

한국인의 행복조사(3차조사)

1 표본설계

가. 모집단

본 조사의 목표 모집단은 조사 시점 기준 대한민국 만 15세 이상의 일반 국민이다. 본 조사의 조사 모집단은 조사가 곤란한 일부 섬 지역, 기숙 시설, 특수사회시설, 관광호텔 등에 거주하는 사람을 제외한 일반 가구의 만 15세 이상 일반 국민이다. 2023년 현재 기준 만 15세 이상의 일반 국민으로 기숙사, 특수사회시설 등의 집단시설 가구, 군복무, 타 지역 및 해외 거주 등으로 장기간 부재중인 가구원은 조사대상에서 제외되며, 구체적인 조사 제외 대상은 다음과 같다.

- 학교, 공장, 병원 등의 기숙사 거주자, 양로원, 고아원, 특수병원 등 특수사회시설 거주자, 현역병, 상근예비역, 대체복무요원 등의 군복무자, 교도소, 소년원, 치료감호소 등의 수감자, 장기 해외 출장·체류자 등

나. 표본추출틀

표본추출틀(sampling frame)은 통계청에서 제공하는 집계구를 활용한다.

다. 모집단 현황

전체 모집단 집계구는 10만 4,292개이고, 일반가구는 2,144만 8,463호이다. 경기도와 서울시는 각각 전체 집계구의 25.5%와 17.8%이고, 가구 수는 24.7%, 18.9%이다.

[표 1] 모집단 현황

구분	집계구 ¹⁾		가구 ²⁾		인구 ³⁾		
	규모 (개)	구성비 (%)	규모 (개)	구성비 (%)	전체 (명)	만 15세 이상 (명)	포함율 (%)
전국	104,292	100.0	21,448,463	100.0	51,738,071	45,650,600	88.2
서울	18,614	17.8	4,046,799	18.9	9,472,127	8,539,361	90.2
부산	6,705	6.4	1,431,365	6.7	3,324,335	2,977,058	89.6
대구	4,873	4.7	1,000,931	4.7	2,387,911	2,114,905	88.6
인천	5,916	5.7	1,183,610	5.5	2,957,044	2,600,269	87.9
광주	3,022	2.9	615,693	2.9	1,475,262	1,285,953	87.2
대전	3,010	2.9	640,085	3.0	1,479,740	1,303,254	88.1
울산	2,272	2.2	451,432	2.1	1,120,753	972,106	86.7
세종	757	0.7	145,295	0.7	366,227	294,620	80.4
경기	26,553	25.5	5,290,662	24.7	13,652,529	11,884,722	87.1
강원	3,195	3.1	674,728	3.1	1,521,890	1,357,358	89.2
충북	3,263	3.1	695,611	3.2	1,624,764	1,433,591	88.2
충남	4,466	4.3	915,634	4.3	2,175,960	1,908,289	87.7
전북	3,691	3.5	772,471	3.6	1,787,053	1,585,951	88.7
전남	4,224	4.1	777,358	3.6	1,778,124	1,575,583	88.6
경북	5,346	5.1	1,156,645	5.4	2,635,314	2,347,447	89.1
경남	6,917	6.6	1,378,982	6.4	3,305,931	2,891,856	87.5
제주	1,468	1.4	271,162	1.3	673,107	578,277	85.9

1) 집계구 : 2021년 6월 기준 통계청 자료

2) 가 구 : 2022년 7월 기준 인구총조사 자료

3) 인구 : 2022년 7월 기준 인구총조사 자료

라. 표본 층화

층화(stratification)는 표본추출틀의 단위들을 동질적 특성을 가진 집단으로 구분하는 것을 의미한다. 본 조사의 표본추출틀인 집계구는 기초단위구를 결합시켰기 때문에 국가승인통계 조사에 한해 제한적으로 사용할 수 있는 조사구(통계조사를 위해 행정 읍, 면, 동별로 일정한 기준이 되는 수 또는 면적에 근사하게 적당히 구획한 구역단위)와 유사한 특성을 지니고 있으며, 국가승인통계에서 검증된 층화 방식을 벤치마킹(benchmarking)할 수 있다.

국가승인통계의 일반적인 층화 방식은 8개 특별·광역시, 9개 도 지역으로 1차 층화하고, 9개 도 지역에 대해서는 2차 층화로 동부와 읍면부로 구분한다. 광역시 중 읍면 지역이 존재하는 대구, 인천, 울산, 세종에 대해서는 읍면부에 속한 가구 비중이 낮거나, 모집단 가구수가 작기 때문에 동부와 읍면부를 구분하여 층화하지 않는다. 마지막으로 동부의 경우에는 다시 아파트와 일반 조사구 특성으로 구분한다.

