< 외부 공개용 >

2024년(7패널) 가계동향조사 표본설계 보고서

2023. 10.

통계청
조사관리국
표 본 과

【 목 차 】

Ⅰ. 표본개편 개요

1 조사 개요 ····· 1

2 기존 표본설계 개요 ····· 1

3 이번 표본설계 방향 ····· 2

Ⅱ. 모집단 및 표본추출틀

1 모집단 ····· 3

2 표본추출틀 ····· 3

Ⅲ. 층화 · 균형변수

1 분석 배경 ····· 7

2 층화변수 ····· 7

3 균형변수 ····· 7

Ⅳ. 표본규모

1 전체 표본규모 ····· 11

2 설계층별 표본조사구 배분 ····· 12

3 표본조사구별 최대 · 예비표본규모 산정 ····· 13

Ⅴ. 표본추출

1 표본조사구 추출 ····· 15

2 표본가구 선정 ····· 15

Ⅵ. 가중값 · 추정

1 가중값 ····· 16

2 추정 ····· 18

Ⅰ 표본설계 개요

1 조사 개요

○ **(목적)** 가구별 수입과 지출의 실태를 조사하여, 소득과 소비 수준의 측정 및 분석에 필요한 기초자료를 제공

○ **(대상)** 전국의 모든 1인 이상 일반가구(농림어가 포함)

○ **(주기)** 매월

○ **(항목)** 조사대상 월 한 달 동안의 소득·지출 및 가구·가구원에 관한 항목

○ **(공표)** 분기 통계는 매 분기 말 익익월(5·8·11·2월), 연간 지출은 익년 2월 공표

2 기존 표본설계 개요

○ **(표본체계)** 전용표본, 6-6-6 연동표본체계

- 표본 가구는 진입 후 6개월 동안 조사에 참여한 후 6개월의 휴식기를 가지며, 그다음 6개월 동안 다시 조사에 참여한 뒤 최종 종료

- 매월 전체 표본의 1/12이 진입-종료, 1/12이 재진입-휴식으로 변동되어, 전월 대비 83.3%, 전분기 대비 62.5%, 전년 대비 58.6%의 표본이 유지

○ **(설계층)** 27개 지역층

- 특·광역시는 시 단위(7개), 세종을 포함한 도지역은 동/읍면 구분(20개)

○ **(표본규모)** 총 900개 조사구, 조사구 당 8가구 → 총 7,200가구 목표

- 안정적인 응답 표본 확보를 위해 과대추출(원표본 8가구 + 예비표본 6~12가구) 적용

○ **(추출방법)** 층화2단집락추출

- 일단계(PSU=조사구) : 설계층별 조사구를 균형표본추출(큐브 방법)

- 이단계(SSU=가구) : 표본조사구 내 시작가구를 단순임의추출

< 6-6-6 연동표본체계 >

	패널	p_i												p_{i+1}											
	그룹	1		2		3		4		5		6		4		5		6		1		2		3	
	구역	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
23년	1월	a5							h5	i5	j5	k5	l5	g6							b4	c4	d4	e4	f4
	2월	a5	b5							i5	j5	k5	l5	g6	h6							c4	d4	e4	f4
	3월	a5	b5	c5							j5	k5	l5	g6	h6	i6							d4	e4	f4
	4월	a5	b5	c5	d5							k5	l5	g6	h6	i6	j6							e4	f4
	5월	a5	b5	c5	d5	e5							l5	g6	h6	i6	j6	k6							f4
	6월	a5	b5	c5	d5	e5	f5							g6	h6	i6	j6	k6	l6						
	7월		b5	c5	d5	e5	f5	g5							h6	i6	j6	k6	l6	a6					
	8월			c5	d5	e5	f5	g5	h5							i6	j6	k6	l6	a6	b6				
	9월				d5	e5	f5	g5	h5	i5							j6	k6	l6	a6	b6	c6			
	10월					e5	f5	g5	h5	i5	j5							k6	l6	a6	b6	c6	d6		
	11월						f5	g5	h5	i5	j5	k5							l6	a6	b6	c6	d6	e6	
	12월							g5	h5	i5	j5	k5	l5							a6	b6	c6	d6	e6	f6
24년	1월	a7							h5	i5	j5	k5	l5	g6							b6	c6	d6	e6	f6
	2월	a7	b7							i5	j5	k5	l5	g6	h6							c6	d6	e6	f6
	3월	a7	b7	c7							j5	k5	l5	g6	h6	i6							d6	e6	f6
	4월	a7	b7	c7	d7							k5	l5	g6	h6	i6	j6							e6	f6
	5월	a7	b7	c7	d7	e7							l5	g6	h6	i6	j6	k6							f6
	6월	a7	b7	c7	d7	e7	f7							g6	h6	i6	j6	k6	l6						
	7월		b7	c7	d7	e7	f7	g7							h6	i6	j6	k6	l6	a6					
	8월			c7	d7	e7	f7	g7	h7							i6	j6	k6	l6	a6	b6				
	9월				d7	e7	f7	g7	h7	i7							j6	k6	l6	a6	b6	c6			
	10월					e7	f7	g7	h7	i7	j7							k6	l6	a6	b6	c6	d6		
	11월						f7	g7	h7	i7	j7	k7							l6	a6	b6	c6	d6	e6	
	12월							g7	h7	i7	j7	k7	l7							a6	b6	c6	d6	e6	f6
25년	1월	a7							h7	i7	j7	k7	l7	g8							b6	c6	d6	e6	f6
	2월	a7	b7							i7	j7	k7	l7	g8	h8							c6	d6	e6	f6
	3월	a7	b7	c7							j7	k7	l7	g8	h8	i8							d6	e6	f6
	4월	a7	b7	c7	d7							k7	l7	g8	h8	i8	j8							e6	f6
	5월	a7	b7	c7	d7	e7							l7	g8	h8	i8	j8	k8							f6
	6월	a7	b7	c7	d7	e7	f7							g8	h8	i8	j8	k8	l8						
	7월		b7	c7	d7	e7	f7	g7							h8	i8	j8	k8	l8	a8					
	8월			c7	d7	e7	f7	g7	h7							i8	j8	k8	l8	a8	b8				
	9월				d7	e7	f7	g7	h7	i7							j8	k8	l8	a8	b8	c8			
	10월					e7	f7	g7	h7	i7	j7							k8	l8	a8	b8	c8	d8		
	11월						f7	g7	h7	i7	j7	k7							l8	a8	b8	c8	d8	e8	
	12월							g7	h7	i7	j7	k7	l7							a8	b8	c8	d8	e8	f8

