



2022 한국인의 행복조사
(2차조사)
표본설계내역서



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, 관광 호텔 등에 거주하는 사람을 제외한 일반 가구의 만 15세 이상 일반 국민이다. 2022년 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 4,107개이고, 일반가구는 2,092만 6,710호이다. 경기도와 서울시는 각각 전체 집계구의 25.2%와 18.0%이고, 가구 수는 24.4%, 19.0%이다.

[표 1] 모집단 현황

구분	집계구 ¹⁾		가구 ²⁾		인구 ³⁾		
	규모 (개)	구성비 (%)	규모 (개)	구성비 (%)	전체	만 15세 이상	포함율 (%)
전국	104,107	100.0	20,926,710	100.0	51,829,136	45,574,979	87.9
서울	18,718	18.0	3,982,290	19.0	9,586,195	8,611,078	89.8
부산	6,742	6.5	1,405,037	6.7	3,349,016	2,992,289	89.3
대구	4,916	4.7	985,816	4.7	2,410,700	2,128,002	88.3
인천	5,856	5.6	1,147,200	5.5	2,945,454	2,581,393	87.6
광주	3,013	2.9	599,217	2.9	1,477,573	1,282,819	86.8
대전	3,014	2.9	631,208	3.0	1,488,435	1,304,662	87.7
울산	2,283	2.2	444,087	2.1	1,135,423	981,130	86.4
세종	727	0.7	139,106	0.7	353,933	284,093	80.3
경기	26,223	25.2	5,098,431	24.4	13,511,676	11,716,926	86.7
강원	3,177	3.1	661,039	3.2	1,521,763	1,353,349	88.9
충북	3,267	3.1	678,922	3.2	1,632,088	1,436,166	88.0
충남	4,468	4.3	892,222	4.3	2,176,636	1,901,808	87.4
전북	3,716	3.6	755,575	3.6	1,802,766	1,593,568	88.4
전남	4,210	4.0	761,518	3.6	1,788,807	1,580,047	88.3
경북	5,361	5.1	1,131,819	5.4	2,644,757	2,348,066	88.8
경남	6,958	6.7	1,350,155	6.5	3,333,056	2,904,701	87.1
제주	1,458	1.4	263,068	1.3	670,858	574,882	85.7

1) 집계구 : 2020년 6월 기준 통계청 자료

2) 가 구 : 2021년 7월 기준 인구총조사 자료

3) 인 구 : 2021년 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 특·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 2개 권역으로 세분화한다. 또한, 조사구의 경우에는 아파트 조사구에는 아파트만이, 일반 조사구에는 일반 주택(다세대, 단독, 연립 등)만이 존재하지만, 집계구는 아파트와 일반 주택이 혼용되어 있기 때문에 이를 구분하기 위해 아파트의 비율이 60% 이상인 경우에는 아파트 집계구로, 60% 미만인 경우에는 일반 집계구로 정의한다. 이상의 층화변수와 내용을 정리하면 다음 [표 2] ~ [표 3]과 같다.

[표 2] 서울, 경기의 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	중·북부	부천시, 성남시, 광명시, 의왕시, 안양시, 과천시, 구리시, 군포시, 하남시, 광주시, 여주시, 시흥시, 고양시, 김포시, 파주시, 의정부시, 포천시, 동두천시, 양주시, 남양주시, 가평군, 양평군, 연천군
	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, 일반	22
	경기(2), 기타 도(8)	동부	아파트, 일반	20
		읍면부	-	10
합계				52

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

[표 4] 층별 모집단 집계구 현황

시도	동부		읍면부	합계	
	아파트	일반			
전국	51,261	34,184	18,662	104,107	
서울	9,162	9,556	-	18,718	
권역	북서부	1,299	1,921	-	3,220
	북동부	2,889	2,949	-	5,838
	남서부	2,636	3,024	-	5,660
	남동부	2,338	1,662	-	4,000
부산	3,981	2,761	-	6,742	
대구	2,985	1,931	-	4,916	
인천	3,422	2,434	-	5,856	
광주	2,109	904	-	3,013	
대전	1,815	1,199	-	3,014	
울산	1,442	841	-	2,283	
세종	574	153	-	727	
경기	14,569	7,373	4,281	26,223	
권역	중북부	9,022	4,722	2,447	16,191
	남부	5,547	2,651	1,834	10,032
강원	1,174	713	1,290	3,177	
충북	1,188	710	1,369	3,267	
충남	1,387	661	2,420	4,468	
전북	1,705	923	1,088	3,716	
전남	1,040	554	2,616	4,210	
경북	1,640	1,174	2,547	5,361	
경남	2,757	1,597	2,604	6,958	
제주	311	700	447	1,458	

