 여러 핵심적 내용들을 포괄하여 매년 조사하는 반복핵심설문(replicating core questions), 국가간비교사회조사(cross-national survey research)인 ISSP(International Social Survey Programme)⁴⁾와 EASS(East Asian Social Survey)⁵⁾ 그리고 연구자들이 제안하는 모듈 등을 유기적으로 결합하는 형태로 설문을 구성한다.

한국종합사회조사의 질적 우수성은 한국종합사회조사의 높은 대표성과 조사의 반복성에 기인한다: (1) 높은 대표성: 한국종합사회조사는 응답자 수보다 응답자의 대표성이 중요하다는 조사원칙에 충실한 자료이다. (2) 조사의 반복성: 한국종합사회조사 핵심 문항들의 반복조사가 갖는 누적성은 사회의 안정성과 변동성 모두를 연구하는데 매우 중요하고, 개별 연도별로 진행되는 설문보다 결합한 누적 자료(pooled cross-sectional data)로서의 가치가 더욱 크다.

미국의 연구재단(NSF)에서 50년동안 재정적으로 지원받은 미국종합사회조사(General Social Survey)가 미국 사회변동을 이해하기 위한 길잡이 노릇을 하고 있듯이, 한국종합사회조사의 활발하고 광범위한 활용을 생각할 때, 한국종합사회조사 또한 한국 사회변동을 파악할 수 있는 토대가 되리라 예상한다.

1) 석현호, 문용갑, 구혜란, 김상욱, 김제운, 은기수, 이명진, 이윤석, 임영규, 정기선, 최문경. (2005). 한국종합사회조사 2003. 성균관대학교출판부.

2) 차종천, 김상욱, 남춘호, 문용갑, 박경숙, 이정환, 이해경, 정기선. (2005). 사회조사의 업그레이드: 한국종합사회조사(KGSS) 출발. 성균관대학교출판부.

3) Smith, Tom W., Jibum Kim, Achim Koch, and Alison Park. (2006). Social science research and the general social surveys. *Comparative Sociology*, 5(1), 33-43.

4) ISSP는 전 세계 43개 국가가 참여하여 국제비교자료를 생산하고 있다. 한국에서는 성균관대학교 동아시아학술원 서베이리서치센터가 회원기관이다(<http://www.issp.org>).

5) EASS는 GSS 형태의 사회조사를 수행 중인 동아시아 4개국(한국, 일본, 중국, 대만)의 공동조사이다(<http://www.eassda.org>). 2016년, 2018년 EASS는 성균관대학교 서베이리서치센터와 서울대학교 사회발전연구소가 공동으로 진행했다.

자료

한국종합사회조사의 모집단은 전국 만 18세 이상 성인 남녀로, 다단계지역확률표집(multi-stage area probability sampling) 방법에 의거해 표본을 추출하며, 현지조사는 대면 면접(person to person interview) 방식으로 수행한다. 한국종합사회조사는 엄격한 확률표집 원칙을 지키면서도 응답률이 다른 조사와 비교하여 상대적으로 매우 높은 수준이다. 하지만, 전반적 조사환경의 어려움을 반영하듯⁶⁾, 한국종합사회조사의 응답률 또한 2016년 이후 40-50% 정도가 되었다. 각 년도 응답자 수와 응답률은 다음과 같다: 2003년 1,315(66%); 2004년 1,312(66%); 2005년 1,613(65%); 2006년 1,605(64%); 2007년 1,431(57%); 2008년 1,507(60%); 2009년 1,599(64%); 2010년 1,576(65%); 2011년 1,535(61%); 2012년 1,396(56%); 2013년 1,294(52%); 2014년 1,370(55%); 2016년 1,051(47%); 2018년 1,031(43%). 2021년 1,205(50%)

2003-2021 누적원자료는 각 년도 자료를 통합하여 하나의 파일로 만들어 이용자들이 KGSS를 쉽게 이용할 수 있도록 하였다. 각 년도 파일을 조화화(harmonized) 하는 과정에서, 기존 공개된 자료에서 다른 변수임에도 불구하고, 변수 이름이 같은 경우 누적자료집(누적원자료)에서는 변수명을 새로 부여하였다(예, 2013년 조사 당시의 변수명 INTERNET이 누적데이터 상에는 INTRNT13으로 변경). 이런 경우는 코드북에 포함된 설문지에서 하나의 설문 문항에 두 개의 변수 명을 제공하였다. 2003-2013년도와 2003-2014년도 누적원자료는 공개된 각 년도 자료를 취합하여 만들었으며, 누적원자료는 공개된 연도별 데이터 중 오류를 수정한 내용을 포함하고 있다. 2003-2016 누적자료부터는 연도별 자료를 제공하고 있지 않다. 누적자료집은 모든 년도 문항에 대한 통계표와 더불어, 부록으로 변수명 기입 설문지, 조사원 지침서, KGSS 데이터 이용 문헌, ISSP, EASS, 및 연구자 제안 설문 모듈 현황, 현재까지의 KGSS 참여 교수와 학생 명단, KGSS 설문구성 현황, KGSS 색인을 포함한다. 아래는 누적자료집 이용자들을 위하여, 누적자료집에서 보여주는 표를 설명한다.

표 671 청와대 지도층에 대한 신뢰

[설문 문항] 다음은 우리나라의 주요 사회기관들입니다. 귀하는 이 기관들을 이끌어가는 사람들을 어느 정도 신뢰하는지 말씀해 주십시오... 청와대

[변수명] CONBLUHO

응답 범주	변수값	03-06	2007	2008	2009	2010	2011	2012	2013	2014	2016	2018	2021	전체
매우 신뢰	1	5.7	7.4	6.8	9.4	11.9	8.6	6.5	14.0	10.9	9.7	16.1	9.3	8.7
다소 신뢰	2	47.5	47.6	43.8	43.7	46.8	50.0	42.7	52.7	40.8	41.7	54.6	48.6	46.7
거의 신뢰하지 않음	3	42.3	42.1	47.3	44.4	38.1	36.8	48.3	30.1	47.6	46.7	26.8	40.1	41.4
모르겠다/무응답	-8	4.5	2.9	2.1	2.6	3.2	4.6	2.5	3.2	0.7	1.9	2.5	2.0	3.1
비해당	-1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
N		5845	1431	1508	1599	1576	1535	1396	1294	1370	1051	1031	1205	20841

* 보기카드 제공(05-21)

** 2003년 설문 문항: 다음은 우리나라의 주요 사회기관들입니다. 이 기관을 이끌어 가는 사람들에 대하여 귀하는 어느 정도 신뢰하는지 말씀해 주십시오... 청와대

** 2004년, 2005년 설문 문항: 다음은 우리나라의 주요 사회기관들입니다. 귀하는 이 기관들을 이끌어가는 사람들을 어느 정도 신뢰하는지 말씀해 주십시오... 청와대

*** 핵심어: 사회기관, 신뢰, 청와대

표는 제목, 문항정보, 빈도표, 보기카드 제공 여부, 설문 혹은 응답범주의 변화 등으로 구성되어 있다. 표 제목에는 일련번호와 설문 문항에 대한 간략한 설명이 표기되어 있고, 그 아래는 전체 설문 문항과 변수명을 보여준다. 빈도표는 응답 범주와 그에 해당하는 변수값, 그리고 2003년부터 2021년 중 문항이 조사된 연도에 응답분포가 기재되어 있다. 지면상의 제약으로 2003, 2004, 2005, 2006년을 묶어서 2003-2006로 표기했다. 2003-2006에 들어간 퍼센트는 각 년도에 모두 물어본 문항이 아닌 경우 비해당이 많게 되어 퍼센트가 특정 연도의 퍼센트를 올바르게 반영할 수 없는 문제점이 있기 때문에 원자료를 이용해야 한다. 빈도의

6) 김지범, 장덕현, 김솔이, 강정환. (2019). 조사동향: 여론조사에 대한 조사. 조사연구, 20(1), 131-164.

수치는 퍼센트로 소수점 한자리까지 표시하였다. 둘째자리에서 반올림했기 때문에 실제로 0이 아닌 경우도 0으로 표기된 경우가 있다(예, 0.04%인 경우 0으로 표기). 빈도표 하단에는 각 연도에 응답한 전체 응답자 수(N)가 있어서, 분포와 응답자 수를 가지고 해당 응답범주에 몇 명이 응답하였는지 알 수 있다. 표의 퍼센트는 가중치를 이용하였다.

기타사항은 보기카드, 설문문항 및 응답범주의 변화, 그리고 핵심어(keyword)를 포함한다. 문항에 보기카드가 제공된 경우(2016부터 모든 문항에 보기카드 제공)에는 주석에 이를 표기하였고, 조사된 여러 연도에 보기카드가 모두 제공되지 않았을 경우에는 보기카드가 제공된 연도만 추가로 표기하였다(예, 05-18은 보기카드가 2005년부터 2018년까지 제공된 경우로, 2003년과 2004년에는 보기카드가 응답자에게 제공되지 않았다는 것을 의미한다). 같은 변수명을 이용하였어도, 설문 문항 및 응답 범주에 변화가 있는 경우, 그 차이를 비교 할 수 있도록 모두 기록하였다. 마지막으로 핵심어는 최대한 설문 문항에 사용된 단어를 그대로 차용해서 만들고자 하였지만, 그렇지 않은 경우도 있다. 또한, 핵심어를 누적자료집의 색인에서 찾으려면, 같은 키워드에 해당하는 다른 변수들을 파악할 수 있다.