* 패널 $i=1,3,5,\cdots$ 신규 진입 표본 6개월 휴식 후 재진입 표본

3 이번 표본설계 방향

○ **(표본추출틀)** ‘21년 기준 인구총조사(등록센서스)을 기초로 구축된 최신 표본추출틀을 활용하여 표본 대표성을 제고

○ **(균형변수)** 인구총조사, 통계등록부 등을 이용해 구축한 보조변수를 조사자료에 연계한 후, 소득·지출과의 관계에 대한 분석 결과를 바탕으로 균형변수 선정

○ **(표본규모)** 모집단 분포와 응답 현황을 고려하여 조사구·가구 규모 검토

Ⅱ 모집단 및 표본추출틀

1 모집단

① 목표모집단

○ 조사 기간 내에 대한민국에 거주하고 있는 모든 일반 가구

② 조사모집단

○ ‘21년 11월 1일 기준 인구총조사(등록센서스)로 파악된, 국내 거주하는 보통·섬·아파트조사구(1, 2, A) 내 모든 가구

- 목표모집단 범위가 일반 가구인 점을 고려하여, 기숙시설·특수사회시설조사구(병영, 병원, 교도소 등) 내 집단 가구는 조사모집단에서 제외

< 조사구특성별 조사모집단 현황 >

구분	2020년			2021년				
	조사구 (개)	가구 (천개)	인구 (천명)	조사구 (개)	가구 (천개)	증감 (천개)	증감률 (%)	인구 (천명)
전체	384,686	21,485	51,829	389,865	22,023	538	2.5	51,738
조사모집단	384,669	21,485	51,070	389,848	22,023	538	2.5	51,091
보통(1)	187,899	10,531	21,855	189,173	10,700	169	1.6	21,599
섬(2)	913	37	75	912	37	0	-0.1	72
아파트(A)	195,857	10,917	29,140	199,763	11,285	369	3.4	29,419
기숙·특수(9)	17	-	759	17	-	-	-	647

2 표본추출틀

① 단위

○ **(구축 단위)** 1단계 추출단위(psu)인 조사구를 기초단위로 구축

○ **(포함 대상)** 조사모집단 중 비연계조사구를 제외한 보통·아파트(1, A)조사구

- 조사모집단 중 0.2%를 차지하는 섬조사구는 조사단위에 접촉하는데 소요되는 시간과 비용이 과다한 점을 고려하여 제외

- 등록센서스상 인구·가구 정보는 있으나 공간 정보는 없는 비연계조사구(0.4%) 제외

- 최종 표본추출틀에 포함된 가구는 전체 조사모집단 대비 약 99.4%로 안정적이나, 섬 조사구가 다수 존재하는 전남은 포함률이 96.7%로 다소 낮은 수준

< 시도별 조사모집단 및 표본추출틀 현황 >

(단위: 개, %)

시도	조사모집단		표본추출틀		
	조사구수	가구수	조사구수	가구수	가구 포함률
전국	389,848	22,022,753	388,604	21,895,478	99.4
서울	72,345	4,191,172	72,334	4,178,144	99.7
부산	26,552	1,447,194	26,543	1,442,238	99.7
대구	17,173	1,013,488	17,166	1,011,013	99.8
인천	21,002	1,216,719	20,868	1,206,001	99.1
광주	10,600	625,511	10,596	624,069	99.8
대전	10,958	648,593	10,891	646,492	99.7
울산	8,341	460,289	8,313	458,242	99.6
세종	2,669	147,347	2,665	146,883	99.7
경기	93,952	5,492,782	93,870	5,464,913	99.5
강원	12,959	682,635	12,941	679,642	99.6
충북	12,841	713,434	12,837	710,084	99.5
충남	17,009	947,235	16,967	940,786	99.3
전북	14,409	785,218	14,366	781,761	99.6
전남	15,243	788,992	14,683	762,765	96.7
경북	22,166	1,178,040	22,135	1,171,633	99.5
경남	26,214	1,405,646	26,053	1,396,134	99.3
제주	5,415	278,458	5,376	274,678	98.6

② 보조변수

○ **(인총 변수)** '21년 가구 부문 표본추출틀의 인구·가구·거처 관련 변수* 활용

* '21년 등록센서스 전수조사와 '20년 인구총조사 표본조사(20%)를 이용해 구축

○ **(자체 변수)** 통계등록부 및 행정자료를 연계·가공하여 소득·지출과 관련성이 높은 조사구 특성 변수를 별도 구축

- 평균등록경상소득: 취업활동통계등록부를 보완*하여 구축한 등록고용소득**과 행정자료 기반 공적연금소득***을 가구 → 조사구 단위로 집계하여 구축

* 일자리별 소득 결측에 대해 위계적 선형모형, 일반 선형모형 등을 이용하여 추정·대체

** 통계등록부와 행정자료를 통해 파악된 일자리별 근로소득 및 사업소득

*** 기초노령연금, 국민연금, 직역연금(공무원·군인·사학·별정우체국)의 개인별 연금수령액

- 중위공시지가: 관측값이 '0'인 조사구에 대해, 지역과 거처종류를 대체군으로 한 계층적 회귀대체를 실시하여 보완

③ 주요 보조변수 분포 개관

○ **(가구원수)** 시도-동부가 시도-읍면부에 비해 3인·4인 이상 가구 비중이 높은 편으로, 그 차이는 세종에서 특히 크게 나타남

- 특광역시 중 인천과 울산은 1인 가구의 비중이 상대적으로 낮은 편

< 설계층별 가구원수 분포 >

(단위: 천 개, %)