본 조사에서는 국가승인통계의 층화방식을 동일하게 활용하되, 가구와 인구수가 많은 서울과 경기 지역은 지리적 인접성과 특성을 고려하여 각각 4개 권역과 3개 권역으로 세분화한다. 또한, 조사구의 경우에는 아파트 조사구에는 아파트만이, 일반 조사구에는 일반 주택(다세대, 단독, 연립 등)만이 존재하지만, 집계구는 아파트와 일반주택이 혼재되어 있기 때문에 이를 구분하기 위해 아파트의 비율이 60% 이상인 집계구는 아파트집계구로, 60% 미만인 집계구는 일반집계구로 정의한다. 이상의 층화변수와 내용을 정리하면 다음 [표 2] ~ [표 3]과 같다.

[표 2] 서울, 경기 권역 구분

권역	해당 지역	
서울	북서부	은평구, 서대문구, 마포구, 용산구, 종로구, 중구
	북동부	노원구, 성북구, 중랑구, 도봉구, 강북구, 동대문구, 성동구, 광진구
	남서부	강서구, 양천구, 금천구, 영등포구, 구로구, 관악구, 동작구
	남동부	서초구, 강남구, 송파구, 강동구
경기	북부	고양시, 김포시, 파주시, 의정부시, 포천시, 동두천시, 양주시, 남양주시 가평군, 양평군, 연천군
	중부	부천시, 성남시, 광명시, 의왕시, 안양시, 과천시, 구리시, 군포시, 하남시 광주시, 시흥시
	남부	수원시, 오산시, 용인시, 이천시, 안성시, 안산시, 평택시, 화성시

[표 3] 모집단 층화 요약

구분	1차 층화 (지역)	2차 층화 (동·읍면)	3차 층화 (주택유형)	총수
인구총조사 집계구	서울(4), 광역시(7)	동부	아파트, 일반주택	22
	경기(3), 기타 도(8)	동부	아파트, 일반주택	9
		읍면부	-	24
합계				55

한편, 층별 모집단 집계구 및 가구 현황은 다음 [표 4] ~ [표 5]와 같다.

[표 4] 층별 모집단 집계구 현황

시도	동부		읍면부 ⁴⁾	합계	
	아파트	일반			
전국	51,927	33,787	18,578	104,292	
서울	9,234	9,380	-	18,614	
권역	북서부	1,337	1,870	-	3,207
	북동부	2,858	2,933	-	5,791
	남서부	2,675	2,955	-	5,630
	남동부	2,364	1,622	-	3,986
부산	4,021	2,684	-	6,705	
대구	2,988	1,885	-	4,873	
인천	3,499	2,417	-	5,916	
광주	2,128	894	-	3,022	
대전	1,807	1,203	-	3,010	
울산	1,440	832	-	2,272	
세종	600	157	-	757	
경기	14,851	7,401	4,301	26,553	
권역	북부	4,003	1,510	2,097	7,610
	중부	5,181	3,199	374	8,754
	남부	5,667	2,692	1,830	10,189
강원	1,211	694	1,290	3,195	
충북	1,215	695	1,353	3,263	
충남	1,404	662	2,400	4,466	
전북	1,715	917	1,059	3,691	
전남	1,061	550	2,613	4,224	
경북	1,650	1,165	2,531	5,346	
경남	2,790	1,550	2,577	6,917	
제주	313	701	454	1,468	