부록은 변수명을 포함한 설문지(부록 1), 조사원들이 조사를 수행할 때 알아야 할 내용을 적은 지침서(부록 2), KGSS를 이용한 문헌목록(부록 3), 각 년도 모듈(부록 4), 참여 교수 및 학생 명단(부록 5), 어떤 문항들이 언제 몇 번 물어보았는지를 파악할 수 있는 설문구성 현황(부록 6), 그리고 변수와 관련된 색인(부록 7)을 포함하고 있다.

조사내용 및 방법

1. 조사내용

한국종합사회조사는 지속적으로 반복핵심 조사, ISSP 주제모듈 조사, EASS 주제모듈 조사, 연구자 제안 주제모듈 조사 등을 해마다 필요에 따라 유기적으로 통합하여 시행하였다. 이를 통하여 학문적으로 한국사회에 대한 다양한 분야 연구, 사회과학 조사 방법론 연구, 한국 사회 변동 연구, 국가간비교연구를 가능하게 했으며, 사회 전 분야에서 활용되며,⁷⁾ 교육 자료로 많이 활용되고 있다.

각 조사의 성격과 내용을 구체적으로 설명하면 다음과 같다.

1) 반복핵심 조사

반복핵심 조사는 자주 반복적으로 물어보는 항목으로서, 그 설문은 각종 사회 인구학적 배경변인들을 비롯하여 정치·경제·사회의 핵심적 요소들을 파악하기 위한 140여 개의 문항들로 구성된다. 이 설문은 학술적 중요성이 현격한 주요 사회적 사안들을 반복적·지속적으로 조사함으로써 한국사회의 안정성과 변동성을 파악한다.

주요 항목별 구체적인 반복핵심 설문 문항은 아래와 같다.

- (1) 정치 관련 문항: 국가자긍심(KRPROUD 2003-2021), 정치성향(PARTYLR 2003-2021), 정치만족도(SATPOL 2003-2014, 2018-2021), 정치전망(POLPROS 2004-2014), 지지정당(PRTYID 2003-2021), 선호정당(PRTYPR 2003-2009, 2013-2021), 국정운영 평가(CURGOV 2003-2018)

북한·민족통일 문제

북한 인식 NORTHWHO 2003-2021

북한이탈주민 수용 NKACCEPT 2011-2016

남북 통일 필요성 UNIFI 2003, 2006-2021

- (2) 경제 관련 문항: 경제만족도(SATECO 2003-2008, 2011-2014), 경제전망(ECOPROS 2003-2014)

- (3) 사회 관련 문항:

사회적 신뢰 및 공정성

일반인 공정성 FAIR 2003, 2005, 2007-2009, 2011-2021

사회 신뢰 RELIABLE 2004-2009, 2011-2014, 2018-2021

사람들 신뢰 CANTRUST 2004, 2007-2008, 2010-2012, 2014-2021

일반인 신뢰 TRUST 2003, 2005-2006, 2009, 2013-2014

사회집단을 이끌어가는 지도자에 대한 신뢰(2003-2021)

대기업 지도층(CONBUS), 국회 지도층(CONLEGIS), 대법원 지도층(CONJUDGE),

학계 지도층(CONSCI), 군대 지도층(CONARMY), 금융기관 지도층(CONFINAN),

청와대 지도층(CONBLUHO), 시민운동단체 지도층(CONCSO),

종교계 지도층(CONCLERG), 교육계 지도층(CONEDUC),

노동조합 지도층(CONLABOR), 신문사 지도층(CONPRESS),

TV방송국 지도층(CONTV), 의료계 지도층(CONMEDIC),

중앙정부부처 지도층(CONGOVT), 지방정부부처 지도층(CONLGOVT),

여론조사기관 운영자(CONRCH 2013-2021), 검찰청 지도층(CONPROS 2021),

통계청 지도층(CONSTAT 2021)

각종 사회문제 해결을 위한 정부 노력에 대한 평가

7) 뉴스를 통한 한국종합사회조사 http://kgss.skku.edu/?page_id=144

소득격차 해소 정부 책임 GOVEQINC 2003, 2009-2011, 2014, 2021

(4) 배경 변수 문항:

응답자

연령 AGE 2003-2021
성별 SEX 2003-2021
혼인상태 MARITAL 2003-2021
월 평균 가구 소득 INCOME/INCOM0 2003-2021
학력 EDUC 2003-2021
취업여부 INCMR 2003-2021, EMPLY 2003-2021
미취업사유 WHYNOE 2003-2021
산업 INDUSTRY 2003-2021, INDUS08 2003-2021, INDUS16 2003-2021
직업 OCC 2003-2021, OCC08 2003-2021, OCC16 2003-2018

배우자

연령 OLD2 2012-2021
성별 GENDER2 2012-2021
학력 SPEDUC 2003-2021
동거여부 AWAY2 2012-2021
비동거사유 WHYAWY2 2012-2021
취업여부 SPINCMR 2003-2021, EMPLY2 2012-2021
미취업사유 WHYNOE2 2012-2021
산업 SPIND 2003-2021, SPIND08 2003-2021, SPIND16 2003-2021
직업 SPOCC 2003-2021, SPOCC08 2003-2021, SPOCC16 2003-2018

가구주

응답자와의 관계 HDHEAD 2003-2021
가구원 번호 HHWHO 2003-2021
연령 HHAGE 2003-2021
성별 HHSEX 2003-2021
혼인상태 HHMARI 2003-2021
취업여부 HHEMPLY 2003-2021
비동거사유 HHWHYAWY 2003-2021
동거여부 HHAWAY 2003-2021
미취업사유 HHWHYNOE 2003-2021

부모의 교육수준

부친학력: 최종학교 PAEDUC 2003-2021
모친학력: 최종학교 MAEDUC 2003-2021

가구원수

동거가족 수 HOMPOP 2003-2021
별거가족 수 SEPAPOP 2003-2021
가족 아닌 동거인 수 UNRELAT 2003-2021

종교여부 및 종교 참석 빈도

응답자 종교 RELIG 2003-2021
응답자 신앙심 RELITEN 2003-2021

종교집회 참석 빈도 ATTEND 2003-2009, 2011-2021

계층귀속

주관적 계층 의식 (10점 척도) RANK 2003-2021
주관적 계층 의식 (6점 척도) CLASS 2003-2005
주관적 계층 의식 (6점 척도) CLASS06 2006-2009

취업상태

응답자 피고용 여부 EMPED 2003-2021
응답자 피고용 고용형태 WGSTAT 2003-2021
응답자 피고용 전일제, 시간제 구분 WGPTFL 2003-2021
응답자 피고용 일주일 평균 근로시간 WGWKHR 2003-2021
응답자 피고용 고용 지속성 WGSTAB 2003-2004, 2006-2021
응답자 피고용 직장의 고용인 수 WGNOEMP 2003, 2005, 2013-2021
응답자 노조 가입 여부 WGUNN 2003-2021
응답자 노조 과거 가입 경험 WGUNNP 2004-2021
응답자 자영업 자영 형태 SFSTAT 2003-2021
응답자 자영업 고용인 수 SFNOHI 2003-2021
응답자 자영업 전일제, 시간제 구분 SFPTFL 2004-2021
응답자 자영업 일주일 평균 근로시간 SFWKHR 2004-2021
응답자 자영업 현업 지속 기간 SFTENYR, SFTENMO 2005, 2007, 2009-2021
응답자 부하직원 수 NUMSUP 2009-2021
응답자 부하직원 유무 EMSUBO 2003-2021
응답자 월평균 근로소득 RINCOM0/RINCOME 2003-2021
응답자 취업 시 근로 외 소득 유무 RXTINC/RXTINC0/RXTINCM 2008-2021
응답자 미취업 일주일 간 구직 노력 UNSRCH 2003-2004, 2006-2021
응답자 미취업 소득 유무 UNXTINC 2008-2021
응답자 미취업 월평균 소득 UNXTINCM 2008-2021
응답자 미취업 근로 경험 여부 UNWKEX 2011-2021
응답자 미취업 최근 근로 지위 UNSTAT 2003-2021
응답자 미취업 최근 고용인 수 UNNOHI 2004-2021
응답자 미취업 당시 부하직원 유무 UNSUBO 2004-2021
응답자 미취업 당시 평균 근무 시간 UNWKHR 2003-2021
배우자 피고용 여부 SPEMPED 2003-2021
배우자 피고용 고용형태 SPWGSTAT 2003-2021
배우자 피고용 전일제, 시간제 구분 SPWGPTFL 2003-2021
배우자 피고용 일주일 평균 근로시간 SPWGWKHR 2003-2021
배우자 피고용 고용 지속성 SPWGSTAB 2003-2021
배우자 노조 가입 여부 SPWGUNN 2003-2021
배우자 노조 과거 가입 경험 SPWGUNNP 2004-2021
배우자 자영업 자영 형태 SPSFSTAT 2003-2021
배우자 자영업 고용원 수 SPSFNOHI 2003-2021
배우자 자영업 전일제, 시간제 구분 SPSFPTFL 2004-2021
배우자 자영업 평균 근로시간 SPSFWKHR 2004-2021
배우자 부하직원 수 SPNUMSUP 2009-2021
배우자 부하직원 유무 SPEMSUBO 2004-2021

배우자 월평균 근로소득 SPINCOME/SPINCOM0 2003-2021
 배우자 근로 외 소득 유무 SPXTINC 2008-2021
 배우자 월평균 근로 외 소득 SPXTINCM/SPXTINC0 2008-2021
 배우자 미취업 일주일 간 구직 노력 SPUNSRCH 2003-2004, 2006-2021
 배우자 미취업 소득 유무 SPUNXTN 2008-2021
 배우자 미취업 월평균 소득 SPUNXTN0/SPUNXTNM 2008-2021
 배우자 미취업 근로 경험 유무 SPUNWKEX 2011-2021
 배우자 미취업 최근 근로 지위 SPUNSTAT 2003-2021
 배우자 미취업 당시 고용인 수 SPUNNOHI 2004-2021
 배우자 미취업 당시 부하직원 유무 SPUNSUBO 2004-2021
 배우자 미취업 당시 평균 근로 시간 SPUNWKHR 2003-2021

고용지위

응답자 고용부문 WGSECT 2003-2021
 배우자 고용부문 SPWGSECT 2003-2021

2) 국가간비교사회조사

한국종합사회조사는 국내의 사회변동뿐만 아니라 ISSP, EASS에 참여하여 국가 간 비교 연구 자료를 수집하고 있다.