설계층		가구 수					가구 구성비			
		전체	1인	2인	3인	4인 이상	1인	2인	3인	4인 이상
전국		21,895	7,386	6,195	4,223	4,091	33.7	28.3	19.3	18.7
서울		4,178	1,563	1,089	792	734	37.4	26.1	19.0	17.6
부산		1,442	493	427	278	244	34.2	29.6	19.3	16.9
대구		1,011	333	285	202	191	32.9	28.2	20.0	18.9
인천		1,206	364	339	255	247	30.2	28.1	21.2	20.5
광주		624	217	170	115	122	34.7	27.2	18.5	19.6
대전		646	245	172	116	115	37.9	26.5	17.9	17.7
울산		458	136	130	99	93	29.7	28.4	21.5	20.4
세종	동부	101	25	23	23	29	25.0	22.9	23.1	29.0
	읍면부	46	21	13	6	6	46.2	27.7	13.6	12.5
경기	동부	4,530	1,320	1,192	984	1,035	29.1	26.3	21.7	22.8
	읍면부	935	298	283	175	179	31.9	30.3	18.7	19.1
강원	동부	408	149	120	73	66	36.4	29.5	17.9	16.1
	읍면부	272	99	95	43	35	36.5	34.8	15.8	13.0
충북	동부	406	144	112	75	74	35.6	27.7	18.4	18.3
	읍면부	304	115	99	48	43	37.7	32.6	15.7	14.0
충남	동부	434	154	112	81	86	35.6	25.8	18.7	19.9
	읍면부	507	185	170	80	72	36.4	33.5	15.8	14.3
전북	동부	552	197	156	99	100	35.7	28.3	18.0	18.1
	읍면부	230	85	86	33	27	36.9	37.2	14.3	11.6
전남	동부	330	110	100	61	59	33.3	30.2	18.6	18.0
	읍면부	433	159	152	66	56	36.7	35.2	15.1	12.9
경북	동부	607	214	179	113	102	35.2	29.4	18.5	16.8
	읍면부	564	210	199	87	68	37.2	35.2	15.5	12.0
경남	동부	886	285	247	180	173	32.2	27.9	20.3	19.5
	읍면부	510	174	169	88	79	34.1	33.2	17.2	15.5
제주	동부	197	65	53	37	42	33.2	26.8	18.7	21.3
	읍면부	77	25	25	13	14	32.9	32.2	17.4	17.5

○ **(거처종류)** 아파트 비중은 특·광역시 및 시도-동부가 시도-읍면부에 비해 높은 편이며, 특히 세종에서 동부·읍면부 간 차이가 큰 것으로 나타남

- 서울·인천·경기·제주에는 다세대주택이 많아 '그 외'의 비중이 다른 지역에 비해 높은 편이며, 세종-동부·대구·부산은 대형아파트 비중이 높은 편임

< 설계층별 거처종류 분포 >

(단위: 천 개, %)

설계층		가구 수						가구 구성비				
		전체	단독	소형 APT	중형 APT	대형 APT	그 외	단독	소형 APT	중형 APT	대형 APT	그 외
전국		21,895	6,631	4,457	5,578	1,248	3,982	30.3	20.4	25.5	5.7	18.2
서울		4,178	1,108	717	784	290	1,279	26.5	17.2	18.8	6.9	30.6
부산		1,442	333	304	418	114	275	23.1	21.1	29.0	7.9	19.1
대구		1,011	333	219	282	81	95	33.0	21.7	27.9	8.0	9.4
인천		1,206	187	268	330	72	349	15.5	22.2	27.4	6.0	29.0
광주		624	166	174	203	40	41	26.7	27.9	32.5	6.4	6.6
대전		646	226	128	189	38	65	34.9	19.8	29.3	6.0	10.0
울산		458	134	90	161	25	48	29.3	19.6	35.2	5.4	10.6
세종	동부	101	1	32	57	9	2	0.6	31.4	56.9	9.2	1.8
	읍면부	46	27	7	6	2	5	58.1	15.2	12.3	4.4	9.9
경기	동부	4,530	854	1,017	1,398	335	926	18.9	22.4	30.9	7.4	20.4
	읍면부	935	336	155	196	28	220	35.9	16.6	21.0	3.0	23.5
강원	동부	408	131	116	115	17	29	32.1	28.4	28.2	4.2	7.2
	읍면부	272	167	42	31	2	31	61.3	15.5	11.2	0.8	11.2
충북	동부	406	132	113	112	20	29	32.5	27.8	27.7	5.0	7.1
	읍면부	304	165	57	47	4	33	54.1	18.7	15.3	1.2	10.6
충남	동부	434	117	109	137	25	47	26.9	25.1	31.5	5.7	10.8
	읍면부	507	283	82	77	9	55	55.8	16.2	15.2	1.9	10.9
전북	동부	552	171	155	157	30	38	31.0	28.2	28.4	5.5	6.9
	읍면부	230	173	23	14	0	19	75.4	9.9	6.2	0.1	8.3
전남	동부	330	88	91	111	12	28	26.5	27.6	33.6	3.7	8.6
	읍면부	433	289	52	49	4	39	66.7	12.1	11.3	0.9	9.0
경북	동부	607	202	126	183	29	67	33.2	20.8	30.1	4.8	11.0
	읍면부	564	355	74	76	4	55	62.9	13.1	13.6	0.7	9.8
경남	동부	886	262	203	296	43	82	29.6	22.9	33.4	4.9	9.3
	읍면부	510	272	75	112	9	43	53.2	14.7	21.9	1.8	8.4
제주	동부	197	68	27	34	4	64	34.7	13.6	17.2	2.1	32.3
	읍면부	77	54	2	3	0	19	69.4	2.5	3.6	0.1	24.5