[표 5] 층별 모집단 가구 현황

시도	동부		읍면부	합계	
	아파트	일반			
전국	9,594,043	7,638,181	3,694,486	20,926,710	
서울	1,727,489	2,254,801	-	3,982,290	
권역	북서부	252,332	446,313	-	698,645
	북동부	552,010	676,993	-	1,229,003
	남서부	505,089	751,683	-	1,256,772
	남동부	418,058	379,812	-	797,870
부산	784,431	620,606	-	1,405,037	
대구	564,070	421,746	-	985,816	
인천	632,133	515,067	-	1,147,200	
광주	405,552	193,665	-	599,217	
대전	351,896	279,312	-	631,208	
울산	267,198	176,889	-	444,087	
세종	105,853	33,253	-	139,106	
경기	2,616,252	1,619,415	862,764	5,098,431	
권역	중북부	1,626,699	1,032,415	493,719	3,152,833
	남부	989,553	587,000	369,045	1,945,598
강원	233,169	161,281	266,589	661,039	
충북	232,735	155,725	290,462	678,922	
충남	257,305	153,387	481,530	892,222	
전북	329,776	198,110	227,689	755,575	
전남	203,494	116,358	441,666	761,518	
경북	313,919	268,712	549,188	1,131,819	
경남	510,467	340,142	499,546	1,350,155	
제주	58,304	129,712	75,052	263,068	

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 2.2명으로 계산할 경우 총 표본가구원 수는 16,500명이며, 이 경우 목표 허용오차는 95% 신뢰수준에서 $\pm 1\%p$ 이내이다(95% 신뢰수준에서 $\pm 0.76\%p$).

[표 6] 표본 크기별 표본오차(95% 신뢰수준)

표본크기(개)	1,000	2,000	3,000	5,000	8,000	10,000	15,000	16,500
표본오차(%)	3.10	2.19	1.79	1.39	1.10	0.98	0.80	0.76

바. 표본 배분

본 조사의 표본크기는 7,500가구인데, 집계구당 10가구를 선정하기 때문에 750개 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 가구수는 약 18~25%의 비중을 차지하는 반면 세종, 울산에서는 집계구와 가구수가 적어 시도 간의 표본추출률·조사표본의 크기가 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 7] 배분방법에 따른 표본배분 결과

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	135	1,350	2.7	85	850	3.4	125	1,250	2.8	
부산	49	490	4.5	51	510	4.4	48	480	4.6	
대구	35	350	5.3	43	430	4.8	36	360	5.3	
인천	42	420	4.9	47	470	4.6	42	420	4.9	
광주	22	220	6.7	34	340	5.4	24	240	6.5	
대전	22	220	6.7	34	340	5.4	24	240	6.5	
울산	16	160	7.9	29	290	5.9	20	200	7.1	
세종	5	50	14.1	17	170	7.7	10	100	10.0	
경기	189	1,890	2.3	100	1,000	3.2	173	1,730	2.4	
강원	23	230	6.6	35	350	5.3	25	250	6.3	
충북	24	240	6.5	35	350	5.3	26	260	6.2	
충남	32	320	5.6	41	410	4.9	34	340	5.4	
전북	27	270	6.1	38	380	5.1	29	290	5.9	
전남	30	300	5.8	40	400	5.0	32	320	5.6	
경북	39	390	5.1	45	450	4.7	39	390	5.1	
경남	50	500	4.5	52	520	4.4	49	490	4.5	
제주	10	100	10.0	24	240	6.5	14	140	8.4	
요약	최소	5	50	2.3	17	170	3.2	10	100	2.4
	최대	189	1,890	14.1	100	1,000	7.7	173	1,730	10.0
	평균	44	441	6.0	44	441	5.0	44	441	6.0

[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 시도

의 조사 결과 해석에 어려움이 따를 수 있다. 마지막으로 제공된 비례배분법은 층별 표본 크기의 격차를 줄이면서 시도별 대표성을 충족시킬 수 있다. 이러한 점을 고려하여 제공된 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 최종 결과

시도	집계구				가구				
	동부		읍면부	합계	동부		읍면부	합계	
	아파트	일반			아파트	일반			
전국	362	238	150	750	3,620	2,380	1,500	7,500	
서울	42	43	-	85	420	430	-	850	
권역	북서부	6	9	-	15	60	90	-	150
	북동부	13	14	-	27	130	140	-	270
	남서부	12	13	-	25	120	130	-	250
	남동부	11	7	-	18	110	70	-	180
부산	30	21	-	51	300	210	-	510	
대구	26	17	-	43	260	170	-	430	
인천	27	20	-	47	270	200	-	470	
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	56	28	16	100	560	280	160	1,000	
권역	중북부	35	18	9	62	350	180	90	620
	남부	21	10	7	38	210	100	70	380
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	17	10	11	38	170	100	110	380	
전남	10	5	25	40	100	50	250	400	
경북	14	10	21	45	140	100	210	450	
경남	21	12	19	52	210	120	190	520	
제주	5	12	7	24	50	120	70	240	

사. 표본 추출

1차 층으로 시도(서울 4개 권역, 경기 2개 권역)와 동/읍면부를 결합하여 31개 층을 구분하고, 2차 층으로 아파트/일반 주택 특성을 결합하여 구성된 최종 52개 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. 모수 추정

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 사후층화보정의 단계를 거쳐서 산출한다.