(1) ISSP 주제모듈 조사

ISSP는 국제비교 연구 공동 작업이다(<http://www.issp.org/>). 1984년에 창립된 국제사회조사 연대기구로서, 2018년 현재 43개 국가가 그 회원국으로 참여하고 있다. 이 기구의 회원국들은 매년 개최하는 총회에서 공동으로 조사 주제 및 설문문항을 결정하고 조사한다. 이 조사로부터 산출되는 자료는 2018년 4월까지 9,079건의 출판물에 이용되었는데, 국제비교 연구에 있어 가장 중요한 데이터의 하나로 그 가치를 널리 인정받고 있다.⁸⁾

성균관대학교 동아시아학술원 서베이리서치센터는 2003년 ISSP 총회에서 한국을 대표하는 회원기관으로 공식 지정되었고, 2003년 국가정체성 모듈부터 자료를 모아오고 있다. ISSP가 지금까지 개발하여 공동으로 조사한 주제모듈은 모두 11가지이다(참조 <표 1>).

8) Haller, M., Jowell, R., & Smith T. W. (2009). *The international social survey program: Charting the globe*. Routledge.

김지범. (2016). 국제사회조사프로그램: International Social Survey Programme (ISSP). KIPA조사포럼, 16, 50-57.

<표 1> ISSP 주제모듈, 1985-2021

ISSP 주제 모듈	ISSP 주제 모듈연도					한국 참여연도	
정부의 역할(Role of Government)	1985	1990	1996	2006	2016	2006	2016
사회연결망과 사회적자원(Social Networks and Social Resources)*	1986	2001	2017				
사회불평등(Social Inequality)	1987	1992	1999	2009		2009	
가족과 성 역할의 변화(Family & Changing Gender Role)	1988	1994	2002	2012		2012	
노동지향(Work Orientations)	1989	1997	2005	2015		2005	
종교(Religion)	1991	1998	2008	2018		2008	2018
환경(Environment)	1993	2000	2010	2020		2010	2021
국가정체성(National Identity)	1995	2003	2013			2003	2013
시민권(Citizenship)	2004	2014				2004	2014
레저와 스포츠(Leisure Time & Sports)	2007					2007	
건강(Health)	2011					2012	

* 1986년에는 사회연결망 및 지지체계(Social Networks and Support Systems), 2001년에는 사회연결망II(Social Networks II)라는 모듈 명칭을 사용하였다.

최근 및 향후 주제로는 2021년 「건강과 건강보호」, 2022년 「가족과 성 역할의 변화」이다. 이 모듈들 가운데 「레저와 스포츠」를 제외하고는 모두 수년 주기(약 5-10년)로 이미 2-5회씩 순환적으로 조사되었다. 순환 조사되는 모듈에는 통상적으로 예전 모듈 문항의 2/3 가량이 포함되며, 새 문항의 도입이나 기존 문항의 수정은 새로이 부각되는 사회적·학술적 중요성에 의거해서 회원국들 간의 심도있는 토의를 거쳐 결정된다. 2021년까지 ISSP가 지정해놓은 모듈들은 총 11개이므로 향후 더 이상 새로운 모듈이 지정되지 않는 한 이 모듈들은 대체로 10년 간격으로 순환 조사될 예정이다. ISSP는 주제 모듈들을 이처럼 수년 간격으로 반복 조사함으로써 세계의 변동을 국가 간 혹은 대륙 간 비교 맥락에서 분석할 수 있는 자료를 산출해낸다. ISSP의 2021년 「건강과 건강보호」, 2022년 「가족과 성 역할의 변화」의 주요 각 주제모듈에서 다루어질 주요 내용들을 보다 자세히 소개하면 다음과 같다.

① 2021 「건강과 건강보호 II」 모듈

건강 상태, 건강 행위, 건강에 대한 믿음, 의료 보험, 의료 접근 및 이용, 건강 보호에 대한 권리, 건강 보호 체계에 대한 신뢰, 건강 보호 체계 만족도, 백신에 대한 태도, 코로나19 등

② 2022 「가족과 성 역할의 변화 V」 모듈

여성의 일에 대한 태도, 남성과 여성의 역할 분배, 양육에 대한 태도, 결혼과 이혼 태도, 이혼 선호도, 가족에 대한 돌봄 여부, 친척에 대한 돌봄 책임 소재, 일과 생활 균형 정도, 가사 분담, 여성 지도자 역할에 대한 태도 등

(2) EASS 주제모듈 조사

EASS는 동아시아 지역에서 General Social Survey(GSS) 형태의 사회조사를 시행하는 4개국(한국, 일본, 중국, 대만)의 기관들 사이에 동아시아 비교 연구를 도모할 목적으로 한국의 주도로 지난 2003년에 신설된 동아시아 사회조사 연대기구이다. EASS는 지금까지 6번의 조사(2006, 2016년 「Family in East Asia」; 2008년 「Globalization & Culture in East Asia」; 2010년 「Health in East Asia」; 2012년 「Network Social Capital in East Asia」)를 수행하였다. 2015년에 조사 예정이었던 5번째 모듈 주제 「Work Life in East Asia」는 한국과 일본에서 수행되지 못했다. 2018년은 「Globalization & Culture in East Asia」를 2008년에 이어 두 번째 반복 조사했다. 2021년은 2010년에 조사했던 「Health in East Asia」를 다시 조사했다.

EASS 설립 후 서베이리서치센터는 EASS 전체를 대표하여 EASS 원자료의 정선(cleaning), 통합 및 조화화 과정을 거쳐 자료를 구축하고 전 세계에 확산시키는 데이터아카이빙 기관으로서, 주도적인 역할을 수행해오고 있다(EASSDA, <http://www.eassda.org/>). EASSDA의 지금까지 신청건수는 총 4,393건으로 국가별 모듈별 신청 건수는 다음과 같다: (1) 국가별 신청 현황: 중국 1,493건, 한국 566건, 일본 525건, 미국 403건, 영국 236건, 대만 231건, 홍콩 135건 등을 포함한 세계 여러 국가에서 자료 사용을 요청하였다. (2) 모듈별 신청 현황: Family(2006) 1,121건, Culture(2008) 1,112건, Health(2010) 994건, Social networks(2012) 835건, Work life(2015) 133건, Family(2016) 97건, 알 수 없음 101건이다.

향후 EASS는 더 이상 새로운 모듈 주제를 지정하지 않고 기존의 5가지 모듈들을 각각 10년 주기로 반복해서 조사하게 될 것으로 예상하고 있다. 2021년 「Health in East Asia」조사의 주요 내용은 다음과 같다.

① 2020 「Health in East Asia」(동아시아인의 건강) II」 모듈

신체적 건강상태, 정신 건강, 미래가 나아지지 않을 것이라는 절망감, 만성질환, 체격, 흡연, 음주, 건강검진, 병원 방문 빈도, 중병 치료비 지불 능력, 건강관리, 건강보험, 대체의학, 사회적 지원, 사회 신뢰도, 거주 지역의 환경오염도, 거주 지역 공공시설의 적절성, 유행병 백신 접종 여부, 가족 돌봄 필요와 돌봄 관리, 나이 드는 것에 대한 걱정, 중독, 체형, 코로나19 등