Ⅲ 층화 · 균형변수

1 분석 배경

○ 관심변수와 상관성이 높은 보조변수가 표본추출틀에 존재한다면 이를 고려하여 표본추출함으로써 효율적인 표본을 얻을 수 있음

○ 지난 표본설계에서는 27개 지역을 설계층으로 1차 추출단위(조사구)를 추출하되, 설계층별 추출 시 큐브 방법을 이용한 균형표본추출을 적용

→ 소득·지출과 보조변수 간의 관계를 통계적으로 분석함으로써 기존 보조변수 활용(층화·균형변수)의 적절성을 평가하고 필요한 경우 개선

2 층화변수

○ '21년 조사자료에서 27개 지역별 소득·지출을 분석한 결과, 지역별 평균값의 차이가 통계적으로 유의한 것으로 확인, 층화변수가 유효한 것으로 판단

3 균형변수

① 분석 대상 보조변수

○ '21년 조사자료에서 소득·지출에 유의미한 영향을 주는 가구 특성을 확인한 후, 그와 유사한 조사구 단위 보조변수를 균형변수 선정을 위한 분석에 포함

② 분석을 위한 데이터 전처리

○ **(조사자료·추출틀 연계)** '21년 연간 조사자료의 총 12,391개 응답가구 중 '21년 표본추출틀에 연계 가능한 12,364개 가구(99.8%)를 대상으로 분석

- 주소 오류 등으로 표본추출틀 연계가 불가능한 27개 가구는 분석 대상에서 제외

○ **(로그 변환)** 관심변수인 소득·지출과 보조변수 중 평균 등록경상소득, 중위 공시지가는 오른쪽으로 꼬리가 긴 로그-정규분포에 가까워, 로그 변환 후 분석

○ **(권역)** 조사구의 관할 사무소 코드와 읍면·동 구분 코드를 결합하여 구성

- 설계층별 최대 7개(경기-동부)에서 최소 1개(광역시 및 세종·제주)로, 전국 총 78개

○ **(거처종류)** 조사구별 대표 거처종류를 정하되, 특·광역시 및 도지역 동부는 9~10개 범주로, 도지역 읍면부 및 제주 지역은 5~7개 범주로 병합한 후 결정

< 설계층별 거처종류 범주화 >

<table>
<tr><th colspan="2">거처종류</th><th>특·광역시,
경기/충북/충남/
전북/경남 동부</th><th>세종/강원/전남/
경북 동부</th><th>세종/경기/
충남/경남 읍면부
제주 동부</th><th>강원/충북/
전북/전남/경북/
제주 읍면부</th></tr>
<tr><td rowspan="2">단독</td><td>일반</td><td colspan="4">① 일반 단독주택</td></tr>
<tr><td>다가구, 영업겸용</td><td colspan="4">② 다가구, 영업겸용 단독주택</td></tr>
<tr><td rowspan="2">기타</td><td>연립, 다세대, 기타</td><td colspan="4">③ 기타 주택</td></tr>
<tr><td>주택 이외 거처</td><td colspan="4">④ 기타</td></tr>
<tr><td rowspan="6">아파트</td><td>40㎡ 이하</td><td>⑤ 소형APT1</td><td>⑤ 소형APT1</td><td rowspan="2">⑤ 소형APT</td><td rowspan="2">⑤ 소형APT</td></tr>
<tr><td>40～60㎡ 이하</td><td>⑥ 소형APT2</td><td>⑥ 소형APT2</td></tr>
<tr><td>60～85㎡ 이하</td><td>⑦ 중형APT1</td><td>⑦ 중형APT1</td><td>⑥ 중형APT</td><td rowspan="4">⑥ 중대형APT</td></tr>
<tr><td>85～102㎡ 이하</td><td>⑧ 중형APT2</td><td>⑧ 중형APT2</td><td rowspan="3">⑦ 중대형APT</td></tr>
<tr><td>102～135㎡ 이하</td><td>⑨ 대형APT1</td><td rowspan="2">⑨ 대형APT</td></tr>
<tr><td>135㎡ 초과</td><td>⑩ 대형APT2</td></tr>
</table>

○ **(최종 분석 변수)** 조사된 가구에 대해 총 15개의 조사구 단위 보조변수 연계

< 관심변수 및 보조변수 개요 >

<table>
<tr><th colspan="2">구분</th><th>변수명</th><th>타입</th><th>처리</th></tr>
<tr><td colspan="2" rowspan="2">관심변수
(가구 단위)</td><td>소득</td><td>연속</td><td>로그 변환</td></tr>
<tr><td>가계지출</td><td>연속</td><td>로그 변환</td></tr>
<tr><td rowspan="13">보조변수
(조사구 단위)</td><td rowspan="10">가구
특성</td><td>권역</td><td>범주</td><td>사무소코드 × 읍/면동구분</td></tr>
<tr><td>거처종류</td><td>범주</td><td>설계층별 5～10개 재분류</td></tr>
<tr><td>평균가구원수</td><td>연속</td><td></td></tr>
<tr><td>1인가구비율</td><td>연속</td><td></td></tr>
<tr><td>미취학아동포함가구비율</td><td>연속</td><td></td></tr>
<tr><td>학령기인구포함가구비율</td><td>연속</td><td></td></tr>
<tr><td>고령가구비율</td><td>연속</td><td></td></tr>
<tr><td>자가소유가구비율</td><td>연속</td><td></td></tr>
<tr><td>농림어가비율</td><td>연속</td><td></td></tr>
<tr><td>중위공시지가</td><td>연속</td><td>로그 변환</td></tr>
<tr><td>평균등록경상소득</td><td>연속</td><td>로그 변환</td></tr>
<tr><td rowspan="4">인구
특성</td><td>평균연령</td><td>연속</td><td></td></tr>
<tr><td>30～49세인구비율</td><td>연속</td><td></td></tr>
<tr><td>30～49세대졸이상비율</td><td>연속</td><td></td></tr>
<tr><td>대졸이상비율</td><td>연속</td><td></td></tr>
</table>