4) 부산, 대구, 인천, 울산, 세종은 읍면부가 존재하지만 사회생활환경 및 행정적 측면에서 동부와 유사하고 그 비율도 적기 때문에 동부에 통합

[표 5] 층별 모집단 가구 현황

시도	동부		읍면부 ⁵⁾	합계	
	아파트	일반			
전국	21,448,463	9,931,226	7,744,107	3,773,130	
서울	4,046,799	1,782,187	2,264,612	-	
권역	북서부	709,339	264,796	444,543	-
	북동부	1,242,101	558,699	683,402	-
	남서부	1,283,436	527,244	756,192	-
	남동부	811,923	431,448	380,475	-
부산	1,431,365	810,506	620,859	-	
대구	1,000,931	576,696	424,235	-	
인천	1,183,610	657,258	526,352	-	
광주	615,693	418,393	197,300	-	
대전	640,085	355,926	284,159	-	
울산	451,432	270,644	180,788	-	
세종	145,295	113,057	32,238	-	
경기	5,290,662	2,724,881	1,673,182	892,599	
권역	북부	1,512,690	738,972	345,498	428,220
	중부	1,751,640	949,621	718,935	83,084
	남부	2,026,332	1,036,288	608,749	381,295
강원	674,728	243,550	160,497	270,681	
충북	695,611	243,612	155,018	296,981	
충남	915,634	267,574	154,939	493,121	
전북	772,471	339,639	204,012	228,820	
전남	777,358	211,189	116,909	449,260	
경북	1,156,645	324,752	273,609	558,284	
경남	1,378,982	530,196	342,550	506,236	
제주	271,162	61,166	132,848	77,148	

5) 부산, 대구, 인천, 울산, 세종은 읍면부가 존재하지만 사회생활환경 및 행정적 측면에서 동부와 유사하고 그 비율도 적기 때문에 동부에 통합

마. 표본 크기

표본크기는 통계공표 수준, 작성되는 통계의 목표 오차, 가용 예산과 조사 기간 등을 고려하여 결정되어야 한다. 이러한 상황들을 종합적으로 고려하여 7,500가구를 표본크기로 한다. 가구당 적격 가구원 수를 2.0명으로 계산할 경우 총 표본가구원 수는 15,000명이며, 이 경우 목표 허용오차는 95% 신뢰수준에서 $\pm 1\%p$ 이내이다 (95% 신뢰수준에서 $\pm 0.80\%p$).

[표 6] 표본크기에 따른 표본오차(95% 신뢰수준)

표본크기(개)	1,000	2,000	3,000	5,000	8,000	10,000	15,000
표본오차(%)	3.10	2.19	1.79	1.39	1.10	0.98	0.80

바. 표본 배분

본 조사의 표본크기는 7,500가구인데, 집계구당 10가구를 선정하기 때문에 750개의 집계구가 필요하다. 집계구당 10가구를 조사하는 이유는 통계의 정도(정밀성·precision)와 조사 비용, 현장조사 업무 관리의 편의성 등을 종합적으로 고려하였다. 앞선 [표 1]에서 서울과 경기의 집계구와 가구 수는 약 17~25%의 비중을 차지하는 반면 세종, 울산에서는 집계구와 가구 수가 적어 시도 간 표본추출률·조사표본의 크기가 약 20배 정도 차이가 난다. 시도별 또는 권역별로 주요 사회적 이슈에 대한 태도가 상이하게 나타나기 때문에 시도별 또는 권역별 특성 분석을 위해 모집단 구성비를 크게 해치지 않는 범위에서 모수 추정치의 정도(precision)가 유사하도록 조정할 필요가 있다. 이러한 관점에서 1차 층인 시도에 대한 표본배분(sample allocation)은 비례배분, 제곱근 비례배분, 우선할당 후 비례배분 방법을 비교·검토한다.

① 비례배분법

$$n_h = n \times \frac{N_h}{\sum_{h=1}^H N_h}$$

N_h 는 h 층의 모집단 집계구수 ($h = 1, 2, 3, \dots, H$)