- 설계 가중치 : 집계구 추출률의 역수에 집계구 내 가구 추출률의 역수 곱함

$$W^{(1)_{hj}} = \frac{\sum_{i=1}^{N_h} M_{hi}}{n_h M_{hi}} \times \frac{M_{hi}}{m_{hi}}$$

- w_{hij} : h 층, i 집계구, j 번째 가구의 설계 가중치
- M_{hi} : h 층, i 집계구의 모집단 가구수
- N_h : h 층의 모집단 집계구수
- n_h : h 층에 할당된 표본 집계구수
- M_{hi} : h 층, i 집계구 내 실제 가구수
- m_{hi} : h 층, i 집계구 내 표본 가구수

- 무응답보정 가중치 : 무응답 보정은 집계구 단위로 실시하며 응답자와 무응답자의 특성이 유사하다는 가정 하에 응답률의 역수를 무응답 보정 가중치로 계산함

$$w^{(2)_{hi}} = \frac{m_{hi}}{r_{hi}}$$

- w_{hi} : h 층, i 집계구의 무응답 보정 가중치
- m_{hi} : h 층, i 집계구 내 표본 가구수
- r_{hi} : h 층, i 집계구 내 조사완료 가구수

- 가구 가중치 : 설계 가중치와 무응답 가중치를 곱함

$$w^{(3)_{hij}} = w^{(1)_{hij}} \times w^{(2)_{hi}}$$

- $w^{(3)_{hij}}$: h 층, i 집계구 내 j 번째 가구의 가중치

- 가구원 가중치 : 가구 가중치와 가구 내 적격가구원 추출률의 역수를 곱함

$$w_{hijk} = w_{hij} \times \left(\frac{1}{e_{hij}}\right)^{-1}$$

- w_{hijk} : h 층, i 집계구, j 가구, k 가구원의 가중치
- e_{hij} : h 층, i 집계구, j 가구의 적격 가구원 수

- 사후층화 가중치 : 가중치를 적용한 표본 자료의 분포가 모집단의 특성과 일치하도록 보정하기 위해 행정안전부 주민등록 인구통계(시도(17), 연령(2), 연령대(5)) 활용

$$W_{hijk}^f = W_{hijk}^f \times \frac{N_{cde}}{\sum_c \sum_d \sum_e W_{hijk,cde}}$$

- W_{hijk}^f : h 층, i 집계구, j 가구, k 가구원의 사후층화 가중치
- N_{cde} : 사후층의 모집단 인구수 (최신 주민등록 인구통계)
- $W_{hijk,cde}$: 사후층의 가구원 가중치 (W_{hijk})

- 표준화 가중치 : 최종 가중치를 조사완료된 표본크기로 표준화하여 분석 용이성 및 이용자 편의 제고함

$$W_{hijk}^s = W_{hijk}^f \times \frac{n}{\sum W_{hijk,cde}^f}$$

- W_{hijk}^s : h 층, i 집계구, j 가구, k 가구원의 표준화 가중치
- n : 전체 조사완료된 표본크기
- $\sum W_{hijk,cde}^f$: 최종 가중치 전체 합계

② 모수추정식

- 평균 또는 비율 추정

$$\widehat{Y} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^m \sum_{k=1}^K w_{hijk} y_{hijk}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^m \sum_{k=1}^K w_{hijk}}$$

- 모평균에 대한 분산추정

$$\begin{aligned} \widehat{V}(\widehat{Y}) &= \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi\cdot} - \bar{e}_{h\cdot})^2 \\ - f_h &= \frac{N_h}{N} \\ - e_{hi\cdot} &= \left[\sum_{j=1}^m w_{hij} (y_{hij} - \widehat{Y}) \right] / w_{h\cdot\cdot} \\ - \bar{e}_{h\cdot} &= \left(\sum_{i=1}^{n_h} e_{hi\cdot} \right) / n_h \end{aligned}$$

- 표준오차 추정

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

- 95% 신뢰수준에서 오차한계

$$\pm 1.96 \times \widehat{SE}(\widehat{Y})$$

- 상대표준오차 추정

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