③ 전역 균형변수

○ **(방법)** 과거 표본설계에서 전역 균형변수로 활용된 권역과 거처종류에 대해 분산분석을 수행하여 소득·지출과의 관계를 확인

○ **(결과)** 두 변수 모두 전역 균형변수로 유지

- 권역: 대부분의 설계층에서 통계적 유의성을 확인할 수 없었으나, 업무량 배정 등을 고려하여 유지
- 거처종류: 대부분의 설계층에서 통계적 유의성이 확인됨

④ 설계층별 균형변수

○ **(방법)** 회귀분석에서 모형 적합도에 유의한 영향을 주는 독립변수를 설계층별로 분석하고 그 결과를 바탕으로 최대 3개의 균형변수를 선정

- 등록경상소득은 소득·지출과의 높은 상관관계 및 균형변수의 시계열 안정성을 고려하여 특광역시와 시도-동부의 균형변수로 공통 활용
- 전역권역·거처종류 및 특광역시·동부등록경상소득의 활용을 전제(‘include=’ 옵션)로 나머지 변수들에 대해 설계층별 회귀분석 실시
- 전진선택법을 이용해 설계층별 최대 3개(등록경상소득 포함)의 균형변수를 선정

○ **(결과)** 설계층별 균형변수로 최다 선정된 보조변수는 등록경상소득(19회)

< 전역 · 설계층별 균형변수 최종 선정 결과 >

설계층		전역(2개)		설계층별(최대 3개)		
서울		권역	거처종류	평균등록경상소득*	노인가구비율	학령기가구비율
부산		권역	거처종류	평균등록경상소득	평균연령	공시지가
대구		권역	거처종류	평균등록경상소득	노인가구비율	
인천		권역	거처종류	평균등록경상소득	평균연령	
광주		권역	거처종류	평균등록경상소득		
대전		권역	거처종류	평균등록경상소득	1인가구비율	
울산		권역	거처종류	평균등록경상소득	평균가구원수	평균연령
세종	동부	권역	거처종류	평균등록경상소득	3049세대졸비율	
	읍면부	권역	거처종류	평균연령		
경기	동부	권역	거처종류	평균등록경상소득	1인가구비율	학령기가구비율
	읍면부	권역	거처종류	대졸자비율	노인가구비율	학령기가구비율
강원	동부	권역	거처종류	평균등록경상소득	1인가구비율	평균가구원수
	읍면부	권역	거처종류	평균가구원수		
충북	동부	권역	거처종류	평균등록경상소득		
	읍면부	권역	거처종류	평균등록경상소득		
충남	동부	권역	거처종류	평균등록경상소득	평균가구원수	
	읍면부	권역	거처종류	평균등록경상소득		
전북	동부	권역	거처종류	평균등록경상소득		
	읍면부	권역	거처종류	3049세비율	평균연령	
전남	동부	권역	거처종류	평균등록경상소득	평균연령	
	읍면부	권역	거처종류	평균가구원수	학령기가구비율	중위공시지가
경북	동부	권역	거처종류	평균등록경상소득	노인가구비율	평균가구원수
	읍면부	권역	거처종류	평균연령		
경남	동부	권역	거처종류	평균등록경상소득		
	읍면부	권역	거처종류	평균연령		
제주	동부	권역	거처종류	평균등록경상소득	중위공시지가	
	읍면부	권역	거처종류	3049세대졸비율	중위공시지가	

※ 균형표본추출이므로 순서는 의미가 없음

* 특광역시 및 시도 동부에서 공통 균형변수로 활용한 경우 음영 처리

Ⅳ 표본규모

1 전체 표본규모

○ 안정적인 표본체계를 위해 '19년 표본설계 당시 산출한 전체 표본규모 유지

- 전체 표본조사구 규모: 연동그룹당 150개 → 총 6×150 = 900개 조사구

< '19년 표본설계에서의 전체 표본조사구 규모 산정 >

▶ 모든 설계층에 대해 1개 이상의 표본조사구를 배정하기 위한 최소 표본규모 검토 결과, 전체 표본규모가 150개 미만인 경우 세종은 표본이 배정되지 않았음

▶ 모의실험 결과, 표본규모 축소에 따라 모집단과의 특성 변수 오차가 확대되는 것으로 확인되었으며, 연동그룹별 최소표본규모는 150조사구 수준인 것으로 판단

- 목표 가구 규모: 표본조사구당 응답 8가구 → 총 900×8 = 7,200개 가구

< '19년 표본설계에서의 목표 가구 규모 산정 >

▶ 품목별 집락효과($Deff$), 급내상관계수(R_a^2)를 기초로, 통상적인 조사비용의 비율(4:1~6:1)을 고려하면 8가구가 최적으로 판단

설계층		$Deff$	R_a^2	조사비용의 비율(C_{psu} : C_{ssu})별 가구 수			
				4:1	6:1	8:1	10:1
평균		2.09	0.09	7	8	9	10
서울		2.58	0.19	6	7	9	10
부산		2.11	0.10	6	8	9	10
대구		2.49	0.13	6	7	8	9
인천		2.37	0.12	6	7	8	9
광주		2.29	0.11	6	7	8	9
대전		3.07	0.18	4	5	6	7
울산		2.78	0.13	6	7	8	9
세종		1.10	0.00	9	11	13	15
경기	동부	3.04	0.22	5	6	7	8
	읍면부	1.93	0.07	7	8	10	11
강원	동부	1.40	0.04	8	10	11	13
	읍면부	1.35	0.03	7	9	10	12
충북	동부	1.94	0.08	8	9	11	12
	읍면부	1.18	0.01	7	8	10	11
충남	동부	2.70	0.14	6	7	8	9
	읍면부	3.17	0.19	5	6	7	8
전북	동부	2.21	0.11	6	7	8	9
	읍면부	0.64	0.00	11	13	15	17
전남	동부	1.97	0.06	6	7	8	9
	읍면부	1.83	0.07	8	10	12	13
경북	동부	2.66	0.13	5	7	8	8
	읍면부	1.50	0.04	9	11	12	14
경남	동부	1.96	0.09	6	8	9	10
	읍면부	1.79	0.06	7	9	10	12
제주		2.24	0.07	5	6	6	7