② 제곱근 비례배분법

$$n_h = n \times \frac{\sqrt{N_h}}{\sum_{h=1}^H \sqrt{N_h}}$$

③ 우선할당 후 비례배분

$$n_h = n' + (n - n' \times H) \times \frac{N_h}{\sum_{h=1}^H N_h}$$

[표 기] 표본배분 결과 비교

시도	비례배분			제공근 비례배분			우선할당 후 비례배분			
	집계구	가구	기대 표본 오차	집계구	가구	기대 표본 오차	집계구	가구	기대 표본 오차	
전국	750	7,500	1.2	750	7,500	1.2	750	7,500	1.2	
서울	134	1,340	2.7	84	840	3.4	108	1,084	3.0	
부산	48	480	4.6	51	510	4.4	47	471	4.6	
대구	35	350	5.3	43	430	4.8	38	376	5.2	
인천	43	430	4.8	48	480	4.6	43	430	4.8	
광주	22	220	6.7	34	340	5.4	28	281	6.0	
대전	22	220	6.7	34	340	5.4	28	280	6.0	
울산	16	160	7.9	29	290	5.9	24	242	6.4	
세종	5	50	14.1	17	170	7.7	16	164	7.8	
경기	191	1,910	2.3	101	1,010	3.1	149	1,493	2.6	
강원	23	230	6.6	35	350	5.3	29	290	5.9	
충북	23	230	6.6	35	350	5.3	29	293	5.8	
충남	32	320	5.6	41	410	4.9	36	355	5.3	
전북	27	270	6.1	38	380	5.1	32	315	5.6	
전남	30	300	5.8	40	400	5.0	34	343	5.4	
경북	38	380	5.1	45	450	4.7	40	401	5.0	
경남	50	500	4.5	51	510	4.4	48	481	4.6	
제주	11	110	9.5	24	240	6.5	20	201	7.1	
요약	최소	5	50	2.3	17	170	3.1	16	164	2.6
	최대	191	1,910	14.1	101	1,010	7.7	149	1,493	7.8
	평균	44	441	6.2	44	441	5.1	44	441	5.4

[표 7]의 결과를 보면, 비례배분법은 층별 구성비는 모집단과 유사하나 시도별 기대 표본오차(sampling error)가 크고, 우선할당 후 비례배분법은 세종, 제주 등 표본 규모가 작은 시도의 기대 표본오차를 감소시키는 효과가 적어 해당 시도의 조사 결과 해석에 어려움이 따를 수 있다. 마지막으로 제곱근 비례배분법은 층별 표본 크기의 격차를 줄이면서 시도별 대표성을 충족시킬 수 있다. 이러한 점을 고려하여 제곱근 비례배분법을 최종적으로 채택했다. 한편, 서울·경기의 권역, 동부/읍면부, 동부 내 아파트/일반주택 층은 비례배분법을 적용하며, 최종 표본배분 결과는 다음 [표 8]과 같다.

[표 8] 표본배분 결과

시도	집계구				가구				
	동부		읍면부	합계	동부		읍면부	합계	
	아파트	일반			아파트	일반			
전국	365	233	150	750	3,660	2,340	1,500	7,500	
서울	42	42	-	84	420	420	-	840	
권역	북서부	6	9	-	15	60	90	-	150
	북동부	13	13	-	26	130	130	-	260
	남서부	12	13	-	25	120	130	-	250
	남동부	11	7	-	18	110	70	-	180
부산	31	20	-	51	310	200	-	510	
대구	26	17	-	43	260	170	-	430	
인천	28	20	-	48	280	200	-	480	
광주	24	10	-	34	240	100	-	340	
대전	20	14	-	34	200	140	-	340	
울산	18	11	-	29	180	110	-	290	
세종	13	4	-	17	130	40	-	170	
경기	57	28	16	101	570	280	160	1,010	
권역	북부	15	6	8	29	150	60	80	290
	중부	20	12	1	33	200	120	10	330
	남부	22	10	7	39	220	100	70	390
강원	13	8	14	35	130	80	140	350	
충북	13	7	15	35	130	70	150	350	
충남	13	6	22	41	130	60	220	410	
전북	18	9	11	38	180	90	110	380	
전남	10	5	25	40	100	50	250	400	
경북	14	10	21	45	140	100	210	450	
경남	21	11	19	51	210	110	190	510	
제주	5	12	7	24	50	120	70	240	

사. 표본 추출

1차 층으로 시도(서울 4개 권역, 경기 3개 권역)와 동/읍면부를 결합하여 33개 층을 구분하고, 2차 층으로 아파트/일반 주택 특성을 결합하여 구성한 최종 55개 층에 대해 배분된 집계구를 추출한다(표 8) 참고). 추출방식은 각 집락의 크기에 비례하여 표본을 추출하는 확률비례계통추출법(PPS: probability proportional to size systematic sampling)을 적용했으며, 여기서 집락은 집계구를, 집락의 크기는 각 집계구의 가구 수를 의미한다. 또한 내재적 층화를 위해, 층 내 세부 지역(동읍면 코드)을 기준으로 정렬하였다. 집계구 내에서 가구를 체계적으로 추출하고, 가구 내에서 만 15세 이상 가구원을 전수 조사하여 응답 가구원의 특성이 편향되지 않도록 한다.

아. 요일별 할당

2020년~2022년 조사 결과 행복감, 정서, 감정, 일상 재구성 등의 항목에서 측정 기준 요일이 ‘월요일~목요일’ 또는 ‘금요일~일요일’인지에 따라 응답 값의 유의미한 차이가 발견되어, 본 조사 시 요일별 할당을 적용하여 ‘월요일~목요일’에 대한 기준 응답이 약 57%, ‘금요일~일요일’에 대한 기준 응답이 약 43% 정도 수집될 수 있도록 고려하였다.