(3) 연구자 제안 특별 모듈조사

연구자 제안 특별 주제모듈 조사(special module survey)는 연구자가 제안한 모듈로 연구자가 일정 조사비용을 지불하여 한국종합사회조사에 포함시켜 조사를 실시하였다. 이 연구자 제안 특별 주제모듈 조사는 기존에 연구재단으로부터 받는 조사비용의 부족분을 보충하면서, 각 연구자에게 적은 비용으로 연구자의 조사를 가능케 해주는 장점이 있다. 지금까지 연구자 모듈은 다음과 같다: 한국 사회의 불평등과 공정성(2005, 2009, 2014 김상욱), 문화산업과 소비(2007 이호영), 정신건강과 자살(2009 박재현), 근골격계 질환(2010 박재현), 국내이주민(2010 김상욱), 시민권, 이타주의, 기부행위(2011 박재현), 연령주의(2011 김주현), 인권의식 확산(2011, 2018 서창록/구정우), 한국인의 인터넷/SNS 사용 및 중독(2012 황명진), 한국인의 정신건강(2012 박재현), 위험사회(2013 김미숙), 한국인의 인터넷/SNS 활용(2013 이원재), 한국인의 일자리 선택과 직업가치(2013 유홍준), 한국인의 노동가치(2013 류석춘/김형아), 그리고 위험사회(2014 김미숙), 원자력발전/SNS 활용(2016 이원재), 불평등과 정치참여(2016, 2018 김석호, 박원호), 유교/인간본성(2016 이정환/Anna Sun/김지범), 낙관주의/희망(2016 김지범), ISSP 종교(2018 조화순/강정환), EASS(2016, 2018 서울대사발연/성대), 결혼, 가족, 출산에 대한 세대 간 인식 변화(2018 김영미/계봉오), 연령주의(2018 정순돌/김주현), 인권의식 확산(2018 서창록/구정우), 종교(2018 송재룡/유광석), 일자리/소득정책(2018 임동균), 여론조사(2018 Taeku Lee, 김선민, 하상웅), 국민성/지역사회의 활성화(2018 박요성), ISSP 종교/낙관주의(2018 Tom W. Smith/김지범). 기본소득(2021, 유종성), 정치적 다양성/정당(2021, 조화순), 포퓰리즘/민주주

의(2021, 조원빈), 계층과 사회보험(2021, 임동균), 코로나19(2021, 심재만), 유교(2021, Neha Sahgal/
김지범), 조사에 대한 태도(2021, 김지범)

2. 조사 방법

한국종합사회조사의 완성도를 높이기 위하여 서베이리서치센터는 지난 20년간 부단히 노력하여 왔다. 한국종합사회조사 자료를 이용한 연구들이 신뢰할 만한 결과들을 산출해내기 위해서는 무엇보다도 원자료 자체의 대표성과 전반적 조사과정 오류의 축소가 필수적이다. 한국종합사회조사에서는 양질의 자료를 생산하기 위해 이전 연도 조사들의 지식과 경험을 최대한 살리면서 기존의 조사 과정에서 드러난 문제점 또한 적극 수정·보완하는 노력을 지속적으로 해나가려고 한다.

1) 2003-2021 표본설계

2003-2014

(1) 표본추출

한국종합사회조사의 모집단은 전국의 가구에 거주하는 만 18세 이상, 시설(예, 기숙사, 군부대)에 거주하지 않고, 한국어로 소통 가능한 성인 남녀이다. 제1차(2003), 2차(2004) 한국종합사회조사에서 표본의 크기는 2,000명이었으나 제3차(2005) 한국종합사회조사부터는 표본오차를 줄이고 유효표본의 수를 늘리기 위하여 그 크기를 2,500명으로 증대시켰다. 2003년부터 2005년까지는 안전행정부의 자료를 표집틀로 이용하여서, 가구 표집틀을 작성하지 않았다. 표본추출 방법은 다단계지역확률표집방법(multi-stage area probability sampling)을 사용하였다. 이 표집법은 모두 여러 단계에 걸쳐 인구비례확률(PPS)로 층화된 지역별 집락표본을 추출하는 방식인데, 그 세부적 절차는 다음과 같다.

① 최종 표집블럭 수를 200개로 정하고 이를 전국의 광역시·도를 가장 최근 “주민등록 인구통계”에 기초하여 가구 수 분포에 비례하여 배분한 후, 각 광역시·도로부터 배분된 수만큼의 1차표집단위(primary sampling unit)인 동·읍·면을 인구비례확률표집 원칙을 적용하여 무작위로 추출한다.

② 이렇게 추출된 각 동·읍·면으로부터 2차표집단위(secondary sampling unit), 동의 경우에는 통/반, 면의 경우에는 리를 인구비례확률표집 원칙을 적용하여 무작위로 선정한다.

③ 선정된 통/반 또는 리에서 표집원(lister)이 가구의 표집틀(sampling frame)을 전통적 가구명부(traditional listing) 혹은 주민센터의 가구 관련 자료를 얻은 경우는 의존적 가구명부(dependent listing) 작성 방식을 이용한다. 이렇게 작성된 가구 표집틀을 서베이리서치센터에 보내면, 서베이리서치센터는 선정된 지역에서 모두 12개 혹은 13개의 가구를 무작위로 선정한 후 각 학교 조사팀장에게 보낸다.⁹⁾

④ 조사원은 이 표집틀에 근거하여 선정된 가구를 직접 방문하여, 표집된 가구 내에서 만 18세 이상 성인 가구원들의 리스트를 빠짐없이 작성한 후, 이들 가운데 실제 생일 기준으로 태어난 월일이 가장 빠른 가구원을 최종 면접대상자로 삼는다. 예외는 2006년과 2007년으로 생일이 가장 늦은 가구원을 선정하였다.

2016-2021¹⁰⁾

2016년 이후에는 한국종합사회조사의 모집단은 전국의 가구에 거주하는 만 18세 이상, 시설(예, 기숙사, 군부대)에 살지 않고, 한국어로 소통 가능한 성인 남녀이다. 만약 모든 성인들을 포함한 목록이 존재한다면, 이 목록을 이용하여 표본을 직접 추출할 것이다. 하지만, 그러한 목록이 없기 때문에 우리는 지역확률표본추출방법을 이용하여 4단계 과정을 통하여 잠재적인 응답자를 선정하였다. 4단계 중 3단계는 가구를 선택하는 것이고, 마지막 단계는 선정된 각 가구에서 응답자로 적격한 사람을 선정하는 것이다.

9) KGSS 표집틀 참조: 구묘정, 김솔이, 김하영, 김지범. (2014). 가구조사 표집틀 작성: 한국종합사회조사. 조사연구, 15(4), 153-174.

10) Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: Wiley.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Sarndal, C-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

(1) 1차 표집단위(Primary Sampling Unit) 추출

제 1차 표집단위는 동/읍/면이다. 2016년에 100개, 2018년 이후에는 200개의 1차 표집단위를 인구비례 확률(probability proportional to size) 원칙을 적용하여 무작위로 추출하였다.

① 표집틀(sampling frame)

표집틀은 2015년 지방자치단체 행정 구역 및 인구현황(www.mogaha.go.kr)의 3,496개 동/읍/면 목록을 포함하고 있다. 2021년은 2019년도 지방자치단체 행정구역 및 인구현황의 3,510개 동/읍/면 목록을 포함하고 있다.

② 층화(stratification) 및 표본 할당(sample allocation)

추정치의 효율성을 향상시키기 위해 표집틀을 지역 및 도농 간 차이를 고려하여 실질적 층화(explicit stratification)를 하였다. 1차 표집단위들은 각 층(stratum)인 광역시도에서 독립적으로 추출되었다. 각 광역시도의 가구 수에 비례하여 2016년 100개, 2018년 이후에는 200개의 1차 표집단위를 배분하였다.

MOS_{ij} : 광역시/도(i^{th}) 동/읍/면(j^{th})의 가구 수

n_i : 광역시/도(i^{th}) 1차 표집단위 할당 수

N_i : 광역시/도(i^{th}) 전체 동/읍/면 수

$$n_i = \text{표집단위 수 (2006년 100, 2018년 이후 200)} \times \frac{\sum_{j=1}^{N_i} MOS_{ij}}{\sum_{i=1}^{17} \sum_{j=1}^{N_i} MOS_{ij}}$$

참조로 총 동/읍/면 수는 3,496(혹은 3,510)개 이고, 할당된 총 1차 표집단위 수는 2016년 100, 2018년 이후로는 200이다. 이때 표집단위 할당 수는 반올림된 정수가 이용되었다. <표 2>는 각 층인 광역시도의 가구 수와 이에 비례한 1차 표집단위의 할당 수를 보여준다.¹¹⁾

③ 추출(Selection)

1차 표집단위는 각 광역시도의 동/읍/면 목록에서 체계적으로(systematically) 추출하였다. 표집틀은 표본 선정 이전에 광역시도 그리고 구/군으로 구분하여 정리하였다. 체계적 추출 방법(systematic selection method)에 대한 자세한 내용은 Cochran (1977), Kish (1965), 그리고 Sarndal, et al. (1992)에서 찾을 수 있다.

p_{ij} : 광역시/도(i^{th}) 동/읍/면(j^{th})의 추출 확률(inclusion probability)

$$p_{ij} = n_i \times \frac{MOS_{ij}}{\sum_{j=1}^{N_i} MOS_{ij}}$$

11) 2016년 이후 세종특별자치시는 반올림 시에 1차 표집단위가 0이어야 하지만(포함되지 말아야 함을 의미), 세종특별자치시를 표본에 포함시키기 위해서 임의적으로 표집단위 하나를 배정하였다.

(2) 제 2차 표집단위(Secondary Sampling Unit)의 추출

2차 표집단위는 동의 경우 통/반 그리고 면의 경우 통/리로 정의된다. 추출된 각 동/읍/면으로부터 한 개의 2차 표집단위를 인구비례확률표집 원칙을 적용하여 체계적으로(systematically) 선정한다.