2 설계층별 표본조사구 배분

○ **(방법)** 전체 900개 조사구를 시도별로 제곱근 비례배분한 다음, 시도 내에서 동부·읍면부별로 비례배분 실시

- 연동모형을 고려하여 설계층별 표본조사구수가 6의 배수가 되도록 조정

○ **(결과)** 시도 배분에서 경기 지역에 6개 조사구를 추가하는 것이 이상적이었으나, 차감할 지역이 분명하지 않아 현행 유지(※ '25년 8패널 설계 시 재검토)

< 표본조사구 배분 결과 >

(단위: 천 가구, %, 개)

		2023년 6패널		2024년 7패널						
		조사 모집단 가구	최종 배분	조사 모집단 가구	시도별 배분			시도 내 배분		
					제곱근 구성비	제곱근 비례배분	조정 (6배수)	구성비	비례배분	최종 (6배수)
전국		21,407	900	21,933	100.0	900	900	-	-	900
서울		4,115	102	4,178	11.7	105	102	-	-	102
부산		1,417	60	1,442	6.8	62	60	-	-	60
대구		995	54	1,011	5.7	52	54	-	-	54
인천		1,174	54	1,212	6.3	56	54	-	-	54
광주		607	42	624	4.5	41	42	-	-	42
대전		638	42	646	4.6	41	42	-	-	42
울산		452	36	458	3.9	35	36	-	-	36
세종	동부	93	18	101	2.2	20	24	68.6	16	18
	읍면부	48	6	46				31.4	8	6
경기	동부	4,376	96	4,530	13.3	120	114	82.9	95	96
	읍면부	896	18	935				17.1	19	18
강원	동부	398	24	408	4.7	42	42	60.0	25	24
	읍면부	268	18	272				40.0	17	18
충북	동부	395	24	406	4.8	43	42	57.1	24	24
	읍면부	298	18	304				42.9	18	18
충남	동부	422	24	434	5.5	50	48	46.1	22	24
	읍면부	496	24	508				53.9	26	24
전북	동부	535	36	552	5.0	45	48	70.5	34	36
	읍면부	230	12	231				29.5	14	12
전남	동부	322	24	331	5.1	45	48	42.1	20	24
	읍면부	447	24	455				57.9	28	24
경북	동부	591	30	607	6.2	56	54	51.8	28	30
	읍면부	555	24	564				48.2	26	24
경남	동부	863	42	886	6.7	61	60	63.3	38	42
	읍면부	508	18	514				36.7	22	18
제주	동부	192	24	197	3.0	27	30	71.3	21	24
	읍면부	77	6	79				28.7	9	6

3 표본조사구별 최대 · 예비표본규모 산정

○ **(최대표본규모)** 응답 표본 부족*으로 '22년 5패널부터 과대추출 및 가구대체를 도입하여 응답 표본을 추가 확보

* '21년 12월 기준 응답 표본은 4,847가구로, 목표(7,200가구) 대비 약 67.3%

< 5·6패널 최대·예비표본규모 >

설계층		최대	예비	설계층		최대	예비	설계층		최대	예비
서울		17	9	경기	동	15	7	전북	읍면	16	8
부산		15	7		읍면	15	7	전남	동	15	7
대구		15	7	강원	동	15	7		읍면	15	7
인천		16	8		읍면	16	8	경북	동	14	6
광주		15	7	충북	동	15	7		읍면	14	6
대전		14	6		읍면	14	6	경남	동	15	7
울산		16	8	충남	동	18	10		읍면	16	8
세종	동	15	7		읍면	16	8	제주	동	17	9
	읍면	20	12	전북	동	16	8		읍면	16	8

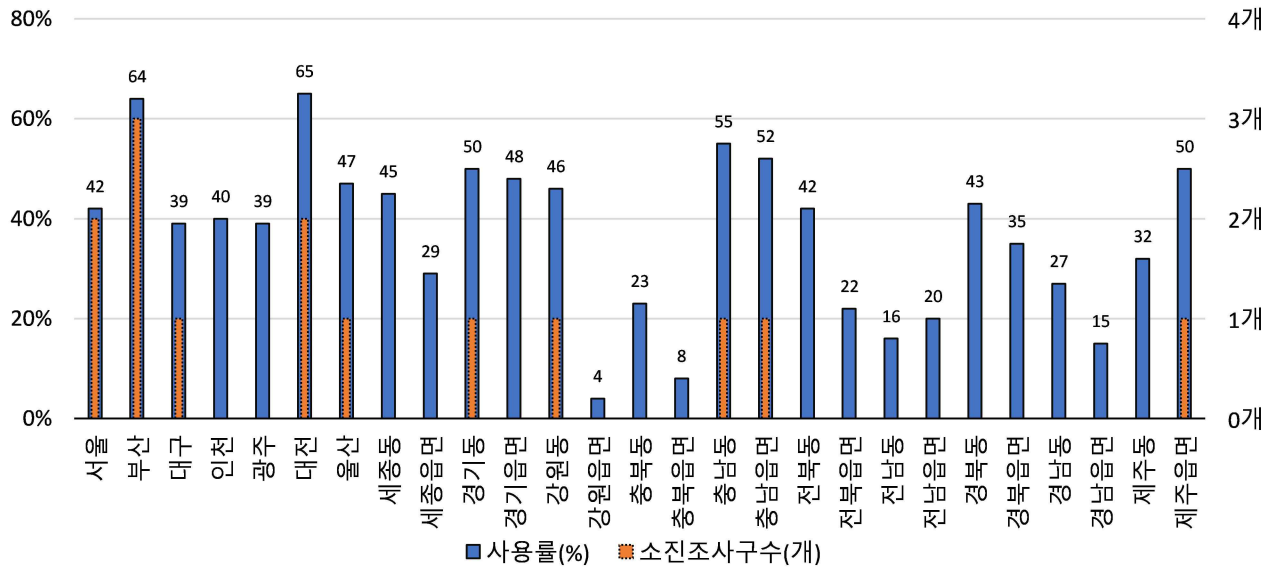
○ **(예비표본 사용 현황)** '23년 6월 기준, 진입 후 6개월차가 되는 5패널 1그룹과 6패널 4그룹의 300개 조사구를 대상으로 예비표본 사용 현황을 분석한 결과,