2 가중치 산출 및 추정

가. 설계가중치

가중치는 표본추출에 따른 추출률 차이와 응답률 및 모집단에 대한 정보를 통해 모집단의 구조와 표본 구조를 일치시켜 추정의 정확도를 높이기 위한 것이다. 이 조사에서 가중치는 설계 가중치와 무응답 조정, 가구원수 보정, 가구원 추출률 보정, 최신 모집단 정보인 행정안전부 주민등록 인구현황 자료를 이용한 사후층화조정의 단계를 거쳐서 산출한다.

각 표본 집계구에 대한 표본 추출틀의 역수와 표본 집계구에서 가구조사 착수율의 역수를 곱하여 다음과 같이 산출하였다.

$$\text{설계가중치} = \frac{S_h}{n_h S_{hi}} \times \frac{M_{hi}}{m_{hi}}$$

- 이때 원칙적으로는 각 층에서 표본 가구들은 모두 동일한 설계가중값을 갖지만, 실제 조사과정에서 표본 추출틀의 집계구 내 가구 수와 실제 가구 수에 차이가 있기 때문에 설계가중값이 동일하지 않을 수 있지만 비슷한 값을 갖게 됨

나. 무응답 조정계수

해당 지역 내 세부 층에서 조사 가구의 설계가중값은 원칙적으로 같지만, 실제 조사에서 10가구를 정확히 지키지 않은 집계구가 있음에 따라 이에 대한 조정을 다음과 같이 하였다.

$$\text{응답률 조정계수} = \frac{m_{hi}}{r_{hi}}$$

다. 벤치마킹 조정계수

조사대상 연도의 모집단을 반영하기 위해 2023년도 주민등록 인구통계의 시도별 (17)*성(2)*연령대(6)별 만 15세 인구 현황 자료와 함께 인구센서스의 1인 가구 비율을 보정하여 레이킹 비 방법(Raking ratio method) 방식으로 가중치를 산출하였다.

$$BF_g = \frac{NPS_g}{\sum_{h=1} \sum_{i=1} W_{hi}^*}$$

라. 최종가중치

최종가중치 = 설계가중치 × 무응답 조정계수 × 벤치마킹 조정계수

※ 사용된 기호

- n_h : 층 h 의 표본 집계구 수
- S_{hi} : 층 h 의 i 번째 집계구에 대한 크기의 측도(해당 집계구의 총 가구 수)
- $S_h = \sum_{i=1}^{N_h} S_{hi}$: 층 h 에서 크기의 측도에 대한 총합
- M_{hi} : 층 h 의 i 번째 집계구 내 가구 수
- m_{hi} : 층 h 의 i 번째 집계구 내 조사착수 가구 수
- r_{hi} : 층 h 의 i 번째 표본집계구 내 조사완료 응답 가구 수

마. 모수 추정

본 조사의 표본설계는 2단계층화집락추출법(two-stage stratified cluster sampling)'을 적용하여 일종의 복합표본설계 기법에 해당되므로 이에 적합한 모수추정식을 적용한다.

○ 모평균 추정

$$\hat{Y} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{K_{hij}} w_{hijk} y_{hijk}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{K_{hij}} w_{hijk}}$$

여기서 h 는 층, i 는 집계구, j 는 가구, k 는 가구원을 나타낸다.

○ 모평균에 대한 분산추정

$$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi..} - \bar{e}_{h...})^2$$

$$- f_h = \frac{N_h}{n_h}, \quad - e_{hi..} = \left[\sum_{j=1}^{m_{hi}} \sum_{k=1}^{K_{hj}} w_{hijk} (y_{hijk} - \widehat{Y}) \right] / w_{...}$$

$$- \bar{e}_{h...} = \left(\sum_{i=1}^{n_h} e_{hi..} \right) / n_h$$

○ 모평균에 대한 표준오차

$$\widehat{SE}(\widehat{Y}) = \sqrt{\widehat{V}(\widehat{Y})}$$

○ 모평균에 대한 변동계수(상대표준오차) 추정

$$\widehat{RSE}(\widehat{Y}) = \frac{\sqrt{\widehat{V}(\widehat{Y})}}{\widehat{Y}} \times 100(\%)$$