MOS_{ijk} : 광역시/도(i^{th}) 동/읍/면(j^{th}) 통/반/리(k^{th})의 가구 수

$p_{k|ij}$: 각 광역시/도(i^{th}) 동/읍/면(j^{th})에서 통/반/리(k^{th})의 조건 추출 확률

$$p_{k|ij} = \frac{MOS_{ijk}}{MOS_{ij}}$$

(3) 가구 추출

선정된 통/반 또는 통/리에서 표집원(sampler)이 가구의 표집틀(sampling frame)을 전통적 가구명부(traditional listing) 혹은 주민센터의 가구 관련 자료를 얻은 경우는 의존적 가구명부(dependent listing) 작성 방식을 이용한다. 이렇게 작성된 가구 표집틀을 서베이리서치센터에 보내면, 서베이리서치센터는 선정된 지역에서, 미리 정해진 가구 수(2016년 16-24가구, 2018년 이후 12가구)를 무작위로 추출하였다.

$p_{l|kij}$: 각 광역시/도(i^{th}) 동/읍/면(j^{th})에서 통/반/리(k^{th})에서 가구의 조건 추출 확률

$$p_{l|kij} = \frac{m}{MOS_{ijk}}$$

가구의 전체 비조건부 추출 확률은 상수로 2016년 이후 KGSS 가구 표본은 동일확률추출방법(equal probability of selection method, 혹은 epsem)으로 선정된 표본이다.

P_{ijkl}

$$\begin{aligned} &= \text{표집단위수 (2016년 100, 2018년 이후 200)} \times \frac{\sum_{j=1}^{N_i} MOS_{ij}}{\sum_{i=1}^{17} \sum_{j=1}^{N_i} MOS_{ij}} \times \frac{MOS_{ij}}{\sum_{j=1}^{N_i} MOS_{ij}} \times \frac{MOS_{ijk}}{MOS_{ij}} \times \frac{m}{MOS_{ijk}} \\ &= \text{표집단위수 (2016년 100, 2018년 이후 200)} \times \frac{m}{\sum_{i=1}^{17} \sum_{j=1}^{N_i} MOS_{ij}} \end{aligned}$$

(4) 가구원 추출

표집된 가구 내에서 만 18세 이상 성인 가구원들의 목록을 작성한 후, 2016년에는 이들 가운데 조사 시점으로부터 생일을 지낸 시기가 가장 가까운 가구원을 선정하는 방법(last birthday method)으로 최종 면접대상자를 선정하였다. 2016년 KGSS 최종 면접자 수는 1,051명이다. 2018년에는 키시테이블(Kish table)을 이용하였고, 2021년에는 가장 최근에 생일을 지낸 가구원을 최종 면접대상자를 선정하였다. KGSS 최종 면접자 수는 2018년 1,031명이고, 2021년 1,205명이다.

<표 2> 층(stratum)의 크기와 표본 할당: 2016, 2018

층	시/도	가구 수	%	2016년 할당 1차 표집단위*	2018년 할당 1차 표집단위*
1	서울특별시	4,194,176	20.23	20	40
2	부산광역시	1,416,648	6.83	7	14
3	대구광역시	970,618	4.68	5	9
4	인천광역시	1,136,280	5.48	5	11
5	광주광역시	573,043	2.76	3	6
6	대전광역시	592,508	2.86	3	6
7	울산광역시	442,250	2.13	2	4
8	세종특별자치시	62,807	0.30	1	1
9	경기도	4,786,718	23.08	23	46
10	강원도	673,978	3.25	3	7
11	충청북도	656,321	3.17	3	6
12	충청남도	871,459	4.20	4	8
13	전라북도	774,562	3.74	4	8
14	전라남도	840,864	4.06	4	8
15	경상북도	1,153,559	5.56	6	11
16	경상남도	1,343,984	6.48	6	13
17	제주특별자치도	246,516	1.19	1	2
전체		20,736,291	100	100	200

* 세종특별자치시에 1개의 1차 표집단위 할당

<표 2> 층(stratum)의 크기와 표본 할당: 2021

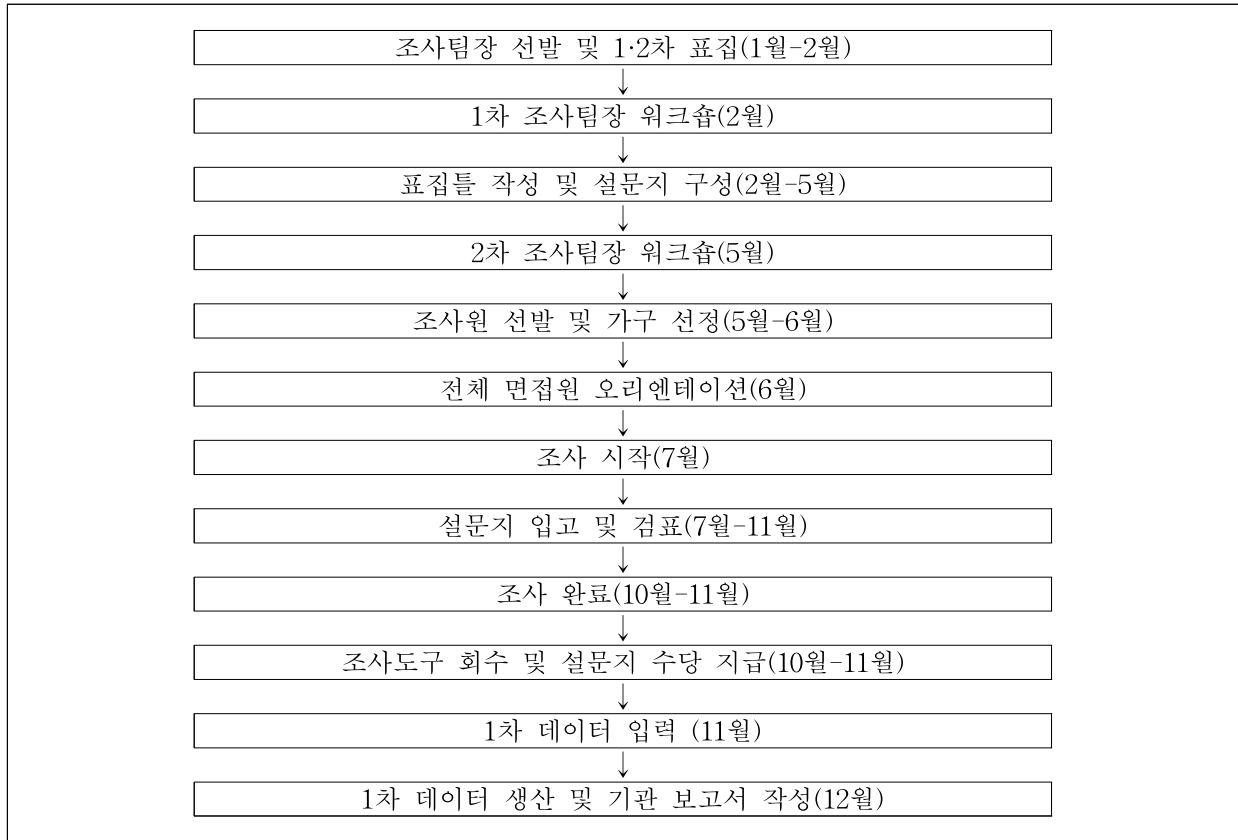
층	시/도	가구 수	%	2021년 할당 1차 표집단위*
1	서울특별시	3,839,765	19.22	39
2	부산광역시	1,363,608	6.83	14
3	대구광역시	957,516	4.79	10
4	인천광역시	1,094,749	5.48	11
5	광주광역시	578,559	2.90	6
6	대전광역시	602,175	3.01	6
7	울산광역시	431,391	2.16	4
8	세종특별자치시	119,029	0.60	1
9	경기도	4,751,498	23.78	48
10	강원도	628,484	3.15	6
11	충청북도	640,977	3.21	6
12	충청남도	851,124	4.26	9
13	전라북도	732,980	3.67	7
14	전라남도	737,406	3.69	7
15	경상북도	1,094,535	5.48	11
16	경상남도	1,306,394	6.54	13
17	제주특별자치도	248,998	1.25	2
전체		19,979,188	100	200

* 세종특별자치시에 1개의 1차 표집단위 할당

2) 현지조사

2003년부터 2018년까지 현지조사는 비용적인 면에서나 자료의 질 면에서도 효과적인 것으로 입증된 이른바 전국대학 조사네트워크(inter-university survey network)를 이용하였다. 전국대학조사네트워크는 전국 대학에 각각 협력교수 1명과 조사팀장 1명 그리고 면접조사원 5~15명으로 구성되는 조사팀을 조직한다(참조: 부록(5)_KGSS 참여 교수·학생 명단). 각 대학의 협력교수는 조사팀을 지도하고, 조사팀장은 담당지역의 표집과 현지조사를 관리하고 감독하는 역할을 맡는다. 현지조사를 보다 원활하게 수행하기 위하여 지난 2004년에 최초로 도입된 ‘지역담당 조사팀장’ 제도는, 현지조사의 완성도를 높이는데 크게 기여하였다. 지역담당 조사팀장은 사회조사방법에 대한 기본적인 지식과 현지조사 경험이 있는 대학생으로 충원하는 것을 원칙으로 한다. 이들을 포함해서 모든 면접조사원들이 한국종합사회조사의 목적과 의의를 이해하고 설문지의 내용과 조사 진행방식 등과 관련된 중요한 사항들을 숙지하는 것은 현지조사를 성공적으로 수행하는데 필수적이다. 2021년은 조사경비와 코로나19 상황으로 인하여 조사업체가 현지 조사를 담당하였다.