- 사용률*은 대전(65%), 부산(64%), 충남동(55%) 순으로 높았으며, 소진조사구**수는 부산(3개), 서울·대전(각 2개), 대구·울산 등(각 1개) 순이었음

* 사용률 = (접촉한 예비표본 수) ÷ (예비표본 수)

** 소진조사구 = 예비표본을 모두 소진한 조사구

< 설계층별 예비가구사용 현황 >



○ **(규모 산정)** 예비표본 사용 현황 등을 고려하여 현행 최대·예비표본규모 유지

< 기존 최대·예비표본규모 산정 방식 >

최대표본규모 = 8×적격률$^{-1}$×응답률$^{-1}$ → **예비표본규모** = 최대표본규모－8

- '23년 상반기 응답 현황과 기존 산정 방식을 바탕으로 최대·예비표본규모를 재산정한 결과, 설계층별로 최소 -1에서 최대 +4의 변동이 있었으나,

- 현시점의 사용·소진 수준이 과도하게 높지는 않고, 재산정 적용 시 업무 변동이 수반되는 점 등을 고려하여 현행(5·6패널) 최대·예비표본규모 유지

< 설계층별 최대·예비표본규모 산정 >

(단위: 개, %)

설계층		'23년 상반기 응답 현황			재산정			최종(5·6패널 동일)	
		예비 사용률	적격률	응답률	최대표본 규모	예비표본 규모	*증감*	최대표본 규모	예비표본 규모
서울		42	89	44	21	13	*4*	17	9
부산		64	86	49	19	11	*4*	15	7
대구		39	87	55	17	9	*2*	15	7
인천		40	91	49	18	10	*2*	16	8
광주		39	88	55	17	9	*2*	15	7
대전		65	83	57	17	9	*3*	14	6
울산		47	91	50	18	10	*2*	16	8
세종	동부	45	86	54	18	10	*3*	15	7
	읍면부	29	76	58	19	11	*-1*	20	12
경기	동부	50	91	49	18	10	*3*	15	7
	읍면부	48	77	56	19	11	*4*	15	7
강원	동부	46	88	54	17	9	*2*	15	7
	읍면부	4	81	67	15	7	*-1*	16	8
충북	동부	23	77	66	16	8	*1*	15	7
	읍면부	8	64	79	16	8	*2*	14	6
충남	동부	55	80	49	21	13	*3*	18	10
	읍면부	52	74	62	18	10	*2*	16	8
전북	동부	42	82	57	18	10	*2*	16	8
	읍면부	22	64	73	18	10	*2*	16	8
전남	동부	16	84	65	15	7	*0*	15	7
	읍면부	20	68	75	16	8	*1*	15	7
경북	동부	43	78	66	16	8	*2*	14	6
	읍면부	35	75	70	16	8	*2*	14	6
경남	동부	27	84	57	17	9	*2*	15	7
	읍면부	15	76	67	16	8	*0*	16	8
제주	동부	32	88	55	17	9	*0*	17	9
	읍면부	50	72	62	19	11	*3*	16	8

Ⅴ 표본추출

1 표본조사구 추출

○ **(전체)** 27개 설계층별로 큐브 방법을 이용한 균형표본추출 실시

- 균형변수 총합에 대한 추정치가 모수와 근사적으로 일치하는 표본이 추출되며,
- 50회 추출 후, 15개 보조변수 분포가 모집단과 가장 유사한 표본을 최종 사용

○ **(그룹)** 전체 표본조사구를 대상으로 동일한 추출방법을 적용하여 6개의 상호 동질적인 그룹으로 분할

○ **(사전 대체)** 가구표본관리지침상 부적격 조사구 사유인 '가구수 과소(< 30)', '타조사 중복'에 해당하는 경우 유사성이 가장 높은 조사구로 사전대체 실시

- 타조사 중복은 51개, 가구수 과소는 15개로, 총 66개(7.3%)가 대체됨

○ **(분포 점검)** 15개 보조변수에 대해 표본추출틀과 전체·그룹별·설계층별 분포를 비교하여 표본의 대표성을 최종 확인

2 표본가구 선정

○ **(표본범위)** 표본조사구별로 시작가구를 단순임의추출함으로써 표본가구 결정

- 인구총조사 가구명부 상 시작가구로부터 적격 8가구가 되는 범위까지를 원표본(1·2구역, 구역 당 적격 4가구)으로 설정하고,
- 원표본 이후의 첫 번째 가구부터 설계층별 예비표본규모를 만족시키는 범위까지를 예비표본으로 설정

○ **(가구대체)** 구역 확정 당시 적격가구가 본조사에서 부적격으로 전환되거나 무응답인 경우, 설정된 예비표본 내에서 가구대체 실시

Ⅵ 가중값 및 추정

1 가중값

① 설계가중값

○ 설계층별 조사구추출률 및 가구추출률의 역수로 구한 기초가중값에 조사구별 예비표본조사비율의 역수를 곱해 산출

설계가중값 (h층 i조사구 j가구)	$w_{hij}^{b} = \frac{M_h}{n_h M_{hi}} \times \frac{M_{hi}}{m_{hi}} \times \frac{m_{hi}}{m_{hi}^{s}} = \frac{M_h}{n_h m_{hi}} \times \frac{m_{hi}}{m_{hi}^{s}}$
M_h	: h층의 인총 전수 조사구의 가구 수
M_{hi}	: h층 i조사구의 전체 가구 수
n_h	: h층의 표본 조사구 수
m_{hi}	: h층 i조사구 내 표본 가구 수(예비표본 포함한 최대표본규모)
m_{hi}^{s}	: h층 i조사구 내 조사 가구 수(원표본+예비표본 중 조사한 가구 수)