[그림 1] 조사추진과정흐름도



위의 [그림 1]은 2016년 37개 대학(2018년 45개 대학)이 참여한 조사추진과정이다. 2016년 전국 37개 대학(2018년 45개 대학)의 협력 교수를 통해 조사팀장 역할을 수행할 학생을 추천 받는다. 센터에서는 1차 표집단위인 동/읍/면 100개(2018년 200개)를 추출하고, 이어서 2차 표집단위인 통/반 혹은 통/리를 선정한다. 각 대학의 조사팀장과 2차 표집단위가 확정된 후 1차 조사팀장 워크숍을 개최한다. 워크숍에서는 조사과정에 대한 전반적인 설명과 조사팀장들의 역할에 대한 안내, 표집틀 작성에 대한 교육을 실시한다. 워크숍이 끝나면 팀장들은 선정된 통/반 혹은 통/리에 직접 방문하여 표집틀을 그리게 된다. 조사팀장이 작성한 표집틀이 모두 수거되면 센터에서는 이를 검토하고 부정확하거나 불완전한 부분을 조사팀장을 통해 수정·보완을 하도록 한다. 센터에서는 조사팀장들이 표집틀을 작성할 동안 설문지 개발에 착수한다. 표집틀이 완성되면, 2차 조사팀장 워크숍을 개최하여 조사원 선발에 대한 교육과 센터에서 개발한 설문지를 바탕으로 사전조사 교육을 실시한다. 조사팀장이 면접원을 선정하고 사전조사를 수행하는 동안 센터에서는 표집틀을 바탕으로 최종 가구를 선정한다. 모든 준비가 끝나면 모든 협력대학의 조사팀장과 면접원을 센터로 소집해 전체 면접원 오리엔테이션을 개최한다. 이때, 센터에서는 조사를 하는 목적과 실질적인 조사 방법에 대해 교육하고, 면접원은 오리엔테이션이 끝난 후에 바로 조사에 착수한다. 2018년에는 조사경비 문제로 인하여, 면접원 교육은 각 대학 팀장의 주관으로 각 대학에서 “전체면접원 교육영상”을 통해 면접원 교육을 실시하였다¹²⁾.

2016년 이후에는 선정된 가구를 홀수번호 가구와 짝수번호 가구로 나누는 표본 분할을 하였다. 조사원은 홀수가구는 A형 설문지를 이용하고, 짝수번호 가구는 B형 설문지를 이용하였다. 이를 통해 처음으로 조사실험을 수행했다.¹³⁾ 면접원이 작성한 설문지는 팀장이 수거하여 검표를 한 뒤에 센터

12) https://youtu.be/XR10X_w9VI8

로 보내고, 센터에서는 받은 모든 설문지를 다시 검토하면서 자료 입력을 시작한다. 조사가 끝나면 설문지와 미완료 기록표, 그리고 응답자 인센티브 등 모든 조사도구를 회수하여 확인한 뒤 설문지 수당을 지급한다. 자료 입력이 끝나면 바로 정선 작업이 시작된다.

전반적 조사과정과 관련된 세부사항은 다음과 같다.

첫째, 조사방법은 면접조사이지만, 조사원이 조사가구를 방문하기 전에 조사가구에 조사참여를 부탁하는 편지를 보내거나 조사에 응답하지 않는 가구 혹은 접촉이 힘든 가구에도 여러 번의 편지를 보내고 있다. 면접대상자와의 접촉 방식에서는 혼합식조사(mixed-mode survey)의 성격을 띠고 있다고 생각할 수 있다.

둘째, 센터 연구원은 ‘지역담당 조사팀장’ 및 조사원이 조사를 수행하는데 어려움을 덜어 주기위해서 조사기간 동안 지속적으로 접촉한다. 조사팀장들이 표집틀을 작성하는데 도움이 되고자, 센터에서는 표집틀 지역 전체 동사무소와 통화를 하고 협조공문을 보냈다. 또한, 아파트관리소의 관리소장과 의 통화 및 협조공문을 통하여 조사원이 아파트/빌라/연립 등의 해당 가구에 쉽게 접촉할 수 있도록 최대한 노력하였다. 또한, 센터 조교와 조사팀장과는 수시로 통화 및 카톡을 통하여 조사팀장 및 조사원의 문제점을 즉각적으로 해결할 수 있도록 하였으며, 특히 2016년과 2018년의 경우 연구책임자와 ‘각 지역대학 조사팀장’의 단체카톡방 개설을 통하여 연구책임자가 현장의 고충 및 문제점을 즉각적으로 해소하려고 하였다.

셋째, 서베이리서치센터는 어떠한 경우에도 면접조사원이 표본을 임의로 대체하는 것(가구 혹은 가구원 대체)을 절대로 허용하지 않는 원칙을 고수하고 있다. 표본가구가 대체되었거나 표본가구내의 면접대상자를 잘못 선정한 사례가 발견되었을 경우에는 예외 없이 재조사를 한다.

넷째, 서베이리서치센터는 조사원의 ‘지속적인 반복방문’을 강조하여 응답률을 최대한 높이려고 노력한다.

다섯째, 현지조사를 통해 수집된 설문지는 곧바로 검표작업을 해서 누락된 응답이 있는지 논리적 일관성이 결여된 응답이 있는지 등을 면밀히 점검하여, 문제가 있는 경우, 전화 또는 재방문을 통한 보충 조사를 즉각적으로 시도한다. 이전 연도에 보통 전체 설문지가 수거된 이후에 전화조사를 통하여 결측이 있는 항목에 대한 전화조사를 하였는데, 2016년과 2018년에는 실시하지 않았다. 이유는 실지 면접조사와 전화 조사간의 시기상의 문제로 인하여 전화조사가 조사시점의 태도 및 행동을 반영하지 못한다는 문제점이 있었고, 결측치를 줄일 수 있는 장점이 있지만 실질적으로 어떤 문항에서 결측이 생기지를 제대로 파악할 수 없는 단점이 있었기 때문이다.¹⁴⁾

여섯째, 현지조사 시 설문이 기입되는 방식에 있어서, 설문지를 이용한 면접 조사방식(Paper and Pencil Interviews, PAPI)을 사용하는 것을 원칙으로 하였다. 하지만, 조사원 면접이 불가능한 경우, 설문지 자기 기입방식(Self-administered Questionnaire, SAQ)을 이용하였다. 2021년에는 테블릿을 이용한 면접 조사방식(Tablet-assisted Personal Interview, TAPI)과 설문지를 이용한 면접 조사방식(PAPI)을 혼합하였는데, 1,205명의 응답자 중 90.4%는 테블릿으로, 9.6%는 종이설문지로 조사가 되었다.

일곱째, KGSS는 보기카드(show card)를 활용하였다. 보기카드란 설문지의 응답 범주들을 별도의 카드에 크고 보기 좋게 인쇄한 것이다. 조사원은 면접진행시 각 질문에 해당하는 보기카드를 응답자에게 보여주면서 면접을 진행함으로써 면접의 효율성을 높이는데 이용하였다. 이전 연도와 달리, 2016년 이후는 전 문항에 대하여 보기카드를 활용하였다. 예로, “귀하는 사람들이 대체로 남을 도우

13) 김지범, 김솔이, 강정환. (2017). 서베이조사실험을 통한 폐쇄형과 개방형 설문 응답 차이: 2016년 한국종합사회조사. *조사연구*, 18(4), 127-147.

14) 결측치가 적은 경우는 조사기관에서 결측치를 줄이려는 노력의 결과 일 수 있다.

김지범, 김솔이. (2015). 무응답이 높은 설문문항: 한국종합사회조사 2003-2010. *조사연구*, 16(4), 147-172.

Kim, J., Son, J., Kwok, P. K., Kang, J., Laken, F., Daquilanea, J., Shin, H. C., & Smith, T. W. (2015). Trends and correlates of income nonresponse: Forty years of the US General Social Survey (GSS). *Journal of The Korean Official Statistics*, 20(1), 1-23.

려 한다고 생각하십니까, 아니면 자신만을 위한다고 생각하십니까?”라는 질문에 대한 설문지 응답 범주는 “① 남을 도우려고 한다 ② 자신만을 위한다 (3) 경우에 따라 다르다 (8) 모르겠다”이다. 한국종합사회조사에서 동그라미 안에 번호가 있는 경우는 조사원이 읽어주고, 괄호로 된 번호는 읽이지 않고 응답자가 대답한 경우만 기록하게 되어 있다. 이전에 혹시라도 교육을 받은 조사원이 실수로 “(3) 경우에 따라 다르다”로 읽어준 경우가 있을 수 있었지만, 이번에는 응답자가 보기카드에서 “① 남을 도우려고 한다 ② 자신만을 위한다”라는 응답범주만 보기 때문에 “(3) 경우에 따라 다르다”를 선정할 가능성이 낮아졌을 가능성이 있다.

끝으로, 지난 20년간의 한국종합사회조사에서는 설문지의 맨 뒤쪽에 소위 <조사원 면접상황 기록표>를 별도로 배치해놓고 면접을 끝낸 조사원으로 하여금 면접 당시의 상황(예, 면접일시, 면접 시 주위 사람 존재 여부, 응답자의 면접 태도, 질문 이해 정도, 접촉 어려움 등)을 즉각적으로 기록하도록 하였다.

3. 자료 이용

1) 자료의 입력 및 처리

한국종합사회조사는 자료의 수집 과정에서뿐만 아니라 자료의 입력, 편집(editing), 부호화(coding), 정제(cleaning), 아카이빙(archiving) 과정에서도 자료의 신뢰성을 높이기 위해 최선의 노력을 다하고 있다.