② 무응답가중값

○ 응답 성향을 고려하여 설정한 무응답층별로 응답률을 추정한 후, 해당 값의 역수를 설계가중값에 곱해 산출

무응답가중값 (g무응답층 h층 i조사구 j가구)	$w_{g,hij}^{r} = w_{hij}^{b} \times \frac{\sum_{\in (s \cap g)} w_{hij}^{b}}{\sum_{\in (r \cap g)} w_{hij}^{b}}$
w_{hij}^{b}	: h층 i조사구 j가구의 설계가중값
s	: 전체 표본
r	: 응답 표본

③ 사후조정가중값

○ 보조변수 주변합에 대한 추정값을 반복비례가중법으로 모수에 수렴시켜 산출

m월 사후가중값 (h사후층 i조사구 j가구)	$w_{m,hij} = w_{hij}^{T}$
w_i^{F}	: 표본 i의 최종 가중값
$w_i^{(T)}$	: 표본 i의 T번째 반복비례 가중값(제약식과 수렴 조건 충족)

제약식	$N_{a..} = \hat{N}_{a..} \left(= \sum_{b=1}^{56} \sum_{c=1}^{10} \hat{N}_{abc} = \Sigma_{\in (r \cap a)} w_{hij} \right)$
	$N_{.b.} = \hat{N}_{.b.} \left(= \sum_{a=1}^{68} \sum_{c=1}^{10} \hat{N}_{abc} = \Sigma_{\in (r \cap b)} w_{hij} \right)$
	$N_{..c} = \hat{N}_{..c} \left(= \sum_{a=1}^{68} \sum_{b=1}^{56} \hat{N}_{abc} \equiv \Sigma_{\in (r \cap c)} w_{hij} \right)$
a, b, c	: 벤치마크 차원(변수), $a=1, \cdots, 68$, $b=1, \cdots, 56$, $c=1, \cdots, 10$
N_{abc}	: 벤치마크 그룹(a, b, c) 모집단 가구수
$\hat{N}_{abc}$	: 벤치마크 그룹(a, b, c) 모집단 가구수에 대한 추정값
r	: 응답 표본
w_{hij}	: h층 i조사구 j가구의 가중값

수렴조건	$\sum_{a=1}^{68} \left\| \hat{N}_{a..}^{(t)} - N_{a..} \right\| + \sum_{b=1}^{56} \left\| \hat{N}_{.b.}^{(t)} - N_{.b.} \right\| + \sum_{c=1}^{10} \left\| \hat{N}_{..c}^{(t)} - N_{..c} \right\| < \varepsilon$
	※ 수렴 조건을 만족할 때까지 반복 수행
$\hat{N}_{abc}^{(t)}$	: 그룹(a, b, c)의 모집단 가구수에 대한 t번째 추정값
ε	: 임계값(= 0.000001)

반복비례가중값 (t번째)	$w_{hij}^{(t,1)} = \begin{cases} w_{hij}^{(0)} \times \dfrac{N_{a..}}{\Sigma_{\in (r \cap a)} w_{hij}^{(0)}}, & for\ t=1 \\ w_{hij}^{(t-1,3)} \times \dfrac{N_{a..}}{\Sigma_{\in (r \cap a)} w_{hij}^{(t-1,3)}}, & for\ t \geq 2 \end{cases}$
	$w_{hij}^{(t,2)} = w_{hij}^{(t,1)} \times \dfrac{N_{.b.}}{\Sigma_{\in (r \cap b)} w_{hij}^{(t,1)}}$
	$w_{hij}^{(t,3)} = w_{hij}^{(t,2)} \times \dfrac{N_{..c}}{\Sigma_{\in (r \cap c)} w_{hij}^{(t,2)}}$
$w_{hij}^{(0)}$	: h층 i조사구 j가구의 무응답 가중값
$w_{hij}^{(t,n)}$	: t번째 반복에서 n번째 차원의 벤치마크 가중값

④ 분기가중값

○ (월별 사후가중값 합÷3)을 반복비례가중법으로 보정

q분기 가중값 (h층 i조사구 j가구)	$w_{q,hij}^{F} = w_{q,hij}^{T}$
$w_{q,hij}$	: 초기가중값($=\frac{1}{3} \sum_{m=q\times3-2}^{q\times3} w_{m,hij}$)
$w_{q,hij}^{T}$	: 제약식과 수렴조건을 만족하는 T번째 반복비례가중값

2 추정

○ (평균) 호르비츠-톰슨(Horvitz-Thompson, HT) 추정량

평균 추정량

$$\hat{\mu} = \bar{y} = \frac{\sum_{h}^{H}\sum_{i}^{n_h}\sum_{j}^{m_{hi}} w_{hij}\, y_{hij}}{\sum_{h}^{H}\sum_{i}^{n_h}\sum_{j}^{m_{hi}} w_{hij}}$$

· w_{hij} : h층 i조사구 j가구의 최종 가중값(분기/연간)

· y_{hij} : h층 i조사구 j가구의 관측값(소득/지출)

○ (분산) 테일러(Taylor) 선형근사 추정량

분산 추정량

$$\widehat{Var}(\bar{y}) = \sum_{h=1}^{H} \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi\cdot} - \bar{e}_{h\cdot\cdot})^2$$

· $e_{hi\cdot} = \left(\sum_{j=1}^{m_{hi}} w_{hij}\,(y_{hij} - \bar{y})\right) / w_{\cdot\cdot\cdot}$

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

· $w_{\cdot\cdot\cdot} = \sum_{h}^{H}\sum_{i}^{n_h}\sum_{j}^{m_{hi}} w_{hij}$

○ 표준오차(SE) 및 상대표준오차(RSE)

표준오차 $\widehat{SE}(\bar{y}) = \sqrt{\widehat{Var}(\bar{y})}$

상대표준오차 $\widehat{RSE}(\bar{y}) = \dfrac{\widehat{SE}(\bar{y})}{\bar{y}} \times 100(\%)$