(1) 자료의 입력 및 정제

센터로 회수된 완료 설문지는 검토 작업을 거쳐, 설문지 응답 범주에 표시된 번호를 각 설문 문항 옆에 표기한다. 이를 기초로 자료입력 전문가를 통해서 자료 입력을 하였다. 이전과 달리, 2016년부터는 데이터를 독립적으로 두 번 입력하여 두 개의 입력자료 간 불일치가 있는 경우 다시 설문지를 확인하는 과정을 거쳐 데이터의 입력 오류를 수정하여 최종 입력 자료를 만들었다. 2018년에는 불일치 입력자료를 수정한 이후에 다시 모든 설문지와 입력자료를 비교하여 일치하는 입력자료도 재차 확인하였다. 2021년 자료는 이전과 달리 처음으로 조사회사에서 개인면접을 수행했다. 2003-2018에는 종이설문지 면접조사(PAPI)였지만, 2021년에는 태블릿PC기반 면접조사(TAPI)와 종이설문지 면접조사(PAPI)를 두 가지 방법을 사용하였다. 이런 이유로 센터에서는 자료를 직접 입력하지 않았으며 조사회사에서 입력된 자료에 기초하여 정제작업을 하였다.

(2) 클리닝 방법

입력된 자료는 SPSS 프로그램의 범위(최소값, 최대값), 교차분석 등을 사용하여 정제 과정을 거친다. 비논리적 응답은 설문지를 다시 확인하여 수정하고, 설문에 해당되지 않는 응답자가 응답한 경우는 비해당으로 처리하였다. 2016년부터는 응답자가 연관된 두 문항에 대하여 상충되는 응답을 한 경우에 있어서 다른 문항을 통하여 어떤 응답이 오류인지를 파악할 수 없는 경우는 주관적으로 판단하여 수정하지 않고, 상충되는 응답을 그대로 공개하려고 하였다. 또한, 가능한 한 오류 수정은 응답자 ID를 이용한 수정이 아니라, 변수를 활용한 수정을 하려고 하였고, 향후 정제 방식의 동일성을 유지하고자 수정된 명령문 파일(syntax file)을 잘 보관하여 클리닝 방식에 있어서 일관성을 유지하려고 하였다. 즉, 향후 데이터에서 반복되는 변수의 수정은 현재 저장된 명령문 파일을 통하여 1차 수정이 가능하도록 하였다.

조사에 응답하는 개인들이 개인정보 및 비밀보장에 더 민감한 상황에서, 2003-2021 누적자료에는 시/도를 제외한 지역정보를 공개하지 않기로 하였고, 조사원의 개인정보(소속학교, 이름 등) 또한 제외하였다. 물론, 연구자들은 개별적으로 지역자료이용신청(<https://kgss.skku.edu/kgss/Geographic.do>)을 통하여 상세한 지역 정보 자료를 이용할 수 있다.

2) 산업 및 직업 분류 과정

2003-2014

산업 및 직업 부호화(coding)는 개방형 설문인 응답자와 배우자의 ‘산업 및 직종(직장명)’, ‘직위’, ‘업무내용’ 3가지 항목 이용하였다(<표 3>). 산업 및 직업 분류는 설문지에 바로 부호화하여 기입하였고 이를 자료 입력전문가를 통해서 입력하였다. 산업 및 직업 변수명은 응답자의 경우 INDUSTRY, OCC이고, 배우자의 경우는 SPIND, SPOCC이다. 코딩 작업은 한 번의 입력으로 완료되었다.

<표 3> 응답자와 배우자의 산업 및 직종, 직위, 업무내용

54. 귀하는 어떤 산업에서 어떤 종류의 일을 하십니까? 구체적으로 말씀해 주십시오. 현재 일을 하지 않고 있다면 가장 최근에 했던 일에 대해서 말씀해 주십시오.

면접원
지침서

54.1 산업 및 직종(직장명) : _____
54.2 직위 : _____
54.3 업무내용 : _____

56. 귀하의 배우자는 어떤 산업에서 어떤 종류의 일을 하십니까? 구체적으로 말씀해 주십시오. 현재 일을 하지 않고 있다면 가장 최근에 했던 일에 대해서 말씀해 주십시오.

면접원
지침서

56.1 산업 및 직종(직장명) : _____
56.2 직위 : _____
56.3 업무내용 : _____

2016-2021

2016-2018년 한국종합사회조사 산업 및 직업분류는 통계청 한국표준산업분류(KSCI)와 한국표준직업분류(KSCO)에 기초하여 분류하였다. 한국표준산업분류는 사업체가 주로 수행하는 산업활동의 유사성에 따라 분류하며, 한국표준직업분류는 수입을 위해 개인이 하고 있는 일(경제활동)의 형태에 따라 분류한다. 산업분류는 통계청 한국표준산업분류 자료집을 참고하고, 직업분류는 통계교육원 직업분류 교육자료(2016년)를 바탕으로 직업분류 원칙, 직업분류 분류체계별 유의사항 등을 활용하였다.

또한 통계청 통계분류포털(https://kssc.kostat.go.kr:8443/ksscNew_web/index.jsp)의 ‘한국표준산업분류, 한국표준직업분류-검색-분류내용보기(해설서)’에서 ‘검색어’를 이용해 ‘분류코드’, ‘설명’, ‘예시’, ‘색인어’를 참고하였다. 통계분류포털 해설서는 해당 검색어를 입력하면 연관된 ‘분류코드’와 ‘분류 항목명’이 검색되고, 유사하지만 입력한 직업/산업에 속하지 않은 경우는 ‘제외’로 안내한다. 기본적으로 KGSS 산업 및 직업 분류는 통계청 자료와 해설서를 참고하지만 세부 규칙이 필요할 경우 KGSS 자체 기준을 따랐다. 위와 같은 과정을 거쳐 최종적으로 산업, 직업코드를 입력하였다. 한국종합사회조사에서 산업분류는 중분류(2자리)로 직업분류는 세분류(4자리)와 세세분류(5자리)이다.

2016-2018 KGSS 산업 및 직업 분류는 먼저 응답자(배우자)의 산업, 직위, 하는 일의 3개 설문 응답을 엑셀 프로그램에 전부 입력하였다. 3개 설문 응답만을 가지고 직업 및 산업을 분류하기 쉽지 않은 경우도 있어 산업 및 직업 부호화(coding)에 성별, 연령, 교육수준, 소득, 취업여부 등의 항목을 산업 및 직업 부호화에 이용하였다. 2018년에는 산업 및 직업 코딩은 2단계로 진행하였다. 1단계 산업직업분류는 두 명의 분류자(코더)가 독립적으로 산업 및 직업 부호값을 기입하고, 2단계 산업직업분류는 제 3의 분류자가 1단계의 두 명의 분류자(코더)가 선정한 값을 비교 검토한 후, 최종 부호값을 결정하였다. 한국종합사회조사 전체 자료의 산업 및 직업 부호값의 일관성을 유지하고자, 2016-2018년 산업 및 직업 분류 기준에 따라 새롭게 2003-2014년 산업 및 직업분류 작업을 하였다.¹⁵⁾

61. 귀하의 아버님께서는 귀하가 16살 때 어떤 산업에서 어떤 종류의 일을 하셨습니까? 구체적으로 말씀해 주십시오.

면접원
지침서

61.1 산업 및 직종(직장명) : _____

61.2 직위 : _____

61.3 업무내용 : _____

___ (-1) 16살 이전에 아버지가 돌아가셨음

2021 KGSS 산업 및 직업 분류는 기본적으로 2016-2018년 산업 및 직업 분류에 기본적으로 동일하나, 2021년에는 응답자가 16살 때 아버지 산업 직업 문항이 추가되었다. 먼저 응답자(배우자)의 산업 및 직업 코딩은 엑셀 프로그램에 입력된 산업, 직위, 하는 일 3개 설문 응답을 성별, 연령, 교육수준, 소득, 취업 여부, 고용 지위 등 항목과 함께 고려하여 산업 및 직업 부호화 하였다. 응답자와 배우자의 산업 및 직업 분류와 달리 응답자가 16살 때 아버지 산업 및 직업 코딩은 산업, 직위, 하는 일 3개의 직접적인 설문 응답 외에 산업 및 직업을 판단할 수 있는 보조 문항으로 교육수준 문항만 활용할 수 있었다. 최종 코딩 직전에 응답자의 연령을 기준으로 세대별로 구분하여 검토하는 작업을 추가하였다. 또한 2021년 산업 및 직업 코딩은 2단계로 진행하였다. 1단계에서 두 명의 코더가 독립적으로 산업 및 직업 분류를 한 코딩 값과 2단계 한 명의 코더가 독립적으로 산업 및 직업 코딩을 한 값을 서로 비교하여 최종 산업 및 직업 코딩 입력값을 정하였다.

3) 가중치(weight)변수

2003-2014

한국종합사회조사는 2012년을 제외하고 가중치를 제공하지 않았다.

2016-2021

한국종합사회조사는 2016년 처음으로 모든 연도 분석에 사용할 수 있는 가중치(변수명 FINALWT)를 제공하였다.

가중치 산정은 다음과 같다

$W_{1,i}$ 를 각 연도 표본의 i 번째 표본 성인에 대한 초기 가중치로 설정한다. 표본 가구단위는 각 연도 조사 자료에서 각각 동일 확률(equal probability)을 가지고 추출되었다. 표본 가구 내에서 적격 가구원 중 한 명의 성인(18세 이상)이 무작위로 추출되었다. 따라서 $W_{1,i}$ 는

$$W_{1,i} = E_i$$

여기서 E_i 는 추출된 표본 가구에서 적격한 성인 가구원 수이며 가구 내 선발 확률의 역수와 같다.¹⁵⁾

무응답편향(non-response bias)과 포함오차(coverage error)가 없다면, 초기 가중치($W_{1,i}$)는 KGSS

15) Kim, ChangHwan, Jibum Kim, and Mihee Ban. (2020). "Do you know what you do for a living? Occupational coding mismatches between coders in the Korean General Social Survey." Research in Social Stratification and Mobility 70. <https://doi.org/10.1016/j.rssm.2019.100467>

15) 2003-2021년 자료에서 혼자 사는 가구의 17명에 대한 연령을 알 수 없었다. 한국종합사회조사의 조사 대상 적격가구원은 18세 이상의 성인이기 때문에 연령을 알 수 없는 경우에 1이라는 가중치를 부여했다. 또한, 2003-2021년 자료에서 여러 사람이 사는 가구 중 몇 명의 가구원의 연령을 알 수 없는 경우, E의 값은 연령결측이 없는 성인 수의 값을 이용하였다.

분석의 가중치가 된다. 초기 가중치에 대한 사후층화 (post-stratification) 보정을 아래 4가지 변수를 사용하였다.

- j: 성별(남, 여)
- k: 나이(18-29세, 30-39세, 40-49세, 50-59세, 60세 이상)
- l: 지역(서울, 경기도, 강원도/제주, 충청도, 경상도, 전라도)
- m: 도시성(도시, 시골)

각 년도에 대해, T_{jklm} 은 성별, 나이, 지역, 도시성의 조합에 대한 모집단 전체이다. 셀 내의(j, k, l, m) i^{th} 번째 사람에 대한 사후 층화 가중치인 $W_{2.ijklm}$ 은 다음과 같다. $W_{1.ijklm}$ 은 위에서 언급한 E_i 이다.

$$W_{2.ijklm} = W_{1.ijklm} \frac{T_{jklm}}{\sum_i W_{1.ijklm}}$$

만일 모집단 전체에 대한 위의 성별/나이/지역/도시성 변수의 교차 셀인 사원 분류표(4-way tables)가 이용 가능하고 각 년도 KGSS에서의 완료된 케이스가 사원 분류표를 사용할 만큼 충분히 크다면(예, 각 셀에 20개 이상), 위의 식을 사용해서 사후 층화된 가중치를 생성할 수 있다. 그러나, 4원 교차표를 활용할 수 있을 정도로 KGSS의 케이스 수가 많지 않고, 인구주택총조사 분포에서 네 가지 변수의 결합 분포를 이용할 수 없기 때문에, 초기 가중치에 대한 사후 층화 보정의 대안적 방법(레이킹 가중법(raking weighting) 혹은 반복비례가중법(iterative proportional weighting))을 이용하기로 결정했다.

레이킹 방법은 위의 4가지 변수의 알려진 주변분포(marginal constraints)를 이용해서 각각의 변수를 차례로 이용해서 표본의 분포를 모집단의 분포에 일치시키는 과정을 반복하는 방법으로 T_{jklm}^* 을 산출한다.

$$\begin{aligned} T_{j++++}^* &= \sum_k \sum_l \sum_m T_{jklm}^* = T_{j++++}, j=1, \dots, J \\ T_{+k+++}^* &= \sum_j \sum_l \sum_m T_{jklm}^* = T_{+k+++}, k=1, \dots, K \\ T_{++l+}^* &= \sum_j \sum_k \sum_m T_{jklm}^* = T_{++l+}, l=1, \dots, L \\ T_{+++m}^* &= \sum_j \sum_k \sum_l T_{jklm}^* = T_{+++m}, m=1, \dots, M \end{aligned}$$

성별, 연령, 지역, 도시성에 대한 인구분포(marginal population distribution)는 인구주택총조사를 이용하였다. <표 4>는 사후층화보정에 이용한 18세 이상 인구의 성별, 연령, 지역, 도시성의 분포이다. <표 4>에서 연도 아래에 세 개의 열은 2005년, 2010년, 2015년, 2020년 인구주택총조사 자료 분포이다. 2005년에 대한 분포는 2003년부터 2007년의 KGSS 자료 보정에 이용한 인구구성비(control total)이다. 마찬가지로, 2010년 구성비는 KGSS의 2008-2012년도에 적용되었고, 2015년 구성비는 KGSS 2013-2018에 적용되었고, 2020년 구성비는 KGSS 2021에 적용되었다.

첫 번째 단계로, 추정량은

$$T_{jklm}^{(1)} = n_{jklm} \frac{T_{j++++}}{n_{j++++}},$$

여기서 n_{jklm} 는 (j, k, l, m) 셀 내 응답 완료 숫자이다. n_{jklm} 이 0일 때 추정되는 셀 내 응답 완료 숫자는 0이다. 사후층화에 이용한 변수를 조합할 때 n_{jklm} 이 0이 되지 않도록 노력하였지만, 네 변수를 이용하기 위해서는 0을 완전히 배제할 수는 없었다. 각 년도에 완료된 인터뷰 숫자인 n 이 $\sum_j \sum_k \sum_l \sum_m n_{jklm}$ 과 같다.

다음 추정치는,

$$T_{jklm}^{(2)} = T_{jklm}^{(1)} \frac{T_{j+++}}{T_{j+++}^{(1)}},$$

그리고, k, l, m 변수에 관해서도 위에서 언급한 방법을 반복한다.¹⁶⁾ 레이킹 방법을 통한 값은 빠르게 수렴되었는데, 가령 2016년 한국종합사회조사는 7번의 반복으로 값이 수렴되었다.

가중치 계산을 시작하기 전에, 2003년부터 2021년의 자료에서 21개의 연령 결측치는 관찰된 분포에 비례한 확률에 의거하여 대치값(imputation values)으로 배정하였다.

(j, k, l, m)셀의 i번째 사람에 대한 레이킹 된 가중치(raked weights)는

$$W_{2,ijklm} = W_{1,ijklm} \frac{T_{ijklm}^*}{\sum_i W_{1,ijklm}}$$

이고 최종 가중치 $W_{3,i}$ (FINALWT)는

$$W_{3,i} = W_{2,i} \frac{n}{\sum_i W_{2,i}}, \text{ 혹은 } \sum_i W_{3,i} = n$$

으로 나타난다. 즉 여기서 마지막 가중치의 합은 각 년도의 응답자 수와 같다.

16) Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass: MIT Press.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data*, 2nd ed. Cambridge, Mass: MIT Press.

Little, R. J. A and Rubin, D. B. (1987). *Statistical Analysis with Missing Values*. NY: Wiley.

<표 4> 18세 이상 인구 특성 분포¹⁾

특성	특성범주	연도			
		2005 ²⁾	2010 ³⁾	2015 ⁴⁾	2020 ⁴⁾
성별	남자	0.491	0.490	0.495	0.495
	여자	0.509	0.510	0.505	0.505
연령	18-29세	0.237	0.209	0.189	0.180
	30-39세	0.227	0.204	0.181	0.156
	40-49세	0.222	0.215	0.207	0.187
	50-59세	0.142	0.172	0.196	0.196
	60세 이상	0.173	0.199	0.227	0.282
지역	서울	0.215	0.207	0.198	0.189
	경기도	0.265	0.282	0.293	0.310
	강원도/제주	0.042	0.041	0.042	0.043
	충청도	0.101	0.103	0.106	0.108
	경상도	0.270	0.264	0.260	0.252
	전라도	0.106	0.102	0.101	0.099
도시성	도시	0.809	0.816	0.815	0.811
	농촌	0.191	0.184	0.185	0.189
주석: 1) 출처: 2005, 2010, 2015, 2020 인구주택총조사 인구구성비 2) 2003-2007 KGSS에 적용된 인구구성비 3) 2008-2012 KGSS에 적용된 인구구성비 4) 2013-2018 KGSS에 적용된 인구구성비 5) 2021 KGSS에 적용된 인구구성비					

4) 누적자료 생산

기존에는 연도별 자료와 통계자료집을 출판하였다. 2016년부터는 연도별 자료와 통계자료집, 그리고 2003-2013, 2003-2014 자료를 한국종합사회조사 홈페이지를 통하여 공개하였고, 누적자료집을 출판하였다. 2016년부터는 연도별 자료집은 출판하지 않고, 지속적으로 누적자료와 누적자료집을 제공하여, 연구자들이 개별적으로 누적자료를 만드는 불편함을 최소화하려고 했다.

5) 자료의 아카이빙

정확한 입력과 정밀한 정선을 거친 한국종합사회조사 자료는 국내외 해외 아카이빙 기관에 저장되어 이용자들에게 무료로 자료를 제공하고 있다.

- ① 한국종합사회조사(KGSS) (<http://kgss.skku.edu/>)
- ② 한국사회과학자료원(KOSSDA) (<http://www.kossda.or.kr/>)
- ③ ZA(Zentralarchiv fuer Empirische Sozialforschung, University of Cologne)
(<http://www.gesis.org/issp>)
- ④ 동아시아사회조사(EASSDA) (<http://eassda.org/>)
- ⑤ ICPSR(Inter-University Consortium for Political & Social Research)
(<http://www.icpsr.umich.edu>)